Differentially Private Matrix Completion through Low-rank Matrix Factorization

Lingxiao Wang
TTIC

Boxin Zhao
University of Chicago

Mladen Kolar
University of Chicago

Abstract

We study the matrix completion problem under joint differential privacy and develop a non-convex low-rank matrix factorization-based method for solving it. Our method comes with strong privacy and utility guarantees, has a linear convergence rate, and is more scalable than the best-known alternative (Chien et al., 2021). Our method achieves the (near) optimal sample complexity for matrix completion required by the non-private baseline and is much better than the best known result under joint differential privacy. Furthermore, we prove a tight utility guarantee that improves existing approaches and removes the impractical resampling assumption used in the literature. Numerical experiments further demonstrate the superiority of our method.

1 INTRODUCTION

The completion of low-rank matrices has been extensively studied in the past decade (Candès and Tao, 2010; Rohde et al., 2011; Chen and Wainwright, 2015; Zheng and Lafferty, 2016; Wang et al., 2017; Chi et al., 2019) due to its wide range of applications in real-world problems, such as recommendation systems (Rennie and Srebro, 2005), clustering (Cai et al., 2008), and sensor localization (Wang et al., 2008). When personalizing recommendations, the goal is to learn the preferences of a set of m users on a collection of n items, using a small portion of the observed ratings. Specifically, given the observed index set $\Omega \subseteq [m] \times [n]$ and the partially observed ratings X^*_{ij}, $(i,j) \in \Omega$, where X^*_{ij} denotes the rating of user i for the item j, the goal is to estimate the user-item rating matrix $\mathbf{X}^* \in \mathbb{R}^{m\times n}$. Assuming that $\mathbf{X}^*$ has low-rank, the goal is to learn two matrices, $\mathbf{U} \in \mathbb{R}^{m\times r}$ and $\mathbf{V} \in \mathbb{R}^{n\times r}$, of rank r ($r \ll \min\{m,n\}$) such that $\mathbf{X}^* \approx \mathbf{U}\mathbf{V}^\top$. The low-rank matrices $\mathbf{U}, \mathbf{V}$ represent the embeddings of users and items, and can be used to predict i-th user's rating on j-th item using $\mathbf{U}_{i*}\mathbf{V}^\top_{*j}$, where $\mathbf{U}_{i*}, \mathbf{V}_{*j}$ denote i-th row and j-th column of $\mathbf{U}$ and $\mathbf{V}$.

Data sets used to train low-rank models contain sensitive personal information, such as personal preferences and location data, which raise serious privacy concerns (Narayanan and Shmatikov, 2010; Calandrino et al., 2011). Privacy-preserving matrix completion (Hardt and Roth, 2012, 2013; Kapralov and Talwar, 2013) aims to achieve differential privacy (DP) (Dwork et al., 2006) for individual records. Unfortunately, these approaches only provide privacy guarantees for a single entry in the low-rank matrix, which is not suitable and can even be detrimental in many real-world problems. For example, in the context of personalized recommendations, ratings from a user are often correlated and can be used to fingerprint this user (Calandrino et al., 2011). Therefore, the privacy guarantee for a single rating may not be strong enough to address each user's privacy concerns. In addition, if we naively strengthen the privacy guarantee for each rating, e.g., by achieving the same DP guarantee for every rating, the DP guarantee (see Definition 3.1) for a single rating implies that a user's predicted preference for a new item cannot be inferred from their own preferences on similar items. As a result, the predicted recommendations generated by a privacy-preserving low-rank matrix completion method are often inaccurate and useless. To address these concerns, Jain et al. (2018); Chien et al. (2021) developed more practical methods for private matrix completion based on user-level privacy or joint DP (Kearns et al., 2014). Specifically, these methods aim to keep all ratings of a user private, rather than a single rating. Joint DP implies that recommendations for a given user cannot be inferred by other users, even if they collude with each other (see Definition 3.2). Therefore, it is a strong privacy guarantee for each user as long as the user herself does not make recommendations given to her public. To achieve more accurate recommendations under the joint DP, previous work (McSherry and Mironov, 2009; Liu et al., 2015; Jain et al., 2018; Chien et al., 2021) proposed using the server and user computation framework. More specifically, a (trusted) server will compute the global model (e.g., the shared item embeddings $\mathbf{V}$) using aggregated information from all users. Then,

Proceedings of the 26th International Conference on Artificial Intelligence and Statistics (AISTATS) 2023, Valencia, Spain. PMLR: Volume 206.

each user can independently compute her local model (e.g., embedding $\mathbf{U}_{i*}$ for user $i \in [m]$) based on the global model and generate accurate recommendations for herself (e.g., $\mathbf{U}_{i*}\mathbf{V}^\top$ for user $i \in [m]$). Since we compute the global model based on the information from all users, we can add sufficient noise to make it private without significantly reducing the accuracy of recommendations.

We investigate the problem of privacy-preserving matrix completion and propose a novel method, named DPLMC, that satisfies joint DP using the server and user computation framework. Previous work on low-rank matrix completion under joint DP has provided methods that are computationally inefficient, offer unsatisfactory utility guarantees, and have much worse sample complexity than non-private methods (McSherry and Mironov, 2009; Liu et al., 2015; Jain et al., 2018; Chien et al., 2021). Our proposed method is based on non-convex low-rank matrix factorization (Tu et al., 2016; Wang et al., 2017) and overcomes the limitations of existing approaches. DPLMC is computationally efficient, has reduced sample complexity requirements, and provides improved utility guarantees. Table 1 offers a comparison with existing strong baselines.

Contributions. We develop a differentially private algorithm for matrix completion that comes with a joint differential privacy guarantee. Unlike the state-of-the-art approach (Chien et al., 2021), our method has a linear convergence rate and is more scalable since it does not require matrix inversion or projection onto the cone of positive semidefinite matrices. Additionally, we have improved the best-known sample complexity of a private method (Chien et al., 2021) by a factor of $\widetilde{O}(r^4)$ and matched the best known sample complexity for non-private methods (Zheng and Lafferty, 2016; Wang et al., 2017). Furthermore, our utility guarantee has improved on the best-known utility bound (Chien et al., 2021) by a factor of $\widetilde{O}(r)$. Notably, our utility analysis does not require the impractical resampling assumption made in Chien et al. (2021). Finally, we have empirically evaluated the performance of the proposed method against the state-of-the-art approach.

Notation. For a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, we use $\mathbf{A}_{i*} \in \mathbb{R}^n$ and $\mathbf{A}_{*j} \in \mathbb{R}^m$ to denote the i-th row and j-th column of $\mathbf{A}$, respectively. The (i,j)-th element of $\mathbf{A}$ is denoted by A_{ij}. The k-th largest singular value of $\mathbf{A}$ is denoted by $\sigma_k(\mathbf{A})$, $\|\mathbf{A}\|_{2,\infty} = \max_{i\in[m]} \|\mathbf{A}_{i*}\|_2$ and $\|\mathbf{A}\|_{\infty,\infty} = \max_{i,j} |A_{ij}|$. For two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^r$, we use $[\mathbf{x}^\top; \mathbf{y}^\top] \in \mathbb{R}^{2\times r}$ to denote the matrix with rows $\mathbf{x}^\top$ and $\mathbf{y}^\top$. For any positive integer m, we use $[m]$ to denote the set $\{1, 2, \ldots, m\}$. For two sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n = O(b_n)$ if there exists $C > 0$ such that $a_n \le Cb_n$, and we write $a_n = \mathbf{\Omega}(b_n)$ if there exists $C > 0$ such that $a_n \ge Cb_n$. We use $\widetilde{O}(\cdot), \widetilde{\mathbf{\Omega}}(\cdot)$ to further hide logarithmic factors and ignore polynomial factors in singular values and the incoherence parameter (c.f. Section 3).

2 RELATED WORK

Our paper is related to the literature on matrix completion under joint DP (McSherry and Mironov, 2009; Liu et al., 2015; Jain et al., 2018; Chien et al., 2021). In particular, McSherry and Mironov (2009) proposed a singular value decomposition-based method without any utility guarantee analysis, while Liu et al. (2015) developed a stochastic gradient Langevin dynamics-based algorithm, which lacks utility guarantees and may require an exponential amount of computation to obtain privacy parameters. Jain et al. (2018) developed a differentially private Frank-Wolfe algorithm with provable utility guarantees, but their method has a sublinear convergence rate and significantly worse sample complexity and utility bounds than existing non-private methods. The method by Chien et al. (2021) involves matrix inversions and projections onto the cone of positive semidefinite matrices, making it computationally expensive, and its utility guarantees depend on the impractical resampling assumption. In contrast, our proposed algorithm for matrix completion under joint DP is efficient and has better sample complexity and utility guarantees compared to existing methods. See Table 1 for an overview.

It is worth noting that our method shares a similar idea with differentially private gradient descent methods (Feldman et al., 2020; Zhou et al., 2021). However, these methods have certain limitations. For instance, Feldman et al. (2020) only consider the convex optimization problem and, therefore, their method cannot deal with the nonconvex optimization problem considered in our paper; see (3.3). On the other hand, while Zhou et al. (2021) study the general nonconvex optimization problem, their method only guarantees convergence to a stationary point at a sublinear rate and cannot exploit the structure of low-rank matrix factorization. Additionally, they can only provide entry-level privacy guarantees instead of the user-level privacy guarantee required for matrix completion.

3 PRELIMINARIES

We introduce the noisy matrix completion problem and the projected gradient descent approach to solving it. We further discuss several differential privacy notions.

3.1 Nonconvex Matrix Completion

We consider the matrix completion problem with noisy observations (Rohde et al., 2011; Koltchinskii et al., 2011; Negahban and Wainwright, 2012). Our primary goal is to recover the unknown low-rank matrix $\mathbf{X}^* \in \mathbb{R}^{m\times n}$ from a set of randomly observed noisy entries. We assume the uniform observation model (Koltchinskii et al., 2011; Negahban and Wainwright, 2012) in which the elements of the

Table 1: Comparison of the different methods to recover $\mathbf{X}^*$ under the context of (ϵ, δ)-joint DP (see Definition 3.2). We report the required sample complexity and utility guarantee, where m is the number of users, n is the number of items, $|\Omega|$ is the number of observations, and r is the rank of $\mathbf{X}^*$. The symbol (†) denotes methods that require a good initialization (see Section 5.3).

Methods	Bounds on $\lvert\Omega\rvert$	Utility	Convergence Rate
Private FW (Jain et al., 2018)	$\widetilde{\mathbf{\Omega}}(nm^{1/2})$	$\widetilde{O}\left(\frac{rn^{1/2}}{(m\epsilon)^{2/5}}\right)$	Sub-linear
Private ALS (†) (Chien et al., 2021)	$\widetilde{\mathbf{\Omega}}(r^6 m)$	$\widetilde{O}\left(\frac{r^5 n^2}{m\lvert\Omega\rvert\epsilon^2}\right)$	Linear
DPLMC (†) (Theorem 5.2)	$\widetilde{\mathbf{\Omega}}(r^2 m)$	$\widetilde{O}\left(\frac{r^4 n^2}{m\lvert\Omega\rvert\epsilon^2}\right)$	Linear

observation matrix $\mathbf{Y} = (Y_{ij}) \in \mathbb{R}^{m\times n}$ are generated as

$$Y_{ij} := \begin{cases} X^*_{ij} + E_{ij}, & \text{for any } (i,j) \in \Omega, \\ *, & \text{otherwise,} \end{cases} \tag{3.1}$$

where $\Omega \subseteq [m] \times [n]$ is the set of observed indices such that for any $(i,j) \in \Omega$, $i \sim$ uniform$([m])$ and $j \sim$ uniform$([n])$, and $\mathbf{E} = (E_{ij}) \in \mathbb{R}^{m\times n}$ is a random noise matrix. Note that we only consider the uniform observation model in this paper, but our methods and results can be extended to the more general weighted sampling model (Negahban and Wainwright, 2012). The uniform observation model has been widely assumed in the literature to recover low-rank matrices, and previous work (Chien et al., 2021) provides theoretical guarantees based on this model. Real-world datasets may not necessarily follow the uniform observation model. Nonetheless, gradient-based algorithms have demonstrated good performance in practice, as shown in previous studies (see, e.g., Zheng and Lafferty, 2016; Wang et al., 2017; Park et al., 2018; Chi et al., 2019).

Recovery of the unknown matrix $\mathbf{X}^*$ of rank $r \ll \min\{m, n\}$ is impossible if $\mathbf{X}^*$ is too sparse (Gross, 2011; Negahban and Wainwright, 2012). Therefore, the following incoherence condition is assumed on $\mathbf{X}^*$ (Candès and Recht, 2009). Let $\mathbf{X}^* = \mathbf{U}^*\mathbf{\Sigma}\mathbf{V}^{*\top}$ be the singular value decomposition of $\mathbf{X}^*$, where $\mathbf{U}^* \in \mathbb{R}^{m\times r}$, $\mathbf{V}^* \in \mathbb{R}^{n\times r}$, $\mathbf{\Sigma} \in \mathbb{R}^{r\times r}$, and $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$ are the singular values of $\mathbf{X}^*$. The incoherence condition states that

$$\|\mathbf{U}^*\|_{2,\infty} \leq \sqrt{\frac{\beta r}{m}} \quad \text{and} \quad \|\mathbf{V}^*\|_{2,\infty} \leq \sqrt{\frac{\beta r}{n}}, \tag{3.2}$$

where β denotes the incoherence parameter. Based on the incoherence condition (3.2), we have $\|\mathbf{X}^*\|_{\infty,\infty} \leq \|\mathbf{U}^*\|_{2,\infty} \cdot \|\mathbf{\Sigma}\|_2 \cdot \|\mathbf{V}^*\|_{2,\infty} \leq \beta r : \sigma_1/\sqrt{mn}$.

3.2 Projected Gradient Descent

Gradient descent-based methods (Tu et al., 2016; Zheng and Lafferty, 2016; Wang et al., 2017; Park et al., 2018; Chi et al., 2019) are widely used to recover $\mathbf{X}^*$ by minimizing the following nonconvex optimization problem

$$\min_{\mathbf{U}\in\mathcal{C}_1, \mathbf{V}\in\mathcal{C}_2} \frac{1}{2p}\big\|\mathcal{P}_\Omega\big(\mathbf{U}\mathbf{V}^\top - \mathbf{Y}\big)\big\|_F^2 + \frac{1}{8}\|\mathbf{U}^\top\mathbf{U} - \mathbf{V}^\top\mathbf{V}\|_F^2, \tag{3.3}$$

where $p = |\Omega|/(mn)$, $\mathcal{P}_\Omega : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$ is a projection operator such that

$$\big(\mathcal{P}_\Omega(\mathbf{Y})\big)_{ij} := \begin{cases} Y_{ij}, & \text{if } (i,j) \in \Omega, \\ 0, & \text{otherwise,} \end{cases}$$

where $\mathcal{C}_1 = \{\mathbf{A} \in \mathbb{R}^{m\times r} \,\big|\, \|\mathbf{A}\|_{2,\infty} \leq \sqrt{\beta r\sigma_1/m}\}$, and $\mathcal{C}_2 = \{\mathbf{A} \in \mathbb{R}^{n\times r} \,\big|\, \|\mathbf{A}\|_{2,\infty} \leq \sqrt{\beta r\sigma_1/n}\}$. To ensure that the produced estimator satisfies the constraints, a projection step is often used (Wang et al., 2017). The (projected) gradient descent-based method enjoys a linear convergence rate and achieves the (near) optimal sample complexity $\widetilde{\mathbf{\Omega}}(r^2 m)$. Given these advantages, we propose to develop our differentially private algorithm based on gradient descent.

3.3 Privacy Notion

We give the definition of the data set and the mechanism that we consider in this paper. Let $S = \{s_1, s_2, \ldots, s_m\}$ be a data set with m users, where s_i denotes the data of user $i \in [m]$. We use S_{-i} to denote the data set without data from the i-th user, and $(s_i; S_{-i})$ denotes the data set obtained by adding data from the i-th user to S_{-i}. Let $\mathcal{M} : \mathcal{S}^m \to \mathcal{R}^m$ be a mechanism that produces m outputs, one for each user, and let $\mathcal{M}_{-i}$ be the set of outputs without the output for the i-th user. The standard (ϵ, δ)-differential privacy provides the privacy guarantee with respect to each entry in the data set (Dwork et al., 2006).

Definition 3.1 ((ϵ, δ)-DP (Dwork et al., 2006))**.** A randomized mechanism $\mathcal{M}$ satisfies (ϵ, δ)-differential privacy if, for any user i, any adjacent datasets $s_i, \widetilde{s}_i \in \mathcal{S}$, any data set for other users $S_{-i} \in \mathcal{S}^{m-1}$, and any output $O \subseteq \mathcal{R}^m$, it holds that

$$\mathbb{P}[\mathcal{M}(s_i; S_{-i}) \in O] \leq e^\epsilon \cdot \mathbb{P}[\mathcal{M}(\widetilde{s}_i; S_{-i}) \in O] + \delta.$$

We aim to develop a method that satisfies joint differential privacy.

Definition 3.2 (Joint Differential Privacy (Kearns et al., 2014))**.** A randomized mechanism $\mathcal{M}$ satisfies (ϵ, δ)-joint differential privacy if, for any user i, any possible values $s_i, \widetilde{s}_i \in \mathcal{S}$, any data set for other users $S_{-i} \in \mathcal{S}^{m-1}$, and any output $O \subseteq \mathcal{R}^{m-1}$, we have

$$\mathbb{P}[\mathcal{M}_{-i}(s_i; S_{-i}) \in O] \leq e^{\epsilon} \cdot \mathbb{P}[\mathcal{M}_{-i}(\widetilde{s}_i; S_{-i}) \in O] + \delta.$$

Although Joint DP is widely used in the literature, it suffers from the loose composition result, which makes it unsuitable for iterative learning algorithms. Instead, we will use the following notion of joint Rényi differential privacy, which is an extension of joint DP. Joint Rényi differential privacy is particularly useful when the data set is accessed by a sequence of randomized mechanisms (Mironov, 2017).

Definition 3.3 (Joint Rényi Differential Privacy)**.** A randomized mechanism $\mathcal{M}$ satisfies (γ, ρ)-joint Rényi differential privacy if, for any user i, any possible values $s_i, \widetilde{s}_i \in \mathcal{S}$, any data set for other users $S_{-i} \in \mathcal{S}^{m-1}$, we have

$$D_\gamma\big(\mathcal{M}_{-i}(s_i; S_{-i}) \| \mathcal{M}_{-i}(\widetilde{s}_i; S_{-i})\big) \leq \rho,$$

where $\gamma > 1, \rho > 0$, and D_γ is the Rényi divergence

$$D_\gamma\big(\mathcal{M}_{-i}(s_i; S_{-i}) \| \mathcal{M}_{-i}(\widetilde{s}_i; S_{-i})\big) := \log \mathbb{E}\big[\big(\mathcal{M}_{-i}(s_i; S_{-i})/\mathcal{M}_{-i}(\widetilde{s}_i; S_{-i})\big)^{\gamma}\big]/(\gamma - 1).$$

First, we will show that our procedure satisfies the joint RDP. Using the results in Mironov (2017), we can subsequently show that it also satisfies the joint DP.

4 PROPOSED METHOD

We present a projected gradient descent-based method for solving the noisy matrix completion problem that has a linear convergence rate, achieves a strong utility guarantee, and satisfies the joint RDP. As we discuss in Section 1, we follow the idea of the server and user computation framework (McSherry and Mironov, 2009; Liu et al., 2015; Jain et al., 2018; Chien et al., 2021) to develop our algorithm: (1) a trusted server obtains a private global model using the aggregated information from all users; (2) each user attains a local model to generate her own recommendations based on the private global model. This framework requires communication between the server and users, and therefore it is important to develop a method with a fast convergence rate so that the number of rounds of communication is small, as the communication cost is the key bottleneck in practice.

Local and global updates. We describe our proposed global and local updates in detail. At the t-th iteration, the gradient descent-based method performs the following update

$$\mathbf{U}^{t+1} = \mathbf{U}^t - \frac{\eta}{p}\mathcal{P}_\Omega\big(\widehat{\mathbf{Y}}^t - \mathbf{Y}\big)\mathbf{V}^t - \frac{\eta}{2}\mathbf{U}^t\big((\mathbf{U}^t)^\top\mathbf{U}^t - (\mathbf{V}^t)^\top\mathbf{V}^t\big), \tag{4.1}$$

$$\mathbf{V}^{t+1} = \mathbf{V}^t - \frac{\eta}{p}\mathcal{P}_\Omega\big(\widehat{\mathbf{Y}}^t - \mathbf{Y}\big)^\top\mathbf{U}^t - \frac{\eta}{2}\mathbf{V}^t\big((\mathbf{V}^t)^\top\mathbf{V}^t - (\mathbf{U}^t)^\top\mathbf{U}^t\big), \tag{4.2}$$

where η is the step size and $\widehat{\mathbf{Y}}^t = \mathbf{U}^t(\mathbf{V}^t)^\top$. The update above is used to solve the nonconvex optimization problem in (3.3). Therefore, we can rewrite the update rule for $\mathbf{U}$ in (4.1) as follows:

$$\begin{bmatrix}(\mathbf{U}_{1*}^{t+1})^\top\\ \cdots\\ (\mathbf{U}_{m*}^{t+1})^\top\end{bmatrix} = \begin{bmatrix}(\mathbf{U}_{1*}^{t})^\top\\ \cdots\\ (\mathbf{U}_{m*}^{t})^\top\end{bmatrix} - \frac{\eta}{p}\begin{bmatrix}\mathcal{P}_{\Omega_{1*}}\big(\widehat{\mathbf{O}}_{1*}^\top\big)\mathbf{V}^t\\ \cdots\\ \mathcal{P}_{\Omega_{m*}}\big(\widehat{\mathbf{O}}_{m*}^\top\big)\mathbf{V}^t\end{bmatrix} - \frac{\eta}{2}\begin{bmatrix}(\mathbf{U}_{1*}^{t})^\top\mathbf{R}^t\\ \cdots\\ (\mathbf{U}_{m*}^{t})^\top\mathbf{R}^t\end{bmatrix}, \tag{4.3}$$

where $\Omega_{i*} = \{j : (i,j) \in \Omega\}$, $\mathbf{R}^t = \sum_{i=1}^m \mathbf{U}_{i*}^t(\mathbf{U}_{i*}^t)^\top - (\mathbf{V}^t)^\top\mathbf{V}^t$, $\widehat{\mathbf{O}}_{i*} = \widehat{\mathbf{Y}}_{i*}^t - \mathbf{Y}_{i*}^t$, and $\widehat{\mathbf{Y}}_{i*}^t = \mathbf{V}^t\mathbf{U}_{i*}^t$. As a result, we propose to perform the following **local** update for user $i \in [m]$ given $\mathbf{V}^t$, $\mathbf{R}^t$ and its own observation $\mathbf{Y}_{i*}$:

$$(\mathbf{U}_{i*}^{t+1})^\top = (\mathbf{U}_{i*}^{t})^\top - \frac{\eta}{p}\mathcal{P}_{\Omega_{i*}}\big((\widehat{\mathbf{Y}}_{i*}^t)^\top - \mathbf{Y}_{i*}^\top\big)\mathbf{V}^t - \frac{\eta}{2}(\mathbf{U}_{i*}^t)^\top\mathbf{R}^t. \tag{4.4}$$

Furthermore, we can rewrite the update rule for $\mathbf{V}$ in (4.2) as follows:

$$\mathbf{V}^{t+1} = \mathbf{V}^t - \frac{\eta}{p}\begin{bmatrix}\mathcal{P}_{\Omega_{1*}}\big(\widehat{\mathbf{O}}_{1*}^\top\big)\\ \cdots\\ \mathcal{P}_{\Omega_{m*}}\big(\widehat{\mathbf{O}}_{m*}^\top\big)\end{bmatrix}^\top\begin{bmatrix}(\mathbf{U}_{1*}^{t})^\top\\ \cdots\\ (\mathbf{U}_{m*}^{t})^\top\end{bmatrix} + \frac{\eta}{2}\mathbf{V}^t\mathbf{R}^t. \tag{4.5}$$

Therefore, we propose to perform the **global** update in (4.5) on the server to obtain $\mathbf{V}^{t+1}$ and $\mathbf{R}^{t+1}$ after receiving $\{\mathcal{P}_{\Omega_{i*}}\big((\widehat{\mathbf{Y}}_{m*}^t)^\top - \mathbf{Y}_{m*}^\top\big)\}_{i\in[m]}$ and $\{\mathbf{U}_{i*}^t\}_{i\in[m]}$ from all users.

Privacy leakage. We can see from (4.4) that the local update of each user depends on the sensitive information of other users only through the parameters $\mathbf{V}^t$ and $\mathbf{R}^t$. Therefore, we can achieve the joint RDP of the aforementioned local and global updates if the server can make $\mathbf{V}^t$ and $\mathbf{R}^t$ private during the global update in (4.5), provided that each user does not make its own recommendation and the local model public. We compute $\mathbf{V}^t$ and $\mathbf{R}^t$ using information from all users, which makes them noise tolerant. Therefore, we can add sufficient random noise to make them private without downgrading the accuracy of the recommendations.

Algorithm 1 Differentially Private Low-rank Matrix Completion via Matrix Factorization (DPLMC)

input Initialization $\mathbf{U}^0, \widetilde{\mathbf{V}}^0$, iteration number T, step size η, projection parameters α_1, α_2, G

1: **for** $t = 0, 1, \ldots, T-1$ **do**
2: **if** $t = 0$ **then**
3: **Communicate (receive)** $\big(\mathcal{P}_{\Omega_{i*}}(\widehat{\mathbf{Y}}^0_{i*} - \mathbf{Y}_{i*})\big)$ to the server, where $\widehat{\mathbf{Y}}^0_{i*} = \widetilde{\mathbf{V}}^0\mathbf{U}^0_{i*}$, $i \in [m]$
4: **end if**
5: **on server do**
6: Compute $\mathbf{U}^t = \big[(\mathbf{U}^t_{1*})^\top; \ldots; (\mathbf{U}^t_{m*})^\top\big]$ and $\mathcal{P}_\Omega\big(\widehat{\mathbf{Y}}^t - \mathbf{Y}\big) = \big[\mathcal{P}_{\Omega_{1*}}\big((\widehat{\mathbf{Y}}^t_{1*})^\top - \mathbf{Y}^\top_{1*}\big); \ldots; \mathcal{P}_{\Omega_{m*}}\big((\widehat{\mathbf{Y}}^t_{m*})^\top - \mathbf{Y}^\top_{m*}\big)\big]$
7: Projection: $\mathcal{P}_\Omega\big(\widehat{\mathbf{Y}}^t - \mathbf{Y}\big) = \mathcal{P}_{\mathcal{C}_\mathbf{Y}}\big(\mathcal{P}_\Omega\big(\widehat{\mathbf{Y}}^t - \mathbf{Y}\big)\big)$, where $\mathcal{C}_\mathbf{Y} = \{\mathbf{A} \in \mathbb{R}^{m\times n} \,\big|\, \|\mathbf{A}\|_{2,\infty} \le G\}$
8: Obtain private $\mathbf{R}$: $\widetilde{\mathbf{R}}^t = \sum_{i=1}^m \mathbf{U}^t_{i*}(\mathbf{U}^t_{i*})^\top - (\widetilde{\mathbf{V}}^t)^\top\widetilde{\mathbf{V}}^t + \mathbf{N}_1$, where $\mathbf{N}_1 \in \mathbb{R}^{r\times r}$ is a symmetric matrix with elements in the upper triangle being i.i.d. $N(0, \nu_1^2)$
9: Obtain private $\mathbf{V}$: $\widetilde{\mathbf{V}}^{t+1} = \widetilde{\mathbf{V}}^t - (\eta/p)\big(\mathcal{P}_\Omega\big(\widehat{\mathbf{Y}}^t - \mathbf{Y}\big)^\top\mathbf{U}^t + \mathbf{N}_2\big) + (\eta/2)\widetilde{\mathbf{V}}^t\widetilde{\mathbf{R}}^t$, where $\mathbf{N}_2 \in \mathbb{R}^{n\times r}$ is a matrix with elements being i.i.d. $N(0, \nu_2^2)$
10: Projection: $\widetilde{\mathbf{V}}^{t+1} = \mathcal{P}_{\mathcal{C}_\mathbf{V}}(\widetilde{\mathbf{V}}^{t+1})$, where $\mathcal{C}_\mathbf{V} = \{\mathbf{A} \in \mathbb{R}^{n\times r} \,\big|\, \|\mathbf{A}\|_{2,\infty} \le \alpha_2\}$
11: **end on server**
12: **Communicate (send)** $(\widetilde{\mathbf{V}}^{t+1}, \widetilde{\mathbf{R}}^t)$ to all users $i \in [m]$
13: **on user** $i \in [m]$ **do**
14: Update: $(\mathbf{U}^{t+1}_{i*})^\top = (\mathbf{U}^t_{i*})^\top - \frac{\eta}{p}\mathcal{P}_{\Omega_{i*}}\big((\widehat{\mathbf{Y}}^t_{i*})^\top - \mathbf{Y}^\top_{i*}\big)\widetilde{\mathbf{V}}^t - \frac{\eta}{2}(\mathbf{U}^t_{i*})^\top\widetilde{\mathbf{R}}^t$
15: Projection: $\mathbf{U}^{t+1}_{i*} = \mathcal{P}_{\mathcal{C}_\mathbf{U}}(\mathbf{U}^{t+1}_{i*})$, $\mathcal{C}_\mathbf{U} = \{\mathbf{A} \in \mathbb{R}^{m\times r} \,\big|\, \|\mathbf{A}\|_{2,\infty} \le \alpha_1\}$
16: Obtain own recommendations: $\widehat{\mathbf{Y}}^{t+1}_{i*} = \widetilde{\mathbf{V}}^{t+1}\mathbf{U}^{t+1}_{i*}$
17: **Communicate (receive)** $\big(\mathcal{P}_{\Omega_{i*}}(\widehat{\mathbf{Y}}^{t+1}_{i*} - \mathbf{Y}_{i*}), \mathbf{U}^{t+1}_{i*}\big)$ to the server
18: **end on user**
19: **end for**

On the other hand, we assume a trusted server in our setting, as is common in the literature (Jain et al., 2018; Chien et al., 2021). Therefore, there is no need to add random noise to make $\{\mathcal{P}_{\Omega_{i*}}\big((\widehat{\mathbf{Y}}^t_{m*})^\top - \mathbf{Y}^\top_{m*}\big)\}_{i\in[m]}$ and $\{\mathbf{U}^t_{i*}\}_{i\in[m]}$ private. This will again improve the accuracy of our recommendations.

4.1 Differentially Private Algorithm

Algorithm 1 details the proposed Differentially Private Low-rank Matrix Factorization (DPLMC) method for solving the noisy matrix completion problem with joint RDP based on the server and user computation framework. At the t-th iteration, the DPLMC method requires one round of communication, i.e., one back and forth communication between the server and m users. The server first obtains $\mathbf{R}^t$ using $\{\mathbf{U}^t_{i*}\}_{i=1}^m$ received from all users and then updates $\mathbf{V}^{t+1}$ according to (4.5). To make $\mathbf{R}^t, \mathbf{V}^{t+1}$ private, we use the Gaussian mechanism (lines 8 and 9 in Algorithm 1). Finally, the server broadcasts the private parameters $\widetilde{\mathbf{V}}^{t+1}$ and $\widetilde{\mathbf{R}}^t$ to all users. After receiving $\widetilde{\mathbf{V}}^{t+1}$ and $\widetilde{\mathbf{R}}^t$, each user $i \in [m]$ updates its own embedding vector according to the local update rule in (4.4) and then generates its own recommendations (lines 14 and 16 in Algorithm 1).

Algorithm 1 uses the projection step (lines 7, 10, 15 in Algorithm 1) for two reasons: (1) the produced estimator satisfies the incoherence constraint (projections in lines 10 and 15), which is crucial for successful recovery and a linear convergence rate; (2) it helps us determine the magnitude of the random noise needed to achieve the joint RDP (projections in lines 7 and 15). Note that these projections can be easily and efficiently implemented using the clipping technique (Abadi et al., 2016).

Compared to DPALS, our method is more efficient since it does do not require inverting a matrix and projecting onto the cone of positive semidefinite matrices.

4.2 Initialization

In practice, random initialization is often sufficient for Algorithm 1 to produce good estimators. This has been observed previously in the non-private setting (Bhojanapalli et al., 2016; Chen et al., 2019). However, to guarantee a linear convergence rate and achieve strong utility guarantees, a good initialization is necessary for Algorithm 1. Algorithm 2 provides such an initial estimator and is motivated by the private singular value decomposition (Dwork et al., 2014; Jain et al., 2018). This algorithm also follows the server and user computation model. The server first computes the covariance matrix $\mathbf{A} = \tau^2\sum_{i=1}^m \mathbf{Y}_{i*}\mathbf{Y}^\top_{i*}/p^2$ using observations from all users. It then obtains a private top-r singular vector $\mathbf{V}^0$ and singular value matrix $\mathbf{\Sigma}^{1/2}$ of $\mathbf{A}$ (see line 4 in Algorithm 2). Differential privacy is achieved in this step using the Gaussian mechanism (Dwork et al., 2014). The

server then broadcasts the private matrices $\mathbf{V}^0$ and $\mathbf{\Sigma}^{1/2}$ to the users. Finally, each user $i \in [m]$ constructs its own initial estimator $\mathbf{U}^0_{i*}$ (see line 10 in Algorithm 2) based on its own observations and the matrices $\mathbf{V}^0$ and $\mathbf{\Sigma}^{1/2}$. Since the initialization of each user depends on other users only through the matrices $\mathbf{V}^0, \mathbf{\Sigma}^{1/2}$ that were constructed on the server in a differentially private way, Algorithm 2 satisfies the joint DP. Similar to Algorithm 1, we use the projection step (see line 1 in Algorithm 2) to determine the magnitude of random noise needed to achieve joint DP. Note that we can also use Algorithm 2 as an initialization scheme for DPALS (Chien et al., 2021) to satisfy its specific initialization requirement in their utility guarantees.

Algorithm 2 Differentially Private Initialization for DPLMC

input Parameters G, τ

1: Projection: $\mathbf{Y} = \mathcal{P}_{\mathcal{C}_\mathbf{Y}}(\mathbf{Y})$, where $\mathcal{C}_\mathbf{Y} = \{\mathbf{A} \in \mathbb{R}^{m\times n} \mid \|\mathbf{A}\|_{2,\infty} \le G\}$
2: **Communicate (receive)** $\mathbf{Y}_{i*}$ to the server for $i \in [m]$
3: **on server do**
4: Obtain private covariance matrix: $\mathbf{A} = \tau^2 \sum_{i=1}^m \mathbf{Y}_{i*}\mathbf{Y}_{i*}^\top/p^2 + \mathbf{N}_0$, where $\mathbf{N}_0 \in \mathbb{R}^{n\times n}$ is a symmetric matrix with elements in the upper triangle following i.i.d. $N(0, \nu_0^2)$
5: Obtain $(\mathbf{V}^0, \mathbf{\Sigma})$ by performing the rank r singular value decomposition of $\mathbf{A}$
6: Let $\widetilde{\mathbf{V}}^0 = \mathbf{V}^0\mathbf{\Sigma}^{1/2}$
7: **end on server**
8: **Communicate (sends)** $(\tau\mathbf{V}^0\mathbf{\Sigma}^{-1/2}/p)$ to all users $i \in [m]$
9: **on user** $i \in [m]$ **do**
10: Obtain $\mathbf{U}^{0\top}_{i*} = \tau\mathbf{Y}^\top_{i*}\mathbf{V}^0\mathbf{\Sigma}^{-1/2}/p$
11: **end on user**

5 MAIN RESULTS

We establish privacy and utility guarantees for the DPLMC method as well as the estimation error of our initialization method.

5.1 Privacy Guarantee

DPLMC can achieve the following privacy guarantee.

Theorem 5.1. Algorithm 1 satisfies $(\gamma, \rho_1 + \rho_2)$-joint RDP with $\rho_1 = 2T\gamma\alpha_1^4/\nu_1^2$ and $\rho_2 = 4T\gamma G^2\alpha_1^2/\nu_2^2$. Furthermore, for any $\epsilon > 0$ and $\delta \in (0, 1)$, it satisfies (ϵ, δ)-joint DP if $\nu_1 = \alpha_1^2\sqrt{8T(\log(1/\delta) + \epsilon)}/(\sqrt{\omega}\epsilon)$ and $\nu_2 = 4\alpha_1 G\sqrt{T(\log(1/\delta) + \epsilon)}/(\sqrt{1-\omega}\epsilon)$ with $\omega \in (0, 1)$.

The variance of the random noise in Algorithm 1 is determined by the parameters α_1 and G, as shown in Theorem 5.1. In practice, the clipping technique can be used to compute projections (Abadi et al., 2016). Additionally, the weight parameter ω in ν_1 and ν_2 determines how the privacy budget ϵ is allocated to the private mechanisms in lines 7 and 9 of Algorithm 1.

The above result does not take into account the potential privacy cost of the initialization. However, in practice, we can always use random initialization without incurring any privacy cost to obtain a reasonable estimate. Alternatively, we can use the private initialization provided in Algorithm 2 and allocate the privacy budget to both the initialization and the main algorithm.

In DPALS (Chien et al., 2021), the resampling (sample splitting) step is used to limit the number of items observed per user to achieve joint DP. However, instead of discarding potentially valuable observations, we propose using a projection-based method to achieve the private guarantees, which is related to the idea of gradient clipping used in differentially private stochastic gradient descent (Abadi et al., 2016).

5.2 Utility Guarantee

We establish the following utility guarantee and linear convergence for the DPLMC method.

Theorem 5.2. Consider the noisy matrix completion problem under the uniform sampling model in (3.1), where the matrix $\mathbf{X}^*$ has rank r and satisfies the incoherence condition (3.2) and the noise E_{ij} is i.i.d. Normal zero mean with variance $\nu^2/(mn)$. There exist constants $\{c_i\}_{i=1}^7$ such that for any $\delta \in (0, 1)$ and privacy budget ϵ, if $\eta = c_1/\sigma_1$, $\nu_1 = \alpha_1^2\sqrt{8T(\log(1/\delta) + \epsilon)}/(\sqrt{\omega}\epsilon)$, $\nu_2 = 4G\alpha_1\sqrt{T(\log(1/\delta) + \epsilon)}/(\sqrt{1-\omega}\epsilon)$, $G = 2\alpha_1\alpha_2 K$, where K is the largest number of observations per user, $\omega = 1/(1 + K/p^2)$, $\alpha_1 = \sqrt{\beta r\sigma_1/m}$, $\alpha_2 = \sqrt{\beta r\sigma_1/n}$, $T = c_2 \log\big(m\epsilon/(\beta\sigma_1)\big)$, $m \ge n$, the number of observations $|\Omega| \ge c_3 r^2 m \log m$, and the initialization $\mathbf{X}^0 = \mathbf{U}^0\widetilde{\mathbf{V}}^{0\top}$ satisfies $\|\mathbf{X}^0 - \mathbf{X}^*\|_F \le c_4\sigma_r$, then with probability at least $1 - c_5/m$ we have

$$\|\mathbf{X}^T - \mathbf{X}^*\|_F^2 \le c_6\frac{(r\sigma_1^2\nu^2 + r^2\beta^2\sigma_1^2\sigma_r)m\log m}{\sigma_r|\Omega|} + c_7\frac{\beta^3\sigma_1^5 r^4 n^2\log(1/\delta)\log^3 m}{\sigma_r|\Omega|m\epsilon^2},$$

where $\mathbf{X}^T = \mathbf{U}^T(\widetilde{\mathbf{V}}^T)^\top$.

The utility guarantee for the DPLMC method consists of two terms. The first term $O\big(r^2 m\log m/|\Omega|\big)$ corresponds to the statistical error for noisy matrix completion and is near optimal. The minimax lower bound is $O(rm\log m/|\Omega|)$ (Negahban and Wainwright, 2012; Koltchinskii et al., 2011). The second term $\widetilde{O}\big(r^4 n^2/(m|\Omega|\epsilon^2)\big)$ is the dominant term and denotes the error introduced by the private mechanism (lines 8 and 9 in Algorithm 1). Furthermore, the sample complexity of the DPLMC method is $O(r^2 m\log m)$,

which matches the best known sample complexity of matrix completion using matrix factorization in the non-private setting (Zheng and Lafferty, 2016; Wang et al., 2017). Compared to the best known sample complexity of $O(r^6 m \log^3 m)$ obtained for matrix completion under the joint DP (Chien et al., 2021), the DPLMC method reduces the sample complexity requirement by a factor of $O(r^2 \log^2 m)$. Furthermore, DPLMC achieves the utility guarantee of $\widetilde{O}(r^4 n^2/(|\Omega| m\epsilon^2))$, which is better by a factor of $\widetilde{O}(r)$ than the utility guarantee $\widetilde{O}(r^5 n^2/(|\Omega| m\epsilon^2))$, provided in Chien et al. (2021). Note that we aim to provide a user-level privacy guarantee, and thus it is reasonable to require more users to obtain good utility guarantees (as the sensitivity is determined by the number of users).

5.3 Initialization

The utility guarantee provided in Theorem 5.1 requires that the initialization satisfies $\|\mathbf{U}^0 \widetilde{\mathbf{V}}^{0\top} - \mathbf{X}^*\|_F \leq c_4 \sigma_r$. The following theorem shows that Algorithm 2 can output private $\mathbf{U}^0, \widetilde{\mathbf{V}}^0$ that satisfy this condition.

Theorem 5.3. For any $\delta \in (0,1)$ and privacy budget ϵ, Algorithm 2 satisfies the (ϵ, δ)-joint DP if $\nu_0 = 4G^2\tau^2\sqrt{\log(1/\delta)}/(p^2\epsilon)$. Furthermore, under the conditions of Theorem 5.2, there exist absolute constants $\{c_i\}_{i=1}^4$, such that if $G^2 = K\alpha_1^2\alpha_2^2$, where K is the largest number of observations per user, $\tau \geq c_1(1 - \sigma_r^2/4\|\mathbf{X}^*\|_F^2)$, and the number of observations $|\Omega| \geq c_2 \max\{r^2 m \log m, r^{5/2} n^{3/2} \log m\}$, then with probability at least $1 - c_3/m$, we have $\|\mathbf{U}_0 \widetilde{\mathbf{V}}_0^\top - \mathbf{X}^*\|_F \leq c_4 \sigma_r$.

When the number of observations $|\Omega| \geq \widetilde{O}(r^{5/2} n^{3/2})$, Algorithm 2 outputs an initial estimate that satisfies the conditions of Theorem 5.1. Furthermore, if $m \geq \widetilde{O}(r^{1/2} n^{3/2})$, Algorithm 2 only requires $|\Omega| \geq \widetilde{O}(r^2 m)$, which matches the sample complexity requirement in Theorem 5.2. Note that Algorithm 2 can reduce to the initialization procedure, i.e., Analyze Gauss (Dwork et al., 2014), used in DPALS if we choose $\tau = 1$ and output $\mathbf{V}^0$ (line 5 in Algorithm 2) as its initialization estimate. However, to achieve its specific initialization requirement, DPALS needs $|\Omega| \geq \max\{\widetilde{O}(r^6 m), \widetilde{O}(r^4 n^{3/2})\}$, which is worse than our sample complexity requirement. We also notice that Algorithm 2 needs the additional requirement on the number of users, i.e., $m \geq \widetilde{O}(r^{1/2} n^{3/2})$, compared with the requirement in our utility guarantee. Therefore, it would be an interesting future research direction to develop a new initialization method to improve or even remove the requirement on m without stringent assumptions on the observed data.

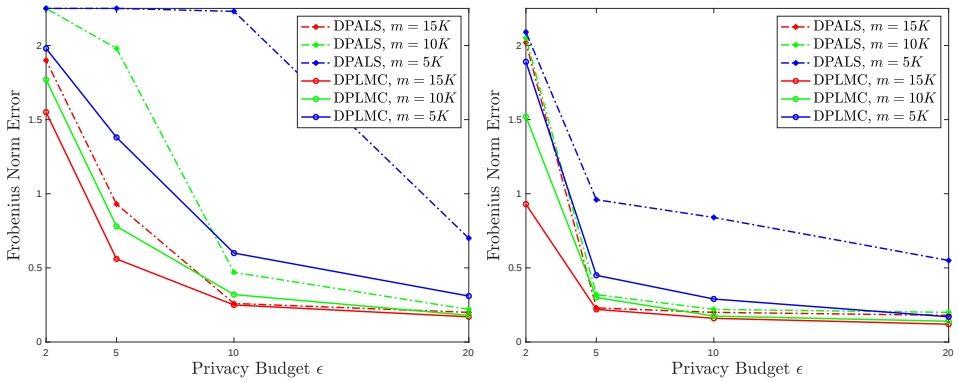
6 NUMERICAL EXPERIMENTS

In this section, we conduct experiments on synthetic data to evaluate the performance of our proposed method, DPLMC. We compare DPLMC with the state-of-the-art approach DPALS (Chien et al., 2021), since it has been shown to outperform other existing baselines such as DPFW (Jain et al., 2018) and Private SVD (McSherry and Mironov, 2009). For both DPALS and DPLMC, we evaluate their performances with both random initialization and the initialization procedure in Algorithm 2.

Synthetic data. We generate the underlying low-rank matrix $\mathbf{X}^* \in \mathbb{R}^{m\times n}$ as $\mathbf{X}^* = \mathbf{U}^*\mathbf{V}^{*\top}$, where each element of $\mathbf{U}^* \in \mathbb{R}^{m\times r}$ and $\mathbf{V}^* \in \mathbb{R}^{n\times r}$ follows i.i.d. standard Gaussian distribution. In this problem, we fix $n = 100$, $r = 5$, and choose m from the set $\{5000, 10000, 15000\}$. In addition, we scale $\mathbf{U}^*$ and $\mathbf{V}^*$ so that $\max\{\|\mathbf{U}^*\|_{2,\infty}, \|\mathbf{V}^*\|_{2,\infty}\} \leq 2$. The observed data matrix is generated according to the uniform observation model (3.1), where E_{ij} is set to 0 in the noiseless case and E_{ij} follows an i.i.d. centered Gaussian distribution with variance $\nu^2 = 1.0$ in the noisy case. We split the fully observed matrix into two datasets: one for training and evaluation, and the other for validation. The validation dataset is used to tune parameters for different algorithms. For the other dataset, we use a fraction of observations (the number of observations is set to $|\Omega| = rm \log m$) to run different algorithms to estimate the underlying matrix, and then we compute the estimation error using the unobserved data in this dataset.

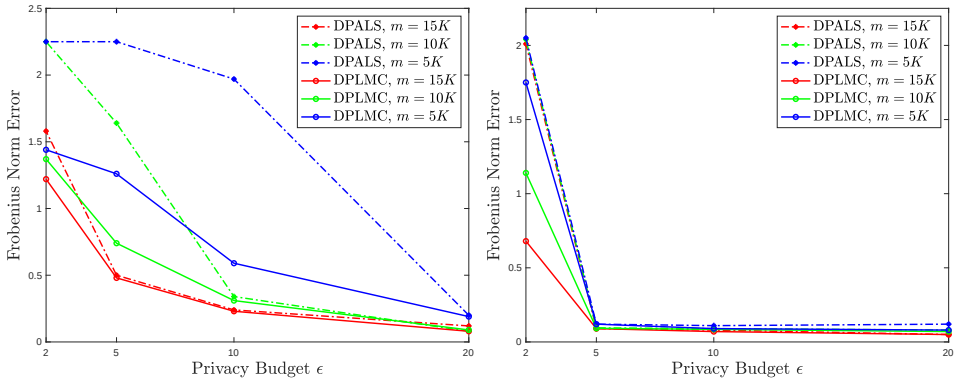
Parameters and privacy tracking. For both DPALS and DPLMC, we choose hyper-parameters on the validation set according to the requirements of their theoretical guarantees. Specifically, for DPALS, we choose the number of iterations T from $\{1, \ldots, 5\}$, set the row clipping and entry clipping parameters to 2 and 4, respectively, and choose the maximum number of ratings per user from the sequence of values $\{30, 31, \ldots, 60\}$. For DPLMC, we choose the iteration number T from the sequence of values $\{10, 15, \ldots, 60\}$ and the step size is chosen from the grid $\{0.05, 0.1, \ldots, 0.3\}$. We choose the projection parameters α_1 and α_2 from $\{2, 3, \ldots, 6\}$ and set $G = \alpha_1\alpha_2$. We set the privacy parameter $\delta = 10^{-5}$ and choose the privacy budget ϵ from $\{2, 5, 10, 20\}$ following previous work (Chien et al., 2021). For both methods, we keep track of the privacy guarantee using RDP and then translate the joint RDP to the standard (ϵ, δ)-joint DP. When using Algorithm 2 to generate the initial estimators, we allocate the privacy budget $\epsilon_{\text{init}} = 1$ for this step when we have a total privacy budget $\epsilon \in \{5, 10, 20\}$. For a total privacy budget $\epsilon = 2$, we allocate the privacy budget $\epsilon_{\text{init}} = 0.2$ for the private initialization step.

Evaluation. We evaluate the performance of different methods using the squared averaged Frobenius norm error $\|\widehat{\mathbf{X}} - \mathbf{X}^*\|_F^2/(mn)$, where $\widehat{\mathbf{X}}$ is the output of the algorithm. In addition, to evaluate the improvements of our method, we generate synthetic data following the same procedure as



(a) Random Initialization ($\nu = 1$) (b) Private Initialization ($\nu = 1$)

Figure 1: Numerical results for DPALS and DPLMC on synthetic data with different numbers of users m in the noisy case are presented. We evaluate the Frobenius norm error versus the privacy budget using both random initialization and private initialization. For private initialization, we use Algorithm 2 with a privacy budget of 1 for $\epsilon \in \{5, 10, 20\}$ and a privacy budget of 0.2 for $\epsilon = 2$. The results are shown in Figure (a) for the Frobenius norm error versus the privacy budget using random initialization and in Figure (b) for the Frobenius norm error versus the privacy budget using private initialization.



(a) Random Initialization ($\nu = 0$) (b) Private Initialization ($\nu = 0$)

Figure 2: Numerical results for DPALS and DPLMC on synthetic data with different numbers of users m in the noisy case are presented. We evaluate the Frobenius norm error versus the privacy budget using both random initialization and private initialization. For private initialization, we use Algorithm 2 with a privacy budget of 1 for $\epsilon \in \{5, 10, 20\}$ and a privacy budget of 0.2 for $\epsilon = 2$. The results are shown in Figure (a) for the Frobenius norm error versus the privacy budget using random initialization and in Figure (b) for the Frobenius norm error versus the privacy budget using private initialization.

before, but varying $r \in \{3, 5, 7\}$. All results are averaged over 10 trials.

Results. We can observe from Figures 1 and 2 that the Frobenius norm error of DPLMC decreases as the number of users increases. This is consistent with the utility guarantees established in Theorem 5.2. More importantly, DPLMC consistently outperforms DPALS in both noisy and noiseless settings across different numbers of users and privacy parameters. Specifically, in the noisy setting, DPLMC significantly outperforms DPALS when we have small privacy budgets and fewer users. In the noiseless setting, both methods can produce accurate estimators when we have large privacy budgets with private initialization. These results demonstrate the superiority of our proposed method. Furthermore, the results show that the proposed initialization algorithm is very helpful for both algorithms to find good estimators, especially when we have small privacy budgets. Figure 3 shows that the utility gap between DPLMC and DPALS increases with increasing rank r, which is consistent with our main results.

7 CONCLUSION AND FUTURE WORK

This paper addresses the problem of privacy-preserving matrix completion, and presents a new algorithm, DPLMC, which is based on low-rank matrix factorization, projected gradient descent, and a server and user computation model. DPLMC satisfies joint RDP, achieves a linear convergence rate, and improves utility guarantees of existing methods.

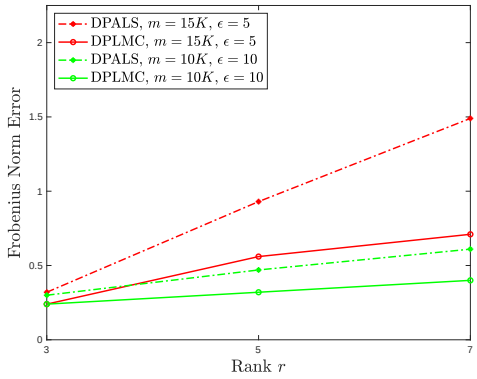


Figure 3: Frobenius norm error for DPALS and DPLMC on synthetic data as the rank of $\boldsymbol{X}^*$ varies. We consider two settings with noisy observations: 1) $m = 15000$, $\epsilon = 5$, and 2) $m = 10000$, $\epsilon = 10$.

Empirical results demonstrate the superiority of our method over the state-of-the-art approach.

Our results rely on a reasonable initial estimate, which can be obtained using the proposed initialization procedure. However, this method makes strong assumptions about the number of users. Therefore, it would be interesting to develop a better initialization approach without any additional assumptions.

While our paper focuses on the uniform observation model, it is important to note that real-world datasets may not adhere to this model. As a result, exploring how different observation models could impact the trade-offs between privacy and utility, both theoretically and practically, would be an interesting avenue for future research.

In this paper, we adopt joint (Rényi) differential privacy, which is suitable for personalized recommendations. However, in many real-world scenarios, the participation of a user may not be sensitive information, while each data record owned by the user may be very sensitive, and we may want to ensure that this data record is private to other users. In such cases, it may be worth considering other privacy notions, such as federated DP (Zheng et al., 2021), and developing corresponding private methods to explore their privacy and utility trade-offs.

Acknowledgments

We thank anonymous reviewers for their helpful comments.

References

M. Abadi, A. Chu, I. J. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang. Deep learning with differential privacy. In *ACM SIGSAC Conference on Computer and Communications Security*, 2016.

S. Bhojanapalli, B. Neyshabur, and N. Srebro. Global optimality of local search for low rank matrix recovery. In *Advances in Neural Information Processing Systems*, 2016.

M. Bun and T. Steinke. Concentrated differential privacy: Simplifications, extensions, and lower bounds. In *Theory of Cryptography Conference*, 2016.

D. Cai, X. He, X. Wu, and J. Han. Non-negative matrix factorization on manifold. In *IEEE International Conference on Data Mining*, 2008.

J. A. Calandrino, A. Kilzer, A. Narayanan, E. W. Felten, and V. Shmatikov. "You might also like": Privacy risks of collaborative filtering. In *IEEE Symposium on Security and Privacy*, 2011.

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

E. J. Candès and T. Tao. The power of convex relaxation: Near-optimal matrix completion. *IEEE Transactions on Information Theory*, 56(5):2053–2080, 2010.

Y. Chen and M. J. Wainwright. Fast low-rank estimation by projected gradient descent: General statistical and algorithmic guarantees. *arXiv preprint arXiv:1509.03025*, 2015.

Y. Chen, Y. Chi, J. Fan, and C. Ma. Gradient descent with random initialization: Fast global convergence for nonconvex phase retrieval. *Mathematical Programming*, 176(1):5–37, 2019.

Y. Chi, Y. M. Lu, and Y. Chen. Nonconvex optimization meets low-rank matrix factorization: An overview. *IEEE Transactions on Signal Processing*, 67(20):5239–5269, 2019.

S. Chien, P. Jain, W. Krichene, S. Rendle, S. Song, A. Thakurta, and L. Zhang. Private alternating least squares: Practical private matrix completion with tighter rates. In *International Conference on Machine Learning*, 2021.

C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography Conference*, 2006.

C. Dwork, K. Talwar, A. Thakurta, and L. Zhang. Analyze gauss: optimal bounds for privacy-preserving principal component analysis. In *ACM Symposium on Theory of Computing*, 2014.

V. Feldman, T. Koren, and K. Talwar. Private stochastic convex optimization: optimal rates in linear time. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, 2020.

D. Gross. Recovering low-rank matrices from few coefficients in any basis. *IEEE Transactions on Information Theory*, 57(3):1548–1566, 2011.

M. Hardt and A. Roth. Beating randomized response on incoherent matrices. In *ACM Symposium on Theory of Computing*, 2012.

M. Hardt and A. Roth. Beyond worst-case analysis in private singular vector computation. In *ACM Symposium on Theory of Computing*, 2013.

P. Jain, O. D. Thakkar, and A. Thakurta. Differentially private matrix completion revisited. In *International Conference on Machine Learning*, 2018.

C. Jin, S. M. Kakade, and P. Netrapalli. Provable efficient online matrix completion via non-convex stochastic gradient descent. In *Advances in Neural Information Processing Systems*, 2016.

M. Kapralov and K. Talwar. On differentially private low rank approximation. In *ACM-SIAM Symposium on Discrete Algorithms*, 2013.

M. Kearns, M. Pai, A. Roth, and J. Ullman. Mechanism design in large games: Incentives and privacy. In *Conference on Innovations in Theoretical Computer Science*, 2014.

V. Koltchinskii, K. Lounici, A. B. Tsybakov, et al. Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *The Annals of Statistics*, 39(5):2302–2329, 2011.

Z. Liu, Y.-X. Wang, and A. Smola. Fast differentially private matrix factorization. In *ACM Conference on Recommender Systems*, 2015.

F. McSherry and I. Mironov. Differentially private recommender systems: Building privacy into the netflix prize contenders. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2009.

I. Mironov. Rényi differential privacy. In *IEEE Computer Security Foundations Symposium*, 2017.

A. Narayanan and V. Shmatikov. Myths and fallacies of" personally identifiable information". *Communications of the ACM*, 53(6):24–26, 2010.

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

D. Park, A. Kyrillidis, C. Caramanis, and S. Sanghavi. Finding low-rank solutions via nonconvex matrix factorization, efficiently and provably. *SIAM Journal on Imaging Sciences*, 11(4):2165–2204, 2018.

J. D. M. Rennie and N. Srebro. Fast maximum margin matrix factorization for collaborative prediction. In *International Conference on Machine Learning*, 2005.

A. Rohde, A. B. Tsybakov, et al. Estimation of high-dimensional low-rank matrices. *The Annals of Statistics*, 39(2):887–930, 2011.

S. Tu, R. Boczar, M. Simchowitz, M. Soltanolkotabi, and B. Recht. Low-rank solutions of linear matrix equations via procrustes flow. In *International Conference on Machine Learning*, 2016.

R. Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge University Press, 2018.

L. Wang, X. Zhang, and Q. Gu. A unified variance reduction-based framework for nonconvex low-rank matrix recovery. In *International Conference on Machine Learning*, 2017.

Z. Wang, S. Zheng, Y. Ye, and S. Boyd. Further relaxations of the semidefinite programming approach to sensor network localization. *SIAM Journal on Optimization*, 19(2): 655–673, 2008.

Q. Zheng and J. Lafferty. Convergence analysis for rectangular matrix completion using burer-monteiro factorization and gradient descent. *arXiv preprint arXiv:1605.07051*, 2016.

Q. Zheng, S. Chen, Q. Long, and W. J. Su. Federated f-differential privacy. In *International Conference on Artificial Intelligence and Statistics*, 2021.

Y. Zhou, S. Wu, and A. Banerjee. Bypassing the ambient dimension: Private {sgd} with gradient subspace identification. In *International Conference on Learning Representations*, 2021.

A Proof of Theorem 5.1

We provide the privacy guarantee for Algorithm 1.To ensure Algorithm 1 satisfies joint RDP, it suffices to show that the computation of $\widetilde{\mathbf{R}}^t$ and $\widetilde{\mathbf{V}}^{t+1}$ satisfy RDP for $t = 0, \ldots, T-1$. Then we can use the composition result to obtain the overall RDP.

To provide the privacy guarantee, we need the following lemmas, which have been established in Mironov (2017).

Lemma A.1. Given a function $q : \mathcal{S}^n \to \mathcal{R}$, the Gaussian Mechanism $\mathcal{M} = q(S) + \mathbf{u}$, where $\mathbf{u} \sim N(0, \nu^2\mathbf{I})$, satisfies $(\gamma, \gamma\Delta^2(q)/(2\nu^2))$-RDP.

Lemma A.2. If k randomized mechanisms $\mathcal{M}_i : \mathcal{S}^n \to \mathcal{R}$ for $i \in [k]$, satisfy (γ, ρ_i)-RDP, then their composition $\big(\mathcal{M}_1(S), \ldots, \mathcal{M}_k(S)\big)$ satisfies $(\gamma, \sum_{i=1}^k \rho_i)$-RDP. Moreover, the input of the i-th mechanism can be based on the outputs of previous $(i-1)$ mechanisms.

We can use the following lemma, proposed in Mironov (2017), to translate the RDP guarantee to the (ϵ, δ)-DP.

Lemma A.3. If a randomized mechanism $\mathcal{M} : \mathcal{S}^n \to \mathcal{R}$ satisfies (γ, ρ)-RDP, then $\mathcal{M}$ satisfies $(\rho + \log(1/\delta)/(\gamma-1), \delta)$-DP for all $\delta \in (0, 1)$.

Given these lemmas, we are ready to provide the privacy guarantee of our proposed method.

We first show that in each iteration $t = 0, \ldots, T-1$, $\widetilde{\mathbf{R}}^t$ and $\widetilde{\mathbf{V}}^{t+1}$ satisfy RDP. Recall that we have

$$\mathbf{R}^t = \sum_{i=1}^m \mathbf{U}_{i*}^t (\mathbf{U}_{i*}^t)^\top - (\widetilde{\mathbf{V}}^t)^\top \widetilde{\mathbf{V}}^t.$$

Therefore, we have the following sensitivity

$$\begin{aligned}
\Delta(\mathbf{R}^t) &= \big\| \mathbf{U}_{j*}^t (\mathbf{U}_{j*}^t)^\top - \mathbf{U}_{j'*}^t (\mathbf{U}_{j'*}^t)^\top \big\|_F \\
&\le \big\| \mathbf{U}_{j*}^t (\mathbf{U}_{j*}^t)^\top \|_F + \big\| \mathbf{U}_{j'*}^t (\mathbf{U}_{j'*}^t)^\top \big\|_F \\
&\le 2\alpha_1^2,
\end{aligned}$$

where the last inequality is due to the projection step, i.e., line 15 in Algorithm 1. Therefore, according to Lemma A.1, we need to add a random Gaussian matrix (symmetric) $\mathbf{N}_1 \in \mathbb{R}^{r \times r}$ with each element (in the upper triangle) following i.i.d. $N(0, \nu_1^2)$. As a result, $\widetilde{\mathbf{R}}^t$ satisfies $(\gamma, 2\gamma\alpha_1^4/\nu_1^2)$-RDP.

On the other hand, we have

$$\mathbf{V}^{t+1} = \widetilde{\mathbf{V}}^t - (\eta/p)\mathcal{P}_\Omega\big(\widehat{\mathbf{Y}}^t - \mathbf{Y}\big)^\top \mathbf{U}^t + (\eta/2)\widetilde{\mathbf{V}}^t \widetilde{\mathbf{R}}^t.$$

To make $\mathbf{V}^{t+1}$ private, we only need to consider the term $q_v = \mathcal{P}_\Omega\big(\widehat{\mathbf{Y}}^t - \mathbf{Y}\big)^\top \mathbf{U}^t$, and thus we have ($q_v'$ is computed as q_v but with the j-th user's data being different)

$$\begin{aligned}
q_v - q_v' &= \begin{bmatrix} \mathcal{P}_{\Omega_{1*}}\big((\widehat{\mathbf{Y}}_{1*}^t)^\top - \mathbf{Y}_{1*}^\top\big) \\ \cdots \\ \mathcal{P}_{\Omega_{j*}}\big((\widehat{\mathbf{Y}}_{j*}^t)^\top - \mathbf{Y}_{j*}^\top\big) \\ \cdots \\ \mathcal{P}_{\Omega_{m*}}\big((\widehat{\mathbf{Y}}_{m*}^t)^\top - \mathbf{Y}_{m*}^\top\big) \end{bmatrix}^\top \begin{bmatrix} (\mathbf{U}_{1*}^t)^\top \\ \cdots \\ (\mathbf{U}_{j*}^t)^\top \\ \cdots \\ (\mathbf{U}_{m*}^t)^\top \end{bmatrix} - \begin{bmatrix} \mathcal{P}_{\Omega_{1*}}\big((\widehat{\mathbf{Y}}_{1*}^t)^\top - \mathbf{Y}_{1*}^\top\big) \\ \cdots \\ \mathcal{P}_{\Omega_{j*}}\big((\widehat{\mathbf{Y}}_{j'*}^t)^\top - \mathbf{Y}_{j'*}^\top\big) \\ \cdots \\ \mathcal{P}_{\Omega_{m*}}\big((\widehat{\mathbf{Y}}_{m*}^t)^\top - \mathbf{Y}_{m*}^\top\big) \end{bmatrix}^\top \begin{bmatrix} (\mathbf{U}_{1*}^t)^\top \\ \cdots \\ (\mathbf{U}_{j'*}^t)^\top \\ \cdots \\ (\mathbf{U}_{m*}^t)^\top \end{bmatrix} \\
&= \begin{bmatrix} \mathcal{P}_{\Omega_{1*}}\big((\widehat{\mathbf{Y}}_{1*}^t)^\top - \mathbf{Y}_{1*}^\top\big) \\ \cdots \\ \mathcal{P}_{\Omega_{j*}}\big((\widehat{\mathbf{Y}}_{j*}^t)^\top - \mathbf{Y}_{j*}^\top\big) \\ \cdots \\ \mathcal{P}_{\Omega_{m*}}\big((\widehat{\mathbf{Y}}_{m*}^t)^\top - \mathbf{Y}_{m*}^\top\big) \end{bmatrix}^\top \begin{bmatrix} (\mathbf{U}_{1*}^t)^\top \\ \cdots \\ (\mathbf{U}_{j*}^t)^\top \\ \cdots \\ (\mathbf{U}_{m*}^t)^\top \end{bmatrix} - \begin{bmatrix} \mathcal{P}_{\Omega_{1*}}\big((\widehat{\mathbf{Y}}_{1*}^t)^\top - \mathbf{Y}_{1*}^\top\big) \\ \cdots \\ \mathcal{P}_{\Omega_{j'*}}\big((\widehat{\mathbf{Y}}_{j'*}^t)^\top - \mathbf{Y}_{j'*}^\top\big) \\ \cdots \\ \mathcal{P}_{\Omega_{m*}}\big((\widehat{\mathbf{Y}}_{m*}^t)^\top - \mathbf{Y}_{m*}^\top\big) \end{bmatrix}^\top \begin{bmatrix} (\mathbf{U}_{1*}^t)^\top \\ \cdots \\ (\mathbf{U}_{j*}^t)^\top \\ \cdots \\ (\mathbf{U}_{m*}^t)^\top \end{bmatrix} \\
&\quad + \begin{bmatrix} \mathcal{P}_{\Omega_{1*}}\big((\widehat{\mathbf{Y}}_{1*}^t)^\top - \mathbf{Y}_{1*}^\top\big) \\ \cdots \\ \mathcal{P}_{\Omega_{j'*}}\big((\widehat{\mathbf{Y}}_{j'*}^t)^\top - \mathbf{Y}_{j'*}^\top\big) \\ \cdots \\ \mathcal{P}_{\Omega_{m*}}\big((\widehat{\mathbf{Y}}_{m*}^t)^\top - \mathbf{Y}_{m*}^\top\big) \end{bmatrix}^\top \begin{bmatrix} (\mathbf{U}_{1*}^t)^\top \\ \cdots \\ (\mathbf{U}_{j*}^t)^\top \\ \cdots \\ (\mathbf{U}_{m*}^t)^\top \end{bmatrix} - \begin{bmatrix} \mathcal{P}_{\Omega_{1*}}\big((\widehat{\mathbf{Y}}_{1*}^t)^\top - \mathbf{Y}_{1*}^\top\big) \\ \cdots \\ \mathcal{P}_{\Omega_{j*}}\big((\widehat{\mathbf{Y}}_{j'*}^t)^\top - \mathbf{Y}_{j'*}^\top\big) \\ \cdots \\ \mathcal{P}_{\Omega_{m*}}\big((\widehat{\mathbf{Y}}_{m*}^t)^\top - \mathbf{Y}_{m*}^\top\big) \end{bmatrix}^\top \begin{bmatrix} (\mathbf{U}_{1*}^t)^\top \\ \cdots \\ (\mathbf{U}_{j'*}^t)^\top \\ \cdots \\ (\mathbf{U}_{m*}^t)^\top \end{bmatrix}.
\end{aligned}$$

Hence, we can get the following sensitivity

$$
\begin{aligned}
\Delta(q_v) \leq & \left\| \left(\begin{bmatrix} \mathcal{P}_{\Omega_{1*}}\big((\widehat{\mathbf{Y}}^t_{1*})^\top - \mathbf{Y}^\top_{1*}\big) \\ \cdots \\ \mathcal{P}_{\Omega_{j*}}\big((\widehat{\mathbf{Y}}^t_{j*})^\top - \mathbf{Y}^\top_{j*}\big) \\ \cdots \\ \mathcal{P}_{\Omega_{m*}}\big((\widehat{\mathbf{Y}}^t_{m*})^\top - \mathbf{Y}^\top_{m*}\big) \end{bmatrix}^\top - \begin{bmatrix} \mathcal{P}_{\Omega_{1*}}\big((\widehat{\mathbf{Y}}^t_{1*})^\top - \mathbf{Y}^\top_{1*}\big) \\ \cdots \\ \mathcal{P}_{\Omega_{j'*}}\big((\widehat{\mathbf{Y}}^t_{j'*})^\top - \mathbf{Y}^\top_{j'*}\big) \\ \cdots \\ \mathcal{P}_{\Omega_{m*}}\big((\widehat{\mathbf{Y}}^t_{m*})^\top - \mathbf{Y}^\top_{m*}\big) \end{bmatrix}^\top \right) \begin{bmatrix} (\mathbf{U}^t_{1*})^\top \\ \cdots \\ (\mathbf{U}^t_{j*})^\top \\ \cdots \\ (\mathbf{U}^t_{m*})^\top \end{bmatrix} \right\|_F \\
& + \left\| \begin{bmatrix} \mathcal{P}_{\Omega_{1*}}\big((\widehat{\mathbf{Y}}^t_{1*})^\top - \mathbf{Y}^\top_{1*}\big) \\ \cdots \\ \mathcal{P}_{\Omega_{j'*}}\big((\widehat{\mathbf{Y}}^t_{j'*})^\top - \mathbf{Y}^\top_{j'*}\big) \\ \cdots \\ \mathcal{P}_{\Omega_{m*}}\big((\widehat{\mathbf{Y}}^t_{m*})^\top - \mathbf{Y}^\top_{m*}\big) \end{bmatrix}^\top \left(\begin{bmatrix} (\mathbf{U}^t_{1*})^\top \\ \cdots \\ (\mathbf{U}^t_{j*})^\top \\ \cdots \\ (\mathbf{U}^t_{m*})^\top \end{bmatrix} - \begin{bmatrix} (\mathbf{U}^t_{1*})^\top \\ \cdots \\ (\mathbf{U}^t_{j'*})^\top \\ \cdots \\ (\mathbf{U}^t_{m*})^\top \end{bmatrix} \right) \right\|_F \\
\leq & \left\| \Big(\mathcal{P}_{\Omega_{j*}}\big((\widehat{\mathbf{Y}}^t_{j*})^\top - \mathbf{Y}^\top_{j*}\big) - \mathcal{P}_{\Omega_{j'*}}\big((\widehat{\mathbf{Y}}^t_{j'*})^\top - \mathbf{Y}^\top_{j'*}\big) \Big)^\top (\mathbf{U}^t_{j*})^\top \right\|_F \\
& + \left\| \Big(\mathcal{P}_{\Omega_{j'*}}\big((\widehat{\mathbf{Y}}^t_{j'*})^\top - \mathbf{Y}^\top_{j'*}\big) \Big)^\top (\mathbf{U}^t_{j*} - \mathbf{U}^t_{j'*})^\top \right\|_F \\
\leq & \, 4G\alpha_1,
\end{aligned}
$$

where the last inequality comes from the projection step, i.e., line 7 in Algorithm 1 and $\mathbf{U}^t$ belongs to $\mathcal{C}_1$. Therefore, according to Lemma A.1, we need to add random Gaussian matrix $\mathbf{N}_2 \in \mathbb{R}^{n\times r}$ with each element following i.i.d. $N(0, \nu_2^2)$. As a result, $\widetilde{\mathbf{V}}^{t+1}$ satisfies $(\gamma, 4\gamma G^2\alpha_1^2/\nu_2^2)$-RDP.

Therefore, by Lemma A.2, we have that after T iterations, Algorithm 1 satisfy $\big(\gamma, 2T\gamma(2G^2\alpha_1^2/\nu_2^2 + \alpha_1^4/\nu_1^2)\big)$-RDP, which implies that Algorithm 1 satisfy $\big(\gamma, \rho_1 + \rho_2\big)$-joint RDP, where $\rho_2 = 4T\gamma G^2\alpha_1^2/\nu_2^2$ and $\rho_1 = 2T\gamma\alpha_1^4/\nu_1^2$.

Next, we translate the joint-RDP to joint-DP according to Lemma A.3. According to the definition of ρ_1 and ρ_2, let $\rho_1 + \rho_2 = \gamma\rho$ and $\rho_1 = \omega\gamma\rho, \rho_2 = (1-\omega)\gamma\rho$ for some $\omega \in (0,1)$. According to Lemma A.3, Algorithm 1 satisfies (ϵ, δ)-joint DP with $\epsilon = \gamma\rho + \log(1/\delta)/(\gamma - 1)$. Therefore, we can choose $\gamma = 1 + \sqrt{\log(1/\delta)/\rho}$ to get the smallest $\epsilon = \rho + 2\sqrt{\log(1/\delta)\rho}$. Thus, we can obtain $\rho = \big(\sqrt{\log(1/\delta) + \epsilon} - \sqrt{\log(1/\delta)}\big)^2$. As a result, we can obtain

$$
\nu_1 = \frac{\sqrt{2T\gamma}\alpha_1^2}{\sqrt{\omega\gamma\rho}} = \frac{\sqrt{2T}\alpha_1^2}{\sqrt{\omega}(\sqrt{\log(1/\delta) + \epsilon} - \sqrt{\log(1/\delta)})} \leq \frac{\alpha_1^2}{\sqrt{\omega}} \frac{\sqrt{8T(\log(1/\delta) + \epsilon)}}{\epsilon}. \tag{A.1}
$$

Furthermore, we have

$$
\nu_2 = \frac{\sqrt{4TG^2\gamma}\alpha_1}{\sqrt{(1-\omega)\gamma\rho}} \leq \frac{\alpha_1}{\sqrt{1-\omega}} \frac{\sqrt{16G^2T(\log(1/\delta) + \epsilon)}}{\epsilon}. \tag{A.2}
$$

Therefore, if we choose $\nu_1 = \alpha_1^2\sqrt{8T(\log(1/\delta) + \epsilon)}/(\sqrt{\omega}\epsilon)$ and $\nu_2 = \alpha_1\sqrt{16G^2T(\log(1/\delta) + \epsilon)}/(\sqrt{1-\omega}\epsilon)$, Algorithm 1 satisfies (ϵ, δ)-joint DP.

B Proof of Theorem 5.2

Note that in the following discussion, we use $\widetilde{\mathbf{V}}$ instead of $\mathbf{V}$ in (3.3) to denote that $\widetilde{\mathbf{V}}$ is private. Furthermore, we have $\|\mathcal{P}_{\Omega_{j*}}\big((\widehat{\mathbf{Y}}^t_{j*})^\top - \mathbf{Y}^\top_{j*}\big)\|_2 \leq 2K\alpha_1\alpha_2$ for all $j \in [m]$, where K is the largest number of observations per user. Since we choose $G = 2K\alpha_1\alpha_2$, the projection step in line 7 will have no effect. According to Algorithm 1, the update rule can be reformulated as follows:

$$
\begin{aligned}
\mathbf{U}^{t+1} &= \mathcal{P}_{\mathcal{C}_1}\bigg(\mathbf{U}^t - \frac{\eta}{p}\mathcal{P}_\Omega\big(\mathbf{U}^t(\widetilde{\mathbf{V}}^t)^\top - \mathbf{Y}\big)\widetilde{\mathbf{V}}^t - \frac{\eta}{2}\mathbf{U}^t\big((\mathbf{U}^t)^\top\mathbf{U}^t - (\widetilde{\mathbf{V}}^t)^\top\widetilde{\mathbf{V}}^t + \mathbf{N}_1\big)\bigg), \\
\widetilde{\mathbf{V}}^{t+1} &= \mathcal{P}_{\mathcal{C}_2}\bigg(\widetilde{\mathbf{V}}^t - \frac{\eta}{p}\big(\mathcal{P}_\Omega\big(\mathbf{U}^t(\widetilde{\mathbf{V}}^t)^\top - \mathbf{Y}\big)^\top\mathbf{U}^t + \mathbf{N}_2\big) - \frac{\eta}{2}\widetilde{\mathbf{V}}^t\big((\widetilde{\mathbf{V}}^t)^\top\widetilde{\mathbf{V}}^t - (\mathbf{U}^t)^\top\mathbf{U}^t - \mathbf{N}_1\big)\bigg).
\end{aligned}
$$

Recall that

$$F(\mathbf{U},\widetilde{\mathbf{V}})=\frac{1}{2p}\big\|\mathcal{P}_\Omega\big(\mathbf{U}\widetilde{\mathbf{V}}^\top-\mathbf{Y}\big)\big\|_F^2+\frac{1}{8}\|\mathbf{U}^\top\mathbf{U}-\widetilde{\mathbf{V}}^\top\widetilde{\mathbf{V}}\|_F^2. \tag{B.1}$$

Therefore, we can obtain

$$\begin{aligned}
&\frac{1}{p}\mathcal{P}_\Omega\big(\mathbf{U}^t(\widetilde{\mathbf{V}}^t)^\top-\mathbf{Y}\big)\widetilde{\mathbf{V}}^t+\frac{1}{2}\mathbf{U}^t\big((\mathbf{U}^t)^\top\mathbf{U}^t-(\widetilde{\mathbf{V}}^t)^\top\widetilde{\mathbf{V}}^t+\mathbf{N}_1\big)=\nabla_{\mathbf{U}}F(\mathbf{U}^t,\widetilde{\mathbf{V}}^t)+\frac{1}{2}\mathbf{U}^t\mathbf{N}_1,\\
&\frac{1}{p}\big(\mathcal{P}_\Omega\big(\mathbf{U}^t(\widetilde{\mathbf{V}}^t)^\top-\mathbf{Y}\big)^\top\mathbf{U}^t+\mathbf{N}_2\big)-\frac{1}{2}\widetilde{\mathbf{V}}^t\big((\widetilde{\mathbf{V}}^t)^\top\widetilde{\mathbf{V}}^t-(\mathbf{U}^t)^\top\mathbf{U}^t-\mathbf{N}_1\big)=\nabla_{\mathbf{V}}F(\mathbf{U}^t,\widetilde{\mathbf{V}}^t)+\frac{1}{p}\mathbf{N}_2+\frac{1}{2}\widetilde{\mathbf{V}}^t\mathbf{N}_1.
\end{aligned}$$

As a result, the update rule in Algorithm 1 is equivalent to:

$$\begin{aligned}
\mathbf{U}^{t+1}&=\mathcal{P}_{\mathcal{C}_1}\big(\mathbf{U}^t-\eta\nabla_{\mathbf{U}}F(\mathbf{U}^t,\widetilde{\mathbf{V}}^t)-(\eta/2)\mathbf{U}^t\mathbf{N}_1\big)=\mathcal{P}_{\mathcal{C}_1}\big(\mathbf{U}^t-\eta\mathbf{G}_{\mathbf{U}}^t\big),\\
\widetilde{\mathbf{V}}^{t+1}&=\mathcal{P}_{\mathcal{C}_2}\big(\widetilde{\mathbf{V}}^t-\eta\nabla_{\mathbf{V}}F(\mathbf{U}^t,\widetilde{\mathbf{V}}^t)-(\eta/p)\mathbf{N}_2-(\eta/2)\widetilde{\mathbf{V}}^t\mathbf{N}_1\big)=\mathcal{P}_{\mathcal{C}_2}\big(\widetilde{\mathbf{V}}^t-\eta\mathbf{G}_{\mathbf{V}}^t\big),
\end{aligned} \tag{B.2}$$

where $\mathbf{G}_{\mathbf{U}}^t=\nabla_{\mathbf{U}}F(\mathbf{U}^t,\widetilde{\mathbf{V}}^t)-\mathbf{U}^t\mathbf{N}_1/2$ and $\mathbf{G}_{\mathbf{V}}^t=\nabla_{\mathbf{V}}F(\mathbf{U}^t,\widetilde{\mathbf{V}}^t)-\mathbf{N}_2/p-\widetilde{\mathbf{V}}^t\mathbf{N}_1/2$.

Let $\mathcal{L}(\mathbf{U}\widetilde{\mathbf{V}}^\top)=(1/2p)\big\|\mathcal{P}_\Omega\big(\mathbf{U}\widetilde{\mathbf{V}}^\top-\widetilde{\mathbf{Y}}\big)\big\|_F^2$ and $\mathbf{Z}=[\mathbf{U};\widetilde{\mathbf{V}}]$, we can rewrite the objective in (B.1) in terms of $\mathbf{Z}$ as follows:

$$\widetilde{F}(\mathbf{Z})=F(\mathbf{U},\widetilde{\mathbf{V}})=\mathcal{L}(\mathbf{U}\widetilde{\mathbf{V}}^\top)+\frac{1}{8}\|\mathbf{U}^\top\mathbf{U}-\widetilde{\mathbf{V}}^\top\widetilde{\mathbf{V}}\|_F^2. \tag{B.3}$$

Therefore, according to the update rule in (B.2), we have the following corresponding gradient estimator (ignoring the projection step)

$$\begin{bmatrix}\nabla_{\mathbf{U}}L(\mathbf{U}\widetilde{\mathbf{V}}^\top)+\frac{1}{2}\mathbf{U}(\mathbf{U}^\top\mathbf{U}-\widetilde{\mathbf{V}}^\top\widetilde{\mathbf{V}})+\frac{1}{2}\mathbf{U}\mathbf{N}_1\\ \nabla_{\mathbf{V}}L(\mathbf{U}\widetilde{\mathbf{V}}^\top)+\frac{1}{2}\widetilde{\mathbf{V}}(\mathbf{U}^\top\mathbf{U}-\widetilde{\mathbf{V}}^\top\widetilde{\mathbf{V}})+\frac{1}{2}\widetilde{\mathbf{V}}\mathbf{N}_1+\frac{1}{p}\mathbf{N}_2\end{bmatrix}=\nabla\widetilde{F}(\mathbf{Z})+\frac{1}{2}\mathbf{Z}\mathbf{N}_1+\frac{1}{p}\widetilde{\mathbf{N}}_2, \tag{B.4}$$

where $\widetilde{\mathbf{N}}_2=[\mathbf{0};\mathbf{N}_2]$.

The following Lemma provides the local curvature condition for the function $\widetilde{F}(\mathbf{Z})$ in (B.3), which has been established in Wang et al. (2017) for nonconvex matrix completion. In the following discussion, we assume $m\geq n$.

Lemma B.1 (Local Curvature Condition)**.** For the objective function $\widetilde{F}(\mathbf{Z})$ in (B.3), where $\mathbf{Z}=[\mathbf{U};\widetilde{\mathbf{V}}]\in\mathbb{R}^{(m+n)\times r}$, $\mathbf{U}\in\mathbb{R}^{m\times r}$ and $\widetilde{\mathbf{V}}\in\mathbb{R}^{n\times r}$, let $\mathbf{X}=\mathbf{U}\widetilde{\mathbf{V}}^\top$, $\mathbf{\Delta}=\mathbf{X}-\mathbf{X}^*$, there exists absolute constants $\{C_i\}_{i=1}^3$, if $|\Omega|\geq C_1rm\log m$, and $\mathbf{\Delta}$ satisfies the following condition

$$\sqrt{\frac{mn}{r}}\frac{\|\mathbf{\Delta}\|_{\infty,\infty}}{\|\mathbf{\Delta}\|_F}\cdot\frac{\|\mathbf{\Delta}\|_*}{\|\mathbf{\Delta}\|_F}\leq\frac{1}{C_2}\sqrt{|\Omega|/(m\log m)}, \tag{B.5}$$

then the following inequality holds with probability at least $1-C_3/d$

$$\begin{aligned}
\langle\nabla\widetilde{F}(\mathbf{Z}),\mathbf{H}\rangle\geq&\frac{\mu}{8}\|\mathbf{X}-\mathbf{X}^*\|_F^2+\frac{\mu\sigma_r}{10}\|\mathbf{H}\|_F^2+\frac{1}{16}\|\mathbf{U}^\top\mathbf{U}-\widetilde{\mathbf{V}}^\top\widetilde{\mathbf{V}}\|_F^2\\
&-\frac{3L+1}{8}\|\mathbf{H}\|_F^4-\left(\frac{4r}{\mu}+\frac{r}{2L}\right)\cdot\|\nabla\mathcal{L}(\mathbf{X}^*)\|_2^2,
\end{aligned}$$

where $\mu=42/43$, $L=44/43$, $\mathbf{R}=\operatorname{argmin}_{\widetilde{\mathbf{R}}\in\mathbb{Q}_r}\|\mathbf{Z}-\mathbf{Z}^*\widetilde{\mathbf{R}}\|_F$ is the optimal rotation with respect to $\mathbf{Z}$, and $\mathbf{H}=\mathbf{Z}-\mathbf{Z}^*\mathbf{R}$. $\mathbf{Z}^*=[\mathbf{U}^*;\mathbf{V}^*]$, $\mathbb{Q}_r$ denotes the set of r by r orthonormal matrices.

The second lemma provides the Local Smoothness Condition.

Lemma B.2 (Local Smoothness Condition)**.** For the objective function $\widetilde{F}(\mathbf{Z})$ in (B.3), where $\mathbf{Z}=[\mathbf{U};\widetilde{\mathbf{V}}]\in\mathbb{R}^{(m+n)\times r}$, $\mathbf{U}\in\mathbb{R}^{m\times r}$ and $\widetilde{\mathbf{V}}\in\mathbb{R}^{n\times r}$, let $\mathbf{X}=\mathbf{U}\widetilde{\mathbf{V}}^\top$, $\mathbf{\Delta}=\mathbf{X}-\mathbf{X}^*$, $\mathbf{G}=[\mathbf{G}_U;\mathbf{G}_V]$, where $\mathbf{G}_U=\nabla_{\mathbf{U}}F(\mathbf{U},\widetilde{\mathbf{V}})-\mathbf{U}\mathbf{N}_1/2$, $\mathbf{G}_V=\nabla_{\mathbf{V}}F\big(\mathbf{U},\widetilde{\mathbf{V}}\big)-\mathbf{N}_2/p-\widetilde{\mathbf{V}}\mathbf{N}_1/2$. Under the same conditions on $|\Omega|$ and $\mathbf{\Delta}$ as in Lemma B.1, we have the following inequality holds with probability at least $1-C/m$

$$\begin{aligned}
\|\mathbf{G}\|_F^2\leq&24L^2\|\mathbf{X}^*-\mathbf{X}\|_F^2\cdot\big\|\mathbf{Z}\big\|_2^2+12r\big\|\nabla\mathcal{L}(\mathbf{X}^*)\big\|_2^2\cdot\big\|\mathbf{Z}\big\|_2^2+6\big\|\mathbf{U}^\top\mathbf{U}-\widetilde{\mathbf{V}}^\top\widetilde{\mathbf{V}}\big\|_F^2\cdot\big\|\mathbf{Z}\big\|_2^2\\
&+2\big\|\mathbf{Z}\big\|_2^2\cdot\big\|\mathbf{N}_1\big\|_F^2+\frac{3}{p^2}\big\|\mathbf{N}_2\big\|_F^2,
\end{aligned}$$

where $L=44/43$ and C is an absolute constant.

Given the local curvature and local smooth conditions, i.e., Lemma B.1 and Lemma B.2, we are ready to prove the main result.

Let $\mathbf{Z}^t = [\mathbf{U}^t; \widetilde{\mathbf{V}}^t]$ and $\mathbf{R}^t = \operatorname{argmin}_{\mathbf{R}\in\mathbb{Q}_r} \|\mathbf{Z}^t - \mathbf{Z}^*\mathbf{R}\|_F$ as the optimal rotation with respect to $\mathbf{Z}^t$, where $\mathbb{Q}_r$ denotes the set of r by r orthonormal matrices. Denote $\mathbf{H}^t = \mathbf{Z}^t - \mathbf{Z}^*\mathbf{R}^t$ and $\mathbf{G}^t = [\mathbf{G}_U^t; \mathbf{G}_V^t]$. By induction, for any $t \geq 0$, we assume $\mathbf{Z}^t \in \mathcal{B}(C_4\sqrt{\sigma_r})$. Therefore, we have

$$\begin{aligned}
\|\mathbf{H}^{t+1}\|_F^2 &\leq \|\mathcal{P}_{\mathcal{C}_1}(\mathbf{U}^t - \eta\mathbf{G}_U^t) - \mathbf{U}^*\mathbf{R}^t\|_F^2 + \|\mathcal{P}_{\mathcal{C}_2}(\widetilde{\mathbf{V}}^t - \eta\mathbf{G}_V^t) - \mathbf{V}^*\mathbf{R}^t\|_F^2 \\
&\leq \|\mathbf{U}^t - \eta\mathbf{G}_U^t - \mathbf{U}^*\mathbf{R}^t\|_F^2 + \|\widetilde{\mathbf{V}}^t - \eta\mathbf{G}_V^t - \mathbf{V}^*\mathbf{R}^t\|_F^2 \\
&= \|\mathbf{H}^t\|_F^2 - 2\eta\langle\mathbf{G}^t, \mathbf{H}^t\rangle + \eta^2\|\mathbf{G}^t\|_F^2 \\
&= \|\mathbf{H}^t\|_F^2 - 2\eta\langle\nabla\widetilde{F}(\mathbf{Z}^t) + \frac{1}{2}\mathbf{Z}^t\mathbf{N}_1 + \frac{1}{p}\mathbf{N}_2, \mathbf{H}^t\rangle + \eta^2\|\mathbf{G}^t\|_F^2, \qquad \text{(B.6)}
\end{aligned}$$

where the first inequality follows from the definition of $\mathbf{H}^t$, the second inequality follows from the non-expansive property of the projection $\mathcal{P}_{\mathcal{C}_i}$ onto $\mathcal{C}_i$ and the fact that $\mathbf{U}^* \in \mathcal{C}_1, \mathbf{V}^* \in \mathcal{C}_2$, the second equality is due to (B.4). Suppose condition (B.5) holds, and thus according to Lemma B.1, we have

$$\begin{aligned}
\langle\nabla\widetilde{F}(\mathbf{Z}^t), \mathbf{H}^t\rangle \geq\ & \frac{\mu}{8}\|\mathbf{X}^t - \mathbf{X}^*\|_F^2 + \frac{\mu\sigma_r}{10}\|\mathbf{H}^t\|_F^2 + \frac{1}{16}\|(\mathbf{U}^t)^\top\mathbf{U}^t - (\widetilde{\mathbf{V}}^t)^\top\widetilde{\mathbf{V}}^t\|_F^2 \\
& - \frac{3L+1}{8}\|\mathbf{H}^t\|_F^4 - \left(\frac{4r}{\mu} + \frac{r}{2L}\right)\cdot\|\nabla\mathcal{L}(\mathbf{X}^*)\|_2^2. \qquad \text{(B.7)}
\end{aligned}$$

Furthermore, by Lemma B.2, we have

$$\begin{aligned}
\mathbb{E}\|\mathbf{G}^t\|_F^2 \leq\ & 24L^2\|\mathbf{X}^* - \mathbf{X}^t\|_F^2 \cdot \big\|\mathbf{Z}^t\big\|_2^2 + 12r\big\|\nabla\mathcal{L}(\mathbf{X}^*)\big\|_2^2 \cdot \big\|\mathbf{Z}^t\big\|_2^2 + 6\big\|(\mathbf{U}^t)^\top\mathbf{U}^t - (\widetilde{\mathbf{V}}^t)^\top\widetilde{\mathbf{V}}^t\big\|_F^2 \cdot \big\|\mathbf{Z}^t\big\|_2^2 \\
& + 2\big\|\mathbf{Z}^t\big\|_2^2 \cdot \big\|\mathbf{N}_1\big\|_F^2 + \frac{3}{p^2}\big\|\mathbf{N}_2\big\|_F^2. \qquad \text{(B.8)}
\end{aligned}$$

In addition, we have

$$\begin{aligned}
\langle\mathbf{Z}^t\mathbf{N}_1 + \frac{1}{p}\mathbf{N}_2, \mathbf{H}^t\rangle &\leq \frac{10}{\mu\sigma_r}\|\mathbf{Z}^t\mathbf{N}_1 + \mathbf{N}_2/p\|_F^2 + \frac{\mu\sigma_r}{40}\|\mathbf{H}^t\|_F^2 \\
&\leq \frac{20}{\mu\sigma_r}\|\mathbf{Z}^t\mathbf{N}_1\|_F^2 + \frac{20}{\mu\sigma_r p^2}\|\mathbf{N}_2\|_F^2 + \frac{\mu\sigma_r}{40}\|\mathbf{H}^t\|_F^2, \qquad \text{(B.9)}
\end{aligned}$$

where the first inequality comes from the Cauchy-Schwarz and Young's inequalities. Note that for any $\mathbf{Z} \in \mathcal{B}(\sqrt{\sigma_r}/4)$, denote $\mathbf{R}$ as the optimal rotation with respect to $\mathbf{Z}$, we have $\|\mathbf{Z}\|_2 \leq \|\mathbf{Z}^*\|_2 + \|\mathbf{Z} - \mathbf{Z}^*\mathbf{R}\|_2 \leq 2\sqrt{\sigma_1}$, and thus we have $\|\mathbf{Z}^t\|_2^2 \leq 4\sigma_1$. Therefore, combining (B.7), (B.8), and (B.9), we can get

$$\begin{aligned}
& -2\eta\langle\nabla\widetilde{F}(\mathbf{Z}^t) + \frac{1}{2}\mathbf{Z}^t\mathbf{N}_1 + \frac{1}{p}\mathbf{N}_2, \mathbf{H}^t\rangle + \eta^2\|\mathbf{G}^t\|_F^2 \\
&\quad\leq -\frac{\mu\eta}{4}\|\mathbf{X}^t - \mathbf{X}^*\|_F^2 - \frac{\eta\mu\sigma_r}{5}\|\mathbf{H}^t\|_F^2 - \frac{\eta}{8}\|(\mathbf{U}^t)^\top\mathbf{U}^t - (\widetilde{\mathbf{V}}^t)^\top\widetilde{\mathbf{V}}^t\|_F^2 \\
&\qquad + \frac{\eta(3L+1)}{4}\|\mathbf{H}^t\|_F^4 + \eta\left(\frac{8r}{\mu} + \frac{r}{L}\right)\cdot\|\nabla\mathcal{L}(\mathbf{X}^*)\|_2^2 \\
&\qquad + 96\eta^2\sigma_1 L^2\|\mathbf{X}^* - \mathbf{X}^t\|_F^2 + 48\eta^2\sigma_1 r\big\|\nabla\mathcal{L}(\mathbf{X}^*)\big\|_2^2 \\
&\qquad + 24\eta^2\sigma_1\big\|(\mathbf{U}^t)^\top\mathbf{U}^t - (\widetilde{\mathbf{V}}^t)^\top\widetilde{\mathbf{V}}^t\big\|_F^2 \\
&\qquad + \left(8\eta^2\sigma_1 + \frac{160\eta\sigma_1}{\mu\sigma_r}\right)\big\|\mathbf{N}_1\big\|_F^2 + \left(\frac{3\eta^2}{p^2} + \frac{40\eta}{\mu\sigma_r p^2}\right)\mathbb{E}\big\|\mathbf{N}_2\big\|_F^2 + \frac{\eta\mu\sigma_r}{20}\|\mathbf{H}^t\|_F^2 \\
&\quad\leq -\frac{3\eta\mu\sigma_r}{20}\|\mathbf{H}^t\|_F^2 + \frac{\eta(3L+1)}{4}\|\mathbf{H}^t\|_F^4 + 20\eta r\|\nabla\mathcal{L}(\mathbf{X}^*)\|_2^2 \\
&\qquad + 200\eta\kappa\big\|\mathbf{N}_1\big\|_F^2 + \frac{50\eta}{\sigma_r p^2}\big\|\mathbf{N}_2\big\|_F^2,
\end{aligned}$$

where $\kappa = \sigma_1/\sigma_r$, and the last inequality is due to the fact that $\eta \leq C_1/\sigma_1$ with $C_1 \leq \min\{\mu/(128L^2), 1/32\}$. Furthermore, we have $\|\mathbf{H}^t\|_F^2 \leq C_2\sigma_r$ with $C_2 \leq 2\mu/(15L+5)$, and thus we have

$$\begin{aligned}
&-2\eta\langle\nabla\widetilde{F}(\mathbf{Z}^t) + \frac{1}{2}\mathbf{Z}^t\mathbf{N}_1 + \frac{1}{p}\mathbf{N}_2, \mathbf{H}^t\rangle + \eta^2\|\mathbf{G}^t\|_F^2 \\
&\qquad \leq -\frac{\eta\mu\sigma_r}{20}\|\mathbf{H}^t\|_F^2 + 20\eta r\|\nabla\mathcal{L}(\mathbf{X}^*)\|_2^2 \\
&\qquad\quad + 200\eta\kappa\big\|\mathbf{N}_1\big\|_F^2 + \frac{50\eta}{\sigma_r p^2}\big\|\mathbf{N}_2\big\|_F^2. \qquad \text{(B.10)}
\end{aligned}$$

Plugging (B.10) into (B.6), we can obtain

$$\begin{aligned}
\|\mathbf{H}^{t+1}\|_F^2 &\leq \bigg(1 - \frac{\eta\mu\sigma_r}{20}\bigg)\|\mathbf{H}^t\|_F^2 + 20\eta r\|\nabla\mathcal{L}(\mathbf{X}^*)\|_2^2 \\
&\quad + 200\eta\kappa\big\|\mathbf{N}_1\big\|_F^2 + \frac{50\eta}{\sigma_r p^2}\big\|\mathbf{N}_2\big\|_F^2 \\
&= \bigg(1 - \frac{\eta\mu\sigma_r}{20}\bigg)\|\mathbf{H}^t\|_F^2 + 20\eta\varepsilon_1 + 50\eta\varepsilon_2,
\end{aligned}$$

where $\varepsilon_1 = r\|\nabla\mathcal{L}(\mathbf{X}^*)\|_2^2$ and $\varepsilon_2 = 4\kappa\|\mathbf{N}_1\|_F^2 + \|\mathbf{N}_2\|_F^2/(\sigma_r p^2)$. As long as we have

$$20\eta\varepsilon_1 + 50\eta\varepsilon_2 \leq C_3\frac{\eta\mu\sigma_r}{20}, \qquad \text{(B.11)}$$

where $C_3 = \min\{1/16, C_2\}$. Thus, we have

$$\|\mathbf{H}^{t+1}\|_F^2 \leq C_3\sigma_r,$$

which satisfies the induction requirement.

Therefore, we can get

$$\|\mathbf{H}^T\|_F^2 \leq \rho^T\|\mathbf{H}^0\|_F^2 + \frac{C_4}{\sigma_r}\varepsilon_1 + \frac{C_5}{\sigma_r}\varepsilon_2,$$

where $\rho = 1 - \eta\mu\sigma_r/20$, and C_4, C_5 are absolute constants.

In addition, according to the observation model in (3.1), each random noise noise E_{jk} follows i.i.d. Gaussian distribution with variance $\nu^2/(mn)$, thus according to Lemma C.4 in Wang et al. (2017), we have

$$\varepsilon_1 \leq C_6\frac{r\nu^2 m\log m}{|\Omega|}, \qquad \text{(B.12)}$$

where C_6 is an absolute constant. Recall that we have

$$\nu_1 = \frac{\alpha_1^2}{\sqrt{\omega}}\frac{\sqrt{8T(\log(1/\delta)+\epsilon)}}{\epsilon} \text{ and } \nu_2 = \frac{\alpha_1^2\alpha_2}{\sqrt{1-\omega}}\frac{\sqrt{64KT(\log(1/\delta)+\epsilon)}}{\epsilon}.$$

Therefore, for ε_2, by the definition of $\mathbf{N}_1$ and $\mathbf{N}_2$ and union bound, we have

$$\begin{aligned}
\varepsilon_2 &= 4\kappa\|\mathbf{N}_1\|_F^2 + \frac{1}{\sigma_r p^2}\|\mathbf{N}_2\|_F^2 \\
&\leq 4\kappa r^2\nu_1^2\log m + \frac{nr}{\sigma_r p^2}\nu_2^2\log m \\
&\leq C_7\frac{\beta^2\sigma_1^2\kappa r^4 T\log(1/\delta)\log m}{\omega m^2\epsilon^2} + C_8\beta^3\sigma_1^3\frac{r^4 KT\log(1/\delta)\log m}{(1-\omega)\sigma_r p^2 m^2\epsilon^2},
\end{aligned}$$

where the last inequality is due to the definition of ν_1, ν_2. Furthermore, under the uniform observation model, we have $K \leq C_8 pn\log m$ holds with probability at least $1 - C_9/m$ by union bound. In addition, we have $\omega = 1/(1+K/p^2)$ and $p = |\Omega|/(mn)$. As a result, we can obtain

$$\varepsilon_2 \leq C_9\beta^3\sigma_1^3\frac{r^4 n^2 T\log(1/\delta)\log^2 m}{\sigma_r m|\Omega|\epsilon^2}$$

and

$$\|\mathbf{H}^T\|_F^2 \leq \rho^T \|\mathbf{H}^0\|_F^2 + C_{10}\frac{r\nu^2 m\log m}{\sigma_r|\Omega|} + C_9\beta^3\sigma_1^3\frac{r^4n^2T\log(1/\delta)\log^2 m}{\sigma_r|\Omega|m\epsilon^2}.$$

Choosing

$$T = O\bigg(\log\bigg(\frac{m^2\epsilon^2}{\beta^3 r^3\kappa^2\sigma_1\log^4 m\log(1/\delta)}\bigg)\bigg),$$

we can obtain

$$\|\mathbf{H}^T\|_F^2 \leq C_{10}\frac{r\nu^2 m\log m}{\sigma_r|\Omega|} + C_{11}\beta^3\sigma_1^3\frac{r^4n^2\log(1/\delta)\log^3 m}{\sigma_r|\Omega|m\epsilon^2}.$$

Thus, according to Lemma E.2, we have

$$\|\mathbf{X}^T - \mathbf{X}^*\|_F^2 \leq C_{12}\frac{r\sigma_1^2\nu^2 m\log m}{\sigma_r|\Omega|} + C_{13}\beta^3\sigma_1^5\frac{r^4n^2\log(1/\delta)\log^3 m}{\sigma_r|\Omega|m\epsilon^2}, \tag{B.13}$$

where $\mathbf{X}^T = \mathbf{U}^T\widetilde{\mathbf{V}}^T$. On the other hand, if condition (B.5) is violated, we have the following extra error term

$$\begin{aligned}\|\mathbf{\Delta}\|_F^2 &\leq C_{14}\big(\sqrt{d_1d_2}\|\mathbf{\Delta}\|_\infty\big)\|\mathbf{\Delta}\|_*\sqrt{\frac{m\log m}{|\Omega|r}}\\ &\leq 2C_{14}\alpha\sqrt{2mn}\|\mathbf{\Delta}\|_F\sqrt{\frac{m\log m}{|\Omega|}},\end{aligned}$$

where $\alpha = \beta r\sigma_1/\sqrt{mn}$, which comes from the incoherence condition of low rank matrices $\mathbf{X}$ and $\mathbf{X}^*$. Hence we can obtain

$$\|\mathbf{\Delta}\|_F^2 \leq 8C_{14}\alpha^2\frac{m\log m}{p} \leq 8C_{14}\frac{\beta^2r^2\sigma_1^2 m\log m}{|\Omega|}. \tag{B.14}$$

As a result, plugging (B.14) into (B.13), we can obtain that

$$\|\mathbf{X}^T - \mathbf{X}^*\|_F^2 \leq C_{15}\frac{(r\sigma_1^2\nu^2 + r^2\beta^2\sigma_1^2\sigma_r)m\log m}{\sigma_r|\Omega|} + C_{13}\beta^3\sigma_1^5\frac{r^4n^2\log(1/\delta)\log^3 m}{\sigma_r|\Omega|m\epsilon^2} \tag{B.15}$$

holds with probability at least $1 - C_{16}/d$, where $\{C_i\}_{i=1}^{16}$ are some constants.

C Proof of Theorem 5.3

We first establish the privacy guarantee. According to Algorithm 2, we have $\mathbf{A} = \mathbf{Z}^\top\mathbf{Z} + \mathbf{N}_0$, where $\mathbf{Z} = \tau\mathbf{Y}/p$, $\mathbf{Y}$ is the observation matrix, τ is a tuning parameter, and $p = |\Omega|/(mn)$. Therefore, the sensitivity of $\mathbf{Z}^\top\mathbf{Z}$ is $2\tau^2G^2/p^2$. According to Gaussian mechanism (Dwork et al., 2014; Bun and Steinke, 2016), if each element in the upper triangle (including the diagonal) of $\mathbf{N}_0$ follows i.i.d. $N(0,\nu_0^2)$ with $\nu_0 = 4G^2\tau^2\sqrt{\log(1/\delta)}/(p^2\epsilon)$, then performing the rank r SVD of $\mathbf{A}$ to obtain $\widetilde{\mathbf{V}}^0 = \mathbf{V}^0\mathbf{\Sigma}^{1/2}$, where $\mathbf{V}^0 \in \mathbb{R}^{m\times r}$ and $\mathbf{\Sigma} \in \mathbb{R}^{r\times r}$, satisfies (ϵ,δ)-DP. Furthermore, each user $i \in [m]$ will obtain its own initialization $\mathbf{U}_{i*}^{0\top} = \tau\mathbf{Y}_{i*}^\top\mathbf{V}^0\mathbf{\Sigma}^{-1/2}/p$. Since $\mathbf{V}^0, \mathbf{\Sigma}$ satisfy (ϵ,δ)-DP and each user holds its own $\mathbf{U}_{i*}^0$, Algorithm 2 satisfies (ϵ,δ)-joint DP.

Next, we will bound $\|\mathbf{U}_0\widetilde{\mathbf{V}}_0^\top - \mathbf{X}^*\|_F$, where $\mathbf{U}_0 = [\mathbf{U}_{1*}^{0\top};\ldots;\mathbf{U}_{m*}^{0\top}]$. In the following discussion, we use $\{C_i\}_{i=1}^{17}$ to denote absolute constants. According to the definition of $\mathbf{U}_0$, $\widetilde{\mathbf{V}}_0$, we have

$$\mathbf{U}_0\widetilde{\mathbf{V}}_0^\top = \mathbf{Z}\mathbf{V}^0\mathbf{\Sigma}^{-1/2}\mathbf{\Sigma}^{1/2}\mathbf{V}^{0\top} = \mathbf{Z}\mathbf{V}^0\mathbf{V}^{0\top}.$$

Let the rank r SVD of $\mathbf{Z}$ to be $\bar{\mathbf{U}} \in \mathbb{R}^{m\times r}$, $\bar{\mathbf{\Sigma}} \in \mathbb{R}^{r\times r}$, $\bar{\mathbf{V}} \in \mathbb{R}^{n\times r}$. Therefore, we can obtain

$$\begin{aligned}\|\mathbf{U}_0\widetilde{\mathbf{V}}_0^\top - \mathbf{X}^*\|_F &= \|\mathbf{Z}\mathbf{V}^0\mathbf{V}^{0\top} - \mathbf{X}^*\|_F\\ &\leq \|\mathbf{Z}\bar{\mathbf{V}}\bar{\mathbf{V}}^\top - \mathbf{X}^*\|_F + \|\mathbf{Z}(\mathbf{V}^0\mathbf{V}^{0\top} - \bar{\mathbf{V}}\bar{\mathbf{V}}^\top)\|_F\\ &\leq \|\bar{\mathbf{U}}\bar{\mathbf{\Sigma}}\bar{\mathbf{V}}^\top - \mathbf{X}^*\|_F + \sqrt{2r}\|\mathbf{Z}\|_2\cdot\|\mathbf{V}^0\mathbf{V}^{0\top} - \bar{\mathbf{V}}\bar{\mathbf{V}}^\top\|_2\\ &\leq 2\sqrt{1-\mu\tau}\|\mathbf{X}^*\|_F + 2\tau\sqrt{3r}\varepsilon\\ &\quad + \frac{\tau\sqrt{2r}}{p}\|\mathcal{P}_\Omega(\mathbf{X}^* + \mathbf{E})\|_2\cdot\|\mathbf{V}^0\mathbf{V}^{0\top} - \bar{\mathbf{V}}\bar{\mathbf{V}}^\top\|_2,\end{aligned} \tag{C.1}$$

where $\varepsilon = \|\mathcal{P}_\Omega(\mathbf{E})/p\|_2 \leq C_1\nu\sqrt{m\log m/|\Omega|}$ by Lemma C.4 in Wang et al. (2017). According to Lemma A.3 in Jin et al. (2016), we have the following holds with probability at least $1 - C_2/m$,

$$\|\mathcal{P}_\Omega(\mathbf{X}^* + \mathbf{E})/p - \mathbf{X}^*\|_2 \leq C_3(\sigma_1 + \nu)\sqrt{\frac{rm\log m}{|\Omega|}}. \tag{C.2}$$

Let $\mathbf{Y}/p - \mathbf{X}^* = \mathbf{E}_0$, according to (C.2), we have $\|\mathbf{E}_0\|_2 \leq \sigma_r^2/(8\sigma_1)$ as long as $|\Omega| \geq C_4(\sigma_1+\nu)^2\sigma_1^2 rm\log m/\sigma_r^4$. Recall that

$$\mathbf{Z}^\top\mathbf{Z} = \frac{\tau^2}{p^2}\mathbf{Y}^\top\mathbf{Y} = \tau^2\mathbf{X}^{*\top}\mathbf{X}^* + \tau^2\mathbf{E}_1,$$

where $\mathbf{E}_1 = \mathbf{E}_0^\top\mathbf{E}_0 + \mathbf{E}_0^\top\mathbf{X}^* + \mathbf{X}^{*\top}\mathbf{E}_0$. Hence, we can obtain that

$$\|\mathbf{E}_1\|_2 \leq \|\mathbf{E}_0\|_2^2 + 2\|\mathbf{E}_0\|_2\cdot\|\mathbf{X}^*\|_2 \leq \frac{\sigma_r^4}{64\sigma_1^2} + \frac{\sigma_r^2}{4} \leq \frac{\sigma_r^2}{2}.$$

Therefore, by Weyl's inequality, we can get $\lambda_r(\mathbf{Z}) - \lambda_{r+1}(\mathbf{Z}) \geq \tau^2\sigma_r^2 - \tau^2\sigma_r^2/2 = \tau^2\sigma_r^2/2$. Furthermore, according to Theorem 4.4.5 in Vershynin (2018), we have the following holds with probability at least $1 - C_5/m$

$$\|\mathbf{N}_0\|_2 \leq C_6\nu_0\sqrt{n} \leq 4C_6\frac{\tau^2\sqrt{n}K\alpha_1^2\alpha_2^2\sqrt{\log(1/\delta)}}{p^2\epsilon} \leq 4C_6\frac{\tau^2 n^{3/2}r^2\beta^2\sigma_1^2\log m\sqrt{\log(1/\delta)}}{|\Omega|\epsilon},$$

where the second inequality comes from the fact that $G^2 = K\alpha_1^2\alpha_2^2$, the last inequality is due to the fact that $K \leq C_7 pn\log m$ with probability at least $1 - C_8/m$. Hence, if $|\Omega| \geq C_9 r^2 n^{3/2}\beta^2\sigma_1^2\log m/(\tau^2\sigma_r^2)$, we have $\|\mathbf{N}_0\|_2 \leq \tau^2\sigma_r^2/4 \leq 2(\lambda_r(\mathbf{Z}) - \lambda_{r+1}(\mathbf{Z}))$. Therefore, according to Theorem 6 in Dwork et al. (2014), we have

$$\|\mathbf{V}^0\mathbf{V}^{0\top} - \bar{\mathbf{V}}\bar{\mathbf{V}}^\top\|_2 \leq \frac{2\|\mathbf{N}_0\|_2}{\lambda_r(\mathbf{Z}) - \lambda_{r+1}(\mathbf{Z})}. \tag{C.3}$$

Plugging the upper bound of $\|\mathbf{N}_0\|_2$ and the lower bound of $\lambda_r(\mathbf{Z}) - \lambda_{r+1}(\mathbf{Z})$ into (C.3), we can obtain

$$\|\mathbf{V}^0\mathbf{V}^{0\top} - \bar{\mathbf{V}}\bar{\mathbf{V}}^\top\|_2 \leq \frac{4\|\mathbf{N}_0\|_2}{\tau^2\sigma_r^2} \leq C_{10}\frac{n^{3/2}r^2\beta^2\sigma_1^2\log m\sqrt{\log(1/\delta)}}{|\Omega|\sigma_r^2\epsilon}, \tag{C.4}$$

Therefore, plugging the upper bound of $\|\mathbf{E}_0\|_2$ and the result in (C.4) into (C.1), we can get

$$\|\mathbf{U}_0\widetilde{\mathbf{V}}_0^\top - \mathbf{X}^*\|_F \leq 2\sqrt{1-\mu\tau}\|\mathbf{X}^*\|_F + C_{11}\tau\nu\sqrt{\frac{r\log m}{|\Omega|}} + C_{12}\tau\frac{n^{3/2}r^{5/2}\beta^2\sigma_1^3\log m\sqrt{\log(1/\delta)}}{|\Omega|\sigma_r^2\epsilon},$$

where the inequality is due to (C.2) that if we have $|\Omega| \geq C_{13}(1+\nu^2)rm\log m$, we have

$$\frac{\tau\sqrt{2r}}{p}\|\mathcal{P}_\Omega(\mathbf{X}^* + \mathbf{E})\|_2 \leq 2\tau\sqrt{2r}\|\mathbf{X}^*\|_2.$$

As a result, if we have $\tau \geq C_{14}(1 - \sigma_r^2/4\|\mathbf{X}^*\|_F^2)$, $|\Omega| \geq C_{15}r^2\nu^2\log m/\sigma_r^2$ and $|\Omega| \geq C_{16}\kappa^3 r^{5/2}n^{3/2}\log m$, where $\kappa = \sigma_1/\sigma_r$, we can obtain that $\|\mathbf{U}_0\widetilde{\mathbf{V}}_0^\top - \mathbf{X}^*\|_F \leq C_{17}\sigma_r$.

D Proof of Lemma B.2

According to the definition of $\mathbf{G}$, we have

$$\begin{aligned}
\|\mathbf{G}\|_F^2 &= \Big\|\nabla_{\mathbf{U}}\mathcal{L}(\mathbf{U}\widetilde{\mathbf{V}}^\top) + \frac{1}{2}\mathbf{U}(\mathbf{U}^\top\mathbf{U} - \widetilde{\mathbf{V}}^\top\widetilde{\mathbf{V}}) + \frac{1}{2}\mathbf{U}\mathbf{N}_1\Big\|_F^2 \\
&\quad + \Big\|\nabla_{\mathbf{V}}\mathcal{L}(\mathbf{U}\widetilde{\mathbf{V}}^\top) + \frac{1}{2}\widetilde{\mathbf{V}}(\mathbf{U}^\top\mathbf{U} - \widetilde{\mathbf{V}}^\top\widetilde{\mathbf{V}}) + \frac{1}{2}\widetilde{\mathbf{V}}\mathbf{N}_1 + \frac{1}{p}\mathbf{N}_2\Big\|_F^2 \\
&\le 3\Big\|\nabla_{\mathbf{U}}\mathcal{L}(\mathbf{U}\widetilde{\mathbf{V}}^\top) + \frac{1}{2}\mathbf{U}(\mathbf{U}^\top\mathbf{U} - \widetilde{\mathbf{V}}^\top\widetilde{\mathbf{V}})\Big\|_F^2 + 3\Big\|\nabla_{\mathbf{V}}\mathcal{L}(\mathbf{U}\widetilde{\mathbf{V}}^\top) + \frac{1}{2}\widetilde{\mathbf{V}}(\mathbf{U}^\top\mathbf{U} - \widetilde{\mathbf{V}}^\top\widetilde{\mathbf{V}})\Big\|_F^2 \\
&\quad + 3\Big\|\frac{1}{2}\mathbf{U}\mathbf{N}_1\Big\|_F^2 + 3\Big\|\frac{1}{2}\widetilde{\mathbf{V}}\mathbf{N}_1\Big\|_F^2 + 3\Big\|\frac{1}{p}\mathbf{N}_2\Big\|_F^2 \\
&\le 6\big\|\nabla_{\mathbf{U}}\mathcal{L}(\mathbf{U}\widetilde{\mathbf{V}}^\top)\big\|_F^2 + 6\big\|\nabla_{\mathbf{V}}\mathcal{L}(\mathbf{U}\widetilde{\mathbf{V}}^\top)\big\|_F^2 + 3\big\|\mathbf{U}^\top\mathbf{U} - \widetilde{\mathbf{V}}^\top\widetilde{\mathbf{V}}\big\|_F^2 \cdot \big(\|\mathbf{U}\|_2^2 + \|\widetilde{\mathbf{V}}\|_2^2\big) \\
&\quad + \frac{3}{4}\big(\|\mathbf{U}\|_2^2 + \|\widetilde{\mathbf{V}}\|_2^2\big) \cdot \|\mathbf{N}_1\|_F^2 + \frac{3}{p^2}\|\mathbf{N}_2\|_F^2 \\
&\le 6\big\|\nabla_{\mathbf{U}}\mathcal{L}(\mathbf{U}\widetilde{\mathbf{V}}^\top)\big\|_F^2 + 6\big\|\nabla_{\mathbf{V}}\mathcal{L}(\mathbf{U}\widetilde{\mathbf{V}}^\top)\big\|_F^2 + 6\big\|\mathbf{U}^\top\mathbf{U} - \widetilde{\mathbf{V}}^\top\widetilde{\mathbf{V}}\big\|_F^2 \cdot \|\mathbf{Z}\|_2^2 \\
&\quad + 2\|\mathbf{Z}\|_2^2 \cdot \|\mathbf{N}_1\|_F^2 + \frac{3}{p^2}\|\mathbf{N}_2\|_F^2, \qquad \text{(D.1)}
\end{aligned}$$

where the first inequality comes from the facts that $\|\mathbf{A}+\mathbf{B}\|_F^2 \le 2\|\mathbf{A}\|_F^2 + 2\|\mathbf{B}\|_F^2$ and $\|\mathbf{A}\mathbf{B}\|_F \le \|\mathbf{A}\|_2 \cdot \|\mathbf{B}\|_F$, and the last inequality is due to the fact that $\max\{\|\mathbf{U}\|_2, \|\widetilde{\mathbf{V}}\|_2\} \le \|\mathbf{Z}\|_2$. In addition, we have

$$\begin{aligned}
\big\|\nabla_{\mathbf{U}}\mathcal{L}(\mathbf{U}\widetilde{\mathbf{V}}^\top)\big\|_F^2 + \big\|\nabla_{\mathbf{V}}\mathcal{L}(\mathbf{U}\widetilde{\mathbf{V}}^\top)\big\|_F^2 &= \big\|\nabla\mathcal{L}(\mathbf{X})\widetilde{\mathbf{V}}\big\|_F^2 + \big\|\nabla\mathcal{L}(\mathbf{X})^\top\mathbf{U}\big\|_F^2 \\
&\le 2\big\|\big(\nabla\mathcal{L}(\mathbf{X}) - \nabla\mathcal{L}(\mathbf{X}^*)\big)\widetilde{\mathbf{V}}\big\|_F^2 + 2r\big\|\nabla\mathcal{L}(\mathbf{X}^*)\big\|_2^2 \cdot \|\widetilde{\mathbf{V}}\|_2^2 \\
&\quad + 2\big\|\big(\nabla\mathcal{L}(\mathbf{X}) - \nabla\mathcal{L}(\mathbf{X}^*)\big)^\top\mathbf{U}\big\|_F^2 + 2r\big\|\nabla\mathcal{L}(\mathbf{X}^*)\big\|_2^2 \cdot \|\mathbf{U}\|_2^2 \\
&\le 2L^2\|\mathbf{X}^* - \mathbf{X}\|_F^2 \cdot \big(\|\widetilde{\mathbf{V}}\|_2^2 + \|\mathbf{U}\|_2^2\big) + 2r\big\|\nabla\mathcal{L}(\mathbf{X}^*)\big\|_2^2 \cdot \|\mathbf{Z}\|_2^2 \\
&\le 4L^2\|\mathbf{X}^* - \mathbf{X}\|_F^2 \cdot \|\mathbf{Z}\|_2^2 + 2r\big\|\nabla\mathcal{L}(\mathbf{X}^*)\big\|_2^2 \cdot \|\mathbf{Z}\|_2^2, \qquad \text{(D.2)}
\end{aligned}$$

where the first equality is due to the definition of $\mathcal{L}$, the second inequality is due to the restricted strong convexity and smoothness conditions of $\mathcal{L}$, which hold with probability at least $1 - C/m$ for an absolute constant C and has been provided given condition (B.5). Plugging (D.2) into (D.1), we can get

$$\begin{aligned}
\|\mathbf{G}\|_F^2 \le 24L^2\|\mathbf{X}^* - \mathbf{X}\|_F^2 \cdot \|\mathbf{Z}\|_2^2 + 12r\big\|\nabla\mathcal{L}(\mathbf{X}^*)\big\|_2^2 \cdot \|\mathbf{Z}\|_2^2 + 6\big\|\mathbf{U}^\top\mathbf{U} - \widetilde{\mathbf{V}}^\top\widetilde{\mathbf{V}}\big\|_F^2 \cdot \|\mathbf{Z}\|_2^2 \\
+ 2\|\mathbf{Z}\|_2^2 \cdot \|\mathbf{N}_1\|_F^2 + \frac{3}{p^2}\|\mathbf{N}_2\|_F^2.
\end{aligned}$$

E Auxiliary Lemmas

Let $d(\mathbf{Z}, \mathbf{Z}') = \min_{\mathbf{R}\in\mathbb{Q}_r}\|\mathbf{Z} - \mathbf{Z}'\mathbf{R}\|_F$, where $\mathbb{Q}_r$ is the set of r-by-r orthonormal matrices. We use the following lemmas in our proofs, which are provided in Tu et al. (2016).

Lemma E.1. For any matrices $\mathbf{Z}, \mathbf{Z}' \in \mathbb{R}^{(d_1+d_2)\times r}$, we have the following inequality

$$d^2(\mathbf{Z}, \mathbf{Z}') \le \frac{1}{2(\sqrt{2}-1)\sigma_r^2(\mathbf{Z}')}\|\mathbf{Z}\mathbf{Z}^\top - \mathbf{Z}'\mathbf{Z}'^\top\|_F^2.$$

Lemma E.2. For any matrices $\mathbf{Z}, \mathbf{Z}' \in \mathbb{R}^{(d_1+d_2)\times r}$, which satisfy $d(\mathbf{Z}, \mathbf{Z}') \le \|\mathbf{Z}'\|_2/4$, we have the following inequality

$$\|\mathbf{Z}\mathbf{Z}^\top - \mathbf{Z}'\mathbf{Z}'^\top\|_F \le \frac{9}{4}\|\mathbf{Z}'\|_2 \cdot d(\mathbf{Z}, \mathbf{Z}').$$