Minimax Regret for Cascading Bandits

Daniel Vial
UT Austin & UIUC
dvial@utexas.edu

Sujay Sanghavi
UT Austin & Amazon
sanghavi@mail.utexas.edu

Sanjay Shakkottai
UT Austin
sanjay.shakkottai@utexas.edu

R. Srikant
UIUC
rsrikant@illinois.edu

Abstract

Cascading bandits is a natural and popular model that frames the task of learning to rank from Bernoulli click feedback in a bandit setting. For the case of unstructured rewards, we prove matching upper and lower bounds for the problem-independent (i.e., gap-free) regret, both of which strictly improve the best known. A key observation is that the hard instances of this problem are those with small mean rewards, i.e., the small click-through rates that are most relevant in practice. Based on this, and the fact that small mean implies small variance for Bernoullis, our key technical result shows that variance-aware confidence sets derived from the Bernstein and Chernoff bounds lead to optimal algorithms (up to log terms), whereas Hoeffding-based algorithms suffer order-wise suboptimal regret. This sharply contrasts with the standard (non-cascading) bandit setting, where the variance-aware algorithms only improve constants. In light of this and as an additional contribution, we propose a variance-aware algorithm for the structured case of linear rewards and show its regret strictly improves the state-of-the-art.

1 Introduction

The cascading click model describes users interacting with ranked lists, such as search results or online advertisements [Craswell et al., 2008]. In this model, there is a set of items $[L] = \{1, \dots, L\}$. The user is given a list of K items $\mathbf{A} = (\mathbf{a}_1, \dots, \mathbf{a}_K)$, sequentially examines the list, and clicks on the first attractive item (if any). If a click occurs, the user leaves without examining the subsequent items. It is assumed that the e-th item has an attraction probability $\bar{w}(e)$ (which we also call the mean reward), and that the random (Bernoulli) clicks are conditionally independent given $\mathbf{A}$.

Cascading bandits, introduced concurrently by Kveton et al. [2015a] and Combes et al. [2015], are a sequential learning version of the model where the mean rewards $\{\bar{w}(e)\}_{e=1}^{L}$ are initially unknown. At each round $t \in [n]$, the learner chooses an action, which is a list of items $\mathbf{A}_t = (\mathbf{a}_1^t, \dots, \mathbf{a}_K^t)$. As in the basic click model, the user scans the list and clicks on the first attractive item (if any). Thus, if the $\mathbf{C}_t$-th item is clicked, the learner knows that the user was not attracted to $\{\mathbf{a}_k^t\}_{k=1}^{\mathbf{C}_t - 1}$ but was attracted to $\mathbf{a}_{\mathbf{C}_t}^t$. However, the learner receives *no* feedback on the items $\{\mathbf{a}_k^t\}_{k=\mathbf{C}_t+1}^{K}$ that the user did not examine before leaving. The objective for the learner is to choose the sequence $\{\mathbf{A}_t\}_{t=1}^{n}$ to maximize the expected number of clicks, or equivalently, minimize the regret defined in (1).

This work provides problem-independent (i.e., gap-free) regret bounds for cascading bandits that strictly improve the state-of-the-art. In the case of unstructured rewards, our results provide the first minimax-optimal regret bounds (up to log terms). Our key insight is that, compared to the standard bandit problem, the reward variance plays an outsized role in the gap-free analysis. In particular, we show that *for cascading bandits, the worst-case problem instances are those with low mean rewards –*

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

Table 1: Problem-independent upper bounds and minimax lower bounds for cascading bandits with L total items, K recommended items, horizon n, and feature dimension d. The tabular columns assume unstructured mean reward. The linear column assumes reward is linear in unit-norm features and does not suppress any L dependencies in the $\tilde{O}(\cdot)$ notation. See Appendix A for other related work.

	Tabular case		**Linear case**
Paper	**Upper bound**	**Lower bound**	**Upper bound**
Zong et al. [2016]	none	none	$\tilde{O}(\sqrt{nd^2K^2})$
Wang and Chen [2017]	$\tilde{O}(\sqrt{nLK})$	none*	none
Lattimore et al. [2018]	$\tilde{O}(\sqrt{nLK^3})$	none*	none
Li and Zhang [2018]	none	none	$\tilde{O}(\sqrt{nd^2K})^{\S}$
Zhong et al. [2021]	$\tilde{O}(\sqrt{nLK})$	$\Omega(\sqrt{nL/K})^{\dagger}$	$\tilde{O}(\sqrt{nd^2K^3\min\{d, \log L\}})$
Kveton et al. [2022]	$\tilde{O}(\sqrt{nLK})^{\ddagger}$	none	none
Ours	$\tilde{O}(\sqrt{nL})$ (Thm 2)	$\Omega(\sqrt{nL})^{\dagger}$ (Thm 1)	$\tilde{O}(\sqrt{nd(d+K)})^{\S}$ (Thm 4)

*These papers contain minimax lower bounds but for click models that are distinct from cascading bandits.
†These bounds assume L is large compared to K and n is large compared to L and K.
‡This bound holds for a Bayesian notion of regret and includes a dependence on the prior not shown above.
§These bounds assume n is large compared to d and K to suppress some additive $o(\sqrt{n})$ terms.

namely, $\bar{w}(e) \leq 1/K$, where the variance $\bar{w}(e)(1 - \bar{w}(e))$ is also small. We emphasize that these worst-case instances are not pathological; rather, they model the low click-through rates that prevail in practice [Richardson et al., 2007]. Further, we argue that *adapting to low variance is crucial to cope with the worst-case instances*. Put differently, algorithms should be *variance-aware*, i.e., more exploitative when the variance is small. We provide the intuition behind this key insight in Section 1.2 and show how to formalize it with a proof sketch in Section 5.

More specifically, our first goal is to establish upper bounds on the problem-independent regret (i.e., the maximum over $\bar{w}$) for cascading bandit algorithms, as well as minimax lower bounds (i.e., the infimum over algorithms of their problem-independent regret). As shown in Table 1, the tightest existing such bounds are $\tilde{O}(\sqrt{nLK})$ and $\Omega(\sqrt{nL/K})$, respectively. What is surprising is that, not only is there a gap between the bounds, but they *increase* and *decrease* in K, respectively. In other words, the following fundamental question is unresolved: *as K grows, does the problem become harder* (as suggested by the upper bound) *or easier* (as suggested by the lower bound)*?*

As discussed by Zhong et al. [2021], this question lacks an obvious answer. On the one hand, larger K means that the learner needs to identify more good items, which hints at a harder problem. On the other hand, the learner receives more feedback as K grows, which intuitively makes the problem easier. As we show later, variance-aware algorithms are the key to resolving this tradeoff.

In addition to this basic version of the model – hereafter, the *tabular case*, where no structure is assumed for the mean rewards – we are also interested in the *linear case*, where $\bar{w}(e) = \langle \phi(e), \theta \rangle$ for some known feature map $\phi : [L] \to \mathbb{R}^d$ and unknown parameter vector $\theta \in \mathbb{R}^d$. A second goal of this work is to apply our insights from the tabular case to the linear one, in hopes of improving the best known upper bound $\tilde{O}(\sqrt{nd^2K})$ (see Table 1).

1.1 Main contributions

Tabular case. First, we show no algorithm can achieve $o(\sqrt{nL})$ regret uniformly across $\bar{w}$, assuming L is large compared to K and n is large compared to L (see Theorem 1). Next, we consider three algorithms: `CascadeKL-UCB`, `CascadeUCB-V`, and `CascadeUCB1` (the first and third are due to Kveton et al. [2015a]; `CascadeUCB-V` is new). All three rank the L items using upper confidence bounds (UCBs) and choose $\mathbf{A}_t$ as the K highest ranked items. They differ in the choice of UCB. As the names suggest, `CascadeKL-UCB` and `CascadeUCB-V` use `KL-UCB` [Garivier and Cappé, 2011, Maillard et al., 2011, Cappé et al., 2013] and `UCB-V` [Audibert et al., 2009], respectively. Both are variance-aware, in the sense that their respective UCBs are derived from the Chernoff and Bernstein inequalities. We show that both algorithms have near-optimal regret $\tilde{O}(\sqrt{nL})$ for any $\bar{w}$ (see Theorem 2). In contrast, `CascadeUCB1` relies on the Hoeffing-style `UCB1` [Auer et al., 2002], and we show that it suffers suboptimal regret $\Omega(\sqrt{nLK})$ on some $\bar{w}$ (see Theorem 3).

In summary, we prove (i) the minimax regret is $\tilde{\Theta}(\sqrt{nL})$ for cascading bandits, (ii) the variance-aware algorithms `CascadeKL-UCB` and `CascadeUCB-V` are minimax-optimal up to log terms, and (iii) the variance-unaware algorithm `CascadeUCB1` is decidely *sub*optimal. Moreover, note from Table 1 that we strictly improve *both* the upper and lower bounds for this problem.

Discussion. There are two surprising aspects to these results. First, the minimax bound $\tilde{\Theta}(\sqrt{nL})$ shows that (in a worst-case sense) the number of recommended items K plays *no role*.[1] In other words, the aforementioned tradeoff (identifying more good items but receiving more feedback) is perfectly balanced when the correct algorithm (e.g., `CascadeKL-UCB`) is employed.

The second (and arguably more surprising) aspect is that `CacadeKL-UCB` and `CascadeUCB-V` are optimal but `CascadeUCB1` is not. This stands in contrast to the standard L-armed bandit problem, where the analogous algorithms `KL-UCB`, `UCB-V`, and `UCB1` all achieve the minimax regret $\tilde{\Theta}(\sqrt{nL})$, and the main advantage of the former two is only to improve the constants in the gap-dependent bounds. We discuss the intuition behind this contrast in Section 1.2.

Linear case. Motivated by these findings, we also propose a variance-aware algorithm for the linear case called `CascadeWOFUL`.[2] `CascadeWOFUL` proceeds in two steps. First, we use the Hoeffding-style UCBs of Abbasi-Yadkori et al. [2011] to upper bound the mean rewards $\bar{w}(e)$, and thus the Bernoulli variances $\bar{w}(e)(1-\bar{w}(e))$, with high probability. Second, we use these Hoeffding UCBs as proxies for the true variances in the `WOFUL` algorithm of Zhou et al. [2021], which computes a variance-weighted estimate of θ that enjoys Bernstein-style concentration. In Theorem 4, we show the regret of `CascadeWOFUL` is $\tilde{O}(\sqrt{nd(d+K)})$ for large n, which improves existing bounds by a factor of at least $\sqrt{\min\{d, K\}}$ (see Table 1).

1.2 Why do variance-aware algorithms succeed but variance-unaware algorithms fail?

To answer this question, we focus on the tabular case and contrast the standard L-armed bandit setting with the cascading one. We first recall how the $\tilde{O}(\sqrt{nL})$ bound for L-armed bandits is derived. For simplicity, we restrict to instances with $\bar{w}(1) = p$ and $\bar{w}(2) = \cdots = \bar{w}(L) = p - \Delta$ for some $p \in (0,1)$ and $\Delta \in (0,p)$. In this case, `UCB1` plays each of the $L-1$ suboptimal items $1/\Delta^2$ times, up to constants and log factors [Auer et al., 2002]. Each such play costs regret Δ, for a total regret that scales as L/Δ. Alternatively, regret can simply be bounded by Δn, since the n plays incur at most Δ regret each. Combining the bounds gives $\min\{L/\Delta, \Delta n\}$. The worst case occurs when the two bounds are equal, i.e., when $\Delta = \sqrt{L/n}$, which implies $\sqrt{nL}$ regret.

In contrast, `UCB-V` plays each suboptimal item σ^2/Δ^2 times (in an order sense), where $\sigma^2 \leq p$ is the variance of the Bernoulli$(p-\Delta)$ reward [Audibert et al., 2009]. Therefore, the argument of the previous paragraph shows that regret grows as $\min\{Lp/\Delta, \Delta n\}$. Here the worst case occurs when p is non-vanishing and $\Delta = \sqrt{L/n}$, which gives the same regret scaling as `UCB1`. A similar argument holds for `KL-UCB`, because the number of plays grows as $1/d(p-\Delta, p)$ [Cappé et al., 2013] and one can show $d(p-\Delta, p) = \Omega(\Delta^2/p)$ (see Claim 1 in Appendix G).

The analysis is more complicated for cascading bandits, because regret is nonlinear in the mean rewards and the amount of feedback is random. To oversimplify things, we draw an analogy with the above and assume $\bar{w}(1) = \cdots = \bar{w}(K) = p$ and $\bar{w}(K+1) = \cdots = \bar{w}(L) = p - \Delta$. In this case, `CascadeUCB1` similarly plays the $L-K$ suboptimal items $1/\Delta^2$ times each, which costs L/Δ regret when $L \gg K$. However, we can no longer bound regret by Δn, because the total number of plays depends on the random number of items the user examines at each round, which is roughly $\min\{1/p, K\}$ (the mean of Geometric(p) random variable, truncated to the maximum K). Thus, the Δn bound inflates to $\Delta n \min\{1/p, K\}$, which gives $\min\{L/\Delta, \Delta n \min\{1/p, K\}\}$ regret. For $p \leq 1/K$ and $\Delta = \sqrt{L/(nK)}$, this yields the best known bound $\sqrt{nLK}$. We emphasize that, unlike the previous paragraph, *the worst case here occurs when p (i.e., the click-through rate) is small*.

On the other hand, the analogous bound for `CascadeUCB-V` and `CascadeKL-UCB` scales as $\min\{Lp/\Delta, \Delta n \min\{1/p, K\}\}$. Crucially, the factor of p in the first term – which arises due to the variance-aware nature of the algorithms – offsets the factor $1/p$ in the second term. Thus, in the hard

[1] The $\tilde{O}(\cdot)$ notation hides $\log K$ terms, but they can be bounded by $\log L$ while retaining $\tilde{O}(\sqrt{nL})$ regret.

[2] `WOFUL` stands for weighted optimism in the face of uncertainty for linear bandits.

case $p \leq 1/K$, the bound becomes $\min\{L/(\Delta K), \Delta n K\}$. Here the worst case is $\Delta = \sqrt{L/(nK^2)}$, which yields the minimax regret $\sqrt{nL}$ that we establish in Theorems 1 and 2.

2 Preliminaries

In this section, we precisely formulate our problem. A cascading bandit instance is defined by the triple $(L, K, \bar{w})$, where $L \in \mathbb{N}$ is the total number of items, $K \in [L] = \{1, \ldots, L\}$ is the number of items the learner displays to the user at each round, and $\bar{w} \in [0,1]^L$ is the vector of attraction probabilities. We define a sequential game as follows. At each round $t \in [n]$, the learner chooses an *action* $\mathbf{A}_t = (\mathbf{a}_1^t, \ldots, \mathbf{a}_K^t)$, where $\mathbf{a}_k^t \in [L]$ and $\mathbf{a}_k^t \neq \mathbf{a}_{k'}^t$ when $k \neq k'$ (i.e., $\mathbf{A}_t$ is an ordered list of K distinct items). The user sequentially examines this list, clicks on the first item that attracts them, and stops examining items after clicking. Mathematically, we denote the first attractive item by $\mathbf{C}_t = \inf\{k \in [K] : \mathbf{w}_t(\mathbf{a}_k^t) = 1\}$, where $\mathbf{w}_t(e) \sim \text{Bernoulli}(\bar{w}(e))$ for each $e \in [L]$. If no item is clicked, i.e., if $\mathbf{w}_t(\mathbf{a}_k^t) = 0$ for all k, we set $\mathbf{C}_t = \infty$. Note the learner only observes the realizations corresponding to the items that the user examined, i.e., $\{\mathbf{w}_t(\mathbf{a}_k^t) : k \in [\min\{\mathbf{C}_t, K\}]\}$.

We denote by $\mathcal{H}_t = \cup_{s=1}^{t-1}\{\mathbf{A}_s\} \cup \{\mathbf{w}_s(\mathbf{a}_k^s) : k \in [\min\{\mathbf{C}_s, K\}]\}$ the *history* of actions and observations before time t.[3] We let $\mathbb{P}_t(\cdot) = \mathbb{P}(\cdot|\mathcal{H}_t \cup \{\mathbf{A}_t\})$ and $\mathbb{E}_t[\cdot] = \mathbb{E}[\cdot|\mathcal{H}_t \cup \{\mathbf{A}_t\}]$ denote conditional probability and expectation given the history and current action. In the cascading bandit model, it is assumed that the feedback $\{\mathbf{w}_t(\mathbf{a}_k^t)\}_{k=1}^K$ (i.e., the presence or absence of clicks) is conditionally independent given $\mathcal{H}_t$ and $\mathbf{A}_t$. Therefore, the conditional click probability is

$$\mathbb{P}_t(\mathbf{C}_t < \infty) = 1 - \mathbb{P}_t(\mathbf{C}_t = \infty) = 1 - \mathbb{P}_t(\cap_{k=1}^K \{\mathbf{w}_t(\mathbf{a}_k^t) = 0\}) = 1 - \prod_{k=1}^K (1 - \bar{w}(\mathbf{a}_k^t)).$$

Mappings from $\mathcal{H}_t$ to $\mathbf{A}_t$ are called *policies*. Let Π be the set of all policies. Given an instance $\bar{w} \in [0,1]^L$, a policy $\pi \in \Pi$, and a horizon $n \in \mathbb{N}$, the expected number of clicks is

$$\mathbb{E}_{\pi,\bar{w}}\left[\sum_{t=1}^n \mathbb{1}(\mathbf{C}_t < \infty)\right] = \mathbb{E}_{\pi,\bar{w}}\left[\sum_{t=1}^n \mathbb{P}_t(\mathbf{C}_t < \infty)\right] = \mathbb{E}_{\pi,\bar{w}}\left[\sum_{t=1}^n \left(1 - \prod_{k=1}^K (1 - \bar{w}(\mathbf{a}_k^t))\right)\right],$$

where $\mathbb{1}$ is the indicator function. Let $A^* = (a_1^*, \ldots, a_K^*)$ be any action $A = (a_1, \ldots, a_K)$ that maximizes the click probability $1 - \prod_{k=1}^K (1 - \bar{w}(a_k))$. Our goal is to minimize *regret*, which is the expected difference in the number of clicks between π and the policy that always plays A^*, i.e.,

$$R_{\pi,\bar{w}}(n) = \mathbb{E}_{\pi,\bar{w}}\left[\sum_{t=1}^n \left(\prod_{k=1}^K (1 - \bar{w}(\mathbf{a}_k^t)) - \prod_{k=1}^K (1 - \bar{w}(a_k^*))\right)\right]. \tag{1}$$

3 Results for the tabular case

We can now state our tabular results (the proofs are discussed in Section 5). First, we have a minimax lower bound showing no algorithm can achieve $o(\sqrt{nL})$ uniformly over the mean rewards $\bar{w}$.

Theorem 1. *Suppose $N \triangleq L/K \in \{4, 5, \ldots\}$ and $n \geq L$. Then for any policy $\pi \in \Pi$, there exists a mean reward vector $\bar{w} \in [0,1]^L$ such that $R_{\pi,\bar{w}}(n) = \Omega(\sqrt{nL})$.*

Remark 1. *The proof of Theorem 1 is essentially a reduction to Lattimore et al. [2018]'s lower bound for the so-called document-based click model. Their proof and ours both use the assumption $L/K \in \mathbb{N}$ to simplify the analysis, which involves partitioning the L items into K subsets of size L/K each. When $L/K \notin \mathbb{N}$, one of the subsets will have fewer items, which makes the analysis more cumbersome; however, this does not fundamentally alter either result.*

Remark 2. *The theorem also requires $L \geq 4K$, which is not very restrictive since $L \gg K$ in typical applications. However, this assumption does eliminate an interesting analytical regime, namely, when $K \to L$. We conjecture the minimax lower bound is $\Omega(\sqrt{n(L-K)})$ in this case, since any algorithm obtains zero regret when $K = L$ and there are no suboptimal items.*

[3] Our notation mostly follows Kveton et al. [2015a], but we clarify that their definition of $\mathcal{H}_t$ includes $\mathbf{A}_t$.

We next consider the algorithms `CascadeKL-UCB` and `CascadeUCB1` from Kveton et al. [2015a], along with a new one called `CascadeUCB-V`. All follow a similar template, which is given in Algorithm 1. This is a natural generalization of upper confidence bound (UCB) algorithms from the standard L-armed bandit setting. At each round $t \in [n]$, it computes UCBs $\mathbf{U}_t(e)$ in a manner to be specified shortly, then chooses $\mathbf{A}_t$ as the K items with the highest UCBs (in order of UCB). After observing the click feedback $\mathbf{C}_t$, the algorithm increments the number of observations $\mathbf{T}_t(e)$ and updates the empirical mean $\hat{\mathbf{w}}_{\mathbf{T}_t(e)}(e)$ for each item e that the user examined.

Algorithm 1: General UCB algorithm for tabular cascading bandits

Initialize number of observations $\mathbf{T}_t(e) = 0$ and empirical mean $\hat{\mathbf{w}}_0(e) = 0$ for each $e \in [L]$
for $t = 1, \ldots, n$ **do**
  Compute $\mathbf{U}_t(e)$ for each $e \in [L]$ (by (2), (3), and (4) for `CascadeKL-UCB`, `CascadeUCB-V`, and `CascadeUCB1`, respectively)
  for $k = 1, \ldots, K$ **do**
    Let $\mathbf{a}_k^t = \arg\max_{e \in [L] \setminus \{\mathbf{a}_i^t\}_{i=1}^{k-1}} \mathbf{U}_t(e)$ be the item with the k-th highest UCB
  Play $\mathbf{A}_t = (\mathbf{a}_1^t, \ldots, \mathbf{a}_K^t)$ and observe $\mathbf{C}_t = \inf\{k \in [K] : \mathbf{w}_t(\mathbf{a}_k^t) = 1\}$ (where $\inf \emptyset = \infty$)
  Let $\mathbf{T}_t(e) = \mathbf{T}_{t-1}(e)$ for each $e \in [L]$
  for $k = 1, \ldots, \min\{\mathbf{C}_t, K\}$ **do**
    $e \leftarrow \mathbf{a}_k^t$, $\mathbf{T}_t(e) \leftarrow \mathbf{T}_t(e) + 1$, $\hat{\mathbf{w}}_{\mathbf{T}_t(e)}(e) \leftarrow (\mathbf{T}_{t-1}(e)\hat{\mathbf{w}}_{\mathbf{T}_{t-1}(e)}(e) + \mathbb{1}(\mathbf{C}_t = k))/\mathbf{T}_t(e)$

For `CascadeKL-UCB`, the UCBs $\mathbf{U}_t(e)$ are computed as follows:

$$\mathbf{U}_t(e) = \max\{u \in [0,1] : d(\hat{\mathbf{w}}_{\mathbf{T}_{t-1}(e)}(e), u) \leq \log(f(t))/\mathbf{T}_{t-1}(e)\}, \tag{2}$$

where $f(t) = t(\log t)^3$ and $d(p,q) = p\log(p/q) + (1-p)\log((1-p)/(1-q))$ is the relative entropy between Bernoullis with means $p, q \in [0,1]$. The set in (2) is a confidence interval for $\bar{w}(e)$, which is derived from the Chernoff bound. For `CascadeUCB-V`, the UCBs are instead given by

$$\mathbf{U}_t(e) = \hat{\mathbf{w}}_{\mathbf{T}_{t-1}(e)}(e) + \sqrt{4\hat{\mathbf{v}}_{\mathbf{T}_{t-1}(e)}(e)\log(t)/\mathbf{T}_{t-1}(e)} + 6\log(t)/\mathbf{T}_{t-1}(e), \tag{3}$$

where $\hat{\mathbf{v}}_s(e) = \hat{\mathbf{w}}_s(e)(1 - \hat{\mathbf{w}}_s(e))$ is the empirical variance from s observations of item e. This UCB is derived from the coarser – but crucially, still variance-aware – Bernstein inequality. Finally, for `CascadeUCB1`, the UCBs are derived from the Hoeffing bound and computed as follows:

$$\mathbf{U}_t(e) = \hat{\mathbf{w}}_{\mathbf{T}_{t-1}(e)}(e) + c_{t,\mathbf{T}_{t-1}(e)}, \quad \text{where} \quad c_{t,s} = \sqrt{1.5\log(t)/s}. \tag{4}$$

We can now show the variance-aware UCBs are nearly optimal, while `CascadeUCB1` is suboptimal.

Theorem 2. *Suppose π is `CascadeKL-UCB` or `CascadeUCB-V`, i.e., the policy from Algorithm 1 with the UCBs given by (2) or (3). Then $R_{\pi,\bar{w}}(n) = \tilde{O}(\sqrt{nL})$ for any $\bar{w} \in [0,1]^L$.*

Remark 3. *The reader may wonder why we proposed `CascadeUCB-V`, since the `CascadeKL-UCB` bound is enough to establish the minimax regret. The main reason is to demonstrate that variance-awareness alone (no additional information encoded by `KL-UCB`) is enough to achieve the optimal regret, which helps motivate our linear algorithm. Furthermore, we show empirically in Section 6 that, while `CascadeUCB-V` is inferior to `CascadeKL-UCB` in terms of regret, its closed form nature leads to quicker computation, while still improving the regret of `CascadeUCB1`.*

Theorem 3. *Suppose $n \geq \max\{LK, 49K^4\}$, $L \geq 800K$, and π is `CascadeUCB1`, i.e., the policy from Algorithm 1 with the UCBs given by (4). Then $R_{\pi,\bar{w}}(n) = \Omega(\sqrt{nLK})$ for some $\bar{w} \in [0,1]^L$.*

Remark 4. *The assumed bounds on n and L simplify the calculations and can be improved (see Remark 9 in Appendix J). The main point of Theorem 3 is to show that, unlike the variance-aware algorithms, `CascadeUCB1` cannot satisfy the conclusion of Theorem 2 for* all *choices of n and L.*

4 Results for the linear case

In light of the previous section, we seek a variance-aware algorithm for the linear case. Our method is based on the `WOFUL` algorithm of Zhou et al. [2021], which was designed for the standard linear

bandit setting (the case $K = 1$). In this section, we review `WOFUL`, discuss how to overcome its limitations in the cascading setting, explain our `CascadeWOFUL` algorithm, and bound its regret.

Existing algorithm. In Section 4.3 of their work, Zhou et al. [2021] consider the following problem. As above, there is a set of items $[L]$, a known feature map $\phi : [L] \to \mathbb{R}^d$, and an unknown parameter vector $\theta \in \mathbb{R}^d$. Successive plays of $e \in [L]$ give i.i.d. rewards with mean $\bar{w}(e) = \langle \phi(e), \theta \rangle$ and variance upper bounded by σ_e^2. Thus, at each round $t \in [n]$, the learner chooses $\mathbf{a}^t \in [L]$ and receives a random reward $r_t = \langle \phi(\mathbf{a}^t), \theta \rangle + \eta_t$, where $\mathbb{E}[\eta_t | \mathbf{a}^t] = 0$ and $\mathbb{E}[\eta_t^2 | \mathbf{a}^t] \leq \sigma_{\mathbf{a}^t}^2$. For this setting, the authors proposed the `WOFUL` algorithm, which is based on the (unweighted) `OFUL` algorithm [Abbasi-Yadkori et al., 2011]. At each round $t \in [n]$, `WOFUL` chooses the item

$$\mathbf{a}^t = \operatorname*{arg\,max}_{e \in [L]} \left(\langle \phi(e), \hat{\theta}_t \rangle + \alpha \|\phi(e)\|_{\mathbf{\Lambda}_t^{-1}} \right), \tag{5}$$

where $\alpha > 0$ is an exploration parameter, $\|x\|_B = \sqrt{x^\mathsf{T} B x}$ is the norm induced by a positive definite matrix B, and $\hat{\theta}_t$ is the regularized and variance-weighted least-squares estimate given by

$$\hat{\theta}_t = \mathbf{\Lambda}_t^{-1} \sum_{s=1}^{t-1} \phi(\mathbf{a}^s) r_s / \sigma_{\mathbf{a}^s}^2, \quad \text{where} \quad \mathbf{\Lambda}_t = I + \sum_{s=1}^{t-1} \phi(\mathbf{a}^s) \phi(\mathbf{a}^s)^\mathsf{T} / \sigma_{\mathbf{a}^s}^2. \tag{6}$$

To gain some intuition, we assume momentarily that $d = L$ and $\phi(e)$ is the e-th standard basis vector, i.e., the vector with 1 in the e-th coordinate and 0 elsewhere. In this case, one can easily calculate

$$\langle \phi(e), \hat{\theta}_t \rangle = \frac{\sum_{s \in [t-1] : \mathbf{a}^s = e} r_s / \sigma_e^2}{1 + \mathbf{T}_{t-1}(e) / \sigma_e^2}, \quad \|\phi(e)\|_{\mathbf{\Lambda}_t^{-1}} = \sqrt{\frac{1}{1 + \mathbf{T}_{t-1}(e) / \sigma_e^2}}.$$

Therefore, for large $\mathbf{T}_{t-1}(e)$ (large enough that $1 + \mathbf{T}_{t-1}(e)/\sigma_e^2 \approx \mathbf{T}_{t-1}(e)/\sigma_e^2$), we have

$$\langle \phi(e), \hat{\theta}_t \rangle + \alpha \|\phi(e)\|_{\mathbf{\Lambda}_t^{-1}} \approx \frac{\sum_{s \in [t-1] : \mathbf{a}^s = e} r_s}{\mathbf{T}_{t-1}(e)} + \alpha \sqrt{\frac{\sigma_e^2}{\mathbf{T}_{t-1}(e)}}.$$

Similar to `UCB-V` (3), the right side of this equation is the empirical mean plus an exploration bonus that grows with the variance upper bound σ_e^2 and decays in $\mathbf{T}_{t-1}(e)$. Hence, the term inside the $\arg\max$ in (5) can be interpreted as a Bernstein-style UCB. The analysis of this UCB relies on a novel concentration inequality for vector-valued martingales [Zhou et al., 2021, Theorem 2], which is a Bernstein analogue of the Hoeffding-style bound due to Abbasi-Yadkori et al. [2011].

Limitations. The fact that `WOFUL` (roughly) generalizes `UCB-V` is promising. However, computing (6) requires knowledge of the variance upper bounds σ_e^2, and nontrivial bounds on the variance are rarely available in practice. We sidestep this issue with three simple observations: (i) cascading bandits only involve Bernoulli rewards, (ii) for Bernoulli rewards, variances are upper bounded by means, and (iii) these means can be learned efficiently since they are linearly-parameterized. This suggests the following algorithm: first, compute Hoeffding-style UCBs; second, treat these UCBs as upper bounds for the true means, and thus the true variances, in `WOFUL`.

Proposed algorithm. Algorithm 2 formalizes this approach. It contains three steps. First, step 1 defines Hoeffding-style UCBs $\mathbf{U}_{t,H}$ as in Abbasi-Yadkori et al. [2011]. Next, step 2 uses $\mathbf{U}_{t,H}$ as an upper bound for the variance and computes the Bernstein-style UCBs $\mathbf{U}_{t,B}$ as in `WOFUL`. Finally, step 3 chooses $\mathbf{A}_t$ as the K items with the highest $\mathbf{U}_{t,B}$, analogous to Algorithm 1.

Two technical clarifications are in order. First, observe that in step 1, we clip the variance bound $\mathbf{U}_{t,H}$ below by $1/K$. We do so to ensure that $\mathbf{\Lambda}_{t,B}$ (which inverts $\mathbf{U}_{t,H}$) remains bounded. Additionally, we note the choice $1/K$ is precisely motivated by Section 1.2, which shows this is a critical threshold for the small click-through rate. Second, $\mathbf{\Lambda}_{t,B}$ uses the regularizer KI. This is to ensure that the regularizer is large enough compared to the summands $\phi(\mathbf{a}_k^s)\phi(\mathbf{a}_k^s)^\mathsf{T} / \mathbf{U}_{s,H}(\mathbf{a}_k^s)$, which scale as K in the worst case where the variance upper bound $\mathbf{U}_{s,H}(\mathbf{a}_k^s)$ is clipped to $1/K$.

Remark 5. *Zong et al. [2016] proposed a `WOFUL`-style cascading algorithm called `CascadeLinUCB` (see Appendix C), but they set the variance upper bound to fixed $\sigma^2 > 0$. In fact, they remark "ideally, σ^2 should be the variance of the observation noises," which `CascadeWOFUL` essentially* learns.

Remark 6. *Appendix B contains an improved version of `CascadeWOFUL`. It is more efficient (for example, the inverses are iteratively updated via Sherman-Morrison), satisfies the same theoretical guarantee as Algorithm 2, and includes some tweaks that improve performance in practice. The tradeoff is that Algorithm 2 is simpler to explain, which is why we prefer it for the main text.*

Algorithm 2: `CascadeWOFUL` for linear cascading bandits

Input: exploration parameters $\{\alpha_{t,H}, \alpha_{t,B}\}_{t=1}^n$
for $t = 1, \dots, n$ **do**

Step 1: Define the (clipped) Hoeffding-style UCBs

$$\mathbf{U}_{t,H}(\cdot) = \max\left\{\langle \phi(\cdot), \hat{\theta}_{t,H}\rangle + \alpha_{t,H}\|\phi(\cdot)\|_{\mathbf{\Lambda}_{t,H}^{-1}}, 1/K\right\}, \quad \text{where}$$

$$\hat{\theta}_{t,H} = \mathbf{\Lambda}_{t,H}^{-1}\sum_{s=1}^{t-1}\sum_{k=1}^{\min\{\mathbf{C}_s,K\}}\phi(\mathbf{a}_k^s)\mathbf{w}_s(\mathbf{a}_k^s),\ \mathbf{\Lambda}_{t,H} = I + \sum_{s=1}^{t-1}\sum_{k=1}^{\min\{\mathbf{C}_s,K\}}\phi(\mathbf{a}_k^s)\phi(\mathbf{a}_k^s)^\mathsf{T}$$

Step 2: Define the Bernstein-style UCBs

$$\mathbf{U}_{t,B}(\cdot) = \langle \phi(\cdot), \hat{\theta}_{t,B}\rangle + \alpha_{t,B}\|\phi(\cdot)\|_{\mathbf{\Lambda}_{t,B}^{-1}}, \quad \text{where}$$

$$\hat{\theta}_{t,B} = \mathbf{\Lambda}_{t,B}^{-1}\sum_{s=1}^{t-1}\sum_{k=1}^{\min\{\mathbf{C}_s,K\}}\frac{\phi(\mathbf{a}_k^s)\mathbf{w}_s(\mathbf{a}_k^s)}{\mathbf{U}_{s,H}(\mathbf{a}_k^s)},\ \mathbf{\Lambda}_{t,B} = KI + \sum_{s=1}^{t-1}\sum_{k=1}^{\min\{\mathbf{C}_s,K\}}\frac{\phi(\mathbf{a}_k^s)\phi(\mathbf{a}_k^s)^\mathsf{T}}{\mathbf{U}_{s,H}(\mathbf{a}_k^s)}$$

Step 3: Let $\mathbf{a}_k^t = \arg\max_{e\in[L]\setminus\{\mathbf{a}_i^t\}_{i=1}^{k-1}} \mathbf{U}_{t,B}(e)$ be the item with the k-th highest UCB, play $\mathbf{A}_t = (\mathbf{a}_1^t, \dots, \mathbf{a}_K^t)$, observe $\mathbf{C}_t = \inf\{k \in [K] : \mathbf{w}_t(\mathbf{a}_k^t) = 1\}$ (as in Algorithm 1)

Regret bound. We can now state our main result for the linear case. Here $\mathscr{B}_d(M) = \{x \in \mathbb{R}^d : \|x\|_2 \le M\}$ denotes the Euclidean ball of radius $M > 0$ in $\mathbb{R}^d$.[4]

Theorem 4. *Suppose $\bar{w} \in [0,1]^L$, $\phi : [L] \to \mathscr{B}_d(1)$, and $\theta \in \mathscr{B}_d(1)$ satisfy $\bar{w}(e) = \langle \phi(e), \theta\rangle$ for all $e \in [L]$. Let π be the policy of Algorithm 2 with inputs $\alpha_{t,H} = \sqrt{d\log(1+tK/d) + 2\log(n)} + 1$ and $\alpha_{t,B} = 8\sqrt{d\log(1+tK/d)\log(n^3K)} + 4\sqrt{K}\log(n^3K) + \sqrt{K}$. Then*

$$R_{\pi,\bar{w}}(n) = \tilde{O}\left(\min\left\{\sqrt{nd(d+K)} + n^{\frac{1}{6}}d^{\frac{7}{6}}(d+K)^{\frac{1}{2}}K^{\frac{1}{3}}, \sqrt{n}\max\{d,K\}\min\{d,K\}^{1/3}\right\}\right).$$

Note this bound is completely independent of the number of items L. For large n (typically the case of interest), it becomes $\tilde{O}(\sqrt{nd(d+K)})$, which improves the best known bound of $\tilde{O}(\sqrt{nd^2K})$ [Li and Zhang, 2018]. The theorem also establishes $\tilde{O}(\sqrt{n}\max\{d,K\}\min\{d,K\}^{1/3})$ regret *uniformly* in n (i.e., without additive $o(\sqrt{n})$ terms which may dominate for small n), which improves the best known uniform-n bound of $\tilde{O}(\sqrt{n}dK)$ [Zong et al., 2016]. See Table 1 for more details.

5 Overview of the analysis

We next discuss the key ideas behind our proofs. The details are deferred to Appendices E-J.

Theorem 2 proof. For simplicity, we assume the arms are ordered by their means, i.e., $\bar{w}(1) \ge \cdots \ge \bar{w}(L)$. Under this assumption, the optimal action is $[K]$. We call those items *optimal* and $[L] \setminus [K]$ *suboptimal*. We let $\mathcal{E}_t$ be the "bad event" that the empirical and true means differ substantially at time t, $\bar{\mathcal{E}}_t$ its complement, $G_{e,e^*,t}$ the event that suboptimal $e > K$ was chosen in favor of optimal $e^* \le K$ and subsequently examined by the user, and $\Delta_{e,e^*} = \bar{w}(e^*) - \bar{w}(e)$ the *reward gap*. Then as in Appendix A.1 of Kveton et al. [2015a], we "linearize" regret as follows:

$$R_{\pi,\bar{w}}(n) \le \mathbb{E}\left[\sum_{t\le n}\sum_{e>K}\sum_{e^*\le K}\Delta_{e,e^*}\mathbb{1}(\bar{\mathcal{E}}_t, G_{e,e^*,t})\right] + \sum_{t\le n}\mathbb{P}(\mathcal{E}_t). \tag{7}$$

The second term is small due to concentration. For the first term, define $\Delta = \bar{w}(K)\sqrt{L/n}$ and assume $n \gg L$ (so $\Delta \ll \bar{w}(K)$).[5] For each $e > K$, let $\mathcal{K}_\Delta(e) = \{e^* \le K : \Delta_{e,e^*} \le \Delta\}$ be the

[4] Our results hold with minor modification if $\phi : [L] \to \mathscr{B}_d(M_1)$ and $\theta \in \mathscr{B}_d(M_2)$ for general $M_1, M_2 > 0$. However, as in prior work, the modified algorithm needs to know an upper bound on M_2.

[5] The full proof addresses the cases $n \not\gg L$ and $\bar{w}(K) = 0$ (here we implicitly assume the latter to invert Δ).

optimal items with small gap relative to e and $\bar{\mathcal{K}}_\Delta(e) = [K] \setminus \mathcal{K}_\Delta(e)$ the other optimal items. Then

$$\sum_{e^* \le K} \Delta_{e,e^*} \mathbb{1}(\bar{\mathcal{E}}_t, G_{e,e^*,t}) \le \Delta \sum_{e^* \in \mathcal{K}_\Delta(e)} \mathbb{1}(G_{e,e^*,t}) + \sum_{e^* \in \bar{\mathcal{K}}_\Delta(e)} \Delta_{e,e^*} \mathbb{1}(\bar{\mathcal{E}}_t, G_{e,e^*,t}).$$

Therefore, we can upper bound the first term in (7) by

$$\Delta \sum_{t \le n} \mathbb{E}\left[\sum_{e > K} \sum_{e^* \in \mathcal{K}_\Delta(e)} \mathbb{1}(G_{e,e^*,t})\right] + \mathbb{E}\left[\sum_{t \le n} \sum_{e > K} \sum_{e^* \in \bar{\mathcal{K}}_\Delta(e)} \Delta_{e,e^*} \mathbb{1}(\bar{\mathcal{E}}_t, G_{e,e^*,t})\right]. \tag{8}$$

For the first term in (8), the inner double summation is the number of $e > K$ that were chosen in favor of $e^* \in \mathcal{K}_\Delta(e)$, and subsequently examined by the user, at round t. Denote this number by $\mathbf{O}_t$. Recall $\bar{w}(e^*) \ge \bar{w}(K)$ (by the assumed ordering) and $\Delta \ll \bar{w}(K)$, so for $e^* \in \mathcal{K}_\Delta(e)$, we have

$$\bar{w}(e) = \bar{w}(e^*) - \Delta_{e,e^*} \ge \bar{w}(e^*) - \Delta \ge \bar{w}(K) - \Delta \gtrapprox \bar{w}(K).$$

Hence, $\mathbf{O}_t$ is bounded by the number of items e with $\bar{w}(e) \approx \bar{w}(K)$ that the user examined. Since the user stops examining at the first attractive item, this number is dominated by a Geometric$(\bar{w}(K))$ random variable. Therefore, the first term in (8) is $\tilde{O}(\Delta n / \bar{w}(K)) = \tilde{O}(\sqrt{nL})$.

The second term in (8) accounts for choosing e instead of e^* when $\Delta_{e,e^*} \ge \Delta$ and the empirical means are concentrated. Building upon the intuition of Section 1.2, we exploit the variance-awareness of `KL-UCB` and `UCB-V` (and use $\bar{w}(e) \le \bar{w}(K)$ by the assumed ordering) to bound this term by $\tilde{O}(L\bar{w}(e)(1 - \bar{w}(e))/\Delta) \le \tilde{O}(L\bar{w}(K)/\Delta)$. Thus, by choice of Δ, this term is $\tilde{O}(\sqrt{nL})$ as well.

Theorem 4 proof. We decompose regret similar to (7) and (8), though with different choices of $\mathcal{E}_t$ and Δ.[6] To bound $\mathbb{P}(\mathcal{E}_t)$, we use the aforementioned result of Abbasi-Yadkori et al. [2011] to show that the Hoeffding UCBs $\mathbf{U}_{t,H}$ upper bound the variances with high probability, then prove a guarantee for the least-squares estimate $\hat{\theta}_{t,B}$ using the Bernstein-style bound from Zhou et al. [2021]. We bound the first term in (8) using the exact same logic as the tabular case. The second term has a more complicated analysis, but (as in the tabular case) it amounts to bounding the number of times that items Δ-far from optimal are chosen when $\hat{\theta}_{t,B}$ is well-concentrated. For this, we adapt techniques from the standard linear bandit setting (the case $K = 1$) to general $K \in [L]$.

Theorem 1 proof. We let $\mathcal{H}'_t = \cup_{s=1}^{t-1} \{\mathbf{A}_s, \mathbf{w}_s(\mathbf{a}_1^s), \ldots, \mathbf{w}_s(\mathbf{a}_K^s)\}$ be the *entire* history before time t, which includes unobserved rewards $\mathbf{w}_s(\mathbf{a}_k^s), k > \mathbf{C}_s$. We also let Π' be the policies that map $\mathcal{H}'_t$ to $\mathbf{A}_t$. Note $\mathcal{H}_t \subset \mathcal{H}'_t$, so $\Pi \subset \Pi'$. For any $\pi \in \Pi'$ and $\bar{w} \in [0,1]^L$, we define

$$R'_{\pi,\bar{w}}(n) = \mathbb{E}_{\pi,\bar{w}}\left[\sum_{t=1}^{n} \sum_{k=1}^{K} (\bar{w}(a_k^*) - \bar{w}(\mathbf{a}_k^t))\right], \tag{9}$$

where $a_k^* = \arg\max_{e \in [L] \setminus \{a_i^*\}_{i=1}^{k-1}} \bar{w}(e)$. Then a lower bound linearization analogous to (7) shows that for any $p \in [0,1]$ and $\bar{w} \in [0,p]^L$, $R_{\pi,\bar{w}}(n) \ge (1-p)^{K-1} R'_{\pi,\bar{w}}(n)$, which implies

$$\inf_{\pi \in \Pi} \sup_{\bar{w} \in [0,1]^L} R_{\pi,\bar{w}}(n) \ge \inf_{\pi \in \Pi'} \sup_{\bar{w} \in [0,p]^L} R_{\pi,\bar{w}}(n) \ge (1-p)^{K-1} \inf_{\pi \in \Pi'} \sup_{\bar{w} \in [0,p]^L} R'_{\pi,\bar{w}}(n). \tag{10}$$

If we choose $p = 1$, the $\inf\sup$ at right is the minimax regret for the document-based model that was analyzed by Lattimore et al. [2018] (see Remark 1), but this makes (10) vacuous. On the other hand, by choosing $p = O(1/K)$ (again, the small click-through rate of Section 1.2), so that the term $(1-p)^{K-1}$ in (10) is $\Omega(1)$, we can modify their analysis to prove Theorem 1.

Theorem 3 proof. We linearize the regret similar to the proof of Theorem 1, then define a problem instance reminiscent of Section 1.2 and (roughly) follow the intuition for `UCB1` therein.

6 Experiments

Before closing, we conduct experiments on both synthetic and real data. Some details regarding experimental setup are deferred to Appendix C. Code is available in the supplementary material.

[6] In fact, the proofs of Theorems 2 and 4 both rely on a more general gap-free regret decomposition for cascading bandits (Lemma 1 in Appendix F), which to our knowledge is novel and may be of independent interest.

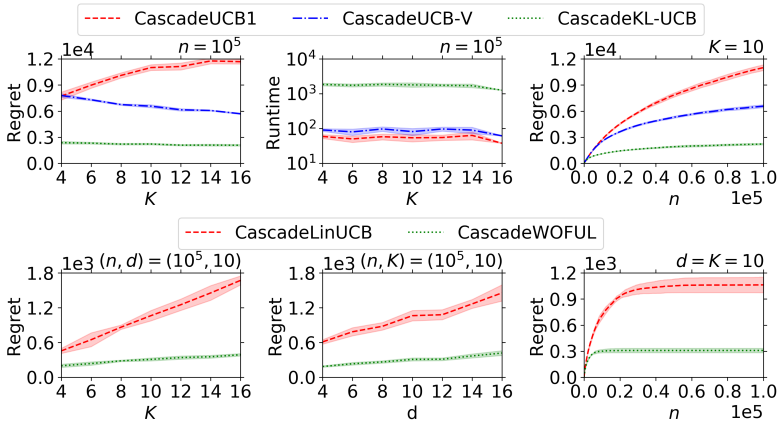

Figure 1: Results for synthetic data (tabular on top, linear on bottom, $L = 100$ in both)

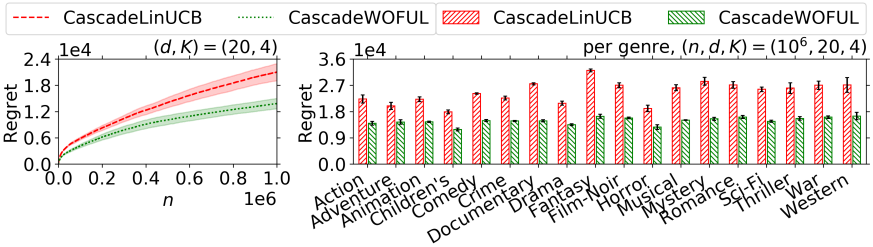

Figure 2: Results for MovieLens data [Harper and Konstan, 2015]

Synthetic data. We let $L = 100$ and $K \in \{2i\}_{i=2}^{8}$. For each K, we sample $\bar{w}(e)$ uniformly in $[\frac{2}{3K}, \frac{1}{K}]$ for $e \leq K$ and in $[0, \frac{1}{3K}]$ for $e > K$. Note this yields a positive gap with the small click-through rate of Section 1.2. The top left plot in Figure 1 shows the regret at $n = 10^5$ for the tabular albums of Section 3 (the shaded regions are the standard deviations across five trials). As predicted by Theorems 2 and 3, the `CascadeUCB1` curve grows with K, while the variance-aware curves do not. The top middle plot shows the time spent computing UCBs for the same experiment, which confirms the behavior mentioned in Remark 3. For the linear case, we vary $d \in \{2i\}_{i=2}^{8}$, generate the same $\bar{w}$, then compute unit-norm vectors θ and $\phi(e)$ satisfying $\bar{w}(e) = \langle \phi(e), \theta \rangle$ (see Appendix C). We compare `CascadeWOFUL` to `CascadeLinUCB` [Zong et al., 2016], which Li and Zhang [2018] showed has the best existing regret guarantee. As suggested by Theorem 4, the left and middle plots on the bottom of Figure 1 show that our algorithm's regret has superior dependence on K and d. The rightmost plots show that regret is sublinear in n for the median values $K = d = 10$.

Real data. We replicate the first experiment from Zong et al. [2016] on the MovieLens-1M dataset (`grouplens.org/datasets/movielens/1m/`), which contains user ratings for $L \approx 4000$ movies. In brief, the setup is as follows. First, we use their default choices $d = 20$ and $K = 4$. Next, we divide the ratings into train and test sets based on the user who provided the rating. From the training data and a rank-d SVD approximation, we learn a feature mapping ϕ from movies to the probability that a uniformly random training user rated the movie more than three stars. Finally, we run the algorithms as above, except at round $t \in [n]$, we sample a uniformly random user $\mathbf{J}_t$ from the test set and define $\mathbf{w}_t(\mathbf{a}_k^t) = W(\mathbf{J}_t, \mathbf{a}_k^t)$, where $W(j, a) = \mathbb{1}(\text{user } j \text{ rated movie } a \text{ more than 3 stars})$. In other words, instead of the independent Bernoulli clicks of Section 2, we observe the actual feedback of user $\mathbf{J}_t$. We point the reader to Section 4 of Zong et al. [2016] and Appendix C for further details. The left plot of Figure 2 shows that `CascadeWOFUL` outperforms `CascadeLinUCB` across n, eventually incurring less than 66% of the regret. In addition to this setup from Zong et al. [2016], we reran the experiment while restricting the set of items to movies of a particular genre, for each of 18 genres in the dataset. This is intended to model platforms like Netflix that recommend movies in various categories. The right plot shows that `CascadeWOFUL` is superior for all genres; for some genres (e.g., fantasy) its regret is about half of `CascadeLinUCB`'s. Moreover, our experiments indicate that `CascadeWOFUL`

improves `CascadeLinUCB` more dramatically for genres with smaller click-through rates (see Figure 3 and surrounding discussion in Appendix C), which reinforces a key message of this paper.

7 Conclusion

In this work, we proved matching upper and lower bounds for the problem-independent regret of tabular cascading bandits and an upper bound for the linear case, all of which improve the best known. Our results suggest some interesting future directions, such as proving minimax lower bounds for the linear case and revisiting Thompson sampling for cascading bandits [Zhong et al., 2021] in light of our variance-aware insight; see Appendix D for details. Finally, we note the paper is theoretical and has no immediate societal impact. Nevertheless, we urge caution for the negative impacts that could arise in practice. For example, our MovieLens experiments involved training on a subset of users, which could cause poor recommendations for demographics underrepresented in the training set.

Acknowledgments and Disclosure of Funding

This work was partially supported by ONR Grant N00014-19-1-2566, NSF TRIPODS Grant 1934932, NSF Grants CCF 22-07547, CCF 19-34986, CNS 21-06801, 2019844, 2112471, 2107037, the Machine Learning Lab (MLL) at UT Austin, and the Wireless Networking and Communications Group (WNCG) Industrial Affiliates Program. We thank Advait Parulekar for helpful discussions.

References

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. *Advances in neural information processing systems*, 24, 2011.

Jean-Yves Audibert, Rémi Munos, and Csaba Szepesvári. Exploration–exploitation tradeoff using variance estimates in multi-armed bandits. *Theoretical Computer Science*, 410(19):1876–1902, 2009.

Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2):235–256, 2002.

Sébastien Bubeck and Nicolo Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Machine Learning*, 5(1):1–122, 2012.

Olivier Cappé, Aurélien Garivier, Odalric-Ambrym Maillard, Rémi Munos, and Gilles Stoltz. Kullback-leibler upper confidence bounds for optimal sequential allocation. *The Annals of Statistics*, pages 1516–1541, 2013.

Wang Chi Cheung, Vincent Tan, and Zixin Zhong. A thompson sampling algorithm for cascading bandits. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 438–447. PMLR, 2019.

Aleksandr Chuklin, Ilya Markov, and Maarten de Rijke. *Click Models for Web Search*. Morgan & Claypool Publishers, 2015.

Richard Combes, Stefan Magureanu, Alexandre Proutiere, and Cyrille Laroche. Learning to rank: Regret lower bounds and efficient algorithms. In *Proceedings of the 2015 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, pages 231–244, 2015.

Nick Craswell, Onno Zoeter, Michael Taylor, and Bill Ramsey. An experimental comparison of click position-bias models. In *Proceedings of the 2008 international conference on web search and data mining*, pages 87–94, 2008.

Varsha Dani, Thomas P Hayes, and Sham M Kakade. Stochastic linear optimization under bandit feedback. In *Conference on Learning Theory*, pages 2137–2143, 2008.

David A Freedman. On tail probabilities for martingales. *the Annals of Probability*, pages 100–118, 1975.

Aurélien Garivier and Olivier Cappé. The kl-ucb algorithm for bounded stochastic bandits and beyond. In *Proceedings of the 24th annual conference on learning theory*, pages 359–376. JMLR Workshop and Conference Proceedings, 2011.

F Maxwell Harper and Joseph A Konstan. The movielens datasets: History and context. *Acm transactions on interactive intelligent systems (tiis)*, 5(4):1–19, 2015.

Gaurush Hiranandani, Harvineet Singh, Prakhar Gupta, Iftikhar Ahamath Burhanuddin, Zheng Wen, and Branislav Kveton. Cascading linear submodular bandits: Accounting for position bias and diversity in online learning to rank. In *Uncertainty in Artificial Intelligence*, pages 722–732. PMLR, 2020.

Branislav Kveton, Csaba Szepesvari, Zheng Wen, and Azin Ashkan. Cascading bandits: Learning to rank in the cascade model. In *International Conference on Machine Learning*, pages 767–776. PMLR, 2015a.

Branislav Kveton, Zheng Wen, Azin Ashkan, and Csaba Szepesvari. Combinatorial cascading bandits. *Advances in Neural Information Processing Systems*, 28, 2015b.

Branislav Kveton, Ofer Meshi, Zhen Qin, and Masrour Zoghi. On the value of prior in online learning to rank. In *The 25th International Conference on Artificial Intelligence and Statistics*, 2022.

Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

Tor Lattimore, Branislav Kveton, Shuai Li, and Csaba Szepesvari. Toprank: A practical algorithm for online stochastic ranking. *Advances in Neural Information Processing Systems*, 31, 2018.

Chang Li and Maarten De Rijke. Cascading non-stationary bandits: online learning to rank in the non-stationary cascade model. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, pages 2859–2865, 2019.

Shuai Li and Shengyu Zhang. Online clustering of contextual cascading bandits. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018.

Shuai Li, Baoxiang Wang, Shengyu Zhang, and Wei Chen. Contextual combinatorial cascading bandits. In *International conference on machine learning*, pages 1245–1253. PMLR, 2016.

Shuai Li, Tor Lattimore, and Csaba Szepesvári. Online learning to rank with features. In *International Conference on Machine Learning*, pages 3856–3865. PMLR, 2019.

Xutong Liu, Jinhang Zuo, Siwei Wang, Carlee Joe-Wong, John Lui, and Wei Chen. Batch-size independent regret bounds for combinatorial semi-bandits with probabilistically triggered arms or independent arms. *arXiv preprint arXiv:2208.14837*, 2022.

Odalric-Ambrym Maillard, Rémi Munos, and Gilles Stoltz. A finite-time analysis of multi-armed bandits problems with kullback-leibler divergences. In *Proceedings of the 24th annual Conference On Learning Theory*, pages 497–514. JMLR Workshop and Conference Proceedings, 2011.

Matthew Richardson, Ewa Dominowska, and Robert Ragno. Predicting clicks: estimating the click-through rate for new ads. In *Proceedings of the 16th international conference on World Wide Web*, pages 521–530, 2007.

Lingda Wang, Huozhi Zhou, Bingcong Li, Lav R Varshney, and Zhizhen Zhao. Near-optimal algorithms for piecewise-stationary cascading bandits. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3365–3369. IEEE, 2021.

Qinshi Wang and Wei Chen. Improving regret bounds for combinatorial semi-bandits with probabilistically triggered arms and its applications. *Advances in Neural Information Processing Systems*, 30, 2017.

Zixin Zhong, Wang Chi Chueng, and Vincent YF Tan. Thompson sampling algorithms for cascading bandits. *Journal of Machine Learning Research*, 22:1–66, 2021.

Dongruo Zhou, Quanquan Gu, and Csaba Szepesvari. Nearly minimax optimal reinforcement learning for linear mixture markov decision processes. In *Conference on Learning Theory*, pages 4532–4576. PMLR, 2021.

Masrour Zoghi, Tomas Tunys, Mohammad Ghavamzadeh, Branislav Kveton, Csaba Szepesvari, and Zheng Wen. Online learning to rank in stochastic click models. In *International Conference on Machine Learning*, pages 4199–4208. PMLR, 2017.

Shi Zong, Hao Ni, Kenny Sung, Nan Rosemary Ke, Zheng Wen, and Branislav Kveton. Cascading bandits for large-scale recommendation problems. In *Proceedings of the Thirty-Second Conference on Uncertainty in Artificial Intelligence*, pages 835–844, 2016.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes] The claims are accurate. Furthermore, we provide references to the formal results alongside each claim stated in the introduction.
   (b) Did you describe the limitations of your work? [Yes] The main limitation is that our lower bounds (Theorems 1 and 3) require some additional assumptions. These are stated clearly in the theorem statements and discussed in Remarks 1 and 4.
   (c) Did you discuss any potential negative societal impacts of your work? [Yes] Section 7 addresses societal impact.
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes] As mentioned above, the assumptions are clearly stated in the theorem statements.
   (b) Did you include complete proofs of all theoretical results? [Yes] Section 1.2 discusses the high-level intuition behind our proofs, Section 5 contains a more detailed proof sketch, and the appendices include complete proofs.
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] Complete code to recreate all plots is available in the supplementary material.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] The experimental setup is discussed in detail in Appendix C, and other training details can be found in the code.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes] All plots include error bars.
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] The experiments were run over several hours on a laptop, so no significant resources were used. Nevertheless, Appendix C mentions the approximate runtime needed to recreate the figures.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes] We only used the MovieLens dataset, which we cited and also provided a link to.
   (b) Did you mention the license of the assets? [Yes] The link to the MovieLens dataset contains a README with license information.
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]

5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

A Details on related work

In addition to the papers listed in Table 1, several others have considered cascading bandits and variants. In this appendix, we separately discuss the relevant work in the tabular and linear cases. We also point the reader to Chuklin et al. [2015] for a survey of click models, and to Chapter 32 of Lattimore and Szepesvári [2020] for an introduction to bandit-style ranking problems.

Tabular case. As mentioned in Section 1, Kveton et al. [2015a] and Combes et al. [2015] introduced cascading bandits concurrently (though the latter work did not use that name and their model is more general). Both proved problem-dependent regret bounds of the form $(L-K)\log(n)/\Delta$, where Δ is a certain notion of mean reward gap. Kveton et al. [2015b] provided gap-free bounds for a model whose reward structure is more general than that of cascading bandits. However, these results involve other problem-dependent quantities, so they can be much larger than ours for worst case mean reward vectors $\bar{w}$. Zoghi et al. [2017] generalized cascading bandits – essentially, by generalizing the underlying click model – and derived gap-dependent bounds. Lattimore et al. [2018] further generalized the underlying click model and established gap-free bounds (those shown in Table 1). Li and De Rijke [2019] and Wang et al. [2021] proved gap-dependent bounds for a non-stationary variant of the cascading bandit model, where $\bar{w}$ changes arbitrarily at M rounds. Theorem 3 of the latter also reports the gap-free lower bound $\Omega(\sqrt{nL(M+1)})$, which in our case $M=0$ appears tighter than the bound $\sqrt{nL/K}$ that the main text stated is the best known. However, their proof shows that the $\Omega(\cdot)$ notation hides a term which is exponentially small in K. Finally, the concurrent work by Liu et al. [2022] also examines the role of variance in cascading bandits (as a special case of a more general model), though their focus is on gap-dependent regret bounds.

Linear case. Li et al. [2016] considered the linear case of the model from Kveton et al. [2015b]. Analogous to Kveton et al. [2015b]'s results in tabular case, their regret bounds can be much worse than ours in the worst case. Li and Zhang [2018] proposed a model where a random user arrives at each round and the parameter vector θ depends on that user. When specialized to our case (i.e., the case of a single θ), their algorithm reduces to `CascadeLinUCB` [Zong et al., 2016], and they sharpen the bound from Zong et al. [2016] by a factor of $\sqrt{K}$ (at the cost of some $o(\sqrt{n})$ additive terms, i.e., their bound does not hold uniformly in n); see Table 1 for details. Hiranandani et al. [2020] studied a model that generalizes cascading bandits to account for position bias, but their regret bounds include problem-dependent quantities. Li et al. [2019] established $O(\sqrt{nd\log LK})$ regret for a generalized click model reminiscent of the ones studied by Zoghi et al. [2017] and Lattimore et al. [2018] in the tabular case. Note this bound can be arbitrarily worse than those in Table 1 due to the $\log L$ factor. Moreover, Li et al. [2019]'s experiments show the algorithm attaining their upper bound performs worse than `CascadeLinUCB` when their general model is specialized to cascading bandits (see their Figure 2), which is another reason why we used the latter for experiments. Finally, we note the work by Zhong et al. [2021] mentioned above is an extended version of the paper by Cheung et al. [2019].

B Improved version of `CascadeWOFUL`

Algorithm 3 provides the version of `CascadeWOFUL` that was mentioned in Remark 6. There are four changes from Algorithm 2. First, Step 1 defines the Bernstein UCB $\mathbf{U}_{t,B}(e)$ from Algorithm 2, along with an (unclipped) Hoeffding UCB $\tilde{\mathbf{U}}_{t,H}(e)$. We then define the UCB $\mathbf{U}_t(e)$ as the minimum of these two UCBs and 1. The basic intuition is that $\mathbf{U}_{t,B}(e)$, $\tilde{\mathbf{U}}_{t,H}(e)$, and 1 all upper bound $\bar{w}(e)$ (the former two with high probability; the latter almost surely), so we should use the tightest of the three for the final UCB. Second, Step 2 chooses the action $\mathbf{A}_t$ similar to Algorithm 2, though it is greedy with respect to $\mathbf{U}_t(e)$ instead of $\mathbf{U}_{t,B}(e)$. Third, Steps 3 and 4 update the inverses $\mathbf{\Lambda}_{t,H}^{-1}$ and $\mathbf{\Lambda}_{t,B}^{-1}$ iteratively, which reduces their computational complexity from d^3 to d^2. Furthermore, note that when updating the latter, we clip $\tilde{\mathbf{U}}_{t,H}(\mathbf{a}_k^t)$ below by $1/K$ (as in Algorithm 2) and above by 1 (following the same intuition described above) when defining the normalized features $\phi_t(\mathbf{a}_k^t)$. Fourth, Steps 5 and 6 iteratively update the estimates $\hat{\theta}_{t,H}$ and $\hat{\theta}_{t,B}$, again to save computation. In Appendix H.3, we sketch a proof that this version also satisfies the guarantee of Theorem 4.

Algorithm 3: Improved version of `CascadeWOFUL` (recommended over Algorithm 2 in practice)

Input: exploration parameters $\{\alpha_{t,H}, \alpha_{t,B}\}_{t=1}^n$
Initialize $\hat{\theta}_{1,H} = 0$, $\mathbf{\Lambda}_{1,H} = I$, $\mathbf{\Lambda}_{1,H}^{-1} = I$, $\hat{\theta}_{1,B} = 0$, $\mathbf{\Lambda}_{1,B} = KI$, $\mathbf{\Lambda}_{1,B}^{-1} = I/K$
for $t = 1, \dots, n$ **do**

Step 1: compute upper confidence bounds (UCBs)
for $e = 1, \dots, L$ **do**
$\tilde{\mathbf{U}}_{t,H}(e) = \langle \phi(e), \hat{\theta}_{t,H} \rangle + \alpha_{t,H} \|\phi(e)\|_{\mathbf{\Lambda}_{t,H}^{-1}}$ (Hoeffding-style UCB)
$\mathbf{U}_{t,B}(e) = \langle \phi(e), \hat{\theta}_{t,B} \rangle + \alpha_{t,B} \|\phi(e)\|_{\mathbf{\Lambda}_{t,B}^{-1}}$ (Bernstein-style UCB)
$\mathbf{U}_t(e) = \min\{\tilde{\mathbf{U}}_{t,H}(e), \mathbf{U}_{t,B}(e), 1\}$ (tightest of the two UCBs and the trivial UCB)

Step 2: play greedily with respect to UCBs (similar to Algorithms 1 and 2)
for $k = 1, \dots, K$ **do**
$\mathbf{a}_k^t = \arg\max_{e \in [L] \setminus \{\mathbf{a}_i^t\}_{i=1}^{k-1}} \mathbf{U}_t(e)$ (k-th highest UCB; ties broken arbitrarily)
Choose $\mathbf{A}_t = (\mathbf{a}_1^t, \dots, \mathbf{a}_K^t)$, observe $\mathbf{C}_t = \inf\{k \in [K] : \mathbf{w}_t(\mathbf{a}_k^t) = 1\}$ (where $\inf \emptyset = \infty$)

Step 3: update $\mathbf{\Lambda}_{\cdot,H}$, $\mathbf{\Lambda}_{\cdot,H}^{-1}$, $\mathbf{\Lambda}_{\cdot,B}$, and $\mathbf{\Lambda}_{\cdot,B}^{-1}$ (the inverses via Sherman-Morrison)
Initialize $\mathbf{\Lambda}_{t+1,H} = \mathbf{\Lambda}_{t,H}$, $\mathbf{\Lambda}_{t+1,H}^{-1} = \mathbf{\Lambda}_{t,H}^{-1}$, $\mathbf{\Lambda}_{t+1,B} = \mathbf{\Lambda}_{t,B}$, and $\mathbf{\Lambda}_{t+1,B}^{-1} = \mathbf{\Lambda}_{t,B}^{-1}$
for $k = 1, \dots, \min\{\mathbf{C}_t, K\}$ **do**
$\mathbf{\Lambda}_{t+1,H} \leftarrow \mathbf{\Lambda}_{t+1,H} + \phi(\mathbf{a}_k^t)\phi(\mathbf{a}_k^t)^\top$
$\mathbf{\Lambda}_{t+1,H}^{-1} \leftarrow \mathbf{\Lambda}_{t+1,H}^{-1} - \mathbf{\Lambda}_{t+1,H}^{-1}\phi(\mathbf{a}_k^t)\phi(\mathbf{a}_k^t)^\top \mathbf{\Lambda}_{t+1,H}^{-1} / (1 + \phi(\mathbf{a}_k^t)^\top \mathbf{\Lambda}_{t+1,H}^{-1}\phi(\mathbf{a}_k^t))$
$\mathbf{U}_{t,H}(\mathbf{a}_k^t) = \min\{\max\{\tilde{\mathbf{U}}_{t,H}(\mathbf{a}_k^t), 1/K\}, 1\}$ (clipped Hoeffding UCB)
$\phi_t(\mathbf{a}_k^t) = \phi(\mathbf{a}_k^t)/\sqrt{\mathbf{U}_{t,H}(\mathbf{a}_k^t)}$ (normalized feature vector)
$\mathbf{\Lambda}_{t+1,B} \leftarrow \mathbf{\Lambda}_{t+1,B} + \phi_t(\mathbf{a}_k^t)\phi_t(\mathbf{a}_k^t)^\top$
$\mathbf{\Lambda}_{t+1,B}^{-1} \leftarrow \mathbf{\Lambda}_{t+1,B}^{-1} - \mathbf{\Lambda}_{t+1,B}^{-1}\phi_t(\mathbf{a}_k^t)\phi_t(\mathbf{a}_k^t)^\top \mathbf{\Lambda}_{t+1,B}^{-1} / (1 + \phi_t(\mathbf{a}_k^t)^\top \mathbf{\Lambda}_{t+1,B}^{-1}\phi_t(\mathbf{a}_k^t))$

Step 4: compute $\hat{\theta}_{t+1,H}$ and $\hat{\theta}_{t+1,B}$ (the two cases correspond to "click" and "no click")
if $\mathbf{C}_t < \infty$ **then**
$\hat{\theta}_{t+1,H} = \mathbf{\Lambda}_{t+1,H}^{-1}(\mathbf{\Lambda}_{t,H}\hat{\theta}_{t,H} + \phi(\mathbf{a}_{\mathbf{C}_t}^t))$, $\hat{\theta}_{t+1,B} = \mathbf{\Lambda}_{t+1,B}^{-1}(\mathbf{\Lambda}_{t,B}\hat{\theta}_{t,B} + \phi_t(\mathbf{a}_{\mathbf{C}_t}^t))$
else
$\hat{\theta}_{t+1,H} = \mathbf{\Lambda}_{t+1,H}^{-1}\mathbf{\Lambda}_{t,H}\hat{\theta}_{t,H}$, $\hat{\theta}_{t+1,B} = \mathbf{\Lambda}_{t+1,B}^{-1}\mathbf{\Lambda}_{t,B}\hat{\theta}_{t,B}$

C Details on experiments

Tabular algorithms. For `CascadeUCB-V` and `CascadeUCB1`, respectively, we compute the UCBs $\mathbf{U}_t(e)$ as shown in (3) and (4), respectively, then take a minimum with the resulting UCB and 1 (analogous to Algorithm 3). For `CascadeKL-UCB` (which lacks a closed form), the fact that $d(\hat{\mathbf{w}}_{\mathbf{T}_{t-1}(e)}(e), \cdot)$ is increasing on $[\hat{\mathbf{w}}_{\mathbf{T}_{t-1}(e)}(e), 1]$ implies that we can estimate the UCB (2) via binary search, as shown in Algorithm 4 (we chose $\texttt{tol} = 10^{-4}$ in experiments).

Linear algorithms. `CascadeLinUCB` [Zong et al., 2016] is shown in Algorithm 5. As mentioned in Remark 5, it is a `WOFUL`-style algorithm that uses fixed $\sigma^2 > 0$ for the variances. We set $\sigma^2 = 1$ (as in their experiments) and choose the exploration parameter $\alpha = \sqrt{d\log(1 + nK/d) + 2\log(nK)} + 1$ (which is the choice dictated by their Theorem 1 in the case $\sigma^2 = 1$ and $\|\theta\|_2 \le 1$). For the experiments, we use an efficient implementation of Algorithm 5 analogous to Algorithm 3.

For `CascadeWOFUL`, we use the Hoeffding parameter suggested by Theorem 4, namely,

$$\alpha_{t,H} = \beta_{t,H} + 1, \quad \text{where} \quad \beta_{t,H} = \sqrt{d\log(1 + tK/d) + 2\log(n)}.$$

Note the term $\beta_{t,H}$ accounts for concentration of the self-normalized estimate $\hat{\theta}_{t,H}$ of θ [Abbasi-Yadkori et al., 2011], while the term 1 accounts for the bias introduced by the regularizer I used to compute $\hat{\theta}_{t,H}$. On the other hand, we found the Bernstein parameter $\alpha_{t,B}$ stated in Theorem 4 is too conservative in practice, so we instead chose $\alpha_{t,B} = \beta_{t,H} + \sqrt{K}$, i.e., we use the same concentration

Algorithm 4: Estimating the `KL-UCB` (2) via binary search

$\ell = \hat{\mathbf{w}}_{\mathbf{T}_{t-1}(e)}(e), u = 1$
while $u - \ell > \texttt{tol}$ **do**
  if $d(\hat{\mathbf{w}}_{\mathbf{T}_{t-1}(e)}(e), (u+\ell)/2) > \log(f(t))/\mathbf{T}_{t-1}(e)$ **then** $u = (u+\ell)/2$
  else $\ell = (u+\ell)/2$
return $u \approx \max\{u' \in [0,1] : d(\hat{\mathbf{w}}_{\mathbf{T}_{t-1}(e)}(e), u') \leq \log(f(t))/\mathbf{T}_{t-1}(e)\}$

Algorithm 5: `CascadeLinUCB` for linear cascading bandits [Zong et al., 2016]

Input: exploration parameter α, variance parameter σ^2
for $t = 1, \ldots, n$ **do**
  Compute $\mathbf{U}_t(e) = \min\{\langle \phi(e), \hat{\theta}_t \rangle + \alpha \|\phi(e)\|_{\mathbf{\Lambda}_t^{-1}}, 1\}$ for each $e \in [L]$, where

$$\hat{\theta}_t = \mathbf{\Lambda}_t^{-1} \sum_{s=1}^{t-1} \sum_{k=1}^{\min\{\mathbf{C}_s, K\}} \phi(\mathbf{a}_k^s) \mathbf{w}_s(\mathbf{a}_k^s)/\sigma^2, \ \mathbf{\Lambda}_t = I + \sum_{s=1}^{t-1} \sum_{k=1}^{\min\{\mathbf{C}_s, K\}} \phi(\mathbf{a}_k^s)\phi(\mathbf{a}_k^s)^\mathsf{T}/\sigma^2$$

  Let $\mathbf{a}_k^t = \arg\max_{e \in [L] \setminus \{\mathbf{a}_i^t\}_{i=1}^{k-1}} \mathbf{U}_t(e)$ be the item with the k-th highest UCB, play $\mathbf{A}_t = (\mathbf{a}_1^t, \ldots, \mathbf{a}_K^t)$, observe $\mathbf{C}_t = \inf\{k \in [K] : \mathbf{w}_t(\mathbf{a}_k^t) = 1\}$

term but change the other term to account for the KI regularizer in $\mathbf{\Lambda}_{t,B}$. Therefore, compared to the theorem, the experimental choice of $\alpha_{t,B}$ uses a smaller concentration parameter. We believe this is reasonable because the choice dictated by the theorem is likely loose. This is because the theoretical choice is inherited from the Bernstein analysis of Zhou et al. [2021], which is based on the Hoeffding analysis of Dani et al. [2008], which Abbasi-Yadkori et al. [2011] improved in the Hoeffding case. Put differently, state-of-the-art Bernstein bounds seem to be looser, which leads to a larger Bernstein parameter in Theorem 4 that hamstrings the practical performance `CascadeWOFUL` compared to `CascadeLinUCB`.[7] Therefore, we use a tighter parameter for a fairer comparison.

Synthetic θ and ϕ. For the linear case of the synthetic data experiments, we randomly sample $\bar{w}(e)$ as described in Section 6, then construct θ and ϕ as follows. First, we generate θ and γ_1 uniformly on the sphere. Next, we define $\gamma_2 = \gamma_1 - \langle \gamma_1, \theta \rangle \theta$ and $\bar{\theta} = \gamma_2 / \|\gamma_2\|_2$. Note that θ and $\bar{\theta}$ are on the unit sphere and are orthogonal. Therefore, if we define $\phi(e) = \bar{w}(e)\theta + \sqrt{1 - \bar{w}(e)^2}\bar{\theta}$, we obtain

$$\langle \phi(e), \theta \rangle = \bar{w}(e)\|\theta\|_2^2 + \sqrt{1 - \bar{w}(e)^2}\langle \bar{\theta}, \theta \rangle = \bar{w}(e),$$
$$\|\phi(e)\|_2^2 = \bar{w}(e)^2\|\theta\|_2^2 + (1 - \bar{w}(e)^2)\|\bar{\theta}\|_2^2 + 2\bar{w}(e)\sqrt{1 - \bar{w}(e)^2}\langle \theta, \bar{\theta} \rangle = 1.$$

In summary, the construction above yields unit-norm features $\{\phi(e)\}_{e=1}^L$ and a unit-norm parameter θ such that $\langle \phi(e), \theta \rangle = \bar{w}(e)$ for each $e \in [L]$.[8]

MovieLens, all genre experiment. From the ratings data, we extract $W_{all} \in \{0,1\}^{m \times L}$, where $m = 6040$ is the number of users, $L = 3952$ is the number of movies, and $W_{all}(j, e) = 1$ if and only if the j-th user rated the e-th movie more than three stars. For each experimental trial, we randomly partition the rows of W_{all} into $W_{train} \in \{0,1\}^{(m/2) \times L}$ and $W_{test} \in \{0,1\}^{(m/2) \times L}$, i.e., we partition the users into train and test sets. We compute a rank-d SVD approximation $W_{train} \approx U \Sigma V^\mathsf{T}$, where $U \in \mathbb{R}^{(m/2) \times d}$ and $V \in \mathbb{R}^{L \times d}$ have orthogonal columns and $\Sigma \in \mathbb{R}^{d \times d}$ is diagonal and nonnegative. Let $\{\tilde{\phi}(e)\}_{e=1}^L$ be the rows of $V\Sigma$ (viewed as column vectors), $\iota = \max_{e \in [L]} \|\tilde{\phi}(e)\|_2$, $\phi(e) = \tilde{\phi}(e)/\iota$, $1_{m/2} \in \mathbb{R}^{m/2}$ the column vector of ones, and $\theta = 2\iota U^\mathsf{T} 1_{m/2}/m$. Notice that

$$\phi(e)^\mathsf{T}\theta = \frac{V(e,:)\Sigma U^\mathsf{T} 1_{m/2}}{m/2} \approx \frac{W_{train}^\mathsf{T}(e,:)1_{m/2}}{m/2} = \frac{\sum_{j=1}^{m/2} W_{train}(j,e)}{m/2},$$

[7] The looser parameter is good enough to prove Theorem 4, so this is mostly an experimental issue.

[8] This may seem overcomplicated; for example, we could have simply generated θ on the unit sphere and set $\phi(e) = \bar{w}(e)\theta$ to ensure $\langle \phi(e), \theta \rangle = \bar{w}(e)$. However, the latter is a bit unsatisfying, because then $\|\phi(e)\|_2 = \bar{w}(e)\|\theta\|_2 = \bar{w}(e)$, so knowledge of the features implies knowledge of the rewards.

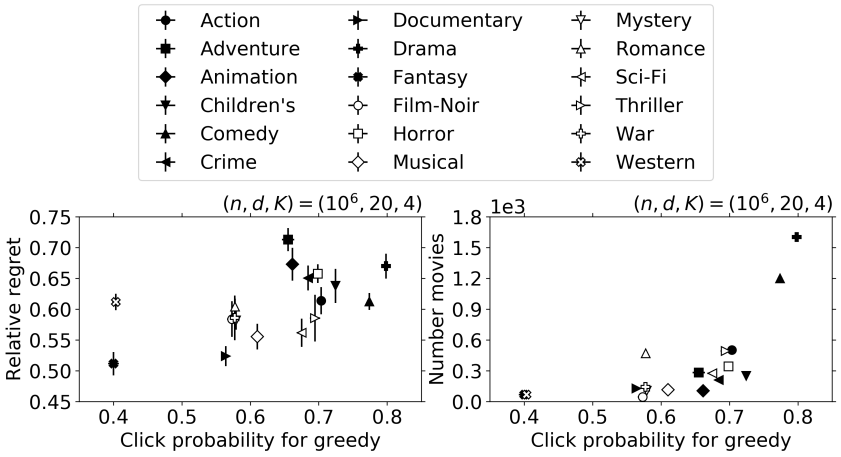

Figure 3: Additional results for MovieLens-1M data

i.e., $\langle \phi(e), \theta \rangle$ is (approximately) the fraction of users in the training set who rated e more than 3 stars. Therefore, if this fraction is roughly similar across train and test sets, and if the implicit low-rank assumption approximately holds, then $\langle \phi(e), \theta \rangle$ is (roughly) the probability that a uniformly random user from the test set rates e more than 3 stars. Consequently, ϕ is a reasonable feature mapping if the Bernoulli click feedback $\{\mathbf{w}_t(\mathbf{a}_k^t)\}_{k=1}^K$ is replaced by $\{W_{test}(\mathbf{J}_t, \mathbf{a}_k^t)\}_{k=1}^K$, where $\mathbf{J}_t$ is sampled uniformly from the $m/2$ users in the test set, as mentioned in Section 6.

Two technical clarifications are in order. First, note that we normalized $\tilde{\phi}(e)$ to ensure $\phi(e)$ has unit norm and absorbed the normalization constant ι into θ. We initially ran the experiment without this normalization, but this yielded $\max_{e \in [L]} \|\phi(e)\|_2 \geq 10$, which violates the assumptions of Theorem 4. In contrast, the normalization ensures $\max_{e \in [L]} \|\phi(e)\|_2 \leq 1$ (as in the theorem), and we found in practice that θ remained in the unit ball (again, as in the theorem) after absorbing the constant ι.

The second clarification is with regards to the optimal policy. In particular, replacing the independent Bernoullis $\{\mathbf{w}_t(\mathbf{a}_k^t)\}_{k=1}^K$ with $\{W_{test}(\mathbf{J}_t, \mathbf{a}_k^t)\}_{k=1}^K$ means that they are no longer conditionally independent given $\{\mathbf{a}_k^t\}_{k=1}^K$. As a consequence, finding the optimal policy is computationally hard. However, as discussed by Zong et al. [2016], one can efficiently compute a constant factor approximation via the greedy algorithm that iteratively adds the item to the optimal set which attracts the most users (from the test set) who are not already attracted to at least one optimal item. We therefore use this greedy poilcy (instead of the globally optimal one) when computing regret.

MovieLens, per genre experiment. For each genre, we run the same experiment but only include the movies whose lists of genres contain the given one. The left plot in Figure 3 shows the regret of `CacadeWOFUL` relative to that of `CascadeLinUCB`, as a function of the click probability for the aforementioned greedy policy that we treat as optimal. Aside from some apparent outliers (e.g., "Western"), the two quantities are roughly correlated. This suggests that `CacadeWOFUL` offers the most dramatic improvement when even the greedy policy struggles for clicks, which is reminiscent of the small click-through rate of Section 1.2. Additionally, we note that the average (across genres) of the relative regret is about 60%, which is better than the 66% figure quoted in Section 6 for the experiment that ignored genres. The reason for this discrepancy seems to be that genres with higher click rates under the near-optimal greedy policy (for which the relative regret is also higher – e.g., "Comedy" and "Drama" in the left plot) are overrepresented in the dataset, as shown at right.

Other information on experiments. Python code to recreate Figures 1, 2, and 3 is included in the supplementary material. All experiments were run on laptop with a 2.4GHz processor and 16GB of memory. The experiments took on the order of hours to complete – specifically, < 20 hours when the three Python files corresponding to synthetic tabular data, synthetic linear data, and real data are executed in parallel, or ≈ 27 hours when executed sequentially via the shell script included with the supplementary material. See the "README.txt" file therein for further details.

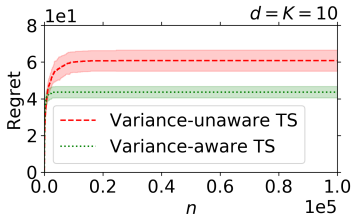

Figure 4: Results for Thompson sampling experiment discussed in Appendix D

D Future directions

As mentioned in the conclusion, our work leaves several problems open.

Linear lower bounds. While our upper and lower bounds match in the tabular case (up to logarithmic terms), we currently lack a lower bound for the linear case. We conjecture that this bound is $\Omega(\sqrt{n}d)$, which would demonstrate that our algorithm's upper bound $\tilde{O}(\sqrt{nd(d+K)})$ is nearly optimal (at least when $K = \tilde{O}(d)$, which we feel is the more reasonable case in applications). This conjecture is based on the observations that (1) our tabular lower bound $\Omega(\sqrt{nL})$ matches the non-cascading (i.e., multi-armed bandit) bound up to constants, and (2) in the linear case, the analogous non-cascading bound is $\Omega(\sqrt{n}d)$ (see, e.g., Chapter 24 of Lattimore and Szepesvári [2020] and the references therein). We have yet to prove this conjecture, but we believe it can be established using the ideas behind our tabular lower bound that were discussed in Section 5 – first, linearize the regret; next, lower bound the $\inf$ over cascading policies Π by the $\inf$ over policies Π' which observe the entire history $\mathcal{H}'_t$; and finally, lower bound the resulting non-cascading regret.

Thompson sampling (TS) for cascading bandits. We focused on UCB algorithms, but TS solutions have also been proposed. An excellent reference is the recent journal paper by Zhong et al. [2021]. Among other results, they prove $\tilde{O}(\sqrt{nLK})$ regret for their tabular algorithm `TS-Cascade`. Note this algorithm explicitly uses the empirical variance $\hat{\mathbf{v}}_t(e)$ as in `UCB-V` (3) (see their Algorithm 2), so our ideas may be able to improve its regret bound to $\tilde{O}(\sqrt{nL})$ (the same improvement Theorem 2 establishes for variance-aware UCB compared to prior work). On the other hand, their linear algorithm `LinTS-Cascade` chooses items e greedily with respect to $\langle \phi(e), \theta^{TS}_{t,H} \rangle$, where $\theta^{TS}_{t,H}$ is a Gaussian TS with variance-*un*aware mean $\hat{\theta}_{t,H}$ and covariance $\mathbf{\Lambda}^{-1}_{t,H}$ (in the notation of our Algorithm 2). Hence, we believe `LinTS-Cascade` can be improved by replacing $\hat{\theta}_{t,H}$ and $\mathbf{\Lambda}^{-1}_{t,H}$ with our variance-aware analogues $\hat{\theta}_{t,B}$ and $\mathbf{\Lambda}^{-1}_{t,B}$. A tentative numerical comparison is shown in Figure 4, which confirms this belief under the experimental setup from the bottom right plot in Figure 1. See Section 7 of Zhong et al. [2021] for a detailed numerical comparison of existing TS and UCB approaches.

E Notes on proofs

The remaining five appendices contain the complete proofs of our theoretical results. We begin with a general upper bound in Appendix F, which is used to prove the upper bounds from Theorems 2 and 4 in Appendices G and H, respectively. We then prove the lower bounds from Theorems 1 and 3 in Appendices I and J, respectively. Throughout, all random variables are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with $\bar{A} = \Omega \setminus A$ denoting the complement of $A \in \mathcal{F}$ and $\sigma(\mathcal{X})$ the σ-algebra generated by a set of random variables $\mathcal{X}$. Other notation is local to each appendix; for example, Δ is a small positive number that takes different values in different appendices.

F General gap-free regret decomposition

As mentioned in the footnotes of Section 5 (and the previous appendix), the proofs of our upper bounds rely on a general regret decomposition (Lemma 1 below). More specifically, this decomposition is intended for use in gap-free regret analyses of cascading bandit algorithms. Its proof is fairly straightforward but contains ideas that are crucial in formalizing the variance-aware insight discussed

intuitively in Section 1.2. Moreover, it is general enough to be the starting point for the proofs of both our tabular and linear upper bounds. Therefore, it may be of independent interest.

Before stating this result, we introduce some notation. We fix an instance $\bar{w} \in [0,1]^L$ and a policy $\pi \in \Pi$ and suppress these objects in our notation; e.g., we write $R(n)$ instead of $R_{\pi,\bar{w}}(n)$. We will also assume the following, which is without loss of generality (after possibly relabeling items).

Assumption 1. *We have* $\bar{w}(1) \geq \cdots \geq \bar{w}(L)$.

Note that under this assumption, the optimal action is $A^* = (1, \ldots, K)$. We will therefore refer to $[K]$ as the *optimal items* and $[L] \setminus [K]$ as the *suboptimal items*.

Next, we recall some notation and basic observations from Kveton et al. [2015a]. For any optimal item $e^* \in [K]$ and suboptimal item $e \in [L] \setminus [K]$, define the reward gap $\Delta_{e,e^*} = \bar{w}(e^*) - \bar{w}(e) \in [0,1]$. Let $\rho_t : [K] \to [K]$ be any $\sigma(\mathcal{H}_t \cup \{\mathbf{A}_t\})$-measurable permutation satisfying $\mathbf{U}_t(\mathbf{a}_k^t) \geq \mathbf{U}_t(\rho_t(k))$ for each $k \in [K]$ (such a permutation exists by Theorem 1 of Kveton et al. [2015a]). For each optimal item $e^* \in [K]$, suboptimal item $e \in [L] \setminus [K]$, and time $t \in [n]$, define the event

$$G_{e,e^*,t} = \cup_{k=1}^K \{\mathbf{a}_k^t = e, \rho_t(k) = e^*, \mathbf{w}_t(\mathbf{a}_1^t) = \cdots = \mathbf{w}_t(\mathbf{a}_{k-1}^t) = 0\}.$$

Observe that when $G_{e,e^*,t}$ occurs, we have $\mathbf{a}_k^t = e$ and $\rho_t(k) = e^*$ for some $k \in [K]$, where the permutation ρ_t satisfies $\mathbf{U}_t(\mathbf{a}_k^t) \geq \mathbf{U}_t(\rho_t(k))$ for all $k \in [K]$. Therefore, we know that

$$G_{e,e^*,t} \subset \{\mathbf{U}_t(e) \geq \mathbf{U}_t(e^*)\} \; \forall \, e \in [L] \setminus [K], e^* \in [K], t \in [n]. \tag{11}$$

Furthermore, note that by definition, at most one of the events $\{G_{e,e^*,t}\}_{e^*=1}^K$ can occur for each (e,t) pair. If any of them do, then the reward from item e is observed at time t, so $\mathbf{T}_t(e) = \mathbf{T}_{t-1}(e) + 1$; if none of them do, the reward is not observed and $\mathbf{T}_t(e) = \mathbf{T}_{t-1}(e)$. Consequently, we have

$$\sum_{e^*=1}^K \mathbb{1}(G_{e,e^*,t}) = \mathbb{1}(\mathbf{T}_t(e) = \mathbf{T}_{t-1}(e) + 1) = \mathbf{T}_t(e) - \mathbf{T}_{t-1}(e) \; \forall \, e \in [L] \setminus [K], t \in [n]. \tag{12}$$

Finally, since the rewards from at most K items are observed as time t, (12) implies

$$\sum_{e \in \mathscr{L}} \sum_{e^* \in \mathscr{K}} \mathbb{1}(G_{e,e^*,t}) \leq K \; \forall \, \mathscr{L} \subset [L], \mathscr{K} \subset [K], t \in [n]. \tag{13}$$

We point the reader to Theorem 1 of Kveton et al. [2015a] and its proof for further explanation.

We are now in position to state and prove the main result of this appendix.

Lemma 1. *Fix* $\Delta \in (0, \max\{\bar{w}(K)/2, 1/K\})$ *and, for each* $e \in [L] \setminus [K]$, *denote the optimal items* Δ*-close to* e *by* $\mathcal{K}_\Delta(e) = \{e^* \in [K] : \Delta_{e,e^*} \leq \Delta\}$ *and those* Δ*-far from* e *by* $\bar{\mathcal{K}}_\Delta(e) = [K] \setminus \mathcal{K}_\Delta$. *Let* $\mathcal{E}_t \in \sigma(\mathcal{H}_t \cup \{\mathbf{A}_t\})$ *and* $\mathcal{F}_t \in \mathcal{F}$ *for each* $t \in [n]$, *and define*

$$R_1(n) = \mathbb{E}\left[\sum_{t=1}^n \sum_{e=K+1}^L \sum_{e^* \in \bar{\mathcal{K}}_\Delta(e)} \Delta_{e,e^*} \mathbb{1}(\bar{\mathcal{E}}_t, \bar{\mathcal{F}}_t, G_{e,e^*,t})\right],$$

$$R_2(n) = \sum_{t=1}^n \mathbb{P}(\mathcal{E}_t), \quad R_3(n) = K\mathbb{E}\left[\sum_{t=1}^n \mathbb{1}(\mathcal{F}_t)\right].$$

Then under Assumption 1, $R(n) \leq \Delta n \min\{2/\bar{w}(K), K\} + \sum_{j=1}^3 R_j(n)$.

Remark 7. *When applying Lemma 1, we will choose* $\bar{\mathcal{E}}_t$ *as a "good event" under which the empirical and true mean rewards are close at time* t. *Thus,* R_1 *is the regret incurred from suboptimal items* e *that are chosen in favor of optimal items* $e^* \in \bar{\mathcal{K}}_\Delta(e)$ *with mean reward at least* Δ *higher, on this good event. The term* R_2 *accounts for the failure of the good events, which will occur with low probability due to concentration. The term* R_3 *is needed for the linear case, where* $\mathcal{F}_t \notin \sigma(\mathcal{H}_t \cup \{\mathbf{A}_t\})$ *may occur for our choice of* $\mathcal{F}_t$ *(see the beginning of Appendix H for details).*[9] *The term* $\Delta n \min\{2/\bar{w}(K), K\}$ *accounts choosing* e *instead of* $e^* \in \mathcal{K}_\Delta(e)$ *with gap* $\Delta_{e,e^*} \leq \Delta$, *which is the cascading analogue of the standard (non-cascading) bound* Δn *discussed in the first paragraph of Section 1.2.*

[9] In such cases, $\mathcal{F}_t$ cannot be absorbed into $\mathcal{E}_t$, which leads to an additional factor of K in R_3 compared to R_2.

Proof. By the choice of $\mathcal{E}_t$ and the same logic as Appendix A.1 of Kveton et al. [2015a], we know

$$R(n) \leq \mathbb{E}\left[\sum_{t=1}^{n}\sum_{e=K+1}^{L}\sum_{e^*=1}^{K}\Delta_{e,e^*}\mathbb{1}(\bar{\mathcal{E}}_t, G_{e,e^*,t})\right] + R_2(n). \tag{14}$$

For the first term, we fix t and e and separately analyze regret due to $e^* \in \mathcal{K}_\Delta(e)$ and $e^* \in \bar{\mathcal{K}}_\Delta(e)$.

For $\mathcal{K}_\Delta(e)$, we consider two cases. For the first case, suppose $\bar{w}(K)/2 \geq \max\{\bar{w}(e), 1/K\}$. Then for any $e^* \in [K]$, we know that

$$\Delta_{e,e^*} = \bar{w}(e^*) - \bar{w}(e) \geq \bar{w}(e^*) - \bar{w}(K)/2 \geq \bar{w}(K)/2 = \max\{\bar{w}(K)/2, 1/K\} > \Delta,$$

where the inequalities hold by $\bar{w}(K)/2 \geq \bar{w}(e)$, Assumption 1, and the assumption on Δ, respectively. Therefore, $\mathcal{K}_\Delta(e) = \emptyset$ by definition in the first case. For the second case, suppose $\bar{w}(K)/2 < \max\{\bar{w}(e), 1/K\}$. Then by definition of $\mathcal{K}_\Delta(e)$ and (12),

$$\sum_{e^*\in\mathcal{K}_\Delta(e)}\Delta_{e,e^*}\mathbb{1}(\bar{\mathcal{E}}_t, G_{e,e^*,t}) \leq \Delta\sum_{e^*\in\mathcal{K}_\Delta(e)}\mathbb{1}(G_{e,e^*,t}) \leq \Delta(\mathbf{T}_t(e) - \mathbf{T}_{t-1}(e)).$$

Combining the cases, we conclude that

$$\sum_{e^*\in\mathcal{K}_\Delta(e)}\Delta_{e,e^*}\mathbb{1}(\bar{\mathcal{E}}_t, G_{e,e^*,t}) \leq \Delta(\mathbf{T}_t(e) - \mathbf{T}_{t-1}(e))\mathbb{1}(\bar{w}(K) < 2\max\{\bar{w}(e), 1/K\}). \tag{15}$$

For $\bar{\mathcal{K}}_\Delta(e)$, we use $\bar{\mathcal{E}}_t \cap G_{e,e^*,t} \subset (\bar{\mathcal{E}}_t \cap \bar{\mathcal{F}}_t \cap G_{e,e^*,t}) \cup (\mathcal{F}_t \cap G_{e,e^*,t})$ and $\Delta_{e,e^*} \leq 1$ to write

$$\Delta_{e,e^*}\mathbb{1}(\bar{\mathcal{E}}_t, G_{e,e^*,t}) \leq \Delta_{e,e^*}\mathbb{1}(\bar{\mathcal{E}}_t, \bar{\mathcal{F}}_t, G_{e,e^*,t}) + \mathbb{1}(\mathcal{F}_t, G_{e,e^*,t}).$$

Summing this bound over $e^* \in \bar{\mathcal{K}}_\Delta(e)$ and again using (12), we obtain

$$\sum_{e^*\in\bar{\mathcal{K}}_\Delta(e)}\Delta_{e,e^*}\mathbb{1}(\bar{\mathcal{E}}_t, G_{e,e^*,t}) \leq \sum_{e^*\in\bar{\mathcal{K}}_\Delta(e)}\Delta_{e,e^*}\mathbb{1}(\bar{\mathcal{E}}_t, \bar{\mathcal{F}}_t, G_{e,e^*,t}) + \mathbb{1}(\mathcal{F}_t)(\mathbf{T}_t(e) - \mathbf{T}_{t-1}(e)). \tag{16}$$

Having separately analyzed $\mathcal{K}_\Delta(e)$ and $\bar{\mathcal{K}}_\Delta(e)$ for fixed $e \in [L] \setminus [K]$, we combine (15) and (16), use $[K] = \mathcal{K}_\Delta(e) \cup \bar{\mathcal{K}}_\Delta(e)$ (by definition), sum over $e \in [L] \setminus [K]$, and use (12)-(13) to write

$$\begin{aligned}\sum_{e=K+1}^{L}\sum_{e^*=1}^{K}\Delta_{e,e^*}\mathbb{1}(\bar{\mathcal{E}}_t, G_{e,e^*,t}) \leq \Delta&\sum_{e=K+1}^{L}(\mathbf{T}_t(e) - \mathbf{T}_{t-1}(e))\mathbb{1}(\bar{w}(K) < 2\max\{\bar{w}(e), 1/K\}) \\ &+ \sum_{e=K+1}^{L}\sum_{e^*\in\bar{\mathcal{K}}_\Delta(e)}\Delta_{e,e^*}\mathbb{1}(\bar{\mathcal{E}}_t, \bar{\mathcal{F}}_t, G_{e,e^*,t}) + \mathbb{1}(\mathcal{F}_t)K.\end{aligned}$$

Summing over $t \in [n]$ and taking expectation, the last two terms become $R_1(n)$ and $R_3(n)$, respectively. Plugging into (14), we have therefore shown

$$R(n) \leq \Delta\sum_{t=1}^{n}\mathbb{E}\left[\sum_{e=K+1}^{L}(\mathbf{T}_t(e) - \mathbf{T}_{t-1}(e))\mathbb{1}(\bar{w}(K) < 2\max\{\bar{w}(e), 1/K\})\right] + \sum_{j=1}^{3}R_j(n).$$

Thus, it suffices to upper bound the remaining expectation term by $\min\{2/\bar{w}(K), K\}$. Toward this end, first note that if $\bar{w}(K) < 2/K$, then we can (naively) upper bound the indicators by 1 and use (12)-(13) to bound this term by $K = \min\{2/\bar{w}(K), K\}$, which completes the proof. Therefore, we assume for the remainder that $\bar{w}(K) \geq 2/K$. Under this assumption, consider any $e \in [L] \setminus [K]$ for which $\bar{w}(K) < 2\max\{\bar{w}(e), 1/K\}$ (i.e., for which the indicator is 1). Then we must have $\bar{w}(e) > \bar{w}(K)/2$, because if instead $\bar{w}(e) \leq \bar{w}(K)/2$, we obtain a contradiction:

$$\bar{w}(K) < 2\max\{\bar{w}(e), 1/K\} \leq 2\max\{\bar{w}(K)/2, 1/K\} = \bar{w}(K).$$

In summary, if $e \in [L] \setminus [K]$ satisfies $\bar{w}(K) < 2\max\{\bar{w}(e), 1/K\}$ (i.e., if its indicator is 1), then it also satisfies $\bar{w}(e) > \bar{w}(K)/2$, and therefore $e \in \mathcal{L} \triangleq \{e' \in [L] : \bar{w}(e') \geq \bar{w}(K)/2\}$. Hence,

$$\sum_{e=K+1}^{L}(\mathbf{T}_t(e) - \mathbf{T}_{t-1}(e))\mathbb{1}(\bar{w}(K) < 2\max\{\bar{w}(e), 1/K\}) \leq \sum_{e\in\mathcal{L}}(\mathbf{T}_t(e) - \mathbf{T}_{t-1}(e)) \triangleq \mathbf{O}_t.$$

Next, we let $\mathbf{B}_t = (\mathbf{b}_1^t, \dots, \mathbf{b}_{|\mathbf{B}_t|}^t)$ be the sublist of $\mathbf{A}_t$ which contains only items from $\mathcal{L}$, with the same relative ordering as $\mathbf{A}_t$. Notice that $\mathbf{O}_t$ is the number of $e \in \mathcal{L}$ whose reward $\mathbf{w}_t(e)$ was observed at time t. Therefore, we can rewrite and then upper bound this quantity as follows:

$$\mathbf{O}_t = \sum_{k=1}^{K} \mathbb{1}(\mathbf{a}_k^t \in \mathcal{L}) \prod_{i=1}^{k-1} (1 - \mathbf{w}_t(\mathbf{a}_i^t)) \leq \sum_{k=1}^{|\mathbf{B}_t|} \prod_{i=1}^{k-1} (1 - \mathbf{w}_t(\mathbf{b}_i^t)).$$

By the conditional independence assumption (see Section 2) and the definition of $\mathcal{L}$, we then obtain

$$\mathbb{E}_t[\mathbf{O}_t] \leq \sum_{k=1}^{|\mathbf{B}_t|} \prod_{i=1}^{k-1} (1 - \bar{w}(\mathbf{b}_i^t)) \leq \sum_{k=1}^{|\mathbf{B}_t|} (1 - \bar{w}(K)/2)^{k-1} \leq \sum_{k=0}^{\infty} (1 - \bar{w}(K)/2)^k = 2/\bar{w}(K).$$

Taking expectation shows $\mathbb{E}[\mathbf{O}_t] \leq 2/\bar{w}(K) = \min\{2/\bar{w}(K), K\}$, which completes the proof. □

G Proof of Theorem 2

We prove the bound for `CascadeKL-UCB` and discuss how to modify the proof for `CascadeUCB-V` in Appendix G.1. Throughout, we adopt the notation of Appendix F and assume the following.

Assumption 2. *We have $\bar{w}(1) \geq \dots \geq \bar{w}(L)$ and $n > L$.*

As in Appendix F, the first inequality is without loss of generality. Thus, Assumption 2 is *with* loss of generality only when the second inequality fails, i.e., only when $n \leq L$. But in this case, we can naively bound regret by $n \leq \sqrt{nL}$, so the conclusion of Theorem 2 is immediate.

To begin the proof, we will invoke Lemma 1 with a particular choice of Δ, $\mathcal{E}_t$, and $\mathcal{F}_t$. First, we choose $\Delta = \max\{\bar{w}(K)/2, 1/K\}\sqrt{L/n}$. Next, for each $e \in [L] \setminus [K]$ and $e^* \in \bar{\mathcal{K}}_\Delta(e)$, we let

$$\tau_{e,e^*} = 8\bar{w}(e^*) \max\{\log(n^2 K), \log(f(n))\}/\Delta_{e,e^*}^2.$$

Note that $\Delta_{e,e^*} \geq \Delta > 0$ (by $e^* \in \bar{\mathcal{K}}_\Delta(e)$ and choice of Δ, respectively), so this quantity is well-defined. Intuitively, it will upper bound the number of plays needed to distinguish the suboptimal item e from the optimal item e^*.[10] Finally, for each $t \in [n]$, we define

$$\mathcal{E}_t = \cup_{e=K+1}^{L} \cup_{e^* \in \bar{\mathcal{K}}_\Delta(e)} \{\mathbf{U}_t(e^*) \leq \mathbf{U}_t(e), \mathbf{T}_{t-1}(e) \geq \tau_{e,e^*}\}, \quad \mathcal{F}_t = \emptyset.$$

Observe that $\Delta \in (0, \max\{\bar{w}(K)/2, 1/K\})$ (by Assumption 2), $\mathcal{E}_t \in \sigma(\mathcal{H}_t \cup \{\mathbf{A}_t\})$, and $\mathcal{F}_t \in \mathcal{F}$, as required by Lemma 1. Therefore, we can specialize the bound from that lemma to obtain

$$R(n) \leq \Delta n \min\{2/\bar{w}(K), K\} + \sum_{j=1}^{3} R_j(n) = \sqrt{nL} + \sum_{j=1}^{2} R_j(n).$$

In light of this inequality and Assumption 2 (the latter of which implies $K \leq L \leq \sqrt{nL}$), it suffices to show $R_j(n) = \tilde{O}(\sqrt{nL} + L + K)$ for each $j \in [2]$. We do so in the next two lemmas.

Lemma 2. *Under Assumption 2, $R_1(n) \leq 24\sqrt{nL} \max\{\log(n^2 K), \log(f(n))\}(\log(n) + 1) + L$.*

Proof. If $n = K = 1$, then $\Delta_{e,e^*} \leq 1$ and (13) imply $R_1(n) \leq nK = 1$, so the bound is immediate. Hence, we assume $\max\{n, K\} > 1$. Fix $e \in [L] \setminus [K]$. Then for any $e^* \in \bar{\mathcal{K}}_\Delta(e)$, we know

$$\bar{\mathcal{E}}_t \subset \{\mathbf{U}_t(e^*) > \mathbf{U}_t(e)\} \cup \{\mathbf{T}_{t-1}(e) < \tau_{e,e^*}\}.$$

Combined with (11) and the choice $\mathcal{F}_t = \emptyset$, we conclude

$$\begin{aligned}\bar{\mathcal{E}}_t \cap \bar{\mathcal{F}}_t \cap G_{e,e^*,t} &\subset (\{\mathbf{U}_t(e^*) > \mathbf{U}_t(e)\} \cup \{\mathbf{T}_{t-1}(e) < \tau_{e,e^*}\}) \cap \Omega \cap G_{e,e^*,t} \\ &= \{\mathbf{T}_{t-1}(e) < \tau_{e,e^*}\} \cap G_{e,e^*,t}.\end{aligned}$$

[10] This bound is likely loose for some problem instances, but choosing a sharper τ_{e,e^*} only improves log terms in the ultimate regret bound (which holds for all problem instances).

Therefore, the regret that e contributes to R_1 is at most (the expected value of)

$$\sum_{t=1}^{n} \sum_{e^* \in \bar{\mathcal{K}}_\Delta(e)} \Delta_{e,e^*} \mathbb{1}(\mathbf{T}_{t-1}(e) < \tau_{e,e^*}) \mathbb{1}(G_{e,e^*,t}). \tag{17}$$

Next, let $C = 8\max\{\log(n^2K), \log(f(n))\} > 0$ denote the term in the definition of τ_{e,e^*} that does not depend on the (e, e^*) pair. Then for any $e^* \in \bar{\mathcal{K}}_\Delta(e)$, we can write

$$\frac{\Delta_{e,e^*}\tau_{e,e^*}}{C} = \frac{\bar{w}(e^*)}{\Delta_{e,e^*}} = \frac{\bar{w}(e)}{\Delta_{e,e^*}} + 1 \le \frac{\bar{w}(K)}{\Delta_{e,e^*}} + 1 \le 2\sqrt{\frac{n}{L}} + 1 \le 3\sqrt{\frac{n}{L}},$$

where the first and third inequalities hold by Assumption 2, the second inequality uses $\Delta_{e,e^*} > \Delta \ge (\bar{w}(K)/2)\sqrt{L/n}$ (by definition of $\bar{\mathcal{K}}_\Delta(e)$ and Δ, respectively), and the equalities hold by definition. Next, using the bound $\mathbb{1}(x < y) \le y/x$ for $x, y > 0$, we write

$$\mathbb{1}(\mathbf{T}_{t-1}(e) < \tau_{e,e^*})\mathbb{1}(\mathbf{T}_{t-1}(e) > 0) \le \tau_{e,e^*}\mathbb{1}(\mathbf{T}_{t-1}(e) > 0)/\mathbf{T}_{t-1}(e),$$

where by convention $\mathbb{1}(\mathbf{T}_{t-1}(e) > 0)/\mathbf{T}_{t-1}(e) = 0$ when $\mathbf{T}_{t-1}(e) = 0$. Finally, note that for any $[0, 1]$-valued random variable $\mathbf{X}$, we have

$$\mathbf{X} = \mathbf{X}\mathbb{1}(\mathbf{T}_{t-1}(e) = 0) + \mathbf{X}\mathbb{1}(\mathbf{T}_{t-1}(e) > 0) \le \mathbb{1}(\mathbf{T}_{t-1}(e) = 0) + \mathbf{X}\mathbb{1}(\mathbf{T}_{t-1}(e) > 0).$$

Combining the previous three inequalities (with $\mathbf{X} = \Delta_{e,e^*}\mathbb{1}(\mathbf{T}_{t-1}(e) < \tau_{e,e^*})$ in the third) gives

$$\begin{aligned}
\Delta_{e,e^*}\mathbb{1}(\mathbf{T}_{t-1}(e) < \tau_{e,e^*}) &\le \mathbb{1}(\mathbf{T}_{t-1}(e) = 0) + \Delta_{e,e^*}\mathbb{1}(\mathbf{T}_{t-1}(e) < \tau_{e,e^*})\mathbb{1}(\mathbf{T}_{t-1}(e) > 0) \\
&\le \mathbb{1}(\mathbf{T}_{t-1}(e) = 0) + \Delta_{e,e^*}\tau_{e,e^*}\mathbb{1}(\mathbf{T}_{t-1}(e) > 0)/\mathbf{T}_{t-1}(e) \\
&\le \mathbb{1}(\mathbf{T}_{t-1}(e) = 0) + 3C\sqrt{n/L}\mathbb{1}(\mathbf{T}_{t-1}(e) > 0)/\mathbf{T}_{t-1}(e) \triangleq \mathbf{V}_t(e).
\end{aligned}$$

Using this bound and (12), we conclude that (17) is upper bounded by

$$\begin{aligned}
&\sum_{t=1}^{n} \mathbf{V}_t(e) \sum_{e^* \in \bar{\mathcal{K}}_\Delta(e)} \mathbb{1}(G_{e,e^*,t}) \le \sum_{t=1}^{n} \mathbf{V}_t(e)\mathbb{1}(\mathbf{T}_t(e) = \mathbf{T}_{t-1}(e) + 1) \\
&= \sum_{t=1}^{n} \mathbb{1}(\mathbf{T}_t(e) = 1, \mathbf{T}_{t-1}(e) = 0) + 3C\sqrt{\frac{n}{L}} \sum_{t=1}^{n} \frac{\mathbb{1}(\mathbf{T}_t(e) = \mathbf{T}_{t-1}(e) + 1, \mathbf{T}_{t-1}(e) > 0)}{\mathbf{T}_{t-1}(e)}.
\end{aligned}$$

Therefore, because $\mathbf{T}_t(e) = 1, \mathbf{T}_{t-1}(e) = 0$ can only occur for one $t \in [n]$, and because

$$\sum_{t=1}^{n} \frac{\mathbb{1}(\mathbf{T}_t(e) = \mathbf{T}_{t-1}(e) + 1, \mathbf{T}_{t-1}(e) > 0)}{\mathbf{T}_{t-1}(e)} = \sum_{s=1}^{\mathbf{T}_{n-1}(e)} \frac{1}{s} \le \sum_{s=1}^{n} \frac{1}{s} \le \log(n) + 1,$$

we conclude that (17) (and thus the regret that e contributes to R_1) is upper bounded by

$$1 + 3C\sqrt{n/L}(\log(n) + 1) = 1 + 24\sqrt{n/L}\max\{\log(n^2K), \log(f(n))\}(\log(n) + 1).$$

Finally, summing this bound over $e \in [L] \setminus [K]$ completes the proof. □

Remark 8. *The bound in Lemma 2 can be sharpened in terms of constants and log terms using ideas from Kveton et al. [2015a], but this requires a more complicated proof. We opted for the simpler proof above because its ideas generalize more easily to the linear case.*

Lemma 3. *Under Assumption 2, $R_2(n) \le L + 7K \log\log n$.*

To prove this lemma, we require the following claim.

Claim 1. *For any $p, q \in (0, 1)$, $d(p, q) \ge (p - q)^2/(2\max\{p, q\})$.*

Proof. We prove the bound assuming $p < q$ (it follows analogously for $p > q$ and is immediate when $p = q$). For each $\Delta \in (-q, 1 - q)$, we define $g(\Delta) = d(q + \Delta, q)$. Then one can compute

$$g'(\Delta) = \log\left(\frac{q + \Delta}{q}\right) - \log\left(\frac{1 - q - \Delta}{1 - q}\right), \quad g''(\Delta) = \frac{1}{(q + \Delta)(1 - q - \Delta)} \ge \frac{1}{q + \Delta}.$$

Thus, $g(0) = g'(0) = 0$, so by Taylor's theorem with remainder, we can find $\Delta \in (p - q, 0)$ such that

$$d(p, q) = g(p - q) = \frac{g''(\Delta)(p - q)^2}{2} \ge \frac{(p - q)^2}{2(q + \Delta)} \ge \frac{(p - q)^2}{2q} = \frac{(p - q)^2}{2\max\{p, q\}}.$$ □

Proof of Lemma 3. Let $\mathcal{C}_t = \cup_{e^*=1}^{K}\{\mathbf{U}_t(e^*) < \bar{w}(e^*)\}$. Then by definition,

$$\mathcal{E}_t \subset (\mathcal{E}_t \cap \bar{\mathcal{C}}_t) \cup \mathcal{C}_t \subset \left(\cup_{e=K+1}^{L} \cup_{e^* \in \bar{\mathcal{K}}_\Delta(e)} \{\mathbf{U}_t(e) \geq \bar{w}(e^*), \mathbf{T}_{t-1}(e) \geq \tau_{e,e^*}\}\right) \cup \mathcal{C}_t.$$

Therefore, by the union bound, we obtain that

$$R_2(n) \leq \sum_{t=1}^{n} \sum_{e=K+1}^{L} \sum_{e^* \in \bar{\mathcal{K}}_\Delta(e)} \mathbb{P}(\mathbf{U}_t(e) \geq \bar{w}(e^*), \mathbf{T}_{t-1}(e) \geq \tau_{e,e^*}) + \sum_{t=1}^{n} \mathbb{P}(\mathcal{C}_t). \tag{18}$$

As shown in Appendix A.2 of Kveton et al. [2015a], the second term is at most $7K \log\log n$. Hence, it suffices to bound the first term by L. Toward this end, first note

$$\mathbb{P}(\mathbf{U}_t(e) \geq \bar{w}(e^*), \mathbf{T}_{t-1}(e) \geq \tau_{e,e^*}) \leq \sum_{s=\lceil \tau_{e,e^*} \rceil}^{t-1} \mathbb{P}(\mathbf{U}_t(e) \geq \bar{w}(e^*), \mathbf{T}_{t-1}(e) = s). \tag{19}$$

Now fix $s \in \{\lceil \tau_{e,e^*} \rceil, \ldots, t-1\}$. We claim

$$\{\mathbf{U}_t(e) \geq \bar{w}(e^*), \mathbf{T}_{t-1}(e) = s\} \subset \{\hat{\mathbf{w}}_s(e) \geq \bar{w}(e) + \Delta_{e,e^*}/2\}. \tag{20}$$

Suppose instead that $\mathbf{U}_t(e) \geq \bar{w}(e^*)$, $\mathbf{T}_{t-1}(e) = s$, and $\hat{\mathbf{w}}_s(e) < \bar{w}(e) + \Delta_{e,e^*}/2$ all occur. Then

$$d(\hat{\mathbf{w}}_s(e), \mathbf{U}_t(e)) = d(\hat{\mathbf{w}}_{\mathbf{T}_{t-1}(e)}(e), \mathbf{U}_t(e)) \leq \frac{\log f(t)}{\mathbf{T}_{t-1}(e)} = \frac{\log f(t)}{s} \leq \frac{\log f(n)}{\tau_{e,e^*}} \leq \frac{\Delta_{e,e^*}^2}{8\bar{w}(e^*)}, \tag{21}$$

where we used $\mathbf{T}_{t-1}(e) = s \geq \tau_{e,e^*}$ and the definitions of $\mathbf{U}_t(e)$ and τ_{e,e^*}. But by Lemma 10.2(c) of Lattimore and Szepesvári [2020] and Claim 1, $\mathbf{U}_t(e) \geq \bar{w}(e^*) > \bar{w}(e) + \Delta_{e,e^*}/2 > \hat{\mathbf{w}}_s(e)$ implies

$$d(\hat{\mathbf{w}}_s(e), \mathbf{U}_t(e)) \geq d(\hat{\mathbf{w}}_s(e), \bar{w}(e^*)) > d(\bar{w}(e) + \Delta_{e,e^*}/2, \bar{w}(e^*)) \geq \frac{(\Delta_{e,e^*}/2)^2}{2\bar{w}(e^*)} = \frac{\Delta_{e,e^*}^2}{8\bar{w}(e^*)},$$

with contradicts (21). This completes the proof of (20). We then write

$$\begin{aligned}
&\mathbb{P}(\mathbf{U}_t(e) \geq \bar{w}(e^*), \mathbf{T}_{t-1}(e) = s) \\
&\quad \leq \mathbb{P}(\hat{\mathbf{w}}_s(e) \geq \bar{w}(e) + \Delta_{e,e^*}/2) \leq \exp(-sd\left(\bar{w}(e) + \Delta_{e,e^*}/2, \bar{w}(e)\right)) \\
&\quad \leq \exp(-s\Delta_{e,e^*}^2/(8\bar{w}(e^*))) \leq \exp(-\tau_{e,e^*}\Delta_{e,e^*}^2/(8\bar{w}(e^*))) \leq 1/(n^2K),
\end{aligned}$$

where we used (20), the Chernoff bound, Claim 1, the choice $s \geq \tau_{e,e^*}$, and the definition of τ_{e,e^*}, respectively. Since $t \leq n$, this implies (19) is upper bounded by $1/(nK)$. Summing over $t \in [n]$, $e \in [L] \setminus [K]$ and $e^* \in \bar{\mathcal{K}}_\Delta(e)$ shows the first term in (18) is upper bounded by L, as desired. □

G.1 Analysis for `CascadeUCB-V`

For `CascadeUCB-V`, we invoke Lemma 1 with the same choice of Δ, $\mathcal{E}_t$, and $\mathcal{F}_t$, but we change the definition of τ_{e,e^*} to $\tau_{e,e^*} = C\bar{w}(e^*)/\Delta_{e,e^*}^2$, where here $C = 30\log(nK)$. As above, it suffices to show $R_j(n) = \tilde{O}(\sqrt{nL} + L + K)$ for each $j \in [2]$. For $j = 1$, this follows as in the proof of Lemma 2 (with a different C). For $j = 2$, by the same logic used to prove Lemma 3, it suffices to show

$$\sum_{t=1}^{n} \sum_{e^*=1}^{K} \mathbb{P}(\mathbf{U}_t(e^*) < \bar{w}(e^*)) = \tilde{O}(\sqrt{nL} + L + K), \tag{22}$$

$$\mathbb{P}(\mathbf{U}_t(e) \geq \bar{w}(e^*), \mathbf{T}_{t-1}(e) = s) \leq \frac{1}{n^2K} \; \forall\, e \in [L] \setminus [K], e^* \in \bar{\mathcal{K}}_\Delta(e), \tau_{e,e^*} \leq s \leq t \leq n. \tag{23}$$

To prove (22), we note that for any $t \in [n]$ and $e^* \in [K]$,

$$\begin{aligned}
\mathbb{P}(\mathbf{U}_t(e^*) < \bar{w}(e^*)) &= \sum_{s=1}^{t-1} \mathbb{P}(\mathbf{U}_t(e^*) < \bar{w}(e^*), \mathbf{T}_{t-1}(e) = s) \\
&\leq \sum_{s=1}^{t-1} \mathbb{P}\left(\hat{\mathbf{w}}_s(e^*) + \sqrt{\frac{4\hat{\mathbf{v}}_s(e^*)\log t}{s}} + \frac{6\log t}{s} < \bar{w}(e^*)\right) < 3t^{-1},
\end{aligned}$$

where the last bound holds because each summand is at most $3t^{-2}$ by Theorem 1 of Audibert et al. [2009]. Thus, the left side of (22) is at most $3K\sum_{t=1}^{n} t^{-1} = \tilde{O}(K)$, as desired.

To prove (23), fix such an e, e^*, s, and t, and suppose $\mathbf{U}_t(e) \geq \bar{w}(e^*)$ and $\mathbf{T}_{t-1}(e) = s$. Then by definition of $\mathbf{U}_t(e)$ and $\hat{\mathbf{v}}_s(e)$ (see (3) and ensuing discussion), we have

$$\hat{\mathbf{w}}_s(e) + \sqrt{\frac{4\hat{\mathbf{w}}_s(e)\log t}{s}} + \frac{6\log t}{s} \geq \mathbf{U}_t(e) \geq \bar{w}(e^*) = \bar{w}(e) + \Delta_{e,e^*}. \tag{24}$$

Next, we consider two cases. First, if $\bar{w}(e) = 0$, then $\hat{\mathbf{w}}_s(e) = 0$ and $\bar{w}(e^*) = \Delta_{e,e^*}$, so by (24),

$$s \leq 6\log(t)/\Delta_{e,e^*} = 6\log(t)\bar{w}(e^*)/\Delta_{e,e^*}^2 < \tau_{e,e^*},$$

which contradicts the choice of s and implies (23). The second case is $\bar{w}(e) > 0$. Here we observe that when (24) holds, one of the following inequalities must also hold:

$$\hat{\mathbf{w}}_s(e) \geq \bar{w}(e) + \frac{2\Delta_{e,e^*}}{5}, \quad \sqrt{\frac{4\hat{\mathbf{w}}_s(e)\log t}{s}} \geq \frac{2\Delta_{e,e^*}}{5}, \quad \frac{6\log t}{s} \geq \frac{\Delta_{e,e^*}}{5}. \tag{25}$$

However, the third inequality cannot hold, because it and $\bar{w}(e^*) > \Delta_{e,e^*}$ (where the latter inequality is by definition in the case $\bar{w}(e) > 0$) together imply

$$s \leq 30\log(t)/\Delta_{e,e^*} < 30\bar{w}(e^*)\log(t)/\Delta_{e,e^*}^2 \leq 30\bar{w}(e^*)\log(nK)/\Delta_{e,e^*}^2 = \tau_{e,e^*},$$

which contradicts the choice of s. Furthermore, if the second inequality holds, then

$$\hat{\mathbf{w}}_s(e) \geq \frac{s\Delta_{e,e^*}^2}{25\log t} \geq \frac{\tau_{e,e^*}\Delta_{e,e^*}^2}{25\log t} \geq \bar{w}(e^*) > \bar{w}(e) + \frac{2\Delta_{e,e^*}}{5},$$

so the first must also hold. Therefore, we have shown that if $\mathbf{U}_t(e) \geq \bar{w}(e^*)$ and $\mathbf{T}_{t-1}(e) = s$ both occur, then the first inequality in (25) must hold. It follows that

$$\begin{aligned}\mathbb{P}(\mathbf{U}_t(e) \geq \bar{w}(e^*), \mathbf{T}_{t-1}(e) = s) &\leq \mathbb{P}(\hat{\mathbf{w}}_s(e) \geq \bar{w}(e) + 2\Delta_{e,e^*}/5) \\ &\leq \exp(-2\tau_{e,e^*}\Delta_{e,e^*}^2/(25\bar{w}(e^*))) \leq (nK)^{-2},\end{aligned}$$

where the second inequality uses the Chernoff bound and Claim 1 similar to the proof of Lemma 3. This is tighter than the bound in (23) that we set out to prove so it concludes the analysis.

H Proof of Theorem 4

As in Appendix G, we adopt the notation of Appendix F. We will also assume the following.

Assumption 3. *There exists $\phi : [L] \to \mathscr{B}_d(1)$ and $\theta \in \mathscr{B}_d(1)$ such that $\bar{w}(e) = \langle \theta, \phi(e)\rangle$ for each $e \in [L]$. Furthermore, we have $\bar{w}(1) \geq \cdots \geq \bar{w}(L)$ and $n > d(d+K)$.*

Analogous to the discussion of Assumption 2 in Appendix G, Assumption 3 is stronger than the assumptions of Theorem 4 only when $n \leq d(d+K)$. But in this case, we have

$$R(n) \leq n \leq \sqrt{nd(d+K)} \leq \sqrt{n\max\{d,K\}(2\max\{d,K\})} = \sqrt{2n}\max\{d,K\}, \tag{26}$$

so $R(n) = O(\min\{\sqrt{nd(d+K)}, \sqrt{n}\max\{d,K\}\})$, which is sharper than the desired bound.

Next, similar to Appendix G, we will invoke Lemma 1. First, we set

$$\Delta = \max\{\bar{w}(K)/2, 1/K\}\sqrt{d(d+K)/n} \in (0, \max\{\bar{w}(K)/2, 1/K\}),$$

where the upper bound holds by Assumption 3. Next, we let $\eta_s^k = \mathbf{w}_s(\mathbf{a}_k^s) - \bar{w}(\mathbf{a}_k^s)$ denote the k-th noise random variable at time s. We define the events

$$\mathcal{E}_{t,H} = \cup_{\tau=1}^{t}\left\{\left\|\sum_{s=1}^{\tau-1}\sum_{k=1}^{\min\{\mathbf{C}_s,K\}}\phi(\mathbf{a}_k^s)\eta_s^k\right\|_{\mathbf{\Lambda}_{\tau,H}^{-1}} > \beta_{\tau,H}\right\},$$

$$\mathcal{E}_{t,B} = \cup_{\tau=1}^{t}\left\{\left\|\sum_{s=1}^{\tau-1}\sum_{k=1}^{\min\{\mathbf{C}_s,K\}}\phi(\mathbf{a}_k^s)\eta_s^k/\mathbf{U}_{s,H}(\mathbf{a}_k^s)\right\|_{\mathbf{\Lambda}_{\tau,B}^{-1}} > \beta_{\tau,B}\right\},$$

$$\mathcal{F}_{t,H} = \{\det(\mathbf{\Lambda}_{t+1,H}) \geq 2\det(\mathbf{\Lambda}_{t,H})\}, \quad \mathcal{F}_{t,B} = \{\det(\mathbf{\Lambda}_{t+1,B}) \geq 2\det(\mathbf{\Lambda}_{t,B})\},$$

where, with $\alpha_{t,H}$ and $\alpha_{t,B}$ defined as in the theorem statement, we let

$$\beta_{t,H} = \alpha_{t,H} - 1 = \sqrt{d\log(1+tK/d) + 2\log(n)},$$
$$\beta_{t,B} = \alpha_{t,B} - \sqrt{K} = 8\sqrt{d\log(1+tK/d)\log(n^3K)} + 4\sqrt{K}\log(n^3K).$$

Finally, we set $\mathcal{E}_t = \mathcal{E}_{t,H} \cup \mathcal{E}_{t,B}$ and $\mathcal{F}_t = \mathcal{F}_{t,H} \cup \mathcal{F}_{t,B}$. Observe that $\mathcal{E}_t \in \sigma(\mathcal{H}_t \cup \{\mathbf{A}_t\})$, though $\mathcal{F}_t$ need not belong to this sub-σ-algebra because it depends on the random rewards (i.e., the presence or absence of clicks) at time t. Therefore, we use Lemma 1 to write

$$R(n) \leq \Delta n \min\{2/\bar{w}(K), K\} + \sum_{j=1}^{3} R_j(n) = \sqrt{nd(d+K)} + \sum_{j=1}^{3} R_j(n). \tag{27}$$

We claim that, in light of this bound, it suffices to show

$$R_j(n) = \tilde{O}\left(\sqrt{nd(d+K)} + n^{1/6}d^{7/6}(d+K)^{1/2}K^{1/3} + dK\right) \ \forall\, j \in [3]. \tag{28}$$

Indeed, if (28) holds, then since $dK \leq d(d+K) \leq \sqrt{nd(d+K)}$ by Assumption 3, (27) will imply

$$R(n) = \tilde{O}\left(\sqrt{nd(d+K)} + n^{1/6}d^{7/6}(d+K)^{1/2}K^{1/3}\right), \tag{29}$$

which is precisely the first term in the minimum of the Theorem 4 regret bound. Therefore, we only need to show that each summand in (29) is bounded by a constant multiple of the second term in the minimum of Theorem 4. For the first summand, this follows as in the third inequality of (26). For the second summand, we use Assumption 3 and $d + K \leq 2\max\{d, K\}$ to write

$$n^{1/6}d^{7/6}(d+K)^{1/2}K^{1/3} \leq \sqrt{n}d^{5/6}(d+K)^{1/6}K^{1/3} \leq 2^{1/6}\sqrt{n}\max\{d,K\}^{1/6}d^{5/6}K^{1/3},$$

which completes the proof by separately considering the cases the cases $d \geq K$ and $d < K$.

Hence, to prove the theorem, it only remains to prove (28). Toward this end, we first state two claims. The first is accuracy guarantee for the least-squares estimates $\langle \hat{\theta}_{t,\cdot}, \phi(e)\rangle$, which also asserts that $\mathbf{U}_{t,\cdot}(e)$ is a valid UCB (on the good event $\bar{\mathcal{E}}_{t,\cdot}$). The second is a version of the so-called elliptical potential lemma (see Lemma 11 and ensuing historical discussion in Abbasi-Yadkori et al. [2011]). Both claims are essentially known but are stated in forms convenient for our purposes, so the proofs are deferred to Appendix H.1. Here and moving forward, for any $t \in [n]$ and $e \in [L]$, we use $\phi_t(e) = \phi(e)/\sqrt{\mathbf{U}_{t,H}(e)}$ to denote the normalized features for item e at time t.

Claim 2. *Under Assumption 3, for any $t \in [n]$, $\tau \in [t]$, and $e \in [L]$, on the event $\bar{\mathcal{E}}_{t,H}$ we have*

$$|\langle \hat{\theta}_{\tau,H}, \phi(e)\rangle - \bar{w}(e)| \leq \alpha_{\tau,H}\|\phi(e)\|_{\mathbf{\Lambda}_{\tau,H}^{-1}}, \quad \mathbf{U}_{\tau,H}(e) \geq \bar{w}(e),$$

and similarly, on the event $\bar{\mathcal{E}}_{t,B}$ we have

$$|\langle \hat{\theta}_{\tau,B}, \phi(e)\rangle - \bar{w}(e)| \leq \alpha_{\tau,B}\|\phi(e)\|_{\mathbf{\Lambda}_{\tau,B}^{-1}}, \quad \mathbf{U}_{\tau,B}(e) \geq \bar{w}(e).$$

Claim 3. *Under Assumption 3, for any $t \in [n]$, we have*

$$\sum_{s=1}^{t} \mathbb{1}(\bar{\mathcal{F}}_s) \sum_{k=1}^{\min\{\mathbf{C}_s, K\}} \|\phi(\mathbf{a}_k^s)\|^2_{\mathbf{\Lambda}_{s,H}^{-1}} \leq 4\log\left(\frac{\det(\mathbf{\Lambda}_{t+1,H})}{\det(\mathbf{\Lambda}_{1,H})}\right) \leq 4d\log\left(1 + \frac{tK}{d}\right),$$
$$\sum_{s=1}^{t} \mathbb{1}(\bar{\mathcal{F}}_s) \sum_{k=1}^{\min\{\mathbf{C}_s, K\}} \|\phi_s(\mathbf{a}_k^s)\|^2_{\mathbf{\Lambda}_{s,B}^{-1}} \leq 4\log\left(\frac{\det(\mathbf{\Lambda}_{t+1,B})}{\det(\mathbf{\Lambda}_{1,B})}\right) \leq 4d\log\left(1 + \frac{tK}{d}\right).$$

We can now state and prove three lemmas that together establish (28).

Lemma 4. *Under Assumption 3,*

$$R_1(n) \leq 16d\left(\frac{8\sqrt{n}\alpha_{n,B}^2}{\sqrt{d(d+K)}} + \frac{3n^{1/6}K^{1/3}\alpha_{n,B}^{4/3}\alpha_{n,H}^{2/3}}{(d(d+K))^{1/6}}\right)\log\left(1 + \frac{nK}{d}\right)$$
$$= \tilde{O}\left(\sqrt{nd(d+K)} + n^{1/6}d^{7/6}(d+K)^{1/2}K^{1/3}\right).$$

Proof. First, we fix $t \in [n]$, $e \in [L] \setminus [K]$, and $e^* \in \bar{\mathcal{K}}_\Delta(e)$, and we assume that $\bar{\mathcal{E}}_t$, $\bar{\mathcal{F}}_t$, and $G_{e,e^*,t}$ all hold. Under this assumption, we can write

$$\mathbf{U}_{t,B}(e) \le \bar{w}(e) + 2\alpha_{t,B}\|\phi(e)\|_{\mathbf{\Lambda}_{t,B}^{-1}} = \bar{w}(e) + 2\alpha_{t,B}\sqrt{\mathbf{U}_{t,H}(e)}\|\phi_t(e)\|_{\mathbf{\Lambda}_{t,B}^{-1}}, \tag{30}$$

where we used Claim 2 and the definition of ϕ_t. We similarly have

$$\begin{aligned}\mathbf{U}_{t,H}(e) &\le \max\left\{\bar{w}(e) + 2\alpha_{t,H}\|\phi(e)\|_{\mathbf{\Lambda}_{t,H}^{-1}}, \frac{1}{K}\right\} \\ &\le 2\left(\max\left\{\frac{\bar{w}(K)}{2}, \frac{1}{K}\right\} + \alpha_{t,H}\|\phi(e)\|_{\mathbf{\Lambda}_{t,H}^{-1}}\right),\end{aligned} \tag{31}$$

where the first inequality again holds by Claim 2 and the second uses $\bar{w}(e) \le \bar{w}(K)$ by Assumption 3 and the fact that, for any $x, y, z \ge 0$,

$$2(\max\{x/2, z\} + y) = \max\{x + 2y, 2z + 2y\} \ge \max\{x + 2y, z\}.$$

Combining the two inequalities with the bound $\sqrt{x+y} \le \sqrt{x} + \sqrt{y}$ for $x, y \ge 0$, we obtain

$$\begin{aligned}\mathbf{U}_{t,B}(e) &\le \bar{w}(e) + 2\alpha_{t,B}\sqrt{2\max\{\bar{w}(K)/2, 1/K\}}\|\phi_t(e)\|_{\mathbf{\Lambda}_{t,B}^{-1}} \\ &\quad + 2\alpha_{t,B}\sqrt{2\alpha_{t,H}\|\phi(e)\|_{\mathbf{\Lambda}_{t,H}^{-1}}}\|\phi_t(e)\|_{\mathbf{\Lambda}_{t,B}^{-1}} \\ &\le \bar{w}(e) + \gamma_B\|\phi_t(e)\|_{\mathbf{\Lambda}_{t,B}^{-1}} + \gamma_H\sqrt{\|\phi(e)\|_{\mathbf{\Lambda}_{t,H}^{-1}}}\|\phi_t(e)\|_{\mathbf{\Lambda}_{t,B}^{-1}},\end{aligned} \tag{32}$$

where we defined $\gamma_B = 2\alpha_{n,B}\sqrt{2\max\{\bar{w}(K)/2, 1/K\}}$ and $\gamma_H = 2\alpha_{n,B}\sqrt{2\alpha_{n,H}}$. Furthermore, note that by Claim 2, we have $\mathbf{U}_{t,B}(e^*) \ge \bar{w}(e^*) = \bar{w}(e) + \Delta_{e,e^*}$ on $\bar{\mathcal{E}}_t$. Combined with the previous inequality and (11), we conclude that on $\bar{\mathcal{E}}_t \cap G_{e,e^*,t}$, we have

$$\begin{aligned}\Delta_{e,e^*} &\le \mathbf{U}_{t,B}(e^*) - \bar{w}(e) \le \mathbf{U}_{t,B}(e) - \bar{w}(e) \\ &\le \gamma_B\|\phi_t(e)\|_{\mathbf{\Lambda}_{t,B}^{-1}} + \gamma_H\sqrt{\|\phi(e)\|_{\mathbf{\Lambda}_{t,H}^{-1}}}\|\phi_t(e)\|_{\mathbf{\Lambda}_{t,B}^{-1}}.\end{aligned} \tag{33}$$

Next, notice that one of the summands at right must exceed $\Delta_{e,e^*}/2$ (else, their sum will be less than Δ_{e,e^*}, a contradiction), so one of the following events must occur:

$$\mathcal{A}_{e,e^*,t} = \left\{\Delta_{e,e^*}^2 \le (2\gamma_B)^2\|\phi_t(e)\|_{\mathbf{\Lambda}_{t,B}^{-1}}^2\right\},$$

$$\mathcal{B}_{e,e^*,t} = \left\{\Delta_{e,e^*}^{4/3} \le (2\gamma_H)^{4/3}\|\phi(e)\|_{\mathbf{\Lambda}_{t,H}^{-1}}^{2/3}\|\phi_t(e)\|_{\mathbf{\Lambda}_{t,B}^{-1}}^{4/3}\right\}.$$

In summary, we have shown $\bar{\mathcal{E}}_t \cap G_{e,e^*,t} \subset \mathcal{A}_{e,e^*,t} \cup \mathcal{B}_{e,e^*,t}$, which implies

$$\mathbb{1}(\bar{\mathcal{E}}_t, \bar{\mathcal{F}}_t, G_{e,e^*,t}) = \mathbb{1}(\bar{\mathcal{E}}_t, G_{e,e^*,t})\mathbb{1}(\bar{\mathcal{F}}_t, G_{e,e^*,t}) \le (\mathbb{1}(\mathcal{A}_{e,e^*,t}) + \mathbb{1}(\mathcal{B}_{e,e^*,t}))\mathbb{1}(\bar{\mathcal{F}}_t, G_{e,e^*,t}).$$

Furthermore, since $\mathbb{1}(x \le y) \le y/x$ for $x, y > 0$ and $2^{4/3} = 16^{1/3} < 27^{1/3} = 3$, we know that

$$\mathbb{1}(\mathcal{A}_{e,e^*,t}) + \mathbb{1}(\mathcal{B}_{e,e^*,t}) \le \frac{4\gamma_B^2}{\Delta_{e,e^*}^2}\|\phi_t(e)\|_{\mathbf{\Lambda}_{t,B}^{-1}}^2 + \frac{3\gamma_H^{4/3}}{\Delta_{e,e^*}^{4/3}}\|\phi(e)\|_{\mathbf{\Lambda}_{t,H}^{-1}}^{2/3}\|\phi_t(e)\|_{\mathbf{\Lambda}_{t,B}^{-1}}^{4/3}.$$

Combining the previous two inequalities and using $\Delta_{e,e^*} \ge \Delta$ for $e^* \in \bar{\mathcal{K}}_\Delta(e)$ by definition, we get

$$\Delta_{e,e^*}\mathbb{1}(\bar{\mathcal{E}}_t, \bar{\mathcal{F}}_t, G_{e,e^*,t}) \le \left(\frac{4\gamma_B^2}{\Delta}\|\phi_t(e)\|_{\mathbf{\Lambda}_{t,B}^{-1}}^2 + \frac{3\gamma_H^{4/3}}{\Delta^{1/3}}\|\phi(e)\|_{\mathbf{\Lambda}_{t,H}^{-1}}^{2/3}\|\phi_t(e)\|_{\mathbf{\Lambda}_{t,B}^{-1}}^{4/3}\right)\mathbb{1}(\bar{\mathcal{F}}_t, G_{e,e^*,t}).$$

Summing both sides over $e^* \in \bar{\mathcal{K}}_\Delta(e)$ and using (12), we conclude

$$\begin{aligned}&\sum_{e^* \in \bar{\mathcal{K}}_\Delta(e)} \Delta_{e,e^*}\mathbb{1}(\bar{\mathcal{E}}_t, \bar{\mathcal{F}}_t, G_{e,e^*,t}) \\ &\le \mathbb{1}(\bar{\mathcal{F}}_t)\left(\frac{4\gamma_B^2}{\Delta}\|\phi_t(e)\|_{\mathbf{\Lambda}_{t,B}^{-1}}^2 + \frac{3\gamma_H^{4/3}}{\Delta^{1/3}}\|\phi(e)\|_{\mathbf{\Lambda}_{t,H}^{-1}}^{2/3}\|\phi_t(e)\|_{\mathbf{\Lambda}_{t,B}^{-1}}^{4/3}\right)\mathbb{1}(\mathbf{T}_t(e) = \mathbf{T}_{t-1}(e) + 1).\end{aligned}$$

Now observe the right side is nonzero if and only if $e = \mathbf{a}_k^t$ for some $k \in [\min\{\mathbf{C}_t, K\}]$. Therefore, summing over $e \in [L] \setminus [K]$ and using nonnegativity of the summands gives

$$\sum_{e=K+1}^{L} \sum_{e^* \in \bar{\mathcal{K}}_\Delta(e)} \Delta_{e,e^*} \mathbb{1}(\bar{\mathcal{E}}_t, \bar{\mathcal{F}}_t, G_{e,e^*,t})$$
$$\leq \frac{4\gamma_B^2}{\Delta} \mathbb{1}(\bar{\mathcal{F}}_t) \sum_{k=1}^{\min\{\mathbf{C}_t,K\}} \|\phi_t(\mathbf{a}_k^t)\|^2_{\mathbf{\Lambda}_{t,B}^{-1}} + \frac{3\gamma_H^{4/3}}{\Delta^{1/3}} \mathbb{1}(\bar{\mathcal{F}}_t) \sum_{k=1}^{\min\{\mathbf{C}_t,K\}} \|\phi(\mathbf{a}_k^t)\|^{2/3}_{\mathbf{\Lambda}_{t,H}^{-1}} \|\phi_t(\mathbf{a}_k^t)\|^{4/3}_{\mathbf{\Lambda}_{t,B}^{-1}}.$$

Summed over $t \in [n]$, the remaining summations are both upper bounded by $4d\log(1 + nK/d)$ (by Claim 3, along with Holder's inequality for the second summation). We thus obtain

$$R_1(n) \leq 4d \left(\frac{4\gamma_B^2}{\Delta} + \frac{3\gamma_H^{4/3}}{\Delta^{1/3}} \right) \log\left(1 + \frac{nK}{d}\right). \tag{34}$$

Finally, recall that by definition, we have

$$\Delta = \max\{\bar{w}(K)/2, 1/K\}\sqrt{d(d+K)/n} \geq \sqrt{d(d+K)/(nK^2)}.$$

Plugging the middle and right expressions into the Δ and $\Delta^{1/3}$ terms in (34), respectively, and recalling $\gamma_B = 2\alpha_{n,B}\sqrt{2\max\{\bar{w}(K)/2, 1/K\}}$ and $\gamma_H = 2\alpha_{n,B}\sqrt{2\alpha_{n,H}}$, yields the first upper bound of the lemma. The second follows since $\alpha_{n,B} = \tilde{O}(\sqrt{d} + \sqrt{K})$ and $\alpha_{n,B} = \tilde{O}(\sqrt{d})$. □

Lemma 5. *Under Assumption 3, $R_2(n) \leq 4K + 1$.*

To prove Lemma 5, we require the following Bernstein-style concentration result. It is essentially implied by Theorem 2 of Zhou et al. [2021], but the proof requires some minor modifications to accommodate the fact that, unlike their work, we do not assume an almost-sure upper bound on the conditional variances $\mathbb{E}[\eta_t^2|\mathcal{G}_t]$. The details of these changes are discussed in Appendix H.2.

Claim 4. *Fix $L, R, \lambda, \delta > 0$.*[11] *Suppose $\{\mathcal{G}_t\}_{t=1}^T$ is a filtration and $\{\mathbf{x}_t, \eta_t\}_{t=1}^T$ is a stochastic process such that, for each $t \in [T]$, $\mathbf{x}_t$ is $\mathscr{B}_d(L)$-valued and $\mathcal{G}_t$-measurable, η_t is $[-R, R]$-valued and $\mathcal{G}_{t+1}$-measurable, and $\mathbb{E}[\eta_t|\mathcal{G}_t] = 0$. For each $t \in [T]$, define $\beta_t = 8\sqrt{d\log(1 + tL^2/(d\lambda))\log(4t^2/\delta)} + 4R\log(4t^2/\delta)$ and $\mathbf{Z}_t = \lambda I + \sum_{i=1}^t \mathbf{x}_i\mathbf{x}_i^\mathsf{T}$. Then*

$$\mathbb{P}\left(\cup_{t=1}^T \left\{ \left\| \sum_{i=1}^t \mathbf{x}_i\eta_i \right\|_{\mathbf{Z}_t^{-1}} > \beta_t \right\} \cap \cap_{t=1}^T \{\mathbb{E}[\eta_t^2|\mathcal{G}_t] \leq 1\} \right) \leq \delta.$$

Proof of Lemma 5. We first observe that by definition, the following holds for each $t \in [n]$:

$$\mathbb{P}(\mathcal{E}_t) = \mathbb{P}(\mathcal{E}_{t,H} \cup \mathcal{E}_{t,B}) = \mathbb{P}(\mathcal{E}_{t,H} \cup (\mathcal{E}_{t,B} \cap \bar{\mathcal{E}}_{t,H})) = \mathbb{P}(\mathcal{E}_{t,H}) + \mathbb{P}(\mathcal{E}_{t,B} \cap \bar{\mathcal{E}}_{t,H}).$$

Thus, because $R_2(n) = \sum_{t=1}^n \mathbb{P}(\mathcal{E}_t)$, it suffices to show

$$\mathbb{P}(\mathcal{E}_{t,H}) \leq 1/n, \quad \mathbb{P}(\mathcal{E}_{t,B} \cap \bar{\mathcal{E}}_{t,H}) \leq 4K/n \quad \forall\, t \in [n]. \tag{35}$$

We fix $t \in [n]$ for the remainder of the proof and establish the two bounds in (35) in turn.

To prove the first bound, for each $i \in [(t-1)K]$, we define

$$\rho(i) = \left\lfloor 1 + \frac{i-1}{K} \right\rfloor \in [t-1], \quad \sigma(i) = i - (\rho(i) - 1)K \in [K],$$

$$\tilde{\phi}_{i,H} = \phi(\mathbf{a}_{\sigma(i)}^{\rho(i)})\mathbb{1}(\sigma(i) \leq \mathbf{C}_{\rho(i)}), \quad \tilde{\eta}_{i,H} = \eta_{\rho(i)}^{\sigma(i)}, \quad \tilde{\mathbf{\Lambda}}_{i,H} = I + \sum_{j=1}^i \tilde{\phi}_{j,H}\tilde{\phi}_{j,H}^\mathsf{T}.$$

[11] To unify the presentation with Zhou et al. [2021], we adopt their notation for this claim only. In particular, notice that L is an upper bound on $\|\mathbf{x}_t\|_2$, not the total number of items (as in the rest of the paper).

In words, $\tilde{\eta}_{i,H}$ is the i-th of the noises $\{\eta_s^k\}_{(s,k)\in[t-1]\times[K]}$ under the lexicographical ordering.[12] Analogously, $\tilde{\phi}_{i,H}$ is the i-th element of $\{\phi(\mathbf{a}_k^s)\mathbb{1}(k \le \mathbf{C}_s)\}_{(s,k)\in[t-1]\times[K]}$ under the same ordering. We define the i-th bad event with respect to this ordering by

$$\tilde{\mathcal{E}}_{i,H} = \left\{ \left\| \sum_{j=1}^{i} \tilde{\phi}_{j,H}\tilde{\eta}_{j,H} \right\|^2_{\tilde{\mathbf{\Lambda}}_{i,H}^{-1}} > 2\log\left(\frac{\sqrt{\det(\tilde{\mathbf{\Lambda}}_{i,H})/\det(\tilde{\mathbf{\Lambda}}_{0,H})}}{1/n} \right) \right\}.$$

Now to prove the first bound in (35), we will show $\mathcal{E}_{t,H} \subset \cup_{i=0}^{(t-1)K}\tilde{\mathcal{E}}_{i,H}$ and $\mathbb{P}(\cup_{i=0}^{(t-1)K}\tilde{\mathcal{E}}_{i,H}) \le 1/n$. For the first result, we observe that for any $\tau \in [t]$,

$$\sum_{j=1}^{(\tau-1)K} \tilde{\phi}_{j,H}\tilde{\eta}_{j,H} = \sum_{s=1}^{\tau-1}\sum_{k=1}^{K} \phi(\mathbf{a}_k^s)\mathbb{1}(k \le \mathbf{C}_s)\eta_s^k = \sum_{s=1}^{\tau-1}\sum_{k=1}^{\min\{\mathbf{C}_t,K\}} \phi(\mathbf{a}_k^s)\eta_s^k,$$

and similarly, $\tilde{\mathbf{\Lambda}}_{(\tau-1)K} = \mathbf{\Lambda}_{\tau,H}$. Therefore, if $\mathcal{E}_{t,H}$ occurs, we can find $\tau \in [t]$ such that

$$\left\| \sum_{j=1}^{(\tau-1)K} \tilde{\phi}_{j,H}\tilde{\eta}_{j,H} \right\|_{\tilde{\mathbf{\Lambda}}_{(\tau-1)K,H}^{-1}} = \left\| \sum_{s=1}^{\tau-1}\sum_{k=1}^{\min\{\mathbf{C}_t,K\}} \phi(\mathbf{a}_k^s)\eta_s^k \right\|_{\mathbf{\Lambda}_{\tau,H}^{-1}} > \beta_{\tau,H}.$$

Furthermore, observe that by $\tilde{\mathbf{\Lambda}}_{(\tau-1)K} = \mathbf{\Lambda}_{\tau,H}$ and Claim 3, we have

$$\beta_{\tau,H}^2 = 2\log\left(\frac{\sqrt{(1+\tau K/d)^d}}{1/n} \right) \ge 2\log\left(\frac{\sqrt{\det(\tilde{\mathbf{\Lambda}}_{(\tau-1)K,H})/\det(\tilde{\mathbf{\Lambda}}_{0,H})}}{1/n} \right). \tag{36}$$

Combining the previous two inequalities, we obtain the first desired result:

$$\mathcal{E}_{t,H} \subset \cup_{\tau=1}^{t}\tilde{\mathcal{E}}_{(\tau-1)K,H} = \cup_{i\in\{0,K,\ldots,(t-1)K\}}\tilde{\mathcal{E}}_{i,H} \subset \cup_{i=0}^{(t-1)K}\tilde{\mathcal{E}}_{i,H}. \tag{37}$$

Finally, let $\mathscr{G}_{i,H}$ denote the σ-algebra generated by $\{\tilde{\phi}_{i',H}\}_{i'=1}^{i} \cup \{\tilde{\eta}_{i',H}\}_{i'=1}^{i-1}$. Note that each $\tilde{\phi}_{i,H}$ is $\mathscr{G}_{i,H}$-measurable and each $\tilde{\eta}_{i,H}$ is $[-1,1]$-valued and $\mathscr{G}_{i+1,H}$-measurable with $\mathbb{E}[\tilde{\eta}_{i,H}|\mathscr{G}_{i,H}] = 0$. Thus, applying Theorem 1 of Abbasi-Yadkori et al. [2011] with parameters $R = 1$ and $\delta = 1/n$ (in their notation), we conclude that $\mathbb{P}(\cup_{i=0}^{(t-1)K}\tilde{\mathcal{E}}_{i,H}) \le 1/n$, as desired.

For the second bound in (35), we first define the normalized versions of $\tilde{\phi}_{i,H}$, $\tilde{\eta}_{i,H}$, and $\tilde{\mathbf{\Lambda}}_{i,H}$:

$$\tilde{\phi}_{i,B} = \tilde{\phi}_{i,H}/\sqrt{\mathbf{U}_{\rho(i),H}(\mathbf{a}_{\sigma(i)}^{\rho(i)})},\ \tilde{\eta}_{i,B} = \tilde{\eta}_{i,H}/\sqrt{\mathbf{U}_{\rho(i),H}(\mathbf{a}_{\sigma(i)}^{\rho(i)})},\ \tilde{\mathbf{\Lambda}}_{i,B} = KI + \sum_{j=1}^{i}\tilde{\phi}_{j,B}\tilde{\phi}_{j,B}^{\mathsf{T}}.$$

Next, for any $i \le nK$, it is easily verified that $n^3K \ge 4i^2/(4K/n)$. Combined with the fact that $\rho(i)+1 \ge 1+(i-1)/K \ge i/K$, we can write

$$\beta_{\rho(i)+1,B} \ge \beta_{i/K,B} \ge 8\sqrt{d\log\left(1+\frac{i}{d}\right)\log\left(\frac{4i^2}{4K/n}\right)} + 4\sqrt{K}\log\left(\frac{4i^2}{4K/n}\right) \triangleq \beta'_{i,B}. \tag{38}$$

Using (38) instead of (36) but otherwise following the approach leading to (37), we analogously obtain $\mathcal{E}_{t,B} \subset \cup_{i=0}^{(t-1)K}\tilde{\mathcal{E}}_{i,B}$, where here we define

$$\tilde{\mathcal{E}}_{i,B} = \left\{ \left\| \sum_{j=1}^{i} \tilde{\phi}_{j,B}\tilde{\eta}_{j,B} \right\|_{\tilde{\mathbf{\Lambda}}_{i,B}^{-1}} > \beta'_{i,B} \right\}.$$

Again analogous to the above, we let $\mathscr{G}_{i,B}$ be the σ-algebra generated by $\{\tilde{\phi}_{i',B}\}_{i'=1}^{i} \cup \{\tilde{\eta}_{i',B}\}_{i'=1}^{i-1}$. Then each $\tilde{\phi}_{i,B}$ is $\mathscr{G}_{i,B}$-measurable and $\mathscr{B}_d(\sqrt{K})$-valued, while each $\tilde{\eta}_{i,B}$ is $[-\sqrt{K},\sqrt{K}]$-valued and $\mathscr{G}_{i+1,B}$-measurable with $\mathbb{E}[\tilde{\eta}_{i,B}|\mathscr{G}_{i,B}] = 0$. Furthermore, on $\bar{\mathcal{E}}_{t,H}$, Claim 2 ensures

$$\mathbb{E}[\tilde{\eta}_{i,B}^2|\mathscr{G}_{i,B}] = \frac{\mathbb{E}[(\eta_{\rho(i)}^{\sigma(i)})^2|\mathscr{G}_{i,B}]}{\mathbf{U}_{\rho(i),H}(\mathbf{a}_{\sigma(i)}^{\rho(i)})} = \frac{\bar{w}(\mathbf{a}_{\sigma(i)}^{\rho(i)})(1-\bar{w}(\mathbf{a}_{\sigma(i)}^{\rho(i)}))}{\mathbf{U}_{\rho(i),H}(\mathbf{a}_{\sigma(i)}^{\rho(i)})} \le \frac{\bar{w}(\mathbf{a}_{\sigma(i)}^{\rho(i)})}{\mathbf{U}_{\rho(i),H}(\mathbf{a}_{\sigma(i)}^{\rho(i)})} \le 1. \tag{39}$$

[12] The ordering $\prec_{\text{lex}}$ over $[t-1]\times[K]$ that satisfies $(s_1,k_1) \prec_{\text{lex}} (s_2,k_2)$ if $s_1 < s_2$ or $s_1 = s_2$ and $k_1 < k_2$.

Taken together, we have shown

$$\mathcal{E}_{t,B} \cap \bar{\mathcal{E}}_{t,H} \subset \cup_{i=0}^{(t-1)K} \tilde{\mathcal{E}}_{i,B} \cap \cap_{i=0}^{(t-1)K} \{\mathbb{E}[\tilde{\eta}_{i,B}^2 | \mathscr{G}_{i,B}] \leq 1\}.$$

Applying Claim 4 with $L = R = \sqrt{K}$, $\lambda = K$, and $\delta = 4K/n$ (recall the definition of $\beta'_{i,B}$ in (38)), we obtain that the probability of the event at right is at most $4K/n$, as desired. □

Lemma 6. *Under Assumption 3,* $R_3(n) \leq 2dK \log_2(1 + nK/d)$.

Proof. Let $\mathcal{T}_H = \{t \in [n] : \det(\mathbf{\Lambda}_{t+1,H}) \geq 2\det(\mathbf{\Lambda}_{t,H})\}$ and $\mathcal{T}_B = \{t \in [n] : \det(\mathbf{\Lambda}_{t+1,B}) \geq 2\det(\mathbf{\Lambda}_{t,B})\}$. Note that $R_2(n) = K\mathbb{E}[|\mathcal{T}_H \cup \mathcal{T}_B|] \leq K\mathbb{E}[|\mathcal{T}_H| + |\mathcal{T}_B|]$. Furthermore, observe

$$2^{|\mathcal{T}_H|} \leq \prod_{t \in \mathcal{T}_H} \frac{\det(\mathbf{\Lambda}_{t+1,H})}{\det(\mathbf{\Lambda}_{t,H})} \leq \prod_{t=1}^{n} \frac{\det(\mathbf{\Lambda}_{t+1,H})}{\det(\mathbf{\Lambda}_{t,H})} = \frac{\det(\mathbf{\Lambda}_{n+1,H})}{\det(\mathbf{\Lambda}_{1,H})} \leq \left(1 + \frac{nK}{d}\right)^d,$$

where we used Claim 3. Therefore, $|\mathcal{T}_H| \leq d\log_2(1 + nK/d)$. The analogous argument gives the same bound for $|\mathcal{T}_B|$, so this completes the proof. □

H.1 Proofs of Claims 2 and 3

To prove the claims, we begin with two general results. Throughout, we use the notation $A \succ B$ and $A \succeq B$ when $A - B$ is positive definite and positive semidefinite, respectively.

Claim 5. *Let* $\lambda > 0$, $\{x_i\}_{i=1}^m \subset \mathscr{B}_d(\sqrt{\lambda})$, *and* $B = \lambda I + \sum_{i=1}^m x_i x_i^\mathsf{T}$. *Then (i)* $B \succ 0$, *(ii)* $\|y\|_{B^{-1}} \leq \|y\|_2/\sqrt{\lambda}\ \forall\, y \in \mathbb{R}^d$, *(iii)* $\mathrm{Tr}(B) \leq \lambda(d+m)$, *(iv)* $\det(B) \leq (\lambda(1 + m/d))^d$.

Proof. Let $\{\lambda_j\}_{j=1}^d$ and $\{v_j\}_{j=1}^d$ be the eigenvalues and eigenvectors of B. Then for any $j \in [d]$,

$$\lambda_j = \lambda_j \|v_j\|_2^2 = v_j^\mathsf{T}(\lambda_j v_j) = v_j^\mathsf{T}(Bv_j) = v_j^\mathsf{T}\left(\lambda v_j + \sum_{i=1}^m x_i x_i^\mathsf{T} v_j\right) = \lambda + \sum_{i=1}^m (x_i^\mathsf{T} v_j)^2. \tag{40}$$

Therefore, $\min_{j\in[d]} \lambda_j \geq \lambda > 0$, which implies (i) and ensures that B is invertible. Next, let V be the orthogonal matrix with columns $\{v_j\}_{j=1}^d$. Then by unitary invariance, we know that

$$\sum_{j=1}^d (v_j^\mathsf{T} y)^2 = \sum_{j=1}^d (V^\mathsf{T} y)_j^2 = \|V^\mathsf{T} y\|_2^2 = \|y\|_2^2\ \forall\, y \in \mathbb{R}^d. \tag{41}$$

Combining (41) and the above bound $\min_{j\in[d]} \lambda_j \geq \lambda$ establishes (ii):

$$\|y\|_{B^{-1}}^2 = y^\mathsf{T} B^{-1} y = y^\mathsf{T}\left(\sum_{j=1}^d v_j v_j^\mathsf{T}/\lambda_j\right) y = \sum_{j=1}^d (v_j^\mathsf{T} y)^2/\lambda_j \leq \sum_{j=1}^d (v_j^\mathsf{T} y)^2/\lambda = \|y\|_2^2/\lambda.$$

Similarly, using (40), (41), and $\max_{i\in[m]} \|x_i\|_2 \leq \sqrt{\lambda}$, we obtain (iii):

$$\mathrm{Tr}(B) = \sum_{j=1}^d \lambda_j \leq \lambda d + \sum_{i=1}^m \sum_{j=1}^d (x_i^\mathsf{T} v_j)^2 \leq \lambda d + m \max_{i\in[m]} \|x_i\|_2^2 \leq \lambda(d+m)$$

Finally, (iii) and the inequality of the arithmetic and geometric means together yield (iv):

$$\det(B)^{1/d} = \left(\prod_{j=1}^d \lambda_j\right)^{1/d} \leq \sum_{j=1}^d \lambda_j/d = \mathrm{Tr}(B)/d \leq \lambda(1 + m/d).$$ □

Claim 6. *Let* $\lambda > 0$, $\{x_i\}_{i=1}^m \subset \mathscr{B}_d(\sqrt{\lambda})$, *and* $B_j = \lambda I + \sum_{i=1}^{j-1} x_i x_i^\mathsf{T}$ *for each* $j \in [m+1]$. *Then (i)* $\sum_{j=1}^m \|x_j\|_{B_j^{-1}}^2 \leq 2\log(\det(B_{m+1})/\det(B_1))$ *and (ii) for any* $y \in \mathbb{R}^d$ *and* $j_1, j_2 \in [m]$ *such that* $j_1 \geq j_2$, *we have* $\|y\|_{B_{j_1}^{-1}} \leq \|y\|_{B_{j_2}^{-1}} \leq \sqrt{\det(B_{j_1})/\det(B_{j_2})}\|y\|_{B_{j_1}^{-1}}$.

Proof. By part (ii) of Claim 5 and $x_j \in \mathscr{B}_d(\sqrt{\lambda})$, we know $\|x_j\|_{B_j^{-1}} \le 1$ for any $j \in [m]$. Hence, by Lemma 11 of Abbasi-Yadkori et al. [2011], we obtain (i):

$$\sum_{j=1}^{m} \|x_j\|^2_{B_j^{-1}} = \sum_{j=1}^{m} \min\{1, \|x_j\|^2_{B_j^{-1}}\} \le 2\log(\det(B_{m+1})/\det(B_1)).$$

For (ii), note that $y^\mathsf{T}(B_{j_1} - B_{j_2})y = \sum_{i=j_2}^{j_1-1}(x_i^\mathsf{T} y)^2 \ge 0$, so $B_{j_1} \succeq B_{j_2}$. This implies $B_{j_2}^{-1} \succeq B_{j_1}^{-1}$ and yields the bound $\|y\|_{B_{j_1}^{-1}} \le \|y\|_{B_{j_2}^{-1}}$. For the other bound in (ii), let $\Xi = \det(B_{j_1})/\det(B_{j_2})$. Then Lemma 12 of Abbasi-Yadkori et al. [2011] implies $y^\mathsf{T} B_{j_1} y \le \Xi y^\mathsf{T} B_{j_2} y$, so $B_{j_2} \succeq B_{j_1}/\Xi$. Therefore, $\Xi B_{j_1}^{-1} \succeq B_{j_2}^{-1}$, so $\|y\|_{B_{j_2}^{-1}} \le \|y\|_{\Xi B_{j_1}^{-1}} = \sqrt{\Xi}\|y\|_{B_{j_1}^{-1}}$. □

Proof of Claim 2. Fix $t \in [n]$, $\tau \in [t]$, and $e \in [L]$, and assume $\bar{\mathcal{E}}_{t,H}$ holds. Then

$$\theta = \mathbf{\Lambda}_{\tau,H}^{-1}\left((\mathbf{\Lambda}_{\tau,H} - I)\theta + \theta\right) = \mathbf{\Lambda}_{\tau,H}^{-1}\left(\sum_{s=1}^{\tau-1}\sum_{k=1}^{\min\{\mathbf{C}_s,K\}} \phi(\mathbf{a}_k^s)\bar{w}(\mathbf{a}_k^s) + \theta\right),$$

where we used the definition of $\mathbf{\Lambda}_{\tau,H}$ and Assumption 3. Therefore, we can write

$$\|\hat{\theta}_{\tau,H} - \theta\|_{\mathbf{\Lambda}_{\tau,H}} = \left\|\sum_{s=1}^{\tau-1}\sum_{k=1}^{\min\{\mathbf{C}_s,K\}} \phi(\mathbf{a}_k^s)\eta_k^s - \theta\right\|_{\mathbf{\Lambda}_{\tau,H}^{-1}} \le \beta_{\tau,H} + \|\theta\|_{\mathbf{\Lambda}_{\tau,H}^{-1}} \le \beta_{\tau,H} + \|\theta\|_2 \le \alpha_{\tau,H},$$

where the inequalities hold by the triangle inequality and the definition of $\bar{\mathcal{E}}_{t,H}$; part (ii) of Claim 5, which applies with parameter $\lambda = 1$ by definition of $\mathbf{\Lambda}_{\tau,H}$ and Assumption 3; and Assumption 3 and the definition of $\beta_{\tau,H}$; respectively. By this inequality and Cauchy-Schwarz, we obtain

$$|\langle\hat{\theta}_{\tau,H}, \phi(e)\rangle - \bar{w}(e)| = |\langle\hat{\theta}_{\tau,H} - \theta, \phi(e)\rangle| \le \|\hat{\theta}_{\tau,H} - \theta\|_{\mathbf{\Lambda}_{\tau,H}}\|\phi(e)\|_{\mathbf{\Lambda}_{\tau,H}^{-1}} \le \alpha_{\tau,H}\|\phi(e)\|_{\mathbf{\Lambda}_{\tau,H}^{-1}}.$$

As an immediate consequence, we see that

$$\mathbf{U}_{\tau,H}(e) = \max\{\langle\hat{\theta}_{\tau,H}, \phi(e)\rangle + \alpha_{\tau,H}\|\phi(e)\|_{\mathbf{\Lambda}_{\tau,H}^{-1}}, 1/K\} \ge \max\{\bar{w}(e), 1/K\} \ge \bar{w}(e). \tag{42}$$

The remaining bounds follow similarly, except we must invoke Claim 5 with $\lambda = K$ and use the fact that $\|\phi_t(e)\|_2 = \|\phi(e)\|_2/\sqrt{\mathbf{U}_{t,H}(e)} \le \sqrt{K}$ by definition and Assumption 3. □

Proof of Claim 3. Fix $t \in [n]$ and let $\mathbf{\Lambda}_{s,H}^k = \mathbf{\Lambda}_{s,H} + \sum_{i=1}^{k-1}\phi(\mathbf{a}_i^s)\phi(\mathbf{a}_i^s)^\mathsf{T}$ for each $s \in [t]$ and $k \in [\min\{\mathbf{C}_s, K\}]$. Then on the event $\bar{\mathcal{F}}_s$, we can apply part (ii) of Claim 6 twice (with parameter $\lambda = 1$ by definition of $\mathbf{\Lambda}_{s,H}^k$ and Assumption 3) to obtain that for any $y \in \mathbb{R}^d$,

$$\|y\|^2_{\mathbf{\Lambda}_{s,H}^{-1}} \le 2\|y\|^2_{\mathbf{\Lambda}_{s+1,H}^{-1}} \le 2\|y\|^2_{(\mathbf{\Lambda}_{s,H}^k)^{-1}}.$$

Using this bound, nonnegativity, and part (i) of Claim 6, we get

$$\sum_{s=1}^{t}\mathbb{1}(\bar{\mathcal{F}}_s)\sum_{k=1}^{\min\{\mathbf{C}_s,K\}}\|\phi(\mathbf{a}_k^s)\|^2_{\mathbf{\Lambda}_{s,H}^{-1}} \le 2\sum_{s=1}^{t}\sum_{k=1}^{\min\{\mathbf{C}_s,K\}}\|\phi(\mathbf{a}_k^s)\|^2_{(\mathbf{\Lambda}_{s,H}^k)^{-1}} \le 4\log\left(\frac{\det(\mathbf{\Lambda}_{t+1,H})}{\det(\mathbf{\Lambda}_{1,H})}\right).$$

Observing $\det(\mathbf{\Lambda}_{1,H}) = 1$ by definition and $\det(\mathbf{\Lambda}_{t+1,H}) \le (1 + tK/d)^d$ by part (iv) of Claim 5 (again, with $\lambda = 1$) completes the proof of the first bound. For the second, one proceeds similarly except uses $\lambda = K$ when invoking Claims 5 and 6 and $\det(\mathbf{\Lambda}_{1,B}) = K^d$ by definition. □

H.2 Proof of Claim 4

As discussed before the statement of the claim, the proof follows Zhou et al. [2021] with some minor changes that are necessary because we do not assume an almost-sure bound on $\mathbb{E}[\eta_t^2|\mathcal{G}_t]$. For brevity, we only explain the changes, and to unify the presentation, we adopt their notation (which is slightly inconsistent with the rest of Appendix H) for Appendix H.2 only. First, we replace their Lemma 11 (a version of Freedman's inequality that assumes the almost-sure bound) with the following.

Lemma 7. *Fix $M, v, \delta > 0$. Suppose $\{\mathcal{G}_t\}_{t=1}^{\infty}$ is a filtration and $\{x_t\}_{t=1}^{\infty}$ is a stochastic process such that, for each $t \in \mathbb{N}$, x_t is $[-M, M]$-valued and $\mathcal{G}_t$-measurable with $\mathbb{E}[x_t|\mathcal{G}_{t-1}] = 0$. Let $\Gamma = \sqrt{2v\log(1/\delta)} + 2M\log(1/\delta)/3$. Then*

$$\mathbb{P}\left(\cup_{t=1}^{\infty}\left(\left\{\sum_{i=1}^{t} x_i > \Gamma\right\} \cap \left\{\sum_{i=1}^{t} \mathbb{E}[x_i^2|\mathcal{G}_i] \le v\right\}\right)\right) \le \delta. \tag{43}$$

This result follows from the standard form of Freedman's inequality [Freedman, 1975], which says the left side of (43) is at most $\exp(-(\Gamma^2/2)/(v + M\Gamma/3)$, along with the lower bound

$$\Gamma^2/2 = \sqrt{v\log(1/\delta)/2}\Gamma + M\log(1/\delta)\Gamma/3 \ge v\log(1/\delta) + M\log(1/\delta)\Gamma/3.$$

Next, in the proof of their Lemma 13, their inequality (29) need not hold almost-surely, but it does hold on $\cap_{t=1}^{T}\{\mathbb{E}[\eta_t^2|\mathcal{G}_t] \le 1\}$ (with constant $\sigma^2 = 1$). Therefore, using our Lemma 7 instead of their Lemma 11 but otherwise following the proof of their Lemma 13, one can show

$$\mathbb{P}\left(\cup_{t=1}^{T}\left\{\sum_{i=1}^{t} \frac{\eta_i \mathbf{x}_i^\top \mathbf{Z}_{i-1}\mathbf{d}_{i-1}}{1+w_i^2}\mathcal{E}_{i-1} > \frac{3\beta_t^2}{8}\right\} \cap \cap_{t=1}^{T}\{\mathbb{E}[\eta_t^2|\mathcal{G}_t] \le 1\}\right) \le \frac{\delta}{2}, \tag{44}$$

where $\mathbf{d}_t = \sum_{i=1}^{t}\mathbf{x}_i\eta_i$, $w_t = \|\mathbf{x}_t\|_{\mathbf{Z}_{t-1}^{-1}}$, and $\mathcal{E}_t = \mathbb{1}(\|\mathbf{d}_s\|_{\mathbf{Z}_{s-1}^{-1}} \le \beta_s \ \forall\, s \in [t])$. By analogously modifying the proof of their Lemma 14, one can also show

$$\mathbb{P}\left(\cup_{t=1}^{T}\left\{\sum_{i=1}^{t} \frac{\eta_i^2 w_i^2}{1+w_i^2} > \frac{\beta_t^2}{4}\right\} \cap \cap_{t=1}^{T}\{\mathbb{E}[\eta_t^2|\mathcal{G}_t] \le 1\}\right) \le \frac{\delta}{2}. \tag{45}$$

Finally, the inductive argument from the end of Appendix B.1 in Zhou et al. [2021] implies

$$\cup_{t=1}^{T}\left\{\left\|\sum_{i=1}^{t}\mathbf{x}_i\eta_i\right\|_{\mathbf{Z}_t^{-1}} > \beta_t\right\} \subset \cup_{t=1}^{T}\left\{\sum_{i=1}^{t} \frac{\eta_i \mathbf{x}_i^\top \mathbf{Z}_{i-1}\mathbf{d}_{i-1}}{1+w_i^2}\mathcal{E}_i > \frac{3\beta_t^2}{8}\right\} \cup \left\{\sum_{i=1}^{t} \frac{\eta_i^2 w_i^2}{1+w_i^2} > \frac{\beta_t^2}{4}\right\},$$

so taking the intersection with $\cap_{t=1}^{T}\{\mathbb{E}[\eta_t^2|\mathcal{G}_t] \le 1\}$ and using (44) and (45) completes the proof.

H.3 Proof sketch for Algorithm 3

Compared to Algorithm 2, the proof for Algorithm 3 requires three changes. The first change is that in Claim 2, we replace the inequality $\mathbf{U}_{\tau,H}(e) \ge \bar{w}(e)$ with $\min\{\mathbf{U}_{\tau,H}(e), \tilde{\mathbf{U}}_{\tau,H}(e)\} \ge \bar{w}(e)$. To prove this stronger form of the claim, we first use an argument like (42) from the proof of the claim to conclude $\tilde{\mathbf{U}}_{\tau,H}(e) \ge \bar{w}(e)$; the other inequality follows, since

$$\mathbf{U}_{\tau,H}(e) = \min\{\max\{\tilde{\mathbf{U}}_{\tau,H}(e), 1/K\}, 1\} \ge \min\{\tilde{\mathbf{U}}_{\tau,H}(e), 1\} \ge \min\{\bar{w}(e), 1\} = \bar{w}(e).$$

The second change is in the proof of Lemma 4. The inequality (30) therein holds by the same logic since $\mathbf{U}_{t,B}(e)$ is defined identically in the two algorithms. Next, (31) continues to hold because it is an upper bound on $\mathbf{U}_{t,H}(e)$, whose definition in Algorithm 3 was only changed to include a minimum with 1. Thus, (32), which is a consequence (30)-(31), continues to hold as well; moreover, since $\mathbf{U}_t(e) \le \mathbf{U}_{t,B}(e)$ by definition in Algorithm 3, the upper bound in (32) holds for $\mathbf{U}_t(e)$ as well. Next, we observe that by the stronger form of Claim 2 (the one just discussed) and $\bar{w}(e^*) \le 1$ that $\mathbf{U}_t(e^*) \ge \bar{w}(e^*)$ on $\bar{\mathcal{E}}_t$. Therefore, the (33) still holds, with $\mathbf{U}_{t,B}$ replaced by $\mathbf{U}_t$. The remainder of the proof does not rely on the form of the UCB, so it goes through identically. Finally, the third change is that in the proof of Lemma 5, we again invoke the stronger form of Claim 2, in particular, to establish (39). The remainder of the Lemma 5 proof is identical.

I Proof of Theorem 1

Set $p = 1/(2K)$ and $\Delta = \sqrt{L/(4nK^2)}$. Note that by assumption $n \ge L$, we have $\Delta \in [0, p]$. Let $\mathscr{W}_{p,\Delta}$ denote the set of vectors $\bar{w} \in [0, p]^L$ in which $\bar{w}(e) = (p+\Delta)/2$ for exactly K (optimal) items e and $\bar{w}(e) = p/2$ for the remaining (suboptimal) items. Define Π' and $R'_{\pi,\bar{w}}$ as in Section 5. Then

using $\mathscr{W}_{p,\Delta} \subset [0,1]^L$, Theorem 1 of Kveton et al. [2015a] (see the proof of Theorem 4 therein for a similar application), Bernoulli's inequality, and $\Pi \subset \Pi'$, respectively, we can write

$$\begin{aligned}\inf_{\pi\in\Pi}\sup_{\bar{w}\in[0,1]^L} R_{\pi,\bar{w}}(n) &\geq \inf_{\pi\in\Pi}\sup_{\bar{w}\in\mathscr{W}_{p,\Delta}} R_{\pi,\bar{w}}(n) \geq (1-p)^{K-1}\inf_{\pi\in\Pi}\sup_{\bar{w}\in\mathscr{W}_{p,\Delta}} R'_{\pi,\bar{w}}(n)\\ &\geq \inf_{\pi\in\Pi}\sup_{\bar{w}\in\mathscr{W}_{p,\Delta}} R'_{\pi,\bar{w}}(n)/2 \geq \inf_{\pi\in\Pi'}\sup_{\bar{w}\in\mathscr{W}_{p,\Delta}} R'_{\pi,\bar{w}}(n)/2.\end{aligned}$$

Now fix $\pi \in \Pi'$. By the previous inequality, it suffices to show $R'_{\pi,\bar{w}}(n) = \Omega(\sqrt{nL})$ for some $\bar{w} \in \mathscr{W}_{p,\Delta}$. For this, we essentially follow Lattimore et al. [2018], but there are some changes that tricky to explain (because our notation – which follows Kveton et al. [2015a] – differs drastically, and because we noticed some small typos in their analysis). For clarity, we therefore provide a complete proof, but we emphasize that it is essentially restated from Lattimore et al. [2018].

First, we introduce some notation. Recall $N = L/K \in \mathbb{N}$, and for each $m \in [N]^K$, define

$$\begin{gathered}\mathcal{L}_i = \{(i-1)N+1,\dots,iN\}, \quad a_i^*(m) = (i-1)N + m(i),\\ \mathcal{L}_i(m) = \mathcal{L}_i \setminus a_i^*(m)\ \forall\, i\in[K], \quad \mathcal{L}(m) = \cup_{i=1}^K \mathcal{L}_i(m),\\ A^*(m) = \{a_i^*(m)\}_{i=1}^K = [L]\setminus\mathcal{L}(m),\\ \bar{w}_m(e) = \frac{p+\Delta\mathbb{1}(e\in A^*(m))}{2}\ \forall\, e\in[L].\end{gathered}$$

In words, we partitioned $[L]$ into groups $\mathcal{L}_i$. Each vector $m \in [N]^K$ defines an instance $\bar{w}_m$ where one item $a_i^*(m)$ in each group i is $(\Delta/2)$-better than the rest of the group $\mathcal{L}_i(m)$, and where $A^*(m)$ is the optimal action ($\mathcal{L}(m)$ are the suboptimal items). More precisely, note each such instance has K items with $\bar{w}(e) = (p+\Delta)/2$ and $\bar{w}(e) = p/2$ for the other items. Therefore, $\{\bar{w}_m\}_{m\in[N]^K} \subset \mathscr{W}_{p,\Delta}$, so it suffices to show $R'_m(n) \triangleq R'_{\pi,\bar{w}_m}(n) = \Omega(\sqrt{nL})$ for some $m \in [N]^K$.

Toward this end, for any $e \in [L]$, $t \in [n]$, and $\mathcal{L} \subset [L]$, we let $\mathbf{T}'_t(e) = \sum_{s=1}^t \mathbb{1}(e \in \mathbf{A}_s)$ denote the number of times e was played up to and including time t and $\mathbf{T}'_t(\mathcal{L}) = \sum_{e\in\mathcal{L}} \mathbf{T}'_t(e)$. For $m \in [N]^K$, we define $\mathbb{E}_m = \mathbb{E}_{\pi,\bar{w}_m}$ to be expectation under instance $\bar{w}_m$. Then by definition (9), we have

$$R'_m(n) = \frac{\Delta}{2}\mathbb{E}_m\left[\sum_{t=1}^n\sum_{k=1}^K \mathbb{1}(\mathbf{a}_k^t \in \mathcal{L}(m))\right] = \frac{\Delta}{2}\mathbb{E}_m[\mathbf{T}'_n(\mathcal{L}(m))]. \tag{46}$$

Next, as observed by Lattimore et al. [2018], we can rewrite $\mathbf{T}'_n(\mathcal{L}(m))$ in two useful ways:

$$\mathbf{T}'_n(\mathcal{L}(m)) = nK - \mathbf{T}'_n(A^*(m)) = \sum_{i=1}^K (n - \mathbf{T}'_n(a_i^*(m))), \quad \mathbf{T}'_n(\mathcal{L}(m)) = \sum_{i=1}^K \mathbf{T}'_n(\mathcal{L}_i(m)).$$

Therefore, adding the two expressions together, we obtain

$$\mathbf{T}'_n(\mathcal{L}(m)) = \frac{2\mathbf{T}'_n(\mathcal{L}(m))}{2} = \frac{\sum_{i=1}^K (n - \mathbf{T}'_n(a_i^*(m)) + \mathbf{T}'_n(\mathcal{L}_i(m)))}{2}. \tag{47}$$

For each $i \in [K]$, let $\tau_i = \min\{n, \min\{t \in [n] : \mathbf{T}'_t(\mathcal{L}_i) \geq n\}$. Then if $\tau_i = n$, we have

$$n - \mathbf{T}'_n(a_i^*(m)) + \mathbf{T}'_n(\mathcal{L}_i(m)) \geq n - \mathbf{T}'_n(a_i^*(m)) = n - \mathbf{T}'_{\tau_i}(a_i^*(m)),$$

while if $\tau_i < n$, we use $\mathbf{T}'_n(a_i^*(m)) \leq n$, monotocity of $\mathbf{T}'_\cdot$, and the definition of τ_i to write

$$\begin{aligned}n - \mathbf{T}'_n(a_i^*(m)) + \mathbf{T}'_n(\mathcal{L}_i(m)) &\geq \mathbf{T}'_n(\mathcal{L}_i(m)) \geq \mathbf{T}'_{\tau_i}(\mathcal{L}_i(m))\\ &= \mathbf{T}'_{\tau_i}(\mathcal{L}_i) - \mathbf{T}'_{\tau_i}(a_i^*(m)) \geq n - \mathbf{T}'_{\tau_i}(a_i^*(m)).\end{aligned}$$

Therefore, in both cases, we have the lower bound $n - \mathbf{T}'_{\tau_i}(a_i^*(m))$. Combined with (46) and (47), and lower bounding a maximum by an average, we can therefore write

$$\max_{m\in[N]^K} R'_m(n) \geq \frac{\Delta}{4}\left(nK - \frac{1}{N^K}\sum_{i=1}^K\sum_{m\in[N]^K}\mathbb{E}_m[\mathbf{T}'_{\tau_i}(a_i^*(m))]\right). \tag{48}$$

Fix $i \in [K]$, and for any $m \in [N]^{K-1}$ and $z \in [N]$, let $g_i(m,z) = [\cdots\ m(i-1)\ z\ m(i+1)\ \cdots] \in [N]^K$. Note that $\{g_i(m,z) : m \in [N]^{K-1}, z \in [N]\} = [N]^K$. Also, recalling $a_i^*(\tilde{m}) = (i-1)N + \tilde{m}(i)$ for any $\tilde{m} \in [N]^K$ by definition, we see that $a_i^*(g_i(m,z)) = (i-1)N + z$. Thus, we obtain

$$\sum_{m \in [N]^K} \mathbb{E}_m[\mathbf{T}'_{\tau_i}(a_i^*(m))] = \sum_{m \in [N]^{K-1}} \sum_{z \in [N]} \mathbb{E}_{g_i(m,z)}[\mathbf{T}'_{\tau_i}((i-1)N + z)]. \tag{49}$$

Now fix $m \in [N]^{K-1}$ and let $\bar{w}_{m'}$ be the instance with

$$A^*(m') = \{(j-1)N + m(j)\}_{j=1}^{i-1} \cup \{jN + m(j)\}_{j=i}^{K-1}, \quad \bar{w}_{m'}(e) = \frac{p + \Delta \mathbb{1}(e \in A^*(m'))}{2}.$$

Observe that, for any $z \in [N]$, $\bar{w}_{m'}$ is the same as $\bar{w}_{g_i(m,z)}$ except $(i-1)N + z$ is only optimal in the former instance (since $A^*(m') \subset \{1, \ldots, (i-1)N, iN+1, \ldots, L\}$). Denote expectation in the former instance by $\mathbb{E}_{m'}$. Then for any $z \in [N]$, following Lattimore et al. [2018], we can use Pinsker's inequality and the chain rule for KL-divergence (see also Section 3.3 of Bubeck and Cesa-Bianchi [2012]) to obtain

$$\begin{aligned}\mathbb{E}_{g_i(m,z)}[\mathbf{T}'_{\tau_i}((i-1)N + z)] &\le \mathbb{E}_{m'}[\mathbf{T}'_{\tau_i}((i-1)N + z)] \\ &\quad + n\sqrt{\mathbb{E}_{m'}[\mathbf{T}'_{\tau_i}((i-1)N + z)] d(p/2, (p+\Delta)/2)/2}.\end{aligned}$$

By Claim 7 below, $\Delta \le p$, and $p \le 1/2$, we have $d(p/2, (p+\Delta)/2) \le \Delta^2/p$. Combining with the previous inequality, then summing over $z \in [N]$ and using Cauchy-Schwarz, we obtain

$$\sum_{z \in [N]} \mathbb{E}_{g_i(m,z)}[\mathbf{T}'_{\tau_i}((i-1)N + z)] \le \mathbb{E}_{m'}[\mathbf{T}'_{\tau_i}(\mathcal{L}_i)] + n\sqrt{N\mathbb{E}_{m'}[\mathbf{T}'_{\tau_i}(\mathcal{L}_i)]\Delta^2/(2p)}. \tag{50}$$

Now by definition of τ_i, we know that $\mathbf{T}'_{\tau_i}(\mathcal{L}_i) = \mathbf{T}'_{\tau_i - 1}(\mathcal{L}_i) + (\mathbf{T}'_{\tau_i}(\mathcal{L}_i) - \mathbf{T}'_{\tau_i - 1}(\mathcal{L}_i)) \le n + K$. Therefore, we can further upper bound the right side of (50) by

$$n + K + n\sqrt{\frac{N(n+K)\Delta^2}{2p}} = n + K + \frac{nN}{2}\sqrt{\frac{n+K}{n}} \le nN\left(\frac{1}{N} + \frac{1}{N^2} + \frac{\sqrt{1+1/N}}{2}\right) < \frac{9nN}{10},$$

where the equality holds by definitions $p = 1/(2K)$, $\Delta = \sqrt{L/(4nK^2)}$, and $N = L/K$; the first inequality by definition $N = L/K$ and assumption $n \ge L$ (so $K = L/N \le n/N = nN/N^2$); and the second by $N \ge 4$. Combining with (48), (49), (50), and the definition of Δ gives

$$\max_{m \in [N]^K} R'_m(n) \ge \frac{\Delta}{4}\left(nK - \frac{1}{N^K} \cdot K \cdot N^{K-1} \cdot \frac{9nN}{10}\right) = \frac{\Delta nK}{40} = \frac{\sqrt{nL}}{80}.$$

Claim 7. *For any $p, q \in (0,1)$, $d(p,q) \le (p-q)^2/(q(1-q))$.*

Proof. Since $(p-q)/q, (q-p)/(1-q) \ge -1$, we can use $\log(1+x) \le x\ \forall\, x \ge -1$ to write

$$d(p,q) \triangleq p\log\left(1 + \frac{p-q}{q}\right) + (1-p)\log\left(1 + \frac{q-p}{1-q}\right) \le p\left(\frac{p-q}{q}\right) + (1-p)\left(\frac{q-p}{1-q}\right).$$

Finally, a simple calculation shows that the right side equals the desired upper bound. □

J Proof of Theorem 3

Let $\chi \ge 4$ be an absolute constant to be chosen later. Set $p = 1/(2K)$, $\Delta = \sqrt{L/(\chi nK)}$, $\bar{w}(e^*) = p\ \forall\, e^* \in [K]$, and $\bar{w}(e) = p - \Delta\ \forall\, e \in [L] \setminus [K]$. Note $p \in [\Delta, 1/2]$ by the assumption $n \ge LK$ and the choice $\chi \ge 4$, so $\bar{w} \in [0,1]^L$. Let π denote `CascadeUCB1` and $R(n) = R_{\pi,\bar{w}}(n)$. For each $t \in [n]$ and $\mathcal{L} \subset [L]$, define $\mathbf{T}_t(\mathcal{L}) = \sum_{e \in \mathcal{L}} \mathbf{T}_t(e)$. Then similar to the beginning of Appendix I, we can use Theorem 1 of Kveton et al. [2015a] and Bernoulli's inequality to write

$$R(n) \ge \Delta\mathbb{E}[\mathbf{T}_n([L] \setminus [K])]/2 = \sqrt{L/(4\chi nK)}\mathbb{E}[\mathbf{T}_n([L] \setminus [K])].$$

Now define the events

$$\mathcal{E} = \{\mathbf{T}_n([L] \setminus [K]) < nK/4\}, \quad \mathcal{F} = \{\mathbf{T}_n([L]) < nK/2\},$$

$$\mathcal{G} = \{\mathbf{T}_n([K]) > nK/4\}, \quad \mathcal{G}_{e^*} = \{\mathbf{T}_n(e^*) > n/4\} \; \forall \, e^* \in [K].$$

Note that $\mathbf{T}_n([L] \setminus [K]) \geq nK/4$ on the event $\bar{\mathcal{E}}$. Substituting into the previous inequality gives

$$R(n) \geq \sqrt{L/(4\chi nK)}\mathbb{E}[\mathbf{T}_n([L] \setminus [K])\mathbb{1}(\bar{\mathcal{E}})] \geq \sqrt{nLK/(64\chi)}(1 - \mathbb{P}(\mathcal{E})). \tag{51}$$

To upper bound $\mathbb{P}(\mathcal{E})$, first notice that when $\mathcal{E}$ and $\bar{\mathcal{F}}$ both occur, we have

$$\mathbf{T}_n([K]) = \mathbf{T}_n([L]) - \mathbf{T}_n([L] \setminus [K]) > (nK/2) - (nK/4) = nK/4,$$

so $\mathcal{E} \cap \bar{\mathcal{F}} \subset \mathcal{G}$. Also observe $\mathcal{G} \subset \cup_{e^*=1}^K \mathcal{G}_{e^*}$. Combining and applying the union bound,

$$\mathbb{P}(\mathcal{E}) = \mathbb{P}(\mathcal{E} \cap \mathcal{F}) + \mathbb{P}(\mathcal{E} \cap \bar{\mathcal{F}}) \leq \mathbb{P}(\mathcal{F}) + \mathbb{P}(\mathcal{E} \cap \mathcal{G}) \leq \mathbb{P}(\mathcal{F}) + \sum_{e^*=1}^K \mathbb{P}(\mathcal{E} \cap \mathcal{G}_{e^*}). \tag{52}$$

For the first term, let $\mathbf{O}_t = \mathbf{T}_t([L]) - \mathbf{T}_{t-1}([L])$ denote the number of observations at t. Then

$$\mathbb{E}_t[\mathbf{O}_t] = \sum_{k=1}^K \prod_{i=1}^{k-1}(1 - \bar{w}(\mathbf{a}_k^t)) \geq \sum_{k=1}^K (1-p)^{k-1} = \frac{1-(1-p)^K}{p} \geq \frac{1 - \exp(-pK)}{p},$$

so by definition $p = 1/(2K)$, we have $\mathbb{E}_t[\mathbf{O}_t] \geq 2(1 - \exp(-1/2))K > 3K/4$. Thus, we obtain

$$\mathbb{E}[\mathbf{T}_n([L])] = \mathbb{E}\left[\sum_{t=1}^n (\mathbf{T}_t([L]) - \mathbf{T}_{t-1}([L]))\right] = \sum_{t=1}^n \mathbb{E}[\mathbb{E}_t[\mathbf{O}_t]] \geq 3nK/4.$$

Hence, because $\mathbf{T}_n([L])$ is $[nK]$-valued, we can use Markov's inequality to write

$$\mathbb{P}(\mathcal{F}) = \mathbb{P}(nK - \mathbf{T}_n([L]) > nK/2) \leq \frac{\mathbb{E}[nK - \mathbf{T}_n([L])]}{nK/2} \leq \frac{nK - 3nK/4}{nK/2} = \frac{1}{2}. \tag{53}$$

Now fix $e^* \in [K]$ and suppose $\mathcal{G}_{e^*}$ occurs. Then $\mathbf{T}_n(e^*) \geq \lceil n/4 \rceil$, in which case we can find $t \geq \lceil n/4 \rceil$ such that $\mathbf{T}_{t-1}(e^*) = \lceil n/4 \rceil - 1$ and $\mathbf{U}_t(e^*)$ is among the K highest UCBs. Mathematically, we can express the latter event as $|\{e \in [L] : \mathbf{U}_t(e) \leq \mathbf{U}_t(e^*)\}| \geq L - K$, which implies

$$\begin{aligned} 3L/4 \leq L - 2K &\leq |\{e \in [L] : \mathbf{U}_t(e) \leq \mathbf{U}_t(e^*)\}| - K \\ &\leq |\{e \in [L] \setminus [K] : \mathbf{U}_t(e) \leq \mathbf{U}_t(e^*)\}| \triangleq \mathcal{L}_{e^*}, \end{aligned}$$

where we also used $L \geq 8K$ by assumption. On the other hand, when $\mathcal{E}$ occurs, there must exist $e \in \mathcal{L}_{e^*}$ such that $\mathbf{T}_{t-1}(e) < nK/(4|\mathcal{L}_{e^*}|)$, because otherwise, we can write $\mathbf{T}_n([L] \setminus [K]) \geq \mathbf{T}_{t-1}(\mathcal{L}_{e^*}) \geq nK/4$, which contradicts $\mathcal{E}$. Combining these observations, we see that when both $\mathcal{G}_{e^*}$ and $\mathcal{E}$ occur, we can find $t \geq \lceil n/4 \rceil$ and $e \in \mathcal{L}_{e^*} \subset [L] \setminus [K]$ such that

$$\mathbf{T}_{t-1}(e^*) = \lceil n/4 \rceil - 1, \quad \mathbf{U}_t(e^*) \geq \mathbf{U}_t(e), \quad \mathbf{T}_{t-1}(e) \in [\lfloor nK/(4|\mathcal{L}_{e^*}|) \rfloor] \subset [\lfloor nK/(3L) \rfloor].$$

Therefore, taking three union bounds, we obtain

$$\mathbb{P}(\mathcal{E} \cap \mathcal{G}_{e^*}) \leq \sum_{e=K+1}^L \sum_{t=\lceil \frac{n}{4} \rceil}^n \sum_{s=1}^{\lfloor \frac{nK}{3L} \rfloor} \mathbb{P}(\mathbf{T}_{t-1}(e^*) = \lceil n/4 \rceil - 1, \mathbf{U}_t(e^*) \geq \mathbf{U}_t(e), \mathbf{T}_{t-1}(e) = s). \tag{54}$$

Now fix e, t, and s as in the triple summation. Then by definition of $\mathbf{U}_t$ (4), we know that

$$\begin{aligned} &\mathbf{T}_{t-1}(e^*) = \lceil n/4 \rceil - 1, \quad \mathbf{U}_t(e^*) \geq \mathbf{U}_t(e), \quad \mathbf{T}_{t-1}(e) = s \\ \Rightarrow \quad &\hat{\mathbf{w}}_{\lceil \frac{n}{4} \rceil - 1}(e^*) + c_{t, \lceil \frac{n}{4} \rceil - 1} \geq \hat{\mathbf{w}}_s(e) + c_{t,s}. \end{aligned}$$

Note that if the event at bottom occurs, one of the following inequalities must hold:

$$\hat{\mathbf{w}}_{\lceil \frac{n}{4} \rceil - 1}(e^*) \geq p + c_{t, \lceil \frac{n}{4} \rceil - 1}, \; \hat{\mathbf{w}}_s(e) \leq p - \Delta - \sqrt{5/6}c_{t,s}, \; \Delta \geq (1 - \sqrt{5/6})c_{t,s} - 2c_{t, \lceil \frac{n}{4} \rceil - 1}. \tag{55}$$

Indeed, if none of these inequalities hold, then we obtain a contradiction:

$$\hat{\mathbf{w}}_{\lceil \frac{n}{4} \rceil - 1}(e^*) + c_{t, \lceil \frac{n}{4} \rceil - 1} < p + 2c_{t, \lceil \frac{n}{4} \rceil - 1} < \hat{\mathbf{w}}_s(e) + \Delta + 2c_{t, \lceil \frac{n}{4} \rceil - 1} + \sqrt{5/6}c_{t,s} < \hat{\mathbf{w}}_s(e) + c_{t,s}.$$

However, Claim 8 below shows the third inequality in (55) cannot occur, provided we choose χ large. Furthermore, by the union and Hoeffding bounds and the definition $c_{t,s} = \sqrt{(3/2)\log(t)/s}$, the probability that either of the first two inequalities holds is at most $t^{-2(3/2)} + t^{-2(3/2)(5/6)} = t^{-3} + t^{-5/2}$. Plugging into (54) and using Claim 9 below and the assumption $n \geq 49K^4$ gives

$$\mathbb{P}(\mathcal{E} \cap \mathcal{G}_{e^*}) \leq \frac{nK}{3} \cdot \sum_{t=\lceil n/4 \rceil}^{\infty} (t^{-3} + t^{-5/2}) \leq \frac{nK}{3} \cdot 10n^{-3/2} = \frac{10K}{3\sqrt{n}} \leq \frac{10}{21K}. \tag{56}$$

Combining with (52) and (53) shows $\mathbb{P}(\mathcal{E}) \leq 41/42$; substituting into (51) gives the desired bound.

Remark 9. *If we replace the above constant $5/6$ with $1-\delta/3$ for fixed $\delta > 0$ (we chose $\delta = 1/2$ above), the third inequality in (55) again fails for appropriate χ, and the Hoeffding bound gives $t^{-2(3/2)(1-\delta/3)} = t^{\delta-3}$ for the probability of the first two. Following the argument above then shows $\mathbb{P}(\mathcal{E} \cap \mathcal{G}_{e^*}) = O(Kn^{\delta-1})$, which is $O(1/K)$ as in (56) when $n = \Omega(K^{2/(1-\delta)})$. Thus, the assumption $n = \Omega(K^4)$ can be weakened to $n = \Omega(K^\eta)$ for fixed $\eta > 2$.*

Claim 8. *Under the assumptions of Theorem 3, there exists an absolute constant $\chi \geq 4$ such that, for any $t \geq \lceil n/4 \rceil$ and $s \leq nK/(3L)$, we have $(1-\sqrt{5/6})c_{t,s} - 2c_{t,\lceil \frac{n}{4} \rceil - 1} > \sqrt{L/(\chi nK)}$.*

Proof. By assumption $L \geq 800K$ and $n \geq LK$, we know $n \geq 800K^2 \geq 800$. The fact that $n \geq 100$ ensures $\lceil n/4 \rceil - 1 \geq (n/4) - 1 \geq 6n/25$; combined with $L \geq 800K$, we obtain

$$2c_{t,\lceil \frac{n}{4} \rceil - 1} = \sqrt{\frac{6\log t}{\lceil n/4 \rceil - 1}} \leq \sqrt{\frac{25\log t}{n}} = \sqrt{\frac{800\log t}{32n}} \leq \sqrt{\frac{L\log t}{32nK}}.$$

For $s \leq nK/(3L)$, we also know $c_{t,s} = \sqrt{3\log(t)/(2s)} \geq \sqrt{9L\log(t)/(2nK)}$, so we conclude

$$(1-\sqrt{5/6})c_{t,s} - 2c_{t,\lceil \frac{n}{4} \rceil - 1} \geq \sqrt{\frac{L\log t}{nK}}\left((1-\sqrt{5/6})\sqrt{9/2} - \sqrt{1/32}\right).$$

The term in parentheses on the right side is positive. Therefore, using the above bound $n \geq 800$ and the choice $t \geq n/4$, the result follows for $\chi > 1/(\log(200)((1-\sqrt{5/6})\sqrt{9/2} - \sqrt{1/32})^2) \geq 4$. $\square$

Claim 9. *Under the assumptions of Theorem 3, $\sum_{t=\lceil \frac{n}{4} \rceil}^{\infty}(t^{-3} + t^{-5/2}) \leq 10n^{-3/2}$.*

Proof. Let $T = \lceil n/4 \rceil - 1$. As in Claim 8, we know $T \geq 6n/25 > 0$. Thus, for any $c > 1$,

$$\sum_{t=T+1}^{\infty} t^{-c} = \sum_{t=T+1}^{\infty} \int_{x=t-1}^{t} t^{-c}dx \leq \sum_{t=T+1}^{\infty} \int_{x=t-1}^{t} x^{-c}dx = \int_{x=T}^{\infty} x^{-c}dx = \frac{T^{1-c}}{c-1}.$$

Using this inequality for $c \in \{3, 5/2\}$ and the bound $T \geq 6n/25$, we thus obtain

$$\sum_{t=T+1}^{\infty}(t^{-3} + t^{-5/2}) \leq \frac{T^{-2}}{2} + \frac{T^{-3/2}}{3/2} \leq \frac{7}{6}T^{-3/2} \leq \frac{7}{6}\left(\frac{25}{6}\right)^{3/2} n^{-3/2} \leq 10n^{-3/2}. \qquad \square$$