A Rank-Corrected Procedure for Matrix Completion with Fixed Basis Coefficients

Weimin Miao*, Shaohua Pan† and Defeng Sun‡

November 9, 2012

Abstract

In this paper, we address low-rank matrix completion problems with fixed basis coefficients, which include the low-rank correlation matrix completion in various fields such as the financial market and the low-rank density matrix completion from the quantum state tomography. For this class of problems, the efficiency of the common nuclear norm penalized estimator for recovery may be challenged. Here, with a reasonable initial estimator, we propose a rank-corrected procedure to generate an estimator of high accuracy and low rank. For this new estimator, we establish a non-asymptotic recovery error bound and analyze the impact of adding the rank-correction term on improving the recoverability. We also provide necessary and sufficient conditions for rank consistency in the sense of Bach [3], in which the concept of constraint nondegeneracy in matrix optimization plays an important role. As a byproduct, our results provide a theoretical foundation for the majorized penalty method of Gao and Sun [25] and Gao [24] for structured low-rank matrix optimization problems.

Keywords: matrix completion, fixed basis coefficients, low-rank, convex optimization, rank consistency, constraint nondegeneracy.

1 Introduction

The low-rank matrix completion is referred to recover an unknown low-rank matrix, exactly or approximately, from the under-sampled observations with or without noises. This problem is of considerable interest in many application areas, from machine learning

*Department of Mathematics, National University of Singapore, 10 Lower Kent Ridge Road, Singapore 119076 (miaoweimin@nus.edu.sg).

†Department of Mathematics, South China University of Technology, Tianhe District of Guangzhou City, China (shhpan@scut.edu.cn).

‡Department of Mathematics and Risk Management Institute, National University of Singapore, 10 Lower Kent Ridge Road, Singapore 119076 (matsundf@nus.edu.sg). This author's research is supported in part by Academic Research Fund under grant R-146-000-149-112.

to quantum state tomography. A basic idea to address a low-rank matrix completion problem is to minimize the rank of a matrix subject to certain constraints involving observations. Given that the direct minimization of rank function is generally NP-hard, a widely-used convex relaxation approach is to replace the rank function with the nuclear norm as the latter is the convex envelope of the rank function over a unit ball of the spectral norm [19].

Nuclear norm minimization (NNM) has been observed to provide a low-rank solution in practice for a long time (see, e.g., [52, 51, 19]). The first theoretical characterization for the minimum rank solution of the NNM was given by Recht, Fazel and Parrilo [61], with the help of the concept of Restricted Isometric Property (RIP). Recognizing that the matrix completion problem does not obey the RIP, Candès and Recht [8] introduced the concept of incoherence property and proved that most low-rank matrices can be exactly recovered from a surprisingly small number of noiseless observations of randomly sampled entries via the NNM. The bound of the number of sampled entries was later improved to be near-optimal by Candés and Tao [9] through a counting argument. Such a bound was also obtained by Keshavan et al. [35] for their proposed OptSpace algorithm. Later, Gross [28] sharpened the bound by employing a novel technique from quantum information theory developed in [29], with the extension of noiseless observations of entries to coefficients relative to any basis. This technique was also adapted by Recht [60]. All the above results focus on noiseless matrix completion. The matrix completion with noise was first addressed by Candés and Plan [7]. More recently, nuclear norm penalized estimators for matrix completion with noise have been well studied by Koltchinskii, Lounici and Tsybakov [42], Negahban and Wainwright [55], and Klopp [38]. Besides the nuclear norm, several other penalties for matrix completion have also been studied in [65, 37, 41, 67, 23].

The NNM has been demonstrated to be a successful approach to encourage a low-rank solution in many situations. However, the efficiency of the NNM may be challenged under general sampling schemes. For example, the conditions characterized by Bach [3] for rank consistency of the nuclear norm penalized least squares estimator may not be satisfied. In particular, for matrix completion problems, Salakhutdinov and Srebro [66] showed that when certain rows and/or columns are sampled with high probability, the NNM may fail in the sense that the number of observations required for recovery is much more than the setting of most matrix completion problems. Negahban and Wainwright [55] also pointed out the impact of such heavy sampling schemes on the recovery error bound. As a remedy for this, a weighted nuclear norm (trace norm), based on row- and column-marginals of the sampling distribution, was suggested in [55, 66, 22] if the prior information on sampling distribution is available.

When the true matrix possesses a symmetric/Hermitian positive semidefinite structure, the impact of general sampling schemes on the recoverability of the NNM is more remarkable. In this situation, the nuclear norm reduces to the trace and thus only depends on diagonal entries rather than all entries as the rank function does. As a result, if diagonal entries are heavily sampled, the ability of the NNM to promote a low-rank

solution, as well as the recoverability, will be highly weakened. This phenomenon is fully reflected in the widely-used correlation matrix completion problem, for which the nuclear norm becomes a constant and completely loses effectiveness for matrix recovery. Another example of particular interest in quantum state tomography is to recover a density matrix of a quantum system from Pauli measurements (see, e.g., [29, 21, 71]). A density matrix is a Hermitian positive semidefinite matrix of trace one. Obviously, if the constraints of positive semidefiniteness and trace one are simultaneously imposed on the NNM, the nuclear norm completely fails in promoting a low-rank solution. Thus, one of the two constraints has to be abandoned in the NNM and then be restored in the post-processing stage. In fact, this idea has been explored in [29, 21] and the numerical results there indicated its relative efficiency though it is at best sub-optimal.

In this paper, with a strong motivation to optimally address the difficulties in correlation and density matrix completion problems, we propose a low-rank matrix completion model with fixed basis coefficients. In our setting, for any given basis of the matrix space, a few basis coefficients of the true matrix are assumed to be fixed due to a certain structure or some prior information, and the rest are allowed to be observed with noises under general sampling schemes. Certainly, one can apply the nuclear norm penalized technique to our model. The challenge is that, as argued earlier, this may not yield a desired low-rank solution with small estimation errors. Here, we introduce a rank-correction step to address this critical issue provided that a reasonable initial estimator is available. A satisfactory choice of the initial estimator is the nuclear norm penalized estimator or one of its analogies. The rank-correction step solves a convex "nuclear norm − rank-correction term + proximal term" regularized least squares problem with fixed basis coefficients (and the possible positive semidefinite constraint). The rank-correction term is a linear term constructed from the initial estimator, and the proximal term is a quadratic term added to ensure the boundness of the solution to the convex problem. The resulting convex matrix optimization problem can be solved by the efficient algorithms recently developed in [32, 33, 34] even for large-scale cases.

The idea of using a two-stage or even multi-stage procedure is not brand new for dealing with sparse recovery in the statistical and machine learning literature. The l_1-norm penalized least squares method, also known as the Lasso [68], is very attractive and popular for variable selection in statistics, thanks to the invention of the fast and efficient LARS algorithm [12]. On the other hand, the l_1-norm penalty has long been known by statisticians to yield biased estimators and cannot attain the estimation optimality [14, 18]. The issue of bias can be overcome by nonconvex penalization methods, see, e.g., [44, 13, 74]. A multi-stage procedure naturally occurs if the nonconvex problem obtained is solved by an iterative algorithm [78]. In particular, once a good initial estimator is used, a two-stage estimator is enough to achieve the desired asymptotic efficiency, e.g., the adaptive Lasso proposed by Zou [77]. There are also a number of important papers in this line on variable selection, including [44, 50, 75, 31, 76, 49, 15], to name only a few. For a broad overview, the interested readers are referred to the recent survey papers [16, 17]. It is natural to extend the ideas from the vector case to the matrix case. Recently, Bach [3] made an important step in extending the adaptive Lasso of Zou

[77] to the matrix case for seeking rank consistency under general sampling schemes. However, it is not clear how to apply Bach's idea to our matrix completion model with fixed basis coefficients since the required rate of convergence of the initial estimator for achieving asymptotic properties is no longer valid as far as we can see. More critically, there are numerical difficulties in efficiently solving the resulting optimization problems. Such difficulties also occur when the reweighted nuclear norm proposed by Mohan and Fazel [54] is applied to the rectangular matrix completion problems.

The rank-correction step to be proposed in this paper is for the purpose to overcome the above difficulties. This approach is inspired by the majorized penalty method recently proposed by Gao and Sun [25] for solving structured matrix optimization problems with a low-rank constraint. For our proposed rank-correction step, we provide a non-asymptotic recovery error bound in the Frobenius norm, following a similar argument adopted by Klopp in [38]. The obtained error bound indicates that adding the rank-correction term could help to substantially improve the recoverability. As the estimator is expected to be of low-rank, we also study the asymptotic property — rank consistency in the sense of Bach [3], under the setting that the matrix size is assumed to be fixed. This setting may not be ideal for analyzing asymptotic properties for matrix completion, but it does allow us to take the crucial first step to gain insights into the limitation of the nuclear norm penalization. Among others, the concept of constraint nondegeneracy for conic optimization problem plays a key role in our analysis. Interestingly, our results of recovery error bound and rank consistency suggest a consistent criterion for constructing a suitable rank-correction function. In particular, for the correlation and density matrix completion problems, we prove that the rank consistency automatically holds for a broad selection of rank-correction functions. To achieve better performance for recovery, the rank-correction step may be iteratively used for several times, especially when the sample ratio is relatively low. Finally, we remark that our results can also be used to provide a theoretical foundation for the majorized penalty method of Gao and Sun [25] and Gao [24] for structured low-rank matrix optimization problems.

This paper is organized as follows. In Section 2, we introduce the observation model of matrix completion with fixed basis coefficients and the formulation of the rank-correction step. In Section 3, we establish a non-asymptotic recovery error bound and discuss the impact of the rank-correction term on recovery. Section 4 provides necessary and sufficient conditions for rank consistency. Section 5 is devoted to the construction of the rank-correction function. In Section 6, we report numerical results to validate the efficiency of our proposed rank-corrected procedure. We conclude this paper in Section 7. All proofs are left in the Appendix.

Notation. Here we provide a brief summary of the notation used in this paper.

- Let $\mathbb{R}^{n_1\times n_2}$ and $\mathbb{C}^{n_1\times n_2}$ denote the space of all $n_1\times n_2$ real and complex matrices, respectively. Let $\mathcal{S}^n(\mathcal{S}^n_+, \mathcal{S}^n_{++})$ denote the set of all $n\times n$ real symmetric (positive semidefinite, positive definite) matrices and $\mathcal{H}^n(\mathcal{H}^n_+, \mathcal{H}^n_{++})$ denote the set of all $n\times n$ Hermitian (positive semidefinite, positive definite) matrices. Let $\mathbb{S}^n$ ($\mathbb{S}^n_+$, $\mathbb{S}^n_{++}$) represent $\mathcal{S}^n$ ($\mathcal{S}^n_+$, $\mathcal{S}^n_{++}$) for the real case and $\mathcal{H}^n$ ($\mathcal{H}^n_+$, $\mathcal{H}^n_{++}$) for the complex case.

- Let $\mathbb{V}^{n_1\times n_2}$ represent $\mathbb{R}^{n_1\times n_2}$, $\mathbb{C}^{n_1\times n_2}$, $\mathcal{S}^n$ or $\mathcal{H}^n$. We define $n := \min(n_1, n_2)$ for the previous two cases and stipulate $n_1 = n_2 = n$ for the latter two cases. Let $\mathbb{V}^{n_1\times n_2}$ be endowed with the trace inner product $\langle\cdot,\cdot\rangle$ and its induced norm $\|\cdot\|_F$, i.e., $\langle X, Y\rangle := \mathrm{Re}\big(\mathrm{Tr}(X^{\mathbb{T}}Y)\big)$ for $X, Y \in \mathbb{V}^{n_1\times n_2}$, where "Tr" stands for the trace of a matrix and "Re" means the real part of a complex number.

- For the real case, $\mathbb{O}^{n\times k}$ denotes the set of all $n \times k$ real matrices with orthonormal columns, and for the complex case, $\mathbb{O}^{n\times k}$ denotes the set of all $n \times k$ complex matrices with orthonormal columns. When $k = n$, we write $\mathbb{O}^{n\times k}$ as $\mathbb{O}^n$ for short.

- The notation $^{\mathbb{T}}$ denotes the transpose for the real case and the conjugate transpose for the complex case. The notation * means the adjoint of operator.

- For any given vector x, $\mathrm{Diag}(x)$ denotes a rectangular diagonal matrix of suitable size with the i-th diagonal entry being x_i.

- For any $x \in \mathbb{R}^n$, let $\|x\|_2$ and $\|x\|_\infty$ denote the Euclidean norm and the maximum norm, respectively. For any $X \in \mathbb{V}^{n_1\times n_2}$, let $\|X\|$ and $\|X\|_*$ denote the spectral norm and the nuclear norm, respectively.

- The notations $\overset{a.s.}{\to}$, $\overset{p}{\to}$ and $\overset{d}{\to}$ mean almost sure convergence, convergence in probability and convergence in distribution, respectively. We write $x_m = O_p(1)$ if x_m is bounded in probability.

- For any set K, let $|K|$ denote the cardinality of K and let $\delta_K(x)$ denote the indicator function of K, i.e., $\delta_K(x) = 0$ if $x \in K$, and $\delta_K(x) = +\infty$ otherwise. Let I_n denote the $n \times n$ identity matrix.

2 Problem formulation

In this section, we formulate the model of the matrix completion problem with fixed basis coefficients, and then propose a rank-correction step for solving this class of problems.

2.1 The observation model

Let $\{\Theta_1, \ldots, \Theta_d\}$ be a given orthonormal basis of the given real inner product space $\mathbb{V}^{n_1\times n_2}$. Then, any matrix $X \in \mathbb{V}^{n_1\times n_2}$ can be uniquely expressed in the form of $X = \sum_{k=1}^d \langle\Theta_k, X\rangle\Theta_k$, where $\langle\Theta_k, X\rangle$ is called the basis coefficient of X relative to Θ_k. Let $\overline{X} \in \mathbb{V}^{n_1\times n_2}$ be the unknown low-rank matrix to be recovered. In some practical applications, for example, the correlation and density matrix completion, a few basis coefficients of the unknown matrix $\overline{X}$ are fixed (or assumed to be fixed) due to a certain structure or reliable prior information. Throughout this paper, we let $\alpha \subseteq \{1, 2, \ldots, d\}$ denote the set of the indices relative to which the basis coefficients are fixed, and β denote the complement of α in $\{1, 2, \ldots, d\}$, i.e., $\alpha \cap \beta = \emptyset$ and $\alpha \cup \beta = \{1, \ldots, d\}$. We define $d_1 := |\alpha|$ and $d_2 := |\beta|$.

When a few basis coefficients are fixed, one only needs to observe the rest for recovering the unknown matrix $\overline{X}$. Assume that we are given a collection of m noisy observations of the basis coefficients relative to $\{\Theta_k : k \in \beta\}$ in the following form

$$y_i = \left\langle \Theta_{\omega_i}, \overline{X} \right\rangle + \nu \xi_i, \quad i = 1, \ldots, m, \tag{1}$$

where ω_i are the indices randomly sampled from the index set β, ξ_i are the independent and identically distributed (i.i.d.) noises with $\mathbb{E}(\xi_i) = 0$ and $\mathbb{E}(\xi_i^2) = 1$, and $\nu > 0$ controls the magnitude of noise. Unless otherwise stated, we assume a general weighted sampling (with replacement) scheme with the sampling distributions of ω_i as follows.

Assumption 1 *The indices $\omega_1, \ldots, \omega_m$ are i.i.d. copies of a random variable ω that has a probability distribution Π over $\{1, \ldots, d\}$ defined by*

$$\Pr(\omega = k) = \begin{cases} 0 & \text{if } k \in \alpha, \\ p_k > 0 & \text{if } k \in \beta. \end{cases}$$

Note that each $\Theta_k, k \in \beta$ is assumed to be sampled with a positive probability in this sampling scheme. In particular, when the sampling probability of all $k \in \beta$ are equal, i.e., $p_k = 1/d_2 \ \forall\, k \in \beta$, we say that the observations are sampled uniformly at random.

Next, we present some examples of low-rank matrix completion problems in the above settings.

(1) Correlation matrix completion. A correlation matrix is an $n \times n$ real symmetric or Hermitian positive semidefinite matrix with all diagonal entries being ones. Let e_i be the vector with the i-th entry being one and the others being zeros. Then, $\langle e_i e_i^{\mathbb{T}}, \overline{X} \rangle = \overline{X}_{ii} = 1 \ \forall\, 1 \le i \le n$. The recovery of a correlation matrix is based on the observations of entries. For the real case, $\mathbb{V}^{n_1 \times n_2} = \mathcal{S}^n$, $d = n(n+1)/2$, $d_1 = n$,

$$\Theta_\alpha = \left\{ e_i e_i^{\mathbb{T}} \;\middle|\; 1 \le i \le n \right\} \quad \text{and} \quad \Theta_\beta = \left\{ \frac{1}{\sqrt{2}} (e_i e_j^{\mathbb{T}} + e_j e_i^{\mathbb{T}}) \;\middle|\; 1 \le i < j \le n \right\};$$

and for the complex case, $\mathbb{V}^{n_1 \times n_2} = \mathcal{H}^n$, $d = n^2$, $d_1 = n$,

$$\Theta_\alpha = \left\{ e_i e_i^{\mathbb{T}} \;\middle|\; 1 \le i \le n \right\} \text{ and } \Theta_\beta = \left\{ \frac{1}{\sqrt{2}} (e_i e_j^{\mathbb{T}} + e_j e_i^{\mathbb{T}}), \frac{\sqrt{-1}}{\sqrt{2}} (e_i e_j^{\mathbb{T}} - e_j e_i^{\mathbb{T}}) \;\middle|\; i < j \right\}.$$

Here, $\sqrt{-1}$ represents the imaginary unit. Of course, one may fix some off-diagonal entries in specific applications.

(2) Density matrix completion. A density matrix of dimension $n = 2^l$ for some positive integer l is an $n \times n$ Hermitian positive semidefinite matrix with trace one. In quantum state tomography, one aims to recover a density matrix from Pauli measurements (observations of the coefficients relative to the Pauli basis) [29, 21], given by

$$\Theta_\alpha = \left\{ \frac{1}{\sqrt{n}} I_n \right\} \text{ and } \Theta_\beta = \left\{ \frac{1}{\sqrt{n}} (\sigma_{s_1} \otimes \cdots \otimes \sigma_{s_l}) \;\middle|\; (s_1, \ldots, s_l) \in \{0,1,2,3\}^l \right\} \Big\backslash \Theta_\alpha,$$

where "$\otimes$" means the Kronecker product of two matrices and

$$\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\ \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},\ \sigma_2 = \begin{pmatrix} 0 & -\sqrt{-1} \\ \sqrt{-1} & 0 \end{pmatrix},\ \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

are the Pauli matrices. In this setting, $\mathbb{V}^{n_1\times n_2} = \mathcal{H}^n$, $\mathrm{Tr}(\overline{X}) = \langle I_n, \overline{X}\rangle = 1$, $d = n^2$, and $d_1 = 1$.

(3) Rectangular matrix completion. Assume that a few entries of a rectangular matrix are known and let $\mathcal{I}$ be the index set of these entries. One aims to recover this rectangular matrix from the observations of the rest entries. In the real case, $\mathbb{V}^{n_1\times n_2} = \mathbb{R}^{n_1\times n_2}$, $d = n_1n_2$, $d_1 = |\mathcal{I}|$,

$$\Theta_\alpha = \left\{e_ie_j^{\mathbb{T}} \mid (i,j)\in\mathcal{I}\right\} \quad \text{and} \quad \Theta_\beta = \left\{e_ie_j^{\mathbb{T}} \mid (i,j)\notin\mathcal{I}\right\};$$

and in the complex case, $\mathbb{V}^{n_1\times n_2} = \mathbb{C}^{n_1\times n_2}$, $d = 2n_1n_2$, $d_1 = 2|\mathcal{I}|$,

$$\Theta_\alpha = \left\{e_ie_j^{\mathbb{T}}, \sqrt{-1}e_ie_j^{\mathbb{T}} \mid (i,j)\in\mathcal{I}\right\} \quad \text{and} \quad \Theta_\beta = \left\{e_ie_j^{\mathbb{T}}, \sqrt{-1}e_ie_j^{\mathbb{T}} \mid (i,j)\notin\mathcal{I}\right\}.$$

Now we introduce some linear operators that are frequently used in the subsequent sections. For any given index set $\pi \subseteq \{1,\ldots,d\}$, say α and β, we define the linear operators $\mathcal{R}_\pi$: $\mathbb{V}^{n_1\times n_2} \to \mathbb{R}^{|\pi|}$ and $\mathcal{P}_\pi$: $\mathbb{V}^{n_1\times n_2} \to \mathbb{V}^{n_1\times n_2}$, respectively, by

$$\mathcal{R}_\pi(X) := \big(\langle\Theta_k, X\rangle\big)_{k\in\pi}^{\mathbb{T}} \ \text{ and } \ \mathcal{P}_\pi(X) := \sum_{k\in\pi}\langle\Theta_k, X\rangle\Theta_k, \quad X\in\mathbb{V}^{n_1\times n_2}. \tag{2}$$

It is easy to see that $\mathcal{P}_\pi = \mathcal{R}_\pi^*\mathcal{R}_\pi$. Define the self-adjoint operators $\mathcal{Q}_\beta : \mathbb{V}^{n_1\times n_2} \to \mathbb{V}^{n_1\times n_2}$ and $\mathcal{Q}_\beta^\dagger : \mathbb{V}^{n_1\times n_2} \to \mathbb{V}^{n_1\times n_2}$ associated with the sampling probability, respectively, by

$$\mathcal{Q}_\beta(X) := \sum_{k\in\beta} p_k\langle\Theta_k, X\rangle\Theta_k \quad \text{and} \quad \mathcal{Q}_\beta^\dagger(X) := \sum_{k\in\beta}\frac{1}{p_k}\langle\Theta_k, X\rangle\Theta_k, \quad X\in\mathbb{V}^{n_1\times n_2}. \tag{3}$$

One may easily verify that the operators $\mathcal{Q}_\beta$, $\mathcal{Q}_\beta^\dagger$ and $\mathcal{P}_\beta$ satisfy the following relations

$$\mathcal{Q}_\beta\mathcal{Q}_\beta^\dagger = \mathcal{Q}_\beta^\dagger\mathcal{Q}_\beta = \mathcal{P}_\beta, \ \ \mathcal{P}_\beta\mathcal{Q}_\beta = \mathcal{Q}_\beta\mathcal{P}_\beta = \mathcal{Q}_\beta, \ \ \mathcal{Q}_\beta^\dagger\mathcal{R}_\alpha^* = 0. \tag{4}$$

Let Ω be the multiset of all sampled indices from the index set β, i.e., $\Omega := \{\omega_1,\ldots,\omega_m\}$. With a slight abuse on notation, we define the sampling operator $\mathcal{R}_\Omega$: $\mathbb{V}^{n_1\times n_2} \to \mathbb{R}^m$ associated with Ω by

$$\mathcal{R}_\Omega(X) := \big(\langle\Theta_{\omega_1}, X\rangle, \ldots, \langle\Theta_{\omega_m}, X\rangle\big)^{\mathbb{T}}, \quad X\in\mathbb{V}^{n_1\times n_2}.$$

Then, the observation model (1) can be expressed in the following vector form

$$y = \mathcal{R}_\Omega(\overline{X}) + \nu\xi, \tag{5}$$

where $y = (y_1,\ldots,y_m)^{\mathbb{T}} \in \mathbb{R}^m$ and $\xi = (\xi_1,\ldots,\xi_m)^{\mathbb{T}} \in \mathbb{R}^m$ denote the observation vector and the noise vector, respectively.

2.2 The rank-correction step

In many situations, the nuclear norm is able to encourage a low-rank solution for matrix recovery, but its efficiency may be challenged if the observations are sampled at random obeying a general distribution such as the one considered in [66]. The setting of fixed basis coefficients in our matrix completion model can be regarded to be under an extreme sampling scheme. In particular, for the correlation and density matrix completion, the nuclear norm completely loses its efficiency since in this case it reduces to a constant. In order to overcome the shortcomings of the nuclear norm penalization, we propose a rank-correction step to generate an estimator in pursuit of a better recovery performance.

For convenience of discussions, in the rest of this paper, for any given $X \in \mathbb{V}^{n_1\times n_2}$, we denote by $\sigma(X) = \big(\sigma_1(X),\ldots,\sigma_n(X)\big)^{\mathbb{T}}$ the singular value vector of X arranged in the nonincreasing order and define

$$\mathbb{O}^{n_1,n_2}(X) := \big\{(U,V) \in \mathbb{O}^{n_1}\times\mathbb{O}^{n_2} \;\big|\; X = U\mathrm{Diag}\big(\sigma(X)\big)V^{\mathbb{T}}\big\}.$$

In particular, when $\mathbb{V}^{n_1\times n_2} = \mathbb{S}^n$, we denote by $\lambda(X) = \big(\lambda_1(X),\ldots,\lambda_n(X)\big)^{\mathbb{T}}$ the eigenvalue vector of X with $|\lambda_1(X)| \ge \ldots \ge |\lambda_n(X)|$ and define

$$\mathbb{O}^{n}(X) := \big\{P \in \mathbb{O}^{n} \;\big|\; X = P\mathrm{Diag}(\lambda(X))P^{\mathbb{T}}\big\}.$$

Before stating our rank-correction step, we introduce the concept of spectral operator associated to a symmetric vector-valued function. A function $f:\mathbb{R}^n\to\mathbb{R}^n$ is said to be symmetric if

$$f(x) = Q^{\mathbb{T}}f(Qx) \quad \forall\,\text{signed permutation matrix } Q \text{ and } x\in\mathbb{R}^n,$$

where a signed permutation matrix is a real matrix that contains exactly one nonzero entry 1 or -1 in each row and column and 0 elsewhere. From this definition, we see that

$$f_i(x) = 0 \quad \text{if } x_i = 0.$$

The spectral operator $F:\mathbb{V}^{n_1\times n_2}\to\mathbb{V}^{n_1\times n_2}$ associated with the function f is defined by

$$F(X) := U\mathrm{Diag}\big(f(\sigma(X))\big)V^{\mathbb{T}}, \tag{6}$$

where $(U,V)\in\mathbb{O}^{n_1,n_2}(X)$ and $X\in\mathbb{V}^{n_1\times n_2}$. From [10, Theorems 3.1 & 3.6], the symmetry of f guarantees the well-definiteness of the spectral operator F, and the continuous differentiability of f implies the continuous differentiability of F. When $\mathbb{V}^{n_1\times n_2} = \mathbb{S}^n$, we have that

$$F(X) = P\mathrm{Diag}\big(f(|\lambda(X)|)\big)\big(P\mathrm{Diag}(s(X))\big)^{\mathbb{T}},$$

where $P\in\mathbb{O}^n(X)$ and $s(X)\in\mathbb{R}^n$ with the i-th component $s_i(X) = -1$ if $\lambda_i(X) < 0$ and $s_i(X) = 1$ otherwise. In particular for the positive semidefinite case, both U and V in (6) reduce to P. For more details on spectral operators, the readers may refer to the PhD thesis [10].

Given a spectral operator $F\colon \mathbb{V}^{n_1\times n_2} \to \mathbb{V}^{n_1\times n_2}$ and an initial estimator $\widetilde{X}_m$ for the unknown matrix $\overline{X}$, say the nuclear norm penalized least squares estimator or one of its analogies, our rank-correction step is to solve the convex optimization problem

$$\begin{array}{ll} \min\limits_{X\in\mathbb{V}^{n_1\times n_2}} & \dfrac{1}{2m}\|y-\mathcal{R}_\Omega(X)\|_2^2+\rho_m\Big(\|X\|_*-\langle F(\widetilde{X}_m),X\rangle+\dfrac{\gamma_m}{2}\|X-\widetilde{X}_m\|_F^2\Big) \\ \text{s.t.} & \mathcal{R}_\alpha(X)=\mathcal{R}_\alpha(\overline{X}), \end{array} \tag{7}$$

where $\rho_m>0$ and $\gamma_m\ge 0$ are the regularization parameters depending on the number of observations. The last quadratic proximal term is added to guarantee the boundness of the solution to (7). If the function $\|X\|_*-\langle F(\widetilde{X}_m),X\rangle$ is level-bounded, one may simply set $\gamma_m=0$. Clearly, when $F\equiv 0$ and $\gamma_m=0$, the problem (7) reduces to the nuclear norm penalized least squares problem. In the sequel, we call $-\langle F(\widetilde{X}_m),X\rangle$ the rank-correction term. If the true matrix is known to be positive semidefinite, we add the constraint $X\in\mathbb{S}^n_+$ to (7). Thus, the rank-correction step is to solve the convex conic optimization problem

$$\begin{array}{ll} \min\limits_{X\in\mathbb{S}^n} & \dfrac{1}{2m}\|y-\mathcal{R}_\Omega(X)\|_2^2+\rho_m\Big(\langle I-F(\widetilde{X}_m),X\rangle+\dfrac{\gamma_m}{2}\|X-\widetilde{X}_m\|_F^2\Big) \\ \text{s.t.} & \mathcal{R}_\alpha(X)=\mathcal{R}_\alpha(\overline{X}),\quad X\in\mathbb{S}^n_+. \end{array} \tag{8}$$

For this case, we assume that the initial estimator $\widetilde{X}_m$ belongs to $\mathbb{S}^n_+$ as the projection of any estimator onto $\mathbb{S}^n_+$ can approximate the true matrix $\overline{X}$ better.

The rank-correction step above is inspired by the majorized penalty approach recently proposed by Gao and Sun [25] for solving the rank constrained matrix optimization problem:

$$\min_{X\in K}\big\{h(X):\ \mathrm{rank}(X)\le r\big\}, \tag{9}$$

where $r\ge 1$, $h:\mathbb{V}^{n_1\times n_2}\to\mathbb{R}$ is a given continuous function and $K\in\mathbb{V}^{n_1\times n_2}$ is a closed convex set. Note that for any $X\in\mathbb{V}^{n_1\times n_2}$, the constraint $\mathrm{rank}(X)\le r$ is equivalent to

$$0=\sigma_{r+1}(X)+\cdots+\sigma_n(X)=\|X\|_*-\|X\|_{(r)},$$

where $\|X\|_{(r)}:=\sigma_1(X)+\cdots+\sigma_r(X)$ denotes the Ky Fan r-norm. The central idea of the majorized penalty approach is to solve the following penalized version of (9):

$$\min_{X\in K}\ h(X)+\rho\big(\|X\|_*-\|X\|_{(r)}\big),$$

where $\rho>0$ is the penalty parameter. With the current iterate X^k, the majorized penalty approach yields the next iterate X^{k+1} by solving the convex optimization problem

$$\min_{X\in K}\ \widehat{h}^k(X)+\rho\Big(\|X\|_*-\langle G^k,X\rangle+\frac{\gamma_k}{2}\|X-X^k\|_F^2\Big), \tag{10}$$

where $\gamma_k\ge 0$, G^k is a subgradient of the convex function $\|X\|_{(r)}$ at X^k, and $\widehat{h}^k$ is a convex majorization function of h at X^k. Comparing with (7), one may notice that our proposed rank-correction step is close to one step of the majorized penalty approach.

Due to the structured randomness of matrix completion, we expect that the estimator generated from the rank-correction step possesses some favorable properties for recovery. The key issue is how to construct the rank-correction function F to make such improvements possible. In the next two sections, we provide theoretical supports to our proposed rank-correction step, from which some important guidelines on the construction of F can be captured.

Henceforth, we let $\widehat{X}_m$ denote the estimator generated from the rank-correction step (7) or (8) for the corresponding cases and let $r = \mathrm{rank}(\overline{X}) \geq 1$. Throughout this paper, for any $X \in \mathbb{V}^{n_1\times n_2}$ and any $(U, V) \in \mathbb{O}^{n_1,n_2}(X)$, we write $U = [U_1 \;\; U_2]$ and $V = [V_1 \;\; V_2]$ with $U_1 \in \mathbb{O}^{n_1\times r}$, $U_2 \in \mathbb{O}^{n_1\times(n_1-r)}$, $V_1 \in \mathbb{O}^{n_2\times r}$ and $V_2 \in \mathbb{O}^{n_2\times(n_2-r)}$. Meanwhile, for any $X \in \mathbb{S}^n_+$ and any $P \in \mathbb{O}^n(X)$, we write $P = [P_1 \;\; P_2]$ with $P_1 \in \mathbb{O}^{n\times r}$ and $P_2 \in \mathbb{O}^{n\times(n-r)}$.

3 Error bounds

In this section, we aim to derive a recovery error bound in the Frobenius norm for the rank-correction step and discuss the impact of the rank-correction term on the obtained bound. The following analysis focuses on the recovery of a rectangular matrix or a symmetric/Hermitian matrix. All the results obtained in this section are applicable to the positive semidefinite case since adding more prior information can only improve recoverability.

We first introduce the orthogonal decomposition $\mathbb{V}^{n_1\times n_2} = T \oplus T^\perp$ with

$$T := \big\{X \in \mathbb{V}^{n_1\times n_2} \;\big|\; X = X_1 + X_2 \text{ with } \mathrm{col}(X_1) \subseteq \mathrm{col}(\overline{X}), \mathrm{row}(X_2) \subseteq \mathrm{row}(\overline{X})\big\},$$
$$T^\perp := \big\{X \in \mathbb{V}^{n_1\times n_2} \;\big|\; \mathrm{row}(X) \perp \mathrm{row}(\overline{X}) \text{ and } \mathrm{col}(X) \perp \mathrm{col}(\overline{X})\big\},$$

where $\mathrm{row}(X)$ and $\mathrm{col}(X)$ denote the row space and column space of the matrix X, respectively. Let $\mathcal{P}_T : \mathbb{V}^{n_1\times n_2} \to \mathbb{V}^{n_1\times n_2}$ and $\mathcal{P}_{T^\perp} : \mathbb{V}^{n_1\times n_2} \to \mathbb{V}^{n_1\times n_2}$ be the orthogonal projection operators onto the subspaces T and $T^\perp$, respectively. It is not hard to verify that

$$\mathcal{P}_T(X) = \overline{U}_1\overline{U}_1^{\mathbb{T}}X + X\overline{V}_1\overline{V}_1^{\mathbb{T}} - \overline{U}_1\overline{U}_1^{\mathbb{T}}X\overline{V}_1\overline{V}_1^{\mathbb{T}} \;\text{ and }\; \mathcal{P}_{T^\perp}(X) = \overline{U}_2\overline{U}_2^{\mathbb{T}}X\overline{V}_2\overline{V}_2^{\mathbb{T}} \tag{11}$$

for any $X \in \mathbb{V}^{n_1\times n_2}$ and $(\overline{U}, \overline{V}) \in \mathbb{O}^{n_1,n_2}(\overline{X})$. Define a_m and b_m, respectively, by

$$a_m := \|\overline{U}_1\overline{V}_1^{\mathbb{T}} - \mathcal{P}_T(F(\widetilde{X}_m) + \gamma_m\widetilde{X}_m))\| \;\text{ and }\; b_m := 1 - \|\mathcal{P}_{T^\perp}(F(\widetilde{X}_m) + \gamma_m\widetilde{X}_m))\|. \tag{12}$$

Note that the first term in the objective function of (7) can be rewritten as

$$\frac{1}{2m}\|y - \mathcal{R}_\Omega(X)\|_2^2 = \frac{1}{2m}\left\|\mathcal{R}_\Omega(X - \overline{X})\right\|_2^2 - \frac{\nu}{m}\left\langle\mathcal{R}_\Omega^*(\xi), X\right\rangle.$$

Using the optimality of $\widehat{X}_m$ to the problem (7), we obtain the following result.

Theorem 1 *Assume that* $\|\mathcal{P}_{T^\perp}(F(\widetilde{X}_m)+\gamma_m\widetilde{X}_m)\|<1$. *For any given* $\kappa>1$, *if*

$$\rho_m \geq \frac{\kappa\nu}{b_m}\Big\|\frac{1}{m}\mathcal{R}^*_\Omega(\xi)\Big\|, \tag{13}$$

then the following inequality holds:

$$\frac{1}{2m}\big\|\mathcal{R}_\Omega(\widehat{X}_m-\overline{X})\big\|_2^2 \leq \sqrt{2r}\Big(a_m+\frac{b_m}{\kappa}\Big)\rho_m\|\widehat{X}_m-\overline{X}\|_F+\frac{\rho_m\gamma_m}{2}\left(\|\overline{X}\|_F^2-\|\widehat{X}_m\|_F^2\right). \tag{14}$$

Theorem 1 shows that, to derive an error bound on $\|\widehat{X}_m-\overline{X}\|_F$, we only need to establish the relation between $\|\widehat{X}_m-\overline{X}\|_F^2$ and $\frac{1}{m}\|\mathcal{R}_\Omega(\widehat{X}_m-\overline{X})\|_2^2$. It is well known that the sampling operator $\mathcal{R}_\Omega$ does not satisfy the RIP, but it has a similar property with high probability under certain conditions (see, e.g., [55, 42, 38, 46]). For deriving such a property, here, we impose a bound restriction on the true matrix $\overline{X}$ in the form of $\|\mathcal{R}_\beta(\overline{X})\|_\infty \leq c$. This condition is very mild since a bound is often known in some applications such as in the correlation and density matrix completion. Correspondingly, we add the bound constraint $\|\mathcal{R}_\beta(X)\|_\infty \leq c$ to the problem (7) in the rank-correction step. Since the feasible set is bounded in this case, we simply set $\gamma_m = 0$ and let $\widehat{X}^c_m$ denote the estimator generated from the rank-correction step in this case.

The above boundedness setting is similar to the one adopted by Klopp [38] for the nuclear norm penalized least squares estimator. A slight difference is that the upper bound is imposed on the basis coefficients of $\overline{X}$ relative to $\{\Theta_k : k\in\beta\}$ rather than all the entries of $\overline{X}$. It is easy to see that if the bound is not too tight, the estimator $\widehat{X}^c_m$ is the same as $\widehat{X}_m$. Therefore, we next derive the recovery error bound of $\widehat{X}^c_m$ instead of $\widehat{X}_m$, by following Klopp's arguments in [38], which are also in line with the work done by Negahban and Wainwright [55].

Let μ_1 be a constant to control the smallest sampling probability for observations as

$$p_k \geq (\mu_1 d_2)^{-1} \quad \forall\, k\in\beta. \tag{15}$$

It follows from Assumption 1 that $\mu_1 \geq 1$ and in particular $\mu_1 = 1$ for the uniform sampling. Note that the magnitude of μ_1 does not depend on d_2 or the matrix size. By the definition of $\mathcal{Q}_\beta$, we then have

$$\langle\mathcal{Q}_\beta(\Delta),\Delta\rangle \geq (\mu_1 d_2)^{-1}\|\Delta\|_F^2 \quad \forall\,\Delta\in\{\Delta\in\mathbb{V}^{n_1\times n_2} \mid \mathcal{R}_\alpha(\Delta)=0\}. \tag{16}$$

Let $\{\epsilon_1,\ldots,\epsilon_m\}$ be an i.i.d. Rademacher sequence, i.e., an i.i.d. sequence of Bernoulli random variables taking the values 1 and -1 with probability $1/2$. Define

$$\vartheta_m := \mathbb{E}\Big\|\frac{1}{m}\mathcal{R}^*_\Omega(\epsilon)\Big\| \quad \text{with} \quad \epsilon=(\epsilon_1,\ldots,\epsilon_m)^{\mathbb{T}}. \tag{17}$$

Then, we can obtain a similar result to [38, Lemma 12] by showing that the sampling operator $\mathcal{R}_\Omega$ satisfies some approximate RIP for the matrices in the following set

$$K(r) := \Big\{\Delta\in\mathbb{V}^{n_1\times n_2}\,\Big|\, \mathcal{R}_\alpha(\Delta)=0,\ \|\mathcal{R}_\beta(\Delta)\|_\infty=1,\ \|\Delta\|_*\leq\sqrt{r}\|\Delta\|_F,\ \langle\mathcal{Q}_\beta(\Delta),\Delta\rangle \geq \sqrt{\frac{64\log(n_1+n_2)}{\log(2)m}}\Big\}.$$

Lemma 2 *For all matrices $\Delta \in K(r)$, with probability at least $1-2/(n_1+n_2)$, we have*

$$\frac{1}{m}\|\mathcal{R}_\Omega(\Delta)\|_2^2 \geq \frac{1}{2}\langle \mathcal{Q}_\beta(\Delta), \Delta\rangle - 128\mu_1 d_2 r \vartheta_m^2.$$

Now, by combining Theorem 1 and Lemma 2, we obtain the following result.

Theorem 3 *Assume that $\|\mathcal{P}_{T^\perp}(F(\widetilde{X}_m))\| < 1$ and $\|\mathcal{R}_\beta(\overline{X})\|_\infty \leq c$ for some constant c. If ρ_m is chosen to satisfy (13), then there exists a numerical constant C such that*

$$\frac{\|\widehat{X}_m^c - \overline{X}\|_F^2}{d_2} \leq C \max\left\{\mu_1^2 d_2 r\left(\left(a_m + \frac{b_m}{\kappa}\right)^2 \rho_m^2 + \frac{\kappa^2(a_m+b_m)^2}{(\kappa-1)^2 b_m^2} c^2 \vartheta_m^2\right), c^2\mu_1\sqrt{\frac{\log(n_1+n_2)}{m}}\right\}$$

with probability at least $1-2/(n_1+n_2)$.

In order to choose a parameter ρ_m such that (13) holds, we need to estimate $\|\frac{1}{m}\mathcal{R}_\Omega^*(\xi)\|$. For this purpose, we make the following assumption on the noises.

Assumption 2 *The i.i.d. noise variables ξ_i are sub-exponential, i.e., there exist positive constants c_1, c_2 and c_3 such that for all $t > 0$, $\Pr(|\xi_i| \geq t) \leq c_1 \exp(-c_2 t^{c_3})$.*

The noncommutative Bernstein inequality is a useful tool for the study of matrix completion problems. It provides bounds of the probability that the sum of random matrices deviates from its mean in the operator norm (see, e.g., [60, 69, 28]). Recently, the noncommutative Bernstein inequality was extended by replacing bounds of the operator norm of matrices with bounds of the Orlicz norms (see [41, 42]). Given any $s \geq 1$, the ψ_s Orlicz norm of a random variable θ is defined by

$$\|\theta\|_{\psi_s} := \inf\left\{t > 0 \;\middle|\; \mathbb{E}\exp(|\theta|^s/t^s) \leq 2\right\}.$$

The Orlicz norms are useful to characterize the tail behavior of random variables. The following noncommutative Bernstein inequality is taken from [40, Corollary 2.1].

Proposition 4 *Let $Z_1, \ldots, Z_m \in \mathbb{V}^{n_1\times n_2}$ be independent random matrices with mean zero. Suppose that $\max\left\{\left\|\|Z_i\|\right\|_{\psi_s}, 2\mathbb{E}^{\frac{1}{2}}(\|Z_i\|^2)\right\} < \varpi_s$ for some constant ϖ_s. Define*

$$\sigma_Z := \max\left\{\left\|\frac{1}{m}\sum_{i=1}^m \mathbb{E}(Z_i Z_i^{\mathbb{T}})\right\|^{1/2}, \left\|\frac{1}{m}\sum_{i=1}^m \mathbb{E}(Z_i^{\mathbb{T}} Z_i)\right\|^{1/2}\right\}.$$

Then, there exists a constant C such that for all $t > 0$, with probability at least $1-\exp(-t)$,

$$\left\|\frac{1}{m}\sum_{i=1}^m Z_i\right\| \leq C\max\left\{\sigma_Z\sqrt{\frac{t+\log(n_1+n_2)}{m}}, \varpi_s\left(\log\frac{\varpi_s}{\sigma_Z}\right)^{1/s}\frac{t+\log(n_1+n_2)}{m}\right\}.$$

It is known that a random variable is sub-exponential if and only its ψ_1 Orlicz norm is finite [53]. To apply the noncommutative Bernstein inequality, we let μ_2 be a constant such that

$$\max\left\{\left\|\sum_{k\in\beta} p_k\Theta_k\Theta_k^{\mathbb{T}}\right\|,\ \left\|\sum_{k\in\beta} p_k\Theta_k^{\mathbb{T}}\Theta_k\right\|\right\} \le \frac{\mu_2}{n}. \tag{18}$$

Notice that since $\mathrm{Tr}\big(\sum_{k\in\beta} p_k\Theta_k\Theta_k^{\mathbb{T}}\big) = \mathrm{Tr}\big(\sum_{k\in\beta} p_k\Theta_k^{\mathbb{T}}\Theta_k\big) = 1$, the lower bound of the term on the left-hand side is $1/n$. This implies that $\mu_2 \ge 1$. In the following, we also assume that the magnitude of μ_2 does not depend on the matrix size. For example, $\mu_2 = 1$ for the correlation matrix completion under uniform sampling and the density matrix completion described in Section 2. The following result extends [42, Lemma 2] and [38, Lemmas 5 & 6] from the standard basis to an arbitrary orthonormal basis. A similar result can also be found in [55, Lemma 6].

Lemma 5 *Under Assumption 2, there exists a constant C^* (only depending on the ψ_1 Orlicz norm of ξ_k) such that for all $t > 0$, with probability at least $1 - \exp(-t)$,*

$$\left\|\frac{1}{m}\mathcal{R}_\Omega^*(\xi)\right\| \le C^* \max\left\{\sqrt{\frac{\mu_2(t+\log(n_1+n_2))}{mn}}, \frac{\log(n)(t+\log(n_1+n_2))}{m}\right\}. \tag{19}$$

In particular, when $m \ge n\log^3(n_1+n_2)/\mu_2$, we also have

$$\mathbb{E}\left\|\frac{1}{m}\mathcal{R}_\Omega^*(\xi)\right\| \le C^*\sqrt{\frac{2e\mu_2\log(n_1+n_2)}{mn}}. \tag{20}$$

Since Bernoulli random variables are sub-exponential, the right-hand side of (20) provides an upper bound of ϑ_m defined by (17). Now, we choose $t = \log(n_1+n_2)$ in Lemma 5 for achieving an optimal order bound. With this choice, when $m \ge 2n\log^2(n_1+n_2)/\mu_2$, the first term in the maximum of (19) dominates the second term. Hence, for any given $\kappa > 1$, by choosing

$$\rho_m = \frac{\kappa\nu}{b_m}C^*\sqrt{\frac{2\mu_2\log(n_1+n_2)}{mn}}, \tag{21}$$

from Theorem 3 and Lemma 5, we obtain the following main result for recovery error bound.

Theorem 6 *Assume that $\|\mathcal{P}_{T^\perp}(F(\widetilde{X}_m))\| < 1$, $\|\mathcal{R}_\beta(\overline{X})\|_\infty \le c$ for some constant c, and Assumption 2 holds. For any given $\kappa > 1$, if ρ_m is chosen according to (21), then there exists a numerical constant C' such that, when $m \ge n\log^3(n_1+n_2)/\mu_2$,*

$$\frac{\|\widehat{X}_m^c - \overline{X}\|_F^2}{d_2} \le C'\max\left\{\left[\left(1+\kappa\frac{a_m}{b_m}\right)^2\nu^2 + \left(\frac{\kappa}{\kappa-1}\right)^2\left(1+\frac{a_m}{b_m}\right)^2 c^2\right]\frac{\mu_1^2\mu_2 d_2 r\log(n_1+n_2)}{mn},\right.$$
$$\left. c^2\mu_1\sqrt{\frac{\log(n_1+n_2)}{m}}\right\} \tag{22}$$

with probability at least $1 - 3/(n_1+n_2)$.

When the matrix size is large, the second term in the maximum of (22) is negligible, compared with the first term. Thus, Theorem 6 indicates that for any rank-correction function such that $\|\mathcal{P}_{T^\perp}(F(\widetilde{X}_m))\| < 1$, one only needs samples with size of order $d_2 r \log(n_1+n_2)/n$ to control the recovery error. Note that d_2 is of order $n_1 n_2$ in general. Hence, the order of sample size needed is roughly the degree of freedom of a rank r matrix up to a logarithmic factor in the matrix size. In addition, it is very interesting to notice that the value of κ (or the value of ρ_m) has a substantial influence on the recovery error bound. The first term in the maximum of (22) is a sum of two parts related to ν and c, respectively. The part related to ν will increase as κ increases provided $a_m/b_m > 0$, while the part related to c will slightly decreases to its limit as κ increases.

Theorem 6 also reveals the impact of the rank-correction term on recovery error. Note that the value of a_m/b_m fully depends on the rank-correction function F when an initial estimator $\widetilde{X}_m$ is given. A smaller value of a_m/b_m brings a smaller error bound and potentially leads to a smaller recovery error for the rank-correction step. Note that for any given $\varepsilon_1 \geq 0$ and $0 \leq \varepsilon_2 < 1$, we have

$$\frac{a_m}{b_m} \leq \frac{\varepsilon_1}{1-\varepsilon_2} \quad \text{if} \quad \big\|\mathcal{P}_T\big(F(\widetilde{X}_m)\big) - \overline{U}_1\overline{V}_1^{\mathbb{T}}\big\| \leq \varepsilon_1 \text{ and } \big\|\mathcal{P}_{T^\perp}\big(F(\widetilde{X}_m)\big)\big\| \leq \varepsilon_2.$$

In particular, if $F \equiv 0$, then the estimator of the rank-correction step reduces to the nuclear norm penalized least squares estimator with $a_m/b_m = 1$. Thus, Theorem 6 shows that, with a suitable rank-correction function F, the estimator generated from the rank-correction step for recovery is very likely to perform better than the nuclear norm penalized least squares estimator. In addition, this observation also provides us clues on how to construct a good rank-correction function, to be discussed in Section 5.

4 Rank consistency

In this section we study the asymptotic behavior of the rank of the estimator $\widehat{X}_m$ for both the rectangular case and the positive semidefinite case. Theorem 6 shows that under mild conditions, the distribution of $\widehat{X}_m$ becomes more and more concentrated to the true matrix $\overline{X}$. Due to the low-rank structure of $\overline{X}$, we expect that the estimator $\widehat{X}_m$ has the same low-rank property as $\overline{X}$. For this purpose, we consider the rank consistency in the sense of Bach [3] under the setting that the matrix size is fixed.

Definition 1 *An estimator X_m of the true matrix $\overline{X}$ is said to be rank consistent if*

$$\lim_{m\to\infty} \Pr\big(\mathrm{rank}(X_m) = \mathrm{rank}(\overline{X})\big) = 1.$$

Throughout this section we make the following assumptions:

Assumption 3 *The spectral operator F is continuous at $\overline{X}$.*

Assumption 4 *The initial estimator $\widetilde{X}_m$ satisfies $\widetilde{X}_m \xrightarrow{p} \overline{X}$ as $m \to \infty$.*

In addition, we also need the following properties of the operator $\mathcal{R}_\Omega$ and its adjoint $\mathcal{R}_\Omega^*$.

Lemma 7 **(i)** *For any given $X \in \mathbb{V}^{n_1 \times n_2}$, the random matrix $\frac{1}{m}\mathcal{R}_\Omega^*\mathcal{R}_\Omega(X) \xrightarrow{a.s.} \mathcal{Q}_\beta(X)$.*
(ii) *The random vector $\frac{1}{\sqrt{m}}\mathcal{R}_{\alpha\cup\beta}\mathcal{R}_\Omega^*(\xi) \xrightarrow{d} N\big(0, \mathrm{Diag}(p)\big)$, where $p = (p_1, \ldots, p_d)^{\mathbb{T}}$.*

Epi-convergence in distribution is useful in proving the convergence in distribution of minimizers or ε_m-minimizers. The following epi-convergence result of Knight [39, Theorem 1] is particularly useful in this regard (see also [30, Proposition 9]).

Proposition 8 *Let $\{\Phi_m\}$ be a sequence of random lower-semicontinuous functions that epi-converges in distribution to Φ. Assume that*

(i) *$\widehat{x}_m$ is an ε_m-minimizer of Φ_m, i.e., $\Phi_m(\widehat{x}_m) \le \inf \Phi_m(x) + \varepsilon_m$, where $\varepsilon_m \xrightarrow{p} 0$;*

(ii) *$\widehat{x}_m = O_p(1)$;*

(iii) *the function Φ has a unique minimizer $\overline{x}$.*

Then, $\widehat{x}_m \xrightarrow{d} \overline{x}$. In addition, if Φ is a deterministic function, then $\widehat{x}_m \xrightarrow{p} \overline{x}$.

We know from [27] that $\widehat{x}_m$ is guaranteed to be $O_p(1)$ when all Φ_m are convex functions and Φ has a unique minimizer. For more details on epi-convergence in distribution, one may refer to King and Wets [36], Geyer [26], Pflug [57] and Knight [39]. In order to apply the epi-convergence theorem to a constrained optimization problem, we need to transform the constrained optimization problem into an unconstrained one by using the indicator function of the feasible set. This leads to the epi-convergence issue of the sum of two sequences of functions. Thus, we need the following epi-convergence result stated in [57, Lemma 1].

Proposition 9 *Let $\{\Phi_m\}$ be a sequence of random lower-semicontinuous functions and $\{\Psi_m\}$ be a sequence of deterministic lower-semicontinuous functions. If either of the following two assumptions holds:*

(i) *Φ_m epi-converges in distribution to Φ and Ψ_m converges to Ψ with respect to the topology of uniform convergence on compact sets;*

(ii) *Φ_m converges in distribution to Φ with respect to the topology of uniform convergence on compact sets and Ψ_m epi-converges to Ψ,*

then $\Phi_m + \Psi_m$ epi-converges in distribution to $\Phi + \Psi$.

Based on the above epi-convergence results, we can analyze the asymptotic behavior of optimal solutions of a sequence of constrained optimization problems. The following result is a direct consequence of the above epi-convergence results and Lemma 7.

Theorem 10 *If $\rho_m \to 0$ and $\gamma_m = O_p(1)$, then $\widehat{X}_m \xrightarrow{p} \overline{X}$ as $m \to \infty$.*

Then, according to Theorem 10 and the lower semi-continuity of the rank function, it is straightforward to obtain:

Corollary 11 *If* $X_m \xrightarrow{p} \overline{X}$, *then* $\lim_{m\to\infty} \Pr\big(\mathrm{rank}(X_m) \geq \mathrm{rank}(\overline{X})\big) = 1$.

In what follows, we focus on the characterization of necessary and sufficient conditions for rank consistency of $\widehat{X}_m$. The idea is similar to that of [3] for the nuclear norm penalized least squares estimator. Note that, unlike for the recover error bound, adding more constraints may break the rank consistency. Therefore, we separate the discussion into the rectangular case (recovering a rectangular matrix or a symmetric/Hermitian matrix) and the positive semidefinite case (recovering a positive/Hermitian semidefinite matrix) below.

4.1 The rectangular case

Since we have established that $\widehat{X}_m \xrightarrow{p} \overline{X}$, we only need to focus on some neighborhood of $\overline{X}$ for the discussion about the rank consistency of $\widehat{X}_m$. First, we take a look at a local property of the rank function via the directional derivative of the singular value functions.

Let $\sigma_i'(X;\cdot)$ denote the directional derivative function of the i-th largest singular value function $\sigma_i(\cdot)$ at X. From [45, Section 5.1] and [11, Proposition 6], for $\mathbb{V}^{n_1\times n_2} \ni H \to 0$,

$$\sigma_i(X+H) - \sigma_i(X) - \sigma_i'(X;H) = O(\|H\|_F^2), \quad i = 1,\ldots,n. \tag{23}$$

Recall that $r = \mathrm{rank}(\overline{X})$. From [11, Proposition 6], we have

$$\sigma_{r+1}'(\overline{X};H) = \|\overline{U}_2^{\mathbb{T}} H \overline{V}_2\|, \quad H \in \mathbb{V}^{n_1\times n_2}.$$

This leads to the following result for the perturbation of the rank function. A similar result can also be found in [3, Proposition 18], whose proof is more involved.

Lemma 12 *Let* $\overline{\Delta} \in \mathbb{V}^{n_1\times n_2}$ *satisfy* $\overline{U}_2^{\mathbb{T}}\overline{\Delta}\,\overline{V}_2 \neq 0$. *Then, for all* $\rho \neq 0$ *sufficiently small and* Δ *sufficiently close to* $\overline{\Delta}$, $\mathrm{rank}(\overline{X}+\rho\Delta) > \mathrm{rank}(\overline{X})$.

To guarantee the efficiency of the rank-correction term on encouraging a low-rank solution, the parameter ρ_m should not decay too fast. Define $\widehat{\Delta}_m := \rho_m^{-1}(\widehat{X}_m - \overline{X})$. Then, for a slow decay on ρ_m, we can establish the following result.

Proposition 13 *If* $\rho_m \to 0, \sqrt{m}\rho_m \to \infty$ *and* $\gamma_m = O_p(1)$, *then* $\widehat{\Delta}_m \xrightarrow{p} \widehat{\Delta}$, *where* $\widehat{\Delta}$ *is the unique optimal solution to the following convex optimization problem*

$$\begin{array}{ll} \min\limits_{\Delta\in\mathbb{V}^{n_1\times n_2}} & \frac{1}{2}\langle \mathcal{Q}_\beta(\Delta),\Delta\rangle + \langle \overline{U}_1\overline{V}_1^{\mathbb{T}} - F(\overline{X}),\Delta\rangle + \|\overline{U}_2^{\mathbb{T}}\Delta\overline{V}_2\|_* \\ \mathrm{s.t.} & \mathcal{R}_\alpha(\Delta) = 0. \end{array} \tag{24}$$

Note that $\widehat{X}_m = \overline{X} + \rho_m \widehat{\Delta}_m$. From Corollary 11, Lemma 12 and Proposition 13, we see that the condition $\overline{U}_2^{\mathbb{T}} \widehat{\Delta} \overline{V}_2 = 0$ is necessary for the rank consistency of $\widehat{X}_m$. From the following property of the unique solution $\widehat{\Delta}$ to (24), we can derive a more detailed necessary condition for rank consistency as stated in Theorem 15 below.

Lemma 14 *Let $\widehat{\Delta}$ be the optimal solution to (24). Then $\overline{U}_2^{\mathbb{T}} \widehat{\Delta} \overline{V}_2 = 0$ if and only if the linear system*

$$\overline{U}_2^{\mathbb{T}} \mathcal{Q}_\beta^\dagger (\overline{U}_2 \Gamma \overline{V}_2^{\mathbb{T}}) \overline{V}_2 = \overline{U}_2^{\mathbb{T}} \mathcal{Q}_\beta^\dagger \big(\overline{U}_1 \overline{V}_1^{\mathbb{T}} - F(\overline{X}) \big) \overline{V}_2 \tag{25}$$

has a solution $\widehat{\Gamma} \in \mathbb{V}^{(n_1-r)\times(n_2-r)}$ with $\|\widehat{\Gamma}\| \le 1$. Moreover, in this case,

$$\widehat{\Delta} = \mathcal{Q}_\beta^\dagger \big(\overline{U}_2 \widehat{\Gamma}\, \overline{V}_2^{\mathbb{T}} - \overline{U}_1 \overline{V}_1^{\mathbb{T}} + F(\overline{X}) \big). \tag{26}$$

Theorem 15 *If $\rho_m \to 0$, $\sqrt{m}\rho_m \to \infty$ and $\gamma_m = O_p(1)$, then a necessary condition for the rank consistency of $\widehat{X}_m$ is that the linear system (25) has a solution $\widehat{\Gamma} \in \mathbb{V}^{(n_1-r)\times(n_2-r)}$ with $\|\widehat{\Gamma}\| \le 1$.*

By making a slight modification for the necessary condition in Theorem 15, we provide a sufficient condition for the rank consistency of the estimator $\widehat{X}_m$ as follows.

Theorem 16 *If $\rho_m \to 0, \sqrt{m}\rho_m \to \infty$ and $\gamma_m = O_p(1)$, then a sufficient condition for the rank consistency of the estimator $\widehat{X}_m$ is that the linear system (25) has a unique solution $\widehat{\Gamma} \in \mathbb{V}^{(n_1-r)\times(n_2-r)}$ with $\|\widehat{\Gamma}\| < 1$.*

4.2 The positive semidefinite case

For the positive semidefinite case, we first need the following Slater condition.

Assumption 5 *There exists some $X^0 \in \mathbb{S}^n_{++}$ such that $\mathcal{R}_\alpha(X^0) = \mathcal{R}_\alpha(\overline{X})$.*

Proposition 17 *If $\rho_m \to 0, \sqrt{m}\rho_m \to \infty$ and $\gamma_m = O_p(1)$, then $\widehat{\Delta}_m \xrightarrow{p} \widehat{\Delta}$, where $\widehat{\Delta}$ is the unique optimal solution to the following convex optimization problem*

$$\begin{array}{ll} \min\limits_{\Delta \in \mathbb{S}^n} & \dfrac{1}{2} \langle \mathcal{Q}_\beta(\Delta), \Delta \rangle + \langle I_n - F(\overline{X}), \Delta \rangle \\ \text{s.t.} & \mathcal{R}_\alpha(\Delta) = 0, \quad \overline{P}_2^{\mathbb{T}} \Delta \overline{P}_2 \in \mathbb{S}^{n-r}_+. \end{array} \tag{27}$$

For the optimal solution $\widehat{\Delta}$ to (27), we also have the following further characterization.

Lemma 18 *Let $\widehat{\Delta}$ be the optimal solution to (27). Then $\overline{P}_2^{\mathbb{T}} \widehat{\Delta} \overline{P}_2 = 0$ if and only if the linear system*

$$\overline{P}_2^{\mathbb{T}} \mathcal{Q}_\beta^\dagger (\overline{P}_2 \Lambda \overline{P}_2^{\mathbb{T}}) \overline{P}_2 = \overline{P}_2^{\mathbb{T}} \mathcal{Q}_\beta^\dagger \big(I_n - F(\overline{X}) \big) \overline{P}_2 \tag{28}$$

has a solution $\widehat{\Lambda} \in \mathbb{S}^{n-r}_+$. Moreover, in this case,

$$\widehat{\Delta} = \mathcal{Q}_\beta^\dagger \big(\overline{P}_2 \widehat{\Lambda}\, \overline{P}_2^{\mathbb{T}} - I_n + F(\overline{X}) \big). \tag{29}$$

Note that Lemma 12 still holds for the positive semidefinite case if $\overline{U}_2^{\mathbb{T}}\Delta\overline{V}_2$ is replaced by $\overline{P}_2^{\mathbb{T}}\Delta\overline{P}_2$. Therefore, in line with the rectangular case, from Lemma 18, we have the following necessary condition for rank consistency.

Theorem 19 *If $\rho_m \to 0, \sqrt{m}\rho_m \to \infty$ and $\gamma_m = O_p(1)$, then a necessary condition for the rank consistency of $\widehat{X}_m$ is that the linear system (28) has a solution $\widehat{\Lambda} \in \mathbb{S}_+^{n-r}$.*

Similarly to Theorem 16, we have the following sufficient condition for rank consistency for the positive semidefinite case.

Theorem 20 *If $\rho_m \to 0, \sqrt{m}\rho_m \to \infty$ and $\gamma_m = O_p(1)$, then a sufficient condition for the rank consistency of $\widehat{X}_m$ is that the linear system (28) has a unique solution $\widehat{\Lambda} \in \mathbb{S}_{++}^{n-r}$.*

4.3 Constraint nondegeneracy and rank consistency

In this subsection, with the help of constraint nondegeneracy, we provide conditions to guarantee that the linear systems (25) and (28) have a unique solution. The concept of constraint nondegeneracy was pioneered by Robinson [62] and later extensively developed by Bonnans and Shapiro [5]. Consider the following constrained optimization problem

$$\min_{X\in\mathbb{V}^{n_1\times n_2}} \Big\{\Phi(X) + \Psi(X) :\ \mathcal{A}(X) - b \in K\Big\}, \tag{30}$$

where $\Phi : \mathbb{V}^{n_1\times n_2} \to \mathbb{R}$ is a continuously differentiable function, $\Psi : \mathbb{V}^{n_1\times n_2} \to \mathbb{R}$ is a convex function, $\mathcal{A} : \mathbb{V}^{n_1\times n_2} \to \mathbb{R}^l$ is a linear operator, $b \in \mathbb{R}^l$ is a given vector and $K \subseteq \mathbb{R}^l$ is a closed convex set. Let $\widehat{X}$ be a given feasible point of (30) and $\widehat{z} := \mathcal{A}(\widehat{X}) - b$. When Ψ is differentiable at $\widehat{X}$, we say that the constraint nondegeneracy holds at $\widehat{X}$ if

$$\mathcal{A}\,\mathbb{V}^{n_1\times n_2} + \mathrm{lin}\big(\mathcal{T}_K(\widehat{z})\big) = \mathbb{R}^l, \tag{31}$$

where $\mathcal{T}_K(\widehat{z})$ denotes the tangent cone of K at $\widehat{z}$ and $\mathrm{lin}(\mathcal{T}_K(\widehat{z}))$ denotes the largest linearity space contained in $\mathcal{T}_K(\widehat{z})$, i.e., $\mathrm{lin}(\mathcal{T}_K(\widehat{z})) = \mathcal{T}_K(\widehat{z}) \cap (-\mathcal{T}_K(\widehat{z}))$. When the function Ψ is nondifferentiable, we can rewrite the optimization problem (30) equivalently as

$$\min_{(X,t)\in\mathbb{V}^{n_1\times n_2}\times\mathbb{R}} \Big\{\Phi(X) + t :\ \widetilde{\mathcal{A}}(X,t) \in K \times \mathrm{epi}\Psi\Big\},$$

where $\mathrm{epi}\Psi := \{(X,t) \in \mathbb{V}^{n_1\times n_2} \times \mathbb{R} \mid \Psi(X) \le t\}$ denotes the epigraph of Ψ and $\widetilde{\mathcal{A}} : \mathbb{V}^{n_1\times n_2} \times \mathbb{R} \to \mathbb{R}^l \times \mathbb{V}^{n_1\times n_2} \times \mathbb{R}$ is a linear operator defined by

$$\widetilde{\mathcal{A}}(X,t) := \begin{pmatrix} \mathcal{A}(X) - b \\ X \\ t \end{pmatrix}, \qquad (X,t) \in \mathbb{V}^{n_1\times n_2} \times \mathbb{R}.$$

From (31) and [64, Theorem 6.41], the constraint nondegeneracy holds at $(\widehat{X},\widehat{t})$ with $\widehat{t} = \Psi(\widehat{X})$ if

$$\widetilde{\mathcal{A}}\begin{pmatrix} \mathbb{V}^{n_1\times n_2} \\ \mathbb{R} \end{pmatrix} + \begin{pmatrix} \mathrm{lin}\big(\mathcal{T}_K(\widehat{X})\big) \\ \mathrm{lin}\big(\mathcal{T}_{\mathrm{epi}\Psi}(\widehat{X},\widehat{t})\big) \end{pmatrix} = \begin{pmatrix} \mathbb{R}^l \\ \mathbb{V}^{n_1\times n_2} \\ \mathbb{R} \end{pmatrix}.$$

By the definition of $\widetilde{\mathcal{A}}$, it is not difficult to verify that this condition is equivalent to

$$[\mathcal{A} \;\; 0]\big(\mathrm{lin}(\mathcal{T}_{\mathrm{epi}\Psi}(\widehat{X},\widehat{t}))\big) + \mathrm{lin}\big(\mathcal{T}_K(\widehat{X})\big) = \mathbb{R}^l. \tag{32}$$

By letting $\Psi = \|\cdot\|_*, \mathcal{A} = \mathcal{R}_\alpha$ and $K = \{0\}$, one can see that the problem (7) takes the form of (30). By the expression of $\mathcal{T}_{\mathrm{epi}\Psi}(\overline{X},\bar{t})$ with $\bar{t} = \|\overline{X}\|_*$ (e.g., see [32]), we see that for the problem (7), the condition (32) reduces to

$$\mathcal{R}_\alpha\big(\mathcal{T}(\overline{X})\big) = \mathbb{R}^{d_1}, \tag{33}$$

where

$$\mathcal{T}(\overline{X}) = \big\{H \in \mathbb{V}^{n_1\times n_2} \;\big|\; \overline{U}_2^{\mathbb{T}} H \overline{V}_2 = 0\big\}. \tag{34}$$

Hence, we say that the constraint nondegeneracy holds at $\overline{X}$ to the problem (7) if the condition (33) holds. By letting $\Psi = \delta_{\mathbb{S}^n_+}, \mathcal{A} = \mathcal{R}_\alpha$ and $K = \{0\}$, we can see that the problem (8) takes the form of (30) , and now that the condition (32) reduces to

$$\mathcal{R}_\alpha\big(\mathrm{lin}(\mathcal{T}_{\mathbb{S}^n_+}(\overline{X}))\big) = \mathbb{R}^{d_1}. \tag{35}$$

Thus, we say that the constraint nondegeneracy holds at $\overline{X}$ to the problem (8) if the condition (35) holds. From Arnold's characterization of the tangent cone $\mathcal{T}_{\mathbb{S}^n_+}(\overline{X}) = \big\{H \in \mathbb{S}^n \mid \overline{P}_2^{\mathbb{T}} H \overline{P}_2 \in \mathbb{S}^{n-r}_+\big\}$ in [2], we can write the linearity space $\mathrm{lin}(\mathcal{T}_{\mathbb{S}^n_+}(\overline{X}))$ explicitly as

$$\mathrm{lin}(\mathcal{T}_{\mathbb{S}^n_+}(\overline{X})) = \big\{H \in \mathbb{S}^n \;\big|\; \overline{P}_2^{\mathbb{T}} H \overline{P}_2 = 0\big\}.$$

Interestingly, for some special matrix completion problems, the constraint nondegeneracy automatically hold at $\overline{X}$, as stated in the following proposition.

Proposition 21 *For the following matrix completion problems:*

(i) *the covariance matrix completion with partial positive diagonal entries being fixed , in particular, the correlation matrix completion with all diagonal entries being fixed as ones;*

(ii) *the density matrix completion with its trace being fixed as one,*

the constraint nondegeneracy (35) holds at $\overline{X}$.

Next, we take a closer look at the solutions to the linear systems (25) and (28). Define linear operators $\mathcal{B}_1 : \mathbb{V}^{r\times r} \to \mathbb{V}^{(n_1-r)\times(n_2-r)}$ and $\mathcal{B}_2 : \mathbb{V}^{(n_1-r)\times(n_2-r)} \to \mathbb{V}^{(n_1-r)\times(n_2-r)}$ associated with $\overline{X}$, respectively, by

$$\mathcal{B}_1(Y) := \overline{U}_2^{\mathbb{T}} \mathcal{Q}_\beta^\dagger(\overline{U}_1 Y \overline{V}_1^T)\overline{V}_2 \;\; \text{and} \;\; \mathcal{B}_2(Z) := \overline{U}_2^{\mathbb{T}} \mathcal{Q}_\beta^\dagger(\overline{U}_2 Z \overline{V}_2^T)\overline{V}_2, \tag{36}$$

where $Y \in \mathbb{V}^{r\times r}$ and $Z \in \mathbb{V}^{(n_1-r)\times(n_2-r)}$. Note that the operator $\mathcal{B}_2$ is self-adjoint and positive semidefinite according to the definition of $\mathcal{Q}_\beta^\dagger$. Let $\widehat{g}(\overline{X})$ be the vector in $\mathbb{R}^r$ defined by

$$\widehat{g}(\overline{X}) := \big(1 - f_1(\sigma(\overline{X})), \ldots, 1 - f_r(\sigma(\overline{X}))\big)^{\mathbb{T}}. \tag{37}$$

Then, by the definition of the spectral operator F, we can rewrite (25) in the following concise form

$$\mathcal{B}_2(\Gamma) = \mathcal{B}_1\big(\mathrm{Diag}(\widehat{g}(\overline{X}))\big), \quad \Gamma \in \mathbb{V}^{(n_1-r)\times(n_1-r)}. \tag{38}$$

For the positive semidefinite case $\mathbb{V}^{n_1\times n_2} = \mathbb{S}^n$ and $\overline{X} \in \mathbb{S}^n_+$, both $\overline{U}_i$ and $\overline{V}_i$ reduce to $\overline{P}_i$ for $i = 1, 2$. In this case, the linear system (28) can be concisely written as

$$\mathcal{B}_2(\Lambda) = \mathcal{B}_2(I_{n-r}) + \mathcal{B}_1\big(\mathrm{Diag}(\widehat{g}(\overline{X}))\big), \quad \Lambda \in \mathbb{S}^{n-r}. \tag{39}$$

Proposition 22 *For the rectangular case, if the constraint nondegeneracy (33) holds at $\overline{X}$ to the problem (7), then the linear operators $\mathcal{B}_2$ defined by (36) is self-adjoint and positive definite. Meanwhile, for the positive semidefinite case, if the constraint nondegeneracy (35) holds at $\overline{X}$ to the problem (8), then the linear operators $\mathcal{B}_2$ is also self-adjoint and positive definite.*

According to Proposition 22, the constraint nondegeneracy at $\overline{X}$ to the problem (7) and (8), respectively, implies that the linear system (25) has a unique solution $\widehat{\Gamma} = \mathcal{B}_2^{-1}\mathcal{B}_1\big(\mathrm{Diag}(\widehat{g}(\overline{X}))\big)$ and the linear system (28) has a unique solution $\widehat{\Lambda} = I_{n-r} + \mathcal{B}_2^{-1}\mathcal{B}_1\big(\mathrm{Diag}(\widehat{g}(\overline{X}))\big)$. Then, from Proposition 22 together with Theorems 16 and 20, we can obtain the following main result for rank consistency.

Theorem 23 *Suppose that $\rho_m \to 0$, $\sqrt{m}\rho_m \to \infty$ and $\gamma_m = O_p(1)$. For the rectangular case, if the constraint nondegeneracy (33) holds at $\overline{X}$ to the problem (7) and*

$$\big\|\mathcal{B}_2^{-1}\mathcal{B}_1\big(\mathrm{Diag}(\widehat{g}(\overline{X}))\big)\big\| < 1, \tag{40}$$

then the estimator $\widehat{X}_m$ generated from the rank-correction step (7) is rank consistent. For the positive semidefinite case, if the constraint nondegeneracy (35) holds at $\overline{X}$ to the problem (8) and

$$I_{n-r} + \mathcal{B}_2^{-1}\mathcal{B}_1\big(\mathrm{Diag}(g_r(\overline{X}))\big) \in \mathbb{S}^{n-r}_{++}, \tag{41}$$

then the estimator $\widehat{X}_m$ generated from the rank-correction step (8) is rank consistent.

From Theorem 23, it is not difficult to see that there exists some threshold $\bar{\varepsilon} > 0$ (depending on $\overline{X}$) such that the condition (40) holds if $|1 - f_i(\sigma(\overline{X}))| \le \bar{\varepsilon}\ \forall 1 \le i \le r$. In other words, when $F(\overline{X})$ is sufficiently close to $\overline{U}_1\overline{V}_1^{\mathbb{T}}$, the condition (40) holds automatically and so does the rank consistency. Thus, Theorem 23 provides us a guideline to construct a suitable rank-correction function for rank consistency. This is another important aspect of what we can benefit from the rank-correction step, besides the reduction of recovery error discussed in Section 3.

The next theorem shows that for the covariance (correlation) and density matrix completion problems with fixed basis coefficients described in Proposition 21, if observations are sampled uniformly at random, the rank consistency can be guaranteed for a broad class of rank-correction functions F.

Theorem 24 *For the covariance (correlation) and density matrix completion problems defined in Proposition 21 under uniform sampling, if* $\rho_m \to 0$, $\sqrt{m}\rho_m \to \infty$, $\gamma_m = O_p(1)$ *and* F *is a spectral operator associated with a symmetric function* $f : \mathbb{R}^n \to \mathbb{R}^n$ *such that*

$$\begin{cases} f_i(x) > 0 & \text{if } x_i > 0, \\ f_i(x) = 0 & \text{if } x_i = 0, \end{cases} \qquad \forall\, x \in \mathbb{R}^n_+ \text{ and } \forall\, i = 1, \ldots, n, \tag{42}$$

then the estimator $\widehat{X}_m$ *generated from the rank-correction step is rank consistent.*

5 Construction of the rank-correction function

In this section, we focus on the construction of a suitable rank-correction function F based on the results obtained in Sections 3 and 4. As can be seen from Theorem 6, a smaller value of a_m/b_m potentially leads to a smaller recovery error. Thus, we desire a construction of the rank-correction function such that $F(\widetilde{X}_m)$ is close to $\overline{U}_1\overline{V}_1^{\mathbb{T}}$. Meanwhile, according to Theorem 23, we also desire that $F(\overline{X})$ is close to $\overline{U}_1\overline{V}_1^{\mathbb{T}}$ for rank consistency. Notice that a reasonable initial estimator $\widetilde{X}_m$ should not deviate too much from the true matrix $\overline{X}$. Therefore, the above two criteria consistently suggest a natural idea to construct a rank-correction function F, if possible, such that

$$F(X) \to \overline{U}_1\overline{V}_1^{\mathbb{T}} \quad \text{as} \quad X \to \overline{X}. \tag{43}$$

Next, we proceed the construction of the rank-correction function F for the rectangular case. For the positive semidefinite case, one may just replace the singular value decomposition with the eigenvalue decomposition and conduct exactly the same analysis.

5.1 The rank is known

If the rank of the true matrix $\overline{X}$ is known in advance, we construct the rank-correction function F by

$$F(X) := U_1V_1^{\mathbb{T}}, \tag{44}$$

where $(U, V) \in \mathbb{O}^{n_1,n_2}(X)$ and $X \in \mathbb{V}^{n_1\times n_2}$. Note that F defined by (44) is not a spectral operator over the whole space of $\mathbb{V}^{n_1\times n_2}$, but in a neighborhood of $\overline{X}$ it is indeed a spectral operator and is actually twice continuously differentiable (see, e.g., [11, Proposition 8]). Hence, it satisfies the criterion (43). With this rank-correction function, the rank-correction step is essentially the same as one step of the majorized penalty method developed in [25]. By Theorem 23, we immediately obtain the following result.

Corollary 25 *Suppose that the rank of the true matrix* $\overline{X}$ *is known and the constraint nondegeneracy holds at* $\overline{X}$. *If* $\rho_m \to 0$, $\sqrt{m}\rho_m \to \infty$, $\gamma_m = O_p(1)$ *and* F *is chosen by (44), then the estimator* $\widehat{X}_m$ *generated from the rank-correction step is rank consistent.*

5.2 The rank is unknown

If the rank of the true matrix $\overline{X}$ is unknown, then the rank-correction function F cannot be defined by (44). What we will do is to construct a spectral operator F to imitate the case when the rank is known. Here, we propose F to be a spectral operator

$$F(X) := U\mathrm{Diag}\big(f(\sigma(X))\big)V^{\mathbb{T}} \tag{45}$$

associated with the symmetric function $f : \mathbb{R}^n \to \mathbb{R}^n$ defined by

$$f_i(x) = \begin{cases} \phi\left(\dfrac{x_i}{\|x\|_\infty}\right) & \text{if } x \in \mathbb{R}^n\backslash\{0\}, \\ 0 & \text{if } x = 0, \end{cases} \tag{46}$$

where $(U, V) \in \mathbb{O}^{n_1,n_2}(X)$, $X \in \mathbb{V}^{n_1\times n_2}$, and the scalar function $\phi : \mathbb{R} \to \mathbb{R}$ takes the form

$$\phi(t) := \mathrm{sgn}(t)(1+\varepsilon^\tau)\frac{|t|^\tau}{|t|^\tau+\varepsilon^\tau}, \quad t \in \mathbb{R}, \tag{47}$$

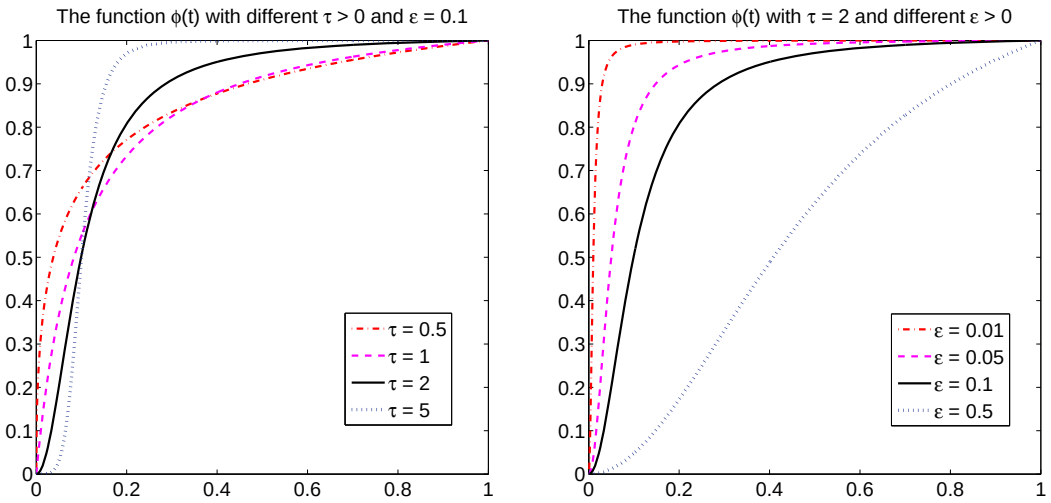
for some $\tau > 0$ and $\varepsilon > 0$. By noting that for each t, $\phi(t) \to \mathrm{sgn}(t)$ as $\varepsilon \downarrow 0$, we directly obtain the following result.

Corollary 26 *Suppose that the constraint nondegeneracy holds at $\overline{X}$. If $\rho_m \to 0$, $\sqrt{m}\rho_m \to \infty$, $\gamma_m = O_p(1)$, then for any given $\tau > 0$, there exists some $\bar{\varepsilon} > 0$ such that for any F defined by (45), (46) and (47) with $0 < \varepsilon \le \bar{\varepsilon}$, the estimator $\widehat{X}_m$ generated from the rank-correction step is rank consistent.*

Corollary 26 indicates that one needs to choose a small $\varepsilon > 0$ in pursuit of rank consistency. Meanwhile, we also need to take care of the influence of a small $\varepsilon > 0$ on the recovery error bound which depends on the value of a_m/b_m. Certainly, we desire $a_m \approx 0$ and $b_m \approx 1$. This motivates us to choose a function ϕ, if possible, such that

$$\phi\left(\frac{\sigma_i(\widetilde{X}_m)}{\sigma_1(\widetilde{X}_m)}\right) \approx \begin{cases} 1 & \text{if } 1 \le i \le r, \\ 0 & \text{if } r+1 \le i \le n. \end{cases} \tag{48}$$

This is also why we normalize the function ϕ defined by (47) in the interval $t \in [0, 1]$ such that $\phi(0) = 0$ and $\phi(1) = 1$. However, as indicated by Corollary 11, the initial estimator $\widetilde{X}_m$ is very possible to have a higher rank than $\overline{X}$ when it approaches to $\overline{X}$. It turns out that when $\varepsilon > 0$ is tiny, $\phi\big(\sigma_i(\widetilde{X}_m)/\sigma_1(\widetilde{X}_m)\big) \approx 1$ for $r+1 \le i \le \mathrm{rank}(\widetilde{X}_m)$, which violates our desired property (48). As a result, $\varepsilon > 0$ should be chosen to be small but balanced. Notice that $\phi(\varepsilon) = (1+\varepsilon^\tau)/2 \approx 1/2$ if $\varepsilon > 0$ is small and $\tau > 0$ is not too small. Thus, the value of ε can be regarded as a divide of confidence on whether $\sigma_i(\widetilde{X}_m)$ is believed to come from a nonzero singular values of $\overline{X}$ with perturbation — positive confidence if $\sigma_i(\widetilde{X}_m) > \varepsilon\sigma_1(\widetilde{X}_m)$ and negative confidence if $\sigma_i(\widetilde{X}_m) < \varepsilon\sigma_1(\widetilde{X}_m)$. On the other hand, the parameter $\tau > 0$ mainly controls the shape of the function ϕ over $t \in [0, 1]$. The function ϕ is concave if $0 < \tau \le 1$ and S-shaped with a single inflection



(a) $\varepsilon = 0.1$ with different $\tau > 0$ (b) $\tau = 2$ with different $\varepsilon > 0$

Figure 1: Shapes of the function ϕ with different $\tau > 0$ and $\varepsilon > 0$

point at $\left(\frac{\tau-1}{\tau+1}\right)^{1/\tau}\varepsilon$ if $\tau > 1$. Moreover, the steepness of the function ϕ increases when τ increases. In particular, if $0 < \varepsilon < 1$ and τ is very large, ϕ is very close to the step function taking the value 0 if $0 \le t < \varepsilon$ and the value 1 if $\varepsilon < t \le 1$. In this case, there exists some ε such that the desired property (48) can be achieved and that the corresponding rank-correction function F is very close to the one defined by (44). Thus, it seems to be a good idea to choose an S-shaped function ϕ with a large τ. However, in practice, the parameter ε should be pre-determined. Since $\mathrm{rank}(\overline{X})$ is unknown and the singular values of $\widetilde{X}_m$ are unpredictable, it is hard to choose a suitable ε in advance, and hence, it will be too risky to choose a large τ for recovery. As a result, one has to be somewhat conservative to choose τ, sacrificing some optimality of recovery in exchange for robustness strategically. If the initial estimator is generated from the nuclear norm penalized least squares problem, we recommend the choices $\tau = 1$ or 2 and $\varepsilon = 0.01 \sim 0.1$ as these choices show stable performance for plenty of problems, as validated in Section 6.

We also remark that for the positive semidefinite case, the rank-correction function defined by (45), (46) and (47) is related to the reweighted trace norm for the matrix rank minimization proposed by Fazel et al. [20, 54]. The reweighted trace norm in [20, 54] for the positive semidefinite case is $\langle (X^k + \varepsilon I_n)^{-1}, X\rangle$, which arises from the derivative of the surrogate function $\log\det(X + \varepsilon I_n)$ of the rank function at an iterate X^k, where ε is a small positive constant. Meanwhile, in our proposed rank-correction step, if we choose $\tau = 1$, then $I_n - \frac{1}{1+\varepsilon}F(\widetilde{X}_m) = \varepsilon'(\widetilde{X}_m + \varepsilon' I_n)^{-1}$ with $\varepsilon' = \varepsilon\|\widetilde{X}_m\|$. Superficially, similarity occurs; however, it is notable that ε' depends on $\widetilde{X}_m$, which is different from the constant ε in [20, 54]. More broadly speaking, the rank-correction function F defined by (45), (46)

and (47) is not a gradient of any real-valued function. This distinguishes our proposed rank-correction step from the reweighted trace norm minimization in [20, 54] even for the positive semidefinite case.

6 Numerical experiments

In this section, we validate the power of our proposed rank-corrected procedure on the recovery by applying it to the positive semidefinite matrix completion problems. In solving the optimization problem in the rank-correction step (8), we adopted the code developed by Jiang et al. [32] for large scale linearly constrained convex semidefinite programming problems. The implemented code is based on an inexact version of the accelerated proximal gradient method [56, 4]. All tests were run in MATLAB under Windows 7.0 operating system on an Intel Core(TM) i7-2720 QM 2.20GHz CPU with 8.00GB memory.

For convenience, in the sequel, the NNPLS estimator and the RCS estimator, respectively, stand for the estimators from the nuclear norm penalized least squares problem (i.e., the problem (8) with $F \equiv 0$ and $\gamma_m = 0$) and the rank-correction step. Let X_m be an estimator. The **relative error** (**relerr** for short) of X_m is defined by

$$\text{relerr} = \frac{\|X_m - \overline{X}\|_F}{\max(10^{-8}, \|\overline{X}\|_F)}.$$

6.1 Influence of fixed basis coefficients on the recovery

In this subsection, we take the correlation matrix completion for example to test the performance of the NNPLS estimator and the RCS estimator with different patterns of fixed basis coefficients. We randomly generated the true matrix $\overline{X}$ by the following command:

```
M = randn(n,r);  ML = weight*M(:,1:k);  M(:,1:k) = ML; Xtemp = M*M';
       D = diag(1./sqrt(diag(Xtemp))); X_bar = D*Gtemp*D
```

where the parameter `weight` is used to control the relative magnitude difference between the first `k` largest eigenvalues and the other nonzero eigenvalues. In our experiment, we set `weight` $= 5$ and `k` $= 1$, and took $\overline{X}$ = `X_bar` with dimension `n` $= 1000$ and rank `r` $= 5$. We randomly fixed partial diagonal and off-diagonal entries of $\overline{X}$ and sampled the rest entries uniformly at random with i.i.d. Gaussian noise at the noise level 10%.

In Figure 2, we plot the curves of the relative error and the rank of the NNPLS estimator and the RCS estimator with different patterns of fixed entries. In the captions of the subfigures, **diag** means the number of fixed diagonal entries and **non-diag** means the number of fixed off-diagonal entries. The subfigures on the left-hand side and the right-hand side show the performance of the NNPLS estimator and the RCS estimator, respectively. For the RCS estimator, the rank-correction function F is defined by (45),

(46) and (47) with $\tau = 2$ and $\varepsilon = 0.02$, and the initial $\widetilde{X}_m$ is chosen from those points of the corresponding subfigures on the left-hand side such that $\big|\|y - \mathcal{R}_\Omega(\widetilde{X}_m)\|_2/\|y\|_2 - 0.1\big|$ attains the smallest value.

From the subfigures on the left-hand side, we observe that as the number of fixed diagonal entries increases, the parameter ρ_m for the smallest recovery error deviates more and more from the one for attaining the true rank. In particular, when $\mathbf{diag} = n$, the NNPLS estimator reduces to the (constrained) least squares estimator so that one cannot benefit from the NNPLS estimator for encouraging a low-rank solution. This implies that the NNPLS estimator does not possess the rank consistency when some entries are fixed. However, the subfigures on the right-hand side indicate that the RCS estimator can yield a solution with the correct rank as well as a desired small recovery error simultaneously, with the parameter ρ_m in a large interval. This exactly validates the theoretical result of Theorem 24 for rank consistency.

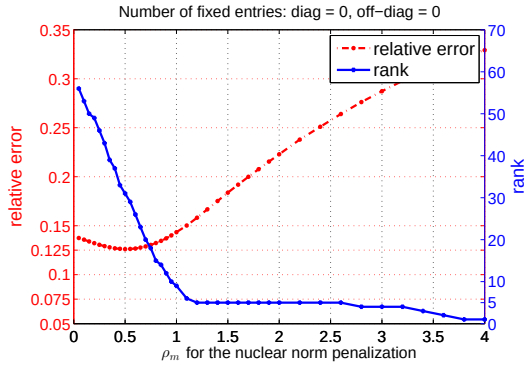
6.2 Performance of different rank-correction functions for recovery

In this subsection, we test the performance of different rank-correction functions for recovering a correlation matrix. We randomly generated the true matrix $\overline{X}$ by the command in Subsection 6.1 with $\mathtt{n} = 1000$, $\mathtt{r} = 10$, $\mathtt{weight} = 2$ and $\mathtt{k} = 5$. We fixed all the diagonal entries of $\overline{X}$ and sampled partial off-diagonal entries uniformly at random with i.i.d. Gaussian noise at the noise level 10%. We chose the (nuclear norm penalized) least squares estimator to be the initial estimator $\widetilde{X}_m$. In Figure 3, we plot four curves corresponding to the rank-correction functions F defined by (45), (46) and (47) with $\tau = 2$ and different ε, and another two curves corresponding to the rank-correction functions F defined by (44) at $\widetilde{X}_m$ (i.e., $\widetilde{U}_1\widetilde{V}_1^{\mathbb{T}}$) and $\overline{X}$ (i.e., $\overline{U}_1\overline{V}_1^{\mathbb{T}}$), respectively. The values of a_m, b_m and the optimal recovery error with different ρ_m are listed in Table 1.

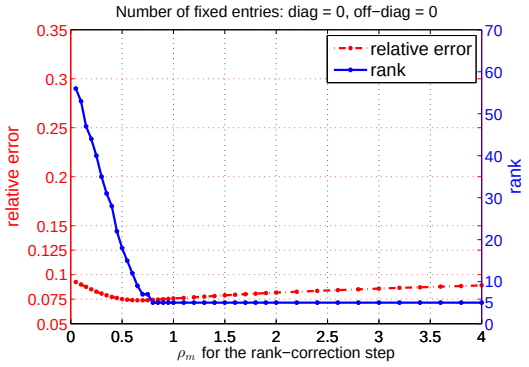
As can be seen from Figure 3, when ρ_m increases, the recovery error decreases with the rank and then increases after the correct rank is attained, except for the case $\overline{U}_1\overline{V}_1^{\mathbb{T}}$. This validates our discussion about the recovery error at the end of Section 3. Moreover, for a smaller ε, the curve of recovery error changes more gently, though a certain optimality in the sense of recovery error is sacrificed. This means that the choice of a relatively small ε, say 0.01 or 0.02, is more robust for those ill-conditioned problems. From Table 1, we see that a smaller a_m/b_m corresponds to a better optimal recovery error. It is worthwhile to point out that, even if a_m/b_m is larger than 1, the performance of the RCS estimator for recovery is still much better than that of the NNPLS estimator.

6.3 Performance for different matrix completion problems

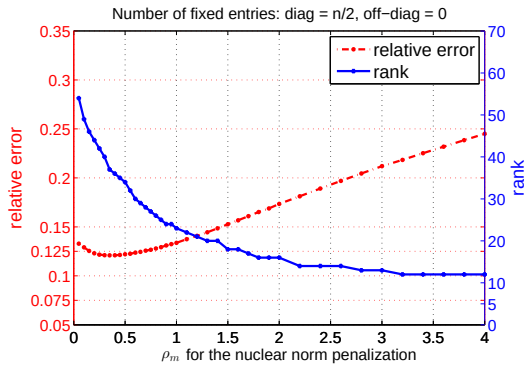
In this subsection, we test the performance of the RCS estimator for the covariance and density matrix completion problems. As can be seen from Figure 2, a good choice of the parameter ρ_m for the RCS estimator could be the smallest one such that the rank becomes stable. Such a parameter ρ_m can be found by the bisection search method. This is actually what we benefit from rank consistency. In the following numerical experiments,



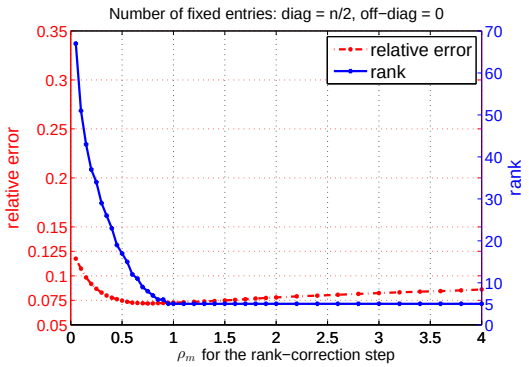
(a) Nuclear norm: diag =0, off-diag = 0



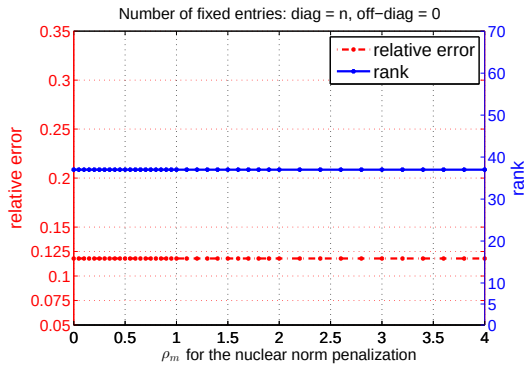
(b) Rank correction step: diag=0, off-diag=0



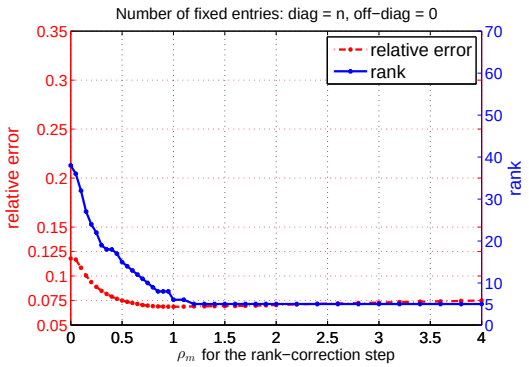
(c) Nuclear norm: diag =n/2, off-diag = 0



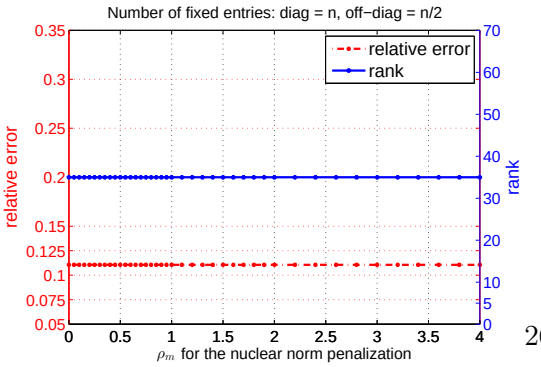
(d) Rank correction step: diag=n/2, off-diag=0



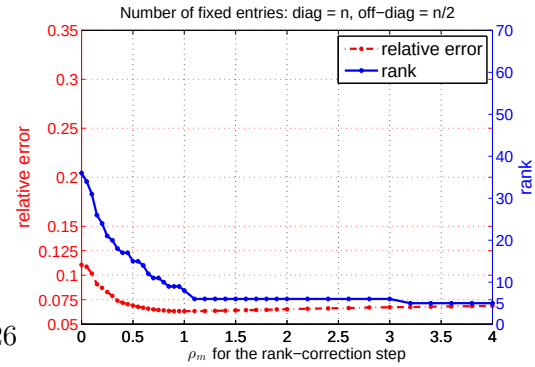
(e) Nuclear norm: diag =n, off-diag = 0



(f) Rank correction step: diag=n, off-diag=0



(g) Nuclear norm: diag =n, off-diag = n/2



(h) Rank correction step: diag=n, off-diag=n/2

Figure 2: Influence of fixed basis coefficients on recovery (sample ratio = 6.38%)

Table 1: Influence of the rank-correction term on the recovery error

rank-correction function	a_m	b_m	a_m/b_m	optimal relerr
zero function	1	1	1	10.85%
$\varepsilon = 0.01, \tau = 2$	0.1420	0.2351	0.6038	5.96%
$\varepsilon = 0.02, \tau = 2$	0.1459	0.5514	0.2646	5.80%
$\varepsilon = 0.05, \tau = 2$	0.1648	0.8846	0.1863	5.75%
$\varepsilon = 0.1, \tau = 2$	0.2399	0.9681	0.2478	5.77%
$\widetilde{U}_1\widetilde{V}_1^{\mathbb{T}}$ (initial)	0.1445	0.9815	0.1472	5.75%
$\overline{U}_1\overline{V}_1^{\mathbb{T}}$ (true)	0	1	0	2.25%

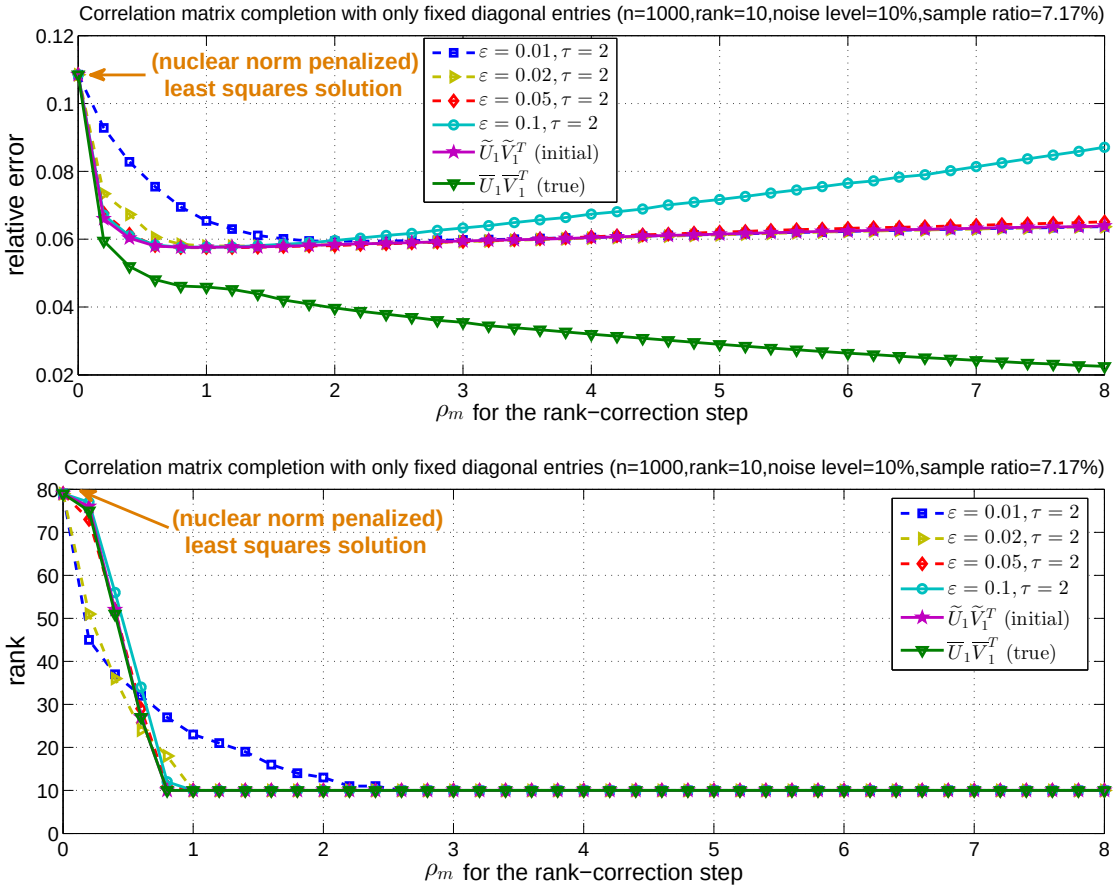


Figure 3: Influence of the rank-correction term on the recovery

we apply the above strategy to find a suitable ρ_m for the RCS estimator, and choose the rank-correction function F defined by (45), (46) and (47) with $\tau = 2$ and $\varepsilon = 0.02$.

We first take the covariance matrix completion for example to test the performance of the RCS estimator with different initial estimators $\widetilde{X}_m$. The true matrix $\overline{X}$ is generated by the command in Subsection 6.1 with `n` $= 500$, `r` $= 5$, `weight` $= 3$ and `k` $= 1$ except that `D = eye(n)`. We depict the numerical results in Figure 4, where the dash curves represent the relative recovery error and the rank of the NNPLS estimator with different ρ_m, and the solid curves represent the relative recovery error and the rank of the RCS estimator with $\widetilde{X}_m$ chosen to be the corresponding NNPLS estimator. As can be seen from Figure 4, the RCS estimator substantially improves the quality of the NNPLS estimator in terms of both the recovery error and the rank. We also observe that when the initial $\widetilde{X}_m$ has a large deviation from the true matrix, the quality of the RCS estimator may still not be satisfied. Thus, it is natural to ask whether further rank-correction steps could improve the quality. The answer can be found from Table 2 below, where the numerical results of the covariance matrix completion are reported. We also report the numerical results of the density matrix completion in Table 3.

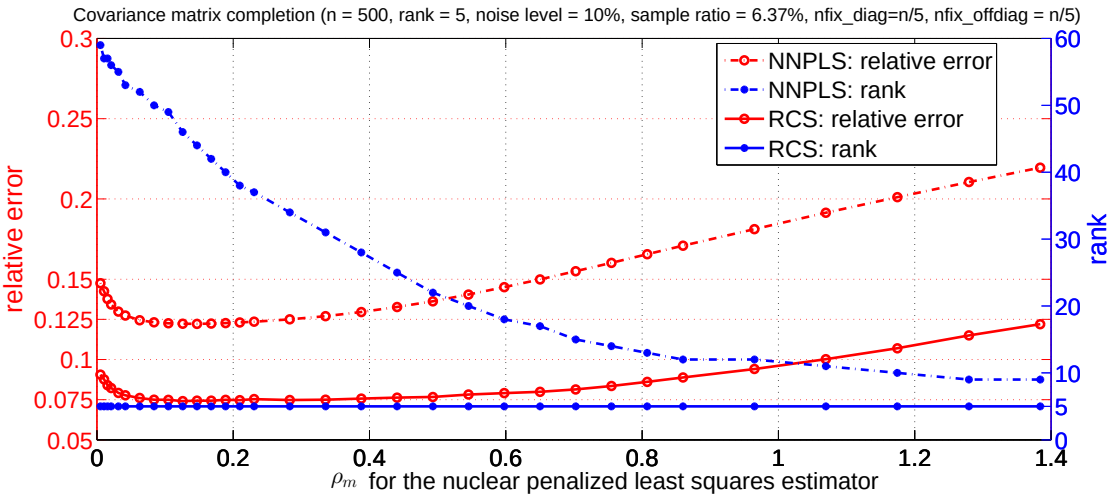


Figure 4: Performance of the RCS estimator with different initial $\widetilde{X}_m$

For the covariance matrix completion problems, we generated the true matrix $\overline{X}$ by the command in Subsection 6.1 with `n` $= 1000$, `weight` $= 3$ and `k` $= 1$ except that `D = eye(n)`. The rank of $\overline{X}$ and the number of fixed diagonal and non-diagonal entries of $\overline{X}$ are reported in the first and the second columns of Table 2, respectively. We sampled partial off-diagonal entries uniformly at random with i.i.d. Gaussian noise at the noise level 10%. The first RCS estimator is using the NNPLS estimator as the initial estimator $\widetilde{X}_m$, and the second (third) RCS estimator is using the first (second) RCS estimator as the initial estimator $\widetilde{X}_m$. From Table 2, we see that when the sample ratio is reasonable, one rank-correction step is enough to yield a desired result. Meanwhile, when the sample

Table 2: Performance for covariance matrix completion problems with $n = 1000$

r	diag/ off-diag	sample ratio	NNPLS	1st RCS	2st RCS	3rd RCS
			relerr (rank)	relerr(rank)	relerr (rank)	relerr (rank)
5	1000/0	2.40%	1.95e-1 (47)	1.27e-1 (5)	1.18e-1 (5)	1.12e-1 (5)
	1000/0	7.99%	6.10e-2 (51)	3.41e-2 (5)	3.37e-2 (5)	3.36e-2 (5)
	500/50	2.39%	2.01e-1 (45)	1.10e-1 (5)	9.47e-2 (5)	8.97e-2 (5)
	500/50	7.98%	7.19e-2 (32)	3.77e-2 (5)	3.59e-2 (5)	3.58e-2 (5)
10	1000/0	5.38%	1.32e-1 (74)	7.68e-2 (10)	7.39e-2 (10)	7.36e-2 (10)
	1000/0	8.96%	9.18e-2 (78)	5.15e-2 (10)	5.08e-2 (10)	5.08e-2 (10)
	500/100	5.37%	1.58e-1 (57)	8.66e-2 (10)	7.74e-2 (10)	7.60e-2 (10)
	500/100	8.96%	1.02e-1 (49)	5.36e-2 (10)	5.24e-2 (10)	5.25e-2 (10)

Table 3: Performance for density matrix completion problems with $n = 1024$

noise	r	noise level	sample ratio	NNPLS1			NNPLS2			RCS		
				fidelity	relerr	rank	fidelity	relerr	rank	fidelity	relerr	rank
statistical	3	10.0%	1.5%	0.697	2.59e-1	3	0.955	2.50e-1	3	0.987	1.02e-1	3
		10.0%	4.0%	0.915	8.04e-2	3	0.997	6.84e-2	3	0.998	4.13e-2	3
	5	10.0%	2.0%	0.550	3.71e-1	5	0.908	4.23e-1	5	0.972	1.61e-1	5
		10.0%	5.0%	0.889	1.03e-1	5	0.995	9.18e-2	5	0.997	4.91e-2	5
mixed	3	12.4%	1.5%	0.654	2.93e-1	3	0.957	2.43e-1	3	0.988	1.06e-1	3
		12.4%	4.0%	0.832	1.49e-1	3	0.995	8.14e-2	3	0.997	6.41e-2	3
	5	12.4%	2.0%	0.521	3.95e-1	5	0.912	4.09e-1	5	0.977	1.51e-1	5
		12.5%	5.0%	0.817	1.61e-1	5	0.987	1.01e-1	5	0.996	7.09e-2	5

ratio is very low, especially if some off-diagonal entries are further fixed, one or two more rank-correction steps can still improve the quality of estimation.

For the density matrix completion problems, we generated the true density matrix $\overline{X}$ by the following command:

```
M = randn(n,r)+i*randn(n,r);  ML = weight*M(:,1:k);  M(:,1:k) = ML;
          Xtemp = M*M';  X_bar = Xtemp/sum(diag((Xtemp))).
```

During the testing, we set `n` $= 1024$, `weight` $= 2$ and `k` $= 1$, and sampled partial Pauli measurements except the trace of $\overline{X}$ uniformly at random with i.i.d. Gaussian noise at the noise level 10%. Besides the above statistical noise, we further added the depolarizing noise, which frequently appears in quantum systems, with strength 0.01. This case is labeled as the mixed noise in the last four rows of Table 3. We remark here that the depolarizing noise differs from our assumption on noise since it does not have randomness. One may refer to [29, 21] for details of the quantum depolarizing channel. In Table 3, the (squared) **fidelity** is a measure of the closeness of two quantum states, defined by $\|\widehat{X}_m^{1/2}\overline{X}^{1/2}\|_*^2$, the NNPLS1 estimator means the NNPLS estimator by dropping the trace one constraint, and the NNPLS2 estimator means the one obtained by normalizing the NNPLS1 estimator to be of trace one. Note that the NNPLS2 estimator was ever used by Flammia et al. [21]. Table 3 shows that the RCS estimator is superior to the NNPLS2 estimator in terms of both the fidelity and the relative error.

7 Conclusions

In this paper, we proposed a rank-corrected procedure for low-rank matrix completion problems with fixed basis coefficients. This approach can substantially overcome the limitation of the nuclear norm penalization for recovering a low-rank matrix. We studied the impact of adding the rank-correction term on both the reduction of the recovery error bounds and the rank consistency (in the sense of Bach [3]). Due to the presence of fixed basis coefficients, constraint nondegeneracy plays an important role in our analysis. Extensive numerical experiments show that our approach can significantly improve the recovery performance in the sense of both the recovery error and the rank, compared with the nuclear norm penalized least square estimator. As a byproduct, our results also provide a theoretical foundation for the majorized penalty method of Gao and Sun [25] and Gao [24] for structured low-rank matrix optimization problems.

Our proposed rank-correction step also allows additional constraints according to other possible prior information. In particular, for additional linear constraints, all the theoretical results in this paper hold with slight modifications. In order to better fit the under-sampling setting of matrix completion, in the future work, it would be of great interest to extend the asymptotic rank consistency results to the case that the matrix size is allowed to grow. It would also be interesting to extend this approach to deal with other low-rank matrix problems.

Acknowledgments

The authors would like to thank Dr. Kaifeng Jiang for helpful discussions on using the accelerated proximal gradient method to solve the density matrix completion problem.

Appendix

Proof of Theorem 1

Let $\Delta_m := \widehat{X}_m - \overline{X}$. Since $\widehat{X}_m$ is optimal to (7) and $\overline{X}$ is feasible to (7), it follows that

$$\begin{aligned}\frac{1}{2m}\|\mathcal{R}_\Omega(\Delta_m)\|_2^2 \le \Big\langle \frac{\nu}{m}\mathcal{R}_\Omega^*(\xi), \Delta_m\Big\rangle - \rho_m\big(\|\widehat{X}_m\|_* - \|\overline{X}\|_* - \langle F(\widetilde{X}_m) + \gamma_m\widetilde{X}_m, \Delta_m\rangle\big) \\ + \frac{\rho_m\gamma_m}{2}\big(\|\overline{X}\|_F^2 - \|\widehat{X}_m\|_F^2\big).\end{aligned} \tag{49}$$

Then, it follows from (13) that

$$\begin{aligned}\Big\langle \frac{\nu}{m}\mathcal{R}_\Omega^*(\xi), \Delta_m\Big\rangle &\le \nu\Big\|\frac{1}{m}\mathcal{R}_\Omega^*(\xi)\Big\|\big(\|\mathcal{P}_T(\Delta_m)\|_* + \|\mathcal{P}_{T^\perp}(\Delta_m)\|_*\big) \\ &\le \frac{\rho_m b_m}{\kappa}\big(\|\mathcal{P}_T(\Delta_m)\|_* + \|\mathcal{P}_{T^\perp}(\Delta_m)\|_*\big).\end{aligned} \tag{50}$$

From the directional derivative of the nuclear norm at $\overline{X}$ (see [72, Theorem 1]), we have

$$\|\widehat{X}_m\|_* - \|\overline{X}\|_* \ge \langle \overline{U}_1\overline{V}_1^{\mathbb{T}}, \Delta_m\rangle + \|\overline{U}_2^{\mathbb{T}}\Delta_m\overline{V}_2\|_*.$$

This, together with equations (11) and (12), implies that

$$\begin{aligned}&\|\widehat{X}_m\|_* - \|\overline{X}\|_* - \langle F(\widetilde{X}_m)+\gamma_m\widetilde{X}_m, \Delta_m\rangle \\ &\ge \langle \overline{U}_1\overline{V}_1^{\mathbb{T}}, \Delta_m\rangle + \|\overline{U}_2^{\mathbb{T}}\Delta_m\overline{V}_2\|_* - \langle F(\widetilde{X}_m)+\gamma_m\widetilde{X}_m, \Delta_m\rangle \\ &= \langle \overline{U}_1\overline{V}_1^{\mathbb{T}} - \mathcal{P}_T(F(\widetilde{X}_m)+\gamma_m\widetilde{X}_m), \Delta_m\rangle + \|\mathcal{P}_{T^\perp}(\Delta_m)\|_* - \langle \mathcal{P}_{T^\perp}(F(\widetilde{X}_m)+\gamma_m\widetilde{X}_m), \Delta_m\rangle \\ &= \langle \overline{U}_1\overline{V}_1^{\mathbb{T}} - \mathcal{P}_T(F(\widetilde{X}_m)+\gamma_m\widetilde{X}_m), \mathcal{P}_T(\Delta_m)\rangle + \|\mathcal{P}_{T^\perp}(\Delta_m)\|_* - \langle \mathcal{P}_{T^\perp}(F(\widetilde{X}_m)+\gamma_m\widetilde{X}_m), \mathcal{P}_{T^\perp}(\Delta_m)\rangle \\ &\ge -\|\overline{U}_1\overline{V}_1^{\mathbb{T}} - \mathcal{P}_T(F(\widetilde{X}_m)+\gamma_m\widetilde{X}_m)\|\|\mathcal{P}_T(\Delta_m)\|_* + \big(1 - \|\mathcal{P}_{T^\perp}(F(\widetilde{X}_m+\gamma_m\widetilde{X}_m)\|\big)\|\mathcal{P}_{T^\perp}(\Delta_m)\|_* \\ &= -a_m\|\mathcal{P}_T(\Delta_m)\|_* + b_m\|\mathcal{P}_{T^\perp}(\Delta_m)\|_*.\end{aligned} \tag{51}$$

By substituting (51) and (50) into (49), we obtain that

$$\begin{aligned}\frac{1}{2m}\|\mathcal{R}_\Omega(\Delta_m)\|_2^2 \le \rho_m\left(\Big(a_m + \frac{b_m}{\kappa}\Big)\|\mathcal{P}_T(\Delta_m)\|_* - \frac{\kappa-1}{\kappa}b_m\|\mathcal{P}_{T^\perp}(\Delta_m)\|_*\right) \\ + \frac{\rho_m\gamma_m}{2}(\|\overline{X}\|_F^2 - \|\widehat{X}_m\|_F^2).\end{aligned} \tag{52}$$

Note that $\mathrm{rank}(\mathcal{P}_T(\Delta_m)) \le 2r$. Hence, $\|\mathcal{P}_T(\Delta_m)\|_* \le \sqrt{2r}\|\mathcal{P}_T(\Delta_m)\|_F \le \sqrt{2r}\|\Delta_m\|_F$, and the desired result follows from (52). Thus, we complete the proof.

Proof of Lemma 2.
The proof is similar to that of [38, Lemma 12]. We need to show that the event

$$E = \left\{ \exists\, \Delta \in K(r) \text{ such that } \left| \frac{1}{m}\|\mathcal{R}_\Omega(\Delta)\|_2^2 - \langle \mathcal{Q}_\beta(\Delta), \Delta\rangle \right| \geq \frac{1}{2}\langle \mathcal{Q}_\beta(\Delta), \Delta\rangle + 128\mu_1 d_2 r \vartheta_m^2 \right\}$$

occurs with probability less than $2/(n_1+n_2)$. For any given $\varepsilon > 0$, we decompose $K(r)$ as

$$K(r) = \bigcup_{k=1}^{\infty} \left\{ \Delta \in K(r) \mid 2^{k-1}\varepsilon \leq \langle \mathcal{Q}_\beta(\Delta), \Delta\rangle \leq 2^k \varepsilon \right\}.$$

For any $a > 0$, let $K(r,a) := \{\Delta \in K(r) \mid \langle \mathcal{Q}_\beta(\Delta), \Delta\rangle \leq a\}$. Then we get $E \subseteq \bigcup_{k=1}^{\infty} E_k$ with

$$E_k = \left\{ \exists\, \Delta \in K(r, 2^k\varepsilon) \text{ such that } \left| \frac{1}{m}\|\mathcal{R}_\Omega(\Delta)\|_2^2 - \langle \mathcal{Q}_\beta(\Delta), \Delta\rangle \right| \geq 2^{k-2}\varepsilon + 128\mu_1 d_2 r \vartheta_m^2 \right\}.$$

Then, we need to estimate the probability of each event E_k. Define

$$Z_a := \sup_{\Delta \in K(r,a)} \left| \frac{1}{m}\|\mathcal{R}_\Omega(\Delta)\|_2^2 - \langle \mathcal{Q}_\beta(\Delta), \Delta\rangle \right|.$$

Notice that for any $\Delta \in \mathbb{V}^{n_1 \times n_2}$,

$$\frac{1}{m}\|\mathcal{R}_\Omega(\Delta)\|_2^2 = \frac{1}{m}\sum_{i=1}^{m} \langle \Theta_{\omega_i}, \Delta\rangle^2 \overset{a.s.}{\to} \mathbb{E}(\langle \Theta_{\omega_i}, \Delta\rangle^2) = \langle \mathcal{Q}_\beta(\Delta), \Delta\rangle.$$

Since $\|\mathcal{R}_\beta(\Delta)\|_\infty \leq 1$ for all $\Delta \in K(r)$, from Massart's Hoeffding type concentration inequality [48, Theorem 9] for suprema of empirical processes, we have

$$\Pr\left(Z_a \geq \mathbb{E}(Z_a) + t\right) \leq \exp(-mt^2/2) \quad \forall\, t > 0. \tag{53}$$

Next, we use the standard Rademacher symmetrization in the theory of empirical processes to further derive an upper bound of $\mathbb{E}(Z_a)$. Let $\{\epsilon_1, \ldots, \epsilon_m\}$ be a Rademacher sequence. Then, we have

$$\begin{aligned}
\mathbb{E}(Z_a) &= \mathbb{E}\left(\sup_{\Delta \in K(r,a)} \left| \frac{1}{m}\sum_{i=1}^{m} \langle \Theta_{\omega_i}, \Delta\rangle^2 - \mathbb{E}\big(\langle \Theta_{\omega_i}, \Delta\rangle^2\big) \right| \right) \\
&\leq 2\mathbb{E}\left(\sup_{\Delta \in K(r,a)} \left| \frac{1}{m}\sum_{i=1}^{m} \epsilon_i \langle \Theta_{\omega_i}, \Delta\rangle^2 \right| \right) \leq 8\mathbb{E}\left(\sup_{\Delta \in K(r,a)} \left| \frac{1}{m}\sum_{i=1}^{m} \epsilon_i \langle \Theta_{\omega_i}, \Delta\rangle \right| \right) \\
&= 8\mathbb{E}\left(\sup_{\Delta \in K(r,a)} \left| \frac{1}{m}\langle \mathcal{R}_\Omega^*(\epsilon), \Delta\rangle \right| \right) \leq 8\mathbb{E}\left\| \frac{1}{m}\mathcal{R}_\Omega^*(\epsilon) \right\| \left(\sup_{\Delta \in K(r,a)} \|\Delta\|_* \right),
\end{aligned} \tag{54}$$

where the first inequality follows from the symmetrization theorem (e.g., see [70, Lemma 2.3.1] and [6, Theorem 14.3]) and the second inequality follows from the contraction

theorem (e.g., see [43, Theorem 4.12] and [6, Theorem 14.4]). Moreover, from (16), we have

$$\|\Delta\|_* \le \sqrt{r}\|\Delta\|_F \le \sqrt{\mu_1 r d_2 \langle \mathcal{Q}_\beta(\Delta), \Delta\rangle} \le \sqrt{\mu_1 r d_2 a} \quad \forall\, \Delta \in K(r,a). \tag{55}$$

Combining (54) and (55) with the definition of ϑ_m in (17), we obtain that

$$\mathbb{E}(Z_a) + \frac{a}{8} \le 8\vartheta_m\sqrt{\mu_1 r d_2 a} + \frac{a}{8} \le 128\mu_1 r d_2 \vartheta_m^2 + \frac{a}{4}.$$

Then, by choosing $t = a/8$ in (53), it follows that

$$\Pr\left(Z_a \ge \frac{a}{4} + 128\mu_1 r d_2 \vartheta_m^2\right) \le \Pr\left(Z_a \ge \mathbb{E}(Z_a) + \frac{a}{8}\right) \le \exp\left(-\frac{ma^2}{128}\right).$$

This implies that $\Pr(E_k) \le \exp(-4^k\varepsilon^2 m/128)$. Then, by choosing $\varepsilon = \sqrt{\frac{64\log(n_1+n_2)}{\log(2)m}}$ and using $e^x \ge 1 + x > x$, we have

$$\begin{aligned}\Pr(E) \ &\le \sum_{k=1}^{\infty}\Pr(E_k) \le \sum_{k=1}^{\infty}\exp\left(-\frac{4^k\varepsilon^2 m}{128}\right) < \sum_{k=1}^{\infty}\exp\left(-\frac{\log(4)k\varepsilon^2 m}{128}\right)\\ &\le \frac{\exp(-\log(2)m\varepsilon^2/64)}{1-\exp(-\log(2)\varepsilon^2 m/64)} = \frac{1}{n_1+n_2-1}.\end{aligned}$$

Thus, we complete the proof.

Proof of Theorem 3

The proof is similar to that of [38, Theorem 3]. Let $\Delta_m^c := \widehat{X}_m^c - \overline{X}$. By noting that $\gamma_m = 0$ in this case, from (52), we have

$$\left(a_m + \frac{b_m}{\kappa}\right)\|\mathcal{P}_T(\Delta_m^c)\|_* - \frac{\kappa-1}{\kappa}b_m\|\mathcal{P}_{T^\perp}(\Delta_m^c)\|_* \ge 0.$$

Then, by setting $t_m := \frac{\kappa}{\kappa-1}(1+\frac{a_m}{b_m})$, together with the above inequality, we obtain that

$$\|\Delta_m^c\|_* \le \|\mathcal{P}_T(\Delta_m^c)\|_* + \|\mathcal{P}_{T^\perp}(\Delta_m^c)\|_* \le t_m\|\mathcal{P}_T(\Delta_m^c)\|_* \le \sqrt{2r}t_m\|\Delta_m^c\|_F. \tag{56}$$

Let $c_m := \|\mathcal{R}_\beta(\Delta_m^c)\|_\infty$. Clearly, $c_m \le 2c$. We proceed the discussions by two cases:

Case 1. Suppose that $\langle \mathcal{Q}_\beta(\Delta_m^c), \Delta_m^c\rangle \le c_m^2\sqrt{\frac{64\log(n_1+n_2)}{\log(2)m}}$. From (16), we obtain that

$$\frac{\|\Delta_m^c\|_F^2}{d_2} \le 4b^2\mu_1\sqrt{\frac{64\log(n_1+n_2)}{\log(2)m}}.$$

Case 2. Suppose that $\langle \mathcal{Q}_\beta(\Delta_m^c), \Delta_m^c\rangle > c_m^2\sqrt{\frac{64\log(n_1+n_2)}{\log(2)m}}$. Then, from (56), we have $\Delta_m^c/c_m \in K(2t_m^2 r)$. Together with Lemma 2, it follows that

$$\frac{1}{2}\langle \mathcal{Q}_\beta(\Delta_m^c), \Delta_m^c\rangle \le \frac{1}{m}\|\mathcal{R}_\Omega(\Delta_m^c)\|_2^2 + 128c_m^2 t_m^2\mu_1 d_2 r\vartheta_m^2.$$

Combining the last inequality with Theorem 1 and equation (16), we obtain that

$$\begin{aligned}\frac{\|\Delta_m^c\|_F^2}{2d_2} &\le \frac{\mu_1}{2}\langle \mathcal{Q}_\beta(\Delta_m^c), \Delta_m^c\rangle \le \frac{\mu_1}{m}\|\mathcal{R}_\Omega(\Delta_m^c)\|_2^2 + 128c_m^2 t_m^2 \mu_1^2 d_2 r \vartheta_m^2 \\ &\le 2\sqrt{2r}\Big(a_m + \frac{b_m}{\kappa}\Big)\mu_1\rho_m\|\Delta_m^c\|_F + 128c_m^2 t_m^2 \mu_1^2 d_2 r \vartheta_m^2 \\ &\le \frac{\|\Delta_m^c\|_F^2}{4d_2} + 8\Big(a_m + \frac{b_m}{\kappa}\Big)^2 \mu_1^2\rho_m^2 r d_2 + 128c_m^2 t_m^2 \mu_1^2 d_2 r \vartheta_m^2.\end{aligned}$$

By plugging in t_m, we have that there exists some constant C_1 such that

$$\frac{\|\Delta_m^c\|_F^2}{d_2} \le C_1\mu_1^2 d_2 r\left(\Big(a_m + \frac{b_m}{\kappa}\Big)^2\rho_m^2 + \frac{\kappa^2(a_m+b_m)^2}{(\kappa-1)^2 b_m^2}c^2\vartheta_m^2\right).$$

This, together with Case 1, completes the proof.

Proof of Lemma 5.
Recall that $\frac{1}{m}\mathcal{R}_\Omega^*(\xi) = \frac{1}{m}\sum_{i=1}^m \xi_i\Theta_{\omega_i}$. Let $Z_i := \xi_i\Theta_{\omega_i}$. Since $\mathbb{E}(\xi_i) = 0$, the independence of ξ_i and Θ_{ω_i} implies that $\mathbb{E}(Z_i) = 0$. Since $\|\Theta_{\omega_i}\|_F = 1$, we have that

$$\|Z_i\| \le \|Z_i\|_F = |\xi_i|\|\Theta_{\omega_i}\|_F = |\xi_i|.$$

It follows that $\big\|\|Z_i\|\big\|_{\psi_1} \le \|\xi_i\|_{\psi_1}$. Thus, $\big\|\|Z_i\|\big\|_{\psi_1}$ is finite since ξ_i is sub-exponential. Meanwhile, $\mathbb{E}^{\frac{1}{2}}(\|Z_i\|^2) \le \mathbb{E}^{\frac{1}{2}}(\|Z_i\|_F^2) = \mathbb{E}^{\frac{1}{2}}(\xi_i^2) = 1$. We also have

$$\mathbb{E}\big(Z_i Z_i^{\mathbb{T}}\big) = \mathbb{E}\big(\xi_i^2\Theta_{\omega_i}\Theta_{\omega_i}^{\mathbb{T}}\big) = \mathbb{E}\big(\Theta_{\omega_i}\Theta_{\omega_i}^{\mathbb{T}}\big) = \sum_{k\in\beta} p_k\Theta_k\Theta_k^{\mathbb{T}}.$$

The calculation of $\mathbb{E}\big(Z_i^{\mathbb{T}}Z_i\big)$ is similar. From (18), we obtain that $\sqrt{1/n} \le \sigma_Z \le \sqrt{\mu_2/n}$. Then, applying the noncommutative Bernstein inequality yields (19). The proof of (20) is exactly the same as the proof of Lemma 6 in [38]. For simplicity, we omit the proof.

Proof of Lemma 7
(i) From the definition of the sampling operator $\mathcal{R}_\Omega$ and its adjoint $\mathcal{R}_\Omega^*$, we have

$$\frac{1}{m}\mathcal{R}_\Omega^*\mathcal{R}_\Omega(X) = \frac{1}{m}\sum_{i=1}^m \langle\Theta_{\omega_i}, X\rangle\,\Theta_{\omega_i}.$$

This is an average value of m i.i.d. random matrices $\langle\Theta_{\omega_i}, X\rangle\Theta_{\omega_i}$. It is easy to see that $\mathbb{E}\big(\langle\Theta_{\omega_i}, X\rangle\Theta_{\omega_i}\big) = \mathcal{Q}_\beta(X)$. The result then follows directly from the strong law of large numbers.
(ii) From the definition of $\mathcal{R}_\Omega^*$ and $\mathcal{R}_{\alpha\cup\beta}$, it is immediate to obtain that

$$\frac{1}{\sqrt{m}}\mathcal{R}_{\alpha\cup\beta}\mathcal{R}_\Omega^*(\xi) = \frac{1}{\sqrt{m}}\mathcal{R}_{\alpha\cup\beta}\Big(\sum_{i=1}^m \xi_i\Theta_{\omega_i}\Big) = \frac{1}{\sqrt{m}}\sum_{i=1}^m \xi_i\mathcal{R}_{\alpha\cup\beta}(\Theta_{\omega_i}).$$

Since $\mathbb{E}(\xi_i) = 0$ and $\mathbb{E}(\xi_i^2) = 1$, from the independence of ξ_i and $\mathcal{R}_{\alpha\cup\beta}(\Theta_{\omega_i})$, we have $\mathbb{E}\big(\xi_i \mathcal{R}_{\alpha\cup\beta}(\Theta_{\omega_i})\big) = 0$ and $\mathrm{cov}\big(\xi_i, \mathcal{R}_{\alpha\cup\beta}(\Theta_{\omega_i})\big) = p_i$. Applying the central limit theorem then yields the desired result.

Proof of Theorem 10

Let Φ_m denote the objective function of (7) and K denote the feasible set. Then, the problem (7) can be concisely written as

$$\min_{X\in\mathbb{V}^{n_1\times n_2}} \Phi_m(X) + \delta_K(X).$$

By Assumptions 3 and 4 and Lemma 7, we have that the convex functions Φ_m converges pointwise in probability to the convex function Φ, where $\Phi(X) := \frac{1}{2}\|\mathcal{Q}_\beta(X - \overline{X})\|_2^2$ for any $X \in \mathbb{V}^{n_1\times n_2}$. As a direct extension of Rockafellar [63, Theorem 10.8], Andersen and Gill [1, Theorem II.1] proved that the pointwise convergence in probability of a sequence of random convex function implies the convergence in probability (and thus in distribution) with respect to the topology of uniform convergence on compact sets. Then, from Proposition 9 we obtain that $\Phi_m + \delta_K$ epi-converges in distribution to $\Phi + \delta_K$. Note that $\overline{X}$ is the unique minimizer of $\Phi(X) + \delta_K(X)$ since $\Phi(X)$ is strongly convex over the feasible set K. Thus,we complete the proof by applying Proposition 8.

Proof of Lemma 12

By replacing X and H in (23) with $\overline{X}$ and $\rho\Delta$, respectively, and noting that $\sigma_{r+1}(\overline{X}) = 0$, we have $\sigma_{r+1}(\overline{X} + \rho\Delta) - \|\overline{U}_2^{\mathbb{T}}(\rho\Delta)\overline{V}_2\| = O(\|\rho\Delta\|_F^2)$. Since $\overline{U}_2^{\mathbb{T}}\overline{\Delta}\,\overline{V}_2 \neq 0$, for any $\rho \neq 0$ sufficiently small and Δ sufficiently close to $\overline{\Delta}$,

$$\begin{aligned}\frac{\sigma_{r+1}(\overline{X} + \rho\Delta)}{|\rho|} &= \|\overline{U}_2^{\mathbb{T}}\Delta\overline{V}_2\| + O(|\rho|\|\Delta\|_F^2)\\ &\geq \|\overline{U}_2^{\mathbb{T}}\overline{\Delta}\,\overline{V}_2\| - \|\overline{U}_2^{\mathbb{T}}(\Delta - \overline{\Delta})\overline{V}_2\| + O(|\rho|\|\Delta\|_F^2)\\ &\geq \frac{1}{2}\|\overline{U}_2^{\mathbb{T}}\overline{\Delta}\,\overline{V}_2\| > 0.\end{aligned}$$

This implies that $\mathrm{rank}(\overline{X} + \rho\Delta) > r$.

Proof of Proposition 13

By letting $\Delta := \rho_m^{-1}(X - \overline{X})$ in the optimization problem (24), one can easily see that $\widehat{\Delta}_m$ is the optimal solution to

$$\begin{aligned}\min_{\Delta\in\mathbb{V}^{n_1\times n_2}}\ & \frac{1}{2m}\|\mathcal{R}_\Omega(\Delta)\|_2^2 - \frac{\nu}{m\rho_m}\langle\mathcal{R}_\Omega^*(\xi), \Delta\rangle + \frac{1}{\rho_m}\big(\|\overline{X} + \rho_m\Delta\|_* - \|\overline{X}\|_*\big)\\ &- \langle F(\widetilde{X}_m), \Delta\rangle + \frac{\rho_m\gamma_m}{2}\|\Delta\|_F^2 + \gamma_m\langle\overline{X} - \widetilde{X}_m, \Delta\rangle \qquad (57)\\ \text{s.t.}\quad & \mathcal{R}_\alpha(\Delta) = 0.\end{aligned}$$

Let Φ_m and Φ denote the objective functions of (57) and (24), respectively. Let K denote the feasible set of (24). By the definition of directional derivative and [72, Theorem 1],

$$\lim_{\rho_m\to 0}\frac{1}{\rho_m}\big(\|\overline{X} + \rho_m\Delta\|_* - \|\overline{X}\|_*\big) = \langle\overline{U}_1\overline{V}_1^{\mathbb{T}}, \Delta\rangle + \|\overline{U}_2^{\mathbb{T}}\Delta\overline{V}_2\|_*.$$

Then, by combining Assumptions 3 and 4 with Lemma 7, we obtain that Φ_m converges pointwise in probability to Φ. By using the same argument as in the proof of Theorem 10, we obtain that $\Phi_m + \delta_K$ epi-converges in distribution to $\Phi + \delta_K$. Moreover, the optimal solution to (24) is unique due to the strong convexity of Φ over the feasible set K. Therefore, we complete the proof by applying Proposition 8 on the epi-convergence.

Proof of Lemma 14

Assume that $\overline{U}_2^{\mathbb{T}}\widehat{\Delta}\overline{V}_2 = 0$. Since $\widehat{\Delta}$ is the optimal solution to (24), from the optimality condition, the subdifferential of $\|X\|_*$ at 0, and [63, Theorem 23.7], we obtain that there exist some $\widehat{\Gamma} \in \mathbb{V}^{(n_1-r)\times(n_2-r)}$ with $\|\widehat{\Gamma}\| \leq 1$ and $\widehat{\eta} \in \mathbb{R}^{d_1}$ such that

$$\begin{cases} \mathcal{Q}_\beta(\widehat{\Delta}) + \overline{U}_1\overline{V}_1^{\mathbb{T}} - F(\overline{X}) - \mathcal{R}_\alpha^*(\widehat{\eta}) - \overline{U}_2\widehat{\Gamma}\,\overline{V}_2^{\mathbb{T}} = 0, \\ \mathcal{R}_\alpha(\widehat{\Delta}) = 0. \end{cases} \tag{58}$$

Then, according to (4), we can easily obtain (26) by applying the operator $\mathcal{Q}_\beta^\dagger$ to the first equation of (58) and using the second equation. By further combining (26) and $\overline{U}_2^{\mathbb{T}}\widehat{\Delta}\overline{V}_2 = 0$, we obtain that $\widehat{\Gamma}$ is a solution to the linear system (25).

Conversely, if the linear system (25) has a solution $\widehat{\Gamma}$ with $\|\widehat{\Gamma}\| \leq 1$, then it is easy to check that (58) is satisfied with $\widehat{\Delta}$ taking the form of (26) and $\widehat{\eta} = \mathcal{R}_\alpha\big(\overline{U}_1\overline{V}_1^{\mathbb{T}} - F(\overline{X}) - \overline{U}_2\widehat{\Gamma}\,\overline{V}_2^{\mathbb{T}}\big)$. Consequently, $\overline{U}_2^{\mathbb{T}}\widehat{\Delta}\overline{V}_2 = 0$ follows directly from the equations (25) and (26).

Proof of Theorem 16

The estimator $\widehat{X}_m$ is the optimal solution to (7) if and only if there exist a subgradient $\widehat{G}_m$ of the nuclear norm at $\widehat{X}_m$ and a vector $\widehat{\eta}_m \in \mathbb{R}^{d_1}$ such that $(\widehat{X}_m, \widehat{\eta}_m)$ satisfies the KKT conditions:

$$\begin{cases} \dfrac{1}{m}\mathcal{R}_\Omega^*\big(\mathcal{R}_\Omega(\widehat{X}_m) - y\big) + \rho_m\big(\widehat{G}_m - F(\widetilde{X}_m) + \gamma_m(\widehat{X}_m - \widetilde{X}_m)\big) - \mathcal{R}_\alpha^*(\widehat{\eta}_m) = 0, \\ \mathcal{R}_\alpha(\widehat{X}_m) = \mathcal{R}_\alpha(\overline{X}). \end{cases} \tag{59}$$

Let $(\widehat{U}_m, \widehat{V}_m) \in \mathbb{O}^{n_1,n_2}(\widehat{X}_m)$. From Theorem 10 and Corollary 11, we obatain that $\mathrm{rank}(\widehat{X}_m) \geq r$ with probability one. When $\mathrm{rank}(\widehat{X}_m) \geq r$ holds, then from the characterization of the subdifferential of the nuclear norm [72, 73], we have that $\widehat{G}_m = \widehat{U}_{m,1}\widehat{V}_{m,1}^{\mathbb{T}} + \widehat{U}_{m,2}\widehat{\Gamma}_m\widehat{V}_{m,2}^{\mathbb{T}}$ for some $\widehat{\Gamma}_m \in \mathbb{V}^{(n_1-r)\times(n_2-r)}$ satisfying $\|\widehat{\Gamma}_m\| \leq 1$. Moreover, if $\|\widehat{\Gamma}_m\| < 1$, then $\mathrm{rank}(\widehat{X}_m) = r$. Since $\widehat{X}_m \xrightarrow{p} \overline{X}$, by [11, Proposition 8] we have $\widehat{U}_{m,1}\widehat{V}_{m,1}^{\mathbb{T}} \xrightarrow{p} \overline{U}_1\overline{V}_1^{\mathbb{T}}$. Together with Lemma 7, the equation (5) and Lemma 14, it is not hard to obtain that

$$\begin{aligned} &\frac{1}{m\rho_m}\mathcal{R}_\Omega^*\big(\mathcal{R}_\Omega(\widehat{X}_m) - y\big) + \widehat{U}_{m,1}\widehat{V}_{m,1}^{\mathbb{T}} - F(\widetilde{X}_m) + \gamma_m(\widehat{X}_m - \widetilde{X}_m) \\ &\qquad \xrightarrow{p} \mathcal{Q}_\beta(\widehat{\Delta}) + \overline{U}_1\overline{V}_1^{\mathbb{T}} - F(\overline{X}) = \overline{U}_2\widehat{\Gamma}\overline{V}_2^{\mathbb{T}}, \end{aligned} \tag{60}$$

where the equality follows from (26) and $\widehat{\Gamma}$ is the unique optimal solution to (25). Then, by applying the operator $\mathcal{Q}_\beta^\dagger$ to (59), we obtain from (60) that

$$\overline{U}_2^{\mathbb{T}}\mathcal{Q}_\beta^\dagger(\widehat{U}_{m,2}\widehat{\Gamma}_m\widehat{V}_{m,2}^{\mathbb{T}})\overline{V}_2 \xrightarrow{p} \overline{U}_2^{\mathbb{T}}\mathcal{Q}_\beta^\dagger(\overline{U}_2\widehat{\Gamma}\overline{V}_2^{\mathbb{T}})\overline{V}_2. \tag{61}$$

Since $\widehat{X}_m \xrightarrow{p} \overline{X}$, according to [11, Proposition 7], there exist two sequences of matrices $Q_{m,U} \in \mathbb{O}^{n_1-r}$ and $Q_{m,V} \in \mathbb{O}^{n_2-r}$ such that

$$\widehat{U}_{m,2}Q_{m,U} \xrightarrow{p} \overline{U}_2 \quad \text{and} \quad \widehat{V}_{m,2}Q_{m,V} \xrightarrow{p} \overline{V}_2. \tag{62}$$

Moreover, the uniqueness of the solution to the linear system (25) is equivalent to the non-singularity of its linear operator. By combining (61) and (62), we obtain that $Q_{m,U}^{\mathbb{T}}\widehat{\Gamma}_m Q_{m,V} \xrightarrow{p} \widehat{\Gamma}$. Hence, we obtain that $\|\widehat{\Gamma}_m\| < 1$ with probability one since $\|\widehat{\Gamma}\| < 1$. As discussed above, it follows that $\mathrm{rank}(\widehat{X}_m) = r$ with probability one.

Proof of Proposition 17

It is easy to verify that $\widehat{\Delta}_m$ is the optimal solution to

$$\begin{aligned} \min_{\Delta \in \mathbb{S}^n} \ & \frac{1}{2m}\|\mathcal{R}_\Omega(\Delta)\|_2^2 - \frac{\nu}{m\rho_m}\langle \mathcal{R}_\Omega^*(\xi), \Delta\rangle + \langle I_n - F(\widetilde{X}_m), \Delta\rangle + \frac{\rho_m\gamma_m}{2}\|\Delta\|_F^2 \\ & \qquad + \gamma_m\langle \overline{X} - \widetilde{X}_m, \Delta\rangle \\ \text{s.t.} \ & \Delta \in K_m := \rho_m^{-1}(E \cap \mathbb{S}_+^n - \overline{X}), \end{aligned} \tag{63}$$

where $E := \{X \in \mathbb{S}^n \mid \mathcal{R}_\alpha(X) = \mathcal{R}_\alpha(\overline{X})\}$. Let Φ_m and Φ denote the objective functions of (63) and (27), respectively. Then the convex functions Φ_m converges pointwise in probability to the convex function Φ. Moreover, by considering the upper limit and lower limit of the family of feasible sets K_m, we know that K_m converges in the sense of Painlevé-Kuratowski to the tangent cone $\mathcal{T}_{E\cap\mathbb{S}_+^n}(\overline{X})$ (see [64, 5]). Note that the Slater condition implies that E and $\mathbb{S}_+^n$ cannot be separated. Then, from [64, Theorem 6.42], we have $\mathcal{T}_{E\cap\mathbb{S}_+^n}(\overline{X}) = \mathcal{T}_E(\overline{X}) \cap \mathcal{T}_{\mathbb{S}_+^n}(\overline{X})$. Elearly, $\mathcal{T}_E(\overline{X}) = \{\Delta \in \mathbb{S}^n \mid \mathcal{R}_\alpha(\Delta) = 0\}$. Moreover, by Arnold [2],

$$\mathcal{T}_{\mathbb{S}_+^n}(\overline{X}) = \{\Delta \in \mathbb{S}^n \mid \overline{P}_2^{\mathbb{T}}\Delta\overline{P}_2 \in \mathbb{S}_+^{n-r}\}.$$

Since epi-convergence of functions corresponds to set convergence of their epigraphs [64], we obtain that δ_{K_m} epi-converges to $\delta_{\mathcal{T}_{E\cap\mathbb{S}_+^n}} = \delta_{\mathcal{T}_E} + \delta_{\mathcal{T}_{\mathbb{S}_+^n}}$. Then, from Proposition 9, $\Phi_m + \delta_{K_m}$ epi-converges in distribution to $\Phi + \delta_{\mathcal{T}_E} + \delta_{\mathcal{T}_{\mathbb{S}_+^n}}$. In addition, the optimal solution to (27) is unique due to the strong convexity of Φ over the feasible set $E \cap \mathbb{S}_+^n$. Therefore, we complete the proof by applying Proposition 8 on the epi-convergence.

Proof of Lemma 18

Note that the Slater condition also holds for the problem (27). (One may check the point $X^0 - \overline{X}$.) Hence, $\widehat{\Delta}$ is the optimal solution to (27) if and only if there exists $(\widehat{\zeta}, \widehat{\Lambda}) \in \mathbb{R}^{d_1} \times \mathbb{S}^{n-r}$ such that

$$\begin{cases} \mathcal{Q}_\beta(\widehat{\Delta}) + I_n - F(\overline{X}) - \mathcal{R}_\alpha^*(\widehat{\zeta}) - \overline{P}_2\widehat{\Lambda}\overline{P}_2^{\mathbb{T}} = 0, \\ \mathcal{R}_\alpha(\widehat{\Delta}) = 0, \\ \overline{P}_2^{\mathbb{T}}\widehat{\Delta}\overline{P}_2 \in \mathbb{S}_+^{n-r}, \ \widehat{\Lambda} \in \mathbb{S}_+^{n-r}, \ \langle \overline{P}_2^{\mathbb{T}}\widehat{\Delta}\overline{P}_2, \widehat{\Lambda}\rangle = 0. \end{cases} \tag{64}$$

Applying the operator $\mathcal{Q}_\beta^\dagger$ to the first equation of (64) yields the equality (29). Suppose that $\overline{P}_2^{\mathbb{T}}\widehat{\Delta}\overline{P}_2 = 0$. Then, it is immediate to obtain from (26) that $\widehat{\Lambda}$ is a solution to the linear system (28).

Conversely, if the linear system (28) has a solution $\widehat{\Lambda} \in \mathbb{S}_+^{n-r}$, then it is easy to check that (64) is satisfied with $\widehat{\Delta}$ taking the form of (29) and $\widehat{\zeta} = \mathcal{R}_\alpha\big(I_n - F(\overline{X}) - \overline{P}_2\widehat{\Lambda}\,\overline{P}_2^{\mathbb{T}}\big)$. Then, $\overline{P}_2^{\mathbb{T}}\widehat{\Delta}\overline{P}_2 = 0$ directly follows from (29) and the first equation of (64).

Proof of Theorem 20

The Slater condition implies that $\widehat{X}_m$ is the optimal solution to (8) if and only if there exists multipliers $(\widehat{\zeta}_m, \widehat{S}_m) \in \mathbb{R}^{d_1} \times \mathbb{S}^n$ such that $(\widehat{X}_m, \widehat{\zeta}_m, \widehat{S}_m)$ satisfy the KKT conditions:

$$\begin{cases} \dfrac{1}{m}\mathcal{R}_\Omega^*\big(\mathcal{R}_\Omega(\widehat{X}_m) - y\big) + \rho_m\big(I_n - F(\widetilde{X}_m) + \gamma_m(\widehat{X}_m - \widetilde{X}_m)\big) - \mathcal{R}_\alpha^*(\widehat{\zeta}_m) - \widehat{S}_m = 0, \\ \mathcal{R}_\alpha(\widehat{X}_m) = \mathcal{R}_\alpha(\overline{X}), \\ \widehat{X}_m \in \mathbb{S}_+^n,\ \widehat{S}_m \in \mathbb{S}_+^n,\ \langle \widehat{X}_m, \widehat{S}_m\rangle = 0. \end{cases} \tag{65}$$

The third equation of (65) implies that $\widehat{X}_m$ and $\widehat{S}_m$ can have a simultaneous eigenvalue decomposition. Let $\widehat{P}_m \in \mathbb{O}^n(\widehat{X}_m)$. From Theorem 10 and Corollary 11, we know that $\mathrm{rank}(\widehat{X}_m) \geq r$ with probability one. When $\mathrm{rank}(\widehat{X}_m) \geq r$ holds, we can write $\widehat{S}_m = \widehat{P}_{m,2}\widehat{\Lambda}_m\widehat{P}_{m,2}^{\mathbb{T}}$ for some diagonal matrix $\widehat{\Lambda}_m \in \mathbb{S}_+^{n-r}$. In addition, if $\widehat{\Lambda}_m \in \mathbb{S}_{++}^{n-r}$, then $\mathrm{rank}(\widehat{X}_m) = r$. Since $\widehat{X}_m \xrightarrow{p} \overline{X}$, according to [11, Proposition 1], there exists a sequence of matrices $Q_m \in \mathbb{O}^{n-r}$ such that $\widehat{P}_{m,2}Q_m \xrightarrow{p} \overline{P}_2$. Then, using the similar arguments to the proof of Theorem 16, we obtain that $Q_m^{\mathbb{T}}\widehat{\Lambda}_m Q_m \xrightarrow{p} \widehat{\Lambda}$. Since $\widehat{\Lambda} \in \mathbb{S}_{++}^n$, we have $\widehat{\Lambda}_m \in \mathbb{S}_{++}^n$ with probability one. Thus, we complete the proof.

Proof of Proposition 21

For the real covariance matrix case, the proof is given in [58, Lemma 3.3] and [59, Proposition 2.1]. For the complex covariance matrix case, one can use the similar arguments to prove the result.

We next consider the density matrix case. Suppose that $\overline{X}$ satisfies the density constraint, i.e., $\mathcal{R}_\alpha(\overline{X}) = \mathrm{Tr}(\overline{X}) = 1$. Note that for any $t \in \mathbb{R}$, we have $t\overline{X} \in \mathrm{lin}\big(\mathcal{T}_{\mathcal{H}_+^n}(\overline{X})\big)$. This, along with $\mathrm{Tr}(\overline{X}) = 1$, implies that $\mathrm{Tr}\big(\mathrm{lin}(\mathcal{T}_{\mathcal{H}_+^n}(\overline{X}))\big) = \mathcal{R}_\alpha\big(\mathrm{lin}(\mathcal{T}_{\mathcal{H}_+^n}(\overline{X}))\big) = \mathbb{R}$. This means that the constraint nondegeneracy condition (35) holds.

Proof of Proposition 22

We prove for the rectangular case by contradiction. Assume that there exists some $\mathbb{V}^{(n_1-r)\times(n_2-r)} \ni \overline{\Gamma} \neq 0$ such that $\mathcal{B}_2(\overline{\Gamma}) = \overline{U}_2^{\mathbb{T}}\mathcal{Q}_\beta^\dagger(\overline{U}_2\overline{\Gamma}\,\overline{V}_2^{\mathbb{T}})\overline{V}_2 = 0$. By noting that $\mathcal{Q}_\beta^\dagger$ is a self-adjoint and positive semidefinite operator, we obtain $(\mathcal{Q}_\beta^\dagger)^{1/2}(\overline{U}_2\overline{\Gamma}\,\overline{V}_2^{\mathbb{T}}) = 0$. It follows that $\mathcal{P}_\beta(\overline{U}_2\overline{\Gamma}\,\overline{V}_2^{\mathbb{T}}) = 0$. This, together with $\overline{\Gamma} \neq 0$, implies that $\mathcal{P}_\alpha(\overline{U}_2\overline{\Gamma}\,\overline{V}_2^{\mathbb{T}}) = \overline{U}_2\overline{\Gamma}\,\overline{V}_2^{\mathbb{T}} \neq 0$ and moreover $\mathcal{R}_\alpha(\overline{U}_2\overline{\Gamma}\,\overline{V}_2^{\mathbb{T}}) \neq 0$. However, for any $H \in \mathcal{T}(\overline{X})$, we have

$$\langle \mathcal{R}_\alpha(\overline{U}_2\overline{\Gamma}\,\overline{V}_2^{\mathbb{T}}), \mathcal{R}_\alpha(H)\rangle = \langle \mathcal{P}_\alpha(\overline{U}_2\overline{\Gamma}\,\overline{V}_2^{\mathbb{T}}), H\rangle = \langle \overline{U}_2\overline{\Gamma}\,\overline{V}_2^{\mathbb{T}}, H\rangle = \langle \overline{\Gamma}, \overline{U}_2^{\mathbb{T}}H\overline{V}_2\rangle = 0.$$

Thus, the constraint nondegeneracy condition (33) implies that $\mathcal{R}_\alpha(\overline{U}_2\overline{\Gamma}\,\overline{V}_2^{\mathbb{T}}) = 0$. This leads to a contradiction. Therefore, the linear operator $\mathcal{B}_2$ is positive definite. The proof for the positive semidefinite case is similar.

Proof of Theorem 24

From Propositions 21 and 22, for both cases, the linear system (28) has a unique solution $\widehat{\Lambda}$. Moreover, uniform sampling yields $\mathcal{Q}_\beta^\dagger = \mathcal{P}_\beta/d_2$. Thus, from (28), we get

$$\widehat{\Lambda} - \overline{P}_2^{\mathbb{T}}\mathcal{P}_\alpha(\overline{P}_2\widehat{\Lambda}\overline{P}_2^{\mathbb{T}})\overline{P}_2 = \overline{P}_2^{\mathbb{T}}\mathcal{P}_\beta(\overline{P}_2\widehat{\Lambda}\overline{P}_2^{\mathbb{T}})\overline{P}_2 = \overline{P}_2^{\mathbb{T}}\mathcal{P}_\beta\big(I_n - F(\overline{X})\big)\overline{P}_2. \tag{66}$$

We first prove the covariance matrix completion by contradiction. Without loss of generality, we assume that the first l diagonal entries are fixed and positive. Then, for any $X \in \mathbb{S}_+^n$, $\mathcal{P}_\alpha(X)$ is the diagonal matrix whose first l diagonal entries are $X_{ii}, 1 \le i \le l$ respectively and the other entries are 0. Assume that $\widehat{\Lambda} \notin \mathbb{S}_{++}^{n-r}$, i.e., $\lambda_{\min}(\widehat{\Lambda}) \le 0$, where $\lambda_{\min}(\cdot)$ denotes the smallest eigenvalue. Then, we have

$$\lambda_{\min}(\widehat{\Lambda}) = \lambda_{\min}(\overline{P}_2\widehat{\Lambda}\overline{P}_2^{\mathbb{T}}) \le \lambda_{\min}\big(\mathcal{P}_\alpha(\overline{P}_2\widehat{\Lambda}\overline{P}_2^{\mathbb{T}})\big) \le \lambda_{\min}\big(\overline{P}_2^{\mathbb{T}}\mathcal{P}_\alpha(\overline{P}_2\widehat{\Lambda}\overline{P}_2^{\mathbb{T}})\overline{P}_2\big),$$

where the equality follows from the fact that $\widehat{\Lambda}$ and $\overline{P}_2\widehat{\Lambda}\overline{P}_2^{\mathbb{T}}$ have the same nonzero eigenvalues, the first inequality follows from the fact that the vector of eigenvalues is majorized by the vector of diagonal entries (e.g., see [47, Theorem 9.B.1]), and the second inequality follows from the Courant-Fischer minmax theorem (e.g., see [47, Theorem 20.A.1]). As a result, the left-hand side of (66) is not positive definite. However, the right-hand side of (66) can be written as

$$\overline{P}_2^{\mathbb{T}}\mathcal{P}_\beta\big(I_n - F(\overline{X})\big)\overline{P}_2 = \overline{P}_2^{\mathbb{T}}\left(\mathcal{P}_\beta(I_n) - F(\overline{X}) + \mathcal{P}_\alpha(F(\overline{X}))\right)\overline{P}_2 = \overline{P}_2^{\mathbb{T}}\big(\mathcal{P}_\beta(I_n) + \mathcal{P}_\alpha(F(\overline{X}))\big)\overline{P}_2,$$

where the second equality follows from the fact that $\overline{P}_2^{\mathbb{T}}F(\overline{X})\overline{P}_2 = 0$. Since $\operatorname{rank}(\overline{X}) = r$, with the choice (42) of F, we have that for any $1 \le i \le l$,

$$\overline{X}_{ii} = \sum_{j=1}^{r} \lambda_j(\overline{X})|\overline{P}_{ij}|^2 > 0 \quad \text{implies} \quad \big(F(\overline{X})\big)_{ii} = \sum_{j=1}^{r} f_i\big(\lambda_j(\overline{X})\big)|\overline{P}_{ij}|^2 > 0.$$

Moreover, $\mathcal{P}_\beta(I_n)$ is the diagonal matrix with the last $n-r$ diagonal entries being ones and the other entries being zeros. Thus, $\mathcal{P}_\beta(I_n) + \mathcal{P}_\alpha\big(F(\overline{X})\big)$ is a diagonal matrix with all positive diagonal entries. It follows that the right-hand side of (66) is positive definite. Thus, we obtain a contradiction. Hence, $\widehat{\Lambda} \in \mathbb{S}_{++}^{n-r}$. Then, from Theorem 20, we obtain the rank consistency.

For the density matrix completion, $\mathcal{P}_\alpha(\cdot) = \frac{1}{n}\operatorname{Tr}(\cdot)I_n$. By further using $\overline{P}_2^{\mathbb{T}}F(\overline{X})\overline{P}_2 = 0$ and $\mathcal{P}_\beta(I_n) = 0$, we can rewrite (66) as

$$\widehat{\Lambda} - \frac{1}{n}\operatorname{Tr}(\widehat{\Lambda})I_{n-r} = \frac{1}{n}\operatorname{Tr}(F(\overline{X}))I_{n-r}.$$

By taking the trace on both sides, we obtain that $\widehat{\Lambda} = \frac{1}{r}\mathrm{Tr}(F(\overline{X}))I_{n-r}$. Since $\overline{X}$ is a density matrix of rank r, with the choice (42) of F, we have that

$$\mathrm{Tr}(\overline{X}) = \sum_{i=1}^{n}\sum_{j=1}^{r}\lambda_j(\overline{X})|\overline{P}_{ij}|^2 = 1 \quad \text{implies} \quad \mathrm{Tr}\big(F(\overline{X})\big) = \sum_{i=1}^{n}\sum_{j=1}^{r} f_i\big(\lambda_j(\overline{X})\big)|\overline{P}_{ij}|^2 > 0.$$

It follows that $\widehat{\Lambda} \in \mathbb{S}^{n-r}_{++}$ and thus we obtain the rank consistency.

References

[1] P.K. Andersen and R.D. Gill. Cox's regression model for counting processes: a large sample study. *The Annals of Statistics*, 10(4):1100–1120, 1982.

[2] V.I. Arnold. On matrices depending on parameters. *Russian Mathematical Surveys*, 26(2):29–43, 1971.

[3] F.R. Bach. Consistency of trace norm minimization. *Journal of Machine Learning Research*, 9:1019–1048, 2008.

[4] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.

[5] J.F. Bonnans and A. Shapiro. *Perturbation analysis of optimization problems*. Springer Verlag, 2000.

[6] P. Bühlmann and S. Van De Geer. *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer-Verlag New York Inc, 2011.

[7] E.J. Candes and Y. Plan. Matrix completion with noise. *Proceedings of the IEEE*, 98(6):925–936, 2010.

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

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

[10] C. Ding. *An introduction to a class of matrix optimization problems*. PhD thesis, National University of Singapore, 2012.

[11] C. Ding, D.F. Sun, and K.C. Toh. An introduction to a class of matrix cone programming. *Preprint available at* http://www.math.nus.edu.sg/~matsundf/IntroductionMCP_revised_Mar_26_12.pdf, 2010.

[12] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *The Annals of statistics*, 32(2):407–499, 2004.

[13] J. Fan. Comments on "Wavelets in statistics: A review" by A. Antoniadis. *Statistical Methods & Applications*, 6(2):131–138, 1997.

[14] J. Fan and R. Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456):1348–1360, 2001.

[15] J. Fan and J. Lv. Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(5):849–911, 2008.

[16] J. Fan and J. Lv. A selective overview of variable selection in high dimensional feature space. *Statistica Sinica*, 20(1):101, 2010.

[17] J. Fan, J. Lv, and L. Qi. Sparse high dimensional models in economics. *Annual Review of Economics*, 3:291, 2011.

[18] J. Fan and H. Peng. Nonconcave penalized likelihood with a diverging number of parameters. *The Annals of Statistics*, 32(3):928–961, 2004.

[19] M. Fazel. *Matrix rank minimization with applications*. PhD thesis, Stanford University, 2002.

[20] M. Fazel, H. Hindi, and S.P. Boyd. Log-det heuristic for matrix rank minimization with applications to hankel and euclidean distance matrices. In *American Control Conference, 2003. Proceedings of the 2003*, volume 3, pages 2156–2162. Ieee, 2003.

[21] S.T. Flammia, D. Gross, Y.K. Liu, and J. Eisert. Quantum tomography via compressed sensing: error bounds, sample complexity, and efficient estimators. *Arxiv preprint* arXiv:1205.2300, 2012.

[22] R. Foygel, R. Salakhutdinov, O. Shamir, and N. Srebro. Learning with the weighted trace-norm under arbitrary sampling distributions. *Arxiv preprint* arXiv:1106.4251, 2011.

[23] R. Foygel and N. Srebro. Concentration-based guarantees for low-rank matrix reconstruction. *Arxiv preprint* arXiv:1102.3923, 2011.

[24] Y. Gao. *Structured low rank matrix optimization problems: a penalized approach*. PhD thesis, National University of Singapore, 2010.

[25] Y. Gao and D.F. Sun. A majorized penalty approach for calibrating rank constrained correlation matrix problems. *Preprint available at* `http://www.math.nus.edu.sg/~matsundf/MajorPen_May5.pdf`, 2010.

[26] C.J. Geyer. On the asymptotics of constrained M-estimation. *The Annals of Statistics*, pages 1993–2010, 1994.

[27] C.J. Geyer. On the asymptotics of convex stochastic optimization. *Unpublished manuscript*, 1996.

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

[29] D. Gross, Y.K. Liu, S.T. Flammia, S. Becker, and J. Eisert. Quantum state tomography via compressed sensing. *Physical Review Letters*, 105(15):150401, 2010.

[30] C. Han and P.C.B. Phillips. Gmm with many moment conditions. *Econometrica*, 74(1):147–192, 2006.

[31] J. Huang, S. Ma, and C.H. Zhang. Adaptive lasso for sparse high-dimensional regression models. *Statistica Sinica*, 18(4):1603, 2010.

[32] K. Jiang, D. Sun, and K.C. Toh. An inexact accelerated proximal gradient method for large scale linearly constrained convex SDP. *SIAM Journal on Optimization*, 22(3):1042–1064, 2012.

[33] K. Jiang, D.F. Sun, and K.C. Toh. A partial proximal point algorithm for nuclear norm regularized matrix least squares problems. *Preprint available at* http://www.math.nus.edu.sg/~matsundf/PPA_Smoothing-2.pdf, 2012.

[34] K. Jiang, D.F. Sun, and K.C. Toh. Solving nuclear norm regularized and semidefinite matrix least squares problems with linear equality constraints. *Preprint available at* http://www.math.nus.edu.sg/~matsundf/PPA_Semismooth-3.pdf, 2012.

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

[36] A.J. King and R.J.B. Wets. Epi-consistency of convex stochastic programs. *Stochastics: An International Journal of Probability and Stochastic Processes*, 34(1-2):83–92, 1991.

[37] O. Klopp. Rank penalized estimators for high-dimensional matrices. *Electronic Journal of Statistics*, 5:1161–1183, 2011.

[38] O. Klopp. Noisy low-rank matrix completion with general sampling distribution. *Arxiv preprint* arXiv:1203.0108, 2012.

[39] K. Knight. Epi-convergence in distribution and stochastic equi-semicontinuity. *Unpublished manuscript*, 1999.

[40] V. Koltchinskii. *Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems: Ecole Dâeté de Probabilités de Saint-FlourXXXVIII-2008*, volume 2033. Springer, 2011.

[41] V. Koltchinskii. Von Neumann entropy penalization and low-rank matrix estimation. *The Annals of Statistics*, 39(6):2936–2973, 2012.

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

[43] M. Ledoux and M. Talagrand. *Probability in Banach Spaces: Isoperimetry and Processes*, volume 23. Springer, 1991.

[44] C. Leng, Y. Lin, and G. Wahba. A note on the lasso and related procedures in model selection. *Statistica Sinica*, 16(4):1273, 2006.

[45] A.S. Lewis and H.S. Sendov. Nonsmooth analysis of singular values. Part II: Applications. *Set-Valued Analysis*, 13(3):243–264, 2005.

[46] Y.K. Liu. Universal low-rank matrix recovery from Pauli measurements. *Arxiv preprint* arXiv:1103.2816, 2011.

[47] A.W. Marshall, I. Olkin, and B. Arnold. *Inequalities: Theory of Majorization and Its Applications*. Springer Verlag, 2010.

[48] P. Massart. About the constants in Talagrand's concentration inequalities for empirical processes. *The Annals of Probability*, pages 863–884, 2000.

[49] N. Meinshausen. Relaxed lasso. *Computational Statistics & Data Analysis*, 52(1):374–393, 2007.

[50] N. Meinshausen and P. Bühlmann. High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, 34(3):1436–1462, 2006.

[51] M. Mesbahi. On the rank minimization problem and its control applications. *Systems & Control Letters*, 33(1):31–36, 1998.

[52] M. Mesbahi and G.P. Papavassilopoulos. On the rank minimization problem over a positive semidefinite linear matrix inequality. *Automatic Control, IEEE Transactions on*, 42(2):239–243, 1997.

[53] V.D. Milman and G. Schechtman. *Asymptotic Theory of Finite Dimensional Normed Spaces*, volume 1200. Springer Verlag, 1986.

[54] K. Mohan and M. Fazel. Reweighted nuclear norm minimization with application to system identification. In *American Control Conference (ACC), 2010*, pages 2953–2959. IEEE, 2010.

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

[56] Y. Nesterov. A method of solving a convex programming problem with convergence rate $O(1/k^2)$. In *Soviet Mathematics Doklady*, volume 27, pages 372–376, 1983.

[57] G.C. Pflug. Asymptotic stochastic programs. *Mathematics of Operations Research*, pages 769–789, 1995.

[58] H. Qi and D.F. Sun. A quadratically convergent newton method for computing the nearest correlation matrix. *SIAM Journal on Matrix Analysis and Applications*, 28(2):360, 2006.

[59] H. Qi and D.F. Sun. An augmented Lagrangian dual approach for the H-weighted nearest correlation matrix problem. *IMA Journal of Numerical Analysis*, 31(2):491–511, 2011.

[60] B. Recht. A simpler approach to matrix completion. *Journal of Machine Learning Research*, 12:3413–3430, 2011.

[61] B. Recht, M. Fazel, and P.A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Review*, 52(3):471–501, 2010.

[62] S.M. Robinson. Local structure of feasible sets in nonlinear programming, Part II: Nondegeneracy. *Mathematical Programming at Oberwolfach II*, pages 217–230, 1984.

[63] R.T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[64] R.T. Rockafellar and R.J.B. Wets. *Variational Analysis*, volume 317. Springer Verlag, 1998.

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

[66] R. Salakhutdinov and N. Srebro. Collaborative filtering in a non-uniform world: learning with the weighted trace norm. *Arxiv preprint* arXiv:1002.2780, 2010.

[67] N. Srebro, J.D.M. Rennie, and T. Jaakkola. Maximum-margin matrix factorization. *Advances in neural information processing systems*, 17(5):1329–1336, 2005.

[68] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[69] J.A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, pages 1–46, 2011.

[70] A.W. Van Der Vaart and J.A. Wellner. *Weak Convergence and Empirical Processes*. Springer Verlag, 1996.

[71] Y. Wang. Asymptotic equivalence of quantum state tomography and trace regression. *Preprint available at* http://pages.stat.wisc.edu/~yzwang/paper/QuantumTomographyTrace.pdf, 2012.

[72] G.A. Watson. Characterization of the subdifferential of some matrix norms. *Linear Algebra and its Applications*, 170:33–45, 1992.

[73] G.A. Watson. On matrix approximation problems with Ky Fan k norms. *Numerical Algorithms*, 5(5):263–272, 1993.

[74] C.H. Zhang. Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 38(2):894–942, 2010.

[75] P. Zhao and B. Yu. On model selection consistency of lasso. *Journal of Machine Learning Research*, 7(2):2541, 2007.

[76] S. Zhou, S. Van De Geer, and P. Bühlmann. Adaptive Lasso for high dimensional regression and Gaussian graphical modeling. *Arxiv preprint* arXiv:0903.2515, 2009.

[77] H. Zou. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476):1418–1429, 2006.

[78] H. Zou and R. Li. One-step sparse estimates in nonconcave penalized likelihood models. *The Annals of Statistics*, 36(4):1509, 2008.