True Impact of Cascade Length in Contextual Cascading Bandits

Hyun-jun Choi
Seoul National University
nschj1@snu.ac.kr

Joongkyu Lee
Seoul National University
jklee0717@snu.ac.kr

Min-hwan Oh
Seoul National University
minoh@snu.ac.kr

Abstract

We revisit the contextual cascading bandit, where a learning agent recommends an ordered list (*cascade*) of items, and a user scans the list sequentially, stopping at the first attractive item. Although cascading bandits underpin various applications including recommender systems and search engines, the role of the cascade length K in shaping regret has remained unclear. Contrary to prior results that regret grows with K, we prove that regret actually *decreases* once K is large enough. Leveraging this insight, we design a new upper-confidence-bound algorithm built on online mirror descent that attains the sharpest known regret upper bound, $\tilde{\mathcal{O}}\big(\min\{K\bar{p}^{K-1}, 1\}d\sqrt{T}\big)$ for contextual cascading bandits. To complement this new regret upper bound, we provide a nearly matching lower bound of $\Omega\big(\min\{K\underline{p}^{K-1}, 1\}d\sqrt{T}\big)$, where $0 \le \underline{p} \le \bar{p} < 1$. Together, these results fully characterize how regret truly scales with K, thereby closing the theoretical gap for contextual cascading bandits. Finally, comprehensive experiments validate our theoretical results and show the effectiveness of our proposed method.

1 Introduction

Cascading bandits have broad applications in online recommender systems, search engines, and social media. In this model, at each round, the agent selects an ordered list (a *cascade*) of K items from a ground set of N items. The user examines each item in order and decides whether to click it or skip to the next. The round stops at the first click or when all K items have been examined without any click. Therefore, the agent receives the partial click feedback only for the observed items in the given cascade. Cascading bandits have been extensively studied from the non-contextual multi-armed formulation [11, 12, 26] to contextual variants incorporating item (and user) features [19, 18, 28, 7, 26, 20, 21], providing a widely used framework for modeling sequential user interactions with multiple items.

In contextual cascading bandits, the effect of the cascade length K on the regret bound remains theoretically unclear. While previous studies suggest that the regret either scales polynomially or logarithmically with K (see Table 1), this contradicts the intuition that a longer cascade provides more opportunities to collect feedback and may therefore reduce regret. The gap between theory and intuition regarding how the cascade length affects the regret has been recognized [28, 7, 20, 21], and recent studies [7, 21] have narrowed this gap, though it has yet to be fully resolved.

Li et al. [19] first introduced the contextual cascading bandits and derived a regret upper bound that scales as $\tilde{O}(\sqrt{K})$ by exploiting the Lipschitz continuity of the expected reward. Recent studies have further refined the dependence of regret on the cascade length by analyzing different click feedback models. Liu et al. [20] considered the case where the click feedback follows a linear model, whereas Choi et al. [7] and Liu et al. [21] focused on the logistic model. These three works investigated a common structural property that the gradient of the expected reward function can be expressed as

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

Table 1: Comparisons of algorithms for cascading bandit. N is the number of ground arms, K is a cascade length, d is a dimension of context vectors and T is total rounds. Here, $0 \leq \underline{p} \leq \bar{p} < 1$ (See Definition 5.2) and $\kappa \in (0, 1/4]$ (See Assumption 5.3).

Algorithm / Paper	Model	Bound	Dependence on K
`CascadeKL-UCB` [11]	(Non-contextual)	$\mathcal{O}\left((N-K)\frac{\Delta(1+\log(1/\Delta))}{D_{KL}(p-\Delta\|\|p)}\log T\right)^*$	Decreasing
Lower Bound [11]	(Non-contextual)	$\Omega\left((N-K)\frac{\Delta}{D_{KL}(p-\Delta\|\|p)}\log T\right)^*$	Decreasing
`C`^{3}`-UCB` [19]	Linear	$\mathcal{O}(d\sqrt{KT}\log T)$	Increasing
`LinTS-Cascade` [28]	Linear	$\mathcal{O}(d^{3/2}K\sqrt{T}\log T)$	Increasing
`CascadeWOFUL` [26]	Linear	$\mathcal{O}(\sqrt{d^2T + dTK}\log(KT))$	Increasing
`VAC`^{2}`-UCB` [20]	Linear	$\mathcal{O}(d\sqrt{T}\log(KT))$	Increasing
`UCB-CCA` [7]	Logistic	$\mathcal{O}(\frac{1}{\kappa}d\sqrt{T}\log(KT))$	Increasing
`UCB-CCA+` [7]	Logistic	$\mathcal{O}(d\sqrt{T}\log(KT))$	Increasing
`CLogUCB` [21]	Logistic	$\mathcal{O}(d\sqrt{\frac{1}{\kappa}KT}\log(KT))$	Increasing
`VA-CLogUCB` [21]	Logistic	$\mathcal{O}(d\sqrt{KT}\log(KT))$	Increasing
`EVA-CLogUCB` [21]	Logistic	$\mathcal{O}(d\sqrt{T}\log(KT))$	Increasing
`UCB-CLB` (**this work**, Theorem 5.6)	Logistic**	$\mathcal{O}\left(\min\left\{K\bar{p}^{K-1}, 1\right\}d\sqrt{T}\log(KT)\right)$	Decreasing for large K
Lower Bound (**this work**, Theorem 5.7)	Logistic	$\Omega\left(\min\left\{K\underline{p}^{K-1}, 1\right\}d\sqrt{T}\right)$	Decreasing for large K

* These non-contextual results are gap-dependent regret bounds under a symmetric instance, where each optimal item has a click probability p and each suboptimal item has $p - \Delta$ for some $\Delta \in (0, p)$. $D_{KL}(p\|q)$ denotes the Kullback–Leibler (KL) divergence between two Bernoulli distributions with means p and q.
** Although we present our main results for the logistic model—reflecting the binary "click" feedback observed in practice—the analysis also carries over to the linear model, hence our results are comparable to existing linear model results, as well as logistic model results.

a product of non-click probabilities. Specifically, Choi et al. [7] combined this property with the optimistic exposure swapping technique, which places the most uncertain item first in the cascade, while Liu et al. [20] and Liu et al. [21] leveraged the triggering probability equivalence technique, which links the probability that an item is observed to the random event of that item being observed. These approaches yielded tighter regret bounds of order $\mathcal{O}(d\sqrt{T}\log(KT))$, demonstrating that the regret no longer grows polynomially with the cascade length but still grows logarithmically.

On the other hand, the *non-contextual* cascading bandit literature [11] provides a more explicit characterization of how the cascade length affects regret. Kveton et al. [11] established gap-dependent upper and lower bounds, explicitly scaling with the difference between total number of arms and cascade length $(N-K)$. Under a symmetric instance, where each optimal item has a click probability p and each suboptimal item has $p - \Delta$ for some $\Delta \in (0, p)$, $\mathcal{O}\big((N-K)\frac{\Delta(1+\log(1/\Delta))}{D_{KL}(p-\Delta\|p)}\log T\big)$ regret upper bound is shown, along with a matching lower bound (up to logarithmic factors of Δ) of $\Omega\big((N-K)\frac{\Delta}{D_{KL}(p-\Delta\|p)}\log T\big)$. These results indicate that the regret decreases as K increases (i.e., as K approaches N). Yet, a direct translation of these bounds to contextual cascading bandits is unclear, since regret bounds typically exhibit no explicit dependence on N in contextual settings.

In this paper, we address this long-standing open question in the contextual cascading bandits:

What is the true impact of the cascade length on the regret bound?

We show that the regret bound of contextual cascading bandits shrinks to zero for sufficiently large cascade length. Our analysis begins with the contextual *logistic* cascading bandits where the binary click feedback follows a logistic model. Within this framework, we propose a new UCB algorithm with an online mirror descent method effectively exploiting the cascading structure and integrating the optimistic exposure swapping technique proposed by Choi et al. [7]. To establish its theoretical guarantee, we derive the tightest known regret upper bound for contextual cascading bandits, $\tilde{\mathcal{O}}(\min\{K\bar{p}^{K-1}, 1\}\, d\sqrt{T})$, overcoming the technical challenges inherent to the cascade structure. Unlike previous results [7, 21] that scale as $\tilde{\mathcal{O}}(d\sqrt{T})$, our bound introduces the multiplicative factor

$K\bar{p}^{K-1}$, revealing that the regret can decrease with larger cascade length. This result is further supported by a matching problem-dependent regret lower bound, which confirms the correct K-scalability of our upper bound. To the best of our knowledge, this is the first lower bound analysis for contextual cascading bandits. Finally, the proposed analysis is directly applicable to the contextual cascading *linear* bandits, where the regret bound $\tilde{\mathcal{O}}(\min\{K\bar{p}^{K-1}, 1\}\, d\sqrt{T})$ remains valid, demonstrating that the $K\bar{p}^{K-1}$ term captures an intrinsic property of the cascade structure rather than a peculiarity of the feedback model.

Our main contributions are summarized as follows.

- We propose a UCB algorithm for cascading logistic bandits and establish the T-step regret upper bound of $\tilde{\mathcal{O}}(\min\left\{K\bar{p}^{K-1}, 1\right\} d\sqrt{T})$ where $0 \leq \bar{p} < 1$ (Theorem 5.6).
- To our best knowledge, this regret upper bound is the *tightest* bound among all the existing regret bounds for contextual cascading bandits. In contrast to previous studies that suggest the bound increases with K, our finding demonstrates that the regret bound decreases with sufficiently large K.
- By leveraging online mirror descent for parameter estimation in cascading logistic bandits, our algorithm achieves constant per-round computational and storage costs, independent of T, thereby ensuring computational efficiency.
- We also derive an N-independent lower bound on the regret in cascading logistic bandits as $\Omega(\min\left\{Kp^{K-1}, 1\right\} d\sqrt{T})$ (in Theorem 5.7), where $0 \leq \underline{p} \leq \bar{p} < 1$. To the best of our knowledge, this is the first derivation of a lower bound in contextual cascading bandits.
- We show that the regret bound derived in the logistic setting also holds in the linear model under mild assumptions, highlighting the generality of our theoretical results.

2 Inefficient Dependence on Cascade Length

In contextual cascading bandits, the true impact of cascade length on the regret bound still remains unresolved to this day. As summarized in Table 1, all existing regret upper bounds for contextual cascading bandits [19, 28, 26, 20, 7, 21] are either $\tilde{\mathcal{O}}(Kd\sqrt{T})$ or $\tilde{\mathcal{O}}(d\sqrt{T})$[1]. However, simple experiments cast doubt on the implication of K dependence in these existing results. We conduct experiments to observe how cumulative regret evolves with varying cascade length K. This experiment is based on the *MovieLens 100K* dataset[2], and we defer the experimental details to Section 6. We gradually increase K, with all other conditions held constant across bandit instances. As shown in Figure 1, our result presents a counterexample: cumulative regret decreases as K increases for all existing methods. Hence, the previous theoretical claim that regret either worsens or remains unaffected by increasing K is not supported by experimental results.

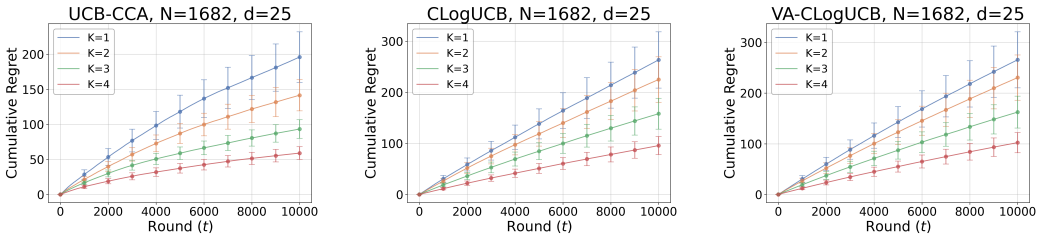

Figure 1: Cumulative regret over time for `UCB-CCA` [7], `CLogUCB` [21] and `VA-CLogUCB` [21] with $N = 1682, d = 25$, and varying cascade lengths $K \in \{1, 2, 3, 4\}$. In all methods, increasing the cascade length leads to a clear reduction in cumulative regret?" .

Furthermore, the claim that regret grows with K is also counterintuitive. By definition, regret is the difference between the expected reward of the optimal cascade and that of the cascade selected by the

[1] Still, these results retain a logarithmic dependency on K.

[2] Available at `https://grouplens.org/datasets/movielens/100k/`.

agent. In cascading bandits, the expected reward of a cascade is generally defined as the probability that at least one arm in the cascade yields positive feedback[3]. When all other conditions are kept constant, increasing the cascade length provides the agent with more opportunities to receive positive feedback. This leads the expected reward of the selected cascade to increase. For the same reason, the expected reward of the optimal cascade also increases as the cascade length increases.

Our intuition and experiments suggest that prior claims about how cascade length affects regret are incomplete and possibly loose. Surprisingly, explicit discussion of this issue is scarce in the literature. We therefore establish a rigorous, refined regret bound that reveals how cascade length fundamentally influences regret in cascading bandits.

3 Preliminaries

3.1 Notation

Define $[n]$ as the set of positive integers from 1 to n. Let $|\cdot|$ be the length of a sequence or the cardinality of a set. For a vector $x \in \mathbb{R}^d$, we denote the ℓ_2-norm of x as $\|x\|$ and the V-weighted norm of x for a positive-definite matrix V as $\|x\|_V = \sqrt{x^\top V x}$. The determinant and trace of a matrix V are $\det(V)$ and $\mathrm{trace}(V)$, respectively. $\lambda_{\min}(V)$ denotes the minimum eigenvalue of a matrix V. $I_d \in \mathbb{R}^{d\times d}$ is an identity matrix.

3.2 Problem Setting

We begin by outlining the core structure of the contextual cascading bandit, followed by an extension that parameterizes the Bernoulli feedback using a logistic function.

The contextual cascading bandit framework models the interaction between a learning agent and the environment as follows: given a set $\mathcal{I}$ that consists of total N arms, the agent offers a list of K distinct arms to the environment, which we refer to as a *cascade*. The set of all possible cascades is denoted by $\Pi := \{(i_1, \ldots, i_K) : i_1, \ldots, i_K \in \mathcal{I}, i_k \neq i_m \text{ for any } k \neq m\}$. At each round t, the environment reveals a contextual feature vector $x_{t,i} \in \mathbb{R}^d$ for each arm $i \in \mathcal{I}$, collectively denoted by $X_t := \{x_{t,i}\}_{i\in\mathcal{I}}$. Based on the history $\mathcal{H}_t$ and the revealed feature vectors X_t, the agent selects a cascade $C_t = (i_{t1}, i_{t2}, \ldots, i_{tK}) \in \Pi$, where i_{tk} is the k-th arm in C_t. For any arm $i_{tk} \in C_t$, we define $y_{t,i_{tk}} \in \{0, 1\}$ as the binary feedback provided by the environment to the agent, where 1 indicates positive feedback (i.e., click). Starting from the first arm i_{t1}, the agent sequentially observes feedback for each arm in C_t. Upon receiving feedback $y_{t,i_{tk}} = 1$, the round immediately terminates, halting further observations for remaining arms $i_{t,k+1}, \ldots, i_{tK}$. Let $k_t := \min\left(\{m \in [K] : y_{t,i_{tm}} = 1\} \cup \{K + 1\}\right)$, which denotes the position of the first clicked item in round t, or $K + 1$ if no click occurs. We define the list of observed arms in round t as $O_t := (i_{tk} : 1 \leq k \leq \min(k_t, K))$.

The reward of agent who chooses cascade C_t in round t is defined as 1 if at least one arm in C_t yields feedback of 1; otherwise, the reward is 0. Formally, this can be expressed as:

$$r_t(C_t) = \max_{i\in C_t} y_{t,i} = \bigvee_{i\in C_t} y_{t,i} = 1 - \prod_{i\in C_t} (1 - y_{t,i}).$$

This type of reward structure is referred to as a disjunctive model [12, 19], in which the agent receives a reward if any arm within the cascade provides positive feedback. This model is well suited to recommender systems, where success is achieved if at least one recommended item satisfies the user.

Building on the framework described above, we study the *cascading logistic bandit*, which integrates a logistic parametric model to account for the Bernoulli feedback. Since the logistic model is better suited to handle binary feedback than the linear model, it has gained attention in recent cascading bandit literature [7, 21]. Specifically, given the history $\mathcal{H}_t := (X_\tau, C_\tau, O_\tau, Y_\tau)_{\tau<t} \cup (X_t, C_t)$, the feedback $y_{t,i}$ for all $i \in \mathcal{I}$ are modeled as mutually independent Bernoulli random variables. Let $\sigma_1(z) = \exp(z)/(1 + \exp(z))$ and $\sigma_0(z) = 1 - \sigma_1(z)$. The conditional expectation of $y_{t,i}$ is parameterized using a logistic function $\mathbb{E}\left[y_{t,i}|\mathcal{H}_t\right] = \sigma_1(x_{t,i}^\top \theta^*)$, where $\theta^* \in \mathbb{R}^d$ is an *unknown*

[3]This is commonly referred to as cascading bandits with a disjunctive objective.

time-invariant parameter. The expected reward of C_t is then given by:

$$\mathbb{E}[r_t(C_t) \mid \mathcal{H}_t] = 1 - \prod_{i \in C_t} (1 - \sigma_1(x_{t,i}^\top \theta^*)) =: f_t(C_t, \theta^*). \tag{1}$$

The optimal action $C_t^* \in \operatorname{argmax}_{C \in \Pi} f_t(C, \theta^*)$ in round t is defined as the cascade that maximizes the expected reward. The objective is to maximize the cumulative expected reward over T rounds by efficiently learning the unknown parameter θ^*. To evaluate the performance of an online learning bandit algorithm, the regret, defined as the gap between the expected cumulative reward of the optimal cascade C_t^* and that achieved by the algorithm's selection C_t over T rounds, is used. The formal definition of regret is as follows:

$$\mathcal{R}(T) := \mathbb{E}\left[\sum_{t=1}^{T} r_t(C_t^*) - r_t(C_t)\right] = \mathbb{E}\left[\sum_{t=1}^{T} f_t(C_t^*, \theta^*) - f_t(C_t, \theta^*)\right], \tag{2}$$

where the last equality is from the law of total expectation and the definition of the expected reward.

4 Algorithm

In this section, we introduce our algorithm, **U**pper **C**onfidence **B**ound for **C**ascading **L**ogistic **B**andits (`UCB-CLB`), leveraging the widely used UCB technique [4, 1, 17] to find an optimistic action based on a well-established confidence set. The pseudocode of `UCB-CLB` is presented in Algorithm 1. `UCB-CLB` consists of three key components.

First, it constructs a confidence set using parameter estimates obtained via online mirror descent (OMD), as detailed in Section 4.1. The use of OMD provides a computationally and memory-efficient alternative to maximum likelihood estimation (MLE), which has been adopted in logistic bandits [27, 15]. To the best of our knowledge, `UCB-CLB` is the first algorithm to incorporate OMD in the cascading bandit setting. Second, Section 4.2 explains how `UCB-CLB` efficiently selects a cascade by ensuring optimism through an exploration bonus. Rather than solving combinatorial optimization over $\binom{N}{K}$ arms to find the cascade that maximize the estimated rewards, `UCB-CLB` reduces this process to a simpler top-K arm selection procedure. Finally, Section 4.3 discusses the incorporation of *optimistic exposure swapping*, originally proposed by Choi et al. [7] adapted to our problem setting. This technique aims to mitigate an issue arising from unobserved feedback, by strategically placing arms with higher uncertainty in the cascade.

Our main design goal is to develop a cascading bandit algorithm whose regret decreases for sufficiently large cascade length K. Since our setting assumes a logistic model, we leverage insights from prior work [8, 2, 9] on contextual logistic bandits, where the leading term of the regret is known to be independent of a problem-dependent factor κ, formally defined in Assumption 5.3. To achieve both goals, our algorithm incorporates two key components: `DO-SWAP` for cascade restructuring and online parameter estimation via OMD.

4.1 Efficient Online Parameter Estimation

In the existing literature on logistic and multinomial logistic bandits [8, 2, 10, 3, 15], the maximum likelihood estimation (MLE) approach has been the common method for parameter estimation. To reduce computational and memory overhead, recent studies [9, 27, 13] have proposed online parameter estimation methods based on algorithms such as online Newton step and online mirror descent (OMD). Building on these advances [27, 13, 14], we adopt an OMD-based estimator tailored to the cascading bandit setting. This adaptation achieves constant computational complexity per round, providing substantial computational efficiency over MLE-based approaches.

The parameter θ is updated for each observation $i_{tk} \in O_t$. Let $\theta_{t,1} := \theta_t$ and $H_{t,1} := H_t$, where

$$H_t = \sum_{\tau=1}^{t-1} \sum_{k \in |O_\tau|} \dot{\sigma}_1(x_{\tau,i_{\tau k}}^\top \theta_{\tau,k+1}) x_{\tau,i_{\tau k}} x_{t,i_{\tau k}}^\top + \lambda I_d.$$

Then, for $k \in [|O_t|]$, we update the parameter $\theta_{t,k+1}$ using an online mirror descent [24] as follows:

$$\theta_{t,k+1} = \operatorname*{argmin}_{\theta \in \Theta} \left\{ \frac{1}{2\eta} \|\theta - \theta_{t,k}\|_{\tilde{H}_{t,k}}^2 + \langle \theta, \nabla \ell_{t,i_{tk}}(\theta_{t,k}) \rangle \right\}, \quad \Theta = \{\theta \in \mathbb{R}^d : \|\theta\|_2 \le 1\}, \tag{3}$$

Algorithm 1 UCB-CLB

Input: penalty λ, radius β_t, step size η.
Initialize $\theta_1 \in \Theta$, $H_1 = \lambda I_d$.
for $t = 1, \dots, T$ **do**
  Compute $\{u_{t,i} = x_{t,i}^\top \theta_t + \beta_t \|x_{t,i}\|_{H_t^{-1}}\}_{i \in \mathcal{I}}$.
  Select $C_t' \in \operatorname{argmax}_{C \in \Pi} \tilde{f}_t(C)$ by Eq.(5).
  $C_t \leftarrow$ DO-SWAP(C_t', θ_t, H_t).
  Play C_t and receive feedback tuple (O_t, Y_t).
  $\theta_{t,1} \leftarrow \theta_t$, $H_{t,1} \leftarrow H_t$
  for $k = 1, \dots, |O_t|$ **do**
    $\tilde{H}_{t,k} \leftarrow H_{t,k} + \eta \dot{\sigma}_1(x_{t,i_{tk}}^\top \theta_{t,k}) x_{t,i_{tk}} x_{t,i_{tk}}^\top$
    Update $\theta_{t,k+1}$ by Eq.(3)
    $H_{t,k+1} \leftarrow H_{t,k} + \dot{\sigma}_1(x_{t,i_{tk}}^\top \theta_{t,k+1}) x_{t,i_{tk}} x_{t,i_{tk}}^\top$.
  end for
  $\theta_{t+1} \leftarrow \theta_{t,|O_t|+1}$, $H_{t+1} \leftarrow H_{t,|O_t|+1}$
end for

Algorithm 2 DO-SWAP

Input: cascade C_t, parameter θ_t, gram matrix H_t.
Find $i_t^{(1)}$ and $i_t^{(2)}$ by Eq.(6).
Swap the positions of i_{t1} and $i_t^{(1)}$.
if $i_t^{(1)} \neq i_t^{(2)}$ **then**
  Swap the positions of i_{t2} and $i_t^{(2)}$.
end if
Output: swapped cascade $(i_t^{(1)}, i_t^{(2)}, \dots)$.

where $\tilde{H}_{t,k} = H_{t,k} + \eta \nabla \sigma(x_{t,i_{tk}}^\top \theta_{t,k}) x_{t,i_{tk}} x_{t,i_{tk}}^\top$. Here, for arm i in round t under parameter θ, the loss is defined as $\ell_{t,i}(\theta) := -\sum_{y \in \{0,1\}} \mathbb{I}\{y_{t,i} = y\} \log \sigma_y(x_{t,i}^\top \theta)$, and its gradient is given by $\nabla \ell_{t,i}(\theta) = \left(\sigma_1(x_{t,i}^\top \theta) - y_{t,i}\right) x_{t,i}$.

To efficiently solve the optimization problem in Eq.(3), we perform a single projected gradient step.

$$\theta'_{t,k+1} = \theta_{t,k} - \eta \tilde{H}_{t,k}^{-1} \nabla \ell_{t,i_{tk}}(\theta_{t,k}), \quad \theta_{t,k+1} \in \operatorname*{argmin}_{\theta \in \Theta} \left\|\theta - \theta'_{t,k+1}\right\|_{\tilde{H}_{t,k}}.$$

Next, we update the gram matrix as follows:

$$H_{t,k+1} = H_{t,k} + \dot{\sigma}_1(x_{t,i_{tk}}^\top \theta_{t,k+1}) x_{t,i_{tk}} x_{t,i_{tk}}^\top.$$

This process is repeated in every round, ensuring the parameters and gram matrices are consistently updated for all observed arms.

The optimization problem in Eq.(3) requires a computational cost of only $\mathcal{O}(Kd^3)$, which is completely independent of the round t. Since we update the parameter K times per round, the total computational cost is $\mathcal{O}(K^2 d^3)$. For storage costs, the estimator does not need to store all historical data because both $\tilde{H}_{t,k}$ and $H_{t,k}$ can be updated incrementally, requiring only $\mathcal{O}(d^2)$ storage.

4.2 Efficient Optimistic Expected Reward

We leverage the UCB technique to compute an optimistic action based on estimates of each arm. We compute our optimistic estimate $u_{t,i}$ for all $t \in [T]$ and $i \in \mathcal{I}$ as follows:

$$u_{t,i} = x_{t,i}^\top \theta_t + \beta_t \|x_{t,i}\|_{H_t^{-1}}. \tag{4}$$

where $\beta_t \geq 0$ represents a confidence radius, with its specific value being provided in Section 5.2 to ensure the necessary statistical guarantees. We define $\tilde{f}_t(C)$ to be the optimistic expected reward of the cascade C in round t based on $u_{t,i}$:

$$\tilde{f}_t(C) := 1 - \prod_{i \in C} \sigma_0(u_{t,i}). \tag{5}$$

Then, the agent identifies the cascade that maximizes $\tilde{f}_t$ in each round. As $\sigma_0(\cdot)$ is a monotonically decreasing function, maximizing $\tilde{f}t$ simplifies to selecting the top-K arms with the highest optimistic estimates $u_{t,i}$ from $i \in \mathcal{I}$.

Remark 1 (Comparison to Liu et al. [21]). We compare our method, UCB-CLB, with the UCB algorithms introduced by Liu et al. [21], namely VA-CLogUCB and EVA-CLogUCB. Similar to our approach, both algorithms adopt a bonus-based method to induce exploration. However, our algorithm offers key

advantages in terms of computational efficiency beyond statistical efficiency. Unlike VA-CLogUCB and EVA-CLogUCB, which rely on MLE, UCB-CLB leverages online parameter estimation, significantly reducing computational and storage costs. Furthermore, VA-CLogUCB requires a non-convex projection, which could be NP-hard, to ensure statistical guarantees. EVA-CLogUCB eliminates this non-convex optimization process by introducing a burn-in stage of $\mathcal{O}(\log T)$ rounds to construct a convex nonlinearity-restricted region. However, EVA-CLogUCB requires an additional assumption that feature vectors remain time-invariant. Thus, our UCB-CLB applies to more general settings.

4.3 Doubly Optimistic Exposure Swapping

In this section, we introduce the *doubly optimistic exposure swapping* (DO-SWAP), which is a technique that provides an alternative approach for handling unobserved feedback. We summarize the process of DO-SWAP in Algorithm 2. In cascading bandits, there exists a critical challenge that the learning agent cannot access feedback information for all arms in C_t, but only for the arms in O_t, i.e., the observed arms. Consequently, the Gram matrix H_t is constructed solely from the feature information of the observed arms—whose number may be strictly smaller than the cascade length—and is used to construct the confidence set. As a result, there is an issue where the leftover sum (from $|O_t|$ to $|C_t|$) becomes out of control [19]. To address this issue, we exploit the doubly optimistic exposure swapping technique first introduced in Choi et al. [7].

DO-SWAP follows the following process. After finding $C_t \in \operatorname{argmax}_{C\in\Pi} \tilde{f}_t(C)$, the agent swaps i_{t1} and i_{t2}, the first and second arms in C_t, with $i_t^{(1)}$ and $i_t^{(2)}$, respectively, where

$$i_t^{(1)} = \operatorname*{argmax}_{i\in C_t} \dot{\sigma}_1(x_{t,i}^\top \theta_t) \left\|x_{t,i}\right\|_{H_t^{-1}}, \quad i_t^{(2)} = \operatorname*{argmax}_{i\in C_t} \left\|x_{t,i}\right\|_{H_t^{-1}}. \tag{6}$$

As a side note, when $i_t^{(1)} = i_t^{(2)}$, it suffices for the agent to swap $i_t^{(1)}$ with the first positioned arm $i_{t,1} \in C_t$. Since the arm in the first position is always observed ($i_t^{(1)} \in O_t$), the following inequality holds and effectively avoids the aforementioned out-of-control issue:

$$\sum_{i\in C_t} \dot{\sigma}_1(x_{t,i}^\top \theta_t) \left\|x_{t,i}\right\|_{H_t^{-1}} = K \max_{i\in C_t} \dot{\sigma}_1(x_{t,i}^\top \theta_t) \left\|x_{t,i}\right\|_{H_t^{-1}} = K\dot{\sigma}_1(x_{t,i_t^{(1)}}^\top \theta_t) \left\|x_{t,i_t^{(1)}}\right\|_{H_t^{-1}}.$$

Additionally, considering that the arm with the largest $\left\|x_{t,i}\right\|_{H_t^{-1}}$ in C_t is placed in the second position, the following equation holds:

$$\sum_{i\in C_t} \left\|x_{t,i}\right\|_{H_t^{-1}}^2 = K \max_{i\in C_t} \left\|x_{t,i}\right\|_{H_t^{-1}}^2 = K \left\|x_{t,i_t^{(2)}}\right\|_{H_t^{-1}}^2.$$

A detailed discussion on how DO-SWAP affects the regret analysis is included in Appendix C.1.

5 Regret Analysis

5.1 Regularity Condition

Before presenting our main theoretical results, we first introduce the regularity assumptions.

Assumption 5.1. $\|x_{t,i}\| \le 1$ for all round t and arms $i \in \mathcal{I}$, and also $\|\theta^*\| \le 1$.

Assumption 5.1 ensures that the regret bound is independent of the scale of the feature vector and parameter, which is a standard assumption commonly adopted in the contextual bandit literature [1, 19, 22]. Under Assumption 5.1, we define the following constant related to the feedback probabilities:

Definition 5.2. Under Assumption 5.1, we define the following constants as:

$$\bar{p} := \sup_{x:\|x\|\le 1} \sigma_0(x^\top \theta^*), \quad \underline{p} := \inf_{x:\|x\|\le 1} \sigma_0(x^\top \theta^*).$$

These two constants play a critical role in expressing the true scaling of regret as the cascade length K becomes sufficiently large. (See Theorem 5.6 and Theorem 5.7.)

Assumption 5.3. There exists $\kappa > 0$ such that $\min_{\theta:\|\theta\|\le 1} \sigma_1(x_{t,i}^\top \theta)\sigma_0(x_{t,i}^\top \theta) \ge \kappa$, for every arm $i \in \mathcal{I}$ and all round t.

The problem-dependent factor κ typically appears in the combinatorial logistic bandit literature [6, 22, 23] and is adapted from standard link-function conditions in the generalized linear contextual bandit literature [17].. Note that a smaller κ indicates a larger deviation from the linear model.

5.2 Confidence Set

In this section, we aim to establish a theoretically grounded confidence set for the parameter estimate $\theta_{t,k}$, which plays a crucial role in constructing the upper confidence bound in our algorithm. Since we update the parameter $\theta_{t,k}$ and the hessian matrix $H_{t,k}$ for every t and $k \in [|O_t|]$, we can directly apply the online confidence bounds previously studied in the recent work [27, 13, 14].

Proposition 5.4 (Online parameter confidence set, Zhang and Sugiyama 27, Lee and Oh 13, 14)**.** *Let $\delta \in (0, 1]$, and denote the assortment size as M. Under Assumption 5.1, set the step size $\eta = \frac{1}{2}\log 2 + 2$ and penalty parameter $\lambda = 84\sqrt{2}d\eta$. For each update time $t \in [T]$, we define the following confidence set with $\beta_t(\delta) = \mathcal{O}(\sqrt{d}\log T)$:*

$$\mathcal{C}_t(\delta) := \{\theta \in \Theta : \|\theta_t - \theta\|_{H_t} \leq \beta_t(\delta)\},$$

Then, we have $\mathbb{P}(\forall t \geq 1, \theta^* \in \mathcal{C}_t(\delta)) \geq 1 - \delta$.

Applying this result to our problem setting, each assortment in our case contains a single item (since the agent receives binary feedback). Hence, our problem setting is a special case with $M = 1$. Up to round t, the algorithm performs one parameter update per observation, resulting in a total of $\sum_{\tau=1}^{t} |O_\tau|$ updates. Hence, we obtain the following result:

Corollary 5.5. *Under the same setting with Proposition 5.4, for all $t \in [T]$ and $k \in [|O_t|]$, with probability at least $1 - \delta$, we have $\theta^* \in \mathcal{C}^{\text{ON}}_{t,k}(\delta)$, where*

$$\mathcal{C}^{\text{ON}}_{t,k}(\delta) := \left\{\theta \in \Theta : \|\theta_{t,k} - \theta\|_{H_{t,k}} \leq \beta_{t,k}(\delta) = \mathcal{O}(\sqrt{d}\log(tK))\right\}.$$

5.3 Regret Upper Bound

Theorem 5.6 (Regret of UCB-CLB)**.** *Suppose $d \geq K$. Set the step size $\eta = \frac{1}{2}\log 2 + 2$ and penalty $\lambda = \max\{84\sqrt{2}d\eta, K\}$. Let $\delta \in (0, 1]$. With probability at least $1 - \delta$, and under Assumptions 5.1 and 5.3, UCB-CLB ensures*

$$\mathcal{R}(T) = \tilde{\mathcal{O}}\left(\min\left\{K\bar{p}^{K-1}d\sqrt{T} + K\bar{p}^{K-1}\frac{d^2}{\kappa \underline{p}} + K\bar{p}^{K-1}\frac{d^2}{\kappa}, d\sqrt{T} + \frac{d^2}{\kappa}\right\}\right).$$

Discussion of Theorem 5.6. This theorem highlights the interesting impact of cascade length K on regret, providing new insights. The dominant term of the regret upper bound in Theorem 5.6 scales as $\tilde{\mathcal{O}}\big(\min\{K\bar{p}^{K-1}, 1\}\, d\sqrt{T}\big)$, which exhibits a *plateau-then-decreasing* behavior with respect to the cascade length K. To interpret this behavior, consider the term $K\bar{p}^{K-1}$. Since K is a positive integer, $K\bar{p}^{K-1}$ is monotonically decreasing in K when $\bar{p} \leq e^{-1}$. For larger values of $\bar{p}$ (e.g., $e^{-1} < \bar{p} < 1$), this term first increases from $K = 1$ and then eventually decreases, achieving its maximum at $K = \frac{1}{\log(1/\bar{p})}$. However, the regret bound depends on the truncated quantity $\min\{K\bar{p}^{K-1}, 1\}$. Because of this truncation, the overall regret bound is non-increasing in K: it remains flat for small K and decreases once K becomes sufficiently large. This offers an intuitive understanding of how an increasing cascade length positively affects regret performance under such conditions. The proof of Theorem 5.6 is provided in Appendix C.1.

Remark 2 (Applicability to Linear Models). A regret bound of $\tilde{\mathcal{O}}(\min\{K\bar{p}^{K-1}, 1\}d\sqrt{T})$ can also be derived for the cascading *linear* bandits. The full derivation is provided in Appendix E. Existing cascading linear bandit algorithms [19, 18, 26, 20] can attain a regret bound of $\tilde{\mathcal{O}}(K\bar{p}^{K-1})$ when combined with a swapping technique similar to DO-SWAP. In particular, the arm maximizing the confidence width $\|x_{t,i}\|_{V_t^{-1}}$ where $V_t = \sum_{\tau=1}^{t-1}\sum_{i\in O_\tau} x_{\tau,i}x_{\tau,i}^\top + \lambda I$ is placed in the first position of the cascade. Deriving this bound in the linear setting requires an assumption that $\bar{p} < 1$; that is, the probability of receiving positive feedback is assumed to be less than 1. This is because in the linear model, the feedback probability is defined as $x_{t,i}^\top\theta^* \in [0, 1]$ in [19, 26]. The assumption of $\bar{p}$ being less than 1 is often considered mild in practice— that is, every arm considered as candidates has at least a (even very) small positive probability of being clicked.

5.4 Regret Lower Bound

Theorem 5.7 (Regret lower bound of contextual cascading bandits)**.** *Let d be divisible by 4, and suppose that Assumption 5.1 holds. Suppose $T \geq C \cdot d^4$ for some constant $C > 0$. Then, for any*

policy π, there exists a problem instance such that the worst-case expected regret of π is lower bounded as follows:

$$\sup_{\theta} \mathbb{E}_{\theta}^{\pi} \left[\mathcal{R}_{\theta}(T)\right] = \Omega\left(\min\left\{K\underline{p}^{K-1}, 1\right\} \cdot d\sqrt{T}\right).$$

The key observation is that the dominant term in our upper bound in Theorem 5.6 matches the lower bound up to logarithmic factors, lower-order terms and a gap of $\bar{p}$ and $\underline{p}$, indicating that the `UCB-CLB` achieves a near optimal regret upper bound. The proof of Theorem 5.7 is deferred to Appendix D.

Remark 3 (Comparison to Kveton et al. [11]). The lower bound for *non-contextual* cascading bandits established by Kveton et al. [11] in Theorem 4 is $\Omega((N-K)\underline{p}^{K-1})$, ignoring constant factors. In the non-contextual setting, regret necessarily depends on the total number of arms N, whereas in the contextual setting, both upper and lower bounds of regret scale with the feature dimension d, not N. Hence, there is a need for an N-independent lower bound for contextual cascading bandits. To this end, Theorem 5.7 provides the first N-independent lower bound for contextual cascading bandits.

6 Numerical Study

In this section, we empirically evaluate the performance of our proposed algorithm, `UCB-CLB`, and compare it against three UCB-based baselines— `UCB-CCA` [7], `CLogUCB` [21], and `VA-CLogUCB` [21]—in the contextual cascading logistic bandit setting. We conduct simulated experiments with a real-world dataset: *MovieLens 100K* dataset. We transformed the dataset into a binary feedback setting, assigning a label of 1 to ratings of 4 or 5, and 0 to ratings of 1, 2 or 3. To construct feature representations, we applied truncated SVD extracting 5-dimensional embeddings for 943 users and 1682 movies. For each user t and movie i, we computed the feature vector $x_{t,i} \in \mathbb{R}^{25}$ by taking the outer product of their embeddings and vectorizing the result. We choose the random unknown parameter $\theta^* \in \mathbb{R}^{25}$. In the online setting, a user is sampled uniformly at random in each round t, and the agent selects an optimal movie from all 1682 movies ($N = 1682$). Additional experimental details are provided in Appendix G. To evaluate the effect of cascade length, we compare cumulative regret under two settings, $K = 5$ and $K = 10$, with all other parameters held constant. The running time is reported here only for $K = 10$.

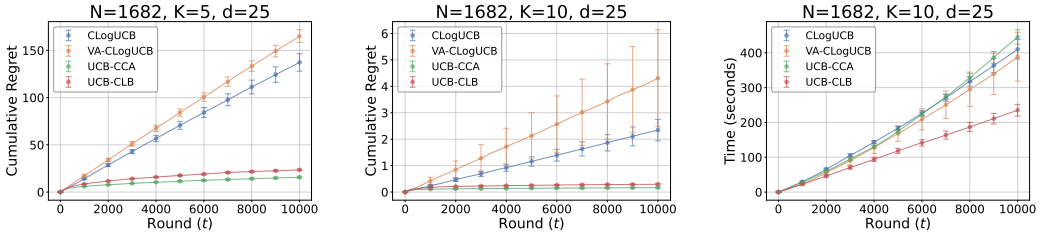

Figure 2: Cumulative regret for varying cascade lengths ($K = 5, 10$) and cumulative running time for $K = 10$ with $N = 1682$ and $d = 25$. Error bars indicate standard error, and all results are averaged over 5 random seeds.

Cumulative Regret Comparison. Figure 2 (left two plots) presents cumulative regret curves for different algorithms under $K = 5$ and $K = 10$. Among the logistic methods, our `UCB-CLB` and `UCB-CCA` achieve better regret than `CLogUCB` and `VA-CLogUCB` for both settings. We attribute this to the small κ (Assumption 5.3) in this experiments, which weakens the performance of $\frac{1}{\kappa}$-dependent methods (`UCB-CCA`, `CLogUCB`). Furthermore, as the cascade length K increases, all algorithms exhibit lower cumulative regret, which is consistent with our theoretical findings.

Computational Efficiency. Unlike MLE-based methods (`UCB-CCA`, `CLogUCB` and `VA-CLogUCB`), which incur high per-round computational costs due to solving complex optimization problems, our `UCB-CLB` avoids this issue and is highly scalable in large-scale cascading bandit settings. As shown in the rightmost plot of Figure 2, the running time of MLE-based methods grows exponentially with T. In contrast, `UCB-CLB` remains computationally efficient, exhibiting a runtime that scales linearly.

7 Conclusion

In this paper, we revisit the contextual cascading bandit problem and resolve the long-standing question of how the cascade length K affects regret. We first empirically show that, unlike prior theoretical results that regret grows with K, the regret actually decreases once K becomes sufficiently large. Motivated by this observation, we propose UCB-CLB, an OMD-based UCB algorithm that achieves the tightest known regret bound of $\tilde{\mathcal{O}}(\min\left\{K\bar{p}^{K-1}, 1\right\} d\sqrt{T})$, and a nearly matching lower bound. Our findings provide new theoretical and empirical insights into how longer cascades can enhance statistical efficiency in contextual cascading bandits.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2022-NR071853, RS-2023-00222663, RS-2024-00406908, and RS-2025-25420849), by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. RS-2025-02263754), and by AI-Bio Research Grant through Seoul National University.

References

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

[2] Marc Abeille, Louis Faury, and Clément Calauzènes. Instance-wise minimax-optimal algorithms for logistic bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 3691–3699. PMLR, 2021.

[3] Sanae Amani and Christos Thrampoulidis. Ucb-based algorithms for multinomial logistic regression bandits. *Advances in Neural Information Processing Systems*, 34:2913–2924, 2021.

[4] Peter Auer. Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, 3(Nov):397–422, 2002.

[5] Xi Chen, Yining Wang, and Yuan Zhou. Dynamic assortment optimization with changing contextual information. *Journal of machine learning research*, 21(216):1–44, 2020.

[6] Wang Chi Cheung and David Simchi-Levi. Assortment optimization under unknown multinomial logit choice models. *arXiv preprint arXiv:1704.00108*, 2017.

[7] Hyunjun Choi, Rajan Udwani, and Min-hwan Oh. Cascading contextual assortment bandits. *Advances in Neural Information Processing Systems*, 36, 2023.

[8] Louis Faury, Marc Abeille, Clément Calauzènes, and Olivier Fercoq. Improved optimistic algorithms for logistic bandits. In *International Conference on Machine Learning*, pages 3052–3060. PMLR, 2020.

[9] Louis Faury, Marc Abeille, Kwang-Sung Jun, and Clément Calauzènes. Jointly efficient and optimal algorithms for logistic bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 546–580. PMLR, 2022.

[10] Kwang-Sung Jun, Lalit Jain, Blake Mason, and Houssam Nassif. Improved confidence bounds for the linear logistic model and applications to bandits. In *International Conference on Machine Learning*, pages 5148–5157. PMLR, 2021.

[11] 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, 2015.

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

[13] Joongkyu Lee and Min-hwan Oh. Nearly minimax optimal regret for multinomial logistic bandit. *Advances in Neural Information Processing Systems*, 37:109003–109065, 2024.

[14] Joongkyu Lee and Min-hwan Oh. Improved online confidence bounds for multinomial logistic bandits. In *Forty-second International Conference on Machine Learning*, 2025.

[15] Junghyun Lee, Se-Young Yun, and Kwang-Sung Jun. Improved regret bounds of (multinomial) logistic bandits via regret-to-confidence-set conversion. In *International Conference on Artificial Intelligence and Statistics*, pages 4474–4482. PMLR, 2024.

[16] Junghyun Lee, Se-Young Yun, and Kwang-Sung Jun. A unified confidence sequence for generalized linear models, with applications to bandits. *Advances in Neural Information Processing Systems*, 37:124640–124685, 2024.

[17] Lihong Li, Yu Lu, and Dengyong Zhou. Provably optimal algorithms for generalized linear contextual bandits. In *International Conference on Machine Learning*, pages 2071–2080. PMLR, 2017.

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

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

[20] Xutong Liu, Jinhang Zuo, Siwei Wang, John CS Lui, Mohammad Hajiesmaili, Adam Wierman, and Wei Chen. Contextual combinatorial bandits with probabilistically triggered arms. In *International Conference on Machine Learning*, pages 22559–22593. PMLR, 2023.

[21] Xutong Liu, Xiangxiang Dai, Xuchuang Wang, Mohammad Hajiesmaili, and John Lui. Combinatorial logistic bandits. *arXiv preprint arXiv:2410.17075*, 2024.

[22] Min-hwan Oh and Garud Iyengar. Thompson sampling for multinomial logit contextual bandits. *Advances in Neural Information Processing Systems*, 32, 2019.

[23] Min-hwan Oh and Garud Iyengar. Multinomial logit contextual bandits: Provable optimality and practicality. In *Proceedings of the AAAI conference on artificial intelligence*, pages 9205–9213, 2021.

[24] Francesco Orabona. A modern introduction to online learning. *arXiv preprint arXiv:1912.13213*, 2019.

[25] Lijing Qin, Shouyuan Chen, and Xiaoyan Zhu. Contextual combinatorial bandit and its application on diversified online recommendation. In *Proceedings of the 2014 SIAM International Conference on Data Mining*, pages 461–469. SIAM, 2014.

[26] Daniel Vial, Sujay Sanghavi, Sanjay Shakkottai, and R Srikant. Minimax regret for cascading bandits. *Advances in Neural Information Processing Systems*, 35:29126–29138, 2022.

[27] Yu-Jie Zhang and Masashi Sugiyama. Online (multinomial) logistic bandit: Improved regret and constant computation cost. *Advances in Neural Information Processing Systems*, 36, 2024.

[28] Zixin Zhong, Wang Chi Chueng, and Vincent YF Tan. Thompson sampling algorithms for cascading bandits. *The Journal of Machine Learning Research*, 22(1):9915–9980, 2021.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction clearly state that the regret decreases when K is sufficiently large. This claim is theoretically supported by Theorems 5.6 and 6.1, and empirically validated by the experiments.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: In Section 5.3, we discuss the limitation that the linear model requires the probability of positive feedback to be bounded away from zero. However, we argue that this is a mild assumption in practice, as such events are rare in real-world scenarios.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: All theoretical results are stated with explicit assumptions. Complete proofs are in the supplementary material, with intuitions provided in the main paper.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: We describe all experimental settings in supplementary materials (See Appendix G) and provide code for full reproducibility.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: We use public data and provide anonymized code and instructions for reproducibility.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: The main paper provides a summary of all experimental settings, and full details including data processing and hyper-parameters are provided in the supplementary material (See Appendix G).

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: Error bars in all plots indicate the standard error (SE), as stated in the main text.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Details about compute resources, including hardware specifications and runtime, are provided in the supplementary material (See Appendix G).

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We comply with the NeurIPS Code of Ethics and identify no ethical concerns in our work.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: The paper discusses potential benefits for personalization and risks such as bias and privacy concerns, despite being primarily theoretical (See Appendix I).

    Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: Our paper does not involve any data or models that pose a risk of misuse.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: We used the MovieLens 100K dataset, which is publicly available for non-commercial use. We have cited the dataset and its source appropriately in the paper.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.

- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: We will release the full code with documentation to support reproducibility.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: We did not use crowdsourcing or human subjects in this work.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: We did not involve human subjects or crowdsourced data in our research.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.

- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [NA]

    Justification: We used LLMs only for writing and editing.

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.

A Notations

We provide the notations used throughout the appendix. The total number of rounds is denoted by T, the total number of arms by N, the cascade length by K, and the dimension of the feature vectors by d. The logistic function $\sigma_1(z)$ and its complement $\sigma_0(z)$ are defined as follows:

$$\sigma_1(z) := \frac{\exp(z)}{1+\exp(z)} \quad \text{and} \quad \sigma_0(z) := \frac{1}{1+\exp(z)} = 1 - \sigma_1(z),$$

for any $z \in \mathbb{R}$. The first and second derivatives of $\sigma_1(z)$ with respect to z are given by

$$\dot{\sigma}_1(z) := \sigma_0(z)\sigma_1(z),$$
$$\ddot{\sigma}_1(z) := \sigma_0(z)\sigma_1(z)(1 - 2\sigma_1(z)).$$

Also, $V_t = \sum_{\tau=1}^{t-1} \sum_{i \in O_\tau} x_{\tau,i} x_{\tau,i}^\top + \lambda I_d$.

Table 2 provides a summary of the symbols and their descriptions used throughout this section.

Table 2: Symbols

Symbol	Description
$\mathcal{I}$	set of all N arms
Π	set of all possible cascades
C_t	cascade chosen by our algorithm in round t
O_t	a list of observed arm in round t
i_{tk}	k-th arm in C_t in round t
$x_{t,i}$	feature vector for arm i given at round t
$y_{t,i}$	feedback for arm i given at round t
$f_t(C, \theta^*)$	$1 - \prod_{i \in C} \sigma_0(x_{t,i}^\top \theta^*)$, expected reward of the cascade C at round t
$\ell_{t,i}(\theta)$	$-\sum_{y \in \{0,1\}} \mathbb{I}\{y_{t,i} = y\} \log \sigma_y(x_{t,i}^\top \theta)$, loss function at round t
$\nabla \ell_{t,i}(\theta)$	$\big(\sigma_1(x_{t,i}^\top \theta) - y_{t,i}\big) x_{t,i}$, the first derivatives of $\ell_{t,i}(\theta)$ with respect to θ
$\nabla^2 \ell_{t,i}(\theta)$	$\dot{\sigma}_1(x_{t,i}^\top \theta) x_{t,i} x_{t,i}^\top$, the second derivatives of $\ell_{t,i}(\theta)$ with respect to θ
λ	regularization parameter
η	step size parameter
$\beta_{t,k}(\delta)$	$\mathcal{O}\left(\sqrt{d}\log(tK)\right)$, confidence radius at round t ($\beta_t = \beta_{t,1}$)
$u_{t,i}$	$x_{t,i}^\top \theta_t + \beta_t(\delta) \|x_{t,i}\|_{H_t^{-1}}$, optimistic estimate for arm i at round t,
$\tilde{f}_t(C)$	$1 - \prod_{i \in C} \sigma_0(u_{t,i})$, optimistic expected reward of the cascade C at round t

B Exploration Bonus

Lemma B.1. *Let θ_t be the online estimate as defined in Equation (3). Let $\mathcal{C}_t^{\text{ON}}(\delta)$ be a confidence set with a confidence radius $\beta_t(\delta)$ in Corollary 5.5. Define $u_{t,i} := x_{t,i}^\top \theta_t + \beta_t(\delta) \|x_{t,i}\|_{H_t^{-1}}$. Then, under the event $\{\forall t \geq 1, \theta^* \in \mathcal{C}_t^{\text{ON}}(\delta)\}$, for all $t \in [T]$ and $i \in \mathcal{I}$, the following holds:*

$$0 \leq u_{t,i} - x_{t,i}^\top \theta^* \leq 2\beta_t(\delta) \|x_{t,i}\|_{H_t^{-1}}.$$

Proof of Lemma B.1. We begin by bounding the estimation error as follows:

$$\begin{aligned} \left|x_{t,i}^\top \theta_t - x_{t,i}^\top \theta^*\right| &= \left|x_{t,i}^\top (\theta_t - \theta^*)\right| \\ &\overset{\text{(i)}}{\leq} \|x_{t,i}\|_{H_t^{-1}} \|\theta_t - \theta^*\|_{H_t} \\ &\overset{\text{(ii)}}{\leq} \beta_t(\delta) \|x_{t,i}\|_{H_t^{-1}}. \end{aligned}$$

where inequality (i) comes from Hölder's inequality and inequality (ii) is obtained by applying Corollary 5.5. Thus, the following inequality holds:

$$0 \leq x_{t,i}^\top \theta_t + \beta_t(\delta) \left\| x_{t,i} \right\|_{H_t^{-1}} - x_{t,i}^\top \theta^* = u_{t,i} - x_{t,i}^\top \theta^* \leq 2\beta_t(\delta) \left\| x_{t,i} \right\|_{H_t^{-1}}.$$

□

Lemma B.2. *For any given $C \in \Pi$, the following holds:*

$$f_t(C, \theta^*) \leq \tilde{f}_t(C).$$

Proof of Lemma B.2. Recall that $f_t(C, \theta^*) := 1 - \prod_{i \in C} \sigma_0(x_{t,i}^\top \theta^*)$ for any given $C \in \Pi$. Due to the fact that $\sigma_0(z)$ is a monotonically decreasing function of $z \in \mathbb{R}$ and $x_{t,i}^\top \theta^* \leq u_{t,i}$ for all $i \in C$ by Lemma B.1, we directly obtain the desired result. □

C Regret Upper Bound of UCB-CLB

C.1 Proof of Theorem 5.6

The proofs of all lemmas stated in this section to establish Theorem 5.6 are deferred to Appendix C.2.

Theorem 5.6. *Suppose $d \geq K$. Set the step size $\eta = \frac{1}{2}\log 2 + 2$ and penalty parameter $\lambda = \max\{84\sqrt{2}d\eta, K\}$. Let $\delta \in (0, 1]$. Then, under Assumption 5.1, UCB-CLB ensures*

$$\mathcal{R}(T) = \tilde{\mathcal{O}}\left(\min\left\{K\bar{p}^{K-1}d\sqrt{T} + K\bar{p}^{K-1}\frac{d^2}{\kappa \underline{p}} + K\bar{p}^{K-1}\frac{d^2}{\kappa}, d\sqrt{T} + \frac{d^2}{\kappa}\right\}\right).$$

with probability at least $1 - \delta$.

Proof of Theorem 5.6. We provide the detailed proof of the regret upper bound stated in 5.6. To this end, we separately analyze the two regimes that dominate the minimum term: (i) the K-dependent regime leading to $\tilde{\mathcal{O}}(K\bar{p}^{K-1}d\sqrt{T} + K\bar{p}^{K-1}\frac{d^2}{\kappa \underline{p}} + K\bar{p}^{K-1}\frac{d^2}{\kappa})$, and (ii) the K-independent regime yielding $\tilde{\mathcal{O}}(d\sqrt{T} + \frac{d^2}{\kappa})$. The combination of these results establishes the upper bound in Theorem 5.6.

In the following, we assume the good event $\mathcal{E}_\delta := \{\forall t \in [T] \text{ and } \forall k \in [|O_t|], \theta^* \in \mathcal{C}_{t,k}^{\texttt{CON}}(\delta)\}$ to hold, which happens with probability at least $1 - \delta$ according to Corollary 5.5. Recall that the optimistic expected reward for the cascade $C = (i_1, \ldots, i_K)$ in round t is defined as $\tilde{f}_t(C) := 1 - \prod_{i \in C_t} \sigma_0(u_{t,i})$ where $u_{t,i} := x_{t,i}^\top \theta_t + \beta_t(\delta) \left\| x_{t,i} \right\|_{H_t^{-1}}$ for all $i \in \mathcal{I}$. Also, recall that $C_t = (i_{t1}, \ldots, i_{tK})$ is the cascade selected by UCB-CLB in round t.

1. Deriving $\mathcal{R}(T) = \tilde{\mathcal{O}}(K\bar{p}^{K-1}d\sqrt{T} + K\bar{p}^{K-1}\frac{d^2}{\kappa \underline{p}} + K\bar{p}^{K-1}\frac{d^2}{\kappa})$.

Now, we bound the regret $\mathcal{R}(T)$ as follows:

$$\begin{aligned}
\mathcal{R}(T) &= \mathbb{E}\left[\sum_{t=1}^T f_t(C_t^*, \theta^*) - f_t(C_t, \theta^*)\right] \overset{\text{(i)}}{\leq} \mathbb{E}\left[\sum_{t=1}^T \tilde{f}_t(C_t^*) - f_t(C_t, \theta^*)\right] \\
&\overset{\text{(ii)}}{\leq} \mathbb{E}\left[\sum_{t=1}^T \tilde{f}_t(C_t) - f_t(C_t, \theta^*)\right] = \mathbb{E}\left[\sum_{t=1}^T \prod_{i \in C_t} \sigma_0(x_{t,i}^\top \theta^*) - \prod_{i \in C_t} \sigma_0(u_{t,i})\right] \\
&= \mathbb{E}\left[\sum_{t=1}^T \sum_{i_{tk} \in C_t} \prod_{l=1}^{k-1} \sigma_0(x_{t,i_{tl}}^\top \theta^*) \left(\sigma_1(u_{t,i_{tk}}) - \sigma_1(x_{t,i_{tk}}^\top \theta^*)\right) \prod_{m=k+1}^{K} \sigma_0(u_{t,i_{tm}})\right] \\
&\overset{\text{(iii)}}{\leq} \mathbb{E}\left[\sum_{t=1}^T \sum_{i \in C_t} \prod_{j \in C_t \setminus \{i\}} \sigma_0(x_{t,j}^\top \theta^*) \left(\sigma_1(u_{t,i}) - \sigma_1(x_{t,i}^\top \theta^*)\right)\right] \\
&\overset{\text{(iv)}}{\leq} \bar{p}^{K-1}\mathbb{E}\left[\sum_{t=1}^T \sum_{i \in C_t} \left(\sigma_1(u_{t,i}) - \sigma_1(x_{t,i}^\top \theta^*)\right)\right] \qquad (7)
\end{aligned}$$

where inequality (i) is obtained by applying Lemma B.2, inequality (ii) holds due to the definition of $C_t := \operatorname{argmax}_{C\in\Pi} \tilde{f}_t(C)$, inequality (iii) follows from the fact that $\sigma_0(\cdot)$ is a monotonically decreasing function and Lemma B.1 which guarantees $u_{t,i} \geq x_{t,i}^\top \theta^*$ for all rounds t and arm i, and inequality (iv) holds by the definition of $\bar{p}$ in Definition 5.2. Notably, the product term in Equation (7) plays a critical role in deriving the desired result, showing that the regret upper bound vanishes as the cascade length K becomes sufficiently large.

Then, we decompose the prediction error in Equation (7) using a second-order Taylor expansion, a standard technique widely employed in the logistic bandit literature [2]. This yields:

$$
\begin{aligned}
&\mathcal{R}(T) \\
&= \bar{p}^{K-1}\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i\in C_t} \dot{\sigma}_1(x_{t,i}^\top\theta^*)\left(u_{t,i} - x_{t,i}^\top\theta^*\right) + \frac{\ddot{\sigma}_1(z_{t,i})}{2}\left(u_{t,i} - x_{t,i}^\top\theta^*\right)^2\right] \\
&\overset{\text{(i)}}{\leq} \bar{p}^{K-1}\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i\in C_t} 2\beta_t(\delta)\dot{\sigma}_1(x_{t,i}^\top\theta^*)\left\|x_{t,i}\right\|_{H_t^{-1}} + \frac{\beta_t^2(\delta)}{5}\left\|x_{t,i}\right\|_{H_t^{-1}}^2\right] \\
&\overset{\text{(ii)}}{\leq} 2\bar{p}^{K-1}\beta_T(\delta)\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i\in C_t} \dot{\sigma}_1(x_{t,i}^\top\theta^*)\left\|x_{t,i}\right\|_{H_t^{-1}}\right] + \frac{\bar{p}^{K-1}\beta_T^2(\delta)}{5}\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i\in C_t}\left\|x_{t,i}\right\|_{H_t^{-1}}^2\right] \qquad (8)
\end{aligned}
$$

where $z_{t,i}$ is a convex combination of $u_{t,i}$ and $x_{t,i}$. The above inequality (i) is due to the fact that $\ddot{\sigma}(\cdot) \leq 0.1$ and from Lemma B.1 which states $u_{t,i} - x_{t,i}^\top\theta^* \leq 2\beta_t(\delta)\left\|x_{t,i}\right\|_{H_t^{-1}}$ for all $t \in [T]$ and $i \in \mathcal{I}$, and inequality (ii) holds since $\beta_t(\delta)$ is monotonically increasing with respect to t.

To control the first term on the right-hand side of Equation (8), we decompose the first summation term as follows:

$$
\begin{aligned}
&\sum_{t=1}^{T}\sum_{i\in C_t} \dot{\sigma}_1(x_{t,i}^\top\theta^*)\left\|x_{t,i}\right\|_{H_t^{-1}} \\
&= \sum_{t=1}^{T}\sum_{i\in C_t} \dot{\sigma}_1(x_{t,i}^\top\theta_t)\left\|x_{t,i}\right\|_{H_t^{-1}} + \sum_{t=1}^{T}\sum_{i\in C_t}\left(\dot{\sigma}_1(x_{t,i}^\top\theta^*) - \dot{\sigma}_1(x_{t,i}^\top\theta_t)\right)\left\|x_{t,i}\right\|_{H_t^{-1}} \\
&\overset{\text{(i)}}{\leq} \sum_{t=1}^{T}\sum_{i\in C_t} \dot{\sigma}_1(x_{t,i}^\top\theta_t)\left\|x_{t,i}\right\|_{H_t^{-1}} + \sum_{t=1}^{T}\sum_{i\in C_t}\frac{1}{4}\left|x_{t,i}^\top\theta^* - x_{t,i}^\top\theta_t\right|\left\|x_{t,i}\right\|_{H_t^{-1}} \\
&\overset{\text{(ii)}}{\leq} \sum_{t=1}^{T}\sum_{i\in C_t} \dot{\sigma}_1(x_{t,i}^\top\theta_t)\left\|x_{t,i}\right\|_{H_t^{-1}} + \sum_{t=1}^{T}\sum_{i\in C_t}\frac{1}{4}\left\|\theta^* - \theta_t\right\|_{H_t}\left\|x_{t,i}\right\|_{H_t^{-1}}^2 \\
&\overset{\text{(iii)}}{\leq} \sum_{t=1}^{T}\sum_{i\in C_t} \dot{\sigma}_1(x_{t,i}^\top\theta_t)\left\|x_{t,i}\right\|_{H_t^{-1}} + \frac{\beta_T(\delta)}{4}\sum_{t=1}^{T}\sum_{i\in C_t}\left\|x_{t,i}\right\|_{H_t^{-1}}^2. \qquad (9)
\end{aligned}
$$

where inequality (i) is obtained by applying Lemma C.2, inequality (ii) holds by applying Hölder's inequality, and inequality (iii) holds by applying Corollary 5.5 and the fact that $\beta_t(\delta)$ is monotonic increasing with respect to t.

Plugging Equation (9) into Equation (8) yields the following:

$$
\begin{aligned}
\mathcal{R}(T) \leq\ & 2\bar{p}^{K-1}\beta_T(\delta)\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i\in C_t} \dot{\sigma}_1(x_{t,i}^\top\theta_t)\left\|x_{t,i}\right\|_{H_t^{-1}}\right] \\
&+ \frac{7}{10}\bar{p}^{K-1}\beta_T^2(\delta)\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i\in C_t}\left\|x_{t,i}\right\|_{H_t^{-1}}^2\right]. \qquad (10)
\end{aligned}
$$

Before diving into bounding Equation (10), we first discuss the challenge in theoretical analysis of contextual cascading bandits and the swapping technique to address it. The general strategy for completing the proof is to utilize the elliptical potential lemma Abbasi-Yadkori et al. [1], a standard

lemma in the contextual bandit literature. However, it is challenging to directly apply the existing elliptical potential lemma in the contextual cascading bandit setting. As seen in Equation (10), we need to bound the term that accumulates all arms in C_t, while the Gram matrix H_t only accumulates the feature information of the observed arms. As a result, there is an issue in which the leftover sum (from $|O_t|+1$ to $|C_t|$) becomes out of control [19]. This limitation arises from the fact that the agent receives feedback only for the observed arms O_t, making it difficult to control the contribution of unobserved arms in the regret analysis. To address this issue, we exploit the optimistic exposure swapping technique first introduced in Choi et al. [7].

Let $C_t' \in \operatorname{argmax}_{C\in\Pi} \tilde{f}_t(C)$. The expected reward $f_t(C,\theta^*)$, as defined in Equation (1), is invariant under permutations of the arms in the cascade. Exploiting this property, the *doubly optimistic exposure swapping* technique rearranges the cascade C_t' by placing the two selected arms at the top, yielding a new cascade C_t without affecting the expected reward [7]. Let $i_t^{(1)}$ and $i_t^{(2)}$ be the arms in C_t' with the highest values of $\dot{\sigma}(x_{t,i}^\top\theta_t)\,\|x_{t,i}\|_{H_t^{-1}}$ and $\|x_{t,i}\|_{H_t^{-1}}$, respectively. We construct the cascade C_t by placing these arms in the first and second positions, followed by the remaining arms in C_t': $C_t := (i_t^{(1)}, i_t^{(2)}) \,\|\, \left(C_t' \setminus \{i_t^{(1)}, i_t^{(2)}\}\right)$, where $A \,\|\, B$ denotes the concatenation of two lists A and B. We denote the k-th arm in C_t as i_{tk}. As a side note, it is possible for the arm that should be placed in the first position to be the same as the arm that should be placed in the second position. In such cases, the arm is placed in the first position. This does not impact the regret upper bound derived in this section. For ease of analysis, we henceforth assume that the two arms are distinct.

Since the arm in the first position is always observed ($i_{t1} \in O_t$), the following inequality holds and effectively avoids the aforementioned out-of-control issue:

$$\sum_{i\in C_t} \dot{\sigma}_1(x_{t,i}^\top\theta_t)\,\|x_{t,i}\|_{H_t^{-1}} = K \max_{i\in C_t} \dot{\sigma}_1(x_{t,i}^\top\theta_t)\,\|x_{t,i}\|_{H_t^{-1}} = K\dot{\sigma}_1(x_{t,i_{t1}}^\top\theta_t)\,\|x_{t,i_{t1}}\|_{H_t^{-1}}\,. \tag{11}$$

Additionally, considering that the arm with the largest $\|x_{t,i}\|_{H_t^{-1}}$ in C_t is placed in the second position, the following equation holds:

$$\sum_{i\in C_t} \|x_{t,i}\|_{H_t^{-1}}^2 = K \max_{i\in C_t} \|x_{t,i}\|_{H_t^{-1}}^2 = K\,\|x_{t,i_{t2}}\|_{H_t^{-1}}^2\,. \tag{12}$$

Applying Equation (11) and Equation (12) to Equation (10), we obtain the following result:

$$\begin{aligned}\mathcal{R}(T) \le{}& 2K\bar{p}^{K-1}\beta_T(\delta)\mathbb{E}\left[\sum_{t=1}^{T} \dot{\sigma}_1(x_{t,i_{t1}}^\top\theta_t)\,\|x_{t,i_{t1}}\|_{H_t^{-1}}\right] \\ &+ \frac{7}{10}K\bar{p}^{K-1}\beta_T^2(\delta)\mathbb{E}\left[\sum_{t=1}^{T} \|x_{t,i_{t2}}\|_{H_t^{-1}}^2\right].\end{aligned} \tag{13}$$

Now, we first focus on the first summation term on the right-hand side of Equation (13). Still, directly applying the elliptical potential lemma [1] to the above equation is challenging due to a dependency mismatch: the weight term $\dot{\sigma}_1(x_{t,i_{t1}}^\top\theta_t)$ depends on the parameter θ_t, while the Gram matrix H_t^{-1} is computed using the parameter $\{\theta_{\tau,k+1}\}_{\tau\in[t-1],k\in[1,|O_\tau|]}$. To address this issue, we adopt the decomposition technique inspired by Lee et al. [16] whose work focuses on generalized linear bandits, and we adapt it to our cascading bandits. Specifically, we define an intermediary parameter $\tilde{\theta}_t$ as:

$$\tilde{\theta}_t := \operatorname*{argmin}_{\theta\in\mathcal{C}_{t,2}^{\mathtt{ON}}(\delta)} \dot{\sigma}_1(x_{t,i_{t1}}^\top\theta).$$

Note that $\dot{\sigma}_1(x_{t,i_{t1}}^\top\theta^*), \dot{\sigma}_1(x_{t,i_{t1}}^\top\theta_{t,2}) \ge \dot{\sigma}_1(x_{t,i_{t1}}^\top\tilde{\theta}_t)$ under the good event $\mathcal{E}_\delta$ and by applying Corollary 5.5. Using this intermediary parameter $\tilde{\theta}_t$, we decompose the summation term on the right-hand side of Equation (13) as follows:

$$\begin{aligned}&\sum_{t=1}^{T} \dot{\sigma}_1(x_{t,i_{t1}}^\top\theta_t)\,\|x_{t,i_{t1}}\|_{H_t^{-1}} \\ &= \sum_{t=1}^{T} \dot{\sigma}_1(x_{t,i_{t1}}^\top\tilde{\theta}_t)\,\|x_{t,i_{t1}}\|_{H_t^{-1}} + \sum_{t=1}^{T} \left(\dot{\sigma}_1(x_{t,i_{t1}}^\top\theta_t) - \dot{\sigma}_1(x_{t,i_{t1}}^\top\tilde{\theta}_t)\right)\|x_{t,i_{t1}}\|_{H_t^{-1}}\,.\end{aligned} \tag{14}$$

For the second term on the right-hand side of Equation (14), the following result holds:

$$
\begin{aligned}
&\sum_{t=1}^{T}\left(\dot{\sigma}_1(x_{t,i_{t1}}^\top\theta_t)-\dot{\sigma}_1(x_{t,i_{t1}}^\top\tilde{\theta}_t)\right)\|x_{t,i_{t1}}\|_{H_t^{-1}}\\
&\overset{\text{(i)}}{\leq}\sum_{t=1}^{T}\frac{1}{4}\left|x_{t,i_{t1}}^\top(\theta_t-\tilde{\theta}_t)\right|\|x_{t,i_{t1}}\|_{H_t^{-1}}\\
&\overset{\text{(ii)}}{\leq}\frac{1}{4}\sum_{t=1}^{T}\left\|\theta_t-\tilde{\theta}_t\right\|_{H_t}\cdot\|x_{t,i_{t1}}\|_{H_t^{-1}}^2\\
&\overset{\text{(iii)}}{\leq}\frac{1}{4}\sum_{t=1}^{T}\left(\|\theta_t-\theta^*\|_{H_t}+\|\theta^*-\theta_{t,2}\|_{H_t}+\left\|\theta_{t,2}-\tilde{\theta}_t\right\|_{H_t}\right)\|x_{t,i_{t1}}\|_{H_t^{-1}}^2\\
&\overset{\text{(iv)}}{\leq}\frac{1}{4}\sum_{t=1}^{T}\left(\|\theta_t-\theta^*\|_{H_t}+\|\theta^*-\theta_{t,2}\|_{H_{t,2}}+\left\|\theta_{t,2}-\tilde{\theta}_t\right\|_{H_{t,2}}\right)\|x_{t,i_{t1}}\|_{H_t^{-1}}^2\\
&\overset{\text{(v)}}{\leq}\frac{3\beta_{T,K}(\delta)}{4}\sum_{t=1}^{T}\|x_{t,i_{t1}}\|_{H_t^{-1}}^2\\
&\overset{\text{(vi)}}{\leq}\frac{3\beta_{T,K}(\delta)}{4}\kappa^{-1}\sum_{t=1}^{T}\|x_{t,i_{t1}}\|_{V_t^{-1}}^2\\
&\overset{\text{(vii)}}{\leq}\frac{3\beta_{T,K}(\delta)}{4}\kappa^{-1}2d\log\left(1+\frac{T}{d\lambda}\right).
\end{aligned}
\tag{15}
$$

where inequality (i) is obtained by applying Lemma C.2, inequality (ii) follows from Hölder's inequality, inequality (iii) is derived using the triangle inequality, inequality (iv) holds since $H_t \preceq H_{t,2}$, inequality (v) follows from Corollary 5.5 along with the monotonicity of $\beta_{t,k}(\delta)$ with respect to t and k, inequality (vi) is valid due to $H_t \succeq \kappa V_t$ ($V_t = \sum_{\tau=1}^{t-1}\sum_{i\in O_\tau} x_{\tau,i}x_{\tau,i}^\top + \lambda I_d$), and finally, (vii) results from Lemma F.2 with the identification $n_t = |O_t|$, $z_{t,k} = x_{t,i_{tk}}$, and $Z_t = V_t$.

To handle the first term on the right-hand side of Equation (14), we establish the following bound:

$$
\begin{aligned}
\sum_{t=1}^{T}\dot{\sigma}_1(x_{t,i_{t1}}^\top\tilde{\theta}_t)\|x_{t,i_{t1}}\|_{H_t^{-1}} &\overset{\text{(i)}}{\leq}\sum_{t=1}^{T}\sqrt{\dot{\sigma}_1(x_{t,i_{t1}}^\top\theta^*)}\sqrt{\dot{\sigma}_1(x_{t,i_{t1}}^\top\theta_{t,2})}\|x_{t,i_{t1}}\|_{H_t^{-1}}\\
&\overset{\text{(ii)}}{\leq}\sqrt{\sum_{t=1}^{T}\dot{\sigma}_1(x_{t,i_{t1}}^\top\theta^*)}\sqrt{\sum_{t=1}^{T}\dot{\sigma}_1(x_{t,i_{t1}}^\top\theta_{t,2})\|x_{t,i_{t1}}\|_{H_t^{-1}}^2}\\
&\overset{\text{(iii)}}{\leq}\sqrt{\frac{T}{4}}\sqrt{\sum_{t=1}^{T}\dot{\sigma}_1(x_{t,i_{t1}}^\top\theta_{t,2})\|x_{t,i_{t1}}\|_{H_t^{-1}}^2}\\
&\leq\sqrt{\frac{T}{4}}\sqrt{\sum_{t=1}^{T}\left\|\sqrt{\dot{\sigma}_1(x_{t,i_{t1}}^\top\theta_{t,2})}x_{t,i_{t1}}\right\|_{H_t^{-1}}^2}\\
&\overset{\text{(iv)}}{\leq}\sqrt{\frac{T}{4}}\sqrt{2d\log\left(1+\frac{T}{d\lambda}\right)}
\end{aligned}
\tag{16}
$$

where inequality (i) is valid due to $\dot{\sigma}_1(x_{t,i_{t1}}^\top\theta^*), \dot{\sigma}_1(x_{t,i_{t1}}^\top\theta_{t,2}) \geq \dot{\sigma}_1(x_{t,i_{t1}}^\top\tilde{\theta}_t)$, inequality (ii) results from the Cauchy-Schwarz inequality, inequality (iii) is due to the fact that $\dot{\sigma}_1(\cdot) \leq \frac{1}{4}$, and inequality (iv) follows from Lemma F.2 with $n_t = |O_t|$, $z_{t,k} = \sqrt{\dot{\sigma}_1(x_{t,i_{tk}}^\top\theta_{t,k+1})}x_{t,i_{tk}}$, and $Z_t = H_t$.

Plugging Equation (15) and Equation (16) into Equation (14), we have:

$$
\sum_{t=1}^{T}\dot{\sigma}_1(x_{t,i_{t1}}^\top\theta_t)\|x_{t,i_{t1}}\|_{H_t^{-1}} \leq \frac{3\beta_{T,K}(\delta)}{2}\kappa^{-1}d\log\left(1+\frac{T}{d\lambda}\right)+\sqrt{\frac{dT}{2}\log\left(1+\frac{T}{d\lambda}\right)}. \tag{17}
$$

Substituting Equation (17) into Equation (13), this yields the following bound:

$$\begin{aligned}\mathcal{R}(T) \leq \sqrt{2}K\bar{p}^{K-1}\beta_T(\delta)\sqrt{dT\log\left(1+\frac{T}{d\lambda}\right)} \\ + 3K\bar{p}^{K-1}\kappa^{-1}\beta_T(\delta)\beta_{T,K}d\log\left(1+\frac{T}{d\lambda}\right) \\ + \frac{7}{10}K\bar{p}^{K-1}\beta_T^2(\delta)\mathbb{E}\left[\sum_{t=1}^{T}\|x_{t,i_{t2}}\|^2_{H_t^{-1}}\right]. \end{aligned} \tag{18}$$

To complete the regret bound, it remains to bound the last term on the right-hand side of Equation (18), which involves the cumulative weighted norm of the second-position features. Unlike the first position, the arm in the second position is not always observed. To address this issue, we present the following lemma. The proof of Lemma C.1 is deferred to Appendix C.2.1.

Lemma C.1. *Recall from Definition 5.2 that $\underline{p}$ denotes the lower bound of $\sigma_0(x^\top\theta^*)$ for any $x \in X$. Suppose C_t is constructed under the doubly optimistic swapping technique placing the arm with the largest $\|x_{t,i}\|_{H_t^{-1}}$ in the second position of C_t Then, we have the following inequality:*

$$\mathbb{E}\left[\sum_{t=1}^{T}\|x_{t,i_{t2}}\|^2_{H_t^{-1}}\right] \leq \underline{p}^{-1}\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{1}\{|O_t|\geq 2\}\cdot\|x_{t,i_{t2}}\|^2_{H_t^{-1}}\right].$$

Applying Lemma C.1 to Equation (18), we obtain the following bound:

$$\begin{aligned}\mathbb{E}\left[\sum_{t=1}^{T}\|x_{t,i_{t2}}\|^2_{H_t^{-1}}.\right] &\overset{\text{(i)}}{\leq} \underline{p}^{-1}\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{1}\{|O_t|\geq 2\}\cdot\|x_{t,i_{t2}}\|^2_{H_t^{-1}}\right] \\ &\overset{\text{(ii)}}{\leq} (\underline{p}\kappa)^{-1}\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{1}\{|O_t|\geq 2\}\cdot\|x_{t,i_{t2}}\|^2_{V_t^{-1}}\right] \\ &\overset{\text{(iii)}}{\leq} (\underline{p}\kappa)^{-1}2d\log\left(1+\frac{2T}{d\lambda}\right). \end{aligned} \tag{19}$$

where inequality (i) is obtained by applying Lemma C.1, inequality (ii) holds since $H_t \succeq \kappa V_t$, and inequality (iii) follows from Lemma F.3 with the identification $n_t = |O_t|$, $z_{t,k} = x_{t,i_{tk}}$, and $Z_t = V_t$.

Finally, combining Equation (19) with Equation (18) yields the claimed result:

$$\begin{aligned}\mathcal{R}(T) &\leq \sqrt{2}K\bar{p}^{K-1}\beta_T(\delta)\sqrt{dT\log\left(1+\frac{T}{d\lambda}\right)} \\ &\quad + 3K\bar{p}^{K-1}\kappa^{-1}\beta_T(\delta)\beta_{T,K}d\log\left(1+\frac{T}{d\lambda}\right) \\ &\quad + \frac{7}{5}K\bar{p}^{K-1}(\underline{p}\kappa)^{-1}\beta_T^2(\delta)d\log\left(1+\frac{2T}{d\lambda}\right) \\ &= \tilde{\mathcal{O}}\left(K\bar{p}^{K-1}d\sqrt{T} + K\bar{p}^{K-1}\frac{d^2}{\kappa\underline{p}} + K\bar{p}^{K-1}\frac{d^2}{\kappa}\right). \end{aligned}$$

2. Deriving $\mathcal{R}(T) = \tilde{\mathcal{O}}(d\sqrt{T} + \frac{d^2}{\kappa})$.

Now, we bound the regret $\mathcal{R}(T)$ as follows:

$$\begin{aligned}\mathcal{R}(T) &= \mathbb{E}\left[\sum_{t=1}^{T} f_t(C_t^*,\theta^*) - f_t(C_t,\theta^*)\right] \overset{\text{(i)}}{\leq} \mathbb{E}\left[\sum_{t=1}^{T}\tilde{f}_t(C_t^*) - f_t(C_t,\theta^*)\right] \\ &\overset{\text{(ii)}}{\leq} \mathbb{E}\left[\sum_{t=1}^{T}\tilde{f}_t(C_t) - f_t(C_t,\theta^*)\right] = \mathbb{E}\left[\sum_{t=1}^{T}\prod_{i\in C_t}\sigma_0(x_{t,i}^\top\theta^*) - \prod_{i\in C_t}\sigma_0(u_{t,i})\right] \\ &= \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i_{tk}\in C_t}\prod_{l=1}^{k-1}\sigma_0(x_{t,i_{tl}}^\top\theta^*)\left(\sigma_1(u_{t,i_{tk}}) - \sigma_1(x_{t,i_{tk}}^\top\theta^*)\right)\prod_{m=k+1}^{K}\sigma_0(u_{t,i_{tm}})\right] \end{aligned} \tag{20}$$

We define an intermediary parameter $\tilde{\theta}_{t,i}$ as follows:

$$\tilde{\theta}_{t,i} := \operatorname*{argmin}_{\theta \in \cup_{\tau \in [t,T], k \in [K]} \mathcal{C}^{\texttt{OM}}_{\tau,k}(\delta)} \dot{\sigma}_1(x_{t,i}^\top \theta), \quad \forall t \in [T], i \in \mathcal{I}$$

This ensures that the gram matrix H_t satisfies the following lower bound:

$$H_t \succeq \sum_{\tau=1}^{t-1} \sum_{i \in O_t} \dot{\sigma}_1(x_{\tau,i}^\top \tilde{\theta}_{\tau,i}) x_{\tau,i} x_{\tau,i}^\top + \lambda I_d =: L_t.$$

The following inequality chain bound the deviation between the true gradient term $\dot{\sigma}1(x_{t,i}^\top \theta^*)$ and $\dot{\sigma}1(x_{t,i}^\top \tilde{\theta}_{t,i})$.

$$\begin{aligned}
\dot{\sigma}_1(x_{t,i}^\top \theta^*) &= \dot{\sigma}_1(x_{t,i}^\top \tilde{\theta}_{t,i}) + \left(\dot{\sigma}_1(x_{t,i}^\top \theta^*) - \dot{\sigma}_1(x_{t,i}^\top \tilde{\theta}_{t,i})\right) \\
&\overset{\text{(i)}}{\leq} \dot{\sigma}_1(x_{t,i}^\top \tilde{\theta}_{t,i}) + \frac{1}{4} \left| x_{t,i}^\top (\theta^* - \tilde{\theta}_{t,i}) \right| \\
&\overset{\text{(ii)}}{\leq} \dot{\sigma}_1(x_{t,i}^\top \tilde{\theta}_{t,i}) + \frac{1}{4} \|x_{t,i}\|_{H_t^{-1}} \|\theta^* - \tilde{\theta}_{t,i}\|_{H_t} \\
&\leq \dot{\sigma}_1(x_{t,i}^\top \tilde{\theta}_{t,i}) + \frac{\beta_{T,K}(\delta)}{4} \|x_{t,i}\|_{H_t^{-1}}
\end{aligned} \tag{21}$$

where inequality (i) is from Lemma C.2, inequality (ii) is obtained by applying the Cauchy-Schwarz inequality, and inequality (iii) is due to Lemma B.1.

Building on the gradient-deviation bound above, we derive the following inequality to control the difference between the UCB estimate and the true expected click probability as follows:

$$\begin{aligned}
&\sigma_1(u_{t,i}) - \sigma_1(x_{t,i}^\top \theta^*) \\
&= \dot{\sigma}_1(x_{t,i}^\top \theta^*) \left(u_{t,i} - x_{t,i}^\top \theta^*\right) + \frac{\ddot{\sigma}_1(z_{t,i})}{2} \left(u_{t,i} - x_{t,i}^\top \theta^*\right)^2 \\
&\overset{\text{(i)}}{\leq} 2\beta_t(\delta) \dot{\sigma}_1(x_{t,i}^\top \theta^*) \|x_{t,i}\|_{H_t^{-1}} + \frac{\beta_t^2(\delta)}{5} \|x_{t,i}\|_{H_t^{-1}}^2 \\
&\overset{\text{(ii)}}{\leq} 2\beta_T(\delta) \dot{\sigma}_1(x_{t,i}^\top \tilde{\theta}_{t,i}) \|x_{t,i}\|_{H_t^{-1}} + \frac{7\beta_{T,K}^2(\delta)}{10} \|x_{t,i}\|_{H_t^{-1}}^2 \\
&\leq 2\beta_T(\delta) \dot{\sigma}_1(x_{t,i}^\top \tilde{\theta}_{t,i}) \|x_{t,i}\|_{L_t^{-1}} + \frac{7\beta_{T,K}^2(\delta)}{10} \|x_{t,i}\|_{H_t^{-1}}^2
\end{aligned}$$

where inequality (i) is obtained by applying Lemma B.1, inequality (ii) is by Equation (21).

Define $\zeta_{t,i} := 2\beta_T(\delta) \dot{\sigma}_1(x_{t,i}^\top \tilde{\theta}_{t,i}) \|x_{t,i}\|_{L_t^{-1}}$ and $\xi_{t,i} := \frac{7\beta_{T,K}^2(\delta)}{10} \|x_{t,i}\|_{H_t^{-1}}^2$. Then, we can bound $Equation$ (20) as follows:

$$\begin{aligned}
\mathcal{R}(T) &\leq \mathbb{E}\left[\sum_{t=1}^{T} \sum_{i_{tk} \in C_t} \prod_{l=1}^{k-1} \sigma_0(x_{t,i_{tl}}^\top \theta^*) \left(\zeta_{t,i_{tk}} + \xi_{t,i_{tk}}\right) \prod_{m=k+1}^{K} \sigma_0(u_{t,i_{tm}})\right] \\
&\leq \underbrace{\mathbb{E}\left[\sum_{t=1}^{T} \sum_{i_{tk} \in C_t} \prod_{l=1}^{k-1} \sigma_0(x_{t,i_{tl}}^\top \theta^*) \zeta_{t,i_{tk}} \prod_{m=k+1}^{K} \sigma_0(x_{t,i_{tm}}^\top \theta^*)\right]}_{\text{Term 1}} \\
&\quad + \underbrace{\mathbb{E}\left[\sum_{t=1}^{T} \sum_{i_{tk} \in C_t} \prod_{l=1}^{k-1} \sigma_0(x_{t,i_{tl}}^\top \theta^*) \xi_{t,i_{tk}}\right]}_{\text{Term 2}}
\end{aligned}$$

We first bound the Term 1 as follows:

$$
\begin{aligned}
\text{Term 1} &\overset{\text{(i)}}{\leq} \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i_{tk}\in C_t}\prod_{l=1}^{k-1}\sigma_0(x_{t,i_{tl}}^\top\theta^*)\zeta_{t,i_{tk}}\frac{\sqrt{\prod_{m=k+1}^{K}\sigma_0(x_{t,i_{tm}}^\top\theta^*)\sigma_1(x_{t,i_{tk}}^\top\theta^*)}}{\sqrt{\sigma_1(x_{t,i_{tk}}^\top\theta^*)\sigma_0(x_{t,i_{tk}}^\top\theta^*)}}\right] \\
&\overset{\text{(ii)}}{\leq} \mathbb{E}\left[\sum_{t=1}^{T}\sqrt{\sum_{i_{tk}\in C_t}\frac{\left(\prod_{l=1}^{k-1}\sigma_0(x_{t,i_{tl}}^\top\theta^*)\right)^2\zeta_{t,i_{tk}}^2}{\sigma_1(x_{t,i_{tk}}^\top\theta^*)\sigma_0(x_{t,i_{tk}}^\top\theta^*)}}\sqrt{\sum_{i_{tk}\in C_t}\prod_{m=k+1}^{K}\sigma_0(x_{t,i_{tm}}^\top\theta^*)\sigma_1(x_{t,i_{tk}}^\top\theta^*)}\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T}\sqrt{\sum_{i_{tk}\in C_t}\frac{\prod_{l=1}^{k-1}\sigma_0(x_{t,i_{tl}}^\top\theta^*)\zeta_{t,i_{tk}}^2}{\sigma_1(x_{t,i_{tk}}^\top\theta^*)\sigma_0(x_{t,i_{tk}}^\top\theta^*)}}\sqrt{1-\prod_{i\in C_t}\sigma_0(x_{t,i}^\top\theta^*)}\right] \\
&\leq \mathbb{E}\left[\sum_{t=1}^{T}\sqrt{\sum_{i_{tk}\in C_t}\frac{\prod_{l=1}^{k-1}\sigma_0(x_{t,i_{tl}}^\top\theta^*)\zeta_{t,i_{tk}}^2}{\sigma_1(x_{t,i_{tk}}^\top\theta^*)\sigma_0(x_{t,i_{tk}}^\top\theta^*)}}\right] \\
&\overset{\text{(iii)}}{\leq} \sqrt{T}\cdot\sqrt{\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i_{tk}\in C_t}\frac{\prod_{l=1}^{k-1}\sigma_0(x_{t,i_{tl}}^\top\theta^*)\zeta_{t,i_{tk}}^2}{\sigma_1(x_{t,i_{tk}}^\top\theta^*)\sigma_0(x_{t,i_{tk}}^\top\theta^*)}\right]}
\end{aligned}
\tag{22}
$$

where inequality (i) follows from the fact that $\sigma(z) \leq \sqrt{\sigma(z)}$ for $z \in (0,1)$ and by multiplying and dividing each summand for $i_{t,k} \in C_t$ with the corresponding $\sqrt{\sigma_1(x_{t,i_{tk}}^\top\theta)}$ and $\sqrt{\sigma_1(x_{t,i_{tk}}^\top\theta)\,\sigma_0(x_{t,i_{tk}}^\top\theta^*)}$, inequality (ii) is due to the Cauchy-Schwarz inequality, and inequality (iii) is obtained by applying the Cauchy-Schwarz inequality and Jensen's inequality. Each term in the summation of Equation (22) is bounded in the following inequality

$$
\begin{aligned}
\frac{\zeta_{t,i_{tk}}^2}{\sigma_1(x_{t,i_{tk}}^\top\theta^*)\sigma_0(x_{t,i_{tk}}^\top\theta^*)} &= \frac{4\beta_T^2(\delta)\dot{\sigma}_1^2(x_{t,i_{tk}}^\top\tilde{\theta}_{t,i_{tk}})\left\|x_{t,i_{tk}}\right\|_{L_t^{-1}}^2}{\sigma_1(x_{t,i_{tk}}^\top\theta^*)\sigma_0(x_{t,i_{tk}}^\top\theta^*)} \\
&\overset{\text{(i)}}{\leq} \frac{4\beta_T^2(\delta)\dot{\sigma}_1(x_{t,i_{tk}}^\top\tilde{\theta}_{t,i_{tk}})\left\|x_{t,i_{tk}}\right\|_{L_t^{-1}}^2}{\sigma_1(x_{t,i_{tk}}^\top\theta^*)\sigma_0(x_{t,i_{tk}}^\top\theta^*)} \\
&= 4\beta_T^2(\delta)\dot{\sigma}_1(x_{t,i_{tk}}^\top\theta^*)\dot{\sigma}_1(x_{t,i_{tk}}^\top\tilde{\theta}_{t,i_{tk}})\left\|x_{t,i_{tk}}\right\|_{L_t^{-1}}^2
\end{aligned}
\tag{23}
$$

where inequality (i) is due to the definition of the intermediary parameter $\tilde{\theta}_{t,i_{tk}}$. Substituting Equation (23) into Equation (22) yields the following bound, and the remaining step is to apply the triggering probability equivalence (TPE) technique of Liu et al. [20, 21]:

$$
\begin{aligned}
\text{Term 1} &\leq 2\beta_T(\delta)\sqrt{T}\cdot\sqrt{\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i_{tk}\in C_t}\prod_{l=1}^{k-1}\sigma_0(x_{t,i_{tl}}^\top\theta^*)\dot{\sigma}_1(x_{t,i_{tk}}^\top\tilde{\theta}_{t,i_{tk}})\left\|x_{t,i_{tk}}\right\|_{L_t^{-1}}^2\right]} \\
&= 2\beta_T(\delta)\sqrt{T}\cdot\sqrt{\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i_{tk}\in C_t}\mathbb{E}\left[\prod_{l=1}^{k-1}\sigma_0(x_{t,i_{tl}}^\top\theta^*)\mid\mathcal{H}_t\right]\dot{\sigma}_1(x_{t,i_{tk}}^\top\tilde{\theta}_{t,i_{tk}})\left\|x_{t,i_{tk}}\right\|_{L_t^{-1}}^2\right]} \\
&= 2\beta_T(\delta)\sqrt{T}\cdot\sqrt{\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i_{tk}\in C_t}\mathbb{E}\left[\mathbb{I}\{i_{tk}\in O_t\}\mid\mathcal{H}_t\right]\dot{\sigma}_1(x_{t,i_{tk}}^\top\tilde{\theta}_{t,i_{tk}})\left\|x_{t,i_{tk}}\right\|_{L_t^{-1}}^2\right]} \\
&= 2\beta_T(\delta)\sqrt{T}\cdot\sqrt{\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i_{tk}\in C_t}\mathbb{I}\{i_{tk}\in O_t\}\dot{\sigma}_1(x_{t,i_{tk}}^\top\tilde{\theta}_{t,i_{tk}})\left\|x_{t,i_{tk}}\right\|_{L_t^{-1}}^2\right]} \\
&= 2\beta_T(\delta)\sqrt{T}\cdot\sqrt{\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i\in O_t}\dot{\sigma}_1(x_{t,i}^\top\tilde{\theta}_{t,i})\left\|x_{t,i}\right\|_{L_t^{-1}}^2\right]}.
\end{aligned}
$$

Next, we bound the Term 2 as follows:

$$\begin{aligned}
\text{Term 2} &= \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i_{tk}\in C_t}\prod_{l=1}^{k-1}\sigma_0(x_{t,i_{tl}}^\top\theta^*)\left(\xi_{t,i_{tk}}\right)\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i_{tk}\in C_t}\mathbb{E}\left[\prod_{l=1}^{k-1}\sigma_0(x_{t,i_{tl}}^\top\theta^*)\mid\mathcal{H}_t\right]\xi_{t,i_{tk}}\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i_{tk}\in C_t}\mathbb{E}\left[\mathbb{I}\{i_{tk}\in O_t\}\mid\mathcal{H}_t\right]\xi_{t,i_{tk}}\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i_{tk}\in C_t}\mathbb{I}\{i_{tk}\in O_t\}\xi_{t,i_{tk}}\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i\in O_t}\xi_{t,i}\right] \\
&= \frac{7\beta_{T,K}^2(\delta)}{10}\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i\in O_t}\|x_{t,i}\|_{H_t^{-1}}^2\right].
\end{aligned}$$

Thus, we have:

$$\begin{aligned}
\mathcal{R}(T) &\le 2\beta_T(\delta)\sqrt{T}\cdot\sqrt{\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i\in O_t}\dot{\sigma}_1(x_{t,i}^\top\tilde{\theta}_{t,i})\|x_{t,i}\|_{L_t^{-1}}^2\right]}+\frac{7\beta_{T,K}^2(\delta)}{10}\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i\in O_t}\|x_{t,i}\|_{H_t^{-1}}^2\right] \\
&\le 2\beta_T(\delta)\sqrt{T}\cdot\sqrt{2d\log\left(1+\frac{TK}{d\lambda}\right)}+\frac{7\beta_{T,K}^2(\delta)}{10\kappa}\left(2d\log\left(1+\frac{TK}{d\lambda}\right)\right) \\
&= \tilde{\mathcal{O}}\left(d\sqrt{T}+\frac{d^2}{\kappa}\right)
\end{aligned}$$

Combining the bounds for Term 1 and Term 2 and applying Lemma F.4 and Lemma F.7 yield the unified regret bound claimed in Theorem 5.6. □

C.2 Proof of Lemmas for Theorem 5.6

C.2.1 Proof of Lemma C.1

Lemma C.1. *Recall from Definition 5.2 that $\underline{p}$ denotes the lower bound of $\sigma_0(x^\top\theta^*)$ for any $x\in X$. Suppose C_t is constructed under the doubly optimistic swapping technique placing the arm with the largest $\|x_{t,i_{tk}}\|_{H_t^{-1}}$ in the second position of C_t. The following inequality holds:*

$$\mathbb{E}\left[\sum_{t=1}^{T}\|x_{t,i_{t2}}\|_{H_t^{-1}}^2\right]\le\underline{p}^{-1}\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{1}\{|O_t|\ge 2\}\cdot\|x_{t,i_{t2}}\|_{H_t^{-1}}^2\right].$$

Proof. By the law of total expectation, we have:

$$\mathbb{E}\left[\sum_{t=1}^{T}\|x_{t,i_{t2}}\|_{H_t^{-1}}^2\right]=\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{E}\left[\|x_{t,i_{t2}}\|_{H_t^{-1}}^2\mid\mathcal{H}_t\right]\right].$$

Now we introduce some definitions related to the probability of observing a specific arm, which is used exclusively in this proof. Let p_{t,C_t} be the probability of the second-position arm of C_t being observed in round t. Also, let $p^*=\min_{1\le t\le T}\min_{C\in\Pi}p_{t,C_t}$. When C_t is fixed, p_{t,C_t} is the probability that $|O_t|\ge 2$, and thus

$$\mathbb{E}\left[\frac{1}{p_{t,C_t}}\mathbb{1}\left\{|O_t|\ge 2\;\middle|\;C_t\right\}\right]=1.$$

Then we have:

$$\begin{aligned}
\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{E}\left[\|x_{t,i_{t2}}\|^2_{H_t^{-1}}\,|\mathcal{H}_t\right]\right] &= \mathbb{E}\left[\sum_{t=1}^{T}\mathbb{E}\left[\|x_{t,i_{t2}}\|^2_{H_t^{-1}}\cdot\mathbb{E}\left[\frac{1}{p_{t,C_t}}\mathbb{1}\left\{|O_t|\geq 2\right\}\,\middle|\,C_t\right]\,\middle|\,\mathcal{H}_t\right]\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T}\mathbb{E}\left[\|x_{t,i_{t2}}\|^2_{H_t^{-1}}\cdot\frac{1}{p_{t,C_t}}\mathbb{1}\left\{|O_t|\geq 2\right\}\,\middle|\,\mathcal{H}_t\right]\right] \\
&\leq \frac{1}{p^*}\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{E}\left[\mathbb{1}\left\{|O_t|\geq 2\right\}\|x_{t,i_{t2}}\|^2_{H_t^{-1}}\,\middle|\,\mathcal{H}_t\right]\right] \\
&= \frac{1}{p^*}\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{1}\left\{|O_t|\geq 2\right\}\|x_{t,i_{t2}}\|^2_{H_t^{-1}}\right] \\
&= \frac{1}{p}\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{1}\left\{|O_t|\geq 2\right\}\|x_{t,i_{t2}}\|^2_{H_t^{-1}}\right]
\end{aligned}$$

where the last inequality follows from Definition 5.2. □

C.2.2 Proof of Lemmas for Lemma C.2

Lemma C.2. *Given $x \in \mathbb{R}^d$, for any $\theta_1, \theta_2 \in \mathbb{R}^d$,*

$$\dot{\sigma}_1(x^\top\theta_1) - \dot{\sigma}_1(x^\top\theta_2) \leq \frac{1}{4}|x^\top(\theta_1-\theta_2)|.$$

Proof. We start from Lemma E.3 of Lee et al. [16] (restated in Lemma F.5):

$$\begin{aligned}
\dot{\sigma}_1(x^\top\theta_1) - \dot{\sigma}_1(x^\top\theta_2) &\leq |\sigma_1(x^\top\theta_1) - \sigma_1(x^\top\theta_2)| \\
&\leq \dot{\sigma}_1(z)|x^\top(\theta_1-\theta_2)| \\
&\leq \frac{1}{4}|x^\top(\theta_1-\theta_2)|.
\end{aligned}$$

□

D Regret Lower Bound

In this section, we provide the proof of Theorem 5.7. To begin with, we construct a hard instance, inspired by Chen et al. [5], Lee and Oh [13].

D.1 Adversarial Construction and Bayes Risk

Let $\epsilon \in (0, 1/d\sqrt{d})$ be a small positive constant, to be specified later. For each subset $V \subseteq [d]$, we define the parameter $\theta_V \in \mathbb{R}^d$ as follows: $[\theta_V]_j = \epsilon$ for all $j \in V$, and $[\theta_V]_j = 0$ for all $j \notin V$. Next, we define the parameter set:

$$\theta \in \Theta := \{\theta_V : V \in \mathcal{V}_{d/4}\} := \{\theta_V : V \subseteq [d], |V| = d/4\},$$

where $\mathcal{V}_k$ represents the collection of all subsets of $[d]$ whose size is k. Let d be divisible by 4.

Let context vectors remain invariant across all rounds t. For each $U \in \mathcal{V}_{d/4}$, we construct K identical context vectors x_U as follows:

$$[x_U]_j = 1/\sqrt{d} \quad \text{for } j \in U; \quad [x_U]_j = 0 \quad \text{for } j \notin U.$$

Note that since there are K identical context vectors, the total number of arms is $N = K \cdot \binom{d}{d/4}$.

For any $V, U \in \mathcal{V}_{d/4}$, the boundedness assumption (Assumption 5.1) is satisfied as follows:

$$\|\theta_V\|_2 \leq \sqrt{d\epsilon^2} \leq 1, \quad \|x_U\|_2 \leq \sqrt{d \cdot 1/d} = 1.$$

Moreover, the worst-case expected regret of any policy π can be lower bounded by the "average" regret over a uniform prior over Θ as follows:

$$\begin{aligned}\sup_{\theta} \mathbb{E}_{\theta}^{\pi}\left[\mathcal{R}_{\theta}(T)\right] &= \sup_{\theta} \mathbb{E}_{\theta}^{\pi} \sum_{t=1}^{T} f(C^*, \theta) - f(C_t, \theta) \\ &\geq \max_{\theta_V} \mathbb{E}_{\theta_V}^{\pi} \sum_{t=1}^{T} f(C^*, \theta_V) - f(C_t, \theta_V) \\ &\geq \frac{1}{|\mathcal{V}_{d/4}|} \sum_{V \in \mathcal{V}_{d/4}} \mathbb{E}_{\theta_V}^{\pi} \sum_{t=1}^{T} f(C^*, \theta_V) - f(C_t, \theta_V). \end{aligned} \tag{24}$$

This simplifies the problem of lower bounding the worst-case regret of any policy to the task of lower bounding the *Bayes risk* for the constructed parameter set.

D.2 Main Proof of Theorem 5.7

Proof of Theorem 5.7. For any sequence of cascades $\{C_t\}_{t=1}^{T}$ produced by policy π, we denote an alternative sequence $\{\tilde{C}_t\}_{t=1}^{T}$ that results in lower regret under the parameterization.

Let $x_{U_{t1}}, \ldots, x_{U_{tK}}$ be the distinct feature vectors contained in cascades C_t, where $U_{t1}, \ldots, U_{tK} \in \mathcal{V}_{d/4}$. Denote $U_t^* = \operatorname{argmax}_{U \in \{U_{t1}, \ldots, U_{tK}\}} x_U^\top \theta_V$, where θ_V is the underlying parameter.

Then, leveraging the properties of the cascading structure and the non-decreasing nature of the sigmoid function σ_1, we make the following observation:

Proposition D.1. *For all $V \in \mathcal{V}_{d/4}$, we have*

$$\mathbb{E}_{\theta_V}^{\pi} \sum_{t=1}^{T} f(C^*, \theta_V) - f(C_t, \theta_V) \geq \min\left\{K\underline{p}^{K-1}, 1\right\} \cdot \underbrace{\mathbb{E}_{\theta_V}^{\pi} \sum_{t=1}^{T} \left(\sigma_1(x_V^\top \theta_V) - \sigma_1(x_{U_t^*}^\top \theta_V)\right)}_{\textit{regret for logistic bandits}}$$

Proof of Proposition D.1.

$$\begin{aligned}&\mathbb{E}_{\theta_V}^{\pi} \sum_{t=1}^{T} f(C^*, \theta_V) - f(C_t, \theta_V) \\ &= \mathbb{E}_{\theta_V}^{\pi} \sum_{t=1}^{T} \left(1 - \prod_{k=1}^{K} \sigma_0(x_V^\top \theta_V)\right) - \left(1 - \prod_{k=1}^{K} \sigma_0(x_{U_{tk}}^\top \theta_V)\right) \\ &= \mathbb{E}_{\theta_V}^{\pi} \sum_{t=1}^{T} \sum_{k=1}^{K} \left(\prod_{l=1}^{k-1} \sigma_0(x_{U_{tl}}^\top \theta_V)\right) \left(\sigma_1(x_V^\top \theta_V) - \sigma_1(x_{U_{tk}}^\top \theta_V)\right) \left(\sigma_0(x_V^\top \theta_V)\right)^{K-k} \\ &\geq \underline{p}^{K-1} \mathbb{E}_{\theta_V}^{\pi} \sum_{t=1}^{T} \sum_{k=1}^{K} \left(\sigma_1(x_V^\top \theta_V) - \sigma_1(x_{U_{tk}}^\top \theta_V)\right) && (\underline{p} = \inf_x \sigma_0\left(x^\top \theta_V\right)) \\ &\geq K\underline{p}^{K-1} \mathbb{E}_{\theta_V}^{\pi} \sum_{t=1}^{T} \left(\sigma_1(x_V^\top \theta_V) - \sigma_1(x_{U_t^*}^\top \theta_V)\right). && (\sigma_1(x_{U_t^*}^\top \theta_V) \geq \sigma_1(x_{U_{tk}}^\top \theta_V)) \end{aligned}$$

Note that, by definition, we have

$$\underline{p} = \inf_x \sigma_0\left(x^\top \theta_V\right) = \frac{1}{1 + e^{x_V^\top \theta_V}} = \frac{1}{1 + e^{\epsilon\sqrt{d}/4}} \leq \frac{1}{2}. \qquad (\epsilon > 0)$$

This directly implies that $K\underline{p}^{K-1} \leq 1$ for all $K \geq 1$. Therefore, we can conclude that

$$\mathbb{E}_{\theta_V}^{\pi} \sum_{t=1}^{T} f(C^*, \theta_V) - f(C_t, \theta_V) \geq \min\left\{K\underline{p}^{K-1}, 1\right\} \cdot \mathbb{E}_{\theta_V}^{\pi} \sum_{t=1}^{T} \left(\sigma_1(x_V^\top \theta_V) - \sigma_1(x_{U_t^*}^\top \theta_V)\right).$$

This completes the proof. □

Hence, it is sufficient to establish a lower bound on the regret for logistic bandits. To this end, we introduce the following proposition.

Proposition D.2 (Regret lower bound for MNL bandits, an intermediate result of Theorem in Lee and Oh 13)**.** *Denote M as the maximum assortment size in MNL bandits. Let d be divisible by 4. Suppose $T \geq C \cdot d^4(1+M)/M$ for some constant $C > 0$. Then, in the uniform reward setting, for any policy π, there exists a worst-case problem instance with $N = \Theta(M \cdot 2^d)$ items such that the worst-case expected regret of π is lower bounded as follows:*

$$\frac{1}{|\mathcal{V}_{d/4}|} \sum_{V \in \mathcal{V}_{d/4}} \mathbb{E}^{\pi}_{\theta_V} \left[\mathcal{R}^{MNL}_{\theta_V}(T) \right] = \Omega\left(\frac{\sqrt{M}}{M+1} \cdot d\sqrt{T} \right).$$

By applying Proposition D.2 with the assortment size set to 1, i.e., $M = 1$, we obtain

$$\sup_{\theta} \mathbb{E}^{\pi}_{\theta} \left[\mathcal{R}_{\theta}(T) \right] = \Omega\left(\min\left\{ K\underline{p}^{K-1}, 1 \right\} \cdot d\sqrt{T} \right),$$

which concludes the proof of Theorem 5.7. □

E Tighter Regret Analysis for Cascading Linear Bandits

E.1 Problem Formulation

Apart from the linear feedback model described below, the overall problem formulation remains identical to the cascading logistic bandits setting presented in Section 3. Each arm $i \in \mathcal{I}$ has an associated Bernoulli feedback $y_{t,i} \in \{0, 1\}$ at round t. Given history $\mathcal{H}_t$, we assume that the feedbacks $\{y_{t,i}\}_{i \in \mathcal{I}}$ are conditionally independent and satisfy the linear model:

$$\mathbb{E}[y_{t,i} \mid \mathcal{H}_t] = x_{t,i}^{\top} \theta^*$$

where $\theta^* \in \mathbb{R}^d$ is an unknown time-invariant parameter under the assumption that $\|x_{t,i}\| \leq 1, \forall t \in [T], i \in \mathcal{I}$ and $\|\theta^*\|_2 \leq 1$, as stated in Assumption 5.1.

We further assume that the expected feedback is strictly positive for all t and i , i.e., $x_{t,i}^{\top}\theta^* > 0$, which excludes the case where the probability of receiving positive feedback is exactly zero. This condition excludes only those context vectors lying on the hyperplane orthogonal to θ^*, which forms a measure-zero subset in $\mathbb{R}^d$. This is a mild assumption that holds almost surely when the contexts are drawn from any continuous or non-degenerate distribution. In practical recommendation settings, each item typically has a non-zero probability of receiving positive feedback from users. Hence, the assumption is commonly satisfied in real-world applications, as every item in the recommendation pool is expected to have at least minimal relevance to some users.

In Appendix E, we overload the notation $\sigma_1(z)$ to denote the identity function, i.e., $\sigma_1(z) := z$. Also, we reuse $\sigma_0(z) := 1 - \sigma_1(z)$ for consistency. Additionally, we overload the notations $\bar{p}$ and $\underline{p}$ from Definition 5.2, while noting that their interpretation differs under the linear model. Specifically, since $\sigma_0(z) = 1 - z$, and thus $\bar{p} = \sup_{\|x\| \leq 1}(1 - x^{\top}\theta^*)$ and $\underline{p} = \inf_{\|x\| \leq 1}(1 - x^{\top}\theta^*)$. Under the linear model ($\sigma_1(z) = z$), the expected reward is defined as follows:

$$f_t(C_t, \theta^*) := 1 - \prod_{i \in C_t} (1 - \sigma_1(x_{t,i}^{\top}\theta^*)) = 1 - \prod_{i \in C_t} (1 - x_{t,i}^{\top}\theta^*). \tag{25}$$

E.2 Algorithm for Cascading Linear Bandits

Algorithm 3 `UCB-CLinB`

Input: penalty λ, radius γ_t.
Initialize $\theta_1 \in \Theta$, $V_1 = \lambda I_d$.
for $t = 1, \dots, T$ **do**
  Compute $\{u_{t,i}^{(L)} = x_{t,i}^\top \hat{\theta}_t + \gamma_t \|x_{t,i}\|_{V_t^{-1}}\}_{i \in \mathcal{I}}$.
  Select $C_t' \in \operatorname{argmax}_{C \in \Pi} \tilde{f}_t(C)$ by (5).
  $C_t \leftarrow$ `DO-SWAP`(C_t', θ_t, H_t).
  Play C_t and receive feedback tuple (O_t, Y_t).
  Update $\hat{\theta}_t$ by Equation (26)
  $V_{t+1} \leftarrow V_t + \sum_{i \in O_t} x_{t,i} x_{t,i}^\top$.
end for

Algorithm 4 `OE-SWAP-L`

Input: cascade $C = \{i_1, \dots, i_K\}$, gram matrix V_t.
Find $i_t^{(1)} \in \operatorname{argmax}_{i \in C} \|x_{t,i}\|_{V_t^{-1}}$.
Swap the positions of i_1 and $i_t^{(1)}$.
Output: swapped cascade $(i_t^{(1)}, \dots)$.

In this section, we present the **UCB**-type algorithm for **C**ascading **Lin**ear **B**andits (`UCB-CLinB`), which combines UCB-type exploration with a linear reward model and optimistic exposure swapping technique introduced by Choi et al. [7]. The pseudocode is shown in Algorithm 3 and consists of three core steps.

First, it estimates the unknown parameter θ^* using ℓ_2-regularized least-squares regression based on observed arm–feedback pairs, constructing a confidence set based on the confidence ellipsoid result from Abbasi-Yadkori et al. [1]. (See Appendix E.2.1) Second, using this estimate and confidence set, we defines an upper confidence bound on the expected feedback of each arm and selects a cascade that maximizes the optimistic expected reward (See Appendix E.2.2). Finally, we apply the optimistic exposure swapping technique (See Appendix E.2.3), originally introduced by Choi et al. [7] to deal with the partial feedback in cascading bandits. This strategy promotes the arm with highest uncertainty to the top position in the cascade. While conceptually similar to `DO-SWAP` used in cascading logistic bandits, this technique differs in the criterion for selecting which arm to swap, reflecting the structure of the linear reward model.

E.2.1 Parameter Estimation & Confidence Set for Cascading Linear Bandits

To estimate the unknown parameter θ^*, we use ridge regression on data collected at each round t, $\{(x_{\tau,i}, y_{\tau,i})\}_{i \in [O_\tau], \tau \in [t-1]}$. Leveraging the collected data, we obtain an ℓ_2-regularized least squares estimate of θ^* with regularization parameter $\lambda > 0$:

$$\hat{\theta}_t = \left(X_t^\top X_t + \lambda I\right)^{-1} X_t^\top Y_t, \tag{26}$$

where X_t is the $(\sum_{\tau=1}^{t-1} |O_\tau|) \times d$ matrix whose rows are $x_{\tau,i}^\top$, and Y_t is a column vector with entries $y_{\tau,i}$, for $i \in [O_\tau], \tau \in [t-1]$. Let

$$V_t = \sum_{\tau=1}^{t-1} \sum_{i \in O_\tau} x_{\tau,i} x_{\tau,i}^\top + \lambda I. \tag{27}$$

Note that $V_t \in \mathbb{R}^{d \times d}$ is a symmetric positive definite matrix. We apply Lemma F.1 to the confidence ellipsoid result in Theorem 2 of Abbasi-Yadkori et al. [1], which provides a confidence bound for the ℓ_2-regularized least squares estimate in the linear model. This yields Corollary E.1, which establishes that the true parameter θ^* lies in the confidence set $\mathcal{C}_t^{(L)}(\delta)$ with high probability under the cascading linear bandit setting.

Corollary E.1. *For all $t \in [T]$, with probability at least $1 - \delta$, we have $\theta^* \in \mathcal{C}_t^{(L)}(\delta)$, where*

$$\mathcal{C}_t^{(L)}(\delta) := \{\theta \in \Theta : \|\hat{\theta}_t - \theta\|_{V_t} \le \gamma_t(\delta)\}.$$

Here, $\gamma_t(\delta) = \mathcal{O}(\sqrt{d} \log(tK))$.

E.2.2 Optimistic Expected Reward for Cascading Linear Bandits

Now, we define an upper confidence bound for the expected feedback of any arm $i \in \mathcal{I}$ as

$$u_{t,i}^{(L)} = \min\left\{x_{t,i}^\top \hat{\theta}_t + \gamma_t(\delta)\left\|x_{t,i}\right\|_{V_t^{-1}},\ 1\right\}. \tag{28}$$

Then, we define $\tilde{f}_t^{(L)}(C)$ to be the optimistic expected reward of the cascade C in round t based on $u_{t,i}^{(L)}$:

$$\tilde{f}_t^{(L)}(C) := 1 - \prod_{i \in C} \sigma_0(u_{t,i}^{(L)}). \tag{29}$$

Then, the agent identifies the cascade that maximizes $\tilde{f}_t^{(L)}$ in each round. To ensure that the above exploration bonus approach remains effective, it is crucial to guarantee that $\tilde{f}_t^{(L)}(C)$ is indeed optimistic relative to the true expected reward $f_t(C, \theta^*)$. This requirement is addressed and formalized in Lemma E.2 and Lemma E.3, demonstrating that `UCB-CLinB` maintains the necessary optimism to effectively explore.

Lemma E.2. *Let $\hat{\theta}_t$ be the ℓ_2-regularized least squares estimate as defined in Equation* (26)*. Let $\mathcal{C}_t^{(L)}(\delta)$ be a confidence set with a confidence radius $\gamma_t(\delta)$ in Corollary E.1. Define $u_{t,i}^{(L)} = \min\{x_{t,i}^\top \hat{\theta}_t + \gamma_t(\delta)\left\|x_{t,i}\right\|_{V_t^{-1}},\ 1\}$ for $t \in [T]$ and $i \in \mathcal{I}$. Then, under the event $\{\forall t \geq 1, \theta^* \in \mathcal{C}_t^{(L)}(\delta)\}$, the following holds for all $t \in [T]$ and $i \in \mathcal{I}$:*

$$0 \leq u_{t,i}^{(L)} - x_{t,i}^\top \theta^* \leq 2\gamma_t(\delta)\|x_{t,i}\|_{V_t^{-1}}.$$

Proof of Lemma E.2. We begin by bounding the estimation error as follows:

$$\begin{aligned}
\left|x_{t,i}^\top \hat{\theta}_t - x_{t,i}^\top \theta^*\right| &= \left|x_{t,i}^\top (\hat{\theta}_t - \theta^*)\right| \\
&\overset{\text{(i)}}{\leq} \left\|x_{t,i}\right\|_{V_t^{-1}} \left\|\hat{\theta}_t - \theta^*\right\|_{V_t} \\
&\overset{\text{(ii)}}{\leq} \gamma_t(\delta)\left\|x_{t,i}\right\|_{V_t^{-1}}.
\end{aligned}$$

where inequality (i) follows from Hölder's inequality and inequality (ii) is obtained by applying Corollary E.1. Since $1 - x_{t,i}^\top \theta^* \geq 0$, the following inequality holds:

$$0 \leq x_{t,i}^\top \hat{\theta}_t + \gamma_t(\delta)\left\|x_{t,i}\right\|_{V_t^{-1}} - x_{t,i}^\top \theta^* = u_{t,i}^{(L)} - x_{t,i}^\top \theta^* \leq 2\gamma_t(\delta)\left\|x_{t,i}\right\|_{V_t^{-1}}.$$

□

Lemma E.3. *For any given $C \in \Pi$, the following holds:*

$$f_t(C, \theta^*) \leq \tilde{f}_t^{(L)}(C).$$

Proof of Lemma E.3. Due to the fact that $\sigma_0(z) = 1 - z$ is a monotonically decreasing function of $z \in \mathbb{R}$ and $x_{t,i}^\top \theta^* \leq u_{t,i}^{(L)}$ for $t \in [T], i \in C$ by Lemma E.2, we directly obtain the desired result. □

E.2.3 Optimistic Exposure Swapping for Cascading Linear Bandits

We describe **O**ptimistic **E**xposure **SWAP**ping for cascading **L**inear bandits (`OE-SWAP-L`) which is a technique handling unobserved feedback in cascading linear bandits and first introduced by Choi et al. [7]. The overall procedure of `OE-SWAP-L` is summarized in Algorithm 4. At each round t, the agent employing `OE-SWAP-L` identifies a single arm $i_t^{(1)} \in \operatorname{argmax}_{i \in C}\left\|x_{t,i}\right\|_{V_t^{-1}}$ from the given cascade C, and swaps it with the first-position arm in C.

E.3 Proof of Theorem E.4

Theorem E.4 provides the regret upper bound for `UCB-CLinB`. It demonstrates that `UCB-CLinB` achieves a regret of order $\tilde{\mathcal{O}}(K\bar{p}^{K-1}d\sqrt{T})$. Notably, this bound exhibits the same dependence on the cascade length K and the constant $\bar{p}$ as in the logistic setting, This highlights that the factor $K\bar{p}^{K-1}$, originally derived in the analysis of cascading logistic bandits, also naturally arises in the linear case.

Theorem E.4. *Set the penalty parameter $\lambda \geq K$. Let $\delta \in (0, 1]$. Then, under Assumption 5.1, with probability at least $1-\delta$, `UCB-CLinB` ensures*

$$\mathcal{R}(T) = \tilde{\mathcal{O}}\left(\min\left\{K\bar{p}^{K-1}, 1\right\} d\sqrt{T}\right).$$

Proof. In the following, we assume the good event $\{\forall t \in [T] \text{ and }, \theta^* \in \mathcal{C}_t^{(L)}(\delta)\}$ to hold, which occurs with probability at least $1-\delta$ according to Corollary E.1.

1. Deriving $\mathcal{R}(T) = \tilde{\mathcal{O}}(K\bar{p}^{K-1}d\sqrt{T})$.

The derivation of the regret upper bound up to Equation (7) in Appendix C.1 remains identical for both logistic and linear reward models. We then bound the cumulative regret $\mathcal{R}(T)$ as follows:

$$\begin{aligned}
\mathcal{R}(T) &= \mathbb{E}\left[\sum_{t=1}^{T} f_t(C_t^*, \theta^*) - f_t(C_t, \theta^*)\right] \overset{\text{(i)}}{\leq} \mathbb{E}\left[\sum_{t=1}^{T} \tilde{f}_t^{(L)}(C_t^*) - f_t(C_t, \theta^*)\right] \\
&\leq \mathbb{E}\left[\sum_{t=1}^{T} \tilde{f}_t^{(L)}(C_t) - f_t(C_t, \theta^*)\right] = \mathbb{E}\left[\sum_{t=1}^{T} \prod_{i\in C_t} \sigma_0(x_{t,i}^\top \theta^*) - \prod_{i\in C_t} \sigma_0(u_{t,i}^{(L)})\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T} \sum_{i_{tk}\in C_t} \prod_{l=1}^{k-1} \sigma_0(x_{t,i_{tl}}^\top \theta^*) \left(\sigma_1(u_{t,i_{tk}}^{(L)}) - \sigma_1(x_{t,i_{tk}}^\top \theta^*)\right) \prod_{m=k+1}^{K} \sigma_0(u_{t,i_{tm}}^{(L)})\right] \\
&\leq \mathbb{E}\left[\sum_{t=1}^{T} \sum_{i\in C_t} \prod_{j\in C_t\setminus\{i\}} \sigma_0(x_{t,j}^\top \theta^*) \left(\sigma_1(u_{t,i}^{(L)}) - \sigma_1(x_{t,i}^\top \theta^*)\right)\right] \qquad (30) \\
&\leq \bar{p}^{K-1}\mathbb{E}\left[\sum_{t=1}^{T} \sum_{i\in C_t} \left(\sigma_1(u_{t,i}^{(L)}) - \sigma_1(x_{t,i}^\top \theta^*)\right)\right]
\end{aligned}$$

where inequality is obtained by applying Lemma E.3. This is because the steps rely only on the structure of the expected reward in the cascade model and do not yet involve any assumptions specific to the nonlinearity of the logistic model. Since $\sigma_1(z) = z$ in this section, we have:

$$\begin{aligned}
\mathcal{R}(T) &\leq \bar{p}^{K-1}\mathbb{E}\left[\sum_{t=1}^{T} \sum_{i\in C_t} \left(u_{t,i}^{(L)} - x_{t,i}^\top \theta^*\right)\right] \\
&\overset{\text{(i)}}{=} 2\bar{p}^{K-1}\mathbb{E}\left[\sum_{t=1}^{T} \sum_{i\in C_t} \gamma_t(\delta) \left\|x_{t,i}\right\|_{V_t^{-1}}\right] \\
&\overset{\text{(ii)}}{\leq} 2\bar{p}^{K-1}\gamma_T(\delta)\mathbb{E}\left[\sum_{t=1}^{T} \sum_{i\in C_t} \left\|x_{t,i}\right\|_{V_t^{-1}}\right] \\
&\overset{\text{(iii)}}{\leq} 2K\bar{p}^{K-1}\gamma_T(\delta)\mathbb{E}\left[\sum_{t=1}^{T} \left\|x_{t,i_{t1}}\right\|_{V_t^{-1}}\right] \\
&\overset{\text{(iv)}}{\leq} 2K\bar{p}^{K-1}\gamma_T(\delta)\mathbb{E}\left[\sqrt{T}\sqrt{\sum_{t=1}^{T} \left\|x_{t,i_{t1}}\right\|_{V_t^{-1}}^2}\right] \\
&\overset{\text{(iv)}}{\leq} 2K\bar{p}^{K-1}\gamma_T(\delta)\sqrt{2dT\log\left(1+\frac{T}{d\lambda}\right)} \\
&= \tilde{O}(K\bar{p}^{K-1}d\sqrt{T})
\end{aligned}$$

where inequality (i) is derived by Lemma E.2, inequality (ii) is due to the fact that $\gamma_t(\delta)$ is monotonically increasing with respect to t, inequality (iii) comes from `OE-SWAP-L`, inequality (iv) results from the Cauchy-Schwarz inequality, and inequality (iv) is obtained by applying Lemma F.2 with the identification $n_t = |O_t|$, $z_{t,k} = x_{t,i_{tk}}$, and $Z_t = V_t$.

2. Deriving $\mathcal{R}(T) = \tilde{\mathcal{O}}(d\sqrt{T})$.

Starting from Equation (30), we rederive the regret bound as follows:

$$\begin{aligned}
\mathcal{R}(T) &\leq \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i\in C_t}\prod_{j\in C_t\setminus\{i\}}\sigma_0(x_{t,j}^\top\theta^*)\left(u_{t,i}^{(L)} - x_{t,i}^\top\theta^*\right)\right] \\
&\leq \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i_{tk}\in C_t}\prod_{l=1}^{k-1}\sigma_0(x_{t,i_{tl}}^\top\theta^*)\left(u_{t,i_{tk}}^{(L)} - x_{t,i_{tk}}^\top\theta^*\right)\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i_{tk}\in C_t}\mathbb{E}\left[\prod_{l=1}^{k-1}\sigma_0(x_{t,i_{tl}}^\top\theta^*)\mid\mathcal{H}_t\right]\left(u_{t,i_{tk}}^{(L)} - x_{t,i_{tk}}^\top\theta^*\right)\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i_{tk}\in C_t}\mathbb{E}\left[\mathbb{I}\{i_{tk}\in O_t\}\mid\mathcal{H}_t\right]\left(u_{t,i_{tk}}^{(L)} - x_{t,i_{tk}}^\top\theta^*\right)\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i_{tk}\in C_t}\mathbb{I}\{i_{tk}\in O_t\}\left(u_{t,i_{tk}}^{(L)} - \sigma_1(x_{t,i_{tk}}^\top\theta^*)\right)\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i_{tk}\in O_t}\left(u_{t,i_{tk}}^{(L)} - x_{t,i_{tk}}^\top\theta^*\right)\right] \\
&\leq 2\gamma_T(\delta)\sqrt{2dT\log\left(1+\frac{T}{d\lambda}\right)} = \tilde{\mathcal{O}}(d\sqrt{T})
\end{aligned}$$

where the equalities employ the TPE technique introduced by Liu et al. [20, 21], and the final inequality follows from Lemma E.2, the monotonicity of $\gamma_t(\delta)$, and the application of Lemma F.4 and Lemma F.7. □

F Auxiliary Lemmas

Lemma F.1. *For any $1 \leq \tau \leq t$ and $1 \leq k \leq n_\tau \leq K$, $\|z_{\tau,k}\|_2 \leq 1$. For some $\lambda > 0$, let $Z_t = \sum_{\tau=1}^{t-1}\sum_{k=1}^{n_\tau} z_{\tau,k}z_{\tau,k}^\top + \lambda I_d$. Then,*

$$\det(Z_t) \leq \left(\lambda + \frac{\sum_{\tau=1}^{t-1} n_\tau}{d}\right)^d.$$

In particular, since $n_\tau \leq K$ for all $\tau \in [t]$, we have the simplified bound

$$\det(Z_t) \leq \left(\lambda + \frac{(t-1)K}{d}\right)^d.$$

Proof. We observe that $Z_t = \sum_{\tau=1}^{t-1}\sum_{k=1}^{n_\tau} z_{\tau,k}z_{\tau,k}^\top + \lambda I_d$ can be rewritten in a form analogous to $\sum_s x_s x_s^\top + \lambda I_d$ as stated in Lemma 10 of Abbasi-Yadkori et al. [1] (restated in Lemma F.7) by interpreting the index s as ranging over $\sum_{\tau=1}^{t-1}|n_\tau|$ instances. Hence, Lemma 10 of Abbasi-Yadkori et al. [1] can be directly applied by treating each pair (τ, k) as an independent index. □

Lemma F.2. *Suppose $n_t \in [K]$ for all t. Let $\{z_{t,k}\}_{t\leq\infty,k\in[n_t]}$ be a sequence in $\mathbb{R}^d$ such that $\|z_{t,k}\|_2 \leq 1$ and define $Z_t := \sum_{\tau=1}^{t-1}\sum_{k=1}^{n_\tau} z_{\tau,k}z_{\tau,k}^\top + \lambda I_d$ where $\lambda \geq 1$. Then, we have that*

$$\sum_{t=1}^{T}\|z_{t,1}\|_{Z_t^{-1}}^2 \leq 2d\log\left(1+\frac{T}{d\lambda}\right).$$

Proof. We follow the proof of Lemma 11 of Abbasi-Yadkori et al. [1]. Since $\lambda \geq 1$ and $\|z_{t,k}\|_2 \leq 1$ for all t and k, the following holds:

$$\|z_{t,k}\|^2_{Z_t^{-1}} \leq \frac{\|z_{t,k}\|_2^2}{\lambda_{\min}(Z_t)} \leq \frac{\|z_{t,k}\|_2^2}{\lambda_{\min}(\lambda I_d)} \leq \frac{\|z_{t,k}\|_2^2}{\lambda} \leq 1.$$

Using that $z \leq 2\log(1+z)$ for any $z \in [0,1]$, we have:

$$\begin{aligned}
\sum_{t=1}^{T} \|z_{t,1}\|^2_{Z_t^{-1}} &\leq \sum_{t=1}^{T} \log\left(1 + \|z_{t,1}\|^2_{Z_t^{-1}}\right) \\
&\leq \log \prod_{t=1}^{T} \left(1 + \|z_{t,1}\|^2_{Z_t^{-1}}\right) \\
&\leq \log \prod_{t=1}^{T} \left(1 + \|z_{t,1}\|^2_{\bar{Z}_t^{-1}}\right), \qquad (31)
\end{aligned}$$

where $\bar{Z}_t := \sum_{\tau=1}^{t-1} z_{\tau,1} z_{\tau,1}^\top + \lambda I_d \preceq Z_t$. Also, the following holds:

$$\begin{aligned}
\det(\bar{Z}_t) &= \det\left(\bar{Z}_{t-1} + z_{t,1} z_{t,1}^\top\right) \\
&\overset{\text{(i)}}{=} \det(\bar{Z}_{t-1}) \cdot \det\left(I_d + (\bar{Z}_{t-1}^{-1/2} z_{t,1})(\bar{Z}_{t-1}^{-1/2} z_{t,1})^\top\right) \\
&\overset{\text{(ii)}}{\geq} \det(\bar{Z}_{t-1}) \left(1 + \|z_{t,1}\|^2_{\bar{Z}_{t-1}^{-1}}\right) \\
&\overset{\text{(iii)}}{\geq} \det(\lambda I_d) \prod_{\tau=1}^{t-1} \left(1 + \|z_{\tau,1}\|^2_{\bar{Z}_\tau^{-1}}\right). \qquad (32)
\end{aligned}$$

where inequality (i) is by the fact that $V + U = V^{1/2}(I + V^{-1/2}UV^{-1/2})V^{1/2}$ for a symmetric positive definite matrix V, inequality (ii) is due to Lemma F.6, and inequality (iii) is by repeatedly applying inequality (ii). Then, by applying Equation (32) to Equation (31), we have:

$$\begin{aligned}
\sum_{t=1}^{T} \|z_{t,1}\|^2_{Z_t^{-1}} &\leq \log \frac{\det(\bar{Z}_{T+1})}{\det(\lambda I_d)} \\
&\overset{\text{(i)}}{\leq} 2d \log\left(1 + \frac{T}{d\lambda}\right)
\end{aligned}$$

where inequality (i) is obtained by applying Lemma F.7. □

Lemma F.3. *Suppose $n_t \in [K]$ for all t. Let $\{z_{t,k}\}_{t \leq \infty, k \in [n_t]}$ be a sequence in $\mathbb{R}^d$ such that $\|z_{t,k}\|_2 \leq 1$, and define $Z_t := \sum_{\tau=1}^{t-1} \sum_{k=1}^{n_\tau} z_{\tau,k} z_{\tau,k}^\top + \lambda I_d$ where $\lambda \geq 1$. Then, we obtain*

$$\sum_{t=1}^{T} \mathbb{1}\{n_t \geq 2\} \|z_{t,2}\|^2_{Z_t^{-1}} \leq 2d \log\left(1 + \frac{2T}{d\lambda}\right).$$

Proof. We follow the proof of Lemma 11 of Abbasi-Yadkori et al. [1]. Since $\lambda \geq 1$ and $\|z_{t,k}\|_2 \leq 1$ for all t and k, the following holds:

$$\|z_{t,k}\|^2_{Z_t^{-1}} \leq \frac{\|z_{t,k}\|_2^2}{\lambda_{\min}(Z_t)} \leq \frac{\|z_{t,k}\|_2^2}{\lambda_{\min}(\lambda I_d)} \leq \frac{\|z_{t,k}\|_2^2}{\lambda} \leq 1.$$

Let $\mathcal{T} := \{t \in [T] : k_t \geq 2\}$. Using that $z \leq 2\log(1+z)$ for any $z \in [0,1]$, we have:

$$\begin{aligned}
\sum_{t=1}^{T} \mathbb{1}\{n_t \geq 2\} \|z_{t,2}\|^2_{Z_t^{-1}} &\leq \sum_{t=1}^{T} \log\left(1 + \mathbb{1}\{n_t \geq 2\} \|z_{t,2}\|^2_{Z_t^{-1}}\right) \\
&\leq \sum_{t\in\mathcal{T}} \log\left(1 + \|z_{t,2}\|^2_{Z_t^{-1}}\right) \\
&\leq \log \prod_{t\in\mathcal{T}} \left(1 + \|z_{t,2}\|^2_{Z_t^{-1}}\right) \\
&\leq \log \prod_{t\in\mathcal{T}} \left(1 + \|z_{t,2}\|^2_{\tilde{Z}_t^{-1}}\right), \qquad (33)
\end{aligned}$$

where $\tilde{Z}_t := \sum_{\substack{\tau\in\mathcal{T},\\ \tau<t}} z_{\tau,2} z_{\tau,2}^\top + \lambda I_d \preceq Z_t$. Also, the following holds:

$$\det(\tilde{Z}_t) \geq \det(\lambda I_d) \prod_{\substack{\tau\in\mathcal{T},\\ \tau<t}} \left(1 + \|z_{\tau,1}\|^2_{\bar{Z}_\tau^{-1}}\right), \qquad (34)$$

where Equation (34) follows from applying the same determinant update steps as in Equation (32). Then, by substituting Equation (34) into Equation (33), we have:

$$\begin{aligned}
\sum_{t=1}^{T} \mathbb{1}\{n_t \geq 2\} \|z_{t,2}\|^2_{Z_t^{-1}} &\leq \log \frac{\det(\bar{Z}_{T+1})}{\det(\lambda I_d)} \\
&\overset{\text{(i)}}{\leq} 2d \log\left(1 + \frac{T}{d\lambda}\right)
\end{aligned}$$

where inequality (i) is obtained by applying Lemma F.7. □

Lemma F.4 (Lemma 4.2 of Qin et al. [25]). *For all $t = 1, 2, \ldots, T$, let $S_t \subseteq [N]$ be a subset of size at most K, and define*

$$V_n := \lambda I_d + \sum_{t=1}^{n-1} \sum_{i\in S_t} x_{t,i} x_{t,i}^\top.$$

If $\lambda \geq K$ and $\|x_{t,i}\|_2 \leq 1$ for all t and i, then

$$\sum_{t=1}^{n} \sum_{i\in S_t} \|x_{t,i}\|^2_{V_t^{-1}} \;\leq\; 2\log\det(V_n) - 2\log\det(\lambda I_d)$$

Lemma F.5 (Lemma E.3 from Lee et al. [16]). *Given $x \in \mathbb{R}^d$, for any $\theta_1, \theta_2 \in R^d$,*

$$|\dot{\sigma}_1(x^\top\theta_1) - \dot{\sigma}_1(x^\top\theta_2)| \leq |\sigma_1(x^\top\theta_1) - \sigma_1(x^\top\theta_2)|.$$

Lemma F.6 (Lemma A.3 from Li et al. [19]). *Let $x_i \in \mathbb{R}^{d\times 1}$, $1 \leq i \leq n$. Then we have*

$$\det\left(I + \sum_{i=1}^{n} x_i x_i^\top\right) \geq 1 + \sum_{i=1}^{n} \|x_i\|_2^2.$$

Lemma F.7 (Lemma 10 from Abbasi-Yadkori et al. [1]). *Suppose $x_1, x_2, \ldots, x_n \in \mathbb{R}^d$ and for any $1 \leq i \leq n$, $\|x_i\|_2 \leq 1$. Let*

$$\bar{V}_n = \sum_{i=1}^{n} x_n x_n^\top + \lambda I \quad \textit{for some } \lambda > 0.$$

Then, we have

$$\det(\bar{V}_n) \leq \left(\lambda + \frac{n}{d}\right)^d.$$

G Numerical Study Details

G.2 Experiments on Simulated Data with MovieLens 100K

Dataset. We use the MovieLens $100K$ dataset, which contains $100,000$ ratings (on a 1–5 scale) from 943 users for $1,682$ movies. Since the ratings are not binary, we convert them into binary labels by assigning feedback 1 for ratings ≥ 4 and 0 for ratings ≤ 3. After preprocessing, the dataset contains $55,375$ interactions across 943 users and $1,682$ movies.

Feature extraction. Let $M \in \mathbb{R}^{943\times 1682}$ denote the user–movie rating matrix after binarization. We apply truncated SVD with rank $r = 5$ to obtain $M \approx U\Sigma V^\top$. We treat $U\Sigma \in \mathbb{R}^{943\times 5}$ as the user feature matrix and $V^\top \in \mathbb{R}^{5\times 1682}$ as the movie feature matrix. Let $u_t \in \mathbb{R}^5$ denote the t-th row vector of $U\Sigma$, and $v_i \in \mathbb{R}^5$ denote the i-th column vector of $V^\top$. For each user–movie pair (t, i), we construct the 25-dimensional context vector $x_{t,i} = \mathrm{vec}(u_t \otimes v_i)$, where $\otimes$ denotes the outer product.

Online evaluation. We empirically evaluate the performance of our proposed algorithm, `UCB-CLB`, and compare it with three baselines: `UCB-CCA` [7], `CLogUCB` [21], and `VA-CLogUCB` [21] At each round $t \in [T]$, we sample a user uniformly at random from the 943 users, and the agent selects a movie from a pool of $N = 1682$ candidates. Because the norm of θ^* is large (e.g., $\|\theta^*\|_2 \geq 5$), the condition number κ is small. To evaluate the effect of the cascade length K, we report the cumulative regret under two settings, $K = 5$ and $K = 10$, while keeping all other parameters fixed.

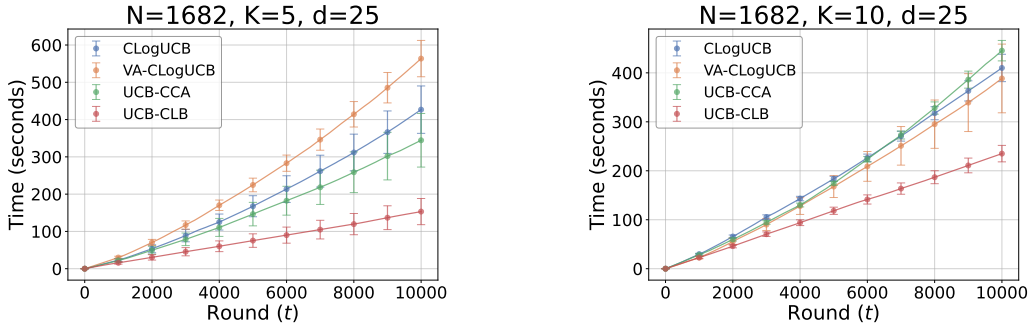

Figure 3: Cumulative running time for varying cascade lengths ($K = 5, 10$) with $N = 1642$ and $d = 25$. Error bars indicate standard error, and all results are averaged over over 5 random seeds.

We evaluate the computational efficiency of the algorithms by plotting the cumulative running time (in seconds) over 10000 rounds for two cascade lengths, $K = 5$ and $K = 10$, with $N = 1682$ and $d = 25$. As shown in the figures, `UCB-CLB` consistently maintains a significantly lower computational cost compared to `UCB-CCA`, `CLogUCB` and `VA-CLogUCB` across different values of K. This efficiency reflects the algorithmic design of `UCB-CLB`, which avoids costly recomputation by leveraging the linear reward structure and an efficient swapping strategy, and employs OMD updates with a per-round cost independent of t. In contrast, `UCB-CCA`, `CLogUCB` and `VA-CLogUCB` rely on MLE updates, whose computational cost increases with each round. Overall, the results demonstrate that `UCB-CLB` achieves strong computational efficiency without sacrificing performance.

Computational resources. All experiments were conducted on a server equipped with an Intel® Xeon® Gold 6526R CPU (16 cores, 2.8GHz, 37.5MB cache, 3UPI, 195W).

H Limitations

While our work revisits contextual cascading bandits under both logistic and linear reward models, these modeling choices come with certain limitations. First, we assume that the feedback for each arm follows either a logistic or linear function of its context. More flexible models could be explored in future work. Second, we adopt the assumption that the feedback from different arms within a cascade is independent, which may not always hold in practice. Finally, in the linear feedback setting, we additionally assume that the expected feedback is strictly greater than zero for all arms and rounds, to ensure that the regret decreases for sufficiently large cascade lengths. While this assumption is mild

in practice, it may not always be satisfied. Nevertheless, we believe that this work offers significant contributions by revisiting the contextual cascading bandit framework and providing improved regret analysis with carefully designed algorithms.

I Broader Impacts

This work revisits contextual cascading bandits with improved theoretical guarantees under logistic and linear models, which can directly support real-world applications such as personalized recommendation, content ranking, and interactive decision-making. In these real-world applications, users are often exposed to long sequences of items—sometimes tens or hundreds—before making a decision. Our algorithms are designed to remain robust and practically effective even in long-sequence settings with partial feedback, which are common in real-world recommendation scenarios. Nevertheless, such systems should be deployed with care, as they may reinforce biases or reduce content diversity.