HIGH-DIMENSIONAL ANALYSIS ON MATRIX DECOMPOSITION WITH APPLICATION TO CORRELATION MATRIX ESTIMATION IN FACTOR MODELS

WU BIN

(B.Sc., ZJU, China)

A THESIS SUBMITTED
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
DEPARTMENT OF MATHEMATICS
NATIONAL UNIVERSITY OF SINGAPORE
2014

To my parents

DECLARATION

I hereby declare that the thesis is my original work and it has been written by me in its entirety.

I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Wu Bin

January 2014

Acknowledgements

I would like to express my sincerest gratitude to my supervisor Professor Sun Defeng for his professional guidance during these past five and a half years. He has patiently given me the freedom to pursue interesting research and also consistently provided me with prompt and insightful feedbacks that usually point to promising directions. His inexhaustible enthusiasm for research and optimistic attitude to difficulties have impressed and influenced me profoundly. Moreover, I am very grateful for his financial support for my fifth year's research.

I have benefited a lot from the previous and present members in the optimization group at Department of Mathematics, National University of Singapore. Many thanks to Professor Toh Kim-Chuan, Professor Zhao Gongyun, Zhao Xinyuan, Liu Yongjin, Wang Chengjing, Li Lu, Gao Yan, Ding Chao, Miao Weimin, Jiang Kaifeng, Gong Zheng, Shi Dongjian, Li Xudong, Du Mengyu and Cui Ying. I cannot imagine a better group of people to spend these days with. In particular, I would like to give my special thanks to Ding Chao and Miao Weimin. Valuable comments and constructive suggestions from the extensive discussions with them were extremely illuminating and helpful. Additionally, I am also very thankful to

Li Xudong for his help and support in coding.

I would like to convey my great appreciation to National University of Singapore for offering me the four-year President's Graduate Fellowship, and to Department of Mathematics for providing me the conference financial assistance of the 21st International Symposium on Mathematical Programming (ISMP) in Berlin, the final half year financial support, and most importantly the excellent research conditions. My appreciation also goes to the Computer Centre in National University of Singapore for providing the High Performance Computing (HPC) service that greatly facilitates my research.

My heartfelt thanks are devoted to all my dear friends, especially Ding Chao, Miao Weimin, Hou Likun and Sun Xiang, for their companionship and encouragement during these years. It is you guys who made my Ph.D. study a joyful and memorable journey.

As always, I owe my deepest gratitude to my parents for their constant and unconditional love and support throughout my life. Last but not least, I am also deeply indebted to my fiancée, Gao Yan, for her understanding, tolerance, encouragement and love. Meeting, knowing, and falling in love with her in Singapore is unquestionably the most beautiful story that I have ever experienced.

Wu Bin

January, 2014

Contents

Summary

In this thesis, we conduct high-dimensional analysis on the problem of low-rank and sparse matrix decomposition with fixed and sampled basis coefficients. This problem is strongly motivated by high-dimensional correlation matrix estimation coming from a factor model used in economic and financial studies, in which the underlying correlation matrix is assumed to be the sum of a low-rank matrix and a sparse matrix respectively due to the common factors and the idiosyncratic components, and the fixed basis coefficients are the diagonal entries.

We consider both of the noiseless and noisy versions of this problem. For the noiseless version, we develop exact recovery guarantees provided that certain standard identifiability conditions for the low-rank and sparse components are assumed to be satisfied. These probabilistic recovery results are especially in accordance with the high-dimensional setting because only a vanishingly small fraction of samples is already sufficient when the intrinsic dimension increases. For the noisy version, inspired by the successful recent development on the adaptive nuclear semi-norm penalization technique for noisy low-rank matrix completion [98, 99], we propose a two-stage rank-sparsity-correction procedure and then examine its

recovery performance by establishing, for the first time up to our knowledge, a non-asymptotic probabilistic error bound under the high-dimensional scaling.

As a main application of our theoretical analysis, we specialize the aforementioned two-stage correction procedure to deal with the correlation matrix estimation problem with missing observations in strict factor models where the sparse component is known to be diagonal. By virtue of this application, the specialized recovery error bound and the convincing numerical results show the superiority of the two-stage correction approach over the nuclear norm penalization.

List of Notations

- Let $\mathbb{R}^n$ be the linear space of all n-dimensional real vectors and $\mathbb{R}^n_+$ be the n-dimensional positive orthant. For any x and $y \in \mathbb{R}^n$, the notation $x \geq 0$ means that $x \in \mathbb{R}^n_+$, and the notation $x \geq y$ means that $x - y \geq 0$.

- Let $\mathbb{R}^{n_1 \times n_2}$ be the linear space of all $n_1 \times n_2$ real matrices and $\mathcal{S}^n$ be the linear space of all $n \times n$ real symmetric matrices.

- Let $\mathbb{V}^{n_1 \times n_2}$ represent the finite dimensional real Euclidean space $\mathbb{R}^{n_1 \times n_2}$ or $\mathcal{S}^n$ with $n := \min\{n_1, n_2\}$. Suppose that $\mathbb{V}^{n_1 \times n_2}$ is equipped with the trace inner product $\langle X, Y \rangle := \mathrm{Tr}(X^T Y)$ for X and Y in $\mathbb{V}^{n_1 \times n_2}$, where "Tr" stands for the trace of a squared matrix.

- Let $\mathcal{S}^n_+$ denote the cone of all $n \times n$ real symmetric and positive semidefinite matrices. For any X and $Y \in \mathcal{S}^n$, the notation $X \succeq 0$ means that $X \in \mathcal{S}^n_+$, and the notation $X \succeq Y$ means that $X - Y \succeq 0$.

- Let $\mathcal{O}^{n \times r}$ (where $n \geq r$) represent the set of all $n \times r$ real matrices with orthonormal columns. When $n = r$, we write $\mathcal{O}^{n \times r}$ as $\mathcal{O}^n$ for short.

- Let I_n denote the $n \times n$ identity matrix, $\mathbf{1}$ denote the vector of proper dimension whose entries are all ones, and e_i denote the i-th standard basis vector of proper dimension whose entries are all zeros except the i-th being one.

- For any $x \in \mathbb{R}^n$, let $\|x\|_p$ denote the vector ℓ_p-norm of x, where $p = 0$, 1, 2, or ∞. For any $X \in \mathbb{V}^{n_1 \times n_2}$, let $\|X\|_0$, $\|X\|_1$, $\|X\|_\infty$, $\|X\|_F$, $\|X\|$ and $\|X\|_*$ denote the matrix ℓ_0-norm, the matrix ℓ_1-norm, the matrix ℓ_∞-norm, the Frobenius norm, the spectral (or operator) norm and the nuclear norm of X, respectively.

- The Hardamard product between vectors or matrices is denoted by "$\circ$", i.e., for any x and $y \in \mathbb{R}^n$, the i-th entry of $x \circ y \in \mathbb{R}^n$ is $x_i y_i$; for any X and $Y \in \mathbb{V}^{n_1 \times n_2}$, the (i,j)-th entry of $X \circ Y \in \mathbb{V}^{n_1 \times n_2}$ is $X_{ij} Y_{ij}$.

- Define the function sign : $\mathbb{R} \to \mathbb{R}$ by $\mathrm{sign}(t) = 1$ if $t > 0$, $\mathrm{sign}(t) = -1$ if $t < 0$, and $\mathrm{sign}(t) = 0$ if $t = 0$, for $t \in \mathbb{R}$. For any $x \in \mathbb{R}^n$, let $\mathrm{sign}(x)$ be the sign vector of x, i.e., $[\mathrm{sign}(x)]_i = \mathrm{sign}(x_i)$, for $i = 1, \ldots, n$. For any $X \in \mathbb{V}^{n_1 \times n_2}$, let $\mathrm{sign}(X)$ be the sign matrix of X where $[\mathrm{sign}(X)]_{ij} = \mathrm{sign}(X_{ij})$, for $i = 1, \ldots, n_1$ and $j = 1, \ldots, n_2$.

- For any $x \in \mathbb{R}^n$, let $|x| \in \mathbb{R}^n$ be the vector whose i-th entry is $|x_i|$, $x^\downarrow \in \mathbb{R}^n$ be the vector of entries of x being arranged in the non-increasing order $x_1^\downarrow \geq \cdots \geq x_n^\downarrow$, and $x^\uparrow \in \mathbb{R}^n$ be the vector of entries of x being arranged in the non-decreasing order $x_1^\uparrow \leq \cdots \leq x_n^\uparrow$. For any index set $\mathcal{J} \subseteq \{1, \ldots, n\}$, we use $|\mathcal{J}|$ to represent the cardinality of $\mathcal{J}$, i.e., the number of elements in $\mathcal{J}$. Moreover, we use $x_\mathcal{J} \in \mathbb{R}^{|\mathcal{J}|}$ to denote the sub-vector of x indexed by $\mathcal{J}$.

- Let $\mathcal{X}$ and $\mathcal{Y}$ be two finite dimensional real Euclidean spaces with Euclidean norms $\|\cdot\|_\mathcal{X}$ and $\|\cdot\|_\mathcal{Y}$, respectively, and $\mathcal{A} : \mathcal{X} \to \mathcal{Y}$ be a linear operator. Define the spectral (or operator) norm of $\mathcal{A}$ by $\|\mathcal{A}\| := \sup_{\|x\|_\mathcal{X}=1} \|\mathcal{A}(x)\|_\mathcal{Y}$.

Denote the range space of $\mathcal{A}$ by $\text{Range}(\mathcal{A}) := \{\mathcal{A}(x) \,|\, x \in \mathcal{X}\}$. Let $\mathcal{A}^*$ represent the adjoint of $\mathcal{A}$, i.e., $\mathcal{A}^* : \mathcal{Y} \to \mathcal{X}$ is the unique linear operator such that $\langle \mathcal{A}(x), y \rangle = \langle x, \mathcal{A}^*(y) \rangle$ for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$.

- Let $\mathbb{P}[\cdot]$ denote the probability of any given event, $\mathbb{E}[\cdot]$ denote the expectation of any given random variable, and $\text{cov}[\cdot]$ denote the covariance matrix of any given random vector.

- For any sets A and B, $A \setminus B$ denotes the relative complement of B in A, i.e., $A \setminus B := \{x \in A \,|\, x \notin B\}$.

Chapter 1

Introduction

High-dimensional structured recovery problems have attracted much attention in diverse fields such as statistics, machine learning, economics and finance. As its name suggests, the high-dimensional setting requires that the number of unknown parameters is comparable to or even much larger than the number of observations. Without any further assumption, statistical inference in this setting is faced with overwhelming difficulties – it is usually impossible to obtain a consistent estimate since the estimation error may not converge to zero with the dimension increasing, and what is worse, the relevant estimation problem is often underdetermined and thus ill-posed. The statistical challenges with high-dimensionality have been realized in different areas of sciences and humanities, ranging from computational biology and biomedical studies to data mining, financial engineering and risk management. For a comprehensive overview, one may refer to [52]. In order to make the relevant estimation problem meaningful and well-posed, various types of embedded low-dimensional structures, including sparse vectors, sparse and structured matrices, low-rank matrices, and their combinations, are imposed on the model. Thanks to these simple structures, we are able to treat high-dimensional problems in low-dimensional parameter spaces.

1.1 Problem and motivation

This thesis studies the problem of high-dimensional low-rank and sparse matrix decomposition with fixed and sampled basis coefficients. Specifically, this problem aims to recover an unknown low-rank matrix and an unknown sparse matrix from a small number of noiseless or noisy observations of the basis coefficients of their sum. In some circumstances, the sum of the unknown low-rank and sparse components may also have a certain structure so that some of its basis coefficients are known exactly in advance, which should be taken into consideration as well.

Such a matrix decomposition problem appears frequently in a lot of practical settings, with the low-rank and sparse components having different interpretations depending on the concrete applications, see, for example, [32, 21, 1] and references therein. In this thesis, we are particularly interested in the high-dimensional correlation matrix estimation problem with missing observations in factor models. As a tool for dimensionality reduction, factor models have been widely used both theoretically and empirically in economics and finance. See, e.g., [108, 109, 46, 29, 30, 39, 47, 48, 5]. In a factor model, the correlation matrix can be decomposed into a low-rank component corresponding to several common factors and a sparse component resulting from the idiosyncratic errors. Since any correlation matrix is a real symmetric and positive semidefinite matrix with all the diagonal entries being ones, the setting of fixed basis coefficients naturally occurs. Moreover, extra reliable prior information on certain off-diagonal entries or basis coefficients of the correlation matrix may also be available. For example, in a correlation matrix of exchange rates, the correlation coefficient between the Hong Kong dollar and the United States dollar can be fixed to one because of the linked exchange rate system implemented in Hong Kong for the stabilization purpose, which yields additional fixed basis coefficients.

Recently, there are plenty of theoretical researches focused on high-dimensional

low-rank and sparse matrix decomposition in both of the noiseless [32, 21, 61, 73, 89, 33, 124] and noisy [135, 73, 1] cases. To the best of our knowledge, however, the recovery performance under the setting of simultaneously having fixed and sampled basis coefficients remains unclear. Thus, we will go one step further to fill this gap by providing both exact and approximate recovery guarantees in this thesis.

1.2 Literature review

In the last decade, we have witnessed a lot of exciting and extraordinary progress in theoretical guarantees of high-dimensional structured recovery problems, such as compressed sensing for exact recovery of sparse vectors [27, 26, 43, 42], sparse linear regression using the LASSO for exact support recovery [95, 133, 121] and analysis of estimation error bound [96, 13, 102], low-rank matrix recovery for the noiseless case [105, 106] and the noisy case [24, 100] under different assumptions, like restricted isometry property (RIP), null space conditions, and restricted strong convexity (RSC), on the mapping of linear measurements, exact low-rank matrix completion [25, 28, 104, 68] with the incoherence conditions, and noisy low-rank matrix completion [101, 79] based on the notion of RSC. The establishment of these theoretical guarantees depends heavily on the convex nature of the corresponding formulations of the above problems, or specifically, the utilization of the ℓ_1-norm and the nuclear norm as the surrogates respectively for the sparsity of a vector and the rank of a matrix.

Given some information on a matrix that is formed by adding an unknown low-rank matrix to an unknown sparse matrix, the problem of retrieving the low-rank and sparse components can be viewed as a natural extension of the aforementioned sparse or low-rank structured recovery problems. Enlightened by the previous tremendous success of the convex approaches in using the ℓ_1-norm and

the nuclear norm, the "nuclear norm plus ℓ_1-norm" approach was first studied by Chandrasekaran et al. [32] for the case that the entries of the sum matrix are fully observed without noise. Their analysis is built on the notion of rank-sparsity incoherence, which is useful to characterize both fundamental identifiability and deterministic sufficient conditions for exact decomposition. Slightly later than the pioneered work [32] was released, Candès et al. [21] considered a more general setting with missing observations, and made use of the previous results and analysis techniques for the exact matrix completion problem [25, 104, 68] to provide probabilistic guarantees for exact recovery when the observation pattern is chosen uniformly at random. However, a non-vanishing fraction of entries is still required to be observed according to the recovery results in [21], which is almost meaningless in high-dimensional setting. Recently, Chen et al. [33] sharpened the analysis used in [21] to further the related research along this line. They established the first probabilistic exact decomposition guarantees that allow a vanishingly small fraction of observations. Nevertheless, as far as we know, there is no existing literature that concerns about recovery guarantees for this exact matrix decomposition problem with both fixed and sampled entries. In addition, it is worthwhile to mention that the problem of exact low-rank and diagonal matrix decomposition without any missing observation was investigated by Saunderson et al. [112], with interesting connections to the elliptope facial structure problem and the ellipsoid fitting problem, but the fully-observed model is too restricted.

All the recovery results reviewed above focus on the noiseless case. In a more realistic setting, the observed entries of the sum matrix are corrupted by a small amount of noise. This noisy low-rank and sparse matrix decomposition problem was first addressed by Zhou et al. [135] with a constrained formulation and later studied by Hsu et al. [73] in both of the constrained and penalized formulations. Very recently, Agarwal et al. [1] adopted the "nuclear norm plus ℓ_1-norm" penalized

least squares formulation and analyzed this problem based on the unified framework with the notion of RSC introduced in [102]. However, a full observation of the sum matrix is necessary for the recovery results obtained in [135, 73, 1], which may not be practical and useful in many applications.

Meanwhile, the nuclear norm penalization approach for noisy matrix completion was noticed to be significantly inefficient in some circumstances, see, e.g., [98, 99] and references therein. The similar challenges may yet be expected in the "nuclear norm plus ℓ_1-norm" penalization approach for noisy matrix decomposition. Therefore, how to go beyond the limitation of the nuclear norm in the noisy matrix decomposition problem also deserves our researches.

1.3 Contributions

From both of the theoretical and practical points of view, the main contributions of this thesis consist of three parts, which are summarized as follows.

Firstly, we study the problem of exact low-rank and sparse matrix decomposition with fixed and sampled basis coefficients. Based on the well-accepted "nuclear norm plus ℓ_1-norm" approach, we formulate this problem into convex programs, and then make use of their convex nature to establish exact recovery guarantees under the assumption of certain standard identifiability conditions for the low-rank and sparse components. Since only a vanishingly small fraction of samples is required as the intrinsic dimension increases, these probabilistic recovery results are particularly desirable in the high-dimensional setting. Although the analysis involved follows from the existing framework of dual certification, such recovery guarantees can still serve as the noiseless counterparts of those for the noisy case.

Secondly, we focus on the problem of noisy low-rank and sparse matrix decomposition with fixed and sampled basis coefficients. Inspired by the successful

recent development on the adaptive nuclear semi-norm penalization technique for noisy low-rank matrix completion [98, 99], we propose a two-stage rank-sparsity-correction procedure, and then examine its recovery performance by deriving, for the first time up to our knowledge, a non-asymptotic probabilistic error bound under the high-dimensional scaling. Moreover, as a by-product, we explore and prove a novel form of restricted strong convexity for the random sampling operator in the context of noisy low-rank and sparse matrix decomposition, which plays an essential and profound role in the recovery error analysis.

Thirdly, we specialize the aforementioned two-stage correction procedure to deal with the correlation matrix estimation problem with missing observations in strict factor models where the sparse component turns out to be diagonal. In this application, we provide a specialized recovery error bound and point out that this bound coincides with the optimal one in the best cases when the rank-correction function is constructed appropriately and the initial estimator is good enough, where by "optimal" we mean the circumstance that the true rank is known in advance. This fascinating finding together with the convincing numerical results indicates the superiority of the two-stage correction approach over the nuclear norm penalization.

1.4 Thesis organization

The remaining parts of this thesis are organized as follows. In Chapter 2, we introduce some preliminaries that are fundamental in the subsequent discussions, especially including a brief introduction on Bernstein-type inequalities for independent random variables and random matrices. In Chapter 3, we summarize the performance in terms of estimation error for the Lasso and related estimators in the context of high-dimensional sparse linear regression. In particular, we propose

a new Lasso-related estimator called the corrected Lasso. We then present non-asymptotic estimation error bounds for the Lasso-related estimators followed by a quantitative comparison. This study sheds light on the usage of the two-stage correction procedure in Chapter 5 and Chapter 6. In Chapter 4, we study the problem of exact low-rank and sparse matrix decomposition with fixed and sampled basis coefficients. After formulating this problem into concrete convex programs based on the "nuclear norm plus ℓ_1-norm" approach, we establish probabilistic exact recovery guarantees in the high-dimensional setting if certain standard identifiability conditions for the low-rank and sparse components are satisfied. In Chapter 5, we focus on the problem of noisy low-rank and sparse matrix decomposition with fixed and sampled basis coefficients. We propose a two-stage rank-sparsity-correction procedure via convex optimization, and then examine its recovery performance by developing a novel non-asymptotic probabilistic error bound under the high-dimensional scaling with the notion of restricted strong convexity. Chapter 6 is devoted to applying the specialized two-stage correction procedure, in both of the theoretical and computational aspects, to correlation matrix estimation with missing observations in strict factor models. Finally, we make the conclusions and point out several future research directions in Chapter 7.

Chapter 2

Preliminaries

In this chapter, we introduce some preliminary results that are fundamental in the subsequent discussions.

2.1 Basics in matrix analysis

This section collects some elementary but useful results in matrix analysis.

Lemma 2.1. *For any $X, Y \in \mathcal{S}_+^n$, it holds that*

$$\|X - Y\| \leq \max\{\|X\|, \|Y\|\}.$$

Proof. Since $X \succeq 0$ and $Y \succeq 0$, we have $X - Y \preceq X$ and $Y - X \preceq Y$. The proof then follows. □

Lemma 2.2. *Assume that $Z \in \mathbb{V}^{n_1 \times n_2}$ has at most k_1 nonzero entries in each row and at most k_2 nonzero entries in each column, where k_1 and k_2 are integers satisfying $0 \leq k_1 \leq n_1$ and $0 \leq k_2 \leq n_2$. Then we have*

$$\|Z\| \leq \sqrt{k_1 k_2}\|Z\|_\infty.$$

Proof. Notice that the spectral norm has the following variational characterization

$$\|Z\| = \sup\left\{x^T Z y \,\middle|\, \|x\|_2 = \|y\|_2 = 1,\, x \in \mathbb{R}^{n_1},\, y \in \mathbb{R}^{n_2}\right\}.$$

Then by using the Cauchy-Schwarz inequality, we obtain that

$$\begin{aligned}
\|Z\| &= \sup_{\|x\|_2=1,\, \|y\|_2=1} \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} Z_{ij}x_i y_j \\
&\le \sup_{\|x\|_2=1} \left(\sum_{i=1}^{n_1}\sum_{j:Z_{ij}\neq 0} x_i^2\right)^{\frac{1}{2}} \sup_{\|y\|_2=1} \left(\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} Z_{ij}^2 y_j^2\right)^{\frac{1}{2}} \\
&\le \sup_{\|x\|_2=1} \left(\sum_{i=1}^{n_1}\sum_{j:Z_{ij}\neq 0} x_i^2\right)^{\frac{1}{2}} \sup_{\|y\|_2=1} \left(\sum_{i:Z_{ij}\neq 0}\sum_{j=1}^{n_2} y_j^2\right)^{\frac{1}{2}} \|Z\|_\infty = \sqrt{k_1 k_2}\|Z\|_\infty,
\end{aligned}$$

where the last equality is due to the assumption. This completes the proof. □

2.2 Bernstein-type inequalities

In probability theory, the laws of large numbers state that the sample average of independent and identically distributed (i.i.d.) random variables is, under certain mild conditions, close to the expected value with high probability. As an extension, concentration inequalities provide probability bounds to measure how much a function of independent random variables deviates from its expectation. Among these inequalities, the Bernstein-type inequalities on sums of independent random variables or random matrices are the most basic and useful ones. We first start with the classical Bernstein's inequality [11].

Lemma 2.3. *Let $z_1, \ldots, z_m$ be independent random variables with mean zero. Assume that $|z_i| \le K$ almost surely for all $i = 1, \ldots, m$. Let $\varsigma_i^2 := \mathbb{E}[z_i^2]$ and $\varsigma^2 := \frac{1}{m}\sum_{i=1}^m \varsigma_i^2$. Then for any $t > 0$, we have*

$$\mathbb{P}\left[\left|\sum_{i=1}^m z_i\right| > t\right] \le 2\exp\left(-\frac{t^2/2}{m\varsigma^2 + Kt/3}\right).$$

Consequently, it holds that

$$\mathbb{P}\left[\left|\sum_{i=1}^{m} z_i\right| > t\right] \leq \begin{cases} 2\exp\left(-\frac{3}{8}\frac{t^2}{m\varsigma^2}\right), & \text{if } t \leq \frac{m\varsigma^2}{K}, \\ 2\exp\left(-\frac{3}{8}\frac{t}{K}\right), & \text{if } t > \frac{m\varsigma^2}{K}. \end{cases}$$

The assumption of boundedness in Lemma 2.3 is so restricted that many interesting cases are excluded, for example, the case when random variables are Gaussian. In fact, this assumption can be relaxed to include random variables with at least exponential tail decay. Such random variables are called sub-exponential.

Given any $s \geq 1$, let $\psi_s(x) := \exp(x^s) - 1$, for $x \geq 0$. The Orlicz ψ_s-norm (see, e.g., [118, pp. 95] and [81, Appendix A.1]) of any random variable z is defined as

$$\|z\|_{\psi_s} := \inf\left\{t > 0 \,|\, \mathbb{E}\psi_s(|z|/t) \leq 1\right\} = \inf\left\{t > 0 \,|\, \mathbb{E}\exp(|z|^s/t^s) \leq 2\right\}. \tag{2.1}$$

It is known that there are several equivalent definitions to define a sub-exponential random variable (cf. [120, Subsection 5.2.4]). One of these equivalent definitions is based on the Orlicz ψ_1-norm, which is also called the sub-exponential norm.

Definition 2.1. *A random variable z is called sub-exponential if there exists a constant $K > 0$ such that $\|z\|_{\psi_1} \leq K$.*

The Orlicz norms are useful to characterize the tail behavior of random variables. Below we state a Bernstein-type inequality for sub-exponential random variables [120, Proposition 5.16].

Lemma 2.4. *Let $z_1, \ldots, z_m$ be independent sub-exponential random variables with mean zero. Suppose that $\|z_i\|_{\psi_1} \leq K$ for all $i = 1, \ldots, m$. Then there exists a constant $C > 0$ such that for every $w = (w_1, \ldots, w_m)^T \in \mathbb{R}^m$ and every $t > 0$, we have*

$$\mathbb{P}\left[\left|\sum_{i=1}^{m} w_i z_i\right| > t\right] \leq 2\exp\left\{-C\min\left(\frac{t^2}{K^2\|w\|_2^2}, \frac{t}{K\|w\|_\infty}\right)\right\}.$$

Next, we introduce the powerful noncommutative Bernstein-type inequalities on random matrices, which play important roles in studying low-rank matrix recovery problems [104, 68, 115, 81, 83, 82, 101, 79]. The goal of these inequalities is to bound the tail probability on the spectral norm of the sum of independent zero-mean random matrices. The origin of these results can be traced back to the noncommutative version of the Chernoff bound developed by Ahlswede and Winter [2, Theorem 18] based on the Golden–Thompson inequality [64, 113]. Within the Ahlswede–Winter framework [2], different matrix extensions of the classical Bernstein's inequality were independently derived in [104, 68, 115] for random matrices with bounded spectral norm. The following standard noncommutative Bernstein inequality with slightly tighter constants is taken from [104, Theorem 4] .

Lemma 2.5. *Let $Z_1, \ldots, Z_m \in \mathbb{R}^{n_1 \times n_2}$ be independent random matrices with mean zero. Denote $\varsigma_i^2 := \max\{\|\mathbb{E}[Z_i Z_i^T]\|, \|\mathbb{E}[Z_i^T Z_i]\|\}$ and $\varsigma^2 := \frac{1}{m}\sum_{i=1}^m \varsigma_i^2$. Suppose that $\|Z_i\| \le K$ almost surely for all $i = 1, \ldots, m$. Then for every $t > 0$, we have*

$$\mathbb{P}\left[\left\|\sum_{i=1}^m Z_i\right\| > t\right] \le (n_1 + n_2)\exp\left(-\frac{t^2/2}{m\varsigma^2 + Kt/3}\right).$$

As a consequence, it holds that

$$\mathbb{P}\left[\left\|\sum_{i=1}^m Z_i\right\| > t\right] \le \begin{cases} (n_1+n_2)\exp\left(-\dfrac{3}{8}\dfrac{t^2}{m\varsigma^2}\right), & \text{if } t \le \dfrac{m\varsigma^2}{K}, \\ (n_1+n_2)\exp\left(-\dfrac{3}{8}\dfrac{t}{K}\right), & \text{if } t > \dfrac{m\varsigma^2}{K}. \end{cases}$$

Very recently, the noncommutative Bernstein-type inequalities were extended by replacing the assumption of bounded spectral norm with bounded Orlicz ψ_s-norm [81, 83, 82, 101, 79]. The next lemma is tailored from [81, pp. 30].

Lemma 2.6. *Let $Z_1, \ldots, Z_m \in \mathbb{R}^{n_1 \times n_2}$ be independent random matrices with mean zero. Suppose that $\max\left\{\left\|\|Z_i\|\right\|_{\psi_s}, 2\mathbb{E}^{\frac{1}{2}}[\|Z_i\|^2]\right\} \le K_s$ for some constant $K_s > 0$*

and for all $i = 1, \ldots, m$. *Define*

$$\varsigma := \max\left\{ \left\| \frac{1}{m} \sum_{i=1}^{m} \mathbb{E}\left[Z_i Z_i^T\right] \right\|^{\frac{1}{2}}, \ \left\| \frac{1}{m} \sum_{i=1}^{m} \mathbb{E}\left[Z_i^T Z_i\right] \right\|^{\frac{1}{2}} \right\}.$$

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

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

As noted by Vershynin [119], the following slightly weaker but simpler noncommutative Bernstein-type inequality can be established under the bounded Orlicz ψ_1-norm assumption. Its proof depends on a noncommutative Chernoff bound [104, Theorem 10] and an upper bound of the moment generating function of subexponential random variables [120, Lemma 5.15].

Lemma 2.7. *Let* $Z_1, \ldots, Z_m \in \mathbb{R}^{n_1 \times n_2}$ *be independent random matrices with mean zero. Suppose that* $\left\| \|Z_i\| \right\|_{\psi_1} \leq K$ *for some constant* $K > 0$ *and for all* $i = 1, \ldots, m$. *Then there exists a constant* $C > 0$ *such that for any* $t > 0$, *we have*

$$\mathbb{P}\left[\left\| \sum_{i=1}^{m} Z_i \right\| > t \right] \leq (n_1 + n_2) \exp\left\{ -C \min\left(\frac{t^2}{mK^2}, \ \frac{t}{K} \right) \right\}.$$

Proof. Define the independent symmetric random matrices by

$$W_i := \begin{bmatrix} 0 & Z_i \\ Z_i^T & 0 \end{bmatrix}, \quad i = 1, \ldots, m.$$

Since $\|Z_i\| \equiv \|W_i\|$, it holds that $\left\| \|Z_i\| \right\|_{\psi_1} = \left\| \|W_i\| \right\|_{\psi_1}$, for all $i = 1, \ldots, m$. Moreover, the spectral norm of $\sum_{i=1}^{m} Z_i$ is equal to the maximum eigenvalue of $\sum_{i=1}^{m} W_i$. By using [104, Theorem 10], we have that for all $\tau > 0$,

$$\mathbb{P}\left[\left\| \sum_{i=1}^{m} Z_i \right\| > t \right] = \mathbb{P}\left[\sum_{i=1}^{m} W_i \not\preceq t I_{n_1+n_2} \right] \leq (n_1+n_2) \exp(-\tau t) \prod_{i=1}^{m} \left\| \mathbb{E}\left[\exp(\tau W_i) \right] \right\|.$$

Since $0 \preceq \exp(\tau W_i) \preceq \exp(\tau\|W_i\|)I_{n_1+n_2}$, we know that

$$\|\mathbb{E}\left[\exp(\tau W_i)\right]\| \leq \mathbb{E}\exp(\tau\|W_i\|).$$

According to [120, Lemma 5.15], there exists some constants $c_1, c_2 > 0$ such that for $|\tau| \leq c_1/K$, it holds that

$$\mathbb{E}\exp(\tau\|W_i\|) \leq \exp\left(c_2\tau^2\left\|\|W_i\|\right\|_{\psi_1}^2\right) \leq \exp\left(c_2\tau^2K^2\right), \quad i = 1, \ldots, m.$$

Putting all these together gives that

$$\mathbb{P}\left[\left\|\sum_{i=1}^m Z_i\right\| > t\right] \leq (n_1+n_2)\exp\left(-t\tau + c_2m\tau^2K^2\right).$$

The optimal choice of the parameter τ is obtained by minimizing $-t\tau + c_2m\tau^2K^2$ as a function of τ subject to the bound constraint $0 \leq \tau \leq c_1/K$, yielding that $\tau^* = \min\{t/(2c_2mK^2), c_1/K\}$ and

$$\mathbb{P}\left[\left\|\sum_{i=1}^m Z_i\right\| > t\right] \leq (n_1+n_2)\exp\left\{-\min\left(\frac{t^2}{4c_2mK^2}, \frac{c_1t}{2K}\right)\right\}.$$

The proof is then completed by choosing the constant $C = \min\{1/(4c_2), c_1/2\}$. □

2.3 Random sampling model

In the problem of low-rank matrix recovery from randomly sampled entries, the model of uniform sampling without replacement is a commonly-used assumption for the recovery results established in [25, 28, 23, 76, 77, 104, 68, 21]. As its name suggests, sampling is called without replacement if each sample is selected at random from the population set and it is not put back. Moreover, a subset of size m obtained by sampling uniformly at random without replacement from the set of size d (assuming that $m \leq d$) means that this sample subset is chosen uniformly at random from all the size-m subsets of the population set.

Evidently, samples selected from the model of sampling without replacement are not independent of each other, which could make the relevant analysis complicated. Therefore, in the proof of recovery results given in [25, 28, 23, 76, 77, 21], a Bernoulli sampling model, where each element from the population set is selected independently with probability equal to p, was studied as a proxy for the model of uniform sampling without replacement. The theoretical guarantee for doing so relies heavily on the equivalence between these two sampling models – the failure probability of recovery under the uniform sampling model without replacement is closely approximated by the failure probability of recovery under the Bernoulli sampling model with $p = \frac{m}{d}$ (see, e.g., [26, Section II.C]). Due to this equivalence, Bernoulli sampling was also directly assumed in some recent work related to the problem of low-rank matrix recovery from randomly sampled entries [89, 33].

Another popular proxy is the model of uniform sampling with replacement, where a sample chosen uniformly at random from the population set is put back and then a second sample is chosen uniformly at random from the unchanged population set. In this sampling model, samples are independent, and any element of the population set may be chosen more than once in general. Just similar to Bernoulli sampling, with the same sample size, the failure probability of recovery under uniform sampling without replacement is upper bounded by the failure probability of recovery under uniform sampling with replacement (cf. [104, Proposition 3] and [68, Section II.A]). This makes the novel techniques for analyzing the matrix completion problem in [104, 68] via the powerful noncommutative Bernstein-type inequalities possible. In addition, it is worth mentioning that sampling with replacement is also widely used for the problem of noisy low-rank matrix recovery in the statistics community [107, 83, 82, 101, 79].

Lastly, it is interesting to note that a noncommutative Bernstein-type inequality was established by [70] in the context of sampling without replacement.

2.4 Tangent space to the set of rank-constrained matrices

Let $X \in \mathbb{V}^{n_1 \times n_2}$ be given and admit a reduced singular value decomposition (SVD) given by $X = U_1 \Sigma V_1^T$, where r is the rank of X, $U_1 \in \mathcal{O}^{n_1 \times r}$, $V_1 \in \mathcal{O}^{n_2 \times r}$, and $\Sigma \in \mathbb{R}^{r \times r}$ is the diagonal matrix with the non-zero singular values of X being arranged in the non-increasing order. The tangent space to the set of rank-constrained matrices $\{Z \in \mathbb{V}^{n_1 \times n_2} \,|\, \mathrm{rank}(Z) \leq r\}$ at X is the linear span of all matrices with either the same row-space as X or the same column-space as X (cf. [32, Section 3.2] and [31, Section 2.3]). Specifically, the tangent space $\mathcal{T}$ and its orthogonal complement $\mathcal{T}^\perp$ can be expressed as

$$\mathcal{T} = \begin{cases} \left\{U_1 B^T + A V_1^T \,|\, A \in \mathbb{R}^{n_1 \times r},\, B \in \mathbb{R}^{n_2 \times r}\right\}, & \text{if } \mathbb{V}^{n_1 \times n_2} = \mathbb{R}^{n_1 \times n_2}, \\ \left\{U_1 A^T + A U_1^T \,|\, A \in \mathbb{R}^{n \times r}\right\}, & \text{if } \mathbb{V}^{n_1 \times n_2} = \mathcal{S}^n, \end{cases} \tag{2.2}$$

and

$$\mathcal{T}^\perp = \begin{cases} \left\{C \in \mathbb{V}^{n_1 \times n_2} \,|\, U_1^T C = 0,\, C V_1 = 0\right\}, & \text{if } \mathbb{V}^{n_1 \times n_2} = \mathbb{R}^{n_1 \times n_2}, \\ \left\{C \in \mathbb{V}^{n_1 \times n_2} \,|\, U_1^T C = 0\right\}, & \text{if } \mathbb{V}^{n_1 \times n_2} = \mathcal{S}^n. \end{cases} \tag{2.3}$$

Moreover, choose U_2 and V_2 such that $U = [U_1, U_2]$ and $V = [V_1, V_2]$ are both orthogonal matrices. Notice that $U = V$ when $X \in \mathcal{S}_+^n$. For any $Z \in \mathbb{V}^{n_1 \times n_2}$,

$$\begin{aligned} Z &= [U_1, U_2][U_1, U_2]^T Z [V_1, V_2][V_1, V_2]^T \\ &= [U_1, U_2] \begin{bmatrix} U_1^T Z V_1 & U_1^T Z V_2 \\ U_2^T Z V_1 & 0 \end{bmatrix} [V_1, V_2]^T + [U_1, U_2] \begin{bmatrix} 0 & 0 \\ 0 & U_2^T Z V_2 \end{bmatrix} [V_1, V_2]^T \\ &= \left(U_1 U_1^T Z + Z V_1 V_1^T - U_1 U_1^T Z V_1 V_1^T\right) + U_2 U_2^T Z V_2 V_2^T. \end{aligned}$$

Then the orthogonal projections of any $Z \in \mathbb{V}^{n_1 \times n_2}$ onto $\mathcal{T}$ and $\mathcal{T}^\perp$ are given by

$$\mathcal{P}_{\mathcal{T}}(Z) = U_1 U_1^T Z + Z V_1 V_1^T - U_1 U_1^T Z V_1 V_1^T \tag{2.4}$$

and

$$\mathcal{P}_{\mathcal{T}^\perp}(Z) = U_2 U_2^T Z V_2 V_2^T. \tag{2.5}$$

This directly implies that for any $Z \in \mathbb{V}^{n_1 \times n_2}$,

$$\|\mathcal{P}_{\mathcal{T}}(Z)\| \leq 2\|Z\| \quad \text{and} \quad \|\mathcal{P}_{\mathcal{T}^\perp}(Z)\| \leq \|Z\|. \tag{2.6}$$

Note that $\mathcal{P}_{\mathcal{T}}$ and $\mathcal{P}_{\mathcal{T}^\perp}$ are both self-adjoint, and $\|\mathcal{P}_{\mathcal{T}}\| = \|\mathcal{P}_{\mathcal{T}^\perp}\| = 1$. With these definitions, the subdifferential of the nuclear norm $\|\cdot\|_*$ at the given X (see, e.g., [122]) can be characterized as follows

$$\partial\|X\|_* = \big\{ Z \in \mathbb{V}^{n_1 \times n_2} \,|\, \mathcal{P}_{\mathcal{T}}(Z) = U_1 V_1^T \text{ and } \|\mathcal{P}_{\mathcal{T}^\perp}(Z)\| \leq 1 \big\}.$$

Chapter 3

The Lasso and related estimators for high-dimensional sparse linear regression

This chapter is devoted to summarizing the performance in terms of estimation error for the Lasso and related estimators in the context of high-dimensional sparse linear regression. In particular, we propose a new Lasso-related estimator called the corrected Lasso, which is enlightened by a two-step majorization technique for nonconvex regularizers. We then present non-asymptotic estimation error bounds for the Lasso-related estimators under different assumptions on the design matrix. Finally, we make a quantitative comparison among these error bounds, which sheds light on the two-stage correction procedure later studied in Chapter 5 and applied in Chapter 6.

3.1 Problem setup and estimators

In this section, we start with the problem of high-dimensional sparse linear regression, and then introduce the formulations of the Lasso and related estimators.

3.1.1 The linear model

Suppose that $\bar{x} \in \mathbb{R}^n$ is an unknown vector with the supporting index set $S := \{j \,|\, \bar{x}_j \neq 0,\, j = 1, \ldots, n\}$. Let s be the cardinality of S, i.e., $s := |S|$. Suppose also that $\bar{x}$ is a sparse vector in the sense that $s \ll n$. In this chapter, we consider the statistical linear model, in which we observe a response vector of m noisy measurements $y = (y_1, \ldots, y_m)^T \in \mathbb{R}^m$ of the form

$$y = A\bar{x} + \xi, \tag{3.1}$$

where $A \in \mathbb{R}^{m \times n}$ is referred to as the design matrix or covariate matrix, and $\xi = (\xi_1, \ldots, \xi_m)^T \in \mathbb{R}^m$ is a vector containing random noises. Given the data y and A, the problem of high-dimensional sparse linear regression seeks an accurate and sparse estimate of $\bar{x}$ based on the observation model (3.1), where there are much more variables than observations, i.e., $n \gg m$.

For convenience of the following discussion, we assume that the noise vector ξ are of i.i.d. zero-mean sub-Gaussian entries. In particular, the class of sub-Gaussian random variables contains the centered Gaussian and all bounded random variables. The definition for a sub-Gaussian random variable is borrowed from [18, Definition 1.1 in Chapter 1], [62, Section 12.7] and [120, Subsection 5.2.3].

Definition 3.1. *A random variable z is called sub-Gaussian if there exists a constant $\varsigma \in [0, +\infty)$ such that for all $t \in \mathbb{R}$,*

$$\mathbb{E}[\exp(tz)] \leq \exp\left(\frac{\varsigma^2 t^2}{2}\right).$$

The exponent $\varsigma(z)$ of the sub-Gaussian random variable z is defined as

$$\varsigma(z) := \inf\left\{\varsigma \geq 0 \,|\, \mathbb{E}[\exp(tz)] \leq \exp\left(\varsigma^2 t^2/2\right) \text{ for all } t \in \mathbb{R}\right\}.$$

Assumption 3.1. *The additive noise vector $\xi \in \mathbb{R}^m$ has i.i.d. entries that are zero-mean and sub-Gaussian with exponent $\nu > 0$.*

Since sub-Gaussian random variables are sub-exponential, a tighter version of the large deviation inequality in Lemma 2.4 is satisfied (cf. [18, Chapter 1], [62, Section 12.7] and [120, Subsection 5.2.3]).

Lemma 3.1. *Under Assumption 3.1, for any fixed $w \in \mathbb{R}^m$, it holds that,*

$$\mathbb{P}\left[|\langle w, \xi\rangle| \geq t\right] \leq 2\exp\left(-\frac{t^2}{2\nu^2\|w\|_2^2}\right), \quad \forall\ t > 0.$$

3.1.2 The Lasso and related estimators

In the high-dimensional setting, the classical ordinary least squares (OLS) estimator usually fails, because it may not be well-defined, and most importantly, it does not take the structural assumption that $\bar{x}$ is sparse into account. In order to effectively utilizes the sparsity assumption, the ℓ_1-norm penalized least squares estimator, also known as the least absolute shrinkage and selection operator (Lasso) [114], is defined as

$$\mathring{x} \in \arg\min_{x\in\mathbb{R}^n} \frac{1}{2m}\|y - Ax\|_2^2 + \rho_m\|x\|_1, \tag{3.2}$$

where $\rho_m > 0$ is a penalization parameter that controls the tradeoff between the least squares fitting and the sparsity level. The Lasso is very popular and powerful for sparse linear regression in statistics and machine learning, partially owing to the superb invention of the efficient LARS algorithm [45]. Under some conditions [59, 95, 133, 136, 126, 121, 22, 116, 44], the Lasso is capable of exactly recovering the true supporting index set, which is a favorable property in model selection. However, it is well-known that the ℓ_1-norm penalty can create unnecessary biases for large coefficients, leading to a remarkable disadvantage that the optimal estimation accuracy is hardly achieved [4, 51].

For the purpose of reducing or removing biasedness, nonconvex penalization methods were suggested and studied, see e.g., [49, 4, 51, 56, 91, 127, 131, 55, 129,

132]. To facilitate the following analysis, we focus on a concise form of nonconvex penalized least-squares given by

$$\min_{x\in\mathbb{R}^n} \frac{1}{2m}\|y-Ax\|_2^2 + \rho_m \sum_{j=1}^n h(|x_j|), \tag{3.3}$$

where $h : \mathbb{R}_+ \to \mathbb{R}_+$ is a non-decreasing concave function with $h(0) = 0$. Let the left derivative and the right derivative of h be denoted by h'_- and h'_+, and set $h'_-(0) = h'_+(0)$. Clearly, the monotonicity and concavity of h implies that $0 \le h'_+(t_2) \le h'_-(t_2) \le h'_+(t_1) \le h'_-(t_1)$ for any $0 \le t_1 < t_2$. By utilizing the classical Majorization-Minimization (MM) algorithm [35, 36, 72, 84, 75, 125], the nonconvex problem (3.3) may be solved iteratively via certain multi-stage convex relaxations. In particular, a two-stage procedure has been shown to enjoy the desired asymptotic oracle properties, provided that the initial estimator is good enough [136, 137]. In this chapter, we consider two majorization techniques using the Lasso as the initial estimator.

The first one is to majorize the concave penalty function based on a separable local linear approximation [137]:

$$\sum_{j=1}^n h(|x_j|) \le \sum_{j=1}^n \big[h(|\mathring{x}_j|) + \breve{w}_j(|x_j| - |\mathring{x}_j|)\big] = \sum_{j=1}^n h(|\mathring{x}_j|) + \langle \breve{w}, |x| - |\mathring{x}|\rangle,$$

where $\breve{w}_j \in [h'_+(|\mathring{x}_j|),\, h'_-(|\mathring{x}_j|)]$. Note that the weight vector $\breve{w} \ge 0$. This majorization results in a convex relaxation of the nonconvex problem (3.3) given as

$$\breve{x} \in \arg\min_{x\in\mathbb{R}^n} \frac{1}{2m}\|y-Ax\|_2^2 + \rho_m\langle \breve{w}, |x|\rangle, \tag{3.4}$$

which we call the weighted Lasso. The weighted Lasso includes the adaptive Lasso [136] as a special case, where $\breve{x}_j$ is automatically set to 0 if $\breve{w}_j = +\infty$.

The second one is the two-step majorization initiated by [98], in which the first step is in accordance with a nonseparable local linear approximation:

$$\sum_{j=1}^n h(|x_j|) \le \sum_{j=1}^n \big[h(|\mathring{x}|_j^{\downarrow}) + \varpi_j(|x|_j^{\downarrow} - |\mathring{x}|_j^{\downarrow})\big] = \sum_{j=1}^n h(|\mathring{x}_j|) + \langle \varpi, |x|^{\downarrow} - |\mathring{x}|^{\downarrow}\rangle,$$

where $\varpi_j \in [h'_+(|\mathring{x}|_j^{\downarrow}), h'_-(|\mathring{x}|_j^{\downarrow})]$ satisfying that $\varpi_j = \varpi_{j'}$ if $|\mathring{x}|_j^{\downarrow} = |\mathring{x}|_{j'}^{\downarrow}$. Then $\varpi = \varpi^{\uparrow} \geq 0$, and thus $\langle \varpi, |x|^{\downarrow} \rangle$ is in general not a convex function about x. In this case, we further assume that $0 < h'_+(0) < +\infty$. Define $\mu := h'_+(0)$ and $\omega := \mathbf{1} - \varpi/\mu \in \mathbb{R}^n$. Observe that

$$\langle \varpi, |x|^{\downarrow} \rangle = \langle \mu\mathbf{1} - (\mu\mathbf{1} - \varpi), |x|^{\downarrow} \rangle = \mu(\|x\|_1 - \|x\|_\omega),$$

where $\|\cdot\|_\omega$ is the ω-weighted function defined by $\|x\|_\omega := \sum_{j=1}^n \omega_j |x|_j^{\downarrow}$ for any $x \in \mathbb{R}^n$. The ω-weighted function is indeed a norm because $\omega = \omega^{\downarrow} \geq 0$. This shows that $\langle \varpi, |\cdot|^{\downarrow} \rangle$ is a difference of two convex functions. Therefore, the second step is to linearize the ω-weighted function as follows:

$$\|x\|_\omega \geq \|\mathring{x}\|_\omega + \langle \hat{w}, x - \mathring{x} \rangle,$$

for any $\hat{w} \in \partial\|\mathring{x}\|_\omega$, where $\partial\|\cdot\|_\omega$ is the subdifferential of the convex function $\|\cdot\|_\omega$ and its characterization can be found in [98, Theorem 2.5]. A particular choice of the subgradient $\hat{w}$ is

$$\hat{w} = \operatorname{sign}(\mathring{x}) \circ \omega_{\pi^{-1}},$$

where π is a permutation of $\{1, \ldots, n\}$ such that $|\mathring{x}|^{\downarrow} = |\mathring{x}|_\pi$, i.e., $|\mathring{x}|_j^{\downarrow} = |\mathring{x}|_{\pi(j)}$ for $j = 1, \ldots, n$, and π^{-1} be the inverse of π. Note that $\operatorname{sign}(\hat{w}) = \operatorname{sign}(\mathring{x})$ and $0 \leq |\hat{w}| \leq \mathbf{1}$. Finally, the two-step majorization yields another convex relaxation of the nonconvex problem (3.3) formulated as

$$\hat{x} \in \arg\min_{x \in \mathbb{R}^n} \frac{1}{2m}\|y - Ax\|_2^2 + \hat{\rho}_m\big(\|x\|_1 - \langle \hat{w}, x \rangle\big), \tag{3.5}$$

where $\hat{\rho}_m := \rho_m \mu$. Since $0 \leq |\hat{w}| \leq \mathbf{1}$, the penalization term $\|x\|_1 - \langle \hat{w}, x \rangle$ is indeed a semi-norm. Similar to the terminologies used in [98, 99], we refer to this estimator as the corrected Lasso, the linear term $-\langle \hat{w}, x \rangle$ as the sparsity-correction term, and the penalization technique as the adaptive ℓ_1 semi-norm penalization technique. To our knowledge, there is not yet any study on the corrected Lasso before.

Another remedy to deal with the biasedness issue is a simple two-stage procedure: apply the OLS estimator to the model, i.e., the supporting index set, selected by the Lasso (see, e.g., [22, 134, 117, 10]). This two-stage estimator is called the thresholded Lasso, which is specifically defined by

$$\tilde{x}_{\mathcal{J}} \in \arg \min_{x_{\mathcal{J}} \in \mathbb{R}^{|\mathcal{J}|}} \frac{1}{2}\|y - A_{\mathcal{J}} x_{\mathcal{J}}\|_2^2 \quad \text{and} \quad \tilde{x}_{\mathcal{J}^c} = 0, \tag{3.6}$$

where $\mathcal{J} := \{j \,|\, |\mathring{x}|_j \geq T,\, j = 1, \ldots, n\}$ for a given threshold $T \geq 0$, and $A_{\mathcal{J}} \in \mathbb{R}^{m \times |\mathcal{J}|}$ is the matrix consisting of the columns of A indexed by $\mathcal{J}$. For the case when the index set $\mathcal{J}$ turns out to be the true supporting index set S, we call this estimator the oracle thresholded Lasso. Obviously, the estimation error bound for the oracle thresholded Lasso is the best that can be expected. Other similar two-stage procedures can be found in [94, 130].

In the rest of this chapter, under different assumptions on the design matrix, we derive non-asymptotic estimation error bounds for the Lasso (3.2), the weighted Lasso (3.4), the corrected Lasso (3.5) and the oracle thresholded Lasso (3.6) by following the unified framework provided in [102] for high-dimensional analysis of M-estimators with decomposable regularizers. Through a quantitative comparison among these error bounds, we verify the estimation performance of the weighted Lasso and the corrected Lasso by revealing that both of them are able to reduce the estimation error bound significantly compared to that for the Lasso and get very close to the optimal estimation error bound achieved by the oracle thresholded Lasso. This comparison sheds light on the two-stage correction procedure later studied in Chapter 5 and applied in Chapter 6.

3.2 Deterministic design

In this section, we consider estimation error bounds for the aforementioned Lasso-related estimators when the design matrix is deterministic or non-random.

In the beginning, we make two assumptions on the design matrix. The first one is essentially equivalent to the restricted eigenvalue (RE) condition originally developed in [13]. For a given index subset $\mathcal{J} \subseteq \{1, \ldots, n\}$ and a given constant $\alpha \geq 1$, define the set

$$\mathcal{C}(\mathcal{J}; \alpha) := \{\delta \in \mathbb{R}^n \mid \|\delta_{\mathcal{J}^c}\|_1 \leq \alpha \|\delta_{\mathcal{J}}\|_1\}.$$

Assumption 3.2. *There exists a constant $\lambda_{S,\kappa} > 0$ such that*

$$\frac{1}{m}\|A\delta\|_2^2 \geq \lambda_{S,\kappa}\|\delta\|_2^2, \quad \forall\ \delta \in \mathcal{C}(S; \alpha_\kappa),$$

where $\alpha_\kappa := (\kappa+1)/(\kappa-1)$ for some given constant $\kappa > 0$. In this case, we say that the design matrix $A \in \mathbb{R}^{m\times n}$ satisfies the RE condition over the true supporting index set S with parameters $(\alpha_\kappa, \lambda_{S,\kappa})$.

Various conditions, other than the RE condition, for analyzing the recovery or estimation performance of ℓ_1-norm based methods include the restricted isometry property (RIP) [27], the sparse Riesz condition [128], and the incoherent design condition [96]. As shown in [13], the RE condition is one of the weakest and hence the most general conditions in literature imposed on the design matrix to establish error bounds for the Lasso estimator. One may refer to [116] for an extensive study of different types of restricted eigenvalue or compatibility conditions.

The second assumption is the column-normalized condition for the design matrix. This condition does not incur any loss of generality, because the linear model (3.1) can always be appropriately rescaled to such a normalized setting. For $j = 1, \ldots, n$, denote the j-th column of A by A_j.

Assumption 3.3. *The columns of the design matrix A are normalized such that $\|A_j\|_2 \leq \sqrt{m}$ for all $j = 1, \ldots, n$.*

Under Assumption 3.1, 3.2 and 3.3, we first derive an estimation error bound for the Lasso according to the unified framework by [102]. This kind of estimation

performance analysis has been theoretically considered by the statistics community. See, e.g., [19, 20, 80, 96, 13, 22, 102].

Lemma 3.2. *Consider the linear model (3.1) under Assumption 3.2 with a given constant $\kappa > 1$. If $\rho_m \geq \kappa \|\frac{1}{m} A^T \xi\|_\infty$, then the Lasso estimator (3.2) satisfies the bound*

$$\|\mathring{x} - \bar{x}\|_2 \leq 2\left(1 + \frac{1}{\kappa}\right) \frac{\sqrt{s}\rho_m}{\lambda_{S,\kappa}}.$$

Proof. Let $\mathring{\delta} := \mathring{x} - \bar{x}$. Since $\mathring{x}$ is an optimal solution to problem (3.2) and $\bar{x}$ satisfies the linear model (3.1), we have that

$$\frac{1}{2m}\|A\mathring{\delta}\|_2^2 \leq \left\langle \frac{1}{m} A^T \xi, \mathring{\delta} \right\rangle - \rho_m\big(\|\bar{x} + \mathring{\delta}\|_1 - \|\bar{x}\|_1\big). \tag{3.7}$$

Since $\rho_m \geq \kappa \|\frac{1}{m} A^T \xi\|_\infty$ with $\kappa > 1$, it holds that

$$\left\langle \frac{1}{m} A^T \xi, \mathring{\delta} \right\rangle \leq \left\| \frac{1}{m} A^T \xi \right\|_\infty \|\mathring{\delta}\|_1 \leq \frac{\rho_m}{\kappa} \|\mathring{\delta}\|_1 = \frac{\rho_m}{\kappa}\big(\|\mathring{\delta}_S\|_1 + \|\mathring{\delta}_{S^c}\|_1\big). \tag{3.8}$$

From the characterization of $\partial\|\bar{x}\|_1$, we derive that

$$\|\bar{x} + \mathring{\delta}\|_1 - \|\bar{x}\|_1 \geq \big\langle \mathrm{sign}(\bar{x}_S), \mathring{\delta}_S \big\rangle + \sup_{\eta \in \mathbb{R}^{n-s},\, \|\eta\|_\infty \leq \mathbf{1}} \big\langle \eta, \mathring{\delta}_{S^c} \big\rangle \geq -\|\mathring{\delta}_S\|_1 + \|\mathring{\delta}_{S^c}\|_1. \tag{3.9}$$

Then by substituting (3.8) and (3.9) into (3.7), we obtain that

$$\begin{aligned} \frac{1}{2m}\|A\mathring{\delta}\|_2^2 &\leq \frac{\rho_m}{\kappa}\big(\|\mathring{\delta}_S\|_1 + \|\mathring{\delta}_{S^c}\|_1\big) + \rho_m\big(\|\mathring{\delta}_S\|_1 - \|\mathring{\delta}_{S^c}\|_1\big) \\ &= \rho_m\left(1 + \frac{1}{\kappa}\right)\|\mathring{\delta}_S\|_1 - \rho_m\left(1 - \frac{1}{\kappa}\right)\|\mathring{\delta}_{S^c}\|_1. \end{aligned} \tag{3.10}$$

Thus, $\mathring{\delta} \in \mathcal{C}(S; \alpha_\kappa)$. Since $\kappa > 1$, Assumption 3.2 and (3.10) yield that

$$\frac{1}{2}\lambda_{S,\kappa}\|\mathring{\delta}\|_2^2 \leq \rho_m\left(1 + \frac{1}{\kappa}\right)\|\mathring{\delta}_S\|_1 \leq \sqrt{s}\rho_m\left(1 + \frac{1}{\kappa}\right)\|\mathring{\delta}\|_2,$$

which completes the proof. □

Proposition 3.3. *Let $\kappa > 1$ and $c > 0$ be given constants. Consider the linear model (3.1) under Assumption 3.1, 3.2 and 3.3. If the penalization parameter*

$$\rho_m = \kappa\nu\sqrt{2+2c}\sqrt{\frac{\log n}{m}},$$

then with probability at least $1-2n^{-c}$, the Lasso estimator (3.2) satisfies the bound

$$\|\mathring{x} - \bar{x}\|_2 \leq \sqrt{1+c}(1+\kappa)\frac{2\sqrt{2}\nu}{\lambda_{S,\kappa}}\sqrt{\frac{s\log n}{m}}.$$

Proof. In light of Lemma 3.2, it suffices to show that $\rho_m = \kappa\nu\sqrt{2+2c}\sqrt{\log n/m} \geq \kappa\|\frac{1}{m}A^T\xi\|_\infty$ with probability at least $1-2n^{-c}$. By using Assumption 3.1, Lemma 3.1 and Assumption 3.3, for $j = 1, \ldots, n$, we have the tail bound

$$\mathbb{P}\left[\frac{|\langle A_j, \xi\rangle|}{m} \geq t\right] \leq 2\exp\left(-\frac{m^2t^2}{2\nu^2\|A_j\|_2^2}\right) \leq 2\exp\left(-\frac{mt^2}{2\nu^2}\right).$$

Choose $t^* = \nu\sqrt{2+2c}\sqrt{\log n/m}$. Then taking the union bound gives that

$$\mathbb{P}\left[\left\|\frac{1}{m}A^T\xi\right\|_\infty \geq t^*\right] \leq 2n\exp\left(-\frac{mt^{*2}}{2\nu^2}\right) = \frac{2}{n^c}.$$

This completes the proof. □

We next derive an estimation error bound for the weighted Lasso in a similar way to that for the Lasso. Analogous results for the adaptive Lasso can be found in [117]. Define

$$\check{w}_S^{\max} := \max_{j\in S}\check{w}_j \quad \text{and} \quad \check{w}_{S^c}^{\min} := \min_{j\in S^c}\check{w}_j. \tag{3.11}$$

Lemma 3.4. *Consider the linear model (3.1) under Assumption 3.2 with a given constant $\kappa > 1$. Suppose that $\check{w}_{S^c}^{\min} \geq \check{w}_S^{\max}$ and $\rho_m\check{w}_{S^c}^{\min} \geq \kappa\|\frac{1}{m}A^T\xi\|_\infty$. Then the weighted Lasso estimator (3.4) satisfies the bound*

$$\|\check{x} - \bar{x}\|_2 \leq 2\left(\frac{\check{w}_S^{\max}}{\check{w}_{S^c}^{\min}} + \frac{1}{\kappa}\right)\frac{\sqrt{s}\rho_m\check{w}_{S^c}^{\min}}{\lambda_{S,\kappa}}.$$

Proof. Let $\check{\delta} := \check{x} - \bar{x}$. Since $\check{x}$ is an optimal solution to problem (3.4) and $\bar{x}$ satisfies the linear model (3.1), we have that

$$\frac{1}{2m}\|A\check{\delta}\|_2^2 \leq \Big\langle \frac{1}{m}A^T\xi, \check{\delta}\Big\rangle - \rho_m\big(\langle \check{w}, |\bar{x}+\check{\delta}|\rangle - \langle \check{w}, |\bar{x}|\rangle\big). \tag{3.12}$$

Since $\rho_m \check{w}_{S^c}^{\min} \geq \kappa \|\frac{1}{m}A^T\xi\|_\infty$ with $\kappa > 1$, it holds that

$$\Big\langle \frac{1}{m}A^T\xi, \check{\delta}\Big\rangle \leq \Big\|\frac{1}{m}A^T\xi\Big\|_\infty \|\check{\delta}\|_1 \leq \frac{\rho_m \check{w}_{S^c}^{\min}}{\kappa}\|\check{\delta}\|_1 = \frac{\rho_m \check{w}_{S^c}^{\min}}{\kappa}\big(\|\check{\delta}_S\|_1 + \|\check{\delta}_{S^c}\|_1\big). \tag{3.13}$$

From the characterization of $\partial\langle \check{w}, |\bar{x}|\rangle$, we derive that

$$\begin{aligned}
\langle \check{w}, |\bar{x}+\check{\delta}|\rangle - \langle \check{w}, |\bar{x}|\rangle &\geq \big\langle \check{w}_S \circ \operatorname{sign}(\bar{x}_S), \check{\delta}_S\big\rangle + \sup_{\eta\in\mathbb{R}^{n-s},\, |\eta|\leq \check{w}_{S^c}} \big\langle \eta, \check{\delta}_{S^c}\big\rangle \\
&\geq -\langle \check{w}_S, |\check{\delta}_S|\rangle + \langle \check{w}_{S^c}, |\check{\delta}_{S^c}|\rangle \\
&\geq -\check{w}_S^{\max}\|\check{\delta}_S\|_1 + \check{w}_{S^c}^{\min}\|\check{\delta}_{S^c}\|_1.
\end{aligned} \tag{3.14}$$

Then by substituting (3.13) and (3.14) into (3.12), we obtain that

$$\begin{aligned}
\frac{1}{2m}\|A\check{\delta}\|_2^2 &\leq \frac{\rho_m \check{w}_{S^c}^{\min}}{\kappa}\big(\|\check{\delta}_S\|_1 + \|\check{\delta}_{S^c}\|_1\big) + \rho_m\big(\check{w}_S^{\max}\|\check{\delta}_S\|_1 - \check{w}_{S^c}^{\min}\|\check{\delta}_{S^c}\|_1\big) \\
&= \rho_m \check{w}_{S^c}^{\min}\left(\frac{\check{w}_S^{\max}}{\check{w}_{S^c}^{\min}} + \frac{1}{\kappa}\right)\|\check{\delta}_S\|_1 - \rho_m \check{w}_{S^c}^{\min}\left(1 - \frac{1}{\kappa}\right)\|\check{\delta}_{S^c}\|_1,
\end{aligned} \tag{3.15}$$

which, together with the assumption that $\check{w}_S^{\max}/\check{w}_{S^c}^{\min} \leq 1$, implies that $\check{\delta} \in \mathcal{C}(S;\alpha_\kappa)$. Since $\kappa > 1$, Assumption 3.2 and (3.15) yield that

$$\frac{1}{2}\lambda_{S,\kappa}\|\check{\delta}\|_2^2 \leq \rho_m \check{w}_{S^c}^{\min}\left(\frac{\check{w}_S^{\max}}{\check{w}_{S^c}^{\min}} + \frac{1}{\kappa}\right)\|\check{\delta}_S\|_1 \leq \sqrt{s}\rho_m \check{w}_{S^c}^{\min}\left(\frac{\check{w}_S^{\max}}{\check{w}_{S^c}^{\min}} + \frac{1}{\kappa}\right)\|\check{\delta}\|_2.$$

This completes the proof. □

Proposition 3.5. *Let $\kappa > 1$ and $c > 0$ be given constants. Consider the linear model (3.1) under Assumption 3.1, 3.2 and 3.3. Suppose that $\check{w}_{S^c}^{\min} \geq \check{w}_S^{\max}$ and $\check{w}_{S^c}^{\min} > 0$. If the penalization parameter is chosen as*

$$\rho_m \check{w}_{S^c}^{\min} = \kappa\nu\sqrt{2+2c}\sqrt{\frac{\log n}{m}},$$

then with probability at least $1 - 2n^{-c}$*, the weighted Lasso estimator (3.4) satisfies the bound*

$$\|\check{x} - \bar{x}\|_2 \leq \sqrt{1+c}\left(1 + \frac{\check{w}_S^{\max}}{\check{w}_{S^c}^{\min}}\kappa\right)\frac{2\sqrt{2}\nu}{\lambda_{S,\kappa}}\sqrt{\frac{s\log n}{m}}.$$

Proof. From Lemma 3.4, it suffices to show that $\rho_m \check{w}_{S^c}^{\min} = \kappa\nu\sqrt{2+2c}\sqrt{\log n/m} \geq \kappa\|\frac{1}{m}A^T\xi\|_\infty$ with probability at least $1 - 2n^{-c}$, which has been established in the proof of Proposition 3.3. □

We then turn to derive an estimation error bound for the corrected Lasso in a similar way to that for the Lasso. Define

$$a_S := \|\mathrm{sign}(\bar{x}_S) - \hat{w}_S\|_\infty \quad \text{and} \quad a_{S^c} := 1 - \|\hat{w}_{S^c}\|_\infty. \tag{3.16}$$

Lemma 3.6. *Consider the linear model (3.1) under Assumption 3.2 with a given constant* $\kappa > 1$*. Suppose that* $\hat{a}_{S^c} \geq \hat{a}_S$ *and* $\hat{\rho}_m\hat{a}_{S^c} \geq \kappa\|\frac{1}{m}A^T\xi\|_\infty$*. Then the corrected Lasso estimator (3.5) satisfies the bound*

$$\|\hat{x} - \bar{x}\|_2 \leq 2\left(\frac{\hat{a}_S}{\hat{a}_{S^c}} + \frac{1}{\kappa}\right)\frac{\sqrt{s}\rho_m\hat{a}_{S^c}}{\lambda_{S,\kappa}}.$$

Proof. Let $\hat{\delta} := \hat{x} - \bar{x}$. From the characterization of $\partial\|\bar{x}\|_1$, we derive that

$$\begin{aligned}
\|\bar{x} + \hat{\delta}\|_1 - \|\bar{x}\|_1 - \langle\hat{w}, \hat{\delta}\rangle &\geq \langle\mathrm{sign}(\bar{x}_S) - \hat{w}_S, \hat{\delta}_S\rangle + \|\hat{\delta}_{S^c}\|_1 - \langle\hat{w}_{S^c}, \hat{\delta}_{S^c}\rangle \\
&\geq -\|\mathrm{sign}(\bar{x}_S) - \hat{w}_S\|_\infty\|\hat{\delta}_S\|_1 + \big(1 - \|\hat{w}_{S^c}\|_\infty\big)\|\hat{\delta}_{S^c}\|_1 \\
&= -\hat{a}_S\|\hat{\delta}_S\|_1 + \hat{a}_{S^c}\|\hat{\delta}_{S^c}\|_1.
\end{aligned}$$

Then the proof can be obtained in a similar way to that of Lemma 3.4. □

Proposition 3.7. *Let* $\kappa > 1$ *and* $c > 0$ *be given constants. Consider the linear model (3.1) under Assumption 3.1, 3.2 and 3.3. Suppose that* $\hat{a}_{S^c} \geq \hat{a}_S$ *and* $\hat{a}_{S^c} > 0$*. If the penalization parameter is chosen as*

$$\hat{\rho}_m\hat{a}_{S^c} = \kappa\nu\sqrt{2+2c}\sqrt{\frac{\log n}{m}},$$

then with probability at least $1-2n^{-c}$*, the corrected Lasso estimator (3.5) satisfies the bound*

$$\|\hat{x}-\bar{x}\|_2 \le \sqrt{1+c}\left(1+\frac{\hat{a}_S}{\hat{a}_{S^c}}\kappa\right)\frac{2\sqrt{2}\nu}{\lambda_{S,\kappa}}\sqrt{\frac{s\log n}{m}}.$$

Proof. From Lemma 3.6, it suffices to show that $\hat{\rho}_m\hat{a}_{S^c} = \kappa\nu\sqrt{2+2c}\sqrt{\log n/m} \ge \kappa\|\frac{1}{m}A^T\xi\|_\infty$ with probability at least $1-2n^{-c}$, which has been established in the proof of Proposition 3.3. □

Lastly, we provide an estimation error bound for the oracle thresholded Lasso, which serves as an evidence to evaluate how well the weighted Lasso and the corrected Lasso can perform.

Proposition 3.8. *Under the assumptions in Proposition 3.3, with probability at least* $1-2n^{-c}$*, the oracle thresholded Lasso estimator (3.6) satisfies the bound*

$$\|\tilde{x}-\bar{x}\|_2 \le \frac{2\sqrt{2}\nu}{\lambda_{S,\kappa}}\sqrt{\frac{s(c\log n+\log s)}{m}}.$$

Proof. Let $\tilde{\delta} := \tilde{x}-\bar{x}$. Since $\mathcal{J}=S$, by following a similar way to (3.7), we have

$$\frac{1}{2m}\|A_S\tilde{\delta}_S\|_2^2 \le \left\langle\frac{1}{m}A_S^T\xi,\tilde{\delta}_S\right\rangle \le \left\|\frac{1}{m}A_S^T\xi\right\|_\infty\|\tilde{\delta}_S\|_1 \le \sqrt{s}\left\|\frac{1}{m}A_S^T\xi\right\|_\infty\|\tilde{\delta}_S\|_2. \quad (3.17)$$

Note that $\tilde{\delta}_{S^c}=0$. Hence, $\tilde{\delta}\in\mathcal{C}(S;\alpha_\kappa)$. From Assumption 3.2 and (3.17), it follows that

$$\|\tilde{x}-\bar{x}\|_2 = \|\tilde{\delta}_S\|_2 \le \frac{2\sqrt{s}}{\lambda_{S,\kappa}}\left\|\frac{1}{m}A_S^T\xi\right\|_\infty.$$

Then it suffices to show that $\|\frac{1}{m}A_S^T\xi\|_\infty \le \sqrt{2}\nu\sqrt{(c\log n+\log s)/m}$ with probability at least $1-2n^{-c}$, which can be established in a similar way to the proof of Proposition 3.3. □

3.3 Gaussian design

In this section, we consider estimation error bounds for the aforementioned Lasso-related estimators in the case of correlated Gaussian design assumed as follows.

Assumption 3.4. *The design matrix $A \in \mathbb{R}^{m\times n}$ is formed by independently sampling each row from the n-dimensional multivariate Gaussian distribution $N(0, \Sigma)$.*

The key factor for establishing estimation error bound is the RE condition of design matrices. Built on the Gordon's Minimax Lemma from the theory of Gaussian random processes [66], a slightly stronger version of the RE condition was shown to hold for such a class of correlated Gaussian design matrices [103, Theorem 1]. Let $\Sigma^{\frac{1}{2}}$ denote the square root of Σ and define $\Sigma_{\max} := \max_{j=1,\ldots,n} \Sigma_{jj}$.

Lemma 3.9. *Under Assumption 3.4, there exist absolute constants $c_1 > 0$ and $c_2 > 0$ such that with probability at least $1 - c_1 \exp(-c_2 m)$, it holds that*

$$\frac{\|A\delta\|_2}{\sqrt{m}} \geq \frac{1}{4}\|\Sigma^{\frac{1}{2}}\delta\|_2 - 9\sqrt{\Sigma_{\max}}\sqrt{\frac{\log n}{m}}\|\delta\|_1, \quad \forall\ \delta \in \mathbb{R}^n.$$

If a similar kind of RE condition is further assumed to hold for the covariance matrix Σ, then it follows immediately from Lemma 3.9 that the correlated Gaussian design matrix A satisfies the RE condition in the sense of Assumption 3.2.

Assumption 3.5. *There exists a constant $\lambda'_{S,\kappa} > 0$ such that*

$$\|\Sigma^{\frac{1}{2}}\delta\|_2^2 \geq \lambda'_{S,\kappa}\|\delta\|_2^2, \quad \forall\ \delta \in \mathcal{C}(S; \alpha_\kappa),$$

where $\alpha_\kappa := (\kappa + 1)/(\kappa - 1)$ for some given constant $\kappa > 0$. In this case, we say that the covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$ satisfies the RE condition over the true supporting index set S with parameters $(\alpha_\kappa, \lambda'_{S,\kappa})$.

Corollary 3.10. *Suppose that Assumption 3.4 and 3.5 hold. Then there exist absolute constants $c_1 > 0$ and $c_2 > 0$ such that if the sample size meets that*

$$m \geq \frac{[72(1 + \alpha_\kappa)]^2\Sigma_{\max}}{\lambda'_{S,\kappa}} s \log n,$$

with probability at least $1 - c_1 \exp(-c_2 m)$, the design matrix A satisfies the RE condition over S with parameters $(\alpha_\kappa, \lambda_{S,\kappa})$ where $\lambda_{S,\kappa} := \lambda'_{S,\kappa}/64$.

Proof. For any $\delta \in \mathcal{C}(S;\alpha_\kappa)$, we have that

$$\|\delta\|_1 = \|\delta_S\|_1 + \|\delta_{S^c}\|_1 \le (1+\alpha_\kappa)\|\delta_S\|_1 \le (1+\alpha_\kappa)\sqrt{s}\|\delta\|_2,$$

Then due to Assumption 3.5 and Lemma 3.9, there exist absolute constants $c_1 > 0$ and $c_2 > 0$ such that with probability at least $1 - c_1\exp(-c_2 m)$, we have

$$\frac{\|A\delta\|_2}{\sqrt{m}} \ge \left[\frac{1}{4}\sqrt{\lambda'_{S,\kappa}} - 9\sqrt{\Sigma_{\max}}(1+\alpha_\kappa)\sqrt{\frac{s\log n}{m}}\right]\|\delta\|_2 \ge \frac{1}{8}\sqrt{\lambda'_{S,\kappa}}\|\delta\|_2,$$

for all $\delta \in \mathcal{C}(S;\alpha_\kappa)$, provided that $m \ge [72(1+\alpha_\kappa)]^2(\Sigma_{\max}/\lambda'_{S,\kappa})s\log n$. □

Another ingredient in establishing estimation error bound is the column-normalized condition in the sense of Assumption 3.3, which can be verified by exploiting an exponential tail inequality for the χ^2 random variable modified from [85, Lemma 1 and Comments] with a suitable change of variables.

Lemma 3.11. *Let z_m be a centralized χ^2 random variable with m degrees of freedom. Then for any $t > 0$,*

$$\mathbb{P}\left[z_m - m \ge \frac{mt}{2} + \frac{mt^2}{8}\right] \le \exp\left(-\frac{mt^2}{16}\right),$$

which implies that for any $0 < t \le 4$,

$$\mathbb{P}\left[z_m \ge m(1+t)\right] \le \exp\left(-\frac{mt^2}{16}\right).$$

Proof. The first part is directly from [85, Lemma 1 and the Comments] with a suitable change of variables. The second part follows by noting that $m(1+t) \ge m(1+t/2+t^2/8)$ when $0 < t \le 4$. □

Corollary 3.12. *Suppose that Assumption 3.4 holds. If the sample size $m \ge (1+c)\log n$ for a given constant $c > 0$, then $\max_{j=1,\ldots,n}\|A_j\|_2 \le \sqrt{5\Sigma_{\max} m}$ with probability at least $1 - n^{-c}$.*

Proof. Under Assumption 3.4, $A_j \in \mathbb{R}^m$ is a random vector of i.i.d. entries from the univariate Gaussian distribution $N(0, \Sigma_{jj})$, for all $j = 1, \ldots, n$. Thus, $\|A_j\|_2^2/\Sigma_{jj}$ follows the χ^2-distribution with m degrees of freedom. By using Lemma 3.11, we obtain that for any $0 < t \leq 4$,

$$\mathbb{P}\left[\frac{\|A_j\|_2^2}{\Sigma_{\max}} \geq m(1+t)\right] \leq \mathbb{P}\left[\frac{\|A_j\|_2^2}{\Sigma_{jj}} \geq m(1+t)\right] \leq \exp\left(-\frac{mt^2}{16}\right).$$

Consequently, if $m \geq (1+c)\log n$, choosing $t^* = 4\sqrt{1+c}\sqrt{\log n/m}$ and taking the union bound give that

$$\begin{aligned}\mathbb{P}\left[\max_{j=1,\ldots,n} \|A_j\|_2^2 \geq 5\Sigma_{\max} m\right] &\leq \mathbb{P}\left[\max_{j=1,\ldots,n} \|A_j\|_2^2 \geq \Sigma_{\max} m(1+t^*)\right] \\ &\leq n\exp\left(-\frac{mt^{*2}}{16}\right) = \frac{1}{n^c},\end{aligned}$$

which completes the proof. □

In the remaining part of this section, we state the specialized estimation error bounds for the Lasso, the weighted Lasso, the corrected Lasso and the oracle thresholded Lasso with respect to Gaussian design under Assumption 3.4 and 3.5.

Proposition 3.13. *Let $\kappa > 1$ and $c > 0$ be given constants. Consider the linear model (3.1) under Assumption 3.1, 3.4 and 3.5. Let $\lambda_{S,\kappa} := \lambda'_{S,\kappa}/64$. If the penalization parameter is chosen as*

$$\rho_m = \kappa\nu\sqrt{1+c}\sqrt{10\Sigma_{\max}}\sqrt{\frac{\log n}{m}}$$

and the sample size meets that

$$m \geq \max\left\{\frac{[72(1+\alpha_\kappa)]^2\Sigma_{\max}}{\lambda'^2_{S,\kappa}} s\log n,\ (1+c)\log n\right\}, \tag{3.18}$$

then there exist absolute constants $c_1 > 0$ and $c_2 > 0$ such that with probability at least $1 - c_1\exp(-c_2 m) - 3n^{-c}$, the Lasso estimator (3.2) satisfies the bound

$$\|\mathring{x} - \bar{x}\|_2 \leq \sqrt{1+c}(1+\kappa)\frac{2\sqrt{10\Sigma_{\max}}\nu}{\lambda_{S,\kappa}}\sqrt{\frac{s\log n}{m}}.$$

Proof. The proof can be obtained by using Proposition 3.3, Corollary 3.10 and Corollary 3.12. □

Proposition 3.14. *Let $\kappa > 1$ and $c > 0$ be given constants. Consider the linear model (3.1) under Assumption 3.1, 3.4 and 3.5. Suppose that $\check{w}_{S^c}^{\min} \geq \check{w}_S^{\max}$ and $\check{w}_{S^c}^{\min} > 0$, where $\check{w}_S^{\max}$ and $\check{w}_{S^c}^{\min}$ are defined by (3.11). Let $\lambda_{S,\kappa} := \lambda'_{S,\kappa}/64$. If the penalization parameter is chosen as*

$$\rho_m \check{w}_{S^c}^{\min} = \kappa\nu\sqrt{1+c}\sqrt{10\Sigma_{\max}}\sqrt{\frac{\log n}{m}}$$

and the sample size meets (3.18), then there exist absolute constants $c_1 > 0$ and $c_2 > 0$ such that with probability at least $1 - c_1\exp(-c_2 m) - 3n^{-c}$, the weighted Lasso estimator (3.4) satisfies the bound

$$\|\check{x} - \bar{x}\|_2 \leq \sqrt{1+c}\left(1 + \frac{\check{w}_S^{\max}}{\check{w}_{S^c}^{\min}}\kappa\right)\frac{2\sqrt{10\Sigma_{\max}}\nu}{\lambda_{S,\kappa}}\sqrt{\frac{s\log n}{m}}.$$

Proof. The proof can be obtained by using Proposition 3.5, Corollary 3.10 and Corollary 3.12. □

Proposition 3.15. *Let $\kappa > 1$ and $c > 0$ be given constants. Consider the linear model (3.1) under Assumption 3.1, 3.4 and 3.5. Suppose that $\hat{a}_{S^c} \geq \hat{a}_S$ and $\hat{a}_{S^c} > 0$, where $\hat{a}_S$ and $\hat{a}_{S^c}$ are defined by (3.16). Let $\lambda_{S,\kappa} := \lambda'_{S,\kappa}/64$. If the penalization parameter is chosen as*

$$\hat{\rho}_m \hat{a}_{S^c} = \kappa\nu\sqrt{1+c}\sqrt{10\Sigma_{\max}}\sqrt{\frac{\log n}{m}}$$

and the sample size meets (3.18), then there exist absolute constants $c_1 > 0$ and $c_2 > 0$ such that with probability at least $1 - c_1\exp(-c_2 m) - 3n^{-c}$, the corrected Lasso estimator (3.5) satisfies the bound

$$\|\hat{x} - \bar{x}\|_2 \leq \sqrt{1+c}\left(1 + \frac{\hat{a}_S}{\hat{a}_{S^c}}\kappa\right)\frac{2\sqrt{10\Sigma_{\max}}\nu}{\lambda_{S,\kappa}}\sqrt{\frac{s\log n}{m}}.$$

Proof. The proof can be obtained by using Proposition 3.7, Corollary 3.10 and Corollary 3.12. □

Proposition 3.16. *Under the assumptions in Proposition 3.13, if the sample size meets (3.18), then there exist absolute constants $c_1 > 0$ and $c_2 > 0$ such that with probability at least $1 - c_1 \exp(-c_2 m) - 3n^{-c}$,the oracle thresholded Lasso estimator (3.6) satisfies the bound*

$$\|\tilde{x} - \bar{x}\|_2 \leq \frac{2\sqrt{10\Sigma_{\max}}\nu}{\lambda_{S,\kappa}} \sqrt{\frac{s(c \log n + \log s)}{m}}.$$

Proof. This follows from Proposition 3.8, Corollary 3.10 and Corollary 3.12. □

3.4 Sub-Gaussian design

In this section, we consider estimation error bounds for the aforementioned Lasso-related estimators in the case of correlated sub-Gaussian design. Before providing the concrete assumption on the design matrix, we need some additional definitions. The first one is taken from [18, Lemma 1.2 and Definition 2.1 in Chapter 1], while the rest two are borrowed from [110, Definition 5].

Definition 3.2. *A sub-Gaussian random variable z with exponent $\varsigma(z) \geq 0$ is called strictly sub-Gaussian if $\mathbb{E}[z^2] = [\varsigma(z)]^2$.*

Definition 3.3. *A random vector $Z \in \mathbb{R}^n$ is called isotropic if for all $w \in \mathbb{R}^n$,*

$$\mathbb{E}[\langle Z, w\rangle^2] = \|w\|_2^2.$$

Definition 3.4. *A random vector $Z \in \mathbb{R}^n$ is called ψ_2-bounded with an absolute constant $\beta > 0$ if for all $w \in \mathbb{R}^n$,*

$$\|\langle Z, w\rangle\|_{\psi_2} \leq \beta \|w\|_2,$$

where $\|\cdot\|_{\psi_2}$ is the Orlicz ψ_2-norm defined by (2.1). In this case, the absolute constant β is called a ψ_2-constant of the random vector Z.

The next lemma shows that any strictly sub-Gaussian random matrix consists of independent, isotropic and ψ_2-bounded rows.

Lemma 3.17. *Suppose that $\Gamma \in \mathbb{R}^{m\times n}$ is a random matrix of i.i.d. strictly sub-Gaussian entries with exponent 1. Then the rows of Γ are independent, isotropic and ψ_2-bounded random vectors in $\mathbb{R}^n$ with a common ψ_2-constant $\beta > 0$.*

Proof. Let $Z \in \mathbb{R}^n$ be any row of Γ. Then according to [18, Lemma 2.1 in Chapter 1], for any fixed $w \in \mathbb{R}^n$, $\langle Z, w\rangle$ is a strictly sub-Gaussian random variable with exponent $\|w\|_2$. This implies that $\mathbb{E}[\langle Z, w\rangle^2] = \|w\|_2^2$. Moreover, it follows from [120, Lemma 5.5] that $\|\langle Z, w\rangle\|_{\psi_2} \leq \beta\|w\|_2$ for some absolute constant $\beta > 0$. □

The setting of correlated sub-Gaussian design is stated below, where the RE condition in the sense of Assumption 3.5 is needed for the "covariance matrix" Σ.

Assumption 3.6. *The design matrix $A \in \mathbb{R}^{m\times n}$ can be expressed as $A = \Gamma\Sigma^{\frac{1}{2}}$, where $\Gamma \in \mathbb{R}^{m\times n}$ is a random matrix of i.i.d. strictly sub-Gaussian entries with exponent 1, and $\Sigma \in \mathbb{R}^{n\times n}$ is a positive semidefinite matrix satisfying the RE condition over the true supporting index set S with parameters $(\alpha_\kappa, \lambda'_{S,\kappa})$ and $(3\alpha_\kappa, \lambda''_{S,\kappa})$ for some given constant $\kappa > 1$ and $\alpha_\kappa := (\kappa+1)/(\kappa-1)$.*

Remark 3.1. *Under Assumption 3.6, it follows from Lemma 3.17 that the rows of Γ are independent, isotropic and ψ_2-bounded random vectors in $\mathbb{R}^n$ with a common ψ_2-constant $\beta > 0$.*

Based on geometric functional analysis, the correlated sub-Gaussian design matrix A was shown to inherit the RE condition in the sense of Assumption 3.2 from that for the "covariance matrix" Σ [110, Theorem 6].

Lemma 3.18. *Suppose that Assumption 3.6 holds. For a given $0 < \vartheta < 1$, set*

$$d := s + s\Sigma_{\max}\frac{(12\alpha_\kappa)^2(3\alpha_\kappa+1)}{\vartheta^2\lambda''_{S,\kappa}},$$

where $\Sigma_{\max} := \max_{j=1,\ldots,n} \Sigma_{jj}$. *Let* $\beta > 0$ *be the* ψ_2*-constant in Remark 3.1. If the sample size meets that*

$$m \geq \frac{2000 \min\{d, n\}\beta^4}{\vartheta^2} \log\left(\frac{180n}{\min\{d, n\}\vartheta}\right),$$

then with probability at least $1 - 2\exp(-\vartheta^2 m/2000\beta^4)$*, the design matrix* A *satisfies the RE condition over* S *with parameters* $(\alpha_\kappa, \lambda_{S,\kappa})$*, where* $\lambda_{S,\kappa} \geq (1-\vartheta)^2 \lambda'_{S,\kappa}$.

Proof. The proof follows from Remark 3.1 and [110, Theorem 6]. □

The column-normalized condition in the sense of Assumption 3.3 can be verified by utilizing an exponential tail bound for positive semidefinite quadratic forms of sub-Gaussian random vectors developed in [74, Theorem 2.1 and Remark 2.2].

Lemma 3.19. *Let* $Z \in \mathbb{R}^m$ *be a random vector of i.i.d. sub-Gaussian entries with exponent* 1*. Then for any* $t > 0$*,*

$$\mathbb{P}\left[\|Z\|_2^2 \geq m + \frac{mt}{2} + \frac{mt^2}{8}\right] \leq \exp\left(-\frac{mt^2}{16}\right),$$

which implies that for any $0 < t \leq 4$*,*

$$\mathbb{P}\left[\|Z\|_2^2 \geq m(1+t)\right] \leq \exp\left(-\frac{mt^2}{16}\right).$$

Proof. The first part is directly from [74, Theorem 2.1 and Remark 2.2] with a suitable change of variables. The second part follows by noting that $m(1+t) \geq m(1 + t/2 + t^2/8)$ when $0 < t \leq 4$. □

Corollary 3.20. *Suppose that the design matrix* $A = \Gamma\Sigma^{\frac{1}{2}}$*, where* $\Gamma \in \mathbb{R}^{m\times n}$ *is a random matrix of i.i.d. sub-Gaussian entries with exponent* 1 *and* $\Sigma \in \mathbb{R}^{n\times n}$ *is a positive semidefinite matrix with* $\Sigma_{\max} := \max_{j=1,\ldots,n} \Sigma_{jj}$*. If the sample size* $m \geq (1+c)\log n$ *for a given constant* $c > 0$*, then* $\max_{j=1,\ldots,n} \|A_j\|_2 \leq \sqrt{5\Sigma_{\max} m}$ *with probability at least* $1 - n^{-c}$.

Proof. From the structure of the design matrix A, we know that $A_j = \Gamma(\Sigma^{\frac{1}{2}})_j \in \mathbb{R}^m$ is a random vector of i.i.d. sub-Gaussian entries with exponent $\|(\Sigma^{\frac{1}{2}})_j\|_2 = \sqrt{\Sigma_{jj}}$, for all $j = 1, \ldots, n$ (cf. [18, Chapter 1] and [62, Section 12.7]). Then the proof can be obtained in a similar way to that of Corollary 3.12 by using Lemma 3.19. □

In the rest of this section, we present the specialized estimation error bounds for the Lasso, the weighted Lasso, the corrected Lasso and the oracle thresholded Lasso with respect to sub-Gaussian design under Assumption 3.6. Define $\Sigma_{\max} := \max_{j=1,\ldots,n} \Sigma_{jj}$.

Proposition 3.21. *Let $\kappa > 1$ and $c > 0$ be given constants. Consider the linear model (3.1) under Assumption 3.1 and 3.6. For any given $0 < \vartheta < 1$, let d, β and $\lambda_{S,\kappa}$ be defined in Lemma 3.18. If the penalization parameter is chosen as*

$$\rho_m = \kappa\nu\sqrt{1+c}\sqrt{10\Sigma_{\max}}\sqrt{\frac{\log n}{m}}$$

and the sample size meets that

$$m \geq \max\left\{\frac{2000\min\{d,n\}\beta^4}{\vartheta^2}\log\left(\frac{180n}{\min\{d,n\}\vartheta}\right), (1+c)\log n\right\}, \tag{3.19}$$

then the Lasso estimator (3.2) satisfies the bound

$$\|\mathring{x} - \bar{x}\|_2 \leq \sqrt{1+c}(1+\kappa)\frac{2\sqrt{10\Sigma_{\max}}\nu}{\lambda_{S,\kappa}}\sqrt{\frac{s\log n}{m}}$$

with probability at least $1 - 2\exp(-\vartheta^2 m/2000\beta^4) - 3n^{-c}$.

Proof. The proof can be obtained by applying Proposition 3.3, Lemma 3.18 and Corollary 3.20. □

Proposition 3.22. *Let $\kappa > 1$ and $c > 0$ be given constants. Consider the linear model (3.1) under Assumption 3.1 and 3.6. Suppose that $\check{w}_{S^c}^{\min} \geq \check{w}_S^{\max}$ and $\check{w}_{S^c}^{\min} >$*

0, where $\check{w}_S^{\max}$ *and* $\check{w}_{S^c}^{\min}$ *are defined by (3.11). For any given* $0 < \vartheta < 1$*, let* d*,* β *and* $\lambda_{S,\kappa}$ *be defined in Lemma 3.18. If the penalization parameter is chosen as*

$$\rho_m \check{w}_{S^c}^{\min} = \kappa\nu\sqrt{1+c}\sqrt{10\Sigma_{\max}}\sqrt{\frac{\log n}{m}}$$

and the sample size meets (3.19), then the weighted Lasso estimator (3.4) satisfies the bound

$$\|\check{x} - \bar{x}\|_2 \leq \sqrt{1+c}\left(1 + \frac{\check{w}_S^{\max}}{\check{w}_{S^c}^{\min}}\kappa\right)\frac{2\sqrt{10\Sigma_{\max}}\nu}{\lambda_{S,\kappa}}\sqrt{\frac{s\log n}{m}}$$

with probability at least $1 - 2\exp(-\vartheta^2 m/2000\beta^4) - 3n^{-c}$*.*

Proof. The proof can be obtained by applying Proposition 3.5, Lemma 3.18 and Corollary 3.20. □

Proposition 3.23. *Let* $\kappa > 1$ *and* $c > 0$ *be given constants. Consider the linear model (3.1) under Assumption 3.1 and 3.6. Suppose that* $\hat{a}_{S^c} \geq \hat{a}_S$ *and* $\hat{a}_{S^c} > 0$*, where* $\hat{a}_S$ *and* $\hat{a}_{S^c}$ *are defined by (3.16). For any given* $0 < \vartheta < 1$*, let* d*,* β *and* $\lambda_{S,\kappa}$ *be defined in Lemma 3.18. If the penalization parameter is chosen as*

$$\hat{\rho}_m \hat{a}_{S^c} = \kappa\nu\sqrt{1+c}\sqrt{10\Sigma_{\max}}\sqrt{\frac{\log n}{m}}$$

and the sample size meets (3.19), then the corrected Lasso estimator (3.5) satisfies the bound

$$\|\hat{x} - \bar{x}\|_2 \leq \sqrt{1+c}\left(1 + \frac{\hat{a}_S}{\hat{a}_{S^c}}\kappa\right)\frac{2\sqrt{10\Sigma_{\max}}\nu}{\lambda_{S,\kappa}}\sqrt{\frac{s\log n}{m}}$$

with probability at least $1 - 2\exp(-\vartheta^2 m/2000\beta^4) - 3n^{-c}$*.*

Proof. The proof can be obtained by applying Proposition 3.7, Lemma 3.18 and Corollary 3.20. □

Proposition 3.24. *Under the assumptions in Proposition 3.21, if the sample size meets (3.19), then the oracle thresholded Lasso estimator (3.6) satisfies the bound*

$$\|\tilde{x} - \bar{x}\|_2 \leq \frac{2\sqrt{10\Sigma_{\max}}\nu}{\lambda_{S,\kappa}}\sqrt{\frac{s(c\log n + \log s)}{m}}$$

with probability at least $1 - 2\exp(-\vartheta^2 m/2000\beta^4) - 3n^{-c}$.

Proof. This follows from Proposition 3.8, Lemma 3.18 and Corollary 3.20. □

3.5 Comparison among the error bounds

In this section, we make a quantitative comparison among the estimation error bounds for the aforementioned Lasso-related estimators. For simplicity, we only focus on the case when the design matrix is deterministic.

Evidently, when the weight vector $\check{w}$ for the weighted Lasso (3.4) is chosen such that $\check{w}_S^{\max} \ll \check{w}_{S^c}^{\min}$, the corresponding estimation error bound provided in Proposition 3.5 will at best become

$$\|\check{x} - \bar{x}\|_2 \leq \sqrt{1+c}\left(1 + \frac{\check{w}_S^{\max}}{\check{w}_{S^c}^{\min}}\kappa\right)\frac{2\sqrt{2}\nu}{\lambda_{S,\kappa}}\sqrt{\frac{s\log n}{m}} \to \sqrt{1+c}\frac{2\sqrt{2}\nu}{\lambda_{S,\kappa}}\sqrt{\frac{s\log n}{m}}.$$

In comparison with the estimation error bounds for the Lasso (3.2) and the oracle thresholded Lasso (3.6) provided in Proposition 3.3 and Proposition 3.8, respectively, this best possible error bound for the weighted Lasso enjoys a significant reduction from the error bound for the Lasso in light of

$$\frac{\text{the best possible error bound for the weighted Lasso}}{\text{the error bound for the Lasso}} = \frac{1}{1+\kappa} \quad \text{with} \quad \kappa > 1,$$

while at the same time it is very close to the optimal estimation error bound achieved by the oracle thresholded Lasso since

$$\frac{\text{the best possible error bound for the weighted Lasso}}{\text{the error bound for the oracle thresholded Lasso}} = \sqrt{\frac{\log n + c\log n}{\log s + c\log n}} \approx 1,$$

as long as the probability-controlling parameter c is not too small. The same conclusions also hold for the estimation error bound of the corrected Lasso (3.5) provided in Proposition 3.7.

Finally, we take the two-stage adaptive Lasso procedure [136] as an illustration. Recall that the adaptive Lasso is equipped with the weight vector $\check{w} = 1/|\mathring{x}|^{\gamma}$, where $\gamma > 0$ is a given parameter. In detail, $\check{w}_j = 1/|\mathring{x}_j|^{\gamma}$ for $j = 1, \ldots, n$, while we set $\check{w}_j = +\infty$ and thus $\check{x}_j = 0$ if $\mathring{x}_j = 0$. With the most common choice of the parameter $\gamma = 1$, we have

$$\check{w}_S^{\max} = \frac{1}{\min_{j \in S} |\mathring{x}|_j}, \quad \check{w}_{S^c}^{\min} = \frac{1}{\max_{j \in S^c} |\mathring{x}|_j}, \quad \text{and} \quad \frac{\check{w}_S^{\max}}{\check{w}_{S^c}^{\min}} = \frac{\max_{j \in S^c} |\mathring{x}|_j}{\min_{j \in S} |\mathring{x}|_j}.$$

Roughly speaking, when the Lasso estimator performs good enough in the sense that $\max_{j \in S^c} |\mathring{x}|_j \ll \min_{j \in S} |\mathring{x}|_j$, then the two-stage adaptive Lasso estimator is able to imitate the ideal behavior of the oracle thresholded Lasso estimator as if the true supporting index set S were known in advance.

Chapter 4

Exact matrix decomposition from fixed and sampled basis coefficients

In this chapter, we study the problem of exact low-rank and sparse matrix decomposition with fixed and sampled basis coefficients. We begin with the model setting and assumption, and then formulate this problem into concrete convex programs based on the "nuclear norm plus ℓ_1-norm" approach. Owing to the convex nature of the proposed optimization problems, we provide exact recovery guarantees if certain standard identifiability conditions for the low-rank and sparse components are satisfied. Lastly, we establish the probabilistic recovery results via a standard dual certification procedure. Although the analysis involved follows from the existing framework, these recovery guarantees can still be considered as the noiseless counterparts of those for the noisy case addressed in Chapter 5.

4.1 Problem background and formulation

In this section, we introduce the background on the problem of exact low-rank and sparse matrix decomposition with fixed and sampled basis coefficients, and then

propose convex optimization formulations that we study in this chapter.

Let the set of the standard orthonormal basis of the finite dimensional real Euclidean space $\mathbb{V}^{n_1\times n_2}$ be denoted by $\Theta := \{\Theta_1, \ldots, \Theta_d\}$, where d is the dimension of $\mathbb{V}^{n_1\times n_2}$. Specifically, when $\mathbb{V}^{n_1\times n_2} = \mathbb{R}^{n_1\times n_2}$, we have $d = n_1 n_2$ and

$$\Theta = \left\{ e_i e_j^T \,\middle|\, 1 \le i \le n_1,\ 1 \le j \le n_2 \right\}; \tag{4.1}$$

when $\mathbb{V}^{n_1\times n_2} = \mathcal{S}^n$, we have $d = n(n+1)/2$ and

$$\Theta = \left\{ e_i e_i^T \,\middle|\, 1 \le i \le n \right\} \bigcup \left\{ \frac{1}{\sqrt{2}}\left(e_i e_j^T + e_j e_i^T\right) \,\middle|\, 1 \le i < j \le n \right\}. \tag{4.2}$$

Then any matrix $Z \in \mathbb{V}^{n_1\times n_2}$ can be uniquely represented as

$$Z = \sum_{j=1}^{d} \langle \Theta_j, Z \rangle \, \Theta_j, \tag{4.3}$$

where $\langle \Theta_j, Z \rangle$ is called the basis coefficient of Z with respect to Θ_j.

Suppose that an unknown matrix $\overline{X} \in \mathbb{V}^{n_1\times n_2}$ can be decomposed into the sum of an unknown low-rank matrix $\overline{L} \in \mathbb{V}^{n_1\times n_2}$ and an unknown sparse matrix $\overline{S} \in \mathbb{V}^{n_1\times n_2}$, that is,

$$\overline{X} = \overline{L} + \overline{S},$$

where both of the components may be of arbitrary magnitude, and by "sparse" we mean that a few basis coefficients of the matrix $\overline{S}$ are nonzero. In this chapter, we assume that a number of basis coefficients of the unknown matrix $\overline{X}$ are fixed and the nonzero basis coefficients of the sparse component $\overline{S}$ only come from these fixed basis coefficients. Due to a certain structure or some reliable prior information, these assumptions are actually of practical interest. For example, the unknown matrix $\overline{X}$ is a correlation matrix with strict factor structure where the sparse component $\overline{S}$ is a diagonal matrix (see, e.g., [3, 123, 16]). Under these assumptions, we focus on the problem on how and when we are able to exactly recover the low-rank component $\overline{L}$ and the sparse component $\overline{S}$ by uniformly sampling a few basis coefficients with replacement from the unfixed ones of the unknown matrix $\overline{X}$.

Throughout this chapter, for the unknown matrix $\overline{X}$, we define $\mathcal{F} \subseteq \{1, \ldots, d\}$ to be the fixed index set corresponding to the fixed basis coefficients and $\mathcal{F}^c := \{1, \ldots, d\} \setminus \mathcal{F}$ to be the unfixed index set associated with the unfixed basis coefficients, respectively. Denote by d_s the cardinality of $\mathcal{F}^c$, that is, $d_s := |\mathcal{F}^c|$. Let Γ be the supporting index set of the sparse component $\overline{S}$, i.e., $\Gamma := \{j \,|\, \langle \Theta_j, \overline{S} \rangle \neq 0,\, j = 1, \ldots, d\}$. Denote $\Gamma_0 := \{j \,|\, \langle \Theta_j, \overline{S} \rangle = 0,\, j \in \mathcal{F}\} = \mathcal{F} \setminus \Gamma$. Below we summarize the model assumption adopted in this chapter.

Assumption 4.1. *The supporting index set Γ of the sparse component $\overline{S}$ is contained in the fixed index set $\mathcal{F}$, i.e., $\Gamma \subseteq \mathcal{F}$. The sampled indices are drawn uniformly at random with replacement from the unfixed index set $\mathcal{F}^c$.*

Under Assumption 4.1, it holds that $\mathcal{F} = \Gamma \cup \Gamma_0$ with $\Gamma \cap \Gamma_0 = \emptyset$ and $\Gamma^c := \{1, \ldots, d\} \setminus \Gamma = \Gamma_0 \cup \mathcal{F}^c$ with $\Gamma_0 \cap \mathcal{F}^c = \emptyset$. In addition, it is worthwhile to note that $\overline{S}$ cannot be exactly recovered if the basis coefficients of $\overline{X}$ with respect to Γ are not entirely known after the sampling procedure. This is the reason why we require $\Gamma \subseteq \mathcal{F}$ in Assumption 4.1.

4.1.1 Uniform sampling with replacement

When the fixed basis coefficients are not sufficient, we need to observe some of the rest for recovering the low-rank component $\overline{L}$ and the sparse component $\overline{S}$.

Let $\Omega := \{\omega_l\}_{l=1}^m$ be the multiset of indices sampled uniformly with replacement[1] from the unfixed index set $\mathcal{F}^c$ of the unknown matrix $\overline{X}$. Then the elements in Ω are i.i.d. copies of a random variable ω following the uniform distribution over $\mathcal{F}^c$, i.e., $\mathbb{P}[\omega = j] = 1/d_s$ for all $j \in \mathcal{F}^c$, where $d_s = |\mathcal{F}^c|$. Define the sampling operator $\mathcal{R}_\Omega : \mathbb{V}^{n_1 \times n_2} \to \mathbb{V}^{n_1 \times n_2}$ associated with the multiset Ω by

$$\mathcal{R}_\Omega(Z) := \sum_{l=1}^{m} \langle \Theta_{\omega_l}, Z \rangle \, \Theta_{\omega_l}, \quad Z \in \mathbb{V}^{n_1 \times n_2}. \tag{4.4}$$

[1]More details on random sampling model can be found in Section 2.3.

Note that the j-th basic coefficient $\langle \Theta_j, \mathcal{R}_\Omega(Z)\rangle$ of $\mathcal{R}_\Omega(Z)$ is zero unless $j \in \Omega$. For any $j \in \Omega$, $\langle \Theta_j, \mathcal{R}_\Omega(Z)\rangle$ is equal to $\langle \Theta_j, Z\rangle$ times the multiplicity of j in Ω. Although $\mathcal{R}_\Omega$ is still self-adjoint, it is in general not an orthogonal projection operator because repeated indices in Ω are likely to exist.

In addition, without causing any ambiguity, for any index subset $\mathcal{J} \subseteq \{1, \ldots, d\}$, we also use $\mathcal{J}$ to denote the subspace of the matrices in $\mathbb{V}^{n_1 \times n_2}$ whose supporting index sets are contained in the index subset $\mathcal{J}$. Let $\mathcal{P}_\mathcal{J} : \mathbb{V}^{n_1 \times n_2} \to \mathbb{V}^{n_1 \times n_2}$ be the orthogonal projection operator over $\mathcal{J}$, i.e.,

$$\mathcal{P}_\mathcal{J}(Z) := \sum_{j \in \mathcal{J}} \langle \Theta_j, Z\rangle \, \Theta_j, \quad Z \in \mathbb{V}^{n_1 \times n_2}. \tag{4.5}$$

Notice that $\mathcal{P}_\mathcal{J}$ is self-adjoint and $\|\mathcal{P}_\mathcal{J}\| = 1$.

With these notations, it follows from Assumption 4.1 that $\mathcal{P}_{\mathcal{F}^c}(\overline{S}) = 0$, $\mathcal{R}_\Omega(\overline{S}) = 0$ and $\mathcal{R}_\Omega(\overline{L}) = \mathcal{R}_\Omega(\overline{X})$. Then we can formulate the recovery problem considered in this chapter via convex optimization.

4.1.2 Convex optimization formulation

Suppose that Assumption 4.1 holds. Given the fixed data $\mathcal{P}_\mathcal{F}(\overline{X})$ and the sampled data $\mathcal{R}_\Omega(\overline{X})$, we wish to exactly recover the low-rank component $\overline{L}$ and the sparse component $\overline{S}$ by solving the following convex optimization problem

$$\begin{array}{ll} \min\limits_{L,\, S \in \mathbb{V}^{n_1 \times n_2}} & \|L\|_* + \rho \|S\|_1 \\ \text{s.t.} & \mathcal{P}_\mathcal{F}(L + S) = \mathcal{P}_\mathcal{F}(\overline{X}), \\ & \mathcal{P}_{\mathcal{F}^c}(S) = 0, \\ & \mathcal{R}_\Omega(L) = \mathcal{R}_\Omega(\overline{X}). \end{array} \tag{4.6}$$

Here $\rho \geq 0$ is a parameter that controls the tradeoff between the low-rank and sparse components. If, in addition, the true low-rank component $\overline{L}$ and the true sparse component $\overline{S}$ are known to be symmetric and positive semidefinite (e.g., $\overline{X}$

is a covariance or correlation matrix resulting from a factor model), we consider to solve the following convex conic optimization problem

$$\begin{aligned} \min \quad & \langle I_n, L\rangle + \rho\|S\|_1 \\ \text{s.t.} \quad & \mathcal{P}_{\mathcal{F}}(L+S) = \mathcal{P}_{\mathcal{F}}(\overline{X}), \\ & \mathcal{P}_{\mathcal{F}^c}(S) = 0, \\ & \mathcal{R}_{\Omega}(L) = \mathcal{R}_{\Omega}(\overline{X}), \\ & L \in \mathcal{S}^n_+,\ S \in \mathcal{S}^n_+. \end{aligned} \tag{4.7}$$

Indeed, the ℓ_1-norm has been shown as a successful surrogate for the sparsity (i.e., the number of nonzero entries) of a vector in compressed sensing [27, 26, 43, 42], while the nuclear norm has been observed and then demonstrated to be an effective surrogate for the rank of a matrix in low-rank matrix recovery [97, 57, 105, 25]. Based on these results, the "nuclear norm plus ℓ_1-norm" approach was studied recently as a tractable convex relaxation for the "low-rank plus sparse" matrix decomposition, and a number of theoretical guarantees provide conditions under which this heuristic is capable of exactly recovering the low-rank and sparse components from the completely fixed data (i.e., $\mathcal{F}^c = \emptyset$) [32, 21, 61, 73] or the completely sampled data (i.e., $\mathcal{F} = \emptyset$) [21, 89, 33]. In the rest of this chapter, we are interested in establishing characterization when the solution to problem (4.6) or (4.7) turns out to be the true low-rank component $\overline{L}$ and the true sparse component $\overline{S}$ with both partially fixed and partially sampled data.

4.2 Identifiability conditions

Generally speaking, the low-rank and sparse decomposition problem is ill-posed in the absent of any further assumptions. Even though the completely fixed data is given, it is still possible that these two components are not identifiable. For instance, the low-rank component may be sparse, or the sparse component may

has low-rank. In these two natural identifiability problems, the decomposition is usually not unique. Therefore, additional conditions should be imposed on the low-rank and sparse components in order to enhance their identifiability from the given data.

For the purpose of avoiding the first identifiability problem, we require that the low-rank component $\overline{L}$ should not have too sparse singular vectors. To achieve this, we borrow the standard notion of incoherence introduced in [25] for matrix completion problem. Essentially, the incoherence assumptions control the dispersion degree of the information contained in the column space and row space of the matrix $\overline{L}$. Suppose that the matrix $\overline{L}$ of rank r has a reduced SVD

$$\overline{L} = U_1 \Sigma V_1^T, \tag{4.8}$$

where $U_1 \in \mathcal{O}^{n_1 \times r}$, $V_1 \in \mathcal{O}^{n_2 \times r}$, and $\Sigma \in \mathbb{R}^{r \times r}$ is the diagonal matrix with the non-zero singular values of $\overline{L}$ being arranged in the non-increasing order. Notice that $U_1 = V_1$ when $\overline{L} \in \mathcal{S}_+^n$. Mathematically, the incoherence of the low-rank component $\overline{L}$ can be described as follows.

Assumption 4.2. *The low-rank component $\overline{L}$ of rank r is incoherent with parameters μ_0 and μ_1. That is, there exist some μ_0 and μ_1 such that*

$$\max_{1 \le i \le n_1} \left\| U_1^T e_i \right\|_2 \le \sqrt{\mu_0 \frac{r}{n_1}}, \quad \max_{1 \le j \le n_2} \left\| V_1^T e_j \right\|_2 \le \sqrt{\mu_0 \frac{r}{n_2}},$$

and

$$\left\| U_1 V_1^T \right\|_\infty \le \mu_1 \sqrt{\frac{r}{n_1 n_2}}.$$

Since $\|U_1^T\|_F^2 = \|V_1^T\|_F^2 = r$, $\|U_1^T e_i\|_2 \le 1$ and $\|V_1^T e_i\|_2 \le 1$, we know that $1 \le \mu_0 \le \frac{\max\{n_1, n_2\}}{r}$. Moreover, by using the Cauchy-Schwarz inequality and the fact that $\|U_1 V_1^T\|_F^2 = r$, we can see that $1 \le \mu_1 \le \mu_0 \sqrt{r}$.

Let $\mathcal{T}$ and $\mathcal{T}^\perp$ be the tangent space and its orthogonal complement defined in the same way as (2.2) and (2.3), respectively. Choose U_2 and V_2 such that

$U = [U_1, U_2]$ and $V = [V_1, V_2]$ are both orthogonal matrices. Notice that $U = V$ when $\overline{L} \in \mathcal{S}^n_+$. The orthogonal projection $\mathcal{P}_{\mathcal{T}}$ onto $\mathcal{T}$ and the orthogonal projection $\mathcal{P}_{\mathcal{T}^\perp}$ onto $\mathcal{T}^\perp$ are given by (2.4) and (2.5), respectively. Then it follows from Assumption 4.2 that for any $1 \le i \le n_1$ and any $1 \le j \le n_2$,

$$\begin{aligned}\left\|\mathcal{P}_{\mathcal{T}}\left(e_i e_j^T\right)\right\|_F^2 &= \left\langle \mathcal{P}_{\mathcal{T}}\left(e_i e_j^T\right), e_i e_j^T \right\rangle \\ &= \left\|U_1^T e_i\right\|_2^2 + \left\|V_1^T e_j\right\|_2^2 - \left\|U_1^T e_i\right\|_2^2 \left\|V_1^T e_j\right\|_2^2 \\ &\le \left\|U_1^T e_i\right\|_2^2 + \left\|V_1^T e_j\right\|_2^2 \le \mu_0 r \frac{(n_1+n_2)}{n_1 n_2},\end{aligned}$$

and for any $1 \le i < j \le n$ (where $n = n_1 = n_2$),

$$\begin{aligned}\left\|\mathcal{P}_{\mathcal{T}}\left(e_i e_j^T + e_j e_i^T\right)\right\|_F^2 &= \left\langle \mathcal{P}_{\mathcal{T}}\left(e_i e_j^T + e_j e_i^T\right), e_i e_j^T + e_j e_i^T \right\rangle \\ &\le \left\|U_1^T e_i\right\|_2^2 + \left\|U_1^T e_j\right\|_2^2 + \left\|V_1^T e_i\right\|_2^2 + \left\|V_1^T e_j\right\|_2^2 \le \frac{4\mu_0 r}{n}.\end{aligned}$$

Thus in our general setting, for any $j \in \{1, \ldots, d\}$, we have

$$\left\|\mathcal{P}_{\mathcal{T}}(\Theta_j)\right\|_F^2 \le \mu_0 r \frac{(n_1+n_2)}{n_1 n_2}. \tag{4.9}$$

As noted by [32], simply bounding the number of nonzero entries in the sparse component does not suffice, since the sparsity pattern also plays an important role in guaranteeing the identifiability. In order to prevent the second identifiability problem, we require that the sparse component $\overline{S}$ should not be too dense in each row and column. This was also called the "bounded degree" (i.e., bounded number of nonzeros per row/column) assumption used in [32, 31].

Assumption 4.3. *The sparse component $\overline{S}$ has at most k nonzero entries in each row and column, for some integer $0 \le k \le \max\{n_1, n_2\}$.*

Geometrically, the following lemma states that the angle (cf. [37] and [38, Chapter 9]) between the tangent space $\mathcal{T}$ of the low-rank component $\overline{L}$ and the supporting space Γ of the sparse component $\overline{S}$ is bounded away from zero. As we

will see later, this is extremely crucial for unique decomposition. An analogous result can be found from [33, Lemma 10].

Lemma 4.1. *Under Assumption 4.2 and 4.3, for any $Z \in \mathbb{V}^{n_1\times n_2}$, we have*

$$\|\mathcal{P}_{\mathcal{T}}\mathcal{P}_{\Gamma}\mathcal{P}_{\mathcal{T}}(Z)\|_F \le \left(\sqrt{\frac{\mu_0 rk}{n_1}} + \sqrt{\frac{\mu_0 rk}{n_2}}\right)\|\mathcal{P}_{\mathcal{T}}(Z)\|_F.$$

Proof. From (2.4) and the choice of U_2, for any $Z \in \mathbb{V}^{n_1\times n_2}$, we can write

$$\mathcal{P}_{\mathcal{T}}(Z) = U_1U_1^TZ + U_2U_2^TZV_1V_1^T.$$

Then we have

$$\|\mathcal{P}_{\mathcal{T}}(Z)\|_F^2 = \left\|U_1U_1^TZ\right\|_F^2 + \left\|U_2U_2^TZV_1V_1^T\right\|_F^2 = \left\|U_1^TZ\right\|_F^2 + \left\|U_2^TZV_1\right\|_F^2. \quad (4.10)$$

On the one hand, for any $1 \le j \le n_2$, by using the Cauchy-Schwarz inequality and Assumption 4.2, we obtain that

$$\begin{aligned}\left\|U_1U_1^TZe_j\right\|_\infty &= \max_{1\le i\le n_1}\left|e_i^TU_1U_1^TZe_j\right| \\ &\le \max_{1\le i\le n_1}\left\|U_1^Te_i\right\|_2\left\|U_1^TZe_j\right\|_2 \le \sqrt{\frac{\mu_0 r}{n_1}}\left\|U_1^TZe_j\right\|_2,\end{aligned}$$

which, together with Assumption 4.3, yields that

$$\left\|\mathcal{P}_{\Gamma}\big(U_1U_1^TZ\big)e_j\right\|_2 \le \sqrt{k}\left\|U_1U_1^TZe_j\right\|_\infty \le \sqrt{\frac{\mu_0 rk}{n_1}}\left\|U_1^TZe_j\right\|_2.$$

This gives that

$$\begin{aligned}\|\mathcal{P}_{\Gamma}\big(U_1U_1^TZ\big)\|_F^2 &= \sum_{1\le j\le n_2}\left\|\mathcal{P}_{\Gamma}\big(U_1U_1^TZ\big)e_j\right\|_2^2 \\ &\le \frac{\mu_0 rk}{n_1}\sum_{1\le j\le n_2}\left\|U_1^TZe_j\right\|_2^2 = \frac{\mu_0 rk}{n_1}\left\|U_1^TZ\right\|_F^2. \quad (4.11)\end{aligned}$$

On the other hand, for any $1 \le i \le n_1$, from the Cauchy-Schwarz inequality and Assumption 4.2, we know that

$$\begin{aligned}\left\|e_i^TU_2U_2^TZV_1V_1^T\right\|_\infty &= \max_{1\le j\le n_2}\left|e_i^TU_2U_2^TZV_1V_1^Te_j\right| \\ &\le \max_{1\le j\le n_2}\left\|V_1^Te_j\right\|_2\left\|e_i^TU_2U_2^TZV_1\right\|_2 \le \sqrt{\frac{\mu_0 r}{n_2}}\left\|e_i^TU_2U_2^TZV_1\right\|_2,\end{aligned}$$

which, together with Assumption 4.3, implies that

$$\left\|e_i^T\mathcal{P}_\Gamma\left(U_2U_2^TZV_1V_1^T\right)\right\|_2 \le \sqrt{k}\left\|e_i^TU_2U_2^TZV_1V_1^T\right\|_\infty \le \sqrt{\frac{\mu_0 rk}{n_2}}\left\|e_i^TU_2U_2^TZV_1\right\|_2.$$

It then follows that

$$\begin{aligned}\|\mathcal{P}_\Gamma\left((U_2U_2^TZV_1V_1^T\right)\|_F^2 &= \sum_{1\le i\le n_1}\left\|e_i^T\mathcal{P}_\Gamma\left(U_2U_2^TZV_1V_1^T\right)\right\|_2^2\\ &\le \frac{\mu_0 rk}{n_2}\sum_{1\le i\le n_1}\left\|e_i^TU_2U_2^TZV_1\right\|_2^2\\ &= \frac{\mu_0 rk}{n_2}\left\|U_2U_2^TZV_1\right\|_F^2 = \frac{\mu_0 rk}{n_2}\left\|U_2^TZV_1\right\|_F^2. \end{aligned} \tag{4.12}$$

Thus, by combining (4.10), (4.11) and (4.12), we derive that

$$\begin{aligned}\|\mathcal{P}_\mathcal{T}\mathcal{P}_\Gamma\mathcal{P}_\mathcal{T}(Z)\|_F &\le \|\mathcal{P}_\Gamma\mathcal{P}_\mathcal{T}(Z)\|_F\\ &\le \|\mathcal{P}_\Gamma\left(U_1U_1^TZ\right)\|_F + \|\mathcal{P}_\Gamma\left(U_2U_2^TZV_1V_1^T\right)\|_F\\ &\le \sqrt{\frac{\mu_0 rk}{n_1}}\left\|U_1^TZ\right\|_F + \sqrt{\frac{\mu_0 rk}{n_2}}\left\|U_2^TZV_1\right\|_F\\ &\le \left(\sqrt{\frac{\mu_0 rk}{n_1}} + \sqrt{\frac{\mu_0 rk}{n_2}}\right)\|\mathcal{P}_\mathcal{T}(Z)\|_F,\end{aligned}$$

which completes the proof. □

The next lemma plays a similar role as Lemma 4.1 in identifying the low-rank and sparse components. Basically, it says that for any matrix in Γ, the operator $\mathcal{P}_\mathcal{T}$ does not increase the matrix ℓ_∞-norm. This implies that the tangent space $\mathcal{T}$ of the low-rank component $\overline{L}$ and the supporting space Γ of the sparse component $\overline{S}$ has only trivial intersection, i.e., $\mathcal{T}\cap\Gamma = \{0\}$. One may refer to [33, (21)] for an analogous inequality.

Lemma 4.2. *Suppose that Assumption 4.2 and 4.3 hold. Then for any $Z \in \mathbb{V}^{n_1\times n_2}$, we have*

$$\|\mathcal{P}_\mathcal{T}\mathcal{P}_\Gamma(Z)\|_\infty \le \left(\sqrt{\frac{\mu_0 rk}{n_1}} + \sqrt{\frac{\mu_0 rk}{n_2}} + \frac{\mu_0 rk}{\sqrt{n_1n_2}}\right)\|\mathcal{P}_\Gamma(Z)\|_\infty.$$

Proof. For any $Z \in \mathbb{V}^{n_1 \times n_2}$, it follows from (2.4) that

$$\begin{aligned}
&\|\mathcal{P}_{\mathcal{T}}\mathcal{P}_{\Gamma}(Z)\|_{\infty} \\
&\leq \left\|U_1U_1^T\mathcal{P}_{\Gamma}(Z)\right\|_{\infty} + \left\|\mathcal{P}_{\Gamma}(Z)V_1V_1^T\right\|_{\infty} + \left\|U_1U_1^T\mathcal{P}_{\Gamma}(Z)V_1V_1^T\right\|_{\infty} \\
&\leq \max_{1\leq i\leq n_1}\left\|U_1U_1^Te_i\right\|_2 \max_{1\leq j\leq n_2}\left\|\mathcal{P}_{\Gamma}(Z)e_j\right\|_2 + \max_{1\leq i\leq n_1}\left\|e_i^T\mathcal{P}_{\Gamma}(Z)\right\|_2 \max_{1\leq j\leq n_2}\left\|V_1V_1^Te_j\right\|_2 \\
&\quad + \max_{1\leq i\leq n_1}\left\|U_1U_1^Te_i\right\|_2\|\mathcal{P}_{\Gamma}(Z)\| \max_{1\leq j\leq n_2}\left\|V_1V_1^Te_j\right\|_2 \\
&\leq \sqrt{\frac{\mu_0 r}{n_1}}\sqrt{k}\left\|\mathcal{P}_{\Gamma}(Z)\right\|_{\infty} + \sqrt{k}\left\|\mathcal{P}_{\Gamma}(Z)\right\|_{\infty}\sqrt{\frac{\mu_0 r}{n_2}} + \sqrt{\frac{\mu_0 r}{n_1}}k\|\mathcal{P}_{\Gamma}(Z)\|_{\infty}\sqrt{\frac{\mu_0 r}{n_2}} \\
&= \left(\sqrt{\frac{\mu_0 rk}{n_1}} + \sqrt{\frac{\mu_0 rk}{n_2}} + \frac{\mu_0 rk}{\sqrt{n_1 n_2}}\right)\|\mathcal{P}_{\Gamma}(Z)\|_{\infty},
\end{aligned}$$

where the first inequality comes from the triangular inequality, the second inequality is due to the Cauchy-Schwarz inequality, and the third inequality is a consequence of Assumption 4.2, 4.3 and Lemma 2.2. This completes the proof. □

4.3 Exact recovery guarantees

Inheriting the success from the nuclear norm and the ℓ_1-norm in recovering "simple object" such as low-rank matrix and sparse vector, the "nuclear norm plus ℓ_1-norm" approach has recently been proved to be able to exactly recover the low-rank and sparse components in the problem of "low-rank plus sparse" matrix decomposition from the completely fixed data (i.e., $\mathcal{F}^c = \emptyset$) [32, 21, 61, 73] or the completely sampled data (i.e., $\mathcal{F} = \emptyset$) [21, 89, 33]. In this section, we will establish such exact recovery guarantees when the given data consists of both fixed and sampled basic coefficients in the sense of Assumption 4.1 with $\mathcal{F} \neq \emptyset$ and $\mathcal{F}^c \neq \emptyset$, provided, of course, that the identifiability conditions (i.e., Assumption 4.2 and Assumption 4.3) are satisfied together with the rank of the low-rank component and the sparsity level of the sparse component being reasonably small. We first present the recovery theorem for problem (4.6).

Theorem 4.3. *Let $\overline{X} = \overline{L} + \overline{S} \in \mathbb{V}^{n_1 \times n_2}$ be an unknown matrix such that Assumption 4.2 holds at the low-rank component $\overline{L}$ and that Assumption 4.3 holds at the sparse component $\overline{S}$ with*

$$\sqrt{\frac{\mu_0 rk}{n_1}} + \sqrt{\frac{\mu_0 rk}{n_2}} + \frac{\mu_0 rk}{\sqrt{n_1 n_2}} \leq \frac{1}{16}.$$

Under Assumption 4.1 with $\mathcal{F} \neq \emptyset$ and $\mathcal{F}^c \neq \emptyset$, for any absolute constant $c > 0$, if the sample size satisfies

$$m \geq 1024(1+c)\max\{\mu_1^2, \mu_0\} r \max\left\{\frac{(n_1+n_2)}{n_1 n_2} d_s,\, 1\right\} \log^2(2n_1 n_2),$$

then with probability at least

$$1 - (2n_1 n_2)^{-c} - \frac{3}{2}\log(2n_1 n_2)\left[(2n_1 n_2)^{-c} + 2(n_1 n_2)^{-c} + (n_1+n_2)^{-c}\right],$$

the optimal solution to problem (4.6) with the tradeoff parameter $\frac{48}{13}\frac{\mu_1\sqrt{r}}{\sqrt{n_1 n_2}} \leq \rho \leq \frac{4\mu_1\sqrt{r}}{\sqrt{n_1 n_2}}$ is unique and equal to $(\overline{L}, \overline{S})$.

Theorem 4.3 reveals the power of the convex optimization formulation (4.6) for the problem of exact low-rank and sparse matrix decomposition with fixed and sampled basis coefficients. Firstly, it is worth pointing out that the restriction on the rank r of $\overline{L}$ and the sparsity level k of $\overline{S}$ is fairly mild. For example, a more restricted condition $\mu_0 rk \ll \min\{n_1, n_2\}$ is quite likely to hold for a strict factor model in which $\overline{S}$ is known to be a diagonal matrix and consequently $k = 1$. Secondly, when d_s, the cardinality of the unfixed index set $\mathcal{F}^c$, is significantly greater than $\max\{n_1, n_2\}$, the vanishing fraction of samples, i.e., $\frac{m}{d_s}$, is already sufficient for exact recovery, which is particularly desirable in the high-dimensional setting where $m \ll d_s$. Lastly, it is interesting to note that the conclusion of the above theorem holds for a range of values of the tradeoff parameter ρ. From the computational point of view, this is probably an attractive advantage that may allow the utilization of simple numerical algorithms, such as the bisection method, for searching an appropriate ρ when the involved μ_1 and r are unknown.

Analogously, the following theorem provides exact recovery guarantee for problem (4.7) when the low-rank and sparse components are both assumed to be symmetric and positive semidefinite.

Theorem 4.4. *Let $\overline{X} = \overline{L} + \overline{S} \in \mathcal{S}^n$ be an unknown matrix such that Assumption 4.2 holds at the low-rank component $\overline{L} \in \mathcal{S}^n_+$ and that Assumption 4.3 holds at the sparse component $\overline{S} \in \mathcal{S}^n_+$ with*

$$2\sqrt{\frac{\mu_0 rk}{n}} + \frac{\mu_0 rk}{n} \leq \frac{1}{16}.$$

Under Assumption 4.1 with $\mathcal{F} \neq \emptyset$ and $\mathcal{F}^c \neq \emptyset$, for any absolute constant $c > 0$, if the sample size satisfies

$$m \geq 1024(1+c)\max\{\mu_1^2, \mu_0\} r \max\left\{\frac{2d_s}{n}, 1\right\} \log^2\big(2n^2\big),$$

then the optimal solution to problem (4.7) with the tradeoff parameter $\frac{48}{13}\frac{\mu_1\sqrt{r}}{n} \leq \rho \leq \frac{4\mu_1\sqrt{r}}{n}$ is unique and equal to $(\overline{L}, \overline{S})$ with probability at least

$$1 - \big(\sqrt{2}n\big)^{-2c} - \frac{3}{2}\log\big(2n^2\big)\big[\big(\sqrt{2}n\big)^{-2c} + 2n^{-2c} + (2n)^{-c}\big].$$

4.3.1 Properties of the sampling operator

Before proceeding to establish the recovery theorems, we need to introduce some critical properties of the sampling operator $\mathcal{R}_\Omega$ defined in (4.4), where Ω is a multiset of indices with size m sampled uniformly with replacement from the unfixed index set $\mathcal{F}^c$. Recall that $d_s = |\mathcal{F}^c|$ and that the fixed index set $\mathcal{F}$ is partitioned into $\Gamma = \{j \,|\, \langle \Theta_j, \overline{S}\rangle \neq 0,\, j = 1, \ldots, d\}$ (with the assumption that $\Gamma \subseteq \mathcal{F}$) and $\Gamma_0 = \{j \,|\, \langle \Theta_j, \overline{S}\rangle = 0,\, j \in \mathcal{F}\}$.

Intuitively, it is desirable to control the maximum number of repetitions of any index in Ω so that more information of the true unknown matrix $\overline{X}$ could be obtained via sampling. Thanks to the model of sampling with replacement and the

noncommutative Bernstein inequality for random matrices with bounded spectral norm, we can show that the maximum duplication in Ω, which is also the spectral norm of $\mathcal{R}_\Omega$, is at most of order $\log(n_1 n_2)$ with high probability. A analogous result was proved in [104, Proposition 5] by applying a standard Chernoff bound for the Bernoulli distribution (cf. [71]).

Proposition 4.5. *For any $c > 0$, if the number of samples satisfies $m < \frac{8}{3}(1+c)d_s \log(2n_1n_2)$, then with probability at least $1-(2n_1n_2)^{-c}$, it holds that*

$$\left\| \mathcal{R}_\Omega - \frac{m}{d_s}\mathcal{P}_{\mathcal{F}^c} \right\| \leq \frac{8}{3}(1+c)\log(2n_1n_2).$$

Consequently, with the same probability, we have

$$\left\| \mathcal{P}_{\Gamma_0} + \frac{d_s}{m}\mathcal{R}_\Omega \right\| \leq \|\mathcal{P}_{\Gamma_0 \cup \mathcal{F}^c}\| + \frac{8}{3}(1+c)\log(2n_1n_2)\frac{d_s}{m} < \frac{16}{3}(1+c)\log(2n_1n_2)\frac{d_s}{m}.$$

Proof. For any uniform random variable ω over $\mathcal{F}^c$, define the random operator $\mathcal{Z}_\omega : \mathbb{V}^{n_1\times n_2} \to \mathbb{V}^{n_1\times n_2}$ associated with ω by

$$\mathcal{Z}_\omega(Z) := \langle \Theta_\omega, Z\rangle \Theta_\omega - \frac{1}{d_s}\mathcal{P}_{\mathcal{F}^c}(Z), \quad Z \in \mathbb{V}^{n_1\times n_2}.$$

Note that $\mathcal{Z}_\omega$ is self-adjoint, i.e., $\mathcal{Z}_\omega^* = \mathcal{Z}_\omega$. From (4.4), we write that

$$\mathcal{R}_\Omega - \frac{m}{d_s}\mathcal{P}_{\mathcal{F}^c} = \sum_{l=l}^{m}\left(\langle \Theta_{\omega_l}, \cdot\rangle \Theta_{\omega_l} - \frac{1}{d_s}\mathcal{P}_{\mathcal{F}^c}\right) = \sum_{l=l}^{m} \mathcal{Z}_{\omega_l}.$$

By using (4.5), we can verify that

$$\mathbb{E}[\mathcal{Z}_\omega] = 0 \ \text{ and } \ \|\mathcal{Z}_\omega\| \leq 1 =: K.$$

Moreover, a direct calculation shows that for any $Z \in \mathbb{V}^{n_1\times n_2}$,

$$\mathcal{Z}_\omega^*\mathcal{Z}_\omega(Z) = \mathcal{Z}_\omega\mathcal{Z}_\omega^*(Z) = \left(1 - \frac{2}{d_s}\right)\langle \Theta_\omega, Z\rangle\Theta_\omega + \frac{1}{d_s^2}\mathcal{P}_{\mathcal{F}^c}(Z).$$

As a consequence, we obtain that

$$\mathbb{E}[\mathcal{Z}_\omega^*\mathcal{Z}_\omega] = \mathbb{E}[\mathcal{Z}_\omega\mathcal{Z}_\omega^*] = \left(\frac{1}{d_s} - \frac{1}{d_s^2}\right)\mathcal{P}_{\mathcal{F}^c},$$

which yields that

$$\|\mathbb{E}[\mathcal{Z}_\omega^*\mathcal{Z}_\omega]\| = \|\mathbb{E}[\mathcal{Z}_\omega\mathcal{Z}_\omega^*]\| \le \frac{1}{d_s} - \frac{1}{d_s^2} \le \frac{1}{d_s} =: \varsigma^2.$$

Let $t^* := \frac{8}{3}(1+c)\log(2n_1n_2)$. If $m < \frac{8}{3}(1+c)d_s\log(2n_1n_2)$, then $t^* > \frac{m\varsigma^2}{K}$. Since $\{\mathcal{Z}_{\omega_l}\}_{l=1}^m$ are i.i.d. copies of $\mathcal{Z}_\omega$, from Lemma 2.5, we know that

$$\mathbb{P}\left[\left\|\sum_{l=1}^m \mathcal{Z}_{\omega_l}\right\| > t^*\right] \le 2n_1n_2\exp\left(-\frac{3}{8}\frac{t^*}{K}\right) \le (2n_1n_2)^{-c}.$$

Since $\|\mathcal{P}_{\Gamma_0\cup\mathcal{F}^c}\| = 1 < \frac{8}{3}(1+c)\log(2n_1n_2)\frac{d_s}{m}$, the proof is completed by using the triangular inequality. □

When the fixed index set $\mathcal{F} = \emptyset$, i.e., Ω is sampled from the whole index set, it has been shown that the operator $\frac{d}{m}\mathcal{P}_\mathcal{T}\mathcal{R}_\Omega\mathcal{P}_\mathcal{T}$ is very close to its expectation $\mathcal{P}_\mathcal{T}\mathcal{P}_{\mathcal{F}^c}\mathcal{P}_\mathcal{T}$ with high probability for the Bernoulli sampling [25, Theorem 4.1] and the uniform sampling with replacement [104, Theorem 6], if the number of samples is sufficiently large. The next proposition generalizes these results to the case that $\mathcal{F} \neq \emptyset$. One may refer to [33, Lemma 2] for an analog in the Bernoulli model.

Proposition 4.6. *Under Assumption 4.2, for any $c > 0$, if the number of samples satisfies $m \ge \frac{8}{3}(1+c)\max\{\mu_0 r\frac{(n_1+n_2)}{n_1n_2}d_s, 1\}\log(2n_1n_2)$, then with probability at least $1-(2n_1n_2)^{-c}$, it holds that*

$$\left\|\frac{d_s}{m}\mathcal{P}_\mathcal{T}\mathcal{R}_\Omega\mathcal{P}_\mathcal{T} - \mathcal{P}_\mathcal{T}\mathcal{P}_{\mathcal{F}^c}\mathcal{P}_\mathcal{T}\right\| \le \sqrt{\frac{8}{3}(1+c)\max\left\{\mu_0 r\frac{(n_1+n_2)}{n_1n_2}d_s,\, 1\right\}\frac{\log(2n_1n_2)}{m}}.$$

Furthermore, if Assumption 4.3 also holds, with the same probability, we have

$$\begin{aligned}&\left\|\mathcal{P}_\mathcal{T}\left(\mathcal{P}_{\Gamma_0} + \frac{d_s}{m}\mathcal{R}_\Omega\right)\mathcal{P}_\mathcal{T} - \mathcal{P}_\mathcal{T}\right\| \\ &\le \sqrt{\frac{8}{3}(1+c)\max\left\{\mu_0 r\frac{(n_1+n_2)}{n_1n_2}d_s,\, 1\right\}\frac{\log(2n_1n_2)}{m}} + \sqrt{\frac{\mu_0 rk}{n_1}} + \sqrt{\frac{\mu_0 rk}{n_2}}.\end{aligned}$$

Proof. Let ω be a uniform random variable over $\mathcal{F}^c$. Define the random operator $\mathcal{Z}_\omega : \mathbb{V}^{n_1\times n_2} \to \mathbb{V}^{n_1\times n_2}$ associated with ω by

$$\mathcal{Z}_\omega(Z) := \langle \mathcal{P}_\mathcal{T}(\Theta_\omega), Z\rangle \mathcal{P}_\mathcal{T}(\Theta_\omega) - \frac{1}{d_s}\mathcal{P}_\mathcal{T}\mathcal{P}_{\mathcal{F}^c}\mathcal{P}_\mathcal{T}(Z), \quad Z \in \mathbb{V}^{n_1\times n_2}.$$

Note that $\mathcal{Z}_\omega$ is self-adjoint, that is, $\mathcal{Z}_\omega^* = \mathcal{Z}_\omega$. According to (4.3) and (4.4), we have the following decomposition

$$\mathcal{P}_\mathcal{T}\mathcal{R}_\Omega\mathcal{P}_\mathcal{T} - \frac{m}{d_s}\mathcal{P}_\mathcal{T}\mathcal{P}_{\mathcal{F}^c}\mathcal{P}_\mathcal{T} = \sum_{l=1}^{m}\left(\langle \mathcal{P}_\mathcal{T}(\Theta_{\omega_l}), \cdot\rangle \mathcal{P}_\mathcal{T}(\Theta_{\omega_l}) - \frac{1}{d_s}\mathcal{P}_\mathcal{T}\mathcal{P}_{\mathcal{F}^c}\mathcal{P}_\mathcal{T}\right) = \sum_{l=l}^{m}\mathcal{Z}_{\omega_l}.$$

From (4.5) and the linearity of $\mathcal{P}_\mathcal{T}$, we derive that

$$\begin{aligned}\mathbb{E}[\mathcal{Z}_\omega] &= \sum_{j\in\mathcal{F}^c}\frac{1}{d_s}\langle \mathcal{P}_\mathcal{T}(\Theta_j), \cdot\rangle \mathcal{P}_\mathcal{T}(\Theta_j) - \frac{1}{d_s}\mathcal{P}_\mathcal{T}\mathcal{P}_{\mathcal{F}^c}\mathcal{P}_\mathcal{T} \\ &= \frac{1}{d_s}\mathcal{P}_\mathcal{T}\left(\sum_{j\in\mathcal{F}^c}\langle \Theta_j, \mathcal{P}_\mathcal{T}(\cdot)\rangle\Theta_j\right) - \frac{1}{d_s}\mathcal{P}_\mathcal{T}\mathcal{P}_{\mathcal{F}^c}\mathcal{P}_\mathcal{T} = 0.\end{aligned}$$

Since $\langle \mathcal{P}_\mathcal{T}(\Theta_\omega), \cdot\rangle \mathcal{P}_\mathcal{T}(\Theta_\omega)$ and $\mathcal{P}_\mathcal{T}\mathcal{P}_{\mathcal{F}^c}\mathcal{P}_\mathcal{T}$ are both self-adjoint positive semidefinite linear operators, we know from Lemma 2.1 and (4.9) that

$$\begin{aligned}\|\mathcal{Z}_\omega\| &\le \max\left\{\|\langle \mathcal{P}_\mathcal{T}(\Theta_\omega), \cdot\rangle \mathcal{P}_\mathcal{T}(\Theta_\omega)\|, \frac{1}{d_s}\|\mathcal{P}_\mathcal{T}\mathcal{P}_{\mathcal{F}^c}\mathcal{P}_\mathcal{T}\|\right\} \\ &= \max\left\{\|\mathcal{P}_\mathcal{T}(\Theta_\omega)\|_F^2, \frac{1}{d_s}\right\} \le \max\left\{\mu_0 r\frac{(n_1+n_2)}{n_1n_2}, \frac{1}{d_s}\right\} =: K.\end{aligned}$$

Moreover, by using Lemma 2.1 and (4.9) again, we obtain that

$$\begin{aligned}\left\|\mathbb{E}\left[\mathcal{Z}_\omega^2\right]\right\| &= \left\|\mathbb{E}\left[\left(\langle \mathcal{P}_\mathcal{T}(\Theta_\omega), \cdot\rangle \mathcal{P}_\mathcal{T}(\Theta_\omega)\right)^2\right] - \frac{1}{d_s^2}\mathcal{P}_\mathcal{T}\mathcal{P}_{\mathcal{F}^c}\mathcal{P}_\mathcal{T}\mathcal{P}_{\mathcal{F}^c}\mathcal{P}_\mathcal{T}\right\| \\ &= \left\|\|\mathcal{P}_\mathcal{T}(\Theta_\omega)\|_F^2\frac{1}{d_s}\mathcal{P}_\mathcal{T}\mathcal{P}_{\mathcal{F}^c}\mathcal{P}_\mathcal{T} - \frac{1}{d_s^2}\mathcal{P}_\mathcal{T}\mathcal{P}_{\mathcal{F}^c}\mathcal{P}_\mathcal{T}\mathcal{P}_{\mathcal{F}^c}\mathcal{P}_\mathcal{T}\right\| \\ &\le \max\left\{\|\mathcal{P}_\mathcal{T}(\Theta_\omega)\|_F^2\frac{1}{d_s}\|\mathcal{P}_\mathcal{T}\mathcal{P}_{\mathcal{F}^c}\mathcal{P}_\mathcal{T}\|, \frac{1}{d_s^2}\|\mathcal{P}_\mathcal{T}\mathcal{P}_{\mathcal{F}^c}\mathcal{P}_\mathcal{T}\mathcal{P}_{\mathcal{F}^c}\mathcal{P}_\mathcal{T}\|\right\} \\ &\le \frac{1}{d_s}\max\left\{\mu_0 r\frac{(n_1+n_2)}{n_1n_2}, \frac{1}{d_s}\right\} =: \varsigma^2.\end{aligned}$$

Choose $t^* := \sqrt{\frac{8}{3}(1+c)\log(2n_1n_2)\max\{\mu_0 r\frac{(n_1+n_2)}{n_1n_2}, \frac{1}{d_s}\}\frac{m}{d_s}}$. Then we have $t^* \le \frac{m\varsigma^2}{K}$ if $m \ge \frac{8}{3}(1+c)\log(2n_1n_2)\max\{\mu_0 r\frac{(n_1+n_2)}{n_1n_2}d_s, 1\}$. Since $\{\mathcal{Z}_{\omega_l}\}_{l=1}^m$ are i.i.d. copies of

$\mathcal{Z}_\omega$, by applying Lemma 2.5, we get that

$$\mathbb{P}\left[\left\|\sum_{l=1}^{m}\mathcal{Z}_{\omega_l}\right\| > t^*\right] \leq 2n_1n_2\exp\left(-\frac{3}{8}\frac{t^{*2}}{m\varsigma^2}\right) \leq (2n_1n_2)^{-c},$$

which completes the first part of the proof.

Furthermore, recall that $\Gamma \subseteq \mathcal{F}$, $\Gamma_0 = \mathcal{F} \setminus \Gamma$ and $\Gamma^c = \Gamma_0 \cup \mathcal{F}^c$. Thus, we have $\mathcal{P}_\mathcal{T} = \mathcal{P}_\mathcal{T}\mathcal{P}_\Gamma\mathcal{P}_\mathcal{T} + \mathcal{P}_\mathcal{T}\mathcal{P}_{\Gamma_0}\mathcal{P}_\mathcal{T} + \mathcal{P}_\mathcal{T}\mathcal{P}_{\mathcal{F}^c}\mathcal{P}_\mathcal{T}$. Then the second part of the proof follows by applying the triangular inequality and Lemma 4.1. □

The following proposition is an extension of [104, Lemma 8] and [21, Lemma 3.1] to include the case that $\mathcal{F} \neq \emptyset$. A similar result for the Bernoulli model was provided in [33, Lemma 13].

Proposition 4.7. *Let $Z \in \mathcal{T}$ be a fixed $n_1 \times n_2$ matrix. Under Assumption 4.2, for any $c > 0$, if the number of samples satisfies $m \geq \frac{32}{3}(1+c)\mu_0 r\frac{(n_1+n_2)}{n_1n_2}d_s\log(n_1n_2)$, it holds that*

$$\left\|\frac{d_s}{m}\mathcal{P}_\mathcal{T}\mathcal{R}_\Omega(Z) - \mathcal{P}_\mathcal{T}\mathcal{P}_{\mathcal{F}^c}(Z)\right\|_\infty \leq \sqrt{\frac{16(1+c)\mu_0 r(n_1+n_2)d_s\log(n_1n_2)}{3n_1n_2m}}\|\mathcal{P}_{\mathcal{F}^c}(Z)\|_\infty$$

with probability at least $1 - 2(n_1n_2)^{-c}$. Moreover, if Assumption 4.3 also holds, with the same probability, we have

$$\begin{aligned}&\left\|\mathcal{P}_\mathcal{T}\left(\mathcal{P}_{\Gamma_0} + \frac{d_s}{m}\mathcal{R}_\Omega\right)(Z) - Z\right\|_\infty \\ &\leq \left(\sqrt{\frac{16(1+c)\mu_0 r(n_1+n_2)d_s\log(n_1n_2)}{3n_1n_2m}} + \sqrt{\frac{\mu_0 rk}{n_1}} + \sqrt{\frac{\mu_0 rk}{n_2}} + \frac{\mu_0 rk}{\sqrt{n_1n_2}}\right)\|Z\|_\infty.\end{aligned}$$

Proof. Let $j \in \{1,\ldots,d\}$ be a fixed index. For any uniform random variable ω over $\mathcal{F}^c$, let z_ω be a random variable associated with ω defined by

$$z_\omega := \left\langle \Theta_j, \frac{d_s}{m}\langle\Theta_\omega, Z\rangle\mathcal{P}_\mathcal{T}(\Theta_\omega) - \frac{1}{m}\mathcal{P}_\mathcal{T}\mathcal{P}_{\mathcal{F}^c}(Z)\right\rangle.$$

By using (4.4) and the linearity of $\mathcal{P}_\mathcal{T}$, we have the following decomposition

$$\left\langle \Theta_j, \frac{d_s}{m}\mathcal{P}_\mathcal{T}\mathcal{R}_\Omega(Z) - \mathcal{P}_\mathcal{T}\mathcal{P}_{\mathcal{F}^c}(Z)\right\rangle = \sum_{l=1}^{m} z_{\omega_l}.$$

From (4.5), we know that $\mathbb{E}[z_\omega] = 0$. Note that

$$\begin{aligned}
|z_\omega| &\le \frac{d_s}{m}\big|\langle\Theta_\omega, Z\rangle\langle\Theta_j, \mathcal{P}_\mathcal{T}(\Theta_\omega)\rangle\big| + \frac{1}{m}\big|\langle\Theta_j, \mathcal{P}_\mathcal{T}\mathcal{P}_{\mathcal{F}^c}(Z)\rangle\big| \\
&= \frac{d_s}{m}\big|\langle\Theta_\omega, Z\rangle\langle\mathcal{P}_\mathcal{T}(\Theta_j), \mathcal{P}_\mathcal{T}(\Theta_\omega)\rangle\big| + \frac{1}{m}\bigg|\sum_{j'\in\mathcal{F}^c}\langle\Theta_{j'}, Z\rangle\langle\mathcal{P}_\mathcal{T}(\Theta_j), \mathcal{P}_\mathcal{T}(\Theta_{j'})\rangle\bigg| \\
&\le \frac{d_s}{m}\max_{j'\in\mathcal{F}^c}\big|\langle\Theta_{j'}, Z\rangle\big|\Big(\|\mathcal{P}_\mathcal{T}(\Theta_j)\|_F\|\mathcal{P}_\mathcal{T}(\Theta_\omega)\|_F + \|\mathcal{P}_\mathcal{T}(\Theta_j)\|_F\max_{j'\in\mathcal{F}^c}\|\mathcal{P}_\mathcal{T}(\Theta_{j'})\|_F\Big) \\
&\le 2\mu_0 r\frac{(n_1+n_2)}{n_1n_2}\frac{d_s}{m}\max_{j'\in\mathcal{F}^c}\big|\langle\Theta_{j'}, Z\rangle\big| \le 2\sqrt{2}\mu_0 r\frac{(n_1+n_2)}{n_1n_2}\frac{d_s}{m}\|\mathcal{P}_{\mathcal{F}^c}(Z)\|_\infty =: K,
\end{aligned}$$

where the first equality follows from (4.5) and the linearity of $\mathcal{P}_\mathcal{T}$, the second inequality is a consequence of the Cauchy-Schwarz inequality and the triangular inequality, the third inequality is due to (4.9), and the last inequality is from (4.1) and (4.2). In addition, a simple calculation gives that

$$\begin{aligned}
\mathbb{E}\big[z_\omega^2\big] &= \frac{d_s^2}{m^2}\mathbb{E}\big[\langle\Theta_\omega, Z\rangle^2\langle\Theta_j, \mathcal{P}_\mathcal{T}(\Theta_\omega)\rangle^2\big] - \frac{1}{m^2}\langle\Theta_j, \mathcal{P}_\mathcal{T}\mathcal{P}_{\mathcal{F}^c}(Z)\rangle^2 \\
&\le \frac{d_s}{m^2}\sum_{j'\in\mathcal{F}^c}\langle\Theta_{j'}, Z\rangle^2\langle\Theta_j, \mathcal{P}_\mathcal{T}(\Theta_{j'})\rangle^2 = \frac{d_s}{m^2}\sum_{j'\in\mathcal{F}^c}\langle\Theta_{j'}, Z\rangle^2\langle\mathcal{P}_\mathcal{T}(\Theta_j), \Theta_{j'}\rangle^2 \\
&\le \frac{d_s}{m^2}\max_{j'\in\mathcal{F}^c}\langle\Theta_{j'}, Z\rangle^2\sum_{j'\in\mathcal{F}^c}\langle\mathcal{P}_\mathcal{T}(\Theta_j), \Theta_{j'}\rangle^2 = \frac{d_s}{m^2}\max_{j'\in\mathcal{F}^c}\langle\Theta_{j'}, Z\rangle^2\|\mathcal{P}_{\mathcal{F}^c}\mathcal{P}_\mathcal{T}(\Theta_j)\|_F^2 \\
&\le \mu_0 r\frac{(n_1+n_2)}{n_1n_2}\frac{d_s}{m^2}\max_{j'\in\mathcal{F}^c}\langle\Theta_{j'}, Z\rangle^2 \le 2\mu_0 r\frac{(n_1+n_2)}{n_1n_2}\frac{d_s}{m^2}\|\mathcal{P}_{\mathcal{F}^c}(Z)\|_\infty^2 =: \varsigma^2,
\end{aligned}$$

where the third equality is owing to (4.5), the third inequality follows from (4.9), and the last inequality is a consequence of (4.1) and (4.2). With the choice of $t^* := \sqrt{\frac{16}{3}(1+c)\log(n_1n_2)\mu_0 r\frac{(n_1+n_2)}{n_1n_2}\frac{d_s}{m}}\|\mathcal{P}_{\mathcal{F}^c}(Z)\|_\infty$, we have $t^* \le \frac{m\varsigma^2}{K}$ if $m \ge \frac{32}{3}(1+c)\log(n_1n_2)\mu_0 r\frac{(n_1+n_2)}{n_1n_2}d_s$. Since $\{z_{\omega_l}\}_{l=1}^m$ are i.i.d. copies of z_ω, by applying Lemma 2.3, we obtain that

$$\mathbb{P}\left[\left|\sum_{l=1}^m z_{\omega_l}\right| > t^*\right] \le 2\exp\left(-\frac{3}{8}\frac{t^{*2}}{m\varsigma^2}\right) \le 2(n_1n_2)^{-(1+c)}.$$

The first part of the proof follows by taking the union bound of d ($\le n_1n_2$) terms.

Moreover, since $Z \in \mathcal{T}$ and $\mathcal{P}_\mathcal{T} = \mathcal{P}_\mathcal{T}\mathcal{P}_\Gamma + \mathcal{P}_\mathcal{T}\mathcal{P}_{\Gamma_0} + \mathcal{P}_\mathcal{T}\mathcal{P}_{\mathcal{F}^c}$, the second part of the proof is completed by using the triangular inequality and Lemma 4.2. □

The next proposition is a generalization of [25, Theorem 6.3] and [104, Theorem 7] to involve the case that $\mathcal{F} \neq \emptyset$. One may refer to [33, Lemma 12] for an analogous result based on the Bernoulli model.

Proposition 4.8. *Let $Z \in \mathbb{V}^{n_1 \times n_2}$ be a fixed matrix. For any $c > 0$, it holds that*

$$\left\| \left(\frac{d_s}{m} \mathcal{R}_\Omega - \mathcal{P}_{\mathcal{F}^c} \right)(Z) \right\| \leq \sqrt{\frac{8(1+c)\max\{d_s(n_1+n_2),\, n_1 n_2\}\log(n_1+n_2)}{3m}} \|\mathcal{P}_{\mathcal{F}^c}(Z)\|_\infty$$

with probability at least $1-(n_1+n_2)^{-c}$ provided that the number of samples satisfies $m \geq \frac{8}{3}(1+c)\frac{(d_s+\sqrt{n_1 n_2})^2}{\max\{d_s(n_1+n_2),\, n_1 n_2\}} \log(n_1+n_2)$. In addition, if Assumption 4.3 also holds, with the same probability, we have

$$\begin{aligned} &\left\| \left(\mathcal{P}_{\Gamma_0} + \frac{d_s}{m} \mathcal{R}_\Omega - \mathcal{I} \right)(Z) \right\| \\ &\leq \left(\sqrt{\frac{8(1+c)\max\{d_s(n_1+n_2),\, n_1 n_2\}\log(n_1+n_2)}{3m}} + k \right) \|Z\|_\infty, \end{aligned}$$

where $\mathcal{I} : \mathbb{V}^{n_1 \times n_2} \to \mathbb{V}^{n_1 \times n_2}$ is the identity operator.

Proof. Let ω be a uniform random variable over $\mathcal{F}^c$ and $Z_\omega \in \mathbb{V}^{n_1 \times n_2}$ be a random matrix associated with ω defined by

$$Z_\omega := \frac{d_s}{m} \langle \Theta_\omega, Z \rangle \Theta_\omega - \frac{1}{m} \mathcal{P}_{\mathcal{F}^c}(Z).$$

By using (4.4), we get that

$$\left(\frac{d_s}{m} \mathcal{R}_\Omega - \mathcal{P}_{\mathcal{F}^c} \right)(Z) = \sum_{l=1}^{m} \left(\frac{d_s}{m} \langle \Theta_{\omega_l}, Z \rangle \Theta_{\omega_l} - \frac{1}{m} \mathcal{P}_{\mathcal{F}^c}(Z) \right) = \sum_{l=1}^{m} Z_{\omega_l}.$$

From (4.5), we know that $\mathbb{E}[Z_\omega] = 0$. According to Lemma 2.2, we can check that

$$\|Z_\omega\| \leq \left(\frac{d_s}{m} + \frac{\sqrt{n_1 n_2}}{m} \right) \|\mathcal{P}_{\mathcal{F}^c}(Z)\|_\infty.$$

Moreover, from (4.1) and (4.2), we derive that

$$\left\| \mathbb{E}\left[\langle \Theta_\omega, Z \rangle^2 \Theta_\omega^T \Theta_\omega \right] \right\| \leq \begin{cases} \dfrac{1}{d_s} \max\{n_1, n_2\} \|\mathcal{P}_{\mathcal{F}^c}(Z)\|_\infty^2, & \text{if } \mathbb{V}^{n_1 \times n_2} = \mathbb{R}^{n_1 \times n_2}, \\ \dfrac{1}{d_s} (n_1+n_2) \|\mathcal{P}_{\mathcal{F}^c}(Z)\|_\infty^2, & \text{if } \mathbb{V}^{n_1 \times n_2} = \mathcal{S}^n, \end{cases}$$

which, together with Lemma 2.1 and Lemma 2.2, gives that

$$\begin{aligned}\left\|\mathbb{E}\left[Z_\omega^T Z_\omega\right]\right\| &= \frac{1}{m^2}\left\|\mathbb{E}\left[d_s^2\langle\Theta_\omega, Z\rangle^2\Theta_\omega^T\Theta_\omega\right] - \mathcal{P}_{\mathcal{F}^c}(Z)^T\mathcal{P}_{\mathcal{F}^c}(Z)\right\| \\ &\le \frac{1}{m^2}\max\left\{\left\|\mathbb{E}\left[d_s^2\langle\Theta_\omega, Z\rangle^2\Theta_\omega^T\Theta_\omega\right]\right\|,\ \left\|\mathcal{P}_{\mathcal{F}^c}(Z)\right\|^2\right\} \\ &\le \frac{1}{m^2}\max\left\{d_s(n_1+n_2),\ n_1n_2\right\}\|\mathcal{P}_{\mathcal{F}^c}(Z)\|_\infty^2.\end{aligned}$$

A similar calculation also holds for $\|\mathbb{E}[Z_\omega Z_\omega^T]\|$. Thus, we have

$$K := \frac{d_s+\sqrt{n_1n_2}}{m}\|\mathcal{P}_{\mathcal{F}^c}(Z)\|_\infty \quad \text{and} \quad \varsigma^2 := \frac{\max\{d_s(n_1+n_2), n_1n_2\}}{m^2}\|\mathcal{P}_{\mathcal{F}^c}(Z)\|_\infty^2.$$

By taking $t^* := \sqrt{\frac{8}{3}(1+c)\log(n_1+n_2)\frac{\max\{d_s(n_1+n_2),n_1n_2\}}{m}}\|\mathcal{P}_{\mathcal{F}^c}(Z)\|_\infty$, we have $t^* \le \frac{m\varsigma^2}{K}$ if $m \ge \frac{8}{3}(1+c)\log(n_1+n_2)\frac{(d_s+\sqrt{n_1n_2})^2}{\max\{d_s(n_1+n_2),n_1n_2\}}$. Since $\{Z_{\omega_l}\}_{l=1}^m$ are i.i.d. copies of Z_ω, from Lemma 2.5, we know that

$$\mathbb{P}\left[\left\|\sum_{l=1}^m Z_{\omega_l}\right\| > t^*\right] \le (n_1+n_2)\exp\left(-\frac{3}{8}\frac{t^{*2}}{m\varsigma^2}\right) \le (n_1+n_2)^{-c}.$$

This completes the first part of the proof.

In addition, note that $\mathcal{I} = \mathcal{P}_\Gamma + \mathcal{P}_{\Gamma_0} + \mathcal{P}_{\mathcal{F}^c}$. By applying Lemma 2.2, we obtain that$\|\mathcal{P}_\Gamma(Z)\| \le k\|\mathcal{P}_\Gamma(Z)\|_\infty$. Then the second part of the proof follows from the triangular inequality. □

4.3.2 Proof of the recovery theorems

In the literature on the problem of low-rank matrix recovery (see, e.g., [25, 28, 104, 68, 21, 89, 33, 124]), one popular strategy for establishing exact recovery results is first to provide dual certificates that certify the unique optimality of some related convex optimization problems, and then to show the existence of such dual certificates probabilistically by certain interesting but technical constructions. The proof of the recovery theorems in this chapter is along this line.

Sufficient optimality conditions

The first step of the proof is to characterize deterministic sufficient conditions, which are also verifiable with high probability in the assumed model, such that the convex optimization problems (4.6) and (4.7) have unique optimal solution. Here the convex nature of these two optimization problems plays a critical role.

Proposition 4.9. *Suppose that the tradeoff parameter satisfies* $0 < \rho < 1$ *and the sample size satisfies* $m \leq d_s$. *Suppose also that* $\|\mathcal{P}_{\Gamma_0} + \frac{d_s}{m}\mathcal{R}_\Omega\| \leq \gamma_1$ *for some* $\gamma_1 > 1$ *and that* $\|\mathcal{P}_\mathcal{T}(\mathcal{P}_{\Gamma_0} + \frac{d_s}{m}\mathcal{R}_\Omega)\mathcal{P}_\mathcal{T} - \mathcal{P}_\mathcal{T}\| \leq 1 - \gamma_2$ *for some* $0 < \gamma_2 < 1$. *Then under Assumption 4.1,* $(\overline{L}, \overline{S})$ *is the unique optimal solution to problem (4.6) if there exist dual certificates* A *and* $B \in \mathbb{V}^{n_1 \times n_2}$ *such that*

(a) $A \in \mathrm{Range}(\mathcal{P}_{\Gamma_0})$ *with* $\|A\|_\infty \leq \dfrac{3}{4}\rho$, *and* $B \in \mathrm{Range}(\mathcal{R}_\Omega)$,

(b) $\big\|U_1 V_1^T - \mathcal{P}_\mathcal{T}\big(\rho\,\mathrm{sign}(\overline{S}) + A + B\big)\big\|_F \leq \dfrac{\rho}{4}\dfrac{\sqrt{\gamma_2}}{\gamma_1}$,

(c) $\big\|\mathcal{P}_{\mathcal{T}^\perp}\big(\rho\,\mathrm{sign}(\overline{S}) + A + B\big)\big\| \leq \dfrac{3}{4}$.

Proof. Note that the constraints of problem (4.6) are all linear. Then any feasible solution is of the form $(\overline{L} + \Delta_L, \overline{S} + \Delta_S)$ with

$$\mathcal{P}_\mathcal{F}(\Delta_L + \Delta_S) = 0, \quad \mathcal{P}_{\mathcal{F}^c}(\Delta_S) = 0 \ \text{ and } \ \mathcal{R}_\Omega(\Delta_L) = 0. \tag{4.13}$$

We will show that the objective function of problem (4.6) at any feasible solution $(\overline{L} + \Delta_L, \overline{S} + \Delta_S)$ increases whenever $(\Delta_L, \Delta_S) \neq 0$, hence proving that $(\overline{L}, \overline{S})$ is the unique optimal solution. Choose $W_L \in \mathcal{T}^\perp$ and $W_S \in \Gamma^c = \Gamma_0 \cup \mathcal{F}^c$ such that

$$\begin{cases} \|W_L\| = 1, & \langle W_L, \Delta_L\rangle = \|\mathcal{P}_{\mathcal{T}^\perp}(\Delta_L)\|_*, \\ \|W_S\|_\infty = 1, & \langle W_S, \Delta_S\rangle = \|\mathcal{P}_{\Gamma^c}(\Delta_S)\|_1. \end{cases} \tag{4.14}$$

The existence of such W_L and W_S is guaranteed by the duality between $\|\cdot\|_*$ and $\|\cdot\|$, and that between $\|\cdot\|_1$ and $\|\cdot\|_\infty$. Moreover, it holds that

$$U_1 V_1^T + W_L \in \partial\|\overline{L}\|_* \quad \text{and} \quad \mathrm{sign}(\overline{S}) + W_S \in \partial\|\overline{S}\|_1. \tag{4.15}$$

Let $Q_L := \rho\,\mathrm{sign}(\overline{S}) + A + B$ and $Q_S := \rho\,\mathrm{sign}(\overline{S}) + A + C$ for any fixed $C \in \mathrm{Range}(\mathcal{P}_{\mathcal{F}^c})$ with $\|C\|_\infty \le \frac{3}{4}\rho$. Since $\rho\,\mathrm{sign}(\overline{S}) + A \in \mathcal{F} = \Gamma \cup \Gamma_0$, $B \in \mathrm{Range}(\mathcal{R}_\Omega)$ and $C \in \mathrm{Range}(\mathcal{P}_{\mathcal{F}^c})$, we know from (4.13) that

$$\langle Q_L, \Delta_L\rangle + \langle Q_S, \Delta_S\rangle = \langle \rho\,\mathrm{sign}(\overline{S}) + A, \Delta_L + \Delta_S\rangle + \langle B, \Delta_L\rangle + \langle C, \Delta_S\rangle = 0. \quad (4.16)$$

In addition, observe that $A \in \Gamma_0$ with $\|A\|_\infty \le \frac{3}{4}\rho$, $C \in \mathcal{F}^c$, $\Gamma^c = \Gamma_0 \cup \mathcal{F}^c$ and $\Gamma_0 \cap \mathcal{F}^c = \emptyset$. Consequently, we have $\mathcal{P}_{\Gamma^c}(Q_S) = A + C$ and $\|A + C\|_\infty \le \max\{\|A\|_\infty, \|C\|_\infty\} \le \frac{3}{4}\rho$. Then a direct calculation yields that

$$\begin{aligned}
&\|\overline{L} + \Delta_L\|_* + \rho\|\overline{S} + \Delta_S\|_1 - \|\overline{L}\|_* - \rho\|\overline{S}\|_1 \\
\ge\ & \langle U_1V_1^T + W_L, \Delta_L\rangle + \rho\langle \mathrm{sign}(\overline{S}) + W_S, \Delta_S\rangle \\
=\ & \langle U_1V_1^T + W_L - Q_L, \Delta_L\rangle + \langle \rho\,\mathrm{sign}(\overline{S}) + \rho W_S - Q_S, \Delta_S\rangle \\
=\ & \langle U_1V_1^T - \mathcal{P}_{\mathcal{T}}(Q_L), \mathcal{P}_{\mathcal{T}}(\Delta_L)\rangle + \langle W_L - \mathcal{P}_{\mathcal{T}^\perp}(Q_L), \mathcal{P}_{\mathcal{T}^\perp}(\Delta_L)\rangle \\
&\quad + \langle \rho\,\mathrm{sign}(\overline{S}) - \mathcal{P}_{\Gamma}(Q_S), \mathcal{P}_{\Gamma}(\Delta_S)\rangle + \langle \rho W_S - \mathcal{P}_{\Gamma^c}(Q_S), \mathcal{P}_{\Gamma^c}(\Delta_S)\rangle \\
\ge\ & -\big\|U_1V_1^T - \mathcal{P}_{\mathcal{T}}(Q_L)\big\|_F\|\mathcal{P}_{\mathcal{T}}(\Delta_L)\|_F + \big(1 - \|\mathcal{P}_{\mathcal{T}^\perp}(Q_L)\|\big)\|\mathcal{P}_{\mathcal{T}^\perp}(\Delta_L)\|_* \\
&\quad - \big\|\rho\,\mathrm{sign}(\overline{S}) - \mathcal{P}_{\Gamma}(Q_S)\big\|_F\|\mathcal{P}_{\Gamma}(\Delta_S)\|_F + \big(\rho - \|\mathcal{P}_{\Gamma^c}(Q_S)\|_\infty\big)\|\mathcal{P}_{\Gamma^c}(\Delta_S)\|_1 \\
\ge\ & -\frac{\rho}{4}\frac{\sqrt{\gamma_2}}{\gamma_1}\|\mathcal{P}_{\mathcal{T}}(\Delta_L)\|_F + \frac{1}{4}\|\mathcal{P}_{\mathcal{T}^\perp}(\Delta_L)\|_* + \frac{\rho}{4}\|\mathcal{P}_{\Gamma^c}(\Delta_S)\|_1, \qquad (4.17)
\end{aligned}$$

where the first inequality is due to (4.15), the first equality follows from (4.16), the second inequality uses the Hölder's inequality and (4.14), and the last inequality is a consequence of the conditions (b) and (c) and the fact that $\mathcal{P}_\Gamma(Q_S) = \rho\,\mathrm{sign}(\overline{S})$ and $\mathcal{P}_{\Gamma^c}(Q_S) = A + C$ with $\|A + C\|_\infty \le \frac{3}{4}\rho$.

Next, we need to show that $\|\mathcal{P}_{\mathcal{T}}(\Delta_L)\|_F$ cannot be too large. Recall that Ω is a multiset sampled from $\mathcal{F}^c$. Since $\mathcal{P}_{\mathcal{F}^c}(\Delta_S) = 0$ according to (4.13), it follows from (4.4) that $\mathcal{R}_\Omega(\Delta_S) = 0$. This, together with (4.13) and the fact that $\Gamma_0 \subseteq \mathcal{F}$, gives that $(\mathcal{P}_{\Gamma_0} + \frac{d_s}{m}\mathcal{R}_\Omega)(\Delta_L + \Delta_S) = 0$. Since $\Gamma^c = \Gamma_0 \cup \mathcal{F}^c$ and $\mathcal{P}_{\mathcal{F}^c}(\Delta_S) = \mathcal{R}_\Omega(\Delta_S) = 0$,

we have $\mathcal{P}_{\Gamma^c}(\Delta_S) = \mathcal{P}_{\Gamma_0}(\Delta_S) = (\mathcal{P}_{\Gamma_0} + \frac{d_s}{m}\mathcal{R}_\Omega)(\Delta_S)$ and

$$\begin{aligned}\|\mathcal{P}_{\Gamma^c}(\Delta_S)\|_1 \geq \|\mathcal{P}_{\Gamma^c}(\Delta_S)\|_F &= \left\|\left(\mathcal{P}_{\Gamma_0} + \frac{d_s}{m}\mathcal{R}_\Omega\right)(\Delta_S)\right\|_F = \left\|\left(\mathcal{P}_{\Gamma_0} + \frac{d_s}{m}\mathcal{R}_\Omega\right)(\Delta_L)\right\|_F \\ &\geq \left\|\left(\mathcal{P}_{\Gamma_0} + \frac{d_s}{m}\mathcal{R}_\Omega\right)\mathcal{P}_{\mathcal{T}}(\Delta_L)\right\|_F - \left\|\left(\mathcal{P}_{\Gamma_0} + \frac{d_s}{m}\mathcal{R}_\Omega\right)\mathcal{P}_{\mathcal{T}^\perp}(\Delta_L)\right\|_F. \quad (4.18)\end{aligned}$$

On the one hand, from the assumption on $\|\mathcal{P}_{\Gamma_0} + \frac{d_s}{m}\mathcal{R}_\Omega\|$, we get that

$$\left\|\left(\mathcal{P}_{\Gamma_0} + \frac{d_s}{m}\mathcal{R}_\Omega\right)\mathcal{P}_{\mathcal{T}^\perp}(\Delta_L)\right\|_F \leq \gamma_1\|\mathcal{P}_{\mathcal{T}^\perp}(\Delta_L)\|_F. \quad (4.19)$$

On the other hand, we notice that

$$\begin{aligned}&\left\|\left(\mathcal{P}_{\Gamma_0} + \frac{d_s}{m}\mathcal{R}_\Omega\right)\mathcal{P}_{\mathcal{T}}(\Delta_L)\right\|_F^2 = \left\langle \mathcal{P}_{\mathcal{T}}(\Delta_L), \left(\mathcal{P}_{\Gamma_0} + \frac{d_s^2}{m^2}\mathcal{R}_\Omega^2\right)\mathcal{P}_{\mathcal{T}}(\Delta_L)\right\rangle \\ &\geq \left\langle \mathcal{P}_{\mathcal{T}}(\Delta_L), \left(\mathcal{P}_{\Gamma_0} + \frac{d_s}{m}\mathcal{R}_\Omega\right)\mathcal{P}_{\mathcal{T}}(\Delta_L)\right\rangle \\ &= \left\langle \mathcal{P}_{\mathcal{T}}(\Delta_L), \mathcal{P}_{\mathcal{T}}(\Delta_L) + \left[\mathcal{P}_{\mathcal{T}}\left(\mathcal{P}_{\Gamma_0} + \frac{d_s}{m}\mathcal{R}_\Omega\right)\mathcal{P}_{\mathcal{T}} - \mathcal{P}_{\mathcal{T}}\right]\mathcal{P}_{\mathcal{T}}(\Delta_L)\right\rangle \\ &\geq \|\mathcal{P}_{\mathcal{T}}(\Delta_L)\|_F^2 - (1-\gamma_2)\|\mathcal{P}_{\mathcal{T}}(\Delta_L)\|_F^2 = \gamma_2\|\mathcal{P}_{\mathcal{T}}(\Delta_L)\|_F^2, \quad (4.20)\end{aligned}$$

where the first equality follows from the observation that $\mathrm{Range}(\mathcal{P}_{\Gamma_0}) \cap \mathrm{Range}(\mathcal{R}_\Omega) = \emptyset$, the first inequality is due to the fact that $\|\mathcal{R}_\Omega\| \geq 1$ and $d_s/m \geq 1$, and the last inequality is a consequence of the assumption on $\|\mathcal{P}_{\mathcal{T}}(\mathcal{P}_{\Gamma_0} + \frac{d_s}{m}\mathcal{R}_\Omega)\mathcal{P}_{\mathcal{T}} - \mathcal{P}_{\mathcal{T}}\|$. Then combining (4.18), (4.19) and (4.20) yields that

$$\|\mathcal{P}_{\Gamma^c}(\Delta_S)\|_1 \geq \sqrt{\gamma_2}\|\mathcal{P}_{\mathcal{T}}(\Delta_L)\|_F - \gamma_1\|\mathcal{P}_{\mathcal{T}^\perp}(\Delta_L)\|_F. \quad (4.21)$$

Finally, from (4.17) and (4.21), we obtain that

$$\begin{aligned}&\|\overline{L} + \Delta_L\|_* + \rho\|\overline{S} + \Delta_S\|_1 - \|\overline{L}\|_* - \rho\|\overline{S}\|_1 \\ &\geq -\frac{\rho}{4}\|\mathcal{P}_{\mathcal{T}^\perp}(\Delta_L)\|_F - \frac{\rho}{4\gamma_1}\|\mathcal{P}_{\Gamma^c}(\Delta_S)\|_1 + \frac{1}{4}\|\mathcal{P}_{\mathcal{T}^\perp}(\Delta_L)\|_* + \frac{\rho}{4}\|\mathcal{P}_{\Gamma^c}(\Delta_S)\|_1 \\ &\geq \left(\frac{1}{4} - \frac{\rho}{4}\right)\|\mathcal{P}_{\mathcal{T}^\perp}(\Delta_L)\|_* + \frac{\rho}{4}\left(1 - \frac{1}{\gamma_1}\right)\|\mathcal{P}_{\Gamma^c}(\Delta_S)\|_1,\end{aligned}$$

which is strictly positive unless $\mathcal{P}_{\mathcal{T}^\perp}(\Delta_L) = 0$ and $\mathcal{P}_{\Gamma^c}(\Delta_S) = 0$, provided that $\rho < 1$ and $\gamma_1 > 1$. Now assume that $\mathcal{P}_{\mathcal{T}^\perp}(\Delta_L) = \mathcal{P}_{\Gamma^c}(\Delta_S) = 0$, or equivalently that $\Delta_L \in \mathcal{T}$ and $\Delta_S \in \Gamma$. Since $\Gamma_0 = \Gamma^c \cap \mathcal{F}$ and $\mathcal{P}_{\mathcal{F}}(\Delta_L + \Delta_S) = 0$, it holds that $\mathcal{P}_{\Gamma_0}(\Delta_L) = \mathcal{P}_{\Gamma_0}(\Delta_S) = 0$ and thus $\big(\mathcal{P}_{\Gamma_0} + \frac{d_s}{m}\mathcal{R}_\Omega\big)(\Delta_L) = 0$. From the assumption on $\|\mathcal{P}_{\mathcal{T}}(\mathcal{P}_{\Gamma_0} + \frac{d_s}{m}\mathcal{R}_\Omega)\mathcal{P}_{\mathcal{T}} - \mathcal{P}_{\mathcal{T}}\|$, we know that the operator $\mathcal{P}_{\mathcal{T}}\big(\mathcal{P}_{\Gamma_0} + \frac{d_s}{m}\mathcal{R}_\Omega\big)\mathcal{P}_{\mathcal{T}}$ is invertible on $\mathcal{T}$. Therefore, it follows from $\Delta_L \in \mathcal{T}$ that $\Delta_L = 0$. This, together with (4.13), implies that $\Delta_S = 0$. In conclusion, $\|\overline{L} + \Delta_L\|_* + \rho\|\overline{S} + \Delta_S\|_1 > \|\overline{L}\|_* + \rho\|\overline{S}\|_1$ unless $\Delta_L = \Delta_S = 0$. This completes the proof. □

It is worth mentioning that Proposition 4.9 could be regarded as a variation of the first-order sufficient optimality conditions for problem (4.6), which, based on the subdifferential of the nuclear norm at $\overline{L}$ and the subdifferential of the ℓ_1-norm at $\overline{S}$, require that the restriction of the operator $\mathcal{P}_{\mathcal{T}}\big(\mathcal{P}_{\Gamma_0} + \frac{d_s}{m}\mathcal{R}_\Omega\big)\mathcal{P}_{\mathcal{T}}$ to $\mathcal{T}$ is invertible and that there exist dual matrices A, B and $C \in \mathbb{V}^{n_1 \times n_2}$ obeying

$$\begin{cases} A \in \mathrm{Range}(\mathcal{P}_{\Gamma_0}),\ B \in \mathrm{Range}(\mathcal{R}_\Omega),\ C \in \mathrm{Range}(\mathcal{P}_{\mathcal{F}^c}), \\ \mathcal{P}_{\mathcal{T}}\big(\rho\,\mathrm{sign}(\overline{S}) + A + B\big) = U_1 U_1^T, \\ \big\|\mathcal{P}_{\mathcal{T}^\perp}\big(\rho\,\mathrm{sign}(\overline{S}) + A + B\big)\big\| < 1, \\ \|A\|_\infty < \rho,\ \|C\|_\infty < \rho. \end{cases}$$

or equivalently that there exist dual matrices A and $B \in \mathbb{V}^{n_1 \times n_2}$ satisfying

$$\begin{cases} A \in \mathrm{Range}(\mathcal{P}_{\Gamma_0}) \text{ with } \|A\|_\infty < \rho,\ B \in \mathrm{Range}(\mathcal{R}_\Omega), \\ \mathcal{P}_{\mathcal{T}}\big(\rho\,\mathrm{sign}(\overline{S}) + A + B\big) = U_1 U_1^T, \\ \big\|\mathcal{P}_{\mathcal{T}^\perp}\big(\rho\,\mathrm{sign}(\overline{S}) + A + B\big)\big\| < 1. \end{cases}$$

Correspondingly, the next proposition provides an analogous variation of the first-order sufficient optimality conditions for problem (4.7).

Proposition 4.10. *Suppose that the assumptions in Proposition 4.9 hold. Then under Assumption 4.1, $(\overline{L}, \overline{S})$ is the unique optimal solution to problem (4.7) if there exist dual certificates A and $B \in \mathcal{S}^n$ such that*

(a) $A \in \mathrm{Range}(\mathcal{P}_{\Gamma_0})$ *with* $\|A\|_\infty \leq \dfrac{3}{4}\rho$, *and* $B \in \mathrm{Range}(\mathcal{R}_\Omega)$,

(b) $\big\|U_1U_1^T - \mathcal{P}_{\mathcal{T}}\big(\rho\,\mathrm{sign}(\overline{S}) + A + B\big)\big\|_F \leq \dfrac{\rho}{4}\dfrac{\sqrt{\gamma_2}}{\gamma_1}$,

(c) $\big\|\mathcal{P}_{\mathcal{T}^\perp}\big(\rho\,\mathrm{sign}(\overline{S}) + A + B\big)\big\| \leq \dfrac{3}{4}$.

Proof. For any feasible solution (L, S) to problem (4.7), let $(\Delta_L, \Delta_S) := (L - \overline{L}, S - \overline{S})$. Notice that (4.13) holds for such (Δ_L, Δ_S). Since $\overline{L} \in \mathcal{S}_+^n$, the reduced SVD (4.8) can be rewritten as $\overline{L} = U_1\Sigma U_1^T$ with $U = V$, where $U = [U_1, U_2]$ and $V = [V_1, V_2]$ are orthogonal matrices. Then $L \in \mathcal{S}_+^n$ implies that $U_2^T L U_2 = U_2^T\overline{L}U_2 + U_2^T\Delta_L U_2 = U_2^T\Delta_L U_2 \in \mathcal{S}_+^{n-r}$. By taking $W_L = U_2U_2^T$ and $W_S = \mathrm{sign}(\mathcal{P}_{\Gamma^c}(\Delta_S))$, we can easily check that (4.14) holds. Thus, the proof can be obtained in a similar way to that of Proposition 4.9. We omit it here. □

From the similarity between Proposition 4.9 and Proposition 4.10, we can see that as long as the proof of Theorem 4.3 or Theorem 4.4 is established, then the other proof will follow in the same way. Therefore, for the sake of simplicity, we will only focus on the proof of Theorem 4.3 by constructing the dual certificates for problem (4.6) based on Proposition 4.9 in the following discussion. Below we state a useful remark for Proposition 4.9.

Remark 4.1. *According to Proposition 4.5 and Proposition 4.6, for any* $c > 0$, *if* $m \geq \frac{128}{3\sqrt{2}}(1+c)\max\{\mu_0 r\frac{(n_1+n_2)}{n_1n_2}d_s, 1\}\log(2n_1n_2)$ *and* $\sqrt{\frac{\mu_0 rk}{n_1}} + \sqrt{\frac{\mu_0 rk}{n_2}} \leq \frac{1}{16}$, *then with probability at least* $1 - 2(2n_1n_2)^{-c}$, *we have*

$$\left\|\mathcal{P}_{\Gamma_0} + \frac{d_s}{m}\mathcal{R}_\Omega\right\| \leq \frac{\sqrt{2}n_1n_2d_s}{8\max\{r(n_1+n_2)d_s,\, n_1n_2\}} \tag{4.22}$$

and

$$\left\|\mathcal{P}_{\mathcal{T}}\left(\mathcal{P}_{\Gamma_0} + \frac{d_s}{m}\mathcal{R}_\Omega\right)\mathcal{P}_{\mathcal{T}} - \mathcal{P}_{\mathcal{T}}\right\| \leq \frac{\sqrt{\sqrt{2}}}{4} + \frac{1}{16} < \frac{1}{2}.$$

Therefore, the condition (b) of Proposition 4.9 can be replaced by

$$\big\|U_1V_1^T - \mathcal{P}_{\mathcal{T}}\big(\rho\,\mathrm{sign}(\overline{S}) + A + B\big)\big\|_F \leq \rho\frac{\max\{r(n_1+n_2)d_s,\, n_1n_2\}}{n_1n_2d_s}. \tag{4.23}$$

Construction of the dual certificates

The second step of the proof of Theorem 4.3 is to demonstrate the existence of the dual certificates A and B for problem (4.6) that satisfy the conditions listed in Proposition 4.9. To achieve this goal, we apply the so-called golfing scheme, an elegant and powerful technique first designed in [69] and later used in [104, 68, 21, 89, 33], to construct such dual certificates. Mathematically, the golfing scheme could be viewed as a "correct and sample" recursive procedure such that the desired error decreases exponentially fast (cf. [68, 33]).

Next, we introduce the golfing scheme in details. Let the sampled multiset Ω be decomposed into p partitions of size q, where the multiset corresponding to the j-th partition is denoted by Ω_j. Then the sample size $m = pq$. Notice that in the model of sampling with replacement, these partitions are independent of each other. Set the matrix $Y_0 := 0$ and define the matrix Y_j recursively as follows

$$Y_j := Y_{j-1} + \left(\mathcal{P}_{\Gamma_0} + \frac{d_s}{q}\mathcal{R}_{\Omega_j}\right)\left[U_1V_1^T - \rho\mathcal{P}_{\mathcal{T}}\big(\operatorname{sign}(\overline{S})\big) - \mathcal{P}_{\mathcal{T}}(Y_{j-1})\right],$$

where $j = 1, \ldots, p$. Let the error in the j-th step be defined by

$$E_j := U_1V_1^T - \rho\mathcal{P}_{\mathcal{T}}\big(\operatorname{sign}(\overline{S})\big) - \mathcal{P}_{\mathcal{T}}(Y_j), \quad \text{for } j = 0, \ldots, p.$$

Then E_j, for $j = 1, \ldots, p$, takes the recursive form of

$$\begin{aligned} E_j &= \left[\mathcal{P}_{\mathcal{T}} - \mathcal{P}_{\mathcal{T}}\left(\mathcal{P}_{\Gamma_0} + \frac{d_s}{q}\mathcal{R}_{\Omega_j}\right)\mathcal{P}_{\mathcal{T}}\right]\left[U_1V_1^T - \rho\mathcal{P}_{\mathcal{T}}\big(\operatorname{sign}(\overline{S})\big) - \mathcal{P}_{\mathcal{T}}(Y_{j-1})\right] \\ &= \left[\mathcal{P}_{\mathcal{T}} - \mathcal{P}_{\mathcal{T}}\left(\mathcal{P}_{\Gamma_0} + \frac{d_s}{q}\mathcal{R}_{\Omega_j}\right)\mathcal{P}_{\mathcal{T}}\right](E_{j-1}), \end{aligned} \tag{4.24}$$

and Y_p can be represented as

$$Y_p = \sum_{j=1}^{p}\left(\mathcal{P}_{\Gamma_0} + \frac{d_s}{q}\mathcal{R}_{\Omega_j}\right)(E_{j-1}) = A_p + B_p, \tag{4.25}$$

where A_p and B_p are the dual certificates constructed by

$$A_p := \sum_{j=1}^{p}\mathcal{P}_{\Gamma_0}(E_{j-1}) \quad \text{and} \quad B_p := \frac{d_s}{q}\sum_{j=1}^{p}\mathcal{R}_{\Omega_j}(E_{j-1}). \tag{4.26}$$

It can be immediately seen that $A_p \in \mathrm{Range}(\mathcal{P}_{\Gamma_0})$ and $B_p \in \mathrm{Range}(\mathcal{R}_\Omega)$.

Verification of the dual certificates

As a consequence of Remark 4.1, it suffices to verify that for problem (4.6), the dual certificates A_p and B_p constructed in (4.26) satisfy the inequality (4.23) as well as the conditions (a) and (c) of Proposition 4.9.

First of all, we recall the assumptions below. Suppose that Assumption 4.2 and Assumption 4.3 hold with

$$\sqrt{\frac{\mu_0 rk}{n_1}} + \sqrt{\frac{\mu_0 rk}{n_2}} + \frac{\mu_0 rk}{\sqrt{n_1 n_2}} \leq \frac{1}{16}. \tag{4.27}$$

Since $1 \leq \mu_1 \leq \mu_0\sqrt{r}$, the assumption (4.27) implies that

$$k \leq \frac{1}{16}\sqrt{\frac{n_1 n_2}{\mu_1^2 r}}. \tag{4.28}$$

Moreover, the tradeoff parameter is chosen as follows

$$\frac{48}{13}\sqrt{\frac{\mu_1^2 r}{n_1 n_2}} \leq \rho \leq 4\sqrt{\frac{\mu_1^2 r}{n_1 n_2}}. \tag{4.29}$$

Then, we state the following probabilistic inequalities related to the random sampling operator. Note that these inequalities have already been prepared in the previous subsection. Let $c > 0$ be an arbitrarily given absolute constant. If the size of each partition satisfies

$$q \geq q_1 := \frac{128}{3\sqrt{2}}(1+c)\max\left\{\mu_0 r\frac{(n_1+n_2)}{n_1 n_2}d_s,\, 1\right\}\log(2n_1 n_2),$$

with probability at least $1-(2n_1n_2)^{-c}$, it follows from Proposition 4.6 and (4.27) that

$$\left\|\mathcal{P}_{\mathcal{T}}\left(\mathcal{P}_{\Gamma_0} + \frac{d_s}{q}\mathcal{R}_{\Omega_j}\right)\mathcal{P}_{\mathcal{T}} - \mathcal{P}_{\mathcal{T}}\right\| \leq \frac{\sqrt{\sqrt{2}}}{4} + \frac{1}{16} < \frac{1}{2}, \tag{4.30}$$

for all $j = 1, \ldots, p$. In addition, it can be seen from (4.24) that $E_{j-1} \in \mathcal{T}$ and E_{j-1} is independent of Ω_j, for all $j = 1, \ldots, p$. Then we know that Proposition 4.7 and

Proposition 4.8 are both applicable to E_{j-1} and Ω_j. On the one hand, according to Proposition 4.7 and (4.27), when the size of each partition satisfies

$$q \geq q_2 := \frac{256}{3\sqrt{2}}(1+c)\mu_0 r \frac{(n_1+n_2)}{n_1 n_2} d_s \log(n_1 n_2),$$

with probability at least $1-2(n_1n_2)^{-c}$, it holds that

$$\begin{aligned}\left\| \mathcal{P}_{\mathcal{T}}\left(\mathcal{P}_{\Gamma_0} + \frac{d_s}{q}\mathcal{R}_{\Omega_j}\right)(E_{j-1}) - E_{j-1}\right\|_\infty &\leq \left(\frac{\sqrt{\sqrt{2}}}{4} + \frac{1}{16}\right)\|E_{j-1}\|_\infty \\ &< \frac{1}{2}\|E_{j-1}\|_\infty. \end{aligned} \tag{4.31}$$

On the other hand, provided that the size of each partition satisfies

$$\begin{aligned} q \geq q_3' &:= \frac{512}{3}(1+c)\mu_1^2 r \frac{(\sqrt{n_1n_2}+d_s)^2 \max\{(n_1+n_2)d_s,\, n_1n_2\}}{\max\{\sqrt{n_1n_2}d_s,\, n_1n_2\}^2}\log(n_1+n_2) \\ &\geq \frac{8}{3}(1+c)\frac{(\sqrt{n_1n_2}+d_s)^2}{\max\{(n_1+n_2)d_s,\, n_1n_2\}}\log(n_1+n_2), \end{aligned}$$

by applying Proposition 4.8, we obtain that with probability at least $1-(n_1+n_2)^{-c}$,

$$\begin{aligned}\left\|\left(\mathcal{P}_{\Gamma_0} + \frac{d_s}{q}\mathcal{R}_{\Omega_j} - \mathcal{I}\right)(E_{j-1})\right\| &\leq \left(\frac{1}{8\sqrt{\mu_1^2 r}}\frac{\max\{\sqrt{n_1n_2}d_s,\, n_1n_2\}}{\sqrt{n_1n_2}+d_s} + k\right)\|E_{j-1}\|_\infty \\ &\leq \left(\frac{1}{8}\sqrt{\frac{n_1n_2}{\mu_1^2 r}} + k\right)\|E_{j-1}\|_\infty. \end{aligned} \tag{4.32}$$

Since $\sqrt{n_1n_2}+d_s \leq 2\max\{\sqrt{n_1n_2}, d_s\}$, we further have

$$q_3 := \frac{2048}{3}(1+c)\mu_1^2 r \max\left\{\frac{(n_1+n_2)}{n_1n_2}d_s,\, 1\right\}\log(n_1+n_2) \geq q_3'.$$

Before proceeding to the verification, we introduce the following notation. That is, for any $j' = 1, \ldots, p$, denote

$$\begin{aligned} &\prod_{j=1}^{j'}\left[\mathcal{P}_{\mathcal{T}} - \mathcal{P}_{\mathcal{T}}\left(\mathcal{P}_{\Gamma_0} + \frac{d_s}{q}\mathcal{R}_{\Omega_j}\right)\mathcal{P}_{\mathcal{T}}\right] \\ :=& \left[\mathcal{P}_{\mathcal{T}} - \mathcal{P}_{\mathcal{T}}\left(\mathcal{P}_{\Gamma_0} + \frac{d_s}{q}\mathcal{R}_{\Omega_{j'}}\right)\mathcal{P}_{\mathcal{T}}\right]\cdots\left[\mathcal{P}_{\mathcal{T}} - \mathcal{P}_{\mathcal{T}}\left(\mathcal{P}_{\Gamma_0} + \frac{d_s}{q}\mathcal{R}_{\Omega_1}\right)\mathcal{P}_{\mathcal{T}}\right], \end{aligned}$$

where the order of the above multiplications is important because of the non-commutativity of these operators.

For the inequality (4.23): Observe that $E_0 = U_1V_1^T - \rho\mathcal{P}_{\mathcal{T}}(\operatorname{sign}(\overline{S}))$. Since $\operatorname{rank}(\overline{L}) = r$, we have $\|U_1V_1^T\|_F = \sqrt{r}$. Moreover, it follows from the non-expansivity of the metric projection $\mathcal{P}_{\mathcal{T}}$ and Assumption 4.3 that $\|\mathcal{P}_{\mathcal{T}}(\operatorname{sign}(\overline{S}))\|_F \leq \|\operatorname{sign}(\overline{S})\|_F \leq k$. Hence, we know that

$$\big\|E_0\big\|_F \leq \big\|U_1V_1^T\big\|_F + \rho\big\|\mathcal{P}_{\mathcal{T}}\big(\operatorname{sign}(\overline{S})\big)\big\|_F \leq \sqrt{r} + \rho k. \tag{4.33}$$

Due to the golfing scheme, we have the following exponential convergence

$$\begin{aligned}
&\big\|U_1V_1^T - \mathcal{P}_{\mathcal{T}}\big(\rho\operatorname{sign}(\overline{S}) + A_p + B_p\big)\big\|_F \\
&= \big\|E_p\big\|_F = \left\|\left\{\prod_{j=1}^{p}\left[\mathcal{P}_{\mathcal{T}} - \mathcal{P}_{\mathcal{T}}\left(\mathcal{P}_{\Gamma_0} + \frac{d_s}{q}\mathcal{R}_{\Omega_j}\right)\mathcal{P}_{\mathcal{T}}\right]\right\}(E_0)\right\|_F \\
&\leq \left[\prod_{j=1}^{p}\left\|\mathcal{P}_{\mathcal{T}} - \mathcal{P}_{\mathcal{T}}\left(\mathcal{P}_{\Gamma_0} + \frac{d_s}{q}\mathcal{R}_{\Omega_j}\right)\mathcal{P}_{\mathcal{T}}\right\|\right]\big\|E_0\big\|_F \leq 2^{-p}(\sqrt{r} + \rho k) < 2^{-p}(\sqrt{r} + 1),
\end{aligned}$$

where the first equality is from the definition of E_p, the second equality is from (4.24), the second inequality is from (4.30) and (4.33), and the last inequality is from (4.28) and (4.29). By taking

$$p := \frac{3}{2}\log(2n_1n_2) > \log_2(2n_1n_2),$$

and using (4.29) (together with the fact that $\mu_1 \geq 1$) and the inequality of arithmetic and geometric means, we have

$$2^{-p}(\sqrt{r} + 1) < \frac{2\sqrt{r}}{2n_1n_2} < \rho\frac{2r}{\sqrt{n_1n_2}} \leq \rho\frac{\max\{r(n_1+n_2)d_s,\, n_1n_2\}}{n_1n_2d_s}.$$

This verifies the inequality (4.23).

For the condition (a): Notice that $\|U_1V_1^T\|_\infty \leq \sqrt{\frac{\mu_1^2 r}{n_1n_2}}$ under Assumption 4.2 and that $\|\mathcal{P}_{\mathcal{T}}(\operatorname{sign}(\overline{S}))\|_\infty = \|\mathcal{P}_{\mathcal{T}}\mathcal{P}_{\Gamma}(\operatorname{sign}(\overline{S}))\|_\infty \leq \sqrt{\frac{\mu_0 rk}{n_1}} + \sqrt{\frac{\mu_0 rk}{n_2}} + \frac{\mu_0 rk}{\sqrt{n_1n_2}}$

according to Lemma 4.2. Since $E_0 = U_1 V_1^T - \rho \mathcal{P}_{\mathcal{T}}(\operatorname{sign}(\overline{S}))$, we get that

$$\begin{aligned}\big\|E_0\big\|_\infty &\le \big\|U_1 V_1^T\big\|_\infty + \rho \big\|\mathcal{P}_{\mathcal{T}}\big(\operatorname{sign}(\overline{S})\big)\big\|_\infty \\ &\le \sqrt{\frac{\mu_1^2 r}{n_1 n_2}} + \rho\bigg(\sqrt{\frac{\mu_0 r k}{n_1}} + \sqrt{\frac{\mu_0 r k}{n_2}} + \frac{\mu_0 r k}{\sqrt{n_1 n_2}}\bigg) \le \frac{1}{3}\rho, \end{aligned} \tag{4.34}$$

where the last inequality is from (4.27) and (4.29). According to the golfing scheme, we derive that

$$\begin{aligned}\|A_p\|_\infty \le \sum_{j=1}^{p} \big\|\mathcal{P}_{\Gamma_0}(E_{j-1})\big\|_\infty &\le \sum_{j=1}^{p} \big\|E_{j-1}\big\|_\infty \\ &= \sum_{j=1}^{p} \bigg\| \bigg\{ \prod_{i=1}^{j-1} \bigg[\mathcal{P}_{\mathcal{T}} - \mathcal{P}_{\mathcal{T}} \bigg(\mathcal{P}_{\Gamma_0} + \frac{d_s}{q} \mathcal{R}_{\Omega_i} \bigg) \mathcal{P}_{\mathcal{T}} \bigg] \bigg\} (E_0) \bigg\|_\infty \\ &\le \sum_{j=1}^{p} 2^{-(j-1)} \big\|E_0\big\|_\infty < \frac{2}{3}\rho < \frac{3}{4}\rho, \end{aligned} \tag{4.35}$$

where the first inequality is from (4.26), the first equality is from (4.24), the third inequality is from (4.31) and the fact that $E_{j-1} \in \mathcal{T}$ for all $j = 1, \dots, p$, and the fourth inequality is from (4.34). This verifies the condition (a).

For the condition (c): Firstly, as a consequence of (2.6), Assumption 4.3, Lemma 2.2, (4.28) and (4.29), it holds that

$$\big\|\mathcal{P}_{\mathcal{T}^\perp}\big(\rho \operatorname{sign}(\overline{S})\big)\big\| \le \rho \big\|\operatorname{sign}(\overline{S})\big\| \le \rho k \big\|\operatorname{sign}(\overline{S})\big\|_\infty \le \frac{1}{4}.$$

Secondly, by applying the golfing scheme, we obtain that

$$\begin{aligned}\big\|\mathcal{P}_{\mathcal{T}^\perp}\big(A_p + B_p\big)\big\| &\le \sum_{j=1}^{p} \bigg\| \mathcal{P}_{\mathcal{T}^\perp} \bigg(\mathcal{P}_{\Gamma_0} + \frac{d_s}{q} \mathcal{R}_{\Omega_j} \bigg) (E_{j-1}) \bigg\| \\ &= \sum_{j=1}^{p} \bigg\| \mathcal{P}_{\mathcal{T}^\perp} \bigg(\mathcal{P}_{\Gamma_0} + \frac{d_s}{q} \mathcal{R}_{\Omega_j} - \mathcal{I} \bigg) (E_{j-1}) \bigg\| \\ &\le \sum_{j=1}^{p} \bigg\| \bigg(\mathcal{P}_{\Gamma_0} + \frac{d_s}{q} \mathcal{R}_{\Omega_j} - \mathcal{I} \bigg) (E_{j-1}) \bigg\| \\ &\le \sum_{j=1}^{p} \bigg(\frac{1}{8} \sqrt{\frac{n_1 n_2}{\mu_1^2 r}} + k \bigg) \|E_{j-1}\|_\infty \le \bigg(\frac{1}{8} \sqrt{\frac{n_1 n_2}{\mu_1^2 r}} + k \bigg) \frac{2}{3}\rho \le \frac{1}{2}, \end{aligned}$$

where the first inequality is from (4.25) and the triangular inequality, the first equality is from the fact that $E_{j-1} \in \mathcal{T}$ for all $j = 1, \ldots, p$, the second inequality is from (2.6), the third inequality is from (4.32), the fourth inequality is from (4.35), and the fifth inequality is from (4.28) and (4.29). Then we have

$$\left\|\mathcal{P}_{\mathcal{T}^\perp}\left(\rho\,\mathrm{sign}(\overline{S}) + A_p + B_p\right)\right\| \leq \left\|\mathcal{P}_{\mathcal{T}^\perp}\left(\rho\,\mathrm{sign}(\overline{S})\right)\right\| + \left\|\mathcal{P}_{\mathcal{T}^\perp}\left(A_p + B_p\right)\right\| \leq \frac{3}{4},$$

which verifies the condition (c).

In conclusion, for any absolute constant $c > 0$, if the total sample size m is large enough such that

$$m \geq 1024(1+c)\max\{\mu_1^2, \mu_0\} r \max\left\{\frac{(n_1+n_2)}{n_1 n_2} d_s,\, 1\right\} \log^2(2n_1n_2),$$

then it follows that $m \geq p\max\{q_1, q_2, q_3\}$, and thus all of the inequalities (4.22), (4.30), (4.31) and (4.32) hold with probability at least

$$1 - (2n_1n_2)^{-c} - \frac{3}{2}\log(2n_1n_2)\left[(2n_1n_2)^{-c} + 2(n_1n_2)^{-c} + (n_1+n_2)^{-c}\right],$$

by the union bound. This completes the proof of Theorem 4.3. Due to the similarity between Proposition 4.9 and Proposition 4.10, the proof of Theorem 4.4 can be obtained in the same way.

Chapter 5

Noisy matrix decomposition from fixed and sampled basis coefficients

In this chapter, we focus on the problem of noisy low-rank and sparse matrix decomposition with fixed and sampled basis coefficients. We first introduce some problem background mainly on the observation model, and then propose a two-stage rank-sparsity-correction procedure via convex optimization, which is inspired by the successful recent development on the adaptive nuclear semi-norm penalization technique. By exploiting the notion of restricted strong convexity, a novel non-asymptotic probabilistic error bound under the high-dimensional scaling is established to examine the recovery performance of the proposed procedure.

5.1 Problem background and formulation

In this section, we present the model of the problem of noisy low-rank and sparse matrix decomposition with fixed and sampled basis coefficients, and formulate this problem into convex programs by applying the adaptive nuclear semi-norm penalization technique developed in [98, 99] and the adaptive ℓ_1 semi-norm penalization

technique used in (3.5).

Suppose that we want to estimate an unknown matrix $\overline{X} \in \mathbb{V}^{n_1 \times n_2}$ of low-dimensional structure in the sense that it is equal to the sum of an unknown low-rank matrix $\overline{L} \in \mathbb{V}^{n_1 \times n_2}$ and an unknown sparse matrix $\overline{S} \in \mathbb{V}^{n_1 \times n_2}$. As motivated by the high-dimensional correlation matrix estimation problem coming from a strict or approximate factor model used in economic and financial studies (see, e.g., [3, 50, 8, 53, 6, 34, 54, 90, 7]), we further assume that some basis coefficients of the unknown matrix $\overline{X}$ are fixed. Throughout this chapter, for the unknown matrix $\overline{X}$, we let $\mathcal{F} \subseteq \{1, \ldots, d\}$ denote the fixed index set corresponding to the fixed basis coefficients and $\mathcal{F}^c = \{1, \ldots, d\} \setminus \mathcal{F}$ denote the unfixed index set associated with the unfixed basis coefficients, respectively. We define $d_s := |\mathcal{F}^c|$.

5.1.1 Observation model

When the fixed basis coefficients are too few to draw any meaningful statistical inference, we need to observe some of the rest for accurately estimating the unknown matrix $\overline{X}$ as well as the low-rank component $\overline{L}$ and the sparse component $\overline{S}$.

We now describe the noisy observation model under a general weighted scheme for non-uniform sampling with replacement that we consider in this chapter. Recall that $\Theta = \{\Theta_1, \ldots, \Theta_d\}$ represents the set of the standard orthonormal basis of the finite dimensional real Euclidean space $\mathbb{V}^{n_1 \times n_2}$. In detail, Θ is given by (4.1) with $d = n_1 n_2$ when $\mathbb{V}^{n_1 \times n_2} = \mathbb{R}^{n_1 \times n_2}$, and Θ is given by (4.2) with $d = n(n+1)/2$ when $\mathbb{V}^{n_1 \times n_2} = \mathcal{S}^n$. Suppose that we are given a collection of m noisy observations $\{(\omega_l, y_l)\}_{l=1}^m$ of the basis coefficients of the unknown matrix $\overline{X}$ with respect to the unfixed basis $\{\Theta_j \,|\, j \in \mathcal{F}^c\}$ of the following form

$$y_l = \langle \Theta_{\omega_l}, \overline{X} \rangle + \nu \xi_l, \quad l = 1, \ldots, m, \tag{5.1}$$

where $\Omega := \{\omega_l\}_{l=1}^m$ is the multiset of indices sampled with replacement[1] from the unfixed index set $\mathcal{F}^c$, $\{\xi_l\}_{l=1}^m$ are i.i.d. additive noises with $\mathbb{E}[\xi_l] = 0$ and $\mathbb{E}[\xi_l^2] = 1$, and $\nu > 0$ is the noise magnitude. For notational simplicity, we define the sampling operator $\mathcal{R}_\Omega : \mathbb{V}^{n_1 \times n_2} \to \mathbb{R}^m$ associated with the multiset Ω by

$$\mathcal{R}_\Omega(Z) := (\langle \Theta_{\omega_1}, Z\rangle, \ldots, \langle \Theta_{\omega_m}, Z\rangle)^T, \quad Z \in \mathbb{V}^{n_1 \times n_2}. \tag{5.2}$$

Then the observation model (5.1) can be rewritten in the following vector form

$$y = \mathcal{R}_\Omega(\overline{X}) + \nu\xi, \tag{5.3}$$

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

Suppose further that the elements of the index set Ω are i.i.d. copies of a random variable ω with the probability distribution Π over $\mathcal{F}^c$ being defined by $\mathbb{P}[\omega = j] := \pi_j > 0$ for all $j \in \mathcal{F}^c$. In particular, this general weighted sampling scheme is called uniform if Π is a uniform distribution, i.e., $\pi_j = 1/d_s$ for all $j \in \mathcal{F}^c$. Let the linear operator $\mathcal{Q}_{\mathcal{F}^c} : \mathbb{V}^{n_1 \times n_2} \to \mathbb{V}^{n_1 \times n_2}$ associated with the unfixed index set $\mathcal{F}^c$ and the sampling probability distribution Π be defined by

$$\mathcal{Q}_{\mathcal{F}^c}(Z) := \sum_{j \in \mathcal{F}^c} \pi_j \langle \Theta_j, Z\rangle \Theta_j, \quad Z \in \mathbb{V}^{n_1 \times n_2}. \tag{5.4}$$

Notice that $\mathcal{Q}_{\mathcal{F}^c}$ is self-adjoint and $\|\mathcal{Q}_{\mathcal{F}^c}\| = \max_{j \in \mathcal{F}^c} \pi_j$.

In addition, for any index subset $\mathcal{J} \subseteq \{1, \ldots, d\}$, we define two linear operators $\mathcal{R}_\mathcal{J} : \mathbb{V}^{n_1 \times n_2} \to \mathbb{R}^{|\mathcal{J}|}$ and $\mathcal{P}_\mathcal{J} : \mathbb{V}^{n_1 \times n_2} \to \mathbb{V}^{n_1 \times n_2}$ by

$$\mathcal{R}_\mathcal{J}(Z) := (\langle \Theta_j, Z\rangle)_{j \in \mathcal{J}}^T, \quad \text{and} \quad \mathcal{P}_\mathcal{J}(Z) := \sum_{j \in \mathcal{J}} \langle \Theta_j, Z\rangle \Theta_j, \quad Z \in \mathbb{V}^{n_1 \times n_2}.$$

[1] One may refer to Section 2.3 for more details on random sampling model.

5.1.2 Convex optimization formulation

The proposed convex formulation for the problem of noisy low-rank and sparse matrix decomposition with fixed and sampled basis coefficients is inspired by the successful recent development on the adaptive nuclear semi-norm penalization technique for noisy low-rank matrix completion [98, 99].

Let $(\mathring{L}, \mathring{S})$ be a pair of initial estimators of the true low-rank and sparse components $(\overline{L}, \overline{S})$. For instance, $(\mathring{L}, \mathring{S})$ can be obtained from the nuclear and ℓ_1 norms penalized least squares (NLPLS) problem

$$\begin{array}{ll} \min\limits_{L,\, S \in \mathbb{V}^{n_1 \times n_2}} & \dfrac{1}{2m}\|y - \mathcal{R}_\Omega(L+S)\|_2^2 + \rho_{L_0}\|L\|_* + \rho_{S_0}\|S\|_1 \\ \text{s.t.} & \mathcal{R}_{\mathcal{F}}(L+S) = \mathcal{R}_{\mathcal{F}}(\overline{X}),\ \|L\|_\infty \le b_L,\ \|S\|_\infty \le b_S, \end{array} \tag{5.5}$$

where $\rho_{L_0} \ge 0$ and $\rho_{S_0} \ge 0$ are the penalization parameters that control the rank of the low-rank component and the sparsity level of the sparse component, and the upper bounds $b_L > 0$ and $b_S > 0$ are two priori estimates of the entry-wise magnitude of the true low-rank and sparse components. We then aim to estimate $(\overline{L}, \overline{S})$ by solving the following convex optimization problem

$$\begin{array}{ll} \min\limits_{L,\, S \in \mathbb{V}^{n_1 \times n_2}} & \dfrac{1}{2m}\|y - \mathcal{R}_\Omega(L+S)\|_2^2 + \rho_L\big(\|L\|_* - \langle F(\mathring{L}), L\rangle\big) + \rho_S\big(\|S\|_1 - \langle G(\mathring{S}), S\rangle\big) \\ \text{s.t.} & \mathcal{R}_{\mathcal{F}}(L+S) = \mathcal{R}_{\mathcal{F}}(\overline{X}),\ \|L\|_\infty \le b_L,\ \|S\|_\infty \le b_S, \end{array} \tag{5.6}$$

where $\rho_L \ge 0$ and $\rho_S \ge 0$ are the penalization parameters that play the same roles as ρ_{L_0} and ρ_{S_0}, $F : \mathbb{V}^{n_1 \times n_2} \to \mathbb{V}^{n_1 \times n_2}$ is a spectral operator associated with a symmetric function $f : \mathbb{R}^n \to \mathbb{R}^n$, and $G : \mathbb{V}^{n_1 \times n_2} \to \mathbb{V}^{n_1 \times n_2}$ is an operator generated from a symmetric function $g : \mathbb{R}^d \to \mathbb{R}^d$. The detailed constructions of the operators F and G as well as the related functions f and g are deferred to Section 5.3. If, in addition, the true low-rank and sparse components $(\overline{L}, \overline{S})$ are known to be symmetric and positive semidefinite (e.g., $\overline{X}$ is a covariance or correlation matrix with factor structure), we consider to solve the following convex

conic optimization problem

$$\begin{aligned} \min \quad & \frac{1}{2m}\|y - \mathcal{R}_\Omega(L+S)\|_2^2 + \rho_L\langle I_n - F(\mathring{L}), L\rangle + \rho_S\big(\|S\|_1 - \langle G(\mathring{S}), S\rangle\big) \\ \text{s.t.} \quad & \mathcal{R}_\mathcal{F}(L+S) = \mathcal{R}_\mathcal{F}(\overline{X}),\ \|L\|_\infty \le b_L,\ \|S\|_\infty \le b_S,\ L \in \mathcal{S}_+^n,\ S \in \mathcal{S}_+^n. \end{aligned} \tag{5.7}$$

Intuitively, the bound constraints that control the "spikiness" of the low-rank and sparse components serve as a noisy version of identifiability conditions[2] in noisy observation models, where approximate recovery is the best possible outcome that can be expected. In fact, this type of spikiness control has recently been shown to be critical in the analysis of the nuclear norm penalization approach for noisy matrix completion [101, 79] and noisy matrix decomposition [1]. In some practical applications, such entry-wise bounds are often available from the assumed structure or prior estimation. For instance, when the true unknown matrix $\overline{X}$ is a correlation matrix originating from a factor model, both of the prescribed bounds b_L and b_S can be at most set to 1.

As coined by [98, 99], the spectral operator F is called the rank-correction function and the linear term $-\langle F(\mathring{L}), L\rangle$ is called the rank-correction term. Accordingly, we call the operator G the sparsity-correction function and the linear term $-\langle G(\mathring{S}), S\rangle$ the sparsity-correction term. Hence, problem (5.6) or (5.7) is referred to as the rank-sparsity-correction step. Moreover, if F and G are chosen such that $\|L\|_* - \langle F(\mathring{L}), L\rangle\big)$ and $\|S\|_1 - \langle G(\mathring{S}), S\rangle$ are both semi-norms, we call the solution $(\widehat{L}, \widehat{S})$ to problem (5.6) or (5.7) the adaptive nuclear and ℓ_1 semi-norms penalized least squares (ANLPLS) estimator. Obviously, the ANLPLS estimator (5.6) includes the NLPLS estimator (5.5) as a special case when $F \equiv 0$ and $G \equiv 0$. With the NLPLS estimator being selected as the initial estimator $(\mathring{L}, \mathring{S})$, it is plausible that the ANLPLS estimator obtained from this two-stage procedure may produce a better recovery performance as long as the correction functions F and G are constructed suitably. In the next section, we derive a recovery error

[2] A noiseless version of identifiability conditions is introduced in Section 4.2.

bound for the ANLPLS estimator, which provides some important guidelines on the construction of F and G so that the prospective recovery improvement may become possible.

5.2 Recovery error bound

In this section, we examine the recovery performance of the proposed rank-sparsity-correction step by establishing a non-asymptotic recovery error bound in Frobenius norm for the ANLPLS estimator. The derivation follows the arguments in [101, 79, 98, 99] for noisy matrix completion, which are in line with the unified framework depicted in [102] for high-dimensional analysis of M-estimators with decomposable regularizers. For the sake of simplicity, we only focus on studying problem (5.6) in the following discussion. All the analysis involved in this section is also applicable to problem (5.7) because the additional positive semidefinite constraints would only lead to better recoverability.

Now we suppose that the true low-rank component $\overline{L}$ is of rank r and admits a reduced SVD

$$\overline{L} = \overline{U}_1 \overline{\Sigma} \overline{V}_1^T,$$

where $\overline{U}_1 \in \mathcal{O}^{n_1 \times r}$, $\overline{V}_1 \in \mathcal{O}^{n_2 \times r}$, and $\overline{\Sigma} \in \mathbb{R}^{r \times r}$ is the diagonal matrix with the non-zero singular values of $\overline{L}$ being arranged in the non-increasing order. Choose $\overline{U}_2$ and $\overline{V}_2$ such that $\overline{U} = \big[\overline{U}_1, \overline{U}_2\big]$ and $\overline{V} = \big[\overline{V}_1, \overline{V}_2\big]$ are both orthogonal matrices. Notice that $\overline{U} = \overline{V}$ when $\overline{L} \in \mathcal{S}_+^n$. Then the tangent space $\mathcal{T}$ (to the set $\{L \in \mathbb{V}^{n_1 \times n_2} \,|\, \mathrm{rank}(L) \leq r\}$ at $\overline{L}$) and its orthogonal complement $\mathcal{T}^\perp$ are defined in the same way as (2.2) and (2.3). Moreover, the orthogonal projection $\mathcal{P}_\mathcal{T}$ onto $\mathcal{T}$ and the orthogonal projection $\mathcal{P}_{\mathcal{T}^\perp}$ onto $\mathcal{T}^\perp$ are given by (2.4) and (2.5).

Suppose also that $\overline{S}$ has k nonzero entries, i.e., $\|\overline{S}\|_0 = k$. Let Γ be the tangent space to the set $\{S \in \mathbb{V}^{n_1 \times n_2} \,|\, \|S\|_0 \leq k\}$ at $\overline{S}$. Then $\Gamma = \{S \in \mathbb{V}^{n_1 \times n_2} \,|\, \mathrm{supp}_S \subseteq$

$\mathrm{supp}_{\overline{S}}\}$, where $\mathrm{supp}_S := \{j \,|\, \langle \Theta_j, S\rangle \neq 0,\, j = 1,\ldots,d\}$ for all $S \in \mathbb{V}^{n_1\times n_2}$. (cf. [32, Section 3.2] and [31, Section 2.3]). Denote the orthogonal complement of Γ by $\Gamma^\perp$. Note that the orthogonal projection onto Γ is given by $\mathcal{P}_\Gamma = \mathcal{P}_{\mathrm{supp}_{\overline{S}}}$ and the orthogonal projection $\mathcal{P}_{\Gamma^\perp}$ onto $\Gamma^\perp$ is given by $\mathcal{P}_{\Gamma^\perp} = \mathcal{P}_{\mathrm{supp}_{\overline{S}}{}^c}$.

We first introduce the parameters a_L and a_S, respectively, by

$$a_L := \frac{1}{\sqrt{r}}\big\|\overline{U}_1\overline{V}_1^T - F(\mathring{L})\big\|_F \quad \text{and} \quad a_S := \frac{1}{\sqrt{k}}\big\|\mathrm{sign}(\overline{S}) - G(\mathring{S})\big\|_F. \tag{5.8}$$

Note that $a_L = 1$ and $a_S = 1$ for the NLPLS estimator where $F \equiv 0$ and $G \equiv 0$. As can be seen later, these two parameters are very important in the subsequent analysis since they embody the effect of the correction terms on the resultant recovery error bound.

Being an optimal solution to problem (5.6), the ANLPLS estimator $(\widehat{L}, \widehat{S})$ satisfies the following preliminary error estimation.

Proposition 5.1. *Let $(\widehat{L}, \widehat{S})$ be an optimal solution to problem (5.6). Denote $\widehat{\Delta}_L := \widehat{L} - \overline{L}$, $\widehat{\Delta}_S := \widehat{S} - \overline{S}$ and $\widehat{\Delta} := \widehat{\Delta}_L + \widehat{\Delta}_S$. For any given $\kappa_L > 1$ and $\kappa_S > 1$, if ρ_L and ρ_S satisfy*

$$\rho_L \geq \kappa_L\nu\left\|\frac{1}{m}\mathcal{R}_\Omega^*(\xi)\right\| \quad \text{and} \quad \rho_S \geq \kappa_S\nu\left\|\frac{1}{m}\mathcal{R}_\Omega^*(\xi)\right\|_\infty, \tag{5.9}$$

then we have

$$\frac{1}{2m}\big\|\mathcal{R}_\Omega(\widehat{\Delta})\big\|_2^2 \leq \rho_L\sqrt{r}\Big(a_L + \frac{\sqrt{2}}{\kappa_L}\Big)\big\|\widehat{\Delta}_L\big\|_F + \rho_S\sqrt{k}\Big(a_S + \frac{1}{\kappa_S}\Big)\big\|\widehat{\Delta}_S\big\|_F.$$

Proof. Since $(\widehat{L}, \widehat{S})$ is optimal and $(\overline{L}, \overline{S})$ is feasible to problem (5.6), it follows from (5.3) that

$$\begin{aligned}\frac{1}{2m}\big\|\mathcal{R}_\Omega(\widehat{\Delta})\big\|_2^2 \leq \Big\langle\frac{\nu}{m}\mathcal{R}_\Omega^*(\xi), \widehat{\Delta}\Big\rangle - \rho_L\Big(\big\|\widehat{L}\big\|_* - \big\|\overline{L}\big\|_* - \big\langle F(\mathring{L}), \widehat{\Delta}_L\big\rangle\Big)\\ - \rho_S\Big(\big\|\widehat{S}\big\|_1 - \big\|\overline{S}\big\|_1 - \big\langle G(\mathring{S}), \widehat{\Delta}_S\big\rangle\Big).\end{aligned} \tag{5.10}$$

According to the Hölder's inequality and (5.9), we obtain that

$$\begin{aligned}&\Big\langle \frac{\nu}{m}\mathcal{R}^*_\Omega(\xi), \widehat{\Delta}\Big\rangle \le \nu\Big\|\frac{1}{m}\mathcal{R}^*_\Omega(\xi)\Big\|\big\|\widehat{\Delta}_L\big\|_* + \nu\Big\|\frac{1}{m}\mathcal{R}^*_\Omega(\xi)\Big\|_\infty\big\|\widehat{\Delta}_S\big\|_1\\ &\le \frac{\rho_L}{\kappa_L}\Big(\big\|\mathcal{P}_{\mathcal{T}}(\widehat{\Delta}_L)\big\|_* + \big\|\mathcal{P}_{\mathcal{T}^\perp}(\widehat{\Delta}_L)\big\|_*\Big) + \frac{\rho_S}{\kappa_S}\Big(\big\|\mathcal{P}_{\Gamma}(\widehat{\Delta}_S)\big\|_1 + \big\|\mathcal{P}_{\Gamma^\perp}(\widehat{\Delta}_S)\big\|_1\Big). \quad (5.11)\end{aligned}$$

From the directional derivative of the nuclear norm at $\overline{L}$ (see, e.g., [122]) and the directional derivative of the ℓ_1-norm at $\overline{S}$, we know that

$$\begin{cases}\big\|\widehat{L}\big\|_* - \big\|\overline{L}\big\|_* \ge \langle \overline{P}_1\overline{P}_1^T, \widehat{\Delta}_L\rangle + \big\|\mathcal{P}_{\mathcal{T}^\perp}(\widehat{\Delta}_L)\big\|_*,\\ \big\|\widehat{S}\big\|_1 - \big\|\overline{S}\big\|_1 \ge \langle \mathrm{sign}(\overline{S}), \widehat{\Delta}_S\rangle + \big\|\mathcal{P}_{\Gamma^\perp}(\widehat{\Delta}_S)\big\|_1.\end{cases}$$

This, together with the definitions of a_L and a_S in (5.8), implies that

$$\begin{aligned}\big\|\widehat{L}\big\|_* - \big\|\overline{L}\big\|_* - \big\langle F(\mathring{L}), \widehat{\Delta}_L\big\rangle &\ge \langle \overline{P}_1\overline{P}_1^T, \widehat{\Delta}_L\rangle + \big\|\mathcal{P}_{\mathcal{T}^\perp}(\widehat{\Delta}_L)\big\|_* - \big\langle F(\mathring{L}), \widehat{\Delta}_L\big\rangle\\ &\ge -\big\|\overline{P}_1\overline{P}_1^T - F(\mathring{L})\big\|_F\big\|\widehat{\Delta}_L\big\|_F + \big\|\mathcal{P}_{\mathcal{T}^\perp}(\widehat{\Delta}_L)\big\|_*\\ &= -a_L\sqrt{r}\big\|\widehat{\Delta}_L\big\|_F + \big\|\mathcal{P}_{\mathcal{T}^\perp}(\widehat{\Delta}_L)\big\|_*, \quad (5.12)\end{aligned}$$

and

$$\begin{aligned}\big\|\widehat{S}\big\|_1 - \big\|\overline{S}\big\|_1 - \big\langle G(\mathring{S}), \widehat{\Delta}_S\big\rangle &\ge \big\langle \mathrm{sign}(\overline{S}), \widehat{\Delta}_S\big\rangle + \big\|\mathcal{P}_{\Gamma^\perp}(\widehat{\Delta}_S)\big\|_1 - \big\langle G(\mathring{S}), \widehat{\Delta}_S\big\rangle\\ &\ge -\big\|\mathrm{sign}(\overline{S}) - G(\mathring{S})\big\|_F\big\|\widehat{\Delta}_S\big\|_F + \big\|\mathcal{P}_{\Gamma^\perp}(\widehat{\Delta}_S)\big\|_1\\ &= -a_S\sqrt{k}\big\|\widehat{\Delta}_S\big\|_F + \big\|\mathcal{P}_{\Gamma^\perp}(\widehat{\Delta}_S)\big\|_1. \quad (5.13)\end{aligned}$$

By substituting (5.11), (5.12) and (5.13) into (5.10), we get that

$$\begin{aligned}\frac{1}{2m}\big\|\mathcal{R}_\Omega(\widehat{\Delta})\big\|_2^2 \le \rho_L\Big[a_L\sqrt{r}\big\|\widehat{\Delta}_L\big\|_F + \frac{1}{\kappa_L}\big\|\mathcal{P}_{\mathcal{T}}(\widehat{\Delta}_L)\big\|_* - \Big(1-\frac{1}{\kappa_L}\Big)\big\|\mathcal{P}_{\mathcal{T}^\perp}(\widehat{\Delta}_L)\big\|_*\Big]\\ + \rho_S\Big[a_S\sqrt{k}\big\|\widehat{\Delta}_S\big\|_F + \frac{1}{\kappa_S}\big\|\mathcal{P}_{\Gamma}(\widehat{\Delta}_S)\big\|_1 - \Big(1-\frac{1}{\kappa_S}\Big)\big\|\mathcal{P}_{\Gamma^\perp}(\widehat{\Delta}_S)\big\|_1\Big]. \quad (5.14)\end{aligned}$$

Since $\mathrm{rank}(\mathcal{P}_{\mathcal{T}}(\widehat{\Delta}_L)) \le 2r$ and $\|\mathcal{P}_{\Gamma}(\widehat{\Delta}_S)\|_0 \le k$, we have that

$$\big\|\mathcal{P}_{\mathcal{T}}(\widehat{\Delta}_L)\big\|_* \le \sqrt{2r}\big\|\widehat{\Delta}_L\big\|_F \quad \text{and} \quad \big\|\mathcal{P}_{\Gamma}(\widehat{\Delta}_S)\big\|_1 \le \sqrt{k}\big\|\widehat{\Delta}_S\big\|_F. \quad (5.15)$$

By combining (5.14) and (5.15) together with the assumptions that $\kappa_L > 1$ and $\kappa_S > 1$, we complete the proof. □

As pointed out in [111], the nuclear norm penalization approach for noisy matrix completion can be significantly inefficient under a general non-uniform sampling scheme, especially when certain rows or columns are sampled with very high probability. This unsatisfactory recovery performance may still exist for the proposed rank-sparsity-correction step. To avoid such a situation, we need to impose additional assumptions on the sampling distribution for the observations from $\mathcal{F}^c$. The first one is to control the smallest sampling probability.

Assumption 5.1. *There exists an absolute constant $\mu_1 \geq 1$ such that*

$$\pi_j \geq \frac{1}{\mu_1 d_s}, \quad \forall\, j \in \mathcal{F}^c.$$

Notice that $\mu_1 \geq 1$ is due to $\sum_{j\in\mathcal{F}^c} \pi_j = 1$ and $d_s = |\mathcal{F}^c|$. In particular, $\mu_1 = 1$ for the uniform sampling. Moreover, the magnitude of μ_1 does not depend on d_s or the matrix dimension. From (5.4) and Assumption 5.1, we derive that

$$\langle \mathcal{Q}_{\mathcal{F}^c}(\Delta), \Delta\rangle \geq (\mu_1 d_s)^{-1}\|\Delta\|_F^2, \quad \forall\, \Delta \in \{\Delta \in \mathbb{V}^{n_1\times n_2} \,|\, \mathcal{R}_{\mathcal{F}}(\Delta) = 0\}. \tag{5.16}$$

For the purpose of deriving a recovery error bound from Proposition 5.1, it is necessary to build a bridge that connects the term $\frac{1}{m}\|\mathcal{R}_\Omega(\widehat{\Delta})\|_2^2$ and its expectation $\langle \mathcal{Q}_{\mathcal{F}^c}(\widehat{\Delta}), \widehat{\Delta}\rangle$. The most essential ingredient for building such a bridge is the notion of restricted strong convexity (RSC) proposed in [102], which stems from the restricted eigenvalue (RE) condition formulated in [13] in the context of sparse linear regression. Fundamentally, the RSC condition says that the sampling operator $\mathcal{R}_\Omega$ is almost strongly convex when restricted to a certain subset. So far, several different forms of RSC have been proven to hold for the noisy matrix completion problem in [101, Theorem 1], [79, Lemma 12] and [98, Lemma 3.2]. However, whether or not an appropriate form of RSC holds for the problem of noisy matrix decomposition with fixed and sampled basic coefficients remains an open question. We give an affirmative answer to this question in the next theorem, whose proof follows the lines of the proofs of [79, Lemma 12] and [98, Lemma 3.2].

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

$$\vartheta_L := \mathbb{E}\left\|\frac{1}{m}\mathcal{R}_\Omega^*(\epsilon)\right\| \quad \text{and} \quad \vartheta_S := \mathbb{E}\left\|\frac{1}{m}\mathcal{R}_\Omega^*(\epsilon)\right\|_\infty. \tag{5.17}$$

Theorem 5.2. *Suppose that Assumption 5.1 holds. Given any positive numbers p_1, p_2, q_1, q_2 and t, define*

$$K(p,q,t) := \left\{ \Delta = \Delta_L + \Delta_S \;\middle|\; \begin{array}{l} \|\Delta_L\|_* \le p_1\|\Delta_L\|_F + p_2\|\Delta_S\|_F,\ \Delta_L \in \mathbb{V}^{n_1\times n_2}, \\ \|\Delta_S\|_1 \le q_1\|\Delta_L\|_F + q_2\|\Delta_S\|_F,\ \Delta_S \in \mathbb{V}^{n_1\times n_2}, \\ \mathcal{R}_{\mathcal{F}}(\Delta) = 0,\ \|\Delta\|_\infty = 1,\ \|\Delta_L\|_F^2 + \|\Delta_S\|_F^2 \ge t\mu_1 d_s \end{array} \right\},$$

where $p := (p_1, p_2)$ and $q := (q_1, q_2)$. Denote $\vartheta_m := \left(\vartheta_L^2 p_1^2 + \vartheta_L^2 p_2^2 + \vartheta_S^2 q_1^2 + \vartheta_S^2 q_2^2\right)^{\frac{1}{2}}$, where ϑ_L and ϑ_S are defined in (5.17). Then for any θ, τ_1 and τ_2 satisfying

$$\theta > 1, \quad 0 < \tau_1 < 1 \quad \textit{and} \quad 0 < \tau_2 < \frac{\tau_1}{\theta}, \tag{5.18}$$

it holds that for all $\Delta \in K(p,q,t)$,

$$\frac{1}{m}\|\mathcal{R}_\Omega(\Delta)\|_2^2 \ge \langle \mathcal{Q}_{\mathcal{F}^c}(\Delta), \Delta\rangle - \frac{\tau_1}{\mu_1 d_s}\left(\|\Delta_L\|_F^2 + \|\Delta_S\|_F^2\right) - \frac{32}{\tau_2}\mu_1 d_s \vartheta_m^2 \tag{5.19}$$

with probability at least $1 - \frac{\exp[-(\tau_1-\theta\tau_2)^2 mt^2/32]}{1-\exp[-(\theta^2-1)(\tau_1-\theta\tau_2)^2 mt^2/32]}$. In particular, given any constant $c > 0$, the inequality (5.19) holds with probability at least $1 - \frac{(n_1+n_2)^{-c}}{1-2^{-(\theta^2-1)c}}$ if taking $t = \sqrt{\frac{32c\log(n_1+n_2)}{(\tau_1-\theta\tau_2)^2 m}}$.

Proof. Let p_1, p_2, q_1, q_2 and t be any given positive numbers. For any θ, τ_1 and τ_2 satisfying (5.18), we will show that the event

$$E := \left\{ \exists\ \Delta \in K(p,q,t) \text{ such that } \begin{array}{r} \left|\frac{1}{m}\|\mathcal{R}_\Omega(\Delta)\|_2^2 - \langle \mathcal{Q}_{\mathcal{F}^c}(\Delta), \Delta\rangle\right| \\ \ge \frac{\tau_1}{\mu_1 d_s}\left(\|\Delta_L\|_F^2 + \|\Delta_S\|_F^2\right) + \frac{32}{\tau_2}\mu_1 d_s \vartheta_m^2 \end{array} \right\}$$

happens with probability less than $\frac{\exp[-(\tau_1-\theta\tau_2)^2mt^2/32]}{1-\exp[-(\theta^2-1)(\tau_1-\theta\tau_2)^2mt^2/32]}$. We first decompose the set $K(p,q,t)$ into

$$K(p,q,t)=\bigcup_{j=1}^{\infty}\left\{\Delta\in K(p,q,t)\,\middle|\,\theta^{j-1}t\le\frac{1}{\mu_1 d_s}\big(\|\Delta_L\|_F^2+\|\Delta_S\|_F^2\big)\le\theta^j t\right\}.$$

For any $s\ge t$, we further define

$$K(p,q,t,s):=\left\{\Delta\in K(p,q,t)\,\middle|\,\frac{1}{\mu_1 d_s}\big(\|\Delta_L\|_F^2+\|\Delta_S\|_F^2\big)\le s\right\}.$$

Then it is not difficult to see that $E\subseteq\bigcup_{j=1}^{\infty}E_j$ with

$$E_j:=\left\{\begin{array}{l}\exists\ \Delta\in K(p,q,t,\theta^j t)\text{ such that}\\ \left|\dfrac{1}{m}\|\mathcal{R}_\Omega(\Delta)\|_2^2-\langle\mathcal{Q}_{\mathcal{F}^c}(\Delta),\Delta\rangle\right|\ge\tau_1\theta^{j-1}t+\dfrac{32}{\tau_2}\mu_1 d_s\vartheta_m^2\end{array}\right\}$$

Thus, it suffices to estimate the probability of each simpler event E_j and then apply the union bound. Let

$$Z_s:=\sup_{\Delta\in K(p,q,t,s)}\left|\frac{1}{m}\|\mathcal{R}_\Omega(\Delta)\|_2^2-\langle\mathcal{Q}_{\mathcal{F}^c}(\Delta),\Delta\rangle\right|.$$

For any $\Delta\in\mathbb{V}^{n_1\times n_2}$, the strong laws of large numbers and (5.4) yield that

$$\frac{1}{m}\|\mathcal{R}_\Omega(\Delta)\|_2^2=\frac{1}{m}\sum_{l=1}^{m}\langle\Theta_{\omega_l},\Delta\rangle^2\xrightarrow{a.s.}\mathbb{E}\big[\langle\Theta_{\omega_l},\Delta\rangle^2\big]=\langle\mathcal{Q}_{\mathcal{F}^c}(\Delta),\Delta\rangle$$

as $m\to\infty$. Since $\|\Delta\|_\infty=1$ for all $\Delta\in K(p,q,t)$, it follows from (5.4) that for all $1\le l\le m$ and $\Delta\in K(p,q,t)$,

$$\big|\langle\Theta_{\omega_l},\Delta\rangle^2-\mathbb{E}\big[\langle\Theta_{\omega_l},\Delta\rangle^2\big]\big|\le\max\big\{\langle\Theta_{\omega_l},\Delta\rangle^2,\,\mathbb{E}\big[\langle\Theta_{\omega_l},\Delta\rangle^2\big]\big\}\le 2.$$

Then according to Massart's Hoeffding-type concentration inequality [17, Theorem 14.2] (see also [92, Theorem 9]), we know that

$$\mathbb{P}\big[Z_s\ge\mathbb{E}[Z_s]+\varepsilon\big]\le\exp\left(-\frac{m\varepsilon^2}{32}\right),\quad\forall\,\varepsilon>0.\tag{5.20}$$

Next, we estimate an upper bound of $\mathbb{E}[Z_s]$ by using the standard Rademacher symmetrization in the theory of empirical processes. Let $\{\epsilon_1, \ldots, \epsilon_m\}$ be a Rademacher sequence. Then we have

$$\begin{aligned}
\mathbb{E}[Z_s] &= \mathbb{E}\left[\sup_{\Delta\in K(p,q,t,s)} \left|\frac{1}{m}\sum_{l=1}^{m}\langle\Theta_{\omega_l},\Delta\rangle^2 - \mathbb{E}\big[\langle\Theta_{\omega_l},\Delta\rangle^2\big]\right|\right] \\
&\le 2\mathbb{E}\left[\sup_{\Delta\in K(p,q,t,s)} \left|\frac{1}{m}\sum_{l=1}^{m}\epsilon_l\langle\Theta_{\omega_l},\Delta\rangle^2\right|\right] \\
&\le 8\mathbb{E}\left[\sup_{\Delta\in K(p,q,t,s)} \left|\frac{1}{m}\sum_{l=1}^{m}\epsilon_l\langle\Theta_{\omega_l},\Delta\rangle\right|\right] = 8\mathbb{E}\left[\sup_{\Delta\in K(p,q,t,s)} \left|\left\langle\frac{1}{m}\mathcal{R}_\Omega^*(\epsilon),\Delta\right\rangle\right|\right] \\
&\le 8\mathbb{E}\left[\sup_{\Delta\in K(p,q,t,s)} \left(\left\|\frac{1}{m}\mathcal{R}_\Omega^*(\epsilon)\right\|\|\Delta_L\|_* + \left\|\frac{1}{m}\mathcal{R}_\Omega^*(\epsilon)\right\|_\infty\|\Delta_S\|_1\right)\right] \\
&\le 8\mathbb{E}\left[\sup_{\Delta\in K(p,q,t,s)} \left\|\frac{1}{m}\mathcal{R}_\Omega^*(\epsilon)\right\|\|\Delta_L\|_* + \sup_{\Delta\in K(p,q,t,s)} \left\|\frac{1}{m}\mathcal{R}_\Omega^*(\epsilon)\right\|_\infty\|\Delta_S\|_1\right] \\
&\le 8\mathbb{E}\left\|\frac{1}{m}\mathcal{R}_\Omega^*(\epsilon)\right\|\left(\sup_{\Delta\in K(p,q,t,s)}\|\Delta_L\|_*\right) + 8\mathbb{E}\left\|\frac{1}{m}\mathcal{R}_\Omega^*(\epsilon)\right\|_\infty\left(\sup_{\Delta\in K(p,q,t,s)}\|\Delta_S\|_1\right),
\end{aligned} \tag{5.21}$$

where the first inequality is due to the symmetrization theorem (see, e.g., [118, Lemma 2.3.1] and [17, Theorem 14.3]) and the second inequality follows from the contraction theorem (see, e.g., [87, Theorem 4.12] and [17, Theorem 14.4]). Notice that for any $u \ge 0$, $v \ge 0$ and $\Delta \in K(p,q,t,s)$,

$$\begin{aligned}
u\|\Delta_L\|_F + v\|\Delta_S\|_F &\le \frac{1}{2}\frac{8\mu_1 d_s}{\tau_2}\big(u^2+v^2\big) + \frac{1}{2}\frac{\tau_2}{8\mu_1 d_s}\big(\|\Delta_L\|_F^2 + \|\Delta_S\|_F^2\big) \\
&\le \frac{4\mu_1 d_s}{\tau_2}\big(u^2+v^2\big) + \frac{1}{16}\tau_2 s,
\end{aligned}$$

where the first inequality is due to the inequality of arithmetic and geometric means. From (5.21), (5.17), the definition of $K(p,q,t)$ and the above inequality,

we derive that

$$\begin{aligned}\mathbb{E}[Z_s] &\leq 8\left[\sup_{\Delta\in K(p,q,t,s)}\vartheta_L\big(p_1\|\Delta_L\|_F+p_2\|\Delta_S\|_F\big)+\sup_{\Delta\in K(p,q,t,s)}\vartheta_S\big(q_1\|\Delta_L\|_F+q_2\|\Delta_S\|_F\big)\right]\\ &\leq \frac{32}{\tau_2}\mu_1 d_s\big(\vartheta_L^2p_1^2+\vartheta_L^2p_2^2+\vartheta_S^2q_1^2+\vartheta_S^2q_2^2\big)+\tau_2 s=\frac{32}{\tau_2}\mu_1 d_s\vartheta_m^2+\tau_2 s. \qquad (5.22)\end{aligned}$$

Then it follows from (5.22) and (5.20) that

$$\mathbb{P}\left[Z_s\geq\frac{\tau_1}{\theta}s+\frac{32}{\tau_2}\mu_1 d_s\vartheta_m^2\right]\leq\mathbb{P}\left[Z_s\geq\mathbb{E}[Z_s]+\Big(\frac{\tau_1}{\theta}-\tau_2\Big)s\right]\leq\exp\left[-\Big(\frac{\tau_1}{\theta}-\tau_2\Big)^2\frac{ms^2}{32}\right].$$

This, together with the choice of $s=\theta^j t$, implies that

$$\mathbb{P}[E_j]\leq\exp\left[-\frac{1}{32}\theta^{2(j-1)}(\tau_1-\theta\tau_2)^2mt^2\right].$$

Since $\theta>1$, by using the simple fact that $\theta^j\geq 1+j(\theta-1)$ for any $j\geq 1$, we obtain that

$$\begin{aligned}\mathbb{P}[E]&\leq\sum_{j=1}^{\infty}\mathbb{P}[E_j]\leq\sum_{j=1}^{\infty}\exp\left[-\frac{1}{32}\theta^{2(j-1)}(\tau_1-\theta\tau_2)^2mt^2\right]\\ &\leq\exp\left[-\frac{1}{32}(\tau_1-\theta\tau_2)^2mt^2\right]\sum_{j=1}^{\infty}\exp\left[-\frac{1}{32}\big(\theta^{2(j-1)}-1\big)(\tau_1-\theta\tau_2)^2mt^2\right]\\ &\leq\exp\left[-\frac{1}{32}(\tau_1-\theta\tau_2)^2mt^2\right]\sum_{j=1}^{\infty}\exp\left[-\frac{1}{32}(j-1)\big(\theta^2-1\big)(\tau_1-\theta\tau_2)^2mt^2\right]\\ &=\frac{\exp\left[-(\tau_1-\theta\tau_2)^2mt^2/32\right]}{1-\exp\left[-(\theta^2-1)(\tau_1-\theta\tau_2)^2mt^2/32\right]}.\end{aligned}$$

In particular, for any given constant $c>0$, taking $t=\sqrt{\frac{32c\log(n_1+n_2)}{(\tau_1-\theta\tau_2)^2m}}$ yields that

$$\frac{\exp\left[-(\tau_1-\theta\tau_2)^2mt^2/32\right]}{1-\exp\left[-(\theta^2-1)(\tau_1-\theta\tau_2)^2mt^2/32\right]}=\frac{(n_1+n_2)^{-c}}{1-(n_1+n_2)^{-(\theta^2-1)c}}\leq\frac{(n_1+n_2)^{-c}}{1-2^{-(\theta^2-1)c}}.$$

The proof is completed. □

Thanks to Theorem 5.2, we are able to derive a further error estimation from Proposition 5.1. Here we measure the recovery error using the joint squared Frobenius norm with respective to the low-rank and sparse components. The derivation is inspired by the proofs of [79, Theorem 3] and [98, Theorem 3.3].

Proposition 5.3. *Suppose that $(\widehat{L}, \widehat{S})$ is an optimal solution to problem (5.6). Let $\widehat{\Delta}_L := \widehat{L} - \overline{L}$, $\widehat{\Delta}_S := \widehat{S} - \overline{S}$ and $\widehat{\Delta} := \widehat{\Delta}_L + \widehat{\Delta}_S$. Then under Assumption 5.1, there exist some positive absolute constants c_0, c_1, c_2 and C_0 such that if ρ_L and ρ_S are chosen according to (5.9) for any given $\kappa_L > 1$ and $\kappa_S > 1$, it holds that either*

$$\frac{\|\widehat{\Delta}_L\|_F^2 + \|\widehat{\Delta}_S\|_F^2}{d_s} \le C_0 \mu_1 (b_L + b_S)^2 \sqrt{\frac{\log(n_1+n_2)}{m}}$$

or

$$\begin{aligned}\frac{\|\widehat{\Delta}_L\|_F^2 + \|\widehat{\Delta}_S\|_F^2}{d_s} \le C_0 \mu_1^2 d_s \Bigg\{ & c_0^2 \left[\rho_L^2 r \Big(a_L + \frac{\sqrt{2}}{\kappa_L}\Big)^2 + \rho_S^2 k \Big(a_S + \frac{1}{\kappa_S}\Big)^2 \right] \\ & + \vartheta_L^2 (b_L + b_S)^2 \Big(\frac{\kappa_L}{\kappa_L - 1}\Big)^2 \left[r\big(a_L + \sqrt{2}\big)^2 + k \Big(\frac{\rho_S}{\rho_L}\Big)^2 \Big(a_S + \frac{1}{\kappa_S}\Big)^2 \right] \\ & + \max\left\{ \vartheta_S^2 (b_L + b_S)^2,\ \frac{b_L^2}{\mu_1^2 d_s^2} \right\} \Big(\frac{\kappa_S}{\kappa_S - 1}\Big)^2 \left[r \Big(\frac{\rho_L}{\rho_S}\Big)^2 \Big(a_L + \frac{\sqrt{2}}{\kappa_L}\Big)^2 + k\big(a_S + 1\big)^2 \right] \Bigg\}\end{aligned}$$

with probability at least $1 - c_1(n_1+n_2)^{-c_2}$, where ϑ_L and ϑ_S are defined in (5.17).

Proof. With the choices of ρ_L and ρ_S in (5.9), we get from (5.14) that

$$\begin{aligned}& \max\left\{ \rho_L \Big(1 - \frac{1}{\kappa_L}\Big) \|\mathcal{P}_{\mathcal{T}^\perp}(\widehat{\Delta}_L)\|_*,\ \rho_S \Big(1 - \frac{1}{\kappa_S}\Big) \|\mathcal{P}_{\Gamma^\perp}(\widehat{\Delta}_S)\|_1 \right\} \\ & \le \rho_L \Big(a_L \sqrt{r} \|\widehat{\Delta}_L\|_F + \frac{1}{\kappa_L} \|\mathcal{P}_{\mathcal{T}}(\widehat{\Delta}_L)\|_* \Big) + \rho_S \Big(a_S \sqrt{k} \|\widehat{\Delta}_S\|_F + \frac{1}{\kappa_S} \|\mathcal{P}_{\Gamma}(\widehat{\Delta}_S)\|_1 \Big),\end{aligned}$$

which, together with (5.15), implies that

$$\begin{cases} \|\widehat{\Delta}_L\|_* \le \dfrac{\kappa_L}{\kappa_L - 1} \left[\sqrt{r}\big(a_L + \sqrt{2}\big) \|\widehat{\Delta}_L\|_F + \sqrt{k} \dfrac{\rho_S}{\rho_L} \Big(a_S + \dfrac{1}{\kappa_S}\Big) \|\widehat{\Delta}_S\|_F \right], \\ \|\widehat{\Delta}_S\|_1 \le \dfrac{\kappa_S}{\kappa_S - 1} \left[\sqrt{r} \dfrac{\rho_L}{\rho_S} \Big(a_L + \dfrac{\sqrt{2}}{\kappa_L}\Big) \|\widehat{\Delta}_L\|_F + \sqrt{k}\big(a_S + 1\big) \|\widehat{\Delta}_S\|_F \right]. \end{cases} \tag{5.23}$$

Let $\hat{b} := \|\widehat{\Delta}\|_\infty$. Then it follows from the bound constraints on the low-rank and sparse components that

$$\hat{b} \le \|\widehat{\Delta}_L\|_\infty + \|\widehat{\Delta}_S\|_\infty \le 2(b_L + b_S).$$

For any fixed constants $c > 0$ and θ, τ_1, τ_2 satisfying (5.18), it suffices to consider the case that

$$\|\widehat{\Delta}_L\|_F^2 + \|\widehat{\Delta}_S\|_F^2 \geq \hat{b}^2 \mu_1 d_s \sqrt{\frac{32c\log(n_1+n_2)}{(\tau_1-\theta\tau_2)^2 m}}.$$

In this case, we know from (5.23) that $\widehat{\Delta}/\hat{b} \in K(p,q,t)$, where $t = \sqrt{\frac{32c\log(n_1+n_2)}{(\tau_1-\theta\tau_2)^2 m}}$, and $p = (p_1, p_2)$ and $q = (q_1, q_2)$ are respectively given by

$$\begin{cases} p_1 := \sqrt{r}\big(a_L + \sqrt{2}\big)\dfrac{\kappa_L}{\kappa_L - 1}, & p_2 := \sqrt{k}\dfrac{\rho_S}{\rho_L}\Big(a_S + \dfrac{1}{\kappa_S}\Big)\dfrac{\kappa_L}{\kappa_L - 1}, \\ q_1 := \sqrt{r}\dfrac{\rho_L}{\rho_S}\Big(a_L + \dfrac{\sqrt{2}}{\kappa_L}\Big)\dfrac{\kappa_S}{\kappa_S - 1}, & q_2 := \sqrt{k}\big(a_S + 1\big)\dfrac{\kappa_S}{\kappa_S - 1}. \end{cases} \tag{5.24}$$

Let $\vartheta_m := \big(\vartheta_L^2 p_1^2 + \vartheta_L^2 p_2^2 + \vartheta_S^2 q_1^2 + \vartheta_S^2 q_2^2\big)^{\frac{1}{2}}$, where ϑ_L and ϑ_S are defined in (5.17). Due to (5.16) and Theorem 5.2, we obtain that with probability at least $1 - \frac{(n_1+n_2)^{-c}}{1-2^{-(\theta^2-1)c}}$,

$$\frac{\big\|\widehat{\Delta}\big\|_F^2}{d_s} \leq \frac{\mu_1}{m}\big\|\mathcal{R}_\Omega(\widehat{\Delta})\big\|_2^2 + \frac{\tau_1}{d_s}\Big(\big\|\widehat{\Delta}_L\big\|_F^2 + \big\|\widehat{\Delta}_S\big\|_F^2\Big) + \frac{32}{\tau_2}\mu_1^2 d_s \vartheta_m^2 \hat{b}^2. \tag{5.25}$$

According to Proposition 5.1, we have that

$$\begin{aligned} &\frac{\mu_1}{m}\big\|\mathcal{R}_\Omega(\widehat{\Delta})\big\|_2^2 \leq 2\mu_1\rho_L\sqrt{r}\Big(a_L + \frac{\sqrt{2}}{\kappa_L}\Big)\big\|\widehat{\Delta}_L\big\|_F + 2\mu_1\rho_S\sqrt{k}\Big(a_S + \frac{1}{\kappa_S}\Big)\big\|\widehat{\Delta}_S\big\|_F \\ &\leq \frac{\mu_1^2 d_s}{\tau_3}\left[\rho_L^2 r\Big(a_L + \frac{\sqrt{2}}{\kappa_L}\Big)^2 + \rho_S^2 k\Big(a_S + \frac{1}{\kappa_S}\Big)^2\right] + \frac{\tau_3}{d_s}\Big(\big\|\widehat{\Delta}_L\big\|_F^2 + \big\|\widehat{\Delta}_S\big\|_F^2\Big). \end{aligned} \tag{5.26}$$

where the second inequality results from the inequality of arithmetic and geometric means, and τ_3 is an arbitrarily given constant that satisfies $0 < \tau_3 < (1-\tau_1)/2$. In addition, since $\big\|\widehat{\Delta}_L\big\|_\infty \leq 2b_L$, we then derive from (5.23) that

$$\begin{aligned} \big\|\widehat{\Delta}\big\|_F^2 &\geq \big\|\widehat{\Delta}_L\big\|_F^2 + \big\|\widehat{\Delta}_S\big\|_F^2 - 2\big\|\widehat{\Delta}_L\big\|_\infty\big\|\widehat{\Delta}_S\big\|_1 \\ &\geq \big\|\widehat{\Delta}_L\big\|_F^2 + \big\|\widehat{\Delta}_S\big\|_F^2 - 4b_L\big(q_1\big\|\widehat{\Delta}_L\big\|_F + q_2\big\|\widehat{\Delta}_S\big\|_F\big) \\ &\geq \big\|\widehat{\Delta}_L\big\|_F^2 + \big\|\widehat{\Delta}_S\big\|_F^2 - \frac{4}{\tau_3}b_L^2\big(q_1^2 + q_2^2\big) - \tau_3\Big(\big\|\widehat{\Delta}_L\big\|_F^2 + \big\|\widehat{\Delta}_S\big\|_F^2\Big). \end{aligned} \tag{5.27}$$

where the third inequality is a consequence of the inequality of arithmetic and

geometric means. Combining (5.25), (5.26) and (5.27) gives that

$$\frac{\|\widehat{\Delta}_L\|_F^2 + \|\widehat{\Delta}_S\|_F^2}{d_s} \leq \frac{\mu_1^2 d_s}{1-(\tau_1+2\tau_3)}\Bigg\{ \frac{1}{\tau_3}\Bigg[\rho_L^2 r\Big(a_L + \frac{\sqrt{2}}{\kappa_L}\Big)^2 + \rho_S^2 k\Big(a_S + \frac{1}{\kappa_S}\Big)^2\Bigg] + \frac{32}{\tau_2}\vartheta_m^2\hat{b}^2 + \frac{4}{\tau_3\mu_1^2 d_s^2} b_L^2\big(q_1^2+q_2^2\big)\Bigg\}.$$

Recall that $\vartheta_m^2 = \vartheta_L^2 p_1^2 + \vartheta_L^2 p_2^2 + \vartheta_S^2 q_1^2 + \vartheta_S^2 q_2^2$. By plugging this together with (5.24) into the above inequality and choosing τ_1, τ_2, τ_3 and θ to be absolute constants, we complete the proof. □

Next, we need the second assumption on the sampling distribution, which controls the largest sampling probability for the observations from $\mathcal{F}^c$.

Assumption 5.2. *There exists an absolute constant $\mu_2 \geq 1$ such that*

$$\pi_j \leq \frac{\mu_2}{d_s}, \quad \forall\ j \in \mathcal{F}^c.$$

Observe that $\mu_2 \geq 1$ is a consequence of $\sum_{j\in\mathcal{F}^c}\pi_j = 1$ and $d_s = |\mathcal{F}^c|$. In particular, $\mu_2 = 1$ for the uniform sampling. Moreover, the magnitude of μ_2 is independent of d_s or the matrix dimension. By using (5.4) and the orthogonality of the basis Θ, we obtain that for any $Z \in \mathbb{V}^{n_1\times n_2}$,

$$\begin{aligned}\|\mathcal{Q}_{\mathcal{F}^c}(Z)\|_F^2 &= \sum_{j\in\mathcal{F}^c}\pi_j^2\langle\Theta_j, Z\rangle^2\|\Theta_j\|_F^2 \\ &\leq \max_{j\in\mathcal{F}^c}\pi_j^2\sum_{j\in\mathcal{F}^c}\langle\Theta_j, Z\rangle^2 = \max_{j\in\mathcal{F}^c}\pi_j^2\|\mathcal{P}_{\mathcal{F}^c}(Z)\|_F^2 \leq \max_{j\in\mathcal{F}^c}\pi_j^2\|Z\|_F^2,\end{aligned}$$

which, together with Assumption 5.2, gives that

$$\|\mathcal{Q}_{\mathcal{F}^c}\| \leq \max_{j\in\mathcal{F}^c}\pi_j \leq \frac{\mu_2}{d_s}. \tag{5.28}$$

In order to obtain probabilistic upper bounds on $\|\frac{1}{m}\mathcal{R}_\Omega^*(\xi)\|$ and $\|\frac{1}{m}\mathcal{R}_\Omega^*(\xi)\|_\infty$ so that the penalization parameters ρ_L and ρ_S can be chosen explicitly according to

(5.9), we assume that the noise vector ξ are of i.i.d. sub-exponential[3] entries. This will facilitate us to apply the Bernstein-type inequalities introduced in Section 2.2.

Assumption 5.3. *The i.i.d. entries ξ_l of the noise vector ξ are sub-exponential, i.e., there exists a constant $M > 0$ such that $\|\xi_l\|_{\psi_1} \leq M$ for all $l = 1, \ldots, m$, where $\|\cdot\|_{\psi_1}$ is the Orlicz ψ_1-norm defined by (2.1).*

Let e denote the exponential constant. We first provide probabilistic upper bounds on $\|\frac{1}{m}\mathcal{R}^*_\Omega(\xi)\|$ and its expectation, which extend [79, Lemma 5 and Lemma 6] to allow the existence of fixed basis. Similar results can be found in [83, Lemma 2], [101, Lemma 6] and [98, Lemma 3.5].

Lemma 5.4. *Under Assumption 5.2 and 5.3, there exists a positive constant C_1 that depends only on the Orlicz ψ_1-norm of ξ_l, such that for every $t > 0$,*

$$\left\|\frac{1}{m}\mathcal{R}^*_\Omega(\xi)\right\| \leq C_1 \max\left\{\sqrt{\frac{\mu_2 N[t+\log(n_1+n_2)]}{md_s}}, \frac{\log(n)[t+\log(n_1+n_2)]}{m}\right\}$$

with probability at least $1-\exp(-t)$, where $N := \max\{n_1, n_2\}$ and $n := \min\{n_1, n_2\}$. Moreover, when $m \geq \big(d_s \log^2(n)\log(n_1+n_2)\big)/(\mu_2 N)$, we also have that

$$\mathbb{E}\left\|\frac{1}{m}\mathcal{R}^*_\Omega(\xi)\right\| \leq C_1\sqrt{\frac{2e\mu_2 N\log(n_1+n_2)}{md_s}}.$$

Proof. Recall that $\Omega = \{\omega_l\}_{l=1}^m$ are i.i.d. copies of the random variable ω with probability distribution Π over $\mathcal{F}^c$. For $l = 1, \ldots, m$, define the random matrix Z_{ω_l} associated with ω_l by

$$Z_{\omega_l} := \xi_l \Theta_{\omega_l}.$$

Then $Z_{\omega_1}, \ldots, Z_{\omega_m}$ are i.i.d. random matrices, and we have the following decomposition

$$\frac{1}{m}\mathcal{R}^*_\Omega(\xi) = \frac{1}{m}\sum_{l=1}^m \xi_l \Theta_{\omega_l} = \frac{1}{m}\sum_{l=1}^m Z_{\omega_l}.$$

[3] See Definition 2.1 for the definition of a sub-exponential random variable.

Notice that ξ_l and Θ_{ω_l} are independent. Since $\mathbb{E}[\xi_l] = 0$, we get that $\mathbb{E}[Z_{\omega_l}] = \mathbb{E}[\xi_l]\mathbb{E}[\Theta_{\omega_l}] = 0$. Moreover, $\|\Theta_{\omega_l}\|_F = 1$ implies that

$$\|Z_{\omega_l}\| \leq \|Z_{\omega_l}\|_F = |\xi_l|\|\Theta_{\omega_l}\|_F = |\xi_l|,$$

which, together with Assumption 5.3, yields that

$$\big\|\|Z_{\omega_l}\|\big\|_{\psi_1} \leq \|\xi_l\|_{\psi_1} \leq M.$$

In addition, it follows from $\mathbb{E}[\xi_l^2] = 1$ that

$$\mathbb{E}^{\frac{1}{2}}\big[\|Z_{\omega_l}\|^2\big] \leq \mathbb{E}^{\frac{1}{2}}\big[\xi_l^2\big] = 1.$$

Furthermore, using $\mathbb{E}[\xi_l^2] = 1$ and the independence between ξ_l and Θ_{ω_l} gives that

$$\begin{aligned}\big\|\mathbb{E}\big[Z_{\omega_l}Z_{\omega_l}^T\big]\big\| = \big\|\mathbb{E}\big[\xi_l^2\Theta_{\omega_l}\Theta_{\omega_l}^T\big]\big\| = \big\|\mathbb{E}\big[\Theta_{\omega_l}\Theta_{\omega_l}^T\big]\big\| = \bigg\|\sum_{j\in\mathcal{F}^c}\pi_j\Theta_j\Theta_j^T\bigg\| \\ \leq \max_{j\in\mathcal{F}^c}\pi_j\bigg\|\sum_{j\in\mathcal{F}^c}\Theta_j\Theta_j^T\bigg\| \leq \frac{\mu_2\max\{n_1,n_2\}}{d_s},\end{aligned}$$

where the first inequality results from the positive semidefiniteness of $\Theta_j\Theta_j^T$, and the second inequality is a consequence of (4.1), (4.2) and Assumption 5.2. A similar calculation also holds for $\|\mathbb{E}[Z_{\omega_l}^T Z_{\omega_l}]\|$. Meanwhile, since $\mathrm{Tr}(\sum_{j\in\mathcal{F}^c}\pi_j\Theta_j\Theta_j^T) = \mathrm{Tr}(\sum_{j\in\mathcal{F}^c}\pi_j\Theta_j^T\Theta_j) = 1$ due to $\sum_{j\in\mathcal{F}^c}\pi_j = 1$ and $\|\Theta_j\|_F = 1$, we have

$$\begin{aligned}\max\Big\{\big\|\mathbb{E}\big[Z_{\omega_l}Z_{\omega_l}^T\big]\big\|,\, \big\|\mathbb{E}\big[Z_{\omega_l}^T Z_{\omega_l}\big]\big\|\Big\} = \max\left\{\bigg\|\sum_{j\in\mathcal{F}^c}\pi_j\Theta_j\Theta_j^T\bigg\|,\, \bigg\|\sum_{j\in\mathcal{F}^c}\pi_j\Theta_j^T\Theta_j\bigg\|\right\} \\ \geq \frac{1}{\min\{n_1,n_2\}}.\end{aligned}$$

Therefore, we know that $\sqrt{1/\min\{n_1,n_2\}} \leq \varsigma \leq \sqrt{\mu_2\max\{n_1,n_2\}/d_s}$ and $K_1 = \max\{M,2\}$. By applying Lemma 2.6, we complete the first part of the proof.

The second part of the proof exploits the formula $\mathbb{E}[z] = \int_0^{+\infty}\mathbb{P}(z > x)\mathrm{d}x$ for any nonnegative continuous random variable z. From the first part of Lemma 5.4

together with a suitable change of variables, we can easily derive that

$$\mathbb{P}\left[\left\|\frac{1}{m}\mathcal{R}_\Omega^*(\xi)\right\| > \tau\right] \leq \begin{cases} (n_1+n_2)\exp\left(-\dfrac{md_s}{C_1^2\mu_2 N}\tau^2\right), & \text{if } \tau \leq \tau^*, \\ (n_1+n_2)\exp\left(-\dfrac{m}{C_1\log(n)}\tau\right), & \text{if } \tau > \tau^*, \end{cases} \tag{5.29}$$

where $\tau^* := \frac{C_1\mu_2 N}{d_s\log(n)}$. By using the Hölder's inequality, we get that

$$\mathbb{E}\left\|\frac{1}{m}\mathcal{R}_\Omega^*(\xi)\right\| \leq \left[\mathbb{E}\left\|\frac{1}{m}\mathcal{R}_\Omega^*(\xi)\right\|^{2\log(n_1+n_2)}\right]^{\frac{1}{2\log(n_1+n_2)}}. \tag{5.30}$$

Denote $\phi_1 := \frac{md_s}{C_1^2\mu_2 N}$ and $\phi_2 := \frac{m}{C_1\log(n)}$. Combining (5.29) and (5.30) yields that

$$\begin{aligned}
&\mathbb{E}\left\|\frac{1}{m}\mathcal{R}_\Omega^*(\xi)\right\| \leq \left\{\int_0^{+\infty}\mathbb{P}\left[\left\|\frac{1}{m}\mathcal{R}_\Omega^*(\xi)\right\| > \tau^{\frac{1}{2\log(n_1+n_2)}}\right]\mathrm{d}\tau\right\}^{\frac{1}{2\log(n_1+n_2)}} \\
&\leq \sqrt{e}\left[\int_0^{+\infty}\exp\left(-\phi_1\tau^{\frac{1}{\log(n_1+n_2)}}\right)\mathrm{d}\tau + \int_0^{+\infty}\exp\left(-\phi_2\tau^{\frac{1}{2\log(n_1+n_2)}}\right)\mathrm{d}\tau\right]^{\frac{1}{2\log(n_1+n_2)}} \\
&= \sqrt{e}\Big[\log(n_1+n_2)\phi_1^{-\log(n_1+n_2)}\Gamma(\log(n_1+n_2)) \\
&\qquad + 2\log(n_1+n_2)\phi_2^{-2\log(n_1+n_2)}\Gamma(2\log(n_1+n_2))\Big]^{\frac{1}{2\log(n_1+n_2)}}.
\end{aligned} \tag{5.31}$$

Since the gamma function satisfies the inequality (see, e.g. [78, Proposition 12]):

$$\Gamma(x) \leq \left(\frac{x}{2}\right)^{x-1}, \quad \forall\ x \geq 2,$$

we then obtain from (5.31) that

$$\begin{aligned}
\mathbb{E}\left\|\frac{1}{m}\mathcal{R}_\Omega^*(\xi)\right\| \leq \sqrt{e}\Big[&\log(n_1+n_2)^{\log(n_1+n_2)}\phi_1^{-\log(n_1+n_2)}2^{1-\log(n_1+n_2)} \\
&+ 2\log(n_1+n_2)^{2\log(n_1+n_2)}\phi_2^{-2\log(n_1+n_2)}\Big]^{\frac{1}{2\log(n_1+n_2)}}.
\end{aligned}$$

Observe that $m \geq \big(d_s\log^2(n)\log(n_1+n_2)\big)/(\mu_2 N)$ implies that $\phi_1\log(n_1+n_2) \leq \phi_2^2$. Thus, we have

$$\mathbb{E}\left\|\frac{1}{m}\mathcal{R}_\Omega^*(\xi)\right\| \leq \sqrt{\frac{2e\log(n_1+n_2)}{\phi_1}} = C_1\sqrt{\frac{2e\mu_2 N\log(n_1+n_2)}{md_s}},$$

provided that $\log(n_1+n_2) \geq 1$. This completes the second part of the proof. □

By choosing $t = c_2 \log(n_1 + n_2)$ in the first part of Lemma 5.4 with the same c_2 in Proposition 5.3, we achieve the following order-optimal bound

$$\left\| \frac{1}{m} \mathcal{R}_\Omega^*(\xi) \right\| \leq C_1 \max \left\{ \sqrt{\frac{(1+c_2)\mu_2 N \log(n_1+n_2)}{m d_s}}, \frac{(1+c_2)\log(n)\log(n_1+n_2)}{m} \right\}$$

with probability at least $1 - (n_1 + n_2)^{-c_2}$. When the sample size satisfies

$$m \geq \frac{(1+c_2) d_s \log^2(n) \log(n_1+n_2)}{\mu_2 N},$$

the first term of the above maximum dominates the second term. Hence, for any given $\kappa_L > 1$, we may choose

$$\rho_L = C_1 \kappa_L \nu \sqrt{\frac{(1+c_2)\mu_2 N \log(n_1+n_2)}{m d_s}}.$$

In addition, note that Bernoulli random variables are sub-exponential. Thus, the second part of Lemma 5.4 also gives an upper bound of ϑ_L defined in (5.17).

We then turn to consider probabilistic upper bounds on $\|\frac{1}{m}\mathcal{R}_\Omega^*(\xi)\|_\infty$ and its expectation. As a preparation, the next lemma bounds the maximum number of repetitions of any index in Ω, which is an extension of Proposition 4.5 to the non-uniform sampling model under Assumption 5.2.

Lemma 5.5. *Under Assumption 5.2, for all $c > 0$, if the sample size satisfies $m < \frac{8}{3}(1+c) d_s \log(2n_1 n_2)/\mu_2$, we have*

$$\|\mathcal{R}_\Omega^* \mathcal{R}_\Omega - m\mathcal{Q}_{\mathcal{F}^c}\| \leq \frac{8}{3}(1+c)\log(2n_1 n_2)$$

with probability at least $1 - (2n_1 n_2)^{-c}$. Consequently,

$$\|\mathcal{R}_\Omega^* \mathcal{R}_\Omega\| \leq m\|\mathcal{Q}_{\mathcal{F}^c}\| + \frac{8}{3}(1+c)\log(2n_1 n_2) \leq \frac{16}{3}(1+c)\log(2n_1 n_2).$$

Proof. Let ω be a random variable with probability distribution Π over $\mathcal{F}^c$. Define the random operator $\mathcal{Z}_\omega : \mathbb{V}^{n_1 \times n_2} \to \mathbb{V}^{n_1 \times n_2}$ associated ω by

$$\mathcal{Z}_\omega(Z) := \langle \Theta_\omega, Z \rangle \Theta_\omega - \mathcal{Q}_{\mathcal{F}^c}(Z), \quad Z \in \mathbb{V}^{n_1 \times n_2}.$$

According to (5.2), we have the following decomposition

$$\mathcal{R}_\Omega^*\mathcal{R}_\Omega - m\mathcal{Q}_{\mathcal{F}^c} = \sum_{l=l}^{m} \left(\langle\Theta_{\omega_l}, \cdot\rangle\Theta_{\omega_l} - \mathcal{Q}_{\mathcal{F}^c}\right) = \sum_{l=l}^{m} \mathcal{Z}_{\omega_l}.$$

By using (5.4), we can check that

$$\mathbb{E}[\mathcal{Z}_\omega] = 0 \quad \text{and} \quad \|\mathcal{Z}_\omega\| \le 1 =: K.$$

Observe that $\mathcal{Z}_\omega$ is self-adjoint, i.e., $\mathcal{Z}_\omega^* = \mathcal{Z}_\omega$. Then a direct calculation shows that for any $Z \in \mathbb{V}^{n_1\times n_2}$,

$$\mathcal{Z}_\omega^*\mathcal{Z}_\omega(Z) = \mathcal{Z}_\omega\mathcal{Z}_\omega^*(Z) = (1-2\pi_\omega)\langle\Theta_\omega, Z\rangle\Theta_\omega + \sum_{j\in\mathcal{F}^c} \pi_j^2\langle\Theta_j, Z\rangle\Theta_j.$$

Therefore, we obtain that

$$\mathbb{E}[\mathcal{Z}_\omega^*\mathcal{Z}_\omega] = \mathbb{E}[\mathcal{Z}_\omega\mathcal{Z}_\omega^*] = \sum_{j\in\mathcal{F}^c} \pi_j(1-\pi_j)\langle\Theta_j, \cdot\rangle\Theta_j,$$

which, together with Assumption 5.2, yields that

$$\|\mathbb{E}[\mathcal{Z}_\omega^*\mathcal{Z}_\omega]\| = \|\mathbb{E}[\mathcal{Z}_\omega\mathcal{Z}_\omega^*]\| \le \max_{j\in\mathcal{F}^c} \pi_j(1-\pi_j) \le \frac{\mu_2}{d_s} =: \varsigma^2.$$

Let $t^* := \frac{8}{3}(1+c)\log(2n_1n_2)$. If $m < \frac{8}{3}(1+c)d_s\log(2n_1n_2)/\mu_2$, then $t^* > \frac{m\varsigma^2}{K}$. From Lemma 2.5, we know that

$$\mathbb{P}\left[\left\|\sum_{l=1}^{m} \mathcal{Z}_{\omega_l}\right\| > t^*\right] \le 2n_1n_2 \exp\left(-\frac{3}{8}\frac{t^*}{K}\right) \le (2n_1n_2)^{-c}.$$

Since $m\|\mathcal{Q}_{\mathcal{F}^c}\| \le \frac{m\mu_2}{d_s} \le \frac{8}{3}(1+c)\log(2n_1n_2)$ from (5.28), the proof is completed. □

By using Assumption 5.3 and Lemma 5.5, we derive probabilistic upper bounds on $\|\frac{1}{m}\mathcal{R}_\Omega^*(\xi)\|_\infty$ and its expectation in the following lemma.

Lemma 5.6. *Under Assumption 5.3 and the assumptions in Lemma 5.5, there exists a positive constant C_2 that depends only on the Orlicz ψ_1-norm of ξ_l, such that for every $t > 0$ and every $c > 0$,*

$$\left\|\frac{1}{m}\mathcal{R}_\Omega^*(\xi)\right\|_\infty \le C_2 \max\left\{\sqrt{\frac{(1+c)\log(2n_1n_2)[t+\log(n_1n_2)]}{m^2}}, \frac{t+\log(n_1n_2)}{m}\right\}$$

with probability at least $1 - 2\exp(-t) - (2n_1n_2)^{-c}$. *Moreover, it also holds that*

$$\mathbb{E}\left\|\frac{1}{m}\mathcal{R}_\Omega^*(\xi)\right\|_\infty \leq C_2\frac{\log(3e)+\log(2n_1n_2)}{m}.$$

Proof. For any index (i,j) such that $1 \leq i \leq n_1$, $1 \leq j \leq n_2$ and $(\Theta_l)_{ij} \neq 0$ for some $l \in \mathcal{F}^c$, let $w^{ij} := ((\Theta_{\omega_1})_{ij}, \ldots, (\Theta_{\omega_m})_{ij})^T \in \mathbb{R}^m$. From Lemma 2.4, we know that there exists a constant $C > 0$ such that for any $\tau > 0$,

$$\mathbb{P}\left[\left|\frac{1}{m}\sum_{l=1}^m w_l^{ij}\xi_l\right| > \tau\right] \leq 2\exp\left[-C\min\left(\frac{m^2\tau^2}{M^2\|w^{ij}\|_2^2}, \frac{m\tau}{M\|w^{ij}\|_\infty}\right)\right].$$

Recall that Ω is the multiset of indices sampled with replacement from the unfixed index set $\mathcal{F}^c$ with $d_s = |\mathcal{F}^c|$. By taking a union bound, we get that

$$\mathbb{P}\left[\left\|\frac{1}{m}\mathcal{R}_\Omega^*(\xi)\right\|_\infty > \tau\right] \leq 2d_s\exp\left[-C\min\left(\frac{m^2\tau^2}{M^2\max\|w^{ij}\|_2^2}, \frac{m\tau}{M\max\|w^{ij}\|_\infty}\right)\right],$$

where both of the maximums are taken over all such indices (i,j). Denote the maximum number of repetitions of any index in Ω by $m_{\max}$. Evidently, $m_{\max} \leq \|\mathcal{R}_\Omega^*\mathcal{R}_\Omega\|$. According to Lemma 5.5, for every $c > 0$, it holds that

$$\max\|w^{ij}\|_2^2 \leq m_{\max} \leq \|\mathcal{R}_\Omega^*\mathcal{R}_\Omega\| \leq \frac{16}{3}(1+c)\log(2n_1n_2)$$

with probability at least $1-(2n_1n_2)^{-c}$. Also note that $\max\|w^{ij}\|_\infty \leq 1$. By letting

$$\begin{aligned}-t &:= -C\min\left(\frac{3}{16M^2}\frac{m^2\tau^2}{(1+c)\log(2n_1n_2)}, \frac{m\tau}{M}\right) + \log(n_1n_2)\\ &\geq -C\min\left(\frac{m^2\tau^2}{M^2\max\|w^{ij}\|_2^2}, \frac{m\tau}{M\max\|w^{ij}\|_\infty}\right) + \log(d_s),\end{aligned}$$

we obtain that with probability at least $1 - 2\exp(-t) - (2n_1n_2)^{-c}$,

$$\left\|\frac{1}{m}\mathcal{R}_\Omega^*(\xi)\right\|_\infty \leq M\max\left\{\sqrt{\frac{16(1+c)}{3C}\frac{\log(2n_1n_2)[t+\log(n_1n_2)]}{m^2}}, \frac{1}{C}\frac{t+\log(n_1n_2)}{m}\right\}.$$

Then choosing a new constant C_2 (only depending on the Orlicz ψ_1-norm of ξ_l) in the above inequality completes the first part of the proof.

Next, we proceed to prove the second part. By taking $c = t/\log(2n_1n_2)$ and $\tau = C_2[t + \log(2n_1n_2)]/m$ in the first part of Lemma 5.6, we get that

$$\mathbb{P}\left[\left\|\frac{1}{m}\mathcal{R}_\Omega^*(\xi)\right\|_\infty > \tau\right] \leq \begin{cases} 1, & \text{if } \tau \leq \tau^*, \\ 3\exp\left[-\dfrac{m}{C_2}\tau + \log(2n_1n_2)\right], & \text{if } \tau > \tau^*, \end{cases}$$

where $\tau^* := C_2[\log(3)+\log(2n_1n_2)]/m$. Recall that for any nonnegative continuous random variable z, we have $\mathbb{E}[z] = \int_0^{+\infty}\mathbb{P}(z > x)\mathrm{d}x$. Then it follows that

$$\begin{aligned}\mathbb{E}\left\|\frac{1}{m}\mathcal{R}_\Omega^*(\xi)\right\|_\infty &\leq \int_0^{\tau^*}\mathrm{d}\tau + \int_{\tau^*}^{+\infty} 3\exp\left[-\frac{m}{C_2}\tau + \log(2n_1n_2)\right]\mathrm{d}\tau \\ &= \frac{C_2[\log(3)+\log(2n_1n_2)]}{m} + \frac{C_2}{m} = C_2\frac{\log(3e)+\log(2n_1n_2)}{m},\end{aligned}$$

which completes the second part of the proof. □

To achieve an order-optimal upper bound for $\|\frac{1}{m}\mathcal{R}_\Omega^*(\xi)\|_\infty$, we may take $t = c_2\log(2n_1n_2)$ and $c = c_2$ in the first part of Lemma 5.6, where c_2 is the same as that in Proposition 5.3. With this choice, it follows that

$$\left\|\frac{1}{m}\mathcal{R}_\Omega^*(\xi)\right\|_\infty \leq C_2(1+c_2)\frac{\log(2n_1n_2)}{m}$$

with probability at least $1 - 3(2n_1n_2)^{-c_2}$. Thus, for any given $\kappa_S > 0$, we may set

$$\rho_S = C_2(1+c_2)\kappa_S\nu\frac{\log(2n_1n_2)}{m}.$$

Furthermore, since Bernoulli random variables are sub-exponential, the second part of Lemma 5.6 also gives an upper bound of ϑ_S defined in (5.17).

Now, we are ready to present a non-asymptotic recovery error bound, measured in the joint squared Frobenius norm, of the ANLPLS estimator for the problem of noisy low-rank and sparse matrix decomposition with fixed and sampled basis coefficients under the high-dimensional scaling. Compared to the exact recovery guarantees, i.e., Theorem 4.3 and Theorem 4.4, in the noiseless setting, this

error bound severs as an approximate recovery guarantee for the noisy case. Let a_L and a_S be the parameters defined in (5.8). Below we first define three fundamental terms that compose the recovery error bound

$$\begin{cases} \Upsilon_1 := \mu_2\Big(\sqrt{2}+a_L\kappa_L\Big)^2\,\dfrac{rN\log(n_1+n_2)}{m}+\big(1+a_S\kappa_S\big)^2\,\dfrac{kd_s\log^2(2n_1n_2)}{m^2}, \\ \Upsilon_2 := \mu_2\Big(\sqrt{2}+a_L\Big)^2\,\dfrac{rN\log(n_1+n_2)}{m}+\left(\dfrac{1+a_S\kappa_S}{\kappa_L}\right)^2\,\dfrac{kd_s\log^2(2n_1n_2)}{m^2}, \\ \Upsilon_3 := \mu_2\left(\dfrac{\sqrt{2}+a_L\kappa_L}{\kappa_S}\right)^2\,\dfrac{rN\log(n_1+n_2)}{m}+(1+a_S)^2\,\dfrac{kd_s\log^2(2n_1n_2)}{m^2}. \end{cases} \tag{5.32}$$

Theorem 5.7. *Let $(\widehat{L},\widehat{S})$ be an optimal solution to problem (5.6). Denote $\widehat{\Delta}_L := \widehat{L}-\overline{L}$ and $\widehat{\Delta}_S := \widehat{S}-\overline{S}$. Define $d_s := |\mathcal{F}^c|$, $N := \max\{n_1,n_2\}$ and $n := \min\{n_1,n_2\}$. Under Assumption 5.1, 5.2 and 5.3, there exist some positive absolute constants c_0', c_1', c_2', c_3' and some positive constants C_0', C_L', C_S' (only depending on the Orlicz ψ_1-norm of ξ_l) such that when the sample size satisfies*

$$m \geq c_3'\frac{d_s\log^2(n)\log(n_1+n_2)}{\mu_2 N},$$

if for any given $\kappa_L > 1$ and $\kappa_S > 1$, the penalization parameters ρ_L and ρ_S are chosen as

$$\rho_L = C_L'\kappa_L\nu\sqrt{\frac{\mu_2 N\log(n_1+n_2)}{md_s}} \quad \textit{and} \quad \rho_S = C_S'\kappa_S\nu\frac{\log(2n_1n_2)}{m},$$

then with probability at least $1-c_1'(n_1+n_2)^{-c_2'}$, it holds that either

$$\frac{\big\|\widehat{\Delta}_L\big\|_F^2+\big\|\widehat{\Delta}_S\big\|_F^2}{d_s} \leq C_0'\mu_1(b_L+b_S)^2\sqrt{\frac{\log(n_1+n_2)}{m}}$$

or

$$\frac{\big\|\widehat{\Delta}_L\big\|_F^2+\big\|\widehat{\Delta}_S\big\|_F^2}{d_s} \leq C_0'\mu_1^2\left\{c_0'^2\nu^2\,\Upsilon_1+(b_L+b_S)^2\left[\left(\frac{\kappa_L}{\kappa_L-1}\right)^2\Upsilon_2+\left(\frac{\kappa_S}{\kappa_S-1}\right)^2\Upsilon_3\right]\right\},$$

where Υ_1, Υ_2 and Υ_3 are defined in (5.32).

Proof. From Lemma 5.6, we know that $\vartheta_S = O(\log(2n_1n_2)/m)$, whose order dominates the order of $1/d_s$ in the high-dimensional setting. Then the proof can be completed by applying Proposition 5.3, Lemma 5.4 and Lemma 5.6. □

As can be seen from Theorem 5.7, if the parameters a_L and a_S defined in (5.8) are both less than 1, we will have a reduced recovery error bound for the ANLPLS estimator (5.6) compared to that for the NLPLS estimator (5.5).

5.3 Choices of the correction functions

In this section, we discuss the choices of the rank-correction function F and the sparsity-correction function G in order to make the parameters a_L and a_S defined in (5.8) both as small as possible.

The construction of the rank-correction function F is suggested by [98, 99]. First we define the scalar function $\phi : \mathbb{R} \to \mathbb{R}$ as

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

for some $\tau > 0$ and $\varepsilon > 0$. Then we let $F : \mathbb{V}^{n_1\times n_2} \to \mathbb{V}^{n_1\times n_2}$ be a spectral operator[4] defined by

$$F(Z) := U\mathrm{Diag}\big(f(\sigma(Z))\big)V^T, \quad \forall\ Z \in \mathbb{V}^{n_1\times n_2}$$

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

$$f_j(z) := \begin{cases} \phi\left(\dfrac{z_j}{\|z\|_\infty}\right), & \text{if } z \in \mathbb{R}^n \setminus \{0\}, \\ 0, & \text{if } z = 0, \end{cases}$$

where Z admits a singular value decomposition of the form $Z = U\mathrm{Diag}\big(\sigma(Z)\big)V^T$, $\sigma(Z) := \big(\sigma_1(Z), \ldots, \sigma_n(Z)\big)^T$ is the vector of singular values of Z arranged in the

[4]We refer the reader to [40, 41] for the extensive studies on spectral operators of matrices.

non-increasing order, $U \in \mathcal{O}^{n_1}$ and $V \in \mathcal{O}^{n_2}$ are orthogonal matrices respectively corresponding to the left and right singular vectors, $\mathrm{Diag}(z) \in \mathbb{V}^{n_1 \times n_2}$ represents the diagonal matrix with the j-th main diagonal entry being z_j for $j = 1, \ldots, n$, and $n := \min\{n_1, n_2\}$. In particular, when the initial estimator is the NLPLS estimator (5.5), we may adopt the recommendation of the choices $\varepsilon \approx 0.05$ (within $0.01 \sim 0.1$) and $\tau = 2$ (within $1 \sim 3$) from [99]. With this rank-correction function F, the parameter a_L was shown to be less than 1 under certain conditions in [99].

Enlightened by the above construction of the rank-correction function F, we may construct the sparsity-correction function G as follows. First we define the operator $\mathcal{R}_d : \mathbb{V}^{n_1 \times n_2} \to \mathbb{R}^d$ as

$$\mathcal{R}_d(Z) := (\langle \Theta_1, Z \rangle, \ldots, \langle \Theta_d, Z \rangle)^T, \quad Z \in \mathbb{V}^{n_1 \times n_2}.$$

Then we let $G : \mathbb{V}^{n_1 \times n_2} \to \mathbb{V}^{n_1 \times n_2}$ be the operator

$$G(Z) := \mathcal{R}_d^*\big(g\big(\mathcal{R}_d(Z)\big)\big), \quad Z \in \mathbb{V}^{n_1 \times n_2}$$

generated from the symmetric function $g : \mathbb{R}^d \to \mathbb{R}^d$ given by

$$g_j(z) := \begin{cases} \phi\left(\dfrac{z_j}{\|z\|_\infty}\right), & \text{if } z \in \mathbb{R}^d \setminus \{0\}, \\ 0, & \text{if } z = 0, \end{cases}$$

where $\mathcal{R}_d^* : \mathbb{R}^d \to \mathbb{V}^{n_1 \times n_2}$ is the adjoint of $\mathcal{R}_d$. However, so far we have not established any theoretical guarantee such that the parameter $a_S < 1$.

Chapter 6

Correlation matrix estimation in strict factor models

In this chapter, we apply the correction procedure proposed in Chapter 5 to correlation matrix estimation with missing observations in strict factor models where the sparse component is diagonal. By virtue of this application, the specialized recovery error bound and the convincing numerical results validate the superiority of the two-stage correction approach over the nuclear norm penalization.

6.1 The strict factor model

We start with introducing the strict factor model in this section. Assume that an observable random vector $\mathrm{z} \in \mathbb{R}^n$ has the following linear structure

$$\mathrm{z} = \mathrm{Bf} + \mathrm{e}, \tag{6.1}$$

where $\mathrm{B} \in \mathbb{R}^{n\times r}$ is a deterministic matrix, $\mathrm{f} \in \mathbb{R}^r$ and $\mathrm{e} \in \mathbb{R}^n$ are unknown random vectors uncorrelated with each other. In the terminology of factor analysis (see, e.g., [86, 93]), the matrix B is called the loading matrix, and the components

of the random vectors f and e are referred to as the common factors and the idiosyncratic components, respectively. For the purpose of simplifying the model (6.1), the number of the hidden common factors is usually assumed to be far less than that of the observed variables, i.e., $r \ll n$. In a strict factor model, we assume further that the components of e are uncorrelated. This, together with the uncorrelation between f and e, implies that the covariance matrix of z satisfies

$$\mathrm{cov}[\mathrm{z}] = \mathrm{B}\mathrm{cov}[\mathrm{f}]\mathrm{B}^T + \mathrm{cov}[\mathrm{e}],$$

where $\mathrm{B}\mathrm{cov}[\mathrm{f}]\mathrm{B}^T \in \mathcal{S}_+^n$ is of rank r and $\mathrm{cov}[\mathrm{e}] \in \mathcal{S}_+^n$ is diagonal. Therefore, such a "low-rank plus diagonal" structure should be taken into account when estimating correlation matrices in strict factor models.

6.2 Recovery error bounds

Let $\overline{X} \in \mathcal{S}_+^n$ be the correlation matrix of a certain observable random vector $\mathrm{z} \in \mathbb{R}^n$ obeying the strict factor model (6.1), with the low-rank component $\overline{L} \in \mathcal{S}_+^n$ and the diagonal component $\overline{S} \in \mathcal{S}_+^n$.

Based on the observation model (5.1), we aim to estimate the unknown correlation matrix $\overline{X}$ and its low-rank and diagonal components $(\overline{L}, \overline{S})$ via solving the following convex conic optimization problem

$$\begin{array}{ll} \min & \dfrac{1}{2m}\|y - \mathcal{R}_\Omega(L+S)\|_2^2 + \rho_L\langle I_n - F(\mathring{L}), L\rangle \\ \text{s.t.} & \mathrm{diag}(L+S) = \mathbf{1},\ L \in \mathcal{S}_+^n,\ S \in \mathcal{S}_+^n \cap \mathcal{D}^n, \end{array} \tag{6.2}$$

where $F : \mathcal{S}^n \to \mathcal{S}^n$ is a given rank-correction function, $\mathring{L} \in \mathcal{S}^n$ is an initial estimator of the true low-rank component $\overline{L}$, $\mathrm{diag} : \mathcal{S}^n \to \mathbb{R}^n$ is the linear operator taking the main diagonal of any given symmetric matrix, and $\mathcal{D}^n \subset \mathcal{S}^n$ is set of all diagonal matrices. Clearly, problem (6.2) can be considered as a specialized version of problem (5.7) for strict factor models with the penalization parameter

ρ_S being chosen as 0, because the true sparse component $\overline{S}$ is known to be diagonal. In addition, as a result of the diagonal dominance of matrices in $\mathcal{S}_+^n$, the bound constraints in problem (5.7) for establishing recovery error bounds are no longer needed in problem (6.2).

Notice that the fixed index set $\mathcal{F}$ and the unfixed index set $\mathcal{F}^c$ are now corresponding to the diagonal and the off-diagonal basis coefficients, respectively. Thus, $d_s := |\mathcal{F}^c| = n(n-1)/2$. Since the multiset of indices Ω is sampled from the unfixed index set $\mathcal{F}^c$, we can equivalently reformulate problem (6.2) as

$$\begin{aligned} \min \quad & \frac{1}{2m}\|y - \mathcal{R}_\Omega(L)\|_2^2 + \rho_L\langle I_n - F(\mathring{L}), L\rangle \\ \text{s.t.} \quad & \operatorname{diag}(L) \leq \mathbf{1},\ L \in \mathcal{S}_+^n, \end{aligned} \tag{6.3}$$

in the sense that $\widehat{L}$ solves problem (6.3) if and only if $(\widehat{L}, \widehat{S})$ solves problem (6.2) with $\operatorname{diag}(\widehat{S}) = \mathbf{1} - \operatorname{diag}(\widehat{L})$. As coined by [98, 99], any optimal solution $\widehat{L}$ to problem (6.3) is called the adaptive nuclear semi-norm penalized least squares (ANPLS) estimator. By following the unified framework discussed in [102], the specialized recovery error bound for problem (6.3) can be derived in a similar but simpler way to that in Theorem 5.7. It is worth mentioning that analogous recovery results have also been established in [101, 79, 98, 99] in the context of high-dimensional noisy matrix completion.

Theorem 6.1. *Let $\widehat{L}$ be an optimal solution to problem (6.3). Denote $d_s := |\mathcal{F}^c| = n(n-1)/2$. Under Assumption 5.1, 5.2 and 5.3, there exist some positive absolute constants c_0', c_1', c_2', c_3' and some positive constants C_0', C_L' (only depending on the Orlicz ψ_1-norm of ξ_l) such that when the sample size satisfies*

$$m \geq c_3' \frac{d_s \log^2(n)\log(2n)}{\mu_2 n},$$

if for any given $\kappa_L > 1$, the penalization parameter ρ_L is chosen as

$$\rho_L = C_L' \kappa_L \nu \sqrt{\frac{\mu_2 n \log(2n)}{m d_s}},$$

then with probability at least $1 - c_1'(2n)^{-c_2'}$*, it holds that*

$$\frac{\|\widehat{L} - \overline{L}\|_F^2}{d_s} \leq C_0' \mu_1^2 \mu_2 \left[c_0'^2 \nu^2 \left(\sqrt{2} + a_L \kappa_L\right)^2 + \left(\frac{\kappa_L}{\kappa_L - 1}\right)^2 \left(\sqrt{2} + a_L\right)^2 \right] \frac{nr \log(2n)}{m},$$

where a_L *is the rank-correction parameter defined in (5.8).*

As pointed out in [99, Section 3], if $a_L \ll 1$, the optimal choice of κ_L to minimize the constant part of the recovery error bound for problem (6.3) satisfies $\kappa_L = O(1/\sqrt{a_L})$. Suppose that $\mathring{L}$ is chosen to be the nuclear norm penalized least squares (NPLS) estimator on the first stage, where $F \equiv 0$. According to the relevant analysis given in [99, Section 3], with this optimal choice of κ_L, the resultant recovery error bound for the ANPLS estimator on the second stage can be reduced by around half. This reveals the superiority of the ANPLS estimator over the NPLS estimator. Furthermore, when $\kappa_L = 1/\sqrt{a_L}$ and $a_L \to 0$, it follows from Theorem 6.1 that the recovery error bound for problem (6.3) becomes

$$\frac{\|\widehat{L} - \overline{L}\|_F^2}{d_s} \leq 2C_0' \mu_1^2 \mu_2 \left(c_0'^2 \nu^2 + 1\right) \frac{nr \log(2n)}{m}. \tag{6.4}$$

On the other hand, if the rank of the true low-rank component $\overline{L}$ is known or well-estimated, we may incorporate this useful information and consider the rank-constrained problem

$$\begin{aligned} \min \quad & \frac{1}{2}\|y - \mathcal{R}_\Omega(L)\|_2^2 \\ \text{s.t.} \quad & \operatorname{diag}(L) \leq \mathbf{1},\ \operatorname{rank}(L) \leq r,\ L \in \mathcal{S}_+^n. \end{aligned} \tag{6.5}$$

By means of the unified framework studied in [102], the following recovery result for problem (6.5) can be shown in a similar but simpler way to Theorem 5.7.

Theorem 6.2. *Under Assumption 5.1, 5.2 and 5.3, any optimal solution to the rank-constrained problem (6.5) satisfies the recovery error bound (6.4) with the same constants and probability.*

This interesting connection demonstrates the power of the rank-correction term – the ANPLS estimator is able to meet the best possible recovery error bound as if the true rank were known in advance, provided that the rank-correction function F and the initial estimator $\mathring{L}$ are chosen suitably so that $F(\mathring{L})$ is close to $\overline{U}_1\overline{V}_1^T$ and thus $a_L \ll 1$. In view of this finding, the recovery error bound (6.4) can be regarded as the optimal recovery error bound for problem (6.3).

Next, we consider the specialized recovery error bound for problem (6.2), which is an immediate consequence of Theorem 6.1.

Corollary 6.3. *Let $(\widehat{L}, \widehat{S})$ be an optimal solution to problem (6.2). With the same assumptions, constants and probability in Theorem 6.1, it holds that*

$$\begin{aligned}&\frac{\left\|\widehat{L}-\overline{L}\right\|_F^2+\|\widehat{S}-\overline{S}\|_F^2}{d_s}\\ &\leq C_0'\mu_1^2\mu_2\left[c_0'^2\nu^2\left(\sqrt{2}+a_L\kappa_L\right)^2+\left(\frac{\kappa_L}{\kappa_L-1}\right)^2\left(\sqrt{2}+a_L\right)^2\right]\frac{nr\log(2n)}{m}+\frac{4n}{d_s}.\end{aligned}$$

Proof. From the diagonal dominance of matrices in $\mathcal{S}_+^n$, we know that $\|\widehat{S}-\overline{S}\|_F \leq \|\widehat{S}\|_F+\|\overline{S}\|_F \leq 2\sqrt{n}$. Due to the equivalence between problem (6.2) and problem (6.3), the proof follows from Theorem 6.1. □

6.3 Numerical algorithms

Before proceeding to the testing problems, we first describe the numerical algorithms that we use to solve problem (6.3) and problem (6.5). Specifically, we apply the proximal alternating direction method of multipliers to problem (6.3) where the true rank is unknown, and the spectral projected gradient method to problem (6.5) where the true rank is known.

6.3.1 Proximal alternating direction method of multipliers

To facilitate the following discussion, we consider a more general formulation of problem (6.3) given by

$$\begin{aligned} \min \quad & \frac{1}{2}\|\mathcal{A}(Y) - b\|_2^2 + \langle C, Y\rangle \\ \text{s.t.} \quad & \mathcal{B}(Y) - d \in \mathcal{Q},\ Y \in \mathcal{S}_+^n, \end{aligned} \tag{6.6}$$

where $\mathcal{A} : \mathcal{S}^n \to \mathbb{R}^m$ and $\mathcal{B} : \mathcal{S}^n \to \mathbb{R}^l$ are linear mappings, $b \in \mathbb{R}^m$ and $d \in \mathbb{R}^l$ are given vectors, $C \in \mathcal{S}^n$ is a given matrix, and $\mathcal{Q} := \{0\}^{l_1} \times \mathbb{R}_+^{l_2}$ with $l_1 + l_2 = l$. Notice that problem (6.6) can be equivalently reformulated as

$$\begin{aligned} \min \quad & \frac{1}{2}\|x\|_2^2 + \langle C, Y\rangle \\ \text{s.t.} \quad & \begin{pmatrix} \mathcal{A} \\ \mathcal{B} \end{pmatrix}(Y) - \begin{pmatrix} x \\ z \end{pmatrix} = \begin{pmatrix} b \\ d \end{pmatrix}, \\ & Y \in \mathcal{S}_+^n,\ x \in \mathbb{R}^m,\ z \in \mathcal{Q}, \end{aligned} \tag{6.7}$$

where $x \in \mathbb{R}^m$ and $z \in \mathcal{Q}$ are two auxiliary variables. By a simple calculation, the dual problem of problem (6.7) can be written as

$$\begin{aligned} \max \quad & \langle b, \eta\rangle + \langle d, \zeta\rangle - \frac{1}{2}\|\eta\|_2^2 \\ \text{s.t.} \quad & C - \mathcal{A}^*(\eta) - \mathcal{B}^*(\zeta) - \Lambda = 0, \\ & \Lambda \in \mathcal{S}_+^n,\ \eta \in \mathbb{R}^m,\ \zeta \in \mathcal{Q}^*, \end{aligned} \tag{6.8}$$

where $\mathcal{Q}^* = \mathbb{R}^{l_1} \times \mathbb{R}_+^{l_2}$ is the dual cone of $\mathcal{Q}$.

We next introduce the proximal alternating direction method of multipliers (proximal ADMM) as follows. The interested readers may refer to [58, Appendix B] and references therein for more details on the proximal ADMM and its convergence analysis. Given a penalty parameter $\beta > 0$, the augmented Lagrangian function

of problem (6.7) is defined by

$$L_\beta(Y,x,z;\eta,\zeta) := \frac{1}{2}\|x\|_2^2 + \langle C, Y\rangle - \langle \mathcal{A}(Y) - x - b, \eta\rangle - \langle \mathcal{B}(Y) - z - d, \zeta\rangle \\ + \frac{\beta}{2}\|\mathcal{A}(Y) - x - b\|_2^2 + \frac{\beta}{2}\|\mathcal{B}(Y) - z - d\|_2^2.$$

The basic idea of the classical ADMM [63, 60] is to minimize L_β with respect to (x,z) and then with respect to Y, followed by an update of the multiplier (η,ζ). While minimizing with respect to (x,z) admits a simple closed-form solution, minimizing with respect to Y does not have an easy solution due to the complicated quadratic terms. For the purpose of eliminating these complicated terms, the proximal ADMM introduces a proximal term when minimizing with respect to Y. In detail, the proximal ADMM for solving problem (6.7) consists of the iterations

$$\begin{cases} \big(x^{j+1}, z^{j+1}\big) := \underset{x\in\mathbb{R}^m,\, z\in\mathcal{Q}}{\arg\min}\, L_\beta(Y^j, x, z; \eta^j, \zeta^j), \\ Y^{j+1} := \underset{Y\in\mathcal{S}_+^n}{\arg\min}\, L_\beta(Y, x^{j+1}, z^{j+1}; \eta^j, \zeta^j) + \frac{1}{2}\|Y - Y^j\|_{\mathcal{S}}^2, \\ \eta^{j+1} := \eta^j - \gamma\beta\big(\mathcal{A}(Y^{j+1}) - x^{j+1} - b\big), \\ \zeta^{j+1} := \zeta^j - \gamma\beta\big(\mathcal{B}(Y^{j+1}) - z^{j+1} - d\big), \end{cases}$$

where $\gamma \in (0, (1+\sqrt{5})/2)$ is the step length, $\mathcal{S}$ is a self-adjoint positive semidefinite (not necessarily positive definite) operator on $\mathcal{S}^n$, and $\|\cdot\|_{\mathcal{S}} := \langle \cdot, \mathcal{S}(\cdot)\rangle$. In order to "cancel out" the complicated term $\beta(\mathcal{A}^*\mathcal{A} + \mathcal{B}^*\mathcal{B})$ so that the update for Y can be implemented easily, we may choose

$$\mathcal{S} := \frac{1}{\delta}\mathcal{I} - \beta\big(\mathcal{A}^*\mathcal{A} + \mathcal{B}^*\mathcal{B}\big) \quad \text{with} \quad \frac{1}{\delta} \geq \beta\|\mathcal{A}^*\mathcal{A} + \mathcal{B}^*\mathcal{B}\| > 0,$$

where $\|\cdot\|$ denotes the spectral norm of operators. With such choices of γ and $\mathcal{S}$, we know from [58, Theorem B.1] that $\{(Y^j, x^j, z^j)\}$ converges to an optimal solution to the primal problem (6.7) and (η^j, ζ^j) converges to an optimal solution to the dual problem (6.8). It is easy to see that (x,z) can be directly computed by

$$x^{j+1} = \frac{1}{1+\beta}\Big[\beta\big(\mathcal{A}(Y^j) - b\big) - \eta^j\Big] \quad \text{and} \quad z^{j+1} = \Pi_{\mathcal{Q}}\Big(\mathcal{B}(Y^j) - d - \frac{1}{\beta}\zeta^j\Big),$$

where $\Pi_{\mathcal{Q}}$ is the metric projector over $\mathcal{Q}$. Then Y can be explicitly updated by

$$Y^{j+1} = \delta \Pi_{\mathcal{S}_+^n}(G^j),$$

where $\Pi_{\mathcal{S}_+^n}$ is the metric projector over $\mathcal{S}_+^n$ and

$$G^j := \mathcal{A}^*(\eta^j) + \beta \mathcal{A}^*(x^{j+1} + b) + \mathcal{B}^*(\zeta^j) + \beta \mathcal{B}^*(z^{j+1} + d) - C + \mathcal{S}(Y^j).$$

For the purpose of deriving a reasonable stopping criterion, we construct the dual variable Λ^j at the j-th iteration as

$$\Lambda^j := \frac{1}{\delta} Y^{j+1} - G^j.$$

Then $\Lambda^j = \Pi_{\mathcal{S}_+^n}(G^j) - G^j \succeq 0$. Moreover, a direct calculation yields that

$$\begin{aligned} \Delta^j &:= C - \mathcal{A}^*(\eta^j) - \mathcal{B}^*(\zeta^j) - \Lambda^j \\ &= \frac{1}{\delta}\big(Y^j - Y^{j+1}\big) + \beta \mathcal{A}^*\big(b + x^{j+1} - \mathcal{A}(Y^j)\big) + \beta \mathcal{B}^*\big(d + z^{j+1} - \mathcal{B}(Y^j)\big). \end{aligned}$$

In our implementation, we terminate the proximal ADMM when

$$\max\{R_P^j,\, R_D^j/2,\, R_O^j\} \le \texttt{tol},$$

where $\texttt{tol} > 0$ is a pre-specified accuracy tolerance and

$$\begin{cases} R_P^j := \sqrt{\dfrac{\|\mathcal{A}(Y^j) - x^j - b\|_2^2 + \|\mathcal{B}(Y^j) - z^j - d\|_2^2}{\max\{1, \|b\|_2^2 + \|d\|_2^2\}}} \\ R_D^j := \dfrac{\|\Delta^j\|_F}{\max\{1, \|[\mathcal{A}^*(b),\, \mathcal{B}^*(d)]\|_F\}} + \dfrac{\mathrm{dist}(\zeta^j, \mathcal{Q}^*)}{\max\{1, \|d\|_2\}} \\ R_O^j := \dfrac{\mathrm{dist}(\mathcal{B}(Y^j) - d, \mathcal{Q})}{\max\{1, \|d\|_2\}} \end{cases}$$

respectively measure the relative feasibility of the primal problem (6.7), the dual problem (6.8) and the original problem (6.6) at the j-th iteration, and $\mathrm{dist}(\cdot, \cdot)$ denotes the point-to-set Euclidean distance. It is well-known that the efficiency

of the proximal ADMM depends heavily on the choice of the penalty parameter β. At each iteration, we adjust β according to the ratio R_P^j/R_D^j in order that R_P^j and R_D^j decrease almost simultaneously. Specifically, we double β if $R_P^j/R_D^j > 3$ (meaning that R_P^j converges too slowly), halve β if $R_P^j/R_D^j < 0.1$ (meaning that R_D^j converges too slowly), and keep β unchanged otherwise.

6.3.2 Spectral projected gradient method

By using a change of variable $L = Y^T Y$ with $Y \in \mathbb{R}^{r\times n}$, we can equivalently reformulate the rank-constrained problem (6.5) as

$$\begin{array}{ll} \min & h(Y) := \dfrac{1}{2}\|H \circ (Y^T Y - C)\|_F^2 \\ \text{s.t.} & Y \in \mathbb{R}^{r\times n},\ \|Y_j\|_2 \le 1,\ j = 1,\ldots,n, \end{array} \tag{6.9}$$

where $H \in \mathcal{S}^n$ is the weight matrix such that $H \circ H \circ Z = \mathcal{R}_\Omega^* \mathcal{R}_\Omega(Z)$ for all $Z \in \mathcal{S}^n$, $C \in \mathcal{S}^n$ is any matrix satisfying $H \circ H \circ C = \mathcal{R}_\Omega^*(y)$, and Y_j is the j-th column of Y for $j = 1,\ldots,n$. Denote $\mathcal{B}^{r\times n} := \{Y \in \mathbb{R}^{r\times n} \,|\, \|Y_j\|_2 \le 1,\ j = 1,\ldots,n\}$. From the definition of the weight matrix H, we can see that $\mathrm{diag}(H) = 0$. Therefore, we may assume $\mathrm{diag}(C) = \mathbf{1}$ without loss of generality.

The spectral projected gradient methods (SPGMs) were developed by [14] for the minimization of continuously differentiable functions on nonempty closed and convex sets. This kind of methods extends the classical projected gradient method [65, 88, 12] with two simple but powerful techniques, i.e., the Barzilai-Borwein spectral step length introduced in [9] and the Grippo-Lampariello-Lucidi nonmonotone line search scheme proposed in [67], to speed up the convergence. It was also proven in [14] that the SPGMs are well-defined and any accumulation point of the iterates generated by the SPGMs is a stationary point. For an interesting review of the SPGMs, one may refer to [15].

When using the SPGMs to solve problem (6.9), there are two main computational steps. One is the evaluation of the objective function $h(Y)$ and its gradient $\nabla h(Y) = 2Y[H \circ H \circ (Y^T Y - C)]$. The other is the projection onto the multiple ball-constrained set $\mathcal{B}^{r\times n}$. Since these two steps are both of low computational cost per iteration, the SPGMs are generally very efficient for problem (6.9). This explains why, as recommended by [16], the SPGMs are among the best choices for problem (6.9). The code for the SPGM (Algorithm 2.2 in [14]) implemented in our numerical experiments is downloaded from the website http://www.di.ens.fr/~mschmidt/Software/thesis.html.

6.4 Numerical experiments

In this section, we conduct numerical experiments to test the performance of the specialized two-stage correction procedure when applied to the correlation matrix estimation problem with missing observations in strict factor models. In our numerical experiments, we refer to the NPLS estimator and the ANPLS estimator as the nuclear norm penalized and the adaptive nuclear semi-norm penalized least squares estimators, respectively. Both of these two estimators are obtained by solving problem (6.3) via the proximal ADMM with different choices of the rank-correction function F, the initial point $\mathring{L}$, and the accuracy tolerance `tol`. For the NPLS estimator, we have $F \equiv 0$ and take $\texttt{tol} = 10^{-5}$. For the ANPLS estimator, we construct F according to Section 5.3 with $\varepsilon = 0.05$ and $\tau = 2$, and take $\texttt{tol} = 10^{-6}$. Moreover, we refer to the ORACLE estimator as the solution produced by the SPGM for problem (6.5), or equivalently, problem (6.9), because the true rank is assumed to be known in advance for this case. All the experiments were run in MATLAB R2012a on the Atlas5 cluster under RHEL 5 Update 9 operating system with two Intel(R) Xeon(R) X5550 2.66GHz quad-core

CPUs and 48GB RAM from the High Performance Computing (HPC) service in the Computer Centre, National University of Singapore.

Throughout the reported numerical results, RelErr_Z stands for the relative recovery error between an estimator $\widehat{Z}$ and the true matrix $\overline{Z}$, i.e.,

$$\mathrm{RelErr}_Z := \frac{\|\widehat{Z} - \overline{Z}\|_F}{\max\left\{10^{-10}, \|\overline{Z}\|_F\right\}},$$

which serves as a standard measurement of the recovery ability of different estimators. Hence in our numerical experiments, RelErr_X, RelErr_L and RelErr_S represent the relative recovery errors of $\widehat{X}$, $\widehat{L}$ and $\widehat{S}$, respectively, where $\widehat{X} = \widehat{L} + \widehat{S}$. In order to tune the penalization parameter ρ_L in problem (6.3), we define RelDev_Z to be the relative deviation from the observations at any given matrix Z, i.e.,

$$\mathrm{RelDev}_Z := \frac{\|y - \mathcal{R}_\Omega(Z)\|}{\max\left\{10^{-10}, \|y\|_2\right\}}.$$

We next validate the efficiency of the two-stage ANPLS estimator in terms of relative recovery error by comparing it with the NPLS estimator and the ORACLE estimator via numerical experiments using random data. Throughout this section, the dimension `n` = 1000 and the starting point of the SPGM for the ORACLE estimator is randomly generated by `randn(r,n)`. For reference, we also report the total computing time for each estimator.

6.4.1 Missing observations from correlations

The first testing example is for the circumstance that the missing observations are from the correlation coefficients.

Example 1. *The true low-rank component $\overline{L}$ and the true diagonal component $\overline{S}$ are randomly generated by the following commands*

```
B = randn(n,r);  B(:,1:nl) = eigw*B(:,1:nl);
```

```
L_temp = B*B';  L_weight = trace(L_temp);
randvec = rand(n,1);  S_weight = sum(randvec);
S_temp = SL*(L_weight/S_weight)*diag(randvec);
DD = diag(1./sqrt(diag(L_temp + S_temp)));
L_bar = DD*L_temp*DD; S_bar = DD*S_temp*DD;
X_bar = L_bar + S_bar;
```

where the parameter `eigw` $= 3$ *is used to control the relative magnitude between the first* `nl` *largest eigenvalues and the other nonzero eigenvalues of* $\overline{L}$*, and the parameter* `SL` *is used to control the relative magnitude between* $\overline{S}$ *and* $\overline{L}$*. The observation vector* y *in (5.3) is formed by uniformly sampling from the off-diagonal basic coefficients of* $\overline{X}$ *with i.i.d. Gaussian noise at the noise level* 10%*.*

When solving problem (6.3), we begin with a reasonably small ρ_L chosen based on the same order given in Theorem 6.1, and then search a largest ρ_L such that the relative deviation is less than the noise level and at the same time the rank of L is as small as possible. This tuning strategy for ρ_L is heuristic but usually results in a relatively smaller recovery error according to our experience.

The numerical results for Example 1 with different `r`, `SL` and `nl` are presented in Table 6.1, 6.2, 6.3, 6.4, 6.5 and 6.6. In these tables, "Sample Ratio" denotes the ratio between the number of the sampled off-diagonal basic coefficients and the number of the off-diagonal basic coefficients, i.e.,

$$\text{Sample Ratio} := \frac{m}{n(n-1)/2}.$$

Intuitively, a higher "Sample Ratio" will lead to a lower recovery error and vice versa. The ANPLS$_1$ estimator is the ANPLS estimator using the NPLS estimator as the initial estimator $\mathring{L}$, and the ANPLS$_2$ (ANPLS$_3$) estimator is the ANPLS estimator using the ANPLS$_1$ (ANPLS$_2$) estimator as the initial estimator $\mathring{L}$.

As can be seen from these six tables, when the sample ratio is not extremely small, the ANPLS$_1$ estimator is already remarkable – it reduces the recovery error by more than half compared to the NPLS estimator, and what is more, it performs as well as the ORACLE estimator by achieving the true rank and a quite comparable recovery error. Meanwhile, when the sample ratio is very low, the ANPLS$_1$ estimator is still able to significantly improve the quality of estimation. Besides, we would like to point out that the NPLS estimator with a larger ρ_L could yield a matrix with rank lower than what we reported. However, the corresponding recovery error will inevitably and greatly increase. In addition, it is worthwhile to note that the SPGM for problem (6.5) is highly efficient in terms of solution quality and computational speed for most of the tested cases, while in fact the true rank is generally unknown or even very difficult to be identified, which prohibits its usage in practice. Also, we observe that for a few cases, the SPGM is not stable enough to return an acceptable solution from a randomly generated starting point, which is understandable due to the nonconvex nature of problem (6.5).

6.4.2 Missing observations from data

The second testing example is for the circumstances that the missing observations are from the data generated by a strict factor model.

Suppose that at time $t = 1, \ldots, T$, we are given the observable data $\mathrm{z}_t \in \mathbb{R}^n$ of the random vector $\mathrm{z} \in \mathbb{R}^n$ that satisfies the strict factor model (6.1), i.e.,

$$\mathrm{z}_t = \mathrm{B}\mathrm{f}_t + \mathrm{e}_t,$$

where $\mathrm{B} \in \mathbb{R}^{n\times r}$ is the loading matrix, $\mathrm{f}_t \in \mathbb{R}^r$ and $\mathrm{e}_t \in \mathbb{R}^n$ are the unobservable data of the factor vector $\mathrm{f} \in \mathbb{R}^r$ and the idiosyncratic vector $\mathrm{e} \in \mathbb{R}^n$ at time t, respectively. By putting into a matrix form, we have

$$\mathrm{Z} = \mathrm{B}\mathrm{F} + \mathrm{E},$$

Table 6.1: Recovery performance for Example 1 with $\mathtt{n} = 1000$, $\mathtt{r} = 1$ and $\mathtt{nl} = 1$

SL	Sample Ratio	NPLS				ANPLS$_1$				ANPLS$_3$				ORACLE			
		RelErr			Time	RelErr			Time	RelErr			Time	RelErr			Time
		X	L (Rank)	S		X	L (Rank)	S		X	L (Rank)	S		X	L	S	
1	1%	13.38%	13.42% (26)	8.79%	1668.4	5.86%	5.88% (**8**)	5.11%	2385.8	5.27%	5.29% (**8**)	4.38%	3062.4	5.21%	5.23%	4.17%	2.5
	3%	5.17%	5.20% (51)	7.88%	597.7	2.72%	2.73% (1)	2.23%	1000.7	2.72%	2.73% (1)	2.22%	1065.1	2.72%	2.73%	2.22%	1.7
	5%	4.15%	4.20% (66)	10.00%	439.6	2.00%	2.01% (1)	1.63%	676.8	2.00%	2.01% (1)	1.63%	705.1	2.00%	2.01%	1.63%	1.6
	10%	3.20%	3.26% (82)	9.68%	224.4	1.44%	1.45% (1)	1.17%	360.4	1.44%	1.45% (1)	1.17%	389.1	1.44%	1.45%	1.17%	1.6
	15%	2.95%	3.02% (99)	11.25%	178.6	1.19%	1.19% (1)	0.96%	360.8	1.19%	1.19% (1)	0.96%	390.1	1.19%	1.19%	0.96%	1.3
10	1%	24.98%	25.79% (27)	5.18%	2033.9	8.92%	9.22% (**4**)	2.76%	3241.8	5.96%	6.15% (1)	1.38%	4276.8	**128.86%**	**132.93%**	**15.33%**	8.1
	3%	5.94%	6.14% (51)	2.29%	562.7	2.68%	2.76% (1)	0.59%	1027.4	2.69%	2.77% (1)	0.60%	1134.9	2.69%	2.77%	0.60%	1.8
	5%	4.61%	4.79% (66)	2.45%	443.3	2.04%	2.10% (1)	0.40%	674.5	2.03%	2.10% (1)	0.40%	715.4	2.03%	2.10%	0.40%	1.7
	10%	3.50%	3.70% (97)	3.20%	215.8	1.39%	1.44% (1)	0.26%	307.1	1.39%	1.44% (1)	0.26%	319.4	1.39%	1.44%	0.26%	1.7
	15%	3.48%	3.71% (125)	4.33%	169.1	1.13%	1.16% (1)	0.23%	238.6	1.13%	1.16% (1)	0.23%	245.4	1.13%	1.16%	0.23%	1.6
100	1%	35.41%	61.21% (19)	3.23%	710.9	20.15%	34.92% (**3**)	2.67%	1639.0	7.93%	13.82% (1)	1.43%	2546.8	**112.22%**	**194.06%**	**11.30%**	48.0
	3%	14.32%	21.56% (41)	2.10%	787.0	3.08%	4.67% (1)	0.63%	1043.9	2.13%	3.20% (1)	0.19%	1211.7	**72.00%**	**107.82%**	**2.15%**	7.2
	5%	6.69%	11.88% (11)	1.24%	469.6	1.52%	2.71% (1)	0.33%	601.0	1.21%	2.13% (1)	0.07%	686.4	1.21%	2.13%	0.07%	2.4
	10%	1.37%	2.60% (24)	0.11%	143.8	0.74%	1.41% (1)	0.04%	171.1	0.74%	1.41% (1)	0.04%	181.7	0.74%	1.41%	0.04%	2.0
	15%	1.37%	2.21% (6)	0.18%	145.9	0.74%	1.19% (1)	0.05%	167.6	0.74%	1.19% (1)	0.05%	175.9	0.74%	1.19%	0.05%	2.2

Table 6.2: Recovery performance for Example 1 with $\mathtt{n} = 1000$, $\mathtt{r} = 2$ and $\mathtt{nl} = 1$

SL	Sample Ratio	NPLS			Time	ANPLS_1			Time	ANPLS_3			Time	ORACLE			Time
		RelErr				RelErr				RelErr				RelErr			
		X	L (Rank)	S		X	L (Rank)	S		X	L (Rank)	S		X	L	S	
1	3%	7.28%	7.31% (48)	6.43%	646.7	3.87%	3.88% (2)	3.09%	1038.7	3.83%	3.84% (2)	3.02%	1214.1	3.86%	3.88%	3.06%	4.5
	5%	5.33%	5.37% (62)	7.98%	450.9	2.94%	2.95% (2)	2.32%	656.3	2.93%	2.94% (2)	2.30%	721.9	2.93%	2.94%	2.30%	3.7
	10%	3.85%	3.89% (78)	8.68%	226.5	2.03%	2.04% (2)	1.72%	338.2	2.03%	2.04% (2)	1.70%	374.0	2.04%	2.05%	1.70%	3.6
	15%	3.18%	3.23% (89)	8.84%	164.2	1.63%	1.63% (2)	1.31%	230.9	1.63%	1.63% (2)	1.31%	243.7	1.63%	1.64%	1.32%	3.2
10	3%	9.48%	9.87% (46)	2.53%	563.7	4.37%	4.56% (2)	1.40%	936.5	4.08%	4.24% (2)	0.94%	1138.0	4.11%	4.28%	0.76%	6.1
	5%	6.05%	6.29% (64)	2.14%	451.0	3.01%	3.13% (2)	0.65%	692.7	2.95%	3.06% (2)	0.54%	796.7	2.98%	3.09%	0.56%	4.3
	10%	4.24%	4.44% (97)	3.05%	200.4	2.00%	2.07% (2)	0.42%	315.3	1.98%	2.05% (2)	0.38%	349.5	1.99%	2.05%	0.38%	3.7
	15%	3.67%	3.88% (116)	3.29%	160.8	1.61%	1.67% (2)	0.33%	222.2	1.61%	1.66% (2)	0.32%	241.8	1.60%	1.66%	0.30%	3.6
100	3%	11.51%	21.55% (24)	1.75%	687.7	4.12%	7.76% (2)	0.80%	862.1	3.20%	5.99% (2)	0.47%	997.5	**8.06%**	**15.61%**	**2.88%**	16.1
	5%	6.37%	12.64% (25)	1.72%	335.7	3.07%	6.26% (2)	1.21%	411.1	2.23%	4.49% (2)	0.75%	489.2	1.87%	3.63%	0.20%	8.6
	10%	2.53%	5.00% (20)	0.49%	173.0	1.22%	2.39% (2)	0.19%	206.8	1.15%	2.24% (2)	0.11%	235.0	1.13%	2.20%	0.07%	3.9
	15%	1.87%	3.32% (16)	0.28%	175.7	1.07%	1.89% (2)	0.11%	207.1	1.03%	1.82% (2)	0.08%	227.9	0.99%	1.74%	0.05%	7.6

Table 6.3: Recovery performance for Example 1 with $\mathtt{n} = 1000$, $\mathtt{r} = 2$ and $\mathtt{nl} = 2$

SL	Sample Ratio	NPLS			Time	ANPLS_1			Time	ANPLS_3			Time	ORACLE			Time
		RelErr				RelErr				RelErr				RelErr			
		X	L (Rank)	S		X	L (Rank)	S		X	L (Rank)	S		X	L	S	
1	3%	7.51%	7.54% (44)	5.48%	639.1	3.94%	3.96% (2)	2.98%	875.1	3.94%	3.96% (2)	2.99%	1016.5	3.94%	3.96%	2.99%	2.3
	5%	5.39%	5.42% (59)	6.75%	345.9	2.87%	2.88% (2)	2.20%	473.7	2.87%	2.88% (2)	2.20%	519.7	2.87%	2.89%	2.20%	2.2
	10%	3.62%	3.65% (64)	4.95%	206.6	2.04%	2.05% (2)	1.48%	258.3	2.04%	2.04% (2)	1.48%	273.8	2.04%	2.04%	1.48%	1.8
	15%	3.08%	3.11% (79)	6.77%	153.8	1.62%	1.63% (2)	1.23%	243.5	1.62%	1.63% (2)	1.23%	268.1	1.62%	1.63%	1.23%	1.7
10	3%	12.03%	12.72% (40)	2.72%	850.5	3.97%	4.19% (2)	0.81%	1114.9	3.83%	4.05% (2)	0.67%	1350.1	3.83%	4.05%	0.67%	2.6
	5%	6.46%	6.81% (54)	1.58%	462.5	2.91%	3.06% (2)	0.52%	600.5	2.89%	3.04% (2)	0.50%	683.8	2.89%	3.04%	0.50%	3.0
	10%	3.45%	3.66% (51)	0.78%	177.0	1.94%	2.06% (2)	0.30%	216.2	1.94%	2.06% (2)	0.30%	238.4	1.94%	2.06%	0.30%	1.6
	15%	2.74%	2.90% (61)	0.80%	150.8	1.54%	1.62% (2)	0.24%	172.3	1.54%	1.62% (2)	0.24%	182.6	1.54%	1.62%	0.24%	2.3
100	3%	9.94%	25.63% (25)	1.29%	355.6	3.44%	8.97% (2)	0.71%	485.9	1.86%	4.79% (2)	0.20%	623.6	1.74%	4.46%	0.09%	8.0
	5%	6.2%	14.11% (23)	0.96%	298.9	1.83%	4.12% (2)	0.30%	388.3	1.42%	3.17% (2)	0.08%	474.2	1.42%	3.17%	0.08%	3.1
	10%	3.02%	5.95% (30)	0.54%	200.5	1.20%	2.35% (2)	0.10%	271.9	1.18%	2.30% (2)	0.07%	322.5	1.18%	2.30%	0.07%	2.8
	15%	1.71%	3.65% (33)	0.29%	151.2	0.79%	1.66% (2)	0.05%	184.7	0.78%	1.65% (2)	0.04%	205.2	0.78%	1.65%	0.04%	1.8

Table 6.4: Recovery performance for Example 1 with `n` = 1000, `r` = 3 and `nl` = 1

`SL`	Sample Ratio	NPLS				ANPLS_1				ANPLS_3				ORACLE			
		RelErr			Time	RelErr			Time	RelErr			Time	RelErr			Time
		X	L (Rank)	S		X	L (Rank)	S		X	L (Rank)	S		X	L	S	
1	3%	10.35%	10.39% (45)	7.74%	663.5	5.38%	5.40% (3)	4.52%	964.7	5.16%	5.18% (3)	3.76%	1198.3	5.22%	5.25%	3.82%	5.7
	5%	6.70%	6.73% (62)	6.62%	357.1	3.57%	3.59% (3)	2.93%	529.4	3.55%	3.56% (3)	2.83%	618.2	3.56%	3.58%	2.84%	4.4
	10%	4.46%	4.49% (71)	6.17%	221.4	2.57%	2.59% (3)	2.11%	306.0	2.56%	2.57% (3)	2.07%	343.0	2.56%	2.57%	2.06%	3.7
	15%	3.53%	3.56% (82)	6.51%	167.8	2.05%	2.05% (3)	1.53%	305.5	2.04%	2.05% (3)	1.53%	349.3	2.05%	2.06%	1.54%	3.5
10	3%	13.48%	14.13% (45)	3.51%	551.9	6.78%	7.12% (3)	2.39%	887.7	5.61%	5.88% (3)	1.46%	1140.4	**10.13%**	**10.62%**	**2.81%**	8.9
	5%	8.37%	8.80% (58)	2.44%	435.9	4.05%	4.26% (3)	1.14%	612.9	3.77%	3.96% (3)	0.72%	743.5	3.80%	3.99%	0.71%	5.6
	10%	4.91%	5.17% (83)	2.00%	189.5	2.56%	2.69% (3)	0.60%	263.1	2.51%	2.63% (3)	0.48%	310.7	2.50%	2.62%	0.44%	4.6
	15%	3.58%	3.76% (80)	1.44%	158.7	2.04%	2.13% (3)	0.39%	209.0	2.02%	2.11% (3)	0.36%	230.6	2.00%	2.09%	0.33%	4.2
100	3%	16.63%	30.15% (30)	2.87%	625.8	7.70%	14.10% (**4**)	1.90%	887.1	5.32%	9.71% (3)	1.16%	1129.5	**19.72%**	**36.82%**	**6.94%**	128.6
	5%	7.45%	13.69% (24)	1.23%	434.1	3.97%	7.31% (3)	0.78%	520.7	3.23%	5.93% (3)	0.56%	596.2	2.69%	4.90%	0.19%	21.8
	10%	3.26%	6.47% (25)	0.59%	184.0	1.70%	3.36% (3)	0.27%	215.9	1.49%	2.93% (3)	0.17%	243.7	1.37%	2.69%	0.10%	5.8
	15%	2.12%	3.98% (34)	0.32%	154.9	1.24%	2.33% (3)	0.14%	187.4	1.19%	2.22% (3)	0.11%	211.0	1.12%	2.10%	0.06%	5.3

Table 6.5: Recovery performance for Example 1 with $\mathtt{n} = 1000$, $\mathtt{r} = 3$ and $\mathtt{nl} = 2$

SL	Sample Ratio	NPLS				ANPLS$_1$				ANPLS$_3$				ORACLE			
		RelErr			Time	RelErr			Time	RelErr			Time	RelErr			Time
		X	L (Rank)	S		X	L (Rank)	S		X	L (Rank)	S		X	L	S	
1	3%	9.79%	9.83% (45)	6.53%	689.3	5.18%	5.20% (3)	4.34%	1005.9	4.95%	4.97% (3)	3.88%	1218.1	5.01%	5.04%	3.97%	5.9
	5%	6.55%	6.59% (56)	5.54%	338.5	3.86%	3.87% (3)	2.90%	481.2	3.82%	3.84% (3)	2.78%	566.2	3.83%	3.85%	2.76%	3.8
	10%	4.22%	4.24% (64)	4.30%	213.5	2.50%	2.51% (3)	1.92%	267.6	2.49%	2.50% (3)	1.89%	296.6	2.49%	2.50%	1.86%	3.4
	15%	3.45%	3.47% (71)	4.41%	155.6	2.03%	2.04% (3)	1.49%	226.0	2.02%	2.03% (3)	1.48%	252.1	2.01%	2.02%	1.49%	3.3
10	3%	13.53%	14.40% (40)	2.78%	773.6	6.25%	6.66% (3)	1.56%	1050.7	5.44%	5.79% (3)	1.14%	1305.9	**13.69%**	**14.66%**	**5.63%**	24.8
	5%	8.35%	8.88% (51)	2.26%	411.2	4.22%	4.49% (3)	1.40%	553.8	3.87%	4.12% (3)	1.05%	673.7	3.72%	3.95%	0.63%	7.9
	10%	4.33%	4.60% (49)	1.00%	179.9	2.53%	2.69% (3)	0.49%	231.8	2.48%	2.63% (3)	0.42%	268.3	2.45%	2.60%	0.37%	4.7
	15%	3.30%	3.49% (63)	0.91%	155.3	1.91%	2.02% (3)	0.35%	184.2	1.90%	2.01% (3)	0.34%	204.4	1.90%	2.00%	0.32%	4.1
100	3%	17.34%	36.84% (31)	2.83%	469.8	7.13%	15.30% (**4**)	1.66%	697.2	5.20%	11.04% (3)	0.77%	934.1	**13.15%**	**28.92%**	**4.65%**	99.6
	5%	10.46%	22.43% (31)	2.12%	306.4	3.60%	7.78% (3)	0.91%	414.6	3.08%	6.57% (3)	0.48%	515.8	2.57%	5.44%	0.22%	10.7
	10%	4.62%	9.20% (24)	1.21%	200.4	2.14%	4.28% (3)	0.64%	245.1	1.79%	3.54% (3)	0.43%	286.2	**6.65%**	**13.16%**	**1.50%**	16.3
	15%	1.88%	4.31% (19)	0.34%	129.4	1.15%	2.63% (3)	0.16%	164.4	1.11%	2.53% (3)	0.13%	186.5	0.93%	2.11%	0.05%	4.6

Table 6.6: Recovery performance for Example 1 with $\mathtt{n} = 1000$, $\mathtt{r} = 3$ and $\mathtt{nl} = 3$

SL	Sample Ratio	NPLS				ANPLS$_1$				ANPLS$_3$				ORACLE			
		RelErr			Time	RelErr			Time	RelErr			Time	RelErr			Time
		X	L (Rank)	S		X	L (Rank)	S		X	L (Rank)	S		X	L	S	
1	3%	10.76%	10.81% (40)	6.45%	747.8	5.17%	5.20% (3)	3.66%	989.3	5.09%	5.12% (3)	3.54%	1201.5	5.09%	5.12%	3.54%	2.8
	5%	6.61%	6.65% (51)	4.46%	325.1	3.76%	3.78% (3)	2.65%	422.8	3.77%	3.79% (3)	2.66%	493.7	3.77%	3.79%	2.66%	2.1
	10%	4.27%	4.29% (51)	2.67%	202.6	2.46%	2.48% (3)	1.69%	241.4	2.46%	2.48% (3)	1.69%	261.8	2.46%	2.48%	1.69%	1.7
	15%	3.31%	3.33% (57)	2.63%	139.8	2.03%	2.05% (3)	1.43%	184.2	2.03%	2.05% (3)	1.43%	202.7	2.03%	2.05%	1.43%	2.1
10	3%	20.15%	21.52% (39)	4.45%	934.9	6.27%	6.70% (3)	1.69%	1294.4	5.20%	5.54% (3)	0.83%	1673.4	5.18%	5.53%	0.79%	5.4
	5%	7.72%	8.26% (47)	1.35%	406.8	3.65%	3.91% (3)	0.51%	524.1	3.63%	3.88% (3)	0.50%	622.4	3.63%	3.88%	0.50%	3.2
	10%	4.35%	4.66% (41)	0.84%	179.0	2.37%	2.54% (3)	0.34%	231.2	2.37%	2.54% (3)	0.33%	262.7	2.37%	2.54%	0.34%	2.3
	15%	3.24%	3.48% (40)	0.61%	144.0	1.91%	2.05% (3)	0.25%	171.5	1.91%	2.05% (3)	0.25%	186.5	1.91%	2.05%	0.25%	2.1
100	3%	23.39%	51.79% (35)	4.01%	397.9	10.77%	24.15% (**5**)	2.74%	664.9	4.57%	10.20% (3)	1.04%	937.3	2.94%	6.45%	0.22%	7.7
	5%	12.67%	30.58% (34)	2.27%	283.0	3.84%	9.43% (3)	1.07%	406.5	1.92%	4.61% (3)	0.28%	521.5	1.77%	4.24%	0.15%	5.2
	10%	3.47%	8.68% (33)	0.70%	231.4	1.20%	2.99% (3)	0.18%	301.8	1.09%	2.70% (3)	0.06%	361.6	1.09%	2.70%	0.06%	2.7
	15%	1.88%	5.04% (20)	0.42%	137.4	0.82%	2.17% (3)	0.08%	174.2	0.81%	2.12% (3)	0.04%	202.4	0.81%	2.12%	0.04%	2.3

where $\mathrm{Z} \in \mathbb{R}^{n\times T}$, $\mathrm{F} \in \mathbb{R}^{r\times T}$ and $\mathrm{E} \in \mathbb{R}^{n\times T}$ consist of z_t, f_t and e_t as their columns. In the following example, we randomly simulate the case that cov[f] $= I_r$ and cov[e] is diagonal. After an appropriate rescaling, we have $\overline{L} = \mathrm{BB}^T$ and $\overline{S} = \mathrm{cov[e]}$ with $\mathrm{diag}(\overline{L} + \overline{S}) = \mathbf{1}$. To reach this setting, we may assume that the factor vector f is of uncorrelated standard normal entries and the idiosyncratic vector e follows the centered multivariate normal distribution with covariance matrix $\overline{S}$.

Example 2. *The true low-rank component* $\overline{L}$ *and the true diagonal component* $\overline{S}$ *are obtained in the same way as that in Example 1. By following the same commands in Example 1, the loading matrix* B*, the factor matrix* F*, the idiosyncratic matrix* E*, the data matrix* Z *are generated as follows*

```
B = DD*B;  F = randn(r,T);
E = mvnrnd(zeros(n,1),S_bar,T)';
Z = B*F + E;
```

where `T` *is the length of the time series,* `DD` *and the original* `B` *are inherited from Example 1. We then generate the data matrix with missing observations* `Z_missing` *and its pairwise correlation matrix* `M` *by*

```
Z_missing = Z; nm = missing_rate*n*T; Ind_missing = randperm(n*T);
              Z_missing(Ind_missing(1:nm)) = NaN;
M = nancov(Z_missing','pairwise'); M = M - diag(diag(M)) + eye(n);
```

where `missing_rate` *represents the missing rate of data. Lastly, we identify the missing correlation coefficients from* `M` *if the length of the overlapping remaining data between any two rows of* `Z_missing` *is less than* `trust_par*n`*, where* `trust_par` *is a parameter introduced to determine when a pairwise correlation calculated from* `Z_missing` *is reliable.*

The numerical results for Example 2 are listed in Table 6.7 and 6.8 with different `T`, `trust_par`, `SL` and `missing_rate`. We can observe from these two tables

that our correction procedure loses its effectiveness in general. This is not surprising because all the theoretical guarantees established for the ANPLS estimator are built on the observation model (5.1), where the missing observations are from the correlation coefficients other than from the data generated by a factor model. Moreover, it is interesting to see that the performance of the NPLS estimator is unexpectedly satisfactory in most cases.

Table 6.7: Recovery performance for Example 2 with $\mathtt{n} = 1000$, $\mathtt{r} = 5$, $\mathtt{T} = 6 * \mathtt{n}$ and $\mathtt{trust\_par} = 3.5$

SL	Missing	NPLS				ANPLS$_1$				ANPLS$_3$				ORACLE			
		RelErr			Time	RelErr			Time	RelErr			Time	RelErr			Time
	Rate	X	L (Rank)	S		X	L (Rank)	S		X	L (Rank)	S		X	L	S	
1	0%	3.78%	3.80% (5)	2.36%	90.8	3.76%	3.78% (5)	2.33%	188.5	3.76%	3.78% (5)	2.33%	192.9	3.77%	3.79%	2.33%	3.2
	5%	3.60%	3.62% (5)	2.38%	89.6	3.66%	3.68% (5)	2.52%	209.8	3.66%	3.68% (5)	2.52%	214.2	3.67%	3.69%	2.54%	2.5
	10%	4.16%	4.18% (5)	2.67%	184.3	4.14%	4.17% (5)	2.69%	329.0	4.15%	4.17% (5)	2.69%	333.4	4.15%	4.17%	2.70%	2.8
	15%	4.24%	4.26% (5)	2.91%	138.6	4.21%	4.23% (5)	2.78%	262.2	4.21%	4.23% (5)	2.78%	267.2	4.21%	4.23%	2.78%	3.0
	20%	4.83%	4.85% (5)	3.13%	270.1	4.93%	4.96% (5)	3.24%	448.4	4.93%	4.96% (5)	3.24%	452.7	4.93%	4.96%	3.25%	2.5
5	0%	6.69%	6.85% (5)	1.94%	62.0	6.29%	6.44% (5)	1.49%	73.0	6.29%	6.44% (5)	1.49%	80.1	6.29%	6.44%	1.47%	3.3
	5%	5.83%	5.97% (5)	1.40%	63.1	5.87%	6.01% (5)	1.33%	96.0	5.87%	6.01% (5)	1.33%	103.4	5.87%	6.01%	1.34%	3.1
	10%	5.90%	6.05% (5)	1.49%	94.1	5.77%	5.92% (5)	1.25%	153.0	5.77%	5.92% (5)	1.25%	162.3	5.78%	5.92%	1.25%	3.1
	15%	7.29%	7.45% (5)	1.94%	188.7	7.61%	7.78% (5)	1.99%	293.4	7.61%	7.78% (5)	1.99%	302.2	7.61%	7.78%	2.01%	3.9
	20%	6.31%	6.46% (5)	1.70%	72.4	6.27%	6.42% (5)	1.51%	83.7	6.27%	6.42% (5)	1.51%	92.8	6.28%	6.43%	1.52%	3.1
10	0%	7.41%	7.83% (5)	1.43%	65.8	7.10%	7.49% (5)	0.97%	77.9	7.10%	7.49% (5)	0.96%	87.2	7.11%	7.51%	0.95%	3.8
	5%	7.38%	7.81% (5)	1.38%	66.8	7.06%	7.47% (5)	1.01%	77.6	7.06%	7.47% (5)	1.01%	87.8	7.08%	7.48%	1.00%	4.3
	10%	7.49%	7.96% (5)	1.29%	75.0	7.37%	7.82% (5)	1.04%	87.4	7.37%	7.83% (5)	1.04%	98.4	7.39%	7.85%	1.05%	4.3
	15%	8.53%	9.00% (5)	1.83%	76.3	7.97%	8.40% (5)	1.26%	87.4	7.97%	8.41% (5)	1.25%	98.2	7.99%	8.42%	1.23%	4.0
	20%	7.63%	8.07% (5)	1.33%	75.8	7.62%	8.05% (5)	1.09%	104.5	7.62%	8.06% (5)	1.09%	115.9	7.67%	8.10%	1.11%	2.6

Table 6.8: Recovery performance for Example 2 with `n` = 1000, `r` = 5, `T` = `n` and `trust_par` = 0.8

SL	Missing	NPLS				$ANPLS_1$				$ANPLS_3$				ORACLE			
		RelErr			Time	RelErr			Time	RelErr			Time	RelErr			Time
	Rate	X	L (Rank)	S		X	L (Rank)	S		X	L (Rank)	S		X	L	S	
1	0%	8.94%	8.99% (5)	5.79%	232.3	9.01%	9.05% (5)	5.65%	397.9	9.01%	9.05% (5)	5.65%	415.9	9.01%	9.05%	5.65%	2.9
	3%	8.59%	8.63% (5)	5.63%	238.9	8.93%	8.97% (5)	6.28%	405.9	8.93%	8.97% (5)	6.28%	422.7	8.94%	8.99%	6.34%	2.6
	5%	9.17%	9.22% (5)	6.12%	106.7	9.11%	9.16% (5)	5.83%	171.4	9.11%	9.16% (5)	5.83%	180.9	9.12%	9.16%	5.83%	2.4
	8%	10.65%	10.71% (5)	7.59%	178.9	10.29%	10.34% (5)	6.66%	303.2	10.29%	10.34% (5)	6.66%	317.0	10.28%	10.33%	6.62%	2.6
	10%	10.69%	10.75% (5)	8.45%	297.5	10.49%	10.54% (5)	6.90%	474.9	10.45%	10.50% (5)	6.77%	534.3	10.41%	10.46%	6.69%	11.7
5	0%	14.65%	15.01% (5)	4.43%	86.2	14.00%	14.33% (5)	3.37%	98.0	14.00%	14.34% (5)	3.36%	109.5	14.04%	14.37%	3.33%	2.8
	3%	13.85%	14.18% (5)	3.12%	153.9	14.65%	15.00% (5)	3.40%	242.8	14.66%	15.01% (5)	3.42%	272.2	14.75%	15.10%	3.51%	3.8
	5%	13.96%	14.31% (5)	3.21%	265.8	14.69%	15.06% (5)	3.37%	377.9	14.70%	15.07% (5)	3.39%	416.4	14.78%	15.15%	3.47%	2.9
	8%	14.34%	14.66% (5)	4.17%	302.8	14.72%	15.05% (5)	3.65%	428.8	14.72%	15.05% (5)	3.64%	465.9	14.74%	15.07%	3.64%	3.4
	10%	15.85%	16.24% (5)	5.43%	248.5	15.85%	16.23% (5)	4.14%	356.5	15.83%	16.20% (5)	3.91%	426.1	15.91%	16.29%	3.92%	14.7
10	0%	16.91%	17.87% (5)	3.20%	147.1	17.45%	18.43% (5)	2.54%	210.0	17.47%	18.45% (5)	2.54%	245.5	17.55%	18.54%	2.56%	3.7
	3%	17.07%	18.06% (5)	3.07%	198.0	17.94%	18.97% (5)	2.65%	279.5	17.99%	19.03% (5)	2.68%	320.5	18.16%	19.21%	2.78%	4.8
	5%	20.13%	21.40% (5)	4.21%	92.2	18.85%	20.02% (5)	2.74%	104.2	18.89%	20.06% (5)	2.72%	120.3	19.01%	20.19%	2.69%	2.7
	8%	19.73%	20.83% (5)	4.68%	90.9	18.38%	19.39% (5)	3.04%	103.7	18.42%	19.42% (5)	3.01%	120.0	18.52%	19.53%	2.96%	4.3
	10%	19.34%	20.48% (5)	5.02%	142.9	17.48%	18.48% (5)	3.33%	164.6	17.22%	18.20% (5)	2.87%	201.4	17.35%	18.33%	2.59%	24.6

Chapter 7

Conclusions

This thesis aimed to study the problem of high-dimensional low-rank and sparse matrix decomposition with fixed and sampled basis coefficients in both of the noiseless and noisy settings. For the noiseless case, we provided exact recovery guarantees via the well-accepted "nuclear norm plus ℓ_1-norm" approach, as long as certain standard identifiability conditions for the low-rank and sparse components are assumed to be satisfied. These probabilistic recovery results are significant in the high-dimensional regime since they reveal that only a vanishingly small fraction of observations is already sufficient as the intrinsic dimension increases. Although the involved analysis followed from the existing framework of dual certification, such recovery guarantees can still be regarded as the noiseless counterparts of those in the noisy setting. For the noisy case, enlightened by the successful recent development on the adaptive nuclear semi-norm penalization technique for noisy low-rank matrix completion [98, 99], we proposed a two-stage rank-sparsity-correction procedure and then examined its recovery performance by deriving a non-asymptotic probabilistic error bound under the high-dimensional scaling. This error bound, which has not been obtained until this research, suggests that the improvement on

recovery error could be expected. Finally, we specialized the aforementioned two-stage correction procedure to deal with the correlation matrix estimation problem with missing observations in strict factor models where the sparse component is diagonal. We pointed out that the specialized recovery error bound matches with the optimal one if the rank-correction function is constructed appropriately and the initial estimator is good enough. This fascinating finding as well as the convincing numerical results demonstrates the superiority of the two-stage correction approach over the nuclear norm penalization.

It should be noticed that the work done in this thesis is far from comprehensive. Below we briefly list some research directions that deserve further explorations.

- Is it possible to establish better exact recovery guarantees when applying the adaptive semi-norm techniques for the noisy case to the noiseless case?
- Regarding the non-asymptotic recovery error bound for the ANLPLS estimator provided in Theorem 5.7, the quantitative analysis on how to optimally choose the parameters κ_L and κ_S such that the constant parts of the error bound are minimized is of considerable importance in practice.
- From the computational point of view, how to efficiently tune the penalization parameters ρ_L and ρ_S for the ANLPLS estimator remains a big challenge.
- Whether or not there exist certain suitable forms of rank consistency, support consistiency, or both, for the ANLPLS estimator in the high-dimensional setting is still an open question.

Bibliography

[1] A. Agarwal, S. Negahban, and M. J. Wainwright, *Noisy matrix decomposition via convex relaxation: Optimal rates in high dimensions*, The Annals of Statistics, 40 (2012), pp. 1171–1197.

[2] R. Ahlswede and A. Winter, *Strong converse for identification via quantum channels*, IEEE Transactions on Information Theory, 48 (2002), pp. 569–579.

[3] L. Andersen, J. Sidenius, and S. Basu, *All your hedges in one basket*, Risk magazine, 16 (2003), pp. 67–72.

[4] A. Antoniadis and J. Fan, *Regularization of wavelet approximations (with discussion)*, Journal of the American Statistical Association, 96 (2001), pp. 939–967.

[5] J. Bai, *Inferential theory for factor models of large dimensions*, Econometrica, 71 (2003), pp. 135–171.

[6] J. Bai and K. Li, *Statistical analysis of factor models of high dimension*, The Annals of Statistics, 40 (2012), pp. 436–465.

[7] J. Bai and Y. Liao, *Efficient estimation of approximate factor models via regularized maximum likelihood.* Arxiv preprint arXiv:1209.5911, 2012.

[8] J. Bai and S. Shi, *Estimating high dimensional covariance matrices and its applications*, Annals of Economics and Finance, 12 (2011), pp. 199–215.

[9] J. Barzilai and J. M. Borwein, *Two-point step size gradient methods*, IMA Journal of Numerical Analysis, 8 (1988), pp. 141–148.

[10] A. Belloni and V. Chernozhukov, *Least squares after model selection in high-dimensional sparse models*, Bernoulli, 19 (2013), pp. 521–547.

[11] S. N. Bernstein, *Theory of Probability*, Moscow, 1927.

[12] D. P. Bertsekas, *On the Goldstein-Levitin-Polyak gradient projection method*, IEEE Transactions on Automatic Control, 21 (1976), pp. 174–184.

[13] P. J. Bickel, Y. Ritov, and A. B. Tsybakov, *Simultaneous analysis of Lasso and Dantzig selector*, The Annals of Statistics, 37 (2009), pp. 1705–1732.

[14] E. G. Birgin, J. M. Martínez, and M. Raydan, *Nonmonotone spectral projected gradient methods on convex sets*, SIAM Journal on Optimization, 10 (2000), pp. 1196–1211.

[15] ——, *Spectral projected gradient methods*, in Encyclopedia of Optimization, Springer, 2009, pp. 3652–3659.

[16] R. Borsdorf, N. J. Higham, and M. Raydan, *Computing a nearest correlation matrix with factor structure*, SIAM Journal on Matrix Analysis and Applications, 31 (2010), pp. 2603–2622.

[17] P. Bühlmann and S. A. Van de Geer, *Statistics for High-Dimensional Data: Methods, Theory and Applications*, Springer Series in Statistics, Springer-Verlag, 2011.

[18] V. V. Buldygin and Y. V. Kozachenko, *Metric Characterization of Random Variables and Random Processes*, vol. 188 of Translations of Mathematical Monographs, American Mathematical Society, Providence, RI, 2000.

[19] F. Bunea, A. B. Tsybakov, and M. H. Wegkamp, *Aggregation for Gaussian regression*, The Annals of Statistics, 35 (2007), pp. 1674–1697.

[20] ——, *Sparsity oracle inequalities for the lasso*, Electronic Journal of Statistics, 1 (2007), pp. 169–194.

[21] E. J. Candès, X. Li, Y. Ma, and J. Wright, *Robust principal component analysis?*, Journal of the ACM, 58 (2011), pp. 11:1–11:37.

[22] E. J. Candès and Y. Plan, *Near-ideal model selection by ℓ_1 minimization*, The Annals of Statistics, 37 (2009), pp. 2145–2177.

[23] ——, *Matrix completion with noise*, Proceedings of the IEEE, 98 (2010), pp. 925–936.

[24] ——, *Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements*, EEE Transactions on Information Theory, 57 (2011), pp. 2342–2359.

[25] E. J. Candès and B. Recht, *Exact matrix completion via convex optimization*, Foundations of Computational mathematics, 9 (2009), pp. 717–772.

[26] E. J. Candès, J. Romberg, and T. Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Transactions on Information Theory, 52 (2006), pp. 489–509.

[27] E. J. Candès and T. Tao, *Decoding by linear programming*, IEEE Transactions on Information Theory, 51 (2005), pp. 4203–4215.

[28] ——, *The power of convex relaxation: Near-optimal matrix completion*, IEEE Transactions on Information Theory, 56 (2010), pp. 2053–2080.

[29] G. Chamberlain, *Funds, factors, and diversification in arbitrage pricing models*, Econometrica, (1983), pp. 1305–1323.

[30] G. Chamberlain and M. Rothschild, *Arbitrage, factor structure, and mean-variance analysis on large asset markets*, Econometrica, 51 (1983), pp. 1281–1304.

[31] V. Chandrasekaran, P. A. Parrilo, and A. S. Willsky, *Latent variable graphical model selection via convex optimization*, The Annals of Statistics, 40 (2012), pp. 1935–1967.

[32] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky, *Rank-sparsity incoherence for matrix decomposition*, SIAM Journal on Optimization, 21 (2011), pp. 572–596.

[33] Y. Chen, A. Jalali, S. Sanghavi, and C. Caramanis, *Low-rank matrix recovery from errors and erasures*, IEEE Transactions on Information Theory, 59 (2013), pp. 4324–4337.

[34] I. Choi, *Efficient estimation of factor models*, Econometric Theory, 28 (2012), pp. 274–308.

[35] J. de Leeuw, *Applications of convex analysis to multidimensional scaling*, in Recent Developments in Statistics (European Meeting of Statisticians, Grenoble, 1976), North-Holland, Amsterdam, 1977, pp. 133–145.

[36] J. de Leeuw and W. J. Heiser, *Convergence of correction matrix algorithms for multidimensional scaling*, in Geometric Representations of Relational Data, Mathesis Press, Ann Arbor, MI, 1977, pp. 735–752.

[37] F. Deutsch, *The angle between subspaces of a Hilbert space*, in Approximation theory, wavelets and applications, Kluwer Academic Publishers, 1995, pp. 107–130.

[38] ——, *Best Approximation in Inner Product Spaces*, vol. 7 of CMS Books in Mathematics, Springer-Verlag, New York, 2001.

[39] F. X. Diebold and M. Nerlove, *The dynamics of exchange rate volatility: a multivariate latent factor ARCH model*, Journal of Applied Econometrics, 4 (1989), pp. 1–21.

[40] C. Ding, *An introduction to a class of matrix optimization problems*, PhD thesis, Department of Mathematics, National University of Singapore, 2012. Available at http://www.math.nus.edu.sg/~matsundf/DingChao_Thesis_final.pdf.

[41] C. Ding, D. Sun, J. Sun, and K.-C. Toh, *Spectral operators of matrices*. Arxiv preprint arXiv:1401.2269, 2014.

[42] D. L. Donoho, *Compressed sensing*, IEEE Transactions on Information Theory, 52 (2006), pp. 1289–1306.

[43] ——, *For most large underdetermined systems of linear equations the minimal ℓ_1-norm solution is also the sparsest solution*, Communications on Pure and Applied Mathematics, 59 (2006), pp. 797–829.

[44] C. Dossal, M.-L. Chabanol, G. Peyré, and J. Fadili, *Sharp support recovery from noisy random measurements by ℓ_1-minimization*, Applied and Computational Harmonic Analysis, 33 (2012), pp. 24–43.

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

[46] R. Engle and M. Watson, *A one-factor multivariate time series model of metropolitan wage rates*, Journal of the American Statistical Association, 76 (1981), pp. 774–781.

[47] E. F. Fama and K. R. French, *The cross-section of expected stock returns*, The Journal of Finance, 47 (1992), pp. 427–465.

[48] ——, *Common risk factors in the returns on stocks and bonds*, Journal of Financial Economics, 33 (1993), pp. 3–56.

[49] J. Fan, *Comments on "Wavelets in statistics: A review" by A. Antoniadis*, Journal of the Italian Statistical Society, 6 (1997), pp. 131–138.

[50] J. Fan, Y. Fan, and J. Lv, *High dimensional covariance matrix estimation using a factor model*, Journal of Econometrics, 147 (2008), pp. 186–197.

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

[52] ——, *Statistical challenges with high dimensionality: feature selection in knowledge discovery*, in Proceedings of the International Congress of Mathematicians: Madrid, August 22–30, Invited Lectures, 2006, pp. 595–622.

[53] J. Fan, Y. Liao, and M. Mincheva, *High dimensional covariance matrix estimation in approximate factor models*, The Annals of statistics, 39 (2011), pp. 3320–3356.

[54] ——, *Large covariance estimation by thresholding principal orthogonal complements (with discussion)*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 75 (2013), pp. 603–680.

[55] J. Fan and J. Lv, *Nonconcave penalized likelihood with NP-dimensionality*, IEEE Transactions on Information Theory, 57 (2011), pp. 5467–5484.

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

[57] M. Fazel, *Matrix rank minimization with applications*, PhD thesis, Department of Electrical Engineering, Stanford University, 2002.

[58] M. Fazel, T. K. Pong, D. Sun, and P. Tseng, *Hankel matrix rank minimization with applications in system identification and realization*, SIAM Journal on Matrix Analysis and Applications, 34 (2013), pp. 946–977.

[59] J.-J. Fuchs, *On sparse representations in arbitrary redundant bases*, IEEE Transactions on Information Theory, 50 (2004), pp. 1341–1344.

[60] D. Gabay and B. Mercier, *A dual algorithm for the solution of nonlinear variational problems via finite element approximation*, Computers & Mathematics with Applications, 2 (1976), pp. 17–40.

[61] A. Ganesh, J. Wright, X. Li, E. J. Candès, and Y. Ma, *Dense error correction for low-rank matrices via principal component pursuit*, in International Symposium on Information Theory Proceedings, IEEE, 2010, pp. 1513–1517.

[62] D. J. H. Garling, *Inequalities: A Journey into Linear Analysis*, Cambridge University Press, Cambridge, 2007.

[63] R. Glowinski and A. Marrocco, *Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité, d'une classe de problèmes de Dirichlet non linéaires*, 1975.

[64] S. Golden, *Lower bounds for the helmholtz function*, Physical Review, 137 (1965), pp. B1127–B1128.

[65] A. A. Goldstein, *Convex programming in Hilbert space*, Bulletin of the American Mathematical Society, 70 (1964), pp. 709–710.

[66] Y. Gordon, *Some inequalities for Gaussian processes and applications*, Israel Journal of Mathematics, 50 (1985), pp. 265–289.

[67] L. Grippo, F. Lampariello, and S. Lucidi, *A nonmonotone line search technique for Newton's method*, SIAM Journal on Numerical Analysis, 23 (1986), pp. 707–716.

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

[69] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, *Quantum state tomography via compressed sensing*, Physical Review Letters, 105 (2010), pp. 150401:1–150401:4.

[70] D. Gross and V. Nesme, *Note on sampling without replacing from a finite collection of matrices*. Arxiv preprint arXiv:1001.2738, 2010.

[71] T. Hagerup and C. Rüb, *A guided tour of Chernoff bounds*, Information Processing Letters, 33 (1990), pp. 305–308.

[72] W. J. Heiser, *Convergent computation by iterative majorization: Theory and applications in multidimensional data analysis*, in Recent advances in descriptive multivariate analysis (Exeter, 1992/1993), vol. 2 of Royal Statistical Society Lecture Note Series, Oxford University Press, New York, 1995, pp. 157–189.

[73] D. Hsu, S. M. Kakade, and T. Zhang, *Robust matrix decomposition with sparse corruptions*, IEEE Transactions on Information Theory, 57 (2011), pp. 7221–7234.

[74] ——, *A tail inequality for quadratic forms of subgaussian random vectors*, Electronic Communications in Probability, 17 (2012), pp. 52:1–52:6.

[75] D. R. Hunter and K. Lange, *A tutorial on MM algorithms*, The American Statistician, 58 (2004), pp. 30–37.

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

[77] ——, *Matrix completion from noisy entries*, Journal of Machine Learning Research, 99 (2010), pp. 2057–2078.

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

[79] ——, *Noisy low-rank matrix completion with general sampling distribution*, Bernoulli, 20 (2014), pp. 282–303.

[80] V. Koltchinskii, *Sparsity in penalized empirical risk minimization*, Annales de l'Institut Henri Poincaré – Probabilités et Statistiques, 45 (2009), pp. 7–57.

[81] ——, *Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems*, École d'Été de Probabilités de Saint-Flour XXXVIII-2008, Springer-Verlag, Heidelberg, 2011.

[82] ——, *Von neumann entropy penalization and low-rank matrix estimation*, The Annals of Statistics, 39 (2011), pp. 2936–2973.

[83] 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 (2011), pp. 2302–2329.

[84] K. Lange, D. R. Hunter, and I. Yang, *Optimization transfer using surrogate objective functions (with discussion)*, Journal of computational and graphical statistics, 9 (2000), pp. 1–59.

[85] B. Laurent and P. Massart, *Adaptive estimation of a quadratic functional by model selection*, The Annals of Statistics, 28 (2000), pp. 1302–1338.

[86] D. N. Lawley and A. E. Maxwell, *Factor analysis as a statistical method*, Elsevier, New York, second ed., 1971.

[87] M. Ledoux and M. Talagrand, *Probability in Banach Spaces: Isoperimetry and Processes*, vol. 23 of Ergebnisse der Mathematik und ihrer Grenzgebiete (3), Springer-Verlag, Berlin, 1991.

[88] E. S. Levitin and B. T. Polyak, *Constrained minimization methods*, USSR Computational Mathematics and Mathematical Physics, 6 (1966), pp. 1–50.

[89] X. Li, *Compressed sensing and matrix completion with constant proportion of corruptions*, Constructive Approximation, 37 (2013), pp. 73–99.

[90] X. Luo, *Recovering model structures from large low rank and sparse covariance matrix estimation*. Arxiv preprint arXiv:1111.1133, 2013.

[91] J. Lv and Y. Fan, *A unified approach to model selection and sparse recovery using regularized least squares*, The Annals of Statistics, 37 (2009), pp. 3498–3528.

[92] P. Massart, *About the constants in Talagrand's concentration inequalities for empirical processes*, The Annals of Probability, 28 (2000), pp. 863–884.

[93] A. E. Maxwell, *Factor analysis*, in Encyclopedia of Statistical Sciences (electronic), John Wiley & Sons, New York, 2006.

[94] N. Meinshausen, *Relaxed Lasso*, Computational Statistics & Data Analysis, 52 (2007), pp. 374–393.

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

[96] N. Meinshausen and B. Yu, *Lasso-type recovery of sparse representations for high-dimensional data*, The Annals of Statistics, (2009), pp. 246–270.

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

[98] W. Miao, *Matrix completion models with fixed basis coefficients and rank regularized prbolems with hard constraints*, PhD thesis, Department of Mathematics, National University of Singapore, 2013. Available at http://www.math.nus.edu.sg/~matsundf/PhDThesis_Miao_Final.pdf.

[99] W. Miao, S. Pan, and D. Sun, *A rank-corrected procedure for matrix completion with fixed basis coefficients*. Arxiv preprint arXiv:1210.3709, 2014.

[100] S. Negahban and M. J. Wainwright, *Estimation of (near) low-rank matrices with noise and high-dimensional scaling*, The Annals of Statistics, 39 (2011), pp. 1069–1097.

[101] ——, *Restricted strong convexity and weighted matrix completion: Optimal bounds with noise*, Journal of Machine Learning Research, 13 (2012), pp. 1665–1697.

[102] S. N. Negahban, P. Ravikumar, M. J. Wainwright, and B. Yu, *A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers*, Statistical Science, 27 (2012), pp. 538–557.

[103] G. Raskutti, M. J. Wainwright, and B. Yu, *Restricted eigenvalue properties for correlated gaussian designs*, Journal of Machine Learning Research, 11 (2010), pp. 2241–2259.

[104] B. Recht, *A simpler approach to matrix completion*, Journal of Machine Learning Research, 12 (2011), pp. 3413–3430.

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

[106] B. Recht, W. Xu, and B. Hassibi, *Null space conditions and thresholds for rank minimization*, Mathematical programming, 127 (2011), pp. 175–202.

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

[108] S. A. Ross, *The arbitrage theory of capital asset pricing*, Journal of Economic Theory, 13 (1976), pp. 341–360.

[109] ——, *The capital asset pricing model CAPM, short-sale restrictions and related issues*, The Journal of Finance, 32 (1977), pp. 177–183.

[110] M. Rudelson and S. Zhou, *Reconstruction from anisotropic random measurements*, IEEE Transactions on Information Theory, 59 (2013), pp. 3434–3447.

[111] R. Salakhutdinov and N. Srebro, *Collaborative filtering in a non-uniform world: Learning with the weighted trace norm*, in Advances in Neural Information Processing Systems 23, 2010, pp. 2056–2064.

[112] J. Saunderson, V. Chandrasekaran, P. A. Parrilo, and A. S. Willsky, *Diagonal and low-rank matrix decompositions, correlation matrices, and ellipsoid fitting*, SIAM Journal on Matrix Analysis and Applications, 33 (2012), pp. 1395–1416.

[113] C. J. Thompson, *Inequality with applications in statistical mechanics*, Journal of Mathematical Physics, 6 (1965), pp. 1812–1823.

[114] R. Tibshirani, *Regression shrinkage and selection via the Lasso*, Journal of the Royal Statistical Society: Series B (Methodological), (1996), pp. 267–288.

[115] J. A. Tropp, *User-friendly tail bounds for sums of random matrices*, Foundations of Computational Mathematics, 12 (2012), pp. 389–434.

[116] S. A. van de Geer and P. Bühlmann, *On the conditions used to prove oracle results for the Lasso*, Electronic Journal of Statistics, 3 (2009), pp. 1360–1392.

[117] S. A. van de Geer, P. Bühlmann, and S. Zhou, *The adaptive and the thresholded Lasso for potentially misspecified models (and a lower bound for the Lasso)*, Electronic Journal of Statistics, 5 (2011), pp. 688–749.

[118] A. W. van der Vaart and J. A. Wellner, *Weak Convergence and Empirical Processes: With Applications to Statistics*, Springer Series in Statistics, Springer-Verlag, New York, 1996.

[119] R. Vershynin, *A note on sums of independent random matrices after Ahlswede-Winter*. Preprint available at http://www-personal.umich.edu/~romanv/teaching/reading-group/ahlswede-winter.pdf, 2009.

[120] ——, *Introduction to the non-asymptotic analysis of random matrices*. Arxiv preprint arXiv:1011.3027, 2011.

[121] M. J. Wainwright, *Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (Lasso)*, IEEE Transactions on Information Theory, 55 (2009), pp. 2183–2202.

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

[123] R. Werner and K. Schöttle, *Calibration of correlation matrices – SDP or not SDP*, 2007.

[124] J. Wright, A. Ganesh, K. Min, and Y. Ma, *Compressive principal component pursuit*, Information and Inference: A Journal of the IMA, 2 (2013), pp. 32–68.

[125] T. T. Wu and K. Lange, *The MM alternative to EM*, Statistical Science, 25 (2010), pp. 492–505.

[126] M. Yuan and Y. Lin, *On the non-negative garrotte estimator*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 69 (2007), pp. 143–161.

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

[128] C.-H. Zhang and J. Huang, *The sparsity and bias of the Lasso selection in high-dimensional linear regression*, The Annals of Statistics, 36 (2008), pp. 1567–1594.

[129] C.-H. Zhang and T. Zhang, *A general theory of concave regularization for high-dimensional sparse estimation problems*, Statistical Science, 27 (2012), pp. 576–593.

[130] T. Zhang, *Some sharp performance bounds for least squares regression with L_1 regularization*, The Annals of Statistics, 37 (2009), pp. 2109–2144.

[131] ——, *Analysis of multi-stage convex relaxation for sparse regularization*, Journal of Machine Learning Research, 11 (2010), pp. 1081–1107.

[132] ——, *Multi-stage convex relaxation for feature selection*, Bernoulli, 19 (2013), pp. 2277–2293.

[133] P. Zhao and B. Yu, *On model selection consistency of Lasso*, Journal of Machine Learning Research, 7 (2006), pp. 2541–2563.

[134] S. Zhou, *Thresholding procedures for high dimensional variable selection and statistical estimation*, in Advances in Neural Information Processing Systems 22, 2009, pp. 2304–2312.

[135] Z. Zhou, X. Li, J. Wright, E. J. Candès, and Y. Ma, *Stable principal component pursuit*, in International Symposium on Information Theory Proceedings, IEEE, 2010, pp. 1518–1522.

[136] H. Zou, *The adaptive Lasso and its oracle properties*, Journal of the American statistical association, 101 (2006), pp. 1418–1429.

[137] H. Zou and R. Li, *One-step sparse estimates in nonconcave penalized likelihood models*, The Annals of Statistics, 36 (2008), pp. 1509–1533.

Name: Wu Bin
Degree: Doctor of Philosophy
Department: Mathematics
Thesis Title: High-Dimensional Analysis on Matrix Decomposition with Application to Correlation Matrix Estimation in Factor Models

Abstract

In this thesis, we conduct high-dimensional analysis on the problem of low-rank and sparse matrix decomposition with fixed and sampled basis coefficients. This problem is strongly motivated by high-dimensional correlation matrix estimation coming from a factor model used in economic and financial studies, in which the underlying correlation matrix is assumed to be the sum of a low-rank matrix and a sparse matrix respectively due to the common factors and the idiosyncratic components. For the noiseless version, we provide exact recovery guarantees if certain identifiability conditions for the low-rank and sparse components are satisfied. These probabilistic recovery results are in accordance with the high-dimensional setting because only a vanishingly small fraction of samples is required. For the noisy version, inspired by the successful recent development on the adaptive nuclear semi-norm penalization technique, we propose a two-stage rank-sparsity-correction procedure and examine its recovery performance by establishing a novel non-asymptotic probabilistic error bound under the high-dimensional scaling. We then specialize this two-stage correction procedure to deal with the correlation matrix estimation problem with missing observations in strict factor models where the sparse component is diagonal. In this application, the specialized recovery error bound and the convincing numerical results validate the superiority of the proposed approach.

HIGH-DIMENSIONAL ANALYSIS ON MATRIX DECOMPOSITION WITH APPLICATION TO CORRELATION MATRIX ESTIMATION IN FACTOR MODELS

WU BIN

NATIONAL UNIVERSITY OF SINGAPORE

2014