Bandit Multi-linear DR-Submodular Maximization and Its Applications on Adversarial Submodular Bandits

Zongqi Wan [1 2] **Jialin Zhang** [1 2] **Wei Chen** [3] **Xiaoming Sun** [1 2] **Zhijie Zhang** [4]

Abstract

We investigate the online bandit learning of the monotone multi-linear DR-submodular functions, designing the algorithm `BanditMLSM` that attains $O(T^{2/3}\log T)$ of $(1-1/e)$-regret. Then we reduce submodular bandit with partition matroid constraint and bandit sequential monotone maximization to the online bandit learning of the monotone multi-linear DR-submodular functions, attaining $O(T^{2/3}\log T)$ of $(1-1/e)$-regret in both problems, which improve the existing results. To the best of our knowledge, we are the first to give a sublinear regret algorithm for the submodular bandit with partition matroid constraint. A special case of this problem is studied by Streeter et al. (2009). They prove a $O(T^{4/5})$ $(1-1/e)$-regret upper bound. For the bandit sequential submodular maximization, the existing work proves an $O(T^{2/3})$ regret with a suboptimal $1/2$ approximation ratio (Niazadeh et al., 2021).

1. Introduction

Research on multi-armed bandit problems has developed rapidly in the last two decades. After the classical finite-arm bandit and linear bandit were well-studied in both stochastic and adversarial settings, people were starting to consider more general bandit problems. Submodular bandit is such an object being considered due to its ability to characterize the diminishing property of reward function in realistic applications.

In submodular bandit, an optimizer/decision-maker needs to select a feasible subset each round, then a monotone submodular reward function of this round is determined stochastically or adversarially, and the decision maker obtains a reward according to the reward function of this round. In this paper, we consider the adversarial setting. In the adversarial submodular bandit literature, the meta-action technique proposed by Streeter & Golovin (2008) is commonly used to obtain a sublinear regret algorithm. This technique employs online optimizers for each offline step of an offline algorithm to mimic it in an online manner. This technique reaches $O(T^{2/3})$ $(1-1/e)$-regret with the cardinality constraint (Streeter & Golovin, 2008). Subsequently, Streeter et al. applied this technique to the assignment constraint (Streeter et al., 2009), which is a special partition matroid where the feasible set can only select one item from each partition. They obtained an $O(T^{4/5})$ $(1-1/e)$-regret in this situation. A recent work (Niazadeh et al., 2021) uses a Blackwell algorithm to turn offline greedy algorithms into online regret minimization algorithms. As an application, they reproduced the $O(T^{2/3})$ $(1-1/e)$-regret for the cardinality constraint.

In this paper, we present a different approach for the adversarial submodular bandit. We reduce the submodular bandit into a bandit multi-linear DR-submodular maximization problem. The DR-submodular function is a kind of non-convex function with theoretical guarantees in optimization, which has received much attention in recent years (Bian et al., 2017a;b; Niazadeh et al., 2020). There are several works considering the online full information or bandit feedback learning of the DR-submodular function (Chen et al., 2018; Zhang et al., 2019; Raut et al., 2020; Thang & Srivastav, 2021; Zhang et al., 2022a;b). DR-submodularity is inspired by the submodular set function, and we find it useful for designing submodular bandit algorithms due to its continuity. Specifically, we propose the function class called the multi-linear DR-submodular function. A multi-linear DR-submodular function is a DR-submodular function, and we additionally require it to be a multi-variable polynomial with the degree of each variable not exceeding 1. We propose the algorithm `BanditMLSM` for the bandit maximization of this function class and reach the $(1-1/e)$-regret of $\widetilde{O}(T^{2/3})$, which is far better than the $O(T^{5/6})$ $(1-1/e)$-regret bound achieved on the general bandit DR-submodular maximization problem (Niazadeh et al., 2021). Multi-linear DR-submodular function captures the property of the multi-

[1]Institute of Computing Technology, Chinese Academy of Sciences [2]University of Chinese Academy of Sciences [3]Microsoft Research [4]Center for Applied Mathematics of Fujian Province, School of Mathematics and Statistics, Fuzhou University. Correspondence to: Zhijie Zhang <zzhang@fzu.edu.cn>.

Proceedings of the 40^{th} International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023.

linear extension of a submodular set function. In fact, a multi-linear extension is a special case of multi-linear DR-submodular functions.

Our next goal is to reduce discrete submodular bandit to bandit multi-linear DR-submodular maximization problem. A natural idea is to run `BanditMLSM` on the multi-linear extension of a submodular set function. However, this idea fails to reduce the submodular bandit to bandit multi-linear DR-submodular maximization problem. This is because the function value of the multi-linear extension cannot be estimated unbiasedly while the constraint is not trivial, which is because the value of the multi-linear extension may require obtaining feedback on a set function value $f(S)$ where the set S is outside the constraint (e.g. the cardinality constraint). This is not allowed in the bandit feedback model. To address this issue, we propose a new kind of continuous extension which is also multi-linear DR-submodular. Then we run `BanditMLSM` on that extension. We try our continuous approach on submodular bandit with partition matroid constraint and bandit sequential submodular maximization, generalizing and improving the previous results, see Section 1.2.

More related works There is also some research on the stochastic submodular bandit. A model named linear submodular bandit has been studied (Yue & Guestrin, 2011; Chen et al., 2017). The model assumes that the reward function is a linear combination of several known submodular functions, only the weights of each submodular function are unknown to the decision maker, and the model requires the noisy marginal gain as the stochastic feedback. Many studies focus on the online influence maximization problem (Vaswani et al., 2015; Chen et al., 2016; Wang & Chen, 2017; Wu et al., 2019; Li et al., 2020; Zhang et al., 2022c), where the submodular reward function is induced by an information diffusion process on a social network. In this problem, different feedback models are studied. However, all the studies above assume extra information more than a *full-bandit* feedback model where the decision maker can only observe the reward of the action played. We only notice two works studying the full-bandit feedback model: (Nie et al., 2022; 2023). Nie et al. (2022) studied the bandit monotone submodular maximization with cardinality constraint, attaining $(1-1/e)$-regret of order $O(T^{2/3})$; Nie et al. (2023) design a framework which adapts an α-approximate offline algorithm into a stochastic bandit algorithm with $O(T^{2/3}(\log(T))^{1/3})$ α-regret. The framework needs the offline algorithm to be robust to small errors. Besides the above, Foster & Rakhlin (2021) studied the submodular contextual bandit.

Remark on the stochastic submodular bandit While Nie et al. (2022) make in their paper a weaker assumption that the online reward function need not be a monotone submodular but only need to be monotone submodular in expectation, we find our algorithm can also be applied in this setting even our adversarial submodular bandit model requires the reward function to be monotone submodular. The key observation is, when we apply the adversarial submodular bandit problem on a stochastic submodular bandit environment, we should see the expected submodular function rather than the stochastically realized reward function as the online reward function selected by the adversary, and see the stochastic feedback as an unbiased estimate of the true value of the expected function. We will explain this in Appendix F.

1.1. Bandit Optimization Model

Adversarial bandit optimization problems can be formalized as a repeated game between an optimizer and an adversary. The game lasts for T rounds and T is known to both players. In t-th round, the optimizer chooses an action x_t from an action set $\mathcal{K}$, then the adversary chooses a reward function $f_t \in \mathcal{F}$. The action set $\mathcal{K}$ and the reward function set $\mathcal{F}$ are determined by specific bandit problems. Generally, f_t maps $\mathcal{K}$ to a bounded interval $[0, M] \subseteq \mathbb{R}$. The optimizer gets reward $f_t(x_t)$ and it can only observe the value $f_t(x_t)$, which is called the *bandit* feedback model. Sometimes people also call it the *full-bandit* model to distinguish it from the semi-bandit model where the optimizer can observe more information, in this paper bandit and full-bandit are the same thing.

In this paper, we consider *oblivious* adversary, which means the reward functions f_t can not be adaptively selected according to $x_1, x_2, \ldots, x_t$. In other words, we can think the adversary selects $f_t \in \mathcal{F}$ for each $1 \leq t \leq T$ before the game starts and these functions are not revealed to the optimizer. Our goal is to design a strategy for the optimizer to minimize its cumulative α-*regret* during T rounds,

$$\mathcal{R}_\alpha(T) = \max_{\boldsymbol{x}^* \in \mathcal{K}} \mathbb{E}\left[\sum_{t=1}^{T} (\alpha f_t(x^*) - f_t(x_t))\right].$$

The action set $\mathcal{K}$ could be some structured set, maybe infinite or finite size. For the convenience of subsequent descriptions, we use $\mathcal{S}$ to denote the finite action set of the optimizer and use $\mathcal{K}$ to denote the infinite action set. Given $\mathcal{K}$ and $\mathcal{F}$, we call the game a $(\mathcal{K}, \mathcal{F})$-bandit.

When we consider the bandit multi-linear monotone DR-submodular maximization and bandit DR-submodular maximization in Section 3 and Section 4, we focus on the situation that $\mathcal{K}$ satisfies Assumption 1.1.

Assumption 1.1. We assume $\mathcal{K}$ is a compact convex subset of $\mathbb{R}^d$ containing $\mathbf{0}$, and $\mathcal{K} \subseteq D\mathbb{B}_d$ for some constant D,

where $\mathbb{B}_d$ is a d-dimensional unit ball.

Before further describing the model we are considering, we give several definitions.

Definition 1.2 (Monotonicity)**.** There is a natural partial order on $\mathbb{R}^d$. For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$, if $x_i \geq y_i \ \forall i \in [d]$, then $\boldsymbol{x} \geq \boldsymbol{y}$. For a function $f : \mathbb{R}^d \to \mathbb{R}$, if for any $\boldsymbol{x} \geq \boldsymbol{y}$, $f(\boldsymbol{x}) \geq f(\boldsymbol{y})$, we call f a monotone function.

Definition 1.3 (DR-submodularity)**.** Let $\mathcal{X} = \prod_{i=1}^{d} \mathcal{X}_i$ be a subset of $\mathbb{R}^d$, where $\mathcal{X}_i$ is an interval $[0, a_i]$. A continuous function $f : \mathcal{X} \to \mathbb{R}_+$ is called a DR-submodular function if for any $\boldsymbol{x} \geq \boldsymbol{y}$, $\lambda \in \mathbb{R}^+$, and the i-th base vector $\boldsymbol{e}_i$ for any $i \in [d]$,

$$f(\boldsymbol{x} + \lambda \boldsymbol{e}_i) - f(\boldsymbol{x}) \leq f(\boldsymbol{y} + \lambda \boldsymbol{e}_i) - f(\boldsymbol{y}).$$

Moreover, if f is second-order differentiable, then the DR-submodularity is equivalent to $\frac{\partial^2 f}{\partial x_i \partial x_j} \leq 0, \forall i, j \in [d]$.

Definition 1.4 (Multi-linearity)**.** We say function $f : \mathbb{R}^d \to \mathbb{R}$ a multi-linear function if f is polynomial of d variables, and for any variable x_i, the degree of x_i in each term of f is no more than 1.

Definition 1.5 (Lipschitz condition and smoothness)**.** Let $\|\cdot\|$ be the L^2-norm. For continuous differentiable function $f : \mathbb{R}^d \to \mathbb{R}$, if $|f(\boldsymbol{x}) - f(\boldsymbol{y})| \leq L_1 \|\boldsymbol{x} - \boldsymbol{y}\|$ for any $\boldsymbol{x}, \boldsymbol{y}$, we say f is L_1-lipschitz continuous. If $\|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\| \leq L_2 \|\boldsymbol{x} - \boldsymbol{y}\|$, we say f is L_2-smooth.

With the above definitions, we consider two reward function sets in Section 3 and Section 4:

- $\mathcal{F}_{DS}$: The set of monotone DR-submodular functions, which are L_1-lipschitz continuous, L_2-smooth, and $f(\mathbf{0}) = 0$.
- $\mathcal{F}_{MDS}$: The set of monotone multi-linear DR-submodular functions, which are L_1-lipschitz continuous, and $f(\mathbf{0}) = 0$.

1.2. Our Results

We are the first to consider the $(\mathcal{K}, \mathcal{F}_{MDS})$-bandit, i.e. Bandit Monotone Multi-linear DR-Submodular Maximization (BMMDSM). We observe that the gradient of multi-linear functions can be written as a linear combination of finite function values. Therefore, compared to the standard one-point gradient estimator proposed in (Flaxman et al., 2005) which is used in previous works (Zhang et al., 2019; Niazadeh et al., 2021), we propose a better one-point gradient estimator for monotone multi-linear DR-submodular functions. Along with other techniques including self-concordant barrier and non-oblivious technique, we propose the algorithm `BanditMLSM`, which achieves $(1 - 1/e)$-regret of $\widetilde{O}(T^{2/3})$. Here the $\widetilde{O}$ hides the $\log T$ factor.

As a secondary result, we also improved the $(1 - 1/e)$-regret of general $(\mathcal{K}, \mathcal{F}_{DS})$-bandit. This bandit is studied in (Zhang et al., 2019; Niazadeh et al., 2021), where they gave the $(1 - 1/e)$-regret bounds of $O(T^{8/9})$ and $O(T^{5/6})$ respectively. We proposed the algorithm `BanditDRSM` which achieves the $(1 - 1/e)$-regret of $\widetilde{O}(T^{3/4})$. Compared with their assumptions on functions, we add a new assumption that $f_t(\mathbf{0}) = 0$. Fortunately, this assumption is satisfied by many applications of DR-submodular maximization, including *optimal budget allocation with continuous assignment*, *senser placement*, *softmax extension* and so on (Bian et al., 2017a;b). For the constraint set, they assume $\mathcal{K}$ is downward closed while we do not make this assumption.

Our main contribution is to propose a continuous approach for combinatorial full-bandit, for example, the case where online functions are submodular set functions and the constraint is a partition matroid. Talking about the continuous approach, a natural idea is reducing combinatorial bandit to Bandit Monotone Multi-linear DR-submodular Maximization using the classical multi-linear extension technique. For a submodular set function g over the ground set $G = \{1, 2, \ldots, n\}$, its multi-linear extension $f : [0, 1]^n \to \mathbb{R}^+$ is defined as

$$f(\boldsymbol{x}) = \sum_{S \subseteq G} g(S) \prod_{i \in S} x_i \prod_{i \notin S} (1 - x_i).$$

From the definition of the multi-linear extension, we can see that it needs the value information of set function g over all subsets of G. However, in the submodular bandit, one can only take the action which satisfies the constraint, thus the algorithm can not explore the value information outside the constraint. As a result, Zhang et al. (2019) proved that it is impossible to construct an unbiased estimate of f and the gradient of f. That is to say, classical multi-linear extension is not a good candidate for our goal.

To overcome the above difficulties, we propose other continuous multi-linear DR-submodular extensions which require only the information of the feasible action. We select two submodular bandit problems to clarify our methodology: Bandit Monotone Submodular Maximization with Partition Matroid Constraint (BMSMPM) and Bandit Sequential Submodular Maximization (BSSM). The results are summarized in Table 1. Previous works have studied two special cases of BMSMPM: cardinality constraint (Streeter & Golovin, 2008) and the assignment problem (Streeter et al., 2009). We improve the regret bound of the bandit assignment problem and reproduce an $\widetilde{O}(T^{2/3})$ $(1 - 1/e)$-regret for cardinality constraint. To our best knowledge, we are the first to give a sublinear $(1 - 1/e)$-regret algorithm for bandit monotone submodular maximization with general partition matroid

Table 1. Our results comparing to the previous results.

Problem	Our α and regret	Previous α and regret
Bandit DR-submodular maximization results		
Bandit Multi-linear Monotone DR-Submodular Maximization	Theorem 3.3 $1-1/e, O\left(d^{4/3}T^{2/3}\log(T)\right)$	\
Bandit Monotone DR-Submodular Maximization	Theorem 4.1 $1-1/e, O\left(d^{1/2}T^{3/4}\log(T)\right)$	(Niazadeh et al., 2021) $1-1/e, O(d(\log(d))^{1/6}T^{5/6})$
Applications on adversarial submodular bandits		
Bandit Assignment Problem	Corollary 5.4† $1-1/e, O\left((\lvert G\rvert)^{5/3}T^{2/3}\log(T)\right)$	(Streeter et al., 2009) $1-1/e, O\left(T^{4/5}\right)$‡
Bandit Monotone Submodular Maximization over Partition Matroid	Corollary 5.4 $1-1/e, O\left(\left(\sum_{k=1}^{K} r_k\lvert G_k\rvert\right)^{5/3}T^{2/3}\log T\right)$	\
Bandit Sequential Submodular Maximization¶	Corollary 5.6 $1-1/e, O\left(\lvert G\rvert^{10/3}T^{2/3}\log(T)\right)$	(Niazadeh et al., 2021) $1/2, O(\lvert G\rvert^{5/3}(\log(\lvert G\rvert))^{1/3}T^{2/3})$

† Bandit assignment problem is a special case of Bandit Monotone Submodular Maximization over Partition Matroid where $r_k = 1$ and $G_k = G$, so this regret bound can be directly derived from Corollary 5.4. ‡ In the original paper (Streeter et al., 2009), the regret is written in the form which contains the optimal cumulative reward value, which will continue to be bounded to T usually, leading to a bad dependent on T. So we re-trade off their regret and write it in terms of T so that it can be compared with our regret bound. ¶ Compared with the setting in (Niazadeh et al., 2021), we actually add a new assumption that there is a dummy element in the ground set that always has 0 marginal gain. This assumption can be satisfied easily in realistic applications, see Section 5.3.

constraint. BSSM is motivated by maximizing user engagement on online retailing platforms. It is first studied in (Niazadeh et al., 2021), and their algorithm attains $O(T^{2/3})$ $1/2$-regret while the $1/2$ approximation ratio is not tight. We improve this result to $\widetilde{O}(T^{2/3})$ $(1-1/e)$-regret, leading to the tight approximation ratio.

In summary, we make the following contributions:

- We are the first to study the bandit maximization of multi-linear monotone DR-submodular functions and propose a $\widetilde{O}(T^{2/3})$ $(1-1/e)$-regret algorithm.
- We improve the previous result of bandit maximization of general monotone DR-submodular functions to $\widetilde{O}(T^{3/4})$ $(1-1/e)$-regret by better exploiting the smoothness.
- We propose a continuous approach to reducing combinatorial bandit to multi-linear DR-submodular bandit. Using this continuous approach, we propose the first sublinear regret algorithm for submodular bandit with partition matroid constraint, which also improves the result of a previous work (Streeter et al., 2009) that studied the special case of this problem. We also improve the previous approximation ratio of Bandit Sequential Submodular Maximization from $1/2$ to tight $1-1/e$.

2. Preliminary

2.1. Regularized Follow the Leader and Self-Concordant Functions

Regularized Follow The Leader(RFTL) is a commonly used algorithm for online optimization. While applying on a sequence of vector $\{\boldsymbol{g}_q\}_{q=1}^{Q}$ with constraint $\mathcal{K}$, RFTL outputs a sequence of point $\{\boldsymbol{x}_q\}_{q=1}^{Q}$, where

$$\boldsymbol{x}_1 = \operatorname*{argmin}_{\boldsymbol{x}\in\mathcal{K}} \Phi(\boldsymbol{x})$$
$$\boldsymbol{x}_{q+1} = \operatorname*{argmin}_{\boldsymbol{x}\in\mathcal{K}} \left(\eta\sum_{s=1}^{q}\langle -\boldsymbol{g}_s, \boldsymbol{x}\rangle + \Phi(\boldsymbol{x})\right).$$

Here $\Phi(\boldsymbol{x})$ is an arbitrary regularizer, η is a parameter. In this paper, we use a self-concordant barrier of $\mathcal{K}$ as the regularizer of RFTL. Self-concordant barrier was first proposed in convex optimization literature, and it was introduced to the bandit optimization problem in (Abernethy et al., 2008).

Definition 2.1 (Self-concordant Barrier (Hazan et al., 2016))**.** Let $\mathcal{K} \in \mathbb{R}^d$ be a convex set with non empty interior $\text{int}(\mathcal{K})$. We call the function $\Phi : \text{int}(\mathcal{K}) \longrightarrow \mathbb{R}$ a ν-self-concordant barrier of $\mathcal{K}$ if:
(1) Φ is three-times continuously differentiable, convex, and approaches infinity along any sequence of points approaching the boundary of $\mathcal{K}$;
(2) For every $\boldsymbol{h} \in \mathbb{R}^d$ and $\boldsymbol{x} \in \text{int}(\mathcal{K})$, the following holds:$|\nabla^3\Phi(\boldsymbol{x})[\boldsymbol{h},\boldsymbol{h},\boldsymbol{h}]| \leq 2(\nabla^2\Phi(\boldsymbol{x})[\boldsymbol{h},\boldsymbol{h}])^{3/2}$,

$|\nabla\Phi(x)[\boldsymbol{h}]| \leq \nu^{1/2}(\nabla^2\Phi(\boldsymbol{x})[\boldsymbol{h},\boldsymbol{h}])^{1/2}$. where the third-order differential is defined as $\nabla^3\Phi(\boldsymbol{x})[\boldsymbol{h},\boldsymbol{h},\boldsymbol{h}] := \frac{\partial^3}{\partial t_1\partial t_2\partial t_3}\Phi(x+t_1\boldsymbol{h}+t_2\boldsymbol{h}+t_3\boldsymbol{h})|_{t_1=t_2=t_3=0}$.

Definition 2.2 (Local norm)**.** The Hessian of self-concordant barrier induces a local norm at every $x \in \text{int}(\mathcal{K})$, denoted as $\|\cdot\|_{\Phi,x}$. We denote its dual norm as $\|\cdot\|_{\Phi,\boldsymbol{x},*}$. For any $\boldsymbol{v} \in \mathbb{R}^d$,

$$\|\boldsymbol{v}\|_{\Phi,\boldsymbol{x}} = \sqrt{\boldsymbol{v}^T\nabla^2\Phi(\boldsymbol{x})\boldsymbol{v}}$$
$$\|\boldsymbol{v}\|_{\Phi,\boldsymbol{x},*} = \sqrt{\boldsymbol{v}^T(\nabla^2\Phi(\boldsymbol{x}))^{-1}\boldsymbol{v}}.$$

The following theorem is proved in (Abernethy et al., 2008). It shows that, if we set the regularizer to be a self-concordant barrier of $\mathcal{K}$ and the algorithm can access the unbiased estimator of g_t, then the regret of the generated solution sequence $\{x_q\}_{q=1}^Q$ can be bounded in terms of the local norm of the estimator.

Theorem 2.3 ((Abernethy et al., 2008))**.** *Let $\mathcal{K}$ be a convex set, $\Phi(x)$ be a self-concordant barrier on $\mathcal{K}$, $\{\widetilde{\boldsymbol{g}}_q\}_{q=1}^Q$ be a vector sequence. If $\widetilde{g}_q$ is an unbiased estimation of g_q, then running RFTL on vector sequence $\widetilde{g}_q$ with $\Phi(x)$ as the regularizer will produce a sequence of point $\{\boldsymbol{x}_q\}_{q=1}^Q, x_q \in \mathcal{K}$. For $\{\boldsymbol{x}_q\}_{q=1}^Q$ and any $\boldsymbol{y} \in \mathcal{K}$, we have*

$$\sum_{q=1}^{Q}\mathbb{E}\left[\langle\boldsymbol{g}_q, \boldsymbol{y}-\boldsymbol{x}_q\rangle\right] \leq \eta\sum_{q=1}^{Q}\mathbb{E}\left[\|\widetilde{\boldsymbol{g}}_q\|^2_{\Phi,\boldsymbol{x}_q,*}\right] + \frac{\Phi(\boldsymbol{y})-\Phi(\boldsymbol{x}_1)}{\eta}$$

2.2. Ellipsoid Gradient Estimator

Ellipsoid gradient estimator is proposed in (Abernethy et al., 2008), where the authors use it along with the tool from Section 2.1 to design an $\widetilde{O}(\sqrt{T})$ regret algorithm for bandit linear optimization. For a continuous function $f : \mathbb{R}^d \to \mathbb{R}$ and an invertible matrix $\boldsymbol{H} \in \mathbb{R}^{d\times d}$, we define the $\boldsymbol{H}$-smoothed version of f.

Definition 2.4 ($\boldsymbol{H}$-smoothed function)**.** For function $f(\boldsymbol{x}) : \mathbb{R}^d \to \mathbb{R}$ and invertible matrix $\boldsymbol{H} \in \mathbb{R}^{d\times d}$, we call $f^{\boldsymbol{H}}(\boldsymbol{x})$ an $\boldsymbol{H}$-smoothed version of $f(\boldsymbol{x})$, where

$$f^{\boldsymbol{H}}(\boldsymbol{x}) = \mathop{\mathbb{E}}_{\boldsymbol{v}\sim\mathbb{B}_d}\left[f(\boldsymbol{x}+\boldsymbol{H}\boldsymbol{v})\right].$$

Here $\boldsymbol{v} \sim \mathbb{B}_d$ means that $\boldsymbol{v}$ is sampled from the unit ball $\mathbb{B}_d$ uniformly at random.

There is a surprising fact that there is an unbiased estimator of $\nabla f^{\boldsymbol{H}}(\boldsymbol{x})$ for any $\boldsymbol{x}$, and the estimator uses only one query to the value oracle of f.

Lemma 2.5 (Ellipsoid estimator (Abernethy et al., 2008))**.** *Let $\boldsymbol{H} \in \mathcal{R}^{d\times d}$ be an invertible matrix, $f(\boldsymbol{x}) : \mathbb{R}^d \to \mathbb{R}$ be an arbitrary function. Then*

$$\nabla f^{\boldsymbol{H}}(\boldsymbol{x}) = d\mathop{\mathbb{E}}_{\boldsymbol{v}\sim\mathbb{S}_{d-1}}\left[f(\boldsymbol{x}+\boldsymbol{H}\boldsymbol{v})\boldsymbol{H}^{-1}\boldsymbol{v}\right].$$

Here $\boldsymbol{v} \sim \mathbb{S}_{d-1}$ means that $\boldsymbol{v}$ is sampled from the $(d-1)$-dimensional unit sphere $\mathbb{S}_{d-1}$ uniformly at random.

For linear f, $f^{\boldsymbol{H}}(\boldsymbol{x}) = f(\boldsymbol{x})$, so Lemma 2.5 gives a one-sample unbiased estimator of the gradient of the linear function. The ellipsoid gradient estimator is usually used along with RFTL with a self-concordant regularizer Φ of $\mathcal{K}$. When the invertible matrix $\boldsymbol{H}$ is set to be $(\nabla^2\Phi(\boldsymbol{x}))^{-1/2}$ and $\boldsymbol{x} \in \text{int}(\mathcal{K})$, the sampled action $\boldsymbol{x}+\boldsymbol{H}\boldsymbol{v}$ is located in the surface of a so-called **Dikin ellipsoid** centered at $\boldsymbol{x}$, i.e. $\{\boldsymbol{x}' \mid \|\boldsymbol{x}'-\boldsymbol{x}\|_{\Phi,\boldsymbol{x}} \leq 1\}$. The fact that Dikin ellipsoid is entirely contained in $\mathcal{K}$ is useful for reducing regret.

2.3. Non-oblivious Techniques for Monotone DR-Submodular Maximization

The non-oblivious technique was first proposed to improve the approximation ratio of the solution returned by a local search algorithm. The idea is to run a local search on an auxiliary function rather than the original objective, and the local optima of the auxiliary function have a higher approximation ratio, thus the search algorithm will return a better solution.

In monotone submodular maximization literature, Filmus & Ward (2014) improved the approximation ratio of the greedy algorithm to $1-1/e$ using the non-oblivious technique. Zhang et al. (2022a) generalized this result to the continuous DR-submodular maximization problem, improving the approximation ratio of projected gradient ascent to $1-1/e$. For a monotone DR-submodular function $f(\boldsymbol{x})$ satisfying $f(\boldsymbol{0}) = 0$, they consider following auxiliary function,

$$F(\boldsymbol{x}) = \int_0^1 \frac{e^{z-1}}{z}f(z\cdot\boldsymbol{x})dz. \tag{1}$$

We need the following lemma about the auxiliary function proved in their paper.

Lemma 2.6 (Auxiliary function(Zhang et al., 2022a))**.** *Let f be a monotone DR-submodular function defined on $\mathcal{X}$ and $f(\boldsymbol{0}) = 0$, $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$. Let F be defined as (1). Then*

$$\nabla F(\boldsymbol{x}) = \int_0^1 e^{z-1}\nabla f(z\cdot\boldsymbol{x})dz \tag{2}$$

and the following inequality holds,

$$\langle\boldsymbol{y}-\boldsymbol{x}, \nabla F(\boldsymbol{x})\rangle \geq (1-1/e)f(\boldsymbol{y}) - f(\boldsymbol{x}).$$

Algorithm 1 `BanditMLSM`(η, L, Φ)

Input: block size L, block number $Q = T/L$, learning rate η, self-concordant barrier Φ

initiate $\boldsymbol{x}_1 \in \text{int}(\mathcal{K})$ such that $\nabla\Phi(\boldsymbol{x}_1) = 0$
for $q = 1, 2, \ldots, Q$ **do**
  Draw $t_q \sim \text{Unif}\{(q-1)L+1, (q-1)L+2, \ldots, qL\}$
  for $t = (q-1)L+1, (q-1)L+2, \ldots, qL$ **do**
    if $t = t_q$ **then**
      $\boldsymbol{H}_q = \left(\nabla^2\Phi(\boldsymbol{x}_q)\right)^{-1/2}$
      sample z_q from $\mathbf{Z}$ where $P(\mathbf{Z} \le z) = \int_0^z \frac{e^{u-1}}{1-e^{-1}}\mathbb{I}\left[u \in [0,1]\right] du$
      draw $\boldsymbol{v}_q \sim \mathbb{S}_{d-1}$
      draw $\boldsymbol{u}_q$ from $\{\mathbf{0}, \boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_d\}$ with probability: $\Pr(\boldsymbol{u}_q = \mathbf{0}) = \frac{1}{2}$, $\Pr(\boldsymbol{u}_q = \boldsymbol{e}_i) = \frac{1}{2d}$
      play $\boldsymbol{y}_{t_q} = z_q \cdot \boldsymbol{x}_q + z_q\langle \boldsymbol{H}_q\boldsymbol{v}_q, \boldsymbol{u}_q\rangle \boldsymbol{u}_q$
      Set $\widetilde{l}_q(\boldsymbol{H}_q\boldsymbol{v}_q)$ as (5)
      $\widetilde{\nabla}\overline{F}_q(\boldsymbol{x}_q) \leftarrow d \cdot \widetilde{l}_q(\boldsymbol{H}_q\boldsymbol{v}_q)\boldsymbol{H}_q^{-1}\boldsymbol{v}_q$
      $\boldsymbol{x}_{q+1} \leftarrow \operatorname{argmin}_{\boldsymbol{x}\in\mathcal{K}} \sum_{s=1}^{q}\langle -\eta\widetilde{\nabla}F_s(\boldsymbol{x}_s), \boldsymbol{x}\rangle + \Phi(\boldsymbol{x})$
    else
      play $\boldsymbol{y}_t = \boldsymbol{x}_q$
    end if
  end for
end for

3. Bandit Monotone Multi-linear DR-Submodular Maximization

In this section, we present our algorithm `BanditMLSM` for BMMDSM. The pseudo-code is shown in Algorithm 1. For some technical reason we will explain later, we divide the whole T rounds into Q equal-size blocks, and each block has L consecutive rounds. Here Q and L are to be determined later, $L = T/Q$. without loss of generality, we assume both L and Q are integers. We define the average function $\overline{f}_q(\boldsymbol{x})$ of each block,

$$\overline{f}_q(\boldsymbol{x}) = \frac{1}{L}\sum_{t=(q-1)L+1}^{qL} f_t(\boldsymbol{x}). \tag{3}$$

Let $\overline{F}_q(\boldsymbol{x})$ be the auxiliary function of $\overline{f}_q(\boldsymbol{x})$,

$$\overline{F}_q(\boldsymbol{x}) = \int_0^1 \frac{e^{z-1}}{zL}\sum_{t=(q-1)L+1}^{qL} f_t(z\cdot\boldsymbol{x})dz. \tag{4}$$

In high level, `BanditMLSM` runs RFTL with a self-concordant regularizer $\Phi(\boldsymbol{x})$ on the vector sequence $\{\nabla\overline{F}_q(\boldsymbol{x}_q)\}_{q=1}^Q$ and controls the regret w.r.t. the linear function sequence $\{l_q\}_{q=1}^Q$ where $l_q(\boldsymbol{u}) := \langle \boldsymbol{u}, \nabla\overline{F}_q(\boldsymbol{x}_q)\rangle$. Now the question is how to estimate $\nabla\overline{F}_q(\boldsymbol{x}_q)$. Recall Lemma 2.5, we can estimate $\nabla\overline{F}_q(\boldsymbol{x}_q) = \nabla l_q(\mathbf{0})$ with the ellipsoid estimator by querying one function value of $l_q(\boldsymbol{u})$. That is, we fix an invertible matrix $\boldsymbol{H}_q = (\nabla^2\Phi(\boldsymbol{x}_q))^{-1/2}$, sample a random direction $\boldsymbol{v}_q$ in the $(d-1)$-dimensional sphere, then query $l_q(\boldsymbol{H}_q\boldsymbol{v}_q)$, and return $\widetilde{\nabla}\overline{F}_q(\boldsymbol{x}_q) := d\cdot l_q(\boldsymbol{H}_q\boldsymbol{v}_q)\boldsymbol{H}_q^{-1}\boldsymbol{v}_q$ as the estimate.

The problem here is that we cannot query l_q directly. The algorithm can only query the function value of f_t by playing the corresponding action in round t. We construct the unbiased estimator of $l_q(\boldsymbol{H}_q\boldsymbol{v}_q)$ as follows. First, we sample a uniformly random $t_q \in [(q-1)L+1, qL] \cap \mathbb{Z}$ and sample z_q from the distribution Z where $\Pr(Z \le z) = \int_0^z \frac{e^{u-1}}{1-e^{-1}}\mathbb{I}\left[u\in[0,1]\right]du$. Then we pick a vector $\boldsymbol{u}_q$ from the set $\{\mathbf{0}, \boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_d\}$ following the distribution: $\Pr(\boldsymbol{u}_q = \mathbf{0}) = \frac{1}{2}$, $\Pr(\boldsymbol{u}_q = \boldsymbol{e}_i) = \frac{1}{2d}$. Then we play $\boldsymbol{y}_{t_q} := z_q\boldsymbol{x}_q + z_q\langle \boldsymbol{H}_q\boldsymbol{v}_q, \boldsymbol{u}_q\rangle\boldsymbol{u}_q$ in round t_q to obtain the feedback $f_{t_q}(\boldsymbol{y}_{t_q})$. We replace $l_q(\boldsymbol{H}_q\boldsymbol{v}_q)$ with an estimate

$$\widetilde{l}_q(\boldsymbol{H}_q\boldsymbol{v}_q) := \begin{cases} -2(1-1/e)\dfrac{d}{z_q}\cdot f_{t_q}(\boldsymbol{y}_{t_q}) & \text{if } \boldsymbol{u}_q = \mathbf{0}, \\ 2(1-1/e)\dfrac{d}{z_q}\cdot f_{t_q}(\boldsymbol{y}_{t_q}) & \text{if } \boldsymbol{u}_q \neq \mathbf{0}. \end{cases} \tag{5}$$

If $z_q = 0$, we define $\widetilde{l}_q(\boldsymbol{H}_q\boldsymbol{v}_q) := 0$. The following Lemma shows that $\widetilde{l}_q(\boldsymbol{H}_q\boldsymbol{v}_q)$ is an unbiased estimator of $l_q(\boldsymbol{H}_q\boldsymbol{v}_q)$. Its proof is deferred to Appendix B. In the rounds other than t_q in block q, we play $\boldsymbol{y}_t := \boldsymbol{x}_q$ output by RFTL at the end of $(q-1)$-th block to exploit the regret bound of RFTL.

Lemma 3.1. *Let $\mathcal{H}_{q-1}$ be the history of the algorithm in the first q blocks, that is, the realization of $t_s, z_s, \boldsymbol{v}_s, \boldsymbol{u}_s, \forall s \le q$. Then $\mathbb{E}[\widetilde{l}_q(\boldsymbol{H}_q\boldsymbol{v}_q) \mid \mathcal{H}_{q-1}, \boldsymbol{v}_q] = l_q(\boldsymbol{H}_q\boldsymbol{v}_q)$.*

So $\widetilde{\nabla}\overline{F}_q(\boldsymbol{x})$ is actually defined as

$$\widetilde{\nabla}\overline{F}_q(\boldsymbol{x}) := d\cdot\widetilde{l}_q(\boldsymbol{H}_q\boldsymbol{v}_q)\boldsymbol{H}_q^{-1}\boldsymbol{v}_q. \tag{6}$$

We show that $\widetilde{\nabla}\overline{F}_q(\boldsymbol{x}_q)$ is an unbiased estimator of $\nabla\overline{F}_q(\boldsymbol{x}_q)$, and its dual local norm is $O(d^4)$ in the following lemma. The proof is deferred to Appendix B.

Lemma 3.2. *The following properties hold for $\widetilde{\nabla}\overline{F}_q(\boldsymbol{x}_q)$:*

(i) $\mathbb{E}\left[\widetilde{\nabla}\overline{F}_q(\boldsymbol{x}_q) \mid \mathcal{H}_{q-1}\right] = \nabla\overline{F}_q(\boldsymbol{x}_q)$,

(ii) $\mathbb{E}\left[\|\widetilde{\nabla}\overline{F}_q(\boldsymbol{x}_q)\|^2_{\boldsymbol{x}_q,*} \mid \mathcal{H}_{q-1}\right] \le 4(1-1/e)^2L_1^2D^2d^4$.

Note that, to estimate $\nabla\overline{F}_q(\boldsymbol{x}_q)$, we must sample an action that is far from $\boldsymbol{x}_q$. This means we cannot do exploration and exploitation at the same time, which is different from the linear bandit. This is the reason why previous works on bandit submodular maximization and our work divide

rounds into blocks. We need to do the exploitation in most of the rounds of a block to maintain the regret bound.

Now recall Theorem 2.3, running RFTL with a self-concordant barrier of $\mathcal{K}$ will generate a series of action $\{x_q\}_{q=1}^{Q}$, which has low regret w.r.t. the linear function sequence $\langle \cdot\ ,\ \nabla \overline{F}_q(\boldsymbol{x}_q)\rangle$. $\overline{F}_q$ is the auxiliary function of the block average of $\{f_t\}_{t=1}^{T}$. In block q, our algorithm plays $\boldsymbol{y}_t = \boldsymbol{x}_q$ most of the time. Intuitively, the rerget of $\boldsymbol{y}_t$ w.r.t. function sequence $\langle \cdot, \nabla F_t(\boldsymbol{y_t})\rangle$ is low, where F_t is the auxiliary function of f_t. By Lemma 2.6, we can bound the $(1-1/e)$-regret of `BanditMLSM`. The proof of Theorem 3.3 is deferred to Appendix B.

Theorem 3.3. *Set $\eta = d^{-4}T^{-2/3}$, $L = d^{-2}T^{1/3}$ in Algorithm 1, if Φ is a ν-self-concordant barrier of $\mathcal{K}$, then the expected $(1-1/e)$-regret of Algorithm 1 can be bounded as*

$$\mathcal{R}_{1-1/e}(T) \leq O(\nu d^{4/3}T^{2/3}\log T).$$

About the computational complexity The computational cost mainly comes from two tasks: (1) Calculating the inverse and square root of the Hessian matrix of the regularizer; (2) Minimizing the convex function over a convex body. These tasks are commonly performed, so `BanditMLSM` can be implemented efficiently.

4. Bandit DR-submodular Maximization

Combining RFTL with a self-concordant barrier and non-oblivious technique, we can also improve the result of the general bandit DR-submodular maximization problem where the online reward functions are not required to be multi-linear functions. Due to the space limitation, the algorithmic details and the proof are deferred to the Appendix C. Here we only give the regret bound of our algorithm.

Theorem 4.1. *If there is a ν-self-concordant barrier of $\mathcal{K}$. Then there is an algorithm that attains the following regret upper bound in any $(\mathcal{K}, \mathcal{F}_{DS})$-bandit instance:*

$$\mathcal{R}_{1-1/e}(T) \leq O(\nu d^{1/2}T^{3/4}\log T).$$

5. A Continuous Approach for Submodular Full-Bandit

In this section, we show reductions from two selected submodular full-bandit problems to the bandit multi-linear DR-submodular maximization problem. All proofs in this section are deferred to Appendix E due to space limitations.

5.1. Reduction Framework

A natural reduction for our task is to consider the multi-linear extension of the submodular function. That is, we consider the multi-linear extension of each submodular set function, running the `BanditMLSM` on the function sequence of the multi-linear extensions. If we could estimate the function value of the multi-linear extension unbiasedly by using only one query to the corresponding discrete submodular function, then we would complete the reduction successfully. This idea is already considered in (Zhang et al., 2019). However, it does not work in the full-bandit setting here. The main reason is that the definition of multi-linear extension uses information of the values of the submodular set function on all subsets, including those not satisfying the constraint. This makes it impossible to find an unbiased estimator for the multi-linear extension under bandit feedback setting. To address this problem, Zhang et al. (2019) consider a relaxed responsive bandit model, where they allow the algorithm to query the function value of an infeasible action and gain zero reward. Through this relaxation, they prove a $O(T^{8/9})$ $(1-1/e)$-regret upper bound for bandit submodular maximization with a matroid constraint. We do not make this relaxation and consider the original full-bandit model, that is, the algorithm must play a feasible action each round.

Algorithm 2 `MLSMWrapper`$(\eta, L, \Phi, \mathrm{EXT})$

Input: learning rate η, block size L, self-concordant barrier Φ, an extension mapping EXT

1: **for** $t = 1, 2, \ldots, T$ **do**
2: Get $\boldsymbol{y}_t$ from `BanditMLSM4PS`(η, L, Φ)
3: Sample S_t from distribution $\mathrm{EXT}(\boldsymbol{y}_t)$
4: Play S_t and feed $g_t(S_t)$ back to `BanditMLSM4PS`(η, L, Φ)
5: **end for**

Assume we want to transform a $(\mathcal{S}, \mathcal{G})$-bandit to a bandit multi-linear DR-submodular maximization instance, where $\mathcal{S}$ is a finite set and we use $g_t \in \mathcal{G}$ to denote the online reward function. The central component of our reduction framework is a mapping from a product of standard simplexes, denoted as $\mathcal{K}$, to the set of all distributions over $\mathcal{S}$, denoted as $\Delta(\mathcal{S})$. The d-dimensional standard simplex is a set $\{(x_1, x_2, \ldots, x_d) \mid x_1 + \cdots + x_d \leq 1, x_i \geq 0, \forall i\}$.

We denote the extension mapping as $\mathrm{EXT} : \mathcal{K} \to \Delta(\mathcal{S})$. The dimension d of the set $\mathcal{K}$ varies with different $\mathcal{S}$ and $\mathcal{G}$. The extension mapping naturally defines an extension of any function $g \in G$, that is, $f(\boldsymbol{x}) = \mathbb{E}_{S\in\mathrm{EXT}(\boldsymbol{x})}[g(S)]$. This extension has a good property, if we sample an element $S \in \mathcal{S}$ according to the distribution $\mathrm{EXT}(\boldsymbol{x})$, then $g(S)$ is an unbiased estimator of $f(\boldsymbol{x})$. The idea is to run `BanditMLSM` on such extensions $\{f_t\}_{t=1}^{T}$ of the online functions sequence $\{g_t\}_{t=1}^{T}$. When we received an action $\boldsymbol{y}_t$ from `BanditMLSM`, we sample an action $S_t \in \mathcal{S}$ from $\mathrm{EXT}(\boldsymbol{y}_t)$, and feed $g_t(S_t)$ back to `BanditMLSM`. However, if we replace the $f_t(\boldsymbol{y}_{t_q})$ with $g_t(S_t)$ in (5), the

estimator $\widetilde{l}(\boldsymbol{H}_q\boldsymbol{v}_q)$ can be unbounded when z_q is very small which makes the regret uncontrollable. Fortunately, when $\mathcal{K}$ is a product of simplexes, we can slightly modify `BanditMLSM` to address this problem. We denote the modified algorithm as `BanditMLSM4PS`. In this algorithm, we use another estimator to substitute (5) when $z_q < \frac{1}{2}$. That is, we draw $\boldsymbol{u}_q \in \{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_d\}$ uniformly at random. Then we let $\boldsymbol{y}_{t_q} = z_q\boldsymbol{x}_q$ or $\boldsymbol{y}_{t_q} = z_q\boldsymbol{x}_q + \frac{1}{2}\boldsymbol{u}_q$ with equal probability. The estimator is set to be $\widetilde{l}_q(\boldsymbol{H}_q\boldsymbol{v}_q) :=$
$\begin{cases} -4(1-1/e)d\langle \boldsymbol{H}_q\boldsymbol{v}_q, \boldsymbol{u}_q\rangle f_{t_q}(\boldsymbol{y}_{t_q}) & \text{if } \boldsymbol{y}_{t_q} = z_q\boldsymbol{x}_q, \\ 4(1-1/e)d\langle \boldsymbol{H}_q\boldsymbol{v}_q, \boldsymbol{u}_q\rangle f_{t_q}(\boldsymbol{y}_{t_q}) & \text{if } \boldsymbol{y}_{t_q} = z_q\boldsymbol{x}_q + \frac{1}{2}\boldsymbol{u}_q. \end{cases}$
To show the estimator is feasible, we need to prove that $z_q\boldsymbol{x}_q + \frac{1}{2}\boldsymbol{u}_q \in \mathcal{K}$ such that the value $f_{t_q}(z_q\boldsymbol{x}_q + \frac{1}{2}\boldsymbol{u}_q)$ can be observed in bandit feedback model. Consider the simplex to which the basis vector $\boldsymbol{u}_q$ belongs, without loss of generality, we assume that $\boldsymbol{u}_q = \boldsymbol{e}_1$ and $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_{d_1}$ form the basis of the simplex. Then $x_1 + \ldots + x_{d_1} \leq 1$, which means $z_q\sum_{i=1}^{d_1} x_i < \frac{1}{2}$, therefore $\frac{1}{2} + z_q\sum_{i=1}^{d_1} x_i \leq 1$, $z_q\boldsymbol{x}_q + \frac{1}{2}\boldsymbol{u}_q \in \mathcal{K}$.

The reduction algorithm is shown in Algorithm 2 and the detailed pseudo-code of `BanditMLSM4PS` can be found in Algorithm 4 of Appendix E. For product simplexes, we give an $O(d)$-self-concordant barrier in Appendix D. To obtain the regret guarantee, we need to make sure that the extension induced by the extension mapping satisfies the assumption `BanditMLSM4PS` requires. Formally, we prove the following lemma.

Lemma 5.1. *For a finite set $\mathcal{S}$, and a function family $\mathcal{G} \subseteq \mathcal{S}^{\mathbb{R}_+}$, where $\mathcal{S}^{\mathbb{R}_+}$ is the set of all functions that map element in $\mathcal{S}$ to $\mathbb{R}^+$. If there is an extension mapping EXT : $\mathcal{K} \to \Delta(\mathcal{S})$ satisfying following conditions:*

1. *$\mathcal{K} \subseteq \mathbb{R}^d$ is a product of standard simplexes.*
2. *For any $g \in \mathcal{G}$, $f(\boldsymbol{x}) = \mathbb{E}_{S\in EXT(\boldsymbol{x})}[g(S)]$ is a multi-linear, monotone, DR-submodular function, and f is L_1-lipschitz continuous, $f(\boldsymbol{0}) = 0$.*
3. *For any $S \in \mathcal{S}$, there exist $\boldsymbol{x} \in \mathcal{K}$ such that EXT$(\boldsymbol{x}) = \boldsymbol{1}_s$. Where $\boldsymbol{1}_S$ assign probability 1 to S and 0 to other elements of $\mathcal{S}$.*

then the algorithm `MLSMWrapper` *attains expected $(1-1/e)$-regret*

$$\mathcal{R}_{1-1/e}(T) \leq O\left(d^{5/3}T^{2/3}\log(T)\right)$$

on $(\mathcal{S}, \mathcal{G})$-bandit.

5.2. Bandit Monotone Submodular Maximization with Partition Matroid Constraint

We consider a $(\mathcal{S}_{PM}, \mathcal{G}_{MS})$-bandit this section, here $\mathcal{S}_{PM}$ is a partition matroid, and $\mathcal{G}_{MS}$ is the family of monotone submodular set function. We assume the functions in $\mathcal{G}_{MS}$ take value 0 on the empty set.

Definition 5.2 (Partition Matroid)**.** Let G be a finite ground set. A set system $\mathcal{S} \subseteq 2^G$ is called a partition matroid if there exist $K > 0$ and positive integers $r_1, r_2, \ldots, r_K$ such that G can be partitioned into K subsets $G = \bigcup_{k=1}^K G_k$, and $\mathcal{S} = \{A \mid A \in 2^G \text{ and } |A \cap G_k| \leq r_k \ \forall\, k\}$.

By Lemma 5.1, all we need is to find an appropriate extension mapping. Let Δ_d be a d-dimensional standard simplex, Let $\mathcal{K} = \prod_{k=1}^K \left(\prod_{i=1}^{r_k} \Delta^{k,i}_{|G_k|}\right)$ be the product of standard simplexes. Here $\Delta^{k,i}_{|G_k|}$ is a $|G_k|$-dimensional standard simplex and (k,i) is the index of this simplex. Next, we construct an extension mapping $\text{EXT}_{PM} : \mathcal{K} \to \mathcal{S}_{PM}$.

For $\boldsymbol{x} \in \mathcal{K}$, write $\boldsymbol{x} = (x_{k,i,s})_{(k,i,s)\in\Lambda}$, $\Lambda = \{(k,i,s) \mid 1 \leq k \leq K, 1 \leq i \leq r_k, s \in G_k, k, i \in \mathbb{N}\}$ is the index set. $x_{k,i,s}$ means the coordinate of the simplex $\Delta^{k,i}_{|G_k|}$, and $\boldsymbol{x} \in \mathbb{R}_+^{\sum_{k=1}^K r_k|G_k|}$ satisfies $\sum_{s\in G_k} x_{k,i,s} \leq 1,\ \forall k, i$. We see the point in the standard simplex $\Delta^{k,i}_{|G_k|}$ as a probability distribution over $G_k \cup \{\circ\}$ where $\circ \notin G_k$ is an extra element which means no element in G_k is chosen. We sample elements according to the coordinate of each simplex independently, then $\boldsymbol{x}$ can be seen as a probability distribution over the set $\Omega := \prod_{k=1}^K (G_k \cup \{\circ\})^{r_k}$, we use pre-$\text{EXT}_{PM}(\boldsymbol{x})$ to denote this distribution on Ω. We now define a mapping $\rho : \Omega \longrightarrow \mathcal{S}$ as follows. For $\omega \in \Omega$, assume ω can be represented as $\omega = (\omega_{k,i})_{(k,i)\in\Gamma}$, where $\omega_{k,i} \in G_k \cup \{\circ\}$ and $\Gamma = \{(k,i) \mid 1 \leq k \leq K, 1 \leq i \leq r_k, k, i \in \mathbb{N}\}$ is the index set. Then $\rho(\omega) = \{\omega_{k,i} \mid (k,i) \in \Gamma\}\backslash\{\circ\}$.

It's easy to check $\rho(\omega) \in \mathcal{S}$. Thus, for $\boldsymbol{x}$, we first sample an $\omega \sim$ pre-$\text{EXT}_{PM}(\boldsymbol{x})$, then map the sample to $\rho(\omega) \in \mathcal{S}$. This process defines a distribution over $\mathcal{S}$. We let this distribution be $\text{EXT}_{PM}(\boldsymbol{x})$.

Lemma 5.3. *For $\mathcal{G}_{MS}$, the extension mapping $EXT_{PM} : \mathcal{K} \to \Delta(\mathcal{S}_{PM})$ satisfies the conditions in Lemma 5.1. Moreover, $\mathcal{K}$ is in a $\sum_{k=1}^K r_k$ dimensional real vector space. For any $g \in \mathcal{G}_{MS}$, the continuous extension $f(\boldsymbol{x}) = \mathbb{E}_{S\in EXT_{PM}(\boldsymbol{x})}[g(S)]$ is $M\sqrt{\sum_{k=1}^K r_k|G_k|}$-lipschitz.*

Corollary 5.4. *There is an algorithm attaining the expected $(1-1/e)$-regret of $\mathcal{R}_{1-1/e}(T) \leq O\left(\left(\sum_{k=1}^K r_k|G_k|\right)^{5/3} T^{2/3}\log T\right)$ on any $(\mathcal{S}_{PM}, \mathcal{G}_{MS})$-bandit.*

5.3. Bandit Sequential Submodular Maximization

Bandit sequential submodular maximization is first studied in (Niazadeh et al., 2021). It is motivated by online retailing platforms where the platform needs to show its products in sequence. There are many types of customers who have different patience and preference. Rarely customers will

see all the products in the list. They will stop browsing the product after seeing some products according to their patience, and the click probability after a customer sees a set of products is submodular. This situation can be formalized into an $(\mathcal{S}_{OL}, \mathcal{G}_{SS})$-bandit. Let G be the ground set of all products, and the constraint $\mathcal{S}_{OL}$ is the set of all ordered lists of length $|G|$ consisting of elements in G. $\mathcal{G}_{SS}$ consists of function $g : \mathcal{S}_{OL} \to [0, M]$ in this form,

$$g(S) = \sum_{i=1}^{|G|} \lambda_i g_i(\{S_j \mid j \leq i\}),$$

where λ_i's with $\lambda_i \geq 0$ are positive weights, g_i's are monotone submodular set functions, and S_j is the i-th element in the ordered list S. If we interpret $g(S)$ as a click probability, then $M = 1$.

For technical reasons, we assume that there is a dummy element $\circ$ in G, which has 0 marginal gain for all g_i. That is, $\forall i, \forall S \subseteq G$, we have $g_i(S \cup \{\circ\}) = g_i(S)$. We denote $G' = G \backslash \{\circ\}$ This assumption can be satisfied by adding a non-clickable item that is not related to the products to G'.

Next, we construct an extension mapping EXT_{SS}. Let $\mathcal{K}$ be the cartesian product of standard simplexes, $\mathcal{K} = \prod_{i=1}^{|G|} \Delta^i_{|G'|}$. We see $\boldsymbol{x} \in \mathcal{K}$ as $|G|$ probability distributions over G, the component of $\boldsymbol{x} \in \mathcal{K}$ in $\Delta^i_{|G'|}$ represents the distribution of the i-th element of the ordered list, all these distributions are independent. Any $\boldsymbol{x} \in \mathcal{K}$ can be seen as a distribution over $\mathcal{S}_{OL}$. Let this distribution be $\text{EXT}_{SS}(\boldsymbol{x})$.

Lemma 5.5. *For $\mathcal{G}_{SS}$, the extension mapping $EXT_{SS} : \mathcal{K} \to \Delta(\mathcal{S}_{OL})$ satisfies the conditions in Lemma 5.1. Moreover, $\mathcal{K}$ is in a $|G|^2 - |G|$ dimensional real vector space. For any $g \in \mathcal{G}_{SS}$, the continuous extension $f(\boldsymbol{x}) = \mathbb{E}_{S \in EXT_{SS}(\boldsymbol{x})}[g(S)]$ is $M|G|$-lipschitz.*

Corollary 5.6. *There is an algorithm for attaining the expected $(1 - 1/e)$-regret of $\mathcal{R}_{1-1/e}(T) \leq O\left((|G|)^{10/3} T^{2/3} \log T\right)$ on any $(\mathcal{S}_{OL}, \mathcal{G}_{SS})$-bandit.*

6. Conclusion

In this paper, we propose two bandit algorithms, `BanditMLSM` for monotone multilinear DR-submodular functions and `BanditDRSM` for general monotone DR-submodular functions. We then show an approach to design the $\widetilde{O}(T^{2/3})$ $(1-1/e)$-regret algorithm for two special combinatorial full-bandits submodular maximization problems, that is, reducing the combinatorial bandits to a multilinear DR-submodular bandit.

There are some remaining open problems that need to be studied. Firstly, we notice that $\widetilde{O}(T^{2/3})$-type regret bounds show up frequently in the submodular bandit literature. However, as far as we know no one has proved or disproved the optimality of this bound, which may be an interesting and challenging problem. Secondly, we still know less about the relationship between the combinatorial constraint and sublinear regret. For submodular set functions, we show that one can achieve sublinear regret with partition matroid constraint in this paper. However, we conjecture that it does not hold for all matroid constraints. How to characterize such a relationship is also a fascinating open question.

Acknowledgements

We thank the anonymous reviewers for their suggestions in the presentation of the article. This work was supported in part by the National Natural Science Foundation of China Grants No. 61832003, 62272441.

References

Abernethy, J., Hazan, E. E., and Rakhlin, A. Competing in the dark: An efficient algorithm for bandit linear optimization. In *21st Annual Conference on Learning Theory, COLT 2008*, pp. 263–273, 2008.

Bian, A., Levy, K. Y., Krause, A., and Buhmann, J. M. Continuous dr-submodular maximization: structure and algorithms. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp. 486–496, 2017a.

Bian, A. A., Mirzasoleiman, B., Buhmann, J., and Krause, A. Guaranteed non-convex optimization: Submodular maximization over continuous domains. In *Artificial Intelligence and Statistics*, pp. 111–120. PMLR, 2017b.

Chen, L., Krause, A., and Karbasi, A. Interactive submodular bandit. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp. 140–151, 2017.

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

Chen, W., Hu, W., Li, F., Li, J., Liu, Y., and Lu, P. Combinatorial multi-armed bandit with general reward functions. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, pp. 1659–1667, 2016.

Filmus, Y. and Ward, J. Monotone submodular maximization over a matroid via non-oblivious local search. *SIAM Journal on Computing*, 43(2):514–542, 2014.

Flaxman, A. D., Kalai, A. T., and McMahan, H. B. Online convex optimization in the bandit setting: gradient descent without a gradient. In *Proceedings of the sixteenth*

annual ACM-SIAM symposium on Discrete algorithms, pp. 385–394, 2005.

Foster, D. P. and Rakhlin, A. On submodular contextual bandits. *arXiv preprint arXiv:2112.02165*, 2021.

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

Li, S., Kong, F., Tang, K., Li, Q., and Chen, W. Online influence maximization under linear threshold model. *Advances in Neural Information Processing Systems*, 33: 1192–1204, 2020.

Nesterov, Y. and Nemirovskii, A. *Interior-point polynomial algorithms in convex programming*. SIAM, 1994.

Niazadeh, R., Roughgarden, T., and Wang, J. R. Optimal algorithms for continuous non-monotone submodular and dr-submodular maximization. *The Journal of Machine Learning Research*, 21(1):4937–4967, 2020.

Niazadeh, R., Golrezaei, N., Wang, J. R., Susan, F., and Badanidiyuru, A. Online learning via offline greedy algorithms: Applications in market design and optimization. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pp. 737–738, 2021.

Nie, G., Agarwal, M., Umrawal, A. K., Aggarwal, V., and Quinn, C. J. An explore-then-commit algorithm for submodular maximization under full-bandit feedback. In *The 38th Conference on Uncertainty in Artificial Intelligence*, 2022.

Nie, G., Nadew, Y. Y., Zhu, Y., Aggarwal, V., and Quinn, C. J. A framework for adapting offline algorithms to solve combinatorial multi-armed bandit problems with bandit feedback. *arXiv preprint arXiv:2301.13326*, 2023.

Raut, P. S., Sadeghi, O., and Fazel, M. Online dr-submodular maximization with stochastic cumulative constraints. *arXiv preprint arXiv:2005.14708*, 2020.

Saha, A. and Tewari, A. Improved regret guarantees for online smooth convex optimization with bandit feedback. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pp. 636–642. JMLR Workshop and Conference Proceedings, 2011.

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

Streeter, M., Golovin, D., and Krause, A. Online learning of assignments. In *Proceedings of the 22nd International Conference on Neural Information Processing Systems*, pp. 1794–1802, 2009.

Thang, N. K. and Srivastav, A. Online non-monotone dr-submodular maximization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 9868–9876, 2021.

Vaswani, S., Lakshmanan, L., Schmidt, M., et al. Influence maximization with bandits. *arXiv preprint arXiv:1503.00024*, 2015.

Wang, Q. and Chen, W. Improving regret bounds for combinatorial semi-bandits with probabilistically triggered arms and its applications. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp. 1161–1171, 2017.

Wu, Q., Li, Z., Wang, H., Chen, W., and Wang, H. Factorization bandits for online influence maximization. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 636–646, 2019.

Yue, Y. and Guestrin, C. Linear submodular bandits and their application to diversified retrieval. In *Proceedings of the 24th International Conference on Neural Information Processing Systems*, pp. 2483–2491, 2011.

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

Zhang, Q., Deng, Z., Chen, Z., Hu, H., and Yang, Y. Stochastic continuous submodular maximization: Boosting via non-oblivious function. In *International Conference on Machine Learning*, pp. 26116–26134. PMLR, 2022a.

Zhang, Q., Deng, Z., Chen, Z., Zhou, K., Hu, H., and Yang, Y. Online learning for non-monotone submodular maximization: From full information to bandit feedback. *arXiv preprint arXiv:2208.07632*, 2022b.

Zhang, Z., Chen, W., Sun, X., and Zhang, J. Online influence maximization with node-level feedback using standard offline oracles. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 9153–9161, 2022c.

A. Technical Lemmas

This section provides some technical lemmas that will be used in the proofs of this appendix later.

Definition A.1 (Minkowski function and Minkowski set)**.** Let $\mathcal{K}$ be a compact convex set, the Minkowski function $\pi_{\boldsymbol{x}} : \mathcal{K} \to \mathbb{R}$ parameterized by a pole $\boldsymbol{x} \in \text{int}(\mathcal{K})$ is defined as $\pi_{\boldsymbol{x}}(\boldsymbol{y}) \triangleq \inf\{t \geq 0 \mid x + t^{-1}(y - x) \in \mathcal{K}\}$. Given $\delta \in \mathbb{R}^+$ and $\boldsymbol{x}_1 \in \text{int}(\mathcal{K})$, we define the Minkowski set $\mathcal{K}_{\gamma,\boldsymbol{x}_1} \triangleq \{\boldsymbol{x} \in \mathcal{K} \mid \pi_{\boldsymbol{x}_1}(\boldsymbol{x}) \leq (1+\gamma)^{-1}\}$.

The following lemma provides an upper bound of the difference between the function value of a self-concordant barrier at two different points.

Lemma A.2 ((Nesterov & Nemirovskii, 1994))**.** *Let Φ be a ν-self-concordant barrier over a compact convex set $\mathcal{K}$, then for all $x, y \in$ int$(\mathcal{K})$:*

$$\Phi(y) - \Phi(x) \leq \nu \log \frac{1}{1 - \pi_x(y)}.$$

The following lemma is already proved in (Abernethy et al., 2008), we include the proof for completeness.

Lemma A.3 ((Abernethy et al., 2008))**.** *Let $\mathcal{K}$ be a compact convex set, $\boldsymbol{x} \in$ int$(\mathcal{K})$ with diameter D, $\boldsymbol{x}^* \in \mathcal{K}$ and $\hat{\boldsymbol{x}}^* \triangleq \text{argmin}_{\boldsymbol{z} \in \mathcal{K}_{\gamma,\boldsymbol{x}}} \|\boldsymbol{z} - \boldsymbol{x}^*\|$ be the projection of $\boldsymbol{x}^*$ onto the Minkowski set $\mathcal{K}_{\gamma,\boldsymbol{x}}$, then*

$$\|\boldsymbol{x}^* - \hat{\boldsymbol{x}}^*\| \leq \gamma D$$

Proof. Consider the point $\boldsymbol{y}$ in the segment $[\boldsymbol{x}, \boldsymbol{x}^*]$ satisfying $\frac{\|\boldsymbol{y}-\boldsymbol{x}\|}{\|\boldsymbol{x}^*-\boldsymbol{x}\|} = \frac{1}{1+\gamma}$. Since $\boldsymbol{x} + (1+\gamma)(\boldsymbol{y} - \boldsymbol{x}) = \boldsymbol{x}^* \in \mathcal{K}$, we can deduce that $\boldsymbol{y} \in \mathcal{K}_{\gamma,\boldsymbol{x}}$. Thus,

$$\|\hat{\boldsymbol{x}}^* - \boldsymbol{x}^*\| \leq \|\boldsymbol{y} - \boldsymbol{x}^*\| = \left(1 - \frac{1}{1+\gamma}\right)\|\boldsymbol{x}^* - \boldsymbol{x}\| \leq \gamma D.$$

□

The following two lemmas show that the average auxiliary functions and the $\boldsymbol{H}$-smoothed functions both inherent good properties of the original online functions. And they will be used later.

Lemma A.4. *If $\forall t \in [(q-1)L + 1, qL]$, f_t is twice differentiable, L_1-lipschitz and L_2-smooth, monotone, DR-submodular, then following holds for the average functions $\overline{f}_q$, $\overline{F}_q$.*

(i) $\overline{f}_q$ is L_1-lipschitz and L_2-smooth.

(ii) $\overline{f}_q$ is a monotone DR-submodular function.

(iii) $\overline{F}_q$ is $\frac{L_2}{e}$-smooth.

(iv) $\overline{F}_q$ is a monotone DR-submodular function.

Proof. (i)

$$\begin{aligned}
\|\overline{f}_q(\boldsymbol{x}) - \overline{f}_q(\boldsymbol{y})\| &= \frac{1}{L}\left\|\sum_{t=(q-1)L+1}^{qL} f_t(\boldsymbol{x}) - \sum_{t=(q-1)L+1}^{qL} f_t(\boldsymbol{y})\right\| \\
&\leq \frac{1}{L}\sum_{t=(q-1)L+1}^{qL} \|f_t(\boldsymbol{x}) - f_t(\boldsymbol{y})\| \\
&\leq \frac{1}{L}\sum_{t=(q-1)L+1}^{qL} L_1\|\boldsymbol{x} - \boldsymbol{y}\| = L_1\|\boldsymbol{x} - \boldsymbol{y}\|
\end{aligned}$$

$$\begin{aligned}\|\nabla \overline{f}_q(\boldsymbol{x}) - \nabla \overline{f}_q(\boldsymbol{y})\| &= \frac{1}{L}\left\|\sum_{t=(q-1)L+1}^{qL} \nabla f_t(\boldsymbol{x}) - \sum_{t=(q-1)L+1}^{qL} \nabla f_t(\boldsymbol{y})\right\| \\ &\leq \frac{1}{L}\sum_{t=(q-1)L+1}^{qL} \|\nabla f_t(\boldsymbol{x}) - \nabla f_t(\boldsymbol{y})\| \\ &\leq \frac{1}{L}\sum_{t=(q-1)L+1}^{qL} L_2\|\boldsymbol{x}-\boldsymbol{y}\| \leq L_2\|\boldsymbol{x}-\boldsymbol{y}\|\end{aligned}$$

(ii) For any $i \in [d]$,

$$\frac{\partial \overline{f}_q}{\partial x_i}(\boldsymbol{x}) = \frac{1}{L}\sum_{t=(q-1)L+1}^{qL} \frac{\partial f_t}{\partial x_i}(\boldsymbol{x}) \geq 0$$

For any $i \in [d], j \in [d]$,

$$\frac{\partial^2}{\partial x_i \partial x_j}\overline{f}_q(\boldsymbol{x}) = \frac{1}{L}\sum_{t=(q-1)L+1}^{qL} \frac{\partial^2 f_t}{\partial x_i \partial x_j}(\boldsymbol{x}) \leq 0$$

Thus $\overline{f}_q$ is monotone DR-submodular.

(iii)

$$\begin{aligned}\|\nabla \overline{F}_q(\boldsymbol{x}) - \nabla \overline{F}_q(\boldsymbol{y})\| &= \left\|\nabla \int_0^1 \frac{e^{z-1}}{zL}\sum_{t=(q-1)L+1}^{qL} f_t(z\cdot\boldsymbol{x})dz - \nabla\int_0^1 \frac{e^{z-1}}{zL}\sum_{t=(q-1)L+1}^{qL} f_t(z\cdot\boldsymbol{y})dz\right\| \\ &= \left\|\int_0^1 \frac{e^{z-1}}{L}\sum_{t=(q-1)L+1}^{qL} \nabla f_t(z\cdot\boldsymbol{x})dz - \int_0^1 \frac{e^{z-1}}{L}\sum_{t=(q-1)L+1}^{qL} \nabla f_t(z\cdot\boldsymbol{y})dz\right\| \\ &\leq \int_0^1 \frac{e^{z-1}}{L}\sum_{t=(q-1)L+1}^{qL} \|\nabla f_t(z\cdot\boldsymbol{x})dz - \nabla f_t(z\cdot\boldsymbol{y})\|dz \\ &\leq \int_0^1 \frac{e^{z-1}}{L}\sum_{t=(q-1)L+1}^{qL} L_2 z\|\boldsymbol{x}-\boldsymbol{y}\|dz \\ &= L_2\int_0^1 ze^{z-1}dz\|\boldsymbol{x}-\boldsymbol{y}\| = \frac{L_2}{e}\|\boldsymbol{x}-\boldsymbol{y}\|\end{aligned}$$

(iv) For any $i \in [d]$,

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{x}_i}\overline{F}_q(\boldsymbol{x}) &= \frac{\partial}{\partial \boldsymbol{x}_i}\int_0^1 \frac{e^{z-1}}{zL}\sum_{t=(q-1)L+1}^{qL} f_t(z\cdot\boldsymbol{x})dz \\ &= \int_0^1 \frac{e^{z-1}}{L}\sum_{t=(q-1)L+1}^{qL} \frac{\partial}{\partial \boldsymbol{x}_i} f_t(z\cdot\boldsymbol{x})dz \\ &\geq 0\end{aligned}$$

For any $i \in [d], j \in [d]$,

$$\frac{\partial}{\partial \boldsymbol{x}_i\boldsymbol{x}_j}\overline{F}_q(\boldsymbol{x}) = \int_0^1 \frac{ze^{z-1}}{L}\sum_{t=(q-1)L+1}^{qL} \frac{\partial}{\partial \boldsymbol{x}_i\boldsymbol{x}_j} f_t(z\cdot\boldsymbol{x})dz$$

$$\leq 0$$

Thus $\overline{F}_q$ is monotone and DR-submodular.

□

Lemma A.5. *Following properties hold for $\boldsymbol{H}$-smoothed version of a twice differentiable function $f(\boldsymbol{x})$.*

(i) If $f(\boldsymbol{x})$ is a monotone DR-submodular function, then for any invertible matrix $\boldsymbol{H}$, its $\boldsymbol{H}$-smoothed version $f^{\boldsymbol{H}}(\boldsymbol{x})$ is a monotone DR-submodular function.

(ii) If $f(\boldsymbol{x})$ is L_1-lipschitz continuous and L_2-smooth, then $f^{\boldsymbol{H}}(\boldsymbol{x})$ is L_1-lipschitz continuous and L_2-smooth.

Proof. (i) By Leibnez integral rule, for any $i \in [d]$,

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{x}_i} f^{\boldsymbol{H}}(\boldsymbol{x}) &= \int_{\boldsymbol{v}\in\mathbb{B}_d} \frac{1}{\text{Vol}(\mathbb{B}_d)} \frac{\partial}{\partial \boldsymbol{x}_i} f(\boldsymbol{x}+\boldsymbol{H}\boldsymbol{v}) d\boldsymbol{v} \\ &\geq 0\end{aligned}$$

The last inequality is because $\frac{\partial}{\partial \boldsymbol{x}_i} f(\boldsymbol{x}+\boldsymbol{H}\boldsymbol{v}) \geq 0$ for any $i \in [d]$.

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{x}_i \boldsymbol{x}_j} f^{\boldsymbol{H}}(\boldsymbol{x}) &= \int_{\boldsymbol{v}\in\mathbb{B}_d} \frac{1}{\text{Vol}(\mathbb{B}_d)} \frac{\partial}{\partial \boldsymbol{x}_i \boldsymbol{x}_j} f(\boldsymbol{x}+\boldsymbol{H}\boldsymbol{v}) d\boldsymbol{v} \\ &\leq 0\end{aligned}$$

The last inequality is because $\frac{\partial}{\partial \boldsymbol{x}_i \boldsymbol{x}_j} f(\boldsymbol{x}+\boldsymbol{H}\boldsymbol{v}) \leq 0$ for any $i, j \in [d]$.

(ii)

$$\begin{aligned}f^{\boldsymbol{H}}(\boldsymbol{x}) - f^{\boldsymbol{H}}(\boldsymbol{y}) &= \int_{\boldsymbol{v}\in\mathbb{B}_d} \frac{1}{\text{Vol}(\mathbb{B}_d)} \left(f(\boldsymbol{x}+\boldsymbol{H}\boldsymbol{v}) - f(\boldsymbol{y}+\boldsymbol{H}\boldsymbol{v})\right) d\boldsymbol{v} \\ &\leq \int_{\boldsymbol{v}\in\mathbb{B}_d} \frac{1}{\text{Vol}(\mathbb{B}_d)} L_1 \|\boldsymbol{x}+\boldsymbol{H}\boldsymbol{v}-\boldsymbol{y}-\boldsymbol{H}\boldsymbol{v}\| d\boldsymbol{v} \\ &= L_1 \|\boldsymbol{x}-\boldsymbol{y}\|\end{aligned}$$

Thus, $f^{\boldsymbol{H}}(\boldsymbol{x})$ is L_1-lipschitz continuous.

$$\begin{aligned}\nabla f^{\boldsymbol{H}}(\boldsymbol{x}) - \nabla f^{\boldsymbol{H}}(\boldsymbol{y}) &= \nabla \int_{\boldsymbol{v}\in\mathbb{B}_d} \frac{1}{\text{Vol}(\mathbb{B}_d)} \left(f(\boldsymbol{x}+\boldsymbol{H}\boldsymbol{v}) - f(\boldsymbol{y}+\boldsymbol{H}\boldsymbol{v})\right) d\boldsymbol{v} \\ &= \int_{\boldsymbol{v}\in\mathbb{B}_d} \frac{1}{\text{Vol}(\mathbb{B}_d)} \nabla \left(f(\boldsymbol{x}+\boldsymbol{H}\boldsymbol{v}) - f(\boldsymbol{y}+\boldsymbol{H}\boldsymbol{v})\right) d\boldsymbol{v} \\ &\leq \int_{\boldsymbol{v}\in\mathbb{B}_d} \frac{1}{\text{Vol}(\mathbb{B}_d)} L_2 \|\boldsymbol{x}-\boldsymbol{y}\| d\boldsymbol{v} \\ &= L_2 \|\boldsymbol{x}-\boldsymbol{y}\|\end{aligned}$$

□

B. Missing Proofs in Section 3

We first show a property of multi-linear functions, which is the key observation of our estimator for the gradient of multi-linear functions.

Lemma B.1. *If $f : \mathcal{K} \to \mathbb{R}$ is a multi-linear function, where $\mathcal{K} \subseteq \mathbb{R}^d$, then for any basis vector $\boldsymbol{e}_i, i \in [d]$ and any $\boldsymbol{x} \in \mathcal{K}$ and $\lambda > 0$ satisfying $\boldsymbol{x} + \lambda \boldsymbol{e}_i \in \mathcal{K}$. We have,*

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}_i} = \frac{f(\boldsymbol{x}+\lambda \boldsymbol{e}_i) - f(\boldsymbol{x})}{\lambda} \tag{7}$$

Proof. When we fix the components of $\boldsymbol{x}$ except $\boldsymbol{x}_i$, f is a linear function of $\boldsymbol{x}_i$. Thus (7) directly comes from the linearity. □

Lemma 3.1. *Let $\mathcal{H}_q$ be the history of the algorithm in the first q blocks, that is, the realization of $t_s, z_s, \boldsymbol{v}_s, \boldsymbol{u}_s, \forall s \leq q$. Then $\mathbb{E}[\widetilde{l}_q(\boldsymbol{H}_q\boldsymbol{v}_q) \mid \mathcal{H}_{q-1}, \boldsymbol{v}_q] = l_q(\boldsymbol{H}_q\boldsymbol{v}_q)$.*

Proof of Lemma 3.1. When $z_q > 0$, we have

$$\begin{aligned}
\mathbb{E}[\widetilde{l}_q(\boldsymbol{H}_q\boldsymbol{v}_q) \mid \mathcal{H}_{q-1}, \boldsymbol{v}_q, t_q, z_q] &= \frac{1}{2}(-2(1-1/e)\frac{d}{z_q} \cdot f_{t_q}(z_q\boldsymbol{x}_q)) + \sum_{i=1}^{d} \frac{1}{2d}(2(1-1/e)\frac{d}{z_q} \cdot f_{t_q}(z_q\boldsymbol{x}_q + z_q\langle \boldsymbol{H}_q\boldsymbol{v}_q, \boldsymbol{e}_i\rangle \boldsymbol{e}_i) \\
&= (1-1/e)\sum_{i=1}^{d} \frac{1}{z_q}\left(f_{t_q}(z_q\boldsymbol{x}_q + z_q\langle \boldsymbol{H}_q\boldsymbol{v}_q, \boldsymbol{e}_i\rangle \boldsymbol{e}_i) - f_{t_q}(z_q\boldsymbol{x}_q)\right) \\
&= (1-1/e)\sum_{i=1}^{d} \frac{1}{z_q} z_q\langle \boldsymbol{H}_q\boldsymbol{v}_q, \boldsymbol{e}_i\rangle \frac{\partial f_{t_q}}{\partial \boldsymbol{x}_i}(z_q \cdot \boldsymbol{x}_q) \\
&= (1-1/e)\sum_{i=1}^{d} \langle \boldsymbol{H}_q\boldsymbol{v}_q, \boldsymbol{e}_i\rangle\langle \boldsymbol{e}_i, \nabla f_{t_q}(z_q \cdot \boldsymbol{x}_q)\rangle \\
&= (1-1/e)\langle \boldsymbol{H}_q\boldsymbol{v}_q, \nabla f_{t_q}(z_q \cdot \boldsymbol{x}_q)\rangle
\end{aligned}$$

Then we take the expectations over t_q and z_q, note the value of $\widetilde{l}(\boldsymbol{H}_q\boldsymbol{v}_q)$ when $z_q = 0$ does not affect the result of the integral since $z_q = 0$ is a zero measured event.

$$\begin{aligned}
\mathbb{E}[\widetilde{l}_q(\boldsymbol{H}_q\boldsymbol{v}_q) \mid \mathcal{H}_{q-1}, \boldsymbol{v}_q] &= \sum_{t_q=(q-1)L+1}^{qL} \int_0^1 \Pr(t_q, z_q)\, \mathbb{E}[\widetilde{l}_q(\boldsymbol{H}_q\boldsymbol{v}_q) \mid \mathcal{H}_{q-1}, \boldsymbol{v}_q, t_q, z_q] dz_q \\
&= \sum_{t_q=(q-1)L+1}^{qL} \int_0^1 \frac{1}{L}\frac{e^{z_q-1}}{1-1/e}(1-1/e)\langle \boldsymbol{H}_q\boldsymbol{v}_q, \nabla f_{t_q}(z_q \cdot \boldsymbol{x}_q)\rangle dz_q \\
&= \left\langle \boldsymbol{H}_q\boldsymbol{v}_q, \int_0^1 \frac{e^{z_q-1}}{L} \sum_{t_q=(q-1)L+1}^{qL} \nabla f_{t_q}(z_q \cdot \boldsymbol{x}_q) dz_q \right\rangle
\end{aligned}$$

Since

$$\begin{aligned}
\nabla \overline{F}_q(\boldsymbol{x}_q) &= \nabla \int_0^1 \frac{e^{z-1}}{zL} \sum_{t=(q-1)L+1}^{qL} f_t(z \cdot \boldsymbol{x}_q) dz \\
&= \int_0^1 \frac{e^{z-1}}{L} \sum_{t=(q-1)L+1}^{qL} \nabla f_t(z \cdot \boldsymbol{x}_q) dz
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbb{E}[\widetilde{l}_q(\boldsymbol{H}_q\boldsymbol{v}_q) \mid \mathcal{H}_{q-1}, \boldsymbol{v}_q] &= \langle \boldsymbol{H}_q\boldsymbol{v}_q, \nabla \overline{F}_q(\boldsymbol{x}_q)\rangle \\
&= l_q(\boldsymbol{H}_q\boldsymbol{v}_q)
\end{aligned}$$

□

Lemma 3.2. *The following properties hold for $\widetilde{\nabla}\overline{F}_q(\boldsymbol{x}_q)$*

(i) $\mathbb{E}\left[\widetilde{\nabla}\overline{F}_q(\boldsymbol{x}_q) \mid \mathcal{H}_{q-1}\right] = \nabla \overline{F}_q(\boldsymbol{x}_q)$

(ii) $\mathbb{E}\left[\|\widetilde{\nabla}\overline{F}_q(\boldsymbol{x}_q)\|^2_{\boldsymbol{x}_q,*} \mid \mathcal{H}_{q-1}\right] \leq 4(1-1/e)^2 L_1^2 D^2 d^4$

Proof of Lemma 3.2.

(i)

$$\begin{aligned}
\mathbb{E}\left[\widetilde{\nabla}\overline{F}_q(\boldsymbol{x}_q) \mid \mathcal{H}_{q-1}\right] &= \int_{\boldsymbol{v}_q\in\mathbb{S}_{d-1}} d\cdot\mathbb{E}[\widetilde{l}_q(\boldsymbol{H}_q\boldsymbol{v}_q) \mid \mathcal{H}_{q-1},\boldsymbol{v}_q]\boldsymbol{H}_q^{-1}\boldsymbol{v}_q d\boldsymbol{v}_q \\
&= \int_{\boldsymbol{v}_q\in\mathbb{S}_{d-1}} \frac{1}{\mathrm{Vol}(\mathbb{S}_{d-1})} d\cdot l_q(\boldsymbol{H}_q\boldsymbol{v}_q)\boldsymbol{H}_q^{-1}\boldsymbol{v}_q d\boldsymbol{v}_q \qquad (8)\\
&= \mathop{\mathbb{E}}_{\boldsymbol{v}_q\sim\mathbb{S}_{d-1}}[d\cdot l_q(\boldsymbol{H}_q\boldsymbol{v}_q)\boldsymbol{H}_q^{-1}\boldsymbol{v}_q] \\
&= \nabla l_q^{\boldsymbol{H}_q}(\mathbf{0}) \qquad (9)\\
&= \nabla l_q(\mathbf{0}) \qquad (10)\\
&= \nabla\overline{F}_q(\boldsymbol{x}_q)
\end{aligned}$$

(8) is due to Lemma 3.1, (9) is due to Lemma 2.5, (10) is because that l_q is a linear function.

(ii)

$$\begin{aligned}
&\mathbb{E}\left[\|\widetilde{\nabla}\overline{F}_q(\boldsymbol{x}_q)\|^2_{\boldsymbol{x}_q,*} \mid \mathcal{H}_{q-1}\right] \\
&= \mathbb{E}\left[\frac{1}{2}(1-1/e)^2(2d^2)^2\frac{1}{z_q^2}f_{t_q}(z_q\cdot\boldsymbol{x}_q)^2\cdot\boldsymbol{v}_q^T\boldsymbol{H}_q\Phi(\boldsymbol{x}_q)^{-1}\boldsymbol{H}_q^{-1}\boldsymbol{v}_q \mid \mathcal{H}_{q-1}\right] \\
&\quad+\sum_{i=1}^{d}\mathbb{E}\left[\frac{1}{2d}(1-1/e)^2 4d^4\frac{1}{z_q^2}f_{t_q}\left(z_q\cdot\boldsymbol{x}_q+z_q\langle\boldsymbol{H}_q\boldsymbol{v}_q,\boldsymbol{e}_i\rangle\boldsymbol{e}_i\right)^2\boldsymbol{v}_q^T\boldsymbol{H}_q\Phi(\boldsymbol{x}_q)\boldsymbol{H}_q^{-1}\boldsymbol{v}_q \mid \mathcal{H}_{q-1}\right] \\
&\le 2(1-1/e)^2d^4\frac{L_1^2z_q^2\|\boldsymbol{x}_q\|^2}{z_q^2}\mathbb{E}\left[\boldsymbol{v}_q^T\boldsymbol{H}_q^{-1}\Phi(\boldsymbol{x}_q)^{-1}\boldsymbol{H}_q^{-1}\boldsymbol{v}_q \mid \mathcal{H}_{q-1}\right] \\
&\quad+2(1-1/e)^2d^4\frac{L_1^2z_q^2\|\boldsymbol{x}_q+\langle\boldsymbol{H}_q\boldsymbol{v}_q,\boldsymbol{e}_i\rangle\boldsymbol{e}_i\|^2}{z_q^2}\mathbb{E}\left[\boldsymbol{v}_q^T\boldsymbol{H}_q^{-1}\Phi(\boldsymbol{x}_q)^{-1}\boldsymbol{H}_q^{-1}\boldsymbol{v}_q \mid \mathcal{H}_{q-1}\right] \qquad (11)\\
&= 4(1-1/e)^2d^4L_1^2D^2\,\mathbb{E}\left[\boldsymbol{v}_q^T(\Phi(\boldsymbol{x}_q)^{-1/2})^{-1}\Phi(\boldsymbol{x}_q)^{-1}(\Phi(\boldsymbol{x}_q)^{-1/2})^{-1}\boldsymbol{v}_q \mid \mathcal{H}_{q-1}\right] \\
&\le 4(1-1/e)^2d^4L_1^2D^2\|\boldsymbol{v}_q\|_2^2 \\
&= 4(1-1/e)^2d^4L_1^2D^2 \qquad (12)
\end{aligned}$$

Inequality (11) is because f_{t_q} is L_1-lipschitz continuous and $\boldsymbol{x}_q+\langle\boldsymbol{H}_q\boldsymbol{v}_q,\boldsymbol{e}_i\rangle\boldsymbol{e}_i$ is in the Dikin ellipsoid $\{\boldsymbol{x} \mid \|\boldsymbol{x}-\boldsymbol{x}_q\|_{\Phi,\boldsymbol{x}_q}\le 1\}$, which is contained in $\mathcal{K}$. (12) is because $\boldsymbol{v}_q\in\mathbb{S}_{d-1}$, thus $\|\boldsymbol{v}_q\|=1$.

□

Theorem 3.3. *Set $\eta=d^{-4}T^{-2/3}$, $L=d^{-2}T^{1/3}$, $Q=T/L=d^2T^{2/3}$ in Algorithm 1, if Φ is a ν-self-concordant barrier of $\mathcal{K}$, then the expected $(1-1/e)$-regret of Algorithm 1 can be bounded as*

$$\mathcal{R}_{1-1/e}(T)\le(4(1-1/e)^2L_1^2D^2+M)d^{4/3}T^{2/3}+(1-1/e)L_1D+\nu d^{4/3}T^{2/3}\log(T)$$

Proof of Theorem 3.3. Set $\widetilde{\boldsymbol{g}}_q=\widetilde{\nabla}\overline{F}_q(\boldsymbol{x}_q)$, $\boldsymbol{g}_q=\nabla\overline{F}_q(\boldsymbol{x}_q)$ in Theorem 2.3. We have proved $\mathbb{E}\left[\widetilde{\nabla}\overline{F}_q(\boldsymbol{x}_q) \mid \mathcal{H}_{q-1}\right]=\nabla\overline{F}_q(\boldsymbol{x}_q)$. Let $\hat{\boldsymbol{x}}^*\triangleq\mathrm{argmin}_{\boldsymbol{x}\in\mathcal{K}_{\gamma,\boldsymbol{x}_1}}\|\boldsymbol{x}^*-\boldsymbol{x}\|$ be the projection of $\boldsymbol{x}^*$ onto the Minkowski set $\mathcal{K}_{\gamma,\boldsymbol{x}_1}$ defined in Definition A.1, here the pole is $\boldsymbol{x}_1$, and γ is a parameter to be determined later, $\boldsymbol{x}^*=\mathrm{argmin}_{\boldsymbol{x}\in\mathcal{K}}\sum_{t=1}^{T}f_t(\boldsymbol{x})$. We have

$$\begin{aligned}\sum_{q=1}^{Q}\mathbb{E}\left[\langle\nabla\overline{F}_q(\boldsymbol{x}_q),\hat{\boldsymbol{x}}^*-\boldsymbol{x}_q\rangle\mid\mathcal{H}_{q-1}\right]&\leq\eta\sum_{q=1}^{Q}\mathbb{E}\left[\|\widetilde{\nabla}\overline{F}_q(\boldsymbol{x}_q)\|^2_{\boldsymbol{x}_q,*}\mid\mathcal{H}_{q-1}\right]+\frac{\Phi(\hat{\boldsymbol{x}}^*)-\Phi(\boldsymbol{x}_1)}{\eta}\\&\leq 4(1-1/e)^2L_1^2D^2\eta d^4Q+\frac{\Phi(\hat{\boldsymbol{x}}^*)-\Phi(\boldsymbol{x}_1)}{\eta}\\&\leq 4(1-1/e)^2L_1^2D^2\eta d^4Q+\frac{\nu\log(\frac{1}{1-(1+\gamma)^{-1}})}{\eta}\end{aligned}\tag{13}$$

The last inequality is because $\pi_{\boldsymbol{x}_1}(\hat{\boldsymbol{x}}^*)\leq(1+\gamma)^{-1}$ and Lemma A.2. Since $\overline{f}_q$ is a monotone DR-submodular function by Lemma A.4 and $\overline{F}_q$ is its auxiliary function, we lower bound the left hand side of (13) by Lemma 2.6,

$$\begin{aligned}\sum_{q=1}^{Q}\mathbb{E}\left[\langle\nabla\overline{F}_q(\boldsymbol{x}_q),\hat{\boldsymbol{x}}^*-\boldsymbol{x}_q\rangle\mid\mathcal{H}_{q-1}\right]&\geq\sum_{q=1}^{Q}\mathbb{E}\left[\left((1-1/e)\overline{f}_q(\hat{\boldsymbol{x}}^*)-\overline{f}_q(\boldsymbol{x}_q)\right)\mid\mathcal{H}_{q-1}\right]\\&=\sum_{q=1}^{Q}\sum_{t=(q-1)L+1}^{qL}\frac{1}{L}\mathbb{E}\left[\left((1-1/e)f_t(\hat{\boldsymbol{x}}^*)-f_t(\boldsymbol{x}_q)\right)\mid\mathcal{H}_{q-1}\right]\\&=\frac{1}{L}\sum_{t=1}^{T}\mathbb{E}\left[(1-1/e)f_t(\hat{\boldsymbol{x}}^*)-f_t\left(\boldsymbol{x}_{\lceil\frac{t}{L}\rceil}\right)\mid\mathcal{H}_{\lceil\frac{t}{L}\rceil-1}\right]\\&=\underbrace{\frac{1}{L}\sum_{t=1}^{T}\mathbb{E}\left[(1-1/e)f_t(\hat{\boldsymbol{x}}^*)-(1-1/e)f_t(\boldsymbol{x}^*)\mid\mathcal{H}_{\lceil\frac{t}{L}\rceil-1}\right]}_{(A)}\\&\quad+\frac{1}{L}\sum_{t=1}^{T}\mathbb{E}\left[(1-1/e)f_t(\boldsymbol{x}^*)-f_t\left(\boldsymbol{y}_t\right)\mid\mathcal{H}_{\lceil\frac{t}{L}\rceil-1}\right]\\&\quad+\underbrace{\frac{1}{L}\sum_{q=1}^{Q}\mathbb{E}\left[f_{t_q}(\boldsymbol{y}_{t_q})-f_{t_q}\left(\boldsymbol{x}_{\lceil\frac{t_q}{L}\rceil}\right)\mid\mathcal{H}_{q-1}\right]}_{(B)}\end{aligned}$$

Since $|f_t(\hat{\boldsymbol{x}}^*)-f_t(\boldsymbol{x}^*)|\leq L_1\|\hat{\boldsymbol{x}}^*-\boldsymbol{x}^*\|\leq L_1\gamma D$ and $|f_{t_q}(\boldsymbol{y}_{t_q})-f_{t_q}\left(\boldsymbol{x}_{\lceil\frac{t_q}{L}\rceil}\right)|\leq M$, we have

$$|(A)|\leq(1-1/e)\frac{L_1}{L}\gamma DT\qquad|(B)|\leq\frac{MQ}{L}\tag{14}$$

Therefore,

$$\begin{aligned}\mathbb{E}\left[\sum_{t=1}^{T}(1-1/e)f_t(\boldsymbol{x}^*)-f_t(\boldsymbol{y}_t)\right]&\leq L\sum_{q=1}^{Q}\mathbb{E}\left[\langle\nabla\overline{F}_q(\boldsymbol{x}_q),\hat{\boldsymbol{x}}^*-\boldsymbol{x}_q\rangle\mid\mathcal{H}_{q-1}\right]-L\times(A)-L\times(B)\\&\leq 4(1-1/e)^2L_1^2D^2\eta d^4T+\frac{\nu L\log(\frac{1}{1-(1+\gamma)^{-1}})}{\eta}+(1-1/e)L_1\gamma DT+MQ\end{aligned}$$

The last inequality is because of (13) and (14). set $\eta=d^{-8/3}T^{-1/3}$, $L=d^{-4/3}T^{1/3}$, $Q=T/L=d^{4/3}T^{2/3}$, $\gamma=\frac{1}{T}$, we have,

$$\begin{aligned}\mathbb{E}\left[\sum_{t=1}^{T}(1-1/e)f_t(\boldsymbol{x}^*)-f_t(\boldsymbol{y}_t)\right]&\leq(4(1-1/e)^2L_1^2D^2+M)d^{4/3}T^{2/3}+(1-1/e)L_1D+\nu d^{4/3}T^{2/3}\log(T+1)\\&=O(\nu d^{4/3}T^{2/3}\log(T))\end{aligned}$$

□

Algorithm 3 `BanditDRSM`(η, δ, L, Φ)

Input: Smoothing radius δ, block size L, block number $Q = T/L$, learning rate η,self-concordant barrier Φ

initiate $\boldsymbol{x}_1 \in \text{int}(\mathcal{K})$ such that $\nabla\Phi(\boldsymbol{x}_1) = 0$
for $q = 1, 2, \ldots, Q$ **do**
  Draw $t_q \sim \text{Unif}\{(q-1)L+1, (q-1)L+2, \ldots, qL\}$
  for $t = (q-1)L+1, (q-1)L+2, \ldots, qL$ **do**
    if $t = t_q$ **then**
      $\boldsymbol{H}_q = \left(\nabla^2\Phi(\boldsymbol{x}_q)\right)^{-1/2}$
      sample z_q from $\mathbf{Z}$ where $P(\mathbf{Z} \le z) = \int_0^z \frac{e^{u-1}}{1-e^{-1}}\mathbb{I}\left[u \in [0,1]\right] du$
      draw $\boldsymbol{v}_q \sim \mathbb{S}_{d-1}$
      play $\boldsymbol{y}_t = z_q \cdot \boldsymbol{x}_q + \delta z_q \cdot \boldsymbol{H}_q \boldsymbol{v}_q$
      $\widetilde{\nabla}\overline{F}_q(\boldsymbol{x}_q) \leftarrow (1-1/e)\frac{d}{\delta z_q} f_{t_q}(\boldsymbol{y}_t)\boldsymbol{H}_q^{-1}\boldsymbol{v}_q$
      $\boldsymbol{x}_{q+1} \leftarrow \underset{\boldsymbol{x}\in\mathcal{K}}{\text{argmin}} \sum_{s=1}^{q}\langle -\eta\widetilde{\nabla}\overline{F}_s(\boldsymbol{x}_s), \boldsymbol{x}\rangle + \Phi(\boldsymbol{x})$
    else
      play $\boldsymbol{y}_t = \boldsymbol{x}_q$
    end if
  end for
end for

C. Bandit DR-submodular Maximization

In this section we present our algorithm `BanditDRSM` for general bandit monotone DR-submodular maximization, the pseudocode is shown in Algorithm 3. `BanditDRSM` is very similar to `BanditMLSM`, it also divides T rounds into Q equal size blocks. We use again $\overline{f}_q(\boldsymbol{x})$ and $\overline{F}_q(\boldsymbol{x})$ to denote the average function of q-th block and the auxiliary function of it, defined as (3) and (4). `BanditDRSM` runs RFTL with self-concordant regularizer on vector sequence $\{\nabla\overline{F}_q(\boldsymbol{x}_q)\}_{q=1}^{Q}$. Here the difference compared with `BanditMLSM` is, we cannot find an unbiased estimator for $\nabla\overline{F}_q(\boldsymbol{x}_q)$. We use the ellipsoid estimator directly to estimate $\nabla\overline{F}_q^{\delta\boldsymbol{H}_q}(\boldsymbol{x}_q)$, the gradient of the $\delta\boldsymbol{H}_q$-smoothed function, here $\boldsymbol{H}_q = (\nabla^2\Phi(\boldsymbol{x}_q))^{-1/2}$ is the same as `BanditMLSM`, δ is a parameter to be determined. Specifically, in block q, we select a uniform random exploration round $t_q \in [(q-1)L+1, qL] \cap \mathbb{Z}$, a random direction $\boldsymbol{v}_q \in \mathbb{S}_{d-1}$, $z_q \sim Z$ where $\Pr(Z \le z) = \int_0^z \frac{e^{u-1}}{1-e^{-1}}\mathbb{I}[u \in \{0,1\}]du$. In round t_q, we play $\boldsymbol{y}_{t_q} = z_q \cdot \boldsymbol{x}_q + \delta z_q \cdot \boldsymbol{H}_q\boldsymbol{v}_q$ and feedback the gradient estimate as follow to RFTL and define the estimator $\widetilde{\nabla}\overline{F}(\boldsymbol{x}_q)$.

$$\widetilde{\nabla}\overline{F}(\boldsymbol{x}_q) := (1-1/e)\frac{d}{z_q\delta} \cdot f_{t_q}(\boldsymbol{y}_{t_q})\boldsymbol{H}_q^{-1}\boldsymbol{v}_q. \tag{15}$$

$\boldsymbol{y}_{t_q} = z_q(\boldsymbol{x}_q + \delta\boldsymbol{H}_q\boldsymbol{v}_q)$, if we let $\delta \le 1$, then $\boldsymbol{x}_q + \delta\boldsymbol{H}_q\boldsymbol{v}_q$ is in the Dikin ellipsoid $\{\boldsymbol{x} \mid \|x - x_q\|_{\Phi,\boldsymbol{x}_q} \le 1\}$, therefore $\boldsymbol{x}_q + \delta\boldsymbol{H}_q^{-1}\boldsymbol{v}_q \in \mathcal{K}$. Since $\mathbf{0} \in \mathcal{K}$, $z_q \in [0,1]$ and $\mathcal{K}$ is convex, $\boldsymbol{y}_{t_q} \in \mathcal{K}$. When $z_q = 0$, we define $\widetilde{\nabla}\overline{F}(\boldsymbol{x}_q) := 0$, since $\Pr(z_q = 0) = 0$, the value of $\widetilde{\nabla}\overline{F}(\boldsymbol{x}_q)$ when $z_q = 0$ does not matter.

We prove that $\widetilde{\nabla}\overline{F}(\boldsymbol{x}_q)$ is an unbiased gradient estimator for the $\delta\boldsymbol{H}_q$-smoothed function $\overline{F}_q^{\delta\boldsymbol{H}_q}(\boldsymbol{x}_q)$. Moreover, the dual local norm of the estimator can be bounded as $O(\frac{d^2}{\delta^2})$. To formalize the above arguments, we have the following lemma.

Lemma C.1. *Let $\widetilde{\nabla}\overline{F}_q(\boldsymbol{x}_q)$ be defined as (15). Assume f_t for $t \in [(q-1)L+1, qL]$ is L_1-lipschitz, $f_t(\mathbf{0}) = 0$, then the following hold,*

(i) $\mathbb{E}\left[\widetilde{\nabla}\overline{F}_q(\boldsymbol{x}_q) \mid \mathcal{H}_{q-1}\right] = \nabla\overline{F}_q^{\delta\boldsymbol{H}_q}(\boldsymbol{x}_q)$.

(ii) $\|\widetilde{\nabla}\overline{F}_q(\boldsymbol{x}_q)\|_{\boldsymbol{x}_q,*}^2 \le \frac{(1-e)^2d^2L_1^2D^2}{\delta^2}$

Proof. (i) Let $\boldsymbol{H} = \delta \boldsymbol{H}_q$ in Lemma 2.5, let $f_{t_q,z_q}(\boldsymbol{x}) \triangleq f_{t_q}(z_q \cdot \boldsymbol{x})$,we have

$$\mathbb{E}\left[\widetilde{\nabla}\overline{F}_q(\boldsymbol{x}_q) \mid \mathcal{H}_{q-1}, t_q, z_q\right] = \frac{1-1/e}{z_q}\nabla f^{\delta \boldsymbol{H}_q}_{t_q,z_q}(\boldsymbol{x}_q)$$

Thus,

$$\begin{aligned}
\mathbb{E}\left[\widetilde{\nabla}\overline{F}_q(\boldsymbol{x}_q) \mid \mathcal{H}_{q-1}\right] &= \sum_{t_q=(q-1)L+1}^{qL} \int_0^1 \Pr\left(t_q, z_q \mid \mathcal{H}_{q-1}\right) \frac{1-1/e}{z_q}\nabla f^{\delta \boldsymbol{H}_q}_{t_q,z_q}(\boldsymbol{x}_q) dz_q \\
&= \sum_{t_q=(q-1)L+1}^{qL} \int_0^1 \frac{e^{z_q-1}}{(1-1/e)L}\frac{1-1/e}{z_q}\nabla f^{\delta \boldsymbol{H}_q}_{t_q,z_q}(\boldsymbol{x}_q) dz_q \\
&= \sum_{t_q=(q-1)L+1}^{qL} \int_0^1 \frac{e^{z_q-1}}{z_q L}\nabla f^{\delta \boldsymbol{H}_q}_{t_q,z_q}(\boldsymbol{x}_q) dz_q \\
&= \nabla \overline{F}^{\delta \boldsymbol{H}_q}_q(\boldsymbol{x}_q)
\end{aligned}$$

(ii)

$$\begin{aligned}
\|\widetilde{\nabla}\overline{F}_q(\boldsymbol{x}_q)\|^2_{\boldsymbol{x}_q,*} &= (1-1/e)^2 \frac{d^2}{\delta^2 z_q^2} f^2_{t_q}(z_q \cdot \boldsymbol{x}_q + \delta z_q \cdot \boldsymbol{H_q v}_q)\boldsymbol{v}_q^T \boldsymbol{H}_q^{-1}\left(\nabla^2\Phi(\boldsymbol{x}_q)\right)^{-1}\boldsymbol{H}_q^{-1}\boldsymbol{v}_q \\
&\le (1-1/e)^2 \frac{d^2}{\delta^2 z_q^2} z_q^2 L_1^2 \|\boldsymbol{x}_q + \delta \boldsymbol{H}_q \boldsymbol{v}_q\|^2 \|\boldsymbol{v}_q\|^2 \\
&\le \frac{(1-e)^2 d^2 L_1^2 D^2}{\delta^2}
\end{aligned}$$

The first inequality is because f_{t_q} is L_1-lipschitz continuous.

□

Intuitively, we can control the regret of $\{\boldsymbol{x}_q\}_{q=1}^Q$ w.r.t. the linear function sequence $\{\langle \cdot, \nabla \overline{F}^{\delta \boldsymbol{H}_q}_q(\boldsymbol{x}_q)\rangle\}_{q=1}^Q$ by using Theorem 2.3. In Lemma A.5, We proved that $\overline{f}^{\delta \boldsymbol{H}_q}_q$ is also DR-submodular. Since $\overline{F}^{\delta \boldsymbol{H}_q}_q(\boldsymbol{x}_q)$ is the auxiliary function of $\overline{f}^{\delta \boldsymbol{H}_q}_q$, this allows us to control the $(1-1/e)$-regret of $\{\boldsymbol{x}_q\}$ w.r.t. $\{\overline{f}^{\delta \boldsymbol{H}_q}_q\}$ by using Lemma 2.6. A key observation here is $\|\overline{f}^{\delta \boldsymbol{H}_q}_q - \overline{f}_q\|_\infty \le O(\delta^2)$ assuming the online functions are smooth, which means we can bound the $(1-1/e)$-regret of $\{\boldsymbol{x}_q\}$ w.r.t. $\{\overline{f}_q\}$ in term of the $(1-1/e)$-regret w.r.t. $\{\overline{f}^{\delta \boldsymbol{H}_q}_q\}$ with an extra $O(\delta^2)$ additive term. Previous works (Zhang et al., 2019; Niazadeh et al., 2021) use the FKM estimator proposed in (Flaxman et al., 2005), where the sample sphere is fixed(which can be seen as a special case of the ellipsoid estimator when $\boldsymbol{H}_q = I$), to prevent the sample action jump out $\mathcal{K}$, they must run their algorithm on a smaller interior $\mathcal{K}_\delta$ which is δ-far from $\partial\mathcal{K}$. So this only guarantees the regret competing with the point in $\mathcal{K}_\delta$, this adds an $O(\delta)$ term to the overall regret, which is bigger than $O(\delta^2)$ since the δ is set to $o(1)$ latter.

With this improved gradient estimator and non-oblivious technique, we prove a $\widetilde{O}(T^{3/4})$ $(1-1/e)$-regret of `BanditDRSM`.

Theorem 4.1 (restatement)**.** *Set $\eta = D^{-2}d^{-1}T^{-1/2}$, $\delta = d^{1/4}T^{-1/8}$, $L = d^{-1/2}T^{1/4}$, $Q = T/L = d^{1/2}T^{3/4}$ in Algorithm 3. If Φ is a ν-self concordant function of $\mathcal{K}$, then the expected $(1-1/e)$-regret of Algorithm 3 can be bounded as*

$$\mathcal{R}_{1-1/e}(T) \le O(\nu d^{1/2}T^{3/4}\log(T))$$

Proof of Theorem 4.1. Let $\hat{\boldsymbol{x}}^* = \operatorname{argmin}_{\boldsymbol{x}\in\mathcal{K}_{\gamma,\boldsymbol{x}_1}} \|\boldsymbol{x}^* - \boldsymbol{x}\|$, where $\boldsymbol{x}^* = \operatorname{argmax}_{\boldsymbol{x}\in\mathcal{K}} \sum_{t=1}^T f_t(\boldsymbol{x}^*)$. Let $\boldsymbol{g}_q = \nabla \overline{F}^{\delta \boldsymbol{H}_q}_q$, $\widetilde{\boldsymbol{g}}_q = \widetilde{\nabla}\overline{F}_q(\boldsymbol{x}_q)$, $\boldsymbol{y} = \hat{\boldsymbol{x}}^*$ in Theorem 2.3. Since we proved $\widetilde{\nabla}\overline{F}_q(\boldsymbol{x}_q)$ is an unbiased estimate of $\nabla \overline{F}^{\delta \boldsymbol{H}_q}_q(\boldsymbol{x}_q)$ in Lemma C.1, we have

$$\sum_{q=1}^Q \mathbb{E}\left[\langle \nabla \overline{F}^{\delta \boldsymbol{H}_q}_q(\boldsymbol{x}_q), \hat{\boldsymbol{x}}^* - \boldsymbol{x}_q\rangle \mid \mathcal{H}_{q-1}\right] \le \eta \sum_{q=1}^Q \mathbb{E}\left[\|\widetilde{\nabla}\overline{F}_q(\boldsymbol{x}_q)\|^2_{\Phi,\boldsymbol{x}_q,*} \mid \mathcal{H}_{q-1}\right] + \frac{\Phi(\hat{\boldsymbol{x}}^*) - \Phi(\boldsymbol{x}_1)}{\eta}$$

$$\leq \eta \sum_{q=1}^{Q} \frac{(1-e)^2 d^2 L_1^2 D^2}{\delta^2} + \frac{\Phi(\hat{\boldsymbol{x}}^*) - \Phi(\boldsymbol{x}_1)}{\eta}$$
$$\leq \frac{(1-e)^2 \eta Q d^2 L_1^2 D^2}{\delta^2} + \frac{\nu \log(\frac{1}{1-(1+\gamma)^{-1}})}{\eta}$$

Since $\nabla \overline{F}_q^{\delta \boldsymbol{H}_q}(\boldsymbol{x})$ is monotone DR-submodular due to Lemma A.5, and it is the auxiliary function of $\overline{f}_q^{\delta \boldsymbol{H}_q}(\boldsymbol{x})$, by Lemma 2.6, we have

$$\sum_{q=1}^{Q} \mathbb{E}\left[\langle \nabla \overline{F}_q^{\delta \boldsymbol{H}_q}(\boldsymbol{x}_q), \hat{\boldsymbol{x}}^* - \boldsymbol{x}_q \rangle \mid \mathcal{H}_{q-1}\right] \geq \sum_{q=1}^{Q} \mathbb{E}\left[(1-1/e)\overline{f}_q^{\delta \boldsymbol{H}_q}(\hat{\boldsymbol{x}}^*) - \overline{f}_q^{\delta \boldsymbol{H}_q}(\boldsymbol{x}_q) \mid \mathcal{H}_{q-1}\right]$$

The RHS can be further decomposed into several terms,

$$\begin{aligned}
&\sum_{q=1}^{Q} \mathbb{E}\left[(1-1/e)\overline{f}_q^{\delta \boldsymbol{H}_q}(\hat{\boldsymbol{x}}^*) - \overline{f}_q^{\delta \boldsymbol{H}_q}(\boldsymbol{x}_q) \mid \mathcal{H}_{q-1}\right] \\
&= \underbrace{\sum_{q=1}^{Q} \mathbb{E}\left[(1-1/e)\overline{f}_q^{\delta \boldsymbol{H}_q}(\hat{\boldsymbol{x}}^*) - (1-1/e)\overline{f}_q^{\delta \boldsymbol{H}_q}(\boldsymbol{x}^*) \mid \mathcal{H}_{q-1}\right]}_{(A)} + \underbrace{\sum_{q=1}^{Q} \mathbb{E}\left[(1-1/e)\overline{f}_q^{\delta \boldsymbol{H}_q}(\boldsymbol{x}^*) - (1-1/e)\overline{f}_q(\boldsymbol{x}^*) \mid \mathcal{H}_{q-1}\right]}_{(B)} \\
&\quad + \sum_{q=1}^{Q} \mathbb{E}\left[(1-1/e)\overline{f}_q(\boldsymbol{x}^*) - \overline{f}_q(\boldsymbol{x}_q) \mid \mathcal{H}_{q-1}\right] + \underbrace{\sum_{q=1}^{Q} \mathbb{E}\left[\overline{f}_q(\boldsymbol{x}_q) - \overline{f}_q^{\delta \boldsymbol{H}_q}(\boldsymbol{x}_q) \mid \mathcal{H}_{q-1}\right]}_{(C)}
\end{aligned} \tag{16}$$

Bounding (A): Since $f_t(\boldsymbol{x})$ is L_1-lipschitz continuous for any t, $\overline{f}_q$ is also L_1-lipschitz continuous by Lemma A.4, thus $\overline{f}_q^{\delta \boldsymbol{H}_q}$ is L_1-lipschitz continuous by Lemma A.5. Since $\|\hat{\boldsymbol{x}}^* - \boldsymbol{x}^*\| \leq \gamma D$ by Lemma A.3,

$$\begin{aligned}
\sum_{q=1}^{Q} \mathbb{E}\left[(1-1/e)\overline{f}_q^{\delta \boldsymbol{H}_q}(\hat{\boldsymbol{x}}^*) - (1-1/e)\overline{f}_q^{\delta \boldsymbol{H}_q}(\boldsymbol{x}^*) \mid \mathcal{H}_{q-1}\right] &\geq -\sum_{q=1}^{Q}(1-1/e)\,\mathbb{E}\left[|\overline{f}_q^{\delta \boldsymbol{H}_q}(\hat{\boldsymbol{x}}^*) - \overline{f}_q^{\delta \boldsymbol{H}_q}(\boldsymbol{x}^*)| \mid \mathcal{H}_{q-1}\right] \\
&\geq -\sum_{q=1}^{Q}(1-1/e)L_1\gamma D = -(1-1/e)L_1\gamma DQ
\end{aligned} \tag{17}$$

Bounding (B): Since $f_t(\boldsymbol{x})$ is L_2-smooth for any t, by Lemma A.4 and Lemma A.5, $\overline{f}_q^{\delta \boldsymbol{H}_q}$ is L_2-smooth. Thus,

$$\begin{aligned}
\overline{f}_q^{\delta \boldsymbol{H}_q}(\boldsymbol{x}^*) - \overline{f}_q(\boldsymbol{x}^*) &= \frac{1}{\mathrm{Vol}(\mathbb{B}_d)} \int_{\boldsymbol{v} \in \mathbb{B}_d} \overline{f}_q(\boldsymbol{x}^* + \delta \boldsymbol{H}_q \boldsymbol{v}) - \overline{f}_q(\boldsymbol{x}^*) d\boldsymbol{v} \\
&\geq \frac{1}{\mathrm{Vol}(\mathbb{B}_d)} \int_{\boldsymbol{v} \in \mathbb{B}_d} \langle \nabla \overline{f}_q(\boldsymbol{x}^*), \delta \boldsymbol{H}_q \boldsymbol{v} \rangle - \frac{L_2}{2}\|\delta \boldsymbol{H}_q \boldsymbol{v}\|^2 d\boldsymbol{v} \\
&= \frac{1}{\mathrm{Vol}(\mathbb{B}_d)} \left\langle \nabla \overline{f}_q(\boldsymbol{x}^*), \delta \boldsymbol{H}_q \int_{v \in \mathbb{B}_d} \boldsymbol{v} d\boldsymbol{v} \right\rangle - \frac{1}{\mathrm{Vol}(\mathbb{B}_d)} \int_{v \in \mathbb{B}_d} \frac{L_2}{2}\|\delta \boldsymbol{H}_q \boldsymbol{v}\|^2 d\boldsymbol{v} \\
&\geq -\frac{1}{\mathrm{Vol}(\mathbb{B}_d)} \int_{\boldsymbol{v} \in \mathbb{B}_d} \frac{L_2}{2}\delta^2 D^2 d\boldsymbol{v} \\
&\geq -\frac{L_2 \delta^2 D^2}{2}
\end{aligned}$$

Therefore,

$$\sum_{q=1}^{Q} \mathbb{E}\left[(1-1/e)\overline{f}_q^{\delta \boldsymbol{H}_q}(\hat{\boldsymbol{x}}^*) - (1-1/e)\overline{f}_q(\hat{\boldsymbol{x}}^*) \mid \mathcal{H}_{q-1}\right] \geq -\frac{(1-1/e)L_2 \delta^2 D^2 Q}{2} \tag{18}$$

Bounding (C): Similarly,

$$\begin{aligned}\overline{f}_q(\boldsymbol{x}_q) - \overline{f}_q^{\delta \boldsymbol{H}_q}(\boldsymbol{x}_q) &= \frac{1}{\text{Vol}(\mathbb{B}_d)} \int_{\boldsymbol{v}\in\mathbb{B}_d} \overline{f}_q(\boldsymbol{x}_q) - \overline{f}_q(\boldsymbol{x}_q + \delta \boldsymbol{H}_q \boldsymbol{v}) d\boldsymbol{v} \\ &\geq \frac{1}{\text{Vol}(\mathbb{B}_d)} \int_{\boldsymbol{v}\in\mathbb{B}_d} \langle \nabla \overline{f}_q(\boldsymbol{x}_q), \delta \boldsymbol{H}_q \boldsymbol{v} \rangle - \frac{L_2}{2} \|\delta \boldsymbol{H}_q \boldsymbol{v}\|^2 d\boldsymbol{v} \\ &\geq -\frac{L_2 \delta^2 D^2}{2}\end{aligned}$$

Therefore,

$$\sum_{q=1}^{Q} \mathbb{E}\left[\overline{f}_q(\boldsymbol{x}_q) - \overline{f}_q^{\delta \boldsymbol{H}_q}(\boldsymbol{x}_q) \mid \mathcal{H}_{q-1}\right] \geq -\frac{L_2 \delta^2 D^2 Q}{2} \tag{19}$$

Put (17),(18),(19) in (16) and rearrange it,

$$\begin{aligned}&\sum_{q=1}^{Q} \mathbb{E}\left[(1-1/e)\overline{f}_q(\boldsymbol{x}^*) - \overline{f}_q(\boldsymbol{x}_q) \mid \mathcal{H}_{q-1}\right] \\ &\leq \sum_{q=1}^{Q} \mathbb{E}\left[\langle \nabla \overline{F}_q^{\delta \boldsymbol{H}_q}(\boldsymbol{x}_q), \hat{\boldsymbol{x}}^* - \boldsymbol{x}_q \rangle \mid \mathcal{H}_{q-1}\right] + (1-1/e) L_1 \gamma D Q + \frac{(1-1/e) L_2 \delta^2 D^2 Q}{2} + \frac{L_2 \delta^2 D^2 Q}{2} \\ &\leq \frac{(1-e)^2 \eta Q d^2 L_1^2 D^2}{\delta^2} + \frac{\nu \log(\frac{1}{1-(1+\gamma)^{-1}})}{\eta} + (1-1/e) L_1 \gamma D Q + \frac{(2-1/e) L_2 \delta^2 D^2 Q}{2}\end{aligned}$$

Then we bound the expected regret,

$$\begin{aligned}\mathcal{R}_{1-1/e}(T) &= \sum_{t=1}^{T} \mathbb{E}\left[(1-1/e) f_t(\boldsymbol{x}^*) - f_t(\boldsymbol{y}_t) \mid \mathcal{H}_{\lceil \frac{t}{L} \rceil - 1}\right] \\ &= \sum_{t=1}^{T} \mathbb{E}\left[(1-1/e) f_t(\boldsymbol{x}^*) - f_t(\boldsymbol{x}_{\lceil \frac{t}{L} \rceil}) \mid \mathcal{H}_{\lceil \frac{t}{L} \rceil - 1}\right] + \sum_{q=1}^{Q} \mathbb{E}\left[f_{t_q}(\boldsymbol{x}_q) - f_{t_q}(\boldsymbol{y}_{t_q}) \mid \mathcal{H}_{q-1}\right] \\ &\leq L \sum_{q=1}^{Q} \mathbb{E}\left[(1-1/e)\overline{f}_q(\boldsymbol{x}^*) - \overline{f}_q(\boldsymbol{x}_q) \mid \mathcal{H}_{q-1}\right] + MQ \\ &\leq \frac{(1-e)^2 \eta L Q d^2 L_1^2 D^2}{\delta^2} + \frac{\nu L \log(\frac{1}{1-(1+\gamma)^{-1}})}{\eta} + (1-1/e) L_1 \gamma D L Q + \frac{(2-1/e) L_2 \delta^2 D^2 L Q}{2} + MQ \\ &= \frac{(1-e)^2 \eta d^2 L_1^2 D^2 T}{\delta^2} + \frac{\nu \log(\frac{1}{1-(1+\gamma)^{-1}}) L}{\eta} + (1-1/e) L_1 \gamma D T + \frac{(2-1/e) L_2 \delta^2 D^2 T}{2} + MQ\end{aligned}$$

Set $\eta = D^{-2} d^{-1} T^{-1/2}$, $\delta = D^{-1/2} d^{1/4} T^{-1/8}$, $L = D^{-1} d^{-1/2} T^{1/4}$, $Q = T/L = D d^{1/2} T^{3/4}$, $\gamma = \frac{1}{T}$

$$\begin{aligned}R_{1-1/e}(T) &\leq (1-e)^2 L_1^2 D d^{1/2} T^{3/4} + \nu D d^{1/2} T^{3/4} \log(T+1) + (1-1/e) L_1 D \\ &\quad + \frac{(2-1/e) L_2 \delta^2 D d^{1/2} T^{3/4}}{2} + M D d^{1/2} T^{3/4} \\ &= O(\nu d^{1/2} T^{3/4} \log(T))\end{aligned}$$

□

The idea of using a self-concordant regularizer RFTL on smooth online functions is motivated by (Saha & Tewari, 2011). Where the authors studied the bandit convex optimization problem, and they find that RFTL with the self-concordant regularizer works well when the convex functions are smooth. We find this idea also works here in the bandit DR-submodular maximization problem.

D. Self-Concordant Barrier of Product Simplexes

In this section, we give a self-concordant barrier for the product simplex, which is a cartesian product of several simplexes.

Let $\mathcal{K}$ be the product of n simplexes, and their dimensions are $d_1, d_2, \ldots, d_n$ respectively. We write $\mathcal{K}$ as

$$\mathcal{K} = \prod_{i=1}^{n} \Delta_{d_i}.$$

For $\boldsymbol{x} \in \mathcal{K}$, we represent it as $\boldsymbol{x} = (x_{1,1}, x_{1,2}, \ldots, x_{1,d_1}, x_{2,1}, \ldots, x_{2,d_2}, \ldots, x_{n,d_n})$. $\boldsymbol{x} \in \mathcal{K}$ iff

$$\begin{cases} x_{i,j} \geq 0, & \forall 1 \leq i \leq n \text{ and } 1 \leq j \leq d_i \\ \sum_{j=1}^{d_i} x_{i,j} \leq 1, & \forall 1 \leq i \leq n \end{cases}$$

Define the function $\Phi : \text{int}(\mathcal{K}) \to \mathbb{R}$,

$$\Phi(\boldsymbol{x}) = -\sum_{i=1}^{n} \log(1 - \vec{1}_{d_i}^T \cdot \boldsymbol{x}_i) - \sum_{i=1}^{n} \sum_{j=1}^{d_i} \log(x_{i,j}).$$

Here $\vec{1}_{d_i} = \underbrace{(1, 1, \ldots, 1)}_{d_i}^T$, $\boldsymbol{x}_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,d_i})$. We prove that Φ is a n-self-concordant barrier of $\mathcal{K}$.

Lemma D.1. *$\Phi(\boldsymbol{x})$ is a $\sum_{i=1}^{n}(d_i + 1)$-self-concordant barrier of $\mathcal{K}$.*

Proof. It's easy to see that $\Phi(\boldsymbol{x})$ is three-times continuously differentiable and approaches infinity alone any sequence of points approaching the boundary of $\mathcal{K}$. We first calculate the gradient and the hessian matrix of Φ.

$$\begin{aligned} \frac{\partial \Phi}{\partial x_{i,j}}(\boldsymbol{x}) &= -\sum_{i=1}^{n} \frac{\partial \log(1 - \vec{1}_{d_i}^T \cdot \boldsymbol{x}_i)}{\partial x_{i,j}} - \sum_{i=1}^{n} \sum_{j=1}^{d_i} \frac{\partial \log(x_{i,j})}{\partial x_{i,j}} \\ &= \frac{1}{1 - \vec{1}^T \cdot \boldsymbol{x}_i} - \frac{1}{x_{i,j}} \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 \Phi(\boldsymbol{x})}{\partial x_{i_1,j_1} \partial x_{i_2,j_2}}(\boldsymbol{x}) &= \frac{\partial (1 - \vec{1}^T \cdot \boldsymbol{x}_{i_1})^{-1}}{\partial x_{i_1,j_1}} - \frac{\partial x_{i_2,j_2}^{-1}}{\partial x_{i_1,j_1}} \\ &= \frac{1}{(1 - \vec{1}^T \cdot \boldsymbol{x}_{i_1})^2} \mathbb{I}[i_1 = i_2] + \frac{1}{x_{i_1,j_1}^2} \mathbb{I}[i_1 = i_2, j_1 = j_2] \end{aligned}$$

For any direction $\boldsymbol{h} = (h_{1,1}, \ldots, h_{1,d_1}, h_{2,1}, \ldots, h_{n,d_n})^T$,

$$\begin{aligned} \boldsymbol{h}^T \nabla^2 \Phi(\boldsymbol{x}) \boldsymbol{h} &= \sum_{i_1=1}^{n} \sum_{j_1=1}^{d_i} \sum_{i_2=1}^{n} \sum_{j_2=1}^{d_i} h_{i_1,j_1} h_{i_2,j_2} \frac{\partial^2 \Phi(\boldsymbol{x})}{\partial x_{i_1,j_1} \partial x_{i_2,j_2}}(\boldsymbol{x}) \\ &= \sum_{i_1=1}^{n} \sum_{j_1=1}^{d_i} \sum_{i_2=1}^{n} \sum_{j_2=1}^{d_i} \left(\frac{h_{i_1,j_1} h_{i_2,j_2}}{(1 - \vec{1}^T \cdot \boldsymbol{x}_{i_1})^2} \mathbb{I}[i_1 = i_2] + \frac{h_{i_1,j_1} h_{i_2,j_2}}{x_{i_1,j_1}^2} \mathbb{I}[i_1 = i_2, j_1 = j_2] \right) \\ &= \sum_{i=1}^{n} \frac{(\sum_{j=1}^{d_i} h_{i,j})^2}{(1 - \vec{1}^T \cdot \boldsymbol{x}_{i_1})^2} + \sum_{i=1}^{n} \sum_{j=1}^{d_i} \frac{h_{i,j}^2}{x_{i,j}^2} \geq 0 \end{aligned}$$

Therefore, Φ is convex. Next, we check the condition 2 of Definition 2.1. Let $\boldsymbol{h} = (h_{1,1}, h_{1,d_1}, h_{2,1}, \ldots, h_{2,d_2}, \ldots, h_{n,d_n})$, $\boldsymbol{h}_i = (h_{i,1}, \ldots, h_{i,d_i})$. Then,

$$
\begin{aligned}
&\nabla^3 \Phi(\boldsymbol{x})[\boldsymbol{h}, \boldsymbol{h}, \boldsymbol{h}] \\
&= \frac{\partial^3}{\partial t_1 \partial t_2 \partial t_3} \left(-\sum_{i=1}^{n} \log(1 - \vec{1}_{d_i}^T \cdot \boldsymbol{x}_i - t_1 \vec{1}_{d_i}^T \cdot \boldsymbol{h}_i - t_2 \vec{1}_{d_i}^T \cdot \boldsymbol{h}_i - t_3 \vec{1}_{d_i}^T \cdot \boldsymbol{h}_i) - \sum_{i=1}^{n} \sum_{j=1}^{d_i} \log(x_{i,j} + (t_1 + t_2 + t_3) h_{i,j}) \right) \Bigg|_{t_1 = t_2 = t_3 = 0} \\
&= \sum_{i=1}^{n} \frac{2(\vec{1}_{d_i}^T \cdot \boldsymbol{h}_i)^3}{(1 - \vec{1}_{d_i}^T \cdot \boldsymbol{x}_i - t_1 \vec{1}_{d_i}^T \cdot \boldsymbol{h}_i - t_2 \vec{1}_{d_i}^T \cdot \boldsymbol{h}_i - t_3 \vec{1}_{d_i}^T \cdot \boldsymbol{h}_i)^3} - \sum_{i=1}^{n} \sum_{j=1}^{d_i} \frac{2 h_{i,j}^3}{(x_{i,j} + (t_1 + t_2 + t_3) h_{i,j})^3} \Bigg|_{t_1 = t_2 = t_3 = 0} \\
&= \sum_{i=1}^{n} \frac{2(\vec{1}_{d_i}^T \cdot \boldsymbol{h}_i)^3}{(1 - \vec{1}_{d_i}^T \cdot \boldsymbol{x}_i)^3} - \sum_{i=1}^{n} \sum_{j=1}^{d_i} \frac{2 h_{i,j}^3}{x_{i,j}^3}
\end{aligned}
$$

We check the first inequality in the condition 2 of Definition 2.1.

$$
\begin{aligned}
2(\nabla^2 \Phi(\boldsymbol{x})[\boldsymbol{h}, \boldsymbol{h}])^{3/2} &= 2 \left(\sum_{i=1}^{n} \frac{(\sum_{j=1}^{d_i} h_{i,j})^2}{(1 - \vec{1}^T \cdot \boldsymbol{x}_{i_1})^2} + \sum_{i=1}^{n} \sum_{j=1}^{d_i} \frac{h_{i,j}^2}{x_{i,j}^2} \right)^{3/2} \\
&= 2 \left(\sum_{i=1}^{n} \frac{(\sum_{j=1}^{d_i} h_{i,j})^2}{(1 - \vec{1}^T \cdot \boldsymbol{x}_{i_1})^2} + \sum_{i=1}^{n} \sum_{j=1}^{d_i} \frac{h_{i,j}^2}{x_{i,j}^2} \right) \left(\sum_{i=1}^{n} \frac{(\sum_{j=1}^{d_i} h_{i,j})^2}{(1 - \vec{1}^T \cdot \boldsymbol{x}_{i_1})^2} + \sum_{i=1}^{n} \sum_{j=1}^{d_i} \frac{h_{i,j}^2}{x_{i,j}^2} \right)^{1/2} \\
&= 2 \left(\sum_{i=1}^{n} \frac{(\sum_{j=1}^{d_i} h_{i,j})^2}{(1 - \vec{1}^T \cdot \boldsymbol{x}_{i_1})^2} \left(\sum_{i=1}^{n} \frac{(\sum_{j=1}^{d_i} h_{i,j})^2}{(1 - \vec{1}^T \cdot \boldsymbol{x}_{i_1})^2} + \sum_{i=1}^{n} \sum_{j=1}^{d_i} \frac{h_{i,j}^2}{x_{i,j}^2} \right)^{1/2} \right. \\
&\quad \left. + \sum_{i=1}^{n} \sum_{j=1}^{d_i} \frac{h_{i,j}^2}{x_{i,j}^2} \left(\sum_{i=1}^{n} \frac{(\sum_{j=1}^{d_i} h_{i,j})^2}{(1 - \vec{1}^T \cdot \boldsymbol{x}_{i_1})^2} + \sum_{i=1}^{n} \sum_{j=1}^{d_i} \frac{h_{i,j}^2}{x_{i,j}^2} \right)^{1/2} \right) \\
&\geq 2 \left(\sum_{i=1}^{n} \left| \frac{(\sum_{j=1}^{d_i} h_{i,j})^3}{(1 - \vec{1}^T \cdot \boldsymbol{x}_{i_1})^3} \right| + \sum_{i=1}^{n} \sum_{j=1}^{d_i} \left| \frac{h_{i,j}^3}{x_{i,j}^3} \right| \right) \\
&\geq \left| \sum_{i=1}^{n} \frac{2(\vec{1}_{d_i}^T \cdot \boldsymbol{h}_i)^3}{(1 - \vec{1}_{d_i}^T \cdot \boldsymbol{x}_i)^3} - \sum_{i=1}^{n} \sum_{j=1}^{d_i} \frac{2 h_{i,j}^3}{x_{i,j}^3} \right| = |\nabla^3 \Phi(\boldsymbol{x})[\boldsymbol{h}, \boldsymbol{h}, \boldsymbol{h}]|
\end{aligned}
$$

Then we check the inequality between $\nabla \Phi(\boldsymbol{x})[\boldsymbol{h}]$ and $\nabla^2 \Phi(\boldsymbol{x})[\boldsymbol{h}, \boldsymbol{h}]$.

$$
\begin{aligned}
|\nabla \Phi(\boldsymbol{x})[\boldsymbol{h}]| &= |\boldsymbol{h}^T \nabla \Phi(\boldsymbol{x})| \\
&\leq \sum_{i=1}^{n} \left| \frac{\vec{1}_{d_i}^T \cdot \boldsymbol{h}_i}{1 - \vec{1}^T \cdot \boldsymbol{x}_i} \right| + \sum_{i=1}^{n} \sum_{j=1}^{d_i} \left| \frac{h_{i,j}}{x_{i,j}} \right| \\
&\leq \sqrt{\sum_{i=1}^{n} (d_i + 1)} \left(\sum_{i=1}^{n} \frac{(\sum_{j=1}^{d_i} h_{i,j})^2}{(1 - \vec{1}^T \cdot \boldsymbol{x}_{i_1})^2} + \sum_{i=1}^{n} \sum_{j=1}^{d_i} \frac{h_{i,j}^2}{x_{i,j}^2} \right)^{1/2} \\
&= \left(\sum_{i=1}^{n} (d_i + 1) \right)^{1/2} \left(\nabla^2 \Phi(\boldsymbol{x})[\boldsymbol{h}, \boldsymbol{h}] \right)^{1/2}
\end{aligned}
$$

Therefore, $\Phi(\boldsymbol{x})$ is a $(\sum_{i=1}^{n} (d_i + 1))$-self-concordant barrier of $\mathcal{K}$. □

Algorithm 4 `BanditMLSM4PS`(η, L, Φ)

Input: block size L, block number $Q = T/L$, learning rate η, potential function Φ

1: initiate $\boldsymbol{x}_1 \in \text{int}(\mathcal{K})$ such that $\nabla\Phi(\boldsymbol{x}_1) = 0$
2: **for** $q = 1, 2, \ldots, Q$ **do**
3: Draw $t_q \sim \text{Unif}\{(q-1)L+1, (q-1)L+2, \ldots, qL\}$
4: **for** $t = (q-1)L+1, (q-1)L+2, \ldots, qL$ **do**
5: **if** $t = t_q$ **then**
6: $\boldsymbol{H}_q = \left(\nabla^2\Phi(\boldsymbol{x}_q)\right)^{-1/2}$
7: sample z_q from $\mathbf{Z}$ where $P(\mathbf{Z} < z) = \int_0^z \frac{e^{u-1}}{1-e^{-1}}\mathbb{I}\left[u \in [0,1]\right] du$
8: draw $\boldsymbol{v}_q \sim \mathbb{S}_{d-1}$
9: **if** $z_q \geq \frac{1}{2}$ **then**
10: draw $\boldsymbol{u}_q$ from $\{\mathbf{0}, \boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_d\}$ with probability: $\Pr(\boldsymbol{u}_q = \mathbf{0}) = \frac{1}{2}$, $\Pr(\boldsymbol{u}_q = \boldsymbol{e}_i) = \frac{1}{2d}$
11: $\boldsymbol{y}_{t_q} \leftarrow z_q \cdot \boldsymbol{x}_q + z_q\langle \boldsymbol{H}_q\boldsymbol{v}_q, \boldsymbol{u}_q\rangle\boldsymbol{u}_q$
12: $\widetilde{l}_q(\boldsymbol{H}_q\boldsymbol{v}_q) \leftarrow \begin{cases} -2(1-1/e)\dfrac{d}{z_q} \cdot f_{t_q}(\boldsymbol{y}_{t_q}) & \text{if } \boldsymbol{u}_q = \mathbf{0}, \\ 2(1-1/e)\dfrac{d}{z_q} \cdot f_{t_q}(\boldsymbol{y}_{t_q}) & \text{if } \boldsymbol{u}_q \neq \mathbf{0}. \end{cases}$
13: **else**
14: draw $\boldsymbol{u}_q$ from $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_d\}$ uniformly at random
15: let $\boldsymbol{y}_{t_q} = z_q\boldsymbol{x}_q + \frac{1}{2}\boldsymbol{u}_q$ or $\boldsymbol{y}_{t_q} = z_q\boldsymbol{x}_q$ with equal probability.
16: play $\boldsymbol{y}_{t_q}$ and observe the feedback $f_{t_q}(\boldsymbol{y}_{t_q})$
17: $\widetilde{l}_q(\boldsymbol{H}_q\boldsymbol{v}_q) \leftarrow \begin{cases} -4(1-1/e)d \cdot \langle \boldsymbol{H}_q\boldsymbol{v}_q, \boldsymbol{u}_q\rangle f_{t_q}(\boldsymbol{y}_{t_q}) & \text{if } \boldsymbol{y}_{t_q} = z_q\boldsymbol{x}_q, \\ 4(1-1/e)d \cdot \langle \boldsymbol{H}_q\boldsymbol{v}_q, \boldsymbol{u}_q\rangle f_{t_q}(\boldsymbol{y}_{t_q}) & \text{if } \boldsymbol{y}_{t_q} = z_q\boldsymbol{x}_q + \dfrac{1}{2}\boldsymbol{u}_q. \end{cases}$
18: **end if**
19: $\widetilde{\nabla}F_q(\boldsymbol{x}_q) \leftarrow d \cdot \widetilde{l}_q(\boldsymbol{H}_q\boldsymbol{v}_q)\boldsymbol{H}_q^{-1}\boldsymbol{v}_q$
20: $\boldsymbol{x}_{q+1} \leftarrow \operatorname{argmin}_{\boldsymbol{x}\in\mathcal{K}} \sum_{s=1}^{q}\langle -\eta\widetilde{\nabla}F_s(\boldsymbol{x}_s), \boldsymbol{x}\rangle + \Phi(\boldsymbol{x})$
21: **else**
22: $\boldsymbol{y}_t \leftarrow \boldsymbol{x}_q$,
23: sample S_t from $\text{EXT}(\boldsymbol{y}_t)$ and play S_t.
24: **end if**
25: **end for**
26: **end for**

E. Missing Proofs in Section 5

The detailed pseudo-code of `BanditMLSM4PS` is shown in Algorithm 4, the only difference between `BanditMLSM4PS` and `BanditMLSM` is the line 9 to line 17 in Algorithm 4.

E.1. Proof of Lemma 5.1

In Algorithm 2, the algorithm `MLSMWrapper` feeds $g_{t_q}(S_{t_q})$ back to `BanditMLSM4PS` to replace the value $f_{t_q}(\boldsymbol{y}_{t_q})$. Therefore `MLSMWrapper` are actually using a new estimator for $l_q(\boldsymbol{H}_q\boldsymbol{v}_q)$, we denote the new estimator $\widetilde{l}'(\boldsymbol{H}_q\boldsymbol{v}_q)$. If $z_q \geq \frac{1}{2}$:

$$\widetilde{l}'_q(\boldsymbol{H}_q\boldsymbol{v}_q) := \begin{cases} -2(1-1/e)\dfrac{d}{z_q} \cdot g_{t_q}(S_{t_q}) & \text{if } \boldsymbol{u}_q = \mathbf{0}, \\ 2(1-1/e)\dfrac{d}{z_q} \cdot g_{t_q}(S_{t_q}) & \text{if } \boldsymbol{u}_q \neq \mathbf{0}. \end{cases}$$

else:

$$\widetilde{l}'_q(\boldsymbol{H}_q\boldsymbol{v}_q) := \begin{cases} -4(1-1/e)d\cdot\langle \boldsymbol{H}_q\boldsymbol{v}_q, \boldsymbol{u}_q\rangle g_{t_q}(S_{t_q}) & \text{if } \boldsymbol{y}_{t_q} = z_q\boldsymbol{x}_q, \\ 4(1-1/e)d\cdot\langle \boldsymbol{H}_q\boldsymbol{v}_q, \boldsymbol{u}_q\rangle g_{t_q}(S_{t_q}) & \text{if } \boldsymbol{y}_{t_q} = z_q\boldsymbol{x}_q + \frac{1}{2}\boldsymbol{u}_q. \end{cases} \tag{20}$$

We first show $\widetilde{l}(\boldsymbol{H}_q\boldsymbol{v}_q)$ is an unbiased estimator of $l(\boldsymbol{H}_q\boldsymbol{v}_q)$.

Lemma E.1. *The estimator $\widetilde{l}'_q(\boldsymbol{H}_q\boldsymbol{v}_q)$ is an unbiased estimator for $l_q(\boldsymbol{H}_q\boldsymbol{v}_q)$, that is,*

$$\mathbb{E}\left[\widetilde{l}'_q(\boldsymbol{H}_q\boldsymbol{v}_q) \mid \mathcal{H}_q, \boldsymbol{v}_q\right] = l_q(\boldsymbol{H}_q\boldsymbol{v}_q)$$

Proof. Condition on $\mathcal{H}_{q-1}, \boldsymbol{v}_q, t_q, z_q, \boldsymbol{u}_q$. If $z_q \geq \frac{1}{2}$ and $\boldsymbol{u}_q = 0$,

$$\begin{aligned}\mathbb{E}\left[\widetilde{l}'_q(\boldsymbol{H}_q\boldsymbol{v}_q) \mid \mathcal{H}_{q-1}, \boldsymbol{v}_q, t_q, \boldsymbol{z}_q, \boldsymbol{u}_q\right] &= -2(1-1/e)\frac{d}{z_q}\,\mathbb{E}[g_{t_q}(S_{t_q}) \mid \mathcal{H}_{q-1}, \boldsymbol{v}_q, t_q, \boldsymbol{z}_q, \boldsymbol{u}_q] \\ &= -2(1-1/e)\frac{d}{z_q} f_{t_q}(z_q\boldsymbol{x}_q).\end{aligned}$$

The last equality is because that $S_{t_q} \sim \text{EXT}(z_q\boldsymbol{x}_q)$ and $f_{t_q}(\boldsymbol{x}) = \mathbb{E}_{S\sim\text{EXT}(\boldsymbol{x})}[g_{t_q}(S)]$. If $z_q \geq \frac{1}{2}$ and $\boldsymbol{u}_q \neq 0$,

$$\begin{aligned}\mathbb{E}\left[\widetilde{l}'_q(\boldsymbol{H}_q\boldsymbol{v}_q) \mid \mathcal{H}_{q-1}, \boldsymbol{v}_q, t_q, \boldsymbol{z}_q, \boldsymbol{u}_q\right] &= 2(1-1/e)\frac{d}{z_q}\,\mathbb{E}[g_{t_q}(S_{t_q}) \mid \mathcal{H}_{q-1}, \boldsymbol{v}_q, t_q, \boldsymbol{z}_q, \boldsymbol{u}_q] \\ &= 2(1-1/e)\frac{d}{z_q} f_{t_q}(z_q\boldsymbol{x}_q + z_q\langle \boldsymbol{H}_q\boldsymbol{v}_q, \boldsymbol{u}_q\rangle\boldsymbol{u}_q).\end{aligned}$$

Condition on $\mathcal{H}_{q-1}, \boldsymbol{v}_q, t_q, z_q$, then

$$\begin{aligned}\mathbb{E}\left[\widetilde{l}'_q(\boldsymbol{H}_q\boldsymbol{v}_q) \mid \mathcal{H}_{q-1}, \boldsymbol{v}_q, t_q, \boldsymbol{z}_q\right] &= \frac{1}{2}(-2(1-1/e)\frac{d}{z_q} f_{t_q}(z_q\boldsymbol{x}_q)) + \sum_{i=1}^{d}\frac{1}{d}2(1-1/e)\frac{d}{z_q} f_{t_q}(z_q\boldsymbol{x}_q + z_q\langle \boldsymbol{H}_q\boldsymbol{v}_q, \boldsymbol{u}_q\rangle\boldsymbol{u}_q) \\ &= (1-1/e)\langle \boldsymbol{H}_q\boldsymbol{v}_q, \nabla f_{t_q}(z_q\cdot\boldsymbol{x}_q)\rangle\end{aligned}$$

where the last equality is already proved in the proof of Lemma 3.1.

If $z_q < \frac{1}{2}$,

$$\begin{aligned}&\mathbb{E}\left[\widetilde{l}'_q(\boldsymbol{H}_q\boldsymbol{v}_q) \mid \mathcal{H}_{q-1}, \boldsymbol{v}_q, z_q, t_q, \boldsymbol{u}_q, \boldsymbol{y}_{t_q}\right] \\ &= \begin{cases} -4(1-1/e)d\cdot\langle \boldsymbol{H}_q\boldsymbol{v}_q, \boldsymbol{u}_q\rangle\,\mathbb{E}[g_{t_q}(S_{t_q}) \mid \mathcal{H}_{q-1}, \boldsymbol{v}_q, z_q, t_q, \boldsymbol{u}_q, \boldsymbol{y}_{t_q}] & \text{if } \boldsymbol{y}_{t_q} = z_q\boldsymbol{x}_q, \\ 4(1-1/e)d\cdot\langle \boldsymbol{H}_q\boldsymbol{v}_q, \boldsymbol{u}_q\rangle\,\mathbb{E}[g_{t_q}(S_{t_q}) \mid \mathcal{H}_{q-1}, \boldsymbol{v}_q, z_q, t_q, \boldsymbol{u}_q, \boldsymbol{y}_{t_q}] & \text{if } \boldsymbol{y}_{t_q} = z_q\boldsymbol{x}_q + \frac{1}{2}\boldsymbol{u}_q. \end{cases} \\ &= \begin{cases} -4(1-1/e)d\cdot\langle \boldsymbol{H}_q\boldsymbol{v}_q, \boldsymbol{u}_q\rangle f_{t_q}(\boldsymbol{y}_{t_q}) & \text{if } \boldsymbol{y}_{t_q} = z_q\boldsymbol{x}_q, \\ 4(1-1/e)d\cdot\langle \boldsymbol{H}_q\boldsymbol{v}_q, \boldsymbol{u}_q\rangle f_{t_q}(\boldsymbol{y}_{t_q}) & \text{if } \boldsymbol{y}_{t_q} = z_q\boldsymbol{x}_q + \frac{1}{2}\boldsymbol{u}_q. \end{cases}\end{aligned}$$

Condition on $\mathcal{H}_{q-1}, \boldsymbol{v}_q, t_q, z_q$,

$$\begin{aligned}&\mathbb{E}\left[\widetilde{l}'_q(\boldsymbol{H}_q\boldsymbol{v}_q) \mid \mathcal{H}_{q-1}, \boldsymbol{v}_q, t_q, \boldsymbol{z}_q\right] \\ &= \sum_{i=1}^{d}\frac{1}{d}\left(\frac{1}{2}\cdot(-4(1-1/e)d\cdot\langle \boldsymbol{H}_q\boldsymbol{v}_q, \boldsymbol{e}_i\rangle f(z_q\boldsymbol{x}_q)) + \frac{1}{2}\cdot(4(1-1/e)d\cdot\langle \boldsymbol{H}_q\boldsymbol{v}_q, \boldsymbol{e}_i\rangle f(z_q\boldsymbol{x}_q + \frac{1}{2}\boldsymbol{e}_i))\right) \\ &= \sum_{i=1}^{d} 2(1-1/e)\langle \boldsymbol{H}_q\boldsymbol{v}_q, \boldsymbol{e}_i\rangle\left(f(z_q\boldsymbol{x}_q + \frac{1}{2}\boldsymbol{e}_i) - f(z_q\boldsymbol{x}_q)\right)\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{d} 2(1-1/e)\langle \boldsymbol{H}_q \boldsymbol{v}_q, \boldsymbol{e}_i\rangle \frac{1}{2}\frac{\partial f}{\partial x_i}(z_q \boldsymbol{x}_q) \\
&= (1-1/e)\sum_{i=1}^{d}\langle \boldsymbol{H}_q \boldsymbol{v}_q, \boldsymbol{e}_i\rangle\langle \boldsymbol{e}_i, \nabla f(z_q \boldsymbol{x}_q)\rangle \\
&= (1-1/e)\langle \boldsymbol{H}_q \boldsymbol{v}_q, \nabla f_{t_q}(z_q \cdot \boldsymbol{x}_q)\rangle
\end{aligned}
$$

Combining with the case $z_q \geq \frac{1}{2}$, we proved this equation whatever z_q is:

$$
\mathbb{E}\left[\widetilde{l}'_q(\boldsymbol{H}_q \boldsymbol{v}_q) \mid \mathcal{H}_{q-1}, \boldsymbol{v}_q, t_q, \boldsymbol{z}_q\right] = (1-1/e)\langle \boldsymbol{H}_q \boldsymbol{v}_q, \nabla f_{t_q}(z_q \cdot \boldsymbol{x}_q)\rangle
$$

Then follow the calculation in Lemma 3.1, we can prove

$$
\mathbb{E}[\widetilde{l}'_q(\boldsymbol{H}_q \boldsymbol{v_q}) \mid \mathcal{H}_{q-1}, \boldsymbol{v}_q] = l_q(\boldsymbol{H}_q \boldsymbol{v}_q).
$$

□

Lemma 5.1. *For a finite set $\mathcal{S}$, and a function family $\mathcal{G} \subseteq \mathcal{S}^{\mathbb{R}_+}$, where $\mathcal{S}^{\mathbb{R}_+}$ is the set of all functions that map element in $\mathcal{S}$ to $\mathbb{R}^+$. If there is an extension mapping EXT : $\mathcal{K} \to \Delta(\mathcal{S})$ satisfying following conditions:*

1. *$\mathcal{K} \subseteq \mathbb{R}^d$ is a product of standard simplexes.*
2. *For any $g \in \mathcal{G}$, $f(\boldsymbol{x}) = \mathbb{E}_{S \in EXT(\boldsymbol{x})}[g(S)]$ is a multi-linear, monotone, DR-submodular function, and f is L_1-lipschitz continuous, $f(\mathbf{0}) = 0$.*
3. *For any $s \in \mathcal{S}$. Here exist $\boldsymbol{x} \in \mathcal{K}$ such that EXT$(\boldsymbol{x}) = \mathbf{1}_s$. Where $\mathbf{1}_S$ assign probability 1 to S and 0 to other elements of $\mathcal{S}$.*

then the algorithm `MLSMWrapper` *attains expected $(1-1/e)$-regret $\mathcal{R}_{1-1/e}(T) \leq O\left(d^{5/3}T^{2/3}\log(T)\right)$ on $(\mathcal{S}, \mathcal{G})$-bandit.*

Proof of Lemma 5.1. We first note that $g_t(S_t)$ is an unbiased estimator of $f_t(\boldsymbol{y}_t)$ by the definition of f_t. The analysis is the same as the analysis of Algorithm 1 except that Algorithm 2 is actually using a new estimator $\widetilde{l}'_q(\boldsymbol{H}_q \boldsymbol{v}_q)$for $l(\boldsymbol{H}_q \boldsymbol{v}_q)$. We first bound this new estimator.

If $z_q \geq \frac{1}{2}$,

$$
\begin{aligned}
|\widetilde{l}'_q(\boldsymbol{H}_q \boldsymbol{v}_q)| &= 2(1-1/e)d\frac{g_{t_q}(S_{t_q})}{z_q} \\
&\leq 4(1-1/e)dM
\end{aligned}
$$

If $z_q < \frac{1}{2}$,

$$
\begin{aligned}
|\widetilde{l}'_q(\boldsymbol{H}_q \boldsymbol{v}_q)| &= 4(1-1/e)d\langle \boldsymbol{H}_q \boldsymbol{v}_q, \boldsymbol{u}_q\rangle g_{t_q}(S_{t_q}) \\
&\leq 4(1-1/e)dM
\end{aligned}
$$

The inequality is because that $z_q \boldsymbol{x}_q + \langle \boldsymbol{H}_q \boldsymbol{v}_q, \boldsymbol{u}_q\rangle \boldsymbol{u}_q \in \mathcal{K}$, and $\boldsymbol{u}_q$ is a basis vector. Therefore $\langle \boldsymbol{H}_q \boldsymbol{v}_q, \boldsymbol{u}_q\rangle \leq D_\infty$, here the D_∞ is the ∞-norm diameter of $\mathcal{K}$. Since $\mathcal{K}$ is a cartesian product of standard simplexes, $D_\infty = 1$.

Let $\widetilde{\nabla}\overline{F}'_q(\boldsymbol{x}_q)$ be the estimator replacing $\widetilde{l}'_q(\boldsymbol{H}_q \boldsymbol{v}_q)$ with $\widetilde{l}_q(\boldsymbol{H}_q \boldsymbol{v}_q)$, that is

$$
\widetilde{\nabla}\overline{F}'_q(\boldsymbol{x}_q) = d \cdot \widetilde{l}'_q(\boldsymbol{H}_q \boldsymbol{v}_q)\boldsymbol{H}_q^{-1}\boldsymbol{v}_q
$$

We bound the dual local norm of $\widetilde{\nabla}\overline{F}'_q(\boldsymbol{x}_q)$

$$
\mathbb{E}\left[\|\widetilde{\nabla}\overline{F}'_q(\boldsymbol{x}_q)\|_{\Phi, \boldsymbol{x}_q, *} \mid \mathcal{H}_{q-1}\right] = \mathbb{E}\left[d^2 \cdot (\widetilde{l}(\boldsymbol{H}_q \boldsymbol{v}_q))^2 \boldsymbol{v}_s^T \boldsymbol{H}_q^{-1}\Phi(\boldsymbol{x}_q)\boldsymbol{H}_q^{-1}\boldsymbol{v}_q \mid \mathcal{H}_{q-1}\right]
$$

$$\begin{aligned}
&\leq 16(1-1/e)^2 d^4 M^2 \,\mathbb{E}\left[\boldsymbol{v}_q^T \Phi(\boldsymbol{x}_q)^{-1/2} \Phi(\boldsymbol{x}_q) \Phi(\boldsymbol{x}_q)^{-1/2} \boldsymbol{v}_q \mid \mathcal{H}_{q-1}\right] \\
&\leq 16(1-1/e)^2 d^4 M^2 \|\boldsymbol{v}_q\|^2 \\
&\leq 16(1-1/e)^2 d^4 M^2
\end{aligned}$$

Since we have proved $\mathbb{E}[\widetilde{l}'_q(\boldsymbol{H}_q \boldsymbol{v}_q) \mid \mathcal{H}_{q-1}, \boldsymbol{v}_q] = l_q(\boldsymbol{H}_q \boldsymbol{v}_q)$, then follow the proof of Lemma 3.2 (i), we can prove

$$\mathbb{E}\left[\widetilde{\nabla}\overline{F}'_q(\boldsymbol{x}_q) \mid \mathcal{H}_{q-1}\right] = \nabla \overline{F}_q(\boldsymbol{x}_q)$$

Then follow the proof of Theorem 3.3, we have for any $\boldsymbol{x}^* \in \mathcal{K}$,

$$\begin{aligned}
\mathbb{E}\left[\sum_{t=1}^{T}(1-1/e) f_t(\boldsymbol{x}^*) - f_t(\boldsymbol{y}_t)\right] &\leq \eta L \sum_{q=1}^{Q} \mathbb{E}\left[\|\widetilde{\nabla}\overline{F}'_q(\boldsymbol{x}_q)\|^2_{\Phi,\boldsymbol{x}_q,*} \mid \mathcal{H}_{q-1}\right] + (1-1/e) L_1 D + MQ + \frac{\nu L \log(\frac{1}{\delta})}{\eta} \\
&\leq 16(1-1/e)^2 M^2 d^4 \eta T + (1-1/e) L_1 D + MQ + \frac{\nu L \log(T)}{\eta}
\end{aligned}$$

Let $S^* = \operatorname{argmax}_{S \in \mathcal{S}} \sum_{t=1}^{T} g_t(S)$, $\boldsymbol{x}^*$ be the point satisfies $\mathrm{EXT}(\boldsymbol{x}^*) = \mathbf{1}_{S^*}$, then $f_t(\boldsymbol{x}^*) = g_t(S^*)$.

Let $\mathcal{H}'_t$ be the history of the first t-rounds, including the realization of $\boldsymbol{v}_q, t_q, z_q, \boldsymbol{u}_q$, $\forall q \leq \lceil \frac{t}{L} \rceil$ and the realization of S_k, $\forall k \leq t$. Since $f_t(\boldsymbol{y}_t) = \mathbb{E}[g_t(S_t) \mid \mathcal{H}_{t-1}]$, we have,

$$\begin{aligned}
\mathbb{E}\left[\sum_{t=1}^{T}(1-1/e) g_t(S^*) - g_t(S_t)\right] &= \mathbb{E}\left[\sum_{t=1}^{T}(1-1/e) f_t(\boldsymbol{x}^*) - \mathbb{E}[g_t(S_t) \mid \mathcal{H}_{t-1}]\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T}(1-1/e) f_t(\boldsymbol{x}^*) - f_t(\boldsymbol{y}_t)\right] \\
&\leq 16(1-1/e)^2 M^2 d^4 \eta T + (1-1/e) L_1 + MQ + \frac{\nu L \log(T)}{\eta}
\end{aligned}$$

If we use the self-concordant barrier described in Appendix D as the input Φ here, by Lemma D.1, $\nu = O(d)$. Then we set $\eta = d^{-7/3} T^{-1/3}$, $L = d^{-5/3} T^{1/3}$, $Q = T/L = d^{5/3} T^{2/3}$. Then

$$\mathcal{R}_{1-1/e}(T) = \mathbb{E}\left[\sum_{t=1}^{T}(1-1/e) g_t(S^*) - g_t(S_t)\right] = O\left(d^{5/3} T^{2/3} \log(T)\right).$$

□

E.2. Proof of Lemma 5.3 and Corollary 5.4

Before proving Lemma 5.3, we first prove a useful lemma.

Lemma E.2. *Let $g : 2^G \longrightarrow \mathbb{R}_+$ be a monotone submodular set function, S is a subset of G, $s_1, s_2 \in G$ and there is no any other restriction on s_1 and s_2, they may be the same element or not, and they may be in S or not. Then $g(S \cup \{s_1, s_2\}) - g(S \cup \{s_1\}) \leq g(S \cup \{s_2\}) - g(S)$.*

Proof. If $s_1 \in S$ and $s_2 \notin S$, then $g(S \cup \{s_1, s_2\}) - g(S \cup \{s_1\}) = g(S \cup \{s_2\}) - g(S)$. If $s_1 \notin S$ and $s_2 \in S$, then $g(S \cup \{s_1, s_2\}) - g(S \cup \{s_1\}) = g(S \cup \{s_1\}) - g(S \cup \{s_1\}) = 0$ and $g(S \cup \{s_2\}) - g(S) = g(S) - g(S) = 0$. If $s_1 \in S, s_2 \in S$, then $g(S \cup \{s_1, s_2\}) - g(S \cup \{s_1\}) = g(S \cup \{s_2\}) - g(S) = g(S) - g(S) = 0$, the inequality holds.

If $s_1 \notin S, s_2 \notin S$ and $s_1 \neq s_2$, then the result holds due to the submodularity of g. If $s_1 = s_2 \notin S$, then $g(S \cup \{s_1, s_2\}) - g(S \cup \{s_1\}) = 0$ and $g(S \cup \{s_2\}) - g(S) \geq 0$ by the monotonicity of g. □

Lemma 5.3. *For $\mathcal{G}_{MS}$, the extension mapping $EXT_{PM} : \mathcal{K} \to \Delta(\mathcal{S}_{PM})$ satisfies the conditions in Lemma 5.1. Moreover, $\mathcal{K}$ is in a $\sum_{k=1}^{K} r_k |G_k|$ dimensional real vector space. For any $g \in \mathcal{G}_{MS}$, the continuous extension $f(\boldsymbol{x}) = \mathbb{E}_{s \in EXT(\boldsymbol{x})}[g(s)]$ is $M\sqrt{\sum_{k=1}^{K} r_k |G_k|}$-lipschitz.*

Proof of Lemma 5.3. We first prove that for any $S \in \mathcal{S}_{PM}$, there is a $\boldsymbol{x} \in \mathcal{K}$ such that $\mathrm{EXT}_{PM}(\boldsymbol{x}) = \mathbf{1}_S$. For any $S \in \mathcal{S}_{PM}$, it can be partitioned into $S = \cup_{k=1}^{K} S_k$ such that $S_k \subseteq G_k$ and $|S_k| \leq r_k$. For any k, we select $|S_k|$ standard simplexes $\Delta_{G_k}^{k,i},\ i \in [|S_k|]$, and we assign probability 1 to the elements in S_k respectively in these standard simplexes. Thus the condition 1 and 3 of Lemma 5.1 is satisfied, the dimension of $\mathcal{K}$ is obvious. It's enough to show that $f(\boldsymbol{x})$ satisfies the condition 2.

$\mathrm{EXT}_{PM}(\mathbf{0})$ assigns probability 1 to the empty set, thus $f(\mathbf{0}) = 0$.

Now we check the multi-linearity of f. Consider a sample $\omega \in \Omega$, the probability of ω is $\Pr(\omega) = \prod_{k=1}^{K} \prod_{i=1}^{r_k} \Pr(\omega_{k,i})$, and $\Pr(\omega_{k,i}) \in \{x_{k,i,s} \mid s \in G_k\} \cup \{1 - \sum_{s \in G_k} x_{k,i,s}\}$, thus $\Pr(\omega)$ is multi-linear with respect to the variables $x_{k,i,s}$. Then we write $f(\boldsymbol{x})$ as follows,

$$\begin{aligned} f(\boldsymbol{x}) &= \mathop{\mathbb{E}}_{S \sim \mathrm{EXT}_{PM}(\boldsymbol{x})}[g(S)] \\ &= \sum_{S \in \mathcal{S}_{PM}} \Pr(S) g(S) \\ &= \sum_{S \in \mathcal{S}_{PM}} \sum_{\omega \in \rho^{-1}(S)} \Pr(\omega) g(S) \\ &= \sum_{\omega \in \Omega} \Pr(\omega) g(\rho(\omega)) \end{aligned}$$

Since $g(\rho(\omega))$ is a constant independent from $\boldsymbol{x}$, $f(\boldsymbol{x})$ is a linear combination of multi-linear functions, thus $f(\boldsymbol{x})$ is also multi-linear.

Since $f(\boldsymbol{x})$ is multi-linear, its partial derivative is

$$\frac{\partial f}{\partial x_{k,i,s}}(\boldsymbol{x}) = \frac{f\left(\boldsymbol{x} \vee (1 - \sum_{s' \in G_k, s' \neq s} x_{k,i,s'}) \boldsymbol{e}_{k,i,s}\right) - f(\boldsymbol{x} \wedge \bar{\boldsymbol{e}}_{k,i,s})}{1 - \sum_{s' \in G_k, s' \neq s} x_{k,i,s'}}$$

Here $\wedge$ is the coordinate-wise minimal, and $\vee$ is the coordinate-wise maximal. $\boldsymbol{e}_{k,i,s}$ is the basis vector which takes 1 only for the component indexed (k,i,s), and 0 for the other components. $\bar{\boldsymbol{e}}_{k,i,s}$ is the vector that take 0 for the component indexed (k,i,s) and 1 for the other components.

We define two mappings $\rho_{\vee}^{k,i,s}, \rho_{\wedge}^{k,i,s} : \Omega \longrightarrow \Omega$. For $\omega = (\omega_{k',i'})_{k',i' \in \Gamma}$, $\Gamma = \{(k',i') \mid 1 \leq k' \leq K, 1 \leq i' \leq r_k,\ i,k \in \mathbb{N}\}$. $\rho_{\vee}^{k,i,s}$ and $\rho_{\wedge}^{k,i,s}$ only change the component indexed (k,i), for any $(k',i') \neq (k,i)$, $(\rho_{\vee}^{k,i,s}(\omega))_{k',i'} = (\rho_{\wedge}^{k,i,s}(\omega))_{k',i'} = \omega_{k',i'}$. For the component indexed (k,i), let

$$\left(\rho_{\vee}^{k,i,s}(\omega)\right)_{k,i} = \begin{cases} s & \text{if } \omega_{k,i} = \circ \\ \omega_{k,i} & \text{if } \omega_{k,i} \neq \circ \end{cases} \tag{21}$$

and

$$\left(\rho_{\wedge}^{k,i,s}(\omega)\right)_{k,i} = \begin{cases} \circ & \text{if } \omega_{k,i} = s \\ \omega_{k,i} & \text{if } \omega_{k,i} \neq s \end{cases} \tag{22}$$

We make the following important claim

Claim E.3.

$$f\left(\boldsymbol{x} \vee \left(1 - \sum_{s' \in G_k, s' \neq s} x_{k,i,s'}\right) \boldsymbol{e}_{k,i,s}\right) = \mathop{\mathbb{E}}_{\omega \sim \textit{pre-EXT}_{PM}(\boldsymbol{x})} \left[g\left(\rho\left(\rho_{\vee}^{k,i,s}(\omega)\right)\right)\right] \tag{23}$$

and

$$f(\boldsymbol{x} \wedge \bar{\boldsymbol{e}}_{k,i,s}) = \mathop{\mathbb{E}}_{\omega \sim \textit{pre-EXT}_{PM}(\boldsymbol{x})} \left[g\left(\rho\left(\rho_{\wedge}^{k,i,s}(\omega)\right)\right)\right]. \tag{24}$$

proof of Claim E.3. Let $\boldsymbol{x}^{\vee,k,i,s} := \boldsymbol{x} \vee \left(1 - \sum_{s' \in G_k, s' \neq s} x_{k,i,s'}\right) \boldsymbol{e}_{k,i,s}$.

$$f(\boldsymbol{x}^{\vee,k,i,s}) = \mathop{\mathbb{E}}_{\omega \sim \text{pre-EXT}_{PM}(\boldsymbol{x}^{\vee,k,i,s})} [g(\rho(\omega))]$$

Let $\omega \sim \text{pre-EXT}_{PM}(\boldsymbol{x})$, it's enough to show $\rho_{\vee}^{k,i,s}(\omega) \sim \text{pre-EXT}(\boldsymbol{x}^{\vee,k,i,s})$. Let $\omega_{\vee} \sim \text{pre-EXT}_{PM}(\boldsymbol{x}^{\vee,k,i,s})$. For $(k',i') \neq (k,i)$ and $s' \in G_{k'}$, $\Pr((\omega_{\vee})_{k',i'} = s') = \Pr((\rho_{\vee}^{k,i,s})_{k',i'} = s') = x_{k',i',s'}$, and $\Pr((\omega_{\vee})_{k',i'} = \circ) = \Pr((\rho_{\vee}^{k,i,s}(\omega))_{k',i'} = \circ) = 1 - \sum_{s' \in G_{k'}} x_{k',i',s'}$. For (k,i)-component and $s' \neq s$, $\Pr((\omega_{\vee})_{k,i} = s') = x_{k,i,s'} = \Pr((\rho_{\vee}^{k,i,s}(\omega))_{k,i} = s')$. For s and $\circ$,

$$\Pr((\omega_{\vee})_{k,i} = s) = 1 - \sum_{s' \in G_k, s' \neq s} x_{k,i,s'}, \qquad \Pr((\omega_{\vee})_{k,i} = \circ) = 0$$

$(\rho_{\vee}^{k,i,s}(\omega))_{k,i} = s$ whenever $\omega \in \{s, \circ\}$, thus

$$\Pr((\rho_{\vee}^{k,i,s}(\omega))_{k,i} = s) = 1 - \sum_{s' \in G_k, s' \neq s} x_{k,i,s'} = \Pr((\omega_{\vee})_{k,i} = s)$$

Since $\rho_{\vee}^{k,i,s}(\omega))_{k,i}$ never be $\circ$, $\Pr(\rho_{\vee}^{k,i,s}(\omega))_{k,i} = \circ) = 0$. Thus $\rho_{\vee}^{k,i,s}(\omega) \sim \text{pre-EXT}(\boldsymbol{x}^{\vee,k,i,s})$, and

$$f(\boldsymbol{x}^{\vee,k,i,s}) = \mathop{\mathbb{E}}_{\omega \sim \text{pre-EXT}_{PM}(\boldsymbol{x}^{\vee,k,i,s})} [g(\rho(\omega))] = \mathop{\mathbb{E}}_{\omega \sim \text{pre-EXT}_{PM}(\boldsymbol{x})} \left[g\left(\rho\left(\rho_{\vee}^{k,i,s}(\omega)\right)\right)\right].$$

In brief, whenever $\omega_{k,i} \in \{s, \circ\}$, $\rho_{\vee}^{k,i,s}(\omega) = s$. That is, $\Pr((\rho_{\vee}^{k,i,s}(\omega))_{k,i} = s) = \Pr(\omega_{k,i} \in \{s, \circ\}) = 1 - \Pr(\omega_{k,i} \notin \{s, \circ\}) = 1 - \sum_{s' \neq s, s' \in G_k} x_{k,i,s'}$ which is the same as a sample in $\text{pre-EXT}_{PM}(\boldsymbol{x}^{\vee,k,i,s})$.

For (24), we can define $\boldsymbol{x}^{\wedge,k,i,s} := \boldsymbol{x}\bar{\boldsymbol{e}}_{k,i,s}$ and let $\omega_{\wedge} \sim \text{pre-EXT}_{PM}(\boldsymbol{x}^{\wedge,k,i,s})$. One can check that for $(k',i',s') \neq (k,i,s)$, $\Pr((\rho_{\wedge}^{k,i,s}(\omega))_{k',i'} = s') = \Pr((\omega_{\wedge})_{k',i'} = s') = x_{k',i',s'}$ and $\Pr((\rho_{\wedge}^{k,i,s}(\omega))_{k',i'} = \circ) = \Pr((\omega_{\wedge})_{k',i'} = \circ) = 1 - \sum_{s' \in G_{k'}} x_{k',i',s'}$. For (k,i,s),

$$\Pr((\rho_{\wedge}^{k,i,s}(\omega))_{k,i} = \circ) = \Pr((\omega_{\wedge})_{k,i} = \circ) = 1 - \sum_{s' \in G_k, s' \neq s} x_{k,i,s'}$$

and

$$\Pr((\rho_{\wedge}^{k,i,s}(\omega))_{k,i} = s) = \Pr((\rho_{\wedge}^{k,i,s}(\omega))_{k,i} = s) = 0.$$

So $\rho_{\wedge}^{k,i,s}(\omega) \sim \text{pre-EXT}(\boldsymbol{x}^{\wedge,k,i,s})$ and (24) holds. □

If $\omega_{k,i} = \circ$ or $\omega_{k,i} = s$, one can check that $\left(\rho_{\vee}^{k,i,s}(\omega)\right)_{k,i} = s$ and $\left(\rho_{\wedge}^{k,i,s}(\omega)\right)_{k,i} = \omega_{k,i} = \circ$, therefore $\rho(\rho_{\vee}^{k,i,s}(\omega)) = \rho(\rho_{\wedge}^{k,i,s}(\omega)) \cup \{s\}$, so $\rho\left(\rho_{\wedge}^{k,i,s}(\omega)\right) \subseteq \rho\left(\rho_{\vee}^{k,i,s}(\omega)\right)$. if $\omega_{k,i} \notin \{\circ, s\}$, then $\rho_{\vee}^{k,i,s}(\omega) = \rho_{\wedge}^{k,i,s}(\omega)$, so $\rho\left(\rho_{\wedge}^{k,i,s}(\omega)\right) = \rho\left(\rho_{\vee}^{k,i,s}(\omega)\right)$. Thus we proved $\rho\left(\rho_{\wedge}^{k,i,s}(\omega)\right) \subseteq \rho\left(\rho_{\vee}^{k,i,s}(\omega)\right)$ for any ω. Since g is monotone, we have

$$\frac{\partial f}{\partial x_{k,i,s}}(\boldsymbol{x}) = \frac{\mathop{\mathbb{E}}_{\omega \sim \text{pre-EXT}_{PM}(\boldsymbol{x})} \left[g\left(\rho\left(\rho_{\vee}^{k,i,s}(\omega)\right)\right) - g\left(\rho\left(\rho_{\wedge}^{k,i,s}(\omega)\right)\right)\right]}{1 - \sum_{s' \in G_k, s' \neq s} x_{k,i,s'}} \geq 0$$

Note that $\rho_{\wedge}^{k,i,s}(\omega) \neq \rho_{\vee}^{k,i,s}(\omega)$ only happens when $\omega_{k,i} = s$ or $\omega_{k,i} = \circ$, thus,

$$\begin{aligned}\left|\frac{\partial f}{\partial x_{k,i,s}}(\boldsymbol{x})\right| &\leq \frac{M \Pr(\omega_{k,i} \in \{s, \circ\})}{1 - \sum_{s' \in G_k, s' \neq s} x_{k,i,s'}} \\ &= \frac{M(1 - \sum_{s' \in G_k, s' \neq s} x_{k,i,s'})}{1 - \sum_{s' \in G_k, s' \neq s} x_{k,i,s'}} = M\end{aligned}$$

which shows that $\|\nabla f\|_\infty \leq M$. Since $\nabla f \in \mathbb{R}^{\sum_{k=1}^K r_k|G_k|}$, $\|\nabla f\|_2 \leq \sqrt{\sum_{k=1}^K r_k|G_k|}\|\nabla f\|_\infty = M\sqrt{\sum_{k=1}^K r_k|G_k|}$. Therefore, f is $M\sqrt{\sum_{k=1}^K r_k|G_k|}$-lipschitz continuous.

Since the partial derivative of a multi-linear function is also multi-linear, $\frac{\partial f}{\partial x_{k,i,s}}(\boldsymbol{x})$ is multi-linear for any k, i, s. Then the second derivative of f can be writen as

$$\begin{aligned}
\frac{\partial^2 f}{\partial x_{k_1,i_1,s_1}\partial x_{k_2,i_2,s_2}}(\boldsymbol{x}) &= \frac{\frac{\partial f}{\partial x_{k_2,i_2,s_2}}\left(\boldsymbol{x} \vee (1-\sum_{s'\in G_{k_1},s'\neq s_1} x_{k,i,s'})\boldsymbol{e}_{k_1,i_1,s_1}\right) - \frac{\partial f}{\partial x_{k_2,i_2,s_2}}\left(\boldsymbol{x} \wedge \bar{\boldsymbol{e}}_{k_1,i_1,s_1}\right)}{1-\sum_{s'\in G_{k_1},s'\neq s_1} x_{k_1,i_1,s'}} \\
&= \frac{f\left(\boldsymbol{x} \vee (1-\sum_{s'\in G_{k_1},s'\neq s_1} x_{k',i',s'})\boldsymbol{e}_{k_1,i_1,s_1} \vee (1-\sum_{s'\in G_{k_1},s'\neq s_2} x_{k',i',s'})\boldsymbol{e}_{k_2,i_2,s_2}\right)}{(1-\sum_{s'\in G_{k_2},s'\neq s_1} x_{k_1,i_1,s'})(1-\sum_{s'\in G_{k_2},s'\neq s_2} x_{k_2,i_2,s'})} \\
&\quad - \frac{f\left((\boldsymbol{x} \wedge \bar{\boldsymbol{e}}_{k_2,i_2,s_2}) \vee (1-\sum_{s'\in G_{k_1},s'\neq s_2} x_{k',i',s'})\boldsymbol{e}_{k_1,i_1,s_1}\right)}{(1-\sum_{s'\in G_{k_1},s'\neq s_1} x_{k_1,i_1,s'})(1-\sum_{s'\in G_{k_2},s'\neq s_2} x_{k_2,i_2,s'})} \\
&\quad - \frac{f\left((\boldsymbol{x} \vee (1-\sum_{s'\in G_{k_2},s'\neq s_2} x_{k',i',s'})\boldsymbol{e}_{k_2,i_2,s_2}) \wedge \bar{\boldsymbol{e}}_{k_1,i_1,s_1}\right)}{(1-\sum_{s'\in G_{k_2},s'\neq s_1} x_{k_1,i_1,s'})(1-\sum_{s'\in G_{k_2},s'\neq s_2} x_{k_2,i_2,s'})} \\
&\quad + \frac{f\left(\boldsymbol{x} \wedge \bar{\boldsymbol{e}}_{k_2,i_2,s_2} \wedge \bar{\boldsymbol{e}}_{k_1,i_1,s_1}\right)}{(1-\sum_{s'\in G_{k_2},s'\neq s_1} x_{k_1,i_1,s'})(1-\sum_{s'\in G_{k_2},s'\neq s_2} x_{k_2,i_2,s'})}
\end{aligned}$$

To prove the DR-submodularity of f, We define 4 mappings $\rho_{\vee,\vee}^{k_1,i_1,s_1,k_2,i_2,s_2}, \rho_{\vee,\wedge}^{k_1,i_1,s_1,k_2,i_2,s_2}, \rho_{\wedge,\vee}^{k_1,i_1,s_1,k_2,i_2,s_2}, \rho_{\wedge,\wedge}^{k_1,i_1,s_1,k_2,i_2,s_2} : \Omega \longrightarrow \Omega$.

$$\begin{aligned}
\rho_{\vee,\vee}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega) &= \rho_{\vee}^{k_1,i_1,s_1}\left(\rho_{\vee}^{k_2,i_2,s_2}(\omega)\right) & \rho_{\vee,\wedge}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega) &= \rho_{\vee}^{k_1,i_1,s_1}\left(\rho_{\wedge}^{k_2,i_2,s_2}(\omega)\right) \\
\rho_{\wedge,\vee}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega) &= \rho_{\wedge}^{k_1,i_1,s_1}\left(\rho_{\vee}^{k_2,i_2,s_2}(\omega)\right) & \rho_{\wedge,\wedge}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega) &= \rho_{\wedge}^{k_1,i_1,s_1}\left(\rho_{\wedge}^{k_2,i_2,s_2}(\omega)\right)
\end{aligned}$$

Same as Claim E.3, we have

$$\begin{aligned}
&\frac{\partial f}{\partial x_{k_1,i_1,s_1}\partial x_{k_2,i_2,s_2}}(\boldsymbol{x}) \\
&= \frac{\mathop{\mathbb{E}}_{\omega\sim\text{pre-EXT}_{PM}(\boldsymbol{x})}\left[\rho\left(\rho_{\vee,\vee}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right) - \rho\left(\rho_{\vee,\wedge}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right)\right]}{(1-\sum_{s'\in G_{k_2},s'\neq s_1} x_{k_1,i_1,s'})(1-\sum_{s'\in G_{k_2},s'\neq s_2} x_{k_2,i_2,s'})} \\
&\qquad + \frac{\mathop{\mathbb{E}}_{\omega\sim\text{pre-EXT}_{PM}(\boldsymbol{x})}\left[-\rho\left(\rho_{\wedge,\vee}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right) + \left(\rho_{\wedge,\wedge}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right)\right]}{(1-\sum_{s'\in G_{k_2},s'\neq s_1} x_{k_1,i_1,s'})(1-\sum_{s'\in G_{k_2},s'\neq s_2} x_{k_2,i_2,s'})}.
\end{aligned}$$

We first consider the situation where $k_1 = k_2$ and $i_1 = i_2$. Recall that $\Pr(\omega_{k_1,i_1}) \in \{x_{k_1,i_1,s} \mid s \in G_k\} \cup \{1 - \sum_{s\in G_{k_1}} x_{k_1,i_1,s}\}$, so $\frac{\partial f}{\partial x_{k_1,i_1,s_1}\partial x_{k_2,i_2,s_2}}\Pr(\omega_{k_1,i_1}) = 0$ when $k_1 = k_2$ and $i_1 = i_2$. In this situation,

$$\begin{aligned}
\frac{\partial f}{\partial x_{k_1,i_1,s_1}\partial x_{k_2,i_2,s_2}}(\boldsymbol{x}) &= \sum_{\omega\in\Omega} g(\rho(\omega))\frac{\partial f}{\partial x_{k_1,i_1,s_1}\partial x_{k_2,i_2,s_2}}\Pr(\omega) \\
&= \sum_{\omega\in\Omega} g(\rho(\omega))\left(\prod_{\substack{k'\in[K],i'\in[r_{k'}]\\ k'\neq k_1 \text{ or } i'\neq i_1}} \Pr(\omega_{k',i'})\right)\frac{\partial f}{\partial x_{k_1,i_1,s_1}\partial x_{k_2,i_2,s_2}}\Pr(\omega_{k_1,i_1}) \qquad (25) \\
&= 0 \qquad \text{if } k_1 = k_2 \text{ and } i_1 = i_2
\end{aligned}$$

Then we consider the situation that $k_1 \neq k_2$ or $i_1 \neq i_2$,

Case 1 : $k_1 \neq k_2$ **or** $i_1 \neq i_2$, $\omega_{k_1,i_1} \in \{s_1, \circ\}$ **and** $\omega_{k_2,i_2} \in \{s_2, \circ\}$. In this case,

$$\left(\rho_{\vee,\vee}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right)_{k_1,i_1} = s_1 \qquad \left(\rho_{\vee,\vee}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right)_{k_2,i_2} = s_2$$
$$\left(\rho_{\vee,\wedge}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right)_{k_1,i_1} = s_1 \qquad \left(\rho_{\vee,\wedge}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right)_{k_2,i_2} = \circ$$
$$\left(\rho_{\wedge,\vee}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right)_{k_1,i_1} = \circ \qquad \left(\rho_{\wedge,\vee}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right)_{k_2,i_2} = s_2$$
$$\left(\rho_{\wedge,\wedge}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right)_{k_1,i_1} = \circ \qquad \left(\rho_{\wedge,\wedge}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right)_{k_2,i_2} = \circ$$

$\rho_{\vee,\vee}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega), \rho_{\vee,\vee}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega), \rho_{\vee,\vee}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega), \rho_{\vee,\vee}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)$ are equal in all but above two components k_1, i_2 and k_2, i_2. Thus

$$\rho\left(\rho_{\vee,\vee}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right) = \rho\left(\rho_{\wedge,\wedge}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right) \cup \{s_1, s_2\}$$
$$\rho\left(\rho_{\vee,\wedge}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right) = \rho\left(\rho_{\wedge,\wedge}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right) \cup \{s_1\}$$
$$\rho\left(\rho_{\wedge,\vee}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right) = \rho\left(\rho_{\wedge,\wedge}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right) \cup \{s_2\}$$

For monotone submodular g, by Lemma E.2,

$$g\left(\rho\left(\rho_{\wedge,\wedge}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right) \cup \{s_1, s_2\}\right) - g\left(\rho\left(\rho_{\wedge,\wedge}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right) \cup \{s_1\}\right)$$
$$\geq g\left(\rho\left(\rho_{\wedge,\wedge}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right) \cup \{s_2\}\right) - g\left(\rho\left(\rho_{\wedge,\wedge}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right)\right)$$

thus,

$$\rho\left(\rho_{\vee,\vee}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right) - \rho\left(\rho_{\vee,\wedge}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right) - \rho\left(\rho_{\wedge,\vee}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right) + \left(\rho_{\wedge,\wedge}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right) \leq 0$$

Case 2 : $k_1 \neq k_2$ **or** $i_1 \neq i_2$, $\omega_{k_1,i_1} \in \{s_1, \circ\}$ **and** $\omega_{k_2,i_2} \notin \{s_2, \circ\}$. In this case,

$$\rho_{\vee}^{k_2,i_2,s_2}(\omega) = \rho_{\wedge}^{k_2,i_2,s_2}(\omega) = \omega$$

thus,

$$\rho_{\vee,\vee}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega) = \rho_{\vee}^{k_1,i_1,s_1}(\omega) \qquad \rho_{\vee,\wedge}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega) = \rho_{\vee}^{k_1,i_1,s_1}(\omega)$$
$$\rho_{\wedge,\vee}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega) = \rho_{\wedge}^{k_1,i_1,s_1}(\omega) \qquad \rho_{\wedge,\wedge}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega) = \rho_{\wedge}^{k_1,i_1,s_1}(\omega)$$

and

$$\rho\left(\rho_{\vee,\vee}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right) - \rho\left(\rho_{\vee,\wedge}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right) - \rho\left(\rho_{\wedge,\vee}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right) + \left(\rho_{\wedge,\wedge}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right) = 0$$

Case 3 : $k_1 \neq k_2$ **or** $i_1 \neq i_2$, $\omega_{k_1,i_1} \notin \{s_1, \circ\}$ **and** $\omega_{k_2,i_2} \in \{s_2, \circ\}$. Since $(k_1, i_1) \neq (k_2, i_2)$, $\rho_{\vee}^{k_2,i_2,s_2}$ and $\rho_{\wedge}^{k_2,i_2,s_2}$ do not change the k_1, i_1 component of ω, that is,

$$\left(\rho_{\vee}^{k_2,i_2,s_2}(\omega)\right)_{k_1,i_1} = \left(\rho_{\wedge}^{k_2,i_2,s_2}(\omega)\right)_{k_1,i_1} = \omega_{k_1,i_1}$$

Therefore,

$$\rho_{\vee,\vee}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega) = \rho_{\vee}^{k_2,i_2,s_2}(\omega) \qquad \rho_{\vee,\wedge}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega) = \rho_{\wedge}^{k_2,i_2,s_2}(\omega)$$
$$\rho_{\wedge,\vee}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega) = \rho_{\vee}^{k_2,i_2,s_2}(\omega) \qquad \rho_{\wedge,\wedge}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega) = \rho_{\wedge}^{k_2,i_2,s_2}(\omega)$$

and

$$\rho\left(\rho_{\vee,\vee}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right) - \rho\left(\rho_{\vee,\wedge}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right) - \rho\left(\rho_{\wedge,\vee}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right) + \left(\rho_{\wedge,\wedge}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right) = 0$$

Case 4 : $k_1 \neq k_2$ **or** $i_1 \neq i_2$, $\omega_{k_1,i_1} \notin \{s_1, \circ\}$ **and** $\omega_{k_2,i_2} \notin \{s_2, \circ\}$. In this case,

$$\rho_{\vee,\vee}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega) = \rho_{\vee,\wedge}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega) = \rho_{\wedge,\vee}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega) = \rho_{\wedge,\wedge}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega) = \omega$$

thus

$$\rho\left(\rho_{\vee,\vee}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right) - \rho\left(\rho_{\vee,\wedge}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right) - \rho\left(\rho_{\wedge,\vee}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right) + \left(\rho_{\wedge,\wedge}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right) = 0.$$

In all 4 cases above, whatever ω is,

$$\rho\left(\rho_{\vee,\vee}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right) - \rho\left(\rho_{\vee,\wedge}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right) - \rho\left(\rho_{\wedge,\vee}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right) + \left(\rho_{\wedge,\wedge}^{k_1,i_1,s_1,k_2,i_2,s_2}(\omega)\right) \leq 0$$

holds. Thus,

$$\frac{\partial f}{\partial x_{k_1,i_1,s_1}\partial x_{k_2,i_2,s_2}}(\boldsymbol{x}) \leq 0 \qquad \text{if } k_1 \neq k_2 \text{ or } i_1 \neq i_2. \tag{26}$$

Combining (25) and (26), $\frac{\partial f}{\partial x_{k_1,i_1,s_1}\partial x_{k_2,i_2,s_2}}(\boldsymbol{x}) \leq 0$ for any $k_1, i_1, s_1, k_2, i_2, s_2$, which shows the DR-submodularity of f.

□

Corollary 5.4. *There is an algorithm attaining the expected* $(1-1/e)$*-regret of*

$$\mathcal{R}_{1-1/e}(T) \leq O\left(\left(\sum_{k=1}^{K} r_k|G_k|\right)^{5/3} T^{2/3}\log T\right)$$

on any $(\mathcal{S}_{PM}, \mathcal{G}_{MS})$*-bandit.*

Proof of Corollary 5.4. Since EXT_{PM} satisfies the conditions in Lemma 5.1 and the dimension of $\mathcal{K}$ is $d = \sum_{k=1}^{K} r_k|G_k|$. This is a direct corollary of Lemma 5.1. □

E.3. Proof of Lemma 5.5 and Corollary 5.6

Lemma 5.5. *For* $\mathcal{G}_{SS}$*, the extension mapping* $EXT_{SS} : \mathcal{K} \to \Delta(\mathcal{S}_{OL})$ *satisfies the conditions in Lemma 5.1. Moreover,* $\mathcal{K}$ *is in a* $|G|^2 - |G|$ *dimensional real vector space. For any* $g \in \mathcal{G}_{SS}$*, the continuous extension* $f(\boldsymbol{x}) = \mathbb{E}_{s\in EXT(\boldsymbol{x})}[g(s)]$ *is* $M|G|$*-lipschitz.*

Proof of Lemma 5.5. The condition 1 and 3 of Lemma 5.1 are obviously satisfied. Now we check the multi-linearity of f. For $\boldsymbol{x} \in \mathcal{K}$, we write $\boldsymbol{x} = (x_{i,s})_{i\in|G|,s\in G'}$. Consider $S \in \mathcal{S}_{OL}$, given $\boldsymbol{x}$, the probability of S in distribution $\mathrm{EXT}_{SS}(\boldsymbol{x})$ is $\Pr(S) = \prod_{i=1}^{|G|}\Pr(S_i)$. For any $1 \leq i \leq |G|$, $\Pr(S_i) \in \{x_{i,s} \mid s \in G'\} \cup \{1 - \sum_{s\in G'} x_{i,s}\}$, thus $\Pr(S)$ is multi-linear with respect to the variables $x_{i,s}$. Then we write $f(\boldsymbol{x})$ as follows,

$$\begin{aligned} f(\boldsymbol{x}) &= \mathop{\mathbb{E}}_{S\sim\mathrm{EXT}_{SS}(\boldsymbol{x})}[g(S)] \\ &= \sum_{S\in\mathcal{S}_{OL}} \Pr(S)g(S) \end{aligned}$$

Since $g(S)$ is a constant independent from $\boldsymbol{x}$, $f(\boldsymbol{x})$ is a linear combination of multi-linear functions, thus $f(\boldsymbol{x})$ is also multi-linear.

$\mathrm{EXT}(\mathbf{0})$ assigns probability 1 to the ordered list $\{\circ\}^{|G|}$, thus $f(\mathbf{0}) = g(\{\circ\}^{|G|}) = 0$. Next we check the monotonicity and DR-submodularity of $f(\boldsymbol{x})$.

Define $\rho_{\vee}^{i,s}, \rho_{\wedge}^{i,s} : \mathcal{S}_{OL} \longrightarrow \mathcal{S}_{OL}$. $\rho_{\vee}^{i,s}(S) \neq S$ only when the i-th position of S is $\circ$, $\rho_{\vee}^{i,s}(S)$ change the i-th position of S to s and keep other positions unchanged. $\rho_{\wedge}^{i,s}(S) \neq S$ only when the i-th position of S is s, $\rho_{\vee}^{i,s}(S)$ change the i-th position of S to $\circ$ and keep other positions unchanged. Then,

$$\frac{\partial f}{\partial x_{i,s}}(\boldsymbol{x}) = \frac{\mathop{\mathbb{E}}_{S\sim\mathrm{EXT}_{SS}(\boldsymbol{x})}\left[g(\rho_{\vee}^{i,s}(S)) - g(\rho_{\wedge}^{i,s}(S))\right]}{1 - \sum_{s'\in G, s'\neq s} x_{i,s'}}$$

Let $S^{\leq k}$ be the set containing the first k elements in the ordered list S. Then $(\rho_{\wedge}^{i,s}(S))^{\leq k} \subseteq (\rho_{\vee}^{i,s}(S))^{\leq k}$, $\forall k$, recall g_k is monotone and $\lambda_k > 0$ for any k,

$$\frac{\partial f}{\partial x_{i,s}}(\boldsymbol{x}) = \frac{\mathop{\mathbb{E}}_{S\sim \mathrm{EXT}_{SSM}(\boldsymbol{x})}\left[\sum_{k=1}^{|G_k|}\lambda_k\left(g_k((\rho_{\vee}^{i,s}(S))^{\leq k}) - g((\rho_{\wedge}^{i,s}(S))^{\leq k})\right)\right]}{1-\sum_{s'\in G, s'\neq s} x_{i,s'}} \geq 0$$

Thus f is monotone. Since $\Pr(\rho_{\vee}^{i,s}(S) \neq \rho_{\wedge}^{i,s}(S)) \leq 1 - \sum_{s'\in G, s'\neq s} x_{i,s'}$ and $\frac{\partial f}{\partial x_{i,s}}(\boldsymbol{x}) \leq M$. Thus $\|\nabla f(\boldsymbol{x})\|_\infty \leq M$, $\|\nabla f(\boldsymbol{x})\|_2 \leq M\sqrt{|G|(|G|-1)} \leq M|G|$, $f(\boldsymbol{x})$ is $M|G|$-lipschitz.

We then define

$$\rho_{\vee,\vee}^{i_1,s_1,i_2,s_2}(S) = \rho_{\vee}^{i_1,s_1}\left(\rho_{\vee}^{i_2,s_2}(S)\right) \qquad \rho_{\vee,\wedge}^{i_1,s_1,i_2,s_2}(S) = \rho_{\vee}^{i_1,s_1}\left(\rho_{\wedge}^{i_2,s_2}(S)\right)$$
$$\rho_{\wedge,\vee}^{i_1,s_1,i_2,s_2}(S) = \rho_{\wedge}^{i_1,s_1}\left(\rho_{\vee}^{i_2,s_2}(S)\right) \qquad \rho_{\wedge,\wedge}^{i_1,s_1,i_2,s_2}(S) = \rho_{\wedge}^{i_1,s_1}\left(\rho_{\wedge}^{i_2,s_2}(S)\right)$$

Then,

$$\frac{\partial f}{\partial x_{i_1,s_1}\partial x_{i_2,s_2}}(\boldsymbol{x}) = \frac{\mathop{\mathbb{E}}_{S\sim \mathrm{EXT}_{SSM}(\boldsymbol{x})}\left[g(\rho_{\vee,\vee}^{i_1,s_1,i_2,s_2}(S)) - g(\rho_{\vee,\wedge}^{i_1,s_1,i_2,s_2}(S)) - g(\rho_{\wedge,\vee}^{i_1,s_1,i_2,s_2}(S)) + g(\rho_{\wedge,\wedge}^{i_1,s_1,i_2,s_2}(S))\right]}{(1-\sum_{s'\in G, s'\neq s_1} x_{i_1,s'})(1-\sum_{s'\in G, s'\neq s_2} x_{i_2,s'})}$$

We then prove $g(\rho_{\vee,\vee}^{i_1,s_1,i_2,s_2}(S)) - g(\rho_{\vee,\wedge}^{i_1,s_1,i_2,s_2}(S)) - g(\rho_{\wedge,\vee}^{i_1,s_1,i_2,s_2}(S)) + g(\rho_{\wedge,\wedge}^{i_1,s_1,i_2,s_2}(S)) \leq 0$ for any $S \in \mathcal{S}$. It's enough to prove $g_i((\rho_{\vee,\vee}^{i_1,s_1,i_2,s_2}(S))^{\leq i}) - g_i((\rho_{\vee,\wedge}^{i_1,s_1,i_2,s_2}(S))^{\leq i}) - g_i((\rho_{\wedge,\vee}^{i_1,s_1,i_2,s_2}(S))^{\leq i}) + g_i((\rho_{\wedge,\wedge}^{i_1,s_1,i_2,s_2}(S))^{\leq i}) \leq 0$ for any $i \in [|G|]$. Note that if $\max\{i_1,i_2\} > i$ then $g_i((\rho_{\vee,\vee}^{i_1,s_1,i_2,s_2}(S))^{\leq i}) - g_i((\rho_{\vee,\wedge}^{i_1,s_1,i_2,s_2}(S))^{\leq i}) - g_i((\rho_{\wedge,\vee}^{i_1,s_1,i_2,s_2}(S))^{\leq i}) + g_i((\rho_{\wedge,\wedge}^{i_1,s_1,i_2,s_2}(S))^{\leq i}) = 0$, so we now consider the case $\max\{i_1,i_2\} \leq i$.

Case 1 : $i_1 = i_2$. In this case, since $\Pr(S_{i_1}) \in \{x_{i_1,s} \mid s \in G'\} \cup \{1 - \sum_{s\in G'} x_{i_1,s}\}$, $\frac{\partial^2 \Pr(S_{i_1})}{\partial x_{i_1,s_1}\partial x_{i_2,s_2}} = 0$ when $i_1 = i_2$. we have

$$\begin{aligned}
\frac{\partial^2 f}{\partial x_{i_1,s_1}\partial x_{i_2,s_2}}(\boldsymbol{x}) &= \frac{\partial^2}{\partial x_{i_1,s_1}\partial x_{i_2,s_2}} \sum_{S\in\mathcal{S}_{OL}} g(S)\prod_{i=1}^{|G|}\Pr(S_i) \\
&= \sum_{S\in\mathcal{S}_{OL}} g(S)\left(\prod_{i\neq i_1}\Pr(S_i)\right)\frac{\partial^2 \Pr(S_{i_1})}{\partial x_{i_1,s_1}\partial x_{i_2,s_2}} \\
&= 0
\end{aligned}$$

Case 2 : $i_1 \neq i_2, S_{i_1} \in \{s_1, \circ\}$ **and** $S_{i_2} \in \{s_2, \circ\}$. In this case,

$$\left(\rho_{\vee,\vee}^{i_1,s_1,i_2,s_2}(S)\right)_{i_1} = s_1 \qquad \left(\rho_{\vee,\vee}^{i_1,s_1,i_2,s_2}(S)\right)_{i_2} = s_2$$
$$\left(\rho_{\vee,\wedge}^{i_1,s_1,i_2,s_2}(S)\right)_{i_1} = s_1 \qquad \left(\rho_{\vee,\wedge}^{i_1,s_1,i_2,s_2}(S)\right)_{i_2} = \circ$$
$$\left(\rho_{\wedge,\vee}^{i_1,s_1,i_2,s_2}(S)\right)_{i_1} = \circ \qquad \left(\rho_{\wedge,\vee}^{i_1,s_1,i_2,s_2}(S)\right)_{i_2} = s_2$$
$$\left(\rho_{\wedge,\wedge}^{i_1,s_1,i_2,s_2}(S)\right)_{i_1} = \circ \qquad \left(\rho_{\wedge,\wedge}^{i_1,s_1,i_2,s_2}(S)\right)_{i_2} = \circ$$

$\rho_{\vee,\vee}^{1,s_1,i_2,s_2}(S), \rho_{\vee,\vee}^{i_1,s_1,i_2,s_2}(S), \rho_{\vee,\vee}^{i_1,s_1,i_2,s_2}(S), \rho_{\vee,\vee}^{i_1,s_1,i_2,s_2}(S)$ are equal in all but above two components i_2 and i_2. Thus,

$$\left(\rho_{\vee,\vee}^{i_1,s_1,i_2,s_2}(S)\right)^{\leq i} = \left(\rho_{\wedge,\wedge}^{i_1,s_1,i_2,s_2}(S)\right)^{\leq i} \cup \{s_1,s_2\}$$
$$\left(\rho_{\vee,\wedge}^{i_1,s_1,i_2,s_2}(S)\right)^{\leq i} = \left(\rho_{\wedge,\wedge}^{i_1,s_1,i_2,s_2}(S)\right)^{\leq i} \cup \{s_1\}$$

$$\left(\rho_{\wedge,\vee}^{i_1,s_1,i_2,s_2}(S)\right)^{\leq i} = \left(\rho_{\wedge,\wedge}^{i_1,s_1,i_2,s_2}(S)\right)^{\leq i} \cup \{s_2\}$$

By Lemma E.2,

$$g_i((\rho_{\vee,\vee}^{i_1,s_1,i_2,s_2}(S))^{\leq i}) - g_i((\rho_{\vee,\wedge}^{i_1,s_1,i_2,s_2}(S))^{\leq i}) - g_i((\rho_{\wedge,\vee}^{i_1,s_1,i_2,s_2}(S))^{\leq i}) + g_i((\rho_{\wedge,\wedge}^{i_1,s_1,i_2,s_2}(S))^{\leq i}) \leq 0$$

Then $\frac{\partial^2 f}{\partial x_{i_1,s_1}\partial x_{i_2,s_2}}(\boldsymbol{x}) \leq 0$.

Case 3 : $i_1 \neq i_2, S_{i_1} \notin \{s_1, \circ\}$ **or** $S_{i_2} \notin \{s_2, \circ\}$: If $S_{i_1} \notin \{s_1, \circ\}$, then $\rho_{\vee,\vee}^{i_1,s_1,i_2,s_2}(S) = \rho_{\wedge,\vee}^{i_1,s_1,i_2,s_2}(S)$ and $\rho_{\vee,\wedge}^{i_1,s_1,i_2,s_2}(S) = \rho_{\wedge,\wedge}^{i_1,s_1,i_2,s_2}(S)$. If $S_{i_2} \notin \{s_2, \circ\}$, then $\rho_{\vee,\vee}^{i_1,s_1,i_2,s_2}(S) = \rho_{\vee,\wedge}^{i_1,s_1,i_2,s_2}(S)$ and $\rho_{\wedge,\vee}^{i_1,s_1,i_2,s_2}(S) = \rho_{\wedge,\wedge}^{i_1,s_1,i_2,s_2}(S)$. Either way, we have,

$$g_i((\rho_{\vee,\vee}^{i_1,s_1,i_2,s_2}(S))^{\leq i}) - g_i((\rho_{\vee,\wedge}^{i_1,s_1,i_2,s_2}(S))^{\leq i}) - g_i((\rho_{\wedge,\vee}^{i_1,s_1,i_2,s_2}(S))^{\leq i}) + g_i((\rho_{\wedge,\wedge}^{i_1,s_1,i_2,s_2}(S))^{\leq i}) = 0.$$

Then $\frac{\partial^2 f}{\partial x_{i_1,s_1}\partial x_{i_2,s_2}}(\boldsymbol{x}) = 0$.

In all cases, $\frac{\partial f}{\partial x_{i_1,s_1}\partial x_{i_2,s_2}}(\boldsymbol{x}) \leq 0$, thus $f(\boldsymbol{x})$ is DR-submodular. □

Corollary 5.6. *There is an algorithm for attaining the expected* $(1-1/e)$*-regret of*

$$\mathcal{R}_{1-1/e}(T) \leq O\left((|G|)^{10/3}T^{2/3}\log T\right)$$

on any $(\mathcal{S}_{OL}, \mathcal{G}_{SS})$*-bandit.*

Proof of Corollary 5.6. Since EXT_{SS} satisfies the conditions in Lemma 5.1 and the dimension of $\mathcal{K}$ is $d = |G|(|G|-1) = O(|G|^2)$. This is a direct corollary of Lemma 5.1. □

F. Remark on the Stochastic Submodular Bandit

In this section, we show how our algorithms for adversarial setting can be applied to the stochastic setting proposed by Nie et al. (2022). We take the stochastic monotone submodular bandit with cardinality constraint investigated in (Nie et al., 2022) as an example.

Stochastic submodular bandit model In the stochastic model, there is an unknown distribution $\mathcal{D}$, its support is a set of set functions, we denote the set $supp(\mathcal{D})$. Any set function $g \in supp(\mathcal{D})$ is defined on the power set of the ground set G, mapping a subset of G to a reward between $[0,1]$, that is, $g: 2^G \to [0,1]$. In t-th round, the reward function g'_t is drawn from $\mathcal{D}$ and we can only select a subset $S_t \subseteq G$ such that $|S_t| \leq k$, which is a cardinality constraint. Then we gain reward $g'_t(S_t)$. Note that the model does not need g'_t to be a monotone submodular function, but requires $g = \mathbb{E}_{g'_t\sim\mathcal{D}}[g'_t]$ to be monotone submodular. Our goal is to minimize the $(1-1/e)$-regret:

$$\mathcal{R}_{1-1/e}^{sto}(T) = (1-\frac{1}{e})T \cdot \max_{S^*\subseteq G, |S^*|\leq k} g(S^*) - \mathbb{E}\left[\sum_{t=1}^{T} g'_t(S_t)\right] = (1-\frac{1}{e})T \cdot \max_{S^*\subseteq G, |S^*|\leq k} g(S^*) - \mathbb{E}\left[\sum_{t=1}^{T} g(S_t)\right]$$

The last equality is because the randomness of S_t is independent of the randomness of g'_t.

Apply our algorithm on stochastic bandit model Since cardinality constraint is a special case of the partition matroid constraint, we use our algorithm in Section 5.2. While applying our algorithm on the stochastic model, we see $g_t = g = \mathbb{E}_{g'_t\sim\mathcal{D}}[g'_t]$ as the online function selected by the adversary, thus the online functions are monotone submodular. Note that, to obtain a regret bound w.r.t. the online reward function $\{g_t\}_{t=1}^T$, we need to query the value of g_t at some subset S_t in round t. However, since the algorithm is actually running on the stochastically realized function sequence $\{g'_t\}_{t=1}^T$, if we query the function value of S_t, the feedback is $g'_t(S_t)$ rather than $g_t(S_t) = g(S_t)$. Fortunately, this is not a big issue since $g'_t(S_t)$ is an unbiased estimate of $g_t(S_t)$. Now we go back to the proof of Lemma 5.1, and replace all the $g_{t_q}(S_{t_q}) = g(S_{t_q})$ with $g'_{t_q}(S_{t_q})$. Since $g'_{t_q}(S_{t_q}) \leq 1$, it won't affect our bound for the dual local norm of the gradient

estimator. And since the randomness of the stochastic function g'_{t_q} is independent of all the randomness introduced in our algorithm and $[E][g'_{t_q}(S_{t_q})] = g_{t_q}(S_{t_q})$, the new gradient estimator constructed by replacing $g_{t_q}(S_{t_q})$ with $g'_{t_q}(S_{t_q})$ is still an unbiased estimator. Thus, as the same as the proof of our algorithm for adversarial submodular bandit with partition matroid constraint, we have the same regret bound,

$$\mathcal{R}^{adv}_{1-1/e}(T) = \max_{S^* \subseteq G, |S^*| \leq k} \mathbb{E}\left[(1-\frac{1}{e})\sum_{t=1}^{T} g(S^*) - \sum_{t=1}^{T} g(S_t)\right] \leq O((k|G|)^{5/3}T^{2/3}\log T)$$

That is,

$$(1-\frac{1}{e})T \cdot \max_{S^* \subseteq G, |S^*| \leq k} g(S^*) - \mathbb{E}\left[\sum_{t=1}^{T} g(S_t)\right] \leq O((k|G|)^{5/3}T^{2/3}\log T)$$