Homomorphic Matrix Completion

Xiao-Yang Liu[1*], **Zechu Li**[2,*], **Xiaodong Wang**[1]
[1]Department of Electrical Engineering, Columbia University, New York,
[2]Department of Computer Science, Columbia University, New York,
{xl2427, zl2993, xw2008}@columbia.edu

Abstract

In recommendation systems, global positioning, system identification, and mobile social networks, it is a fundamental routine that a server completes a low-rank matrix from an observed subset of its entries. However, sending data to a cloud server raises up the data privacy concern due to eavesdropping attacks and the single-point failure problem, e.g., the Netflix prize contest was canceled after a privacy lawsuit. In this paper, we propose a homomorphic matrix completion algorithm for privacy-preserving purpose. First, we formulate a *homomorphic matrix completion* problem where a server performs matrix completion on cyphertexts, and propose an encryption scheme that is fast and easy to implement. Secondly, we prove that the proposed scheme satisfies the *homomorphism property* that decrypting the recovered matrix on cyphertexts will obtain the target matrix (on plaintexts). Thirdly, we prove that the proposed scheme satisfies an (ϵ, δ)-differential privacy property. While with similar level of privacy guarantee, we reduce the best-known error bound $O(\sqrt[10]{n_1^3 n_2})$ to EXACT recovery at a price of more samples. Finally, on synthetic data and real-world data, we show that both homomorphic nuclear-norm minimization and alternating minimization algorithms achieve accurate recoveries on cyphertexts, verifying the homomorphism property.

1 Introduction

The recurring low-rank matrix completion problem [4, 5, 20, 28, 11, 27, 44] concerns completing a low-rank matrix from a randomly observed subset of entries. It has wide applications in recommendation systems (collaborative filtering) [1, 45, 24], computer vision [2, 14, 25], global positioning [46], sensory data analysis in Internet of Things [23, 32, 31], system identification, network data analysis [49], mobile social networks [21, 36], etc. Existing works [5, 8] have demonstrated a remarkable fact: if an $n \times n$ matrix with rank $r \ll n$ satisfies a certain incoherence property, then with high probability, it is possible to exactly recover the matrix from $O(nr\mathbf{poly}\log n) \ll n^2$ entries using polynomial-time algorithms. Intuitively, one needs roughly $(2nr - r^2)$ parameters [5] (this is the dimension of the tangent space to the manifold of rank-r matrices) to fix an $n \times n$ matrix of rank r, and the sampling randomness introduces a $\log n$ factor due to a coupon collector's effect. The information theoretical lower bound is $\Omega(nr \log n)$ [5], while the tightest known upper bound is $O(nr \log^2 n)$ [8] with another $\log n$ factor from the Golfing scheme used by the recovery algorithms.

The low-rank matrix completion problem usually deals with large-scale matrices that involve extensive computations, while in mobile computing, smart devices usually outsource such a huge computation task to a cloud server. However, revealing data to a server or releasing anonymized data raises up privacy concerns [21, 42, 40], e.g., the recommendation contest Netflix prize was canceled after privacy lawsuit [33]. There are two major obstructive factors: anonymization in data publishing is still vulnerable, and storing sensitive data on a cloud server may encounter the single-point of failure

*Equal contribution.

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

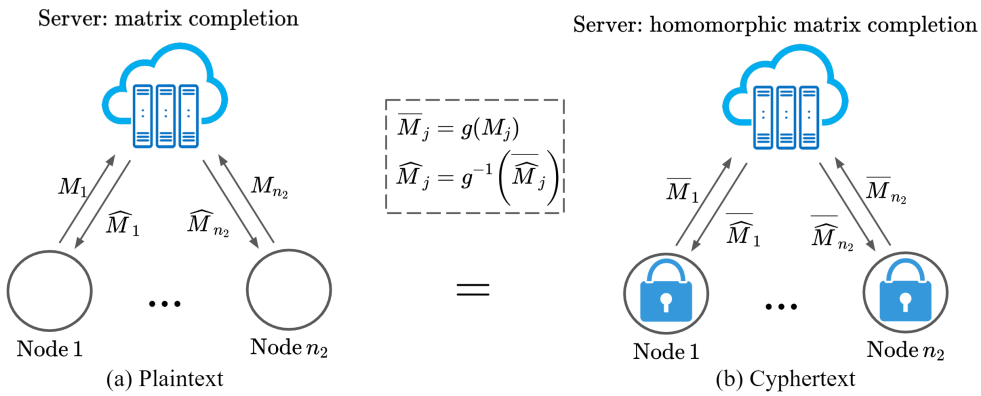

Figure 1: Matrix completion on plaintext versus homomorphic matrix completion on cyphertext.

(SPOF) problem, say hackers. Existing works [18, 17, 9] address the privacy concern in various ways, e.g., a popular approach is to [18] add noise to the data, therefore making a tradeoff between the recovery accuracy and the level of privacy.

In cloud computing and distributed systems, the homomorphism property [12, 43] allows computations to be carried out on cyphertexts, generating an encrypted result which, when decrypted, matches the result of operations performed on the corresponding plaintexts. In this manner, **homomorphic encryption securely chains together different services without sacrificing recovery accuracy, but at a price of more samples**. There are several partially homomorphic crypto-systems, and also a number of fully homomorphic crypto-systems [12, 43]. In addition, the homomorphic property can also be used to create many other secure systems, for example secure voting systems, collision-resistant hash functions, private information retrieval schemes [41], etc.

In this paper, we integrate the large-scale distributed matrix completion task with a homomorphic encryption-decryption scheme, which guarantees the EXACT recovery and differential privacy at a price of more samples. First, we define the *homomorphic matrix completion problem* that ensures data privacy by preserving a similarly homomorphism property between plaintexts and cyphertexts. Specifically, we propose a homomorphic encryption-decryption scheme, in which each node performs local encryption and decryption, and uploads an encrypted incomplete vector to a server that carries out the matrix completion computation. Then, we theoretically prove that the proposed scheme satisfies the homomorphism and differential privacy properties — reducing the best-known error bound $O(\sqrt[10]{n_1^3 n_2})$ [18] to EXACT recovery. Finally, based on synthetic and real-world data, we show that the homomorphic nuclear-norm minimization and alternating minimization algorithms achieve accurate recoveries on both cyphertexts and plaintexts, verifying the homomorphism property.

2 Homomorphic Matrix Completion Problem

First, we formally define the homomorphic matrix completion problem. Then, we introduce an notation of privacy by adapting the join (ϵ, δ) differential privacy, which is a subspace-aware variant.

2.1 Notations and Preliminaries

Notations: Let $\boldsymbol{e}_i$ denote the i-th standard basis, $\boldsymbol{I}_k$ denote the $k \times k$ identity matrix, and $\mathcal{I}$ denote the identity linear operator. For matrix $\boldsymbol{X}$, the (i, j)-th element is $\boldsymbol{X}_{ij}$ or $\boldsymbol{X}(i, j)$, the j-th column is $\boldsymbol{X}_j$, and the transpose is $\boldsymbol{X}^\top$. The concatenation of two matrices $\boldsymbol{A} \in \mathbb{R}^{n_1 \times n_2}$ and $\boldsymbol{B} \in \mathbb{R}^{n_1 \times n_3}$ with the same number of rows is denoted by $[\boldsymbol{A},\ \boldsymbol{B}] \in \mathbb{R}^{n_1 \times (n_2 + n_3)}$. By *with high probability* (w.h.p.) we mean that with probability at least $1 - c_1 n^{-c_2}$ for some positive constants c_1, c_2. Let $\mathcal{N}(0, \sigma^2)$ denote a Gaussian distribution with mean 0 and standard deviation σ. We use an overline to represent the encrypted version of a variable: variables before encryption are called *plaintexts*, e.g., $\boldsymbol{X}$, while the encrypted variables are called *cyphertexts*, e.g., $\overline{\boldsymbol{X}}$.

Let $\boldsymbol{M}_\Omega \in \mathbb{R}^{n_1 \times n_2}$ denote the observed entries of a data matrix $\boldsymbol{M} \in \mathbb{R}^{n_1 \times n_2}$, where $\Omega \subseteq \{(1, 1), (1, 2), ..., (n_1, n_2)\}$ indicates the observed entries. We define a linear operator

$\mathcal{P}_\Omega : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^{n_1 \times n_2}$ to represent the observation model as follows

$$[\mathcal{P}_\Omega(\boldsymbol{M})]_{ij} = \begin{cases} \boldsymbol{M}_{ij}, \text{ if } (i,j) \in \Omega \\ 0, \text{ otherwise.} \end{cases} \tag{1}$$

Assuming the true data matrix $\boldsymbol{M}$ is low-rank, i.e., $\text{rank}(\boldsymbol{M}) = r \ll \min(n_1, n_2)$. The (compact) singular value decomposition (SVD) is $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$, where $\boldsymbol{U} \in \mathbb{R}^{n_1 \times r}$ represents r left singular vectors (a basis of the column subspace), $\boldsymbol{V} \in \mathbb{R}^{n_2 \times r}$ represents r right singular vectors (a basis of the row subspace), and $\boldsymbol{\Sigma} = \text{diag}([\sigma_1, \sigma_1, \cdots, \sigma_r]) \in \mathbb{R}^{r \times r}$ with singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq 0$. The ℓ_2-norm of a vector is $||\boldsymbol{x}||_2$, while the Frobenius norm and nuclear norm of $\boldsymbol{M}$ are $||\boldsymbol{M}||_F = \sqrt{\sum_{i,j} |\boldsymbol{M}_{ij}|^2}$ and $||\boldsymbol{M}||_* = \sum_{i=1}^r \sigma_i$, respectively. The operator norm (spectral norm) of a matrix and a linear operator $\mathcal{L}$ are defined as follows

$$||\boldsymbol{M}|| \triangleq \sup_{\boldsymbol{x} \in \mathbb{R}^{n_2},\ ||\boldsymbol{x}||_2 \leq 1} ||\boldsymbol{M}\boldsymbol{x}||_2 = \sigma_1, \text{ and } ||\mathcal{L}|| \triangleq \sup_{||\boldsymbol{X}||_F \leq 1} ||\mathcal{L}(\boldsymbol{X})||_F. \tag{2}$$

Definition 1. ***(Column subspace and null space** [26]) Let $\boldsymbol{A} \in \mathbb{R}^{n_1 \times n_2}$. The set $\mathcal{S}(\boldsymbol{A}) = \{\boldsymbol{b} \in \mathbb{R}^{n_1} |\ \boldsymbol{b} = \boldsymbol{A}\boldsymbol{x},\ \boldsymbol{x} \in \mathbb{R}^{n_2}\}$ is the column space or range of $\boldsymbol{A}$, and the set **Ker**$(\boldsymbol{A}) = \{\boldsymbol{x} \in \mathbb{R}^{n_2} |\ \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}\}$ is the kernel or (right) null space of $\boldsymbol{A}$.*

The null space (kernel space) of operator $\mathcal{P}_\Omega$ is **Ker**$(\mathcal{P}_\Omega) = \{\boldsymbol{Z} \in \mathbb{R}^{n_1 \times n_2} \mid \mathcal{P}_\Omega(\boldsymbol{Z}) = \boldsymbol{0}\}$, which is denoted as $\Omega^\perp$ [2]. Let $\Omega \sim$ **Uni**(m) denote a set with m entries, which is sampled uniformly from all sets of m entries, and $\Omega \sim$ **Ber**(p) denote a set with $\mathbb{E}|\Omega| = m$ entries, where each entry is sampled independently according to a Bernoulli model with $p = m/(n_1 n_2)$.

Let $\boldsymbol{P_U}$ and $\boldsymbol{P_V}$ denote the orthogonal projections onto the column and row space of $\boldsymbol{M}$, respectively

$$\boldsymbol{P_U} = \sum_{i \in [r]} \boldsymbol{u}_i \boldsymbol{u}_i^\top = \boldsymbol{U}\boldsymbol{U}^\top, \quad \boldsymbol{P_V} = \sum_{i \in [r]} \boldsymbol{v}_i \boldsymbol{v}_i^\top = \boldsymbol{V}\boldsymbol{V}^\top. \tag{3}$$

Define an orthogonal decomposition $\mathbb{R}^{n_1 \times n_2} = T \oplus T^\perp$, where T is the linear space spanned by matrices with the same column space or row space as $\boldsymbol{M}$, and $T^\perp$ is its orthogonal complement that consists of matrices with row-space orthogonal to the row-space $\boldsymbol{V}$ and column-space orthogonal to the column-space $\boldsymbol{U}$. T can be expressed as follows

$$T = \{\boldsymbol{U}\boldsymbol{A}^\top + \boldsymbol{B}\boldsymbol{V}^\top \mid \boldsymbol{A} \in \mathbb{R}^{n_1 \times r},\ \boldsymbol{B} \in \mathbb{R}^{n_2 \times r}\}. \tag{4}$$

The orthogonal projection $\mathcal{P}_T$ onto T and the orthogonal projection onto $T^\perp$ are as follows

$$\begin{aligned} \mathcal{P}_T(\boldsymbol{X}) &= \boldsymbol{P_U}\boldsymbol{X} + \boldsymbol{X}\boldsymbol{P_V} - \boldsymbol{P_U}\boldsymbol{X}\boldsymbol{P_V}, \\ \mathcal{P}_{T^\perp}(\boldsymbol{X}) &= (\mathcal{I} - \mathcal{P}_T)(\boldsymbol{X}) = (\boldsymbol{I}_{n_1} - \boldsymbol{P_U})\boldsymbol{X}(\boldsymbol{I}_{n_2} - \boldsymbol{P_V}). \end{aligned} \tag{5}$$

2.2 Problem Formulation for Homomorphic Matrix Completion

We are interested in completing large-scale matrices and want to outsource this compute-intensive task from mobile devices to a cloud server. Here we aim to preserve the matrix entries from leakage, which is the key concern for recommendation systems as in Netflix's privacy lawsuit [33].

Distributed matrix completion problem on plaintexts. Assume that there are n_2 nodes with limited computing power and a cloud server with superior computing power. The j-th node has an attribute vector $\boldsymbol{M}_j \in \mathbb{R}^{n_1}$, $j = 1, ..., n_2$, however, it is incomplete and the observed entries are indexed by a set $\Omega_j \subseteq \{(1,j), (2,j), ..., (n_1,j)\}$. We assume that the true values of these n_2 vectors form a low-rank matrix $\boldsymbol{M} \in \mathbb{R}^{n_1 \times n_2}$ with rank $r \ll \min(n_1, n_2)$, the ℓ_2-norms of the attribute vectors is bounded by L, i.e., $\max_{1 \leq j \leq n_2} ||\boldsymbol{M}_j||_2 \leq L$, and the observation set $\Omega = \bigcup_{j=1,...,n_2} \Omega_j$. Nodes upload their incomplete vectors to a cloud server to carry out a matrix completion task

$$\text{Find a matrix } \boldsymbol{X} \in \mathbb{R}^{n_1 \times n_2}, \text{ s.t. } \mathcal{P}_\Omega(\boldsymbol{X}) = \mathcal{P}_\Omega(\boldsymbol{M}),\ \text{rank}(\boldsymbol{X}) \leq r, \tag{6}$$

where $\Omega \sim$ **Uni**(m) and r may be unknown. Without loss of generality, we assume that $n_1 \leq n_2$ from now on. Note that our formulation also includes the case [35] where a matrix is distributed into blocks and then is completed in parallel.

[2] We adopt the notation $\Omega^\perp$ since **Ker**$(\mathcal{P}_\Omega)$ corresponds to a set of matrices vanishing at $\Omega^\perp$.

Homomorphic matrix completion problem on cyphertexts. In cloud computing, the homomorphism property allows computations to be carried out on cyphertexts, generating an encrypted result which, when decrypted, matches the result of operations performed on the plaintext. Following this paradigm, we define a homomorphic matrix completion problem that ensures data privacy. As shown in Fig. 1, this novel framework consists of three main steps:

1) each node locally encrypts as $\mathcal{P}_{\Omega_j}(\overline{\boldsymbol{M}}_j) = \mathcal{P}_{\Omega_j}(g(\boldsymbol{M}_j))$ with its private keys, $j = 1, ..., n_2$, and uploads $\mathcal{P}_{\Omega_j}(\overline{\boldsymbol{M}}_j)$ to a cloud server that later forms an incomplete matrix $\mathcal{P}_{\Omega}(\overline{\boldsymbol{M}}) \in \mathbb{R}^{n_1 \times n_2}$;
2) the cloud server solves the following matrix completion problem given $\mathcal{P}_{\Omega}(\overline{\boldsymbol{M}})$ and sends back the recovered vector $\widehat{\overline{\boldsymbol{M}}}_j$ to the j-th node, $j = 1, ..., n_2$,
$$\text{Find a matrix } \overline{\boldsymbol{X}} \in \mathbb{R}^{n_1 \times n_2}, \text{ s.t. } \mathcal{P}_{\Omega}(\overline{\boldsymbol{X}}) = \mathcal{P}_{\Omega}(\overline{\boldsymbol{M}}), \text{ rank}(\overline{\boldsymbol{X}}) \leq \overline{r}, \tag{7}$$
where $\overline{r} = \text{rank}(\overline{\boldsymbol{M}})$ may be slightly bigger than r due to by the encryption scheme $g(\cdot)$.
3) each node locally decrypts its own vector using private keys, i.e., $\widehat{\boldsymbol{M}}_j = g^{-1}(\widehat{\overline{\boldsymbol{M}}}_j)$, $j = 1, ..., n_2$.

2.3 Notions of Privacy

We introduce a new variant of differential privacy for low-rank matrices.

2.3.1 Differential Privacy

Let $D = \{d_1, ..., d_n\}$ be a dataset of n entries and $\mathcal{T}$ be a fixed domain, where each entry $d_j \in \mathcal{T}$ encodes potentially sensitive information about node j. Let $\mathcal{A} : \mathcal{T}^n \to \mathcal{O}^n$ be an algorithm that operates on dataset D and produces n outputs, one for each node j and from a set of possible output $\mathcal{O}$. Let D_{-j} denote the dataset D without the entry of the j-th node, and similarly $\mathcal{A}_{-j}(D)$ denote the set of outputs without the output for the j-th node. Let $(d_j; D_{-j})$ denote the dataset obtained by adding a data entry d_j to the dataset D_{-j}.

The (ϵ, δ)-differential privacy and joint (ϵ, δ)-differential privacy [19] are given in the following.

Definition 2. *((ϵ, δ)-differential privacy [10]). An algorithm $\mathcal{A}$ satisfies (ϵ, δ)-differential privacy if for any node j, any two possible values of data entry $d_j, d'_j \in \mathcal{T}$ for node j, any tuple of data entries for all other nodes $D_{-j} \in \mathcal{T}^{n-1}$, and any output set $O \subseteq \mathcal{O}^n$, we have*
$$\mathbb{P}_{\mathcal{A}}[\mathcal{A}(d_j; D_{-j}) \in O] \leq e^{\epsilon} \cdot \mathbb{P}_{\mathcal{A}}[\mathcal{A}(d'_j; D_{-j}) \in O] + \delta. \tag{8}$$

Definition 3. *(Joint (ϵ, δ)-differential privacy [19]). An algorithm $\mathcal{A}$ satisfies (ϵ, δ)-joint differential privacy if for any node j, any two possible values of data entry $d_j, d'_j \in \mathcal{T}$ for node j, any tuple of data entries for all other nodes $D_{-j} \in \mathcal{T}^{n-1}$, and any output set $O \subseteq \mathcal{O}^{n-1}$, we have*
$$\mathbb{P}_{\mathcal{A}}[\mathcal{A}_{-j}(d_j; D_{-j}) \in O] \leq e^{\epsilon} \cdot \mathbb{P}_{\mathcal{A}}[\mathcal{A}_{-j}(d'_j; D_{-j}) \in O] + \delta. \tag{9}$$

Intuitively, an algorithm $\mathcal{A}$ satisfies (ϵ, δ)-differential privacy if for any node j and dataset D, $\mathcal{A}(D)$ and D_{-j} do not reveal "much" information about d_j. For low-rank matrices, [18] used a relaxed notion *joint (ϵ, δ)-differential privacy*: an algorithm $\mathcal{A}$ satisfies joint (ϵ, δ)-differential privacy if for any node j and dataset D, $\mathcal{A}_{-j}(D)$ (the output for the other $n-1$ nodes) and D_{-j} (data entries of the other $n-1$ nodes) do not reveal "much" information about d_j. Relaxing (ϵ, δ)-differential privacy to joint (ϵ, δ)-differential privacy is reasonable for the matrix completion problem since the j-th column for the j-th node can reveal a lot of information about d_j.

2.3.2 Differential Privacy for Low-rank Matrix Completion

We would like to point out that joint (ϵ, δ)-differential privacy in Def. 3 can be further refined. For a low-rank matrix $\boldsymbol{M}$, its column subspace $\mathcal{S}(\boldsymbol{M})$ is *global information*, which is shared across all n_2 nodes and can be easily inferred from $\mathcal{A}_{-j}(D)$ and D_{-j}. Note that the differential privacy notion aims to protect individual information, rather than global information. We adapt it to low-rank matrices by excluding the shared column subspace.

Low-rank matrices have linearly dependent columns, and this dependency is reflected in the fact that they share a common column subspace. Formally, a rank-r matrix $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\top}$ can be expressed

Algorithm 1 Homomorphic matrix completion at the cloud server

Input: parameters n_1, n_2, r, k.

Output: matrix $\boldsymbol{K} \in \mathbb{R}^{n_1 \times k}$ as public keys, the recovered matrix $\overline{\widehat{\boldsymbol{X}}} \in \mathbb{R}^{n_1 \times n_2}$ (cyphertexts).

1: Generate a random matrix $\boldsymbol{K} \in \mathbb{R}^{n_1 \times k}$ and broadcast $\boldsymbol{K}$ to all n_2 nodes;
2: **until** received all n_2 encrypted vectors $\mathcal{P}_{\Omega_j}(\overline{\boldsymbol{M}}_j)$ (line 4 in Alg. 2) **do**
3: Carry out a matrix completion task in (7) and obtain $\overline{\widehat{\boldsymbol{X}}} \in \mathbb{R}^{n_1 \times n_2}$;
4: Send the recovered vector $\overline{\widehat{\boldsymbol{X}}}_j \in \mathbb{R}^{n_1}$ back to the j-th node, $j = 1, ..., n_2$.
5: **end**

Algorithm 2 Homomorphic matrix completion at node j, for $j = 1, ..., n_2$

Input: an incomplete vector $\mathcal{P}_{\Omega_j}(\boldsymbol{M}_j)$, observation set Ω_j, and parameters n_1, r, k.

Output: an recovered vector $\widehat{\boldsymbol{X}}_j$ (plaintexts).

1: **until** received $\boldsymbol{K} \in \mathbb{R}^{n_1 \times k}$ from the server (line 1 in Alg. 1) **do**
2: Generate k random numbers $\boldsymbol{R}_j \overset{\text{i.i.d}}{\sim} \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_k)$;
3: Perform local encryption as $\mathcal{P}_{\Omega_j}(\overline{\boldsymbol{M}}_j) = \mathcal{P}_{\Omega_j}(\boldsymbol{M}_j) + \mathcal{P}_{\Omega_j}(\boldsymbol{K}\boldsymbol{R}_j)$;
4: Upload $\mathcal{P}_{\Omega_j}(\overline{\boldsymbol{M}}_j)$ to the cloud server;
5: **end**
6: **until** received the recovered vector $\overline{\widehat{\boldsymbol{X}}}_j$ from the cloud server (line 4 in Alg. 1) **do**
7: Using $\boldsymbol{R}_j$ and $\boldsymbol{K}$, decrypt $\overline{\widehat{\boldsymbol{X}}}_j$ to obtain $\widehat{\boldsymbol{X}}_j$, i.e., $\widehat{\boldsymbol{X}}_j = \overline{\widehat{\boldsymbol{X}}}_j - \boldsymbol{K}\boldsymbol{R}_j$.
8: **end**

as $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{C}$ where $\boldsymbol{U} \in \mathbb{R}^{n_1 \times r}$ and $\boldsymbol{C} = \boldsymbol{\Sigma}\boldsymbol{V}^\top \in \mathbb{R}^{r \times n_2}$; alternatively, a column can be expressed as $\boldsymbol{M}_j = \boldsymbol{U}\boldsymbol{C}_j$, for $j = 1, ..., n_2$, where $\boldsymbol{C}_j$ is the coefficient vector (individual information) of the j-th node in the column subspace with basis $\boldsymbol{U}$ (global information).

The following subspace-aware joint (ϵ, δ)-differential privacy considers the coefficient vectors $\boldsymbol{C}_j$ for $j = 1, ..., n_2$, i.e., D in Def. 3 corresponds to the coefficient matrix $\boldsymbol{C} \in \mathbb{R}^{r \times n_2}$.

Definition 4. *(Subspace-aware joint (ϵ, δ)-differential privacy). Assume n_2 nodes' data vector form a rank-r matrix $\boldsymbol{M} \in \mathbb{R}^{n_1 \times n_2}$ with $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\top = \boldsymbol{U}\boldsymbol{C}$ where $\boldsymbol{U} \in \mathbb{R}^{n_1 \times r}$ and $\boldsymbol{C} = \boldsymbol{S}\boldsymbol{V}^\top \in \mathbb{R}^{r \times n_2}$. A matrix completion algorithm $\mathcal{A}$ satisfies subspace-aware (ϵ, δ)-joint differential privacy if for any node j, any two possible coefficient vectors $\boldsymbol{C}_j, \boldsymbol{C}'_j \in \mathbb{R}^r$ for node j, any tuple of coefficient vectors for all other nodes $\boldsymbol{C}_{-j} \in \mathbb{R}^{r \times (n_2 - 1)}$, and any output set $O \subseteq \mathbb{R}^{r \times n_2}$ that consists of estimated coefficient vectors in a column subspace with basis $\boldsymbol{U}$, we have*

$$\mathbb{P}_{\mathcal{A}}[\mathcal{A}_{-j}(\boldsymbol{C}_j; \boldsymbol{C}_{-j} | \boldsymbol{U}) \in O] \leq e^{\epsilon} \cdot \mathbb{P}_{\mathcal{A}}[\mathcal{A}_{-j}(\boldsymbol{C}'_j; \boldsymbol{C}_{-j} | \boldsymbol{U}) \in O] + \delta. \tag{10}$$

3 Novel Homomorphic Framework for Matrix Completion

We propose a homomorphic encryption-decryption scheme: a node performs local encryption and decryption, and uploads an encrypted vector to a server to perform the matrix completion computation.

3.1 Our Idea: Hiding a Low-rank Data Matrix in a Larger Subspace

To preserve the privacy of a low-rank data matrix $\boldsymbol{M} \in \mathbb{R}^{n_1 \times n_2}$ with rank r, our idea is to hide $\boldsymbol{M}$ (lies in an r-dimensional subspace) into a larger subspace of dimension $\overline{r}$, such that $\overline{r} \geq r$ and $r, \overline{r} \ll n_1$. A sound approach would be enlarging the original subspace of the data matrix (i.e., the plaintext) as follows: a cloud server generates a random matrix $\boldsymbol{K} \in \mathbb{R}^{n_1 \times k}$ as public keys, $k \ll n_1$, and broadcasts $\boldsymbol{K}$ to all n_2 nodes; then, node j generates k random numbers as private keys $\boldsymbol{R}_j \in \mathbb{R}^k$, and encrypts its vector $\boldsymbol{M}_j \in \mathbb{R}^{n_1}$ as follows (a version with missing entries will be given in (12))

$$\overline{\boldsymbol{M}}_j = \boldsymbol{M}_j + \boldsymbol{K}\boldsymbol{R}_j, \ j = 1, ..., n_2; \quad \text{Equivalently, } \overline{\boldsymbol{M}} = \boldsymbol{M} + \boldsymbol{R}\boldsymbol{R}. \tag{11}$$

In the encryption scheme (11), $\boldsymbol{M}$ is added up with $\boldsymbol{K}\boldsymbol{R}$, resulting in a matrix $\overline{\boldsymbol{M}}$ with rank $\overline{r} \leq r + k$. Since $\overline{r} \ll n_1$, $\overline{\boldsymbol{M}}$ is also low-rank, it is possible to recover $\overline{\boldsymbol{M}}$ from a subset of entries.

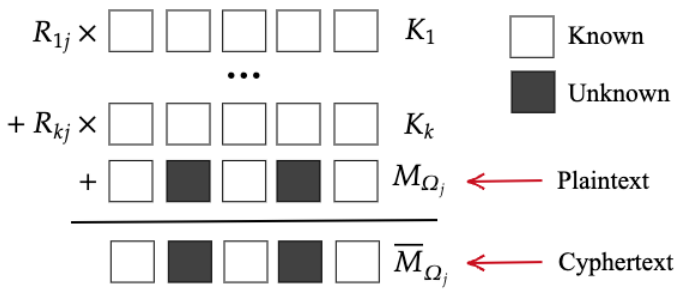

Figure 2: Our encryption method. The sets of missing entries are the same for plaintext and cyphertext.

3.2 Proposed Homomorphic Encryption-Decryption Scheme

We propose a homomorphic encryption-decryption scheme that consists of the following steps, while the pseudocodes are summarzied in Alg. 1 and Alg. 2.

- First, in line 1 of Alg. 1, the cloud server generates a random matrix $\boldsymbol{K} \in \mathbb{R}^{n_1 \times k}$ as public keys, then broadcasts $\boldsymbol{K}$ to all n_2 nodes.
- Second, in lines 1-5 of Alg. 2, after receiving $\boldsymbol{K} \in \mathbb{R}^{n_1 \times k}$ from the server (line 1 in Alg. 1), the j-th node locally carries out an encryption with k private keys (i.e., $\boldsymbol{R}_j \in \mathbb{R}^k$). As shown in Fig. 2, the j-th node locally encrypts its incomplete vector $\mathcal{P}_{\Omega_j}(\boldsymbol{M}_j)$ as follows
$$\mathcal{P}_{\Omega_j}(\overline{\boldsymbol{M}}_j) = \mathcal{P}_{\Omega_j}(\boldsymbol{M}_j) + \mathcal{P}_{\Omega_j}(\boldsymbol{K}\boldsymbol{R}_j),\ j = 1, ..., n_2, \tag{12}$$
where $\boldsymbol{R}_j \overset{\text{i.i.d}}{\sim} \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_k)$, $\mathcal{P}_{\Omega_j}(\boldsymbol{K}\boldsymbol{R}_j)$ means keeping the entries in Ω_j and setting the entries in the complement set of Ω_j to be zeros, thus $\mathcal{P}_{\Omega_j}(\overline{\boldsymbol{M}}_j)$ has the same set of missing entries as $\mathcal{P}_{\Omega_j}(\boldsymbol{M}_j)$. Note that $\boldsymbol{R}_j$ is stored locally, which are private keys that will NOT be shared with any other node. Then, each node uploads its encrypted vector $\mathcal{P}_{\Omega_j}(\overline{\boldsymbol{M}}_j)$ to the cloud server.
- Third, in lines 2-5 of Alg. 1, after receiving all n_2 encrypted vectors $\mathcal{P}_{\Omega_j}(\overline{\boldsymbol{M}}_j)$, $j = 1, ..., n_2$, the server forms an incomplete matrix $\overline{\boldsymbol{M}}_\Omega$ with $\Omega = \bigcup_{j=1,...,n_2} \Omega_j$. Then, the server carries out a matrix completion task in (7), and sends the recovered vector $\overline{\widehat{\boldsymbol{X}}}_j$ back to the j-th node, $j = 1, ..., n_2$.
- Finally, in lines 11-13 of Alg. 2, using the locally stored private keys $\boldsymbol{R}_j$, and the public keys $\boldsymbol{K}$, the j-th node decrypts its own vector, i.e., $\widehat{\boldsymbol{X}}_j = g^{-1}(\overline{\widehat{\boldsymbol{X}}}_j) = \overline{\widehat{\boldsymbol{X}}}_j - \boldsymbol{K}\boldsymbol{R}_j$, $j = 1, ..., n_2$.

4 Homomorphism Property Holds at Price of More Samples

We prove that the homomorphism property holds for the proposed scheme, which guarantees exact recovery on cyphertexts at a cost of more samples. The detailed proofs are given in Appx. B.

Overview: Starting from a necessary and sufficient condition in Lemma 1, we relax to a sufficient condition in Lemma 3 for the homomorphism property to hold. Then, we provide a homomorphic version of *Rudelson Selection Estimation Theorem* in Theorem 2 that guarantees Lemma 3 with high probability. Therefore, we obtain a sample complexity for EXACT recovery in Theorem 3, where our interesting finding is that *the homomorphism property holds at price of more samples*.

4.1 Sufficient Condition for Low-rank Matrix Completion

We start from a necessary and sufficient condition for low-rank matrix completion. Note that a similar necessary and sufficient condition for sparse vector recovery is discussed in compressive sensing [3, 7, 44]. Here, we apply a similar argument to obtain Lemma 1.

We define a set of matrices with rank at most r and a rank-descent cone as follows
$$\begin{cases} \mathcal{M} = \{\boldsymbol{X} \in \mathbb{R}^{n_1 \times n_2} : \text{rank}(\boldsymbol{X}) \leq r\}, \\ \mathcal{D}_{\mathcal{M}}(\boldsymbol{M}) = \{t(\boldsymbol{X} - \boldsymbol{M}) \in \mathbb{R}^{n_1 \times n_2} : \text{rank}(\boldsymbol{X}) \leq r,\ t \geq 0\}, \end{cases} \tag{13}$$

where $\mathcal{M}$ is the closure of the manifold of rank-r matrices. Accordingly, for $\overline{\boldsymbol{M}}$, we have

$$\begin{cases} \overline{\mathcal{M}} = \{\boldsymbol{X} \in \mathbb{R}^{n_1 \times n_2} : \text{rank}(\boldsymbol{X}) \leq \overline{r}\}, \\ \mathcal{D}_{\overline{\mathcal{M}}}(\overline{\boldsymbol{M}}) = \{t(\boldsymbol{X} - \overline{\boldsymbol{M}}) \in \mathbb{R}^{n_1 \times n_2} : \text{rank}(\boldsymbol{X}) \leq \overline{r},\ t \geq 0\}. \end{cases} \tag{14}$$

Lemma 1. *(**Necessary and sufficient condition for low-rank matrix completion**) $\boldsymbol{M}$ is the unique optimal solution to (6) if and only if $\Omega^{\perp} \cap \mathcal{D}_{\mathcal{M}}(\boldsymbol{M}) = \{\boldsymbol{0}\}$, where $\Omega^{\perp}$ denotes **Ker**$(\mathcal{P}_{\Omega})$.*

Geometric interpretation: $\boldsymbol{M}$ is the unique optimal solution to problem (6) if and only if starting from $\boldsymbol{M}$, the rank of $\boldsymbol{M} + \boldsymbol{D}$ increases for all directions $\boldsymbol{D} \in \Omega^{\perp}$, where $\boldsymbol{D}$ is nonzero. Therefore, the homomorphism property of low-rank matrix completion in problem (7) holds if

$$\Omega^{\perp} \cap \mathcal{D}_{\mathcal{M}}(\boldsymbol{M}) = \{\boldsymbol{0}\} = \Omega^{\perp} \cap \mathcal{D}_{\overline{\mathcal{M}}}(\overline{\boldsymbol{M}}). \tag{15}$$

Lemma 2. *([13, 6] (Theorem 6.1)) Let $\boldsymbol{M} = \boldsymbol{U}\Sigma\boldsymbol{V}^{\top}$ be the compact SVD of matrix $\boldsymbol{M}$. The tangent cone $T_{\mathcal{M}}(\boldsymbol{M})$ of the set $\mathcal{M}$ at $\boldsymbol{M}$ is a linear subspace given by*

$$T_{\mathcal{M}}(\boldsymbol{M}) = \{\boldsymbol{U}\boldsymbol{A}^{\top} + \boldsymbol{B}\boldsymbol{V}^{\top} \mid \boldsymbol{A} \in \mathbb{R}^{n_1 \times r}, \boldsymbol{B} \in \mathbb{R}^{n_2 \times r}\} \triangleq T, \tag{16}$$

and its complementary space is denoted by $T^{\perp}$.

Since the rank-descent cone is a subset of the tangent cone defined in (16) ([15], Theorem 4.8), $\mathcal{D}_{\mathcal{M}}(\boldsymbol{M}) \subseteq T$, and $\mathcal{D}_{\overline{\mathcal{M}}}(\overline{\boldsymbol{M}}) \subseteq \overline{T}$, we relax (15) to the following sufficient condition.

Lemma 3. *A sufficient condition for the homomorphic property of matrix completion under the proposed scheme in Alg. 1 and Alg. 2 is $\Omega^{\perp} \cap \overline{T} = \{\boldsymbol{0}\}$.*

Interpretation: if $\Omega^{\perp} \cap \overline{T} = \{\boldsymbol{0}\}$ holds, then we know that $\overline{\boldsymbol{M}} = \boldsymbol{M} + \boldsymbol{K}\boldsymbol{R}$ is the unique optimal solution to problem (7) and $\boldsymbol{M}$ is the unique optimal solution to problem (6). Since $\overline{\boldsymbol{M}} = \boldsymbol{M} + \boldsymbol{K}\boldsymbol{R}$ is a *one-to-one mapping*, a decryption scheme $\overline{\boldsymbol{M}} - \boldsymbol{K}\boldsymbol{R}$ will return the desired true matrix $\boldsymbol{M}$.

4.2 Homomorphic Version of Rudelson Selection Estimation Theorem

The Rudelson selection estimation theorem [37] investigates the number of random points needed to bring a convex body into a nearly isotropic position. Such an approximate isometry property is fundamentally useful to characterize the number of entries needed to complete a low-rank matrix.

Definition 5. *(Coherence) Let $\boldsymbol{U} \in \mathbb{R}^{n \times r}$ be the r left singular vectors of $\boldsymbol{M}$ (corresponds to the column subspace $\mathcal{S}(\boldsymbol{M})$) and $\boldsymbol{P}_{\boldsymbol{U}}$ be the orthogonal projection onto $\boldsymbol{U}$. Then the coherence of $\boldsymbol{U}$ (or $\mathcal{S}(\boldsymbol{M})$, respectively) is defined as*

$$\mu(\mathcal{S}(\boldsymbol{M})) = \mu(\boldsymbol{U}) \triangleq \frac{n}{r} \max_{1 \leq i \leq n} ||\boldsymbol{P}_{\boldsymbol{U}}\boldsymbol{e}_i||_2^2 = \frac{n}{r} \max_{1 \leq i \leq n} ||\boldsymbol{U}^{\top}\boldsymbol{e}_i||_2^2, \tag{17}$$

since $(\boldsymbol{U}^{\top}\boldsymbol{U})^{-1} = \boldsymbol{I}$ and $\boldsymbol{U}$ is orthonormal.

The concept "coherence" measures the relationship between a low-dimensional space and the observation operator $\mathcal{P}_{\Omega}$, namely the cosine (with a scaling factor $\frac{n}{r}$) of the principal angle between the low-dimensional space and a standard basis. $\boldsymbol{M}$ is said to satisfy the *standard incoherence* condition with parameter μ_0 if

$$\mu(\boldsymbol{U}) \leq \mu_0, \quad \text{and} \quad \mu(\boldsymbol{V}) \leq \mu_0. \tag{18}$$

A small μ_0 ensures that the information of the row/column spaces of $\boldsymbol{M}$ is not too concentrated on a small number of rows/columns. It characterizes the contribution of an entry in recovering $\boldsymbol{M}$: a small μ_0 means that each entry provides approximated the same amount of information.

Theorem 1. *(**Rudelson selection estimation theorem** [3]) Assume that $\Omega \sim$ **Ber**(p) with $p = \Theta(\frac{m}{n_1 n_2})$, and $\boldsymbol{M}$ obeys the standard incoherence condition (18) with parameter μ_0. There is a constant C_R such that for $\beta > 1$, with probability $\geq 1 - 3n_2^{-\beta}$, we have*

$$||p^{-1}\mathcal{P}_T\mathcal{P}_{\Omega}\mathcal{P}_T - \mathcal{P}_T|| \leq C_R\sqrt{\frac{\mu_0 n_2 r(\beta \log n_2)}{m}} \triangleq \epsilon < 1. \tag{19}$$

We derive the following homomorphic variant of the Rudelson selection estimation theorem [37] and will use it to guarantee Lemma 3. Our new contribution here is to derive the conditions when the approximate isometry property will hold simultaneously for both cyphertexts and plaintexts.

Theorem 2. ***(Homomorphic version of Rudelson selection estimation theorem)*** *Assume that $\Omega \sim$ **Ber**(p) with $p = \Theta(\frac{m}{n_1 n_2})$, $\boldsymbol{M}$ and $\overline{\boldsymbol{M}}$ satisfy the standard incoherence condition (18) with parameter μ_0 and $\overline{\mu}_0$, respectively. Under the proposed scheme in Alg. 1 and Alg. 2, there are constants C_R, C'_R such that for $\beta > 1$, with probability $\geq 1 - 3n_2^{-\beta}$, we have*

$$\begin{aligned} &(\textit{cyphertext})\ \ ||p^{-1}\mathcal{P}_{\overline{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\overline{T}} - \mathcal{P}_{\overline{T}}|| \leq C'_R\sqrt{\frac{n_2\overline{\mu}_0\overline{r}(\beta\log n_2)}{m}} \triangleq \epsilon' < 1,\ \textit{which implies} \\ &(\textit{plaintext})\ \ ||p^{-1}\mathcal{P}_{T}\mathcal{P}_{\Omega}\mathcal{P}_{T} - \mathcal{P}_{T}|| \leq C_R\sqrt{\frac{n_2\mu_0 r(\beta\log n_2)}{m}} \triangleq \epsilon < 1. \end{aligned} \tag{20}$$

Note that $||p^{-1}\mathcal{P}_{\overline{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\overline{T}} - \mathcal{P}_{\overline{T}}|| < 1$ implies that the sufficient condition $\Omega^{\perp} \cap \overline{T} = \{\mathbf{0}\}$ holds.

4.3 Sample Complexity for EXACT Recovery

Then, we prove Theorem 3 that the homormophism property holds for the proposed scheme in Alg. 1 and Alg. 2, provided that there are sufficient number of observations ($|\Omega|$ is large enough).

Theorem 3. *For Alg. 1 and Alg. 2, with probability $\geq 1 - 3n_2^{-\beta}$, the homomorphism property holds if $p \geq \frac{C_0\overline{\mu}_0\overline{r}(\beta\log n_2)}{n_1}$, where C_0 is a positive constant.*

Next, we characterize the coherence change of $\overline{\mu}_0$ and provide the sample complexity for the EXACT recovery in Alg. 1 and Alg. 2.

Lemma 4. *The new coherence under the proposed scheme in Alg. 1 and Alg. 2 satisfies*

$$\overline{\mu}_0 \leq \frac{r}{\overline{r}}\mu_0 + C\max(\frac{k}{\overline{r}}, \frac{\log n_2}{\overline{r}}),\ \ \textit{with probability at least}\ 1 - cn_2^{-3}\log n_2. \tag{21}$$

Combining Theorem 3 and Lemma 4, we characterize the required number of entries.

Corollary 1. *For Alg. 1 and Alg. 2, with probability $\geq 1 - 6n_2^{-\beta} - cn_2^{-3}\log n_2$, the homomorphism property holds if $p \geq \frac{C_0(r\mu_0 + C\max(k,\log n_2))(\beta\log n_2)}{n_1}$, where c, C_0, C are positive constants.*

5 Differential Privacy Property Holds

In this section, we show that the differential privacy holds for the proposed scheme in Alg. 1 and Alg. 2. It is well-known that one can achieve (ϵ, δ)-differential privacy by adding Gaussian noise.

Definition 6. ***(Privacy loss as a random variable** [10]) Considering a mechanism $\mathcal{A}$ on a pair of databases D, D'. For an outcome $o \in \mathcal{O}$, the privacy loss on o is defined as the logarithmic ratio between the probability to observe o on input D compared to that on input D':*

$$\mathcal{L}^{(o)}_{\mathcal{A}(D)||\mathcal{A}(D')} = \ln\frac{\mathbb{P}(\mathcal{A}(D) = o)}{\mathbb{P}(\mathcal{A}(D') = o)}, \tag{22}$$

where $\mathbb{P}(\mathcal{A}(D) = o)$ is a probability density over a continuous set $\mathcal{O}$.

Two potential issues of the proposed scheme in Alg. 1 and Alg. 2 is the projection recovery and the rank value r may be unknown. Namely, one can do a corresponding projection to obtain the real data after a single encryption. Therefore, we execute the proposed scheme twice and introduce two parameters σ_1 and σ_2:

- First-round encryption: the server randomly generats a matrix $\boldsymbol{K}^1 \in \mathbb{R}^{n_1\times k}$ and each node generates k random numbers $\boldsymbol{R}^1_j \overset{\text{i.i.d}}{\sim} \mathcal{N}(\mathbf{0}, \sigma_1^2\boldsymbol{I}_k)$. Then, we have $\mathcal{P}_{\Omega}(\overline{\boldsymbol{M}}) = \mathcal{P}_{\Omega}(\boldsymbol{M}) + \mathcal{P}_{\Omega}(\boldsymbol{K}^1\boldsymbol{R}^1)$.
- Second-round encryption: the server obtains the column space of $\overline{\boldsymbol{M}}$ as $\boldsymbol{K}^2 \in \mathbb{R}^{n_1\times\overline{r}}$ with $\overline{r} = r + k$ and then each node generates $r + k$ random numbers $\boldsymbol{R}^2_j \overset{\text{i.i.d}}{\sim} \mathcal{N}(\mathbf{0}, \sigma_2^2\boldsymbol{I}_{r+k})$. Then we have $\mathcal{P}_{\Omega}(\overline{\overline{\boldsymbol{M}}}) = \mathcal{P}_{\Omega}(\overline{\boldsymbol{M}}) + \mathcal{P}_{\Omega}(\boldsymbol{K}^2\boldsymbol{R}^2)$.

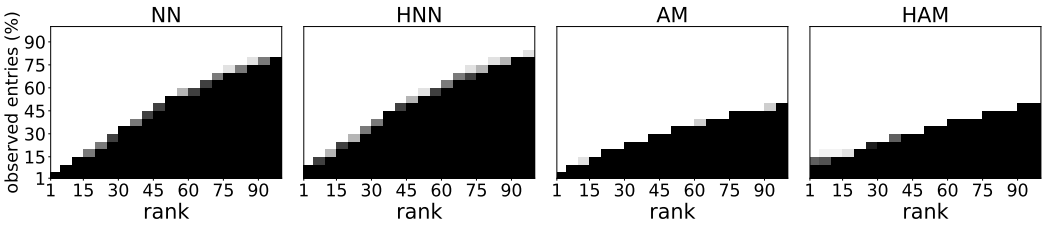

Figure 3: Comparing NN and AM algorithms with their homomorphic versions. The figure plots the success rates within 10 trials, where the white and black cells mean "success" and "fail". The trial is "success" if RSE $\leq 10^{-5}$. We set $k = 10$ in Alg. 1 and Alg. 2

Theorem 4 states that the proposed scheme satisfies the subspace-aware joint (ϵ, δ)-differential privacy in Section 2.3.2. The detailed proofs are given in Appx. C, where the key is to quantify σ under which the random variable privacy loss in (22) is bounded by ϵ, with probability at least $1 - \delta$.

Theorem 4. *Let $\epsilon \in (0, 1)$ and $c^2 > 2\ln(1.25/\delta)$. Assume the true data matrix $\boldsymbol{M} \in \mathbb{R}^{n_1 \times n_2}$ has rank r and each column has bounded ℓ_2-norm, i.e., $\Delta = \max_{1 \leq j \leq n_2} ||\boldsymbol{M}_j||_2 \leq L$. Let $\boldsymbol{R}_j^1 \sim \mathcal{N}(\boldsymbol{0}, \sigma_1^2 \boldsymbol{I}_k)$ with $\sigma_1 \geq 2cL\sqrt{2\ln(2/\delta)}/\epsilon$ and $\boldsymbol{R}_j^2 \sim \mathcal{N}(\boldsymbol{0}, \sigma_2^2 \boldsymbol{I}_{k+r})$ with $\sigma_2 \geq 2cL\sqrt{2\ln(2/\delta)}/\epsilon$, then the encryption and decryption scheme in Alg. 1 and Alg. 2, satisfies the subspace-aware joint (ϵ, δ)-differential privacy property.*

A substantial improvement is: for the same level of privacy (the same ϵ, δ parameter in the above joint (ϵ, δ)-DP property), our algorithms are able to achieve EXACT recovery. Note that by proving the homomorphism property and characterising the sample complexity, we reduce the error bound $O(\sqrt[10]{n_1^3 n_2})$ from [18] to ZERO since we have EXACT recovery.

6 Performance Evaluation

We evaluate the proposed scheme on synthetic data and real-world datasets using two matrix completion algorithms [39, 16], verifying the homomorphism property of the proposed scheme.

6.1 Experimental Settings

Datasets. We experiment with synthetic data and real-world datasets. The synthetic data is generated randomly according to the low-rank $1,000 \times 1,000$ matrix model and serves as well-controlled inputs for verification. The real-world datasets include two benchmark datasets for recommendation systems, namely the *MovieLens10M (Top 400)*[3] and *Netflix (Top 400)* datasets. The MovieLens dataset contains ratings of 400 most rated movies made by approximately $7,000$ users, and the Netflix dataset contains ratings of 400 most rated movies made by approximately 480 thousand users.

Matrix completion algorithms. For the matrix completion on the server, we use nuclear-norm minimization (NN) and alternating minimization (AM) algorithms. In Section 6.2, we compare both algorithms with their homomorphic versions. In Section 6.3, on the real-world datasets, we also include the private Frank-Wolf (FW) algorithm [18] for comparison.

Performance metric. We measure the recovery error via the relative square root error RSE $= \frac{||\widehat{\boldsymbol{M}} - \boldsymbol{M}||_F}{||\boldsymbol{M}||_F}$. All experiments are executed for ten times and we report the average results.

6.2 Results on Synthetic Data

We experiment with randomly generated low-rank matrices on NN and AM algorithms and their homomorphic versions HNN and HAM. We vary the rank r of the generated matrix and the percentage of observed entries from 1, 5, to 95. As shown in Fig. 6.2, we observe two trends: 1) for a certain rank r, the success rate increases as the percentage of observed entries increases; and 2) for a certain percentage of observed entries, the success rate decreases as the rank r increases. On the other hand, we find that the HNN and HAM need slightly more observed entries to reach the success threshold,

[3] https://movielens.org/

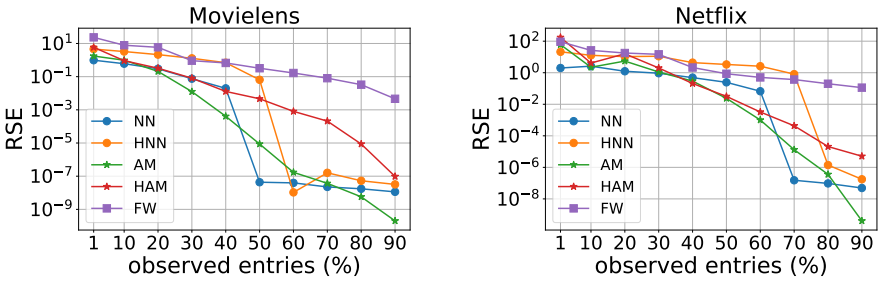

Figure 4: Results on MovieLens10M and Netflix datasets. We vary the percentage of observed entries and measure the RSE recovery error.

which verifies Theorem 3 that the scheme guarantees exact recovery at a cost of more samples. As an interpretation, the homomorphic version is to hide the plaintext matrix into a larger space, namely from rank r to rank $r + k$. In this case, given that we set $k = 10$ for the experiments, we find that the results of HNN and HAM can be obtained by shifting the results of their counterparts left one grid.

6.3 Results on MovieLens10M and Netflix Datasets

Fig. 4 shows the results on MovieLens10M and Neflix datasets. For the newly introduced compared algorithm FW, we set the privacy parameter $\epsilon = 2\log(1/\delta)$ and $\delta = 10^{-6}$. For the NN and AM algorithms, the setting is the same in Section 6.2.

First of all, we observe that the homomorphic algorithms can achieve significantly lower recovery errors than the error of FW algorithm. This points out the difference between the proposed scheme and existing strategies, in which we do not sacrifice the recovery error to improve the privacy. On the other hand, we find that the homomorphic algorithms can reach the same level of recovery error as the vanilla algorithms on plaintexts, but need more samples. Such a performance is consistent with our theoretical proofs and our observations in Section 6.2. Moreover, we analyze the impact of increasing the percentage of observed entries on three types of algorithms, as shown in Fig. 4. For AM and FW algorithms, the recovery error decreases smoothly as the percentage increases (note that the y-axis decreasing in $\log$). However, the NN algorithm demonstrates a significant error drop as we increase the percentage of observed entries.

7 Conclusion and Future Work

This work studied the problem of privacy-preserving data completion in a distributed manner. To address the privacy concern, we define the homomorphic matrix completion problem and propose a homomorphic encryption-decryption scheme. Unlike existing works that preserve privacy by sacrificing recovery accuracy, our work guarantees the EXACT recovery while making a tradeoff between privacy and the number of samples. Then, we theoretically prove that the proposed scheme satisfies the homomorphism and differential privacy properties. Experimentally, we show that the proposed scheme is compatible with two matrix completion algorithms, namely the nuclear norm minimization and alternating minimization, and verify the homomorphism property.

In the future, it would be interesting to extend this homomophic framework to the tensor completion problem [29, 30]. It would also be practically interesting to study federated learning application [22] and develop high-performance implementations for high-dimensional data analysis [48, 47, 34, 48].

Acknowledgement

Xiao-Yang Liu would like to thank Prof. John Wright (Department of Electrical Engineering at Columbia University) for his insightful sharing.

References

[1] J. Bennett and S. Lanning. The Netflix prize. In *Proceedings of KDD Cup and Workshop*, volume 2007, page 35. New York, NY, USA, 2007.

[2] R. S. Cabral, F. Torre, J. P. Costeira, and A. Bernardino. Matrix completion for multi-label image classification. In *Advances in Neural Information Processing Systems*, pages 190–198, 2011.

[3] E. J. Candès. Mathematics of sparsity (and a few other things). In *Proceedings of the International Congress of Mathematicians, Seoul, South Korea*, volume 123, 2014.

[4] E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717, 2009.

[5] 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.

[6] T. P. Cason, P.-A. Absil, and P. Van Dooren. Iterative methods for low rank approximation of graph similarity matrices. *Linear Algebra and its Applications*, 438(4):1863–1882, 2013.

[7] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky. The convex geometry of linear inverse problems. *Foundations of Computational Mathematics*, 12(6):805–849, 2012.

[8] Y. Chen. Incoherence-optimal matrix completion. *IEEE Transactions on Information Theory*, 61(5):2909–2923, 2015.

[9] 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*, pages 1877–1887. PMLR, 2021.

[10] C. Dwork and A. Roth. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014.

[11] A. Elmahdy, J. Ahn, C. Suh, and S. Mohajer. Matrix completion with hierarchical graph side information. *Advances in Neural Information Processing Systems*, 33:9061–9074, 2020.

[12] C. Gentry. Fully homomorphic encryption using ideal lattices. *ACM STOC*, 9:169–178, 2009.

[13] J. Harris. *Algebraic geometry: a first course*, volume 133. Springer Science & Business Media, 2013.

[14] Dat T. Huynh and Ehsan Elhamifar. Interactive multi-label cnn learning with partial labels. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9420–9429, 2020.

[15] J. Jahn. *Introduction to the theory of nonlinear optimization*. Springer Science & Business Media, 2007.

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

[17] P. Jain, J. Rush, A. Smith, S. Song, and A. Guha Thakurta. Differentially private model personalization. *Advances in Neural Information Processing Systems*, 34, 2021.

[18] P. Jain, O. D. Thakkar, and A. Thakurta. Differentially private matrix completion revisited. In *International Conference on Machine Learning*, pages 2220–2229, 2018.

[19] M. Kearns, M. Pai, A. Roth, and J. Ullman. Mechanism design in large games: Incentives and privacy. In *Proceedings of the Conference on Innovations in Theoretical Computer Science*, pages 403–410. ACM, 2014.

[20] R.H. Keshavan, A. Montanari, and S. Oh. Matrix completion from a few entries. *IEEE Transactions on Information Theory*, 56(6):2980–2998, 2010.

[21] L. Kong, L. He, X.-Y. Liu, Y. Gu, M.-Y. Wu, and X. Liu. Privacy-preserving compressive sensing for crowdsensing based trajectory recovery. In *IEEE 35th International Conference on Distributed Computing Systems (ICDCS)*, pages 31–40, 2015.

[22] L. Kong, X.-Y. Liu, H. Sheng, P. Zeng, and G. Chen. Federated tensor mining for secure industrial internet of things. *IEEE Transactions on Industrial Informatics*, 2019.

[23] L. Kong, M. Xia, X.-Y. Liu, M.-Y. Wu, and X. Liu. Data loss and reconstruction in sensor networks. In *Proceedings IEEE INFOCOM*, pages 1654–1662. IEEE, 2013.

[24] Y. Koren, S. Rendle, and R. Bell. Advances in collaborative filtering. *Recommender Systems Handbook*, pages 91–142, 2022.

[25] Kaustav Kundu and Joseph Tighe. Exploiting weakly supervised visual patterns to learn from partial annotations. *Advances in Neural Information Processing Systems*, 33:561–572, 2020.

[26] A. J. Laub. *Matrix analysis for scientists and engineers*. Siam, 2005.

[27] Z. Li, B. Ding, C. Zhang, N. Li, and J. Zhou. Federated matrix factorization with privacy guarantee. *Proceedings of the VLDB Endowment*, 15(4):900–913, 2021.

[28] G. Liu, Q. Liu, and X. Yuan. A new theory for matrix completion. *Advances in Neural Information Processing Systems*, 30, 2017.

[29] X.-Y. Liu, S. Aeron, V. Aggarwal, and X. Wang. Low-tubal-rank tensor completion using alternating minimization. *IEEE Transactions on Information Theory*, 66(3):1714–1737, 2019.

[30] X.-Y. Liu, S. Aeron, V. Aggarwal, X. Wang, and M.-Y. Wu. Adaptive sampling of rf fingerprints for fine-grained indoor localization. *IEEE Transactions on Mobile Computing*, 15(10):2411–2423, 2015.

[31] X.-Y. Liu and X. Wang. Ls-decomposition for robust recovery of sensory big data. *IEEE Transactions on Big Data*, 4(4):542–555, 2017.

[32] X.-Y. Liu, Y. Zhu, L. Kong, C. Liu, Y. Gu, A. V Vasilakos, and M.-Y. Wu. CDC: Compressive data collection for wireless sensor networks. *IEEE Transactions on Parallel and Distributed Systems*, 26(8):2188–2197, 2014.

[33] S. Lohr. Netflix cancels contest after concerns are raised about privacy. Web page: `http://www.nytimes.com/2010/03/13/technology/13netflix.html?mcubz=0`. *The New Yorks Times*, Mar. 12, 2010.

[34] H. Lu, T. Zhang, and X.-Y. Liu. High-performance homomorphic matrix completion on gpus. In *IEEE 21st International Conference on High Performance Computing and Communications*, pages 1627–1634. IEEE, 2019.

[35] Lester W Mackey, Ameet Talwalkar, and Michael I Jordan. Distributed matrix completion and robust factorization. *Journal of Machine Learning Research*, 16(1):913–960, 2015.

[36] S. Rallapalli, L. Qiu, Y. Zhang, and Y.-C. Chen. Exploiting temporal stability and low-rank structure for localization in mobile networks. In *Proceedings of the International Conference on Mobile Computing and Networking*, pages 161–172. ACM, 2010.

[37] M. Rudelson. Random vectors in the isotropic position. *Journal of Functional Analysis*, 164(1):60–72, 1999.

[38] R. Schneider and A. Uschmajew. Convergence results for projected line-search methods on varieties of low-rank matrices via łojasiewicz inequality. *SIAM Journal on Optimization*, 25(1):622–646, 2015.

[39] S. Shalev-Shwartz, A. Gonen, and O. Shamir. Large-scale convex minimization with a low-rank constraint. In *International Conference on Machine Learning*, 2011.

[40] V. Singhal and T. Steinke. Privately learning subspaces. *Advances in Neural Information Processing Systems*, 34, 2021.

[41] H. Sun and S.A. Jafar. The capacity of private information retrieval. *IEEE Transactions on Information Theory*, 2017.

[42] Jalaj Upadhyay. The price of privacy for low-rank factorization. *Advances in Neural Information Processing Systems*, 31, 2018.

[43] M. Van Dijk, C. Gentry, S. Halevi, and V. Vaikuntanathan. Fully homomorphic encryption over the integers. *Springer, Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 24–43, 2010.

[44] J. Wright and Y. Ma. *High-dimensional data analysis with low-dimensional models: Principles, computation, and applications*. Cambridge University Press, 2022.

[45] Q. Wu, H. Zhang, X. Gao, J. Yan, and H. Zha. Towards open-world recommendation: An inductive model-based collaborative filtering approach. In *International Conference on Machine Learning*, pages 11329–11339. PMLR, 2021.

[46] Q. Ye, J. Cheng, H. Du, X. Jia, and J. Zhang. A matrix-completion approach to mobile network localization. In *Proceedings of the 15th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pages 327–336. ACM, 2014.

[47] T. Zhang, X.-Y. Liu, X. Wang, and A. Walid. cuTensor-Tubal: Efficient primitives for tubal-rank tensor learning operations on gpus. *IEEE Transactions on Parallel and Distributed Systems*, 31(3):595–610, 2019.

[48] T. Zhang, H. Lu, and X.-Y. Liu. High-performance homomorphic matrix completion on multiple gpus. *IEEE Access*, 8:25395–25406, 2020.

[49] Y. Zhang, M. Roughan, W. Willinger, and L. Qiu. Spatio-temporal compressive sensing and internet traffic matrices. In *ACM SIGCOMM Computer Communication Review*, volume 39, pages 267–278. ACM, 2009.

Broader Impact Statement

This paper is within the area of private machine learning, which calls for privacy-preserving data completion by proposing a homomorphic encryption-decryption scheme. Due to the wide application areas of the matrix completion problem, this work may have broad practical impact in recommendation systems, global positioning, system identification and mobile social networks, etc.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes]
   (c) Did you discuss any potential negative societal impacts of your work? [N/A] Not aware of foresee negative impacts.
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes]
   (b) Did you include complete proofs of all theoretical results? [Yes]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes]
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes]
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

A Preliminaries

Notations: Lower case bold letters denote vectors, while capital bold face letters denote matrices, e.g., $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{X} \in \mathbb{R}^{n_1 \times n_2}$. Let $\boldsymbol{e}_i$ denote the i-th standard basis. For matrix $\boldsymbol{X}$, its (i,j)-th element is $\boldsymbol{X}_{ij}$ or $\boldsymbol{X}(i,j)$ and its j-th column is $\boldsymbol{X}_j$. The identity matrix $\boldsymbol{I}$ is a square matrix, sometimes we use $\boldsymbol{I}_n$ to specify its size $n \times n$, and the identity linear operator is denoted by $\mathcal{I}$. The transpose of a vector/matrix is indicated by a superscript $^\top$, e.g., $\boldsymbol{x}^\top$ and $\boldsymbol{X}^\top$. The concatenation of two matrices $\boldsymbol{A} \in \mathbb{R}^{n_1 \times n_2}$ and $\boldsymbol{B} \in \mathbb{R}^{n_1 \times n_3}$ is denoted by $[\boldsymbol{A},\ \boldsymbol{B}] \in \mathbb{R}^{n_1 \times (n_2+n_3)}$. By *with high probability* (w.h.p.) we mean that with probability at least $1 - c_1 n^{-c_2}$ for some constants $c_1, c_2 > 0$. Denote the Gaussian distribution by $\mathcal{N}(0, \sigma^2)$, with mean 0 and standard deviation σ. We use an overline to represent the encrypted version of a variable. Variables before encryption are called *plaintexts*, e.g., $\boldsymbol{X}$, while the encrypted variables are called *cyphertexts*, e.g., $\overline{\boldsymbol{X}}$.

We use an overline to represent the encrypted version of a variable on a cloud server. Variables before encryption are called *plaintexts*, e.g., $\boldsymbol{X}$, while the corresponding encrypted variables are called *cyphertexts*, e.g., $\overline{\boldsymbol{X}}$. Let set $\Omega \subseteq \{(1,1),(1,2),...,(n_1,n_2)\}$ index the observed entries. We represent the observed entries as $\boldsymbol{M}_\Omega$, while for analysis purpose we define a corresponding linear operator $\mathcal{P}_\Omega : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^{n_1 \times n_2}$ to represent the observation model as follows

$$(\boldsymbol{M}_\Omega)_{ij} = [\mathcal{P}_\Omega(\boldsymbol{M})]_{ij} = \begin{cases} \boldsymbol{M}_{ij}, \text{ if } (i,j) \in \Omega \\ 0, \text{ otherwise.} \end{cases} \tag{23}$$

We assume the true matrix $\boldsymbol{M}$ is low-rank, i.e., $\text{rank}(\boldsymbol{M}) = r \ll \min(n_1, n_2)$. The singular value decomposition (SVD) is $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\top$, where $\boldsymbol{U} \in \mathbb{R}^{n_1 \times r}$ denotes the r left singular vectors (corresponding to the column subspace), $\boldsymbol{V} \in \mathbb{R}^{n_2 \times r}$ denotes the r right singular vectors (corresponding to the row subspace), and $\boldsymbol{S} = \text{diag}(\sigma_i) \in \mathbb{R}^{r \times r}$ where σ_i is the i-th largest singular value and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq 0$. The nuclear norm of $\boldsymbol{M}$ is $||\boldsymbol{M}||_* = \sum_{i=1}^{r} \sigma_i$. The ℓ_2-norm of a vector is $||\boldsymbol{x}||_2$, while the Frobenius norm of a matrix is $||\boldsymbol{M}||_F = \sqrt{\sum_{i,j} |\boldsymbol{M}_{ij}|^2}$. The operator norm (spectral norm) of a matrix is equal to its first singular value, i.e., for $\boldsymbol{M} \in \mathbb{R}^{n_1 \times n_2}$,

$$||\boldsymbol{M}|| \triangleq \sup_{\boldsymbol{x} \in \mathbb{R}^{n_2},\ ||\boldsymbol{x}||_2 \leq 1} ||\boldsymbol{M}\boldsymbol{x}||_2 = \sigma_1(\boldsymbol{M}), \tag{24}$$

while more generally, the operator norm of a linear operator $\mathcal{L}$ is

$$||\mathcal{L}|| = \sup_{||\boldsymbol{X}||_F \leq 1} ||\mathcal{L}(\boldsymbol{X})||_F. \tag{25}$$

Definition 7. ***(Column subspace and null space** [26]) Let $\boldsymbol{A} \in \mathbb{R}^{n_1 \times n_2}$. The set*

$$\mathcal{S}(\boldsymbol{A}) = \{\boldsymbol{b} \in \mathbb{R}^{n_1} |\ \boldsymbol{b} = \boldsymbol{A}\boldsymbol{x},\ \boldsymbol{x} \in \mathbb{R}^{n_2}\} \tag{26}$$

is the column space or range of $\boldsymbol{A}$, and the set

$$\textbf{\textit{Ker}}(\boldsymbol{A}) = \{\boldsymbol{x} \in \mathbb{R}^{n_2} |\ \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}\} \tag{27}$$

is the kernel or (right) null space of $\boldsymbol{A}$.

Definition 8. ***(Row subspace and null space** [26]) Let $\boldsymbol{A} \in \mathbb{R}^{n_1 \times n_2}$. The set*

$$\mathcal{S}(\boldsymbol{A}^\top) = \{\boldsymbol{d} \in \mathbb{R}^{n_2} |\ \boldsymbol{d} = \boldsymbol{A}^\top \boldsymbol{y},\ \boldsymbol{y} \in \mathbb{R}^{n_1}\} \tag{28}$$

is the row space of $\boldsymbol{A}$, and the set

$$\textbf{\textit{Ker}}(\boldsymbol{A}^\top) = \{\boldsymbol{y} \in \mathbb{R}^{n_1} |\ \boldsymbol{A}^\top \boldsymbol{y} = \boldsymbol{0}\} \tag{29}$$

is the left null space of $\boldsymbol{A}$.

According to Definition 7, the kernel space of the linear operator $\mathcal{P}_\Omega$ is $\textbf{Ker}(\mathcal{P}_\Omega) = \{\boldsymbol{Z} \in \mathbb{R}^{n_1 \times n_2} \mid \mathcal{P}_\Omega(\boldsymbol{Z}) = \boldsymbol{0}\}$, which is denoted as $\Omega^\perp$ for simplicity. We adopt the notation $\Omega^\perp$ since $\textbf{Ker}(\mathcal{P}_\Omega)$ equals to a space on the complement set of Ω.

Let $\Omega \sim \textbf{Uni}(m)$ denote a set with m entries, which is sampled uniformly from all sets of m entries, and $\Omega \sim \textbf{Ber}(p)$ denote a set with $\mathbb{E}|\Omega| = m$ entries, each sampled independently according to the Bernoulli model in Definition 9.

Definition 9. ***(Bernoulli model [4]).*** *Let $\{\gamma_{ij}, i = 1, .., n_1, j = 1, ..., n_2\}$ be a sequence of independent identically distributed (i.i.d.) binary Bernoulli random variables with*

$$\mathbb{P}(\gamma_{ij} = 1) = p \triangleq \frac{m}{n_1 n_2}, \quad \mathbb{P}(\gamma_{ij} = 0) = 1 - p. \tag{30}$$

Correspondingly, we define a set

$$\Omega = \{(i, j) |\ \gamma_{ij} = 1\} \text{ with } \mathbb{E}|\Omega| = m. \tag{31}$$

Definition 10. ***(Random orthogonal model [4])*** *The set $\{\boldsymbol{u}_j \in \mathbb{R}^n,\ j = 1, ..., r\}$ is selected uniformly at random among all sets of r orthonormal vectors.*

B Proofs for Section 4

B.1 Proof of Lemma 1

Proof. Problem (6) is equivalent to the following problem

$$\text{Find a matrix } \boldsymbol{D} \in \mathbb{R}^{n_1 \times n_2} \quad \text{s.t. rank}(\boldsymbol{M} + \boldsymbol{D}) \leq r,\ \boldsymbol{D} \in \Omega^{\perp}, \tag{32}$$

where $\boldsymbol{D} \in \Omega^{\perp} = \{\boldsymbol{D} \in \mathbb{R}^{n_1 \times n_2} :\ \mathcal{P}_{\Omega}(\boldsymbol{M} + \boldsymbol{D}) = \mathcal{P}_{\Omega}(\boldsymbol{M})\}$ (i.e., $\mathcal{P}_{\Omega}(\boldsymbol{D}) = \boldsymbol{0}$)) is equivalent to the constraint $\mathcal{P}_{\Omega}(\boldsymbol{X}) = \mathcal{P}_{\Omega}(\boldsymbol{M})$ in (6).

Then, $\boldsymbol{M}$ is the unique optimal solution to (6) is equivalent to $\boldsymbol{D} = \boldsymbol{0}$. We know that rank$(\boldsymbol{M} + \boldsymbol{D}) \leq r$ is equivalent to $\boldsymbol{D} \in \mathcal{D}_{\mathcal{M}}(\boldsymbol{M})$ according to (13). Therefore, $\boldsymbol{D} = \boldsymbol{0}$ is equivalent to $\Omega^{\perp} \cap \mathcal{D}_{\mathcal{M}}(\boldsymbol{M}) = \{\boldsymbol{0}\}$, namely, set $\mathcal{D}_{\mathcal{M}}(\boldsymbol{M})$ and set $\Omega^{\perp}$ intersect uniquely at $\boldsymbol{0}$. The lemma is proved.

Interested readers can check a similar proof for Lemma 3.34 (page 116) of [44].

□

B.2 Proof of Lemma 3

We provide a geometric interpretation of the homomorphism property in the context of matrix completion problems.

- The homomorphism property under the scheme in Alg. 1 and Alg. 2 can be mathematically interpreted as: $\boldsymbol{M}$ is the unique optimal solution to problem (6), while simultaneously $\boldsymbol{M} + \boldsymbol{K}\boldsymbol{R}$ is the unique optimal solution to problem (7). According to Lemma 1, the necessary and sufficient condition would be $\Omega^{\perp} \cap \mathcal{D}_{\mathcal{M}}(\boldsymbol{M}) = \{\boldsymbol{0}\}$ and $\Omega^{\perp} \cap \mathcal{D}_{\overline{\mathcal{M}}}(\overline{\boldsymbol{M}}) = \{\boldsymbol{0}\}$.
- Since it always holds that $\mathcal{D}_{\mathcal{M}}(\boldsymbol{M}) \subseteq T$ and $\mathcal{D}_{\overline{\mathcal{M}}}(\overline{\boldsymbol{M}}) \subseteq \overline{T}$ according to Lemma 5 and Lemma 6, a sufficient condition for the homomorphic property is that $\Omega^{\perp} \cap \overline{T} = \{\boldsymbol{0}\}$ and $\Omega^{\perp} \cap T = \{\boldsymbol{0}\}$ hold simultaneously. The scheme in Alg. 1 and Alg. 2 satisfies that $T \subseteq \overline{T}$, thus $\Omega^{\perp} \cap \overline{T} = \{\boldsymbol{0}\}$ implies $\Omega^{\perp} \cap T = \{\boldsymbol{0}\}$.
- If the sampling probability p is large enough, with high probability, it is guaranteed that $\Omega^{\perp} \cap \overline{T} = \{\boldsymbol{0}\}$, which also implies $\Omega^{\perp} \cap T = \{\boldsymbol{0}\}$ since $T \subseteq \overline{T}$.

First, according to Lemma 5 and Lemma 6, we have

$$\mathcal{D}_{\mathcal{M}}(\boldsymbol{M}) = \textbf{cone}\{\mathcal{M} - \{\boldsymbol{M}\}\} \subseteq T,\ \ \mathcal{D}_{\overline{\mathcal{M}}}(\overline{\boldsymbol{M}}) = \textbf{cone}\{\overline{\mathcal{M}} - \{\overline{\boldsymbol{M}}\}\} \subseteq \overline{T}. \tag{33}$$

Lemma 5 states that the tangent cone of the set $\mathcal{M}$ evaluated at $\boldsymbol{M}$ has a closed-form expression, namely, it is just the space T in (4). Lemma 6 claims that the tangent cone of the set $\mathcal{M}$ evaluated at $\boldsymbol{M}$ is slightly larger than the cone $\mathbf{cone}(\mathcal{M} - \{\mathbf{M}\})$, where $\{\boldsymbol{M}\}$ denotes a set.

Definition 11. *([38]) A matrix $\Xi \in \mathbb{R}^{n_1 \times n_2}$ is tangent to $\mathcal{M}$ at $\boldsymbol{M}$ if there exists a sequence $\{\boldsymbol{M}^n\}$ contained in $\mathcal{M}$ and converging to $\boldsymbol{M}$, and a sequence $\{a_n\}$ of nonnegative numbers, such that the sequence $a_n(\boldsymbol{M}^n - \boldsymbol{M})$ converges to Ξ. Then, the tangent cone of the set $\mathcal{M}$ at point $\boldsymbol{M}$, denoted by $T_{\mathcal{M}}(\boldsymbol{M})$, is formally defined as follows*

$$T_{\mathcal{M}}(\boldsymbol{M}) = \{\Xi \in \mathbb{R}^{n_1 \times n_2} \mid \exists \boldsymbol{M}^n \subseteq \mathcal{M},\ \{a_n\} \subseteq \mathbb{R}^+,\ s.t.\ \boldsymbol{M}^n \to \boldsymbol{M},\ a_n(\boldsymbol{M}^n - \boldsymbol{M}) \to \Xi\}. \tag{34}$$

Lemma 5. *[6] (Theorem. 6.1)) Let $\boldsymbol{M} = \boldsymbol{U}\Sigma\boldsymbol{V}^\top$ be the skinny SVD (or reduced SVD) of matrix $\boldsymbol{M}$. The tangent cone $T_{\mathcal{M}}(\boldsymbol{M})$ of the set $\mathcal{M} = \{\boldsymbol{X} \in \mathbb{R}^{n_1 \times n_2} : rank(\boldsymbol{X}) \leq r\}$ at $\boldsymbol{M}$ is a linear subspace given by*

$$T_{\mathcal{M}}(\boldsymbol{M}) = \{\boldsymbol{U}\boldsymbol{A}^\top + \boldsymbol{B}\boldsymbol{V}^\top \mid \boldsymbol{A} \in \mathbb{R}^{n_1 \times r}, \boldsymbol{B} \in \mathbb{R}^{n_2 \times r}\} \triangleq T, \tag{35}$$

where the complementary space is denoted by $T^\perp$.

Lemma 6. *([15], Theorem 4.8) Let $\mathcal{M}$ be a non-empty subset of a real normed space. If $\mathcal{M}$ is star-shaped w.r.t. some $\boldsymbol{M} \in \mathcal{M}$, i.e., $t(\mathcal{M} - \{\boldsymbol{M}\}) \subseteq \mathcal{M} - \{\boldsymbol{M}\}$ for all $t \in [0, 1]$, then if follows*

$$\mathcal{D}_{\mathcal{M}}(\boldsymbol{M}) = \boldsymbol{cone}(\mathcal{M} - \{\boldsymbol{M}\}) \subseteq T_{\mathcal{M}}(\boldsymbol{M}). \tag{36}$$

Second, we point out that the scheme in Alg. 1 and Alg. 2 satisfies $T \subseteq \overline{T}$. It leads to our argument that $\Omega^\perp \cap \overline{T} = \{\boldsymbol{0}\}$ implies $\Omega^\perp \cap T = \{\boldsymbol{0}\}$, since $\boldsymbol{0}$ belongs to $\Omega^\perp$, $\overline{T}$ and T by definition. This condition can be interpreted as follows: with the same observation set Ω, recovering $\overline{\boldsymbol{M}} = \boldsymbol{M} + \boldsymbol{K}\boldsymbol{R}$ (a matrix in a larger subspace) is more difficult than recovering $\boldsymbol{M}$. This is formulated in Lemma 3.

B.3 Proof of Theorem 2

Proof. Applying Theorem 1 for cyphertext $\overline{\boldsymbol{M}}$, then we obtain that with probability at least $1 - 3n_2^{-\beta}$,

$$||p^{-1}\mathcal{P}_{\overline{T}}\mathcal{P}_\Omega\mathcal{P}_{\overline{T}} - \mathcal{P}_{\overline{T}}|| \leq C'_R\sqrt{\frac{n_2\overline{\mu}_0\overline{r}(\beta\log n_2)}{m}} \triangleq \epsilon', \tag{37}$$

provided that $\epsilon' < 1$.

Then, under the proposed scheme in Alg. 1 and Alg. 2, we show that $\epsilon' < 1$ implies $\epsilon < 1$, and we have $\overline{\boldsymbol{M}} = \boldsymbol{M} + \boldsymbol{K}\boldsymbol{R}$ and $\overline{r} > r$.

Combining the coherence change in Section 4.3, we can easily know that $\mu_0 < \overline{\mu}_0$. Since $r < \overline{r} = \text{rank}(\overline{\boldsymbol{M}})$, we obtain that $\mu_0 r < \overline{\mu}\overline{r}$. Therefore, there are constants C_R, C'_R such that for $\beta > 1$, $\epsilon' < 1$ implies $\epsilon < 1$. The proof is completed. □

B.4 Proof of Theorem 3

Proof. Recall that Lemma 3 states that a sufficient condition for the homomorphic property is $\Omega^\perp \cap \overline{T} = \{\boldsymbol{0}\}$. We use Theorem 2 to derive a sufficient condition $p \geq \frac{C_0\overline{\mu}_0\overline{r}(\beta\log n_2)}{n_1}$.

According to Theorem 2, if $\Omega \sim \textbf{Ber}(p)$, with probability at least $1 - 3n_2^{-\beta}$, we have

$$||\mathcal{P}_T - p^{-1}\mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T|| \leq \epsilon < 1, \quad ||\mathcal{P}_{\overline{T}} - p^{-1}\mathcal{P}_{\overline{T}}\mathcal{P}_\Omega\mathcal{P}_{\overline{T}}|| \leq \epsilon' < 1, \tag{38}$$

provided that $p \geq \frac{C_0\overline{\mu}_0\overline{r}(\beta\log n_2)}{n_1}$.

Note that $\mathcal{I} = \mathcal{P}_\Omega + \mathcal{P}_{\Omega^\perp}$, where $\mathcal{P}_{\Omega^\perp}$ denote the projection onto $\Omega^\perp$, then

$$\mathcal{P}_{\overline{T}} - p^{-1}\mathcal{P}_{\overline{T}}\mathcal{P}_\Omega\mathcal{P}_{\overline{T}} = \mathcal{P}_{\overline{T}} - p^{-1}\mathcal{P}_{\overline{T}}(\mathcal{I} - \mathcal{P}_{\Omega^\perp})\mathcal{P}_{\overline{T}} = p^{-1}(\mathcal{P}_{\overline{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\overline{T}} - (1-p)\mathcal{P}_{\overline{T}}), \tag{39}$$

then by the triangle inequality of the operator norm, we obtain that

$$||\mathcal{P}_{\overline{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\overline{T}}|| \leq p||\mathcal{P}_{\overline{T}} - p^{-1}\mathcal{P}_{\overline{T}}\mathcal{P}_\Omega\mathcal{P}_{\overline{T}}|| + (1-p)||\mathcal{P}_{\overline{T}}|| < \epsilon' p + (1-p), \tag{40}$$

where we have $||\mathcal{P}_{\overline{T}}|| = 1$ according to Eq. 24, since for $\overline{\boldsymbol{X}} \in \overline{T}$, we have $||\mathcal{P}_{\overline{T}}(\overline{\boldsymbol{X}})|| = \overline{\boldsymbol{X}}$.

Since $||\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\overline{T}}||^2 = ||\mathcal{P}_{\overline{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\overline{T}}|| \leq ||\mathcal{P}_{\overline{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\overline{T}}||$, then $||\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\overline{T}}|| < \sqrt{1 - p + \epsilon' p} < 1$. Note that $||\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\overline{T}}|| < 1$ implies that $\Omega^\perp \cap \overline{T} = \{\boldsymbol{0}\}$, which holds by contradiction: if there exists a nonzero matrix $\boldsymbol{X} \in \Omega^\perp \cap \overline{T}$, then $||\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\overline{T}}(\boldsymbol{X})|| = \boldsymbol{X}$ and $||\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\overline{T}}|| = 1$.

We have shown that if $p \geq \frac{C_0\overline{\mu}_0\overline{r}(\beta\log n_2)}{n_1}$, with probability at least $1 - 6n_2^{-\beta}$, it holds for $\Omega^\perp \cap \overline{T} = \{\boldsymbol{0}\}$. Combining Lemma 3, the proof is completed. □

B.5 Proof of Lemma 4

We provide key lemmas of coherence change under elementary operations (namely, union of two subspaces), which will be used to characterize the coherence change under the scheme in Alg. 1 and Alg. 2.

The concept "coherence" measures the relationship between a low-dimensional space and the observation operator $\mathcal{P}_\Omega$, namely the cosine (with a scaling factor $\frac{n}{r}$) of the principal angle between the low-dimensional space and a standard basis.

Definition 12. *(Coherence [5]) Let $\boldsymbol{U} \in \mathbb{R}^{n\times r}$ be the r left singular vectors of $\boldsymbol{M}$ (corresponds to the column subspace $\mathcal{S}(\boldsymbol{M})$) and $\boldsymbol{P_U}$ be the orthogonal projection onto $\boldsymbol{U}$. Then the coherence of $\boldsymbol{U}$ (or $\mathcal{S}(\boldsymbol{M})$, respectively) is defined as*

$$\mu(\mathcal{S}(\boldsymbol{M})) = \mu(\boldsymbol{U}) \triangleq \frac{n}{r}\max_{i=1,\dots,n} ||\boldsymbol{P_U}\boldsymbol{e}_i||_2^2 = \frac{n}{r}\max_{i=1,\dots,n} ||\boldsymbol{U}(\boldsymbol{U}^\top\boldsymbol{U})^{-1}\boldsymbol{U}^\top\boldsymbol{e}_i||_2^2 = \frac{n}{r}\max_{i=1,\dots,n} ||\boldsymbol{U}^\top\boldsymbol{e}_i||_2^2, \tag{41}$$

since $(\boldsymbol{U}^\top\boldsymbol{U})^{-1} = \boldsymbol{I}$ and $\boldsymbol{U}$ is orthonormal.

Remark 1. *Note that the left singular vectors $\boldsymbol{U}$ (or right singular vectors $\boldsymbol{V}$, respectively) are not uniquely defined and all candidates are unitary transform to each other, i.e., $\boldsymbol{M} = (\boldsymbol{UW})\boldsymbol{S}(\boldsymbol{VW})^\top = \boldsymbol{USV}^\top$ where $\boldsymbol{W} \in \mathbb{R}^{r\times r}$ is an orthonormal matrix, i.e., $\boldsymbol{WW}^\top = \boldsymbol{W}^\top\boldsymbol{W} = \boldsymbol{I}$. The coherence of the column subspace $\mu(\mathcal{S}(\boldsymbol{M}))$ (or the row subspace $\mu(\mathcal{S}(\boldsymbol{M}^\top))$, respectively) is uniquely defined, since the coherence is invariant under unitary transformation. Specifically, given two orthonormal matrices $\boldsymbol{W} \in \mathbb{R}^{r\times r}$ and $\boldsymbol{U} \in \mathbb{R}^{n_1\times r}$, then $\mu(\boldsymbol{UW}) = \mu(\boldsymbol{U})$, since $||\boldsymbol{P}_{(\boldsymbol{UW})}\cdot\boldsymbol{e}_i||_2 = ||\boldsymbol{UW}((\boldsymbol{UW})^\top\boldsymbol{UW})^{-1}(\boldsymbol{UW})^\top\boldsymbol{e}_i||_2 = ||\boldsymbol{U}^\top\boldsymbol{e}_i||_2$. Therefore, we will use $\mu(\boldsymbol{U})$ instead of $\mu(\mathcal{S}(\boldsymbol{M}))$.*

Lemma 7. ***(Incoherence change under the union operation)*** *Let $r = r_1 + r_2$, the matrix $\boldsymbol{U} = [\boldsymbol{U}_1, \boldsymbol{U}_2] \in \mathbb{R}^{n\times(r_1+r_2)}$ be the concatenation of two non-overlapping orthonormal matrices $\boldsymbol{U}_1 \in \mathbb{R}^{n\times r_1}$ and $\boldsymbol{U}_2 \in \mathbb{R}^{n\times r_2}$. Then, we have $\mu(\boldsymbol{U}) \le \frac{r_1}{r}\mu(\boldsymbol{U}_1) + \frac{r_2}{r}\mu(\boldsymbol{U}_2)$, and it applies to the union of two subspaces, i.e., $\mu(\mathcal{S}(\boldsymbol{U})) \le \frac{r_1}{r}\mu(\mathcal{S}(\boldsymbol{U}_1)) + \frac{r_2}{r}\mu(\mathcal{S}(\boldsymbol{U}_2))$.*

Proof. From Def. 12, we have

$$\begin{aligned}\mu(\boldsymbol{U}) &= \frac{n}{r}\max_{i=1}^{n} ||[\boldsymbol{U}_1, \boldsymbol{U}_2]^\top\boldsymbol{e}_i||_2^2 = \frac{n}{r}\max_{i=1}^{n} \left(||\boldsymbol{U}_1^\top(:,i)||_2^2 + ||\boldsymbol{U}_2^\top(:,i)||_2^2\right)\\ &\le \frac{n}{r}\max_{i=1}^{n} ||\boldsymbol{U}_1^\top(:,i)||_2^2 + \frac{n}{r}\max_{j=1}^{n} ||\boldsymbol{U}_2^\top(:,j)||_2^2\\ &\le \frac{r_1}{r}\mu(\boldsymbol{U}_1) + \frac{r_2}{r}\mu(\boldsymbol{U}_2),\end{aligned} \tag{42}$$

where $\mu(\boldsymbol{U}_1)$ and $\mu(\boldsymbol{U}_2)$ are given in Def. 12. For the union of subspaces, one can easily verify the incoherence change according to Remark 1. □

Definition 13. ***(Random orthogonal model** [5]) The set $\{\boldsymbol{u}_j \in \mathbb{R}^n,\ j = 1, ..., r\}$ is selected uniformly at random among all sets of r orthonormal vectors.*

Lemma 8. ***(Random subspaces span incoherence subspaces** [5]) Set $r' = \max(r, \log n)$. There exist positive constants C and c, the random orthogonal model in Def. 13 obeys*

$$\max_i ||\mathcal{P}_{\boldsymbol{U}}\boldsymbol{e}_i||_2^2 = \max_i ||\boldsymbol{U}^\top\boldsymbol{e}_i||_2^2 \le Cr'/n, \ \text{ which yields } \mu(\boldsymbol{U}) \le C\cdot\max(1, \frac{\log n}{r}), \tag{43}$$

with probability $1 - cn^{-3}\log n$.

Here we formally prove Lemma 4.

Proof. The scheme in Alg. 1 and Alg. 2 can be represented as $\overline{\boldsymbol{M}} = \boldsymbol{M} + \boldsymbol{KR}$. Let $\boldsymbol{M} = \boldsymbol{USV}^\top$ and $\overline{\boldsymbol{M}} = \overline{\boldsymbol{U}}\ \overline{\boldsymbol{S}}\ \overline{\boldsymbol{V}}^\top$, we have the column subspaces $\mathcal{S}(\boldsymbol{M}) = \mathcal{S}(\boldsymbol{U})$, $\mathcal{S}(\overline{\boldsymbol{M}}) = \mathcal{S}(\overline{\boldsymbol{U}})$, and the row subspaces $\mathcal{S}(\boldsymbol{M}^\top) = \mathcal{S}(\boldsymbol{V})$, $\mathcal{S}(\overline{\boldsymbol{M}}^\top) = \mathcal{S}(\overline{\boldsymbol{V}})$. Let $\boldsymbol{KR} = \boldsymbol{U}'\boldsymbol{S}'(\boldsymbol{V}')^\top$. Since $\boldsymbol{K}$ and $\boldsymbol{R}$ are randomly generated matrices, thus we know that $\boldsymbol{U}'$ and $\boldsymbol{V}'$ obey the random orthogonal model in Def. 13.

First, with probability 1, we have

$$\mathcal{S}(\boldsymbol{U}) \subseteq \mathcal{S}(\overline{\boldsymbol{U}}), \quad \text{specifically, } \mathcal{S}(\overline{\boldsymbol{U}}) = \mathcal{S}(\boldsymbol{U}) \cup \mathcal{S}(\boldsymbol{K}) = \mathcal{S}(\boldsymbol{U}) \cup \mathcal{S}(\boldsymbol{U}'), \tag{44}$$

and

$$\mathcal{S}(\boldsymbol{V}) \subseteq \mathcal{S}(\overline{\boldsymbol{V}}), \quad \text{specifically, } \mathcal{S}(\overline{\boldsymbol{V}}) = \mathcal{S}(\boldsymbol{V}) \cup \mathcal{S}(\boldsymbol{R}^\top) = \mathcal{S}(\boldsymbol{V}) \cup \mathcal{S}(\boldsymbol{V}'). \tag{45}$$

Secondly, we bound the coherence changes. Note that with probability 1, $\text{rank}(\overline{\boldsymbol{M}}) = \overline{r} = r + k$. From (44), we know that $\overline{\boldsymbol{U}} \in \mathbb{R}^{n_1 \times \overline{r}}$ can be represented as the union of $\boldsymbol{U} \in \mathbb{R}^{n_1 \times r}$ and $\boldsymbol{U}' \in \mathbb{R}^{n_1 \times k}$, i.e., $\overline{\boldsymbol{U}} = [\boldsymbol{U}\ \boldsymbol{U}']$, where $\boldsymbol{U}'$ obeys the random orthogonal model and the non-overlapping requirement in Lemma 7. From Lemma 8, we know that $\mu(\boldsymbol{U}') \leq C \cdot \max(1, \frac{\log n_1}{k})$.

According to Lemma 7, we then obtain that

$$\mu(\overline{\boldsymbol{U}}) \leq \frac{r}{\overline{r}}\mu_0 + \frac{k}{\overline{r}} \cdot C \max(1, \frac{\log n_1}{k}) \leq \frac{r}{\overline{r}}\mu_0 + C \max(\frac{k}{\overline{r}}, \frac{\log n_1}{\overline{r}}). \tag{46}$$

Similarly, for the row subspace, we have

$$\mu(\overline{\boldsymbol{V}}) \leq \frac{r}{\overline{r}}\mu_0 + C \max(\frac{k}{\overline{r}}, \frac{\log n_2}{\overline{r}}), \tag{47}$$

Therefore,

$$\overline{\mu}_0 \leq \frac{r}{\overline{r}}\mu_0 + C \max(\frac{k}{\overline{r}}, \frac{\log n_2}{\overline{r}}). \tag{48}$$

□

C Proof for Section 5

C.1 Proof of Theorem 4

A potential issue of the proposed scheme in Alg. 1 and Alg. 2 is the projection recovery. Namely, one can do a corresponding projection to obtain the real data after a single encryption. Therefore, we execute the proposed scheme twice and introduce two parameters σ_1 and σ_2 in Theorem 4:

- First-round encryption: we randomly generate a matrix $\boldsymbol{K}^1 \in \mathbb{R}^{n_1 \times k}$ and k random numbers $\boldsymbol{R}^1 \overset{\text{i.i.d}}{\sim} \mathcal{N}(\boldsymbol{0}, \sigma_1^2 \boldsymbol{I}_k)$. Then we obtain $\mathcal{P}_\Omega(\overline{M}) = \mathcal{P}_\Omega(M) + \mathcal{P}_\Omega(\boldsymbol{K}^1\boldsymbol{R}^1)$.
- Second-round encryption: we obtain the column space of $\overline{M}$ as $\boldsymbol{K}^2 \in \mathbb{R}^{n_1 \times k}$ and k random numbers $\boldsymbol{R}^2 \overset{\text{i.i.d}}{\sim} \mathcal{N}(\boldsymbol{0}, \sigma_2^2 \boldsymbol{I}_k)$. Then we obtain $\mathcal{P}_\Omega(\overline{\overline{M}}) = \mathcal{P}_\Omega(\overline{M}) + \mathcal{P}_\Omega(\boldsymbol{K}^2\boldsymbol{R}^2)$.

As an overview of the proof strategy, we first detail the proof for a single-round encryption case shown in Theorem 5. Note that Theorem 4 can be easily derived by applying Theorem 5 twice.

Theorem 5. *(Gaussian mechanism [10]) Given an arbitrary n-dimensional function $\mathcal{A} : \mathcal{T}^n \to \mathcal{O}^n$, define its ℓ_2-norm sensitivity as $\Delta = \max_{dist(D,D')=1} ||\mathcal{A}(D) - \mathcal{A}(D')||_2$ where dist$(D, D') = 1$ means that $D, D' \in \mathcal{T}^n$ differ from each other by one data entry. The Gaussian mechanism with parameter σ adds Gaussian noise $\boldsymbol{w} \overset{i.i.d}{\sim} \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$ to the output. Let $\epsilon \in (0, 1)$ and $c^2 > 2\ln(1.25/\delta)$, then the Gaussian mechanism, i.e., $\mathcal{A}(D) + \boldsymbol{w}$, with parameter $\sigma \geq c\Delta\sqrt{2\ln(2/\delta)}/\epsilon$ satisfies (ϵ, δ)-differential privacy.*

Proof. Let $\boldsymbol{v} \in \mathcal{O}^n$ be a vector with $||\boldsymbol{v}||_2 \leq \Delta$. For a pair of databases $D, D' \in \mathcal{T}^n$ with dist$(D, D') = 1$, we would like to obsecure the vector $\boldsymbol{v} = \mathcal{A}(D) - \mathcal{A}(D')$ by adding Gaussian noise $\boldsymbol{w} \overset{\text{i.i.d}}{\sim} \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$, i.e., $\mathcal{A}(D) + \boldsymbol{w}$. Thus an adversary user will not be able to differentiate D from D' by observing $\mathcal{A}(D) + \boldsymbol{w}$ or $\mathcal{A}(D') + \boldsymbol{w}'$. According to the joint (ϵ, δ)-differential privacy in Def. 3, the key is to quantify σ under which the random variable *privacy loss* in (22) is bounded by ϵ, with probability at least $1 - \delta$. Explicitly, we would like to show that

$$\begin{aligned} \left|\mathcal{L}^{(o)}_{\mathcal{M}(D)||\mathcal{M}(D')}\right| &= \left|\ln \frac{\mathbb{P}(\mathcal{A}(D) = o)}{\mathbb{P}(\mathcal{A}(D') = o)}\right| = \left|\ln \frac{\exp(-\frac{1}{2\sigma^2}||o - \mathcal{A}(D)||_2^2)}{\exp(-\frac{1}{2\sigma^2}||o - \mathcal{A}(D')||_2^2)}\right| \\ &= \left|\ln \frac{\exp(-\frac{1}{2\sigma^2}||\boldsymbol{w}||_2^2)}{\exp(-\frac{1}{2\sigma^2}||\boldsymbol{w} + \boldsymbol{v}||_2^2)}\right| = \left|\frac{1}{2\sigma^2}(||\boldsymbol{w}||_2^2 - ||\boldsymbol{w} + \boldsymbol{v}||_2^2)\right| \leq \epsilon \end{aligned} \tag{49}$$

holds with with probability at least $1-\delta$.

Note that $\boldsymbol{w} \overset{\text{i.i.d}}{\sim} \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$, combining the fact that *the distribution of a spherically symmetric normal is independent of the orthogonal basis from which its constituent normals are drawn*, we choose to work in a basis that is aligned with $\boldsymbol{v}$. For a basis $\{\boldsymbol{B}_1, \boldsymbol{B}_2, ..., \boldsymbol{B}_n\}$ (or alternatively represented as a matrix $\boldsymbol{B} \in \mathbb{R}^{n\times n}$) with $\boldsymbol{B}_1$ being parallel to $\boldsymbol{w}$, we draw $\boldsymbol{w}$ by drawing signed scalars $\lambda_i \sim \mathcal{N}(0, \sigma^2)$ for $i = 1, ..., n$, then defining $\boldsymbol{w}^i = \lambda_i \boldsymbol{B}_i \in \mathbb{R}^n$ and $\boldsymbol{w} = \sum_{i=1}^{n} \boldsymbol{w}^i \in \mathbb{R}^n$. Consider the right triangle with base $\boldsymbol{v} + \boldsymbol{w}^i$ and edge $\sum_{i=2}^{n} \boldsymbol{w}^i \in \mathbb{R}^n$, its hypotenuse is $\boldsymbol{v} + \boldsymbol{w}$. Then, we have

$$\begin{aligned} ||\boldsymbol{w}||_2^2 &= \sum_{i=1}^{n} ||\boldsymbol{w}^i||_2^2, \\ ||\boldsymbol{v}+\boldsymbol{w}||_2^2 &= ||\boldsymbol{v}+\boldsymbol{w}^1||_2^2 + \sum_{i=2}^{n} ||\boldsymbol{w}^i||_2^2 = (||\boldsymbol{v}||_2 + \lambda_1)^2 + \sum_{i=2}^{n} ||\boldsymbol{w}^i||_2^2, \\ ||\boldsymbol{v}+\boldsymbol{w}||_2^2 - ||\boldsymbol{w}||_2^2 &= ||\boldsymbol{v}||_2^2 + 2\lambda_1 ||\boldsymbol{v}||_2, \end{aligned} \tag{50}$$

where we used the fact $\boldsymbol{B}_1$ is parallel to $\boldsymbol{v}$ and $||\boldsymbol{w}^1||_2^2 = \lambda_1^2$.

Recall that $||\boldsymbol{v}||_2 \leq \Delta$ and $\lambda_i \sim \mathcal{N}(0, \sigma^2)$ for $i = 1, ..., n$, the bound in (49) becomes

$$\left| \ln \frac{\exp(-\frac{1}{2\sigma^2}||\boldsymbol{w}||_2^2)}{\exp(-\frac{1}{2\sigma^2})||\boldsymbol{w}+\boldsymbol{v}||_2^2} \right| = \left| \frac{1}{2\sigma^2}(||\boldsymbol{w}||_2^2 - ||\boldsymbol{w}+\boldsymbol{v}||_2^2) \right| \leq \left| \frac{1}{2\sigma^2}(2\lambda_1\Delta + \Delta^2) \right| \leq \epsilon. \tag{51}$$

The quantity in (51) is bounded by ϵ whenever $|\lambda_1| < \frac{\sigma^2\epsilon}{\Delta} - \frac{\Delta}{2}$.

Set $\sigma = c\Delta/\epsilon$ and $c^2 > 2\ln(1.25/\delta)$. Let us partition $\mathbb{R}$ into $\mathbb{R} = R_1 \cup R_2$, where $R_1 = \{\lambda \in \mathbb{R} \mid |\lambda| \leq c\Delta/\epsilon\}$ and $R_2 = \{\lambda \in \mathbb{R} \mid |\lambda| > c\Delta/\epsilon\}$, and define $\mathcal{O} = O_1 \cup O_2$ where

$$\begin{aligned} O_1 &= \{\mathcal{A}(x) + \lambda | \lambda \in R_1\}, \\ O_2 &= \{\mathcal{A}(x) + \lambda | \lambda \in R_2\}. \end{aligned} \tag{52}$$

To ensure the privacy loss is bounded by ϵ with probability at least $1-\delta$, one requires

$$\mathbb{P}\left[|\lambda| \geq \frac{\sigma^2\epsilon}{\Delta} - \frac{\Delta}{2}\right] < \delta, \tag{53}$$

or alternatively, one requires

$$\mathbb{P}\left[\lambda \geq \frac{\sigma^2\epsilon}{\Delta} - \frac{\Delta}{2}\right] < \delta/2. \tag{54}$$

Taking the integration of (51), one obtains that

$$\begin{aligned} \mathbb{P}[\mathcal{A}(x) + \lambda \in O] &= \mathbb{P}[\mathcal{A}(x) + \lambda \in O_1] + \mathbb{P}[\mathcal{A}(x) + \lambda \in O_2] \\ &\leq \mathbb{P}[\mathcal{A}(x) + \lambda \in O_1] + \delta \\ &\leq e^{\epsilon} \cdot \mathbb{P}[\mathcal{A}(y) + \lambda \in O_1] + \delta, \end{aligned} \tag{55}$$

yielding the joint (ϵ, δ)-differntial privacy for the Gaussian mechanism. □