Online Nonsubmodular Minimization with Delayed Costs: From Full Information to Bandit Feedback

Tianyi Lin [*1] **Aldo Pacchiano** [*2] **Yaodong Yu** [*1] **Michael. I. Jordan** [1 3]

Abstract

Motivated by applications to online learning in sparse estimation and Bayesian optimization, we consider the problem of online unconstrained nonsubmodular minimization with delayed costs in both full information and bandit feedback settings. In contrast to previous works on online unconstrained submodular minimization, we focus on a class of nonsubmodular functions with special structure, and prove regret guarantees for several variants of the online and approximate online bandit gradient descent algorithms in static and delayed scenarios. We derive bounds for the agent's regret in the full information and bandit feedback setting, even if the delay between choosing a decision and receiving the incurred cost is unbounded. Key to our approach is the notion of (α, β)-regret and the extension of the generic convex relaxation model from El Halabi & Jegelka (2020), the analysis of which is of independent interest. We conduct and showcase several simulation studies to demonstrate the efficacy of our algorithms.

1. Introduction

With machine learning systems increasingly being deployed in real-world settings, there is an urgent need for online learning algorithms that can minimize cumulative costs over the long run, even in the face of complete uncertainty about future outcomes. There exist a myriad of works that deal with this setting, most prominently in the area of online learning and bandits (Cesa-Bianchi & Lugosi, 2006; Lattimore & Szepesvári, 2020). The majority of this literature deals with problems where the decisions are taken from either a small set (such as in the multi armed bandit framework (Auer, 2002)), a continuous decision space (as in linear bandits (Auer, 2002; Dani et al., 2008)) or in the case the decision set is combinatorial in nature, the response is often assumed to maintain a simple functional relationship with the input (e.g., linear (Cesa-Bianchi & Lugosi, 2012)).

In this paper, we depart from these assumptions and explore what we believe is a more realistic type of model for the setting where the actions can be encoded as selecting a subset of a universe of size n. We study a sequential interaction between an agent and the world that takes place in rounds. At the beginning of round t, the agent chooses a *subset* $S^t \subseteq [n]$ (e.g., selecting the set of products in a factory (McCormick, 2005)), after which the agent suffers cost $f_t(S^t)$ such that f_t is an $\alpha-$weakly DR-submodular and $\beta-$weakly DL-supermodular function (Lehmann et al., 2006). The agent then may receive extra information about f_t as feedback, for example in the full information setting the agent observes the whole function f_t and in the bandit feedback scenario the learner does not receive any extra information about f_t beyond the value of $f_t(S^t)$. The standard metric to measure an online learning algorithm is *regret* (Blum & Mansour, 2007): the regret at time T is the difference between $\sum_{t=1}^T f_t(S^t)$ that is the total cost achieved by the algorithm and $\min_{x \in A} \sum_{t=1}^T f_t(x)$ that is the total cost achieved by the best fixed action in hindsight. A *no-regret* learning algorithm is one that achieves sublinear regret (as a function of T). Many no-regret learning algorithms have been developed based on online convex optimization toolbox (Zinkevich, 2003; Kalai & Vempala, 2005; Shalev-Shwartz & Singer, 2006; Hazan et al., 2007; Shalev-Shwartz, 2011; Arora et al., 2012; Hazan, 2016) many of them achieving minimax-optimal regret bounds for different cost functions even when these are produced by the world in an adversarial fashion. However, many online decision-making problems remain open, for example when the decision space is discrete and large (e.g., exponential in the number of problem parameters) and the cost functions are nonlinear (Hazan & Kale, 2012).

To the best of our knowledge, Hazan & Kale (2012) were the first to investigate non-parametric online learning in combinatorial domains by considering the setting where the costs f_t are all submodular functions. In this formulation the decision space is the set of all subsets of a set of

[*]Equal contribution [1]Department of Electrical Engineering and Computer Science, UC Berkeley [2]Microsoft Research, NYC [3]Department of Statistics, UC Berkeley. Correspondence to: Tianyi Lin <darren_lin@berkeley.edu>.

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

n elements; and the cost functions are *submodular*. They provided no-regret algorithms for both the full information and bandit settings. Their chief innovation was to propose a computationally efficient algorithm for online submodular learning that resolved the exponential computational and statistical dependence on n suffered by all previous approaches (Hazan & Kale, 2012). These results served as a catalyst for a rich and expanding research area (Streeter & Golovin, 2008; Jegelka & Bilmes, 2011; Buchbinder et al., 2014; Chen et al., 2018c; Roughgarden & Wang, 2018; Chen et al., 2018b; Cardoso & Cummings, 2019; Anari et al., 2019; Harvey et al., 2020; Thang & Srivastav, 2021; Matsuoka et al., 2021).

Even though submodularity can be used to model a few important typical cost functions that arise in machine learning problems (Boykov et al., 2001; Boykov & Kolmogorov, 2004; Narasimhan et al., 2005; Bach, 2010), it is an insufficient assumption for many other applications where the cost functions do not satisfy submodularity, e.g., structured sparse learning (El Halabi & Cevher, 2015), batch Bayesian optimization (González et al., 2016; Bogunovic et al., 2016), Bayesian A-optimal experimental design (Bian et al., 2017), column subset selection (Sviridenko et al., 2017) and so on. In this work we aim to fill in this gap. In view of all this, we consider the following question:

Can we design online learning algorithms when the cost functions are nonsubmodular?

This paper provides an affirmative answer to this question by demonstrating that online/bandit approximate gradient descent algorithm can be directly extended from online submodular minimization (Hazan & Kale, 2012) to online nonsubmodular minimization when each cost functions f_t satisfy the regularity condition in El Halabi & Jegelka (2020).

Moreover, in online decision-making there is often a significant delay between decision and feedback. This delay has an adverse effect on the characterization between marketing feedback and an agent's decision (Quanrud & Khashabi, 2015; Héliou et al., 2020). For example, a click on an ad can be observed within seconds of the ad being displayed, but the corresponding sale can take hours or days to occur. We extend all of our algorithms to the delayed feedback setting by leveraging a pooling strategy recently introduced by Héliou et al. (2020) into the framework of online/bandit approximate gradient descent.

Contribution. First, we introduce a new notion of (α, β)-regret which allows for analyzing no-regret online learning algorithms when the loss functions are nonsubmodular. We then propose two randomized algorithms for both the full-information and bandit feedback settings respectively with the regret bounds in expectation and high-probability sense. We then combine the aforementioned algorithms with the pooling strategy found in (Héliou et al., 2020) and prove that the resulting algorithms are no-regret even when the delays are unbounded (cf. Assumption 5.1). Specifically, when the delay d_t satisfies $d_t = o(t^\gamma)$, we establish a $O(\sqrt{nT^{1+\gamma}})$ regret bound in full-information setting and a $O(nT^{\frac{2+\gamma}{3}})$ regret bound in bandit feedback setting. To our knowledge, this is the first theoretical guarantee for no-regret learning in online nonsubmodular minimization with delayed costs. Experimental results on sparse learning with synthetic data confirm our theoretical findings.

It is worth comparing our results with that in the existing works (El Halabi & Jegelka, 2020; Hazan & Kale, 2012; Héliou et al., 2020). First of all, the results concerning online nonsubmodular minimization are not a straightforward consequence of El Halabi & Jegelka (2020). Indeed, it is natural yet nontrivial to identify the notion of (α, β)-regret under which formal guarantees can be established for the nonsubmodular case. This notion does not appear before and appears to be a novel idea and an interesting conceptual contribution. Further, our results provide the first theoretical guarantee for no-regret learning in online and bandit nonsubmodular minimization and generalize the results in Hazan & Kale (2012). Even though the online and bandit learning algorithms and regret analysis share the similar spirits with the context of Hazan & Kale (2012), the proof techniques are different since we need to deal with the nonsubmodular case with (α, β)-regret. Finally, we are not aware of any results on online and bandit combinatorial optimization with delayed costs. Héliou et al. (2020) focused on the gradient-free game-theoretical learning with delayed costs where the action sets are *continuous* and *bounded*. Thus, their results can not imply ours. The only component that two works share is the pooling strategy which has been a common algorithmic component to handle the delays. Even though the pooling strategy is crucial to our delayed algorithms, we make much efforts to combine them properly and prove (α, β)-regret bound of our new algorithms.

Notation. We let $[n]$ be the set $\{1, 2, \ldots, n\}$ and $\mathbb{R}^n_+$ be the set of all vectors in $\mathbb{R}^n$ with nonnegative components. We denote $2^{[n]}$ as the set of all subsets of $[n]$. For a set $S \subseteq [n]$, we let $\chi_S \in \{0, 1\}^n$ be the characteristic vector satisfying that $\chi_S(i) = 1$ for each $i \in S$ and $\chi_S(i) = 0$ for each $i \notin S$. For a function $f : 2^{[n]} \mapsto \mathbb{R}$, we denote the marginal gain of adding an element i to S by $f(i \mid S) = f(S \cup \{i\}) - f(S)$. In addition, f is normalized if $f(\emptyset) = 0$ and nondecreasing if $f(A) \leq f(B)$ for $A \subseteq B$. For a vector $x \in \mathbb{R}^n$, its Euclidean norm refers to $\|x\|$ and its i-th entry refers to x_i. We denote the support set of x by $\text{supp}(x) = \{i \in [n] : x_i \neq 0\}$ and, by abuse of notation, we let x define a set function $x(S) = \sum_{i \in S} x_i$. We let P_S be the projection onto a closed set S and $\text{dist}(x, S) = \inf_{y \in S} \|x - y\|$ denotes the distance between x and S. A pair of parameters $(\alpha, \beta) \in \mathbb{R}_+ \times \mathbb{R}_+$ in the regret refer to approximation factors of the corresponding offline setting. Lastly, $a = O(b(\alpha, \beta, n, T))$ refers to an upper bound $a \leq C \cdot b(\alpha, \beta, n, T)$ where $C > 0$ is independent of α, β, n and T.

2. Related Work

The offline nonsubmodular optimization with different notions of approximate submodularity has recently received a lot of attention. Most research focused on the maximization of nonsubmodular set functions, emerging as an important paradigm for studying real-world application problems (Das & Kempe, 2011; Horel & Singer, 2016; Chen et al., 2018a; Kuhnle et al., 2018; Hassidim & Singer, 2018; Elenberg et al., 2018; Harshaw et al., 2019). In contrast, we are aware of relatively few investigations into the minimization of nonsubmodular set functions. An interesting example is the ratio problem (Bai et al., 2016) where the objective function to be minimized is the ratio of two set functions and is thus nonsubmodular in general. Note that the ratio problem does not admit a constant factor approximation even when two set functions are submodular (Svitkina & Fleischer, 2011). However, if the objective function to be minimized is approximately modular with bounded curvature, the optimal approximation algorithms exist even when the constrain sets are assumed (Iyer et al., 2013). Another typical example is the minimization of the difference of two submodular functions, where some approximation algorithms were proposed in Iyer & Bilmes (2012) and Kawahara et al. (2015) but without any approximation guarantee. Very recently, El Halabi & Jegelka (2020) provided a comprehensive treatment of optimal approximation guarantees for minimizing nonsubmodular set functions, characterized by how close the function is to submodular. Our work is close to theirs and our results can be interpreted as the extension of El Halabi & Jegelka (2020) to online learning with delayed feedback.

Another line of relevant works comes from online learning literature and focuses on no-regret algorithms in different settings with delayed costs. In the context of online convex optimization, Quanrud & Khashabi (2015) proposed an extension of online gradient descent (OGD) where the agent performs a batched gradient update the moment gradients are received and proved that OGD achieved a regret bound of $O(\sqrt{T + D_T})$ where D_T is the total delay over a horizon T. However, their batch update approach can not be extended to bandit convex optimization since it does not work with stochastic estimates of the received gradient information (or when attempting to infer such information from realized costs). This issue was posted by Zhou et al. (2017) and recently resolved by Héliou et al. (2020) who proposed a new pooling strategy based on a priority queue. The effect of delay was also discussed in the multi-armed bandit (MAB) literature under different assumptions (Joulani et al., 2013; 2016; Vernade et al., 2017; Pike-Burke et al., 2018; Thune et al., 2019; Bistritz et al., 2019; Zhou et al., 2019; Zimmert & Seldin, 2020; Gyorgy & Joulani, 2021). In particular, Thune et al. (2019) proved the regret bound in adversarial MABs with the cumulative delay and Gyorgy & Joulani (2021) studied the adaptive tuning to delays and data in this setting. Further, Joulani et al. (2016) and Zimmert & Seldin (2020) also investigated adaptive tuning to the unknown sum of delays while Bistritz et al. (2019) and Zhou et al. (2019) gave further results in adversarial and linear contextual bandits respectively. However, the algorithms developed in the aforementioned works have little to do with online nonsubmodular minimization with delayed costs.

3. Preliminaries and Technical Background

We present the basic setup for minimizing structured nonsubmodular functions, including motivating examples and convex relaxation based on Lovász extension. We extend the offline setting to online setting and (α, β)-regret which is important to the subsequent analysis.

3.1. Structured nonsubmodular function

Minimizing a set function $f : 2^{[n]} \mapsto \mathbb{R}$ is NP-hard in general but is solved exactly with *submodular* structure in polynomial time (Iwata, 2003; Grötschel et al., 2012; Lee et al., 2015) and in strongly polynomial time (Schrijver, 2000; Iwata et al., 2001; Iwata & Orlin, 2009; Orlin, 2009; Lee et al., 2015). More specifically, f is submodular if it satisfies the diminishing returns (DR) property as follows,

$$f(i \mid A) \geq f(i \mid B), \quad \text{for all } A \subseteq B, \ i \in [n] \setminus B. \quad (1)$$

Further, f is modular if the inequality in Eq. (1) holds as an equality and is supermodular if

$$f(i \mid B) \geq f(i \mid A), \quad \text{for all } A \subseteq B, \ i \in [n] \setminus B.$$

Relaxing these inequalities will bring us the notions of weak DR-submodularity/supermodularity that were introduced by Lehmann et al. (2006) and revisited in the machine learning literature (Bian et al., 2017). Formally, we have

Definition 3.1 *A set function $f : 2^{[n]} \mapsto \mathbb{R}$ is α-weakly DR-submodular with $\alpha > 0$ if*

$$f(i \mid A) \geq \alpha f(i \mid B), \quad \text{for all } A \subseteq B, \ i \in [n] \setminus B.$$

Similarly, f is β-weakly DR-supermodular with $\beta > 0$ if

$$f(i \mid B) \geq \beta f(i \mid A), \quad \text{for all } A \subseteq B, \ i \in [n] \setminus B.$$

We say that f is (α, β)-weakly DR-modular if both of the above two inequalities hold true.

The above notions of weak DR-submodularity (or weak DR-supermodularity) generalize the notions of submodularity (or supermodularity); indeed, we have f is submodular (or supermodular) if and only if $\alpha = 1$ (or $\beta = 1$). They are also special cases of more general notions of weak submodularity (or weak supermodularity) (Das & Kempe, 2011) and we refer to Bogunovic et al. (2018, Proposition 1) and El Halabi et al. (2018, Proposition 8) for the details. For an overview of the approximate submodularity, we refer to Bian et al. (2017, Section 6) and El Halabi & Jegelka (2020, Figure 1). In addition, the parameters $1 - \alpha$ and $1 - \beta$ are referred to as *generalized inverse curvature* and *generalized curvature* respectively (Bian et al., 2017; Bogunovic et al., 2018) and can be interpreted as the extension of inverse curvature and curvature (Conforti & Cornuéjols, 1984) for submodular and supermodular functions. Intuitively, these parameters quantify how far the function f is from being a submodular (or supermodular) function.

Recently, El Halabi & Jegelka (2020) have proposed and studied the problem of minimizing a class of structured nonsubmodular functions as follows,

$$\min_{S \subseteq [n]} f(S) := \bar{f}(S) - \underline{f}(S), \quad (2)$$

where $\bar{f}$ and $\underline{f}$ are both normalized (i.e., $\bar{f}(\emptyset) = \underline{f}(\emptyset) = 0$)[1] and nondecreasing, $\bar{f}$ is α-weakly DR-submodular and $\underline{f}$ is β-weakly DR-supermodular. Note that the problem in Eq. (2) is challenging; indeed, f is neither weakly DR-submodular nor weakly DR-supermodular in general since the weak DR-submodularity (or weak DR-supermodularity) are only valid for monotone functions.

It is worth mentioning that Eq. (2) is not necessarily theoretically artificial but encompasses a wide range of applications. We present two typical examples which can be formulated in the form of Eq. (2) and refer to El Halabi & Jegelka (2020, Section 4) for more details.

Example 3.1 (Structured Sparse Learning) *We aim to estimate a sparse parameter vector whose support satisfies a particular structure and commonly formulate such problems as* $\min_{x \in \mathbb{R}^n} \ell(x) + \lambda f(\text{supp}(x))$*, where* $\ell : \mathbb{R}^n \mapsto \mathbb{R}$ *is a loss function and* $f : 2^{[n]} \mapsto \mathbb{R}$ *is a set function favoring the desirable supports. Existing approaches such as (Bach, 2010) proposed to replace the discrete regularization function* $f(\text{supp}(x))$ *by its closest convex relaxation and is computationally tractable only when* f *is submodular. However, this problem is often better modeled by a nonsubmodular regularizer in practice (El Halabi & Cevher, 2015). An alternative formulation of structured sparse learning problems is*

$$\min_{S \subseteq [n]} \lambda f(S) - h(S), \quad (3)$$

where $h(S) = \ell(0) - \min_{\text{supp}(x) \subseteq S} \ell(x)$*. Note that Eq. (3) can be reformulated into the form of Eq. (2) under certain conditions; indeed,* h *is a normalized and nondecreasing function and El Halabi & Jegelka (2020, Proposition 5) has shown that* h *is weakly DR-modular if* ℓ *is smooth, strongly convex and is generated from random data. Examples of weakly DR-submodular regularizers* f *include the ones used in time-series and cancer diagnosis (Rapaport et al., 2008) and healthcare (Sakaue, 2019).*

Example 3.2 (Batch Bayesian Optimization) *We aim to optimize an unknown expensive-to-evaluate noisy function* ℓ *with as few batches of function evaluations as possible. The evaluation points are chosen to maximize an acquisition function – the variance reduction function (González et al., 2016) – subject to a cardinality constraint. Maximizing the variance reduction may be phrased as a special instance of the problems in Eq. (2) in the form of* $\min_{S \subseteq [n]} \lambda |S| - G(S)$*, where* $G : 2^{[n]} \mapsto \mathbb{R}$ *is the variance reduction function defined accordingly and El Halabi & Jegelka (2020, Proposition 6) has shown that it is also non-decreasing and weakly DR-modular. This formulation allows to include nonlinear costs with (weak) decrease in marginal costs (economies of scale) with some applications in the sensor placement.*

3.2. Convex relaxation based on the Lovász extension

The Lovász extension (Lovász, 1983) is a toolbox commonly used for minimizing a submodular set function $f : 2^{[n]} \mapsto \mathbb{R}$. It is a continuous interpolation of f on the unit hypercube $[0, 1]^n$ and can be minimized efficiently since it is *convex* if and only if f is submodular. The minima of the Lovász extension also recover the minima of f.

Before the formal argument, we define a maximal chain of $[n]$; that is, $\{A_0, \ldots, A_n\}$ is a maximal chain if $\emptyset = A_0 \subseteq A_1 \subseteq \ldots \subseteq A_n = [n]$. Formally, we have

Definition 3.2 *Given a submodular function* f*, the Lovász extension is the function* $f_L : [0, 1]^n \mapsto \mathbb{R}$ *given by* $f_L(x) = \sum_{i=0}^n \lambda_i f(A_i)$ *where* $\{A_0, \ldots, A_n\}$ *is a maximal chain*[2] *of* $[n]$ *so that* $\sum_{i=0}^n \lambda_i \chi_{A_i} = x$ *and* $\sum_{i=0}^n \lambda_i = 1$ *where* $\chi_{A_i}(j) = 1$ *for* $\forall j \in A_i$ *and* $\chi_{A_i}(j) = 0$ *for* $\forall j \notin S$.

Even though Definition 3.2 implies that $f_L(\chi_S) = f(S)$ for all $S \subseteq [n]$, it remains unclear how to find the chain or the coefficients. The preceding discussion defines the Lovász extension in an equivalent way that is more amenable for computing the subgradient of f_L.

Let $x = (x_1, x_2, \ldots, x_n) \in [0, 1]^n$ and we define that $\pi : [n] \mapsto [n]$ is the sorting permutation of $\{x_1, x_2, \ldots, x_n\}$ where $\pi(i) = j$ implies that x_j is the i-th largest element. By definition, we have $1 \geq x_{\pi(1)} \geq \ldots \geq x_{\pi(n)} \geq 0$ and let $x_{\pi(0)} = 1$ and $x_{\pi(n+1)} = 0$ for simplicity. Then, we set

[1]In general, we can let $\bar{f}(S) \leftarrow \bar{f}(S) - \bar{f}(\emptyset)$ and $\underline{f}(S) \leftarrow \underline{f}(S) - \underline{f}(\emptyset)$ which will not change the minimization problem.

[2]Both the chain and the set of λ_i may depend on the input x.

$\lambda_i = x_{\pi(i)} - x_{\pi(i+1)}$ for all $0 \leq i \leq n$ and let $A_0 = \emptyset$ and $A_i = \{\pi(1), \ldots, \pi(i)\}$ for all $i \in [n]$. We also have

$$\sum_{i=0}^{n} \lambda_i \chi_{A_i} = \sum_{i=0}^{n} (x_{\pi(i)} - x_{\pi(i+1)})(\chi_{A_{i-1}} + e_{\pi(i)}) = \sum_{i=1}^{n} e_{\pi(i)} \sum_{j=i}^{n} (x_{\pi(j)} - x_{\pi(j+1)}) = x.$$

As such, we obtain that $f_L(x) = \sum_{i=1}^{n} x_{\pi(i)} f(\pi(i) \mid A_{i-1})$ where $x_{\pi(1)} \geq x_{\pi(2)} \geq \ldots \geq x_{\pi(n)}$ are the sorted entries in decreasing order, $A_0 = \emptyset$ and $A_i = \{\pi(1), \ldots, \pi(i)\}$ for all $i \in [n]$. Then, the classical results (Edmonds, 2003; Fujishige, 2005) suggest that the subgradient g of f_L at any $x \in [0,1]^n$ can be computed by simply sorting the entries in decreasing order and taking

$$g_{\pi(i)} = f(A_i) - f(A_{i-1}), \text{ for all } i \in [n]. \tag{4}$$

Since f_L is convex if and only if f is submodular, we can apply the convex optimization toolbox here. Recently, El Halabi & Jegelka (2020) have shown that the similar idea can be extended to nonsubmodular optimization in Eq. (2).

More specifically, we can define the convex closure f_C for any nonsubmodular function f; indeed, $f_C : [0,1]^n \mapsto \mathbb{R}$ is the point-wise largest convex function which always lower bounds f. By definition, f_C is the *tightest* convex extension of f and $\min_{S \subseteq [n]} f(S) = \min_{x \in [0,1]^n} f_C(x)$. In general, it is NP-hard to evaluate and optimize f_C (Vondrák, 2007). Fortunately, El Halabi & Jegelka (2020) demonstrated that the Lovász extension f_L approximates f_C such that the vector computed using the approach in Edmonds (2003) and Fujishige (2005) approximates the subgradient of f_C. We summarize their results in the following proposition and provide the proofs in Appendix A for completeness.

Proposition 3.1 *Focusing on Eq. (2), we let* $x \in [0,1]^n$ *with* $x_{\pi(1)} \geq \ldots \geq x_{\pi(n)}$ *and* $g_{\pi(i)} = f(A_i) - f(A_{i-1})$ *for all* $i \in [n]$ *where* $A_0 = \emptyset$ *and* $A_i = \{\pi(1), \ldots, \pi(i)\}$ *for all* $i \in [n]$*. Then, we have*

$$f_L(x) = g^\top x \geq f_C(x), \tag{5}$$

and

$$g(A) = \sum_{i \in A} g_i \leq \tfrac{1}{\alpha} \bar{f}(A) - \beta \underline{f}(A), \text{ for all } A \subseteq [n], \tag{6}$$

and

$$g^\top z \leq \tfrac{1}{\alpha} \bar{f}_C(z) + \beta(-\underline{f})_C(z), \text{ for all } z \in [0,1]^n. \tag{7}$$

Proposition 3.1 highlights how f_L approximates f_C; indeed, we see from Eq. (5) and Eq. (7) that $f_C(x) \leq f_L(x) \leq \frac{1}{\alpha} \bar{f}_C(x) + \beta(-\underline{f})_C(x)$ for all $x \in [0,1]^n$. As such, it gives the key insight for analyzing the offline algorithms in El Halabi & Jegelka (2020) and will play an important role in the subsequent analysis of our paper.

3.3. Online nonsubmodular minimization

We consider online nonsubmodular minimization which extends the offline problem in Eq. (2) to the online setting. In particular, an adversary first chooses structured nonsubmodular functions $f_1, f_2, \ldots, f_T : 2^{[n]} \mapsto \mathbb{R}$ given by

$$f_t(S) := \bar{f}_t(S) - \underline{f}_t(S), \text{ for all } S \subseteq [n], \ t \in [T], \tag{8}$$

where $\bar{f}_t$ and $\underline{f}_t$ are normalized and non-decreasing, $\bar{f}_t$ is α-weakly DR-submodular and $\underline{f}_t$ is β-weakly DR-supermodular. In each round $t = 1, 2, \ldots, T$, the agent chooses S^t and observes the incurred loss $f_t(S^t)$ after committing to her decision. Throughout the horizon $[0, T]$, one aims to minimize the regret – the difference between $\sum_{t=1}^{T} f_t(S^t)$ and the loss at the best fixed solution in hindsight, i.e., $S_\star^T = \text{argmin}_{S \subseteq [n]} \sum_{t=1}^{T} f_t(S)$ – which is defined by[3]

$$R(T) = \sum_{t=1}^{T} f_t(S^t) - \sum_{t=1}^{T} f_t(S_\star^T). \tag{9}$$

An algorithm is *no-regret* if $R(T)/T \to 0$ as $T \to +\infty$ and *efficient* if it computes each decision set S^t in polynomial time. In the context, the regret is used when the minimization for a known cost, i.e., $\min_{S \subseteq [n]} f(S)$, can be solved exactly. However, solving the optimization problem in Eq. (2) with nonsubmodular costs is NP-hard regardless of any multiplicative constant factor (Iyer & Bilmes, 2012; Trevisan, 2014). Thus, it is necessary to consider a bicriteria-like approximation guarantee with the factors $\alpha, \beta > 0$ as El Halabi & Jegelka (2020) suggested. In particular, (α, β) are bounds on the quality of a solution S returned by a given offline algorithm compared to the optimal solution $\bar{S}$; that is, $f(S) \leq \frac{1}{\alpha} \bar{f}(S_\star) - \beta \underline{f}(S_\star)$. Such bicriteria-like approximation is optimal: El Halabi & Jegelka (2020, Theorem 2) has shown that no algorithm with subexponential number of value queries can improve on it in the oracle model.

Our goal is to analyze *online approximate gradient descent algorithm and its bandit variant* for online nonsubmodular minimization. Let (α, β) be the approximation factors attained by an offline algorithm that solves $\min_{S \subseteq [n]} f(S)$ for a known nonsubmodular function f in Eq. (2). The (α, β)-*regret* compares to the best solution that can be expected in polynomial time and is defined by

$$R_{\alpha,\beta}(T) = \sum_{t=1}^{T} f_t(S^t) - \sum_{t=1}^{T} (\tfrac{1}{\alpha} \bar{f}_t(S_\star^T) - \beta \underline{f}_t(S_\star^T)), \tag{10}$$

where $S_\star^T = \text{argmin}_{S \subseteq [n]} \sum_{t=1}^{T} f_t(S)$. It is analogous to the α-regret which is widely used in online constrained submodular minimization (Jegelka & Bilmes, 2011) and online submodular maximization (Streeter & Golovin, 2008).

[3] If the sets S^t are chosen by a randomized algorithm, we consider the expected regret over the randomness.

Algorithm 1 Online Approximate Gradient Descent

Initialization: the point $x^1 \in [0,1]^n$ and the stepsize $\eta > 0$;
for $t = 1, 2, \ldots$ **do**
  Let $x^t_{\pi(1)} \geq \ldots x^t_{\pi(n)}$ be the sorted entries in the decreasing order with $A^t_i = \{\pi(1), \ldots, \pi(i)\}$ for all $i \in [n]$ and $A^t_0 = \emptyset$. Let $x^t_{\pi(0)} = 1$ and $x^t_{\pi(n+1)} = 0$.
  Let $\lambda^t_i = x^t_{\pi(i)} - x^t_{\pi(i+1)}$ for all $0 \leq i \leq n$.
  Sample S^t from the distribution $\mathbb{P}(S^t = A^t_i) = \lambda^t_i$ for all $0 \leq i \leq n$ and observe the new loss function f_t.
  Compute $g^t_{\pi(i)} = f_t(A^t_i) - f_t(A^t_{i-1})$ for all $i \in [n]$.
  Compute $x^{t+1} = P_{[0,1]^n}(x^t - \eta g^t)$.

Algorithm 2 Bandit Approximate Gradient Descent

Initialization: the point $x^1 \in [0,1]^n$ and the stepsize $\eta > 0$; the exploration probability $\mu \in (0,1)$.
for $t = 1, 2, \ldots, T$ **do**
  Let $x^t_{\pi(1)} \geq \ldots x^t_{\pi(n)}$ be the sorted entries in decreasing order with $A^t_i = \{\pi(1), \ldots, \pi(i)\}$ for all $i \in [n]$ and $A^t_0 = \emptyset$. Let $x^t_{\pi(0)} = 1$ and $x^t_{\pi(n+1)} = 0$.
  Let $\lambda^t_i = x^t_{\pi(i)} - x^t_{\pi(i+1)}$ for all $0 \leq i \leq n$.
  Sample S^t from the distribution $\mathbb{P}(S^t = A^t_i) = (1-\mu)\lambda^t_i + \frac{\mu}{n+1}$ for all $0 \leq i \leq n$ and observe the loss $f_t(S^t)$.
  Compute $\hat{f}^t_i = \frac{\mathbf{1}(S^t = A^t_i)}{(1-\mu)\lambda^t_i + \mu/(n+1)} f_t(S^t)$ for all $0 \leq i \leq n$.
  Compute $\hat{g}^t_{\pi(i)} = \hat{f}^t_i - \hat{f}^t_{i-1}$ for all $i \in [n]$.
  Compute x^{t+1} $x^{t+1} = P_{[0,1]^n}(x^t - \eta g^t)$.

As mentioned before, we consider the algorithmic design in both *full information* and *bandit feedback* settings. In the former one, the agent is allowed to have unlimited access to the value oracles of $f_t(\cdot)$ after choosing S^t in each round t. In the latter one, the agent only observes the incurred loss at the point that she has chosen in each round t, i.e., $f_t(S^t)$, and receives no other information.

4. Online Approximation Algorithm

We analyze online approximate gradient descent algorithm and its bandit variant for regret minimization when the nonsubmodular cost functions are in the form of Eq (8). Due to space limit, we defer the proofs to Appendix B and C.

4.1. Full information setting

Let $[0,1]^n$ be the unit hypercube and the cost function on $[0,1]^n$ corresponding to f_t is the function $(f_t)_C$ that is the convex closure of f_t. Equipped with Proposition 3.1, we can compute approximate subgradients of $(f_t)_C$ such that the online gradient descent (Zinkevich, 2003) is applicable.

This leads to Algorithm 1 which performs one-step projected gradient descent that yields x^t and then samples S^t from the distribution λ over $\{A_i\}_{i=0}^n$ encoded by x^t. It is worth mentioning that $\lambda^t_i = x^t_{\pi(i)} - x^t_{\pi(i+1)}$ for all $0 \leq i \leq n$ and λ is thus completely independent of f_t. This guarantees that Algorithm 1 is valid in online manner since f_t is realized after the decision maker chooses S^t. One of the advantages of Algorithm 1 is that it does not require the value of α and β which can be hard to compute in practice. We summarize our results for Algorithm 1 in the following theorem.

Theorem 4.1 *Suppose the adversary chooses nonsubmodular functions in Eq.* (8) *satisfying* $\bar{f}_t([n]) + \underline{f}_t([n]) \leq L$. *Fixing* $T \geq 1$ *and letting* $\eta = \frac{\sqrt{n}}{L\sqrt{T}}$ *in Algorithm 1, we have* $\mathbb{E}[R_{\alpha,\beta}(T)] = O(\sqrt{nT})$ *and* $R_{\alpha,\beta}(T) = O(\sqrt{nT} + \sqrt{T\log(1/\delta)})$ *with probability* $1-\delta$.

Remark 4.2 *Theorem 4.1 demonstrates that Algorithm 1 is regret-optimal for our setting; indeed, our setting includes online unconstrained submodular minimization as a special case where* (α,β)*-regret becomes standard regret in Eq.* (9) *and Hazan & Kale (2012) shows that Algorithm 1 is optimal up to constants. Our theoretical result also extends the results in Hazan & Kale (2012) from submodular cost functions to nonsubmodular cost functions in Eq.* (8) *using the* (α,β)*-regret instead of the standard regret in Eq.* (9).

4.2. Bandit feedback setting

In contrast with the full-information setting, the agent only observes the loss function f_t at her action S^t, i.e., $f_t(S^t)$, in bandit feedback setting. This is a more challenging setup since the agent does not have full access to the new loss function f_t at each round t yet.

Despite the bandit feedback, we can compute an unbiased estimator of the gradient g^t in Algorithm 1 using the technique of importance weighting and try to implement a stochastic version of Algorithm 1. More specifically, we notice that $\hat{f}^t_i = \frac{\mathbf{1}(S^t = A^t_i)}{\lambda^t_i} f_t(S^t)$ is unbiased for estimating $f_t(A^t_i)$ for all $0 \leq i \leq n$. Thus, $\hat{g}^t_{\pi(i)} = \hat{f}^t_i - \hat{f}^t_{i-1}$ for all $i \in [n]$ gives us an unbiased estimator of the gradient g^t. However, the variance of the estimator $\hat{g}$ could be undesirably large since the values of λ^t_i may be arbitrarily small.

To resolve this issue, we can sample S^t from a mixture distribution that combines (with probability $1-\mu$) samples from λ^t and (with probability μ) samples from the uniform distribution over $\{A^t_i\}_{i=0}^n$. This guarantees that the variance of $\hat{f}^t_i$ is upper bounded by $O(n^2/\mu)$. The similar idea has been employed in Hazan & Kale (2012) for online submodular minimization. Then, we conduct the careful analysis for the estimators $\hat{g}^t$ such that the scale of the variance is taken into account. Note that our analysis is different from the standard analysis in Flaxman et al. (2005) which seems oversimplified for our setting and results in worse regret of $O(T^{3/4})$ compared to our result in the following theorem.

Theorem 4.3 *Suppose the adversary chooses nonsubmodular functions f_t in Eq. (8) satisfying $\bar{f}_t([n]) + \underline{f}_t([n]) \leq L$. Fixing $T \geq 1$ and letting $(\eta, \mu) = (\frac{1}{LT^{2/3}}, \frac{n}{T^{1/3}})$ in Algorithm 2, we have $\mathbb{E}[R_{\alpha,\beta}(T)] = O(nT^{\frac{2}{3}})$ and $R_{\alpha,\beta}(T) = O(nT^{\frac{2}{3}} + \sqrt{n\log(1/\delta)}T^{\frac{2}{3}})$ with probability $1-\delta$.*

Remark 4.4 *Theorem 4.3 demonstrates that Algorithm 2 is no-regret for our setting even when only the bandit feedback is available, further extending the results in Hazan & Kale (2012) from submodular cost functions to nonsubmodular cost functions in Eq. (8) using the (α, β)-regret instead of the standard regret in Eq. (9).*

5. Online Delayed Approximation Algorithm

We investigate Algorithm 1 and 2 for regret minimization even when the delay between choosing an action and receiving the incurred cost exists and can be unbounded.

5.1. The general framework

The general online learning framework with large delay that we consider can be represented as follows. In each round $t = 1, \ldots, T$, the agent chooses the decision $S^t \subseteq [n]$ and this generates a loss $f_t(S^t)$. Simultaneously, S^t triggers a *delay* $d_t \geq 0$ which determines the round $t + d_t$ at which the information about f_t will be received. Finally, the agent receives the information about f_t from all previous rounds $\mathcal{R}_t = \{s : s + d_s = t\}$.

The above model has been stated in an abstract way as the basis for the regret analysis. The information about f_t is determined by whether the setting is full information or bandit feedback. Our blanket assumptions for the stream of the delays encountered will be:

Assumption 5.1 *The delays $d_t = o(t^\gamma)$ for some $\gamma < 1$.*

Assumption 5.1 is not theoretically artificial but uncovers that long delays are observed in practice (Chapelle, 2014); indeed, the data statistics from real-time bidding company suggested that more than 10% of the conversions were ≥ 2 weeks old. More specifically, Chapelle (2014) showed that the delays in online advertising have long-tail distributions when conditioning on context and feature variables available to the advertiser, thus justifying the existence of unbounded delays. Note that Assumption 5.1 is mild and the delays can even be adversarial as in Quanrud & Khashabi (2015).

5.2. Full information setting

At the round t, the agent receives the loss function $f_s(\cdot)$ for $\mathcal{R}_t = \{s : s + d_s = t\}$ after committing her decision, i,e., gets to observe $f_s(A_i^t)$ for all $s \in \mathcal{R}_t$ and all $0 \leq i \leq n$. To let Algorithm 1 handle these delays, the first thing to note is that the set $\mathcal{R}_t$ received at a given round might be empty, i.e., we could have $\mathcal{R}_t = \emptyset$ for some $t \geq 1$. Following up the pooling strategy in Héliou et al. (2020), we assume that, as information is received over time, the agent adds it to an information pool $\mathcal{P}_t$ and then uses the oldest information available in the pool (where "oldest" stands for the time at which the information was generated).

Algorithm 3 Delay Online Approximate Gradient Descent

Initialization: the point $x^1 \in [0,1]^n$ and the stepsize $\eta_t > 0$; $\mathcal{P}_0 \leftarrow \emptyset$ and $f_\infty = 0$.
for $t = 1, 2, \ldots$ **do**
  Let $x^t_{\pi(1)} \geq \ldots x^t_{\pi(n)}$ be the sorted entries in the decreasing order with $A_i^t = \{\pi(1), \ldots, \pi(i)\}$ for all $i \in [n]$ and $A_0^t = \emptyset$. Let $x^t_{\pi(0)} = 1$ and $x^t_{\pi(n+1)} = 0$.
  Let $\lambda_i^t = x^t_{\pi(i)} - x^t_{\pi(i+1)}$ for all $0 \leq i \leq n$.
  Sample S^t from the distribution $\mathbb{P}(S^t = A_i^t) = \lambda_i^t$ for $0 \leq i \leq n$ and observe the new loss function f_t.
  Compute $g^t_{\pi(i)} = f_t(A_i^t) - f_t(A_{i-1}^t)$ for all $i \in [n]$ and then trigger a delay $d_t \geq 0$.
  Let $\mathcal{R}_t = \{s : s + d_s = t\}$ and $\mathcal{P}_t \leftarrow \mathcal{P}_{t-1} \cup \mathcal{R}_t$. Take $q_t = \min \mathcal{P}_t$ and set $\mathcal{P}_t \leftarrow \mathcal{P}_t \setminus \{q_t\}$.
  Compute x^{t+1} using Eq. (11).

Since no information is available at $t = 0$, we have $\mathcal{P}_0 = \emptyset$ and update the agent's information pool recursively: $\mathcal{P}_t = \mathcal{P}_{t-1} \cup \mathcal{R}_t \setminus \{q_t\}$ where $q_t = \min(\mathcal{P}_{t-1} \cup \mathcal{R}_t)$ denotes the oldest round from which the agent has unused information at round t. As Héliou et al. (2020) pointed out, this scheme can be seen as a priority queue where $\{f_s(\cdot), s \in \mathcal{R}_t\}$ arrives at time t and is assigned in order; subsequently, the oldest information is utilized at first. An important issue that arises in the above computation is that, it may well happen that the agent's information pool $\mathcal{P}_t$ is empty at time t (e.g., if we have $d_1 > 0$ at time $t = 1$). Following the convention that $\inf \emptyset = +\infty$, we set $q_t = +\infty$ and $g^\infty = 0$ (since it is impossible to have information at time $t = +\infty$). Under this convection, the computation of a new iterate x^{t+1} at time t can be written more explicitly form as follows,

$$x^{t+1} = \begin{cases} x^t & \text{if } \mathcal{P}_t = \emptyset, \\ P_{[0,1]^n}(x^t - \eta_t g^{q_t}), & \text{otherwise.} \end{cases} \tag{11}$$

We present a delayed variant of Algorithm 1 in Algorithm 3. There is no information aggregation here but the updates of x^{t+1} follows the pooling policy induced by a priority queue. We summarize our results in the following theorem.

Theorem 5.2 *Suppose the adversary chooses nonsubmodular functions in Eq. (8) satisfying $\bar{f}_t([n]) + \underline{f}_t([n]) \leq L$ and let the delays satisfy Assumption 5.1. Fixing $T \geq 1$ and letting $\eta_t = \frac{\sqrt{n}}{L\sqrt{t^{1+\gamma}}}$ in Algorithm 3, we have $\mathbb{E}[R_{\alpha,\beta}(T)] = O(\sqrt{nT^{1+\gamma}})$ and $R_{\alpha,\beta}(T) = O(\sqrt{nT^{1+\gamma}} + \sqrt{T\log(1/\delta)})$ with probability $1-\delta$.*

Remark 5.3 *Theorem 5.2 demonstrates that Algorithm 3 is no-regret if Assumption 5.1 hold. To our knowledge, this*

Algorithm 4 Delay Bandit Approximate Gradient Descent

1: **Initialization:** the point $x^1 \in [0,1]^n$ and the stepsize $\eta_t > 0$; $\mathcal{P}_0 \leftarrow \emptyset$ and $f_\infty = 0$; the exploration probability $\mu_t \in (0,1)$.
2: **for** $t = 1, 2, \ldots$ **do**
3: Let $x^t_{\pi(1)} \geq \ldots x^t_{\pi(n)}$ be the sorted entries in the decreasing order with $A^t_i = \{\pi(1), \ldots, \pi(i)\}$ for all $i \in [n]$ and $A^t_0 = \emptyset$. Let $x^t_{\pi(0)} = 1$ and $x^t_{\pi(n+1)} = 0$.
4: Let $\lambda^t_i = x^t_{\pi(i)} - x^t_{\pi(i+1)}$ for all $0 \leq i \leq n$.
5: Sample S^t from the distribution $\mathbb{P}(S^t = A^t_i) = (1 - \mu_t)\lambda^t_i + \frac{\mu_t}{n+1}$ for $0 \leq i \leq n$ and observe the loss $f_t(S^t)$.
6: Compute $\hat{f}^t_i = \frac{\mathbf{1}(S^t = A^t_i)}{(1-\mu_t)\lambda^t_i + \mu_t/(n+1)} f_t(S^t)$.
7: Compute $\hat{g}^t_{\pi(i)} = \hat{f}^t_i - \hat{f}^t_{i-1}$ for all $i \in [n]$ and then trigger a delay $d_t \geq 0$.
8: Let $\mathcal{R}_t = \{s : s + d_s = t\}$ and $\mathcal{P}_t \leftarrow \mathcal{P}_{t-1} \cup \mathcal{R}_t$. Take $q_t = \min \mathcal{P}_t$ and set $\mathcal{P}_t \leftarrow \mathcal{P}_t \setminus \{q_t\}$.
9: Compute x^{t+1} using Eq. (12).

is the first theoretical guarantee for no-regret learning in online nonsubmodular minimization with delayed costs and also complement similar results for online convex optimization with delayed costs (Quanrud & Khashabi, 2015).

5.3. Bandit feedback setting

As we have done in the previous section, we will make use of an unbiased estimator $\hat{g}$ of the gradient for the bandit feedback setting. However, we only receive the old estimator $\hat{g}^{q_t}$ at the round t due to the delay d_t. Following the same reasoning as in the full information setting, the computation of a new iterate x^{t+1} at time t can be written more explicitly form as follows,

$$x^{t+1} = \begin{cases} x^t & \text{if } \mathcal{P}_t = \emptyset, \\ P_{[0,1]^n}(x^t - \eta_t \hat{g}^{q_t}), & \text{otherwise.} \end{cases} \tag{12}$$

Algorithm 4 follows the same template as Algorithm 3 but substituting the exact gradients with the gradient estimator. We summarize our results in the following theorem.

Theorem 5.4 *Suppose the adversary chooses nonsubmodular functions in Eq. (8) satisfying $\bar{f}_t([n]) + \underline{f}_t([n]) \leq L$ and let the delays satisfy Assumption 5.1. Fixing $T \geq 1$ and letting $(\eta_t, \mu_t) = (\frac{1}{Lt^{(2+\gamma)/3}}, \frac{n}{t^{(1-\gamma)/3}})$ in Algorithm 4, we have $\mathbb{E}[R_{\alpha,\beta}(T)] = O(nT^{\frac{2+\gamma}{3}})$ and $R_{\alpha,\beta}(T) = O(nT^{\frac{2+\gamma}{3}} + \sqrt{n\log(1/\delta)}T^{\frac{4-\gamma}{6}})$ with probability $1-\delta$.*

Remark 5.5 *Theorem 5.4 demonstrates that Algorithm 4 attains the regret of $nT^{\frac{2+\gamma}{3}}$ which is worse that that of $\sqrt{nT^{1+\gamma}}$ for Algorithm 3 and reduces to that of $nT^{\frac{2}{3}}$ for Algorithm 2. Since $\gamma < 1$ is assumed, Algorithm 4 is the first no-regret bandit learning algorithm for online nonsubmodular minimization with delayed costs to our knowledge.*

6. Experiments

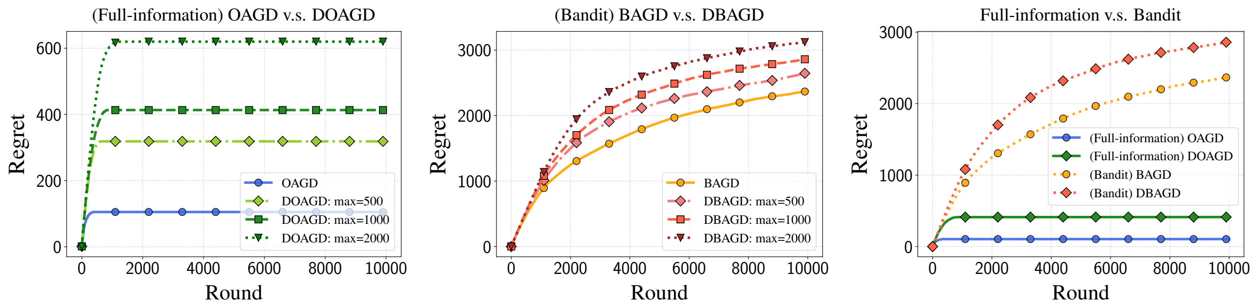
We conduct the numerical experiments on structured sparse learning problems and include Algorithm 1-4, which we refer to as OAGD, BAGD, DOAGD, and DBAGD. All the experiments are implemented in Python 3.7 with a 2.6 GHz Intel Core i7 and 16GB of memory. For all our experiments, we set total number of rounds $T = 10,000$, dimension $d = 10$, number of samples (for round t) $n = 100$, and sparse parameter $k = 2$. For OAGD and DOAGD, we set the default step size $\eta_o = \sqrt{n}/(L\sqrt{T})$ (as described in Theorem 4.1). For BAGD and DBAGD, we set the default step size $\eta_b = 1/(LT^{2/3})$ (as described in Theorem 4.3).

Our goal is to estimate the sparse vector $x^\star \in \mathbb{R}^d$ using the structured nonsubmodular model (see Example 3.1). Following the setup in El Halabi & Jegelka (2020), we let the function f^r be the regularization in Eq. (3) such that $f(S) = f^r(S) = \max(S) - \min(S) + 1$ for all $S \neq \emptyset$ and $f^r(\emptyset) = 0$. We generate true solution $x^\star \in \mathbb{R}^d$ with k consecutive 1's and other $n-k$ elements are zeros. We define the function $h_t(S)$ for the round t as follows: let $y_t = A_t x^\star + \epsilon_t$ where each row of $A_t \in \mathbb{R}^{n\times d}$ is an i.i.d. Gaussian vector and each entry of $\epsilon_t \in \mathbb{R}^n$ is sampled from a normal distribution with standard deviation equals to 0.01. Then, we define the square loss $\ell_t(x) = \|A_t x - y_t\|_2^2$ and let $h_t(S) = \ell_t(0) - \min_{\text{supp}(x)\subseteq S} \ell_t(x)$. We consider the constant delays in our experiments, i.e., the delay $\max_t d_t \leq d$ for all $t \geq 1$ where $d > 0$ is a constant.

Figure 1 summarizes some of experimental results. Indeed, we see from Figure 1(a) that the bigger delays lead to worse regret for the full-information setting which confirms Theorem 4.1 and 5.2. The result in Figure 1(b) demonstrates the similar phenomenon for the bandit feedback setting which confirms Theorem 4.3 and 5.4. Further, Figure 1(c) demonstrates the effect of bandit feedback and delay simultaneously; indeed, OAGD and DOAGD perform better than BAGD and DBAGD since the regret will increase if only the bandit feedback is available. We implement all the algorithms with varying step sizes and summarize the results in Figure 2 and 3. In the former one, we use step sizes $\eta_o/2 = \sqrt{n}/(2L\sqrt{T})$ for OAGD and DOAGD and $\eta_b/2 = 1/(2LT^{2/3})$ for BAGD and DBAGD. In the latter one, we use step sizes $\eta_o/5 = \sqrt{n}/(5L\sqrt{T})$ for OAGD and DOAGD, and $\eta_b/5 = 1/(5LT^{2/3})$ for BAGD and DBAGD. Figure 1-3 demonstrate that our proposed algorithms are not sensitive to the step size choice.

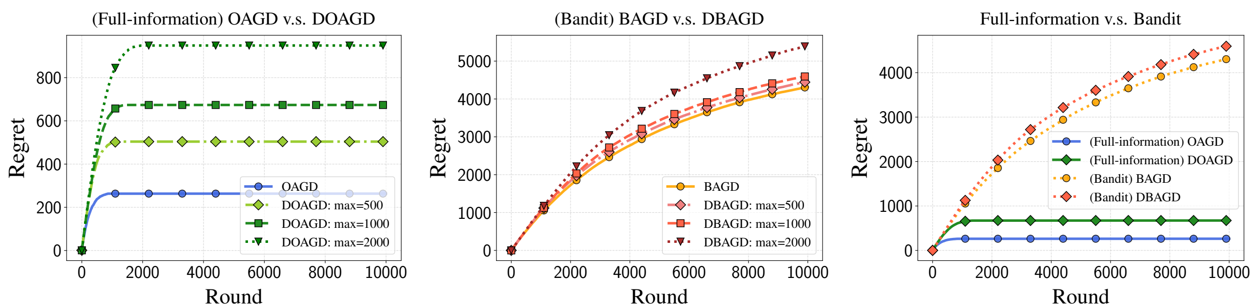
7. Concluding Remarks

This paper studied online nonsubmodular minimization with special structure through the lens of (α,β)-regret and the extension of generic convex relaxation model. We proved that online approximate gradient descent algorithm and its



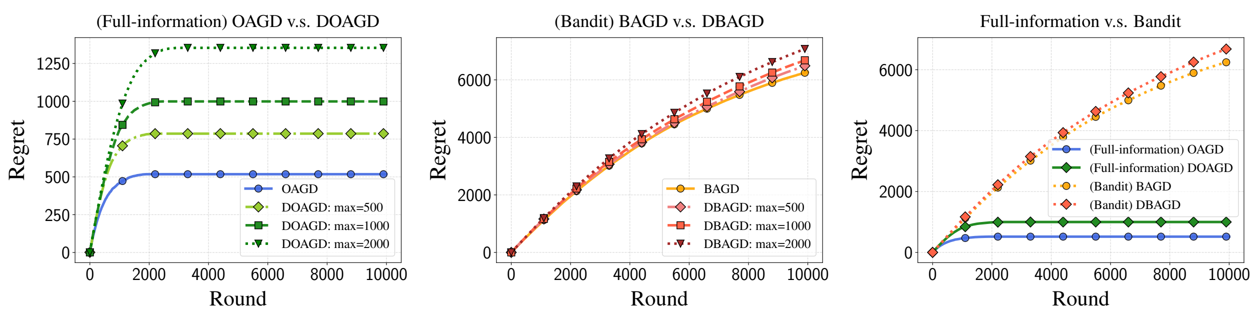
(a) (Full Information) Effect of delay. (b) (Bandit Feedback) Effect of delay. (c) Effect of bandit feedback and delay.

Figure 1. Comparison of our algorithms on sparse learning with delayed costs. In (a) and (b), we examine the effect of delay in the full-information and bandit settings respectively where the maximum delay $d \in \{500, 1000, 2000\}$. In (c), we examine the effect of bandit feedback by comparing the online algorithm with its bandit version where the maximum delay $d = 500$.



(a) (Full Information) Effect of delay. (b) (Bandit Feedback) Effect of delay. (c) Effect of bandit feedback and delay.

Figure 2. Comparison of our algorithms on sparse learning with delayed costs and step size $\eta = \sqrt{n}/(2L\sqrt{T})$ for OAGD and DOAGD, and $\eta = 1/(2LT^{2/3})$ for BAGD and DBAGD. Note that (a), (b) and (c) follow the same setup as Figure 1.



(a) (Full Information) Effect of delay. (b) (Bandit Feedback) Effect of delay. (c) Effect of bandit feedback and delay.

Figure 3. Comparison of our algorithms on sparse learning with delayed costs and step size $\eta = \sqrt{n}/(5L\sqrt{T})$ for OAGD and DOAGD, and $\eta = 1/(5LT^{2/3})$ for BAGD and DBAGD. Note that (a), (b) and (c) follow the same setup as Figure 1.

bandit variant adapted for the convex relaxation model could achieve the bounds of $O(\sqrt{nT})$ and $O(nT^{\frac{2}{3}})$ in terms of (α, β)-regret respectively. We also investigated the delayed variants of two algorithms and proved new regret bounds where the delays can even be unbounded. More specifically, if delays satisfy $d_t = o(t^\gamma)$ with $\gamma < 1$, we showed that our proposed algorithms achieve the regret bound of $O(\sqrt{nT^{1+\gamma}})$ and $O(nT^{\frac{2+\gamma}{3}})$ for full-information setting and bandit setting respectively. Simulation studies validate our theoretical findings in practice.

Acknowledgements

This work was supported in part by the Mathematical Data Science program of the Office of Naval Research under grant number N00014-18-1-2764 and by the Vannevar Bush Faculty Fellowship program under grant number N00014-21-1-2941. The work of Michael Jordan is also partially supported by NSF Grant IIS-1901252.

References

Anari, N., Haghtalab, N., Naor, S., Pokutta, S., Singh, M., and Torrico, A. Structured robust submodular maximization: Offline and online algorithms. In *AISTATS*, pp. 3128–3137. PMLR, 2019.

Arora, S., Hazan, E., and Kale, S. The multiplicative weights update method: A meta-algorithm and applications. *Theory of Computing*, 8(1):121–164, 2012.

Auer, P. Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, 3(Nov):397–422, 2002.

Bach, F. Structured sparsity-inducing norms through submodular functions. In *NeurIPS*, pp. 118–126, 2010.

Bai, W., Iyer, R., Wei, K., and Bilmes, J. Algorithms for optimizing the ratio of submodular functions. In *ICML*, pp. 2751–2759. PMLR, 2016.

Bian, A. A., Buhmann, J. M., Krause, A., and Tschiatschek, S. Guarantees for greedy maximization of non-submodular functions with applications. In *ICML*, pp. 498–507. PMLR, 2017.

Bistritz, I., Zhou, Z., Chen, X., Bambos, N., and Blanchet, J. Online EXP3 learning in adversarial bandits with delayed feedback. In *NeurIPS*, pp. 11349–11358, 2019.

Blum, A. and Mansour, Y. From external to internal regret. *Journal of Machine Learning Research*, 8(6), 2007.

Bogunovic, I., Scarlett, J., Krause, A., and Cevher, V. Truncated variance reduction: A unified approach to Bayesian optimization and level-set estimation. In *NeurIPS*, pp. 1507–1515, 2016.

Bogunovic, I., Zhao, J., and Cevher, V. Robust maximization of non-submodular objectives. In *AISTATS*, pp. 890–899. PMLR, 2018.

Boykov, Y. and Kolmogorov, V. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9):1124–1137, 2004.

Boykov, Y., Veksler, O., and Zabih, R. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11): 1222–1239, 2001.

Buchbinder, N., Feldman, M., and Schwartz, R. Online submodular maximization with preemption. In *SODA*, pp. 1202–1216. SIAM, 2014.

Cardoso, A. R. and Cummings, R. Differentially private online submodular minimization. In *AISTATS*, pp. 1650–1658. PMLR, 2019.

Cesa-Bianchi, N. and Lugosi, G. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

Cesa-Bianchi, N. and Lugosi, G. Combinatorial bandits. *Journal of Computer and System Sciences*, 78(5):1404–1422, 2012.

Chapelle, O. Modeling delayed feedback in display advertising. In *KDD*, pp. 1097–1105, 2014.

Chen, L., Feldman, M., and Karbasi, A. Weakly submodular maximization beyond cardinality constraints: Does randomization help greedy? In *ICML*, pp. 804–813. PMLR, 2018a.

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

Chen, L., Hassani, H., and Karbasi, A. Online continuous submodular maximization. In *AISTATS*, pp. 1896–1905. PMLR, 2018c.

Conforti, M. and Cornuéjols, G. Submodular set functions, matroids and the greedy algorithm: tight worst-case bounds and some generalizations of the Rado-Edmonds theorem. *Discrete Applied Mathematics*, 7(3):251–274, 1984.

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

Das, A. and Kempe, D. Submodular meets spectral: Greedy algorithms for subset selection, sparse approximation and dictionary selection. In *ICML*, pp. 1057–1064, 2011.

Dughmi, S. Submodular functions: Extensions, distributions, and algorithms. a survey. *ArXiv Preprint: 0912.0322*, 2009.

Edmonds, J. Submodular functions, matroids, and certain polyhedra. In *Combinatorial Optimization—Eureka, You Shrink!*, pp. 11–26. Springer, 2003.

El Halabi, M. and Cevher, V. A totally unimodular view of structured sparsity. In *AISTATS*, pp. 223–231. PMLR, 2015.

El Halabi, M. and Jegelka, S. Optimal approximation for unconstrained non-submodular minimization. In *ICML*, pp. 3961–3972. PMLR, 2020.

El Halabi, M., Bach, F., and Cevher, V. Combinatorial penalties: Which structures are preserved by convex relaxations? In *AISTATS*, pp. 1551–1560. PMLR, 2018.

Elenberg, E. R., Khanna, R., Dimakis, A. G., and Negahban, S. Restricted strong convexity implies weak submodularity. *The Annals of Statistics*, 46(6B):3539–3568, 2018.

Flaxman, A. D., Kalai, A. T., and McMahan, H. B. Online convex optimization in the bandit setting: gradient descent without a gradient. In *SODA*, pp. 385–394, 2005.

Freedman, D. A. On tail probabilities for martingales. *The Annals of Probability*, pp. 100–118, 1975.

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

González, J., Dai, Z., Hennig, P., and Lawrence, N. Batch Bayesian optimization via local penalization. In *AISTATS*, pp. 648–657. PMLR, 2016.

Grötschel, M., Lovász, L., and Schrijver, A. *Geometric algorithms and combinatorial optimization*, volume 2. Springer Science & Business Media, 2012.

Gyorgy, A. and Joulani, P. Adapting to delays and data in adversarial multi-armed bandits. In *ICML*, pp. 3988–3997. PMLR, 2021.

Harshaw, C., Feldman, M., Ward, J., and Karbasi, A. Submodular maximization beyond non-negativity: Guarantees, fast algorithms, and applications. In *ICML*, pp. 2634–2643. PMLR, 2019.

Harvey, N., Liaw, C., and Soma, T. Improved algorithms for online submodular maximization via first-order regret bounds. In *NeurIPS*, pp. 123–133. Curran Associates, Inc., 2020.

Hassidim, A. and Singer, Y. Optimization for approximate submodularity. In *NeurIPS*, pp. 394–405, 2018.

Hazan, E. Introduction to online convex optimization. *Foundations and Trends in Optimization*, 2(3-4):157–325, 2016.

Hazan, E. and Kale, S. Online submodular minimization. *Journal of Machine Learning Research*, 13(10), 2012.

Hazan, E., Agarwal, A., and Kale, S. Logarithmic regret algorithms for online convex optimization. *Machine Learning*, 69(2-3):169–192, 2007.

Héliou, A., Mertikopoulos, P., and Zhou, Z. Gradient-free online learning in continuous games with delayed rewards. In *ICML*, pp. 4172–4181. PMLR, 2020.

Hoeffding, W. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

Horel, T. and Singer, Y. Maximization of approximately submodular functions. In *NeurIPS*, pp. 3045–3053, 2016.

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

Iwata, S. and Orlin, J. B. A simple combinatorial algorithm for submodular function minimization. In *SODA*, pp. 1230–1237, 2009.

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

Iyer, R. and Bilmes, J. Algorithms for approximate minimization of the difference between submodular functions, with applications. In *UAI*, pp. 407–417, 2012.

Iyer, R. K., Jegelka, S., and Bilmes, J. A. Curvature and optimal algorithms for learning and minimizing submodular functions. In *NeurIPS*, pp. 2742–2750, 2013.

Jegelka, S. and Bilmes, J. Online submodular minimization for combinatorial structures. In *ICML*, pp. 345–352, 2011.

Joulani, P., Gyorgy, A., and Szepesvári, C. Online learning under delayed feedback. In *ICML*, pp. 1453–1461. PMLR, 2013.

Joulani, P., Gyorgy, A., and Szepesvári, C. Delay-tolerant online convex optimization: Unified analysis and adaptive-gradient algorithms. In *AAAI*, pp. 1744–1750, 2016.

Kalai, A. and Vempala, S. Efficient algorithms for online decision problems. *Journal of Computer and System Sciences*, 71(3):291–307, 2005.

Kawahara, Y., Iyer, R., and Bilmes, J. On approximate non-submodular minimization via tree-structured supermodularity. In *AISTATS*, pp. 444–452. PMLR, 2015.

Kuhnle, A., Smith, J. D., Crawford, V., and Thai, M. Fast maximization of non-submodular, monotonic functions on the integer lattice. In *ICML*, pp. 2786–2795. PMLR, 2018.

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

Lee, Y. T., Sidford, A., and Wong, S. C.-W. A faster cutting plane method and its implications for combinatorial and convex optimization. In *FOCS*, pp. 1049–1065. IEEE, 2015.

Lehmann, B., Lehmann, D., and Nisan, N. Combinatorial auctions with decreasing marginal utilities. *Games and Economic Behavior*, 55(2):270–296, 2006.

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

Matsuoka, T., Ito, S., and Ohsaka, N. Tracking regret bounds for online submodular optimization. In *AISTATS*, pp. 3421–3429. PMLR, 2021.

McCormick, S. T. Submodular function minimization. *Handbooks in Operations Research and Management Science*, 12:321–391, 2005.

Narasimhan, M., Jojic, N., and Bilmes, J. Q-Clustering. In *NeurIPS*, pp. 979–986, 2005.

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

Pike-Burke, C., Agrawal, S., Szepesvari, C., and Grunewalder, S. Bandits with delayed, aggregated anonymous feedback. In *ICML*, pp. 4105–4113. PMLR, 2018.

Quanrud, K. and Khashabi, D. Online learning with adversarial delays. In *NeurIPS*, pp. 1270–1278, 2015.

Rapaport, F., Barillot, E., and Vert, J.-P. Classification of arrayCGH data using fused SVM. *Bioinformatics*, 24 (13):i375–i382, 2008.

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

Sakaue, S. Greedy and IHT algorithms for non-convex optimization with monotone costs of non-zeros. In *AISTATS*, pp. 206–215. PMLR, 2019.

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

Shalev-Shwartz, S. Online learning and online convex optimization. *Foundations and Trends in Machine Learning*, 4(2):107–194, 2011.

Shalev-Shwartz, S. and Singer, Y. Convex repeated games and Fenchel duality. In *NIPS*, pp. 1265–1272, 2006.

Streeter, M. and Golovin, D. An online algorithm for maximizing submodular functions. In *NeurIPS*, pp. 1577–1584, 2008.

Sviridenko, M., Vondrák, J., and Ward, J. Optimal approximation for submodular and supermodular optimization with bounded curvature. *Mathematics of Operations Research*, 42(4):1197–1218, 2017.

Svitkina, Z. and Fleischer, L. Submodular approximation: Sampling-based algorithms and lower bounds. *SIAM Journal on Computing*, 40(6):1715–1737, 2011.

Thang, N. K. and Srivastav, A. Online non-monotone DR-submodular maximization. In *AAAI*, pp. 9868–9876, 2021.

Thune, T. S., Cesa-Bianchi, N., and Seldin, Y. Nonstochastic multiarmed bandits with unrestricted delays. In *NeurIPS*, pp. 6538–6547, 2019.

Trevisan, L. Inapproximability of combinatorial optimization problems. *Paradigms of Combinatorial Optimization: Problems and New Approaches*, pp. 381–434, 2014.

Vernade, C., Cappé, O., and Perchet, V. Stochastic bandit models for delayed conversions. In *UAI*. AUAI Press, 2017.

Vondrák, J. *Submodularity in Combinatorial Optimization*. PhD thesis, Charles University, 2007.

Zhou, Z., Mertikopoulos, P., Bambos, N., Glynn, P., and Tomlin, C. Countering feedback delays in multi-agent learning. In *NeurIPS*, pp. 6172–6182, 2017.

Zhou, Z., Xu, R., and Blanchet, J. Learning in generalized linear contextual bandits with stochastic delays. In *NeurIPS*, pp. 5197–5208, 2019.

Zimmert, J. and Seldin, Y. An optimal algorithm for adversarial bandits with arbitrary delays. In *AISTATS*, pp. 3285–3294. PMLR, 2020.

Zinkevich, M. Online convex programming and generalized infinitesimal gradient ascent. In *ICML*, pp. 928–936, 2003.

A. Proof of Proposition 3.1

We have

$$f_C(x) = \max_{g \in \mathbb{R}^d, \rho \in \mathbb{R}} \left\{ g^\top x + \rho : g(A) + \rho \leq f(A), \forall A \subseteq [n] \right\}. \tag{13}$$

First, we prove Eq. (5) using the definition of f_L and Eq. (13). Indeed, we have $f_L(x) = g^\top x$ where we let $x \in [0,1]^n$ with $x_{\pi(1)} \geq \ldots \geq x_{\pi(n)}$ and $g_{\pi(i)} = f(\pi(i) \mid A_{i-1})$ for all $i \in [n]$. Then, it suffices to show that $g^\top x \geq \tilde{g}^\top x + \tilde{\rho}$ in which $\tilde{g}(A) + \tilde{\rho} \leq f(A)$ for all $A \subseteq [n]$. We have

$$\begin{aligned} g^\top x - (\tilde{g}^\top x + \tilde{\rho}) &= \sum_{i=1}^n x_{\pi(i)} \left(f(\pi(i) \mid A_{i-1}) - \tilde{g}_{\pi(i)} \right) - \tilde{\rho} \\ &= \sum_{i=1}^{n-1} \left(x_{\pi(i)} - x_{\pi(i+1)} \right) \left(f(A_i) - \tilde{g}(A_i) \right) + x_{\pi(n)} \left(f([n]) - \tilde{g}([n]) \right) - \tilde{\rho}. \end{aligned}$$

Since $\tilde{g}(A) + \tilde{\rho} \leq f(A)$ for all $A \subseteq [n]$, we have

$$f([n]) - \tilde{g}([n]) \geq \tilde{\rho}, \quad f(A_i) - \tilde{g}(A_i) \geq \tilde{\rho}, \quad \text{for all } i \in [n].$$

Putting these pieces together with $x_{\pi(1)} \geq \ldots \geq x_{\pi(n)}$ yields that

$$g^\top x - (\tilde{g}^\top x + \tilde{\rho}) \geq \sum_{i=1}^{n-1} \left(x_{\pi(i)} - x_{\pi(i+1)} \right) \tilde{\rho} + x_{\pi(n)} \tilde{\rho} - \tilde{\rho} = (x_{\pi(1)} - 1) \tilde{\rho}.$$

Since $x \in [0,1]^n$, we have $x_{\pi(1)} \leq 1$. Since $\tilde{g}(A) + \tilde{\rho} \leq f(A)$ for all $A \subseteq [n]$ and $f(\emptyset) = 0$, we derive by letting $A = \emptyset$ that $\tilde{\rho} \leq f(\emptyset) - \tilde{g}(\emptyset) \leq 0$. This implies the desired result.

Further, we prove Eq. (6) using the definition of weak DR-submodularity. Indeed, we have $g(A) = \sum_{i \in A} g_i$. Since $g_{\pi(i)} = f(\pi(i) \mid A_{i-1})$ for all $i \in [n]$, we have

$$g(A) = \sum_{\pi(i) \in A} f(\pi(i) \mid A_{i-1}) = \sum_{\pi(i) \in A} \left(\bar{f}(\pi(i) \mid A_{i-1}) - \underline{f}(\pi(i) \mid A_{i-1}) \right).$$

Since $\bar{f}$ is α-weakly DR-submodular, $\underline{f}$ is β-weakly DR-supermodular and $A \cap A_{i-1} \subseteq A_{i-1}$, we have

$$\bar{f}(\pi(i) \mid A \cap A_{i-1}) \geq \alpha \bar{f}(\pi(i) \mid A_{i-1}), \quad \underline{f}(\pi(i) \mid A_{i-1}) \geq \beta \underline{f}(\pi(i) \mid A \cap A_{i-1}). \tag{14}$$

Putting these pieces together yields that

$$g(A) \leq \sum_{\pi(i) \in A} \left(\tfrac{1}{\alpha} \bar{f}(\pi(i) \mid A \cap A_{i-1}) - \beta \underline{f}(\pi(i) \mid A \cap A_{i-1}) \right).$$

Then, we have

$$\begin{aligned} g(A) &\leq \tfrac{1}{\alpha} \left(\sum_{i=1}^n \left(\bar{f}(A \cap A_i) - \bar{f}(A \cap A_{i-1}) \right) \right) - \beta \left(\sum_{i=1}^n \left(\underline{f}(A \cap A_i) - \underline{f}(A \cap A_{i-1}) \right) \right) \\ &= \tfrac{1}{\alpha} \bar{f}(A) - \beta \underline{f}(A), \quad \text{for all } A \subseteq [n]. \end{aligned}$$

This implies the desired result.

Finally, we prove Eq. (7) using Eq. (13). Indeed, we have $g = \bar{g} - \underline{g}$ where $\bar{g}_{\pi(i)} = \bar{f}(\pi(i) \mid A_{i-1})$ and $\underline{g}_{\pi(i)} = \underline{f}(\pi(i) \mid A_{i-1})$ for all $i \in [n]$. For any $A \subseteq [n]$, we obtain by using Eq. (14) that

$$\bar{g}(A) \leq \tfrac{1}{\alpha} \left(\sum_{\pi(i) \in A} \bar{f}(\pi(i) \mid A \cap A_{i-1}) \right) = \tfrac{1}{\alpha} \left(\sum_{i=1}^n \left(\bar{f}(A \cap A_i) - \bar{f}(A \cap A_{i-1}) \right) \right) = \tfrac{1}{\alpha} \bar{f}(A),$$

$$-\underline{g}(A) \leq -\beta \left(\sum_{\pi(i) \in A} \underline{f}(\pi(i) \mid A \cap A_{i-1}) \right) = -\beta \left(\sum_{i=1}^n \left(\underline{f}(A \cap A_i) - \underline{f}(A \cap A_{i-1}) \right) \right) = -\beta \underline{f}(A).$$

Equivalently, we have $\alpha\bar{g}(A) + 0 \leq \bar{f}(A)$ and $\frac{1}{\beta}(-\underline{g}(A)) + 0 \leq -\underline{f}(A)$ for any $A \subseteq [n]$. Using Eq. (13), we have

$$\alpha\bar{g}^\top z + 0 \leq \bar{f}_C(z), \quad \tfrac{1}{\beta}(-\underline{g})^\top z + 0 \leq (-\underline{f})_C(z), \quad \text{for all } z \in [0,1]^n.$$

Since $g = \bar{g} - \underline{g}$ and $\alpha, \beta > 0$, we have $g^\top z \leq \frac{1}{\alpha}\bar{f}_C(z) + \beta(-\underline{f})_C(z)$. This implies the desired result.

B. Regret Analysis for Algorithm 1

In this section, we present several technical lemmas for analyzing the regret minimization property of Algorithm 1. We also give the missing proof of Theorem 4.1.

B.1. Technical lemmas

We provide two technical lemmas for Algorithm 1. The first lemma gives a bound on the vector g^t and the difference between x^t and any fixed $x \in [0,1]^n$.

Lemma B.1 *Suppose that the iterates $\{x^t\}_{t\geq 1}$ and the vectors $\{g^t\}_{t\geq 1}$ be generated by Algorithm 1 and $x \in [0,1]^n$ and let $f_t = \bar{f}_t - \underline{f}_t$ satisfy that $\bar{f}_t([n]) + \underline{f}_t([n]) \leq L$ for all $t \geq 1$ and both $\bar{f}_t$ and $\underline{f}_t$ are nondecreasing. Then, we have $\|x^t - x\| \leq \sqrt{n}$ and $\|g^t\| \leq L$ for all $t \geq 1$.*

Proof. Since $x^t \in [0,1]^n$ and $x \in [0,1]^n$ is fixed, we have $\|x^t - x\| \leq \sqrt{\sum_{i=1}^n 1} = \sqrt{n}$ for all $t \geq 1$. By the definition of g^t, we have $g^t_{\pi(i)} = f_t(A_i^t) - f_t(A_{i-1}^t)$ for all $i \in [n]$ where $A_i^t = \{\pi(1), \ldots, \pi(i)\}$ for all $i \in [n]$. Then, we have

$$\|g^t\| \leq \sum_{i=1}^n |f_t(A_i^t) - f_t(A_{i-1}^t)| \leq \sum_{i=1}^n |\bar{f}_t(A_i^t) - \bar{f}_t(A_{i-1}^t)| + \sum_{i=1}^n |\underline{f}_t(A_i^t) - \underline{f}_t(A_{i-1}^t)|.$$

Since $\bar{f}_t$ and $\underline{f}_t$ are both normalized and non-decreasing, we have

$$\sum_{i=1}^n |\bar{f}_t(A_i^t) - \bar{f}_t(A_{i-1}^t)| + \sum_{i=1}^n |\underline{f}_t(A_i^t) - \underline{f}_t(A_{i-1}^t)| = \bar{f}_t([n]) + \underline{f}_t([n]) \leq L.$$

Putting these pieces together yields that $\|g^t\| \leq L$ for all $t = 1, 2, \ldots, T$. ☐

The second lemma gives a key inequality for analyzing Algorithm 1.

Lemma B.2 *Suppose that the iterates $\{x^t\}_{t\geq 1}$ are generated by Algorithm 1 and $x \in [0,1]^n$ and let $f_t = \bar{f}_t - \underline{f}_t$ satisfy that $\bar{f}_t([n]) + \underline{f}_t([n]) \leq L$ for all $t \geq 1$. Then, we have*

$$\sum_{t=1}^T \mathbb{E}[(f_t)_L(x^t)] \leq \left(\sum_{t=1}^T \tfrac{1}{\alpha}(\bar{f}_t)_C(x) + \beta(-\underline{f}_t)_C(x)\right) + \tfrac{n}{2\eta} + \tfrac{\eta L^2 T}{2}.$$

Proof. Since $x^{t+1} = P_{[0,1]^n}(x^t - \eta g^t)$, we have

$$(x - x^{t+1})^\top (x^{t+1} - x^t + \eta g^t) \geq 0, \quad \text{for all } x \in [0,1]^n.$$

Rearranging the above inequality and using the fact that $\eta > 0$, we have

$$(x^{t+1} - x)^\top g^t \leq \frac{1}{\eta}(x - x^{t+1})^\top (x^{t+1} - x^t) = \tfrac{1}{2\eta}\left(\|x - x^t\|^2 - \|x - x^{t+1}\|^2 - \|x^t - x^{t+1}\|^2\right). \tag{15}$$

Using Young's inequality, we have

$$(x^t - x^{t+1})^\top g^t \leq \tfrac{1}{2\eta}\|x^t - x^{t+1}\|^2 + \tfrac{\eta}{2}\|g^t\|^2. \tag{16}$$

Combining Eq. (15) and Eq. (16) yields that

$$(x^t - x)^\top g^t \leq \tfrac{1}{2\eta}\left(\|x - x^t\|^2 - \|x - x^{t+1}\|^2\right) + \tfrac{\eta}{2}\|g^t\|^2.$$

Since $f_t = \bar{f}_t - \underline{f}_t$ where $\bar{f}_t$ and $\underline{f}_t$ are both non-decreasing, $\bar{f}_t$ is α-weakly DR-submodular and $\underline{f}_t$ is β-weakly DR-supermodular, Proposition 3.1 implies that

$$(x^t - x)^\top g^t \geq (f_t)_L(x^t) - \left(\tfrac{1}{\alpha}(\bar{f}_t)_C(x) + \beta(-\underline{f}_t)_C(x)\right).$$

By Lemma B.1, we have $\|g^t\| \leq L$ for all $t = 1, 2, \ldots, T$. Then, we have

$$(f_t)_L(x^t) \leq \tfrac{1}{\alpha}(\bar{f}_t)_C(x) + \beta(-\underline{f}_t)_C(x) + \tfrac{1}{2\eta}\left(\|x - x^t\|^2 - \|x - x^{t+1}\|^2\right) + \tfrac{\eta L^2}{2}.$$

Summing up the above inequality over $t = 1, 2, \ldots, T$ and using $\|x^1 - x\| \leq \sqrt{n}$ (cf. Lemma B.1), we have

$$\sum_{t=1}^T (f_t)_L(x^t) \leq \left(\sum_{t=1}^T \tfrac{1}{\alpha}(\bar{f}_t)_C(x) + \beta(-\underline{f}_t)_C(x)\right) + \tfrac{n}{2\eta} + \tfrac{\eta L^2 T}{2}.$$

Taking the expectation of both sides yields the desired inequality. □

B.2. Proof of Theorem 4.1

By the definition of the Lovász extension, we have

$$(f_t)_L(x^t) = \sum_{i=1}^{n-1}(x^t_{\pi(i)} - x^t_{\pi(i+1)})f_t(A_i^t) + (1 - x^t_{\pi(1)})f_t(A_0^t) + x^t_{\pi(n)}f_t(A_n^t) = \sum_{i=0}^n \lambda_i^t f_t(A_i^t).$$

By the update formula, we have $\mathbb{E}[f_t(S^t) \mid x^t] = (f_t)_L(x^t)$ which implies that $\mathbb{E}[f_t(S^t)] = \mathbb{E}[(f_t)_L(x^t)]$. By the definition of the convex closure, we obtain that the convex closure of a set function f agrees with f on all the integer points (Dughmi, 2009, Page 4, Proposition 3.3). Letting $S_\star^T = \mathrm{argmin}_{S \subseteq [n]} \sum_{t=1}^T f_t(S)$, we have $S_\star^T$ is an integer point and

$$(\bar{f}_t)_C(\chi_{S_\star^T}) = \bar{f}_t(S_\star^T), \quad (-\underline{f}_t)_C(\chi_{S_\star^T}) = -\beta\underline{f}_t(S_\star^T),$$

which implies that

$$\tfrac{1}{\alpha}(\bar{f}_t)_C(\chi_{S_\star^T}) + \beta(-\underline{f}_t)_C(\chi_{S_\star^T}) = \tfrac{1}{\alpha}\bar{f}_t(S_\star^T) - \beta\underline{f}_t(S_\star^T).$$

Putting these pieces together and letting $x = \chi_{S_\star^T}$ in the inequality of Lemma B.2 yields that

$$\sum_{t=1}^T \mathbb{E}[f_t(S^t)] \leq \left(\sum_{t=1}^T \tfrac{1}{\alpha}\bar{f}_t(S_\star^T) - \beta\underline{f}_t(S_\star^T)\right) + \tfrac{n}{2\eta} + \tfrac{\eta L^2 T}{2}.$$

Plugging the choice of $\eta = \frac{\sqrt{n}}{L\sqrt{T}}$ into the above inequality yields that $\mathbb{E}[R_{\alpha,\beta}(T)] = O(\sqrt{nT})$ as desired.

We proceed to derive a high probability bound using the concentration inequality. In particular, we review the Hoeffding inequality (Hoeffding, 1963) and refer to Cesa-Bianchi & Lugosi (2006, Appendix A) for a proof. The following proposition is a restatement of Cesa-Bianchi & Lugosi (2006, Corollary A.1).

Proposition B.3 *Let $X_1, \ldots, X_n$ be independent real-valued random variables such that for each $i = 1, \ldots, n$, there exist some $a_i \leq b_i$ such that $\mathbb{P}(a_i \leq X_i \leq b_i) = 1$. Then for every $\epsilon > 0$, we have*

$$\begin{aligned}
\mathbb{P}\left(\sum_{i=1}^n X_i - \mathbb{E}\left[\sum_{i=1}^n X_i\right] > +\epsilon\right) &\leq \exp\left(-\frac{2\epsilon^2}{\sum_{i=1}^n (b_i - a_i)^2}\right), \\
\mathbb{P}\left(\sum_{i=1}^n X_i - \mathbb{E}\left[\sum_{i=1}^n X_i\right] < -\epsilon\right) &\leq \exp\left(-\frac{2\epsilon^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).
\end{aligned}$$

Since the sequence of points $x^1, x^2, \ldots, x^T$ is obtained by several deterministic gradient descent steps, we have this sequence is purely deterministic. Since each of S^t is obtained by independent randomized rounding on the point x^t, we have the sequence of random variables $X_t = f_t(S^t)$ is independent. By definition of f_t, we have

$$|X_t| = |\bar{f}_t(S^t) - \underline{f}_t(S^t)| \leq \bar{f}_t(S^t) + \underline{f}_t(S^t).$$

Since $\bar{f}_t$ and $\underline{f}_t$ are non-decreasing and $\bar{f}_t([n]) + \underline{f}_t([n]) \leq L$ for all $t \geq 1$, we have $\mathbb{P}(-L \leq X_i \leq L) = 1$ for all $t \geq 1$. Then, by Proposition B.3, we have

$$\mathbb{P}\left(\sum_{i=1}^n f_t(S^t) - \mathbb{E}\left[\sum_{i=1}^n f_t(S^t)\right] > \epsilon\right) \leq \exp\left(-\frac{\epsilon^2}{2nL^2}\right).$$

Equivalently, we have $\sum_{i=1}^n f_t(S^t) - \mathbb{E}[\sum_{i=1}^n f_t(S^t)] \leq L\sqrt{2T\log(1/\delta)}$ with probability at least $1-\delta$. This together with $\mathbb{E}[R_{\alpha,\beta}(T)] = O(\sqrt{nT})$ yields that $R_{\alpha,\beta}(T) = O(\sqrt{nT} + \sqrt{T\log(1/\delta)})$ with probability at least $1-\delta$ as desired.

C. Regret Analysis for Algorithm 2

In this section, we present several technical lemmas for analyzing the regret minimization property of Algorithm 2. We also give the missing proofs of Theorem 4.3.

C.1. Technical lemmas

We provide several technical lemmas for Algorithm 2. The first lemma is analogous to Lemma B.1 and gives a bound on the vector $\hat{g}^t$ (in expectation) and the difference between x^t and any fixed $x \in [0,1]^n$.

Lemma C.1 *Suppose that the iterates $\{x^t\}_{t\geq 1}$ and the vectors $\{\hat{g}^t\}_{t\geq 1}$ be generated by Algorithm 2 and $x \in [0,1]^n$ and let $f_t = \bar{f}_t - \underline{f}_t$ satisfy that $\bar{f}_t([n]) + \underline{f}_t([n]) \leq L$ for all $t \geq 1$ and both $\bar{f}_t$ and $\underline{f}_t$ are nondecreasing. Then, we have $\|x^t - x\| \leq \sqrt{n}$ for all $t \geq 1$ and*

$$\mathbb{E}[\hat{g}^t \mid x^t] = g^t, \quad \mathbb{E}[\|\hat{g}^t\|^2 \mid x^t] \leq \tfrac{8n^2L^2}{\mu}, \quad \|\hat{g}^t\|^2 \leq \tfrac{2(n+1)^2L^2}{\mu^2}.$$

where we have $g^t_{\pi(i)} = f_t(A^t_i) - f_t(A^t_{i-1})$ for all $i \in [n]$.

Proof. Using the same argument as in Lemma B.1, we have $\|x^t - x\| \leq \sqrt{n}$ for all $t \geq 1$. By the definition of $\hat{g}^t$, we have

$$\hat{g}^t_{\pi(i)} = \left(\tfrac{\mathbf{1}(S^t = A^t_i)}{(1-\mu)\lambda^t_i + \frac{\mu}{n+1}} - \tfrac{\mathbf{1}(S^t = A^t_{i-1})}{(1-\mu)\lambda^t_{i-1} + \frac{\mu}{n+1}}\right) f_t(S^t), \quad \text{for all } i \in [n].$$

This together with the sampling scheme for S^t implies that

$$\mathbb{E}[\hat{g}^t_{\pi(i)} \mid x^t] = f_t(A^t_i) - f_t(A^t_{i-1}), \quad \text{for all } i \in [n],$$

Since $g^t_{\pi(i)} = f_t(A^t_i) - f_t(A^t_{i-1})$ for all $i \in [n]$, we have $\mathbb{E}[\hat{g}^t \mid x^t] = g^t$. Since $f_t = \bar{f}_t - \underline{f}_t$ satisfy that $\bar{f}_t([n]) + \underline{f}_t([n]) \leq L$ for all $t \geq 1$ and $\bar{f}_t$ and $\underline{f}_t$ are both normalized and non-decreasing, we have

$$\mathbb{E}[\|\hat{g}^t\|^2 \mid x^t] \leq \sum_{i=0}^n \tfrac{2(f_t(A^t_i))^2}{(1-\mu)\lambda^t_i + \frac{\mu}{n+1}} \leq \tfrac{2(n+1)^2L^2}{\mu} \leq \tfrac{8n^2L^2}{\mu}.$$

Further, let $S^t = A_{i_t}$ in the round t, we can apply the same argument and obtain that

$$\|\hat{g}^t\|^2 \leq 2\left(\tfrac{f_t(A^t_{i_t})}{(1-\mu)\lambda^t_{i_t} + \frac{\mu}{n+1}}\right)^2 \leq \tfrac{2(n+1)^2L^2}{\mu^2}.$$

This completes the proof. □

The second lemma is analogous to Lemma B.2 and gives a key inequality for analyzing Algorithm 2.

Lemma C.2 *Suppose that the iterates $\{x^t\}_{t\geq 1}$ are generated by Algorithm 2 and $x \in [0,1]^n$ and let $f_t = \bar{f}_t - \underline{f}_t$ satisfy that $\bar{f}_t([n]) + \underline{f}_t([n]) \leq L$ for all $t \geq 1$. Then, we have*

$$\sum_{t=1}^T \mathbb{E}[(f_t)_L(x^t)] \leq \left(\sum_{t=1}^T \tfrac{1}{\alpha}(\bar{f}_t)_C(x) + \beta(-\underline{f}_t)_C(x)\right) + \tfrac{n}{2\eta} + \tfrac{4n^2L^2\eta T}{\mu}.$$

Proof. Using the same argument as in Lemma B.2, we have

$$(x^t - x)^\top \hat{g}^t \leq \tfrac{1}{2\eta}\left(\|x - x^t\|^2 - \|x - x^{t+1}\|^2\right) + \tfrac{\eta}{2}\|\hat{g}^t\|^2.$$

By Lemma C.1, we have $\mathbb{E}[\hat{g}^t \mid x^t] = g^t$ and $\mathbb{E}[\|\hat{g}^t\|^2 \mid x^t] \leq \frac{8n^2L^2}{\mu}$ for all $t \geq 1$. This implies that

$$(x^t - x)^\top g^t \leq \tfrac{1}{2\eta}\left(\|x - x^t\|^2 - \mathbb{E}[\|x - x^{t+1}\|^2 \mid x^t]\right) + \tfrac{4n^2L^2\eta}{\mu}.$$

Since $f_t = \bar{f}_t - \underline{f}_t$ where $\bar{f}_t$ and $\underline{f}_t$ are both non-decreasing, $\bar{f}_t$ is α-weakly DR-submodular and $\underline{f}_t$ is β-weakly DR-supermodular, Proposition 3.1 implies that

$$(x^t - x)^\top g^t \geq (f_t)_L(x^t) - \left(\tfrac{1}{\alpha}(\bar{f}_t)_C(x) + \beta(-\underline{f}_t)_C(x)\right).$$

By Lemma B.1, we have $\|g^t\| \leq L$ for all $t = 1, 2, \ldots, T$. Then, we have

$$(f_t)_L(x^t) \leq \tfrac{1}{\alpha}(\bar{f}_t)_C(x) + \beta(-\underline{f}_t)_C(x) + \tfrac{1}{2\eta}\left(\|x - x^t\|^2 - \mathbb{E}[\|x - x^{t+1}\|^2 \mid x^t]\right) + \tfrac{4n^2L^2\eta}{\mu}.$$

Taking the expectation of both sides and summing up the resulting inequality over $t = 1, 2, \ldots, T$, we have

$$\sum_{t=1}^T \mathbb{E}[(f_t)_L(x^t)] \leq \left(\sum_{t=1}^T \tfrac{1}{\alpha}(\bar{f}_t)_C(x) + \beta(-\underline{f}_t)_C(x)\right) + \tfrac{1}{2\eta}\|x - x^1\|^2 + \tfrac{4n^2L^2\eta T}{\mu}.$$

Using $\|x^1 - x\| \leq \sqrt{n}$ (cf. Lemma C.1) yields the desired inequality. □

To prove the high probability bound, we require the following concentration inequality. In particular, we review the Bernstein inequality for martingales (Freedman, 1975) and refer to Cesa-Bianchi & Lugosi (2006, Appendix A) for a proof. The following proposition is a consequence of Cesa-Bianchi & Lugosi (2006, Lemma A.8).

Proposition C.3 *Let $X_1, \ldots, X_n$ be a bounded martingale difference sequence with respect to the filtration $\mathcal{F} = (\mathcal{F}_i)_{1\leq i\leq n}$ such that $|X_i| \leq K$ for each $i = 1, \ldots, n$. We also assume that $\mathbb{E}[\|X_{i+1}\|^2 \mid \mathcal{F}_i] \leq V$ for each $i = 1, \ldots, n-1$. Then for every $\delta > 0$, we have*

$$\mathbb{P}\left(\left|\sum_{i=1}^n X_i - \mathbb{E}[X_i \mid \mathcal{F}_{i-1}]\right| > \sqrt{2TV\log(1/\delta)} + \tfrac{\sqrt{2}}{3}K\log(1/\delta)\right) \leq \delta.$$

Then we provide our last lemma which significantly generalizes Lemma C.2 for deriving the high-probability bounds.

Lemma C.4 *Suppose that the iterates $\{x^t\}_{t\geq 1}$ are generated by Algorithm 2 with $\mu = \frac{n}{T^{1/3}}$ and $x \in [0,1]^n$ and let $f_t = \bar{f}_t - \underline{f}_t$ satisfy that $\bar{f}_t([n]) + \underline{f}_t([n]) \leq L$ for all $t \geq 1$. Fixing a sufficiently small $\delta \in (0,1)$ and letting $T > \log^{\frac{3}{2}}(1/\delta)$. Then, we have*

$$\sum_{t=1}^T (f_t)_L(x^t) \leq \left(\sum_{t=1}^T \tfrac{1}{\alpha}\bar{f}_t(S) - \beta\underline{f}_t(S)\right) + \tfrac{n}{2\eta} + \tfrac{4n^2L^2\eta T}{\mu} + 12LT^{\frac{2}{3}}\sqrt{n^2 + n\log(1/\delta)} + 6\eta L^2 T\sqrt{n\log(1/\delta)},$$

with probability at least $1 - 3\delta$.

Proof. Using the same argument as in Lemma C.2, we have

$$(x^t - x)^\top \hat{g}^t \leq \tfrac{1}{2\eta}\left(\|x - x^t\|^2 - \|x - x^{t+1}\|^2\right) + \tfrac{\eta}{2}\|\hat{g}^t\|^2,$$

and

$$(x^t - x)^\top g^t \geq (f_t)_L(x^t) - \left(\tfrac{1}{\alpha}(\bar{f}_t)_C(x) + \beta(-\underline{f}_t)_C(x)\right).$$

For simplicity, we define $e_t = \hat{g}^t - g^t$. Then, we have

$$(f_t)_L(x^t) - \left(\tfrac{1}{\alpha}(\bar{f}_t)_C(x) + \beta(-\underline{f}_t)_C(x)\right) \leq (x - x^t)^\top e_t + \tfrac{1}{2\eta}\left(\|x - x^t\|^2 - \|x - x^{t+1}\|^2\right) + \tfrac{\eta}{2}\|\hat{g}^t\|^2 \tag{17}$$

Summing up Eq. (17) over $t = 1, 2, \ldots, T$ and using $\|x^1 - x\| \leq \sqrt{n}$ and $\mathbb{E}[\|\hat{g}^t\|^2 \mid x^t] \leq \frac{8n^2L^2}{\mu}$ for all $t \geq 1$ (cf. Lemma C.1), we have

$$\begin{aligned}\sum_{t=1}^T (f_t)_L(x^t) \quad &\leq \quad \left(\sum_{t=1}^T \tfrac{1}{\alpha}(\bar{f}_t)_C(x) + \beta(-\underline{f}_t)_C(x) + (x - x^t)^\top e_t + \tfrac{\eta}{2}(\|\hat{g}^t\|^2 - \mathbb{E}[\|\hat{g}^t\|^2 \mid x^t])\right) \\ &\quad + \tfrac{n}{2\eta} + \tfrac{4n^2L^2\eta T}{\mu}.\end{aligned}$$

By the definition of the convex closure, we obtain that the convex closure of a set function f agrees with f on all the integer points (Dughmi, 2009, Page 4, Proposition 3.3). Letting $S \subseteq [n]$, we have $(\bar{f}_t)_C(\chi_S) = f_t(S)$ and $(-\underline{f}_t)_C(\chi_S) = -\beta\underline{f}_t(S)$ which implies that

$$\tfrac{1}{\alpha}(\bar{f}_t)_C(\chi_S) + \beta(-\underline{f}_t)_C(\chi_S) = \tfrac{1}{\alpha}f_t(S) - \beta\underline{f}_t(S).$$

Letting $x = \chi_S$, we have

$$\sum_{t=1}^T (f_t)_L(x^t) \leq \left(\sum_{t=1}^T \tfrac{1}{\alpha}\bar{f}_t(S) - \beta\underline{f}_t(S)\right) + \tfrac{n}{2\eta} + \tfrac{4n^2L^2\eta T}{\mu} + \underbrace{\sum_{t=1}^T (\chi_S - x^t)^\top e_t}_{\mathbf{I}} + \tfrac{\eta}{2}\underbrace{\left(\sum_{t=1}^T \|\hat{g}^t\|^2 - \mathbb{E}[\|\hat{g}^t\|^2 \mid x^t]\right)}_{\mathbf{II}}. \tag{18}$$

In what follows, we prove the high probability bounds for the terms **I** and **II** in the above inequality.

Bounding I. Consider the random variables $X_t = (x^t)^\top \hat{g}^t$ for all $1 \leq t \leq T$ that are adapted to the natural filtration generated by the iterates $\{x_t\}_{t \geq 1}$. By Lemma C.1 and the Hölder's inequality, we have

$$|X_t| \leq \|\hat{g}^t\|_1 \|x^t\|_\infty \leq 2\left|\frac{f_t(A^t_{i_t})}{(1-\mu)\lambda^t_{i_t} + \frac{\mu}{n+1}}\right| \leq \frac{2(n+1)L}{\mu}$$

Since $\mu = \frac{n}{T^{1/3}}$, we have $|X_t| \leq 4LT^{\frac{1}{3}}$. Further, we have

$$\mathbb{E}[X_t^2 \mid x_t] \leq \mathbb{E}[\|\hat{g}^t\|_1^2 \|x^t\|_\infty^2 \mid x_t] = \sum_{i=0}^n \frac{4(f_t(A^t_i))^2}{(1-\mu)\lambda^t_i + \frac{\mu}{n+1}} \leq \frac{2(n+1)^2L^2}{\mu} \leq 8nL^2T^{\frac{1}{3}}.$$

Since $\mathbb{E}[\hat{g}^t \mid x^t] = g^t$ and $e_t = \hat{g}^t - g^t$, Proposition C.3 implies that

$$\mathbb{P}\left(\left|\sum_{t=1}^T (x^t)^\top e_t\right| > 4LT^{\frac{2}{3}}\sqrt{n\log(1/\delta)} + 2LT^{\frac{1}{3}}\log(1/\delta)\right) \leq \delta.$$

Since $T > \log^{\frac{3}{2}}(1/\delta)$, we have $T^{\frac{2}{3}}\sqrt{\log(1/\delta)} \geq T^{\frac{1}{3}}\log(1/\delta)$. This implies that

$$\mathbb{P}\left(\left|\sum_{t=1}^T (x^t)^\top e_t\right| > 6LT^{\frac{2}{3}}\sqrt{n\log(1/\delta)}\right) \leq \delta.$$

Similarly, we fix a set $S \subseteq [n]$ and consider the random variable $X_t = (\chi_S)^\top \hat{g}^t$ for all $1 \leq t \leq T$ that are adapted to the natural filtration generated by the iterates $\{x_t\}_{t \geq 1}$. By repeating the above argument with $\frac{\delta}{2^n}$, we have

$$\mathbb{P}\left(\left|\sum_{t=1}^T (\chi_S)^\top e_t\right| > 6LT^{\frac{2}{3}}\sqrt{n\log(2^n/\delta)}\right) \leq \tfrac{\delta}{2^n}.$$

By taking a union bound over the 2^n choices of S, we obtain that

$$\mathbb{P}\left(\left|\sum_{t=1}^T (\chi_S)^\top e_t\right| > 6LT^{\frac{2}{3}}\sqrt{n\log(2^n/\delta)}\right) \leq \delta, \quad \text{for any } S \subseteq [n].$$

Since $\sqrt{n\log(2^n/\delta)} \leq \sqrt{n^2 + n\log(1/\delta)}$, we have $\mathbf{I} \leq 12LT^{\frac{2}{3}}\sqrt{n^2 + n\log(1/\delta)}$ with probability at least $1 - 2\delta$.

Bounding II. Consider the random variables $X_t = \|\hat{g}^t\|^2$ for all $1 \leq t \leq T$ that are adapted to the natural filtration generated by the iterates $\{x^t\}_{t\geq 1}$. By Lemma C.1, we have $|X_t| \leq \frac{2(n+1)^2L^2}{\mu^2}$. Since $\mu = \frac{n}{T^{1/3}}$, we have $|X_t| \leq 8L^2T^{2/3}$. Further, we have

$$\mathbb{E}[X_t^2 \mid x_t] \leq \sum_{i=0}^n \frac{2(f_t(A_i^t))^4}{((1-\mu)\lambda_i^t + \frac{\mu}{n+1})^3} \leq \frac{2(n+1)^4L^4}{\mu^3} \leq 32nL^4T.$$

Applying Proposition C.3, we have

$$\mathbb{P}\left(\left|\sum_{t=1}^T \|\hat{g}^t\|^2 - \mathbb{E}[\|\hat{g}^t\|^2 \mid x^t]\right| > 8L^2T\sqrt{n\log(1/\delta)} + 4L^2T^{\frac{2}{3}}\log(1/\delta)\right) \leq \delta.$$

Since $T > \log^{\frac{3}{2}}(1/\delta)$, we have $T\sqrt{\log(1/\delta)} \geq T^{\frac{2}{3}}\log(1/\delta)$. This implies that

$$\mathbb{P}\left(\left|\sum_{t=1}^T \|\hat{g}^t\|^2 - \mathbb{E}[\|\hat{g}^t\|^2 \mid x^t]\right| > 12L^2T\sqrt{n\log(1/\delta)}\right) \leq \delta.$$

Therefore, we conclude that $\mathbf{II} \leq 12L^2T\sqrt{n\log(1/\delta)}$ with probability at least $1 - \delta$.

Putting these pieces together with Eq. (18) yields that

$$\sum_{t=1}^T (f_t)_L(x^t) \leq \left(\sum_{t=1}^T \tfrac{1}{\alpha}\bar{f}_t(S) - \beta\underline{f}_t(S)\right) + \tfrac{n}{2\eta} + \tfrac{4n^2L^2\eta T}{\mu} + 12LT^{\frac{2}{3}}\sqrt{n^2 + n\log(1/\delta)} + 6\eta L^2T\sqrt{n\log(1/\delta)},$$

with probability at least $1 - 3\delta$. □

C.2. Proof of Theorem 4.3

By the definition of the Lovász extension and λ^t, we have

$$(f_t)_L(x^t) = \sum_{i=1}^{n-1}(x^t_{\pi(i)} - x^t_{\pi(i+1)})f_t(A_i^t) + (1 - x^t_{\pi(1)})f_t(A_0^t) + x^t_{\pi(n)}f_t(A_n^t) = \sum_{i=0}^n \lambda_i^t f_t(A_i^t).$$

By the update formula of S^t, we have

$$\mathbb{E}[f_t(S^t) \mid x^t] - (f_t)_L(x^t) = \mu\sum_{i=0}^n \left(\frac{1}{n+1} - \lambda_i^t\right)f_t(A_i^t) \leq \mu\sum_{i=0}^n \left(\frac{1}{n+1} + \lambda_i^t\right)|f_t(A_i^t)|.$$

Since $f_t = \bar{f}_t - \underline{f}_t$ satisfy that $\bar{f}_t([n]) + \underline{f}_t([n]) \leq L$ for all $t \geq 1$ and $\bar{f}_t$ and $\underline{f}_t$ are both normalized and non-decreasing, we have

$$\mathbb{E}[f_t(S^t) \mid x^t] - (f_t)_L(x^t) \leq L\mu\sum_{i=0}^n \left(\frac{1}{n+1} + \lambda_i^t\right) = 2L\mu. \tag{19}$$

which implies that

$$\mathbb{E}[f_t(S^t)] - \mathbb{E}[(f_t)_L(x^t)] \leq 2L\mu.$$

Using the same argument as in Theorem 4.1, we have

$$\tfrac{1}{\alpha}(\bar{f}_t)_C(\chi_{S_\star^T}) + \beta(-\underline{f}_t)_C(\chi_{S_\star^T}) = \tfrac{1}{\alpha}f_t(S_\star^T) - \beta\underline{f}_t(S_\star^T), \quad \text{where } S_\star^T = \operatorname*{argmin}_{S\subseteq[n]}\sum_{t=1}^T f_t(S).$$

Putting these pieces together and letting $x = \chi_{S_\star^T}$ in the inequality of Lemma C.2 yields that

$$\sum_{t=1}^T \mathbb{E}[f_t(S^t)] \leq \left(\sum_{t=1}^T \tfrac{1}{\alpha}\bar{f}_t(S_\star^T) - \beta \underline{f}_t(S_\star^T)\right) + \tfrac{n}{2\eta} + \tfrac{4n^2L^2\eta T}{\mu} + 2LT\mu.$$

Plugging the choice of $\eta = \frac{1}{LT^{2/3}}$ and $\mu = \frac{n}{T^{1/3}}$ into the above inequality yields that $\mathbb{E}[R_{\alpha,\beta}(T)] = O(nT^{\frac{2}{3}})$ as desired.

We proceed to derive a high probability bound using Lemma C.4. Indeed, we first consider the case of $T < 2\log^{\frac{3}{2}}(1/\delta)$. Since $f_t = \bar{f}_t - \underline{f}_t$ satisfy that $\bar{f}_t([n]) + \underline{f}_t([n]) \leq L$ for all $t \geq 1$, we have

$$R_{\alpha,\beta}(T) \leq \sum_{t=1}^T f_t(S^t) - \sum_{t=1}^T (\tfrac{1}{\alpha}\bar{f}_t(S_\star^T) - \beta \underline{f}_t(S_\star^T)) \leq \left(1 + \tfrac{1}{\alpha} + \beta\right) LT = O(T^{\frac{2}{3}}\sqrt{\log(1/\delta)}).$$

For the case of $T \geq 2\log^{\frac{3}{2}}(1/\delta)$, we obtain by combining Lemma C.4 with Eq. (19) that

$$\begin{aligned}\sum_{t=1}^T \mathbb{E}[f_t(S^t) \mid x^t] &\leq \left(\sum_{t=1}^T \tfrac{1}{\alpha}\bar{f}_t(S) - \beta \underline{f}_t(S)\right) \\ &\quad + \tfrac{n}{2\eta} + 2nLT^{\frac{2}{3}} + 4nL^2\eta T^{\frac{4}{3}} + 12LT^{\frac{2}{3}}\sqrt{n^2 + n\log(1/\delta)} + 6\eta L^2 T\sqrt{n\log(1/\delta)},\end{aligned}$$

with probability at least $1-3\delta$. Then, it suffices to bound the term $\sum_{t=1}^T f_t(S^t) - \sum_{t=1}^T \mathbb{E}[f_t(S^t) \mid x^t]$ using Proposition C.3. Consider the random variables $X_t = f_t(S^t)$ for all $1 \leq t \leq T$ that are adapted to the natural filtration generated by the iterates $\{x^t\}_{t\geq 1}$. Since $f_t = \bar{f}_t - \underline{f}_t$ satisfy that $\bar{f}_t([n]) + \underline{f}_t([n]) \leq L$ for all $t \geq 1$, we have $|X_t| \leq L$. Further, we have $\mathbb{E}[X_t^2 \mid x_t] \leq L^2$. Applying Proposition C.3, we have

$$\mathbb{P}\left(\left|\sum_{t=1}^T f_t(S^t) - \mathbb{E}[f_t(S^t) \mid x^t]\right| > L\sqrt{2T\log(1/\delta)} + \tfrac{L}{2}\log(1/\delta)\right) \leq \delta.$$

Since $T > \log^{\frac{3}{2}}(1/\delta)$, we have $\sqrt{2T\log(1/\delta)} \geq \frac{1}{2}\log(1/\delta)$. This implies that

$$\mathbb{P}\left(\left|\sum_{t=1}^T f_t(S^t) - \mathbb{E}[f_t(S^t) \mid x^t]\right| > 3L\sqrt{T\log(1/\delta)}\right) \leq \delta.$$

Therefore, we conclude that $\sum_{t=1}^T f_t(S^t) - \sum_{t=1}^T \mathbb{E}[f_t(S^t) \mid x^t] \leq 3L\sqrt{T\log(1/\delta)}$ with probability at least $1-\delta$. Putting these pieces together yields that

$$\begin{aligned}\sum_{t=1}^T f_t(S^t) &\leq \left(\sum_{t=1}^T \tfrac{1}{\alpha}\bar{f}_t(S) - \beta \underline{f}_t(S)\right) + \tfrac{n}{2\eta} + 3L\sqrt{T\log(1/\delta)} + 2nLT^{\frac{2}{3}} \\ &\quad + 4nL^2\eta T^{\frac{4}{3}} + 12nLT^{\frac{2}{3}} + 12LT^{\frac{2}{3}}\sqrt{n\log(1/\delta)} + 6\eta L^2 T\sqrt{n\log(1/\delta)},\end{aligned}$$

with probability at least $1-4\delta$. Plugging the choice of $\eta = \frac{1}{LT^{2/3}}$ yields that

$$\sum_{t=1}^T f_t(S^t) \leq \left(\sum_{t=1}^T \tfrac{1}{\alpha}\bar{f}_t(S) - \beta \underline{f}_t(S)\right) + \tfrac{37}{2}nLT^{\frac{2}{3}} + 21LT^{\frac{2}{3}}\sqrt{n\log(1/\delta)},$$

with probability at least $1-4\delta$. Letting $S = S_\star^T = \text{argmin}_{S\subseteq[n]}\sum_{t=1}^T f_t(S)$ and changing δ to $\frac{\delta}{4}$ yields that $R_{\alpha,\beta}(T) = O(nT^{\frac{2}{3}} + \sqrt{n\log(1/\delta)}T^{\frac{2}{3}})$ with probability at least $1-\delta$ as desired.

D. Regret Analysis for Algorithm 3

In this section, we present several technical lemmas for analyzing the regret minimization property of Algorithm 3. We also give the missing proofs of Theorem 5.2.

D.1. Technical lemmas

We provide one technical lemma for Algorithm 3 which is analogues to Lemma B.2. It gives a key inequality for analyzing the regret minimization property of Algorithm 3. Note that the results in Lemma B.1 still hold true for the iterates $\{x^t\}_{t \geq 1}$ and $\{g^t\}_{t \geq 1}$ generated by Algorithm 3.

Lemma D.1 *Suppose that the iterates $\{x^t\}_{t \geq 1}$ are generated by Algorithm 3 and $x \in [0,1]^n$ and let $f_t = \bar{f}_t - \underline{f}_t$ satisfy that $\bar{f}_t([n]) + \underline{f}_t([n]) \leq L$ for all $t \geq 1$. Then, we have*

$$\sum_{t=1}^{T} \mathbb{E}[(f_t)_L(x^t)] \leq \left(\sum_{t=1}^{T} \tfrac{1}{\alpha}(\bar{f}_t)_C(x) + \beta(-\underline{f}_t)_C(x)\right) + \tfrac{n}{2\eta_{\bar{T}}} + \tfrac{L^2}{2}\left(\sum_{t=1}^{\bar{T}} \eta_t\right) + L^2\left(\sum_{t=1}^{\bar{T}}\sum_{s=q_t}^{t-1} \eta_s\right),$$

where $\bar{T} > 0$ in the above inequality satisfies that $q_{\bar{T}} = T$.

Proof. Using the same argument as in Lemma B.2, we have

$$(x^t - x)^\top g^{q_t} \leq \tfrac{1}{2\eta_t}\left(\|x - x^t\|^2 - \|x - x^{t+1}\|^2\right) + \tfrac{\eta_t}{2}\|g^{q_t}\|^2.$$

Since $f_t = \bar{f}_t - \underline{f}_t$ where $\bar{f}_t$ and $\underline{f}_t$ are both normalized and non-decreasing, $\bar{f}_t$ is α-weakly DR-submodular and $\underline{f}_t$ is β-weakly DR-supermodular, Proposition 3.1 implies that

$$(x^{q_t} - x)^\top g^{q_t} \geq (f_{q_t})_L(x^{q_t}) - \left(\tfrac{1}{\alpha}(\bar{f}_{q_t})_C(x) + \beta(-\underline{f}_{q_t})_C(x)\right).$$

By Lemma B.1, we have $\|g^t\| \leq L$ for all $t \geq 1$. Then, we have

$$(f_{q_t})_L(x^{q_t}) \leq \tfrac{1}{\alpha}(\bar{f}_{q_t})_C(x) + \beta(-\underline{f}_{q_t})_C(x) + \tfrac{1}{2\eta_t}\left(\|x - x^t\|^2 - \|x - x^{t+1}\|^2\right) + L\|x^{q_t} - x^t\| + \tfrac{\eta_t L^2}{2}. \tag{20}$$

Further, we have

$$\|x^{q_t} - x^t\| \leq \sum_{s=q_t}^{t-1} \eta_s \|g^s\| \leq L\left(\sum_{s=q_t}^{t-1} \eta_s\right). \tag{21}$$

Plugging Eq. (21) into Eq. (20) yields that

$$(f_{q_t})_L(x^{q_t}) \leq \tfrac{1}{\alpha}(\bar{f}_{q_t})_C(x) + \beta(-\underline{f}_{q_t})_C(x) + \tfrac{1}{2\eta_t}\left(\|x - x^t\|^2 - \|x - x^{t+1}\|^2\right) + \tfrac{\eta_t L^2}{2} + L^2\left(\sum_{s=q_t}^{t-1} \eta_s\right). \tag{22}$$

For a fixed horizon $T \geq 1$, we have $q_{\bar{T}} = T$ for some $\bar{T} \geq T$. Then, by summing up Eq. (22) over $t = 1, 2, \ldots, \bar{T}$ and using $\|x^t - x\| \leq \sqrt{n}$ for all $t \geq 1$ (cf. Lemma B.1) and that $\{\eta_t\}_{t \geq 1}$ is nonincreasing, we have

$$\sum_{t=1}^{\bar{T}} (f_{q_t})_L(x^{q_t}) \leq \left(\sum_{t=1}^{\bar{T}} \tfrac{1}{\alpha}(\bar{f}_{q_t})_C(x) + \beta(-\underline{f}_{q_t})_C(x)\right) + \tfrac{n}{2\eta_{\bar{T}}} + \tfrac{L^2}{2}\left(\sum_{t=1}^{\bar{T}} \eta_t\right) + L^2\left(\sum_{t=1}^{\bar{T}}\sum_{s=q_t}^{t-1} \eta_s\right).$$

Since $q_{\bar{T}} = T$ and our pooling policy is induced by a priority queue (note that $f_{q_t} = \bar{f}_{q_t} = \underline{f}_{q_t} = 0$ if $\mathcal{P}_t = \emptyset$), we have

$$\begin{aligned}
\sum_{t=1}^{\bar{T}} (f_{q_t})_L(x^{q_t}) &= \sum_{t=1}^{T} (f_t)_L(x^t), \\
\sum_{t=1}^{\bar{T}} \tfrac{1}{\alpha}(\bar{f}_{q_t})_C(x) + \beta(-\underline{f}_{q_t})_C(x) &= \sum_{t=1}^{T} \tfrac{1}{\alpha}(\bar{f}_t)_C(x) + \beta(-\underline{f}_t)_C(x).
\end{aligned}$$

Therefore, we conclude that

$$\sum_{t=1}^{T} (f_t)_L(x^t) \leq \left(\sum_{t=1}^{T} \tfrac{1}{\alpha}(\bar{f}_t)_C(x) + \beta(-\underline{f}_t)_C(x)\right) + \tfrac{n}{2\eta_T} + \tfrac{L^2}{2}\left(\sum_{t=1}^{\bar{T}} \eta_t\right) + L^2\left(\sum_{t=1}^{\bar{T}}\sum_{s=q_t}^{t-1} \eta_s\right).$$

Taking the expectation of both sides yields the desired inequality. $\square$

D.2. Proof of Theorem 5.2

By Héliou et al. (2020, Corollary 1), we have $t - q_t = o(t^\gamma)$ under Assumption 5.1; in particular, we have $t - q_t = o(t)$ and $q_t = \Theta(t)$. Since $q_{\bar{T}} = T$, we have $T = \Theta(\bar{T})$ which implies that $\bar{T} = \Theta(T)$. Recall that $\eta_t = \frac{\sqrt{n}}{L\sqrt{t^{1+\gamma}}}$, we have

$$\begin{aligned}
\frac{n}{2\eta_{\bar{T}}} &= \frac{L\sqrt{n\bar{T}^{1+\gamma}}}{2} = O(\sqrt{nT^{1+\gamma}}), \\
\frac{L^2}{2}\left(\sum_{t=1}^{\bar{T}} \eta_t\right) &= \frac{\sqrt{n}L}{2}\left(\sum_{t=1}^{\bar{T}} \frac{1}{\sqrt{t^{1+\gamma}}}\right) \leq \frac{\sqrt{n}L}{1-\gamma}\sqrt{\bar{T}^{1-\gamma}} = O(\sqrt{nT^{1-\gamma}}), \\
L^2\left(\sum_{t=1}^{\bar{T}}\sum_{s=q_t}^{t-1} \eta_s\right) &\leq L^2\left(\sum_{t=1}^{\bar{T}}(t-q_t)\eta_{q_t}\right) = O\left(\sqrt{n}L\sum_{t=1}^{\bar{T}}\frac{1}{\sqrt{t^{1-\gamma}}}\right) = O(L\sqrt{n\bar{T}^{1+\gamma}}) = O(\sqrt{nT^{1+\gamma}}),
\end{aligned}$$

Putting these pieces together with Lemma D.1 yields that

$$\sum_{t=1}^{T}\mathbb{E}[(f_t)_L(x^t)] - \left(\sum_{t=1}^{T}\tfrac{1}{\alpha}(\bar{f}_t)_C(x) + \beta(-\underline{f}_t)_C(x)\right) = O(\sqrt{nT^{1+\gamma}}). \tag{23}$$

By the definition of the Lovász extension, we have

$$(f_t)_L(x^t) = \sum_{i=1}^{n-1}(x^t_{\pi(i)} - x^t_{\pi(i+1)})f_t(A^t_i) + (1 - x^t_{\pi(1)})f_t(A^t_0) + x^t_{\pi(n)}f_t(A^t_n).$$

By the update formula, we have $\mathbb{E}[f_t(S^t) \mid x^t] = (f_t)_L(x^t)$ which implies that $\mathbb{E}[f_t(S^t)] = \mathbb{E}[(f_t)_L(x^t)]$. Further, by using the same argument as in Theorem 4.1, we have

$$\tfrac{1}{\alpha}(\bar{f}_t)_C(\chi_{S^T_\star}) + \beta(-\underline{f}_t)_C(\chi_{S^T_\star}) = \tfrac{1}{\alpha}\bar{f}_t(S^T_\star) - \beta\underline{f}_t(S^T_\star).$$

Putting these pieces together and letting $x = \chi_{S^T_\star}$ in Eq. (23) yields that

$$\sum_{t=1}^{T}\mathbb{E}[f_t(S^t)] - \left(\sum_{t=1}^{T}\tfrac{1}{\alpha}\bar{f}_t(S^T_\star) - \beta\underline{f}_t(S^T_\star)\right) = O(\sqrt{nT^{1+\gamma}}).$$

which implies that $\mathbb{E}[R_{\alpha,\beta}(T)] = O(\sqrt{nT^{1+\gamma}})$ as desired.

We proceed to derive a high probability bound using the concentration inequality in Proposition B.3. Indeed, we have

$$\mathbb{P}\left(\sum_{i=1}^{n} f_t(S^t) - \mathbb{E}\left[\sum_{i=1}^{n} f_t(S^t)\right] > \epsilon\right) \leq \exp\left(-\frac{\epsilon^2}{2nL^2}\right).$$

Equivalently, we have $\sum_{i=1}^n f_t(S^t) - \mathbb{E}[\sum_{i=1}^n f_t(S^t)] \leq L\sqrt{2T\log(1/\delta)}$ with probability at least $1-\delta$. This together with $\mathbb{E}[R_{\alpha,\beta}(T)] = O(\sqrt{nT^{1+\gamma}})$ yields that $R_{\alpha,\beta}(T) = O(\sqrt{nT^{1+\gamma}} + \sqrt{T\log(1/\delta)})$ with probability at least $1-\delta$.

E. Regret Analysis for Algorithm 4

In this section, we present several technical lemmas for analyzing the regret minimization property of Algorithm 4. We also give the missing proofs of Theorem 5.4.

E.1. Technical lemmas

We provide two technical lemmas for Algorithm 4 which are analogues to Lemma C.2 and C.4. It gives a key inequality for analyzing the regret minimization property of Algorithm 3. Note that the results in Lemma C.1 still hold true for the iterates $\{x^t\}_{t\geq 1}$ and $\{\hat{g}^t\}_{t\geq 1}$ generated by Algorithm 4.

Lemma E.1 *Suppose that the iterates* $\{x^t\}_{t\geq 1}$ *are generated by Algorithm 4 and* $x \in [0,1]^n$ *and let* $f_t = \bar{f}_t - \underline{f}_t$ *satisfy that* $\bar{f}_t([n]) + \underline{f}_t([n]) \leq L$ *for all* $t \geq 1$*. Then, we have*

$$\sum_{t=1}^T \mathbb{E}[(f_t)_L(x^t)] \leq \left(\sum_{t=1}^T \tfrac{1}{\alpha}(\bar{f}_t)_C(x) + \beta(-\underline{f}_t)_C(x)\right) + \tfrac{n}{2\eta_T} + 4n^2L^2\left(\sum_{t=1}^{\bar{T}} \tfrac{\eta_t}{\mu_{q_t}}\right) + 4nL^2\left(\sum_{t=1}^{\bar{T}}\sum_{s=q_t}^{t-1} \tfrac{\eta_s}{\mu_s}\right),$$

where $\bar{T} > 0$ *in the above inequality satisfies that* $q_{\bar{T}} = T$*.*

Proof. Using the same argument as in Lemma B.2, we have

$$(x^t - x)^\top \hat{g}^{q_t} \leq \tfrac{1}{2\eta_t}\left(\|x - x^t\|^2 - \|x - x^{t+1}\|^2\right) + \tfrac{\eta_t}{2}\|\hat{g}^{q_t}\|^2.$$

Since our pooling policy is induced by a priority queue, $\hat{g}^{q_t}$ has never been used before updating x^{t+1}. Thus, we have $\mathbb{E}[\hat{g}^{q_t} \mid x^t] = \mathbb{E}[\hat{g}^{q_t} \mid x^{q_t}]$ and $\mathbb{E}[\|\hat{g}^{q_t}\|^2 \mid x^t] = \mathbb{E}[\|\hat{g}^{q_t}\|^2 \mid x^{q_t}]$. By Lemma C.1, we have $\mathbb{E}[\hat{g}^{q_t} \mid x^{q_t}] = g^{q_t}$ and $\mathbb{E}[\|\hat{g}^{q_t}\|^2 \mid x^{q_t}] \leq \frac{8n^2L^2}{\mu_{q_t}}$ for all $t \geq 1$. Putting these pieces together yields that

$$(x^t - x)^\top g^{q_t} \leq \tfrac{1}{2\eta_t}\left(\|x - x^t\|^2 - \mathbb{E}[\|x - x^{t+1}\|^2 \mid x^t]\right) + \tfrac{4n^2L^2\eta_t}{\mu_{q_t}}.$$

Since $f_t = \bar{f}_t - \underline{f}_t$ where $\bar{f}_t$ and $\underline{f}_t$ are both normalized and non-decreasing, $\bar{f}_t$ is α-weakly DR-submodular and $\underline{f}_t$ is β-weakly DR-supermodular, Proposition 3.1 implies that

$$(x^{q_t} - x)^\top g^{q_t} \geq (f_{q_t})_L(x^{q_t}) - \left(\tfrac{1}{\alpha}(\bar{f}_{q_t})_C(x) + \beta(-\underline{f}_{q_t})_C(x)\right).$$

By Lemma B.1, we have $\|g^t\| \leq L$ for all $t \geq 1$. Then, we have

$$(f_{q_t})_L(x^{q_t}) \leq \tfrac{1}{\alpha}(\bar{f}_{q_t})_C(x) + \beta(-\underline{f}_{q_t})_C(x) + \tfrac{1}{2\eta_t}\left(\|x - x^t\|^2 - \mathbb{E}[\|x - x^{t+1}\|^2 \mid x^t]\right) + L\|x^{q_t} - x^t\| + \tfrac{4n^2L^2\eta_t}{\mu_{q_t}}. \quad (24)$$

Further, by Lemma C.1, we have

$$\|x^{q_t} - x^t\| \leq \sum_{s=q_t}^{t-1} \eta_s\|\hat{g}^s\| \leq 2(n+1)L\left(\sum_{s=q_t}^{t-1}\tfrac{\eta_s}{\mu_s}\right). \quad (25)$$

Plugging Eq. (25) into Eq. (24) yields that

$$\begin{aligned}(f_{q_t})_L(x^{q_t}) &\leq \tfrac{1}{\alpha}(\bar{f}_{q_t})_C(x) + \beta(-\underline{f}_{q_t})_C(x) \\ &\quad + \tfrac{1}{2\eta_t}\left(\|x - x^t\|^2 - \mathbb{E}[\|x - x^{t+1}\|^2 \mid x^t]\right) + \tfrac{4n^2L^2\eta_t}{\mu_{q_t}} + 4nL^2\left(\sum_{s=q_t}^{t-1}\tfrac{\eta_s}{\mu_s}\right).\end{aligned}$$

By using the same argument as in Lemma D.1, we have

$$\begin{aligned}\sum_{t=1}^T (f_t)_L(x^t) &\leq \left(\sum_{t=1}^T \tfrac{1}{\alpha}(\bar{f}_t)_C(x) + \beta(-\underline{f}_t)_C(x)\right) + \sum_{t=1}^{\bar{T}} \tfrac{1}{2\eta_t}\left(\|x - x^t\|^2 - \mathbb{E}[\|x - x^{t+1}\|^2 \mid x^t]\right) \\ &\quad + 4n^2L^2\left(\sum_{t=1}^{\bar{T}}\tfrac{\eta_t}{\mu_{q_t}}\right) + 4nL^2\left(\sum_{t=1}^{\bar{T}}\sum_{s=q_t}^{t-1}\tfrac{\eta_s}{\mu_s}\right).\end{aligned}$$

Taking the expectation of both sides of the above inequality and using $\|x^t - x\| \leq \sqrt{n}$ for all $t \geq 1$ (cf. Lemma C.1) and that $\{\eta_t\}_{t\geq 1}$ is nonincreasing yields the desired inequality. □

Then, we provide our second lemma which significantly generalizes Lemma E.1 for deriving the high-probability bounds.

Lemma E.2 *Suppose that the iterates* $\{x^t\}_{t\geq 1}$ *are generated by Algorithm 4 with* $\eta_t = \frac{1}{Lt^{(2+\gamma)/3}}$, $\mu_t = \frac{n}{t^{(1-\gamma)/3}}$ *and* $x \in [0,1]^n$ *and let* $f_t = \bar{f}_t - \underline{f}_t$ *satisfy that* $\bar{f}_t([n]) + \underline{f}_t([n]) \leq L$ *for all* $t \geq 1$. *Fixing a sufficiently small* $\delta \in (0,1)$ *and letting* $T > \log^{\frac{3}{2+\gamma}}(1/\delta)$. *Then, we have*

$$\begin{aligned}\sum_{t=1}^T (f_t)_L(x^t) \quad \leq \quad & \left(\sum_{t=1}^T \tfrac{1}{\alpha}\bar{f}_t(S) - \beta \underline{f}_t(S)\right) + \tfrac{n}{2\eta_{\bar{T}}} + 4n^2L^2\left(\sum_{t=1}^{\bar{T}} \tfrac{\eta_t}{\mu_{q_t}}\right) + 4nL^2\left(\sum_{t=1}^{\bar{T}}\sum_{s=q_t}^{t-1} \tfrac{\eta_s}{\mu_s}\right) \\ & + 12L\bar{T}^{\frac{4-\gamma}{6}}\sqrt{n^2 + n\log(1/\delta)} + 6L\sqrt{n\bar{T}\log(1/\delta)},\end{aligned}$$

with probability at least $1 - 3\delta$ *where* $\bar{T} > 0$ *in the above inequality satisfies that* $q_{\bar{T}} = T$.

Proof. Using the same argument as in Lemma E.1, we have

$$(x^t - x)^\top \hat{g}^{q_t} \leq \tfrac{1}{2\eta_t}\left(\|x - x^t\|^2 - \|x - x^{t+1}\|^2\right) + \tfrac{\eta_t}{2}\|\hat{g}^{q_t}\|^2,$$

and

$$(x^{q_t} - x)^\top g^{q_t} \geq (f_{q_t})_L(x^{q_t}) - \left(\tfrac{1}{\alpha}(\bar{f}_{q_t})_C(x) + \beta(-\underline{f}_{q_t})_C(x)\right).$$

For simplicity, we define $e_t = \hat{g}^t - g^t$. By Lemma B.1, we have $\|g^t\| \leq L$ for all $t \geq 1$. Then, we have

$$\begin{aligned}& (f_{q_t})_L(x^{q_t}) - \left(\tfrac{1}{\alpha}(\bar{f}_{q_t})_C(x) + \beta(-\underline{f}_{q_t})_C(x)\right) \\ \leq \quad & (x - x^t)^\top e_{q_t} + \tfrac{1}{2\eta_t}\left(\|x - x^t\|^2 - \|x - x^{t+1}\|^2\right) + L\|x^{q_t} - x^t\| + \tfrac{\eta_t}{2}\|\hat{g}^{q_t}\|^2.\end{aligned} \tag{26}$$

Plugging Eq. (25) into Eq. (26) yields that

$$\begin{aligned}& (f_{q_t})_L(x^{q_t}) - \left(\tfrac{1}{\alpha}(\bar{f}_{q_t})_C(x) + \beta(-\underline{f}_{q_t})_C(x)\right) \\ \leq \quad & (x - x^t)^\top e_{q_t} + \tfrac{1}{2\eta_t}\left(\|x - x^t\|^2 - \|x - x^{t+1}\|^2\right) + \tfrac{\eta_t}{2}(\|\hat{g}^t\|^2 - \mathbb{E}[\|\hat{g}^t\|^2 \mid x^t]) + \tfrac{4n^2L^2\eta_t}{\mu_{q_t}} + 4nL^2\left(\sum_{s=q_t}^{t-1}\tfrac{\eta_s}{\mu_s}\right).\end{aligned}$$

By using the same argument as in Lemma D.1, we have

$$\begin{aligned}\sum_{t=1}^T (f_t)_L(x^t) \quad \leq \quad & \left(\sum_{t=1}^T \tfrac{1}{\alpha}(\bar{f}_t)_C(x) + \beta(-\underline{f}_t)_C(x)\right) + \sum_{t=1}^{\bar{T}}(x - x^t)^\top e_{q_t} + \sum_{t=1}^{\bar{T}} \tfrac{\eta_t}{2}(\|\hat{g}^t\|^2 - \mathbb{E}[\|\hat{g}^t\|^2 \mid x^t]) \\ & + \tfrac{n}{2\eta_{\bar{T}}} + 4n^2L^2\left(\sum_{t=1}^{\bar{T}}\tfrac{\eta_t}{\mu_{q_t}}\right) + 4nL^2\left(\sum_{t=1}^{\bar{T}}\sum_{s=q_t}^{t-1}\tfrac{\eta_s}{\mu_s}\right).\end{aligned}$$

By the definition of the convex closure, we obtain that the convex closure of a set function f agrees with f on all the integer points (Dughmi, 2009, Page 4, Proposition 3.3). Letting $S \subseteq [n]$, we have $(\bar{f}_t)_C(\chi_S) = f_t(S)$ and $(-\underline{f}_t)_C(\chi_S) = -\beta\underline{f}_t(S)$ which implies that

$$\tfrac{1}{\alpha}(\bar{f}_t)_C(\chi_S) + \beta(-\underline{f}_t)_C(\chi_S) = \tfrac{1}{\alpha}f_t(S) - \beta\underline{f}_t(S).$$

Letting $x = \chi_S$, we have

$$\begin{aligned}\sum_{t=1}^T (f_t)_L(x^t) \quad \leq \quad & \left(\sum_{t=1}^T \tfrac{1}{\alpha}\bar{f}_t(S) - \beta\underline{f}_t(S)\right) + \underbrace{\sum_{t=1}^{\bar{T}}(\chi_S - x^t)^\top e_{q_t}}_{\mathbf{I}} + \underbrace{\left(\sum_{t=1}^{\bar{T}}\tfrac{\eta_t}{2}(\|\hat{g}^t\|^2 - \mathbb{E}[\|\hat{g}^t\|^2 \mid x^t])\right)}_{\mathbf{II}} \\ & + \tfrac{n}{2\eta_{\bar{T}}} + 4n^2L^2\left(\sum_{t=1}^{\bar{T}}\tfrac{\eta_t}{\mu_{q_t}}\right) + 4nL^2\left(\sum_{t=1}^{\bar{T}}\sum_{s=q_t}^{t-1}\tfrac{\eta_s}{\mu_s}\right).\end{aligned} \tag{27}$$

In what follows, we prove the high probability bounds for the terms **I** and **II** in the above inequality.

Bounding I. Consider the random variables $X_t = (x^t)^\top \hat{g}^{q_t}$ for all $1 \leq t \leq \bar{T}$ that are adapted to the natural filtration generated by the iterates $\{x_t\}_{t\geq 1}$. By Lemma C.1 and the Hölder's inequality, we have

$$|X_t| \leq \|\hat{g}^{q_t}\|_1 \|x^t\|_\infty \leq \frac{2(n+1)L}{\mu_t}.$$

Since $\mu = \frac{n}{t^{(1-\gamma)/3}}$, we have $|X_t| \leq 4L\bar{T}^{\frac{1-\gamma}{3}}$ for all $1 \leq t \leq \bar{T}$. Further, we have

$$\mathbb{E}[X_t^2 \mid x_t] \leq \mathbb{E}[\|\hat{g}^t\|_1^2 \|x^t\|_\infty^2 \mid x_t] \leq \frac{2(n+1)^2L^2}{\mu_t} \leq 8nL^2\bar{T}^{\frac{1-\gamma}{3}}.$$

Since $\mathbb{E}[\hat{g}^{q_t} \mid x^t] = g^{q_t}$ and $e_{q_t} = \hat{g}^{q_t} - g^{q_t}$, Proposition C.3 implies that

$$\mathbb{P}\left(\left|\sum_{t=1}^{\bar{T}} (x^t)^\top e_{q_t}\right| > 4L\bar{T}^{\frac{4-\gamma}{6}}\sqrt{n\log(1/\delta)} + 2L\bar{T}^{\frac{1-\gamma}{3}}\log(1/\delta)\right) \leq \delta.$$

Since $\bar{T} \geq T > \log^{\frac{3}{2+\gamma}}(1/\delta)$, we have $\bar{T}^{\frac{4-\gamma}{6}}\sqrt{\log(1/\delta)} \geq \bar{T}^{\frac{1-\gamma}{3}}\log(1/\delta)$. This implies that

$$\mathbb{P}\left(\left|\sum_{t=1}^{\bar{T}} (x^t)^\top e_{q_t}\right| > 6L\bar{T}^{\frac{4-\gamma}{6}}\sqrt{n\log(1/\delta)}\right) \leq \delta.$$

Similarly, we fix a set $S \subseteq [n]$ and consider the random variable $X_t = (\chi_S)^\top \hat{g}^t$ for all $1 \leq t \leq \bar{T}$ that are adapted to the natural filtration generated by the iterates $\{x_t\}_{t\geq 1}$. By repeating the above argument with $\frac{\delta}{2^n}$, we have

$$\mathbb{P}\left(\left|\sum_{t=1}^{\bar{T}} (\chi_S)^\top e_{q_t}\right| > 6L\bar{T}^{\frac{4-\gamma}{6}}\sqrt{n\log(2^n/\delta)}\right) \leq \frac{\delta}{2^n}.$$

By taking a union bound over the 2^n choices of S, we obtain that

$$\mathbb{P}\left(\left|\sum_{t=1}^{\bar{T}} (\chi_S)^\top e_{q_t}\right| > 6L\bar{T}^{\frac{4-\gamma}{6}}\sqrt{n\log(2^n/\delta)}\right) \leq \delta, \quad \text{for any } S \subseteq [n].$$

Since $\sqrt{n\log(2^n/\delta)} \leq \sqrt{n^2 + n\log(1/\delta)}$, we have $\mathbf{I} \leq 12L\bar{T}^{\frac{4-\gamma}{6}}\sqrt{n^2 + n\log(1/\delta)}$ with probability at least $1 - 2\delta$.

Bounding II. Consider the random variables $X_t = \frac{\eta_t}{2}\|\hat{g}^{q_t}\|^2$ for all $1 \leq t \leq \bar{T}$ that are adapted to the natural filtration generated by the iterates $\{x^t\}_{t\geq 1}$. By Lemma C.1, we have $|X_t| \leq \frac{(n+1)^2L^2\eta_t}{\mu_t^2}$. Since $\eta_t = \frac{1}{Lt^{(2+\gamma)/3}}$ and $\mu_t = \frac{n}{t^{(1-\gamma)/3}}$, we have $|X_t| \leq 4L$. Further, we have

$$\mathbb{E}[X_t^2 \mid x_t] \leq \frac{(n+1)^4L^4\eta_t^2}{2\mu_t^3} \leq 8nL^2.$$

Applying Proposition C.3, we have

$$\mathbb{P}\left(\left|\sum_{t=1}^{\bar{T}} \frac{\eta_t}{2}(\|\hat{g}^{q_t}\|^2 - \mathbb{E}[\|\hat{g}^{q_t}\|^2 \mid x^t])\right| > 4L\sqrt{n\bar{T}\log(1/\delta)} + 2L\log(1/\delta)\right) \leq \delta.$$

Since $\bar{T} \geq T > \log^{\frac{3}{2+\gamma}}(1/\delta)$, we have $\sqrt{\bar{T}\log(1/\delta)} \geq \log(1/\delta)$. This implies that

$$\mathbb{P}\left(\left|\sum_{t=1}^{\bar{T}} \frac{\eta_t}{2}(\|\hat{g}^{q_t}\|^2 - \mathbb{E}[\|\hat{g}^{q_t}\|^2 \mid x^t])\right| > 6L\sqrt{n\bar{T}\log(1/\delta)}\right) \leq \delta.$$

Therefore, we conclude that $\mathbf{II} \leq 6L\sqrt{n\bar{T}\log(1/\delta)}$ with probability at least $1 - \delta$. Putting these pieces together with Eq. (27) yields that

$$\begin{aligned}\sum_{t=1}^{T}(f_t)_L(x^t) \quad &\leq \quad \left(\sum_{t=1}^{T}\tfrac{1}{\alpha}\bar{f}_t(S) - \beta\underline{f}_t(S)\right) + \frac{n}{2\eta_{\bar{T}}} + 4n^2L^2\left(\sum_{t=1}^{\bar{T}}\frac{\eta_t}{\mu_{q_t}}\right) + 4nL^2\left(\sum_{t=1}^{\bar{T}}\sum_{s=q_t}^{t-1}\frac{\eta_s}{\mu_s}\right) \\ &+12L\bar{T}^{\frac{4-\gamma}{6}}\sqrt{n^2 + n\log(1/\delta)} + 6L\sqrt{n\bar{T}\log(1/\delta)},\end{aligned}$$

with probability at least $1 - 3\delta$. □

E.2. Proof of Theorem 5.4

By Héliou et al. (2020, Corollary 1), we have $t - q_t = o(t^\gamma)$ under Assumption 5.1; in particular, we have $t - q_t = o(t)$ and $q_t = \Theta(t)$. Since $q_{\bar{T}} = T$, we have $T = \Theta(\bar{T})$ which implies that $\bar{T} = \Theta(T)$. Recall that $\eta_t = \frac{1}{Lt^{(2+\gamma)/3}}$ and $\mu_t = \frac{n}{t^{(1-\gamma)/3}}$, we have

$$\begin{aligned}
\frac{n}{2\eta_{\bar{T}}} &= \frac{nL\bar{T}^{\frac{2+\gamma}{3}}}{2} = O(nT^{\frac{2+\gamma}{3}}), \\
4n^2L^2\left(\sum_{t=1}^{\bar{T}} \frac{\eta_t}{\mu_{q_t}}\right) &= 4nL\left(\sum_{t=1}^{\bar{T}} \frac{(q_t)^{\frac{1-\gamma}{3}}}{t^{\frac{2+\gamma}{3}}}\right) = O\left(nL\sum_{t=1}^{\bar{T}} \frac{1}{t^{\frac{1+2\gamma}{3}}}\right) = O(nL\bar{T}^{\frac{2-2\gamma}{3}}) = O(nT^{\frac{2-2\gamma}{3}}), \\
4nL^2\left(\sum_{t=1}^{\bar{T}}\sum_{s=q_t}^{t-1} \frac{\eta_s}{\mu_s}\right) &\leq 4L\left(\sum_{t=1}^{\bar{T}} (t-q_t)\frac{\eta_{q_t}}{\mu_t}\right) = O\left(L\sum_{t=1}^{\bar{T}} \frac{1}{t^{\frac{1-\gamma}{3}}}\right) = O(L\bar{T}^{\frac{2+\gamma}{3}}) = O(T^{\frac{2+\gamma}{3}}),
\end{aligned}$$

Putting these pieces together with Lemma E.1 yields that

$$\sum_{t=1}^{T} \mathbb{E}[(f_t)_L(x^t)] - \left(\sum_{t=1}^{T} \tfrac{1}{\alpha}(\bar{f}_t)_C(x) + \beta(-\underline{f}_t)_C(x)\right) = O(nT^{\frac{2+\gamma}{3}}). \tag{28}$$

By using the similar argument as in Theorem 4.3, we have

$$\mathbb{E}[f_t(S^t) \mid x^t] - (f_t)_L(x^t) \leq L\mu_t \sum_{i=0}^{n}\left(\tfrac{1}{n+1} + \lambda_i^t\right) = 2L\mu_t. \tag{29}$$

which implies that

$$\sum_{t=1}^{T} \mathbb{E}[f_t(S^t)] - \mathbb{E}[(f_t)_L(x^t)] \leq 2L\sum_{t=1}^{T} \mu_t = O(nT^{\frac{2+\gamma}{3}}).$$

Using the same argument as in Theorem 4.1, we have

$$\tfrac{1}{\alpha}(\bar{f}_t)_C(\chi_{S_\star^T}) + \beta(-\underline{f}_t)_C(\chi_{S_\star^T}) = \tfrac{1}{\alpha}f_t(S_\star^T) - \beta\underline{f}_t(S_\star^T), \quad \text{where } S_\star^T = \operatorname*{argmin}_{S\subseteq[n]} \sum_{t=1}^{T} f_t(S).$$

Putting these pieces together and letting $x = \chi_{S_\star^T}$ in Eq. (28) yields that

$$\sum_{t=1}^{T} \mathbb{E}[f_t(S^t)] - \left(\sum_{t=1}^{T} \tfrac{1}{\alpha}\bar{f}_t(S_\star^T) - \beta\underline{f}_t(S_\star^T)\right) = O(nT^{\frac{2+\gamma}{3}}).$$

which implies that $\mathbb{E}[R_{\alpha,\beta}(T)] = O(nT^{\frac{2+\gamma}{3}})$ as desired.

We proceed to derive a high probability bound using Lemma E.2. Indeed, we first consider the case of $T < 2\log^{\frac{3}{2+\gamma}}(1/\delta)$. Since $f_t = \bar{f}_t - \underline{f}_t$ satisfy that $\bar{f}_t([n]) + \underline{f}_t([n]) \leq L$ for all $t \geq 1$, we have

$$R_{\alpha,\beta}(T) \leq \sum_{t=1}^{T} f_t(S^t) - \sum_{t=1}^{T}(\tfrac{1}{\alpha}\bar{f}_t(S_\star^T) - \beta\underline{f}_t(S_\star^T)) \leq \left(1 + \tfrac{1}{\alpha} + \beta\right)LT = O(T^{\frac{4-\gamma}{6}}\sqrt{\log(1/\delta)}).$$

For the case of $T \geq 2\log^{\frac{3}{2+\gamma}}(1/\delta)$, we obtain by combining Lemma E.2 with Eq. (29) that

$$\begin{aligned}
\sum_{t=1}^{T} \mathbb{E}[f_t(S^t) \mid x^t] \leq &\left(\sum_{t=1}^{T} \tfrac{1}{\alpha}\bar{f}_t(S) - \beta\underline{f}_t(S)\right) + \tfrac{n}{2\eta_{\bar{T}}} + 2L\left(\sum_{t=1}^{T}\mu_t\right) + 4n^2L^2\left(\sum_{t=1}^{\bar{T}} \frac{\eta_t}{\mu_{q_t}}\right) \\
&+ 4nL^2\left(\sum_{t=1}^{\bar{T}}\sum_{s=q_t}^{t-1} \frac{\eta_s}{\mu_s}\right) + 12L\bar{T}^{\frac{4-\gamma}{6}}\sqrt{n^2 + n\log(1/\delta)} + 6L\sqrt{n\bar{T}\log(1/\delta)},
\end{aligned}$$

with probability at least $1-3\delta$. Then, it suffices to bound the term $\sum_{t=1}^T f_t(S^t) - \sum_{t=1}^T \mathbb{E}[f_t(S^t) \mid x^t]$ using Proposition C.3. By using the same argument as in Theorem 4.3, we have

$$\mathbb{P}\left(\left|\sum_{t=1}^T f_t(S^t) - \mathbb{E}[f_t(S^t) \mid x^t]\right| > 3L\sqrt{T\log(1/\delta)}\right) \leq \delta,$$

which implies that $\sum_{t=1}^T f_t(S^t) - \sum_{t=1}^T \mathbb{E}[f_t(S^t) \mid x^t] \leq 3L\sqrt{T\log(1/\delta)}$ with probability at least $1-\delta$. Putting these pieces together yields that

$$\begin{aligned}\sum_{t=1}^T f_t(S^t) \leq & \left(\sum_{t=1}^T \tfrac{1}{\alpha}\bar{f}_t(S) - \beta \underline{f}_t(S)\right) + \tfrac{n}{2\eta_{\bar{T}}} + 2L\left(\sum_{t=1}^T \mu_t\right) + 4n^2L^2\left(\sum_{t=1}^{\bar{T}} \tfrac{\eta_t}{\mu_{q_t}}\right) \\ & + 4nL^2\left(\sum_{t=1}^{\bar{T}}\sum_{s=q_t}^{t-1} \tfrac{\eta_s}{\mu_s}\right) + 3L\sqrt{T\log(1/\delta)} + 12L\bar{T}^{\frac{4-\gamma}{6}}\sqrt{n^2 + n\log(1/\delta)} + 6L\sqrt{n\bar{T}\log(1/\delta)},\end{aligned}$$

with probability at least $1-4\delta$. Plugging the choices of $\eta_t = \frac{1}{Lt^{(2+\gamma)/3}}$ and $\mu_t = \frac{n}{t^{(1-\gamma)/3}}$ and $\bar{T} = \Theta(T)$ yields that

$$\sum_{t=1}^T f_t(S^t) - \left(\sum_{t=1}^T \tfrac{1}{\alpha}\bar{f}_t(S) - \beta \underline{f}_t(S)\right) = O\left(nT^{\frac{2+\gamma}{3}} + \sqrt{n\log(1/\delta)}T^{\frac{4-\gamma}{6}}\right),$$

with probability at least $1-4\delta$. Letting $S = S_\star^T = \mathrm{argmin}_{S\subseteq[n]} \sum_{t=1}^T f_t(S)$ and changing δ to $\frac{\delta}{4}$ yields that $R_{\alpha,\beta}(T) = O(nT^{\frac{2+\gamma}{3}} + \sqrt{n\log(1/\delta)}T^{\frac{4-\gamma}{6}})$ with probability at least $1-\delta$ as desired.