AIMS Mathematics

http://www.aimspress.com/journal/Math

AIMS Mathematics, 8(9): 21439–21462.
DOI: 10.3934/math.20231092
Received: 22 March 2023
Revised: 26 June 2023
Accepted: 28 June 2023
Published: 05 July 2023

Research article

Pilot estimators for a kind of sparse covariance matrices with incomplete heavy-tailed data

Huimin Li and Jinru Wang*

Department of Mathematics, Beijing University of Technology, Beijing 100124, China

* **Correspondence:** Email: wangjinru@bjut.edu.cn.

Abstract: This paper investigates generalized pilot estimators of covariance matrix in the presence of missing data. When the random samples have only bounded fourth moment, two kinds of generalized pilot estimators are provided, the generalized Huber estimator and the generalized truncated mean estimator. In addition, we construct thresholding generalized pilot estimator for a kind of sparse covariance matrices and establish the convergence rates in terms of probability under spectral and Frobenius norms respectively. Moreover, the convergence rates in sense of expectation are also given under an extra condition. Finally, simulation studies are conducted to demonstrate the superiority of our method.

Keywords: sparse covariance matrix; incomplete data; generalized pilot estimator; convergence rate
Mathematics Subject Classification: 62H12, 62J10

1. Introduction

Let $\mathbf{X}$ be a p-dimensional random vector. Estimating its covariance matrix $\mathbf{\Sigma} = (\sigma_{uv})_{p\times p}$ is of interest in high-dimensional statistics (Mendelson and Zhivotovskiy [1], Dendramis et al. [2] and Zhang et al. [3]). Until now, a commonly adopted strategy for evaluating the covariance matrix has been to impose sparse structure on itself (Belomestny [4], Kang and Deng [5], Bettache et al. [6] and Liang et al. [7]).

If $\mathbf{X}$ is sub-Gaussian, Bickel and Levina [8], Cai and Liu [9] and Cai and Zhou [10] considered all the rows or columns of the covariance matrix belonging to l_q-ball, weighted l_q-ball or weak l_q-ball as a kind of sparse assumption. Moreover, they proposed the corresponding thresholding estimators and established the convergence rates in sense of probability or expectation respectively.

When each component of $\mathbf{X} = (X_1, \cdots, X_p)^T$ is subject to heavy-tailed distribution, i.e., the distribution of X_u satisfies $\int_{\mathbb{R}} e^{tx} dF_u(x) = \infty$ for $t > 0$, Avella-Medina et al. [11] introduced a pilot estimator $\tilde{\mathbf{\Sigma}} = (\tilde{\sigma}_{uv})_{p\times p}$ satisfying

$$\mathbb{P}\left\{|\tilde{\sigma}_{uv} - \sigma_{uv}| \geq C_0 \sqrt{(\log p)/n},\ \exists 1 \leq u, v \leq p\right\} \leq \varepsilon_{n,p}$$

for positive constant C_0 and $\log p = o(n)$. Where $\varepsilon_{n,p}$ is a deterministic positive sequence and satisfies $\lim_{n,p\to\infty} \varepsilon_{n,p} = 0$. Avella-Medina et al. pointed out the sample covariance matrix

$$\hat{\mathbf{\Sigma}} = \frac{1}{n}\sum_{k=1}^{n}(\mathbf{X}_k - \bar{\mathbf{X}})(\mathbf{X}_k - \bar{\mathbf{X}})^T \text{ with } \bar{\mathbf{X}} = \frac{1}{n}\sum_{k=1}^{n}\mathbf{X}_k$$

must be a pilot estimator if $\mathbf{X}_1, \cdots, \mathbf{X}_n$ are i.i.d. sub-Gaussian random samples. In addition, some other pilot estimators were provided under bounded fourth moment assumption. The authors also considered convergence rate of the thresholding pilot estimator in terms of probability when the rows or columns of the covariance matrix are in weighted l_q-ball.

However, missing data (also called incomplete data) always occurs in high-dimensional sampling setting, see Hawkins et al. [12], Lounici [13] and Loh and Wain-wright [14]. Instead of obtaining whole i.i.d. samples $\mathbf{X}_1, \cdots, \mathbf{X}_n$, one can only collect some parts of them. Let the vector $\mathbf{S}_i \in \{0, 1\}^p$ $(i = 1, \cdots, n)$ denote by

$$S_{iu} = \begin{cases} 1, & \text{if } X_{iu} \text{ is observed;} \\ 0, & \text{if } X_{iu} \text{ is missing,} \end{cases}$$

where X_{iu} and S_{iu} are the u-th coordinate of the $\mathbf{X}_i$ and $\mathbf{S}_i$ respectively. This paper denotes the samples with missing values by $\mathbf{X}_i^* = (X_{i1}^*, \cdots, X_{ip}^*)^T$ where $X_{iu}^* = X_{iu}S_{iu}$. The following missing mechanism introduced by Cai and Zhang [15] is adopted.

Assumption 1.1 (Missing completely at random)**.** $\mathbf{S} = \{\mathbf{S}_1, \cdots, \mathbf{S}_n\}$ *can be either deterministic or random and is independent of* $\mathbf{X} = \{\mathbf{X}_1, \cdots, \mathbf{X}_n\}$.

Define

$$n_{uv}^* = \sum_{i=1}^{n} S_{iu}S_{iv},$$

i.e., n_{uv}^* is the number of the u-th and v-th entries of $\mathbf{X}_i^*$ being both observed. For convenience, let

$$n_u^* = n_{uu}^*,\ n_{\min}^* = \min_{u,v} n_{uv}^*.$$

Then, it is easy to see

$$n_{\min}^* \leq n_{uv}^* \leq \min\{n_u^*, n_v^*\} \leq n.$$

Meanwhile, the generalized sample mean $\bar{\mathbf{X}}^* = (\bar{X}_u^*)_{1\leq u\leq p}$ is defined by

$$\bar{X}_u^* = \frac{1}{n_u^*}\sum_{i=1}^{n} X_{iu}S_{iu}$$

and the generalized sample covariance matrix $\hat{\mathbf{\Sigma}}^* = (\hat{\sigma}_{uv}^*)$ is given by

$$\hat{\sigma}_{uv}^* = \frac{1}{n_{uv}^*}\sum_{i=1}^{n}(X_{iu} - \bar{X}_u^*)(X_{iv} - \bar{X}_v^*)S_{iu}S_{iv}. \tag{1.1}$$

Our goal is to construct the thresholding estimator of the sparse covariance matrix $\mathbf{\Sigma}$ based on incomplete heavy-tailed data. Furthermore, the convergence rates of the thresholding estimator are investigated in terms of probability and expectation respectively.

The rest of paper is organized as follows. Section 2 introduces the definition of generalized pilot estimator based on missing data. Then under bounded fourth moment assumption, two kinds of generalized pilot estimators are given. In Section 3, we construct the thresholding generalized pilot estimator and explore its convergence rates in sense of probability under spectral and Frobenius norms respectively. In Section 4, the convergence rates are given in terms of expectation under an extra mild condition. Then, Section 5 investigates the numerical performances of the thresholding generalized Huber pilot estimator and thresholding generalized truncated mean pilot estimator respectively and compares these two estimators with the adaptive thresholding estimator proposed by Cai and Zhang [15].

2. Generalized pilot estimator

Definition 2.1. *Any symmetric matrix* $\tilde{\mathbf{\Sigma}}^* = (\tilde{\sigma}^*_{uv})_{p\times p}$ *based on incomplete data* $\mathbf{X}^*_1, \cdots, \mathbf{X}^*_n$ *is said to be a generalized pilot estimator of* $\mathbf{\Sigma}$*, if for* $L > 0$ *there exists constant* $C_0(L)$ *such that*

$$\mathbb{P}\left\{|\tilde{\sigma}^*_{uv} - \sigma_{uv}| \geq C_0(L)\sqrt{(\log p)/n^*_{uv}},\ \exists 1 \leq u, v \leq p\right\} = O(p^{-L}) \tag{2.1}$$

holds with $\log p = o(n^*_{\min})$.

Remark 2.1. *If one can obtain complete data, the generalized pilot estimator defined by (2.1) coincides with the pilot estimator proposed by Avella-Medina et al. [11] except* $\varepsilon_{n,p}$ *is replaced by* $O(p^{-L})$.

Remark 2.2. *If* $\mathbf{X}$ *is a sub-Gaussian random vector and the items* σ_{uu} $(u = 1, \cdots p)$ *of* $\mathbf{\Sigma}$ *are uniformly bounded, the generalized sample covariance matrix* $\hat{\mathbf{\Sigma}}^*$ *given by (1.1) must be a generalized pilot estimator of* $\mathbf{\Sigma}$.

In fact, Theorem 3.1 in [15] tells that for any $0 < x \leq 1$ *there exists constants* $C,\ c > 0$ *such that*

$$\mathbb{P}\left\{|\hat{\sigma}^*_{uv} - \sigma_{uv}| \geq x\sqrt{\sigma_{uu}\sigma_{vv}}\right\} \leq C\exp\left(-cn^*_{uv}x^2\right). \tag{2.2}$$

By $n^*_{\min} \leq n^*_{uv}$ *and* $\log p = o(n^*_{\min})$*, one knows* $\log p = o(n^*_{uv})$*. If* $x = \sqrt{(2+L)\log p/(cn^*_{uv})}$ *with* $L > 0$*, (2.2) reduces to*

$$\mathbb{P}\left\{|\hat{\sigma}^*_{uv} - \sigma_{uv}| \geq C_0(L)\sqrt{(\log p)/n^*_{uv}}\right\} \leq Cp^{-(L+2)}$$

with $C_0(L) = \sqrt{(2+L)\sigma_{uu}\sigma_{vv}/c}$.

Furthermore,

$$\mathbb{P}\left\{|\hat{\sigma}^*_{uv} - \sigma_{uv}| \geq C_0(L)\sqrt{(\log p)/n^*_{uv}},\ \exists 1 \leq u, v \leq p\right\} \leq Cp^{-L}.$$

Therefore, $\hat{\mathbf{\Sigma}}^*$ *given by (1.1) is a generalized pilot estimator of* $\mathbf{\Sigma}$.

We introduce the following theorem in order to provide two other kinds of generalized pilot estimators under bounded fourth moment assumption.

Theorem 2.1. *Suppose* $\max_{1\le u\le p}\mathbb{E}|X_u|^4 \le k^4$, $\log p = o(n^*_{\min})$, $\mathbb{E}X_u = \mu_u$, $\mathbb{E}(X_uX_v) = \mu_{uv}$, *and Assumption 1.1 holds. If* $\tilde{\mu}^*_u$ *and* $\tilde{\mu}^*_{uv}$ *satisfy*

$$(i)\ \mathbb{P}\left\{|\tilde{\mu}^*_u - \mu_u| > ck\sqrt{((2+L)\log p)/n^*_u}\right\} = O(p^{-(2+L)}); \tag{2.3}$$

$$(ii)\ \mathbb{P}\left\{|\tilde{\mu}^*_{uv} - \mu_{uv}| > ck^2\sqrt{((2+L)\log p)/n^*_{uv}}\right\} = O(p^{-(2+L)}) \tag{2.4}$$

with absolute constants $L > 0$ *and* $c \ge 1$ *then* $\tilde{\mathbf{\Sigma}}^* = (\tilde{\sigma}^*_{uv})_{p\times p} := (\tilde{\mu}^*_{uv} - \tilde{\mu}^*_u\tilde{\mu}^*_v)_{p\times p}$ *must be a generalized pilot estimator of* $\mathbf{\Sigma}$.

Proof. Let $K := ck\sqrt{2+L}$. Thus,

$$\mathbb{P}\left\{|\tilde{\mu}^*_u - \mu_u| \ge K\sqrt{(\log p)/n^*_u},\ \exists 1 \le u \le p\right\} = O(p^{-(1+L)}) \tag{2.5}$$

thanks to condition (2.3). Moreover,

$$\mathbb{P}\left\{|(\tilde{\mu}^*_u - \mu_u)(\tilde{\mu}^*_v - \mu_v)| \ge K^2(\log p)/\sqrt{n^*_u n^*_v},\ \exists 1 \le u, v \le p\right\} = O(p^{-(1+L)}). \tag{2.6}$$

Similarly, one derives

$$\mathbb{P}\left\{|\tilde{\mu}^*_{uv} - \mu_{uv}| \ge K^2\sqrt{(\log p)/n^*_{uv}},\ \exists 1 \le u, v \le p\right\} = O(p^{-L}) \tag{2.7}$$

due to (2.4) and $c \ge 1$.

By $\max_{1\le u\le p}\mathbb{E}|X_u|^4 \le k^4$, one obtains $|\mu_u| \le (\mathbb{E}|X_u|^4)^{1/4} \le k$ $(u = 1, \cdots, p)$ and

$$\begin{aligned}|\tilde{\mu}^*_u\tilde{\mu}^*_v - \mu_u\mu_v| &\le |\mu_v(\tilde{\mu}^*_u - \mu_u)| + |\mu_u(\tilde{\mu}^*_v - \mu_v)| + |(\tilde{\mu}^*_u - \mu_u)(\tilde{\mu}^*_v - \mu_v)|\\ &\le k(|\tilde{\mu}^*_u - \mu_u| + |\tilde{\mu}^*_v - \mu_v|) + |(\tilde{\mu}^*_u - \mu_u)(\tilde{\mu}^*_v - \mu_v)|\\ &\le c^{-1}K(|\tilde{\mu}^*_u - \mu_u| + |\tilde{\mu}^*_v - \mu_v|) + |(\tilde{\mu}^*_u - \mu_u)(\tilde{\mu}^*_v - \mu_v)|,\end{aligned}$$

the above last inequality follows from $K \ge ck$. Thus, one concludes

$$\begin{aligned}&\mathbb{P}\left\{|\tilde{\mu}^*_u\tilde{\mu}^*_v - \mu_u\mu_v| > c^{-1}K^2\left(\sqrt{(\log p)/n^*_u} + \sqrt{(\log p)/n^*_v}\right) + K^2(\log p)/\sqrt{n^*_u n^*_v},\ \exists 1 \le u, v \le p\right\}\\ &\quad = O(p^{-(1+L)})\end{aligned}$$

thanks to (2.5) and (2.6).

Since $n^*_{uv} \le \min\{n^*_u, n^*_v\}$, the above result reduces to

$$\mathbb{P}\left\{|\tilde{\mu}^*_u\tilde{\mu}^*_v - \mu_u\mu_v| > 2c^{-1}K^2\sqrt{(\log p)/n^*_{uv}} + K^2(\log p)/n^*_{uv},\ \exists 1 \le u, v \le p\right\} = O(p^{-(1+L)}).$$

Furthermore, according to $\log p = o(n^*_{\min})$ and $n^*_{\min} \le n^*_{uv}$, one knows $\log p = o(n^*_{uv})$. Therefore, $(\log p)/n^*_{uv} \le ((\log p)/n^*_{uv})^{1/2}$ and

$$\mathbb{P}\left\{|\tilde{\mu}^*_u\tilde{\mu}^*_v - \mu_u\mu_v| > (2c^{-1} + 1)K^2\sqrt{(\log p)/n^*_{uv}},\ \exists 1 \le u, v \le p\right\} = O(p^{-(1+L)}) \tag{2.8}$$

hold.

Note that

$$|\tilde{\sigma}^*_{uv} - \sigma_{uv}| = |(\tilde{\mu}^*_{uv} - \tilde{\mu}^*_u\tilde{\mu}^*_v) - (\mu_{uv} - \mu_u\mu_v)| \leq |\tilde{\mu}^*_{uv} - \mu_{uv}| + |\tilde{\mu}^*_u\tilde{\mu}^*_v - \mu_u\mu_v|.$$

Then,

$$\mathbb{P}\left\{|\tilde{\sigma}^*_{uv} - \sigma_{uv}| > 2(c^{-1}+1)K^2\sqrt{(\log p)/n^*_{uv}},\ \exists 1 \leq u, v \leq p\right\} = O(p^{-L})$$

follows from (2.7) and (2.8).

Hence, $\tilde{\mathbf{\Sigma}}^*$ is a generalized pilot estimator of $\mathbf{\Sigma}$ with $C_0(L) = 2ck^2(1+c)(2+L)$. □

We shall give two generalized pilot estimators based on incomplete heavy-tailed samples.

Denote Huber function by

$$\psi_\alpha(x) = \alpha\psi\left(\frac{x}{\alpha}\right),$$

where $\alpha > 0$ and $\psi(x) = \begin{cases} x, & |x| \leq 1, \\ \text{sign } x, & |x| > 1 \end{cases}$. For any constant $L > 0$, let $(\tilde{\mu}^*_H)_u$ $(u = 1, \cdots, p)$ satisfy

$$\sum_{i=1}^{n} \psi_{\alpha_u}(X^*_{iu} - (\tilde{\mu}^*_H)_u)S_{iu} = 0 \tag{2.9}$$

with $\alpha_u := \sqrt{n^*_u\zeta^2/(2+L)\log p}$ and $\zeta \geq \sqrt{DX_u}$. Similarly, $(\tilde{\mu}^*_H)_{uv}$ $(u, v = 1, \cdots, p)$ satisfies

$$\sum_{i=1}^{n} \psi_{\alpha_{uv}}(X^*_{iu}X^*_{iv} - (\tilde{\mu}^*_H)_{uv})S_{iu}S_{iv} = 0 \tag{2.10}$$

with $\alpha_{uv} := \sqrt{n^*_{uv}\zeta_1^2/(2+L)\log p}$ and $\zeta_1 \geq \sqrt{D(X_uX_v)}$. Then, we have the following estimator.

Example 2.1 (Generalized Huber estimator)**.** *Suppose conditions of Theorem 2.1 hold, then* $\tilde{\mathbf{\Sigma}}^*_H := \left((\tilde{\mu}^*_H)_{uv} - (\tilde{\mu}^*_H)_u(\tilde{\mu}^*_H)_v\right)_{p\times p}$ *is a generalized pilot estimator of* $\mathbf{\Sigma}$*, where* $(\tilde{\mu}^*_H)_j$ $(j = u, v)$ *and* $(\tilde{\mu}^*_H)_{uv}$ *are defined by (2.9) and (2.10).*

Proof. With the definition of X^*_{iu}, (2.9) is equivalent to

$$\sum_{i\in A_u} \psi_{\alpha_u}(X_{iu} - (\tilde{\mu}^*_H)_u) = 0, \tag{2.11}$$

where $A_u = \{i : S_{iu} \neq 0\}$. Obviously, $|A_u| = \sum_{i=1}^n S_{iu}$. By the definition of n^*_u, we have $|A_u| = n^*_u$.

Similarly, we find (2.10) is equivalent to

$$\sum_{i\in A_{uv}} \psi_{\alpha_{uv}}(X_{iu}X_{iv} - (\tilde{\mu}^*_H)_{uv}) = 0 \tag{2.12}$$

with $A_{uv} = \{i : S_{iu}S_{iv} \neq 0\}$ and $|A_{uv}| = n^*_{uv}$.

By $\max_{1\leq u\leq p} \mathbb{E}|X_u|^4 \leq k^4$, we get

$$DX_u \leq \mathbb{E}|X_u|^2 \leq (\mathbb{E}|X_u|^4)^{1/2} \leq k^2.$$

On the other hand,

$$D(X_uX_v) \leq \mathbb{E}|X_uX_v|^2 \leq (\mathbb{E}|X_u|^4\mathbb{E}|X_v|^4)^{1/2} \leq k^4$$

due to Cauchy-Schwarz inequality. Thus,

$$\alpha_u = \sqrt{n_u^* k^2/(2+L)\log p},\ \alpha_{uv} = \sqrt{n_{uv}^* k^4/(2+L)\log p}. \tag{2.13}$$

Obviously, it holds

$$\frac{(2+L)\log p}{n_u^*} \leq \frac{(2+L)\log p}{n_{\min}^*} < \frac{1}{8},\ \frac{(2+L)\log p}{n_{uv}^*} \leq \frac{(2+L)\log p}{n_{\min}^*} < \frac{1}{8}$$

thanks to $n_{\min}^* \leq n_{uv}^* \leq n_u^*$ and $\log p = o(n_{\min}^*)$.

According to (2.11)–(2.13) and Theorem 5 in [16], we know that if $(2+L)\log p/n_u^* \leq 1/8$ and $(2+L)\log p/n_{uv}^* \leq 1/8$ then

$$\mathbb{P}\left\{|(\tilde{\mu}_H^*)_u - \mu_u| > 4k\sqrt{((2+L)\log p)/n_u^*}\right\} = O(p^{-(2+L)}),$$
$$\mathbb{P}\left\{|(\tilde{\mu}_H^*)_{uv} - \mu_{uv}| > 4k^2\sqrt{((2+L)\log p)/n_{uv}^*}\right\} = O(p^{-(2+L)}),$$

i.e., $(\tilde{\mu}_H^*)_u$ and $(\tilde{\mu}_H^*)_{uv}$ reach the expected results (2.3) and (2.4). □

In order to give another generalized pilot estimator, let $(\tilde{\mu}_T^*)_u$ $(u = 1, \cdots, p)$, $(\tilde{\mu}_T^*)_{uv}$ $(u, v = 1, \cdots, p)$ be defined by

$$(\tilde{\mu}_T^*)_u := \frac{1}{n_u^*}\sum_{i=1}^{n} X_{iu}^* \mathbf{1}\left\{|X_{iu}^*| \leq \beta\sqrt{\frac{n_u^*}{(2+L)\log p}}\right\}, \tag{2.14}$$

$$(\tilde{\mu}_T^*)_{uv} := \frac{1}{n_{uv}^*}\sum_{i=1}^{n} X_{iu}^* X_{iv}^* \mathbf{1}\left\{|X_{iu}^* X_{iv}^*| \leq \beta_1\sqrt{\frac{n_{uv}^*}{(2+L)\log p}}\right\} \tag{2.15}$$

respectively where $L > 0, \beta \geq \sqrt{\mathbb{E}|X_u|^2}$ and $\beta_1 \geq \sqrt{\mathbb{E}|X_uX_v|^2}$. Then, we have the second estimator.

Example 2.2 (Generalized truncated mean estimator)**.** *Suppose conditions of Theorem 2.1 hold. Then,* $\tilde{\mathbf{\Sigma}}_T^* := \left((\tilde{\mu}_T^*)_{uv} - (\tilde{\mu}_T^*)_u(\tilde{\mu}_T^*)_v\right)_{p\times p}$ *is a generalized pilot estimator of* $\mathbf{\Sigma}$*, where* $(\tilde{\mu}_T^*)_j$ $(j = u, v)$ *and* $(\tilde{\mu}_T^*)_{uv}$ *are defined by (2.14) and (2.15).*

Proof. We first show $(\tilde{\mu}_T^*)_u$ satisfies (2.3). According to $\max\limits_{1\leq u\leq p}\mathbb{E}|X_u|^4 \leq k^4$, we have

$$\mathbb{E}|X_u|^2 \leq (\mathbb{E}|X_u|^4)^{1/2} \leq k^2.$$

So, (2.14) is equivalent to

$$(\tilde{\mu}_T^*)_u = \frac{1}{n_u^*}\sum_{i\in A_u} X_{iu}\mathbf{1}\left\{|X_{iu}| \leq k\sqrt{\frac{n_u^*}{(2+L)\log p}}\right\}$$

where $A_u = \{i : S_{iu} \neq 0\}$.

Let $a := k\sqrt{n_u^*/(2+L)\log p}$. We derive

$$\begin{aligned}|(\tilde{\mu}_T^*)_u - \mu_u| &= \left|\frac{1}{n_u^*}\sum_{i\in A_u} X_{iu}\mathbf{1}\{|X_{iu}|\le a\} - \frac{1}{n_u^*}\sum_{i\in A_u}\mathbb{E}X_u\right| \\ &= \left|\frac{1}{n_u^*}\sum_{i\in A_u}\Big(X_{iu}\mathbf{1}\{|X_{iu}|\le a\} - \mathbb{E}(X_{iu}\mathbf{1}\{|X_{iu}|\le a\})\Big) - \frac{1}{n_u^*}\sum_{i\in A_u}\mathbb{E}(X_{iu}\mathbf{1}\{|X_{iu}|> a\})\right|.\end{aligned}$$

Therefore, upon combining $\mathbb{E}|X_u|^2 \le k^2$ and $|A_u| = n_u^*$

$$\begin{aligned}|(\tilde{\mu}_T^*)_u - \mu_u| &\le \left|\frac{1}{n_u^*}\sum_{i\in A_u}\Big(X_{iu}\mathbf{1}\{|X_{iu}|\le a\} - \mathbb{E}(X_{iu}\mathbf{1}\{|X_{iu}|\le a\})\Big)\right| + \left|\frac{1}{n_u^*}\sum_{i\in A_u}\frac{k^2}{a}\right| \\ &= \left|\frac{1}{n_u^*}\sum_{i\in A_u}\Big(X_{iu}\mathbf{1}\{|X_{iu}|\le a\} - \mathbb{E}(X_{iu}\mathbf{1}\{|X_{iu}|\le a\})\Big)\right| + \frac{k^2}{a}. \qquad (2.16)\end{aligned}$$

According to $\mathbb{E}(X_{iu}^2\mathbf{1}\{|X_{iu}|\le a\}) \le k^2$ and Bernstein's inequality in [17],

$$\mathbb{P}\left\{\left|\frac{1}{n_u^*}\sum_{i\in A_u}\Big(X_{iu}\mathbf{1}\{|X_{iu}|\le a\} - \mathbb{E}(X_{iu}\mathbf{1}\{|X_{iu}|\le a\})\Big)\right| \le k\sqrt{\frac{2t}{n_u^*}} + \frac{at}{3n_u^*}\right\} \ge 1 - 2\exp(-t) \qquad (2.17)$$

for any $t > 0$.

By (2.16) and (2.17) and taking $t = (2+L)\log p$, we have

$$\mathbb{P}\left\{|(\tilde{\mu}_T^*)_u - \mu_u| \le k\sqrt{\frac{2(2+L)\log p}{n_u^*}} + \frac{a(2+L)\log p}{3n_u^*} + \frac{k^2}{a}\right\} \ge 1 - 2p^{-(2+L)}.$$

Substituting $a = k\sqrt{n_u^*/(2+L)\log p}$ into the above inequality, we know

$$\mathbb{P}\left\{|(\tilde{\mu}_T^*)_u - \mu_u| > 4k\sqrt{((2+L)\log p)/n_u^*}\right\} = O(p^{-(2+L)})$$

which is the expected condition (2.3) of Theorem 2.1.

Similarly, we can derive

$$\mathbb{P}\left\{|(\tilde{\mu}_T^*)_{uv} - \mu_{uv}| > 4k^2\sqrt{((2+L)\log p)/n_{uv}^*}\right\} = O(p^{-(2+L)})$$

i.e., the condition (2.4) of Theorem 2.1 holds. □

3. Convergence rates in terms of probability

We introduce the thresholding function and the space of sparse covariance matrices.

Definition 3.1. *For any constant $\lambda > 0$, a real valued function $\tau_\lambda(\cdot)$ is said to be thresholding function if*

(i) $\tau_\lambda(z) = 0,\ |z| \le \lambda$*;*
(ii) $|\tau_\lambda(z) - z| \le \lambda$*;*
(iii) $|\tau_\lambda(z)| \le c_0|y|$ *for* $|z-y| \le \lambda$ *and the constant* $c_0 > 0$.

In fact, many functions satisfy conditions (i)–(iii). For example, the soft thresholding function $\tau_\lambda(z) = \text{sign}(z)(|z| - \lambda)_+$, the adaptive lasso thresholding function $\tau_\lambda(z) = z(1 - |\lambda/z|^\eta)$ with $\eta \geq 1$ and the smoothly clipped absolute deviation thresholding rule proposed by Rothman et al. [18].

This paper considers the following class of covariance matrices introduced by [15]

$$\mathcal{H}(s_{n,p}) := \left\{ \mathbf{\Sigma} = (\sigma_{uv})_{p\times p} > 0 : \max_v \sum_{u=1}^{p} \min\left\{ \sqrt{\sigma_{uu}\sigma_{vv}}, \frac{|\sigma_{uv}|}{\sqrt{(\log p)/n}} \right\} \leq s_{n,p} \right\}.$$

Next, we define the thresholding generalized pilot estimator $(\tilde{\mathbf{\Sigma}}^*)^\tau = ((\tilde{\sigma}^*_{uv})^\tau)_{p\times p}$ and consider its convergence rates in terms of probability under spectral and Frobenius norms respectively over the parametric space $\mathcal{H}(s_{n,p})$.

Let $\tilde{\mathbf{\Sigma}}^* = (\tilde{\sigma}^*_{uv})$ be a generalized pilot estimator and define

$$(\tilde{\sigma}^*_{uv})^\tau := \tau_{\lambda_{uv}}(\tilde{\sigma}^*_{uv}), \tag{3.1}$$

where $\tau_{\lambda_{uv}}(\cdot)$ is the thresholding function with

$$\lambda_{uv} = \delta \sqrt{\frac{\tilde{\sigma}^*_{uu}\tilde{\sigma}^*_{vv} \log p}{n^*_{uv}}}. \tag{3.2}$$

The constant δ will be specified in the proving process of Lemma 3.1.

The following lemma is useful for inferring Theorem 3.1 and Theorem 4.1.

Lemma 3.1. *Suppose* $\min_u \sigma_{uu} \geq \gamma > 0$, $\log p = o(n^*_{\min})$ *and Assumption 1.1 hold. Denote the events* $Q_1,\ Q_2$ *as*

$$Q_1 := \{|\tilde{\sigma}^*_{uv} - \sigma_{uv}| \leq \lambda_{uv},\ \forall 1 \leq u, v \leq p\}, \tag{3.3}$$

$$Q_2 := \{\tilde{\sigma}^*_{uu}\tilde{\sigma}^*_{vv} \leq 2\sigma_{uu}\sigma_{vv},\ \forall 1 \leq u, v \leq p\}. \tag{3.4}$$

Then, for any $L > 0$

(i) there exists $C_1(L) > 0$ *such that*

$$|\tau_{\lambda_{uv}}(\tilde{\sigma}^*_{uv}) - \sigma_{uv}| \leq C_1(L) \sqrt{\frac{\log p}{n^*_{\min}}} \min\left\{ \sqrt{\sigma_{uu}\sigma_{vv}}, \frac{|\sigma_{uv}|}{\sqrt{(\log p)/n}} \right\},\ \forall 1 \leq u, v \leq p$$

holds under the event $Q_1 \cap Q_2$.

(ii) $\mathbb{P}(Q_1 \cap Q_2) \geq 1 - O(p^{-L})$.

Proof. (i) Under the event Q_1, one knows

$$|\tau_{\lambda_{uv}}(\tilde{\sigma}^*_{uv}) - \sigma_{uv}| \leq c_0|\sigma_{uv}|, \tag{3.5}$$

$$|\tau_{\lambda_{uv}}(\tilde{\sigma}^*_{uv}) - \sigma_{uv}| \leq |\tau_{\lambda_{uv}}(\tilde{\sigma}^*_{uv}) - \tilde{\sigma}^*_{uv}| + |\tilde{\sigma}^*_{uv} - \sigma_{uv}| \leq 2\lambda_{uv} \tag{3.6}$$

thanks to conditions (ii) and (iii) of Definition 3.1.

Define

$$\delta := \frac{\sqrt{2}C_0(L)}{\gamma} \tag{3.7}$$

where $C_0(L)$ is given in Definition 2.1. By (3.2) when the event Q_2 happens as well (3.6) reduces to

$$|\tau_{\lambda_{uv}}(\tilde{\sigma}^*_{uv}) - \sigma_{uv}| \le 2\delta \sqrt{\frac{\tilde{\sigma}^*_{uu}\tilde{\sigma}^*_{vv}\log p}{n^*_{uv}}} \le C(L)\sqrt{\frac{\sigma_{uu}\sigma_{vv}\log p}{n^*_{uv}}}. \tag{3.8}$$

According to (3.5) and (3.8), one obtains

$$\begin{aligned}|\tau_{\lambda_{uv}}(\tilde{\sigma}^*_{uv}) - \sigma_{uv}| &\le \min\left\{c_0|\sigma_{uv}|, C(L)\sqrt{\frac{\sigma_{uu}\sigma_{vv}\log p}{n^*_{uv}}}\right\} \\ &\le C_1(L)\sqrt{\frac{\log p}{n^*_{uv}}}\min\left\{\sqrt{\sigma_{uu}\sigma_{vv}}, \frac{|\sigma_{uv}|}{\sqrt{(\log p)/n^*_{uv}}}\right\}\end{aligned}$$

under the event $Q_1 \cap Q_2$. Therefore,

$$|\tau_{\lambda_{uv}}(\tilde{\sigma}^*_{uv}) - \sigma_{uv}| \le C_1(L)\sqrt{\frac{\log p}{n^*_{\min}}}\min\left\{\sqrt{\sigma_{uu}\sigma_{vv}}, \frac{|\sigma_{uv}|}{\sqrt{(\log p)/n}}\right\}$$

holds due to $n^*_{\min} \le n^*_{uv} \le n$. This reaches the conclusion (i) of Lemma 3.1.

(ii) In order to show $\mathbb{P}(Q_1 \cap Q_2) \ge 1 - O(p^{-L})$, one first estimates $\mathbb{P}(Q_2^c)$.

Clearly,

$$\tilde{\sigma}^*_{uu}\tilde{\sigma}^*_{vv} - \sigma_{uu}\sigma_{vv} = (\tilde{\sigma}^*_{uu} - \sigma_{uu})\tilde{\sigma}^*_{vv} + (\tilde{\sigma}^*_{vv} - \sigma_{vv})\tilde{\sigma}^*_{uu} - (\tilde{\sigma}^*_{uu} - \sigma_{uu})(\tilde{\sigma}^*_{vv} - \sigma_{vv})$$

and

$$\tilde{\sigma}^*_{uu}\tilde{\sigma}^*_{vv} \le \sigma_{uu}\sigma_{vv} + |\tilde{\sigma}^*_{uu} - \sigma_{uu}| \mid \tilde{\sigma}^*_{vv} \mid + |\tilde{\sigma}^*_{vv} - \sigma_{vv}||\tilde{\sigma}^*_{uu}| + |\tilde{\sigma}^*_{uu} - \sigma_{uu}||\tilde{\sigma}^*_{vv} - \sigma_{vv}|.$$

Define the event

$$E := \left\{|\tilde{\sigma}^*_{uv} - \sigma_{uv}| \le C_0(L)\sqrt{(\log p)/n^*_{uv}},\ \forall 1 \le u, v \le p\right\}.$$

Since $\tilde{\mathbf{\Sigma}}^* = (\tilde{\sigma}^*_{uv})_{p\times p}$ is a generalized pilot estimator of $\mathbf{\Sigma} = (\sigma_{uv})_{p\times p}$ then one gets

$$\mathbb{P}(E) = 1 - O(p^{-L}).$$

By $\log p = o(n^*_{\min})$ and $n^*_{\min} \le n^*_{uv}$, one knows $\log p = o(n^*_{uv})$. Furthermore, it holds

$$\tilde{\sigma}^*_{uu}\tilde{\sigma}^*_{vv} \le \sigma_{uu}\sigma_{vv} + \frac{\gamma}{4}(\sigma_{vv} + \frac{\gamma}{2}) + \frac{\gamma}{4}(\sigma_{uu} + \frac{\gamma}{2}) + \frac{\gamma^2}{4} \le \sigma_{uu}\sigma_{vv} + \frac{\sigma_{uu}\sigma_{vv}}{2} + \frac{\gamma^2}{2} \le 2\sigma_{uu}\sigma_{vv}$$

under the event E because of $\min_u \sigma_{uu} \ge \gamma$. Hence,

$$\mathbb{P}(Q_2^c) \le \mathbb{P}(E^c) = O(p^{-L}). \tag{3.9}$$

Next to estimate $\mathbb{P}(Q_1^c)$. One observes $\tilde{\sigma}^*_{uu}\tilde{\sigma}^*_{vv} \ge \sigma_{uu}\sigma_{vv} - |\tilde{\sigma}^*_{uu} - \sigma_{uu}||\tilde{\sigma}^*_{vv}| - |\tilde{\sigma}^*_{vv} - \sigma_{vv}||\tilde{\sigma}^*_{uu}| - |\tilde{\sigma}^*_{uu} - \sigma_{uu}||\tilde{\sigma}^*_{vv} - \sigma_{vv}|$ and it follows that

$$\tilde{\sigma}^*_{uu}\tilde{\sigma}^*_{vv} \ge \sigma_{uu}\sigma_{vv} - \frac{\gamma}{8}(\sigma_{vv} + \frac{\gamma}{2}) - \frac{\gamma}{8}(\sigma_{uu} + \frac{\gamma}{2}) - \frac{\gamma^2}{8} \ge \frac{3}{4}\sigma_{uu}\sigma_{vv} - \frac{\gamma^2}{4} \ge \frac{\gamma^2}{2}$$

holds true on the event E due to $\min_u \sigma_{uu} \geq \gamma$ and $\log p = o(n^*_{uv})$. Hence,

$$\mathbb{P}\left\{\tilde{\sigma}^*_{uu}\tilde{\sigma}^*_{vv} \geq \frac{\gamma^2}{2},\ \forall 1 \leq u, v \leq p\right\} \geq \mathbb{P}(E) = 1 - O(p^{-L}). \tag{3.10}$$

Let $\lambda_{uv} = \delta\sqrt{\tilde{\sigma}^*_{uu}\tilde{\sigma}^*_{vv}\log p / n^*_{uv}}$ given by (3.2). It can be shown that

$$\begin{aligned}\mathbb{P}(Q_1^c) &= \mathbb{P}\left\{\frac{|\tilde{\sigma}^*_{uv} - \sigma_{uv}|}{\sqrt{\tilde{\sigma}^*_{uu}\tilde{\sigma}^*_{vv}}} > \delta\sqrt{\frac{\log p}{n^*_{uv}}},\ \exists 1 \leq u, v \leq p\right\} \\ &\leq \mathbb{P}\left\{|\tilde{\sigma}^*_{uv} - \sigma_{uv}| > \delta\sqrt{\frac{\gamma^2 \log p}{2n^*_{uv}}},\ \exists 1 \leq u, v \leq p\right\} + O(p^{-L}) \end{aligned} \tag{3.11}$$

follows from (3.10).

Note that $\tilde{\mathbf{\Sigma}}^* = (\tilde{\sigma}^*_{uv})$ is a generalized pilot estimator of $\mathbf{\Sigma}$. Then, one derives

$$\mathbb{P}\left\{|\tilde{\sigma}^*_{uv} - \sigma_{uv}| > \delta\sqrt{\frac{\gamma^2 \log p}{2n^*_{uv}}},\ \exists 1 \leq u, v \leq p\right\} = O(p^{-L})$$

thanks to $\delta = \sqrt{2}C_0(L)/\gamma$ defined in (3.7). Substituting the above result into (3.11) gives

$$\mathbb{P}(Q_1^c) = O(p^{-L}).$$

Combining this with (3.9), one obtains the stated result

$$\mathbb{P}(Q_1 \cap Q_2) \geq 1 - \mathbb{P}(Q_1^c) - \mathbb{P}(Q_2^c) \geq 1 - O(p^{-L}). \qquad \square$$

Finally, we give the upper bounds of $\|(\tilde{\mathbf{\Sigma}}^*)^\tau - \mathbf{\Sigma}\|_{2,F}$ in terms of probability and $\|\mathbf{A}\|_{2,F}$ denotes the spectral and Frobenius norms of matrix $\mathbf{A}$ respectively.

Theorem 3.1. *Suppose* $\min_u \sigma_{uu} \geq \gamma > 0$, $\log p = o(n^*_{\min})$ *and Assumption 1.1 hold. Then,*

$$(i)\ \inf_{\mathbf{\Sigma}\in\mathcal{H}(s_{n,p})} \mathbb{P}\left\{\|(\tilde{\mathbf{\Sigma}}^*)^\tau - \mathbf{\Sigma}\|_2 \leq C_1(L)s_{n,p}\sqrt{\frac{\log p}{n^*_{\min}}}\right\} \geq 1 - O\left(p^{-L}\right);$$

$$(ii)\ \inf_{\mathbf{\Sigma}\in\mathcal{H}(s_{n,p})} \mathbb{P}\left\{\frac{1}{\sqrt{p}}\|(\tilde{\mathbf{\Sigma}}^*)^\tau - \mathbf{\Sigma}\|_F \leq C_1(L)s_{n,p}\sqrt{\frac{\log p}{n^*_{\min}}}\right\} \geq 1 - O\left(p^{-L}\right);$$

$$(iii)\ \inf_{\substack{\mathbf{\Sigma}\in\mathcal{H}(s_{n,p}) \\ \max_u \sigma_{uu} \leq M}} \mathbb{P}\left\{\frac{1}{\sqrt{p}}\|(\tilde{\mathbf{\Sigma}}^*)^\tau - \mathbf{\Sigma}\|_F \leq C_1(L)\sqrt{M}\sqrt{\frac{s_{n,p}\log p}{n^*_{\min}}}\right\} \geq 1 - O\left(p^{-L}\right).$$

Proof. (i) Define the event $Q := Q_1 \cap Q_2$, where $Q_1,\ Q_2$ are given by (3.3) and (3.4) respectively. Then, it is easy to see

$$\|(\tilde{\mathbf{\Sigma}}^*)^\tau - \mathbf{\Sigma}\|_1 := \max_v \sum_{u=1}^{p} |\tau_{\lambda_{uv}}(\tilde{\sigma}^*_{uv}) - \sigma_{uv}| \leq C_1(L)s_{n,p}\sqrt{\frac{\log p}{n^*_{\min}}} \tag{3.12}$$

thanks to Lemma 3.1 and $\mathbf{\Sigma} \in \mathcal{H}(s_{n,p})$.

Gersgorin theorem tells $\|(\tilde{\mathbf{\Sigma}}^*)^\tau - \mathbf{\Sigma}\|_2 \leq \|(\tilde{\mathbf{\Sigma}}^*)^\tau - \mathbf{\Sigma}\|_1$ and this combining (3.12) implies

$$\|(\tilde{\mathbf{\Sigma}}^*)^\tau - \mathbf{\Sigma}\|_2 \leq C_1(L) s_{n,p} \sqrt{\frac{\log p}{n^*_{\min}}}$$

on the event Q.

On the other hand, Lemma 3.1 tells $\mathbb{P}(Q) \geq 1 - O(p^{-L})$. Hence, Theorem 3.1(i) holds.

(ii) One observes

$$\frac{1}{p}\|(\tilde{\mathbf{\Sigma}}^*)^\tau - \Sigma\|_F^2 = \frac{1}{p}\sum_{v=1}^{p}\sum_{u=1}^{p} |\tau_{\lambda_{uv}}(\tilde{\sigma}^*_{uv}) - \sigma_{uv}|^2 \leq \max_v \sum_{u=1}^{p} |\tau_{\lambda_{uv}}(\tilde{\sigma}^*_{uv}) - \sigma_{uv}|^2$$

and it follows

$$\frac{1}{p}\|(\tilde{\mathbf{\Sigma}}^*)^\tau - \Sigma\|_F^2 \leq (C_1(L))^2 \frac{\log p}{n^*_{\min}} \max_v \sum_{u=1}^{p} \left(\min\left\{\sqrt{\sigma_{uu}\sigma_{vv}}, \frac{|\sigma_{uv}|}{\sqrt{(\log p)/n}}\right\}\right)^2 \tag{3.13}$$

on the event Q according to Lemma 3.1.

Note that $\max_v \sum_{u=1}^{p} |a_{uv}|^2 \leq (\max_v \sum_{u=1}^{p} |a_{uv}|)^2$. Then, (3.13) reduces to

$$\begin{aligned}\frac{1}{p}\|(\tilde{\mathbf{\Sigma}}^*)^\tau - \Sigma\|_F^2 &\leq (C_1(L))^2 \frac{\log p}{n^*_{\min}} \left(\max_v \sum_{u=1}^{p} \min\left\{\sqrt{\sigma_{uu}\sigma_{vv}}, \frac{|\sigma_{uv}|}{\sqrt{(\log p)/n}}\right\}\right)^2 \\ &\leq (C_1(L))^2 s_{n,p}^2 \frac{\log p}{n^*_{\min}}\end{aligned}$$

as long as $\mathbf{\Sigma} \in \mathcal{H}(s_{n,p})$.

Therefore, conclusion (ii) reaches since Lemma 3.1 says $\mathbb{P}(Q) \geq 1 - O(p^{-L})$.

(iii) By $\max_u \sigma_{uu} \leq M$, one knows

$$\min\left\{\sqrt{\sigma_{uu}\sigma_{vv}}, \frac{|\sigma_{uv}|}{\sqrt{(\log p)/n}}\right\} \leq M.$$

Furthermore, it holds

$$\begin{aligned}\frac{1}{p}\|(\tilde{\mathbf{\Sigma}}^*)^\tau - \Sigma\|_F^2 &\leq M(C_1(L))^2 \frac{\log p}{n^*_{\min}} \max_v \sum_{u=1}^{p} \min\left\{\sqrt{\sigma_{uu}\sigma_{vv}}, \frac{|\sigma_{uv}|}{\sqrt{(\log p)/n}}\right\} \\ &\leq M(C_1(L))^2 s_{n,p} \frac{\log p}{n^*_{\min}}\end{aligned}$$

under the event Q due to (3.13) and $\mathbf{\Sigma} \in \mathcal{H}(s_{n,p})$.

Thus, the claim (iii) follows from Lemma 3.1 immediately. □

Remark 3.1. *Theorem 3.1(i) generalizes the result of [15] which requires* $\mathbf{X}$ *is the sub-Gaussian random vector. In addition, if* $n^*_{\min} = n$*, Theorem 3.1(i) yields the result of [11] thanks to the parametric class* $\mathcal{H}(s_{n,p})$ *containing the class of sparse covariance matrices defined in [11].*

Remark 3.2. *From the proving process of Theorem 3.1(i), we find that*

$$\|(\tilde{\mathbf{\Sigma}}^*)^\tau - \mathbf{\Sigma}\|_1 \leq C_1(L) s_{n,p} \sqrt{\frac{\log p}{n^*_{\min}}}$$

under the event Q.

Furthermore, let $\|\mathbf{A}\|_\omega$ *denote the matrix* l_ω*-operator norm of* $\mathbf{A}$*, Lemma 7.2 in [19] tells*

$$\|\mathbf{A}\|_\omega \leq \|\mathbf{A}\|_1 (1 \leq \omega \leq \infty)$$

for any symmetric matrix $\mathbf{A}$*. Hence,*

$$\|(\tilde{\mathbf{\Sigma}}^*)^\tau - \mathbf{\Sigma}\|_\omega \leq C_1(L) s_{n,p} \sqrt{\frac{\log p}{n^*_{\min}}}$$

holds under the event Q.

Then, using Lemma 3.1 indicates

$$\inf_{\mathbf{\Sigma} \in \mathcal{H}(s_{n,p})} \mathbb{P}\left\{ \|(\tilde{\mathbf{\Sigma}}^*)^\tau - \mathbf{\Sigma}\|_\omega \leq C_1(L) s_{n,p} \sqrt{\frac{\log p}{n^*_{\min}}} \right\} \geq 1 - O\left(p^{-L}\right).$$

4. Convergence rates in terms of expectation

This section studies the convergence rates of the thresholding generalized pilot estimator $(\tilde{\mathbf{\Sigma}}^*)^\tau$ in terms of expectation over $\mathcal{H}(s_{n,p})$.

We introduce the following technical lemma.

Lemma 4.1. *Let* $\min_u \sigma_{uu} \geq \gamma > 0$, $\log p = o(n^*_{\min})$, $p \geq (n^*_{\min})^\xi (\xi > 0)$, $\mathbb{E}|\tilde{\sigma}^*_{uv} - \sigma_{uv}|^2 \leq M$ *and Assumption 1.1 holds. Then,*

$$(i) \sup_{\mathbf{\Sigma} \in \mathcal{H}(s_{n,p})} \int_{(Q_1 \cap Q_2)^c} \max_v \sum_{u=1}^p \min\left\{ \sqrt{\sigma_{uu}\sigma_{vv}}, \frac{|\sigma_{uv}|}{\sqrt{(\log p)/n}} \right\} dP \lesssim s_{n,p} \sqrt{\frac{\log p}{n^*_{\min}}};$$

$$(ii) \sup_{\mathbf{\Sigma} \in \mathcal{H}(s_{n,p})} \int_{(Q_1 \cap Q_2)^c} \left(\max_v \sum_{u=1}^p \min\left\{ \sqrt{\sigma_{uu}\sigma_{vv}}, \frac{|\sigma_{uv}|}{\sqrt{(\log p)/n}} \right\} \right)^{\frac{1}{2}} dP \lesssim \sqrt{\frac{s_{n,p} \log p}{n^*_{\min}}};$$

$$(iii) \sup_{\mathbf{\Sigma} \in \mathcal{H}(s_{n,p})} \int_{(Q_1 \cap Q_2)^c} \max_v \sum_{u=1}^p |\tilde{\sigma}^*_{uv} - \sigma_{uv}| dP \lesssim \sqrt{\frac{\log p}{n^*_{\min}}}.$$

Where Q_1, Q_2 *are defined by (3.3) and (3.4).* $x \lesssim y$ *denotes* $x \leq cy$ *with a absolute constant* $c > 0$.

Proof. Denote $Q := Q_1 \cap Q_2$ and

$$I_{n,p} := \sup_{\mathbf{\Sigma} \in \mathcal{H}(s_{n,p})} \int_{Q^c} \max_v \sum_{u=1}^p \min\left\{ \sqrt{\sigma_{uu}\sigma_{vv}}, \frac{|\sigma_{uv}|}{\sqrt{(\log p)/n}} \right\} dP.$$

Then, $I_{n,p} \leq s_{n,p}\mathbb{P}(Q^c)$.

According to Lemma 3.1, one knows $\mathbb{P}(Q^c) \leq O(p^{-L})$ and $I_{n,p} \lesssim s_{n,p}p^{-L}$. Taking $L = \xi^{-1} + 3 > 0$, one obtains

$$p^{-L} \leq (n^*_{\min})^{-L\xi} \leq (n^*_{\min})^{-1} \leq \sqrt{(\log p)/n^*_{\min}} \tag{4.1}$$

due to $p \geq (n_{\min})^{\xi}$. Hence, it follows that

$$I_{n,p} \lesssim s_{n,p}\sqrt{\frac{\log p}{n^*_{\min}}},$$

which is the desired conclusion (i).

Similarly, the definition of $\mathcal{H}(s_{n,p})$ and Lemma 3.1 imply

$$\sup_{\Sigma\in\mathcal{H}(s_{n,p})}\int_{Q^c}\left(\max_v \sum_{u=1}^{p} \min\left\{\sqrt{\sigma_{uu}\sigma_{vv}}, \frac{|\sigma_{uv}|}{\sqrt{(\log p)/n}}\right\}\right)^{1/2} dP \lesssim \sqrt{s_{n,p}}\; p^{-L}.$$

Moreover, the above result combining (4.1) concludes (ii).

To show (iii), Hölder inequality tells

$$\begin{aligned}\sup_{\Sigma\in\mathcal{H}(s_{n,p})}\int_{Q^c}\max_v\sum_{u=1}^{p}|\tilde{\sigma}^*_{uv}-\sigma_{uv}|dP &= \sup_{\Sigma\in\mathcal{H}(s_{n,p})}\mathbb{E}\left\{\left(\max_v\sum_{u=1}^{p}|\tilde{\sigma}^*_{uv}-\sigma_{uv}|\right)I(Q^c)\right\}\\ &\leq \sup_{\Sigma\in\mathcal{H}(s_{n,p})}\left\{\mathbb{E}\left(\max_v\sum_{u=1}^{p}|\tilde{\sigma}^*_{uv}-\sigma_{uv}|\right)^2\right\}^{1/2}\{\mathbb{P}(Q^c)\}^{1/2}.\end{aligned}$$

On the other hand, it holds

$$\left(\max_v\sum_{u=1}^{p}|\tilde{\sigma}^*_{uv}-\sigma_{uv}|\right)^2 \leq p\sum_{u,v=1}^{p}|\tilde{\sigma}^*_{uv}-\sigma_{uv}|^2.$$

Furthermore, one obtains

$$\mathbb{E}\left(\max_v\sum_{u=1}^{p}|\tilde{\sigma}^*_{uv}-\sigma_{uv}|\right)^2 \lesssim p^3$$

due to the given condition $\mathbb{E}|\tilde{\sigma}^*_{uv}-\sigma_{uv}|^2 \leq M$. Hence,

$$\sup_{\Sigma\in\mathcal{H}(s_{n,p})}\int_{Q^c}\max_v\sum_{u=1}^{p}|\tilde{\sigma}^*_{uv}-\sigma_{uv}|dP \lesssim p^{\frac{3}{2}}\{\mathbb{P}(Q^c)\}^{\frac{1}{2}} \lesssim p^{\frac{3-L}{2}} \tag{4.2}$$

follows from $\mathbb{P}(Q^c) \leq O(p^{-L})$.

By $L = \xi^{-1} + 3$ and the assumptions $p \geq (n^*_{\min})^{\xi}$, one finds

$$p^{\frac{3-L}{2}} = p^{-\frac{1}{2\xi}} \leq (n^*_{\min})^{-\frac{1}{2}} \leq \sqrt{\frac{\log p}{n^*_{\min}}}.$$

Substituting this into (4.2) implies the expected result (iii) holding. □

Theorem 4.1. *Let* $(\tilde{\mathbf{\Sigma}}^*)^\tau = ((\tilde{\sigma}^*_{uv})^\tau)_{p\times p}$ *given by (3.1),* $\mathbb{E}|\tilde{\sigma}^*_{uv} - \sigma_{uv}|^2 \le M$, $\min_u \sigma_{uu} \ge \gamma > 0$ *and Assumption 1.1 holds. If* $\log p = o(n^*_{\min})$, $p \ge (n^*_{\min})^\xi (\xi > 0)$ *and* $s_{n,p} \gtrsim 1$ *then*

$$(i) \sup_{\mathbf{\Sigma}\in\mathcal{H}(s_{n,p})} \mathbb{E}\|(\tilde{\mathbf{\Sigma}}^*)^\tau - \mathbf{\Sigma}\|_2 \lesssim s_{n,p}\sqrt{\frac{\log p}{n^*_{\min}}};$$

$$(ii) \sup_{\mathbf{\Sigma}\in\mathcal{H}(s_{n,p})} \mathbb{E}\frac{1}{\sqrt{p}}\|(\tilde{\mathbf{\Sigma}}^*)^\tau - \mathbf{\Sigma}\|_F \lesssim s_{n,p}\sqrt{\frac{\log p}{n^*_{\min}}};$$

$$(iii) \sup_{\substack{\mathbf{\Sigma}\in\mathcal{H}(s_{n,p})\\ \max_u \sigma_{uu}\le M}} \mathbb{E}\frac{1}{\sqrt{p}}\|(\tilde{\mathbf{\Sigma}}^*)^\tau - \mathbf{\Sigma}\|_F \lesssim \sqrt{s_{n,p}}\sqrt{\frac{\log p}{n^*_{\min}}}.$$

Proof. (i) Let the event $Q := Q_1 \cap Q_2$ where Q_1, Q_2 are given by (3.3) and (3.4). Then, by Gersgorin theorem $\|(\tilde{\mathbf{\Sigma}}^*)^\tau - \mathbf{\Sigma}\|_2 \le \|(\tilde{\mathbf{\Sigma}}^*)^\tau - \mathbf{\Sigma}\|_1$ we have

$$\sup_{\mathbf{\Sigma}\in\mathcal{H}(s_{n,p})} \mathbb{E}\|(\tilde{\mathbf{\Sigma}}^*)^\tau - \mathbf{\Sigma}\|_2 \le \sup_{\mathbf{\Sigma}\in\mathcal{H}(s_{n,p})} \int_Q \|(\tilde{\mathbf{\Sigma}}^*)^\tau - \mathbf{\Sigma}\|_1 dP + \sup_{\mathbf{\Sigma}\in\mathcal{H}(s_{n,p})} \int_{Q^c} \|(\tilde{\mathbf{\Sigma}}^*)^\tau - \mathbf{\Sigma}\|_1 dP.$$

Clearly, $\|(\tilde{\mathbf{\Sigma}}^*)^\tau - \mathbf{\Sigma}\|_1 := \max_v \sum_{u=1}^p |\tau_{\lambda_{uv}}(\tilde{\sigma}^*_{uv}) - \sigma_{uv}|$ and it follows

$$\|(\tilde{\mathbf{\Sigma}}^*)^\tau - \mathbf{\Sigma}\|_1 \lesssim s_{n,p}\sqrt{\frac{\log p}{n^*_{\min}}}$$

under the event Q thanks to Lemma 3.1 and the definition of $\mathcal{H}(s_{n,p})$. Hence,

$$\sup_{\mathbf{\Sigma}\in\mathcal{H}(s_{n,p})} \int_Q \|(\tilde{\mathbf{\Sigma}}^*)^\tau - \mathbf{\Sigma}\|_1 dP \lesssim s_{n,p}\sqrt{\frac{\log p}{n^*_{\min}}}.$$

Then, we just need to show

$$J_{n,p} := \sup_{\mathbf{\Sigma}\in\mathcal{H}(s_{n,p})} \int_{Q^c} \|(\tilde{\mathbf{\Sigma}}^*)^\tau - \mathbf{\Sigma}\|_1 dP \lesssim s_{n,p}\sqrt{\frac{\log p}{n^*_{\min}}} \tag{4.3}$$

for finishing the proof of (i).

According to condition (iii) of Definition 3.1, we obtain $|\tau_{\lambda_{uv}}(\tilde{\sigma}^*_{uv})| \le c_0|\tilde{\sigma}^*_{uv}|$ and

$$|\tau_{\lambda_{uv}}(\tilde{\sigma}^*_{uv}) - \sigma_{uv}| \le |\tau_{\lambda_{uv}}(\tilde{\sigma}^*_{uv})| + |\sigma_{uv}| \le c_0|\tilde{\sigma}^*_{uv}| + |\sigma_{uv}| \le c_0|\tilde{\sigma}^*_{uv} - \sigma_{uv}| + (c_0+1)|\sigma_{uv}|.$$

By $|\sigma_{uv}| \le \sqrt{\sigma_{uu}\sigma_{vv}}$ and $\log p = o(n^*_{\min})$, we know

$$|\sigma_{uv}| \le \min\left\{\sqrt{\sigma_{uu}\sigma_{vv}}, \frac{|\sigma_{uv}|}{\sqrt{(\log p)/n^*_{\min}}}\right\} \le \min\left\{\sqrt{\sigma_{uu}\sigma_{vv}}, \frac{|\sigma_{uv}|}{\sqrt{(\log p)/n}}\right\}$$

due to $n^*_{\min} \le n$. Hence, it holds

$$|\tau_{\lambda_{uv}}(\tilde{\sigma}^*_{uv}) - \sigma_{uv}| \le c_0|\tilde{\sigma}^*_{uv} - \sigma_{uv}| + (c_0+1)\min\left\{\sqrt{\sigma_{uu}\sigma_{vv}}, \frac{|\sigma_{uv}|}{\sqrt{(\log p)/n}}\right\} \tag{4.4}$$

and

$$J_{n,p} \lesssim \sup_{\Sigma\in\mathcal{H}(s_{n,p})} \int_{Q^c} \max_v \sum_{u=1}^{p} |\tilde{\sigma}^*_{uv} - \sigma_{uv}| dP$$
$$+ \sup_{\Sigma\in\mathcal{H}(s_{n,p})} \int_{Q^c} \max_v \sum_{u=1}^{p} \min\left\{ \sqrt{\sigma_{uu}\sigma_{vv}}, \frac{|\sigma_{uv}|}{\sqrt{(\log p)/n}} \right\} dP.$$

Therefore, (4.3) follows from Lemma 4.1(i), (iii) and $s_{n,p} \gtrsim 1$. This reaches (i).

(ii) To show (ii), we observe

$$\sup_{\Sigma\in\mathcal{H}(s_{n,p})} \mathbb{E}\frac{1}{\sqrt{p}}\|(\tilde{\mathbf{\Sigma}}^*)^\tau - \mathbf{\Sigma}\|_F \le \sup_{\Sigma\in\mathcal{H}(s_{n,p})} \int_Q \frac{1}{\sqrt{p}}\|(\tilde{\mathbf{\Sigma}}^*)^\tau - \mathbf{\Sigma}\|_F dP + \sup_{\Sigma\in\mathcal{H}(s_{n,p})} \int_{Q^c} \frac{1}{\sqrt{p}}\|(\tilde{\mathbf{\Sigma}}^*)^\tau - \mathbf{\Sigma}\|_F dP.$$

Clearly,

$$\frac{1}{p}\|(\tilde{\mathbf{\Sigma}}^*)^\tau - \Sigma\|_F^2 = \frac{1}{p}\sum_{v=1}^{p}\sum_{u=1}^{p} |\tau_{\lambda_{uv}}(\tilde{\sigma}^*_{uv}) - \sigma_{uv}|^2 \le \max_v \sum_{u=1}^{p} |\tau_{\lambda_{uv}}(\tilde{\sigma}^*_{uv}) - \sigma_{uv}|^2. \tag{4.5}$$

According to Lemma 3.1, we have

$$\frac{1}{p}\|(\tilde{\mathbf{\Sigma}}^*)^\tau - \Sigma\|_F^2 \lesssim \frac{\log p}{n^*_{\min}} \max_v \sum_{u=1}^{p} \left(\min\left\{ \sqrt{\sigma_{uu}\sigma_{vv}}, \frac{|\sigma_{uv}|}{\sqrt{(\log p)/n}} \right\} \right)^2 \tag{4.6}$$

on the event Q. Furthermore,

$$\sup_{\Sigma\in\mathcal{H}(s_{n,p})} \int_Q \frac{1}{\sqrt{p}}\|(\tilde{\mathbf{\Sigma}}^*)^\tau - \mathbf{\Sigma}\|_F dP$$
$$\lesssim \sup_{\Sigma\in\mathcal{H}(s_{n,p})} \int_Q \left\{ \frac{\log p}{n^*_{\min}} \left(\max_v \sum_{u=1}^{p} \min\left\{ \sqrt{\sigma_{uu}\sigma_{vv}}, \frac{|\sigma_{uv}|}{\sqrt{(\log p)/n}} \right\} \right)^2 \right\}^{1/2} dP$$
$$\le s_{n,p}\sqrt{\frac{\log p}{n^*_{\min}}}$$

holds due to the definition of $\mathcal{H}(s_{n,p})$.

Hence, it suffices to prove

$$\sup_{\Sigma\in\mathcal{H}(s_{n,p})} \int_{Q^c} \frac{1}{\sqrt{p}}\|(\tilde{\mathbf{\Sigma}}^*)^\tau - \mathbf{\Sigma}\|_F dP \lesssim s_{n,p}\sqrt{\frac{\log p}{n^*_{\min}}}. \tag{4.7}$$

By (4.4), we find

$$|\tau_{\lambda_{uv}}(\tilde{\sigma}^*_{uv}) - \sigma_{uv}|^2 \lesssim |\tilde{\sigma}^*_{uv} - \sigma_{uv}|^2 + \left(\min\left\{ \sqrt{\sigma_{uu}\sigma_{vv}}, \frac{|\sigma_{uv}|}{\sqrt{(\log p)/n}} \right\} \right)^2.$$

Substituting the above inequality into (4.5) leads to

$$\frac{1}{p}\|(\tilde{\mathbf{\Sigma}}^*)^\tau - \Sigma\|_F^2 \lesssim \max_v \sum_{u=1}^{p} |\tilde{\sigma}^*_{uv} - \sigma_{uv}|^2 + \max_v \sum_{u=1}^{p} \left(\min\left\{ \sqrt{\sigma_{uu}\sigma_{vv}}, \frac{|\sigma_{uv}|}{\sqrt{(\log p)/n}} \right\} \right)^2. \tag{4.8}$$

Since $\sqrt{|a|+|b|} \le \sqrt{|a|} + \sqrt{|b|}$ and $(\max_v \sum_{u=1}^p |a_{uv}|^2)^{\frac{1}{2}} \le \max_v \sum_{u=1}^p |a_{uv}|$, we obtain

$$\frac{1}{\sqrt{p}}\|(\tilde{\mathbf{\Sigma}}^*)^\tau - \Sigma\|_F \lesssim \max_v \sum_{u=1}^p |\tilde{\sigma}^*_{uv} - \sigma_{uv}| + \max_v \sum_{u=1}^p \min\left\{\sqrt{\sigma_{uu}\sigma_{vv}}, \frac{|\sigma_{uv}|}{\sqrt{(\log p)/n}}\right\}.$$

Thus, (4.7) follows from Lemma 4.1(i), (iii) and $s_{n,p} \gtrsim 1$.

(iii) By $\max_u \sigma_{uu} \le M$, we obtain

$$\min\left\{\sqrt{\sigma_{uu}\sigma_{vv}}, \frac{|\sigma_{uv}|}{\sqrt{(\log p)/n}}\right\} \le M. \tag{4.9}$$

On the other hand, (4.6), (4.9) and $\mathbf{\Sigma} \in \mathcal{H}(s_{n,p})$ tells

$$\frac{1}{p}\|(\tilde{\mathbf{\Sigma}}^*)^\tau - \Sigma\|_F^2 \lesssim \frac{\log p}{n^*_{\min}} \max_v \sum_{u=1}^p \min\left\{\sqrt{\sigma_{uu}\sigma_{vv}}, \frac{|\sigma_{uv}|}{\sqrt{(\log p)/n}}\right\} \le s_{n,p}\frac{\log p}{n^*_{\min}}$$

under the event Q. Therefore,

$$\sup_{\substack{\mathbf{\Sigma}\in\mathcal{H}(s_{n,p}) \\ \max_u \sigma_{uu}\le M}} \int_Q \frac{1}{\sqrt{p}}\|(\tilde{\mathbf{\Sigma}}^*)^\tau - \mathbf{\Sigma}\|_F dP \lesssim \sqrt{s_{n,p}}\sqrt{\frac{\log p}{n^*_{\min}}}. \tag{4.10}$$

Using (4.8) and (4.9), we have

$$\begin{aligned}\frac{1}{\sqrt{p}}\|(\tilde{\mathbf{\Sigma}}^*)^\tau - \mathbf{\Sigma}\|_F &\lesssim \left(\max_v \sum_{u=1}^p \min\left\{\sqrt{\sigma_{uu}\sigma_{vv}}, \frac{|\sigma_{uv}|}{\sqrt{(\log p)/n}}\right\}\right)^{1/2} + \left(\max_v \sum_{u=1}^p |\tilde{\sigma}^*_{uv} - \sigma_{uv}|^2\right)^{1/2} \\ &\lesssim \left(\max_v \sum_{u=1}^p \min\left\{\sqrt{\sigma_{uu}\sigma_{vv}}, \frac{|\sigma_{uv}|}{\sqrt{(\log p)/n}}\right\}\right)^{1/2} + \max_v \sum_{u=1}^p |\tilde{\sigma}_{uv} - \sigma_{uv}|.\end{aligned}$$

Hence, it holds

$$\sup_{\substack{\mathbf{\Sigma}\in\mathcal{H}(s_{n,p}) \\ \max_u \sigma_{uu}\le M}} \int_{Q^c} \frac{1}{\sqrt{p}}\|(\tilde{\mathbf{\Sigma}}^*)^\tau - \mathbf{\Sigma}\|_F dP \lesssim \sqrt{s_{n,p}}\sqrt{\frac{\log p}{n^*_{\min}}} \tag{4.11}$$

due to Lemma 4.1(ii), (iii) and $s_{n,p} \gtrsim 1$. Finally, conclusion (iii) follows from (4.10) to (4.11). □

Remark 4.1. *The upper bound of Theorem 4.1(i) is optimal due to Proposition 3.1 in [15]. In addition, Theorem 4.1(i) performs better than Theorem 3.1 of [15] which requires* **X** *to be sub-Gaussian.*

Remark 4.2. *From the proving process of Theorem 4.1(i), we observe*

$$\sup_{\mathbf{\Sigma}\in\mathcal{H}(s_{n,p})} \mathbb{E}\|(\tilde{\mathbf{\Sigma}}^*)^\tau - \mathbf{\Sigma}\|_1 \le s_{n,p}\sqrt{\frac{\log p}{n^*_{\min}}}.$$

Note that $\|\mathbf{A}\|_\omega \leq \|\mathbf{A}\|_1 (1 \leq \omega \leq \infty)$ *for any symmetric matrix* $\mathbf{A}$. *Then,*

$$\sup_{\Sigma \in \mathcal{H}(s_{n,p})} \mathbb{E}\|(\tilde{\mathbf{\Sigma}}^*)^\tau - \mathbf{\Sigma}\|_\omega \leq s_{n,p} \sqrt{\frac{\log p}{n^*_{\min}}}$$

holds.

Remark 4.3. *The condition*

$$\mathbb{E}|\tilde{\sigma}^*_{uv} - \sigma_{uv}|^2 \leq M \tag{4.12}$$

in Lemma 4.1 and Theorem 4.1 is mild. In fact, the generalized Huber estimator (Example 2.1) and generalized truncated mean estimator (Example 2.2) both satisfy (4.12). The details can be found in Appendix.

5. Simulation studies

Let $(\tilde{\mathbf{\Sigma}}^*_H)^\tau$ and $(\tilde{\mathbf{\Sigma}}^*_T)^\tau$ be defined by (3.1) and (3.2), this section investigates the numerical properties and performances of the estimators $(\tilde{\mathbf{\Sigma}}^*_H)^\tau$, $(\tilde{\mathbf{\Sigma}}^*_T)^\tau$ and compares these two estimators with the adaptive thresholding estimator $\hat{\mathbf{\Sigma}}^{at}$ proposed by [15]. The following two types of sparse covariance matrices are considered:

Model 1. (Rothman et al., [18]) $\mathbf{\Sigma} = (\sigma_{uv})_{p\times p}$ with $\sigma_{uv} = \max\{1 - |u - v|/5, 0\}$.

Model 2. (Cai and Zhang, [15]) $\mathbf{\Sigma} = \mathbf{I}_p + (\mathbf{D} + \mathbf{D}^T)/(\|\mathbf{D} + \mathbf{D}^T\|_2 + 0.01)$, where $\mathbf{D} = (d_{uv})_{p\times p}$ is given by $d_{uu} = 0$ $(u = 1, \cdots, p)$ and

$$d_{uv} = \begin{cases} 1, & \text{with probability } 0.1; \\ 0, & \text{with probability } 0.8; \\ -1, & \text{with probability } 0.1; \end{cases} \quad \text{for } u \neq v.$$

Under each model we generate random samples $\mathbf{X}_i \in \mathbb{R}^p$ $(i = 1, \cdots, n)$ by two different scenarios:

(i) $\mathbf{X}_i$ are independently drawn from multivariate t-distribution $t_\nu(0, \mathbf{\Sigma})$ with freedom $\nu = 4.5$;

(ii) $\mathbf{X}_i$ are independently drawn from multivariate skewed t-distribution $st_\nu(0, \mathbf{\Sigma}, \epsilon)$ with freedom $\nu = 5$ and skew parameter $\epsilon = 10$.

In each simulation setting we adopt the following two cases of the missingness for the data matrix $\mathbf{Y} = (\mathbf{X}_1, \cdots, \mathbf{X}_n)$ with $\mathbf{X}_i = (X_{i1}, \cdots, X_{ip})^T$ which proposed by Cai and Zhang [15]. The first case is missing uniformly and completely at random(MUCR) in which every entries X_{ik} are observed with probability $0 < \rho \leq 1$. The second case is missing not uniformly but completely at random(MCR) in which $\mathbf{Y}$ is divided into four equal-size parts,

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_{11} & \mathbf{Y}_{12} \\ \mathbf{Y}_{21} & \mathbf{Y}_{22} \end{bmatrix}, \mathbf{Y}_{11},\ \mathbf{Y}_{12},\ \mathbf{Y}_{21},\ \mathbf{Y}_{22} \in \mathbb{R}^{\frac{p}{2}\times\frac{n}{2}}$$

where every entries of $\mathbf{Y}_{11},\ \mathbf{Y}_{22}$ are observed with probability $0 < \rho^{(1)} \leq 1$ every entries of $\mathbf{Y}_{12},\ \mathbf{Y}_{21}$ are observed with probability $0 < \rho^{(2)} \leq 1$.

Moreover, for each procedure we set $p = 50, 200, 300$ and $n = 50, 100, 200$ respectively and 50 replications are used. Meanwhile, we choose the soft thresholding rule and measure the errors by the spectral and Frobenius norm respectively in each setting. The tuning parameter in thresholding estimator is chosen by 10-fold cross-validation which is explained in Section 4 of Cai and Zhang [15], and unspecified tuning parameters in the generalized pilot estimator are chosen by the method suggested in Section 6 of Avella-Medina et al. [11].

Tables 1 and 2 demonstrate that thredsholding estimators $(\tilde{\mathbf{\Sigma}}_H^*)^\tau$ and $(\tilde{\mathbf{\Sigma}}_T^*)^\tau$ perform better than the adaptive thresholding estimator $\hat{\mathbf{\Sigma}}^{at}$ under both MUCR and MCR settings. Moreover, thresholding generalized Huber estimator $(\tilde{\mathbf{\Sigma}}_H^*)^\tau$ outperforms thresholding generalized truncated mean estimator $(\tilde{\mathbf{\Sigma}}_T^*)^\tau$. We also find that the errors decrease if sample size n gets larger. Meanwhile, we observe that the errors under Model 1 is larger than under Model 2 since the covariance matrix in Model 1 is more dense than in Model 2. All these numerical results are consistent with our theoretical results.

Table 1. Means errors (with standard errors in parentheses) for three kinds of thresholding estimators with t-distribution.

	Spectral norm			Frobenius norm		
(p, n)	$\hat{\mathbf{\Sigma}}^{at}$	$(\tilde{\mathbf{\Sigma}}_H^*)^\tau$	$(\tilde{\mathbf{\Sigma}}_T^*)^\tau$	$\hat{\mathbf{\Sigma}}^{at}$	$(\tilde{\mathbf{\Sigma}}_H^*)^\tau$	$(\tilde{\mathbf{\Sigma}}_T^*)^\tau$
Model 1, MUCR, $\rho = 0.5$						
(50, 50)	6.59(0.09)	4.21(0.03)	5.14(0.01)	11.13(0.09)	8.28(0.02)	9.24(0.02)
(50, 200)	3.78(0.04)	2.07(0.06)	2.28(0.02)	5.66(0.01)	3.15(0.04)	4.06(0.01)
(200, 100)	5.62(0.02)	3.67(0.02)	4.42(0.03)	16.17(0.02)	13.02(0.03)	13.91(0.05)
(200, 200)	4.46(0.03)	2.59(0.02)	3.02(0.01)	11.39(0.03)	8.49(0.01)	9.45(0.07)
(300, 200)	4.97(0.02)	2.92(0.07)	3.63(0.04)	15.80(0.04)	12.73(0.03)	13.68(0.06)
Model 2, MUCR, $\rho = 0.5$						
(50, 50)	5.72(0.03)	3.58(0.02)	4.29(0.01)	9.12(0.08)	5.34(0.02)	6.93(0.02)
(50, 200)	3.45(0.06)	1.65(0.01)	2.12(0.02)	5.23(0.03)	2.17(0.04)	3.41(0.06)
(200, 100)	5.31(0.03)	3.26(0.08)	4.13(0.04)	11.88(0.01)	8.54(0.03)	10.07(0.04)
(200, 200)	3.89(0.02)	1.92(0.01)	2.96(0.03)	8.96(0.01)	5.46(0.02)	6.84(0.05)
(300, 200)	4.27(0.02)	2.14(0.02)	3.48(0.04)	11.51(0.02)	8.12(0.03)	9.72(0.05)
Model 1, MCR, $\rho^{(1)} = 0.8,\ \rho^{(2)} = 0.2$						
(50, 50)	6.42(0.01)	4.12(0.01)	4.93(0.02)	10.79(0.02)	7.90(0.09)	9.12(0.03)
(50, 200)	3.65(0.04)	1.96(0.03)	2.16(0.02)	5.42(0.04)	2.86(0.01)	3.96(0.02)
(200, 100)	5.47(0.06)	3.55(0.02)	4.32(0.04)	15.66(0.02)	12.93(0.02)	13.77(0.06)
(200, 200)	4.19(0.05)	2.38(0.01)	2.74(0.04)	11.15(0.01)	8.32(0.02)	9.26(0.05)
(300, 200)	4.52(0.03)	2.63(0.05)	3.25(0.05)	15.48(0.04)	12.56(0.01)	13.18(0.07)
Model 2, MCR, $\rho^{(1)} = 0.8,\ \rho^{(2)} = 0.2$						
(50, 50)	5.48(0.01)	3.31(0.09)	4.28(0.03)	8.97(0.08)	5.19(0.01)	6.62(0.02)
(50, 200)	3.03(0.02)	1.31(0.02)	1.97(0.01)	5.12(0.03)	2.02(0.04)	3.47(0.03)
(200, 100)	5.15(0.04)	3.15(0.02)	4.09(0.03)	11.56(0.04)	8.25(0.03)	9.86(0.04)
(200, 200)	3.54(0.03)	1.72(0.01)	2.68(0.03)	8.61(0.01)	5.36(0.02)	6.73(0.06)
(300, 200)	3.96(0.07)	2.08(0.03)	3.04(0.05)	11.35(0.02)	8.01(0.02)	9.69(0.07)

Table 2. Means errors (with standard errors in parentheses) for three kinds of thresholding estimators with skewed t-distribution.

	Spectral norm			Frobenius norm		
(p, n)	$\hat{\mathbf{\Sigma}}^{at}$	$(\tilde{\mathbf{\Sigma}}_H^*)^\tau$	$(\tilde{\mathbf{\Sigma}}_T^*)^\tau$	$\hat{\mathbf{\Sigma}}^{at}$	$(\tilde{\mathbf{\Sigma}}_H^*)^\tau$	$(\tilde{\mathbf{\Sigma}}_T^*)^\tau$
Model 1, MUCR, ρ = 0.5						
(50, 50)	8.98(0.04)	7.34(0.03)	8.03(0.04)	12.44(0.02)	9.74(0.01)	10.79(0.02)
(50, 200)	4.69(0.07)	3.15(0.08)	3.68(0.01)	6.06(0.08)	3.83(0.03)	4.65(0.03)
(200, 100)	8.17(0.03)	6.88(0.04)	7.53(0.03)	18.32(0.05)	15.41(0.05)	16.93(0.06)
(200, 200)	5.93(0.01)	4.54(0.09)	5.11(0.01)	12.58(0.06)	9.93(0.02)	11.07(0.05)
(300, 200)	6.98(0.02)	5.85(0.02)	6.06(0.04)	18.19(0.04)	15.34(0.03)	16.51(0.06)
Model 2, MUCR, ρ = 0.5						
(50, 50)	7.67(0.01)	5.41(0.03)	6.34(0.04)	10.31(0.08)	8.68(0.08)	9.26(0.03)
(50, 200)	4.15(0.02)	1.88(0.06)	2.65(0.03)	5.49(0.01)	3.24(0.04)	4.01(0.02)
(200, 100)	7.42(0.01)	5.14(0.09)	6.01(0.02)	15.45(0.02)	13.73(0.03)	14.39(0.05)
(200, 200)	4.88(0.03)	2.64(0.02)	3.60(0.08)	10.63(0.01)	8.89(0.02)	9.37(0.06)
(300, 200)	5.41(0.01)	3.17(0.02)	4.32(0.06)	15.30(0.02)	13.36(0.04)	13.95(0.05)
Model 1, MCR, $\rho^{(1)} = 0.8,\ \rho^{(2)} = 0.2$						
(50, 50)	8.75(0.02)	7.16(0.04)	7.84(0.04)	12.23(0.09)	9.46(0.02)	10.55(0.01)
(50, 200)	4.36(0.08)	3.03(0.05)	3.59(0.03)	5.80(0.01)	3.62(0.06)	4.42(0.02)
(200, 100)	7.99(0.03)	6.52(0.02)	7.38(0.04)	18.08(0.03)	15.29(0.03)	16.54(0.06)
(200, 200)	5.81(0.02)	4.34(0.03)	4.96(0.01)	12.36(0.02)	9.64(0.04)	10.78(0.05)
(300, 200)	6.69(0.07)	5.48(0.01)	5.92(0.03)	18.12(0.03)	15.25(0.05)	16.49(0.07)
Model 2, MCR, $\rho^{(1)} = 0.8,\ \rho^{(2)} = 0.2$						
(50, 50)	7.58(0.02)	5.22(0.04)	6.27(0.05)	10.22(0.03)	8.36(0.03)	8.83(0.02)
(50, 200)	4.06(0.06)	1.84(0.02)	2.48(0.05)	5.26(0.04)	2.95(0.06)	3.72(0.03)
(200, 100)	7.14(0.02)	5.06(0.07)	5.95(0.04)	15.34(0.01)	13.36(0.05)	14.07(0.06)
(200, 200)	4.76(0.02)	2.42(0.01)	3.36(0.08)	10.46(0.03)	8.77(0.03)	9.16(0.06)
(300, 200)	5.30(0.01)	3.08(0.02)	4.18(0.06)	15.08(0.09)	13.06(0.05)	13.74(0.05)

6. Conclusions

In this paper, we propose the generalized pilot estimator in the presence of incomplete heavy-tailed data. Moreover, two kinds of generalized pilot estimators are provided under the bounded fourth moment assumption while lots of previous studies hinged upon the sub-Gaussian condition. In addition, we establish the thresholding pilot estimator for a family of sparse covariance matrices and give the convergence rates in terms of probability and expectation respectively.

In the future, we may consider the compositional data with missing data under lower bounded moment assumption by referring Li et al. [20]. Moreover, we can adopt the different methods to estimate the sparse covariance matrix with incomplete data such as the proximal distance algorithm [21] or continuous matrix shrinkage [22].

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This paper is supported by the National Natural Science Foundation of China (No. 12171016).

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. S. Mendelson, N. Zhivotovskiy, Robust covariance estimation under $L_4 - L_2$ norm equivalence, *Ann. Statist.*, **48** (2020), 1648–1664. https://doi.org/10.1214/19-AOS1862
2. Y. Dendramis, L. Giraitis, G. Kapetanios, Estimation of time-varying covariance matrices for large datasets, *Economet. Theory*, **37** (2021), 1100–1134. https://doi.org/10.1017/S0266466620000535
3. Y. Zhang, J. Tao, Y. Lv, G. Wang, An improved DCC model based on large-dimensional covariance matrices estimation and its applications, *Symmetry*, **15** (2023), 953. https://doi.org/10.3390/sym15040953
4. D. Belomestny, M. Trabs, A. Tsybakov, Sparse covariance matrix estimation in high-dimensional deconvolution, *Bernoulli*, **25** (2019), 1901–1938. https://doi.org/10.3150/18-BEJ1040A
5. X. Kang, X. Deng, On variable ordination of Cholesky-based estimation for a sparse covariance matrix, *Canad. J. Stat.*, **49** (2021), 283–310. https://doi.org/10.1002/cjs.11564
6. N. Bettache, C. Butucea, M. Sorba, Fast nonasymptotic testing and support recovery for large sparse Toeplitz covariance matrices, *J. Multivariate Anal.*, **190** (2022), 104883. https://doi.org/10.1016/j.jmva.2021.104883
7. W. Liang, Y. Wu, H. Chen, Sparse covariance matrix estimation for ultrahigh dimensional data, *Stat*, **11** (2022), e479. https://doi.org/10.1002/sta4.479
8. P. J. Bickel, E. Levina, Covariance regularization by thresholding, *Ann. Statist.*, **36** (2008), 2577–2604. https://doi.org/10.1214/08-AOS600
9. T. Cai, W. Liu, Adaptive thresholding for sparse covariance matrix estimation, *J. Amer. Stat. Assoc.*, **106** (2011), 672–684. https://doi.org/10.1198/jasa.2011.tm10560
10. T. T. Cai, H. H. Zhou, Optimal rates of convergence for sparse covariance matrix estimation, *Ann. Statist.*, **40** (2012), 2389–2420. https://doi.org/10.1214/12-AOS998
11. M. Avella-Medina, H. Battery, J. Fan, Q. Li, Robust estimation of high-dimensional covariance and precision matrices, *Biometrika*, **105** (2018), 271–284. https://doi.org/10.1093/biomet/asy011
12. R. D. Hawkins, G. C. Hon, B. Ren, Next-generation genomics: an intergrative approach, *Nat. Rev. Genet.*, **11** (2010), 476–486. https://doi.org/10.1038/nrg2795

13. K. Lounici, Sparse principal component analysis with missing observations, In: C. Houdré, D. Mason, J. Rosiński, J. Wellner, *High dimensional probability VI*, Progress in Probability, **66** (2013), 327–356. https://doi.org/10.1007/978-3-0348-0490-5_20

14. P. L. Loh, M. J. Wainwright, High-dimensional regression with noisy and missing data: Provable guarantees with non-convexity, *Ann. Statist.*, **40** (2012), 1637–1664. https://doi.org/10.1214/12-AOS1018

15. T. T. Cai, A. Zhang, Minimax rate-optimal estimation of high-dimensional covariance matrices with incomplete data, *J. Multivariate Anal.*, **150** (2016), 55–74. https://doi.org/10.1016/j.jmva.2016.05.002

16. J. Fan, Q. Li, Y. Wang, Estimation of high-dimensional mean regression in absence of symmetry and light-tail assumptions, *J. R. Stat. Soc. B*, **79** (2017), 247–265. https://doi.org/10.1111/rssb.12166

17. M. Pascal, *Concentration inequalities and model selection*, Berlin, Heidelberg: Springer, 2007. https://doi.org/10.1007/978-3-540-48503-2

18. A. J. Rothman, E. Levina, J. Zhu, Generalized thresholding of large covariance matrices, *J. Am. Stat. Assoc.*, **104** (2009), 177–186. https://doi.org/10.1198/jasa.2009.0101

19. T. T. Cai, W. Liu, H. H. Zhou, Estiamtion sparse precision matrix: optimal rates of covariacne and adaptive estimation, *Ann. Statist.*, **44** (2016), 455–488. https://doi.org/10.1214/13-AOS1171

20. D. Li, A. Srinivasan, Q. Chen, L. Xue, Robust covariance matrix estimation for high-dimensional compositional data with application to sales data analysis, *J. Bus. Econ. Stat.*, in press. https://doi.org/10.1080/07350015.2022.2106990

21. J. Xu, K. Lange, A proximal distance algorithm for likelihood-based sparse covariance estimation, *Biometrika*, **109** (2022), 1047–1066. https://doi.org/10.1093/biomet/asac011

22. F. Xie, J. Cape, C. E. Priebe, Y. Xu, Bayesian sparse spiked covariance model with a continuous matrix shrinkage prior, *Bayesian Anal.*, **17** (2022), 1193–1217. https://doi.org/10.1214/21-BA1292

Appendix

In order to show Example 2.1 and Example 2.2 satisfying condition (4.12), we introduce a Proposition A.1.

Proposition A.1. *Let* $\max\limits_{1\le u\le p}\mathbb{E}|X_u|^4 \le k^4$, $\mathbb{E}X_u = \mu_u$, $\mathbb{E}(X_uX_v) = \mu_{uv}$, *and Assumption 1.1 holds. If* $\tilde{\mu}_u^*$ *and* $\tilde{\mu}_{uv}^*$ *satisfy*

$$|\tilde{\mu}_u^*| \lesssim |A|^{-1}\sum_{i\in A}|X_{iu}|, \tag{A.1}$$

$$|\tilde{\mu}_{uv}^*| \lesssim |B|^{-1}\sum_{i\in B}|X_{iu}X_{iv}| \tag{A.2}$$

where $A, B \subseteq \{1, \cdots, n\}$. *Then,* $\tilde{\mathbf{\Sigma}}^* = (\tilde{\sigma}_{uv}^*)_{p\times p} = (\tilde{\mu}_{uv}^* - \tilde{\mu}_u^*\tilde{\mu}_v^*)_{p\times p}$ *obeys (4.12).*

Proof. It suffices to prove

$$|\sigma_{uv}| \lesssim 1, \tag{A.3}$$

$$\mathbb{E}|\tilde{\sigma}^*_{uv}|^2 \lesssim 1. \tag{A.4}$$

By $\max\limits_{1\le u\le p} \mathbb{E}|X_u|^4 \le k^4$, one knows $\mathbb{E}|X_u| \le (\mathbb{E}|X_u|^4)^{1/4} \le k$ and

$$\mathbb{E}|X_uX_v| \le (\mathbb{E}X_u^2)^{\frac{1}{2}}(\mathbb{E}X_v^2)^{\frac{1}{2}} \le (\mathbb{E}|X_u|^4)^{\frac{1}{4}}(\mathbb{E}|X_v|^4)^{\frac{1}{4}} \le k^2.$$

Thus, it holds

$$|\sigma_{uv}| \le \mathbb{E}|X_uX_v| + (\mathbb{E}|X_u|)(\mathbb{E}|X_v|) \le 2k^2 \lesssim 1$$

which reaches (A.3).

For (A.4), one observes

$$\mathbb{E}|\tilde{\sigma}^*_{uv}|^2 = \mathbb{E}|\tilde{\mu}^*_{uv} - \tilde{\mu}^*_u\tilde{\mu}^*_v|^2 \lesssim \mathbb{E}|\tilde{\mu}^*_{uv}|^2 + \mathbb{E}|\tilde{\mu}^*_u\tilde{\mu}^*_v|^2. \tag{A.5}$$

According to (A.1) and Jensen's inequality, it follows

$$\mathbb{E}|\tilde{\mu}^*_u|^4 \lesssim \mathbb{E}\left(\frac{1}{|A|}\sum_{i\in A}|X_{iu}|\right)^4 \le \mathbb{E}|X_u|^4 \le k^4.$$

Furthermore, upon combining Cauchy–Schwarz inequality leads to

$$\mathbb{E}|\tilde{\mu}^*_u\tilde{\mu}^*_v|^2 \le \left(\mathbb{E}|\tilde{\mu}^*_u|^4\mathbb{E}|\tilde{\mu}^*_v|^4\right)^{\frac{1}{2}} \le k^4 \lesssim 1. \tag{A.6}$$

Similarly, (A.2) implies

$$\mathbb{E}|\tilde{\mu}^*_{uv}|^2 \lesssim \mathbb{E}\left(\frac{1}{|B|}\sum_{i\in B}|X_{iu}X_{iv}|\right)^2 \le \mathbb{E}|X_uX_v|^2.$$

By Cauchy-Schwarz inequality and $\max\limits_{1\le u\le p} \mathbb{E}|X_u|^4 \le k^4$, one finds

$$\mathbb{E}|X_uX_v|^2 \le (\mathbb{E}|X_u|^4)^{\frac{1}{2}}\left(\mathbb{E}|X_v|^4\right)^{\frac{1}{2}} \le k^4.$$

Hence,

$$\mathbb{E}|\tilde{\mu}^*_{uv}|^2 \lesssim k^4 \lesssim 1. \tag{A.7}$$

Finally, the expected conclusion (A.4) follows from (A.5)–(A.7). This completes the proof of Proposition A.1. □

Now, based on Proposition A.1 we verify two kinds of generalized pilot estimators (Example 2.1 and Example 2.2) satisfying (4.12).

For the generalized truncated mean estimator $\tilde{\mathbf{\Sigma}}^*_T := \left((\tilde{\mu}^*_T)_{uv} - (\tilde{\mu}^*_T)_u(\tilde{\mu}^*_T)_v\right)_{p\times p}$, it is easy to see that $(\tilde{\mu}^*_T)_u$, $(\tilde{\mu}^*_T)_{uv}$ obey (A.1), (A.2) respectively.

By (2.14), we know

$$|(\tilde{\mu}^*_T)_u| \le \frac{1}{n^*_u}\sum_{i=1}^{n}|X^*_{iu}| = \frac{1}{n^*_u}\sum_{i=1}^{n}|X_{iu}S_{iu}| = \frac{1}{n^*_u}\sum_{i\in A_u}|X_{iu}|$$

where $A_u = \{i : S_{iu} \neq 0\}$ and $|A_u| = n_u^*$. Similarly, it holds

$$|(\tilde{\mu}_T^*)_{uv}| \leq \frac{1}{n_{uv}^*} \sum_{i \in A_{uv}} |X_{iu} X_{iv}|$$

with $A_{uv} = \{i : S_{iu} S_{iv} \neq 0\}$ and $|A_{uv}| = n_{uv}^*$. The above two inequalities imply $(\tilde{\mu}_T^*)_u$, $(\tilde{\mu}_T^*)_{uv}$ satisfying (A.1), (A.2) respectively.

In fact, it is hard to check the generalized Huber estimator

$$\tilde{\mathbf{\Sigma}}_H^* := \left((\tilde{\mu}_H^*)_{uv} - (\tilde{\mu}_H^*)_u (\tilde{\mu}_H^*)_v\right)_{p \times p}$$

satisfying (4.12) due to the structures of $(\tilde{\mu}_H^*)_u$ and $(\tilde{\mu}_H^*)_{uv}$ being unclear. But we can consider a special case.

Proposition A.2. *Let $A_u = \{i : S_{iu} \neq 0\}$, $A_{uv} = \{i : S_{iu} S_{iv} \neq 0\}$. If α_u, α_{uv} defined in (2.9) and (2.10) obey*

$$\alpha_u > \max_{i \in A_u} X_{iu} - \min_{i \in A_u} X_{iu}, \ \alpha_{uv} > \max_{i \in A_{uv}} X_{iu} X_{iv} - \min_{i \in A_{uv}} X_{iu} X_{iv}$$

respectively. Then, $(\tilde{\mu}_H^*)_u$, $(\tilde{\mu}_H^*)_{uv}$ satisfy (A.1), (A.2).

Proof. For $i \in A_u$, it holds

$$X_{iu} - \left(\max_{i \in A_u} X_{iu} - \alpha_u\right) \geq \min_{i \in A_u} X_{iu} - \max_{i \in A_u} X_{iu} + \alpha_u > 0,$$
$$X_{iu} - \left(\min_{i \in A_u} X_{iu} + \alpha_u\right) \leq \max_{i \in A_u} X_{iu} - \min_{i \in A_u} X_{iu} - \alpha_u < 0.$$

Obviously, (2.9) is equivalent to

$$\sum_{i \in A_u} \psi_{\alpha_u}(X_{iu} - (\tilde{\mu}_H^*)_u) = 0.$$

By the definition of $\psi_{\alpha_u}(x)$, we have

$$\sum_{i \in A_u} \psi_{\alpha_u}\left(X_{iu} - \left(\max_{i \in A_u} X_{iu} - \alpha_u\right)\right) > 0,$$
$$\sum_{i \in A_u} \psi_{\alpha_u}\left(X_{iu} - \left(\min_{i \in A_u} X_{iu} + \alpha_u\right)\right) < 0.$$

Note that $\sum_{i \in A_u} \psi_{\alpha_u}(X_{iu} - (\tilde{\mu}_H^*)_u)$ is the continuous and decreasing function about $(\tilde{\mu}_H^*)_u$. Then, the solution of equation $\sum_{i \in A_u} \psi_{\alpha_u}(X_{iu} - (\tilde{\mu}_H^*)) = 0$ belongs to the interval $(\max_{i \in A_u} X_{iu} - \alpha_u, \min_{i \in A_u} X_{iu} + \alpha_u)$.

Hence, we obtain $\max_{\{i \in A_u\}} X_{iu} - \alpha_u < (\tilde{\mu}_H^*)_u < \min_{\{i \in A_u\}} X_{iu} + \alpha_u$ and

$$-\alpha_u < X_{iu} - (\tilde{\mu}_H^*)_u < \alpha_u.$$

Furthermore, the above inequality and definition of $\psi_{\alpha_u}(x)$ implies

$$\sum_{i \in A_u} \psi_{\alpha_u}(X_{iu} - (\tilde{\mu}_H^*)_u) = \sum_{i \in A_u} (X_{iu} - (\tilde{\mu}_H^*)_u) = \sum_{i \in A_u} X_{iu} - n_u^* (\tilde{\mu}_H^*)_u.$$

Therefore, $(\tilde{\mu}_H^*)_u = (n_u^*)^{-1} \sum_{i \in A_u} X_{iu}$ satisfies (A.1).

Following the similar discussion, we can derive $(\tilde{\mu}_H^*)_{uv}$ satisfying (A.2) with

$$\alpha_{uv} > \max_{i \in A_{uv}} X_{iu} X_{iv} - \min_{i \in A_{uv}} X_{iu} X_{iv}.$$

□

In fact, the condition in Proposition A.2 is easy to satisfy, since $\log p = o(n^*_{\min})$ and $n^*_{\min} \leq n^*_{uv} \leq n^*_u$ lead to large enough α_u and α_{uv}.