ONLINE LOW RANK MATRIX COMPLETION

Prateek Jain
Google Research
Bangalore, India
prajain@google.com

Soumyabrata Pal
Google Research
Bangalore, India
soumyabrata@google.com

ABSTRACT

We study the problem of *online* low-rank matrix completion with M users, N items and T rounds. In each round, the algorithm recommends one item per user, for which it gets a (noisy) reward sampled from a low-rank user-item preference matrix. The goal is to design a method with sub-linear regret (in T) and nearly optimal dependence on M and N. The problem can be easily mapped to the standard multi-armed bandit problem where each item is an *independent* arm, but that leads to poor regret as the correlation between arms and users is not exploited. On the other hand, exploiting the low-rank structure of reward matrix is challenging due to non-convexity of the low-rank manifold. We first demonstrate that the low-rank structure can be exploited using a simple explore-then-commit (ETC) approach that ensures a regret of $O(\text{polylog}(\mathsf{M} + \mathsf{N})\mathsf{T}^{2/3})$. That is, roughly only $\text{polylog}(\mathsf{M} + \mathsf{N})$ item recommendations are required per user to get a non-trivial solution. We then improve our result for the rank-1 setting which in itself is quite challenging and encapsulates some of the key issues. Here, we propose OCTAL (Online Collaborative filTering using iterAtive user cLustering) that guarantees nearly optimal regret of $O(\text{polylog}(\mathsf{M} + \mathsf{N})\mathsf{T}^{1/2})$. OCTAL is based on a novel technique of clustering users that allows iterative elimination of items and leads to a nearly optimal minimax rate.

1 INTRODUCTION

Collaborative filtering based on low-rank matrix completion/factorization techniques are the cornerstone of most modern recommendation systems (Koren, 2008). Such systems model the underlying user-item affinity matrix as a low-rank matrix, use the acquired user-item recommendation data to estimate the low-rank matrix and subsequently, use the matrix estimate to recommend items for each user. Several existing works study this *offline* setting (Candès & Recht, 2009; Deshpande & Montanari, 2012; Jain et al., 2013; Chen et al., 2019; Abbe et al., 2020). However, typical recommendation systems are naturally *online* and interactive – they recommend items to users and need to adapt quickly based on users' feedback. The goal of such systems is to quickly identify each user's preferred set of items, so it is necessary to identify the best items for each user instead of estimating the entire affinity matrix. Moreover, items/users are routinely added to the system, so it should be able to quickly adapt to new items/users by using only a small amount of recommendation feedback.

In this work, we study this problem of the online recommendation system. In particular, we study the online version of low-rank matrix completion with the goal of identifying top few items for each user using say only logarithmic many exploratory recommendation rounds for each user. In each round (out of T rounds) we predict one item (out of N items) for each user (out of M users) and obtain feedback/reward for each of the predictions – e.g. did the users view the recommended movie. The goal is to design a method that has asymptotically similar reward to a method that can pick *best* items for each user. As mentioned earlier, we are specifically interested in the setting where $\mathsf{T} \ll \mathsf{N}$ i.e. the number of recommendation feedback rounds is much smaller than the total number of items.

Moreover, we assume that the expected reward matrix is low-rank. That is, if $\mathbf{R}_{ij}^{(t)}$ is the reward obtained in the tth round for predicting item-j for user-i, then $\mathbb{E}\mathbf{R}_{ij}^{(t)} = \mathbf{P}_{ij}$, where $\mathbf{P} \in \mathbb{R}^{\mathsf{M} \times \mathsf{N}}$ is a low-rank matrix. A similar low rank reward matrix setting has been studied in *online* multi-dimensional learning problems (Katariya et al., 2017b; Kveton et al., 2017; Trinh et al., 2020). But in

these problems, the goal is to find the matrix entry with the highest reward. Instead, our goal is to recommend good items to *all* users which is a significantly harder challenge. A trivial approach is to ignore the underlying low-rank structure and solve the problem using standard multi-arm bandit methods. That is, model each *(user, item)* pair as an *arm*. Naturally, that would imply exploration of almost all the items for *each* user, which is also reflected in the regret bound (averaged over users) of $O(\sqrt{\mathsf{N}\mathsf{T}})$ (Remark 1). That is, as expected, the regret bound is vacuous when the number of recommendation rounds $\mathsf{T} \ll \mathsf{N}$.

In contrast, most of the existing online learning techniques that leverage structure amongst arms assume a parametric form for the reward function and require the reward function to be convex in the parameters (Shalev-Shwartz et al., 2011; Bubeck, 2011). Thus, due to the non-convexity of the manifold of low-rank matrices, such techniques do not apply in our case. While there are some exciting recent approaches for *non-convex online learning* (Agarwal et al., 2019; Suggala & Netrapalli, 2020), they do not apply to the above mentioned problem.

Our Techniques and Contributions: We first present a method based on the explore-then-commit (ETC) approach (Algorithm 2). For the first few rounds, the algorithm runs a pure exploration strategy by sampling random items for each user. We use the data obtained by pure exploration to learn a good estimate of the underlying reward matrix $\mathbf{P}$; this result requires a slight modification of the standard matrix completion result in Chen et al. (2019). We then run the exploitation rounds based on the current estimate. In particular, for the remaining rounds, the algorithm commits to the arm with the highest estimated reward for each user. With the ETC algorithm, we achieve a regret bound of $O(\mathrm{polylog}(\mathsf{M} + \mathsf{N})\mathsf{T}^{2/3})$ (Thm. 1). This bound is able to get rid of the dependence on N, implying non-trivial guarantees even when $\mathsf{T} \ll \mathsf{N}$. That is, we require only $\mathrm{polylog}(\mathsf{M} + \mathsf{N})$ exploratory recommendations per user. However, the dependence of the algorithm on T is sub-optimal. To address this, we study the special but critical case of rank-one reward matrix. The rank-1 setting is itself technically challenging (Katariya et al., 2017b) and encapsulates many key issues. We provide a novel algorithm OCTAL in Algorithm 3 and a modified version in Algorithm 8 that achieves nearly optimal regret bound of $O(\mathrm{polylog}(\mathsf{M} + \mathsf{N})\mathsf{T}^{1/2})$ (see Theorems 2 and E). The key insight is that in rank-one case, we need to cluster users based on their true latent representation to ensure low regret.

Our method OCTAL consists of multiple phases of exponentially increasing number of rounds. Each phase refines the current estimate of relevant sub-matrices of the reward matrix using standard matrix completion techniques. Using the latest estimate, we jointly refine the cluster of users and the estimate of the best items for users in each cluster. We can show that the regret in each phase decreases very quickly and since the count of phases required to get the correct clustering is small, we obtain our desired regret bounds. Finally, we show that our method achieves a regret guarantee that scales as $\widetilde{O}(\mathsf{T}^{1/2})$ with the number of rounds T. We also show that the dependence on T is optimal (see Theorem 3). Below we summarize our main contributions ($\widetilde{O}(\cdot)$ hides logarithmic factors):

- We formulate the online low rank matrix completion problem and define the appropriate notion of regret to study. We propose Algorithm 2 based on the explore-then-commit approach that suffers a regret of $\widetilde{O}(\mathsf{T}^{2/3})$. (Thm. 1)
- We propose a novel algorithm OCTAL (Online Collaborative filTering using iterAtive user cLustering) and a modified version (Alg. 3 and Alg. 8 respectively) for the special case of rank-1 reward matrices and guarantee a regret of $\widetilde{O}(\mathsf{T}^{1/2})$. Importantly, our algorithms provide non-trivial regret guarantees in the practical regime of $\mathsf{M} \gg \mathsf{T}$ and $\mathsf{N} \gg \mathsf{T}$ i.e when the number of users and items is much larger than the number of recommendation rounds. Moreover, OCTAL does not suffer an issue of large cold-start time (possibly large exploration period) as in the ETC algorithm.
- We conducted detailed empirical study of our proposed algorithms (see Appendix A) on synthetic and multiple real datasets, and demonstrate that our algorithms can achieve significantly lower regret than methods that do not use collaboration between users. Furthermore, we show that for rank-1 case, while it is critical to tune the exploration period in ETC (as a function of rounds and sub-optimality gaps) and is difficult in practice (Lattimore & Szepesvári, 2020)[Ch. 6], OCTAL still suffers lower regret without such side-information (see Figures 1a and 2b).

Technical Challenges: For rank-1 case, i.e., $\mathbf{P} = \mathbf{u}\mathbf{v}^{\mathsf{T}}$, one can cluster users in two bins: ones with $\mathbf{u}_i \geq 0$ and ones with $\mathbf{u}_i < 0$. Cluster-2 users dislike the best items for Cluster-1 users and vice-versa. Thus, we require algorithms that can learn this cluster structure and exploit the fact that within the same cluster, relative ranking of all items remain the same. This is a challenging information to

exploit, and requires learning the latent structure. Note that an algorithm that first attempts to cluster all users and then optimize the regret will suffer the same $T^{2/3}$ dependence in the worst case as in the ETC algorithm; this is because the difficulty of clustering the users are widely different. Instead, our proposed OCTAL algorithm (Algorithms 3 and 8) in each iteration/phase performs two tasks: i) it tries to eliminate some of the items similar to standard phase elimination method (Lattimore & Szepesvári, 2020)[Ch 6, Ex 6.8] for users that are already clustered, ii) it simultaneously tries to grow the set of clustered users. For partial clustering, we first apply low rank matrix completion guarantees over carefully constructed reward sub-matrices each of which correspond to a cluster of users and a set of active items that have high rewards for all users in the same cluster, and then use the partial reward matrix for eliminating some of the items in each cluster.

1.1 RELATED WORKS

To the best of our knowledge, we provide first rigorous online matrix completion algorithms. But, there are several closely related results/techniques in the literature which we briefly survey below.

A very similar setting was considered in Sen et al. (2017) where the authors considered a multi-armed bandit problem with L contexts and K arms with context dependent reward distributions. The authors assumed that the $L \times K$ reward matrix is low rank and can be factorized into non-negative components which allowed them to use recovery guarantees from non-negative matrix factorization. Moreover, the authors only showed ETC algorithms that resulted in $T^{2/3}$ regret guarantees. Our techniques can be used to improve upon the existing guarantees in Sen et al. (2017) in two ways 1) Removing the assumption of the low rank components being non-negative as we use matrix completion with entry-wise error guarantees. 2) The dependence on T can be improved from $T^{2/3}$ to $T^{1/2}$ when the reward matrix $\mathbf{P}$ is rank-1.

Multi-dimensional online decision making problems namely stochastic rank-1 matrix bandits was introduced in Katariya et al. (2017b;a); Trinh et al. (2020). In their settings, at each round $t \in [T]$, the learning agent can choose one row and one column and observes a reward corresponding to an entry of a rank-1 matrix. Here, the regret is defined in terms of the best *(row ,column)* pair which corresponds to the best arm. This setting was extended to the rank r setting (Kveton et al., 2017), 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). Although these papers provide tight regret guarantees, they cannot be translated to our problem. This is because, we solve a significantly different problem with an underlying rank-1 reward matrix $\mathbf{P}$ where we need to minimize the regret for all users (rows of $\mathbf{P}$) jointly. Hence, it is essential to find the entries (columns) of $\mathbf{P}$ with large rewards for each user(row) of $\mathbf{P}$; contrast this with the multi-dimensional online learning problem where it is sufficient to infer only the entry ((row,column) pair) in the matrix/tensor with the highest reward. Since the rewards for each user have different gaps, the analysis becomes involved for our OCTAL algorithm. Finally, (Dadkhahi & Negahban, 2018; Zhou et al., 2020) also consider our problem setting but they only provide heuristic algorithms without any theoretical guarantees.

Another closely 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). In particular, these papers were the first to motivate and theoretically analyze the collaborative framework with the restriction that the same item cannot be recommended more than once to the same user. Here a significantly stricter cluster structure assumption is made over users where users in same cluster have similar preferences. Such models are restrictive as they provide theoretical guarantees only on a very relaxed notion of regret (termed *pseudo-regret*).

In the past decade, several papers have studied the problem of *offline* low rank matrix completion on its own (Mazumder et al., 2010; 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 (Xu et al., 2013; Ahn et al., 2018; 2021; Elmahdy et al., 2020; Jo & Lee, 2021; Zhang et al., 2022). Some of these results namely the ones that provide $\| \cdot \|_\infty$ norm guarantees on the estimated matrix can be adapted into Explore-Then-Commit (ETC) style algorithms (see Sec. 4). Finally, there is significant amount of related theoretical work for online non-convex learning (Suggala & Netrapalli, 2020; Yang et al., 2018; Huang et al., 2020) and empirical work for online Collaborative Filtering (Huang et al., 2020; Lu et al., 2013; Zhang et al., 2015) but they do not study the regret in online matrix completion setting.

2 PROBLEM DEFINITION

Notations: We write $[m]$ to denote the set $\{1, 2, \ldots, m\}$. For a vector $\mathbf{v} \in \mathbb{R}^m$, $\mathbf{v}_i$ denotes the ith element; for any set $\mathcal{U} \subseteq [m]$, let $\mathbf{v}_\mathcal{U}$ denote the vector $\mathbf{v}$ restricted to the indices in $\mathcal{U}$. $\mathbf{A}_i$ denotes the ith row of $\mathbf{A}$ and $\mathbf{A}_{ij}$ denotes the (i, j)-th element of matrix $\mathbf{A}$. $[n]$ denotes the set $\{1, 2, \ldots, n\}$. For any set $\mathcal{U} \subset [m], \mathcal{V} \subset [n]$, $\mathbf{A}_{\mathcal{U},\mathcal{V}}$ denotes the matrix $\mathbf{A}$ restricted to the rows in $\mathcal{U}$ and columns in $\mathcal{V}$. Also, let $\|\mathbf{A}\|_{2\to\infty}$ be the maximum ℓ_2 norm of the rows of $\mathbf{A}$ and $\|\mathbf{A}\|_\infty$ be the absolute value of the largest entry in $\mathbf{A}$. We write $\mathbb{E}X$ to denote the expectation of a random variable X.

Consider a system with a set of M users and N items. Let $\mathbf{P} = \mathbf{U}\mathbf{V}^\mathsf{T} \in \mathbb{R}^{\mathsf{M}\times\mathsf{N}}$ be the unknown reward matrix of rank $r < \min(\mathsf{M}, \mathsf{N})$ where $\mathbf{U} \in \mathbb{R}^{\mathsf{M}\times r}$ and $\mathbf{V} \in \mathbb{R}^{\mathsf{N}\times r}$ denote the latent embeddings corresponding to users and items respectively. In other words, we can denote $\mathbf{P}_{ij} \triangleq \langle \mathbf{u}_i, \mathbf{v}_j \rangle$ where $\mathbf{u}_i, \mathbf{v}_j \in \mathbb{R}^r$ denotes the r-dimensional embeddings of i-th user and the j-th item, respectively. Often, we will also use the SVD decomposition of $\mathbf{P} = \bar{\mathbf{U}}\Sigma\bar{\mathbf{V}}$ where $\bar{\mathbf{U}} \in \mathbb{R}^{\mathsf{M}\times r}, \bar{\mathbf{V}} \in \mathbb{R}^{\mathsf{N}\times r}$ are orthonormal matrices i.e. $\bar{\mathbf{U}}^\mathsf{T}\bar{\mathbf{U}} = \mathbf{I}$ and $\bar{\mathbf{V}}^\mathsf{T}\bar{\mathbf{V}} = \mathbf{I}$ and $\Sigma \triangleq \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r) \in \mathbb{R}^{r\times r}$ is a diagonal matrix. We will denote the condition number of the matrix $\mathbf{P}$ by $\kappa \triangleq (\max_i \lambda_i)(\min_i \lambda_i)^{-1}$.

Consider a system that recommends one item to every user, in each round $t \in [\mathsf{T}]$. Let, $\mathbf{R}^{(t)}_{u\rho_u(t)}$ be the reward for recommending item $\rho_u(t) \in [\mathsf{N}]$ for user u. Also, let:

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

where $\mathbf{E}^{(t)}_{u\rho_u(t)}$ denotes the unbiased additive noise. Each element of $\{\mathbf{E}^{(t)}_{u\rho_u(t)}\}_{u\in[\mathsf{M}],t\in[\mathsf{T}]}$ is assumed to be i.i.d. zero mean sub-gaussian random variables with variance proxy σ^2. That is, $\mathbb{E}[\mathbf{E}^{(t)}_{u\rho_u(t)}] = 0$ and $\mathbb{E}[\exp(s\mathbf{E}^{(t)}_{u\rho_u(t)})] \leq \exp(\sigma^2 s^2/2)$ for all $u \in [\mathsf{M}], t \in [\mathsf{T}]$. The goal is to minimize the expected regret where the expectation is over randomness in rewards and the algorithm:

$$\mathsf{Reg}(\mathsf{T}) \triangleq \frac{\mathsf{T}}{\mathsf{M}}\sum_{u\in[\mathsf{M}]}\max_{j\in[\mathsf{N}]}\mathbf{P}_{uj} - \mathbb{E}[\sum_{t\in[\mathsf{T}]}\frac{1}{\mathsf{M}}\sum_{u\in[\mathsf{M}]}\mathbf{R}^{(t)}_{u\rho_u(t)}]. \tag{2}$$

In this problem, the interesting regime is $(\mathsf{N}, \mathsf{M}) \gg \mathsf{T}$ as is often the case for most practical recommendation systems. Here, treating each user separately will lead to vacuous regret bounds as each item needs to be observed at least once by each user to find the best item for each user. However, low-rank structure of the rewards can help share information about items across users.

Remark 1. *If $\mathsf{T} \gg \mathsf{N}$, then we can treat each user as a separate multi-armed bandit problem. In that case, in our setting, the well-studied Upper Confidence Bound (UCB) algorithm achieves an expected regret of at most $O(\sigma\sqrt{\mathsf{NT}\log\mathsf{T}})$ (Theorem 2.1 in Bubeck & Cesa-Bianchi (2012)).*

3 PRELIMINARIES

Let us introduce a different observation model from (1). Consider an unknown rank r matrix $\mathbf{P} \in \mathbb{R}^{\mathsf{M}\times\mathsf{N}}$. For each entry $i \in [\mathsf{M}], j \in [\mathsf{N}]$, we observe:

$$\mathbf{P}_{ij} + \mathbf{E}_{ij} \text{ with probability } p, \quad 0 \text{ with probability } 1 - p, \tag{4}$$

where $\mathbf{E}_{ij}$ are independent zero mean sub-gaussian random variables with variance proxy $\sigma^2 > 0$. We now introduce the following result from Chen et al. (2019):

Lemma 1 (Theorem 1 in Chen et al. (2019)). *Let rank $r = O(1)$ matrix $\mathbf{P} \in \mathbb{R}^{d\times d}$ with SVD decomposition $\mathbf{P} = \bar{\mathbf{U}}\Sigma\bar{\mathbf{V}}^\mathsf{T}$ satisfy $\|\bar{\mathbf{U}}\|_{2,\infty} \leq \sqrt{\mu r/d}, \|\bar{\mathbf{V}}\|_{2,\infty} \leq \sqrt{\mu r/d}$ and condition number $\kappa = O(1)$. Let $1 \geq p \geq C\mu^2 d^{-1}\log^3 d$ for some sufficiently large constant $C > 0$, $\sigma = O\left(\sqrt{\frac{pd}{\mu^3\log d}}\|\mathbf{P}\|_\infty\right)$. Suppose we observe noisy entries of $\mathbf{P}$ according to observation model in (4). Then, with probability exceeding $1 - O(d^{-3})$, we can compute a matrix $\widehat{\mathbf{P}}$ by using Algorithm 4 (Appendix B) with parameters $(\mathcal{U} = [\mathsf{M}], \mathcal{V} = [\mathsf{N}], \sigma^2, r, p)$ s.t.,*

$$\|\widehat{\mathbf{P}} - \mathbf{P}\|_\infty \leq O\left(\frac{\sigma}{\min_i \lambda_i} \cdot \sqrt{\frac{\mu d\log d}{p}}\|\mathbf{P}\|_\infty\right). \tag{5}$$

Algorithm 1 ESTIMATE

Require: Set of users $\mathcal{U} \subseteq [\mathsf{M}]$, set of items $\mathcal{V} \subseteq [\mathsf{N}]$, total rounds m, set of indices $\Omega \subseteq \mathcal{U} \times \mathcal{V}$, rounds in each iteration $b = \max_{u \in \mathcal{U}} |v \in \mathcal{V} \mid (u, v) \in \Omega|$, regularization parameter λ. Index of round t is relative to the first round when the algorithm is invoked; hence $t = 1, 2, \dots, m$.

1: **for** $\ell = 1, 2, \dots, m/b$ **do**
2:     For all $(i, j) \in \Omega$, set $\mathsf{Mask}_{ij} = 0$.
3:     **for** $\ell' = 1, 2, \dots, b$ **do**
4:         **for** each user $u \in \mathcal{U}$ in round $t = (\ell - 1)b + \ell'$ **do**
5:             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 $\mathbf{R}^{(t)}_{u\rho_u(t)}$.
6:         **end for**
7:     **end for**
8: **end for**
9: 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. $\mathbf{Z}_{uj} = \mathrm{avg}\{\mathbf{R}^{(t)}_{u\rho_u(t)}$ for $t \in [m] \mid \rho_u(t) = j\}$. Discard all other observations corresponding to indices not in Ω.
10: 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 $\mathcal{V}^{(1)}, \mathcal{V}^{(2)}, \dots, \mathcal{V}^{(k)}$ 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 $\mathcal{U}$.*
11: **for** $q \in [k]$ **do**
12:     Solve convex program
$$\min_{\mathbf{Q}^{(q)} \in \mathbb{R}^{|\mathcal{U}| \times |\mathcal{V}^{(q)}|}} \frac{1}{2} \sum_{(i,j) \in \Omega^{(q)}} \left(\mathbf{Q}^{(q)}_{i\pi(j)} - \mathbf{Z}_{ij}\right)^2 + \lambda \|\mathbf{Q}^{(q)}\|_\star, \tag{3}$$

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

Note there are several difficulties in using Lemma 1 directly in our setting which are discussed below:

Remark 2 (Matrix Completion for Rectangular Matrices). *Lemma 1 is described for square matrices and a trivial approach to use Lemma 1 for rectangular matrices with M rows and N columns (say $\mathsf{N} \geq \mathsf{M}$) by appending $\mathsf{N} - \mathsf{M}$ zero rows leads to an undesirable $(\mathsf{N}/\mathsf{M})^{1/2}$ factor (Lemma 5) in the error bound (the $(\mathsf{N}/\mathsf{M})^{1/2}$ factor does not arise if we care about spectral/Frobenius norm instead of L_∞ norm). One way to resolve the issue is to partition the columns into N/M groups by assigning each column into one of the groups uniformly at random. Thus, we create N/M matrices which are almost square and apply Lemma 1 to recover an estimate that is close in L_∞ norm. Thus we can recover an estimate of the entire matrix which is close in L_∞ norm up to the desired accuracy without suffering the undesirable $(\mathsf{N}/\mathsf{M})^{1/2}$ factor (Lemma 6 and Steps 10-12 in Algorithm 1).*

Remark 3 (Observation Models). *The observation model in equation 4 is significantly different from equation 1. In the former, a noisy version of each element of $\mathbf{P}$ is observed independently with probability p while in the latter, in each round $t \in [\mathsf{T}]$, for each user $u \in [\mathsf{M}]$, we observe noisy version of a chosen element $\rho_u(t)$. Our approach to resolve this discrepancy theoretically is to first sample a set Ω of indices according to equation 4 and subsequently use equation 1 to observe the indices in Ω (see Steps 3-6 in Algorithm 1 and Corollary 1). Of course, this implies obtaining observations corresponding to indices in a super-set of Ω (see Step 5 in Algorithm 1) and only using the observations in Ω for obtaining an estimate of the underlying matrix. In practice, this is not necessary and we can use all the observed indices to obtain the estimate in Step 12 of Algorithm 1.*

Remark 4 (Repetition and Median Tricks). *The smallest error that is possible to achieve by using Lemma 1 is by substituting $p = 1$ and thereby obtaining $\|\widehat{\mathbf{P}} - \mathbf{P}\|_\infty \leq O\Big(\sigma(\min_i \lambda_i)^{-1} \cdot \sqrt{\mu d \log d} \|\mathbf{P}\|_\infty\Big)$ and moreover, the probability of failure is polynomially small in the dimension d; however, this is insufficient when d is not large enough. Two simple tricks allow us to resolve this issue: 1) First we can obtain repeated observations from the same entry of the reward matrix and take its average; s repetitions can bring down the noise variance to σ^2/s 2) Second, we can use the*

median trick where we obtain several independent estimates of the reward matrix and compute the element-wise median to boost the success probability (see proof of Lemma 2).

We address all these issues (see Appendix B for detailed proofs) and arrive at the following lemma:

Lemma 2. *Let rank $r = O(1)$ reward matrix $\mathbf{P} \in \mathbb{R}^{\mathsf{M} \times \mathsf{N}}$ with SVD decomposition $\mathbf{P} = \bar{\mathbf{U}} \mathbf{\Sigma} \bar{\mathbf{V}}^{\mathsf{T}}$ satisfy $\|\bar{\mathbf{U}}\|_{2,\infty} \leq \sqrt{\mu r / \mathsf{M}}, \|\bar{\mathbf{V}}\|_{2,\infty} \leq \sqrt{\mu r / \mathsf{N}}$ and condition number $\kappa = O(1)$. Let $d_1 = \max(\mathsf{M}, \mathsf{N})$, $d_2 = \min(\mathsf{M}, \mathsf{N})$ such that $d_2 = \Omega(\mu r \log(r d_2))$ and $1 \geq p \geq C \mu^2 d_2^{-1} \log^3 d_2$ for sufficiently large constant $C > 0$. Suppose we observe noisy entries of $\mathbf{P}$ according to observation model in (1). For any positive integer $s > 0$ satisfying $\frac{\sigma}{\sqrt{s}} = O\left(\sqrt{\frac{p d_2}{\mu^3 \log d_2}} \|\mathbf{P}\|_\infty\right)$, there exists an algorithm $\mathcal{A}$ with parameters s, p, σ that uses $m = O\left(s \log(\mathsf{MN}\delta^{-1})(\mathsf{N}p + \sqrt{\mathsf{N}p \log \mathsf{M}\delta^{-1}})\right)$ rounds to compute a matrix $\widehat{\mathbf{P}}$ such that with probability exceeding $1 - O(\delta \log(\mathsf{MN}\delta^{-1}))$*

$$\|\mathbf{P} - \widehat{\mathbf{P}}\|_\infty \leq O\left(\frac{\sigma r}{\sqrt{s d_2}} \sqrt{\frac{\mu^3 \log d_2}{p}}\right). \tag{6}$$

Remark 5. *Alg. $\mathcal{A}$ repeats the following process $O(\log(\mathsf{MN}\delta^{-1}))$ times: 1) sample subset of indices $\Omega \subseteq [\mathsf{M}] \times [\mathsf{N}]$ such that every $(i,j) \in [\mathsf{M}] \times [\mathsf{N}]$ is inserted into Ω independently with probability p. 2) By setting $b = \max_{i \in [\mathsf{M}]} |j \in [\mathsf{N}] \mid (i,j) \in \Omega|$, Algorithm $\mathcal{A}$ invokes Alg. 1 with total rounds bs, number of rounds in each iteration b, set Ω, set of users $[\mathsf{M}]$, items $[\mathsf{N}]$ and regularization parameter $\lambda = C_\lambda \sigma \sqrt{\min(\mathsf{M}, \mathsf{N})p}$ for a suitable constant $C_\lambda > 0$ in order to compute an estimate of $\mathbf{P}$. The final estimate $\widehat{\mathbf{P}}$ is computed by taking an entry-wise median of each individual estimate obtained as output from several invocations of Alg. 1. Alternatively, Alg. $\mathcal{A}$ is detailed in Alg. 7 in Appendix B.*

Note that the total number of noisy observations made from the matrix $\mathbf{P}$ is $m \cdot \mathsf{M} \geq \mathsf{MN} \cdot p \cdot s$. Therefore, informally speaking, the average number of observations per index is $p \cdot s$ which results in an error of $\widetilde{O}(\sigma / \sqrt{sp})$ ignoring other terms (contrast with error $\widetilde{O}(\sigma / \sqrt{p})$ in equation 5.)

Remark 6 (Setting parameters s, p in lemma 2)**.** *Lemma 2 has three input parameters namely $s \in \mathbb{Z}, 0 \leq p \leq 1$ and $0 \leq \delta \leq 1$. For any set of input parameters (η, ν), our goal is to set s, p, δ as functions of known σ, r, μ, d_2 such that we can recover $\|\mathbf{P} - \widehat{\mathbf{P}}\|_\infty \leq \eta$ with probability $1 - \nu$ for which the conditions on σ and p are satisfied. From (17), we must have $\sqrt{sp} = \frac{c\sigma r}{\sqrt{d_2}} \frac{\sqrt{\mu^3 \log d_2}}{\eta}$ for some appropriate constant $c > 0$. If $r = O(1)$ and $\eta \leq \|\mathbf{P}\|_\infty$, then an appropriate choice of c also satisfies the condition $\frac{\sigma}{\sqrt{s}} = O\left(\sqrt{\frac{p d_2}{\mu^3 \log d_2}} \|\mathbf{P}\|_\infty\right)$. More precisely, we are going to set $p = C\mu^2 d_2^{-1} \log^3 d_2$ and $s = \left\lceil \left(\frac{c\sigma r \sqrt{\mu}}{\eta \log d_2}\right)^2 \right\rceil$ in order to obtain the desired guarantee.*

4 EXPLORE-THEN-COMMIT (ETC) ALGORITHM

In this section, we present an Explore-Then-Commit (ETC) based algorithm for online low-rank matrix completion. The algorithm has two disjoint phases of exploration and exploitation. We will first jointly explore the set of items for all users for a certain number of rounds and compute an estimate $\widehat{\mathbf{P}}$ of the reward matrix $\mathbf{P}$. Subsequently, we commit to the estimated best item found for each user and sample the reward of the best item for the remaining rounds in the exploitation phase for that user. Note that the exploration phase involves using a matrix completion estimator in order to estimate the entire reward matrix $\mathbf{P}$ from few observed entries. Our regret guarantees in this framework is derived by carefully balancing exploration phase length and the matrix estimation error (detailed proof provided in Appendix C).

Theorem 1. *Consider the rank-r online matrix completion problem with M users, N items, T recommendation rounds. Set $d_2 = \min(\mathsf{M}, \mathsf{N})$. Let $\mathbf{R}_{u\rho_u(t)}^{(t)}$ be the reward in each round, defined as in equation 1. Suppose $d_2 = \Omega(\mu r \log(r d_2))$. Let $\mathbf{P} \in \mathbb{R}^{\mathsf{M} \times \mathsf{N}}$ be the expected reward matrix that satisfies the conditions stated in Lemma 2 , and let σ^2 be the noise variance in rewards. Then, Algorithm 2, applied to the online rank-r matrix completion problem guarantees the following regret:*

$$\mathsf{Reg}(\mathsf{T}) = O\left(\left(\mathsf{T}^{\frac{2}{3}}(\sigma^2 r^2 \|\mathbf{P}\|_\infty)^{\frac{1}{3}}\left(\frac{\mu^3 \mathsf{N} \log d_2}{d_2}\right)^{1/3} + \frac{\mathsf{N}\mu^2 \|\mathbf{P}\|_\infty}{d_2}\right) \log^5(\mathsf{MNT}) + \frac{\|\mathbf{P}\|_\infty}{\mathsf{T}^2}\right). \tag{7}$$

Algorithm 2 ETC ALGORITHM

Require: users M, items N, rounds T, noise σ^2, rank r of $\mathbf{P}$, upper bound on magnitude of expected
rewards $\|\mathbf{P}\|_\infty$, no. of estimates $f = O(\log(MNT))$.

1: Set $d_2 = \min(M, N)$ and $v = (N\|\mathbf{P}\|_\infty)^{-2/3}\left(\frac{T\sigma r}{\sqrt{d_2}}\sqrt{\mu^3 \log d_2}\right)^{2/3}$. Set $p = C\mu^2 d_2^{-1}\log^3 d_2$,
   $s = \lceil vp^{-1}\rceil$ and $\lambda = C_\lambda \sigma\sqrt{d_2 p}$ for some constants $C, C_\lambda > 0$.
2: **for** $k = 1, 2, \ldots, f$ **do**
3:    For each tuple of indices $(i, j) \in [M] \times [N]$, independently set $\delta_{ij} = 1$ with probability p and
      $\delta_{ij} = 0$ with probability $1 - p$.
4:    Denote $\Omega = \{(i, j) \in [M] \times [N] \mid \delta_{ij} = 1\}$ and $b = \max_{i \in [M]} | \mid j \in [N] \mid (i, j) \in \Omega|$ to be
      the maximum number of index tuples in a particular row. Set total number of rounds to be bs.
5:    Compute the kth estimate $\widehat{\mathbf{P}}^{(k)} = \text{ESTIMATE}([M], [N], bs, \Omega, b, \lambda)$. # *(Algorithm 1 is used to*
      recommend items to every user for bs rounds.
6: **end for**
7: Compute final estimate estimate $\widehat{\mathbf{P}}$ by taking the entry-wise median of $\widehat{\mathbf{P}}^{(1)}, \widehat{\mathbf{P}}^{(2)}, \ldots, \widehat{\mathbf{P}}^{(f)}$.
8: **for** each of remaining rounds **do**
9:    Recommend $\text{argmax}_{j \in [N]}\widehat{\mathbf{P}}_{ij}$ for each user $i \in [M]$. # *Number of remaining rounds is $T - bsf$.*
10: **end for**

Remark 7 (Non-trivial regret bounds). *Theorem 1 provides non-trivial regret guarantees in the key
regime when $N \gg T$ and $M > N$ where the regret scales only logarithmically on M, N. This is
intuitively satisfying since in each round we are obtaining M observations, so more users translate
to more information which in-turn allows better understanding of the underlying reward matrix.
However, the dependence of regret on T (namely $T^{2/3}$) is sub-optimal. In the subsequent section, we
provide a novel algorithm to obtain regret guarantees with $T^{1/2}$ for rank-1 $\mathbf{P}$.*

Remark 8 (Gap dependent bounds). *Define the minimum gap to be $\Delta = \min_{u \in [M]}\left|\mathbf{P}_{u\pi_u(1)} - \mathbf{P}_{u\pi_u(2)}\right|$ where $\pi_u(1), \pi_u(2)$ corresponds to the items with the highest
and second highest reward for user u respectively. If the quantity Δ is known then it is possible to
design ETC algorithms where length of the exploration phase is tuned accordingly in order to obtain
regret bounds that scale logarithmically with the number of rounds T.*

5 OCTAL ALGORITHM

In this section we present our algorithm OCTAL (Algorithm 3) for online matrix completion where
the reward matrix $\mathbf{P}$ is rank 1. The set of users is described by a latent vector $\mathbf{u} \in \mathbb{R}^M$ and the set of
items is described by a latent vector $\mathbf{v} \in \mathbb{R}^N$. Thus $\mathbf{P} = \mathbf{u}\mathbf{v}^\mathsf{T}$ with SVD decomposition $\mathbf{P} = \bar{\lambda}\bar{\mathbf{u}}\bar{\mathbf{v}}^\mathsf{T}$.

Algorithm Overview: Our first key observation is that as $\mathbf{P}$ is rank-one, we can partition the set of
users into two disjoint clusters $\mathcal{C}_1, \mathcal{C}_2$ where $\mathcal{C}_1 \equiv \{i \in [M] \mid \mathbf{u}_i \geq 0\}$ and $\mathcal{C}_2 \equiv [M] \setminus \mathcal{C}_1$. Clearly,
for all users $u \in \mathcal{C}_1$, the item that results in maximum reward is $j_{\max} = \text{argmax}_{t \in [N]}\mathbf{v}_t$. On the other
hand, for all users $u \in \mathcal{C}_2$, the item that results in maximum reward is $j_{\min} = \text{argmin}_{t \in [N]}\mathbf{v}_t$. Thus,
if we can identify $\mathcal{C}_1, \mathcal{C}_2$ and estimate items with high reward (identical for users in the same cluster)
using few recommendations per user, we can ensure low regret.

But, initially $\mathcal{C}_1, \mathcal{C}_2$ are unknown, so all users are *unlabelled* i.e., their cluster is unknown. In
each phase (the outer loop indexed by ℓ), Algorithm 3 tries to label at least a few unlabelled users
correctly. This is achieved by progressively refining estimate $\widetilde{\mathbf{Q}}$ of the reward matrix $\mathbf{P}$ restricted to
the unlabelled users and all items (Step 12). Subsequently, unlabelled users for which the difference
in maximum and minimum reward (inferred from estimated reward matrix) is large are labelled
(Step 19). At the same time, in Step 13 users labelled in previous phases are partitioned into two
clusters (denoted by $\mathcal{M}^{(\ell,1)}$ and $\mathcal{M}^{(\ell,2)}$) and for each of them, the algorithm refines an estimate of
two distinct sub-matrices of the reward matrix $\mathbf{P}$ by recommending items only from a refined set
($\mathcal{N}^{(\ell,1)}$ and $\mathcal{N}^{(\ell,2)}$ respectively) containing the best item ($j_{\max}$ or $j_{\min}$). We also identify a small set
of *good* items for each labelled user (including users labelled in previous phases), which correspond
to large estimated rewards. We partition all these users into two clusters ($\mathcal{M}^{(\ell+1,1)}$ and $\mathcal{M}^{(\ell+1,2)}$)
such that the set of *good* items for users in different clusters are disjoint. We can prove that such a
partitioning is possible; users in same cluster have same sign of user embedding.

Algorithm 3 OCTAL (ONLINE COLLABORATIVE FILTERING USING ITERATIVE USER CLUSTERING)

Require: Number of users M, items N, rounds T, noise σ^2, bound on the entry-wise magnitude of expected rewards $\|\mathbf{P}\|_\infty$, incoherence μ.

1: Set $\mathcal{M}^{(1,1)} = \mathcal{M}^{(1,2)} = \phi$ and $\mathcal{B}^{(1)} = [\mathsf{M}]$. Set $\mathcal{N}^{(1,1)} = \mathcal{N}^{(1,2)} = \phi$. Set $f = O(\log(\mathsf{MNT}))$ and suitable constants $a, c, C, C', C_\lambda > 0$.

2: **for** $\ell = 1, 2, \ldots,$ **do**

3:     Set $\Delta_\ell = C' 2^{-\ell} \min\left(\|\mathbf{P}\|_\infty, \frac{\sigma\sqrt{\mu}}{\log \mathsf{N}}\right)$.

4:     **for** $k = 1, 2, \ldots, f$ **do**

5:         **for** each pair of non-null sets $(\mathcal{B}^{(\ell)}, \mathsf{N}), (\mathcal{M}^{(\ell,1)}, \mathcal{N}^{(\ell,1)}), (\mathcal{M}^{(\ell,2)}, \mathcal{N}^{(\ell,2)}) \subseteq [\mathsf{M}] \times [\mathsf{N}]$ **do**

6:             Denote $(\mathcal{T}^{(1)}, \mathcal{T}^{(2)})$ to be the considered pair of sets and $i \in \{0, 1, 2\}$ to be its index.

7:             Set $d_{2,i} = \min(|\mathcal{T}^{(1)}|, |\mathcal{T}^{(2)}|)$. Set $p_{\ell,i} = C\mu^2 d_{2,i}^{-1} \log^3 d_{2,i}$ and $s_{\ell,i} = \left\lceil \left(\frac{c\sigma\sqrt{\mu}}{\Delta_\ell \log d_{2,i}}\right)^2 \right\rceil$.

8:             For each tuple of indices $(u, v) \in \mathcal{T}^{(1)} \times \mathcal{T}^{(2)}$, independently set $\delta_{uv} = 1$ with probability $p_{\ell,i}$ and $\delta_{uv} = 0$ with probability $1 - p_{\ell,i}$.

9:             Denote $\Omega^{(i)} = \{(u, v) \in \mathcal{T}^{(1)} \times \mathcal{T}^{(2)} \mid \delta_{uv} = 1\}$ and $b_{\ell,i} = \max_{u \in \mathcal{U}} |v \in \mathcal{V} \mid (u, v) \in \Omega|$. Set total number of rounds to be $m_{\ell,i} = b_{\ell,i} s_{\ell,i}$.

10:         **end for**

11:         Set $m_\ell = \max_{i \in \{0,1,2\}} m_{\ell,i}$.

12:         Compute $\widetilde{\mathbf{Q}}^{(\ell,k)} = \text{ESTIMATE}(|\mathcal{B}^{(\ell)}|, [\mathsf{N}], m_\ell, \Omega^{(0)}, b_{\ell,0}, \lambda = C_\lambda \sigma \sqrt{d_{2,0} p_\ell})$. # *Algorithm 1 is used to recommend items to every user in $\mathcal{B}^{(\ell)}$ for m_ℓ rounds.*

13:         For $i \in \{1, 2\}$, compute $\widetilde{\mathbf{P}}^{(\ell,i,f)} = \text{ESTIMATE}(|\mathcal{M}^{(\ell,i)}|, |\mathcal{N}^{(\ell,i)}|, m_\ell, \Omega^{(i)}, b_{\ell,i}, \lambda = C_\lambda \sigma \sqrt{d_{2,i} p_\ell})$. # *Algorithm 1 recommends items to every user in $\mathcal{M}^{(\ell,i)}$ for m_ℓ rounds.*

14:     **end for**

15:     Compute $\widetilde{\mathbf{Q}}^{(\ell)} = $ Entrywise Median$(\{\widetilde{\mathbf{Q}}^{(\ell,k)}\}_{k=1}^f)$, $\widetilde{\mathbf{P}}^{(\ell,i)} = $ Entrywise Median$(\{\widetilde{\mathbf{P}}^{(\ell,i,k)}\}_{k=1}^f)$ for $i \in \{1, 2\}$.

16:     Set $\mathcal{B}^{(\ell+1)} \equiv \left\{u \in \mathcal{B}^{(\ell)} \mid \left|\max_{t \in [\mathsf{N}]} \widetilde{\mathbf{Q}}_{ut}^{(\ell)} - \min_{t \in [\mathsf{N}]} \widetilde{\mathbf{Q}}_{ut}^{(\ell)}\right| \leq 2a\Delta_\ell\right\}$

17:     Compute $\mathcal{T}_u^{(\ell+1)} = \{j \in [\mathsf{N}]\} \mid \widetilde{\mathbf{Q}}_{uj}^{(\ell)} + \Delta_\ell > \max_{t \in [\mathsf{N}]} \widetilde{\mathbf{Q}}_{ut}^{(\ell)}\}$ for all $u \in \mathcal{B}^{(\ell)} \setminus \mathcal{B}^{(\ell+1)}$.

18:     For $i \in \{1, 2\}$, for all users $u \in \mathcal{M}^{(\ell,i)}$, compute $\mathcal{T}_u^{(\ell+1)} = \{j \in \mathcal{N}^{(\ell,i)} \mid \widetilde{\mathbf{P}}_{uj}^{(\ell,i)} + \Delta_\ell > \max_{t \in \mathcal{N}^{(\ell,i)}} \widetilde{\mathbf{P}}_{ut}^{(\ell,i)}\}$.

19:     Set v to be any user in $[\mathsf{M}] \setminus \mathcal{B}^{(\ell+1)}$. Set $\mathcal{M}^{(\ell+1,1)} = \{u \in [\mathsf{M}] \setminus \mathcal{B}^{(\ell+1)} \mid \mathcal{T}_u^{(\ell+1)} \cap \mathcal{T}_v^{(\ell+1)} \neq \phi\}$. Set $\mathcal{M}^{(\ell+1,2)} = [\mathsf{M}] \setminus (\mathcal{B}^{(\ell+1)} \cup \mathcal{M}^{(\ell+1,1)})$.

20:     Compute $\mathcal{N}^{(\ell+1,1)} = \bigcap_{u \in \mathcal{M}^{(\ell+1,1)}} \mathcal{T}_u^{(\ell+1)}$, $\mathcal{N}^{(\ell+1,2)} = \bigcap_{u \in \mathcal{M}^{(\ell+1,2)}} \mathcal{T}_u^{(\ell+1)}$.

21:     For $i \in \{1, 2\}$, if $|\mathcal{M}^{(\ell+1,i)}| \leq \frac{\mathsf{M}}{\sqrt{\mathsf{T}}}$, then set $\mathcal{B}^{(\ell+1)} \leftarrow \mathcal{B}^{(\ell+1)} \cup \mathcal{M}^{(\ell+1,i)}$ and $\mathcal{M}^{(\ell+1,i)} \leftarrow \phi$.

22: **end for**

We also prove that the set of good items contain the best item for each labelled user ($j_{\max}$ or $j_{\min}$). So, after each phase, for each cluster of users, we compute the intersection of *good* items over all users in the cluster. This subset of items (*joint good* items) must contain the best item for that cluster and therefore we can discard the other items (Step 20). We can show that all items in the set of *joint good* items ($\mathcal{N}^{(\ell+1,1)}$ and $\mathcal{N}^{(\ell+1,2)}$) have rewards which is close to the reward of the best item. Therefore the algorithm suffers small regret if for each group of labelled users, the algorithm recommends items from the set of *joint good* items (Step 13) in the next phase. We can further show that for the set of unlabelled users, the difference in rewards between the best item and worst item is small and hence the regret for such users is small, irrespective of the recommended item (Step 12). Note that until the number of labelled users is sufficiently large, we do not consider them separately (Step 21). A crucial part of our analysis is to show that for any subset of users and items considered in Step 5, the number of rounds sufficient to recover a good estimate of the expected reward sub-matrix is small irrespective of the number of considered items (if the number of users is sufficiently large).

Remark 9 (Practical considerations). *In general OCTAL (Alg. 3) is computationally faster than the ETC Algorithm (Alg. 2) with a higher exploration length. This is because OCTAL eliminates large chunks of items in every phase and therefore has to solve easier optimization problems; on the other hand, ETC has to solve a low rank matrix completion problem in* MN *variables that becomes slower with the exploration length (datapoints). Moreover, OCTAL algorithm runs in phases with the initial*

phases being very small; hence the users do not have to wait for a long time to even get personalized recommendations like in ETC. These features make OCTAL much more practical than ETC.

To summarize, in Algorithm 3, the entire set of rounds $[\mathsf{T}]$ is partitioned into phases of exponentially increasing length. In each phase, for the set of unlabelled users, we do pure exploration and recommend random items from the set of all possible items (Step 12). The set of labelled users are partitioned into two clusters; for each, we follow a semi-exploration strategy where we recommend random items from a set of *joint good* items (Steps 13). We now introduce the following definition:

Definition 3 ((α, μ)-Local Incoherence). *For $0 \le \alpha \le 1$, a vector $\mathbf{v} \in \mathbb{R}^m$ is (α, μ)-local incoherent if for all sets $\mathcal{U} \subseteq [m]$ satisfying $|\mathcal{U}| \ge \alpha m$, we have $\|\mathbf{v}_{\mathcal{U}}\|_\infty \le \sqrt{\frac{\mu}{|\mathcal{U}|}} \|\mathbf{v}_{\mathcal{U}}\|_2$.*

Local incoherence for a vector $\mathbf{v}$ implies that any sub-vector of $\mathbf{v}$ having a significant size must be incoherent as well. Note that the local incoherence condition is trivially satisfied if the magnitude of each vector entry is bounded from below. We are now ready to state our main result:

Theorem 2. *Consider the rank-1 online matrix completion problem with T rounds, M users s.t. $\mathsf{M} \ge \sqrt{\mathsf{T}}$ and N items. Denote $d_2 = \min(\mathsf{M}, \mathsf{N})$. Let $\mathbf{R}^{(t)}_{u \rho_u(t)}$ be the reward in each round, defined as in equation 1. Let σ^2 be the noise variance in rewards and let $\mathbf{P} \in \mathbb{R}^{\mathsf{M} \times \mathsf{N}}$ be the expected reward matrix with SVD decomposition $\mathbf{P} = \lambda \bar{\mathbf{u}} \bar{\mathbf{v}}^\top$ such that $\bar{\mathbf{u}}$ is $(\mathsf{T}^{-1/2}, \mu)$-locally incoherent, $\|\bar{\mathbf{v}}\|_\infty \le \sqrt{\mu/\mathsf{N}}$, $d_2 = \Omega(\mu \log d_2)$ and $|\bar{\mathbf{v}}_{j_{\min}}| = \Theta(|\bar{\mathbf{v}}_{j_{\max}}|)$. Then, by suitably choosing parameters $\{\Delta_\ell\}_\ell$, positive integers $\{s_{(\ell, 0)}, s_{(\ell, 1)}, s_{(\ell, 2)}\}_\ell$ and $1 \ge \{p_{(\ell, 0)}, p_{(\ell, 1)}, p_{(\ell, 2)}\}_\ell \ge 0$ as described in Algorithm 3, we can ensure a regret guarantee of $\mathsf{Reg}(\mathsf{T}) = O(\sqrt{\mathsf{T}} \|\mathbf{P}\|_\infty + \mathsf{J}\sqrt{\mathsf{TV}})$ where $\mathsf{J} = O\left(\log\left(\frac{1}{\sqrt{\mathsf{T}^{-1}}} \min\left(\|\mathbf{P}\|_\infty, \frac{\sigma\sqrt{\mu}}{\log \mathsf{N}}\right)\right)\right)$ and $\mathsf{V} = \left(\max(1, \frac{\mathsf{N}\sqrt{\mathsf{T}}}{\mathsf{M}})\sigma^2 \mu^3 \log^2(\mathsf{MNT})\right)$.*

Similar to Algorithm 2, Algorithm 3 allows non-trivial regret guarantees even when $\mathsf{N} \gg \mathsf{T}$ provided the number of users is significantly large as well i.e. $\mathsf{M} = \widetilde{\Omega}(\mathsf{N}\sqrt{\mathsf{T}})$.

Remark 10. *Under slightly stronger local incoherence conditions on the vector $\bar{\mathbf{u}}$, we can analyze a modified version of OCTAL (Alg. 8) without requiring users M to be large. When $|\mathcal{C}_1| \approx |\mathcal{C}|_2$, the regret guarantee (Thm. E) scales as $\widetilde{O}(\sqrt{\mathsf{NT}/\mathsf{M}})$. Due to space limitations, details of Algorithm 8 and the proof of Theorem E can be found in Appendix E*

Finally, we show that the above dependence on $\mathsf{N}, \mathsf{M}, \mathsf{T}$ matches the lower bound that we obtain by reduction to the well-known multi-armed bandit problem.

Theorem 3. *Let $\mathbf{P} \in [0, 1]^{\mathsf{M} \times \mathsf{N}}$ be a rank 1 reward matrix and the noise variance $\sigma^2 = 1$. In that case, any algorithm for online matrix completion problem will suffer regret of $\Omega(\sqrt{\mathsf{NTM}^{-1}})$.*

6 CONCLUSIONS

We studied the problem of online rank-one matrix completion in the setting of repeated item recommendations and blocked item recommendations, which should be applicable for several practical recommendation systems. We analyzed an explore-then-commit (ETC) style method which is able to get the regret averaged over users to be nearly independent of number of items. That is, per user, we require only logarithmic many item recommendations to get non-trivial regret bound. But, the dependence on the number of rounds T is sub-optimal. We further improved this dependence by proposing OCTAL that carefully combines exploration, exploitation and clustering for different users/items. Our methods iteratively refines estimate of the underlying reward matrix, while also identifying users which can be recommended certain items confidently. Our algorithms and proof techniques are significantly different than existing bandit learning literature. We believe that our work only scratches the surface of an important problem domain with several open problems. For example, Algorithm 3 requires rank-1 reward matrix. Generalizing the result to rank-r reward matrices would be interesting. Furthermore, relaxing assumptions on the reward matrix like stochastic noise or additive noise should be relevant for several important settings. Finally, collaborative filtering can feed users related items, hence might exacerbate their biases. Our method might actually help mitigate the bias due to explicit exploration, but further investigation into such challenges is important.

REFERENCES

Emmanuel Abbe, Jianqing Fan, Kaizheng Wang, and Yiqiao Zhong. Entrywise eigenvector analysis of random matrices with low expected rank. *Annals of statistics*, 48(3):1452, 2020.

Naman Agarwal, Alon Gonen, and Elad Hazan. Learning in non-convex games with an optimization oracle. In *Conference on Learning Theory*, pp. 18–29. PMLR, 2019.

Junhyung Ahn, Adel Elmahdy, Soheil Mohajer, and Changho Suh. On the fundamental limits of matrix completion: Leveraging hierarchical similarity graphs. *arXiv preprint arXiv:2109.05408*, 2021.

Kwangjun Ahn, Kangwook Lee, Hyunseung Cha, and Changho Suh. Binary rating estimation with graph side information. *Advances in neural information processing systems*, 31, 2018.

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

Guy Bresler, George H. Chen, and Devavrat Shah. 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.

Guy Bresler, Devavrat Shah, and Luis Filipe Voloch. 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.

Sébastien Bubeck. Introduction to online optimization. *Lecture Notes*, 2, 2011.

Sébastien Bubeck and Nicolo Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *arXiv preprint arXiv:1204.5721*, 2012.

Emmanuel J Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717–772, 2009.

Yuxin Chen, Yuejie Chi, Jianqing Fan, Cong Ma, and Yuling Yan. Noisy matrix completion: Understanding statistical guarantees for convex relaxation via nonconvex optimization. *arXiv preprint arXiv:1902.07698*, 2019.

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

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

Adel Elmahdy, Junhyung Ahn, Changho Suh, and Soheil Mohajer. Matrix completion with hierarchical graph side information. *Advances in neural information processing systems*, 33:9061–9074, 2020.

Ken Goldberg, Theresa Roeder, Dhruv Gupta, and Chris Perkins. Eigentaste: A constant time collaborative filtering algorithm. *information retrieval*, 4(2):133–151, 2001.

Botao Hao, Jie Zhou, Zheng Wen, and Will Wei Sun. Low-rank tensor bandits. *arXiv preprint arXiv:2007.15788*, 2020.

Reinhard Heckel and Kannan Ramchandran. 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.

Baihe Huang, Kaixuan Huang, Sham Kakade, Jason D Lee, Qi Lei, Runzhe Wang, and Jiaqi Yang. Optimal gradient-based algorithms for non-concave bandit optimization. *Advances in Neural Information Processing Systems*, 34:29101–29115, 2021.

Xiao-Yu Huang, Bing Liang, and Wubin Li. Online collaborative filtering with local and global consistency. *Information Sciences*, 506:366–382, 2020.

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

Prateek Jain and Purushottam Kar. Non-convex optimization for machine learning. *arXiv preprint arXiv:1712.07897*, 2017.

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

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

Kwang-Sung Jun, Rebecca Willett, Stephen Wright, and Robert Nowak. Bilinear bandits with low-rank structure. In *International Conference on Machine Learning*, pp. 3163–3172. PMLR, 2019.

Sumeet Katariya, Branislav Kveton, Csaba Szepesvári, Claire Vernade, and Zheng Wen. Bernoulli rank-1 bandits for click feedback. *arXiv preprint arXiv:1703.06513*, 2017a.

Sumeet Katariya, Branislav Kveton, Csaba Szepesvari, Claire Vernade, and Zheng Wen. Stochastic rank-1 bandits. In *Artificial Intelligence and Statistics*, pp. 392–401. PMLR, 2017b.

Yehuda Koren. 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.

Branislav Kveton, Csaba Szepesvári, Anup Rao, Zheng Wen, Yasin Abbasi-Yadkori, and S Muthukrishnan. Stochastic low-rank bandits. *arXiv preprint arXiv:1712.04644*, 2017.

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

Jing Lu, Steven Hoi, and Jialei Wang. Second order online collaborative filtering. In *Asian Conference on Machine Learning*, pp. 325–340. PMLR, 2013.

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

Rahul Mazumder, Trevor Hastie, and Robert Tibshirani. Spectral regularization algorithms for learning large incomplete matrices. *The Journal of Machine Learning Research*, 11:2287–2322, 2010.

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

Rajat Sen, Karthikeyan Shanmugam, Murat Kocaoglu, Alex Dimakis, and Sanjay Shakkottai. Contextual bandits with latent confounders: An nmf approach. In *Artificial Intelligence and Statistics*, pp. 518–527. PMLR, 2017.

Shai Shalev-Shwartz et al. Online learning and online convex optimization. *Foundations and trends in Machine Learning*, 4(2):107–194, 2011.

Arun Sai Suggala and Praneeth Netrapalli. Online non-convex learning: Following the perturbed leader is optimal. In *Algorithmic Learning Theory*, pp. 845–861. PMLR, 2020.

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

Joel A Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics*, 12(4):389–434, 2012.

Miao Xu, Rong Jin, and Zhi-Hua Zhou. Speedup matrix completion with side information: Application to multi-label learning. *Advances in neural information processing systems*, 26, 2013.

Lin Yang, Lei Deng, Mohammad H Hajiesmaili, Cheng Tan, and Wing Shing Wong. An optimal algorithm for online non-convex learning. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 2(2):1–25, 2018.

Feng Zhang, Ti Gong, Victor E Lee, Gansen Zhao, and Guangzhi Qu. Simple is beautiful: An online collaborative filtering recommendation solution with higher accuracy. In *Asia-Pacific Web Conference*, pp. 497–508. Springer, 2015.

Qiaosheng Zhang, Geewon Suh, Changho Suh, and Vincent YF Tan. Mc2g: An efficient algorithm for matrix completion with social and item similarity graphs. *IEEE Transactions on Signal Processing*, 2022.

Jie Zhou, Botao Hao, Zheng Wen, Jingfei Zhang, and Will Wei Sun. Stochastic low-rank tensor bandits for multi-dimensional online decision making. *arXiv e-prints*, pp. arXiv–2007, 2020.

Cai-Nicolas Ziegler, Sean M McNee, Joseph A Konstan, and Georg Lausen. Improving recommendation lists through topic diversification. In *Proceedings of the 14th international conference on World Wide Web*, pp. 22–32, 2005.
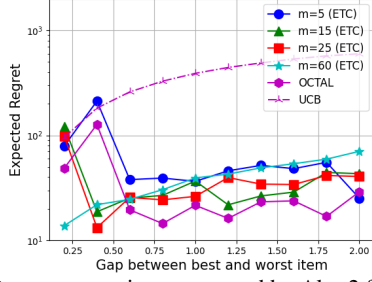
(a) Regret comparison computed by Alg. 2 for exploration periods $(5, 15, 45, 50)$, Alg. 3 and UCB when $\mathsf{T} = 1000$. The average regret is plotted with gap in reward between best and worst item for all users.
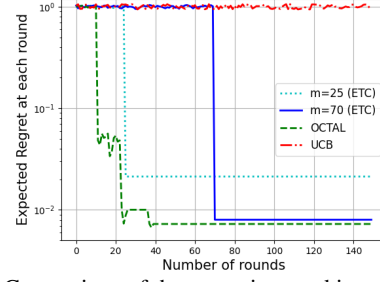
(b) Comparison of the regret incurred in each round $(\mathsf{M}^{-1} \sum_{u \in [\mathsf{M}]} \boldsymbol{\mu}_u^\star - \mathbb{E} \frac{1}{\mathsf{M}} \sum_{u \in [\mathsf{M}]} \mathbf{R}_{u \rho_u(t)}^{(t)})$ by Algorithm 2 (for exploration period $m = 25, 70$), Algorithm 3 and the UCB algorithm when $\mathsf{T} = 1000$.

Figure 1: Comparison of regret incurred by OCTAL algorithm (Algorithm 3, and ETC algorithm 2 against UCB, when expected reward matrix is rank-1. See Section A.1 for data generation process. Clearly, both OCTAL and ETC have significantly lower regret than baseline UCB method which does not exploit reward matrix's low-rank structure.

A EXPERIMENTS

A.1 SYNTHETIC DATASETS

We set the total number of users $\mathsf{M} = 100$, the total number of items $\mathsf{N} = 150$. For the total number of rounds, we look at two settings namely 1) $\mathsf{T} = 1000$ ($\mathsf{T} \gg \mathsf{N}$) and 2) $\mathsf{T} = 100$ ($\mathsf{T} \ll \mathsf{N}$). The former one allows us to demonstrate the tension between exploration and exploitation clearly as the number of rounds is large and moreover, it also allows us to compare with the Upper Confidence Bound (UCB) algorithm (see Remark 1) that does not use the low rank structure of the algorithm. In the latter setting, we demonstrate empirically that our regret bounds are non-trivial even when the number of rounds is significantly smaller than the number of items. We design the unknown reward matrix $\mathbf{P} = \mathbf{u}\mathbf{v}^\mathsf{T}$ by designing $\mathbf{u} \in \mathbb{R}^\mathsf{M}, \mathbf{v} \in \mathbb{R}^\mathsf{N}$ corresponding to the embeddings of the users and items respectively in the following manner: the entries of $\mathbf{u}$ are randomly sampled from the set $\{1, -1\}$. All the entries in the vector $\mathbf{v}$ are uniformly sampled from the interval $[\mathsf{Gap}/2, -\mathsf{Gap}/2]$ where Gap is a parameter unknown to the implemented algorithms. With such a construction, we ensure almost surely that the difference between the reward of the best item and second best item also increases with Gap. The algorithms are allowed to make observations from a noisy version of $\mathbf{P}$ (say $\mathbf{P}_{\mathsf{noisy}}$) where $\mathbf{P}_{\mathsf{noisy}} \leftarrow \mathbf{P} + \mathbf{E}$ where every entry of $\mathbf{E}$ is sampled i.i.d according to $\mathcal{N}(0, 0.1)$.

We vary Gap and run Algorithm 2 with different exploration periods m. For each configuration of Gap, m we run Algorithm 2 10 times and store the regret (see the relevant definition in equation 1) averaged across the 10 runs. In Figure 2b, we compare the regret of Algorithm 2 with different values of Gap as we change the exploration periods for $\mathsf{T} = 1000$. Clearly, with a small exploration period, the algorithm commits to the wrong item more often and with a large exploration period, the cost of exploration is too high. The existence of a sweet spot can be noticed from the U-curves in Figure 2b.

Similarly, we run Algorithm 3 for the same values of Gap, each for 10 times and store the average regret. We also run the Upper Confidence Bound $((1, 1)$-UCB) algorithm Bubeck & Cesa-Bianchi (2012) for each user separately and compute the regret (averaged over 10 simulations). In Figure 1a, we compare the regrets incurred by Algorithm 2, Algorithm 3 and the UCB algorithm when $\mathsf{T} = 1000$. Similarly, in Figure 2a, we compare the regrets incurred by Algorithm 2 and Algorithm 3 when $\mathsf{T} = 100$. Note that when $\mathsf{T} = 1000$, the tension between exploration and exploitation is clearer as the number of exploitation rounds is larger for Algorithm 2; Algorithm 3 performs better than ETC algorithms for all the different values of exploration periods used for many of the different values of Gap. However, even for $\mathsf{T} = 100$, the performance of Algorithm 3 is quite close to that of Algorithm 2 with the best performing exploration period parameter.
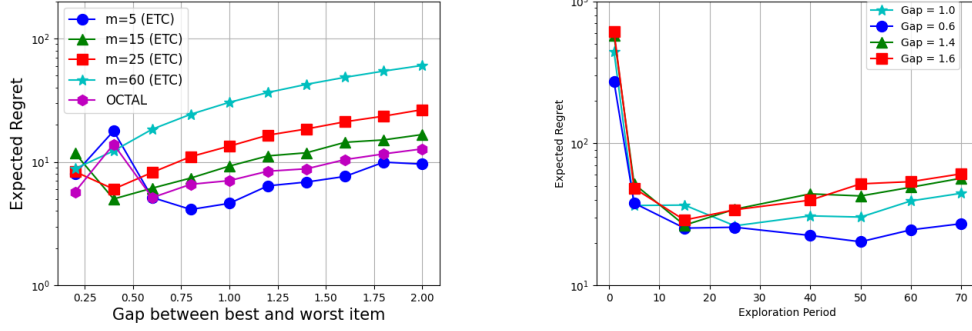
(a) Comparison of the regret computed by Algorithm 2 for exploration periods $5, 15, 25, 60$ ($5, 15$ are the optimal exploration periods for Algorithm 2) and Algorithm 3 when $\mathsf{T} = 100$. The average regret is plotted with gap in reward between best and worst item for all users.

(b) Comparison of the regret computed by Algorithm 2 for 4 different gaps in reward between best and worst item when $\mathsf{T} = 1000$. The average regret is plotted with different exploration periods for four different values of Gap resulting in U-curves.

Figure 2: Simulation results on Regret averaged over a few runs (See Defintion 2) of Algorithms 2 and 3

For Algorithm 3, in phase indexed by ℓ, we use the number of rounds to be $m_\ell = 10 + 2^\ell$. We also make Step 11 in Algorithm 3 more robust by taking those items in $\mathcal{N}^{(\ell+1,1)}, \mathcal{N}^{(\ell+1,2)}$ that are present in at least $2/3$ of the set $\mathcal{T}_u^{(\ell+1)}$ for users u in the sets $\mathcal{M}^{(\ell+1,1)}, \mathcal{M}^{(\ell+1,2)}$ respectively. If, due to some failure event, $\mathcal{N}^{(\ell+1,1)} = \phi$ (or $\mathcal{N}^{(\ell+1,2)} = \phi$) then we set $\mathcal{N}^{(\ell+1,1)}$ ($\mathcal{N}^{(\ell+1,2)}$) to be the item that is present is maximum number of sets $\mathcal{T}_u^{(\ell+1)}$ for users u in the sets $\mathcal{M}^{(\ell+1,1)}$ ($\mathcal{M}^{(\ell+1,2)}$).

In our final experiment in the repeated setting with $\mathsf{T} = 1000$, we plot the regret $\mathsf{M}^{-1} \sum_{u \in [\mathsf{M}]} \max_{j \in [\mathsf{N}]} \mathbf{P}_{uj} - \frac{1}{\mathsf{M}} \sum_{u \in [\mathsf{M}]} \mathbf{P}_{u\rho_u(t)}$ at time t (up to $t < 150$) as t is increased for Algorithms 2, 3 and the UCB Algorithm (the reported values are averaged over 10 simulations). For Algorithm 2, the regret is high at all times during the exploration period and then it experiences a sharp drop in the exploitation period. On the other hand, for Algorithm 3, the regret decreases in each phase and so we observe a more gradual decrease in the regret; thus Algorithm 3 is an anytime algorithm. Also note that the regret of the UCB algorithm does not decrease during the first 150 rounds.

A.2 REAL DATASETS

MovieLens: Next, we demonstrate experimental results on the MovieLens dataset on the lines of Kveton et al. (2017). The MovieLens 1M dataset comprises of 1 million ratings provided by 6K users to 4K movies. We cluster the users into disjoint groups (241) where each cluster represents a unique combination of gender, age group, and occupation in the MovieLens dataset. Moreover, we use a random subset of movies (300) for which the average rating is between 2.5 and 3.5 (in order to avoid movies which are too good or too bad). For each pair of (user group, movie), we take the average of all the ratings provided by users in that group for that movie. Around 33% of the 241×300 ratings matrix could be filled in this manner. In order to complete the matrix, we optimized a convex program for low rank matrix completion (see eq. (6) and (Chen et al., 2019)) by minimizing the MSE with a nuclear norm regularizer.

We randomly take $\mathsf{M} = 128$ user groups, $\mathsf{N} = 128$ movies and take the total rounds $\mathsf{T} = 100$. Note that the number of rounds is less than the number of items (movies) and therefore, the UCB algorithm implemented separately for each user group can only explore and incur a regret of 179.61. For the ETC Algorithm (Alg. 2), as a baseline (for a fair comparison with OCTAL), we consider the rank 1-approximation of the estimated reward matrix $\widehat{\mathbf{P}}$ (ETC Rank-1). From Figure 3a, note that the OCTAL Algorithm (Alg. 3 with the same minor mofications as in the synthetic datasets) outperforms ETC with the rank-1 approximation by a significant margin. In fact, in Figure 3c, we can see that the OCTAL Algorithm has smaller cumulative regret in every round as compared to ETC with rank-1

14

approximation for different exploration periods. All results that are reported are an average of 10 independent runs.

Although the OCTAL algorithm crucially uses the rank 1 structure, it has reasonable performance even when the reward matrix has a larger rank but can still be approximated well with a rank 1 matrix; in our experiment, it turns out that the highest singular value of the ratings matrix is ≈ 10 times larger than the second largest singular value. In figure 3b, we also compare the round-wise incurred regret at different rounds $t \in [100]$.

(a) Comparison of regret at $\mathsf{T} = 100$ for UCB, OCTAL and ETC (with rank-1 approxmiation) for different exploration periods

(b) Comparison of regret at every round for OCTAL and ETC with rank-1 approximation and exploration period $m = 20, 40$

(c) Comparison of cumulative regret until every round for OCTAL and ETC with rank-1 approximation and different exploration periods
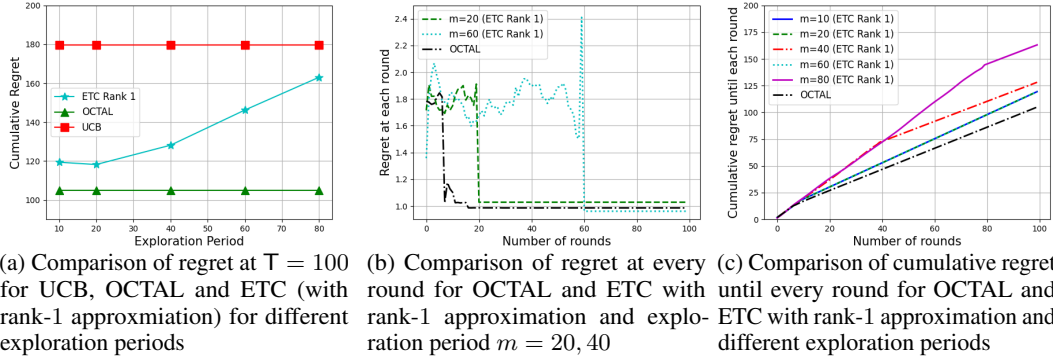
Figure 3: Comparison of regret incurred by OCTAL algorithm (Algorithm 3, ETC algorithm 2 with rank-1 approximation of the estimated matrix against UCB for the MovieLens dataset. In Figure 3a, we show that OCTAL and both versions of ETC have significantly lower regret than baseline UCB method. In Figure 3b, we show that the regret of OCTAL decreases in every phase unlike ETC algorithms. In Figure 3c, we show that the cumulative regret of OCTAL is lower than that of ETC with rank 1 approximation at every round $t \le \mathsf{T}$.

Notice from Figure 3b that the regret of the OCTAL algorithm decreases continuously making it anytime which is extremely useful in practical applications. On the other hand, the regret of the ETC algorithm only decreases after the exploration period is over which is undesirable. In fact, for a large number of initial rounds, the OCTAL algorithm outperforms the ETC algorithms.

Jester Dataset: Next, we consider the Jester dataset (Goldberg et al., 2001) which consists data from 24983 users who have provided ratings for 100 jokes that are between -10.0 to $+10.0$. We select $\mathsf{M} = 100$ users who have rated all the $\mathsf{N} = 100$ jokes and use the corresponding 100×100 ratings matrix as the underlying reward matrix; we select the number of recommendation rounds $\mathsf{T} = 100$ However, this reward matrix is not well-approximated by a low rank matrix unlike the MovieLens dataset since there exists a long tail of singular values with large magnitude. Despite this, we demonstrate the efficacy of our algorithms. First, we run the UCB algorithm for each of the 100 users separately for 100 rounds and thus incur an average regret of 734.66.

As in the Movielens setting, we run Algorithm 2 namely the ETC Algorithm (with rank-1 approximation) for different exploration periods and the OCTAL algorithm (Algorithm 3 with the same minor mofications as in the synthetic datasets) in this setting. Note from Figure 4a that both the ETC and OCTAL algorithm have improved performances as compared to the UCB algorithm. Again, from Figure 4c, the cumulative regret of OCTAL is lower than that of ETC with rank-1 approximation for different exploration periods at every round. In Figure 4b, we show the regret at every round incurred by OCTAL and ETC with rank-1 approximation; the regret of OCTAL decreases in every phase which is useful in practical recommendation systems. Recall that the ETC algorithm with rank-1 approximation has a poor performance since it suffers from the $\mathsf{T}^{2/3}$ dependence in the worst case. Again, we must point out that it is critical to tune the exploration period in ETC (as a function of rounds and sub-optimality gaps - unknown in practice and therefore difficult) (Lattimore & Szepesvári, 2020)[Ch. 6], OCTAL still suffers lower regret without such side-information.

Book-Crossing Dataset: Next we consider the Book-Crossing Dataset (Ziegler et al., 2005) which consists of ratings from 278K users (with demographic information) for 271K books. As in the

(a) Comparison of regret at $T = 100$ for UCB, OCTAL and ETC (with rank-1 approxmiation) for different exploration periods

(b) Comparison of regret at every round for OCTAL and ETC with rank-1 approximation and exploration period $m = 20, 40$

(c) Comparison of cumulative regret until every round for OCTAL and ETC with rank-1 approximation and different exploration periods
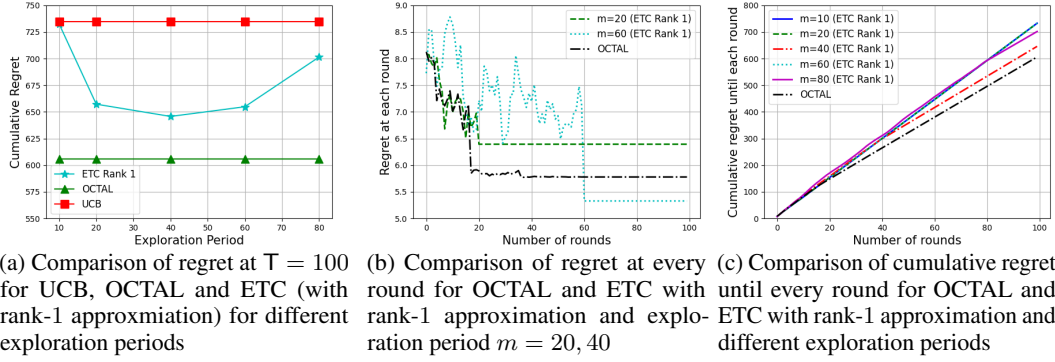
Figure 4: Comparison of regret incurred by OCTAL algorithm (Algorithm 3, ETC algorithm 2 with rank-1 approximation of the estimated matrix against UCB for the Jester dataset. In Figure 4a, we show that OCTAL and both versions of ETC have significantly lower regret than baseline UCB method. In Figure 4b, we show that the regret of OCTAL decreases in every phase unlike ETC algorithms. In Figure 4c, we show that the cumulative regret of OCTAL is lower than that of ETC with rank 1 approximation at every round $t \leq T$.

Movielens setting, we identify $M = 70$ disjoint clusters of users as a unique combination of country and age bin (0-25,25-50,50-75,75-100) that have rated at least 100 books. Next we take $N = 150$ books that have the most ratings. For each user cluster and each chosen book, we take the average rating of all users in the cluster for that book. We select the number of rounds $T = 100$ and as before, we run the UCB algorithm separately for each user, ETC algorithm (Alg. 2) with the rank-1 approximation of the estimated reward matrix, OCTAL Algorthm (Alg. 3 with the same minor mofications as in the synthetic datasets). The results are demonstrated in Figures 5a,5b and 5c. As before the UCB algorithm has the worst performance; OCTAL has a superior performance to ETC with rank-1 approximation for different exploration periods (and at all rounds). The conclusions are very similar to that obtained in the Jester dataset.

(a) Comparison of regret at $T = 100$ for UCB, OCTAL and ETC (with rank-1 approxmiation) for different exploration periods

(b) Comparison of regret at every round for OCTAL and ETC with rank-1 approximation and exploration period $m = 20, 40$

(c) Comparison of cumulative regret until every round for OCTAL and ETC with rank-1 approximation and different exploration periods
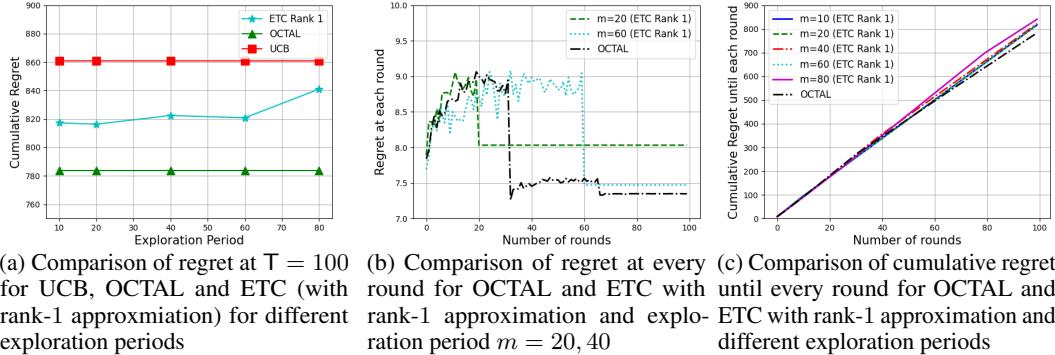
Figure 5: Comparison of regret incurred by OCTAL algorithm (Algorithm 3, ETC algorithm 2 with rank-1 approximation of the estimated matrix against UCB for the Book Crosssing dataset. In Figure 5a, we show that OCTAL and both versions of ETC have significantly lower regret than baseline UCB method. In Figure 5b, we show that the regret of OCTAL decreases in every phase unlike ETC algorithms. In Figure 5c, we show that the cumulative regret of OCTAL is lower than that of ETC with rank 1 approximation at every round $t \leq T$.

B MISSING PROOFS IN SECTION 3

We start with the following corollary:

Algorithm 4 ESTIMATE LOW RANK MATRIX (SUB-MATRIX) OFFLINE (CHEN ET AL., 2019)

Require: rows $\mathcal{U} \subseteq [\mathsf{M}]$, columns $\mathcal{V} \subseteq [\mathsf{N}]$, noise σ^2, rank r of $\mathbf{P}$, bernoulli sampling probability $0 \leq p \leq 1$.

1: Set $d_2 = \min(|\mathcal{U}|, |\mathcal{V}|)$ and $\lambda = C_\lambda \sigma \sqrt{d_2 p}$ for some constant $C_\lambda > 0$.
2: 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 $1 - p$.
3: Denote $\Omega = \{(i, j) \in \mathcal{U} \times \mathcal{V} \mid \delta_{ij} = 1\}$. Observe $\mathbf{Z}_{ij}$ (noisy version of $\mathbf{P}_{ij}$) for all $(i, j) \in \Omega$
4: Solve convex program

$$\min_{\widehat{\mathbf{P}} \in \mathbb{R}^{|\mathcal{U}| \times |\mathcal{V}|}} \frac{1}{2} \sum_{(i,j) \in \Omega} \left(\widehat{\mathbf{P}}_{ij} - \mathbf{Z}_{ij} \right)^2 + \lambda \|\widehat{\mathbf{P}}\|_\star \tag{8}$$

where $\|\widehat{\mathbf{P}}\|_\star$ denotes nuclear norm of matrix $\widehat{\mathbf{P}}$
5: Return $\widehat{\mathbf{P}}$.

Lemma 4 (Theorem 1 in Chen et al. (2019)). *Let $\mathbf{P} = \bar{\mathbf{U}} \boldsymbol{\Sigma} \bar{\mathbf{V}}^\mathsf{T} \in \mathbb{R}^{d \times d}$ such that $\bar{\mathbf{U}} \in \mathbb{R}^{d \times r}, \bar{\mathbf{V}} \in \mathbb{R}^{d \times r}$ and $\boldsymbol{\Sigma} \triangleq \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r) \in \mathbb{R}^{r \times r}$ with $\bar{\mathbf{U}}^\mathsf{T} \bar{\mathbf{U}} = \bar{\mathbf{V}}^\mathsf{T} \bar{\mathbf{V}} = \mathbf{I}$ and $\|\bar{\mathbf{U}}\|_{2,\infty} \leq \sqrt{\mu r / d}, \|\bar{\mathbf{V}}\|_{2,\infty} \leq \sqrt{\mu r / d}$. Let $1 \geq p \geq C \mu^2 d^{-1} \log^3 d$ for some sufficiently large constant $C > 0$, $\sigma = O\left(\sqrt{\frac{pd}{\mu^3 \log d}} \|\mathbf{P}\|_\infty\right)$, rank $r = O(1)$ and condition number $\kappa \triangleq \frac{\max_i \lambda_i}{\min_i \lambda_i} = O(1)$. Then, with probability exceeding $1 - O(d^{-3})$, we can compute a matrix $\widehat{\mathbf{P}} \in \mathbb{R}^{\mathsf{M} \times \mathsf{N}}$ by using Algorithm 4 with parameters $(\mathcal{U} = [\mathsf{M}], \mathcal{V} = [\mathsf{N}], \sigma^2, r, p)$ s.t.,*

$$\|\widehat{\mathbf{P}} - \mathbf{P}\|_\infty \leq O\left(\frac{\sigma}{\min_i \lambda_i} \cdot \sqrt{\frac{\mu d \log d}{p}} \|\mathbf{P}\|_\infty\right). \tag{9}$$

Next, we extend Lemma 4 to rectangular matrices in the following result:

Lemma 5. *Let $\mathbf{P} = \bar{\mathbf{U}} \boldsymbol{\Sigma} \bar{\mathbf{V}}^\mathsf{T} \in \mathbb{R}^{\mathsf{M} \times \mathsf{N}}$ such that $\bar{\mathbf{U}} \in \mathbb{R}^{\mathsf{M} \times r}, \bar{\mathbf{V}} \in \mathbb{R}^{\mathsf{N} \times r}$ and $\boldsymbol{\Sigma} \triangleq \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r) \in \mathbb{R}^{r \times r}$ with $\bar{\mathbf{U}}^\mathsf{T} \bar{\mathbf{U}} = \bar{\mathbf{V}}^\mathsf{T} \bar{\mathbf{V}} = \mathbf{I}$ and $\|\bar{\mathbf{U}}\|_{2,\infty} \leq \sqrt{\mu r / \mathsf{M}}, \|\bar{\mathbf{V}}\|_{2,\infty} \leq \sqrt{\mu r / \mathsf{N}}$. Let $d_1 = \max(\mathsf{M}, \mathsf{N})$ and $d_2 = \min(\mathsf{M}, \mathsf{N})$. Let $1 \geq p \geq C \mu^2 d_1 d_2^{-2} \log^3 d_1$ for some sufficiently large constant $C > 0$, $\sigma = O\left(\sqrt{\frac{pd_2^3}{d_1^2 \mu^3 \log^3 d_1}} \|\mathbf{P}\|_\infty\right)$, rank $r = O(1)$ and condition number $\kappa \triangleq \frac{\max_i \sigma_i}{\min_i \sigma_i} = O(1)$. Then, with probability exceeding $1 - O(d_1^{-3})$, we can compute a matrix $\widehat{\mathbf{P}} \in \mathbb{R}^{\mathsf{M} \times \mathsf{N}}$ by using Algorithm 4 with parameters $(\mathcal{U} = [\mathsf{M}], \mathcal{V} = [\mathsf{N}], \sigma^2, r, p)$ s.t.,*

$$\|\widehat{\mathbf{P}} - \mathbf{P}\|_\infty = O\left(\frac{\sigma r}{\sqrt{d_2}} \left(\frac{d_1}{d_2}\right)^{1/2} \sqrt{\frac{\mu^3 \log d_1}{p}}\right) \tag{10}$$

Proof of Lemma 5. Without loss of generality, let us assume that the matrix $\mathbf{P}$ is tall i.e. $\mathsf{M} \geq \mathsf{N}$. Now, let us construct the matrix

$$\mathbf{Q} = [\mathbf{P} \quad \mathbf{0}_{\mathsf{M} \times \mathsf{M} - \mathsf{N}}] = \bar{\mathbf{U}} \boldsymbol{\Sigma} [\bar{\mathbf{V}}^\mathsf{T} \quad \mathbf{0}_{\mathsf{N} - \mathsf{M}}^\mathsf{T}]$$

where $\mathbf{Q} \in \mathbb{R}^{\mathsf{M} \times \mathsf{M}}$. Clearly, the decomposition $\mathbf{Q} = \bar{\mathbf{U}} \boldsymbol{\Sigma} [\bar{\mathbf{V}}^\mathsf{T} \quad \mathbf{0}_{\mathsf{N} - \mathsf{M}}^\mathsf{T}]$ also coincides with the SVD of $\mathbf{Q}$ since both matrices $\bar{\mathbf{U}}$ and $[\bar{\mathbf{V}}^\mathsf{T} \quad \mathbf{0}_{\mathsf{N} - \mathsf{M}}^\mathsf{T}]^\mathsf{T}$ are orthonormal matrices while $\boldsymbol{\Sigma}$ remains unchanged. In case when $\mathsf{N} > \mathsf{M}$, we can construct $\mathbf{Q}$ similarly by vertically stacking $\mathbf{P}$ with a zero matrix of dimensions $(\mathsf{N} - \mathsf{M}) \times \mathsf{M}$. Hence, generally speaking, let us denote $d_1 = \max(\mathsf{M}, \mathsf{N})$ and $d_2 = \min(\mathsf{M}, \mathsf{N})$.

The matrix $\mathbf{Q}$ is $\bar{\mu}$-incoherent where $\bar{\mu} r (d_1)^{-1} = \mu r d_2^{-1}$ implying that $\bar{\mu} = \mu d_1 / d_2$. Moreover, we also have $\|\mathbf{Q}\|_\infty = \|\mathbf{P}\|_\infty$ implying that $\max_{ij} |\mathbf{P}_{ij}| = \max_{ij} |\mathbf{Q}_{ij}|$. Therefore, by invoking Lemma 1, the sample size must obey

$$p \geq \frac{C \mu^2 d_1}{d_2^2} \log^3(d_1) \quad \text{and} \quad \sigma = O\left(\sqrt{\frac{pd_2^3}{d_1^2 \mu^3 \log^3 d_1}} \|\mathbf{P}\|_\infty\right)$$

Then with probability at least $O(d_1^{-3})$, we can recover a matrix $\widehat{\mathbf{Q}}$ such that

$$\|\widehat{\mathbf{Q}} - \mathbf{Q}\|_\infty \leq O\left(\frac{\sigma}{\min_i \lambda_i} \sqrt{\frac{\mu \frac{d_1}{d_2} d_1 \log d_1}{p}} \|\mathbf{P}\|_\infty\right).$$

17

Algorithm 5 ESTIMATE LOW RANK MATRIX (SUB-MATRIX) OFFLINE - RECTANGULAR MATRICES

Require: rows $\mathcal{U} \subseteq [\mathsf{M}]$, columns $\mathcal{V} \subseteq [\mathsf{N}]$, noise σ^2, rank r of $\mathbf{P}$, bernoulli sampling probability $0 \leq p \leq 1$.
1: Set $d_2 = \min(|\mathcal{U}|, |\mathcal{V}|)$ and $\lambda = C_\lambda \sigma \sqrt{d_2 p}$ for some constant $C_\lambda > 0$.
2: 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 $1 - p$.
3: Denote $\Omega = \{(i,j) \in \mathcal{U} \times \mathcal{V} \mid \delta_{ij} = 1\}$. Observe $\mathbf{Z}_{ij}$ (noisy version of $\mathbf{P}_{ij}$) for all $(i,j) \in \Omega$
4: Without loss of generality, assume $|\mathcal{U}| \leq |\mathcal{V}|$. For each $i \in \mathcal{V}$, independently set ζ_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 $\mathcal{V}^{(q)} = \{i \in \mathcal{V} \mid \zeta_i = q\}$ 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 $\mathcal{U}$.*
5: **for** $q \in [k]$ **do**
6:   Solve convex program
$$\min_{\widehat{\mathbf{P}}^{(q)} \in \mathbb{R}^{|\mathcal{U}| \times |\mathcal{V}^{(q)}|}} \frac{1}{2} \sum_{(i,j) \in \Omega^{(q)}} \left(\widehat{\mathbf{P}}^{(q)}_{i\pi(j)} - \mathbf{Z}_{ij} \right)^2 + \lambda \|\widehat{\mathbf{P}}^{(q)}\|_\star, \tag{11}$$

   where $\|\widehat{\mathbf{P}}^{(q)}\|_\star$ denotes nuclear norm of matrix $\widehat{\mathbf{P}}^{(q)}$ and $\pi(j)$ is index of j in set $\mathcal{V}^{(q)}$.
7: **end for**
8: Return $\widehat{\mathbf{P}} \in \mathbb{R}^{\mathsf{M} \times \mathsf{N}}$ s.t. $\widehat{\mathbf{P}}_{\mathcal{U}, \mathcal{V}^{(q)}} = \widehat{\mathbf{P}}^{(q)}$ for all $q \in [k]$ and for every $(i,j) \notin \mathcal{U} \times \mathcal{V}$, $\widehat{\mathbf{P}}_{ij} = 0$.

Using the fact that $\|\mathbf{P}\|_\infty \leq \max_i \lambda_i \|\bar{\mathbf{U}}\|_{2,\infty} \|\bar{\mathbf{V}}\|_{2,\infty} = \max_i \lambda_i \mu r/\sqrt{d_1 d_2} \implies \|\mathbf{P}\|_\infty / \min_i \lambda_i = \kappa \cdot \mu r/\sqrt{d_1 d_2} = O\left(\mu r/\sqrt{d_1 d_2}\right)$ (using the fact that $\kappa = O(1)$, we obtain a matrix $\widehat{\mathbf{P}}$ such that

$$\|\widehat{\mathbf{P}} - \mathbf{P}\|_\infty \leq O\left(\frac{\sigma \mu r}{\sqrt{d_1 d_2}} \left(\frac{d_1}{d_2}\right)^{1/2} \sqrt{\frac{\mu d_1 \log d_1}{p}} \right) = O\left(\frac{\sigma r}{\sqrt{d_2}} \left(\frac{d_1}{d_2}\right)^{1/2} \sqrt{\frac{\mu^3 \log d_1}{p}} \right).$$

$\square$

Here we recall the discussion in Remark 2. From Lemma 5, we saw that in the guarantee provided in equation 10, the entry-wise error guarantee has an undesirable $(d_1/d_2)^{1/2}$ factor on the right hand side. This is because appending a zero matrix to make the rectangular matrix square leads to an increase in the incoherence factor by $(d_1/d_2)^{1/2}$ since there are many zero entries. This is what leads to the undesirable factor; but we can get around this by splitting the rectangular matrix randomly into approximately square sub-matrices and completing each of them individually. The improved algorithm for rectangular matrices is described in Algorithm 5 - note that in Step 4, we split the rectangular matrix into approximately square matrices and in Steps 5-7, we complete each of the approximately square sub-matrix individually. Finally, in Step 8, we join the estimated sub-matrices to get an estimate of the entire matrix and return it. In the main paper, the corresponding steps are done in Lines 10-14 in Algorithm 1. We provide a formal proof in the following lemma:

Lemma 6. *Let the matrix $\mathbf{P} \in \mathbb{R}^{\mathsf{M} \times \mathsf{N}}$ satisfy the conditions as stated in Lemma 5. Let $d_1 = \max(\mathsf{M}, \mathsf{N})$, $d_2 = \min(\mathsf{M}, \mathsf{N})$ such that $d_2 = \Omega(\mu r \log(r d_2))$, $1 \geq p \geq C \mu^2 d_2^{-1} \log^3 d_2$ for some sufficiently large constant $C > 0$ and $\sigma = O\left(\sqrt{\frac{p d_2}{\mu^3 \log d_2}} \|\mathbf{P}\|_\infty \right)$. We can use Algorithm 5 with parameters $(\mathcal{U} = [\mathsf{M}], \mathcal{V} = [\mathsf{N}], \sigma^2, r, p)$ to compute a matrix $\widehat{\mathbf{P}} \in \mathbb{R}^{\mathsf{M} \times \mathsf{N}}$ such that for any pair of indices $(i,j) \in [\mathsf{M}] \times [\mathsf{N}]$,*

$$|\widehat{\mathbf{P}}_{ij} - \mathbf{P}_{ij}| \leq O\left(\frac{\sigma r}{\sqrt{d_2}} \sqrt{\frac{\mu^3 \log d_2}{p}} \right) \tag{12}$$

with probability exceeding $1 - O(d_2^{-3})$. Again, with probability exceeding $1 - O(d_1 d_2^{-4})$, the output matrix $\widehat{\mathbf{P}}$ also satisfies $\|\mathbf{P} - \widehat{\mathbf{P}}\|_\infty \leq O\left(\frac{\sigma r}{\sqrt{d_2}} \sqrt{\frac{\mu^3 \log d_2}{p}} \right)$.

Proof. Let us assume that the matrix $\mathbf{P}$ is tall i.e. $\mathsf{M} \geq \mathsf{N}$. Now, let us partition the set of rows into $\frac{\mathsf{M}}{\mathsf{N}}$ groups by assigning each group uniformly at random to each row. Notice that the expected

number of rows in each group is N and by using Chernoff bound, the number of rows in each group lies in the interval $[\frac{\mathsf{N}}{2}, \frac{3\mathsf{N}}{2}]$ with probability at least $1 - 2\exp(-\mathsf{N}/12)$. When $\mathsf{M} \le \mathsf{N}$, we partition the set of columns in a similar manner into N/M groups so the number of columns in each group lies in the interval $[\frac{\mathsf{M}}{2}, \frac{3\mathsf{M}}{2}]$ with probability at least $1 - 2\exp(-\mathsf{M}/12)$. Hence, generally speaking, let us denote $d_1 = \max(\mathsf{M}, \mathsf{N})$ and $d_2 = \min(\mathsf{M}, \mathsf{N})$; we constructed d_1/d_2 sub-matrices of $\mathbf{P}$ denoted by $\mathbf{P}^{(1)}, \mathbf{P}^{(2)}, \dots, \mathbf{P}^{(d_1/d_2)}$. With probability at least $1 - 2\exp(-d_2/12)$, the matrices $\mathbf{P}^{(1)}, \mathbf{P}^{(2)}, \dots, \mathbf{P}^{(d_1/d_2)}$ are approximately square i.e. the ratio of their dimensions is a constant and lie in the interval $[1/2, 3/2]$.

Let us analyze the guarantees on estimating $\mathbf{P}^{(1)}$. The analysis for other matrices follow along similar lines. Note that $\mathbf{P}^{(1)} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}_{\mathsf{sub}}^{\mathsf{T}}$ where $\mathbf{V}_{\mathsf{sub}}$ denotes the $\mathsf{M}' \times r$ matrix where the rows in $\mathbf{V}_{\mathsf{sub}}$ corresponds to the M' rows in $\mathbf{V}$ assigned to $\mathbf{P}^{(1)}$.

First we will bound from below the minimum eigenvalue of the matrix $\mathbf{V}_{\mathsf{sub}}^{\mathsf{T}}\mathbf{V}_{\mathsf{sub}}$. Note that every row of $\mathbf{V}$ is independently sampled with probability $p \triangleq d_2/d_1$ for the matrix $\mathbf{P}^{(1)}$. Hence, we have

$$\frac{1}{p}\mathbf{V}_{\mathsf{sub}}^{\mathsf{T}}\mathbf{V}_{\mathsf{sub}} = \frac{1}{p}\sum_{i\in[d_1]}\delta_i\mathbf{V}_i\mathbf{V}_i^{\mathsf{T}} = \sum_{i\in[d_1]}\mathbf{W}^{(i)}$$

where δ_i denotes the indicator random variable which is true when $\mathbf{V}_i$ (the ith row of $\mathbf{V}$) is chosen for $\mathbf{P}^{(1)}$ and $\mathbf{W}^{(i)} = \frac{1}{p}\delta_i\mathbf{V}_i\mathbf{V}_i^{\mathsf{T}}$. Notice that the random matrices $\mathbf{W}^{(i)}$ are independent with $\mathbb{E}\mathbf{W}^{(i)} = \mathbf{V}_i\mathbf{V}_i^{\mathsf{T}}$. Hence we define $\mathbf{Z}^{(i)} = \mathbf{W}^{(i)} - \mathbb{E}\mathbf{W}^{(i)}$ satisfying $\mathbb{E}\mathbf{Z}^{(i)} = 0$. Moreover, for all $i \in [d_1]$, we have $\|\mathbf{Z}^{(i)}\|_2 \le \left(1 + \frac{1}{p}\right)\|\mathbf{V}_i\mathbf{V}_i^{\mathsf{T}}\|_2 \le \left(1 + \frac{1}{p}\right)\max_i\|\mathbf{V}_i\|_2^2 \le \frac{2\mu r}{pd_1}$. Next, we can show the following:

$$\|\sum_{i\in[d_1]}\mathbf{Z}^{(i)}(\mathbf{Z}^{(i)})^{\mathsf{T}}\|_2 \le \|\left(\frac{1}{p} - 1\right)\sum_{i\in[d_1]}(\mathbf{V}_i\mathbf{V}_i^{\mathsf{T}})(\mathbf{V}_i\mathbf{V}_i^{\mathsf{T}})^{\mathsf{T}}\|_2 \le \|\left(\frac{1}{p} - 1\right)\|\mathbf{V}_i\|^2\sum_{i\in[d_1]}(\mathbf{V}_i\mathbf{V}_i^{\mathsf{T}})\|_2$$
$$\le \frac{\mu r}{pd_1}\lambda_{\max}(\mathbf{V}^{\mathsf{T}}\mathbf{V}).$$

Similarly, we will also have

$$\|\sum_{i\in[d_1]}(\mathbf{Z}^{(i)})^{\mathsf{T}}\mathbf{Z}^{(i)}\|_2 \le \frac{\mu r}{pd_1}\lambda_{\max}(\mathbf{V}^{\mathsf{T}}\mathbf{V}) \le \frac{\mu r}{pd_1}$$

where we used that $\mathbf{V}^{\mathsf{T}}\mathbf{V}$ is orthogonal. Therefore, by using Bernstein's inequality for matrices (Theorem 1.6 in Tropp (2012)), we have with probability at least $1 - \delta$,

$$\|\frac{1}{p}\mathbf{V}_{\mathsf{sub}}^{\mathsf{T}}\mathbf{V}_{\mathsf{sub}} - \mathbf{V}^{\mathsf{T}}\mathbf{V}\| \le \frac{2\mu r}{3pd_1}\log\frac{2r}{\delta} + \sqrt{\frac{\mu r}{pd_1}\log\frac{2r}{\delta}}.$$

Hence, by using Weyl's inequality, we will have with probability $1 - \delta$

$$\lambda_{\min}(\mathbf{V}_{\mathsf{sub}}^{\mathsf{T}}\mathbf{V}_{\mathsf{sub}}) \ge p - \frac{2\mu r}{3d_1}\log\frac{2r}{\delta} - \sqrt{\frac{p\mu r}{d_1}\log\frac{2r}{\delta}}.$$

Hence, by substituting $\delta = d_2^{-3}$ we have that with probability at least $1 - d_2^{-3}$, if $d_2 = \Omega(\mu r\log(rd_2))$, then $\lambda_{\min}(\mathbf{V}_{\mathsf{sub}}^{\mathsf{T}}\mathbf{V}_{\mathsf{sub}}) \ge p/2$ (since $p = d_1/d_2$) implying that $\mathbf{V}_{\mathsf{sub}}^{\mathsf{T}}\mathbf{V}_{\mathsf{sub}}$ is invertible. Also, under the same condition, note that we can show similarly that $\lambda_{\max}(\mathbf{V}_{\mathsf{sub}}^{\mathsf{T}}\mathbf{V}_{\mathsf{sub}}) \le 3p/2$ implying that the condition number of each sub-matrix also stays $O(1)$ with probability at least $1 - d_2^{-3}$. Now, note that $\mathbf{V}_{\mathsf{sub}}$ is not orthogonal and therefore, we have

$$\mathbf{P}^{(1)} = \mathbf{U}\boldsymbol{\Sigma}(\mathbf{V}_{\mathsf{sub}}^{\mathsf{T}}\mathbf{V}_{\mathsf{sub}})^{1/2}(\mathbf{V}_{\mathsf{sub}}^{\mathsf{T}}\mathbf{V}_{\mathsf{sub}})^{-1/2}\mathbf{V}_{\mathsf{sub}}^{\mathsf{T}} = \mathbf{U}\widehat{\mathbf{U}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{V}}(\mathbf{V}_{\mathsf{sub}}^{\mathsf{T}}\mathbf{V}_{\mathsf{sub}})^{-1/2}\mathbf{V}_{\mathsf{sub}}^{\mathsf{T}}$$

where $\widehat{\mathbf{U}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{V}}$ is the SVD of the matrix $\boldsymbol{\Sigma}(\mathbf{V}_{\mathsf{sub}}^{\mathsf{T}}\mathbf{V}_{\mathsf{sub}})^{1/2}$. Since $\widehat{\mathbf{U}}$ is orthogonal, $\widehat{\mathbf{U}}\mathbf{U}$ is orthogonal as well. Similarly, $(\widehat{\mathbf{V}}(\mathbf{V}_{\mathsf{sub}}^{\mathsf{T}}\mathbf{V}_{\mathsf{sub}})^{-1/2}\mathbf{V}_{\mathsf{sub}}^{\mathsf{T}})^{\mathsf{T}}$ is orthogonal as well whereas $\widehat{\boldsymbol{\Sigma}}$ is diagonal. Hence $\mathbf{U}\widehat{\mathbf{U}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{V}}(\mathbf{V}_{\mathsf{sub}}^{\mathsf{T}}\mathbf{V}_{\mathsf{sub}})^{-1/2}\mathbf{V}_{\mathsf{sub}}^{\mathsf{T}}$ indeed corresponds to the SVD of $\mathbf{P}^{(1)}$ and we only need to

argue about the incoherence of $\widehat{\mathbf{U}}\mathbf{U}$ and $\widehat{\mathbf{V}}(\mathbf{V}_{\mathsf{sub}}^{\mathsf{T}}\mathbf{V}_{\mathsf{sub}})^{-1/2}\mathbf{V}_{\mathsf{sub}}^{\mathsf{T}}$. Notice that $\max_i \|(\mathbf{U}\widehat{\mathbf{U}})^{\mathsf{T}}\mathbf{e}_i\| \le$ $\|\mathbf{U}^{\mathsf{T}}\mathbf{e}_i\| \le \sqrt{\frac{\mu r}{d_2}}$. On the other hand,

$$\max_i \|\widehat{\mathbf{V}}(\mathbf{V}_{\mathsf{sub}}^{\mathsf{T}}\mathbf{V}_{\mathsf{sub}})^{-1/2}\mathbf{V}_{\mathsf{sub}}^{\mathsf{T}}\mathbf{e}_i\| \le \max_i \|(\mathbf{V}_{\mathsf{sub}}^{\mathsf{T}}\mathbf{V}_{\mathsf{sub}})^{-1/2}\mathbf{V}_{\mathsf{sub}}^{\mathsf{T}}\mathbf{e}_i\|$$

$$\le \frac{\|\mathbf{V}_{\mathsf{sub}}\|_{2,\infty}}{\sqrt{\lambda_{\min}(\mathbf{V}_{\mathsf{sub}}^{\mathsf{T}}\mathbf{V}_{\mathsf{sub}})}} \le \frac{\|\mathbf{V}\|_{2,\infty}}{\sqrt{p/2}} \le \sqrt{\frac{2\mu r}{d_2}}.$$

Hence, with probability $1 - 2\exp(-d_2/12) - O(d_2^{-3})$, we can recover an estimate $\widehat{\mathbf{P}}^{(1)}$ and apply Lemma 5 to conclude that (recall that the rank of $\mathbf{P}^{(1)}$ is $O(1)$, we have proved that the condition number of $\mathbf{P}^{(1)}$ is $O(1)$ w.p. at least $1 - O(d_2^{-3})$, the incoherence factor of $\mathbf{P}^{(1)}$ has increased by a factor of at most 2 w.p. at least $1 - O(d_2^{-3})$ and the ratio of the dimensions of $\mathbf{P}^{(1)}$ is in the interval $[1/2, 3/2]$ w.p. at least $1 - \exp(-d_2/12)$ implying that the conditions of Lemma 5 are satisfied for $\mathbf{P}^{(1)}$ w.h.p)

$$\|\mathbf{P}^{(1)} - \widehat{\mathbf{P}}^{(1)}\|_\infty \le O\Big(\frac{\sigma r}{\sqrt{d_2}}\sqrt{\frac{\mu^3 \log d_2}{p}}\Big).$$

as long as the conditions stated in the Lemma are satisfied. This implies that for any pair of indices $(i,j) \in [\mathsf{M}] \times [\mathsf{N}]$, we must have

$$|\widehat{\mathbf{P}}_{ij} - \mathbf{P}_{ij}| \le O\Big(\frac{\sigma r}{\sqrt{d_2}}\sqrt{\frac{\mu^3 \log d_2}{p}}\Big) \tag{13}$$

with probability exceeding $1 - O(d_2^{-3})$

Finally, by taking a union bound over all the d_1/d_2 partitions, we can compute estimates of all the sub-matrices $\mathbf{P}^{(1)}, \mathbf{P}^{(2)}, \dots, \mathbf{P}^{(d_1/d_2)}$ that have similar guarantees as above with probability at least $1 - 2d_1 d_2^{-1}\exp(-d_2/12) - O(d_1 d_2^{-4})$. Hence, by combining all the estimates, we can obtain a final estimate $\widehat{\mathbf{P}}$ of the matrix $\mathbf{P}$ that satisfies

$$\|\mathbf{P} - \widehat{\mathbf{P}}\|_\infty \le O\Big(\frac{\sigma r}{\sqrt{d_2}}\sqrt{\frac{\mu^3 \log d_2}{p}}\Big).$$

$\square$

Notice that the theoretical guarantees presented in Lemmas 4, 5 and 6 hold for the offline low rank matrix completion problem when each (noisy) entry of the low rank matrix $\mathbf{P} \in \mathbb{R}^{\mathsf{M} \times \mathsf{N}}$ is observed independently with some probability p (Bernoulli sampling model with probability p - see eq. 4). To summarize, if the observed set of entries is $\Omega \subseteq [\mathsf{M}] \times [\mathsf{N}]$ (each index $(i,j) \in [\mathsf{M}] \times [\mathsf{N}]$ is present in Ω with probability p), then we can solve several convex optimization problems (see Steps 5-7 in Algorithm 5) restricted to observations in Ω to compute an estimate $\widehat{\mathbf{P}}$ of the low rank matrix $\mathbf{P}$.

However, in our problem setting (Section 2), recall that in each round $t \in [\mathsf{T}]$, for each user $u \in [\mathsf{M}]$, some item $\rho_u(t)$ is recommended. Our goal is to translate the theoretical guarantees under Bernoulli sampling model with probability p (for some pre-determined p) to our setting. A simple approach (as mentioned in Remark 3) is the following (described in Algorithm 6): we first sample a set of indices $\Omega \subseteq [\mathsf{M}] \times [\mathsf{N}]$ such that each index $(i,j) \in [\mathsf{M}] \times [\mathsf{N}]$ is present in Ω with probability p (Step 2 in Algorithm 6). Subsequently, we aim to obtain a single observation (noisy) corresponding to each pair of indices in Ω; in each round, for each user $i \in [\mathsf{M}]$, we recommend an item $j \in [\mathsf{N}]$ such that $(i,j) \in \Omega$ and has not been observed yet . If such an item is unavailable for the user i, we recommend any arbitrary item j such that $(i,j) \notin \Omega$ (Step 5 in Algorithm 6) but ignore the observation while reconstruction of the reward matrix; note that this is unnecessary in practice. Once we have obtained observations corresponding to all pairs of indices in Ω (we discard all observations corresponding to entries not in Ω), we can solve the convex optimization problems restricted to the observed entries in Ω (Steps 8-10 in Algorithm 6) to compute an estimate $\widehat{\mathbf{P}}$ of the reward matrix $\mathbf{P}$ and apply the theoretical guarantees in Lemma 6 directly. The total number of rounds needed will

---

Algorithm 6 ESTIMATE REWARD MATRIX (SUB-MATRIX) ONLINE

Require: users $\mathcal{U} \subseteq [\mathsf{M}]$, items $\mathcal{V} \subseteq [\mathsf{N}]$, noise σ^2, rank r of $\mathbf{P}$, bernoulli sampling probability $0 \leq p \leq 1$. Index of round t is relative to the first round when the algorithm is invoked; hence $t = 1, 2, \dots$.
1: Set $d_2 = \min(|\mathcal{U}|, |\mathcal{V}|)$ and $\lambda = C_\lambda \sigma \sqrt{d_2 p}$ for some constant $C_\lambda > 0$.
2: For each tuple of indices $(i, j) \in [\mathsf{M}] \times [\mathsf{N}]$, independently set $\delta_{ij} = 1$ with probability p and $\delta_{ij} = 0$ with probability $1 - p$.
3: Denote $\Omega = \{(i, j) \in \mathcal{U} \times \mathcal{V} \mid \delta_{ij} = 1\}$ and $m = \max_{i \in [\mathsf{M}]} | \mid j \in \mathcal{V} \mid (i, j) \in \Omega|$ to be the maximum number of index tuples in a particular row. For all $(i, j) \in \Omega$, set $\mathsf{Mask}_{ij} = 0$.
4: **for** rounds $t = 1, 2, \dots, m$ **do**
5:     For each user $u \in \mathcal{U}$, 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 $\mathbf{R}^{(t)}_{u\rho_u(t)}$.
6: **end for**
7: Without loss of generality, assume $|\mathcal{U}| \leq |\mathcal{V}|$. For each $i \in \mathcal{V}$, independently set ζ_i to be a value in the set $[\lceil |\mathcal{V}|/|\mathcal{U}| \rceil]$ uniformly at random. Partition indices in $\mathcal{V}$ into $\mathcal{V}^{(1)}, \mathcal{V}^{(2)}, \dots, \mathcal{V}^{(k)}$ where $k = \lceil |\mathcal{V}|/|\mathcal{U}| \rceil$ and $\mathcal{V}^{(q)} = \{i \in \mathcal{V} \mid \zeta_i = q\}$ 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 $\mathcal{U}$.*
8: **for** $q \in [k]$ **do**
9:     Solve convex program
$$\min_{\widehat{\mathbf{P}}^{(q)} \in \mathbb{R}^{|\mathcal{U}| \times |\mathcal{V}^{(q)}|}} \frac{1}{2} \sum_{(i,j) \in \Omega^{(q)}} \left(\widehat{\mathbf{P}}^{(q)}_{i\pi(j)} - \mathbf{Z}_{ij} \right)^2 + \lambda \|\widehat{\mathbf{P}}^{(q)}\|_\star, \tag{14}$$

    where $\|\widehat{\mathbf{P}}^{(q)}\|_\star$ denotes nuclear norm of matrix $\widehat{\mathbf{P}}^{(q)}$ and $\pi(j)$ is index of j in set $\mathcal{V}^{(q)}$.
10: **end for**
11: Return $\widehat{\mathbf{P}} \in \mathbb{R}^{\mathsf{M} \times \mathsf{N}}$ s.t. $\widehat{\mathbf{P}}_{\mathcal{U}, \mathcal{V}^{(q)}} = \widehat{\mathbf{P}}^{(q)}$ for all $q \in [k]$ and for every $(i, j) \notin \mathcal{U} \times \mathcal{V}$, $\widehat{\mathbf{P}}_{ij} = 0$.

---

be $\max_{i \in [\mathsf{M}]} |(i, j) \in \Omega \mid j \in [\mathsf{N}]|$ i.e. the maximum number of indices in a particular row present in Ω (which we will show to be bounded from above with high probability in the following corollary). This is because we wait until we have one noisy observation for every pair of indices in Ω.

In Algorithm 1 in the main paper, the aforementioned procedure is incorporated in Steps 10-14.

Corollary 1. *Consider a set $\Omega \subseteq [\mathsf{M}] \times [\mathsf{N}]$ of indices such that every index $(i, j) \in [\mathsf{M}] \times [\mathsf{N}]$ is present in Ω independently with probability p. Consider algorithm 6 with parameters $(\mathcal{U} = [\mathsf{M}], \mathcal{V} = [\mathsf{N}], \sigma^2, r, p)$ that recommends items to users according to Step 5 from the set Ω. Suppose the rank r reward matrix $\mathbf{P}$ and parameters p, σ satisfies the conditions stated in Lemma 6. In that case, using $m = O\left(\mathsf{N}p + \sqrt{\mathsf{N}p \log \mathsf{M} \delta^{-1}}\right)$ rounds, Algorithm 6 is able to compute a matrix $\widehat{\mathbf{P}}$ such that for any $(i, j) \in [\mathsf{M}] \times [\mathsf{N}]$, we have with probability exceeding $1 - \delta - O(d_2^{-3})$*

$$|\widehat{\mathbf{P}}_{ij} - \mathbf{P}_{ij}| \leq O\left(\frac{\sigma r}{\sqrt{d_2}} \sqrt{\frac{\mu^3 \log d_2}{p}} \right). \tag{15}$$

Proof of Corollary 1. Recall that $d_1 = \max(\mathsf{M}, \mathsf{N})$ and $d_2 = \min(\mathsf{M}, \mathsf{N})$. Suppose, with some parameter $p = \Omega(\mu^2 d_2^{-1} \log^3 d_2)$, we sample a set $\Omega \in [\mathsf{M}] \times [\mathsf{N}]$ of indices (Step 2 in Algorithm 6). Let us define the event $\mathcal{F}_1$ which is true when the maximum number of indices observed in some row $m \triangleq \max_{i \in [\mathsf{M}]} |(i, j) \in \Omega \mid j \in [\mathsf{N}]|$ satisfies $m = \Omega\left(\mathsf{N}p + \sqrt{\mathsf{N}p \log \mathsf{M} \delta^{-1}}\right)$. Recall that our goal is to obtain a single noisy reward observation for every (user,item) pair whose corresponding indices is present in Ω. In Step 5 of Algorithm 6, in each round, one item is recommended to every user -clearly the total number of rounds needed to achieve our goal is m.

We will bound the probability of the event $\mathcal{F}_1$ from above by using Chernoff bound. Let us denote the number of items observed for user $i \in [\mathsf{M}]$ to be Y_i i.e. $Y_i = |(i, j) \in \Omega \mid j \in [\mathsf{N}]|$. Algorithm $\mathcal{A}$ can then obtain the noisy entries of $\mathbf{P}$ corresponding to the set Ω by doing the following: in each round, for each user $i \in [\mathsf{M}]$, if there is an unobserved tuple of indices $(i, j) \in \Omega$, then $\mathcal{A}$ will recommend j to user i and obtain an noisy observation $\mathbf{P}_{ij} + \mathbf{E}_{ij}$; on the other hand, if there no unobserved entry, then $\mathcal{A}$ will simply recommend a random item j such that $(i, j) \notin \Omega$ (Step 5 in Algorithm

6). Subsequently, Algorithm 6 discards the observations corresponding to indices not in Ω and only utilizes the observations in Ω to compute an estimate of the reward matrix $\mathbf{P}$. Notice that each of the random variables $Y_1, Y_2, \ldots, Y_{\mathsf{M}} \sim \mathsf{Binomial}(\mathsf{N}, p)$ and are independent. By using Chernoff bound, we have that for each $i \in [\mathsf{M}]$,

$$\Pr\left(\cup_{i \in [\mathsf{M}]} |Y_i - \mathsf{N}p| \geq \mathsf{N}p\epsilon\right) \leq 2\mathsf{M}\exp\left(-\frac{\epsilon^2 \mathsf{N}p}{3}\right)$$
$$\implies Y_i \leq \mathsf{N}p + O\left(\sqrt{\mathsf{N}p\log\mathsf{M}\delta^{-1}}\right) \text{ for all } i \in [\mathsf{M}]$$

with probability $1 - \delta$ implying that $\Pr(\mathcal{F}_1) \leq \delta$. Let $\mathcal{F}_2$ be the event when the recovered matrix $\widehat{\mathbf{P}}$ does not satisfy the guarantee on $\|\widehat{\mathbf{P}} - \mathbf{P}\|_\infty$ as stated in Lemma 6 equation 12 given a set of observed indices Ω sampled according to equation 4. From Lemma 6, we know that $\Pr(\mathcal{F}_2) = O(d_2^{-3})$ where $d_1 = \max(\mathsf{M}, \mathsf{N})$. Hence we can conclude $\Pr(\mathcal{F}_1 \cup \mathcal{F}_2) \leq \Pr(\mathcal{F}_1) + \Pr(\mathcal{F}_2) = \delta + O(d_2^{-3})$. Note that conditioned on the event that $\mathcal{F}_1$ does not hold true, we need only $m = O\left(\mathsf{N}p + \sqrt{\mathsf{N}p\log\mathsf{M}\delta^{-1}}\right)$ rounds to obtain a set of noisy observations corresponding to $\Omega' \supseteq \Omega$.

Hence, to conclude, with only $m = O\left(\mathsf{N}p + \sqrt{\mathsf{N}p\log\mathsf{M}\delta^{-1}}\right)$ rounds, Alg. 6 can obtain a single noisy reward observation corresponding to each set of indices in Ω (where every tuple of indices is present with probability p); note that observations for indices outside Ω are discarded completely. Subsequently, the noisy reward observations in Ω are used to compute an estimate $\widehat{\mathbf{P}}$ of the reward matrix $\mathbf{P}$ with the (Steps 8-11 in Algorithm 6) theoretical guarantees presented in Lemma 6 holding true. This completes the proof of the corollary. □

Remark 11. *[Remark 3 in Chen et al. (2019)] Note the failure probability in Corollary 1, Lemmas 6, 5 and 4 is $O(d_2^{-3})$. However the constant 3 can be replaced by any arbitrary constant c for example $c = 100$ without any change in the guarantees on $\|\widehat{\mathbf{P}} - \mathbf{P}\|_\infty$. Hence, the guarantees presented in Lemma 6 and Corollary 1 hold with probability at least $1 - O(d_2^{-c})$ for any arbitrary constant c.*

As mentioned in Remark 4, the drawbacks of Lemma 6 and Corollary 1 are that the smallest error that is possible to achieve by using Lemma 1 is by substituting $p = 1$ and thereby obtaining $\|\widehat{\mathbf{P}} - \mathbf{P}\|_\infty \leq \zeta = O\left(\sigma(\min_i \lambda_i)^{-1} \cdot \sqrt{\mu d \log d}\|\mathbf{P}\|_\infty\right)$ and moreover, the probability of failure is polynomially small in d_2 (see Remark 11). The former issue is substantial when we need an estimate $\widehat{\mathbf{P}}$ of the low rank matrix $\mathbf{P}$ with a smaller entry-wise estimation error than ζ. The second issue is substantial when the shorter dimension d_2 is small. We provide an improved algorithm (Algorithm 7) with technical modifications to fix the two aforementioned issues. Below, we describe our main ideas at a high level first:

Repeated recommendations for small estimation error: Recall that our strategy for computing an estimate $\widehat{\mathbf{P}}$ with theoretical guarantees was the following two-step procedure 1) obtain a subset of indices $\Omega \subseteq [\mathsf{M}] \times [\mathsf{N}]$ such that each index (i, j) is present with probability $p > 0$ 2) obtain noisy reward observations corresponding to all indices ((user,item) pairs) in Ω - in each round, for each user $i \in [\mathsf{M}]$ we recommended an item $j \in [\mathsf{N}]$ such that $(i, j) \in \Omega$ if such a j exists otherwise we recommend any item $j \in [\mathsf{N}]$ such that $(i, j) \notin \Omega$.

In our problem, an algorithm has the flexibility of recommending an item more than once to the same user. Suppose we have sampled a subset of indices $\Omega \subseteq [\mathsf{M}] \times [\mathsf{N}]$ as described above and is fixed. We modify the second step in the following way: for some fixed $s > 0$, our modified aim is to obtain s noisy observations corresponding to all entries ((user,item) pairs) in Ω. To do so, we repeat the process of recommending items to users corresponding to indices in Ω s times (See Step 5 in Algorithm 7). Clearly, if the total number of rounds required to obtain noisy reward observations corresponding to all pairs of indices in Ω once is m (bounded from above w.h.p - see Corollary 1), then the total number of rounds required for obtaining s observations corresponding to each entry in Ω is ms. As before, all reward observations outside the indices in Ω are discarded. For each entry $(i, j) \in \Omega$, we take $\mathbf{Z}_{ij}$ to be the average of the s reward observations corresponding to the recommendation of item j to user u- clearly, the variance proxy of the averaged observation is σ^2/s. Thus, we can again solve similar convex optimization problems as before (see Steps 13-15 in Algorithm 7) to obtain the same theoretical guarantee as in Corollary 1 but with noise variance

---

Algorithm 7 ESTIMATE REWARD MATRIX (SUB-MATRIX) ONLINE - IMPROVED ALGORITHM

Require: users $\mathcal{U} \subseteq [\mathsf{M}]$, items $\mathcal{V} \subseteq [\mathsf{N}]$, noise σ^2, rank r of $\mathbf{P}$, bernoulli sampling probability $0 \leq p \leq 1$, repetitions s and number of independent estimates f. Index of round t is relative to the first round when the algorithm is invoked; hence $t = 1, 2, \ldots$.

1: Set $d_2 = \min(|\mathcal{U}|, |\mathcal{V}|)$ and $\lambda = C_\lambda \sigma \sqrt{d_2 p}$ for some constant $C_\lambda > 0$.

2: **for** $z = 1, 2, \ldots, f$ **do**

3: For each tuple of indices $(i, j) \in [\mathsf{M}] \times [\mathsf{N}]$, independently set $\delta_{ij} = 1$ with probability p and $\delta_{ij} = 0$ with probability $1 - p$.

4: Denote $\Omega = \{(i, j) \in \mathcal{U} \times \mathcal{V} \mid \delta_{ij} = 1\}$ and $m_z = \max_{i \in \mathcal{U}} | \, |j \in \mathcal{V} \mid (i, j) \in \Omega|$ to be the maximum number of index tuples in a particular row.

5: **for** $\ell = 1, 2, \ldots, s$ **do**

6: For all $(i, j) \in \Omega$, set $\mathsf{Mask}_{ij} = 0$.

7: **for** $\ell' = 1, 2, \ldots, m_z$ **do**

8:   For each user $u \in \mathcal{U}$ in round $t = \sum_{z'=1}^{z-1} sm_{z'} + (\ell - 1)s + \ell'$, 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 $\mathbf{R}_{u\rho_u(t)}^{(t)}$.

9: **end for**

10: **end for**

11: For each tuple $(u, j) \in \Omega$, compute $\mathbf{Z}_{uj}$ to be average of $\{\mathbf{R}_{u\rho_u(t)}^{(t)}$ for $t \in [\sum_{z'=1}^{z-1} sm_{z'}, \sum_{z'=1}^{z} sm_{z'}] \mid \rho_u(t) = j\}$. *# $\mathbf{Z}_{uj}$ is the average of s independent observations at each index $(u, j) \in \Omega$ for a fixed z*

12: Without loss of generality, assume $|\mathcal{U}| \leq |\mathcal{V}|$. For each $i \in \mathcal{V}$, independently set ζ_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 $\mathcal{V}^{(q)} = \{i \in \mathcal{V} \mid \zeta_i = q\}$ 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 $\mathcal{U}$.*

13: **for** $q \in [k]$ **do**

14: Solve convex program

$$\min_{\widehat{\mathbf{P}}^{(q)} \in \mathbb{R}^{|\mathcal{U}| \times |\mathcal{V}^{(q)}|}} \frac{1}{2} \sum_{(i,j) \in \Omega^{(q)}} \left(\widehat{\mathbf{P}}_{i\pi(j)}^{(q)} - \mathbf{Z}_{ij}\right)^2 + \lambda \|\widehat{\mathbf{P}}^{(q)}\|_\star, \qquad (16)$$

where $\|\widehat{\mathbf{P}}^{(q)}\|_\star$ denotes nuclear norm of matrix $\widehat{\mathbf{P}}^{(q)}$ and $\pi(j)$ is index of j in set $\mathcal{V}^{(q)}$.

15: **end for**

16: Compute $\widehat{\mathbf{P}}^{(z)} \in \mathbb{R}^{\mathsf{M} \times \mathsf{N}}$ s.t. $\widehat{\mathbf{P}}_{\mathcal{U}, \mathcal{V}^{(q)}}^{(z)} = \mathbf{Q}^{(q)}$ for all $q \in [k]$ and for every $(i, j) \notin \mathcal{U} \times \mathcal{V}$, $\widehat{\mathbf{P}}_{ij}^{(z)} = 0$. *#We are computing independent estimates $\widehat{\mathbf{P}}^{(z)}$*

17: **end for**

18: Obtain estimate $\widehat{\mathbf{P}}$ by taking the entry-wise median of $\widehat{\mathbf{P}}^{(1)}, \widehat{\mathbf{P}}^{(2)}, \ldots, \widehat{\mathbf{P}}^{(f)}$. Return $\widehat{\mathbf{P}}$.

σ^2/s. This intuition is formalized in Lemma 2 below. In the main paper, the process of repeatedly recommending the items in a fixed subset of indices Ω is described in the For Loop in Steps 1-8 in Algorithm 1 - note that Ω is fixed in For Loop in Step 1 and in beginning of every iteration, we restart the recommendation procedure.

Independent estimates and entry-wise median for small error probability:

To increase the probability of success, we can compute f independent estimates of $\mathbf{P}$ (Line 2 and Line 16 in Algorithm 7) namely $\widehat{\mathbf{P}}^{(1)}, \widehat{\mathbf{P}}^{(2)}, \ldots, \widehat{\mathbf{P}}^{(f)}$ and compute their entry-wise median i.e. $\widehat{\mathbf{P}}_{ij} = \mathsf{median}(\widehat{\mathbf{P}}_{ij}^{(1)}, \ldots, \widehat{\mathbf{P}}_{ij}^{(f)})$ for all $(i, j) \in [\mathsf{M}] \times [\mathsf{N}]$. From our previous argument, the number of rounds for obtaining a single estimate $\widehat{\mathbf{P}}$ of the matrix $\mathbf{P}$ is $O(ms)$ (where the noise variance is reduced to σ^2/s) with high probability. Hence the total number of rounds for obtaining f estimates will be $\widetilde{O}(msf)$ with high probability. If we set $f = O(\log(\mathsf{MN}\delta^{-1}))$ for some fixed $\delta > 0$, then by taking a union bound over all MN entries, we can show that our guarantees hold with probability at least $1 - \delta$ (thus we remove the dependence of the failure probability on d_2). Again, this intuition is formalized in Lemma 2 below. In Algorithm 2, note that the f independent estimates are computed in Lines 2-6 and the entry-wise median is computed in Line 7. In Algorithm 3, the f independent estimates for each sub-matrix are computed in Lines 4-10 and the entry-wise median is computed in Line 15.

Lemma 7 (Restatement of Lemma 2). *Let rank $r = O(1)$ reward matrix $\mathbf{P} \in \mathbb{R}^{M \times N}$ with SVD decomposition $\mathbf{P} = \bar{\mathbf{U}}\boldsymbol{\Sigma}\bar{\mathbf{V}}^{\mathsf{T}}$ satisfy $\|\bar{\mathbf{U}}\|_{2,\infty} \leq \sqrt{\mu r/M}, \|\bar{\mathbf{V}}\|_{2,\infty} \leq \sqrt{\mu r/N}$ and condition number $\kappa = O(1)$. Let $d_1 = \max(M, N)$ and $d_2 = \min(M, N)$ such that $1 \geq p \geq C\mu^2 d_2^{-1} \log^3 d_2$ for sufficiently large constant $C > 0$. Suppose we observe noisy entries of $\mathbf{P}$ according to observation model in (1). For any positive integer $s > 0$ satisfying $\frac{\sigma}{\sqrt{s}} = O\Big(\sqrt{\frac{pd_2}{\mu^3 \log d_2}}\|\mathbf{P}\|_\infty\Big)$, there exists an algorithm $\mathcal{A}$ (See Algorithm 7 with input parameters $\mathcal{U} = [M], \mathcal{V} = N, \sigma^2, r, p, s, f$) with parameters s, p, σ that uses $m = O\Big(s\log(MN\delta^{-1})(Np + \sqrt{Np \log M\delta^{-1}})\Big)$ rounds to recommend items for users and compute a matrix $\hat{\mathbf{P}}$ such that with probability exceeding $1 - O(\delta \log(MN\delta^{-1}))$, we have*

$$\|\mathbf{P} - \hat{\mathbf{P}}\|_\infty \leq O\Big(\frac{\sigma r}{\sqrt{sd_2}}\sqrt{\frac{\mu^3 \log d_2}{p}}\Big). \tag{17}$$

Proof.

Claim 1. *Fix any index $(i, j) \in [M] \times [N]$. Then, if the conditions in the Lemma statement are satisfied, for any $z \in [f]$ in Line 2 of Alg. 7, the computed estimate $\hat{\mathbf{P}}^{(z)}$ in Line 16 in Alg. 7 satisfies the following: 1) with probability exceeding $9/10$, we have*

$$|\hat{\mathbf{P}}_{ij}^{(z)} - \mathbf{P}_{ij}| \leq O\Big(\frac{\sigma r}{\sqrt{sd_2}}\sqrt{\frac{\mu^3 \log d_2}{p}}\Big)$$

2) the total number of rounds required to compute $\hat{\mathbf{P}}^{(z)}$ is given by $sm_z = O\Big(s(Np + \sqrt{Np \log(M\delta^{-1})})\Big)$.

Proof. **Repeated Recommendations for small estimation error:** Consider Lines 5-10 in Algorithm 7 prior to which we have a fixed subset of indices $\Omega \subseteq [M] \times [N]$ where every index tuple (i, j) is present in Ω with probability p. In Line 8, for each user $i \in [M]$, Algorithm 7 recommends an item $j \in [N]$ such that $(i, j) \in \Omega$ if such a j exists otherwise we recommend any item $j \in [N]$ such that $(i, j) \notin \Omega$. Because of the For Loop in Line 5, Algorithm 7 does the aforementioned step s times for the same fixed subset of indices Ω. After discarding all observations corresponding to indices outside Ω, at the end of the For Loop in Line 10, we have s noisy iid observations (sub-gaussian random variables) $\mathbf{R}_{ij}^{(t_1)}, \mathbf{R}_{ij}^{(t_2)}, \ldots, \mathbf{R}_{ij}^{(t_s)}$ each with expectation $\mathbf{P}_{ij}$ and variance proxy σ^2 i.e. we have for all $h \in [s]$

$$\mathbb{E} \exp\Big(\lambda(\mathbf{R}_{ij}^{(t_h)} - \mathbf{P}_{ij})\Big) \leq \exp(\sigma^2 \lambda^2/2) \ \forall \lambda \in \mathbb{R}.$$

In Line 11 in Algorithm 7, we take $\mathbf{Z}_{ij} = s^{-1}\sum_{h=1}^{s} \mathbf{R}_{ij}^{(t_h)}$. Clearly, we have $\mathbb{E}\mathbf{Z}_{ij} = \mathbf{P}_{ij}$ and further, $\mathbf{Z}_{ij}$ has a variance proxy of σ^2/s as shown below:

$$\mathbb{E} \exp\Big(\lambda(\mathbf{Z}_{ij} - \mathbf{P}_{ij})\Big) = \mathbb{E} \prod_{h=1}^{s} \exp\Big(\lambda s^{-1}(\mathbf{R}_{ij}^{(t_h)} - \mathbf{P}_{ij})\Big)$$

$$= \prod_{h=1}^{s} \mathbb{E} \exp\Big(\lambda s^{-1}(\mathbf{R}_{ij}^{(t_h)} - \mathbf{P}_{ij})\Big) \leq \exp(\sigma^2 \lambda^2/2s) \ \forall \lambda \in \mathbb{R}.$$

Therefore, for any $z \in [f]$, if we solve the convex programs in Lines 13-16 of Algorithm 7, then we directly apply the guarantees in Corollary 1 to conclude that the computed matrix in Line 16 $\hat{\mathbf{P}}^{(z)}$ satisfies the following: for any $(i, j) \in [M] \times [N]$, we have with probability exceeding $1 - \delta - O(d_2^{-c})$ (see Remark 11 - we can set c such that the failure probability $\delta + O(d_2^{-c}) < 1/10$)

$$|\hat{\mathbf{P}}_{ij}^{(z)} - \mathbf{P}_{ij}| \leq O\Big(\frac{\sigma r}{\sqrt{sd_2}}\sqrt{\frac{\mu^3 \log d_2}{p}}\Big) \tag{18}$$

where with $m_z = \max_{i \in [M]} | \{j \in [N] \mid (i,j) \in \Omega\}|$, the total number of rounds required to compute $\widehat{\mathbf{P}}^{(z)}$ is given by $sm_z = O\Big(s(\mathsf{N}p + \sqrt{\mathsf{N}p\log(\mathsf{M}\delta^{-1})})\Big)$. Also, note that since the new and reduced noise variance proxy is σ^2/s, hence the condition $\sigma = O\Big(\sqrt{\frac{pd_2}{\mu^3 \log d_2}}\|\mathbf{P}\|_\infty\Big)$ stated in Lemma 6 (referred to in Corollary 1 that is being invoked) translates to $\sigma/\sqrt{s} = O\Big(\sqrt{\frac{pd_2}{\mu^3 \log d_2}}\|\mathbf{P}\|_\infty\Big)$. This is the new condition on σ that is stated in the current lemma. $\square$

Claim 2. *Fix any index $(i,j) \in [\mathsf{M} \times [\mathsf{N}]]$. Suppose the conditions in the Lemma statement hold true. Consider the final estimate $\widehat{\mathbf{P}}$ computed by taking the entry-wise median of $\widehat{\mathbf{P}}^{(1)}, \widehat{\mathbf{P}}^{(2)}, \ldots, \widehat{\mathbf{P}}^{(f)}$. With $f = O(\log(\mathsf{MN}\delta^{-1}))$, the $(ij)^{\text{th}}$ entry of the final estimate $\widehat{\mathbf{P}}_{ij}$ satisfies*

$$|\widehat{\mathbf{P}}_{ij} - \mathbf{P}_{ij}| = O\Big(\frac{\sigma r}{\sqrt{sd_2}}\sqrt{\frac{\mu^3 \log d_2}{p}}\Big)$$

with probability $1 - \delta/\mathsf{MN}$.

Proof. **Independent estimates and entry-wise median for small probability of error:** Consider Lines 2,3 and 18 in Algorithm 7. Algorithm 7 basically computes f independent estimates $\widehat{\mathbf{P}}^{(1)}, \widehat{\mathbf{P}}^{(2)}, \ldots, \widehat{\mathbf{P}}^{(f)}$ of the reward matrix $\mathbf{P}$ in order to boost the probability of success. Furthermore, we can compute a final estimate $\widehat{\mathbf{P}}$ (Line 18) by computing the entry-wise median of the matrix estimates $\widehat{\mathbf{P}}^{(1)}, \widehat{\mathbf{P}}^{(2)}, \ldots, \widehat{\mathbf{P}}^{(f)}$ i.e. for all $(i,j) \in [\mathsf{M}] \times [\mathsf{N}]$, we compute $\widehat{\mathbf{P}}_{ij} = \mathsf{median}(\widehat{\mathbf{P}}_{ij}^{(1)}, \ldots, \widehat{\mathbf{P}}_{ij}^{(f)})$.

Our goal is to increase the probability of success to $1 - \delta$ and we will show that the entry-wise median of the f estimates $\widehat{\mathbf{P}}$ does have the increased probability of success. As mentioned in the claim statement, fix the pair of indices (i,j). Consider the random variable

$$Y^{(z)} = \mathbf{1}\Big[|\widehat{\mathbf{P}}_{ij}^{(z)} - \mathbf{P}_{ij}| = O\Big(\frac{\sigma r}{\sqrt{sd_2}}\sqrt{\frac{\mu^3 \log d_2}{p}}\Big)\Big]$$

which is 1 with probability at least $9/10$ (from Claim 1). Since the estimates $\widehat{\mathbf{P}}^{(1)}, \widehat{\mathbf{P}}^{(2)}, \ldots, \widehat{\mathbf{P}}^{(f)}$ are independently computed , the random variables $Y^{(1)}, Y^{(2)}, \ldots, Y^{(f)}$ are independent as well. Hence consider the median $\widehat{\mathbf{P}}_{ij} = \mathsf{median}(\widehat{\mathbf{P}}_{ij}^{(1)}, \widehat{\mathbf{P}}_{ij}^{(2)}, \ldots, \widehat{\mathbf{P}}_{ij}^{(f)})$. Note that the median $\widehat{\mathbf{P}}_{ij}$ will satisfy $|\widehat{\mathbf{P}}_{ij} - \mathbf{P}_{ij}| = O\Big(\frac{\sigma r}{\sqrt{sd_2}}\sqrt{\frac{\mu^3 \log d_2}{p}}\Big)$ if at least half of the random variables $Y^{(1)}, Y^{(2)}, \ldots, Y^{(f)}$ are non-zero. Hence, we can apply Chernoff bound directly to state that $\Pr(\sum_{i=1}^{f} Y^{(i)} < f/2) \le 2\exp(-4f/75)$ Therefore, by setting $f = O(\log(\mathsf{MN}\delta^{-1}))$, we must have that $\Pr(\sum_{i=1}^{f} Y^{(i)} < f/2) \le \delta/\mathsf{MN}$. Hence we must have that $\widehat{\mathbf{P}}_{ij}$ satisfies

$$|\widehat{\mathbf{P}}_{ij} - \mathbf{P}_{ij}| = O\Big(\frac{\sigma r}{\sqrt{sd_2}}\sqrt{\frac{\mu^3 \log d_2}{p}}\Big) \tag{19}$$

with probability $1 - \delta/\mathsf{MN}$. $\square$

Now, to complete the proof of the lemma, by taking a union bound over all indices, we must have that

$$\|\widehat{\mathbf{P}} - \mathbf{P}\|_\infty \le O\Big(\frac{\sigma r}{\sqrt{sd_2}}\sqrt{\frac{\mu^3 \log d_2}{p}}\Big) \tag{20}$$

with probability at least $1 - \delta$. Also, note that the total number of rounds needed to compute these estimates is at most $\sum_z m_z sf$. From Corollary 1, we know that for any $z \in [f]$, m_z is bounded from above by $O(\mathsf{N}p + \sqrt{\mathsf{N}p\log(\mathsf{M}\delta^{-1})})$ with probability $1 - \delta$. Hence, by taking a union bound over all $z \in [f]$, the total number of rounds needed to compute all the f estimates must be bounded from above by

$$\sum_z m_z sf = O\Big(sf(\mathsf{N}p + \sqrt{\mathsf{N}p\log(\mathsf{M}\delta^{-1})})\Big)$$

with probability at least $1 - \delta f$. Therefore, by a union bound over both failure events, the total failure probability is $1 - O(\delta + \delta \log(\mathsf{MN}\delta^{-1})) = 1 - O(\delta \log(\mathsf{MN}\delta^{-1}))$. This completes the proof of the lemma.

$\square$

C MISSING PROOFS IN SECTION 4

Theorem (Restatement of Theorem 1). *Consider the rank-r online matrix completion problem with* M *users,* N *items,* T *recommendation rounds. Set* $d_2 = \min(\mathsf{M}, \mathsf{N})$. *Let* $\mathbf{R}^{(t)}_{u\rho_u(t)}$ *be the reward in each round, defined as in equation 1. Suppose* $d_2 = \Omega(\mu r \log(r d_2))$. *Let* $\mathbf{P} \in \mathbb{R}^{\mathsf{M} \times \mathsf{N}}$ *be the expected reward matrix that satisfies the conditions stated in Lemma 2 , and let* σ^2 *be the noise variance in rewards. Then, Algorithm 2, applied to the online rank-r matrix completion problem guarantees the following regret:*

$$\mathsf{Reg}(\mathsf{T}) = O\Big(\Big(\mathsf{T}^{\frac{2}{3}}(\sigma^2 r^2 \|\mathbf{P}\|_\infty)^{\frac{1}{3}}\Big(\frac{\mu^3 \mathsf{N} \log d_2}{d_2}\Big)^{1/3} + \frac{\mathsf{N}\mu^2}{d_2} \log^3 d_2\Big) \log^2(\mathsf{MNT}) + \frac{\|\mathbf{P}\|_\infty}{\mathsf{T}^2}\Big). \tag{21}$$

Proof of Theorem 1. Suppose we explore for a period of S rounds such that the exploration period (Steps 3-7 in Algorithm 2) succeeds with a probability of $1 - \nu$ i.e. conditioned on the event that the exploration period succeeds, we obtain an estimate $\widehat{\mathbf{P}}$ of the reward matrix $\mathbf{P}$ satisfying $\|\mathbf{P} - \widehat{\mathbf{P}}\|_\infty \le \eta$. In that case, conditioned on the event that the exploration period succeeds, at each step of the online algorithm in the exploitation phase (Step 9 in Algorithm 2), we suffer a regret of at most 2η. To see this, fix any user $u \in [\mathsf{M}]$. Let $i = \mathsf{argmax}_{t \in [\mathsf{N}]} \mathbf{P}_{ut}$ and $i' = \mathsf{argmax}_t \widehat{\mathbf{P}}_{ut}$ be the items with the largest rewards in the matrices $\mathbf{P}, \widehat{\mathbf{P}}$ respectively for the user u. In that case, using the fact that $\|\mathbf{P} - \widehat{\mathbf{P}}\|_\infty \le \eta$, we have

$$\mathbf{P}_{ui'} - \mathbf{P}_{ui} = \mathbf{P}_{ui'} - \widehat{\mathbf{P}}_{ui'} + \widehat{\mathbf{P}}_{ui'} - \widehat{\mathbf{P}}_{ui} + \widehat{\mathbf{P}}_{ui} - \mathbf{P}_{ui} \ge -2\eta.$$

Conditioned on the event that the exploration fails (and therefore the exploration stage as well), the regret at each step can be bounded from above by $2\|\mathbf{P}\|_\infty$. In that case, we have

$$\mathsf{Reg}(\mathsf{T}) = \frac{T}{\mathsf{M}} \sum_{u \in [\mathsf{M}]} \boldsymbol{\mu}^\star_u - \sum_{t \in [\mathsf{T}]} \frac{1}{\mathsf{M}} \sum_{u \in [\mathsf{M}]} \mathbf{P}_{u\pi_u(t)} \le \max_{u \in [\mathsf{M}]} \Big(\mathsf{T}\boldsymbol{\mu}^\star_u - \sum_{t \in [\mathsf{T}]} \mathbf{P}_{u\pi_u(t)}\Big)$$

$$\le 2\mathsf{S}\|\mathbf{P}\|_\infty + 2(\mathsf{T} - \mathsf{S})\eta \Pr(\text{Exploration succeeds}) + 2(\mathsf{T} - \mathsf{S})\|\mathbf{P}\|_\infty \Pr(\text{Exploration fails})$$

$$\le 2\mathsf{S}\|\mathbf{P}\|_\infty + 2\mathsf{T}\eta + 2\mathsf{T}\nu\|\mathbf{P}\|_\infty.$$

Now, with $d_2 = \min(\mathsf{M}, \mathsf{N})$ satisfying $d_2 = \Omega(\mu r \log(r d_2))$, we can directly use Lemma 2 with $\mathsf{S} = O(s \log(\mathsf{MN}\delta^{-1})(\mathsf{N}p + \sqrt{\mathsf{N}p \log \mathsf{M}\delta^{-1}})$ such that $\eta = O(\frac{\sigma r}{\sqrt{sd_2}}\sqrt{\frac{\mu^3 \log d_2}{p}})$ and $\nu = \delta \log(\mathsf{MN}\delta^{-1})$. Hence, we have

$$\mathsf{Reg}(\mathsf{T}) \le O\Big(\underbrace{s \log(\mathsf{MN}\delta^{-1})(\mathsf{N}p + \sqrt{\mathsf{N}p \log \mathsf{M}\delta^{-1}})\|\mathbf{P}\|_\infty}_{\text{Term 1}} + \underbrace{\mathsf{T}O\Big(\frac{\sigma r}{\sqrt{sd_2}}\sqrt{\frac{\mu^3 \log d_2}{p}}\Big)}_{\text{Term 2}}$$

$$+ \mathsf{T}(\delta \log(\mathsf{MN}\delta^{-1}))\|\mathbf{P}\|_\infty.$$

For the sake of simplicity, we ignore some of the logarithmic terms and lower order terms in the regret while choosing the value of sp that makes the regret small. To be precise, sp is chosen by minimizing the quantity $\mathsf{N}sp\|\mathbf{P}\|_\infty + \mathsf{T}\Big(\frac{\sigma r \sqrt{\mu^3 \log d_2}}{\sqrt{spd_2}}\Big)$. It can be seen that the aforementioned quantity is minimized when the two terms are equal giving us $\mathsf{N}sp\|\mathbf{P}\|_\infty = \mathsf{T}\Big(\frac{\sigma r \sqrt{\mu^3 \log d_2}}{\sqrt{spd_2}}\Big)$ implying that $sp = (\mathsf{N}\|\mathbf{P}\|_\infty)^{-2/3}\Big(\mathsf{T}\frac{\sigma r \sqrt{\mu^3 \log d_2}}{\sqrt{d_2}}\Big)^{2/3}$. Moreover, we will also choose $\delta = (\mathsf{MNT})^{-4}$. Also,

notice that $\mathsf{N}p \geq 1$ since $p \geq C\mu^2 d_2^{-1} \log^3 d_2$ (for some appropriate constant $C \geq 1$) and $\mu \geq 1$; therefore $\mathsf{N}p + \sqrt{\mathsf{N}p \log \mathsf{M}\delta^{-1}} = O(\mathsf{N}p\sqrt{\log \mathsf{M}\delta^{-1}})$. Subsequently, we have that

$$\mathsf{Reg}(\mathsf{T}) = O\Big(\underbrace{\mathsf{T}^{2/3}(\sigma^2 r^2 \|\mathbf{P}\|_\infty)^{1/3}\Big(\frac{\mu^3 \mathsf{N} \log d_2}{d_2}\Big)^{1/3} \log^2(\mathsf{MNT})}_{\text{Bound 1}} + \|\mathbf{P}\|_\infty \mathsf{T}^{-2}\Big). \qquad (22)$$

There exists an edge case when $sp \leq C\mu^2 d_2^{-1} \log^3 d_2$. Then we can substitute $s = 1$ and $p = C\mu^2 d_2^{-1} \log^3 d_2$. In that case, Term 2 will still be bounded by Bound 1 in equation 22 since $sp \leq C\mu^2 d_2^{-1} \log^3 d_2$. On the other hand Term 1 will now be bounded by $O(\frac{\mathsf{N}\mu^2}{d_2} \log^3 d_2 \log^2(\mathsf{MNT})\|\mathbf{P}\|_\infty)$. Hence, our regret will be bounded by

$$\mathsf{Reg}(\mathsf{T}) = O\Big(\mathsf{T}^{2/3}(\sigma^2 r^2 \|\mathbf{P}\|_\infty)^{1/3}\Big(\frac{\mu^3 \mathsf{N} \log d_2}{d_2}\Big)^{1/3} \log^2(\mathsf{MNT}) + \frac{\mathsf{N}\mu^2}{d_2} \log^5(\mathsf{MNT})\|\mathbf{P}\|_\infty + \|\mathbf{P}\|_\infty \mathsf{T}^{-2}\Big).$$

$$\square$$

D MISSING PROOFS IN SECTION 5

In this section, we will consider the reward matrix $\mathbf{P} = \mathbf{u}\mathbf{v}^\mathsf{T} = \lambda \bar{\mathbf{u}}\bar{\mathbf{v}}^\mathsf{T} \in \mathbb{R}^{\mathsf{M}\times\mathsf{N}}$ to be a rank-1 matrix with $\|\bar{\mathbf{u}}\|_2 = \|\bar{\mathbf{v}}\|_2 = 1$. We make the following mild assumptions in line with Theorem 2: (recall that $j_{\max} = \arg\max_i \mathbf{v}_i$ and $j_{\min} = \arg\min_i \mathbf{v}_i$

1. If $\mathsf{M} = \mathsf{T}^\zeta$ for $\zeta > \frac{1}{2}$, then the vector $\bar{\mathbf{u}}$ is $(\mathsf{M}^{-1/2\zeta}, \mu)$-incoherent i.e. for any subset $\mathcal{U} \subseteq [\mathsf{M}]$ of indices such that $|\mathcal{U}| \geq \mathsf{M}^{1-(2\zeta)^{-1}}$, we must have $\|\bar{\mathbf{v}}_\mathcal{U}\|_\infty \leq \sqrt{\frac{\mu}{|\mathcal{U}|}}\|\bar{\mathbf{v}}_\mathcal{U}\|_2$.

2. $\beta = \max\Big(\Big|\frac{\bar{\mathbf{v}}_{j_{\max}}}{\bar{\mathbf{v}}_{j_{\min}}}\Big|, \Big|\frac{\bar{\mathbf{v}}_{j_{\min}}}{\bar{\mathbf{v}}_{j_{\max}}}\Big|\Big)$ for some positive constant $\beta > 0$. Note that if we represent $\mathbf{P} = \mathbf{u}\mathbf{v}^\mathsf{T}$ where $\mathbf{u}$ is the user embedding and $\mathbf{v}$ is the item embedding, then we have $\beta = \max\Big(\Big|\frac{\mathbf{v}_{j_{\max}}}{\mathbf{v}_{j_{\min}}}\Big|, \Big|\frac{\mathbf{v}_{j_{\min}}}{\mathbf{v}_{j_{\max}}}\Big|\Big)$. In the definition of β, the second term can be larger than the first because of the absolute value.

In each phase ℓ, we will maintain three groups of users (user groups $\mathcal{B}^{(\ell)}, \mathcal{M}^{(\ell,1)}, \mathcal{M}^{(\ell,2)}$ are disjoint and form a partition of $[\mathsf{M}]$) and three corresponding groups of items (not necessarily disjoint) namely 1) $(\mathcal{B}^{(\ell)}, [\mathsf{N}])$ (initialized by $([\mathsf{M}], [\mathsf{N}])$, 2) $(\mathcal{M}^{(\ell,1)}, \mathcal{N}^{(\ell,1)})$ (initialized with (ϕ, ϕ)) and 3) $(\mathcal{M}^{(\ell,2)}, \mathcal{N}^{(\ell,2)})$ (initialized with (ϕ, ϕ)) such that $\mathcal{B}^{(\ell)} \cup \mathcal{M}^{(\ell,1)} \cup \mathcal{M}^{(\ell,2)} = [\mathsf{M}]$ and $\mathcal{N}^{(\ell,1)}, \mathcal{N}^{(\ell,2)} \subseteq [\mathsf{N}]$. Suppose we have a sequence of tuples $(m_\ell, \Delta_\ell)_\ell$ that is going to be characterized precisely later. In every phase indexed by ℓ, we will compute three matrices $\widetilde{\mathbf{Q}}^{(\ell)}, \widetilde{\mathbf{P}}^{(\ell,1)}, \widetilde{\mathbf{P}}^{(\ell,2)} \in \mathbb{R}^{\mathsf{M}\times\mathsf{N}}$ (using m_ℓ rounds for each) that correspond to estimates of three relevant sub-matrices of $\mathbf{P}$ namely $\mathbf{P}_{\mathcal{B}^{(\ell)},[\mathsf{N}]}, \mathbf{P}_{\mathcal{M}^{(\ell,1)},\mathcal{N}^{(\ell,1)}}$ and $\mathbf{P}_{\mathcal{M}^{(\ell,2)},\mathcal{N}^{(\ell,2)}}$ respectively. We will define the event $\mathcal{E}_1^{(\ell)}$ when the following holds true

$$\Big\|\widetilde{\mathbf{Q}}^{(\ell)}_{\mathcal{B}^{(\ell)},[\mathsf{N}]} - \mathbf{P}_{\mathcal{B}^{(\ell)},[\mathsf{N}]}\Big\|_\infty \leq \Delta_\ell/2$$

$$\Big\|\widetilde{\mathbf{P}}^{(\ell,1)}_{\mathcal{M}^{(\ell,1)},\mathcal{N}^{(\ell,1)}} - \mathbf{P}_{\mathcal{M}^{(\ell,1)},\mathcal{N}^{(\ell,1)}}\Big\|_\infty \leq \Delta_\ell/2$$

$$\Big\|\widetilde{\mathbf{P}}^{(\ell,2)}_{\mathcal{M}^{(\ell,2)},\mathcal{N}^{(\ell,2)}} - \mathbf{P}_{\mathcal{M}^{(\ell,2)},\mathcal{N}^{(\ell,2)}}\Big\|_\infty \leq \Delta_\ell/2$$

by using a number of rounds m_ℓ that is bounded from above by $O\Big(\max(1, \frac{\mathsf{N}}{\mathsf{M}^{1-(2\zeta)^{-1}}})\frac{\sigma^2\mu^3 \log^2(\mathsf{MNT})}{\Delta_\ell}\Big)$. The randomness stems from the inherent randomness in the algorithm.

At each phase indexed by ℓ, we will compute a temporary set of active items $\mathcal{T}_u^{(\ell+1)}$ for all users $u \in [\mathsf{N}]$. We update $\mathcal{B}^{(\ell+1)} \equiv \Big\{ u \in \mathcal{B}^{(\ell)} \mid \Big|\max_t \widetilde{\mathbf{P}}^{(\ell)}_{ut} - \min_t \widetilde{\mathbf{P}}^{(\ell)}_{ut}\Big| \leq 2a\Delta_\ell \Big\}$ where an appropriate

value of a will be determined later. This corresponds to the sets of users whose rewards are almost similar across items and hence, it does not no matter which items are picked in the next phase. We set $\mathcal{T}_u^{(\ell+1)} = [N]$ for all users $u \in \mathcal{B}^{(\ell+1)}$ and for the rest of the users (namely $u \in [N] \setminus \mathcal{B}^{(\ell+1)}$), we set $\mathcal{T}_u^{(\ell+1)}$ according to Steps 17,18 in Algorithm 3 for combined users in the set $\mathcal{B}^{(\ell)} \setminus \mathcal{B}^{(\ell+1)}$, $\mathcal{M}^{(\ell,1)}$ and $\mathcal{M}^{(\ell,2)}$ respectively. Subsequently, we update the following groups of users and corresponding items:

1. We will set v to be any user in $[M] \setminus \mathcal{B}^{(\ell+1)}$. Subsequently, we will define the set $\mathcal{M}^{(\ell+1,1)} = \{u \in [M] \setminus \mathcal{B}^{(\ell+1)} \mid \mathcal{T}_u^{(\ell+1)} \cap \mathcal{T}_v^{(\ell+1)} \neq \phi\}$. We update $\mathcal{N}^{(\ell+1,1)} = \bigcap_{u \in \mathcal{M}^{(\ell+1,1)}} \mathcal{T}_u^{(\ell+1)}$.

2. Finally, we will identify the set $\mathcal{M}^{(\ell+1,2)} \equiv [M] \setminus (\mathcal{B}^{(\ell+1)} \cup \mathcal{M}^{(\ell+1,1)})$ to be the remaining users. As before, we update $\mathcal{N}^{(\ell+1,2)} = \bigcap_{u \in \mathcal{M}^{(\ell+1,2)}} \mathcal{T}_u^{(\ell+1)}$.

For any phase indexed by ℓ, we define the event $\mathcal{E}_2^{(\ell)}$ such that for every user $u \in \mathcal{M}^{(\ell,1)}$, the following holds:

$$\left| \mathbf{P}_{ut} - \max_{t' \in [N]} \mathbf{P}_{ut} \right| \leq 2\Delta_{\ell-1} \text{ for all } t \in \mathcal{N}^{(\ell,1)}, u \in \mathcal{M}^{(\ell,1)}$$

$$\left| \mathbf{P}_{ut} - \max_{t' \in [N]} \mathbf{P}_{ut'} \right| \leq 2\Delta_{\ell-1} \text{ for all } t \in \mathcal{N}^{(\ell,2)}, u \in \mathcal{M}^{(\ell,2)}$$

$$\left| \mathbf{P}_{ut} - \max_{t' \in [N]} \mathbf{P}_{ut'} \right| \leq (2a+2)\Delta_{\ell-1} \text{ for all } t \in [N], u \in \mathcal{B}^{(\ell)}$$

Define the event $\mathcal{E}_3^{(\ell)}$ which is true when $\mathcal{M}^{(\ell,1)} \subseteq \mathcal{C}_1$, $j_{\max} \in \mathcal{N}^{(\ell,1)}$ and $\mathcal{M}^{(\ell,2)} \subseteq \mathcal{C}_2$, $j_{\min} \in \mathcal{N}^{(\ell,2)}$. Finally, let $\mathcal{E}_4^{(\ell)}$ be the event which is true when $\mathcal{M}^{(\ell,2)} \subseteq \mathcal{C}_1$, $j_{\min} \in \mathcal{N}^{(\ell,1)}$ and $\mathcal{M}^{(\ell,1)} \subseteq \mathcal{C}_2$, $j_{\max} \in \mathcal{N}^{(\ell,2)}$.

We now present a series of Lemma required for the proof of Theorem 2.

Lemma 8. *Consider two users u, v such that $u \in \mathcal{C}_1$ and $v \in \mathcal{C}_2$. Suppose the event $\mathcal{E}_1^{(\ell)}$ is true and furthermore, either $\mathcal{E}_3^{(\ell)}$ or $\mathcal{E}_4^{(\ell)}$ is true. In that case, $\mathcal{T}_u^{(\ell+1)} \cap \mathcal{T}_v^{(\ell+1)} \neq \phi$ only if either $\max_{t \in [N]} \mathbf{P}_{ut} - \min_{t \in [N]} \mathbf{P}_{ut} \leq 4\Delta_\ell$ or $\max_{t \in [N]} \mathbf{P}_{vt} - \min_{t \in [N]} \mathbf{P}_{vt} \leq 4\Delta_\ell$ holds true.*

Proof of Lemma 8. Since either $\mathcal{E}_3^{(\ell)}$ or $\mathcal{E}_4^{(\ell)}$ holds true, it must happen that $j_{\max} \in \mathcal{N}^{(\ell,1)}$ for all $u \in \mathcal{C}_1$ and $j_{\min} \in \mathcal{N}^{(\ell,2)}$ for all $u \in \mathcal{C}_2$. Further, suppose there exists an item $j \in [N]$ such that $j \in \mathcal{T}_u^{(\ell+1)} \cap \mathcal{T}_v^{(\ell+1)}$. Hence this implies that

$$-\widetilde{\mathbf{P}}_{uj}^{(\ell,1)} + \max_{t \in \mathcal{N}^{(\ell,1)}} \widetilde{\mathbf{P}}_{ut}^{(\ell,1)} \leq \Delta_\ell$$

$$\implies -\widetilde{\mathbf{P}}_{uj}^{(\ell,1)} + \mathbf{P}_{uj} - \mathbf{P}_{uj} + \mathbf{P}_{uj_{\max}} - \mathbf{P}_{uj_{\max}}$$
$$+ \widetilde{\mathbf{P}}_{uj_{\max}}^{(\ell,1)} - \widetilde{\mathbf{P}}_{uj_{\max}}^{(\ell,1)} + \max_{t \in \mathcal{N}^{(\ell,1)}} \widetilde{\mathbf{P}}_{ut}^{(\ell,1)} \leq \Delta_\ell$$

$$\implies \mathbf{P}_{uj_{\max}} - \mathbf{P}_{uj} \leq 2\Delta_\ell$$

where we used the following facts: (a) $-\mathbf{P}_{uj_{\max}} + \widetilde{\mathbf{P}}_{uj_{\max}}^{(\ell,1)} \geq -\Delta_\ell/2$ (b) $-\widetilde{\mathbf{P}}_{uj_{\max}}^{(\ell,1)} + \max_{t \in \mathcal{N}^{(\ell,1)}} \widetilde{\mathbf{P}}_{ut}^{(\ell,1)} \geq 0$ (c) $-\widetilde{\mathbf{P}}_{uj}^{(\ell,1)} + \mathbf{P}_{uj} \geq -\Delta_\ell/2$. Similarly, we must have $\mathbf{P}_{vj_{\min}} - \mathbf{P}_{vj} \leq 2\Delta_\ell$. Without loss of generality, assume that $\mathbf{u}_u \leq -\mathbf{u}_v$. In that case, we will have

$$\mathbf{P}_{uj_{\max}} - \mathbf{P}_{uj_{\min}} \leq \mathbf{P}_{uj_{\max}} - \mathbf{P}_{uj} + \mathbf{P}_{uj} - \mathbf{P}_{uj_{\min}}$$
$$\leq \mathbf{P}_{uj_{\max}} - \mathbf{P}_{uj} - \mathbf{P}_{vj} + \mathbf{P}_{vj_{\min}} \leq 4\Delta_\ell$$

where we used the fact that $\mathbf{P}$ is a rank-1 matrix and hence $\mathbf{P}_{uj} - \mathbf{P}_{uj_{\min}} = \mathbf{u}_u(\mathbf{v}_j - \mathbf{v}_{j_{\min}}) \leq -\mathbf{u}_v(\mathbf{v}_j - \mathbf{v}_{j_{\min}}) = -\mathbf{P}_{vj} + \mathbf{P}_{vj_{\min}}$. This leads to a contradiction and therefore completes the proof of the lemma. $\square$

Lemma 9. *Recall that in Step 16 of Algorithm 3, we have* $\mathcal{B}^{(\ell+1)} \equiv \Big\{ u \in \mathcal{B}^{(\ell)} \mid \left| \max_{t \in [\mathsf{N}]} \widetilde{\mathbf{Q}}_{ut}^{(\ell)} - \min_{t \in [\mathsf{N}]} \widetilde{\mathbf{Q}}_{ut}^{(\ell)} \right| \le (2a+1)\Delta_\ell \Big\}$ *for* $a > 3$*. Conditioned on the event* $\mathcal{E}_1^{(\ell)}$*, the set of users* $\mathcal{B}^{(\ell+1)}$ *must contain all users* $u \in [\mathsf{N}]$ *satisfying* $\max_{t \in [\mathsf{N}]} \mathbf{P}_{ut} - \min_{t \in [\mathsf{N}]} \mathbf{P}_{ut} \le 2a\Delta_\ell$*. Furthermore, all users* $u \in \mathcal{B}^{(\ell+1)}$ *must satisfy* $\max_{t \in [\mathsf{N}]} \mathbf{P}_{ut} - \min_{t \in [\mathsf{N}]} \mathbf{P}_{ut} \le (2a+2)\Delta_\ell$*.*

Proof of Lemma 9. Recall that $\mathcal{B}^{(\ell+1)} \equiv \Big\{ u \in \mathcal{B}^{(\ell)} \mid \max_{t \in [\mathsf{N}]} \widetilde{\mathbf{Q}}_{ut}^{(\ell)} - \min_{t \in [\mathsf{N}]} \widetilde{\mathbf{Q}}_{ut}^{(\ell)} \le (2a+1)\Delta_\ell \Big\}$ for some appropriate constant $a > 0$ that will be chosen later. Suppose for a user $u \in \mathcal{B}^{(\ell)}$, we have $\max_{t \in [\mathsf{N}]} \mathbf{P}_{ut} - \min_{t \in [\mathsf{N}]} \mathbf{P}_{ut} \le 2a\Delta_\ell$. Let $t_1 = \arg\max_{t \in [\mathsf{N}]} \widetilde{\mathbf{Q}}_{\mathbf{ut}}^{(\ell)}$, $t_2 = \arg\min_{t \in [\mathsf{N}]} \widetilde{\mathbf{Q}}_{\mathbf{ut}}^{(\ell)}$ and $t_3 = \arg\min_{t \in [\mathsf{N}]} \mathbf{P}_{ut}$. Since $\mathcal{E}_1^{(\ell)}$ is true, we must have

$$\left| \widetilde{\mathbf{Q}}_{ut_1}^{(\ell)} - \mathbf{P}_{ut_1} \right| \le \Delta_\ell/2, \quad \left| \widetilde{\mathbf{Q}}_{ut_1}^{(\ell)} - \mathbf{P}_{ut_1} \right| \le \Delta_\ell/2,$$
$$\text{and} \quad \left| \widetilde{\mathbf{Q}}_{ut_2}^{(\ell)} - \mathbf{P}_{ut_2} \right| \le \Delta_\ell/2.$$

Therefore, we must have

$$\widetilde{\mathbf{Q}}_{ut_1}^{(\ell)} - \widetilde{\mathbf{Q}}_{ut_2}^{(\ell)}$$
$$\le \widetilde{\mathbf{Q}}_{ut_1}^{(\ell)} - \mathbf{P}_{ut_1} + \mathbf{P}_{ut_1} - \max_{t \in [\mathsf{N}]} \mathbf{P}_{ut} + \max_{t \in [\mathsf{N}]} \mathbf{P}_{ut}$$
$$- \min_{t \in [\mathsf{N}]} \mathbf{P}_{ut} + \min_{t \in [\mathsf{N}]} \mathbf{P}_{ut} - \mathbf{P}_{ut_2} + \mathbf{P}_{ut_2} - \widetilde{\mathbf{Q}}_{ut_2}^{(\ell)} \le (2a+1)\Delta_\ell$$

where we used the fact that $\mathbf{P}_{ut_1} - \max_{t \in [\mathsf{N}]} \mathbf{P}_{ut} \le 0$ and $\min_{t \in [\mathsf{N}]} \mathbf{P}_{ut} - \mathbf{P}_{ut_2} \le 0$. Now, consider a user u such that $\widetilde{\mathbf{Q}}_{ut_1}^{(\ell)} - \widetilde{\mathbf{Q}}_{ut_2}^{(\ell)} \le (2a+1)\Delta_\ell$. In that case, we will have

$$\max_{t \in [\mathsf{N}]} \mathbf{P}_{ut} - \min_{t \in [\mathsf{N}]} \mathbf{P}_{ut}$$
$$\le \max_{t \in [\mathsf{N}]} \mathbf{P}_{ut} - \mathbf{P}_{ut_1} + \mathbf{P}_{ut_1} - \widetilde{\mathbf{Q}}_{ut_1}^{(\ell)} + \widetilde{\mathbf{Q}}_{ut_1}^{(\ell)} - \widetilde{\mathbf{Q}}_{ut_2}^{(\ell)}$$
$$+ \widetilde{\mathbf{Q}}_{ut_2}^{(\ell)} - \widetilde{\mathbf{Q}}_{ut_3}^{(\ell)} + \widetilde{\mathbf{Q}}_{ut_3}^{(\ell)} - \min_{t \in [\mathsf{N}]} \mathbf{P}_{ut} \le (2a+2)\Delta_\ell$$

where we used the fact that $\max_{t \in [\mathsf{N}]} \mathbf{P}_{ut} - \mathbf{P}_{ut_1} \le 0$ and $\widetilde{\mathbf{Q}}_{ut_2}^{(\ell)} - \widetilde{\mathbf{Q}}_{ut_3}^{(\ell)} \le 0$. $\square$

Lemma 10. *Suppose the event* $\mathcal{E}_1^{(\ell)}$ *is true and furthermore, either* $\mathcal{E}_3^{(\ell)}$ *or* $\mathcal{E}_4^{(\ell)}$ *is true. In that case, for every user* $u \in [\mathsf{M}] \setminus \mathcal{B}^{(\ell+1)}$*, we will have that* $\arg\max_{t \in [\mathsf{N}]} \mathbf{P}_{ut} \in \mathcal{T}_u^{(\ell+1)}$*.*

Proof of Lemma 10. Without loss of generality, let us assume that $\mathcal{E}_3^{(\ell)}$ is true. Hence, this implies that $\mathcal{M}^{(\ell,1)} \subseteq \mathcal{C}_1$ and for every user $u \in \mathcal{M}^{(\ell,1)}$, $j_{\max} \in \mathcal{N}^{(\ell,1)}$ and furthermore, $\mathcal{M}^{(\ell,2)} \subseteq \mathcal{C}_2$ and for every user $u \in \mathcal{M}^{(\ell,2)}$, $j_{\min} \in \mathcal{N}^{(\ell,2)}$. Recall that for every user $u \in \mathcal{M}^{(\ell,1)}$, we compute $\mathcal{T}_u^{(\ell+1)} = \{j \in \mathcal{N}^{(\ell,1)} \mid \widetilde{\mathbf{P}}_{uj}^{(\ell,1)} + \Delta_\ell > \max_{t \in \mathcal{N}^{(\ell,1)}} \widetilde{\mathbf{P}}_{ut}^{(\ell,1)} \}$ for all $u \in \mathcal{M}^{(\ell,1)}$. Clearly $\arg\max_{t \in \mathcal{T}_u^{(\ell+1)}} \widetilde{\mathbf{P}}_{ut}^{(\ell,1)} = \arg\max_{t \in \mathcal{N}_u^{(\ell,1)}} \widetilde{\mathbf{P}}_{ut}^{(\ell,1)}$. Let $t_1 = \arg\max_{t \in [\mathsf{N}]} \mathbf{P}_{ut}^{(\ell,1)}$ and $t_2 = \arg\max_{t \in \mathcal{N}^{(\ell,1)}} \widetilde{\mathbf{P}}_{ut}^{(\ell,1)}$. By using triangle inequality, we will have

$$\widetilde{\mathbf{P}}_{ut_1}^{(\ell,1)} - \widetilde{\mathbf{P}}_{ut_2}^{(\ell,1)} \le \widetilde{\mathbf{P}}_{ut_1}^{(\ell,1)} - \mathbf{P}_{ut_1} + \mathbf{P}_{ut_1}$$
$$- \mathbf{P}_{ut_2} + \mathbf{P}_{ut_2} - \widetilde{\mathbf{P}}_{ut_2}^{(\ell,1)} \le \Delta_\ell$$

where we used the fact that $|\widetilde{\mathbf{P}}_{ut_1}^{(\ell,1)} - \mathbf{P}_{ut_1}| \le \Delta_\ell/2$, $|\widetilde{\mathbf{P}}_{ut_2}^{(\ell,1)} - \mathbf{P}_{ut_2}| \le \Delta_\ell/2$ and $\mathbf{P}_{ut_1} - \mathbf{P}_{ut_2} \le 0$. Hence, $t_1 \in \mathcal{T}_u^{(\ell+1)}$ for any user $u \in \mathcal{M}^{(\ell,1)}$. By a similar analysis this holds for any user $u \in \mathcal{M}^{(\ell,2)} \cup (\mathcal{B}^{(\ell)} \setminus \mathcal{B}^{(\ell+1)})$ as well. $\square$

Lemma 11. *Suppose the event* $\mathcal{E}_1^{(\ell)}$ *is true and furthermore, either* $\mathcal{E}_3^{(\ell)}$ *or* $\mathcal{E}_4^{(\ell)}$ *is true. In that case the event* $\mathcal{E}_2^{(\ell+1)}$ *is true and one of* $\mathcal{E}_3^{(\ell+1)}, \mathcal{E}_4^{(\ell+1)}$ *is also true.*

Proof of Lemma 11. From Lemma 9 (and using the fact that $a > 3$ in Step 16 of Algorithm 3), we know that $\mathcal{B}^{(\ell+1)}$ must contain all users $u \in [\mathsf{N}]$ satisfying $\max_{t\in[\mathsf{N}]} \mathbf{P}_{ut} - \min_{t\in[\mathsf{N}]} \mathbf{P}_{ut} \le 4\Delta_\ell$. Therefore, from Lemma 8, there cannot exist users $u, v \in [\mathsf{N}] \setminus \mathcal{B}^{(\ell+1)}$ such that $u \in \mathcal{C}_1, v \in \mathcal{C}_2$ and $\mathcal{T}_u^{(\ell+1)} \cap \mathcal{T}_v^{(\ell+1)} \ne \phi$. Without loss of generality, suppose the user v described in Line 10 in Algorithm 3 is in $\mathcal{C}_1$. Since $\mathcal{M}^{(\ell+1,1)}$ contains all those users not in $\mathcal{B}^{(\ell+1)}$ such that $\mathcal{T}_u^{(\ell+1)} \cap \mathcal{T}_u^{(\ell+1)} \ne \phi$, $\mathcal{M}^{(\ell+1,1)}$ can only contain users in $\mathcal{C}_1$. Conversely, for any user $u \in \mathcal{C}_1 \setminus \mathcal{B}^{(\ell+1)}$, $j_{\max} \in \mathcal{T}_u^{(\ell+1)}$ (using Lemma 10) and therefore $u \in \mathcal{M}^{(\ell+1,1)}$. In that case, it must happen that $\mathcal{M}^{(\ell+1,1)} = \mathcal{C}_1 \setminus \mathcal{B}^{(\ell+1)}$ and similarly $\mathcal{M}^{(\ell+1,2)} = \mathcal{C}_2 \setminus \mathcal{B}^{(\ell+1)}$. Therefore, again by using Lemma 10, $j_{\max} \in \mathcal{N}^{(\ell+1,1)}$ for all users $u \in \mathcal{M}^{(\ell+1,1)}$ and $j_{\min} \in \mathcal{N}^{(\ell+1,2)}$ for all users $u \in \mathcal{M}^{(\ell+1,2)}$ implying that $\mathcal{E}_3^{(\ell+1)}$ holds true. On the other hand, if $v \in \mathcal{C}_2$, then by a similar analysis, $\mathcal{E}_4^{(\ell+1)}$ holds true.

Next, we will show that $\mathcal{E}_2^{(\ell+1)}$ is also going to hold true. From Lemma 9, we know that all users $u \in \mathcal{B}^{(\ell+1)}$ must satisfy $\max_{t\in[\mathsf{N}]} \mathbf{P}_{ut} - \min_{t\in[\mathsf{N}]} \mathbf{P}_{ut} \le (2a+2)\Delta_\ell$. Now, without loss of generality, assume that $\mathcal{E}_3^{(\ell)}$ is true. Consider any item $j \in \mathcal{N}^{(\ell+1,1)}$ implying that $j \in \mathcal{T}_u^{(\ell+1)}$ for all users $u \in \mathcal{M}^{(\ell+1,1)}$. Let $t_1 = \arg\max_{t\in\mathcal{T}_u^{(\ell+1)}} \widetilde{\mathbf{P}}_{ut}$. We must have

$$\mathbf{P}_{ut_1} - \mathbf{P}_{uj} = \mathbf{P}_{ut_1} - \widetilde{\mathbf{P}}_{ut_1} + \widetilde{\mathbf{P}}_{ut_1} - \widetilde{\mathbf{P}}_{uj} + \widetilde{\mathbf{P}}_{uj} - \mathbf{P}_{uj}.$$

On the other hand, since $j_{\max} \in \mathcal{T}_u^{(\ell+1)}$, we also have

$$\mathbf{P}_{uj_{\max}} - \mathbf{P}_{ut_1} = \mathbf{P}_{uj_{\max}} - \widetilde{\mathbf{P}}_{uj_{\max}}^{(\ell,1)} + \widetilde{\mathbf{P}}_{uj_{\max}}^{(\ell,1)}$$
$$- \widetilde{\mathbf{P}}_{ut_1}^{(\ell,1)} + \widetilde{\mathbf{P}}_{ut_1}^{(\ell,1)} - \mathbf{P}_{ut_1}.$$

Adding up the previous two equations, we get that

$$\mathbf{P}_{uj_{\max}} - \mathbf{P}_{uj} = \mathbf{P}_{uj_{\max}} - \widetilde{\mathbf{P}}_{uj_{\max}}^{(\ell,1)} + \widetilde{\mathbf{P}}_{uj_{\max}}^{(\ell,1)} - \widetilde{\mathbf{P}}_{ut_1}^{(\ell,1)}$$
$$+ \widetilde{\mathbf{P}}_{ut_1}^{(\ell,1)} - \widetilde{\mathbf{P}}_{uj}^{(\ell,1)} + \widetilde{\mathbf{P}}_{uj}^{(\ell,1)} - \mathbf{P}_{uj} \le 2\Delta_\ell$$

where we used the fact that $\widetilde{\mathbf{P}}_{ut_1}^{(\ell,1)} - \widetilde{\mathbf{P}}_{uj} \le \Delta_\ell$, $\widetilde{\mathbf{P}}_{uj_{\max}}^{(\ell,1)} - \widetilde{\mathbf{P}}_{ut_1}^{(\ell,1)} \le 0$, $\mathbf{P}_{uj_{\max}} - \widetilde{\mathbf{P}}_{uj_{\max}}^{(\ell,1)} \le \Delta_\ell/2$ and $\widetilde{\mathbf{P}}_{uj}^{(\ell,1)} - \mathbf{P}_{uj} \le \Delta_\ell/2$. By showing a similar analysis for all users $u \in \mathcal{M}^{(\ell+1,2)}$ and all items $j \in \mathcal{N}^{(\ell+1,2)}$. we prove that $\mathcal{E}_2^{(\ell+1)}$ holds true. $\qquad\square$

We will say that a user has been ignored in the phase indexed by ℓ under the following circumstances:

1. The size of the set of users $\mathcal{M}^{(\ell,1)}$ (or $\mathcal{M}^{(\ell,2)}$) in Step 19 is smaller than $\mathsf{M}^{1-(2\varsigma)^{-1}}$. In that case, those users are put back in the set $\mathcal{B}^{(\ell)}$ (Step 21). In that case the users in the set $\mathcal{M}^{(\ell,1)}$ (respectively $\mathcal{M}^{(\ell,2)}$) will be ignored.

2. The size of the set of users $\mathcal{B}^{(\ell)}$ in Step 6 is smaller than $\mathsf{M}^{1-(2\varsigma)^{-1}}$. In this case, even if it is added with $\mathcal{M}^{(\ell,1)}$ or $\mathcal{M}^{(\ell,2)}$, it size becomes at most $2\mathsf{M}^{1-(2\varsigma)^{-1}}$ (the size of the added set in Step 21 must also be less than $\mathsf{M}^{1-(2\varsigma)^{-1}}$ due to the first point above). In that case, the users in $\mathcal{B}^{(\ell)}$ will be ignored.

Corollary 2. *1. Suppose the event $\mathcal{E}_2^{(\ell)}$ is true for all iterations indexed by ℓ. In that case, in any phase indexed by ℓ, for every user $u \in [\mathsf{M}]$ that is not ignored in Algorithm 3 in the ℓth iteration, Algorithm 3 only recommends items that have reward at most $(2a+2)\Delta_{\ell-1}$ smaller than the reward of the best item $\arg\max_{j\in[\mathsf{N}]} \mathbf{P}_{uj}$.*

2. *Furthermore, at each iteration indexed by ℓ, at most $2\mathsf{M}^{1-(2\varsigma)^{-1}}$ users are ignored in Algorithm 3.*

Proof. The proof of the first part follows directly from Lemma 11 and the definition of $\mathcal{E}_2^{(\ell)}$. Note that for the first phase ($\ell = 1$), we have $\mathcal{B}^{(1)} = [\mathsf{M}]$ and therefore $\mathcal{E}_1^{(1)}$ being true implies that $\mathcal{E}_2^{(2)}$ is true and one of $\mathcal{E}_3^{(2)}$ or $\mathcal{E}_4^{(2)}$ is true as well.

Let us now move on to the proof of the second part. Since the three sets $\mathcal{M}^{(\ell,1)}, \mathcal{M}^{(\ell,2)}, \mathcal{B}^{(\ell)}$ (before Step 12) partition the set of users $[\mathsf{M}]$, at most two of them can be smaller than $\mathsf{M}^{1-(2\varsigma)^{-1}}$ and can be ignored. Hence, the total number of ignored users can be at most $2\mathsf{M}^{1-(2\varsigma)^{-1}}$. ∎

Lemma 12 (Incoherence). *Suppose $a > 3\max(1+\beta, 1+\frac{1}{\beta})$ in Step 6 of Algoithm 3. For any iteration in Algorithm 3 indexed by $\ell > 0$, we will have*

$$\frac{\|\mathbf{v}_{\mathcal{N}^{(\ell+1,1)}}\|_\infty}{\|\mathbf{v}_{\mathcal{N}^{(\ell+1,1)}}\|_2} \leq \frac{3\max(\beta,\beta^{-1})}{2\sqrt{|\mathcal{N}^{(\ell+1,1)}|}} \;\; and \;\; \frac{\|\mathbf{v}_{\mathcal{N}^{(\ell+1,2)}}\|_\infty}{\|\mathbf{v}_{\mathcal{N}^{(\ell+1,2)}}\|_2} \leq \frac{3\max(\beta,\beta^{-1})}{2\sqrt{|\mathcal{N}^{(\ell+1,2)}|}}$$

Proof. Consider any user $u \in [\mathsf{M}]$ such that 1) $u \in \mathcal{B}^{(\ell)}$ in the ℓth iteration but in the $(\ell+1)^{\text{th}}$ iteration, $u \in [\mathsf{N}] \setminus \mathcal{B}^{(\ell+1)}$ and 2) $\mathbf{u}_u > 0$. In this case, we must have $\mathbf{u}_u(\mathbf{v}_{j_{\max}} - \mathbf{v}_{j_{\min}}) \geq 2a\Delta_\ell$ from Lemma 9 where $a > 3\max(1+\beta, 1+\frac{1}{\beta})$. Now consider any item $j \in [\mathsf{N}]$ such that $\mathbf{P}_{uj_{\max}} - \mathbf{P}_{uj} > 2\Delta_\ell$.

In that case, by triangle inequality we have

$$\widetilde{\mathbf{P}}^{(\ell,1)}_{uj_{\max}} - \widetilde{\mathbf{P}}^{(\ell,1)}_{u} > \widetilde{\mathbf{P}}^{(\ell,1)}_{uj_{\max}} - \mathbf{P}_{uj_{\max}} + \mathbf{P}_{uj_{\max}} - \mathbf{P}_{uj} + \widetilde{\mathbf{P}}^{(\ell,1)}_{uj_{\max}} > 2\Delta_\ell - \frac{\Delta_\ell}{2} - \frac{\Delta_\ell}{2} > \Delta_\ell$$

and therefore $j \notin \mathcal{T}^{(\ell+1)}_u$. Hence consider any item $j \in [\mathsf{N}]$ that belongs to $\mathcal{T}^{(\ell+1)}_u$; it must happen that $\mathbf{u}_u(\mathbf{v}_{j_{\max}} - \mathbf{v}_j) \leq 2\Delta_\ell \leq \frac{\mathbf{u}_u(\mathbf{v}_{j_{\max}} - \mathbf{v}_{j_{\min}})}{a}$ implying that $\mathbf{v}_j \geq \mathbf{v}_{j_{\max}}\left(1 - \frac{(\beta+1)}{a}\right)$. Due to our choice of a, we get that $\mathbf{v}_j \geq \frac{2\mathbf{v}_{j_{\max}}}{3}$. Therefore $\|\mathbf{v}_{\mathcal{T}^{(\ell+1)}_u}\|_\infty \geq \frac{2\mathbf{v}_{j_{\max}}}{3}$. Since we construct the sets $\mathcal{N}^{(\ell+1,1)}$ and $\mathcal{N}^{(\ell+1,2)}$ by taking intersections of the sets $\mathcal{T}^{(\ell+1)}_u$ of items for users in $\mathcal{M}^{(\ell,1)}$ and $\mathcal{M}^{(\ell,2)}$ respectively, one of these sets (say $\mathcal{N}^{(\ell+1,1)}$) must correspond to users with positive embedding while the other will correspond to users with a negative embedding (see Lemma 11). Hence, we will have $\|\mathbf{v}_{\mathcal{N}^{(\ell+1,1)}}\|_2 \geq \sqrt{|\mathcal{N}^{(\ell+1,1)}|}\frac{2|\mathbf{v}_{j_{\max}}|}{3}$ and $\|\mathbf{v}_{\mathcal{N}^{(\ell+1,1)}}\|_\infty \leq \max(\beta,\beta^{-1})|\mathbf{v}_{j_{\max}}|$ thus completing the proof. By an analogous argument, the conclusion of the lemma is also true for users $u \in [\mathsf{M}]$ with a negative embedding i.e. $\mathbf{u}_u < 0$. ∎

Estimation of matrix $\mathbf{P}_{\mathcal{B}^{(\ell)},\mathsf{N}}$ in Step 3 of Algorithm 3 in ℓ^{th} iteration: We will only provide guarantees on estimation of the matrix $\mathbf{P}_{\mathcal{B}^{(\ell)},\mathsf{N}}$ under the condition that the number of users in $\mathcal{B}^{(\ell)}$ is at least $\mathsf{M}^{1-(2\varsigma)^{-1}}$ (recall that $\mathsf{T}^\varsigma = \mathsf{M}$). Suppose the SVD of the matrix $\mathbf{P}$ is $\lambda\bar{\mathbf{u}}\bar{\mathbf{v}}$ satisfying $\|\bar{\mathbf{u}}\| = 1, \|\bar{\mathbf{v}}\| = 1$. In that case, SVD of the matrix $\mathbf{P}_{\mathcal{B}^{(\ell)},\mathsf{N}}$ is $\lambda\|\bar{\mathbf{u}}_{\mathcal{B}^{(\ell)}}\|_2 \frac{\bar{\mathbf{u}}_{\mathcal{B}^{(\ell)}}}{\|\bar{\mathbf{u}}_{\mathcal{B}^{(\ell)}}\|_2}\bar{\mathbf{v}}$. For brevity, we will write $\widetilde{\mathbf{u}} = \frac{\bar{\mathbf{u}}_{\mathcal{B}^{(\ell)}}}{\|\bar{\mathbf{u}}_{\mathcal{B}^{(\ell)}}\|_2}$. From our local incoherence assumption (we know that $\bar{\mathbf{u}}$ is $(\mathsf{T}^{-1/2}, \mu)$-locally incoherent), we know that

$$\|\widetilde{\mathbf{u}}\|_2 = 1, \; \|\widetilde{\mathbf{u}}\|_\infty \leq \sqrt{\frac{\mu}{|\mathcal{B}^{(\ell)}|}} \text{ and } \|\bar{\mathbf{v}}\|_\infty \leq \sqrt{\frac{\mu}{\mathsf{N}}}.$$

By directly using the result in Lemma 2 for the $r = 1$ setting (rank 1) restricted to users in $\mathcal{B}^{(\ell)}$ and items in $[\mathsf{N}]$, we can recover an estimate $\widetilde{\mathbf{Q}}^{(\ell)}_{\mathcal{B}^{(\ell)},\mathsf{N}}$ of $\mathbf{P}_{\mathcal{B}^{(\ell)},\mathsf{N}}$ using $m_\ell = s_\ell \log(\mathsf{MN}\delta^{-1})(\mathsf{N}p_\ell + \sqrt{\mathsf{N}p_\ell \log \mathsf{M}\delta^{-1}}) = O(s_\ell \log(\mathsf{MN}\delta^{-1})(\mathsf{N}p_\ell \sqrt{\log \mathsf{M}\delta^{-1}}))$ rounds (where s_ℓ and p_ℓ corresponds to the parameters s, p in Lemma 2) such that

$$\|\widetilde{\mathbf{Q}}^{(\ell)}_{\mathcal{B}^{(\ell)},\mathsf{N}} - \mathbf{P}_{\mathcal{B}^{(\ell)},\mathsf{N}}\|_\infty \leq O\left(\frac{\sigma}{\sqrt{s_\ell d'_2}}\sqrt{\frac{\mu^3 \log d'_2}{p_\ell}}\right)$$

with probability $1 - O(\delta \log(\mathsf{MN}\delta^{-1}))$ where $d'_2 = \min(\mathsf{M}^{1-(2\varsigma)^{-1}}, \mathsf{N})$. In other words, by denoting $q_\ell = p_\ell s_\ell$, we have that by using $m_\ell = O(q_\ell \mathsf{N} \log^2(\mathsf{MNT}))$ rounds, we get

$$\|\widetilde{\mathbf{Q}}^{(\ell)}_{\mathcal{B}^{(\ell)},\mathsf{N}} - \mathbf{P}_{\mathcal{B}^{(\ell)},\mathsf{N}}\|_\infty \leq O\left(\frac{\sigma\sqrt{\mu^3 \log d'_2}}{\sqrt{q_\ell d'_2}}\right) \tag{23}$$

with probability at least $1 - \mathsf{T}^{-4}$. Hence, if we need an estimate $\widetilde{\mathbf{Q}}_{\mathcal{B}^{(\ell)},\mathsf{N}}^{(\ell)}$ such that $\|\widetilde{\mathbf{Q}}_{\mathcal{B}^{(\ell)},\mathsf{N}}^{(\ell)} - \mathbf{P}_{\mathcal{B}^{(\ell)},\mathsf{N}}\|_\infty \leq \Delta_\ell$ for some input parameter Δ_ℓ, then $\frac{\sigma\sqrt{\mu^3 \log d_2'}}{\sqrt{q_\ell d_2'}} \leq c'\Delta_\ell$ for some appropriate constant c' implying that $q_\ell = \frac{\sigma^2 \mu^3 \log d_2'}{c'^2 d_2' \Delta_\ell^2}$ and therefore, the total number of rounds is $O\Big(\max(1, \frac{\mathsf{N}}{\mathsf{M}^{1-(2\zeta)^{-1}}}) \frac{\sigma^2 \mu^3 \log^2(\mathsf{MNT})}{\Delta_\ell^2} \Big)$ with probability $1 - \mathsf{T}^{-4}$. By using the result in Lemma 2, we will need to use s_ℓ, p_ℓ such that $p_\ell \geq C\mu^2 (d_2')^{-1} \log^3 d_2'$ and $\frac{\sigma}{\sqrt{s_\ell}} = O\Big(\sqrt{\frac{p_\ell d_2'}{\mu^3 \log d_2'}}\|\mathbf{P}\|_\infty\Big)$. The latter condition implies that $q_\ell = s_\ell p_\ell = \frac{\sigma^2 \mu^3 \log d_2'}{c_2^2 d_2' \|\mathbf{P}\|_\infty^2}$ for some constant c_2. Putting everything together, we have that a sufficient value of q_ℓ is $\frac{\sigma^2 \mu^3 \log d_2'}{c_2^2 d_2' \|\mathbf{P}\|_\infty^2}$ provided $\Delta_\ell \leq \|\mathbf{P}\|_\infty$.

Estimation of matrix $\mathbf{P}_{\mathcal{M}^{(\ell,1)},\mathcal{N}^{(\ell,1)}}$ (and similarly $\mathbf{P}_{\mathcal{M}^{(\ell,2)},\mathcal{N}^{(\ell,2)}}$) in ℓ^{th} iteration: In this case again, we will show guarantees only when $\left|\mathcal{M}^{(\ell,1)}\right| \geq \mathsf{M}^{1-(2\zeta)^{-1}}$ (recall that $\mathsf{M}^\zeta = \mathsf{T}$). There are two sub-cases in this setting:

1. ($\mathsf{N} \geq |\mathcal{N}^{(\ell,1)}| \geq \mathsf{M}^{1-(2\zeta)^{-1}}$). In this case, as in the analysis for estimating $\mathbf{P}_{\mathcal{B}^{(\ell)},\mathsf{N}}$, we can find an estimate $\widetilde{\mathbf{P}}^{(\ell,1)} \in \mathbb{R}^{\mathsf{M} \times \mathsf{N}}$ such that $\|\widetilde{\mathbf{P}}_{\mathcal{M}^{(\ell,1)},\mathcal{N}^{(\ell,1)}}^{(\ell,1)} - \mathbf{P}_{\mathcal{M}^{(\ell,1)},\mathcal{N}^{(\ell,1)}}\| \leq \Delta_\ell$ for an input parameter Δ_ℓ using $O\Big(\max(1, \frac{\mathsf{N}}{\mathsf{M}^{1-(2\zeta)^{-1}}}) \frac{\sigma^2 \mu^3 \log^2(\mathsf{MNT})}{\Delta_\ell^2} \Big)$ rounds with probability $1 - \mathsf{T}^{-4}$ (by using Lemma 2 for users in $\mathcal{M}^{(\ell,1)}$ and items in $\mathcal{N}^{(\ell,1)}$). In the above analysis, we used the fact that after clustering (see Lemma 12), we have $|\sqrt{\mathcal{N}^{(\ell,1)}}|\|\mathbf{v}_{\mathcal{N}^{(\ell,1)}}\|_\infty = O(\|\mathbf{v}_{\mathcal{N}^{(\ell,1)}}\|_2)$. In addition, we also used the fact that the $\mathbf{u}$ is $(\mathsf{T}^{-1/2}, \mu)$-locally incoherent.

2. ($|\mathcal{N}^{(\ell,1)}| \leq \mathsf{M}^{1-(2\zeta)^{-1}}$) This is the more interesting case. Again, let us write the SVD of $\mathbf{P}_{\mathcal{M}^{(\ell,1)},\mathcal{N}^{(\ell,1)}} = \widetilde{\lambda}\widetilde{\mathbf{u}}\widetilde{\mathbf{v}}$ with $\widetilde{\mathbf{u}} \in \mathbb{R}^{|\mathcal{M}^{(\ell,1)}|}, \widetilde{\mathbf{v}} \in \mathbb{R}^{|\mathcal{N}^{(\ell,1)}|}$. We know that $\widetilde{\mathbf{u}}$ is μ-incoherent (since $|\mathcal{M}^{(\ell,1)}| \geq \mathsf{M}^{1-(2\zeta)^{-1}}$) and moreover $\widetilde{\mathbf{v}}$ is $O(1)$-incoherent (see lemma 12). Again, by directly using the result in Lemma 2 for the $r = 1$ setting (rank 1) for users in $\mathcal{M}^{(\ell,1)}$ and items in $\mathcal{N}^{(\ell,1)}$), we can recover an estimate $\widehat{\mathbf{P}}_{\mathcal{M}^{(\ell,1)},\mathcal{N}^{(\ell,1)}}^{(\ell,1)}$ of $\mathbf{P}_{\mathcal{M}^{(\ell,1)},\mathcal{N}^{(\ell,1)}}$ (where $\widehat{\mathbf{P}}^{(\ell,1)} \in \mathbb{R}^{\mathsf{M} \times \mathsf{N}}$) using $m_\ell = O(s_\ell \log(\mathsf{MN}\delta^{-1})(|\mathcal{N}^{(\ell,1)}|p_\ell\sqrt{\log \mathsf{M}\delta^{-1}}))$ rounds (where s_ℓ and p_ℓ corresponds to the parameters s, p in Lemma 2) such that

$$\|\widehat{\mathbf{P}}_{\mathcal{M}^{(\ell,1)},\mathcal{N}^{(\ell,1)}}^{(\ell,1)} - \mathbf{P}_{\mathcal{M}^{(\ell,1)},\mathcal{N}^{(\ell,1)}}\|_\infty \leq O\Big(\frac{\sigma}{\sqrt{s_\ell |\mathcal{N}^{(\ell,1)}|}}\sqrt{\frac{\mu^3 \log |\mathcal{N}^{(\ell,1)}|}{p_\ell}}\Big)$$

with probability $1 - O(\delta \log \delta^{-1})$. As before, let us denote $q_\ell = p_\ell s_\ell$. We have that by using $m_\ell = O(q_\ell |\mathcal{N}^{(\ell,1)}| \log^2 \mathsf{MNT})$ rounds, we get

$$\|\widetilde{\mathbf{P}}_{\mathcal{M}^{(\ell,1)},\mathcal{N}^{(\ell,1)}}^{(\ell,1)} - \mathbf{P}_{\mathcal{M}^{(\ell,1)},\mathcal{N}^{(\ell,1)}}\|_\infty \leq O\Big(\frac{\sigma\sqrt{\mu^3 \log |\mathcal{N}^{(\ell,1)}|}}{\sqrt{q_\ell |\mathcal{N}^{(\ell,1)}|}}\Big) \tag{24}$$

with probability at least $1 - \mathsf{T}^{-4}$. Hence, if we need an estimate $\widetilde{\mathbf{P}}_{\mathcal{M}^{(\ell,1)},\mathcal{N}^{(\ell,1)}}^{(\ell,1)}$ such that $\|\widetilde{\mathbf{P}}_{\mathcal{M}^{(\ell,1)},\mathcal{N}^{(\ell,1)}}^{(\ell,1)} - \mathbf{P}_{\mathcal{M}^{(\ell,1)},\mathcal{N}^{(\ell,1)}}\|_\infty \leq \Delta_\ell$ for some input parameter Δ_ℓ, then $\frac{\sigma\sqrt{\mu^3 \log |\mathcal{N}^{(\ell,1)}|}}{\sqrt{q_\ell |\mathcal{N}^{(\ell,1)}|}} \leq c'\Delta_\ell$ for some appropriate constant c' implying that a sufficient value of $q_\ell = \frac{\sigma^2 \mu^3 \log \mathsf{N}}{c'^2 |\mathcal{N}^{(\ell,1)}|\Delta_\ell^2}$ and therefore, the total number of rounds is $O\Big(\frac{\sigma^2 \mu^3 \log \mathsf{N} \log^2(\mathsf{MNT})}{\Delta_\ell^2}\Big)$ with probability $1 - \mathsf{T}^{-4}$. Again, by using the result in Lemma 2, we will need to use s_ℓ, p_ℓ such that $p_\ell \geq C\mu^2 (|\mathcal{N}^{(\ell,1)}|)^{-1} \log^3 |\mathcal{N}^{(\ell,1)}|$ and $\frac{\sigma}{\sqrt{s_\ell}} = O\Big(\sqrt{\frac{p_\ell |\mathcal{N}^{(\ell,1)}|}{\mu^3 \log |\mathcal{N}^{(\ell,1)}|}}\|\mathbf{P}\|_\infty\Big)$. Again, the latter condition implies that $q_\ell = s_\ell p_\ell = \frac{\sigma^2 \mu^3 \log |\mathcal{N}^{(\ell,1)}|}{c_2^2 |\mathcal{N}^{(\ell,1)}|\|\mathbf{P}\|_\infty^2}$ for some constant c_2. Putting everything together, we have that a sufficient value of q_ℓ is $\frac{\sigma^2 \mu^3 \log |\mathcal{N}^{(\ell,1)}|}{c_2^2 |\mathcal{N}^{(\ell,1)}|\Delta_\ell^2}$ provided $\Delta_\ell \leq \|\mathbf{P}\|_\infty$.

Remark 12. *There is an edge case when $\left|\mathcal{N}^{(\ell,1)}\right|$ becomes so small that the condition $\left|\mathcal{N}^{(\ell,1)}\right| = \Omega(\mu \log \left|\mathcal{N}^{(\ell,1)}\right|)$ is not satisfied. Hence, we must have $\left|\mathcal{N}^{(\ell,1)}\right| = O(\mu)$. If this becomes true, then in theory, a condition in Lemma 2 becomes violated and low rank matrix completion will not work. However, an easy fix in this situation is to implement the standard Upper-Confidence Bound algorithm for each user in $\mathcal{M}^{(\ell,1)}$ separately until the end of rounds (from the phase ℓ when the condition becomes false) with the set of items $\mathcal{N}^{(\ell,1)}$ (and also for users in $\mathcal{B}^{(\ell)}$ that are inserted into $\mathcal{M}^{(\ell,1)}$ in subsequent phases). From Remark 1, this step will only add a regret of $O(\sigma\sqrt{\left|\mathcal{N}^{(\ell,1)}\right| \mathsf{T} \log \mathsf{T}}) = O(\sigma\sqrt{\mu \mathsf{T} \log \mathsf{T}})$ which is a significantly lower order term than the regret guarantee proved in Theorem 2.*

Again, by using the result in Lemma 2, we will use s_ℓ to be the minimum positive integer satisfying $\frac{\sigma}{\sqrt{s_\ell}} = O\left(\Delta_\ell \sqrt{\frac{|\mathcal{N}^{(\ell,1)}|}{\mu^3 \log |\mathcal{N}^{(\ell,1)}|}}\right)$ and $\sqrt{p_\ell} = \frac{\sigma r}{\sqrt{s_\ell |\mathcal{N}^{(\ell,1)}|}} \frac{\sqrt{\mu^3 \log |\mathcal{N}^{(\ell,1)}|}}{c\Delta_\ell}$ for some appropriate constant $c > 0$.

Now, we are ready to state and prove the main theorems:

Theorem (Restatement of Theorem 2). *Consider the rank-1 online matrix completion problem with T rounds, M users s.t. $\mathsf{M} \geq \sqrt{\mathsf{T}}$ and N items. Denote $d_2 = \min(\mathsf{M}, \mathsf{N})$. Let $\mathbf{R}^{(t)}_{u\rho_u(t)}$ be the reward in each round, defined as in equation 1. Let σ^2 be the noise variance in rewards and let $\mathbf{P} \in \mathbb{R}^{\mathsf{M} \times \mathsf{N}}$ be the expected reward matrix with SVD decomposition $\mathbf{P} = \lambda \bar{\mathbf{u}}\bar{\mathbf{v}}^\mathsf{T}$ such that $\bar{\mathbf{u}}$ is $(\mathsf{T}^{-1/2}, \mu)$-locally incoherent, $\|\bar{\mathbf{v}}\|_\infty \leq \sqrt{\mu/\mathsf{N}}$, $d_2 = \Omega(\mu \log d_2)$ and $|\bar{\mathbf{v}}_{j_{\min}}| = \Theta(|\bar{\mathbf{v}}_{j_{\max}}|)$. Then, by suitably choosing parameters $\{\Delta_\ell\}_\ell$, positive integers $\{s_{(\ell,0)}, s_{(\ell,1)}, s_{(\ell,2)}\}_\ell$ and $1 \geq \{p_{(\ell,0)}, p_{(\ell,1)}, p_{(\ell,2)}\}_\ell \geq 0$ as described in Algorithm 3, we can ensure a regret guarantee of $\mathrm{Reg}(\mathsf{T}) = O(\sqrt{\mathsf{T}}\|\mathbf{P}\|_\infty + \mathsf{J}\sqrt{\mathsf{T}V})$ where $\mathsf{J} = O\left(\log\left(\frac{1}{\sqrt{\mathsf{T}^{-1}}}\min\left(\|\mathbf{P}\|_\infty, \frac{\sigma\sqrt{\mu}}{\log \mathsf{N}}\right)\right)\right)$ and $V = \left(\max(1, \frac{\mathsf{N}\sqrt{\mathsf{T}}}{\mathsf{M}})\sigma^2\mu^3 \log^2(\mathsf{MNT})\right)$.*

Proof of Theorem 2. In the ℓth iteration, we set $\Delta_\ell = C'2^{-\ell}\min\left(\|\mathbf{P}\|_\infty, \frac{\sigma\sqrt{\mu}}{\log \mathsf{N}}\right)$ for some appropriate constant $C' > 0$. In the ℓth iteration, our goal is to obtain estimates of the matrices $\mathbf{P}_{\mathcal{B}^{(\ell)},\mathsf{N}}$ (using Lemma 2 with parameters $(s_{(\ell,0)}, p_{(\ell,0)})$ corresponding to s, p in Lemma 2), $\mathbf{P}_{\mathcal{M}^{(\ell,1)},\mathcal{N}^{(\ell,1)}}$ (using Lemma 2 with parameters $(s_{(\ell,1)}, p_{(\ell,1)})$) and $\mathbf{P}_{\mathcal{M}^{(\ell,2)},\mathcal{N}^{(\ell,2)}}$ (using Lemma 2 with parameters $(s_{(\ell,2)}, p_{(\ell,2)})$ up to an error of $\Delta_\ell/2$ with high probability.

From Lemma 2, for any sub-matrix $\mathbf{P}_{\mathsf{sub}} \in \mathbb{R}^{\mathsf{M}' \times \mathsf{N}'}$ of $\mathbf{P}_{\mathsf{sub}}$, we can compute an estimate $\widehat{\mathbf{P}}_{\mathsf{sub}}$ satisfying $\|\widehat{\mathbf{P}}_{\mathsf{sub}} - \mathbf{P}_{\mathsf{sub}}\|_\infty \leq \Delta_\ell/2$ with high probability by setting the parameters s_ℓ, p_ℓ such that

$$\frac{\sigma r}{\sqrt{s_\ell d'_2}}\sqrt{\frac{\mu^3 \log d'_2}{p_\ell}} = c\Delta_\ell \tag{25}$$

for some constant $c > 0$ where $d'_2 = \min(\mathsf{M}', \mathsf{N}')$. The largest possible error Δ_ℓ that is possible to obtain by using the matrix completion technique is obtained by substituting $s = 1$ and $p = C\mu^2(d'_2)^{-1}\log^3 d'_2$; we obtain that $\Delta_\ell \leq \frac{\sigma r\sqrt{\mu}}{\log d'_2}$. Note that $d'_2 \leq \mathsf{N}$ (in both cases when $\mathsf{M} \geq \mathsf{N}$ and vice-versa) and therefore, if we choose $\Delta_\ell \leq C'\min\left(\frac{\sigma\sqrt{\mu}}{\log \mathsf{N}}\right)$ for some appropriate constant $C' > 0$, then there exists s_ℓ and p_ℓ for which we can obtain $\|\widehat{\mathbf{P}}_{\mathsf{sub}} - \mathbf{P}_{\mathsf{sub}}\|_\infty \leq \Delta_\ell/2$ with high probability. Moreover, if $\Delta_\ell \leq \|\mathbf{P}\|_\infty$, we can also ensure that $\frac{\sigma}{\sqrt{s_\ell}} = O\left(\sqrt{\frac{p_\ell|\mathcal{N}^{(\ell,1)}|}{\mu^3 \log |\mathcal{N}^{(\ell,1)}|}}\|\mathbf{P}\|_\infty\right)$. Hence, we can ignore any lower bounds on p_ℓ and upper bounds on σ as they are automatically satisfied due to the instantiation of Δ_ℓ.

Also from the analysis above (see equations 23 and 24), we know that by using a total number of rounds in the ℓth phase that is bounded from above by $O\left(\max(1, \frac{\mathsf{N}}{\mathsf{M}^{1-(2\varsigma)^{-1}}})\frac{\sigma^2\mu^3 \log \mathsf{N}\log^2(\mathsf{MNT})}{\Delta_\ell^2}\right)$ (see the analysis for the sufficient number of rounds for estimating $\mathbf{P}_{\mathcal{B}^{(\ell)},\mathsf{N}}$ in the ℓth iteration) , we

have with probability $1 - O(\mathsf{T}^{-4})$ (the event $\mathcal{E}_1^{(\ell)}$ for the ℓth iteration)

$$\left\|\widetilde{\mathbf{Q}}^{(\ell)}_{\mathcal{B}^{(\ell)},[\mathsf{N}]} - \mathbf{P}_{\mathcal{B}^{(\ell)},[\mathsf{N}]}\right\|_\infty \le \frac{\Delta_\ell}{2} \tag{26}$$

$$\left\|\widetilde{\mathbf{P}}^{(\ell,1)}_{\mathcal{M}^{(\ell,1)},\mathcal{N}^{(\ell,1)}} - \mathbf{P}_{\mathcal{M}^{(\ell,1)},\mathcal{N}^{(\ell,1)}}\right\|_\infty \le \frac{\Delta_\ell}{2} \tag{27}$$

$$\left\|\widetilde{\mathbf{P}}^{(\ell,2)}_{\mathcal{M}^{(\ell,2)},\mathcal{N}^{(\ell,2)}} - \mathbf{P}_{\mathcal{M}^{(\ell,2)},\mathcal{N}^{(\ell,2)}}\right\|_\infty \le \frac{\Delta_\ell}{2}. \tag{28}$$

We condition on the events $\mathcal{E}_1^{(\ell)}$ being true for all ℓ. The probability that there exists any ℓ such that the event $\mathcal{E}_1^{(\ell)}$ is false is $O(\mathsf{T}^{-4})$; hence the probability that $\mathcal{E}_1^{(\ell)}$ is true for all ℓ is at least $1 - O(\mathsf{T}^{-3})$ (the total number of iterations can be at most T). In case, one such event $\mathcal{E}_1^{(\ell)}$ is false, we can bound the regret trivially by $O(\mathsf{T}^{-3}\|\mathbf{P}\|_\infty)$. By an inductive argument (See Lemma 10 and the first part of Corollary 2), we know that the events $\mathcal{E}_2^{(\ell)}$ will be true for all ℓ. We will also denote the set of rounds in phase ℓ by $\mathcal{T}_\ell \subseteq [\mathsf{T}]$ (therefore $|\mathcal{T}_\ell| = m_\ell$). Let us compute the regret restricted to the rounds in $\mathcal{T}^{(\ell)}$ conditioned on the events $\mathcal{E}_1^{(\ell)}, \mathcal{E}_2^{(\ell)}$ being true for all ℓ as follows:

$$\frac{m_\ell}{\mathsf{M}} \sum_{u\in[\mathsf{M}]} \boldsymbol{\mu}_u^\star - \sum_{t\in\mathcal{T}^{(\ell)}} \frac{1}{\mathsf{M}} \sum_{u\in[\mathsf{M}]} \mathbf{P}^{(t)}_{u\rho_u(t)}$$

$$= \underbrace{\frac{m_\ell}{\mathsf{M}} \sum_{u\in[\mathsf{M}]:u\text{ not ignored}} \boldsymbol{\mu}_u^\star - \sum_{t\in\mathcal{T}^{(\ell)}} \frac{1}{\mathsf{M}} \sum_{u\in[\mathsf{M}]:u\text{ not ignored}} \mathbf{P}^{(t)}_{u\rho_u(t)}}_{\text{Term 1}}$$

$$+ \underbrace{\frac{m_\ell}{\mathsf{M}} \sum_{u\in[\mathsf{M}]:u\text{ ignored}} \boldsymbol{\mu}_u^\star - \sum_{t\in\mathcal{T}^{(\ell)}} \frac{1}{\mathsf{M}} \sum_{u\in[\mathsf{M}]:u\text{ ignored}} \mathbf{P}^{(t)}_{u\rho_u(t)}}_{\text{Term 2}}$$

The second term can be bounded from above by $O(m_\ell \mathsf{M}^{1-(2\zeta)^{-1}}\|\mathbf{P}\|_\infty/\mathsf{M}) = O(m_\ell \mathsf{M}^{-(2\zeta)^{-1}}\|\mathbf{P}\|_\infty)$ (see second part of Corollary 2). The first term can be bounded by using Corollary 2 conditioned on the events $\mathcal{E}_1^{(\ell)}, \mathcal{E}_2^{(\ell)}$ being true for all ℓ,

$$(2a+1)\Delta_{\ell-1}m_\ell = O\Big(\max(1, \frac{\mathsf{N}}{\mathsf{M}^{1-(2\zeta)^{-1}}})\frac{\sigma^2\mu^3 \log \mathsf{N} \log^2(\mathsf{MNT})}{\Delta_\ell}\Big). \tag{29}$$

Putting everything together, we can now bound the total regret as follows (using $\Delta_{\ell-1} = 2\Delta_\ell$):

$$\mathsf{Reg}(\mathsf{T}) = \frac{\mathsf{T}}{\mathsf{M}} \sum_{u\in[\mathsf{M}]} (\mu_u^\star - \sum_{t\in[\mathsf{T}]} \mathbb{E}\mathbf{R}^{(t)}_{u\rho_u(t)})$$

$$= \sum_\ell O\Big(\Delta_\ell m_\ell \mid \mathcal{E}_1^{(\ell)}, \mathcal{E}_2^{(\ell)} \text{ is true for all } \ell\Big) + O(\mathsf{M}^{-(2\zeta)^{-1}}\|\mathbf{P}\|_\infty m_\ell \mid \mathcal{E}_1^{(\ell)}, \mathcal{E}_2^{(\ell)} \text{ is true for all } \ell) + O(\mathsf{T}^{-3}\|\mathbf{P}_\infty\|)$$

$$\le \sum_\ell O\Big(\Delta_\ell m_\ell \mid \mathcal{E}_1^{(\ell)}, \mathcal{E}_2^{(\ell)} \text{ is true for all } \ell\Big) + O(\sqrt{\mathsf{T}}\|\mathbf{P}\|_\infty)$$

where we used the fact that $\sum_\ell m_\ell = \mathsf{T}$ and $\mathsf{T}^\zeta = \mathsf{M}$. As stated in the theorem, let $\mathsf{V} = \Big(\max(1, \frac{\mathsf{N}}{\mathsf{M}^{1-(2\zeta)^{-1}}})\sigma^2\mu^3 \log \mathsf{N} \log^2(\mathsf{MNT})$. Moving on, we can decompose the first term into two parts:

$$\mathsf{Reg}(\mathsf{T}) \le O(\sqrt{\mathsf{T}}\|\mathbf{P}\|_\infty) + O\Big(\sum_{\ell:\Delta_\ell \le \Phi} \Delta_\ell m_\ell \mid \mathcal{E}_1^{(\ell)}, \mathcal{E}_2^{(\ell)} \text{ is true for all } \ell\Big)$$

$$+ O\Big(\sum_{\ell:\Delta_\ell > \Phi} \Delta_\ell \mathsf{V}\Delta_\ell^{-2} \mid \mathcal{E}_1^{(\ell)}, \mathcal{E}_2^{(\ell)} \text{ is true for all } \ell\Big)$$

$$\le O(\sqrt{\mathsf{T}}\|\mathbf{P}\|_\infty) + \mathsf{T}\Phi + O\Big(\sum_{\ell:\Delta_\ell > \Phi} \mathsf{V}\Delta_\ell^{-1}\Big)$$

Since we chose $\Delta_\ell = C' 2^{-\ell} \min\left(\|\mathbf{P}\|_\infty, \frac{\sigma\sqrt{\mu}}{\log N}\right)$ for some constant $C' > 0$, the maximum number of phases ℓ for which $\Delta_\ell > \Phi$ can be bounded from above by $J = O\left(\log\left(\frac{1}{\Phi}\min\left(\|\mathbf{P}\|_\infty, \frac{\sigma\sqrt{\mu}}{\log N}\right)\right)\right)$. Hence, we have

$$\mathsf{Reg}(\mathsf{T}) \le O(\sqrt{\mathsf{T}}\|\mathbf{P}\|_\infty) + O(\mathsf{T}\Phi) + O\left(\mathsf{JV}\Phi^{-1}\right)$$

$$= O(\sqrt{\mathsf{T}}\|\mathbf{P}\|_\infty) + O(\mathsf{J}\sqrt{\mathsf{TV}})$$

where we substituted $\Phi = \sqrt{\mathsf{VT}^{-1}}$ and hence $J = O\left(\log\left(\frac{1}{\sqrt{\mathsf{VT}^{-1}}}\min\left(\|\mathbf{P}\|_\infty, \frac{\sigma\sqrt{\mu}}{\log N}\right)\right)\right)$ in the final step. $\qquad\square$

Theorem (Restatement of Theorem 3). *Let $\mathbf{P}$ be a rank 1 matrix such that $0 \le \mathbf{P}_{ij} \le 1$ for all $i, j \in [\mathsf{M}] \times [\mathsf{N}]$ and the noise variance $\sigma^2 = 1$. In that case, for problem P1, any algorithm will suffer a regret of at least $\Omega(\sqrt{\mathsf{NTM}^{-1}})$.*

Proof of Theorem 3. Recall that the well-known multi-armed bandit (MAB) problem is a special case of our setting when the number of users $\mathsf{M} = 1$. We can cast the reward matrix as a $1 \times \mathsf{N}$ matrix $\mathbf{P}$ whose i^{th} element is denoted by $\mathbf{P}_i$. In that case, the expected regret $\mathsf{Reg}_{\mathsf{MAB}}(\mathsf{T})$ over a time period of T steps is $\mathsf{T}\max_{j\in[\mathsf{N}]}\mathbf{P}_j - \sum_{t\in[\mathsf{T}]}\mathbf{P}_{\rho(t)}$ where $\rho(t)$ is the item recommended at time t. For the MAB problem, it is known that any algorithm must suffer a regret of at least $\Omega(\sqrt{\mathsf{NT}})$ (see Theorem 15.2 in Lattimore & Szepesvári (2020)). Consider our problem with M users and N items when all users $u \in [\mathsf{M}]$ are identical and furthermore, in each round, the item recommended for the ith user can depend on the rewards obtained for users $1, 2, \dots, i-1$ in that round and the rewards obtained in the previous rounds. In that case, any algorithm that achieves a regret $\mathsf{Reg}(\mathsf{T})$ in the repeated setting and will achieve a regret of $\mathsf{M}^{-1}\mathsf{Reg}_{\mathsf{MAB}}(\mathsf{MT})$ in the MAB problem and vice-versa. Hence, we have that $\mathsf{Reg}(\mathsf{T}) = \mathsf{M}^{-1}\mathsf{Reg}_{\mathsf{MAB}}(\mathsf{MT}) = \Omega(\sqrt{(\mathsf{NT}/\mathsf{M})})$.

$\qquad\square$

E SMALL NUMBER OF USERS M (PROOF OF THEOREM E)

As in section D, we will consider the reward matrix $\mathbf{P} = \lambda\bar{\mathbf{u}}\bar{\mathbf{v}}^\mathsf{T} \in \mathbb{R}^{\mathsf{M}\times\mathsf{N}}$ to be a rank-1 matrix with $\|\bar{\mathbf{u}}\|_2 = \|\bar{\mathbf{v}}\|_2 = 1$. We make the following assumptions where we use the notations presented in Section D:

1. We assume that the vector $\bar{\mathbf{u}}$ is $(\min(|\mathcal{C}_1|, |\mathcal{C}_2|), \mu)$-incoherent. This is stricter than the first assumption presented in Section D.

2. $\beta = \max\left(\left|\frac{\bar{\mathbf{v}}_{j_{\max}}}{\bar{\mathbf{v}}_{j_{\min}}}\right|, \left|\frac{\bar{\mathbf{v}}_{j_{\min}}}{\bar{\mathbf{v}}_{j_{\max}}}\right|\right)$ for some positive constant $\beta > 0$. Note that if we represent $\mathbf{P} = \mathbf{u}\mathbf{v}^\mathsf{T}$ where $\mathbf{u}$ is the user embedding and $\mathbf{v}$ is the item embedding, then we have $\beta = \max\left(\left|\frac{\mathbf{v}_{j_{\max}}}{\mathbf{v}_{j_{\min}}}\right|, \left|\frac{\mathbf{v}_{j_{\min}}}{\mathbf{v}_{j_{\max}}}\right|\right)$.

Here, we have a slight modification of the phased algorithm (see Algorithm 8). In each phase ℓ, we will maintain three groups of users and items $(\mathcal{B}^{(\ell)}, \mathcal{N}^{(\ell,1)}\cup\mathcal{N}^{(\ell,2)}), (\mathcal{M}^{(\ell,1)}, \mathcal{N}^{(\ell,1)})$ and $(\mathcal{M}^{(\ell,2)}, \mathcal{N}^{(\ell,2)})$ such that $\mathcal{B}^{(\ell)} \cup \mathcal{M}^{(\ell,1)} \cup \mathcal{M}^{(\ell,2)} = [\mathsf{M}]$ and $\mathcal{N}^{(\ell,1)}, \mathcal{N}^{(\ell,2)} \subseteq [\mathsf{N}]$. Here $\mathcal{B}^{(\ell)}, \mathcal{M}^{(\ell,1)}, \mathcal{M}^{(\ell,2)}$ are initialized by $[\mathsf{M}], \phi, \phi$ respectively and $\mathcal{N}^{(\ell,1)}, \mathcal{N}^{(\ell,2)}$ are initialized by $[\mathsf{N}], \phi$ respectively.

Suppose we have a sequence of tuples $(m_\ell, \Delta_\ell)_\ell$ that is going to be characterized precisely later. In every phase indexed by ℓ, we will compute two matrices $\widetilde{\mathbf{P}}^{(\ell,1)}, \widetilde{\mathbf{P}}^{(\ell,2)} \in \mathbb{R}^{\mathsf{M}\times\mathsf{N}}$ (using m_ℓ rounds for each) that correspond to estimates of two relevant sub-matrices of $\mathbf{P}$ namely $\mathbf{P}_{\mathcal{M}^{(\ell,1)}\cup\mathcal{B}^{(\ell)},\mathcal{N}^{(\ell,1)}}$ and $\mathbf{P}_{\mathcal{M}^{(\ell,2)}\cup\mathcal{B}^{(\ell)},\mathcal{N}^{(\ell,2)}}$ respectively. We will define the event $\mathcal{E}_1^{(\ell)}$ when the following holds true

$$\left\|\widetilde{\mathbf{P}}_{\mathcal{M}^{(\ell,1)}\cup\mathcal{B}^{(\ell)},\mathcal{N}^{(\ell,1)}} - \mathbf{P}_{\mathcal{M}^{(\ell,1)}\cup\mathcal{B}^{(\ell)},\mathcal{N}^{(\ell,1)}}\right\|_\infty \le \Delta_\ell/2$$

$$\left\|\widetilde{\mathbf{P}}_{\mathcal{M}^{(\ell,2)}\cup\mathcal{B}^{(\ell)},\mathcal{N}^{(\ell,2)}} - \mathbf{P}_{\mathcal{M}^{(\ell,2)}\cup\mathcal{B}^{(\ell)},\mathcal{N}^{(\ell,2)}}\right\|_\infty \le \Delta_\ell/2$$

Algorithm 8 OCTAL ALGORITHM FOR SMALL NUMBER OF USERS M

Require: Number of users M, items N, rounds T, noise σ^2, bound on the entry-wise magnitude of expected rewards $\|\mathbf{P}\|_\infty$, incoherence μ.
1: Set $\mathcal{M}^{(1,1)} = \mathcal{M}^{(1,2)} = \phi$ and $\mathcal{B}^{(1)} = [\mathsf{M}]$. Set $\mathcal{N}^{(1,1)} = \mathcal{N}^{(1,2)} = \phi$. Set $f = O(\log(\mathsf{MNT}))$ and suitable constants $a, c, C, C', C_\lambda > 0$.
2: **for** $\ell = 1, 2, \ldots,$ **do**
3:     Set $\Delta_\ell = C' 2^{-\ell} \min\left(\|\mathbf{P}\|_\infty, \frac{\sigma\sqrt{\mu}}{\log \mathsf{N}}\right)$.
4:     **for** $k = 1, 2, \ldots, f$ **do**
5:         **for** each pair of non-null sets $(\mathcal{M}^{(\ell,1)} \cup \mathcal{B}^{(\ell)}, \mathcal{N}^{(\ell,1)}), (\mathcal{M}^{(\ell,2)} \cup \mathcal{B}^{(\ell)}, \mathcal{N}^{(\ell,2)}) \subseteq [\mathsf{M}] \times [\mathsf{N}]$ **do**
6:             Denote $(\mathcal{T}^{(1)}, \mathcal{T}^{(2)})$ to be the considered pair of sets and $i \in \{1, 2\}$ to be its index.
7:             Set $d_{2,i} = \min(|\mathcal{T}^{(1)}|, |\mathcal{T}^{(2)}|)$. Set $p_{\ell,i} = C\mu^2 d_{2,i}^{-1} \log^3 d_{2,i}$ and $s_{\ell,i} = \left\lceil \left(\frac{c\sigma\sqrt{\mu}}{\Delta_\ell \log d_{2,i}}\right)^2 \right\rceil$.
8:             For each tuple of indices $(u, v) \in \mathcal{T}^{(1)} \times \mathcal{T}^{(2)}$, independently set $\delta_{uv} = 1$ with probability $p_{\ell,i}$ and $\delta_{uv} = 0$ with probability $1 - p_{\ell,i}$.
9:             Denote $\Omega^{(i)} = \{(u, v) \in \mathcal{T}^{(1)} \times \mathcal{T}^{(2)} \mid \delta_{uv} = 1\}$ and $b_{\ell,i} = \max_{u \in \mathcal{U}} |v \in \mathcal{V} \mid (u, v) \in \Omega|$. Set total number of rounds to be $m_{\ell,i} = b_{\ell,i} s_{\ell,i}$.
10:         **end for**
11:         For $i \in \{1, 2\}$, compute $\widetilde{\mathbf{P}}^{(\ell,i,f)} = \text{ESTIMATE}(m_{\ell,i}, b_{\ell,i}, \Omega^{(i)}, |\mathcal{M}^{(\ell,i)}|, |\mathcal{N}^{(\ell,i)}|, \lambda = C_\lambda \sigma\sqrt{d_{2,i} p_{\ell,i}})$. *# Algorithm 1 recommends items to every user in $\mathcal{M}^{(\ell,i)} \cup \mathcal{B}^{(\ell)}$ for $m_{\ell,i}$ rounds. Since users in $\mathcal{B}^{(\ell)}$ get recommended items for $m_{\ell,1} + m_{\ell,2}$ rounds, for $i \in \{1, 2\}$ we can recommend arbitrary items in $\mathcal{N}^{(\ell,i)}$ for users in $\mathcal{M}^{(\ell,i)}$ for the additional $m_{(\ell,3-i)}$ rounds.*
12:     **end for**
13:     Compute $\widetilde{\mathbf{P}}^{(\ell,i)} = \text{Entrywise Median}(\{\widetilde{\mathbf{P}}^{(\ell,i,k)}\}_{k=1}^f)$ for $i \in \{1, 2\}$.
14:     Set $\mathcal{B}^{(\ell+1)} \subseteq \mathcal{B}^{(\ell)}$ to be the set

$$\left\{u \in \mathcal{B}^{(\ell)} \mid \left|\max\left(\max_{t \in \mathcal{N}^{(\ell,1)}} \widetilde{\mathbf{P}}_{ut}^{(\ell,1)}, \max_{t \in \mathcal{N}^{(\ell,2)}} \widetilde{\mathbf{P}}_{ut}^{(\ell,2)}\right) - \min\left(\min_{t \in \mathcal{N}^{(\ell,1)}} \widetilde{\mathbf{P}}_{ut}^{(\ell,1)}, \min_{t \in \mathcal{N}^{(\ell,2)}} \widetilde{\mathbf{P}}_{ut}^{(\ell,2)}\right)\right| \leq 2a\Delta_\ell\right\}.$$

15:     Set $\mathsf{A}_u = \max(\max_{t \in \mathcal{N}^{(\ell,1)}} \widetilde{\mathbf{P}}_{ut}^{(\ell,1)}, \max_{t \in \mathcal{N}^{(\ell,2)}} \widetilde{\mathbf{P}}_{ut}^{(\ell,2)})$. Compute $\mathcal{T}_u^{(\ell+1)} = \{j \in \mathcal{N}^{(\ell,1)}\} \mid \widetilde{\mathbf{P}}_{uj}^{(\ell,1)} + \Delta_\ell > \mathsf{A}_u\} \cup \{j \in \mathcal{N}^{(\ell,2)}\} \mid \widetilde{\mathbf{P}}_{uj}^{(\ell,2)} + \Delta_\ell > \mathsf{A}_u\}$ for all $u \in \mathcal{B}^{(\ell)} \setminus \mathcal{B}^{(\ell+1)}$.
16:     For $i \in \{1, 2\}$, for all users $u \in \mathcal{M}^{(\ell,i)}$, compute $\mathcal{T}_u^{(\ell+1)} = \{j \in \mathcal{N}^{(\ell,i)} \mid \widetilde{\mathbf{P}}_{uj}^{(\ell,i)} + \Delta_\ell > \max_{t \in \mathcal{N}^{(\ell,i)}} \widetilde{\mathbf{P}}_{ut}^{(\ell,i)}\}$.
17:     Set v to be any user in $[\mathsf{M}] \setminus \mathcal{B}^{(\ell+1)}$. Set $\mathcal{M}^{(\ell+1,1)} = \{u \in [\mathsf{M}] \setminus \mathcal{B}^{(\ell+1)} \mid \mathcal{T}_u^{(\ell+1)} \cap \mathcal{T}_v^{(\ell+1)} \neq \phi\}$. Set $\mathcal{M}^{(\ell+1,2)} = [\mathsf{M}] \setminus (\mathcal{B}^{(\ell+1)} \cup \mathcal{M}^{(\ell+1,1)})$.
18:     Compute $\mathcal{N}^{(\ell+1,1)} = \bigcap_{u \in \mathcal{M}^{(\ell+1,1)}} \mathcal{T}_u^{(\ell+1)}$, $\mathcal{N}^{(\ell+1,2)} = \bigcap_{u \in \mathcal{M}^{(\ell+1,2)}} \mathcal{T}_u^{(\ell+1)}$.
19: **end for**

---

by using a number of rounds m_ℓ that is bounded from above by $O\left(\max(1, \frac{\mathsf{N}}{\mathsf{M}^{1-(2\zeta)^{-1}}}) \frac{\sigma^2 \mu^3 \log^2(\mathsf{MNT})}{\Delta_\ell}\right)$.

Thus, the only difference with the analysis in Section D is that for the users in $\mathcal{B}^{(\ell)}$, we consider the items in $\mathcal{N}^{(\ell,1)} \cup \mathcal{N}^{(\ell,2)}$ (which can be shown to comprise the best arms for both clusters); therefore we combine the users in $\mathcal{B}^{(\ell)}$ with users in $\mathcal{M}^{(\ell,1)}$ and users in $\mathcal{M}^{(\ell,2)}$ separately to obtain the matrix estimates. Since users in $\mathcal{B}^{(\ell)}$ need to recommended more items in each phases, we can recommend arbitrary items to users in $\mathcal{M}^{(\ell,1)}$ ($\mathcal{M}^{(\ell,1)}$) from $\mathcal{N}^{(\ell,1)}$ ($\mathcal{N}^{(\ell,1)}$ respectively.)

At each phase indexed by ℓ, we will compute a temporary set of active items $\mathcal{T}_u^{(\ell+1)}$ for all users $u \in [\mathsf{N}]$. We update $\mathcal{B}^{(\ell+1)}$ as given in Step 14 of Algorithm 8. We update the sets of users $\mathcal{M}^{(\ell+1,1)}$, $\mathcal{M}^{(\ell+1,2)}$ as in Section D (see Steps 17,18 in Algorithm 8). For any phase indexed by ℓ, we define the event $\mathcal{E}_2^{(\ell)}$ such that for every user $u \in \mathcal{M}^{(\ell,1)}$, the following holds:

$$\left|\mathbf{P}_{ut} - \max_{t \in [\mathsf{N}]} \mathbf{P}_{ut}\right| \leq 2\Delta_{\ell-1} \text{ for all } t \in \mathcal{N}^{(\ell,1)}, u \in \mathcal{M}^{(\ell,1)}$$

$$\left|\mathbf{P}_{ut} - \max_{t \in [\mathsf{N}]} \mathbf{P}_{ut}\right| \leq 2\Delta_{\ell-1} \text{ for all } t \in \mathcal{N}^{(\ell,2)}, u \in \mathcal{M}^{(\ell,2)}$$

$$\left|\mathbf{P}_{ut} - \max_{t \in [\mathsf{N}]} \mathbf{P}_{ut}\right| \leq (2a+2)\Delta_{\ell-1} \text{ for all } t \in \mathcal{N}^{(\ell,1)} \cup \mathcal{N}^{(\ell,2)}, u \in \mathcal{B}^{(\ell)}.$$

As before, we define the events $\mathcal{E}_3^{(\ell)}$ which is true when $\mathcal{M}^{(\ell,1)} \subseteq \mathcal{C}_1$, $j_{\max} \in \mathcal{N}^{(\ell,1)}$ and $\mathcal{M}^{(\ell,2)} \subseteq \mathcal{C}_2$, $j_{\min} \in \mathcal{N}^{(\ell,2)}$. Finally, let $\mathcal{E}_4^{(\ell)}$ be the event which is true when $\mathcal{M}^{(\ell,2)} \subseteq \mathcal{C}_1$, $j_{\min} \in \mathcal{N}^{(\ell,1)}$ and $\mathcal{M}^{(\ell,1)} \subseteq \mathcal{C}_2$, $j_{\max} \in \mathcal{N}^{(\ell,2)}$. By a similar analysis as in Section D, we can show that Lemmas 8, 9, 10 and 11, the first part of 2 (note that no users are *ignored* in this analysis) and 12 all hold true with very minor modifications wherever required.

Estimation of matrix $\mathbf{P}_{\mathcal{M}^{(\ell,1)} \cup \mathcal{B}^{(\ell)}, \mathcal{N}^{(\ell,1)}}$ (and similarly $\mathbf{P}_{\mathcal{M}^{(\ell,2)} \cup \mathcal{B}^{(\ell)}, \mathcal{N}^{(\ell,2)}}$) in ℓ^{th} iteration: In this case, notice that the size of $\left|\mathcal{M}^{(\ell,1)} \cup \mathcal{B}^{(\ell)}\right| \geq \min(|\mathcal{C}_1|, |\mathcal{C}_2|)$ since the combination of the two sets must comprise the users of the cluster $\mathcal{M}^{(\ell,1)}$ corresponds to. There are two sub-cases in this setting:

1. $(\mathsf{N} \geq |\mathcal{N}^{(\ell,1)}| \geq \min(|\mathcal{C}_1|, |\mathcal{C}_2|))$. In this case, as in the analysis for estimating $\mathbf{P}_{\mathcal{B}^{(\ell)}, \mathsf{N}}$, we can find an estimate $\widetilde{\mathbf{P}}_{\mathcal{M}^{(\ell,1)} \cup \mathcal{B}^{(\ell)}, \mathcal{N}^{(\ell,1)}} \in \mathbb{R}^{\mathsf{M} \times \mathsf{N}}$ such that $\left\|\widetilde{\mathbf{P}}_{\mathcal{M}^{(\ell,1)} \cup \mathcal{B}^{(\ell)}, \mathcal{N}^{(\ell,1)}} - \mathbf{P}_{\mathcal{M}^{(\ell,1)} \cup \mathcal{B}^{(\ell)}, \mathcal{N}^{(\ell,1)}}\right\|_{\infty} \leq \Delta_\ell/2$ for an input parameter Δ_ℓ using $O\left(\max(1, \frac{\mathsf{N}}{\min(|\mathcal{C}_1|, |\mathcal{C}_2|)}) \frac{\sigma^2 \mu^3 \log^2(\mathsf{MNT})}{\Delta_\ell^2}\right)$ rounds with probability $1 - \mathsf{T}^{-4}$ (by using Lemma 2 for users in $\mathcal{M}^{(\ell,1)}$ and items in $\mathcal{N}^{(\ell,1)}$). In the above analysis, we used Lemma 12 and the fact that the $\mathbf{u}$ is $(\mathsf{M}^{-1} \min(|\mathcal{C}_1|, |\mathcal{C}_2|), \mu)$-locally incoherent.

2. $(|\mathcal{N}^{(\ell,1)}| \leq \min(|\mathcal{C}_1|, |\mathcal{C}_2|))$ Again, by a similar analysis as in Section D and by using Lemma 12 and the fact that the $\mathbf{u}$ is $(\mathsf{M}^{-1} \min(|\mathcal{C}_1|, |\mathcal{C}_2|), \mu)$-locally incoherent, the total number of sufficient rounds is $O\left(\frac{\sigma^2 \mu^3 \log \mathsf{N} \log^2(\mathsf{MNT})}{\Delta_\ell^2}\right)$ with probability $1 - \mathsf{T}^{-4}$.

Now, we are ready to state and prove the main theorems:

Theorem. *Consider the rank-1 online matrix completion problem with M users, N items, T recommendation rounds such that $\mathsf{M} \geq \sqrt{\mathsf{T}}$. Denote $d_2 = \min(\mathsf{M}, \mathsf{N})$. Let $\mathbf{R}_{u\rho_u(t)}^{(t)}$ be the reward in each round, defined as in equation 1. Let σ^2 be the noise variance in rewards and let $\mathbf{P} \in \mathbb{R}^{\mathsf{M} \times \mathsf{N}}$ be the expected reward matrix with SVD decomposition $\mathbf{P} = \lambda \bar{\mathbf{u}} \bar{\mathbf{v}}^\top$ such that $\bar{\mathbf{u}}$ is $(\min(|\mathcal{C}_1|, |\mathcal{C}_2|), \mu)$-locally incoherent, $\|\bar{\mathbf{v}}\|_\infty \leq \sqrt{\mu/\mathsf{N}}$, $d_2 = \Omega(\mu \log d_2)$ and $|\bar{\mathbf{v}}_{j_{\min}}| = \Theta(|\bar{\mathbf{v}}_{j_{\max}}|)$. Then, by suitably choosing parameters $\{\Delta_\ell\}_\ell$, positive integers $\{s_{(\ell,1)}, s_{(\ell,2)}\}_\ell$ and $1 \geq \{p_{(\ell,1)}, p_{(\ell,2)}\}_\ell \geq 0$ as described in Algorithm 3, we can ensure a regret guarantee of $\mathsf{Reg}(\mathsf{T}) = O(\sqrt{\mathsf{T}} \|\mathbf{P}\|_\infty + \mathsf{J}\sqrt{\mathsf{T}\mathsf{V}})$ where $\mathsf{J} = O\left(\log\left(\frac{1}{\sqrt{\mathsf{V}\mathsf{T}^{-1}}} \min\left(\|\mathbf{P}\|_\infty, \frac{\sigma\sqrt{\mu}}{\log \mathsf{N}}\right)\right)\right)$ and $\mathsf{V} = \left(\max(1, \frac{\mathsf{N}}{\min(|\mathcal{C}_1|, |\mathcal{C}_2|)}) \sigma^2 \mu^3 \log^3(\mathsf{MNT})\right)$.*

Proof of Theorem E. The proof follows on similar lines as the proof in Theorem 2. The only change is in m_ℓ which is reflected in the result (more specifically in V). Also, note that in Algorithm 8, no users are *ignored* and so we do not need to bound the regret for ignored users.

□