Simulation comparisons between Bayesian and de-biased estimators in low-rank matrix completion

The Tien Mai (✉ t.t.mai@medisin.uio.no)

University of Oslo <https://orcid.org/0000-0002-3514-9636>

Research Article

Keywords: Low-rank matrix, matrix completion, Bayesian method, de-biased estimator, confidence interval

Posted Date: July 6th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-660262/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Simulation comparisons between Bayesian and de-biased estimators in low-rank matrix completion

The Tien Mai^a

^a Oslo Centre for Biostatistics and Epidemiology, Department of Biostatistics, University of Oslo, Norway.

ARTICLE HISTORY

Compiled June 26, 2021

ABSTRACT

In this paper we perform numerous numerical studies for the problem of low-rank matrix completion. We compare the Bayesian approaches and a recently introduced de-biased estimator which provides a useful way to build confidence intervals of interest. From a theoretical viewpoint, the de-biased estimator comes with a sharp minimax-optimal rate of estimation error whereas the Bayesian approach reaches this rate with an additional logarithmic factor. Our simulation studies show originally interesting results that the de-biased estimator is just as strong as the Bayesian estimators. Moreover, Bayesian approaches are much more stable and can outperform the de-biased estimator in the case of small samples. In additions, we also find that the coverage rate of the confidence intervals revealed by the de-biased estimator for an entry is absolutely lower than of the considered credible interval. These suggest further theoretical studies on the estimation error and the concentration for Bayesian methods as they are being quite limited up to present.

KEYWORDS

Low-rank matrix; matrix completion; Bayesian method; de-biased estimator; uncertainty quantification; confidence interval.

1. Introduction

The goal of low-rank matrix completion is to recover a low-rank matrix from its partially observed entries. This problem has recently received and increased attention due to the emergence of several challenging applications, such as recommender systems [1,2] (particularly the famous Netflix challenge [3]), genotype imputation [4,5], image processing [6–8] and quantum state tomography [9–11]. Different approaches from frequentist to Bayesian methods have been conducted in this problem, both from theoretical and computational point of views, see for example [12–25].

From a frequentist point of view, most of the recent methods are usually based on penalized optimization. A seminal result can be found in [13,14] for exact matrix completion (noiseless case) and further developed in the noisy case in [12,15,26]. Some efficient algorithms had also been studied, for example see [16,27,28]. More particularly, in the notable work [15], the authors studied nuclear-norm penalized estimators and provided reconstruction errors for their methods. Moreover, they also showed that these errors are minimax-optimal (up to a logarithmic factor). Note that the average

quadratic error on the entries of a rank- r matrix size $m \times p$ from n -observations can not be better than: $r \max(m, p)/n$ [15].

More recently, in a remarkable work [24], de-biased estimators have been proposed for the problem of noisy low-rank matrix completion. The estimation accuracy of this estimator is shown to be sharp in the sense that it reaches the minimax-optimal rate without any additional logarithmic factor. A sharp bound has also been obtained by a different estimator in [29], however its statistical inference is not known up to our knowledge. More importantly, the confident intervals on the reconstruction of entries of the underlying matrix are also provided by using the de-biased estimators in the work by [24]. It is noted that conducting uncertainty quantification for matrix completion is not straightforward. This is because, in general, the solutions for matrix completion do not admit closed-form and the distributions of the estimates returned by the state-of-the-art algorithms are hard to derive.

On the other hand, uncertainty quantification can be assessed normally from a Bayesian perspective that is the unknown matrix is considered as a random variable through a prior distribution and the statistical inference can be made by the posterior distribution, for example the credible intervals. Bayesian methods have been also studied for low-rank matrix completion mainly from a computational viewpoint [18–21, 30–33]. Most Bayesian estimators are based on conjugate priors which allow to use Gibbs sampling [20, 30] or Variational Bayes methods [19] and these algorithms are fast enough to deal with and actually tested on large datasets like Netflix [3] or MovieLens [34]. However, the theoretical understanding of Bayesian estimators is quite limited, up to our knowledge, [22] and [25] are the only prominent examples. More specifically, they showed that a Bayesian estimator with a low-rank factorization prior reaches the minimax-optimal rate up to a logarithmic factor and the paper [25] further show that the same rate can be obtained by using a Variational Bayesian estimator.

In this paper, to understand the performance of Bayesian approaches vs. the de-biased estimators, we perform various numerical comparisons on the estimation accuracy (the mean square error, the normalized mean square error and the prediction error, see Section 3) between the de-biased estimator in [24] and the Bayesian methods [25] for which the statistical properties have been well studied thoroughly. Furthermore, we examine in details the behaviour of the confidence intervals revealed by the de-biased estimator against the Bayesian credible intervals. Interestingly, it is noted that recent works [35, 36] shows that Bayesian methods are now the most accurate method in practical recommender systems. Although Bayesian method has become common in matrix completion, its uncertainty quantification in this problem (e.g. credible interval for recovering entries of the matrix) has received much more limited attention in the literature.

Results from simulation comparisons release originally interesting messages. More specifically, the de-biased estimator is just as good as the Bayesian estimators on the estimation accuracy, although it is completely successful in improving the ancestry estimator that being de-biased. On the other hand, the Bayesian approaches are much more stable than the de-biased method and in addition they outperform the de-biased estimator especially in case of small samples. Moreover, we find that the coverage rates of the 95% confidence intervals disclosed by the de-biased estimator are lower than the 89% equal-tailed credible intervals. These evidences suggest that the Bayesian estimators may actually reach the minimax-optimal rate sharply and the log-term could be due to the technical proofs (the PAC-Bayesian bounds technique). Furthermore, the concentration rate of the corresponding Bayesian posterior showed in [25] with a log-term could be sharp.

The rest of the paper is structured as follows. In Section 2 we present the low-rank matrix completion problem, then introduce the de-biased estimator as well as the disclosed confidence interval and provide details on the considered Bayesian estimators. In Section 3, simulation studies comparing the different methods are presented. We discuss and conclude our results in the final section.

2. Low-rank matrix completion

2.1. Model

Let $M^* \in \mathbb{R}^{m \times p}$ be an unknown rank- r matrix of interest. We observe a random subset of noisy entries of M^* as

$$Y_{ij} = M_{ij}^* + \mathcal{E}_{ij}, \quad (i, j) \in \Omega \quad (1)$$

where Ω is a subset of indices $\{1, \dots, m\} \times \{1, \dots, p\}$ and $\mathcal{E}_{ij} \sim \mathcal{N}(0, \sigma^2)$ are independently generated noise at the location (i, j) . We assume that the data are missing uniformly at random. Then, the problem of estimating M^* with $n = |\Omega| < mp$ is called the (noisy) low-rank matrix completion problem.

Let $\mathcal{P}_\Omega(\cdot) : \mathbb{R}^{m \times p} \mapsto \mathbb{R}^{m \times p}$ be the orthogonal projection onto the observed entries in the index set Ω that

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

Notations: For a matrix $A \in \mathbb{R}^{m \times p}$, $\|A\|_F = \sqrt{\text{trace}(A^\top A)}$ denotes its Frobenius norm and $\|A\|_* = \text{trace}(\sqrt{A^\top A})$ denotes its nuclear norm. $[a \pm b]$ is used to denote the interval $[a - b, a + b]$. We use Id_q to denote the identity matrix of dimension $q \times q$.

2.2. The de-biased estimator

Let $\hat{M}$ be either the solution of the following nuclear norm regularization [27]

$$\min_{Z \in \mathbb{R}^{m \times p}} \frac{1}{2} \|\mathcal{P}_\Omega(Z - Y)\|_F^2 + \lambda \|Z\|_*,$$

or of the following factorization minimization [28]

$$\min_{U \in \mathbb{R}^{m \times r}, V \in \mathbb{R}^{p \times r}} \frac{1}{2} \|\mathcal{P}_\Omega(Y - UV^\top)\|_F^2 + \frac{\lambda}{2} \|U\|_F^2 + \frac{\lambda}{2} \|V\|_F^2, \quad (2)$$

where $\lambda > 0$ is a tuning parameter.

Given an estimator $\hat{M}$ as above, the de-biased estimator [24] is defined as

$$M^{db} := \text{Pr}_{\text{rank}-r} \left[\hat{M} - \mathcal{P}_\Omega(\hat{M} - Y) \right], \quad (3)$$

where $\text{Pr}_{\text{rank}-r}(B) = \arg \min_{A: \text{rank}(A) \leq r} \|A - B\|_F$ is the projection onto the set of rank- r matrices.

Remark 1. The estimation accuracy of the de-biased estimator, provided in Theorem 3 in [24] under some assumptions, is that $\|M^{db} - M^*\|_F^2 \leq c \max(m, p) r \sigma^2 / n$ without any extra log-term and c is universal numerical constant.

2.2.1. Confidence interval

Let $\hat{M} = \hat{U} \hat{\Sigma} \hat{V}^\top$ be the singular values decomposition of $\hat{M}$. Put

$$v_{ij} := \sigma^2 \left[U_{i\cdot}^{db} (U^{db\top} U^{db})^{-1} U_{i\cdot}^{db\top} + V_{j\cdot}^{db} (V^{db\top} V^{db})^{-1} V_{j\cdot}^{db\top} \right] mp/n, \quad (4)$$

where

$$U^{db} = \hat{U}(\hat{\Sigma} + \lambda Id_r)^{1/2} \text{ and } V^{db} = \hat{V}(\hat{\Sigma} + \lambda Id_r)^{1/2}.$$

Then, given a significance level $\alpha \in (0, 1)$, the following interval

$$\left[M_{ij}^{db} \pm \Phi^{-1}(1 - \alpha/2) \sqrt{v_{ij}} \right]$$

is a nearly accurate two-sided $(1 - \alpha)$ confidence interval of M_{ij}^* , where $\Phi(\cdot)$ is the CDF of the standard normal distribution. This is given in Corollary 1 in [24].

2.3. The Bayesian estimators

The Bayesian estimator studied in [22, 25] is given by

$$M^B := \int M \rho_\lambda(M|Y) dM$$

where

$$\rho_\lambda(M|Y) \propto L(Y|M)^\lambda \pi(M)$$

is the posterior and $L(Y|M)^\lambda$ is the likelihood raising by λ . Here $\lambda \in (0, 1)$ is a tuning parameter and $\pi(M)$ is the prior distribution.

2.3.1. Priors

A popular choice for the priors in Bayesian matrix completion is to assign conditional Gaussian priors to $U \in \mathbb{R}^{m \times K}$ and $V \in \mathbb{R}^{p \times K}$ with

$$M = UV^\top = \sum_{k=1}^K U_{\cdot k} V_{\cdot k}^\top,$$

for a fixed integer $K \leq \min(m, p)$. More specifically, for $k \in \{1, \dots, K\}$, independently

$$\begin{aligned} U_{\cdot k} &\sim \mathcal{N}(0, \gamma_k Id_m), \\ V_{\cdot k} &\sim \mathcal{N}(0, \gamma_k Id_p), \\ \gamma_k^{-1} &\sim \Gamma(a, b), \end{aligned} \quad (5)$$

where Id_q is the identity matrix of dimension $q \times q$ and a, b are some tuning parameters. This kind of prior is conjugate for which the conditional posteriors can be derived explicitly in closed form and allows to use the Gibbs sampler, see [20] for details. Some reviews and discussion on low-rank factorization priors can be found in [17,30].

Remark 2. In the case that the rank r is not known, it is natural to take K as large as possible, e.g. $K = \min(m, p)$ but this may be computationally prohibitive if K is large.

Remark 3. The estimation error for this Bayesian estimator, under some assumptions, is given in Corollary 4.2 in [25] that $\|M^B - M^*\|_F^2 \leq \max(m, p)r\sigma^2/n$ with additional (multiplicative) log-term by $\log(n \max(m, p))$. It is noted that, in [22], the rate is also reached with additional (multiplicative) log-term by $\log(\min(m, p))$ under general sampling distribution however the paper considered some truncations on the prior.

For a given rank- r , we propose to consider the following prior, called fixed-rank-prior,

$$\begin{aligned} U_{.k} &\sim \mathcal{N}(0, Id_m), \\ V_{.k} &\sim \mathcal{N}(0, Id_p), \end{aligned} \tag{6}$$

for $k = 1, \dots, r$. This prior is a simplified version of the above prior. We note that for $K > r$ the Gibbs sampler of the fixed-rank-prior will be faster than Gibbs sampler for the above prior. However, interestingly, results from simulation for the Bayesian estimator with this prior are slightly better than the one based on the above prior at some point.

Remark 4. We remark that the theoretical estimation error for the Bayesian estimator with the fixed-rank-prior given in (6) remains unchanged following by Corollary 4.2 in [25].

2.3.2. Credible intervals

Using Bayesian approach, the credibility intervals for the matrix and their functions (e.g. entries) can be easily constructed using the Markov Chain Monte Carlo (MCMC) technique. Here, we focus on the equal-tailed credible interval for an entry.

More precisely, the credible intervals are reported using the 89% equal-tailed intervals which is suggested by [37,38] for small posterior samples as in our situations with 500 posterior samples. We noted that, in practice e.g. as in [20] as the data are too big to draw a reasonable size sample, the authors state that to draw only 500 observations from the Gibbs Sampler took 90 hours for the Netflix dataset. Thus we focus on the 89% equal-tailed credible intervals for 500 posterior samples. It is however noted that to obtain 95% intervals, an effective posterior sample size of at least 10.000 is recommended [37], which is computationally prohibited to run on all of our simulations. A few examples with 10.000 posterior samples are examined in Figure 1.

3. Numerical simulation studies

3.1. Experimental designs

In order to access the performance of different estimators, a series of experiments were conducted with simulated data:

- Setting I: In the first setting, we fix $m = 100$ and a rank- r matrix $M_{m \times p}^*$ is generated as the product of two rank- r matrices,

$$M^* = U_{m \times r}^* (V_{p \times r}^*)^\top,$$

where the entries of U^* and V^* are i.i.d $\mathcal{N}(0, 1)$. With a missing rate $\tau = 20\%, 50\%$ and 80% , the entries of the matrix M^* are observed using a uniform sampling. This sampled set is then corrupted by noise as in (1), where the $\mathcal{E}_i$ are i.i.d $\mathcal{N}(0, 1)$. We alternate the other dimension by taking $p = 100$ and $p = 1000$. The rank r is varied between $r = 2$ and $r = 5$.

- Setting II: The second series of simulations is similar to the first one, except that the matrix M^* is no longer rank- r , but it can be well approximated by a rank- r matrix:

$$M^* = U_{m \times r}^* (V_{m \times r}^*)^\top + \frac{1}{10} (A_{m \times 50}) (B_{m \times 50})^\top$$

where the entries of A and B are i.i.d $\mathcal{N}(0, 1)$.

Remark 5. We note that for second series of simulations, with approximate low-rank matrices, the theory of the de-biased estimator can not be used whereas theoretical guarantees for Bayesian estimators are still validated, see [22,25]. The setting I follows exactly the minimax-optimal regime and thus it will allow to access the accuracy of the considered estimators.

For each setup, we generate 50 data sets (simulation replicates) and report the average and the standard deviation for a measure of error of each estimator over the replicates. The behavior of an estimator (say $\widehat{M}$) is evaluated through the mean squared error (MSE) per entry

$$\text{MSE} := \frac{1}{mp} \|\widehat{M} - M^*\|_F^2$$

and the normalized mean square error (NMSE)

$$\text{NMSE} := \frac{\|\widehat{M} - M^*\|_F^2}{\|M^*\|_F^2}.$$

We also measure the prediction error by using

$$\text{Pred} := \frac{\|\mathcal{P}_{\bar{\Omega}}(\widehat{M} - M^*)\|_F^2}{mp - n},$$

where $\bar{\Omega}$ is the set of un-observed entries.

We compare the de-biased estimator (denoted by ‘**d.b**’), the Bayesian estimator with the fixed-rank-prior (6) (denoted by ‘**f.Bayes**’) and the Bayesian estimator with the (flexible rank) prior (5) (denoted by ‘**Bayes**’). As a by-product in calculating the de-biased estimator through the Alternating Least Squares estimator (2), we also report the results for this estimator, denoted it by ‘**als**’.

The ‘als’ estimator is available from the R package ‘softImpute’ [27] and is used with defaulted options. The ‘d.b’ estimator is run with $\lambda = 2.5\sigma\sqrt{mp}$ as in [24]. The ‘f.Bayes’ and ‘Bayes’ estimators are used with tuning parameter $\lambda = 1/(4\sigma^2)$ and parameters for the prior of ‘Bayes’ estimator are $K = 10, a = 1, b = 1/100$. The Gibbs samplers for these two Bayesian estimators are run with 500 steps and 100 burn-in steps.

3.2. Results on estimation accuracy

From the results in Tables 1 and 2, it is clear to see that the de-biased estimator significantly outperforms its ancestry estimator being de-biased. Whereas, the de-biased estimator is just as good as the Bayesian methods in some cases.

More specifically, in Table 1, the de-biased estimator is equally comparative to Bayesian estimators in the case with high rates of observation (say $\tau = 20\%$ or 50%). With the case of highly missing rate $\tau = 80\%$, the de-biased estimator returns highly unstable results, this may be because its ancestry estimator (here is the als estimator) is unstable with small observation. However, when the dimension of the matrix increases, the differences between the de-biased estimator and the Bayesian estimators become smaller. This is also recognized for the setting of approximate low-rank matrices as in Table 2 and in Table 3, 4.

The ‘f.Bayes’ method returns the best results quite often in terms of all considered errors (MSE, NMSE and Pred) in setting of exact low-rank matrices. However, it is noted that for the setting with true underlying matrix being approximately low-rank, in Table 2 and 4, the ‘Bayes’ approach is slightly better than the ‘f.Bayes’ approach at some point. This can be explained as the ‘Bayes’ approach employed a kind of approximate low-rank prior through the Gamma prior on the variance of the factor matrices and thus it is able to adapt to the approximate low-rankness.

3.3. Results on uncertainty quantification

To examine the uncertainty quantification across the methods, we simulate a matrix as in Section 3.1 then we repeat the observation process 50 times. More precisely, we obtain 50 data sets by replicating the observation of 20%, 50% and 80% entries of the matrix M^* using a uniform sampling and then each sampled set is corrupted by noise as in (1), where the $\mathcal{E}_i$ are i.i.d $\mathcal{N}(0, 1)$.

Table 3 and 4 have the empirical coverage rate of the confidence intervals and of the credible intervals as well as the estimation errors of all methods over 50 independent experiments. More precisely, we report the 95% confidence intervals for the de-biased method. The credible intervals is reported using the 89% equal-tailed intervals which is suggested by [37,38] for small posterior samples as in our situations with 500 posterior samples. We noted that to obtain 95% intervals, an effective posterior sample size of at least 10,000 is recommended [37]. A few examples from Setting I with 10000 posterior samples are given in Figure 1.

A noteworthy conclusion from the results in Table 3 and 4 is that the coverage

rates of the 89% credible intervals are significantly higher than those of the 95% confidence intervals revealed by the de-biased method. The credible intervals of the 'f.Bayes' approach show a slightly better coverage rate than those based on the 'Bayes' approach. It is also noted that in the setting of approximate low-rankness, Table 4, where we do not have theoretical guarantee for the de-biased estimator, the coverage rate of the 95% confidence intervals is very low while the credible intervals still do their job quite well. These results further explain why Bayesian methods yield better results in accuracy as in Table 1 and 2.

In Figure 1, we compare the limiting Gaussian distribution of the de-biased estimator and posterior samples for the 'f.Bayes' method against the true entries of interest. It is shown that the limiting Gaussian distribution of the de-biased estimator yields a slightly sharper tail distribution compared to the distribution of the posterior samples. In addition, Figure 2 displays the Q-Q (quantile-quantile) plots of 10000 posterior samples of some entries vs. the standard Gaussian random variables. It is evident that the posterior distributions of these entries reasonably well match the standard Gaussian distribution.

4. Discussion and Conclusion

In this paper, we have provided extensive numerical comparisons between the de-biased estimator and the Bayesian estimators in the problem of low-rank matrix completion. Results from numerical simulations draw a systematic picture on the behaviour of these estimators originally. More specifically, on the estimation accuracy, the de-biased estimator is just comparable to the Bayesian estimators whereas the Bayesian estimators are much more stable and in some cases can outperform the de-biased estimator especially in the small samples regime. Moreover, the credible intervals reasonably cover the underlying entries quite well and slightly better than the confidence intervals in exact low-rank matrix completion. However, in the case of approximate low-rankness, the confidence intervals revealed by the de-biased estimators are no longer work well. These results are interesting and can be served as a guideline to researchers as well as practitioners in many areas.

These findings suggest that the considered Bayesian estimators may actually reach the minimax-optimal rate of convergence without additional logarithmic factor. These log-terms could be due to the PAC-Bayesian bounds technique that used to prove the theoretical properties of the Bayesian estimator. On the other hand, as shown in [25], the same rate with log-term is proved for the concentration of the corresponding posterior and we conjecture that this rate could also be improved due to the coverage of credible intervals. These are important questions that remain open up to our knowledge.

Last but not least, it is also important to do the comparisons with the Variational Bayesian (VB) method in [19] where its theoretical guarantees are given in [25], because this method is very popular for matrix completion with large dataset. This will be the objective of our future work. However, we would like to note that, in a preprint [30], the authors had performed some comparisons between the Bayesian approach and the VB method. The message from their works is that we can expect that VB should be more or less as accurate as Bayes, maybe slightly less, but that the credibility intervals would be wrong (see e.g Figure 3 in [30]).

Funding

This research was supported by the European Research Council grant number 742158.

Availability of data and codes

The R codes and data used in the numerical experiments are available at: <https://github.com/tienmt/UQMC>.

Acknowledgements

The author would like to thank Pierre Alquier for kindly providing useful feedbacks on a first draft of this paper.

Table 1. Simulation results for Setting I (exact low-rank). The mean and the standard deviation (in parentheses) of each error between the simulation replicates are presented. (MSE: average of mean square error; NMSE: average of normalized mean square error; Pred: average of prediction error; als: the alternating least squares estimator; d.b: the de-biased estimator; f.Bayes: the Bayesian estimator with the fixed-rank-prior; Bayes: the Bayesian estimator with flexible rank prior.)

Errors	$r = 2, p = 100, \tau = 20\%$				$r = 5, p = 100, \tau = 20\%$			
	als	d.b	f.Bayes	Bayes	als	d.b	f.Bayes	Bayes
MSE	0.808 (.012)	0.051 (.004)	0.051 (.003)	0.051 (.003)	0.828 (.013)	0.130 (.005)	0.130 (.006)	0.130 (.006)
NMSE	0.399 (.056)	0.025 (.004)	0.025 (.004)	0.025 (.004)	0.167 (.014)	0.026 (.003)	0.026 (.002)	0.026 (.003)
Pred	0.054 (.004)	0.054 (.004)	0.054 (.004)	0.054 (.004)	0.148 (.009)	0.148 (.009)	0.148 (.009)	0.149 (.010)
	$r = 2, p = 100, \tau = 50\%$				$r = 5, p = 100, \tau = 50\%$			
	als	d.b	f.Bayes	Bayes	als	d.b	f.Bayes	Bayes
MSE	0.548 (.011)	0.088 (.007)	0.088 (.007)	0.089 (.007)	0.632 (.014)	0.233 (.012)	0.235 (.012)	0.238 (.012)
NMSE	0.290 (.040)	0.046 (.007)	0.046 (.007)	0.047 (.008)	0.130 (.014)	0.048 (.006)	0.048 (.005)	0.049 (.006)
Pred	0.093 (.007)	0.093 (.007)	0.094 (.007)	0.095 (.008)	0.265 (.016)	0.265 (.016)	0.268 (.016)	0.272 (.016)
	$r = 2, p = 100, \tau = 80\%$				$r = 5, p = 100, \tau = 80\%$			
	als	d.b	f.Bayes	Bayes	als	d.b	f.Bayes	Bayes
MSE	0.695 (.784)	0.545 (.805)	0.286 (.027)	0.294 (.028)	3.999 (.900)	3.755 (.912)	1.083 (.085)	1.417 (.228)
NMSE	0.335 (.344)	0.261 (.354)	0.141 (.020)	0.145 (.024)	0.813 (.177)	0.763 (.181)	0.222 (.029)	0.289 (.051)
Pred	0.618 (.979)	0.618 (.981)	0.304 (.031)	0.313 (.033)	4.748 (1.123)	4.328 (1.116)	1.221 (.102)	1.602 (.265)
	$r = 2, p = 1000, \tau = 20\%$				$r = 5, p = 1000, \tau = 20\%$			
	als	d.b	f.Bayes	Bayes	als	d.b	f.Bayes	Bayes
MSE	0.806 (.004)	0.028 (.001)	0.028 (.001)	0.028 (.001)	0.816 (.005)	0.070 (.001)	0.070 (.001)	0.070 (.001)
NMSE	0.407 (.041)	0.014 (.001)	0.014 (.001)	0.014 (.001)	0.162 (.010)	0.013 (.001)	0.014 (.001)	0.014 (.001)
Pred	0.029 (.001)	0.029 (.001)	0.029 (.001)	0.029 (.001)	0.075 (.002)	0.075 (.002)	0.076 (.002)	0.076 (.002)
	$r = 2, p = 1000, \tau = 50\%$				$r = 5, p = 1000, \tau = 50\%$			
	als	d.b	f.Bayes	Bayes	als	d.b	f.Bayes	Bayes
MSE	0.523 (.004)	0.046 (.001)	0.046 (.001)	0.046 (.001)	0.564 (.004)	0.119 (.002)	0.120 (.002)	0.120 (.002)
NMSE	0.266 (.025)	0.023 (.002)	0.023 (.002)	0.024 (.002)	0.115 (.007)	0.024 (.001)	0.024 (.001)	0.024 (.001)
Pred	0.048 (.002)	0.048 (.001)	0.048 (.001)	0.048 (.001)	0.129 (.003)	0.129 (.003)	0.129 (.003)	0.129 (.003)
	$r = 2, p = 1000, \tau = 80\%$				$r = 5, p = 1000, \tau = 80\%$			
	als	d.b	f.Bayes	Bayes	als	d.b	f.Bayes	Bayes
MSE	0.312 (.005)	0.133 (.004)	0.134 (.004)	0.134 (.004)	0.580 (.156)	0.435 (.161)	0.404 (.012)	0.406 (.012)
NMSE	0.174 (.089)	0.068 (.009)	0.068 (.009)	0.069 (.009)	0.116 (.035)	0.087 (.035)	0.081 (.006)	0.081 (.006)
Pred	0.175 (.258)	0.139 (.005)	0.140 (.005)	0.140 (.005)	0.475 (.195)	0.474 (.196)	0.436 (.014)	0.438 (.014)

Table 2. Simulation results for Setting II (approximate low-rank). The mean and the standard deviation (in parentheses) of each error between the simulation replicates are presented. (MSE: average of mean square error; NMSE: average of normalized mean square error; Pred: average of prediction error; als: the alternating least squares estimator; d.b: the de-biased estimator; f.Bayes: the Bayesian estimator with the fixed-rank-prior; Bayes: the Bayesian estimator with flexible rank prior.)

Errors	approximate rank-2, $p = 100, \tau = 20\%$				approximate rank-5, $p = 100, \tau = 20\%$			
	als	d.b	f.Bayes	Bayes	als	d.b	f.Bayes	Bayes
MSE	0.915 (.012)	0.539 (.017)	0.538 (.017)	0.537 (.017)	0.944 (.014)	0.593 (.015)	0.592 (.015)	0.592 (.015)
NMSE	0.373 (.047)	0.220 (.029)	0.219 (.028)	0.218 (.028)	0.172 (.014)	0.108 (.009)	0.108 (.009)	0.108 (.009)
Pred	0.583 (.022)	0.583 (.022)	0.579 (.022)	0.580 (.022)	0.718 (.030)	0.718 (.030)	0.713 (.030)	0.712 (.030)
	approximate rank-2, $p = 100, \tau = 50\%$				approximate rank-5, $p = 100, \tau = 50\%$			
	als	d.b	f.Bayes	Bayes	als	d.b	f.Bayes	Bayes
MSE	0.822 (.016)	0.593 (.018)	0.588 (.016)	0.587 (.017)	0.956 (.021)	0.760 (.023)	0.742 (.022)	0.743 (.022)
NMSE	0.324 (.038)	0.234 (.028)	0.232 (.027)	0.232 (.027)	0.173 (.015)	0.137 (.013)	0.134 (.012)	0.134 (.012)
Pred	0.642 (.024)	0.642 (.024)	0.633 (.021)	0.633 (.022)	0.912 (.034)	0.913 (.034)	0.882 (.033)	0.883 (.033)
	approximate rank-2, $p = 100, \tau = 80\%$				approximate rank-5, $p = 100, \tau = 80\%$			
	als	d.b	f.Bayes	Bayes	als	d.b	f.Bayes	Bayes
MSE	1.344 (.872)	1.275 (.887)	0.842 (.035)	0.846 (.034)	4.747 (1.044)	4.603 (1.064)	1.724 (.084)	1.854 (.193)
NMSE	0.555 (.370)	0.527 (.375)	0.346 (.036)	0.348 (.037)	0.876 (.181)	0.849 (.184)	0.320 (.023)	0.344 (.043)
Pred	1.430 (1.090)	1.431 (1.091)	0.902 (.040)	0.906 (.039)	5.685 (1.303)	5.305 (1.300)	1.963 (.100)	2.109 (.223)
	approximate rank-2, $p = 1000, \tau = 20\%$				approximate rank-5, $p = 1000, \tau = 20\%$			
	als	d.b	f.Bayes	Bayes	als	d.b	f.Bayes	Bayes
MSE	0.909 (.004)	0.522 (.012)	0.522 (.012)	0.501 (.009)	0.923 (.005)	0.552 (.011)	0.552 (.010)	0.525 (.009)
NMSE	0.363 (.030)	0.208 (.017)	0.208 (.017)	0.200 (.016)	0.168 (.009)	0.101 (.006)	0.100 (.006)	0.095 (.005)
Pred	0.546 (.013)	0.546 (.013)	0.545 (.013)	0.532 (.011)	0.615 (.013)	0.615 (.013)	0.613 (.012)	0.594 (.011)
	approximate rank-2, $p = 1000, \tau = 50\%$				approximate rank-5, $p = 1000, \tau = 50\%$			
	als	d.b	f.Bayes	Bayes	als	d.b	f.Bayes	Bayes
MSE	0.785 (.005)	0.547 (.010)	0.546 (.010)	0.545 (.010)	0.847 (.008)	0.624 (.012)	0.622 (.012)	0.619 (.012)
NMSE	0.316 (.025)	0.220 (.018)	0.219 (.018)	0.219 (.017)	0.153 (.009)	0.113 (.007)	0.112 (.007)	0.111 (.007)
Pred	0.571 (.010)	0.571 (.010)	0.569 (.010)	0.569 (.010)	0.694 (.014)	0.694 (.014)	0.688 (.014)	0.687 (.013)
	approximate rank-2, $p = 1000, \tau = 80\%$				approximate rank-5, $p = 1000, \tau = 80\%$			
	als	d.b	f.Bayes	Bayes	als	d.b	f.Bayes	Bayes
MSE	0.767 (.014)	0.678 (.015)	0.663 (.015)	0.662 (.015)	1.140 (.018)	1.068 (.019)	0.992 (.020)	0.992 (.020)
NMSE	0.307 (.030)	0.271 (.027)	0.265 (.026)	0.265 (.026)	0.207 (.012)	0.194 (.011)	0.180 (.010)	0.180 (.010)
Pred	0.709 (.017)	0.708 (.017)	0.689 (.016)	0.689 (.016)	1.175 (.023)	1.175 (.023)	1.080 (.023)	1.080 (.023)

Figure 1. Plot to compare the limiting Gaussian distributions of the de-biased estimator and the histograms of the 10000 posterior samples for some entries. The dotted line is the true value of the entries. Top row (from left to right, 3 figures) is the results from Setting I with $r = 2, p = 100, \tau = 50\%$. Bottom row (from left to right, 3 figures) is the results from Setting I with $r = 2, p = 1000, \tau = 50\%$

Table 3. Simulation results for Setting I on coverage rate of the confidence intervals and of the credible intervals of the entries as well as the mean square errors (standard deviation is given in parentheses).

missing rate	coverage rate			MSE		
	CI db	CI f.Bayes	CI Bayes	MSE db	MSE f.Bayes	MSE Bayes
$r = 2, p = 100$						
$\tau = 20\%$	0.948 (.008)	0.999 (.001)	0.993 (.002)	0.051 (.003)	0.051 (.003)	0.051 (.003)
$\tau = 50\%$	0.940 (.010)	0.991 (.002)	0.998 (.001)	0.088 (.006)	0.088 (.007)	0.089 (.008)
$\tau = 80\%$	0.886 (.058)	0.985 (.005)	0.984 (.005)	0.498 (.664)	0.290 (.028)	0.295 (.030)
$r = 2, p = 1000$						
$\tau = 20\%$	0.948 (.005)	0.999 (.0002)	0.994 (.001)	0.028 (.001)	0.028 (.001)	0.028 (.001)
$\tau = 50\%$	0.945 (.004)	0.999 (.0002)	0.993 (.002)	0.046 (.001)	0.046 (.001)	0.046 (.001)
$\tau = 80\%$	0.929 (.005)	0.994 (.001)	0.989 (.001)	0.364 (.977)	0.137 (.004)	0.137 (.005)
$r = 5, p = 100$						
$\tau = 20\%$	0.947 (.005)	0.999 (.0004)	0.991 (.002)	0.128 (.006)	0.155 (.007)	0.128 (.006)
$\tau = 50\%$	0.931 (.007)	0.997 (.001)	0.990 (.002)	0.234 (.012)	0.239 (.012)	0.235 (.013)
$\tau = 80\%$	0.698 (.085)	0.968 (.005)	0.945 (.039)	2.768 (1.407)	1.799 (.117)	1.490 (.156)
$r = 5, p = 1000$						
$\tau = 20\%$	0.947 (.002)	0.999 (.0001)	0.993 (.001)	0.070 (.002)	0.077 (.002)	0.070 (.001)
$\tau = 50\%$	0.940 (.002)	0.999 (.0002)	0.992 (.001)	0.120 (.003)	0.148 (.003)	0.120 (.003)
$\tau = 80\%$	0.897 (.009)	0.990 (.001)	0.987 (.001)	0.418 (.011)	0.418 (.011)	0.411 (.011)

Table 4. Simulation results for Setting II on coverage rate of the confidence intervals and of the credible intervals of the entries as well as the mean square errors (standard deviation is given in parentheses).

missing rate	coverage rate			MSE		
	CI db	CI f.Bayes	CI Bayes	db	f.Bayes	Bayes
approximate rank-2, $p = 100$						
$\tau = 20\%$	0.408 (.003)	0.679 (.003)	0.547 (.004)	0.554 (.004)	0.553 (.004)	0.552 (.004)
$\tau = 50\%$	0.494 (.004)	0.784 (.004)	0.655 (.005)	0.575 (.009)	0.569 (.008)	0.569 (.008)
$\tau = 80\%$	0.608 (.025)	0.898 (.006)	0.818 (.008)	1.116 (.685)	0.830 (.029)	0.835 (.030)
approximate rank-2, $p = 1000$						
$\tau = 20\%$	0.313 (.001)	0.545 (.001)	0.541 (.021)	0.502 (.001)	0.501 (.001)	0.489 (.005)
$\tau = 50\%$	0.393 (.001)	0.654 (.002)	0.531 (.007)	0.559 (.002)	0.558 (.002)	0.557 (.002)
$\tau = 80\%$	0.505 (.002)	0.763 (.003)	0.663 (.003)	0.673 (.006)	0.659 (.006)	0.659 (.006)
approximate rank-5, $p = 100$						
$\tau = 20\%$	0.610 (.004)	0.888 (.002)	0.766 (.003)	0.617 (.007)	0.641 (.009)	0.614 (.008)
$\tau = 50\%$	0.681 (.005)	0.935 (.003)	0.847 (.004)	0.749 (.019)	0.812 (.020)	0.735 (.017)
$\tau = 80\%$	0.541 (.036)	0.942 (.006)	0.890 (.031)	4.946 (.819)	2.434 (.129)	1.984 (.269)
approximate rank-5, $p = 1000$						
$\tau = 20\%$	0.501 (.001)	0.793 (.001)	0.702 (.009)	0.544 (.002)	0.551 (.002)	0.542 (.002)
$\tau = 50\%$	0.556 (.001)	0.841 (.001)	0.724 (.007)	0.619 (.003)	0.636 (.003)	0.614 (.004)
$\tau = 80\%$	0.671 (.008)	0.921 (.002)	0.858 (.003)	0.676 (.007)	0.714 (.006)	0.659 (.006)

Figure 2. Q-Q (quantile-quantile) plot to compare the 10000 posterior samples for some entries vs. the standard normal distribution. Top row (from left to right, 3 figures) is the results from Setting I with $r = 2, p = 100, \tau = 50\%$. Bottom row (from left to right, 3 figures) is the results from Setting I with $r = 2, p = 1000, \tau = 50\%$

References

- [1] Xiong L, Chen X, Huang TK, et al. Temporal collaborative filtering with bayesian probabilistic tensor factorization. In: Proceedings of the 2010 SIAM international conference on data mining; SIAM; 2010. p. 211–222.
- [2] Adomavicius G, Tuzhilin A. Context-aware recommender systems. In: Recommender systems handbook. Springer; 2011. p. 217–253.
- [3] Bennett J, Lanning S. The netflix prize. In: Proceedings of KDD cup and workshop; Vol. 2007; 2007. p. 35.
- [4] Chi EC, Zhou H, Chen GK, et al. Genotype imputation via matrix completion. *Genome research*. 2013;23(3):509–518.
- [5] Jiang B, Ma S, Causey J, et al. Sparrec: An effective matrix completion framework of missing data imputation for gwas. *Scientific reports*. 2016;6(1):1–15.
- [6] Cabral R, De la Torre F, Costeira JP, et al. Matrix completion for weakly-supervised multi-label image classification. *IEEE transactions on pattern analysis and machine intelligence*. 2014;37(1):121–135.
- [7] Luo Y, Liu T, Tao D, et al. Multiview matrix completion for multilabel image classification. *IEEE Transactions on Image Processing*. 2015;24(8):2355–2368.
- [8] He K, Sun J. Image completion approaches using the statistics of similar patches. *IEEE transactions on pattern analysis and machine intelligence*. 2014;36(12):2423–2435.
- [9] Gross D, Liu YK, Flammia ST, et al. Quantum state tomography via compressed sensing. *Physical review letters*. 2010;105(15):150401.
- [10] Mai TT, Alquier P. Pseudo-bayesian quantum tomography with rank-adaptation. *Journal of Statistical Planning and Inference*. 2017;184:62–76.
- [11] Mai TT. Efficient adaptive mcmc implementation for pseudo-bayesian quantum tomography. *arXiv preprint arXiv:210600577*. 2021;.
- [12] Candès EJ, Plan Y. Matrix completion with noise. *Proceedings of the IEEE*. 2010; 98(6):925–936.
- [13] Candès EJ, Recht B. Exact matrix completion via convex optimization. *Found Comput Math*. 2009;9(6):717–772. Available from: <http://dx.doi.org/10.1007/s10208-009-9045-5>.
- [14] Candès EJ, Tao T. The power of convex relaxation: near-optimal matrix completion. *IEEE Trans Inform Theory*. 2010;56(5):2053–2080. Available from: <http://dx.doi.org/10.1109/TIT.2010.2044061>.
- [15] Koltchinskii V, Lounici K, Tsybakov AB. Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *Ann Statist*. 2011;39(5):2302–2329. Available from: <http://dx.doi.org/10.1214/11-AOS894>.
- [16] Recht B, Ré C. Parallel stochastic gradient algorithms for large-scale matrix completion. *Math Program Comput*. 2013;5(2):201–226. Available from: <http://dx.doi.org/10.1007/s12532-013-0053-8>.
- [17] Alquier P. Bayesian methods for low-rank matrix estimation: short survey and theoretical study. In: *Algorithmic Learning Theory 2013*; Springer; 2013. p. 309–323.
- [18] Lawrence ND, Urtasun R. Non-linear matrix factorization with gaussian processes. In: *Proceedings of the 26th Annual International Conference on Machine Learning*; ACM; 2009. p. 601–608.
- [19] Lim YJ, Teh YW. Variational bayesian approach to movie rating prediction. In: *Proceedings of KDD Cup and Workshop*; Vol. 7; 2007. p. 15–21.
- [20] Salakhutdinov R, Mnih A. Bayesian probabilistic matrix factorization using markov chain monte carlo. In: *Proceedings of the 25th international conference on Machine learning*; ACM; 2008. p. 880–887.
- [21] Zhou M, Wang C, Chen M, et al. Nonparametric bayesian matrix completion. *Proc IEEE SAM*. 2010;.
- [22] Mai TT, Alquier P. A bayesian approach for noisy matrix completion: Optimal rate under general sampling distribution. *Electron J Statist*. 2015;9(1):823–841. Available from: <https://doi.org/10.1214/15-EJS1020>.

- [23] Mai TT. Bayesian matrix completion with a spectral scaled student prior: theoretical guarantee and efficient sampling. arXiv preprint arXiv:210408191. 2021;.
- [24] Chen Y, Fan J, Ma C, et al. Inference and uncertainty quantification for noisy matrix completion. *Proceedings of the National Academy of Sciences*. 2019;116(46):22931–22937.
- [25] Alquier P, Ridgway J. Concentration of tempered posteriors and of their variational approximations. *The Annals of Statistics*. 2020;48(3):1475–1497.
- [26] Negahban S, Wainwright MJ. Restricted strong convexity and weighted matrix completion: optimal bounds with noise. *J Mach Learn Res*. 2012;13:1665–1697.
- [27] Mazumder R, Hastie T, Tibshirani R. Spectral regularization algorithms for learning large incomplete matrices. *The Journal of Machine Learning Research*. 2010;11:2287–2322.
- [28] Hastie T, Mazumder R, Lee JD, et al. Matrix completion and low-rank svd via fast alternating least squares. *The Journal of Machine Learning Research*. 2015;16(1):3367–3402.
- [29] Klopp O. Matrix completion by singular value thresholding: sharp bounds. *Electronic journal of statistics*. 2015;9(2):2348–2369.
- [30] Alquier P, Cottet V, Chopin N, et al. Bayesian matrix completion: prior specification and consistency. arXiv preprint arXiv:14061440. 2014;.
- [31] Cottet V, Alquier P. 1-bit matrix completion: Pac-bayesian analysis of a variational approximation. *Machine Learning*. 2018;107(3):579–603.
- [32] Babacan SD, Luessi M, Molina R, et al. Sparse bayesian methods for low-rank matrix estimation. *IEEE Transactions on Signal Processing*. 2012;60(8):3964–3977.
- [33] Yang L, Fang J, Duan H, et al. Fast low-rank bayesian matrix completion with hierarchical gaussian prior models. *IEEE Transactions on Signal Processing*. 2018;66(11):2804–2817.
- [34] Harper FM, Konstan JA. The movielens datasets: History and context. *Acm transactions on interactive intelligent systems (tiis)*. 2015;5(4):1–19.
- [35] Rendle S, Zhang L, Koren Y. On the difficulty of evaluating baselines: A study on recommender systems. arXiv preprint arXiv:190501395. 2019;.
- [36] Rendle S, Krichene W, Zhang L, et al. Neural collaborative filtering vs. matrix factorization revisited. In: *Fourteenth ACM Conference on Recommender Systems*; 2020. p. 240–248.
- [37] Kruschke J. *Doing bayesian data analysis: A tutorial with r, jags, and stan*. Academic Press; 2014.
- [38] McElreath R. *Statistical rethinking: A bayesian course with examples in r and stan*. CRC press; 2020.