Uncertainty Quantification For Low-Rank Matrix Completion With Heterogeneous and Sub-Exponential Noise

Vivek F. Farias
Massachusetts Institute of Technology

Andrew A. Li
Carnegie Mellon University

Tianyi Peng
Massachusetts Institute of Technology

Abstract

The problem of low-rank matrix completion with heterogeneous and sub-exponential (as opposed to homogeneous and Gaussian) noise is particularly relevant to a number of applications in modern commerce. Examples include panel sales data and data collected from web-commerce systems such as recommendation engines. An important unresolved question for this problem is characterizing the distribution of estimated matrix entries under common low-rank estimators. Such a characterization is essential to any application that requires quantification of uncertainty in these estimates and has heretofore only been available under the assumption of homogenous Gaussian noise. Here we characterize the distribution of estimated matrix entries when the observation noise is heterogeneous sub-exponential and provide, as an application, explicit formulas for this distribution when observed entries are Poisson or Binary distributed.

1. Introduction

Consider the problem of *low-rank matrix completion*: there exists a low-rank matrix that we seek to recover, having observed only a subset of its entries, each perturbed by additive noise. A rich stream of research over the past two decades has essentially solved this problem – there exist efficient algorithms which achieve order-optimal recovery guarantees under provably-minimal assumptions (Candès and Recht 2009, Candes and Plan 2010, Keshavan et al. 2010). Further advances have yielded (and continue to yield) algorithmic improvements (Mazumder et al. 2010, Jain et al. 2013, Tanner and Wei 2016, Dong et al. 2021), and a deeper understanding of the optimization landscape itself (Ge et al. 2016, Zhu et al. 2017).

Naturally, these algorithms have been applied in a vast array of applications, including recommendation systems, bioinformatics, network localization, and modern commerce (Su and Khoshgoftaar 2009, Natarajan and Dhillon 2014, So and Ye 2007, Amjad and Shah 2017, Farias et al. 2021a), just to name a few. Now many of these applications require, in addition to scalability and accuracy, the ability to quantify the *uncertainty* of an estimator – for example, something as seemingly-straightforward as confidence intervals on the estimated entries of a matrix.

Such an *uncertainty quantification* procedure, analogous to existing procedures for problems like linear regression, would ideally (a) apply to a commonly-used estimator, (b) require no more additional computation than the estimator itself, and (c) be justified by a (limiting) distributional characterization. Given the volume and success of the research just described, it is perhaps surprising that this problem has been largely unsolved (see the Related Work for past progress).

Fortunately, there was a recent "breakthrough." Applying newer techniques such as the *leave-one-out* technique and fine-grained entry-wise analysis (Ma et al. 2018, Ding and Chen 2020, Abbe et al. 2020), Chen et al. (2019, 2020) proposed an uncertainty quantification technique for matrix completion, which satisfies the three "ideal" conditions above, in the case of *homogeneous Gaussian noise*. Further progress in Xia and Yuan (2021) extended this to homogeneous sub-Gaussian noise.

Toward "Realistic" Noise: A gap still exists when we seek to apply these inferential results in practice, since many applications have more sophisticated noise models (namely, heterogeneous and sub-exponential noise). For example, in discrete panel sales data, the observation for sales at a location during a period of time is commonly modeled as *Poisson* with a certain

Proceedings of the 25$^{\text{th}}$ International Conference on Artificial Intelligence and Statistics (AISTATS) 2022, Valencia, Spain. PMLR: Volume 151.

Noise Model	Entry-wise Uncertainty
Gaussian	$\sigma^2(u_i^{*2} + v_j^{*2})/p$
Poisson	$M_{ij}^*(u_i^*\|u^*\|_3^3 + v_j^*\|v^*\|_3^3)/p$
Binary	$M_{ij}^*(u_i^*\|u^*\|_3^3 + v_j^*\|v^*\|_3^3 - M_{ij}^*\|u^*\|_4^4 - M_{ij}^*\|v^*\|_4^4)/p$

Table 1: A comparison of uncertainty formulas for different noise models when $r = 1$, i.e., $M^* = \sigma_1 u^* v^{*\top}$. See details in Section 4.

expected sales rate (Amjad and Shah 2017, Shi et al. 2014). Similarly in web-commerce systems, data indicating clicks or purchases is often binary and modeled as *Bernoulli* random variables (Ansari and Mela 2003, Grover and Srinivasan 1987).

Thus motivated, in this work we establish the first uncertainty quantification results for matrix completion with *heterogenous and sub-exponential noise*. Precisely, we characterize the distribution of recovered matrix entries from common estimators. An application of our results can already be seen in Table 1, where we have derived explicit formulas under Poisson and Binary noise, which are distinctive from the homogeneous Gaussian noise case already existing in the literature. In addition, we demonstrate the quality of our procedure through experiments on real sales data. The proof of our main result generalizes the proof framework in (Chen et al. 2019), leveraging recent results for sub-exponential matrix completion from McRae and Davenport (2019), and a new high-dimension concentration bound (Lemma 1), which may be of independent interest.

Related Work: This paper is related to at least three streams of work. The first is, naturally, *uncertainty quantification in matrix completion.* Besides the works described above, prior approaches to this were based on either (a) converting recovery guarantees on matrix norms to confidence regions (Carpentier et al. 2015, 2018), (b) the Bayesian formulation of matrix completion (Salakhutdinov and Mnih 2008, Fazayeli et al. 2014, Tanaka 2021, Alquier et al. 2015), or (c) deep-learning-bsaed methods (Lakshminarayanan et al. 2016, Zeldes et al. 2017). The second stream relates to *sub-exponential matrix completion.* McRae and Davenport (2019) established guarantees on the Frobenius error $\|\hat{M} - M^*\|_{\mathrm{F}}$; Farias et al. (2021b) established entry-wise error guarantees. This work makes takes one step further with an entry-wise distributional characterization of the error. Finally, there is a line of work, in multi-variate linear regression or PCA, advocating the use of *heteroskedasticity-robust variance estimators* instead of homoskedasticity estimators, since the former are more robust to heterogeneous noise (Long and Ervin 2000, Hayes and Cai 2007, Imbens and Kolesar 2016, Cribari-Neto and Maria da Glória 2014, Zhang et al. 2018). Our work is in the same spirit, but in the context of matrix completion.

Notation: The sub-exponential norm of a random variable X is defined as $\|X\|_{\psi_1} := \inf\{t > 0 : \mathbb{E}(\exp(|X|/t)) \leq 2\}$. For a matrix $A \in \mathbb{R}^{m\times n}$, we abbreviate $\sum_{(i,j)\in[m]\times[n]} A_{ij}$ as $\sum_{ij} A_{ij}$ when no ambiguity exists. We require a few matrix norms: $\|A\|_{2,\infty}^2 := \max_i \sum_j A_{ij}^2$, $\|A\|_{\max} = \max_{ij}|A_{ij}|$, and $\|A\|_{\mathrm{F}}^2 = \sum_{ij} A_{ij}^2$. The spectral norm is denoted $\|A\|_2$.

2. Model

Let $M^* \in \mathbb{R}^{m\times n}$ be a rank-r matrix, where $m \leq n$ without loss of generality. Let $O = M^* + E$ be the realization of M^* corrupted by a noise matrix $E \in \mathbb{R}^{m\times n}$. We observe $P_\Omega(O)$, which is the subset of entries of O restricted to an observation set $\Omega \subset [m]\times[n]$:

$$P_\Omega(O)_{ij} = \begin{cases} O_{ij} & (i,j)\in\Omega \\ 0 & (i,j)\notin\Omega \end{cases}.$$

The *matrix completion* problem is to recover M^* from this noisy and partial observation $P_\Omega(O)$.

Let $M^* = U^*\Sigma^*V^{*\top}$ be the SVD of M^*. Here, $\Sigma^* \in \mathbb{R}^{r\times r}$ is a diagonal matrix with singular values $\sigma_{\max} = \sigma_1^* \geq \sigma_2^* \geq \ldots \geq \sigma_r^* = \sigma_{\min}$; and $U^* \in \mathbb{R}^{m\times r}, V^* \in \mathbb{R}^{n\times r}$ contain the left and right-singular vectors. Let $\kappa = \sigma_{\max}/\sigma_{\min}$ be the condition number of M^*.

We will make three assumptions. The first two are, by this point, canonical in the matrix completion literature (Candes and Plan 2010, Keshavan et al. 2010, Ma et al. 2018, Abbe et al. 2020, Chen et al. 2019):

Assumption 1 (Uniform Sampling)**.** *Each element of $[m]\times[n]$ is included in Ω independently, and with probability p.*

Assumption 2 (Incoherence)**.**

$$\|U^*\|_{2,\infty} \leq \sqrt{\frac{\mu r}{m}} \quad and \quad \|V^*\|_{2,\infty} \leq \sqrt{\frac{\mu r}{n}} \qquad (1)$$

Finally, our third assumption is a *generalization* of the independent (and often homogeneous), sub-Gaussian noise that is typically assumed in the literature (Chen et al. 2019, Xia and Yuan 2021). As described above, this generalization enables a host of practical applications, such as those arising in counting data and panel sales data (Amjad and Shah 2017, Ansari and Mela 2003).

Assumption 3 (Independent Sub-exponential Noise)**.** *The entries of E are independent, mean-zero random variables with variances σ_{ij}^2, and are also independent from Ω. Furthermore, $\|E_{ij}\|_{\psi_1} \le L$ for every (i,j), where $\|\cdot\|_{\psi_1}$ is the sub-exponential norm.*

3. Algorithm

In this section, we describe a "de-biased" estimator M^{d} for M^*. This was originally proposed in (Chen et al. 2019), where the uncertainty quantification for M^{d} is characterized under homogeneous, Gaussian noise. Motivated by practical applications, we study new uncertainty quantification formulas for M^{d} under heterogenous sub-exponential noise.

To begin, consider a natural least-square estimator for M^*

$$\hat{M} \triangleq \underset{M' \in \mathbb{R}^{m\times n}, \mathrm{rank}(M')=r}{\arg\min} \frac{1}{2p} \|P_\Omega(O - M')\|_{\mathrm{F}}^2 \tag{2}$$

Here, $\hat{M}$ is the projection of M into the set of rank-r matrices in regard to Euclidean distance (restricted on the set Ω).

Directly solving Eq. (2) turns out to be a challenge task. A popular method is to represent $M' = XY^\top$ where $X \in \mathbb{R}^{m\times r}, Y \in \mathbb{R}^{n\times r}$ are low-rank factors, and solve the following non-convex regularized optimization problem

$$\underset{X\in R^{n_1\times r}, Y \in R^{n_2 \times r}}{\mathrm{minimize}} f(X,Y) \tag{3}$$

where

$$f(X,Y) := \frac{1}{2p}\left\|P_\Omega(XY^T - O)\right\|_{\mathrm{F}}^2 + \frac{\lambda}{2p}\|X\|_{\mathrm{F}}^2 + \frac{\lambda}{2p}\|Y\|_{\mathrm{F}}^2 .$$

With proper initializations, simple first-order methods are often sufficient to solve Eq. (3) (Sun and Luo 2016). The regularizer $\lambda > 0$ here is used to promote additional structure properties. For example, when gradient descent is performed, a positive λ is critical for analyzing the convergence properties and also helps to achieve a balance between X and Y (Chen et al. 2020).

However, the use of λ also introduces additional bias to the estimator in Eq. (3), which has been a major obstacle to analyze the uncertainty quantification properties. (Chen et al. 2019) proposes a de-bias procedure to remove the bias brought by λ, based on the solution of Eq. (3). The algorithm is summarized below[1].

Algorithm 1 Gradient Descent with De-bias

Input: $P_\Omega(O)$

1: **Spectral initialization**: $X^0 = U\sqrt{\Sigma}, Y^0 = V\sqrt{\Sigma}$ where $U\Sigma V^\top$ is the top-r partial SVD decomposition of $\frac{1}{p}P_\Omega(O)$.

2: **Gradient updates**: **for** $t = 0, 1, \dots, t_\star - 1$ **do**

$$X^{t+1} = X^t - \frac{\eta}{p}[P_\Omega(X^tY^{t\top} - O)Y^t + \lambda X^t];$$
$$Y^{t+1} = Y^t - \frac{\eta}{p}[P_\Omega(X^tY^{t\top} - O)^TX^t + \lambda Y^t]$$

where η determines the learning rate.

3: **De-bias**:

$$X^{\mathrm{d}} = X^{t_\star}\left(I_r + \frac{\lambda}{p}\left(X^{t_\star\top}X^{t_\star}\right)^{-1}\right)^{1/2} \tag{5}$$

$$Y^{\mathrm{d}} = Y^{t_\star}\left(I_r + \frac{\lambda}{p}\left(Y^{t_\star\top}Y^{t_\star}\right)^{-1}\right)^{1/2} \tag{6}$$

Output: $M^{\mathrm{d}} = X^{\mathrm{d}}Y^{\mathrm{d}\top}$

Steps 1 and 2 in Algorithm 1 form a typical gradient descent procedure for solving Eq. (3). The de-biasing step, i.e. Eqs. (5) and (6) in Algorithm 1, is critical for enabling the uncertainty quantification analysis.

We will use the remainder of this section (which can be skipped without loss of continuity) to provide some intuition for the peculiar form of Eqs. (5) and (6) based on first-order conditions. Consider an example with $p = 1$ (no entry is missing). Since O is fully observed, let $O = U_r\Sigma_rV_r^\top + U_{n-r}\Sigma_{n-r}V_{n-r}^\top$ be the SVD of O, where Σ_r corresponds to the largest r singular values and Σ_{n-r} corresponds to the remaining one. Then it follows that the optimal solution of Eq. (2) is $\hat{M} = U_r\Sigma_rV_r^\top$ (Eckart and Young 1936).

Next, consider the regularized objective Eq. (3). We can derive that the optimal solution (X,Y) for Eq. (3) has the form

$$X = U_r(\Sigma_r - \lambda I_r)^{1/2}, \quad Y = V_r(\Sigma_r - \lambda I_r)^{1/2}.$$

In fact, this can be verified from the first-order condi-

[1] We assume $\lambda \asymp L\log(n)\sqrt{np}, t^\star \asymp n^{23}, \eta \asymp 1/(n^6\kappa^3\sigma_{\max})$ throughout the paper, if not specified explicitly.

tions,

$$\begin{aligned}\frac{\partial f(X,Y)}{\partial X} &= (XY^\top - O)Y + \lambda X \\ &= (U_r(\Sigma_r - \lambda I)V_r^\top - O)Y + \lambda X \\ &= (U_{n-r}\Sigma_{n-r}V_{n-r}^\top - \lambda U_r V_r^\top)Y + \lambda X \\ &\overset{(i)}{=} -\lambda U_r V_r^\top V_r(\Sigma_r - \lambda I_r)^{1/2} + \lambda X \\ &\overset{(ii)}{=} 0,\end{aligned}$$

where in (i) we use that $V_{n-r}^\top Y = V_{n-r}^\top V_r(\Sigma_r - \lambda I_r)^{1/2} = 0$, and in (ii) we use that $V_r^\top V_r = I_r$. Similarly $\frac{\partial f(X,Y)}{\partial Y} = 0$ also vanishes.

Then, this justifies the particular de-biased form in Eqs. (5) and (6):

$$\begin{aligned}X^{\mathrm{d}} &= X(I_r + \lambda(\Sigma_r - \lambda I_r)^{-1})^{1/2} \\ &= X(\Sigma_r(\Sigma_r - \lambda I_r)^{-1})^{1/2} \\ &= U_r(\Sigma_r - \lambda I_r)^{1/2}(\Sigma_r - \lambda I_r)^{-1/2}\Sigma_r^{1/2} \\ &= U_r\Sigma_r^{1/2}.\end{aligned}$$

Similarly, $Y^{\mathrm{d}} = V_r\Sigma_r^{1/2}$. Thus $X^{\mathrm{d}}Y^{\mathrm{d}\top} = U_r\Sigma_r V_r^\top$ is the desired optimal solution of Eq. (2).

4. Results

We can now state our main result: an uncertainty quantification for M^{d} under heterogeneous, sub-exponential noise.

Theorem 1. *Assume $mp \gg \kappa^4\mu^2 r^2 \log^3 n$ and $L\log(n)\sqrt{n/p} \ll \sigma_{\min}/\sqrt{\kappa^4 \mu r \log n}$. Then for every $(i,j) \in [m]\times[n]$, we have*

$$\sup_{t\in\mathbb{R}} \left| P\left\{ \frac{M_{ij}^{\mathrm{d}} - M_{ij}^*}{s_{ij}} \le t \right\} - \Phi(t) \right| \lesssim s_{ij}^{-3}\frac{L^2\mu^3 r^3}{m^2 p} + s_{ij}^{-1}\left(\frac{L^2\log^3(n)\mu r\kappa^5}{p\sigma_{\min}} + \frac{L\mu^2 r^2 \log^2(n)\kappa^4}{pm} \right) + \frac{1}{m^{10}},$$

where $\Phi(\cdot)$ is the CDF of the standard Gaussian, and $s_{ij} > 0$ is given by

$$s_{ij}^2 := \frac{\sum_{l=1}^m \sigma_{lj}^2 \left(\sum_{k=1}^r U_{ik}^* U_{lk}^*\right)^2 + \sum_{l=1}^n \sigma_{il}^2 \left(\sum_{k=1}^r V_{lk}^* V_{jk}^*\right)^2}{p}. \tag{7}$$

To quickly parse this result, note that a typical scaling of the parameters would see $m = \Theta(n)$, $np \gtrsim \log^6(n)$, $\mu = r = \kappa = L = O(1)$, $\sigma_{\min} = \Omega(n)$, $\sigma_{ij} = \Omega(1)$, and $\|V_{j,\cdot}^*\| = \|U_{i,\cdot}^*\| = \Omega(\sqrt{1/n})$. Theorem 1 would then imply that

$$\frac{M_{ij}^{\mathrm{d}} - M_{ij}^*}{s_{ij}} \xrightarrow{d} \mathcal{N}(0,1) \tag{8}$$

where s_{ij} is defined in Eq. (7). This is precisely the type of characterization we sought at the outset. The form of s_{ij}, as defined in Eq. (7), is of course critical to the characterization, and probably best understood via a few examples[2]:

1. Homogeneous Gaussian Noise. First as a sanity check, when $E_{ij} \sim \mathcal{N}(0,\sigma^2)$, Theorem 1 reduces to the same variance formula as Theorem 2 in (Chen et al. 2019):

$$s_{ij}^2 = \frac{\sigma^2(\|U_{i,\cdot}^*\|^2 + \|V_{j,\cdot}^*\|^2)}{p}. \tag{9}$$

2. Poisson Noise. When the observations are Poisson, i.e. $O_{ij} \sim \mathrm{Poisson}(M_{ij}^*)$, the variance of the noise E_{ij} is $\sigma_{ij}^2 = \mathrm{Var}(O_{ij} - M_{ij}^*) = M_{ij}^*$. Then applying Theorem 1, we have that $M_{ij}^{\mathrm{d}} - M_{ij}^* \sim \mathcal{N}(0, s_{ij}^2)$ where

$$s_{ij}^2 = \frac{\sum_{l=1}^m M_{lj}^* \left(\sum_{k=1}^r U_{ik}^* U_{lk}^*\right)^2 + \sum_{l=1}^n M_{il}^* \left(\sum_{k=1}^r V_{lk}^* V_{jk}^*\right)^2}{p}. \tag{10}$$

A special case is when $r = 1$ and $M^* = \sigma_1 u^* v^{*\top}$, for which we have

$$\begin{aligned}s_{ij}^2 &= \frac{\sum_{l=1}^m M_{lj}^* \left(u_l^* u_i^*\right)^2 + \sum_{l=1}^n M_{il}^* \left(v_l^* v_j^*\right)^2}{p} \\ &= \frac{\sum_{l=1}^m \sigma_1 u_l^* v_j^* \left(u_l^* u_i^*\right)^2 + \sum_{l=1}^n \sigma_1 u_i^* v_l^* \left(v_l^* v_j^*\right)^2}{p} \\ &= \frac{\sigma_1 v_j^* u_i^{*2}\sum_{l=1}^m u_l^{*3} + \sigma_1 u_i^* v_j^{*2}\sum_{l=1}^n v_l^{*3}}{p} \\ &= \frac{M_{ij}^*(u_i^*\|u^*\|_3^3 + v_j^*\|v^*\|_3^3)}{p},\end{aligned}$$

which corresponds to the formula in Table 1.

3. Binary Noise. Finally, binary observations occur frequently in applications. For example, in a recommender system or e-commerce platform, $O_{ij} \in \{0,1\}$ can represent whether the ith user viewed (or purchased) the jth item (or product) (Ansari and Mela 2003, Grover and Srinivasan 1987, Farias and Li 2019). A common noise model for such observations is to assume the O_{ij} are Bernoulli random variables with mean M_{ij}^*, i.e., $O_{ij} \sim \mathrm{Ber}(M_{ij}^*)$.

With such binary observations, the variance of the noise E_{ij} is $\sigma_{ij}^2 = \mathrm{Var}(O_{ij} - M_{ij}^*) = M_{ij}^*(1 - M_{ij}^*)$.

[2]See more discussion for the form of s_{ij} in Section 5.

Then s_{ij} takes the form

$$s_{ij}^2 = \frac{\sum_{l=1}^m M_{lj}^*(1-M_{lj}^*)\left(\sum_{k=1}^r U_{ik}^* U_{lk}^*\right)^2}{p} + \frac{\sum_{l=1}^n M_{il}^*(1-M_{il}^*)\left(\sum_{k=1}^r V_{lk}^* V_{jk}^*\right)^2}{p}.$$

When $r=1$ and $M^* = \sigma_1 u^* v^{*\top}$, we have

$$\begin{aligned} s_{ij}^2 &= \frac{\sum_{l=1}^m \sigma_1 u_l^* v_j^*(1-\sigma_1 u_l^* v_j^*)\left(u_l^* u_i^*\right)^2}{p} \\ &+ \frac{\sum_{l=1}^n \sigma_1 u_i^* v_l^*(1-\sigma_1 u_i^* v_l^*)\left(v_l^* v_j^*\right)^2}{p} \\ &= \frac{M_{ij}^*(u_i^*\|u^*\|_3^3 + v_j^*\|v^*\|_3^3 - M_{ij}^*(\|u^*\|_4^4 + \|v^*\|_4^4))}{p}. \end{aligned}$$

Empirical Inference: In practice, the underlying U^* and V^* are not known, and thus s_{ij} cannot be computed exactly. We propose the use of the corresponding empirical estimators to estimate s_{ij} for the purposes of inference. Let $M^{\mathrm{d}} = U^{\mathrm{d}}\Sigma^{\mathrm{d}} V^{\mathrm{d}\top}$ be the SVD of M^{d}. For example, in the Poisson noise scenario, we would use the following empirical estimator for s_{ij}:

$$\hat{s}_{ij}^2 = \frac{\sum\limits_{l=1}^m M_{lj}^{\mathrm{d}}\left(\sum\limits_{k=1}^r U_{ik}^{\mathrm{d}} U_{lk}^{\mathrm{d}}\right)^2 + \sum\limits_{l=1}^n M_{il}^{\mathrm{d}}\left(\sum\limits_{k=1}^r V_{lk}^{\mathrm{d}} V_{jk}^{\mathrm{d}}\right)^2}{p}.$$

In cases where σ_{kl} is also unknown, we let $\hat{E}_{ij} = O_{ij} - M_{ij}^{\mathrm{d}}$ be the empirical estimator for the noise. This procedure (i.e. the use of empirical estimators) can be justified via the following result:

Corollary 1. *Follow the settings in Theorem 1. Assume that $\forall (i,l), \sigma_{il} = \Theta(L)$ and*

$$s_{ij} \gtrsim L^2\mu^2 r^2 \kappa^5 \log^4(n)\left(\frac{1}{\sigma_{\min} p} + \frac{1}{mp} + \frac{1}{m^{2/3}p^{1/3}}\right).$$

Let

$$\hat{s}_{ij}^2 = \frac{\sum\limits_{l=1,(l,j)\in\Omega}^m \frac{1}{p}\hat{E}_{lj}^2\left(\sum_{k=1}^r U_{ik}^{\mathrm{d}} U_{lk}^{\mathrm{d}}\right)^2}{p} + \frac{\sum\limits_{l=1,(i,l)\in\Omega}^n \frac{1}{p}\hat{E}_{il}^2\left(\sum_{k=1}^r V_{lk}^{\mathrm{d}} V_{jk}^{\mathrm{d}}\right)^2}{p}$$

be the empirical estimator of s_{ij}. Then under the same assumptions made in Theorem 1, we have that

$$\sup_{t\in\mathbb{R}}\left|P\left\{\frac{M_{ij}^{\mathrm{d}} - M_{ij}^*}{\hat{s}_{ij}} \le t\right\} - \Phi(t)\right| = o(1).$$

Additional justification for this procedure is given as experiments later on. Note that in the typical case, $m = \Theta(n)$, $np \gtrsim \log^6(n)$, $\mu = r = \kappa = L = O(1)$, $\sigma_{\min} = \Omega(n)$, $\sigma_{ij} = \Omega(1)$, and $\|V_{j,\cdot}^*\| = \|U_{i,\cdot}^*\| = \Omega(\sqrt{1/n})$ would then imply $s_{ij} = \Omega(1/\sqrt{np})$, which is sufficient to apply Corollary 1 since the lower bound condition in Corollary 1,

$$s_{ij} = \tilde{\Omega}\left(\frac{1}{np} + \frac{1}{n^{2/3}p^{1/3}}\right),$$

is well satisfied.

Aside: When $s_{ij} \approx 0$. Curious readers may note that s_{ij} may be too small for Theorem 1 and Corollary 1 to apply. In this case, although the Gaussian approximation in Theorem 1 does not hold, an entry-wise error bound still holds, and may be sufficient for many applications (see the Appendix[3] for details):

$$|M_{ij}^{\mathrm{d}} - M_{ij}^*| = \tilde{O}\left(s_{ij} + \frac{1}{np}\right).$$

An uncertainty characterization when $s_{ij} \approx 0$ involves a second-order error analysis and remains an open question.

5. Proof Overview

In this section, we present the proof framework of Theorem 1 (see details in Appendix A). In order to extend to heterogeneous sub-exponential noise from homogeneous Gaussian, we generalize the proof of (Chen et al. 2019) with the help of recent sub-exponential matrix completion results (McRae and Davenport 2019) and a sub-exponential variant of matrix Bernstein inequality (Lemma 1).

Similar to Chen et al. (2019), our proof is based on the leave-one-out technique that has been recently used for providing breakthrough bounds for entry-wise analysis in matrix completion problems (see (Ma et al. 2018) as well as (Ding and Chen 2020, Abbe et al. 2020, Chen et al. 2020)).

We establish the following key results to characterize the decomposition of low-rank factors $(X^{\mathrm{d}}, Y^{\mathrm{d}})$, as a heterogeneous sub-exponential generalization of Theorem 5 in Chen et al. (2019).

Theorem 2. *Assume $mp \gg \kappa^4\mu^2 r^2 \log^3 n$ and $L\log(n)\sqrt{\frac{n}{p}} \ll \frac{\sigma_{\min}}{\sqrt{\kappa^4\mu r\log n}}$. There exists a rotation matrix $H^{\mathrm{d}} \in \mathrm{O}^{r\times r}$ and $\Phi_X \in \mathbb{R}^{m\times r}, \Phi_Y \in \mathbb{R}^{n\times r}$ such*

[3]See Appendix in the online full version `https://arxiv.org/abs/2110.12046`.

that the following holds with probability $1 - O(n^{-10})$,

$$X^{\mathrm{d}} H^{\mathrm{d}} - X^* = \frac{1}{p} P_\Omega(E) Y^* (Y^{*T} Y^*)^{-1} + \Phi_X$$
$$Y^{\mathrm{d}} H^{\mathrm{d}} - Y^* = \frac{1}{p} P_\Omega(E)^T X^* (X^{*T} X^*)^{-1} + \Phi_Y$$

where

$$\max\left\{\|\Phi_X\|_{2,\infty}, \|\Phi_Y\|_{2,\infty}\right\} \lesssim \frac{L\log n}{\sqrt{p}\sigma_{\min}}\left(\frac{L\log n}{\sigma_{\min}}\sqrt{\frac{\kappa^9 \mu r n \log n}{p}} + \sqrt{\frac{\kappa^7 \mu^3 r^3 \log^2 n}{mp}}\right).$$

Proof. At a high level, the proof of Theorem 2 follows a similar proof of Theorem 5 in (Chen et al. 2019), but with replacements that employ more fine-grained analyses of E for whenever the Gaussianity of E is used in (Chen et al. 2019). These analyses aim to address the sub-exponentiality and heterogeneity of E, with the help of the following two lemmas.

Lemma 1. *Given* k *independent random* $m_1 \times m_2$ *matrices* $X_1, X_2, \dots, X_k$ *with* $E[X_i] = 0$. *Let*

$$V := \max\left(\left\|\sum_{i=1}^{k} E[X_i X_i^T]\right\|, \left\|\sum_{i=1}^{k} E[X_i^T X_i]\right\|\right).$$

Suppose $\|\|X_i\|\|_{\psi_1} \le B$ *for* $i \in [k]$. *Then,*

$$\|X_1 + X_2 + \ldots + X_k\| \lesssim \sqrt{V\log(k(m_1+m_2))} + B\log(k(m_1+m_2))\log(k)$$

with probability $1 - O(k^{-c})$ *for any constant* c.

Lemma 2. *Suppose* $E \in R^{m\times n}$ $(m \le n)$ *whose entries are independent and centered. Suppose* $\|E_{ij}\|_{\psi_1} \le L$ *for any* $(i,j) \in [m]\times[n]$. *Let* $\Omega \in [m]\times[n]$ *be the subset of indices where each index* (i,j) *is included in* Ω *independently with probability* p. *Suppose* $np \ge c_0 \log^3 n$ *for some sufficient large constant* c_0. *then, with probability* $1 - O(n^{-11})$,

$$\left\|\frac{1}{p} P_\Omega(E)\right\| \le CL\sqrt{\frac{n}{p}}.$$

Here, Lemma 1 is a generalization of matrix Bernstein inequality in Theorem 6.1.1 of (Tropp et al. 2015). Lemma 2 is an implication of Lemma 4 in (McRae and Davenport 2019).

Equipped with Lemmas 1 and 2, the desired bounds for sub-exponential E can be established. Following we provide an example of using Lemma 1 to bound $\|X^{*T} P_\Omega(E) Y^*\|$, which is critical for obtaining the bounds in Theorem 2.

To begin, note that

$$X^{*T} P_\Omega(E) Y^* = \sum_{k=1}^{m}\sum_{l=1}^{n} X^*_{k,\cdot} Y^{*\top}_{l,\cdot} \delta_{k,l} E_{k,l}$$

where $\delta_{k,l} \sim \mathrm{Ber}(p)$ indicates whether $(k,l) \in \Omega$. Let $A_{k,l} := X^*_{k,\cdot} Y^{*\top}_{l,\cdot} \delta_{k,l} E_{k,l}$ for $k \in [m], l \in [n]$. Then,

$$\|X^{*T} P_\Omega(E) Y^*\| = \left\|\sum_{k,l} A_{k,l}\right\|.$$

Note that $A_{k,l} \in \mathbb{R}^{r\times r}$ are independent zero-mean random matrices and we aim to invoke Lemma 1 to bound $\|\sum_{k,l} A_{k,l}\|$. Let

$$V := \max\left(\left\|\sum_{k,l} E[A_{k,l} A_{k,l}^\top]\right\|, \left\|\sum_{k,l} E[A_{k,l}^\top A_{k,l}]\right\|\right)$$
$$B := \max_{k,l} \|\|A_{k,l}\|\|_{\psi_1}.$$

Note that

$$\begin{aligned}\left\|\sum_{k,l} E[A_{k,l} A_{k,l}^\top]\right\| &\le \sum_{k,l} \|E[A_{k,l} A_{k,l}^\top]\| \\ &= \sum_{k,l} \sigma_{k,l}^2 p \|X^*_{k,\cdot}\|^2 \|Y^*_{l,\cdot}\|^2 \\ &\overset{(i)}{\le} 2L^2 p \|X^*\|_{\mathrm{F}}^2 \|Y^*\|_{\mathrm{F}}^2\end{aligned}$$

where in (i) we use the fact that $E(x^2) \le 2\|x\|_{\psi_1}^2$ for an sub-exponential zero-mean random variable x. Similarly, the bounds can be established for $\|\sum_{k,l} E[A_{k,l}^\top A_{k,l}]\|$. Hence $V \le 2L^2 p \|X^*\|_{\mathrm{F}}^2 \|Y^*\|_{\mathrm{F}}^2$.

Then, consider

$$\begin{aligned} B &:= \max_{k,l} \|\|A_{k,l}\|\|_{\psi_1} \\ &\le \max_{k,l} \|E_{k,l}\delta_{k,l}\|_{\psi_1} \|X^*\|_{2,\infty} \|Y^*\|_{2,\infty} \\ &\overset{(i)}{\le} L\|X^*\|_{2,\infty} \|Y^*\|_{2,\infty}\end{aligned}$$

where (i) we use that $\|E_{k,l}\delta_{k,l}\|_{\psi_1} \le \|E_{k,l}\|_{\psi_1} \le L$. Then apply Lemma 1, with probability $1 - O(n^{-11})$, we obtain the desired bound for $\|X^{*\top} P_\Omega(E) Y^*\|$

$$\begin{aligned}\left\|\sum_{k,l} A_{k,l}\right\| &\lesssim \sqrt{V\log(n)} + B\log^2(n) \\ &\overset{(i)}{\lesssim} \sqrt{p} L \sigma_{\max} r \sqrt{\log(n)} + \frac{\mu r}{n} L \sigma_{\max} \log^2(n) \\ &\overset{(ii)}{\lesssim} \sqrt{p} L \sigma_{\max} r \sqrt{\log(n)}\end{aligned}$$

where in (i) we use that $\|X^*\|_F^2 = \|Y^*\|_F^2 \leq \sigma_{\max} r$ and the incoherence condition Eq. (1), in (ii) we use that $mp \gg \kappa^4\mu^2 r^2 \log^3 n$.

To establish a similar bound for $\|X^{*\top} P_\Omega(E) Y^*\|$, Chen et al. (2019) uses the Gaussianity of E, which is not applicable here.

Similar to this example, we apply Lemma 1 and Lemma 2 with more fine grained analyses to address the sub-exponentiality and heterogeneity of E. See Appendix A.3 for full details.

□

Note that the error of $M^{\mathrm{d}} - M^*$ is closely related to the errors of low-rank factors $X^{\mathrm{d}}H^{\mathrm{d}} - X^*, Y^{\mathrm{d}}H^{\mathrm{d}} - Y^*$ through the following

$$\begin{aligned}M^{\mathrm{d}} - M^* &= X^{\mathrm{d}}H^{\mathrm{d}}H^{\mathrm{d}\top}Y^{\mathrm{d}\top} - X^*Y^{*\top} \\ &= (X^{\mathrm{d}}H^{\mathrm{d}} - X^*)Y^{*\top} + X^*(Y^{\mathrm{d}}H^{\mathrm{d}} - Y^*)^\top \\ &\quad - (X^{\mathrm{d}}H^{\mathrm{d}} - X^*)(Y^{\mathrm{d}}H^{\mathrm{d}} - Y^*)^\top \\ &\overset{(i)}{\approx} (X^{\mathrm{d}}H^{\mathrm{d}} - X^*)Y^{*\top} + X^*(Y^{\mathrm{d}}H^{\mathrm{d}} - Y^*)^\top\end{aligned}$$

where in (i) we ignore the second-order error term $(X^{\mathrm{d}}H^{\mathrm{d}} - X^*)(Y^{\mathrm{d}}H^{\mathrm{d}} - Y^*)^\top$. Note that Theorem 2 implies that

$$\begin{aligned}X^{\mathrm{d}}H^{\mathrm{d}} - X^* &\approx \frac{1}{p}P_\Omega(E)Y^*(Y^{*T}Y^*)^{-1} \\ Y^{\mathrm{d}}H^{\mathrm{d}} - Y^* &\approx \frac{1}{p}P_\Omega(E)^\top X^*(X^{*T}X^*)^{-1}\end{aligned}$$

Plug this into the decomposition of $M^{\mathrm{d}} - M^*$, we have

$$\begin{aligned}M^{\mathrm{d}} - M^* &\approx \frac{1}{p}P_\Omega(E)Y^*(Y^{*T}Y^*)^{-1}Y^{*\top} \\ &\quad + \frac{1}{p}X^*(X^{*T}X^*)^{-1}X^{*\top}P_\Omega(E) \\ &= \frac{1}{p}P_\Omega(E)V^*V^{*\top} + \frac{1}{p}U^*U^{*\top}P_\Omega(E).\end{aligned}$$

The results of Theorem 1 then follow from the above approximation and the use of Berry-Esseen type of inequalities.[4] In particular,

$$\begin{aligned}s_{ij}^2 &= Var(M_{ij}^d - M_{ij}^*) \\ &\overset{(i)}{\approx} \frac{1}{p}Var(e_i^\top P_\Omega(E)V^*V^{*\top}e_j) \\ &\quad + \frac{1}{p}Var(e_i^\top U^*U^{*\top}P_\Omega(E)e_j) \\ &= \frac{\sum_{l=1}^m \sigma_{lj}^2\left(\sum_{k=1}^r U_{ik}^*U_{lk}^*\right)^2 + \sum_{l=1}^n \sigma_{il}^2\left(\sum_{k=1}^r V_{lk}^*V_{jk}^*\right)^2}{p}\end{aligned}$$

[4]Note that this decomposition also can be interpreted geometrically (to the first-order) as the projection of $P_\Omega(E)$ to the tangent space of M^* in the rank-r manifold.

where (i) assumes the near-independence of the corresponding terms. See Appendix A.5 for full rigorous details.

6. Experiments

We evaluate the results in Theorem 1 for synthetic data under multiple settings. We then compare the performances of various uncertainty quantification formulas in real data.[5]

Synthetic Data. We generate an ensemble of instances. Each instance consists of a few parameters: (i) (m, n): the size of M^*; (ii) r: the rank of M^*; (iii) p: the probability of an entry being observed; (iv) $\bar{M}^*$: the entry-wise mean of M^* ($\bar{M}^* = \frac{1}{mn}\sum_{ij} M_{ij}^*$).

Given $(m, n, r, p, \bar{M}^*)$, we follow the typical procedures of generating random non-negative low-rank matrices in (Cemgil 2008, Farias et al. 2021b). Each instance is generated in two steps: (i) Generate M^*: let $U^* \in \mathbb{R}^{m\times r}, V^* \in \mathbb{R}^{n\times r}$ be random matrices with independent entries from Gamma$(2, 1)$. Set $M^* = kU^*V^{*\top}$ where $k \in \mathbb{R}$ is picked such that $\frac{1}{mn}\sum_{ij} M_{ij}^* = \bar{M}^*$. (ii) Generate $P_\Omega(O)$: then O_{ij} = Poisson(M_{ij}^*) and entries in Ω is sampled independently with probability p.[6]

We first verify the entry-wise distributional characterization $M_{ij}^{\mathrm{d}} - M_{ij}^* \sim \mathcal{N}(0, s_{ij}^2)$ where s_{ij} is specified in Eq. (7). See a demonstration of the Gaussian approximality of the empirical distribution $(M_{ij}^{\mathrm{d}} - M_{ij}^*)/s_{ij}$ in Fig. 1. Given an instance, we compute the coverage rate (the percentage of coverage of entries) that correponds to the 95% confidence interval, where an "coverage" of an entry (i, j) occurs if

$$M_{ij}^{\mathrm{d}} \in [M_{ij}^* - 1.96s_{ij}, M_{ij}^* + 1.96s_{ij}].$$

The average coverage rates under different settings are shown in Table 2. The closeness of the results (ranging from 91% − 95%) to the "true" coverage rate 95% suggests the applicability of inference based on our variance formula. The trends in Table 2 are also consistent with the intuition: the performance starts to degrade when r increases, p decreases, and the noise to signal ratio increases (decrease of $\bar{M}^*$).

Real Data. Next, we study a real dataset consisting of daily sales for 1115 units with 942 days (Rossmann 2021). To compare different uncertainty quantification formulas, we consider the *coverage rate maximization*

[5]The source code is provided in `https://github.com/TianyiPeng/Uncertainty-Quantification-For-Low-Rank-Matrix-Completion`.

[6]Here, we focus on the results of Poisson noise, where the results under the binary noise are similar in the experiments.

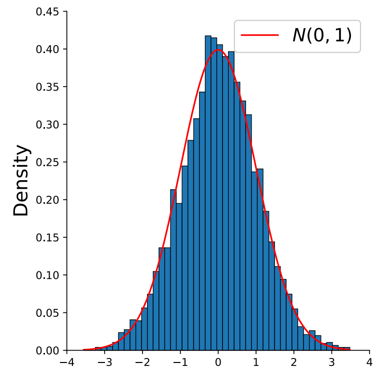

Figure 1: Empirical distribution of $(M^{\mathrm{d}}_{11} - M^*_{11})/s_{11}$ with $m = n = 300, r = 2, p = 0.6$, and $\bar{M}^* = 20$.

$(r, p, \bar{M}^*)$	**Coverage Rate**
(3, 0.3, 5)	0.936 ($\pm$ 0.003)
(3, 0.3, 20)	0.945 ($\pm$ 0.004)
(3, 0.6, 5)	0.947 ($\pm$ 0.003)
(3, 0.6, 20)	0.949 ($\pm$ 0.003)
(6, 0.3, 5)	0.910 ($\pm$ 0.002)
(6, 0.3, 20)	0.934 ($\pm$ 0.002)
(6, 0.6, 5)	0.934 ($\pm$ 0.003)
(6, 0.6, 20)	0.943 ($\pm$ 0.003)

Table 2: Coverage rates for different $(r, p, \bar{M}^*)$ with $m = n = 500$. The empirical mean and empirical standard deviation are reported over 100 instances.

task that aims to maximize the coverage rate given the total interval length constraint.

Coverage Rate Maximization. In particular, given a uncertainty quantification formula, suppose one can provide an interval predictor $[a_{ij}, b_{ij}]$ for each entry (i, j) in a set $\tilde{\Omega}$. The "coverage" of (i, j) occurs if $M_{ij} \in [a_{ij}, b_{ij}]$ where M_{ij} is the true value of entry (i, j). The task aims to maximize the coverage rate given that the total length of intervals $\sum_{(i,j)\in\tilde{\Omega}} b_{ij} - a_{ij}$ is constrained by a budget threshold α:

$$\begin{aligned} \text{maximize} \quad & \frac{\sum_{(i,j)\in\tilde{\Omega}} \mathbb{1}(M_{ij} \in [a_{ij}, b_{ij}])}{|\tilde{\Omega}|} \\ \text{subject to} \quad & \sum_{(i,j)\in\tilde{\Omega}} b_{ij} - a_{ij} \le \alpha \end{aligned} \tag{11}$$

We are interested in comparing the performances of the above task using different variance predictors s_{ij}, either provided by Eq. (9) with the homogeneous Gaussian noise assumption (Theorem 2 in Chen et al. (2019)), or by our Theorem 1, capable of addressing the heterogeneous sub-exponential noise. Note that both results in Chen et al. (2019) and our Theorem 1 predict that $M^{\mathrm{d}}_{ij} \sim \mathcal{N}(M^*_{ij}, s^2_{ij})$. With this distributional assumption, we tackle Eq. (11) by a greedy algorithm that achieves the maximal expected coverage rate with the budget constraint. Specifically, with given $\{M^{\mathrm{d}}_{ij}, s_{ij}\}$, we provide the interval predictors $\{[a_{ij}, b_{ij}]\}$ by solving the following problem:

$$\begin{aligned} \underset{a_{ij}, b_{ij}}{\text{maximize}} \quad & \frac{\sum_{(i,j)\in\tilde{\Omega}} \mathbb{E}_{M_{ij}\sim\mathcal{N}(M^{\mathrm{d}}_{ij}, s^2_{ij})}(M_{ij} \in [a_{ij}, b_{ij}])}{|\tilde{\Omega}|} \\ \text{subject to} \quad & \sum_{(i,j)\in\tilde{\Omega}} b_{ij} - a_{ij} \le \alpha \end{aligned}$$

Experiment Results. In the experiment, the low-rankness of the dataset is verified and the "true" rank, as well as the "true" underlying matrix M, is predetermined through the spectrum of singular value decomposition.

We split the entries uniformly into a training set Ω (with probability p) and a test set $\tilde{\Omega}$. We use the observations in Ω to learn M^{d} with Algorithm 1. Let $M^{\mathrm{d}} = U^{\mathrm{d}}\Sigma^{\mathrm{d}}V^{\mathrm{d}\top}$ be the SVD of M^{d}.

The empirical variance $s^{\mathrm{Gaussian}}_{ij}$ for homogeneous Gaussian noise is computed by (Chen et al. 2019)

$$(s^{\mathrm{Gaussian}}_{ij})^2 = \frac{\hat{\sigma}^2(\|U^{\mathrm{d}}_{i,\cdot}\|^2 + \|V^{\mathrm{d}}_{j,\cdot}\|^2)}{p}$$

where $\hat{\sigma}^2 := \sum_{(i,j)\in\Omega}(O_{ij} - M^{\mathrm{d}}_{ij})^2/|\Omega|$ is the empirical estimator for the noise variance.

We then compute the empirical variance $s^{\mathrm{Poisson}}_{ij}$ for Poisson noise

$$(s^{\mathrm{Poisson}}_{ij})^2 = \frac{\sum_{l=1}^{m} M^{\mathrm{d}}_{lj}\left(\sum_{k=1}^{r} U^{\mathrm{d}}_{ik}U^{\mathrm{d}}_{lk}\right)^2 + \sum_{l=1}^{n} M^{\mathrm{d}}_{il}\left(\sum_{k=1}^{r} V^{\mathrm{d}}_{lk}V^{\mathrm{d}}_{jk}\right)^2}{p}.$$

Given $M^{\mathrm{d}}, s^{\mathrm{Poisson}}_{ij}$ and $s^{\mathrm{Gaussian}}_{ij}$, the coverage rate maximization task is evaluated in the test set $\tilde{\Omega}$. The results for various budgets α are reported in Fig. 2. The Poisson noise formula shows a higher coverage rate than the homogeneous Gaussian formula, as the former is more robust to addressing heterogeneous noises in sales data. This improvement tends to vanish with more presences of missing entries, which might be due to the degrading accuracy of matrix completion and variance estimation when p decreases.

7. Conclusion

We solved the uncertainty quantification problem for matrix completion with heterogeneous and sub-exponential noise. The error variance of a common

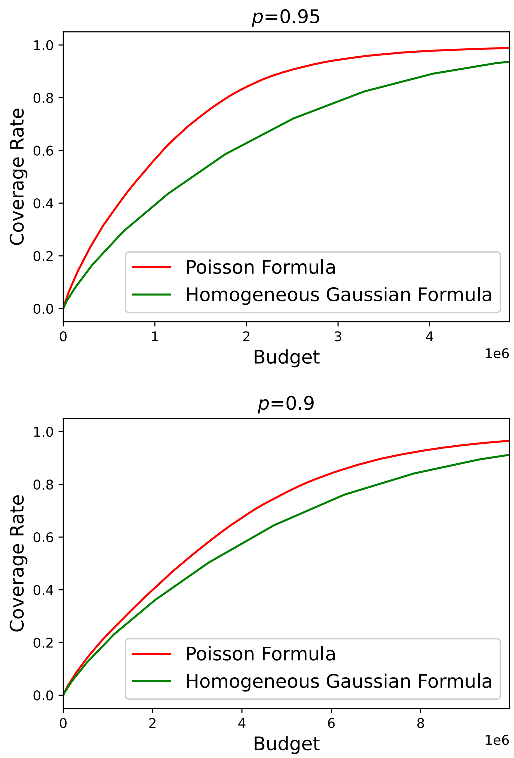

Figure 2: Coverage rates for difference variance formulas correspond to the total-interval-length budget.

estimator was determined and the asymptotical normality with inference results were established. The explicit formulas for various scenarios such as Poisson noise and Binary noise were analyzed. Experimental results showed significant improvements of our new uncertainty quantification formulas over existing ones.

One exciting direction for further work is in assuming less restrictive Ω. As in most of the matrix completion literature, we made the uniform sampling assumption for Ω, which may not be applicable in some practical applications. The study of uncertainty quantification for matrix completion with non-uniform sampling patterns is especially valuable, given the recent progress on deterministic matrix completion, e.g., (Chatterjee 2020).

There are still many exciting questions awaiting for exploration. To name a few,

- **Generalization to different loss functions.** The current form of uncertainty is tailored to the estimator under l_2 loss, the generalization to uncertainty quantification under different loss functions is of great interest. This generalization might require some non-trivial and more refined extensions of the current framework (e.g. showing an entry-wise guarantee for new types of estimators). We also note that there is a concurrent work Chen et al. (2021) tackling the statistical inference for 1-bit matrix completion under a special linear case $M^*_{ij} = a_i + b_j$, which significantly simplifies the problem but could provide useful insights.
- **Dependence on κ.** Optimizing the dependence on κ is an interesting direction to study, particularly given the recent progress on the noiseless matrix completion problem (e.g., Tong et al. (2021) proposed an alternating scaled gradient descent method with the iteration steps independent of κ).
- **Less restrictive Ω.** As most of the matrix completion literature, we made the uniform sampling assumption for Ω, which may not be applicable in some practical applications. We think the study of uncertainty quantification for matrix completion with non-uniform sampling patterns is especially valuable, given the recent progress on deterministic matrix completion, e.g., (Chatterjee 2020).
- **Correlated noise E.** The generalization from independent noises to scenarios that noises may be correlated is also of practical interest. Since leave-one-out techniques seemingly need the independence assumption, new techniques beyond leave-one-out may need to be developed.

Acknowledgements

We thank the anonymous reviewers for their valuable comments in AISTATS 2022. Andrew A. Li is supported by a BP America Junior Faculty Chair.

References

Abbe E, Fan J, Wang K, Zhong Y, et al. (2020) Entrywise eigenvector analysis of random matrices with low expected rank. *Annals of Statistics* 48(3):1452–1474.

Alquier P, et al. (2015) A bayesian approach for noisy matrix completion: Optimal rate under general sampling distribution. *Electronic Journal of Statistics* 9(1):823–841.

Amjad MJ, Shah D (2017) Censored demand estimation in retail. *Proceedings of the ACM on Measurement and Analysis of Computing Systems* 1(2):1–28.

Ansari A, Mela CF (2003) E-customization. *Journal of marketing research* 40(2):131–145.

Candes EJ, Plan Y (2010) Matrix completion with noise. *Proceedings of the IEEE* 98(6):925–936.

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

Carpentier A, Eisert J, Gross D, Nickl R (2015) Uncertainty quantification for matrix compressed sensing and quantum tomography problems. *arXiv preprint arXiv:1504.03234* .

Carpentier A, Klopp O, Löffler M, Nickl R (2018) Adaptive confidence sets for matrix completion. *Bernoulli* 24(4A):2429–2460.

Cemgil AT (2008) Bayesian inference for nonnegative matrix factorisation models. *Computational intelligence and neuroscience* 2009.

Chatterjee S (2020) A deterministic theory of low rank matrix completion. *IEEE Transactions on Information Theory* 66(12):8046–8055.

Chen Y, Chi Y, Fan J, Ma C, Yan Y (2020) Noisy matrix completion: Understanding statistical guarantees for convex relaxation via nonconvex optimization. *SIAM journal on optimization* 30(4):3098–3121.

Chen Y, Fan J, Ma C, Yan Y (2019) Inference and uncertainty quantification for noisy matrix completion. *Proceedings of the National Academy of Sciences* 116(46):22931–22937.

Chen Y, Li C, Xu G (2021) A note on statistical inference for noisy incomplete 1-bit matrix. *arXiv preprint arXiv:2105.01769* .

Cribari-Neto F, Maria da Glória AL (2014) New heteroskedasticity-robust standard errors for the linear regression model. *Brazilian Journal of Probability and Statistics* 28(1):83–95.

Ding L, Chen Y (2020) Leave-one-out approach for matrix completion: Primal and dual analysis. *IEEE Transactions on Information Theory* 66(11):7274–7301.

Dong S, Absil PA, Gallivan K (2021) Riemannian gradient descent methods for graph-regularized matrix completion. *Linear Algebra and its Applications* 623:193–235.

Eckart C, Young G (1936) The approximation of one matrix by another of lower rank. *Psychometrika* 1(3):211–218.

Esseen CG (1942) On the liapunov limit error in the theory of probability. *Ark. Mat. Astr. Fys.* 28:1–19.

Farias V, Li A, Peng T (2021a) Learning treatment effects in panels with general intervention patterns. *Advances in Neural Information Processing Systems* 34.

Farias V, Li AA, Peng T (2021b) Near-optimal entrywise anomaly detection for low-rank matrices with subexponential noise. *International Conference on Machine Learning*, 3154–3163 (PMLR).

Farias VF, Li AA (2019) Learning preferences with side information. *Management Science* 65(7):3131–3149.

Fazayeli F, Banerjee A, Kattge J, Schrodt F, Reich PB (2014) Uncertainty quantified matrix completion using bayesian hierarchical matrix factorization. *2014 13th International Conference on Machine Learning and Applications*, 312–317 (IEEE).

Ge R, Lee JD, Ma T (2016) Matrix completion has no spurious local minimum. *arXiv preprint arXiv:1605.07272* .

Grover R, Srinivasan V (1987) A simultaneous approach to market segmentation and market structuring. *Journal of Marketing Research* 24(2):139–153.

Hayes AF, Cai L (2007) Using heteroskedasticity-consistent standard error estimators in ols regression: An introduction and software implementation. *Behavior research methods* 39(4):709–722.

Imbens GW, Kolesar M (2016) Robust standard errors in small samples: Some practical advice. *Review of Economics and Statistics* 98(4):701–712.

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

Keshavan RH, Montanari A, Oh S (2010) Matrix completion from a few entries. *IEEE transactions on information theory* 56(6):2980–2998.

Lakshminarayanan B, Pritzel A, Blundell C (2016) Simple and scalable predictive uncertainty estimation using deep ensembles. *arXiv preprint arXiv:1612.01474* .

Long JS, Ervin LH (2000) Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician* 54(3):217–224.

Ma C, Wang K, Chi Y, Chen Y (2018) Implicit regularization in nonconvex statistical estimation: Gradient descent converges linearly for phase retrieval, matrix completion, and blind deconvolution. *Foundations of Computational Mathematics* 1–182.

Mazumder R, Hastie T, Tibshirani R (2010) Spectral regularization algorithms for learning large incomplete matrices. *The Journal of Machine Learning Research* 11:2287–2322.

McRae AD, Davenport MA (2019) Low-rank matrix completion and denoising under poisson noise. *arXiv preprint arXiv:1907.05325* .

Natarajan N, Dhillon IS (2014) Inductive matrix completion for predicting gene–disease associations. *Bioinformatics* 30(12):i60–i68.

Rossmann (2021) Rossmann store sales. `https://www.kaggle.com/c/rossmann-store-sales`, accessed: 2021-05-26.

Salakhutdinov R, Mnih A (2008) Bayesian probabilistic matrix factorization using markov chain monte carlo. *Proceedings of the 25th international conference on Machine learning*, 880–887.

Shi J, Katehakis MN, Melamed B, Xia Y (2014) Production-inventory systems with lost sales and compound poisson demands. *Operations Research* 62(5):1048–1063.

So AMC, Ye Y (2007) Theory of semidefinite programming for sensor network localization. *Mathematical Programming* 109(2):367–384.

Stewart GW (1977) On the perturbation of pseudo-inverses, projections and linear least squares problems. *SIAM review* 19(4):634–662.

Su X, Khoshgoftaar TM (2009) A survey of collaborative filtering techniques. *Advances in artificial intelligence* 2009.

Sun R, Luo ZQ (2016) Guaranteed matrix completion via non-convex factorization. *IEEE Transactions on Information Theory* 62(11):6535–6579.

Tanaka M (2021) Bayesian matrix completion approach to causal inference with panel data. *Journal of Statistical Theory and Practice* 15(2):1–22.

Tanner J, Wei K (2016) Low rank matrix completion by alternating steepest descent methods. *Applied and Computational Harmonic Analysis* 40(2):417–429.

Tong T, Ma C, Chi Y (2021) Low-rank matrix recovery with scaled subgradient methods: Fast and robust convergence without the condition number. *IEEE Transactions on Signal Processing* 69:2396–2409.

Tropp JA, et al. (2015) An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning* 8(1-2):1–230.

Vershynin R (2018) *High-dimensional probability: An introduction with applications in data science*, volume 47 (Cambridge university press).

Xia D, Yuan M (2021) Statistical inferences of linear forms for noisy matrix completion. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 83(1):58–77.

Zeldes Y, Theodorakis S, Solodnik E, Rotman A, Chamiel G, Friedman D (2017) Deep density networks and uncertainty in recommender systems. *arXiv preprint arXiv:1711.02487* .

Zhang AR, Cai TT, Wu Y (2018) Heteroskedastic pca: Algorithm, optimality, and applications. *arXiv preprint arXiv:1810.08316* .

Zhu Z, Li Q, Tang G, Wakin MB (2017) The global optimization geometry of low-rank matrix optimization. *arXiv preprint arXiv:1703.01256* .