Post-Processed Posterior for Banded Covariances

Kwangmin Lee
Bayesian Lab

my1989@snu.ac.kr

May 20, 2021

Overview

1. Banded Covariance Inference
2. Post-Processed Posterior
3. Asymptotic Analysis
4. Numerical Studies

- Model :

$$X_1, \ldots, X_n \overset{i.i.d.}{\sim} N_p(0, \Sigma)$$

- Inverse-Wishart ($\pi(\Sigma) \propto |\Sigma|^{-\nu/2} \exp(-tr(\Sigma^{-1}\Lambda)/2)$) has been used for prior.
  -$IW(O_{p\times p}, p+1)$ is the multivariate Jeffreys prior distribution.
  -If $\Sigma \sim IW(I, 2p+1)$, then $\rho_{ij} \sim Unif(-1, 1)$ for all i, j.
  -Huang and Wand (2013) proposed a mixture of inverse-Wishart as a prior.
- Decision-theoretic analysis on prior distributions for Σ (Lee and Lee; 2018)
  -When $p < n$ and $||A_n||_2 \vee (\nu_n - 2p) = o(n)$, $IW(A_n, \nu_n)$ prior attains minimax optimal rate.
  -When $p \geq n$, δ_{I_p} prior attains minimax optimal rate.
- High-dimensional covariance problem
  $\Rightarrow$ Assume a low-dimensinal structure.

Banded Covariance Inference

-Banded covariance model

$$\{\Sigma = (\sigma_{ij}) : \sigma_{ij} = 0 \text{ if } |i-j| > k\}$$

-Banded sample covariance (Bickel and Levina; 2008) :

$$B_k(S_n) = (s_{n,ij} I(|i-j| \leq k)),\ S_n = (s_{n,ij}) \text{ is the sample covariance}$$

- Convergence rate is obtained for bandable covariance
- Application for stationary time series (Wu and Pourahmadi; 2009).
- Application for time-varying autoregressive moving average model (Wiesel and Globerson; 2012).

(Pros and Cons)

-Computing is fast.

-It lacks interval estimators.

-No minimax convergence rates for the banded covariance, and it is not minimax or nearly minimax in for the bandable covariance.

Banded Covariance Inference

The banded covariance model is a special case of covariance graphical model.

$\longrightarrow$ Methods for the covariance graphical model can be used for the banded covariance model.

- Frequentist methods

  -Dual maximum likelihood estimator (Kauermann; 1996)

  -Maximum likelihood estimator (ICF algorithm by Chaudhuri et al. (2007))

- Bayesian methods

  -G-inverse Wishart (Silva and Ghahramani; 2009)

$$\pi(\Sigma) \propto |\Sigma|^{-(\delta+2p)/2} \exp\{-tr(\Sigma^{-1}U)/2\} I(\Sigma \in M^{+}(\mathcal{G}))$$

  -Wishart for covariance graph (Khare and Rajaratnam; 2011)

$$\Sigma = LDL^T, \ \pi(L, D) \propto \exp(-(tr((LDL^T)^{-1}U) + \sum_{i=1}^{p} \alpha_i \log D_{ii})/2) I((L, D) \in \Theta_{\mathcal{G}})$$

(Pros and Cons)

-They all lack an asymptotic justification.

-Bayesian methods is computationally inefficient but provide interval estimators.

-Chaudhuri et al. (2007) provides interval estimation, but only justified in the fixed p case.

Banded Covariance Inference

There are no methods for banded covariances to satisfy the three conditions simultaneously.

- Computationally efficient.
- Providing interval estimators.
- Asymptotically optimal.

Overview

1. Banded Covariance Inference
2. Post-Processed Posterior
3. Asymptotic Analysis
4. Numerical Studies

- Model

$$X_1, \ldots, X_n \overset{i.i.d.}{\sim} N_p(0, \Sigma),$$

where $\Sigma \in \mathcal{B}_{p,k} = \{\Sigma = (\sigma_{ij}) \in \mathcal{C}_p : \sigma_{ij} = 0 \text{ if } |i-j| > k\}$, and $\mathcal{C}_p$ is the set of all positive definite $p \times p$ matrices.

- Algorithm

-Step 1. Initial posterior computing

Set initial prior $IW(B_0, \nu_0)$ and draw the initial posterior samples as

$$\Sigma^{(1)}, \ldots, \Sigma^{(N)} \overset{i.i.d.}{\sim} \pi^i(\Sigma \mid \mathbb{X}_n) = IW(B_0 + nS_n, \nu_0 + n).$$

-Step 2. Post-processing

Transform the initial posterior sample using

$$B_k^{(\epsilon_n)}(\Sigma^{(i)}) = \begin{cases} B_k(\Sigma^{(i)}) + (\epsilon_n - \lambda_{\min}(B_k(\Sigma^{(i)})))I_p, & \lambda_{\min}(B_k(\Sigma^{(i)})) < \epsilon_n \\ B_k(\Sigma^{(i)}), & \lambda_{\min}(B_k(\Sigma^{(i)})) \geq \epsilon_n \end{cases}.$$

- Distribution of banding Post-Processed Posterior (PPP)

$$B_k^{(\epsilon_n)}(\Sigma^{(1)}) \ldots, B_k^{(\epsilon_n)}(\Sigma^{(N)}) \overset{i.i.d.}{\sim} \pi^i(\cdot \mid \mathbb{X}_n) \circ (B_k^{(\epsilon)})^{-1} := \mathbb{P}^{pp}.$$

Similar Ideas of Transforming Initial Posterior Sample

The idea of transforming the initial posterior sample has been used.

- Notation

  $\Theta_0(\subset \Theta)$: Parameter space (Unconstrained parameter space)

  $\psi : \Theta \mapsto \Theta_0$: Post-processing function

- Dunson and Neelon (2003)

  $\Theta_0 = \{(x_1, \ldots, x_p) \in \Theta = \mathbb{R}^p : x_1 \leq \ldots \leq x_p\}$,ψ : isotonic regression transformation

- Lin and Dunson (2014)

  $\Theta = C[0,1]$, $\Theta_0 = C[0,1] \cap \mathcal{M}[0,1]$. $\psi(w(t)) = argmin_{F \in \Theta_0} \int_0^1 \{w(t) - F(t)\} G_0(t)$

- Chakraborty and Ghosal (2020)

  Θ : measurable function on $[0,1]$, $\Theta_0 = \mathcal{M}[0,1]$

  -Let $f(x) = \sum_{j=1}^{J} \theta_j I(x \in I_j)$, where $I_j = (\xi_{j-1}, \xi_j), 0 = \xi_0 < \ldots < \xi_J = 1$.

  -$f \in \Theta_0$, iff $\theta_1 \leq \ldots \leq \theta_J$.

  -Apply method of Dunson and Neelon (2003) to $(\theta_1, \ldots, \theta_J)$.

  -Posterior convergence rate is obtained, as $J \longrightarrow \infty$.

Similar Ideas of Transforming Initial Posterior Sample

- Projected posterior (Patra and Dunson; 2018)
  Θ : complete separable Banach space, Θ_0 : closed and convex subset of Θ,
  ψ : projection from Θ to Θ_0
  -Existence of data-dependent prior for the projected posterior.
  -The convergence rate of the projected posterior is at least that of the original posterior.
- The post-processed posterior for the banded covariance is the projected posterior under the Frobenius norm without positive definite condition.
- The idea of the post-processed posterior is the same as the projected posterior. The difference lies in the minimax consideration.

We compare the post-processed posterior with a conventional posterior in terms of interval estimation.

- $\Sigma = (\sigma_{ij})$, $\theta_1(\Sigma) = (\sigma_{ij}, |i-j| \leq k)$, $\theta_2(\Sigma) = (\sigma_{ij}, |i-j| > k)$
- Conventional Bayesian method
  It is the Bayesian method imposing a prior distribution on banded covariance matrices (θ_1) directly. This method is computationally intractable.
- $[\theta_1(\Sigma) \mid \mathbb{X}_n]_C$: a posterior distribution of the conventional Bayesian method.
  $[\theta_1(\Sigma) \mid \mathbb{X}_n]_{PPP}$: the post-processed posterior distribution.
  $[\theta_1(\Sigma) \mid \mathbb{X}_n]_{PPP,0}$: the post-processed posterior distribution without positive-definite adjustment.

We want to show that HPD (highest posterior density) region of $[\theta_1 \mid \mathbb{X}_n]_{PPP}$ is interpreted in terms of $[\theta_1 \mid \mathbb{X}_n]_C$. $(1-\alpha)100\%$ highest posterior density region of the post-processed posterior is asymptotically on the average an $(1-\alpha)100\%$ credible set of the conventional posterior.

Coverage Probability of Post-Processed Posterior

Theorem

Suppose A1 and A2 hold. If $C_{1-\alpha,n}$ is the highest posterior density regions of $[\theta_1 \mid \mathbb{X}_n]_{PPP}$ and p is fixed, then $\lim_{n \longrightarrow \infty} E_{\Sigma_0}\{\mathbb{P}[\theta_1(\Sigma) \in C_{1-\alpha,n} \mid \mathbb{X}_n]_C\} = 1 - \alpha.$

A1. (Bernstein-von Mises condition)

$$\begin{aligned}
\lim_{n \longrightarrow \infty} E_{\Sigma_0} ||[n^{1/2}(\theta_1(\Sigma) - \hat{\theta}_1) \mid \mathbb{X}_n]_C - N(0, \mathcal{I}_{11}^{-1}\{\theta_1(\Sigma_0), 0\})||_{TV} &= 0, \\
\lim_{n \longrightarrow \infty} E_{\Sigma_0} ||[n^{1/2}(\theta_1(\Sigma) - \hat{\theta}_1^*) \mid \mathbb{X}_n]_{PPP,0} - N(0, \mathcal{I}_{11 \cdot 2}^{-1}\{\theta_1(\Sigma_0), 0\})||_{TV} &= 0,
\end{aligned}$$

where $(\hat{\theta}_1^*, \hat{\theta}_2^*)^T = argmax_{\theta_1, \theta_2} \log p\{\mathbb{X}_n \mid \Sigma(\theta_1, \theta_2)\}$, $\hat{\theta}_1 = argmax_{\theta_1} \log p\{\mathbb{X}_n \mid \Sigma(\theta_1, 0)\}$ and $\mathcal{I}(\theta_1, \theta_2)$ is the Fisher-information matrix.

A2. As $n \longrightarrow \infty$,

$$\begin{aligned}
n^{1/2} L_n'\{\theta_1(\Sigma_0), 0\} &\xrightarrow{d} N[0, \mathcal{I}\{\theta_1(\Sigma_0), 0\}], \\
(\hat{\theta}_1^*, \hat{\theta}_2^*) &\xrightarrow{P} \{\theta_1(\Sigma_0), 0\}, \\
\hat{\theta}_1 &\xrightarrow{P} \theta_1(\Sigma_0),
\end{aligned}$$

and $L_n''\{\theta_1(\Sigma_0), 0\}$ is continuous, where $L_n(\theta_1, \theta_2) = \log p\{\mathbb{X}_n \mid \Sigma(\theta_1, \theta_2)\}$.

Overview

1. Banded Covariance Inference
2. Post-Processed Posterior
3. Asymptotic Analysis
4. Numerical Studies

P-risk: A New Framework for Bayesian Minimax Rate (Lee and Lee; 2018)

- Parameter and Parameter space : $\theta \in \Theta$
- Observation : $X \mid \theta \sim p(x \mid \theta)$, $X \in \mathcal{X}$
- Action and Action space ($\mathcal{A}$): An action is a probability measure on Θ, i.e. a posterior is an action.
- Decision rule ($\Pi : \mathcal{X} \mapsto \mathcal{A}$) : A prior on Θ is a decision rule. A prior with an observation determines an action, the posterior.
- P-loss ($\Theta \times \mathcal{A} \mapsto \mathbb{R}^+$) : $\mathcal{L}(\theta_0, \pi(\cdot \mid X)) := \mathbb{E}^\pi(L(\theta, \theta_0) \mid X)$
- P-risk ($\Theta \times \Pi \mapsto \mathbb{R}^+$) : $\mathcal{R}(\theta_0, \pi) := \mathbb{E}_{\theta_0}\mathbb{E}^\pi(L(\theta, \theta_0) \mid X)$

P-risk: A Extended Framework for Bayesian Minimax Rate

- Parameter and Parameter space : $\theta \in \Theta_0$
- Observation : $X \mid \theta \sim p(x \mid \theta)$, $X \in \mathcal{X}$
- Action and Action space ($\mathcal{A}$): An action is a probability measure on Θ, i.e. a post-processed posterior is an action.
- Decision rule ($(\Pi \times \mathcal{F}) : \mathcal{X} \mapsto \mathcal{A}$) :A pair of initial prior on Θ and post-processing function is a decision rule. A pair of prior and post-processing function with an observation determines an action, the post-processed posterior.
- P-loss ($\Theta \times \mathcal{A} \mapsto \mathbb{R}^+$) : $\mathcal{L}(\theta_0, \pi^{pp}(\cdot \mid X)) := \mathbb{E}^{\pi^{pp}}(L(\theta, \theta_0) \mid X)$
- P-risk ($\Theta \times (\Pi \times \mathcal{F}) \mapsto \mathbb{R}^+$): $\mathcal{R}(\theta_0, (\pi, f)) := \mathbb{E}_{\theta_0}\mathbb{E}^{\pi^{pp}}(L(\theta, \theta_0) \mid X)$

Decision rule of the extended framework includes that of the original framework.

P-risk Minimax Rate

- r_n is the P-risk minimax rate if and only if

$$\inf_{(\pi,f)\in\Pi\times\mathcal{F}} \sup_{\theta_0\in\Theta_0} \mathbb{E}_{\theta_0}\mathbb{E}^{\pi^{pp}}(L(f(\theta),\theta_0) \mid \mathbb{X}_n) \asymp r_n, \text{ as } n \longrightarrow \infty.$$

- A pair of prior and post-processing function (π^*, f^*) attains the P-risk minimax rate if

$$\sup_{\theta_0\in\Theta_0} \mathbb{E}_{\theta_0}\mathbb{E}^{\pi^*}(L(f^*(\theta),\theta_0) \mid \mathbb{X}_n) \asymp r_n, \text{ as } n \longrightarrow \infty.$$

Convergence rate of the post-processing posterior

Theorem (Convergence rate of the banding PPP)

Let the prior π^i *of* Σ *be* $IW_p(A_n, \nu_n)$*. If* $A_n \in \mathcal{B}_{p,k}$ *and* $n/4 \geq (M_0^{1/2} M_1^{-1} \log p) \vee k \vee ||A_n|| \vee (\nu_n - 2p)$*, then*

$$\sup_{\Sigma_0 \in \mathcal{B}_{p,k}} E_{\Sigma_0}\{E^{\pi^i}(||B_k^{(\epsilon_n)}(\Sigma) - \Sigma_0||^2 \mid \mathbb{X}_n)\} \leq C(\log k)^2 \frac{k + \log p}{n},$$

where $\epsilon_n^2 = O\{(\log k)^2(k + \log p)/n\}$*, and* C *depends on* M_0 *and* M_1*.*

Theorem (Minimax lower bound for banded covariances)

If $n/2 \geq [\min\{(M_0 - M_1)^2, 1\} \log p] \vee k$*, then*

$$\inf_{(\pi,f)\in\Pi^*} \sup_{\Sigma_0 \in \mathcal{B}_{p,k}} E_{\Sigma_0}\{E^{\pi}(||f(\Sigma) - \Sigma_0||^2 \mid \mathbb{X}_n)\} \geq C\frac{k + \log p}{n},$$

where C *depends on* M_0 *and* M_1*.*

The banding post-processed posterior is nearly optimal since its convergence rate has only $(\log k)^2$ factor up to a minimax lower bound.

Overview

1. Banded Covariance Inference
2. Post-Processed Posterior
3. Asymptotic Analysis
4. Numerical Studies

A Simulation Study (Banded Covariance)

Let $\Sigma_0^{(0)*} = (\sigma_{0,ij}^{(1)})_{p\times p}$, where

$$\sigma_{0,ij}^{(1)} = \begin{cases} 1, & 1 \le i = j \le p \\ 0.6|i-j|^{-(1+0.1)}, & 1 \le i \ne j \le p. \end{cases}$$

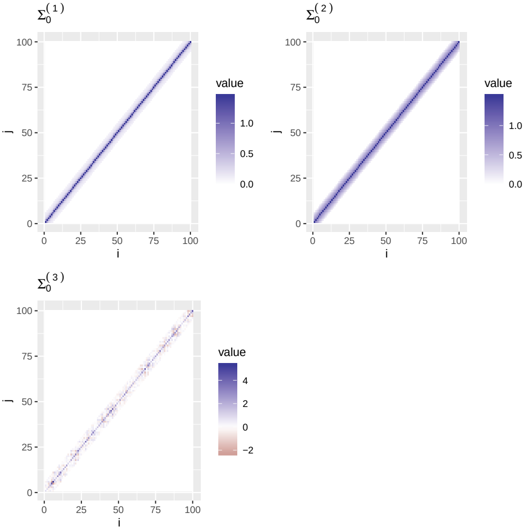
- $\Sigma_0^{(1)*} = B_{k_0}(\Sigma_0^{(0)*})$
- $\Sigma_0^{(2)*} = (\sigma_{0,ij}^{(2)})_{p\times p}$, where $\sigma_{0,ij}^{(2)} = \{1 - |i-j|/(k_0+1)\} \wedge 0$ for any $1 \le i, j \le p$.
- $\Sigma_0^{(3)*} = L_0 D_0 L_0^T$, where

$$L_{ij}^0 = \begin{cases} 1, & 1 \le i = j \le p \\ l_{ij}, & 0 < i - j \le k_0 \\ 0, & \text{otherwise,} \end{cases}$$

  l_{ij} are independent sample from $N(0,1)$, and $D_0 = diag(d_{ii})$ is a diagonal matrix where d_{ii} is independent sample from $IG(5,1)$, the inverse-gamma distribution with the shape parameter 5 and the scale parameter 1.
- $\Sigma_0^{(i)} = \Sigma_0^{(i)*} + [0.5 - \{\lambda_{\min}(\Sigma_0^{(i)*})\}]I_p,\ i = 1,2,3$

A Simulation Study (Banded Covariance)

We set $k_0 = 5$ for the definitions of $\Sigma_0^{(1)}$, $\Sigma_0^{(2)}$ and $\Sigma_0^{(3)}$.



Selection of Tuning Parameters

We select the tuning parameters using the leave-one-out cross-validation method (Gelman et al.; 2014)

- Post-processed posterior

$$argmax_{(k,\epsilon_n)}\hat{R}(k,\epsilon_n) = argmax_{(k,\epsilon_n)}\sum_{i=1}^{n}\log\frac{1}{S}\sum_{s=1}^{S}p\{X_i \mid B_k^{\epsilon_n}(\Sigma_s)\}\frac{\pi^i(\Sigma_s \mid \mathbb{X}_{n,-i})}{\pi^i(\Sigma_s \mid \mathbb{X}_n)},$$

where $\mathbb{X}_{n,-i} = (X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n)$, $\Sigma_1,\ldots,\Sigma_S$ are posterior sample from $\pi^i(\cdot \mid \mathbb{X}_n)$ and $p\{\cdot \mid B_k^{\epsilon_n}(\Sigma)\}$ is the multivariate normal density. $\hat{R}(k,\epsilon_n)$ is the estimator of $\sum_{i=1}^{n}\mathbb{E}^{\pi^{pp}}(\log p(X_i \mid B_k^{\epsilon_n}(\Sigma)) \mid \mathbb{X}_{n,-i})$.

- Frequentist

$$argmax_{(k,\epsilon_n)}\hat{R}_f(k,\epsilon_n) = argmax_{(k,\epsilon_n)}\sum_{i=1}^{n}\log p\{X_i \mid h(\mathbb{X}_{n,-i};k,\epsilon_n)\},$$

where $h(\mathbb{X}_n;k,\epsilon_n)$ is a frequentist estimator of Σ based on $\mathbb{X}_n$, bandwidth k and an adjustment parameter ϵ_n.

The performance of each estimator : $\frac{1}{S}\sum_{s=1}^{S}||\Sigma_0 - \hat{\Sigma}^{(s)}||$,

where $\hat{\Sigma}^{(s)}$ is a point estimate based on the sth simulated data set.

		$n = 25$			$n = 50$			$n = 100$	
	$\Sigma_0^{(1)}$	$\Sigma_0^{(2)}$	$\Sigma_0^{(3)}$	$\Sigma_0^{(1)}$	$\Sigma_0^{(2)}$	$\Sigma_0^{(3)}$	$\Sigma_0^{(1)}$	$\Sigma_0^{(2)}$	$\Sigma_0^{(3)}$
Banding post-processed posterior	3.67	4.62	5.63	2.16	3.01	3.61	1.48	1.94	2.34
G-inverse Wishart	3.60	5.79	6.83	3.28	5.21	6.08	2.77	4.4	5.16
Wishart for covariance graph	4.56	6.85	6.08	2.07	4.36	4.81	1.41	2.9	4.96
Banded sample covariance	3.38	4.5	5.66	2.19	2.8	3.42	1.51	1.9	2.23
Dual maximum likelihood estimator	3.96	6.41	7.67	3.9	6.28	7.55	3.33	5.23	6.33
Maximum likelihood estimator	4.96	4.78	6.92	2.31	2.52	3.4	1.42	1.76	2.17

To show the performance as an interval estimator for a functional, we consider the conditional mean estimation.

- Let $X^{(new)} \overset{i.i.d.}{\sim} N(0, \Sigma_0)$, which is independent $\mathbb{X}_n$.
- Conditional mean given $X_{-p}^{(new)}$, which is denoted by $\psi(\Sigma_0; X_{-p}^{(new)}) := (\Sigma_0)_{p,-p}(\Sigma_0)_{-p,-p}X_{-p}^{(new)}$, is to be estimated.
- Bayesian methods
  Given (post processed) posterior sample $\Sigma_1, \ldots, \Sigma_S$, the interval estimator for $\psi(\Sigma_0; X^{(new)})$ is given by quantiles of $\{\psi(\Sigma_i; X_{-p}^{(new)}); i = 1, \ldots, S\}$.
- Maximum likelihood estimator
  Apply asymptotic normality of MLE and delta method.

$$\psi(\Sigma^{MLE}; X^{(new)}) \pm z_{\alpha/2} \nabla^T \psi(\Sigma^{MLE}; X^{(new)}) \mathcal{I}(\Sigma^{MLE}) \nabla \psi(\Sigma^{MLE}; X^{(new)}).$$

Coverage probabilities and lengths of interval estimates of the conditional mean for banded covariances $\Sigma_0^{(1)}$, $\Sigma_0^{(2)}$ and $\Sigma_0^{(3)}$. The average lengths of intervals are represented in parentheses.

		$n = 25$			$n = 50$	
	$\Sigma_0^{(1)}$	$\Sigma_0^{(2)}$	$\Sigma_0^{(3)}$	$\Sigma_0^{(1)}$	$\Sigma_0^{(2)}$	$\Sigma_0^{(3)}$
Banding post-processed posterior	96.7% (2.54)	95.5% (2.27)	94.3% (3.66)	98.3% (2.05)	96.7% (1.93)	98.5% (3.32)
G-inverse Wishart	44.7% (1.02)	49.1% (0.96)	45.6% (1.55)	60.2% (0.67)	61.4% (0.68)	60.7% (1.02)
Wishart for covariance graph	99.2% (2.97)	99.7% (3.24)	97.4% (3.69)	97.7% (1.49)	99.3% (1.85)	91.5% (1.8)
Maximum likelihood estimator	100% (10.67)	100% (13.61)	100% (32.88)	100% (3.12)	100% (4.75)	99.9% (8.82)
		$n = 100$				
Post-processed posterior	95.6% (1.21)	96.9% (1.55)	98.7% (2.75)			
G-inverse Wishart	74.4% (0.57)	74.2% (0.56)	73.1% (0.81)			
Wishart for covariance graph	93.3% (0.92)	97.3% (1.11)	88.3% (1.09)			
Maximum likelihood estimator	99.8% (1.66)	100% (2.84)	100% (5.38)			

Conclusion

- The post-processed posterior method is computationally efficient.
- The post-processed posterior has nearly minimax convergence rate.
- The simulation study shows that the post-processed posterior spectral norm error is smaller than the other methods, and the credible interval attains nominal coverage probability.

References I

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

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**(2): 455–488.

Cai, T. T. and Zhou, H. H. (2010). Optimal rates of convergence for covariance matrix estimation, *The Annals of Statistics* **38**(4): 2118–2144.

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

Cai, T. T. and Zhou, H. H. (2012b). Optimal rates of convergence for sparse covariance matrix estimation, *The Annals of Statistics* **40**(5): 2389–2420.

Chakraborty, M. and Ghosal, S. (2020). Convergence rates for bayesian estimation and testing in monotone regression, *arXiv preprint arXiv:2008.01244* .

Chaudhuri, S., Drton, M. and Richardson, T. S. (2007). Estimation of a covariance matrix with zeros, *Biometrika* **94**(1): 199–216.

Donoho, D. L., Gavish, M. and Johnstone, I. M. (2018). Optimal shrinkage of eigenvalues in the spiked covariance model, *Annals of statistics* **46**(4): 1742.

References II

Dunson, D. B. and Neelon, B. (2003). Bayesian inference on order-constrained parameters in generalized linear models, *Biometrics* **59**(2): 286–295.

Fan, J., Liao, Y. and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements, *Journal of the Royal Statistical Society. Series B, Statistical methodology* **75**(4).

Gelman, A., Hwang, J. and Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models, *Statistics and computing* **24**(6): 997–1016.

Ghosal, S. and Van der Vaart, A. (2017). *Fundamentals of nonparametric Bayesian inference*, Vol. 44, Cambridge University Press.

Hu, A. and Negahban, S. (2017). Minimax estimation of bandable precision matrices, *Advances in Neural Information Processing Systems*, pp. 4888–4896.

Huang, A. and Wand, M. P. (2013). Simple marginally noninformative prior distributions for covariance matrices, *Bayesian Analysis* **8**(2): 439–452.

Kauermann, G. (1996). On a dualization of graphical Gaussian models, *Scandinavian journal of statistics* pp. 105–116.

References III

Khare, K. and Rajaratnam, B. (2011). Wishart distributions for decomposable covariance graph models, *The Annals of Statistics* **39**(1): 514–555.

Lee, K. and Lee, J. (2017). Estimating large precision matrices via modified Cholesky decomposition, *Statistica Sinica* **forthcoming**: arXiv:1707.01143.

Lee, K. and Lee, J. (2018). Optimal Bayesian minimax rates for unconstrained large covariance matrices, *Bayesian Analysis* **13**(4): 1215–1233.

Lin, L. and Dunson, D. B. (2014). Bayesian monotone regression using Gaussian process projection, *Biometrika* **101**(2): 303–317.

Patra, S. and Dunson, D. B. (2018). Constrained bayesian inference through posterior projections, *arXiv preprint arXiv:1812.05741* .

Silva, R. and Ghahramani, Z. (2009). The hidden life of latent variables: Bayesian learning with mixed graph models, *Journal of Machine Learning Research* **10**(Jun): 1187–1238.

Wiesel, A. and Globerson, A. (2012). Covariance estimation in time varying ARMA processes, *Sensor Array and Multichannel Signal Processing Workshop (SAM), 2012 IEEE 7th*, IEEE, pp. 357–360.

Wu, W. B. and Pourahmadi, M. (2009). Banding sample autocovariance matrices of stationary processes, *Statistica Sinica* pp. 1755–1768.