# Approximate computation of the posterior mean in large coloured graphical Gaussian models

April 25, 2016

Keywords:

**Abstract**

# 1 Introduction

In this paper, we consider graphical Gaussian models with symmetry constraints. Symmetry restrictions in the multivariate Gaussian have a long history dating back to Wilks (1946) and the reader is referred to Lauritzen and Gehrmann (2012) for a complete list of references. In their thesis, Hylleberg et al. (1993) consider Gaussian models with conditional independence and symmetry restrictions that could be described by a group action, that is, graphical Gaussian models with group symmetry restrictions. This was followed by Andersen et al. (1995) and Madsen (2000). More recently, Højsgaard, S. and S. Lauritzen (2008) considered graphical Gaussian models with symmetry constraints not necessarily described by a group action. Rather those symmetries are described by coloured graphs $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with skeleton $G = (V, E)$ where $V$ is the set of vertices, $E$ the set of undirected edges, $\mathcal{V}$ is the set of colour classes for the vertices and $\mathcal{E}$ the set of colour classes for the edges. The models for $X = (X_v, v \in V)$ are therefore also called coloured graphical Gaussian models. The symmetry is given by the equality of certain entries either in the covariance, the correlation or the precision matrices. Coloured graphical Gaussian models have two main advantages. First they may reflect true or imposed symmetries. For example, variables could represent characteristics of twins (see Frets heads data set, Frets 1921) and therefore be assumed to be equal. Second, since conditional independences imply that certain entries of the precision matrix are set to zero, these restrictions combined with the symmetry restrictions reduce the number of free parameters and facilitate inference in high-dimensional models. Højsgaard, S. and S. Lauritzen (2008) developed algorithms to compute the maximum likelihood estimate of the covariance, correlation or precision matrix.

In Massam et al. (2015), the authors considered the coloured graphical Gaussian model with symmetry restrictions on the precision matrix and they did so from a Bayesian perspective. Given a sample $X_1, \ldots, X_n$ from the coloured graphical Gaussian model with precision matrix $K$, the distribution of $\sum_{i=1}^{n} X_i X_i^t$ is a Wishart distribution, which is a natural exponential family with canonical parameter $K$. A convenient prior for $K$ is therefore the Diaconis-Ylvisaker (1979) distribution. Massam et al. (2015) gave a method to sample from this distribution in order to estimate $K$ as the mean of the posterior Diaconis-Ylvisaker conjugate distribution. As illustrated in that paper, the estimates are accurate but it is difficult to do the computations for models with more than 30 variables. The accuracy decrease and the computational time increases with $p$, the number of variables, and it is therefore not possible to estimate the posterior mean for large models.

In order to be able to give a Bayesian estimate of the posterior mean of the precision matrix for high-dimensional models, in this paper, we consider distributed estimation. The idea behind distributed estimation is that one considers the conditional or marginal models for each vertex $v \in V$ and its neighbours, and show that some of the parameters, say $\theta_{(v)}, v \in V$, of the local models are the same as some of the components of the parameter $\theta$ of the given, i.e. global, model. Then one estimates $\theta_{(v)}$ in the local model and then combines the local estimates to give an estimate of the parameter $\theta$. In this paper, the parameter that we want to estimate is the precision matrix $\theta = K$. Distributed estimation is usually done using local conditional models. Meng et al (2014) were the first ones to use local marginal rather than conditional models in the context of graphical Gaussian models. It is easy to find the unnormalized density of the $X_{\{v\}\cup\mathcal{N}_v}$-marginal distribution of the Diaconis-Ylvisaker prior and using the results of Massam et al. (2015) we know how to sample from this distribution. Like Meng et al. (2014), we will therefore also use local marginal distributions for the DY prior but we will do so in a Bayesian context. Having obtained our Bayesian estimate of $K$ using local marginal models, we will study its asymptotic properties. We will do so first under the traditional asymptotic conditions, i.e. when the sample size $n$ goes to infinity and the number of variables $p$ is fixed and second under the double asymptotic regime when both $n$ and $p$ go to infinity.

The study of the asymptotic properties, for $p$ fixed, of the Bayesian estimate goes back to Bickel & Yahav (1969) who proved the convergence of the normalized posterior density to the appropriate normal density as well as the consistency and efficiency of the posterior mean. Since then, a lot of research has been devoted to Bayesian asymptotics for $p$ fixed. One of the most recent and well-known work in that area is Ghosal et al. (1995). For both $p$ and $n$ going to infinity, Ghosal (2000) studied the consistency and asymptotic normality, under certain conditions, of the posterior distribution of the natural parameter for an exponential family when the dimension of the parameter grows with the sample size. The author also indicates that under additional conditions, the difference between normalized posterior mean of the canonical parameter and the normalized sample mean tends to 0 in probability.

We will prove in this paper first that, for $p$ fixed, our estimate is consistent and asymptotically normally distributed, second that, for both $p$ and $n$ going to infinity, under certain boundedness conditions and for $\frac{p^{13}(\log p)^2}{\sqrt{n}} \to 0$, our estimate tends to the true value of the parameter with probability tending to 1. For $p$ fixed our arguments are classical arguments adapted to our distributed estimate. Under the double asymptotic regime, there are three main features to our proofs. For each local model, we follow an argument similar to that given in Ghosal (2000). We therefore need to verify that our DY conjugate prior and our sampling distribution satisfy the conditions and properties assumed by Ghosal (2000) in his arguments. The second feature is that in the process of proving that the norm of the difference between our estimate and the true value of the parameter tends to 0, we need to prove that asymptotically, our sampling distribution asymptotically satisfies so-called cumulant-boundedness condition. To do so, we use an argument similar to that developed by Gao and Carrol (2016) who, in turn, were inspired by the sharp deviation bounds given by Spokoiny and Zhilova (2013) for $n$ fixed. Finally, we have to combine the results obtained for each local model to show our result for the estimate of the global parameter.

In Section 2, we recall definitions and basic properties of coloured graphical models and distributed computing. We also recall the scheme for sampling from the posterior coloured G-Wishart. In Section 3, we study the asymptotic properties of our estimate when $p$ is fixed. Section 4 is the most important section with the study of the asymptotic properties under the double asymptotic regime. In Section 5 , we illustrate the efficacy of our method to obtain the posterior mean of $K$ through several simulated examples. We demonstrate numerically how our method can scale up to any dimension by looking at coloured graphical Gaussian model governed by a $10 \times 10$ grid. Even though our theoretical results indicate that we need to have $\frac{p^{13}(\log p)^2}{\sqrt{n}} \to 0$, in practice, we found that we obtain very accurate results for $n \approx$????. Section 6 contains proffs of some ancillary results.

# 2 Preliminaries

## 2.1 Colored graphical models

Suppose $X_1, X_2, \ldots, X_n$ be independent and identically distributed $p$-dimensional random variables following a multivariate normal distribution $N_p(0, \Sigma)$. Let $K = \Sigma^{-1}$ be the inverse of the covariance matrix and $G = (V, E)$ be an undirected graph where $V = \{1, 2, \ldots, p\}$ and $E$ are the sets of vertices and edges, respectively. For any $U$, a subset of $V$, define $E_U$ as the set of all edges in $E$ with both endpoints in $U$. For $X = (X_v, v \in V)$, we say that the distribution of $X$ is Markov with respect to $G$ if $X_i \perp X_j | X_{V \setminus \{i,j\}}$, where $i \neq j$. Such models for $X$ are called graphical Gaussian models. Since conditional independence of the variables $X_i$ and $X_j$ is equivalent to $K_{ij} = 0$, if we denote $P_G$ as the cone of positive definite matrices with zero $(i, j)$ entry whenever the edge $(i, j)$ does not belong to $E$, then the graphical Gaussian model Markov with respect to $G$ can be represented as

$$\mathcal{N}_G = \{N(0, \Sigma) | K \in P_G\}. \tag{1}$$ Markov

Højsgaard & Lauritzen [2008] introduced the colored graphical Gaussian models with additional symmetry on $K$. Let $\mathcal{V}= \{V_1, V_2, \ldots, V_T\}$ from a partition of $V$ and $\mathcal{E}= \{E_1, E_2, \ldots, E_S\}$ from a partition of $E$. If all the vertices belonging to an element $V_i$ of $\mathcal{V}$ have the same color, we say $\mathcal{V}$ is a colouring of $V$. Similarly if all the edges belonging to an element $E_i$ of $\mathcal{E}$ have the same color, we say that $\mathcal{E}$ is a coloring of $E$. We call $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ a colored graph. Furthermore, if the model (1) is imposed with the following additional restrictions

(a) if $m$ is a vertex class in $\mathcal{V}$, then for all $i \in m$, $K_{ii}$ are equal, and

(b) if $s$ is an edge class in $\mathcal{E}$, then for all $(i, j) \in s$, $K_{ij}$ are equal,

then the model is defined as a coloured graphical Gaussian model RCON$(\mathcal{V}, \mathcal{E})$ and denoted as

$$\mathcal{N}_{\mathcal{G}} = \{N(0, \Sigma) | K \in P_{\mathcal{G}}\}$$

where $P_{\mathcal{G}}$ is the cone of positive symmetric matrix with zero and coloured constraints.

We operate within a Bayesian framework. The prior for $K$ will be the colored $G$-Wishart with density

$$\pi(K|\delta, D) = \frac{1}{I_G(\delta, D)} (\det K)^{(\delta-2)/2} \exp\{-\frac{1}{2} tr(KD) 1_{K \in P_{\mathcal{G}}},$$

where $\delta > 0$ and $D$, a symmetric positive definite $p \times p$ matrix, are the hyper parameters of the prior distribution and $I_{\mathcal{G}}(\delta, D)$ is the normalizing constant, namely,

$$I_{\mathcal{G}}(\delta, D) = \int_{P_{\mathcal{G}}} (\det K)^{(\delta-2)/2} \exp\{-\frac{1}{2}tr(KD)dK.$$

In the previous expression, $tr(\cdot)$ represents the trace and $\det K$ represents the determinant of a matrix $K$.

Massam et al. [2015] proposed a sampling scheme for the colored $G$-Wishart distribution. This sampling method is based on the Metropolis - Hastings algorithm and the Choleskey decomposition of matrices. The authors first consider the Choleskey decomposition of $D^{-1}$ and $K \in P_{\mathcal{G}}$. Write $D^{-1} = Q^tQ$, $K = \Phi^t\Phi$ and $\psi = \Phi Q^{-1}$, where $Q = (Q_{ij})_{1\leq i\leq j\leq p}$ and $\Phi = (\Phi_{ij})_{1\leq i\leq j\leq p}$ are upper triangular matrices with real positive diagonal entries.

We are interested in the posterior mean of $K$ as an estimator of $K$.

## 2.2 Local relaxed marginal model

Massam et al. [2015] developed a Metropolis-Hastings (MH) algorithm to sample from the posterior colored $G$-Wishart and obtain the estimate of the posterior mean of $K$. However, for a large colored graph, the algorithm has a computational challenge because of the matrix completion step. The main purpose of the current work is to circumvent those computational difficulties. In order to do that, we employ local computation approach similar to that has been done for the MLE in large graphical Gaussian models (see Meng et al. [2014]). In what follows, we describe the developed approach.

For a given vertex $i \in V$, define the set of immediate neighbors of vertex $i$ as $I_i = \{j|(i,j) \in E\}$. Consider two types of estimators: the one-hop estimator and two-hop estimator. For the one-hop estimator, let $N_i = \{i\} \cup I_i$, while $N_i = \{i\} \cup I_i \cup \{k|(k,j) \in E$ and $j \in I_i\}$ for the two-hop estimator. Consider the local marginal model over $X^i = X_{N_i}$, this is a Gaussian model with precision matrix denoted by $K^i$. Then

$$K^i = (\Sigma_{N_i,N_i})^{-1} = K_{N_i,N_i} - K_{N_i,V\setminus N_i}[K_{V\setminus N_i,V\setminus N_i}]^{-1}K_{V\setminus N_i,N_i}. \qquad (2)$$ inverse

Define the buffer set $B_i = \{j|j \in N_i$ and $I_j \cap (V\setminus N_i) \neq \emptyset\}$ and the protected set $\mathcal{P}_i = N_i\setminus B_i$. Due to the Markov Property $X_{\mathcal{P}_i} \perp X_{V\setminus N_i}|X_{B_i}$, we have

$$K^i_{\mathcal{P}_i,V\setminus N_i} = 0. \qquad (3)$$ conditional

By decomposing $N_i$ into $\mathcal{P}_i$ and $B_i$, the equation (2) becomes

$$\begin{aligned}
&\begin{pmatrix} K^i_{\mathcal{P}_i,\mathcal{P}_i} & K^i_{\mathcal{P}_i,B_i} \\ K^i_{B_i,\mathcal{P}_i} & K^i_{B_i,B_i} \end{pmatrix} \\
= &\begin{pmatrix} K_{\mathcal{P}_i,\mathcal{P}_i} & K_{\mathcal{P}_i,B_i} \\ K_{B_i,\mathcal{P}_i} & K_{B_i,B_i} \end{pmatrix} - \begin{pmatrix} K_{\mathcal{P}_i,V\backslash N_i} \\ K_{B_i,V\backslash N_i} \end{pmatrix} (K_{V\backslash N_i,V\backslash N_i})^{-1} \begin{pmatrix} K_{V\backslash N_i,\mathcal{P}_i} & K_{V\backslash N_i,B_i} \end{pmatrix} \\
= &\begin{pmatrix} K_{\mathcal{P}_i,\mathcal{P}_i} & K_{\mathcal{P}_i,B_i} \\ K_{B_i,\mathcal{P}_i} & K_{B_i,B_i} \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ 0 & K_{B_i,V\backslash N_i}(K_{V\backslash N_i,V\backslash N_i})^{-1}K_{V\backslash N_i,B_i} \end{pmatrix}
\end{aligned}$$

where the 0's in the matrix above follows from the identity (3). Therefore, we obtain the following relationships

$$K^i_{\mathcal{P}_i,\mathcal{P}_i} = K_{\mathcal{P}_i,\mathcal{P}_i}, \qquad K^i_{\mathcal{P}_i,B_i} = K_{\mathcal{P}_i,B_i} \tag{4}$$ relax1

$$K^i_{B_i,B_i} = K_{B_i,B_i} - K_{B_i,V\backslash N_i}(K_{V\backslash N_i,V\backslash N_i})^{-1}K_{V\backslash N_i,B_i}. \tag{5}$$ relax2

which shows the local parameters of $K^i$ indexed by $(\mathcal{P}_i, \mathcal{P}_i)$ and $(\mathcal{P}_i, B_i)$ are equal to the global ones and the local parameters of $K^i$ indexed by $(B_i, B_i)$ are modified by $K_{B_i,V\backslash N_i}(K_{V\backslash N_i,V\backslash N_i})^{-1}K_{V\backslash N_i,B_i}$. Based on these observations, the local relaxed local model can be defined as follows. First, a relaxed edge set $R_i$ is defined as $R_i = E_{N_i} \cup \{B_i \times B_i\}$ and the local relaxed undirected graph is defined as $G_i = (N_i, R_i)$. According to (4) and (5), the relaxed zero and colored constraints on $K^i$ require

(a) if $j \in N_i \backslash B_i$ and $j \in V_k$, $k = 1, 2, \ldots, T$, then the entries $K^i_{jj}$ are equal,

(b) if $j \in B_i$, then the entries $K^i_{jj}$ are different from each other,

(c) if $(j, h) \in R_i \backslash \{B_i \times B_i\}$ and $(j, h) \in E_k$, $k = 1, 2, \ldots, S$, then the entries $K^i_{jh}$ are equal,

(d) if $(j, h) \in \{B_i \times B_i\}$, then the entries $K^i_{jh}$ are different from each other,

(e) for $j, h \in N_i$, if $(j, h) \notin R_i$, then the entries $K^i_{jh} = 0$.

The local relaxed colored graph is denoted by $\mathcal{G}_i = (\mathcal{V}_i, \mathcal{E}_i)$ where $\mathcal{V}_i$ is the coloring of vertex set $N_i$ and $\mathcal{E}_i$ is the coloring of edge set $R_i$. We use the notation $A^t$ for the transpose matrix of $A$. In each local $\mathcal{G}_i$, $i \in V$, we use the method proposed by Massam et al. [2015] to obtain the Bayesian estimator $\tilde{K}^i$, the posterior mean of $K^i$ with prior distribution the colored $G$-Wishart. The authors first express the density of the colored $G$-Wishart in terms of the Cholesky components of $K^i$ scaled by $D^i$. Then they consider the Cholesky components of $(D^i)^{-1}$ and $K^i$, written as $(D^i)^{-1} = (Q^i)^t Q^i$ and $K^i = (\Phi^i)^t \Phi^i$. Let $\Psi^i = \Phi^i (Q^i)^{-1}$. Finally, they use the MH algorithm to obtain the samples of $\Psi^i$ and so the samples of $K^i$. The sample mean of $K^i$ will be used to estimate $K^i$. In the global model, denote $\theta_{V_k}$, $k = 1, 2, \ldots, T$, as the common value of $K_{jj}$ for all $(j, j) \in V_k$ and denote $\theta_{E_k}$, $k = 1, 2, \ldots, S$, as the common value of $K_{jh}$ for all $(j, h) \in E_k$. Define the global parameter as $\theta = (\theta_{V_1}, \theta_{V_2}, \ldots, \theta_{V_T}, \theta_{E_1}, \theta_{E_2}, \ldots, \theta_{E_S})^t$ and the corresponding composite estimator as $\tilde{\theta} = (\tilde{\theta}_{V_1}, \tilde{\theta}_{V_2}, \ldots, \tilde{\theta}_{V_T}, \tilde{\theta}_{E_1}, \tilde{\theta}_{E_2}, \ldots, \tilde{\theta}_{E_S})^t$. The true value of $\theta$ is denoted by $\theta_0$. In each local model $\mathcal{G}_i$, $i \in V$, we define the local parameter as $\theta^i = (\theta^i_1, \theta^i_2, \ldots, \theta^i_{S_i})^t$ and the corresponding local estimator as $\tilde{\theta}^i = (\tilde{\theta}^i_1, \tilde{\theta}^i_2, \ldots, \tilde{\theta}^i_{S_i})^t$. The true value of $\theta^i$ is denoted by $\theta^i_0$. Furthermore, we introduce a $\sum\limits_{i=1}^{p} S_i$ dimensional vector

$$\bar{\theta} = ((\tilde{\theta}^1)^t, (\tilde{\theta}^2)^t, \ldots, (\tilde{\theta}^p)^t)^t \tag{6}$$

thetabar

and its true value $\bar{\theta}_0$. After obtaining the local estimators, a composite estimate of $\tilde{\theta}$ can be constructed by extracting the non-zero estimators of $K^i_{ii}$ and $K^i_{ij}$, $j \in I_i$, in $\tilde{\theta}^i$. Therefore, our composite estimate $\tilde{\theta}$ is defined as

$$\tilde{\theta}_{V_k} = \frac{1}{|V_k|} \sum_{i \in V_k} \sum_{j=1}^{S_i} \tilde{\theta}^i_j \mathbf{1}_{\theta^i_j = \theta_{V_k}} = g_{V_k}(\bar{\theta}), \qquad k = 1, 2, \ldots, T,$$

and

$$\tilde{\theta}_{E_k} = \frac{1}{2|E_k|} \sum_{i \in G_k} \sum_{j=1}^{S_i} \tilde{\theta}^i_j \mathbf{1}_{\theta^i_j = \theta_{E_k}} = g_{E_k}(\bar{\theta}), \qquad k = 1, 2, \ldots, S,$$

where $G_k = \{i | \exists h \in N_i, (i, h) \in E_k\}$ and $\mathbf{1}_A$ is a indicator function of the set $A$. Define $g(\bar{\theta}) = (g_{V_1}(\bar{\theta}), g_{V_2}(\bar{\theta}), \ldots, g_{V_T}(\bar{\theta}), g_{E_1}(\bar{\theta}), g_{V_2}(\bar{\theta}), \ldots, g_{V_S}(\bar{\theta}))^t = \tilde{\theta}$. The first order derivative of $g(\bar{\theta})$ is a $(S+T) \times (\sum\limits_{i=1}^{p} S_i)$ matrix with the following expression

$$\frac{\partial g(\bar{\theta})}{\partial (\bar{\theta})^t} = \begin{pmatrix} \frac{\partial g_{V_1}(\bar{\theta})}{\partial \theta^1_1} & \cdots & \frac{\partial g_{V_1}(\bar{\theta})}{\partial \theta^1_{S_1}} & \frac{\partial g_{V_1}(\bar{\theta})}{\partial \theta^2_1} & \cdots & \frac{\partial g_{V_1}(\bar{\theta})}{\partial \theta^2_{S_2}} & \frac{\partial g_{V_1}(\bar{\theta})}{\partial \theta^p_1} & \cdots & \frac{\partial g_{V_1}(\bar{\theta})}{\partial \theta^p_{S_p}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_{V_T}(\bar{\theta})}{\partial \theta^1_1} & \cdots & \frac{\partial g_{V_T}(\bar{\theta})}{\partial \theta^1_{S_1}} & \frac{\partial g_{V_T}(\bar{\theta})}{\partial \theta^2_1} & \cdots & \frac{\partial g_{V_T}(\bar{\theta})}{\partial \theta^2_{S_2}} & \frac{\partial g_{V_T}(\bar{\theta})}{\partial \theta^p_1} & \cdots & \frac{\partial g_{V_T}(\bar{\theta})}{\partial \theta^p_{S_p}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & & \ddots & \vdots \\ \frac{\partial g_{E_1}(\bar{\theta})}{\partial \theta^1_1} & \cdots & \frac{\partial g_{E_1}(\bar{\theta})}{\partial \theta^1_{S_1}} & \frac{\partial g_{E_1}(\bar{\theta})}{\partial \theta^2_1} & \cdots & \frac{\partial g_{E_1}(\bar{\theta})}{\partial \theta^2_{S_2}} & \frac{\partial g_{E_1}(\bar{\theta})}{\partial \theta^p_1} & \cdots & \frac{\partial g_{E_1}(\bar{\theta})}{\partial \theta^p_{S_p}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_{E_S}(\bar{\theta})}{\partial \theta^1_1} & \cdots & \frac{\partial g_{E_S}(\bar{\theta})}{\partial \theta^1_{S_1}} & \frac{\partial g_{E_S}(\bar{\theta})}{\partial \theta^2_1} & \cdots & \frac{\partial g_{E_S}(\bar{\theta})}{\partial \theta^2_{S_2}} & \frac{\partial g_{E_S}(\bar{\theta})}{\partial \theta^p_1} & \cdots & \frac{\partial g_{E_S}(\bar{\theta})}{\partial \theta^p_{S_p}} \end{pmatrix}$$

where $\frac{\partial g_{V_k}(\bar{\theta})}{\partial \theta^i_j} = \frac{1}{|V_k|}$ if $\theta^i_j = \theta_{V_k}$ and $i \in V_k$, $\frac{\partial g_{E_k}(\bar{\theta})}{\partial \theta^i_j} = \frac{1}{2|E_k|}$ if $\theta^i_j = \theta_{E_k}$ and $i \in G_k$.

## 3 The asymptotic property of the Bayesian estimator $\tilde{\theta}$ when $p$ is fixed and $n \to \infty$

Let $\xrightarrow{\pounds}$ and $\xrightarrow{p}$ denote the convergence in distribution and in probability, respectively. In each local model corresponding to the vertex $i \in V$, let $L^i(\theta^i)$ and $l^i(\theta^i)$ denote the likelihood and log likelihood, respectively. The Fisher information is denoted by $I(\theta^i) = E_{\theta^i}[\frac{\partial}{\partial\theta^i}\log L(\theta^i|X^i)(\frac{\partial}{\partial\theta^i}\log L(\theta^i|X^i))^t]$. Define a $S_i$ dimensional vector $U_{ij} = \frac{1}{\sqrt{n}}I^{-1}(\theta_0^i)\frac{\partial l^i(\theta^i|X_j^i)}{\partial\theta^i}|_{\theta^i=\theta_0^i}$ for $j = 1,\ldots,n$ and $i = 1,\ldots,p$, a $\sum_{i=1}^{p} S_i$ dimensional vector $U_j = (U_{1j}^t, U_{2j}^t, \ldots, U_{pj}^t)^t$ and $\bar{G} = nCov(U_1)$. The following Theorem 3.1 shows that the global estimator has the property of asymptotic normality when the number of parameters $p$ is fixed and the sample size $n$ goes to infinity.

⟨pfix⟩

**Theorem 3.1** *Let $\theta_0$, $\tilde{\theta}$ and $\bar{G}$ be defined above. Then*

$$\sqrt{n}(\tilde{\theta} - \theta_0) \xrightarrow{\pounds} N(0, A) \qquad as \qquad n \to \infty$$

*where* $A = \frac{\partial g(\bar{\theta})}{\partial\bar{\theta}^t}\bar{G}(\frac{\partial g(\bar{\theta})}{\partial\bar{\theta}^t})^t$.

**Proof.** For any $i \in V$, we have that $\sqrt{n}(\tilde{\theta}^i - \theta_0^i) = \sqrt{n}(\tilde{\theta}^i - T^i) + \sqrt{n}(T^i - \theta_0^i)$ where $T^i = \theta_0^i + \frac{1}{n}I^{-1}(\theta_0^i)\frac{\partial l^i(\theta^i)}{\partial\theta^i}|_{\theta^i=\theta_0^i}$. It then follows from Theorem 8.3 in Lehmann & Casella [1998] that $\sqrt{n}(\tilde{\theta}^i - T^i) \xrightarrow{p} 0$ and $\sqrt{n}(T^i - \theta_0^i) \xrightarrow{\pounds} N(0, I^{-1}(\theta_0^i))$ as $n \to \infty$. Furthermore, we have

$$\sqrt{n}(T^i - \theta_0^i) = \frac{1}{\sqrt{n}}I^{-1}(\theta_0^i)\frac{\partial l^i(\theta^i)}{\partial\theta^i}|_{\theta^i=\theta_0^i} = \frac{1}{\sqrt{n}}I^{-1}(\theta_0^i)\sum_{j=1}^{n}\frac{\partial l^i(\theta^i|X_j^i)}{\partial\theta^i}|_{\theta^i=\theta_0^i} = \sum_{j=1}^{n}U_{ij}.$$

Since $p$ is fixed, then $\sqrt{n}(\bar{\theta} - \bar{\theta}_0)$ has the same limiting distribution as $\sum_{j=1}^{n} U_j$ as $n \to \infty$, where $\bar{\theta}$ is defined as in (6). Clearly the $U_j, j = 1, 2, \ldots, n$, are i.i.d. We are going to compute their mean and variance. First, since $E[U_{ij}] = E[\frac{1}{\sqrt{n}}I^{-1}(\theta_0^i)\frac{\partial l^i(\theta^i|X_j^i)}{\partial\theta^i}|_{\theta^i=\theta_0^i}] = 0$, it follows that $E[U_j] = 0$ for $j = 1, 2, \cdots, n$. Second, we will compute the $(\sum_{i=1}^{p} S_i) \times (\sum_{i=1}^{p} S_i)$ covariance matrix $Cov(U_1)$ with $(i,k)$ entry

$$\begin{aligned} Cov(U_{i1}, U_{k1}^t) &= Cov(\frac{1}{\sqrt{n}}I^{-1}(\theta_0^i)\frac{\partial l^i(\theta^i|X_1^i)}{\partial\theta^i}|_{\theta^i=\theta_0^i}, (\frac{1}{\sqrt{n}}I^{-1}(\theta_0^k)\frac{\partial l^k(\theta^k|X_1^k)}{\partial\theta^k}|_{\theta^k=\theta_0^k})^t) \\ &= \frac{1}{n}I^{-1}(\theta_0^i)Cov(\frac{\partial l^i(\theta^i|X_1^i)}{\partial\theta^i}|_{\theta^i=\theta_0^i}, (\frac{\partial l^k(\theta^k|X_1^k)}{\partial\theta^k}|_{\theta^k=\theta_0^k})^t)I^{-1}(\theta_0^k) \\ &= \frac{1}{n}I^{-1}(\theta_0^i)E[\frac{\partial l^i(\theta^i|X_1^i)}{\partial\theta^i}|_{\theta^i=\theta_0^i}(\frac{\partial l^k(\theta^k|X_1^k)}{\partial\theta^k}|_{\theta^k=\theta_0^k})^t]I^{-1}(\theta_0^k). \qquad (7) \end{aligned}$$

variance

For each $j$, $j = 1, 2, \ldots, S_i$, let $\delta^i_j$ be the $S_i \times S_i$ matrix with entries $(\delta^i_j)_{hl} = 1$ if $K^i_{hl} = \theta^i_j$ and 0 otherwise. We rewrite the $(l, h)$ entry of $K^i$ as $K^i_{l_j h_j}$ if $K^i_{lh} = \theta^i_j$ and denote $\tau^i_j, j = 1, 2, \cdots, S_j$, as the numbers of $K^i_{l_j h_j}$ and $(X^i_1)_{l_j}(X^i_1)_{h_j}$. Therefore,

$$\theta^i_0 = (\frac{1}{|\tau^i_1|} tr(\delta^i_j K^i_0), \frac{1}{|\tau^i_2|} tr(\delta^i_2 K^i_0), \cdots, \frac{1}{|\tau^i_{S_i}|} tr(\delta^i_{S_j} K^i_0))^t. \tag{8}$$ k and theta

Since $X^i$ has a multivariate normal distribution $N(0, (K^i_0)^{-1})$, we have

$$\frac{\partial l^i(\theta^i|X^i_1)}{\partial \theta^i_j}|_{\theta^i=\theta^i_0} = \frac{1}{2} tr(\delta^i_j (K^i_0)^{-1}) - \frac{1}{2} tr(\delta^i_j X^i_1 (X^i_1)^t).$$

Therefore, the $(q, m)$ entry of $E\big[\frac{\partial l^i(\theta^i|X^i_1)}{\partial \theta^i}|_{\theta^i=\theta^i_0} (\frac{\partial l^k(\theta^k|X^k_1)}{\partial \theta^k}|_{\theta^k=\theta^k_0})^t\big]$ in (7) is

$$\begin{aligned}
& E\Big[\frac{\partial l^i(\theta^i|X^i_1)}{\partial \theta^i_q}|_{\theta^i=\theta^i_0} \frac{\partial l^k(\theta^k|X^k_1)}{\partial \theta^k_m}|_{\theta^k=\theta^k_0}\Big] \\
= & \frac{1}{4} tr(\delta^i_q \Sigma^i_0) \times tr(\delta^k_m \Sigma^k_0) - \frac{1}{4} tr(\delta^i_q \Sigma^i_0) \times tr(\delta^k_m E[X^k_1 (X^k_1)^t]) \\
& - \frac{1}{4} tr(\delta^k_m \Sigma^k_0) \times tr(\delta^i_q E[X^i_1 (X^i_1)^t]) + \frac{1}{4} E[tr(\delta^i_q X^i_1 (X^i_1)^t) \times tr(\delta^k_m X^k_1 (X^k_1)^t)] \\
= & \frac{1}{4} tr(\delta^i_q \Sigma^i_0) \times tr(\delta^k_m \Sigma^k_0) - \frac{1}{4} tr(\delta^i_q \Sigma^i_0) \times \tau^k_m E[(X^k_1)_{1_m} (X^k_1)_{h_m}] \\
& - \frac{1}{4} tr(\delta^k_m \Sigma^k_0) \times \tau^i_q E[(X^i_1)_{l_q} (X^i_1)_{h_q}] + \frac{1}{4} \tau^i_q \tau^k_m E[(X^k_1)_{l_m} (X^k_1)_{h_m} (X^i_1)_{l_q} (X^i_1)_{h_q}]
\end{aligned}$$

where $\Sigma^i_0 = (K^i_0)^{-1}$ and $\Sigma^k_0 = (K^k_0)^{-1}$. According to Isserlis' Theorem (see Lemma 6.1 in the Appendix), we have that

$$\begin{aligned}
& E[(X^k_1)_{l_m} (X^k_1)_{h_m} (X^i_1)_{l_q} (X^i_1)_{h_q}] \\
= & E[(X^k_1)_{l_m} (X^k_1)_{h_m}] E[(X^i_1)_{l_q} (X^i_1)_{h_q}] + E[(X^k_1)_{l_m} (X^i_1)_{l_q}] E[(X^k_1)_{h_m} (X^i_1)_{h_q}] \\
& + E[(X^k_1)_{l_m} (X^i_1)_{h_q}] E[(X^k_1)_{h_m} (X^i_1)_{l_q}] \\
= & (\Sigma_0)_{l_m h_m} (\Sigma_0)_{l_q h_q} + (\Sigma_0)_{l_m l_q} (\Sigma_0)_{h_m h_q} + (\Sigma_0)_{l_m h_q} (\Sigma_0)_{h_m l_q}
\end{aligned}$$

Therefore,

$$\begin{aligned}
& E\Big[\frac{\partial l^i(\theta^i|X^i_1)}{\partial \theta^i_q}|_{\theta^i=\theta^i_0} \frac{\partial l^k(\theta^k|X^k_1)}{\partial \theta^k_m}|_{\theta^k=\theta^k_0}\Big] \\
= & \frac{1}{4} tr(\delta^i_q \Sigma^i_0) \times tr(\delta^k_m \Sigma^k_0) - \frac{1}{4} tr(\delta^i_q \Sigma^i_0) \times \tau^k_m (\Sigma^k_0)_{l_m h_m} - \frac{1}{4} tr(\delta^k_m \Sigma^k_0) \times \tau^i_q (\Sigma^i_0)_{l_q h_q} \\
& + \frac{1}{4} \tau^i_q \tau^k_m [(\Sigma_0)_{l_m h_m} (\Sigma_0)_{l_q h_q} + (\Sigma_0)_{l_m l_q} (\Sigma_0)_{h_m h_q} + (\Sigma_0)_{l_m h_q} (\Sigma_0)_{h_m l_q}]
\end{aligned} \tag{9}$$ well

By the Multivariate Central Limit Theorem, we have $\sqrt{n}(\bar{\theta}-\bar{\theta}_0) \xrightarrow{\pounds} N(0,\bar{G})$ as $n \to \infty$, where $\bar{G} = nCov(U_1)$ and each entry of $nCov(U_1)$ is well-defined in (9). Finally, using the Delta method, since $\tilde{\theta} = g(\bar{\theta})$, we have that

$$\sqrt{n}(\tilde{\theta}-\theta_0) \xrightarrow{\pounds} N(0,A)$$

where $A = \frac{\partial g(\bar{\theta})}{\partial \theta^t}\bar{G}(\frac{\partial g(\bar{\theta})}{\partial \theta^t})^T$. ■

We now want to give a result similar to Theorem 3.1 above but with the MLE replacing the posterior mean. We consider the same local models that we have used to compute our composite posterior mean $\tilde{\theta}$. Instead of evaluating $\tilde{\theta}^i$ for each model, we compute the local MLE $\hat{\theta}^i$ of $\theta^i$ and obtain a composite MLE, which is denoted by $\hat{\theta}$.

?⟨covariance⟩?

**Theorem 3.2** *Let $\hat{\theta}$ be the composite MLE. Then*

$$\sqrt{n}(\hat{\theta}-\theta_0) \xrightarrow{\pounds} N(0,A) \qquad as \qquad n \to 0$$

*where $A$ was defined as in Theorem 3.1 above.*

**Proof.** For any $i \in V$, we use the well known result for MLE as follows

$$\sqrt{n}(\hat{\theta}^i-\theta_0^i) = \frac{1}{\sqrt{n}} I^{-1}(\theta_0^i)\sum_{j=1}^{n}\frac{\partial l^i(\theta^i|X_j^i)}{\partial \theta^i}\Big|_{\theta^i=\theta_0^i} + R^i \tag{10}$$ MLE

where $R^i \xrightarrow{p} 0$ as $n \to \infty$. Comparing identity (10) with (7) in Theorem 3.1, it is easy to get the desired result. ■

We thus see that the composite Bayesian estimator $\tilde{\theta}$ has the same limiting distribution as the composite MLE $\hat{\theta}$.

# 4 Properties of the Bayesian estimator under the double asymptotic regimes $p \to \infty$ and $n \to \infty$

In this section, we study the consistency of the global estimator $\tilde{\theta}$ when both $p$ and $n$ go to infinity. For a vector $x = (x_1, x_2, \ldots, x_p)$, let $||x||$ stand for its Euclidean norm $(\sum_{i=1}^{p} x_i^2)^{1/2}$. For a square $p \times p$ matrix $A$, let $||A||$ stand for its operator norm defined by $\sup\{||Ax|| : ||x|| \leq 1\}$, $||A||_F$ stand for its Frobenius norm defined by $||A||_F = (\sum_{j=1}^{p}\sum_{k=1}^{p}|a_{jk}|^2)^{\frac{1}{2}}$, and denote $\lambda(A)$, $\lambda_{min}(A)$ and $\lambda_{max}(A)$ as the eigenvalues, the smallest eigenvalues and largest eigenvalues of $A$, respectively. The vector obtained by stacking columnwise the entries of $A$ is denoted by $vec(A)$. Let $I_p$ be the identity matrix with $p$ dimension. In the local model $\mathcal{G}_i$, $i \in V$, we write the density of $X_j^i$, $j = 1, 2, \ldots, n$, as

$$f(X_j^i; K^i) \quad = \quad \frac{(\det K^i)^{\frac{1}{2}} \exp\big\{ -\frac{1}{2} tr(K^i X_j^i (X_j^i)^t)\big\}}{(2\pi)^{\frac{p_i}{2}}} \mathbf{1}_{K^i \in P_{\mathcal{G}_i}} \tag{11}$$

density

where $p_i = |N_i|$. The normalized and non-normalized local colored $G$-Wishart distribution of $K^i$ is denoted by

$$\pi^i(K^i|\delta^i, D^i) = \frac{1}{I^i_{\mathcal{G}_i}(\delta^i, D^i)} (\det K^i)^{(\delta-2)/2} \exp\{-\frac{1}{2} tr(K^i D^i)\} \mathbf{1}_{K^i \in P_{\mathcal{G}_i}},$$

where $I^i_{\mathcal{G}_i}(\delta^i, D^i)$ is the normalizing constant for the local model, and

$$\pi_0^i(K^i|\delta^i, D^i) = (\det K^i)^{(\delta-2)/2} \exp\{-\frac{1}{2} tr(K^i D^i)\} \mathbf{1}_{K^i \in P_{\mathcal{G}_i}},$$

respectively. In order to obtain our results, we will follow the argument similar to that of Ghosal [2000] who gives the asymptotic distribution of posterior mean when both the dimension $p$ of the model and the sample size $n$ go to $\infty$. Ghosal consider $x$ from a natural exponential family written under the form

$$f(x; \theta) = \exp[x^t \theta - \psi(\theta)]$$

where $x$ is the canonical statistic, $\theta$ is the canonical parameter and $\psi(\theta)$ is the cumulant generating function. To follow the notations of Ghosal [2000], we define an $S_i$ dimensional vector

$$Y_j^i = -\frac{1}{2}(tr(\delta_1^i X_j^i (X_j^i)^t), tr(\delta_2^i X_j^i (X_j^i)^t), \ldots, tr(\delta_{S_i}^i X_j^i (X_j^i)^t))^t, \tag{12}$$

sufficient

where $\delta_1^i, \delta_2^i, \cdots, \delta_{S_i}^i$ are indicator matrices for each color class. We reparameterize the equation (11) in terms of the canonical parameter vector $\theta^i$ and the canonical statistic $Y_j^i$. We get that

$$\begin{aligned} f(X_j^i; K^i) \quad &= \quad \exp\big[-\frac{1}{2} tr(K^i X_j^i (X_j^i)^t) + \frac{1}{2} \log\det K^i - \frac{\pi}{2} \log 2\pi\big] \mathbf{1}_{K^i \in P_{\mathcal{G}_i}} \\ &= \quad \exp\big[\sum_{k=1}^{S_i} \frac{1}{\tau_k} tr(\delta_k^i K^i) tr(-\frac{1}{2} \delta_k^i X_j^i (X_j^i)^t) + \frac{1}{2} \log\det K^i - \frac{\pi}{2} \log 2\pi\big] \mathbf{1}_{K^i \in P_{\mathcal{G}_i}} \\ &= \quad \exp\big[\sum_{k=1}^{S_i} \theta_k^i (Y_j^i)_k + \frac{1}{2} \log\det K^i - \frac{\pi}{2} \log 2\pi\big] \mathbf{1}_{K^i \in P_{\mathcal{G}_i}} \\ &= \quad \exp\big[(Y_j^i)^t \theta^i - \psi(\theta^i)\big] \end{aligned} \tag{13}$$

density1

where $\psi(\theta^i) = (-\frac{1}{2} \log(\det K^i) + \frac{p_i}{2} \log(2\pi)) \mathbf{1}_{K^i \in P_{\mathcal{G}_i}}$ is the cumulant generating function, $(Y_j^i)_k$ represents the $k$th element of $Y_j^i$ and the third equality above is obtained by definitions (8) and (12). Since (13) represents a natural exponential family distribution, we have that

$$\mu^i = \psi'(\theta_0^i) \qquad \text{and} \qquad F^i = \psi''(\theta_0^i) \tag{14}$$ F

are the mean vector and the covariance matrix of $Y_j^i$, respectively. Note that $F^i$ is also the Fisher information for the canonical parameter $\theta^i$ and is positive semidefinite. Let $J^i$ be a square root of $F^i$, i.e. $J^i(J^i)^t = F^i$. Let

$$V_j^i = (J^i)^{-1}(Y_j^i - E_{\theta^i}(Y_j^i)) \tag{15}$$ stand

be the standardized version of the canonical statistic. Following Ghosal [2000], for any constant $c$, $c > 0$, we define

$$B_{1n}^i(c) = \sup\{E_{\theta^i}|a^t V_j^i|^3 : a \in \mathbb{R}^{S_i}, ||a|| = 1, ||J^i(\theta^i - \theta_0^i)||^2 \leq \frac{cS_i}{n}\}$$

and

$$B_{2n}^i(c) = \sup\{E_{\theta^i}|a^t V_j^i|^4 : a \in \mathbb{R}^{S_i}, ||a|| = 1, ||J^i(\theta^i - \theta_0^i)||^2 \leq \frac{cS_i}{n}\}.$$

Define

$$u^i = \sqrt{n}J^i(\theta^i - \theta_0^i),$$

then $\theta^i = \theta_0^i + n^{-1/2}(J^i)^{-1}u^i$. Therefore, the likelihood ratio can be written as a function of $u^i$ in the following form

$$Z_n^i(u^i) = \frac{\prod_{j=1}^n f(Y_j^i;\theta^i)}{\prod_{j=1}^n f(Y_j^i;\theta_0^i)} = \exp\{\sqrt{n}(\bar{Y}^i)^t(J^i)^{-1}u^i - n[\psi(\theta_0^i + n^{-\frac{1}{2}}(J^i)^{-1}u^i) - \psi(\theta_0^i)]\} \tag{16}$$ ?znu?

where $\bar{Y}^i = \frac{1}{n}\sum_{j=1}^n Y_j^i$. Furthermore, we denote

$$\Delta_n^i = \sqrt{n}(J^i)^{-1}(\bar{Y}^i - \mu^i). \tag{17}$$ delta

The following three conditions will be assumed.

(1) The orders of $\log p$ and $\log n$ are the same, i.e. $\frac{\log p}{\log n} \to \lambda > 0$ as $n \to \infty$ (see similar condition in Ghosal [2000]).

(2) There exists two constants $\kappa_1$ and $\kappa_2$ such that $0 < \kappa_1 \leq \lambda_{min}(K_0) < \lambda_{max}(K_0) \leq \kappa_2 < \infty$.

(3) For any $i \in V$, the numbers of the entries $K_{jk}^i$ in the same color class is bounded, i.e. there exists a constant $\zeta$ such that for any $1 \leq l \leq S_i$, $\max_i |\{K_{jk}^i | K_{jk}^i = \theta_l^i\}| \leq \zeta$.

(4) The dimension $p$ of the matrix $K$ is allowed to grow slowly with $n$ satisfying $\frac{p^{13}(\log p)^2}{n^{\frac{1}{2}}} \to 0$.

Remark: Condition (2) implies $0 < \frac{1}{\kappa_2} \le \lambda_{min}(\Sigma_0) < \lambda_{max}(\Sigma_0) \le \frac{1}{\kappa_1} < \infty$. By the interlacing property of eigenvalues, we have that $0 < \frac{1}{\kappa_2} \le \lambda_{min}((\Sigma_0)_{N_i,N_i}) < \lambda_{max}((\Sigma_0)_{N_i,N_i}) \le \frac{1}{\kappa_1} < \infty$ where $N_i$ is defined as in section 2.2. Therefore, $0 < \kappa_1 \le \lambda_{min}((\Sigma_0)_{N_i,N_i})^{-1} < \lambda_{max}((\Sigma_0)_{N_i,N_i})^{-1} \le \kappa_2 < \infty$. By the definition (2), for any $i \in V$, we have $0 < \kappa_1 \le \lambda_{min}(K_0^i) < \lambda_{max}(K_0^i) \le \kappa_2 < \infty$.

We will show in Proposition 4.2 that if Condition (2) is satisfied, then every entry of $K_0^i$ is bounded. We will use this property throughout the paper. Our aim in this section is to prove that under the assumption (1), (2) (3) and (4), for $n^{-\frac{1}{2}}p^{13}(\log p)^{\frac{1}{2}} \to 0$, the composite estimator $\tilde{\theta}$ tends to $\theta_0$ in Frobenius norm with probability tending to 1. We state this now in Theorem 4.1.

⟨converge⟩

**Theorem 4.1** *Under Conditions (1)-(4), there exists a constant $\bar{c}$ such that*

$$P(||\tilde{\theta} - \theta_0|| \le \Big\{\kappa_2^2\big[\frac{3a^2p^3}{n} + \frac{p^2(p+1)}{2n}A(p,n,\bar{c})\big]\Big\}^{\frac{1}{2}}) \ge 1 - 10.4\exp\{-\frac{1}{6}p^2\log p + \log p\}.$$

*where*

$$A(p,n,\bar{c}) = c_5(\bar{c})\frac{p^{13}\log p}{\sqrt{n}} + \exp[-c_9(\bar{c})p^2\log p] + \frac{2}{\sqrt{2\pi}}p^{-3a^2+2} + \sqrt{3a^2}\frac{2}{\sqrt{2\pi}}p^{-3a^2+1},$$

*and $c_5(\bar{c})$ and $c_9(\bar{c})$ are constants.*

**Proof.** In this theorem, we study the consistency of $\tilde{\theta}$ in the context of Frobenius norm. In order to do this, first, we evaluate the norm $||\tilde{\theta}^i - \theta_0^i||^2$ in each local model. Since $||\sqrt{n}J^i(\tilde{\theta}^i - \theta_0^i)||^2 = n(\tilde{\theta}^i - \theta_0^i)^t(J^i)^tJ^i(\tilde{\theta}^i - \theta_0^i) \ge n\lambda_{min}(F^i)||\tilde{\theta}^i - \theta_0^i||^2$, we obtain

$$\begin{aligned} ||\tilde{\theta}^i - \theta_0^i||^2 &\le \frac{1}{n\lambda_{min}(F^i)}||\sqrt{n}J^i(\tilde{\theta}^i - \theta_0^i)||^2 \\ &= \frac{1}{n\lambda_{min}(F^i)}||\Delta_n^i + \int u^i[\pi_*^i(u^i) - \phi(u^i;\Delta_n^i, I_{S_i})]du^i||^2 \qquad \text{by Lemma 4.8} \\ &\le \frac{1}{n\lambda_{min}(F^i)}\Big(||\Delta_n^i||^2 + ||\int u^i[\pi_*^i(u^i) - \phi(u^i;\Delta_n^i, I_{S_i})]du^i||^2\Big) \qquad (18) \end{aligned}$$

T1

where $\phi(\cdot; v, \Sigma)$ stands for the multivariate normal density of $N(v, \Sigma)$ and $\pi_*^i(u^i)$ stands for the posterior distribution of $u^i$. Next, for every element of the vector $\int u^i[\pi_*^i(u^i) - \phi(u^i;\Delta_n^i, I_{S_i})]du^i$ in (18), we will find out it's upper bound. Denote $u^i = (u_1^i, u_2^i, \cdots, u_{Si}^i)^t$. Then for the $j$th element of $\int u^i[\pi_*^i(u^i) - \phi(u^i;\Delta_n^i, I_{S_i})]du^i$, we have that

$$\int u_j^i[\pi_*^i(u^i) - \phi(u^i;\Delta_n^i, I_{S_i})]du^i \le \int ||u^i||[\pi_*^i(u^i) - \phi(u^i;\Delta_n^i, I_{S_i})]du^i \qquad (19)$$

absolute

Set $\bar{c} = \max\{c, C\}$, where $c$ is defined in Lemma 4.6 and $C$ is defined in Lemma 4.7. According to the argument of Theorem 2.3 in Ghosal [2000], the integral $\int ||u^i||[\pi_*^i(u^i) - \phi(u^i; \Delta_n^i, I_{S_i})]du^i$ in (19) can be bounded by a sum of three integrands as follows.

$$\begin{aligned}
&\int ||u^i|| \times |\pi_*^i(u^i) - \phi(u^i; \Delta_n^i, I_{S^i})|du^i \\
\leq\ & \frac{\int_{||u^i||^2 \leq \bar{c}M(p)} ||u^i|| \cdot |\pi^i(\theta_0^i + n^{-\frac{1}{2}}(J^i)^{-1}u^i)Z_n^i(u^i) - \pi^i(\theta_0^i)\tilde{Z}_n^i(u^i)|du^i}{\int \pi^i(\theta_0^i)\tilde{Z}_n^i(u^i)du^i} \\
&+[\int \pi^i(\theta_0^i)\tilde{Z}_n^i(u^i)du^i]^{-1} \int_{||u^i||^2 > \bar{c}M(p)} ||u^i|| \cdot Z_n^i(u^i)\pi^i(\theta_0 + n^{-\frac{1}{2}}(J^i)^{-1}u^i)du^i \\
&+ \int_{||u^i||^2 > \bar{c}M(p)} ||u^i||\phi(u^i; \Delta_n^i, I_{S_i})du^i,
\end{aligned}$$

By Lemmas 4.5, 4.6 and 4.7, every element of $\int u^i[\pi_*^i(u^i) - \phi(u^i; \Delta_n^i, I_{S_i})]du^i$ in (18) can be bounded by

$$A(p, n, \bar{c}) = c_5(\bar{c})\frac{p^{13}\log p}{\sqrt{n}} + \exp[-c_9(\bar{c})p^2 \log p] + \frac{2}{\sqrt{2\pi}}p^{-3a^2+2} + \sqrt{3a^2}\frac{2}{\sqrt{2\pi}}p^{-3a^2+1}$$

with a probability greater than $1 - 10.4\exp\{-\frac{1}{6}p^2\}$. Consequently,

$$\int u_j^i[\pi_*^i(u^i) - \phi(u^i; \Delta_n^i, I_{S_i})]du^i \quad \leq \quad A(p, n, \bar{c}).$$

Since the dimension of $\int u^i[\pi_*^i(u^i) - \phi(u^i; \Delta_n^i, I_{S_i})]du^i$ is $S_i$, from the inequality (18), we get

$$||\tilde{\theta}^i - \theta_0^i||^2 \quad \leq \quad \frac{1}{\lambda_{min}(F^i)}\Big(\frac{||\Delta_n^i||^2}{n} + \frac{S_i}{n}A(p, n, \bar{c})\Big)$$

with a probability greater than $1 - 10.4\exp\{-\frac{1}{6}p^2\}$. Finally, we will estimate the Frobenius norm $||\tilde{\theta} - \theta_0||$ for the composite estimator $\tilde{\theta}$ in terms of $||\tilde{\theta}^i - \theta_0^i||$ from the local model. By Lemma 4.1, for any $i \in V$, $\lambda_{min}(F^i) \geq \frac{1}{\kappa_2^2}$. Therefore, we have

$$\begin{aligned}
||\tilde{\theta} - \theta_0|| &\leq ||\bar{\theta} - \bar{\theta}_0|| \\
&\leq (\sum_{i=1}^p ||\tilde{\theta}^i - \theta_0^i||^2)^{\frac{1}{2}} \qquad \text{by triangle inequality} \\
&\leq \Big\{\sum_{i=1}^p \Big[\frac{1}{\lambda_{min}(F^i)}\big(\frac{||\Delta_n^i||^2}{n} + \frac{S_i}{n}A(p, n, \bar{c})\big)\Big]\Big\}^{\frac{1}{2}} \\
&\leq \Big\{\kappa_2^2\big[\frac{p||\Delta_n^i||^2}{n} + \frac{p^2(p+1)}{2n}A(p, n, \bar{c})\big]\Big\}^{\frac{1}{2}}
\end{aligned}$$

According to Lemma 4.4, we have $||\Delta_n^i||^2 \leq 3a^2p^2$ with a probability greater than $1 - 10.4\exp\{-\frac{1}{6}p^2\}$. Therefore,

$$||\tilde{\theta} - \theta_0|| \quad \leq \quad \Big\{\kappa_2^2\big[\frac{p3a^2p^2}{n} + \frac{p^2(p+1)}{2n}A(p,n,\bar{c})\big]\Big\}^{\frac{1}{2}}$$

with a probability greater than $1 - 10.4p\exp\{-\frac{1}{6}p^2\}$ by the Bonferroni inequality. Furthermore, Condition (4) implies $A(p,n,\bar{c}) \to 0$, $\frac{p^2(p+1)}{2n} \to 0$ and $\frac{p^3}{n} \to 0$. It follows $||\tilde{\theta} - \theta_0|| \to 0$ with a probability greater than $1 - 10.4\exp\{-\frac{1}{6}p^2 + \log p\} \to 1$. ■

⟨Fbound⟩ **Proposition 4.1** *Let $F^i$ be defined in definition (14) for any $i \in V$, then there exists two constants $\rho_1$ and $\rho_2$ such that $\frac{1}{\kappa_2^2} \leq \lambda_{min}(F^i) \leq \lambda_{max}(F^i) \leq \frac{1}{\kappa_1^2}$.*

**Proof.** Let $G^i$ be the Fisher information matrix for the uncolored graphical models e.g. $G^i = \psi_u''(\theta^i)$ where $\psi_u(\theta^i) = (-\frac{1}{2}\log(\det K^i) + \frac{p_i}{2}\log(2\pi))\mathbf{1}_{K^i \in P_{G_i}}$. Let $\tau$ and $\varpi$ be the numbers of eigenvalues of $G^i$ and $F^i$. Since it is a linear projection from $G^i$ to $F^i$, then $\tau > \varpi$. Under Condition (2) and by Proposition (6.1) in Appendix, for any $l$, $1 \leq l \leq \varpi$, we have

$$\frac{1}{\kappa_2^2} \leq \min\{\frac{1}{\lambda_j(G^i)\lambda_k(G^i)}|1 \leq j,k \leq \tau\} \leq \lambda_l(F_i) \leq \max\{\frac{1}{\lambda_j(G^i)\lambda_k(G^i)}|1 \leq j,k \leq \tau\} \leq \frac{1}{\kappa_1^2}.$$

■

⟨entrybound⟩ **Proposition 4.2** *For any $i \in V$, let $K_{\alpha\beta}^{i,0}$ be the $(\alpha,\beta)$ entry of $K_0^i$. Then $|K_{\alpha\beta}^{i,0}| \leq \kappa_2$.*

**Proof.** By Condition (2), we have $\lambda_{max}(K_0^i) \leq \kappa_2$ for any $i \in V$. Therefore, $\kappa_2 - \lambda_j(K_0^i), j = 1,2,\cdots,p_i$, are the eigenvalues of $\kappa_2 I_{p_i} - K_0^i$. Since $\lambda_{max}(K_0^i) \leq \kappa_2$, then $\kappa_2 \geq \lambda_j', j = 1,2,\cdots,p_i$. It follows that $\kappa_2 I_{p_i} - K_0^i$ is a positive semidefinite matrix. Since the diagonal elements of a positive semidefinite $bI_{p_i} - K_0^i$ are all non negative, then $\kappa_2 - K_{\alpha\alpha}^{i,0} \geq 0, \alpha = 1,2,\ldots,p_i$. It follows $0 < K_{\alpha\alpha}^0 \leq \kappa_2$. Since $K_0^i$ is a positive definite matrix, then each 2 by 2 principal sub matrices

$$\begin{pmatrix} K_{\alpha\alpha}^{i,0} & K_{\alpha\beta}^{i,0} \\ K_{\beta\alpha}^{i,0} & K_{\beta\beta}^{i,0} \end{pmatrix}$$

of $K_0^i$ are positive definite. Therefore, $K_{\alpha\alpha}^{i,0}K_{\beta\beta}^{i,0} - (K_{\alpha\beta}^{i,0})^2 > 0$, from which we get $|K_{\alpha\beta}^{i,0}| < (K_{\alpha\alpha}^{i,0}K_{\beta\beta}^{i,0})^{1/2} < \kappa_2$. ■

⟨trace⟩ **Proposition 4.3** *For any $i \in V$, we have the trace of $F^i$ satisfies $tr(F^i) = O(p^2)$ and the determinant $det(F^i)$ satisfies $\log(det(F^i)) = O(p^2)$.*

**Proof.** Since $\frac{\partial^2 \psi(\theta^i)}{\partial\theta^i_j \partial\theta^i_k} = \frac{1}{2}tr(\delta^i_j \Sigma^i_0 \delta^i_k \Sigma^i_0)$, then $tr(F^i) = \frac{1}{2}\sum_{j=1}^{S_i} tr((\delta^i_j\Sigma^i_0)^2)$. Furthermore, by Condition (3), $tr(\delta^i_j\Sigma^i_0)$ is bounded. Therefore, $tr((\delta^i_j\Sigma^i_0)^2)$ is bounded. It follows

$$tr(F^i) = \frac{1}{2}\sum_{j=1}^{S_i} tr((\delta^i_j\Sigma^i_0)^2) \leq \frac{1}{2}\frac{p_i(p_i+1)}{2} tr((\delta^i_j\Sigma^i_0)^2) \leq \frac{1}{2}\frac{p(p+1)}{2} tr((\delta^i_j\Sigma^i_0)^2) = O(p^2).$$

Next, let us consider $\log(det(F^i))$. Since

$$det(F^i) = \prod_{j=1}^{S_i} \lambda_j(F^i) \leq \Big(\frac{\sum_{j=1}^{S_i}\lambda_j(F^i)}{S_i}\Big)^{S_i} = \Big(\frac{tr(F^i)}{S_i}\Big)^{S_i},$$

then

$$\log(det(F^i)) \leq S_i \log\frac{tr(F^i)}{S_i} \leq \frac{p_i(p_i+1)}{2}\log\frac{\frac{1}{2}\frac{p_i(p_i+1)}{2}tr((\delta^i_j\Sigma^i_0)^2)}{\frac{p_i(p_i+1)}{2}} = O(p^2).$$

The proposition is proved. ■

⟨prior⟩

**Proposition 4.4** *For any $i \in V$, we have $\log \pi^i_0(K^i_0) \geq -\frac{1}{2}p_i\kappa_2 + \frac{\delta-2}{2}p_i\log\kappa_1$ when $D^i = I_{p_i}$.*

**Proof.**

$$\begin{aligned}
\pi^i_0(K^i_0) &= \exp\Big\{-\frac{1}{2}tr(K^i_0 I_{p_i}) + \frac{\delta-2}{2}\log(det(K^i_0))\Big\} \\
&= \exp\Big\{-\frac{1}{2}\sum_{j=1}^{p_i}\lambda_j(K^i_0) + \frac{\delta-2}{2}\log\prod_{j=1}^{p_i}\lambda_j(K^i_0)\Big\} \\
&\geq \exp\Big\{-\frac{1}{2}p_i\kappa_2 + \frac{\delta-2}{2}p_i\log\kappa_1\Big\}.
\end{aligned}$$

■

⟨lipschiz⟩

**Proposition 4.5** *(Lipschitz continuity) For any $i \in V$ and any constant $c$, there exists a constant $M_1$ such that $|\log\pi^i(\theta^i) - \log\pi^i(\theta^i_0)| \leq M_1 p||\theta^i - \theta^i_0||$ when $||\theta^i - \theta^i_0|| \leq \sqrt{||(F^i)^{-1}||cp^2\log p/n} \to 0$.*

**Proof.** By mean value theorem, we have

$$\begin{aligned}
|\log\pi^i(\theta^i) - \log\pi^i(\theta^i_0)| &= |\log\pi^i_0(\theta^i) - \log\pi^i_0(\theta^i_0)| \\
&= |(\theta^i - \theta^i_0)^t \frac{\partial\log\pi^i_0(\theta^i)}{\partial\theta^i}|_{\theta^i=\check{\theta}^i}| \\
&\leq ||\theta^i - \theta^i_0|| \cdot ||\frac{\partial\log\pi^i_0(\theta^i)}{\partial\theta^i}|_{\theta^i=\check{\theta}^i}|| \\
&= ||\theta^i - \theta^i_0|| \cdot \sqrt{\sum_{j=1}^{S_i}\Big[-\frac{1}{2}tr(\delta^i_j D^i) + \frac{\delta-2}{2}tr(\delta^i_j\Sigma^i_0)\Big]^2}
\end{aligned}$$

where $\breve{\theta}^i$ is the point on the line segment joining $\theta^i$ and $\theta^i_0$. Since $tr(\delta^i_j D^i)$ and $tr(\delta^i_j \Sigma^i_0)$ are bounded. Therefore, there exists a constant $M_1$ such that

$$\sqrt{\sum_{j=1}^{S_i}\Big[-\frac{1}{2}tr(\delta^i_j D^i)+\frac{\delta-2}{2}tr(\delta^i_j\Sigma^i_0)\Big]^2}\leq\sqrt{\frac{p_i(1+p_i)}{2}M_1^2}\leq\sqrt{\frac{p(1+p)}{2}M_1^2}=M_1p.$$

■

⟨exp⟩

**Proposition 4.6** *For any $i \in V$, let $Y^i_j$ and $V^i_j$ be defined in* (12) *and* (15)*, respectively. Then $B^i_{1n}(c) = O(p^9)$ and $B^i_{2n}(c) = O(p^{12})$.*

**Proof.** Let $B_{\alpha\beta}$, $\alpha, \beta \in \{1, 2, \ldots, S_i\}$, be the entries of $(J^i)^{-1}$. Define $b = \max\{|B_{\alpha\beta}|; \alpha, \beta \in \{1, 2, \ldots, S_i\}\}$. Then for the vectors $Y^i_j = (Y^i_{j1}, Y^i_{j2}, \ldots, Y^i_{jS_i})^t$ and $a = (a_1, a_2, \ldots, a_{S_i})^t$, the following property holds for $h = 1, 2, 3, 4$

$$\begin{aligned}
E|a^t(J^i)^{-1}Y^i_j|^h &\leq E\Bigg[(|a_1|,|a_2|,\ldots,|a_{S_i}|)\begin{pmatrix} b\sum_{k=1}^{S_i}|Y^i_{jk}| \\ b\sum_{k=1}^{S_i}|Y^i_{jk}| \\ \vdots \\ b\sum_{k=1}^{S_i}|Y^i_{jk}|\end{pmatrix}\Bigg]^h \\
&= E\Big[(b\sum_{k=1}^{S_i}|Y^i_{jk}|)\sum_{k=1}^{S_i}|a_k|\Big]^h \\
&\leq E\Big[b(\sum_{k=1}^{S_i}|Y^i_{jk}|)\sqrt{S_i}||a||\Big]^h \quad \text{by Cauchy - Schwarz inequality} \\
&\leq b^h(S_i)^{h/2}E\Big[\sum_{k=1}^{S_i}|Y^i_{jk}|\Big]^h \\
&\leq b^h(S_i)^{h/2}E\Big[\sum_{k_1=1}^{S_i}\cdots\sum_{k_h=1}^{S_i}|Y^i_{jk_1}|\cdots|Y^i_{jk_h}|\Big]
\end{aligned}$$

Each entry of $\theta^i$ is bounded when $||J^i(\theta^i - \theta^i_0)||^2 \leq \frac{cS_i}{n} \to 0$. By Lemma 6.2 in Appendix, we have $E\Big[|Y^i_{jk_1}|\cdots|Y^i_{jk_h}|\Big]$ is bounded for $h = 1, 2, 3, 4$. Therefore, $E|a^t(J^i)^{-1}Y^i_j|^h = O(p_i^{3h})$. Similarly, $|a^t(J^i)^{-1}E(Y^i_j)|^h = O(p_i^{3h})$. Hence, we have

$$\begin{aligned}
E|a^tV^i_j|^3 &= E|a^t(J^i)^{-1}Y^i_j - a^t(J^i)^{-1}E(Y^i_j)|^3 \\
&\leq E|a^t(J^i)^{-1}Y^i_j|^3 + 3|a^t(J^i)^{-1}E(Y^i_j)|E(a^t(J^i)^{-1}Y^i_j)^2 \\
&\quad +3(a^t(J^i)^{-1}E(Y^i_j))^2E|a^t(J^i)^{-1}Y^i_j| + |(a^t(J^i)^{-1}E(Y^i_j))^3| \\
&= O(p_i^9) = O(p^9).
\end{aligned}$$

A similar argument deduces $E|a^t V_j^i|^4 = O(p^{12})$. By the definition $B_{1n}^i(c)$ and $B_{2n}^i(c)$, the desired result follows. ■

⟨boundcondition⟩

**Lemma 4.1** *Let $Y_1^i$ be defined in* (12) *and $G_Y^i(\gamma^i) = \log E(e^{(\gamma^i)^t Y_1^i})$ be the cumulant generating function of $Y_1^i$, for any $i \in V$, there exist constants $\eta$ and $C_1$ such that with $||\gamma^i|| \leq \eta$, the absolute value of all the third derivatives of $G_Y^i(\gamma^i)$ satisfy $\left|\frac{\partial^3 G_Y^i(\gamma^i)}{\partial\gamma_k^i \partial\gamma_l^i \partial\gamma_m^i}\right| \leq C_1$ for all $k, l, m$ such that $1 \leq k, l, m \leq S_i$ under Condition (2) and (3).*

**Proof.** Let $\gamma^i$ be a $S_i$ dimensional vector, by Theorem 3.2.3 in Muirhead [1982], the moment generating function of $Y_1^i$ is

$$M^i(\gamma^i) = E\{\exp[(\gamma^i)^t Y_1^i]\} = \det(I_{p_i} + T^i(\gamma^i)\Sigma_0^i)^{-\frac{1}{2}}$$

where $T^i(\gamma^i)$ is a $p_i \times p_i$ matrix with $T_{\alpha\beta}^i = \gamma_k^i$ if $K_{\alpha\beta}^i = \theta_k^i$. Therefore, the cumulant generating function $G_Y^i(\gamma^i)$ of $Y_1^i$ is given by

$$G_Y^i(\gamma^i) = \log M^i(\gamma^i) = -\frac{1}{2}\log\det(I_{p_i} + T^i(\gamma^i)\Sigma_0^i).$$

It is easy to obtain the first, second and third derivative of the cumulant generating function $G_Y^i(\gamma^i)$, which can be expressed as

$$\begin{aligned}
\frac{\partial G_Y^i(\gamma^i)}{\partial\gamma_k^i} &= -\frac{1}{2}tr\Big((I_{p_i} + T^i(\gamma^i)\Sigma_0^i)^{-1}(\delta_k^i\Sigma_0^i)\Big), \\
\frac{\partial^2 G_Y^i(\gamma^i)}{\partial\gamma_k^i\partial\gamma_l^i} &= \frac{1}{2}tr\Big(\delta_k^i\Sigma_0^i(I_{p_i} + T^i(\gamma^i)\Sigma_0^i)^{-1}(\delta_l^i\Sigma_0^i)(I_{p_i} + T^i(\gamma^i)\Sigma_0^i)^{-1}\Big) \text{ and} \\
\frac{\partial^3 G_Y^i(\gamma^i)}{\partial\gamma_k^i\partial\gamma_l^i\partial\gamma_m^i} &= -\frac{1}{2}tr\Big(\delta_k^i\Sigma_0^i(I_{p_i} + T^i(\gamma^i)\Sigma_0^i)^{-1}(\delta_m^i\Sigma_0^i)(I_{p_i} + T^i(\gamma^i)\Sigma_0^i)^{-1}(\delta_l^i\Sigma_0^i)(I_{p_i} + T^i(\gamma^i)\Sigma_0^i)^{-1} \\
&\quad + \delta_k^i\Sigma_0^i(I_{p_i} + T^i(\gamma^i)\Sigma_0^i)^{-1}(\delta_l^i\Sigma_0^i)(I_{p_i} + T^i(\gamma^i)\Sigma_0^i)^{-1}(\delta_m^i\Sigma_0^i)(I_{p_i} + T^i(\gamma^i)\Sigma_0^i)^{-1}\Big),
\end{aligned}$$

respectively. First, Condition (3) implies $\lambda_{max}(\Sigma_0^i) \leq \frac{1}{\kappa_1}$. By Proposition 4.2, the absolute value of each element of $\Sigma_0^i$ is bounded by $\frac{1}{\kappa_1}$. Next, by $\sum_{j=1}^{p} |\lambda_j(A)| \leq ||A||_F$ and $||AB|| \leq ||AB||_F \leq ||A||_F||B||$ for any two $p \times p$ symmetric matrix, we have that $|\lambda_j(T^i(\gamma^i)\Sigma_0^i)| \leq ||T^i(\gamma^i)||_F||\Sigma_0^i|| \leq \eta\frac{1}{\kappa_1}$. It implies $1 - \eta\frac{1}{\kappa_1} \leq \lambda(I_{p_i} + T^i(\gamma^i)\Sigma_0^i) \leq 1 + \eta\frac{1}{\kappa_1}$. Moreover, according to Lemma 6.3 in the Appendix, $I_{p_i} + T^i(\gamma^i)\Sigma_0^i$ is a positive definite. Therefore, by Proposition 4.2 again, the absolute value of each element of $(I_{p_i} + T^i(\gamma^i)\Sigma_0^i)^{-1}$ is bounded. Finally, combining the above results and Condition (3), for any $i \in V$, there exists a constant $C_1$ such that $|\frac{\partial^3 G_Y^i(\gamma^i)}{\partial\gamma_k^i\partial\gamma_l^i\partial\gamma_m^i}| \leq C_1$ for any $k, m, l$. ■

⟨center⟩

**Lemma 4.2** *For any $i \in V$, let $\bar{U}_1^i = (J^i)^{-1}(Y_1^i - \mu^i)$, then there exist constants $\eta$ and $C_2$ with $||\gamma^i|| \leq \eta$, the absolute value of all the third derivatives of the cumulant* *generating function* $G^i_{\bar{U}}(\gamma^i)$ *of* $\bar{U}^i_1$ *satisfy* $\left|\frac{\partial^3 G^i_{\bar{U}}(\gamma^i)}{\partial\gamma^i_k\partial\gamma^i_l\partial\gamma^i_m}\right| \leq C_2$ *for all* $k, l, m$ *such that* $1 \leq k, l, m \leq S_i$ *under Condition (2) and (3).*

**Proof.** The cumulant generating function of $\bar{U}^i_1$ is

$$\begin{aligned} G^i_{\bar{U}}(\gamma^i) &= \log E[e^{(\gamma^i)^t (J^i)^{-1}(Y^i_1-\mu^i)}] = \log E[e^{(\gamma^i)^t (J^i)^{-1}Y^i_1-(\gamma^i)^t(J^i)^{-1}\mu^i}] \\ &= \log E[e^{(\gamma^i)^t (J^i)^{-1}Y^i_1}] + \log e^{-(\gamma^i)^t(J^i)^{-1}\mu^i} \\ &= \log E[e^{(\gamma^i)^t (J^i)^{-1}Y^i_1}] - (\gamma^i)^t(J^i)^{-1}\mu^i \\ &= G^i_Y((J^i)^{-1}\gamma^i) - (\gamma^i)^t(J^i)^{-1}\mu^i \end{aligned}$$

By lemma 4.1, there exists a constant $C_2$ such that $|\frac{\partial^3 G^i_{\bar{U}}(\gamma^i)}{\partial\gamma^i_k\partial\gamma^i_l\partial\gamma^i_m}| \leq C_2$ for $||\gamma^i|| \leq \eta$. ■

In the following proof, we let $M(p) = p^2 \log p$. The following result holds for a random vector when its cumulant generating function satisfies some conditions, which was first established in Lemma 1 of Gao & Carroll [2015].

⟨Xin⟩ **Lemma 4.3** *Let* $\bar{U}^i_j = (J^i)^{-1}(Y^i_j - \mu^i)$, $j = 1, 2, \ldots, n$, *and* $\Delta^i_n = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\bar{U}^i_j$ *as defined in equation* (17). *Then for any arbitrary constant* $a$ *such that* $a^2 > 1$, *if* $\frac{C_2 p^3}{3\sqrt{n}} \leq a - 1$, *we have that the cumulant generating function* $G^i_{\Delta^i_n}(\gamma^i) = \log\left(E\{\exp[(\gamma^i)^t\Delta^i_n]\}\right) \leq a^2||\gamma^i||^2/2$ *for* $||\gamma^i|| < p$ *under condition (2) and (3).*

**Proof.** By a Taylor expansion of $G^i_{\bar{U}}(\gamma^i)$ around 0, there exists a vector $\gamma^{i,*}$ on the line segment between 0 and $\gamma^i$ such that

$$\begin{aligned} G^i_{\bar{U}}(\gamma^i) &= G^i_{\bar{U}}(0) + \sum_{k=1}^{S_i}(\frac{\partial G^i_{\bar{U}}(\gamma^i)}{\partial\gamma^i_k}|_{\gamma^i=0})\gamma^i_k + \frac{1}{2}\sum_{k=1}^{S_i}\sum_{l=1}^{S_i}(\frac{\partial^2 G^i_{\bar{U}}(\gamma^i)}{\partial\gamma^i_k\partial\gamma^i_l}|_{\gamma^i=0})\gamma^i_k\gamma^i_l \\ &+\frac{1}{6}\sum_{k=1}^{S_i}\sum_{l=1}^{S_i}\sum_{m=1}^{S_i}(\frac{\partial^3 G^i_{\bar{U}}(\gamma^i)}{\partial\gamma^i_k\partial\gamma^i_l\partial\gamma^i_m}|_{\gamma^i=\gamma^{i,*}})\gamma^i_k\gamma^i_l\gamma^i_m. \end{aligned}$$

Since $\bar{U}^i_j$ has zero mean and identity covariance matrices, then $\frac{\partial G^i_{\bar{U}}(\gamma^i)}{\partial\gamma^i_k}|_{\gamma^i=0} = 0$, $\frac{\partial^2 G^i_{\bar{U}}(\gamma^i)}{\partial\gamma^i_k\partial\gamma^i_l}|_{\gamma^i=0} = 1$ for $k = l$ and $\frac{\partial^2 G^i_{\bar{U}}(\gamma^i)}{\partial\gamma^i_k\partial\gamma^i_l}|_{\gamma^i=0} = 0$ for $k \neq l$. Furthermore, since $G^i_{\bar{U}}(0) = 0$, we have

$$G^i_{\bar{U}}(\gamma^i) = \frac{1}{2}(\gamma^i)^t\gamma^i + \frac{1}{6}\sum_{k=1}^{S_i}\sum_{l=1}^{S_i}\sum_{m=1}^{S_i}(\frac{\partial^3 G^i_{\bar{U}}(\gamma^i)}{\partial\gamma^i_k\partial\gamma^i_l\partial\gamma^i_m}|_{\gamma^i=\gamma^{i,*}})\gamma^i_k\gamma^i_l\gamma^i_m.$$

By the definition (17), we have $\Delta_n^i = \frac{1}{\sqrt{n}}\sum\limits_{j=1}^{n}\bar{U}_j^i$. Since the moment generating function of $\bar{U}_j^i$ is $\exp G_{\bar{U}}^i(\gamma^i)$, then the moment generating function of $\Delta_n^i$ is

$$\begin{aligned}
E[e^{(\gamma^i)^t\Delta_n^i}] &= E[e^{(\gamma^i)^t\frac{1}{\sqrt{n}}\sum\limits_{j=1}^{n}\bar{U}_j^i}] = \prod_{j=1}^{n} E[e^{(\frac{\gamma^i}{\sqrt{n}})^t\bar{U}_j^i}] \\
&= \prod_{j=1}^{n}\exp\Big\{\frac{1}{2}(\frac{\gamma^i}{\sqrt{n}})^t\frac{\gamma^i}{\sqrt{n}} + \frac{1}{6}\sum_{k=1}^{S_i}\sum_{l=1}^{S_i}\sum_{m=1}^{S_i}(\frac{\partial^3 G_{\bar{U}}^i(\frac{\gamma^i}{\sqrt{n}})}{\partial\gamma_k^i\partial\gamma_l^i\partial\gamma_m^i}|_{\gamma^i=\gamma^{i,*}})\frac{\gamma_k^i}{\sqrt{n}}\frac{\gamma_l^i}{\sqrt{n}}\frac{\gamma_m^i}{\sqrt{n}}\Big\} \\
&= \exp\Big\{\frac{1}{2}n(\frac{\gamma^i}{\sqrt{n}})^t\frac{\gamma^i}{\sqrt{n}} + \frac{1}{6}n\sum_{k=1}^{S_i}\sum_{l=1}^{S_i}\sum_{m=1}^{S_i}(\frac{\partial^3 G_{\bar{U}}^i(\frac{\gamma^i}{\sqrt{n}})}{\partial\gamma_k^i\partial\gamma_l^i\partial\gamma_m^i}|_{\gamma^i=\gamma^{i,*}})\frac{\gamma_k^i}{\sqrt{n}}\frac{\gamma_l^i}{\sqrt{n}}\frac{\gamma_m^i}{\sqrt{n}}\Big\} \\
&= \exp\Big\{\frac{1}{2}(\gamma^i)^t\gamma^i + \frac{1}{6}\frac{1}{\sqrt{n}}\sum_{k=1}^{S_i}\sum_{l=1}^{S_i}\sum_{m=1}^{S_i}(\frac{\partial^3 G_{\bar{U}}^i(\frac{\gamma^i}{\sqrt{n}})}{\partial\gamma_k^i\partial\gamma_l^i\partial\gamma_m^i}|_{\gamma^i=\gamma^{i,*}})\gamma_k^i\gamma_l^i\gamma_m^i\Big\}
\end{aligned}$$

Since $||\gamma^{i,*}|| < ||\gamma^i||$, we have $||\frac{\gamma^{i,*}}{\sqrt{n}}|| < ||\frac{\gamma^i}{\sqrt{n}}|| < \frac{p}{\sqrt{n}}$. Moreover, Condition (4) implies $\frac{p}{\sqrt{n}} \to 0$, we can thus choose a constant $\eta_1$ small enough such that $||\gamma^i|| \le \eta_1 \le \eta$. Therefore, by Lemma 4.2, there exists a constant $C_2$ such that $\left|\frac{\partial^3 G_{\bar{U}}^i(\gamma^i)}{\partial\gamma_k^i\partial\gamma_l^i\partial\gamma_m^i}\right| \le C_2$. It follows

$$\begin{aligned}
E[e^{(\gamma^i)^t\Delta_n^i}] &\le \exp\Big\{\frac{1}{2}(\gamma^i)^t\gamma^i + \frac{1}{6}\frac{C_2}{\sqrt{n}}\sum_{k=1}^{S_i}\sum_{l=1}^{S_i}\sum_{m=1}^{S_i}\gamma_k^i\gamma_l^i\gamma_m^i\Big\} \\
&= \exp\Big\{\frac{1}{2}(\gamma^i)^t\gamma^i\big[1 + \frac{1}{3}\frac{C_2}{\sqrt{n}}\sum_{m=1}^{S_i}\gamma_m^i\big]\Big\}.
\end{aligned}$$

Therefore, for any arbitrary constant $a$ such that $a^2 > 1$, if $\frac{1}{3}\frac{C_2}{\sqrt{n}}\sum\limits_{m=1}^{S_i}\gamma_m^i \le a^2 - 1$, then we have

$$\log E[e^{(\gamma^i)^t\eta^i}] \le a^2||\gamma^i||^2/2.$$

Actually, the inequality $\frac{1}{3}\frac{C_2}{\sqrt{n}}\sum\limits_{m=1}^{S_i}\gamma_m^i \le a^2 - 1$ holds under Condition (4). Since $||\gamma^i|| < p$, we have $|\gamma_m^i| \le ||\gamma^i|| < p$ for any $1 \le m \le S_i$. Therefore, according to Condition (4), we have

$$\frac{1}{3}\frac{C_2}{\sqrt{n}}\sum_{m=1}^{S_i}\gamma_m^i = O(\frac{p^3}{\sqrt{n}}) = o(1).$$

It implies $\frac{1}{3}\frac{C_2}{\sqrt{n}}\sum\limits_{m=1}^{S_i}\gamma_m^i \le a^2 - 1$ for any constant $a$ such that $a^2 > 1$. ∎

⟨max⟩

**Lemma 4.4** *Under Condition (2)-(4), for any $i \in V$, there exists a constant $a$, $a^2 > 1$, such that*

$$P\{||\Delta_n^i||^2 > 3a^2p^2\} \leq 10.4\exp\{-\frac{1}{6}p^2\}$$

*where $\Delta_n^i$ is defined as in* (17).

**Proof.** According to Lemma 4.3, we have

$$\log\big(E\{\exp[(\gamma^i)^t\Delta_n^i]\}\big) \leq a^2||\gamma^i||^2/2 \qquad for \qquad ||\gamma^i|| \leq p$$

where $a$ is a constant with $a^2 > 1$. Let $g = ap$ and $t_1^i = a\gamma^i$, then the subsequent inequality holds

$$\log[E\{\exp((t_1^i)^t\frac{\Delta_n^i}{a})\}] \leq ||t_1^i||^2/2 \qquad for \qquad ||t_1^i|| \leq g.$$

Next we apply the large deviation result from Corollary 3.2 in Spokoiny & Zhilova [2013]. Following the notations in Spokoiny & Zhilova [2013], we introduce $w_c^i$ satisfying the equation $\frac{w_c^i(1+w_c^i)}{[1+(w_c^i)^2]^{\frac{1}{2}}} = gS_i^{-1/2}$. Based on $w_c^i$, we define $x_c^i = 0.5S_i[(w_c^i)^2 - \log(1+(w_c^i)^2)]$. Since $g^2 = a^2p^2 > \frac{p^2+p}{2} \geq S_i$, by the arguments in Spokoiny & Zhilova [2013], we have $x_c^i > \frac{1}{4}g^2 = \frac{1}{4}a^2p^2$. Let $x = \frac{1}{6}p^2$, then $\frac{S_i}{6.6} \leq \frac{p^2+p}{2\times 6.6} < x < x_c^i$. By Corollary 3.2 in Spokoiny & Zhilova [2013], the following inequality holds

$$P(||\frac{\Delta_n^i}{a}||^2 \geq S_i + 6.6\times\frac{1}{6}p^2) \leq 2e^{-\frac{1}{6}p^2} + 8.4e^{-x_c^i},$$

which implies $P(||\frac{\Delta_n^i}{a}||^2 \geq 3p^2) \leq 10.4e^{-\frac{1}{6}p^2}$. Hence, $P(||\Delta_n^i||^2 \geq 3a^2p^2) \leq 10.4e^{-\frac{1}{6}p^2}$, which means $||\Delta_n^i||^2 = O_p(p^2)$. ■

Denote $\tilde{Z}_n^i(u^i) = \exp[(u^i)^t\Delta_n^i - \frac{1}{2}||u^i||^2]$.

⟨firsttermnorm⟩

**Lemma 4.5** *If $\frac{p^{12}(\log p)^{\frac{1}{2}}}{\sqrt{n}} \to 0$, then for any given $i \in V$ and a given constant $c$, there exists a constant $c_5(c)$ such that*

$$P\Big(\frac{\int_{||u^i||^2\leq cM(p)}||u^i||\cdot|\pi^i(\theta_0^i + n^{-\frac{1}{2}}(J^i)^{-1}u^i)Z_n^i(u^i) - \pi^i(\theta_0^i)\tilde{Z}_n^i(u^i)|du^i}{\int\pi^i(\theta_0^i)\tilde{Z}_n^i(u^i)du^i} \leq c_5(c)\frac{p^{13}\log p}{\sqrt{n}}\Big)$$

$$> \quad 1 - 10.4\exp\{-\frac{1}{6}p^2\}.$$

**Proof.** Let $Q^i$ denote the set $\{u^i | ||u^i||^2 \leq cM(p)\}$. We get that

$$
\begin{aligned}
& [\int \pi^i(\theta_0^i)\tilde{Z}_n^i(u^i)du^i]^{-1} \int_Q ||u^i|| \cdot |\pi^i(\theta_0^i + n^{-1/2}(J^i)^{-1}u^i)Z_n^i(u^i) - \pi^i(\theta_0^i)\tilde{Z}_n^i(u^i)|du^i \\
= & [\int \pi^i(\theta_0^i)\tilde{Z}_n^i(u^i)du^i]^{-1} \int_{Q^i} ||u^i|| \cdot |\frac{\pi^i(\theta_0^i + n^{-1/2}(J^i)^{-1}u^i)}{\pi^i(\theta_0^i)} Z_n^i(u^i) - \tilde{Z}_n^i(u^i)|\pi^i(\theta_0^i)du^i \\
= & \frac{\int_{Q^i} ||u^i|| \cdot |\frac{\pi^i(\theta_0^i + n^{-1/2}(J^i)^{-1}u^i)}{\pi^i(\theta_0^i)} Z_n^i(u^i) - Z_n^i(u^i) + Z_n^i(u^i) - \tilde{Z}_n^i(u^i)|\pi^i(\theta_0^i)du^i}{\int \pi^i(\theta_0^i)\tilde{Z}_n^i(u^i)du^i} \\
= & \frac{\int_{Q^i} ||u^i|| \cdot |\frac{\pi^i(\theta_0^i + n^{-1/2}(J^i)^{-1}u^i)}{\pi^i(\theta_0^i)} Z_n^i(u^i) - Z_n^i(u^i) + Z_n^i(u^i) - \tilde{Z}_n^i(u^i)|du^i}{\int \tilde{Z}_n^i(u^i)du^i} \\
\leq & \frac{\int_{Q^i} ||u^i|| \cdot |(\frac{\pi^i(\theta_0^i + n^{-1/2}(J^i)^{-1}u^i)}{\pi^i(\theta_0^i)} - 1)Z_n^i(u^i)|du^i + \int_{Q^i} ||u^i|| \cdot |Z_n^i(u^i) - \tilde{Z}_n^i(u^i)|du^i}{\int \tilde{Z}_n^i(u^i)du^i} \\
\leq & \frac{\sup_{u^i \in Q^i} \left\{||u^i|| \cdot |\frac{\pi^i(\theta_0^i + n^{-1/2}(J^i)^{-1}u^i)}{\pi^i(\theta_0^i)} - 1|\right\} \int_{Q^i} Z_n^i(u^i)du^i}{\int \tilde{Z}_n^i(u^i)du^i} + \frac{\int_{Q^i} ||u^i|| \cdot |Z_n^i(u^i) - \tilde{Z}_n^i(u^i)|du^i}{\int \tilde{Z}_n^i(u^i)du^i}.
\end{aligned}
$$

Since

$$
\begin{aligned}
cM(p) \geq & \; ||u^i||^2 = ||\sqrt{n}J^i(\theta^i - \theta_0^i)||^2 = (\sqrt{n}J^i(\theta^i - \theta_0^i))^t \sqrt{n}J^i(\theta^i - \theta_0^i) \\
= & \; n(\theta^i - \theta_0^i)^t J^t J(\theta^i - \theta_0^i) = n(\theta^i - \theta_0^i)^t F^i(\theta^i - \theta_0^i) \\
\geq & \; n\lambda_{min}(F^i)(\theta^i - \theta_0^i)^t(\theta^i - \theta_0^i) = n\lambda_{min}(F^i)||\theta^i - \theta_0^i||^2
\end{aligned}
$$

then

$$
||\theta^i - \theta_0^i|| \leq \sqrt{\frac{cM(p)}{n\lambda_{min}(F^i)}} = \sqrt{\frac{cM(p)||(F^i)^{-1}||}{n}}.
$$

By Proposition 4.1, we have $\kappa_1^2 \leq ||(F^i)^{-1}|| \leq \kappa_2^2$. Since $\frac{p^{12}(\log p)^{\frac{1}{2}}}{\sqrt{n}} \to 0$, then $\frac{p^2 \log p}{n} \to 0$. Therefore, $||\theta^i - \theta_0^i|| \to 0$. Using the fact $|e^x - 1| \leq 2|x|$ for sufficiently small $|x|$, we obtain

$$
\begin{aligned}
& \sup_{u^i \in Q^i} \left\{||u^i|| \cdot |\frac{\pi^i(\theta_0^i + n^{-1/2}(J^i)^{-1}u^i)}{\pi^i(\theta_0^i)} - 1|\right\} \\
\leq & \; 2 \sup_{u^i \in Q^i} \{||u^i|| \cdot |\log \frac{\pi^i(\theta_0^i + n^{-1/2}(J^i)^{-1}u^i)}{\pi^i(\theta_0^i)}|\} \\
\leq & \; 2\sqrt{cM(p)}M_1 p||\theta^i - \theta_0^i|| \qquad \text{by Proposition 4.5} \\
\leq & \; 2\sqrt{cM(p)}M_1 p\sqrt{\frac{cM(p)||(F^i)^{-1}||}{n}} \\
\leq & \; \frac{2cM_1\kappa_2 M(p)p}{\sqrt{n}}
\end{aligned}
$$

where $M_1$ is a constant. We also have that

$$\begin{aligned}
\frac{\int_{Q^i} Z_n^i(u^i)du^i}{\int \tilde{Z}_n^i(u^i)du^i} &= \frac{\int_{Q^i} Z_n^i(u^i)du^i + \int_{Q^i} \tilde{Z}_n(u^i)du^i - \int_{Q^i} \tilde{Z}_n^i(u^i)du^i}{\int \tilde{Z}_n^i(u^i)du^i} \\
&= \frac{\int_{Q^i} \tilde{Z}_n^i(u^i)du^i + \int_{Q^i} [Z_n^i(u^i) - \tilde{Z}_n^i(u^i)]du^i}{\int \tilde{Z}_n^i(u^i)du^i} \\
&\leq \frac{\int \tilde{Z}_n^i(u^i)du^i + \int_{Q^i} |Z_n^i(u^i) - \tilde{Z}_n^i(u^i)|du^i}{\int \tilde{Z}_n^i(u^i)du^i} \\
&\leq 1 + \Big(\int \tilde{Z}_n^i(u^i)du^i\Big)^{-1} \int_{Q^i} |Z_n^i(u^i) - \tilde{Z}_n^i(u^i)|du^i.
\end{aligned}$$

According to Lemma **??** in the Appendix, we can obtain

$$\Big(\int \tilde{Z}_n^i(u^i)du^i\Big)^{-1} \int_{Q^i} |Z_n^i(u^i) - \tilde{Z}_n^i(u^i)|du^i \leq f^i(||\Delta_n^i||, c) \tag{20}$$ ff

where

$$\begin{aligned}
f^i(||\Delta_n^i||, c) &= \varphi_n^i(c)[p^2 + \big(1 - 2\varphi_n^i(c)\big)^{-1}||\Delta_n^i||^2]\Big(1 - 2\varphi_n^i(c)\Big)^{-(\frac{p_i^2}{2}+1)} \\
&\quad \times \exp\Big\{\frac{\varphi_n^i(c)||\Delta_n^i||^2}{1 - 2\varphi_n^i(c)}\Big\},
\end{aligned}$$

and

$$\varphi_n^i(c) = \frac{1}{6}[n^{-\frac{1}{2}}\big(cM(p)\big)^{\frac{1}{2}} B_{1n}^i(0) + n^{-1}cM(p)B_{2n}^i(c\frac{M(p)}{S_i})]$$

Furthermore, since $||u^i|| \leq \sqrt{cM(p)}$, by the inequality (20), it is easy to see that

$$\frac{\int_{Q^i} ||u^i|| \cdot |Z_n^i(u^i) - \tilde{Z}_n^i(u^i)|du^i}{\int \tilde{Z}_n^i(u^i)du^i} \leq \sqrt{cM(p)} f(||\Delta_n^i||, c).$$

Combining the above results, we can show that the LHS in (20) is bounded by

$$R_1(||\Delta_n^i||, c) = \frac{2cM_1\kappa_2 M(p)p}{\sqrt{n}}[1 + f^i(||\Delta_n^i||, c)] + \sqrt{cM(p)} f^i(||\Delta_n^i||, c).$$

According to Proposition 4.6, we have $B_{1n}^i(0) = O(p^9)$ and $B_{2n}^i(c\frac{M(p)}{S_i}) = O(p^{12})$. Therefore, there exist two constants $c_1$ and $c_2$ such that

$$\begin{aligned}
\varphi_n^i(c) &\leq \frac{1}{6}[n^{-\frac{1}{2}}\big(cM(p)\big)^{\frac{1}{2}} c_1 p^9 + n^{-1}cM(p)c_2 p^{12}] \\
&= \frac{1}{6}[\sqrt{c}\frac{\sqrt{p^2 \log p}}{\sqrt{n}} c_1 p^9 + \frac{cp^2 \log p}{n} c_2 p^{12}] \\
&= \frac{1}{6}[\sqrt{c}c_1 \frac{p^{10}\sqrt{\log p}}{\sqrt{n}} + c_2 \frac{p^{14} \log p}{n}] \\
&= \frac{1}{6}\frac{p^{10}\sqrt{\log p}}{\sqrt{n}}[\sqrt{c}c_1 + c_2 \frac{p^4\sqrt{\log p}}{\sqrt{n}}].
\end{aligned} \tag{21}$$ cc

Since the first term in (21) is the dominating term, then there exists a constant $c_3(c)$ such that $\varphi_n^i \leq c_3(c)\frac{p^{10}\sqrt{\log p}}{\sqrt{n}}$. Since $\frac{p^{12}(\log p)^{\frac{1}{2}}}{\sqrt{n}} \to 0$, then immediately $\varphi_n^i(c) \to 0$. Furthermore, using the fact $(1-x)^{-1} \leq 2$ and $-\log(1-x) \leq 2x$ for sufficiently small $x$, we have $[1-2\varphi_n^i(c)]^{-1} \leq 2$ and $e^{-(\frac{p_i^2}{2}+1)\log\left(1-2\varphi_n^i(c)\right)} \leq e^{(\frac{p_i^2}{2}+1)4\varphi_n^i(c)}$. Therefore, the following inequality holds

$$f^i(||\Delta_n^i||, c) \leq \varphi_n^i(c)[p^2 + 2||\Delta_n^i||^2]\exp\{(\frac{p^2}{2}+1)4\varphi_n^i(c)\}\exp\left\{2\varphi_n^i(c)||\Delta_n^i||^2\right\}.$$

According to Lemma 4.4, we see that $P(||\Delta_n^i||^2 \leq 3a^2p^2) > 1 - 10.4\exp\{-\frac{1}{6}p^2\}$. Therefore,

$$f^i(||\Delta_n^i||, c) \quad \leq \quad c_3(c)\frac{p^{10}\sqrt{\log p}}{\sqrt{n}}[p^2 + 6a^2p^2]\exp\left\{c_3(c)\frac{p^{10}\sqrt{\log p}}{\sqrt{n}}(6a^2p^2 + 2p^2 + 4)\right\}$$

with a probability greater than $1-10.4\exp\{-\frac{1}{6}p^2\}$. Since $\frac{p^{12}\sqrt{\log p}}{\sqrt{n}} \to 0$, then $\frac{p^{10}\sqrt{\log p}}{\sqrt{n}}(6a^2p^2 + 2p^2 + 2) \to 0$. Therefore, $\exp\left\{c_3(c)\frac{p^{10}\sqrt{\log p}}{\sqrt{n}}(4a^2p^2 + 2p^2 + 2)\right\} < 2$. It follows

$$f^i(||\Delta_n^i||, c) \quad \leq \quad 2(1+6a^2)c_3(c)\frac{p^{12}\sqrt{\log p}}{\sqrt{n}}$$

with a probability greater than $1 - 10.4\exp\{-\frac{1}{6}p^2\}$. Let $c_4(c) = 2(1+6a^2)c_3(c)$, then $f^i(||\Delta_n^i||, c) \leq c_4(c)\frac{p^{12}\sqrt{\log p}}{\sqrt{n}}$ with a probability greater than $1 - 10.4\exp\{-\frac{1}{6}p^2\}$. Furthermore, we can get

$$\begin{aligned} R_1(||\Delta_n^i||, c) &= \frac{2cM_1\kappa_2 M(p)p}{\sqrt{n}}[1 + c_4(c)\frac{p^{12}\sqrt{\log p}}{\sqrt{n}}] + \sqrt{cM(p)}c_4(c)\frac{p^{12}\sqrt{\log p}}{\sqrt{n}} \\ &= \frac{2cM_1\kappa_2 p^3\log p}{\sqrt{n}} + c_4(c)\frac{2cM_1\kappa_2 p^{15}\log^{\frac{3}{2}} p}{n} + \sqrt{c}c_4(c)\frac{p^{13}\log p}{\sqrt{n}} \\ &= \frac{2cM_1\kappa_2 p^3\log p}{\sqrt{n}} + \frac{p^{13}\log p}{\sqrt{n}}[c_4(c)\frac{2cM_1\kappa_2 p^2\log^{\frac{1}{2}} p}{\sqrt{n}} + \sqrt{c}c_4(c)] \quad (22) \end{aligned}$$ b3

with a probability greater than $1 - 10.4\exp\{-\frac{1}{6}p^2\}$. It is easy to see that the third term in (22) is the dominating term. Therefore, there exists a constant $c_5(c)$ such that $R_1(||\Delta_n^i||, c) \leq c_5(c)\frac{p^{13}\log p}{\sqrt{n}}$ with a probability greater than $1 - 10.4\exp\{-\frac{1}{6}p^2\}$. The proof is completed. ■

⟨secondtermnorm⟩

**Lemma 4.6** *There exists a constant $c$ and a constant $c_9(c)$ such that for any given $i \in V$,*

$$
\begin{aligned}
& P\Big(\frac{\int_{||u^i||^2 > cM(p)} ||u^i||\pi^i(\theta_0^i + n^{-\frac{1}{2}}(J^i)^{-1}u^i)Z_n^i(u^i)du^i}{\int \pi^i(\theta_0^i)\tilde{Z}_n^i(u^i)du^i} \le \exp[-c_9(c)p^2\log p]\Big) \\
> \;& 1 - 10.4\exp\{-\frac{1}{6}p^2\}.
\end{aligned}
$$

**Proof.** Let

$$
\begin{aligned}
R_2(||\Delta_n^i||, c) &= \frac{\int_{||u^i||^2 > cM(p)} ||u^i||\pi^i(\theta_0^i + n^{-\frac{1}{2}}(J^i)^{-1}u^i)Z_n^i(u^i)du^i}{\int \pi^i(\theta_0^i)\tilde{Z}_n^i(u^i)du^i} \\
&= \frac{\int_{||u^i||^2 > cM(p)} ||u^i||\frac{\pi^i(\theta_0^i + n^{-\frac{1}{2}}(J^i)^{-1}u^i)}{\pi^i(\theta_0^i)}Z_n^i(u^i)du^i}{\int \tilde{Z}_n^i(u^i)du^i} \\
&= \frac{\int_{||u^i||^2 > cM(p)} ||u^i||\frac{\pi^i(\theta_0^i + n^{-\frac{1}{2}}(J^i)^{-1}u^i)}{\pi^i(\theta_0^i)}Z_n^i(u^i)du^i}{(2\pi)^{S_i/2}\exp[\frac{||\Delta_n^i||^2}{2}]}
\end{aligned}
$$

According to Lemma 2.2 in Ghosal [2000], we have that $Z_n^i(u^i) \le \exp[-\frac{1}{4}cp^2\log p]$ with a probability greater than $1 - 10.4\exp\{-\frac{1}{6}p^2\}$. Therefore, we obtain that

$$
\begin{aligned}
R_2(||\Delta_n^i||, c) &\le \frac{\exp[-\frac{1}{4}cp^2\log p]}{(2\pi)^{S_i/2}\exp[\frac{||\Delta_n^i||^2}{2}]}\int_{||u^i||^2 > cM(p)} ||u^i||\frac{\pi^i(\theta_0^i + n^{-\frac{1}{2}}(J^i)^{-1}u^i)}{\pi^i(\theta_0^i)}du^i \\
&= \frac{\exp[-\frac{1}{4}cp^2\log p]}{(2\pi)^{S_i/2}\exp[\frac{||\Delta_n^i||^2}{2}]}\int_{||u^i||^2 > cM(p)} ||u^i||\frac{\pi_0^i(\theta_0^i + n^{-\frac{1}{2}}(J^i)^{-1}u^i)}{\pi_0^i(\theta_0^i)}du^i \\
&= \frac{\exp[-\frac{1}{4}cp^2\log p]}{(2\pi)^{S_i/2}\exp[\frac{||\Delta_n^i||^2}{2}]} \\
&\quad \times \int_{||\sqrt{n}J^i(\theta^i - \theta_0^i)||^2 > cM(p)} ||\sqrt{n}J^i(\theta^i - \theta_0^i)||\frac{\pi_0^i(\theta^i)}{\pi_0^i(\theta_0^i)}n^{S_i/2}det(J^i)d\theta^i \\
&\le \frac{n^{S_i/2}det(J^i)\exp[-\frac{1}{4}cp^2\log p]}{\pi_0^i(\theta_0^i)(2\pi)^{S_i/2}\exp[\frac{||\Delta_n^i||^2}{2}]}\int_{||\sqrt{n}J^i(\theta^i - \theta_0^i)||^2 > cM(p)} ||\sqrt{n}J^i(\theta^i - \theta_0^i)||\pi_0^i(\theta^i)d\theta^i \\
&\le \exp[\frac{S_i}{2}\log n + \frac{1}{2}\log(det(F^i)) - \frac{1}{4}cp^2\log p \\
&\quad + \log\int_{||\sqrt{n}J^i(\theta^i - \theta_0^i)||^2 > cM(p)} ||\sqrt{n}J^i(\theta^i - \theta_0^i)||\pi_0^i(\theta^i)d\theta^i - \log\pi_0^i(\theta_0^i) - \frac{S_i}{2}\log 2\pi - \frac{||\Delta_n^i||^2}{2}] \\
&\le \exp[\frac{S_i}{2}\log n + \frac{1}{2}\log(det(F^i)) - \frac{1}{4}cp^2\log p - \log\pi_0^i(\theta_0^i) \\
&\quad + \log\int_{||\sqrt{n}J^i(\theta^i - \theta_0^i)||^2 > cM(p)} ||\sqrt{n}J^i(\theta^i - \theta_0^i)||\pi_0^i(\theta^i)d\theta^i]
\end{aligned}
$$

with a probability greater than $1 - 10.4\exp\{-\frac{1}{6}p^2\}$. By Proposition 4.4 and Lemma 6.4 in the Appendix, we have that

$$R_2(||\Delta_n^i||, c) \leq \exp[\frac{S_i}{2}\log n + \frac{1}{2}\log(det(F^i)) - \frac{1}{4}cp^2\log p + \frac{1}{2}p_i\kappa_2 - \frac{\delta-2}{2}p_i\log\kappa_1 + M_7p^2\log p]$$

with a probability greater than $1 - 10.4\exp\{-\frac{1}{6}p^2\}$. By Condition (1), $\log n$ and $\log p$ are of the same order. Furthermore, Proposition 4.3 implies $\log(\det(F^i)) = O(p^2)$. Therefore, there exists a constant $c_6$ such that $\log(\det(F^i)) \leq c_6p^2$. It follows the RHS in (23) is bounded by the following term with a probability greater than $1-10.4\exp\{-\frac{1}{6}p^2\}$.

$$\exp[\frac{p(p+1)}{4}\log p + \frac{1}{2}c_6p^2 - \frac{1}{4}cp^2\log p + \frac{1}{2}p_i\kappa_2 - \frac{\delta-2}{2}p_i\log\kappa_1 + M_7p^2\log p]$$

Furthermore, there exists a constant $c_8$ such that

$$R_2(||\Delta_n^i||, c) \leq \exp[\tfrac{p(p+1)}{4}\log p - \tfrac{1}{4}cp^2\log p + M_7p^2\log p + c_8p^2\log p]$$

with a probability greater than $1 - 10.4\exp\{-\frac{1}{6}p^2\}$. We can choose a constant $c$ big enough such that $c_9(c) = \frac{1}{4} - \frac{1}{4}c + c_8 + M_7 < 0$. It immediately implies $R_2(||\Delta_n^i||, c) \leq \exp[-c_9(c)p^2\log p]$ with a probability greater than $1 - 10.4\exp\{-\frac{1}{6}p^2\}$. ■

⟨thirdtermnorm⟩

**Lemma 4.7** *For any given $i \in V$ and any constant $C$ with $C > 9a^2$, $a^2 > 1$, we have*

$$P(\int_{||u^i||^2 > CM(p)} ||u^i||\phi(u^i; \Delta_n^i, I_{S_i})du^i \leq \frac{2}{\sqrt{2\pi}}p^{-3a^2+2} + \sqrt{3a^2}\frac{2}{\sqrt{2\pi}}p^{-3a^2+1}) > 1 - 10.4\exp\{-\frac{1}{6}p^2\}$$

**Proof.** First we observe that

$$\begin{aligned} &\int_{||u^i||^2>CM(p)} ||u^i||\phi(u^i; \Delta_n^i, I_{S_i})du^i \\ \leq &\int_{||u^i||^2>CM(p)} (||u^i - \Delta_n^i|| + ||\Delta_n^i||)\phi(u^i; \Delta_n^i, I_{S_i})du^i \\ \leq &\int_{||u^i||^2>CM(p)} (||u^i - \Delta_n^i||)\phi(u^i; \Delta_n^i, I_{S_i})du^i + \int_{||u^i||^2>CM(p)} ||\Delta_n^i||\phi(u^i; \Delta_n^i, I_{S_i})du^i. \end{aligned}$$

Let $v^i = u^i - \Delta_n^i$, since $||v^i||^2 + ||\Delta_n^i||^2 \geq ||v^i + \Delta_n^i||^2 = ||u^i||^2 > CM(p)$, then immediately $||v^i||^2 > CM(p) - ||\Delta_n^i||^2$. By Lemma 4.4, we can see that $||\Delta_n^i||^2 \leq 3a^2p^2$ with a probability greater than $1 - 10.4\exp\{-\frac{1}{6}p^2\}$ with $a^2 > 1$. Therefore, we can get $||v^i||^2 > CM(p) - ||\Delta_n^i||^2 > CM(p) - 3a^2p^2 > (9a^2 - 3a^2)p^2 \log p = 6a^2p^2 \log p$ with a probability greater than $1 - 10.4\exp\{-\frac{1}{6}p^2\}$. Thus the following inequality holds with a probability greater than $1 - 10.4\exp\{-\frac{1}{6}p^2\}$.

$$\begin{aligned}
& \int_{||u^i||^2 > CM(p)} (||u^i - \Delta_n^i||)\phi(u^i; \Delta_n^i, I_{S_i}) du^i \\
= & \int_{||v^i + \Delta_n^i||^2 > CM(p)} ||v^i||\phi(v^i; 0, I_{S_i}) dv^i \leq \int_{||v^i||^2 > 6a^2M(p)} ||v^i||\phi(v^i; 0, I_{S_i}) dv^i \\
\leq & \int_{||v^i||^2 > 6a^2M(p)} \sum_{j=1}^{S_i} |v_j^i|\phi(v^i; 0, I_{S_i}) dv^i = \sum_{j=1}^{S_i} \int_{||v^i||^2 > 6a^2M(p)} |v_j^i|\phi(v^i; 0, I_{S_i}) dv^i
\end{aligned}$$

Since $||v^i||^2 > 6a^2M(p)$, there exists a $k \in \{1, 2, \cdots, S_i\}$ such that $(v^i_j)^2 > \frac{6a^2M(p)}{S_i}$. Therefore,

$$
\begin{aligned}
& \sum_{j=1}^{S_i} \int_{||v^i||^2 > 6a^2M(p)} |v^i_j| \phi(v^i; 0, I_{S_i}) dv^i \\
\leq \; & \sum_{j=1}^{S_i} \prod_{j \neq k} \int_{-\infty}^{+\infty} \int_{(v^i_k)^2 > 6a^2 \frac{M(p)}{S_i}} |v^i_j| \phi(v^i; 0, I_{S_i}) dv^i \\
= \; & \prod_{j \neq k} \int_{-\infty}^{+\infty} \int_{(v^i_k)^2 > 6a^2 \frac{M(p)}{S_i}} |v^i_k| \phi(v^i; 0, I_{S_i}) dv^i + \sum_{j \neq k} \prod_{j \neq k} \int_{-\infty}^{+\infty} \int_{(v^i_k)^2 > 6a^2 \frac{M(p)}{S_i}} |v^i_j| \phi(v^i; 0, I_{S_i}) dv^i \\
= \; & \int_{(v^i_k)^2 > 6a^2 \frac{M(p)}{S_i}} |v^i_k| \frac{1}{\sqrt{2\pi}} e^{-\frac{(v^i_k)^2}{2}} dv^i_k \\
& + \sum_{j \neq k} \int_{(v^i_k)^2 > 6a^2 \frac{M(p)}{S_i}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(v^i_k)^2}{2}} dv^i_k \int_{-\infty}^{\infty} |v^i_j| \frac{1}{\sqrt{2\pi}} e^{-\frac{(v^i_j)^2}{2}} dv^i_j \\
= \; & 2 \int_{v^i_k > \sqrt{6a^2 \frac{M(p)}{S_i}}} v^i_k \frac{1}{\sqrt{2\pi}} e^{-\frac{(v^i_k)^2}{2}} dv^i_k \\
& + \sum_{j \neq k} 2 \int_{v^i_k > \sqrt{6a^2 \frac{M(p)}{S_i}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(v^i_k)^2}{2}} dv^i_k [2 \int_0^{\infty} v^i_j \frac{1}{\sqrt{2\pi}} e^{-\frac{(v^i_j)^2}{2}} dv^i_j] \\
< \; & 2 \int_{v^i_k > \sqrt{6a^2 \frac{M(p)}{S_i}}} v^i_k \frac{1}{\sqrt{2\pi}} e^{-\frac{(v^i_k)^2}{2}} dv^i_k \\
& + \sum_{j \neq k} 2 \int_{v^i_k > \sqrt{6a^2 \frac{M(p)}{S_i}}} v^i_k \frac{1}{\sqrt{2\pi}} e^{-\frac{(v^i_k)^2}{2}} dv^i_k [2 \int_0^{\infty} v^i_j \frac{1}{\sqrt{2\pi}} e^{-\frac{(v^i_j)^2}{2}} dv^i_j] \\
= \; & 2 \int_{v^i_k > \sqrt{6a^2 \frac{M(p)}{S_i}}} v^i_k \frac{1}{\sqrt{2\pi}} e^{-\frac{(v^i_k)^2}{2}} dv^i_k [1 + \sum_{j \neq k} 2 \int_0^{\infty} v^i_j \frac{1}{\sqrt{2\pi}} e^{-\frac{(v^i_j)^2}{2}} dv^i_j] \\
= \; & 2 \frac{1}{\sqrt{2\pi}} e^{-\frac{6a^2M(p)}{2S_i}} [1 + (S_i - 1) 2 \frac{1}{\sqrt{2\pi}}] \leq 2S_i \frac{1}{\sqrt{2\pi}} e^{-\frac{6a^2M(p)}{2S_i}} \leq 2p^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{6a^2M(p)}{2p^2}} \\
\leq \; & 2p^2 \frac{1}{\sqrt{2\pi}} p^{-3a^2}
\end{aligned}
$$

with a probability greater than $1 - 10.4 \exp\{-\frac{1}{6}p^2\}$. Similarly, there exists a $l \in \{1, 2, \cdots, S_i\}$ such that

$$
\begin{aligned}
\int_{||u^i||^2 > CM(p)} ||\Delta^i_n|| \phi(u^i; \Delta^i_n, I_{S_i}) du^i \; & \leq \; ||\Delta^i_n|| 2 \int_{v^i_l > \sqrt{6a^2 \frac{M(p)}{S_i}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(v^i_l)^2}{2}} dv^i_l \\
& = \; ||\Delta^i_n|| \frac{2}{\sqrt{2\pi}} p^{-3a^2} \\
& \leq \; \sqrt{3a^2} p \frac{2}{\sqrt{2\pi}} p^{-3a^2}
\end{aligned}
$$

with a probability greater than $1 - 10.4 \exp\{-\frac{1}{6}p^2\}$. Hence, the desired result follows.

■

⟨allnorm⟩

**Lemma 4.8** *For a given $i \in V$, we have*

$$\sqrt{n}J^i(\tilde{\theta}^i - \theta_0^i) = \Delta_n^i + \int u^i[\pi_*^i(u^i) - \phi(u^i; \Delta_n^i, I_{S_i})]du^i$$

*where $\pi_*^i(u^i)$ is the posterior distribution of $u^i$.*

**Proof.** Let $\pi_*^i(\theta^i)$ be the posterior distribution of $\theta^i$. Therefore, $\pi_*^i(u^i) = \pi_*^i(\theta_0^i + n^{-\frac{1}{2}}(J^i)^{-1}u^i)|n^{-\frac{1}{2}}(J^i)^{-1}|$. Thus

$$\begin{aligned}
\tilde{\theta}^i &= \int \theta^i \cdot \pi_*^i(\theta^i)d\theta^i \\
&= \int (\theta_0^i + n^{-\frac{1}{2}}(J^i)^{-1}u^i)\pi_*^i(\theta_0^i + n^{-\frac{1}{2}}(J^i)^{-1}u^i)|n^{-1/2}(J^i)^{-1}|d\theta^i \\
&= \int (\theta_0^i + n^{-\frac{1}{2}}(J^i)^{-1}u^i)\pi_*^i(u^i)du^i \\
&= \theta_0^i + n^{-\frac{1}{2}}(J^i)^{-1}\int u^i\pi_*^i(u^i)du^i.
\end{aligned}$$

It follows

$$\sqrt{n}J^i(\tilde{\theta}^i - \theta_0^i) = \int u^i\pi_*^i(u^i)du^i.$$

On the other hand, the following equations hold

$$\begin{aligned}
\int u^i\phi(u^i; \Delta_n^i, I_{S_i})du &= \int (u^i - \Delta_n^i + \Delta_n^i)\phi(u^i; \Delta_n^i, I_{S_i})du^i \\
&= \int (u^i - \Delta_n^i)\phi(u^i; \Delta_n^i, I_{S_i})du^i + \Delta_n^i\int \phi(u^i; \Delta_n^i, I_{S_i})du^i \\
&= \Delta_n^i.
\end{aligned}$$

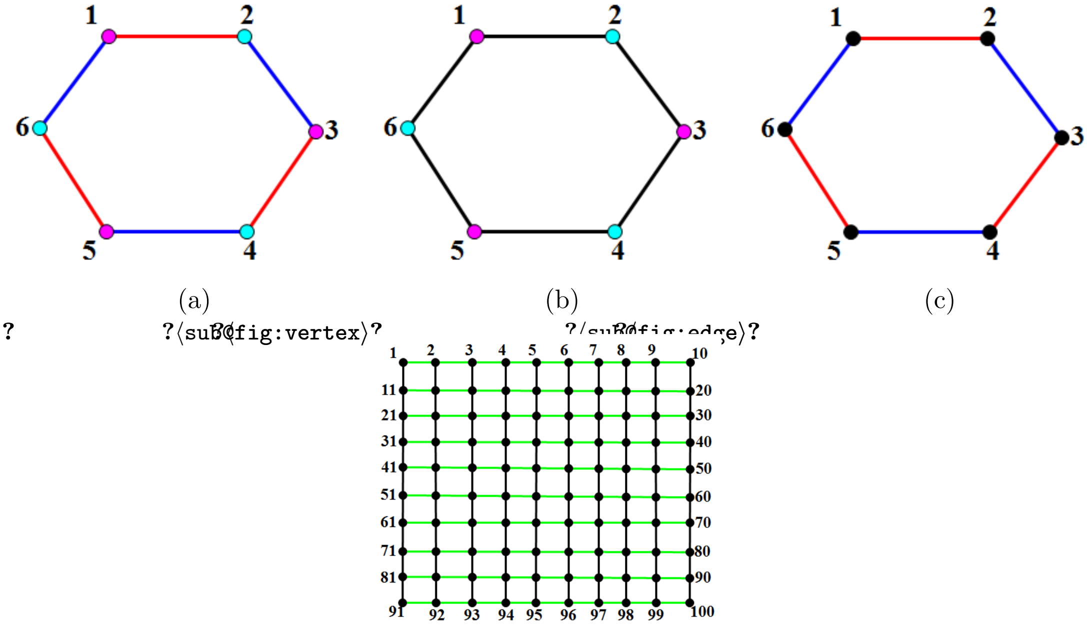
We thus have

$$\begin{aligned}
\sqrt{n}J^i(\tilde{\theta}^i - \theta_0^i) - \Delta_n^i &= \int u^i\pi_*^i(u^i)du^i - \int u^i\phi(u^i; \Delta_n^i, I_{S_i})du^i \\
&= \int u^i[\pi_*^i(u^i) - \phi(u^i; \Delta_n^i, I_{S_i})]du^i.
\end{aligned}$$

■

# 5 Simulations

In order to investigate the performance of our proposed composite Bayesian estimators under different scenarios, we conducted a number of numerical experiments. First,




(a) (b) (c)

g:vertex and edge⟩? ?⟨sub@fig:vertex⟩? ?⟨sub@fig:edge⟩?

(d)

?⟨sub@fig:vertex⟩?

Figure 1: Cycles of length 6 with the three different patterns of colouring that we use for the cycles of length $p = 20$ and $p = 30$. Black vertices or edges indicate different arbitrary colours.

⟨fig:1⟩

we evaluate the composite 1-hop Bayesian (MBE−1hop) estimator, the composite 2-hop Bayesian (MBE−2hop) estimator and the global Bayesian estimators (GBE) for three colored cycles. Moreover, We also compare them with the global maximum likelihood estimators (GMLE) for the three colored cycles when the sample size are different. For a lattice colored graph, we compare the 1-hop composite maximum likelihood estimators (MMLE-1hop), the 1-hop composite Bayesian estimator with the global maximum likelihood estimators. The normalized mean squared errors (NMSE) are defined as $\frac{||\hat{K}-K||^2}{||K||^2}$.

Figure 1 (a)-(c) show three different patterns of colouring of an cycle of order 6. In particular, we show the simulation results for $p = 20$ and $p = 30$. The parameters for these figures are listed in Table 1. Furthermore, a square lattice colored graph with $p = 10 \times 10 = 100$ vertices is illustrated in Figure 1 (d). For this figure, the parameters are chosen as $K_{i+10(j-1),i+1+10(j-1)} = 1$ for $i = 1, 2, \ldots, 9$ and $j = 1, 2, \ldots, 10$, $K_{i+10(j-1),i+10j} = 1 + 0.01i + 0.1j$ for $i = 1, 2, \ldots, 10$ and $j = 1, 2, \ldots, 9$

and $K_{i,i} = 10 + 0.01i$ for $i = 1, 2, \ldots, 100$. The posterior mean estimates are based on 5000 iterations after the first 1000 burn-in iterations.

Table 1: The parameters chosen for the matrix $K$ for producing figure 1.

| parameters | Figure 1 (a) | Figure 1 (b) | Figure 1 (c) |
|---|---|---|---|
| $K_{ii}$ $(i = 1, 3, \ldots, 2p-1)$ | 0.1 | 0.1 | 0.1+0.1i |
| $K_{ii}$ $(i = 2, 4, \ldots, p)$ | 0.03 | 0.3 | 0.03+0.01i |
| $K_{i,i+1} = K_{i+1,i}$ $(i = 1, 3, \ldots, 2p-1)$ | 0.01 | 0.01+0.001i | 0.01 |
| $K_{i,i+1} = K_{i+1,i}$ $(i = 2, 4, \ldots, p-2)$ | 0.02 | 0.01+0.002i | 0.02 |
| $K_{1p} = K_{p1}$ | 0.02 | 0.01 | 0.02 |

⟨table:parameters⟩

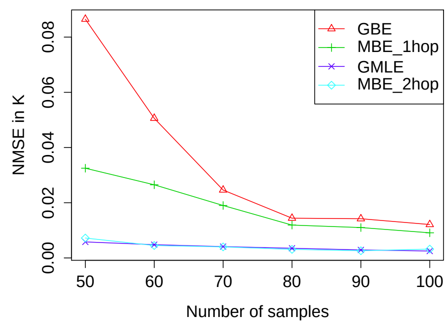
Table 2 shows $NMSE(K, \hat{K})$ for the comparison between the three colored models on the simulated examples when $p = 20$ and $p = 30$, averaged over 100 simulations. For each graph, we generated 100 datasets from the $N(0, K^{-1})$ distribution. The $NMSE$ and the standard derivations are shown in the 3th 4th and 5th columns of Table 2. Standard errors are indicated in parentheses. From the results shown on Table 2, our proposed composite 1-hop and 2-hop Bayesian estimators perform very well comparing the global Bayesian estimators. Computation was performed on a 2 core 4 threads with i5-4200U, 2.3 GHZ chips and 8GB of RAM, running on Windows 8. The average computing time shows in the 3th, 4th and 5th columns in minutes on Table 3. As shown in the simulation result on Table 3, the total computing time of composite Bayesian estimation is much smaller than the global Bayesian estimation.

Figure 2 shows the NMSE curves for the colored graphs in Figure 1 when the sample sizes are from 50-100. The averaged NMSE is illustrated over 100 simulations. In summary, our proposed estimators can perform better as the sample size increase and the NMSE for the composite 1-hop Bayesian estimator always smaller than the NMSE for the composite 2-hop Bayesian estimator.

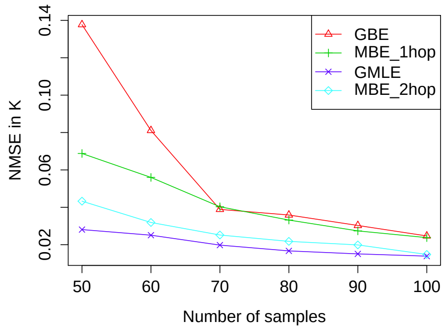
# 6 Appendix

The following Isserlis'Theorem give us the general formula for the moments of the multivariate normal distribution in terms of its covariance matrix and also shows that all the moments of the multivariate normal distribution are finite.

⟨Isserlis⟩

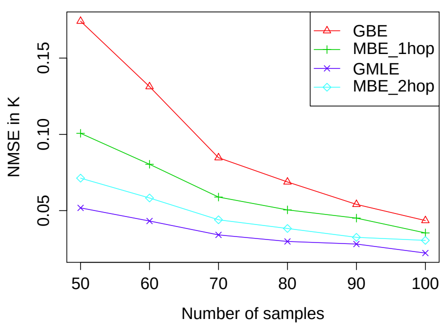



(a)

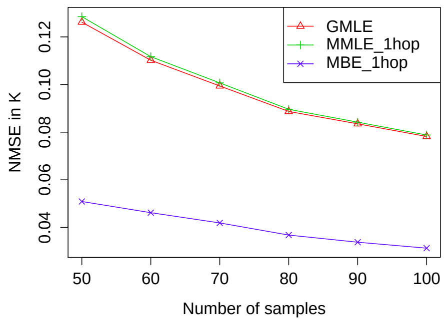



(b)

?@fig:vertex and edge 20⟩? ⟩?




(c)




(d)

?⟨sub@fig:edge 20⟩? ?⟨sub@fig:lattice⟩?

Figure 2: NMSE in $K$ for different colored graphical models. (a) NMSE for the colored graph in Figure 1 (a) when $p = 20$. (b) NMSE for the colored graph in Figure 1 (b) when $p = 20$. (c) NMSE for the colored graph in Figure 1 (c) when $p = 20$. (d) NMSE for the colored lattice graph in Figure 1 (d) when $p = 100$.

⟨fig:2⟩

Table 2: $NMSE(K, \hat{K})$ for the three colored models when $p = 20$ and $p = 30$.

| | | NMSE | | |
|---|---|---|---|---|
| $p$ | $\mathcal{G}$ | MBE_1hop | MBE_2hop | GBE |
| | (a) | 0.0162 (0.0155) | 0.0032 (0.0027) | 0.0110 (0.0102) |
| 20 | (b) | 0.0256 (0.0153) | 0.0148 (0.0058) | 0.0237 (0.0189) |
| | (c) | 0.0375 (0.0283) | 0.0305 (0.0142) | 0.0308 (0.0241) |
| | (a) | 0.0098 (0.0070) | 0.0017(0.0014) | 0.0317 (0.0571) |
| 30 | (b) | 0.0234 (0.0088) | 0.0151(0.0054) | 0.0482 (0.0533) |
| | (c) | 0.0379 (0.0127) | 0.0308 (0.0086) | 0.0823 (0.0257) |

⟨table:2⟩

Table 3: Timing for the three colored models when $p = 20$ and $p = 30$.

| | | Timing | | |
|---|---|---|---|---|
| $p$ | $\mathcal{G}$ | MBE_1hop | MBE_2hop | GBE |
| | (a) | 0.365 | 3.410 | 21.875 |
| 20 | (b) | 1.047 | 3.353 | 16.249 |
| | (c) | 0.944 | 3.054 | 15.513 |
| | (a) | 1.442 | 4.952 | 83.965 |
| 30 | (b) | 1.538 | 4.557 | 80.255 |
| | (c) | 1.504 | 4.509 | 79.918 |

⟨table:3⟩

**Lemma 6.1** ***(Isserlis'Theorem)*** *Let $X = (X_1, X_2, \cdots, X_n)$ be the random variables following the multivariate normal distribution $N_p(0, \Sigma)$, then*

$$E[X_{a_1} X_{a_2} \cdots X_{a_{2n}}] = \sum_{\sigma} A^{(\sigma)}$$

*and*

$$E[X_{a_1} X_{a_2} \cdots X_{a_{2n-1}}] = 0$$

*where the sum is over every partition $\sigma$ of $\{1, 2, \ldots, 2n\}$ into $n$ disjoint pairs $(\sigma(2k-1), \sigma(2k))$ such that $\sigma(2k-1) < \sigma(2k)$, for $k = 1, 2, \ldots, n$, and $\sigma(2k-1) < \sigma(2k+1)$ for $k = 1, 2, \ldots, n-1$. For each partition $\sigma$, $A^{(\sigma)} = \prod_{k=1}^{n} \Sigma_{\sigma(2k-1)\sigma(2k)}$.*

⟨aa⟩

**Lemma 6.2** *Let $Y_j^i$ be defined in* (12) *and denote $Y_j^i = (Y_{j1}, Y_{j2}, \cdots, Y_{jS_i})^t$, under Condition (2), we have $E\Big[|Y_{jk_1}^i| \cdots |Y_{jk_h}^i|\Big]$ is bounded for $h = 1, 2, 3, 4$.*

**Proof.** Since $Y_{jk}^i = -\frac{1}{2} tr(\delta_k^i X_j^i (X_j^i)^t)$, by Isserlis' Theorem (see Lemma 6.1 in Appendix), the moments of every entry of $X_j^i (X_j^i)^t$ is finite. By Condition (3), $E(Y_{jk}^i)$ is bounded and $E(Y_{jk}^i)^2$ is also bounded. By Hölder's inequality ($E[|XY|] \leq (E[|X|^p])^{\frac{1}{p}} (E[|Y|^q])^{\frac{1}{q}}$ on wiki), when $h = 1$, $E(|Y_{jk_1}^i|) \leq (E(Y_{jk_1}^i)^2)^{\frac{1}{2}}$ is bounded. When $h = 2$, we have

$$E(|Y_{jk_1}^i||Y_{jk_2}^i|) \leq (E(Y_{jk_1}^i)^2)^{\frac{1}{2}} (E(Y_{jk_2}^i)^2)^{\frac{1}{2}}.$$

It follows $E(|Y_{jk_1}^i||Y_{jk_2}^i|)$ is bounded. When $h = 3$, we have

$$E(|Y_{jk_1}^i||Y_{jk_2}^i||Y_{jk_3}^i|) \leq (E(|Y_{jk_1}^i||Y_{jk_2}^i|)^2)^{\frac{1}{2}} (E(Y_{jk_3}^i)^2)^{\frac{1}{2}}.$$

Since $E(|Y_{jk_1}^i||Y_{jk_2}^i|)$ is bounded, then $E(|Y_{jk_1}^i||Y_{jk_2}^i|)^2$ is also bounded. Therefore, $E(|Y_{jk_1}^i||Y_{jk_2}^i||Y_{jk_3}^i|)$ is bounded. Consequently, $E\Big[|Y_{jk_1}^i| \cdots |Y_{jk_h}^i|\Big]$ is bounded for $h = 1, 2, 3, 4$. ■

⟨eigen⟩

**Proposition 6.1** *Let $E$ be a Euclidean space and let $F \subset E$ be a linear subspace. Let $p_F$ denote the orthogonal projection of $E$ onto $F$. Let $g$ be a linear symmetric operator $g : E \to E$ and consider the linear application $f$ of $F$ into itself defined by*

$$f : x \in F \to f(x) = p_F \circ g(x)$$

*Then, we have that if $\lambda_1 < \lambda_2 < \cdots < \lambda_m$ are the eigenvalues of $g$ and $\mu_1 < \mu_2 < \cdots < \mu_n$ are the eigenvalues of $f$, $n < m$, then for any $j = 1, 2, \cdots, m$, the following inequality holds*

$$\min\{\mu_k | k = 1, 2, \cdots, n\} \leq \lambda_j \leq \max\{\mu_k | k = 1, 2, \cdots, n\}.$$

**Proof.** We prove is first for $m = dim(F) = dim(F) - 1$. Let $e = (e_1, e_2, \cdots, e_m)$ be an orthonormal basis of $F$ such that basis the matrix representative of $f$ is a diagonal $[f]_e^e = diag(\lambda_1, \lambda_2, \cdots, \lambda_m)$ and let $e_0 \in E$ be such that $e' = (e_0, e_1, e_2, \cdots, e_m)$ is an orthonormal basis of $E$. Then in that basis, the matrix representative of $g$ is

$$[g]_{e'}^{e'} = \begin{pmatrix} a & b_1 & \cdots & b_{S_i} \\ b_1 & \lambda_1 & 0 & 0 \\ \cdots & \cdots & \ddots & 0 \\ b_{S_i} & \cdots & 0 & \lambda_m \end{pmatrix}. \quad (23)$$

?diag2?

We see here that the matrix representative of $g$ is a submatrix of the matrix representative of $f$. By the interlacing property of the eigenvalues, we have

$$\mu_0 \leq \lambda_1 \leq \mu_1 \leq \cdots \leq \mu_1 \leq \lambda_1 \leq \mu_m.$$

If $dim(E) - dim(F) > 1$, we iterate the process by induction on $dim(E) - dim(F)$ and complete the proof. ■

⟨positive⟩

**Lemma 6.3** *For any $i \in V$, let $T^i(\gamma^i)$ be a symmetric matrix with dimension $p_i$. Then there exists a constant $\eta$ such that with $||T^i(\gamma^i)||_F \leq \eta$, the matrix $I_{p_i} + T^i\Sigma_0^i$ is positive definite.*

**Proof.** If we want to show $I_{p_i} + T^i(\gamma^i)\Sigma_0^i$ is positive definite, it is equivalent to show for any non zero vector $z$ with dimension $p_i$, $z^t(I_{p_i} + T^i\Sigma_0^i)z$ is positive. By Cauchy Schwarz inequality, we have

$$\begin{aligned} |< T^i(\gamma^i)z, \Sigma_0^i z >| &\leq ||T^i(\gamma^i)z|| \times ||\Sigma_0^i z|| \leq ||T^i(\gamma^i)|| \times ||z|| \times ||\Sigma_0^i|| \times ||z|| \\ &\leq ||z||^2 ||T^i(\gamma^i)||_F \times \frac{1}{\kappa_1} \leq \eta ||z||^2 \frac{1}{\kappa_1} \end{aligned}$$

Therefore,

$$\begin{aligned} z^t(I_{p_i} + T^i(\gamma^i)\Sigma_0^i)z &= z^t I_{p_i} z + z^t T^i(\gamma^i)\Sigma_0^i z = ||z||^2 + < T^i(\gamma^i)z, \Sigma_0^i z > \\ &\geq ||z||^2 - ||z||^2 \eta \frac{1}{\kappa_1} = [1 - \eta \frac{1}{\kappa_1}]||z||^2. \end{aligned}$$

We can thus choose a constant $\eta$, such that $\eta < \kappa_1$. It follows $z^t(I_{p_i} + T^i\Sigma_0^i)z \geq ||z||^2 > 0$ when $||T^i(\gamma^i)||_F \leq \eta$. ■

⟨He⟩

**Lemma 6.4** *For any $i \in V$, $\log \int ||\sqrt{n}J^i(\theta^i - \theta_0^i)||\pi_0^i(\theta^i)d\theta^i \leq \exp[M_7 p^2 \log p]$ with $M_7$ is a constant.*

**Proof.**

$$\begin{aligned} &\int ||\sqrt{n}J^i(\theta^i - \theta_0^i)||\pi_0^i(\theta^i)d\theta^i \\ \leq\ & n^{\frac{1}{2}} \frac{1}{\kappa_1} \frac{2^{S-s+p}}{p^p\Gamma(S)} [\prod_{r=1}^{s} (\frac{2}{\tau_r})^{\frac{k_r}{2}+1} \Gamma(\frac{k_r}{2}+1)](M_0 p\Gamma(\alpha p + S) + 2\Gamma(\alpha p + 1 + S)) \\ \leq\ & n^{\frac{1}{2}} \frac{1}{\kappa_1} \frac{2^{S-s+p}}{p^p\Gamma(S)} [\prod_{r=1}^{s} (\frac{2}{\tau_r})^{\frac{k_r}{2}+1} \Gamma(\frac{k_r}{2}+1)] M_2 p\Gamma(\alpha p + 1 + S) \end{aligned}$$

Therefore,

$$\begin{aligned}
&\log\{\int ||\sqrt{n}J^i(\theta^i-\theta^i_0)||\pi^i_0(\theta^i)d\theta^i\} \\
\leq\ & \frac{1}{2}\log n - \log\kappa_1 + (S-s+p)\log 2 - p\log p - \log\Gamma(S) + \log M_2 + \log p + \log\Gamma(\alpha p+1+S) \\
&+\sum_{r=1}^{s}[(\frac{k_r}{2}+1)\log\frac{2}{\tau_r} + \log\Gamma(\frac{k_r}{2}+1)]
\end{aligned}$$

Since $\log n$ and $\log p$ is the same order and $k_r \leq p$, we have that

$$\begin{aligned}
&\exp[\log\{\int ||\sqrt{n}J^i(\theta^i-\theta^i_0)||\pi^i_0(\theta^i)d\theta^i\}] \\
\leq\ & \exp[\frac{1}{2}\log p + (S-s+p)\log 2 + \log p + \log\Gamma(\alpha p+1+S) + p(\frac{p}{2}+1)\log 2 + p\log\Gamma(\frac{p}{2}+1) + M_3] \\
\leq\ & \exp[\frac{3}{2}\log p + (S-s+p)\log 2 + \log\Gamma(\alpha p+1+S) + p(\frac{p}{2}+1)\log 2 + p\log\Gamma(\frac{p}{2}+1) + M_3]
\end{aligned}$$

By Sterling's approximation, we have $\log n! = n\log n + O(\log n)$. Therefore, there exist two constant $M_5$ and $M_6$ such that

$$\log\Gamma(\alpha p+1+S) \leq \log\Gamma(\alpha p+1+\frac{p(p+1)}{2}) \leq \log(\alpha p+p^2)! \leq M_5p^2\log p$$

and

$$\log\Gamma(\frac{k_r}{2}+1) \leq \log\Gamma(\frac{p}{2}+1) \leq \log p! \leq M_6 p\log p.$$

Combining all results above, we obtain that

$$\begin{aligned}
&\exp[\log\{\int ||\sqrt{n}J^i(\theta^i-\theta^i_0)||\pi^i_0(\theta^i)d\theta^i\}] \\
\leq\ & \exp[\frac{3}{2}\log p + [\frac{p(p+1)}{2}+p]\log 2 + M_5p^2\log p + p(\frac{p}{2}+1)\log 2 + M_6p^2\log p + M_3] \\
\leq\ & \exp[M_7p^2\log p],
\end{aligned}$$

where $M_7$ is a constant. ■

# References


Bickel69 Bickel, P. J., and Yahav, J. A. (1969). Some contributions to the asymptotic theory of Bayes solutions. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 11(4), 257-276.

Xin15 Gao, X. and Carroll, R. J. (2015). Data Integration with High Dimensionality. submitted.

Ghosal00 Ghosal, S. (2000). Asymptotic normality of posterior distributions for exponential families when the number of parameters tends to infinity. Journal of Multivariate Analysis, 74(1), 49-68.

?Lau07? Højsgaard, S. and Lauritzen, S. L. (2007). Inference in graphical Gaussian models with edge and vertex symmetries with the gRc package for R. Journal of Statistical Software, 23(6), 1-26.

Lau08 Højsgaard, S. and Lauritzen, S. L. (2008). Graphical Gaussian models with edge and vertex symmetries. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 70(5), 1005-1027.

Ibragimov73 Ibragimov, I. A., and Khas' Minskii, R. Z. (1973). Asymptotic behavior of some statistical estimators II. Limit theorems for the a posteriori density and Bayes' estimators. Theory of Probability and Its Applications, 18(1), 76-91.

Lehmann98 Lehmann, E. and Casella, G. (1998). Theory of Point Estimation. Springer, Berlin.

Helene15 Massam, H., Li, Q. and Gao, X. (2015). Bayesian precision matrix estimation for graphical Gaussian models with edge and vertex symmetries. arXiv preprint arXiv:1506.04347v1.

Meng14 Meng, Z., Wei, D., Wiesel, A., and Hero, A. O. (2014). Marginal likelihoods for distributed parameter estimation of Gaussian graphical models. Signal Processing, IEEE Transactions on, 62(20), 5425-5438.

Mui82 Muirhead, R. J. (1982). Aspects of Multivariate Statistical Theory. John Wiley and Sons.

?Orte70? Ortega, J. M., and Rheinboldt, W. C. (1970). Iterative solution of nonlinear equations in several variables (Vol. 30). Siam.

?Portnoy88? Portnoy, S. (1988). Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. The Annals of Statistics, 16(1), 356-366.

Spokoiny13 Spokoiny, V., and Zhilova, M. (2013). Sharp deviation bounds for quadratic forms. Mathematical Methods of Statistics, 22(2), 100-113.

?Van98? Van der Vaart, A. W. (1998). Asymptotic Statistics. Cambridge University Press.

?Wang86? Wang, S. D., Kuo, T. S., and Hsu, C. F. (1986). Trace bounds on the solution of the algebraic matrix Riccati and Lyapunov equation. Automatic Control, IEEE Transactions on, 31(7), 654-656.

?Orte70? Ortega, J. M., and Rheinboldt, W. C. (1970). Iterative solution of nonlinear equations in several variables (Vol. 30). Siam.