Robust Matroid Bandit Optimization: Near-Optimal Rates under Adversarial Contamination*

Youming Tao[a,b], Xiuzhen Cheng[a], Falko Dressler[b], Zhipeng Cai[c], Dongxiao Yu[a,*]

[a]*Shandong University, China*
[b]*TU Berlin, Germany*
[c]*Georgia State University, USA*

Abstract

We study the matroid bandit optimization problem, a fundamental and broadly applicable framework for combinatorial multi-armed bandits where the action space is constrained by a matroid. In particular, we address the challenge of designing algorithms that remain effective under adversarial contamination of feedback rewards, which may severely degrade performance or even mislead existing methods. Our main contribution is an efficient and robust algorithm named ROMM, which builds upon the principle of optimistic matroid maximization and leverages robust statistical estimators to assess base arm quality in polynomial time. Under the ϵ-contamination model, we establish lower bounds and and prove that ROMM achieves near-optimal regret guarantees up to polylogarithmic factors. Our analysis further reveals a sharp phase transition between the low and high contamination regimes. Notably, ROMM can tolerate up to a universal constant fraction of corrupted feedback, which is optimal under mild conditions. Finally, we validate our theoretical findings with numerical experiments that demonstrate the effectiveness of the proposed method.

Keywords: Multi-Armed Bandits, Combinatorial Optimization, Matroids, Robustness.

1. Introduction

Combinatorial optimization is a classical area with wide-ranging practical applications, including resource allocation [1] and network routing [2]. Modern instances of these problems are often so massive that even mildly polynomial-time solutions are infeasible. Fortunately, many significant problems, such as finding a minimum spanning tree, admit efficient greedy solutions. Such problems can often be formulated as optimization over a *matroid* [3], a combinatorial structure that captures the concept of independence and is closely tied to computational tractability. Notably, it is well established that maximizing a modular function subject to a matroid constraint can be solved greedily if and only if feasible solutions correspond to the independent sets of a matroid [4]. Matroids are ubiquitous in applications because they generalize many forms of independence, including linear independence and forest structures in graphs.

In this paper, we consider a more realistic setting of learning how to maximize a *stochastic* modular function on a matroid. Specifically, the modular function is defined as the sum of the weights of up to K items selected from the ground set E of N items. Each item $a \in E$ is associated with an unknown, unbounded σ-sub-Gaussian distribution $\mathcal{D}_a$, and selecting item a yields a sample from $\mathcal{D}_a$. These distributions $\mathcal{D}_a$'s are initially unknown and must be learned through repeated interactions with the environment.

Many real-world optimization problems can be modeled in our setting. For example, consider building a minimum spanning tree for network routing [2]. If the delays on the network links are stochastic and their distributions is known, the optimal routing corresponds to a classical minimum spanning tree. However, when the delay distributions are

*A preliminary version of this paper was presented in the 30th International Computing and Combinatorics Conference (COCOON 2024).
*Corresponding Author.
Email addresses: tao@ccs-labs.org (Youming Tao), xzcheng@sdu.edu.cn (Xiuzhen Cheng), dressler@ccs-labs.org (Falko Dressler), zcai@gsu.edu (Zhipeng Cai), dxyu@sdu.edu.cn (Dongxiao Yu)

unknown, one must learn them over time by observing link delays. This introduces a trade-off between exploration and exploitation, which is characteristic for *stochastic multi-armed bandits* framework [5] where a learner seeks to optimize decisions based on stochastic feedback. Motivated by this connection, we refer to this problem setting as *matroid bandit optimization*.

Prior Arts. Matroid bandit optimization has attracted significant attention in the past decade, with investigations into various aspects such as regret minimization [6, 7, 8], pure exploration [9], algorithm efficiency [10] and differential privacy guarantees [11]. However, existing works commonly assume feedback is sampled faithfully from the underlying distributions. In practice, this assumption is often violated as feedback may be corrupted due to system malfunction or adversarial attacks [12]. In such cases, standard bandit algorithm suffer degraded performance. Recently, there has been a surge of interest in developing bandit algorithms that are robust to data corruption, including work on classical multi-armed bandit [12, 13, 14, 15] and more general combinatorial settings [16]. Although matroid bandits are a special case of combinatorial bandits, general-purpose algorithms are often too computationally intensive. By exploiting the greedy structure of matroids, much more efficient solutions can be achieved [6]. Thus, designing efficient and robust matroid bandit algorithms under adversarial contamination remains an open and important challenge.

Our Contributions. This paper focuses on the robustness of matroid bandit optimization in the presence of adversarial corruption. We study regret minimization under the ϵ-*contamination model*, where an ϵ fraction of the feedback is corrupted by an arbitrary adversary. Our main contributions are as follows:

- We introduce a robust algorithmic framework for matroid bandit optimization, called `ROMM`, which is based on the principle of optimistic matroid maximization. It incorporates a generic robust mean estimation subroutine, `RME`, for which we formalize the requirements and provide several concrete implementations.
- We establish both instance-dependent and instance-independent regret lower bounds under the ϵ-contamination model, characterizing the fundamental limits of performance with respect to the contamination level ϵ. As a byproduct, our results yield the first instance-independent lower bound for uncontaminated matroid bandits, resolving an open problem posed by [6].
- Through theoretical analysis, we show that `ROMM` achieves near-optimal regret bounds in both instance-dependent and instance-independent settings. Our results uncover a sharp phase transition between low and high contamination regimes: when ϵ is smaller than the minimum suboptimality gap $\Delta_{\min}$, `ROMM` maintains sublinear regret, whereas for larger ϵ (up to a universal constant $\frac{1}{4}$), the regret scales linearly with the horizon T.
- We perform numerical experiments on specific matroid instances to demonstrate that `ROMM` significantly outperforms non-robust baselines. Our experiments empirically confirm the phase transition phenomenon and validate the practical effectiveness of the proposed method.

2. Preliminaries

2.1. Combinatorial Optimization over a Matroid

Let $\mathcal{M} = (E, \mathcal{I})$ denote a matroid, where $E = \{1, \cdots, N\}$ is a finite ground set of N items, and $\mathcal{I}$ is a collection of subsets of E called the *independent sets*. A subset $A \subseteq E$ is independent if $A \in \mathcal{I}$.

Definition 1 (Matroid [3])**.** A pair $\mathcal{M}(E, \mathcal{I})$ is a matroid if the following properties hold:

1. The empty set is independent, i.e., $\emptyset \in \mathcal{I}$.
2. Hereditary property: If $A \in \mathcal{I}$ and $A' \subset A$, then $A' \in \mathcal{I}$.
3. Exchange property: If $A, A' \in \mathcal{I}$ with $|A| > |A'|$, then there exists $a \in A \setminus A'$ such that $A' \cup a \in \mathcal{I}$.

An independent set $A \in \mathcal{I}$ is called a **basis** if it is maximal, i.e., it is not a proper subset of any other independent set in $\mathcal{I}$. Let $\mathcal{B}$ denote the set of all bases of $\mathcal{M}$. It is a classical result that all bases in a matroid have the same cardinality [3], known as the **rank** of the matroid, denoted by K; that is, $|A| = K$ for all $A \in \mathcal{B}$.

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 define the weight vector $\mu = (\mu_1, \mu_2, \ldots, \mu_N) \in (\mathbb{R}^+)^N$. The goal is to find a basis $A^* \in \mathcal{B}$ that maximizes the total weight:

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

This optimization can be efficiently solved using the greedy algorithm 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 the matroid bandit setting, a learner interacts with a matroid bandit instance over T rounds. Each item $a \in E$ is referred to as a **base arm** and is associated with an unknown feedback distribution $\mathcal{D}_a$ instead of a fixed weight. A basis $A \in \mathcal{B}$ is referred to as a **super arm**. At each round $t \in [T]$, each base arm a generates feedback $x_a(t) \sim \mathcal{D}_a$. We assume each $\mathcal{D}_a$ is an unbounded σ-sub-Gaussian distribution with mean μ_a, which generalizes the bounded distributions used in earlier works [6, 7].

The feedback sequence $\{x_a(t)\}_{t=1}^T$ is i.i.d. for each base arm a, but the values $\{x_a(t)\}_{a=1}^N$ may be arbitrarily correlated across arms at any fixed time t. We adopt the *semi-bandit feedback model* [17], in which, at each round t, the learner selects a super arm $A(t) \coloneqq \{a_1(t), a_2(t), \cdots, a_K(t)\} \in \mathcal{B}$ and observes both the total reward $r(t) \coloneqq \sum_{a \in A(t)} x_a(t)$ and the individual feedback $\{(a, x_a(t)) | a \in A(t)\}$.

The objective is to minimize the expected cumulative **regret** over T rounds, defined as:

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

where $A^\star \coloneqq \arg\max_{A \in \mathcal{B}} \sum_{a \in A} \mu_a$ is the optimal super arm. Let $A^\star \coloneqq \{a_1^\star, \cdots, a_K^\star\}$ such that $\mu_{a_1^\star} \geq \cdots \geq \mu_{a_K^\star}$. We call any $a \in A^\star$ an **optimal** base arm, and any $a \in \overline{A^\star} \coloneqq E \backslash A^\star$ a **sub-optimal** base arm. For any sub-optimal $a \in \overline{A^\star}$ and optimal base arm $a_k^\star \in A^\star$, define the **gap**:

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

Let

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

be the set of indices of optimal arms whose mean is higher than that of a, and define

$$H_a \coloneqq |\mathcal{H}_a|. \tag{5}$$

The **sub-optimality gap** of arm a is then defined as

$$\Delta_a \coloneqq \Delta_{a,H_a}. \tag{6}$$

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 $a \in A(t)$.

2.3. Contamination Model

We now consider a setting in which the learner observes corrupted feedback rather than the true values. Specifically, for any base arm $a \in E$ and any round t, an adversary may replace $x_a(t)$ with an arbitrary value $y_a(t)$. The adversary is constrained by an *ϵ-contamination rate*, meaning that no more than an ϵ fraction of the feedback for each arm may be corrupted by round t:

$$\frac{\sum_{i=1}^{t}\{\mathbb{1}_{x_a(i)\neq y_a(i)}\}}{t} \leq \epsilon, \quad \text{for } \forall t \in [T]. \tag{7}$$

This setting corresponds to the well-known *ϵ-contamination model*, which has been extensively studied in the robust statistics literature [18, 19]. It generalizes the *ϵ-Huber contamination model* [20], commonly used in robust estimation (see also [21, 22]).

The contamination model in (7) is quite general as it allows the adversary to corrupt feedback arbitrarily, provided the overall corruption rate remains below ϵ. Moreover, the adversary is retrospective: it may adapt its strategy based on the learner's past, present, or even future actions, thus significantly complicating the learning process.

In this work, our goal is to design a robust learning algorithm for matroid bandit optimization under the contamination model defined in (7). We emphasize that although the observed feedback may be corrupted, the actual rewards received by the learner are based on the true feedback. Therefore, the regret defined in (2) remains valid as the performance metric.

3. Robust Matroid Bandit Optimization Framework: ROMM

3.1. Matroid Bandit Optimization Framework

Our framework is a robust extension of the Optimistic Matroid Maximization (OMM) algorithm developed in [6], which is grounded in the optimistic principle for decision-making under uncertainty [23]. Accordingly, we refer to our framework as Robust Optimistic Matroid Maximization (ROMM).

The key idea behind 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 round t, the algorithm substitutes the weight/expected feedback μ_a of each base arm a with an optimistic upper confidence bound (UCB) estimate $U_a(t)$, which is typically computed as the estimated feedback mean plus a confidence interval. When there is no contamination, using the empirical mean as an estimator suffices to achieve the order-optimal regret.

However, empirical mean is highly sensitive to outliers. Even a single outlier can deviate the empirical mean arbitrarily. This vulnerability renders empirical mean estimators ineffective. To overcome this, our framework replaces the empirical mean with a robust mean estimator (RME) that can tolerate a limited fraction of corruptions.

To formalize this requirement, we introduce the notion of an (ϵ, δ)-robust mean estimator.

Definition 2 ((ϵ, δ)-robust mean estimator((ϵ, δ)-RME))**.** 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

$$|\texttt{RME}(S_n) - \mu| \leq I(\epsilon, \delta, n) := C \cdot \frac{\sigma}{1 - 2\epsilon} \left(\sqrt{\frac{\log \frac{1}{\delta}}{n}} + \epsilon \sqrt{\log \frac{1}{\delta}} \right), \tag{8}$$

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

Remark 1. While (ϵ, δ)-RME mitigates adversarial corruption, it also introduces a trade-off that impacts exploration. Specifically, robust estimation reduces the effective sample size from n to roughly $(1 - 2\epsilon)^2 n$, inflating the confidence interval. This leads to more conservative estimates and prolonged exploration. Moreover, adversarial contamination limits how accurately each base arm can be learned, as reflected in the non-vanishing additive error term of order $\widetilde{O}(\frac{\sigma\epsilon}{1-2\epsilon})$ in the confidence bound. These effects illustrate the inherent tension between robustness and statistical efficiency in matroid bandit optimization.

We provide concrete examples of robust mean estimators satisfying Definition 2 in the next subsection. With this foundation, we now introduce our ROMM framework for matroid bandit optimization under the ϵ-contamination model, as described in Algorithm 2. At a high level, ROMM operates in each round t by performing the following steps:

1. For each base arm $a \in E$, compute its UCB estimate $U_a(t)$ using a robust mean estimator and its corresponding confidence interval [lines 7-8].
2. Sort all base arms in descending order of their UCB values [line 9].
3. Construct the super arm $A(t)$ by greedily selecting arms according to this order, while maintaining matroid independence [lines 10-12].
4. Pull the selected super arm $A(t)$ and observe the individual feedbacks $x_a(t)_{a \in A(t)}$ [lines 13-14].

A subtle but important detail is that the confidence interval $I(\epsilon, \delta, n)$ in the robust mean estimator contains an additive term independent of the sample size n. Since this term is the same across all base arms, it does not affect the relative ordering of UCB values and is therefore omitted in the implementation of the algorithm.

3.2. Instantiations of (ϵ, δ)-robust mean estimator

A key component of the ROMM framework is the (ϵ, δ)-robust mean estimator (RME), which is critical for tolerating adversarially corrupted feedback. While Definition 2 formally specifies the properties an RME must satisfy, its practical implementation remains to be addressed. Fortunately, the definition is broad enough to encompass several well-studied estimators from the robust statistics literature. Below, we present three concrete instantiations of (ϵ, δ)-robust mean estimators, all of which satisfy the robustness criteria in Definition 2.

- *Median (Med) [21]:* Find the median of all the sample points.
- *Trimmed Mean (TM) [24]:* Trim the smallest and largest ϵ fraction of points from the sample and calculate the mean of the remaining points.
- *Shorth Mean (SM) [12]:* 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.

We provide the formal performance guarantees of these estimators under the ϵ-contamination model.

Theorem 1. Let $S = \{z_1, \cdots, z_n\} \subset \mathbb{R}$ be i.i.d. samples drawn from a σ-sub-Gaussian distribution with mean μ and let S_C be an ϵ-contaminated version of S. For $\epsilon < \frac{1}{4}$ and $0 < \delta < 1$, each of the following bounds holds with probability at least $1 - \delta$:

$$|\texttt{Med}(S_C) - \mu| \leq e\sigma\left(\sqrt{\frac{\log\frac{4}{\delta}}{2n}} + \epsilon\right),$$

$$|\texttt{TM}(S_C) - \mu| \leq \frac{\sigma}{1-2\epsilon}\left(\sqrt{\frac{2\log\frac{4}{\delta}}{n}} + 4\epsilon\sqrt{3\log\frac{4}{\delta}}\right), \tag{9}$$

$$|\texttt{SM}(S_C) - \mu| \leq \frac{\sigma}{1-2\epsilon}\left(\sqrt{\frac{2\log\frac{4}{\delta}}{n}} + \frac{2\epsilon(3-4\epsilon)}{1-\epsilon}\sqrt{3\log\frac{4}{\delta}}\right).$$

These results are based on existing analyses in the literature. The bound for the median estimator originates from [21], and the explicit concentration inequality presented here is taken directly from [13, Equation (2)]. The bound for the trimmed mean estimator is given in [12, Theorem1], which builds on foundational results from[24]. Lastly, the bound for the shorth mean estimator is established in [12, Theorem 2].

3.3. Theoretical Learning Performance Guarantees

Let $\Delta_{\min} \coloneqq \min_{a \in \overline{A^\star}} \Delta_a$ denote the minimum sub-optimality gap. We begin by analyzing the performance of ROMM in the regime of small contamination.

Theorem 2 (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 satisfies:

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

Proof of Theorem 2. Let R_t denote the instantaneous regret at round t:

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

Then, the total regret $\mathcal{R}_T$ satisfies:

$$\mathcal{R}_T = \sum_{t=1}^{T} \mathbb{E}[R_t] \leq \sum_{t=1}^{T} \mathbb{E}\left[\sum_{a \in \overline{A^\star}} \sum_{k=1}^{H_a} \Delta_{a,k} \cdot \mathbb{1}_{a,k}(t)\right] = \sum_{a \in \overline{A^\star}} \sum_{k=1}^{H_a} \Delta_{a,k} \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}_{a,k}(t)\right], \tag{10}$$

where the inequality follows from (A.2) in Lemma 3.

Recall that in the ROMM framework (Algorithm 2), we denote by $\widehat{\mu}_a(t)$ the estimated mean of base arm a at the end of round t, and by $T_a(t)$ the number of times arm a has been pulled by round t. Then confidence interval for the robust estimator is given by

$$I(\epsilon, \delta, n) = C \cdot \frac{\sigma}{1-2\epsilon}\left(\sqrt{\frac{\log(\frac{1}{\delta})}{n}} + \epsilon\sqrt{\log(\frac{1}{\delta})}\right).$$

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 6), we have,

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

Substituting into (10), we split the regret into two parts:

$$\mathcal{R}_T \leq \sum_{a\in\overline{A^\star}}\sum_{k=1}^{H_a}\Delta_{a,k}\mathbb{E}\Big[\sum_{t=1}^{T}\mathbb{1}_{a,k}(t)\cdot\mathbb{1}\{\Lambda_{t,a}\}\Big] + \sum_{a\in\overline{A^\star}}\sum_{k=1}^{H_a}\Delta_{a,k}\mathbb{E}\Big[\sum_{t=1}^{T}\mathbb{1}_{a,k}(t)\cdot\mathbb{1}\{\overline{\Lambda_{t,a}}\}\Big]. \tag{11}$$

The second term in (11) is bounded by:

$$\sum_{a\in\overline{A^\star}}\sum_{k=1}^{H_a}\Delta_{a,k}\mathbb{E}[\sum_{t=1}^{T}\mathbb{1}_{a,k}(t)\cdot\mathbb{1}\{\overline{\Lambda_{t,a}}\}] \leq \frac{\pi^2}{6}\sum_{a\in\overline{A^\star}}\sum_{k=1}^{H_a}\Delta_{a,k}, \tag{12}$$

For the first term in (11), a sub-optimal arm a is chosen over its optimal counterpart $a^* \coloneqq a_k^\star$ only if:

$$\mu_a + \frac{2\sigma}{1-2\epsilon}\sqrt{\frac{3\log(t)}{T_a(t-1)}} + \frac{\sigma}{1-2\epsilon}\epsilon\sqrt{3\log(t)} \geq U_a(t) \geq U_{a^*}(t) \geq \mu_{a^*} - \frac{\sigma}{1-2\epsilon}\sqrt{3\log(t)}, \tag{13}$$

implying

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

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

$$\Delta_{a,k} > I(\epsilon, t^{-3}, T_a(t-1)). \tag{14}$$

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 (14):

$$T_a(t-1) > \frac{48C^2\sigma^2\log(t)}{\Delta_{a,k}^2}. \tag{15}$$

Note that, when $\epsilon \leq \frac{\Delta_{\min}}{4\Delta_{\min}+4\sqrt{3}C\sigma\sqrt{\log(T)}}$, (15) holds for all $a \in E$. Define $\tau_{a,k}(t) = \frac{48C^2\sigma^2\log(t)}{\Delta_{a,k}^2}$. Then we know that, when $T_a(t-1) > \tau_{a,k}(t)$, the algorithm must choose $a_k^\star$ instead of a. In other words, when event $\Lambda_{t,a}$ holds, a can only be chosen in rounds where $T_a(t-1) \leq \tau_{a,k}(t)$. Based on this, the first term in (11) can be bounded as follows:

$$\begin{aligned}
\sum_{a\in\overline{A^\star}}\sum_{k=1}^{H_a}\Delta_{a,k}\mathbb{E}\Big[\sum_{t=1}^{T}\mathbb{1}_{a,k}(t)\cdot\mathbb{1}\{\Lambda_{t,a}\}\Big] &\leq \sum_{t=1}^{T}\sum_{a\in\overline{A^\star}}\sum_{k=1}^{H_a}\Delta_{a,k}\mathbb{E}[\mathbb{1}_{a,k}(t)\cdot\mathbb{1}\{T_a(t-1)\leq\tau_{a,k}(t)\}] \\
&\leq \max\sum_{t=1}^{T}\sum_{a\in\overline{A^\star}}\sum_{k=1}^{H_a}\Delta_{a,k}\mathbb{1}_{a,k}(t)\mathbb{1}\{T_a(t-1)\leq\tau_{a,k}(t)\} \\
&= \max\sum_{a\in\overline{A^\star}}\left(\sum_{k=1}^{H_a}\Delta_{a,k}\sum_{t=1}^{T}\mathbb{1}_{a,k}(t)\mathbb{1}\{T_a(t-1)\leq\tau_{a,k}(t)\}\right).
\end{aligned} \tag{16}$$

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 (16) as follows,

$$\begin{aligned}
\sum_{a\in\overline{A^\star}}\sum_{k=1}^{H_a}\Delta_{a,k}\mathbb{E}\Big[\sum_{t=1}^{T}\mathbb{1}_{a,k}(t)\cdot\mathbb{1}\{\Lambda_{1,t,a},\Lambda_{2,t,a}\}\Big] &\leq \sum_{a\in\overline{A^\star}}\left[\Delta_{a,1}\tau_{a,1}(T)+\sum_{k=2}^{H_a}\Delta_{a,k}(\tau_{a,k}(T)-\tau_{a,k-1}(T))\right] \\
&= 48C^2\sigma^2\log(t)\sum_{a\in\overline{A^\star}}\left[\Delta_{a,1}\frac{1}{\Delta_{a,1}^2}+\sum_{k=2}^{H_a}\Delta_{a,k}\left(\frac{1}{\Delta_{a,k}^2}-\frac{1}{\Delta_{a,k-1}^2}\right)\right] \\
&= 48C^2\sigma^2\log(t)\sum_{a\in\overline{A^\star}}\left(\sum_{k=1}^{H_a-1}\frac{\Delta_{a,k}-\Delta_{a,k+1}}{\Delta_{a,k}^2}+\frac{1}{\Delta_{a,H_a}}\right) \\
&\leq 48C^2\sigma^2\log(t)\sum_{a\in\overline{A^\star}}\left(\sum_{k=1}^{H_a-1}\frac{\Delta_{a,k}-\Delta_{a,k+1}}{\Delta_{a,k}\Delta_{a,k+1}}+\frac{1}{\Delta_{a,H_a}}\right) \\
&= 48C^2\sigma^2\log(t)\sum_{a\in\overline{A^\star}}\left[\sum_{k=1}^{H_a-1}\left(\frac{1}{\Delta_{a,k+1}}-\frac{1}{\Delta_{a,k}}\right)+\frac{1}{\Delta_{a,H_a}}\right] \\
&= 48C^2\sigma^2\log(t)\sum_{a\in\overline{A^\star}}\left(\frac{2}{\Delta_{a,H_a}}-\frac{1}{\Delta_{a,1}}\right) \qquad (17) \\
&< 48C^2\sigma^2\log(t)\sum_{a\in\overline{A^\star}}\frac{2}{\Delta_{a,H_a}}. \qquad (18)
\end{aligned}$$

Finally, by combining equation (18) and (12) together, we get

$$\mathcal{R}_T \leq 96C^2\sigma^2\sum_{a\in\overline{A^\star}}\frac{\log(t)}{\Delta_a}+\frac{\pi^2}{6}\sum_{a\in\overline{A^\star}}\sum_{k=1}^{H_a}\Delta_{a,k}, \qquad (19)$$

which concludes the proof. □

Theorem 3 (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

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

Proof of Theorem 3. 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 (10), $\mathcal{R}_T$ is bounded for any λ as:

$$\mathcal{R}_T \leq \sum_{a\in\overline{A^\star}}\sum_{k=1}^{H_{a,\lambda}}\Delta_{a,k}\mathbb{E}\Big[\sum_{t=1}^{T}\mathbb{1}_{a,k}(t)\Big]+\sum_{a\in\overline{A^\star}}\sum_{k=H_{a,\lambda}+1}^{H_a}\Delta_{a,k}\mathbb{E}\Big[\sum_{t=1}^{T}\mathbb{1}_{a,k}(t)\Big]. \qquad (20)$$

The first term in (20) can be bounded similarly to (19):

$$\begin{aligned}
\sum_{a\in\overline{A^\star}}\sum_{k=1}^{H_{a,\lambda}}\Delta_{a,k}\mathbb{E}\Big[\sum_{t=1}^{T}\mathbb{1}_{a,k}(t)\Big] &\leq 96C^2\sigma^2\sum_{a\in\overline{A^\star}}\frac{\log(T)}{\Delta_{a,H_{a,\lambda}}}+\frac{\pi^2}{6}\sum_{a\in\overline{A^\star}}\sum_{k=1}^{H_{a,\lambda}}\Delta_{a,k} \\
&< \frac{96C^2\sigma^2(N-K)\log(T)}{\lambda}+\frac{\pi^2(N-K)K}{6}.
\end{aligned}$$

The second term in (20) can be bounded trivially as:

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

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 (20) together, we obtain that

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

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

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

which concludes the proof. □

Remark 2. The above Theorem 2 and Theorem 3 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, [6] 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 4 (Instance-Dependent Regret Upper Bound)**.** If $\epsilon \leq \frac{1}{4}$, the instance-dependent expected cumulative regret of ROMM is at most

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

Proof of Theorem 4. Note that our argument for bounding $\mathbb{E}[T_a(t-1)]$ in Theorem 2 works under the following condition

$$\epsilon \leq \frac{\Delta_a}{4\Delta_a+4\sqrt{3}C\sigma\sqrt{\log(T)}}. \tag{21}$$

Let $\mathcal{S}$ be the set of base arms satisfying the condition (21). The arguments in the proof of Theorem 2 show that

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

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

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

Therefore, the total regret should be bounded as follows

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

which concludes the proof. □

Theorem 5 (Instance-Independent Regret Upper Bound)**.** If $\epsilon \leq \frac{1}{4}$, the instance-independent expected cumulative regret of ROMM is at most

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

Proof of Theorem 5. The proof is similar to that of Theorem 4. Specifically, for any base arm $a \in \mathcal{S}$, Theorem 3 itself applies, i.e.,

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

For any base arm $a \in \overline{\mathcal{S}}$, we have the same bound we derived in the proof of Theorem 4 hold, that is

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

By combining the bounds above, we obtain

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

which concludes the proof. □

Remark 3. The above Theorem 4 and Theorem 5 illustrates that, in general, ROMM incurs a linear regret term of $\tilde{O}(\frac{\epsilon}{1-\epsilon} KT)$ 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 [6].

Remark 4. 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 5 recovers the state-of-the-art bound of $\widetilde{O}(\sqrt{NT} + \frac{\epsilon}{1-4\epsilon} T)$ for ϵ-contaminated MAB in [12]. That is, our bound can be seen as a generalization of the previous MAB bound to the matroid bandit case.

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 [6]. Let $P_1, P_2, \cdots, P_K$ be a partition of the ground set E, such that,

$$\bigcup_{k=1}^{K} P_k = E, \text{ and } P_i \bigcap P_j = \emptyset \text{ for } \forall i, j \in [K]. \tag{22}$$

The family of independent sets is defined as

$$\mathcal{I} = \{I \subseteq E : |I \cap P_k| \leq 1, \forall k \in [K]\}. \tag{23}$$

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:

$$\mu_a = \begin{cases} \frac{1}{2}, & a = \min_{i \in P_k} i, \\ \frac{1}{2} - g_a, & \text{otherwise}, \end{cases} \tag{24}$$

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 [6] 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.,

$$|P_1| = |P_2| = \cdots = |P_K| = N/K. \tag{25}$$

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 $0 < c < 1$. [1]

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 [6, 25, 26, 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

$$\mathcal{R}_{T,P_k} \geq \Omega\left(\sum_{a \in \overline{A^*} \cap P_k} \frac{\log T}{\Delta_a} + \frac{\epsilon}{1-\epsilon} T \right). \tag{26}$$

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

Proof of Lemma 1. We denote by $\mathcal{R}(\pi, \nu, T)$ the expected cumulative regret for algorithm π over instance ν in T rounds.
Contamination setting: 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

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

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

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

where $\mathcal{B}\Big(([N/K] \times \mathbb{R})^T\Big)$ 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

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

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,

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

Combining these two inequalities, we have

$$\begin{aligned}\mathcal{R}(\pi, \nu, T) + \mathcal{R}(\pi, \nu^i, T) &\geq \frac{T\Delta_{\bar{a}_i}}{2}\left(\mathbb{P}_{\pi\nu}\left(N_{\bar{a}_i}(T) \geq \frac{T}{2}\right) + \mathbb{P}_{\pi\nu^i}\left(N_{\bar{a}_i}(T) \leq \frac{T}{2}\right)\right) \\ &\geq \frac{T\Delta_{\bar{a}_i}}{4} \exp\left(-\mathrm{KL}\left(\mathbb{P}^T_{\pi\nu} \| \mathbb{P}^T_{\pi\nu^i}\right)\right), \end{aligned} \tag{27}$$

where in the second inequality we use the probabilistic Pinsker's inequality [28]. By classical divergence decomposition lemma [29, Lemma 15.1], we have

$$\begin{aligned}\mathrm{KL}\left(\mathbb{P}^T_{\pi\nu} \| \mathbb{P}^T_{\pi\nu^i}\right) &= \sum_a \mathbb{E}_{\pi\nu}\left[N_a(T)\right] \mathrm{KL}\left(g^{\nu}_a \| g^{\nu^i}_a\right) \\ &= \mathbb{E}_{\pi\nu}[N_{\bar{a}_i}(T)] \mathrm{KL}(g^{\nu}_{\bar{a}_i} \| g^{\nu^i}_{\bar{a}_i}), \end{aligned} \tag{28}$$

where the second equality is due to the fact that ν and ν^i only differs in $\overline{a}_i$. By combing equation (27) and (28), we get

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

Note that, according to Lemma 5, by setting $\Delta_i = c \cdot \frac{2\epsilon}{1-\epsilon} \leq \frac{1}{2}$ for some constant $c < 1$, we have $\mathrm{TV}(f^{\nu}_{\overline{a}_i} \| f^{\nu^i}_{\overline{a}_i}) \leq \frac{\Delta_i}{2} \leq c \cdot \frac{\epsilon}{1-\epsilon}$, which leads to $\mathrm{KL}(g^{\nu}_{\overline{a}_i} \| g^{\nu^i}_{\overline{a}_i}) = 0$. Therefore,

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

Recall the instance-dependent lower bound result for multi-armed Bernoulli bandit, see e.g., [29], 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

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

which concludes the proof. □

Theorem 6 (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

$$\mathcal{R}_T \geq \Omega\left(\sum_{a \in \overline{A^\star}} \frac{\log T}{\Delta_a} + \frac{K\epsilon}{1-\epsilon}T\right). \tag{29}$$

Proof of Theorem 6. 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

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

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

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 7 (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

$$\mathcal{R}_T \geq \Omega\left(\sqrt{(N-K)KT} + \frac{K\epsilon}{1-\epsilon}T\right). \tag{30}$$

Proof of Theorem 7. Actually, the contaminated lower bound of $\Omega(\frac{K\epsilon}{1-\epsilon}T)$ we provided in the proof of Theorem 6 is independent on instances. So next we study lower bound for non-contaminated matroid bandit optimization, which was left as an open problem in [6]. 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:

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

where in the first inequality we apply Lemma 4 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

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

which concludes the proof. □

Remark 5. When $\epsilon = 0$, Theorem 7 establishes the instance-independent lower bound for non-contamination matroid bandit optimization, which solves the open problem left in [6].

Remark 6. Theorem 6 and Theorem 7 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 4 and Theorem 5 are nearly optimal with respect to the dominant term T up to $poly(\log T)$ factors.

5. Numerical Experiments

5.1. Experimental Setup

Implemented Algorithms. We evaluate the practical performance of the proposed algorithm `ROMM`, instantiated with the three robust mean estimators discussed in Section 3.2: Median (`Med`), Trimmed Mean (`TM`), and Shorth Mean (`SM`). As a baseline, we compare against the non-robust `OMM` algorithm from [6], which achieves near-optimal regret in the absence of contamination.

Matroid Bandit Environment. We consider a partition matroid bandit instance of rank $K = 3$ over $N = 9$ base arms. The arms are divided into three disjoint partitions P_1, P_2, P_3, each containing 3 arms. The mean rewards in each partition are configured as follows:

- P_1: $\mu = 0.9,\ 0.5,\ 0.1$
- P_2: $\mu = 0.8,\ 0.5,\ 0.2$
- P_3: $\mu = 0.7,\ 0.5,\ 0.3$

At each round, feedback from each arm is sampled from a Gaussian distribution with standard deviation $\sigma = 0.5$.

Contamination Model. We adopt the ϵ-Huber contamination model to generate adversarial feedback and vary the contamination level $\epsilon \in 0,\ 0.0001,\ 0.001,\ 0.01,\ 0.05,\ 0.1,\ 0.15,\ 0.2$. A strong adversary is used to maximally mislead the learner: when contamination occurs, the optimal arm in each partition returns a fixed negative value (-10^8), while sub-optimal arms return a fixed positive value (10^8), effectively inverting their observed performance.

Performance Measurement. We measure performance using cumulative regret over time. Each experiment is repeated over 10 independent trials using seeds $65283 + i$ for the i-th run. We report the mean cumulative regret along with standard deviations to capture variability across trials.

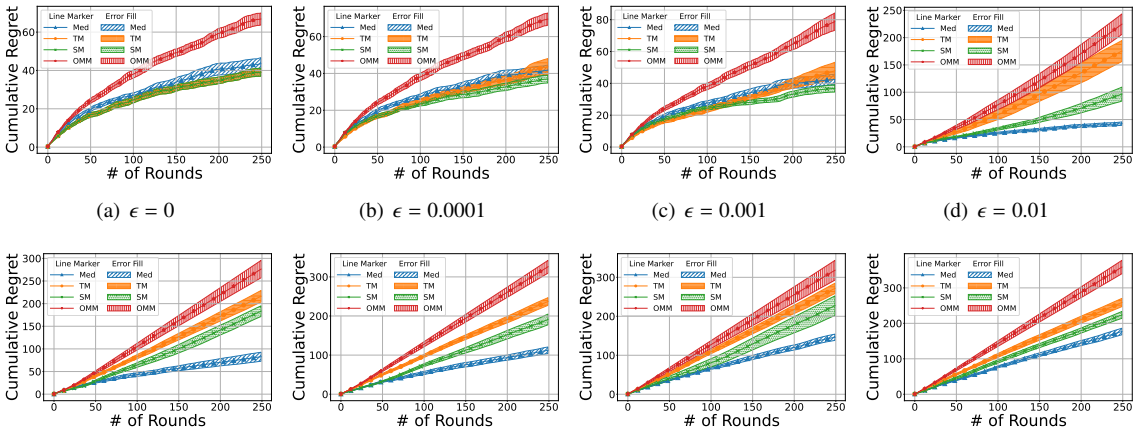
5.2. Results and Discussions

The results in Figure 1 highlight the effectiveness of `ROMM` across varying contamination levels. In the clean setting ($\epsilon = 0$) presented in Figure 1(a), all `ROMM` variants slightly outperform the non-robust baseline `OMM`, achieving lower cumulative regret. This suggests that robust mean estimators can enhance statistical efficiency even in the absence of contamination, potentially due to their ability to suppress outliers in the reward distribution. In contrast, `OMM` was primarily designed in [6] under the assumption of bounded rewards (e.g., in the [0, 1] interval), and may suffer in more general settings with unbounded or heavier-tailed rewards.

As contamination increases, the performance gap becomes more pronounced. While the regret of `OMM` grows rapidly, the robust variants of `ROMM` remain significantly more stable. Among them, the Median estimator (`Med`)

consistently achieves the lowest regret across contamination levels, followed by Shorth Mean (SM) and Trimmed Mean (TM).

Moreover, the results reveal a clear phase transition between the low and high contamination regimes. In the no-contamination and small contamination settings (Figure 1(a)-1(c)), the regret grows sublinearly—approximately logarithmically—with the number of rounds, consistent with theoretical guarantees. In contrast, for higher contamination levels (Figure 1(d)-1(h)), the regret grows nearly linearly, indicating that robust estimators are forced to continue exploring without being able to confidently distinguish optimal arms due to adversarial corruption. This transition aligns with the additive linear regret term predicted by our theoretical findings.



(a) $\epsilon = 0$ (b) $\epsilon = 0.0001$ (c) $\epsilon = 0.001$ (d) $\epsilon = 0.01$

(e) $\epsilon = 0.05$ (f) $\epsilon = 0.1$ (g) $\epsilon = 0.15$ (h) $\epsilon = 0.2$

Figure 1: Comparison of cumulative regret for ROMM and OMM. **Top:** Feedback generated from small contamination case (including no contamination) with contamination rate ϵ varying in $\{0, 0.0001, 0.001, 0.01\}$. **Bottom:** Feedback generated from large contamination case with contamination rate ϵ varying in $\{0.05, 0.1, 0.15, 0.2\}$.

6. Conclusion

In this paper, we studied the problem of matroid bandit optimization under the ϵ-contamination model, where a fraction of the feedback 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.

Acknowledgements

Youming 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. Dongxiao Yu and Xiuzhen Cheng were supported in part by the Major Basic Research Program of Shandong Provincial Natural Science Foundation under Grant

ZR2022ZD02. Dongxiao Yu was also supported in part by the National Natural Science Foundation of China (NSFC) under Grant 62122042.

Appendix A. Useful Facts for (Matroid) Bandits

Lemma 2 (Bijection [6])**.** 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:

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

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 [6])**.** 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

$$\mathbb{E}[R_t] \leq \sum_{a \in \overline{A^\star}} \sum_{k=1}^{H_a} \Delta_{a,k} \mathbb{1}_{a,k}(t), \tag{A.2}$$

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

$$\mathbb{1}_{a,k}(t) := \mathbb{1}\{\exists i : a_i(t) = a, \iota(a_i(t)) = a_k^\star\}. \tag{A.3}$$

Moreover,

$$\sum_{a \in \overline{A^\star}} \sum_{k=1}^{H_a} \mathbb{1}_{a,k}(t) \leq K, \forall t \in [T], \tag{A.4}$$

$$\sum_{k=1}^{H_a} \mathbb{1}_{a,k}(t) \leq 1, \forall t \in [T], a \in \overline{A^\star}. \tag{A.5}$$

Lemma 4 (Instance-Independent Lower Bound for Stochastic n-Multi-Armed Bandits [29, theorem15.2])**.** There exists a stochastic n-armed bandit instance such that the expected regret of any algorithm is $\Omega\left(\sqrt{(n-1)T}\right)$.

Appendix B. Probability and Statistics Tools

Lemma 5 (Theorem 5.1 in [30])**.** 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

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

Lemma 6 (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

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

References

[1] J. Zuo, C. Joe-Wong, Combinatorial multi-armed bandits for resource allocation, in: Proceedings of the 55th Annual Conference on Information Sciences and Systems (CISS), IEEE, 2021, pp. 1–4.

[2] R. Gallager, A minimum delay routing algorithm using distributed computation, IEEE transactions on communications 25 (1) (1977) 73–85.

[3] J. G. Oxley, Matroid theory, Vol. 3, Oxford University Press, USA, 2006.

[4] A. Schrijver, et al., Combinatorial optimization: polyhedra and efficiency, Vol. 24, Springer, 2003.

[5] P. Auer, N. Cesa-Bianchi, P. Fischer, Finite-time analysis of the multiarmed bandit problem, Machine learning 47 (2) (2002) 235–256.

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

[7] M. S. Talebi, A. Proutiere, An optimal algorithm for stochastic matroid bandit optimization, in: Proceedings of the 16th International Conference on Autonomous Agents & Multiagent Systems (AAMAS), 2016, pp. 548–556.

[8] Z. Huang, Y. Xu, B. Hu, Q. Wang, J. Pan, Thompson sampling for combinatorial semi-bandits with sleeping arms and long-term fairness constraints, arXiv preprint arXiv:2005.06725 (2020).

[9] L. Chen, A. Gupta, J. Li, Pure exploration of multi-armed bandit under matroid constraints, in: Proceedings of the 29th Conference on Learning Theory (COLT), 2016, pp. 647–669.

[10] P. Perrault, V. Perchet, M. Valko, Exploiting structure of uncertainty for efficient matroid semi-bandits, in: Proceedings of the 36th International Conference on Machine Learning (ICML), 2019, pp. 5123–5132.

[11] K. Chandak, B. Hu, N. Hegde, Differentially private algorithms for efficient online matroid optimization, in: Conference on Lifelong Learning Agents, PMLR, 2023, pp. 66–88.

[12] L. Niss, A. Tewari, What you see may not be what you get: Ucb bandit algorithms robust to *varepsilon*-contamination, in: Conference on Uncertainty in Artificial Intelligence, PMLR, 2020, pp. 450–459.

[13] S. Kapoor, K. K. Patel, P. Kar, Corruption-tolerant bandit learning, Machine Learning 108 (4) (2019) 687–715.

[14] D. Basu, O.-A. Maillard, T. Mathieu, Bandits corrupted by nature: Lower bounds on regret and robust optimistic algorithm, arXiv preprint arXiv:2203.03186 (2022).

[15] Y. Wu, X. Zhou, Y. Tao, D. Wang, On private and robust bandits, Advances in Neural Information Processing Systems 36 (2024).

[16] H. Xu, J. Li, Simple combinatorial algorithms for combinatorial bandits: Corruptions and approximations, in: Uncertainty in Artificial Intelligence, PMLR, 2021, pp. 1444–1454.

[17] J.-Y. Audibert, S. Bubeck, G. Lugosi, Regret in online combinatorial optimization, Mathematics of Operations Research 39 (1) (2014) 31–45.

[18] M. Charikar, J. Steinhardt, G. Valiant, Learning from untrusted data, in: Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, 2017, pp. 47–60.

[19] G. LUGOSI, S. MENDELSON, Robust multivariate mean estimation: The optimality of trimmed mean, The Annals of Statistics 49 (1) (2021) 393–410.

[20] P. J. Huber, Robust statistics, Vol. 523, John Wiley & Sons, 2004.

[21] K. A. Lai, A. B. Rao, S. Vempala, Agnostic estimation of mean and covariance, in: 2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS), IEEE, 2016, pp. 665–674.

[22] A. Prasad, S. Balakrishnan, P. Ravikumar, A robust univariate mean estimator is all you need, in: International Conference on Artificial Intelligence and Statistics, PMLR, 2020, pp. 4034–4044.

[23] R. Munos, The optimistic principle applied to games, optimization and planning: Towards foundations of monte-carlo tree search, Foundations and Trends in Machine Learning 7 (1) (2014) 1–130.

[24] L. Liu, T. Li, C. Caramanis, High dimensional robust estimation of sparse models via trimmed hard thresholding, arXiv preprint arXiv:1901.08237 (2019).

[25] D. Basu, C. Dimitrakakis, A. Tossou, Privacy in multi-armed bandits: Fundamental definitions and lower bounds, arXiv preprint arXiv:1905.12298 (2019).

[26] X. Chen, K. Zheng, Z. Zhou, Y. Yang, W. Chen, L. Wang, (locally) differentially private combinatorial semi-bandits, in: Proceedings of the 37th International Conference on Machine Learning (ICML), 2020, pp. 1757–1767.

[27] Y. Tao, Y. Wu, P. Zhao, D. Wang, Optimal rates of (locally) differentially private heavy-tailed multi-armed bandits, in: Proceedings of the 25th International Conference on Artificial Intelligence and Statistics, 2022, pp. 1546–1574.

[28] T. Lattimore, C. Szepesvári, An information-theoretic approach to minimax regret in partial monitoring, in: Proceedings of the 32nd Conference on Learning Theory (COLT), 2019, pp. 2111–2139.

[29] T. Lattimore, C. Szepesvári, Bandit algorithms, Cambridge University Press, 2020.

[30] M. Chen, C. Gao, Z. Ren, Robust covariance and scatter matrix estimation under huber's contamination model, The Annals of Statistics 46 (5) (2018) 1932–1960.