Statistica Sinica **31** (2021), 173-196
doi:https://doi.org/10.5705/ss.202018.0476

ESTIMATING LARGE PRECISION MATRICES VIA MODIFIED CHOLESKY DECOMPOSITION

Kyoungjae Lee and Jaeyong Lee

Inha University and Seoul National University

Abstract: We introduce a k-banded Cholesky prior for estimating high-dimensional bandable precision matrices using a modified Cholesky decomposition. The bandable assumption is imposed on the Cholesky factor of the decomposition. We obtain the P-loss convergence rate under the spectral norm and the matrix ℓ_∞-norm, as well as the minimax lower bounds. Because the P-loss convergence rate is stronger than the posterior convergence rate, the rates obtained are also posterior convergence rates. Furthermore, when the true precision matrix is a k_0-banded matrix, for some finite k_0, we obtain the minimax rate. The established convergence rates for bandable precision matrices are slightly slower than the minimax lower bounds, but are the fastest of the existing Bayesian approaches. Simulation results show that the performance of the proposed method is better than or comparable to that of competitive estimators.

Key words and phrases: Modified Cholesky decomposition, P-loss convergence rate, precision matrix.

1. Introduction

Today, it is not uncommon to find that the number of variables p in a data set is much larger than the sample size n. Such high-dimensional data sets arise in studies on genomics, climatology, fMRI, and neuroimaging, among many others. In this study, we estimate the precision matrix (i.e., the inverse of the covariance matrix) for high-dimensional data.

When the number of variables p tends to infinity as $n \longrightarrow \infty$, and is possibly larger than n, the traditional sample covariance fails to converge to the true covariance marix (Johnstone and Lu (2009)). Thus, it is necessary to assume certain constraints on the covariance to obtain a consistent estimator in an ultra high-dimensional setting, $\log p = o(n)$. These constraints include sparse, bandable assumptions or lower-dimensional structures such as the sparse spiked covariance and factor model. The minimax convergence rates under the sparsity

Corresponding author: Kyoungjae Lee, Department of Statistics, Inha University, Nam-gu, Incheon, South Korea. E-mail: leekjstat@gmail.com.

or bandable assumption on a covariance/precision matrix have been established by Bickel and Levina (2008a), Bickel and Levina (2008b), Cai, Zhang and Zhou (2010), Cai and Zhou (2012a), Cai and Zhou (2012b), Xue and Zou (2013), Cai, Liu and Zhou (2016), and Hu and Negahban (2017), among others. Bickel and Levina (2008b) and Verzelen (2010) obtained convergence rates for precision matrices under the sparsity or bandable assumption using a Cholesky decomposition. The convergence rates for lower-dimensional structures of covariance matrices, such as the factor model (Fan, Fan and Lv (2008)) and sparse spiked covariance model (Cai, Ma and Wu (2015)), have also been explored. Cai, Liang and Zhou (2015) and Fan, Rigollet and Wang (2015) derived the minimax convergence rates for the functionals of the covariance matrices. Cai, Ren and Zhou (2016) provide a comprehensive review on convergence rates for large matrices.

From a Bayesian perspective, although the posterior convergence rates for large covariance or precision matrices have been investigated, few works have done so for high-dimensional settings. Banerjee and Ghosal (2015) showed the posterior convergence rate for a precision matrix under the sparsity assumption, using a mixture prior for off-diagonal elements of the precision matrix to assign exactly zero. To estimate bandable precision matrices, Banerjee and Ghosal (2014) used the G-Wishart prior on the precision matrix to establish the posterior convergence rate. Xiang, Khare and Ghosh (2015) extended the result of Banerjee and Ghosal (2014) to decomposable graphical models, which include bandable precision matrices as a special case. Pati et al. (2014) considered the posterior convergence rate for covariance estimations using a sparse factor model, obtaining nearly optimal rates, the minimax rates with $(\log n)^{1/2}$ factor, when the number of true factors is bounded. Gao and Zhou (2015) derived the optimal posterior convergence rate for covariance matrices under a sparse spiked covariance model. Cao, Khare and Ghosh (2016) considered the sparse Cholesky factor of the precision matrix and proved the strong model selection consistency and convergence rate. The above results assume an ultra high-dimensional setting, $\log p = o(n)$, or a variant thereof. Recently, Gao and Zhou (2016) derived Bernstein-von Mises theorems for functionals of the covariance matrix and its inverse under conditions such as $p = o(n)$ or $p^3 = o(n)$.

Lee and Lee (2018) proposed a new decision theoretical framework for prior selection and obtained the Bayesian minimax rate of the unconstrained covariance matrix under the spectral norm, for all rates of p. They also obtained the Bayesian minimax rates under the Frobenius norm, Bregman divergence, and squared log-determinant loss when $p \leq n^{1/2}$ or $p = o(n)$. They showed that

when $p > n/2$, there is no better prior than the point mass prior δ_{I_p}, in terms of the induced posterior convergence rate. This implies that restrictions on the covariance or precision matrix are needed to obtain consistent estimators.

In this study, we consider a class of bandable precision matrices and a modified Cholesky decomposition (MCD) in an ultra high-dimensional setting, and derive the P-loss convergence rates under the spectral norm and the matrix ℓ_∞-norm. Because the P-loss convergence rate implies the traditional posterior convergence rate, it can be viewed as the posterior convergence rate. The bandable assumption is imposed on the lower triangular matrix from the MCD, which is called the Cholesky factor. Bickel and Levina (2008b) used a similar assumption, and their parameter space is a special case of ours. Our work is also closely related to the works of Banerjee and Ghosal (2014) and Xiang, Khare and Ghosh (2015), who also considered bandable precision matrices. However, we obtain the minimax rate when the true precision matrix is k_0-banded, for some finite k_0. Furthermore, when the true precision matrix is bandable, the convergence rate obtained using the proposed method is faster than those obtained in using existing methods. To the best of our knowledge, this is the fastest rate obtained using a Bayesian method for bandable precision matrices. Although our parameter space is not the same as that of Banerjee and Ghosal (2014), the two are closely related. Proposition 1 describes this relationship. This study is also related to the work of Cao, Khare and Ghosh (2016). However, they considered only sparse Cholesky factors in which most elements are exactly zero, which does not cover the class of bandable Cholesky factors considered here. Furthermore, we show the minimax lower bounds for precision matrices under the bandable assumption on the Cholesky factor. The lower bounds are derived under the spectral norm and matrix ℓ_∞-norm. Recently, Liu and Ren (2017) obtained a sharper lower bound for the spectral norm under the bandable assumption on the Cholesky factor, concurrently with our work.

The rest of the paper is organized as follows. In Section 2, we define our model, the matrix norms, the parameter class, and the decision theoretic prior selection. The convergence rates for precision matrices under the spectral norm and matrix ℓ_∞-norm are shown in Section 3. In Section 4, we propose a practical choice of the bandwidth, and in Section 5, we present a simulation study. Section 6 concludes the paper. All proofs of the main results are provided in the online Supplementary Material.

2. Preliminaries

2.1. Norms and notation

For any constants a and b, $a \vee b$ and $a \wedge b$ denote the maximum and minimum of, respectively, a and b. For any positive sequences a_n and b_n, we denote $a_n = o(b_n)$ if $a_n/b_n \longrightarrow 0$ as $n \to \infty$. We denote $a_n \asymp b_n$ if there exist positive constants C_1 and C_2, such that $C_1 \leq a_n/b_n \leq C_2$, for all sufficiently large n, and $a_n \lesssim b_n$ if there exists a positive constant C, such that $a_n \leq C b_n$, for all sufficiently large n. For any $p \times p$ matrix A, $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the minimum and maximum eigenvalues, respectively, of the matrix A.

For any p-dimensional vector a, we define the vector norms as follows: $\|a\|_1 := \sum_{i=1}^p |a_i|$, $\|a\|_2 := (\sum_{i=1}^p a_i^2)^{1/2}$, and $\|a\|_{\max} := \max_{1 \leq i \leq p} |a_i|$. We define the operator norms for the matrices using these norms. Let $A = (a_{ij})$ be a $p \times p$ matrix. Then, the spectral norm (or matrix ℓ_2-norm) is defined by

$$\|A\| := \sup_{\substack{x \in \mathbb{R}^p \\ \|x\|_2 = 1}} \|Ax\|_2 \;=\; (\lambda_{\max}(A^T A))^{1/2}.$$

We define the matrix ℓ_1-norm, matrix ℓ_∞-norm, and Frobenius norm as

$$\|A\|_1 := \sup_{\substack{x \in \mathbb{R}^p \\ \|x\|_1 = 1}} \|Ax\|_1 \;=\; \max_j \sum_{i=1}^p |a_{ij}|,$$

$$\|A\|_\infty := \sup_{\substack{x \in \mathbb{R}^p \\ \|x\|_{\max} = 1}} \|Ax\|_{\max} \;=\; \max_i \sum_{j=1}^p |a_{ij}|,$$

$$\|A\|_F := \left(\sum_{i=1}^p \sum_{j=1}^p a_{ij}^2 \right)^{1/2},$$

respectively. The max norm for the matrices is defined as $\|A\|_{\max} := \max_{i,j} |a_{ij}|$.

2.2. The model and the prior

Suppose we observe a data set from the p-dimensional normal distribution

$$X_1, \ldots, X_n \overset{i.i.d.}{\sim} N_p(0, \Omega_n^{-1}), \tag{2.1}$$

where Ω_n is a $p \times p$ positive-definite matrix. We assume that $p = p_n$ is a function of n, increasing to ∞ as $n \to \infty$. Let $\mathbf{X}_n = (X_1, \ldots, X_n)^T$ and $\Omega_{0,n}$ be the $n \times p$ data matrix and the $p \times p$ true precision matrix, respectively.

For a $p \times p$ positive-definite matrix Ω_n, the MCD guarantees a unique lower-triangular matrix $A_n = (a_{jl})$ and a unique diagonal matrix $D_n = diag(d_j)$ exist, such that

$$\Omega_n = (I_p - A_n)^T D_n^{-1} (I_p - A_n), \tag{2.2}$$

where $a_{jj} = 0$ and $d_j > 0$, for all $j = 1, \ldots, p$. The MCD has a nice autoregressive interpretation that enables simple and effective inferences. Consider a latent variable $\epsilon \sim N_p(0, D_n)$ and a random vector $Y \sim N_p(0, \Omega_n^{-1})$, and note that the relationship $(I_p - A_n)Y \stackrel{d}{=} \epsilon$ holds because they have the same distributions. Therefore model (2.1) with precision matrix (2.2) is equivalent to the following autoregressive model:

$$\begin{aligned} X_{\cdot,1} \mid d_1 &\sim N_n(0, d_1 I_n), \\ X_{\cdot,j} \mid a_j, d_j, X_{\cdot,1:(j-1)} &\sim N_n(X_{\cdot,1:(j-1)} a_j,\, d_j I_n),\ j = 2, \ldots, p, \end{aligned} \tag{2.3}$$

where $a_j = (a_{j1}, \ldots, a_{j,j-1})^T \in \mathbb{R}^{j-1}$, and $X_{\cdot,j} \in \mathbb{R}^n$ and $X_{\cdot,1:(j-1)} \in \mathbb{R}^{n \times (j-1)}$ are submatrices of $\mathbf{X}_n$, consisting of the jth column and the $1, \ldots, (j-1)$th columns, respectively. We denote the submatrix of $\mathbf{X}_n$ consisting of the $a, \ldots, b$th columns as $X_{\cdot,a:b}$, for any positive integer $a \leq b$. With a slight abuse of notation, if $a \leq 0$ and $b > 0$, $X_{\cdot,a:b} := X_{\cdot,(a \vee 1):b} = X_{\cdot,1:b}$ to define a proper column position. The zero patterns in the Cholesky factor A_n and model (2.3) rely on the order of the variables. Here, we assume there is a known natural ordering of the variables as is common in the literature; see Bickel and Levina (2008b), Shojaie and Michailidis (2010), Khare et al. (2016), Banerjee and Ghosal (2014), and Cao, Khare and Ghosh (2016).

Bickel and Levina (2008b) approximated the precision matrix by considering only the k closest regressors in the regression interpretation (2.3), which is the same as assuming that the lower-triangular matrix A_n in the MCD is a k-banded lower-triangular matrix. This approximation assumes that, based on the given ordering of the variables, only the k closest previous variables affect the current variable. Note that the resulting precision matrix $\Omega_n = (I_p - A_n)^T D_n^{-1} (I_p - A_n)$ also becomes a k-banded matrix.

In this paper, we propose the following prior:

$$\begin{aligned} &\pi(a_{jl}) \propto 1, \quad l = (j-k) \vee 1, \ldots, j-1, \text{ and } \pi(a_{jl}) = \delta_0, \text{ otherwise}, \\ &\pi(d_j) \ \propto \ d_j^{-\nu_0/2-1} I(0 < d_j < M), \quad j = 1, \ldots, p, \end{aligned} \tag{2.4}$$

for some nonnegative constants M and ν_0, where ν_0 can depend on n. Here, δ_0 is the Dirac measure at zero. Note that the degenerate prior of a_{jl} when $l < (j-k) \vee 1$ is due to the k-banded Cholesky (k-BC) factor assumption. We call the prior defined in (2.4) the k-BC prior. The appropriate conditions on M and ν_0 are discussed in Section 3. The prior in (2.4) leads to the following joint posterior distribution:

$$\begin{aligned} d_j \mid \mathbf{X}_n &\overset{ind.}{\sim} IG^{Tr}\left(d_j \mid \frac{n_j}{2}, \frac{n}{2}\widehat{d}_{jk}, d_j \le M\right), \ j = 1, \ldots, p, \\ a_j^{(k)} \mid d_j, \mathbf{X}_n &\overset{ind.}{\sim} N_{j-1\wedge k}\left(a_j^{(k)} \middle| \widehat{a}_j^{(k)}, \ d_j\big(X^T_{\cdot,(j-k):(j-1)} X_{\cdot,(j-k):(j-1)}\big)^{-1}\right), \\ & \qquad j = 2, \ldots, p, \end{aligned} \tag{2.5}$$

where $n_j = n + \nu_0 - (j - 1 \wedge k) - 4$, $a_j^{(k)} = (a_{j,(j-k\vee 1)}, \ldots, a_{j,j-1})^T$,

$$\begin{aligned} \widehat{a}_j^{(k)} &= (X^T_{\cdot,(j-k):(j-1)} X_{\cdot,(j-k):(j-1)})^{-1} X^T_{\cdot,(j-k):(j-1)} X_{\cdot,j}, \\ \widehat{d}_{jk} &= n^{-1} X^T_{\cdot,j}\big(I_n - X_{\cdot,(j-k):(j-1)} \\ &\qquad (X^T_{\cdot,(j-k):(j-1)} X_{\cdot,(j-k):(j-1)})^{-1} X^T_{\cdot,(j-k):(j-1)}\big) X_{\cdot,j}, \end{aligned} \tag{2.6}$$

for $j = 2, \ldots, p$, and $\widehat{d}_{1k} = n^{-1}\|X_{\cdot,1}\|_2^2$. We denote $IG^{Tr}(X \mid a, b, A)$ as the truncated version of $IG(X \mid a, b)$ on support A, where $IG(X \mid a, b)$ is the density function of the inverse-gamma random variable X with shape and rate parameters a and b, respectively. Then, $N_p(X \mid \mu, \Sigma)$ is the density function of the p-dimensional normal random variable X with mean vector μ and covariance matrix Σ.

Remark 1. The main results in Section 3 still hold for the prior

$$a_j^{(k)} \mid d_j \overset{ind.}{\sim} N_{j-1\wedge k}\left(m_j, d_j B_j\right), \quad j = 2, \ldots, p, \tag{2.7}$$

with certain bounded conditions on $\|m_j\|_2$ and $\|B_j^{-1}\|$. This includes the prior (2.4) as a special case, with $m_j = 0$ and $B_j = diag(b = \infty)$. However, we omit the proofs, presenting only the results for the prior in (2.4), for simplicity of notation.

Remark 2. The prior for d_j has a compact support in order to deal with the P-loss defined in Section 2.4. If we focus on the posterior convergence rate rather than the P-loss convergence rate, the prior $\pi(d_j) \propto d_j^{-\nu_0/2-1}$ is sufficient to establish the main results in Section 3.

The zero pattern of the Cholesky factor is related to a directed acyclic graph

(Rütimann and Bühlmann (2009)). Using the k-BC prior in (2.4) implies that we approximate the true model by means of a directed Gaussian graphical model. Thus, our method can be applied to such models, but we do not examine this further here. For more information about graphical models, see Lauritzen (1996), Koller and Friedman (2009), and Rütimann and Bühlmann (2009).

2.3. Parameter class

For a given constant $\epsilon_0 > 0$ and a decreasing function $\gamma(k) \to 0$ as $k \longrightarrow \infty$, we define a class of precision matrices, as follows:

$$\mathcal{U}(\epsilon_0, \gamma) = \mathcal{U}_p(\epsilon_0, \gamma) = \Big\{\Omega = (I_p - A)^T D^{-1}(I_p - A) \in \mathcal{C}_p : \tag{2.8}$$
$$\epsilon_0 \le \lambda_{\min}(\Omega) \le \lambda_{\max}(\Omega) \le \epsilon_0^{-1}, \|A - B_k(A)\|_\infty \le \gamma(k),\ \forall 0 < k \le p-1\Big\},$$

where $\mathcal{C}_p$ is the class of all $p \times p$-dimensional positive-definite matrices, and $A = (a_{ij})$ is a lower-triangular matrix from the MCD of Ω, and $B_k(A) := (b_{ij} = a_{ij} I(|i-j| \le k), 1 \le i, j \le p)$. Thus, $\|A - B_k(A)\|_\infty \le \gamma(k)$ is equivalent to $\max_{1 \le i \le p} \sum_{j<i-k} |a_{ij}| \le \gamma(k)$. We consider the following classes of $\gamma(k)$:

1. (polynomially decreasing) $\gamma(k) = Ck^{-\alpha}$, for some $\alpha > 0$ and $C > 0$;
2. (exponentially decreasing) $\gamma(k) = Ce^{-\beta k}$, for some $\beta > 0$ and $C > 0$; and
3. (exact banding) $\gamma(k) = 0$, for some $k_0 > 0$ and all $k > k_0$.

Banerjee and Ghosal (2014) considered a similar parameter space for a precision matrix, defined as

$$\mathcal{U}^*(\epsilon_0, \gamma) = \mathcal{U}_p^*(\epsilon_0, \gamma) = \Bigg\{\Omega = (\omega_{ij}) \in \mathcal{C}_p : 0 < \epsilon_0 \le \lambda_{\min}(\Omega) \le \lambda_{\max}(\Omega) \le \epsilon_0^{-1},$$
$$\max_{1 \le i \le p} \sum_{j:|i-j|>k} |\omega_{ij}| \le \gamma(k),\ \forall 0 < k \le p-1\Bigg\}.$$

If we consider an exact banding $\gamma(k)$ or an exponentially decreasing $\gamma(k)$ with $\beta > \log(\epsilon_0^{-2} + 1)$, the two classes $\mathcal{U}(\epsilon_0, \gamma)$ and $\mathcal{U}^*(\epsilon_0, \gamma)$ are *equivalent* in terms of the convergence rates over these classes. For polynomially decreasing $\gamma(k)$, with $\alpha > 1$, $\Omega \in \mathcal{U}(\epsilon_0, \gamma)$ does not guarantee $\Omega \in \mathcal{U}^*(\epsilon_0, \gamma)$; nevertheless, the two classes remain related. The following proposition describes this relation; the proof is given in the Supplementary Material.

Proposition 1. *Suppose γ is a decreasing function defined on positive integers. If γ is exponentially decreasing with $\gamma(k) = Ce^{-\beta k}$, with $\beta > \log(\epsilon_0^{-2} + 1)$ and $C > 0$, or exact banding for some $k_0 > 0$, then*

$$\mathcal{U}(\epsilon_0, C_1\gamma) \subseteq \mathcal{U}^*(\epsilon_0, \gamma) \subseteq \mathcal{U}(\epsilon_0, C_2\gamma),$$

and if $\gamma(k) = Ck^{-\alpha}$ with $\alpha > 1$, then

$$\mathcal{U}(\epsilon_0, \gamma) \subseteq \mathcal{U}^*(\epsilon_0, C_3\gamma'),$$

where $\gamma'(k) = Ck^{-(\alpha-1)}$, for some positive constants C_1, C_2, and C_3 not depending on p.

2.4. Bayesian minimax rate

The posterior convergence rate is the most commonly used measure for the asymptotic concentration of the posterior around the true parameter (Ghosal, Ghosh and van der Vaart (2000); Ghosal and van der Vaart (2007)). However, even though the concept of a posterior convergence rate is used to justify priors, defining the best possible posterior convergence rate is difficult. Motivated by this difficulty, a new decision theoretic framework for prior selection was suggested by Lee and Lee (2018).

Consider a prior $\pi(\Omega)$ as a decision rule, and define the P-loss as

$$\mathcal{L}(\Omega_{0,n}, \pi) = \mathbb{E}^{\pi}\left(d(\Omega, \Omega_{0,n}) \mid \mathbf{X}_n\right),$$

where $d(\Omega, \Omega')$ is a pseudometric on a set of positive-definite matrices, $\Omega_{0,n}$ is the true precision matrix, and $\mathbb{E}^{\pi}(\cdot|\mathbf{X}_n)$ is the expectation under the posterior of Ω when the prior π and observation $\mathbf{X}_n$ are given. The P-risk is defined as

$$\mathcal{R}(\Omega_{0,n}, \pi) = \mathbb{E}_{0n}\mathbb{E}^{\pi}\left(d(\Omega, \Omega_{0,n}) \mid \mathbf{X}_n\right), \tag{2.9}$$

where $\mathbb{E}_{0n} = \mathbb{E}_{\Omega_{0,n}}$ denotes the expectation with respect to $X_1, \ldots, X_n \overset{i.i.d.}{\sim} N_p(0, \Omega_{0,n}^{-1})$. Let Π_n be the class of all priors on $\mathcal{C}_p$. Then, the Bayesian minimax rate of the posterior for the class $\mathcal{C}_p^* \subset \mathcal{C}_p$ and the space of prior distributions $\Pi_n^* \subset \Pi_n$ is naturally defined as a sequence r_n, such that

$$\inf_{\pi \in \Pi_n^*} \sup_{\Omega_{0,n} \in \mathcal{C}_p^*} \mathbb{E}_{0n}\mathcal{L}\big(\Omega_{0,n}, \pi(\cdot|\mathbf{X}_n)\big) \asymp r_n.$$

If a prior π^* satisfies

$$\sup_{\Omega_{0,n}\in\mathcal{C}_p^*} \mathbb{E}_{0n}\mathcal{L}\big(\Omega_{0,n}, \pi(\cdot|\mathbf{X}_n)\big) \lesssim a_n,$$

then π^* is said to have a P-loss convergence rate a_n, and if a_n has the same rate as the Bayesian minimax rate (i.e., $a_n \asymp r_n$), π^* is said to achieve the Bayesian minimax rate. This new decision-theoretic view of the posterior analysis makes optimal properties conceptually transparent, even if the class of priors and the parameter space are constrained. It also makes it possible to study the optimality properties of pseudo-posteriors, such as the consensus Monte Carlo Scott et al. (2016). The P-loss convergence rate is a stronger measure than the posterior convergence rate, and a frequentist minimax lower bound is also a Bayesian minimax lower bound, in general. See Proposition A.1 and Proposition A.2 in Lee and Lee (2018).

3. Main Results

3.1. P-loss convergence rate and Bayesian minimax lower bound under the spectral norm

In this subsection, we establish the Bayesian minimax lower and upper bounds under the spectral norm. The P-loss convergence rate based on the k-BC prior (2.4) is one of the main results of this study. The rate obtained in Theorem 2 is slightly slower than the rate of a frequentist minimax lower bound given in Theorem 1. The proofs of the theorems are given in the Supplementary Material.

Theorem 1. *Consider model* (2.1) *with* $p \le \exp(cn)$*, for some constant* $c > 0$*. Assume* $\Omega_{0,n} \in \mathcal{U}(\epsilon_0, \gamma)$*, which is defined at* (2.8)*, for a given* $\epsilon_0 > 0$ *and a decreasing function* γ*.*

(i) If there exists a constant $k_0 > 0$ *such that* $\gamma(k) = 0$*, for all* $k \ge k_0$*, we have*

$$\inf_{\widehat{\Omega}_n} \sup_{\Omega_{0,n}\in\mathcal{U}(\epsilon_0,\gamma)} \mathbb{E}_{0n}\|\widehat{\Omega}_n - \Omega_{0,n}\| \gtrsim \left(\frac{\log p}{n}\right)^{1/2},$$

where $\widehat{\Omega}_n$ *denotes an arbitrary estimator of* $\Omega_{0,n}$*.*

(ii) If $\gamma(k) = Ce^{-\beta k}$ *for some constants* $\beta > 0$ *and* $C > 0$*, then we have*

$$\inf_{\widehat{\Omega}_n} \sup_{\Omega_{0,n}\in\mathcal{U}(\epsilon_0,\gamma)} \mathbb{E}_{0n}\|\widehat{\Omega}_n - \Omega_{0,n}\| \gtrsim \min\left\{\left(\frac{\log(n\vee p)}{n}\right)^{1/2}, \left(\frac{p}{n}\right)^{1/2}\right\}.$$

(iii) If $\gamma(k) = Ck^{-\alpha}$ *for some constants* $\alpha > 0$ *and* $C > 0$*, then we have*

$$\inf_{\widehat{\Omega}_n} \sup_{\Omega_{0,n} \in \mathcal{U}(\epsilon_0, \gamma)} \mathbb{E}_{0n} \|\widehat{\Omega}_n - \Omega_{0,n}\| \gtrsim \min \left\{ \left(\frac{\log p}{n} \right)^{1/2} + n^{-\alpha/(2\alpha+1)}, \left(\frac{p}{n} \right)^{1/2} \right\}.$$

The estimation of a precision matrix using a polynomially banded Cholesky factor under the spectral norm has been studied by Bickel and Levina (2008b), although they do not consider a minimax lower bound. Verzelen (2010) obtained a minimax lower bound, but he considered the sparse Cholesky factor under the Frobenius norm.

Cai and Yuan (2016) estimated a covariance operator for random variables on a lattice graph under the spectral norm. They used an exponentially (and polynomially) bandable assumption for the covariance operator. In the one-dimensional lattice case, interestingly, the minimax lower bound in Cai and Yuan (2016) coincides with the minimax lower bound in Theorem 1 (ii). This makes sense, because the two classes are equivalent, by Proposition 1.

Remark 3. Because a frequentist minimax lower bound is also a P-loss minimax lower bound, Theorem 1 yields a P-loss minimax lower bound. For the proof of this argument, see Proposition A.2 in Lee and Lee (2018).

Remark 4. Recently, Liu and Ren (2017) obtained the minimax lower bound with respect to the spectral norm, concurrently with our work. Their lower bound is sharper than that in (iii) of Theorem 1, and they show that it is the minimax rate. However, Liu and Ren (2017) considered only polynomially decreasing $\gamma(k) = Ck^{-\alpha}$, whereas we also provide lower bounds for exponentially decreasing and exactly banded $\gamma(k)$. Furthermore, we consider the "small p" case, $p = o(n)$, whereas Liu and Ren (2017) assumed $n = O(p)$. Specifically, under conditions $\log p = O(n)$ and $n = O(p)$, they proved that

$$\inf_{\widehat{\Omega}_n} \sup_{\Omega_{0,n} \in \mathcal{U}(\epsilon_0, \gamma)} \mathbb{E}_{0n} \|\widehat{\Omega}_n - \Omega_{0,n}\| \gtrsim \left(\frac{\log p}{n} \right)^{1/2} + n^{(-2\alpha+1)/4\alpha}, \tag{3.1}$$

where $\gamma(k) = Ck^{-\alpha}$, for some constants $\alpha > 1/2$ and $C > 0$. Under their assumptions, the lower bound in (iii) of Theorem 1 is quite close to that in (3.1).

The P-loss convergence rate of the k-BC prior in (2.4) under the spectral norm is given in the following theorem.

Theorem 2. *Consider model* (2.1) *and the* k*-BC prior in* (2.4) *for the precision matrix* $\Omega_n = (I_p - A_n)^T D_n^{-1} (I_p - A_n)$*, with* $M \geq 9\epsilon_0^{-1}$ *and* $\nu_0 = o(n)$*, for a*

Table 1. A summary of the P-loss convergence rates and the minimax lower bounds under the spectral norm for various γ. The second column shows the P-loss convergence rate in Theorem 2 with an optimal choice of k.

Type of γ	P-loss convergence rate	Minimax lower bound
$\gamma(k) = 0$ for $k > k_0$	$\left(\frac{\log(n\vee p)}{n}\right)^{1/2}$	$\left(\frac{\log p}{n}\right)^{1/2}$
$\gamma(k) = Ce^{-\beta k}$, $\beta > 0$	$(\log n)^{3/4}\left(\frac{\log(n\vee p)}{n}\right)^{1/2}$	$\left(\frac{\log(n\vee p)}{n}\right)^{1/2}$ if $p \geq \log n$
$\gamma(k) = Ck^{-\alpha}$	$\left(\frac{\log p}{n}\right)^{(4\alpha-3)/8\alpha} + n^{-(4\alpha-3)/(8\alpha+4)}$ if $p \geq n^{1/(2\alpha)}$, $\alpha > 1$	$\left(\frac{\log p}{n}\right)^{1/2} + n^{-(2\alpha-1)/4\alpha}$, $\alpha > \frac{1}{2}$

given constant $\epsilon_0 > 0$. *If* $k^{3/2}(k + \log(n \vee p)) = O(n)$, $k + \log p = o(n)$ *and* $1 \leq k \leq p - 1$, *and*

$$\sup_{\Omega_{0,n} \in \mathcal{U}(\epsilon_0,\gamma)} \mathbb{E}_{0n}\mathbb{E}^{\pi}\big(\|\Omega_n - \Omega_{0,n}\| \mid \mathbf{X}_n\big) \lesssim k^{3/4}\left[\left(\frac{k + \log(n \vee p)}{n}\right)^{1/2} + \gamma(k)\right],$$

where $\mathcal{U}(\epsilon_0, \gamma)$ *is defined in* (2.8), *and* $\sum_{m=1}^{\infty} \gamma(m) < \infty$.

Note that because we impose the k-BC prior on the precision matrix Ω_n, the posterior for Ω_n is also supported on k-banded positive-definite matrices.

In the proof, we use a divide-and-conquer strategy to deal with the P-loss convergence rate, where we decompose it into small terms, which are easier to handle; that is,

$$\mathbb{E}_{0n}\mathbb{E}^{\pi}\big(\|\Omega_n - \Omega_{0,n}\| \mid \mathbf{X}_n\big) \leq \mathbb{E}_{0n}\mathbb{E}^{\pi}\big(\|\Omega_n - \widehat{\Omega}_{nk}\| \mid \mathbf{X}_n\big) + \mathbb{E}_{0n}\|\widehat{\Omega}_{nk} - \Omega_{0,n}\|,$$

where $\widehat{\Omega}_{nk}$ is a frequentist estimator of $\Omega_{0,n}$ that is a k-banded positive-definite matrix. For the first term, we use concentration inequalities for the posteriors of the parameters around certain frequentist estimators. For the second term, techniques for the frequentist convergence rate can be adopted.

Table 1 shows the P-loss convergence rates and minimax lower bounds under the spectral norm for various types of γ in $\mathcal{U}(\epsilon_0, \gamma)$. In the second row, we assume k_0 is fixed. The second column shows the P-loss convergence rates with optimal choices of k that minimize the convergence rates in Theorem 2. Optimal values of k are k_0, $(2\beta)^{-1}\log n$, and $\min\{n^{1/(2\alpha+1)}, (n/\log p)^{1/(2\alpha)}\}$ for the second, third, and fourth rows, respectively. Table 1 shows that the P-loss convergence rate with optimal k coincides with, or is quite close to, the minimax lower bound for every setting. Note that the P-loss convergence rate in the fourth row is equal to

the rate of the minimax lower bound up to $\min\{n^{(3\alpha-1)/[\alpha(8\alpha+4)]}, (n/\log p)^{3/(8\alpha)}\}$.

3.2. P-loss convergence rate and Bayesian minimax lower bound under the matrix ℓ_∞-norm

In this subsection, we establish the upper and lower bounds for the Bayesian minimax rate under the matrix ℓ_∞-norm. The P-loss convergence rate obtained in Theorem 4 is slightly slower than the rate of the minimax lower bound given in Theorem 3. However, our convergence rate is the fastest for bandable precision matrices using existing Bayesian methods. The proofs of the theorems are given in the Supplementary Material.

Theorem 3. *Consider model* (2.1)*, and let* $p \leq \exp(cn)$*, for some constant* $c > 0$*. Assume* $\Omega_{0,n} \in \mathcal{U}(\epsilon_0, \gamma)$*, defined in* (2.8)*, for given* $\epsilon_0 > 0$ *and a decreasing function* γ*.*
(i) If there exists a constant k_0 *such that* $\gamma(k) = 0$*, for all* $k \geq k_0$*, we have*

$$\inf_{\widehat{\Omega}_n} \sup_{\Omega_{0,n} \in \mathcal{U}(\epsilon_0,\gamma)} \mathbb{E}_{0n} \|\widehat{\Omega}_n - \Omega_{0,n}\|_\infty \gtrsim \left(\frac{\log p}{n}\right)^{1/2}.$$

(ii) If $\gamma(k) = Ce^{-\beta k}$*, for some constants* $\beta > 0$ *and* $C > 0$*, then we have*

$$\inf_{\widehat{\Omega}_n} \sup_{\Omega_{0,n} \in \mathcal{U}(\epsilon_0,\gamma)} \mathbb{E}_{0n} \|\widehat{\Omega}_n - \Omega_{0,n}\|_\infty \gtrsim \min\left\{\left(\frac{\log p \, \log n}{n}\right)^{1/2}, \frac{p}{\sqrt{n}}\right\}.$$

(iii) If $\gamma(k) = Ck^{-\alpha}$*, for some constants* $\alpha > 0$ *and* $C > 0$*, then we have*

$$\inf_{\widehat{\Omega}_n} \sup_{\Omega_{0,n} \in \mathcal{U}(\epsilon_0,\gamma)} \mathbb{E}_{0n} \|\widehat{\Omega}_n - \Omega_{0,n}\|_\infty \gtrsim \min\left\{\left(\frac{\log p}{n}\right)^{\alpha/(2\alpha+1)} + n^{-\alpha/(2\alpha+2)}, \frac{p}{\sqrt{n}}\right\}.$$

Theorem 4. *Consider model* (2.1) *and the k-BC prior* (2.4) *for the precision matrix* $\Omega_n = (I_p - A_n)^T D_n^{-1}(I_p - A_n)$*, with* $M \geq 9\epsilon_0^{-1}$ *and* $\nu_0 = o(n)$*, for a given constant* $\epsilon_0 > 0$*. If* $k(k + \log(n \vee p)) = O(n)$*,* $k + \log p = o(n)$ *and* $1 \leq k \leq p-1$*, then*

$$\sup_{\Omega_{0,n} \in \mathcal{U}(\epsilon_0,\gamma)} \mathbb{E}_{0n} \mathbb{E}^\pi\big(\|\Omega_n - \Omega_{0,n}\|_\infty \mid \mathbf{X}_n\big) \lesssim k\left[\left(\frac{k + \log(n \vee p)}{n}\right)^{1/2} + \gamma(k)\right],$$

where $\mathcal{U}(\epsilon_0, \gamma)$ *is defined in* (2.8)*, and* $\sum_{m=1}^{\infty} \gamma(m) < \infty$*.*

Table 2 shows the P-loss convergence rates and minimax lower bounds under

Table 2. P-loss convergence rates and minimax lower bounds under the matrix ℓ_∞-norm for various γ. The second column shows the P-loss convergence rate in Theorem 4 with the optimal choice of k.

Type of γ	P-loss convergence rate	Minimax lower bound
$\gamma(k) = 0$ for $k > k_0$	$\left(\frac{\log(n\vee p)}{n}\right)^{1/2}$	$\left(\frac{\log p}{n}\right)^{1/2}$
$\gamma(k) = Ce^{-\beta k}$	$\log n \left(\frac{\log(n\vee p)}{n}\right)^{1/2}$	$\left(\frac{\log p \log n}{n}\right)^{1/2}$ if $p \geq (\log n \log p)^{1/2}$
$\gamma(k) = Ck^{-\alpha}$	$\left(\frac{\log p}{n}\right)^{(\alpha-1)/2\alpha} + n^{-(\alpha-1)/(2\alpha+1)}$ if $p \geq n^{1/(2\alpha+2)}, \alpha > 1$	$\left(\frac{\log p}{n}\right)^{\alpha/(2\alpha+1)} + n^{-\alpha/(2\alpha+2)}, \alpha > 0$

the matrix ℓ_∞-norm for various γ in $\mathcal{U}(\epsilon_0, \gamma)$. As in Table 1, we assume k_0 is fixed, and present the P-loss convergence rates with optimal choices of k that minimize the convergence rates in Theorem 4. The optimal values of k are the same as those in Section 3.1. From Table 2, we can see that the P-loss convergence rate with optimal k coincides with, or is quite close to, a minimax lower bound for every setting.

Remark 5. The P-loss convergence rate in Theorem 4 is sharper than the posterior convergence rate of Banerjee and Ghosal (2014). If we consider an exponentially decreasing or exact banding $\gamma(k)$, then the parameter spaces of the two works are equivalent, by Proposition 1. In that case, the convergence rate obtained in Theorem 4 is equal to or faster than that of Banerjee and Ghosal (2014). When $\gamma(k) = Ck^{-\alpha}$, we have $\mathcal{U}(\epsilon_0, \gamma) \subseteq \mathcal{U}^*(\epsilon_0, \gamma')$, where $\gamma'(k) = C'k^{-(\alpha-1)}$, for some constant $C' > 0$, by Proposition 1. Thus, the rate obtained in Theorem 4 can be directly compared with that in Banerjee and Ghosal (2014) under the parameter class $\mathcal{U}(\epsilon_0, \gamma) \cap \mathcal{U}^*(\epsilon_0, \gamma')$. With the optimal choice of k for each result, the former is

$$\left(\frac{\log p}{n}\right)^{(\alpha-1)/2\alpha} + n^{-(\alpha-1)/(2\alpha+1)},$$

and the latter is $(\log p/n)^{(2\alpha-5)/(4\alpha)}$. The rate obtained in Theorem 4 is faster than that in Banerjee and Ghosal (2014) by factors $n^{(4\alpha+5)/[4\alpha(2\alpha+1)]}$ and $(n/\log p)^{3/(4\alpha)}$ when $n^{1/(2\alpha+2)} \leq p \leq \exp(n^{1/(2\alpha+1)})$ and $p \geq \exp(n^{1/(2\alpha+1)})$, respectively.

3.3. Frequentist convergence rates and posterior convergence rates

In this subsection, we obtain the frequentist convergence rate and the traditional posterior convergence rate of the k-BC prior defined in (2.4). For the

frequentist convergence rate, we propose the following plug-in estimator:

$$\widehat{\Omega}_{nk}^{LL} = (I_p - \mathbb{E}^{\pi}(A_n|\mathbf{X}_n))^T \mathbb{E}^{\tilde{\pi}}(D_n^{-1}|\mathbf{X}_n)(I_p - \mathbb{E}^{\pi}(A_n|\mathbf{X}_n)), \tag{3.2}$$

where $\mathbb{E}^{\tilde{\pi}}(\cdot \mid \mathbf{X}_n)$ are the posterior means from the nontruncated posteriors,

$$\tilde{\pi}(d_j \mid \mathbf{X}_n) = IG\left(d_j \mid \frac{n_j}{2}, \frac{n}{2}\widehat{d}_{jk}\right), \quad j = 1, \ldots, p.$$

The plug-in estimator $\widehat{\Omega}_{nk}^{LL}$ is more convenient than the posterior mean $\mathbb{E}^{\pi}(\Omega_n \mid \mathbf{X}_n)$ in practice, owing to its simple form. Note that $\mathbb{E}^{\pi}(a_j^{(k)} \mid d_j, \mathbf{X}_n) = \widehat{a}_j^{(k)}$ and $\mathbb{E}^{\tilde{\pi}}(d_j^{-1} \mid \mathbf{X}_n) = n_j \widehat{d}_{jk}^{-1}/n$. As a justification for the using the nontruncated posterior mean, in Corollary 1, we show that $\widehat{\Omega}_{nk}^{LL}$ achieves the same rate as the P-loss convergence rate. The proof of Corollary 1 is given in the Supplementary Material.

From Proposition A.1 of Lee and Lee (2018), a P-loss convergence rate is a posterior convergence rate. Thus, Corollary 2 follows from Proposition A.1 of Lee and Lee (2018), which means the rates obtained in Theorem 2 and Theorem 4 in this paper are also posterior convergence rates.

Corollary 1. *Consider model* (2.1) *and* $\mathcal{U}(\epsilon_0, \gamma)$ *defined in* (2.8)*, and assume* $k + \log p = o(n)$*,* $\sum_{m=1}^{\infty} \gamma(m) < \infty$*,* $\nu_0 = o(n)$*, and* $1 \le k \le p$*. If* $k^{3/2}(k + \log(n \vee p)) = O(n)$*, then*

$$\sup_{\Omega_{0,n} \in \mathcal{U}(\epsilon_0, \gamma)} \mathbb{E}_{0n} \|\widehat{\Omega}_{nk}^{LL} - \Omega_{0,n}\| \lesssim k^{3/4} \left[\left(\frac{k + \log(n \vee p)}{n} \right)^{1/2} + \gamma(k) \right].$$

If $k(k + \log(n \vee p)) = O(n)$*, then*

$$\sup_{\Omega_{0,n} \in \mathcal{U}(\epsilon_0, \gamma)} \mathbb{E}_{0n} \|\widehat{\Omega}_{nk}^{LL} - \Omega_{0,n}\|_{\infty} \lesssim k \left[\left(\frac{k + \log(n \vee p)}{n} \right)^{1/2} + \gamma(k) \right].$$

Corollary 2. *Consider model* (2.1)*,* $\mathcal{U}(\epsilon_0, \gamma)$ *defined in* (2.8)*, and the k-BC prior* (2.4)*, with* $M \ge 9\epsilon_0^{-1}$ *and* $\nu_0 = o(n)$*. Assume* $k + \log p = o(n)$*,* $\sum_{m=1}^{\infty} \gamma(m) < \infty$*, and* $1 \le k \le p$*. If* $k^{3/2}(k + \log(n \vee p)) = O(n)$ *and* $\epsilon_n = k^{3/4}[(k + \log(n \vee p))/n)^{1/2} + \gamma(k)]$*, then for any* $M_n \to \infty$ *as* $n \to \infty$*,*

$$\sup_{\Omega_{0,n} \in \mathcal{U}(\epsilon_0, \gamma)} \mathbb{E}_{0n} \Big[\pi\big(\|\Omega_n - \Omega_{0,n}\| \ge M_n \epsilon_n \mid \mathbf{X}_n \big) \Big] \longrightarrow 0.$$

If $k(k + \log(n \vee p)) = O(n)$ *and* $\epsilon_n^* = k\big[\left((k + \log(n \vee p))/n\right)^{1/2} + \gamma(k) \big]$*, then for*

any $M_n \to \infty$ *as* $n \to \infty$,

$$\sup_{\Omega_{0,n} \in \mathcal{U}(\epsilon_0, \gamma)} \mathbb{E}_{0n}\Big[\pi\big(\|\Omega_n - \Omega_{0,n}\|_\infty \geq M_n \epsilon_n^* \mid \mathbf{X}_n\big)\Big] \longrightarrow 0.$$

4. Choice of the Bandwidth k

We suggest using the posterior mode of k as a practical choice of the bandwidth k. Using Theorem 2 and Theorem 4, we can calculate the optimal *rate* of the bandwidth k by minimizing the P-loss convergence rate, when the rate of $\gamma(k)$ is given. However, this does not provide a proper choice of k in practice, because $\gamma(k)$ is unknown.

Let $\pi(k)$ be a prior distribution for the bandwidth k, and let $f(\mathbf{X}_n \mid A_n, D_n, k)$ be the likelihood function based on the observation $\mathbf{X}_n$. In Section 5, the prior distribution of k is set as $\pi(k) \propto \exp(-k^4)$, as in Banerjee and Ghosal (2014). The marginal posterior for k is easily derived as

$$\begin{aligned}
&\pi(k \mid \mathbf{X}_n) \\
&\propto \pi(k) \int\int f(\mathbf{X}_n \mid A_n, D_n, k)\pi(A_n, D_n \mid k) dA_n dD_n \\
&\propto \pi(k) \prod_{j=2}^{p} \det\left(X^T_{\cdot,(j-k):(j-1)} X_{\cdot,(j-k):(j-1)}/(2\pi)\right)^{-1/2} \Gamma\left(\frac{n_j}{2}\right)\left(\frac{n}{2}\widehat{d}_{jk}\right)^{-n_j/2} \\
&\quad \times \prod_{j=1}^{p} F_{IG}\left(M \;\Big|\; \frac{n_j}{2}, \frac{n\widehat{d}_{jk}}{2}\right), \qquad (4.1)
\end{aligned}$$

by routine calculations, where $\pi(A_n, D_n \mid k)$ denotes the k-BC prior (2.4), $\det(\cdot)$ is the determinant function, $\Gamma(\cdot)$ is the gamma function, and $F_{IG}(M \mid a, b)$ is a cumulative distribution function of $IG(a, b)$. Because the marginal posterior (4.1) has a simple analytic form, the posterior mode, say $\hat{k}$, can be easily obtained. The performance of $\hat{k}$ is described through comparisons with other approaches in the next section.

Note that the Cholesky-based Bayes estimator $\widehat{\Omega}^{LL}_{nk}$ is similar to the banded estimator (Bickel and Levina (2008b)), $\widehat{\Omega}^{BL}_{nk}$. The major difference between the two estimators is the choice of the bandwidth parameter k. Bickel and Levina (2008b) proposed a resampling scheme to estimate the oracle k that minimizes

$$R(k) = \mathbb{E}_{0n}\|\widehat{\Omega}^{BL}_{nk} - \Omega_{0,n}\|_1. \qquad (4.2)$$

To approximate the risk in (4.2), they randomly divide n observations into two groups, with sizes $n_1 = n/3$ and $n_2 = n - n_1$, respectively. They calculate the banded estimator based on the first group, say $\widehat{\Omega}^{BL}_{1,nk}$, and used it for $\widehat{\Omega}^{BL}_{nk}$ in (4.2). The second group is used to approximate $\Omega_{0,n}$ in (4.2), but Bickel and Levina (2008b) do not indicate which estimator to use. Because the sample precision matrix is computationally unstable for large p, we use the banded estimator with $K = 20$ based on the second group, say $\widehat{\Omega}^{BL,(t)}_{2,nK}$, in the simulation study. In the same way, a tth random split gives $\widehat{\Omega}^{BL,(t)}_{1,nk}$ and $\widehat{\Omega}^{BL,(t)}_{2,nk}$, for $t = 1, \ldots, T$. Finally, the risk (4.2) is approximated by

$$\widehat{R}(k) = \frac{1}{T}\sum_{t=1}^{T} \|\widehat{\Omega}^{BL,(t)}_{1,nk} - \widehat{\Omega}^{BL,(t)}_{2,nK}\|_1, \tag{4.3}$$

and the bandwidth k is selected as $\hat{k}^{BL} = \text{argmin}_{0\le k\le K}\, \widehat{R}(k)$, where $K = 20$. Note that $\hat{k}^{BL}$ does not need to minimize the ℓ_1-norm risk in practice, because its consistency is not guaranteed.

5. Simulation Study

We investigate the performance of the proposed Bayes estimator $\widehat{\Omega}^{LL}_{nk}$, defined in (3.2), and the posterior mode $\hat{k}$. Then, we compare the performance of the Bayes estimator based on the G-Wishart prior $\widehat{\Omega}^{BG}_{nk}$ (Banerjee and Ghosal (2014)) and the banded estimator $\widehat{\Omega}^{BL}_{nk}$ (Bickel and Levina (2008b)) in various scenarios. For the proposed estimator $\widehat{\Omega}^{LL}_{nk}$, we use $\nu_0 = 2$ throughout this section.

Banerjee and Ghosal (2014) proposed two Bayes estimators, corresponding to the Stein loss and the squared-error loss, respectively. We examine the performance of two Bayes estimators, say $\widehat{\Omega}^{BG1}_{nk}$ and $\widehat{\Omega}^{BG2}_{nk}$, with $\delta = 3$. For these estimators, the bandwidth k is chosen using the posterior mode in Banerjee and Ghosal (2014), $\hat{k}^{BG}$.

For the banded Cholesky-based estimator proposed by (Bickel and Levina (2008b)), we tried two different bandwidth estimators, $\hat{k}^{BL}$ and $\hat{k}$, in order to compare their relative performance.

The spectral norm, matrix ℓ_∞-norm, and Frobenius norm are used as loss functions. The sample sizes are set to $n = 100, 200$, and 500, and the dimensions are set to $p = 100, 200$, and 500. For each setting, the values of the loss function,

$$\|\widehat{\Omega}^{(s)}_{nk} - \Omega_{0,n}\|, \quad s = 1, \ldots, 100, \tag{5.1}$$

are calculated based on 100 simulated data items, for each method $\widehat{\Omega}_{nk}$ and loss function $\|\cdot\|$, where $\Omega_{0,n}$ denotes the true precision matrix. The mean and standard deviation of (5.1) are used as summary statistics. We consider the following true precision matrices.

Example 1. ($AR(1)$ process) Assume the true covariance matrix $\Sigma_{0,n} = (\sigma_{0,ij})$ is given by

$$\sigma_{0,ij} = \rho^{|i-j|},\ 1 \le i, j \le p,$$

with $\rho = 0.3$. Then, the true precision matrix is a banded matrix with an $AR(1)$ process structure.

Example 2. ($AR(4)$ process) Assume the true precision matrix $\Omega_{0,n} = (\omega_{0,ij})$ is given by

$$\begin{aligned}\omega_{0,ij} = {} & I(|i-j| = 0) + 0.4 \cdot I(|i-j| = 1) + 0.2 \cdot I(|i-j| = 2) \\ & + 0.2 \cdot I(|i-j| = 3) + 0.1 \cdot I(|i-j| = 4).\end{aligned}$$

Thus, the true precision matrix is a banded matrix with an $AR(4)$ process structure. Furthermore, it is always positive definite because of the diagonally dominant property.

Example 3. (Long-range dependence) In the last example, the true precision matrix is not a bandable matrix in $\mathcal{U}(\epsilon_0, \gamma)$. Consider a fractional Gaussian noise model, where the true covariance matrix $\Sigma_{0,n} = (\sigma_{0,ij})$ is given by

$$\sigma_{0,ij} = \frac{1}{2}\left(||i-j|+1|^{2H} - 2|i-j|^{2H} + ||i-j|-1|^{2H}\right),\ 1 \le i, j \le p,$$

with $H \in [0.5, 1]$. The Hurst parameter H indicates the dependency of the process. Here, $H = 0.5$ implies white noise, and H near 1 denotes long-range dependence. We chose $H = 0.7$. In this case, the true precision matrix does not belong to the bandable class.

Tables 3–5 show the simulation results for the above three examples, and Figure 1 shows the performance of each estimator when the true precision matrix is an $AR(4)$ process and $(n, p) = (500, 500)$. We omit the estimator $\widehat{\Omega}_{nk}^{BG2}$, because its performance is similar to that of $\widehat{\Omega}_{nk}^{BG1}$ in all scenarios, where BG in Tables 3–5 and Figure 1 represents $\widehat{\Omega}_{nk}^{BG1}$.

We also report the summary statistics for the estimated bandwidths, $\hat{k}$, $\hat{k}^{BG}$, and $\hat{k}^{BL}$, for the $AR(1)$ and $AR(4)$ models in Table 6 and Table 7, respectively.

Table 3. Simulation results for the $AR(1)$ model. For each n and p, the mean and standard deviation (in parentheses) of three loss functions (the spectral norm, matrix ℓ_∞-norm, and Frobenius norm) are calculated. Columns $BL1$ and $BL2$ show the results for banded estimators with bandwidths $\hat{k}^{BL}$ and $\hat{k}$, respectively.

			LL	BG	BL1	BL2
$n = 100$	$p = 100$	$\|\cdot\|$	0.720 (0.139)	0.759 (0.141)	1.217 (0.391)	0.786 (0.146)
		$\|\cdot\|_\infty$	0.913 (0.176)	0.957 (0.177)	1.905 (0.813)	0.989 (0.184)
		$\|\cdot\|_F$	2.382 (0.171)	2.447 (0.187)	3.837 (1.067)	2.503 (0.196)
	$p = 200$	$\|\cdot\|$	0.802 (0.140)	0.842 (0.140)	1.294 (0.353)	0.873 (0.145)
		$\|\cdot\|_\infty$	1.025 (0.180)	1.071 (0.180)	2.044 (0.716)	1.108 (0.186)
		$\|\cdot\|_F$	3.395 (0.165)	3.487 (0.179)	5.471 (0.322)	3.567 (0.188)
	$p = 500$	$\|\cdot\|$	0.910 (0.147)	0.951 (0.146)	1.504 (0.417)	0.985 (0.152)
		$\|\cdot\|_\infty$	1.151 (0.181)	1.196 (0.181)	2.412 (0.928)	1.239 (0.188)
		$\|\cdot\|_F$	5.377 (0.172)	5.521 (0.186)	9.070 (2.353)	5.647 (0.353)
$n = 200$	$p = 100$	$\|\cdot\|$	0.482 (0.090)	0.498 (0.094)	0.585 (0.165)	0.507 (0.096)
		$\|\cdot\|_\infty$	0.619 (0.117)	0.636 (0.121)	0.815 (0.315)	0.646 (0.124)
		$\|\cdot\|_F$	1.673 (0.110)	1.696 (0.116)	1.991 (0.442)	1.714 (0.119)
	$p = 200$	$\|\cdot\|$	0.537 (0.098)	0.556 (0.100)	0.644 (0.154)	0.567 (0.102)
		$\|\cdot\|_\infty$	0.685 (0.127)	0.706 (0.130)	0.896 (0.277)	0.718 (0.133)
		$\|\cdot\|_F$	2.374 (0.113)	2.406 (0.121)	2.851 (0.544)	2.432 (0.124)
	$p = 500$	$\|\cdot\|$	0.594 (0.124)	0.615 (0.080)	0.747 (0.156)	0.626 (0.082)
		$\|\cdot\|_\infty$	0.755 (0.108)	0.777 (0.109)	1.054 (0.326)	0.792 (0.111)
		$\|\cdot\|_F$	3.762 (0.104)	3.813 (0.111)	4.692 (0.866)	3.855 (0.114)
$n = 500$	$p = 100$	$\|\cdot\|$	0.287 (0.045)	0.292 (0.046)	0.309 (0.055)	0.295 (0.047)
		$\|\cdot\|_\infty$	0.368 (0.060)	0.373 (0.063)	0.404 (0.084)	0.376 (0.064)
		$\|\cdot\|_F$	1.053 (0.065)	1.060 (0.067)	1.110 (0.110)	1.064 (0.067)
	$p = 200$	$\|\cdot\|$	0.314 (0.045)	0.321 (0.046)	0.340 (0.065)	0.324 (0.047)
		$\|\cdot\|_\infty$	0.405 (0.059)	0.412 (0.061)	0.445 (0.101)	0.415 (0.062)
		$\|\cdot\|_F$	1.489 (0.073)	1.497 (0.074)	1.565 (0.172)	1.503 (0.074)
	$p = 500$	$\|\cdot\|$	0.340 (0.042)	0.347 (0.042)	0.359 (0.051)	0.350 (0.043)
		$\|\cdot\|_\infty$	0.436 (0.053)	0.444 (0.053)	0.465 (0.078)	0.448 (0.053)
		$\|\cdot\|_F$	2.352 (0.069)	2.365 (0.070)	2.433 (0.188)	2.375 (0.071)

Two remarks related to the simulation results are in order. First, it seems that the proposed Bayes estimator $\hat{k}$ is practically comparable with or better than the method of Banerjee and Ghosal (2014). Because our theoretical results and those of Banerjee and Ghosal (2014) are based on an optimal choice of k, which depends on unknown parameters, the practical performance when using the posterior mode $\hat{k}$ is of independent interest. Based on our simulation, the performance of $\widehat{\Omega}_{nk}^{LL}$ is better, in general, than that of $\widehat{\Omega}_{nk}^{BG1}$. Furthermore, based on Tables 6 and 7, $\hat{k}^{BG}$ tended to underestimate the true bandwidth, and $\hat{k}$

Table 4. Simulation results for the $AR(4)$ model. For each n and p, the mean and standard deviation (in parentheses) of three loss functions (the spectral norm, matrix ℓ_∞-norm, and Frobenius norm) are calculated. Columns $BL1$ and $BL2$ show the results for banded estimators with bandwidths $\hat{k}^{BL}$ and $\hat{k}$, respectively.

			LL	BG	BL1	BL2
$n = 100$	$p = 100$	$\|\cdot\|$	1.510 (0.040)	1.475 (0.041)	1.481 (0.340)	1.473 (0.041)
		$\|\cdot\|_\infty$	1.854 (0.058)	1.826 (0.061)	2.446 (0.607)	1.827 (0.063)
		$\|\cdot\|_F$	5.130 (0.070)	5.050 (0.065)	4.189 (0.620)	5.046 (0.065)
	$p = 200$	$\|\cdot\|$	1.541 (0.034)	1.506 (0.035)	1.668 (0.395)	1.504 (0.035)
		$\|\cdot\|_\infty$	1.899 (0.065)	1.873 (0.069)	2.873 (0.678)	1.874 (0.071)
		$\|\cdot\|_F$	7.312 (0.072)	7.196 (0.068)	6.015 (0.840)	7.191 (0.068)
	$p = 500$	$\|\cdot\|$	1.564 (0.029)	1.530 (0.030)	1.884 (0.368)	1.528 (0.030)
		$\|\cdot\|_\infty$	1.938 (0.052)	1.913 (0.056)	3.061 (0.654)	1.915 (0.057)
		$\|\cdot\|_F$	11.610 (0.076)	11.426 (0.072)	9.620 (1.288)	11.417 (0.072)
$n = 200$	$p = 100$	$\|\cdot\|$	1.313 (0.314)	1.461 (0.027)	0.843 (0.168)	1.288 (0.325)
		$\|\cdot\|_\infty$	1.616 (0.273)	1.734 (0.050)	1.366 (0.284)	1.596 (0.280)
		$\|\cdot\|_F$	4.477 (0.980)	4.949 (0.049)	2.513 (0.260)	4.431 (0.972)
	$p = 200$	$\|\cdot\|$	0.972 (0.289)	1.482 (0.027)	0.482 (0.171)	0.934 (0.299)
		$\|\cdot\|_\infty$	1.343 (0.245)	1.759 (0.043)	1.457 (0.280)	1.319 (0.249)
		$\|\cdot\|_F$	4.574 (1.357)	7.047 (0.054)	3.528 (0.336)	4.502 (1.356)
	$p = 500$	$\|\cdot\|$	0.871 (0.052)	1.499 (0.023)	1.015 (0.169)	0.840 (0.059)
		$\|\cdot\|_\infty$	1.300 (0.097)	1.800 (0.041)	1.643 (0.303)	1.291 (0.117)
		$\|\cdot\|_F$	6.083 (0.345)	11.200 (0.058)	5.686 (0.554)	6.001 (0.226)
$n = 500$	$p = 100$	$\|\cdot\|$	0.501 (0.139)	1.052 (0.395)	0.450 (0.084)	0.513 (0.122)
		$\|\cdot\|_\infty$	0.767 (0.151)	1.281 (0.355)	0.733 (0.144)	0.784 (0.137)
		$\|\cdot\|_F$	1.663 (0.444)	3.573 (1.324)	1.439 (0.118)	1.676 (0.411)
	$p = 200$	$\|\cdot\|$	0.447 (0.063)	0.807 (0.255)	0.481 (0.067)	0.473 (0.067)
		$\|\cdot\|_\infty$	0.722 (0.080)	1.083 (0.229)	0.768 (0.094)	0.754 (0.088)
		$\|\cdot\|_F$	1.939 (0.068)	3.764 (1.245)	2.010 (0.107)	1.985 (0.075)
	$p = 500$	$\|\cdot\|$	0.493 (0.081)	0.737 (0.044)	0.530 (0.084)	0.522 (0.085)
		$\|\cdot\|_\infty$	0.784 (0.111)	1.036 (0.056)	0.835 (0.117)	0.820 (0.116)
		$\|\cdot\|_F$	3.069 (0.067)	5.151 (0.456)	3.189 (0.141)	3.139 (0.076)

outperforms $\hat{k}^{BG}$. Second, our selection scheme for k is comparable with that of Bickel and Levina (2008b); however, a comparison of the two is not straightforward. The $BL1$ and $BL2$ columns in Tables 3–5 show the results for the banded estimators of Bickel and Levina (2008b), with k chosen using $\hat{k}^{BL}$ and $\hat{k}$, respectively. Based on Tables 3 and 5, $BL2$ outperforms $BL1$; however, in the second scenario (Table 4), it is difficult to determine which of the two performs best. From Tables 6 and 7, we can see that $\hat{k}^{BL}$ seems to overestimate the true bandwidth, whereas $\hat{k}$ underestimates the true bandwidth. However, when the sample size n is large ($n = 500$), $\hat{k}$ estimates the true bandwidth quite well,

Table 5. Simulation results for the fractional Gaussian noise model. For each n and p, the mean and standard deviation (in parentheses) of three loss functions (the spectral norm, matrix ℓ_∞-norm, and Frobenius norm) are calculated. Columns $BL1$ and $BL2$ show the resultSs for banded estimators with bandwidths $\hat{k}^{BL}$ and $\hat{k}$, respectively.

			LL	BG	BL1	BL2
$n = 100$	$p = 100$	$\|\cdot\|$	0.837 (0.149)	0.880 (0.149)	1.232 (0.357)	0.911 (0.154)
		$\|\cdot\|_\infty$	1.588 (0.194)	1.636 (0.193)	2.284 (0.698)	1.676 (0.200)
		$\|\cdot\|_F$	2.879 (0.194)	2.955 (0.211)	3.980 (0.908)	3.030 (0.222)
	$p = 200$	$\|\cdot\|$	0.931 (0.147)	0.973 (0.146)	1.326 (0.335)	1.008 (0.152)
		$\|\cdot\|_\infty$	1.752 (0.188)	1.800 (0.187)	2.495 (0.654)	1.843 (0.194)
		$\|\cdot\|_F$	4.100 (0.178)	4.208 (0.194)	5.806 (1.236)	4.315 (0.204)
	$p = 500$	$\|\cdot\|$	1.043 (0.156)	1.085 (0.155)	1.493 (0.348)	1.124 (0.161)
		$\|\cdot\|_\infty$	1.929 (0.196)	1.976 (0.194)	2.800 (0.740)	2.025 (0.202)
		$\|\cdot\|_F$	6.478 (0.185)	6.648 (0.202)	9.321 (0.945)	6.817 (0.212)
$n = 200$	$p = 100$	$\|\cdot\|$	0.601 (0.096)	0.622 (0.096)	0.659 (0.126)	0.634 (0.098)
		$\|\cdot\|_\infty$	1.269 (0.137)	1.293 (0.138)	1.360 (0.217)	1.307 (0.140)
		$\|\cdot\|_F$	2.287 (0.123)	2.318 (0.131)	2.435 (0.278)	2.347 (0.134)
	$p = 200$	$\|\cdot\|$	0.665 (0.111)	0.687 (0.111)	0.719 (0.116)	0.699 (0.113)
		$\|\cdot\|_\infty$	1.400 (0.148)	1.424 (0.148)	1.484 (0.175)	1.440 (0.150)
		$\|\cdot\|_F$	3.244 (0.131)	3.289 (0.139)	3.465 (0.289)	3.330 (0.143)
	$p = 500$	$\|\cdot\|$	0.733 (0.092)	0.755 (0.092)	0.801 (0.104)	0.769 (0.094)
		$\|\cdot\|_\infty$	1.526 (0.127)	1.551 (0.127)	1.646 (0.181)	1.568 (0.129)
		$\|\cdot\|_F$	5.141 (0.117)	5.212 (0.124)	5.547 (0.464)	5.277 (0.128)
$n = 500$	$p = 100$	$\|\cdot\|$	0.406 (0.047)	0.434 (0.043)	0.436 (0.046)	0.420 (0.049)
		$\|\cdot\|_\infty$	1.004 (0.076)	1.004 (0.073)	1.046 (0.074)	1.020 (0.077)
		$\|\cdot\|_F$	1.722 (0.064)	1.875 (0.066)	1.869 (0.084)	1.742 (0.066)
	$p = 200$	$\|\cdot\|$	0.433 (0.045)	0.467 (0.044)	0.468 (0.046)	0.448 (0.046)
		$\|\cdot\|_\infty$	1.086 (0.073)	1.128 (0.073)	1.130 (0.074)	1.105 (0.074)
		$\|\cdot\|_F$	2.433 (0.063)	2.649 (0.065)	2.647 (0.085)	2.460 (0.065)
	$p = 500$	$\|\cdot\|$	0.461 (0.043)	0.495 (0.039)	0.498 (0.040)	0.476 (0.044)
		$\|\cdot\|_\infty$	1.164 (0.073)	1.203 (0.062)	1.209 (0.061)	1.183 (0.074)
		$\|\cdot\|_F$	3.856 (0.062)	4.189 (0.071)	4.202 (0.085)	3.900 (0.064)

whereas $\hat{k}^{BL}$ still overestimates, with a relatively larger variance than that of $\hat{k}$.

6. Discussion

We have proposed a k-BC prior (2.4) for bandable precision matrices based on an MCD. The P-loss convergence rates for precision matrices under the spectral norm and matrix ℓ_∞-norm were established. Although the P-loss convergence rates are slightly slower than that of the Bayesian minimax lower bounds, the proposed approach attains a faster posterior convergence rate than those of other Bayesian methods. Simulation results indicate that its practical performance is

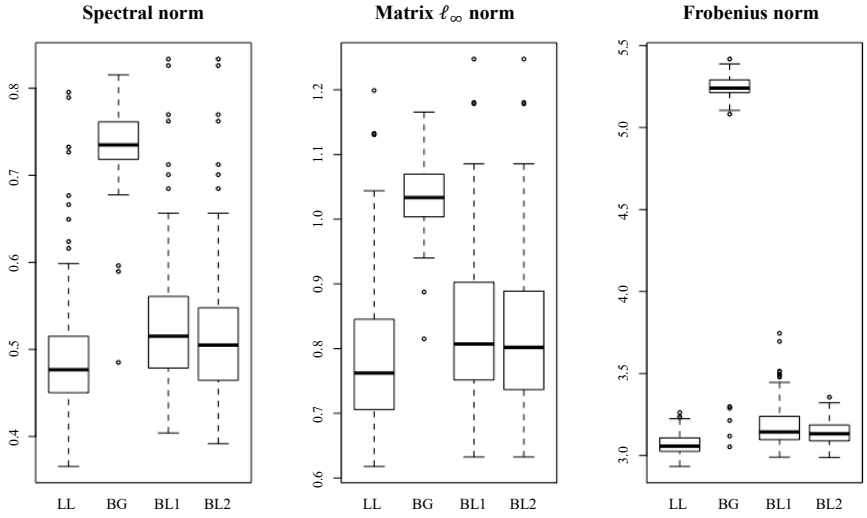

Figure 1. The average errors for an $AR(4)$ process structure precision matrix under the spectral norm, matrix ℓ_∞-norm, and Frobenius norm. The sample size n and the dimensionality p are both set as 500.

Table 6. The mean and standard deviation (in parentheses) of the estimated bandwidths for the $AR(1)$ model ($k_0 = 1$) in Example 1.

		$\hat{k}$	$\hat{k}^{BG}$	$\hat{k}^{BL}$
$n = 100$	$p = 100$	1.000 (0.000)	1.000 (0.000)	3.060 (1.638)
	$p = 200$	1.000 (0.000)	1.000 (0.000)	3.080 (1.454)
	$p = 500$	1.000 (0.000)	1.000 (0.000)	3.360 (1.667)
$n = 200$	$p = 100$	1.000 (0.000)	1.000 (0.000)	1.650 (0.999)
	$p = 200$	1.000 (0.000)	1.000 (0.000)	1.700 (0.882)
	$p = 500$	1.000 (0.000)	1.000 (0.000)	1.890 (0.952)
$n = 500$	$p = 100$	1.000 (0.000)	1.000 (0.000)	1.170 (0.377)
	$p = 200$	1.000 (0.000)	1.000 (0.000)	1.170 (0.403)
	$p = 500$	1.000 (0.000)	1.000 (0.000)	1.100 (0.333)

comparable to or better than that of competitive approaches.

Several extensions to this work are possible, related to the bandwidth k. First, our theoretical results depend on the unknown parameter of $\gamma(k)$. To choose an optimal k, we need to know the rate of $\gamma(k)$. Thus, developing an adaptive procedure that simultaneously attains a reasonable convergence rate, regardless of $\gamma(k)$, is one possible extension. Second, the theoretical property of

Table 7. The mean and standard deviation (in parenthesis) of estimated values of bandwidth for $AR(4)$ model ($k_0 = 4$) in Example 2.

		$\hat{k}$	$\hat{k}^{BG}$	$\hat{k}^{BL}$
$n = 100$	$p = 100$	1.000 (0.000)	1.000 (0.000)	5.130 (1.060)
	$p = 200$	1.000 (0.000)	1.000 (0.000)	5.170 (1.074)
	$p = 500$	1.000 (0.000)	1.000 (0.000)	5.240 (1.074)
$n = 200$	$p = 100$	1.440 (0.833)	1.000 (0.000)	4.570 (0.686)
	$p = 200$	2.560 (0.833)	1.000 (0.676)	4.530 (0.717)
	$p = 500$	3.090 (0.288)	1.000 (0.000)	4.660 (0.794)
$n = 500$	$p = 100$	3.700 (0.461)	2.000 (1.005)	4.240 (0.571)
	$p = 200$	4.000 (0.000)	2.740 (0.676)	4.110 (0.314)
	$p = 500$	4.000 (0.000)	3.050 (0.219)	4.140 (0.377)

the posterior mode $\hat{k}$ is unexplored. Here, a theoretical result similar to Theorem 4 in Bickel and Levina (2008a) can be investigated.

Supplementary Material

The online Supplementary Material provides proofs for the main and other auxiliary results.

Acknowledgements

We would like to thank an associate editor and two referees for their valuable comments which have led to improvements of an earlier version of the paper. Kyoungjae Lee was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2019R1F1A1059483) and INHA UNIVERSITY Research Grant. Jaeyong Lee was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2018R1A2A3074973).

References

Banerjee, S. and Ghosal, S. (2014). Posterior convergence rates for estimating large precision matrices using graphical models. *Electronic Journal of Statistics* **8**, 2111–2137.

Banerjee, S. and Ghosal, S. (2015). Bayesian structure learning in graphical models. *Journal of Multivariate Analysis* **136**, 147–162.

Bickel, P. J. and Levina, E. (2008a). Covariance regularization by thresholding. *The Annals of Statistics* **36**, 2577–2604.

Bickel, P. J. and Levina, E. (2008b). Regularized estimation of large covariance matrices. *The Annals of Statistics* **36**, 199–227.

Cai, T. T., Liang, T. and Zhou, H. H. (2015). Law of log determinant of sample covariance matrix and optimal estimation of differential entropy for high-dimensional gaussian distributions. *Journal of Multivariate Analysis* **137**, 161–172.

Cai, T. T., Liu, W. and Zhou, H. H. (2016). Estimating sparse precision matrix: Optimal rates of convergence and adaptive estimation. *The Annals of Statistics* **44**, 455–488.

Cai, T. T., Ma, Z. and Wu, Y. (2015). Optimal estimation and rank detection for sparse spiked covariance matrices. *Probability Theory and Related Fields* **161**, 781–815.

Cai, T. T., Ren, Z. and Zhou, H. H. (2016). Estimating structured high-dimensional covariance and precision matrices: Optimal rates and adaptive estimation. *Electronic Journal of Statistics* **10**, 1–59.

Cai, T. T. and Yuan, M. (2016). Minimax and adaptive estimation of covariance operator for random variables observed on a lattice graph. *Journal of the American Statistical Association* **111**, 253–265.

Cai, T. T., Zhang, C.-H. and Zhou, H. H. (2010). Optimal rates of convergence for covariance matrix estimation. *The Annals of Statistics* **38**, 2118–2144.

Cai, T. T. and Zhou, H. H. (2012a). Minimax estimation of large covariance matrices under ℓ_1-norm. *Statistica Sinica* **22**, 1319–1349.

Cai, T. T. and Zhou, H. H. (2012b). Optimal rates of convergence for sparse covariance matrix estimation. *The Annals of Statistics* **40**, 2389–2420.

Cao, X., Khare, K. and Ghosh, M. (2016). Posterior graph selection and estimation consistency for high-dimensional Bayesian dag models. *arXiv:1611.01205* .

Fan, J., Fan, Y. and Lv, J. (2008). High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics* **147**, 186–197.

Fan, J., Rigollet, P. and Wang, W. (2015). Estimation of functionals of sparse covariance matrices. *The Annals of Statistics* **43**, 2706.

Gao, C. and Zhou, H. H. (2015). Rate-optimal posterior contraction for sparse pca. *The Annals of Statistics* **43**, 785–818.

Gao, C. and Zhou, H. H. (2016). Bernstein-von mises theorems for functionals of the covariance matrix. *Electronic Journal of Statistics* **10**, 1751–1806.

Ghosal, S., Ghosh, J. K. and van der Vaart, A. W. (2000). Convergence rates of posterior distributions. *The Annals of Statistics* **28**, 500–531.

Ghosal, S. and van der Vaart, A. (2007). Posterior convergence rates of dirichlet mixtures at smooth densities. *The Annals of Statistics* **35**, 697–723.

Hu, A. and Negahban, S. (2017). Minimax estimation of bandable precision matrices. In *Advances in Neural Information Processing Systems*, 4895–4903.

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**, 682–693.

Khare, K., Oh, S., Rahman, S. and Rajaratnam, B. (2016). A convex framework for high-dimensional sparse Cholesky based covariance estimation. *arXiv preprint arXiv: 1610.02436*.

Koller, D. and Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques*. MIT Press.

Lauritzen, S. L. (1996). *Graphical Models*. Oxford University Press Inc., New York.

Lee, K. and Lee, J. (2018). Optimal Bayesian minimax rates for unconstrained large covariance

matrices. *Bayesian Analysis* **13**, 1211–1229.

Liu, Y. and Ren, Z. (2017). Minimax estimation of large precision matrices with bandable Cholesky factor. *ArXiv e-prints*.

Pati, D., Bhattacharya, A., Pillai, N. S. and Dunson, D. (2014). Posterior contraction in sparse Bayesian factor models for massive covariance matrices. *The Annals of Statistics* **42**, 1102–1130.

Rütimann, P. and Bühlmann, P. (2009). High dimensional sparse covariance estimation via directed acyclic graphs. *Electronic Journal of Statistics* **3**, 1133–1160.

Scott, S. L., Blocker, A. W., Bonassi, F. V., Chipman, H. A., George, E. I., and McCulloch, R. E. (2016). Bayes and big data: The consensus monte carlo algorithm. *International Journal of Management Science and Engineering Management* **11**, 78–88.

Shojaie, A. and Michailidis, G. (2010). Penalized likelihood methods for estimation of sparse high-dimensional directed acyclic graphs. *Biometrika* **97**, 519–538.

Verzelen, N. (2010). Adaptive estimation of covariance matrices via cholesky decomposition. *Electronic Journal of Statistics* **4**, 1113–1150.

Xiang, R., Khare, K. and Ghosh, M. (2015). High dimensional posterior convergence rates for decomposable graphical models. *Electronic Journal of Statistics* **9**, 2828–2854.

Xue, L. and Zou, H. (2013). Minimax optimal estimation of general bandable covariance matrices. *Journal of Multivariate Analysis* **116**, 45–51.

Kyoungjae Lee

Department of Statistics, Inha University, Nam-gu, Incheon, South Korea.

E-mail: leekjstat@gmail.com

Jaeyong Lee

Department of Statistics, Seoul National University, Gwanak-gu, Seoul, South Korea.

E-mail: leejyc@gmail.com

(Received February 2018; accepted February 2019)