Online Matrix Completion: A Collaborative Approach with Hott Items

Dheeraj Baby [*1] **Soumyabrata Pal** [*2]

Abstract

We investigate the low rank matrix completion problem in an online setting with M users, N items, T rounds, and an unknown rank-r reward matrix $\mathbf{R} \in \mathbb{R}^{\mathsf{M}\times\mathsf{N}}$. This problem has been well-studied in the literature (Jain & Pal, 2023; Dadkhahi & Negahban, 2018; Sen et al., 2017; Zhou et al., 2020) and has several applications in practice. In each round, we recommend S carefully chosen distinct items to every user and observe noisy rewards. In the regime where $\mathsf{M}, \mathsf{N} \gg \mathsf{T}$, we propose two distinct computationally efficient algorithms for recommending items to users and analyze them under the benign *hott items* assumption 1) First, for $\mathsf{S} = 1$, under additional incoherence/smoothness assumptions on $\mathbf{R}$, we propose the phased algorithm PHASEDCLUSTERELIM. Our algorithm obtains a near-optimal per-user regret of $\widetilde{O}(\mathsf{NM}^{-1}(\Delta^{-1} + \Delta_{\mathsf{hott}}^{-2}))$ where $\Delta_{\mathsf{hott}}, \Delta$ are problem-dependent gap parameters with $\Delta_{\mathsf{hott}} \gg \Delta$ almost always. 2) Second, we consider a simplified setting with $\mathsf{S} = r$ where we make significantly milder assumptions on $\mathbf{R}$. Here, we introduce another phased algorithm, DETERMINANTELIM, to derive a regret guarantee of $\widetilde{O}(\mathsf{NM}^{-1/r}\Delta_{\mathsf{det}}^{-1}))$ where Δ_{det} is another problem-dependent gap. Both algorithms crucially use collaboration among users to jointly eliminate sub-optimal items for groups of users successively in phases, but with distinctive and novel approaches.

1. Introduction

Collaborative Filtering via matrix completion (Koren, 2008) is a fundamental framework for large-scale recommendation systems that cater to millions of users and items jointly. Usually collaborative filtering is an online problem since the recommendation system can update its suggestions based on the history of observed ratings from users. Thus, there is a clear tension between exploration and exploitation - the system needs to learn the user preferences accurately through a diverse set of recommendations and yet recommend meaningful likeable items quickly. Recently, to study the aforementioned exploration/exploitation dilemma theoretically, (Jain & Pal, 2023; Pal et al., 2023) have modelled the problem on lines of online matrix completion that is equivalent to the well-known Multi-Armed Bandit (MAB) optimization (Lattimore & Szepesvári, 2020) for multiple agents (representing users) jointly over a common set of items (representing arms).

We consider the same setting of online matrix completion under bandit feedback with M users, N items, T rounds, and an associated low rank reward matrix $\mathbf{R} \in \mathbb{R}^{\mathsf{M}\times\mathsf{N}}$. This setup was first studied theoretically in (Jain & Pal, 2023) and a close variant in (Sen et al., 2017) - at each round, *one* item is recommended to every user. Note that the recommended item in a particular round can be different for each user and we only observe noisy rewards for the recommended item-user pairs. The main difficulty in obtaining optimal algorithms in this setting is the latent structure which makes it challenging to characterize sharp confidence widths on non-uniform sampling of items. On the other hand, greedy algorithms exploiting the low rank of reward matrix were proposed in (Jain & Pal, 2023),(Sen et al., 2017) that achieved a regret guarantee of $\widetilde{O}(\mathsf{NM}^{-1}\Delta^{-2})$ where Δ is the minimum sub-optimality gap across users and items (Definition 2). Exploiting collaboration among users i.e. aggregating their feedback lead to a significantly smaller effective number of items per user - $O(\mathsf{NM}^{-1})$. However, because of the greedy nature, the achievable regret guarantee is sub-optimal in its dependence on the gap parameter Δ.

To alleviate this issue, an optimal algorithm with a regret guarantee of $\widetilde{O}(\mathsf{NM}^{-1}\Delta^{-1})$ was proposed (Jain & Pal, 2023) only for rank-1 reward matrices. The key idea was that for rank-1 setting, users can be partitioned into two latent clusters - users in the same cluster have the same preference ordering of items while users in different clusters have the exact opposite preference ordering of items. While this property was crucially used for algorithm design in rank-1 case, it does not extend for higher rank. In fact, even

[*]Equal contribution [1]Dept. of Computer Science, UC Santa Barbara, California, United States [2]Adobe, Bangalore, India. Correspondence to: Dheeraj Baby <dheeraj@ucsb.edu>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024. Copyright 2024 by the author(s).

when the reward matrix is rank-2, the possible orderings of items across users is polynomially large in N (Asteris et al., 2014). To get around this, in (Maillard & Mannor, 2014; Pal et al., 2023; Gentile et al., 2014; Lee et al., 2023), the authors provided optimal algorithms by assuming that users can be partitioned into a small number of latent clusters - users in same cluster have identical preferences across items. However, such an assumption can be very restrictive in practice. On the flip side, (Dadkhahi & Negahban, 2018; Zhou et al., 2020) studied the same problem with a weak low rank assumption on the reward matrix - they proposed computationally efficient algorithms but with no theoretical guarantees.

For tractability of analysis, we adopt a relatively weak hott items (aka separability) assumption (A1) widely used in literature - for instance, in non-negative matrix factorization (NMF) (Arora et al., 2012; Recht et al., 2012; Donoho & Stodden, 2003), bandit optimization (Sen et al., 2017; Kveton et al., 2017) and mixture models (Pal & Mazumdar, 2022). The assumption only posits that that the best (most preferred) item of every user can belong to a small set of particular items called *hott* items [1]. Consider a weak notion of latent user clusters based on the best item - users with the same best item belong to same cluster. Our assumption imply that number of such clusters is small. However, unlike the rank-1 setting, users in the same cluster can have vastly different ordering of remaining items. In the seminal work on topic modelling via LDA, (Blei et al., 2003) showed that such an assumption is satisfied approximately by several real-world recommendation datasets.

Techniques and Contributions: In this work, we make two contributions related to online matrix completion for rank > 1 reward matrices - the goal is to improve regret guarantees of greedy algorithms proposed in (Jain & Pal, 2023; Sen et al., 2017).

- First, if number of hott items is small (**A1**) and with additional standard incoherence/smoothness assumptions (**A3**) on reward matrix $\mathbf{R}$ with rank $r > 1$, we propose Alg. PHASEDCLUSTERELIM. Initially, our algorithm finds a set of r distinct *opinionated users* - users who have very strong preferences for a particular hott item over other items. *This is easy since we can expect few users to be very opinionated in most applications.* We consider this set of identified *opinionated users* as labels - the goal is to cluster users based on best item and thus associate the true label [2] with each of the more challenging *non-opinionated* users. Thus, we attempt, in phases, to associate each *non-opinionated* user to at-least one of the labels. In each phase, we assign multiple labels to each *non-opinionated* user encoding the uncertainty. Since there are r opinionated users, the above label association defines r overlapping groups/clusters of users corresponding to each label. For each group of users, we find a common set of *good* items that are high rewarding. Next, we estimate relevant sub-matrices of the reward-matrix from feedback data - based on the estimates, we refine items for each user group. Algorithm PHASEDCLUSTERELIM obtains a *near-optimal* regret guarantee described in Theorem 1. We remark that no prior knowledge of gaps is required to invoke our algorithms.
- Our second contribution is to propose a theoretically sound algorithm under significantly milder assumption of non-negative reward matrix (**A2**) satisfying the hott items property. To this end, following (Kveton et al., 2017), we consider a simplified setting where r items are recommended to every user at each round. The set of items recommended at each round can vary across users. We wish to ensure that there is at-least one item in the recommended set that is high rewarding [3]. The hott items assumption (**A1**) ensures that the unique highest rewarding set of r items that can be commonly recommended to any user is the set of r hott items. Our proposed algorithm DETERMINANTELIM eliminates subsets of r items in successive phases and achieves a regret guarantee described in Theorem 2. For $r = 1$ (rank one reward matrix), our regret guarantees recover the results in (Jain & Pal, 2023) and are thus optimal.

Technical Challenges: Here, we highlight the challenges in analysis of Alg. PHASEDCLUSTERELIM. Recall that at beginning, we find a set of r *opinionated users* with distinct hott items as best item and use them as labels. The *opinionated users* have strong preferences towards their corresponding best item and are thus easy to identify. For labelling *non-opinionated* users, the challenge is that their preferences are more spread out - this makes labelling difficult. Our algorithm, in each phase, assigns the set of *non-opinionated* users into r overlapping groups. There are three key steps 1) First, we show that each group of users (in any phase) is a superset of a cluster of users all having the same best item. 2) Second, we show that in every phase the joint set of good items for each user group (corresponding to a label) is high rewarding for all users in the group 3) Fnally, we show that the best hott item for every user survives in the union of the joint set of good items for the groups that the user belongs to. Therefore, we refine the set of items gradually with each phase (jointly across many users) and obtain our guarantees. Later, in Section 3, we also demonstrate with an illustrative example why simple item elimination

[1] An example of hott items - in a large library of films, only a small number have the highest of ratings provided by users.

[2] To elaborate, true label corresponds to the *opinionated* user sharing the same best (hott) item as the non-opinionated user.

[3] This setting models situations where a handful of movies are recommended by a streaming website and the user is satisfied if at-least one of the recommendations is of interest to the user

strategies fail - necessitating our novel approach.

For the analysis of Alg. DETERMINANTELIM, we use the fact that the hott items are, in some sense, the unique highest rewarding set of r items that can be recommended to any user. Hence, we track subsets of r items and eliminate them succesively to converge to the set of r hott items. This was also observed in (Kveton et al., 2017) where a similar problem was studied with the goal of finding the largest entry of a low rank matrix in an online fashion. However, in their setting, it was possible to choose both row and column of reward matrix in each round. However, we need to recommend to every user r items at each round. *This is challenging because the difficulty of identifying preferred items can vary significantly across users.* Here, we jointly estimate how *good* each group of r items is and eliminate sub-optimal groups of r-items based on the estimates - these estimates are based on determinant of relevant sub-matrices. Our key idea (Lemma 13) is to exploit redundancy in recommending overlapping sets of r items at every round and characterize the sufficient rounds in each phase for ensuring nice concentration properties of our estimates (Lemma 14). Due to space restrictions, a description of other loosely related work has been delegated to Appendix.

2. Problem Formulation and Results

2.1. Formulation

Notations: We write $[m]$ to denote the set $\{1, 2, \ldots, m\}$. $\mathbf{M}_i$ denotes the ith row of $\mathbf{M}$ and $\mathbf{M}_{ij}$ or $\mathbf{M}_{i,j}$ denotes the (i, j)-th element of matrix $\mathbf{M}$. For any set $\mathcal{U} \subset [m], \mathcal{V} \subset [n]$, $\mathbf{M}_{\mathcal{U},\mathcal{V}}$ denotes the matrix $\mathbf{M}$ restricted to the rows in $\mathcal{U}$ and columns in $\mathcal{V}$. Similarly, $\mathbf{M}_{\mathcal{U}}$ denotes the matrix $\mathbf{M}$ restricted to the rows in $\mathcal{U}$. We write $\det(\mathbf{M})$ to denote the determinant of a matrix $\mathbf{M}$. We will assume that the elements in sets $\mathcal{I}, \mathcal{J}$ will be ordered in ascending order. We will refer as $d-$row (analogously $d-$ column) a set of d distinct rows (columns respectively) of a matrix. We denote the set $\mathcal{S}_d \equiv \{\mathbf{v} \in [0, 1]^d \mid ||\mathbf{v}||_1 \leq 1\}$ to be the d-dimensional simplex; $\mathcal{S}_{n,d}$ denotes the set of $n \times d$ matrices whose rows belong to $\mathcal{S}_d$ i.e. $\mathcal{S}_{n,d} = \{\mathbf{M} \in [0, 1]^{n\times d} \mid \mathbf{M}_i \in \mathcal{S}_d\}$. Let $\Pi_k(\mathcal{A})$ denote the set of all k-subsets of set $\mathcal{A}$ with distinct elements. We use $\mathbb{E}X$ to denote the expectation of a random variable X. $\widetilde{O}(\cdot)$ hides logarithmic dependencies on $\mathsf{M}, \mathsf{N}, \mathsf{T}$. Notations such as $\widetilde{\mathbf{Q}}$ denotes an estimate of a quantity $\mathbf{Q}$.

Consider a set of M users, N items and a horizon of T rounds. We have an associated unknown rank-r expected reward matrix $\mathbf{R} \in [-1, 1]^{\mathsf{M}\times\mathsf{N}}$ such that $\mathbf{R} = \mathbf{U}\mathbf{V}^\mathsf{T}$ where $\mathbf{U} \in \mathbb{R}^{\mathsf{M}\times r}$, $\mathbf{V} \in \mathbb{R}^{\mathsf{N}\times r}$ denotes the user embedding matrix and item embedding matrix respectively. At each round $t \in [\mathsf{T}]$, for each user $u \in [\mathsf{M}]$, S carefully chosen distinct items $\{\rho_u(t, j)\}_{j=1}^{\mathsf{S}} \in [\mathsf{N}]$ are recommended and we observe S rewards denoted by the random variables $\{\mathbf{P}^{(t)}_{u\rho_u(t,j)}\}_{j=1}^{\mathsf{S}}$ such that

$$\mathbf{P}^{(t)}_{u\rho_u(t,j)} = \mathbf{R}_{u\rho_u(t,j)} + \mathbf{E}^{(t)}_{u\rho_u(t,j)}. \tag{1}$$

Here $\mathbf{E}^{(t)}_{u\rho_u(t,j)}$ denotes the additive noise that are i.i.d zero-mean sub-Gaussian random variables with variance proxy at most σ^2. We consider the practical regime where $\mathsf{M}, \mathsf{N} \gg \mathsf{T}$ i.e. the number of users and items are much larger than the number of rounds. Let ordering $\pi_u : [\mathsf{N}] \to [\mathsf{N}]$ sort the rewards for each user $u \in [\mathsf{M}]$ in descending order, i.e., for any $i < j$, $\mathbf{R}_{u\pi_u(i)} \geq \mathbf{R}_{u\pi_u(j)}$. We make the following assumption:

Assumption 1 (**A1** Hott items and vectors)**.** *We assume that there exists a set of r unknown distinct ordered indices $\mathcal{A} \subseteq [\mathsf{N}], |\mathcal{A}| = r$ such that all the item embedding vectors $\{\mathbf{V}_i\}_{i\in[\mathsf{N}]}$ lie within the convex hull of $\{\mathbf{V}_i\}_{i\in\mathcal{A}} \cup \{\mathbf{0}\}$. We will call the items corresponding to indices in $\mathcal{A}$ hott items and the rows in $\mathbf{V}_\mathcal{A}$ to be hott vectors.*

As mentioned before, Assumption 1, also sometimes known as separability assumption in existing literature has been used widely. In fact, as pointed out in (Donoho & Stodden, 2003; Arora et al., 2012), an approximate separability condition is regarded as a fairly benign assumption that holds in many practical contexts in machine learning such as LDA in information retrieval (Blei et al., 2003). We first show that with Assumption 1, for each user $u \in [\mathsf{M}]$, the highest and lowest rewarding items $\pi_u(1), \pi_u(\mathsf{N}) \in \mathcal{A}$ are hott items.

Lemma 1. *For any $u \in [\mathsf{M}]$, it must happen that $\pi_u(1), \pi_u(\mathsf{N}) \in \mathcal{A}$.*

Thus, we can partition the users $\mathcal{T}^{(1)}, \mathcal{T}^{(2)}, \ldots, \mathcal{T}^{(r)}$ into r clusters such that $\forall i \in [r], \mathcal{T}^{(i)} \triangleq \{u \in [\mathsf{M}] \mid \pi_u(1) = \mathcal{A}_i\}$ is the set of users whose best item is the ith hott item. For some κ, we can write the minimum cluster size to be $\min_{i\in[r]} \left|\mathcal{T}^{(i)}\right| \geq \kappa\mathsf{M}r^{-1}$. We will now consider two distinct (although related) notions of regret under different additional assumptions on the reward matrix $\mathbf{R}$:

Setting 1 (General Regret): We consider $\mathsf{S} = 1$ i.e. at every round $t \in [\mathsf{T}]$, for each user $u \in [\mathsf{M}]$, a single item $\rho_u(t)$ is recommended to user u. The objective is to design an algorithm (possibly randomized) for the recommendation policy so that the expected regret defined below is minimized:

$$\mathsf{Reg} \triangleq \frac{1}{\mathsf{M}}\Big(\sum_{t\in[\mathsf{T}]}\sum_{u\in[\mathsf{M}]} \mathbf{R}_{u\pi_u(1)} - \sum_{t\in[\mathsf{T}]}\sum_{u\in[\mathsf{M}]} \mathbf{R}_{u\rho_u(t)}\Big) \tag{2}$$

We make the following additional assumptions on the expected reward matrix $\mathbf{R}$:

Assumption 2 (**A2**)**.** *We will assume r, σ, κ are positive constants that do not scale with $\mathsf{M}, \mathsf{N}, \mathsf{T}$.*

Note that Assumption 2 is made only for simplicity of exposition/analysis and is not necessary for our theoretical guarantees or the algorithm. It is routine to generalize our results for a non-constant r, σ, κ and leads to a polynomial dependence on the parameters in the regret guarantee.

Assumption 3 (**A3** Assumptions on reward matrix $\mathbf{R}$)**.** *We assume that $\mathbf{R}$ with SVD decomposition $\mathbf{R} = \widetilde{\mathbf{U}}\mathbf{\Sigma}\widetilde{\mathbf{V}}^{\mathsf{T}}$ satisfies the following properties 1) (Condition Number) $\mathbf{R}$ has rank r and has non zero singular values $\lambda_1 > \lambda_2 > \cdots > \lambda_r$ with $\lambda_1/\lambda_r = O(1)$ 2) (μ-incoherence) $\left|\left|\widetilde{\mathbf{U}}\right|\right|_{2,\infty} \leq \sqrt{\mu r/\mathsf{N}}$ and $\left|\left|\widetilde{\mathbf{V}}\right|\right|_{2,\infty} \leq \sqrt{\mu r/\mathsf{M}}$ for some $\mu = O(1)$. 3) (Subset Strong Smoothness (a)) For some constant $\beta > 0$ and for any subset of indices $\mathcal{S} \subseteq [\mathsf{N}], \mathcal{S} = \mathcal{T}^{(j)}$ (corresponding to some cluster of users), we must have $\mathbf{x}^{\mathsf{T}}\widetilde{\mathbf{U}}_{\mathcal{S}}^{\mathsf{T}}\widetilde{\mathbf{U}}_{\mathcal{S}}\mathbf{x} \geq \beta$ for all unit norm vectors $\mathbf{x} \in \mathbb{R}^r$. 4) (Subset Strong Smoothness (b)) For some α satisfying $\alpha \log \mathsf{M} = \Omega(1)$, $\gamma = \widetilde{O}(1)$, for any subset of indices $\mathcal{S} \subseteq [\mathsf{M}], |\mathcal{S}| \geq \gamma r$, we must have $\mathbf{x}^{\mathsf{T}}\widetilde{\mathbf{V}}_{\mathcal{S}}^{\mathsf{T}}\widetilde{\mathbf{V}}_{\mathcal{S}}\mathbf{x} \geq \alpha\, |\mathcal{S}|\, /\mathsf{M}$ for all $\mathbf{x} \in \mathbb{R}^r, ||\mathbf{x}||_2 = 1$.*

Assumption 3, also proposed in (Pal et al., 2023) states that matrix $\mathbf{R}$ must be well-conditioned and incoherent - two properties that are commonly used for invoking theoretical guarantees on offline low rank matrix completion (Chen et al., 2019; Jain et al., 2013). In essence, Assumption 3 allows us to invoke low rank matrix completion guarantees to suit large sub-matrices of $\mathbf{R}$. The subset strong smoothness assumptions on $\mathbf{U}, \mathbf{V}$ suffice to prove that any sub-matrix of $\mathbf{R}$ of a reasonable size also satisfies the low condition number and incoherence properties (see (Pal et al., 2023)). Now, we define some problem-dependent gap parameters:

Definition 1 (Hott item gaps)**.** *Let $\mathcal{A}$ be the set of hott items. For any user $u \in [\mathsf{M}]$, define the hott item gap $\Delta_u \triangleq \mathbf{R}_{u\pi_u(1)} - \max_{v \in \mathcal{A} \setminus \pi_u(1)} \mathbf{R}_{uv}$ to be the minimum gap for user u between the best item and any other hott item. For each $i \in [r]$, we will call the user $u^\star(i) \triangleq \mathsf{argmax}_{u \in \mathcal{T}^{(i)}} \Delta_u$ with the largest hott item gap to be the most opinionated user in the ith cluster. Now, define $\Delta_{\mathsf{hott}} \triangleq \min_{i \in [r]} \Delta_{u^\star(i)}$ to be hott item gap minimized across the most opinionated users across each cluster.*

Definition 2 (Minimum Reward gap)**.** *Define $\Delta \triangleq \min_{u \in [\mathsf{M}]} \mathbf{R}_{u\pi_u(1)} - \mathbf{R}_{u\pi_u(2)}$ to be the smallest gap between the best item and second best item minimized across all users.*

In most settings, the hott item gap Δ_{hott} is significantly larger than Δ – the hott item gap needs to be satisfied by only a *single* user from every latent cluster. To make this clear, consider the illustrative reward matrix in equation (4) in Sec. 3 - for this matrix, $\Delta_{\mathsf{hott}} = 1$ and $\Delta = 2\epsilon/3$ from definitions - the latter quantity can be made as small as desired by setting ϵ appropriately. We are now ready to present our first main result. All proofs are deferred to Appendix C.

2.2. Main Results

Theorem 1. *Consider the online matrix completion problem with M users, N items, T rounds such that at round $t \in [\mathsf{T}]$, we observe reward $\{\mathbf{P}^{(t)}_{u\rho_u(t)}\}_{u \in [\mathsf{M}]}$ as in eq. ((1)) with $\mathsf{S} = 1$ (i.e only one item recommended per user). Let $\mathbf{R} \in \mathbb{R}^{\mathsf{M} \times \mathsf{N}}$ be the expected rank-r reward matrix. Suppose Assumptions 1, 2, 3 are true. Then Alg.1 (*PHASEDCLUSTERELIM*) guarantees the regret Reg (eq. (2)) to be $\widetilde{O}\Big(\max\left(1, \mathsf{N}\mathsf{M}^{-1}\right)\left(\Delta_{\mathsf{hott}}^{-2} + \Delta^{-1}\right) + 1\Big)$.*

Importantly, the regret guarantee above depends on the gap parameters $\Delta, \Delta_{\mathsf{hott}}$ - the dependence on number of rounds T is poly-logarithmic and is hidden within $\widetilde{O}(\cdot)$ notation [4] For readability, we have also hidden the polynomial dependence on r, σ, κ that are assumed to be constants (**A2**).

The regret guarantee in Theorem 1 has an optimal linear dependence on $\mathsf{N}\mathsf{M}^{-1}, \Delta^{-1}$ and a quadratic dependence on $\Delta_{\mathsf{hott}}^{-1}$; it is significantly better than the Δ^{-2} dependence achieved by the greedy algorithm proposed in (Jain & Pal, 2023). Again, we stress that the hott item gap Δ_{hott} is significantly larger than the minimum reward gap Δ from definition. While the quadratic dependence on $\Delta_{\mathsf{hott}}^{-1}$ is unwanted, Δ_{hott} being large should make its contribution to the regret small in practice.

Remark 1. *The optimality of the dependence on $\mathsf{N}, \mathsf{M}, \Delta$ can be observed from a simple example. Suppose all users are equivalent - hence, rows of $\mathbf{R}$ are identical and in a single round, users come in a sequential fashion too. This is strictly easier and equivalent to a multi-armed bandit (MAB) problem with N items and MN rounds up to normalization. Standard MAB literature imply that asymptotic lower bound for our setting is $\widetilde{O}(\mathsf{N}\mathsf{M}^{-1}\Delta^{-1})$ - see(Lattimore & Szepesvári, 2020), Chapter 16.*

Remark 2. *Note that our guarantees in Theorem 1 can also be extended easily to the setting where the number of hott vectors is $t > r$. In the extreme case, when all the item embedding vectors lie on the surface of a sphere, the number of hott vectors is same as total number of items. However, if number of hott vectors is smaller, then we have non-trivial regret guarantees. Hence, we can also think of our result as a more fine-grained parameterization in terms of the number of hott vectors which can be small in several practical applications.*

Setting 2 (Simplified Regret): Next, we consider a related

[4] Gap-dependent regret bounds with logarithmic dependence on T are stronger than gap-free worst case regret bounds with fractional power dependence on T. (see for example (Lattimore & Szepesvári, 2020), Theorem 7.1). Usually, we can convert gap-dependent regret bounds to worst-case regret bounds by setting the gap parameter appropriately as function of N, T.

setting with $\mathsf{S} = r$ i.e. at each round $t \in [\mathsf{T}]$, for each user $u \in [\mathsf{M}]$, r distinct items $\rho_u(t,1), \ldots, \rho_u(t,r) \in [\mathsf{N}]$ are recommended and we observe the corresponding noisy rewards. Now, instead of Assumption 3, we make the following milder assumption (similar to (Kveton et al., 2017)) that the latent embeddings of users and items are non-negative:

Assumption 4. *The user (item) embedding matrix* $\mathbf{U}$ *(*$\mathbf{V}$*) belongs to the set* $\mathcal{S}_{\mathsf{M},d}$ *(*$\mathcal{S}_{\mathsf{N},d}$*).*

In this setting, we consider the simplified regret:

$$\mathsf{Reg}_{\mathsf{Simple}} \triangleq \frac{1}{\mathsf{M}}\Big(\sum_{u \in [\mathsf{M}]} \mathsf{T}\mathbf{R}_{u\pi_u(1)} - \sum_{\substack{t \in [\mathsf{T}] \\ u \in [\mathsf{M}]}} \max_{j \in [\mathsf{r}]} \mathbf{R}_{u\rho_u(t,j)} \Big). \tag{3}$$

Note that at each round t, the simplified regret for a user u is small if one of the recommended items $\{\rho_u(t,j)\}_{j \in [r]}$ is close to the best item $\pi_u(1)$. In general the simplified notion of regret (eq. (3)) is easier to optimize than the general regret (eq. (2)) - more precisely, if there exists an algorithm guaranteeing the regret $\mathsf{Reg} = O(h)$ for some h that is a function of $\mathsf{M}, \mathsf{N}, \mathsf{T}$ by recommending items with $\mathsf{S} = 1$, then the same algorithm can also guarantee $\mathsf{Reg}_{\mathsf{Simple}} = O(h)$ with $\mathsf{S} = r$. However, as shown below, our theoretical guarantees for $\mathsf{Reg}_{\mathsf{Simple}}$ hold under significantly milder assumptions than that for Reg. Due to Assumption 1, we can conclude that the squared determinant $\mathsf{det}^2(\mathbf{V}_{\mathcal{A}})$ of the item embedding matrix restricted to the hott items is larger than $\mathsf{det}^2(\mathbf{V}_{\mathcal{A}'})$ for any $\mathcal{A}' \neq \mathcal{A}, |\mathcal{A}'| = r$ (see (Kveton et al., 2017)). Thus, we can define the following:

Definition 3 (Determinant gap)**.** *We define the determinant sub-optimality gap* Δ_{det} *as follows:* $\Delta_{\mathsf{det}} = \mathsf{det}^2(\mathbf{V}_{\mathcal{A}}) - \max_{\mathcal{A}' \neq \mathcal{A}, |\mathcal{A}'|=r} \mathsf{det}^2(\mathbf{V}_{\mathcal{A}'})$.

We now state our second main result:

Theorem 2. *Consider the online matrix completion problem with* M *users,* N *items* T *rounds such that at round* $t \in [\mathsf{T}]$*, we observe noisy reward* $\{\mathbf{P}^{(t)}_{u\rho_u(t,j)}\}_{u \in [\mathsf{M}], j \in [r]}$ *as in eq. ((1)) with* $\mathsf{S} = r$*, noise variance proxy* $\sigma^2 > 0$*. Let* $\mathbf{R} \in \mathbb{R}^{\mathsf{M} \times \mathsf{N}}$ *be the expected rank-r reward matrix. Suppose Assumptions 1, 4 are true. Then Alg.3 (*DETERMINANTELIM*) guarantees regret* $\mathsf{Reg}_{\mathsf{Simple}}$ *(eq. (3)) to be* $\widetilde{O}\Big(\frac{\sigma^2 \mathsf{N}}{c_{\mathsf{avg}} c_{\max} \mathsf{M}^{1/r} \Delta_{\mathsf{det}}} + 1\Big)$ *where* $c_{\mathsf{avg}} \triangleq \binom{\mathsf{M}}{r}^{-1} \sum_{\mathcal{I} \subset [\mathsf{M}] || \mathcal{I}|=r} \mathsf{det}^2(\mathbf{U}_{\mathcal{I}})$ *and* $c_{\max} \triangleq \mathsf{det}^2(\mathbf{V}_{\mathcal{A}})$.

As in Theorem 1, the poly-logarithmic dependence on T and the poly dependence on r, σ, κ (assumed to be constants) are subsumed within $\widetilde{O}(\cdot)$. The regret guarantee in Theorem 2 has a linear dependence on the inverse of determinant gap $\Delta_{\mathsf{det}}^{-1}$. For rank-$r > 1$, we do suffer a cost in form of the factor $\mathsf{M}^{-1/r}$ in the regret. For $r = 1$ (rank-1), the two proposed notions of regret are equivalent and the guarantee in Theorem 2 recovers the guarantees in (Jain & Pal, 2023) for rank-1 setting.

Remark 3. *Note that our proposed algorithms are computationally efficient with run-times having polynomial dependence on* M, N*. Their run-times do have an exponential dependence on the rank-r and hence, the proposed algorithms are computationally feasible for small values of rank r. Our poly dependence on* M, N *stems from the use of low rank matrix completion algorithms as a sub-module (see Step 17 in Algorithm 1) - the convex relaxation based approach used crucially for strong theoretical guarantees is computationally less efficient for matrix completion. However, in practice, a number of highly efficient techniques have been proposed for optimizing matrix completion - for example, see (Recht & Ré, 2013; Teflioudi et al., 2012). Substituting such estimators for invoking matrix completion can allow us to scale significantly.*

3. Algorithm PHASEDCLUSTERELIM and Analysis of Reg (Theorem1)

Due to space restrictions, certain modules of algorithm are delegated to Appendix C. Below, we provide an overview of the main components of Alg. PHASEDCLUSTERELIM (see Algorithm 1) and their analysis. The algorithm is divided into two distinct stages, each of which runs in phases of exponentially increasing length. Note that we ensure every user $u \in [\mathsf{M}]$ is recommended one item at each round in the `ESTIMATE`.

Stage 1 (Identifying a set of *opinionated users* via independent set): Our first key step is to identify a set of r users such that any pair among them clearly do not share the same best item. We find these set of r users in the following way: for the first few rounds, we explore i.e. for every user, we recommend randomly sampled items (See Remark 4) via the call to `ESTIMATE` and store the corresponding noisy observations (Line 3). The rows and columns of the reward matrix is given by the first and second arguments of `ESTIMATE` respectively (which are set equal to $[\mathsf{M}]$ and $[\mathsf{N}]$ in Line 3). At this point, we have a partially observed (noisy) reward matrix. Due to its low rank, we can invoke guarantees of well-known low rank matrix completion estimators that allow us to compute an estimate of the entire reward matrix from the few observed entries with an error guarantee (equal to k_ℓ, see Line 3) on each entry. The task of matrix completion is also carried out within `ESTIMATE`. Based on the estimated rewards, we maintain a candidate item set for each user (denoted by $\mathcal{T}_u^{\ell+1}$ at phase ℓ, Line 4). This item set is used to construct a graph with users as nodes such that two users are connected iff their candidate item sets have a non-empty intersection. The idea of the graph is to cluster similar users (in terms of their latent embeddings) together. As the phase number increases, we

Algorithm 1 PHASEDCLUSTERELIM (Iterative Multi-Label User Classification and Joint Item Elimination)

1: Initialize G to be a fully connected graph of M users. Set $\mathcal{T}_u^0 = [\mathsf{N}]$ for all users $u \in [\mathsf{M}]$, phase number $\ell = 0$, $k_0 = 1$ and $c_0 = 3k_0$.
// Stage 1: Identifying a set of *opinionated users* via independent set
2: **while** `CheckIndependentSet`(G) returns False **do**
3: $\tilde{\mathbf{R}}^{\ell+1} \leftarrow$ `ESTIMATE`$([\mathsf{M}], [\mathsf{N}], k_\ell)$
4: Update the item set for each user as $\mathcal{T}_u^{\ell+1} = \{x \in \mathcal{T}_u^\ell : \tilde{\mathbf{R}}^{\ell+1}_{u,\tilde{\pi}_u(1)} - \tilde{\mathbf{R}}^{\ell+1}_{u,x} \leq c_\ell\}$
5: Update the graph G such that two users u and v are connected iff $\mathcal{T}_u^{\ell+1} \cap \mathcal{T}_v^{\ell+1} \neq \Phi$; Set $\ell \leftarrow \ell + 1$, $k_\ell \leftarrow k_{\ell-1}/2$ and $c_\ell = 3k_\ell$
6: **end while**
7: Let $u_1, \ldots, u_r$ be the users returned by the last call to `CheckIndependentSet`(G)
8: Set $k_\ell \leftarrow k_{\ell-1}/10$ and $c_\ell = 3k_\ell$.
9: $\tilde{\mathbf{R}}^{\ell+1} \leftarrow$ `ESTIMATE_SIMPLE`$([\mathsf{M}], [\mathsf{N}], k_\ell)$
// Stage 2: Joint Elimination of items via expansion and intersection
10: **while** time horizon of the game is not reached **do**
11: Update the item set for each user as $\mathcal{T}_u^{\ell+1} = \{x \in \mathcal{T}_u^\ell : \tilde{\mathbf{R}}^{\ell+1}_{u,\tilde{\pi}_u(1)} - \tilde{\mathbf{R}}^{\ell+1}_{u,x} \leq c_\ell\}$
12: Update the graph G such that two users u and v are connected iff $\mathcal{T}_u^{\ell+1} \cap \mathcal{T}_v^{\ell+1} \neq \Phi$; Set $\ell \leftarrow \ell + 1$; $k_\ell \leftarrow k_{\ell-1}/2$ and $c_\ell = 3k_\ell$
13: Update candidate item sets by calling `ExpanditemSets`$(G, (u_1, \ldots, u_r), c_{\ell-1}, \tilde{\mathbf{R}}^\ell)$
14: Let $\mathcal{T}_{C_i}^\ell$ be the common good item set for all users that are connected to user u_i
15: **for** each cluster $i \in [r]$ **do**
16: $\mathcal{U} \leftarrow \{u : \mathcal{T}_u^\ell \cap \mathcal{T}_{C_i}^\ell \neq \Phi\}$ // Recommend items to users in $\mathcal{U}$ by calling ESTIMATE as below
17: $\tilde{\mathbf{R}}^{\ell+1}_{\mathcal{U},\mathcal{T}_{C_i}^\ell} \leftarrow$ `ESTIMATE`$(\mathcal{U}, \mathcal{T}_{C_i}^\ell, k_\ell)$.
//ESTIMATE collects data for $O(k_\ell^{-2} \max(1, \mathsf{NM}^{-1}))$ rounds and estimates matrix $\mathbf{R}_{\mathcal{U},\mathcal{T}_{C_i}^\ell}$
18: **end for**
19: **end while**

get finer estimates of the reward matrix (the argument k_ℓ in the call to `ESTIMATE` controls the estimation accuracy level) . Consequently the size of the candidate item sets become small and many of the nodes in the graph get disconnected, thereby surfacing out similar users. We prove that after $\log(1/\Delta_{\text{hott}})$ phases, an independent set of size r arises in the graph, where Δ_{hott} is as in Definition 1. Since the independent set has size r, there must exist at-least r users that are mutually disconnected in the graph. We call such users *opinionated* users since it can be shown that their hott-item gap is at-least Δ_{hott} (see Defn. 1). We prove that at every phase, the best item for *any* user belongs to their candidate item sets with high probability. Consequently, the set of r opinionated users must have distinct best items.

Lines 8-9 are necessary for technical reasons to ensure that if an item is not in a user's candidate item set, its reward for that user is sufficiently lower than the reward of the best item.

Remark 4. *The recommendation of items to users is being done inside the* `ESTIMATE_SIMPLE` *function (Algorithm 2) in L17 of Algorithm* PHASEDCLUSTERELIM. *On invoking* `ESTIMATE_SIMPLE` *function , the algorithm recommends randomly sampled items from the input subset of items in the second argument of the function to the input subset of users in the first argument (the third argument determine the number of rounds for which to recommend). To invoke the theoretical guarantees of offline low rank matrix completion (Chen et al., 2019), the recommendation strategy needs to be slightly modified (see Sec. 3 in (Jain & Pal, 2023)) to account for our setting. Furthermore, the theoretical estimator used is the convex relaxation approach wherein we minimize the MSE with a nuclear norm regularizer - for details, see the more sophisticated Algorithm 7 (*`ESTIMATE`*) in Appendix, for which, sharp theoretical guarantees can be invoked.*

After identifying r opinionated users, our strategy is to utilize the opinionated users as labels for r groups i.e. if we say that a user $u \in [\mathsf{M}]$ has similarity with a subset of opinionated users, then the best item of u must be among the best items (hott items) of the aforementioned subset of opinionated users. The correct label for each user is determined by the opinionated user who shares the same best item. If a user is connected to multiple opinionated users, we assign multiple labels to capture the uncertainty in their best item. At most, we assign r labels per user, but we aim to reduce the number of labels as we gather more information. This brings us to the second stage of our algorithm:

Stage 2 (Joint Elimination of Items via Expansion and Intersection): Our strategy for refining the labels is to explore carefully while exploiting the user-user and item-item similarities learned from the previous phase. User-user similarity is leveraged by forming groups of users with the same assigned label (for non-opinionated user, this means that the label under consideration is present in its multi-label set). There can be r such (possibly non-disjoint) groups. For concreteness, we restrict our attention to a group corresponding to label u_i hereof, where u_i is an opinionated user

with $\mathcal{A}_i$ as its best hott-item (see Assumption 1). Thus all users in this group have connection to the user u_i. Such a group corresponding to a label u_i is stored in $\mathcal{U}$ in Line 16 of Algorithm 1. Due to our graph design, the set of all users whose best item is $\mathcal{A}_i$ must be a *subset* of $\mathcal{U}$. To exploit the item-item similarity we can take the intersection of the candidate item sets of the users in a group (see Line 14). This set of items must be high rewarding for all users within that group. However there can be multiple users (for eg. the user u_i) in $\mathcal{U}$ with best item being $\mathcal{A}_i$. So the intersection can potentially eliminate the best item $\mathcal{A}_i$. This can happen because of the subset condition noted before, there can also be users in $\mathcal{U}$ that do not have $\mathcal{A}_i$ in its candidate item set. Specifically, a non-opinionated user u with its best item not equal to $\mathcal{A}_i$ can be connected to u_i (or equivalently has a label u_i) due to some $x \in \mathcal{T}_{u_i}^{\ell} \cap \mathcal{T}_u^{\ell}$ with $x \neq \mathcal{A}_i$. In order to resolve this issue, we expand (Line 13) the candidate item set for each non-opinionated user slightly so that provably: 1) the expanded set will contain the true best item (hott item) of the user u and the item $\mathcal{A}_i$ and 2) the set of labels assigned to user u is same as that before the expansion. As mentioned before, best item for *any* user belongs to their candidate item sets with high probability. Subsequently, taking intersection of the candidate item sets of all users with label u_i assigned to them (denoted by $\mathcal{T}_{C_i}^{\ell}$ in Line 14) will ensure that the item $\mathcal{A}_i$ will not be eliminated from $\mathcal{T}_{C_i}^{\ell}$. So we can randomly explore within the sub-matrix defined by the rows $\mathcal{U}$ and columns $\mathcal{T}_{C_i}^{\ell}$ via the call to `ESTIMATE` in Line 17 with a finer estimation error k_ℓ. This ensures that the regret due to the exploration is low enough and at the same time the best item for each user is not eliminated when updating the candidate item sets via the refined estimates of reward (sub) matrix in Line 11. The process is repeated across phases, increasing confidence in user similarity and rewarding items. This is reflected in the algorithm by a decrease in multi-labelled users as well as a shrinkage of candidate item set sizes.

Regret Analysis: Stage 1 of our algorithm is greedy i.e. we incur worst-case regret until we have found a set of r *opinionated users*. However, we only need to find one *opinionated user* per hott item and therefore the regret for Stage 1 scales as $\widetilde{O}(\max(1, \mathsf{N}\mathsf{M}^{-1})\Delta_{\mathsf{hott}}^{-2})$. For the subsequent stages, we can show that in an phase with length $\widetilde{O}(\max(1, \mathsf{N}\mathsf{M}^{-1})\Delta^{-2})$ (needed for getting the estimates of the relevant sub-matrices up to error $O(\Delta)$), the regret incurred per round is at most $\Delta \geq \Delta$. Combining the above and the fact that the total number of phases is at most $O(\log \mathsf{T})$, we obtain the regret described in Theorem 1.

To ensure that expanding the item sets for non-opinionated users does not increase regret, we work with latent embeddings of users and items. Consider the scenario where a user u is connected to opinionated users u_i and u_j with their respective best items $\mathcal{A}_i$ and $\mathcal{A}_j$. Let $\tilde{\mathcal{T}}_v^{\ell}$ denote the candidate item sets for any user v before Line 13. Suppose $\mathcal{A}_i$ is not in $\tilde{\mathcal{T}}_u^{\ell}$ but there exists x in $\tilde{\mathcal{T}}_u^{\ell} \cap \tilde{\mathcal{T}} u_i{}^{\ell}$ due to their connection. The call to `ExpanditemSets` ensures that both $\mathcal{A}_i$ and $\mathcal{A}_j$ are in $\mathcal{T}_u^{\ell}$. We aim to demonstrate control over $\mathbf{R}_{u,\mathcal{A}_j} - \mathbf{R}_{u,\mathcal{A}_i}$ since $\mathcal{A}_i$ will be recommended to user u during the `ESTIMATE` call. Let e and e' denote the latent embedding maps of users and items respectively. We begin with the regret decomposition

$$\mathbf{R}_{u,\mathcal{A}_j} - \mathbf{R}_{u,\mathcal{A}_i} = e(u)^T e'(\mathcal{A}_j - \mathcal{A}_i) = e(u)^T e'(\mathcal{A}_j - x)$$

$$+ e(u_i)^T e'(\mathcal{A}_i - x) + e(u_i + u)^T e'(x - \mathcal{A}_i)$$
$$\leq \left(\mathbf{R}_{u,\mathcal{A}_j} - \mathbf{R}_{u,x}\right) + \left(\mathbf{R}_{u_i,\mathcal{A}_i} - \mathbf{R}_{u_i,x}\right).$$

We show that for a hypothetical user defined by the embedding $e(u_i + u)$, its best item must be $\mathcal{A}_i$, due to which the last inequality follows. Since $x \in \tilde{\mathcal{T}}_u^{\ell} \cap \tilde{\mathcal{T}}_{u_i}^{\ell}$, even before the expansion step, the last two terms can also be controlled (at the level $O(k_\ell)$). This also guarantees that $\mathcal{A}_i$ will belong to the modified item set of user u after a slight expansion of the item set.

Illustrative example on why simple item elimination fails: Now, we provide a working example to motivate why a careful design of algorithm is required to attain optimal regret. Suppose for simplicity we know the value of $\Delta_{\text{hott}} = \Theta(1)$. Then by running Stage 1 $\Theta(1)$ number of times, we can guarantee that the suboptimality of any candidate item for any user is at-most $\Delta_{\text{hott}}/4$ (for eg. see Lemma 5). Consider the following claim which we eventually show to be false.

Proposition 4. *Suppose we run Stage 1 of Alg.* PHASED-CLUSTERELIM *so that the sub-optimality of each candidate item is at-most $\Delta_{hott}/4$. Then if two users u and v do not share the same optimal item, their items sets maintained at the current round will be disjoint.*

Before disproving the proposition, we note that if it was true, then it will imply a simple item elimination strategy as follows. We can run Stage 1 for a constant number of times. Then we cluster together all users who have non-empty intersection of their candidate items and play items that are present in the intersection of their candidate item sets. Afterwards we can refine the intersection of their candidate item sets via matrix completion. This will also guarantee that the regret will remain controlled since the intersection of the item sets do not eliminate the best item of users within a cluster. Unfortunately the proposition above can be false as shown by the counter-example below.

Consider a ground truth reward matrix as follows:

$$R = \begin{bmatrix} 1 & 0 & 1/3 \\ 0 & 1 & 2/3 \\ p & p-\epsilon & p-2\epsilon/3 \end{bmatrix} \tag{4}$$

Algorithm 2 ESTIMATE_SIMPLE (Function for estimating reward sub-matrix restricted to a subset of users and items)

Input: Set of users $\mathcal{U} \subseteq [\mathsf{M}]$, $\mathcal{V} \subseteq [\mathsf{N}]$, desired entry-wise error η.
for round index $1, 2, \ldots, O(\eta^{-2}\max(1, \mathsf{NM}^{-1}))$ **do**
  For each user $u \in \mathcal{U}$, recommend to u an item $v \in \mathcal{V}$ sampled randomly. Store the observed feedback.
end for
Use offline Matrix Completion with observed feedback as input data and return an estimate of the matrix $\mathbf{R}_{\mathcal{U},\mathcal{V}}$.
//There exist several standard algorithms for offline Low Rank Matrix Completion with partial noisy observations. Two well-known approaches are 1) Minimizing MSE with nuclear norm regularizer 2) Alternating Minimization (see (Jain & Kar, 2017) and references therein).

with the corresponding user and item embeddings denoted by $U = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ p & p-\epsilon \end{pmatrix}$ and $V = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1/3 & 2/3 \end{pmatrix}$ where $\epsilon < 1/4$ and $1/4 < p < 1$. Let user corresponding to row i of reward matrix be denoted by u_i and item corresponding to column i be denoted by v_i. Here the hott-topics are items v_1 and v_2 and the hott-topic gap $\Delta_{\text{hott}} = 1$ and $\Delta = 2\epsilon/3$. Further u_1 and u_2 are the opinionated users as defined in Proposition 6. By Lemma 5, we attain a sub-optimality of the candidate item set of $1/4$ by running Stage 1 until $k_\ell = 1/20$. Then the candidate item for user u_1 is $\{v_1\}$, u_2 is $\{v_2\}$ and u_3 may contain $\{v_1, v_2, v_3\}$. Thus Proposition 4 only holds true for opinionated users and can be false for non-opinionated user u_3 since the item sets for u_3 and u_1 intersect although their optimal items are different.

Takeaway message. The primary challenge in our algorithm lies in how to recommend items to less opinionated users while still exploiting their mutual similarities. This required us to do careful identification of relevant sub-matrices that are forwarded to downstream estimation (function ESTIMATE Alg. 2) - here, noisy entries of the sub-matrix are observed to collect data for sub-matrix completion task.

4. Algorithm DETERMINANTELIM and Guarantees on $\mathsf{Reg}_{\mathsf{Simple}}$ (Theorem2)

In this section, we consider the setting when at each round $t \in [\mathsf{T}]$, for each user $u \in [\mathsf{M}]$, r distinct items $\rho_u(t,1), \rho_u(t,2), \ldots, \rho_u(t,r) \in [\mathsf{N}]$ are recommended with the goal of minimizing the simplified regret $\mathsf{Reg}_{\mathsf{Simple}}$ defined in equation (3). In this setting, we make benign assumptions on the reward matrix $\mathbf{R} = \mathbf{U}\mathbf{V}^\mathsf{T}$ namely Assumptions 1 and 4. Note that for any r-row $\mathcal{I}$ and r-column $\mathcal{C}$, we must have that $\det(\mathbf{R}_{\mathcal{I},\mathcal{C}}) = \det(\mathbf{U}_\mathcal{I})\det(\mathbf{V}_\mathcal{C})$ where $\mathbf{R}_{\mathcal{I},\mathcal{C}}$ is the sub-matrix of $\mathbf{R}$ restricted to rows in $\mathcal{I}$ and columns in $\mathcal{C}$. This implies that $\det(\mathbf{R}_{\mathcal{I},\mathcal{C}})$ is a scaled version of $\det(\mathbf{V}_\mathcal{C})$. Recall that $\mathcal{A} \subseteq [\mathsf{N}]$ corresponds to the indices of the r hott columns. Notice that determinant of a matrix corresponds to the signed volume determined by the matrix rows and zero vector. The hott items assumption ensures that the volume of $\mathbf{V}$ restricted to indices in $\mathcal{A}$ is largest - hence, $\det^2(\mathbf{V}_\mathcal{A}) > \det^2(\mathbf{V}_{\mathcal{A}'})$ for any r-column $\mathcal{A}' \neq \mathcal{A}$ (see for eg. (Kveton et al., 2017)) and therefore, we must have that $\det^2(\mathbf{R}_{\mathcal{I},\mathcal{A}}) > \det^2(\mathbf{R}_{\mathcal{I},\mathcal{A}'})$ for any r-row $\mathcal{I}$. This property is exploited crucially by our proposed algorithm DETERMINANTELIM to gradually eliminate sub-optimal r-columns for this setting and converge to the hott item indices $\mathcal{A}$. Moreover, for any r-column $\mathcal{A}' \neq \mathcal{A}$, we define sub-optimality gap of the r-column $\mathcal{A}'$ to be $\det^2(\mathbf{V}_\mathcal{A}) - \det^2(\mathbf{V}_{\mathcal{A}'})$.

Algorithm 3 DETERMINANTELIM (Collaborative Filtering via Phased Elimination of r-columns)

Require: Set of users $[\mathsf{M}]$, set of items N, rounds T, noise variance $\sigma^2 > 0$.
Initialize $\mathcal{B}^{(1)} = \Pi_r(\mathsf{N})$ and $d_1 = O(\mathsf{NM}^{-1/r}C(r))$ where $C(r) = 2^{r+1}r^{1+r/2}$.
for $\ell = 1, 2, \ldots$ **do**
  For each user $u \in [\mathsf{M}]$, sample uniformly at random d_ℓ items from $\cup_{\mathcal{J}\in\mathcal{B}^{(\ell)}}\mathcal{J}$. Store the sampled d_ℓ items in the set $\mathcal{T}_u^{(\ell)}$. // Beginning of a phase
  For the next $2\sigma^2 d_\ell$ rounds, for each user $u \in [\mathsf{M}]$, for each item z in $\mathcal{T}_u^{(\ell)}$, recommend some $\mathcal{J} \in \mathcal{B}^{(\ell)}$ $2\sigma^2$ times such that $z \in \mathcal{J}$. // Recommending items to users
  Initialize $\widetilde{\mathbf{R}}^{(\ell,1)}, \widetilde{\mathbf{R}}^{(\ell,2)} \in \mathbf{0}^{\mathsf{M}\times\mathsf{N}}$ to be all zero matrices. For each user $u \in [\mathsf{M}]$, for each item $z \in \mathcal{T}_u^{(\ell)}$, we store $\widetilde{\mathbf{R}}_{uz}^{(\ell,1)}$ ($\widetilde{\mathbf{R}}_{uz}^{(\ell,2)}$) to be the mean of the first σ^2 (remaining σ^2) observations corresponding to the recommendation of item z to user u.
  For each $\mathcal{J} \in \mathcal{B}^{(\ell)}$, compute $\widetilde{\mu}_\mathcal{J}$ as in equation (5). // Estimate avg. determinant
  Compute $\mathcal{B}^{(\ell+1)}$ to be $\{\mathcal{J} \in \mathcal{B}^{(\ell)} \mid \widetilde{\mu}_\mathcal{J} \geq \max_{\mathcal{J}'\in\mathcal{B}^{(\ell)}}\widetilde{\mu}_{\mathcal{J}'} - 2\cdot 2^{-\ell}\}$.
  Update $d_{\ell+1} \leftarrow 4d_\ell$. // End of a phase
end for

Algorithm DETERMINANTELIM is initialized with the set of all possible r-columns. The algorithm runs in phases of exponentially increasing length and gradually eliminates r-columns that are sub-optimal in every phase. At the be-

ginning of ℓth phase, suppose $\mathcal{B}^{(\ell)}$ is the set of surviving r-columns. Consider the following quantity which will be central to our algorithm design:

$$\mu_{\mathcal{J}} \triangleq \frac{1}{\binom{\mathsf{M}}{r}} \sum_{\mathcal{I}\subset[\mathsf{M}]||\mathcal{I}|=r} \mathsf{det}^2(\mathbf{R}_{\mathcal{I},\mathcal{J}})$$

For a fixed set of r items $\mathcal{J}$, $\mu_{\mathcal{J}}$ is the average of determinant of all $\binom{\mathsf{M}}{r}$ $r\times r$ sub-matrices of $\mathbf{R}$ formed by the r-column $\mathcal{J}$ and each of the $\binom{\mathsf{M}}{r}$ r-rows of $\mathbf{R}$. Clearly, we have $\mu_{\mathcal{J}} \leq \mu_{\mathcal{A}}$. Hence our next goal will be to estimate $\mu_{\mathcal{J}}$ for each r-column $\mathcal{J} \in \mathcal{B}^{(\ell)}$ and eliminate r-columns in $\mathcal{B}^{(\ell)}$ that have low estimated values. By doing so we hope to gradually identify the r-column defined by $\mathcal{A}$.

Denote the set of columns that belongs to some surviving r-column in $\mathcal{B}^{(\ell)}$ to be $\mathcal{Y} \triangleq \cup_{\mathcal{J}\in\mathcal{B}^{(\ell)}}\mathcal{J}$. To compute an estimate $\widetilde{\mu}_{\mathcal{J}}$ of $\mu_{\mathcal{J}}$ for all $\mathcal{J} \in \mathcal{B}^{(\ell)}$, for every user, a set $\mathcal{T}_u^{(\ell)} \subseteq \mathcal{Y}$ of d_ℓ distinct items are uniformly sampled at random from $\mathcal{Y}$. Then, throughout the phase of length $2\sigma^2 d_\ell$, we recommend items in $\mathcal{T}_u^{(\ell)}$ to user u and construct two independent estimates $\widetilde{\mathbf{R}}^{(\ell,1)}, \widetilde{\mathbf{R}}^{(\ell,2)}$ of the reward matrix $\mathbf{R}$ (Lines 3-5). Finally, for a fixed set of r surviving items $\mathcal{J} \in \mathcal{B}^{(\ell)}$, let $\mathcal{H}_{\mathcal{J}} \subseteq [\mathsf{M}]$ be the set of users for which $\mathcal{J} \subset \mathcal{T}_u^{(\ell)}$ i.e., the items in $\mathcal{J}$ are present in the sampled items $\mathcal{T}_u^{(\ell)}$ for every user $u \in \mathcal{H}_{\mathcal{J}}$, We compute estimate $\widetilde{\mu}_{\mathcal{J}}$ of $\mu_{\mathcal{J}}$ as (Line 3):

$$\widetilde{\mu}_{\mathcal{J}} \triangleq \frac{1}{\binom{n_{\mathcal{J}}}{r}} \sum_{\mathcal{I}\subset\mathcal{H}_{\mathcal{J}}||\mathcal{I}|=r} \mathsf{det}(\widetilde{\mathbf{R}}^{(\ell,1)}_{\mathcal{I},\mathcal{J}}) \cdot \mathsf{det}(\widetilde{\mathbf{R}}^{(\ell,2)}_{\mathcal{I},\mathcal{J}}) \quad (5)$$

We prove that in phase ℓ, it suffices to choose the scaled length of the phase $d_\ell = O(\mathsf{N}\mathsf{M}^{-1/r}C(r)2^{2\ell})$ where $C(r) = 2^{r+1}r^{1+r/2}$. Based on our estimates, we can eliminate r-columns from $\mathcal{B}^{(\ell)}$ as shown in Line 3.

Note that the quality of the estimate $\widetilde{\mu}_{\mathcal{J}}$ improves as $|\mathcal{H}_{\mathcal{J}}|$increases. One of the key insights is to show that $|\mathcal{H}_{\mathcal{J}}| \geq n_\ell$ for a fixed n_ℓ if d_ℓ satisfies the condition in the following lemma.

Lemma 2. *Fix $n_\ell > 0, 0 \leq \delta \leq 1$. For each user $u \in [\mathsf{M}]$, suppose d_ℓ items are chosen for recommendation (as in Line 3 of Alg. 3) in phase ℓ. With probability at least $1-\delta$, for every r-column in $\mathcal{B}^{(\ell)}$, we obtain noisy observations from at least n_ℓ distinct users provided $\frac{d_\ell}{\mathsf{N}-d_\ell} \geq \Big(\mathsf{M}^{-1}(12\log\delta^{-1} + r\log\mathsf{N} + 2n_\ell)\Big)^{1/r}$.*

We choose $n_\ell = \widetilde{O}(C(r)2^{2\ell})$ where $C(r) = 2^{r+1}r^{1+r/2}$. That implies that a sufficient value of $d_\ell = O(\mathsf{N}\mathsf{M}^{-1/r}C(r)2^{2\ell})$ for which the high probability event in Lemma 2 holds true. Once, we have computed an estimate $\widetilde{\mu}_{\mathcal{J}}$ of $\mu_{\mathcal{J}}$ for all surviving sets of r-columns $\mathcal{J} \in \mathcal{B}^{(\ell)}$, we can eliminate r-columns from $\mathcal{B}^{(\ell)}$ to compute $\mathcal{B}^{(\ell+1)}$ based on a high probability confidence width ϵ_ℓ (Line 3 in Alg. 3). More precisely, at the end of the ℓth phase, we have

$$\mathcal{B}^{(\ell+1)} = \{\mathcal{J} \in \mathcal{B}^{(\ell)} \mid \widetilde{\mu}_{\mathcal{J}} \geq \max_{\mathcal{J}'\in\mathcal{B}^{(\ell)}} \widetilde{\mu}_{\mathcal{J}'} - 2\epsilon_\ell\}$$

where $\epsilon_\ell = 2^{-\ell}$ with high probability, we can show that the set of hott items $\mathcal{A}$ always survive and furthermore, the sub-optimality gap of surviving r-columns decreases exponentially in every phase. Our argument can also be associated with the regret incurred by each user at every round. More precisely, we have for any user $u \in [\mathsf{M}]$:

$$\mathbf{R}_{u\pi_u(1)} - \max_{j\in[r]} \mathbf{R}_{u\rho_u(t,j)} \leq 6r^{5/2}\Big(\frac{\mathsf{det}^2(\mathbf{V}_{\mathcal{A}}) - \mathsf{det}^2(\mathbf{V}_{\mathcal{J}})}{\mathsf{det}^2(\mathbf{V}_{\mathcal{A}})}\Big).$$

Thus we can bound the regret incurred by the users at each phase as we keep eliminating sub-optimal r-columns. Since sub-optimal r-columns are eliminated in each phase, we can also show the regret to decrease exponentially with phase index. Combining all the arguments, we obtain the regret guarantee provided in Theorem 2.

5. Conclusion and Future Works

In this work we studied online low rank matrix completion under the hott item assumption. Our framework is a significant relaxation of the settings studied in prior works that derive theoretical guarantees. We provide two novel algorithms with theoretical guarantees that exploit the hott items in distinct ways. The bounds in this paper often improve or closely compare with bounds that are derived in much restrictive setups. Though our main focus was on the statistical rate, we seek to improve the computational complexity of our methods in future (see for example Remark 3).

Impact Statement

The work is of theoretical nature and we do not see any negative societal consequences.

References

Abbe, E., Fan, J., Wang, K., and Zhong, Y. Entrywise eigenvector analysis of random matrices with low expected rank. *Annals of statistics*, 48(3):1452, 2020.

Ahn, J., Elmahdy, A., Mohajer, S., and Suh, C. On the fundamental limits of matrix completion: Leveraging hierarchical similarity graphs. *arXiv preprint arXiv:2109.05408*, 2021.

Ahn, K., Lee, K., Cha, H., and Suh, C. Binary rating estimation with graph side information. *Advances in neural information processing systems*, 31, 2018.

Arora, S., Ge, R., Kannan, R., and Moitra, A. Computing a nonnegative matrix factorization–provably. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pp. 145–162, 2012.

Asteris, M., Papailiopoulos, D., and Dimakis, A. Nonnegative sparse pca with provable guarantees. In *International Conference on Machine Learning*, pp. 1728–1736. PMLR, 2014.

Bhojanapalli, S. and Jain, P. Universal matrix completion. In *International Conference on Machine Learning*, pp. 1881–1889. PMLR, 2014.

Blei, D. M., Ng, A. Y., and Jordan, M. I. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan): 993–1022, 2003.

Bresler, G. and Karzand, M. Regret bounds and regimes of optimality for user-user and item-item collaborative filtering. *arXiv:1711.02198*, 2019.

Bresler, G., Chen, G. H., and Shah, D. A latent source model for online collaborative filtering. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'14, pp. 3347–3355, 2014.

Bresler, G., Shah, D., and Voloch, L. F. Collaborative filtering with low regret. In *Proceedings of the 2016 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Science*, SIGMETRICS '16, pp. 207–220, New York, NY, USA, 2016. ACM.

Chen, Y., Chi, Y., Fan, J., Ma, C., and Yan, Y. Noisy matrix completion: Understanding statistical guarantees for convex relaxation via nonconvex optimization. *Siam J.Opt*, 2019.

Dadkhahi, H. and Negahban, S. Alternating linear bandits for online matrix-factorization recommendation. *arXiv preprint arXiv:1810.09401*, 2018.

Deshpande, Y. and Montanari, A. Linear bandits in high dimension and recommendation systems. In *Allerton Conference on Communication, Control, and Computation*, pp. 1750–1754, October 2012.

Donoho, D. and Stodden, V. When does non-negative matrix factorization give a correct decomposition into parts? *Advances in neural information processing systems*, 16, 2003.

Elmahdy, A., Ahn, J., Suh, C., and Mohajer, S. Matrix completion with hierarchical graph side information. *Advances in neural information processing systems*, 33: 9061–9074, 2020.

Gentile, C., Li, S., and Zappella, G. Online clustering of bandits. In *International Conference on Machine Learning*, pp. 757–765. PMLR, 2014.

Hao, B., Zhou, J., Wen, Z., and Sun, W. W. Low-rank tensor bandits. *arXiv preprint arXiv:2007.15788*, 2020.

Heckel, R. and Ramchandran, K. The sample complexity of online one-class collaborative filtering. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, pp. 1452–1460. JMLR.org, 2017.

Huang, B., Huang, K., Kakade, S., Lee, J. D., Lei, Q., Wang, R., and Yang, J. Optimal gradient-based algorithms for non-concave bandit optimization. *Advances in Neural Information Processing Systems*, 34:29101–29115, 2021.

Huleihel, W., Pal, S., and Shayevitz, O. Learning user preferences in non-stationary environments. In *International Conference on Artificial Intelligence and Statistics*, pp. 1432–1440. PMLR, 2021.

Jain, P. and Kar, P. Non-convex optimization for machine learning. *Foundations and Trends in Machine Learning*, 2017.

Jain, P. and Pal, S. Online low rank matrix completion. In *ICLR*, 2023.

Jain, P., Netrapalli, P., and Sanghavi, S. Low-rank matrix completion using alternating minimization. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pp. 665–674, 2013.

Jo, C. and Lee, K. Discrete-valued latent preference matrix estimation with graph side information. In *International Conference on Machine Learning*, pp. 5107–5117. PMLR, 2021.

Jun, K.-S., Willett, R., Wright, S., and Nowak, R. Bilinear bandits with low-rank structure. In *International Conference on Machine Learning*, pp. 3163–3172. PMLR, 2019.

Katariya, S., Kveton, B., Szepesvári, C., Vernade, C., and Wen, Z. Bernoulli rank-1 bandits for click feedback. *arXiv preprint arXiv:1703.06513*, 2017a.

Katariya, S., Kveton, B., Szepesvari, C., Vernade, C., and Wen, Z. Stochastic rank-1 bandits. In *Artificial Intelligence and Statistics*, pp. 392–401. PMLR, 2017b.

Koren, Y. Factorization meets the neighborhood: A multifaceted collaborative filtering model. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '08, pp. 426–434, New York, NY, USA, 2008. ACM.

Kveton, B., Szepesvári, C., Rao, A., Wen, Z., Abbasi-Yadkori, Y., and Muthukrishnan, S. Stochastic low-rank bandits. *arXiv preprint arXiv:1712.04644*, 2017.

Lattimore, T. and Szepesvári, C. *Bandit algorithms*. Cambridge University Press, 2020.

Lee, C.-W., Liu, Q., Abbasi-Yadkori, Y., Jin, C., Lattimore, T., and Szepesvári, C. Context-lumpable stochastic bandits. *arXiv preprint arXiv:2306.13053*, 2023.

Lu, Y., Meisami, A., and Tewari, A. Low-rank generalized linear bandit problems. In *International Conference on Artificial Intelligence and Statistics*, pp. 460–468. PMLR, 2021.

Maillard, O.-A. and Mannor, S. Latent bandits. In *International Conference on Machine Learning*, pp. 136–144. PMLR, 2014.

Negahban, S. and Wainwright, M. J. Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. *The Journal of Machine Learning Research*, 13(1):1665–1697, 2012.

Pal, S. and Mazumdar, A. On learning mixture models with sparse parameters. In *International Conference on Artificial Intelligence and Statistics*, pp. 9182–9213. PMLR, 2022.

Pal, S., Suggala, A. S., Shanmugam, K., and Jain, P. Optimal algorithms for latent bandits with cluster structure, 2023. URL `https://arxiv.org/abs/2301.07040`.

Recht, B. and Ré, C. Parallel stochastic gradient algorithms for large-scale matrix completion. *Mathematical Programming Computation*, 5(2):201–226, 2013.

Recht, B., Re, C., Tropp, J., and Bittorf, V. Factoring nonnegative matrices with linear programs. *Advances in neural information processing systems*, 25, 2012.

Sen, R., Shanmugam, K., Kocaoglu, M., Dimakis, A., and Shakkottai, S. Contextual bandits with latent confounders: An nmf approach. In *Artificial Intelligence and Statistics*, pp. 518–527. PMLR, 2017.

Stojanovic, S., Jedra, Y., and Proutiere, A. Spectral entry-wise matrix estimation for low-rank reinforcement learning. *Advances in Neural Information Processing Systems*, 36, 2024.

Teflioudi, C., Makari, F., and Gemulla, R. Distributed matrix completion. In *2012 ieee 12th international conference on data mining*, pp. 655–664. IEEE, 2012.

Trinh, C., Kaufmann, E., Vernade, C., and Combes, R. Solving bernoulli rank-one bandits with unimodal thompson sampling. In *Algorithmic Learning Theory*, pp. 862–889. PMLR, 2020.

Vershynin, R. High-dimensional probability. *University of California, Irvine*, 10:11, 2020.

Zhang, Q., Suh, G., Suh, C., and Tan, V. Y. Mc2g: An efficient algorithm for matrix completion with social and item similarity graphs. *IEEE Transactions on Signal Processing*, 2022.

Zhou, J., Hao, B., Wen, Z., Zhang, J., and Sun, W. W. Stochastic low-rank tensor bandits for multi-dimensional online decision making. *arXiv e-prints*, pp. arXiv–2007, 2020.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes - Section 2]
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes - Section 2]
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [No]
2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Yes - Theorems 1 and 2 including Assumptions 1,2,3,4.]
   (b) Complete proofs of all theoretical results. [Yes - Appendix D]
   (c) Clear explanations of any assumptions. [Yes - Section 2]
3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [No]
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes - Appendix C]
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Not Applicable]
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Yes]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. [Not Applicable]

(b) The license information of the assets, if applicable. [Not Applicable]
(c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable]
(d) Information about consent from data providers/curators. [Not Applicable]
(e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. [Not Applicable]
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

Algorithm 4 PHASEDELIMSIMPLIFIED (A simplified version of Alg. PHASEDCLUSTERELIM)

Require: Number of users M, number of items N. Input parameter B. Number of hott item vectors d, Gap Parameter Δ.

For B rounds, for each user $u \in [\mathsf{M}]$, recommend a random item in $[\mathsf{N}]$. Compute an estimate $\widetilde{\mathbf{P}}$ of $\mathbf{P}$.
For each user $u \in [\mathsf{M}]$, compute $\mathcal{T}_u = \{j' \in [\mathsf{N}] \mid \max_{j \in [\mathsf{N}]} \widetilde{\mathbf{P}}_{uj} - \widetilde{\mathbf{P}}_{uj'} \leq \Delta\}$.
Run k-means algorithm on the set of users $[\mathsf{M}]$ where each user u is assigned the embedding $\widetilde{\mathbf{P}}[u][\cup_{u \in [\mathsf{M}]} \mathcal{T}_u]$ i.e. the corresponding rewards estimate restricted to the items in $\cup_{u \in [\mathsf{M}]} \mathcal{T}_u$.
For each cluster of users $\mathcal{C}_i$ obtained, we compute $\mathcal{N}^{(i,0)} = \cap_{u \in \mathcal{C}_i} \mathcal{T}_u$ (or a robust version where an item needs to survive in 70% of $\mathcal{T}_u$'s for $u \in \mathcal{C}_i$ to survive in $\mathcal{N}^{(i,0)}$).
for $\ell = 1, 2, \ldots,$ **do**
  Set $\mathsf{B} \leftarrow 3\mathsf{B}$ and $\Delta = \Delta/3$
  For B rounds, for each $i \in [d]$ and for each user $u \in \mathcal{C}_i$, recommend a random item in $\mathcal{N}^{(i,\ell-1)}$.
  For each cluster of users $\mathcal{C}_i$, use a standard matrix completion method to compute an estimate $\widetilde{\mathbf{P}}^{(i,\ell)} \in \mathbb{R}^{\mathsf{M} \times \mathsf{N}}$.
  For each cluster $\mathcal{C}_i$ and for each user u in $\mathcal{C}_i$, compute $\mathcal{T}_u = \{j' \in \mathcal{C}_i \mid \max_{j \in \mathcal{N}^{(i)}} \widetilde{\mathbf{P}}_{uj} - \widetilde{\mathbf{P}}_{uj'} \leq \Delta\}$
  For each $i \in [d]$, compute $\mathcal{N}^{(i,\ell)} = P_{u \in \mathcal{C}_i} \mathcal{T}_u$.
end for

A. Other Related Work

There are several other loosely related works/results that we briefly survey in this section.

The problem of multi-dimensional bandit optimization where the goal is to assume a structure among the items (such as a matrix or tensor with low rank) was introduce in (Katariya et al., 2017b;a; Trinh et al., 2020). In the simplest version of this problem, in the rank-1 setting at each round $t \in [\mathsf{T}]$, an agent selects one row and one column and receives a reward corresponding to the entry of a rank-1 matrix. The regret in this setting is defined with respect to the best *(row,column)* pair, which corresponds to the best item. Since then, this setting has been extended to include rank r (Kveton et al., 2017; Stojanovic et al., 2024) which is most relevant to our setting in terms of techniques. Furthermore, this line of work also extends to rank 1 multi-dimensional tensors (Hao et al., 2020), bilinear bandits (Jun et al., 2019; Huang et al., 2021), and generalized linear bandits (Lu et al., 2021). However, these guarantees cannot be applied to our problem directly in most cases. Note that our goal is to minimize the regret for all users (rows of $\mathbf{R}$) jointly. In our problem, it is crucial to identify the entries (columns) of $\mathbf{R}$ with high rewards for each user (row), in contrast to the multi-dimensional online learning problem, where only the entry with the highest reward needs to be inferred.

A related line of work is the theoretical model for User-based Collaborative Filtering (CF) studied in (Bresler et al., 2014; 2016; Heckel & Ramchandran, 2017; Bresler & Karzand, 2019; Huleihel et al., 2021). To the best of knowledge, these papers first motivated and theoretically analyzed the collaborative framework with the practically relevant restriction that the same item cannot be recommended more than once to the same user. Here, similar to (Pal et al., 2023), a latent cluster structure is assumed to exist across users such that users in same cluster have identical expected rewards. These models are quite restrictive in a theoretical sense as they provide guarantees only on a very relaxed notion of regret (termed *pseudo-regret*).

In the offline setting, several papers have studied the problem of low rank matrix completion with partially observed entries (Negahban & Wainwright, 2012; Chen et al., 2019; Deshpande & Montanari, 2012; Abbe et al., 2020; Jain et al., 2013; Jain & Kar, 2017) and also in the presence of side information such as social graphs or similarity graphs (Ahn et al., 2018; 2021; Elmahdy et al., 2020; Jo & Lee, 2021; Zhang et al., 2022). Some of these results can be converted into greedy algorithms for the online problem (as has been done in (Jain & Pal, 2023)) - we can solve the offline problem in a particular number of rounds and subsequently commit to the computed estimate of the reward matrix.

B. Experiments

We conduct synthetic experiments to demonstrate the efficacy of our PHASEDCLUSTERELIM algorithm that runs in phases. All experiments have been conducted on Colab Machines with a system RAM of 12GB and hard disk space of 23.7GB.

Simplified Algorithm: Having practical considerations, we propose a simplified version of our PHASEDCLUSTERELIM algorithm (denoted as Algorithm PHASEDELIMSIMPLIFIED - ALg. 4). There are two main advantages of Algorithm 4 from implementation point of view - 1) the runtime is polynomial in the rank of the reward matrix 2) the number of

hyperparameters to tune is much small. At a high level, in Algorithm 4, we proposed a simplification where the clustering step is decoupled (Step 3) from the exploration/exploitation step (Steps 6-11). In other words, the clustering step is done only once at the beginning. Note that such an algorithm (Algorithm 3) is simplified (easy to understand and implement) and hyperparameter-scarce. On the other hand, because of the decoupling step, it will not share the near optimal regret guarantees of Algorithm 1. Nevertheless, Algorithm 4 is expected to be significantly better than the standard greedy algorithm because of the initial clustering step.

Let us elaborate. In Algorithm PHASEDELIMSIMPLIFIED, the overall idea is to first recommend random items to every user in the first few rounds and then compute an initial estimate of the low rank matrix. Based on this initial estimate, we run k-means with $k = d$ and cluster the set of users into d clusters. For each cluster of users, we find a small active set of items that are good for all users in the same cluster. The main workhorse is that users in the same cluster have the same hott item as the best item. Next, Algorithm PHASEDELIMSIMPLIFIED runs in phases of exponentially increasing length where for each cluster of users, we again recommend random items (from the active set) to every user in the same cluster. Next, we run low rank matrix completion to compute an estimate of the sub-matrix restricted to each cluster of users and its corresponding good items. Based on this estimate, we prune the set of good items further. The goal is to converge to the best item for every cluster of users.

Baselines for Comparison: We compare with two distinct baselines 1) Alternating Minimization (AM) - This algorithm was proposed for online low rank matrix completion in (Dadkhahi & Negahban, 2018) based on an alternating minimization recipe. Here, assuming the reward matrix to be of low rank i.e. $\mathbf{P} = \mathbf{U}\mathbf{V}^{\mathsf{T}}$, estimates of $\mathbf{U}, \mathbf{V}$ are improved gradually. However, it is important to note that no theoretical guarantees were provided in (Dadkhahi & Negahban, 2018). 2) Explore Then Commit (ETC) Algorithm - This greedy algorithm was proposed in (Jain & Pal, 2023). In this algorithm, we recommend random items to every user for an initial few rounds (say m - exploration rounds) and compute an estimate of the expected reward matrix. Based on the low rank estimate, we recommend the estimated best item for every user in the remaining rounds.

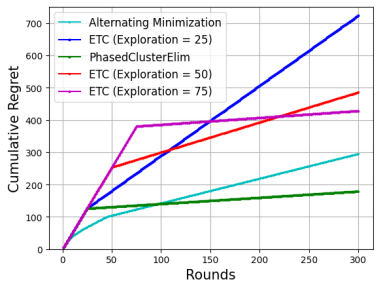
We consider an instance with $\mathsf{M} = 200$ users, $\mathsf{N} = 200$, items, rank $r = 4$ across $\mathsf{T} = 300$ rounds. Here, we consider the expected reward matrix $\mathbf{P} = \mathbf{U}\mathbf{V}^{\mathsf{T}}$ constructed in the following way: 1) For every $i \in [\mathsf{M}]$, in the ith row of $\mathbf{U}$, there is a 1 in the $(i\%\mathsf{C})^{\mathsf{th}}$ position and 0 elsewhere in the row. 2) The first $\mathsf{N} - r$ columns of $\mathbf{V}$ are obtained by generating each entry from $\mathcal{N}(0, 1)$ independently and subsequently, the negative entries are trimmed to zero. Suppose the maximum positive entry in the first $\mathsf{N} - r$ columns of $\mathbf{V}$ is α. In that case, the final r columns are an identity matrix multiplied by a factor of 2α. This ensures the hott item property - every column in $\mathbf{V}$ can be written as a convex combination of the final r columns. Now, with this reward matrix, at each round, we recommend one item to every user such that a noisy reward is obtained (eq. (1)) with gaussian noise having variance $\sigma^2 = 0.25$.

We run Algorithm PHASEDELIMSIMPLIFIED (Alg. 4 - the simplified version of Alg. PHASEDCLUSTERELIM) in this set-up (with an initial exploration period of 25) along with the AM algorithm and the ETC algorithm. The number of exploration rounds in ETC is a hyper-parameter and we run the ETC algorithm with number of exploration rounds $m = 25, 50, 75$. For the AM algorithm, we use the same hyper-parameters as suggested in (Dadkhahi & Negahban, 2018) for the setting when $\mathbf{U}, \mathbf{V}$ have entries sampled independently from $\mathcal{N}(0, 1)$.

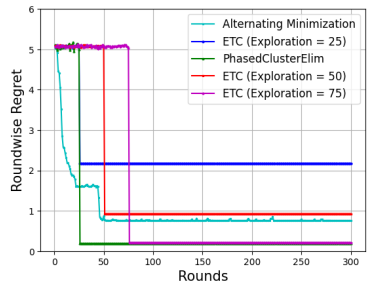
Results and Insights: All algorithms have been run 5 times and the average regret has been reported across the 5 runs in Figures 1a and 1b. Note that from Figure 1a our algorithm incurs a much lower cumulative regret compared to the Alternating Minimization algorithm and the ETC (greedy) algorithm with different exploration periods (small and large). The picture is clearer in 1b - both the ETC algorithms and our algorithm have an initial exploratory period but our algorithm is able to make use of the exploration period significantly better. Note that the ETC algorithm with exploration period=25 rounds has a higher regret per round than our algorithm with a similar exploration period=25 rounds. ETC algorithms with higher exploration period have smaller exploitation regret per round but their exploration cost is larger. The AM algorithm also has a similarly large regret.

C. Missing Algorithmic Modules and Proofs

The missing algorithmic modules from Section 3 are preseneted in Algorithms 5, 6 and 7.



(a) Regret accrued by Alg. 4, Explore-Then-Commit algorithm (Jain & Pal, 2023) for exploration periods $(25, 50, 75)$, Alternating Minimization (AM) algorithm (Dadkhahi & Negahban, 2018) when $\mathsf{T} = 300$. The cumulative regret is plotted with number of rounds.



(b) Comparison of the regret incurred in each round $(\mathsf{M}^{-1}\sum_{u\in[\mathsf{M}]}\mathbf{P}_{u\pi_u(1)} - \frac{1}{\mathsf{M}}\sum_{u\in[\mathsf{M}]}\mathbf{P}^{(t)}_{u\rho_u(t)})$ by Algorithm 4 and compared with Explore-Then-Commit algorithm (Jain & Pal, 2023) for exploration periods $(25, 50, 75)$, Alternating Minimization (AM) algorithm (Dadkhahi & Negahban, 2018).

Figure 1: Comparison of regret incurred by Algorithm 4 (a simplified version of our proposed phased elimination algorithm namely Algorithm PHASEDCLUSTERELIM) with baselines such as Explore-Then-Commit (Greedy) algorithm and Alternating Minimization (AM) algorithm. Clearly, Algorithm 4 outperforms the other baselines significantly.

Algorithm 5 `CheckIndependentSet`: Algorithm to check existence of an independent set of size d

Input: A graph G

for any combination of d vertices $(u_1, \ldots, u_d)$ from the graph **do**
  if there is no edge between any pair u_i and u_j **then**
    return (True , $(u_1, \ldots, u_d)$)
  end if
end for
return False

Algorithm 6 `ExpanditemSets`: Algorithm to update the items sets of users that are connected to multiple clusters

Input: A graph $G(V, E)$; Candidate item sets $\mathcal{T}_u^\ell$ for every user; Seed users $u_1, \ldots, u_d$, suboptimality level $c_{\ell-1}$, estimated reward matrix $\tilde{R}^\ell$

for each user u that is connected to more than one user in $u_{1:d}$ **do**
  Let $\mathcal{C} \leftarrow \{u_i : \mathcal{T}_{u_i}^\ell \cap \mathcal{T}_u^\ell \neq \Phi\}$
  `//Expanding the item set for users in` D `slightly`
  Set $\mathcal{T}_u^\ell \leftarrow \{x : \tilde{R}^\ell_{u,\tilde{\pi}_u(1)} - \tilde{R}^\ell_{u,x} \leq 12k_{\ell-1}$ and $x \in \mathcal{T}^\ell_{C_i}$ for some $i \in \mathcal{C}\}$
end for
return the updated item sets

C.1. Proofs for generalized regret setting (Setting 1 from Section 2)

Throughout the proofs, we suppose that the reward matrix satisfies the assumptions detailed in Section 2 for Setting 1. However, we generalise Assumption 1 as below so that we allow d number of hott-topics where $d \geq r$. All the results of Setting 1 can be recovered by setting $d = r$, where recall that r is the rank of the reward matrix.

Assumption 5 (**B1** Hott items and vectors)**.** *We assume that there exists a set of d unknown distinct ordered indices $\mathcal{A} \subseteq [\mathsf{N}], |\mathcal{A}| = d$ such that all the item embedding vectors $\{\mathbf{V}_i\}_{i\in[\mathsf{N}]}$ lie within the convex hull of $\{\mathbf{V}_i\}_{i\in\mathcal{A}} \cup \{\mathbf{0}\}$. We will call the items corresponding to indices in $\mathcal{A}$ hott items and the rows in $\mathbf{V}_\mathcal{A}$ to be hott vectors. Further, we suppose that $r = O(1)$.*

Thus moving forward, the reward matrix is assumed to satisfy Assumptions 5, 2 and 3.

We have the following mtarix completion guarantee for Algorithm 7 which is a direct consequence of Remark 6 from (Jain & Pal, 2023).

Proposition 5 (Matrix completion guarantee)**.** *Let $d_1 = \max(\mathsf{M}, \mathsf{N})$ and $d_2 = \min(\mathsf{M}, \mathsf{N})$. Set $p = C\mu^2 d_2^{-1}\log^3 d_2$,*

Algorithm 7 ESTIMATE from (Jain & Pal, 2023)

Require: Set of users $\mathcal{U} \subseteq [\mathsf{M}]$, set of items $\mathcal{V} \subseteq [\mathsf{N}]$, sub-optimality level η, sampling probability p, regularization parameter λ, rounds control parameter s, candidate item sets of users $\mathcal{T}_u$ for all $u \in [M]$. Index of round t is relative to the first round when the algorithm is invoked; hence $t = 1, 2, \ldots, m$.

For each tuple of indices $(i, j) \in \mathcal{U} \times \mathcal{V}$, independently set $\delta_{ij} = 1$ with probability p and $\delta_{ij} = 0$ with probability $1 - p$.
Denote $\Omega = \{(i, j) \in \mathcal{U} \times \mathcal{V} \mid \delta_{ij} = 1\}$ and $b = \max_{i \in [\mathcal{U}]} \mid |j \in [\mathcal{V}] \mid (i, j) \in \Omega|$ to be the maximum number of index tuples in a particular row.
Set $b = |\mathcal{V}|$ and $m = bs$.
for $\ell = 1, 2, \ldots, m/b$ **do**
  For all $(i, j) \in \Omega$, set $\mathsf{Mask}_{ij} = 0$.
  for $\ell' = 1, 2, \ldots, b$ **do**
    for each user $u \in \mathcal{U}$ in round $t = (\ell - 1)b + \ell'$ **do**
      Recommend an item $\rho_u(t)$ in $\{j \in \mathcal{V} \mid (u, j) \in \Omega, \mathsf{Mask}_{uj} = 0\}$ and set $\mathsf{Mask}_{u\rho_u(t)} = 1$. If not possible then recommend any item $\rho_u(t)$ in $\mathcal{V}$ s.t. $(u, \rho_u(t)) \notin \Omega$. Observe $\mathbf{R}^{(t)}_{u\rho_u(t)}$.
    end for
    for each user $u \notin \mathcal{U}$ in round $t = (\ell - 1)b + \ell'$ **do**
      Recommend an item randomly from its candidate item set $\mathcal{T}_u$ and observe its noisy reward.
    end for
  end for
end for
For each $(u, j) \in \Omega$, compute $\mathbf{Z}_{uj}$ to be average of $\lfloor m/b \rfloor$ observations corresponding to user u being recommended item j i.e. $\mathbf{Z}_{uj} = \text{avg}\{\mathbf{R}^{(t)}_{u\rho_u(t)}$ for $t \in [m] \mid \rho_u(t) = j\}$. Discard all other observations corresponding to indices not in Ω.
Without loss of generality, assume $|\mathcal{U}| \leq |\mathcal{V}|$. For each $i \in \mathcal{V}$, independently set δ_i to be a value in the set $[\lceil |\mathcal{V}|/|\mathcal{U}| \rceil]$ uniformly at random. Partition indices in $\mathcal{V}$ into $\mathcal{V}^{(1)}, \mathcal{V}^{(2)}, \ldots, \mathcal{V}^{(k)}$ where $k = \lceil |\mathcal{V}|/|\mathcal{U}| \rceil$ and $\mathcal{V}^{(q)} = \{i \in \mathcal{V} \mid \delta_i = q\}$ for each $q \in [k]$. Set $\Omega^{(q)} \leftarrow \Omega \cap (\mathcal{U} \times \mathcal{V}^{(q)})$ for all $q \in [k]$. *#If $|\mathcal{U}| \geq |\mathcal{V}|$, we partition the indices in $\mathcal{U}$.*
for $q \in [k]$ **do**
  Solve convex program

$$\min_{\mathbf{Q}^{(q)} \in \mathbb{R}^{|\mathcal{U}| \times |\mathcal{V}^{(q)}|}} \frac{1}{2} \sum_{(i,j) \in \Omega^{(q)}} \left(\mathbf{Q}^{(q)}_{i\pi(j)} - \mathbf{Z}_{ij} \right)^2 + \lambda \|\mathbf{Q}^{(q)}\|_\star,$$

  where $\|\mathbf{Q}^{(q)}\|_\star$ denotes nuclear norm of matrix $\mathbf{Q}^{(q)}$ and $\pi(j)$ is index of j in set $\mathcal{V}^{(q)}$.
end for
return $\widetilde{\mathbf{Q}} \in \mathbb{R}^{|\mathcal{U}| \times |\mathcal{V}|}$ s.t. $\widetilde{\mathbf{Q}}_{\mathcal{U}, \mathcal{V}^{(q)}} = \mathbf{Q}^{(q)}$ for all $q \in [k]$ and for every $(i, j) \notin \mathcal{U} \times \mathcal{V}$, $\widetilde{\mathbf{Q}}_{ij} = 0$.

$s = \left\lceil \left(\frac{c\sigma r \sqrt{\mu}}{k_\ell \log d_2} \right)^2 \right\rceil$ and $\lambda = C_\lambda \sigma \sqrt{d_2 p}$ for a suitable constants $c, C, C_\lambda > 0$. Let $\hat{\mathbf{R}} = \texttt{ESTIMATE}(\mathsf{M}, \mathsf{N}, k_\ell, p, \lambda, s)$. Then with probability at-least $1 - \delta$, we have that

$$\|\hat{\mathbf{R}} - \mathbf{R}\|_\infty \leq k_\ell$$

Further the total number of rounds Algorithm 7 runs is bounded by $\tilde{O}\left(\max(1, \mathsf{N}/\mathsf{M})/k_\ell^2\right)$, where $\tilde{O}$ hides multiplicative factors of $\log(1/\delta)$.

Remark 5. *Note that the matrix completion guarantee in Proposition 5 implies that a reasonable estimate of the reward matrix $\mathbf{R}$ can be obtained if we have $\widetilde{O}(\mathsf{NM}sp)$ randomly sampled observations. Intuitively, every entry of $\mathbf{R}$ is used with probability p - if used, then s noisy observations are made corresponding to that entry and the average of those observations is taken to reduce the variance. The bottom-line is that we do not need to see all the entries of the reward matrix - usually p is set to be of the order of $1/\min(\mathsf{M}, \mathsf{N})$ and therefore only a small number of entries are actually sufficient for us to compute an estimate of the reward matrix.*

Remark 6. *In our algorithm, it will be often necessary to estimate sub-matrices of the reward matrix. To do so, the sub-matrix like the entire reward matrix $\mathbf{R}$ needs to satisfy certain incoherence and condition number guarantees - they*

need to be small. In this regard, Assumption 3 ensures that the incoherence and condition number of all sub-matrices of a reasonable size are bounded by at most a constant multiplicative factor of the incoherence and condition number of the original reward matrix $\mathbf{R}$ *(see (Pal et al., 2023) for example).*

First we proceed to bound the regret of stage 1 in Algorithm 1

Lemma 3. *Let the phase* ℓ *be such that the Algorithm 1 is operating in Stage 1. Then with probability at-least* $1-\delta$*, we have that* $\pi_u(1) \in \mathcal{T}_u^{\ell+1}$ *for all users* u*.*

Proof. Consider a user u. We have

$$\begin{aligned}\tilde{\mathbf{R}}^{\ell+1}_{u,\tilde{\pi}_u(1)} - \tilde{\mathbf{R}}^{\ell+1}_{u,\pi_u(1)} &= \tilde{\mathbf{R}}^{\ell+1}_{u,\tilde{\pi}_u(1)} - \mathbf{R}_{u,\tilde{\pi}_u(1)} + \mathbf{R}_{u,\pi_u(1)} - \tilde{\mathbf{R}}^{\ell+1}_{u,\pi_u(1)} + \mathbf{R}_{u,\tilde{\pi}_u(1)} - \mathbf{R}_{u,\pi_u(1)} \\ &\leq 2k_\ell,\end{aligned}$$

with probability at-least $1-\delta$ where in the last line we used Proposition 5 and the fact that $\mathbf{R}_{u,\tilde{\pi}_u(1)} - \mathbf{R}_{u,\pi_u(1)} \leq 0$.

□

Lemma 3 implies that with high probability, in any epoch that belongs to Stage 1, the best item for each user survives in its candidate set.

The following terminology will be central to the arguments in the proofs.

Terminology: We use the terms phase and epoch interchangeably. For the rest of the proof, we fix the quantities. Let ℓ_0 be the value of epoch ℓ in Algorithm 1 when the control reaches line 7. For epochs $\ell < \ell_0$, we fix $k_\ell := 2^{-\ell}$ for $\ell \leq \ell_0$; $k_{\ell_0} = k_{\ell_0-1}/10$; $k_\ell = k_{\ell-1}/2$ for $\ell \geq \ell_0+1$. Let $c_\ell = 3k_\ell$ used in Algorithm 1. Stage 1 runs from epochs 0 to ℓ_0; All epochs greater than or equal to ℓ_0+1 belongs to Stage 2. For notational simplicity we take the hott-topic set $\mathcal{A} = [d]$.

The following Proposition is a direction consequence of Definition 1.

Proposition 6. *There exists* d *opinionated users* $v_{1:d}$ *such that for any* v_i*, any hott-topic* $x \neq \pi_{v_i}(1)$ *is well separated:*

$$\mathbf{R}_{v_i,\pi_{v_i}(1)} - \mathbf{R}_{v_i,x} \geq \Delta_{hott}.$$

We call the users $v_{1:d}$ *as opinionated users. We remark that both* $v_{1:d}$ *and* Δ *need not be known ahead of time.*

Lemma 4. *Consider an item* $x \in [\mathsf{M}]$*, a user* u *and epoch* $\ell \leq \ell_0$*. Suppose* $x \notin \mathcal{T}_u^{\ell+1}$*. Then with probability at-least* $1-\delta$*, we have that*

$$\mathbf{R}_{u,\pi_u(1)} - \mathbf{R}_{u,x} > k_\ell.$$

Proof. Since $x \notin \mathcal{T}_u^{\ell+1}$,

$$\begin{aligned}3k_\ell &< \tilde{\mathbf{R}}_{u,\tilde{\pi}_u(1)} - \tilde{\mathbf{R}}_{u,x} \\ &\leq_{(a)} \tilde{\mathbf{R}}_{u,\tilde{\pi}_u(1)} - \mathbf{R}_{u,\tilde{\pi}_u(1)} + \mathbf{R}_{u,x} - \tilde{\mathbf{R}}_{u,x} + \mathbf{R}_{u,\pi_u(1)} - \mathbf{R}_{u,x} \\ &\leq_{(b)} 2k_\ell + \mathbf{R}_{u,\pi_u(1)} - \mathbf{R}_{u,x},\end{aligned}$$

where in line (a) we used the fact that $\mathbf{R}_{u,\pi_u(1)} \geq \mathbf{R}_{u,\tilde{\pi}_u(1)}$ and in line (b) we used Proposition 5.

Thus if an item x gets eliminated by the end of epoch ℓ, we are guaranteed with high probability that

$$\mathbf{R}_{u,\pi_u(1)} - \mathbf{R}_{u,x} > k_\ell.$$

□

Lemma 5. *Consider an epoch* ℓ *such that* $x \in \mathcal{T}_u^{\ell+1}$ *for a user* u*. Then with probability at-least* $1-\delta$*, it holds that*

$$\mathbf{R}_{u,\pi_u(1)} - \mathbf{R}_{u,x} \leq 5k_\ell.$$

Proof. We have

$$\begin{aligned}\mathbf{R}_{u,\pi_u(1)} - \mathbf{R}_{u,x} &= \mathbf{R}_{u,\pi_u(1)} - \tilde{\mathbf{R}}^{\ell+1}_{u,\pi_u(1)} + \tilde{\mathbf{R}}^{\ell+1}_{u,x} - \mathbf{R}_{u,x} + \tilde{\mathbf{R}}^{\ell+1}_{u,\pi_u(1)} - \tilde{\mathbf{R}}^{\ell+1}_{u,\tilde{\pi}_u(1)} + \mathbf{R}^{\ell+1}_{u,\tilde{\pi}_u(1)} - \tilde{\mathbf{R}}^{\ell+1}_{u,x} \\ &\leq 2k_\ell + 3k_\ell \\ &= 5k_\ell,\end{aligned}$$

with probability at-least $1-\delta$, where we used Proposition 5 and the fact that $x \in \mathcal{T}_u^{\ell+1}$

□

Lemma 6. *Consider an epoch ℓ_0 such that the call to Algorithm 5 in Line 2 of Algorithm 1 returns True for the first time. Let $u_{1:d}$ be the users returned by Algorithm 5. Then it holds with probability at-least $1-\delta$ that:*

1. *Best item for user u_i is different from the best item for user u_j, $j \neq i$.*
2. *$\ell_0 \leq \lceil \log\left(\frac{40}{\Delta_{hott}}\right)\rceil$, where Δ_{hott} is as in Proposition 6.*

Proof. By Lemma 3, the best item for any user u survives in its candidate set $T_u^{\ell_0}$ with probability at-least $1-\delta$. Then if two users u_i and u_j have the same best item, then there will be an edge between them. This is a contradiction to the fact that $u_{1:d}$ forms an independent set. This proves the first statement.

Consider opinionated users $V := \{v_1, \ldots, v_d\}$ as per Proposition 6. Epoch ℓ_0 is the first time Line 2 of Algorithm 1 returns True. So at epoch $\ell_0 - 1$ there must exist a user $u \in V$ and a hott-topic $x \neq \pi_u(1)$ such that $x \in \mathcal{T}_u^{\ell_0-1}$. Now combining Proposition 6 and Lemma 5 we have

$$\begin{aligned}\Delta_{\text{hott}} &\leq \mathbf{R}_{u,\pi_u(1)} - \mathbf{R}_{u,x} \\ &\leq 5k_{\ell_0-2},\end{aligned}$$

with probability at-least $1-\delta$. Rearranging the last display yields the second statement of the lemma.

□

Next, we provide a crude upper bound on Stage 1 regret of Algorithm 1

Lemma 7 (Stage 1 regret bound)**.** *Let ℓ_0 be the number of epochs until the completion of Stage 2. Then with probability at-least $1-\delta$, the average regret incurred by Algorithm 1 during Stage 1 is bounded by*

$$\textit{Reg}_{\textit{Stage 1}} = \tilde{O}(\max(1, \mathsf{N}/\mathsf{M})/\Delta^2_{hott} + 1),$$

with probability at-least $1-\delta$. Here Δ_{hott} is as in Proposition 6.

Proof. Due to Proposition 5, the total number of rounds Stage 1 operates can be bounded by

$$\begin{aligned}\sum_{l=1}^{\ell_0} \tilde{O}\left(\max(1, \mathsf{N}/\mathsf{M})/k_\ell^2\right) &= \sum_{l=1}^{\ell_0} \tilde{O}\left(\max(1, \mathsf{N}/\mathsf{M})2^{2\ell}\right) \\ &= \tilde{O}\left(\max(1, \mathsf{N}/\mathsf{M})2^{2\ell_0} + 1\right) \\ &= \tilde{O}(\max(1, \mathsf{N}/\mathsf{M})/\Delta^2_{\text{hott}} + 1),\end{aligned}$$

where the last line uses Lemma 6.

In Stage 1, the algorithm incurs $O(1)$ regret for any user since the rewards are assumed to be bounded. This concludes the lemma.

□

Lemma 8. *Suppose before the call to Algorithm 6, the best item of every user belongs to its corresponding candidate sets. Then the best item for each user survives in the candidate item sets created by the end of Algorithm 6.*

Consider the seed users $u_{i:d}$ given as input to the algorithm. Since they form an independent set, the best item of these users must be different. WLOG we assume that user u_i has hott-topic i as its best item.

Further consider a user u that is connected to more than one user in $u_{1:d}$ at Line 1 of Algorithm 6. Let u_i be one such user that the user u is connected to. Then in the final item set $\mathcal{T}_u^\ell$ maintained by the algorithm (Line 3), it is guaranteed that the hott-topic item $i \in \mathcal{T}_u^\ell$.

Proof. The first time Algorithm 6 gets called within Algorithm 1 is the epoch $\ell_0 + 1$. So the current epoch must obey $\ell \geq \ell_0 + 1$.

Let $u_1, \ldots, u_d$ be the set of users maintained at Line 7 of Algorithm 1. These are basically given as the "seed users" to Algorithm 6. Let the current epoch be ℓ. Since the best item was not eliminated before the call to Algorithm 6, we have that $i \in \mathcal{T}_{u_i}^\ell$.

Note that we are in in stage 2. The independent set of users were first discovered before the start of epoch ℓ_0. (i.e based on the items sets $\mathcal{T}_\cdot^{\ell_0}$). Then due to Lemma 4, we conclude that for any hott-topic $j \in [d] \setminus \{i\}$

$$\mathbf{R}_{u_i,i} - \mathbf{R}_{u_i,j} > k_{\ell_0 - 1} \tag{6}$$

Next, we are going to show that for the user u, the expansion step at Line 3 ensures that the hott-topic i belongs to the updated item set of the user u.

Let the itemset for the user u before invoking Algorithm 6 be denoted as $\tilde{\mathcal{T}}_u^\ell$. By the premise of the lemma, $\tilde{\mathcal{T}}_u^\ell \cap \mathcal{T}_{u_i}^\ell \neq \Phi$.

In the first case where $i \in \tilde{\mathcal{T}}_u^\ell$, we can surely say that $i \in \mathcal{T}_u^\ell$ as well since in Line 3 we are only setting the acceptable sub-optimality level to be 12 times the sub-optimality level used to create $\tilde{\mathcal{T}}_u^\ell$ in the first place.

So in the rest of the proof we consider the case when $i \notin \tilde{\mathcal{T}}_u^\ell$. Since $\tilde{\mathcal{T}}_u^\ell \cap \mathcal{T}_{u_i}^\ell \neq \Phi$, there must exist an item x such that $x \in \tilde{\mathcal{T}}_u^\ell \cap \mathcal{T}_{u_i}^\ell$. Note that this item x cannot be a hott-topic because $\mathcal{T}_{u_i}^\ell$ does not contain any other hott-topic $j \neq i$.

Let $j \neq i$ be the best item for user u. Then by Lemma 5, we conclude that

$$\mathbf{R}_{u_i,i} - \mathbf{R}_{u_i,x} \leq 5k_{\ell-1}, \tag{7}$$

and

$$\mathbf{R}_{u,j} - \mathbf{R}_{u,x} \leq 5k_{\ell-1}. \tag{8}$$

Note that due to our low rank reward matrix assumption, the reward of any user v for an item k, denoted by $\mathbf{R}_{v,k}$ can be represented as $\mathbf{R}_{v,k} = e(v)^T e'(k)$ where e, e' denotes the latent embedding of user and item respectively.

Hence the previous equations can be restated as

$$e(u_i)^T e'(i) - e(u_i)^T e'(x) \leq 5k_{\ell-1},$$

and

$$e(u)^T e'(j) - e(u)^T e'(x) \leq 5k_{\ell-1}. \tag{9}$$

Adding the last two displays yields that

$$\begin{aligned} e(u_i+u)^T e'(x-j) &\geq e(u_i)^T e'(i-j) - 10k_{\ell-1} \\ &>_{(a)} k_{\ell_0-1} - 10k_{\ell-1} \\ &\geq k_{\ell_0-1} - 10k_{\ell_0} \\ &= 0, \end{aligned}$$

where in line (a) we used Eq.(6) and in the following lines we used the fact that the current epoch $\ell \geq \ell_0 + 1$ since we are in stage 2 and hence $k_\ell \leq k_{\ell_0} = k_{\ell_0-1}/10$ (Line 8 of Algorithm 1).

Thus we conclude that for a hypothetical user defined by the embedding $e(u_i + u)$, the item x is strictly better than the hott-topic j.

For any hott-topic $p \notin \{i, j\}$, since $\mathbf{R}_{u,p} \leq \mathbf{R}_{u,j}$ we have that $e(u)^T e'(p) - e(u)^T e'(x) \leq 5k_\ell$ due to Eq.(9).

Hence by proceeding similarly as before we conclude that for the user defined by the embedding $e(u_i + u)$, the item x is strictly better than any hott-topic in $[d] - \{i\}$.

So the best item for this user cannot be any hott-topic in $[d] - \{i\}$. Consequently we conclude that the best item for the user defined by the embedding $e(u_i + u)$ must be i.

Now we look at how much the item i is suboptimal for the user u.

$$\begin{aligned} \mathbf{R}_{u,j} - \mathbf{R}_{u,i} &= \mathbf{R}_{u,j} - \mathbf{R}_{u,x} + \mathbf{R}_{u,x} - \mathbf{R}_{u,i} \\ &\leq 5k_{\ell-1} + \mathbf{R}_{u,x} - \mathbf{R}_{u,i} \\ &= 5k_{\ell-1} + e(u_i+u)^T e'(x-i) + \mathbf{R}_{u_i,i} - \mathbf{R}_{u_i,x} \\ &\leq 10k_{\ell-1} + e(u_i+u)^T e'(x-i) \\ &\leq 10k_{\ell-1}, \end{aligned} \tag{10}$$

where we used Eq. (7) and (8) and the fact that item i is best for the user defined by the embedding $e(u_i + u)$.

To make further progress, we assume the following: $|\tilde{\mathbf{R}}^{\ell}_{u,i} - \mathbf{R}_{u,i}| \leq k_{\ell-1}$ with probability at-least $1 - \delta$.

This is trivially true for the first epoch in stage 2 which is $\ell_0 + 1$. Because at epoch $\ell_0 + 1$ we have that $|\tilde{\mathbf{R}}^{\ell_0+1}_{y,h} - \mathbf{R}_{y,h}| \leq k_{\ell_0}$ for all users y and all items h. This follows from the fact that until epoch ℓ_0, we estimate the entire reward matrix globally.

We will proceed to show that $|\tilde{\mathbf{R}}^{\ell+1}_{u,i} - \mathbf{R}_{u,i}| \leq k_\ell$ as well.

Recall that $\pi_u(1) = j$. We have

$$\begin{aligned} \tilde{\mathbf{R}}^{\ell}_{u,\tilde{\pi}_u(1)} - \tilde{\mathbf{R}}^{\ell}_{u,i} &= \tilde{\mathbf{R}}^{\ell}_{u,\tilde{\pi}_u(1)} - \mathbf{R}_{u,\tilde{\pi}_u(1)} + \mathbf{R}_{u,\tilde{\pi}_u(1)} - \mathbf{R}_{u,\pi_u(1)} + \mathbf{R}_{u,\pi_u(1)} - \mathbf{R}_{u,i} + \mathbf{R}_{u,i} - \tilde{\mathbf{R}}^{\ell}_{u,i} \\ &\leq \tilde{\mathbf{R}}^{\ell}_{u,\tilde{\pi}_u(1)} - \mathbf{R}_{u,\tilde{\pi}_u(1)} + \mathbf{R}_{u,\pi_u(1)} - \mathbf{R}_{u,i} + \mathbf{R}_{u,i} - \tilde{\mathbf{R}}^{\ell}_{u,i} \\ &\leq 2k_{\ell-1} + \mathbf{R}_{u,\pi_u(1)} - \mathbf{R}_{u,i} \\ &\leq 12k_{\ell-1}, \end{aligned}$$

where in the last line we used Eq.(10).

Thus we conclude that the expansion step in Line 3 of Algorithm 6 ensures that $i \in \mathcal{T}^{\ell}_u$.

Now to prove the induction hypothesis that $|\tilde{\mathbf{R}}^{\ell+1}_{u,i} - \mathbf{R}_{u,i}| \leq k_\ell$, we look at the for loop in Line 15 of Algorithm 1 when processing the cluster i and during the epoch ℓ. Note that $u \in \mathcal{U}$ by virtue of the expansion step. Consequently in the matrix estimation step in Line 17 of Algorithm 1, we end up estimating the entry $\mathbf{R}_{u,i}$ with errror at-most k_ℓ. This proves the induction hypothesis.

The best item for user u was already present in its candidate item set before the start of the Algorithm 6. This candidate item set was constructed by accumulating all items with empirical sub-optimality level of $3k_{\ell-1}$. Now noting that the expansion step cannot eliminate the best item of the user u concludes the proof.

□

Lemma 9. *(stage 2 best item survival) Consider an epoch ℓ within stage 2 (so $\ell \geq \ell_0 + 1$). Suppose for any user u, $\pi_u(1) \in \mathcal{T}_u^\ell$. Then with probability at-least $1 - \delta$ we have that $\pi_u(1) \in \mathcal{T}_u^{\ell+1}$.*

Proof. Updating the candidate item set in Stage 2 happens at two places. In Line 13 of Algorithm 1 via call to `ExpanditemSets` as well as in Line 11. In Lemma 8, we showed that the call to Algorithm 6 does not eliminate the best item of any user.

So we only need to argue that the step in Line 11 of Algorithm 1 also does not eliminate the best item of any user. To do this, consider a user v with $\pi_v(1) = i$. This implies that the user v is connected to u_i and $v \in \mathcal{U}$ in Line 16 of Algorithm 1 (when processing cluster i) since the best item is not eliminated prior to epoch ℓ. Since by Lemma 8, the users that are connected to user u_i have the item i in their candidate item set, we conclude that $i \in \mathcal{T}_{C_i}^\ell$ at Line 14 of Algorithm 1. This implies that after the call to `ESTIMATE` in Line 17, we must have that $|\mathbf{R}_{u,i}^{\ell+1} - \mathbf{R}_{u,i}| \leq k_\ell$ via Proposition 5.

Now putting it all together yields,

$$\begin{aligned}\tilde{\mathbf{R}}_{u,\tilde{\pi}_u(1)}^{\ell+1} - \tilde{\mathbf{R}}_{u,\pi_u(1)}^{\ell+1} &= \tilde{\mathbf{R}}_{u,\tilde{\pi}_u(1)}^{\ell+1} - \mathbf{R}_{u,\tilde{\pi}_u(1)} + \mathbf{R}_{u,\pi_u(1)} - \tilde{\mathbf{R}}_{u,\pi_u(1)}^{\ell+1} + \mathbf{R}_{u,\tilde{\pi}_u(1)} - \mathbf{R}_{u,\pi_u(1)} \\ &\leq 2k_\ell,\end{aligned}$$

with probability at-least $1 - \delta$ where the last line follows by Proposition 5. Hence the best item is not eliminated with high probability within Stage 2.

□

Lemma 10. *(instantaneous regret in Stage 2) Consider an epoch ℓ withing Stage 2 of Algorithm 1. Then with probability at-least $1 - \delta$, we have that for any user u,*

$$\mathbf{R}_{u,\pi_u(1)} - \mathbf{R}_{u,\rho_u(t)} \leq k_{\ell-1},$$

where t is a time in epoch ℓ and $\rho_u(t)$ is the item recommended to user u as per Algorithm 7.

Proof. Let $\mathcal{T}_u^\ell$ be the item set maintained by Algorithm 1 before the call in Line 17. We know from Lemma 9 that $\pi_u(1) \in \mathcal{T}_u^\ell$ with probability at-least $1 - \delta$. Further due to Line 11 in Algorithm 1 and Line 3 in Algorithm 6, we have that $\tilde{\mathbf{R}}_{u,\pi_u(1)}^\ell - \tilde{\mathbf{R}}_{u,x}^\ell \leq 12k_{\ell-1}$ for any $x \in \mathcal{T}_u^\ell$.

So

$$\begin{aligned}\mathbf{R}_{u,\pi(1)} - \mathbf{R}_{u,\rho_u(t)} &= \mathbf{R}_{u,\pi(1)} - \tilde{\mathbf{R}}_{u,\pi(1)}^\ell + \tilde{\mathbf{R}}_{u,\rho_u(t)}^\ell - \mathbf{R}_{u,\rho_u(t)} + \tilde{\mathbf{R}}_{u,\pi(1)}^\ell - \tilde{\mathbf{R}}_{u,\tilde{\pi}(1)}^\ell + \tilde{\mathbf{R}}_{u,\tilde{\pi}(1)}^\ell - \tilde{\mathbf{R}}_{u,\rho_u(t)}^\ell \\ &\leq \mathbf{R}_{u,\pi(1)} - \tilde{\mathbf{R}}_{u,\pi(1)}^\ell + \tilde{\mathbf{R}}_{u,\rho_u(t)}^\ell - \mathbf{R}_{u,\rho_u(t)} + \tilde{\mathbf{R}}_{u,\tilde{\pi}(1)}^\ell - \tilde{\mathbf{R}}_{u,\rho_u(t)}^\ell \\ &\leq 2k_{\ell-1} + 12k_{\ell-1} \\ &= 14k_{\ell-1},\end{aligned}$$

with probability at-least $1 - \delta$ where we used the matrix completion guarantee from proposition 5.

□

The following result is a direct consequence of Assumption 2.

Proposition 7. *For each hott-topic $i \in [d]$ there is at-least $\kappa\mathsf{M}$ users that have their best item as i, where $\kappa = O(1)$.*

Lemma 11. *Consider an epoch ℓ in Stage 2. Then Algorithm 7 when called in Line 17 of Algorithm 1 executes for a total of $\tilde{O}(\max(1, \mathsf{N}/\mathsf{M})/k_\ell^2)$ rounds.*

Proof. Consider a hott-topic $i \in [d]$. By Lemma 9, the best item for each user survives in its candidate item set. Hence we have that $|\mathcal{U}| \geq \kappa\mathsf{M}$ by Proposition 7, where $\mathcal{U}$ is as in Line 16 of Algorithm 1 while processing cluster C_i.

Next, we proceed to bound the number of rounds run by `ESTIMATE` when it is called at Line 17 of Algorithm 1.

Here we have $|\mathcal{U}| \geq \kappa\mathsf{M}$ and $|\mathcal{T}^{\ell}_{D_i}| \leq \mathsf{N}$. In the case where $\min(|\mathcal{U}|, |\mathcal{T}^{\ell}_{D_i}|) = |\mathcal{T}^{\ell}_{D_i}|$, then by Proposition 5 we have that the number of rounds run by `ESTIMATE` is $\tilde{O}(1/k_\ell^2)$. In the other case where $\min(|\mathcal{U}|, |\mathcal{T}^{\ell}_{D_i}|) = |\mathcal{U}|$, then the number of rounds is $\tilde{O}((\mathsf{N}/|\mathcal{U}|)/k_\ell^2) = \tilde{O}((\mathsf{N}/\mathsf{M})/k_\ell^2)$ since $|\mathcal{U}| \geq \kappa\mathsf{M}$. This concludes the proof.

□

Definition 8. *The minimum sub-optimality gap is defined as:*

$$\Delta := \min_{u \in [\mathsf{M}]} \min_{j \in [\mathsf{N}], j \neq \pi_u(1)} \mathbf{R}_{u,\pi_u(1)} - \mathbf{R}_{u,j}.$$

Lemma 12 (Stage 2 regret bound)**.** *Assume Δ is as defined in Definition 8. Then the total regret incurred in Stage 2 is bounded by $\tilde{O}(\max(1, \mathsf{N}/\mathsf{M})/\Delta)$ with probability at-least $1 - \delta \log T$.*

Proof. Due to Lemma 5, with probability at-least $1 - \delta$, the candidate item sets will only contain the best item for each user whenever $5k_\ell \leq \Delta$ (see Definition 8. Consequently we can say that for epochs $\ell \geq \log(1/(5\Delta))$ the algorithm do not suffer any regret at all.

So we only need to bound the regret till the epoch $\ell^* := \log(1/(5\Delta))$. Again by Lemma 5 and Lemma 11, the regret incurred in an epoch ℓ within stage 2 is bounded by $\tilde{O}(\max(1, \mathsf{N}/\mathsf{M})/k_\ell)$ with probability greater than $1 - \delta$. Thus the total regret across all epochs in stage 2 is

$$\begin{aligned}
\mathbf{R}_{\text{stage 2}} &= \sum_{\ell=\ell_0+1}^{\ell^*} \tilde{O}(\max(1, \mathsf{N}/\mathsf{M})/k_\ell) \\
&= \tilde{O}(\max(1, \mathsf{N}/\mathsf{M})/k_{\ell^*}) \\
&= \tilde{O}(\max(1, \mathsf{N}/\mathsf{M})/\Delta),
\end{aligned}$$

where line 2 is due to the halving progression of k_ℓ. Taking a union bound of the failure probability across all epochs concludes the proof.

□

We are now ready to present the proof of the main result in Setting 1.

Proof of Theorem 1 The proof of the theorem is an immediate consequence of Lemmas 7 and 12. With probability at-least $1 - 2\delta \log T$, we can bound the total regret by $\widetilde{O}\Big(\max\Big(1, \mathsf{N}\mathsf{M}^{-1}\Big)\Big(\Delta_{\mathsf{hott}}^{-2} + \Delta^{-1}\Big) + 1\Big)$. Now doing a change of variables from $2\delta \log T$ to δ yields the theorem.

C.2. Proof of Lemma 1

Proof of Lemma 1. Let us fix a particular user $u \in [\mathsf{M}]$. Assume without loss of generality that $\mathbf{v}^{(1)}$ is the item in $\mathcal{C}$ with the highest reward i.e. $\langle \mathbf{u}, \mathbf{v}^{(1)} \rangle \geq \langle \mathbf{u}, \mathbf{v} \rangle$ for all $\mathbf{v} \in \mathcal{C} \setminus \{\mathbf{v}^{(1)}\}$ In that case for any item vector $\mathbf{z} = \sum_{\mathbf{v} \in \mathcal{C}} \alpha_{\mathbf{v}} \mathbf{v} \in \mathcal{V} \setminus \mathcal{C}$ (non-negative coefficients $\{\alpha_{\mathbf{v}}\}_{\mathbf{v} \in \mathcal{C}}$satisfying $\sum_{\mathbf{v} \in \mathcal{C}} \alpha_{\mathbf{v}} = 1$ and $\alpha_{\mathbf{v}} > 0$ for some $\mathbf{v} \in \mathcal{C} \setminus \{\mathbf{v}^{(1)}\}$), we must have that $\langle \mathbf{u}, \mathbf{v}^{(1)} \rangle > \langle \mathbf{u}, \mathbf{z} \rangle$. Hence, $\mathsf{argmax}_{j \in [\mathsf{N}]} \mathbf{R}_{uj} \in \mathcal{A}$ and a similar argument also shows $\mathsf{argmin}_{j \in [\mathsf{N}]} \mathbf{R}_{uj} \in \mathcal{A}$. □

C.3. Proof of Theorem 2

Recall that our goal is to gradually eliminate columns from $\Pi_r([\mathsf{N}])$ and converge to the set of hott items induced by $\mathcal{A}$ in as few rounds as possible. We propose a successive elimination algorithm (Algorithm DETERMINANTELIM) where at each phase indexed by ℓ, we will maintain a surviving set of r columns denoted by $\mathcal{B}^{(\ell)} \subseteq \Pi_r([\mathsf{N}])$ (initialized by $\mathcal{A}^{(1)} = \Pi_r([\mathsf{N}])$). Since at each round, we need to recommend items to every single user, we will design a recommendation

scheme such that in phase ℓ, with high probability, we can ensure that for every r-column in $\mathcal{B}^{(\ell)}$, we have noisy observations corresponding to at least n_ℓ users that are sampled uniformly from $[\mathsf{M}]$. Concurrently, we also need to ensure that the number of sufficient rounds in phase ℓ satisfies the condition in equation (11). We now prove our key lemma using the probabilistic method:

Lemma 13. *Fix $n_\ell > 0, 0 \leq \delta \leq 1$. Let $\mathcal{B} = \bigcup_{\mathcal{J} \in \mathcal{B}^{(\ell)}} \mathcal{J}$ be the set of columns that are surviving in phase ℓ. For each user $u \in [\mathsf{M}]$, suppose we map d_ℓ distinct items in $\mathcal{B}$ sampled uniformly at random for recommendation in phase ℓ. With probability at least $1-\delta$, for every r-column in $\mathcal{B}^{(\ell)}$, we obtain noisy observations from at least n_ℓ distinct users provided*

$$\frac{d_\ell}{\mathsf{N} - d_\ell} \geq \left(\mathsf{M}^{-1}(12 \log \delta^{-1} + r \log \mathsf{N} + 2n_\ell)\right)^{1/r} \tag{11}$$

Proof of Lemma 13. The items sampled for user u is denoted by $\mathcal{T}_u^{(\ell)} \subseteq \mathcal{B}$. Fix a particular r-column $\mathcal{J} \in \mathcal{B}^{(\ell)}$. Suppose $s = |\mathcal{B}|$. The probability that for any user $u \in [\mathsf{M}]$, all items in $\mathcal{J}$ have been sampled i.e. $\mathcal{J} \subseteq \mathcal{T}_u^{(\ell)}$ is

$$p_\ell \triangleq \Pr(\mathcal{J} \subseteq \mathcal{T}_u^{(\ell)}) = \frac{\binom{s-r}{d_\ell - r}}{\binom{s}{d_\ell}} \leq \frac{\binom{s}{d_\ell - r}}{\binom{s}{d_\ell}} \leq \left(\frac{d_\ell}{s - d_\ell}\right)^r.$$

Notice that the expected number of users for which $\mathcal{J} \subseteq \mathcal{T}_u^{(\ell)}$ is $\mathsf{M}p_\ell$. Suppose p_ℓ is set such that $\mathsf{M}p_\ell \geq 2n_\ell$. Since the d items sampled for each user are independent, the probability of the event $\mathcal{E}_{\mathcal{J}}^{(\ell)}$ that $\mathcal{J}$ is not a subset of $\mathcal{T}_u^{(\ell)}$ for at least n_ℓ users is given by (let us denote the random variable corresponding to the number of users in $[\mathsf{M}]$ for which $\mathcal{J} \in \mathcal{T}_u^{(\ell)}$ by X)

$$\Pr(X \leq n_\ell) \leq \Pr(|X - \mathbb{E}X| \geq \mathsf{M}p_\ell - n_\ell) \leq \Pr(|X - \mathbb{E}X| \geq \frac{\mathsf{M}p_\ell}{2}) \leq 2\exp(-\mathsf{M}p_\ell/12)$$

Hence, the probability that $\mathcal{E}_{\mathcal{J}^{(\ell)}}$ is true for some $\mathcal{J} \in \mathcal{B}^{(\ell)}$ is given by

$$\Pr\Big(\bigcup_{\mathcal{J} \in \mathcal{B}^{(\ell)}} \mathcal{E}_{\mathcal{J}}^{(\ell)}\Big) \leq 2\left|\mathcal{B}^{(\ell)}\right| \exp(-\mathsf{M}p_\ell/12) \leq 2\mathsf{N}^r \exp(-\mathsf{M}p_\ell/12).$$

Therefore, for some value of $0 < \delta < 1$, if we have

$$\mathsf{M}p_\ell \geq 12 \log \delta^{-1} + r \log \mathsf{N} + 2n_\ell$$

then with probability $1-\delta$, the event $\bigcap_{\mathcal{J} \in \mathcal{B}^{(\ell)}} \mathcal{E}_{\mathcal{J}}^{(\ell)}$ is true. Hence, the above condition is satisfied if

$$\frac{d_\ell}{s - d_\ell} \geq \left(\mathsf{M}^{-1}(12 \log \delta^{-1} + r \log \mathsf{N} + 2n_\ell)\right)^{1/r}$$

□

We define the event $\mathcal{E}^{(\ell)}$ which is true when for d_ℓ satisfying eq. (11), for every r-column $\mathcal{J}$ in $\mathcal{B}^{(\ell)}$, there will exist at least $n_{\mathcal{J}} \geq n_\ell$ users $u \in [\mathsf{M}]$ for which $\mathcal{J} \in \mathcal{T}_u^{(\ell)}$. From Lemma 13, we know that $\mathcal{E}^{(\ell)}$ is true with probability at least $1-\delta$. Furthermore, due to our sampling approach being invariant to a permutation of the users, conditioned on the value of $n_{\mathcal{J}}$, the corresponding $n_{\mathcal{J}}$ users are sampled uniformly at random from $[\mathsf{M}]$. In Alg. DETERMINANTELIM, in phase ℓ, once the sets $\mathcal{T}_u^{(\ell)}$ have been sampled, every item in $\mathcal{T}_u^{(\ell)}$ is recommended to user u at least $2\sigma^2$ times (See Step 4). From our recommendation strategy, the total number of rounds sufficient in phase ℓ to make all the recommendations necessary in our scheme is $2\sigma^2 d_\ell$. Now, for any r-column $\mathcal{J}$ in $[\mathsf{N}]^r$, define

$$\mu_{\mathcal{J}} \triangleq \frac{1}{\binom{\mathsf{M}}{r}} \sum_{\mathcal{I} \subset [\mathsf{M}] || \mathcal{I}| = r} \mathsf{det}^2(\mathbf{R}_{\mathcal{I},\mathcal{J}}) < \frac{1}{\binom{\mathsf{M}}{r}} \sum_{\mathcal{I} \subset [\mathsf{M}] || \mathcal{I}| = r} \mathsf{det}(\mathbf{R}_{\mathcal{I},\mathcal{A}}) \triangleq \mu_{\mathcal{A}}.$$

Recall $c_{\mathsf{avg}} \triangleq \frac{1}{\binom{\mathsf{M}}{r}} \sum_{\mathcal{I} \subset [\mathsf{M}] || \mathcal{I}| = r} \mathsf{det}^2(\mathbf{U}_{\mathcal{I}})$ is the average of determinants of the user features taken r at a time - implying that $\mu_{\mathcal{J}} = c_{\mathsf{avg}} \mathsf{det}^2(\mathbf{V}_{\mathcal{J}})$. Moreover, $c_{\max} = \mathsf{det}^2(\mathbf{V}_{\mathcal{A}})$ corresponds to the matrix of hott vectors that has the largest

unsigned determinant. For $\mathcal{J} \in \mathcal{B}^{(\ell)}$, let $\mathcal{H}_{\mathcal{J}}$ be the set of users for which items in $\mathcal{J}$ were recommended in phase ℓ according to our sampling scheme in the first and second components respectively. We compute an estimate $\widetilde{\mu}_{\mathcal{J}}$ of $\mu_{\mathcal{J}}$ in the following way:

$$\widetilde{\mu}_{\mathcal{J}} \triangleq \frac{1}{\binom{n_{\mathcal{J}}}{r}} \sum_{\mathcal{I} \subset \mathcal{H}_{\mathcal{J}} || \mathcal{I}| = r} \mathsf{det}(\widetilde{\mathbf{R}}^{(\ell,1)}_{\mathcal{I},\mathcal{J}}) \cdot \mathsf{det}(\widetilde{\mathbf{R}}^{(\ell,2)}_{\mathcal{I},\mathcal{J}})$$

where the matrices $\widetilde{\mathbf{R}}^{(\ell,1)}, \widetilde{\mathbf{R}}^{(\ell,2)}$ are described in Step 5 of Alg. DETERMINANTELIM. We can now characterize the concentration property of the estimate $\widetilde{\mu}_{\mathcal{J}}$ for all $\mathcal{J} \in \mathcal{B}^{(\ell)}$:

Lemma 14. *Fix n_ℓ and condition on $\mathcal{E}^{(\ell)}$. Then, in phase ℓ of Alg.* DETERMINANTELIM*, we must have that for all $\mathcal{J} \in \mathcal{B}^{(\ell)}$, $\widetilde{\mu}_{\mathcal{J}} \in [\mu_{\mathcal{J}} - \epsilon_\ell, \mu_{\mathcal{J}} + \epsilon_\ell]$ for $\epsilon_\ell = \sqrt{(2^{r+1} r^{1+r/2}) \log(\mathsf{N}^r \mathsf{T}) n_\ell^{-1}}$ with probability at least $1 - O(\mathsf{T}^{-2})$.*

Proof of Lemma 14. We also condition on the event $\mathcal{F}^{(\ell)}$ that the random variables corresponding to the mean noise for all rounds t in phase ℓ is bounded within $[-10\sigma\sqrt{b_\ell^{-2}\log(\mathsf{MT})}, 10\sigma\sqrt{b_\ell^{-2}\log(\mathsf{MT})}]$ - event $\mathcal{F}^{(\ell)}$ is true with probability at least $1 - \mathsf{T}^{10}$ (sub-gaussian concentration properties). We set $b_\ell^2 = 100\sigma^2 \log \mathsf{MT}$ such that noise random variables are bounded within $[-1, 1]$ with probability at least $1 - \mathsf{T}^{-10}$.

Note that there are two sources of randomness in $\widetilde{\mu}_{\mathcal{J}}$ - the first source of randomness is due to the noise added in the observations and the second source is due to the randomness in $\mathcal{H}_{\mathcal{J}}$. Note that

$$\begin{aligned}
\mathbb{E}\widetilde{\mu}_{\mathcal{J}} \mid \mathcal{E}^{(\ell)}, \mathcal{F}^{(\ell)} &= \mathbb{E}_{\mathcal{H}_{\mathcal{J}}} \frac{1}{\binom{n_{\mathcal{J}}}{r}} \sum_{\mathcal{I} \subset \mathcal{H}_{\mathcal{J}} || \mathcal{I}| = r} \mathbb{E}\Big[\mathsf{det}(\widetilde{\mathbf{R}}^{(\ell,1)}_{\mathcal{I},\mathcal{J}}) \cdot \mathsf{det}(\widetilde{\mathbf{R}}^{(\ell,2)}_{\mathcal{I},\mathcal{J}}) \mid \mathcal{H}_{\mathcal{J}}, \mathcal{E}^{(\ell)}, \mathcal{F}^{(\ell)}\Big] \\
&= \mathbb{E}_{\mathcal{H}_{\mathcal{J}}} \frac{1}{\binom{n_{\mathcal{J}}}{r}} \sum_{\mathcal{I} \subset \mathcal{H}_{\mathcal{J}} || \mathcal{I}| = r} \mathbb{E}\Big[\mathsf{det}(\widetilde{\mathbf{R}}^{(\ell,1)}_{\mathcal{I},\mathcal{J}}) \mid \mathcal{H}_{\mathcal{J}}, \mathcal{E}^{(\ell)}, \mathcal{F}^{(\ell)}\Big] \cdot \mathbb{E}\Big[\mathsf{det}(\widetilde{\mathbf{R}}^{(\ell,2)}_{\mathcal{I},\mathcal{J}}) \mid \mathcal{H}_{\mathcal{J}}, \mathcal{E}^{(\ell)}, \mathcal{F}^{(\ell)}\Big] \\
&= \mathbb{E}_{\mathcal{H}_{\mathcal{J}}} \Big[\frac{1}{\binom{n_{\mathcal{J}}}{r}} \sum_{\mathcal{I} \subset \mathcal{H}_{\mathcal{J}} || \mathcal{I}| = r} \mathsf{det}^2(\mathbf{R}_{\mathcal{I},\mathcal{J}}) \mid \mathcal{E}^{(\ell)}\Big].
\end{aligned}$$

We use the fact that conditioning on $\mathcal{E}^{(\ell)}$ only modifies the distribution of $n_{\mathcal{J}}$. Here, we also used the fact that $\mathbb{E}\mathsf{det}(\mathbf{Z}) = \mathsf{det}(\mathbb{E}\mathbf{Z})$ for zero mean independent noise which is the case even after conditioning on $\mathcal{F}^{(\ell)}$ (due to symmetricity of the sub-gaussian random variables). Recall that conditioned on $n_{\mathcal{J}}$, the set $\mathcal{H}_{\mathcal{J}}$ is sampled uniformly at random from $\Pi_{n_{\mathcal{J}}}[\mathsf{M}]$ i.e. sets of all distinct users of size $n_{\mathcal{J}}$. Therefore, we can simplify to conclude that $\mathbb{E}\widetilde{\mu}_{\mathcal{J}} \mid \mathcal{E}^{(\ell)}, \mathcal{F}^{(\ell)} = \mu_{\mathcal{J}}$. By the same set of arguments, we can show that $\mathbb{E}\widetilde{\mu}_{\mathcal{J}} = \mu_{\mathcal{J}}$ - hence, we prove that conditioning on the events $\mathcal{E}^{(\ell)}, \mathcal{F}^{(\ell)}$ do not change the expected value of the estimate $\widetilde{\mu}_{\mathcal{J}}$.

Next, we show that conditioned on the events $\mathcal{E}^{(\ell)}, \mathcal{F}^{(\ell)}$ and on $n_{\mathcal{J}}$, the estimate $\widetilde{\mu}_{\mathcal{J}}$ is concentrated around its mean $\mu_{\mathcal{J}}$ - for which we are going to use McDiarmid's inequality (Vershynin, 2020)[Theorem 2.9.1]. The lemma that we state below is a minor modification of the standard McDiarmid's inequality for uniform sampling without replacement.

Lemma 15 (McDiarmid's inequality for sampling without replacement)**.** *Consider a function $f : \mathcal{X}^n \to \mathbb{R}$ which satisfies the bounded difference property i.e. there exists a constant $c > 0$ such that for all $\mathbf{x}, \mathbf{y} \in \mathcal{X}^n$ satisfying for some $i \in [n]$ 1) $\mathbf{y}_j = \mathbf{x}_j$ for $j \neq i$ and 2) $\mathbf{y}_i \neq \mathbf{x}_i$, we have*

$$|f(\mathbf{x}) - f(\mathbf{y})| \leq c.$$

In that case, for random variables $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \in \mathcal{X}$ sampled uniformly at random without replacement, we must have for $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$,

$$\Pr\Big(|f(\mathbf{x}) - \mathbb{E}\mathbf{x}| \geq \epsilon\Big) \leq 2\exp\Big(-\frac{2\epsilon^2}{nc^2}\Big).$$

Remark 7. *Note that the proof of Lemma 15 is a trivial modification of the proof to the standard McDiarmid's inequality. For the standard McDiarmid's inequality, the random variables $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are sampled independently from $\mathcal{X}$ - this is used crucially for the fact that the probability of any instantiation of the random variables $\mathbf{x}_{i+1}, \ldots, \mathbf{x}_n$ does not depend on the values assigned to the random variables $\mathbf{x}_1, \ldots, \mathbf{x}_i$. In fact this is the only (but crucial) use of the independence. For the*

uniform sampling without replacement case, the independence does not hold - yet, the probability of any instantiation of the random variables $\mathbf{x}_{i+1}, \ldots, \mathbf{x}_n$ is simply $1/\binom{|\mathcal{X}|-i}{n-i}$ and therefore does not depend on the values assigned to the random variables $\mathbf{x}_1, \ldots, \mathbf{x}_i$. Therefore all remaining steps hold true.

In our setting, we let $\mathcal{H}_\mathcal{J}$ (with $|\mathcal{H}_\mathcal{J}| = n_\mathcal{J} \geq n_\ell$) be the set of users sampled for items in $\mathcal{J}$. In that case, we define the function of the observations corresponding to the items in $\mathcal{J}$ and users in $\mathcal{H}_\mathcal{J}$ to be the RHS in eq. (5). Note that because of how we set the value of σ^2, conditioned on the event $\mathcal{F}^{(\ell)}$, each entry of $\widetilde{\mathbf{R}}^{(\ell,1)}, \widetilde{\mathbf{R}}^{(\ell,2)}$ are bounded between $[-1, 2]$. Therefore the maximum determinant $\mathsf{det}_{\max}(r)$ of an $r \times r$ sub-matrix of either of $\widetilde{\mathbf{R}}^{(\ell,1)}, \widetilde{\mathbf{R}}^{(\ell,2)}$ must be bounded from above by $2^r r^{r/2}$ (Hadamard's determinant inequality). Therefore, on altering the set of observations corresponding to a particular user in $\mathcal{H}_\mathcal{J}$ by a different set of observations, conditioned on the event $\mathcal{F}^{(\ell)}$, the change in RHS in eq. (5) is at most

$$\binom{n_\mathcal{J}}{r-1}\Big(\binom{n_\mathcal{J}}{r}\Big)^{-1} 2^r r^{r/2} \leq 2^r r^{1+r/2}(n_\mathcal{J} - r)^{-1} \leq 2^{r+1} r^{1+r/2} n_\mathcal{J}^{-1}$$

where we use the fact that $n_\mathcal{J} \geq n_\ell \geq 2r$. Hence, we can conclude by application of McDiarmid's inequality

$$\Pr\Big(|\widetilde{\mu}_\mathcal{J} - \mu_\mathcal{J}| \geq \epsilon_\ell \mid \mathcal{E}^{(\ell)}, \mathcal{F}^{(\ell)}\Big) \leq 2\exp\Big(-\frac{2n_\mathcal{J}\epsilon_\ell^2}{2^{r+1}r^{1+r/2}}\Big).$$

Therefore, by setting

$$\epsilon_\ell = \sqrt{\frac{(2^{r+1}r^{1+r/2})\log(\mathsf{N}^r\mathsf{T})}{n_\ell}},$$

with probability at least $2\mathsf{N}^{-r}\mathsf{T}^{-2}$, the estimate $\widetilde{\mu}_\mathcal{J}$ must belong to the interval $[\mu_\mathcal{J} - \epsilon_\ell, \mu_\mathcal{J} + \epsilon_\ell]$. Therefore, conditioning on events $\mathcal{E}^{(\ell)}, \mathcal{F}^{(\ell)}$ by taking a union bound over all possible r-columns of $[\mathsf{N}]$, we can conclude that for all surviving r-columns $\mathcal{J}$ in $\mathcal{B}^{(\ell)}$, with our algorithm, the estimates $\widetilde{\mu}_\mathcal{J}$ will be within ϵ_ℓ of their mean $\mu_\mathcal{J}$ for our choice of ϵ_ℓ. Since the event $(\mathcal{F}^{(\ell)})^c$ happens with an order-wise smaller probability ($O(\mathsf{T}^{-10})$ in particular), we can ignore the conditioning on $\mathcal{F}^{(\ell)}$.

□

Let us define $\mathcal{G}^{(\ell)}$ if $\widetilde{\mu}_\mathcal{J} \in [\mu_\mathcal{J} - \epsilon_\ell, \mu_\mathcal{J} + \epsilon_\ell]$ in phase ℓ for our choice of ϵ_ℓ in Lemma 14. Next, we show the following lemma (slightly modified from (Kveton et al., 2017)) to characterize instantaneous regret:

Lemma 16. *Suppose for a user $u \in [\mathsf{M}]$, we recommend a set of r distinct items at round t (denoted by $\mathcal{J} \equiv \{\rho_u(t,j)\}_{j\in[r]}$). In that case, we will have*

$$\mathbf{R}_{u\pi_u(1)} - \max_{j\in[r]} \mathbf{R}_{u\rho_u(t,j)} \leq 6r^{5/2}\Big(\frac{\mathsf{det}^2(\mathbf{V}_\mathcal{A}) - \mathsf{det}^2(\mathbf{V}_\mathcal{J})}{\mathsf{det}^2(\mathbf{V}_\mathcal{A})}\Big).$$

Proof of Lemma 16. Consider any permutation $\nu : [r] \to [r]$. We denote the r-dimensional basis vectors by $\{\mathbf{e}_i\}_{i\in[r]}$. Also, let $\mathbf{V}_\mathcal{A}$ be the matrix whose rows are formed by the hott vectors $\{\mathbf{v}^{(a_j)}\}_{j\in[r]}$. Then any item vector $\mathbf{v}'$ can be written as $\mathbf{V}_\mathcal{A}^\mathsf{T}\mathbf{s}$ where $\mathbf{s}$ is a vector with non-negative entries that sum up to at most 1. For the jth item, we will have $\mathbf{v}^{(j)} = \mathbf{V}_\mathcal{A}^\mathsf{T}\mathbf{s}^{(j)}$.

Note that we have for a user u,

$$\begin{aligned}
\mathbf{R}_{u\pi_u(1)} - \max_{j\in[r]} \mathbf{R}_{u\rho_u(t,\nu(j))} &= (\mathbf{u}^{(u)})^\mathsf{T}\mathbf{v}^{(\pi_u(1))} - \max_{j\in[r]}(\mathbf{u}^{(u)})^\mathsf{T}\mathbf{v}^{(\rho_u(t,\nu(j)))} \\
&= \max_{j\in[r]}(\mathbf{u}^{(u)})^\mathsf{T}\mathbf{v}^{(a_j)} - \max_{j\in[r]}(\mathbf{u}^{(u)})^\mathsf{T}\mathbf{v}^{(\rho_u(t,\nu(j)))} \\
&\le \max_{j\in[r]}\Big|(\mathbf{u}^{(u)})^\mathsf{T}\mathbf{v}^{(a_j)} - \max_{j\in[r]}(\mathbf{u}^{(u)})^\mathsf{T}\mathbf{v}^{(\pi_u(t,\rho(j)))}\Big| \\
&\le \sum_{j\in[r]}\Big|(\mathbf{u}^{(u)})^\mathsf{T}\mathbf{v}^{(a_j)} - (\mathbf{u}^{(u)})^\mathsf{T}\mathbf{v}^{(\rho_u(t,\nu(j)))}\Big| \\
&= \Big|\Big|(\mathbf{u}^{(u)})^\mathsf{T}\mathbf{V}_{\mathcal{A}}^\mathsf{T}\Big|\Big|_2 \sum_{j\in[r]}\Big|\Big|\Big(\mathbf{e}_i - \mathbf{s}^{\rho_u(t,\nu(j))}\Big)\Big|\Big|_2 \\
&\le 2r\sum_{j\in[r]}\Big|\Big|\mathbf{e}_i - \mathbf{s}^{\rho_u(t,\nu(j))}\Big|\Big|_2.
\end{aligned}$$

At this point, we can simply use the lemma in (Kveton et al., 2017) which says the following:

Lemma 17 (Lemma 4 in (Kveton et al., 2017))**.** *There must exist a permutation $\nu : [r] \to [r]$ such that*

$$\sum_{j\in[r]}\Big|\Big|\mathbf{e}_i - \mathbf{s}^{\rho_u(t,\nu(j))}\Big|\Big|_2 \le 6r^{3/2}\Big(\frac{\mathsf{det}^2(\mathbf{V}_{\mathcal{A}}) - \mathsf{det}^2(\mathbf{V}_{\mathcal{J}})}{\mathsf{det}^2(\mathbf{V}_{\mathcal{A}})}\Big)$$

Combining the two statements above, we obtain the desired result. □

Now we are ready to prove Thm 2 using the intermediate Lemmas 13, 14 and 16.

Proof of Theorem 2. Recall that in the first phase $\ell = 1$, we initialize the set of r-columns by $\Pi_r([\mathsf{N}])$. In each phase ℓ, we eliminate subsets of r-columns. Suppose at beginning of phase ℓ, we have $\mathcal{B}^{(\ell)}$ to be the set of surviving r-columns and for each $\mathcal{J} \in \mathcal{B}^{(\ell)}$, we compute $\widetilde{\mu}_{\mathcal{J}}$, an estimate of $\mu_{\mathcal{J}}$. For the subsequent phase $\ell + 1$, we compute $\mathcal{B}^{(\ell+1)}$ in the following way:

$$\mathcal{B}^{(\ell+1)} = \{\mathcal{J} \in \mathcal{B}^{(\ell)} \mid \widetilde{\mu}_{\mathcal{J}} \ge \max_{\mathcal{J}'\in\mathcal{B}^{(\ell)}} \widetilde{\mu}_{\mathcal{J}'} - 2\epsilon_\ell\}.$$

First, we show that if $\mathcal{A}$, the set of hott item vectors belong to $\mathcal{B}^{(\ell)}$, then $\mathcal{A}$ should also belong to $\mathcal{B}^{(\ell+1)}$ conditioned on the event $\mathcal{G}^{(\ell)}$. To see this, note that $\mu_{\mathcal{A}} > \mu_{\mathcal{J}}$ for all $\mathcal{J} \neq \mathcal{A}$. Since, conditioned on the event $\mathcal{G}$, we have $\widetilde{\mu}_{\mathcal{J}} \in [\mu_{\mathcal{J}} - \epsilon_\ell, \mu_{\mathcal{J}} + \epsilon_\ell]$, by triangle inequality, it must happen that $\mathcal{A} \in \mathcal{B}^{(\ell+1)}$. Furthermore, for all $\mathcal{J} \in \mathcal{B}^{(\ell+1)}$, it must happen that (Lemma 14)

$$\mu_{\mathcal{J}} \ge \mu_{\mathcal{A}} - 2\epsilon_\ell \implies \mathsf{det}^2(\mathbf{V}_{\mathcal{J}}) \ge \mathsf{det}^2(\mathbf{V}_{\mathcal{A}}) - \frac{2\epsilon_\ell}{c_{\mathsf{avg}}}.$$

Hence, in all rounds t in phase ℓ, the regret incurred by user u (Lemma 16) is bounded by

$$\mathbf{R}_{u\pi_u(1)} - \max_{j\in[r]} \mathbf{R}_{u\rho_u(t,\rho(j))} \le \frac{12r^{5/2}\epsilon_\ell}{c_{\mathsf{avg}}c_{\mathsf{max}}}$$

In phase $\ell \ge 1$, we set $\epsilon_\ell = 2^{-\ell}$ and $\epsilon_0 = 1$. Hence, we have $n_\ell = C(r)2^{2\ell}$ where $C(r) = 2^{r+1}r^{1+r/2}$. Thus, a sufficient value for d_ℓ and the number of rounds in phase ℓ is (Lemma 13)

$$d_\ell = O\Big(\frac{\mathsf{N}}{\mathsf{M}^{1/r}} \cdot C(r)2^{2\ell}\Big) \text{ and } 2\sigma^2 d_\ell = O\Big(\frac{\sigma^2\mathsf{N}\log(\mathsf{MT})}{\mathsf{M}^{1/r}} \cdot C(r)2^{2\ell}\Big) \text{ respectively.}$$

Hence, if the total number of phases is $m = O(\log \mathsf{T})$ within which all sub-optimal r-columns are eliminated by the algorithm, then by combining the above statements, the regret can be computed as

$$\begin{aligned}\mathsf{Reg}_{\mathsf{Simplified}}(\mathsf{T}) &\leq \sum_{\ell=1}^{m} 2\sigma^2 d_\ell \epsilon_\ell + \mathsf{T}\Pr(\cup_\ell (\mathcal{E}^{(\ell)})^c \cup_\ell (\mathcal{F}^{(\ell)})^c \cup_\ell (\mathcal{G}^{(\ell)})^c) \\ &= O\Big(1 + \frac{\sigma^2 \mathsf{N}\log(\mathsf{MT})}{\mathsf{M}^{1/r}} \cdot \frac{2^{r+1} r^{(r+7)/2}}{c_{\mathsf{avg}} c_{\mathsf{max}} \Delta_{\mathsf{det}}}\Big).\end{aligned}$$

□

D. Discussion on Feasibility of Assumptions

Note that the proof analysis of Theorem 1 generalizes easily (without almost any changes) if the number of hott items is $d > r$ as well. We stress that only for ease of exposition, we have considered the number of hott vectors to be r.

In this section we are going to show evidence that Assumptions 1 and 3 are true together for a large number of rank-r reward matrices $\mathbf{R} \in \mathbb{R}^{\mathsf{M}\times\mathsf{N}}$ with the number of hott topics to be $d = 2r$. In particular, suppose we decompose $\mathbf{R} = \mathbf{X}^\mathsf{T}\mathbf{Y}$ where $\mathbf{X} \in \mathbb{R}^{r\times\mathsf{M}}$, $\mathbf{Y} \in \mathbb{R}^{r\times\mathsf{N}}$ are the transposes of user and item embedding matrices respectively. We consider the generative model where

1. All entries of $\mathbf{X}$ are independently generated from $\mathcal{N}(0,1)$.
2. The first r columns of $\mathbf{Y}$ are the identity matrix multiplied by a factor of $10\log(r\mathsf{N})$. The second r columns of $\mathbf{Y}$ are the identity matrix multiplied by a factor of $-10\log(r\mathsf{N})$. All entries in the remaining $\mathsf{N} - 2r$ columns of $\mathbf{Y}$ are uniformly generated from the interval $[0,1]$.

For the first part of our proof, we will show guarantees on the SVD of $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\mathsf{T}$ where $\mathbf{U} \in \mathbb{R}^{r\times r}, \mathbf{V} \in \mathbb{R}^{\mathsf{M}\times r}$ are orthonormal matrices (singular vectors) and $\mathbf{\Sigma} \in \mathbb{R}^{r\times r}$ is a diagonal matrix corresponding to the singular values of $\mathbf{X}$. More specifically, we have the following lemmas from (Pal et al., 2023) - however, we include the proofs for the sake of completeness. We will denote $\mathbf{X}_{|\mathcal{S}}$ to be the matrix $\mathbf{X}$ restricted to the columns in $\mathcal{S}$ and similarly, we denote $\mathbf{X}_{\mathcal{S}}$ to denote the matrix $\mathbf{X}$ restricted to the rows in $\mathcal{S}$. For a matrix $\mathbf{X}$, we will write $\lambda_j(\mathbf{X})$ to denote the jth singular value of $\mathbf{X}$.

Lemma 18. *[(Pal et al., 2023)] Fix $\gamma > 0$. If $\mathbf{x}^\mathsf{T}\mathbf{X}_{|\mathcal{S}}\mathbf{X}_{|\mathcal{S}}^\mathsf{T}\mathbf{x} \geq \alpha\gamma r\lambda_1^2/\mathsf{M}$ for a subset $\mathcal{S} \subseteq [\mathsf{M}], |\mathcal{S}| = \gamma r$ for all unit vectors $\mathbf{x} \in \mathbb{R}^r$, then the minimum eigenvalue of $\mathbf{V}_{\mathcal{S}}^\mathsf{T}\mathbf{V}_{\mathcal{S}} \geq \alpha\gamma r/\mathsf{M}$.*

Proof. If $\mathbf{x}^\mathsf{T}\mathbf{X}_{|\mathcal{S}}\mathbf{X}_{|\mathcal{S}}^\mathsf{T}\mathbf{x} \geq \alpha\gamma r\lambda_1^2/\mathsf{M}$ for a subset $\mathcal{S} \subseteq [\mathsf{M}], |\mathcal{S}| = \gamma r$ for all unit vectors $\mathbf{x} \in \mathbb{R}^r$, then the minimum eigenvalue of $\mathbf{V}_{\mathcal{S}}^\mathsf{T}\mathbf{V}_{\mathcal{S}} \geq \alpha\gamma r/\mathsf{M}$. To see this, note $\mathbf{X}_{|\mathcal{S}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}_{\mathcal{S}}^\mathsf{T}$ implying that $\mathbf{V}_{\mathcal{S}}^\mathsf{T} = (\mathbf{U}\mathbf{\Sigma})^{-1}\mathbf{X}_{|\mathcal{S}}$. Hence, $\mathbf{V}_{\mathcal{S}}^\mathsf{T}\mathbf{V}_{\mathcal{S}} = (\mathbf{U}\mathbf{\Sigma})^{-1}\mathbf{X}_{|\mathcal{S}}\mathbf{X}_{|\mathcal{S}}^\mathsf{T}(\mathbf{U}\mathbf{\Sigma})^{-\mathsf{T}}$ implying that $(\mathbf{V}_{\mathcal{S}}^\mathsf{T}\mathbf{V}_{\mathcal{S}})^{-1} = (\mathbf{U}\mathbf{\Sigma})^\mathsf{T}(\mathbf{X}_{|\mathcal{S}}\mathbf{X}_{|\mathcal{S}}^\mathsf{T})^{-1}(\mathbf{U}\mathbf{\Sigma})$. Taking the operator norm on both sides, we have $\lambda_{\min}(\mathbf{V}_{\mathcal{S}}^\mathsf{T}\mathbf{V}_{\mathcal{S}}) \geq \lambda_1^{-2}\lambda_{\min}(\mathbf{X}_{|\mathcal{S}}\mathbf{X}_{\mathcal{S}}^{|\mathsf{T}})$ implying that $\mathbf{x}^\mathsf{T}\mathbf{V}_{\mathcal{S}}^\mathsf{T}\mathbf{V}_{\mathcal{S}}\mathbf{x} \geq \alpha\gamma r/\mathsf{M}$. □

Lemma 19. *Suppose $r \ll \mathsf{M}$ and further, the entries of $\mathbf{X}$ are generated independently according to $\mathcal{N}(0,1)$. In that case, it must happen that with high probability that (for $\gamma = 16\log\mathsf{M}$)*

1. *$\lambda_1(\mathbf{X})/\lambda_r(\mathbf{X}) = O(1)$*
2. *Hence $||\mathbf{V}||_{2,\infty} \leq 16\sqrt{\frac{r\log\mathsf{M}}{\mathsf{M}}}$*
3. *$\lambda_r(\mathbf{V}_{\mathcal{S}}) \geq \beta$ for some constant $\beta > 0$ where $\mathcal{S}$ corresponds to a nice subset of users.*

Proof. We must have $\sqrt{\mathsf{M}} - \sqrt{r} - t \leq \lambda_r(\mathbf{X}) \leq \lambda_1(\mathbf{X}) \leq \sqrt{\mathsf{M}} + \sqrt{r} + t$ w.p. at least $1 - 2e^{-t^2/2}$ implying that $\sqrt{\mathsf{M}}/2 \leq \lambda_r \leq \lambda_1 \leq 2\sqrt{\mathsf{M}}$; hence we must have $\lambda_1/\lambda_r = O(1)$ w.p. at least $1 - O(e^{-\mathsf{M}})$. Moreover, we have $\mathbf{X}^\mathsf{T}\mathbf{X} = \mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^\mathsf{T}$. Clearly, we must have $||\mathbf{X}||_{\infty,2}\lambda_r^{-1} \leq ||\mathbf{V}||_{2,\infty} \leq ||\mathbf{X}||_{\infty,2}\lambda_1^{-1}$. For any column $\mathbf{X}_{|i}$, we have $\left|\left|\mathbf{X}_{|\mathbf{i}}\right|\right|_2^2$ is a chi-squared random variable with r degrees of freedom. Using standard concentration inequalities for chi-squared random variables, we have $\left|\left|\mathbf{X}_{|\mathbf{i}}\right|\right|_2 \leq 8\sqrt{r\log\mathsf{M}}$ with probability at least $1 - \mathsf{M}^{-2}$. By taking a union bound over all $i \in [\mathsf{M}]$, we have $||\mathbf{X}||_{\infty,2} \leq 8\sqrt{r\log\mathsf{M}}$ w.p. at least $1 - \mathsf{M}^{-1}$. Hence $||\mathbf{V}||_{2,\infty} \leq 16\sqrt{\frac{r\log\mathsf{M}}{\mathsf{M}}}$. Therefore, when $\mathbf{X}$ is a random Gaussian matrix, the first part of Lemma holds true and the second part of Lemma holds true with $\mu = O(\log\mathsf{M})$ with high probability.

On the other hand, for a subset $\mathcal{S} \subseteq [\mathsf{M}]$ such that $\mathcal{S}$ corresponds to a *nice subset of users* we must have that $|\mathcal{S}| \geq c\mathsf{M}$ for some constant $c > 0$ (with Assumption 2 being true). In that case, we have the minimum singular value of $\mathbf{X}_{|\mathcal{S}}$ to be at least $\sqrt{c\mathsf{M}} - \sqrt{r} - t$ w.p. at least $1 - 2e^{-t^2/2}$. Taking $t = \sqrt{4r \log \mathsf{M}}$, we must have the minimum singular value of $\mathbf{X}_{|\mathcal{S}}$ to be at least $\sqrt{c\mathsf{M}}/2$ with probability at least $1 - 2e^{-2r \log \mathsf{M}}$. Taking a union bound over all nice subsets (at most 2^{C} of them where C is the number of clusters), we must have for all unit norm vectors $\mathbf{x} \in \mathbb{R}^r$

$$\mathbf{x}^\mathsf{T} \mathbf{X}_{|\mathcal{S}} \mathbf{X}_{|\mathcal{S}}^\mathsf{T} \mathbf{x} \geq \frac{c^2 \mathsf{M}}{4} = \frac{c^2 \mathsf{M} \cdot 4\mathsf{M}}{16\mathsf{M}} \geq \frac{c^2 \lambda_1^2}{16}$$

with probability at least $1 - O(e^{-\mathsf{M}})$ implying that $\alpha \geq 1/16$. Combining with Lemma 18, we can prove the final conclusion of the lemma. □

Conjecture: We will have that the minimum singular value of any sub-matrix $\mathbf{Y}_{|\mathcal{S}}$ restricted to the columns in $\mathcal{S}$ of size $|\mathcal{S}| = \gamma r$ to be at least $\frac{r\gamma}{4}$. An open problem to a similar effect has been stated in (Bhojanapalli & Jain, 2014).

Discussion: We take $\gamma = 16 \log \mathsf{N}$. Again, for a subset $\mathcal{S} \subseteq [\mathsf{N}], |\mathcal{S}| = \gamma r$, we must have the minimum singular value of $\mathbf{Y}_{|\mathcal{S}}$ to be at least $\sqrt{r}(\sqrt{\gamma} - 1) - t$ with probability at least $1 - 2e^{-t^2/2}$. Taking $t = 2\sqrt{r \log \mathsf{N}}$, we must have the minimum singular value of $\mathbf{Y}_{|\mathcal{S}}$ to be at least $\sqrt{r\gamma}/2$ with probability at least $1 - 2e^{-2r \log \mathsf{N}}$.

Taking a union bound over all such subsets $\mathcal{S}$ of $[\mathsf{M}]$ with size $|\mathcal{S}| = \gamma r$, we have the failure probability to be $\binom{N}{16r \log \mathsf{N}} e^{-2r \log \mathsf{N}}$ which becomes large - more precisely $\binom{\mathsf{N}}{16r \log \mathsf{N}} \approx e^{r \log^2 \mathsf{N}}$ that is, the factor in the exponent is larger by a multiplicative factor of $\log \mathsf{N}$ than the tail probability. At this point we conjecture that such an analysis is weak as the union bound does not capture the dependencies among the sets due to non-empty intersections. In fact, even theoretically, for most sets, the minimum singular value is going to be large. We leave a tight analysis of this case as a future work. However, if the failure probability for this event is indeed smaller than 1, then we must have for all unit norm vectors $\mathbf{x} \in \mathbb{R}^r$

$$\mathbf{x}^\mathsf{T} \mathbf{Y}_{|\mathcal{S}} \mathbf{Y}_{|\mathcal{S}}^\mathsf{T} \mathbf{x} \geq \frac{r\gamma}{4} = \frac{r\gamma \cdot 4\mathsf{N}}{16\mathsf{N}} \geq \frac{r\gamma\lambda_1^2}{16\mathsf{N}}$$

with high probability at least implying that $\alpha \geq 1/16$. Combining with Lemma 18, we can conclude that if the conjecture is true, then $\lambda_r(\widehat{\mathbf{V}}_{\mathcal{S}}) \geq \sqrt{\frac{r\gamma}{16\mathsf{N}}}$.

Recall that $\mathbf{P} = \mathbf{X}^\mathsf{T}\mathbf{Y} = \mathbf{V}\mathbf{\Sigma}\mathbf{U}^\mathsf{T}\widehat{\mathbf{U}}\widehat{\mathbf{\Sigma}}\widehat{\mathbf{V}}$. We take a further SVD of $\mathbf{\Sigma}\mathbf{U}\widehat{\mathbf{U}}\widehat{\mathbf{\Sigma}} = \widetilde{\mathbf{U}}\widetilde{\mathbf{\Sigma}}\widetilde{\mathbf{V}}$. Therefore, we get that $\mathbf{P} = \mathbf{V}\widetilde{\mathbf{U}}\widetilde{\mathbf{\Sigma}}\widetilde{\mathbf{V}}\widehat{\mathbf{V}}$. Note that $\mathbf{V}\widetilde{\mathbf{U}}$ and $\widetilde{\mathbf{V}}\widehat{\mathbf{V}}$ are orthonormal matrices and therefore $\mathbf{V}\widetilde{\mathbf{U}}\widetilde{\mathbf{\Sigma}}\widetilde{\mathbf{V}}\widehat{\mathbf{V}}$ is a valid SVD of $\mathbf{P}$. The properties of Lemmas 19 and 20 are preserved in the modified singular value matrices $\mathbf{V}\widetilde{\mathbf{U}}$ and $\widetilde{\mathbf{V}}\widehat{\mathbf{V}}$ respectively. Since the condition numbers of both $\mathbf{X}, \mathbf{Y}$ are constants, therefore the condition number of $\mathbf{P}$ is also a constant. Moreover, the $\left|\left|\mathbf{V}\widetilde{\mathbf{U}}\right|\right|_{2,\infty} \leq ||\mathbf{V}||_{2,\infty}$ and $\left|\left|\widetilde{\mathbf{V}}\widehat{\mathbf{V}}\right|\right|_{2,\infty} \leq \left|\left|\widehat{\mathbf{V}}\right|\right|_{2,\infty}$. Finally, for any set $\mathcal{S} \subseteq [\mathsf{M}]$, we will have $\lambda_r(\mathbf{V}_{\mathcal{S}}\widetilde{\mathbf{U}}) = \lambda_r(\mathbf{V})$ and a similar property also holds for $\widehat{\mathbf{V}}$.

Lemma 20. *Assume that the above conjecture is true. Suppose $r \ll \mathsf{N}$ and further, the entries of last $\mathsf{N} - 2r$ columns of $\mathbf{Y}$ are generated independently according to $\mathcal{N}(0, 1)$. In that case, it must happen with high probability that (suppose SVD of $\mathbf{Y} = \widehat{\mathbf{U}}\widehat{\mathbf{\Sigma}}\widehat{\mathbf{V}}$)*

1. $\lambda_1(\mathbf{Y})/\lambda_r(\mathbf{Y}) = O(1)$.
2. *Hence* $\left|\left|\widehat{\mathbf{V}}\right|\right|_{2,\infty} \leq 16\sqrt{\frac{r \log \mathsf{M}}{\mathsf{M}}}$

We move on to a analysis for the matrix $\mathbf{Y}$ which is very similar to the matrix $\mathbf{X}$. Note that Lemma 18 holds true for the matrix $\mathbf{Y}$ as well with M replaced by the number of items N. We now show the following lemma:

First of all, we have that the eigen-spectrum of $\mathbf{Y}$ is same as the eigen-spectrum of $\mathbf{Y}^\mathsf{T}$. Notice that for any unit vector $\mathbf{z}$, we have that

$$\sum_{i=1}^{\mathsf{N}} \mathbf{z}^\mathsf{T} \mathbf{Y}_i \mathbf{Y}_i^\mathsf{T} \mathbf{z} = 20 \log(r\mathsf{N}) + \sum_{i=2r}^{\mathsf{N}} \mathbf{z}^\mathsf{T} \mathbf{Y}_i \mathbf{Y}_i^\mathsf{T} \mathbf{z}.$$

Combining with the analysis of Lemma 19, we must have that, with probability at least $1 - 2e^{-t^2/2}$,

$$\sqrt{\mathsf{N}} - \sqrt{r} - t + \sqrt{20 \log(r\mathsf{N})} \leq \lambda_r(\mathbf{Y}) \leq \lambda_1(\mathbf{Y}) \leq \sqrt{\mathsf{N}} + \sqrt{r} + t + \sqrt{20 \log(r\mathsf{N})}.$$

Therefore, we can again conclude that that $\lambda_1(\mathbf{Y})/\lambda_r(\mathbf{Y}) = O(1)$ with probability at least $1 - O(e^{-\mathsf{M}})$. Furthermore, we also have that $||\mathbf{Y}||_{\infty,2} \leq 8\sqrt{r \log \mathsf{N}}$ with probability at least $1 - \mathsf{N}^{-1}$. Again, if the SVD of the matrix $\mathbf{Y} = \widehat{\mathbf{U}}\widehat{\mathbf{\Sigma}}\widehat{\mathbf{V}}$ where $\widehat{\mathbf{U}}, \widehat{\mathbf{V}}$ are orthonormal matrices, then $\left|\left|\widehat{\mathbf{V}}\right|\right|_{2,\infty} \leq ||\mathbf{Y}||_{\infty,2}\lambda_1(\mathbf{Y})^{-1}$. Since $\lambda_1(\mathbf{Y}) \geq \frac{\sqrt{\mathsf{N}}}{4}$ with probability at least $1 - e^{-\mathsf{N}}$, we have that $\left|\left|\widehat{\mathbf{V}}\right|\right|_{2,\infty} \leq 32\sqrt{\frac{r \log \mathsf{N}}{\mathsf{N}}}$.

Consider some $\mathsf{N}' = c\mathsf{N}$ for a suitable constant $c > 1$ which will be decided later. Suppose we randomly generate N' r-dimensional vectors $\mathbf{y}^1, \mathbf{y}^2, \ldots, \mathbf{y}^{\mathsf{N}'}$ such that every entry of each of these vectors is independently generated from $\mathcal{N}(0,1)$. Now, for some fixed $\epsilon > 0$, consider an ϵ-net of the r-dimensional unit ball $\mathcal{S}_{r,\epsilon}$ such that for any unit vector $\mathbf{x} \in \mathbb{R}^r$, there must exist $\mathbf{z} \in \mathcal{S}_{r,\epsilon}$ such that $||\mathbf{x} - \mathbf{z}||_2 \leq \epsilon$. Now, for any unit fixed vector $\mathbf{z} \in \mathcal{S}_{r,\epsilon}$ and any $i \in [\mathsf{N}']$, we have that the random variable $\mathbf{z}^\mathsf{T}\mathbf{y}^i \sim \mathcal{N}(0,1)$. Further note that the random variables $\{\mathbf{z}^\mathsf{T}\mathbf{y}^i\}_{i=1}^{\mathsf{N}'}$ are independent. Now, for any $X \sim \mathcal{N}(0,1)$, we must have that for any constant $\gamma < 1$

$$\Pr(|X| \leq \gamma) \leq \sqrt{\frac{2}{\pi}} \cdot \gamma.$$

Therefore, for a fixed $\gamma > 0$, consider the random variables $J_1, J_2, \ldots, J_{\mathsf{N}'}$ where $J_i = \mathbb{1}[\left|\mathbf{z}^\mathsf{T}\mathbf{y}^i\right| \leq \gamma]$. Clearly, we have that

$$\sum_{i \in [\mathsf{N}']} \mathbb{E}J_i = \sqrt{\frac{2}{\pi}} \cdot \gamma\mathsf{N}'$$

Since the variables $\{J_i\}_i$ are indicator random variables, we can directly use Chernoff bound to conclude that with probability at least $1 - \exp(-\mathsf{N})$, we have $\sum_i J_i \leq \sqrt{\frac{8}{\pi}} \cdot \gamma\mathsf{N}'$. We will choose $\epsilon = 1/3$ and it is well-known that there exists an $\epsilon-$net $\mathcal{S}_{r,\epsilon}$ satisfying $|\mathcal{S}_{r,\epsilon}| \leq 9^r$. Therefore, if we take a union bound over all 9^r vectors in $\mathcal{S}_{r,\epsilon}$, then we must have with probability $1 - 2^r \exp(-\mathsf{N})$,

$$\sum_{\mathbf{z} \in \mathcal{S}_{r,1/3}} \sum_{i \in [\mathsf{N}']} \mathbb{1}[\left|\mathbf{z}^\mathsf{T}\mathbf{y}^i\right| \leq \gamma] \leq \sqrt{\frac{2}{\pi}} \cdot 9^r \cdot \gamma\mathsf{N}'$$

Recall that r is a small constant and therefore, 9^r is a constant as well. Hence, we choose the constant $\gamma > 0$ such that $\sqrt{\frac{2}{\pi}} \cdot 9^r \cdot \gamma \leq 10^{-1}$. Therefore, we have a subset $\mathcal{T} \subseteq \{\mathbf{y}^1, \ldots \mathbf{y}^{\mathsf{N}'}\}$ of least $9\mathsf{N}'/10$ vectors such that for all $\mathbf{y} \in \mathcal{T}$, we have for the constant $\gamma > 0$ and $\epsilon = 1/3$

$$\left|\mathbf{z}^\mathsf{T}\mathbf{y}\right| \geq \gamma \text{ for all } \mathbf{z} \in \mathcal{S}_{r,\epsilon}.$$

We will use the vectors in $\mathcal{T}$ to be the set of vectors for constructing the last $\mathsf{N} - 2r$ columns of $\mathbf{Y}$.

Now, consider any set $\mathcal{S} \subseteq \mathcal{T}$. Consider the matrix $\mathbf{H}$ whose columns are given by the vectors in $\mathcal{S}$. In that case, we have for all $\mathbf{z} \in \mathcal{S}_{r,\epsilon}$,

$$\mathbf{z}^\mathsf{T}(\mathbf{I} - \mathbf{y}\mathbf{y}^\mathsf{T})\mathbf{z} \leq 1 - \gamma.$$