

MATRIX COMPLETION VIA RESIDUAL SPECTRAL MATCHING

BY ZIYUAN CHEN^{1,a}, FANG YAO^{1,b}

¹School of Mathematical Sciences, Center for Statistical Science, Peking University,
^achenziyuan@pku.edu.cn, ^bfyyao@math.pku.edu.cn

Noisy matrix completion has attracted significant attention due to its applications in recommendation systems, signal processing and image restoration. Most existing works rely on (weighted) least squares methods under various low-rank constraints. However, minimizing the sum of squared residuals is not always efficient, as it may ignore the potential structural information in the residuals. In this study, we propose a novel residual spectral matching criterion that incorporates not only the numerical but also locational information of residuals. This criterion is the first in noisy matrix completion to adopt the perspective of low-rank perturbation of random matrices and exploit the spectral properties of sparse random matrices. We derive optimal statistical properties by analyzing the spectral properties of sparse random matrices and bounding the effects of low-rank perturbations and partial observations. Additionally, we propose algorithms that efficiently approximate solutions by constructing easily computable pseudo-gradients. The iterative process of the proposed algorithms ensures convergence at a rate consistent with the optimal statistical error bound. Our method and algorithms demonstrate improved numerical performance in both simulated and real data examples, particularly in environments with high noise levels.

1. Introduction.

1.1. Literature review. Matrix completion aims to reconstruct the target matrix $\mathbf{M} \in \mathbb{R}^{m \times n}$ from partial observations of its entries that are corrupted by random noise, which has applications across different fields. For example, it has been utilized in recommendation systems (Jannach et al., 2016; Ramlatchan et al., 2018), signal processing (Weng and Wang, 2012; Li et al., 2019), and image restoration (Ji et al., 2011; Jia et al., 2022). Recently, matrix completion based statistical models have also been applied to various problems, including wireless sensor networks (Xie et al., 2016; Kortas et al., 2021), causal inference (Kallus, Mao and Udell, 2018; Athey et al., 2021), and policy evaluation (Iweze, 2020; Duan, Li and Xia, 2024).

In the matrix completion model, the sample size is often significantly smaller than the matrix dimensions $m \times n$. Completing the target matrix \mathbf{M} is generally impossible without specific structures and properties. The mostly considered property for the underlying matrix \mathbf{M} is a low-rank structure or related characteristics, such as the rapid decay of singular values. A common approach to addressing the low-rank matrix completion problem is to impose strict low-rank constraints through specific matrix factorization techniques. For instance, Keshavan, Montanari and Oh (2009) introduced a method that employs thin singular value decomposition. The typical method for constraining the rank of matrix \mathbf{M} to be less than s represents it as $\mathbf{M} = \mathbf{L}\mathbf{R}'$, where $\mathbf{L} \in \mathbb{R}^{m \times s}$ and $\mathbf{R} \in \mathbb{R}^{n \times s}$. Although this low-rank representation lacks identifiability, extensive research has examined the properties of least squares methods utilizing this matrix factorization with alternative minimizing schemes.

MSC2020 subject classifications: Primary 62H12; secondary 60B20.

Keywords and phrases: low-rank perturbation, random matrix theory, residual structure, sparse random matrix, spectral property.

This includes the incorporation of regularization terms (Sun and Luo, 2016; Ge, Lee and Ma, 2016) and algorithmic analyses (Jain, Netrapalli and Sanghavi, 2013; Hardt, 2014; Zheng and Lafferty, 2016; Ma et al., 2018; Chen, Liu and Li, 2020). Following the findings of Candès and Recht (2009), another prevalent method has emerged. They demonstrated that exact matrix completion is achievable in the absence of random noise, by relaxing the strict rank constraint to a nuclear norm constraint. Consequently, the nuclear norm constraint is widely used in both practical applications and theoretical analyses due to its convex properties. Various least squares estimators based on the nuclear norm (Candes and Plan, 2010; Rohde and Tsybakov, 2011; Negahban and Wainwright, 2011) and their variants (Koltchinskii, Lounici and Tsybakov, 2011; Cai and Zhang, 2015; Cai and Zhou, 2016) have been thoroughly investigated.

Recent studies have also explored various extensions to traditional noisy matrix completion settings. Works such as Klopp (2014) and Bi et al. (2017) extended the missing data mechanism from “missing completely at random” to “missing at random”. Later, the missing mechanism was further extended to “missing not at random” (Bhattacharya and Chatterjee, 2022; Choi and Yuan, 2023). Chen et al. (2019) and Xia and Yuan (2021) investigated both completion and statistical inference problems. Gui, Barber and Ma (2023) later combined conformal inference techniques with matrix completion. Using additional covariate information, Mao, Chen and Wong (2019) and Jin, Ma and Jiang (2022) proposed a column space decomposition model. Ma et al. (2024) also studied the inference problem related to this model. Xu and Davenport (2016) and Chen, Yang and Yao (2024) introduced dynamic matrix estimators designed for time-varying target matrices. In the absence of low-rank structure, high rank matrix completion (Eriksson, Balzano and Nowak, 2012; Elhamifar, 2016; Ongie et al., 2017) and nonlinear variants (Fan and Chow, 2017; Fan and Cheng, 2018) have also been extensively studied. Additionally, tensor completion has gained popularity due to the development of its efficient algorithms (Zhang et al., 2020; Luo and Zhang, 2023).

1.2. Different perspective from random matrix theory. For the noisy matrix completion problem and its variants, previous works typically treats the noisy matrix completion model as data sampled from the target matrix \mathbf{M} with random noise. The commonly used criterion for evaluating performance in noisy matrix completion is the (weighted) squared error. Least squares methods are often employed in statistical models involving noise, and the resulting estimators generally exhibit good performance. However, the rank structure of matrices relates to both the values and locations of entries, indicating that random noises possesses both numerical and locational characteristics in the noisy matrix completion problem. Applying least squares methods directly overlooks the locational properties of residuals and may not be efficient, especially in high-noise scenarios.

In contrast, we approach the noisy matrix completion problem from a different perspective. We model the data as samples from a low-rank perturbation \mathbf{M} of a random matrix \mathbf{H} consisting of random noise. This perspective motivates us to propose a criterion based on the similarity between the residual matrix and sparse random matrices. Inspired by the fact that singular values capture both numerical and positional information about entries in matrices, we measure the similarity by comparing the distribution of singular values between the residual matrix and sparse random matrices. The properties of singular values of mean zero rectangular random matrices $\mathbf{X} \in \mathbb{R}^{m \times n}$ with $\mathbb{E}[\mathbf{X}_{i,j}^2] = 1$, are directly linked to the empirical spectral distribution (ESD) of sample covariance matrices $\mathbf{S} = \mathbf{X}\mathbf{X}'/m$. This ESD converges to the Marchenko-Pastur (M-P) distribution as $m \rightarrow \infty$ with the ratio $n/m \rightarrow \rho > 0$ (Yin and Krishnaiah, 1986; Bai and Yin, 1988). Extensive research has examined the convergence rate to the M-P distribution (Götze and Tikhomirov, 2004; Tikhomirov, 2011; Bai, Hu and Zhou, 2012), local convergence in various settings (Bourgade, Yau and Yin, 2014; Alt,

Erdos and Krüger, 2017; Bao, Erdős and Schnelli, 2017) and the universality property (Erdos, Yau and Yin, 2010; Erdős, Yau and Yin, 2012; Tao and Vu, 2012; Pillai and Yin, 2014). Benaych-Georges and Nadakuditi (2012) provided the limit values of leading singular values and vectors in low-rank perturbations of random matrices. Subsequent studies have explored additional properties, such as the limit distribution of leading singular values (Veraart et al., 2016; Bao, Ding and Wang, 2021; Bao et al., 2022). The tail distributions of entries in classical random matrix theories are often constrained by bounded high-order moment conditions. Sparse random matrices, where each entry is multiplied by a random variable following a binomial distribution with probability $p_{i,j} \rightarrow 0$, do not satisfy these conditions. For example, the entries in Gaussian sparse random matrices \mathbf{X} with $\mathbb{E}[\mathbf{X}_{i,j}^2] = 1$ have a divergent fourth moment $\mathbb{E}[\mathbf{X}_{i,j}^4] = 3p_{i,j}^{-1}$. Recently, the spectral properties of sparse random matrix have attracted much attention. For example, the local law for sparse sample covariance matrices has been established (Lee, 2021; Tikhomirov and Timushev, 2022; Kafetzopoulos and Maltsev, 2023). The universality property of these matrices is also discussed in several studies (Ding and Yang, 2018; He, 2023; Brailovskaya and van Handel, 2024).

1.3. Our contributions and paper organization. In this paper, we propose a novel residual spectral matching criterion for noisy matrix completion. We adopt the perspective of low-rank perturbation of random matrices and leverage the properties of sparse random noise matrices to design this effective criterion. The criterion incorporates both the locational and numerical information of residuals, unlike the Frobenius norm loss which considers only the numerical aspect. To the best of our knowledge, this is the first approach that accounts for the location of residuals in noisy matrix completion. The proposed criterion is formulated as the weighted L_2 distance between the ordered singular values of the residual matrix and the expected singular values of variance-adjusted sparse Gaussian random matrices. Adjusting the spectral distribution of the residual matrix to resemble that of sparse random matrices provides more structural information than merely minimizing the Frobenius norm of the residual matrix. Our research demonstrates both optimal theoretical properties and improved numerical performance for this criterion, especially at a high noise level.

The main contributions of our study are four-fold. First, a new residual spectral matching criterion is proposed for noisy matrix completion. This criterion is user-friendly and adaptable to various low-rank constraints. Second, the statistical properties of estimators under both strict low-rank and relaxed nuclear norm constraints are provided. We develop an analytical framework that leverages techniques from low-rank perturbation, random missing data, and sparse random matrices. Oracle bounds are presented, and optimal upper bounds are achieved by selecting the appropriate rank constraint and tuning hyper-parameter for the nuclear norm constraint. Third, a general iterative algorithm for minimizing the proposed criterion is presented. An alternating descent scheme and a pseudo-gradient formula are introduced to handle the non-convexity of the criterion and the challenge of formulating explicit gradients. Finally, the properties of algorithm iterations are carefully analysed. We demonstrate that our algorithm achieves the optimal error bound within a finite number of iterations by monitoring both the incoherence condition and the distance between iterative series and the target matrix.

The structure of the paper is as follows. Section 2 provides a brief introduction of the noisy matrix completion problem, and discusses the success of the low-rank perturbation viewpoint and the random matrix theory in the matrix denoising model, a special case of noisy matrix completion. In Section 3, we present the spectral matching criterion, estimation methods and the corresponding algorithm. Section 4 establishes the optimal statistical properties of the proposed estimators and the error analysis for the iterative series in the proposed algorithm. In Section 5, both simulation and real data experiments demonstrate the favorable

performance of the proposed methods over commonly used Frobenius norm methods, particularly when dealing with high noise levels. Conclusions and additional issues are discussed in Section 6. Proof sketches and algorithms are presented in the Appendix, while additional technical lemmas and detailed proofs are provided to an online Supplementary Material for space economy.

2. Preliminary and Motivation.

2.1. Settings for noisy matrix completion and notations. We first introduce some notations to be used throughout the paper. For any matrix $\mathbf{M} \in \mathbb{R}^{m \times n}$, the following norms are defined: $\|\mathbf{M}\|$ for the spectral norm (the largest singular value), $\|\mathbf{M}\|_F$ for the Frobenius norm (the square root of the sum of squared entries), $\|\mathbf{M}\|_*$ for the nuclear norm (the sum of singular values), $\|\mathbf{M}\|_\infty$ for the maximum norm (the largest absolute value of entries), and $\|\mathbf{M}\|_{2,\infty}$ for the largest L_2 norm of the rows (sometimes named as the L_2/L_∞ norm (Ma et al., 2018)). The transpose of \mathbf{M} is denoted by \mathbf{M}' , and $\text{Proj}_s(\mathbf{M})$ indicates the projection of \mathbf{M} onto the space of matrices with rank no more than s . We use the notation $\mathbf{M}_{[1:s]}$ to denote the matrix formed by the first s columns of \mathbf{M} , and $(a)_+$ to represent the positive part of a . Additionally, the notations $a(m) \lesssim b(m)$ means that there exists a constant $c > 0$ such that $|a(m)| \leq c|b(m)|$ and similarly $a(m) \gtrsim b(m)$ means that $|a(m)| \geq c|b(m)|$.

Noisy matrix completion involves recovering the target matrix $\mathbf{M}_0 \in \mathbb{R}^{m \times n}$, $n/m = \rho \in (0, 1]$ from partial observations of its noisy perturbations. Formally speaking, it aims to estimate \mathbf{M}_0 from the observation $\mathbf{Y} \in \mathbb{R}^{m \times n}$

$$\mathbf{Y} = P_\Omega(\mathbf{M}_0 + \mathbf{H}),$$

where $\Omega \in \mathbb{R}^{m \times n}$ is a binary matrix with ones indexing the locations of observed entries, and $\mathbf{H} \in \mathbb{R}^{m \times n}$ is an unknown noise matrix that is independent of \mathbf{M}_0 . The projection $P_\Omega(\cdot)$ preserves the observed entries in Ω and assigns zeros to the unobserved entries.

In previous related works, $\Omega_{i,j}, 1 \leq i \leq m, 1 \leq j \leq n$ are often assumed to follow independent binomial distributions with probability p and $H_{i,j}, 1 \leq i \leq m, 1 \leq j \leq n$ are independent mean zero random variables with variance σ^2 . To separate the target matrix \mathbf{M}_0 from the random noise \mathbf{H} , a key assumption is that \mathbf{M}_0 has a low rank r which is far less than n . The incoherent condition defined in Definition 2.1 is also necessary for \mathbf{M}_0 , to ensure that the missing pattern Ω does not completely disrupt the singular subspace information of \mathbf{M}_0 .

DEFINITION 2.1. A rank- r matrix $\mathbf{M} \in \mathbb{R}^{m \times n}$ is said to be μ -incoherent if

$$\|\mathbf{U}\|_{2,\infty} \leq \sqrt{\frac{\mu r}{m}}, \quad \|\mathbf{V}\|_{2,\infty} \leq \sqrt{\frac{\mu r}{n}},$$

where $\mathbf{M} = \mathbf{U}\mathbf{D}\mathbf{V}'$ is the singular value decomposition.

This incoherent condition of a low-rank matrix \mathbf{M} implies the delocalization of its singular vectors, ensuring that each entry of \mathbf{M} contributes similarly important information to the matrix as a whole. This condition, in turn, provides favorable properties for projection operator $P_\Omega(\cdot)$ onto sparse observations. Specifically, the restricted isometry property (RIP) $(1 - \delta)\|\mathbf{M}\|_F^2 \leq \|p^{-1/2}P_\Omega(\mathbf{M})\|_F^2 \leq (1 + \delta)\|\mathbf{M}\|_F^2$ holds for $P_\Omega(\cdot)$ with high probability, where $\delta \geq 0$ is sufficiently small.

Two mainstream methods and their variations have been proposed and extensively studied for noisy matrix completion. The first type is mostly based on the low-rank factorization $\mathbf{M} = \mathbf{L}\mathbf{R}', \mathbf{L} \in \mathbb{R}^{m \times s}, \mathbf{R} \in \mathbb{R}^{n \times s}$ of a rank s matrix \mathbf{M} . The matrix \mathbf{M}_0 is then estimated by

$$(1) \quad \widehat{\mathbf{M}}_{\text{fac}} = \arg \min_{\mathbf{M} = \mathbf{L}\mathbf{R}', \mathbf{L} \in \mathbb{R}^{m \times s}, \mathbf{R} \in \mathbb{R}^{n \times s}} \|\mathbf{Y} - P_\Omega(\mathbf{L}\mathbf{R}')\|_F^2.$$

Several variations add different regularization terms, such as the Frobenius norm penalty $\lambda(\|\mathbf{L}\|_F^2 + \|\mathbf{R}\|_F^2)$ or the balanced penalty $\lambda\|\mathbf{L}'\mathbf{L} - \mathbf{R}'\mathbf{R}\|_F^2$. The second type constrains the rank of the estimator by controlling its nuclear norm, i.e.,

$$(2) \quad \widehat{\mathbf{M}}_{\text{nuc}} = \arg \min_{\mathbf{M} \in \mathbb{R}^{m \times n}} \|\mathbf{Y} - P_{\Omega}(\mathbf{M})\|_F^2 + \lambda \|\mathbf{M}\|_{*}.$$

In the absence of noise matrix \mathbf{H} , Theorem 1.7 in [Candès and Tao \(2010\)](#) shows that the lower bound for p to achieve exact matrix completion with high probability $1 - \delta$ is $p \gtrsim \mu_0 r m^{-1} \log(m/2\delta)$. For sufficiently large p , the lower bound for the Frobenius norm distance between \mathbf{M}_0 and its estimator $\widehat{\mathbf{M}}$ is $\|\widehat{\mathbf{M}} - \mathbf{M}_0\|_F \gtrsim \sigma \sqrt{mr/p}$. Both estimators mentioned above have been proven to achieve the optimal or near-optimal (up to a logarithmic factor $\log m$) upper bounds. These results can be found in methods based on matrix factorization (e.g., [Keshavan, Montanari and Oh, 2009](#); [Ma et al., 2018](#); [Chen et al., 2020](#)) and for convex relaxed methods (e.g., [Koltchinskii, Lounici and Tsybakov, 2011](#); [Klopp, 2014](#)).

2.2. A motivating example: matrix denoising. The matrix denoising problem aims to separate the underlying signal matrix \mathbf{M}_0 from its noisy version $\mathbf{M}_0 + \mathbf{H}$. In fact, the model can be viewed as a special case of matrix completion with $p = 1$, where $\mathbf{Y} = \mathbf{M}_0 + \mathbf{H}$ is fully observed. When we apply the estimation methods (1) and (2) in parallel to the matrix denoising problem, two estimators for \mathbf{M}_0 are obtained respectively:

$$\widetilde{\mathbf{M}}_{\text{fac}} = \sum_{i=1}^s \sigma_i \mathbf{u}_i \mathbf{v}_i', \quad \widetilde{\mathbf{M}}_{\text{nuc}} = \sum_{i=1}^n (\sigma_i - \lambda/2)_+ \mathbf{u}_i \mathbf{v}_i',$$

where $\mathbf{Y} = \sum_{i=1}^n \sigma_i \mathbf{u}_i \mathbf{v}_i'$ is the singular value decomposition of \mathbf{Y} . While both estimators achieve optimal upper bounds that match the problem's lower bound $\sigma \sqrt{rm}$, we shall demonstrate that neither performs the best in matrix denoising compared to the estimator based on random matrix theory. And the performance gap widens as the noise level rises.

Rather than viewing the matrix denoising model as a target matrix contaminated by random noise, we consider the model as a random matrix \mathbf{H} perturbed by a low-rank matrix \mathbf{M}_0 . Classical random matrix theory ([Bai and Silverstein, 2010](#)) shows that the empirical spectral distribution of $\mathbf{H}'\mathbf{H}/m$ converges to the Marchenko-Pastur distribution. Consequently, the density of the singular values of \mathbf{H}/\sqrt{m} converges to

$$\nu(x) = \frac{1}{\pi \sigma^2} \frac{\sqrt{(\lambda_+^2 - x^2)(x^2 - \lambda_-^2)}}{\rho x} \mathbf{1}_{x \in [\lambda_-, \lambda_+]}, \quad \lambda_{\pm} = \sigma(1 \pm \sqrt{\rho}).$$

The smallest s singular values of \mathbf{H} are found to be larger than $\sqrt{m}\lambda_-$ with high probability. Taking $\widetilde{\mathbf{M}}_{\text{fac}}$ as an example, $\widetilde{\mathbf{H}} = \mathbf{Y} - \widetilde{\mathbf{M}}_{\text{fac}}$ can intuitively be viewed as an estimator of the random noise matrix \mathbf{H} . However, the s smallest singular values of $\widetilde{\mathbf{H}}$ are zero, which differs from those of the random noise matrix \mathbf{H} . This discrepancy indicates that $\widetilde{\mathbf{M}}_{\text{fac}}$ may not be adequate. [Benaych-Georges and Nadakuditi \(2012\)](#) derived the limit distribution of singular values of random matrices with low-rank perturbations. It helps us to construct the estimator $\widetilde{\mathbf{M}}_{\text{rmt}} = \sum_{i=1}^s \widetilde{\sigma}_i \mathbf{u}_i \mathbf{v}_i'$, where

$$\widetilde{\sigma}_i = \left(\frac{1}{2} \left((\sigma_i^2 - \widehat{\sigma}^2(1 + \rho)m) + \sqrt{(\widehat{\sigma}^2(1 + \rho)m - \sigma_i^2)^2 - 4\widehat{\sigma}^4\rho m^2} \right) \right)^{1/2},$$

and $\widehat{\sigma}$ is a \sqrt{m} -consistency estimator of σ . To verify our motivation, We conduct a numerical study to compare the performance of the estimator $\widetilde{\mathbf{M}}_{\text{rmt}}$ with $\widetilde{\mathbf{M}}_{\text{fac}}$ and $\widetilde{\mathbf{M}}_{\text{nuc}}$ in Table 1, which shows the outstanding performance of the estimator $\widetilde{\mathbf{M}}_{\text{rmt}}$ based on random matrix theory, and the increasing performance gap for larger σ .

TABLE 1
The Frobenius norm error ($\times 10^{-2}$) $\|\tilde{\mathbf{M}} - \mathbf{M}_0\|_F / (mn)^{-1/2}$ for $\tilde{\mathbf{M}}_{\text{fac}}$, $\tilde{\mathbf{M}}_{\text{nuc}}$ and $\tilde{\mathbf{M}}_{\text{rmt}}$ across varying noise levels, based on 100 repeated experiments with matrix dimension $m = 500$, $n = 250$ and the rank set to r .

Noise level σ	0.01	0.02	0.05	0.08	0.1	0.2	0.3	0.4	0.5
$r = 5$	$\tilde{\mathbf{M}}_{\text{rmt}}$	0.1724	0.3453	0.8626	1.3796	1.7244	3.4465	5.1645	6.8835
$r = 5$	$\tilde{\mathbf{M}}_{\text{fac}}$	0.1724	0.3453	0.8630	1.3812	1.7275	3.4717	5.2503	7.0831
$r = 5$	$\tilde{\mathbf{M}}_{\text{nuc}}$	0.2699	0.5396	1.3360	2.1219	2.6390	5.1299	7.4743	9.7042
$r = 10$	$\tilde{\mathbf{M}}_{\text{rmt}}$	0.2431	0.4869	1.2154	1.9459	2.4301	4.8577	7.2784	9.6842
$r = 10$	$\tilde{\mathbf{M}}_{\text{fac}}$	0.2431	0.4870	1.2160	1.9480	2.4344	4.8916	7.3943	9.9552
$r = 10$	$\tilde{\mathbf{M}}_{\text{nuc}}$	0.3597	0.7183	1.7776	2.8225	3.5043	6.8005	9.8827	12.782
$r = 20$	$\tilde{\mathbf{M}}_{\text{rmt}}$	0.3415	1.0264	1.8807	2.9057	3.4155	6.9952	10.402	13.600
$r = 20$	$\tilde{\mathbf{M}}_{\text{fac}}$	0.3415	1.0266	1.8816	2.9090	3.4211	7.0432	10.560	13.953
$r = 20$	$\tilde{\mathbf{M}}_{\text{nuc}}$	0.4907	1.4653	2.6650	4.0801	4.7751	9.4669	13.650	17.346

While random matrix theory has achieved significant success in matrix denoising models, its application to noisy matrix completion remains underexplored. The strong performance of the estimator $\tilde{\mathbf{M}}_{\text{rmt}}$ in matrix denoising prompted us to adopt the perspective of low-rank perturbations of random matrices. In the noisy matrix completion problem, we treat the observation $\mathbf{Y} = P_{\Omega}(\mathbf{M}_0 + \mathbf{H})$ as a perturbation of the sparse random matrix $P_{\Omega}(\mathbf{H})$. Thus, finding an optimal estimator $\hat{\mathbf{M}}$ of \mathbf{M}_0 corresponds to approximating $P_{\Omega}(\mathbf{H})$ through the residual matrix $\mathbf{Y} - P_{\Omega}(\hat{\mathbf{M}})$. Unlike matrix denoising, there is no closed-form solution for $\hat{\mathbf{M}}$ in noisy matrix completion. Due to limitation in sample size, approximating each entry of $P_{\Omega}(\mathbf{H})$ is challenging. We leverage the spectral distributions of sparse random matrices, which are stable and converge rapidly to their limit distributions. By using the distance between spectral distributions as a criterion, we can assess whether the residual matrix belongs to the high probability set of sparse random matrices. This approach offers a novel perspective on the noisy matrix completion problem.

3. Methodology and algorithm.

3.1. *Proposed method.* A sparse Gaussian random matrix of size $m \times n$ and missing probability $1 - p_{i,j} \geq 0$, $1 \leq i \leq m$, $1 \leq j \leq n$ can be formulated as

$$\mathbf{M} \in \mathbb{R}^{m \times n}, \quad \mathbf{M}_{i,j} = \Phi_{i,j} \Omega_{i,j}, \quad 1 \leq i \leq m, 1 \leq j \leq n,$$

where $\Phi_{i,j}$ are independent Gaussian random variables with mean zero and variance $1/m$, and $\Omega_{i,j}$ follow independent binomial distributions with success probability $p_{i,j}$. Define the probability space $\mathcal{S}_{m,n,p} = (\mathbb{R}^{m \times n}, \mathcal{B}_{m,n}, \mathcal{P}_{m,n,p})$, where $\mathcal{B}_{m,n}$ is the Borel σ -algebra of $\mathbb{R}^{m \times n}$, and the probability measure $\mathcal{P}_{m,n,p}$ has the density function

$$f_{m,n,p}(\mathbf{M}) = \prod_{i=1}^m \prod_{j=1}^n [p_{i,j} \phi(\sqrt{m} \mathbf{M}_{i,j}) + (1 - p_{i,j}) \delta(\mathbf{M}_{i,j})],$$

where $\phi(x)$ is the density function of standard Gaussian distribution, and $\delta(x)$ is the Dirac delta function. We denote the singular values of \mathbf{M} as $\lambda_1(\mathbf{M}) \geq \lambda_2(\mathbf{M}) \geq \dots \geq \lambda_n(\mathbf{M})$ and combine them into a vector $\boldsymbol{\lambda}(\mathbf{M})$. In the probability space $\mathcal{S}_{m,n,p}$, we define the expectation of the random vector $\boldsymbol{\lambda}(\mathbf{M})$ as $\boldsymbol{\lambda} = \mathbb{E}_{\mathbf{M}} \boldsymbol{\lambda}(\mathbf{M})$. Although deriving the explicit formulation of $\boldsymbol{\lambda}$ is challenging, it can be estimated using a simple Monte Carlo procedure $\hat{\boldsymbol{\lambda}} = \sum_{i=1}^{\ell} \boldsymbol{\lambda}(\mathbf{M}_i)/\ell$ where \mathbf{M}_i , $1 \leq i \leq \ell$ are independent sparse Gaussian random matrices. The entries of $\hat{\boldsymbol{\lambda}}$ are denoted as $\hat{\lambda}_i$, $1 \leq i \leq n$.

We then construct a criterion $l(\mathbf{N}; \boldsymbol{\omega})$ to evaluate whether a matrix $\mathbf{N} \in \mathbb{R}^{m \times n}$ belongs to a high probability set in $\mathcal{S}_{m,n,p}$. Let the ordered singular values of \mathbf{N} be denoted as $\sigma_1(\mathbf{N}) \geq \sigma_2(\mathbf{N}) \geq \dots \geq \sigma_n(\mathbf{N})$, and define the vector $\boldsymbol{\sigma}(\mathbf{N}) = (\sigma_1(\mathbf{N}), \dots, \sigma_n(\mathbf{N}))$. The criterion calculates the weighted L_2 distance between $\boldsymbol{\sigma}(\mathbf{N})$ and $\hat{\boldsymbol{\lambda}}$, i.e.,

$$l(\mathbf{N}; \boldsymbol{\omega}) = \left\| \boldsymbol{\sigma}(\mathbf{N}) - \hat{\boldsymbol{\lambda}} \right\|_{\boldsymbol{\omega}}^2 = \sum_{i=1}^n \omega_i \left(\sigma_i(\mathbf{N}) - \hat{\lambda}_i \right)^2,$$

where $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)$ is the weight satisfying $\sum_{i=1}^n \omega_i = 1$.

For noisy matrix completion, any matrix $\mathbf{M} \in \mathbb{R}^{m \times n}$ close to \mathbf{M}_0 will result in the residual matrix

$$\mathbf{Y} - P_{\Omega}(\mathbf{M}) = P_{\Omega}(\mathbf{M}_0 - \mathbf{M}) + P_{\Omega}(\mathbf{H})$$

being close to $P_{\Omega}(\mathbf{H})$. The latter is located in a high probability set of $\mathcal{S}_{m,n,p}$ after variance adjustment. Therefore, we propose the following method to find an estimator $\widetilde{\mathbf{M}} \in \mathbb{M}$ for \mathbf{M}_0 as

$$(3) \quad \widetilde{\mathbf{M}} = \arg \min_{\mathbf{M} \in \mathbb{M}} l \left(\frac{\mathbf{Y} - P_{\Omega}(\mathbf{M})}{\sqrt{m} \hat{\sigma}(\mathbf{M})}; \boldsymbol{\omega} \right).$$

Here $\hat{\sigma}(\mathbf{M})$ is the estimated working standard deviation, defined as

$$(4) \quad \hat{\sigma}(\mathbf{M}) = \frac{\sum_{i=n/3}^{i=2n/3} \sigma_i(\mathbf{Y} - P_{\Omega}(\mathbf{M}))}{\sqrt{m} \sum_{i=n/3}^{i=2n/3} \lambda_i},$$

where $\sigma_i(\mathbf{Y} - P_{\Omega}(\mathbf{M}))$, $1 \leq i \leq n$ are the ordered singular values of $\mathbf{Y} - P_{\Omega}(\mathbf{M})$. The working estimator $\hat{\sigma}(\mathbf{M})$ approximates the standard deviation σ of the noise by calculating the proportion of the bulk singular values' sum from the residual matrix and sparse random matrices. This approach is efficient because the perturbation $P_{\Omega}(\mathbf{M}_0 - \mathbf{M})$ has less impact on the bulk singular values than on the edge singular values of $P_{\Omega}(\mathbf{H})$.

When the estimation space \mathbb{M} is chosen as $\mathbb{M}_s = \{ \mathbf{M} \mid \mathbf{M} = \mathbf{L}\mathbf{R}', \mathbf{L} \in \mathbb{R}^{m \times s}, \mathbf{R} \in \mathbb{R}^{n \times s} \}$, it results in the matrix factorization based estimator $\widetilde{\mathbf{M}}_{\text{fac}}$

$$(5) \quad \widetilde{\mathbf{M}}_{\text{fac}} = \arg \min_{\mathbf{M} = \mathbf{L}\mathbf{R}', \mathbf{L} \in \mathbb{R}^{m \times s}, \mathbf{R} \in \mathbb{R}^{n \times s}} l \left(\frac{\mathbf{Y} - P_{\Omega}(\mathbf{L}\mathbf{R}')}{\sqrt{m} \hat{\sigma}(\mathbf{L}\mathbf{R}')}; \boldsymbol{\omega} \right).$$

If the space \mathbb{M} is defined as $\mathbb{M}_C = \{ \mathbf{M} \in \mathbb{R}^{m \times n} \mid \|\mathbf{M}\|_* \leq C \}$, we equivalently obtain the convex relaxed nuclear norm estimator $\widetilde{\mathbf{M}}_{\text{nuc}}$

$$(6) \quad \widetilde{\mathbf{M}}_{\text{nuc}} = \arg \min_{\mathbf{M} \in \mathbb{R}^{m \times n}} l \left(\frac{\mathbf{Y} - P_{\Omega}(\mathbf{M})}{\sqrt{m} \hat{\sigma}(\mathbf{M})}; \boldsymbol{\omega} \right) + \lambda \|\mathbf{M}\|_*.$$

It is notable that our proposal, unlike the Frobenius norm based methods, incorporates the spectral properties of the noise matrix rather than solely minimizing the Frobenius norm of the residual matrix. In many statistical problems, random noise lacks meaningful structural information, and minimizing the sum of squared errors often yields optimal results. However, in models involving a fixed low-rank matrix plus a random noise matrix, the low-rank matrix perturbs only a few dominant singular values of the random matrix significantly. Thus, the spectral distribution of the residual matrix in noisy matrix completion can provide additional insights. Particularly, in scenarios where the Signal-to-Noise Ratio (SNR) $\|\mathbf{M}_0\|_F/\|\mathbf{H}\|_F$ is low, disregarding the information encoded in \mathbf{H} and merely reducing the norm of the residual matrix may result in a loss of efficiency.

3.2. Optimal algorithm. Minimizing the term $\|\mathbf{Y} - P_{\Omega}(\mathbf{M})\|_F^2$ in equations (1) and (2) is straightforward due to its strong convexity. However, in our proposed method, minimizing $l((\mathbf{Y} - P_{\Omega}(\mathbf{M}))/\sqrt{m}\hat{\sigma}(\mathbf{M}); \boldsymbol{\omega})$ is more challenging because it is non-convex and lacks of a computable gradient. To address this, we have developed efficient algorithms and their convergence properties will be demonstrated in the next section.

In our algorithms, \mathbf{M} and $\hat{\sigma}(\mathbf{M})$ are updated alternatively in each iteration step. When $p_{i,j} = 1$, $P_{\Omega}(\mathbf{M}) = \mathbf{M}$ and the gradient for \mathbf{M} of can be directly calculated as

$$(7) \quad \nabla_{\mathbf{M}} l\left(\frac{\mathbf{Y} - \mathbf{M}}{\sqrt{m}\hat{\sigma}}; \boldsymbol{\omega}\right) = \frac{1}{\sqrt{m}\hat{\sigma}} \sum_{i=1}^n \omega_i \left(\sigma_i(\mathbf{M}) - \sqrt{m}\hat{\sigma}\hat{\lambda}_i \right) \mathbf{u}_i(\mathbf{M}) \mathbf{v}'_i(\mathbf{M}),$$

where $\mathbf{u}_i(\mathbf{M})$, $\mathbf{v}_i(\mathbf{M})$ are the corresponding left and right singular vectors of $\mathbf{Y} - P_{\Omega}(\mathbf{M})$. In the context of matrix completion where $p_{i,j} < 1$, computing the gradient becomes non-trivial. A pivotal step in our algorithm is finding a direction that closely approximates the true gradient for iterative gradient descent. Inspired by the form of gradient in (7), we define the pseudo-gradient as

$$(8) \quad \hat{\nabla}_{\mathbf{M}} l\left(\frac{\mathbf{Y} - P_{\Omega}(\mathbf{M})}{\sqrt{m}\hat{\sigma}}; \boldsymbol{\omega}\right) = P_{\Omega} \left(\frac{1}{\sqrt{m}\hat{\sigma}} \sum_{i=1}^n \omega_i \left(\sigma_i(\mathbf{M}) - \sqrt{m}\hat{\sigma}\hat{\lambda}_i \right) \mathbf{u}_i(\mathbf{M}) \mathbf{v}'_i(\mathbf{M}) \right).$$

It is straightforward to verify that the pseudo-gradient coincides with the gradient when $p = 1$, and we will later demonstrate its adequacy in our algorithms for $p < 1$. We provide the algorithm for the estimator $\widetilde{\mathbf{M}}_{\text{fac}}$ in Algorithm 1 and defer the algorithm for the estimator $\widetilde{\mathbf{M}}_{\text{nuc}}$ to the Appendix.

Algorithm 1 Iteration Algorithm for $\widetilde{\mathbf{M}}_{\text{fac}}$

Input: Observed data $\mathbf{Y} \in \mathbb{R}^{m \times n}$, missing index matrix $\boldsymbol{\Omega} \in \mathbb{R}^{m \times n}$ and used rank s .

Output: Completion matrix \mathbf{M} .

- 1: $\mathbf{U}^{(0)}, \mathbf{D}^{(0)}, \mathbf{V}^{(0)} \leftarrow \text{svd}(p^{-1}\mathbf{Y})$;
 - 2: $\mathbf{R}^{(0)} \leftarrow \mathbf{U}_{[1:s]}^{(0)} (\mathbf{D}_{[1:s]}^{(0)})^{1/2}$ and $\mathbf{L}^{(0)} \leftarrow \mathbf{V}_{[1:s]}^{(0)} (\mathbf{D}_{[1:s]}^{(0)})^{1/2}$;
 - 3: Sample l standard Gaussian random matrix $\mathbf{M}_i \in \mathbb{R}^{m \times n}$ with singular value vector $\boldsymbol{\lambda}(\mathbf{M}_i)$, $1 \leq i \leq l$;
 - 4: $\hat{\boldsymbol{\lambda}} \leftarrow \sum_{i=1}^l \boldsymbol{\lambda}(\mathbf{M}_i)/l$, $\text{part_sum} \leftarrow \sum_{i=n/3}^{2n/3} \hat{\lambda}_i$;
 - 5: **while** $k \leq K$ **do**
 - 6: $\mathbf{U}^{(k+1)}, \mathbf{D}^{(k+1)}, \mathbf{V}^{(k+1)} \leftarrow \text{svd}(\mathbf{Y} - P_{\Omega}(\mathbf{R}^{(k)} \mathbf{L}^{(k)'}))$;
 - 7: $\hat{\sigma}^{(k+1)} \leftarrow \sum_{i=n/3}^{2n/3} \mathbf{D}_{i,i}^{(k+1)} / \text{part_sum}$;
 - 8: $\hat{\nabla}_{\mathbf{M}}^{(k+1)} \leftarrow \sum_{i=1}^n w_i (\mathbf{D}_{i,i}^{(k+1)} - \hat{\sigma}^{(k+1)} \hat{\lambda}_i) \mathbf{U}_{[i]}^{(k+1)} \mathbf{V}_{[i]}^{(k+1)'}$;
 - 9: $\mathbf{R}^{(k+1)} \leftarrow \mathbf{R}^{(k)} - \eta_{k+1} P_{\Omega}(\hat{\nabla}_{\mathbf{M}}^{(k+1)}) \mathbf{L}^{(k)}$;
 - 10: $\mathbf{L}^{(k+1)} \leftarrow \mathbf{L}^{(k)} - \eta_{k+1} P_{\Omega}(\hat{\nabla}_{\mathbf{M}}^{(k+1)}) \mathbf{R}^{(k)}$;
 - 11: **end while**
 - 12: **return** $\mathbf{M} = \mathbf{R}^{(K+1)} \mathbf{L}^{(K+1)'}$.
-

4. Theoretical results. Several common assumptions prevalent in noisy matrix completion researches are outlined here.

ASSUMPTION 1. The random noise matrix $\mathbf{H} \in \mathbb{R}^{m \times n}$ satisfies that

- (a) $H_{i,j}$, $1 \leq i \leq m, 1 \leq j \leq n$ are independent random variables with mean zero and variance σ^2 .

- (b) $\mathbf{H}_{i,j}$, $1 \leq i \leq m, 1 \leq j \leq n$ have sub-exponential tailed distribution, i.e., there exist constant $c > 0$ such that

$$P(|\mathbf{H}_{i,j}| \geq t) \leq \exp\left\{-\frac{t}{c\sigma}\right\}, \quad \forall t > T > 0.$$

- (c) The standard deviation σ is large enough such that

$$\sigma \gtrsim ((mn)^{-1/2} \|\mathbf{M}_0\|_F) r^{1/2} m^{-1/2} \log^{1/2} n.$$

ASSUMPTION 2. The underlying fixed matrix $\mathbf{M}_0 \in \mathbb{R}^{m \times n}$, $n/m = \rho \leq 1$, with a rank $r \ll n$ and singular value decomposition $\mathbf{M}_0 = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i'$, satisfies that

- (a) The condition number $\kappa = \sigma_1/\sigma_r$ is bounded by a constant,
- (b) The rank r matrix \mathbf{M}_0 is μ_0 -incoherent.

ASSUMPTION 3. Data are missing completely at random with observed probability $p_{i,j} = p$, $1 \leq i \leq m, 1 \leq j \leq n$ and $1 \geq p \gtrsim \mu_0 r m^{-1} \log^\alpha m$ for some large enough constant $\alpha > 0$.

Assumptions 1, 2, and 3 are standard for noisy matrix completion, except for (c) in Assumption 1. Assumption 1.(c) ensures the noise level is sufficiently high to make the structural information of the noise matrix \mathbf{H} useful. This mild assumption applies to a wide range of noise levels, approximately when $\sigma \geq m^{-1} n^{-1/2} \|\mathbf{M}\|_F$. Assumptions 3 can be relaxed to include a missing at random mechanism, where $p_{i,j}$ are not all equivalent and may depend on i . We assume Assumptions 1, 2, and 3 hold throughout this section.

4.1. Statistical properties of matrix factorization estimation. We consider the oracle properties of $\widetilde{\mathbf{M}}_{\text{fac}}$. This involves finding a matrix \mathbf{M} with rank at most s that minimizes the loss function

$$\mathcal{L}(\mathbf{M}; \mathbf{Y}, \Omega) = p^{-1} l\left(\frac{\mathbf{Y} - P_\Omega(\mathbf{M})}{\sqrt{m\hat{\sigma}}(\mathbf{M})}; \omega\right).$$

We verify that the loss function $\mathcal{L}(\mathbf{M}; \mathbf{Y}, \Omega)$ is reasonable by showing that $\mathcal{L}(\mathbf{M}_0; \mathbf{Y}, \Omega)$ is sufficiently small. Lemma 4.1 asserts that at the point \mathbf{M}_0 , the value of this loss function rapidly converges to zero with high probability.

LEMMA 4.1. *With probability larger than $1 - cn^{-3}$, it holds that*

$$\mathcal{L}(\mathbf{M}_0; \mathbf{Y}, \Omega) \leq C_0 \log^{-2} m,$$

with some constant $C_0 \geq 0$ independent of m, n, p, σ and \mathbf{M}_0 .

In Assumption 1, we do not assume a specific distribution for the entries of the noise matrix \mathbf{H} . Classical random matrix theory suggests that the spectrum of \mathbf{H} can be approximated by that of Gaussian random matrices. This property, known as universality, helps in determining suitable criteria for the spectrum of \mathbf{H} . However, a different scenario arises when considering the sparse random matrix $P_\Omega(\mathbf{H})$. A key technique for proving Lemma 4.1 is to establish a similar universality property of sparse random matrices.

Define the estimation space

$$\mathbb{M}_{s, \mu_1} = \{\mathbf{M} \in \mathbb{R}^{m \times n} \mid \text{rank}(\mathbf{M}) \leq s, \mathbf{M} \text{ is } \mu_1\text{-incoherent}\}.$$

Utilizing the matrix factorization $\mathbf{M} = \mathbf{L}\mathbf{R}'$, we can only partially fulfill the conditions such that $\text{rank}(\mathbf{M}) \leq s$ constrained in \mathbb{M}_{s,μ_1} . Although no constraint is explicitly imposed to ensure that \mathbf{M} adheres to any incoherent condition, we will demonstrate in Section 4.3 that the incoherent property is maintained implicitly throughout the iteration process. Therefore, we consider the properties of estimators constrained in the estimation space \mathbb{M}_{s,μ_1} .

Firstly, we focus on the scenario where the rank s of the estimation space \mathbb{M}_{s,μ_1} is properly selected, i.e., $s = r$. Consider the estimator

$$\widetilde{\mathbf{M}} = \arg \min_{\mathbf{M} \in \mathbb{M}_{r,\mu_1}} \mathcal{L}(\mathbf{M}; \mathbf{Y}, \boldsymbol{\Omega}).$$

We provide a lemma demonstrating that any sub-optimal estimator \mathbf{M} within the space \mathbb{M}_{r,μ_1} will result in a sufficiently high value for the loss function $\mathcal{L}(\mathbf{M}; \mathbf{Y}, \boldsymbol{\Omega})$.

LEMMA 4.2. *For $\mathbf{M} \in \mathbb{M}_{r,\mu_1}$, if*

$$\|\mathbf{M}_0 - \mathbf{M}\|_F - (3 + 5\kappa)(2 + 2\sqrt{\rho} + \mathcal{C}^{1/2})\sigma\sqrt{mr/p} \gtrsim (\log^{-1} m)\sigma\sqrt{mr/p}.$$

then with probability larger than $1 - cn^{-3}$,

$$\mathcal{L}(\mathbf{M}; \mathbf{Y}, \boldsymbol{\Omega}) > \mathcal{C}\omega_1.$$

Lemma 4.1 demonstrates that at the point \mathbf{M}_0 , the loss function yields a sufficiently small value. In contrast, Lemma 4.2 illustrates that an inadequate estimator \mathbf{M} results in a significantly large value of the loss function. Integrating these two findings, it is evident that there exists an $\mathbf{M} \in \mathbb{M}_{r,\mu_1}$ for which the loss $\mathcal{L}(\mathbf{M}; \mathbf{Y}, \boldsymbol{\Omega})$ is small. This minimized loss ensures a small estimation error $\|\mathbf{M}_0 - \mathbf{M}\|_F$, leading to Theorem 4.1 that gives the upper bound of the Frobenius norm error $\|\widetilde{\mathbf{M}} - \mathbf{M}_0\|_F$ with high probability.

THEOREM 4.1. *When the weight $\boldsymbol{\omega}$ satisfies that $\omega_1 \gg \log^{-2} m$, with probability larger than $1 - cn^{-3}$,*

$$\|\widetilde{\mathbf{M}} - \mathbf{M}_0\|_F \leq (3 + 5\kappa)(2 + 2\sqrt{\rho} + o(1))\sigma\sqrt{mr/p}.$$

REMARK 4.1. *The upper bound presented in Theorem 4.1 coincides with the minmax lower bound as shown in Theorem 3 in Negahban and Wainwright (2012) and Theorem 5 in Koltchinskii, Lounici and Tsybakov (2011), with an order of $\sigma\sqrt{mr/p}$. This is owing to the ratio $\rho = n/m \leq 1$, and we assume that the condition number κ is bounded by a constant. It has been demonstrated that methods minimizing the squared Frobenius norm of the residual matrix $\mathbf{Y} - P_\Omega(\mathbf{M})$ also achieve optimal bounds. Consequently, our bound aligns with results of those methods. Additionally, we provide an explicit term in the upper bound of Theorem 4.1, elucidating the influence of κ and ρ especially when they are no longer constant and increase as $m \rightarrow \infty$.*

Our method primarily relies on the spectral properties of the random noise matrix, which implies that the noise level cannot be too low. In Assumption 1.(c), we impose a constraint on the lower bound of σ . This can be also viewed as constraining the Signal-to-Noise Ratio (SNR) $\|\mathbf{M}_0\|_F/\|\mathbf{H}\|_F$, ensuring it is not too large. As previously discussed, this assumption holds for a wide range of noise levels, requiring only that the standard deviation σ of the noise exceeds approximately $m^{-1}n^{-1/2}\|\mathbf{M}_0\|_F$. Even if this mild assumption is not satisfied, we can still bound the estimation error. The error bound will be affected by the Frobenius norm of \mathbf{M}_0 if (c) in Assumption 1 is not satisfied, as demonstrated in Corollary 4.1.

COROLLARY 4.1. If Assumption 1.(c) does not hold, when the weight ω satisfies that $\omega_1 \gg \log^{-2} m$, with probability larger than $1 - cn^{-3}$,

$$\|\widetilde{\mathbf{M}} - \mathbf{M}_0\|_F \leq (3 + 5\kappa)(2 + 2\sqrt{\rho} + o(1))\sigma\sqrt{\frac{mr}{p}} + 8(3 + 5\kappa)(\mu_0 + \mu_1)\sqrt{\frac{r^3 \log n}{mnp}}\|\mathbf{M}_0\|_F.$$

REMARK 4.2. The additional term in Corollary 4.1 becomes dominant when the standard deviation σ of the noise is very small. In the extreme case of matrix completion without noise, i.e., $\sigma = 0$, Theorem 1.2 in Candès and Tao (2010) demonstrates that exact completion can be achieved with high probability using (2). In scenarios with minimal noise, uncertainty in the spectral properties of the noise matrix \mathbf{H} can obscure its structural information, reducing the effectiveness of our method. Directly minimizing the Frobenius norm of the residual matrix $\mathbf{Y} - P_\Omega(\mathbf{M})$ is more effective in this case. However, when the noise level is high, which is the more challenging scenario, the advantage of our method becomes evident.

Several studies have investigated the estimation and testing issues concerning the underlying rank r of \mathbf{M}_0 , e.g., Zhang, Shapiro and Xie (2019) and Li et al. (2023). When the noise level is significant, precisely estimating the true rank of \mathbf{M}_0 may not always be feasible. Specifically, there is uncertainty about whether the selected low-rank space \mathbb{M}_{s,μ_1} matches \mathbb{M}_{r,μ_1} . Misjudging the underlying rank of \mathbf{M}_0 results in increased estimation errors. Nevertheless, for $s \geq r$ that is not excessively large, we establish that the Frobenius norm error of the estimator $\widetilde{\mathbf{M}}_s$, defined as

$$\widetilde{\mathbf{M}}_s = \arg \min_{\mathbf{M} \in \mathbb{M}_{s,\mu_1}} \mathcal{L}(\mathbf{M}; \mathbf{Y}, \Omega)$$

can still be controlled at a low level.

COROLLARY 4.2. When $r \leq s \leq mp \log^{-2} m$, the weight ω satisfies that $\omega_1 \gg \log^{-2} m$ and the noise standard deviation $\sigma \geq s^{3/2}r^{-1}m^{-1}n^{-1/2}\log^{1/2} n\|\mathbf{M}_0\|_F$, with probability larger than $1 - cn^{-3}$,

$$\|\widetilde{\mathbf{M}}_s - \mathbf{M}_0\|_F \leq \left((2 + 4\kappa)r^{1/2} + (1 + \kappa)s^{1/2} \right) (2 + 2\sqrt{\rho} + o(1))\sigma\sqrt{m/p}.$$

When s is selected as $s = r$, the upper bound stated in Corollary 4.2 aligns with the results of Theorem 4.1. If $s > r$, the estimation error $\|\widetilde{\mathbf{M}}_s - \mathbf{M}\|_F$ is bounded by the order $\sigma\sqrt{sm/p}$ when both κ and ρ are considered constants. A higher value of s corresponds to a larger estimation space \mathbb{M}_{s,μ_1} , leading to increased errors. An additional source of error arises from including the top singular subspace of the noise matrix in the estimator $\widetilde{\mathbf{M}}_s$, due to the excess rank $s > r$. Frobenius norm based methods tend to capture the entire top subspace of noise when minimizing the Frobenius norm of the residual matrix. In contrast, our proposed method is less affected because it also considers the spectral distribution of the random noise matrix.

4.2. Statistical properties of convex relaxation estimation. In addition to the strictly low-rank constraint $\mathbf{M} = \mathbf{L}\mathbf{R}'$, several studies such as Candès and Recht (2009) suggest that the nuclear norm $\|\mathbf{M}\|_*$ serves as an effective convex relaxation alternative for the exact rank. Our method, which aims to minimize the spectral distance between the residual matrix and sparse random matrices, can employ a similar convex relaxation procedure for low-rank constraints. Formally, we investigate the properties of the estimator $\widetilde{\mathbf{M}}_\lambda$, defined as

$$\widetilde{\mathbf{M}}_\lambda = \arg \min_{\mathbf{M} \in \mathbb{R}^{m \times n}} \mathcal{L}(\mathbf{M}; \mathbf{Y}, \Omega) + \lambda\|\mathbf{M}\|_*.$$

Methods that minimizing the Frobenius norm benefit from the convex relaxation, as the overall estimation problem is strongly convex. This property enhances both theoretical analysis and computational feasibility. Previous techniques demonstrating the performance guarantee of $\widehat{\mathbf{M}}_{\text{nuc}}$ often relies on the convexity of the squared Frobenius norm and the nuclear norm. In contrast, analyzing our method is more challenging due to the non-convexity of $\mathcal{L}(\mathbf{M}; \mathbf{Y}, \Omega)$. When focusing on estimation properties, the estimation space $\mathbb{M}_C = \{\mathbf{M} \in \mathbb{R}^{m \times n} \mid \|\mathbf{M}\|_* \leq C\}$ is more difficult to control compared to \mathbb{M}_{s, μ_1} . This difficulty arises from the ambiguity surrounding the rank of matrices in \mathbb{M}_C . The following theorem provides an error bound for the estimator $\widetilde{\mathbf{M}}_\lambda$ with high probability for any $\lambda \geq 0$.

THEOREM 4.2. *For some $\bar{\omega} > 0$, define that*

$$R_1 = 2(1 + \sqrt{\rho})\sigma\sqrt{m/p}, \quad R_2 = R_1 + \frac{\lambda\sigma^2 m}{\bar{\omega} p} + \left(\frac{\lambda\sigma^2 m}{\bar{\omega} p}\right)^{1/2} R_1^{1/2}.$$

If $\min_{1 \leq i \leq r(R_1 + R_2)/R_1} \omega_i \geq \bar{\omega}$, with probability larger than $1 - cn^{-3}$,

$$\begin{aligned} \|\widetilde{\mathbf{M}}_\lambda - \mathbf{M}_0\|_F &\leq (4 + 5\kappa) \left(\left(\frac{\lambda\sigma^2 m}{\bar{\omega} p} \right)^{1/2} (rR_2 - R_1)^{1/2} + R_1 + R_2 \right) \\ &\quad + ((C_0 + 1)\lambda^{-1} \log^{-2} m + 2rR_2)^{1/2} R_1^{1/2}. \end{aligned}$$

The upper bound in Theorem 4.2 illustrates two different sources of error related to the choice of the tuning parameter λ . One source is the term $(C_0 + 1)\lambda^{-1} \log^{-2} m$, which arises from the flexibility of the estimation space. A higher λ reduces this flexibility, thereby decreasing the generalization error. The other source is the term $\lambda\sigma^2 m / \bar{\omega} p$, which results from the shrinkage of singular values due to nuclear norm regularization. A smaller λ leads to decreased approximation error. These two distinct sources of error create a trade-off when tuning λ to minimize the bound, as illustrated in Corollary 4.3.

COROLLARY 4.3. *When the weight ω satisfies that $\min_{1 \leq i \leq 4r} \omega_i \gg \log^{-2} m$, choosing $\lambda = 2(1 + \sqrt{\rho})\sigma^{-1}p^{1/2}m^{-1/2}\log^{-2} m$, with probability larger than $1 - cn^{-3}$,*

$$\|\widetilde{\mathbf{M}}_\lambda - \mathbf{M}_0\|_F \leq 2(1 + \sqrt{\rho}) \left((\sqrt{6} + 4\sqrt{3} + 5\sqrt{3}\kappa)r^{1/2} + 4(4 + 5\kappa) + o(1) \right) \sigma\sqrt{m/p}.$$

By adjusting the tuning parameter λ , our results in Corollary 4.3 match the minmax lower bound for noisy matrix completion. Estimation properties for methods based on minimizing the squared Frobenius norm of the residual matrix with nuclear norm regularization, have been well understood. For example, assuming \mathbf{M}_0 is a μ -incoherent matrix, Corollary 2 in [Koltchinskii, Lounici and Tsybakov \(2011\)](#) provides a nearly optimal upper bound of the order $\sigma\sqrt{mr\log m/p}$. Compared to results from classical convex relaxation methods, our upper bound no longer contains logarithmic factors.

4.3. Algorithmic properties of iterative procedure. Iterative gradient descent algorithms are proposed to find approximate solutions, as closed-form solutions for the proposed methods are not feasible. Due to the non-convex nature of the loss function $\mathcal{L}(\mathbf{M}; \mathbf{Y}, \Omega)$ and the non-convex constraint $\text{rank}(\mathbf{M}) \leq s$, general optimization theory does not guarantee global convergence. It is essential to accurately measure the performance gap between estimators (often the global minimizer) and the results obtained through iterative algorithms. For the

constraint $\text{rank}(\mathbf{M}) \leq s$, methodologies such as matrix factorization $\mathbf{M} = \mathbf{L}\mathbf{R}'$ with alternating descent schemes, have been comprehensively researched. Therefore, our primary focus is on addressing the challenges arising from the proposed loss function $\mathcal{L}(\mathbf{M}; \mathbf{Y}, \boldsymbol{\Omega})$.

We formally define the iterative procedure for proposed method $\min_{\mathbf{M} \in \mathbb{M}_{s, \mu_1}} \mathcal{L}(\mathbf{M}; \mathbf{Y}, \boldsymbol{\Omega})$. With a feasible initial value $\mathbf{M}^{(0)}$, we define the iterative series $\{\mathbf{M}^{(k)}\}_{k=0,1,\dots}$ and the nuisance series $\{\mathbf{N}^{(k)}\}_{k=1,2,\dots}$ as

$$\begin{aligned}\mathbf{N}^{(k+1)} &= \mathbf{M}^{(k)} - \eta_{k+1} \hat{\nabla} \mathcal{L}(\mathbf{M}^{(k)}; \mathbf{Y}, \boldsymbol{\Omega}), \\ \mathbf{M}^{(k+1)} &= \text{Proj}_s(\mathbf{N}^{(k+1)}), \quad k = 0, 1, \dots,\end{aligned}$$

where

$$\hat{\nabla} \mathcal{L}(\mathbf{M}; \mathbf{Y}, \boldsymbol{\Omega}) = p^{-1} \hat{\nabla}_{\mathbf{M}} l \left(\frac{\mathbf{Y} - P_{\boldsymbol{\Omega}}(\mathbf{M})}{\sqrt{m} \hat{\sigma}}; \boldsymbol{\omega} \right)$$

is the pseudo-gradient given by (8). We also define the singular value decomposition for $\mathbf{M}^{(k)}$ as $\mathbf{M}^{(k)} = \mathbf{U}^{(k)} \mathbf{D}^{(k)} \mathbf{V}^{(k)}$.

There are three primary challenges. First, control the influence by using the pseudo-gradient $\hat{\nabla} \mathcal{L}(\mathbf{M}; \mathbf{Y}, \boldsymbol{\Omega})$ instead of the true gradient. Second, bound the impact of projecting $\mathbf{N}^{(k+1)}$ onto the rank s matrix $\mathbf{M}^{(k+1)}$. Finally, maintain the incoherent condition throughout the iterative process. We examine the scenario where $s = r$ and establish bounds for the iterative series in the proposed algorithms.

We begin our algorithm analysis with the selection of the initial value $\mathbf{M}^{(0)}$. Since $p^{-1} P_{\boldsymbol{\Omega}}(\mathbf{Y})$ is an unbiased estimator of \mathbf{M}_0 , a commonly used initial value for matrix completion problems is the rank- r projection of this estimator, i.e.,

$$\mathbf{M}^{(0)} = \text{Proj}_r(p^{-1} P_{\boldsymbol{\Omega}}(\mathbf{Y})).$$

Error bounds for the Frobenius and infinity norms, as well as an incoherent bound for $\mathbf{M}^{(0)}$, are provided in Lemma 4.3.

LEMMA 4.3. *With probability larger than $1 - cn^{-3}$, it holds that*

$$\begin{aligned}\|\mathbf{M}^{(0)} - \mathbf{M}_0\|_F &\leq C(\rho^{-1/4} \|\mathbf{M}_0\|_{\infty} + \sigma) \kappa \sqrt{\frac{mr}{p}} \\ \|\mathbf{M}^{(0)} - \mathbf{M}_0\|_{\infty} &\leq \|\mathbf{U}\|_{2,\infty} \|\mathbf{V}\|_{2,\infty} (\|\mathbf{M}^{(0)} - \mathbf{M}_0\|_F + \sqrt{r} \|\mathbf{H}\|)(1 + f^{(0)})^2 \\ \|\mathbf{U}^{(0)}\|_{2,\infty} &\leq \|\mathbf{U}\|_{2,\infty} (1 + f^{(0)}), \quad \|\mathbf{V}^{(0)}\|_{2,\infty} \leq \|\mathbf{V}\|_{2,\infty} (1 + f^{(0)}),\end{aligned}$$

where

$$f^{(0)} = C_1 \sqrt{r}(\kappa - 1) + C_2(\kappa + \sqrt{r})(\|\mathbf{M}_0\|_{\infty} + 6\sigma c \log m)^{1/2} \left(\frac{m}{p\sigma_r^2} \right)^{1/2} \log m.$$

In Lemma 4.3, we demonstrate that $\mathbf{M}^{(0)}$ serves as a suitable initial matrix. It is sufficiently close to \mathbf{M}_0 and preserves the crucial incoherence condition without excessively increasing the incoherence coefficient. Using mathematical induction, we establish the main result below that tracks both the error $\mathbf{M}^{(k)} - \mathbf{M}_0$ and the incoherence condition for $\mathbf{M}^{(k)}$, $k = 1, 2, \dots$.

THEOREM 4.3. *With probability larger than $1 - cn^{-3}$, for each $M^{(k)}$ in the iterative series $\{M^{(k)}\}_{k=1,2,\dots}$ with $\|M^{(k)} - M_0\| \geq C_3 \kappa \sigma \sqrt{m/p}$, it holds that*

$$\begin{aligned} \|M^{(k)} - M_0\|_F &\leq \|M^{(0)} - M_0\|_F \prod_{i=0}^{k-1} (1 - \gamma_i(1 - \alpha_i)), \\ \|U^{(k)}\|_{2,\infty} &\leq \|U^{(0)}\|_{2,\infty} \prod_{i=0}^{k-1} \left(1 + C\mu_0 r^4 \frac{\|M^{(i)} - M_0\|_F}{\sigma_r^{(k)}} \gamma_i \alpha_i \log^{3/2} m \right), \\ \|V^{(k)}\|_{2,\infty} &\leq \|V^{(0)}\|_{2,\infty} \prod_{i=0}^{k-1} \left(1 + C\mu_0 r^4 \frac{\|M^{(i)} - M_0\|_F}{\sigma_r^{(k)}} \gamma_i \alpha_i \log^{3/2} m \right), \\ \|M^{(k)} - M_0\|_\infty &\leq 2\|M^{(k)} - M_0\|_F \|U^{(k)}\|_{2,\infty} \|V^{(k)}\|_{2,\infty}, \end{aligned}$$

where we define that $\sigma_r^{(k)}$ is the r th singular value of $M^{(k)}$, $\kappa_r^{(k)}$ is the condition number, $C_2^{(k)} < 1$ is a sufficiently small constant and

$$\begin{aligned} \gamma_k &= \frac{2\eta_{k+1}}{\hat{\sigma}(M^{(k)})^2} \frac{(\|P_\Omega(M^{(k)} - Y)\| - \hat{\sigma}(M^{(k)})\lambda_1)}{\|M^{(k)} - M_0\|_F}, \\ \alpha_k &= \left(C_2^{(k)} + \left(\frac{(\sigma_r^{(k)} \kappa_r^{(k)})^2 (2 - 2r(\|M^{(k)} - M_0\|_F / \sigma_r^{(k)})^{-2})^2}{\|M^{(k)} - M_0\|_F^2} + 1 \right)^{-1/2} \right). \end{aligned}$$

REMARK 4.3. When treating $\rho = 1 - \gamma_i(1 - \alpha_i) < 1$ as a constant, we obtain an exponentially decaying sequence such that $\|M^{(k)} - M_0\|_F \leq \rho^k \|M^{(0)} - M_0\|_F$ holds if $\|M^{(k)} - M_0\| \geq C_3 \kappa \sigma \sqrt{m/p}$. Theorem 2 in Ma et al. (2018) provides a similar decay for the gradient descent of the problem (1). A primary challenge it addresses is the ill-posed representation $M = LR'$ which lacks identifiability. In our theorem, we omit the matrix factorization and the alternating optimal step. We mainly focus on the issues related to the non-convex loss $\mathcal{L}(M; Y, \Omega)$ in our proposed method. Expanding our results to the alternating gradient descent algorithm with the representation $M = LR'$ is straightforward by integrating their techniques.

In Theorem 4.3, we demonstrate that the Frobenius and infinity norms of $M^{(k)} - M_0$ decrease successively while controlling the growth rates of $\|U^{(k)}\|_{2,\infty}$ and $\|V^{(k)}\|_{2,\infty}$. This ensures that with a finite step k , we can obtain an $M^{(k)}$ with small errors in both the Frobenius and infinity norms while maintaining the incoherence condition with a small incoherence coefficient.

COROLLARY 4.4. If $p \gtrsim \mu_0^4 \kappa^4 r^{11} m^{-1} \log^\alpha m$, for sufficiently small step sizes $\eta_k, k = 1, 2, \dots$, there exist $K > 0$ with probability larger than $1 - cn^{-3}$ that

$$\begin{aligned} \|M^{(K)} - M_0\|_F &\lesssim \kappa \sigma \sqrt{mr/p}, \\ \|M^{(K)} - M_0\|_\infty &\lesssim \kappa \sigma \mu_0 \sqrt{r^3 \log^3 m / np}. \end{aligned}$$

REMARK 4.4. Corollary 4.4 demonstrates that $M^{(K)}$ is a reliable approximation for the estimator \tilde{M} . It has the same order in Frobenius error bounds as \tilde{M} , as established in Theorem 4.1. The additional assumption in Corollary 4.4 is slightly more stringent than Assumption 3. Treating μ_0, κ , and r as constants, the lower bound of observation probability $p \gtrsim m^{-1} \log^\alpha m$ still aligns with the conditions specified in Assumption 3.

Corollary 4.4 confirms that the algorithmic error does not dominate the statistical error. Our proposed algorithm achieves this by ensuring a sufficiently small step size in each update and an adequate number of iterations. As a result, there is no gap in the error bound rates between the oracle estimators and the results obtained through numerical methods. This is important because it guarantees the numerical performance of the proposed algorithm for our method.

5. Numerical results. To show the numerical performance of our method, we compare the proposed estimators $\widetilde{\mathbf{M}}_{\text{fac}}$ and $\widetilde{\mathbf{M}}_{\text{nuc}}$, presented in equations (5) and (6), to the Frobenius norm minimizing estimators $\widehat{\mathbf{M}}_{\text{fac}}$ and $\widehat{\mathbf{M}}_{\text{nuc}}$, described in equations (1) and (2). This comparison is conducted using both simulation and real data experiments. For clarity, we denote $\widetilde{\mathbf{M}}_{\text{fac}}$ and $\widetilde{\mathbf{M}}_{\text{nuc}}$ as Estimator 1 and Estimator 2, with their corresponding baselines $\widehat{\mathbf{M}}_{\text{fac}}$ and $\widehat{\mathbf{M}}_{\text{nuc}}$ being Baseline 1 and Baseline 2.

For the implementation details, we select the initial value $\mathbf{M}^{(0)} = \text{Proj}_s(p^{-1}\mathbf{Y})$ for $\widetilde{\mathbf{M}}_{\text{fac}}$ and $\widehat{\mathbf{M}}_{\text{fac}}$, which represents the rank- s projection of the unbiased estimator $p^{-1}\mathbf{Y}$. We also set $\mathbf{M}^{(0)} = p^{-1}\mathbf{Y}$ as the initial value for $\widetilde{\mathbf{M}}_{\text{nuc}}$ and $\widehat{\mathbf{M}}_{\text{nuc}}$. Subsequently, we apply the alternating gradient descent algorithm (Algorithm 2 in Ma et al. (2018)) to obtain $\widetilde{\mathbf{M}}_{\text{fac}}$ and use the singular value shrinkage method to obtain $\widetilde{\mathbf{M}}_{\text{nuc}}$. Our estimators are calculated using the proposed algorithms, utilizing the pseudo-gradient (8) in gradient descent step. Three types of error are considered to measure performance of the estimated \mathbf{M} : 1) Frobenius norm error, $\mathcal{E}_F = \|\mathbf{M} - \mathbf{M}_0\|_F/\sqrt{mn}$; 2) Spectral norm error, $\mathcal{E}_{sp} = \|\mathbf{M} - \mathbf{M}_0\|/\sqrt{m}$; and 3) Maximum norm error, $\mathcal{E}_\infty = \|\mathbf{M} - \mathbf{M}_0\|_\infty$.

5.1. Simulation study. In simulations, the observation matrix $\mathbf{Y} \in \mathbb{R}^{m \times n}$ is generated by $\mathbf{Y} = P_\Omega(\mathbf{M}_0 + \mathbf{H})$. Here Ω is an index matrix with entries following i.i.d. binomial distributions with probability p . Each entry in the noise matrix \mathbf{H} is independently generated as the sum of a mean zero Gaussian random variable with variance $\sigma^2/2$ and a Rademacher random variable multiplied by $\sigma/\sqrt{2}$. The underlying true matrix \mathbf{M}_0 is generated using the singular value decomposition $\mathbf{M}_0 = \mathbf{U}_0 \mathbf{D}_0 \mathbf{V}_0'$. Here $\mathbf{U}_0 \in \mathbb{R}^{m \times r}$ and $\mathbf{V}_0 \in \mathbb{R}^{n \times r}$ are the leading r left and right singular vectors of a randomly generated standard Gaussian matrix $\mathbf{W} \in \mathbb{R}^{m \times n}$. The delocalization property of random Gaussian matrices ensures that \mathbf{M}_0 is μ -incoherent with a small μ . Here $\mathbf{D}_0 \in \mathbb{R}^{r \times r}$ is a diagonal matrix with $[\mathbf{D}_0]_{i,i} = \sigma_i$, $1 \leq i \leq r$, the singular values of \mathbf{M}_0 . Three different true ranks of \mathbf{M}_0 are chosen: $r = 5, 10, 20$, and the singular values σ_i , $1 \leq i \leq r$ are set as

$$\begin{aligned} \text{Case 1: } r = 5, \quad \sigma_i &= (3 + i/3)\sqrt{m}(1 + \sqrt{n/m}), \quad 1 \leq i \leq 5; \\ \text{Case 2: } r = 10, \quad \sigma_i &= (3 + i/3)\sqrt{m}(1 + \sqrt{n/m}), \quad 1 \leq i \leq 10; \\ \text{Case 3: } r = 20, \quad \sigma_i &= (3 + (i+1)/6)\sqrt{m}(1 + \sqrt{n/m}), \quad 1 \leq i \leq 20. \end{aligned}$$

Firstly, we examine how the matrix sizes m, n of \mathbf{M}_0 affect errors of the proposed estimators. We set the standard deviation of \mathbf{H} to $\sigma = 1$ and the observation probability p to 0.2. For the estimators $\widetilde{\mathbf{M}}_{\text{fac}}$ and $\widehat{\mathbf{M}}_{\text{fac}}$, we use $s = r$. The observed data is randomly divided into 80% for training and 20% for validation. The tuning parameter λ in calculating $\widetilde{\mathbf{M}}_{\text{nuc}}$ and $\widehat{\mathbf{M}}_{\text{nuc}}$ is determined by minimizing the mean squared error on the validation set. In Table 2, we present the average Frobenius norm error \mathcal{E}_F from 100 repeated runs, varying the matrix size m from 200 to 800 with $n/m = 0.5$. The proposed estimators $\widetilde{\mathbf{M}}_{\text{fac}}$ and $\widetilde{\mathbf{M}}_{\text{nuc}}$ demonstrate lower Frobenius norm error compared to corresponding baselines. The average spectral norm error \mathcal{E}_{sp} and maximum norm error \mathcal{E}_∞ show similar results and are provided

TABLE 2

The Frobenius norm error \mathcal{E}_F of proposed methods and corresponding baselines for $\sigma = 1$, observation probability $p = 0.2$, rank of M_0 as 5, 10, 20 and matrix size m from 200 to 800 with $n/m = 0.5$.

Rank	Estimator	m=200	m=300	m=400	m=500	m=600	m=700	m=800
$r = 5$	Estimator 1: \tilde{M}_{fac}	0.8422	0.6283	0.5014	0.4374	0.3967	0.3647	0.3418
	Baseline 1: \widehat{M}_{fac}	0.9007	0.6372	0.5103	0.4429	0.3972	0.3678	0.3406
	Estimator 2: \tilde{M}_{nuc}	0.8108	0.5913	0.4860	0.4252	0.3858	0.3566	0.3349
	Baseline 2: \widehat{M}_{nuc}	0.9052	0.6325	0.5157	0.4482	0.4021	0.3675	0.3406
$r = 10$	Estimator 1: \tilde{M}_{fac}	2.5518	1.0856	0.8025	0.6715	0.5947	0.5397	0.4977
	Baseline 1: \widehat{M}_{fac}	3.1489	1.1402	0.8220	0.6826	0.5989	0.5457	0.5008
	Estimator 2: \tilde{M}_{nuc}	2.1028	1.0409	0.7731	0.6565	0.5803	0.5280	0.4974
	Baseline 2: \widehat{M}_{nuc}	2.7178	1.1408	0.8297	0.6896	0.6058	0.5450	0.5007
$r = 20$	Estimator 1: \tilde{M}_{fac}	3.7134	3.0684	1.8963	1.2775	1.0178	0.8797	0.7999
	Baseline 1: \widehat{M}_{fac}	3.8373	3.8161	2.4105	1.3799	1.0568	0.9210	0.8125
	Estimator 2: \tilde{M}_{nuc}	3.7677	2.7677	1.6476	1.1842	0.9719	0.8480	0.7921
	Baseline 2: \widehat{M}_{nuc}	4.4844	4.3773	2.5853	1.5286	1.1511	0.9500	0.8267

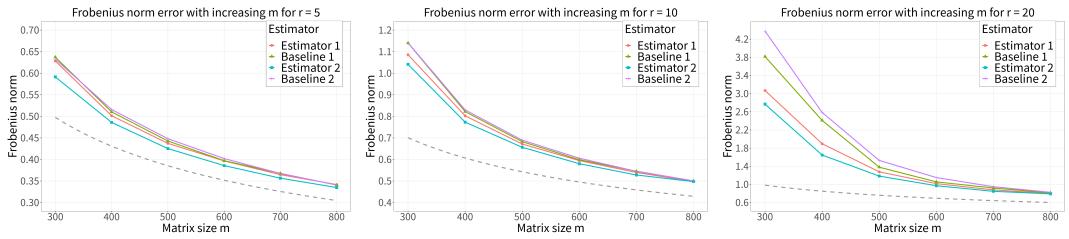


FIG 1. The trend curve of the Frobenius norm error \mathcal{E}_F for our estimators and baselines when matrix size approaches infinity. The plots from left to right correspond to the scenarios $r = 5, 10, 20$.

in the supplementary materials due to space limit. In Figure 1, we illustrate the trend of the Frobenius norm error as the matrix size m approaches infinity. The dashed line represents $p^{-1/2}$ times the Frobenius norm error for \tilde{M}_{rmt} in the matrix denoising model, serving as the optimal lower bound of order $\sigma\sqrt{r/mp}$. As m increases, errors of the proposed methods decrease following the theoretical order, which coincide with the theoretical findings in Theorem 4.1 and Corollary 4.3.

Secondly, we examine the singular value distributions of residual matrices in both the proposed and baseline methods. Figure 2 illustrates that the empirical spectral distributions of the residual matrices for our estimators closely resemble the expected spectral distribution of sparse random matrices, scaled by the estimated working standard deviation (4). In contrast, the residual matrices in the baselines deviate from the behavior of a sparse random matrix, exhibiting outliers in both the top and bottom singular values.

Then, we investigate the influence of misspecifying the true rank in Estimator 1 and Baseline 1, i.e., the effect of the parameter s on the estimators \tilde{M}_{fac} and \widehat{M}_{fac} . Table 3 presents the average Frobenius norm error \mathcal{E}_F for 100 repeated runs, with s varying from the true rank r to $r + 6$, while keeping all other settings as previously described. As s gradually deviates from the true rank r , the estimation errors of both Estimator 1 and Baseline 1 increase. However, the error of our estimator rises at a significantly slower rate, demonstrating greater robustness, which is evident shown by Figure 3. The dashed line represents the behavior of \tilde{M}_{rmt} , similar to what is shown in Figure 1. The average spectral norm error \mathcal{E}_{sp} and maximum norm error \mathcal{E}_∞ show similar patterns and are also provided in the supplementary materials. It shows that as $s \geq r$ increases, the Frobenius norm error remains approximately at the theoretical order of $\sigma\sqrt{(s+r)/mp}$, consistent with the findings in Corollary 4.2.

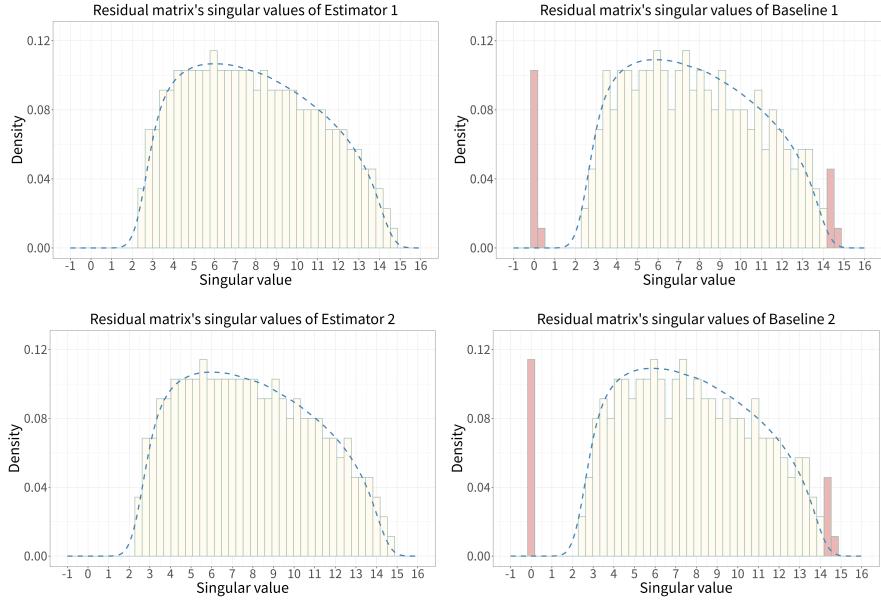


FIG 2. The singular value distributions of residual matrix for our estimators and baselines in the scenarios $r = 10$ and matrix size 500×250 . The left two plots correspond to the results of Estimator 1 and Baseline 1, and the right two plots correspond to the results of Estimator 2 and Baseline 2.

TABLE 3

The Frobenius norm error \mathcal{E}_F of Estimator 1 and Baseline 1 for true rank of M_0 as 5, 10, 20, used rank s from r to $r + 6$ and matrix size 500×250 .

Rank	Estimator	$s = r$	$s = r+1$	$s = r+2$	$s = r+3$	$s = r+4$	$s = r+5$	$s = r+6$
$r = 5$	Estimator 1: \tilde{M}_{fac}	0.4374	0.4907	0.5452	0.5993	0.6516	0.7012	0.7474
	Baseline 1: \tilde{M}_{fac}	0.4429	0.5236	0.5996	0.6740	0.7451	0.8142	0.8827
$r = 10$	Estimator 1: \tilde{M}_{fac}	0.6715	0.7205	0.7712	0.8238	0.8761	0.9286	0.9804
	Baseline 1: \tilde{M}_{fac}	0.6826	0.7633	0.8435	0.9265	1.0101	1.0927	1.1757
$r = 20$	Estimator 1: \tilde{M}_{fac}	1.2775	1.3076	1.3585	1.4156	1.4701	1.5262	1.5784
	Baseline 1: \tilde{M}_{fac}	1.3799	1.4997	1.6380	1.7688	1.8860	1.9985	2.0922

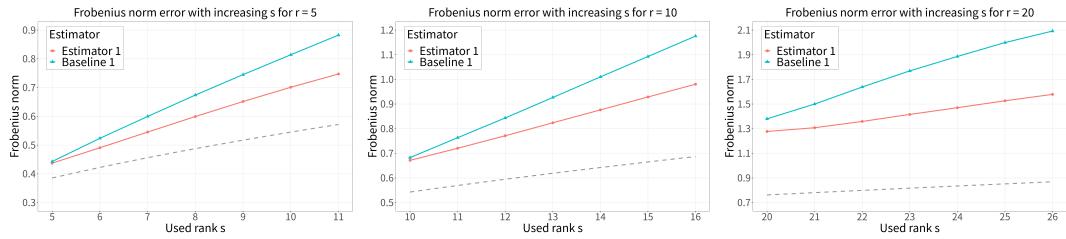


FIG 3. The trend curve of the Frobenius norm error \mathcal{E}_F for Estimator 1 and Baseline 1 when $s \geq r$ increases. The plots from left to right correspond to the scenarios $r = 5, 10, 20$.

Finally, we evaluate the effectiveness of our method compared to baselines under varying noise levels. We use the ratio of the the spectral norm of \mathbf{H} , to the r th singular value of M_0 multiplied by the rooted observation probability, as a metric for assessing the intensity of noise, i.e., $R = \|\mathbf{H}\|/(p^{1/2}\sigma_r(M_0))$. We fix M_0 and vary the standard deviation σ of the noise, from 0.01 to 5. In Figure 4, when the noise level is low (approximately $R \leq 0.5$), the spectral information of the noise matrix \mathbf{H} is difficult to use, making Frobenius norm based methods slightly better. However, when the noise level increases ($R > 0.5$), our method

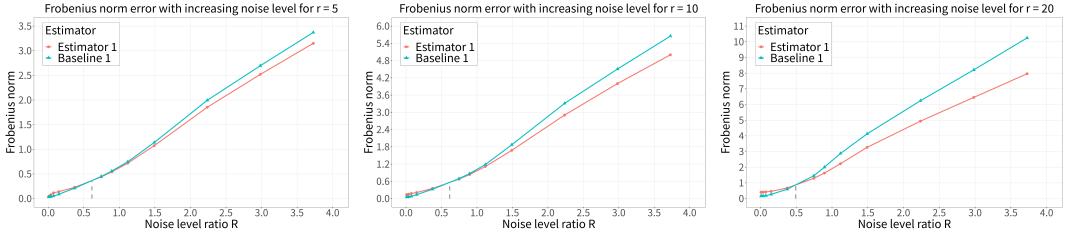


FIG 4. The Frobenius norm error \mathcal{E}_F for Estimator 1 and Baseline 1 when noise level ratio R increases. The plots from left to right correspond to the scenarios $r = 5, 10, 20$.

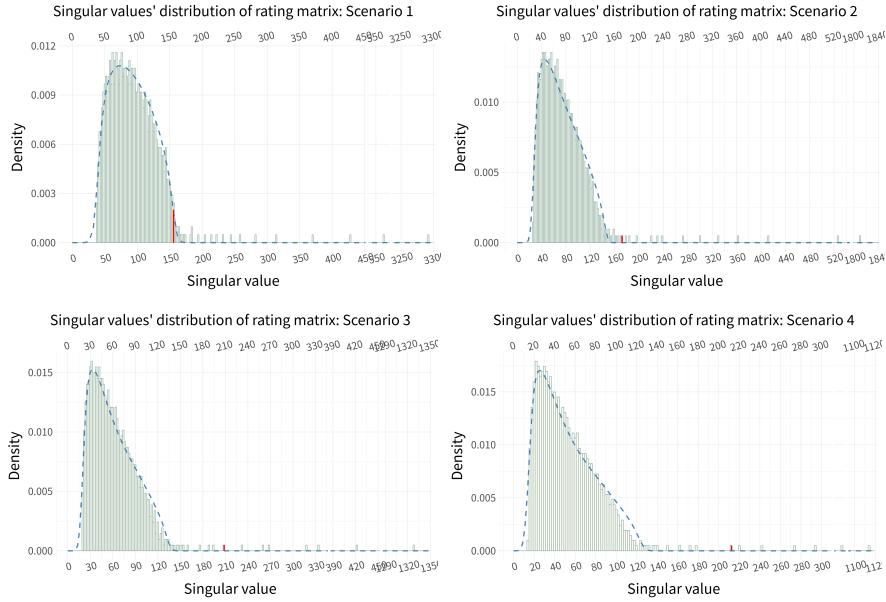


FIG 5. The spectral distribution of the observed rating matrix and selected rank s in each scenario respectively. The dash lines are densities of expectation of singular values of corresponding variance-adjusted sparse random matrices. The red lines shows the selected rank for Estimator 1.

performs significantly more effective. This indicates that our proposed method is especially suitable for scenarios with relatively high noise levels.

5.2. Netflix rating data. We utilize the Netflix Prize Dataset to demonstrate the performance of the proposed method. This dataset is detailed in [Bennett et al. \(2007\)](#) and is publicly available at <https://www.kaggle.com/datasets/netflix-inc/netflix-prize-data>. The dataset contains user ratings (ranging from 1 to 5) for various movies on Netflix from October 1998 to December 2005. Due to the vast number of movies available, each user has rated only a small portion of them. Applying matrix completion directly to this dataset allows for predicting users' ratings for movies they have not viewed, thereby developing a recommendation system. The dataset is preprocessed by retaining movies rated more than 25,000 times, excluding lesser-known films. Four distinct user groups are considered. In scenario 1, we focus on users who are actively engaged in watching and rating movies, retaining a total of 3,036 users, each with more than 700 ratings. In scenario 2 to 4, we consider general users and randomly sample 3,000 users from those with more than 200, 75 and 30 ratings, respectively. This results in a rating matrix with 3,036 users and 1,034 movies for scenario 1 and matrices with 3,000 users and 1,034 movies for scenarios 2 to 4. The observation probabilities in these rating matrices are approximately 0.50, 0.36, 0.23, and 0.12 for scenarios 1 to 4. Figure 5

TABLE 4

The RMSE in the test dataset for the proposed estimators, their corresponding baselines, and modified versions, evaluated across four scenarios of Netflix rating data.

	Estim1	Mod-Estim1	Base1	Mod-Base1	Estim2	Mod-Estim2	Base2	Mod-Base2
Scenario1	0.7680	0.7665	0.7775	0.7761	0.7713	0.7688	0.7722	0.7696
Scenario2	0.8075	0.8058	0.8118	0.8099	0.8119	0.8100	0.8148	0.8123
Scenario3	0.8383	0.8368	0.8421	0.8407	0.8467	0.8443	0.8504	0.8479
Scenario4	0.8490	0.8468	0.8505	0.8493	0.8670	0.8621	0.9053	0.8879

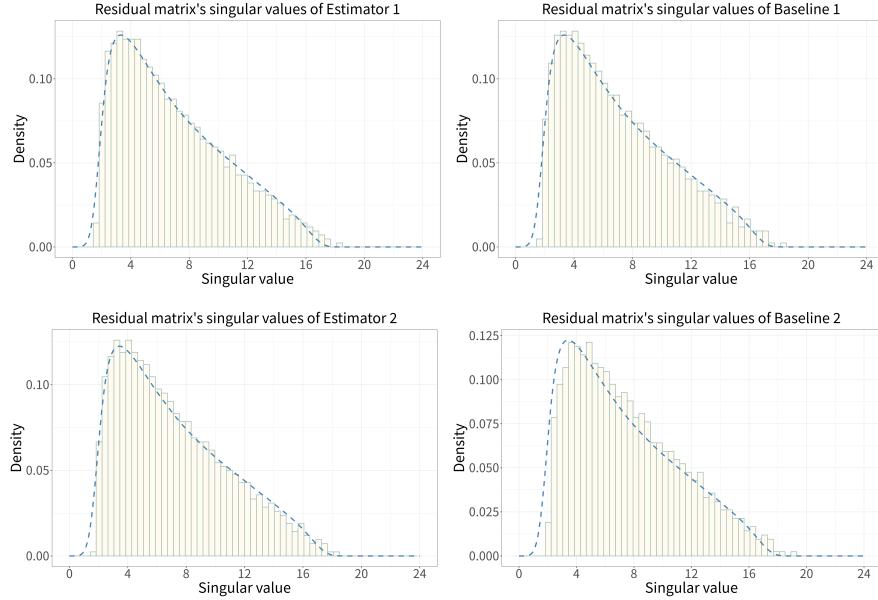


FIG 6. The distribution of singular values of residual matrix for proposed and baseline methods for scenario 4. The left two plots correspond to the results of Estimator 1 and Baseline 1, and the right two plots correspond to the results of Estimator 2 and Baseline 2.

shows the singular value distributions of the observed rating matrices for each scenario. In all cases, the distribution of singular values aligns closely with the expected distribution of the corresponding sparse random matrices after variance adjustment, with only a few outliers due to low-rank perturbations.

The proposed and baseline methods are implemented in the four scenarios as described earlier. The dataset is divided into three parts: training, validation, and testing, with a ratio of 4 : 1 : 1. Hyperparameters s in Estimator 1 and Baseline 1, and λ in Estimator 2 and Baseline 2, are tuned by minimizing the rooted mean squared error (RMSE) on the validation set. For scenarios 1 to 4, Estimator 1 selects ranks $s = 34, 14, 8, 7$ and Baseline 1 selects ranks $s = 34, 15, 8, 7$. Given that ratings fall within the range [1, 5], we truncate the estimated matrices to this interval, referring to them as modified versions. In scenarios 1 to 4, the noise level ratio R can be approximately estimated by our estimators and the selected ranks, to be $R = 0.58, 0.50, 0.48$ and 0.60 . This indicates that the Netflix data is subject to a relatively high noise level. Under high noise conditions ($R > 0.5$), our estimators consistently outperforms the baseline methods in the simulation study. For the Netflix rating data, Table 4 demonstrates that in all scenarios, the proposed estimators and their modified (Mod) versions yield smaller prediction errors on the test set, compared to the corresponding baselines. The empirical spectral distribution of the residual matrices from our estimators also closely aligns with the expected spectral distributions of the corresponding variance-adjusted sparse random matrices, showing a better fit than the baselines. Figure 6 displays the singular value

TABLE 5

The RMSE in the test dataset for the proposed estimators, their corresponding baselines, and modified versions, evaluated across four scenarios of Amazon review data.

	Estim1	Mod-Estim1	Base1	Mod-Base1	Estim2	Mod-Estim2	Base2	Mod-Base2
Scenario1	0.9264	0.9176	0.9364	0.9272	0.9251	0.9168	0.9457	0.9367
Scenario2	0.8441	0.8338	0.8686	0.8576	0.9239	0.8883	0.8656	0.8549
Scenario3	0.9912	0.9805	1.0020	0.9907	0.9912	0.9810	1.0180	1.0072
Scenario4	0.9787	0.9588	0.9898	0.9689	0.9831	0.9642	1.0172	0.9997

distributions of the residual matrices for both our estimators and the baselines in scenario 4 as an example.

5.3. Amazon reviews data. We also compare performance of our estimators with the baselines using the Amazon reviews 2023 dataset (Hou et al., 2024), which is publicly available at <https://amazon-reviews-2023.github.io>. The data collected user reviews in 2023 with rich features including the users rating, text and purchase verification. We consider four scenarios for reviews, each corresponding to a distinct category: movies, books, electronics and automotive. For each scenario, the top 1000 items in each category are considered, with 3000 users randomly selected after data preprocessing. Hence, Four 1000×3000 reviewing matrices are observed with observation probabilities 0.0128, 0.0127, 0.0110, 0.0068 for scenarios 1 to 4 respectively. The data preprocessing and implementation are similar to those for the Netflix rating data and we present the implementation details in the supplementary material. The noise level ratio R for the four scenarios is around $R = 1.16, 1.09, 1.54$ and 1.56 , which means that the dataset is under a higher noise level. Table 5 presents the RMSE on the test data for our estimators, their modified versions and the baselines. The performance gaps between our estimators (including modified versions) and the baselines are larger across all four scenarios.

6. Discussion. This paper introduces a novel spectral matching criterion for noisy matrix completion. We demonstrate that estimators based on this criterion perform well both theoretically and numerically. Efficient algorithms are devised for the proposed criterion, ensuring small Frobenius and maximum norms. Our methods provide advantages over commonly used least squared estimators by utilizing more information from noise, especially in high noise level situations.

Our method can be readily generalized to encompass broader missing mechanisms and more complex noise matrices if the variance structure of the noise is understood. However, expanding our method becomes challenging in scenarios with heterogeneous noise matrices, e.g., the entries of noise matrices have different and unknown variances. If a uniform upper bound for the variances of noise exists, a comparable error bound can still be derived, similar to the results in least squares estimators. However, this bound might be inadequate, especially when the upper bounds for variances are significantly larger than the normalized spectral norm of the noise matrix. Therefore, it is meaningful to develop methods that adapt the proposed approach to more complex noise scenarios in the future.

APPENDIX A: PROOF SKETCHES FOR MAIN THEOREMS

We provide intuitive sketches of technical proofs of key lemmas and theorems presented in Section 4 for ease of readers. Detailed proofs are available in the supplementary material.

Theorem 4.1 and its associated corollaries mainly rely on two pivotal Lemmas 4.1 and 4.2. The proof of Lemma 4.1 is straightforward. Define that the singular values of $P_\Omega(\mathbf{H})/\sqrt{m}\sigma$

as $\theta_1 \geq \dots \geq \theta_n$, the value of loss function can be bounded by

$$\begin{aligned}\mathcal{L}(\mathbf{M}_0; \mathbf{Y}, \Omega) &= p^{-1} \sum_{i=1}^n \omega_i \left(\frac{\sigma}{\widehat{\sigma}(\mathbf{M}_0)} \theta_i - \widehat{\lambda}_i \right)^2 \\ &\leq p^{-1} \sum_{i=1}^n \omega_i (\theta_i - \widehat{\lambda}_i)^2 + p^{-1} \sum_{i=1}^n \omega_i \left(\frac{\sigma}{\widehat{\sigma}(\mathbf{M}_0)} - 1 \right)^2 \theta_i^2 \\ &\leq p^{-1} \sup_{1 \leq i \leq n} (\theta_i - \widehat{\lambda}_i)^2 + p^{-1} \theta_1^2 \left(\frac{\sigma}{\widehat{\sigma}(\mathbf{M}_0)} - 1 \right)^2.\end{aligned}$$

Next, we bound the term $\sup_{1 \leq i \leq n} (\theta_i - \widehat{\lambda}_i)$ and θ_1 by establishing the universality for sparse random matrices. The bias $|\sigma/\widehat{\sigma}(\mathbf{M}_0) - 1|$ is controlled through singular value inequalities for low-rank perturbations.

The key argument in proof of Lemma 4.2 is to demonstrate that when $\|\mathbf{M} - \mathbf{M}_0\|_F$ is large, the leading singular values of $P_\Omega(\mathbf{M} - \mathbf{M}_0)$ are likewise large. Consequently, the spectral distribution of $P_\Omega(\mathbf{H}) + P_\Omega(\mathbf{M}_0 - \mathbf{M})$ deviates significantly from that of a sparse random matrix, resulting in a substantial increase in the loss function. To formally address this issue, we first demonstrate that a substantial magnitude of $|\mathbf{u}'(\mathbf{M} - \mathbf{M}_0)\mathbf{v}|$ for specific directions $\mathbf{u} \in \mathbb{R}^m, \mathbf{v} \in \mathbb{R}^n$ with sufficiently small $\|\mathbf{u}\|_\infty$ or $\|\mathbf{v}\|_\infty$ will induce a notable bias in the spectrum of the residual matrix $P_\Omega(\mathbf{H}) + P_\Omega(\mathbf{M}_0 - \mathbf{M})$, i.e.,

$$\|P_\Omega(\mathbf{H}) + P_\Omega(\mathbf{M}_0 - \mathbf{M})\| \geq |\mathbf{u}'P_\Omega(\mathbf{M} - \mathbf{M}_0)\mathbf{v}| - \|P_\Omega(\mathbf{H})\| \gg \|P_\Omega(\mathbf{H})\|.$$

In this step, the incoherence condition and the constraint of the observation probability p are required. Subsequently, we establish that for any $\mathbf{M} \in \mathbb{M}_{r,\mu_1}$ with a significant error $\|\mathbf{M} - \mathbf{M}_0\|_F$, there must exist certain directions, denoted by $\mathbf{u} \in \mathbb{R}^m, \mathbf{v} \in \mathbb{R}^n$, which lead to a considerable $|\mathbf{u}'(\mathbf{M} - \mathbf{M}_0)\mathbf{v}|$. This is intuitive because when those $|\mathbf{u}'(\mathbf{M} - \mathbf{M}_0)\mathbf{v}|$ for different directions \mathbf{u}, \mathbf{v} are all small, $\|\mathbf{M} - \mathbf{M}_0\|_F$ is also small, given that the rank of $\mathbf{M} - \mathbf{M}_0$ is low.

The proof of Theorem 4.2 appears more complex. Define the singular value decomposition of \mathbf{M}_0 and $\widetilde{\mathbf{M}}_\lambda$ as $\mathbf{U}\mathbf{D}\mathbf{V}'$ and $\mathbf{X}\Theta\mathbf{Y}'$ respectively, and singular values of $\mathbf{U}'(\mathbf{M}_0 - \widetilde{\mathbf{M}}_\lambda)\mathbf{V}$ as $\tau_1 \geq \dots \geq \tau_r$. The nuclear norm of $\widetilde{\mathbf{M}}_\lambda$ can be moderately controlled by $\lambda^{-1}\mathcal{L}(\mathbf{M}_0; \mathbf{Y}, \Omega)$ due to the definition of $\widetilde{\mathbf{M}}_\lambda$. Conversely, the lower bound of $\|\widetilde{\mathbf{M}}_\lambda\|_*$ can be controlled by the singular values τ_i . We partition $\widetilde{\mathbf{M}}_\lambda$ into two components: $\mathbf{M}_1 = \text{Proj}_r(\widetilde{\mathbf{M}}_\lambda)$, representing its projection onto a space of rank r , and $\mathbf{M}_2 = \widetilde{\mathbf{M}}_\lambda - \mathbf{M}_1$, denoting the residual component. Initially, we constrain τ_i for $1 \leq i \leq r$ by acknowledging that a high τ_i would substantially increase the value of the loss function. This approach mirrors the technique employed in the proof of Lemma 4.2. Then, $\|\mathbf{M}_0 - \mathbf{M}_1\|_F$ can be controlled when we also similarly investigate the singular values of $\mathbf{X}'_{[1:r]}(\mathbf{M}_0 - \widetilde{\mathbf{M}}_\lambda)\mathbf{Y}_{[1:r]}$, $\mathbf{X}'_{[1:r]}(\mathbf{M}_0 - \widetilde{\mathbf{M}}_\lambda)\mathbf{V}$ and $\mathbf{U}(\mathbf{M}_0 - \widetilde{\mathbf{M}}_\lambda)\mathbf{Y}_{[1:r]}$. Second, we establish an upper bound for the portion $(\mathbf{M}_0 - \widetilde{\mathbf{M}}_\lambda) - \mathbf{U}\mathbf{U}'(\mathbf{M}_0 - \widetilde{\mathbf{M}}_\lambda)\mathbf{V}\mathbf{V}'$ by combining the aforementioned bounds for the sum of singular values of $\widetilde{\mathbf{M}}_\lambda$ and utilizing the methodology pertaining to relating large singular values to large loss. This helps us to derive the bound for $\|\mathbf{M}_2\|_F$. Finally, we combine the bounds for $\|\mathbf{M} - \mathbf{M}_1\|_F$ and $\|\mathbf{M}_2\|_F$ to establish the theorem.

In section 4.3, we first establish the properties for the initial value $\mathbf{M}^{(0)}$ in Lemma 4.3. The spectral norm error of $p^{-1}P_\Omega(\mathbf{Y})$ can be derived from [Keshavan, Montanari and Oh \(2010\)](#). Then we treat $p^{-1}P_\Omega(\mathbf{Y})$ as the perturbation for \mathbf{M}_0 as follows,

$$p^{-1}P_\Omega(\mathbf{Y}) = \mathbf{M}_0 + (p^{-1}P_\Omega(\mathbf{Y}) - \mathbf{M}_0).$$

The Weyl-Wedin theorem assists in managing the Frobenius norm error for $\mathbf{M}^{(0)}$. We establish some new L_∞ matrix perturbation bounds and apply them here to both control $\|\mathbf{U}^{(0)} - \mathbf{U}\|_{2,\infty}$ and $\|\mathbf{V}^{(0)} - \mathbf{V}\|_{2,\infty}$, and subsequently $\|\mathbf{M}^{(0)} - \mathbf{M}_0\|_\infty$.

The proof of Theorem 4.3 is based on induction and consists of three main parts. First, we verify the efficiency of the pseudo-gradient utilized in our algorithm. Consider the perturbation formula,

$$p^{-1}P_\Omega(\mathbf{M}^{(k)} - \mathbf{M}_0 - \mathbf{H}) = (\mathbf{M}^{(k)} - \mathbf{M}_0) + \left[p^{-1}P_\Omega(\mathbf{M}^{(k)} - \mathbf{M}_0 - \mathbf{H}) - (\mathbf{M}^{(k)} - \mathbf{M}_0) \right].$$

By bounding the spectral norm of $p^{-1}P_\Omega(\mathbf{M}^{(k)} - \mathbf{M}_0 - \mathbf{H}) - (\mathbf{M}^{(k)} - \mathbf{M}_0)$ and applying the established L_∞ matrix perturbation bounds, we can control the L_2/L_∞ norm for the singular vectors of $P_\Omega(\mathbf{M}^{(k)} - \mathbf{M}_0 - \mathbf{H})$ and their distances to the singular vectors of $\mathbf{M}^{(k)} - \mathbf{M}_0$. This allows us to achieve a sufficiently rapid reduction in the Frobenius norm error from $\|\mathbf{M}^{(k)} - \mathbf{M}_0\|_F$ to $\|\mathbf{N}^{(k+1)} - \mathbf{M}_0\|_F$. Second, we demonstrate that the low rank projection $\mathbf{M}^{(k+1)} = \text{Proj}_r(\mathbf{N}^{(k+1)})$ has a minimal impact. This is because the leading singular space of the pseudo-gradient does not significantly deviate from the singular space of $\mathbf{M}^{(k)}$, resulting in only a minor truncation by the projection. Finally, we argue that the incoherent coefficients will not exhibit rapid increase from $\mathbf{M}^{(k)}$ to $\mathbf{M}^{(k+1)}$. Consider the perturbation formula given by

$$\begin{aligned} \mathbf{N}^{(k+1)} = & \left(\mathbf{N}^{(k+1)} - \eta \mathcal{P}_{\mathbf{U}^{(k)}, \perp} \hat{\nabla} \mathcal{L}(\mathbf{M}^{(k)}; \mathbf{Y}, \Omega) \mathcal{P}_{\mathbf{V}^{(k)}, \perp'} \right) \\ & + \eta \mathcal{P}_{\mathbf{U}^{(k)}, \perp} \hat{\nabla} \mathcal{L}(\mathbf{M}^{(k)}; \mathbf{Y}, \Omega) \mathcal{P}_{\mathbf{V}^{(k)}, \perp'}. \end{aligned}$$

The row and column norm of the perturbation term can be controlled. We subsequently employ the findings from Fan, Wang and Zhong (2018) to verify the incoherent condition. Then we can bound the maximum norm of $\mathbf{M}^{(k+1)} - \mathbf{M}_0$, and the induction for the Frobenius norm, L_2/L_∞ norm, and maximum norm is completed.

APPENDIX B: ALGORITHMS FOR PROPOSED ESTIMATORS

We provide the algorithm for the proposed estimator $\tilde{\mathbf{M}}_{\text{nuc}}$ in (6). The algorithm is designed for analyzing algorithm iterations and implementing in both simulated and real-world data.

Algorithm 2 Iteration Algorithm for $\tilde{\mathbf{M}}_{\text{nuc}}$

Input: Observed data $\mathbf{Y} \in \mathbb{R}^{m \times n}$, missing index matrix $\Omega \in \mathbb{R}^{m \times n}$ and tuning parameter λ .

Output: Completion matrix \mathbf{M} .

- 1: $\mathbf{U}^{(0)}, \mathbf{D}^{(0)}, \mathbf{V}^{(0)} \leftarrow \text{svd}(p^{-1}\mathbf{Y})$;
 - 2: $\mathbf{M}^{(0)} \leftarrow \mathbf{U}^{(0)}(\mathbf{D}^{(0)} - \lambda/2)_+ \mathbf{V}^{(0)'};$
 - 3: Sample l standard Gaussian random matrix $\mathbf{M}_i \in \mathbb{R}^{m \times n}$ with singular value vector $\boldsymbol{\lambda}(\mathbf{M}_i), 1 \leq i \leq l$;
 - 4: $\hat{\boldsymbol{\lambda}} \leftarrow \sum_{i=1}^l \boldsymbol{\lambda}(\mathbf{M}_i)/l$, part_sum $\leftarrow \sum_{i=n/3}^{2n/3} \hat{\lambda}_i$;
 - 5: **while** $k \leq K$ **do**
 - 6: $\mathbf{U}^{(k+1)}, \mathbf{D}^{(k+1)}, \mathbf{V}^{(k+1)} \leftarrow \text{svd}(\mathbf{Y} - P_\Omega(\mathbf{M}^{(k)}))$;
 - 7: $\hat{\sigma}^{(k+1)} \leftarrow \sum_{i=n/3}^{2n/3} \mathbf{D}_{i,i}^{(k+1)} / \text{part_sum}$;
 - 8: $\hat{\nabla}_{\mathbf{M}}^{(k+1)} \leftarrow \sum_{i=1}^n w_i(\mathbf{D}_{i,i}^{(k+1)} - \hat{\sigma}^{(k+1)} \hat{\lambda}_i) \mathbf{U}_{[i]}^{(k+1)} \mathbf{V}_{[i]}^{(k+1)'}$;
 - 9: $\mathbf{N}^{(k+1)} \leftarrow \mathbf{M}^{(k)} - \eta_{k+1} P_\Omega(\hat{\nabla}_{\mathbf{M}}^{(k+1)})$;
 - 10: $\mathbf{X}^{(k+1)}, \mathbf{\Lambda}^{(k+1)}, \mathbf{Z}^{(k+1)} \leftarrow \text{svd}(\mathbf{N}^{(k+1)})$;
 - 11: $\mathbf{M}^{(k+1)} \leftarrow \mathbf{X}^{(k+1)}(\mathbf{\Lambda}^{(k+1)} - \lambda/2)_+ \mathbf{Z}^{(k+1)'}$;
 - 12: **end while**
 - 13: **return** $\mathbf{M} = \mathbf{M}^{(K+1)}$.
-

Acknowledgments. Fang Yao is the corresponding author. This research is partially supported by the National Key R&D Program of China (No. 2022YFA1003800), the National Natural Science Foundation of China (No. 12292981, 11931001, 123B2009), the LMAM, the Fundamental Research Funds for the Central Universities and the LMEQF.

SUPPLEMENTARY MATERIAL

The Supplementary Material “SUPPLEMENT TO ‘MATRIX COMPLETION VIA RESIDUAL SPECTRAL MATCHING’ ” contains technical lemmas and detailed proofs for all the presented lemmas, theorems and corollaries. Additional numerical results are also provided.

REFERENCES

- ALT, J., ERDOS, L. and KRÜGER, T. (2017). Local law for random Gram matrices. *Electron. J. Probab* **22** 1–41.
- ATHEY, S., BAYATI, M., DOUDCHENKO, N., IMBENS, G. and KHOSRAVI, K. (2021). Matrix completion methods for causal panel data models. *Journal of the American Statistical Association* **116** 1716–1730.
- BAI, Z., HU, J. and ZHOU, W. (2012). Convergence rates to the Marchenko–Pastur type distribution. *Stochastic Processes and their Applications* **122** 68–92.
- BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral analysis of large dimensional random matrices* **20**. Springer.
- BAI, Z.-D. and YIN, Y.-Q. (1988). Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix. *The Annals of Probability* 1729–1741.
- BAO, Z., DING, X. and WANG, K. (2021). Singular vector and singular subspace distribution for the matrix denoising model. *The Annals of Statistics* **49** 370–392.
- BAO, Z., ERDŐS, L. and SCHNELLI, K. (2017). Local law of addition of random matrices on optimal scale. *Communications in Mathematical Physics* **349** 947–990.
- BAO, Z., DING, X., WANG, J. and WANG, K. (2022). Statistical inference for principal components of spiked covariance matrices. *The Annals of Statistics* **50** 1144–1169.
- BENAYCH-GEORGES, F. and NADAKUDITI, R. R. (2012). The singular values and vectors of low rank perturbations of large rectangular random matrices. *Journal of Multivariate Analysis* **111** 120–135.
- BENNETT, J., LANNING, S. et al. (2007). The netflix prize. In *Proceedings of KDD cup and workshop* **2007** 35. New York, NY, USA.
- BHATTACHARYA, S. and CHATTERJEE, S. (2022). Matrix completion with data-dependent missingness probabilities. *IEEE Transactions on Information Theory* **68** 6762–6773.
- BI, X., QU, A., WANG, J. and SHEN, X. (2017). A group-specific recommender system. *Journal of the American Statistical Association* **112** 1344–1353.
- BOURGADE, P., YAU, H.-T. and YIN, J. (2014). Local circular law for random matrices. *Probability Theory and Related Fields* **159** 545–595.
- BRAILOVSKAYA, T. and VAN HANDEL, R. (2024). Universality and sharp matrix concentration inequalities. *Geometric and Functional Analysis* 1–105.
- CAI, T. T. and ZHANG, A. (2015). ROP: Matrix recovery via rank-one projections. *The Annals of Statistics* 102–138.
- CAI, T. T. and ZHOU, W.-X. (2016). Matrix completion via max-norm constrained optimization. *Electronic Journal of Statistics* **10** 1493–1525.
- CANDES, E. J. and PLAN, Y. (2010). Matrix completion with noise. *Proceedings of the IEEE* **98** 925–936.
- CANDÈS, E. J. and RECHT, B. (2009). Exact Matrix Completion via Convex Optimization. *Found Comput Math* **9** 717–772.
- CANDÈS, E. J. and TAO, T. (2010). The power of convex relaxation: Near-optimal matrix completion. *IEEE transactions on information theory* **56** 2053–2080.
- CHEN, J., LIU, D. and LI, X. (2020). Nonconvex Rectangular Matrix Completion via Gradient Descent Without $l_{2,\infty}$ Regularization. *IEEE Transactions on Information Theory* **66** 5806–5841.
- CHEN, Z., YANG, Y. and YAO, F. (2024). Dynamic Matrix Recovery. *Journal of the American Statistical Association* 1–12.
- CHEN, Y., FAN, J., MA, C. and YAN, Y. (2019). Inference and uncertainty quantification for noisy matrix completion. *Proceedings of the National Academy of Sciences* **116** 22931–22937.
- CHEN, Y., CHI, Y., FAN, J., MA, C. and YAN, Y. (2020). Noisy matrix completion: Understanding statistical guarantees for convex relaxation via nonconvex optimization. *SIAM journal on optimization* **30** 3098–3121.

- CHOI, J. and YUAN, M. (2023). Matrix Completion When Missing Is Not at Random and Its Applications in Causal Panel Data Models. *arXiv preprint arXiv:2308.02364*.
- DING, X. and YANG, F. (2018). A necessary and sufficient condition for edge universality at the largest singular values of covariance matrices. *The Annals of Applied Probability* **28** 1679–1738.
- DUAN, C., LI, J. and XIA, D. (2024). Online Policy Learning and Inference by Matrix Completion. *arXiv preprint arXiv:2404.17398*.
- ELHAMIFAR, E. (2016). High-rank matrix completion and clustering under self-expressive models. *Advances in Neural Information Processing Systems* **29**.
- ERDŐS, L., YAU, H.-T. and YIN, J. (2012). Rigidity of eigenvalues of generalized Wigner matrices. *Advances in Mathematics* **229** 1435–1515.
- ERDOS, L., YAU, H.-T. and YIN, J. (2010). Bulk universality for generalized Wigner matrices. *arXiv preprint arXiv:1001.3453*.
- ERIKSSON, B., BALZANO, L. and NOWAK, R. (2012). High-rank matrix completion. In *Artificial Intelligence and Statistics* 373–381. PMLR.
- FAN, J. and CHENG, J. (2018). Matrix completion by deep matrix factorization. *Neural Networks* **98** 34–41.
- FAN, J. and CHOW, T. (2017). Deep learning based matrix completion. *Neurocomputing* **266** 540–549.
- FAN, J., WANG, W. and ZHONG, Y. (2018). An l_∞ eigenvector perturbation bound and its application to robust covariance estimation. *Journal of Machine Learning Research* **18** 1–42.
- GE, R., LEE, J. D. and MA, T. (2016). Matrix completion has no spurious local minimum. *Advances in neural information processing systems* **29**.
- GÖTZE, F. and TIKHOMIROV, A. (2004). Rate of convergence in probability to the Marchenko-Pastur law. *Bernoulli* **10** 503–548.
- GUI, Y., BARBER, R. and MA, C. (2023). Conformalized matrix completion. *Advances in Neural Information Processing Systems* **36** 4820–4844.
- HARDT, M. (2014). Understanding alternating minimization for matrix completion. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science* 651–660. IEEE.
- HE, Y. (2023). Edge universality of sparse Erdos-Renyi digraphs. *arXiv preprint arXiv:2304.04723*.
- HOU, Y., LI, J., HE, Z., YAN, A., CHEN, X. and MC CAULEY, J. (2024). Bridging language and items for retrieval and recommendation. *arXiv preprint arXiv:2403.03952*.
- IWEZE, D. (2020). A Matrix Completion Approach to Policy Evaluation: Evaluating the Impact of the VCT Scheme on Investment in the UK. Available at SSRN 4561908.
- JAIN, P., NETRAPALLI, P. and SANGHAVI, S. (2013). Low-rank matrix completion using alternating minimization. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing* 665–674.
- JANNACH, D., RESNICK, P., TUZHILIN, A. and ZANKER, M. (2016). Recommender systems—beyond matrix completion. *Communications of the ACM* **59** 94–102.
- JI, H., HUANG, S., SHEN, Z. and XU, Y. (2011). Robust video restoration by joint sparse and low rank matrix approximation. *SIAM Journal on Imaging Sciences* **4** 1122–1142.
- JIA, Z., JIN, Q., NG, M. K. and ZHAO, X.-L. (2022). Non-local robust quaternion matrix completion for large-scale color image and video inpainting. *IEEE Transactions on Image Processing* **31** 3868–3883.
- JIN, H., MA, Y. and JIANG, F. (2022). Matrix completion with covariate information and informative missingness. *Journal of Machine Learning Research* **23** 1–62.
- KAFETZOPOULOS, A. and MALTSEV, A. (2023). Local Marchenko–Pastur law at the hard edge of the sample covariance ensemble. *Journal of Mathematical Physics* **64**.
- KALLUS, N., MAO, X. and UDELL, M. (2018). Causal inference with noisy and missing covariates via matrix factorization. *Advances in neural information processing systems* **31**.
- KESHAVAN, R., MONTANARI, A. and OH, S. (2009). Matrix completion from noisy entries. *Advances in neural information processing systems* **22**.
- KESHAVAN, R. H., MONTANARI, A. and OH, S. (2010). Matrix completion from a few entries. *IEEE transactions on information theory* **56** 2980–2998.
- KLOPP, O. (2014). Noisy low-rank matrix completion with general sampling distribution. *Bernoulli* 282–303.
- KOLTCHINSKII, V., LOUNICI, K. and TSYBAKOV, A. B. (2011). Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *The Annals of Statistics* **39** 2302–2329.
- KORTAS, M., HABACHI, O., BOUALLEGUE, A., MEGHDADI, V., EZZEDINE, T. and CANCES, J.-P. (2021). Robust data recovery in wireless sensor network: a learning-based matrix completion framework. *Sensors* **21** 1016.
- LEE, J. (2021). Higher order fluctuations of extremal eigenvalues of sparse random matrices. *arXiv preprint arXiv:2108.11634*.
- LI, X. P., HUANG, L., SO, H. C. and ZHAO, B. (2019). A survey on matrix completion: Perspective of signal processing. *arXiv preprint arXiv:1901.10885*.

- LI, Z., NIE, F., WANG, R. and LI, X. (2023). Robust rank-one matrix completion with rank estimation. *Pattern Recognition* **142** 109637.
- LUO, Y. and ZHANG, A. R. (2023). Low-rank tensor estimation via riemannian gauss-newton: Statistical optimality and second-order convergence. *The Journal of Machine Learning Research* **24** 18274–18321.
- MA, C., WANG, K., CHI, Y. and CHEN, Y. (2018). Implicit regularization in nonconvex statistical estimation: Gradient descent converges linearly for phase retrieval and matrix completion. In *International Conference on Machine Learning* 3345–3354. PMLR.
- MA, S., NIU, P.-Y., ZHANG, Y. and ZHU, Y. (2024). Statistical Inference For Noisy Matrix Completion Incorporating Auxiliary Information. *Journal of the American Statistical Association* **just-accepted** 1–24.
- MAO, X., CHEN, S. X. and WONG, R. K. (2019). Matrix completion with covariate information. *Journal of the American Statistical Association* **114** 198–210.
- NEGAHBAN, S. and WAINWRIGHT, M. J. (2011). Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *The Annals of Statistics* **39** 1069–1097.
- NEGAHBAN, S. and WAINWRIGHT, M. J. (2012). Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. *The Journal of Machine Learning Research* **13** 1665–1697.
- ONGIE, G., WILLETT, R., NOWAK, R. D. and BALZANO, L. (2017). Algebraic variety models for high-rank matrix completion. In *International Conference on Machine Learning* 2691–2700. PMLR.
- PILLAI, N. S. and YIN, J. (2014). Universality of covariance matrices. *The Annals of Applied Probability* 935–1001.
- RAMLATCHAN, A., YANG, M., LIU, Q., LI, M., WANG, J. and LI, Y. (2018). A survey of matrix completion methods for recommendation systems. *Big Data Mining and Analytics* **1** 308–323.
- ROHDE, A. and TSYBAKOV, A. B. (2011). Estimation of high-dimensional low-rank matrices. *The Annals of Statistics* **39**.
- SUN, R. and LUO, Z.-Q. (2016). Guaranteed matrix completion via non-convex factorization. *IEEE Transactions on Information Theory* **62** 6535–6579.
- TAO, T. and VU, V. (2012). Random covariance matrices: Universality of local statistics of eigenvalues. *The Annals of Probability* 1285–1315.
- TIKHOMIROV, A. (2011). On the Rate of Convergence to the Marchenko–Pastur Distribution. *arXiv preprint arXiv:1110.1284*.
- TIKHOMIROV, A. and TIMUSHEV, D. (2022). Local laws for sparse sample covariance matrices without the truncation condition. *arXiv preprint arXiv:2209.13207*.
- VERAART, J., NOVIKOV, D. S., CHRISTIAENS, D., ADES-ARON, B., SIJBERS, J. and FIEREMANS, E. (2016). Denoising of diffusion MRI using random matrix theory. *Neuroimage* **142** 394–406.
- WENG, Z. and WANG, X. (2012). Low-rank matrix completion for array signal processing. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 2697–2700. IEEE.
- XIA, D. and YUAN, M. (2021). Statistical Inferences of Linear Forms for Noisy Matrix Completion. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*.
- XIE, K., NING, X., WANG, X., XIE, D., CAO, J., XIE, G. and WEN, J. (2016). Recover corrupted data in sensor networks: A matrix completion solution. *IEEE Transactions on Mobile Computing* **16** 1434–1448.
- XU, L. and DAVENPORT, M. (2016). Dynamic matrix recovery from incomplete observations under an exact low-rank constraint. *Advances in Neural Information Processing Systems* **29**.
- YIN, Y. Q. and KRISHNAIAH, P. R. (1986). Limit theorem for the eigenvalues of the sample covariance matrix when the underlying distribution is isotropic. *Theory of Probability & Its Applications* **30** 861–867.
- ZHANG, R., SHAPIRO, A. and XIE, Y. (2019). Statistical rank selection for incomplete low-rank matrices. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 2912–2916. IEEE.
- ZHANG, A. R., LUO, Y., RASKUTTI, G. and YUAN, M. (2020). ISLET: Fast and optimal low-rank tensor regression via importance sketching. *SIAM journal on mathematics of data science* **2** 444–479.
- ZHENG, Q. and LAFFERTY, J. (2016). Convergence analysis for rectangular matrix completion using Burer-Monteiro factorization and gradient descent. *arXiv preprint arXiv:1605.07051*.