BANDIT SUBMODULAR MAXIMIZATION UNDER MATROID CONSTRAINTS: LEARNING COMPRESSED EXCHANGE POLICY

Anonymous authors
Paper under double-blind review

ABSTRACT

We study adversarial bandit maximization of monotone submodular functions under a matroid constraint. For a rank-k matroid on n elements, we give a randomized oracle-polynomial algorithm that makes one feasible value query per round and has expected $(1-1/e)$-regret $\widetilde{O}(n^{1/3}k^{2/3}T^{2/3})$. This is the first sublinear-regret algorithm for adversarial bandit submodular maximization under general matroid constraints.

Technically, we view the problem as learning an exchange policy for the Poisson base walk. This connects the problem to contextual bandits and gives an information-theoretic sublinear-regret guarantee, but directly learning the exponentially many policies requires exponential time and space. We therefore introduce *balanced fractional exchanges*, which compress the policy mixture into a single fractional base while retaining the exchange information needed by the Poisson analysis. This leads to an polynomial time algorithm with the same regret guarantee.

1 INTRODUCTION

Selecting a collection of items whose joint value exhibits diminishing returns is a basic problem in combinatorial optimization. The value of adding a new item is often largest when few related items have already been selected: an additional seed user reaches fewer new individuals after many influential users have been chosen, and an additional sensor supplies less information once nearby locations are already monitored. Monotone submodular functions formalize precisely this phenomenon. They underlie, among many other examples, influence maximization in social networks (Kempe et al., 2003) and information-based sensor placement (Krause et al., 2008).

The offline submodular maximization problem has a long history. For a cardinality constraint, the classical greedy algorithm achieves the factor $1-1/e$ (Nemhauser et al., 1978). Under a general matroid constraint, the elementary greedy algorithm no longer attains this factor; the optimal polynomial-time guarantee was recovered by continuous greedy, together with the multilinear extension and dependent rounding (Vondrák, 2008; Călinescu et al., 2011; Chekuri et al., 2010). Non-oblivious local search provides a discrete alternative with the same approximation factor (Filmus & Ward, 2014; Buchbinder & Feldman, 2024). Most recently, Ganz Rozenman et al. (2026) obtained $1-1/e$ through a different nonhomogeneous Poisson walk approach on matroid bases. Unlike continuous-relaxation algorithms, this walk moves only by feasible single-element exchanges, a feature that will be very helpful in this paper's setting.

In many applications, however, the objective is neither fixed nor known when the decision is made. A platform repeatedly chooses a slate of recommendations or advertisements before observing user response; a campaign selects seed users before the resulting diffusion is realized; and a monitoring system deploys a feasible sensor configuration before its utility is revealed. These examples motivate an online model in which, on each round t, the learner selects an independent set S_t before seeing the current monotone submodular reward function f_t. Under full-information feedback the learner subsequently receives access to the reward function, whereas under full-bandit feedback it observes only the single scalar $f_t(S_t)$. The natural benchmark is the best fixed feasible set in hindsight, weakened by the approximation factor that is unavoidable even offline.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Online and bandit submodular maximization were first studied through online-greedy and meta-action reductions. For cardinality constraints, Streeter & Golovin (2008) obtained $O(T^{2/3})$ $(1-1/e)$-regret; Streeter et al. (2009) developed the assignment setting, which corresponds to a structured partition-matroid constraint, with $O(T^{4/5})$ $(1-1/e)$-regret. Offline-to-online reductions based on Blackwell approachability later gave a general explanation of how robust greedy algorithms can be converted into online algorithms and bandit analogues (Niazadeh et al., 2023). Stochastic full-bandit variants have also been studied under cardinality, knapsack, and general matroid constraints (Nie et al., 2022; 2023; Fourati et al., 2024; Nie et al., 2025).

Turning an offline $(1-1/e)$-approximation into a bandit algorithm faces a basic obstacle: most existing offline methods evaluate auxiliary subsets that need not satisfy the matroid constraint. Such queries are harmless in the offline value-oracle model, where the algorithm may inspect the objective on any subset. In the bandit model, however, the queried set is also the action played by the learner. Under the original, or *strict feasible-query* model, every such set—including an exploratory set used only to construct an estimator—must be independent.

Zhang et al. (2019) made this obstruction precise for the multilinear-extension approach. They showed that, for some matroid, there is no function-independent randomized rounding scheme supported on feasible sets whose single observed value is an unbiased estimate of the multilinear extension for every submodular objective. Motivated by this impossibility, they introduced *responsive bandit submodular maximization*: the learner may query an infeasible set and observe its function value, but receives zero reward on that round. Under this relaxed feedback model, they obtained $O(T^{8/9})$ $(1-1/e)$-regret. However, in the original adversarial bandit model, where every queried set must be feasible, no sublinear $(1-1/e)$-regret guarantee was known for arbitrary monotone submodular rewards under general matroid constraints. Indeed, after establishing such a guarantee for partition matroids, Wan et al. (2023) conjectured that sublinear regret might fail for general matroid constraints.

We revisit this problem under the original strict feasible-query model. Our work is organized around the following question:

> *Can we design a bandit submodular maximization algorithm that achieves sublinear $(1-1/e)$-regret under general matroid constraints in polynomial time?*

We answer this question affirmatively, thereby disproving the conjecture of Wan et al. (2023). Our approach builds on the recent Poisson-process method for offline submodular maximization (Ganz Rozenman et al., 2026). We first formulate its online counterpart as a policy learning problem: the current base of the Poisson walk is the context, a feasible exchange is an action, and each comparator base induces a state-dependent exchange policy. This contextual-bandit viewpoint, which has not appeared in prior work on this problem, gives a possible route to sublinear regret. However, this approach only produce an information-theoretic guarantee, because the number of policies is exponentially large and requires exponential time and space to execute. We therefore introduce a technique to compress the resulting policy mixture into a fractional base and couple it with each current base of the Poisson walk while preserving the required exchange marginals. This new technique yields the desired polynomial-time algorithm with sublinear $(1-1/e)$-regret.

Main result. Our main result gives the optimal polynomial-time approximation factor $1-1/e$ in expected approximate regret, formally defined in (2.2). Concretely, we obtain the following theorem, which is an informal version of Theorem 5.3.

Theorem 1.1. *Let $T \geq 2$, and let $\mathcal{M}$ be a rank-k matroid on n elements. There exists a randomized oracle-polynomial algorithm that makes exactly one feasible value query per round and satisfies*

$$\left(1-\frac{1}{e}\right)\max_{O\in\mathcal{B}}\sum_{t=1}^{T} f_t(O) - \mathbb{E}\left[\sum_{t=1}^{T} f_t(S_t)\right] = \widetilde{O}\left(n^{1/3}k^{2/3}T^{2/3}\right).$$

To our knowledge, this is the first polynomial time sublinear $(1-1/e)$-regret guarantee for bandit monotone submodular maximization in the strict feasible-query bandit problem under general matroid constraint. The dependence on T matches the $T^{2/3}$ scale attained for previous results on uniform matroid and partition matroid (Streeter & Golovin, 2008; Wan et al., 2023).

1.1 Technical Overview

The starting point is the Poisson process approximation algorithm of Ganz Rozenman et al. (2026). It evolves a feasible base through a continuous-time Poisson process, performing at each arrival a single-element exchange that satisfies a suitable local improvement condition. These local conditions imply a differential inequality whose integration yields the $1 - 1/e$ approximation guarantee. This form is particularly well suited to bandit feedback because the subset maintained by the algorithm can remain feasible throughout. The difficulty is that an online learner must choose each exchange before observing the current reward function f_t and hence cannot directly enforce the required local condition. The Poisson analysis localizes the resulting loss: the gap in the final approximation is controlled by the accumulated deficits of the exchanges actually taken. Our task is therefore to learn these exchanges from one feasible query per round.

Why the standard meta-framework does not apply. A standard way to onlineize a sequential approximation algorithm is to associate a no-regret linear optimizer with each update step: on every round, the optimizers generate the updates in sequence, and the optimizer at a given step is trained on the marginal objective induced by the preceding updates (Streeter & Golovin, 2008; Streeter et al., 2009; Niazadeh et al., 2023). This template presumes that each step has a fixed action domain and that the offline comparator can be represented by a fixed action for the corresponding optimizer. The random number of Poisson arrivals can be handled by truncation, but the state dependence of an arrival cannot. Indeed, at a given arrival, both the set of feasible exchanges and their gains depend on the current random base A. Even for a fixed comparator base O, the exchange that certifies the Poisson improvement condition is selected through a Brualdi matching between A and O, and therefore varies with A rather than being a fixed pair (i, j). Thus, an online optimizer whose actions are individual exchanges has no fixed comparator valid across rounds.

First try: an information-theoretic policy learner. To address the above barrier, an action must instead specify how to exchange from every possible current base, that is, it must be a state-dependent exchange policy. For every base O, fix a complete exchange policy π_O: whenever the Poisson walk is at a base A, the policy chooses a uniformly random element $i \in A$ and inserts the element matched to i by a fixed Brualdi bijection from A to O. The policy class thus contains, for every possible hindsight base, a rule that is locally correct at every state that the walk may visit. Treating the current base as the context and a feasible exchange as the action, one may run the Exp4 algorithm over this finite policy class (Auer et al., 2002). With standard exploration, its finite-policy regret bound controls the accumulated Poisson deficits and yields sublinear approximation regret. Although the policy class is exponentially large, the regret depends only logarithmically on its cardinality. Computationally, however, Exp4 maintains one weight per policy and therefore requires exponential time and memory.

Using balanced fractional exchange to compress the policy mixture. To resolve the computational issue, a key observation is that the Poisson process analysis does not require the identity of the mixed comparator. Let W_t be the distribution over comparator policies in the above exponential time algorithm and set

$$x_t \triangleq \sum_{O \in \mathcal{B}} W_t(O) \mathbf{1}_O \in P(\mathcal{M}).$$

When the Brualdi exchanges are averaged according to W_t, the resulting exchange distribution has two simple marginals: the element removed from the current base is uniform, and the element inserted has distribution x_t/k. The Poisson analysis uses only these marginals, not the identity of the policy from which an exchange originated. Thus the exponentially supported distribution W_t can be replaced by the single fractional base x_t. We can therefore learn x_t directly, without maintaining any distribution over bases.

The remaining question is whether these two marginals can be realized at every current base using only feasible exchanges. This is precisely the role of a balanced fractional exchange. Given A and x_t, we construct a nonnegative matrix $Q_A^{x_t}$ supported on pairs (i, j) for which $A - i + j$ is a base and satisfying

$$\sum_{j \in \mathcal{U}} Q_A^{x_t}(i, j) = 1 \quad (i \in A), \qquad \sum_{i \in A} Q_A^{x_t}(i, j) = x_{t,j} \quad (j \in \mathcal{U}).$$

Sampling an exchange from $Q_A^{x_t}/k$ therefore reproduces exactly the deletion and insertion marginals of the policy mixture. We prove that, for every current base A and every $x_t \in P(\mathcal{M})$, such a matrix exists and can be constructed in polynomial time as a transportation flow on the bipartite exchange graph of A.

From balanced exchanges to a linear bandit. The marginal identities of a balanced exchange allow us to aggregate the state-dependent deficits along the Poisson trajectory without tracking the individual comparator-specific exchange rules. Combining these identities with a local linearization, we upper-bound the cumulative Poisson deficit by the regret of an online linear optimization problem over the matroid base polytope: the learner chooses a fractional base x_t, and every hindsight base O is represented by the fixed vertex $\mathbf{1}_O$. The original bandit submodular problem is thereby reduced to a specialized linear bandit problem. Its linear loss is not observed directly, but must be estimated from one feasible value query, which motivates the separate estimator construction below.

A feasible leave-one-out estimator. Constructing the estimator requires a further idea independent of the balanced-exchange representation. The local improvement condition in the original Poisson algorithm is expressed through the gradient gain $\partial_j F(\tau \mathbf{1}_A) - \partial_i F(\tau \mathbf{1}_A)$. Directly sampling its insertion term may require evaluating $f(S + j)$ for $S \sim \tau A$, which need not be feasible when S contains the element i that must be removed before inserting j. The leave-one-out gain avoids this obstruction by sampling $R \sim \tau(A - i)$: both $R + j \subseteq A - i + j$ and $R + i \subseteq A$ are feasible. Moreover, it equals the normalized multilinear jump caused by the exchange and dominates the original gradient gain, so it preserves the Poisson analysis. After linearization leaves only an insertion loss, a fair-coin randomization between querying $R + j$ and R gives a nonnegative unbiased estimate of this loss. Thus the required feedback is obtained from one feasible value query.

The remaining online-learning step is standard: importance weighting turns the one-query statistic into an unbiased estimate of the linear loss, and entropic mirror descent controls its cumulative regret. Together with a capped Poisson simulation and polynomial-time implementations of the balanced exchange flow and the entropy projection, we obtain the mentioned result.

1.2 Related Work

Stochastic submodular bandits. Linear submodular bandits model the unknown utility as a linear combination of known submodular components (Yue & Guestrin, 2011). Interactive submodular bandits assume an RKHS model and observe noisy marginal utilities (Chen et al., 2017). The stochastic full-bandit model is closer to our feedback protocol: the learner observes only the noisy total reward of the selected set. Under a cardinality constraint, Nie et al. (2022) give an explore-then-commit algorithm for monotone rewards. Fourati et al. (2024) improve the dependence on the cardinality bound using stochastic-greedy exploration. Subsequently, Tajdini et al. (2024) establish the first minimax lower bounds and a nearly matching upper bound for a restricted algorithm class. Nie et al. (2023) give a black-box reduction from robust offline approximation algorithms to stochastic full-bandit algorithms, with applications to monotone submodular maximization under cardinality and knapsack constraints. For general matroids, the monotone k_{sub}-submodular result of Nie et al. (2025), specialized to $k_{\mathrm{sub}} = 1$, gives an approximation factor of $1/2$ and $\widetilde{O}(n^{1/3}kT^{2/3})$ expected $1/2$-regret for ordinary set-submodular maximization. Aggarwal (2026) establishes controlled-oracle resilience for Poisson-process submodular maximization, then uses an offline-to-online reduction to obtain stochastic full-bandit guarantees under general matroid constraints. For unconstrained nonmonotone objectives that are submodular in expectation, Fourati et al. (2023) obtain a stochastic full-bandit guarantee. These results assume a fixed mean reward function or other stochastic structure; our setting allows an adversarial reward sequence.

Bandit submodular minimization. For unconstrained submodular minimization, the convexity of the Lovász extension allows a reduction to online convex optimization. Hazan & Kale (2012) give efficient algorithms for both full-information and bandit feedback, with an $O(nT^{2/3})$ regret bound in the adversarial bandit setting. Cardoso & Cummings (2019) develop differentially private variants under both feedback models. Ito (2022) studies best-of-both-worlds guarantees for online submodular minimization.

2 PRELIMINARIES

Submodular functions. For a finite ground set $\mathcal{U}$ and a set $S \subseteq \mathcal{U}$, let $\mathbf{1}_S \in \{0,1\}^{\mathcal{U}}$ denote the incidence vector of S. For an element $u \in \mathcal{U}$, write $S+u \triangleq S \cup \{u\}$ and $S-u \triangleq S \setminus \{u\}$. A function $f : 2^{\mathcal{U}} \to [0,1]$ is *normalized* if $f(\varnothing) = 0$, *monotone* if for any $A \subseteq B$ we have $f(A) \leq f(B)$. We call a set function is *submodular* if it satisfies following property for any $A \subseteq B,\ u \notin B$: $f(A+u) - f(A) \geq f(B+u) - f(B)$.

Matroids. Let $\mathcal{M} = (\mathcal{U}, \mathcal{I})$ be a loopless matroid of rank $k \geq 1$.[1] The family of bases is denoted by $\mathcal{B}$. For a base A, define its feasible single-exchange set

$$\mathcal{E}(A) \triangleq \{(i,j) : i \in A,\ j \in \mathcal{U},\ A - i + j \in \mathcal{B}\}. \tag{2.1}$$

Self-loops (i,i) are included. If $j \in A$ and $(i,j) \in \mathcal{E}(A)$, then necessarily $i = j$. The base polytope is $P(\mathcal{M}) \triangleq \operatorname{conv}\{\mathbf{1}_B : B \in \mathcal{B}\}$. Every $x \in P(\mathcal{M})$ satisfies $0 \leq x_u \leq 1$ and $\sum_{u \in \mathcal{U}} x_u = k$. We need the following standard result in our analysis.

Lemma 2.1 ((Brualdi, 1969)). *For every pair of bases $A, B \in \mathcal{B}$, there is a bijection $h_{A,B} : A \to B$ such that $A - i + h_{A,B}(i) \in \mathcal{B}$ for every $i \in A$. The bijection may be chosen so that $h_{A,B}(i) = i$ for all $i \in A \cap B$.*

Bandit protocol. An oblivious adversary fixes a sequence $f_1, \ldots, f_T : 2^{\mathcal{U}} \to [0,1]$ of normalized, monotone submodular functions before the interaction. In each round $t = 1, \ldots, T$, the learner chooses $S_t \in \mathcal{I}$ without observing f_t, then observes only the scalar $f_t(S_t)$ and receives this value. Because every independent set extends to a base and every f_t is monotone, we have $\max_{S \in \mathcal{I}} \sum_{t=1}^{T} f_t(S) = \max_{O \in \mathcal{B}} \sum_{t=1}^{T} f_t(O)$. Accordingly, for $\alpha \in [0,1]$, the expected α-regret is

$$R_\alpha(T) \triangleq \alpha \max_{O \in \mathcal{B}} \sum_{t=1}^{T} f_t(O) - \mathbb{E}\left[\sum_{t=1}^{T} f_t(S_t)\right]. \tag{2.2}$$

Multilinear extension. Fix a monotone submodular function $f : 2^{\mathcal{U}} \to [0,1]$. Its multilinear extension is $F(x) \triangleq \mathbb{E}_{R \sim x}[f(R)]$, where every element u is included independently with probability x_u. For $S \subseteq \mathcal{U}$ and $\tau \in [0,1]$, the notation $R \sim \tau S$ means that every element of S is included independently with probability τ and no element outside S is included.

3 POISSON BASE WALK WITH GENERAL EXCHANGE DISTRIBUTION

The Poisson-process algorithm of Ganz Rozenman et al. (2026) is an important component of our bandit algorithm. We generalize the algorithm and its analysis to allow general distributions over feasible exchanges. In their original algorithm, the exchange is chosen to maximize the instantaneous gain in the multilinear extension. For simplicity, we call the algorithm of Ganz Rozenman et al. (2026) the GKSS algorithm, after its authors.

Poisson base walk. Fix a monotone submodular function f with multilinear extension F, a rank-k matroid $\mathcal{M}$, a starting time $\varepsilon \in (0,1)$, and an arbitrary starting base A_0. The walk runs a nonhomogeneous Poisson process on $[\varepsilon, 1]$ with intensity $\lambda(s) \triangleq k/s$. For every time s and base A, let $p_s(\cdot \mid A)$ be a measurable distribution supported on the feasible exchanges $\mathcal{E}(A)$. We call each arrival of this Poisson process an *event*: event J occurs at time τ_J, samples a feasible exchange $e_J = (i_J, j_J) \sim p_{\tau_J}(\cdot \mid A_{J-1})$, and updates the base to $A_J = A_{J-1} - i_J + j_J$. The terminal base is A_M, with $A_M = A_0$ when no event occurs.

Let $A(s-)$ denote the base immediately before time s, and let $A(s)$ denote the right-continuous post-event state. In particular, $A(\tau_J-) = A_{J-1}$ and $A(\tau_J) = A_J$. Every state of this walk is a base, so the process requires no rounding. Moreover, the number M of exchange events satisfies

$$M \sim \text{Poisson}(\mu), \qquad \mu \triangleq \int_{\varepsilon}^{1} \frac{k}{s}\, ds = k \log(1/\varepsilon), \tag{3.1}$$

[1] This is without loss of generality: loops occur in no independent set and can be deleted in preprocessing without changing the action space or comparator benchmark.

and hence the expected number of exchanges is $k \log(1/\varepsilon)$.

The GKSS algorithm is the objective-dependent specialization of this Poisson base walk obtained as follows. At an event time s with current base A, it computes a maximum-weight base

$$B_s(A) \in \arg\max_{B \in \mathcal{B}} \sum_{j \in B} \partial_j F(s\mathbf{1}_A). \tag{3.2}$$

By Lemma 2.1, there is a bijection $h_{A,B_s(A)} : A \to B_s(A)$ such that $A - i + h_{A,B_s(A)}(i)$ is a base for every $i \in A$. The GKSS exchange distribution chooses i uniformly from A and sets $j = h_{A,B_s(A)}(i)$. Conditional on A and s, this exchange satisfies, for every comparator base O,

$$\mathbb{E}\left[\partial_j F(s\mathbf{1}_A) - \partial_i F(s\mathbf{1}_A) \mid A, s\right] \geq \frac{1}{k}\left(\sum_{o \in O} \partial_o F(s\mathbf{1}_A) - \sum_{i \in A} \partial_i F(s\mathbf{1}_A)\right). \tag{3.3}$$

Since i is uniform on A and $h_{A,B_s(A)}$ is a bijection, j is uniform on $B_s(A)$. The inequality then follows from the maximality of $B_s(A)$ in (3.2). The GKSS rule satisfies this condition at every event time, which is crucial for achieving the $1 - 1/e$ approximation factor.

The bandit algorithm in Section 5 instantiates this Poisson base walk with a different exchange distribution. Because that distribution is chosen without observing the current reward function, it cannot in general satisfy (3.3) at every event. The remainder of this section analyzes the terminal value when the expected exchange gain falls below the right-hand side of (3.3).

Leave-One-Out exchange gain. For a feasible exchange $(i, j) \in \mathcal{E}(A)$, define the *leave-one-out exchange gain* for a given submodular function f and its multilinear extension F, a time $\tau \in [0, 1]$ as

$$G_f(A, \tau; i, j) \triangleq \mathbb{E}_{R \sim \tau(A-i)}[f(R + j) - f(R + i)]. \tag{3.4}$$

Note that the two sets in this difference are feasible: $R + i \subseteq A$ and $R + j \subseteq A - i + j$. We have following two properties of G_f that are used in the analysis of the Poisson process algorithm.

Lemma 3.1. *For every $(i, j) \in \mathcal{E}(A)$,*

$$F(\tau\mathbf{1}_{A-i+j}) - F(\tau\mathbf{1}_A) = \tau G_f(A, \tau; i, j). \tag{3.5}$$

Lemma 3.2. *For every base A, time $\tau \in [0, 1]$, and feasible exchange $(i, j) \in \mathcal{E}(A)$,*

$$G_f(A, \tau; i, j) \geq \partial_j F(\tau\mathbf{1}_A) - \partial_i F(\tau\mathbf{1}_A). \tag{3.6}$$

For a comparator base O, define the local benchmark

$$\zeta_{f,O}(A, s) \triangleq \frac{1}{k}\left(\sum_{o \in O} \partial_o F(s\mathbf{1}_A) - \sum_{a \in A} \partial_a F(s\mathbf{1}_A)\right), \tag{3.7}$$

and the discount weight $w(s) \triangleq s \exp(-(1 - s))$.

Theorem 3.3. *Fix a comparator base O. For each event J, define the discounted deficit*

$$D_J(O) \triangleq w(\tau_J)\left(\zeta_{f,O}(A(\tau_J-), \tau_J) - G_f(A(\tau_J-), \tau_J; e_J)\right). \tag{3.8}$$

Then

$$\mathbb{E}[f(A(1))] \geq \left(1 - e^{-(1-\varepsilon)}\right) f(O) - \mathbb{E}\left[\sum_{J=1}^{M} D_J(O)\right] \tag{3.9}$$

The role of the deficit term distinguishes Theorem 3.3 from the analysis of Ganz Rozenman et al. (2026). Under their objective-dependent exchange rule, the conditional expected deficit at every event is nonpositive, so the expected cumulative deficit is nonpositive and can be discarded from the lower bound. This gives the coefficient $1 - e^{-(1-\varepsilon)}$, which tends to $1 - 1/e$ as ε tends to 0. In contrast, the Poisson base walk defined here permits a general exchange distribution $p_s(\cdot \mid A)$, which need not satisfy the exchange gain condition of GKSS. Our theorem therefore retains the cumulative deficit explicitly and quantifies exactly how its expected value degrades the approximation guarantee. In the online setting, bounding this term converts the learner's cumulative local error into approximation regret.

4 Balanced Fractional Exchanges

The discussion in Section 1.1 suggests that an exponential mixture of comparator-specific Brualdi rules contains more information than the Poisson analysis needs. Its uniform deletion marginal and fractional insertion marginal can instead be represented by a single point $x \in P(\mathcal{M})$. This section establishes the existence of a distribution over feasible exchanges with exactly these two marginals and shows that it can be computed in polynomial time.

Definition 4.1 (Balanced fractional exchange)**.** Fix a base $A \in \mathcal{B}$ and a point $x \in P(\mathcal{M})$. A *balanced fractional exchange* from A with insertion marginal x is a nonnegative matrix $Q_A^x(i,j), i \in A,\ j \in \mathcal{U}$, satisfying (1) $Q_A^x(i,j) > 0 \implies A - i + j \in \mathcal{B}$; (2) $\sum_{j \in \mathcal{U}} Q_A^x(i,j) = 1$ for every $i \in A$; and (3) $\sum_{i \in A} Q_A^x(i,j) = x_j$ for every $j \in \mathcal{U}$. The corresponding probability distribution on feasible exchanges is $q_A^x(i,j) \triangleq \frac{1}{k} Q_A^x(i,j)$.

The row and column sums both have total mass k, so q_A^x is a probability distribution. Its two marginals are

$$\mathbb{P}_{(i,j)\sim q_A^x}(i = a) = \frac{1}{k} \quad (a \in A), \qquad \mathbb{P}_{(i,j)\sim q_A^x}(j = u) = \frac{x_u}{k} \quad (u \in \mathcal{U}). \tag{4.1}$$

The following theorem establishes the existence and efficient construction of balanced fractional exchanges.

Theorem 4.2. *For every base $A \in \mathcal{B}$ and every fractional base $x \in P(\mathcal{M})$, a balanced fractional exchange Q_A^x exists. Given A and x explicitly and access to an independence oracle for $\mathcal{M}$, such an exchange can be constructed in polynomial time.*

5 The Balanced-Exchange Bandit Algorithm

5.1 Algorithm Overview

To bound the Poisson deficit by a linear loss, we first isolate the insertion contribution to an exchange. Fix a monotone submodular function f, its multilinear extension F, a base A, and a time $\tau \in [0,1]$. For a feasible exchange $(i,j) \in \mathcal{E}(A)$, define the insertion marginal

$$I_f(A, \tau; i, j) \triangleq \mathbb{E}_{R \sim \tau(A-i)}[f(R + j) - f(R)]. \tag{5.1}$$

Fix $x \in P(\mathcal{M})$ with strictly positive coordinates and a balanced matrix $Q = Q_A^x$. Define the average insertion marginal and the corresponding insertion loss by

$$\overline{I}_{f,A,\tau,x}(j) \triangleq \frac{1}{x_j} \sum_{i \in A} Q(i,j) I_f(A, \tau; i, j), \qquad \ell_{f,A,\tau,x}(j) \triangleq 1 - \overline{I}_{f,A,\tau,x}(j). \tag{5.2}$$

The following theorem bounds the expected local deficit by an inner product with the insertion loss vector.

Theorem 5.1. *Let $q_A^x \triangleq Q_A^x / k$ be a balanced exchange distribution. For every comparator base O,*

$$\zeta_{f,O}(A, \tau) - \mathbb{E}_{(i,j)\sim q_A^x}[G_f(A, \tau; i, j)] \le \frac{1}{k} \left\langle x - \mathbf{1}_O, \ell_{f,A,\tau,x} \right\rangle. \tag{5.3}$$

By (5.3), controlling linear regret over $P(\mathcal{M})$ bounds the accumulated Poisson deficits. The bandit update estimates the loss coordinate of the inserted element j using the importance weight $1/x_{t,j}$. To keep these weights bounded, we follow Flaxman et al. (2005) and contract the polytope toward a point with strictly positive coordinates.

For every $u \in \mathcal{U}$, choose a base $B^{(u)}$ containing u, and define $x^\circ \triangleq \frac{1}{n} \sum_{u \in \mathcal{U}} \mathbf{1}_{B^{(u)}}$. Then $x^\circ \in P(\mathcal{M})$ and $x_u^\circ \ge \frac{1}{n}$ for every $u \in \mathcal{U}$. For $\delta \in (0,1)$, let

$$K_\delta \triangleq (1-\delta) P(\mathcal{M}) + \delta x^\circ = \{(1-\delta) y + \delta x^\circ : y \in P(\mathcal{M})\}. \tag{5.4}$$

Every $x \in K_\delta$ belongs to $P(\mathcal{M})$ and satisfies $x_u \ge \delta / n$ for every $u \in \mathcal{U}$. For a comparator base O, define its shrunken representative $u_O^\delta \triangleq (1-\delta)\mathbf{1}_O + \delta x^\circ \in K_\delta$.

Algorithm 1: Poisson Balanced-Exchange Bandit

Input: Matroid $\mathcal{M}$; start base A_0; point x°; parameters $\varepsilon, \gamma, \delta, \eta, L$.

Set $x_1 = x^\circ$ and $\mu = k \log(1/\varepsilon)$.;
for $t = 1, \dots, T$ **do**
  Sample $M_t \sim \mathrm{Poisson}(\mu)$.;
  if $M_t > L$ **then**
    Play $S_t = A_0$, observe $f_t(S_t)$, and set $\widehat{L}_t = 0$.;
  else
    Sample and sort event times $\varepsilon < \tau_{t,1} < \cdots < \tau_{t,M_t} \le 1$ according to (D.28).;
    Set $A_{t,0} = A_0$.;
    for $J = 1, \dots, M_t$ **do**
      Compute a balanced exchange $Q^{x_t}_{A_{t,J-1}}$.;
      Draw $(i_{t,J}, j_{t,J}) \sim Q^{x_t}_{A_{t,J-1}}/k$.;
      Set $A_{t,J} = A_{t,J-1} - i_{t,J} + j_{t,J}$.;
    Draw $Z_t \sim \mathrm{Bernoulli}(\gamma)$.;
    if $Z_t = 0$ **then**
      Play the final base $S_t = A_{t,M_t}$, observe $f_t(S_t)$, and set $\widehat{L}_t = 0$.;
    else
      if $M_t = 0$ **then**
        Play $S_t = A_0$, observe $f_t(S_t)$, and set $\widehat{L}_t = 0$.;
      else
        Draw $J_t \sim \mathrm{Unif}\{1, \dots, M_t\}$.;
        Put $A \triangleq A_{t,J_t-1}$, $\tau \triangleq \tau_{t,J_t}$, and $(i,j) \triangleq (i_{t,J_t}, j_{t,J_t})$.;
        Sample $R \sim \tau(A - i)$ and $C_t \sim \mathrm{Bernoulli}(1/2)$.;
        if $C_t = 1$ **then**
          Play $S_t = R + j$, observe $y_t = f_t(S_t)$, and set $z_t = 2(1 - y_t)$.;
        else
          Play $S_t = R$, observe $y_t = f_t(S_t)$, and set $z_t = 2y_t$.;
        Set $\widehat{L}_t = \dfrac{M_t}{\gamma} \dfrac{w(\tau) z_t}{x_{t,j}} e_j$, where e_j is the jth coordinate vector and $w(s) \triangleq s e^{-(1-s)}$.;
  Set $\widetilde{x}_{t+1,u} = x_{t,u} \exp(-\eta \widehat{L}_{t,u})$ for every $u \in \mathcal{U}$.;
  Compute the entropy projection $x_{t+1} \in \operatorname*{arg\,min}_{x \in K_\delta} D_\Phi(x, \widetilde{x}_{t+1})$..;

In round t, the learner runs the Poisson base walk of Section 3 with the balanced exchange distribution $p_s(\cdot \mid A) = Q^{x_t}_A/k$. Let $\varepsilon, \gamma, \delta \in (0,1)$ be the Poisson starting time, exploration probability, and shrinkage parameter, and let $\eta > 0$ be the learning rate. The process has intensity k/s on $[\varepsilon, 1]$ and mean event count $\mu \triangleq k \log(1/\varepsilon)$. An integer cap L bounds the number of simulated exchanges per round; the regret analysis accounts separately for overflow rounds.

On a nonoverflow round, the walk starts from a fixed base A_0 and samples each exchange from $Q^{x_t}_A/k$ at the current base A. Simulating the trajectory requires only the independence oracle and internal randomness, with no value queries. The learner then exploits with probability $1 - \gamma$ by querying the terminal base, or explores with probability γ. If the trajectory contains at least one event, exploration selects an event uniformly and uses one feasible query to estimate the aggregate linear loss. When the event count is zero or exceeds L, the learner queries A_0 and sets $\widehat{L}_t = 0$. Algorithm 1 gives the full procedure. We next introduce the bandit estimator $\widehat{L}_t$ used in its update.

Bandit estimator. Condition on the selected pre-event base A, time τ, and exchange (i,j). Draw $R \sim \tau(A - i)$, so that both R and $R + j$ are feasible. A fair coin determines which set to query: querying $R + j$ gives the statistic $2(1 - f_t(R + j))$, while querying R gives $2 f_t(R)$. Its conditional mean is $1 - I_{f_t}(A, \tau; i, j)$, the insertion loss for the sampled exchange. Let

$$\ell_{t,J} \triangleq \ell_{f_t, A_{t,J-1}, \tau_{t,J}, x_t} \in [0,1]^{\mathcal{U}} \tag{5.5}$$

be the local loss vector at event J. By (5.3), its inner product with $x_t - \mathbf{1}_O$ bounds the expected local deficit relative to any comparator base O, up to the factor $1/k$. With the discount $w(s) \triangleq s e^{-(1-s)}$,

the aggregate loss is

$$L_t \triangleq \frac{\mathbf{1}\{M_t \le L\}}{k} \sum_{J=1}^{M_t} w(\tau_{t,J})\ell_{t,J}. \tag{5.6}$$

On an exploration round with $1 \le M_t \le L$, we estimate L_t by

$$\widehat{L}_t = \frac{M_t}{\gamma} \frac{w(\tau_{t,J_t})z_t}{x_{t,j_{t,J_t}}} e_{j_{t,J_t}}, \tag{5.7}$$

and set $\widehat{L}_t = 0$ otherwise. After the query, the learner performs the standard exponential weights update and returns to K_δ by the entropy projection displayed in Algorithm 1. *All vector divisions and logarithms in the algorithm are coordinatewise.* The projection restores the matroid base constraints and the uniform positivity condition.

Let $\mathcal{H}_t$ be the sigma-field generated by the interaction history through round $t-1$, the predictable state x_t, and f_t, before the fresh round-t randomization. We have the following guarantees for the estimator $\widehat{L}_t$.

Proposition 5.2 (Estimator guarantees)**.** *Every round of Algorithm 1 makes exactly one feasible value query. Moreover,*

$$\mathbb{E}[\widehat{L}_t \mid \mathcal{H}_t] = \mathbb{E}[L_t \mid \mathcal{H}_t], \qquad \mathbb{E}\left[\sum_{u\in\mathcal{U}} x_{t,u}\widehat{L}_{t,u}^2 \,\middle|\, \mathcal{H}_t\right] \le \frac{4n(\mu^2+\mu)}{k\gamma}.$$

Computational complexity. We will set the event cap to be $L = O(k\log(1/\varepsilon) + \log T)$ and each balanced exchange is computed in polynomial time by Theorem 4.2, thus the per-round computational cost is polynomial in n, k, and $\log T$. The exact entropy projection is a separable strictly convex minimization problem over a shrunken matroid base polytope and can be implemented in polynomial time by Suehiro et al. (2012).

5.2 Regret Analysis

Theorem 5.3. *Let $T \ge 2$, set $\varepsilon \triangleq 1/T$ and $\mu \triangleq k\log T$. If $n(\mu^2+\mu)\log(en) < T$, run Algorithm 1 with*

$$L \triangleq \lceil 2e(\mu + \log T)\rceil, \qquad \delta \triangleq \min\left\{\frac{1}{2}, \frac{1}{\mu T}\right\},$$

and

$$\gamma \triangleq \left(\frac{n(\mu^2+\mu)\log(en)}{T}\right)^{1/3}, \qquad \eta \triangleq \sqrt{\frac{k^2\gamma\log(en)}{2n(\mu^2+\mu)T}}.$$

Then

$$\begin{aligned} R_{1-1/e}(T) &= O\Big(\left[n\big((k\log T)^2 + k\log T\big)\log(en)\right]^{1/3} T^{2/3} + 1\Big) \\ &= \widetilde{O}\Big(n^{1/3}k^{2/3}T^{2/3}\Big). \end{aligned} \tag{5.8}$$

If $n(\mu^2+\mu)\log(en) \ge T$, the same bound is achieved by playing any fixed base.

Proof sketch. The proof starts from Theorem 3.3, which bounds the approximation error of the uncapped Poisson walk by its cumulative exchange deficits. Theorem 5.1 then bounds these deficits by linear losses over the base polytope. We apply negative-entropy online mirror descent on K_δ (Orabona, 2026, Chapter 6): the conditional unbiasedness in Proposition 5.2 transfers its expected regret bound to L_t, and the weighted second-moment bound controls the learning error. We account separately for comparator shrinkage and the reward lost on exploration and overflow rounds. Balancing these terms gives the parameter choices and regret bound below; the intermediate linear-regret bound and the full proof are in the appendix. □

AI Use Statement

Generative LLMs was used as an interactive aid in the development of the proofs. It also helped with the organization and drafting of the proofs, which are verified and polished by the authors. The idea and technical framework are proposed by the human authors.

REFERENCES

Vaneet Aggarwal. Adversarial resilience of Poisson-process submodular maximization over matroids: From robust offline optimization to full-bandit learning. *arXiv preprint arXiv:2608.12134*, 2026. URL https://arxiv.org/abs/2608.12134v1.

Peter Auer, Nicolò Cesa-Bianchi, Yoav Freund, and Robert E. Schapire. The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, 32(1):48–77, 2002. doi: 10.1137/S0097539701398375.

Richard A. Brualdi. Comments on bases in dependence structures. *Bulletin of the Australian Mathematical Society*, 1(2):161–167, 1969.

Niv Buchbinder and Moran Feldman. Deterministic algorithm and faster algorithm for submodular maximization subject to a matroid constraint. In *Proceedings of the 65th IEEE Annual Symposium on Foundations of Computer Science*, pp. 700–712, 2024.

Gruia Călinescu, Chandra Chekuri, Martin Pál, and Jan Vondrák. Maximizing a monotone submodular function subject to a matroid constraint. *SIAM Journal on Computing*, 40(6):1740–1766, 2011.

Adrian Rivera Cardoso and Rachel Cummings. Differentially private online submodular minimization. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pp. 1650–1658, 2019.

Chandra Chekuri, Jan Vondrák, and Rico Zenklusen. Dependent randomized rounding via exchange properties of combinatorial structures. In *Proceedings of the 51st IEEE Annual Symposium on Foundations of Computer Science*, pp. 575–584, 2010.

Lin Chen, Andreas Krause, and Amin Karbasi. Interactive submodular bandit. In *Advances in Neural Information Processing Systems 30*, pp. 141–152, 2017.

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

Abraham D. Flaxman, Adam Tauman Kalai, and H. Brendan McMahan. Online convex optimization in the bandit setting: Gradient descent without a gradient. In *Proceedings of the Sixteenth Annual ACM–SIAM Symposium on Discrete Algorithms*, pp. 385–394, 2005.

Fares Fourati, Vaneet Aggarwal, Christopher John Quinn, and Mohamed-Slim Alouini. Randomized greedy learning for non-monotone stochastic submodular maximization under full-bandit feedback. In *Proceedings of the 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pp. 7455–7471, 2023.

Fares Fourati, Christopher John Quinn, Mohamed-Slim Alouini, and Vaneet Aggarwal. Combinatorial stochastic-greedy bandit. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 12052–12060, 2024.

Amit Ganz Rozenman, Ariel Kulik, Roy Schwartz, and Mohit Singh. A Poisson process for submodular maximization. In *Proceedings of the 58th Annual ACM Symposium on Theory of Computing*, pp. 2152–2163, 2026.

Elad Hazan and Satyen Kale. Online submodular minimization. *Journal of Machine Learning Research*, 13:2903–2922, 2012.

Shinji Ito. Revisiting online submodular minimization: Gap-dependent regret bounds, best of both worlds and adversarial robustness. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 9678–9694, 2022.

David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 137–146, 2003. doi: 10.1145/956750.956769.

Andreas Krause, Ajit Singh, and Carlos Guestrin. Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research*, 9:235–284, 2008.

George L. Nemhauser, Laurence A. Wolsey, and Marshall L. Fisher. An analysis of approximations for maximizing submodular set functions—I. *Mathematical Programming*, 14:265–294, 1978.

Rad Niazadeh, Negin Golrezaei, Joshua R. Wang, Fransisca Susan, and Ashwinkumar Badanidiyuru. Online learning via offline greedy algorithms: Applications in market design and optimization. *Management Science*, 69(7):3797–3817, 2023.

Guanyu Nie, Mridul Agarwal, Abhishek Kumar Umrawal, Vaneet Aggarwal, and Christopher John Quinn. An explore-then-commit algorithm for submodular maximization under full-bandit feedback. In *Proceedings of the 38th Conference on Uncertainty in Artificial Intelligence*, volume 180 of *Proceedings of Machine Learning Research*, pp. 1541–1551, 2022.

Guanyu Nie, Yididiya Y. Nadew, Yanhui Zhu, Vaneet Aggarwal, and Christopher John Quinn. A framework for adapting offline algorithms to solve combinatorial multi-armed bandit problems with bandit feedback. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 26166–26198, 2023.

Guanyu Nie, Vaneet Aggarwal, and Christopher John Quinn. Stochastic k-submodular bandits with full bandit feedback. In *Proceedings of the 24th International Conference on Autonomous Agents and Multiagent Systems*, pp. 2696–2698, 2025.

Francesco Orabona. Online learning: A modern introduction using convex optimization, 2026. arXiv:1912.13213, version 10.

Matthew Streeter and Daniel Golovin. An online algorithm for maximizing submodular functions. In *Advances in Neural Information Processing Systems 21*, pp. 1577–1584, 2008.

Matthew Streeter, Daniel Golovin, and Andreas Krause. Online learning of assignments. In *Advances in Neural Information Processing Systems 22*, pp. 1794–1802, 2009.

Daiki Suehiro, Kohei Hatano, Shuji Kijima, Eiji Takimoto, and Kiyohito Nagano. Online prediction under submodular constraints. In *Proceedings of the 23rd International Conference on Algorithmic Learning Theory*, volume 7568 of *Lecture Notes in Computer Science*, pp. 260–274. Springer, 2012. doi: 10.1007/978-3-642-34106-9_22.

Artin Tajdini, Lalit Jain, and Kevin Jamieson. Nearly minimax optimal submodular maximization with bandit feedback. In *Advances in Neural Information Processing Systems 37*, 2024. doi: 10.52202/079017-3051.

Jan Vondrák. Optimal approximation for the submodular welfare problem in the value oracle model. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing*, pp. 67–74, 2008.

Zongqi Wan, Jialin Zhang, Wei Chen, Xiaoming Sun, and Zhijie Zhang. Bandit multi-linear DR-submodular maximization and its applications on adversarial submodular bandits. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 35491–35524, 2023.

Yisong Yue and Carlos Guestrin. Linear submodular bandits and their application to diversified retrieval. In *Advances in Neural Information Processing Systems 24*, pp. 2483–2491, 2011.

Mingrui Zhang, Lin Chen, Hamed Hassani, and Amin Karbasi. Online continuous submodular maximization: From full-information to bandit feedback. In *Advances in Neural Information Processing Systems 32*, pp. 9210–9220, 2019.