An Optimal Algorithm for Stochastic Matroid Bandit Optimization

Mohammad Sadegh Talebi
Department of Automatic Control
KTH Royal Institute of Technology
Stockholm, SWEDEN
mstms@kth.se

Alexandre Proutiere
Department of Automatic Control
KTH Royal Institute of Technology
Stockholm, SWEDEN
alepro@kth.se

ABSTRACT

The selection of leaders in leader-follower multi-agent systems can be naturally formulated as a matroid optimization problem. In this paper, we investigate the online and stochastic version of such a problem, where in each iteration or round, we select a set of leaders and then observe a random realization of the corresponding reward, i.e., of the system performance. This problem is referred to as a stochastic matroid bandit, a variant of combinatorial multi-armed bandit problems where the underlying combinatorial structure is a matroid. We consider semi-bandit feedback and Bernoulli rewards, and derive a tight and problem-dependent lower bound on the regret of any consistent algorithm. We propose KL-OSM, a computationally efficient algorithm that exploits the matroid structure. We derive a finite-time upper bound of the regret of KL-OSM that improves the performance guarantees of existing algorithms. This upper bound actually matches our lower bound, i.e., KL-OSM is asymptotically optimal. Numerical experiments attest that KL-OSM outperforms state-of-the-art algorithms in practice, and the difference in some cases is significant.

Categories and Subject Descriptors

I.2.6 [**Learning**]: Parameter learning; I.2.11 [**Distributed Artificial Intelligence**]: Multiagent systems

Keywords

Multi-Armed Bandits; Online Learning; Combinatorial Optimization; Matroids; Regret Analysis

1. INTRODUCTION

This work is motivated by the design of leader-follower multi-agent systems where a set of leaders act as external control inputs and have the ability to impact the dynamics of the entire system and in turn its overall performance. Of course, the choice of the set of leaders in these systems critically influences their behaviour. It has been recently shown [9,24] that the leader selection problem could be naturally formulated as a matroid optimization problem. In this paper, we investigate the online and stochastic version of such a problem, where in each iteration or round, we select a set of leaders and then observe a random realization of the corresponding reward, i.e., of the system performance. This problem, referred to as a stochastic matroid Multi-Armed Bandit (MAB) problem, can be seen as a particular instance of combinatorial MAB problems, and is particularly relevant when one wishes to learn as quickly as possible the optimal set of leaders, e.g., in scenarios where this optimal set could evolve over time.

MAB problems [15, 27] constitute the most fundamental model for sequential decision making problems with an exploration vs. exploitation trade-off and have found applications in many fields, including sequential clinical trials, communication systems, economics; see e.g. [6]. In such problems, the decision maker repeatedly selects an arm and observes a realization of the corresponding unknown reward distribution, where each decision is made based on past decisions and observed rewards. The objective is to maximize the expected cumulative reward over some time horizon by balancing exploitation and exploration. Equivalently, the performance of a decision rule or algorithm can be measured through its expected regret, defined as the gap between the expected reward achieved by the algorithm and that achieved by an oracle algorithm always selecting the best arm.

We consider matroid bandits in the stochastic setting as introduced in [19], which are actually a sub-class of combinatorial MAB problems with linear reward functions, defined in, e.g., [8, 10, 13], in which the underlying combinatorial structure is a matroid. Given a set of basic actions E (called ground set), a matroid is a pair $(E, \mathcal{I})$ with some $\mathcal{I} \subset 2^E$ such that $\mathcal{I}$ is an independence system (i.e., it is closed under subset operation) and satisfies the so-called augmentation property (see Definition 1 for a precise definition). Matroid bandits consider weighted matroids, where each element of E is assigned a weight (its average reward). Each arm is then a basis (i.e., an inclusion-wise maximal element of $\mathcal{I}$) of the matroid. The weight of various basic actions are fixed and a priori unknown. The decision maker aims at learning the maximum weight basis by sequentially selecting various arms. Hence, at each round she faces a linear optimization problem under a matroid constraint.

Matroid bandits can be applied beyond leader-follower multi-agent systems. Indeed, matroid structures occur naturally in many problems with practical applications ranging from bidding in ad exchange [29], product search [1], task assignment in crowdsourcing [5], and many other engineering applications. Hence, matroid constraints are quite nat-

Appears in: *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2016), J. Thangarajah, K. Tuyls, C. Jonker, S. Marsella (eds.), May 9–13, 2016, Singapore.*

ural for combinatorial problems. For example, assume that the elements of ground set E are categorized into L disjoint categories. A natural requirement for some applications is to force to choose at most one element from each category. In the context of product search, each category might be a specific brand, whereas for news aggregation, each category may correspond to a news domain. Assume that the decision maker is interested in finding a subset $M \subset E$ while maximizing the total reward and such that at most one element from each category belongs to M. Then, she faces a linear optimization subject to a partition matroid constraint. Another natural type of constraints is to have cardinality constraint on the set M, which is related to the notion of uniform matroid. Another notable instance of matroid constraints appears in the problem of finding the minimum spanning tree in a graph, which arises in various engineering disciplines.

Matroid optimization problems are of special interests in the area of combinatorial optimization both theoretically and practically, due to relative tractability of optimization over matroids. In particular, linear optimization over matroid bases is proven to be greedily solvable. More precisely, a well-known result in combinatorial optimization states that an independence system (see later for a formal definition) is a matroid if and only if the *greedy algorithm* leads to a maximum weight basis; see, e.g., [11]. More general cases have been addressed in, e.g., [4, 25]. Matroid theory brings a two-fold advantage in the corresponding bandit optimization problems: firstly, it is possible to devise computationally efficient algorithms that, in most cases, select arms greedily. Secondly, the corresponding regret analysis is usually more tractable. Despite such advantage, lack of optimal algorithms for matroid bandits in the literature is evident. Here we provide a sequential arm selection algorithm, KL-OSM, and show that it is asymptotically optimal. To the best of our knowledge, KL-OSM constitutes the first optimal algorithm for the online matroid problem considered.

Contributions

(a) We derive an asymptotic (as the time horizon T grows large) lower bound on the regret, satisfied by any algorithm (Theorem 2). This lower bound is tight and problem-dependent and its derivation leverages the theory of optimal control of Markov chains with unknown transition probabilities. To our knowledge, our proposed lower bound constitutes the first fundamental performance limit for matroid bandits.

(b) We propose KL-OSM (KL-based Efficient Sampling for Matroids), which is an index policy that maintains a KL-UCB index [14] for each basic action and is based on the greedy algorithm. Hence, it is provably computationally efficient assuming access to an independence oracle (see Section 3 for a precise definition). Through a finite-time analysis (Theorem 1), we show that KL-OSM attains a regret (asymptotically) growing as the proposed lower bound in Theorem 2. Hence, it is asymptotically optimal. To our best knowledge, this is the first optimal algorithm for this class of combinatorial MABs. Numerical experiments for some specific matroid problems show that KL-OSM significantly outperforms existing algorithms.

The rest of the paper is organized as follows. Section 2 provides an overview of combinatorial MAB problems. Section 3 is devoted to description of our model and a precise statement of the problem. In Section 4, we describe KL-OSM, our proposed algorithm for matroid bandits, and provide a finite-time analysis of its regret. In Section 5, we present a lower bound on the regret of our problem. We present some numerical experiments in Section 6. Finally, Section 7 concludes the paper and provides some future work directions. All proofs are provided in the appendix.

2. RELATED WORK

Combinatorial MAB problems have been an active area of research in recent years. These problems have been extensively studied in the adversarial setting; see, e.g., [3, 7] and references therein. In the stochastic setting, some research contributions investigate generic combinatorial problems, e.g., [8,10,13,20], whereas others mostly concern problems where the set of arms exhibits specific structures, such as fixed-size subsets [2,17], matroid and polymatroid [19,21], or permutations [12, 23]. The proposed algorithms in these works are variants of UCB or KL-UCB algorithms.

Matroid bandits were introduced and studied in [18, 19]. The proposed algorithm, called OMM, is a UCB-type policy relying on the greedy method. Our proposed algorithm is quite similar to OMM, but uses the KL-UCB index instead.

For a generic combinatorial structure and the stochastic setting considered in this paper, the state-of-the-art algorithm is ESCB [10], which achieves a regret upper-bounded by $\mathcal{O}(\frac{d\sqrt{m}}{\Delta_{\min}}\log(T))$ after T rounds. Here, d and m respectively denote the number of basic actions and maximum cardinality of arms, and $\Delta_{\min}$ denotes the smallest gap between the average rewards of the best arm and of a sub-optimal arm. For matroid bandits, OMM achieves a regret scaling at most as $\mathcal{O}(\frac{d-m}{\Delta_{\min}}\log(T))$. The dependence of this bound on (d, m) is tight and cannot be improved. The regret upper bound of our proposed algorithm, KL-OSM, admits the same scaling $\mathcal{O}(\frac{d-m}{\Delta_{\min}}\log(T))$. However, we are able to show that the constant in $\mathcal{O}(.)$ in the case of KL-OSM is strictly smaller. Moreover, we prove that under KL-OSM, the upper bound on the regret cannot be improved.

3. MODEL AND OBJECTIVES

3.1 Matroid Structure

We give a formal definition of matroids and state some useful related results. More details can be found in, e.g., [26, 28].

DEFINITION 1. *Let E be a finite set and $\mathcal{I} \subset 2^E$. The pair $G = (E, \mathcal{I})$ is called a* matroid *if the following conditions hold:*

(i) $\emptyset \in \mathcal{I}$,

(ii) if $X \in \mathcal{I}$ *and* $Y \subseteq X$, *then* $Y \in \mathcal{I}$,

(iii) if $X, Y \in \mathcal{I}$ *with* $|X| > |Y|$, *then there is some element* $\ell \in X \setminus Y$ *such that* $Y \cup \{\ell\} \in \mathcal{I}$.

Any system satisfying conditions (i) and (ii) is called an independence system. Condition (iii) is referred to as the augmentation property. The set E is usually referred to as the *ground set* and the elements of $\mathcal{I}$ are called the *independent sets*. Any (inclusion-wise) maximal independent set is

called a *basis* for matroid G. In other words, if $X \in \mathcal{I}$ is a basis for G, then $X \cup \{\ell\} \notin \mathcal{I}$ for all $\ell \in E \setminus X$.

PROPOSITION 1 ([26]). *Let $G = (E, \mathcal{I})$ be a matroid. Then the following hold:*

(i) All bases of G have the same cardinality, referred to as rank *of G.*

(ii) For all bases X, Y of G, if $\ell \in X \setminus Y$ then there exists $k \in Y \setminus X$ such that $(X \setminus \ell) \cup \{k\}$ is a basis for G. [1]

(iii) For all bases X, Y of G, if $\ell \in X \setminus Y$ then there exists $k \in Y \setminus X$ such that $(Y \setminus k) \cup \{\ell\}$ is a basis for G.

We next provide some examples of matroids. Consider a ground set E with cardinality d. *Uniform matroid* of rank m is $(E, \mathcal{I})$, where $\mathcal{I}$ is the collection of subsets of E with at most m elements. Consider a partition of E given by $\{E_i\}_{i \in [\ell]}$. For some given parameters $k_1, \ldots, k_\ell$, define

$$\mathcal{I} = \{X \subseteq E : |X \cap E_i| \leq k_i, \ \forall i \in [\ell]\}.$$

Then $(E, \mathcal{I})$ is *partition matroid* of rank $\sum_{i \in [\ell]} k_i$. Given an undirected graph $\mathcal{G} = (V, H)$, define

$$\mathcal{I} = \{F \subseteq H : (V, F) \text{ is a forest}\}.$$

Then, it can be shown that $G(\mathcal{G}) = (H, \mathcal{I})$ is a matroid, referred to as *graphic matroid*. Every spanning forest of the graph G is indeed a basis for matroid $G(\mathcal{G})$.

3.1.1 Weighted Matroids

Now we consider a weighted matroid, where each element of the ground set is given a non-negative weight. For any $\ell \in E$, let w_ℓ denote the weight assigned to ℓ. The matroid optimization problem is to find a basis with maximum total weight:

$$\max_{X \in \mathcal{I}} \sum_{\ell \in X} w_\ell. \tag{1}$$

The above problem can be solved efficiently by the GREEDY algorithm, whose pseudo-code is shown in Algorithm 1.

Algorithm 1 GREEDY [26]

Sort weights $w_i, i \in E$. Denote the new ordering by a bijection $k : E \to E$:

$$w_{k(1)} \geq w_{k(2)} \geq \cdots \geq w_{k(d)}.$$

$X \leftarrow \emptyset$
for $i = 1, \ldots, d$ **do**
  if $X \cup \{k(i)\} \in \mathcal{I}$ **then**
    $X \leftarrow X \cup \{k(i)\}$
  end if
end for

Next we determine the complexity of the GREEDY algorithm. Clearly, sorting can be carried out in $\mathcal{O}(d \log(d))$. Furthermore, assume that testing whether a given subset of the ground set E is independent takes $\mathcal{O}(h(d))$ time for some function h. Then, the time complexity of the GREEDY algorithm is $\mathcal{O}(d \log(d) + dh(d))$. In some computational models, it is assumed that an algorithm has access to an *independence oracle*, that is a routine that given $X \subset E$ returns whether $X \in \mathcal{I}$ or not. Under the independence oracle model, the GREEDY algorithm has a time complexity of $\mathcal{O}(d \log(d))$. Hence, a maximum-weight independent set in a matroid can be found in strongly polynomial time under independence oracle model ([28, Corollary 40.1]).

[1]For any set X and element ℓ, by a slight abuse of notation, we write $X \setminus \ell$ to imply $X \setminus \{\ell\}$.

3.2 MAB Model

Consider a finite set of *basic actions* $E = \{1, \ldots, d\}$ and a matroid $G = (E, \mathcal{I})$ of rank m. We consider a combinatorial MAB problem, where each arm M is a basis of G. We let $\mathcal{M}$ denote the set of arms, i.e., the collection of all bases of G. Each arm M is identified with a binary column vector $(M_1, \ldots, M_d)^\top$, and we have $\|M\|_1 = m$, $\forall M \in \mathcal{M}$ since G is of rank m. Time proceeds in rounds. For $i \in E$, $X_i(n)$ denotes the random reward of basic action i in round n. For each i, the sequence $(X_i(n))_{n \geq 1}$ is i.i.d. with Bernoulli distribution of mean θ_i. The rewards across basic actions may be arbitrarily correlated. We denote by $\theta = (\theta_1, \ldots, \theta_d)^\top \in \Theta = [0, 1]^d$ the vector of unknown expected rewards of the various basic actions.

At the beginning of each round n, an algorithm or policy π, selects an arm $M^\pi(n) \in \mathcal{M}$ based on the arms chosen in previous rounds and their observed rewards. The reward of arm $M^\pi(n)$ selected in round n is

$$X^{M^\pi(n)}(n) = \sum_{i \in E} M_i^\pi(n) X_i(n) = M^\pi(n)^\top X(n).$$

We consider semi-bandit feedback, where under policy π and at the end of round n, the outcome of actions $X_i(n)$ for all $i \in M^\pi(n)$ are revealed to the decision maker.[2] The objective is to identify a policy in Π, the set of all feasible policies, which maximizes the cumulative expected reward over a finite time horizon T. Here the expectation is understood with respect to the randomness in the rewards and the possible randomization in the policy. Equivalently, we aim at designing a policy that minimizes regret, where the regret of policy $\pi \in \Pi$ is defined by:

$$R^\pi(T) = \max_{M \in \mathcal{M}} \mathbb{E}[\sum_{n=1}^{T} X^M(n)] - \mathbb{E}[\sum_{n=1}^{T} X^{M^\pi(n)}(n)].$$

Finally, we denote by $\mu_M(\theta) = M^\top \theta$ the expected reward of arm M, and let $M^\star(\theta) \in \mathcal{M}$ be any arm with maximum expected reward:

$$M^\star(\theta) \in \arg\max_{M \in \mathcal{M}} \mu_M(\theta).$$

To simplify the presentation in subsequent analysis, we assume that the elements of the vector θ are distinct, and hence the optimal arm $M^\star(\theta)$ is unique. We further define: $\Delta_M = M^\star(\theta)^\top \theta - \mu_M(\theta)$ for all $M \in \mathcal{M}$.

In subsequent sections, when clear from the context that θ is the underlying parameter, we use $M^\star$ to indicate $M^\star(\theta)$.

4. THE KL-OSM ALGORITHM

In this section, we present KL-OSM which is a natural extension of the KL-UCB algorithm [14] for the described matroid bandit problem.

We introduce the following notation: At time n, we define $t_i(n) = \sum_{s=1}^{n} M_i(s)$ the number of times basic action i has

[2]For brevity, in what follows, for any binary vector z, we write $i \in z$ to denote $z_i = 1$.

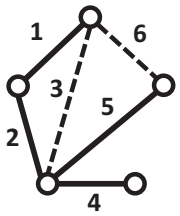


Figure 1: An example for the set $\mathcal{K}_i$ in the case of graphic matroids: Edges shown with solid line correspond to optimal actions. Two sub-optimal actions are shown in dashed line, where $\mathcal{K}_3 = \{1, 2\}$ and $\mathcal{K}_6 = \{1, 2, 5\}$.

been sampled. In round n, we define the empirical mean reward of action i as $\hat{\theta}_i(n) = (1/t_i(n)) \sum_{s=1}^{n} X_i(s) M_i(s)$ if $t_i(n) > 0$ and $\hat{\theta}_i(n) = 0$ otherwise. Our algorithm is an index policy relying on KL-UCB index [14] maintained for each basic action. More precisely, the index of basic action i in round n is denoted by $\omega_i(n)$ and defined as:

$$\omega_i(n) = \max\Big\{q \in [\hat{\theta}_i(n), 1] : t_i(n)\mathrm{kl}(\hat{\theta}_i(n), q) \leq f(n)\Big\},$$

with $f(n) = \log(n) + 3\log(\log(n))$.

In each round $n \geq 1$, the KL-OSM algorithm simply consists in computing indexes $\omega_i(n)$ for all i, and then selecting an arm $M(n)$ by solving

$$M(n) \in \arg\max_{M \in \mathcal{M}} \sum_{i \in M} \omega_i(n),$$

using the GREEDY algorithm. The pseudo-code of KL-OSM is given in Algorithm 2.

Algorithm 2 KL-OSM

for $n \geq 1$ **do**
  Select $M(n) \in \mathrm{argmax}_{M \in \mathcal{M}} \sum_{i \in M} \omega_i(n)$ using GREEDY.
  Play $M(n)$, observe the rewards, and update $t_i(n)$ and $\hat{\theta}_i(n), \forall i \in M(n)$.
end for

Next we provide a finite-time analysis of the regret of KL-OSM. To this aim, for any problem instance we introduce mapping $\sigma : E \backslash M^\star \to M^\star$ such that for any $i \in E \backslash M^\star$:

$$\sigma(i) = \arg\min_{j \in \mathcal{K}_i} \theta_j,$$

where

$$\mathcal{K}_i = \Big\{\ell \in M^\star : (M^\star \setminus \ell) \cup \{i\} \in \mathcal{M}\Big\}.$$

Figure 1 shows an example for the set $\mathcal{K}_i$ for the case of graphic matroids.

It is noted that by Proposition 1, we have that $\mathcal{K}_i \neq \emptyset$ for all $i \in E \setminus M^\star$. Moreover, for any $i \notin M^\star$, if $\ell \in \mathcal{K}_i$, then $\theta_\ell > \theta_i$. We show this claim by contradiction: assume this does not hold, namely, $\theta_\ell < \theta_i$. Consider $M' = (M^\star \backslash \ell) \cup \{i\}$. Then, by Proposition 1, $M' \in \mathcal{M}$. Moreover,

$$\mu_{M'}(\theta) - \mu_{M^\star}(\theta) = \sum_{k \in M'} \theta_k - \sum_{k \in M^\star} \theta_k = \theta_i - \theta_\ell > 0,$$

which contradicts the optimality of $M^\star$. Hence, $\theta_\ell > \theta_i$ for any $\ell \in \mathcal{K}_i$.

The following theorem gives an upper bound on the regret of the KL-OSM algorithm.

THEOREM 1. *For any $\varepsilon > 0$, there exist positive constants C_1, $C_2(\varepsilon)$, and $\beta(\varepsilon)$ such that the regret under algorithm π =*KL-OSM *satisfies:*

$$R^\pi(T) \leq \sum_{i \in E \backslash M^\star} \frac{\theta_{\sigma(i)} - \theta_i}{\mathrm{kl}(\theta_i, \theta_{\sigma(i)})}(1 + \varepsilon)\log(T) + o(\log(T)).$$

Hence,

$$\limsup_{T \to \infty} \frac{R^\pi(T)}{\log(T)} \leq \sum_{i \in E \backslash M^\star} \frac{\theta_{\sigma(i)} - \theta_i}{\mathrm{kl}(\theta_i, \theta_{\sigma(i)})}.$$

REMARK 1. *When the underlying matroid is a uniform matroid, the problem reduces to MAB with multiple plays as studied in [2, 17]. Assume that actions are enumerated such that $\theta_1 \geq \theta_2 \geq \ldots \theta_m > \cdots \geq \theta_d$. Then $M^\star = \{1, 2, \ldots, m\}$ and $\sigma(i) = m$ for all $i \notin M^\star$. Hence, the regret upper bound of Theorem 1 asymptotically coincides with the results provided in [2, 17].*

Next we compare KL-OSM and OMM [19] in terms of their regret upper bounds. OMM achieves a regret upper-bounded by

$$\mathbb{E}[R^\pi(T)] \leq \sum_{i \in E \backslash M^\star} \frac{16}{\Delta_{\min,i}} \log(T) + \mathcal{O}(1),$$

where for any sub-optimal i:

$$\Delta_{\min,i} = \min_{j \in E \backslash M^\star} |\theta_i - \theta_j|.$$

Note that by Pinsker's inequality, we have

$$\mathrm{kl}(\theta_i, \theta_{\sigma(i)}) \geq 2(\theta_i - \theta_{\sigma(i)})^2 \geq 2\Delta_{\min,i}^2.$$

Hence, the regret upper bound for KL-OSM is better than that for OMM. The numerical experiments in Section 6 also show that KL-OSM outperforms OMM in practice.

Implementation.

The KL-OSM algorithm finds a basis with the maximum index using the GREEDY algorithm, whose time complexity under independence oracle model is $\mathcal{O}(d\log(d))$. We also remark that the computation of index $\omega_i(n)$ amounts to finding the roots of a strictly convex and increasing function in one variable (since $z \mapsto \mathrm{kl}(p, z)$ is an increasing function for $z \geq p$). Hence, $\omega_i(n)$ can be computed straightforwardly by a simple line search such as bisection. Therefore, the time complexity of KL-OSM after T rounds is $\mathcal{O}(dT\log(d))$.

5. LOWER BOUND

In this section, we derive a lower bound on the regret of any uniformly good algorithm π for the matroid bandit problem under the case where the rewards across basic actions are independent. We define uniformly good algorithms as in [22]: An algorithm π is uniformly good if and only if $R^\pi(T) = o(T^\alpha)$ for all $\alpha > 0$ and all parameters $\theta \in \Theta$.

The proof of this lower bound uses the theory of optimal control of Markov chains with unknown transition probabilities studied in [16]. Such a technique is also used in [10]

to study the regret lower bound for generic stochastic combinatorial MABs. In what follows, first we state the latter result.

Given $\theta \in \Theta$, define the set of *bad* parameters that cannot be distinguished from θ when selecting arm $M^\star(\theta)$, and for which the arm $M^\star(\theta)$ is sub-optimal:

$$B(\theta) = \{\lambda \in \Theta : \lambda_i = \theta_i, \forall i \in M^\star(\theta),\ \max_M M^\top \lambda > M^\star(\theta)^\top \theta\}.$$

According to [10, Theorem 1], the regret of any uniformly good policy $\pi \in \Pi$ for any $\theta \in \Theta$ satisfies

$$\liminf_{T\to\infty} \frac{R^\pi(T)}{\log(T)} \geq c(\theta),$$

where $c(\theta)$ is the optimal value of the following problem:

$$\inf_{x\geq 0} \sum_{M\in\mathcal{M}} \Delta_M x_M \qquad (2)$$

$$\text{subject to: } \sum_{M\in\mathcal{M}} x_M \sum_{i\in E} M_i \mathrm{kl}(\theta_i, \lambda_i) \geq 1,\ \ \forall \lambda \in B(\theta).$$

Here $\mathrm{kl}(u, v)$ is the Kullback-Leibler divergence between Bernoulli distributions of respective means u and v, i.e.,

$$\mathrm{kl}(u, v) = u\log(u/v) + (1-u)\log((1-u)/(1-v)).$$

Let $x^\star = (x^\star_M, M \in \mathcal{M})$ denote the optimal solution to this problem. It then follows from [16] that the expected number of times that an optimal algorithm plays arm M up to round T will be $x^\star_M \log(T) + o(\log(T))$. Note that the optimal value $c(\theta)$ and solution $x^\star$ are unfortunately not explicit.

Building on this result, in the next theorem we provide an *asymptotic* lower bound on the regret of any uniformly good policy π for the considered matroid optimization problem.

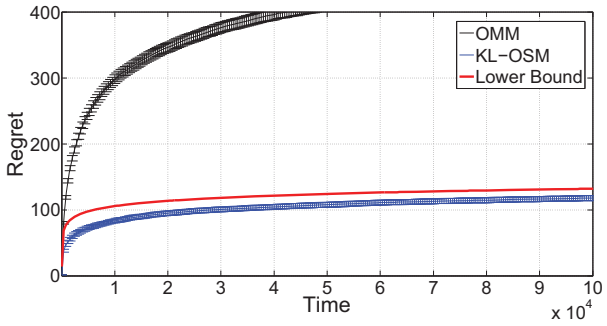
THEOREM 2. *For all $\theta \in \Theta$ and any uniformly good algorithm $\pi \in \Pi$,*

$$\liminf_{T\to\infty} \frac{R^\pi(T)}{\log(T)} \geq \sum_{i\in E\setminus M^\star} \frac{\theta_{\sigma(i)} - \theta_i}{\mathrm{kl}(\theta_i, \theta_{\sigma(i)})}.$$

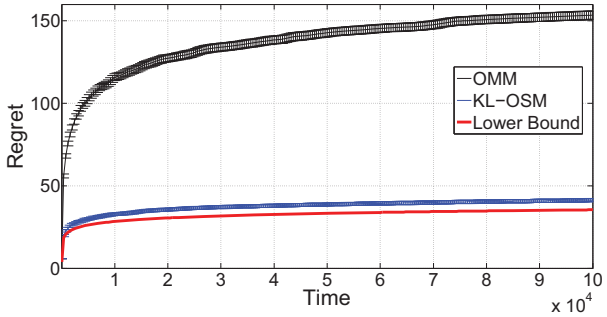
Comparing the result of Theorem 1 with that of Theorem 2, we observe that for the case of Bernoulli rewards, the regret upper bound of KL-OSM asymptotically matches the lower bound. Hence, it is asymptotically optimal. We remark that contrary to the lower bound in [19], the lower bound in Theorem 2 is problem-dependent, namely it holds for any parameter θ and any matroid G. Moreover, in contrast to the lower bound given in [10, Theorem 1], ours is explicit.

6. NUMERICAL EXPERIMENTS

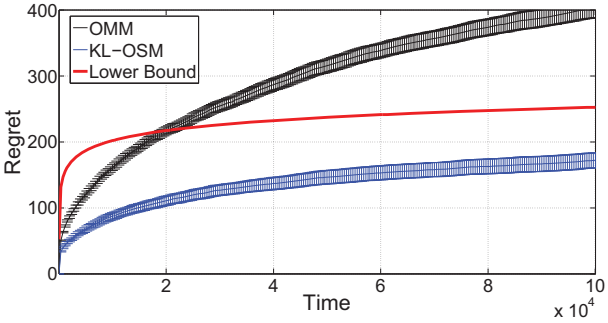
We briefly illustrate the performance under the KL-OSM algorithm for the case of graphic and partition matroids. In our first experiment, we consider spanning trees in the complete graph K_N. We set $N = 5$, in which case by Cayley's formula there are 5^3 spanning trees or arms, and $d = 10$ basic actions. In this experiment we consider two scenarios: 'Scenario 1', in which parameter θ is chosen such that $\theta_i = 0.8$ if $i \in M^\star$ and $\theta_i = 0.6$ otherwise; and 'Scenario 2', where θ is drawn uniformly at random from $[0, 1]^{10}$. In the second experiment, we consider a partition matroid of rank 5 with $d = 10$ basic actions and 4 partitions. Furthermore, parameter θ is drawn uniformly at random from $[0, 1]^{10}$.



(a) Graphic matroid, Scenario 1



(b) Graphic matroid, Scenario 2



(c) Partition matroid

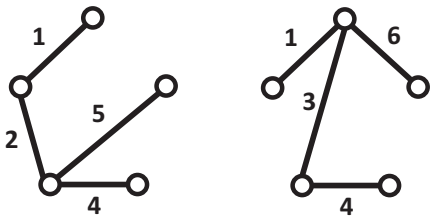
Figure 2: Regret of various algorithms

Figures 2(a)-(c) present the regret vs. time horizon under KL-OSM and OMM for the various cases. In these figures, curves in blue and black show the average over 100 independent runs along with the 95% confidence intervals. Moreover, the curve in red represents the lower bound of Theorem 2 for the corresponding scenario.

We observe that in all experiments, KL-OSM significantly outperforms OMM. The curves in Figures 2(a)-(c) show the regret of KL-OSM is growing at the same rate of the 'lower bound' when the number of rounds grows large, thus verifying the asymptotic optimality of KL-OSM. Finally, we remark that it is not contradictory that the 'lower bound' curve in Figure 2(a) or 2(c) lies above the regret curve of KL-OSM. This is because the lower bound of Theorem 2 holds asymptotically, i.e., when T grows large.

7. CONCLUSION

In this paper we have investigated matroid bandits with Bernoulli rewards. We have provided a tight and problem-



(a) $M^\star$ (b) $M(n)$

Figure 3: An example of bijection τ_n for the case of graphic matroids. In this case: $\tau_n(1) = 1, \tau_n(3) = 2, \tau_n(4) = 4, \tau_n(6) = 5$.

dependent lower bound on the regret. Moreover, we proposed KL-OSM, an efficient algorithm for matroid bandits, and provided a finite-time analysis of its regret. We showed that the regret upper bound of KL-OSM matches the lower bound, and hence it is asymptotically optimal. Moreover, we showed that KL-OSM enjoys a better regret than existing algorithms both theoretically and experimentally. As a future work, we will investigate matroid bandits with more complicated reward functions. Of particular interest is the case where the reward function is a submodular set function.

Acknowledgement

Alexandre Proutiere's research is supported by the ERC FSA grant and the SSF ICT-Psi project.

APPENDIX

A. PROOF OF THEOREM 1

PROOF. Let $T > 0$. Consider round n where $M(n) \neq M^\star$ is selected by the algorithm π =KL-OSM. Using Lemma 1, proven at the end of this section, there exists a bijective mapping τ_n such that: $\tau_n(i) = i$ if $i \in M^\star \cap M(n)$ or $i \in \overline{M^\star \cup M(n)}$. Otherwise, $\tau_n(i) = j$ for some $j \in \mathcal{K}_i' := \mathcal{K}_i \setminus M(n)$. The mapping τ_n simply maps the sub-optimal basic actions of $M(n)$ to the corresponding ones in $M^\star$ that are not chosen by the algorithm at round n. It then follows that

$$\mathbf{1}\{M_i(n) = 1\} = \sum_{j \in \mathcal{K}_i'} \mathbf{1}\{M_i(n) = 1,\ \tau_n(i) = j\}$$

and that $\sum_{j \in \mathcal{K}_i'} \mathbf{1}\{\tau_n(i) = j\} \leq 1$ since τ_n is bijective. An example of mapping τ_n for the case of graphic matroids is shown in Figure 3.

For any $i, j \in E$, define $\Delta_{j,i} = \theta_j - \theta_i$. Then, the regret under policy π=KL-OSM is upper bounded as:

$$\begin{aligned}
R^\pi(T) &\leq \mathbb{E}[\sum_{n=1}^{T} \Delta_{M(n)}] \\
&= \mathbb{E}[\sum_{n=1}^{T} \sum_{i \in E \setminus M^\star} \Delta_{\tau_n(i),i} \mathbf{1}\{M_i(n) = 1\}] \\
&= \mathbb{E}[\sum_{i \in E \setminus M^\star} \sum_{n=1}^{T} \sum_{j \in \mathcal{K}_i'} \Delta_{j,i} \mathbf{1}\{M_i(n) = 1,\ \tau_n(i) = j\}].
\end{aligned}$$

Let $i \in E \setminus M^\star$. Motivated by the design of KL-OSM, namely use of GREEDY to determine $M(n)$ at each round n, we use the following decomposition:

$$\begin{aligned}
\mathbf{1}\{M_i(n) = 1,\ \omega_i(n) \geq \omega_{\tau_n(i)}(n)\} &\leq \mathbf{1}\{\omega_{\tau_n(i)}(n) < \theta_{\tau_n(i)}\} \\
&\quad + \mathbf{1}\{M_i(n) = 1,\ \omega_i(n) \geq \theta_{\tau_n(i)}\}.
\end{aligned}$$

Hence,

$$\begin{aligned}
&\sum_{n=1}^{T} \sum_{j \in \mathcal{K}_i'} \Delta_{j,i} \mathbf{1}\{M_i(n) = 1,\ \tau_n(i) = j\} \\
&\quad \leq \sum_{n=1}^{T} \sum_{j \in \mathcal{K}_i'} \Delta_{j,i} \mathbf{1}\{\tau_n(i) = j,\ \omega_j(n) < \theta_j\} \\
&\quad + \sum_{n=1}^{T} \sum_{j \in \mathcal{K}_i'} \Delta_{j,i} \mathbf{1}\{M_i(n) = 1,\ \tau_n(i) = j,\ \omega_i(n) \geq \theta_j\},
\end{aligned}$$

and therefore,

$$\begin{aligned}
&\mathbb{E}[\sum_{i \in E \setminus M^\star} \sum_{n=1}^{T} \sum_{j \in \mathcal{K}_i'} \Delta_{j,i} \mathbf{1}\{M_i(n) = 1,\ \tau_n(i) = j\} \\
&\leq \mathbb{E}[\sum_{i \in E \setminus M^\star} \sum_{n=1}^{T} \sum_{j \in \mathcal{K}_i'} \mathbf{1}\{\tau_n(i) = j,\ \omega_j(n) < \theta_j\}] \\
&+ \mathbb{E}[\sum_{i \in E \setminus M^\star} \sum_{n=1}^{T} \sum_{j \in \mathcal{K}_i'} \Delta_{j,i} \mathbf{1}\{M_i(n) = 1,\ \tau_n(i) = j,\ \omega_i(n) \geq \theta_j\}],
\end{aligned}$$

since $\Delta_{j,i} \leq 1$. We prove that there exist positive constants C_1, $C_2(\varepsilon)$, and $\beta(\varepsilon)$ such that

$$\begin{aligned}
&\mathbb{E}[\sum_{i \in E \setminus M^\star} \sum_{n=1}^{T} \sum_{j \in \mathcal{K}_i'} \mathbf{1}\{\tau_n(i) = j,\ \omega_j(n) < \theta_j\}] \\
&\quad \leq (d - m) C_1 \log(\log(T)), \qquad (3) \\
&\mathbb{E}[\sum_{i \in E \setminus M^\star} \sum_{n=1}^{T} \sum_{j \in \mathcal{K}_i'} \Delta_{j,i} \mathbf{1}\{M_i(n) = 1,\ \tau_n(i) = j,\ \omega_i(n) \geq \theta_j\}] \\
&\quad \leq \sum_{i \in E \setminus M^\star} (1 + \varepsilon) \frac{\theta_{\sigma(i)} - \theta_i}{\mathrm{kl}(\theta_i, \theta_{\sigma(i)})} f(T) + (d - m) \frac{C_2(\varepsilon)}{T^{\beta(\varepsilon)}}. \qquad (4)
\end{aligned}$$

Hence, we get the announced result:

$$\begin{aligned}
R^\pi(T) &\leq \mathbb{E}[\sum_{i \in E \setminus M^\star} \sum_{n=1}^{T} \sum_{j \in \mathcal{K}_i'} \Delta_{j,i} \mathbf{1}\{M_i(n) = 1,\ \tau_n(i) = j\}] \\
&\leq \sum_{i \in E \setminus M^\star} \frac{\theta_{\sigma(i)} - \theta_i}{\mathrm{kl}(\theta_i, \theta_{\sigma(i)})} (1 + \varepsilon) f(T) \\
&\quad + (d - m) \left(\frac{C_2(\varepsilon)}{T^{\beta(\varepsilon)}} + C_1 \log(\log(T)) \right).
\end{aligned}$$

Inequality (3):
Fix $j \in \mathcal{K}_i'$. By the concentration inequality in [14, Theorem 10], we have

$$\mathbb{P}[\omega_j(n) < \theta_j] \leq \lceil f(n) \log(n) \rceil e^{1 - f(n)},$$

and hence following the same steps as in the proof of [14, Theorem 2], we observe that there exists constant $C_1 \leq 7$

such that $\mathbb{E}[\sum_{n=1}^{T}\mathbf{1}\{\omega_j(n)<\theta_j\}]\leq C_1\log(\log(T))$. It then follows that

$$\sum_{j\in\mathcal{K}_i'}\mathbb{E}[\sum_{n=1}^{T}\mathbf{1}\{\tau_n(i)=j,\ \omega_j(n)<\theta_j\}]\leq C_1(\log(\log(T)))$$

since τ_n for any n is a bijection. As a result:

$$\sum_{i\notin M^\star}\sum_{j\in\mathcal{K}_i'}\mathbb{E}[\sum_{n=1}^{T}\mathbf{1}\{\tau_n(i)=j,\ \omega_j(n)<\theta_j\}]\leq(d-m)C_1(\log(\log(T))).$$

Inequality (4):

For $x,y\in[0,1]$, introduce $\mathrm{kl}^+(x,y)=\mathrm{kl}(x,y)\mathbf{1}\{x<y\}$. Fix $j\in\mathcal{K}_i'$. Observe that the event $\omega_i(n)\geq\theta_j$ implies $\omega_i(n)\geq\theta_{\sigma(i)}$, which further implies that $\mathrm{kl}^+(\hat{\theta}_i(n),\theta_{\sigma(i)})\leq\mathrm{kl}(\hat{\theta}_i(n),\omega_i(n))=f(n)/t_i(n)$.

We let $\hat{\theta}_{i,s}$ denote the empirical average of rewards of action i when it is selected s times. Hence following the similar steps as in the proof of Lemma 7 in [14], we obtain:

$$\begin{aligned}
&\sum_{j\in\mathcal{K}_i'}\sum_{n=1}^{T}\mathbf{1}\{M_i(n)=1,\ \tau_n(i)=j,\ \omega_i(n)\geq\theta_j\}\\
&\leq\sum_{j\in\mathcal{K}_i'}\sum_{n=1}^{T}\mathbf{1}\{M_i(n)=1,\ \tau_n(i)=j,\ \omega_i(n)\geq\theta_{\sigma(i)}\}\\
&\leq\sum_{n=1}^{T}\mathbf{1}\{M_i(n)=1,\ \omega_i(n)\geq\theta_{\sigma(i)}\}\\
&\leq\sum_{n=1}^{T}\mathbf{1}\{M_i(n)=1,\ t_i(n)\mathrm{kl}^+(\hat{\theta}_i(n),\theta_{\sigma(i)})\leq f(n)\}\\
&=\sum_{n=1}^{T}\sum_{s=1}^{n}\mathbf{1}\{M_i(n)=1,\ t_i(n)=s,\ s\mathrm{kl}^+(\hat{\theta}_{i,s},\theta_{\sigma(i)})\leq f(n)\}\\
&\leq\sum_{n=1}^{T}\sum_{s=1}^{n}\mathbf{1}\{M_i(n)=1,\ t_i(n)=s,\ s\mathrm{kl}^+(\hat{\theta}_{i,s},\theta_{\sigma(i)})\leq f(T)\}\\
&=\sum_{s=1}^{T}\mathbf{1}\{s\mathrm{kl}^+(\hat{\theta}_{i,s},\theta_{\sigma(i)})\leq f(T)\}\sum_{n=s}^{T}\mathbf{1}\{M_i(n)=1,\ t_i(n)=s\}\\
&=\sum_{s=1}^{T}\mathbf{1}\{s\mathrm{kl}^+(\hat{\theta}_{i,s},\theta_{\sigma(i)})\leq f(T)\},
\end{aligned}$$

where in the last step, we used the fact that for any s, there is only one round n such that $t_i(n)=s$ and $M_i(n)=1$.

From [14, Lemma 8], we have that (see the arXiv version of the present work for details [30]):

$$\mathbb{E}[\sum_{s=1}^{T}\mathbf{1}\{s\mathrm{kl}^+(\hat{\theta}_{i,s},\theta_{\sigma(i)})\leq f(T)\}]\leq\frac{(1+\varepsilon)f(T)}{\mathrm{kl}(\theta_i,\theta_{\sigma(i)})}+\frac{C_2(\varepsilon)}{T^{\beta(\varepsilon)}},$$

so that

$$\mathbb{E}[\sum_{n=1}^{T}\sum_{j\in\mathcal{K}_i'}\Delta_{j,i}\mathbf{1}\{M_i(n)=1,\ \tau_n(i)=j,\ \omega_i(n)\geq\theta_j\}]\leq\frac{\theta_{\sigma(i)}-\theta_i}{\mathrm{kl}(\theta_i,\theta_{\sigma(i)})}(1+\varepsilon)f(T)+\frac{C_2(\varepsilon)}{T^{\beta(\varepsilon)}}.$$

Summing over $i\in E\backslash M^\star$ completes the proof of inequality (4) and hence concludes the proof. □

LEMMA 1. *For every $M,M'\in\mathcal{M}$, there exists a bijective mapping $\tau_{MM'}:E\to E$, or τ for short, such that $\tau(i)=i$ if $i\in M\cap M'$ or $i\in\overline{M\cup M'}$. Otherwise, $\tau(i)=j$ for some $j\in\mathcal{L}_i$, where*

$$\mathcal{L}_i=\Big\{\ell\in M'\setminus M:(M'\setminus\ell)\cup\{i\}\in\mathcal{M}\Big\}.$$

In particular this lemma implies that for any problem instance, i.e. fixed $M^\star$, and for any $M\in\mathcal{M}$, there exists a bijective mapping $\tau_{MM^\star}$, or τ for short, which maps any optimal basic action of M to itself, and maps any sub-optimal basic action $i\in M$ to some element in $\mathcal{K}_i\setminus M$. Note that such a mapping may not be unique.

PROOF. Let $M,M'\in\mathcal{M}$. We provide an algorithm that outputs $\tau_{MM'}$. To present the algorithm, for any $i\in M\setminus M'$ let us define:

$$\mathcal{D}_i=\Big\{\ell\in M'\setminus M:(M\setminus i)\cup\{\ell\}\in\mathcal{M}\Big\}.$$

The pseudo-code of the algorithm is shown in Algorithm 3.

Algorithm 3 CONSTRUCTION OF MAPPING $\tau_{MM'}$

Initialization:
Set $\tau(i)=i$ for all $i\in M\cap M'$ and $i\in\overline{M\cup M'}$.
Set $S_i=\mathcal{D}_i$ for all $i\in M\setminus M'$. Set $Q=M$.
while $Q\neq M'$ **do**
  Let $i_0\in\mathrm{argmin}_{i\in M\setminus M'}|S_i|$ (ties are broken arbitrarily).
  Select $j\in S_{i_0}$ arbitrarily. Set $\tau(i_0)=j$.
  $Q\leftarrow(Q\setminus i_0)\cup\{j\}$
  for $i\in M\setminus M'$ **do**
    $S_i\leftarrow S_i\setminus j$
  end for
end while
Output τ.

Clearly the algorithm terminates after at most m steps since at each step one element in Q is replaced with some element of M'. In order to guarantee that the output of the algorithm is a bijective mapping it suffices to show at that any step, if $S_i=S_j=\{x\}$ for some $i,j\in B$, then necessarily $i=j$. We prove this claim by contradiction. Assume $S_i=S_j=\{x\}$ and $i\neq j$. Consider basis $M_1=(Q\setminus i)\cup\{x\}$. Note that $M_1\neq M'$ since $j\in M_1$. Hence there must exist $\ell\in M'\setminus M_1$ such that $(M_1\setminus j)\cup\{\ell\}$ is a basis, and by definition $\ell\in S_j$. This is clearly a contradiction since $S_j=\{x\}$ and $x\notin M'\setminus M_1$.

Finally, we show that at each step $\tau(i_0)\in\mathcal{L}_{i_0}$. Observe that $\tau(i_0)\in S_{i_0}$ implies $\tau(i_0)\in\mathcal{D}_{i_0}$ since $S_{i_0}\subseteq\mathcal{D}_{i_0}$. Observe that $\ell\in\mathcal{D}_{i_0}$ implies that $(M\setminus i_0)\cup\{\ell\}$ is a basis, and hence $(M'\setminus\ell)\cup\{i_0\}$ is a basis, so that by definition $\ell\in\mathcal{L}_{i_0}$. This further implies that $\tau(i_0)\in\mathcal{L}_{i_0}$, which concludes the proof. □

B. PROOF OF THEOREM 2

PROOF. For any $M\neq M^\star$ introduce

$$B_M(\theta)=\{\lambda\in\Theta:\lambda_i=\theta_i,\forall i\in M^\star(\theta),\ M^\top\lambda>M^\star(\theta)^\top\theta\}.$$

Observing that $B(\theta)=\cup_{M\neq M^\star}B_M(\theta)$, we equivalently rewrite

problem (2) as

$$\inf_{x\geq 0} \sum_{M\neq M^\star} \Delta_M x_M, \tag{5}$$

subject to:

$$\inf_{\lambda\in B_M(\theta)} \sum_{i\in M\setminus M^\star} \mathrm{kl}(\theta_i,\lambda_i) \sum_{Q\in\mathcal{M}} Q_i x_Q \geq 1,\ \ \forall M\neq M^\star.$$

Fix $i\in E\setminus M^\star$. Recall that $\sigma(i)=\mathrm{argmin}_{j\in\mathcal{K}_i}\theta_j$ and let $M^{(i)}=(M^\star\setminus\sigma(i))\cup\{i\}$. By Proposition 1, $M^{(i)}\in\mathcal{M}$. We may simplify the l.h.s. of the constraint corresponding to arm $M^{(i)}$ in (5) as follows:

$$\begin{aligned}
&\inf_{\lambda\in B_{M^{(i)}}(\theta)} \sum_{j\in M^{(i)}\setminus M^\star} \mathrm{kl}(\theta_j,\lambda_j)\sum_Q Q_j x_Q\\
&\quad= \inf_{\lambda\in B_{M^{(i)}}(\theta)} \mathrm{kl}(\theta_i,\lambda_i)\sum_Q Q_i x_Q\\
&\quad= \inf_{\lambda\in\Theta:\lambda_i>\theta_{\sigma(i)}} \mathrm{kl}(\theta_i,\lambda_i)\sum_Q Q_i x_Q\\
&\quad= \mathrm{kl}(\theta_i,\theta_{\sigma(i)})\sum_Q Q_i x_Q,
\end{aligned}$$

where we used $M^{(i)}\setminus M^\star=\{i\}$. Let $\mathcal{M}^-=\mathcal{M}\setminus(\{M^\star\}\cup\{M^{(i)},i\in E\setminus M^\star\})$. It then follows that

$$\begin{aligned}
c(\theta) \ =\ &\inf_{x\geq 0} \sum_{M\in\mathcal{M}} \Delta_M x_M \qquad (6)\\
&\text{subject to:}\\
&\sum_{M\neq M^\star} M_i x_M \geq \frac{1}{\mathrm{kl}(\theta_i,\theta_{\sigma(i)})},\ \ \forall i\in E\setminus M^\star,\\
&\inf_{\lambda\in B_M(\theta)} \sum_{Q\in\mathcal{M}} x_Q \sum_{i\in E} Q_i \mathrm{kl}(\theta_i,\lambda_i)\geq 1,\ \ \forall M\in\mathcal{M}^-.
\end{aligned}$$

Defining

$$\begin{aligned}
&\mathsf{P1}\colon\quad \inf_{x\geq 0} \sum_{M\neq M^\star} \Delta_M x_M\\
&\text{subject to:}\quad \sum_{M\neq M^\star} M_i x_M \geq \frac{1}{\mathrm{kl}(\theta_i,\theta_{\sigma(i)})},\ \ \forall i\in E\setminus M^\star,
\end{aligned}$$

gives $c(\theta)\geq \mathrm{val}(\mathsf{P1})$ since the feasible region of problem (6) is contained in that of P1. [3]

For any sub-optimal action $i\in E$, introduce $z_i=\sum_M M_i x_M$, and define $z=(z_i,i\in E)$. Next we represent the objective of P1 in terms of z, and give a lower bound for it. Using Lemma 1, there exists a bijective mapping τ_M such that $\tau_M(i)=i$ if $i\in M\cap M^\star$ or $i\in\overline{M\cup M^\star}$. Otherwise, $\tau_M(i)=j$ for some $j\in\mathcal{K}_i\setminus M$. We have:

$$\begin{aligned}
\Delta_M &= \sum_{i\in M}(\theta_{\tau_M(i)}-\theta_i)\\
&= \sum_{i\in M\setminus M^\star}(\theta_{\tau_M(i)}-\theta_i)\\
&= \sum_{i\in E\setminus M^\star} M_i(\theta_{\tau_M(i)}-\theta_i)\\
&\geq \sum_{i\in E\setminus M^\star} M_i(\theta_{\sigma(i)}-\theta_i).
\end{aligned}$$

[3]We use val(P) to denote the optimal value of a given optimization problem P.

Hence,

$$\begin{aligned}
\sum_M x_M\Delta_M &\geq \sum_M x_M \sum_{i\notin M^\star} M_i(\theta_{\sigma(i)}-\theta_i)\\
&= \sum_{i\notin M^\star}(\theta_{\sigma(i)}-\theta_i)z_i.
\end{aligned}$$

Then, defining

$$\begin{aligned}
&\mathsf{P2}\colon\quad \inf_{z\geq 0} \sum_{i\in E\setminus M^\star}(\theta_{\sigma(i)}-\theta_i)z_i\\
&\text{subject to:}\ \ z_i\geq\frac{1}{\mathrm{kl}(\theta_i,\theta_{\sigma(i)})},\ \ \forall i\in E\setminus M^\star,
\end{aligned}$$

yields: $c(\theta)\geq\mathrm{val}(\mathsf{P1})\geq\mathrm{val}(\mathsf{P2})$. The proof is completed by observing that

$$\mathrm{val}(\mathsf{P2})=\sum_{i\in E\setminus M^\star}\frac{\theta_{\sigma(i)}-\theta_i}{\mathrm{kl}(\theta_i,\theta_{\sigma(i)})}.$$

□

REFERENCES

[1] Z. Abbassi, V. S. Mirrokni, and M. Thakur. Diversity maximization under matroid constraints. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 32–40, 2013.

[2] V. Anantharam, P. Varaiya, and J. Walrand. Asymptotically efficient allocation rules for the multiarmed bandit problem with multiple plays-part i: iid rewards. *IEEE Transactions on Automatic Control*, 32(11):968–976, 1987.

[3] J.-Y. Audibert, S. Bubeck, and G. Lugosi. Regret in online combinatorial optimization. *Mathematics of Operations Research*, 39(1):31–45, 2013.

[4] Y. Berstein, J. Lee, H. Maruri-Aguilar, S. Onn, E. Riccomagno, R. Weismantel, and H. Wynn. Nonlinear matroid optimization and experimental design. *SIAM Journal on Discrete Mathematics*, 22(3):901–919, 2008.

[5] J. Bragg, A. Kolobov, M. Mausam, and D. S. Weld. Parallel task routing for crowdsourcing. In *Second AAAI Conference on Human Computation and Crowdsourcing*, 2014.

[6] S. Bubeck and N. Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5(1):1–222, 2012.

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

[8] W. Chen, Y. Wang, and Y. Yuan. Combinatorial multi-armed bandit: General framework and applications. In *Proceedings of the 30th International Conference on Machine Learning (ICML)*, pages 151–159, 2013.

[9] A. Clark, L. Bushnell, and R. Poovendran. On leader selection for performance and controllability in multi-agent systems. In *Proceedings of the 51st Annual Conference on Decision and Control (CDC)*, pages 86–93, 2012.

[10] R. Combes, M. S. Talebi, A. Proutiere, and M. Lelarge. Combinatorial bandits revisited. In *Advances in Neural Information Processing Systems 28 (NIPS)*, pages 2107–2115, 2015.

[11] J. Edmonds. Matroids and the greedy algorithm. *Mathematical programming*, 1(1):127–136, 1971.

[12] Y. Gai, B. Krishnamachari, and R. Jain. Learning multiuser channel allocations in cognitive radio networks: A combinatorial multi-armed bandit formulation. In *IEEE Symposium on New Frontiers in Dynamic Spectrum (DySPAN)*, pages 1–9, 2010.

[13] Y. Gai, B. Krishnamachari, and R. Jain. Combinatorial network optimization with unknown variables: Multi-armed bandits with linear rewards and individual observations. *IEEE/ACM Transactions on Networking*, 20(5):1466–1478, 2012.

[14] A. Garivier and O. Cappé. The kl-ucb algorithm for bounded stochastic bandits and beyond. In *Proceedings of the 24th Annual Conference on Learning Theory (COLT)*, pages 359–376, 2011.

[15] J. C. Gittins. Bandit processes and dynamic allocation indices. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 148–177, 1979.

[16] T. L. Graves and T. L. Lai. Asymptotically efficient adaptive choice of control laws in controlled markov chains. *SIAM J. Control and Optimization*, 35(3):715–743, 1997.

[17] J. Komiyama, J. Honda, and H. Nakagawa. Optimal regret analysis of thompson sampling in stochastic multi-armed bandit problem with multiple plays. In *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, pages 1152–1161, 2015.

[18] B. Kveton, Z. Wen, A. Ashkan, and H. Eydgahi. Matroid bandits: Practical large-scale combinatorial bandits. In *Proceedings of AAAI Workshop on Sequential Decision-Making with Big Data*, 2014.

[19] B. Kveton, Z. Wen, A. Ashkan, H. Eydgahi, and B. Eriksson. Matroid bandits: Fast combinatorial optimization with learning. In *Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 420–429, 2014.

[20] B. Kveton, Z. Wen, A. Ashkan, and C. Szepesvari. Tight regret bounds for stochastic combinatorial semi-bandits. In *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 535–543, 2015.

[21] B. Kveton, Z. Wen, A. Ashkan, and M. Valko. Learning to act greedily: Polymatroid semi-bandits. *arXiv preprint arXiv:1405.7752*, 2014.

[22] T. L. Lai and H. Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6(1):4–22, 1985.

[23] M. Lelarge, A. Proutiere, and M. S. Talebi. Spectrum bandit optimization. In *Proceedings of Information Theory Workshop (ITW)*, pages 34–38, 2013.

[24] F. Lin, M. Fardad, and M. R. Jovanović. Algorithms for leader selection in large dynamical networks: Noise-corrupted leaders. In *Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC)*, pages 2932–2937, 2011.

[25] S. Onn. Convex matroid optimization. *SIAM Journal on Discrete Mathematics*, 17(2):249–253, 2003.

[26] J. G. Oxley. *Matroid theory*, volume 3. Oxford university press, 2006.

[27] H. Robbins. Some aspects of the sequential design of experiments. In *Herbert Robbins Selected Papers*, pages 169–177. Springer, 1985.

[28] A. Schrijver. *Combinatorial Optimization: Polyhedra and Efficiency*. Springer, 2003.

[29] M. Streeter, D. Golovin, and A. Krause. Online learning of assignments. In *Advances in Neural Information Processing Systems 22 (NIPS)*, pages 1794–1802, 2009.

[30] M. S. Talebi and A. Proutiere. An optimal algorithm for stochastic matroid bandit optimization. *arXiv preprint*, 2016.