Electronic Journal of Statistics
Vol. 9 (2015) 2828–2854
ISSN: 1935-7524
DOI: 10.1214/15-EJS1084

High dimensional posterior convergence rates for decomposable graphical models

Ruoxuan Xiang, Kshitij Khare and Malay Ghosh

Department of Statistics, University of Florida
Gainesville, Florida, 32611 USA
e-mail: rxxiang@ufl.edu; kdkhare@stat.ufl.edu; ghoshm@stat.ufl.edu

Abstract: Gaussian concentration graphical models are one of the most popular models for sparse covariance estimation with high-dimensional data. In recent years, much research has gone into development of methods which facilitate Bayesian inference for these models under the standard G-Wishart prior. However, convergence properties of the resulting posteriors are not completely understood, particularly in high-dimensional settings. In this paper, we derive high-dimensional posterior convergence rates for the class of decomposable concentration graphical models. A key initial step which facilitates our analysis is transformation to the Cholesky factor of the inverse covariance matrix. As a by-product of our analysis, we also obtain convergence rates for the corresponding maximum likelihood estimator.

MSC 2010 subject classifications: Primary 62F15; secondary 62G20.
Keywords and phrases: Graphical models, decomposable graph, posterior consistency, high-dimensional data.

Received April 2015.

Contents

1. Introduction

Covariance estimation is a fundamental problem in multivariate statistical inference, and plays a crucial role in many inferential and data analytic methods. For instance, methods such as principal component analysis (PCA), multivariate analysis of variance (MANOVA), classification via linear/quadratic discriminant analysis (LDA/QDA), canonical correlation analysis (CCA) all require estimation of the covariance matrix (or some appropriate function of its entries). In recent years, advances in science and information technology have led to an explosion of "high-dimensional" datasets from a variety of scientific fields. In such datasets, the number of variables p is either of the same order of, or much larger than the number of samples n. It is well-known that in high-dimensional settings, the sample covariance matrix (traditional estimator for the population covariance matrix), can perform rather poorly (see [7, 8, 16, 17, 18]). To address the challenge posed by high-dimensionality, several promising methods have been proposed in the literature. In particular, methods inducing sparsity in the covariance matrix, or in an appropriate function of the covariance matrix, have proven to be very effective in applications.

Perhaps the most well-known and well-studied class of sparsity based models in this context is the class of concentration graphical models introduced in [10]. These models induce sparsity in the concentration matrix (or the inverse covariance matrix). If the underlying distribution is assumed to be multivariate Gaussian, then zeros in the inverse covariance matrix correspond to conditional independence. Hence, concentration graphical models achieve parameter reduction in a naturally interpretable way. To understand the connection with graphs, consider *i.i.d.* vectors $\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_n$ which are drawn from a p-variate normal distribution with mean vector $\mathbf{0}$ and covariance matrix Σ. A given sparsity pattern on $\Omega = \Sigma^{-1}$ can be encoded in terms of a graph G on the set of p variables as follows. If the variables i and j do not share an edge in G, then $\Omega_{ij} = 0$. Hence, a concentration graph model corresponding to a graph G restricts the inverse covariance matrix Ω to a submanifold of the cone of positive definite matrices (referred to as P_G).

There are two major approaches in the literature for analyzing concentration graphical models. The first approach is based on regularized likelihood/pseudo-likelihood using ℓ_1 penalization. A variety of methods using this approach have been proposed (see [3, 7, 12, 13, 15, 20, 24, 27, 29, 35] and the references therein). These optimization based methods often undertake estimation and sparsity selection (selection of G) simultaneously, and have fared well in high-dimensional settings. For a majority of these approaches, high-dimensional estimation/selection consistency of the corresponding estimators has been established.

The second approach is based on the Bayesian paradigm. Dawid and Lauritzen [9] introduced a class of prior distributions called Hyper Inverse Wishart distributions for the covariance matrix $\Sigma = \Omega^{-1}$. The induced class of priors for Ω (supported on P_G) is known as the class of G-Wishart distributions (see [31]). This class of prior distributions is quite useful and popular, and has several desirable properties, including the fact that it forms a Diaconis-Ylvisaker conjugate class of priors [11] for the concentration graph model corresponding to the graph G. Subsequently, several techniques for posterior inference using the G-Wishart distribution have been developed in the literature (see for instance [1, 22, 25, 31, 32, 34]). The subfamily of *decomposable* graphs has featured prominently in the Bayesian literature on concentration graph models (see for instance [9, 23, 28, 31]). In fact, direct sampling and closed form computations of relevant posterior expected values corresponding to the G-Wishart distribution are in general available only if the underlying graph is decomposable.

High-dimensional posterior convergence results have been established for some Bayesian covariance estimation models in the literature. Ghosal [14] proves asymptotic normality of the posterior for exponential family models (which include the multivariate normal model) when p grows much slower than n. Pati et al. [26] establish high-dimensional posterior convergence rates for covariance estimation using Bayesian factor models. Banerjee and Ghosal [5] derive high-dimensional posterior convergence rates for Bayesian concentration graphical models with priors which are obtained by mixing a point mass at zero with an appropriate continuous distribution. However, despite the great interest and activity in Bayesian inference for concentration graphical models with G-Wishart priors, a complete investigation into the important issue of high-dimensional posterior consistency in this setting has not been undertaken. To the best of our knowledge, the only results along these lines can be found in the recent work of Banerjee and Ghosal [4]. Under standard regularity assumptions, the authors in [4] provide posterior convergence rates for banded concentration graphical models (with a G-Wishart prior) in the popular and useful high-dimensional setting where the "true" concentration matrices generating the data are allowed to be approximately banded.

Banded models form a subclass of decomposable concentration graphical models. Furthermore, quantities such as the posterior mean, normalizing constant etc. are available in closed form for decomposable graphical models with G-Wishart priors. These facts led us to investigate whether one can prove consistency results similar to [4] for the class of decomposable concentration graphical models. In this paper we achieve this goal by providing high-dimensional posterior convergence rates for decomposable concentration graphical models with G-Wishart priors, where the "true" concentration matrices generating the data are approximately decomposable (Theorem 3.1). A key initial step in our analysis is expressing the concentration matrix Ω in terms of its Cholesky parameter. The main result is then established using a combination of extensive analytic arguments and distributional results for the Cholesky parameter of decomposable G-Wishart matrices. This approach is quite different from the one taken in [4]. Based on a reviewer's comment, we investigated and found that the approach used in [4] can be easily and directly generalized to decomposable graphs, thereby providing another route to prove Theorem 3.1. Nevertheless, our extensive set on intermediate lemmas (Lemmas 3.2, 3.3, 3.5, 3.6, 3.9) regarding posterior convergence properties of the Cholesky parameter are of independent interest, and may serve as useful tools in other related problems.

The rest of the paper is organized as follows. Section 2 contains a brief overview of relevant concepts from graph theory and matrix theory. In Section 3, we provide the required assumptions, and then state and prove the main result (Theorem 3.1) regarding posterior convergence rates for decomposable concentration graphical models. The proof of Theorem 3.1 is preceded by a series of lemmas which pave the way for establishing Theorem 3.1. One of these lemmas (Lemma 3.7) establishes convergence rates for the maximum likelihood estimator. The proofs of all the lemmas are provided in the appendix.

2. Preliminaries

In this section, we provide required background material from graph theory and graphical models.

2.1. Decomposable graphs and Cholesky decomposition

An undirected graph $G = (V, E)$ consists of a vertex set $V = \{1, ..., p\}$ with an edge set $E \subseteq \{(i, j) \in V \times V, i \neq j\}$, with $(i, j) \in E$ if and only if $(j, i) \in E$. Two vertices $v, v' \in V$ are called adjacent if there is an edge between v and v'. An undirected graph $G = (V, E)$ is called a *complete graph* if all pairs of distinct vertices in V are adjacent, and is called a *cycle* if there exists a permutation $\{v_1, v_2, \cdots, v_p\}$ of V such that $(v_i, v_j) \in E$ if and only if $|i - j| = 1$ or $|i - j| = p - 1$. The *induced subgraph* of $G = (V, E)$ corresponding to $V' \subseteq V$ is an undirected graph with vertex set V' and edge set given by $E' = E \cap (V' \times V')$. A subset V' of V is called a *clique* if the induced subgraph corresponding to V' is a complete graph. See [21, 23] for more details.

For an undirected graph $G = (V, E)$, we denote by M_G the set of all $|V| \times |V|$ matrices $A = (A_{ij})_{1 \leq i,j \leq |V|}$ satisfying $A_{ij} = A_{ji} = 0$ for all pairs $(i, j) \notin E, i \neq j$, and by P_G the set of all $|V| \times |V|$ symmetric positive definite matrices that are in M_G. A graph $G' = (V', E')$ is defined as an induced subgraph of $G = (V, E)$ if $V' \subseteq V$ and $E' = (V' \times V') \cap E$, and is denoted by $G' \subseteq G$. We now recall the definition of decomposable graphs.

Definition 2.1. *([21]) A graph G is defined to be decomposable if it does not contain a cycle of length ≥ 4 as an induced subgraph.*

There are several other characterizations of decomposable graphs. Here is a recursive characterization. A graph $G = (V, E)$ is decomposable if and only if it is either complete (no missing edges) or if there exist non-empty disjoint subsets A, B, C of V such that (i) $A \cup B \cup C = V$ (ii) the graph induced by B is complete (iii) any path from A to C passes through B, and (iv) the graphs induced by $A \cup B$ and $B \cup C$ are decomposable.

Recall that for every positive definite matrix A, there exists a unique lower triangular matrix L (with positive diagonal elements) such that $A = LL^T$. This decomposition of A is known as the Cholesky decomposition and we refer to L as the Cholesky factor of A. The following characterization of decomposable graphs (see [31, Theorem 1]), in terms of Cholesky decomposition, will be useful for our analysis.

Lemma 2.1. *([31]) An undirected graph $G = (V, E)$ is decomposable if and only if there exists a permutation of vertices V such that after reordering the vertices based on this permutation, every $A \in P_G$ factors as $A = LL^T$ where $L \in M_G$ and L is lower triangular with positive diagonal entries. Such a permutation is called a perfect vertex elimination scheme for G.*

The above lemma says that for a decomposable graph G, if the vertices are ordered according to a perfect vertex elimination scheme and $\Omega = LL^T$, $\Omega \in P_G$, then L has the same zero pattern as Ω in its lower triangle. Note that a perfect vertex elimination scheme is not unique and several of them can exist for a given graph G. A variety of algorithms for obtaining perfect vertex elimination schemes (or perfect elimination orderings, as they are known in the computer science literature) are available in the literature. See [30, 33] for instance.

2.2. The G-Wishart distribution

Let $\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_n$ be independent, identically distributed $MVN_p(\mathbf{0}, \Sigma = \Omega^{-1})$ random vectors, where $\boldsymbol{Y}_i = (Y_{i1}, ..., Y_{ip})^T, 1 \leq i \leq n$ and MVN stands for the multivariate Gaussian distribution. For an undirected graph $G = (V, E)$ (with $V = \{1, ..., p\}$), the Gaussian concentration graphical model corresponding to G assumes that $\Omega \in P_G$. Dawid and Lauritzen [9] developed the class of Hyper Inverse Wishart distributions for $\Sigma = \Omega^{-1}$. The class of induced priors for Ω are known as the G-Wishart distributions on P_G. In particular, the G-Wishart distribution with parameter $\delta > 0$ and D positive definite, denoted by $W_G(\delta, D)$, has density proportional to

$$(det(\Omega))^{\delta/2} exp[-tr(D\Omega)/2], \, \Omega \in P_G. \tag{2.1}$$

The class of G-Wishart distributions on P_G form a conjugate family of priors under the Gaussian concentration graphical model corresponding to G. In particular, if the priors on $\Omega \in P_G$ is $W_G(\delta, D)$, then it can be easily shown that the posterior on Ω is $W_G(\delta + n, D + nS)$, where $S = 1/n \sum_{i=1}^n \boldsymbol{Y}_i \boldsymbol{Y}_i^T$ is the sample covariance matrix. If G is decomposable, then quantities such as the mean, mode and normalizing constant for $W_G(\delta, D)$ are available in closed form (see for instance [28]), but if G is non-decomposable, one has to resort to MCMC to estimate these quantities (see for instance [1, 2, 25, 32, 34]).

3. Main results

In this section, we will provide the main high-dimensional posterior convergence result. We start by introducing some required notation. For $x \in \mathbb{R}^p$, $\|x\|_r = (\sum_{j=1}^p |x_j|^r)^{1/r}$ and $\|x\|_\infty = \max_j |x_j|$ denote the standard l_r and l_∞ vector norms. For a $p \times p$ matrix $A = (A_{ij})_{1 \leq i,j \leq p}$ with ordered eigenvalues $|eig_1(A)| \leq ... \leq |eig_p(A)|$, we denote

$$\|A\|_{\max} = \max_{1 \leq i,j \leq p} |A_{ij}|,$$

$$\|A\|_{(r,s)} = \sup\{\|Ax\|_s : \|x\|_r = 1\},$$

where $1 \leq r, s \leq \infty$. In particular, we have

$$\|A\|_{(1,1)} = \max_j \sum_i |A_{ij}|, \, \|A\|_{(\infty,\infty)} = \max_i \sum_j |A_{ij}|, \, \|A\|_{(2,2)} = \{eig_p(A^T A)\}^{1/2}.$$

Also, for a symmetric matrix A, we have $\|A\|_{(1,1)} = \|A\|_{(\infty,\infty)}$ and $\|A\|_{(2,2)} = |eig_p(A)|$. Now, given the graph $G = (V, E)$, with $V = \{1, ..., p\}$, we denote $A^{>i} = (A_{jk})_{i<j,k\leq p,(i,j)\in E,(i,k)\in E}$, the column vectors $A_{\cdot i}^{>} = (A_{ji})_{j>i,(i,j)\in E}$ and $A_{\cdot i}^{\geq} = (A_{ii}, (A_{\cdot i}^{>})^T)^T$. Also,

$$A^{\geq i} = \begin{bmatrix} A_{ii} & (A_{\cdot i}^{>})^T \\ A_{\cdot i}^{>} & A^{>i} \end{bmatrix}.$$

In particular, $A_{\cdot p}^{\geq} = A^{\geq p} = A_{pp}$.

We now provide the model specification and required assumptions. We consider a setting when the number of variables $p = p_n$ increases with the sample size n. Suppose that $\boldsymbol{Y}_1^n, ..., \boldsymbol{Y}_n^n$ are independent and identically distributed random vectors drawn from a $MVN_{p_n}(\mathbf{0}, \bar{\Omega}_n^{-1})$ distribution. Hence, $\{\bar{\Omega}_n\}_{n\geq 1}$ denotes the sequence of true concentration matrices. Let $G_n = (V_n, E_n)$ (with $V_n = \{1, ..., p_n\}$) be a decomposable graph with vertices ordered according to a perfect vertex elimination scheme. We define the matrix $\tilde{\Omega}_n$ by

$$(\tilde{\Omega}_n)_{ij} = \begin{cases} (\bar{\Omega}_n)_{ij} & \text{if } (i,j) \in E_n \\ 0 & \text{otherwise,} \end{cases}$$

and $A_n = \bar{\Omega}_n - \tilde{\Omega}_n$. Hence, $\tilde{\Omega}_n \in M_{G_n}$. The assumption that the vertices of G_n are ordered according to a perfect vertex elimination scheme does not lead to any loss of generality, as $\|\Omega - \bar{\Omega}_n\|_{(\infty,\infty)}$ (see statement of Theorem 3.1) is invariant with respect to any reordering of vertices. Let d_n denote the maximum number of non-zero entries in any row (column) of the symmetric matrix $\tilde{\Omega}_n$. Also, $\bar{P}$ and $\bar{E}$ respectively denote the probability measure and expected value corresponding to the "true" Gaussian model specified above. With the model specification in place, we now provide the required assumptions.

Assumption 1. *The eigenvalues of* $\{\bar{\Omega}_n\}_{n\geq 1}$ *are uniformly bounded, i.e., there exists* $\epsilon_0 > 0$ *such that* $0 < \epsilon_0 \leq eig_1(\bar{\Omega}_n) \leq eig_p(\bar{\Omega}_n) \leq \epsilon_0^{-1} < \infty$, *for every* $n \geq 1$.

Assumption 2. *For* $n \geq 1$, *we use the prior* $W_{G_n}(\delta, D_n)$ *for the concentration matrix* Ω. *Here* $\delta > 0$ *and* D_n *is a positive definite matrix which satisfies:* $eig_p(D_n) \leq a < \infty$, *for every* $n \geq 1$.

Assumption 3. $d_n^5 \log p_n / n \to 0$ *and* $p_n \to \infty$.

Assumption 4. $\|A_n\|_{(\infty,\infty)} \leq \gamma(d_n)$, *where* $d_n^{3/2}\gamma(d_n) \to 0$.

Henceforth, for notational and expositional convenience, we will refer to p_n, $\bar{\Omega}_n$, $\tilde{\Omega}_n$, G_n, d_n, D_n, A_n as $p, \bar{\Omega}, \tilde{\Omega}, G, d, D, A$. Note also that ϵ_0, δ and a above do not depend on n.

The following theorem is the main contribution of this paper and provides the convergence rate for the posterior distribution of Ω. As in [4], we provide the convergence rate under the (∞, ∞) norm, which is a stronger norm than the more standard $(2, 2)$ norm.

Theorem 3.1. *Under Assumptions 1 - 4, for a large enough constant K (not depending on n), the posterior distribution of Ω satisfies:*

$$\bar{E}[Pr\{\|\Omega - \bar{\Omega}\|_{(\infty,\infty)} \geq K\epsilon_n|\boldsymbol{Y}\}] \to 0, \tag{3.1}$$

where $\epsilon_n = d^{5/2}(\log p/n)^{1/2} + d^{3/2}\gamma(d)$.

Remark 3.1. *Based on a reviewer's comment, we investigated and found that Theorem 3.1 for approximate decomposable models can be obtained (with the exact same convergence rate) by a direct extension of the approach in [4].*

We now provide a series of lemmas which will play a crucial role in the proof of Theorem 3.1. The proofs of these lemmas are provided in the appendix. The first lemma provides inequalities involving the various matrix norms introduced earlier in this section, in the context of sparse matrices.

Lemma 3.1. *For any $p \times p$ matrix $A = (A_{ij})_{1\leq i,j\leq p}$ with at most d nonzero elements in each row and each column, we have*

$$\|A\|_{(\infty,\infty)} \leq \sqrt{d}\|A\|_{(2,2)}, \tag{3.2}$$

$$\|A\|_{(2,2)} \leq d\|A\|_{\max}. \tag{3.3}$$

The next lemma provides an expression for the Cholesky parameter $\tilde{L}$ of $\tilde{\Omega}$ in terms of $\tilde{\Sigma} = \tilde{\Omega}^{-1}$. Note that by Assumption 4, for large enough n,

$$\begin{aligned}
&\|\tilde{\Omega}\|_{(2,2)} \leq \|\bar{\Omega} - A\|_{(2,2)} \leq \|\bar{\Omega}\|_{(2,2)} + \|A\|_{(2,2)} \leq \epsilon_0^{-1} + \gamma(d) \leq 2\epsilon_0^{-1}, \text{ and} \\
&eig_1(\tilde{\Omega}) \geq eig_1(\bar{\Omega}) + eig_1(-A) \geq \epsilon_0 - \gamma(d) \geq \epsilon_0/2.
\end{aligned} \tag{3.4}$$

Thus, $\tilde{\Omega} \in P_G$ for large enough n.

Lemma 3.2. *For large enough n, let $\tilde{\Omega} = \tilde{L}\tilde{L}^T$ be the Cholesky decomposition of $\tilde{\Omega} \in P_G$. Let $\tilde{\Sigma} = \tilde{\Omega}^{-1}$. Then $\tilde{L}_{ji} = 0$ for $1 \leq i < j \leq p$ and $(i,j) \notin E$, and*

$$\tilde{L}_{ii} = \sqrt{\frac{1}{\tilde{\Sigma}_{ii} - (\tilde{\Sigma}^{>}_{\cdot i})^T(\tilde{\Sigma}^{>i})^{-1}\tilde{\Sigma}^{>}_{\cdot i}}}, \quad \tilde{L}^{>}_{\cdot i} = -\tilde{L}_{ii}(\tilde{\Sigma}^{>i})^{-1}\tilde{\Sigma}^{>}_{\cdot i}, \tag{3.5}$$

for $1 \leq i \leq p$.

The next lemma establishes distributional properties for the Cholesky parameter of a matrix following a G-Wishart distribution.

Lemma 3.3. *Let $\Omega = LL^T$ be the Cholesky decomposition of $\Omega \in P_G$, where L is a lower triangular matrix. Under the prior distribution $W_G(\delta, D)$ on Ω, the posterior distribution of L is: $L_{ji} = 0$ for $1 \leq i < j \leq p$ and $(i,j) \notin E$; $L^{\geq}_{\cdot i}$ are independent for $1 \leq i \leq p$ and*

$$L^{>}_{\cdot i}|L_{ii} \sim MVN(-(\overleftarrow{S}^{>i})^{-1}\overleftarrow{S}^{>}_{\cdot i}L_{ii}, (\overleftarrow{S}^{>i})^{-1}/n), \quad \textit{for} \quad 1 \leq i \leq p-1, \tag{3.6}$$

$$L^2_{ii} \sim Gamma((n + v_i + \delta)/2 + 1, nc_i/2), \quad \textit{for} \quad 1 \leq i \leq p, \tag{3.7}$$

where $\overleftarrow{S} = S + D/n$, $v_i = dim(L^{>}_{\cdot i})$ and $c_i = \overleftarrow{S}_{ii} - (\overleftarrow{S}^{>}_{\cdot i})^T(\overleftarrow{S}^{>i})^{-1}\overleftarrow{S}^{>}_{\cdot i}$.

After a careful comparison, it can be argued that the above lemma follows from [23, Theorem 4.4]. Nevertheless, we provide a proof of this lemma in the appendix for completeness. The next three lemmas show that the difference between relevant functions of $\overleftarrow{S} = S + D/n$ and the same functions applied to $\tilde{\Sigma}$, converges to zero in $\bar{P}$-probability at a certain rate.

Lemma 3.4. *For a large enough constant M_1 (not depending on n), we have*

$$\bar{P}(\|\overleftarrow{S} - \tilde{\Sigma}\|_{\max} \geq M_1(\sqrt{\log p/n} + \gamma(d))) \leq m_1 p^{2-m_2M_1^2/4} \to 0 \tag{3.8}$$

as $n \to \infty$, where m_1 and m_2 are constants which depend only on ϵ_0 (from Assumption 1).

Lemma 3.5. *For large enough constants M_1 (same as in Lemma 3.4) and M_2 (not depending on n),*

$$\bar{P}(\max_{1\leq i\leq p} \|\overleftarrow{S}^{\geq i} - \tilde{\Sigma}^{\geq i}\|_{(2,2)} \geq M_1(d\sqrt{\log p/n} + \gamma(d))) \to 0, \tag{3.9}$$

$$\bar{P}(\max_{1\leq i\leq p} \|(\overleftarrow{S}^{\geq i})^{-1}\|_{(2,2)} \geq 2\epsilon_0^{-1}) \to 0, \tag{3.10}$$

$$\bar{P}(\max_{1\leq i\leq p} \|(\overleftarrow{S}^{\geq i})^{-1} - (\tilde{\Sigma}^{\geq i})^{-1}\|_{(2,2)} \geq M_2(d\sqrt{\log p/n} + \gamma(d))) \to 0, \tag{3.11}$$

as $n \to \infty$.

Lemma 3.6. *Let $\tilde{c}_i = \tilde{\Sigma}_{ii} - (\tilde{\Sigma}^{>}_{\cdot i})^T(\tilde{\Sigma}^{>i})^{-1}\tilde{\Sigma}^{>}_{\cdot i}$. Then*

$$\epsilon_0/2 \leq \min_{1\leq i\leq p} \tilde{c}_i \leq \max_{1\leq i\leq p} \tilde{c}_i \leq 2\epsilon_0^{-1}. \tag{3.12}$$

Let $c_i = \overleftarrow{S}_{ii} - (\overleftarrow{S}^{>}_{\cdot i})^T(\overleftarrow{S}^{>i})^{-1}\overleftarrow{S}^{>}_{\cdot i}$. Then

$$\bar{P}(\max_{1\leq i\leq p} c_i \geq 2\epsilon_0^{-1}) \to 0,\ \ \bar{P}(\min_{1\leq i\leq p} c_i \leq \epsilon_0/2) \to 0, \tag{3.13}$$

as $n \to \infty$. Also, for a large enough constant M_3 (not depending on n), we have

$$\bar{P}(\max_{1\leq i\leq p} |c_i - \tilde{c}_i| \geq M_3(d\sqrt{\log p/n} + \gamma(d))) \to 0. \tag{3.14}$$

Note that by Assumption 1, $d \leq n$ for large enough n. Hence the size of any clique in G is bounded by n. In this case, it is known that the maximum likelihood estimator exists and is unique (see for instance [21]). Although the main goal of this paper is derivation of the convergence rate for the posterior distribution, using the previous lemmas we derive the convergence rate for the maximum likelihood estimator under Assumptions 1 - 4.

Lemma 3.7 (Convergence Rate for MLE)**.** *For a large enough constant $\tilde{K}$ (not depending on n), the maximum likelihood estimator $\hat{\Omega}$ satisfies*

$$\bar{P}(\|\hat{\Omega} - \bar{\Omega}\|_{(\infty,\infty)} \geq \tilde{K}\epsilon_n) \to 0,$$

as $n \to \infty$.

The proof is provided in the appendix. We need two more lemmas before proceeding to prove Theorem 3.1. The next lemma gives a concentration bound (around the mean value) for the square root of a Gamma random variable.

Lemma 3.8. *If $X^2 \sim Gamma(\alpha, \lambda), X > 0$. Let $\mu = E(X) = \frac{\Gamma(\alpha+1/2)}{\sqrt{\lambda}\Gamma(\alpha)}$ and $\alpha > 1/2$. Then*

$$Pr(|X - \mu| \geq x) \leq \sqrt{2e}\, e^{-\lambda x^2}. \tag{3.15}$$

The final lemma provides the posterior convergence rate for the diagonal entries of the Cholesky factor of Ω.

Lemma 3.9. *For a large enough constant M_5 (not depending on n), we have*

$$Pr(\max_{1 \leq i \leq p} |L_{ii} - \tilde{L}_{ii}| \geq M_5(d\sqrt{\log p/n} + \gamma(d))|\boldsymbol{Y}) \xrightarrow{\bar{P}} 0. \tag{3.16}$$

We would like to remind the reader that the proofs of all the above lemmas are provided in the appendix. With these lemmas in hand, we now provide a proof for Theorem 3.1.

Proof of Theorem 3.1. Note that it suffices to show $Pr\{\|\Omega - \bar{\Omega}\|_{(\infty,\infty)} \geq K\epsilon_n|\boldsymbol{Y}\} \xrightarrow{\bar{P}} 0$ to get the required result. Also, since $\|\tilde{\Omega} - \bar{\Omega}\|_{(\infty,\infty)} \leq \gamma(d)$, by the triangle inequality it suffices to show $Pr\{\|\Omega - \tilde{\Omega}\|_{(\infty,\infty)} \geq K\epsilon_n|\boldsymbol{Y}\} \xrightarrow{\bar{P}} 0$. By Lemma 3.1, we have

$$\begin{aligned}
& Pr\{\|\Omega - \tilde{\Omega}\|_{(\infty,\infty)} \geq K\epsilon_n|\boldsymbol{Y}\} \\
\leq\ & Pr\{\sqrt{d}\|\Omega - \tilde{\Omega}\|_{(2,2)} \geq K\epsilon_n|\boldsymbol{Y}\} \\
\leq\ & Pr\{\|LL^T - \tilde{L}\tilde{L}^T\|_{(2,2)} \geq K\epsilon_n/\sqrt{d}|\boldsymbol{Y}\} \\
\leq\ & Pr\{\|L\|_{(2,2)}\|L - \tilde{L}\|_{(2,2)} + \|\tilde{L}\|_{(2,2)}\|L - \tilde{L}\|_{(2,2)} \geq K\epsilon_n/\sqrt{d}|\boldsymbol{Y}\} \\
\leq\ & Pr\{(2\|\tilde{L}\|_{(2,2)} + \|L - \tilde{L}\|_{(2,2)})\|L - \tilde{L}\|_{(2,2)} \geq K\epsilon_n/\sqrt{d}|\boldsymbol{Y}\} \\
\leq\ & Pr\left\{\|L - \tilde{L}\|_{(2,2)} \geq \frac{K\epsilon_n}{4\|\tilde{L}\|_{(2,2)}\sqrt{d}}\middle|\boldsymbol{Y}\right\} + \\
& Pr\left\{\|L - \tilde{L}\|_{(2,2)} \geq \frac{\sqrt{K}\sqrt{\epsilon_n}}{\sqrt{2}\sqrt[4]{d}}\middle|\boldsymbol{Y}\right\}.
\end{aligned} \tag{3.17}$$

Since $\epsilon_0/2 \leq eig_1(\tilde{\Omega}) \leq eig_p(\tilde{\Omega}) \leq 2\epsilon_0^{-1}$ by (3.4), it follows that $\sqrt{\epsilon_0/2} \leq \|\tilde{L}\|_{(2,2)} = \{eig_p(\tilde{\Omega})\}^{1/2} \leq (\epsilon_0/2)^{-1/2}$. Also by Assumption 3, it follows that $\sqrt{\epsilon_n}/\sqrt[4]{d} \geq \epsilon_n/\sqrt{d}$ for large enough n. In view of these observations and (3.17), it suffices to show that $Pr\{\|L - \tilde{L}\|_{(2,2)} \geq K_1\epsilon_n/\sqrt{d}|\boldsymbol{Y}\} \xrightarrow{\bar{P}} 0$ for a large enough constant K_1. Now, let $\delta_n = \epsilon_n/d^{3/2} = d(\log p/n)^{1/2} + \gamma(d)$. It follows by Lemma 3.1 and the triangle inequality that

$$Pr\{\|L - \tilde{L}\|_{(2,2)} \geq K_1\epsilon_n/\sqrt{d}|\boldsymbol{Y}\} \leq Pr\{d\|L - \tilde{L}\|_{\max} \geq K_1\epsilon_n/\sqrt{d}|\boldsymbol{Y}\}$$

$$\begin{aligned}
&\leq Pr\{\|L-M\|_{\max} \geq K_1\delta_n/4|\boldsymbol{Y}\}+ \\
&\quad Pr\{\|M-\tilde{M}\|_{\max} \geq K_1\delta_n/4|\boldsymbol{Y}\}+ \\
&\quad Pr\{\|\tilde{M}-\tilde{L}\|_{\max} \geq K_1\delta_n/4|\boldsymbol{Y}\}, \qquad (3.18)
\end{aligned}$$

where M and $\tilde{M}$ are $p \times p$ lower triangular matrices defined by

$$M_{\cdot i}^{\geq} = \begin{bmatrix} L_{ii} \\ -(\overleftarrow{S}^{>i})^{-1}\overleftarrow{S}_{\cdot i}^{>}L_{ii} \end{bmatrix}, \tilde{M}_{\cdot i}^{\geq} = \begin{bmatrix} L_{ii} \\ -(\tilde{\Sigma}^{>i})^{-1}\tilde{\Sigma}_{\cdot i}^{>}L_{ii} \end{bmatrix},$$

for $1 \leq i \leq p$, and the rest entries in M and $\tilde{M}$ are 0, i.e., $M_{ji} = \tilde{M}_{ji} = 0$ for $(i,j) \notin E, 1 \leq i < j \leq p$. We will deal with the three expressions on the RHS of (3.18) separately.

First, by the union-sum inequality, we have

$$\begin{aligned}
& Pr\{\|L-M\|_{\max} \geq K_1\delta_n/4|\boldsymbol{Y}\} \\
= \quad & Pr\left\{\max_{(i,j)\in E, 1\leq i<j\leq p} |L_{ji}-M_{ji}| \geq K_1\delta_n/4|\boldsymbol{Y}\right\} \\
= \quad & Pr\left\{\bigcup_{(i,j)\in E, 1\leq i<j\leq p} \{|L_{ji}-M_{ji}| \geq K_1\delta_n/4\}|\boldsymbol{Y}\right\} \\
\leq \quad & pd \max_{(i,j)\in E, 1\leq i<j\leq p} Pr\{|L_{ji}-M_{ji}| \geq K_1\delta_n/4|\boldsymbol{Y}\}, \qquad (3.19)
\end{aligned}$$

where the last inequality follows from the fact that there are at most d neighbors for each vertex i, $1 \leq i \leq p$. Let $c = \max_{1\leq i\leq j\leq p}(\overleftarrow{S}^{\geq i})_{jj}^{-1}$, where $(\overleftarrow{S}^{\geq i})_{jj}^{-1}, j \geq i$ stands for the diagonal element of $(\overleftarrow{S}^{\geq i})^{-1}$ corresponding to the vertex j. Then it follows from (3.19) that

$$\begin{aligned}
& Pr\{\|L-M\|_{\max} \geq K_1\delta_n/4|\boldsymbol{Y}\} \\
\leq \quad & pd \max_{(i,j)\in E, 1\leq i<j\leq p} Pr\left\{\frac{\sqrt{n}|L_{ji}-M_{ji}|}{\sqrt{(\overleftarrow{S}^{\geq i})_{jj}^{-1}}} \geq \frac{\sqrt{n}K_1\delta_n}{4\sqrt{c}}\middle|\boldsymbol{Y}\right\} \\
= \quad & pd \max_{(i,j)\in E, 1\leq i<j\leq p} E\left[Pr\left\{|z_{ji}| \geq \frac{\sqrt{n}K_1\delta_n}{4\sqrt{c}}\middle|\boldsymbol{Y}, L_{ii}\right\}\right], \qquad (3.20)
\end{aligned}$$

where $z_{ji} = \frac{\sqrt{n}(L_{ji}-M_{ji})}{\sqrt{(\overleftarrow{S}^{\geq i})_{jj}^{-1}}}\Big|(\boldsymbol{Y}, L_{ii}) \sim N(0,1)$ by Lemma 3.3 and the expectation in the last equality is taken with respect to the posterior distribution of L_{ii}. Hence,

$$\begin{aligned}
& Pr\{\|L-M\|_{\max} \geq K_1\delta_n/4|\boldsymbol{Y}\} \\
\leq \quad & pd \max_{(i,j)\in E, 1\leq i<j\leq p} E\left[Pr\left\{|z_{ji}| \geq \frac{\sqrt{n}K_1\delta_n}{4\sqrt{c}}\middle|\boldsymbol{Y}, L_{ii}\right\}\right] \\
= \quad & 2pd\left(1-\Phi\left(\frac{\sqrt{n}K_1\delta_n}{4\sqrt{c}}\right)\right)
\end{aligned}$$

$$
\begin{aligned}
&\leq \quad 2pd\exp\{-nK_1^2\delta_n^2/(8c)\} \\
&\leq \quad 2p^{2-K_1^2/(8c)}, \qquad (3.21)
\end{aligned}
$$

where Φ is the standard normal cdf and the second inequality follows from $1-\Phi(t)\leq\exp\{-t^2/2\}$ for $t\geq 0$. Note that for any $\eta>0$,

$$
\begin{aligned}
& \quad \bar{P}(2p^{2-K_1^2/(8c)}>\eta) \\
&\leq \quad \bar{P}(2p^{2-K_1^2/(8c)}>\eta, c<2\epsilon_0^{-1})+\bar{P}(c\geq 2\epsilon_0^{-1}) \\
&\leq \quad \bar{P}(2p^{2-K_1^2\epsilon_0/16}>\eta)+\bar{P}(c\geq 2\epsilon_0^{-1}).
\end{aligned}
$$

With K_1 chosen large enough such that $2-K_1^2\epsilon_0/16<0$, it follows from Assumption 3 that $2p^{2-K_1^2\epsilon_0/16}\to 0$ as $n\to\infty$. It now follows from (3.10) in Lemma 3.5 that

$$
2p^{2-K_1^2/(8c)}\xrightarrow{\bar{P}} 0.
$$

Thus, by (3.21), we get $Pr\{\|L-M\|_{\max}\geq K_1\delta_n/4|\boldsymbol{Y}\}\xrightarrow{\bar{P}} 0$. Hence, the first term in (3.18) has been dealt with.

We now focus on the second term in (3.18). Let $[(\overleftarrow{S}^{>i})^{-1}\overleftarrow{S}^{>}_{\cdot i}]_j$, $j>i$ be the component of $(\overleftarrow{S}^{>i})^{-1}\overleftarrow{S}^{>}_{\cdot i}$ corresponding to vertex j. Recall that $c_i=\overleftarrow{S}_{ii}-(\overleftarrow{S}^{>}_{\cdot i})^T(\overleftarrow{S}^{>i})^{-1}\overleftarrow{S}^{>}_{\cdot i}$ and $\tilde{c}_i=\tilde{\Sigma}_{ii}-(\tilde{\Sigma}^{>}_{\cdot i})^T(\tilde{\Sigma}^{>i})^{-1}\tilde{\Sigma}^{>}_{\cdot i}$. Note that

$$
\begin{aligned}
& \quad \max_{1\leq i<j\leq p,(i,j)\in E}|[(\overleftarrow{S}^{>i})^{-1}\overleftarrow{S}^{>}_{\cdot i}]_j-[(\tilde{\Sigma}^{>i})^{-1}\tilde{\Sigma}^{>}_{\cdot i}]_j| \\
&= \quad \max_{1\leq i<j\leq p,(i,j)\in E}\left|\frac{[(\overleftarrow{S}^{>i})^{-1}\overleftarrow{S}^{>}_{\cdot i}]_j}{c_i}c_i-\frac{[(\tilde{\Sigma}^{>i})^{-1}\tilde{\Sigma}^{>}_{\cdot i}]_j}{\tilde{c}_i}\tilde{c}_i\right| \\
&\leq \quad \max_{1\leq i<j\leq p,(i,j)\in E}\left|\frac{[(\overleftarrow{S}^{>i})^{-1}\overleftarrow{S}^{>}_{\cdot i}]_j}{c_i}\right||c_i-\tilde{c}_i|+ \\
& \quad \max_{1\leq i<j\leq p,(i,j)\in E}\tilde{c}_i\left|\frac{[(\overleftarrow{S}^{>i})^{-1}\overleftarrow{S}^{>}_{\cdot i}]_j}{c_i}-\frac{[(\tilde{\Sigma}^{>i})^{-1}\tilde{\Sigma}^{>}_{\cdot i}]_j}{\tilde{c}_i}\right| \\
&\leq \quad \max_{1\leq i<j\leq p,(i,j)\in E}\tilde{c}_ic_i\left|\frac{[(\overleftarrow{S}^{>i})^{-1}\overleftarrow{S}^{>}_{\cdot i}]_j}{c_i}\right|\left|\frac{1}{c_i}-\frac{1}{\tilde{c}_i}\right|+ \\
& \quad \max_{1\leq i<j\leq p,(i,j)\in E}\tilde{c}_i\left|\frac{[(\overleftarrow{S}^{>i})^{-1}\overleftarrow{S}^{>}_{\cdot i}]_j}{c_i}-\frac{[(\tilde{\Sigma}^{>i})^{-1}\tilde{\Sigma}^{>}_{\cdot i}]_j}{\tilde{c}_i}\right|. \qquad (3.22)
\end{aligned}
$$

Note that the first column of $(\overleftarrow{S}^{\geq i})^{-1}$ is $\begin{bmatrix} 1/c_i \\ -(\overleftarrow{S}^{>i})^{-1}\overleftarrow{S}^{>}_{\cdot i}/c_i \end{bmatrix}$. It follows from (3.22) that

$$
\begin{aligned}
& \max_{1\leq i<j\leq p,(i,j)\in E}|[(\overleftarrow{S}^{>i})^{-1}\overleftarrow{S}^{>}_{\cdot i}]_j-[(\tilde{\Sigma}^{>i})^{-1}\tilde{\Sigma}^{>}_{\cdot i}]_j| \\
&\leq (\max_i c_i\|(\overleftarrow{S}^{\geq i})^{-1}\|_{\max})(\max_i \tilde{c}_i\|(\overleftarrow{S}^{\geq i})^{-1}-(\tilde{\Sigma}^{\geq i})^{-1}\|_{\max})+
\end{aligned}
$$

$$\begin{aligned} &\max_i \tilde{c}_i \|(\overleftarrow{S}^{\geq i})^{-1} - (\tilde{\Sigma}^{\geq i})^{-1}\|_{\max} \\ &= (\max_i \tilde{c}_i \|(\overleftarrow{S}^{\geq i})^{-1} - (\tilde{\Sigma}^{\geq i})^{-1}\|_{\max})(1 + \max_i c_i \|(\overleftarrow{S}^{\geq i})^{-1}\|_{\max}) \end{aligned} \tag{3.23}$$

Let $M_4 = \frac{2M_2}{\epsilon_0}(8\epsilon_0^{-2} + 1)$, where M_2 is as in Lemma 3.5. Then $\frac{M_4\epsilon_0/(2M_2)-1}{2\epsilon_0^{-1}} = 4\epsilon_0^{-1} > 2\epsilon_0^{-1}$. It follows from (3.23), Lemma 3.5 and Lemma 3.6 that,

$$\begin{aligned} & \bar{P}\left(\max_{1\leq i<j\leq p,(i,j)\in E} |[(\overleftarrow{S}^{>i})^{-1}\overleftarrow{S}^{>}_{\cdot i}]_j - [(\tilde{\Sigma}^{>i})^{-1}\tilde{\Sigma}^{>}_{\cdot i}]_j| \geq M_4\delta_n\right) \\ \leq \quad & \bar{P}\left(\max_i \|(\overleftarrow{S}^{\geq i})^{-1} - (\tilde{\Sigma}^{\geq i})^{-1}\|_{\max} \geq M_2\delta_n\right) + \bar{P}\left(\max_i c_i \geq 2\epsilon_0^{-1}\right) + \\ & \bar{P}\left(\max_i \|(\overleftarrow{S}^{\geq i})^{-1}\|_{\max} \geq \frac{M_4\epsilon_0/(2M_2) - 1}{2\epsilon_0^{-1}}\right) \\ \to \quad & 0, \end{aligned} \tag{3.24}$$

as $n \to \infty$. Let $r = \max_{1\leq i<j\leq p,(i,j)\in E} |[(\overleftarrow{S}^{>i})^{-1}\overleftarrow{S}^{>}_{\cdot i}]_j - [(\tilde{\Sigma}^{>i})^{-1}\tilde{\Sigma}^{>}_{\cdot i}]_j|$. Now, by the form of M and $\tilde{M}$, $\max_i \tilde{L}_{ii} \leq (eig_p(\tilde{\Sigma}))^{1/2} \leq (\epsilon_0/2)^{-1/2}$ and (3.18), we have

$$\begin{aligned} & Pr\left\{\|M - \tilde{M}\|_{\max} \geq K_1\delta_n/4|\boldsymbol{Y}\right\} \\ \leq \quad & Pr\left\{\max_i L_{ii} \geq K_1\delta_n/(4r)|\boldsymbol{Y}\right\} \\ \leq \quad & Pr\left\{\max_i \tilde{L}_{ii} + \max_i |L_{ii} - \tilde{L}_{ii}| \geq K_1\delta_n/(4r)|\boldsymbol{Y}\right\} \\ \leq \quad & Pr\left\{\max_i |L_{ii} - \tilde{L}_{ii}| \geq K_1\delta_n/(4r) - (\epsilon_0/2)^{-1/2}|\boldsymbol{Y}\right\}. \end{aligned} \tag{3.25}$$

Note that for any $\eta > 0$,

$$\begin{aligned} & \bar{P}\left(Pr\left\{\max_i |L_{ii} - \tilde{L}_{ii}| \geq K_1\delta_n/(4r) - (2\epsilon_0)^{-1/2}|\boldsymbol{Y}\right\} > \eta\right) \\ \leq & \bar{P}\left(Pr\left\{\max_i |L_{ii} - \tilde{L}_{ii}| \geq K_1\delta_n/(4r) - (2\epsilon_0)^{-1/2}|\boldsymbol{Y}\right\} > \eta, r < M_4\delta_n\right) + \\ & \bar{P}(r \geq M_4\delta_n) \\ \leq & \bar{P}\left(Pr\left\{\max_i |L_{ii} - \tilde{L}_{ii}| \geq K_1/(4M_4) - (2\epsilon_0)^{-1/2}|\boldsymbol{Y}\right\} > \eta\right) + \bar{P}(r \geq M_4\delta_n). \end{aligned}$$

If K_1 is chosen large enough such that $K_1/(4M_4) - (2\epsilon_0)^{-1/2} > 1$, then by Lemma 3.9 and (3.24), it follows that $Pr\{\max_i |L_{ii} - \tilde{L}_{ii}| \geq K_1\delta_n/(4r) - (2\epsilon_0)^{-1/2}|\boldsymbol{Y}\} \xrightarrow{\bar{P}} 0$. By (3.25), we get that $Pr\{\|M - \tilde{M}\|_{\max} \geq K_1\delta_n/4|\boldsymbol{Y}\} \xrightarrow{\bar{P}} 0$. Hence, the second term in (3.18) has been dealt with.

For the last term in (3.18), note that $\tilde{L}^{\geq}_{\cdot i} = \begin{bmatrix} \tilde{L}_{ii} \\ -(\tilde{\Sigma}^{>i})^{-1}\tilde{\Sigma}^{>}_{\cdot i}\tilde{L}_{ii} \end{bmatrix}$ by Lemma 3.2 and $\tilde{M}^{\geq}_{\cdot i} = \begin{bmatrix} L_{ii} \\ -(\tilde{\Sigma}^{>i})^{-1}\tilde{\Sigma}^{>}_{\cdot i}L_{ii} \end{bmatrix}$. Then, by Lemma 3.6 and the fact that

$$\max_{1\leq i\leq p} \|(\tilde{\Sigma}^{\geq i})^{-1}\|_{\max} \leq \max_{1\leq i\leq p} \|(\tilde{\Sigma}^{\geq i})^{-1}\|_{(2,2)} = \max_{1\leq i\leq p} (eig_1(\tilde{\Sigma}^{\geq i}))^{-1} \leq (\epsilon_0/2)^{-1},$$

also note that

$$\begin{aligned}\max_{1\le i<j\le p,(i,j)\in E} |[(\tilde{\Sigma}^{>i})^{-1}\tilde{\Sigma}^{>}_{\cdot i}]_j| &\le \max_{1\le i<j\le p,(i,j)\in E} |[(\tilde{\Sigma}^{>i})^{-1}\tilde{\Sigma}^{>}_{\cdot i}]_j/\tilde{c}_i| \max_i \tilde{c}_i \\ &\le \max_i \|(\tilde{\Sigma}^{\ge i})^{-1}\|_{\max} \max_i \tilde{c}_i,\end{aligned}$$

we have

$$\begin{aligned}& Pr\left\{\|\tilde{M}-\tilde{L}\|_{\max} \ge K_1\delta_n/4|\boldsymbol{Y}\right\} \\ \le\ & Pr\left\{\left(1+\max_{1\le i<j\le p,(i,j)\in E}|[(\tilde{\Sigma}^{>i})^{-1}\tilde{\Sigma}^{>}_{\cdot i}]_j|\right)\max_i|L_{ii}-\tilde{L}_{ii}| \ge K_1\delta_n/4\middle|\boldsymbol{Y}\right\} \\ \le\ & Pr\left\{\left(1+\max_i\|(\tilde{\Sigma}^{\ge i})^{-1}\|_{\max}\max_i\tilde{c}_i\right)\max_i|L_{ii}-\tilde{L}_{ii}| \ge K_1\delta_n/4\middle|\boldsymbol{Y}\right\} \\ \le\ & Pr\left\{(1+4\epsilon_0^{-2})\max_i|L_{ii}-\tilde{L}_{ii}| \ge K_1\delta_n/4\middle|\boldsymbol{Y}\right\}. \qquad (3.26)\end{aligned}$$

It then follows from Lemma 3.9 that $Pr\{\|\tilde{M}-\tilde{L}\|_{\max} \ge K_1\delta_n/4|\boldsymbol{Y}\} \xrightarrow{\bar{P}} 0$ for $K_1 > 4M_5(1+4\epsilon_0^{-2})$. We have now shown that all the three terms on the RHS of (3.18) converge in $\bar{P}$-probability to zero. This completes the proof. $\square$

Appendix

Proof of Lemma 3.1. Banerjee and Ghosal [4] provide a proof of this result for banded matrices. The proof below, for general sparse matrices, is based on similar arguments, and is provided for the sake of completeness.

For any $p\times p$ matrix $A=(A_{ij})_{1\le i,j\le p}$ with at most d nonzero elements in each row and each column, $\|A\|_{(\infty,\infty)} = \max_i \sum_j |A_{ij}| \le \max_i \sqrt{d}\sqrt{\sum_j A_{ij}^2} \le \sqrt{d}\|A\|_{(2,2)}$, where the first inequality follows by Cauchy-Schwarz and the second inequality follows from that $\sqrt{\sum_j A_{ij}^2} = \|e_i^T A\|_2 \le \sup_{\|x\|_2=1}\|Ax\|_2 = \|A\|_{(2,2)}$, where the i^{th} component of e_i is 1 and the rest are 0.

Now, note that $\|A\|^2_{(2,2)} = eig_p(A^TA) \le \|A^TA\|_{(1,1)} \le \|A^T\|_{(1,1)}\|A\|_{(1,1)} = \|A\|_{(\infty,\infty)}\|A\|_{(1,1)}$, where the first inequality follows by the Gershgorin circle theorem. Thus

$$\begin{aligned}\|A\|_{(2,2)} &\le \sqrt{\|A\|_{(1,1)}}\sqrt{\|A\|_{(\infty,\infty)}} \\ &= \sqrt{\max_j\sum_i|A_{ij}|}\sqrt{\max_i\sum_j|A_{ij}|} \\ &\le (\sqrt{d}\|A\|_{\max})^2 \\ &= d\|A\|_{\max}.\end{aligned}$$

This completes the proof. $\square$

Proof of Lemma 3.2. Let R be non-negative definite matrix such that any principal submatrix of R of size $\leq d$ is positive definite. We will show that the function

$$tr(AR) - \log det(A),\ A \in P_G, \tag{3.27}$$

has a unique minimizer, denoted by $\hat{A}$. Furthermore, the Cholesky factor of $\hat{A}$, denoted by $\hat{C}$ satisfies $\hat{C}_{ij} = 0$ for $(i,j) \notin E$ or $i < j$, and

$$\hat{C}_{ii} = \sqrt{\frac{1}{R_{ii} - (R^{>}_{\cdot i})^T (R^{>i})^{-1} R^{>}_{\cdot i}}}, \quad \hat{C}^{>}_{\cdot i} = -\hat{C}_{ii}(R^{>i})^{-1} R^{>}_{\cdot i}. \tag{3.28}$$

To prove this assertion, we first note by Lemma 2.1 that the minimization problem in (3.27) is equivalent to minimizing the function $\ell(C) = tr(CC^T R) - 2\log det(C)$ where C varies over lower triangular matrices in M_G with positive diagonal entries. Now, straightforward matrix algebra and the fact that $C \in M_G$ imply that

$$\begin{aligned}
\ell(C) &= \sum_{i=1}^{p} \left(C^T_{\cdot i} R C_{\cdot i} - 2\log C_{ii} \right) \\
&= \sum_{i=1}^{p} \left((C^{\geq}_{\cdot i})^T R^{\geq i} C^{\geq}_{\cdot i} - 2\log C_{ii} \right) \\
&= \sum_{i=1}^{p} (C^{>}_{\cdot i} + C_{ii}(R^{>i})^{-1} R^{>}_{\cdot i})^T R^{>i} (C^{>}_{\cdot i} + C_{ii}(R^{>i})^{-1} R^{>}_{\cdot i}) + \\
&\quad \sum_{i=1}^{p} C^2_{ii}(R_{ii} - (R^{>}_{\cdot i})^T (R^{>i})^{-1} R^{>}_{\cdot i}) - 2\log C_{ii}.
\end{aligned} \tag{3.29}$$

It follows that the first sum in (3.29) is minimized if and only if

$$C^{>}_{\cdot i} = -C_{ii}(R^{>i})^{-1} R^{>}_{\cdot i},$$

and the second sum is minimized if and only if

$$C_{ii} = \sqrt{\frac{1}{R_{ii} - (R^{>}_{\cdot i})^T (R^{>i})^{-1} R^{>}_{\cdot i}}}.$$

Thus, the assertion in (3.28) holds. Lemma 3.2 now follows by noting that $\tilde{\Omega}$ is the unique minimizer for the function $tr(A\tilde{\Sigma}) - \log det(A)$, $A \in P_G$. □

Proof of Lemma 3.3. The fact that $L_{ji} = 0$ for $1 \leq i < j \leq p$ and $(i,j) \notin E$ follows directly from Lemma 2.1. Under the prior distribution $W_G(\delta, D)$ on Ω, the posterior distribution of Ω is $W_G(\delta + n, nS + D)$:

$$f(\Omega | G, \boldsymbol{Y}) \propto (det(\Omega))^{(\delta+n)/2} exp[-tr((nS + D)\Omega)/2]. \tag{3.30}$$

The Jacobian of the transformation $\Omega \to LL^T$ is $2^p \prod_{i=1}^p L_{ii}^{v_i+1}$, where $v_i = dim(L_{\cdot i}^{>})$, by Lemma 1 of [2]. Hence, the posterior of L is:

$$
\begin{aligned}
&f(L|G, \boldsymbol{Y}) \propto \prod_{i=1}^p L_{ii}^{v_i+1}(L_{ii}^2)^{(\delta+n)/2} exp[-tr(n\tilde{S}\Omega)/2] \\
&= \prod_{i=1}^p L_{ii}^{\delta+n+v_i+1} exp\{-n/2 \sum_{i=1}^p [(L_{\cdot i}^{>} + L_{ii}(\tilde{S}^{>i})^{-1}\tilde{S}_{\cdot i}^{>})^T \tilde{S}^{>i}(L_{\cdot i}^{>} + L_{ii}(\tilde{S}^{>i})^{-1}\tilde{S}_{\cdot i}^{>}) \\
&\quad + L_{ii}^2(\tilde{S}_{ii} - (\tilde{S}_{\cdot i}^{>})^T(\tilde{S}^{>i})^{-1}\tilde{S}_{\cdot i}^{>})]\} \\
&= \prod_{i=1}^p L_{ii}^{\delta+n+v_i+1} exp\{-nL_{ii}^2 c_i/2\} exp\{-n/2 \\
&\quad \times \sum_{i=1}^{p-1} (L_{\cdot i}^{>} + L_{ii}(\tilde{S}^{>i})^{-1}\tilde{S}_{\cdot i}^{>})^T \tilde{S}^{>i}(L_{\cdot i}^{>} + L_{ii}(\tilde{S}^{>i})^{-1}\tilde{S}_{\cdot i}^{>})\},
\end{aligned}
\tag{3.31}
$$

by a similar calculation as (3.29). It follows from (3.31) that $L_{\cdot i}^{\geq}$ are independent for $1 \leq i \leq p$ and

$$
L_{\cdot i}^{>}|L_{ii} \sim MVN(-(\tilde{S}^{>i})^{-1}\tilde{S}_{\cdot i}^{>}L_{ii}, (\tilde{S}^{>i})^{-1}/n), \quad \text{for} \quad 1 \leq i \leq p-1,
$$

$$
L_{ii}^2 \sim Gamma((n + v_i + \delta)/2 + 1, nc_i/2), \quad \text{for} \quad 1 \leq i \leq p,
$$

where $c_i = \tilde{S}_{ii} - (\tilde{S}_{\cdot i}^{>})^T(\tilde{S}^{>i})^{-1}\tilde{S}_{\cdot i}^{>}$. □

Proof of Lemma 3.4. It follows by Assumption 1 and Lemma A.3 of [7] that there exists constants m_1, m_2 and δ depending on ϵ_0 only such that for $1 \leq i, j \leq p$, we have:

$$
\bar{P}(|S_{ij} - \bar{\Sigma}_{ij}| \geq t) \leq m_1 \exp\{-m_2 n t^2\}, \, |t| \leq \delta.
$$

Hence, by the union-sum inequality,

$$
\bar{P}(\|S - \bar{\Sigma}\|_{\max} \geq t) = \bar{P}(\max_{1 \leq i,j \leq p} |S_{ij} - \bar{\Sigma}_{ij}| \geq t) \leq m_1 p^2 \exp\{-m_2 n t^2\}. \tag{3.32}
$$

Recall that $\overleftarrow{S} = S + D/n$ and note $\|D\|_{\max} \leq \|D\|_{(2,2)} \leq a$. Hence, for a large enough M_1 such that $2 - m_2 M_1^2/4 < 0$, we get

$$
\begin{aligned}
& \bar{P}(\|\overleftarrow{S} - \bar{\Sigma}\|_{\max} \geq M_1\sqrt{\log p/n}) \\
\leq\ & \bar{P}(\|S - \bar{\Sigma}\|_{\max} + \|D\|_{\max}/n \geq M_1\sqrt{\log p/n}) \\
\leq\ & \bar{P}(\|S - \bar{\Sigma}\|_{\max} + a/n \geq M_1\sqrt{\log p/n}) \\
\leq\ & \bar{P}(\|S - \bar{\Sigma}\|_{\max} \geq (M_1/2)\sqrt{\log p/n}) \\
\leq\ & m_1 p^{2 - m_2 M_1^2/4} \to 0,
\end{aligned}
$$

as $n \to \infty$. The last inequality follows from (3.32).

Note that

$$\|\tilde{\Sigma} - \bar{\Sigma}\|_{\max} \leq \|\tilde{\Sigma} - \bar{\Sigma}\|_{(2,2)} \leq \|\tilde{\Sigma}\|_{(2,2)}\|\tilde{\Omega} - \bar{\Omega}\|_{(2,2)}\|\bar{\Sigma}\|_{(2,2)} \leq 2\epsilon_0^{-2}\gamma(d).$$

Hence for $M_1 > \max\left(2\epsilon_0^{-2}, \sqrt{\frac{8}{m_2}}\right)$, we get

$$\begin{aligned}
& \bar{P}(\|\overleftarrow{S} - \tilde{\Sigma}\|_{\max} \geq M_1(\sqrt{\log p/n} + \gamma(d))) \\
\leq \quad & \bar{P}(\|\overleftarrow{S} - \bar{\Sigma}\|_{\max} + \|\tilde{\Sigma} - \bar{\Sigma}\|_{\max} \geq M_1(\sqrt{\log p/n} + \gamma(d))) \\
\leq \quad & \bar{P}(\|\overleftarrow{S} - \bar{\Sigma}\|_{\max} \geq M_1\sqrt{\log p/n}) \to 0, \qquad \square
\end{aligned}$$

Proof of Lemma 3.5. Note that for $1 \leq i \leq p$, $\overleftarrow{S}^{\geq i} - \bar{\Sigma}^{\geq i}$ can be at most size $d \times d$. Hence, $\|\overleftarrow{S}^{\geq i} - \bar{\Sigma}^{\geq i}\|_{(2,2)} \leq d\|\overleftarrow{S}^{\geq i} - \bar{\Sigma}^{\geq i}\|_{\max}$ by Lemma 3.1. It follows from Lemma 3.4 that

$$\begin{aligned}
& \bar{P}\left(\max_{1 \leq i \leq p} \|\overleftarrow{S}^{\geq i} - \bar{\Sigma}^{\geq i}\|_{(2,2)} \geq M_1 d\sqrt{\log p/n}\right) \\
\leq \quad & \bar{P}\left(\max_{1 \leq i \leq p} \|\overleftarrow{S}^{\geq i} - \bar{\Sigma}^{\geq i}\|_{\max} \geq M_1\sqrt{\log p/n}\right) \\
\leq \quad & \bar{P}(\|\overleftarrow{S} - \bar{\Sigma}\|_{\max} \geq M_1\sqrt{\log p/n}) \to 0, \qquad (3.33)
\end{aligned}$$

as $n \to \infty$. Note that

$$\max_{1 \leq i \leq p} \|\tilde{\Sigma}^{\geq i} - \bar{\Sigma}^{\geq i}\|_{(2,2)} \leq \max_{1 \leq i \leq p} \|\tilde{\Sigma}^{\geq i}\|_{(2,2)}\|\tilde{\Omega}^{\geq i} - \bar{\Omega}^{\geq i}\|_{(2,2)}\|\bar{\Sigma}^{\geq i}\|_{(2,2)} \leq 2\epsilon_0^{-2}\gamma(d).$$

Hence, by triangle inequality, and the fact that $M_1 > 2\epsilon_0^{-2}$, we get

$$\begin{aligned}
& \bar{P}\left(\max_{1 \leq i \leq p} \|\overleftarrow{S}^{\geq i} - \tilde{\Sigma}^{\geq i}\|_{(2,2)} \geq M_1(d\sqrt{\log p/n} + \gamma(d))\right) \\
\leq \quad & \bar{P}\left(\max_{1 \leq i \leq p} \|\overleftarrow{S}^{\geq i} - \bar{\Sigma}^{\geq i}\|_{(2,2)} \geq M_1 d\sqrt{\log p/n}\right) \to 0. \qquad (3.34)
\end{aligned}$$

Note that by Assumption 1, $d \leq n$ for large enough n. Hence, $(\overleftarrow{S}^{\geq i})^{-1}$ is well-defined for every $1 \leq i \leq p$ for large enough n. Also note that

$$\max_{1 \leq i \leq p} \|(\bar{\Sigma}^{\geq i})^{-1}\|_{(2,2)} = \max_{1 \leq i \leq p} (eig_1(\bar{\Sigma}^{\geq i}))^{-1} \leq (eig_1(\bar{\Sigma}))^{-1} \leq \epsilon_0^{-1},$$

and

$$\|(\overleftarrow{S}^{\geq i})^{-1} - (\bar{\Sigma}^{\geq i})^{-1}\|_{(2,2)} \leq \|(\overleftarrow{S}^{\geq i})^{-1}\|_{(2,2)}\|\overleftarrow{S}^{\geq i} - \bar{\Sigma}^{\geq i}\|_{(2,2)}\|(\bar{\Sigma}^{\geq i})^{-1}\|_{(2,2)}.$$

Thus,

$$\begin{aligned}
& \max_{1 \leq i \leq p} \|(\overleftarrow{S}^{\geq i})^{-1}\|_{(2,2)} \\
\leq \quad & \max_{1 \leq i \leq p} \|(\bar{\Sigma}^{\geq i})^{-1}\|_{(2,2)} +
\end{aligned}$$

$$\begin{aligned} & \max_{1\le i\le p} \|(\overleftarrow{S}^{\ge i})^{-1}\|_{(2,2)} \max_{1\le i\le p} \|\overleftarrow{S}^{\ge i} - \bar{\Sigma}^{\ge i}\|_{(2,2)} \max_{1\le i\le p} \|(\bar{\Sigma}^{\ge i})^{-1}\|_{(2,2)} \\ \le\ & \epsilon_0^{-1}\left(1 + \max_{1\le i\le p} \|(\overleftarrow{S}^{\ge i})^{-1}\|_{(2,2)} \max_{1\le i\le p} \|\overleftarrow{S}^{\ge i} - \bar{\Sigma}^{\ge i}\|_{(2,2)}\right). \end{aligned}$$

It follows that if $\epsilon_0^{-1} \max_{1\le i\le p} \|\overleftarrow{S}^{\ge i} - \bar{\Sigma}^{\ge i}\|_{(2,2)} < 1$, then

$$\max_{1\le i\le p} \|(\overleftarrow{S}^{\ge i})^{-1}\|_{(2,2)} \le \frac{\epsilon_0^{-1}}{1 - \epsilon_0^{-1} \max_{1\le i\le p} \|\overleftarrow{S}^{\ge i} - \bar{\Sigma}^{\ge i}\|_{(2,2)}}.$$

Hence,

$$\begin{aligned} & \bar{P}\left(\max_{1\le i\le p} \|(\overleftarrow{S}^{\ge i})^{-1}\|_{(2,2)} \ge 2\epsilon_0^{-1}\right) \\ \le\ & \bar{P}\left(\max_{1\le i\le p} \|(\overleftarrow{S}^{\ge i})^{-1}\|_{(2,2)} \ge 2\epsilon_0^{-1}, \epsilon_0^{-1} \max_{1\le i\le p} \|\overleftarrow{S}^{\ge i} - \bar{\Sigma}^{\ge i}\|_{(2,2)} < 1\right) + \\ & \bar{P}\left(\epsilon_0^{-1} \max_{1\le i\le p} \|\overleftarrow{S}^{\ge i} - \bar{\Sigma}^{\ge i}\|_{(2,2)} \ge 1\right) \\ \le\ & \bar{P}\left(\frac{\epsilon_0^{-1}}{1 - \epsilon_0^{-1} \max_{1\le i\le p} \|\overleftarrow{S}^{\ge i} - \bar{\Sigma}^{\ge i}\|_{(2,2)}} \ge 2\epsilon_0^{-1}\right) + \\ & \bar{P}\left(\epsilon_0^{-1} \max_{1\le i\le p} \|\overleftarrow{S}^{\ge i} - \bar{\Sigma}^{\ge i}\|_{(2,2)} \ge 1\right). \end{aligned}$$

Now by (3.34), we have for all large n,

$$\begin{aligned} & \bar{P}\left(\max_{1\le i\le p} \|(\overleftarrow{S}^{\ge i})^{-1}\|_{(2,2)} \ge 2\epsilon_0^{-1}\right) \\ \le\ & \bar{P}\left(\max_{1\le i\le p} \|\overleftarrow{S}^{\ge i} - \bar{\Sigma}^{\ge i}\|_{(2,2)} \ge \epsilon_0/2\right) + \\ & \bar{P}\left(\epsilon_0^{-1} \max_{1\le i\le p} \|\overleftarrow{S}^{\ge i} - \bar{\Sigma}^{\ge i}\|_{(2,2)} \ge 1\right) \\ \le\ & 2\bar{P}\left(\max_{1\le i\le p} \|\overleftarrow{S}^{\ge i} - \bar{\Sigma}^{\ge i}\|_{(2,2)} \ge M_1 d\sqrt{\log p/n}\right) \\ \to\ & 0, \end{aligned} \tag{3.35}$$

as $n \to \infty$. Also, let $M_2 = \max(2M_1/\epsilon_0^2, (2\epsilon_0^{-2})^2)$, then $\epsilon_0 M_2/(2\epsilon_0^{-1}) \ge M_1$. Note that

$$\begin{aligned} & \max_{1\le i\le p} \|(\tilde{\Sigma}^{\ge i})^{-1} - (\bar{\Sigma}^{\ge i})^{-1}\|_{(2,2)} \\ \le\ & \max_{1\le i\le p} \|(\tilde{\Sigma}^{\ge i})^{-1}\|_{(2,2)} \|\tilde{\Sigma}^{\ge i} - \bar{\Sigma}^{\ge i}\|_{(2,2)} \|(\bar{\Sigma}^{\ge i})^{-1}\|_{(2,2)} \\ \le\ & (2\epsilon_0^{-2})^2 \gamma(d) \le M_2 \gamma(d) \end{aligned} \tag{3.36}$$

Hence, it follows from (3.34), (3.35) and $\max_{1\le i\le p} \|(\bar{\Sigma}^{\ge i})^{-1}\|_{(2,2)} \le {\epsilon_0}^{-1}$ that

$$\begin{aligned}
& \bar{P}\left(\max_{1\le i\le p} \|(\overleftarrow{S}^{\ge i})^{-1} - (\tilde{\Sigma}^{\ge i})^{-1}\|_{(2,2)} \ge M_2(d\sqrt{\log p/n} + \gamma(d))\right) \\
\le\ & \bar{P}\left(\max_{1\le i\le p} \|(\overleftarrow{S}^{\ge i})^{-1} - (\bar{\Sigma}^{\ge i})^{-1}\|_{(2,2)} \ge M_2 d\sqrt{\log p/n}\right) \\
\le\ & \bar{P}\left(\max_{1\le i\le p} \|(\overleftarrow{S}^{\ge i})^{-1}\|_{(2,2)} \max_{1\le i\le p} \|\overleftarrow{S}^{\ge i} - \bar{\Sigma}^{\ge i}\|_{(2,2)} \ge \epsilon_0 M_2 d\sqrt{\log p/n}\right) \\
\le\ & \bar{P}\left(\max_{1\le i\le p} \|\overleftarrow{S}^{\ge i} - \bar{\Sigma}^{\ge i}\|_{(2,2)} \ge (\epsilon_0 M_2/(2\epsilon_0^{-1}))d\sqrt{\log p/n}\right) + \\
& \bar{P}\left(\max_{1\le i\le p} \|(\overleftarrow{S}^{\ge i})^{-1}\|_{(2,2)} \ge 2\epsilon_0^{-1}\right) \\
\to\ & 0,
\end{aligned} \tag{3.37}$$

as $n \to \infty$. ☐

Proof of Lemma 3.6. Note that the first column of $(\tilde{\Sigma}^{\ge i})^{-1}$ is $\begin{bmatrix} 1/\tilde{c}_i \\ -(\tilde{\Sigma}^{>i})^{-1}\tilde{\Sigma}^{>}_{\cdot i}/\tilde{c}_i \end{bmatrix}$. Also note that $eig_1((\tilde{\Sigma}^{\ge i})^{-1}) \le 1/\tilde{c}_i \le \|(\tilde{\Sigma}^{\ge i})^{-1}\|_{(2,2)}$. Thus

$$\min_i \tilde{c}_i = (\max_i\{1/\tilde{c}_i\})^{-1} \ge (\max_i \|(\tilde{\Sigma}^{\ge i})^{-1}\|_{(2,2)})^{-1} = \min_i eig_1(\tilde{\Sigma}^{\ge i}) \ge \epsilon_0/2,$$

and

$$\max_i \tilde{c}_i = (\min_i\{1/\tilde{c}_i\})^{-1} \le (\min_i\{eig_1((\tilde{\Sigma}^{\ge i})^{-1})\})^{-1} = \max_i \|\tilde{\Sigma}^{\ge i}\|_{(2,2)} \le 2\epsilon_0^{-1}.$$

Similarly,

$$\epsilon_0 \le \min_i \bar{c}_i \le \max_i \bar{c}_i \le \epsilon_0^{-1}.$$

Note that

$$\begin{aligned}
\max_i c_i &\le \max_i \bar{c}_i + \max_i |c_i - \bar{c}_i| \\
&\le \max_i \bar{c}_i + \max_i c_i\bar{c}_i|1/c_i - 1/\bar{c}_i| \\
&\le \epsilon_0^{-1}(1 + \max_i c_i\|(\overleftarrow{S}^{\ge i})^{-1} - (\bar{\Sigma}^{\ge i})^{-1}\|_{\max}),
\end{aligned}$$

which implies that if $\epsilon_0^{-1}\max_i \|(\overleftarrow{S}^{\ge i})^{-1} - (\bar{\Sigma}^{\ge i})^{-1}\|_{\max} < 1$,

$$\max_i c_i \le \frac{\epsilon_0^{-1}}{1 - \epsilon_0^{-1}\max_i \|(\overleftarrow{S}^{\ge i})^{-1} - (\bar{\Sigma}^{\ge i})^{-1}\|_{\max}}.$$

Hence, for all large n,

$$\begin{aligned}
& \bar{P}\left(\max_i c_i \ge 2\epsilon_0^{-1}\right) \\
\le\ & \bar{P}\left(\max_i c_i \ge 2\epsilon_0^{-1}, \epsilon_0^{-1}\max_i \|(\overleftarrow{S}^{\ge i})^{-1} - (\bar{\Sigma}^{\ge i})^{-1}\|_{\max} < 1\right) + \\
& \bar{P}\left(\epsilon_0^{-1}\max_i \|(\overleftarrow{S}^{\ge i})^{-1} - (\bar{\Sigma}^{\ge i})^{-1}\|_{\max} \ge 1\right)
\end{aligned}$$

$$\leq \quad \bar{P}\left(\frac{\epsilon_0^{-1}}{1-\epsilon_0^{-1}\max_i \|(\overleftarrow{S}^{\geq i})^{-1}-(\bar{\Sigma}^{\geq i})^{-1}\|_{\max}} \geq 2\epsilon_0^{-1}\right)+ \bar{P}\left(\epsilon_0^{-1}\max_i \|(\overleftarrow{S}^{\geq i})^{-1}-(\bar{\Sigma}^{\geq i})^{-1}\|_{\max} \geq 1\right)$$

By (3.37), it now follows that for all large n,

$$\begin{aligned}
& \bar{P}\left(\max_i c_i \geq 2\epsilon_0^{-1}\right) \\
\leq \quad & \bar{P}\left(\max_i \|(\overleftarrow{S}^{\geq i})^{-1}-(\bar{\Sigma}^{\geq i})^{-1}\|_{\max} \geq \epsilon_0/2\right)+ \\
& \bar{P}\left(\epsilon_0^{-1}\max_i \|(\overleftarrow{S}^{\geq i})^{-1}-(\bar{\Sigma}^{\geq i})^{-1}\|_{\max} \geq 1\right) \\
\leq \quad & 2\bar{P}\left(\max_i \|(\overleftarrow{S}^{\geq i})^{-1}-(\bar{\Sigma}^{\geq i})^{-1}\|_{\max} \geq M_2 d\sqrt{\log p/n}\right) \\
\to \quad & 0,
\end{aligned}$$

as $n \to \infty$. Now,

$$\begin{aligned}
\max_i |c_i - \bar{c}_i| &= \max_i c_i\bar{c}_i|1/c_i - 1/\bar{c}_i| \\
&\leq (\max_i |c_i-\bar{c}_i| + \max_i \bar{c}_i)\max_i \bar{c}_i \max_i \|(\overleftarrow{S}^{\geq i})^{-1}-(\bar{\Sigma}^{\geq i})^{-1}\|_{\max} \\
&\leq (\max_i |c_i-\bar{c}_i| + \epsilon_0^{-1})\epsilon_0^{-1}\max_i \|(\overleftarrow{S}^{\geq i})^{-1}-(\bar{\Sigma}^{\geq i})^{-1}\|_{\max},
\end{aligned}$$

which again implies that if $\epsilon_0^{-1}\max_i \|(\overleftarrow{S}^{\geq i})^{-1}-(\bar{\Sigma}^{\geq i})^{-1}\|_{\max} < 1$,

$$\max_i |c_i-\bar{c}_i| \leq \frac{\epsilon_0^{-2}\max_i \|(\overleftarrow{S}^{\geq i})^{-1}-(\bar{\Sigma}^{\geq i})^{-1}\|_{\max}}{1-\epsilon_0^{-1}\max_i \|(\overleftarrow{S}^{\geq i})^{-1}-(\bar{\Sigma}^{\geq i})^{-1}\|_{\max}}. \tag{3.38}$$

Let $M_3 = 2\epsilon_0^{-2}M_2$, then

$$\frac{M_3 d\sqrt{\log p/n}}{\epsilon_0^{-2}+M_3 d\sqrt{\log p/n}\epsilon_0^{-1}} \geq \frac{M_3 d\sqrt{\log p/n}}{2\epsilon_0^{-2}} = M_2 d\sqrt{\log p/n},$$

for all large n. It follows from (3.37) and (3.38) that for all large n,

$$\begin{aligned}
& \bar{P}\left(\max_i |c_i-\bar{c}_i| \geq M_3 d\sqrt{\log p/n}\right) \\
& \leq \bar{P}\left(\max_i |c_i-\bar{c}_i| \geq M_3 d\sqrt{\log p/n}, \epsilon_0^{-1}\max_i \|(\overleftarrow{S}^{\geq i})^{-1}-(\bar{\Sigma}^{\geq i})^{-1}\|_{\max} < 1\right) \\
& \quad + \bar{P}\left(\epsilon_0^{-1}\max_i \|(\overleftarrow{S}^{\geq i})^{-1}-(\bar{\Sigma}^{\geq i})^{-1}\|_{\max} \geq 1\right) \\
& \leq \bar{P}\left(\max_i \|(\overleftarrow{S}^{\geq i})^{-1}-(\bar{\Sigma}^{\geq i})^{-1}\|_{\max} \geq \frac{M_3 d\sqrt{\log p/n}}{\epsilon_0^{-2}+M_3 d\sqrt{\log p/n}\epsilon_0^{-1}}\right) \\
& \quad + \bar{P}\left(\epsilon_0^{-1}\max_i \|(\overleftarrow{S}^{\geq i})^{-1}-(\bar{\Sigma}^{\geq i})^{-1}\|_{\max} \geq 1\right)
\end{aligned}$$

$$\leq 2\bar{P}\left(\max_i \|(\overleftarrow{S}^{\geq i})^{-1} - (\bar{\Sigma}^{\geq i})^{-1}\|_{\max} \geq M_2 d\sqrt{\log p/n}\right)$$
$$\to 0, \tag{3.39}$$

as $n \to \infty$. Note that by (3.36),

$$\begin{aligned} \max_i |\tilde{c}_i - \bar{c}_i| &= \max_i \tilde{c}_i\bar{c}_i|1/\tilde{c}_i - 1/\bar{c}_i| \\ &\leq 2\epsilon_0^{-2}\max_i \|(\tilde{\Sigma}^{\geq i})^{-1} - (\bar{\Sigma}^{\geq i})^{-1}\|_{\max} \\ &\leq 2\epsilon_0^{-2} M_2\gamma(d) \\ &= M_3\gamma(d), \end{aligned}$$

we get

$$\begin{aligned} & \bar{P}\left(\max_i |c_i - \tilde{c}_i| \geq M_3(d\sqrt{\log p/n} + \gamma(d))\right) \\ \leq\ & \bar{P}\left(\max_i |c_i - \bar{c}_i| \geq M_3 d\sqrt{\log p/n}\right) \\ \to\ & 0. \end{aligned}$$

Also, by (3.39), and the fact that $\min_i \bar{c}_i \geq \epsilon_0$, it follows that for all large n,

$$\begin{aligned} \bar{P}(\min_i c_i \leq \epsilon_0/2) &\leq \bar{P}(\min_i \bar{c}_i - \max_i |c_i - \bar{c}_i| \leq \epsilon_0/2) \\ &\leq \bar{P}(\max_i |c_i - \bar{c}_i| \geq \epsilon_0/2) \\ &\leq \bar{P}(\max_i |c_i - \bar{c}_i| \geq M_3 d\sqrt{\log p/n}) \\ &\to 0, \end{aligned}$$

as $n \to \infty$. This completes the proof. □

Proof of Lemma 3.7. Note that $\hat{\Omega}$ is the unique minimizer of the function $tr(AS) - \log det(A)$, $A \in P_G$. Since all principal sub matrices of S with size less than or equal to n are positive definite with $\bar{P}$-probability one, it follows from (3.28) (with $R = S$) that the Cholesky factor $\hat{L}$ of $\hat{\Omega}$ satisfies

$$\hat{L}_{ii} = \sqrt{\frac{1}{S_{ii} - (S_{\cdot i}^{>})^T (S^{>i})^{-1} S_{\cdot i}^{>}}}, \quad \hat{L}_{\cdot i}^{>} = -\hat{L}_{ii}(S^{>i})^{-1} S_{\cdot i}^{>}, \ \ 1 \leq i \leq p. \tag{3.40}$$

Now, using (3.32) and repeating the proof of Lemma 3.5 verbatim, except by using S in place of $\overleftarrow{S}$, we obtain (similar to (3.11)) that

$$\bar{P}(\max_{1\leq i\leq p} \|(S^{\geq i})^{-1} - (\tilde{\Sigma}^{\geq i})^{-1}\|_{(2,2)} \geq M_2(d\sqrt{\log p/n} + \gamma(d))) \to 0, \tag{3.41}$$

for a large enough constant M_2 (not depending on n). Now, it follows by (3.40), and the fact

$$S^{\geq i} = \begin{bmatrix} S_{ii} & (S_{\cdot i}^{>})^T \\ S_{\cdot i}^{>} & S^{>i} \end{bmatrix},$$

that the first column of $(S^{\geq i})^{-1}$ is given by $(\hat{L}_{ji}\hat{L}_{ii})_{j\geq i,(i,j)\in E}$. Similarly, by Lemma 3.2, it follows that the first column of $(\tilde{\Sigma}^{\geq i})^{-1}$ is given by $(\tilde{L}_{ji}\tilde{L}_{ii})_{j\geq i,(i,j)\in E}$. By (3.41), we get that

$$\bar{P}\left(\max_{1\leq i\leq j\leq p,(i,j)\in E}|\hat{L}_{ji}\hat{L}_{ii}-\tilde{L}_{ji}\tilde{L}_{ii}|\geq M_2(d\sqrt{\log p/n}+\gamma(d))\right)\to 0, \quad (3.42)$$

as $n\to\infty$. In particular, (3.42) implies that

$$\begin{aligned} & \bar{P}(\max_{1\leq i\leq p}|\hat{L}_{ii}^2-\tilde{L}_{ii}^2|\geq M_2(d\sqrt{\log p/n}+\gamma(d))) \\ = \; & \bar{P}(\max_{1\leq i\leq p}|\hat{L}_{ii}+\tilde{L}_{ii}||\hat{L}_{ii}-\tilde{L}_{ii}|\geq M_2(d\sqrt{\log p/n}+\gamma(d))) \\ \to \; & 0, \end{aligned}$$

as $n\to\infty$. Since $\hat{L}_{ii}>0$ and $\sqrt{\epsilon_0/2}\leq\tilde{L}_{ii}\leq\sqrt{2/\epsilon_0}$ (by Lemma 3.2 and 3.6), we have

$$\bar{P}(\max_{1\leq i\leq p}|\hat{L}_{ii}-\tilde{L}_{ii}|\geq M_2(d\sqrt{\log p/n}+\gamma(d))/\sqrt{\epsilon_0/2})\to 0, \quad (3.43)$$

as $n\to\infty$, which further implies that

$$\bar{P}(\min_{1\leq i\leq p}\hat{L}_{ii}\geq\sqrt{\epsilon_0}/2)\to 1, \quad (3.44)$$

as $n\to\infty$. Note that $\|\tilde{L}\|_{\max}\leq\|\tilde{L}\|_{(2,2)}\leq\sqrt{2/\epsilon_0}$. Hence, we get, for $j>i,(i,j)\in E$,

$$\begin{aligned} |\hat{L}_{ji}\hat{L}_{ii}-\tilde{L}_{ji}\tilde{L}_{ii}| & \geq \hat{L}_{ii}|\hat{L}_{ji}-\tilde{L}_{ji}|-\tilde{L}_{ji}|\hat{L}_{ii}-\tilde{L}_{ii}| \\ & \geq \hat{L}_{ii}|\hat{L}_{ji}-\tilde{L}_{ji}|-\sqrt{2/\epsilon_0}|\hat{L}_{ii}-\tilde{L}_{ii}|. \quad (3.45) \end{aligned}$$

It follows by (3.42), (3.43), (3.44) and (3.45) that

$$\bar{P}\left(\max_{1\leq i<j\leq p,(i,j)\in E}|\hat{L}_{ji}-\tilde{L}_{ji}|\geq M_2^*(d\sqrt{\log p/n}+\gamma(d))\right)\to 0, \quad (3.46)$$

as $n\to 0$, where $M_2^*=\frac{2}{\sqrt{\epsilon_0}}\left(1+\frac{2}{\epsilon_0}\right)M_2$. It follows by (3.43), (3.46) and Lemma 3.1 that

$$\bar{P}(\|\hat{L}-\tilde{L}\|_{(2,2)}\geq(\sqrt{2/\epsilon_0}M_2+M_2^*)d(d\sqrt{\log p/n}+\gamma(d)))\to 0, \quad (3.47)$$

as $n\to\infty$.
Since

$$\|\bar{\Omega}-\tilde{\Omega}\|_{(\infty,\infty)}\leq\gamma(d),$$

it suffices to show that

$$\bar{P}(\|\hat{\Omega}-\tilde{\Omega}\|_{(\infty,\infty)}\geq\tilde{K}\epsilon_n)\to 0,$$

for a large enough $\tilde{K}$.

By Lemma 3.1, we have

$$\begin{aligned}
& \bar{P}(\|\hat{\Omega} - \tilde{\Omega}\|_{(\infty,\infty)} \geq \tilde{K}\epsilon_n) \\
\leq & \bar{P}(\|\hat{\Omega} - \tilde{\Omega}\|_{(2,2)} \geq \tilde{K}\epsilon_n/\sqrt{d}) \\
= & \bar{P}(\|\hat{L}\hat{L}^T - \tilde{L}\tilde{L}^T\|_{(2,2)} \geq \tilde{K}\epsilon_n/\sqrt{d}) \\
\leq & \bar{P}(\|\hat{L}\|_{(2,2)}\|\hat{L} - \tilde{L}\|_{(2,2)} + \|\tilde{L}\|_{(2,2)}\|\hat{L} - \tilde{L}\|_{(2,2)} \geq \tilde{K}\epsilon_n/\sqrt{d}) \\
\leq & \bar{P}((2\|\tilde{L}\|_{(2,2)} + \|\hat{L} - \tilde{L}\|_{(2,2)})\|\hat{L} - \tilde{L}\|_{(2,2)} \geq \tilde{K}\epsilon_n/\sqrt{d}) \\
\leq & \bar{P}\left(\|\hat{L} - \tilde{L}\|_{(2,2)} \geq \frac{\tilde{K}\epsilon_n}{4\|\tilde{L}\|_{(2,2)}\sqrt{d}}\right) + \\
& \bar{P}\left(\|\hat{L} - \tilde{L}\|_{(2,2)} \geq \frac{\sqrt{\tilde{K}}\sqrt{\epsilon_n}}{\sqrt{2}\sqrt[4]{d}}\right).
\end{aligned} \tag{3.48}$$

Since $\epsilon_0/2 \leq eig_1(\tilde{\Omega}) \leq eig_p(\tilde{\Omega}) \leq (\epsilon_0/2)^{-1}$, it follows that $\sqrt{\epsilon_0/2} \leq \|\tilde{L}\|_{(2,2)} = \{eig_p(\tilde{\Omega})\}^{1/2} \leq (\epsilon_0/2)^{-1/2}$. Also by Assumption 3, it follows that $\sqrt{\epsilon_n}/\sqrt[4]{d} \geq \epsilon_n/\sqrt{d}$ for large enough n. In view of these observations and (3.48), it suffices to show that $\bar{P}(\|\hat{L} - \tilde{L}\|_{(2,2)} \geq \tilde{K}_1\epsilon_n/\sqrt{d}) \to 0$ as $n \to \infty$ for a large enough constant $\tilde{K}_1$, which is precisely what has been proved in (3.47). □

Proof of Lemma 3.8. The density of X is given by

$$f(x) = 2\lambda^\alpha/\Gamma(\alpha)x^{2\alpha-1}e^{-\lambda x^2},\ x > 0.$$

Using the Legendre Duplication Formula

$$\Gamma(\alpha)\Gamma(\alpha + 1/2) = \sqrt{\pi}2^{1-2\alpha}\Gamma(2\alpha)$$

and the inequalities (see [19, (1.3)] and [6, Theorem 1.5]),

$$\frac{\alpha}{\sqrt{\alpha + 1/2}} \leq \frac{\Gamma(\alpha + 1/2)}{\Gamma(\alpha)} \leq \frac{\alpha}{\sqrt{\alpha + 1/4}}, \quad \Gamma(\alpha + 1) \leq \sqrt{2\pi}\left(\frac{\alpha + 1/2}{e}\right)^{\alpha+1/2}, \tag{3.49}$$

we get

$$\mu = \frac{\Gamma(\alpha + 1/2)}{\Gamma(\alpha)\sqrt{\lambda}} \geq \frac{\alpha}{\sqrt{\alpha + 1/2}\sqrt{\lambda}}, \tag{3.50}$$

and

$$\begin{aligned}
E[\exp\{\lambda(X - \mu)^2\}] &= \int_0^\infty \frac{2\lambda^\alpha}{\Gamma(\alpha)}x^{2\alpha-1}e^{-\lambda x^2}e^{\lambda(x-\mu)^2}dx \\
&= \int_0^\infty \frac{2\lambda^\alpha}{\Gamma(\alpha)}x^{2\alpha-1}e^{-2\lambda\mu x + \lambda\mu^2}dx \\
&= \frac{2\lambda^\alpha}{\Gamma(\alpha)}\frac{\Gamma(2\alpha)}{(2\lambda\mu)^{2\alpha}}e^{\lambda\mu^2}
\end{aligned}$$

$$= \frac{2\lambda^\alpha}{(2\lambda\mu)^{2\alpha}} \frac{\Gamma(\alpha+1/2)}{\sqrt{\pi}2^{1-2\alpha}} e^{\lambda\mu^2}.$$

Now, using the second inequality in (3.49) with $\alpha - 1/2$ and (3.50), we get

$$\begin{aligned} E[\exp\{\lambda(X-\mu)^2\}] &\leq \frac{2\lambda^\alpha}{(2\alpha\sqrt{\lambda}/\sqrt{\alpha+1/2})^{2\alpha}} \frac{\sqrt{2\pi}(\alpha/e)^\alpha}{\sqrt{\pi}2^{1-2\alpha}} e^{\lambda\mu^2} \\ &\leq \frac{1}{(\alpha/\sqrt{\alpha+1/2})^{2\alpha}} \sqrt{2}(\alpha/e)^\alpha e^\alpha \\ &= \sqrt{2}(1+1/(2\alpha))^\alpha \leq \sqrt{2e}. \end{aligned}$$

Thus, by the Markov inequality,

$$Pr(|X-\mu| \geq x) = Pr(\exp\{\lambda(X-\mu)^2\} \geq e^{\lambda x^2}) \leq \sqrt{2e}\, e^{-\lambda x^2}. \qquad \square$$

Proof of Lemma 3.9. By Lemma 3.3, $L_{ii}^2|\boldsymbol{Y} \sim Gamma(\alpha_i, \lambda_i)$, where $\alpha_i = (n + v_i + \delta)/2 + 1$, $\lambda_i = nc_i/2$, where $c_i = \overleftarrow{S}_{ii} - (\overleftarrow{S}_{\cdot i}^{>})^T (\overleftarrow{S}^{>i})^{-1} \overleftarrow{S}_{\cdot i}^{>}$. By Lemma 3.2, $\tilde{L}_{ii} = 1/\sqrt{\tilde{c}_i}$, where $\tilde{c}_i = \tilde{\Sigma}_{ii} - (\tilde{\Sigma}_{\cdot i}^{>})^T (\tilde{\Sigma}^{>i})^{-1} \tilde{\Sigma}_{\cdot i}^{>}$. Let $\mu_i = E(L_{ii})$. Note that

$$\mu_i \leq \sqrt{E(L_{ii}^2)} = \sqrt{\alpha_i/\lambda_i} = \sqrt{\frac{n+v_i+\delta+2}{nc_i}} \leq (1+\sqrt{(d+\delta+2)/n})\sqrt{1/c_i}.$$

Also, by (3.50) in the proof of Lemma 3.8,

$$\mu_i \geq \frac{\alpha_i}{\sqrt{(\alpha_i+1/2)\lambda_i}} = \frac{n+v_i+\delta+2}{\sqrt{(n+v_i+\delta+3)nc_i}} \geq 1/\sqrt{c_i}.$$

Thus,

$$\max_i |\sqrt{1/c_i} - \mu_i| \leq \sqrt{(d+\delta+2)/n}\sqrt{1/\min_i c_i}.$$

Let $M_6 = 4M_3/\epsilon_0^{3/2}$. It follows that

$$\begin{aligned} &\bar{P}\left(\max_i |\tilde{L}_{ii} - \mu_i| \geq M_6(d\sqrt{\log p/n} + \gamma(d))\right) \\ \leq\ & \bar{P}\left(\max_i |\tilde{L}_{ii} - \sqrt{1/c_i}| \geq (M_6/2)(d\sqrt{\log p/n} + \gamma(d))\right) + \\ & \bar{P}\left(\max_i |\sqrt{1/c_i} - \mu_i| \geq (M_6/2)(d\sqrt{\log p/n} + \gamma(d))\right) \\ =\ & \bar{P}\left(\max_i |\frac{c_i - \tilde{c}_i}{\sqrt{c_i\tilde{c}_i}(\sqrt{c_i}+\sqrt{\tilde{c}_i})}| \geq (M_6/2)(d\sqrt{\log p/n} + \gamma(d))\right) + \\ & \bar{P}\left(\max_i |\sqrt{1/c_i} - \mu_i| \geq (M_6/2)(d\sqrt{\log p/n} + \gamma(d))\right) \\ \leq\ & \bar{P}\left(\frac{\max_i |c_i - \tilde{c}_i|}{\min_i \sqrt{c_i}(\epsilon_0/2) + \min_i c_i\sqrt{\epsilon_0/2}} \geq (M_6/2)(d\sqrt{\log p/n} + \gamma(d))\right) + \\ & \bar{P}\left(\sqrt{(d+\delta+2)/n}\sqrt{1/\min_i c_i} \geq (M_6/2)(d\sqrt{\log p/n} + \gamma(d))\right). \qquad (3.51) \end{aligned}$$

By Lemma 3.6 and Assumption 3, for all large n,

$$\begin{aligned}
& \bar{P}\left(\sqrt{(d+\delta+2)/n}\sqrt{1/\min_i c_i} \geq (M_6/2)(d\sqrt{\log p/n}+\gamma(d))\right) \\
\leq \quad & \bar{P}\left(\min_i c_i \leq \frac{4(d+\delta+2)}{M_6^2 d^2 \log p}\right) \\
\leq \quad & \bar{P}(\min_i c_i \leq \epsilon_0/2) \\
\to \quad & 0,
\end{aligned}$$

as $n \to \infty$, and

$$\begin{aligned}
& \bar{P}\left(\frac{\max_i |c_i - \tilde{c}_i|}{\min_i \sqrt{c_i}(\epsilon_0/2) + \min_i c_i \sqrt{\epsilon_0/2}} \geq (M_6/2)(d\sqrt{\log p/n}+\gamma(d))\right) \\
& \leq \bar{P}(\max_i |c_i - \tilde{c}_i| \geq M_3(d\sqrt{\log p/n}+\gamma(d))) + \\
& \quad \bar{P}(\min_i \sqrt{c_i}(\epsilon_0/2) + \min_i c_i \sqrt{\epsilon_0/2} \leq 2M_3/M_6) \\
& \leq \bar{P}(\max_i |c_i - \tilde{c}_i| \geq M_3(d\sqrt{\log p/n}+\gamma(d))) + \bar{P}(\min_i \sqrt{c_i}\sqrt{\epsilon_0} + \min_i c_i \leq \epsilon_0) \\
& \leq \bar{P}(\max_i |c_i - \tilde{c}_i| \geq M_3(d\sqrt{\log p/n}+\gamma(d))) + \bar{P}(\min_i c_i \leq \epsilon_0/2) \\
& \to 0,
\end{aligned}$$

as $n \to \infty$. It follows from (3.51) that

$$\bar{P}(\max_i |\tilde{L}_{ii} - \mu_i| \geq M_6(d\sqrt{\log p/n}+\gamma(d))) \to 0. \tag{3.52}$$

Let M_5 be large enough such that $1-\epsilon_0(M_5-M_6)^2/4 < 0$ and $M_5 > M_6$. Then, for any $\eta > 0$,

$$\begin{aligned}
& \bar{P}(Pr(\max_i |L_{ii} - \tilde{L}_{ii}| \geq M_5(d\sqrt{\log p/n}+\gamma(d))|\boldsymbol{Y}) > \eta) \\
\leq \quad & \bar{P}(Pr(\max_i |L_{ii} - \tilde{L}_{ii}| \geq M_5(d\sqrt{\log p/n}+\gamma(d))|\boldsymbol{Y}) > \eta, \\
& \max_i |\tilde{L}_{ii} - \mu_i| < M_6(d\sqrt{\log p/n}+\gamma(d))) \\
& +\bar{P}(\max_i |\tilde{L}_{ii} - \mu_i| \geq M_6(d\sqrt{\log p/n}+\gamma(d))) \\
\leq \quad & \bar{P}(Pr(\max_i |L_{ii} - \mu_i| \geq (M_5-M_6)(d\sqrt{\log p/n}+\gamma(d))|\boldsymbol{Y}) > \eta) + \\
& \bar{P}(\max_i |\tilde{L}_{ii} - \mu_i| \geq M_6(d\sqrt{\log p/n}+\gamma(d))).
\end{aligned} \tag{3.53}$$

Note that, by Lemma 3.6,

$$\begin{aligned}
& \bar{P}(Pr(\max_i |L_{ii} - \mu_i| \geq (M_5-M_6)(d\sqrt{\log p/n}+\gamma(d))|\boldsymbol{Y}) > \eta) \\
\leq \quad & \bar{P}(Pr(\max_i |L_{ii} - \mu_i| \geq (M_5-M_6)d\sqrt{\log p/n}|\boldsymbol{Y}) > \eta)
\end{aligned}$$

$$
\begin{aligned}
&\leq \quad \bar{P}(p \max_i Pr(|L_{ii} - \mu_i| \geq (M_5 - M_6) d\sqrt{\log p/n}|\boldsymbol{Y}) > \eta) \\
&\overset{(a)}{\leq} \quad \bar{P}(\sqrt{2e}p \exp\{-\min_i c_i (M_5 - M_6)^2 d^2 \log p/2\} > \eta) \\
&= \quad \bar{P}(\sqrt{2e}p^{1-\min_i c_i (M_5-M_6)^2 d^2/2} > \eta) \\
&\leq \quad \bar{P}(\sqrt{2e}p^{1-\min_i c_i (M_5-M_6)^2 d^2/2} > \eta, \min_i c_i \geq \epsilon_0/2) + \bar{P}(\min_i c_i < \epsilon_0/2) \\
&\leq \quad \bar{P}(\sqrt{2e}p^{1-\epsilon_0 (M_5-M_6)^2 d^2/4} > \eta) + \bar{P}(\min_i c_i < \epsilon_0/2) \\
&\to \quad 0, \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (3.54)
\end{aligned}
$$

as $n \to \infty$, where (a) follows from Lemma 3.8. By (3.52), (3.53) and (3.54), we get

$$Pr(\max_i |L_{ii} - \tilde{L}_{ii}| \geq M_5(d\sqrt{\log p/n} + \gamma(d))|\boldsymbol{Y}) \xrightarrow{\bar{P}} 0.$$

This completes the proof. □

References

[1] Asci, C. and Piccioni, M. (2007). Functionally Compatible Local Characteristics for the Local Specification of Priors in Graphical Models. *Scand. J. Stat.* **34** 829–840. MR2396941

[2] Atay-Kayis, A. and Massam, H. (2005). A Monte Carlo method for computing the marginal likelihood in nondecomposable Gaussian graphical models. *Biometrika* **92** 317–335. MR2201362

[3] Banerjee, O., Ghaoui, L. and d'Aspremont, A. (2008). Model Selection Through Sparse Maximum Likelihood Estimation for Multivariate Gaussian or Binary Data. *J. of Mach. Learn. Res.* **9** 485–516. MR2417243

[4] Banerjee, S. and Ghosal, S. (2014). Posterior convergence rates for estimating large precision matrices using graphical models. *Electron. J. Statist.* **8** 2111–2137. MR3273620

[5] Banerjee, S. and Ghosal, S. (2015). Bayesian structure learning in graphical models. *J. Multivariate Anal.* **136** 147–162. MR3321485

[6] Batir, N. (2008). Inequalities for the gamma function. *Archiv der Mathematik* **91** 554–563. MR2465874

[7] Bickel, P. J. and Levina, E. (2008a). Regularized estimation of large covariance matrices. *Ann. Statist.* **36** 199–227. MR2387969

[8] Bickel, P. J. and Levina, E. (2008b). Covariance regularization by thresholding. *Ann. Statist.* **36** 2577–2604. MR2485008

[9] Dawid, A. P. and Lauritzen, S. L. (1993). Hyper-Markov laws in the statistical analysis of decomposable graphical models. *Ann. Statist.* **12** 1272–1317. MR1241267

[10] Dempster, A. P. (1972). Covariance Selection. *Biometrics* **28** 157–175.

[11] Diaconis, P. and Ylvisaker, D. (1979). Conjugate priors for exponential families. *Ann. Statist.* **7** 269–281. MR0520238

[12] Friedman, J., Hastie, T. and Tibshirani, R. (2007). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **0** 1–10.

[13] Friedman, J., Hastie, T. and Tibshirani, R. (2010). Applications of the lasso and grouped lasso to the estimation of sparse graphical models. *Tech. Report, Stanford Univ.*

[14] Ghosal, S. (2000). Asymptotic normality of posterior distributions for exponential families when the number of parameters tends to infinity. *J. Multivariate Anal.* **74** 49–68. MR1790613

[15] Huang, J., Liu, N., Pourahmadi, M. and Liu, L. (2006). Covariance matrix selection and estimation via penalised normal likelihood. *Biometrika* **93** 85–98. MR2277742

[16] Johnstone, I. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961

[17] Johnstone, I. and Lu, A. (2004). Sparse Principal Components Analysis. *Tech. Report, Stanford Univ.*

[18] Karoui, N. (2007). Tracy-Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *Ann. Prob.* **2** 663–714. MR2308592

[19] Kershaw, D. (1983). Some Extensions of W. Gautschi's Inequalities for the Gamma Function. *Mathematics of Computation* **41** 607–611. MR0717706

[20] Khare, K., Oh, S. and Rajaratnam, R. (2014). A convex pseudo-likelihood framework for high dimensional partial correlation estimation, to appear in *Journal of the Royal Statistical Society B.*

[21] Lauritzen, S. L. (1996). *Graphical Models*, Oxford Univ. Press, New York. MR1419991

[22] Lenkoski, A. (2013). A Direct Sampler for G-Wishart Variates. *Stat* **2** 119–128.

[23] Letac, G. and Massam, H. (2007). Wishart Distributions For Decomposable Graphs. *Ann. Statist.* **35** 1278–1323. MR2341706

[24] Meinshausen, N. and Buehlmann, P. (2006). High dimensional graphs and variable selection with the Lasso. *Ann. Statist.* **34** 1436–1462. MR2278363

[25] Mitsakakis, N., Massam, H. and Escobar, M. (2011). A Metropolis-Hastings based method for sampling from the G-Wishart distribution in Gaussian graphical models. *Electron. J. Statist.* **5** 18–30. MR2763796

[26] Pati, D., Bhattacharya, A., Pillai, N. S. and Dunson, D. (2014). Posterior contraction in sparse Bayesian factor models for massive covariance matrices. *Ann. Statist.* **42** 1102–1130. MR3210997

[27] Peng, J., Wang, P., Zhou, N. and Zhu, J. (2009). Partial Correlation Estimation by Joint Sparse Regression Models. *J. Am. Statist. Assoc.* **5** 735–746. MR2541591

[28] Rajaratnam, B., Massam, H. and Carvalho, C. M. (2008). Flexible covariance estimation in graphical Gaussian models. *Ann. Statist.* **36** 2818–2849. MR2485014

[29] Rocha, G., Zhao, P. and Yu, B. (2008). A path following algorithm for Sparse Pseudo-Likelihood Inverse Covariance Estimation (SPLICE). *Tech. Report 759, Statistics Department, UC Berkeley.*

[30] Rose, D., Tarjan, R. and Lueker, G. (1975). Algorithmic aspects of vertex elimination on graphs, *SIAM J. Comput.* **5** 266–283. MR0408312

[31] Roverato, A. (2000). Cholesky decomposition of a hyper inverse Wishart matrix. *Biometrika* **87** 99–112. MR1766831

[32] Roverato, A. (2002). Hyper inverse Wishart distribution for non-decomposable graphs and its application to Bayesian inference for Gaussian graphical models. *Scand. J. Stat.* **29** 391–411. MR1925566

[33] Tarjan, R. and Yannakakis, M. (1984). Simple linear-time algorithm to test chordality of graphs, test acyclicity of hypergraphs, and selectivity reduce acyclic hypergraphs. *SIAM J. Comput.* **13** 566–579. MR0749707

[34] Wang, H. and Carvalho, C. (2010). Simulation of hyper-inverse Wishart distributions for non-decomposable graphs. *Electron. J. Statist.* **4** 1470–1475. MR2741209

[35] Yuan, M. and Lin, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika* **94** 19–35. MR2367824