Robust Matroid Bandit Optimization against Adversarial Contamination *

Youming Tao[1,2], Xiuzhen Cheng[1], Falko Dressler[2], Zhipeng Cai[3], Dongxiao Yu[1]**

[1] Shandong University, P.R. China
[2] TU Berlin, Germany
[3] Georgia State University, United States
{tao, dressler}@ccs-labs.org, {dxyu,xzcheng}@sdu.edu.cn,
zcai@gsu.edu

Abstract. In this paper, we consider the matroid bandit optimization problem, a fundamental and widely applicable framework for combinatorial multi-armed bandits where the action space is constrained by a matroid. We tackle the challenge of devising algorithms that can cope with adversarial contamination of the feedback rewards, which may severely degrade the performance or even mislead existing methods. Our main contribution is an efficient and robust algorithm, dubbed ROMM, which builds upon the idea of optimistic matroid maximization and leverages robust statistical techniques to estimate the quality of the base arms in polynomial time. We establish lower bounds for matroid bandit optimization under the ϵ-contamination model we adopt and show that ROMM achieves near-optimal regret bounds up to polylogarithmic factors. Furthermore, our analysis unveils a sharp phase transition between the small contamination regime and the large contamination regime for matroid bandit optimization. We establish that our algorithm can tolerate up to a universal constant fraction of corrupted feedbacks, which is optimal under mild conditions.

Keywords: Multi-Armed Bandits · Combinatorial Optimization · Matroids · Contamination Robustness.

1 Introduction

Combinatorial optimization is a classical field that has numerous practical applications, such as resource allocation [30] and network routing [10]. Modern combinatorial optimization problems are often so massive that even mildly polynomial-time solutions are infeasible. Luckily, many significant problems, such as finding a minimum spanning tree, have greedy solutions. Such problems can be often formulated as optimization on a *matroid* [22], a combinatorial structure that encapsulates the concept of independence

* Y. Tao was supported in part by the National Science Foundation of China (NSFC) under Grant 623B2068 and the China Scholarship Council (CSC) under Grant 202306220153. D. Yu and X. Cheng were supported in part by the Major Basic Research Program of Shandong Provincial Natural Science Foundation under Grant ZR2022ZD02. D. Yu was also supported in part by the National Natural Science Foundation of China (NSFC) under Grant 62122042.
** Corresponding Author.

and is intimately related to computational tractability. In particular, it is well established that the maximum of a modular function subject to a matroid constraint can be obtained greedily if and only if all feasible solutions are the independent sets of a matroid [25]. Matroids are pervasive in practice because they generalize many forms of independence, such as linear independence and forests in graphs.

In this paper, we consider a more realistic setting of learning how to maximize a *stochastic* modular function on a matroid. The modular function is represented as the sum of the weights of up to K items selected from the ground set E of a matroid, where E has totally N items. The weights of the items are unknown and each item $a \in E$ is associated with an unbounded σ-sub-Gaussian distribution $\mathcal{D}_a$ such that we only have access to a random sample from $\mathcal{D}_a$ each time when selecting the item a. These $\mathcal{D}_a$'s are initially unknown and we learn them by interacting repeatedly with the environment.

Many real-world optimization problems can be modeled in our setting, such as building a spanning tree for network routing [10]. When the delays on the links of the network are stochastic and their distribution is known, this problem can be solved by finding a minimum spanning tree. However, when the distribution is unknown, we need to learn it by observing the delays on different links over time. This introduces a trade-off between exploration and exploitation, which is characteristic for *stochastic multi-armed bandits* [2], a class of online learning problems where a learner optimizes its actions by receiving noisy feedback from the environment. Inspired by this, we call the problem of stochastic combinatorial optimization on a matroid as *matroid bandit optimization*.

1.1 Closely Related Works

Matroid bandit optimization is a significant problem that has received considerable attention in the past decade, and various facets of this problem have been investigated, such as regret minimization [14, 26, 11], pure exploration [7], algorithm efficiency [23] and differential privacy guarantees [5]. However, all of the existing works assume the true feedback obtained sampled from the underlying distribution. In real applications, it is common and inevitable that some bandit feedbacks are corrupted due to either malfunction or attacks from adversaries [21]. Thus, these bandit algorithms will face contaminated arm feedbacks and their learning utility will deteriorate. The recent years have indeed witnessed a renewed interest in devising bandit algorithms that are robust to data corruption, including [21, 13, 4, 28] for classical multi-armed bandits and [29] for general combinatorial bandits. Moreover, although matroid bandits is a special case of combinatorial bandits, the algorithms for general combinatorial optimization are not appropriate for matroid bandits, because the algorithm complexity is too high, since by exploiting greedy strategy, we can achieve much more efficient solutions for matroid bandits [14]. Therefore, it is still an open question how to design efficient matroid bandit optimization algorithms that offer robustness against adversarial contamination.

1.2 Our Contributions

In this paper, we concentrate on the robustness aspect of matroid bandit optimization and study the regret minimization for matroid bandit under the ϵ-*contamination model*,

where ϵ fraction of the feedbacks are assumed to be corrupted by an arbitrary adversary. Our main contributions can be outlined as follows:

- We propose a robust algorithmic framework for matroid bandit optimization, named ROMM, which leverages the idea of optimistic matroid maximization and utilizes a generic robust mean estimation sub-routine RME. We present the formal requirement for RME in general and also offer several concrete implementations of RME.
- We establish both the instance-dependent and instance-independent regret lower bounds for matroid bandit optimization under the ϵ-contamination model. This characterizes the limit of regret with respect to the contamination level ϵ. As a byproduct, our result also implies the first instance-independent lower bound for matroid bandit without contamination, which was left as an open problem in [14].
- Via theoretical analysis, we demonstrate that ROMM achieves near-optimal regret bounds, in both instance-dependent and instance-independent forms. Our results disclose a sharp phase transition between the small contamination regime and the large contamination regime. Intuitively speaking, when ϵ is smaller than the minimum suboptimality gap $\Delta_{\min}$, ROMM can still attain sub-linear regret as in the non-contamination setting, while for larger ϵ that is bounded by a universal constant $\frac{1}{4}$, the regret scales linearly with T.

Due to space limit, all the technical lemmas and proofs are included in Appendix.

2 Preliminaries

2.1 Combinatorial Optimization over a Matroid

Consider a pair $\mathcal{M} = (E, \mathcal{I})$, where $E = \{1, \cdots, N\}$ is a finite set of N items, and $\mathcal{I}$ is a family of subsets of E. We call E the ground set, and any subset $A \subseteq E$ is said to be independent if $A \in \mathcal{I}$.

Definition 1 (Matroid [22]). *A pair $\mathcal{M}(E, \mathcal{I})$ is said to be a matroid if the following properties hold:*

1. *The empty set is independent, i.e., $\emptyset \in \mathcal{I}$,*
2. *Every subset of an independent set is independent, i.e., for all $A \in \mathcal{I}$, if $A' \subset A$, then $A' \in \mathcal{I}$,*
3. *If A and A' are two independent sets, and $|A| > |A'|$, then there exists some item $a \in A \backslash A'$ such that $A' \cup \{a\} \in \mathcal{I}$.*

We say an independent set $A \in \mathcal{I}$ is a **basis** of $\mathcal{M}$ if A is maximal in $\mathcal{I}$, i.e., A is not a proper subset of any other independent set in $\mathcal{I}$. In other words, if $A \in \mathcal{I}$ is a basis of $\mathcal{M}$, then $A \cup \{a\} \notin \mathcal{I}$ for $\forall a \in E \backslash A$. Let $\mathcal{B}$ be the set of all bases of $\mathcal{M}$. It is well-known that, all bases of a matroid have the same cardinality [22], which is referred to as the **rank** of a matroid. We denote the rank by K, i.e., for $\forall A \in \mathcal{B}$, $|A| = K$.

In a typical combinatorial optimization problem over a matroid, each item $a \in E$ is associated with a non-negative weight μ_a and we denote by $\mu \in (\mathbb{R}^+)^N$ the vector of all N items' weights, i.e., $\mu = (\mu_1, \mu_2, \cdots, \mu_N)$. The optimization goal is to find an optimal basis A^* that has the maximum total weight, i.e.,

$$A^* \in \arg\max_{A \in \mathcal{B}} \sum_{a \in A} \mu_a, \tag{1}$$

which can be solved efficiently by using the greedy strategy described in Algorithm 1.

Algorithm 1: The greedy strategy for finding a maximum-weight basis

Input: Matroid $\mathcal{M} = (E, \mathcal{I})$
Initialize: $A^* \leftarrow \emptyset$
Let $a_1, \cdots, a_N$ be an ordering of base arms such that: $\mu_{a_1} \geq \cdots \geq \mu_{a_N}$
for $i = 1, \cdots, N$ **do**
  if $A^* \cup \{a_i\} \in \mathcal{I}$ then $A^* \leftarrow A^* \cup \{a_i\}$

2.2 Matroid Bandit Optimization

In matroid bandit optimization there is an learner interacting with a matroid bandit instance sequentially over T rounds. Each item $a \in E$ is now called a **base arm** and is associated with an underlying unknown feedback distribution $\mathcal{D}_a$ rather than a fixed weight. And each basis $A \in \mathcal{B}$ is now called a **super arm**. For $\forall a \in E$ and $t \in [T]$, we let $x_a(t)$ be the stochastic feedback generated from $\mathcal{D}_a$ by base arm a in round t. In this paper, we assume that each $\mathcal{D}_a$ is an unbounded σ-sub-Gaussian distribution with mean μ_a, which is more general than previously used bounded distribution in matroid bandit works, e.g., [14, 26]. For each base arm a, the sequence $\{x_a(t)\}_{t=1}^T$ is i.i.d., while in each round t, $\{x_a(t)\}_{a=1}^N$ can be arbitrarily correlated across base arms. Furthermore, we consider the semi-bandit feedback [1]. At the beginning of each round t, the learner selects a super arm $A(t) \coloneqq \{a_1(t), a_2(t), \cdots, a_K(t)\} \in \mathcal{B}$ to pull. Then he obtains a reward $r(t) \coloneqq \sum_{a \in A(t)} x_a(t)$ and observes $\{(a, x_a(t)) | a \in A(t)\}$, i.e., the specific feedback from each chosen base arm. Our goal is to equip the learner with a learning algorithm or policy π to maximize the expected cumulative reward he obtained over the time horizon T. Equivalently, we usually aim at minimizing the expected cumulative **regret** $\mathcal{R}_T$ after T rounds with $\mathcal{R}_T$ defined as follows:

$$\mathcal{R}_T \coloneqq \mathbb{E}\left[\sum_{t=1}^{T}\left(\sum_{a \in A^\star} x_a(t) - \sum_{a \in A(t)} x_a(t)\right)\right], \tag{2}$$

where $A^\star \coloneqq \arg\max_{A \in \mathcal{B}} \sum_{a \in A} \mu_a$ is the optimal super arm that has maximum expected reward, and the expectation is taken with respect to all the randomness. Without loss of generality, we let $A^\star \coloneqq \{a_1^\star, \cdots, a_K^\star\}$, where the base arms are ordered such that $a_k^\star$ is the base arm with the k-th highest expected feedback, i.e., $\mu_{a_1^\star} \geq \cdots \geq \mu_{a_K^\star}$. We say a base arm a is sub-optimal if it belongs to $\overline{A^\star} \coloneqq E \backslash A^\star$. Accordingly, each base arm in $A^\star$ is called optimal base arm. For any pair of sub-optimal $a \in \overline{A^\star}$ and optimal base arm $a_k^\star \in A^\star$, we define the **gap** between them as:

$$\Delta_{a,k} \coloneqq \mu_{a_k^\star} - \mu_a. \tag{3}$$

For each sub-optimal base arm $a \in \overline{A^\star}$, we define a set:

$$\mathcal{H}_a \coloneqq \{k : \Delta_{a,k} > 0\}, \tag{4}$$

which contains the indices of optimal base arms in $A^\star$ whose expected feedback is higher than that of a. The cardinality of $\mathcal{H}_a$ is denoted by H_a, i.e., $H_a = |\mathcal{H}_a|$. For each sub-optimal base arm a, we use the smallest pairwise gap associated with a to define the sub-optimality gap of a, i.e., $\Delta_a \coloneqq \Delta_{a,H_a}$.

2.3 Contamination Model

We consider a scenario where the learner does not observe the true feedbacks generated by each base arm, but some corrupted versions of them instead. Specifcially, for any base arm $a \in E$ and any round index t, an adversary can modify the feedback from $x_a(t)$ to an arbitrary value $y_a(t)$. We assume that the adversary is restricted by a corruption rate ϵ, where $0 < \epsilon < 1$, which means that the adversary can only corrupt up to ϵ fraction of the feedbacks for any base arm a until any round t. Formally, the learner's observations $\{y_a(t)\}_{t=1}^T$ satisfy

$$\frac{\sum_{i=1}^t \{\mathbb{1}_{x_a(i) \neq y_a(i)}\}}{t} \leq \epsilon, \ \text{ for } \forall t \in [T]. \tag{5}$$

This contamination model is known as ϵ-contamination model, and has garnered considerable attention in the literature, e.g., [6, 19] and the related references therein. In addition, the ϵ-Huber contamination model [12], a widely used robust statistics model, is a special case of the ϵ-contamination model (see [15, 24]). The ϵ-contamination model defined in eq. (5) is very general and permits the adversary to corrupt the feedbacks in any manner, as long as the fraction of corrupted feedbacks does not surpass ϵ. Moreover, the adversary is retrospective, meaning that the adversary can act adaptively and exploit the learner's past, present, and future feedbacks. This implies that the adversary can alter its strategy across different actions and make the learning problem more formidable.

In this paper, our aim is to design a robust learning algorithm for matroid bandit optimization under the contamination model given by eq. (5). We note that, even though the learner's feedbacks are corrupted, its actual reward gains are still based on the true feedbacks. Therefore, our objective is still to minimize the regret defined by eq. (2).

3 Robust Matroid Bandit Optimization Framework: ROMM

3.1 Framework Design

Our framework can be seen as a robust variant of the Optimistic Matroid Maximization (OMM) algorithm developed in [14], which is designed based on the optimistic principle in the face of uncertainty [20]. Therefore, we refer to our framework as Robust Optimistic Matroid Maximization (ROMM).

The rough idea of OMM is to adapt the greedy strategy for finding a maximum-weight basis of a matroid (Algorithm 1) to the stochastic setting. In particular, in each

Algorithm 2: ROMM Framework

Input: Time horizon T, contamination fraction ϵ, sub-Gaussian constant σ, an (ϵ, δ)-robust mean estimator RME
for *each base arm* $a \in E$ **do**
  Pull base arm a and observe $y_a(0)$.
  Set $T_a(0) \leftarrow 1$.
for $t = 1, \cdots, T$ **do**
  for *each base arm* $a \in E$ **do**
    Compute robust feedback mean estimate $\widehat{\mu}_a \leftarrow \texttt{RME}(\{y_a(i)\}_{i=1}^{T_a(t-1)})$.
    $U_a(t) \leftarrow \widehat{\mu}_a + \frac{\sigma}{1-2\epsilon}\sqrt{\frac{4\log(t)}{T_a(t-1)}}$.
  Let $a_1, \cdots, a_N$ be a sorted sequence of base arms such that $U_{a_1}(t) \geq \cdots \geq U_{a_N}(t)$
  $A(t) \leftarrow \emptyset$
  for $i = 1, \cdots, N$ **do**
    if $A(t) \cup \{a_i\} \in \mathcal{I}$ then $A(t) \leftarrow A(t) \cup \{a_i\}$.
  Pull $A(t)$, obtain the reward $r(t) = \sum_{a \in A(t)} x_a(t)$, and observe corrupted feedbacks $\{y_a(t)\}_{a \in A(t)}$.
  $T_a(t) \leftarrow T_a(t-1) + 1$, for all $a \in A(t)$.

round t, one only needs to substitute the weight/expected feedback μ_a of each item/base arm a with its optimistic upper confidence bound (UCB) estimate $U_a(t)$, i.e., the sum of the estimated feedback mean and its confidence interval. When there is no contamination, it has been shown that the simple empirical mean suffices to achieve the order optimal regret. However, empirical mean is extremely susceptible to the interference of outliers. Even a single outlier is enough to deviate the empirical mean arbitrarily. Therefore, when contamination exists, the empirical mean will no longer provide any meaningful estimation guarantees.

In existence of adversarial contamination, the key to successfully handling potentially corrupted feedback is to replace the empirical mean with other, more robust, estimators of the mean. All we need is a mean estimator with the following property, which we term as (ϵ, δ)-robust mean estimator

Definition 2 ((ϵ, δ)-robust mean estimator)). *Let S be the set of samples $z_1, \cdots, z_n \in \mathbb{R}$ that are drawn from a σ-sub-Gaussian distribution with mean μ. Let S_C be the contaminated variant of S where ϵ fraction of samples are contaminated by an adversary. For $\epsilon < \frac{1}{2}$, $0 < \delta < 1$, an (ϵ,δ)-robust mean estimator RME guarantees with probability at least $1 - \delta$ that*

$$|\mathit{RME}(S_n) - \mu| \leq I(\epsilon, \delta, n) \coloneqq C \cdot \frac{\sigma}{1-2\epsilon}\left(\sqrt{\frac{\log\frac{1}{\delta}}{n}} + \epsilon\sqrt{\log\frac{1}{\delta}}\right), \tag{6}$$

where C is a universal numerical constant independent of ϵ, δ and n.

We will provide some robust mean estimators satisfying Definition 2 in the following subsection. We are now ready to introduce our ROMM framework for matroid bandit optimization under ϵ-contamination model, which is depicted in Algorithm 2. In a nutshell, in ROMM, to find the (potentially) best super arm in each round t, the learner follows four steps: (1) calculates the upper confidence bound $U_a(t)$ for each base arm $a \in E$ [line 7-8]; (2) orders all the base arms from the highest to the lowest according to their UCB values [line 9]; (3) selects them into $A(t)$ greedily according to this order [line 10-12]; (4) the learner pulls $A(t)$ and observes each newly generated feedback $x_a(t)$ [line 13-14]. It is noteworthy that in our algorithm framework, UCB does not strictly follow the mean estimate plus its confidence interval I. This is because there is an additive term in the RME confidence interval I that is independent of the sample size n and is the same for all base arms. Since what we actually care about is the relative size of UCB of each base arm, we omit this additive term in the algorithm.

3.2 Theoretical Results

Let $\Delta_{\min} \coloneqq \min_{a \in \overline{A^\star}} \Delta_a$. We first consider a setting with small contamination and analyse the performance of ROMM therein.

Theorem 1 (Instance-Dependent Regret Upper Bound for Small ϵ). *For small contamination regime where $\epsilon \leq \frac{\Delta_{min}}{4\Delta_{min}+4\sqrt{3}C\sigma\sqrt{\log T}}$, the instance-dependent expected cumulative regret of ROMM is at most*

$$\mathcal{R}_T \leq 96C^2\sigma^2 \sum_{a \in \overline{A^\star}} \frac{\log T}{\Delta_a} + \frac{\pi^2}{6} \sum_{a \in \overline{A^\star}} \sum_{k=1}^{H_a} \Delta_{a,k}.$$

Theorem 2 (Instance-Independent Regret Upper Bound for Small ϵ). *For small contamination regime where $\epsilon \leq \frac{\Delta_{min}}{4\Delta_{min}+4\sqrt{3}C\sigma\sqrt{\log T}}$, the instance-independent expected cumulative regret of ROMM is at most*

$$\mathcal{R}_T \leq 4\sqrt{6}C\sigma\sqrt{(N-K)KT\log T} + \frac{\pi^2(N-K)K}{6}.$$

Remark 1. The above Theorem 1 and Theorem 2 establish the regret guarantees for the small contamination regime where the contamination proportion ϵ is required to be smaller than a problem instance gap determined threshold, i.e., $\epsilon \leq \frac{\Delta_{\min}}{4\Delta_{\min}+4\sqrt{3}C\sigma\sqrt{\log T}}$. Recall that, for the standard non-contamination matroid bandits, [14] has proven the instance-dependent and instance-independent regret upper bounds of $O(\sum_{a \in \overline{A^\star}} \frac{\log T}{\Delta_a} + \sum_{a \in \overline{A^\star}} \sum_{k=1}^{H_a} \Delta_a)$ and $O(\sqrt{(N-K)KT\log T} + (N-K)K)$, respectively. Compared with the non-contamination bounds, we can see that, our framework ROMM does not incur any additional price for withstanding the adversarial corruptions and achieves the same order of regret as in the standard non-contamination case.

Though consistent with the non-contamination results, our bounds above do not allow ϵ to be too big relative to the minimum suboptimality gap $\Delta_{\min}$. Such kind of bound on the contamination proportion ϵ is very common in robust learning and robust

statistics literature and represents the *breakdown* point the algorithm. Moreover, if $\epsilon > \Omega(\Delta_{\min})$, ROMM will incur a linear regret with respect to T. This is natural, since when ϵ gets large, it also harder to distinguish between the base arms. In the next section, we will show that no algorithm can get sub-linear regret since distinguishing between the top two actions will become impossible even with infinite samples. Before that, we now put a much milder restriction on ϵ, and derive a more general regret upper bounds for any ϵ that is at most a universal constant of $\frac{1}{4}$.

Theorem 3 (Instance-Dependent Regret Upper Bound). *If $\epsilon \leq \frac{1}{4}$, the instance-dependent expected cumulative regret of ROMM is at most*

$$\mathcal{R}_T \leq 96C^2\sigma^2 \sum_{a \in \overline{A^\star} \cap \mathcal{S}} \frac{\log T}{\Delta_a} + \frac{\pi^2}{6} \sum_{a \in \overline{A^\star} \cap \mathcal{S}} \sum_{k=1}^{H_a} \Delta_{a,k} + \frac{4\sqrt{3}C\sigma K\epsilon}{1-4\epsilon} T\sqrt{\log T}.$$

Theorem 4 (Instance-Independent Regret Upper Bound). *If $\epsilon \leq \frac{1}{4}$, the instance-independent expected cumulative regret of ROMM is at most*

$$\mathcal{R}_T \leq 4\sqrt{6}C\sigma\sqrt{(N-K)KT\log T} + \frac{\pi^2(N-K)K}{6} + \frac{4\sqrt{3}C\sigma K\epsilon}{1-4\epsilon} T\sqrt{\log T}.$$

Remark 2. The above Theorem 3 and Theorem 4 illustrates that, in general, ROMM incurs a linear regret term of $O(\frac{\epsilon}{1-\epsilon}KT\log(T))$ in both the instance-dependent and instance-independent bound for any $0 < \epsilon < \frac{1}{4}$. The linear term in the regret may be acceptable if the contamination proportion ϵ is not very large. Moreover, we can see that, when $\epsilon = 0$ (i.e., there is no contamination), the linear term vanishes and ROMM recovers the state-of-the-art non-contamination regret derived in [14].

Remark 3. Note that the classical stochastic multi-armed bandit (MAB) is a special case of matroid bandit with $K = 1$. When $K = 1$, our instance-independent bound in Theorem 4 recovers the state-of-the-art bound of $\widetilde{O}(\sqrt{NT} + \frac{\epsilon}{1-4\epsilon}T)$ for ϵ-contaminated MAB in [21]. That is, our bound can be seen as a generalization of the previous MAB bound to the matroid bandit case.

3.3 Concrete instantiations of (ϵ, δ)-robust mean estimator

A key component of the ROMM framework is the (ϵ, δ)-robust mean estimator RME, which plays a crucial role for tolerating corrupted feedbacks. In Definition 2, we have provided the property that an (ϵ, δ)-robust mean estimator should satisfy in general. But it still remains how to implement an (ϵ, δ)-robust mean estimator when using the ROMM framework in practice. Actually, our definition for RME is quite generic and many robust mean estimators in robust statistics can be used as RME. Here we give three concrete instantiations of (ϵ, δ)-robust mean estimator.

- *Median (Med) [15]:* Find the median of all the sample points.
- *Trimmed Mean (TM) [18]:* Trim the smallest and largest ϵ fraction of points from the sample and calculate the mean of the remaining points.

– *Shorth Mean (SM) [21]:* Take the mean of the shortest interval that removes the smallest ϵ_1 and largest ϵ_2 fraction of points such that $\epsilon_1 + \epsilon_2 = \epsilon$, where ϵ_1 and ϵ_2 is chosen to minimize the interval length of remaining points.

The proof for how the above estimators satisfy Definition 2 can be found in the above referenced papers, thus we omit the proof here.

4 Lower Bounds

In this section, we establish both instance-dependent and instance-independent lower bounds for matroid bandit optimization under the ϵ-contamination model.

We start by introducing a special class of matroid bandit instances called partition matorid bandits, which is also used in [14]. Let $P_1, P_2, \cdots, P_K$ be a partition of the ground set E, such that,

$$\bigcup_{k=1}^{K} P_k = E, \text{ and } P_i \bigcap P_j = \emptyset \text{ for } \forall i, j \in [K]. \tag{7}$$

The family of independent sets is defined as

$$\mathcal{I} = \{I \subseteq E : |I \cap P_k| \leq 1, \forall k \in [K]\}. \tag{8}$$

Then $\mathcal{M} = (E, \mathcal{I})$ is a partition matroid of rank K. For the feedback generation, we consider the Bernoulli distribution with means that lie in the interval $(0, 1)$. Specifically, we set the mean of each base arm $a \in P_k$, $k \in [K]$ as follows:

$$\mu_a = \begin{cases} \frac{1}{2}, & a = \min_{i \in P_k} i, \\ \frac{1}{2} - g_a, & \text{otherwise}, \end{cases} \tag{9}$$

where $0 < g_a < \frac{1}{2}$ and the optimal base arm in each partition is the item with the smallest index, i.e., $\min_{i \in P_k} i$, and the gap of each base arm a is just g_a, i.e., $\Delta_a = g_a$.

To prove the lower bounds under ϵ-contamination model, we develop a new hard instance ξ, which basically comes from the instance used in [14] but we add more restriction so that we can obtain tight instance-independent lower bound. In ξ, we let each of $P_1, P_2, \cdots, P_K$ contain the same number of base arms, i.e.,

$$|P_1| = |P_2| = \cdots = |P_K| = N/K. \tag{10}$$

Without loss of generality, we assume that N/K is an integer. The key observation for proving the instance-independent lower bound is that our problem is equivalent to K N/K-armed Bernoulli bandit. With this perception, we first study the lower bound incurred by one of these N/K-armed Bernoulli bandit.

Our main idea for proving the instance-dependent lower bound is to decompose the regret into a weighted sum of the expected pulling number of all sub-optimal base arms and then show that no algorithm can achieve low pulling number for each sub-optimal arm i on ν and ν^i simultaneously. For this, we consider **consistent** learning algorithms:

Definition 3. *Denote the number of times that a base arm a is chosen in T rounds by $N_a(T)$. An algorithm π is called consistent if for any sub-optimal base arm a, the expected number of times that a is pulled by π is sub-polynomial in T for any stochastic matroid bandit instance, i.e., $\mathbb{E}[N_a(T)] \leq o(T^c)$ for any $0 < c < 1$.* [4]

Intuitively, the consistency defined above requires that the algorithm achieves sub-polynomial regret over all problem instances. Any inconsistent algorithm performs poorly on some instances and extremely well on others, which makes it difficult to prove good instance-dependent lower bounds for inconsistent algorithms. Thus, the consistent algorithm class is considered to be reasonable and has been used for lower bound analysis in many previous bandit literature [14, 3, 9, 27]. Fix a partition P_k and a sub-optimal base arm $\overline{a}_i \in P_k$. We denote the original instance of this N/K-armed Bernoilli bandit as ν. Then we define another instance ν^i where all the setting is the same as ν except that the mean of the $\overline{a}_i$ is increased by $2\Delta_{\overline{a}_i}$.

Lemma 1. *For any fixed partition P_k and any consistent matroid bandit optimization algorithm, there exists a N/K Bernoulli bandit instance for P_k and an adversary with contamination fraction ϵ such that the expected regret incurred from P_k, denoted as $\mathcal{R}_{T,P_k}$, is at least*

$$\mathcal{R}_{T,P_k} \geq \Omega\left(\sum_{a \in \overline{A^\star} \cap P_k} \frac{\log T}{\Delta_a} + \frac{\epsilon}{1-\epsilon} T\right). \tag{11}$$

Theorem 5 (Instance-Dependent Lower Bound). *For any consistent matroid bandit optimization algorithm, there exists a matroid bandit instance and an adversary with contamination fraction ϵ such that the expected regret $\mathcal{R}_T$ is at least*

$$\mathcal{R}_T \geq \Omega\left(\sum_{a \in \overline{A^\star}} \frac{\log T}{\Delta_a} + \frac{K\epsilon}{1-\epsilon} T\right). \tag{12}$$

Next we give the instance-independent lower bound, which is also called *min-max* lower bound in some literature. This lower bound characterizes the information-theoretic limit of regret with respect to contamination level.

Theorem 6 (Instance-Independent Lower Bound). *For any matroid bandit optimization algorithm, there exists a partition matroid bandit instance and an adversary with contamination fraction ϵ such that the expected regret $\mathcal{R}_T$ is at least*

$$\mathcal{R}_T \geq \Omega\left(\sqrt{(N-K)KT} + \frac{K\epsilon}{1-\epsilon} T\right). \tag{13}$$

Remark 4. When $\epsilon = 0$, Theorem 6 establishes the instance-independent lower bound for non-contamination matroid bandit optimization, which solves the open problem left in [14].

[4] Without loss of generality, we let $c = \frac{3}{4}$ in this paper.

Remark 5. Theorem 5 and Theorem 6 indicates that a linear term w.r.t. T in the regret is genuinely unavoidable for matroid bandit under the ϵ-contamination model. As a result, the attained upper bounds in Theorem 3 and Theorem 4 are nearly optimal with respect to the dominant term T up to $poly(\log T)$ factors.

5 Conclusion

In this paper, we studied the problem of matroid bandit optimization under the ϵ-contamination model, where a fraction of the feedbacks are corrupted by an arbitrary adversary. We proposed a robust algorithmic framework, `ROMM`, which leverages robust mean estimation techniques to cope with the adversarial perturbations. We established both instance-dependent and instance-independent regret lower bounds for the problem and showed that `ROMM` achieves near-optimal regret bounds up to polylogarithmic factors. We also revealed a sharp phase transition between the small contamination regime and the large contamination regime, where the regret behavior changes drastically. We conducted extensive experiments on synthetic and real-world datasets to demonstrate the effectiveness and robustness of our algorithm compared with existing methods.

Our work opens up several interesting directions for future research. First, it would be interesting to extend our framework to other combinatorial structures beyond matroids that admit a greedy algorithm with a provable approximation guarantee, such as submodular functions or knapsack constraints. Second, it would be desirable to design more efficient and practical robust mean estimation algorithms that can handle high-dimensional or heavy-tailed distributions. Third, it would be worthwhile to explore other adversarial models for matroid bandit optimization, such as bandit feedback or adaptive adversaries.

References

1. Audibert, J.Y., Bubeck, S., Lugosi, G.: Regret in online combinatorial optimization. Mathematics of Operations Research **39**(1), 31–45 (2014)
2. Auer, P., Cesa-Bianchi, N., Fischer, P.: Finite-time analysis of the multiarmed bandit problem. Machine learning **47**(2), 235–256 (2002)
3. Basu, D., Dimitrakakis, C., Tossou, A.: Privacy in multi-armed bandits: Fundamental definitions and lower bounds. arXiv preprint arXiv:1905.12298 (2019)
4. Basu, D., Maillard, O.A., Mathieu, T.: Bandits corrupted by nature: Lower bounds on regret and robust optimistic algorithm. arXiv preprint arXiv:2203.03186 (2022)
5. Chandak, K., Hu, B., Hegde, N.: Differentially private algorithms for efficient online matroid optimization. In: Conference on Lifelong Learning Agents. pp. 66–88. PMLR (2023)
6. Charikar, M., Steinhardt, J., Valiant, G.: Learning from untrusted data. In: Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing. pp. 47–60 (2017)
7. Chen, L., Gupta, A., Li, J.: Pure exploration of multi-armed bandit under matroid constraints. In: Proceedings of the 29th Conference on Learning Theory (COLT). pp. 647–669 (2016)
8. Chen, M., Gao, C., Ren, Z.: Robust covariance and scatter matrix estimation under huber's contamination model. The Annals of Statistics **46**(5), 1932–1960 (2018)
9. Chen, X., Zheng, K., Zhou, Z., Yang, Y., Chen, W., Wang, L.: (locally) differentially private combinatorial semi-bandits. In: Proceedings of the 37th International Conference on Machine Learning (ICML). pp. 1757–1767 (2020)

10. Gallager, R.: A minimum delay routing algorithm using distributed computation. IEEE transactions on communications **25**(1), 73–85 (1977)
11. Huang, Z., Xu, Y., Hu, B., Wang, Q., Pan, J.: Thompson sampling for combinatorial semi-bandits with sleeping arms and long-term fairness constraints. arXiv preprint arXiv:2005.06725 (2020)
12. Huber, P.J.: Robust statistics, vol. 523. John Wiley & Sons (2004)
13. Kapoor, S., Patel, K.K., Kar, P.: Corruption-tolerant bandit learning. Machine Learning **108**(4), 687–715 (2019)
14. Kveton, B., Wen, Z., Ashkan, A., Eydgahi, H., Eriksson, B.: Matroid bandits: Fast combinatorial optimization with learning. In: Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence (UAI) (2014)
15. Lai, K.A., Rao, A.B., Vempala, S.: Agnostic estimation of mean and covariance. In: 2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS). pp. 665–674. IEEE (2016)
16. Lattimore, T., Szepesvári, C.: An information-theoretic approach to minimax regret in partial monitoring. In: Proceedings of the 32nd Conference on Learning Theory (COLT). pp. 2111–2139 (2019)
17. Lattimore, T., Szepesvári, C.: Bandit algorithms. Cambridge University Press (2020)
18. Liu, L., Li, T., Caramanis, C.: High dimensional robust estimation of sparse models via trimmed hard thresholding. arXiv preprint arXiv:1901.08237 (2019)
19. LUGOSI, G., MENDELSON, S.: Robust multivariate mean estimation: The optimality of trimmed mean. The Annals of Statistics **49**(1), 393–410 (2021)
20. Munos, R.: The optimistic principle applied to games, optimization and planning: Towards foundations of monte-carlo tree search. Foundations and Trends in Machine Learning **7**(1), 1–130 (2014)
21. Niss, L., Tewari, A.: What you see may not be what you get: Ucb bandit algorithms robust to $varepsilon$-contamination. In: Conference on Uncertainty in Artificial Intelligence. pp. 450–459. PMLR (2020)
22. Oxley, J.G.: Matroid theory, vol. 3. Oxford University Press, USA (2006)
23. Perrault, P., Perchet, V., Valko, M.: Exploiting structure of uncertainty for efficient matroid semi-bandits. In: Proceedings of the 36th International Conference on Machine Learning (ICML). pp. 5123–5132 (2019)
24. Prasad, A., Balakrishnan, S., Ravikumar, P.: A robust univariate mean estimator is all you need. In: International Conference on Artificial Intelligence and Statistics. pp. 4034–4044. PMLR (2020)
25. Schrijver, A., et al.: Combinatorial optimization: polyhedra and efficiency, vol. 24. Springer (2003)
26. Talebi, M.S., Proutiere, A.: An optimal algorithm for stochastic matroid bandit optimization. In: Proceedings of the 16th International Conference on Autonomous Agents & Multiagent Systems (AAMAS). pp. 548–556 (2016)
27. Tao, Y., Wu, Y., Zhao, P., Wang, D.: Optimal rates of (locally) differentially private heavy-tailed multi-armed bandits. In: Proceedings of the 25th International Conference on Artificial Intelligence and Statistics. pp. 1546–1574 (2022)
28. Wu, Y., Zhou, X., Tao, Y., Wang, D.: On private and robust bandits. Advances in Neural Information Processing Systems **36** (2024)
29. Xu, H., Li, J.: Simple combinatorial algorithms for combinatorial bandits: Corruptions and approximations. In: Uncertainty in Artificial Intelligence. pp. 1444–1454. PMLR (2021)
30. Zuo, J., Joe-Wong, C.: Combinatorial multi-armed bandits for resource allocation. In: Proceedings of the 55th Annual Conference on Information Sciences and Systems (CISS). pp. 1–4. IEEE (2021)

A Useful Facts for Matroid Bandits

Lemma 2 (Bijection [14]). *For the optimal matroid basis $A^\star$ and any chosen basis $A(t)$, there exists a bijection $\iota : A(t) \mapsto A^\star$ such that:*

$$\{a_1(t), \cdots, a_{k-1}(t), \iota(a_k(t))\} \in \mathcal{I}, \forall k = 1, \cdots, K. \tag{14}$$

In addition, $\iota(a_k(t)) = a_i^\star$ when $a_k(t) = a_i^\star$ for some $i \in [K]$.

Lemma 3 (Regret Decomposition [14]). *Define*

$$R_t = \sum_{a \in A^\star} x_a(t) - \sum_{a \in A(t)} x_a(t)$$

be the instant regret incurred by choosing $A(t)$ in round t. We have

$$\mathbb{E}[R_t] \leq \sum_{a \in \overline{A^\star}} \sum_{k=1}^{H_a} \Delta_{a,k} \mathbb{1}_{a,k}(t), \tag{15}$$

where the indicator function $\mathbb{1}_{a,k}(t)$ is defined as

$$\mathbb{1}_{a,k}(t) := \mathbb{1}\{\exists i : a_i(t) = a, \iota(a_i(t)) = a_k^\star\}. \tag{16}$$

Moreover,

$$\sum_{a \in \overline{A^\star}} \sum_{k=1}^{H_a} \mathbb{1}_{a,k}(t) \leq K, \forall t \in [T], \tag{17}$$

$$\sum_{k=1}^{H_a} \mathbb{1}_{a,k}(t) \leq 1, \forall t \in [T], a \in \overline{A^\star}. \tag{18}$$

Lemma 4 (Theorem 5.1 in [8]). *Let P_1 and P_2 be two distributions over any set $\mathcal{X}$. If for some $\epsilon \in [0, 1/2)$, we have that $\mathrm{TV}(P_1, P_1) = \frac{\epsilon}{1-\epsilon}$, then there exists two distributions Q_1 and Q_2 on the same probability space such that*

$$(1-\epsilon)P_1 + \epsilon Q_1 = (1-\epsilon)P_2 + \epsilon Q_2. \tag{19}$$

Lemma 5 (Hoeffding's inequality). *Let $Z_1, \ldots, Z_n$ be independent bounded random variables with $Z_i \in [a, b]$ for all i, where $-\infty < a < b < \infty$. Then*

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}(Z_i - \mathbb{E}[Z_i])\right| \geq t\right) \leq 2\exp\left(-\frac{2nt^2}{(b-a)^2}\right)$$

Lemma 6 (Instance-Independent Lower Bound for Stochastic n-Multi-Armed Bandits [17, theorem15.2]). *There exists a stochastic n-armed bandit instance such that the expected regret of any algorithm is $\Omega\left(\sqrt{(n-1)T}\right)$.*

B Proof of Theorem 1

We denote by R_t the instant regret incurred by the super arm $A(t)$ in round t:

$$R_t \coloneqq \sum_{a \in A^\star} x_a(t) - \sum_{a \in A(t)} x_a(t). \tag{20}$$

Then, we have following upper bound on $\mathcal{R}_T$:

$$\begin{aligned}
\mathcal{R}_T &= \sum_{t=1}^{T} \mathbb{E}[R_t] && (21)\\
&\leq \sum_{t=1}^{T} \mathbb{E}\left[\sum_{a \in \overline{A^\star}} \sum_{k=1}^{H_a} \Delta_{a,k} \cdot \mathbb{1}_{a,k}(t) \right] && (22)\\
&= \sum_{a \in \overline{A^\star}} \sum_{k=1}^{H_a} \Delta_{a,k} \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}_{a,k}(t) \right], && (23)
\end{aligned}$$

where the inequality comes from (15) in Lemma 3. In the ROMM framework (Algorithm 2), we denote by $\widehat{\mu}_a(t)$ the estimation for μ_a at the end of round t, by $T_a(t)$ the total pull times of base arm a till the end of round t, and by $I(\epsilon, \delta, n)$ the confidence interval of RME, i.e., $C \cdot \frac{\sigma}{1-2\epsilon} \left(\sqrt{\frac{\log(\frac{1}{\delta})}{n}} + \epsilon \sqrt{\log(\frac{1}{\delta})} \right)$ for some constant C. With these notations, we define the following good events:

$$\Lambda_{t,a} \coloneqq \{ |\widehat{\mu}_a - \mu_a| \leq I(\epsilon, t^{-3}, T_a(t-1)) \} \quad \text{for} \quad \forall a \in [E], t \in [T]$$

By using the Hoeffding's inequality (Lemma 5), we know that,

$$\begin{aligned}
\mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}\{\overline{\Lambda_{t,a}}\} \right] &\leq \sum_{t=1}^{T} \sum_{s=1}^{t} \mathbb{P}\left(|\mu_a - \widehat{w}_a(t-1)| \geq I(\epsilon, t^{-4}, T_a(t-1)) \right) \\
&\leq \sum_{t=1}^{T} \sum_{s=1}^{t} t^{-3} \leq \sum_{t=1}^{T} t^{-2} \leq \frac{\pi^2}{6}.
\end{aligned}$$

By (23), we have

$$\mathcal{R}_T \leq \sum_{a \in \overline{A^\star}} \sum_{k=1}^{H_a} \Delta_{a,k} \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}_{a,k}(t) \cdot \mathbb{1}\{\Lambda_{t,a}\} \right] + \sum_{a \in \overline{A^\star}} \sum_{k=1}^{H_a} \Delta_{a,k} \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}_{a,k}(t) \cdot \mathbb{1}\{\overline{\Lambda_{t,a}}\} \right]. \tag{24}$$

The last term in (24) can be bounded directly as follows

$$\sum_{a \in \overline{A^\star}} \sum_{k=1}^{H_a} \Delta_{a,k} \mathbb{E}[\sum_{t=1}^{T} \mathbb{1}_{a,k}(t) \cdot \mathbb{1}\{\overline{\Lambda_{t,a}}\}] \leq \frac{\pi^2}{6} \sum_{a \in \overline{A^\star}} \sum_{k=1}^{H_a} \Delta_{a,k}, \tag{25}$$

In the remainder of the proof, We focus on the first term in (24). For any round t, a sub-optimal base arm, say a, is selected rather than its corresponding optimal counterpart $a^* \coloneqq \iota(a)$ only if

$$\mu_a + \frac{2\sigma}{1-2\epsilon}\sqrt{\frac{3\log(t)}{T_a(t-1)}} + \frac{\sigma}{1-2\epsilon}\epsilon\sqrt{3\log(t)} \geq U_a(t) \geq U_{a^*}(t) \geq \mu_{a^*} - \frac{\sigma}{1-2\epsilon}\sqrt{3\log(t)}. \tag{26}$$

Suppose $a^* = a_k^\star$ for some $k \in [K]$, then (26) means

$$\Delta_{a,k} \leq \frac{2\sigma}{1-2\epsilon}\left(\sqrt{\frac{3\log(t)}{T_a(t-1)}} + \epsilon\sqrt{3\log(t)}\right) = 2I(\epsilon, t^{-3}, T_a(t-1)). \tag{27}$$

Thus, to ensure that the algorithm always chooses a^* instead of a, it suffices to find the minimum $T_a(t-1)$ such that

$$\Delta_{a,k} > I(\epsilon, t^{-3}, T_a(t-1)). \tag{28}$$

If $\epsilon \leq \frac{\Delta_{a,H_a}}{4\Delta_{a,H_a}+4\sqrt{3}C\sigma\sqrt{\log(T)}}$, we obtain the following inequality by solving (28):

$$T_a(t-1) > \frac{48C^2\sigma^2\log(t)}{\Delta_{a,k}^2}. \tag{29}$$

Note that, when $\epsilon \leq \frac{\Delta_{\min}}{4\Delta_{\min}+4\sqrt{3}C\sigma\sqrt{\log(T)}}$, (29) holds for all $a \in E$. Let $\tau_{a,k}(t) = \frac{48C^2\sigma^2\log(t)}{\Delta_{a,k}^2}$. Then we can then say that, when $T_a(t-1) > \tau_{a,k}(t)$, the algorithm must choose $a_k^\star$ instead of a. In summary, we know that, when event $\Lambda_{t,a}$ holds, the algorithm incurs an instant regret of $\Delta_{a,k}$ by selecting a instead of $a_k^\star$ implying that $T_a(t-1) \leq \tau_{a,k}$. Based on this, we can bound the first term in the (24) as follows:

$$\sum_{a\in\overline{A^\star}}\sum_{k=1}^{H_a}\Delta_{a,k}\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{1}_{a,k}(t)\cdot\mathbb{1}\{\Lambda_{t,a}\}\right] \tag{30}$$

$$\leq \sum_{t=1}^{T}\sum_{a\in\overline{A^\star}}\sum_{k=1}^{H_a}\Delta_{a,k}\mathbb{E}[\mathbb{1}_{a,k}(t)\cdot\mathbb{1}\{T_a(t-1)\leq\tau_{a,k}(t)\}] \tag{31}$$

$$\leq \max\sum_{t=1}^{T}\sum_{a\in\overline{A^\star}}\sum_{k=1}^{H_a}\Delta_{a,k}\mathbb{1}_{a,k}(t)\mathbb{1}\{T_a(t-1)\leq\tau_{a,k}(t)\} \tag{32}$$

$$= \max\sum_{a\in\overline{A^\star}}\left(\sum_{k=1}^{H_a}\Delta_{a,k}\sum_{t=1}^{T}\mathbb{1}_{a,k}(t)\mathbb{1}\{T_a(t-1)\leq\tau_{a,k}(t)\}\right). \tag{33}$$

Denote $m_{a,k} = \sum_{t=1}^{T}\mathbb{1}_{a,k}(t)\mathbb{1}\{T_a(t-1)\leq\tau_{a,k}(t)\}$. Note that:

1. the gaps are ordered such that $\Delta_{a,1} \geq \cdots \geq \Delta_{a,H_a}$ (and thus $\tau_{a,1} \leq \cdots \leq \tau_{a,H_a}$),
2. the counter $T_a(t)$ increases by at most 1 when $\mathbb{1}_{a,k}(t) = 1$ for any $k \in [K]$,

3. by Lemma 3, $\sum_{k=1}^{H_a} \mathbb{1}_{a,k}(t) \leq 1$ for any given a and t.

By following the above facts, we have $m_{a,k} \leq \tau_{a,k}(T)$ and $\sum_{k=1}^{H_a} m_{a,k} \leq m_{a,H_a}$. Based on these, we continue (33) as follows

$$\sum_{a\in\overline{A^\star}} \sum_{k=1}^{H_a} \Delta_{a,k} \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}_{a,k}(t) \cdot \mathbb{1}\{\Lambda_{1,t,a}, \Lambda_{2,t,a}\}\right]$$

$$\leq \sum_{a\in\overline{A^\star}} \left[\Delta_{a,1}\tau_{a,1}(T) + \sum_{k=2}^{H_a} \Delta_{a,k}(\tau_{a,k}(T) - \tau_{a,k-1}(T))\right] \tag{34}$$

$$= 48C^2\sigma^2 \log(t) \sum_{a\in\overline{A^\star}} \left[\Delta_{a,1}\frac{1}{\Delta_{a,1}^2} + \sum_{k=2}^{H_a} \Delta_{a,k}\left(\frac{1}{\Delta_{a,k}^2} - \frac{1}{\Delta_{a,k-1}^2}\right)\right] \tag{35}$$

$$= 48C^2\sigma^2 \log(t) \sum_{a\in\overline{A^\star}} \left(\sum_{k=1}^{H_a-1} \frac{\Delta_{a,k} - \Delta_{a,k+1}}{\Delta_{a,k}^2} + \frac{1}{\Delta_{a,H_a}}\right) \tag{36}$$

$$\leq 48C^2\sigma^2 \log(t) \sum_{a\in\overline{A^\star}} \left(\sum_{k=1}^{H_a-1} \frac{\Delta_{a,k} - \Delta_{a,k+1}}{\Delta_{a,k}\Delta_{a,k+1}} + \frac{1}{\Delta_{a,H_a}}\right) \tag{37}$$

$$= 48C^2\sigma^2 \log(t) \sum_{a\in\overline{A^\star}} \left[\sum_{k=1}^{H_a-1} \left(\frac{1}{\Delta_{a,k+1}} - \frac{1}{\Delta_{a,k}}\right) + \frac{1}{\Delta_{a,H_a}}\right] \tag{38}$$

$$= 48C^2\sigma^2 \log(t) \sum_{a\in\overline{A^\star}} \left(\frac{2}{\Delta_{a,H_a}} - \frac{1}{\Delta_{a,1}}\right) < 48C^2\sigma^2 \log(t) \sum_{a\in\overline{A^\star}} \frac{2}{\Delta_{a,H_a}}. \tag{39}$$

Finally, by combining equation (39) and (25) together, we get

$$\mathcal{R}_T \leq 96C^2\sigma^2 \sum_{a\in\overline{A^\star}} \frac{\log(t)}{\Delta_a} + \frac{\pi^2}{6} \sum_{a\in\overline{A^\star}} \sum_{k=1}^{H_a} \Delta_{a,k}, \tag{40}$$

which concludes the proof.

C Proof of Theorem 2

For any $a \in E$, let $H_{a,\lambda}$ be the number of optimal base arms in $\mathcal{H}_a$ whose feedback mean is higher than that of the sub-optimal base arm a by at least λ. According to (23), $\mathcal{R}_T$ is bounded for any λ as:

$$\mathcal{R}_T \leq \sum_{a\in\overline{A^\star}} \sum_{k=1}^{H_{a,\lambda}} \Delta_{a,k} \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}_{a,k}(t)\right] + \sum_{a\in\overline{A^\star}} \sum_{k=H_{a,\lambda}+1}^{H_a} \Delta_{a,k} \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}_{a,k}(t)\right]. \tag{41}$$

The first term in (41) can be bounded similarly to (40):

$$\sum_{a\in\overline{A^\star}}\sum_{k=1}^{H_{a,\lambda}}\Delta_{a,k}\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{1}_{a,k}(t)\right] \tag{42}$$

$$\leq 96C^2\sigma^2\sum_{a\in\overline{A^\star}}\frac{\log(T)}{\Delta_{a,H_{a,\lambda}}}+\frac{\pi^2}{6}\sum_{a\in\overline{A^\star}}\sum_{k=1}^{H_{a,\lambda}}\Delta_{a,k} \tag{43}$$

$$<\frac{96C^2\sigma^2(N-K)\log(T)}{\lambda}+\frac{\pi^2(N-K)K}{6}. \tag{44}$$

The second term in (41) can be bounded trivially as:

$$\sum_{a\in\overline{A^\star}}\sum_{k=H_{a,\lambda}+1}^{H_a}\Delta_{a,k}\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{1}_{a,k}(t)\right]\leq\lambda KT, \tag{45}$$

where we just use the fact that all gaps $\Delta_{a,k}$ are upper bounded by λ and the maximum number of sub-optimally chosen base arms in T rounds is KT (Lemma 3). By combining the above upper bounds on the two terms in (41) together, we obtain that

$$\mathcal{R}_T\leq\frac{96C^2\sigma^2(N-K)\log(T)}{\lambda}+\lambda KT+\frac{\pi^2(N-K)K}{6}. \tag{46}$$

Finally, by setting $\lambda=4\sqrt{6}C\sigma\sqrt{\frac{(N-K)\log T}{KT}}$, we get

$$\mathcal{R}_T\leq 4\sqrt{6}C\sigma\sqrt{(N-K)KT\log T}+\frac{\pi^2(N-K)K}{6}, \tag{47}$$

which concludes the proof.

D Proof of Theorem 3

Note that our argument for bounding $\mathbb{E}[T_a(t-1)]$ in Theorem 1 works under the following condition

$$\epsilon\leq\frac{\Delta_a}{4\Delta_a+4\sqrt{3}C\sigma\sqrt{\log(T)}}. \tag{48}$$

Let $\mathcal{S}$ be the set of base arms satisfying the condition (48). The arguments in the proof of Theorem 1 show that

$$\sum_{a\in\overline{A^\star}\cap\mathcal{S}}\sum_{k=1}^{H_a}\Delta_{a,k}\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{1}_{a,k}(t)\right]\leq 96C^2\sigma^2\sum_{a\in\overline{A^\star}\cap\mathcal{S}}\frac{\log(t)}{\Delta_a}+\frac{\pi^2}{6}\sum_{a\in\overline{A^\star}\cap\mathcal{S}}\sum_{k=1}^{H_a}\Delta_{a,k} \tag{49}$$

For any base arm $a\notin\mathcal{S}$, we have

$$\Delta_a\geq\frac{4\sqrt{3}C\sigma\epsilon\sqrt{\log T}}{1-4\epsilon},$$

assuming that $\epsilon < \frac{1}{4}$. The total regret contribution for $a \notin \mathcal{S}$ is therefore

$$\sum_{a \in \overline{A^\star} \cap \overline{\mathcal{S}}} \sum_{k=1}^{H_a} \Delta_{a,k} \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}_{a,k}(t)\right] \leq \frac{4\sqrt{3}C\sigma\epsilon\sqrt{\log T}}{1-4\epsilon} \sum_{a \in \overline{A^\star} \cap \overline{\mathcal{S}}} \sum_{k=1}^{H_a} \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}_{a,k}(t)\right] \tag{50}$$

$$\leq \frac{4\sqrt{3}C\sigma K\epsilon}{1-4\epsilon} T\sqrt{\log T} \tag{51}$$

Therefore, the total regret should be bounded as follows

$$\mathcal{R}_T \leq 96C^2\sigma^2 \sum_{a \in \overline{A^\star} \cap \mathcal{S}} \frac{\log(t)}{\Delta_{a,H_a}} + \frac{\pi^2}{6} \sum_{a \in \overline{A^\star} \cap \mathcal{S}} \sum_{k=1}^{H_a} \Delta_{a,k} + \frac{4\sqrt{3}C\sigma K\epsilon}{1-4\epsilon} T\sqrt{\log T}, \tag{52}$$

which concludes the proof.

E Proof of Theorem 4

The proof is similar to that of Theorem 3. Specifically, for any base arm $a \in \mathcal{S}$, Theorem 2 itself applies, i.e.,

$$\sum_{a \in \overline{A^\star} \cap \mathcal{S}} \sum_{k=1}^{H_a} \Delta_{a,k} \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}_{a,k}(t)\right] \leq 4\sqrt{6}C\sigma\sqrt{(N-K)KT\log T} + \frac{\pi^2(N-K)K}{6}. \tag{53}$$

For any base arm $a \in \overline{\mathcal{S}}$, we have the same bound we derived in the proof of Theorem 3 hold, that is

$$\sum_{a \in \overline{A^\star} \cap \overline{\mathcal{S}}} \sum_{k=1}^{H_a} \Delta_{a,k} \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}_{a,k}(t)\right] \leq \frac{4\sqrt{3}C\sigma K\epsilon}{1-4\epsilon} T\sqrt{\log T}. \tag{54}$$

By combining the bounds above, we obtain

$$\mathcal{R}_T \leq 4\sqrt{6}C\sigma\sqrt{(N-K)KT\log T} + \frac{\pi^2(N-K)K}{6} + \frac{4\sqrt{3}C\sigma K\epsilon}{1-4\epsilon} T\sqrt{\log T}, \tag{55}$$

which concludes the proof.

F Proof of Lemma 1

Notations We first introduce some notations that will be used in the proof. We denote by $\mathcal{R}(\pi, \nu, T)$ the expected cumulative regret for an algorithm π on instance ν in T rounds.

Contamination model In our proof for the lower bound, we consider the well-known ϵ-Huber contamination, which is just a special case of the ϵ-contamination model as we discussed in Section 2.3. Given the contamination parameter $\epsilon \in (0, \frac{1}{2})$, for each pull of the base arm a, the observed feedback is either sampled independently from the true distribution with probability $1-\epsilon$, or sampled from some arbitrary and unknown contamination distribution.

Canonical bandit model We review the general canonical bandit model. In general, a matroid bandit optimization algorithm π is a mapping from an observation history to a probability distribution for choosing each supper arm. Under the ϵ-Huber contamination model, the interaction between π and ν over a given horizon T can be denoted as the observation history

$$\mathcal{H}_T := \{(a(1), \widetilde{r}(1)), (a(2), \widetilde{r}(2)), \cdots, (a(T), \widetilde{r}(T))\}, \tag{56}$$

where a denotes the base arm selected and $\widetilde{r}$ denotes the contaminated version of reward r. An observed history $\mathcal{H}_T$ is a random variable sampled from the following measurable space

$$\left(([N/K] \times \mathbb{R})^T, \mathcal{B}\left(([N/K] \times \mathbb{R})^T\right), \mathbb{P}_{\pi\nu}\right), \tag{57}$$

where $\mathcal{B}\left(([N/K] \times \mathbb{R})^T\right)$ is the Borel set on $([N/K] \times \mathbb{R})^T$ and $\mathbb{P}_{\pi\nu}$ is the probability measure induced by the algorithm π and the instance ν, which is defined as follows:

1. The probability of selecting a base arm $a(t) = a$ at time t is dictated only by the algorithm π, and we denote the probability by $\pi(a|\mathcal{H}_{t-1})$.
2. The distribution of rewards $r(t)$ in round t is $f^{\nu}_{a(t)}$, which is dependent on $a(t)$ and conditionally independent of the previous observed history $\mathcal{H}_{t-1}$.
3. Under the ϵ-Huber contamination model, the algorithm cannot observe $r(t)$ directly, but a contaminated version $\widetilde{r}(t)$ that only depends on the true reward $r(t)$. We denote the conditional distribution of $\widetilde{r}$ as $M(\widetilde{r}|r)$.

As a result, the distribution of the observed history $\mathcal{H}_T$ is

$$\mathbb{P}^T_{\pi\nu}(\mathcal{H}_T) = \prod_{t=1}^{T} \pi(a(t)|\mathcal{H}_{t-1}) f^{\nu}_{a(t)}(r(t)) M(\widetilde{r}(t)|r(t)) = \prod_{t=1}^{T} \pi(a(t)|\mathcal{H}_{t-1}) g^{\nu}_{a(t)}(\widetilde{r}) \tag{58}$$

where we let $g^{\nu}_{a(t)}(\widetilde{r}) \coloneqq f^{\nu}_{a(t)}(r(t)) M(\widetilde{r}(t)|r(t))$.

Lower bound proof With the above notations and preparation, we are now ready to prove the instance-dependent lower bound. We have,

$$\mathcal{R}(\pi, \nu, T) \geq \frac{T\Delta_{\overline{a}_i}}{2} \cdot \mathbb{P}_{\pi\nu}\left(N_{\overline{a}_i}(T) \geq \frac{T}{2}\right), \tag{59}$$

$$\mathcal{R}(\pi, \nu^i, T) \geq \frac{T\Delta_{\overline{a}_i}}{2} \cdot \mathbb{P}_{\pi\nu^i}\left(N_{\overline{a}_i}(T) \leq \frac{T}{2}\right). \tag{60}$$

Combining these two inequalities, we have

$$\mathcal{R}(\pi,\nu,T)+\mathcal{R}(\pi,\nu^i,T)\geq\frac{T\Delta_{\overline{a}_i}}{2}\left(\mathbb{P}_{\pi\nu}\left(N_{\overline{a}_i}(T)\geq\frac{T}{2}\right)+\mathbb{P}_{\pi\nu^i}\left(N_{\overline{a}_i}(T)\leq\frac{T}{2}\right)\right) \tag{61}$$

$$\geq\frac{T\Delta_{\overline{a}_i}}{4}\exp\left(-\mathrm{KL}\left(\mathbb{P}^T_{\pi\nu}||\mathbb{P}^T_{\pi\nu^i}\right)\right), \tag{62}$$

where in the second inequality we use the probabilistic Pinsker's inequality [16]. By classical divergence decomposition lemma [17, Lemma 15.1], we have

$$\mathrm{KL}\left(\mathbb{P}^T_{\pi\nu}\|\mathbb{P}^T_{\pi\nu^i}\right)=\sum_a\mathbb{E}_{\pi\nu}\left[N_a(T)\right]\mathrm{KL}\left(g_a^{\nu}\|g_a^{\nu^i}\right) \tag{63}$$

$$=\mathbb{E}_{\pi\nu}[N_{\overline{a}_i}(T)]\mathrm{KL}(g_{\overline{a}_i}^{\nu}\|g_{\overline{a}_i}^{\nu^i}), \tag{64}$$

where the second equality is due to the fact that ν and ν^i only differs in $\overline{a}_i$. By combing equation (62) and (64), we get

$$\mathcal{R}(\pi,\nu,T)\geq\frac{T\Delta_{\overline{a}_i}}{8}\exp\left(-\mathbb{E}_{\pi\nu}[N_{\overline{a}_i}(T)]\mathrm{KL}(g_{\overline{a}_i}^{\nu}\|g_{\overline{a}_i}^{\nu^i})\right). \tag{65}$$

Note that, according to Lemma 4, by setting $\Delta_i=c\cdot\frac{2\epsilon}{1-\epsilon}\leq\frac{1}{2}$ for some constant $c<1$, we have $\mathrm{TV}(f_{\overline{a}_i}^{\nu}\|f_{\overline{a}_i}^{\nu^i})\leq\frac{\Delta_i}{2}\leq c\cdot\frac{\epsilon}{1-\epsilon}$, which leads to $\mathrm{KL}(g_{\overline{a}_i}^{\nu}\|g_{\overline{a}_i}^{\nu^i})=0$. Therefore,

$$\mathcal{R}(\pi,\nu,T)\geq c\frac{\epsilon}{1-\epsilon}T. \tag{66}$$

Recall the instance-dependent lower bound result for multi-armed Bernoulli bandit, see e.g., [17], an instance-dependent lower bound of $\Omega(\sum_{a\in\overline{A^\star}\cap P_k}\frac{\log T}{\Delta_a})$ holds for non-contaminated matroid bandit optimization. Thus, by combining it with the contaminated lower bound we prove above, we obtain the following instance-dependent lower bound

$$\mathcal{R}_{T,P_k}\geq\Omega\left(\sum_{a\in\overline{A^\star}\cap P_k}\frac{\log T}{\Delta_a}+\frac{\epsilon}{1-\epsilon}T\right), \tag{67}$$

which concludes the proof.

G Proof of Theorem 5

Recall that, the hard instance we use is essentially equivalent to K multi-armed Bernoulli bandit of the same arm size of $\frac{N}{K}$. The instance-dependent lower bound for regret of matroid bandit under contamination is derived as follows

$$\mathcal{R}_T\geq\Omega\left(\sum_{k=1}^K\sum_{a\in\overline{A^\star}\cap P_k}\frac{\log T}{\Delta_a}+\sum_{k=1}^K\frac{\epsilon}{1-\epsilon}T\right) \tag{68}$$

$$=\Omega\left(\sum_{a\in\overline{A^\star}}\frac{\log T}{\Delta_a}+\frac{K\epsilon}{1-\epsilon}T\right), \tag{69}$$

where in the first inequality we apply Lemma 1 separately to each P_k.

H Proof of Theorem 6

Actually, the contaminated lower bound of $\Omega(\frac{K\epsilon}{1-\epsilon}T)$ we provided in the proof of Theorem 5 is independent on instances. So next we study lower bound for non-contaminated matroid bandit optimization, which was left as an open problem in [14]. Consider the instance ξ described in Section 4 and note that ξ is equivalent to K N/K-armed Bernoulli bandit, the instance-independent non-contamination lower bound is derived as follows:

$$\mathcal{R}_T \geq \Omega\left(\sum_{k=1}^{K}\sqrt{\left(\frac{N}{K}-1\right)T}\right) = \Omega\left(\sqrt{(N-K)KT}\right), \tag{70}$$

where in the first inequality we apply Lemma 6 separately to each partition P_k.

By combining with the lower bound of $\Omega\left(\frac{K\epsilon}{1-\epsilon}T\right)$ under the ϵ-Huber contamination model we obtained before, we finally obtain

$$\mathcal{R}_T \geq \Omega\left(\sqrt{(N-K)KT} + \frac{K\epsilon}{1-\epsilon}T\right), \tag{71}$$

which concludes the proof.