

Estimation of conditional mean operator under the bandable covariance structure

Kwangmin Lee

Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison, Wisconsin, United States of America
e-mail: klee564@wisc.edu

Kyoungjae Lee

Department of Statistics, Sungkyunkwan University, Seoul, Korea
e-mail: leekjstat@gmail.com

Jaeyong Lee

Department of Statistics, Seoul National University, Seoul, Korea
e-mail: leejyc@gmail.com

Abstract: We consider high-dimensional multivariate linear regression models, where the joint distribution of covariates and response variables is a multivariate normal distribution with a bandable covariance matrix. The main goal of this paper is to estimate the regression coefficient matrix, which is a function of the bandable covariance matrix. Although the tapering estimator of covariance has the minimax optimal convergence rate for the class of bandable covariances, we show that it is sub-optimal for the regression coefficient; that is, a minimax estimator for the class of bandable covariances may not be a minimax estimator for its functionals. We propose the blockwise tapering estimator of the regression coefficient, which has the minimax optimal convergence rate for the regression coefficient under the bandable covariance assumption. We also propose a Bayesian procedure called the blockwise tapering post-processed posterior of the regression coefficient and show that the proposed Bayesian procedure has the minimax optimal convergence rate for the regression coefficient under the bandable covariance assumption. We show that the proposed methods outperform the existing methods via numerical studies.

MSC2020 subject classifications: Primary 62C20, 62H12; secondary 62J05.

Keywords and phrases: Bandable covariance, conditional mean, covariance estimation, minimax analysis, post-processed posterior.

Received May 2021.

Contents

1	Introduction	1254
2	Blockwise tapering estimator and minimax analysis	1257
2.1	Notation	1257
2.2	Blockwise tapering estimator	1258

2.3	Minimax analysis of blockwise tapering estimator	1260
3	Blockwise tapering post-processed posterior	1262
4	Numerical studies	1264
4.1	Simulation	1264
4.2	Application to forecasting traffic speed	1267
5	Proofs of theorems and lemma	1270
5.1	Proof of Theorem 2.1	1270
5.2	Proof of Theorem 2.2	1271
5.3	Proofs of Lemma 2.3 and Theorem 2.4	1273
5.4	Proof of Theorem 2.5	1278
6	Discussion	1281
A	Proofs of remaining lemmas and theorems	1282
A.1	Proofs of Theorems 3.1 and 3.2	1282
A.2	Proofs of lemmas in Section 5.1	1285
A.3	Proofs of lemmas in Section 5.2	1287
A.4	Proof of lemma in Section 5.3	1292
A.5	Proof of lemma in Section 5.4	1293
A.6	Proofs of lemmas in Section A.1	1294
	References	1300

1. Introduction

Consider the multivariate linear regression model

$$Y_i = CX_i + \epsilon_i, \quad i = 1, \dots, n,$$

where $Y_i \in \mathbb{R}^q$ is a response vector, $X_i \in \mathbb{R}^{p_0}$ is a covariate vector, $C \in \mathbb{R}^{q \times p_0}$ is a regression coefficient matrix, and $\epsilon_i \in \mathbb{R}^q, i = 1, 2, \dots, n$, are independent and identically distributed error vectors from a q -dimensional distribution with mean zero. The multivariate linear regression model has been used for various fields of applications. For example, [36] analyzed atmospheric data using the model to forecast PM2.5 concentration, and [29] used the model to analyze the genomics data.

For the estimation of the multivariate linear regression coefficient $C \in \mathbb{R}^{q \times p_0}$, one of the most commonly used approaches is a penalized least square method, which finds the minimizer of the following objective function,

$$f(C, \Omega) = n^{-1} \text{trace}((\mathbb{Y} - \mathbb{X}C^T)^T (\mathbb{Y} - \mathbb{X}C^T) \Omega) - \log |\Omega| + P(C, \Omega),$$

where $P(C, \Omega)$ is a penalty term, and the $\mathbb{X} \in \mathbb{R}^{n \times p_0}$ and $\mathbb{Y} \in \mathbb{R}^{n \times q}$ are defined such that the i th row vector of $\mathbb{X} \in \mathbb{R}^{n \times p_0}$ ($\mathbb{Y} \in \mathbb{R}^{n \times q}$) is X_i (Y_i), and Ω is $q \times q$ positive-definite matrix representing the precision matrix of the error vector in the multivariate regression model. The penalized least square method penalizes the objective function when the estimate C deviates from the low dimensional structure which the true coefficient matrix C is assumed to have, and it is

beneficial under high-dimensional settings, where p_0 and q can grow to infinity as $n \rightarrow \infty$.

For example, as the penalty term $P(C, \Omega)$, the weighted sums of absolute value of elements in C and Ω have been suggested in [30], [35] and [24]. These penalty terms provide sparse estimations on the regression coefficient C and Ω .

There are also researches on the penalized estimation with the additional rank condition $rank(C) = r$ for $r < q$. For example, in [10], the coefficient C is decomposed as $C^T = BA^T$ where $B \in \mathbb{R}^{p_0 \times r}$ and $A \in \mathbb{R}^{r \times q}$ with $A^T A = I_r$. Then, they set the group lasso penalty on the row vectors of B to drop irrelevant predictor variables in the regression model. [9] suggested the penalty term as the sum of singular values of C and [33] used the penalty term on the singular vectors of C as well as the sum of singular values of C .

Employing a covariance estimation method is another approach for the estimation of the regression coefficient. The coefficient matrix C can be considered as a function of the joint covariance matrix of the covariate vector $X \in \mathbb{R}^{p_0}$ and the response vector $Y \in \mathbb{R}^q$. Assume that $Z = (X^T, Y^T)^T \in \mathbb{R}^{p_0+q}$ follows a joint distribution with a mean vector μ and a covariance matrix Σ such that

$$\begin{aligned} \mu &= \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix} \\ \Sigma &= \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{pmatrix}, \end{aligned}$$

where $\mu_X \in \mathbb{R}^{p_0}$, $\mu_Y \in \mathbb{R}^q$, $\Sigma_{XX} \in \mathbb{R}^{p_0 \times p_0}$ and $\Sigma_{YY} \in \mathbb{R}^{q \times q}$. Then, we have

$$\begin{aligned} (\mu_0, \Psi_0) &:= (\mu_Y - \Sigma_{YX} \Sigma_{XX}^{-1} \mu_X, \Sigma_{YX} \Sigma_{XX}^{-1}) \\ &= \operatorname{argmin}_{(\mu, \Psi)} \mathbb{E}\{(Y - \mu - \Psi X)(Y - \mu - \Psi X)^T\}, \end{aligned}$$

and $\mu_0 + \Psi_0 x$ is the conditional mean of Y given $X = x$ if Z follows the multivariate Gaussian distribution. Note that μ_0 is the zero vector if we assume that μ_X and μ_Y are zero vectors. In this case, the coefficient matrix C in the multivariate regression model corresponds to $\Psi_0 = \Sigma_{YX} \Sigma_{XX}^{-1}$, which is a function of the covariance matrix Σ and is called *the conditional mean operator*. Thus, estimators for covariance matrices can be used for the estimation of the conditional mean operator, which can be seen as a functional of a covariance matrix. Estimation methods for various functionals of covariance matrices have also been investigated in the literature. [13] suggested an optimal estimator for quadratic and l^r functionals of sparse covariance matrices. [14] considered optimal estimation of $\mu^T \Sigma^{-1} \mu$, where μ is mean vector and Σ is covariance matrix, under the sparsity assumption of $\Sigma^{-1} \mu$.

We need to consider a high-dimensional covariance estimation method when we use a covariance estimator for the multivariate regression model under high-dimensional settings. Suppose Z_1, Z_2, \dots, Z_n are independent and identically generated from a p -dimensional distribution with mean zero and covariance matrix Σ . We refer to the estimation of covariance Σ as high-dimensional covariance estimation when p is assumed to go to infinity as $n \rightarrow \infty$. Since traditional

covariance estimation methods, such as the sample covariance matrix and the Bayesian method by the inverse-Wishart prior, are not consistent when p is larger than n [18, 22], various structural assumptions on covariance matrices have been used to reduce the number of effective parameters. For example, the banded covariances [23], the bandable covariances [2], sparse covariances [5] and sparse spiked covariances [4] have been considered. These structural assumptions can be used in the joint covariance matrix of covariates and response variables when we employ covariance estimation for the multivariate regression under the high-dimensional settings.

In this paper, we consider the multivariate linear regression model, where the joint covariance of the covariate vector and the response vector is a bandable covariance matrix. Under the bandable covariance assumption, the farther apart two variables are, the smaller their covariance is. On the frequentist side, [6, 7] proved that *the tapering estimator of covariance* has the minimax optimal convergence rates for the class of bandable covariances under the spectral norm, Frobenius norm, and matrix l_1 norm. Therefore, a naive approach would be estimating the conditional mean operator based on the tapering estimator of covariance (or other minimax covariance estimators).

Unfortunately, even if a covariance estimator $\hat{\Sigma}$ has the minimax optimal convergence rate for the covariance Σ , it does not imply that $f(\hat{\Sigma})$ has also the minimax optimal convergence rate for $f(\Sigma)$ where f is a function on the space of covariances. Thus, the estimator for $\Sigma_{YX}\Sigma_{XX}^{-1}$ based on the tapering estimator of covariance may not have the minimax optimal convergence rate. Furthermore, there is no Bayesian method achieving the minimax posterior convergence rate for the class of bandable covariances. Note that [31], [20] and [23] proposed Bayesian procedures for banded covariances, but the class of bandable covariances considered in this paper is larger than the class of banded covariances.

We investigate the decision-theoretic property of the tapering estimator when the parameter of interest is the conditional mean operator, $\Sigma_{YX}\Sigma_{XX}^{-1}$, instead of the covariance itself under the normality assumption. As a naive plug-in estimator for the conditional mean, we first define *the tapering estimator of regression coefficient*, a tapering estimator of covariance plugged into the conditional mean operator. However, we show that the tapering estimator of the regression coefficient is sub-optimal under the bandable covariance assumption. This implies that a naive plug-in estimator is not enough to achieve the minimax rate. To resolve this issue, we propose a modified plug-in estimator, *the blockwise tapering estimator of regression coefficient*, and show that it obtains the minimax rate.

As a Bayesian procedure for the conditional mean operator under the bandable covariance assumption, we adopt the post-processed posterior method [23]. A post-processed posterior is a posterior constructed by transforming posterior samples from the initial posterior, which is typically a computationally convenient posterior. This idea is especially useful when it is difficult to impose a prior distribution on a restricted parameter space due to an unknown normalizing constant. For a given parameter space Θ^* , suppose that we are interested in restricted parameter space, $\Theta \subset \Theta^*$. A post-processed posterior can be ob-

tained by generating samples from an initial posterior on Θ^* and post-processing the posterior samples so that the transformed post-processed samples belong to Θ . When the post-processing function is a projection map from Θ^* to Θ , the method is called the posterior projection method. The posterior projection method has been suggested for various settings including [12], [17], [26] and [8], and was investigated in general aspects by [27]. The idea of transforming posterior samples was also used for the inference on covariance or precision matrices in [23] and [1].

We suggest two post-processed posteriors for the conditional mean operator. Both methods use the inverse-Wishart distribution as the initial prior distribution on the unconstrained covariance matrix space and use the tapering function and the blockwise tapering function as the post-processing functions for the conditional mean operator $\Sigma_{YX}\Sigma_{XX}^{-1}$. We present the asymptotic analysis to justify the proposed post-processed posteriors, and show that the post-processed posterior by the blockwise tapering function has the minimax optimal convergence rate.

The rest of the paper is organized as follows. In Section 2, we introduce the blockwise tapering estimator for the inference of the conditional mean operator under the bandable covariance assumption and show that this estimator has the minimax convergence rate. In Section 3, we introduce the post-processed posteriors for the conditional mean operator, and present the posterior convergence rates. Simulation studies and real data analysis are given in Section 4. Proofs of theorems in Sections 2 and 3 are given in Sections 5 and A, respectively. We conclude this paper with a discussion section.

2. Blockwise tapering estimator and minimax analysis

2.1. Notation

Let q, k and l be positive integers with $l \vee k \leq q$. For a $q \times q$ -matrix Σ and positive real numbers σ_{ij} , $1 \leq i, j \leq q$, let $\Sigma = (\sigma_{ij})_{1 \leq i, j \leq q} = (\sigma_{ij})$ when σ_{ij} is equal to the (i, j) element of Σ . We define sub-matrix operators $M_l^{(k)} : \mathbb{R}^{q \times q} \mapsto \mathbb{R}^{k^* \times k^*}$, where $k^* = \{(l + k - 1) \wedge q\} - (l \vee 1) + 1$, and $M_l^{*(k)} : \mathbb{R}^{q \times q} \mapsto \mathbb{R}^{q \times q}$ as

$$\begin{aligned} M_l^{(k)}(\Sigma) &= (\sigma_{ij})_{(l \vee 1) \leq i, j \leq \{(l+k-1) \wedge q\}} \\ M_l^{*(k)}(\Sigma) &= (\sigma_{ij} I[(l \vee 1) \leq i, j \leq \{(l+k-1) \wedge q\}])_{1 \leq i, j \leq q}, \end{aligned}$$

for $\Sigma = (\sigma_{ij})_{1 \leq i, j \leq q}$. Let $\Sigma_{a:b, c:d}$ be the sub-block matrix of $\Sigma \in \mathbb{R}^{q \times q}$ with $(a, a + 1, \dots, b - 1, b)$ rows and $(c, c + 1, \dots, d - 1, d)$ columns for positive integers a, b, c and d with $1 \leq a < b \leq q$ and $1 \leq c < d \leq q$, and let $\Sigma_{a:b} = \Sigma_{a:b, a:b}$. We also let $X_{a:b} = (x_a, x_{a+1}, \dots, x_{b-1}, x_b) \in \mathbb{R}^{b-a+1}$ for a vector $X = (x_1, x_2, \dots, x_q) \in \mathbb{R}^q$ and positive integers a and b with $1 \leq a < b \leq q$.

For a $q \times q$ -matrix $\Sigma = (\sigma_{ij})_{1 \leq i, j \leq q}$ and a positive integer k with $k \leq q$, define the tapering function $T_k(\Sigma)$, which was first defined in [6], as

$$T_k(\Sigma) = (w_{ij}^{(k)} \sigma_{ij})_{1 \leq i, j \leq q},$$

where

$$w_{ij}^{(k)} = \begin{cases} 1, & \text{when } |i - j| \leq k/2 \\ 2 - \frac{|i-j|}{k/2}, & \text{when } k/2 < |i - j| < k. \\ 0, & \text{otherwise} \end{cases}$$

For any sequences a_n and b_n of positive real numbers, we denote $a_n = o(b_n)$ if $\lim_{n \rightarrow \infty} a_n/b_n = 0$, and $a_n = O(b_n)$ if $\limsup_{n \rightarrow \infty} a_n/b_n = C$ for a positive constant C . We denote $a_n \preceq b_n$ if $a_n \leq Cb_n$ for all sufficiently large n and a positive constant C .

Let $\|\Sigma\| = \|\Sigma\|_2 = \{\lambda_{\max}(\Sigma\Sigma^T)\}^{1/2}$ be the spectral norm of a covariance matrix Σ , where $\lambda_{\max}(\Sigma)$ is the maximum eigenvalue of Σ . Given positive integers p and p_0 with $p_0 < p$, let $A_{XX} = A_{1:p_0, 1:p_0}$ and $A_{YX} = A_{(p_0+1):p, 1:p_0}$ for a positive $p \times p$ -matrix A . We also let $\Sigma_{0,XX}$ and $\Sigma_{0,YX}$ denote $(\Sigma_0)_{XX}$ and $(\Sigma_0)_{YX}$, respectively.

2.2. Blockwise tapering estimator

Let n , p and p_0 be positive integers with $p_0 < p$. Suppose Z_1, Z_2, \dots, Z_n are independent and identically distributed from a p -dimensional distribution with mean zero and covariance matrix Σ_0 , where $Z_i = (X_i^T, Y_i^T)^T$, $X_i \in \mathbb{R}^{p_0}$ and $Y_i \in \mathbb{R}^{p-p_0}$ for $i \in \{1, 2, \dots, n\}$. When only the first p_0 elements of Z_i , i.e. X_i , are given, the conditional mean vector for the other $p - p_0$ variables is

$$\Sigma_{0,YX}(\Sigma_{0,XX})^{-1}X_i. \quad (1)$$

The conditional mean operator $\Sigma_{0,YX}(\Sigma_{0,XX})^{-1}$ in (1) is the estimand we focus on in this paper. We define the transformation ψ from a covariance to the conditional mean operator as

$$\psi(\Sigma) := \psi(\Sigma; p_0) = \Sigma_{YX}(\Sigma_{XX})^{-1},$$

for $\Sigma \in \mathcal{C}_p$, where \mathcal{C}_p is the set of all $p \times p$ -dimensional positive definite matrices.

We assume Σ_0 belongs to a class of bandable covariances, \mathcal{F}_α , which is defined as

$$\begin{aligned} \mathcal{F}_\alpha &:= \mathcal{F}_{p,\alpha}(M, M_0, M_1) \\ &= \left\{ \Sigma = (\sigma_{ij})_{1 \leq i, j \leq p} \in \mathcal{C}_p : \sum_{(i,j): |i-j| \geq k} |\sigma_{ij}| \leq Mk^{-\alpha}, \forall k \geq 1, \right. \\ &\quad \left. \lambda_{\max}(\Sigma) \leq M_0, \lambda_{\min}(\Sigma) \geq M_1 \right\}, \end{aligned}$$

for some positive constants $\alpha, M > 0$ and $0 < M_1 < M_0$, where $\lambda_{\min}(\Sigma)$ is the minimum eigenvalue of Σ . [2] and [6] also considered the same class of bandable covariances except the minimum eigenvalue condition.

A natural estimator for $\psi(\Sigma_0)$ is the plug-in estimator, the tapering estimator of covariance plugged into ψ , because the tapering estimator achieves the

minimax optimal convergence rate for the class of bandable covariances under the spectral norm loss [6]. For the positive-definiteness is necessary for the covariance estimator, we modify the tapering estimator of covariance so that it is positive definite and call it *adjusted tapering estimator of covariance*:

$$T_k^{(\epsilon_n)}(S_n) := T_k(S_n) + ([\epsilon_n - \lambda_{\min}\{T_k(S_n)\}] \vee 0)I_p,$$

where $\epsilon_n > 0$ is the positive-definite adjustment parameter, S_n is the sample covariance matrix $\sum_{i=1}^n Z_i Z_i^T/n$, and I_p is the $p \times p$ identity matrix. We call the plug-in estimator with adjusted tapering estimator of covariance *the tapering estimator of regression coefficient*, in short *the tapering estimator*.

Since every column of the tapering estimator is not the zero vector with probability one, the tapering estimator uses all variables in a given covariate vector when the estimator is used as the regression coefficient. In other words, in the variable selection perspective, all variables are selected when the tapering estimator is used. Note that selecting out negligible covariates can increase the accuracy of a regression estimator, and partial correlations between covariates and responses have been used as a criterion for the variable selection [25, 3]. Theorem 2.1 shows that if a covariance matrix is bandable, then so is its inverse matrix. It essentially means that, under the bandable covariance assumption, the variables far away from each other tend to have weak partial correlations. The proof of Theorem 2.1 is given in Section 5.1.

Theorem 2.1. *Suppose $\Sigma_0 \in \mathcal{F}_{p,\alpha}(M, M_0, M_1)$, and let $\Sigma_0^{-1} = (w_{ij})$. There exist some positive constants C and λ depending only on M_0, M_1 and M such that*

$$\max_j \sum_i \{|w_{ij}| : |i - j| > ak \log k\} \leq C(k^{-a\lambda+1} + k^{-\alpha}),$$

for all $a > 0$ and all sufficiently large integer k with $p > k \vee (ak \log k)$.

Note [21] showed that the partial correlation between the i th and the j th variables is

$$r_{ij|[p]\setminus\{i,j\}} = \frac{w_{ij}}{\sqrt{w_{ii}w_{jj}}},$$

where $\Sigma_0^{-1} = (w_{ij})$ and $[p] = \{1, 2, \dots, p\}$. Since $|w_{ii}| \geq M_0^{-1}$ for all $i \in [p]$, by Theorem 2.1, each element in response vector $(Z_i)_{(p_0+1):p}$ has negligibly weak partial correlations with remote covariates $(Z_i)_j$ for $j \leq p_0$ and large $|j - p_0|$. Thus, selecting out these negligible covariates could yield a more accurate estimator for the conditional mean operator.

Based on the above argument, we propose *the blockwise tapering estimator of regression coefficient*, in short *the blockwise tapering estimator*. Let \mathbb{Z} be the set of all integers and $\lfloor x \rfloor = \max\{z \in \mathbb{Z} : z \leq x\}$. For positive real numbers a and ϵ_n , and a positive integer k with $2\lfloor ak \log k \rfloor \leq p_0$, define *the blockwise tapering estimator* as

$$\begin{aligned} \phi(S_n; 2\lfloor ak \log k \rfloor, \epsilon_n) &:= \phi(S_n; p_0, 2\lfloor ak \log k \rfloor, \epsilon_n) \\ &= T_k(S_n)_{YX} \Lambda^{(\epsilon_n)} \{T_k(S_n)_{XX}; 2\lfloor ak \log k \rfloor\}, \quad (2) \end{aligned}$$

where $\Lambda^{(\epsilon_n)}(A; b)$ is defined as for a $p_0 \times p_0$ matrix A

$$\Lambda^{(\epsilon_n)}(A; b) = \begin{pmatrix} O_{(p_0-b) \times (p_0-b)} & O_{(p_0-b) \times b} \\ O_{b \times (p_0-b)} & \{M_{p_0-b+1}^{(b)}(A) + [\epsilon_n - \lambda_{\min}\{M_{p_0-b+1}^{(b)}(A)\}] \vee 0\} I_b \end{pmatrix}^{-1},$$

where $O_{c \times d}$ is the $c \times d$ zero matrix for positive integers c and d . The definition of the blockwise tapering estimator gives

$$\begin{aligned} & \phi(S_n; 2\lfloor ak \log k \rfloor, \epsilon_n) \\ &= \begin{pmatrix} O_{(p-p_0) \times (p_0-2\lfloor ak \log k \rfloor)} & T_k(S_n)_{YX^{(2)}} \{T_k^{(\epsilon_n)}(S_{n, X^{(2)}X^{(2)}})\}^{-1} \end{pmatrix}, \end{aligned} \quad (3)$$

where $T_k(S_n)_{YX^{(2)}} = \{T_k(S_n)\}_{(p_0+1):p, (p_0-2\lfloor ak \log k \rfloor+1):p_0}$ and

$S_{n, X^{(2)}X^{(2)}} = (S_n)_{(p_0-2\lfloor ak \log k \rfloor+1):p_0}$, and we examine the blockwise tapering estimator using the formula (3). The first $p_0 - 2\lfloor ak \log k \rfloor$ columns of the estimator (3) are zero column vector, thus, given a covariate vector $x \in \mathbb{R}^{p_0}$, the blockwise tapering estimator drops the first $p_0 - 2\lfloor ak \log k \rfloor$ covariate variables $x_{1:p_0-2\lfloor ak \log k \rfloor}$. The remaining part of the estimator

$$T_k(S_n)_{YX^{(2)}} \{T_k^{(\epsilon_n)}(S_{n, X^{(2)}X^{(2)}})\}^{-1} \quad (4)$$

operates as the regression coefficient for the remaining covariate variables

$x_{p_0-2\lfloor ak \log k \rfloor+1:p_0}$, and (4) is almost the same as

$\psi(T_k^{(\epsilon_n)}(S_{n, (X^{(2)}, Y)(X^{(2)}, Y)}); 2\lfloor ak \log k \rfloor)$ except for the positive-definiteness adjustment part, where $S_{n, (X^{(2)}, Y)(X^{(2)}, Y)} = (S_n)_{p_0-2\lfloor ak \log k \rfloor+1:p}$.

Note that $\psi(T_k^{(\epsilon_n)}(S_{n, (X^{(2)}, Y)(X^{(2)}, Y)}); 2\lfloor ak \log k \rfloor)$ is the tapering estimator of regression coefficient with data $\{(X_i)_{p_0-2\lfloor ak \log k \rfloor+1:p_0}^T, Y_i^T\}^T; i = 1, \dots, n\}$. These observations show that the blockwise tapering estimator is constructed based on the variable selection and the idea of the tapering estimator of the regression coefficient.

2.3. Minimax analysis of blockwise tapering estimator

In this section, we give the convergence rates of the tapering and blockwise tapering estimators under the normality assumption, i.e., Z_1, Z_2, \dots, Z_n are independent and identically distributed from a p -dimensional Gaussian distribution with mean zero and covariance matrix Σ_0 , which is denoted by $N_p(0, \Sigma_0)$. We also show that the blockwise tapering estimator has the minimax convergence rate under the normality assumption. We use the loss function on $\mathbb{R}^{(p-p_0) \times p_0}$

$$L\{\hat{C}, \psi(\Sigma_0)\} = \|\hat{C} - \psi(\Sigma_0)\|_2, \quad (5)$$

for a pair of parameter $\psi(\Sigma_0)$ and estimator \hat{C} . The loss function gives the upper bound of the estimation error of $E(Y | X = x)$ given $x \in \mathbb{R}^{p_0}$, because

the definition of the operator norm gives

$$\begin{aligned} \|\hat{C}x - E(Y | X = x)\|_2 &= \|\{\hat{C} - \psi(\Sigma_0)\}x\|_2 \\ &\leq L\{\hat{C}, \psi(\Sigma_0)\}\|x\|_2. \end{aligned}$$

We show that the tapering estimator has a sub-optimal convergence rate under the loss function (5), while the blockwise tapering estimator has the minimax optimal convergence rate.

Theorem 2.2 gives the convergence rate of the tapering estimator. If we set ϵ_n such that $p^{1/2}5^{k/2}n \exp(-\lambda n) \preceq \epsilon_n^2 \preceq (k + \log p)/n$, then the convergence rate is $(k + \log p)/n + k^{-2\alpha}$, which is the same rate as the convergence rate of the tapering estimator of covariance [6]. The proof of this theorem is given in Section 5.2.

Theorem 2.2. *Suppose $\Sigma_0 \in \mathcal{F}_{p,\alpha}(M, M_0, M_1)$. Let k be a positive integer. If $k \vee \log p = o(n)$, $\epsilon_n = O(1)$ and $\lfloor k/2 \rfloor > \{4M/\lambda_{\min}(\Sigma_0)\}^{1/\alpha}$, then there exist some positive constants C and λ depending only on M, M_0, M_1 and α such that*

$$\begin{aligned} E_{\Sigma_0}(\|\psi(\Sigma_0) - \psi\{T_k^{(\epsilon_n)}(S_n)\}\|^2) \\ \leq C \left\{ k^{-2\alpha} + \frac{k + \log p}{n} + \epsilon_n^2 + \frac{p^{1/2}5^{k/2} \exp(-\lambda n)}{\epsilon_n^2} \right\}, \end{aligned}$$

for all sufficiently large n .

Next, we show the convergence rate of the blockwise tapering estimator. The blockwise tapering estimator is designed to estimate $\phi(\Sigma_0; 2\lfloor ak \log k \rfloor, 0)$ which approximates $\psi(\Sigma_0)$. Lemma 2.3 gives the approximation error, which is negligible when k is large enough. Based on the approximation error, the convergence rate of the blockwise tapering estimator is given in Theorem 2.4. If we set ϵ_n such that $(ak \log k)^{1/2}5^{k/2}n \exp(-\lambda n) \preceq \epsilon_n^2 \preceq k/n$, the convergence rate of the blockwise tapering estimator is $k/n + k^{-2\{\alpha \wedge (a\tau - 1)\}}$. The proofs of the lemma and theorem are given in Section 5.3.

Lemma 2.3. *Suppose $\Sigma_0 \in \mathcal{F}_{p,\alpha}(M, M_0, M_1)$. There exist some positive constants C and τ depending only on M, M_0 and M_1 such that*

$$\|\psi(\Sigma_0) - T_k(\Sigma_0)_{YX} \Lambda^{(0)}\{T_k(\Sigma_0, XX); 2\lfloor ak \log k \rfloor\}\| \leq C(k^{-\alpha} + k^{-a\tau+1}),$$

for all $a > 0$ and all sufficiently large integers k and p_0 with $\lfloor ak \log k \rfloor/2 \geq k$ and $2\lfloor ak \log k \rfloor < p_0$.

Theorem 2.4. *Suppose $\Sigma_0 \in \mathcal{F}_{p,\alpha}(M, M_0, M_1)$.*

If $k = o(n)$, $\lfloor k/2 \rfloor > \{4M/\lambda_{\min}(\Sigma_0)\}^{1/\alpha}$ and $\epsilon_n = O(1)$, then there exist some positive constants C, λ and τ depending only on M, M_0, M_1 and α such that

$$\begin{aligned} E_{\Sigma_0}(\|\psi(\Sigma_0) - \phi(S_n; 2\lfloor ak \log k \rfloor, \epsilon_n)\|^2) \\ \leq C \left\{ k^{-2\{\alpha \wedge (a\tau - 1)\}} + \frac{k}{n} + \epsilon_n^2 + \frac{(ak \log k)^{1/2}5^{k/2} \exp(-\lambda n)}{\epsilon_n^2} \right\}, \end{aligned}$$

for all $a > 0$ and all sufficiently large n , k and p_0 with $\lfloor ak \log k \rfloor / 2 \geq k$ and $p_0 > 2\lfloor ak \log k \rfloor$.

Remark 1. Contrary to Theorem 2.2, Theorem 2.4 does not require the upper bound of p_0 and p . Recall that the blockwise tapering estimator is designed to estimate $T_k(\Sigma_0)_{YX} \Lambda^{(0)}\{T_k(\Sigma_0, XX); 2\lfloor ak \log k \rfloor\}$, which is independent of p and p_0 . The factor $\Lambda^{(0)}\{T_k(\Sigma_0, XX); 2\lfloor ak \log k \rfloor\}$ is obviously dependent only on $(\Sigma_0)_{(p_0 - 2\lfloor ak \log k \rfloor + 1):p_0}$ by the definition. Furthermore $T_k(\Sigma_0)_{YX}$ depends only on $(\Sigma_0)_{(p_0 - k):(p_0 + k)}$ because (i) $T_k(\Sigma_0)$ is a k -banded matrix and (ii) $T_k(\Sigma_0)_{YX}$ is an off-diagonal block matrix. Thus, $T_k(\Sigma_0)_{YX} \Lambda^{(0)}\{T_k(\Sigma_0, XX); 2\lfloor ak \log k \rfloor\}$ depends only on $(\Sigma_0)_{(p_0 - 2\lfloor ak \log k \rfloor + 1):p_0}$ and $(\Sigma_0)_{(p_0 - k):(p_0 + k)}$. Note that the dimensions of the square matrices $(\Sigma_0)_{(p_0 - 2\lfloor ak \log k \rfloor + 1):p_0}$ and $(\Sigma_0)_{(p_0 - k):(p_0 + k)}$ are $2\lfloor ak \log k \rfloor$ and $2k + 1$, respectively, which are independent of p and p_0 . Since the blockwise tapering estimator is constructed based on the sample covariance estimators on $(\Sigma_0)_{(p_0 - 2\lfloor ak \log k \rfloor + 1):p_0}$ and $(\Sigma_0)_{(p_0 - k):(p_0 + k)}$, p_0 and p does not affect the convergence rate of the blockwise tapering estimator.

Next, we give the lower bound of the minimax risk for the conditional mean operator under the bandable covariance assumption to show that the blockwise tapering estimator is a minimax optimal estimator. Let $\hat{C} = \hat{C}(X_1, X_2, \dots, X_n)$ be an estimator on $\mathbb{R}^{(p-p_0) \times p_0}$. The minimax risk is defined as

$$\inf_{\hat{C}} \sup_{\Sigma_0 \in \mathcal{F}_\alpha} E \|\psi(\Sigma_0) - \hat{C}\|^2.$$

Theorem 2.5 gives a lower bound of the minimax risk as $n^{-2\alpha/(2\alpha+1)}$. If we set k , a and ϵ_n of the blockwise tapering estimator such that $k = n^{1/(2\alpha+1)}$, $a > (\alpha + 1)/\tau$ and $(ak \log k)^{1/2} 5^{k/2} n \exp(-\lambda n) \leq \epsilon_n^2 \leq k/n$, then the convergence rate is the same as the lower bound asymptotically. Thus, the minimax convergence rate is $n^{-2\alpha/(2\alpha+1)}$, and the blockwise tapering estimator attains the convergence rate. The proof of Theorem 2.5 is given in Section 5.4.

Theorem 2.5. *There exist some positive constants C and γ depending only on M , M_0 , M_1 and α such that*

$$\inf_{\hat{C}} \sup_{\Sigma_0 \in \mathcal{F}_\alpha} E \|\psi(\Sigma_0) - \hat{C}\|^2 \geq C n^{-2\alpha/(2\alpha+1)},$$

for all sufficiently large n and p_0 with $p_0 > \gamma n^{1/(2\alpha+1)}$

3. Blockwise tapering post-processed posterior

We propose the Bayesian counterparts of the tapering estimator and the blockwise tapering estimator using the post-processed posterior method [23]. The algorithm for the post-processed posteriors consists of the following two steps.

- (a) (Initial posterior sampling step) First, we obtain the initial conjugate posterior distribution on the unconstrained parameter space. We take the

inverse-Wishart distribution $IW_p(B_0, \nu_0)$ as the initial prior distribution of which density function is

$$\pi^i(\Sigma) \propto |\Sigma|^{-\nu_0/2} e^{-tr(\Sigma^{-1}B_0)/2}, \quad \Sigma \in \mathcal{C}_p,$$

where $B_0 \in \mathcal{C}_p$ and $\nu_0 > 2p$. Then, using the likelihood of Gaussian distribution, we obtain the initial posterior distribution $\pi^i(\Sigma|\mathbb{Z}_n)$ as $IW_p(B_0 + nS_n, \nu_0 + n)$, where n is the number of observations, $S_n = n^{-1} \sum_{i=1}^n Z_i Z_i^T$ and $\mathbb{Z}_n = (Z_1, \dots, Z_n)$. We generate $\Sigma^{(1)}, \Sigma^{(2)}, \dots, \Sigma^{(N)}$ from the initial posterior distribution.

- (b) (Post-processing step) Second, we post-process the samples from the initial posterior distribution with $\psi\{T_k^{(\epsilon_n)}(\cdot)\}$ or $\phi(\cdot; 2\lfloor ak \log k \rfloor, \epsilon_n)$, which are called *the tapering function* and *the blockwise tapering function*.

We call the post-processed posteriors obtained from the post-processing functions *the tapering post-processed posterior* (tapering PPP) and *the blockwise tapering post-processed posterior* (blockwise tapering PPP).

We use the decision-theoretic framework [22, 23] to prove the minimax optimality of the blockwise tapering post-processed posterior. We define P-loss $\mathcal{L}(\cdot, \cdot)$ and P-risk $\mathcal{R}(\cdot, \cdot)$ for the conditional mean operator as

$$\begin{aligned} \mathcal{L}\{\psi(\Sigma_0), \pi^{pp}(\cdot | \mathbb{Z}_n; f)\} &:= E^{\pi^i}(\|\psi(\Sigma_0) - f(\Sigma)\|^2 | \mathbb{Z}_n) \\ \mathcal{R}\{\psi(\Sigma_0), (\pi^i, f)\} &:= E_{\Sigma_0}\{E^{\pi^i}(\|\psi(\Sigma_0) - f(\Sigma)\|^2 | \mathbb{Z}_n)\}, \end{aligned}$$

where $\pi^{pp}(\cdot | \mathbb{Z}_n; f)$ is the post-processed posterior distribution derived from initial prior π^i and post-processing function f , and (π^i, f) is a pair of initial prior π^i and post-processing function f . Theorems 3.1 and 3.2 give the P-risk convergence rates of the tapering and the blockwise tapering post-processed posteriors, respectively. The convergence rates are the same as their frequentist counterparts. The proofs of these theorems are given in Section A.

Theorem 3.1. *Suppose $\Sigma_0 \in \mathcal{F}_{p,\alpha}(M, M_0, M_1)$. Let k be a positive integer, and let the prior π^i of Σ be $IW_p(A_n, \nu_n)$ for $A_n \in \mathcal{C}_p$ and $\nu_n > 2p$. If $\epsilon_n = O(1)$, $\lfloor k/2 \rfloor > \{4M/\lambda_{\min}(\Sigma_0)\}^{1/\alpha}$ and $k \vee \|A_n\| \vee (\nu_n - 2p) \vee \log p = o(n)$, then there exist positive constants C and λ depending only on M, M_0 and M_1 such that*

$$\begin{aligned} &E_{\Sigma_0}\{E^{\pi^i}(\|\psi(\Sigma_0) - \psi\{T_k^{(\epsilon_n)}(\Sigma)\}\|^2 | \mathbb{Z}_n)\} \\ &\leq C\left\{k^{-2\alpha} + \frac{k + \log p}{n} + \epsilon_n^2 + \frac{p^{1/2}5^{k/2} \exp(-\lambda n)}{\epsilon_n^2}\right\}, \end{aligned}$$

for all sufficiently large n and k .

Theorem 3.2. *Suppose $\Sigma_0 \in \mathcal{F}_{p,\alpha}(M, M_0, M_1)$. Let the prior π^i of Σ be $IW_p(A_n, \nu_n)$ for $A_n \in \mathcal{C}_p$ and $\nu_n > 2p$. If $\epsilon_n = O(1)$, $\lfloor k/2 \rfloor > \{4M/\lambda_{\min}(\Sigma_0)\}^{1/\alpha}$ and $k \vee \|A_n\| \vee (\nu_n - 2p) = o(n)$, then there exist positive constants C, τ and λ depending only on M, M_0 and M_1 such that*

$$\begin{aligned} &E_{\Sigma_0}\{E^{\pi^i}(\|\psi(\Sigma_0) - \phi(\Sigma; 2\lfloor ak \log k \rfloor, \epsilon_n)\|^2 | \mathbb{Z}_n)\} \\ &\leq C\left\{k^{-2(\alpha \wedge (a\tau - 1))} + \frac{k}{n} + \epsilon_n^2 + \frac{(ak \log k)^{1/2}5^{k/2} \exp(-\lambda n)}{\epsilon_n^2}\right\}, \end{aligned}$$

then for all $a > 0$ and all sufficiently large n , k and p_0 with $\lfloor ak \log k \rfloor / 2 > k$ and $p_0 > 2 \lfloor ak \log k \rfloor$.

Note that the P-risk minimax lower bound is $n^{-2\alpha/(2\alpha+1)}$ since the P-risk convergence rate is slower than or equal to the frequentist minimax rate [22]. Thus, if we set k , a and ϵ_n of the blockwise tapering post-processed posterior such that $k = n^{1/(2\alpha+1)}$, $(ak \log k)^{1/2} 5^{k/2} n \exp(-\lambda n) \preceq \epsilon_n^2 \preceq k/n$ and $a > (\alpha + 1)/\tau$, then the P-risk convergence rate is the same as the lower bound asymptotically. Thus, the P-risk minimax convergence rate is $n^{-2\alpha/(2\alpha+1)}$, and the blockwise tapering post-processed posterior attains the convergence rate.

4. Numerical studies

4.1. Simulation

We compare the blockwise tapering estimator with the tapering estimator using simulation data. We define the true covariance matrix $\Sigma_0 \in \mathbb{R}^{p \times p}$ as below. Let $\Sigma_0^* = (\sigma_{0,ij}^*)_{1 \leq i, j \leq p}$, where

$$\sigma_{0,ij}^* = \begin{cases} 1, & 1 \leq i = j \leq p \\ \rho^{|i-j|-(\alpha+1)}, & 1 \leq i \neq j \leq p \end{cases},$$

and let $\Sigma_0 = \Sigma_0^* + \{0.5 - \lambda_{\min}(\Sigma_0^*)\} I_p$, which guarantees the minimum eigenvalue of Σ_0 is bounded away from zero. We set $\rho = 0.6$ and $\alpha = 0.1$ for Σ_0 and generate data Z_1, \dots, Z_n from $N_p(0, \Sigma_0)$ independently, where $p \in \{500, 1000\}$ and $n = p/2$. Let $p_0 = 0.8p$ and fix the positive-definite adjustment parameter ϵ_n as 0.5. We define the error reduction value by choosing the blockwise tapering estimator over the tapering estimator as

$$d_f(S_n; k, a) = \|\psi\{T_k^{(\epsilon_n)}(S_n)\} - \psi(\Sigma_0)\| - \|\phi(S_n; 2 \lfloor ak \log k \rfloor, \epsilon_n) - \psi(\Sigma_0)\|.$$

We repeat generating the simulation data T times, and let $\mathbb{Z}_n^{(i)}$ and $S_n^{(i)}$ denote the data and the sample covariance matrix, respectively, in the i th repetition for $i \in \{1, 2, \dots, T\}$. We summarize the error reduction values from the repetitions as t-value

$$t_f(k, a; T) = \frac{\sum_{i=1}^T d_f(S_n^{(i)}; k, a)/T}{[\sum_{i=1}^T \{d_f(S_n^{(i)}; k, a) - \sum_{i=1}^T d_f(S_n^{(i)}; k, a)/T\}^2 / T]^{1/2}},$$

which is the performance measure for the comparison between the tapering and blockwise tapering estimators. We also compare the blockwise tapering PPP with the tapering PPP for the same simulation data. We define the error reduction value by choosing the blockwise tapering PPP as

$$d_b(\mathbb{Z}_n; k, a) = \|\hat{C}^{(TPPP)} - \psi(\Sigma_0)\| - \|\hat{C}^{(bTPPP)} - \psi(\Sigma_0)\|,$$

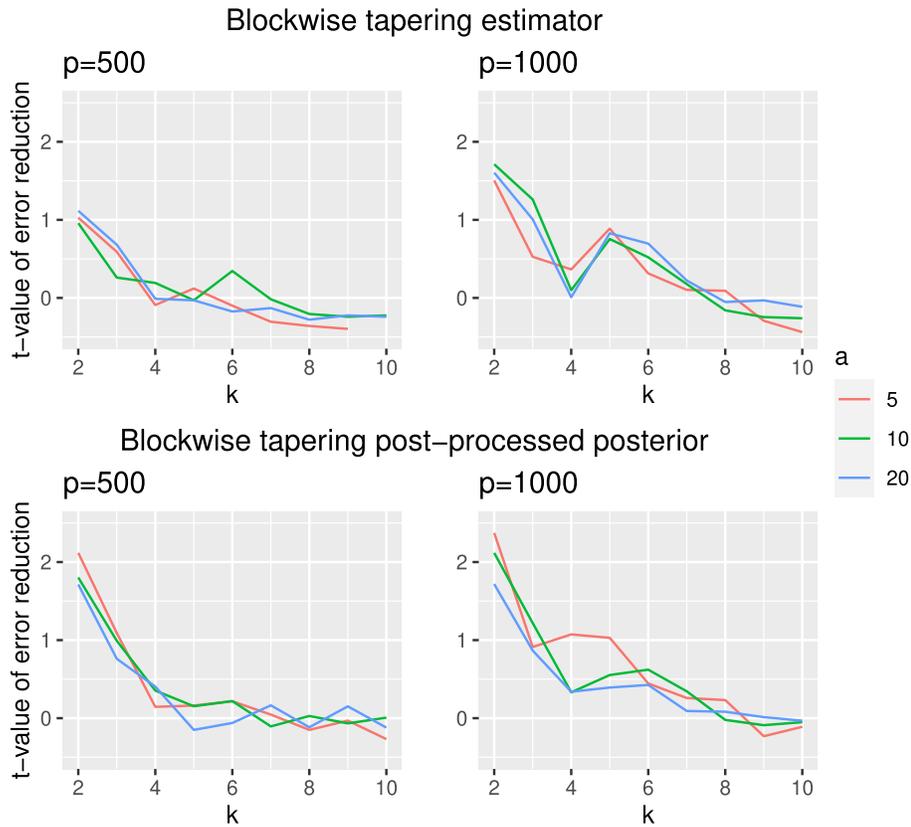


FIG 1. The evaluated $t_f(k, a; 100)$, the summarized error reductions by choosing the blockwise tapering estimator over the tapering estimator, are represented in the upper plots. The dimension of the covariance p is set to 500 and 1000. As the tuning parameters of the methods, $k \in \{2, 3, \dots, 10\}$ and $a \in \{5, 10, 20\}$ are used. The evaluated $t_b(k, a; 100)$, the summarized error reductions by choosing the blockwise tapering post-processed posterior over the tapering post-processed posterior, are represented in the lower plots for the same parameters.

where $\hat{C}^{(TPPP)}$ and $\hat{C}^{(bTPPP)}$ are the posterior means of the tapering PPP and the blockwise tapering PPP, respectively. We define the t-value for T repetitions as

$$t_b(k, a; T) = \frac{\sum_{i=1}^T d_b(\mathbb{Z}_n^{(i)}; k, a)/T}{[\sum_{i=1}^T \{d_b(\mathbb{Z}_n^{(i)}; k, a) - \sum_{i=1}^T d_b(\mathbb{Z}_n^{(i)}; k, a)/T\}^2/T]^{1/2}}.$$

We evaluate $t_f(k, a; 100)$ and $t_b(k, a; 100)$ for $k \in \{2, 3, \dots, 10\}$ and $a \in \{5, 10, 20\}$. For the post-processed posteriors we generate 1000 posterior samples in each setting. We represent the result of the evaluations in Figure 1. When p is large and k is small, the effects of error reductions by the blockwise tapering estimator and the blockwise tapering post-processed posterior increase. Note

that the convergence rates of the tapering estimator and the tapering PPP contain the additional $\log p/n$ term. The effect of the additional term is increased when another term in the convergence rate, k/n , is relatively small. Thus, the error reduction is effective when p is large compared to k . The figure also shows that the tapering estimator is slightly better otherwise. If $\log p$ is not relatively large, one does not need to abandon the covariates $X_{1:(p_0-2\lfloor ak \log k \rfloor)}$ by using the blockwise tapering estimator or the blockwise tapering PPP.

Next, we compare the tapering estimator, blockwise tapering estimator, and their Bayesian versions with two other methods: covariance estimation method and multivariate regression method. A covariance estimator can be used for the estimation of the conditional mean operator by applying the transformation (2). We use the banding estimator [2], dual maximum likelihood estimator [19], and the banding post-processed posterior [23] as covariance estimators for comparison. The multivariate regression method is also used for comparison, since the multivariate linear regression coefficient is the conditional mean operator. We adopt the reduced-rank regression [9], the sparse reduced-rank regression [10] and the method of sparse orthogonal factor regression (SOFAR) [33].

We need to select tuning parameters for the conditional mean operator estimators. Based on the tuning parameter selection process, we divide the estimation methods into three categories: frequentist covariance-based method, post-processed posterior method, and multivariate regression method.

The tapering and blockwise tapering estimators belong to the frequentist covariance-based method, and the process of the tuning parameter selection is as follows. When a covariance estimator is given, the conditional mean operator and the conditional variance are derived, which yield the conditional distribution under the normality assumption. The log-likelihood function of the conditional distribution is used for the leave-one-out cross-validation. Let $\hat{\Sigma}(\mathbb{Z}_{n,-i}, \tau)$ be a frequentist covariance estimator based on $\mathbb{Z}_{n,-i} = (Z_1, \dots, Z_{i-1}, Z_{i+1}, \dots, Z_n)$ given a tuning parameter vector τ . The tuning parameter τ is (k, ϵ) when the tapering estimator is considered and τ is (k, ϵ, a) when the blockwise tapering estimator is considered. The derived conditional mean operator is $\psi\{\hat{\Sigma}(\mathbb{Z}_{n,-i}, \tau)\}$, and the conditional variance is

$$\begin{aligned} \nu\{\hat{\Sigma}(\mathbb{Z}_{n,-i}, \tau)\} &:= \hat{\Sigma}(\mathbb{Z}_{n,-i}, \tau)_{YY} \\ &- \hat{\Sigma}(\mathbb{Z}_{n,-i}, \tau)_{YX} \{\hat{\Sigma}(\mathbb{Z}_{n,-i}, \tau)_{XX}\}^{-1} \hat{\Sigma}(\mathbb{Z}_{n,-i}, \tau)_{XY}. \end{aligned}$$

We select τ as the minimizer of

$$\hat{R}^{(f)}(\tau) = \sum_{i=1}^n \log p[Y_i \mid \psi\{\hat{\Sigma}(\mathbb{Z}_{n,-i}, \tau)\} X_i, \nu\{\hat{\Sigma}(\mathbb{Z}_{n,-i}, \tau)\}],$$

where $(X_i^T, Y_i^T)^T = Z_i$ and $p(x \mid \mu, \Sigma)$ is the density function of the multivariate normal distribution with mean μ and covariance Σ . Since the conditional variance can not be derived from the blockwise tapering estimator, we use the conditional variance from the tapering estimator in this case.

For the tuning parameter selection of the post-processed posterior methods, we use the Bayesian leave-one-out cross-validation method [15] to the

log-likelihood function of the conditional distribution. Let $\Sigma_1^{(i)}, \Sigma_2^{(i)}, \dots, \Sigma_S^{(i)}$ be leave-one-out initial posterior samples which are generated from the initial posterior by $\mathbb{Z}_{n,-i}$ for $i \in \{1, 2, \dots, n\}$. We select the tuning parameter vector τ as the minimizer of

$$\sum_{i=1}^n \log \frac{1}{S} \sum_{s=1}^S p\{Y_i \mid \psi^*(\Sigma_s^{(i)}; \tau) X_i, \nu^*(\Sigma_s^{(i)}; \tau)\},$$

where ψ^* and ν^* are post-processing functions for the conditional mean operator and conditional variance given the tuning parameter τ , respectively. For the post-processing function of the conditional variance ν^* , the banding PPP uses $\nu\{B_k^{(\epsilon_n)}(\Sigma_s^{(i)})\}$, where $B_k^{(\epsilon_n)}$ is the positive-definite adjusted banding operator defined as

$$B_k^{(\epsilon_n)}(\Sigma) = B_k(\Sigma) + ([\epsilon_n - \lambda_{\min}\{B_k(\Sigma)\}] \vee 0) I_p,$$

and the tapering and blockwise tapering PPPs use $\nu\{T_k^{(\epsilon_n)}(\Sigma_s^{(i)})\}$.

For the multivariate regression method, we use 10-fold cross-validation method as [10], [9] and [33] suggested. Note that all the methods contain the rank parameter in the tuning parameters. While we select the rank from $\{0, 1, \dots, 10\}$ for the reduced-rank regression, $\{1, 2, \dots, 10\}$ is considered for the others. Because if the rank is zero, all the three methods coincide, we only consider the [9]'s method for the zero rank case.

We set $p = 200$, $\rho = 0.6$ and $\alpha = 0.1, 0.3$ for Σ_0 and generate Z_1, Z_2, \dots, Z_n from $N_p(0, \Sigma_0)$ independently for $n \in \{100, 200\}$. We repeat generating the simulation data 100 times for each simulation setting. The performance of each method is measured as

$$\frac{1}{100} \sum_{s=1}^{100} \|\psi(\Sigma_0) - \hat{C}_s\|,$$

where \hat{C}_s is the point estimator for the conditional mean operator in the s th repetition. For the post-processed posterior methods, we use the posterior mean as the point estimator. Table 1 gives the simulation error. The tapering estimator and the blockwise tapering estimator, and their Bayesian counterparts are the best in all settings. The multivariate regression methods, i.e. the reduced-rank regression, sparse reduced-rank regression and sparse orthogonal factor regression, are the worst in all settings. Unlike the other covariance-based methods, the bandable or banded covariance structure is not considered in the multivariate regression methods. It appears that the multivariate regression framework does not perform well under the high-dimensional bandable covariance assumption.

4.2. Application to forecasting traffic speed

We apply the proposed methods to multivariate regression analysis for small area spatio-temporal data, and use this application to forecast traffic speed in Yeoui-daero, a road in Seoul.

TABLE 1
Spectral norm errors of estimators for the conditional mean operator.

	$n = 100$		$n = 200$	
	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.1$	$\alpha = 0.3$
Tapering estimator	0.255	0.240	0.206	0.188
Blockwise tapering estimator	0.255	0.240	0.206	0.188
Banding estimator	0.319	0.293	0.290	0.247
Dual maximum likelihood estimator	0.365	0.357	0.319	0.273
Tapering post-processed posterior	0.257	0.246	0.208	0.193
Blockwise tapering post-processed posterior	0.257	0.246	0.208	0.193
Banding post-processed posterior	0.323	0.323	0.303	0.256
Reduced-rank regression	0.509	0.488	0.509	0.488
Sparse reduced-rank regression	3.324	3.309	2.818	2.902
Sparse orthogonal factor regression	1.823	1.819	1.679	1.703

Suppose spatio-temporal data are observed in S spatial regions and T times, where S and T are positive integers. Let $X_{s,t}$ be a random variable at s th spatial index and t th time index, $s = 1, \dots, S$ and $t = 1, \dots, T$. We assume

$$\begin{aligned} & E[\{X_{s_1,t_1} - E(X_{s_1,t_1})\}\{X_{s_2,t_2} - E(X_{s_2,t_2})\}] \\ & \leq r(|t_1 - t_2|), \quad s_1, s_2 \in \{1, \dots, S\}, t_1, t_2 \in \{1, \dots, T\}, \end{aligned} \quad (6)$$

where r is a real-valued function from the non-negative integer space, and is assumed to be a decreasing function. Rearranging $(X_{s,t})_{s=1,\dots,S,t=1,\dots,T}$, we define $Z \in \mathbb{R}^{TS}$ as

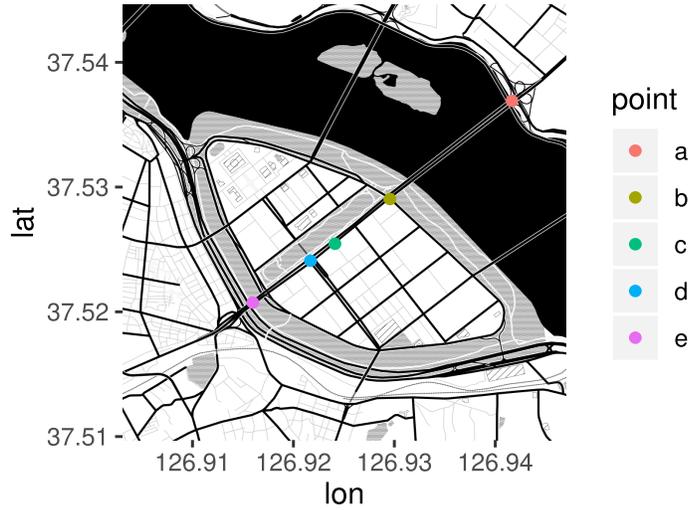
$$Z = (X_{1,1}, X_{2,1}, \dots, X_{S,1}, X_{1,2}, X_{2,2}, \dots, X_{S,T}), \quad (7)$$

and let $E(ZZ^T) = \Sigma_0 = (\sigma_{0,ij})$. We show that Σ_0 is a bandable covariance, if the decreasing rate of $r(x)$ is $x^{-\alpha-1}$. Note that if $mS \leq |i - j| < (m + 1)S$ for $m \in \{0, 1, \dots, T - 1\}$ and $i, j \in \{1, 2, \dots, TS\}$, then the time index difference between Z_i and Z_j is at least m . This observation and assumption (6) give $|\sigma_{0,ij}| \leq r(\lfloor |i - j|/S \rfloor)$ and

$$\begin{aligned} \sup_j \sum_i \{|\sigma_{0,ij}| : |i - j| \geq k\} & \leq \sup_j \sum_i \{|\sigma_{0,ij}| : |i - j| \geq \lfloor k/S \rfloor S\} \\ & \leq \sup_j \sum_{m=\lfloor k/S \rfloor}^{T-1} \sum_{i \in I_m^{(j)}} |\sigma_{0,ij}| \\ & \leq S \sum_{m=\lfloor k/S \rfloor}^{T-1} r(m), \end{aligned}$$

where $I_m^{(j)} = \{i \in \{1, 2, \dots, p\} : \lfloor |i - j|/S \rfloor = m\}$. If the decreasing rate of $r(x)$ is $x^{-\alpha-1}$, then

$$\sup_j \sum_i \{|\sigma_{0,ij}| : |i - j| \geq k\} \leq Ck^{-\alpha},$$



index	1	2	3	4	5	6	7	8
route	a→b	b→c	c→d	d→e	e→d	d→c	c→b	b→a

FIG 2. The eight routes in Yeoui-daero and their index allocation.

for some positive constant C . Thus, Σ_0 is a bandable covariance, and the proposed methods for the conditional mean operator under the bandable covariance assumption can be used to predict $X_{1:S,t_0+1:T}$ given $X_{1:S,1:t_0}$.

Based on the rearrangement (7) and the proposed methods for the conditional mean operator under the bandable covariance assumption, we forecast traffic speed in Yeoui-daero using data from TOPIS [32]. In the traffic speed data in Yeoui-daero, a daily data set consists of observations in 8 spatial indexes and 24-time indexes. Let a daily traffic speed be $(X_{s,t})_{1 \leq s \leq 8, 1 \leq t \leq 24}$, where time index t indicates the time interval from $(t - 1)$ o'clock to t o'clock, and the allocation of the spatial index s is given in Figure 2. We rearrange $(X_{s,t})_{1 \leq s \leq 8, 1 \leq t \leq 24}$ as (7), and apply the proposed estimators to forecast $X_{1:8,18:24}$ given $X_{1:8,1:17}$.

In the data from TOPIS, we use data from January to October in 2020, excluding weekend data sets and missing data sets. We have 172 days observations which are denoted by $Z_1, Z_2, \dots, Z_{172} \in \mathbb{R}^{192}$. To apply the proposed methods, we use mean-centered observations $\tilde{Z}_1, \tilde{Z}_2, \dots, \tilde{Z}_{172} \in \mathbb{R}^{192}$. For the performance measure, let the training data be $Z_{train} = (\tilde{Z}_1, \tilde{Z}_2, \dots, \tilde{Z}_{86})$, and the test data be $Z_{test} = (\tilde{Z}_{87}, \tilde{Z}_{88}, \dots, \tilde{Z}_{172})$. The forecast errors are summarized as

$$\frac{1}{86} \sum_{i=87}^{172} \|\hat{C}(Z_{train})(\tilde{Z}_i)_{1:p_0} - (\tilde{Z}_i)_{(p_0+1):p}\|_2,$$

TABLE 2

The root mean square errors of forecast results for traffic speed in Yeoui-daero.

Method	Error
Tapering estimator	2.80
Blockwise tapering estimator	2.80
Banding estimator	2.85
Dual maximum likelihood estimator	3.33
Tapering post-processed posterior	2.78
Blockwise tapering post-processed posterior	2.78
Banding post-processed posterior	2.87
Reduced-rank regression	3.59
Sparse reduced-rank regression	3.39
Sparse orthogonal factor regression	4.31

where $p_0 = 17 \times 8$ and $\hat{C}(\mathbb{Z}_{train})$ is a point estimator for the conditional mean operator based on \mathbb{Z}_{train} . The summarized forecast errors are represented in Table 2, which shows the tapering and blockwise tapering estimators and their Bayesian versions are the best among all methods.

5. Proofs of theorems and lemma

In this section, the proofs of Theorems 2.1- 2.5 and Lemma 2.3 are given. The proofs of the other theorems, Theorems 3.1 and 3.2, are given in Section A. We give notations for the proofs. Let $\|\Sigma\|_F = \text{tr}(\Sigma\Sigma^T)$ and $\|\Sigma\|_r$ be the Frobenius norm and the matrix r -norm for a covariance matrix Σ , respectively. We also let $W_p(B_0, \nu_0)$ be Wishart distribution of which density function is

$$\pi(\Sigma) \propto |\Sigma|^{(\nu_0 - p - 1)/2} e^{-\text{tr}(B_0^{-1}\Sigma)/2}, \quad \Sigma \in \mathcal{C}_p,$$

where $B_0 \in \mathcal{C}_p$ and $\nu_0 > p - 1$.

For a $q \times q$ -matrix $\Sigma = (\sigma_{ij})_{1 \leq i, j \leq q}$ and a positive integer k with $k \leq q$, define the banding function $B_k(\Sigma)$ as $B_k(\Sigma) = (\sigma_{ij}I(|i - j| \leq k))_{1 \leq i, j \leq q}$.

5.1. Proof of Theorem 2.1

In this subsection, we give the proof of Theorem 2.1. First, we present Lemmas 5.1 and 5.2 which are necessary for the proof of Theorem 2.1. Proofs of these lemmas are given in Section A.

Lemma 5.1. *Let p and k be positive integers with $k < p$. Suppose $\Sigma \in \mathcal{C}_p$ is a k -band matrix, and let $\Sigma^{-1} = (w_{ij})$. For all $a > 0$ and all sufficiently large k with $p > k \vee (ak \log k)$,*

$$\max_j \sum_i \{|w_{ij}| : |i - j| > ak \log k\} \leq -\frac{2C}{\log q} k^{2a \log q + 1},$$

where $q = (\kappa^{1/2} - 1)/(\kappa^{1/2} + 1)$, $C = \{\|\Sigma^{-1}\| \vee (1 + \kappa^{1/2})^2\}/(2\|\Sigma\|)$, and κ is the spectral condition number of Σ .

Lemma 5.2. *Let p and k be positive integers with $k < p$ and suppose $\Sigma_0 \in \mathcal{F}_{p,\alpha}(M, M_0, M_1)$. There exist some positive constants C_1 and C_2 depending only on M, M_0 and M_1 such that*

$$\begin{aligned} \|\Sigma_0^{-1} - B_k(\Sigma_0)^{-1}\|_1 &\leq C_1 k^{-\alpha} \\ \|\Sigma_0^{-1} - T_k(\Sigma_0)^{-1}\|_1 &\leq C_2 (\lfloor k/2 \rfloor)^{-\alpha}, \end{aligned}$$

for all sufficiently large k .

The proofs of these lemmas are given in Section A.

Now we prove Theorem 2.1.

Proof of Theorem 2.1. Let $B_k(\Sigma_0)^{-1} = (w_{ij}^*)$. We have

$$\begin{aligned} \sum_{i \in I_j} |w_{ij}| &\leq \sum_{i \in I_j} |w_{ij}^*| + \sum_{i \in I_j} |w_{ij} - w_{ij}^*| \\ &\leq \sum_{i \in I_j} |w_{ij}^*| + \|\Sigma_0^{-1} - B_k(\Sigma_0)^{-1}\|_1, \end{aligned}$$

where $I_j = \{i \in \{1, 2, \dots, p\} : |i - j| > ak \log k\}$. By Lemma 5.2, there exists some positive constant C_1 depending only on M, M_0 and M_1 such that

$$\|\Sigma_0^{-1} - B_k(\Sigma_0)^{-1}\|_1 \leq C_1 k^{-\alpha}, \tag{8}$$

for all sufficiently large k .

For the upper bound of $\sum_{i \in I_j} |w_{ij}^*|$, Lemma 5.1 gives that, for all sufficiently large k with $p > k \vee (ak \log k)$,

$$\max_j \sum_i \{|w_{ij}^*| : |i - j| > ak \log k\} \leq -\frac{2C_2}{\log q} k^{2\alpha \log q + 1}, \tag{9}$$

where $q = (\kappa^{1/2} - 1)/(\kappa^{1/2} + 1)$, $C_2 = \{\|\Sigma^{-1}\| \vee (1 + \kappa^{1/2})^2\}/(2\|\Sigma\|)$ and κ is the spectral condition number of Σ . By collecting the inequalities (8) and (9), we complete the proof. \square

5.2. Proof of Theorem 2.2

In this subsection, we prove Theorem 2.2 which gives the convergence rate of the tapering estimator. First, we present Lemmas 5.3-5.6 necessary for the proof of Theorem 2.2. Proofs of these lemmas are given in Section A.

Lemma 5.3. *Let Σ and Σ_0 be $p \times p$ -symmetric matrices and $\Sigma^{(\epsilon_n)} = \Sigma + [\{\epsilon_n - \lambda_{\min}(\Sigma)\} \vee 0]I_p$. For all positive integer r and all positive real number $\epsilon_n > 0$,*

$$\|\Sigma^{(\epsilon_n)} - \Sigma_0\|^r \leq 2^{2r-1} \|\Sigma - \Sigma_0\|^r + 4^{r-1} |\epsilon_n|^r.$$

Lemma 5.4. *Let n , k and p be positive integers with $k \leq p$ and suppose $\Sigma_0 \in \mathcal{F}_{p,\alpha}(M, M_0, M_1)$. If $c \leq \lambda_{\min}(\Sigma_0)/2$ and $\lfloor k/2 \rfloor > \{4M/\lambda_{\min}(\Sigma_0)\}^{1/\alpha}$, then*

$$P_{\Sigma_0}[\lambda_{\min}\{T_k(S_n)\} \leq c] \leq 2p5^k \exp(-\lambda n),$$

for some positive constant λ depending only on M_0 and M_1 .

Lemma 5.5. *Let n , p and k be positive integers with $k \leq p$ and suppose $\Sigma_0 \in \mathcal{F}_{p,\alpha}(M, M_0, M_1)$. If $k \vee \log p = o(n)$ and $\epsilon_n = O(1)$, then there exists some positive constant C depending only on M_0 and M_1 such that*

$$E_{\Sigma_0}(\|T_k^{(\epsilon_n)}(S_n) - \Sigma_0\|^4) \leq C,$$

for all sufficiently large n .

Lemma 5.6. *Suppose $\Sigma_0 \in \mathcal{F}_{p,\alpha}(M, M_0, M_1)$. Let q and k be positive integers with $k < q \leq p$, and let $(\Sigma_0)_{1:q,1:q}$ and $\{T_k^{(\epsilon_n)}(S_n)\}_{1:q,1:q}$ denoted by $\Sigma_{0,11}$ and $T_k^{(\epsilon_n)}(S_n)_{11}$, respectively. If $k \vee \log p = o(n)$, $\lfloor k/2 \rfloor > \{4M/\lambda_{\min}(\Sigma_0)\}^{1/\alpha}$ and $\epsilon_n = O(1)$, then there exist some positive constants C and λ depending only on M , M_0 , M_1 and α such that*

$$\begin{aligned} & E_{\Sigma_0}(\|T_k^{(\epsilon_n)}(S_n)_{11}^{-1} - \Sigma_{0,11}^{-1}\|^2) \\ & \leq C \left\{ \frac{k + \log p}{n} + k^{-2\alpha} + \epsilon_n^2 + \frac{p^{1/2}5^{k/2} \exp(-\lambda n)}{\epsilon_n^2} \right\}, \end{aligned}$$

for all sufficiently large n .

The proofs of these lemmas are given in Section A.

Now, we give the proof of Theorem 2.2.

Proof of Theorem 2.2. We have

$$\begin{aligned} & E(\|\Sigma_{0,YX} \Sigma_{0,XX}^{-1} - \psi\{T_k^{(\epsilon_n)}(S_n)\}\|^2) \\ & \leq 2E(\|T_k^{(\epsilon_n)}(S_n)_{YX} - \Sigma_{0,YX}\|^2 \|T_k^{(\epsilon_n)}(S_n)_{XX}^{-1}\|^2) \\ & \quad + 2\|\Sigma_{0,YX}\|^2 E(\|T_k^{(\epsilon_n)}(S_n)_{XX}^{-1} - \Sigma_{0,XX}^{-1}\|^2) \\ & \leq \frac{2^3}{M_1^2} E(\|T_k^{(\epsilon_n)}(S_n)_{YX} - \Sigma_{0,YX}\|^2 I[\lambda_{\min}\{T_k(S_n)_{XX}\} > M_1/2]) \quad (10) \end{aligned}$$

$$+ \frac{2}{\epsilon_n^2} E(\|T_k^{(\epsilon_n)}(S_n)_{YX} - \Sigma_{0,YX}\|^2 I[\lambda_{\min}\{T_k(S_n)_{XX}\} \leq M_1/2]) \quad (11)$$

$$+ 2M_0^2 E(\|T_k^{(\epsilon_n)}(S_n)_{XX}^{-1} - \Sigma_{0,XX}^{-1}\|^2), \quad (12)$$

where the last inequality holds since $\lambda_{\min}\{T_k^{(\epsilon_n)}(S_n)_{XX}\} \geq \lambda_{\min}\{T_k^{(\epsilon_n)}(S_n)\} \geq \epsilon_n$. We have that there exists some positive constant C_1 depending only on M , M_0 , M_1 and α such that

$$(10) \leq \frac{2^3}{M_1^2} E(\|T_k^{(\epsilon_n)}(S_n) - \Sigma_0\|^2)$$

$$\begin{aligned} &\leq \frac{2^3}{M_1^2} \{2^3 E(\|T_k(S_n) - \Sigma_0\|^2) + 4\epsilon_n^2\} \\ &\leq C_1 \left\{ \frac{k + \log p}{n} + k^{-2\alpha} + \epsilon_n^2 \right\}, \end{aligned}$$

for all sufficiently large n . The second and last inequalities hold by Lemma 5.3 and Theorem 2 of [6], respectively. By Lemmas 5.4 and 5.5, there exist some positive constants C_2 and λ_1 depending only on M_0 and M_1 such that

$$\begin{aligned} (11) \quad &\leq \frac{2}{\epsilon_n^2} E(\|T_k^{(\epsilon_n)}(S_n) - \Sigma_0\|^4)^{1/2} P(\lambda_{\min}(T_k(S_n)_{XX}) \leq M_1/2)^{1/2} \\ &\leq \frac{C_2}{\epsilon_n^2} p^{1/2} 5^{k/2} \exp(-\lambda_1 n), \end{aligned}$$

for all sufficiently large n . Finally, by Lemma 5.6, we get the upper bound of (12) as

$$(12) \leq C_3 \left(\frac{k + \log p}{n} + k^{-2\alpha} + \epsilon_n^2 + \frac{p^{1/2} 5^{k/2} \exp(-\lambda_2 n)}{\epsilon_n^2} \right),$$

for some positive constants C_3 and λ_2 depending only on M , M_0 , M_1 and α . Combining the upper bounds of (10), (11) and (12), we complete the proof. \square

5.3. Proofs of Lemma 2.3 and Theorem 2.4

In this subsection, we show the convergence rate of the blockwise tapering estimator by proving Lemma 2.3 and Theorem 2.4. First, we present Lemma 5.7.

Lemma 5.7. *Let q and k be positive integers with $k < q$, and $A \in \mathcal{C}_q$ be a k -band matrix. For positive real numbers a and b with $q > \lfloor ak \log k \rfloor + \lfloor bk \log k \rfloor$, we express A and A^{-1} as*

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, A^{-1} = \Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix},$$

where $A_{22}, \Omega_{22} \in \mathbb{R}^{(\lfloor ak \log k \rfloor + \lfloor bk \log k \rfloor) \times (\lfloor ak \log k \rfloor + \lfloor bk \log k \rfloor)}$. There exist some positive constants λ and C depending only on $\|A\|$ and $\|A^{-1}\|$ such that

$$\|M_{\lfloor ak \log k \rfloor + 1}^{(\lfloor bk \log k \rfloor)}(\Omega_{22}) - M_{\lfloor ak \log k \rfloor + 1}^{(\lfloor bk \log k \rfloor)}(A_{22}^{-1})\| \leq Ck^{-a\lambda+1},$$

for all sufficiently large k and q with $\lfloor ak \log k \rfloor \geq k$ and $q > \lfloor ak \log k \rfloor + \lfloor bk \log k \rfloor$.

The proof of this lemma is given in Section A.

Next, we provide the proofs of Lemma 2.3 and Theorem 2.4.

Proof of Lemma 2.3. There exist some positive constants C_1 and C_2 depending only on M , M_0 and M_1 such that

$$\|\psi(\Sigma_0) - T_k(\Sigma_0)_{YX} T_k(\Sigma_0, XX)^{-1}\|$$

$$\begin{aligned}
 &\leq \|\Sigma_{0,XX}^{-1}\| \|T_k(\Sigma_0)_{YX} - \Sigma_{0,YX}\| \\
 &\quad + \|T_k(\Sigma_0)_{YX}\| \|\Sigma_{0,XX}^{-1} - T_k(\Sigma_{0,XX})^{-1}\| \\
 &\leq M_1^{-1}M(\lfloor k/2 \rfloor)^{-\alpha} + (\|\Sigma_0\| + \|T_k(\Sigma_0) - \Sigma_0\|)\|\Sigma_{0,XX}^{-1} - T_k(\Sigma_{0,XX})^{-1}\| \\
 &\leq M_1^{-1}M(\lfloor k/2 \rfloor)^{-\alpha} + \{M_0 + M(\lfloor k/2 \rfloor)^{-\alpha}\}C_1(\lfloor k/2 \rfloor)^{-\alpha} \\
 &\leq C_2(\lfloor k/2 \rfloor)^{-\alpha},
 \end{aligned}$$

for all sufficiently large k . The third inequality holds by Lemma 5.2 since $\Sigma_{0,XX} \in \mathcal{F}_{p_0,\alpha}(M, M_0, M_1)$. Then, we have

$$\begin{aligned}
 &\|\psi(\Sigma_0) - T_k(\Sigma_0)_{YX}\Lambda^{(0)}\{T_k(\Sigma_{0,XX}); 2\lfloor ak \log k \rfloor\}\| \\
 &\leq \|\psi(\Sigma_0) - T_k(\Sigma_0)_{YX}T_k(\Sigma_{0,XX})^{-1}\| \\
 &\quad + \|T_k(\Sigma_0)_{YX}[\Lambda^{(0)}\{T_k(\Sigma_{0,XX}); 2\lfloor ak \log k \rfloor\} - T_k(\Sigma_{0,XX})^{-1}]\| \\
 &\leq C_2(\lfloor k/2 \rfloor)^{-\alpha} + \|T_k(\Sigma_0)_{YX}[\Lambda^{(0)}\{T_k(\Sigma_{0,XX}); \\
 &\quad 2\lfloor ak \log k \rfloor\} - T_k(\Sigma_{0,XX})^{-1}]\|.
 \end{aligned}$$

For this upper bound, we have

$$\begin{aligned}
 &\|T_k(\Sigma_0)_{YX}[\Lambda^{(0)}\{T_k(\Sigma_{0,XX}); 2\lfloor ak \log k \rfloor\} - T_k(\Sigma_{0,XX})^{-1}]\| \\
 &\leq \|T_k(\Sigma_0)_{YX}[T_k(\Sigma_{0,XX})^{-1} - M_{p_0-\lfloor ak \log k \rfloor+1}^{*(\lfloor ak \log k \rfloor)}\{T_k(\Sigma_{0,XX})^{-1}\}]\| \quad (13) \\
 &\quad + \|T_k(\Sigma_0)_{YX}[M_{p_0-\lfloor ak \log k \rfloor+1}^{*(\lfloor ak \log k \rfloor)}\{T_k(\Sigma_{0,XX})^{-1}\} \\
 &\quad \quad - \Lambda^{(0)}\{T_k(\Sigma_{0,XX}); 2\lfloor ak \log k \rfloor\}]\|, \quad (14)
 \end{aligned}$$

First, we show the upper bound of (13). Since $T_k(\Sigma_0)$ is a k -band matrix, $T_k(\Sigma_0)_{YX}$ is expressed as

$$T_k(\Sigma_0)_{YX} = \begin{bmatrix} O_{k \times (p_0-k)} & \{T_k(\Sigma_0)_{YX}\}_{1:k,(p_0-k+1):p_0} \\ O_{(p-p_0-k) \times (p_0-k)} & O_{(p-p_0-k) \times k} \end{bmatrix}. \quad (15)$$

Since $\lfloor ak \log k \rfloor \geq k$, the definition of $M_{p_0-\lfloor ak \log k \rfloor+1}^{*(\lfloor ak \log k \rfloor)}$ gives

$$\begin{aligned}
 &T_k(\Sigma_{0,XX})^{-1} - M_{p_0-\lfloor ak \log k \rfloor+1}^{*(\lfloor ak \log k \rfloor)}\{T_k(\Sigma_{0,XX})^{-1}\} \\
 &= \begin{bmatrix} B^* & C^* \\ \{T_k(\Sigma_{0,XX})^{-1}\}_{(p_0-k+1):p_0,1:(p_0-\lfloor ak \log k \rfloor)} & O_{k \times \lfloor ak \log k \rfloor} \end{bmatrix},
 \end{aligned}$$

for some $(p_0 - k) \times (p_0 - \lfloor ak \log k \rfloor)$ -matrix B^* and some $(p_0 - k) \times \lfloor ak \log k \rfloor$ -matrix C^* . Thus,

$$\begin{aligned}
 &T_k(\Sigma_0)_{YX}[T_k(\Sigma_{0,XX})^{-1} - M_{p_0-\lfloor ak \log k \rfloor+1}^{*(\lfloor ak \log k \rfloor)}\{T_k(\Sigma_{0,XX})^{-1}\}] \\
 &= \begin{bmatrix} T_{11}^{(B)} & O_{k \times \lfloor ak \log k \rfloor} \\ O_{(p-p_0-k) \times (p_0-\lfloor ak \log k \rfloor)} & O_{(p-p_0-k) \times \lfloor ak \log k \rfloor} \end{bmatrix},
 \end{aligned}$$

where $T_{11}^{(B)} = \{T_k(\Sigma_0)_{YX}\}_{1:k, (p_0-k+1):p_0} \{T_k(\Sigma_0, XX)^{-1}\}_{(p_0-k+1):p_0, 1:(p_0-\lfloor ak \log k \rfloor)}$, and

$$(13) = \|\{T_k(\Sigma_0)_{YX}\}_{1:k, (p_0-k+1):p_0} \{T_k(\Sigma_0, XX)^{-1}\}_{(p_0-k+1):p_0, 1:(p_0-\lfloor ak \log k \rfloor)}\| \leq \|T_k(\Sigma_0)_{YX}\| \|\{T_k(\Sigma_0, XX)^{-1}\}_{(p_0-k+1):p_0, 1:(p_0-\lfloor ak \log k \rfloor)}\|.$$

Let $\{T_k(\Sigma_0, XX)\}^{-1} = (b_{ij})_{1 \leq i, j \leq p_0}$. Every element b_{ij} in $\{T_k(\Sigma_0, 11)^{-1}\}_{(p_0-k+1):p_0, 1:(p_0-\lfloor ak \log k \rfloor)}$ satisfies

$$\begin{aligned} |i - j| &\geq \lfloor ak \log k \rfloor - k + 1 \\ &> \lfloor ak \log k \rfloor / 2, \end{aligned}$$

since $\lfloor ak \log k \rfloor / 2 \geq k$. Thus, by Lemma 5.1, there exist some positive constants C_3 and λ_1 depending only on M_0 and M_1 such that

$$\begin{aligned} &\|\{T_k(\Sigma_0, XX)^{-1}\}_{(p_0-k+1):p_0, 1:(p_0-\lfloor ak \log k \rfloor)}\|_2 \\ &\leq \|\{T_k(\Sigma_0, XX)^{-1}\}_{(p_0-k+1):p_0, 1:(p_0-\lfloor ak \log k \rfloor)}\|_1 \\ &\quad \vee \|\{T_k(\Sigma_0, XX)^{-1}\}_{(p_0-k+1):p_0, 1:(p_0-\lfloor ak \log k \rfloor)}\|_\infty \\ &\leq C_3 k^{-a\lambda_1+1}, \end{aligned}$$

for all sufficiently large k with $p_0 > k \vee \lfloor ak \log k \rfloor / 2$. Since

$$\|T_k(\Sigma_0)_{YX}\| \leq \|\Sigma_0\| + \|\Sigma_0 - T_k(\Sigma_0)\| \leq M_0 + M \lfloor k/2 \rfloor,$$

we get

$$(13) \leq (M_0 + M \lfloor k/2 \rfloor) C_2 k^{-a\lambda_1+1}, \tag{16}$$

for all sufficiently large k with $p_0 > k \vee \lfloor ak \log k \rfloor / 2$.

Next, we show the upper bound of (14). Since $T_k(\Sigma_0)_{YX}$ is expressed as (15), we have

$$(14) \leq \|T_k(\Sigma_0)\| \|\{M_{p_0-\lfloor ak \log k \rfloor+1}^{*(\lfloor ak \log k \rfloor)} \{T_k(\Sigma_0, XX)^{-1}\} - \Lambda^{(0)}\{T_k(\Sigma_0, XX); 2\lfloor ak \log k \rfloor\}\}_{(p_0-k+1):p_0, 1:p_0}\|. \tag{17}$$

Note

$$\begin{aligned} &[M_{p_0-\lfloor ak \log k \rfloor+1}^{*(\lfloor ak \log k \rfloor)} \{T_k(\Sigma_0, XX)^{-1}\} - \Lambda^{(0)}\{T_k(\Sigma_0, XX); \\ &2\lfloor ak \log k \rfloor\}]_{(p_0-k+1):p_0, 1:p_0} \\ &= [O_{k \times (p_0-2\lfloor ak \log k \rfloor)} \quad -D_1^* \quad D_3^* - D_2^*], \end{aligned}$$

where D_1^* , D_2^* and D_3^* are $k \times \lfloor ak \log k \rfloor$ -matrices with

$$\begin{aligned} D_1^* &= [M_{p_0-2\lfloor ak \log k \rfloor+1}^{(2\lfloor ak \log k \rfloor)} \{T_k(\Sigma_0, XX)^{-1}\}]_{(2\lfloor ak \log k \rfloor-k+1):2\lfloor ak \log k \rfloor, 1:\lfloor ak \log k \rfloor} \\ D_2^* &= [M_{p_0-2\lfloor ak \log k \rfloor+1}^{(2\lfloor ak \log k \rfloor)} \{T_k(\Sigma_0, XX)^{-1}\}]_{(2\lfloor ak \log k \rfloor-k+1):2\lfloor ak \log k \rfloor, (\lfloor ak \log k \rfloor+1):2\lfloor ak \log k \rfloor} \\ D_3^* &= \{T_k(\Sigma_0, XX)^{-1}\}_{(p_0-k+1):p_0, (p_0-\lfloor ak \log k \rfloor+1):p_0}. \end{aligned}$$

We have

$$(14) \leq \|T_k(\Sigma_0)\|(\|D_1^*\| + \|D_2^* - D_3^*\|).$$

First, we show the upper bound of $\|D_2^* - D_3^*\|$. Let

$$\begin{aligned} A_{22} &= M_{p_0 - 2\lfloor ak \log k \rfloor + 1}^{(2\lfloor ak \log k \rfloor)} \{T_k(\Sigma_{0,XX})\} \\ \Omega_{22} &= M_{p_0 - 2\lfloor ak \log k \rfloor + 1}^{(2\lfloor ak \log k \rfloor)} \{T_k(\Sigma_{0,XX})^{-1}\}. \end{aligned}$$

By Lemma 5.7, there exist some positive constants C_4 and λ_2 depending only on M_0 and M_1 such that

$$\begin{aligned} \|D_2^* - D_3^*\| &\leq \|M_{\lfloor ak \log k \rfloor + 1}^{(\lfloor ak \log k \rfloor)}(\Omega_{22}) - M_{\lfloor ak \log k \rfloor + 1}^{(\lfloor ak \log k \rfloor)}(A_{22}^{-1})\| \\ &\leq C_3 k^{-a\lambda_2 + 1}, \end{aligned}$$

for all sufficiently large k with $\lfloor ak \log k \rfloor \geq k$ and $p_0 > 2\lfloor ak \log k \rfloor$.

Next, we show the upper bound of $\|D_1^*\|$. Let $M_{p_0 - 2\lfloor ak \log k \rfloor + 1}^{(2\lfloor ak \log k \rfloor)} \{T_k(\Sigma_{0,XX})\}^{-1} = (c_{ij})_{1 \leq i, j \leq 2\lfloor ak \log k \rfloor}$. Every element c_{ij} in D_1^* satisfies

$$\begin{aligned} |i - j| &\geq \lfloor ak \log k \rfloor - k + 1 \\ &\geq \lfloor ak \log k \rfloor / 2, \end{aligned}$$

since $\lfloor ak \log k \rfloor / 2 \geq k$. Thus, Lemma 5.1 gives that there exist some positive constants C_5 and λ_3 depending only on M_0 and M_1 such that

$$\|D_1^*\| \leq C_5 k^{-a\lambda_3 + 1},$$

for all sufficiently large k with $p_0 > k \vee (ak \log k)$. By combining this inequality with (17), we have

$$(14) \leq (M_0 + M(\lfloor k/2 \rfloor)^{-\alpha})(C_2 k^{-a\lambda_2 + 1} + C_3 k^{-a\lambda_3 + 1}). \tag{18}$$

By collecting the inequalities (16) and (18), the proof is completed. □

Proof of Theorem 2.4. We have

$$\begin{aligned} &E(\|\psi(\Sigma_0) - T_k(S_n)_{YX} \Lambda^{(\epsilon_n)} \{T_k(S_n)_{XX}; 2\lfloor ak \log k \rfloor\}\|^2) \\ &\leq 2\|\psi(\Sigma_0) - T_k(\Sigma_0)_{YX} \Lambda^{(0)} \{T_k(\Sigma_{0,XX}); 2\lfloor ak \log k \rfloor\}\|^2 \end{aligned} \tag{19}$$

$$+ 4E(\|T_k(\Sigma_0)_{YX} - T_k(S_n)_{YX}\|^2 \|\Lambda^{(\epsilon_n)} \{T_k(S_n)_{XX}; 2\lfloor ak \log k \rfloor\}\|^2) \tag{20}$$

$$\begin{aligned} &+ 4\|T_k(\Sigma_0)\|^2 E(\|\Lambda^{(0)} \{T_k(\Sigma_{0,XX}); 2\lfloor ak \log k \rfloor\} \\ &- \Lambda^{(\epsilon_n)} \{T_k(S_n)_{XX}; 2\lfloor ak \log k \rfloor\}\|^2). \end{aligned} \tag{21}$$

By Lemma 2.3, there exist some positive constants C_1 and λ_1 depending only on M , M_0 , M_1 and α such that

$$(19) \leq C_1 k^{-2(\alpha \wedge (a\lambda_1 - 1))},$$

for all sufficiently large k with $\lfloor ak \log k \rfloor / 2 \geq k$ and $p_0 > 2\lfloor ak \log k \rfloor$.

Next, we show the upper bound of (20). Let $\tilde{M} := M_{p_0 - 2\lfloor ak \log k \rfloor + 1}^{(2\lfloor ak \log k \rfloor)}$ in this proof. By the definition of $\Lambda^{(\epsilon_n)}$, we have

$$\|\Lambda^{(\epsilon_n)}\{T_k(S_n)_{XX}; 2\lfloor ak \log k \rfloor\}\| = \|T_k^{(\epsilon_n)}[\tilde{M}\{(S_n)_{XX}\}]^{-1}\|,$$

and

$$\begin{aligned} (20) &= 4E(\|T_k(\Sigma_0)_{YX} - T_k(S_n)_{YX}\|^2 \|T_k^{(\epsilon_n)}[\tilde{M}\{(S_n)_{XX}\}]^{-1}\|^2) \\ &\leq \frac{4}{\epsilon_n^2} E\{\|T_k(\Sigma_0)_{YX} - T_k(S_n)_{YX}\|^2 I(\lambda_{\min}[\tilde{M}\{(S_n)_{XX}\}] < M_1/2)\} \\ &\quad + \frac{16}{M_1^2} E(\|T_k(\Sigma_0)_{YX} - T_k(S_n)_{YX}\|^2) \\ &\leq \frac{4}{\epsilon_n^2} E(\|T_k\{M_{p_0-k}^{(2k+1)}(\Sigma_0)\} - T_k\{M_{p_0-k}^{(2k+1)}(S_n)\}\|^4)^{1/2} \\ &\quad \times P\{\lambda_{\min}[\tilde{M}\{(S_n)_{XX}\}] < M_1/2\}^{1/2} \\ &\quad + \frac{16}{M_1^2} E(\|T_k\{M_{p_0-k}^{(2k+1)}(\Sigma_0)\} - T_k\{M_{p_0-k}^{(2k+1)}(S_n)\}\|^2) \\ &\leq \frac{4}{\epsilon_n^2} E(2^3 \|M_{p_0-k}^{(2k+1)}(\Sigma_0) - T_k\{M_{p_0-k}^{(2k+1)}(S_n)\}\|^4 + 2^3 M^4 (\lfloor k/2 \rfloor)^{-4\alpha})^{1/2} \\ &\quad \times P\{\lambda_{\min}[\tilde{M}\{(S_n)_{XX}\}] < M_1/2\}^{1/2} \\ &\quad + \frac{16}{M_1^2} E(2\|M_{p_0-k}^{(2k+1)}(\Sigma_0) - T_k\{M_{p_0-k}^{(2k+1)}(S_n)\}\|^2 + 2M^2 (\lfloor k/2 \rfloor)^{-2\alpha}) \\ &\leq C_2 \left\{ \frac{1}{\epsilon_n^2} (ak \log k)^{1/2} 5^{k/2} \exp(-\lambda_2 n) + \frac{k + \log k}{n} + k^{-2\alpha} \right\}, \end{aligned}$$

for some positive constants C_2 and λ_2 depending only on M , M_0 , M_1 and α . The first inequality holds since

$$\lambda_{\min}(T_k^{(\epsilon_n)}[\tilde{M}\{(S_n)_{XX}\}]) \geq \epsilon_n.$$

The second inequality holds since $T_k(A)_{YX}$ for a matrix A is determined only by $M_{p_0-k}^{(2k+1)}(A)$. Note that $T_k(A)$ is a k -banded matrix and $T_k(A)_{YX}$ is the off-diagonal block matrix. The third inequality holds since

$$\begin{aligned} \|T_k(\Sigma_0) - \Sigma_0\|_1 &\leq \|B_{\lfloor k/2 \rfloor}(\Sigma_0) - \Sigma_0\|_1 \\ &\leq M(\lfloor k/2 \rfloor)^{-\alpha}. \end{aligned} \tag{22}$$

The last inequality holds by Lemmas 5.4, 5.5 and Theorem 2 of [6]. For the upper bound of (21), we have that there exists some positive constant C_3 depending only on M , M_0 and M_1 such that

$$\begin{aligned} &E(\|\Lambda^{(0)}\{T_k(\Sigma_0)_{XX}; 2\lfloor ak \log k \rfloor\} - \Lambda^{(\epsilon_n)}\{T_k(S_n)_{XX}; 2\lfloor ak \log k \rfloor\}\|^2) \\ &= E(\|T_k\{\tilde{M}(\Sigma_0)_{XX}\}^{-1} - T_k^{(\epsilon_n)}[\tilde{M}\{(S_n)_{XX}\}]^{-1}\|^2) \end{aligned}$$

$$\begin{aligned} &\leq 2E(\|\tilde{M}(\Sigma_{0,XX})^{-1} - T_k^{(\epsilon_n)}[\tilde{M}\{(S_n)_{XX}\}]^{-1}\|^2) \\ &\quad + 2\|T_k\{\tilde{M}(\Sigma_{0,XX})\}^{-1} - \tilde{M}(\Sigma_{0,XX})^{-1}\|^2 \\ &\leq 2E(\|\tilde{M}(\Sigma_{0,XX})^{-1} - T_k^{(\epsilon_n)}[\tilde{M}\{(S_n)_{XX}\}]^{-1}\|^2) + C_3([k/2])^{-2\alpha}, \end{aligned}$$

for all sufficiently large k . The first equality is satisfied by the definition of $\Lambda^{(0)}$ and $\Lambda^{(\epsilon_n)}$. The last inequality holds by Lemma 5.2 since $\tilde{M}(\Sigma_{0,XX}) \in \mathcal{F}_{2\lfloor ak \log k \rfloor, \alpha}(M, M_0, M_1)$. We apply Lemma 5.6 and obtain that there exist some positive constants C_4 and λ_3 depending only on M, M_0, M_1 and α such that

$$\begin{aligned} &E(\|\tilde{M}(\Sigma_{0,XX})^{-1} - T_k^{(\epsilon_n)}[\tilde{M}\{(S_n)_{XX}\}]^{-1}\|^2) \\ &\leq C_4 \left\{ \frac{k + \log(2ak \log k)}{n} + \epsilon_n^2 + k^{-2\alpha} + \frac{(2ak \log k)^{1/2} 5^{k/2} \exp(-\lambda_1 n)}{\epsilon_n^2} \right\}, \end{aligned}$$

for all sufficiently large n . Thus, there exists some positive constant C_5 depending only on M, M_0, M_1 and α such that

$$(21) \leq C_5 \left\{ \frac{k + \log(2ak \log k)}{n} + \epsilon_n^2 + k^{-2\alpha} + \frac{(2ak \log k) 5^k \exp(-\lambda_1 n)}{\epsilon_n^2} \right\},$$

for all sufficiently large n and k . Collecting the upper bounds of (19), (20) and (21), we complete the proof. \square

5.4. Proof of Theorem 2.5

In this subsection, we prove Theorem 2.5 which gives the lower bound of the minimax risk for the conditional mean operator under the bandable covariance assumption. For probability measures P_θ and $P_{\theta'}$, we define

$$\begin{aligned} \|P_\theta - P_{\theta'}\|_1 &= \int |p_{\theta'} - p_\theta| d\nu \\ \|P_\theta \wedge P_{\theta'}\| &= \int p_{\theta'} \wedge p_\theta d\nu, \end{aligned}$$

where p_θ and $p_{\theta'}$ are probability density functions of P_θ and $P_{\theta'}$, respectively, with respect to the reference measure ν . First, we provide Lemma 5.8 which is a reformulation of the proof of Lemma 6 in [6].

Lemma 5.8. *Suppose Σ and Σ' are $p \times p$ -positive definite matrices. Let P_Σ be the joint distribution of X_1, X_2, \dots, X_n , which are independent and identically generated from $N_p(0, \Sigma)$. If $\|\Sigma - \Sigma'\|_2 (\|\Sigma^{-1}\|_2 \wedge \|\Sigma'^{-1}\|_2) < 1/2$, then*

$$\|P_\Sigma - P_{\Sigma'}\|_1^2 \leq n(\|\Sigma^{-1}\|_2 \wedge \|\Sigma'^{-1}\|_2)^2 \|\Sigma - \Sigma'\|_F^2.$$

The proof of this lemma is given in Section A.

Proof of Theorem 2.5. Let $\mathcal{C}_{p,p_0} = \{A \in \mathbb{R}^{p \times p} : A_{1:p_0, 1:p_0} \in \mathcal{C}_{p_0}\}$, and $(\mathcal{C}_{p,p_0})^\chi$ be the space of estimators on \mathcal{C}_{p,p_0} , where χ is the sample space. Since for an

arbitrary $B \in \mathbb{R}^{(p-p_0) \times p_0}$ there exists $A_B \in \mathcal{C}_{p,p_0}$ such that $\psi(A_B; p_0) = B$, it suffices to show

$$\inf_{\hat{A} \in (\mathcal{C}_{p,p_0})^{\times}} \sup_{\Sigma_0 \in \mathcal{F}_{p,\alpha}} E(\|\psi(\Sigma_0; p_0) - \psi(\hat{A}; p_0)\|^2) \geq Cn^{-2\alpha/(2\alpha+1)},$$

for some positive constant C . Let $d(\hat{\Sigma}, \Sigma_0) = \|\hat{\Sigma}_{YX} \hat{\Sigma}_{XX}^{-1} - \Sigma_{0,YX} \Sigma_{0,XX}^{-1}\|_2$, which is a semimetric on \mathcal{C}_{p,p_0} . For a positive integer k with $k < p_0/2$, let $\Theta = \{0, 1\}^k$, and let $H(\theta, \theta')$ be the Hamming distance on Θ . We define

$$\Sigma(\theta) = \bar{M}I_p + \tau \sum_{m=1}^k \theta_m B(m + p_0; k), \quad \theta = (\theta_1, \theta_2, \dots, \theta_k) \in \Theta,$$

where $\tau = \{M(2k)^{-\alpha-1}\} \wedge \{(M_0 - \bar{M})/(2k)\}$, $\bar{M} = (M_0 + M_1)/2$, $B(l; k) = (b_{ij})_{1 \leq i, j \leq p}$ and

$$b_{ij} = I(i = l \text{ and } l - 2k \leq j \leq l - 1, \text{ or } j = l \text{ and } l - 2k \leq i \leq l - 1).$$

Note that $\Sigma(\theta) \in \mathcal{F}_{p,\alpha}(M, M_0, M_1)$ for all $\theta \in \Theta$. Thus,

$$\begin{aligned} & \inf_{\hat{A} \in (\mathcal{C}_{p,p_0})^{\times}} \sup_{\Sigma_0 \in \mathcal{F}_{p,\alpha}} E\|\psi(\Sigma_0; p_0) - \psi(\hat{A}; p_0)\|^2 \\ & \geq \inf_{\hat{A} \in (\mathcal{C}_{p,p_0})^{\times}} \sup_{\theta \in \Theta} E\|\psi(\Sigma(\theta); p_0) - \psi(\hat{A}; p_0)\|^2 \\ & = \inf_{\hat{A} \in (\mathcal{C}_{p,p_0})^{\times}} \sup_{\theta \in \Theta} E[d^2\{\Sigma(\theta), \hat{A}\}] \end{aligned} \tag{23}$$

By the Assouad lemma [6, Lemma 4], we have, for all $s > 0$,

$$\sup_{\theta \in \Theta} 2^s E[d^s\{\hat{A}, \Sigma(\theta)\}] \geq \min_{H(\theta, \theta') \geq 1} \frac{d^s\{\Sigma(\theta), \Sigma(\theta')\}}{H(\theta, \theta')} \frac{k}{2} \min_{H(\theta, \theta')=1} \|P_\theta \wedge P_{\theta'}\|, \tag{24}$$

where P_θ is the joint distribution of n independent observations from the multivariate normal distribution with mean zero and covariance $\Sigma(\theta)$.

First, we show the lower bound of

$$\frac{d^s\{\Sigma(\theta), \Sigma(\theta')\}}{H(\theta, \theta')}.$$

For vector $v = \{I(p_0 - k < i \leq p_0)\}_{1 \leq i \leq p_0}$ and all $\theta, \theta' \in \Theta$ with $\theta \neq \theta'$, we have

$$\begin{aligned} d^2\{\Sigma(\theta), \Sigma(\theta')\} & \geq \frac{\|\{\Sigma(\theta)_{YX} - \Sigma(\theta')_{YX}\}(\bar{M}I_{p_0})^{-1}v\|^2}{\|v\|^2} \\ & \geq \bar{M}^{-2} \frac{\|\{\Sigma(\theta)_{YX} - \Sigma(\theta')_{YX}\}v\|^2}{\|v\|^2} \\ & \geq \bar{M}^{-2} H(\theta, \theta') \tau^2 k, \end{aligned}$$

which implies

$$\min_{H(\theta, \theta') \geq 1} \frac{d^2\{\Sigma(\theta), \Sigma(\theta')\}}{H(\theta, \theta')} \geq \bar{M}^{-2}\tau^2k. \quad (25)$$

Next, we consider the lower bound of $\min_{H(\theta, \theta')=1} \|P_\theta \wedge P_{\theta'}\|$. We assume $H(\theta, \theta') = 1$, $2\tau k < \bar{M}$, $2\tau k/(\bar{M} - 2\tau k) < 1/2$ and $(\sqrt{2}n^{1/2}k^{1/2}\tau)/(\bar{M} - 2\tau k) \leq 1/2$. We have

$$\begin{aligned} \|\Sigma(\theta) - \Sigma(\theta')\|_2 &\leq \|\Sigma(\theta) - \Sigma(\theta')\|_1 \\ &\leq 2\tau k, \end{aligned}$$

and

$$\begin{aligned} \|\Sigma(\theta)^{-1}\| &= \lambda_{\min}\{\Sigma(\theta)\}^{-1} \\ &\leq \frac{1}{\bar{M} - \|\Sigma(\theta) - \bar{M}I_p\|} \\ &\leq \frac{1}{\bar{M} - 2\tau k}, \end{aligned}$$

where the first inequality holds by Lemma 4.14 in [23]. Since $2\tau k/(\bar{M} - 2\tau k) < 1/2$, Lemma 5.8 gives

$$\begin{aligned} \min_{H(\theta, \theta')=1} \|P_\theta \wedge P_{\theta'}\| &= 1 - \max_{H(\theta, \theta')=1} \|P_\theta - P_{\theta'}\|_{1/2} \\ &\geq 1 - \frac{n^{1/2}\|\Sigma(\theta) - \Sigma(\theta')\|_F}{2(\bar{M} - 2\tau k)} \\ &\geq 1 - \frac{\sqrt{2}n^{1/2}k^{1/2}\tau}{\bar{M} - 2\tau k}, \end{aligned} \quad (26)$$

where the last inequality holds since $\|\Sigma(\theta) - \Sigma(\theta')\|_F \leq (8k\tau^2)^{1/2}$.

Since $(\sqrt{2}n^{1/2}k^{1/2}\tau)/(\bar{M} - 2\tau k) \leq 1/2$,

$$\min_{H(\theta, \theta')=1} \|P_\theta \wedge P_{\theta'}\| \geq 1/2.$$

Thus, collecting inequalities (23), (24) and (25), we get

$$\inf_{\hat{\Sigma}} \max_{\theta \in \{0,1\}^k} 2^2 \mathbb{E}_\theta d^2\{\hat{\Sigma}, \Sigma(\theta)\} \geq ck^2\tau^2,$$

for some positive constant c . Since $\tau \leq M(2k)^{-\alpha-1}$ we obtain the desired min-max lower bound by setting $k = (\gamma n)^{1/(2\alpha+1)}/2$, where $\gamma = 16M^2/(\bar{M}^2)$.

Finally, we check the assumed conditions on k :

$$\begin{aligned} 2k &< p_0 \\ 2\tau k &< \bar{M} \\ 2\tau k/(\bar{M} - 2\tau k) &< 1/2 \\ \frac{\sqrt{2}n^{1/2}k^{1/2}\tau}{\bar{M} - 2\tau k} &\leq 1/2. \end{aligned}$$

Note $\tau \leq M(2k)^{-\alpha-1}$ and $k = (\gamma n)^{1/(2\alpha+1)}/2$. The first condition is satisfied when $(\gamma n)^{1/(2\alpha+1)} < p_0$. For the other conditions, we have

$$\begin{aligned} 2\tau k &\leq M(\gamma n)^{-\alpha/(2\alpha+1)} \\ 2\tau k/(\bar{M} - 2\tau k) &\leq \frac{M(\gamma n)^{-\alpha/(2\alpha+1)}}{\bar{M} - M(\gamma n)^{-\alpha/(2\alpha+1)}} \\ \frac{\sqrt{2}n^{1/2}k^{1/2}\tau}{\bar{M} - 2\tau k} &\leq \frac{M/\gamma^{1/2}}{\bar{M} - M(\gamma n)^{-\alpha/(2\alpha+1)}}. \end{aligned}$$

The first and second upper bounds of these inequalities can be arbitrary small numbers for all sufficiently large n . The last upper bound is smaller than $1/2$ for all sufficiently large n since $\gamma = 16M^2/(\bar{M}^2)$. Thus, the assumed conditions on k are satisfied for all sufficiently large n . \square

6. Discussion

We have considered the estimation of the conditional mean operator under the bandable covariance assumption, which is useful for the multivariate linear regression when there is a natural order in the variables. We showed that the plug-in estimator by the tapering estimator of covariance, which is the minimax optimal estimator for the class of bandable covariance, is sub-optimal for the conditional mean operator. This observation implies that when a function of the covariance matrix is to be estimated, the plug-in estimator by a minimax optimal covariance estimator may not be optimal. We have proposed the blockwise tapering estimator and the blockwise tapering post-processed posterior as minimax-optimal estimators for the conditional mean operator under the bandable covariance assumption. We constructed the estimators by modifying the tapering estimator and the tapering post-processed posterior to exclude the covariates which have small partial correlations with the response variables. Using the numerical studies, we also showed that the blockwise tapering estimator and the blockwise tapering post-processed posterior have smaller errors when p is large enough.

The tapering estimator and blockwise tapering estimator are constructed from the tapered sample covariance, and we can make similar estimators for the regression coefficient by replacing the tapered sample covariance with the banded sample covariance. Furthermore, the banding versions of these estimators also have the same convergence rates as their tapering versions. When we obtain the convergence rates of the tapering estimator and blockwise tapering estimator (Theorems 2.2 and 2.4), we rely on the concentration inequality for the tapered sample covariance $T_k(S_n)$,

$$E(T_k(S_n - \Sigma_0) > t) \leq C_1 p(1 + C_2)^k \exp(-\lambda n t^2),$$

for some positive constants C_1, C_2 and λ . According to inequality (5.5) in [34], the banded sample covariance $B_k(S_n)$ also has the same form of concentration inequality. Thus, the same convergence rates are obtained for the banding versions of the proposed frequentist methods.

Appendix A: Proofs of remaining lemmas and theorems

In this section, we give the proofs of Theorems 3.1 and 3.2 and lemmas given in Section 5.

A.1. Proofs of Theorems 3.1 and 3.2

In this section, we prove Theorems 3.1 and 3.2 which give the P-risk convergence rates of the tapering and blockwise tapering post-processed posteriors, respectively. First, we present Lemmas A.1-A.4 necessary for the proofs of Theorems 3.1 and 3.2.

Lemma A.1. *Let k and p be positive integers with $k \leq p$. Suppose $\Sigma_0 \in \mathcal{F}_{p,\alpha}(M, M_0, M_1)$, and let the prior π^i of Σ be $IW_p(A_n, \nu_n)$ for $A_n \in \mathcal{C}_p$ and $\nu_n > 2p$. If $k \vee \|A_n\| \vee (\nu_n - 2p) \vee \log p = o(n)$ and $\epsilon_n = O(1)$, then for all sufficiently large n , $E_{\Sigma_0}\{E^{\pi^i}(\|T_k^{(\epsilon_n)}(\Sigma) - \Sigma_0\|^4 | \mathbb{Z}_n)\}$ is bounded above by some positive constant depending only on M, M_0, M_1 and α .*

Lemma A.2. *Suppose the same setting of Lemma A.1. If $c \leq \lambda_{\min}(\Sigma_0)/2$, $\lfloor k/2 \rfloor > \{4M/\lambda_{\min}(\Sigma_0)\}^{1/\alpha}$ and $k \vee \|A_n\| \vee (\nu_n - 2p) \vee \log p = o(n)$, then there exist some positive constants C and λ depending only on M_0 and M_1 such that*

$$E_{\Sigma_0}(P^{\pi^i}[\lambda_{\min}\{T_k(\Sigma)\} \leq c | \mathbb{Z}_n]) \leq Cp5^k \exp(-\lambda n),$$

for all sufficiently large n .

Lemma A.3. *Suppose the same setting of Lemma A.1. If $\|A_n\| \vee (\nu_n - 2p) \vee k \vee \log p = o(n)$ and $\epsilon_n = O(1)$, then there exist some positive constant C depending only on M, M_0, M_1 and α such that*

$$E_{\Sigma_0}\{E^{\pi^i}(\|\Sigma_0 - T_k^{(\epsilon_n)}(\Sigma)\|^2 | \mathbb{Z}_n)\} \leq C\left(k^{-2\alpha} + \frac{k + \log p}{n} + \epsilon_n^2\right),$$

for sufficiently large n .

Lemma A.4. *Suppose the same setting of Lemma A.1. Let q and k be positive constants with $k < q \leq p$, and let $\{T_k^{(\epsilon_n)}(\Sigma)\}_{1:q,1:q}$ and $(\Sigma_0)_{1:q,1:q}$ be denoted by $T_k^{(\epsilon_n)}(\Sigma)_{11}$ and $\Sigma_{0,11}$, respectively. If $\lfloor k/2 \rfloor > \{4M/\lambda_{\min}(\Sigma_0)\}^{1/\alpha}$ and $k \vee \|A_n\| \vee (\nu_n - 2p) \vee \log p = o(n)$, then there exist positive constants C and λ depending only on M, M_0, M_1 and α such that*

$$\begin{aligned} & E_{\Sigma_0}\{E^{\pi^i}(\|\Sigma_{0,11}^{-1} - T_k^{(\epsilon_n)}(\Sigma)_{11}^{-1}\|^2 | \mathbb{Z}_n)\} \\ & \leq C\left\{\frac{k + \log p}{n} + k^{-2\alpha} + \epsilon_n^2 + \frac{p^{1/2}5^{k/2} \exp(-\lambda n)}{\epsilon_n^2}\right\}, \end{aligned}$$

for all sufficiently large n .

The proofs of Lemmas A.1-A.4 are given in Section A.6.

Now we give the proofs of Theorems 3.1 and 3.2.

Proof of Theorem 3.1. Since $\lambda_{\min}\{T_k^{(\epsilon_n)}(\Sigma)_{XX}^{-1}\} \geq \lambda_{\min}\{T_k^{(\epsilon_n)}(\Sigma)^{-1}\} \geq \epsilon_n$, we have

$$\begin{aligned} & E_{\Sigma_0}\{E^{\pi^i}(\|\psi(\Sigma_0) - \psi\{T_k^{(\epsilon_n)}(\Sigma)\}\|^2 \mid \mathbb{Z}_n)\} \\ &= E_{\Sigma_0}\{E^{\pi^i}(\|\Sigma_{0,YX}\Sigma_{0,XX}^{-1} - T_k^{(\epsilon_n)}(\Sigma)_{YX}T_k^{(\epsilon_n)}(\Sigma)_{XX}^{-1}\|^2 \mid \mathbb{Z}_n)\} \\ &\leq 2E_{\Sigma_0}\{E^{\pi^i}(\|T_k^{(\epsilon_n)}(\Sigma)_{YX} - \Sigma_{0,YX}\|^2\|T_k^{(\epsilon_n)}(\Sigma)_{XX}^{-1}\|^2 \mid \mathbb{Z}_n)\} \\ &\quad + 2\|\Sigma_{0,YX}\|^2 E_{\Sigma_0}\{E^{\pi^i}(\|T_k^{(\epsilon_n)}(\Sigma)_{XX}^{-1} - \Sigma_{0,XX}^{-1}\|^2 \mid \mathbb{Z}_n)\} \\ &\leq \frac{2^3}{M_1^2} E_{\Sigma_0}\{E^{\pi^i}(\|T_k^{(\epsilon_n)}(\Sigma)_{YX} - \Sigma_{0,YX}\|^2 \mid \mathbb{Z}_n)\} \tag{27} \\ &\quad + \frac{2}{\epsilon_n^2} E_{\Sigma_0}\{E^{\pi^i}(\|T_k^{(\epsilon_n)}(\Sigma)_{YX} - \Sigma_{0,YX}\|^2 \end{aligned}$$

$$\times I[\lambda_{\min}\{T_k(S_n)\} \leq M_1/2 \mid \mathbb{Z}_n]\} \tag{28}$$

$$+ 2\|\Sigma_{0,YX}\|^2 E_{\Sigma_0}\{E^{\pi^i}(\|T_k^{(\epsilon_n)}(\Sigma)_{XX}^{-1} - \Sigma_{0,XX}^{-1}\|^2 \mid \mathbb{Z}_n)\}. \tag{29}$$

Lemmas A.3 and A.4 give that there exist some positive constants C_1 and λ_1 depending only on M, M_0, M_1 and α such that

$$(27) + (29) \leq C_1 \left\{ \frac{k + \log p}{n} + k^{-2\alpha} + \epsilon_n^2 + \frac{p^{1/2}5^{k/2} \exp(-\lambda_1 n)}{\epsilon_n^2} \right\},$$

for all sufficiently large n . Next we show the upper bound of (28). By Lemmas A.1 and A.2, there exist positive constants C_2 and λ_2 depending only on M, M_0, M_1 and α such that

$$\begin{aligned} (28) &\leq \frac{2}{\epsilon_n^2} E_{\Sigma_0}\{E^{\pi^i}(\|T_k^{(\epsilon_n)}(\Sigma)_{YX} - \Sigma_{0,YX}\|^4 \mid \mathbb{Z}_n)\}^{1/2} \\ &\quad \times E_{\Sigma_0}\{P^{\pi^i}(I[\lambda_{\min}\{T_k(S_n)\} \leq M_1/2 \mid \mathbb{Z}_n])\}^{1/2} \\ &\leq C_2 \frac{p^{1/2}5^{k/2}}{\epsilon_n^2} \exp(-\lambda_2 n), \end{aligned}$$

for all sufficiently large n . Collecting the upper bounds of (27), (28) and (29), we complete the proof. \square

Proof of Theorem 3.2. We have

$$\begin{aligned} & E_{\Sigma_0}\{E^{\pi^i}(\|\psi(\Sigma_0) - T_k(\Sigma)_{YX}\Lambda^{(\epsilon_n)}\{T_k(\Sigma)_{XX}; 2\lfloor ak \log k \rfloor\}\|^2 \mid \mathbb{Z}_n)\} \\ &\leq 2\|\psi(\Sigma_0) - T_k(\Sigma_0)_{YX}\Lambda^{(0)}\{T_k(\Sigma_{0,XX}); 2\lfloor ak \log k \rfloor\}\|^2 \tag{30} \end{aligned}$$

$$\begin{aligned} & + 4E_{\Sigma_0}\{E^{\pi^i}(\|\{T_k(\Sigma_0)_{YX} - T_k(\Sigma)_{YX}\} \\ &\quad \times \Lambda^{(\epsilon_n)}\{T_k(\Sigma)_{XX}; 2\lfloor ak \log k \rfloor\}\|^2 \mid \mathbb{Z}_n)\} \tag{31} \end{aligned}$$

$$\begin{aligned} & + 4\|T_k(\Sigma_0)_{YX}\|^2 E_{\Sigma_0}\{E^{\pi^i}(\Lambda^{(0)}\{T_k(\Sigma_{0,XX}); 2\lfloor ak \log k \rfloor\} \\ &\quad - \Lambda^{(\epsilon_n)}\{T_k(\Sigma)_{XX}; 2\lfloor ak \log k \rfloor\})^2 \mid \mathbb{Z}_n)\}. \tag{32} \end{aligned}$$

Lemma 2.3 gives that there exist positive constants C_1 and λ_1 depending only on M, M_0, M_1 and α such that

$$(30) \leq C_1 k^{-2\{\alpha \wedge (a\lambda_1 - 1)\}}$$

for all sufficiently large k with $\lfloor ak \log k \rfloor / 2 > k$ and $p_0 > 2\lfloor ak \log k \rfloor$. Let $\tilde{M} = M_{p_0 - 2\lfloor ak \log k \rfloor + 1}^{(2\lfloor ak \log k \rfloor)}$ in this proof. By the definition of $\Lambda^{(\epsilon_n)}$, we have

$$\|\Lambda^{(\epsilon_n)}\{T_k(\Sigma_{XX}); 2\lfloor ak \log k \rfloor\}\| = \|T_k^{(\epsilon_n)}\{\tilde{M}(\Sigma_{XX})\}^{-1}\|.$$

Thus, there exist positive constants C_2 and λ_2 depending only on M, M_0, M_1 and α such that

$$\begin{aligned} (31) &\leq 4E_{\Sigma_0}\{E^{\pi^i}(\|T_k(\Sigma_0)_{YX} - T_k(\Sigma)_{YX}\|^2 \|T_k^{(\epsilon_n)}\{\tilde{M}(\Sigma_{XX})\}^{-1}\|^2 \mid \mathbb{Z}_n)\} \\ &\leq \frac{4}{\epsilon_n^2} E_{\Sigma_0}\{E^{\pi^i}(\|T_k(\Sigma_0)_{YX} - T_k(\Sigma)_{YX}\|^2 I[\lambda_{\min}\{\tilde{M}(\Sigma_{XX})\} < M_1/2 \mid \mathbb{Z}_n]) \\ &\quad + \frac{16}{M_1^2} E_{\Sigma_0}\{E^{\pi^i}(\|T_k(\Sigma_0)_{YX} - T_k(\Sigma)_{YX}\|^2 \mid \mathbb{Z}_n)\} \\ &\leq \frac{4}{\epsilon_n^2} E_{\Sigma_0}\{E^{\pi^i}(\|T_k\{M_{p_0-k}^{(2k+1)}(\Sigma_0)\} - T_k\{M_{p_0-k}^{(2k+1)}(\Sigma)\}\|^4 \mid \mathbb{Z}_n)\}^{1/2} \\ &\quad \times E_{\Sigma_0}[P^{\pi^i}\{\lambda_{\min}\{\tilde{M}(\Sigma_{XX})\} < M_1/2 \mid \mathbb{Z}_n\}]^{1/2} \\ &\quad + \frac{16}{M_1^2} E_{\Sigma_0}\{E^{\pi^i}(\|T_k\{M_{p_0-k}^{(2k+1)}(\Sigma_0)\} - T_k\{M_{p_0-k}^{(2k+1)}(\Sigma)\}\|^2 \mid \mathbb{Z}_n)\} \\ &\leq C_2 \left(\frac{1}{\epsilon_n^2} (ak \log k)^{1/2} 5^{k/2} \exp(-\lambda_2 n) + \frac{k + \log k}{n} + k^{-2\alpha} \right), \end{aligned}$$

for all sufficiently large n . The third inequality holds since $\lambda_{\min}\{\tilde{M}(\Sigma_{XX})\} \geq \lambda_{\min}(\Sigma)$. The last inequality holds by Lemma A.1, A.2 and A.3. For the upper bound of (32), we have

$$\begin{aligned} &E_{\Sigma_0}\{E^{\pi^i}(\Lambda^{(0)}\{T_k(\Sigma_{0,XX}); 2\lfloor ak \log k \rfloor\} \\ &\quad - \Lambda^{(\epsilon_n)}\{T_k(\Sigma)_{XX}; 2\lfloor ak \log k \rfloor\})^2 \mid \mathbb{Z}_n)\} \\ &= E_{\Sigma_0}\{E^{\pi^i}(\|T_k\{\tilde{M}(\Sigma_{0,XX})\}^{-1} - T_k^{(\epsilon_n)}\{\tilde{M}(\Sigma_{XX})\}^{-1}\|^2 \mid \mathbb{Z}_n)\} \\ &\leq 2E_{\Sigma_0}\{E^{\pi^i}(\|\tilde{M}(\Sigma_{0,XX})^{-1} - T_k^{(\epsilon_n)}\{\tilde{M}(\Sigma_{XX})\}^{-1}\|^2 \mid \mathbb{Z}_n)\} \tag{33} \\ &\quad + 2\|\tilde{M}(\Sigma_{0,XX})^{-1} - T_k\{\tilde{M}(\Sigma_{0,XX})\}^{-1}\|^2 \tag{34} \end{aligned}$$

where the first equality holds by the definition of $\Lambda^{(0)}$ and $\Lambda^{(\epsilon_n)}$. Since $\tilde{M}(\Sigma_{0,XX}) \in \mathcal{F}_{2\lfloor ak \log k \rfloor, \alpha}(M, M_0, M_1)$, Lemma 5.2 gives that there exists some positive constant C_3 depending only on M, M_0 and M_1 such that

$$(34) \leq C_3 (\lfloor k/2 \rfloor)^{-2\alpha}$$

for all sufficiently large k . Lemma A.4 gives that there exist some positive constants C_4 and λ_3 depending only on M, M_0, M_1 and α such that

$$(33) \leq C_4 \left\{ \frac{k + \log(2ak \log k)}{n} + k^{-2\alpha} + \epsilon_n^2 + \frac{(2ak \log k)^{1/2} 5^{k/2} \exp(-\lambda_3 n)}{\epsilon_n^2} \right\},$$

for all sufficiently large n . Thus, there exists some positive constant C_5 depending only on M, M_0, M_1 and α such that

$$(32) \leq C_5 \left\{ \frac{k + \log(2ak \log k)}{n} + k^{-2\alpha} + \epsilon_n^2 + \frac{(2ak \log k)^{1/2} 5^{k/2} \exp(-\lambda_3 n)}{\epsilon_n^2} \right\},$$

for all sufficiently large n and k . Collecting the upper bounds of (30), (31) and (32), we complete the proof. \square

A.2. Proofs of lemmas in Section 5.1

In this section, we give the proofs of lemmas in Section 5.1.

Proof of Lemma 5.1. Since Σ is a k -band matrix,

$$|w_{ij}| \leq Cq^{2|i-j|/k}, \quad i, j \in \{1, 2, \dots, p\},$$

where $q = (\kappa^{1/2} - 1)/(\kappa^{1/2} + 1)$, $C = \{ \|\Sigma^{-1}\| \vee (1 + \kappa^{1/2})^2 \} / (2\|\Sigma\|)$ and κ is the spectral condition number of Σ defined as $\lambda_{\max}(\Sigma)/\lambda_{\min}(\Sigma)$ [11, Theorem 2.4]. Note that

$$\begin{aligned} \max_j \sum_{\{i: |i-j| > ak \log k\}} q^{2|i-j|/k} &\leq \frac{2}{1 - q^{2/k}} q^{2a \log k} \\ &= \frac{2}{k(1 - q^{2/k})} k^{\log q^{2a+1}}, \end{aligned}$$

and

$$\lim_{k \rightarrow \infty} k(q^{2/k} - 1) = \left(\frac{dq^{2x}}{dx} \right)_{x=0} = 2 \log q.$$

Thus, we get

$$\max_j \sum_{\{i: |i-j| > ak \log k\}} q^{2|i-j|/k} \leq \frac{2}{-\log q} k^{\log q^{2a+1}}$$

for all sufficiently large k . \square

Proof of Lemma 5.2. By Theorem 2.3.4 of [16], we have

$$\|\Sigma_0^{-1} - B_k(\Sigma_0)^{-1}\|_1 \leq \frac{\|B_k(\Sigma_0)^{-1}\|_1^2 \|\Sigma_0 - B_k(\Sigma_0)\|_1}{1 - \|B_k(\Sigma_0)^{-1}\|_1 \|\Sigma_0 - B_k(\Sigma_0)\|_1}, \quad (35)$$

when $\|B_k(\Sigma_0)^{-1}\{\Sigma_0 - B_k(\Sigma_0)\}\|_1 < 1$. This inequality condition holds for all sufficiently large k , since $\|B_k(\Sigma_0)^{-1}\|_1$ is bounded above by a constant and $\|\Sigma_0 - B_k(\Sigma_0)\|_1 \leq Mk^{-\alpha}$. We show the boundedness of $\|B_k(\Sigma_0)^{-1}\|_1$ below. Note that $B_k(\Sigma_0) \in \mathcal{C}_p$ for all sufficiently large k since $\lambda_{\min}\{B_k(\Sigma_0)\} \geq \lambda_{\min}(\Sigma_0) - \|B_k(\Sigma_0) - \Sigma_0\|$ [23, Lemma 4.14]. Thus, Theorem 2.4 of [11] gives

$$\begin{aligned} \|B_k(\Sigma_0)^{-1}\|_1 &\leq \max_j \sum_{i=1}^p |w_{ij}| \\ &\leq C_1 \sup_j \sum_{i=1}^p (q_1^{2/k})^{|i-j|} \\ &\leq \frac{2C_1}{1 - q_1^{2/k}}, \end{aligned} \tag{36}$$

where $B_k(\Sigma_0) = (w_{ij})$, $C_1 = \{\|B_k(\Sigma_0)^{-1}\| \vee (1 + \kappa_1^{1/2})^2\} / (2\|B_k(\Sigma_0)\|)$, $q_1 = (\kappa_1^{1/2} - 1) / (\kappa_1^{1/2} + 1)$ and κ_1 is the spectral condition number of $B_k(\Sigma_0)$. To show that (36) is bounded, it suffices to show that $\|B_k(\Sigma_0)^{-1}\|_2$ is bounded. By Lemma 4.14 in [23], we have

$$\begin{aligned} \|B_k(\Sigma_0)^{-1}\|_2 &= \lambda_{\min}\{B_k(\Sigma_0)\}^{-1} \\ &\leq \frac{1}{\lambda_{\min}(\Sigma_0) - \|B_k(\Sigma_0) - \Sigma_0\|} \\ &\leq \frac{1}{\lambda_{\min}(\Sigma_0) - Mk^{-\alpha}}, \end{aligned} \tag{37}$$

which is bounded for all sufficiently large k . Thus, for all sufficiently large k , $\|B_k(\Sigma_0)^{-1}\{\Sigma_0 - B_k(\Sigma_0)\}\|_1 < 1$ and the inequality (35) is satisfied. By combining (35) and (36), we have that there exists some positive constant C_2 depending only on M , M_0 and M_1 such that

$$\begin{aligned} \|\Sigma_0^{-1} - B_k(\Sigma_0)^{-1}\|_1 &\leq \left(\frac{2C_1}{1 - q_1^{2/k}}\right)^2 \frac{Mk^{-\alpha}}{1 - 2C_1Mk^{-\alpha}/(1 - q_1^{2/k})} \\ &\leq C_2Mk^{-\alpha}, \end{aligned}$$

for all sufficiently large k .

Next, we consider the upper bound of $\|\Sigma_0^{-1} - T_k(\Sigma_0)^{-1}\|_1$. Note that $T_k(\Sigma_0)$ is a k -band matrix and

$$\begin{aligned} \|T_k(\Sigma_0) - \Sigma_0\|_1 &\leq \|B_{\lfloor k/2 \rfloor}(\Sigma_0) - \Sigma_0\|_1 \\ &\leq M(\lfloor k/2 \rfloor)^{-\alpha}. \end{aligned} \tag{38}$$

In the same way as deriving the inequalities (36) and (37), we have

$$\begin{aligned} \|T_k(\Sigma_0)^{-1}\|_1 &\leq \frac{2C_3}{1 - q_2^{2/k}} \\ \|T_k(\Sigma_0)^{-1}\|_2 &\leq \frac{1}{\lambda_{\min}(\Sigma_0) - M(\lfloor k/2 \rfloor)^{-\alpha}}, \end{aligned} \tag{39}$$

where $C_3 = \|T_k(\Sigma_0)^{-1}\| \vee (1 + \kappa_2^{1/2})^2 / (2\|T_k(\Sigma_0)\|)$, $q_2 = (\kappa_2^{1/2} - 1) / (\kappa_2^{1/2} + 1)$ and κ_2 is the spectral condition number of $T_k(\Sigma_0)$. Thus, we have $\|\Sigma_0 - T_k(\Sigma_0)\|_1 < 1$ for all sufficiently large k , and apply Theorem 2.3.4 of [16]. By Theorem 2.3.4 of [16], there exists some positive constant C_4 depending only on M , M_0 and M_1 such that

$$\begin{aligned} \|\Sigma_0^{-1} - T_k(\Sigma_0)^{-1}\|_1 &\leq \frac{\|T_k(\Sigma_0)^{-1}\|_1^2 \|\Sigma_0 - T_k(\Sigma_0)\|_1}{1 - \|T_k(\Sigma_0)^{-1}(\Sigma_0 - T_k(\Sigma_0))\|_1} \\ &\leq C_4 (k/2)^{-\alpha}, \end{aligned}$$

for all sufficiently large k . The last inequality holds by the inequalities (38) and (39). \square

A.3. Proofs of lemmas in Section 5.2

In this section, we give the proofs of the lemmas in Section 5.2. First, we present Lemmas A.5 and A.6 which are necessary for the proofs of these lemmas.

Lemma A.5. *Let p and k be positive integers with $k \leq p$. For an arbitrary covariance matrix $\Sigma \in \mathcal{C}_p$,*

$$\|T_k(\Sigma)\|_r \leq 3 \max_{1 \leq l \leq p} \|M_l^{(k)}(\Sigma)\|_r,$$

where $\|\cdot\|_r$ is the matrix r -norm.

Proof. [6] in Lemma 1 shows

$$T_k(\Sigma) = (k/2)^{-1} \{S^{*(k)}(\Sigma) - S^{*(k/2)}(\Sigma)\},$$

where $S^{*(k)}(\Sigma) = \sum_{l=1-k}^p M_l^{*(k)}(\Sigma)$. Thus, we have

$$\|T_k(\Sigma)\|_r \leq (2/k) (\|S^{*(k)}(\Sigma)\|_r + \|S^{*(k/2)}(\Sigma)\|_r). \tag{40}$$

Note that

$$S^{*(k)}(\Sigma) = \sum_{l=1}^k \sum_{-1 \leq j \leq \lfloor p/k \rfloor} M_{jk+l}^{*(k)}(\Sigma).$$

We have

$$\left\| \sum_{-1 \leq j \leq \lfloor p/k \rfloor} M_{jk+l}^{*(k)}(\Sigma) \right\|_r = \max_{-1 \leq j \leq \lfloor p/k \rfloor} \|M_{jk+l}^{*(k)}(\Sigma)\|_r,$$

for $l \in \{1, 2, \dots, k\}$, since the sum of matrices is a block diagonal matrix. Thus, we get

$$\begin{aligned} \|S^{*(k)}(\Sigma)\|_r &\leq \sum_{l=1}^k \left\| \sum_{-1 \leq j \leq \lfloor p/k \rfloor} M_{jk+l}^{*(k)}(\Sigma) \right\|_r \\ &\leq k \max_{1 \leq l \leq k} \left\| \sum_{-1 \leq j \leq \lfloor p/k \rfloor} M_{jk+l}^{*(k)}(\Sigma) \right\|_r \end{aligned}$$

$$\begin{aligned} &\leq k \max_{1-k \leq l \leq p} \|M_l^{*(k)}(\Sigma)\|_r, \\ &= k \max_{1 \leq l \leq p} \|M_l^{*(k)}(\Sigma)\|_r, \end{aligned} \tag{41}$$

and

$$\|S^{*(k/2)}(\Sigma)\|_r \leq (k/2) \max_{1 \leq l \leq p} \|M_l^{*(k/2)}(\Sigma)\|_r.$$

Since $\|M_l^{*(k/2)}(\Sigma)\|_r \leq \|M_l^{*(k)}(\Sigma)\|_r$ for $l \in \{1, 2, \dots, p\}$, we have

$$\|S^{*(k/2)}(\Sigma)\|_r \leq (k/2) \max_{1-k \leq l \leq p} \|M_l^{*(k)}(\Sigma)\|_r. \tag{42}$$

Collecting (40), (41) and (42), we get

$$\|T_k(\Sigma)\|_r \leq 3 \max_{1 \leq l \leq p} \|M_l^{(k)}(\Sigma)\|_r. \quad \square$$

Lemma A.6. *Let k and r be positive integers, and τ and M_0 be positive real numbers with $\tau \geq k-1$. Suppose $\Sigma_0 \in \mathcal{C}_k$ and $T \sim W_k(\Sigma_0, \tau)$. If $\lambda_{\max}(\Sigma_0) \leq M_0$, then there exists some positive constant C depending only on M_0 and r such that*

$$E(\|T/\tau - \Sigma_0\|_2^r) \leq C \frac{5^k}{\tau^{r/2}}.$$

Proof. Let $\Omega_\tau^{(k)} = \Sigma_0^{-1/2} T \Sigma_0^{-1/2} / \tau$. We have

$$E(\|T/\tau - \Sigma_0\|_2^r) \leq \|\Sigma_0\|_2^r E(\|\Omega_\tau^{(k)} - I_k\|_2^r).$$

Since $\|\Sigma_0\|_2^r$ is bounded by M_0^r , it suffices to show the upper bound of $E(\|\Omega_\tau^{(k)} - I_k\|_2^r)$. By a property of Wishart distribution, we have

$$\Omega_\tau^{(k)} = \Sigma_0^{-1/2} T \Sigma_0^{-1/2} / \tau \sim W_k(I_k / \tau, \tau),$$

and

$$\begin{aligned} &E(\|\Omega_\tau^{(k)} - I_k\|^r) \\ &\leq \int_0^1 x^{r-1} P(\|\Omega_\tau^{(k)} - I_k\| > x) dx + \int_1^\infty x^{r-1} P(\|\Omega_\tau^{(k)} - I_k\| > x) dx \\ &\leq 2 \times 5^k \left\{ \int_0^1 x^{r-1} \exp(-\tau x^2 / 2^\tau) dx + \int_1^\infty x^{r-1} \exp(-\tau x / 2^\tau) dx \right\}, \end{aligned}$$

where the last inequality is satisfied by Lemma 4.4 in [23]. Integration by substitution and the definition of Gamma function give the inequalities

$$\begin{aligned} \int_0^1 x^{r-1} \exp(-\tau x^2 / 2^\tau) dx &\leq \frac{2^{7r/2-1}}{\tau^{r/2}} \int_0^\infty t^{r/2-1} \exp(-t) dt \\ &= \frac{2^{7r/2-1}}{\tau^{r/2}} \Gamma(r/2) \end{aligned}$$

$$\begin{aligned} \int_1^\infty x^{r-1} \exp(-\tau x/2^\tau) dx &\leq \left(\frac{2^\tau}{\tau}\right)^r \int_0^\infty t^{r-1} \exp(-t) dx \\ &= \left(\frac{2^\tau}{\tau}\right)^r \Gamma(r). \end{aligned}$$

Thus,

$$E(\|\Omega_\tau^{(k)} - I_k\|^r) \leq C_1 \frac{5^k}{\tau^{r/2}},$$

for some positive constant C_1 depending only on r . □

Proof of Lemma 5.3. By Lemma 4.14 in [23], we have

$$\begin{aligned} \lambda_{\min}(\Sigma) &= \lambda_{\min}(\Sigma - \Sigma_0 + \Sigma_0) \\ &\geq \lambda_{\min}(\Sigma_0) - \|\Sigma - \Sigma_0\|. \end{aligned}$$

Using this inequality, we get

$$\begin{aligned} \|\Sigma - \Sigma_0\| &\geq \|\Sigma - \Sigma_0\| I(\lambda_{\min}(\Sigma) < 0) \\ &\geq \{ \|\Sigma - \Sigma_0\| - \lambda_{\min}(\Sigma_0) \} I(\lambda_{\min}(\Sigma) < 0) \\ &\geq -\lambda_{\min}(\Sigma) I(\lambda_{\min}(\Sigma) < 0). \end{aligned} \tag{43}$$

We also have

$$\begin{aligned} \{\epsilon_n - \lambda_{\min}(\Sigma)\} \vee 0 &\leq \{\epsilon_n - \lambda_{\min}(\Sigma) I(\lambda_{\min}(\Sigma) < 0)\} \vee 0 \\ &= \epsilon_n - \lambda_{\min}(\Sigma) I(\lambda_{\min}(\Sigma) < 0). \end{aligned} \tag{44}$$

Thus, we have

$$\begin{aligned} &\|\Sigma^{(\epsilon_n)} - \Sigma_0\|^r \\ &= \|\Sigma - \Sigma_0 + [\{\epsilon_n - \lambda_{\min}(\Sigma)\} \vee 0] I_p\|^r \\ &\leq 2^{r-1} \|\Sigma - \Sigma_0\|^r + 2^{r-1} |\epsilon_n - \lambda_{\min}(\Sigma) I(\lambda_{\min}(\Sigma) < 0)|^r \\ &\leq 2^{r-1} \|\Sigma - \Sigma_0\|^r + 4^{r-1} |\epsilon_n|^r + 4^{r-1} |-\lambda_{\min}(\Sigma) I(\lambda_{\min}(\Sigma) < 0)|^r \\ &\leq 2^{r-1} \|\Sigma - \Sigma_0\|^r + 4^{r-1} |\epsilon_n|^r + 4^{r-1} \|\Sigma - \Sigma_0\|^r \\ &\leq 2^{2r-1} \|\Sigma - \Sigma_0\|^r + 4^{r-1} |\epsilon_n|^r, \end{aligned}$$

where the first and third inequalities are satisfied inequalities (44) and (43), respectively. □

Proof of Lemma 5.4. Two inequalities in Lemma 4.14 of [23],

$$\begin{aligned} \lambda_{\min}\{T_k(S_n)\} &\geq \lambda_{\min}\{T_k(\Sigma_0)\} - \|T_k(S_n) - T_k(\Sigma_0)\| \\ \lambda_{\min}\{T_k(\Sigma_0)\} &\geq \lambda_{\min}(\Sigma_0) - \|\Sigma_0 - T_k(\Sigma_0)\|, \end{aligned}$$

imply

$$I[\lambda_{\min}\{T_k(S_n)\} \leq c]$$

$$\begin{aligned} &\leq I\{\lambda_{\min}(\Sigma_0) - \|T_k(\Sigma_0) - \Sigma_0\| - \|T_k(S_n) - T_k(\Sigma_0)\| \leq c\} \\ &\leq I(\|T_k(S_n) - T_k(\Sigma_0)\| \geq c_k), \end{aligned} \quad (45)$$

where $c_k = \lambda_{\min}(\Sigma_0) - M(\lfloor k/2 \rfloor)^{-\alpha} - c$. The last inequality holds by the inequality (38). Since $c \leq \lambda_{\min}(\Sigma_0)/2$ and $\lfloor k/2 \rfloor > \{4M/\lambda_{\min}(\Sigma_0)\}^{1/\alpha}$, we have $c_k \geq \lambda_{\min}(\Sigma_0)/4$. By Lemma A.5, we have

$$\begin{aligned} P(\|T_k(S_n) - T_k(\Sigma_0)\| \geq c_k) &\leq P(\max_{1 \leq l \leq p} \|M_l^{(k)}(S_n - \Sigma_0)\| \geq c_k/3) \\ &\leq P(\|\Sigma_0\| \max_{1 \leq l \leq p} \|\Omega_{n,l}^{*(k)} - I_k\| \geq c_k/3) \\ &\leq P(M_0 \max_{1 \leq l \leq p} \|\Omega_{n,l}^{*(k)} - I_k\| \geq c_k/3) \\ &\leq p \max_{1 \leq l \leq p} P\{\|\Omega_{n,l}^{*(k)} - I_k\| \geq c_k/(3M_0)\}, \end{aligned}$$

where $\Omega_{n,l}^{*(k)} = M_l^{(k)}(\Sigma_0)^{-1/2} M_l^{(k)}(S_n) M_l^{(k)}(\Sigma_0)^{-1/2}$ of which distribution is $W_k(I_k/n, n)$. Since $c_k/(3M_0) < 1$, Lemma 4.4 in [23] gives

$$P\{\|\Omega_{n,l}^{*(k)} - I_k\| \geq c_k/(3M_0)\} \leq 2 \times 5^k \exp[-n\{c_k/(3M_0)\}^2/2^7]$$

Thus, combining this inequality with (45), we get

$$\begin{aligned} E(I[\lambda_{\min}\{T_k(S_n)\} \leq c]) &\leq E\{I(\|T_k(S_n) - T_k(\Sigma_0)\| \geq c_k)\} \\ &\leq p \max_{1 \leq l \leq p} P\{\|\Omega_{n,l}^{*(k)} - I_k\| \geq c_k/(3M_0)\} \\ &\leq 2p5^k \exp(-\lambda n), \end{aligned}$$

for some positive constant λ depending only on M_0 and M_1 . \square

Proof of Lemma 5.5. Note, by Lemmas A.5 and 5.3,

$$\begin{aligned} &E(\|T_k^{\epsilon_n}(S_n) - \Sigma_0\|^4) \\ &\leq 2^7 E(\|T_k(S_n) - \Sigma_0\|^4) + 4^3 \epsilon_n^4 \\ &\leq 2^{10} E(\|T_k(S_n) - T_k(\Sigma_0)\|^4) + 2^{10} \|\Sigma_0 - T_k(\Sigma_0)\|^4 + 4^3 \epsilon_n^4 \\ &\leq 2^{10} 3^4 E(\max_{1 \leq l \leq p} \|M_l^{(k)}(S_n - \Sigma_0)\|^4) + 2^{10} \|\Sigma_0 - T_k(\Sigma_0)\|^4 + 4^3 \epsilon_n^4. \end{aligned}$$

It suffices to show $E(\max_{1 \leq l \leq p} \|M_l^{(k)}(S_n - \Sigma_0)\|^4)$ is bounded above. There exist some positive constants ρ_1 and C_1 depending only on M_0 such that

$$\begin{aligned} &E(\max_{1 \leq l \leq p} \|M_l^{(k)}(S_n - \Sigma_0)\|^4) \\ &\leq x^4 + E\{\max_{1 \leq l \leq p} \|M_l^{(k)}(S_n - \Sigma_0)\|^4 I(\max_{1 \leq l \leq p} \|M_l^{(k)}(S_n - \Sigma_0)\| > x)\} \\ &\leq x^4 + p^{1/2} \max_{1 \leq l \leq p} E(\|M_l^{(k)}(S_n - \Sigma_0)\|^8)^{1/2} P(\max_{1 \leq l \leq p} \|M_l^{(k)}(S_n - \Sigma_0)\| > x)^{1/2} \\ &\leq x^4 + C_1 p^{1/2} \frac{5^{k/2}}{n^2} \{2p5^k \exp(-nx^2 \rho_1)\}^{1/2}, \end{aligned}$$

for all $0 < x < \rho_1$. The last inequality is satisfied by Lemma A.6 and Lemma 3 of [6]. We set $x = \{2(\log p + k) \log 5 / (n\rho_1)\}^{1/2}$, which can be smaller than an arbitrary positive constant for all sufficiently large n . Then, we have

$$\begin{aligned} & E(\max_{1 \leq l \leq p} \|M_l^{(k)}(S_n - \Sigma_0)\|^4) \\ & \leq \left\{ \frac{2(\log p + k) \log 5}{n\rho_1} \right\}^2 + \frac{\sqrt{2}C_1 p 5^k}{n^2} \exp\{-(\log p + k) \log 5\} \\ & = \left\{ \frac{2(\log p + k) \log 5}{n\rho_1} \right\}^2 + \frac{\sqrt{2}C_1}{n^2} \exp(-(\log 5 - 1) \log p), \end{aligned}$$

which is bounded by some positive constant depending only on M_0 for all sufficiently large n . □

Proof of Lemma 5.6. Let $c = \lambda_{\min}(\Sigma_0)/2$.

Since $\lambda_{\min}\{T_k^{(\epsilon_n)}(S_n)_{11}\} \geq \lambda_{\min}\{T_k^{(\epsilon_n)}(S_n)\} \geq \epsilon_n$, we have

$$\begin{aligned} & E(\|T_k^{(\epsilon_n)}(S_n)_{11}^{-1} - \Sigma_{0,11}^{-1}\|^2) \\ & \leq \|\Sigma_{0,11}^{-1}\|^2 E(\|T_k^{(\epsilon_n)}(S_n)_{11}^{-1}\|^2 \|T_k^{(\epsilon_n)}(S_n)_{11} - \Sigma_{0,11}\|^2) \\ & \leq \frac{\|\Sigma_{0,11}^{-1}\|^2}{\epsilon_n^2} E(\|T_k^{(\epsilon_n)}(S_n)_{11} - \Sigma_{0,11}\|^2 I[\lambda_{\min}\{T_k(S_n)\} \leq c]) \end{aligned} \tag{46}$$

$$+ \frac{\|\Sigma_{0,11}^{-1}\|^2}{c^2} E(\|T_k^{(\epsilon_n)}(S_n)_{11} - \Sigma_{0,11}\|^2). \tag{47}$$

By Lemmas 5.4 and 5.5, there exist some positive constants C_1 and λ_1 depending only on M_0 and M_1 such that

$$\begin{aligned} (46) & \leq \frac{1}{M_1^2 \epsilon_n^2} E(\|T_k^{(\epsilon_n)}(S_n)_{11} - \Sigma_{0,11}\|^4)^{1/2} P[\lambda_{\min}\{T_k(S_n)\} \leq c]^{1/2} \\ & \leq \frac{C_1}{M_1^2 \epsilon_n^2} p^{1/2} 5^{k/2} \exp(-\lambda_1 n), \end{aligned} \tag{48}$$

for all sufficiently large n . We also have

$$\begin{aligned} (47) & \leq \frac{1}{M_1^2 c^2} E(\|T_k^{(\epsilon_n)}(S_n) - \Sigma_0\|^2) \\ & \leq \frac{2^3}{M_1^2 c^2} E(\|T_k(S_n) - \Sigma_0\|^2) + \frac{4}{M_1^2 c^2} \epsilon_n^2 \\ & \leq C_2 \left\{ \frac{k + \log p}{n} + k^{-2\alpha} + \epsilon_n^2 \right\}, \end{aligned} \tag{49}$$

for some positive constant C_2 depending only on M, M_0, M_1 and α . The second and last inequalities hold by Lemma 5.3 and Theorem 2 of [6], respectively. Collecting (48) and (49), we complete the proof. □

A.4. Proof of lemma in Section 5.3

In this section, we give the proof of Lemma 5.7 in Section 5.3.

Proof of Lemma 5.7. Let $\{B_{\lfloor ak \log k \rfloor}(\Omega)\}_{12} = \{B_{\lfloor ak \log k \rfloor}(\Omega)\}_{1:q^*,(q^*+1):q}$, where $q^* = q - \lfloor ak \log k \rfloor - \lfloor bk \log k \rfloor$ and $B_{\lfloor ak \log k \rfloor}$ is $\lfloor ak \log k \rfloor$ -banding operator defined in [2]. First, we show the equality

$$M_{\lfloor ak \log k \rfloor+1}^{(\lfloor bk \log k \rfloor)}(A_{22}^{-1}[I - A_{21}\{B_{\lfloor ak \log k \rfloor}(\Omega)\}_{12}]) = M_{\lfloor ak \log k \rfloor+1}^{(\lfloor bk \log k \rfloor)}(A_{22}^{-1}). \tag{50}$$

Since $\lfloor ak \log k \rfloor \geq k$ and A is a k -band matrix, A_{21} is expressed as

$$A_{21} = \begin{bmatrix} O_{\{\lfloor ak \log k \rfloor \times (q^* - \lfloor ak \log k \rfloor)\}} & A^* \\ O_{\{\lfloor bk \log k \rfloor \times (q^* - \lfloor ak \log k \rfloor)\}} & O_{(\lfloor bk \log k \rfloor \times \lfloor ak \log k \rfloor)} \end{bmatrix},$$

for some $\lfloor ak \log k \rfloor \times \lfloor ak \log k \rfloor$ -matrix A^* . Likewise, $\{B_{\lfloor ak \log k \rfloor}(\Omega)\}_{12}$ is expressed as

$$\{B_{\lfloor ak \log k \rfloor}(\Omega)\}_{12} = \begin{bmatrix} O_{\{(q^* - \lfloor ak \log k \rfloor) \times \lfloor ak \log k \rfloor\}} & O_{\{(q^* - \lfloor ak \log k \rfloor) \times \lfloor bk \log k \rfloor\}} \\ B^*_{(\lfloor ak \log k \rfloor \times \lfloor ak \log k \rfloor)} & O_{(\lfloor ak \log k \rfloor \times \lfloor bk \log k \rfloor)} \end{bmatrix},$$

for some $\lfloor ak \log k \rfloor \times \lfloor ak \log k \rfloor$ -matrix B^* . Thus,

$$A_{21}\{B_{\lfloor ak \log k \rfloor}(\Omega)\}_{12} = \begin{bmatrix} A^*B^* & O_{(\lfloor ak \log k \rfloor \times \lfloor bk \log k \rfloor)} \\ O_{(\lfloor bk \log k \rfloor \times \lfloor ak \log k \rfloor)} & O_{(\lfloor bk \log k \rfloor \times \lfloor bk \log k \rfloor)} \end{bmatrix},$$

and $M_{\lfloor ak \log k \rfloor+1}^{(\lfloor bk \log k \rfloor)}(A_{22}^{-1}[A_{21}\{B_{\lfloor ak \log k \rfloor}(\Omega)\}_{12}])$ is the zero matrix, which gives the equality (50).

Next, we show the upper bound of

$$\|M_{\lfloor ak \log k \rfloor+1}^{(\lfloor bk \log k \rfloor)}(A_{22}^{-1}[I - A_{21}\{B_{\lfloor ak \log k \rfloor}(\Omega)\}_{12}]) - M_{\lfloor ak \log k \rfloor+1}^{(\lfloor bk \log k \rfloor)}(\Omega_{22})\|. \tag{51}$$

By the block matrix inversion formula, $\Omega_{22} = A_{22}^{-1}(I - A_{21}\Omega_{12})$, we have

$$\begin{aligned} (51) &= \|M_{\lfloor ak \log k \rfloor+1}^{(\lfloor bk \log k \rfloor)}(A_{22}^{-1}A_{21}[\{B_{\lfloor ak \log k \rfloor}(\Omega)\}_{12} - \Omega_{12}])\| \\ &\leq \|A_{22}^{-1}A_{21}\| \|\{B_{\lfloor ak \log k \rfloor}(\Omega)\}_{12} - \Omega_{12}\| \\ &\leq \lambda_{\min}(A_{22})^{-1} \|A_{21}\| \|B_{\lfloor ak \log k \rfloor}(\Omega) - \Omega\|_1 \\ &\leq \|A\| \|A^{-1}\| \|B_{\lfloor ak \log k \rfloor}(\Omega) - \Omega\|_1. \end{aligned}$$

By Lemma 5.1, there exist some positive constants λ_1 and C_1 depending only on $\|A\|$ and $\|A^{-1}\|$ such that

$$\|B_{\lfloor ak \log k \rfloor}(\Omega) - \Omega\|_1 \leq C_1 k^{-a\lambda_1+1},$$

for all sufficiently large k with $p > k \vee (ak \log k)$. By combining this inequality with the equality (50), we get

$$\|M_{\lfloor ak \log k \rfloor+1}^{(\lfloor bk \log k \rfloor)}(A_{22}^{-1}) - M_{\lfloor ak \log k \rfloor+1}^{(\lfloor bk \log k \rfloor)}(\Omega_{22})\| \leq C_1 \|A\| \|A^{-1}\| k^{-a\lambda_1+1},$$

for all sufficiently large k . □

A.5. Proof of lemma in Section 5.4

In this section, we give the proof of Lemma 5.8 in Section 5.4.

Proof of Lemma 5.8. Let $c = \|\Sigma^{-1}\|_2 \wedge \|\Sigma'^{-1}\|_2$ and $D = \Sigma' - \Sigma$. Without loss of generality we suppose $\|\Sigma^{-1}\|_2 = c$. The Pinsker inequality gives

$$\begin{aligned} \|P_\Sigma - P_{\Sigma'}\|_1^2 &\leq 2KL(P_{\Sigma'} | P_\Sigma) \\ &= n\{tr(\Sigma'\Sigma^{-1}) - \log \det(\Sigma'\Sigma^{-1}) - p\}, \\ &= n\{tr(D\Sigma^{-1}) - \log \det(I_p + D\Sigma^{-1})\}, \end{aligned} \tag{52}$$

where $KL(\cdot | \cdot)$ is the Kullback-Leibler divergence. Before showing the upper bound of (52), we show

$$\log(1 + x) \geq x - x^2, \text{ for } x > -1/2. \tag{53}$$

Let $g(x) = \log(1 + x) - x + x^2$. Note

$$\begin{aligned} g(0) &= 0 \\ g'(x) &< 0, \text{ if } -1/2 < x < 0 \\ g'(x) &> 0, \text{ if } x > 0, \end{aligned}$$

where $g'(x)$ is the derivative of $g(x)$. Thus, inequality (53) holds.

Now we give the upper bound of (52). Note that $D\Sigma^{-1}$ is a similar matrix of $\Sigma^{-1/2}D\Sigma^{-1/2}$ and the set of eigenvalues of $D\Sigma^{-1}$ coincides with that of $\Sigma^{-1/2}D\Sigma^{-1/2}$. Let $\{\lambda_1, \lambda_2, \dots, \lambda_p\}$ be the set of eigenvalues of $D\Sigma^{-1}$. If $\lambda_i > -1/2$ for all $i \in \{1, \dots, p\}$, then inequality (53) gives

$$\begin{aligned} -\log \det(I_p + D\Sigma^{-1}) &= -\sum_{i=1}^p \log(1 + \lambda_i) \\ &\leq -\sum_{i=1}^p \lambda_i + \sum_{i=1}^p \lambda_i^2 \\ &= -tr(D\Sigma^{-1}) + \sum_{i=1}^p \lambda_i^2. \end{aligned}$$

By applying this inequality to (52), we have

$$\|P_\Sigma - P_{\Sigma'}\|_1^2 \leq n \sum_{i=1}^p \lambda_i^2.$$

It suffices to show $\max_{i=1, \dots, p} |\lambda_i| < 1/2$ and $\sum_{i=1}^p \lambda_i^2 \leq c^2 \|D\|_F^2$. Since $\|D\|_2 < 1/(2c)$,

$$\begin{aligned} \max_{i=1, \dots, p} |\lambda_i| &\leq \|\Sigma^{-1/2}D\Sigma^{-1/2}\|_2 \\ &\leq \|D\|_2 \|\Sigma^{-1}\|_2 \\ &< 1/2. \end{aligned}$$

Next, we show the upper bound of $\sum_{i=1}^p \lambda_i^2$. we have $\Sigma^{-1} = UVU^T$ for some orthogonal matrix U and some diagonal matrix V , and

$$\begin{aligned} \sum_{i=1}^p \lambda_i^2 &= \|\Sigma^{-1/2} D \Sigma^{-1/2}\|_F^2 \\ &= \|V U^T D U V\|_F^2 \\ &\leq \|V\|^2 \|U^T D U\|_F^2 \\ &= \|\Sigma^{-1}\|^2 \|D\|_F^2. \end{aligned}$$

The second and third equalities hold since U is an orthogonal matrix. The first inequality holds since V is a diagonal matrix. \square

A.6. Proofs of lemmas in Section A.1

In this section, we give the proofs of the lemmas in Section A.1. First, we present Lemmas A.7 and A.8 which are necessary for the proofs of these lemmas.

Lemma A.7. *Let n, k and r be positive integers and M_0 be a positive real number. Suppose $\Sigma_0 \in \mathcal{C}_k$ and let the prior distribution π on \mathcal{C}_k be $IW(A_n, \tau_n)$ for $A_n \in \mathcal{C}_k$ and $\tau_n > 2k$. If $(\tau_n - 2k) \vee \|A_n\| \vee k = o(n)$ and $\|\Sigma_0\| \leq M_0$, then there exists some positive constant C depending only on M_0 and r such that*

$$E_{\Sigma_0} \{E^\pi(\|\Sigma - \Sigma_0\|^r \mid \mathbb{Z}_n)\} \leq C k^r \frac{5^{3k/2}}{n^{r/2}},$$

for all sufficiently large n .

Proof. Let $\hat{\Sigma} = (nS_n + A_n)/(n + \tau_n - k - 1)$. We have

$$\begin{aligned} E_{\Sigma_0} \{E^\pi(\|\Sigma - \Sigma_0\|^r \mid \mathbb{Z}_n)\} &\leq 2^{r-1} E_{\Sigma_0} \{E^\pi(\|\Sigma - \hat{\Sigma}\|^r \mid \mathbb{Z}_n)\} \tag{54} \\ &\quad + 2^{r-1} E_{\Sigma_0}(\|\hat{\Sigma} - \Sigma_0\|^r). \tag{55} \end{aligned}$$

For the upper bound of (55), we have

$$\begin{aligned} &E_{\Sigma_0}(\|\hat{\Sigma} - \Sigma_0\|^r) \\ &\leq 2^{r-1} E_{\Sigma_0} \left(\left\| \frac{n}{n + \tau_n - k - 1} (S_n - \Sigma_0) \right\|^r \right) + 2^{r-1} \left\| \frac{A_n - (\tau_n - k - 1)\Sigma_0}{n + \tau_n - k - 1} \right\|^r. \end{aligned}$$

Since $(\tau_n - k - 1)/n = O(k/n)$ and $\|A_n\| = o(n)$, there exists some positive constant C_1 depending only on M_0 and r such that

$$\left\| \frac{A_n - (\tau_n - k - 1)\Sigma_0}{n + \tau_n - k - 1} \right\|^r \leq C_1 \left(\frac{k}{n}\right)^r,$$

for all sufficiently large n . Lemma A.6 gives

$$E_{\Sigma_0} \left(\left\| \frac{n}{n + \tau_n - k - 1} (S_n - \Sigma_0) \right\|^r \right) \leq C_2 \frac{5^k}{n^{r/2}},$$

for some positive constant C_2 depending only on M_0 and r . Thus, we have

$$(55) \leq 4^{r-1} C_1 \left(\frac{k}{n}\right)^r + 4^{r-1} C_2 \frac{5^k}{n^{r/2}}. \tag{56}$$

Next, we show the upper bound of (54). We have

$$\begin{aligned}
 & E_{\Sigma_0}\{E^\pi(\|\Sigma - \hat{\Sigma}\|^r \mid \mathbb{Z}_n)\} \\
 &= E_{\Sigma_0}\{E^\pi(\|\hat{\Sigma}^{1/2}(\Omega_{n^*}^{(k)})^{-1}(I_k - \Omega_{n^*}^{(k)})\hat{\Sigma}^{1/2}\|^r \mid \mathbb{Z}_n)\} \\
 &\leq E_{\Sigma_0}\{\|\hat{\Sigma}\|^r E^\pi(\|(\Omega_{n^*}^{(k)})^{-1}\|^r \|I_k - \Omega_{n^*}^{(k)}\|^r \mid \mathbb{Z}_n)\} \\
 &\leq E_{\Sigma_0}\{\|\hat{\Sigma}\|^r E^\pi(\|(\Omega_{n^*}^{(k)})^{-1}\|^{2r} \mid \mathbb{Z}_n)^{1/2} E^\pi(\|I_k - \Omega_{n^*}^{(k)}\|^{2r} \mid \mathbb{Z}_n)^{1/2}\}, \tag{57}
 \end{aligned}$$

where $\Omega_{n^*}^{(k)} = \hat{\Sigma}^{1/2}\Sigma^{-1}\hat{\Sigma}^{1/2}$. Note that $[\Sigma \mid \mathbb{Z}_n] \sim IW_k(A_n + nS_n, n + \tau_n)$. Since $[\Sigma^{-1} \mid \mathbb{Z}_n] \sim W_k\{(A_n + nS_n)^{-1}, n + \tau_n - k - 1\}$ [28], we have $(n + \tau_n - k - 1)\Omega_{n^*}^{(k)} \sim W_k(I_k, n + \tau_n - k - 1)$. Lemma A.6 gives

$$\begin{aligned}
 E^\pi(\|I_k - \Omega_{n^*}^{(k)}\|^{2r} \mid S_n)^{1/2} &\leq C_3 \frac{5^{k/2}}{(n + \tau_n - k - 1)^{r/2}} \\
 &\leq C_3 \frac{5^{k/2}}{n^{r/2}}, \tag{58}
 \end{aligned}$$

for some positive constant C_3 depending only on M_0 and r . Since $[(\Omega_{n^*}^{(k)})^{-1} \mid \mathbb{Z}_n] \sim IW_k\{(n + \tau_n - k - 1)I_k, n + \tau_n\}$, we have

$$\begin{aligned}
 & E^\pi(\|(\Omega_{n^*}^{(k)})^{-1}\|^{2r} \mid \mathbb{Z}_n) \\
 &\leq E^\pi(\left[\sum_{i=1}^k \{(\Omega_{n^*}^{(k)})^{-1}\}_{ii}\right]^{2r} \mid \mathbb{Z}_n) \\
 &\leq k^{2r-1} \sum_{i=1}^k E^\pi[\{(\Omega_{n^*}^{(k)})^{-1}\}_{ii}^{2r} \mid \mathbb{Z}_n] \\
 &\leq k^{2r}(n + \tau_n - k - 1)^{2r} \frac{1}{2^{2r}} \frac{\Gamma\{(n + \tau_n - 2k)/2 - 2r\}}{\Gamma\{(n + \tau_n - 2k)/2\}} \\
 &= k^{2r}(n + \tau_n - k - 1)^{2r} \frac{1}{2^{2r}} \prod_{i=1}^{2r} \{(n + \tau_n - 2k)/2 - i\}^{-1} \\
 &\leq \left(\frac{k}{2}\right)^{2r} \left\{ \frac{n + \tau_n - k - 1}{(n + \tau_n - 2k)/2 - 2r} \right\}^{2r}, \tag{59}
 \end{aligned}$$

where $\{(\Omega_{n^*}^{(k)})^{-1}\}_{ii}$ is the i th diagonal element of $(\Omega_{n^*}^{(k)})^{-1}$. The third inequality holds by Lemma 4.9 of [23]. By collecting (57), (58) and (59), we have that there exists some positive constant C_4 depending only on M_0 and r such that

$$(54) \leq C_4 k^r \frac{5^{k/2}}{n^{r/2}} E_{\Sigma_0}(\|\hat{\Sigma}\|^r), \tag{60}$$

for all sufficiently large n .

For the upper bound of $E_{\Sigma_0}(\|\hat{\Sigma}\|^r)$, there exist some positive constants C_5 and C_6 depending only on M_0 and r such that

$$\begin{aligned} E_{\Sigma_0}(\|\hat{\Sigma}\|^r) &\leq 2^{r-1}E_{\Sigma_0}(\|S_n\|^r) + \frac{2^{r-1}}{n^r}\|A_n\|^r \\ &\leq 4^{r-1}E_{\Sigma_0}(\|S_n - \Sigma_0\|^r) + 4^{r-1}\|\Sigma_0\|^r + \frac{2^{r-1}}{n^r}\|A_n\|^r \\ &\leq C_5 + C_6\frac{5^k}{n^{r/2}}, \end{aligned} \tag{61}$$

for all sufficiently large n . The last inequality holds by Lemma A.6. Collecting (60) and (61), we have

$$(54) \leq C_7 k^r \left(\frac{5^{k/2}}{n^{r/2}} + \frac{5^{3k/2}}{n^r} \right),$$

for some positive constant C_7 depending only on M_0 and r . Combining this inequality with (56), we complete the proof. \square

Lemma A.8. *Suppose the same setting of Lemma A.7. If $(\tau_n - 2k) \vee \|A_n\| \vee k = o(n)$, $\|\Sigma_0\| \leq M_0$ and $x < 4\|\Sigma_0\|$, then there exist some positive constants C and λ depending only on M_0 such that*

$$E_{\Sigma_0}\{P^\pi(\|\Sigma - \Sigma_0\| > x \mid \mathbb{Z}_n)\} \leq C5^k\{\exp(-\lambda nx^2) + \exp(-\lambda n)\},$$

for all sufficiently large n .

Proof. Let $\hat{\Sigma} = (nS_n + A_n)/(n + \tau_n - k - 1)$. We have

$$\begin{aligned} E_{\Sigma_0}\{P^\pi(\|\Sigma - \Sigma_0\| > x \mid \mathbb{Z}_n)\} &\leq E_{\Sigma_0}\{P^\pi(\|\Sigma - \hat{\Sigma}\| > x/2 \mid \mathbb{Z}_n)\} \\ &\quad + P_{\Sigma_0}(\|\hat{\Sigma} - \Sigma_0\| > x/2). \end{aligned}$$

The upper bound of $P_{\Sigma_0}(\|\hat{\Sigma} - \Sigma_0\| > x/2)$ is given as

$$\begin{aligned} &P_{\Sigma_0}(\|\hat{\Sigma} - \Sigma_0\| > x/2) \\ &= P\left(\left\|\frac{n}{n + \tau_n - k - 1}(S_n - \Sigma_0) + \frac{A_n - (\tau_n - k - 1)\Sigma_0}{n + \tau_n - k - 1}\right\| > x/2\right) \\ &\leq P(\|S_n - \Sigma_0\| > x/4) + I\left(\|A_n\|/n + \frac{\tau_n - k - 1}{n + \tau_n - k - 1}\|\Sigma_0\| > x/4\right) \\ &\leq P(\|\Sigma_0\| \|\Omega_n^{(k)} - I_k\| > x/4) + I\left(\|A_n\|/n + \frac{\tau_n - k - 1}{n + \tau_n - k - 1}\|\Sigma_0\| > x/4\right) \\ &= P\{\|\Omega_n^{(k)} - I_k\| > x/(4\|\Sigma_0\|)\} + I\left(\|A_n\|/n + \frac{\tau_n - k - 1}{n + \tau_n - k - 1}\|\Sigma_0\| > x/4\right), \end{aligned}$$

where $\Omega_n^{(k)} = \Sigma_0^{-1/2}S_n\Sigma_0^{-1/2}$, of which distribution is $W_k(I_k/n, n)$. Since $\|A_n\| = o(n)$ and $(\tau_n - k - 1)/(n + \tau_n - k - 1) = o(1)$, we have

$$I\left(\|A_n\|/n + \frac{\tau_n - k - 1}{n + \tau_n - k - 1}\|\Sigma_0\| > x/4\right) = 0,$$

for all sufficiently large n . Since $x < 4\|\Sigma_0\|$, Lemma 4.4 in [23] gives

$$P\{\|\Omega_n^{(k)} - I_k\| > x/(4\|\Sigma_0\|)\} \leq 2 \times 5^k \exp\{-nx^2/(2^{11}\|\Sigma_0\|^2)\}.$$

Thus, we have

$$P_{\Sigma_0}(\|\hat{\Sigma} - \Sigma_0\| > x/2) \leq 2 \times 5^k \exp\{-nx^2/(2^{11}\|\Sigma_0\|^2)\}, \tag{62}$$

for all sufficiently large n .

Next, we show the upper bound of $E_{\Sigma_0}\{P^\pi(\|\Sigma - \hat{\Sigma}\| > x \mid \mathbb{Z}_n)\}$. For arbitrary positive constants R_1 and R_2 , we have

$$\begin{aligned} & E_{\Sigma_0}\{P^\pi(\|\Sigma - \hat{\Sigma}\| > x/2 \mid \mathbb{Z}_n)\} \\ &= E\{P^\pi(\|\hat{\Sigma}^{1/2}(\Omega_{n^*}^{(k)})^{-1}(I_k - \Omega_{n^*}^{(k)})\hat{\Sigma}^{1/2}\| > x/2 \mid \mathbb{Z}_n)\} \\ &\leq E\{P^\pi(\|\hat{\Sigma}\| \|(\Omega_{n^*}^{(k)})^{-1}\| \|I_k - \Omega_{n^*}^{(k)}\| > x \mid \mathbb{Z}_n)\} \\ &\leq E\{P^\pi(R_1\|(\Omega_{n^*}^{(k)})^{-1}\| \|I_k - \Omega_{n^*}^{(k)}\| > x \mid \mathbb{Z}_n)\} + P(\|\hat{\Sigma}\| > R_1) \\ &\leq E\{P^\pi(R_1R_2\|I_k - \Omega_{n^*}^{(k)}\| > x \mid \mathbb{Z}_n)\} \\ &\quad + E\{P^\pi\{\lambda_{\min}(\Omega_{n^*}^{(k)}) < R_2^{-1} \mid \mathbb{Z}_n\} + P(\|\hat{\Sigma}\| > R_1), \end{aligned} \tag{63}$$

where $\Omega_{n^*}^{(k)} = \hat{\Sigma}^{1/2}\Sigma^{-1}\hat{\Sigma}^{1/2}$, of which distribution is $W_k(I_k/(n + \tau_n - k - 1), n + \tau_n - k - 1)$. If we set $R_2^{-1} \leq [1 - \{k/(n + \tau_n - k - 1)\}^{1/2}]^2/4$, Lemma 4.2 in [23] gives that there exists some positive constant λ_1 depending only on M_0 and M_1 such that

$$\begin{aligned} & P^\pi\{\lambda_{\min}(\Omega_n^{(k)}) < R_2^{-1} \mid \mathbb{Z}_n\} \\ &\leq 2 \exp\{-(n + \tau_n - k - 1)[1 - \{k/(n + \tau_n - k - 1)\}^{1/2}]^2/8\} \\ &\leq 2 \exp(-\lambda_1 n), \end{aligned} \tag{64}$$

for all sufficiently large n . If $R_1 \in (\|\Sigma_0\|, 3\|\Sigma_0\|)$, then

$$\begin{aligned} P(\|\hat{\Sigma}\| > R_1) &\leq P(\|\hat{\Sigma} - \Sigma_0\| > R_1 - \|\Sigma_0\|) \\ &\leq 2 \times 5^k \exp\{-n(R_1 - \|\Sigma_0\|)^2/(2^9\|\Sigma_0\|^2)\}, \end{aligned} \tag{65}$$

where the second inequality holds by (62). Since we can set R_2 large enough to satisfy $x/(R_1R_2) < 1$, Lemma 4.4 in [23] gives

$$\begin{aligned} & P^\pi(R_1R_2\|I_k - \Omega_{n^*}^{(k)}\| > x \mid \mathbb{Z}_n) \\ &= P^\pi(\|I_k - \Omega_{n^*}^{(k)}\| > x/(R_1R_2) \mid \mathbb{Z}_n) \\ &\leq 2 \times 5^k \exp\left\{-\frac{(n + \tau_n - k - 1)x^2}{2^7(R_1R_2)^2}\right\}. \end{aligned} \tag{66}$$

By collecting the inequalities (63)-(66), we obtain that there exist some positive constants C_1 and λ_2 depending only on M_0 and M_1 such that

$$E_{\Sigma_0}\{P^\pi(\|\Sigma - \hat{\Sigma}\| > x/2 \mid \mathbb{Z}_n)\} \leq C_1 5^k \{\exp(-\lambda_2 nx^2) + \exp(-\lambda_2 n)\},$$

for all sufficiently large n . Combining this inequality with (62), we complete the proof. \square

Proof of Lemma A.1. By Lemma 5.3 and A.5, we have

$$\begin{aligned}
 & E_{\Sigma_0} \{E^{\pi^i} (\|T_k^{(\epsilon_n)}(\Sigma) - \Sigma_0\|^4 \mid \mathbb{Z}_n)\} \\
 & \leq 2^7 E_{\Sigma_0} (E^{\pi^i} [\|T_k(\Sigma) - \Sigma_0\|^4 \mid \mathbb{Z}_n]) + 2^6 \epsilon_n^4 \\
 & \leq 2^{10} E_{\Sigma_0} \{E^{\pi^i} (\|T_k(\Sigma - \Sigma_0)\|^4 \mid \mathbb{Z}_n)\} + 2^{10} \|T_k(\Sigma_0) - \Sigma_0\|^4 + 2^6 \epsilon_n^4 \\
 & \leq 2^{10} 3^4 E_{\Sigma_0} \{E^{\pi^i} (\max_{1 \leq i \leq p} \|M_i^{(k)}(\Sigma - \Sigma_0)\|^4 \mid \mathbb{Z}_n)\} \\
 & \quad + 2^{10} \{M(\lfloor k/2 \rfloor)^{-\alpha}\}^4 + 2^6 \epsilon_n^4. \tag{67}
 \end{aligned}$$

Next, we have

$$\begin{aligned}
 & E_{\Sigma_0} \{E^{\pi^i} (\max_{1 \leq i \leq p} \|M_i^{(k)}(\Sigma - \Sigma_0)\|^4 \mid \mathbb{Z}_n)\} \\
 & \leq x^4 + E_{\Sigma_0} \{E^{\pi^i} (\max_{1 \leq i \leq p} \|M_i^{(k)}(\Sigma - \Sigma_0)\|^4 \\
 & \quad I(\max_{1 \leq i \leq p} \|M_i^{(k)}(\Sigma - \Sigma_0)\| > x \mid \mathbb{Z}_n))\} \\
 & \leq x^4 + E_{\Sigma_0} \{E^{\pi^i} (\max_{1 \leq i \leq p} \|M_i^{(k)}(\Sigma - \Sigma_0)\|^8 \mid \mathbb{Z}_n)\}^{1/2} \\
 & \quad \times E_{\Sigma_0} \{P^{\pi^i} (\max_{1 \leq i \leq p} \|M_i^{(k)}(\Sigma - \Sigma_0)\| > x \mid \mathbb{Z}_n)\}^{1/2} \\
 & \leq x^4 + p \max_{1 \leq i \leq p} E_{\Sigma_0} \{E^{\pi^i} (\|M_i^{(k)}(\Sigma - \Sigma_0)\|^8 \mid \mathbb{Z}_n)\}^{1/2} \\
 & \quad \times \max_{1 \leq i \leq p} E_{\Sigma_0} \{P^{\pi^i} (\|M_i^{(k)}(\Sigma - \Sigma_0)\| > x \mid \mathbb{Z}_n)\}^{1/2}.
 \end{aligned}$$

Note that $M_i^{(k)}(\Sigma_0) \in \mathcal{F}_{k,\alpha}(M, M_0, M_1)$ and $[M_l^{(k)}(\Sigma) \mid \mathbb{Z}_n] \sim IW_k(M_l^{(k)}(nS_n + A_n), n + \nu_n - 2p + 2k)$ [28]. If $x < 4\|M_l^{(k)}(\Sigma_0)\|$, then Lemmas A.7 and A.8 give that there exist some positive constants C_1 and λ_1 depending only on M_0 and M_1 such that

$$\begin{aligned}
 & p \max_{1 \leq i \leq p} E_{\Sigma_0} \{E^{\pi^i} (\|M_i^{(k)}(\Sigma - \Sigma_0)\|^8 \mid \mathbb{Z}_n)\}^{1/2} \\
 & \quad \max_{1 \leq i \leq p} E_{\Sigma_0} \{P^{\pi^i} (\|M_i^{(k)}(\Sigma - \Sigma_0)\| > x \mid \mathbb{Z}_n)\}^{1/2} \\
 & \leq C_1 p k^4 5^{5k/2} n^{-2} \{\exp(-\lambda_1 n x^2) + \exp(-\lambda_1 n)\}, \tag{68}
 \end{aligned}$$

for all sufficiently large n . By setting $x = 3 \log p / (\lambda_1 n)$, which is smaller than an arbitrary positive constant for sufficiently large n , We show that (68) is bounded above by some positive constant for all sufficiently large n . Since $\epsilon_n = O(1)$, (67) is bounded by some positive constant depending only on M_0, M_1, M and α . \square

Proof of Lemma A.2. By inequality (45), we have

$$I[\lambda_{\min}\{T_k(\Sigma)\} \leq c] \leq I(\|T_k(\Sigma - \Sigma_0)\| \geq c_k)$$

where $c_k = \lambda_{\min}(\Sigma_0) - M(\lfloor k/2 \rfloor)^{-\alpha} - c$. Since $c \leq \lambda_{\min}(\Sigma_0)/2$ and $\lfloor k/2 \rfloor >$

$\{4M/\lambda_{\min}(\Sigma_0)\}^{1/\alpha}$, we have $c_k \geq \lambda_{\min}(\Sigma_0)/4$. There exist some positive constants C and λ depending only on M_0 and M_1 such that

$$\begin{aligned} & E_{\Sigma_0}(P^{\pi^i}[\lambda_{\min}\{T_k(\Sigma)\} \leq c \mid \mathbb{Z}_n]) \\ & \leq E_{\Sigma_0}\{P^{\pi^i}(\lambda_{\min}(\|T_k(\Sigma - \Sigma_0)\|) \geq c_k \mid \mathbb{Z}_n)\} \\ & \leq p \max_{1 \leq l \leq p} E_{\Sigma_0}\{P^{\pi^i}(\|M_l^{(k)}(\Sigma - \Sigma_0)\| \geq c_k/3 \mid \mathbb{Z}_n)\} \\ & \leq pC5^k[\exp\{-\lambda n(c_k/3)^2\} + \exp(-\lambda n)], \end{aligned}$$

for all sufficiently large n . The second inequality is satisfied by Lemma A.5. For the third inequality, Lemma A.8 is used. Note $M_l^{(k)}(\Sigma_0) \in \mathcal{F}_{k,\alpha}(M, M_0, M_1)$, $[M_l^{(k)}(\Sigma) \mid \mathbb{Z}_n] \sim IW_k\{M_l^{(k)}(nS_n + A_n), n + \nu_n - 2p + 2k\}$ and $c_k/3$ satisfies

$$\begin{aligned} c_k/3 & \leq \lambda_{\min}(\Sigma_0) \\ & \leq \lambda_{\min}\{M_l^{(k)}(\Sigma_0)\} \\ & \leq 4\|M_l^{(k)}(\Sigma_0)\|, \end{aligned}$$

for all $l \in \{1, 2, \dots, p\}$. □

Proof of Lemma A.3. By Lemmas 5.3 and A.5, we have

$$\begin{aligned} & E_{\Sigma_0}\{E^{\pi^i}(\|\Sigma_0 - T_k^{(\epsilon_n)}(\Sigma)\|^2 \mid \mathbb{Z}_n)\} \\ & \leq 2^3 E_{\Sigma_0}\{E^{\pi^i}(\|\Sigma_0 - T_k(\Sigma)\|^2 \mid \mathbb{Z}_n)\} + 4\epsilon_n^2 \\ & \leq 2^4 E_{\Sigma_0}\{E^{\pi^i}(\|T_k(\Sigma_0) - T_k(\Sigma)\|^2 \mid \mathbb{Z}_n)\} + 2^4\|T_k(\Sigma_0) - \Sigma_0\|^2 + 4\epsilon_n^2 \\ & \leq 2^4 3^2 E_{\Sigma_0}\{E^{\pi^i}(\max_{1 \leq l \leq p} \|M_l^{(k)}(\Sigma_0 - \Sigma)\|^2 \mid \mathbb{Z}_n)\} \\ & \quad + 2^4(M\lfloor k/2 \rfloor^{-\alpha})^2 + 4\epsilon_n^2. \end{aligned} \tag{69}$$

Note $M_l^{(k)}(\Sigma_0) \in \mathcal{F}_{k,\alpha}(M, M_0, M_1)$, $[M_l^{(k)}(\Sigma) \mid \mathbb{Z}_n] \sim IW_k\{M_l^{(k)}(nS_n + A_n), n + \nu_n - 2p + 2k\}$. Thus, for an arbitrary positive real number x

with $x < 4 \min_{1 \leq l \leq p} \|M_l^{(k)}(\Sigma_0)\|$, there exist some positive constants C_1 and λ_1 depending only on M_0 such that

$$\begin{aligned} & E_{\Sigma_0}\{E^{\pi^i}(\max_{1 \leq l \leq p} \|M_l^{(k)}(\Sigma_0 - \Sigma)\|^2 \mid \mathbb{Z}_n)\} \\ & \leq x^2 + E_{\Sigma_0}[E^{\pi^i}\{\max_{1 \leq l \leq p} \|M_l^{(k)}(\Sigma_0 - \Sigma)\|^2 I \\ & \quad (\max_{1 \leq l \leq p} \|M_l^{(k)}(\Sigma_0 - \Sigma)\| > x) \mid \mathbb{Z}_n\}] \\ & \leq x^2 + p \max_{1 \leq l \leq p} E_{\Sigma_0}\{E^{\pi^i}(\|M_l^{(k)}(\Sigma_0 - \Sigma)\|^4 \mid \mathbb{Z}_n)\}^{1/2} \\ & \quad \times \max_{1 \leq l \leq p} E_{\Sigma_0}\{E^{\pi^i}(\|M_l^{(k)}(\Sigma_0 - \Sigma)\| > x \mid \mathbb{Z}_n)\}^{1/2} \\ & \leq x^2 + C_1 p k^2 \frac{5^3}{n} \{ \exp(-\lambda_1 n x^2) + \exp(-\lambda_1 n) \}, \end{aligned}$$

for all sufficiently large n . The last inequality holds by Lemmas A.7 and A.8. By setting $x^2 = 2(\log p + k \log 5)/n\lambda_1$, which is lower than an arbitrary positive

constant for all sufficiently large n , we have

$$E_{\Sigma_0} \{ E^{\pi^i} (\max_{1 \leq l \leq p} \|M_l^{(k)}(\Sigma_0 - \Sigma)\|^2 \mid \mathbb{Z}_n) \} \leq C_2 \left(\frac{\log p + k}{n} \right),$$

for some positive constants C_2 depending only on M_0 . Combining this inequality with (69), we complete the proof. \square

Proof of Lemma A.4. Let $c = \lambda_{\min}(\Sigma_0)/2$.

Since $\lambda_{\min}\{T_k^{(\epsilon_n)}(\Sigma)_{11}\} \geq \lambda_{\min}\{T_k^{(\epsilon_n)}(\Sigma)\} \geq \epsilon_n$, we have

$$\begin{aligned} & E_{\Sigma_0} \{ E^{\pi^i} (\|\Sigma_{0,11}^{-1} - T_k^{(\epsilon_n)}(\Sigma)_{11}^{-1}\|^2 \mid \mathbb{Z}_n) \} \\ & \leq \|\Sigma_0^{-1}\|^2 E_{\Sigma_0} \{ E^{\pi^i} (\|T_k^{(\epsilon_n)}(\Sigma)_{11}^{-1}\|^2 \|\Sigma_{0,11} - T_k^{(\epsilon_n)}(\Sigma)_{11}\|^2 \mid \mathbb{Z}_n) \} \\ & \leq \frac{1}{M_1^2 c^2} E_{\Sigma_0} \{ E^{\pi^i} (\|\Sigma_{0,11} - T_k^{(\epsilon_n)}(\Sigma)_{11}\|^2 \mid \mathbb{Z}_n) \} \\ & \quad + \frac{1}{M_1^2 \epsilon_n^2} E_{\Sigma_0} \{ E^{\pi^i} (\|\Sigma_{0,11} - T_k^{(\epsilon_n)}(\Sigma)_{11}\|^2 I[\lambda_{\min}\{T_k(\Sigma)\} \leq c] \mid \mathbb{Z}_n) \} \\ & \leq \frac{1}{M_1^2 c^2} E_{\Sigma_0} \{ E^{\pi^i} (\|\Sigma_{0,11} - T_k^{(\epsilon_n)}(\Sigma)_{11}\|^2 \mid \mathbb{Z}_n) \} \tag{70} \\ & \quad + \frac{1}{M_1^2 \epsilon_n^2} E_{\Sigma_0} \{ E^{\pi^i} (\|\Sigma_{0,11} - T_k^{(\epsilon_n)}(\Sigma)_{11}\|^4 \mid \mathbb{Z}_n) \}^{1/2} \end{aligned}$$

$$E_{\Sigma_0} \{ E^{\pi^i} (I[\lambda_{\min}\{T_k(\Sigma)\} \leq c] \mid \mathbb{Z}_n) \}^{1/2}. \tag{71}$$

For the upper bound of (70), Lemma A.3 gives

$$\begin{aligned} E_{\Sigma_0} \{ E^{\pi^i} (\|\Sigma_{0,11} - T_k^{(\epsilon_n)}(\Sigma)_{11}\|^2 \mid \mathbb{Z}_n) \} & \leq E_{\Sigma_0} \{ E^{\pi^i} (\|\Sigma_0 - T_k^{(\epsilon_n)}(\Sigma)\|^2 \mid \mathbb{Z}_n) \} \\ & \leq C_1 \left(\frac{k + \log p}{n} + k^{-2\alpha} + \epsilon_n^2 \right), \end{aligned}$$

for some positive constant C_1 depending only on M , M_0 , M_1 and α . Lemmas A.1 and A.2 gives that there exist positive constants C_2 and λ depending only on M , M_0 , M_1 and α such that

$$(71) \leq C_2 \frac{p^{1/2} 5^{k/2}}{\epsilon_n^2} \exp(-\lambda_1 n),$$

for all sufficiently large n . Collecting the upper bound of (70) and (71), we complete the proof. \square

References

- [1] Bashir, A., Carvalho, C. M., Hahn, P. R., Jones, M. B. et al. (2018). Post-processing posteriors over precision matrices to produce sparse graph estimates, *Bayesian Analysis*. [MR4044846](#)
- [2] Bickel, P. J. and Levina, E. (2008). Regularized estimation of large covariance matrices, *The Annals of Statistics* pp. 199–227. [MR2387969](#)

- [3] Bühlmann, P., Kalisch, M. and Maathuis, M. H. (2010). Variable selection in high-dimensional linear models: partially faithful distributions and the pc-simple algorithm, *Biometrika* **97**(2): 261–278. [MR2650737](#)
- [4] Cai, T., Ma, Z. and Wu, Y. (2015). Optimal estimation and rank detection for sparse spiked covariance matrices, *Probability theory and related fields* **161**(3-4): 781–815. [MR3334281](#)
- [5] Cai, T. T., Ma, Z., Wu, Y. et al. (2013). Sparse PCA: optimal rates and adaptive estimation, *The Annals of Statistics* **41**(6): 3074–3110. [MR3161458](#)
- [6] Cai, T. T. and Zhou, H. H. (2010). Optimal rates of convergence for covariance matrix estimation, *The Annals of Statistics* **38**(4): 2118–2144. [MR2676885](#)
- [7] Cai, T. T. and Zhou, H. H. (2012). Minimax estimation of large covariance matrices under l_1 -norm, *Statistica Sinica* pp. 1319–1349. [MR3027084](#)
- [8] Chakraborty, M. and Ghosal, S. (2020). Convergence rates for Bayesian estimation and testing in monotone regression, *arXiv preprint arXiv:2008.01244*. [MR4280172](#)
- [9] Chen, K., Dong, H. and Chan, K.-S. (2013). Reduced rank regression via adaptive nuclear norm penalization, *Biometrika* **100**(4): 901–920. [MR3142340](#)
- [10] Chen, L. and Huang, J. Z. (2012). Sparse reduced-rank regression for simultaneous dimension reduction and variable selection, *Journal of the American Statistical Association* **107**(500): 1533–1545. [MR3036414](#)
- [11] Demko, S., Moss, W. F. and Smith, P. W. (1984). Decay rates for inverses of band matrices, *Mathematics of computation* **43**(168): 491–499. [MR0758197](#)
- [12] Dunson, D. B. and Neelon, B. (2003). Bayesian inference on order-constrained parameters in generalized linear models, *Biometrics* **59**(2): 286–295. [MR1987395](#)
- [13] Fan, J., Rigollet, P. and Wang, W. (2015). Estimation of functionals of sparse covariance matrices, *Annals of statistics* **43**(6): 2706. [MR3405609](#)
- [14] Fan, J., Weng, H. and Zhou, Y. (2019). Optimal estimation of functionals of high-dimensional mean and covariance matrix, *arXiv preprint arXiv:1908.07460*.
- [15] Gelman, A., Hwang, J. and Vehtari, A. (2014). Understanding predictive information criteria for bayesian models, *Statistics and computing* **24**(6): 997–1016. [MR3253850](#)
- [16] Golub, G. H. and Van Loan, C. F. (2013). Matrix computations, 4th, *Johns Hopkins*. [MR3024913](#)
- [17] Gunn, L. H. and Dunson, D. B. (2005). A transformation approach for incorporating monotone or unimodal constraints, *Biostatistics* **6**(3): 434–449.
- [18] Johnstone, I. M. and Lu, A. Y. (2009). On consistency and sparsity for principal components analysis in high dimensions, *Journal of the American Statistical Association* **104**(486): 682–693. [MR2751448](#)
- [19] Kauermann, G. (1996). On a dualization of graphical gaussian models, *Scandinavian journal of statistics* pp. 105–116. [MR1380485](#)

- [20] Khare, K., Rajaratnam, B. et al. (2011). Wishart distributions for decomposable covariance graph models, *The Annals of Statistics* **39**(1): 514–555. [MR2797855](#)
- [21] Lauritzen, S. L. (1996). *Graphical models*, Vol. 17, Clarendon Press. [MR1419991](#)
- [22] Lee, K. and Lee, J. (2018). Optimal Bayesian minimax rates for unconstrained large covariance matrices, *Bayesian Analysis* **13**(4): 1211–1229. [MR3855369](#)
- [23] Lee, K., Lee, K. and Lee, J. (2020). Post-processed posteriors for banded covariances, *arXiv preprint arXiv:2011.12627*.
- [24] Lee, W. and Liu, Y. (2012). Simultaneous multiple response regression and inverse covariance matrix estimation via penalized gaussian maximum likelihood, *Journal of multivariate analysis* **111**: 241–255. [MR2944419](#)
- [25] Li, R., Liu, J. and Lou, L. (2017). Variable selection via partial correlation, *Statistica Sinica* **27**(3): 983. [MR3699690](#)
- [26] Lin, L. and Dunson, D. B. (2014). Bayesian monotone regression using gaussian process projection, *Biometrika* **101**(2): 303–317. [MR3215349](#)
- [27] Patra, S. and Dunson, D. B. (2018). Constrained Bayesian inference through posterior projections, *arXiv preprint arXiv:1812.05741*. [MR4035485](#)
- [28] Press, S. J. (2012). *Applied multivariate analysis: using Bayesian and frequentist methods of inference*, Courier Corporation.
- [29] Qian, J., Tanigawa, Y., Li, R., Tibshirani, R., Rivas, M. A. and Hastie, T. (2020). Large-scale sparse regression for multiple responses with applications to UK biobank, *BioRxiv*.
- [30] Rothman, A. J., Levina, E. and Zhu, J. (2010). Sparse multivariate regression with covariance estimation, *Journal of Computational and Graphical Statistics* **19**(4): 947–962. [MR2791263](#)
- [31] Silva, R. and Ghahramani, Z. (2009). The hidden life of latent variables: Bayesian learning with mixed graph models, *The Journal of Machine Learning Research* **10**: 1187–1238. [MR2520804](#)
- [32] Transport Operation & Information Service (2021). *Seoul traffic information*. <https://topis.seoul.go.kr/>
- [33] Uematsu, Y., Fan, Y., Chen, K., Lv, J. and Lin, W. (2019). SOFAR: large-scale association network learning, *IEEE Transactions on Information Theory* **65**(8): 4924–4939. [MR3988531](#)
- [34] Xiao, L. and Bunea, F. (2014). On the theoretic and practical merits of the banding estimator for large covariance matrices, *arXiv preprint arXiv:1402.0844*.
- [35] Yin, J. and Li, H. (2011). A sparse conditional Gaussian graphical model for analysis of genetical genomics data, *The annals of applied statistics* **5**(4): 2630. [MR2907129](#)
- [36] Zhao, R., Gu, X., Xue, B., Zhang, J. and Ren, W. (2018). Short period PM_{2.5} prediction based on multivariate linear regression model, *PloS one* **13**(7): e0201011.