BAYESIAN ESTIMATION OF GRAPHICAL GAUSSIAN MODELS WITH EDGES AND VERTICES SYMMETRIES

QIONG LI

A DISSERTATION SUBMITTED TO
THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN MATHEMATICS AND STATISTICS
YORK UNIVERSITY
TORONTO, ONTARIO

November 2016

Abstract

In this thesis, we consider the Bayesian analysis of undirected graphical Gaussian models with edges and vertices symmetries. The graphical Gaussian models with equality constraints on the precision matrix, that is the inverse covariance matrix, were introduced by Højsgaard and Lauritzen [2008] as RCON models. The models can be represented by colored graphs, where edges or vertices have the same color if the corresponding entries of the precision matrix are equal. In this thesis, we define a conjugate prior distribution for the precision matrix in RCON models. For simplicity, we will call this prior the colored G-Wishart distribution.

We begin with the study of the sampling scheme for the colored G-Wishart distribution. This sampling method is based on the Metropolis-Hastings algorithm and the Cholesky decomposition of matrices. In order to assess the accuracy of the Metropolis-Hastings sampling method, we compute the expected values of the precision matrix in the colored G-Wishart distribution for some particular colored graphs: trees, star graphs, a complete graph with 3 vertices and a decomposable model on 4

vertices. Moreover, the simulation results for comparing the true mean of the precision matrix K in the colored G-Wishart distribution with the sample mean of K obtained from our Metropolis-Hastings algorithm are presented.

Next, we propose the distributed Bayesian estimate of the precision matrix for large colored graphical Gaussian models. We also study the asymptotic behavior of our proposed estimate under the regular asymptotic regime where the variable dimension p is fixed and under the double asymptotic regimes where both p and the sample size n go to infinity. The proofs of the asymptotic properties of the distributed estimate are provided.

Evaluating the normalizing constant is important and necessary for obtaining the posterior distribution and the marginal probability of the likelihood. We give three methods, the Monte Carlo method, the importance sampling and the Laplace approximation, for estimating the normalizing constant of the colored G-Wishart distribution. We then apply these methods on the model search for a real dataset using Bayes factors.

Keywords: asymptotic normality, Bayesian estimator, colored G-Wishart distribution, conditional independence, conjugate prior, consistency, marginal model, Metropolis-Hastings, large deviation, symmetry constraint.

Acknowledgements

First and foremost, I would like to express my deepest appreciation to my two supervisors, Professor Hélène Massam and Professor Xin Gao. Their deep and extensive knowledge and invaluable guidance enabled me to understand the amazing area of graphical models. This thesis would not have been possible without their contributions of time, idea and funding. I truly respect them for their brilliant insights and enthusiasm on research.

It is my great pleasure to thank Professors Steven Wang and Dong Liang for serving on the supervisory committee for my Ph.D. program. I am also grateful to all faculty members, staffs and fellow graduate students at the department of Mathematics and Statistics at York University.

Last, but not least, I deeply thank my parents, my parents-in-law and my beloved husband for their continuous supports and encouragement. This thesis is dedicated to my lovely daughter for continuing to show us what is important and what is not.

Table of Contents

List of Tables

List of Figures

1 Introduction and Notations

1.1 Introduction

Graphical Gaussian models, also known as covariance selection models [Dempster, 1972], is a family of probability distributions in which the dependencies or independencies among continuous random variables are expressed by an underlying graph. It provides an efficient framework for compactly modeling the complex joint distribution by means of their conditional dependency graph. Therefore, it has become a powerful tool of modern statistics for analyzing and representing complex high-dimensional data. Such models are commonly used in so many different fields, including biology, medicine, computer vision and statistical physics. In graphical Gaussian models, each vertex represents a random variable, and the absent of edge (i, j) indicates the conditional independence of the variable X_i and the variable X_j given all remaining variables.

Højsgaard and Lauritzen [2008] introduced graphical Gaussian models with edge and vertex symmetries in order to reduce the number of parameters in graphical Gaussian models. The models generalize graphical Gaussian models and can be represented by colored graphs, where vertices or edges are restricted to being identical coloring if the associated parameters are equal. These models are defined as graphical Gaussian models with three different types of equality constraints: equality of specified entries of the precision matrix (RCON), equality of specified entries of the correlation matrix (RCOR) and equality of specified entries of the covariance matrix generated by permutation symmetry (RCOP). The combination of symmetric restrictions and conditional independent restrictions results in the reduction in the number of parameters and makes the models efficient and flexible.

In recent years, many methods which facilitate Bayesian inference have been developed using graphical Gaussian models. Bayesian analysis of undirected graphical Gaussian models has been considered by Dawid and Lauritzen [1993]. In particular, Dawid and Lauritzen [1993] mainly focused on the decomposable graphs and introduced the hyper inverse Wishart distribution as the conjugate prior for the covariance matrix. Roverato [2002] further generalized the hyper inverse Wishart distribution to arbitrary graphs. Atay-Kayis and Massam [2005] as well as Letac and Massam [2007] continued these developments and termed this distribution the G-Wishart as a prior

specified for the inverse covariance matrix. A number of sampling methods for the G-Wishart distribution associated with an arbitrary graph have been proposed [see, e.g., Piccioni, 2000, Mitsakakis et al., 2011, Dobra et al., 2011, Carvalho et al., 2007, Wang and Li, 2012, Lenkoski, 2013]. The existing methods expand the usefulness of the G-Wishart distribution and provide a statistical tool to the estimation of the posterior mean for the covariance matrix in a Bayesian framework.

In this thesis, we work with RCON models in a Bayesian framework. In order to identify the conditional independencies between the random variables in the RCON models, we concern the problem of estimating the elements of the inverse covariance matrix, which commonly referred to as the concentration or precision matrix. Since RCON models belong to an exponential family, we use the Diaconis and Ylvisaker [1979] (henceforth abbreviated DY) conjugate prior for the precision matrix. This yields a distribution similar to the G-Wishart distribution but with the colored constraints on the edges and vertices. We call this distribution the colored G-Wishart distribution and further propose a method to sample from the colored G-Wishart distribution. Our sampling scheme is an adaptation of the Metropolis-Hastings algorithm proposed by Mitsakakis et al. [2011] for the G-Wishart distribution in uncolored models. Extensive numerical experiments demonstrate our proposed sampling method performs very well by comparing the true mean of the precision matrix K

in the colored G-Wishart distribution with the sample mean of K obtained through our Metropolis-Hastings algorithm.

At the end, we propose an efficient algorithm to estimate the precision matrix for high-dimensional setting, which is distributed estimation. The idea behind the distributed estimation is that the estimation of the precision matrix is separated to smaller local models from which we can estimate parts of the parameters of the global model. The estimates of parameters from local models are then combined together to yield an estimate of the global model. This efficient computational method was first proposed by Meng et al. [2014] for the maximum likelihood estimate in graphical Gaussian models and we adapt it here to graphical Gaussian models with edges and vertices symmetries in a Bayesian framework. We consider the asymptotic behavior of our proposed estimators when the number of variables p is fixed and the sample size n grows to infinity. We further derive high-dimensional convergence rates when both p and n are large. We also demonstrate numerically how our method can scale up to any dimension by looking at colored graphical Gaussian models represented by large colored cycles and also by a colored 10×10 grid. The simulation study demonstrates that our method produces statistically efficient estimators of the precision matrix for the colored graphical Gaussian models.

1.2 Notations and Preliminaries

This chapter covers some terminologies and known results which are going to be used in the rest of this thesis. They are involved in graph theory, Gaussian graphical models, colored Gaussian graphical models and the colored G-Wishart prior. Further details and explanations are available in Højsgaard and Lauritzen [2008] and Lauritzen [1996].

1.2.1 Preliminaries

We summarize here the notations to be used throughout this thesis. We write $f(n) = O(g(n))$ if and only if $f(n)/g(n)$ is bounded as $n \to \infty$. We write $f(n) = o(g(n))$ if and only if $f(n)/g(n) \to 0$ as $n \to \infty$. The notation $X_n = O_p(a_n)$ means that, for any $\varepsilon > 0$, there exists a finite $M > 0$ such that $P(|X_n/a_n| > M) < \varepsilon$ for any n. The cardinality of a set A is denoted by $|A|$ and the difference of two sets A and B is denoted by $A \backslash B$. If U and V are square matrices with the same dimension, then we use $tr(UV) = \sum_{i,j} U_{ij}V_{ij}$ to denote the trace of UV and $|U|$ to denote the determinant of the matrix U. Let $\lambda(U)$, $\lambda_{max}(U)$ and $\lambda_{min}(U)$ stand for the eigenvalues, the largest and smallest eigenvalues of U, respectively. Following the standard notation, $U_{A,B}$ represents a submatrix of U with rows indexed by A and columns indexed by B. The identity matrix of order p is denoted by I_p. For

a vector $x = (x_1, x_2, \ldots, x_p)$, $||x||$ stands for its Euclidean norm $(\sum_{i=1}^{p} x_i^2)^{1/2}$. For a square matrix U, $||U||$ stands for its operator norm defined by $\sup\{||Ux|| : ||x|| \leq 1\}$. Let $\xrightarrow{\mathcal{L}}$ and $\xrightarrow{p}$ denote the convergence in distribution and in probability, respectively. The superscript t denotes the transpose.

1.2.2 Graph Theory

A *graph* G is an ordered pair $G = (V, E)$ consisting of a nonempty set V of vertices and an edge set E disjoint from V. A graph is *undirected* if the edge set is composed of unordered vertex pairs. Two vertices v, $v' \in V$ are said to be *adjacent* if there is an edge between v and v'. The degree of a vertex v in a graph G is the number of edges of G incident with v. A graph $G = (V, E)$ is called *complete* if every pair of distinct vertices of G are adjacent. A graph is *connected* if any two vertices are linked by a path. A connected graph without cycles is called a *tree*. A *star* graph is a tree which consists of a single vertex with degree at least 2.

1.2.3 Graphical Gaussian Models

This thesis is concerned with graphical Gaussian models, where the variables in the model are jointly Gaussian. The graphical Gaussian model is also known as a covariance selection model [Dempster, 1972]. Let $X^1, X^2, \ldots, X^n$ be independent

and identically distributed p-dimensional random variables following a multivariate normal distribution $N_p(\mu, \Sigma)$. The inverse covariance matrix Σ^{-1} is called the *precision* matrix and we denote Σ^{-1} by $K = (K_{ij})_{p\times p}$, where K_{ij} stands for the (i, j) entry of K. Since the precision matrix is the primary goal, we can assume that these models are centred $N_p(0, \Sigma)$ without any loss of generality. In multivariate Gaussian analysis, the independence and conditional independence relationships between the variables can be represented by means of an undirected graph $G = (V, E)$, where $V = \{1, 2, \ldots, p\}$ and E are the sets of vertices and edges, respectively. For $X = (X_v, v \in V)$ and for any pair $(i, j) \notin E$, $i \neq j$, if the random variable X_i and X_j are conditionally independent given all the other variables $X_{V\setminus\{i,j\}}$, we say that the distribution of X is *Markov* with respect to G. Such models for X are called graphical Gaussian models. The conditional independences can be denoted by

$$(i, j) \notin E \Rightarrow X_i \perp X_j \mid X_{V\setminus\{i,j\}}.$$

The nonzero elements of K are associated with edges in E. A missing edge (i, j) in E implies $K_{ij} = 0$ and corresponds to the conditional independence of univariate elements X_i and X_j given the remaining elements. Since the conditional independence of the variables X_i and X_j is equivalent to $K_{ij} = 0$, if we let P_G be 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

$$\mathcal{N}_G = \{N(0, \Sigma) | K \in P_G\}. \tag{1.1}$$

Therefore, the joint density of X for the sample size n can be written as

$$p(X|K) = \frac{|K|^{n/2}}{(2\pi)^{pn/2}} \exp\{-\frac{1}{2}\sum_{i=1}^{n}(X^i)^t K X^i\}\mathbf{1}_{K \in P_G},$$

where $\mathbf{1}_A$ is an indicator function of the set A.

1.2.4 Colored Graphical Gaussian Models

The graphical Gaussian models with edge and vertex symmetries, which we here call the colored graphical Gaussian models, have been introduced by Højsgaard and Lauritzen [2008]. These models are defined as graphical Gaussian models with three different types of symmetry constraints: equality of specified entries of the precision matrix K, equality of specified entries of the correlation matrix and equality of specified entries of K generated by a subgroup of the automorphism group of G. These models are denoted as RCON, RCOR and RCOP models for short, respectively. In this thesis, we only consider RCON models. The model can be represented by colored graphs, where edges or vertices have the same color if the corresponding entries of the precision matrix are equal. Now, we define the RCON model as follows. Let $\mathcal{V} = \{V_1, \ldots, V_k\}$ be a partition of V and $\mathcal{E} = \{E_1, \ldots, E_l\}$ be a partition of E. If all

the vertices belonging to a vertex color class V_i of $\mathcal{V}$ have the same color, we say that $\mathcal{V}$ is a coloring of V. Similarly, if all the edges belonging to an edge color class 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 (1.1) is imposed with the following additional restrictions

(C_1) if m is a vertex class in $\mathcal{V}$, then for all $i \in m$, K_{ii} are equal, and

(C_2) 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 colored 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 the positive symmetric matrix with zero and colored constraints. When drawing a colored graph, we use black for color classes with only one element. Thus two vertices displayed in black will be in different color classes. Figure 1.1 illustrates a colored graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{\{1, 2, 3\}, \{4, 5\}, \{6\}, \{7\}\}$ and $\mathcal{E} = \{\{12\}, \{23\}, \{13\}, \{34, 45, 46, 47\}\}$.

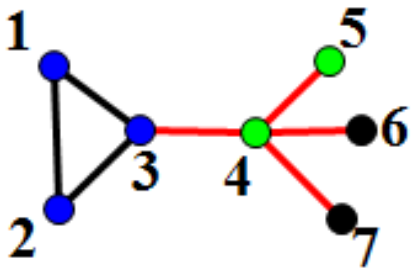


Figure 1.1: The colored graph in an RCON model.

1.2.5 Colored G-Wishart Distributions

In 1979, Diaconis and Ylvisaker [1979] derived the standard conjugate priors for the exponential family distribution. Since the graphical Gaussian model is also a regular exponential family, then the DY conjugate prior for the graphical Gaussian model is called G-Wishart distribution with the density

$$p(K|\delta, D) = \frac{1}{I_G(\delta, D)}|K|^{(\delta-2)/2}\exp\{-\frac{1}{2}tr(KD)\}\mathbf{1}_{K\in P_G} \tag{1.2}$$

where $\delta > 0$ and D, a symmetric positive definite $p \times p$ matrix, are hyperparameters of the prior distribution and $I_G(\delta, D)$ is the normalizing constant, namely,

$$I_G(\delta, D) = \int_{P_G} |K|^{(\delta-2)/2}\exp\{-\frac{1}{2}tr(KD)\}dK. \tag{1.3}$$

Next, we will define a colored G-Wishart distribution for the colored graphical Gaussian models in terms of the G-Wishart distribution. The density of colored G-Wishart

can be written as

$$p(K|\delta, D) = \frac{1}{I_{\mathcal{G}}(\delta, D)}|K|^{(\delta-2)/2}\exp\{-\frac{1}{2}tr(KD)\}\mathbf{1}_{K\in P_{\mathcal{G}}} \tag{1.4}$$

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

$$I_{\mathcal{G}}(\delta, D) = \int_{P_{\mathcal{G}}} |K|^{(\delta-2)/2}\exp\{-\frac{1}{2}tr(KD)\}dK. \tag{1.5}$$

After choosing the colored G-Wishart distribution as the prior, the posterior distribution of K can be expressed as

$$\begin{aligned}\pi^*(K|\delta, D) &\propto |K|^{\frac{n}{2}}\exp\{-\frac{1}{2}tr(K\sum_{i=1}^{n}X^i(X^i)^t)\} \times |K|^{\frac{\delta-2}{2}}\exp\{-\frac{1}{2}tr(KD)\}\mathbf{1}_{K\in P_{\mathcal{G}}} \\ &= |K|^{\frac{n+\delta-2}{2}}\exp\{-\frac{1}{2}tr\big(K(D+\sum_{i=1}^{n}X^i(X^i)^t)\big)\}\mathbf{1}_{K\in P_{\mathcal{G}}}.\end{aligned}$$

1.2.6 Related Theorems and Known Results

Definition 1.2.1 *(Lexicographical order) Given two partially ordered sets A and B, the lexicographical order on the Cartesian product $A \times B$ is defined as $(a_1, b_1) \leq (a_2, b_2)$ if and only if either*

(1) $a_1 < a_2$ or

(2) $a_1 = a_2$ and $b_1 < b_2$.

Theorem 1.2.1 *(**Isserlis' Theorem**) Let $X = (X_1, X_2, \cdots, X_m)$ 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 σ 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 σ, $A^{(\sigma)} = \prod_{k=1}^{n} \Sigma_{\sigma(2k-1)\sigma(2k)}$.

For example,

$$E[X_1 X_2 X_3 X_4] = \Sigma_{12}\Sigma_{34} + \Sigma_{13}\Sigma_{24} + \Sigma_{14}\Sigma_{23}$$

and

$$\begin{aligned} E[X_1 X_2 X_3 X_4 X_5 X_6] &= \Sigma_{12}(E[X_3 X_4 X_5 X_6]) + \Sigma_{13}(E[X_2 X_4 X_5 X_6]) \\ &\quad + \Sigma_{14}(E[X_2 X_3 X_5 X_6]) + \Sigma_{15}(E[X_2 X_3 X_4 X_6]) \\ &\quad + \Sigma_{16}(E[X_2 X_3 X_4 X_5]). \end{aligned}$$

Theorem 1.2.2 *(Cauchy-Schwarz inequality) Let u and v be two vectors in R^n. Then*

$$| < u, v > | \leq ||u|| \times ||v||$$

where $< u, v >$ *denotes the inner product of* u *and* v.

If let $u = (1, 1, \cdots, 1)^t$ be a n-dimensional vector and $v = (|v_1|, |v_2|, \cdots, |v_n|)^t$, then Cauchy-Schwarz inequality implies

$$\sum_{i=1}^{n} |v_i| \leq \sqrt{n}||v||.$$

The Delta Method:

Now suppose $\boldsymbol{\theta} \in R^k$ and we have an asymptotically normal estimator $\hat{\boldsymbol{\theta}}$ such that

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \boldsymbol{\Sigma}).$$

Let $\boldsymbol{\eta} = \boldsymbol{g}(\boldsymbol{\theta})$: $R^k \to R^j$; i.e., $\boldsymbol{\eta} = \boldsymbol{g}(\boldsymbol{\theta}) = (g_1(\boldsymbol{\theta}), g_2(\boldsymbol{\theta}), \ldots, g_j(\boldsymbol{\theta}))^t$, denote the parameter of interest where $\boldsymbol{\eta} \in R^j$ and $j \leq k$. Assume the $\boldsymbol{g}(\boldsymbol{\theta})$ is continuous with continuous first derivatives

$$\frac{\partial \boldsymbol{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta^t}} = \begin{pmatrix} \frac{\partial g_1(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial g_1(\boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial g_1(\boldsymbol{\theta})}{\partial \theta_k} \\ \frac{\partial g_2(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial g_2(\boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial g_2(\boldsymbol{\theta})}{\partial \theta_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_j(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial g_j(\boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial g_j(\boldsymbol{\theta})}{\partial \theta_k} \end{pmatrix}.$$

Then

$$\sqrt{n}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}) = \sqrt{n}(\boldsymbol{g}(\hat{\boldsymbol{\theta}}) - \boldsymbol{g}(\boldsymbol{\theta})) \xrightarrow{\mathcal{L}} N(\mathbf{0}, (\frac{\partial \boldsymbol{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta^t}})\boldsymbol{\Sigma}(\frac{\partial \boldsymbol{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta^t}})^t).$$

1.3 Review of Literature

Many statistical problems require at some point the estimation of population covariance matrices from samples of multivariate data. However, when the number of variables p increases, the number of unknown parameters $\frac{1}{2}p(p+1)$ in the covariance matrix increases quadratically with p. Efficient estimation of population covariance matrices becomes a difficult statistical problem when p increases.

1.3.1 Graphical Gaussian Models

In order to efficiently and parsimoniously estimate the covariance matrix Σ, Dempster [1972] first proposed estimating the covariance matrix parsimoniously by setting off-diagonal elements of the precision matrix $K = \Sigma^{-1}$ to zero. The reason for adopting such a model is that in many problems the precision matrix has a large number of zeros in its off diagonal elements and these should be exploited in the estimation [Cox and Wermuth, 1996]. This model is often referred to as the covariance selection model. Dempster [1972] considered the exponential family of the Gaussian distribution with unknown covariance parameters. The density is represented by the family of continuous densities

$$f(x, \Sigma) = (\frac{1}{2\pi})^{\frac{1}{2}p}(\frac{1}{|\Sigma|})^{\frac{1}{2}} \exp(-\frac{1}{2}x^t\Sigma^{-1}x).$$

It is a representation of the density f as a member of an exponential family of distributions:

$$\exp[\alpha_0 + t(x) + \alpha_{11}t_{11}(x) + \ldots + \alpha_{rs}t_{rs}(x)]$$

with here $\alpha_{ij} = \Sigma_{ij}^{-1}$, $\alpha_0 = -\frac{1}{2}p\log 2\pi - \frac{1}{2}\log|\Sigma|$, $t_{ij}(x) = -x_i x_j$ for $i \neq j$, $t_{ii}(x) = -\frac{1}{2}x_i^2$ and $t(x) = 0$. Suppose there are $m+1$ observations on p random variables. The estimate of the sample covariance matrix is

$$S = \frac{1}{m}\sum_{i=1}^{m+1}(x_i - \bar{x})(x_i - \bar{x})^t,$$

where $\bar{x} = \frac{1}{m+1}\sum_{i=1}^{m+1} x_i$. Let I be a subset of index pairs (j,k) such that K_{jk} is zero, and J be the remaining set of pairs (j,k) such that K_{jk} is not zero. Dempster [1972] chose $\hat{\Sigma}$ to be a positive definite symmetric matrix such that S and $\hat{\Sigma}$ are identical for index pairs (j,k) in J while K is identically 0 for index pairs (j,k) in I.

Several researchers [Cox and Wermuth, 1996, Whittaker, 1990, Lauritzen, 1996] also called the Gaussian model with a pattern of zero constraints in K as the graphical Gaussian model since it represents a pairwise conditional independence structure. A graphical Gaussian model is represented by an undirected $G = (V, E)$, where V contains p vertices and the edge E describes the conditional independence relationship among the random variables $X = \{X_1, X_2 \ldots, X_p\}$. The edge between X_i and X_j is absent if and only if X_i and X_j are conditionally independent given all the other variables $X_{V\setminus\{i,j\}}$, and $K_{ij} = 0$.

1.3.2 Bayesian Methods

A number of articles took a Bayesian approach to graphical Gaussian models. Early work on the Bayesian estimation for graphical Gaussian models has largely focused on decomposable graphs. When the graph G is decomposable, Dawid and Lauritzen [1993] proposed a convenient prior based on the factorisations of the likelihood in terms of the cliques and separators of the underlying graph G. The class of priors is specified over the covariance matrix Σ as well as on the graphical structure G, which is named hyper inverse Wishart distribution. Although the priors enjoy many advantages, such as the computational efficiency due to its conjugate nature and the exact calculation of marginal likelihoods, they are sometimes inflexible since this method can deal only with decomposable graphical models. To solve this, Roverato [2002] generalized the priors over arbitrary graphs and showed that the hyper inverse Wishart prior for the covariance matrix is equivalent to a constrained Wishart prior for the precision matrix. This prior is called the G-Wishart distribution. Although it is straightforward to define a constrained prior distribution for arbitrary graphs in graphical Gaussian models, until recently, the normalizing constants of such distributions could not be exactly computed unless the graph is decomposable. Using an iterative method and some special functions, Uhler et al. [2014] seem to have solved this very difficult problem. Roverato [2002] and Atay-Kayis and Massam [2005] also

proposed different Monte Carlo methods and efficient simulations for estimating the normalizing constants for the non-decomposable graph.

There have been several proposed sampling methods for the G-Wishart distribution. Piccioni [2000] proposed the block Gibbs sampler using the Bayesian iterative proportional scaling. This sampler updates K according to its clique decomposition and matrix inversion. Since identifying all cliques and inverting the large matrix are computationally expensive, it is not suitable for high-dimensional problems. The related sampling method is the rejection sampling method developed by Wang and Carvalho [2010]. Implementation of this method relies on the junction tree representation of graphs through the local computation. In order to avoid the calculation of the posterior normalizing constants, the authors in Mitsakakis et al. [2011] and Dobra et al. [2011] proposed the MH algorithms for the G-Wishart distribution. These methods are based on the matrix decomposition and matrix completion developed in Roverato [2002] and Atay-Kayis and Massam [2005]. A direct sampler for the G-Wishart distribution was recently proposed by Lenkoski [2013] that is closely related to the block Gibbs sampler of Piccioni [2000]. Unlike the block Gibbs sampler, however, each sample is drawn independently from previous samples.

1.3.3 Graphical Gaussian Models with Edge and Vertices Symmetries

Symmetry restrictions for the multivariate Gaussian distribution have a long history dating back to Wilks [1946]. Wilks [1946] was the first to add the symmetry restrictions on the covariance matrix such that the covariance matrix with equal diagonal elements and equal off-diagonal elements. Symmetry restrictions on the multivariate Gaussian distribution are also considered by several authors [Andersson, 1975, Andersson et al., 1983, Jensen, 1988, Olkin, 1969]. Graphical Gaussian models with symmetry restrictions were first considered by Hylleberg et al. [1993]. Hylleberg et al. [1993] combined the conditional independence restrictions with the group symmetry restrictions. Subsequently, Andersen et al. [1995] and Madsen [2000] also considered such models. More recently, Højsgaard and Lauritzen [2008] considered graphical Gaussian models with symmetry constraints not necessarily described by a group action. The symmetry is given by the equality of certain entries either in the covariance, the correlation or the precision matrices. Models for the multivariate random variable $X = (X_i, i \in V)$ Markov with respect to G and with edges and vertices symmetries are called colored graphical Gaussian models. These 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 the variance of the corresponding variables can 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. Højsgaard and Lauritzen [2008] also developed algorithms to compute the maximum likelihood estimate of the covariance, the correlation or the precision matrix.

2 A Sampling Method: Metropolis-Hastings

In this chapter, following what has been done by Mitsakakis et al. [2011], we want to develop a Metropolis-Hastings (MH) algorithm to obtain random samples from the colored G-Wishart distribution. According to Atay-Kayis and Massam [2005], we first consider the Cholesky decomposition of D^{-1} and K. Denote

$$D^{-1} = Q^t Q \qquad \text{and} \qquad K = \Phi^t \Phi \tag{2.1}$$

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 elements. Then, we will express the density of the colored G-Wishart distribution in terms of the new variable

$$\Psi = \Phi Q^{-1}. \tag{2.2}$$

Finally, we use the MH algorithm to generate the samples of Ψ. After drawing the random samples of Ψ, the samples of K can be obtained by $K = Q^t(\Psi^t \Psi)Q$.

The advantage of changes of variables from K to Ψ is that it keeps the samples for K positive definite. Also, by change to Ψ, we found that (see next section)

to sample K, we need to deal with normal and chi-squared distributions which we know how to sample. In the next section, we will derive the density of the colored G-Wishart distribution in terms of the new variable Ψ. The density can be written as a multiplication of the densities of chi-squares, normals and a function of free elements of Ψ. This give us the idea about how to choose the proposal distribution in the MH algorithm.

2.1 The Density of the Colored G-Wishart

We denote

$$v_u(G) = \min\{(i,j) \in u | i \leq j,\ u \in \mathcal{V} \cup \mathcal{E}\}$$

where the minimum is defined according to the lexicographical order and define

$$v(G) = \bigcup_{u \in \mathcal{V} \cup \mathcal{E}} v_u(G).$$

Let $K^{v(G)} = \{K_{ij} |\ (i,j) \in v(G)\}$ be the free entries of K. The zero and coloring constraints on the entries of K associated with a colored graph $\mathcal{G}$ determine the free entries $\Phi^{v(G)} = \{\Phi_{ij} | (i,j) \in v(G)\}$ and $\Psi^{v(G)} = \{\Psi_{ij} | (i,j) \in v(G)\}$ of the matrices Φ and Ψ, respectively. Each non-free entry Φ_{ij} and Ψ_{ij} with $(i,j) \notin v(G)$ is a function of the free entries Φ_{ij} and Ψ_{ij} with $(i,j) \in v(G)$ that precede it in the lexicographical order. The following two propositions give the expression of the non-free entries of

Φ and Ψ in function of the free ones of $\Phi^{v(G)}$ and $\Psi^{v(G)}$ and the free entries of K. The first part of the equations in each proposition can be found in Roverato [2002].

Proposition 2.1.1 *Let $K = \Phi^t\Phi$ be an element of $P_{\mathcal{G}}$, and $(i_u, j_u) = \min\{(i,j) \in u | i \leq j$ and $u \in \mathcal{V} \cup \mathcal{E}\}$ in the lexicographical order. Then the entries Φ_{ij} are such that for $(i,j) \in v(G)$,*

$$\Phi_{ij} = \frac{K_{ij} - \sum_{k=1}^{i-1} \Phi_{ki}\Phi_{kj}}{\Phi_{ii}}. \tag{2.3}$$

For $K_{1k} = 0,\ k = 2, \ldots, p$,

$$\Phi_{1k} = 0.$$

For $K_{ij} = 0, j = 1, \ldots, p,\ i \neq 1$,

$$\Phi_{ij} = -\frac{\sum_{k=1}^{i-1} \Phi_{ki}\Phi_{kj}}{\Phi_{ii}}.$$

For $K_{ij} \neq 0,\ (i,j) \in u \in \mathcal{V} \cup \mathcal{E},\ (i,j) \notin v(G)$,

$$\Phi_{ij} = \frac{\Phi_{i_u j_u}\Phi_{i_u i_u} + \sum_{k=1}^{i_u-1} \Phi_{k i_u}\Phi_{k j_u} - \sum_{k=1}^{i-1} \Phi_{ki}\Phi_{kj}}{\Phi_{ii}}. \tag{2.4}$$

For $i = 1, \ldots, p$,

$$\Phi_{ii} = |\Phi^2_{i_u i_u} + \sum_{k=1}^{i_u-1} \Phi^2_{k i_u} - \sum_{k=1}^{i-1} \Phi^2_{ki}|^{\frac{1}{2}}. \tag{2.5}$$

Proof. The first three equations can be found in Roverato [2002]. We will only prove (2.4) since (2.5) will follow immediately from it. For all $(i,j) \in u \in \mathcal{V} \cup \mathcal{E}$ and $(i,j) \neq (i_u, j_u) \in u$, by (2.3), we have that $K_{i_u j_u} = \sum_{k=1}^{i_u} \Phi_{k i_u} \Phi_{k j_u}$ and in general $K_{ij} = \sum_{k=1}^{i} \Phi_{ki} \Phi_{kj}$. Since $K_{ij} = K_{i_u j_u}$, it follows that

$$\Phi_{i_u j_u} \Phi_{i_u i_u} + \sum_{k=1}^{i_u - 1} \Phi_{k i_u} \Phi_{k j_u} = \Phi_{ii} \Phi_{ij} + \sum_{k=1}^{i-1} \Phi_{ki} \Phi_{kj}.$$

Equations (2.4) and (2.5) follow then immediately. ■

Proposition 2.1.2 *For $K = Q^t(\Psi^t \Psi)Q \in P_{\mathcal{G}}$ with Ψ and Q as defined in definitions (2.1) and (2.2), the entries Ψ_{ij} of Ψ are as follows: for $(r,s) \in v(G)$ and $r \neq s$,*

$$\Psi_{rs} = \sum_{j=r}^{s-1} -\Psi_{rj} \frac{Q_{js}}{Q_{ss}} + \frac{\Phi_{rs}}{Q_{ss}}. \tag{2.6}$$

For $(r,s) \in v(G)$ and $r = s$,

$$\Psi_{ss} = \frac{\Phi_{ss}}{Q_{ss}}.$$

For $K_{rs} = 0$ and $r \neq 1$,

$$\Psi_{rs} \quad = \sum_{j=r}^{s-1} -\Psi_{rj} \frac{Q_{js}}{Q_{ss}} - \sum_{i=1}^{r-1} \left(\frac{\Psi_{ir} + \sum_{j=i}^{r-1} \Psi_{ij} \frac{Q_{jr}}{Q_{rr}}}{\Psi_{rr}} \right) \left(\Psi_{is} + \sum_{j=i}^{s-1} \Psi_{ij} \frac{Q_{js}}{Q_{ss}} \right).$$

For $K_{1s} = 0$,

$$\Psi_{1s} = \sum_{j=1}^{s-1} \left(-\Psi_{1j} \frac{Q_{js}}{Q_{ss}} \right).$$

For $K_{rs} \neq 0, (r,s) \in u \in \mathcal{V} \cup \mathcal{E}, (r,s) \notin v(G)$,

$$\Psi_{rs} = \frac{\Phi_{i_u j_u}\Phi_{i_u i_u} + \sum_{k=1}^{i_u-1} \Phi_{k i_u}\Phi_{k j_u} - \sum_{k=1}^{r-1} \Phi_{kr}\Phi_{ks}}{\Phi_{rr}Q_{ss}} - \sum_{j=r}^{s-1} \Psi_{rj}\frac{Q_{js}}{Q_{ss}}. \tag{2.7}$$

For $s = 1, \ldots, p$,

$$\Psi_{ss} = \frac{|\Phi^2_{i_u i_u} + \sum_{k=1}^{i_u-1} \Phi^2_{k i_u} - \sum_{k=1}^{r-1} \Phi^2_{ks}|^{\frac{1}{2}}}{Q_{ss}}. \tag{2.8}$$

Proof. We will prove (2.7) and therefore (2.8). Since $\Phi = \Psi Q$, for $r \neq s$, we have

$$\Phi_{rs} = \Psi_{rs}Q_{ss} + \sum_{j=r}^{s-1} \Psi_{rj}Q_{js}.$$

On the other hand, by (2.6), we have

$$\Phi_{rs} = \frac{\Phi_{i_u j_u}\Phi_{i_u i_u} + \sum_{k=1}^{i_u-1} \Phi_{k i_u}\Phi_{k j_u} - \sum_{k=1}^{r-1} \Phi_{kr}\Phi_{ks}}{\Phi_{rr}}.$$

It then follows that

$$\Psi_{rs}Q_{ss} + \sum_{j=r}^{s-1} \Psi_{rj}Q_{js} = \frac{\Phi_{i_u j_u}\Phi_{i_u i_u} + \sum_{k=1}^{i_u-1} \Phi_{k i_u}\Phi_{k j_u} - \sum_{k=1}^{r-1} \Phi_{kr}\Phi_{ks}}{\Phi_{rr}}$$

which implies (2.7) and (2.8). ■

In order to induce the density of the colored G-Wishart distribution in terms of Ψ, we need to know the Jacobian of the change of variables from $K^{v(G)}$ to $\Psi^{v(G)}$, which can be achieved through two steps as follows.

Lemma 2.1.1 *Let K be in $P_{\mathcal{G}}$ and v_i^G be the number $j \in \{i, \ldots, p\}$ such that $(i,j) \notin v(G)$. Then the Jacobian of the change of variables $K^{v(G)} \to \Phi^{v(G)}$ as defined in* (2.1) *is*

$$|J(K^{v(G)} \to \Phi^{v(G)})| = 2^{|\mathcal{V}|} \prod_{i=1}^{p} \Phi_{ii}^{p-i+1-v_i^G}$$

where $|\mathcal{V}|$ is the number of vertex color class of G.

Proof. We order the entries of both matrices K and Φ according to the lexicographic order. For $(i,j) \in v(G)$, differentiating (2.3) yields

$$\frac{\partial K_{ii}}{\partial \Phi_{ii}} = 2\Phi_{ii} \ , \qquad \frac{\partial K_{ii}}{\partial \Phi_{ks}} = 0 \text{ for } (k,s) > (i,i) \ ,$$
$$\frac{\partial K_{ij}}{\partial \Phi_{ij}} = \Phi_{ii} \ , \qquad \frac{\partial K_{ij}}{\partial \Phi_{ks}} = 0 \text{ for } (k,s) > (i,j), \ \ i \neq j.$$

Therefore, the Jacobian is an upper-triangular matrix and its determinant is the product of the diagonal elements. The lemma then follows immediately from the fact that for a given $i \in \{1, \ldots, p\}$, the cardinality of the set $\{j | (i,j) \in v(G), \ (i,j) \geq (i,i)\}$ is $p - i + 1 - v_i^G$. ■

Lemma 2.1.2 *Let K be in $P_{\mathcal{G}}$ and $d_i^G = \mid \{j | j \leq i, \ (j,i) \notin v(G)\} \mid$. The Jacobian of the change of variables $\Phi^{v(G)} \to \Psi^{v(G)}$ where Φ and Ψ are defined in* (2.1) *and* (2.2) *is*

$$|J(\Phi^{v(G)} \to \Psi^{v(G)})| = \prod_{i=1}^{p} Q_{ii}^{i-d_i^G}.$$

Proof. We order the entries of both matrices Φ and Ψ according to the lexicographic order. For $(r,s) \in v(G)$, differentiating (2.6), we obtain

$$\frac{\partial \Phi_{rs}}{\partial \Psi_{ss}} = Q_{ss} \ , \qquad\qquad \frac{\partial \Phi_{rs}}{\partial \Psi_{ij}} = 0 \text{ for } (i,j) > (r,s).$$

The Jacobian is thus an upper-triangular matrix and its determinant is the product of the diagonal elements. The lemma follows from the definition of d_i^G. ■

Theorem 2.1.3 *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an arbitrary p-dimensional colored graph. Then the density of the colored G-Wishart distribution expressed in terms of $\Psi^{v(G)}$ is*

$$p(\Psi^{v(G)}|\delta, D) = \frac{2^{|\mathcal{V}|}}{I_G(\delta, D)} \prod_{i=1}^{p} Q_{ii}^{p-v_i^G-d_i^G+\delta-1} \prod_{i=1}^{p} \Psi_{ii}^{p-i-v_i^G+\delta-1} e^{-\frac{1}{2}\sum_{i=1}^{p}\sum_{j=i}^{p}\Psi_{ij}^2}. \qquad (2.9)$$

Proof. The expression of $p(\Phi^{v(G)}|\delta, D)$ above follows immediately from the fact that $|K| = \prod_{i=1}^{p} \Phi_{ii}^2$, that $tr(KD) = \sum_{i=1}^{p}\sum_{j=1}^{p} \Psi_{ij}^2$ and from the expressions of the Jacobians given in Lemmas 2.1.1 and 2.1.2. ■

2.2 Metropolis-Hastings Algorithms

In this section, the MH algorithm we use to obtain the random samples from the density (2.9) is briefly described. First, we note that if we make the further change of variables

$$(\Psi_{ii}, (i,i) \in v(G), \Psi_{ij}, (i,j) \in v(G), i \neq j)$$

$$\mapsto \quad (t_{ii} = \Psi_{ii}^2, (i,i) \in v(G), \Psi_{ij}, (i,j) \in v(G), i \neq j),$$

we get

$$\begin{aligned} & p(t_{ii},(i,i)\in v(G),\Psi_{ij},(i,j)\in v(G),i\neq j|\delta,D) \\ \propto & \prod_{i=1}^{p} t_{ii}^{\frac{p-i-v_i^G+\delta}{2}-1} e^{-\frac{1}{2}\sum_{i=1} t_{ii}} e^{-\frac{1}{2}\sum_{i=1}^{p}\sum_{j=i+1}^{p}\Psi_{ij}^2} \end{aligned}$$

and we observe that $t_{ii}^{\frac{p-i-v_i^G+\delta}{2}-1} e^{-\frac{1}{2}\sum_{i=1} t_{ii}}$ has the form of a chi-square $\chi^2_{p-i-v_i^G+\delta}$ distribution.

We let $\Psi^{[s]}$ and $\Psi^{[s+1]}$ be the current state of the chain and the next state of the chain, respectively. Denote Ψ' as the candidate of $\Psi^{[s+1]}$. We further define

$$\Psi_{v(G)^c} = \Big\{\Psi_{ij}|(i,j)\in v(G)^c\Big\}$$

where $v(G)^c$ is the complement of $v(G)$ in $V\times V$. For $(i,j)\in v(G)$, an element $\Psi_{ij}^{[s]}$ is updated by sampling a candidate value Ψ'_{ij} from a standard normal distribution. For $(i,i)\in v(G)$, a element $\Psi_{ii}^{[s]}$ is updated by sampling a candidate value $(\Psi_{ii}^2)'$ from a $\chi^2_{p-i-v_i^G+\delta}$ distribution. The non-free elements of Ψ' are uniquely defined by the functions of the free elements in Proposition 2.1.1 and Proposition 2.1.2. The Markov chain moves to Ψ' with the acceptance probability

$$\min\Big\{\frac{h[(\Psi')_{v(G)^c}]}{h[(\Psi^{[s]})_{v(G)^c}]},1\Big\},$$

where

$$h(\Psi_{v(G)^c}) = \prod_{(i,i)\in v(G)^c} \Psi_{ii}^{p-i-v_i^G+\delta-1}\exp(-\frac{1}{2}\sum_{(i,j)\in v(G)^c}\Psi_{ij}^2). \tag{2.10}$$

Finally, we can obtain $K^{[s]}$ by $K^{[s]} = Q^t(\Psi^{[s]})^t\Psi^{[s]}Q$.

In our MH algorithm, the candidates are generated independently from the current samples through the proposal density. Therefore, the algorithm gives us an independent MH chain. Since the proposal density is a product of normals and chi-squares, then any state j can be reached from any state i in a finite number of steps for all i and j, and the probability of going from state j in the step t to state j in the next step $t+1$ is positive for any $t \geq 1$. Hence, the MH algorithm constructs an irreducible and aperiodic Markov chain for which the stationary distribution equals to the colored G-Wishart distribution of $\Psi^{v(G)}$.

In this chapter, we adapt the MH algorithm proposed by Mitsakakis et al. [2011] for the G-Wishart distribution to our colored G-Wishart distribution. We may wonder whether there is other method we can adapt from the G-Wishart distribution. For the G-Wishart distribution, Dobra et al. [2011] derived another MH algorithm to generate the samples. In their MH algorithm, they sample the candidate for the free diagonal element ψ_{ii} from a normal distribution $N(\psi_{ii}^{[t]}, \sigma^2)$ truncated below at zero, and sample the candidate for the free off diagonal element ψ_{ij} from a normal distribution $N(\psi_{ij}^{[t]}, \sigma^2)$. This algorithm can be adapted to the colored G-Wishart distribution. However, the chain from their algorithm has very high autocorrelations. Wang and Li [2012] and Lenkoski [2013] also derived two different sampling methods

for the G-Wishart distribution. Both of the methods are based on the block Gibbs sampler. Let $C = \{C_1, C_2, \ldots, C_k\}$ be a set of cliques of the graph. We know that $(\Sigma_{C_i})^{-1} \sim W(\delta, D_{C_i,C_i})$ and $(\Sigma_{C_i})^{-1} = K_{C_i,C_i} - K_{C_i,V\setminus C_i} K^{-1}_{V\setminus C_i,V\setminus C_i} K_{V\setminus C_i,C_i}$ Thus, in each clique C_i, we can first generate a Wishart random matrix A from $W(\delta, D_{C_i,C_i})$ and then set $K_{C_i,C_i} = A + K_{C_i,V\setminus C_i} K^{-1}_{V\setminus C_i,V\setminus C_i} K_{V\setminus C_i,C_i}$. The elements of K are updated according to all cliques until convergence. However, in the colored G-Wishart distribution, there still exists the colored constraints in each clique. We even do not know what is the conditional distribution on the cliques. Therefore, we can't adapt their methods to the colored G-Wishart distribution.

We now have a sampling method for the colored G-Wishart distribution, whether it is a prior or a posterior distribution and thus obtain an estimate of the posterior mean of K. Our simulation results in Chapter 4 will show that the chain has a good mixing, low autocorrelations and the high proximity to the colored G-Wishart distribution. In order to assess the accuracy of our MH algorithm, we would like to have the exact value of the expected value of K under the colored G-Wishart distribution. This is done in the next chapter for some special colored graphs.

3 Expected Values in Some Special Cases

In this chapter, we evaluate the normalizing constants of the colored G-Wishart distribution for some special colored graphs. We consider the colored trees, two colored styles of star graphs, a colored complete graph on three vertices and a simple decomposable colored graph on four vertices. In order to calculate the analytic expression of the normalizing constant $I_G(\delta, D)$, we need to know two special functions, the Bessel function of the third kind $K_\lambda(z)$ and the hypergeometric function ${}_pF_q$. The *Bessel* function of the third kind is defined as

$$K_\lambda(z) = \int_0^\infty u^{2\lambda-1} e^{-\frac{z}{2}(\frac{1}{u^2}+u^2)} du.$$

For some special values of λ, the Bessel function can be given explicitly, for example

$$K_{1/2}(z) = \sqrt{\frac{\pi}{2}} z^{-1/2} e^{-z}, \quad K_{3/2}(z) = \sqrt{\frac{\pi}{2}} (z^{-1/2} + z^{-3/2}) e^{-z},$$
$$K_{5/2}(z) = \sqrt{\frac{\pi}{2}} (z^{-1/2} + 3z^{-3/2} + 3z^{-5/2}) e^{-z}.$$

We will also use the classical formula

$$(\frac{p}{q})^{\frac{\lambda}{2}} K_\lambda(\sqrt{pq}) = \int_0^\infty u^{2\lambda-1} e^{-\frac{1}{2}(\frac{p}{u^2}+qu^2)} du.$$

The *hypergeometric* function ${}_pF_q$ is defined by the power series:

$${}_pF_q(a_1,\ldots,a_p;b_1,\ldots,b_q;z) = \sum_{k=0}^{\infty} \frac{(a_1)_k \cdots (a_p)_k}{(b_1)_k \cdots (b_p)_k} \frac{z^k}{k!}$$

where

$$(a)_k = \begin{cases} 1 & \text{if } k = 0 \\ a(a+1)\cdots(a+k-1) & \text{if } k > 0\ . \end{cases}$$

The derivative of the hypergeometric function ${}_pF_q(a_1,\ldots,a_p;b_1,\ldots,b_q;z)$ is given by

$$\frac{d}{dz}[{}_pF_q(a_1,\ldots,a_p;b_1,\ldots,b_q;z)] = \frac{a_1\cdots a_p}{b_1\cdots b_q}({}_pF_q(a_1{+}1,\ldots,a_p{+}1;b_1{+}1,\ldots,b_q{+}1;z))\ . \tag{3.1}$$

Since the colored G-Wishart distribution can be expressed in an exponential family form, then we can use the property of the cumulant generating function to obtain the mean of K in the colored G-Wishart distribution. For a given colored graph $\mathcal{G}$, the colored G-Wishart distribution as defined in (1.4) and (1.5) can be written as an exponential family of the type

$$f(K;\theta)dK = \exp\{tr(K\theta) - k(\theta)\}\mu(dK)$$

with the generating measure $\mu(dK) = |K|^{(\delta-2)/2}\mathbf{1}_{K\in P_\mathcal{G}}$, $\theta = -\frac{1}{2}D$ and the cumulant generating function $k(\delta, D) = \log I_\mathcal{G}(\delta, D)$. From the classical theory of the exponen-

tial family of distribution, we know that the mean of K in the colored G-Wishart can be obtained by

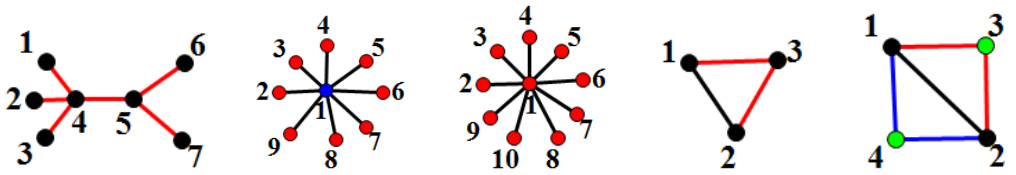
$$E(K) = \frac{\partial\, k(\delta, D)}{\partial\, (-\frac{1}{2}D)} = -2\frac{\partial\, k(\delta, D)}{\partial D}.$$

Therefore, we first need to determine for which values of δ and D the quantity $I_{\mathcal{G}}(\delta, D)$ is finite and calculate the analytic expression of $I_{\mathcal{G}}(\delta, D)$. Then we differentiate this expression of $I_{\mathcal{G}}(\delta, D)$ to get the mean of K in the colored G-Wishart distribution.

We can not do this in general but we will consider several special colored graphs for which we can calculate $I_{\mathcal{G}}(\delta, D)$. For the corresponding RCON models, we will see that when $\delta > 0$ (except in the case of the star graph with all leaves in the same color class where we must have $\delta \geq 1$), the normalizing constant $I_{\mathcal{G}}(\delta, D)$ is finite when the hyperparameter D belongs to the dual $P_{\mathcal{G}}^*$ of $P_{\mathcal{G}}$. For any open convex cone C in R^n, the dual of C is defined as

$$C^* = \{y \in R^n \mid \langle x, y\rangle > 0,\ \forall x \in \bar{C} \setminus \{0\}\}$$

where $\bar{C}$ denotes the closure of C and $\langle x, y\rangle$ denotes the inner product of x and y. In the remainder of this chapter, for each RCON model, we determine the dual $P_{\mathcal{G}}^*$ and compute the value of the normalizing constant $I_{\mathcal{G}}(\delta, D)$. This will allow us, in Chapter 4, to verify the accuracy of our sampling method.



(a) (b) (c) (d) (e)

Figure 3.1: (a) The colored tree. (b) The colored star with the centre vertex of a different color. (c) The colored star with all vertices of the same color. (d) The triangle with two edges of the same color. (e) The decomposable graph with three different colors for the edges.

3.1 Trees with Vertices of Different Colors and Edges of the Same Color

Let $T = (V, E)$ be a tree with vertices of different colors and edges of the same color. An example of such $\mathcal{G}$ is given in Figure 3.1(a). Let $a = (a_i, i = 1, \ldots, p)^t$ with $a_i \geq 0$ and $b \in R$. Let S be the space of symmetric $p \times p$ matrices. We define the mapping

$$m : \ (a, b) \in R^{p+1} \mapsto m(a, b) \in S \tag{3.2}$$

with $m(a, b)$ satisfying the conditions

$$[m(a,b)]_{ii} = a_i, \ [m(a,b)]_{ij} = b = [m(a,b)]_{ji} \text{ for } (i,j) \in E, \ [m(a,b)]_{ij} = 0 \text{ for } (i,j) \notin E.$$

Let $M(\mathcal{G})$ be the linear space of matrices $m(a,b)$ for $(a,b) \in R^{p+1}$ and P be the cone of $p \times p$ symmetric positive definite matrices. Then

$$P_{\mathcal{G}} = M(\mathcal{G}) \cap \mathcal{P} \; . \tag{3.3}$$

Proposition 3.1.1 *Let T be a tree as described above. The dual cone $P_{\mathcal{G}}^*$ is*

$$P_{\mathcal{G}}^* = \{m(a',b') \in M(\mathcal{G}) \mid a' = (a'_i, i = 1, \ldots, p),\ b' \in R,\ |b'| < \frac{1}{p-1} \sum_{(i,j)\in E} \sqrt{a'_i a'_j}\} \; . \tag{3.4}$$

Proof. Let M be the set of $p \times p$ matrices and

$$\begin{aligned} \mathcal{T}_G \;\; = \;\; & \{X \in M \mid X_{ij} = 0, \text{ for } i < j,\ X_{ij} = s_{ij} \neq 0, \text{ for } i > j, (i,j) \in E, \\ & X_{ii} = t_i > 0, i = 1, \ldots, p\} \end{aligned}$$

be the set of upper triangular matrices with positive diagonal elements and nonzero entries $X_{ij}, i > j$ only for $(i,j) \in E$. The vector $s = (s_{ij}, (i,j) \in E)$ belongs to R^{p-1} since a tree with p vertices has $p-1$ edges and $t = (t_i, i = 1, \ldots, p)$ belongs to R^p. It is well-known [see Paulsen et al., 1989, Roverato, 2002] that we can find a perfect elimination scheme enumeration of the vertices of T such that, with this enumeration, $K \in P_{\mathcal{G}}$ can be written as $K = X(t,s)^t X(t,s)$ with $X(t,s) \in \mathcal{T}_G$. Then for $K = K(a,b)$ as in (3.3) we have

$$a_j = t_j^2 + \sum_{i \in E_j} s_{ij}^2,\ b = t_i s_{ij},$$

where (t, s) is the Cholesky parametrization of $K \in P_{\mathcal{G}}$. We can also parameterize $K \in P_{\mathcal{G}}$ with $(t, b) \in (0, +\infty)^p \times R$ using

$$a_j = t_j^2 + b^2 \sum_{i \in E_j} \frac{1}{t_i^2}. \tag{3.5}$$

In this proof and the following one, without loss of generality, we assume that the numbering of the vertices of T follows a perfect elimination scheme ordering. We then say that the last vertex p in that ordering is the root of the tree and we will write

$$E_j = \{i, i < j \mid (i, j) \in E\}.$$

For convenience, we denote by C the right-hand side of equation (3.4).

We show first that $P_{\mathcal{G}}^* \subset C$. Let $D = m(a', b') \in P_{\mathcal{G}}^*$. Using (3.5), we have

$$\langle K, D \rangle = a_1 a_1' + \cdots + a_p a_p' + 2(p-1)bb' = t_p^2 a_p' + Ab^2 + 2Bb + C > 0, \tag{3.6}$$

where

$$A = \sum_{j=1}^{p} \left(\sum_{i \in E_j} \frac{1}{t_i^2} \right) a_j', \quad B = (p-1)b', \quad C = \sum_{i=1}^{p-1} t_i^2 a_i'. \tag{3.7}$$

Now observe that for fixed $i = 1, \ldots, p$ then either $i = p$ and the set $\{j; i \in E_j\}$ is empty since p is the root of the tree, or the set $\{j; i \in E_j\}$ is reduced to one point, say j_i. Therefore we have $\sum_{i \in E_j} a_j' = a_{j_i}$ for $i < p$ and zero for $i = p$. (For the graph in Figure 3.1 (a), we have $j_1 = j_2 = j_5 = j_6 = 7$ and $j_3 = j_4 = 6$) and it follows that

$$A = \sum_{i=1}^{p-1} \frac{1}{t_i^2} a_{j_i}'. \tag{3.8}$$

Let us prove that $a'_j > 0$ for all $j = 1, \ldots, p$. Take $(a_1, \ldots, a_p) \in [0, \infty)^p \setminus \{0, \ldots, 0\}$. Then $K(a, 0) \in \overline{P_{\mathcal{G}}} \setminus \{0\}$ and $\langle K, D \rangle = a_1 a'_1 + \cdots + a_p a'_p > 0$ implies that $a'_j > 0$ for all j. Let us now prove that $Ab^2 + 2Bb + C \geq 0$ for all b. If not, there exists b_0 such that $Ab_0^2 + 2Bb_0 + C < 0$. Since $a'_p > 0$ when t_p is very small and $b = b_0$ in (3.6), we get a contradiction.

Let us prove that

$$|b'| \leq \frac{1}{p-1} \sum_{(i,j) \in E} \sqrt{a'_i a'_j}. \tag{3.9}$$

Since $\forall b,\ Ab^2 + 2Bb + C \geq 0$, we have $B^2 \leq AC$. Now consider the function

$$(t_1, \ldots, t_{p-1}) \mapsto AC$$

and let us compute its minimum A^*C^* on $(0, \infty)^p$. This function AC is homogeneous of degree 0 and therefore if its minimum is reached at $t^* = (t_1^*, \ldots, t_{p-1}^*)$, it will also be reached on κt^* for any $\kappa > 0$. We have for $i = 1, \ldots, p-1$,

$$\frac{\partial}{\partial t_i} AC = 2t_i a'_i A - \frac{2}{t_i^3} a'_{j_i} C = 0$$

and we therefore have

$$t_i^* = \kappa \left(\frac{a'_{j_i}}{a'_i} \right)^{1/4},\ A^* = C^* = \sum_{i=1}^{p-1} \sqrt{a'_i a'_{j_i}} = \sum_{(i,j) \in E} \sqrt{a'_i a'_j}.$$

Since $B^2 \leq AC$ for all $(t_1, \ldots, t_{p-1}) \in (0, \infty)^{p-1}$, we can claim that $B^2 \leq A^*C^*$ or equivalently (3.9).

Let us prove that inequality (3.9) is strict, that is $B^2 = A^*C^*$ is impossible. Suppose that $B^2 = A^*C^*$, i.e. $|b'| = A^*/(p-1) > 0$. Then with $t_i = t_i^*$ we get $Ab^2 + 2Bb + C = A^*(b + \operatorname{sign} b')^2$. Taking $b = -\operatorname{sign} b'$ and $t_i = t_i^*, i = 1, \ldots, p-1$ yields $Ab^2 + 2Bb + C = 0$. Now, letting also $t_p = 0$ in (3.6), we see that the left hand side of (3.6) is zero for an $(a, b) \in \overline{S} \setminus \{0\}$ which is not zero, since $b = \pm 1$. But this can not happen for $D(a', b') \in P_{\mathcal{G}}^*$. Therefore (3.9) is strict and the proof of $P_{\mathcal{G}}^* \subset C$ is completed.

Let us now show that $C \subset P_{\mathcal{G}}^*$. For $D(a', b') \in C$ given, we want to show that $\langle K, D \rangle$ is positive for all $K(a, b) \in \overline{P_{\mathcal{G}}} \setminus \{0\}$. We will do so first for $K(a, b) \in P_{\mathcal{G}}$ and then for $K(a, b) \in \overline{P_{\mathcal{G}}} \setminus (P_{\mathcal{G}} \cup \{0\})$. For $K(a, b) \in P_{\mathcal{G}}$, $t_k > 0$ and $b \in R$. From (3.6), we have

$$\langle K, D \rangle = t_p^2 a_p' + Ab^2 + 2Bb + C = t_p^2 a_p' + A \left[\left(b + \frac{B}{A} \right)^2 + \frac{1}{A^2}(AC - B^2) \right] .$$

We have checked above that $AC - B^2 \geq 0$. Moreover $a_p' > 0$ since $D(a', b') \in C$. It follows immediately that $\langle K, D \rangle > 0$.

Let us now show $\langle K, D \rangle > 0$ for $K(a, b) \in \overline{P_{\mathcal{G}}} \setminus (P_{\mathcal{G}} \cup \{0\})$ that is for $t_1 \ldots t_p = 0$ and $(t_1, \ldots, t_p, b) \neq 0$. We need to show that $\langle K, D \rangle \neq 0$. In fact $0 = a_1 a_1' + \cdots + a_p a_p' + 2(p-1)bb' = \sum_{i=1}^p t_i^2 a_i'$ implies that $t_i = 0$ for all $i = 1, \ldots, p$ since $(a', b') \in C$ implies $a_i' > 0$. On the other hand, since $b = t_i s_{ij}$ implies $b = 0$, this is impossible since we exclude the zero matrix for K. ■

We are now in a position to give the analytic expression of $I_{\mathcal{G}}(\delta, D)$ for trees with vertices of different colors and edges of the same color.

Theorem 3.1.2 *For $\mathcal{G} = \mathcal{T}$ as described above, $\delta > 0$ and $D = m(a', b') \in P_{\mathcal{G}}^*$, the normalizing constant $I_{\mathcal{G}}(\delta, D)$ is finite and equal to*

$$I_{\mathcal{G}}(\delta, D) = 2^{\frac{\delta}{2}+p-1}\Gamma(\frac{\delta}{2})\left(\prod_{i=1}^{p}(a_i')^{d_i-2}\right)^{\frac{\delta}{4}}\int_{-\infty}^{\infty}\left(\prod_{(i,j)\in E}K_{\frac{\delta}{2}}(|b|\sqrt{a_i'a_j'})\right)|b|^{\frac{p\delta}{2}}e^{-(p-1)bb'}db \tag{3.10}$$

where d_i denotes the number of neighbours of the vertex i in the tree (V, E).

For $\delta = 1$, we have

$$I_{\mathcal{G}}(1, D) = (2\pi)^{\frac{p}{2}}\prod_{i=1}^{p}(a_i')^{-\frac{1}{2}}\Big([\sum_{(i,j)\in E}(a_i'a_j')^{\frac{1}{2}} - (p-1)b']^{-1} + [\sum_{(i,j)\in E}(a_i'a_j')^{\frac{1}{2}} + (p-1)b']^{-1}\Big).$$

For $\delta = 3$, let σ_k be the k-th elementary function of the variables $\sqrt{a_i'a_j'}$, $(i,j) \in E$. We have

$$\begin{aligned} I_{\mathcal{G}}(3, D) \;=\; & 2^{\frac{p}{2}+1}\pi^{\frac{p}{2}}\prod_{i=1}^{p}(a_i')^{-\frac{3}{2}}\sum_{k=0}^{p-1}\sigma_k\Gamma(k+1)\big([\sum_{(i,j)\in E}(a_i'a_j')^{\frac{1}{2}} - (p-1)b']^{-(k+1)} \\ & + [\sum_{(i,j)\in E}(a_i'a_j')^{\frac{1}{2}} + (p-1)b']^{-(k+1)}\big). \end{aligned}$$

Proof. In $I_{\mathcal{G}}(\delta, D)$ we make the change of variables (3.5). Switching to these Cholesky coordinates leads to the Jacobian $dadb = 2^p t_1 \dots t_p \, dbdt$. As seen before the new domain of the integration is the product

$$\{(b, t); t_k > 0, b \in R\} = (0, \infty)^p \times R.$$

With the notation A, B, C of (3.7), we have

$$\langle K(a,b), D(a',b')\rangle = 2(p-1)bb' + a_1a'_1 + \cdots + a_pa'_p = t_p^2 a'_p + Ab^2 + 2Bb + C.$$

Using (3.8) for the expression of A, we obtain

$$\begin{aligned}
& I_G(\delta, D) \\
= \;& 2^p \int_{(0,\infty)^p \times R} (t_1 \dots t_p)^{\delta-1} e^{-(p-1)bb'} e^{-\frac{t_p^2 a'_p}{2}} \prod_{i=1}^{p-1} e^{-\frac{t_i^2 a'_i}{2} - \frac{b^2 a'_{j_i}}{2t_i^2}} \, dt_1 \dots dt_p \, db \qquad (3.11) \\
= \;& 2^p \int_0^\infty e^{-\frac{t_p^2 a'_p}{2}} t_p^{\delta-1} dt_p \int_{-\infty}^\infty e^{-(p-1)bb'} \prod_{i=1}^{p-1} \left(K_{\delta/2}(|b|(a'_i a'_{j_i})^{1/2})(|b|\sqrt{a'_{j_i}/a_i}^{\delta/2} \right) db \\
= \;& 2^{p+\frac{\delta}{2}-1} \frac{\Gamma(\delta/2)}{(a'_p)^{\delta/2}} \left(\prod_{i=1}^{p-1} \frac{a'_{j_i}}{a'_i} \right)^{\delta/4} J_\delta(D) \qquad (3.12)
\end{aligned}$$

with the notation

$$J_\delta(D) = \int_{-\infty}^\infty e^{-(p-1)bb'} |b|^{(p-1)\delta/2} \prod_{i=1}^{p-1} K_{\delta/2}(|b|(a'_i a'_{j_i})^{1/2}) db. \qquad (3.13)$$

We now prove by induction that

$$\frac{1}{(a'_p)^2} \times \prod_{i=1}^{p-1} \frac{a'_{j_i}}{a_i} = \prod_{i=1}^{p} (a'_i)^{d_i - 2}. \qquad (3.14)$$

Of course (3.14) is correct for $p = 2$. Suppose that (3.14) is true for any rooted tree with size p. Consider a rooted tree T^* with vertices $\{0, 1, \dots, p\}$ and root p and numbered, as usual, such that $i \prec j$ implies $i \leq j$. Denote T the induced tree with vertices $\{1, \dots, p\}$. Finally denote $d^* = (d_0^*, \dots, d_p^*)$ and $d = (d_1, \dots, d_p)$ the number

of neighbours in T^* and T. Then $d_0^* = 1$, $d_{j_0}^* = 1 + d_{j_0}$ and $d_i^* = d_i$ if $i \neq 0$ and $i \neq j_0$. This implies that

$$\frac{1}{(a'_p)^2} \times \prod_{i=0}^{p-1} \frac{a'_{j_i}}{a_i} = \frac{a'_{j_0}}{a_0} \frac{1}{(a'_p)^2} \times \prod_{i=1}^{p-1} \frac{a'_{j_i}}{a_i} \overset{(1)}{=} \frac{a'_{j_0}}{a_0} \prod_{i=1}^{p} (a'_i)^{d_i - 2} \overset{(2)}{=} \prod_{i=0}^{p} (a'_i)^{d_i^* - 2},$$

where (1) comes from the induction hypothesis and (2) is due to the link between d and d^*. The induction hypothesis is extended to $p+1$ and (3.14) is proved.

We now prove that $J_\delta(D)$ defined as (3.13) converges if $D = m(a', b') \in P_{\mathcal{G}}^*$ where $P_{\mathcal{G}}^*$ is the convex cone defined in Proposition 3.1.1. We write $J_\delta(D)$ as the sum

$$J_\delta(D) = \int_{-\infty}^{0} \dots db + \int_{0}^{+\infty} \dots db. \tag{3.15}$$

When $b \to \pm\infty$, $|b| \to +\infty$. From the formula (1) in [Watson, 1995, Chapter 7.23, page 202], we have

$$K_\lambda(s) \sim_{s \to \infty} \sqrt{\frac{\pi}{2}} \frac{e^{-s}}{s^{1/2}}.$$

We use this fact to analyse the convergence of $J_\delta(D)$. If $D = m(a', b') \in P_{\mathcal{G}}^*$, from the asymptotic formula above, we see that the integrands in both integrals on the right hand side of (3.15), when $|b|$ goes to infinity, behave like $|b|^c e^{-|b|H}$ where, since $m(a', b') \in P_{\mathcal{G}}^*$,

$$H = \sum_{(i,j) \in E}^{p} \sqrt{a'_i a'_j} - (p-1)|b'| \text{sign}(bb') > 0$$

and $c = (p-1)\frac{\delta - 1}{2}$. Since the argument of (3.13) is continuous, both integrals converge at infinity.

To study the convergence of these integrals when $b \to 0$, we recall that

$$2K_\lambda(s) = \int_0^{+\infty} x^{\lambda-1} e^{-\frac{s}{2}(x+\frac{1}{x})} dx.$$

Making the change of variable $u = sx$ in the expression of $2K_\lambda(s)$ we see that

$$K_\lambda(s) \sim_{s\to 0} s^{-\lambda} 2^{\lambda-1}\Gamma(\lambda).$$

Therefore, for both integrals in the RHS of (3.15), the integrand is equivalent to

$$\Big(|b|^{-\frac{\delta}{2}}\Gamma(\frac{\delta}{2})\Big)^{p-1} |b|^{\frac{p\delta}{2}} e^{-(p-1)bb'} = |b|^{\frac{\delta}{2}} e^{-(p-1)bb'}$$

and therefore both integrals converge at 0. The expression (3.10) of the normalizing constant is now proved.

Particularly, we will present the expression of the normalizing constants for $\delta = 1$ and $\delta = 3$. By (3.10), $I_G(1, D) = 2^{p-\frac{1}{2}}\Gamma(\frac{1}{2})(a'_p)^{-\frac{1}{2}}(\prod_{i=1}^{p-1} \frac{a'_{j_i}}{a'_i})^{\frac{1}{4}} J_1(D)$, where

$$\begin{aligned}
J_1(D) &= \int_{-\infty}^{\infty} e^{-(p-1)bb'} |b|^{\frac{p-1}{2}} \prod_{i=1}^{p-1} K_{\frac{1}{2}}(|b|(a'_i a'_{j_i})^{\frac{1}{2}}) db \\
&= \int_{-\infty}^{\infty} e^{-(p-1)bb'} |b|^{\frac{p-1}{2}} (\prod_{i=1}^{p-1} \sqrt{\frac{\pi}{2}} |b|^{-\frac{1}{2}} (a'_i a'_{j_i})^{-\frac{1}{4}} e^{|b|(a'_i a'_{j_i})^{\frac{1}{2}}}) db \\
&= (\frac{\pi}{2})^{\frac{p-1}{2}} \prod_{i=1}^{p-1} (a'_i a'_{j_i})^{-\frac{1}{4}} \int_{-\infty}^{\infty} e^{-(p-1)bb' - |b| \sum_{i=1}^{p-1} (a'_i a'_{j_i})^{\frac{1}{2}}} db.
\end{aligned}$$

We compute the integral

$$\begin{aligned}\int_{-\infty}^{\infty} e^{-(p-1)bb'-|b|\sum_{i=1}^{p-1}(a'_i a'_{j_i})^{\frac{1}{2}}}db &= \int_{-\infty}^{0} e^{-(p-1)bb'+b\sum_{i=1}^{p-1}(a'_i a'_{j_i})^{\frac{1}{2}}}db \\ &\quad + \int_{0}^{\infty} e^{-(p-1)bb'-b\sum_{i=1}^{p-1}(a'_i a'_{j_i})^{\frac{1}{2}}}db \\ &= \frac{1}{\sum_{i=1}^{p-1}(a'_i a'_{j_i}) - (p-1)b'} + \frac{1}{\sum_{i=1}^{p-1}(a'_i a'_{j_i}) + (p-1)b'}\,.\end{aligned}$$

Therefore

$$\begin{aligned}&J_1(D) \\ =\ &(\frac{\pi}{2})^{\frac{p-1}{2}}\prod_{i=1}^{p-1}(a'_i a'_{j_i})^{-\frac{1}{4}}\Big[\Big(\sum_{i=1}^{p-1}(a'_i a'_{j_i})^{\frac{1}{2}} - (p-1)b'\Big)^{-1} + \Big(\sum_{i=1}^{p-1}(a'_i a'_{j_i})^{\frac{1}{2}} + (p-1)b'\Big)^{-1}\Big].\end{aligned}$$

Since $\sum_{i=1}^{p-1}(a'_i a'_{j_i}) = \sum_{(i,j)\in E}(a'_i a'_j)$, this yields the expression of $I_{\mathcal{G}}(1, D)$.

Similarly, from (3.10), $I_G(3,D) = 2^{p+\frac{1}{2}}\Gamma(\frac{3}{2})(a'_p)^{-\frac{3}{2}}\prod_{i=1}^{p-1}(\frac{a'_{j_i}}{a'_i})^{-\frac{3}{4}}J_3(D)$ with

$$\begin{aligned}
J_3(D) &= \int_{-\infty}^{\infty} e^{-(p-1)bb'}|b|^{\frac{3}{2}(p-1)}\prod_{i=1}^{p-1}K_{\frac{3}{2}}(|b|(a'_ia'_{j_i})^{\frac{1}{2}})db \\
&= \int_{-\infty}^{\infty} e^{-(p-1)bb'}|b|^{\frac{3}{2}(p-1)} \\
&\quad \times\prod_{i=1}^{p-1}[\sqrt{\frac{\pi}{2}}(|b|^{-\frac{1}{2}}(a'_ia'_{j_i})^{-\frac{1}{4}}+|b|^{-\frac{3}{2}}(a'_ia'_{j_i})^{-\frac{3}{4}})e^{-|b|(a'_ia'_{j_i})^{\frac{1}{2}}}]db \\
&= \int_{-\infty}^{\infty} e^{-(p-1)bb'}|b|^{\frac{3}{2}(p-1)}(\frac{\pi}{2})^{\frac{p-1}{2}}|b|^{-\frac{3}{2}(p-1)} \\
&\quad \times\prod_{i=1}^{p-1}(a'_ia'_{j_i})^{-\frac{3}{4}}\prod_{i=1}^{p-1}[(|b|(a'_ia'_{j_i})^{\frac{1}{2}}+1)e^{-|b|(a'_ia'_{j_i})^{\frac{1}{2}}}]db \\
&= (\frac{\pi}{2})^{\frac{p-1}{2}}\prod_{i=1}^{p-1}(a'_ia'_{j_i})^{-\frac{3}{4}}\int_{-\infty}^{\infty} e^{-(p-1)bb'-|b|\sum_{i=1}^{p-1}(a'_ia'_{j_i})^{\frac{1}{2}}}\prod_{i=1}^{p-1}(1+|b|(a'_ia'_{j_i})^{\frac{1}{2}})db \\
&= (\frac{\pi}{2})^{\frac{p-1}{2}}\prod_{i=1}^{p-1}(a'_ia'_{j_i})^{-\frac{3}{4}} \\
&\quad \times\int_{-\infty}^{\infty} e^{-(p-1)bb'-|b|\sum_{i=1}^{p-1}(a'_ia'_{j_i})^{\frac{1}{2}}}(1+|b|\sigma_1+|b|^2\sigma_2+\ldots+|b|^{p-1}\sigma_{p-1})db \\
&= (\frac{\pi}{2})^{\frac{p-1}{2}}\prod_{i=1}^{p-1}(a'_ia'_{j_i})^{-\frac{3}{4}}\sum_{k=0}^{p-1}\sigma_k\int_{-\infty}^{\infty} e^{-(p-1)bb'-|b|\sum_{i=1}^{p-1}(a'_ia'_{j_i})^{\frac{1}{2}}}|b|^k db\,,
\end{aligned}$$

where the $\sigma_i = \sigma_i(\sqrt{a'_ia'_{j_i}}, i = 1,\ldots,p-1)$ are the symmetric functions of $\sqrt{a'_ia'_{j_i}}, i = 1,\ldots,p-1$. Since

$$\begin{aligned}
&\int_{-\infty}^{\infty} e^{-(p-1)bb'-|b|\sum_{i=1}^{p-1}(a'_ia'_{j_i})^{\frac{1}{2}}}|b|^{m-1}db \\
&= \Gamma(m)\left[\left(\sum_{i=1}^{p-1}(a'_ia'_{j_i})^{\frac{1}{2}}-(p-1)b'\right)^{-m}+\left(\sum_{i=1}^{p-1}(a'_ia'_{j_i})^{\frac{1}{2}}+(p-1)b'\right)^{-m}\right],
\end{aligned}$$

then

$$I_{\mathcal{G}}(3,D) = 2^{\frac{p}{2}+1}\pi^{\frac{p}{2}}\prod_{i=1}^{p}(a_i')^{-\frac{3}{2}}\sum_{k=0}^{p-1}\sigma_k\Gamma(k+1)$$
$$\times\left[\left(\sum_{i=1}^{p-1}(a_i'a_{j_i}')^{\frac{1}{2}}-(p-1)b'\right)^{-k-1}+\left(\sum_{i=1}^{p-1}(a_i'a_{j_i}')^{\frac{1}{2}}+(p-1)b'\right)^{-k-1}\right],$$

which yields the expression of $I_{\mathcal{G}}(3,D)$. ■

3.2 The Star Graph with Its n Leaves in One Color Class

An example of the star graph with its n leaves in one color class and different colors for the edges and the central node is given in Figure 3.1(b). For $a \in R$, $c \in R$, $b = (b_1, \ldots, b_n)^t \in R^n$, let $L(\mathcal{G})$ be the linear space of matrices of the form

$$l(a,b,c) = \begin{bmatrix} a & b_1 & b_2 & \ldots & b_n \\ b_1 & c & 0 & \ldots & 0 \\ b_2 & 0 & c & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ b_n & 0 & 0 & \ldots & c \end{bmatrix}.$$

It is easy to see that the determinant of $l(a,b,c)$ is

$$|l(a,b,c)| = c^n\left(a - \frac{\|b\|^2}{c}\right) \tag{3.16}$$

and therefore, $P_{\mathcal{G}}$ is the open cone and denoted by

$$P_{\mathcal{G}} = \{l(a,b,c) \in L(\mathcal{G}) : \ c > 0, \ a - \frac{\|b\|^2}{c} > 0\}.$$

The dual cone $P^*(\mathcal{G})$ and the normalizing constant $I_{\mathcal{G}}(\delta, D)$ are given by the following proposition and theorem.

Proposition 3.2.1 *For a star graph with all n leaves in one color class, the dual of $P_{\mathcal{G}}$ is*

$$P_{\mathcal{G}}^* = \{l(a', b', c') \in L(\mathcal{G}) \mid ||b'||^2 \leq na'c'\}. \tag{3.17}$$

Proof. By definition, for the star graph, the dual of $P_{\mathcal{G}}$ is

$$P_{\mathcal{G}}^* = \{D = l(a', b', c') \in M(\mathcal{G}) \mid \langle K, D\rangle > 0, \;\; K \in \bar{P}_{\mathcal{G}} \setminus \{0\}\}.$$

Let β denote the angle between b and b'. Then, since $\cos\beta > -1$, we have

$$\langle K, D\rangle \;\; = \;\; aa' + ncc' + 2||b||||b'||\cos\beta > aa' + ncc' - 2||b||||b'||.$$

Therefore, $2||b||||b'|| < aa' + ncc'$. Since $ac > 0$ and $\frac{4||b||^2||b'||^2}{ac} < \frac{(aa'+ncc')^2}{ac}$, by differentiation with respect to a and c, we see that $\frac{(aa'+ncc')^2}{ac} \geq 4na'c'$ and hence $\langle K, D\rangle > 0$ implies that $||b'||^2 < na'c'$. ■

Theorem 3.2.2 *For a star graph $\mathcal{G}$ with all n leaves in the same color class, $\delta \geq 1$ and $D = l(a', b', c') \in P_{\mathcal{G}}^*$, the normalizing constant of the colored G-Wishart is*

$$I_{\mathcal{G}}(\delta, D) = 2^{\frac{\delta+n\delta+2}{2}}\pi^{n/2} \times a'^{(\frac{\delta}{2}-1)(n-1)} \times \frac{1}{(na'c' - \|b'\|^2)^{(\delta-1)\frac{n}{2}+1}} \times \Gamma((\delta-1)\frac{n}{2}+1)\Gamma(\frac{\delta}{2}).$$

Proof. Let us introduce the matrix

$$A(r,s,t) = \begin{bmatrix} r & s_1 & s_2 & \dots & s_n \\ 0 & t & 0 & \dots & 0 \\ 0 & 0 & t & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & t \end{bmatrix}.$$

If $(a,b,c) \in P_{\mathcal{G}}$, the only triple (r,s,t) such that $t>0$ and $r>0$ and

$$K(a,b,c) = A(r,s,t)A^t(r,s,t) = \begin{bmatrix} r^2+\|s\|^2 & s't \\ ts & t^2 I_n \end{bmatrix}$$

satisfies $r = (a - \frac{\|b\|^2}{c})^{1/2}$, $t = \sqrt{c}$ and $s = \frac{b}{\sqrt{c}}$. A new parameterization of $P_{\mathcal{G}}$ is therefore given by the change of variables (a,b,c) into (r,s,t) with $a = r^2 + \|s\|^2$, $b = ts$ and $c = t^2$, where (r,s,t) belongs to

$$\{(r,s,t); r>0, s \in R^n, t>0\} = (0,\infty) \times R^n \times (0,\infty).$$

With this parameterization, from (3.16), we have $|K| = r^2 t^{2n}$ and $dadbdc = 4rt^{n+1}drdsdt$. Then

$$\begin{aligned} I_{\mathcal{G}}(\delta, D) &= 4\int_0^\infty \int_0^\infty \left(\int_{R^n} e^{\frac{-\|s\|^2 a' - 2t\langle s, b'\rangle}{2}} ds \right) r^{\delta-1} t^{(\delta-1)n+1} e^{\frac{-r^2 a' - nt^2 c'}{2}} dr dt \\ &= 4\left(\frac{\pi}{a'}\right)^{n/2} \int_0^\infty e^{\frac{-nt^2 c'}{2} + \frac{t^2 \|b'\|^2}{2a'}} t^{(\delta-1)n+1} dt \times \int_0^\infty r^{\delta-1} e^{\frac{-r^2 a'}{2}} dr \\ &= \left(\frac{\pi}{a'}\right)^{n/2} \int_0^\infty e^{\frac{-nvc'}{2} + \frac{v\|b'\|^2}{2a'}} v^{(\delta-1)\frac{n}{2}} dv \times \int_0^\infty v^{\frac{\delta}{2}-1} e^{\frac{-va'}{2}} dv \\ &= 2^{\frac{\delta+n\delta+2}{2}} \pi^{n/2} a'^{(\frac{\delta}{2}-1)(n-1)} \frac{1}{(na'c' - \|b'\|^2)^{(\delta-1)\frac{n}{2}+1}} \Gamma((\delta-1)\frac{n}{2}+1)\Gamma(\frac{\delta}{2}). \end{aligned}$$

The proof is completed. ■

3.3 The Star Graph with All Vertices in One Color Class

An example of the star graph with all vertices in one color class and different colors for the edges is given in Figure 3.1(c). This is a special case of the preceding one and therefore, we have immediately that

$$P_{\mathcal{G}} = \{l(a, b, a) \in L(\mathcal{G}) \mid a > 0, a^2 - ||b||^2 > 0\}.$$

Since this is a well-known cone, called the Lorentz cone, we know also that it is self dual and therefore

$$P_{\mathcal{G}}^* = \{l(a', b', a') \in L(\mathcal{G}) \mid a' > 0, (a')^2 - ||b'||^2 > 0\}.$$

It remains to compute $I_{\mathcal{G}}(\delta, D)$.

Theorem 3.3.1 *For the star graph $\mathcal{G}$ with n leaves and all vertices in the same color class, $\delta > 0$ and $D = l(a', b', a') \in P_{\mathcal{G}}^*$, the normalizing constant of the colored G-Wishart is*

$$I_{\mathcal{G}}(\delta, D) = \frac{2^{\frac{(n+1)\delta}{2}-1} C_n \Gamma((n+1)\frac{\delta}{2})}{(n+1)^{\frac{(n+1)\delta}{2}} (a')^{\frac{(n+1)\delta}{2}}} B(\frac{\delta}{2}, \frac{n}{2}) \, {}_2F_1\Big((n+1)\frac{\delta}{4}, (n+1)\frac{\delta}{4} + \frac{1}{2}, \frac{n+\delta}{2}; u\Big)$$

where $u = \left(\frac{2||b'||}{(n+1)a'}\right)^2$ and $B(\frac{\delta}{2}, \frac{n}{2})$ is the Beta function with argument $(\frac{\delta}{2}, \frac{n}{2})$.

Proof. It is easy to see that

$$\begin{aligned} I_{\mathcal{G}}(\delta, D) &= \int_{R^n} \left(\int_{||b||}^{\infty} a^{n\frac{\delta-2}{2}} (a - \frac{||b||^2}{a})^{\frac{\delta-2}{2}} \exp -\frac{1}{2}\{(n+1)aa' + 2\langle b, b'\rangle\} da \right) db \\ &= \int_{R^n} \left(\int_{||b||}^{\infty} a^{(n-1)\frac{\delta-2}{2}} (a^2 - ||b||^2)^{\frac{\delta-2}{2}} \exp -\frac{1}{2}\{(n+1)aa' + 2\langle b, b'\rangle\} da \right) db. \end{aligned}$$

Let us make the change of variables

$$(a, b) \in (||b||, +\infty) \times R^n \mapsto (u, R, \theta) \in (0, 1) \times (0, +\infty) \times S$$

where $b = R\theta$, S is the unit sphere in R^n and $a = \frac{R}{\sqrt{u}}$. We have $dadb = -\frac{1}{2u^{3/2}} RC_n R^{n-1} du dR d\theta$ where C_n is the surface area of S. Then

$$\begin{aligned} & I_{\mathcal{G}}(\delta, D) \\ = & \frac{C_n}{2} \int_S \left[\int_0^{+\infty} \left(\int_0^1 R^{(n-1)\frac{\delta-2}{2}} u^{-(n-1)\frac{\delta-2}{4}} R^{\delta-2} (\frac{1}{u} - 1)^{\frac{\delta-2}{2}} \exp -\{\frac{(n+1)Ra'}{2\sqrt{u}} + R\langle\theta, b'\rangle\} u^{-3/2} du \right) R^n dR \right] d\theta \\ = & \frac{C_n}{2} \int_S \left[\int_0^{+\infty} \left(\int_0^1 R^{(n-1)\frac{\delta-2}{2}} u^{-(n+1)\frac{\delta-2}{4}} R^{\delta-2} (1-u)^{\frac{\delta-2}{2}} \exp -\{\frac{(n+1)Ra'}{2\sqrt{u}} + R\langle\theta, b'\rangle\} u^{-3/2} du \right) R^n dR \right] d\theta \\ = & \frac{C_n}{2} \int_S \left[\int_0^{+\infty} \left(\int_0^1 R^{(n+1)\frac{\delta}{2}-1} u^{-(n+1)\frac{\delta-2}{4}-\frac{3}{2}} (1-u)^{\frac{\delta-2}{2}} \exp -R\{\frac{(n+1)a'}{2\sqrt{u}} + \langle\theta, b'\rangle\} du \right) dR \right] d\theta \\ = & \frac{C_n}{2} \int_S \left[\int_0^1 u^{-(n+1)\frac{\delta-2}{4}-\frac{3}{2}} (1-u)^{\frac{\delta-2}{2}} \left(\int_0^{+\infty} R^{(n+1)\frac{\delta}{2}-1} \exp -R\{\frac{(n+1)a'}{2\sqrt{u}} + \langle\theta, b'\rangle\} dR \right) du \right] d\theta \\ = & \frac{C_n \Gamma((n+1)\frac{\delta}{2})}{2} \int_S \left[\int_0^1 u^{-(n+1)\frac{\delta-2}{4}-\frac{3}{2}} (1-u)^{\frac{\delta-2}{2}} \left(\frac{(n+1)a'}{2\sqrt{u}} + \langle\theta, b'\rangle \right)^{-(n+1)\frac{\delta}{2}} du \right] d\theta \\ = & \frac{C_n \Gamma((n+1)\frac{\delta}{2})}{2} (\frac{(n+1)a'}{2})^{-(n+1)\frac{\delta}{2}} \times \\ & \int_S \left[\int_0^1 u^{-(n+1)\frac{\delta-2}{4}-\frac{3}{2}} (1-u)^{\frac{\delta-2}{2}} u^{(n+1)\frac{\delta}{4}} \left(1 + \frac{2}{(n+1)a'} \sqrt{u} \langle\theta, b'\rangle \right)^{-(n+1)\frac{\delta}{2}} du \right] d\theta \\ = & K_{n,\delta}(a') \int_S \left[\int_0^1 u^{\frac{n}{2}-1} (1-u)^{\frac{\delta}{2}-1} \sum_{k=0}^{\infty} (-1)^k \left(\frac{2\langle\theta, b'\rangle}{(n+1)a'} \right)^k u^{\frac{k}{2}} \frac{\left((n+1)\frac{\delta}{2}\right)_k}{k!} du \right] d\theta \end{aligned}$$

with $K_{n,\delta}(a') = \frac{2^{\frac{(n+1)\delta}{2}-1}C_n\Gamma((n+1)\frac{\delta}{2})}{(n+1)^{\frac{(n+1)\delta}{2}}(a')^{\frac{(n+1)\delta}{2}}}$. Therefore

$$\begin{aligned}
& I_{\mathcal{G}}(\delta, D) \\
= \; & K_{n,\delta}(a') \sum_{k=0}^{\infty} (-1)^k \Big(\frac{2}{(n+1)a'}\Big)^k \frac{\Big((n+1)\frac{\delta}{2}\Big)_k}{k!} \int_0^1 u^{\frac{k+n}{2}-1}(1-u)^{\frac{\delta}{2}-1} du \int_S \langle \theta, b' \rangle^k d\theta \\
= \; & K_{n,\delta}(a') \sum_{k=0}^{\infty} \Big(\frac{2}{(n+1)a'}\Big)^{2k} \frac{\Big((n+1)\frac{\delta}{2}\Big)_{2k}}{(2k)!} \int_0^1 u^{\frac{2k+n}{2}-1}(1-u)^{\frac{\delta}{2}-1} du \int_S \langle \theta, b' \rangle^{2k} d\theta \\
= \; & K_{n,\delta}(a') \sum_{k=0}^{\infty} \Big(\frac{2}{(n+1)a'}\Big)^{2k} \frac{\Big((n+1)\frac{\delta}{2}\Big)_{2k}}{(2k)!} \frac{\Gamma(k+\frac{n}{2})\Gamma(\frac{\delta}{2})}{\Gamma(k+\frac{\delta+n}{2})} \|b'\|^{2k} \frac{(1/2)_k}{(n/2)_k}.
\end{aligned}$$

We now use the fact that $(\alpha)_{2k} = 2^{2k}\Big(\frac{\alpha}{2}\Big)_k\Big(\frac{\alpha+1}{2}\Big)_k$, $\Gamma(\alpha+k) = \Gamma(\alpha)(\alpha)_k$ and

$$\begin{aligned}
(2k)! &= (135...(2k-1))(246...2k) \\
&= 2^k k! 2^k \frac{1}{2}\frac{3}{2}...\frac{2k-1}{2} \\
&= 2^{2k} k! \frac{1}{2}(\frac{1}{2}+1)(\frac{1}{2}+2)...(\frac{1}{2}+(k-1)) \\
&= 2^{2k} k! \Big(\frac{1}{2}\Big)_k.
\end{aligned}$$

Finally, since the integral is rotational symmetric, we take $b' = ||b'||e_1$ so that $\langle \theta, b' \rangle = \theta_1 ||b'||$. Recalling that $d\theta$ is the distribution of $\frac{Z}{||Z||}$ when $Z \sim N(0, I_n)$. Therefore, $\theta_1^2 \sim \text{Beta}(\frac{1}{2}, \frac{n-1}{2})$ where $\theta_1 = \frac{Z_1}{\sqrt{Z_1^2+...+Z_n^2}}$. For $v = \theta_1$, we have

$$\int_S \langle \theta, b' \rangle^{2k} d\theta = \frac{\|b'\|^{2k}}{B(\frac{1}{2}, \frac{n-1}{2})} \int_0^1 v^{k-\frac{1}{2}}(1-v)^{\frac{n-1}{2}-1} dv = \|b'\|^{2k} \frac{(1/2)_k}{(n/2)_k}.$$

Writing $B(\alpha,\beta)$ for the Beta function with argument (α,β), we obtain

$$\begin{aligned}
& I_{\mathcal{G}}(\delta, D) \\
= \;& K_{n,\delta}(a')B(\frac{\delta}{2},\frac{n}{2}) \\
& \times \sum_{k=0}^{\infty}\Big(\frac{2}{(n+1)a'}\Big)^{2k}\frac{2^{2k}}{(2k)!}\Big((n+1)\frac{\delta}{4}\Big)_k\Big((n+1)\frac{\delta}{4}+\frac{1}{2}\Big)_k\frac{\Big(\frac{n}{2}\Big)_k}{\Big(\frac{n+\delta}{2}\Big)_k}\|b'\|^{2k}\frac{(\frac{1}{2})_k}{(\frac{n}{2})_k} \\
= \;& K_{n,\delta}(a')B(\frac{\delta}{2},\frac{n}{2}) \\
& \times \sum_{k=0}^{\infty}\Big(\frac{2}{(n+1)a'}\Big)^{2k}\frac{2^{2k}}{2^{2k}k!(\frac{1}{2})_k}\Big((n+1)\frac{\delta}{4}\Big)_k\Big((n+1)\frac{\delta}{4}+\frac{1}{2}\Big)_k\frac{\Big(\frac{n}{2}\Big)_k}{\Big(\frac{n+\delta}{2}\Big)_k}\|b'\|^{2k}\frac{(\frac{1}{2})_k}{(\frac{n}{2})_k}\,.
\end{aligned}$$

Let $u = \Big(\frac{2||b'||}{(n+1)a'}\Big)^2$. We note that since $D = l(a',b',a') \in P^*_{\mathcal{G}}$, then $u \leq 1$. After obvious simplifications in the expression above, we have

$$\begin{aligned}
I_{\mathcal{G}}(\delta, D) &= K_{n,\delta}(a')B(\frac{\delta}{2},\frac{n}{2})\sum_{k=0}^{\infty}\frac{u^k}{k!}\frac{\Big((n+1)\frac{\delta}{4}\Big)_k\Big((n+1)\frac{\delta}{4}+\frac{1}{2}\Big)_k}{\Big(\frac{n+\delta}{2}\Big)_k} \\
&= K_{n,\delta}(a')B(\frac{\delta}{2},\frac{n}{2})\;{}_2F_1\Big((n+1)\frac{\delta}{4},(n+1)\frac{\delta}{4}+\frac{1}{2},\frac{n+\delta}{2};u\Big)\,.
\end{aligned}$$

The proof is completed. ■

3.4 A Complete Graph on Three Vertices

This graph is represented in Figure 3.1(d). In this case, the cone $P_{\mathcal{G}}$ is the set of positive definite matrices $K = (k_{ij})_{1\leq i,j\leq 3}$ with $k_{13} = k_{23}$. The dual cone $P^*(\mathcal{G})$ and the normalizing constant $I_{\mathcal{G}}(\delta, D)$ are given below.

Proposition 3.4.1 *For the graph in Figure 3.1(d), the dual of $P_{\mathcal{G}}$ is*

$$\begin{aligned} P_{\mathcal{G}}^* &= \{D = (d_{ij})_{1\leq i,j\leq 3} \in S \mid d_{13} = d_{23}, \\ &\qquad d_{ii} > 0, i = 1,2,3,\ d_{12}^2 < d_{11}d_{22},\ 4d_{13}^2 < (d_{11} + d_{22} + 2d_{12})d_{33}\}. \end{aligned}$$

Proof. We write the Cholesky decomposition of K under the form $K = AA^t$ with

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{pmatrix}.$$

Expressing the k_{ij} in terms of the a_{ij} and imposing $k_{13} = k_{23}$ immediately show that we must have $a_{13} = a_{23}$. Then, let $D = (d_{ij})_{1\leq i,j\leq 3}$ with $d_{13} = d_{23}$ since the dual of $P_{\mathcal{G}}$ must be in the same linear space as $P_{\mathcal{G}}$. We also have

$$\begin{aligned} \langle K, D\rangle &= (a_{11}^2 + a_{12}^2 + a_{13}^2)d_{11} + (a_{22}^2 + a_{13}^2)d_{22} + a_{33}^2 d_{33} \\ &\qquad +2(a_{22}a_{12} + a_{13}^2)d_{12} + 4a_{13}a_{33}d_{13} \\ &= a_{13}^2(d_{11} + d_{22} + 2d_{12}) + 4a_{13}a_{33}d_{13} + a_{12}^2 d_{11} + 2a_{12}a_{22}d_{12} \\ &\qquad +a_{11}^2 d_{11} + a_{22}^2 d_{22} + a_{33}^2 d_{33}, \end{aligned}$$

which we view as a quadratic form a^tMa with $a^t = (a_{13}, a_{33}, a_{12}, a_{22}, a_{33})$ and

$$M = \begin{pmatrix} d_{11}+d_{22}+2d_{12} & 2d_{13} & 0 & 0 & 0 \\ 2d_{13} & d_{33} & 0 & 0 & 0 \\ 0 & 0 & d_{11} & d_{12} & 0 \\ 0 & 0 & d_{12} & d_{22} & 0 \\ 0 & 0 & 0 & 0 & d_{11} \end{pmatrix}.$$

Since AA^t is the Cholesky parametrization of $P_{\mathcal{G}}$, clearly $K \in P_{\mathcal{G}}$ if and only if $a_{ii} > 0, i = 1, 2, 3$. If we can prove the following lemma, the condition $M > 0$ will yield the dual cone $P_{\mathcal{G}}^*$.

Lemma 3.4.1 *The $tr(K, D)$ is positive for all $K \in \bar{P}_{\mathcal{G}} \setminus \{0\}$ if and only if the matrix M of the quadratic form $\langle K, D\rangle = a^tMa$ is positive definite.*

Let us now prove the lemma. Clearly if $M > 0$, then $\langle K, D\rangle = a^tMa > 0$ for all $a \in R^5$ and in particular for all a with $a_{ii} > 0$, $i = 1, 2, 3$. Conversely, let $a \in R^5$. Then a can be written as

$$a = (\epsilon_1 a_{11}, \epsilon_2 a_{22}, \epsilon_3 a_{33}, a_{12}, a_{13})^t$$

where ϵ_i is the sign of $a_{ii}, i = 1, 2, 3$ and we have

$$a^tMa = (a_{11}^2+a_{12}^2+a_{13}^2)d_{11}+(a_{22}^2+a_{13}^2)d_{22}+a_{33}^2d_{33}+2(\epsilon_2 a_{22}a_{12}+a_{13}^2)d_{12}+4\epsilon_3 a_{13}a_{33}d_{13}.$$

But this is also equal to $\tilde{a}^t M \tilde{a}$ where

$$\tilde{a}^t = (|a_{11}|, |a_{22|}, |a_{33}|, \epsilon_2 a_{12}, \epsilon_3 a_{13})$$

which is in $P_{\mathcal{G}}$. Therefore $\langle K, D \rangle > 0$ for all $K \in P_{\mathcal{G}}$ if and only if M is positive definite which translates immediately into the conditions defining $P_{\mathcal{G}}^*$ in Proposition 3.4.1. ■

Theorem 3.4.2 *For the colored graph $\mathcal{G}$ as in Figure 3.1(d), $\delta > 0$ and $D \in P_{\mathcal{G}}^*$, the normalizing constant of the colored G-Wishart is*

$$\begin{aligned} I_{\mathcal{G}}(\delta, D) &= 2^{\frac{3\delta+4}{2}} \pi \Gamma(\frac{\delta}{2}) \Big(\Gamma(\frac{\delta+1}{2}) \Big)^2 (d_{11} + d_{22} + 2d_{12})^{\frac{\delta}{2}} \\ &\quad \times [d_{33}(d_{11} + d_{22} + 2d_{12}) - 4d_{13}^2]^{-\frac{\delta+1}{2}} (d_{11}d_{22} - d_{12}^2)^{-\frac{\delta+1}{2}}. \end{aligned}$$

Proof. For the proof of the theorem, it will be convenient to adopt a slightly different form of the parametrization of the Cholesky decomposition of $K = AA^t$ in $P_{\mathcal{G}}$. Let

$$A_{ij} = \begin{cases} \sqrt{a_{ii}} & \text{if } i = j, \\ -a_{ij} & \text{if } i < j, \end{cases}$$

so that

$$(AA^t)_{ij} = \begin{cases} a_{ii} + \sum_{l>i} a_{il}^2 & \text{if } i = j, \\ -a_{ij}\sqrt{a_{jj}} + \sum_{l>max(i,j)} a_{il}a_{jl} & \text{if } i < j. \end{cases}$$

Equating each entry k_{ij} of K to the corresponding entry of AA^t with the constraint that $k_{13} = k_{23}$ shows that

$$k_{11} = a_{11} + a_{12}^2 + a_{13}^2, \qquad k_{12} = -\sqrt{a_{22}}a_{12} + a_{13}a_{23},$$

$$k_{22} = a_{22} + a_{23}^2, \qquad k_{13} = -\sqrt{a_{33}}a_{13},$$

$$k_{33} = a_{33}, \qquad k_{23} = -\sqrt{a_{33}}a_{23}.$$

In particular, we find that $a_{33} > 0$, $a_{13} = a_{23}$ and $k_{12} = -\sqrt{a_{22}}a_{12} + a_{13}^2$. The Jacobian of the transformation from K to A is

$$J = \begin{array}{c} \\ a_{11} \\ a_{12} \\ a_{13} \\ a_{22} \\ a_{33} \end{array} \begin{array}{c} \begin{array}{ccccc} k_{11} & k_{12} & k_{13} & k_{22} & k_{33} \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ * & -\sqrt{a_{22}} & 0 & 0 & 0 \\ * & * & -\sqrt{a_{33}} & 2a_{13} & 0 \\ * & * & * & 1 & 0 \\ * & * & * & * & 1 \end{pmatrix} \end{array}.$$

Therefore, it is easy to see $|J| = |diag(J)| = a_{22}^{1/2} a_{33}^{1/2}$. We now have all the ingredients necessary to calculate the normalizing constant $I_{\mathcal{G}}(\delta, D)$. We have $|K| = a_{11}a_{22}a_{33}$ and

$$\begin{aligned} \langle K, D \rangle &= d_{11}k_{11} + d_{22}k_{22} + d_{33}k_{33} + 2d_{12}k_{12} + 2d_{13}k_{13} + 2d_{23}k_{23} \\ &= d_{11}(a_{11} + a_{12}^2 + a_{13}^2) + d_{22}(a_{22} + a_{23}^2) + d_{33}a_{33} \\ &\quad + 2d_{12}(-\sqrt{a_{22}}a_{12} + a_{13}a_{23}) + 2d_{13}(-a_{13}\sqrt{a_{33}}) + 2d_{23}(-a_{23}\sqrt{a_{33}}) \end{aligned}$$

and so the normalizing constant is

$$\begin{aligned} I_{\mathcal{G}}(\delta, D) &= \int_A a_{11}^{\frac{\delta-2}{2}} a_{22}^{\frac{\delta-1}{2}} a_{33}^{\frac{\delta-1}{2}} \exp\{-\frac{1}{2}d_{11}a_{11} - \frac{1}{2}d_{22}a_{22} - \frac{1}{2}d_{33}a_{33} - \frac{1}{2}d_{11}a_{12}^2 \\ &\quad -\frac{1}{2}(d_{11}+d_{22}+2d_{12})a_{13}^2 + d_{12}\sqrt{a_{22}}a_{12} + 2d_{13}a_{13}\sqrt{a_{33}}\}dA, \end{aligned}$$

where dA denotes the product of all differentials, $a_{ii} > 0$ and $a_{ij} \in R$, $i < j$. The integral with respect to a_{11} is a gamma integral with

$$\int_0^\infty a_{11}^{\frac{\delta-2}{2}} \exp\{-\frac{1}{2}d_{11}a_{11}\}da_{11} = 2^{\frac{\delta}{2}}\Gamma(\frac{\delta}{2})d_{11}^{-\frac{\delta}{2}}.$$

The integrals with respect to a_{12} and a_{13} are Gaussian integrals with

$$\int_{-\infty}^\infty \exp\{-\frac{1}{2}d_{11}a_{12}^2 + d_{12}\sqrt{a_{22}}a_{12}\}da_{12} = \frac{\sqrt{2\pi}}{\sqrt{d_{11}}}\exp\{\frac{d_{12}^2 a_{22}}{2d_{11}}\}$$

and

$$\begin{aligned} &\int_{-\infty}^\infty \exp\{-\frac{1}{2}(d_{11}+d_{22}+2d_{12})a_{13}^2 + 2d_{13}\sqrt{a_{33}}a_{13}\}da_{13} \\ = &\frac{\sqrt{2\pi}}{\sqrt{d_{11}+d_{22}+2d_{12}}}\exp\{\frac{2d_{13}^2 a_{33}}{d_{11}+d_{22}+2d_{12}}\}. \end{aligned}$$

Therefore, it follows

$$
\begin{aligned}
& I_{\mathcal{G}}(\delta, D) \\
= \; & \Gamma(\frac{\delta}{2}) 2^{\frac{\delta}{2}} d_{11}^{-\frac{\delta+1}{2}} 2\pi (d_{11}+d_{22}+2d_{12})^{-\frac{1}{2}} \\
& \int_0^\infty a_{22}^{\frac{\delta-1}{2}} a_{33}^{\frac{\delta-1}{2}} \exp\big\{(-\frac{d_{22}}{2}+\frac{d_{12}^2}{2d_{11}})a_{22} + (-\frac{d_{33}}{2}+\frac{2d_{13}^2}{d_{11}+d_{22}+2d_{12}})a_{33}\big\} da_{22} da_{33} \\
= \; & \Gamma(\frac{\delta}{2}) 2^{\frac{\delta}{2}} d_{11}^{-\frac{\delta+1}{2}} 2\pi (d_{11}+d_{22}+2d_{12})^{-\frac{1}{2}} \\
& \Gamma(\frac{\delta+1}{2}) (\frac{2d_{11}}{d_{11}d_{22}-d_{12}^2})^{\frac{\delta+1}{2}} \Gamma(\frac{\delta+1}{2}) (\frac{2(d_{11}+d_{22}+2d_{12})}{d_{33}(d_{11}+d_{22}+2d_{12})-4d_{13}^2})^{\frac{\delta+1}{2}} \\
= \; & \Gamma(\frac{\delta}{2}) \Gamma^2(\frac{\delta+1}{2}) \pi 2^{\frac{3\delta+4}{2}} (d_{11}+d_{22}+2d_{12})^{\frac{\delta}{2}} \\
& \times [d_{33}(d_{11}+d_{22}+2d_{12})-4d_{13}^2]^{-\frac{\delta+1}{2}} (d_{11}d_{22}-d_{12}^2)^{-\frac{\delta+1}{2}} \; .
\end{aligned}
$$

The proof is completed. ■

3.5 A Decomposable Graph

This colored graph is represented in Figure 3.1(e). Then the cone $P_{\mathcal{G}}$ is the set of matrices of the form

$$
K = \begin{pmatrix}
k_{11} & k_{12} & k_{13} & k_{14} \\
k_{12} & k_{22} & k_{13} & k_{14} \\
k_{13} & k_{13} & k_{33} & 0 \\
k_{14} & k_{14} & 0 & k_{33}
\end{pmatrix} .
$$

Proposition 3.5.1 *For the colored graph $\mathcal{G}$ as in Figure 3.1(e), the dual cone is the set of matrices*

$$\begin{aligned} P_{\mathcal{G}}^* &= \{D = (d_{ij})_{1\le i,j\le 4} \in S \mid d_{23} = d_{13}, d_{24} = d_{14}, d_{44} = d_{33}, d_{11}d_{22} - d_{12}^2 > 0, \\ &\quad d_{11} > 0,\ d_{11} + 2d_{12} + d_{22} > 0,\ d_{33}(d_{11} + 2d_{12} + d_{22}) - 2(d_{13}^2 + d_{14}^2) > 0\}. \end{aligned}$$

Proof. We proceed as in the proof of Proposition 3.4.1. That is, we let $K = AA^t$ be the Cholesky decomposition of K with A upper triangular. Equating the entries of K and AA^t yields

$$a_{23} = a_{13},\ a_{24} = a_{14},\ a_{44} = a_{33}$$

with

$$\begin{aligned} k_{11} = a_{11}^2 + a_{12}^2 + a_{12}^2 + a_{14}^2, \quad k_{12} = a_{12}a_{22} + a_{13}^2 + a_{14}^2,\ k_{13} = a_{13}a_{33}. \quad k_{14} = a_{14}a_{33}, \\ k_{22} = a_{22}^2 + a_{13}^2 + a_{14}^2,\ k_{23} = a_{13}a_{33}, \quad k_{14} = a_{14}a_{33}, \\ k_{33} = a_{33}^2, \qquad k_{34} = 0, \\ k_{44} = a_{33}^2. \end{aligned}$$

Then, ordering $\langle K, D\rangle$ as a polynomial in a_{ij}, we see that

$$\begin{aligned} \langle K, D\rangle &= d_{11}a_{11}^2 + d_{22}a_{22}^2 + 2d_{33}a_{33}^2 + d_{11}a_{12}^2 + 2d_{12}a_{22}a_{12} + a_{13}^2(d_{11} + 2d_{12} + d_{22}) \\ &\quad + 4d_{13}a_{13}a_{33} + a_{14}^2(d_{11} + 2d_{12} + d_{22}) + 4d_{14}a_{14}a_{33} \end{aligned}$$

is a quadratic form and the matrix of this quadratic form is

$$M = \begin{pmatrix} d_{11} & 0 & 0 & 0 & 0 & 0 \\ 0 & d_{22} & d_{12} & 0 & 0 & 0 \\ 0 & d_{12} & d_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & 2d_{33} & 2d_{13} & 2d_{14} \\ 0 & 0 & 0 & 2d_{13} & d_{11}+2d_{12}+d_{22} & 0 \\ 0 & 0 & 0 & 2d_{14} & 0 & d_{11}+2d_{12}+d_{22} \end{pmatrix}.$$

With exactly the same argument as in Proposition 3.4.1, we can show that $\langle K, D\rangle > 0$ for all $K \in \bar{P}_{\mathcal{G}}$ if and only if $M > 0$, i.e. D satisfies the conditions of Proposition 3.5.1. ■

Theorem 3.5.2 *For the colored graph $\mathcal{G}$ as in Figure 3.1(e), $\delta > 0$ and $D \in P_{\mathcal{G}}^*$, the normalizing constant of the colored G-Wishart is*

$$\begin{aligned} I_{\mathcal{G}}(\delta, D) &= 2^{\delta+2}\pi^{\frac{3}{2}}\Gamma(\frac{\delta}{2})\Gamma(\frac{\delta+1}{2})\Gamma(\delta)(d_{11}+d_{22}+2d_{12})^{\delta-1}(d_{11}d_{22}-d_{12}^2)^{-\frac{\delta+1}{2}} \\ &\quad \times[d_{33}(d_{11}+d_{22}+2d_{12})-2(d_{13}^2+d_{14}^2)]^{-\delta}. \end{aligned}$$

Proof. As in the proof of Theorem 3.4.2, it will be convenient to adopt a slightly different parametrization of the Cholesky decomposition of K. Let

$$A_{ij} = \begin{cases} \sqrt{a_{ii}} & \text{if } i = j, \\ -a_{ij} & \text{if } i < j, \end{cases}$$

so that the entries of AA^t are given by

$$(AA^t)_{ij} = \begin{cases} a_{ii} + \sum_{l>i} a_{il}^2 & \text{if } i = j, \\ -a_{ij}\sqrt{a_{jj}} + \sum_{l>j} a_{il}a_{jl} & \text{if } i < j. \end{cases}$$

Equating each entry k_{ij} of K to the corresponding entry of AA^t, we find that

$$k_{11} = a_{11} + a_{12}^2 + a_{13}^2 + a_{14}^2, \qquad k_{12} = -\sqrt{a_{22}}a_{12} + a_{13}a_{23} + a_{14}a_{24},$$

$$k_{13} = -\sqrt{a_{33}}a_{13} + a_{14}a_{34}, \qquad k_{14} = -\sqrt{a_{44}}a_{14},$$

$$k_{22} = a_{22} + a_{23}^2 + a_{24}^2, \qquad k_{23} = -\sqrt{a_{33}}a_{23} + a_{24}a_{34},$$

$$k_{24} = -\sqrt{a_{44}}a_{24}, \qquad k_{33} = a_{33} + a_{34}^2,$$

$$k_{34} = -\sqrt{a_{44}}a_{34}, \qquad k_{44} = a_{44}.$$

This shows that $a_{44} > 0$ and $a_{34} = 0$. Since $a_{33} > 0$ and $k_{13} = k_{23}$, then $a_{13} = a_{23}$. Since $a_{44} > 0$ and $k_{14} = k_{24}$, then $a_{14} = a_{24}$. Since $k_{34} = 0$, then $a_{33} = a_{44}$. Therefore, we obtain that

$$k_{11} = a_{11} + a_{12}^2 + a_{13}^2 + a_{14}^2, \qquad k_{12} = -\sqrt{a_{22}}a_{12} + a_{13}^2 + a_{14}^2,$$

$$k_{13} = k_{23} = -\sqrt{a_{33}}a_{13}, \qquad k_{14} = k_{24} = -\sqrt{a_{33}}a_{14},$$

$$k_{22} = a_{22} + a_{13}^2 + a_{14}^2, \qquad k_{33} = k_{44} = a_{33}.$$

The Jacobian of the transformation from K to A is

$$J = \begin{array}{c} \\ a_{11} \\ a_{12} \\ a_{13} \\ a_{14} \\ a_{22} \\ a_{33} \end{array} \begin{array}{c} \begin{array}{cccccc} k_{11} & k_{12} & k_{13} & k_{14} & k_{22} & k_{33} \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ * & -\sqrt{a_{22}} & 0 & 0 & 0 & 0 \\ * & * & -\sqrt{a_{33}} & 0 & 2a_{13} & 0 \\ * & * & * & -\sqrt{a_{33}} & 2a_{14} & 0 \\ * & * & * & * & 1 & 0 \\ * & * & * & * & * & 1 \end{pmatrix} \end{array}.$$

It is easy to see $|J| = |diag(J)| = a_{22}^{1/2} a_{33}$. We now have all the ingredients necessary to calculate the normalizing constant $I_G(\delta, D)$. Through the change of variables, $K = AA^t$. Then $|K| = a_{11}a_{22}a_{33}^2$,

$$\begin{aligned} \langle K, D \rangle &= d_{11}k_{11} + d_{22}k_{22} + d_{33}k_{33} + d_{44}k_{44} + 2d_{12}k_{12} + 2d_{13}k_{13} + 2d_{14}k_{14} \\ &\quad + 2d_{23}k_{23} + 2d_{24}k_{24} + 2d_{34}k_{34} \\ &= d_{11}(a_{11} + a_{12}^2 + a_{13}^2 + a_{14}^2) + d_{22}(a_{22} + a_{13}^2 + a_{14}^2) + 2d_{33}a_{33} \\ &\quad + 2d_{12}(-\sqrt{a_{22}}a_{12} + a_{13}^2 + a_{14}^2) + 4d_{13}(-a_{13}\sqrt{a_{33}}) + 4d_{14}(-a_{14}\sqrt{a_{33}}) \end{aligned}$$

and so the integral equals

$$
\begin{aligned}
& I_{\mathcal{G}}(\delta, D) \\
= & \int_A a_{11}^{\frac{\delta-2}{2}} a_{22}^{\frac{\delta-1}{2}} a_{33}^{\delta-1} \exp\Big\{ -\frac{1}{2} d_{11} a_{11} - \frac{1}{2} d_{11} a_{12}^2 + d_{12}\sqrt{a_{22}} a_{12} - \frac{1}{2}(d_{11} + d_{22} + 2d_{12}) a_{13}^2 \\
& + 2d_{13} a_{13}\sqrt{a_{33}} - \frac{1}{2}(d_{11} + d_{22} + 2d_{12}) a_{14}^2 + 2d_{14} a_{14}\sqrt{a_{33}} - \frac{1}{2} d_{22} a_{22} - d_{33} a_{33} \Big\} dA,
\end{aligned}
$$

where dA denotes the product of all differentials, $a_{ii} > 0$ and $a_{ij} \in R$, $i < j$. Since the integral with respect to a_{11} is a gamma integral, then

$$
\int_0^\infty a_{11}^{\frac{\delta-2}{2}} \exp\{-\frac{1}{2} d_{11} a_{11}\} da_{11} = 2^{\frac{\delta}{2}} \Gamma(\frac{\delta}{2}) d_{11}^{-\frac{\delta}{2}}.
$$

Since the integrals with respect to a_{12}, a_{13} and a_{14} are normal integrals, then

$$
\int_{-\infty}^\infty \exp\{-\frac{1}{2} d_{11} a_{12}^2 + d_{12}\sqrt{a_{22}} a_{12}\} da_{12} = \frac{\sqrt{2\pi}}{\sqrt{d_{11}}} \exp\{\frac{d_{12}^2 a_{22}}{2d_{11}}\},
$$

$$
\begin{aligned}
& \int_{-\infty}^\infty \exp\{-\frac{1}{2}(d_{11} + d_{22} + 2d_{12}) a_{13}^2 + 2d_{13}\sqrt{a_{33}} a_{13}\} da_{13} \\
= & \frac{\sqrt{2\pi}}{\sqrt{d_{11} + d_{22} + 2d_{12}}} \exp\{\frac{2d_{13}^2 a_{33}}{d_{11} + d_{22} + 2d_{12}}\},
\end{aligned}
$$

and

$$
\begin{aligned}
& \int_{-\infty}^\infty \exp\{-\frac{1}{2}(d_{11} + d_{22} + 2d_{12}) a_{14}^2 + 2d_{14}\sqrt{a_{33}} a_{14}\} da_{14} \\
= & \frac{\sqrt{2\pi}}{\sqrt{d_{11} + d_{22} + 2d_{12}}} \exp\{\frac{2d_{14}^2 a_{33}}{d_{11} + d_{22} + 2d_{12}}\}.
\end{aligned}
$$

Therefore, the integral becomes

$$
\begin{aligned}
& I_{\mathcal{G}}(\delta, D) \\
= \; & \Gamma(\frac{\delta}{2}) 2^{\frac{\delta}{2}} d_{11}^{-\frac{\delta+1}{2}} (2\pi)^{\frac{3}{2}} (d_{11}+d_{22}+2d_{12})^{-1} \\
& \times \int_0^\infty a_{22}^{\frac{\delta-1}{2}} a_{33}^{\frac{\delta-1}{2}} \exp\{(-\frac{d_{22}}{2} + \frac{d_{12}^2}{2d_{11}}) a_{22} + (-d_{33} + \frac{2d_{13}^2 + 2d_{14}^2}{d_{11}+d_{22}+2d_{12}}) a_{33}\} da_{22} da_{33} \\
= \; & \Gamma(\frac{\delta}{2}) 2^{\frac{\delta+3}{2}} d_{11}^{-\frac{\delta+1}{2}} \pi^{\frac{3}{2}} (d_{11}+d_{22}+2d_{12})^{-1} \\
& \times \Gamma(\frac{\delta+1}{2}) (\frac{2d_{11}}{d_{11}d_{22}-d_{12}^2})^{\frac{\delta+1}{2}} \Gamma(\delta) (\frac{d_{11}+d_{22}+2d_{12}}{d_{33}(d_{11}+d_{22}+2d_{12}) - 2(d_{13}^2+d_{14}^2)})^{\delta} \\
= \; & \Gamma(\frac{\delta}{2}) \Gamma(\frac{\delta+1}{2}) \Gamma(\delta) \pi^{\frac{3}{2}} 2^{\delta+2} (d_{11}+d_{22}+2d_{12})^{\delta-1} \\
& \times [d_{33}(d_{11}+d_{22}+2d_{12}) - 2(d_{13}^2+d_{14}^2)]^{-\delta} (d_{11}d_{22}-d_{12}^2)^{-\frac{\delta+1}{2}} \; .
\end{aligned}
$$

The proof is completed. ■

4 Numerical Experiments for the Metropolis-Hastings

In this chapter, we examine the effectiveness of the proposed methods. Simulations are conducted in several scenarios, including various colored graphs, situations of small p and large p.

4.1 Simulation Results for the Special Colored Graphs

In order to illustrate the performance of our MH algorithm, we conduct a numerical experiment for colored graphical models represented by the colored graphs (a) - (e) shown in Figure 3.1. In each case, for the given hyperparameters D and δ, we first derive $\log I_G(\delta, D)$, then obtain the prior mean $E(K)$ under the colored G-Wishart by differentiating $\log I_G(\delta, D)$ with respect to $-\frac{D}{2}$. We then generate the samples from the colored G-Wishart distribution. We run the independent chain for 5000 iterations and discard the first 1000 samples as burn in. Our estimate $\hat{K}$ for K

is the average $\hat{K} = \frac{\sum_{i=1001}^{i=5000} \hat{K}_i}{4000}$ of the remaining 4000 iterations $\hat{K}_i$, $i = 1001, \ldots, 5000$. For arbitrary K and K', we define the normalized mean square error ($NMSE$) and the normalized mean absolutely error ($NMAE$) between K and K' as

$$NMSE(K, K') = \frac{||K - K'||_2^2}{||K'||_2^2}$$

and

$$NMAE(K, K') = \frac{||K - K'||_2}{||K'||_2}$$

where $||K||_2^2$ is the sum of the squares of the elements of K. We repeat the previous experiment 100 times to obtain $\hat{K}^j$, $j = 1, \ldots, 100$ and calculate

$$\overline{NMSE}(\hat{K}, E(K)) = \frac{1}{100} \sum_{j=1}^{100} NMSE(\hat{K}^j, E(K))$$

and

$$\overline{NMAE}(\hat{K}, E(K)) = \frac{1}{100} \sum_{j=1}^{100} NMAE(\hat{K}^j, E(K))$$

where $E(K)$ is obtained by the differentiation of $\log I_{\mathcal{G}}(\delta, D)$ with respect to $-\frac{D}{2}$ at our given hyperparameters D and δ.

For each colored graph in Figure 3.1, for an arbitrary $j \in \{1, \ldots, 100\}$, we give the traceplot of $\log |K_i^j|$, $i = 1000, \ldots, 5000$. The traceplot shows that the independent chain seems to be mixing well. We also give the autocorrelation plot with time-lag h for $\log |K_i^j|$, $i = 1000, \ldots, 5000$, in function of h where, for an arbitrary given j,

we define the autocorrelation coefficient for $Y_i = \log|K_i^j|,\ i = 1000, \ldots, 5000$, to be

$$R_h = \frac{\sum_{i=1000}^{5000-h}(Y_i - \bar{Y})(Y_{i+h} - \bar{Y})}{\sum_{i=1000}^{5000}(Y_i - \bar{Y})^2}.$$

The autocorrelation plots indicate that the samples from the MH algorithm have low autocorrelations. The numerical values of the matrices D, $E(K)$ and $\hat{K}$ as well as the traceplot and autocorrelation plot of $\log(|K|)$ for all five colored graphs in Figure 3.1 are given as follows.

In practice, we usually choose a small value of δ and the identity matrix D as the hyperparamters in the colored G-Wishart distribution to reduce the effect of the priors on the likelihood. In the following simulations, we just randomly choose a δ and D for the five colored graphs in Figure 3.1 to illustrate the performance of the MH algorithm.

Graph in Figure 3.1(a)

For the colored tree in Figure 3.1(a), we choose the hyperparameters $\delta = 1$ and

$$D = \begin{pmatrix} 1 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 5 & 2 & 0 & 0 & 0 \\ 2 & 2 & 2 & 25 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2 & 6 & 2 & 2 \\ 0 & 0 & 0 & 0 & 2 & 3 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 4 \end{pmatrix}.$$

The true mean of K can be computed as

$$E(K) = \begin{pmatrix} 1.1294 & 0 & 0 & -0.0129 & 0 & 0 & 0 \\ 0 & 0.5915 & 0 & -0.0129 & 0 & 0 & 0 \\ 0 & 0 & 0.2578 & -0.0129 & 0 & 0 & 0 \\ -0.0129 & -0.0129 & -0.0129 & 0.0767 & -0.0129 & 0 & 0 \\ 0 & 0 & 0 & -0.0129 & 0.2589 & -0.0129 & -0.0129 \\ 0 & 0 & 0 & 0 & -0.0129 & 0.3699 & 0 \\ 0 & 0 & 0 & 0 & -0.0129 & 0 & 0.2817 \end{pmatrix}.$$

The sample mean of K^j for a random j and the mean of $\hat{K}$ over 100 simulations are

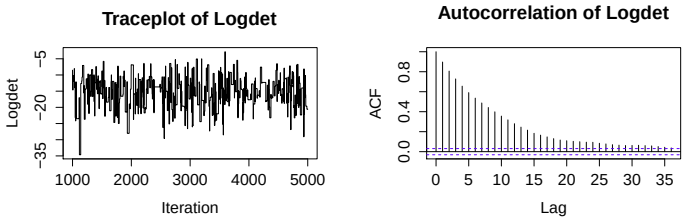
$$\hat{K}^j = \begin{pmatrix} 1.1274 & 0 & 0 & -0.0127 & 0 & 0 & 0 \\ 0 & 0.5961 & 0 & -0.0127 & 0 & 0 & 0 \\ 0 & 0 & 0.2563 & -0.0127 & 0 & 0 & 0 \\ -0.0127 & -0.0127 & -0.0127 & 0.0767 & -0.0127 & 0 & 0 \\ 0 & 0 & 0 & -0.0127 & 0.2594 & -0.0127 & -0.0127 \\ 0 & 0 & 0 & 0 & -0.0127 & 0.3708 & 0 \\ 0 & 0 & 0 & 0 & -0.0127 & 0 & 0.2818 \end{pmatrix}$$

and $E(\hat{K}) = \begin{pmatrix} 1.1287 & 0 & 0 & -0.0163 & 0 & 0 & 0 \\ 0 & 0.5436 & 0 & -0.0163 & 0 & 0 & 0 \\ 0 & 0 & 0.26733 & -0.0163 & 0 & 0 & 0 \\ -0.0163 & -0.0163 & -0.0163 & 0.0875 & -0.0163 & 0 & 0 \\ 0 & 0 & 0 & -0.0163 & 0.2596 & -0.0163 & -0.0163 \\ 0 & 0 & 0 & 0 & -0.0163 & 0.4116 & 0 \\ 0 & 0 & 0 & 0 & -0.0163 & 0 & 0.2825 \end{pmatrix}$.

The traceplot and the autocorrelation plot for $\log(|K|)$ are shown in Figure 4.1.

Graph in Figure 3.1(b)

For the colored star with the centre vertex of a different color in Figure 3.1(b), we choose the hyperparameters $\delta = 3$ and



(a) Traceplot (b) ACF plot

Figure 4.1: (a) Traceplot of $\log(|K|)$ v.s. the number of iterations. (b) Autocorrelation plot of $\log(|K|)$ for Figure 3.1(a).

$$D = \begin{pmatrix} 9 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 1 & 25 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 25 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & 25 & 0 & 0 & 0 & 0 & 0 \\ 4 & 0 & 0 & 0 & 25 & 0 & 0 & 0 & 0 \\ 5 & 0 & 0 & 0 & 0 & 25 & 0 & 0 & 0 \\ 6 & 0 & 0 & 0 & 0 & 0 & 25 & 0 & 0 \\ 7 & 0 & 0 & 0 & 0 & 0 & 0 & 25 & 0 \\ 8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 25 \end{pmatrix}.$$

The true mean of K can be computed as

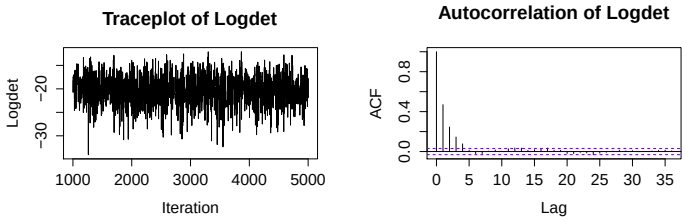
$$E(K) = \begin{pmatrix} 1.4778 & -0.0112 & -0.0225 & -0.0338 & -0.0451 & -0.0563 & -0.0676 & -0.0789 & -0.0902 \\ -0.0112 & 0.1015 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.0225 & 0 & 0.1015 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.0338 & 0 & 0 & 0.1015 & 0 & 0 & 0 & 0 & 0 \\ -0.0451 & 0 & 0 & 0 & 0.1015 & 0 & 0 & 0 & 0 \\ -0.0563 & 0 & 0 & 0 & 0 & 0.1015 & 0 & 0 & 0 \\ -0.0676 & 0 & 0 & 0 & 0 & 0 & 0.1015 & 0 & 0 \\ -0.0789 & 0 & 0 & 0 & 0 & 0 & 0 & 0.1015 & 0 \\ -0.0902 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.1015 \end{pmatrix}.$$

The sample mean of K^j for a random j and the mean of $\hat{K}$ over 100 simulations are

$$\hat{K^j} = \begin{pmatrix} 1.4690 & -0.0113 & -0.0223 & -0.0341 & -0.0455 & -0.0569 & -0.0677 & -0.0796 & -0.0905 \\ -0.0113 & 0.1016 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.0223 & 0 & 0.1016 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.0341 & 0 & 0 & 0.1016 & 0 & 0 & 0 & 0 & 0 \\ -0.0455 & 0 & 0 & 0 & 0.1016 & 0 & 0 & 0 & 0 \\ -0.0569 & 0 & 0 & 0 & 0 & 0.1016 & 0 & 0 & 0 \\ -0.0677 & 0 & 0 & 0 & 0 & 0 & 0.1016 & 0 & 0 \\ -0.0796 & 0 & 0 & 0 & 0 & 0 & 0 & -0.1016 & 0 \\ -0.0905 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.1016 \end{pmatrix} \text{ and}$$

$$E(\hat{K}) = \begin{pmatrix} 1.4697 & -0.0114 & -0.0224 & -0.0340 & -0.0454 & -0.0569 & -0.0670 & -0.0793 & -0.0902 \\ -0.0114 & 0.1016 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.0224 & 0 & 0.1016 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.0340 & 0 & 0 & 0.1016 & 0 & 0 & 0 & 0 & 0 \\ -0.0454 & 0 & 0 & 0 & 0.1016 & 0 & 0 & 0 & 0 \\ -0.0569 & 0 & 0 & 0 & 0 & 0.1016 & 0 & 0 & 0 \\ -0.0670 & 0 & 0 & 0 & 0 & 0 & 0.1016 & 0 & 0 \\ -0.0793 & 0 & 0 & 0 & 0 & 0 & 0 & -0.1016 & 0 \\ -0.0902 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.1016 \end{pmatrix}$$

The traceplot and the autocorrelation plot for $\log(|K|)$ are shown in Figure 4.2.



(a) Traceplot (b) ACF plot

Figure 4.2: Traceplot and Autocorrelation plot of $\log(|K|)$ for Graph in Figure 3.1(b).

Graph in Figure 3.1(c)

For the colored star with all vertices of the same color in Figure 3.1(c), we choose the hyperparameters $\delta = 3$ and

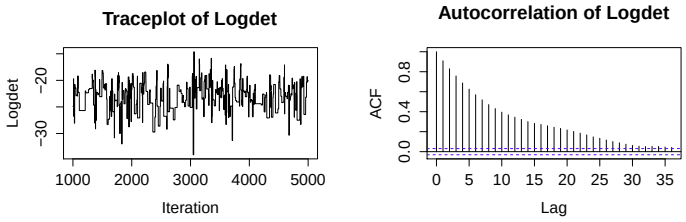
$$D = \begin{pmatrix} 25 & 9 & 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \\ 9 & 25 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 8 & 0 & 25 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 7 & 0 & 0 & 25 & 0 & 0 & 0 & 0 & 0 & 0 \\ 6 & 0 & 0 & 0 & 25 & 0 & 0 & 0 & 0 & 0 \\ 5 & 0 & 0 & 0 & 0 & 25 & 0 & 0 & 0 & 0 \\ 4 & 0 & 0 & 0 & 0 & 0 & 25 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & 0 & 0 & 0 & 25 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 25 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 25 \end{pmatrix}.$$

The true mean of K can be computed as

$$E(K) = \begin{pmatrix} 0.1229 & -0.0013 & -0.0026 & -0.0039 & -0.0052 & -0.0065 & -0.0078 & -0.0091 & -0.0104 & -0.0117 \\ -0.0013 & 0.1229 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.0026 & 0 & 0.1229 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.0039 & 0 & 0 & 0.1229 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.0052 & 0 & 0 & 0 & 0.1229 & 0 & 0 & 0 & 0 & 0 \\ -0.0065 & 0 & 0 & 0 & 0 & 0.1229 & 0 & 0 & 0 & 0 \\ -0.0078 & 0 & 0 & 0 & 0 & 0 & 0.1229 & 0 & 0 & 0 \\ -0.0091 & 0 & 0 & 0 & 0 & 0 & 0 & 0.1229 & 0 & 0 \\ -0.0104 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.1229 & 0 \\ -0.0117 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.1229 \end{pmatrix}.$$

The sample mean of K^j for a random j and the mean of $\hat{K}$ over 100 simulations are

$$\hat{K^j} = \begin{pmatrix} 0.1223 & -0.0012 & -0.0027 & -0.0041 & -0.0055 & -0.0064 & -0.0077 & -0.0090 & -0.0102 & -0.0115 \\ -0.0012 & 0.1223 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.0027 & 0 & 0.1223 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.0041 & 0 & 0 & 0.1223 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.0055 & 0 & 0 & 0 & 0.1223 & 0 & 0 & 0 & 0 & 0 \\ -0.0064 & 0 & 0 & 0 & 0 & 0.1223 & 0 & 0 & 0 & 0 \\ -0.0077 & 0 & 0 & 0 & 0 & 0 & 0.1223 & 0 & 0 & 0 \\ -0.0090 & 0 & 0 & 0 & 0 & 0 & 0 & 0.1223 & 0 & 0 \\ -0.0102 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.1223 & 0 \\ -0.0115 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.1223 \end{pmatrix}$$



(a) Traceplot (b) ACF plot

Figure 4.3: Traceplot and Autocorrelation plot of $\log(|K|)$ for Graph in Figure 3.1(c).

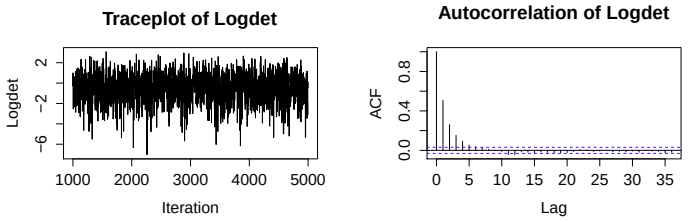
and $E(\hat{K}) =$

$$\begin{pmatrix}
0.1225 & -0.0012 & -0.0028 & -0.0044 & -0.0051 & -0.0063 & -0.0079 & -0.0090 & -0.0103 & -0.0116 \\
-0.0012 & 0.1225 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-0.0028 & 0 & 0.1225 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-0.0044 & 0 & 0 & 0.1225 & 0 & 0 & 0 & 0 & 0 & 0 \\
-0.0051 & 0 & 0 & 0 & 0.1225 & 0 & 0 & 0 & 0 & 0 \\
-0.0063 & 0 & 0 & 0 & 0 & 0.1225 & 0 & 0 & 0 & 0 \\
-0.0079 & 0 & 0 & 0 & 0 & 0 & 0.1225 & 0 & 0 & 0 \\
-0.0090 & 0 & 0 & 0 & 0 & 0 & 0 & 0.1225 & 0 & 0 \\
-0.0103 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.1225 & 0 \\
-0.0116 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.1225
\end{pmatrix}.$$

The traceplot and the autocorrelation plot for $\log(|K|)$ are shown in Figure 4.3.

Graph in Figure 3.1(d)

For the triangle with two edges of the same color in Figure 3.1(d), one hyperpa-



(a) Traceplot (b) ACF plot

Figure 4.4: Traceplot and Autocorrelation plot of $\log(|K|)$ for Graph in Figure 3.1(d).

rameter $\delta = 3$. The other hyperparamete D and the true mean of K are as follows

$$D = \begin{pmatrix} 3 & 1 & 2 \\ 1 & 4 & 2 \\ 2 & 2 & 5 \end{pmatrix} \quad \text{and} \quad E(K) = \begin{pmatrix} 1.8108 & -0.0073 & -0.5517 \\ -0.0073 & 1.4472 & -0.5517 \\ -0.5517 & -0.5517 & 1.2413 \end{pmatrix}.$$

The sample mean of K^j for a random j and the mean of $\hat{K}$ over 100 simulations are

$$\hat{K^j} = \begin{pmatrix} 1.8097 & -0.0075 & -0.5514 \\ -0.0075 & 1.4485 & -0.5514 \\ -0.5514 & -0.5514 & 1.2442 \end{pmatrix} \text{ and}$$

$$E(\hat{K}) = \begin{pmatrix} 1.7900 & -0.0054 & -0.5487 \\ -0.0054 & 1.4443 & -0.5487 \\ -0.5487 & -0.5487 & 1.2487 \end{pmatrix}.$$

The traceplot and the autocorrelation plot for $\log(|K|)$ are shown in Figure 4.4.

Graph in Figure 3.1 (e)

For the decomposable graph with three different colors for edges in Figure 3.1(e), one hyperparameter $\delta = 3$. The other hyperparameter D and the true mean of K are as follows:

$$D = \begin{pmatrix} 2 & 1 & 3 & 4 \\ 1 & 1 & 3 & 4 \\ 3 & 3 & 200 & 0 \\ 4 & 4 & 0 & 200 \end{pmatrix} \quad \text{and} \quad E(K) = \begin{pmatrix} 4.4631 & -3.5368 & -0.0189 & -0.0252 \\ -3.5368 & 8.4631 & -0.0189 & -0.0252 \\ -0.0189 & -0.0189 & 0.0157 & 0 \\ -0.0252 & -0.0252 & 0 & 0.0157 \end{pmatrix}.$$

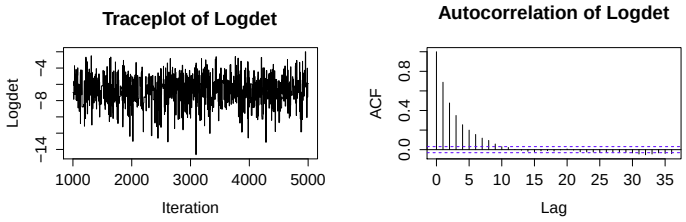
The sample mean of K^j for a random j and the mean of $\hat{K}$ over 100 simulations are

$$\hat{K^j} = \begin{pmatrix} 4.4714 & -3.5386 & -0.0192 & -0.0256 \\ -3.5386 & 8.4658 & -0.0192 & -0.0256 \\ -0.0192 & -0.0192 & 0.0158 & 0 \\ -0.0256 & -0.0256 & 0 & 0.0158 \end{pmatrix}$$

and

$$E(\hat{K}) = \begin{pmatrix} 4.4726 & -3.5374 & -0.0187 & -0.0258 \\ -3.5374 & 8.4658 & -0.0187 & -0.0258 \\ -0.0187 & -0.0187 & 0.0164 & 0 \\ -0.0258 & -0.0258 & 0 & 0.0164 \end{pmatrix}.$$

The traceplot and the autocorrelation plot for $\log(|K|)$ are shown in Figure 4.5.



(a) Traceplot (b) ACF plot

Figure 4.5: Traceplot and Autocorrelation plot of $\log(|K|)$ for Graph in Figure 3.1(e).

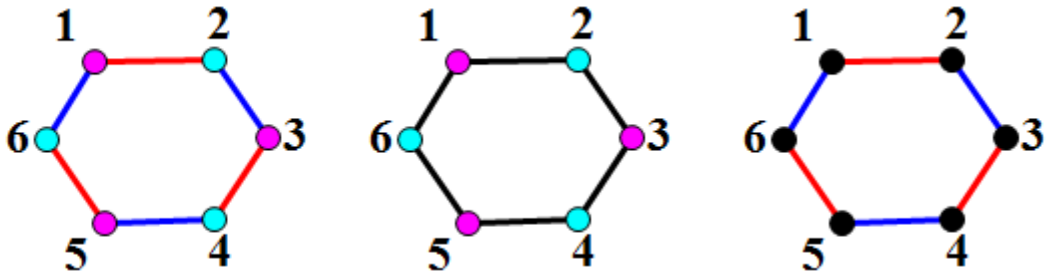
A summary of calculations and results are given in Table 4.1 which, for all different five colored graphs in Figure 3.1, shows the hyperparameter δ we chose for the prior distribution, $\log I_{\mathcal{G}}(\delta, D)$ and the normalized mean square errors. In order to obtain the true mean $E(K)$ of the colored G-Wishart distribution for the colored graph in Figure 3.1(c), we use formula (3.1) to get the derivative of the hypergeometric function ${}_pF_q(a_1, \ldots, a_p; b_1, \ldots, b_q; z)$ defined in Chapter 1. We see that from Table 4.1 the normalized mean square error is of the order of 10^{-3} or less except for the star graph with all leaves of the same color in Figure 3.1(b).

4.2 Posterior Mean from the Simulated Data: $p = 20$, $p = 30$

In order to assess the accuracy of our MH method for larger colored graphs, we generate data from a $N(0, K^{-1})$ distribution with K given in $P_{\mathcal{G}}$. We take the colored G-Wishart distribution with hyperparameters $\delta = 3$ and $D = I$ as the prior

$\mathcal{G}$	δ	$\log I_{\mathcal{G}}(\delta, D)$	$\overline{NMSE}$	$\overline{NMAE}$
(a)	1	$-\frac{1}{2}\sum_{i=1}^{7}\log a'_i + \log\left[\frac{1}{\sum_{i=1}^{6}(a'_i a'_{j_i})^{\frac{1}{2}}-6b'} - \frac{1}{\sum_{i=1}^{6}(a'_i a'_{j_i})^{\frac{1}{2}}+6b'}\right]$	0.0069	0.0517
(b)	3	$\frac{7}{2}\log a' - 9\log(8a'c' - \|b'\|^2)$	0.0187	0.3196
(c)	3	$-15\log a' + \log\ {}_2F_1\left(\frac{15}{2}, 8; 6; \frac{\|b'\|^2}{25a'^2}\right)$	0.0064	0.2080
(d)	3	$\frac{3}{2}\log d - 2\log(d_{33}d - 4d_{13}^2) - 2\log(d_{11}d_{22} - d_{12}^2)$	0.0005	0.0071
(e)	3	$2\log d - 3\log(d_{33}d - 2d_{13}^2 - 2d_{14}^2) - 2\log(d_{11}d_{22} - d_{12}^2)$	0.0009	0.0112

Table 4.1: For the graphs of Figure 3.1 and δ given: analytic expression of $\log I_{\mathcal{G}}(\delta, D)$ where $d = d_{11} + d_{22} + 2d_{12}$, and values of $\overline{NMSE}(\hat{K}, E(K))$ and $\overline{NMAE}(\hat{K}, E(K))$ averaged over 100 experiments.



(a) (b) (c)

Figure 4.6: Cycles of length 6 with three different patterns of coloring that we use for the cycles of length $p = 20$ and $p = 30$. Black vertices or edges indicate different arbitrary colors.

distribution of the precision matrix K. Clearly the posterior distribution will be the colored G-Wishart distribution with parameters $\delta + n$ and $I + nS$ where S is the sample covariance matrix. We will use the sampling method proposed in Chapter 2 to obtain the samples from this posterior distribution.

We conduct our experiment with six different colored graphs. For three of them, the skeleton is a colored cycle of order $p = 20$ and for the other three, the skeleton is a colored cycle of order $p = 30$. For each cycle of order $p = 20$ or $p = 30$, we give three different patterns of coloring which, for the sake of saving space, are illustrated in Figure 4.6 for $p = 6$. The values for the elements of K for all three types of graphs

are given as follows:

$$K_{ii} = 0.1,\ i = 1, 3, \ldots, 2p-1,\ \ K_{ii} = 0.03.\ i = 2, 4, \ldots, p,$$

$$K_{i,i+1} = K_{i+1,i} = 0.01,\ i = 1, 2, \ldots, p-1,\ \ K_{1p} = K_{p1} = 0.01.$$

For each colored graph, we generate 100 datasets from the multivariate normal $N(0, K^{-1})$ distribution. The posterior mean estimates are the sample mean of K based on 5000 iterations after the first 1000 burn-in iterations. We denote $\hat{K} = (\hat{K}_{ij})_{1 \leq i,j \leq p}$ as the posterior mean estimate of K.

Table 4.2 shows $\overline{NMSE}(\hat{K}, K)$ for the three colored models when $p = 20$ and $p = 30$, averaged over 100 simulations. Standard errors are indicated in parentheses. Computations were performed on a 2 core 4 threads with i5-4200U, 2.3 GHZ chips and 8GB of RAM, running on Windows 8. We also give in Table 4.2 the computing time per simulation in minutes.

For the colored graphs of Figure 4.6 with $p = 20$ and $p = 30$, we give the estimates of the elements of K together with their batch standard errors. The estimates and batch standard errors are given below for the elements of K listed in the lexicographic order. Table 4.3 shows the estimates and batch standard errors for the precision matrix K in Figure 4.6(a). Tables 4.4 and 4.5 illustrate the estimates and batch standard errors for the precision matrix K in Figure 4.6(b) when $p = 20$, respectively. Tables 4.6 and 4.7 show the estimates and batch standard errors for the precision

$\mathcal{G}$	$p = 20$		$p = 30$	
	$\overline{NMSE}(\hat{K}, K)$	Time/sim	$\overline{NMSE}(\hat{K}, K)$	Time/sim
Figure 4.6(a)	0.005 (0.003)	19.425	0.040 (0.021)	86.423
Figure 4.6(b)	0.011 (0.003)	18.739	0.033 (0.011)	82.876
Figure 4.6(c)	0.039 (0.021)	16.410	0.080 (0.033)	82.563

Table 4.2: $\overline{NMSE}(\hat{K}, K)$ for the three colored models when $p = 20$ and $p = 30$.

matrix K in Figure 4.6(b) when $p = 30$, respectively. Tables 4.8 and 4.9 show the estimates and batch standard errors for the precision matrix K in Figure 4.6(c) when $p = 20$, respectively. Tables 4.10 and 4.11 show the estimates and batch standard errors for the precision matrix K in Figure 4.6(c) when $p = 30$, respectively.

p	K_{11}	K_{12}	K_{1p}	K_{22}
20	0.1040 (0.0005)	0.0103 (0.0002)	0.0104 (0.0002)	0.0313 (0.0001)
30	0.1223 (0.0009)	0.0121 (0.0004)	0.0125 (0.0004)	0.0361 (0.0003)

Table 4.3: The average estimates and batch standard errors for K in Figure 4.6(a).

0.1072	0.0100	0.0096	0.0322	0.0109	0.0093	0.0102	0.0105	0.0101
0.0103	0.0099	0.0106	0.0100	0.0099	0.0109	0.0104	0.0111	0.0116
0.0104	0.0100	0.0113	0.0115					

Table 4.4: The average estimates for entries of K for Figure 4.6(b) when $p = 20$.

0.0004	0.0005	0.0005	0.0001	0.0005	0.0005	0.0005	0.0005
0.0005	0.0005	0.0005	0.0004	0.0005	0.0004	0.0005	0.0004
0.0005	0.0005	0.0005	0.0005	0.0004	0.0005		

Table 4.5: The batch standard errors for Figure 4.6(b) when $p = 20$.

0.1217	0.0109	0.0126	0.0366	0.0109	0.0118	0.0120	0.0120	0.0115	0.0122
0.0108	0.0121	0.0113	0.0119	0.0125	0.0114	0.0120	0.0112	0.0119	0.0131
0.0115	0.0125	0.0116	0.0132	0.0110	0.0119	0.0119	0.0107	0.0129	0.0119
0.0124	0.0119								

Table 4.6: The average estimates for entries of K for Figure 4.6(b) when $p = 30$.

0.0008	0.0006	0.0006	0.0003	0.0006	0.0006	0.0006	0.0006
0.0006	0.0006	0.0006	0.0006	0.0006	0.0006	0.0006	0.0006
0.0006	0.0006	0.0006	0.0006	0.0006	0.0006	0.0006	0.0006
0.0006	0.0005	0.0006	0.0005	0.0006	0.0006	0.0005	0.0005

Table 4.7: The batch standard errors for Figure 4.6(b) when $p = 30$.

0.1102	0.0106	0.0104	0.0347	0.1135	0.0329	0.1104	0.0335	0.1113	0.0332
0.1103	0.0326	0.1157	0.0330	0.1082	0.0333	0.1083	0.0318	0.1096	0.0326
0.1059	0.0311								

Table 4.8: The average estimates for entries of K for Figure 4.6(c) when $p = 20$.

0.0011	0.0002	0.0002	0.0004	0.0012	0.0003	0.0011	0.0004
0.0012	0.0004	0.0012	0.0003	0.0013	0.0003	0.0012	0.0003
0.0012	0.0004	0.0011	0.0003	0.0012	0.0003		

Table 4.9: The batch standard errors for Figure 4.6(c) when $p = 20$.

0.1295	0.0117	0.0111	0.0384	0.1253	0.0386	0.1266	0.0376	0.1248	0.0357
0.1214	0.0358	0.1209	0.0357	0.1181	0.0358	0.1161	0.0349	0.1126	0.0345
0.1123	0.0339	0.1126	0.0338	0.1136	0.0330	0.1143	0.0323	0.1083	0.0324
0.1077	0.0318								

Table 4.10: The average estimates for entries of K for Figure 4.6(c) when $p = 30$.

0.0013	0.0002	0.0002	0.0003	0.0012	0.0004	0.0011	0.0004
0.0011	0.0003	0.0011	0.0003	0.0013	0.0004	0.0010	0.0004
0.0011	0.0003	0.0010	0.0003	0.0010	0.0003	0.0011	0.0003
0.0010	0.0003	0.0012	0.0003	0.0010	0.0003	0.0010	0.0003

Table 4.11: The batch standard errors for Figure 4.6(c) when $p = 30$.

5 Estimation of Normalizing Constants

The normalizing constant is an important quantity in model comparisons. Although we can not compute the normalizing constant exactly for any arbitrary colored graph, it would be worthwhile to investigate the estimation of normalizing constants in an efficient manner. This chapter introduces three methods for computing the normalizing constant of the colored G-Wishart distribution and illustrates their application in the analysis of Fret's heads data [Frets, 1921].

5.1 Monte Carlo Method

According to (2.9) in Theorem 2.1.3, the normalizing constant of the colored G-Wishart distribution can be written as

$$\begin{aligned}
I_{\mathcal{G}}(\delta, D) &= \int_{P_{\mathcal{G}}} 2^{|\mathcal{V}|} \prod_{i=1}^{p} Q_{ii}^{p-v_i^G-d_i^G+\delta-1} \Big(\prod_{i=1}^{p} \Psi_{ii}^{p-i-v_i^G+\delta-1} \Big) e^{-\frac{1}{2}\sum_{i=1}^{p}\sum_{j=i}^{p}\Psi_{ij}^2} d\Psi^{v(G)} \\
&= 2^{|\mathcal{V}|} \prod_{i=1}^{p} Q_{ii}^{p-v_i^G-d_i^G+\delta-1} \int_{P_{\mathcal{G}}} \Big(\prod_{i=1}^{p} \Psi_{ii}^{p-i-v_i^G+\delta-1} \Big) e^{-\frac{1}{2}\sum_{i=1}^{p}\sum_{j=i}^{p}\Psi_{ij}^2} d\Psi^{v(G)} \\
&= 2^{|\mathcal{V}|} \prod_{i=1}^{p} Q_{ii}^{p-v_i^G-d_i^G+\delta-1} \int_{P_{\mathcal{G}}} \Big(\prod_{(i,i)\in v(G)^c} \Psi_{ii}^{p-i-v_i^G+\delta-1} \Big) e^{-\frac{1}{2}\sum_{(i,j)\in v(G)^c}\Psi_{ij}^2} \\
&\quad \times \Big(\prod_{(i,i)\in v(G)} \Psi_{ii}^{p-i-v_i^G+\delta-1} \Big) e^{-\frac{1}{2}\sum_{(i,i)\in v(G)}\Psi_{ii}^2} e^{-\frac{1}{2}\sum_{(i,j)\in v(G),i\neq j}\Psi_{ij}^2} \\
&\quad \times \prod_{(i,i)\in v(G)} d\Psi_{ii} \prod_{(i,j)\in v(G),i\neq j} d\Psi_{ij}. \qquad (5.1)
\end{aligned}$$

Let $v_1(G) = \{(i,j); (i,j) \in v(G), i \neq j\}$. Then (5.1) becomes

$$\begin{aligned}
I_{\mathcal{G}}(\delta, D) &= 2^{|\mathcal{V}|}(2\pi)^{\frac{|v_1(G)|}{2}} \prod_{i=1}^{p} Q_{ii}^{p-v_i^G-d_i^G+\delta-1} \int_{P_{\mathcal{G}}} h(\Psi_{v(G)^c}) \\
&\quad \times \Big(\prod_{(i,i)\in v(G)} \Psi_{ii}^{p-i-v_i^G+\delta-1} \Big) e^{-\frac{1}{2}\sum_{(i,i)\in v(G)}\Psi_{ii}^2} \\
&\quad \times \prod_{(i,j)\in v(G),i\neq j} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\Psi_{ij}^2} \prod_{(i,i)\in v(G)} d\Psi_{ii} \prod_{(i,j)\in v(G),i\neq j} d\Psi_{ij} \qquad (5.2)
\end{aligned}$$

where $h(\Psi_{v(G)^c})$ is defined as in (2.10). Since $d\Psi_{ii} = \frac{1}{2}\Psi_{ii}^{-1} d(\Psi_{ii}^2)$, then (5.2) becomes

$$\begin{aligned}
I_{\mathcal{G}}(\delta, D) &= 2^{|\mathcal{V}|}(2\pi)^{\frac{|v_1(G)|}{2}} \prod_{i=1}^{p} Q_{ii}^{p-v_i^G-d_i^G+\delta-1} \int_{P_{\mathcal{G}}} h(\Psi_{v(G)^c}) \\
&\quad \times \Big(\prod_{(i,i)\in v(G)} \Psi_{ii}^{p-i-v_i^G+\delta-1} \frac{1}{2} \Psi_{ii}^{-1} \Big) e^{-\frac{1}{2} \sum\limits_{(i,i)\in v(G)} \Psi_{ii}^2} \\
&\quad \times \prod_{(i,j)\in v(G), i\neq j} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\Psi_{ij}^2} \prod_{(i,i)\in v(G)} d\Psi_{ii}^2 \prod_{(i,j)\in v(G), i\neq j} d\Psi_{ij} \\
&= (\frac{1}{2})^{|\mathcal{V}|} 2^{|\mathcal{V}|} (2\pi)^{\frac{|v_1(G)|}{2}} \prod_{i=1}^{p} Q_{ii}^{p-v_i^G-d_i^G+\delta-1} \int_{P_{\mathcal{G}}} h(\Psi_{v(G)^c}) \\
&\quad \times \Big(\prod_{(i,i)\in v(G)} (\Psi_{ii}^2)^{\frac{p-i-v_i^G+\delta}{2}-1} \Big) e^{-\frac{1}{2} \sum\limits_{(i,i)\in v(G)} \Psi_{ii}^2} \\
&\quad \times \prod_{(i,j)\in v(G), i\neq j} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\Psi_{ij}^2} \prod_{(i,i)\in v(G)} d\Psi_{ii}^2 \prod_{(i,j)\in v(G), i\neq j} d\Psi_{ij} \\
&= (2\pi)^{\frac{|v_1(G)|}{2}} \Big(\prod_{(i,i)\in v(G)} 2^{\frac{p-i-v_i^G+\delta}{2}} \Gamma(\frac{p-i-v_i^G+\delta}{2}) \Big) \Big(\prod_{i=1}^{p} Q_{ii}^{p-v_i^G-d_i^G+\delta-1} \Big) \\
&\quad \times \int_{P_{\mathcal{G}}} h(\Psi_{v(G)^c}) \prod_{(i,i)\in v(G)} \frac{1}{2^{\frac{p-i-v_i^G+\delta}{2}} \Gamma(\frac{p-i-v_i^G+\delta}{2})} (\Psi_{ii}^2)^{\frac{p-i-v_i^G+\delta}{2}-1} e^{-\frac{1}{2}\Psi_{ii}^2} \\
&\quad \times \prod_{(i,j)\in v(G), i\neq j} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\Psi_{ij}^2} \prod_{(i,i)\in v(G)} d\Psi_{ii}^2 \prod_{(i,j)\in v(G), i\neq j} d\Psi_{ij}. \qquad (5.3)
\end{aligned}$$

We are now ready to state our result about the normalizing constant for the colored G-Wishart.

Theorem 5.1.1 *Let $\mathcal{G}$ be an arbitrary colored graph and $I_{\mathcal{G}}(\delta, D)$ be the normalizing*

constant of the colored G-Wishart distribution. Then

$$\begin{aligned} I_{\mathcal{G}}(\delta, D) &= (2\pi)^{\frac{|v_1(G)|}{2}} \Big(\prod_{(i,i)\in v(G)} 2^{\frac{p-i-v_i^G+\delta}{2}} \Gamma(\frac{p-i-v_i^G+\delta}{2}) \Big) \\ &\quad \times \Big(\prod_{i=1}^{p} Q_{ii}^{p-v_i^G-d_i^G+\delta-1} \Big) E_{f(\Psi_{v(G)})}[h(\Psi_{v(G)^c})] \end{aligned} \tag{5.4}$$

where $h(\Psi_{v(G)^c})$ *is defined as in* (2.10) *and*

$$\begin{aligned} f(\Psi_{v(G)}) &= \prod_{(i,i)\in v(G)} \frac{1}{2^{\frac{p-i-v_i^G+\delta}{2}} \Gamma(\frac{p-i-v_i^G+\delta}{2})} (\Psi_{ii}^2)^{\frac{p-i-v_i^G+\delta}{2}-1} e^{-\frac{1}{2}\Psi_{ii}^2} \\ &\quad \times \prod_{(i,j)\in v(G), i\neq j} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\Psi_{ij}^2}. \end{aligned}$$

Proof. The proof follows immediately from expression (5.3) for $I_{\mathcal{G}}(\delta, D)$. ■

By the Law of Large Numbers, the Monte Carlo method evaluates the normalizing constant $I_{\mathcal{G}}(\delta, D)$ by the formula $\frac{1}{N}\sum_{l=1}^{N} h(\Psi^l_{v(G)^c})$, where N is a large integer, Ψ_{ii}, $(i,i) \in v(G)$, are the random samples from the independent chi-squared distribution with $p-i-v_i^G+\delta$ degrees of freedom, and Ψ_{ij}, $(i,j) \in v(G)$, $i \neq j$, are the random samples from the independent standard normal distribution.

5.2 Importance Sampling

According to the importance sampler, we are going to estimate the normalizing constant $I_{\mathcal{G}}(\delta, D)$ by $\frac{1}{N}\sum_{l=1}^{N} h(\Psi^l_{v(G)^c}) \frac{f(\Psi^l_{v(G)})}{f_*(\Psi^l_{v(G)})}$. Since Ψ_{ij}, $(i,j) \in v(G)^c$, are well defined

functions for all $(i,j) \in v(G)$, we can compute $h(\Psi^l_{v(G)^c})$, $l = 1, 2, \cdots, N$, after obtaining the samples Ψ_{ij}, $(i,j) \in v(G)$, from the density $f_*(\Psi_{v(G)})$.

5.3 Laplace Approximation

The normalizing constant can be written as

$$I_{\mathcal{G}}(\delta, D) = \int_{K \in P_{\mathcal{G}}} \exp\{P^*_{\delta,D}(K)\} \prod_{(i,j)\in v(G)} dK_{ij}$$

where $P^*_{\delta,D}(K) = \frac{\delta-2}{2} \log |K| - \frac{1}{2} tr(KD)$.

The Laplace approximation to $I_{\mathcal{G}}(\delta, D)$ is

$$\hat{I}_{\mathcal{G}}(\delta, D) = \exp\{P^*_{\delta,D}(\hat{K})\}(2\pi)^{\frac{|v(G)|}{2}} |H_{\delta,D}(\hat{K})|^{-1/2}$$

where $\hat{K} \in P_{\mathcal{G}}$ is the mode of the colored G-Wishart and $H_{\delta,D}(\hat{K})$ is the $|v(G)| \times |v(G)|$ Hessian matrix associated with $-P^*_{\delta,D}(K)$. For each colour class $u \in v(G)$, let A^u be the $p \times p$ diagonal matrix with entries $A^u_{ij} = 1$ if $(i,j) \in u$ and 0 otherwise. Therefore,

$$\frac{\partial^2(-P^*_{\delta,D}(K))}{\partial K_u \partial K_v} = \frac{\delta - 2}{2} tr(A^u \Sigma A^v \Sigma).$$

5.4 Simulations

We illustrate the performance of the Monte Carlo method, the importance sampling method and the Laplace approximation. Let $D = I_p$ be the identity matrix. To evaluate the performance of our Monte Carlo method and the importance sampling method, we take 15000 sample points. The simulation results are shown in Table 5.1 and Table 5.2 using the Monte Carlo method. The $\overline{NMSE}(\hat{I}_{\mathcal{G}}(\delta, D), I_{\mathcal{G}}(\delta, D))$ are presented in Table 5.1 for the five graphs in Fig. 3.1 for different values of δ with 100 replications. Table 5.2 reports $\overline{NMSE}(\hat{I}_{\mathcal{G}}(\delta, D), I_{\mathcal{G}}(\delta, D))$ for the star graphs with all the vertices in the same color for different p and different δ with 100 replications. Standard deviations are shown in parentheses. The performance of our algorithm is very bad for Fig. 3.1(c) and is not good for Fig. 3.1(e) when using the Monte Carlo method.

We use the importance sampling method for Fig. 3.1(c) and Fig. 3.1(e). In Fig. 3.1(c), we sample Ψ_{ii}^2, $(i,i) \in v(G)$, from the chi-square distribution with $k = 8, 10, 12$ degrees of freedom and sample Ψ_{ij}, $(i,j) \in v(G)$ and $i \neq j$, from the normal distribution with zero mean and standard deviations equal to $\sigma = 0.2, 0.5, 0.8$. In Fig. 3.1(e), we sample Ψ_{ii}^2, $(i,i) \in v(G)$, from the chi-square distribution with $k = 4, 5, 6, 7$ degrees of freedom and sample Ψ_{ij}, $(i,j) \in v(G)$ and $i \neq j$, from the normal distribution with zero mean and standard deviations equal to $\sigma = 0.5, 0.7, 0.9$. Table 5.3 and

Table 5.4 show $\overline{NMSE}(\hat{I}_{\mathcal{G}}(\delta, D), I_{\mathcal{G}}(\delta, D))$ with 30 replications and 100 replications for graphs in Fig. 3.1(c) and Fig. 3.1(e) for different k and σ, respectively. Table 5.3 show us the $\overline{NMSE}(\hat{I}_{\mathcal{G}}(\delta, D), I_{\mathcal{G}}(\delta, D))$ is very small when we choose $k = 8$ or 10, and $\sigma = 0.6$ for Fig. 3.1(c). Table 5.4 show us $\overline{NMSE}(\hat{I}_{\mathcal{G}}(\delta, D), I_{\mathcal{G}}(\delta, D))$ is very small when we choose $k = 5$ or 6, and $\sigma = 0.7$ for Fig. 3.1(e). Comparing to the Monte Carlo method, the algorithm of the importance sampling is highly efficient for Fig. 3.1(c) and Fig. 3.1(e).

At the end, we compare the performance of the Monte Carlo method and the Laplace approximation. Table 5.5 and Table 5.6 report $NMSE(\hat{I}_{\mathcal{G}}(\delta, D), I_{\mathcal{G}}(\delta, D))$ of normalizing constants in Fig. 3.1 for different δ when we use the Monte Carlo method and the Laplace approximation, respectively. The results indicate that the Monte Carlo method works very much better than the Laplace approximation for Figs. 3.1(a), (b) and (d). However, for Figs. 3.1(c) and (e), both of two methods don't give a good estimation.

5.5 Real Data Analysis

We use the Monte Carlo method and the Laplace approximation to compute the normalizing constant to perform a model search for Fret's heads dataset in Frets [1921]. The data consist of measurements in millimetres of the length and breadth

δ	Fig. 3.1(a)	Fig. 3.1(b)	Fig. 3.1(c)	Fig. 3.1(d)	Fig. 3.1(e)
1	2.506×10^{-4}	2.178×10^{-4}	587.721	4.703×10^{-5}	11.007
	(0.019)	(0.016)	(25.329)	(0.008)	(1.034)
3	1.018×10^{-4}	7.337×10^{-5}	35.671	2.420×10^{-5}	0.563
	(1.361×10^{-4})	(0.010)	(1.453)	(0.007)	(0.144)
5		6.813×10^{-5}	6.538	1.962×10^{-5}	0.120
		(0.009)	(0.548)	(0.006)	(0.084)
7		9.016×10^{-5}	1.780	1.662×10^{-7}	0.037
		(0.011)	(0.105)	(0.005)	(0.062)

Table 5.1: $\overline{NMSE}(\hat{I}_{\mathcal{G}}(\delta, D), I_{\mathcal{G}}(\delta, D))$ for graphs in Fig. 3.1 using the Monte Carlo method.

of the heads of 25 random pairs of first and second sons. We compare the 12 colored graphs shown in Fig. 5.1. Whittaker [1990] shows the model represented by the uncolored graph in Fig. 5.1(a) fits the data very well. The analyses in Højsgaard and Lauritzen [2008] support the models represented by graphs in Fig. 5.1(b) and Fig. 5.1(c) comparing the saturated model using the likelihood ratio test. The Edwards-Havránek model selection procedure in Gehrmann [2011] arrives at 9 minimally accepted models represented by the graphs in Figs. 5.1(d)-5.1(l). The model

δ	$p=2$	$p=3$	$p=4$	$p=5$
1	0.176	0.994	3.288	9.049
	(0.172)	(0.352)	(0.637)	(1.252)
3	0.017	0.111	0.381	1.001
	(0.037)	(0.075)	(0.121)	(0.194)
5	0.003	0.021	0.075	0.207
	(0.023)	(0.042)	(0.069)	(0.101)
7	5.349×10^{-4}	0.018	0.019	0.053
	(0.013)	(0.027)	(0.044)	(0.069)

Table 5.2: $\overline{NMSE}(\hat{I}_{\mathcal{G}}(\delta, D), I_{\mathcal{G}}(\delta, D))$ for the star graph with all the vertices in the same color using the Monte Carlo method.

presented in Fig. 5.1(l) gives the lowest BIC value among the 9 models.

We here use the Bayes factors to compare the 12 models presented in Fig. 5.1. In the prior colored G-Wishart distribution, the hyperparameters $\delta = 3, 10$ and $D = I_4$. In order to obtain the estimated normalizing constants using the Monte Carlo method, we take 15000 sample points. For the normalizing constant of the posterior distribution, we always use the Monte Carlo method to get the estimate. For the estimation of the normalizing constant of the prior distribution, we use two

k	$\sigma = 0.2$	$\sigma = 0.5$	$\sigma = 0.8$
8	4.528	0.003	0.237
	(3.764)	(0.063)	(0.082)
10	0.726	0.004	0.097
	(0.461)	(0.073)	(0.283)
12	0.932	0.011	0.205
	(0.868)	(0.107)	(0.438)

Table 5.3: $\overline{NMSE}(\hat{I}_{\mathcal{G}}(\delta, D), I_{\mathcal{G}}(\delta, D))$ for the graph in Fig. 3.1(c) using the importance sampling when $\delta = 3$ and $D = I_p$.

methods to compute them. One is the Monte Carlo method while the other one is the Laplace approximation. For both methods, the results of the model selection are the exactly same. Tables 5.7 and 5.8 show the estimation of the normalizing constant using the Monte Carlo method and the Laplace approximation for $\delta = 3$ and $\delta = 10$, respectively. Tables 5.9 and 5.10 report the marginal probability of the posterior distribution using the Monte Carlo method and the Laplace approximation, respectively. Both methods select the same most likely models for the same δ. The most three likely models are Fig. 5.1(k), Fig. 5.1(b) and Fig. 5.1(l) for $\delta = 3$, and for $\delta = 10$, the most three likely models are Fig. 5.1(l), Fig. 5.1(k) and Fig. 5.1(b).

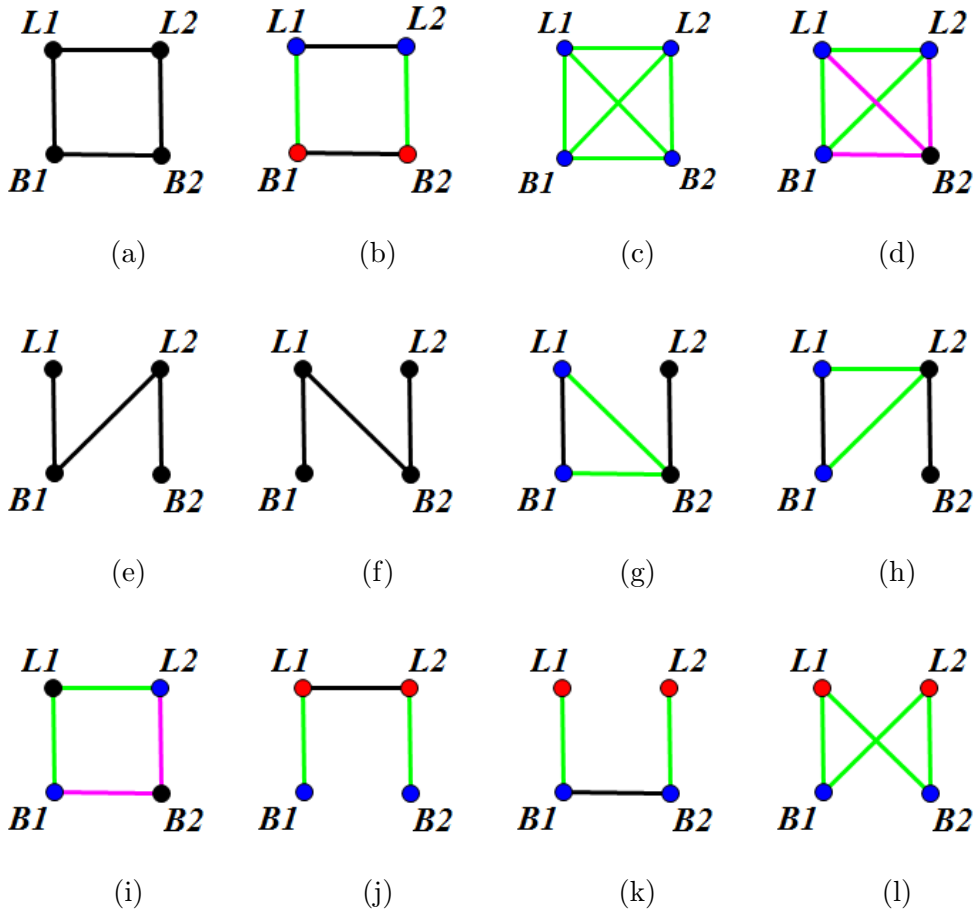


Figure 5.1: Possible colored graphs supported by Fret's heads data. L1: The head length of the eldest son; B1: The head breadth of the eldest son; L2: The head length of the second son; B2: The head breadth of the second son.

k	$\sigma = 0.5$	$\sigma = 0.7$	$\sigma = 0.9$
4	0.108	0.032	0.093
	(0.299)	(0.009)	(0.013)
5	0.097	0.016	0.044
	(0.494)	(0.014)	(0.013)
6	0.020	0.013	0.030
	(0.067)	(0.031)	(0.039)
7	0.219	0.056	0.028
	(1.796)	(0.290)	(0.049)

Table 5.4: $\overline{NMSE}(\hat{I}_{\mathcal{G}}(\delta, D), I_{\mathcal{G}}(\delta, D))$ for the graph in Fig. 3.1(e) using the importance sampling when $\delta = 3$ and $D = I_p$.

δ	Fig. 3.1(a)	Fig. 3.1(b)	Fig. 3.1(c)	Fig. 3.1(d)	Fig. 3.1(e)
3	0.521	0.760	50.426	0.566	0.620
5		0.307	36.465	0.192	0.219
7		0.156	15.247	0.093	0.108

Table 5.5: $NMSE(\hat{I}_{\mathcal{G}}(\delta, D), I_{\mathcal{G}}(\delta, D))$ for the graphs in Fig. 3.1 using the Laplace approximation.

δ	Fig. 3.1(a)	Fig. 3.1(b)	Fig. 3.1(c)	Fig. 3.1(d)	Fig. 3.1(e)
3	1.328×10^{-5}	3.197×10^{-5}	32.416	2.288×10^{-7}	0.568
5		1.048×10^{-4}	6.779	1.050×10^{-5}	0.119
7		1.115×10^{-4}	1.699	6.984×10^{-7}	0.038

Table 5.6: $NMSE(\hat{I}_{\mathcal{G}}(\delta, D), I_{\mathcal{G}}(\delta, D))$ for the graphs in Fig. 3.1 using the Monte Carlo sampling.

	Fig. 5.1(a)	Fig. 5.1(b)	Fig. 5.1(c)	Fig. 5.1(d)
posterior	1.191×10^{-157}	4.966×10^{-152}	9.758×10^{-295}	1.554×10^{-204}
prior (MC)	1.052×10^{4}	33.200	0.443	5.938
prior (Laplace)	843.704	9.469	0.245	2.056
	Fig. 5.1(e)	Fig. 5.1(f)	Fig. 5.1(g)	Fig. 5.1(h)
posterior	6.997×10^{-173}	1.035×10^{-158}	6.981×10^{-216}	4.622×10^{-206}
prior (MC)	2.526×10^{3}	2.530×10^{3}	243.525	242.603
prior (Laplace)	336.589	336.589	47.475	47.475
	Fig. 5.1(i)	Fig. 5.1(j)	Fig. 5.1(k)	Fig. 5.1(l)
posterior	$< 10^{-300}$	$< 10^{-300}$	9.594×10^{-151}	9.285×10^{-156}
prior (MC)	38.623	9.179	8.315	2.282
prior (Laplace)	13.392	3.777	3.777	1.065

Table 5.7: The normalizing constants for the graphs in Fig. 5.1 for $\delta = 3$.

	Fig. 5.1(a)	Fig. 5.1(b)	Fig. 5.1(c)	Fig. 5.1(d)
posterior	5.014×10^{-187}	2.042×10^{-180}	$< 10^{-300}$	2.941×10^{-277}
prior (MC)	1.248×10^{15}	4.665×10^{11}	5.475×10^{8}	3.794×10^{10}
prior (Laplace)	8.088×10^{14}	4.012×10^{11}	4.596×10^{8}	3.080×10^{10}
	Fig. 5.1(e)	Fig. 5.1(f)	Fig. 5.1(g)	Fig. 5.1(h)
posterior	$< 10^{-300}$	3.785×10^{-191}	3.239×10^{-285}	6.423×10^{-280}
prior (MC)	1.608×10^{14}	1.605×10^{14}	7.334×10^{12}	7.389×10^{12}
prior (Laplace)	1.141×10^{14}	1.141×10^{14}	5.689×10^{12}	5.689×10^{12}
	Fig. 5.1(i)	Fig. 5.1(j)	Fig. 5.1(k)	Fig. 5.1(l)
posterior	$< 10^{-300}$	$< 10^{-300}$	2.421×10^{-179}	3.697×10^{-180}
prior (MC)	6.424×10^{11}	6.361×10^{10}	6.297×10^{10}	6.446×10^{9}
prior (Laplace)	5.674×10^{11}	5.659×10^{10}	5.659×10^{10}	5.644×10^{9}

Table 5.8: The normalizing constants for the graphs in Fig. 5.1 for $\delta = 10$.

δ	Fig. 5.1(a)	Fig. 5.1(b)	Fig. 5.1(c)	Fig. 5.1(d)
3	1.133×10^{-161}	1.496×10^{-153}	2.202×10^{-294}	2.618×10^{-205}
10	4.016×10^{-202}	4.378×10^{-192}	$< 10^{-300}$	7.753×10^{-288}
	Fig. 5.1(e)	Fig. 5.1(f)	Fig. 5.1(g)	Fig. 5.1(h)
3	2.770×10^{-176}	4.092×10^{-162}	2.867×10^{-218}	1.905×10^{-208}
10	$< 10^{-300}$	2.358×10^{-205}	4.417×10^{-298}	8.693×10^{-293}
	Fig. 5.1(i)	Fig. 5.1(j)	Fig. 5.1(k)	Fig. 5.1(l)
3	$< 10^{-300}$	$< 10^{-300}$	1.154×10^{-151}	4.068×10^{-156}
10	$< 10^{-300}$	$< 10^{-300}$	3.845×10^{-190}	5.736×10^{-190}

Table 5.9: The marginal probability of the posterior distribution for the graphs in Fig. 5.1 using the Monte Carlo method.

δ	Fig. 5.1(a)	Fig. 5.1(b)	Fig. 5.1(c)	Fig. 5.1(d)
3	1.412×10^{-160}	5.244×10^{-153}	3.975×10^{-294}	7.559×10^{-205}
10	6.200×10^{-202}	5.091×10^{-192}	$< 10^{-300}$	9.549×10^{-288}
	Fig. 5.1(e)	Fig. 5.1(f)	Fig. 5.1(g)	Fig. 5.1(h)
3	2.079×10^{-175}	3.076×10^{-161}	1.471×10^{-217}	9.736×10^{-208}
10	$< 10^{-300}$	3.318×10^{-205}	5.695×10^{-298}	1.129×10^{-292}
	Fig. 5.1(i)	Fig. 5.1(j)	Fig. 5.1(k)	Fig. 5.1(l)
3	$< 10^{-300}$	$< 10^{-300}$	2.540×10^{-151}	8.713×10^{-156}
10	$< 10^{-300}$	$< 10^{-300}$	4.279×10^{-190}	6.551×10^{-190}

Table 5.10: The marginal probability of the posterior distribution for the graphs in Fig. 5.1 using the Laplace approximation.

6 Precision Estimation in High-dimensional Models

6.1 Bayesian Estimation in Large Dimensions

Recently, covariance estimation in graphical Gaussian models in high dimensional settings with relatively small sample sizes has attracted more and more attentions. However, traditional estimation methods often rely on the expensive inference of global models and can not be implemented efficiently. To conquer these challenges, a general framework for distributed algorithms, in which both data and estimation are distributed across the vertices of the underlying graph, is developed based on combining local and inexpensive estimators. The distributed algorithm is usually directly applicable to practically sized problem because of its attractive properties, including low computational cost and low communication cost across the local models.

6.1.1 Distributed Algorithms in Graphical Gaussian Models

Meng et al. [2014] considered the distributed estimation of the precision matrix K in graphical Gaussian models based on the maximum likelihood estimation. The distributed algorithms in Meng et al. [2014] determined the global estimator of K by splitting the global estimator into low-dimensional local estimators. Their approach estimated the local parameters by maximizing the marginal likelihood corresponding to each vertex and its neighborhood in the underlying graph.

Now we briefly introduce their method here. Define the set of immediate neighbours of the given vertex $i \in V$ as $ne(i) = \{j|(i,j) \in E\}$. Then the one-hop neighbourhood and two-hop neighbourhood can be defined as $N_i = \{i\} \cup ne(i)$ and $N_i = \{i\} \cup ne(i) \cup \{k \mid (k,j) \in E, j \in ne(i)\}$, respectively. The local marginal model for $X^i = \{X_v, v \in N_i\}$ is defined as the graphical Gaussian model with the precision matrix denoted by $\mathcal{K}^i$, which can be evaluated through

$$\mathcal{K}^i = (\Sigma_{N_i,N_i})^{-1} = K_{N_i,N_i} - K_{N_i,V\backslash N_i}[K_{V\backslash N_i,V\backslash N_i}]^{-1}K_{V\backslash N_i,N_i}. \tag{6.1}$$

Denote $B_i = \{j|j \in N_i,\ ne(j) \cap (V\backslash N_i) \neq \emptyset\}$ as the buffer set and $\mathcal{P}_i = N_i\backslash B_i$ as the protected set, which are illustrated in Figure 6.1(b). The Markov property of a

random vector X with respect to G implies $K_{\mathcal{P}_i, V\setminus N_i} = 0$. Therefore, we have that

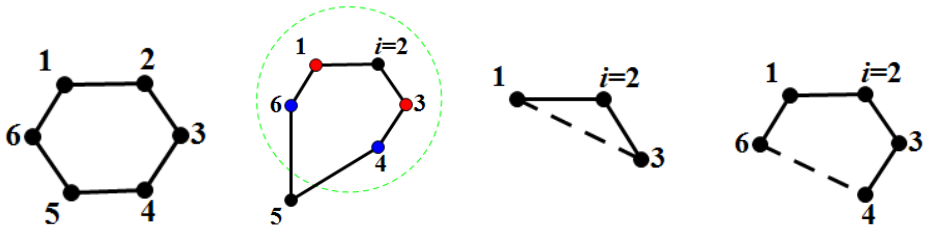
$$\begin{pmatrix} \mathcal{K}^i_{\mathcal{P}_i,\mathcal{P}_i} & \mathcal{K}^i_{\mathcal{P}_i,B_i} \\ \mathcal{K}^i_{B_i,\mathcal{P}_i} & \mathcal{K}^i_{B_i,B_i} \end{pmatrix} = \begin{pmatrix} K_{\mathcal{P}_i,\mathcal{P}_i} & K_{\mathcal{P}_i,B_i} \\ K_{B_i,\mathcal{P}_i} & K_{B_i,B_i} \end{pmatrix}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

This shows that the local parameters of $\mathcal{K}^i$ indexed by $(\mathcal{P}_i, \mathcal{P}_i)$ and $(\mathcal{P}_i, B_i)$ are totally preserved and equal to the corresponding global parameters. However, this claim does not hold for the parameters indexed by (B_i, B_i).

Based on these observations, Meng et al. [2014] defined relaxed local graphs as follows. Denote a local graph corresponding to the given vertex i by $G_i = (N_i, E_i)$ where

$$E_i = E \cap \{\{\mathcal{P}_i \times \mathcal{P}_i\} \cup \{\mathcal{P}_i \times B_i\} \cup \{B_i \times \mathcal{P}_i\}\} \cup \{B_i \times B_i\}.$$

In the local graph G_i, the zero constraints of the edges in $E_i\setminus\{B_i, B_i\}$ are the same as the corresponding constraints in the global graph G. The edges in $B_i \times B_i$ are arbitrary without any constraint. An illustration of these local graphs is given in Figure 6.1(c). Each relaxed local model corresponding to the vertex i is a graphical Gaussian model with respect to the local graph G_i. In each relaxed local model, the MLE $\hat{\mathcal{K}}^i$ will be used to estimate $\mathcal{K}^i$. Then the estimate of local parameters can be obtained by extracting the elements in $\hat{\mathcal{K}}^i$ corresponding to the vertex i and edges adjacent to vertex i. Meng et al. [2014] also proved that the proposed distributed



(a) (b) (c)

Figure 6.1: (a) The underlying graph. (b) The colors of the buffer set and the protected set are blue and red, respectively. The two-hop neighbourhood for vertex i is indicated with dashed contours. (c) The graphs representing the one-hop relaxed model (left) and the two-hop relaxed model (right). Dotted lines denote edges not existing in the original underlying graph.

estimate of K in graphical Gaussian models is asymptotically consistent when the number of variables p is fixed and the sample size n grows to infinity. Furthermore, the convergence rate to the true parameters was derived when both p and n go to infinity.

6.1.2 Bayesian Estimation and Large Deviation

In a Bayesian framework, we consider the distributed estimation of the precision matrix K in large colored graphical Gaussian models. Since the colored G-Wishart distribution is a member of exponential families, we can use the strategy proposed in

Ghosal [2000] to study the asymptotic properties of our distributed estimate. In fact, Ghosal [2000] considered the consistency of the posterior mean of the parameters in an exponential family when both p and n go to infinity. Suppose the p-dimensional independent random samples $X_1, X_2, \ldots, X_n$ are from the exponential family with the density

$$f(x;\theta) = \exp\{x^t\theta - \psi(\theta)\}.$$

Denote $\mu = \psi'(\theta_0)$ and $F = \psi''(\theta_0)$ where θ_0 is the true value of the parameter θ. Let J be a square root of F. Moreover, let $u = \sqrt{n}J(\theta - \theta_0)$ and $\Delta_n = \sqrt{n}J^{-1}(\bar{x} - \mu)$ where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Under some conditions, Ghosal [2000] proved that

$$\int ||u|| \cdot |\pi^*(u) - \phi(u; \Delta_n, I_p)|du \xrightarrow{p} 0 \tag{6.2}$$

where $\pi^*(u)$ is the posterior distribution and $\phi(u; \Delta_n, I_p)$ is the density of multivariate normal $N(\Delta_n, I_p)$. It follows that the posterior mean of the parameter is asymptotically normal and asymptotically efficient when both p and n increase to infinity. The proof can be achieved by expressing the left-hand side of (6.2) as a sum of three terms and locating the corresponding upper bounds for the three terms separately. In each term, the only random component is $||\Delta_n||$ and $||\Delta_n|| = O_p(\sqrt{p})$ by Chebyshev's inequality.

We will use the above technique in our proofs for the local relaxed models. Nevertheless, every vertex in the global graph corresponds to a local relaxed model. If we

allow the number of vertices increases to infinity, there is an infinite number of local models. In order to use Bonferroni inequality to bound the overall error between the distributed estimate and the true value of the parameter θ in probability, we need to know the exact tail probability of $P(||\Delta_n|| > cp)$, where c is a constant. This requires us to establish a new large deviation result for $||\Delta_n||$. Fortunately, Gao and Carroll [2015] proposed the cumulant boundedness conditions for the large deviation results based on the quadratic forms [Spokoiny and Zhilova, 2013]. We next briefly introduce their idea. Let ξ be a random vector and B be a matrix. Then $||B\xi||$ follows a general quadratic form and the deviation probability for quadratic forms $||B\xi||$ was established by Spokoiny and Zhilova [2013] under the following exponential moment condition

$$\log E[\exp\{\gamma^t \xi\}] \leq ||\gamma||^2/2, \qquad \gamma \in R^p, \quad ||\gamma|| \leq g$$

where g is a positive constant.

Gao and Carroll [2015] introduced a cumulant boundedness condition and proved that the exponential moment condition in Spokoiny and Zhilova [2013] can be satisfied asymptotically under the cumulant boundedness condition.

Definition 6.1.1 *[Cumulant Boundedness Condition, Gao and Carroll, 2015] Let C_1, C_2 and C_3 be constants. For a random vector Z of dimension m, let $g(\gamma)$ denote its cumulant generation function, where γ denotes an m dimensional vector. Assume*

the first two derivatives of its cumulant generating function satisfy $|\frac{\partial g(\gamma)}{\partial \gamma_j}|_{\gamma=0}| \leq C_1$ *and* $|\frac{\partial^2 g(\gamma)}{\partial \gamma_j \partial \gamma_k}|_{\gamma=0}| \leq C_2$. *Assume further that there exists a constant* δ *such that with* $||\gamma|| \leq \delta$, *the absolute value of all the third derivatives of its cumulant generating function satisfy* $|\frac{\partial^3 g(\gamma)}{\partial \gamma_j \partial \gamma_k \partial \gamma_l}| \leq C_3$ *for all* $1 \leq j, k, l \leq m$.

Based on the cumulant boundedness condition, Gao and Carroll [2015] proved the following asymptotical exponential moment condition, which provides a toolbox for calculating the deviation probability bound of $||\Delta_n||$ using the large deviation results in Spokoiny and Zhilova [2013].

Theorem 6.1.1 *[Gao and Carroll, 2015] Let* $Z_i, i = 1, 2, \ldots, n$, *be independently distributed random vectors of dimension* m *with zero mean and identity covariance matrices and* $\eta = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i$. *If each random vector* Z_i *satisfies the cumulant boundedness condition with the same bounds and* $p^2 \log p = o(n)$, *then* $\log E[\exp\{\gamma^t \eta\}] \leq a^2 ||\gamma||^2/2$ *for* $||\gamma|| < (p^2 \log p)^{1/2}$ *with some constant* $a^2 > 1$ *when* n *is sufficiently large.*

This result implies that if the cumulant boundedness condition holds, we will be able to apply large deviation results in Spokoiny and Zhilova [2013] to the proof of the asymptotic properties of the distributed estimate in the colored graphical Gaussian models.

6.2 Distributed Estimation in Large Colored Graphical Models

Distributed estimation methods have recently been given for computing the maximum likelihood estimate of the precision matrix K in large graphical Gaussian models. Our aim, in this chapter, is to give a distributed Bayesian estimate of the precision matrix K in large colored graphical Gaussian models. In each local relaxed model, we take the sample posterior mean of the precision matrix as our estimate of the precision matrix. The simulation results show that the distributed Bayesian estimate performs very well when the number of variable p is large.

6.2.1 Local Relaxed Marginal Models

For a given vertex $i \in V$, define the set of immediate neighbors of vertex i as $ne(i) = \{j|(i,j) \in E\}$. For each vertex $i \in V$, we consider two types of neighbourhoods of the vertex i, the so-called one-hop and two-hop neighbourhood. The one-hop neighbourhood $N_i = \{i\} \cup ne(i)$ is made up of the vertex i and the vertices directly connected to it. The two-hop neighbourhood $N_i = \{i\} \cup ne(i) \cup \{k \mid (k,j) \in E, j \in ne(i)\}$ consists of the vertex i, its neighbours and the neighbours of the neighbours. Without the risk of confusion, we let N_i be either a one-hop or two-hop neighbourhood. We consider the local marginal model for $X_{N_i} = \{X_v, v \in N_i\}$

which is abbreviated as X^i. This is a graphical Gaussian model with the precision matrix denoted by $\mathcal{K}^i$. Then we have that

$$\mathcal{K}^i = (\Sigma_{N_i,N_i})^{-1} = K_{N_i,N_i} - K_{N_i,V\backslash N_i}[K_{V\backslash N_i,V\backslash N_i}]^{-1}K_{V\backslash N_i,N_i}. \tag{6.3}$$

Based on the collection of vertices N_i and its complement set $V\backslash N_i$, we partition the set N_i further into two subsets. One is the buffer set denoted as $B_i = \{j | j \in N_i$ and $ne(j) \cap (V\backslash N_i) \neq \emptyset\}$, which are the vertices having edges connecting to the complement of N_i in V. The other is the protected set denoted as $\mathcal{P}_i = N_i\backslash B_i$, which contains the vertices in N_i that are not directly connected to $V\backslash N_i$. Since the distribution of X is Markov with respect to G, then we have that $X_{\mathcal{P}_i} \perp X_{V\backslash N_i}|X_{B_i}$ and it follows that

$$K_{\mathcal{P}_i,V\backslash N_i} = 0. \tag{6.4}$$

Then equation (6.3) becomes

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

where the 0's in the matrix above follows from the identity (6.4). Therefore, we obtain the following relationships

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

$$\mathcal{K}^i_{B_i,B_i} = K_{B_i,B_i} - K_{B_i,V\setminus N_i}(K_{V\setminus N_i,V\setminus N_i})^{-1}K_{V\setminus N_i,B_i}.$$

This shows that the local parameters of $\mathcal{K}^i$ indexed by $(\mathcal{P}_i,\mathcal{P}_i)$ and $(\mathcal{P}_i,B_i)$ are equal to the corresponding global parameters but the same does not hold for the parameters indexed by (B_i,B_i). This important observation above motivates us to use the N_i-marginal local models to estimate those parameters which are identical in both local and global models.

We denote by $\mathcal{G}_i$ the colored graph with vertex set N_i and edge set

$$E_i = E \cap \{\{\mathcal{P}_i \times \mathcal{P}_i\} \cup \{\mathcal{P}_i \times B_i\} \cup \{B_i \times \mathcal{P}_i\}\} \cup \{B_i \times B_i\}.$$

In $\mathcal{G}_i$, the colors of the vertices in $N_i \setminus B_i$ are the same as the corresponding colors in $\mathcal{G}$. The colors of the edges in $E_i\setminus\{B_i,B_i\}$ are the same as the corresponding colors in $\mathcal{G}$. The colors of the vertices in B_i and the edges in $B_i \times B_i$ are arbitrary without any constraint. Let K^i be the precision matrix of this relaxed local marginal model. We thus keep the important relationships

$$\mathcal{K}^i_{\mathcal{P}_i,\mathcal{P}_i} = K_{\mathcal{P}_i,\mathcal{P}_i}, \;\; \mathcal{K}^i_{\mathcal{P}_i,B_i} = K_{\mathcal{P}_i,B_i}$$

and have a local graphical Gaussian model with the canonical parameter K^i on which we can put a local colored G-Wishart distribution as the prior. In each local model Markov with respect to $\mathcal{G}_i$, $i \in \{1, 2, \ldots, p\}$, we use the method developed in Chapter 2 to obtain the Bayesian estimator $\tilde{K}^i$, the sample posterior mean of K^i with the colored G-Wishart prior.

Next, we will show how to construct a distributed Bayesian estimate based on the local models. Let $\theta = (\theta_{V_1}, \theta_{V_2}, \ldots, \theta_{V_T}, \theta_{E_1}, \theta_{E_2}, \ldots, \theta_{E_S})^t$ denote the global parameter, that is the "free" entries of K which represent the vertex class or the edge class, and let θ_0 be its true value. In each local model $\mathcal{G}_i$, we define the local parameter as $\theta^i = (\theta^i_1, \theta^i_2, \ldots, \theta^i_{S_i})^t$, the vector of free entries of K^i, and the corresponding local estimator as $\tilde{\theta}^i$. The true value of θ^i is denoted by θ^i_0. Furthermore, we collapse all the local parameters into one vector

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

and its true value is denoted as $\bar{\theta}_0$. After obtaining the local estimators, a distributed estimate of $\tilde{\theta}$ can be constructed as

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

and

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

where $G_k = \{i | \exists h \in N_i, (i,h) \in E_k\}$. Define the global distributed Bayesian estimate

$$\tilde{\theta} = g(\bar{\theta}) = (g_{V_1}(\bar{\theta}), g_{V_2}(\bar{\theta}), \ldots, g_{V_T}(\bar{\theta}), g_{E_1}(\bar{\theta}), g_{E_2}(\bar{\theta}), \ldots, g_{E_S}(\bar{\theta}))^t.$$

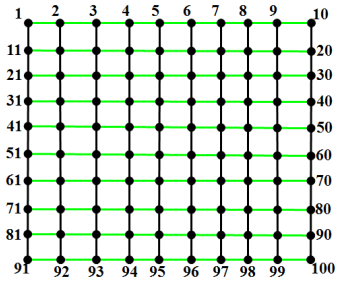
6.2.2 Simulations

In order to illustrate the performance of our proposed distributed Bayesian estimate of K, we conduct a number of experiments using simulated data. For each simulation, we compute the distributed estimator using relaxed local marginal models built on the "one-hop" and on the "two-hop" neighbourhoods of each $i \in \{1, 2, \ldots, p\}$. We choose the colored G-Wishart distribution as the prior with hyperparameters $D^i = I_{p_i}$ and $\delta^i = 3$ for all $i \in \{1, 2, \ldots, p\}$. The corresponding estimators are called the MBE-1hop and MBE-2hop estimates of K respectively. We consider seven different colored graphical Gaussian models. The underlying graph of three of those models are colored cycles of length $p = 20$ with alternate vertex and edge colors as indicated in Figure 4.6 (a), (b) and (c). Three other models have the same type of underlying colored graphs but the colored cycles are of length $p = 30$. The underlying graph of the seventh model is a 10×10 grid with colors as shown in Figure 6.2. For both the colored cycles and the 10×10 grid, black edges or vertices indicate that there is no color constraint on edges or vertices. For the sake of comparisons, for models represented by the colored cycles of order 20 or 30, we also compute the global

Bayesian estimate of the precision matrix K, denoted by GBE, using the method given in Chapter 2. Since asymptotically, the posterior mean of K is expected to be close to the maximum likelihood estimate of K, for all colored models, we also compute the global MLE of K, denoted by GMLE.

The values of $(K_{ij})_{1\leq i,j\leq p}$ used for the simulation for models represented by colored graphs as given in Figure 4.6 (a), (b) and (c) are given in Table 6.1. For the 10×10 colored grid-graph of Figure 6.2, we chose $K_{i+10(j-1),i+1+10(j-1)} = 1$ 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 of K obtained from the MH algorithm are based on 5000 iterations after the first 1000 burn-in iterations.

Table 6.2 shows the normalized mean square error $NMSE(\hat{K}, K) = \frac{||\hat{K}-K||^2}{||K||^2}$ for the six models with the colored cycles as underlying graphs. Values are averaged over 100 data sets from the multivariate normal $N(0, K^{-1})$ distribution. We repeat the simulations 100 times. Standard deviations are shown in parentheses. From these results, we see that our MBE-1hop and MBE-2hop estimates perform very well compared to the global estimate GBE. In Figure 6.3 we give the graphs of $NMSE(\hat{K}, K)$ in the function of the sample size, for different sample sizes ranging from 50 to 100 for the four models with underlying graphs the colored cycles of



(a)

Figure 6.2: The 10×10 colored grid graph. Black vertices or edges indicate different arbitrary colors.

length $p = 20$ and the 10×10 grid. We see that the MLE and the GBE consistently yield the smallest and largest $NMSE(\hat{K}, K)$ respectively with the MBE-1hop and MBE-2hop, and the $NMSE(\hat{K}, K)$ of MBE-2hop is always smaller than that of the MBE-1hop. As expected, as n increases, all $NMSE(\hat{K}, K)$ tend to the same value.

Computations are performed on a 2 core 4 thread processor with i5-4200U, 2.3 GHZ chips and 8 GB of RAM, running on Windows 8. The computing times for the estimates of K are given in minutes in Table 6.3 for the six models with cycles as underlying graphs. We can see that the computation times for the MBE-1hop and MBE 2-hop are much smaller than for the GBE.

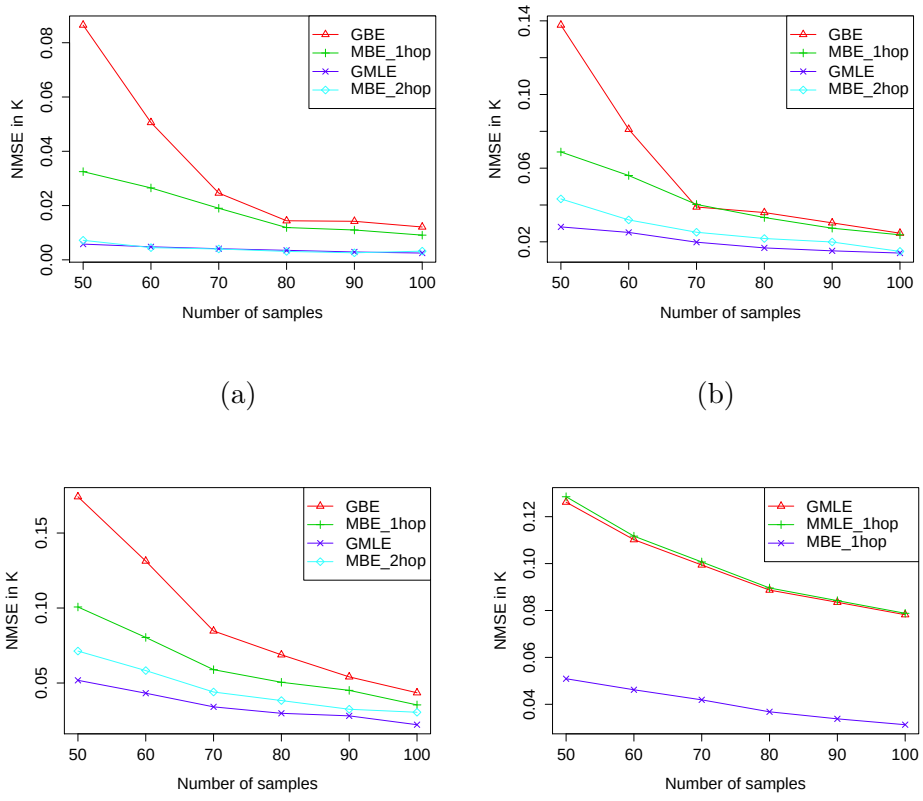


Figure 6.3: NMSE in K for different colored graphical models. (a) NMSE for the colored graph in Figure 4.6 (a) when $p = 20$. (b) NMSE for the colored graph in Figure 4.6 (b) when $p = 20$. (c) NMSE for the colored graph in Figure 4.6 (c) when $p = 20$. (d) NMSE for the colored lattice graph in Figure 6.2 when $p = 100$.

parameters	Figure 4.6 (a)	Figure 4.6 (b)	Figure 4.6 (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 6.1: The parameters chosen for the matrix K for producing Figure 4.6.

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

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

7 Asymptotic Analysis of Distributed Bayesian Estimation

We study the asymptotic behaviour of the distributed estimate proposed in Chapter 6 under the regular asymptotic regime when the number of variables p is fixed and under the double asymptotic regimes when both p and the sample size n are large.

When the number of variables p is fixed, we obtain the limiting distribution of the estimator, which can be used to approximate the density of the estimator in a large samples. The approximate distribution is necessary for statistical inference: confidence interval and hypothesis tests. We can see that the asymptotic variance of the estimator decreases as the sample size n increases. That means the estimator can estimate the true value very well when the sample size is large enough. This also implies the consistency of the distributed estimator.

When the number of variables p grows with the sample size n, we compute the

upper bound of the distance between the estimator and the true value. Under certain growth restrictions on p, the upper bound shows the accuracy of the estimator. Condition (4) below implies that p grows much slower than n when the number of variables in the local models goes to infinity. However, in practice, the local graphs are very sparse and of low dimension. Therefore, condition (4) can be relaxed to condition (4^*) to obtain a smaller upper bound.

We also show in particular, that when the number of parameters in the local models is uniformly bounded, the convergence rate we obtain for the asymptotic consistency, in the Frobenius norm, of our estimate of the precision matrix compares well with the convergence rate in previous literature for the maximum likelihood estimate.

7.1 Bayesian Estimator When p Is Fixed and $n \to \infty$

In each local model corresponding to the vertex $i \in \{1, 2, \ldots, p\}$, let $L^i(\theta^i)$ and $l^i(\theta^i)$ denote the likelihood and log likelihood, respectively. The Fisher information is denoted by $I^i(\theta^i) = E_{\theta^i}[\frac{\partial}{\partial \theta^i} l^i(\theta^i|X^i)[\frac{\partial}{\partial \theta^i} l^i(\theta^i|X^i)]^t]$. Define a S_i-dimensional vector $U_{ij} = \frac{1}{\sqrt{n}}[I^i(\theta_0^i)]^{-1} \frac{\partial l^i(\theta^i|X_j^i)}{\partial \theta^i}\big|_{\theta^i=\theta_0^i}$ 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 $\bar{G} = nCov(U_1)$. For each r, $r = 1, 2, \ldots, S_i$, let δ_r^i be the $S_i \times S_i$ indicator matrix with $(\delta_r^i)_{hl} = 1$ if $K_{hl}^i = \theta_r^i$ and 0 otherwise. The

following theorem shows that the global estimator has the property of asymptotic normality when the number of variables p is fixed and the sample size n goes to infinity.

Theorem 7.1.1 *Let θ_0, $\tilde{\theta}$, $\bar{\theta}$ and $\bar{G}$ be defined above. Then*

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

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 $i \in \{1, 2, \ldots, p\}$, 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 $T^i = \theta_0^i + \frac{1}{n}[I^i(\theta_0^i)]^{-1} \frac{\partial l^i(\theta^i)}{\partial \theta^i}\big|_{\theta^i = \theta_0^i}$. It then follows from Theorem 8.3 in Lehmann and Casella [1998] that $\sqrt{n}(\tilde{\theta}^i - T^i) \xrightarrow{p} 0$. Furthermore, we have

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

with $E[U_j] = 0$ for $j = 1, 2, \cdots, n$. Next, we compute the covariance matrix $Cov(U_1)$ with (i, k) entry

$$Cov(U_{i1}, U_{k1}) = \frac{1}{n}[I^i(\theta_0^i)]^{-1} E\big[\frac{\partial l^i(\theta^i | X_1^i)}{\partial \theta^i}\big|_{\theta^i = \theta_0^i} (\frac{\partial l^k(\theta^k | X_1^k)}{\partial \theta^k}\big|_{\theta^k = \theta_0^k})^t\big] [I^k(\theta_0^k)]^{-1}. \quad (7.1)$$

Based on the definition of the indicator matrix δ_r^i,

$$\theta_0^i = (\frac{1}{|\tau_1^i|} tr(\delta_1^i K_0^i), \frac{1}{|\tau_2^i|} tr(\delta_2^i K_0^i), \cdots, \frac{1}{|\tau_{S_i}^i|} tr(\delta_{S_i}^i K_0^i))^t,$$

where τ_r^i is the numbers of elements belonging to the r-th color class in K_0^i. Since X^i has a multivariate normal distribution $N(0, (K_0^i)^{-1})$, we have

$$\frac{\partial l^i(\theta^i | X_1^i)}{\partial \theta_j^i}\big|_{\theta^i = \theta_0^i} = \frac{1}{2} tr(\delta_j^i (K_0^i)^{-1}) - \frac{1}{2} tr(\delta_j^i X_1^i (X_1^i)^t).$$

Therefore, the (q, m) entry of $E\big[\frac{\partial l^i(\theta^i|X_1^i)}{\partial \theta^i}\big|_{\theta^i=\theta_0^i}(\frac{\partial l^k(\theta^k|X_1^k)}{\partial \theta^k}\big|_{\theta^k=\theta_0^k})^t\big]$ in (7.1) is

$$\begin{aligned} & E\Big[\frac{\partial l^i(\theta^i|X_1^i)}{\partial \theta_q^i}\big|_{\theta^i=\theta_0^i}\frac{\partial l^k(\theta^k|X_1^k)}{\partial \theta_m^k}\big|_{\theta^k=\theta_0^k}\Big] \\ = \; & \frac{1}{4}tr(\delta_q^i\Sigma_0^i)\times tr(\delta_m^k\Sigma_0^k) - \frac{1}{4}tr(\delta_q^i\Sigma_0^i)\times tr(\delta_m^k E[X_1^k(X_1^k)^t]) \\ & -\frac{1}{4}tr(\delta_m^k\Sigma_0^k)\times tr(\delta_q^i E[X_1^i(X_1^i)^t]) + \frac{1}{4}E[tr(\delta_q^i X_1^i(X_1^i)^t)\times tr(\delta_m^k X_1^k(X_1^k)^t)] \end{aligned}$$

where $\Sigma_0^i = (K_0^i)^{-1}$ and $\Sigma_0^k = (K_0^k)^{-1}$. According to Isserlis' Theorem, we have that

$$E[X_{1a}X_{1b}X_{1c}X_{1d}] = (\Sigma_0)_{ab}(\Sigma_0)_{cd} + (\Sigma_0)_{ac}(\Sigma_0)_{bd} + (\Sigma_0)_{ad}(\Sigma_0)_{bc}.$$

Therefore, each entry of $nCov(U_1)$ is well-defined. By Multivariate Central Limit Theorem, we have $\sqrt{n}(\bar{\theta} - \bar{\theta}_0) \xrightarrow{\mathcal{L}} N(0, \bar{G})$ as $n \to \infty$, where $\bar{G} = nCov(U_1)$. As $\tilde{\theta} = g(\bar{\theta})$, based on Delta method, we have that $\sqrt{n}(\tilde{\theta} - \theta_0) \xrightarrow{\mathcal{L}} N(0, A)$ where $A = \frac{\partial g(\bar{\theta})}{\partial \bar{\theta}^t}\bar{G}(\frac{\partial g(\bar{\theta})}{\partial \bar{\theta}^t})^t$. ■

All proofs of lemmas and propositions used in the proofs of theorems of this chapter are given in the Appendix A. We now establish an asymptotic result similar to Theorem 7.1.1 but with the MLE replacing the posterior mean of K. Based on the same local models, we compute the local MLE $\hat{\theta}^i$ of θ^i from the local models and obtain a distributed MLE for the global model, which is denoted by $\hat{\theta}$.

Theorem 7.1.2 *Let $\hat{\theta}$ be the distributed MLE. Then*

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{\mathcal{L}} N(0, A) \quad as \quad n \to 0$$

where A is defined as in Theorem 7.1.1 above.

Proof. For any $i \in \{1, 2, \ldots, p\}$, we use the well known result for MLE as follows

$$\sqrt{n}(\hat{\theta}^i - \theta_0^i) = \frac{1}{\sqrt{n}}[I^i(\theta_0^i)]^{-1} \sum_{j=1}^{n} \frac{\partial l^i(\theta^i | X_j^i)}{\partial \theta^i}\Big|_{\theta^i = \theta_0^i} + R^i \tag{7.2}$$

where $R^i \xrightarrow{p} 0$ as $n \to \infty$. Comparing identity (7.2) with (7.1) in Theorem 7.1.1, the result of Theorem 7.1.2 follows. ■

The distributed MLE is calculated by the method of Meng et al. [2014] using the local relaxed marginal models defined above. We thus see that the distributed Bayesian estimator $\tilde{\theta}$ we proposed has the same limiting distribution as the distributed MLE $\hat{\theta}$.

7.2 Bayesian Estimator When $p \to \infty$ 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 $p \times p$ matrix A, let $||A||_F$ be its Frobenius norm defined by $||A||_F = (\sum_{j=1}^{p} \sum_{k=1}^{p} |a_{jk}|^2)^{\frac{1}{2}}$. In the local model Markov with respect to $\mathcal{G}_i$ as defined in Section 6.2.1 above, we write the density of X_j^i, $j = 1, 2, \ldots, n$, as

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

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

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

where $I^i_{\mathcal{G}_i}(\delta^i, D^i)$ is the normalizing constant. In order to obtain our asymptotic results, we will follow an argument similar to that of Ghosal [2000] which gives the asymptotic distribution of the posterior mean of K when both the dimension p of the model and the sample size n grow to ∞. Ghosal [2000] considers a random variable X with the density belonging to the natural exponential family

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

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

$$Y^i_j = -\frac{1}{2}(tr(\delta^i_1X^i_j(X^i_j)^t), tr(\delta^i_2X^i_j(X^i_j)^t), \ldots, tr(\delta^i_{S_i}X^i_j(X^i_j)^t))^t, \tag{7.3}$$

where $\delta^i_1, \delta^i_2, \cdots, \delta^i_{S_i}$ are indicator matrices for each color class and $j = 1, 2, \ldots, n$. The distribution of Y^i_j is as follows

$$f(Y^i_j; K^i) \propto \exp\big[-\frac{1}{2}tr(K^iX^i_j(X^i_j)^t) + \frac{1}{2}\log|K^i|\big] = \exp\big[(Y^i_j)^t\theta^i - \psi(\theta^i)\big]$$

where $\psi(\theta^i) = -\frac{1}{2}\log|K^i|$ is the cumulant generating function. From standard properties of natural exponential families, we have that

$$\mu^i = \psi'(\theta_0^i) \qquad \text{and} \qquad F^i = \psi''(\theta_0^i) \tag{7.4}$$

are the mean vector and the covariance matrix of Y_j^i, $j = 1, 2, \ldots, n$, respectively. Let J^i be a square root of F^i, i.e. $J^i(J^i)^t = F^i$. Let

$$V_j^i = (J^i)^{-1}(Y_j^i - E_{\theta^i}(Y_j^i)) \tag{7.5}$$

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 also

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

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

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

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

$$\Delta_n^i = \sqrt{n}(J^i)^{-1}(\bar{Y}^i - \mu^i). \tag{7.6}$$

The following four conditions will be assumed.

(1) The orders of $\log p$ and $\log n$ are the same, i.e. $\frac{\log p}{\log n} \to \zeta > 0$ as $n \to \infty$.

(2) There exists two constants κ_1 and κ_2 such that $0 < \kappa_1 \leq \lambda_{min}(K_0) < \lambda_{max}(K_0) \leq \kappa_2 < \infty$.

(3) For any $i \in \{1, 2, \ldots, p\}$, the numbers τ^i of the entries K_{jk}^i in the same color class is bounded.

(4) As $p \to 0$, the sample size satisfies the rate $\frac{p^{13}(\log p)^2}{n^{\frac{1}{2}}} \to 0$.

Remark 7.2.1 *Condition (2) implies* $0 < \frac{1}{\kappa_2} \leq \lambda_{min}(\Sigma_0) < \lambda_{max}(\Sigma_0) \leq \frac{1}{\kappa_1} < \infty$. *By the interlacing property of eigenvalues, we have that* $0 < \frac{1}{\kappa_2} \leq \lambda_{min}((\Sigma_0)_{N_i,N_i}) < \lambda_{max}((\Sigma_0)_{N_i,N_i}) \leq \frac{1}{\kappa_1} < \infty$ *where* N_i *is defined as in Section 6.2.1. Therefore,* $0 < \kappa_1 \leq \lambda_{min}((\Sigma_0)_{N_i,N_i})^{-1} < \lambda_{max}((\Sigma_0)_{N_i,N_i})^{-1} \leq \kappa_2 < \infty$. *By the definition* (6.3), *for any* $i \in \{1, 2, \ldots, p\}$, *we have* $0 < \kappa_1 \leq \lambda_{min}(K_0^i) < \lambda_{max}(K_0^i) \leq \kappa_2 < \infty$.

Our aim in this section is to prove that under Conditions (1)-(4) when both p and n are large, the distributed estimator $\tilde{\theta}$ tends to θ_0 in Frobenius norm with probability tending to 1. Ghosal [2000] considered the consistency of the posterior mean of K

for the exponential family. The convergence rate depends on three expressions which added together yield an upper bound of the overall error $||\tilde{\theta}-\theta_0||$. In each expression, the only random component is $||\Delta_n^i||$ and $||\Delta_n^i|| = O_p(p)$. However, we have an infinite number of local models. In order to use Bonferroni inequality to bound the overall error probability, we need to know the exact tail probability of $P(||\Delta_n^i|| > cp)$, where c is a constant. This leads us to establish a new large deviation result for $||\Delta_n^i||$ in Lemma A.1.3. We now state the asymptotic consistency of our proposed estimator in Theorem 7.2.1.

Theorem 7.2.1 *Under Conditions (1)-(4), there exists a constant c^* such that*

$$||\tilde{\theta} - \theta_0|| \leq c^* \frac{p^{\frac{3}{2}}}{\sqrt{n}}$$

with probability greater than $1 - 10.4 \exp\{-\frac{1}{6}p^2 \log p + \log p\}$.

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 $||\sqrt{n}J^i(\tilde{\theta}^i - \theta_0^i)||^2 = n(\tilde{\theta}^i - \theta_0^i)^t (J^i)^t J^i (\tilde{\theta}^i - \theta_0^i) \geq n\lambda_{min}(F^i)||\tilde{\theta}^i - \theta_0^i||^2$, we obtain

$$\begin{aligned}
||\tilde{\theta}^i - \theta_0^i||^2 &\leq \frac{1}{n\lambda_{min}(F^i)} ||\sqrt{n}J^i(\tilde{\theta}^i - \theta_0^i)||^2 \\
&= \frac{1}{n\lambda_{min}(F^i)} ||\Delta_n^i + \int u^i [\pi_*^i(u^i) - \phi(u^i; \Delta_n^i, I_{S_i})] du^i||^2 \text{ by Lemma A.1.7} \\
&\leq \frac{1}{n\lambda_{min}(F^i)} \Big(||\Delta_n^i||^2 + || \int u^i [\pi_*^i(u^i) - \phi(u^i; \Delta_n^i, I_{S_i})] du^i||^2 \Big) \qquad (7.7)
\end{aligned}$$

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 (7.7), we will find out its upper bound. Denote $u^i = (u_1^i, u_2^i, \cdots, u_{S_i}^i)^t$. 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

$$\int u_j^i[\pi_*^i(u^i) - \phi(u^i; \Delta_n^i, I_{S_i})]du^i \quad \leq \quad \int ||u^i||[\pi_*^i(u^i) - \phi(u^i; \Delta_n^i, I_{S_i})]du^i. \quad (7.8)$$

Let $\tilde{Z}_n^i(u^i) = \exp[(u^i)^t\Delta_n^i - \frac{1}{2}||u^i||^2]$ and $M(p) = p^2 \log p$. 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 (7.8) can be bounded by a sum of three integrals as follows.

$$\begin{aligned}
& \int ||u^i|| \times |\pi_*^i(u^i) - \phi(u^i; \Delta_n^i, I_{S_i})|du^i \\
\leq \quad & \frac{\int_{||u^i||^2 \leq cM(p)} ||u^i|| \cdot |\pi^i(\theta_0^i + n^{-\frac{1}{2}}(J^i)^{-1}u^i)Z_n^i(u^i) - \pi^i(\theta_0^i)\tilde{Z}_n^i(u^i)|du^i}{\int \pi^i(\theta_0^i)\tilde{Z}_n^i(u^i)du^i} \quad (7.9) \\
& + \frac{\int_{||u^i||^2 > cM(p)} ||u^i|| \cdot Z_n^i(u^i)\pi^i(\theta_0 + n^{-\frac{1}{2}}(J^i)^{-1}u^i)du^i}{\int \pi^i(\theta_0^i)\tilde{Z}_n^i(u^i)du^i} \quad (7.10) \\
& + \int_{||u^i||^2 > cM(p)} ||u^i||\phi(u^i; \Delta_n^i, I_{S_i})du^i, \quad (7.11)
\end{aligned}$$

where c is defined in Lemma A.1.5. By Lemmas A.1.4, A.1.5 and A.1.6, $\int ||u^i|| \times |\pi_*^i(u^i) - \phi(u^i; \Delta_n^i, I_{S_i})| du^i$ can be bounded by

$$A(p,n,c) = c_5(c)\frac{p^{13}\log p}{\sqrt{n}} + \exp[-c_9(c)p^2 \log p] + \frac{2}{\sqrt{2\pi}}p^{-4a^2+4} + \sqrt{3a^2}\frac{2}{\sqrt{2\pi}}p^{-4a^2+3}$$

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

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

with probability greater than $1-10.4\exp\{-\frac{1}{6}p^2\}$. 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 (7.7) and Lemma A.1.3, we get

$$||\tilde{\theta}^i - \theta_0^i||^2 \quad \leq \quad \frac{1}{\lambda_{min}(F^i)}\Big(\frac{3a^2p^2}{n} + \frac{S_i}{n}A(p,n,c)\Big)$$

with 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 distributed estimator $\tilde{\theta}$ in terms of $||\tilde{\theta}^i - \theta_0^i||$ from the local model. By Proposition A.1.1, for any $i \in \{1, 2, \ldots, p\}$, $\lambda_{min}(F^i) \geq \frac{1}{\kappa_2^2}$.

Therefore, we have

$$\begin{aligned} ||\tilde{\theta} - \theta_0|| &\leq ||\bar{\theta} - \bar{\theta}_0|| \leq (\sum_{i=1}^{p} ||\tilde{\theta}^i - \theta_0^i||^2)^{\frac{1}{2}} \qquad \text{by triangle inequality} \\ &\leq \Big\{ \sum_{i=1}^{p} \Big[\frac{1}{\lambda_{min}(F^i)} \big(\frac{3a^2p^2}{n} + \frac{S_i}{n} A(p,n,c) \big) \Big] \Big\}^{\frac{1}{2}} \\ &\leq \Big\{ \kappa_2^2 \big[\frac{3a^2p^3}{n} + \frac{p^2(p+1)}{2n} A(p,n,c) \big] \Big\}^{\frac{1}{2}} \end{aligned}$$

with probability greater than $1 - 10.4p \exp\{-\frac{1}{6}p^2\}$ by the Bonferroni inequality. Furthermore, Condition (4) implies $A(p,n,c) \to 0$. Therefore, there exists a constant c^* such that

$$\begin{aligned} ||\tilde{\theta} - \theta_0|| &\leq \Big\{ \kappa_2^2 \big[\frac{3a^2p^3}{n} + \frac{p^2(p+1)}{2n} o(1) \big] \Big\}^{\frac{1}{2}} \\ &\leq c^* \frac{p^{\frac{3}{2}}}{\sqrt{n}} \end{aligned}$$

with probability greater than $1 - 10.4 \exp\{-\frac{1}{6}p^2 + \log p\} \to 1$. ■

7.3 Bayesian Estimator When Dimensions of Local Models Bounded

We investigated in Section 7.2 the asymptotic behaviour of $\tilde{\theta}$ when S_i is unbounded under the double asymptotic regimes. In this Section, we assume that the dimension S_i in the local models is bounded and we will see that for $\tilde{\theta}$ to be close

to θ_0, n must grow as a power of $\log p$ rather than as a power of p. Furthermore, we assume the following conditions:

(4*) The orders of $\log p$ and $\log n$ satisfy that $\frac{\log p}{\log n} \to \zeta$, $0 < \zeta < 1$, as $n \to \infty$.

(5) The number of parameters in each local model is bounded by a constant S^*, i.e. $S_i \leq S^*$, $i \in \{1, 2, \ldots, p\}$.

The main result of this section is Theorem 7.3.1 stated below.

Theorem 7.3.1 *Under Conditions (2), (4*) and (5), there exists a constant c_1^* such that*

$$||\tilde{\theta} - \theta_0|| \leq c_1^* \frac{p^{\frac{1}{2}} \log p}{\sqrt{n}}$$

with probability greater than $1 - 10.4 \exp\{-\frac{1}{6}\log^2 p + \log p\}$.

For the convenience, we now point out the main difference between the proofs of Theorems 7.2.1 and 7.3.1.

(a) Under Conditions (2), (4*) and (5), for any $i \in \{1, 2, \ldots, p\}$, the quantities $\log |F^i|$ in Proposition A.1.3, $B_{1n}^i(c)$ and $B_{2n}^i(c)$, $i = 1, 2, \ldots, p$, in Proposition A.1.6 are all uniformly bounded because the number of parameters in each local model is uniformly bounded and the eigenvalues of K^i are uniformly bounded from above and below.

(b) The equivalent of Lemma A.1.2 under our new boundedness condition is Lemma A.1.8 where $||\gamma^i|| < p$ is replaced by the condition $||\gamma^i|| < \log p$.

(c) The equivalent of Lemma A.1.3 is Lemma A.1.9, the large deviation result is established for $||\Delta_n^i||^2 > 3a^2(\log p)^2$ rather than $||\Delta_n^i||^2 > 3a^2p^2$.

(d) When S_i is unbounded, in Theorem 2.1 of Ghosal [2000], the fact that $||\Delta_n^i||^2 > 3a^2p^2$ with probability $1-\varepsilon$ implies $n||\hat{\theta}^i - \theta_0^i||^2 > bp^2$ with the same probability $1-\varepsilon$, where $\hat{\theta}^i$ is the MLE and b is a constant. For S_i bounded, using the new large deviation result $||\Delta_n^i||^2 > 3a^2(\log p)^2$ in (c) above, we have the new result of $n||\hat{\theta}^i - \theta_0^i||^2 > b'(\log p)^2$ with probability greater than $1 - 10.4\exp\{-\frac{1}{6}(\log p)^2\}$, where b' is a constant (See Lemma A.1.10).

(e) As a consequence of our choice $||\gamma^i|| < \log p$ in (b) above, the threshold $M(p) = p^2 \log p$ in the proof of Theorem 7.2.1 can be replaced by $M(p) = (\log p)^2(\log\log p)$.

Proof. The proof follows the same line as that of Theorem 7.2.1. Our aim is to find the upper bound for the three terms (7.9), (7.10) and (7.11).

1. A bound for (7.9): Under Condition (5), the Lipschitz continuity in Proposition A.1.5 becomes $|\log \pi^i(\theta^i) - \log \pi^i(\theta_0^i)| \leq M_1 S^* ||\theta^i - \theta_0^i||$ when $||\theta^i - \theta_0^i|| \leq \sqrt{||(F^i)^{-1}||cM(p)/n}$. We choose $M(p) = \log^2 p(\log\log p)$, then $\varphi_n^i(c) = O(\frac{\log p(\log\log p)^{\frac{1}{2}}}{\sqrt{n}})$ and $f^i(||\Delta_n^i||, c) = O(\frac{\log^3 p(\log\log p)^{\frac{1}{2}}}{\sqrt{n}})$ in Lemma A.1.4. Therefore, following the same proof of Lemma A.1.4, we have that there exists a constant $c_5'(c)$ such that $R_1(||\Delta_n^i||, c) \leq c_5'(c)\frac{\log^4 p \log\log p}{\sqrt{n}}$ with probability greater than $1 - 10.4\exp\{-\frac{1}{6}\log^2 p\}$. Condition (4^*) implies $\frac{\log^4 p \log\log p}{\sqrt{n}} \to 0$ as $p \to \infty$.

2. A bound for (7.10): According to Lemma A.1.10 and following the same proof of Lemma 2.2 of Ghosal [2000], with probability greater than $1-10.4\exp\{-\frac{1}{6}\log^2 p\}$, we have $Z_n^i(u^i) \leq \exp[-\frac{1}{4}c\log^2 p(\log\log p)]$ on $||u^i||^2 > cM(p)$. Following the same proof as that of Lemma A.1.5, there exists a constant c and a constant $c_9'(c)$ such that

$$R_2(||\Delta_n^i||, c) \leq \exp[-c_9'(c)M(p)]$$

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

3. A bound for (7.11): According to Lemma A.1.6, for $M(p) = \log^2 p(\log\log p)$, we have

$$\int_{||u^i||^2>cM(p)} ||u^i||\phi(u^i; \Delta_n^i, I_{S_i})du^i \leq \frac{2S_i^2}{\sqrt{2\pi}}e^{-\frac{4a^2M(p)}{S_i}} + \frac{2\sqrt{3a^2}S^i\log p}{\sqrt{2\pi}}e^{-\frac{4a^2M(p)}{S_i}}$$

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

Combining the above results, we have

$$A(p,n,c) = c_5'(c)\frac{\log^4 p\log\log p}{\sqrt{n}} + e^{-c_9'(c)M(p)} + \frac{2S_i^2}{\sqrt{2\pi}}e^{-\frac{4a^2M(p)}{S_i}} + \frac{2\sqrt{3a^2}S^i\log p}{\sqrt{2\pi}}e^{-\frac{4a^2M(p)}{S_i}}$$

with probability greater than $1 - 10.4\exp\{-\frac{1}{6}\log^2 p\}$. It follows

$$\begin{aligned} ||\tilde{\theta} - \theta_0|| &\leq \Big\{\kappa_2^2\big[\frac{p3a^2\log^2 p}{n} + \frac{S_i}{n}A(p,n,c\big]\Big\}^{\frac{1}{2}} \\ &\leq c_1^*\frac{\sqrt{p}\log p}{\sqrt{n}} \end{aligned}$$

with probability greater than $1 - 10.4p\exp\{-\frac{1}{6}\log^2 p\}$ by the Bonferroni inequality. The condition (4*) implies $\frac{\sqrt{p}\log p}{\sqrt{n}} \to 0$ as $p \to \infty$. This completes the proof. ■

Remark 7.3.1 *We note that the error bound* $\frac{p^{\frac{1}{2}} \log p}{\sqrt{n}}$ *in Theorem 7.3.1 is smaller than that* $\frac{p^{\frac{3}{2}}}{\sqrt{n}}$ *in Theorem 7.2.1. Also the rate of growth for the sample size is in terms of powers of* $\log p$ *rather than* p *as in Section 7.2.*

Remark 7.3.2 *As in Meng et al. [2014], we assume that the graph structure is known. When* $S_i < S^*$*, the error bound in our case is of the order* $\frac{p \log^2 p}{n}$ *which compares well with the order* $\frac{p \log p}{n}$ *in Meng et al. [2014]. The sample size requirement* $\frac{\log^4 p \log \log p}{\sqrt{n}} \to 0$ *is slightly more demanding than Meng's condition of* $n > \log p$*.*

8 Conclusions and Future Work

In this chapter, we summarize the results in this thesis and also point out some problems for future research.

8.1 Conclusions

In this thesis, we investigated the sampling methods of the colored G-Wishart distribution and the precision matrix estimation under a Bayesian framework.

We proposed a conjugate prior for the colored graphical Gaussian models. Such prior is called the colored G-Wishart distribution. One of the major contributions of the thesis is the proposal of a sampling method from the colored G-Wishart distribution, based on the Metropolis-Hastings algorithm. The proposed sampling method makes it possible to obtain the estimation of the posterior mean of K which is used for the Bayesian estimation of the precision matrix.

In order to illustrate the validity of the proposed Metropolis-Hastings algorithm,

we also investigated the issue of computing the expected value of the precision matrix on the colored G-Wishart distribution for some particular colored graphs. These colored graphs are the colored tree, two types of colored stars, one colored complete graph with 3 vertices and one non-decomposable graph with 4 vertices. For all of these colored graphs, we compared the expected values of the precision matrix K with the sample mean of K obtained from the Metropolis-Hastings algorithm. Through a number of numerical experiments, we found that the Bayesian estimator for the precision matrix provides a good estimate of the true value.

We further developed a fast algorithm for estimating the precision matrix in large colored graphical Gaussian models. The parallel algorithm is suitable for high-dimensional applications and computationally efficient. The distributed algorithm split the high-dimensional global model into p different low-dimensional local models. Each vertex corresponds to one local model based on its neighborhood. In each local model, the estimate of the local precision matrix was obtained by the Metropolis-Hastings algorithm algorithm we proposed. The global estimate was defined as the average of corresponding local estimates in order to satisfy the symmetric property of the precision matrix. We also derived asymptotic properties for the proposed distributed Bayesian estimator, as well as the convergence rate when the dimension of the model is large. Results from simulation studies have shown the accuracy and

efficiency of the proposed estimators in various settings.

Finally, in this thesis, we proposed three methods for estimating the normalizing constants of the colored G-Wishart distribution for any arbitrary colored graph: the Monte Carlo method, the importance sampling and the Laplace approximation. Moreover, we applied these methods to the study of heredity of head dimensions [Frets, 1921]. In this real data analysis, we compute the marginal probability of the data for each colored graph G. The marginal probability is the ratio of the normalizing constant of the posterior distribution and the normalizing constant for the prior distribution. Both the Monte Carlo method and the Laplace approximation were used to estimate the normalizing constant of the colored G-Wishart distribution for different choices of δ.

8.2 Future Work

Our proposed sampling method in Chapter 2 involves the matrix completion step for every update of the Cholesky components of the precision matrix. However, the matrix completion step is conducted iteratively and time-consuming. For the model with increasingly large dimensions, the computational requirement of the matrix completion step becomes increasingly burdensome. An interesting future direction would be developing a new sampling method from the colored G-Wishart distribu-

tion which can reduce the computational complexity. In my thesis, we assume the structure of the colored graphical models is known. Based on this assumption, we developed the Bayesian estimator of the precision matrix. However, we sometimes do not have enough statistical information for the structure of the underlying graphs in practice. Therefore, another topic worth investigating is the model selection. The graphical model selection problem can be reduced to the problem of estimating the zero-pattern and equal-pattern of the precision matrix. The existing method for the simultaneous clustering and feature selection in a regression would be adapted to the graphical model selection.

Bibliography

H. H. Andersen, M. Hojbjerre, D. Sorensen, and P. S. Eriksen. *Linear and graphical models: for the multivariate complex normal distribution*. Springer, 1995.

S. Andersson. Invariant normal models. *The Annals of Statistics*, 3(1):132–154, 1975.

S. A. Andersson, H. K. Brons, and S. T. Jensen. Distribution of eigenvalues in multivariate statistical analysis. *The Annals of Statistics*, 11(2):392–415, 1983.

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

C. M. Carvalho, H. Massam, and M. West. Simulation of hyper-inverse Wishart distributions in graphical models. *Biometrika*, 94(3):647–659, 2007.

D. R. Cox and N. Wermuth. *Multivariate dependencies: Models, analysis and interpretation*. Chapman and Hall, London, 1996.

A. P. Dawid and S. L. Lauritzen. Hyper Markov laws in the statistical analysis of decomposable graphical models. *The Annals of Statistics*, 21(3):1272–1317, 1993.

A. P. Dempster. Covariance selection. *Biometrics*, 28(1):157–175, 1972.

P. Diaconis and D. Ylvisaker. Conjugate priors for exponential families. *The Annals of statistics*, 7(2):269–281, 1979.

A. Dobra, A. Lenkoski, and A. Rodriguez. Bayesian inference for general Gaussian graphical models with application to multivariate lattice data. *Journal of the American Statistical Association*, 106(496):1418–1433, 2011.

J. Faraut and A. Korányi. *Analysis on symmetric cones*. Oxford Mathematical Monographs, The Clarendon Press, Oxford University Press, New York, 1994.

G. P. Frets. Heredity of headform in man. *Genetica*, 3(3):193–400, 1921.

X. Gao and R. J. Carroll. Data integration with high dimensionality. *Manuscript submitted for publication*, 2015.

H. Gehrmann. Lattices of graphical gaussian models with symmetries. *Symmetry*, 3 (3):653–679, 2011.

S. Ghosal. Asymptotic normality of posterior distributions for exponential families

when the number of parameters tends to infinity. *Journal of Multivariate Analysis*, 74(1):49–68, 2000.

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

B. Hylleberg, M.-B. Jensen, and E. Ørnbøl. *Graphical symmetry models*. Aalborg University, Institute for Electronic Systems, Department of Mathematics and Computer Science, 1993.

S. T. Jensen. Covariance hypotheses which are linear in both the covariance and the inverse covariance. *The Annals of Statistics*, 16(1):302–322, 1988.

S. L. Lauritzen. *Graphical models*. Oxford University Press, 1996.

E. L. Lehmann and G. Casella. *Theory of point estimation*. Springer, Berlin, 1998.

A. Lenkoski. A direct sampler for G-Wishart variates. *Stat*, 2(1):119–128, 2013.

G. Letac and H. Massam. Wishart distributions for decomposable graphs. *The Annals of Statistics*, 35(3):1278–1323, 2007.

J. Madsen. Invariant normal models with recursive graphical Markov structure. *Annals of statistics*, 28(4):1150–1178, 2000.

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

N. Mitsakakis, H. Massam, and M. D. Escobar. A Metropolis-Hastings based method for sampling from the G-Wishart distribution in Gaussian graphical models. *Electronic Journal of Statistics*, 5:18–30, 2011.

R. J. Muirhead. *Aspects of multivariate statistical theory*. John Wiley & Sons, 1982.

I. Olkin. Testing and estimation for a circular stationary model. *The Annals of Mathematical Statistics*, 40(4):1358–1373, 1969.

J. M. Ortega and W. C. Rheinboldt. *Iterative solution of nonlinear equations in several variables*. Academic press, Boston, 1970.

V. I. Paulsen, S. C. Power, and R. R. Smith. Schur products and matrix completions. *Journal of functional analysis*, 85(1):151–178, 1989.

M. Piccioni. Independence structure of natural conjugate densities to exponential families and the Gibbs' sampler. *Scandinavian journal of statistics*, 27(1):111–127, 2000.

S. Portnoy. Asymptotic behavior of likelihood methods for exponential families when

the number of parameters tends to infinity. *The Annals of Statistics*, 16(1):356–366, 1988.

A. Roverato. Hyper inverse Wishart distribution for non-decomposable graphs and its application to Bayesian inference for Gaussian graphical models. *Scandinavian Journal of Statistics*, 29(3):391–411, 2002.

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

C. Uhler, A. Lenkoski, and D. Richards. Exact formulas for the normalizing constants of Wishart distributions for graphical models. *arXiv preprint arXiv:1406.4901*, 2014.

H. Wang and C. M. Carvalho. Simulation of hyper-inverse Wishart distributions for non-decomposable graphs. *Electronic Journal of Statistics*, 4:1470–1475, 2010.

H. Wang and S. Z. Li. Efficient Gaussian graphical model determination under G-wishart prior distributions. *Electronic Journal of Statistics*, 6:168–198, 2012.

G. N. Watson. *A treatise on the theory of Bessel functions*. Cambridge university press, 1995.

J. Whittaker. *Graphical models in applied multivariate statistics*. Wiley Publishing, 1990.

S. S. Wilks. Sample criteria for testing equality of means, equality of variances, and equality of covariances in a normal multivariate distribution. *The Annals of Mathematical Statistics*, 17(3):257–281, 1946.

A Appendix

A.1 Proofs Used in Chapter 7

Here we provide the lemmas and their proofs used in Chapter 7. We let

$$\bar{U}_j^i = (J^i)^{-1}(Y_j^i - \mu^i) \tag{A.1}$$

for $i = 1, 2, \ldots, p$, and $j = 1, 2, \ldots, n$. We now want to show the large deviation result for Δ_n^i. To do so, we need to show that the cumulant boundedness condition is satisfied by $\bar{U}_j^i$ (Lemma A.1.1). This will allow us to show that Δ_n^i satisfy the exponential moment condition (Lemma A.1.2). In Lemma A.1.3, we obtain the large deviation result for Δ_n^i.

Lemma A.1.1 *For any $i \in \{1, 2, \ldots, p\}$, there exist constants η and C_2 such that under Conditions (2) and (3), for $||\gamma^i|| \leq \eta$ and for all $1 \leq k, l, m \leq S_i$, the absolute value of all the third derivatives of the cumulant generating function $G^i_{\bar{U}_j^i}(\gamma^i)$ of $\bar{U}_j^i$*

satisfy

$$\Big|\frac{\partial^3 G^i_{\bar{U}^i_j}(\gamma^i)}{\partial\gamma^i_k\partial\gamma^i_l\partial\gamma^i_m}\Big| \le C_2, \quad j = 1, 2 \ldots, n.$$

Proof. Let Y^i_j be defined in (7.3) and $G^i_{Y^i_j}(\gamma^i) = \log E(e^{(\gamma^i)^t Y^i_j})$ be the cumulant generating function of Y^i_j. Let γ^i be a S_i-dimensional vector, by Theorem 3.2.3 in Muirhead [1982], the moment generating function of Y^i_j is

$$M^i(\gamma^i) = E\{\exp[(\gamma^i)^t Y^i_j]\} = |I_{p_i} + T^i(\gamma^i)\Sigma^i_0|^{-\frac{1}{2}}$$

where $T^i(\gamma^i)$ is a $p_i \times p_i$ matrix with $T^i_{\alpha\beta} = \gamma^i_k$ if $K^i_{\alpha\beta} = \theta^i_k$. Therefore, the cumulant generating function $G^i_{Y^i_j}(\gamma^i)$ of Y^i_j is given by

$$G^i_{Y^i_j}(\gamma^i) = \log M^i(\gamma^i) = -\frac{1}{2}\log|I_{p_i} + T^i(\gamma^i)\Sigma^i_0|.$$

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

$$\begin{aligned}
\frac{\partial G^i_{Y^i_j}(\gamma^i)}{\partial\gamma^i_k} &= -\frac{1}{2}tr\Big([I_{p_i} + T^i(\gamma^i)\Sigma^i_0]^{-1}(\delta^i_k\Sigma^i_0)\Big), \\
\frac{\partial^2 G^i_{Y^i_j}(\gamma^i)}{\partial\gamma^i_k\partial\gamma^i_l} &= \frac{1}{2}tr\Big(\delta^i_k\Sigma^i_0[I_{p_i} + T^i(\gamma^i)\Sigma^i_0]^{-1}(\delta^i_l\Sigma^i_0)[I_{p_i} + T^i(\gamma^i)\Sigma^i_0]^{-1}\Big) \text{ and} \\
\frac{\partial^3 G^i_{Y^i_j}(\gamma^i)}{\partial\gamma^i_k\partial\gamma^i_l\partial\gamma^i_m} &= -\frac{1}{2}tr\Big(\delta^i_k\Sigma^i_0[I_{p_i} + T^i(\gamma^i)\Sigma^i_0]^{-1}(\delta^i_m\Sigma^i_0)(I_{p_i} + T^i(\gamma^i)\Sigma^i_0)^{-1}(\delta^i_l\Sigma^i_0) \\
&\quad \times[I_{p_i} + T^i(\gamma^i)\Sigma^i_0]^{-1} + \delta^i_k\Sigma^i_0[I_{p_i} + T^i(\gamma^i)\Sigma^i_0]^{-1}(\delta^i_l\Sigma^i_0)[I_{p_i} + T^i(\gamma^i)\Sigma^i_0]^{-1} \\
&\quad \times(\delta^i_m\Sigma^i_0)[I_{p_i} + T^i(\gamma^i)\Sigma^i_0]^{-1}\Big),
\end{aligned}$$

respectively. First, Condition (2) implies $\lambda_{max}(\Sigma_0^i) \leq \frac{1}{\kappa_1}$. By Proposition A.1.2, the absolute value of each element of Σ_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 $1 - \eta\frac{1}{\kappa_1} \leq \lambda_j(I_{p_i} + T^i(\gamma^i)\Sigma_0^i) \leq 1 + \eta\frac{1}{\kappa_1}$. Moreover, according to Lemma A.1.13, $I_{p_i} + T^i(\gamma^i)\Sigma_0^i$ is a positive definite. Therefore, by Proposition A.1.2 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 $i \in \{1, 2, \ldots, p\}$, there exists a constant C_1 such that $|\frac{\partial^3 G^i_{Y_j^i}(\gamma^i)}{\partial\gamma_k^i\partial\gamma_l^i\partial\gamma_m^i}| \leq C_1$ for any k, m, l. Since the cumulant generating function of $\bar{U}_j^i$ is

$$G^i_{\bar{U}_j^i}(\gamma^i) \;=\; \log E[e^{(\gamma^i)^t(J^i)^{-1}(Y_j^i - \mu^i)}] = G^i_{Y_j^i}((J^i)^{-1}\gamma^i) - (\gamma^i)^t(J^i)^{-1}\mu^i.$$

It follows that there exists a constant C_2 such that $|\frac{\partial^3 G^i_{\bar{U}_j^i}(\gamma^i)}{\partial\gamma_k^i\partial\gamma_l^i\partial\gamma_m^i}| \leq C_2$ for $||\gamma^i|| \leq \eta$. ■

Lemma A.1.2 *Let Δ_n^i and $\bar{U}_j^i$ be as defined in* (7.6) *and* (A.1)*, respectively. Let C_2 be as in Lemma A.1.1. Then, under Conditions (2)-(4), for any arbitrary constant a such that $a^2 > 1$, we have that if $\frac{C_2 p^3}{3\sqrt{n}} \leq a - 1$, then as $n \to \infty$,*

$$G^i_{\Delta_n^i}(\gamma^i) = \log\left(E\{\exp[(\gamma^i)^t\Delta_n^i]\}\right) \leq a^2||\gamma^i||^2/2 \quad \textit{for} \quad ||\gamma^i|| < p. \tag{A.2}$$

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

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

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

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

By the definition (7.6), we have $\Delta^i_n = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\bar{U}^i_j$. Since the moment generating function of $\bar{U}^i_j$ is $\exp G^i_{\bar{U}^i_j}(\gamma^i)$, then the moment generating function of Δ^i_n is

$$\begin{aligned} E[e^{(\gamma^i)^t\Delta^i_n}] &= E[e^{(\gamma^i)^t\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\bar{U}^i_j}] = \prod_{j=1}^{n} E[e^{(\frac{\gamma^i}{\sqrt{n}})^t\bar{U}^i_j}] \\ &= \exp\Big\{\frac{1}{2}(\gamma^i)^t\gamma^i + \frac{1}{6}\frac{1}{\sqrt{n}}\sum_{k=1}^{S_i}\sum_{l=1}^{S_i}\sum_{m=1}^{S_i}\big(\frac{\partial^3 G^i_{\bar{U}^i_j}(\frac{\gamma^i}{\sqrt{n}})}{\partial \gamma^i_k \partial \gamma^i_l \partial \gamma^i_m}|_{\gamma^i=\gamma^{i,*}}\big)\gamma^i_k\gamma^i_l\gamma^i_m\Big\}. \end{aligned}$$

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 $\frac{p}{\sqrt{n}} \to 0$, and thus $||\frac{\gamma^i}{\sqrt{n}}|| \leq \eta$ for n large enough. Therefore, by Lemma A.1.1, there

exists a constant C_2 such that $\left|\frac{\partial^3 G^i_{\bar{U}^i_j}(\frac{\gamma^i}{\sqrt{n}})}{\partial\gamma^i_k\partial\gamma^i_l\partial\gamma^i_m}\right| \le C_2$. It follows

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

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

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

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

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

It implies $\frac{1}{3}\frac{C_2}{\sqrt{n}}\sum_{m=1}^{S_i}\gamma^i_m \le a^2 - 1$ for any constant a with $a^2 > 1$. ■

Lemma A.1.3 *Under Conditions (2)-(4), for any $i \in \{1, 2, \ldots, p\}$ and n sufficiently large, there exists a constant a, $a^2 > 1$, such that*

$$P\{||\Delta^i_n||^2 > 3a^2p^2\} \le 10.4\exp\{-\frac{1}{6}p^2\}$$

where Δ^i_n is defined as in (7.6).

Proof. According to Lemma A.1.2, we have

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

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 and Zhilova [2013]. Following the notations in Spokoiny and 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 and 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 and 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 $||\Delta_n^i||^2 = O_p(p^2)$. ■

The next four lemmas are used to complete the proof of Theorem 7.2.1.

Lemma A.1.4 *Under Conditions (2)-(4), for any given* $i \in \{1, 2, \ldots, p\}$ *and for*

any given constant c, there exists a constant $c_5(c)$ *such that*

$$\frac{\int_{||u^i||^2 \le cM(p)} ||u^i|| \cdot |\pi^i(\theta_0^i + n^{-\frac{1}{2}}(J^i)^{-1}u^i)Z_n^i(u^i) - \pi^i(\theta_0^i)\tilde{Z}_n^i(u^i)|du^i}{\int \pi^i(\theta_0^i)\tilde{Z}_n^i(u^i)du^i} \le c_5 \frac{p^{13}\log p}{\sqrt{n}} \quad \text{(A.3)}$$

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

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

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

Since $cM(p) \ge ||u^i||^2 = ||\sqrt{n}J^i(\theta^i - \theta_0^i)||^2 \ge n\lambda_{min}(F^i)||\theta^i - \theta_0^i||^2$, then $||\theta^i - \theta_0^i|| \le \sqrt{\frac{cM(p)||(F^i)^{-1}||}{n}}$. By Proposition A.1.1, we have $\kappa_1^2 \le ||(F^i)^{-1}|| \le \kappa_2^2$. Based on Condition (4), $\frac{p^2\log p}{n} \to 0$. Therefore, $||\theta^i - \theta_0^i|| \to 0$. Using the fact $|e^x - 1| \le 2|x|$ for sufficiently small $|x|$ and Proposition A.1.5, we obtain

$$\sup_{u^i \in Q^i}\left\{||u^i|| \cdot |\frac{\pi^i(\theta_0^i + n^{-1/2}(J^i)^{-1}u^i)}{\pi^i(\theta_0^i)} - 1|\right\} \le 2\sqrt{cM(p)}M_1 p||\theta^i - \theta_0^i|| \le \frac{2cM_1\kappa_2 M(p)p}{\sqrt{n}}$$

where M_1 is a constant. We also have that

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

According to Lemma 2.3 in Ghosal [2000], we can obtain

$$\Big(\int \tilde{Z}_n^i(u^i)du^i\Big)^{-1} \int_{Q^i} |Z_n^i(u^i) - \tilde{Z}_n^i(u^i)|du^i \leq f^i(||\Delta_n^i||, c) \tag{A.4}$$

where

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

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 (A.4), 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 (A.3) 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 A.1.6, we have $B^i_{1n}(0) = O(p^9)$ and $B^i_{2n}(c\frac{M(p)}{S_i}) = O(p^{12})$. Therefore, there exist two constants c_1 and c_2 such that

$$\begin{aligned}\varphi^i_n(c) &\leq \frac{1}{6}[\frac{\big(cM(p)\big)^{\frac{1}{2}}c_1p^9}{\sqrt{n}} + n^{-1}cM(p)c_2p^{12}] \\ &= \frac{1}{6}\frac{p^{10}\sqrt{\log p}}{\sqrt{n}}[\sqrt{c}c_1 + c_2c\frac{p^4\sqrt{\log p}}{\sqrt{n}}]. \qquad \text{(A.5)}\end{aligned}$$

Since the first term in (A.5) is the dominating term, then there exists a constant $c_3(c)$ such that $\varphi^i_n(c) \leq c_3(c)\frac{p^{10}\sqrt{\log p}}{\sqrt{n}}$. By Condition (4), we have that $\varphi^i_n(c) \to 0$. Furthermore, using the fact $(1-x)^{-1} \leq 2$ and $-\log(1-x) \leq 2x$ for sufficiently small x, we have $[1-2\varphi^i_n(c)]^{-1} \leq 2$ and $e^{-(\frac{p_i^2}{2}+1)\log\left(1-2\varphi^i_n(c)\right)} \leq e^{(\frac{p_i^2}{2}+1)4\varphi^i_n(c)}$. Therefore, the following inequality holds

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

According to Lemma A.1.3, we see that $P(||\Delta^i_n||^2 \leq 3a^2p^2) > 1 - 10.4\exp\{-\frac{1}{6}p^2\}$. Therefore,

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

with probability greater than $1 - 10.4\exp\{-\frac{1}{6}p^2\}$. By Condition (4), we have that $\frac{p^{10}\sqrt{\log p}}{\sqrt{n}}(6a^2p^2 + 2p^2 + 2) \to 0$. Therefore, $\exp\left\{c_3(c)\frac{p^{10}\sqrt{\log p}}{\sqrt{n}}(4a^2p^2 + 2p^2 + 2)\right\} < 2$.

It follows

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

with 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 probability greater than $1 - 10.4\exp\{-\frac{1}{6}p^2\}$. Furthermore, we can get

$$\begin{aligned} R_1(||\Delta_n^i||, c) &= \frac{2cM_1\kappa_2 M(p)p}{\sqrt{n}}[1 + c_4(c)\frac{p^{12}\sqrt{\log p}}{\sqrt{n}}] + \sqrt{cM(p)}c_4(c)\frac{p^{12}\sqrt{\log p}}{\sqrt{n}} \\ &= \frac{2cM_1\kappa_2 p^3\log p}{\sqrt{n}} + \frac{p^{13}\log p}{\sqrt{n}}[c_4(c)\frac{2cM_1\kappa_2 p^2\log^{\frac{1}{2}} p}{\sqrt{n}} + \sqrt{c}c_4(c)] \quad \text{(A.6)} \end{aligned}$$

with probability greater than $1 - 10.4\exp\{-\frac{1}{6}p^2\}$. It is easy to see that the third term in (A.6) 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 probability greater than $1 - 10.4\exp\{-\frac{1}{6}p^2\}$. The proof is completed. ■

Lemma A.1.5 *Under Conditions (2)-(4), there exists a constant c large enough and a constant $c_9(c)$ such that for any given $i \in \{1, 2, \ldots, p\}$,*

$$\frac{\int_{||u^i||^2 > cM(p)} ||u^i||\pi^i(\theta_0^i + n^{-\frac{1}{2}}(J^i)^{-1}u^i)Z_n^i(u^i)du^i}{\int \pi^i(\theta_0^i)\tilde{Z}_n^i(u^i)du^i} \leq \exp[-c_9(c)p^2\log p]$$

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

Proof. Let

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

According to Lemma 2.2 in Ghosal [2000], we have that $Z_n^i(u^i) \leq \exp[-\frac{1}{4}cp^2\log p]$ with probability greater than $1 - 10.4\exp\{-\frac{1}{6}p^2\}$. Let $\pi_0^i(\theta^i)$ denotes the non-normalized local colored G-Wishart distribution. Then we obtain that

$$\begin{aligned} & R_2(||\Delta_n^i||, c) \\ \leq\ & \frac{\exp[-\frac{1}{4}cp^2\log p]}{(2\pi)^{S_i/2}\exp[\frac{||\Delta_n^i||^2}{2}]} \\ & \quad \times \int_{||\sqrt{n}J^i(\theta^i - \theta_0^i)||^2 > cM(p)} ||\sqrt{n}J^i(\theta^i - \theta_0^i)|| \frac{\pi_0^i(\theta^i)}{\pi_0^i(\theta_0^i)} n^{S_i/2}|J^i| d\theta^i \\ \leq\ & \exp[\frac{S_i}{2}\log n + \frac{1}{2}\log|F^i| - \frac{1}{4}cp^2\log p - \log\pi_0^i(\theta_0^i) \\ & \quad + \log\int_{||\sqrt{n}J^i(\theta^i - \theta_0^i)||^2 > cM(p)} ||\sqrt{n}J^i(\theta^i - \theta_0^i)|| \pi_0^i(\theta^i) d\theta^i] \end{aligned} \tag{A.7}$$

with probability greater than $1 - 10.4\exp\{-\frac{1}{6}p^2\}$. By Proposition A.1.4 and Lemma A.1.15, we have that

$$\begin{aligned} R_2(||\Delta_n^i||, c) \quad \leq \quad & \exp[\frac{S_i}{2}\log n + \frac{1}{2}\log|F^i| - \frac{1}{4}cp^2\log p + \frac{1}{2}p_i\kappa_2 - \frac{\delta^i - 2}{2}p_i\log\kappa_1 \\ & + M_7 p^2 \log p] \end{aligned}$$

with 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 A.1.3 implies $\log|F^i| = O(p^2)$. Therefore, there exists a constant c_6 such that $\log|F^i| \leq c_6 p^2$. It follows the RHS in (A.7) is bounded by the following term

$$\exp[\frac{p(p+1)}{4}\log p + \frac{1}{2}c_6 p^2 - \frac{1}{4}cp^2\log p + \frac{1}{2}p_i\kappa_2 - \frac{\delta^i - 2}{2}p_i\log\kappa_1 + M_7 p^2\log p]$$

with probability greater than $1 - 10.4\exp\{-\frac{1}{6}p^2\}$. Furthermore, there exists a constant c_8 such that

$$R_2(||\Delta_n^i||, c) \quad \leq \exp[\tfrac{p(p+1)}{4}\log p - \tfrac{1}{4}cp^2\log p + M_7 p^2\log p + c_8 p^2\log p]$$

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

Lemma A.1.6 *Under Conditions (2)-(4), for any given $i \in \{1, 2, \ldots, p\}$ and for any constant c such that $c > 11a^2$ and $a^2 > 1$, we have*

$$\int_{||u^i||^2 > cM(p)} ||u^i|| \phi(u^i; \Delta_n^i, I_{S_i}) du^i \leq \frac{2}{\sqrt{2\pi}} p^{-4a^2+4} + \sqrt{3a^2} \frac{2}{\sqrt{2\pi}} p^{-4a^2+3}$$

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

Proof. First we observe that

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

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 A.1.3, we can see that $||\Delta_n^i||^2 \leq 3a^2p^2$ with probability greater than $1 - 10.4 \exp\{-\frac{1}{6}p^2\}$ with $a^2 > 1$. As c is chosen that $c > 11a^2$, we can get $||v^i||^2 > cM(p) - ||\Delta_n^i||^2 > cM(p) - 3a^2p^2 > (11a^2 - 3a^2)p^2 \log p = 8a^2p^2 \log p$ with probability greater than $1 - 10.4 \exp\{-\frac{1}{6}p^2\}$. Thus

the following inequality holds with probability greater than $1 - 10.4\exp\{-\frac{1}{6}p^2\}$.

$$\begin{aligned}
&\int_{||u^i||^2>cM(p)} (||u^i - \Delta_n^i||)\phi(u^i; \Delta_n^i, I_{S_i})du^i \\
= &\int_{||v^i+\Delta_n^i||^2>cM(p)} ||v^i||\phi(v^i; 0, I_{S_i})dv^i \leq \sum_{j=1}^{S_i} \int_{||v^i||^2>8a^2M(p)} |v_j^i|\phi(v^i; 0, I_{S_i})dv^i,
\end{aligned}$$

where v_j^i is the j-th element of v^i. We also have that

$$\begin{aligned}
&\sum_{j=1}^{S_i} \int_{||v^i||^2>8a^2M(p)} |v_j^i|\phi(v^i; 0, I_{S_i})dv^i \leq \sum_{j=1}^{S_i}\sum_{k=1}^{S_i} \int_{\mathbb{R}^{S^i-1}} \int_{(v_k^i)^2>8a^2\frac{M(p)}{S_i}} |v_j^i|\phi(v^i; 0, I_{S_i})dv^i \\
= &\ S_i \int_{\mathbb{R}^{S^i-1}} \int_{(v_k^i)^2>8a^2\frac{M(p)}{S_i}} |v_k^i|\phi(v^i; 0, I_{S_i})dv^i + S_i \sum_{j\neq k} \int_{\mathbb{R}^{S^i-1}} \int_{(v_k^i)^2>8a^2\frac{M(p)}{S_i}} |v_j^i|\phi(v^i; 0, I_{S_i})dv^i \\
= &\ S_i \int_{(v_k^i)^2>8a^2\frac{M(p)}{S_i}} |v_k^i| \frac{1}{\sqrt{2\pi}} e^{-\frac{(v_k^i)^2}{2}} dv_k^i \\
&+S_i \sum_{j\neq k} \int_{(v_k^i)^2>8a^2\frac{M(p)}{S_i}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(v_k^i)^2}{2}} dv_k^i \int_{-\infty}^{\infty} |v_j^i| \frac{1}{\sqrt{2\pi}} e^{-\frac{(v_j^i)^2}{2}} dv_j^i \\
= &\ 2S_i \int_{v_k^i>\sqrt{8a^2\frac{M(p)}{S_i}}} v_k^i \frac{1}{\sqrt{2\pi}} e^{-\frac{(v_k^i)^2}{2}} dv_k^i \\
&+S_i \sum_{j\neq k} 2\int_{v_k^i>\sqrt{8a^2\frac{M(p)}{S_i}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(v_k^i)^2}{2}} dv_k^i [2\int_0^{\infty} v_j^i \frac{1}{\sqrt{2\pi}} e^{-\frac{(v_j^i)^2}{2}} dv_j^i] \\
< &\ 2S_i \int_{v_k^i>\sqrt{8a^2\frac{M(p)}{S_i}}} v_k^i \frac{1}{\sqrt{2\pi}} e^{-\frac{(v_k^i)^2}{2}} dv_k^i \\
&+S_i \sum_{j\neq k} 2\int_{v_k^i>\sqrt{8a^2\frac{M(p)}{S_i}}} v_k^i \frac{1}{\sqrt{2\pi}} e^{-\frac{(v_k^i)^2}{2}} dv_k^i [2\int_0^{\infty} v_j^i \frac{1}{\sqrt{2\pi}} e^{-\frac{(v_j^i)^2}{2}} dv_j^i] \\
= &\ 2S_i \int_{v_k^i>\sqrt{8a^2\frac{M(p)}{S_i}}} v_k^i \frac{1}{\sqrt{2\pi}} e^{-\frac{(v_k^i)^2}{2}} dv_k^i \big[1 + \sum_{j\neq k} 2\int_0^{\infty} v_j^i \frac{1}{\sqrt{2\pi}} e^{-\frac{(v_j^i)^2}{2}} dv_j^i\big] \\
= &\ 2S_i \frac{1}{\sqrt{2\pi}} e^{-\frac{8a^2M(p)}{2S_i}} [1 + (S_i - 1)2\frac{1}{\sqrt{2\pi}}] \leq 2S_i^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{8a^2M(p)}{2S_i}} \leq 2p^4 \frac{1}{\sqrt{2\pi}} p^{-4a^2}
\end{aligned}$$

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

$$\int_{||u^i||^2 > cM(p)} ||\Delta_n^i||\phi(u^i; \Delta_n^i, I_{S_i})du^i \leq \sqrt{3a^2}p^3\frac{2}{\sqrt{2\pi}}p^{-4a^2}$$

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

■

Lemma A.1.7 *For a given $i \in \{1, 2, \ldots, p\}$, 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 $q_*^i(\theta^i)$ be the posterior distribution of θ^i. Therefore, we have that

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

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

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

We thus have

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

■

Lemma A.1.8 *(Parallel to Lemma A.1.2) Let Δ_n^i and $\bar{U}_j^i$ be as defined in* (7.6) *and* (A.1)*, respectively. Let C_2 be defined as in Lemma A.1.1. Then under Conditions (2), (4*) and (5), for any arbitrary constant a such that $a^2 > 1$, we have that if $\frac{C_2 \log p}{3\sqrt{n}} \leq a - 1$, for $||\gamma^i|| < \log p$ and n sufficiently large,*

$$G^i_{\Delta_n^i}(\gamma^i) = \log\big(E\{\exp[(\gamma^i)^t \Delta_n^i]\}\big) \leq a^2||\gamma^i||^2/2.$$

Proof. Since $||\gamma^i|| < \log p$ and $\frac{\log p}{\sqrt{n}} \to 0$ by Condition (4*), then $||\frac{\gamma^i}{\sqrt{n}}|| \leq \eta$ where η as given in Lemma A.1.1 is the size of the neighborhood for γ^i. Therefore, by Lemma A.1.1, there exists a constant C_2 such that $\Big|\frac{\partial^3 G^i_{\bar{U}_j^i}(\frac{\gamma^i}{\sqrt{n}})}{\partial\gamma_k^i \partial\gamma_l^i \partial\gamma_m^i}\Big| \leq C_2$. We also have

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

According to Condition (4*), $\frac{\log p}{\sqrt{n}} = o(1)$. Therefore, for any arbitrary constant a such that $a^2 > 1$, $\frac{1}{3}\frac{C_2}{\sqrt{n}} \sum_{m=1}^{S_i} \gamma_m^i \leq a^2 - 1$. Following the argument similar to that of Lemma A.1.2, we obtain

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

■

Lemma A.1.9 *(Parallel to Lemma A.1.3) Under Conditions (2), (4*) and (5), for any $i \in \{1, 2, \ldots, p\}$ and n sufficiently large, there exists a constant a, $a^2 > 1$, such*

that

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

where Δ_n^i *is defined as in* (7.6).

Proof. According to Lemma A.1.8, we have

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

where a is a constant with $a^2 > 1$. Condition (5) implies $a^2 \log^2 p > S_i$. Let x_c^i be defined as in the proof of Lemma A.1.3. Then $x_c^i > \frac{1}{4}a^2 \log^2 p$ and let $x = \frac{1}{6}\log^2 p$, then we have $\frac{S_i}{6.6} < x < x_c^i$. Following similar argument as in the proof of Lemma A.1.3, we can obtain that $P(||\Delta_n^i||^2 \geq 3a^2 \log^2 p) \leq 10.4e^{-\frac{1}{6}\log^2 p}$. ■

Lemma A.1.10 *Let* $\hat{\theta}^i$ *be the MLE of* θ^i *in the* i*-th local model. Under Conditions (2), (4*) and (5), for any* $i \in \{1, 2, \ldots, p\}$,

$$\sqrt{n}||J^i(\hat{\theta}^i - \theta_0^i)|| \leq c' \log p$$

with probability greater than $1 - 10.4\exp\{-\frac{1}{6}\log^2 p\}$, *where* $c' = 1.2\sqrt{3}a\frac{\lambda_{max}(F^i)}{\lambda_{min}^2(F^i)}$.

Proof. Let $B^i(\theta^i) = \psi'(\theta^i) - \bar{Y}^i$ be the negative of the score function. Then the MLE $\hat{\theta}^i$ satisfy the likelihood equation $B^i(\hat{\theta}^i) = 0$. Let $b_n = \frac{\sqrt{3}a\log p}{\sqrt{n}}\frac{\lambda_{max}^{\frac{1}{2}}(F^i)}{\lambda_{min}(F^i)}$ with $a^2 > 1$. We are going to show with probability greater than $1 - 10.4\exp\{-\frac{1}{6}\log^2 p\}$,

for any θ^i on the ball $||\theta^i - \theta_0^i|| = 1.2b_n$, we have

$$(\theta^i - \theta_0^i)^t B^i(\theta^i) > 0. \tag{A.8}$$

Because that according to Theorem 6.3.4 of Ortega and Rheinboldt [1970], this will imply that there exists a root of $B^i(\hat{\theta}^i) = 0$ inside the ball $||\theta^i - \theta_0^i|| \leq 1.2b_n$ and thus with probability greater than $1 - 10.4\exp\{-\frac{1}{6}\log^2 p\}$, $||J^i(\hat{\theta}^i - \theta_0^i)|| \leq \lambda_{max}^{\frac{1}{2}}(F^i)1.2b_n \leq c'\frac{\log p}{\sqrt{n}}$. To complete the proof, it now suffices to show the inequality (A.8) holds. Based on (2.3) in Proposition 2.1 of Portnoy [1988], we have

$$\begin{aligned}
(\theta^i - \theta_0^i)^t B^i(\theta^i) &= (\theta^i - \theta_0^i)^t(\psi'(\theta^i) - \bar{Y}^i) \\
&= (\theta^i - \theta_0^i)^t\mu^i + (\theta^i - \theta_0^i)^t\psi''(\theta_0^i)(\theta^i - \theta_0^i) \\
&\quad + \frac{1}{2}E_{\bar{\theta}^i}[(\theta^i - \theta_0^i)^t V_j^i]^3 - (\theta^i - \theta_0^i)^t\bar{Y}^i \\
&= -(\theta^i - \theta_0^i)^t[\bar{Y}^i - \mu^i] + (\theta^i - \theta_0^i)^t\psi''(\theta_0^i)(\theta^i - \theta_0^i) \\
&\quad + \frac{1}{2}E_{\bar{\theta}^i}[(\theta^i - \theta_0^i)^t V_j^i]^3 \\
&= term1 + term2 + term3,
\end{aligned}$$

where $\mu^i = \psi'(\theta_0^i)$, V_j^i is defined as in (7.5) and $\bar{\theta}^i$ is a point on the line segment between θ^i and θ_0^i. It is easy to see that

$$term2 \geq \lambda_{min}(F^i)||(\theta^i - \theta_0^i)||^2.$$

For $term3$, under Condition (5), from (A.10), (A.11) and Lemma A.1.11, we see that $\sup\{|E_\theta(a^t V_j^i)^3| : ||a|| = 1, ||\theta^i - \theta_0^i||$ is bounded. Since $b_n \to 0$, then $\frac{0.1}{b_n}\lambda_{min}(F^i) \to$

∞. Therefore,

$$\sup\{|E_\theta(a^t V_j^i)^3| : ||a|| = 1, ||\theta^i - \theta_0^i|| = 1.2b_n\} \leq \frac{0.1}{b_n}\lambda_{min}(F^i).$$

It follows

$$term3 \geq -\frac{0.05}{b_n}\lambda_{min}(F^i).$$

In $term1$, there is a random term $\bar{Y}^i - \psi'(\theta_0^i)$. We will now show that

$$term1 \geq \frac{\sqrt{3}a\log p}{\sqrt{n}}\lambda_{max}^{\frac{1}{2}}(F^i)||\theta^i - \theta_0^i||$$

with probability greater than $1 - 10.4\exp\{-\frac{1}{6}\log^2 p\}$. According to Lemma A.1.9, we have $||\Delta_n^i||^2 \leq 3a^2\log^2 p$ with probability greater than $1 - 10.4\exp\{-\frac{1}{6}\log^2 p\}$. Furthermore, since

$$||\Delta_n^i||^2 = ||\sqrt{n}(J^i)^{-1}(\bar{Y}^i - \mu^i)||^2 = n(\bar{Y}^i - \mu^i)^t(F^i)^{-1}(\bar{Y}^i - \mu^i) \geq \frac{n}{\lambda_{max}(F^i)}||\bar{Y}^i - \mu^i||^2,$$

then $\frac{n}{\lambda_{max}(F^i)}||\bar{Y}^i - \mu^i||^2 \leq 3a^2\log^2 p$ with probability greater than $1 - 10.4\exp\{-\frac{1}{6}\log^2 p\}$. It implies $||\bar{Y}^i - \mu^i||^2 \leq \frac{\sqrt{3}a\log p}{\sqrt{n}}\lambda_{max}^{\frac{1}{2}}(F^i)$ with probability greater than $1 - 10.4\exp\{-\frac{1}{6}\log^2 p\}$. Consequently, $term1 \geq \frac{\sqrt{3}a\log p}{\sqrt{n}}\lambda_{max}^{\frac{1}{2}}(F^i)||\theta^i - \theta_0^i||$. Combining the above results, on

the ball of $||\theta^i - \theta_0^i|| = 1.2b_n$, we have

$$\begin{aligned}
& (\theta^i - \theta_0^i)^t B^i(\theta^i) \\
\geq & -\frac{\sqrt{3}a\log p}{\sqrt{n}}\lambda_{max}^{\frac{1}{2}}(F^i)||\theta^i - \theta_0^i|| + \lambda_{min}(F^i)||(\theta^i - \theta_0^i)||^2 - \frac{0.05}{b_n}\lambda_{min}(F^i)||\theta^i - \theta_0^i||^3 \\
= & -3a^2\sqrt{\frac{\log^2 p}{n}}\lambda_{max}^{\frac{1}{2}}(F^i)1.2b_n + \lambda_{min}(F^i)(1.2b_n)^2 - \lambda_{min}(F^i)\frac{0.05}{b_n}(1.2b_n)^3 \\
\geq & \lambda_{min}(F^i)b_n^2[-1.2 + (1.2)^2 - 0.05(1.2)^3] > 0
\end{aligned}$$

with probability greater than $1 - 10.4\exp\{-\frac{1}{6}\log^2 p\}$. Therefore, we proved that $||\hat{\theta}^i - \theta_0^i|| \leq 1.2b_n$ with probability greater than $1 - 10.4\exp\{-\frac{1}{6}\log^2 p\}$. It follows

$$||J^i(\hat{\theta}^i - \theta_0^i)|| \leq \lambda_{max}^{\frac{1}{2}}(F^i)||\hat{\theta}^i - \theta_0^i|| \leq \lambda_{max}^{\frac{1}{2}}(F^i)1.2b_n$$

with probability greater than $1 - 10.4\exp\{-\frac{1}{6}\log^2 p\}$. ■

Proposition A.1.1 *Let F^i be defined in definition* (7.4) *for any $i \in \{1, 2, \ldots, p\}$, then under Condition (2), we have 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 $\psi_u(\theta^i) = (-\frac{1}{2}\log|K^i| + \frac{p_i}{2}\log(2\pi))\mathbf{1}_{K^i \in P_{G_i}}$. Let τ and ϖ be the numbers of eigenvalues of G^i and F^i. Since F^i is a linear projection of G^i onto the space of uncolored symmetric matrices, then $\tau > \varpi$. Under Condition (2) and

by Proposition A.1.7, for any l, $1 \leq l \leq \varpi$, we have

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

■

Proposition A.1.2 *For any $i \in \{1, 2, \ldots, p\}$, let $K_{\alpha\beta}^{i,0}$ be the (α, β) entry of K_0^i. Under Condition (2), we have $|K_{\alpha\beta}^{i,0}| \leq \kappa_2$.*

Proof. By Condition (2), we have $\lambda_{max}(K_0^i) \leq \kappa_2$ for any $i \in \{1, 2, \ldots, p\}$. 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 $\kappa_2 \geq \lambda_j(K_0^i), j = 1, 2, \cdots, p_i$. It follows that $\kappa_2 I_{p_i} - K_0^i$ is a positive semidefinite matrix. Since the diagonal elements of a positive semidefinite $\kappa_2 I_{p_i} - K_0^i$ are all non negative, then $\kappa_2 - K_{\alpha\alpha}^{i,0} \geq 0, \alpha = 1, 2, \ldots, p_i$. It follows $0 < K_{\alpha\alpha}^{i,0} \leq \kappa_2$. 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 $|K_{\alpha\beta}^{i,0}| < (K_{\alpha\alpha}^{i,0} K_{\beta\beta}^{i,0})^{1/2} < \kappa_2$. ■

The next four propositions provide the properties of $\log|F^i|$, the colored G-Wishart prior, the third and fourth moments of the normalized Y_j^i.

Proposition A.1.3 *Under Condition (3), for any $i \in \{1, 2, \ldots, p\}$, we have the trace of F^i satisfies $tr(F^i) = O(p^2)$ and the determinant $|F^i|$ satisfies $\log|F^i| = O(p^2)$.*

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 $\log|F^i|$. Since $|F^i| = \prod_{j=1}^{S_i} \lambda_j(F^i) \leq \left(\frac{\sum_{j=1}^{S_i} \lambda_j(F^i)}{S_i} \right)^{S_i} = \left(\frac{tr(F^i)}{S_i} \right)^{S_i}$, then

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

The proposition is proved. ■

Proposition A.1.4 *Under Condition (2), for any $i \in \{1, 2, \ldots, p\}$, we have*

$$\log \pi^i_0(K^i_0) \geq -\frac{1}{2} p_i \kappa_2 + \frac{\delta^i - 2}{2} p_i \log \kappa_1$$

when $D^i = I_{p_i}$.

Proof. The non-normalized colored G-Wishart distribution can be rewritten as

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

The last inequality due to Condition (2). ■

Proposition A.1.5 *(Lipschitz continuity) For any $i \in \{1, 2, \ldots, p\}$ and any constant c, there exists a constant M_1 such that*

$$|\log \pi^i(\theta^i) - \log \pi^i(\theta_0^i)| \leq M_1 p ||\theta^i - \theta_0^i||$$

when $||\theta^i - \theta_0^i|| \leq \sqrt{||(F^i)^{-1}||cM(p)/n} \to 0$.

Proof. Let $\pi_0^i(\theta^i)$ be the non-normalized colored G-Wishart distribution for the local model. By mean value theorem, we have

$$\begin{aligned} |\log \pi^i(\theta^i) - \log \pi^i(\theta_0^i)| &= |\log \pi_0^i(\theta^i) - \log \pi_0^i(\theta_0^i)| = |(\theta^i - \theta_0^i)^t \frac{\partial \log \pi_0^i(\theta^i)}{\partial \theta^i}|_{\theta^i = \check{\theta}^i}| \\ &= ||\theta^i - \theta_0^i|| \cdot \sqrt{\sum_{j=1}^{S_i} \Big[-\frac{1}{2} tr(\delta_j^i D^i) + \frac{\delta^i - 2}{2} tr(\delta_j^i (\check{K}^i)^{-1}) \Big]^2}, \end{aligned}$$

where $\check{\theta}^i$ is the point on the line segment joining θ^i and θ_0^i. Since $||\theta^i - \theta_0^i|| \to 0$, then $(\check{K}^i)^{-1} \to (K_0^i)^{-1}$. According to Condition (2) and Proposition A.1.2, each entry of $(K_0^i)^{-1}$ is uniformly bounded, then using the similar proof of Lemma A.1.11, each entry $(\check{K}^i)^{-1}$ is uniformly bounded. Therefore, there exists a constant M_1 such that

$$\sqrt{\sum_{j=1}^{S_i} \Big[-\frac{1}{2} tr(\delta_j^i D^i) + \frac{\delta - 2}{2} tr(\delta_j^i \Sigma_0^i) \Big]^2} \leq \sqrt{\frac{p_i(1+p_i)}{2} M_1^2} \leq \sqrt{\frac{p(1+p)}{2} M_1^2} = M_1 p.$$

■

Proposition A.1.6 *For any $i \in V$, let Y_j^i and V_j^i be defined in (7.3) and (7.5), respectively. Then $B_{1n}^i(c) = O(p^9)$ and $B_{2n}^i(c) = O(p^{12})$.*

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

$$\begin{aligned} E_{\theta^i}|a^t(J^i)^{-1}Y_j^i|^h &\leq E_{\theta^i}\Big[(|a_1|,|a_2|,\ldots,|a_{S_i}|)\begin{pmatrix} b\sum_{k=1}^{S_i}|Y_{jk}^i| \\ b\sum_{k=1}^{S_i}|Y_{jk}^i| \\ \vdots \\ b\sum_{k=1}^{S_i}|Y_{jk}^i| \end{pmatrix}\Big]^h \\ &= E_{\theta^i}\Big[(b\sum_{k=1}^{S_i}|Y_{jk}^i|)\sum_{k=1}^{S_i}|a_k|\Big]^h. \end{aligned} \tag{A.9}$$

According to Cauchy-Schwarz inequality, we have that

$$\begin{aligned} E_{\theta^i}|a^t(J^i)^{-1}Y_j^i|^h &\leq E_{\theta^i}\Big[b(\sum_{k=1}^{S_i}|Y_{jk}^i|)\sqrt{S_i}||a||\Big]^h \\ &\leq b^h(S_i)^{h/2}E_{\theta^i}\Big[\sum_{k_1=1}^{S_i}\cdots\sum_{k_h=1}^{S_i}|Y_{jk_1}^i|\cdots|Y_{jk_h}^i|\Big]. \end{aligned} \tag{A.10}$$

According to Lemma A.1.11, each entry of θ^i is bounded when $||J^i(\theta^i - \theta_0^i)||^2 \leq \frac{cS_i}{n} \to 0$. By Lemma A.1.12, we have $E_{\theta^i}\Big[|Y_{jk_1}^i|\cdots|Y_{jk_h}^i|\Big]$ is bounded for $h = 1,2,3,4$. Therefore, $E_{\theta^i}|a^t(J^i)^{-1}Y_j^i|^h = O(p_i^{3h})$. Similarly, $|a^t(J^i)^{-1}E_{\theta^i}(Y_j^i)|^h =$

$O(p_i^{3h})$. Hence, we have

$$\begin{aligned} & E_{\theta^i}|a^t V_j^i|^3 = E_{\theta^i}|a^t (J^i)^{-1} Y_j^i - a^t (J^i)^{-1} E_{\theta^i}(Y_j^i)|^3 \\ \leq \; & E_{\theta^i}|a^t (J^i)^{-1} Y_j^i|^3 + 3|a^t (J^i)^{-1} E_{\theta^i}(Y_j^i)| E_{\theta^i}(a^t (J^i)^{-1} Y_j^i)^2 \\ & +3[a^t (J^i)^{-1} E_{\theta^i}(Y_j^i)]^2 E_{\theta^i}|a^t (J^i)^{-1} Y_j^i| + |[a^t (J^i)^{-1} E_{\theta^i}(Y_j^i)]^3| \quad \text{(A.11)} \\ = \; & O(p_i^9) = O(p^9). \end{aligned}$$

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

Lemma A.1.11 *Let θ_j^i be the j-th element of θ^i, $i = 1, 2, \ldots, p$. Under Condition (2), for $||\theta^i - \theta_0^i|| \leq \varepsilon_1$, we have that $|\theta_j^i| \leq \varepsilon_1 + \kappa_2$.*

Proof. Let $\theta_{j,0}^i$ be the j-th element of θ_0^i, $i = 1, 2, \ldots, p$. Since $||\theta^i - \theta_0^i|| \leq \varepsilon_1$, then $\sqrt{\sum_{j=1}^{S_i} (\theta_j^i - \theta_{j,0}^i)^2} \leq \varepsilon_1$. Therefore, for any $j \in \{1, 2, \ldots, S_i\}$, we have

$$\sqrt{(\theta_j^i - \theta_{j,0}^i)^2} \leq \sqrt{\sum_{j=1}^{S_i} (\theta_j^i - \theta_{j,0}^i)^2} \leq \varepsilon_1.$$

It implies $|\theta_j^i - \theta_{j,0}^i| \leq \varepsilon_1$. By Proposition A.1.2, under Condition (2), we have $|\theta_{j,0}^i| \leq \kappa_2$. It follows $|\theta_j^i| \leq \varepsilon_1 + \kappa_2$. ■

Lemma A.1.12 *Let Y_j^i be defined in (7.3) and denote $Y_j^i = (Y_{j1}, Y_{j2}, \cdots, Y_{jS_i})^t$, under Condition (2) and $||\theta^i - \theta_0^i|| \leq \varepsilon_1$, we have $E_{\theta^i}\left[|Y_{jk_1}^i| \cdots |Y_{jk_h}^i|\right]$ is bounded for $h = 1, 2, 3, 4$.*

Proof. According to Lemma A.1.11, each element of θ^i is bounded. Since $Y^i_{jk} = -\frac{1}{2}tr(\delta^i_k X^i_j (X^i_j)^t)$, by Isserlis' Theorem, the moments of every entry of $X^i_j (X^i_j)^t$ is finite. By Condition (3), $E_{\theta^i}(Y^i_{jk})$ is bounded and $E_{\theta^i}(Y^i_{jk})^2$ is also bounded. By Hölder's inequality, we have $E_{\theta^i}[|XY|] \leq (E_{\theta^i}[|X|^p])^{\frac{1}{p}} (E_{\theta^i}[|Y|^q])^{\frac{1}{q}}$. Therefore, when $h = 1$, $E_{\theta^i}(|Y^i_{jk_1}|) \leq [E_{\theta^i}(Y^i_{jk_1})^2]^{\frac{1}{2}}$ is bounded. When $h = 2$, we have

$$E_{\theta^i}(|Y^i_{jk_1}||Y^i_{jk_2}|) \leq [E_{\theta^i}(Y^i_{jk_1})^2]^{\frac{1}{2}} [E_{\theta^i}(Y^i_{jk_2})^2]^{\frac{1}{2}}.$$

It follows $E_{\theta^i}(|Y^i_{jk_1}||Y^i_{jk_2}|)$ is bounded. When $h = 3$, we have

$$E_{\theta^i}(|Y^i_{jk_1}||Y^i_{jk_2}||Y^i_{jk_3}|) \leq [E_{\theta^i}(|Y^i_{jk_1}||Y^i_{jk_2}|)^2]^{\frac{1}{2}} [E_{\theta^i}(Y^i_{jk_3})^2]^{\frac{1}{2}}.$$

Since $E_{\theta^i}(|Y^i_{jk_1}||Y^i_{jk_2}|)$ is bounded, then $E_{\theta^i}(|Y^i_{jk_1}||Y^i_{jk_2}|)^2$ is also bounded. Therefore, $E_{\theta^i}(|Y^i_{jk_1}||Y^i_{jk_2}||Y^i_{jk_3}|)$ is bounded. Consequently, $E_{\theta^i}\left[|Y^i_{jk_1}| \cdots |Y^i_{jk_h}|\right]$ is bounded for $h = 1, 2, 3, 4$. ■

Proposition A.1.7 *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 $\mu_1 < \mu_2 < \cdots < \mu_m$ are the eigenvalues of g and $\lambda_1 < \lambda_2 < \cdots < \lambda_n$ are the eigenvalues of f, $n < m$, then for any $j = 1, 2, \cdots, n$, the following

inequalities hold

$$\mu_1 \leq \lambda_j \leq \mu_m.$$

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

$$[g]_{e'}^{e'} = \begin{pmatrix} a & b_1 & \cdots & b_n \\ b_1 & \lambda_1 & 0 & 0 \\ \cdots & \cdots & \ddots & 0 \\ b_n & \cdots & 0 & \lambda_n \end{pmatrix}.$$

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

$$\mu_1 \leq \lambda_j \leq \mu_{n+1} \quad j = 1, 2, \ldots, n.$$

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

Lemma A.1.13 *For any $i \in \{1, 2, \ldots, p\}$, let $T^i(\gamma^i)$ be a symmetric matrix with dimension p_i. Then there exists a constant η such that with $||T^i(\gamma^i)||_F \leq \eta$, the matrix $I_{p_i} + T^i(\gamma^i)\Sigma_0^i$ 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, $z^t(I_{p_i} + T^i(\gamma^i)\Sigma_0^i)z$ 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 η, such that $\eta < \kappa_1$. It follows $z^t(I_{p_i} + T^i(\gamma^i)\Sigma_0^i)z \geq ||z||^2 > 0$ when $||T^i(\gamma^i)||_F \leq \eta$. ■

Lemma A.1.14 *Let $K = (K_{ij})_{1 \leq i,j \leq p}$ be $p \times p$ positive semi-definite matrix. Then*

$$||K||_F \leq tr^2(K).$$

Proof. Because K is positive semi-definite, we have $K_{ij}^2 \leq K_{ii}K_{jj}$. Thus

$$\sum_{1 \leq i,j \leq p} K_{ij}^2 \leq \sum_{1 \leq i,j \leq p} K_{ii}K_{jj} = (\sum_{1 \leq i \leq p} K_{ii})(\sum_{1 \leq j \leq p} K_{jj}) = (\sum_{1 \leq i \leq p} K_{ii})^2.$$

■

In order to prove the following Theorem , we start from a finite-dimensional real linear space E of dimension n (thus isomorphic to $\mathbb{R}^n$ but we prefer to avoid the

use of artificial coordinates). We denote by E^* its dual space, that means the set of linear applications $\theta : E \mapsto \mathbb{R}$. We denote $\langle \theta, x \rangle = \theta(x)$. If E is Euclidean, the dual E^* is identified with E and $\langle \theta, x \rangle$ is the scalar product.

Consider a non empty open convex cone C with closure $\overline{C}$ such that C is proper, that is to say such that

$$\overline{C} \cap (-\overline{C}) = \{0\}.$$

The dual cone of C is

$$C^* = \{\theta \in E^* \ ; \ \langle \theta, x \rangle \ \geq 0 \ \forall x \in \overline{C} \setminus \{0\}\}.$$

This is a standard result of convex analysis that C^* is not empty [Faraut and Korányi, 1994]. In general the description of C^* is a non trivial matter.

A polynomial P on E is a function $P : E \mapsto \mathbb{R}$ such that if $e = (e_1, \ldots, e_n)$ is a basis of E and if $x = x_1 e_1 + \cdots + x_n e_n \in E$ then $P(x)$ is a polynomial with respect to the real variables $(x_1, \ldots, x_n)$. Needless to say the definition does not depend on the particular chosen basis e. A polynomial P is homogeneous of degree k if for all $\lambda \in \mathbb{R}$ and all $x \in E$ we have

$$P(\lambda x) = \lambda^k P(x).$$

Theorem A.1.8 *Let C be an open convex and proper cone of E, let P be a homogeneous polynomial on E of degree k and let $\alpha > -n/k$. We assume that $P(x) > 0$*

on C. We choose a Lebesgue measure dx on E. For $\theta \in E^*$ consider the integral

$$L(\theta) = \int_C e^{-\langle \theta, x \rangle} P(x)^\alpha dx \leq \infty.$$

If $\theta \notin C^*$ the integral $L(\theta)$ diverges. If $\theta \in C^*$ denote $H_1 = \{x \in E \ ; \ \langle \theta, x \rangle = 1\}$. Then $\overline{C} \cap H_1$ is compact. In this case $\theta \in C^*$, the integral $L(\theta)$ is finite if and only if $\int_{C \cap H_1} P(x)^\alpha dx$ is finite. Furthermore

$$L(\theta) = \Gamma(\alpha k + n) \int_{C \cap H_1} P(x)^\alpha dx. \tag{A.12}$$

Proof. (personal communication from G. Letac) Suppose that $\theta_0 \in C^*$ and let us show (A.12). Consider the affine hyperplanes H_1 and H_0 of E defined by

$$H_1 = \{x \in E \ ; \ \langle \theta_0, x \rangle = 1\}, \quad H_0 = \{x \in E \ ; \ \langle \theta_0, x \rangle = 0\}.$$

The convex set $\overline{C} \cap H_1$ is compact. To see this let us choose an arbitrary scalar product on E. Observe that the function $u \mapsto \langle \theta_0, u \rangle$ defined on the intersection of $\overline{C}$ with the unit sphere of E is continuous and reaches a minimum $m > 0$ since the set of definition is compact. Thus for all $x \in \overline{C} \cap H_1$ we have

$$\|x\| \leq \frac{1}{m} \langle \theta_0, x \rangle = \frac{1}{m}$$

and the closed set $\overline{C} \cap H_1$ is also bounded, thus compact.

We fix now $h_1 \in H_1$ and we write any element x of E in a unique way as $x = x_0 + x_1 h_1$ where x_1 is a number and x_0 is in H_0. If E is Euclidean, a natural

choice for h_1 is $\theta_0/\|\theta_0\|^2$ although other choices would be possible. We also write $x = (x_0, x_1)$ for short. We denote by $K \subset H_0$ the set of x_0 such that $x_0 + h_1 = (x_0, 1)$ is in $\overline{C} \cap H_1$. Note that K is also compact. We get that $x = (x_0, x_1)$ is in $\overline{C} \setminus \{0\}$ if and only if $y = x_0/x_1 \in K$ and $x_1 > 0$. To see this denote

$$C_1 = \{(x_0, x_1) \ ; \ y = x_0/x_1 \in K, \ x_1 > 0\}.$$

The inclusion $C_1 \subset \overline{C} \setminus \{0\}$ is obvious as well as $\overline{C} \setminus H_0 \subset C_1$. However if $(x_0, 0)$ is in $\overline{C} \cap H_0$ and if $x_0 \neq 0$ this implies that $(\lambda x_0, 0)$ is in $\overline{C} \cap H_0$ for all $\lambda > 0$ and thus $\lambda x_0 \in K$ for all $\lambda > 0$: this contradicts the compactness of K. As a result $(x_0, 0)$ in $\overline{C} \cap H_0$ implies $x_0 = 0$. This implies $\overline{C} \setminus \{0\} = \overline{C} \setminus H_0$ and thus $\overline{C} \setminus \{0\} = C_1$

We are now in position to make the change of variable $(x_0, x_1) \mapsto (y = x_0/x_1, x_1)$ in the integral $L(\theta_0)$ with an easy Jacobian, since $\dim H_0 = n - 1$:

$$dx = dx_0 dx_1 = x_1^{n-1} dy dx_1.$$

We get

$$L(\theta_0) = \int_C e^{-x_1} P(x_0, x_1)^\alpha dx_0 dx_1 = \int_K I(y) dy$$

where

$$I(y) = \int_0^\infty e^{-x_1} P(yx_1, x_1)^\alpha x_1^{n-1} dx_1 = P(y, 1)^\alpha \Gamma(\alpha k + n) = P(y + h_1)^\alpha \Gamma(\alpha k + n)$$

from the homogeneity of the polynomial P. Thus (A.12) is proved.

Suppose that $\theta_0 \notin C^*$. This is saying that there exists $x_0 \in C$ such that $\langle \theta_0, x_0 \rangle \leq 0$. Let us show that $L(\theta_0) = \infty$. Since C is open we may assume that $\langle \theta_0, x_0 \rangle < 0$. Choose an arbitrary scalar product on E. There exists ϵ such that for all x in $B = \{x \ ; \ \|x - x_0\| < \epsilon\}$ we have $x \in C$ and $\langle \theta_0, x \rangle < 0$. Consider the open subcone $C_1 = \{\lambda x \ ; \ x \in B, \ \lambda > 0\}$ of C. We can write

$$L(\theta_0) \geq \int_{C_1} e^{-\langle \theta_0, x \rangle} P(x)^\alpha dx \geq \int_{C_1} P(x)^\alpha dx.$$

Clearly the last integral diverges for $\alpha \geq 0$. For $-n/k < \alpha < 0$ we use the same trick: we parameterize C_1 with the help of the compact set $\overline{C_1} \cap H_1$ by considering the compact set K_1 of $y \in H_0$ such that $y + h_1 \in \overline{C_1} \cap H_1$ and we write

$$\int_{C_1} P(x)^\alpha dx = \int_{K_1} \int_0^\infty P^\alpha(x_0, x_1) dx = \int_{K_1} P(y, 1)^\alpha \left(\int_0^\infty x_1^{\alpha k + n - 1} dx_1 \right) dy = \infty.$$

This proves $\Rightarrow$. ■

Lemma A.1.15 *For any $i \in \{1, 2, \ldots, p\}$, there exists a constant M_7 such that*

$$\log \int_{||\sqrt{n} J^i(\theta^i - \theta_0^i)||^2 > cM(p)} ||\sqrt{n} J^i(\theta^i - \theta_0^i)|| \pi_0^i(\theta^i) d\theta^i \leq \exp[M_7 p^2 \log p].$$

Proof. Without loss of generality, let $\theta_k^i, k = 1, \ldots, s_i$, be the entries of K^i on the diagonal and $\theta_k^i, k = s_i + 1, \ldots, S_i$, be the off-diagonal entries. We assume that

$D^i = I_{p_i}$, which we need later on anyway. Then

$$tr(K^i D^i) = tr(K^i I_{p_i}) = \sum_{k=1}^{s_i} \tau_k^i \theta_k^i = tr(K^i),$$

$$\pi_0^i(\theta^i; \delta^i, I_{p_i}) = \exp\{-\frac{1}{2}\sum_{k=1}^{s_i} \theta_k^i \tau_k^i + \frac{\delta^i - 2}{2} \log|K^i(\theta^i)|\}, \qquad \text{and}$$

$$\begin{aligned} ||\theta^i||^2 &= \sum_{k=1}^{s_i} (\theta_k^i)^2 + \sum_{k=s_i+1}^{S_i} (\theta_k^i)^2 \leq \sum_{k=1}^{S_i} \tau_k^i (\theta_k^i)^2 \\ &= ||K^i(\theta^i)||_F \leq (\sum_{k=1}^{s_i} \tau_k^i \theta_k^i)^2 \qquad \text{by lemma A.1.14,} \end{aligned}$$

where $\tau_k^i = |v_k^i|$ is the number of elements in the color class v_k^i. We therefore have $||\theta^i|| \leq \sum_{k=1}^{s_i} \tau_k^i \theta_k^i$. Let H^i denote the convex cone $P_{\mathcal{G}^i}$ for short.

$$\begin{aligned} &\int_{H^i} ||\sqrt{n} J^i(\theta^i - \theta_0^i)|| \pi_0^i(\theta^i) d\theta^i \\ = &\int_{H^i} ||\sqrt{n} J^i(\theta^i - \theta_0^i)|| \exp\{-\frac{1}{2}\sum_{k=1}^{s_i} \theta_k^i \tau_k^i + \frac{\delta^i - 2}{2} \log|K^i(\theta^i)|\} d\theta^i \\ \leq &\sqrt{||F^i||}\sqrt{n} \int_{H^i} \Big[||\theta^i|| + ||\theta_0^i||\Big] \exp\{-\frac{1}{2}\sum_{k=1}^{s_i} \theta_k^i \tau_k^i + \frac{\delta^i - 2}{2} \log|K^i(\theta^i)|\} d\theta^i \\ = &\sqrt{||F^i||}\sqrt{n} ||\theta_0^i|| \int_{H^i} \exp\{-\frac{1}{2}\sum_{k=1}^{s_i} \theta_k^i \tau_k^i + \frac{\delta^i - 2}{2} \log|K^i(\theta^i)|\} d\theta^i \qquad \text{(A.13)} \\ &+\sqrt{||F^i||}\sqrt{n} \int_{H^i} ||\theta^i|| \exp\{-\frac{1}{2}\sum_{k=1}^{s_i} \theta_k^i \tau_k^i + \frac{\delta^i - 2}{2} \log|K^i(\theta^i)|\} d\theta^i. \qquad \text{(A.14)} \end{aligned}$$

By Proposition A.1.2, we have $||\theta_0^i||^2 \leq S_i \kappa_2^2$. Furthermore, according to Proposition A.1.1, we have $||F^i|| \leq \frac{1}{\kappa_1^2}$. We therefore need to find upper bounds for the integrals in (A.13) and (A.14). These two integrals are of the type $\int_{H^i} f(\theta^i)^{\alpha^i} e^{-tr(\theta^i D^i)} d\theta^i$

where $f(\theta^i)$ is a homogeneous function of order k^i. If n^i is the dimension of the space in which H^i sits, we use the result of Theorem A.1.8. Let $\bar{D}^i$ be the s_i-dimensional vector with entries $\frac{\tau_k^i}{2}$. We have

$$\int_{H^i} \exp\{-\frac{1}{2}\sum_{k=1}^{s_i}\theta_k^i\tau_k^i + \frac{\delta^i-2}{2}\log|K^i(\theta^i)|\}d\theta^i = \int_{H^i} e^{-tr(\bar{D}^i\theta^i)}|K^i(\theta^i)|^{\frac{\delta^i-2}{2}}d\theta^i,$$

and

$$\begin{aligned}&\int_{H^i} ||\theta^i|| \exp\{-\frac{1}{2}\sum_{k=1}^{s_i}\theta_k^i\tau_k^i + \frac{\delta^i-2}{2}\log|K^i(\theta^i)|\}d\theta^i\\ \leq\ &\int_{H^i}(\sum_{k=1}^{s_i}\tau_k^i\theta_k^i)e^{-tr(\bar{D}^i\theta^i)}|K^i(\theta^i)|^{\frac{\delta^i-2}{2}}d\theta^i,\end{aligned}$$

and therefore since $K^i(\theta^i)$ is homogeneous of order p_i, $\sum_{k=1}^{S_i}\tau_k^i\theta_k^i$ is homogeneous of order 1 and 1 is homogeneous of order 0, we have, for $\alpha^i = \frac{\delta^i-2}{2}$

$$\begin{aligned}&\int_{H^i} e^{-tr(\bar{D}^i\theta^i)}|K^i(\theta^i)|^{\frac{\delta^i-2}{2}}d\theta^i = \Gamma(\alpha^i p_i + S_i)\int_{H^i\cap H_1^i}|K^i(\theta^i)|^{\frac{\delta^i-2}{2}}d\theta^i,\\ &\int_{H^i}(\sum_{k=1}^{s_i}\tau_k^i\theta_k^i)e^{-tr(\bar{D}^i\theta^i)}|K^i(\theta^i)|^{\frac{\delta^i-2}{2}}d\theta^i\\ =\ &\Gamma(\alpha^i p_i + 1 + S_i)\int_{H^i\cap H_1^i}(\sum_{k=1}^{s_i}\tau_k^i\theta_k^i)|K^i(\theta^i)|^{\frac{\delta^i-2}{2}}d\theta^i. \qquad (A.15)\end{aligned}$$

However, we do not know how to compute the integrals $\int_{H^i\cap H_1^i}|K^i(\theta^i)|^{\frac{\delta^i-2}{2}}d\theta^i$ and $\int_{H^i\cap H_1^i}(\sum_{k=1}^{s_i}\tau_k^i\theta_k^i)|K^i(\theta^i)|^{\frac{\delta^i-2}{2}}d\theta^i$. The set $H_1^i = \{\theta^i \mid tr(\bar{D}^i\theta^i) = 1\}$ is $H_1^i = \{\theta^i \mid \sum_{k=1}^{s_i}\tau_k^i\theta_k^i = 2\}$. So, we only have one integral, $\int_{H^i\cap H_1^i}|K^i(\theta^i)|^{\frac{\delta^i-2}{2}}d\theta^i$, to compute. But

$$\sum_{k=1}^{s_i}\tau_k^i\theta_k^i = tr(K^i(\theta^i)I_{p_i}) = tr(K^i(\theta^i)) = \sum_{j=1}^{p_i}\lambda_j$$

where the λ_j are the eigenvalues of $K^i(\theta^i)$. Following the inequality between the arithmetic mean and the geometric mean, on $H_1^i \cap H^i$, we have

$$|K^i(\theta^i)| = (\prod_i^{p_i} \lambda_j) \leq \frac{(\sum_{j=1}^{p_i} \lambda_j)^{p_i}}{p_i^{p_i}} = \frac{2^{p_i}}{p_i^{p_i}}$$

and thus

$$\int_{H^i \cap H_1^i} |K^i(\theta^i)|^{\frac{\delta^i - 2}{2}} d\theta^i \leq \frac{2^{p_i}}{p_i^{p_i}} \int_{H^i \cap H_1^i} d\theta^i. \tag{A.16}$$

We are now going to use Theorem A.1.8 in the reverse direction with $f(\theta^i) = 1$ in order to evaluate $\int_{H^i \cap H_1^i} d\theta^i$. We have

$$\int_{H^i} e^{-tr(\bar{D}^i \theta^i)} d\theta^i = \Gamma(0 + S_i) \int_{H^i \cap H_1^i} d\theta^i \tag{A.17}$$

and we are going to majorize $\int_{H^i} e^{-tr(\bar{D}^i \theta^i)} d\theta^i$. We now use the fact that the matrices in H^i are positive definite, thus we have that, for $l = s_i + 1, \ldots, S_i$, $(\theta_l^i)^2 \leq \theta_{t_l}^i \theta_{u_l}^i$ whenever $\theta_l^i = K_{jk}^i$, $j \neq k$ and $\theta_{t_l}^i = K_{jj}^i, \theta_{u_l}^i = K_{kk}^i$. Since the cone H^i is included in the cone P^i of positive definite matrices, we have that, for $l = s_i + 1, \ldots, S_i$, $(\theta_l^i)^2 \leq \theta_{t_l}^i \theta_{u_l}^i$ whenever $\theta_l^i = K_{jk}^i$, $j \neq k$ and $\theta_{t_l}^i = K_{jj}^i, \theta_{u_l}^i = K_{kk}^i$ and thus we can write

$$\begin{aligned}
\int_{H^i} e^{-tr(\bar{D}^i \theta^i)} d\theta^i &\leq \int_0^{+\infty} \cdots \int_0^{+\infty} e^{-\sum_{r=1}^{s_i} \bar{D}_r^i \theta_r^i} \Big[\prod_{l=s_i+1}^{S_i} \int_{-\sqrt{\theta_{t_l}^i \theta_{u_l}^i}}^{\sqrt{\theta_{t_l}^i \theta_{u_l}^i}} d\theta_l^i \Big] \prod_{r=1}^{s_i} d\theta_r^i \\
&= \int_0^{+\infty} \cdots \int_0^{+\infty} e^{-\sum_{r=1}^{s_i} \bar{D}_r^i \theta_r^i} \Big[\prod_{l=s_i+1}^{S_i} 2\sqrt{\theta_{t_l}^i \theta_{u_l}^i} \Big] \prod_{r=1}^{s_i} d\theta_r^i.
\end{aligned}$$

Since we have assumed that D^i is equal to the identity, $\bar{D}^i_r = \frac{\tau^i_r}{2}$ with the τ^i_r being bounded. Then

$$\begin{aligned}
&\int_0^{+\infty} \dots \int_0^{+\infty} e^{-\sum_{r=1}^{s_i} \bar{D}^i_r \theta^i_r} \prod_{r=1}^{s_i} \Big[\prod_{l=s_i+1}^{S_i} 2\sqrt{\theta^i_{t_l} \theta^i_{u_l}} \Big] d\theta^i_r \\
= \;& 2^{S_i - s_i} \prod_{r=1}^{s_i} \int_0^{+\infty} (\theta^i_r)^{\frac{k^i_r}{2}} e^{-\frac{\tau^i_r \theta^i_r}{2}} d\theta^i_r = 2^{S_i - s_i} \prod_{r=1}^{s_i} (\frac{2}{\tau^i_r})^{\frac{k^i_r}{2}+1} \Gamma(\frac{k^i_r}{2}+1)
\end{aligned}$$

where k^i_r is the number of t_l or u_l equal to r in the i-th local model. From the majorization above, (A.17), (A.16) and (A.15) successively, we obtain the following inequalities

$$\begin{aligned}
\int_{H^i \cap H^i_1} d\theta^i &\leq \frac{1}{\Gamma(S_i)} 2^{S_i - s_i} \prod_{r=1}^{s_i} (\frac{2}{\tau^i_r})^{\frac{k^i_r}{2}+1} \Gamma(\frac{k^i_r}{2}+1), \\
\int_{H^i \cap H^i_1} |K^i(\theta^i)|^{\frac{\delta^i - 2}{2}} d\theta^i &\leq \frac{2^{S_i - s_i + p_i}}{p_i^{p_i} \Gamma(S_i)} \prod_{r=1}^{s_i} (\frac{2}{\tau^i_r})^{\frac{k^i_r}{2}+1} \Gamma(\frac{k^i_r}{2}+1), \\
\int_{H^i} (\sum_{k=1}^{S_i} \tau^i_k \theta^i_k) e^{-tr(\bar{D}^i \theta^i)} |K^i(\theta^i)|^{\frac{\delta^i - 2}{2}} d\theta^i &\leq \frac{2^{S_i - s_i + p_i + 1} \Gamma(\alpha^i p_i + 1 + S_i)}{p_i^{p_i} \Gamma(S_i)} \prod_{r=1}^{s_i} (\frac{2}{\tau^i_r})^{\frac{k^i_r}{2}+1} \Gamma(\frac{k^i_r}{2}+1), \\
\text{and} \quad \int_{H^i} e^{-tr(\bar{D}^i \theta^i)} |K^i(\theta^i)|^{\frac{\delta^i - 2}{2}} d\theta^i &\leq \frac{2^{S_i - s_i + p_i} \Gamma(\alpha^i p_i + S_i)}{p_i^{p_i} \Gamma(S_i)} \prod_{r=1}^{s_i} (\frac{2}{\tau^i_r})^{\frac{k^i_r}{2}+1} \Gamma(\frac{k^i_r}{2}+1).
\end{aligned}$$

It follows that

$$\begin{aligned}
& \int ||\sqrt{n} J(\theta^i - \theta^i_0)|| \pi^i_0(\theta^i) d\theta^i \\
\leq \;& n^{\frac{1}{2}} \frac{1}{\kappa_1} \frac{2^{S_i - s_i + p_i}}{p_i^{p_i} \Gamma(S_i)} [\prod_{r=1}^{s_i} (\frac{2}{\tau^i_r})^{\frac{k^i_r}{2}+1} \Gamma(\frac{k^i_r}{2}+1)] (M_0 p_i \Gamma(\alpha^i p_i + S_i) + 2\Gamma(\alpha^i p_i + 1 + S_i)) \\
\leq \;& n^{\frac{1}{2}} \frac{1}{\kappa_1} \frac{2^{S_i - s_i + p_i}}{p_i^{p_i} \Gamma(S_i)} [\prod_{r=1}^{s_i} (\frac{2}{\tau^i_r})^{\frac{k^i_r}{2}+1} \Gamma(\frac{k^i_r}{2}+1)] M_2 p_i \Gamma(\alpha^i p_i + 1 + S_i),
\end{aligned}$$

where M_0 and M_2 are constants. Therefore,

$$\begin{aligned}
& \log\Big\{\int ||\sqrt{n}J^i(\theta^i-\theta_0^i)||\pi_0^i(\theta^i)d\theta^i\Big\} \\
\leq\ & \frac{1}{2}\log n - \log\kappa_1 + (S_i - s_i + p_i)\log 2 - p_i\log p_i - \log\Gamma(S_i) + \log M_2 + \log p_i \\
& + \log\Gamma(\alpha^i p_i + 1 + S_i) + \sum_{r=1}^{s_i}[(\frac{k_r^i}{2}+1)\log\frac{2}{\tau_r^i} + \log\Gamma(\frac{k_r^i}{2}+1)].
\end{aligned}$$

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

$$\begin{aligned}
& \exp[\log\Big\{\int ||\sqrt{n}J^i(\theta^i-\theta_0^i)||\pi_0^i(\theta^i)d\theta^i\Big\}] \\
\leq\ & \exp[\frac{1}{2}\log p + (S_i - s_i + p_i)\log 2 + \log p_i + \log\Gamma(\alpha^i p_i + 1 + S_i) + p_i(\frac{p_i}{2}+1)\log 2 \\
& + p_i\log\Gamma(\frac{p_i}{2}+1) + M_3],
\end{aligned}$$

where M_3 is a constant. 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

$$\log\Gamma(\alpha^i p_i + 1 + S_i) \leq \log\Gamma(\alpha^i p_i + 1 + \frac{p_i(p_i+1)}{2}) \leq \log(2\alpha^i p_i + p_i^2)! \leq M_5 p_i^2\log p_i$$

and

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

Combining all results above, we obtain that

$$\begin{aligned} & \exp\Big(\log\Big\{\int ||\sqrt{n}J^i(\theta^i-\theta_0^i)||\pi_0^i(\theta^i)d\theta^i\Big\}\Big) \\ \leq\ & \exp\Big(\frac{1}{2}\log p+[\frac{p_i(p_i+1)}{2}+p_i]\log 2+\log p_i+M_5p_i^2\log p_i+p_i(\frac{p_i}{2}+1)\log 2 \\ & +M_6p_i^2\log p_i+M_3\Big)\leq \exp[M_7p^2\log p], \end{aligned}$$

where M_7 is a constant. ■