Electronic Journal of Statistics
Vol. 14 (2020) 580–604
ISSN: 1935-7524
https://doi.org/10.1214/19-EJS1669

On the Letac-Massam conjecture and existence of high dimensional Bayes estimators for graphical models

Emanuel Ben-David

US Census Bureau, Center for Statistical Research and Methodology
e-mail: e.h.bendavid@gmail.com

and

Bala Rajaratnam

UC Davis, Department of Statistics
e-mail: brajaratnam01@gmail.com

Abstract: The Wishart distribution defined on the open cone of positive-definite matrices plays a central role in multivariate analysis and multivariate distribution theory. Its domain of parameters is often referred to as the Gindikin set. In recent years, varieties of useful extensions of the Wishart distribution have been proposed in the literature for the purposes of studying Markov random fields and graphical models. In particular, generalizations of the Wishart distribution, referred to as Type I and Type II (graphical) Wishart distributions introduced by Letac and Massam in *Annals of Statistics* (2007) play important roles in both frequentist and Bayesian inference for Gaussian graphical models. These distributions have been especially useful in high-dimensional settings due to the flexibility offered by their multiple-shape parameters. Concerning Type I and Type II Wishart distributions, a conjecture of Letac and Massam concerns the domain of multiple-shape parameters of these distributions. The conjecture also has implications for the existence of Bayes estimators corresponding to these high dimensional priors. The conjecture, which was first posed in the Annals of Statistics, has now been an open problem for about 10 years. In this paper, we give a necessary condition for the Letac and Massam conjecture to hold. More precisely, we prove that if the Letac and Massam conjecture holds on a decomposable graph, then no two separators of the graph can be nested within each other. For this, we analyze Type I and Type II Wishart distributions on appropriate Markov equivalent perfect DAG models and succeed in deriving the aforementioned necessary condition. This condition in particular identifies a class of counterexamples to the conjecture.

MSC 2010 subject classifications: Primary 62-09, 60E05; secondary 62F99.

Keywords and phrases: Graphical models, Wishart distribution, Gaussian distribution.

Received February 2019.

1. Introduction

Inference for graphical models is an important topic of contemporary interest [20], and in this regard, various tools for inference have been proposed in the statistics literature, including establishing sufficient and/or necessary conditions for existence of high dimensional estimators. One important contribution in the area are the families of Type I and Type II Wishart distributions introduced by Letac and Massam (LM, henceforth) [19]. Type II Wishart distributions of Letac-Massam have the distinct advantage of being conjugate priors for the scale parameter of Gaussian graphical models and have the strong hyper Markov property (see Appendix A.2 for the definition). Type I Wishart distributions are weak hyper Markov, a property parallel to the weak hyper Markov property of the hyper Wishart distribution [19]. Both Type I and II Wishart distributions have multiple-shape parameters, in contrast with the classical Wishart distribution which has just one shape parameter that is restricted to the one dimensional Gindikin set: $\Delta = \left\{\frac{1}{2}, \frac{2}{2}, \frac{3}{2}, \dots, \frac{p-1}{2}\right\} \cup \left(\frac{p-1}{2}, +\infty\right)$. The LM conjecture is closely related to generalizations of the Gindikin set and the Gindikin conjecture on the parameter set for Riesz distributions. A complete description of positive Riesz distributions on homogeneous cones was given in [14].

These multiple-shape parameter Wishart distributions are useful for flexible high dimensional inference [21], and have been used for objective Bayesian model selection in Gaussian graphical models [4]. Since the domain of parameters of these high dimensional priors is not fully identified, it is not clear when these distributions yield well-defined and proper priors. The Letac and Massam conjecture (the LM conjecture, henceforth) aims to address this question formally. The LM conjecture is critical for understanding when these priors lead to well-defined Bayes estimators, since the existence of such estimators is not always guaranteed in high dimensional settings with the sample size $n \ll p$. This is due to the fact that, for $n < p$, being a proper posterior probability is determined by specific choice of parameters (see [19, Corollary 4.1]). In this sense resolving the LM conjecture can be viewed as a Bayesian analogue of the frequentist problem of identifying sufficient and necessary conditions for the existence of the maximum likelihood estimator for Gaussian graphical models. The LM conjecture is also relevant to statistical applications of these Wishart distributions as prior distributions. A reason is that in simulations and MCMC methods our knowledge of the parameter space gives us leverage in exploring the space and selecting the desirable parameters. Our main contribution in this paper is to give a necessary condition for the LM conjecture to hold and consequently identify a large subclass of decomposable graphs on which the LM conjecture does not hold.

In what follows, we shall employ the notation introduced in the work of LM [19]. The Type I and Type II Wishart distributions are defined on two cones. These cones, respectively denoted by $\mathrm{Q}_{\mathcal{G}}$ and $\mathrm{P}_{\mathcal{G}}$, correspond to spaces of covariance and inverse-covariance matrices of the Gaussian graphical model over a decomposable graph $\mathcal{G}$, that is, an undirected graph that has no induced cycle of length greater than or equal to four. The cone $\mathrm{P}_{\mathcal{G}}$, a sub-cone of positive definite matrices, is simply the space of inverse-covariance matrices and $\mathrm{Q}_{\mathcal{G}}$ is the space of incomplete covariance matrices. Incomplete, in the sense that only entries along the edges (including the loops) of $\mathcal{G}$ are specified and the rest are unspecified. When $\mathcal{G}$ is complete, that is, for the saturated Gaussian model where no conditional independence assumption is imposed, Type I and Type II Wishart distributions are identical to the classical Wishart and inverse-Wishart distributions. Also, by restricting the multiple-shape parameter to a specific one dimensional space, these distributions reduce to the hyper Wishart distribution introduced by Dawid and Lauritzen [7] and the G-Wishart distribution defined by Roverato in [22] respectively (see [19] for more details).

By having multiple-shape parameter, Type I and Type II Wishart distributions are more flexible than aforementioned prior distributions, but there is a trade-off: for general decomposable graphs, except for homogeneous graphs and paths, the (multiple-shape) parameter spaces are not fully identified. In case of a non-homogeneous decomposable graph $\mathcal{G}$, LM in [19] attempt to identify the parameter spaces by first fixing a perfect order $\mathcal{P}$ of $\mathcal{G}$ and identify two corresponding subsets $A_{\mathcal{P}}$ and $\mathcal{B}_{\mathcal{P}}$ of the corresponding parameter spaces. By taking the union of $A_{\mathcal{P}}$ and $\mathcal{B}_{\mathcal{P}}$ over all perfect orders of $\mathcal{G}$, they conjecture that they can fully cover, and thus identify, the parameter spaces. They prove that the conjecture holds when $\mathcal{G}$ is the path A_4, that is, $v_1 - v_2 - v_3 - v_4$ (this is the simplest non-homogeneous decomposable graph). More recently, Graczyk et al. in [11] prove that the conjecture also holds for Type I Wishart distribution on Q_{A_n}, the cone of partial positive definite matrices on paths of length $n \geq 4$.

In this paper, we approach the LM conjecture in an appropriate perfect DAG setting. To this end, we first transform the Type I and Type II Wishart distributions to distributions that correspond to appropriate perfect DAG models. By analyzing the LM conjecture on the corresponding domains we derive a necessary condition for the the LM conjecture to hold. This in return leads to identifying a large subclass of decomposable graphs that are counterexamples to the LM conjecture.

The organization of the paper is as follows. In §2 we recall some basic definitions and concepts for graphical models, in general, and Gaussian graphical models, in particular. In §3 we provide the reader with definition of Type I and Type II Wishart distributions and formally state the LM conjecture. In §4 we prove the main result of the paper, Theorem 4.1, in which we identify a class of counterexamples to the LM conjecture, via a simple constructive procedure. The proof of this theorem is accomplished after introducing some new notation and concepts and proving several key lemmas.

2. Preliminaries

2.1. Matrix notation

For a finite set V, let $|V|$ denote the cardinality of V. Let $\mathbb{R}^V$ and $\mathbb{R}^{V\times V}$ denote respectively the linear spaces of $|V|$-dimensional vectors $x = (x_i|\, i \in V)$ and

$|V| \times |V|$ real matrices $A = (A_{ij})_{i,j \in V}$. The spaces of $|V| \times |V|$ symmetric and positive definite matrices are respectively denoted by $\mathrm{S}_V(\mathbb{R})$ and $\mathrm{PD}_V(\mathbb{R})$. When $V = \{1, \dots, p\}$, the aforementioned spaces are denoted by $\mathbb{R}^p$, $\mathbb{R}^{p\times p}$, $\mathrm{S}_p(\mathbb{R})$ and $\mathrm{PD}_p(\mathbb{R})$. A positive definite matrix is sometimes denoted by $\Sigma \succ 0$. For $a, b \subseteq V$, let x_a denote the subvector $(x_i |\, i \in a)$ and let A_{ab} denote the $|a| \times |b|$ submatrix $(A_{ij})_{i\in a, j\in b} \in \mathbb{R}^{a\times b}$. For simplicity, A_{aa} is often denoted by A_a. When $b = V \setminus a$, the Schur complement of A_a is defined as $A_{bb|a} = A_b - A_{ba}(A_a)^{-1}A_{ab}$, assuming that A_a in invertible.

2.2. Graph theoretic notation and terminology

We now introduce some preliminaries on graph theory and graphical models. This section closely follows the notation and exposition given in [3]. A graph $\mathcal{G}$ is a pair of objects (V, E), where V and E are two disjoint finite sets representing, respectively, the vertices and the edges of $\mathcal{G}$. An edge $e \in E$ is said to be undirected if e is an unordered pair $\{i, j\}$, or directed if e is an ordered pair (i, j) for some $i, j \in V$. A graph is said to be undirected (directed) if its edges are all undirected (directed). A directed edge $(i, j) \in E$ is denoted by $i \to j$. When $i \to j$ and $i \neq j$ we say that i is a parent of j, and j is a child of i. The set of parents of i is denoted by $\mathrm{pa}\,(i)$, and the set of children of i is denoted by $\mathrm{ch}\,(i)$. The family of i is $\mathrm{fa}\,(i) = \mathrm{pa}\,(i) \cup \{i\}$. For an undirected edge $\{i, j\} \in E$ the vertex i is said to be a neighbor of j, or j a neighbor of i, if $i \neq j$. The set of all neighbors of i is denoted by $\mathrm{ne}\,(i)$. In general two distinct vertices are said to be adjacent, denoted by $i \sim j$, if there exists either a directed or an undirected edge between them. A loop in $\mathcal{G}$ is an ordered pair (i, i), or an unordered pair $\{i, i\}$ in E. For ease of notation, in this paper, we shall always assume that the edge set of each graph contains all the loops; however, we draw the graph without the loops.

A graph $\mathcal{G}' = (V', E')$ is (more precisely, induced) subgraph of $\mathcal{G} = (V, E)$, if $V' \subseteq V$ and $E' = V' \times V' \cap E$. For a set $A \subseteq V$, the subgraph $\mathcal{G}_A = (A, A \times A \cap E)$ is said to be the graph induced by A. A graph $\mathcal{G}$ is said to be complete if every pair of vertices are adjacent. A subset $A \subseteq V$ is said to be a clique if the induced subgraph $\mathcal{G}_A$ is complete and is not contained in any other complete subgraphs of $\mathcal{G}$. A path in $\mathcal{G}$ of length $n \geq 1$ from a vertex i to a vertex j is a finite sequence of distinct vertices $i_0 = i, \dots, i_n = j$ in V such that (i_{t-1}, i_t) or $\{i_{t-1}, i_t\}$ are in E for each $t = 1, \dots, n$. A path is said to be directed if at least one of the edges is directed. We say that i leads to j, denoted by $i \to \cdots \to j$, if there is a directed path from i to j. A graph $\mathcal{G} = (V, E)$ is said to be connected if for any pair of distinct vertices $i, j \in V$ there exists a path between them. An n-cycle in $\mathcal{G}$ is a path of length $n \geq 3$ with the additional requirement that the end points are identical. A directed n-cycle is defined accordingly. A graph is acyclic if it does not have any cycles. An acyclic directed graph, denoted by DAG, is a directed graph with no directed cycles.

Notation. Henceforth in this paper, we denote an undirected graph by $\mathcal{G} = (V, E)$ and a DAG by $\mathcal{D} = (V, F)$. In addition, otherwise stated, we always assume that the vertex set V is $\{1, 2, \ldots, p\}$.

The undirected version of a DAG $\mathcal{D} = (V, F)$, denoted by $\mathcal{D}^{\mathrm{u}} = (V, F^{\mathrm{u}})$, is the undirected graph obtained by replacing all the directed edges of $\mathcal{D}$ by undirected ones. An immorality in $\mathcal{D}$ is an induced subgraph of the form $i \to j \leftarrow k$. Moralizing an immorality entails adding an undirected edge between the pair of parents that have the same children. The moral graph of $\mathcal{D}$, denoted by $\mathcal{D}^m = (V, F^m)$, is the undirected graph obtained by first moralizing each immorality of $\mathcal{D}$ and then making the undirected version of the resulting graph.

Given a DAG, the set of ancestors of a vertex i, denoted by $\mathrm{an}\,(i)$, is the set of those vertices j such that $j \to \cdots \to i$. Similarly, the set of descendants of a vertex i, denoted by $\mathrm{de}\,(i)$, is the set of those vertices k such that $i \to \cdots \to k$. The set of non-descendants of i is $\mathrm{nd}\,(i) = V \setminus (\mathrm{de}\,(i) \cup \{i\})$. A set $A \subseteq V$ is said to be ancestral when A contains the parents of its members. The smallest ancestral set containing a set $B \subseteq V$ is denoted by $\mathrm{An}\,(B)$.

2.3. Decomposable graphs, homogeneous graphs and perfect DAGs

Let $\mathcal{G} = (V, E)$ be a decomposable graph. Here we employ some common notations and definitions from Lauritzen [18]. The order $(C_1, \ldots, C_r)$ of cliques of $\mathcal{G}$ is said to be perfect if, for every $t > 1$, $S_t = (C_1 \cup \cdots \cup C_{t-1}) \cap C_t$ is a separator. Every decomposable graph admits a perfect order of its cliques. Let $(C_1, \ldots, C_r)$ be one such perfect order (of the cliques) of $\mathcal{G}$. With this perfect order we associate:

- the histories $H_t = C_1 \cup C_2 \cup \cdots \cup C_t$, for $t = 1, \ldots, r$,
- the separators $S_t = H_{t-1} \cap C_t$, for $t = 2, \ldots, r$,
- the residuals $R_1 = C_1 \setminus S_2$ and $R_t = C_t \setminus H_{t-1}$ for $t = 2, \ldots, r$.

Let $\mathscr{C}$ denote the set of cliques and let $\mathscr{S}$ denote the set of separators of a decomposable graph $\mathcal{G}$. Let $r' \leq r - 1$ denote the number of distinct separators and $\nu\,(S)$ denote the multiplicity of S, that is, the number of t such that $S_t = S$.

A graph $\mathcal{G}$ is said to be homogeneous if it is decomposable and does not contain the path A_4, as an induced subgraph. Alternatively, a graph $\mathcal{G}$ is homogeneous if and only if for any two adjacent vertices i, j we have

$$\mathrm{cl}(j) \subseteq \mathrm{cl}(i) \ \text{ or } \ \mathrm{cl}(i) \subseteq \mathrm{cl}(j), \tag{2.1}$$

where, $\mathrm{cl}(j) = \mathrm{ne}\,(j) \cup \{j\}$ is the closure of j (see [19, Theorem 2.2]). The reader is referred to [19, 15, 14] for all the common notions and properties of homogeneous graphs.

Decomposable graphs have a deep connection to perfect DAGs. A DAG $\mathcal{D} = (V, F)$ is perfect if it has no immoralities, that is, the parents of all vertices are adjacent, or equivalently if the set of parents of each vertex induces a complete subgraph of $\mathcal{D}$. If $\mathcal{G} = (V, E)$ is decomposable, then there exists a DAG version

$\mathcal{D} = (V, F)$ of $\mathcal{G}$ that is perfect. On the other hand, the undirected version of a perfect DAG is necessarily decomposable [18].

Remark 2.1. In mathematical graph theory, homogeneous graphs are also known as cographs. These graphs can be recognized in linear time algorithms [5, 6].

2.4. Gaussian graphical models

Let $\mathcal{G} = (V, E)$ be an undirected graph with the vertex set $V = \{1, \ldots, p\}$. The Gaussian graphical model over $\mathcal{G}$, denoted by $\mathscr{N}(\mathcal{G})$, is the family of p-variate (non-singular) Gaussian distributions $N_p(0, \Sigma)$ that are Markov with respect to $\mathcal{G}$. A simple characterization of $N_p(0, \Sigma) \in \mathscr{N}(\mathcal{G})$ is that $\left(\Sigma^{-1}\right)_{ij} = 0$ whenever $\{i, j\} \notin E$ (a simple proof of this well-known fact is found in [18, section 5.1]). Let $\Omega = \Sigma^{-1}$ denote the inverse-covariance matrix (also said to be the precision or concentration matrix). Let $\Omega \succ 0$ denote that Ω is positive definite. The space of inverse-covariance matrices for the Gaussian distributions in $\mathscr{N}(\mathcal{G})$ is $\mathrm{P}_{\mathcal{G}} = \{\Omega \succ 0 : \Omega_{ij} = 0 \text{ whenever } \{i, j\} \notin E\}$. The space of covariance matrices, denoted by $\mathrm{PD}_{\mathcal{G}}$, is the inverse of the elements of $\mathrm{P}_{\mathcal{G}}$. It is thus natural to parametrize $\mathscr{N}(\mathcal{G})$ over $\mathrm{P}_{\mathcal{G}}$ or $\mathrm{PD}_{\mathcal{G}}$. The space of covariance matrices $\mathrm{PD}_{\mathcal{G}}$ has a complicated structure, but it can be identified with a simpler space, specifically, its image, denoted by $\mathrm{Q}_{\mathcal{G}}$, under the projection mapping $\Sigma \mapsto \Sigma_E := (\Sigma_{ij} : \{i, j\} \in E)$. By the Grone's result in [12], an element $(A_{ij} : \{i, j\} \in E) \in \mathrm{Q}_{\mathcal{G}}$ is described as a partial matrix (or an incomplete matrix), where the entries along the edge set E are specified and the rest are unspecified, with the property that for each clique C of $\mathcal{G}$ the $|C| \times |C|$ matrix $A_C = (A_{ij} : i, j \in C)$ is positive definite. Under this specification, each element of $\mathrm{Q}_{\mathcal{G}}$ can be completed to a unique positive definite matrix in $\mathrm{PD}_{\mathcal{G}}$. Grone et al. [12, Theorem 4] explicitly provide the bijective mapping $(\Sigma_E \mapsto \Sigma) : \mathrm{Q}_{\mathcal{G}} \to \mathrm{PD}_{\mathcal{G}}$. Now by composing this mapping with the mapping $\left(\Sigma \mapsto \Sigma^{-1}\right) : \mathrm{PD}_{\mathcal{G}} \to \mathrm{P}_{\mathcal{G}}$, we obtain the bijective mapping $\left(\Sigma_E \mapsto \Sigma^{-1}\right) : \mathrm{Q}_{\mathcal{G}} \to \mathrm{P}_{\mathcal{G}}$. The inverse of this mapping is $\left(\Omega \mapsto \Omega_E^{-1}\right) : \mathrm{P}_{\mathcal{G}} \to \mathrm{Q}_{\mathcal{G}}$, using the notation Ω_E^{-1} for $\left(\Omega^{-1}\right)_E$. We emphasize that these mappings are explicit when $\mathcal{G}$ is decomposable. We shall frequently invoke these mappings in subsequent sections.

3. The Letac-Massam Wishart type distributions for decomposable graphs

3.1. Markov ratios and corresponding measures on $\mathrm{Q}_{\mathcal{G}}$ and $\mathrm{P}_{\mathcal{G}}$

Henceforth in this paper, we assume that $\mathcal{G} = (V, E)$ is a decomposable graph and the vertices are labeled $1, 2, \ldots, p$. The primary goal of this section is to provide the reader with an overview of the families of Wishart-Type I and Wishart-Type II distributions introduced in [19]. At the end of this section, we shall formally state the LM conjecture concerning the domain of multiple-shape parameters of these distributions.

Let $C_1, \ldots, C_r$ be a perfect order of $\mathcal{G}$ and let $(S_2, \ldots, S_r)$ be the corresponding sequence of separators, with possible repetitions. For each $\alpha \in \mathbb{R}^r,\ \beta \in \mathbb{R}^{r-1}$ and $\Sigma_E \in \mathrm{Q}_{\mathcal{G}}$, the Markov ratio $H_{\mathcal{G}}(\alpha, \beta, \Sigma_E)$ is defined as follows:

$$H_{\mathcal{G}}(\alpha, \beta, \Sigma_E) = \frac{\prod_{t=1}^{r} \det(\Sigma_{C_t})^{\alpha_t}}{\prod_{t=2}^{r} \det(\Sigma_{S_t})^{\beta_t}}, \tag{3.1}$$

where $\beta_i = \beta_j$ whenever $S_i = S_j$, so essentially (α, β) are restricted to an $(r + r')$-dimensional subspace of $\mathbb{R}^r \times \mathbb{R}^{r-1}$, where r' is the number of distinct separators.

Let $c = (c_1, \ldots, c_r)$ and $s = (s_2, \ldots, s_r)$ where $c_t = |C_t|$ and $s_t = |S_t|$, respectively. Moreover, let $d\Sigma_E$ denote Lebesgue measure on $\mathrm{Q}_{\mathcal{G}}$. Then

$$\mu_{\mathcal{G}}(d\Sigma_E) = H_{\mathcal{G}}(-(c+1)/2, -(s+1)/2, \Sigma_E)\, d\Sigma_E \tag{3.2}$$

is a measure on $\mathrm{Q}_{\mathcal{G}}$. The image of $\mu_{\mathcal{G}}$ under the mapping $\Sigma_E \mapsto \Sigma^{-1} : \mathrm{Q}_{\mathcal{G}} \to \mathrm{P}_{\mathcal{G}}$ is a measure on $\mathrm{P}_{\mathcal{G}}$ given by

$$\nu_{\mathcal{G}}(d\Omega) = H_{\mathcal{G}}\left((c+1)/2, (s+1)/2, \Omega_E^{-1}\right) d\Omega, \tag{3.3}$$

where $d\Omega$ is Lebesgue measure on $\mathrm{P}_{\mathcal{G}}$ [19].

3.2. Type I & Type II Wishart distributions

The Type I and Type II Wishart distributions were introduced in [19]. The Type I Wishart distribution is defined on the cone $\mathrm{Q}_{\mathcal{G}}$. The non-normalized density of this distribution is given by

$$\omega_{\mathrm{Q}_{\mathcal{G}}}(\alpha, \beta, U_E, d\Sigma_E) = \exp\left\{-\mathrm{tr}\left(\Sigma U^{-1}\right)\right\} H_{\mathcal{G}}(\alpha, \beta, \Sigma_E)\, \mu_{\mathcal{G}}(d\Sigma_E),$$

where $(\alpha, \beta) \in \mathbb{R}^r \times \mathbb{R}^{r-1}$ denotes the multiple-shape parameter and $U_E \in \mathrm{Q}_{\mathcal{G}}$ is the scale parameter. Let $\mathcal{A}$ denote the set of (α, β) such that for every $U_E \in \mathrm{Q}_{\mathcal{G}}$

$$\int_{\mathrm{Q}_{\mathcal{G}}} \omega_{\mathrm{Q}_{\mathcal{G}}}(\alpha, \beta, U_E, d\Sigma_E) < \infty \ \text{ and} \tag{A1}$$

$$\int_{\mathrm{Q}_{\mathcal{G}}} \omega_{\mathrm{Q}_{\mathcal{G}}}(\alpha, \beta, U_E, d\Sigma_E) / H_{\mathcal{G}}(\alpha, \beta, U_E) \quad \text{is functionally independent of } U_E. \tag{A2}$$

The normalized version of $\omega_{\mathrm{Q}_{\mathcal{G}}}$, denoted by $W_{\mathrm{Q}_{\mathcal{G}}}$, is then defined for $(\alpha, \beta) \in \mathcal{A}$. In similar fashion, the Type II Wishart distribution is defined on the cone $\mathrm{P}_{\mathcal{G}}$ with the non-normalized density

$$\omega_{\mathrm{P}_{\mathcal{G}}}(\alpha, \beta, U_E, d\Omega) = \exp\left\{-\mathrm{tr}(\Omega U)\right\} H_{\mathcal{G}}\left(\alpha, \beta, \Omega_E^{-1}\right) \nu_{\mathcal{G}}(d\Omega).$$

Let $\mathcal{B}$ be the set of (α, β) such that for every $U_E \in \mathrm{Q}_{\mathcal{G}}$

$$\int_{\mathrm{P}_{\mathcal{G}}} \omega_{\mathrm{P}_{\mathcal{G}}}(\alpha, \beta, U_E, d\Omega) < \infty \ \text{ and} \tag{B1}$$

$$\int_{\mathrm{P}_{\mathcal{G}}} \omega_{\mathrm{P}_{\mathcal{G}}}\left(\alpha, \beta, U_E, d\Omega\right) / H_{\mathcal{G}}\left(\alpha, \beta, U_E\right) \text{ is functionally independent of } U_E. \tag{B2}$$

The normalized version of $\omega_{\mathrm{P}_{\mathcal{G}}}$, denoted by $W_{\mathrm{P}_{\mathcal{G}}}$, is defined for $(\alpha, \beta) \in \mathcal{B}$.

3.3. The LM conjecture for identifying $\mathcal{A}$ and $\mathcal{B}$

An important goal of LM in [19] after defining the Type I and II Wishart distributions is to identify their parameter spaces $\mathcal{A}$ and $\mathcal{B}$, respectively. When the underlying graph $\mathcal{G}$ is homogeneous both $\mathcal{A}$ and $\mathcal{B}$ are fully identified in [19] (see Appendix A.3), but when $\mathcal{G}$ is not homogeneous, these spaces are only partially identified. More precisely, LM in [19] identify a subset of $\mathcal{A}$ and a subset of $\mathcal{B}$ as follows.

Let $\mathcal{P} = (C_1, \cdots, C_r)$ be a given perfect order of $\mathcal{G}$ and $(S_2, \cdots, S_r)$ the corresponding sequence of separators. For each separator $S \in \mathscr{S}$ let $J(\mathcal{P}, S) = \{t : S_t = S\}$. A set associated with $\mathcal{P}$ and $\mathcal{A}$, denoted by $A_{\mathcal{P}}$, is the set of $(\alpha, \beta) \in \mathbb{R}^r \times \mathbb{R}^{r-1}$ such that:

$$\sum_{t \in J(\mathcal{P}, S)} \alpha_t - \nu(s)\, \beta(S) = 0, \tag{A1}$$

for each $S \neq S_2$ (where $\nu(S)$ is the multiplicity of S);

$$\alpha_t - (c_t - 1)/2 > 0, \text{ for each } t = 2, \ldots, r; \tag{A2}$$

$$\alpha_1 - \delta_2 > (s_2 - 1)/2, \quad \text{where } \delta_2 = \sum_{t \in J(\mathcal{P}, S_2)} \alpha_t - \nu(S_2)\, \beta_2. \tag{A3}$$

Similarly, a set associated with $\mathcal{P}$ and $\mathcal{B}$, denoted by $B_{\mathcal{P}}$, is the set of (α, β) such that:

$$\sum_{t \in J(\mathcal{P}, S)} (\alpha_t + (c_t - s_t)/2) - \nu(S)\, \beta(S) = 0, \text{ for each } S \neq S_2; \tag{B1}$$

$$-\alpha_t - (c_t - s_t - 1)/2 > 0, \tag{B2}$$

for each $t = 2, \ldots, r$ and $-\alpha_1 - (c_1 - s_2 - 1)/2 > 0$;

$$-\alpha_1 - (c_1 - s_2 + 1)/2 - \eta_2 > (s_2 - 1)/2, \tag{B3}$$

$$\text{where } \eta_2 = \sum_{t \in J(\mathcal{P}, S_2)} (\alpha_t + (c_t - s_2)/2) - \nu(S_2)\, \beta_2.$$

Theorems 3.3 and Theorem 3.4 in [19] prove that if $\mathcal{G}$ is a non-complete decomposable graph, then $A_{\mathcal{P}} \subseteq \mathcal{A}$ and $B_{\mathcal{P}} \subseteq \mathcal{B}$. Therefore, $\bigcup_{\mathcal{P}} A_{\mathcal{P}} \subseteq \mathcal{A}$ and $\bigcup_{\mathcal{P}} B_{\mathcal{P}} \subseteq \mathcal{B}$, where the subscript $\mathcal{P}$ runs through all perfect orders of $\mathcal{G}$. Except in case of homogeneous graphs in which $\bigcup_{\mathcal{P}} A_{\mathcal{P}} \neq \mathcal{A}$ and $\bigcup_{\mathcal{P}} B_{\mathcal{P}} \neq \mathcal{B}$, LM in [19] conjecture that for non-homogeneous decomposable graphs these are equalities. Their detailed computations [19, Corollary 3.1] show that the conjecture holds on the path A_4. More recently Graczyk et al. in [11] show that for every path

A_n of length $n \geq 4$, the conjecture holds for Type I Wishart distributions (that is, Part (I) below when $\mathcal{G} = A_n$).

The LM Conjecture. Let $\mathcal{G}$ be a non-homogeneous decomposable graph and let $\mathrm{Ord}(\mathcal{G})$ denote the set of perfect orders of $\mathcal{G}$. Then

$$\bigcup_{\mathcal{P} \in \mathrm{Ord}(\mathcal{G})} A_{\mathcal{P}} = \mathcal{A}, \tag{I}$$

$$\bigcup_{\mathcal{P} \in \mathrm{Ord}(\mathcal{G})} B_{\mathcal{P}} = \mathcal{B}. \tag{II}$$

Remark 3.1. For each perfect order $\mathcal{P} = (C_1, \ldots, C_r)$ of the cliques of a decomposable graph $\mathcal{G}$, the sets $A_{\mathcal{P}}$ and $B_{\mathcal{P}}$, as manifolds, are of dimension $r + 1$. Therefore, the LM conjecture does not hold if we can show that there is a non-homogeneous decomposable graph $\mathcal{G}$ such that both $\mathcal{A}$ and $\mathcal{B}$ contain a manifold of dimension greater than $r + 1$.

4. A general class of counterexamples to the LM conjecture

4.1. The main theorem

In this section, we prove the main result of this paper that produces a class of counterexamples to the LM conjecture. The main idea is to express $W_{\mathrm{Q}_{\mathcal{G}}}$ and $W_{\mathrm{P}_{\mathcal{G}}}$ as distributions with respect to a particular perfect DAG version of $\mathcal{G}$. This turns out to be very useful for analyzing $\mathcal{A}$ and $\mathcal{B}$, the spaces of multiple-shape parameters for Type I and II Wishart distributions. Before we state the theorem, we introduce the following notation.

Notation 4.1. *Let $\mathcal{G}$ be a decomposable graph and let $\mathcal{D}$ be a DAG version of it. Then $\mathscr{S}^{\mathcal{D}}$ denotes the set of all separators of $\mathcal{G}$ which are ancestral in $\mathcal{D}$, and $r_{\mathcal{D}}$ denotes the size of $\mathscr{S}^{\mathcal{D}}$.*

Theorem 4.1. *Suppose $\mathcal{G}$ is a decomposable graph with two nested separators, that is, one separator contained in another separator. Then there exists a perfect DAG version $\mathcal{D}$ of $\mathcal{G}$ such that $r_{\mathcal{D}} \geq 2$ and both $\mathcal{A}$ and $\mathcal{B}$ contain a manifold of dimension greater than or equal to $r + r_{\mathcal{D}}$. If, in addition, $\mathcal{G}$ is not homogeneous, then the LM conjecture fails on $\mathcal{G}$.*

Theorem 4.1 now provides a convenient mechanism for constructing graphs that are counterexamples to the LM conjecture. The graphs given by Figure 1(a) and Figure 1(b) are two examples of such graphs. In Figure 1(a), the nested separators are $\{2\}$ and $\{2, 5\}$ and in Figure 1(b) the nested separators are $\{4\} \subset \{3, 4\}$ and $\{5\} \subset \{5, 6\}$. In Appendix A.4 we show how in general one can construct counterexamples from any homogeneous graph $\mathcal{H}$ that has two or more separators.

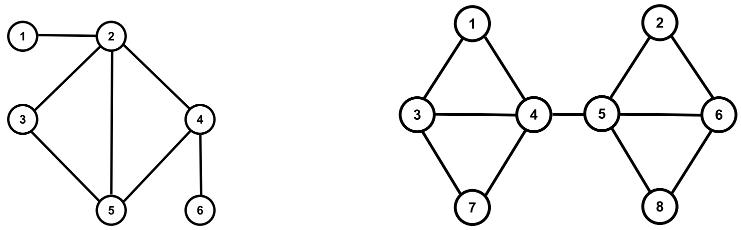


FIG 1. *Two counterexamples to the LM conjecture.*

Remark 4.1. Recall that the Markov ratio $H_{\mathcal{G}}(\alpha, \beta, \Sigma_E)$ in Equation (3.1) is defined for (α, β) in an $(r+r')$-dimensional subspace of $\mathbb{R}^r \times \mathbb{R}^{r-1}$, thus dimension of $\mathcal{A}$ and $\mathcal{B}$ are at most $r+r'$. In Lemma 4.1 we prove that for any decomposable graph $\mathcal{G}$ there exists an induced DAG version $\mathcal{D}$ such that $r_{\mathcal{D}} \geq 1$. Therefore, Theorem 4.1 implies that dimension of $\mathcal{A}$ and $\mathcal{B}$ is greater than or equal to $r+1$, which are also implied by Theorem 3.3 and Theorem 3.4 in [19]. Theorem 4.1 also implies that when $\mathcal{G}$ is a non-complete homogeneous graph both $\mathcal{A}$ and $\mathcal{B}$ are of dimensions $r+r'$. This follows from the fact that, by Remark 4.2, there exists a DAG version $\mathcal{D}$ of $\mathcal{G}$ such that $r_{\mathcal{D}} = r'$. This agrees with Theorem 3.1 and Theorem 3.2 in [19] which fully describe $\mathcal{A}$ and $\mathcal{B}$.

The proof of Theorem 4.1 will be given in subsection 4.3. First, we proceed to introduce tools that will allow us to relate a decomposable graph to a specific perfect DAG version of it.

The proof of Theorem 4.1 heavily relies on the properties of certain perfect DAG versions of decomposable graphs. We proceed to describe such perfect DAG versions.

4.2. Induced perfect DAGs

First, we introduce a definition.

Definition 4.1. Let $\mathcal{P} = (C_1, \ldots, C_r)$ be a perfect order of $\mathcal{G}$. A DAG version $\mathcal{D}$ of $\mathcal{G}$ is said to be induced by $\mathcal{P}$ if the histories $H_1, \ldots, H_r$ are all ancestral in $\mathcal{D}$.

Lemma 4.1. *Let $\mathcal{G}$ be a non-complete decomposable graph and let $\mathcal{P} = (C_1, \ldots, C_r)$ be a perfect order of $\mathcal{G}$. Then every DAG version of $\mathcal{G}$ induced by $\mathcal{P}$ is a perfect DAG. Moreover, there always exists a perfect DAG version $\mathcal{D}$ of $\mathcal{G}$ induced by $\mathcal{P}$ such that S_2 is ancestral in $\mathcal{D}$, thus $r_{\mathcal{D}} \geq 1$.*

Proof. Suppose, to the contrary, that $\mathcal{D}$ is not perfect. Let t be the smallest integer such that $\mathcal{D}_{H_t}$, the induced DAG on H_t, is not perfect. It is clear that $1 < t \leq r$. Let $i \to k \leftarrow j$ be an immorality in $\mathcal{D}_{H_t}$. This, in particular, implies that there are two distinct, cliques C_{t_1} and C_{t_2}, with subscript $t_1, t_2 \leq t$, such that they contain i, k and k, j, respectively. Since t_1 and t_2 are distinct at least one of them say $t_1 \leq t-1$. However, since H_{t-1} is ancestral and j is a parent of $k \in H_{t-1}$ we must have $j \in H_{t-1}$. This contradicts the fact that the induced DAG on H_{t-1} is perfect.

Now we show that in particular there exists a DAG $\mathcal{D}$ induced by $\mathcal{P}$ such that S_2 is ancestral in $\mathcal{D}$. First, consider the case where there are only two cliques. We start with relabeling the vertices in S_2, $H_1 \setminus S_2$ and R_2, respectively, in a decreasing order. Let $\mathcal{D}$ be the DAG version of $\mathcal{G}$ induced by this order, that is, an edge $\{i,j\} \in E$ is converted to a directed edge $i \to j$ if $i > j$. By this construction, then S_2 and H_1 are ancestral in $\mathcal{D}$. Now suppose that such a DAG version exists for any decomposable graph with number of cliques less than $r \geq 3$. By the mathematical induction, there exists a DAG version $\mathcal{D}'$ of $\mathcal{G}_{H_{r-1}}$ such that $S_2, H_1, \ldots, H_{r-2}$ are ancestral in $\mathcal{D}'$. Without loss of generality, we can assume that the vertices in $\mathcal{D}'$ are labeled from $p, \ldots, p-|R_r|$. Let us label the vertices in R_r from $1, \ldots, |R_r|$ and let $\mathcal{D}$ be the DAG version of $\mathcal{G}$ induced by this order. One can easily check that $\mathcal{D}$ has the desired properties. $\square$

Below we state and prove a more particular version of Lemma 4.1 that is required.

Proposition 4.1. *Let $\mathcal{G}$ be a non-homogeneous decomposable graph that has two nested separators $S' \subsetneq S$. Then there exists a perfect order $\mathcal{P}$ of $\mathcal{G}$ such that for the perfect DAG induced by $\mathcal{P}$ the separator S_2 is ancestral in $\mathcal{D}$ and furthermore $r_{\mathcal{D}} \geq 2$.*

Proof. First, we choose a perfect order $\mathcal{P}$ such that the separator $S_2 = S$. Now we slightly modify the construction of $\mathcal{D}$ in the second paragraph of the proof of Lemma 4.1 by relabeling the vertices in $S' \subset S_2$ in a decreasing order. By this labeling, S' as well as S will be ancestral in $\mathcal{D}$. $\square$

Remark 4.2. When $\mathcal{H} = (V, E)$ is a non-complete homogeneous graph there always exists a perfect DAG version $\mathcal{D}$ of $\mathcal{H}$ such that $r_{\mathcal{D}}$ is equal to r', the number of (distinct) separators of $\mathcal{H}$. To construct $\mathcal{D}$, we direct each edge $(i,j) \in E$ either as $i \to j$ if $\mathrm{cl}(j) \subset \mathrm{cl}(i)$ or $j \to i$ if $\mathrm{cl}(i) \subset \mathrm{cl}(j)$ (in case of $\mathrm{cl}(j) = \mathrm{cl}(i)$ arbitrarily we choose one of these directed edges). First note that $i \to k \leftarrow j$ implies that $\mathrm{cl}(i) = \mathrm{cl}(j)$, therefore either $i \to j$ or $j \to i$. Also by construction $\mathcal{D}$ is transitive, that is, $i \to j \to k$ implies $i \to k$. This in particular implies:

(1) $\mathcal{D}$ is acyclic. Since $i_1 \to \cdots \to i_\nu \to i_1$ implies that $\{i_1, \ldots, i_\nu\}$ is a clique of $\mathcal{H}$ that does not intersect any other clique, which is impossible unless $\mathcal{H}$ is complete or disconnected.
(2) Every separator of $\mathcal{H}$ is ancestral in $\mathcal{D}$. On the contrary, suppose that there exists a separator S and a directed path $i \to \cdots \to j \in S$. By transitivity, then $i \to j$. Let C be a clique containing i and j. Since S is a separator there is another clique C' that contains it. However, $\mathrm{cl}(j) \subset \mathrm{cl}(i)$ implies that $C' \subset C$, which is a contradiction.

4.3. Two key lemmas

We proceed with two lemmas essential to the proof of Theorem 4.1. The first lemma yields an expression for the Markov ratio $H_{\mathcal{G}}(-(c+1)/2, -(s+1)/2, \Sigma_E)$, defined by Equation (3.1), in terms of a perfect DAG version of $\mathcal{G}$.

Lemma 4.2. *Suppose that $\mathcal{D}$ is a perfect DAG version of the decomposable graph $\mathcal{G}$. For a vertex j let pa_j denote $|pa(j)|$. Then*

$$H_{\mathcal{G}}\left(-\frac{c+1}{2}, -\frac{s+1}{2}, \Sigma_E\right) = \frac{\prod_{C\in\mathscr{C}} \det\left(\Sigma_C\right)^{-\frac{|C|+1}{2}}}{\prod_{S\in\mathscr{S}} \det\left(\Sigma_S\right)^{-\frac{|S|+1}{2}}} = \prod_{j\in V} \left(\Sigma_{jj|pa(j)}\right)^{-\frac{pa_j+2}{2}} \det\left(\Sigma_{pa(j)}\right)^{-\frac{1}{2}}, \qquad (4.1)$$

where for a positive definite matrix Σ,

$$\Sigma_{jj|pa(j)} = \Sigma_{jj} - \Sigma_{j,pa(j)}(\Sigma_{pa(j)})^{-1}\Sigma_{pa(j),j}.$$

Proof. We shall proceed by the method of mathematical induction. Suppose this is true for any decomposable graph with number of vertices less than p and we prove the lemma for $|V| = p$. The equality trivially holds when $p = 1$. Let us therefore assume that $p > 1$. As before, let r be the number of the cliques and consider the following cases.

1) Suppose that $r = 1$, that is, $\mathcal{G}$ is complete. This implies that $\mathrm{pa}(1) = V \setminus \{1\}$, therefore $\det(\Sigma) = \Sigma_{11|\mathrm{pa}(1)} \det(\Sigma_{\mathrm{pa}(1)})$. Thus, we can write

$$\begin{aligned}
&\frac{\prod_{C\in\mathscr{C}} \det\left(\Sigma_C\right)^{-\frac{|C|+1}{2}}}{\prod_{S\in\mathscr{S}} \det\left(\Sigma_S\right)^{-\frac{|S|+1}{2}}} \\
&= \det\left(\Sigma\right)^{-\frac{p+1}{2}} \\
&= \left(\Sigma_{11|pa(1)} \det\left(\Sigma_{pa(1)}\right)\right)^{-\frac{p+1}{2}} \\
&= \left(\Sigma_{11|pa(1)}\right)^{-\frac{pa_1+2}{2}} \det\left(\Sigma_{pa(1)}\right)^{-\frac{1}{2}} \det\left(\Sigma_{pa(1)}\right)^{-\frac{p}{2}} \\
&= \left(\Sigma_{11|pa(1)}\right)^{-\frac{pa_1+2}{2}} \det\left(\Sigma_{pa(1)}\right)^{-\frac{1}{2}} \prod_{j=2}^{p} \left(\Sigma_{jj|pa(j)}\right)^{-\frac{pa_j+2}{2}} \det\left(\Sigma_{pa(j)}\right)^{-\frac{1}{2}},
\end{aligned}$$

where the last equality uses the induction hypothesis for the induced graph $G_{pa(1)}$.

2) Suppose that $r \geq 2$. By Lemma 4.1 we can assume that $\mathcal{D}$ is induced by some perfect order $\mathcal{P} = (C_1, \ldots, C_r)$. This, in particular, implies that a vertex in R_r has no child. Without loss of generality, we label a such vertex as $1 \in R_r$. We now consider two cases:

a) If the residual $R_r = \{1\}$, then $(C_1, \ldots, C_{r-1})$ and $\mathcal{D}_{V\setminus\{1\}}$ are, respectively, a perfect order and a perfect DAG version of $\mathcal{G}_{V\setminus\{1\}}$. Moreover, $S_r = pa(1)$. Now it follows that

$$\begin{aligned}
&\frac{\prod_{t=1}^{r} \det\left(\Sigma_{C_t}\right)^{-\frac{c_t+1}{2}}}{\prod_{t=2}^{r} \det\left(\Sigma_{S_t}\right)^{-\frac{s_t+1}{2}}} \\
&= \frac{\prod_{t=1}^{r-1} \det\left(\Sigma_{C_t}\right)^{-\frac{c_t+1}{2}}}{\prod_{t=2}^{r-1} \det\left(\Sigma_{S_t}\right)^{-\frac{s_t+1}{2}}} \det\left(\Sigma_{R_r|S_r}\right)^{\frac{c_r+1}{2}} \det\left(\Sigma_{S_r}\right)^{\frac{s_r-c_r}{2}} \\
&= \prod_{j=2}^{p} \left(\Sigma_{jj|pa(j)}\right)^{-\frac{pa_j+2}{2}} \det\left(\Sigma_{pa(j)}\right)^{-\frac{1}{2}} \left(\Sigma_{11|pa(1)}\right)^{-\frac{pa_1+2}{2}} \det\left(\Sigma_{pa(1)}\right)^{-\frac{1}{2}} \\
&= \prod_{j=1}^{p} \left(\Sigma_{jj|pa(j)}\right)^{-\frac{pa_j+2}{2}} \det\left(\Sigma_{pa(j)}\right)^{-\frac{1}{2}}.
\end{aligned}$$

b) If the residual R_r has more than one element, then $(C_1, \ldots, C_{r-1}, C_r \setminus \{1\})$ is a perfect order of $\mathcal{G}_{V\setminus\{1\}}$ with associated separators $S_2, \ldots, S_r$. Using the induction hypothesis we obtain

$$\begin{aligned}
&\prod_{j=1}^{p} \Sigma_{jj|pa(j)}^{-\frac{pa_j+2}{2}} \det \Sigma_{pa(j)}^{-\frac{1}{2}} = \Sigma_{11|pa(1)}^{-\frac{pa_1+2}{2}} \det \Sigma_{pa(j)}^{-\frac{1}{2}} \prod_{j=2}^{p} \Sigma_{jj|pa(j)}^{-\frac{pa_j+2}{2}} \det \Sigma_{pa(j)}^{-\frac{1}{2}} \\
&= \Sigma_{11|C_r\setminus\{1\}}^{-\frac{c_r+1}{2}} \det\left(\Sigma_{C_r\setminus\{1\}}\right)^{-\frac{1}{2}} \det \frac{\prod_{j=1}^{r-1} \det\left(\Sigma_{C_t}\right)^{-\frac{c_t+1}{2}}}{\prod_{j=2}^{r} \det\left(\Sigma_{S_t}\right)^{-\frac{s_t+1}{2}}} \det\left(\Sigma_{C_r\setminus\{1\}}\right)^{-\frac{c_r}{2}} \\
&= \frac{\prod_{j=1}^{r-1} \det\left(\Sigma_{C_t}\right)^{-\frac{c_t+1}{2}}}{\prod_{j=2}^{r} \det\left(\Sigma_{S_t}\right)^{-\frac{s_t+1}{2}}} \det\left(\Sigma_{C_r}\right)^{-\frac{c_r+1}{2}} \\
&= \frac{\prod_{j=1}^{r} \det\left(\Sigma_{C_t}\right)^{-\frac{c_t+1}{2}}}{\prod_{j=2}^{r} \det\left(\Sigma_{S_t}\right)^{-\frac{s_t+1}{2}}} \qquad \square
\end{aligned}$$

Remark 4.3. The Markov ratio in Equation (4.1) is the square root of the Jacobian of the inverse mapping $\left(\Sigma_E \mapsto \Sigma^{-1}\right) : \mathrm{Q}_{\mathcal{G}} \to \mathrm{P}_{\mathcal{G}}$ (see [19, Theorem 2.1]).

Example 4.1. Let us illustrate the result of Lemma 4.2 for the decomposable graph $\mathcal{G}$ and its perfect DAG version $\mathcal{D}$ given in Figure 2(a) and Figure 2(b). Consider the perfect order

$$\mathcal{P} = \left(C_1 = \{3,5,6\}, C_2 = \{4,5,6\}, C_3 = \{2,6\}, C_4 = \{1,4\}\right).$$

The separators of $\mathcal{G}$ are $S_2 = \{5,6\}$, $S_3 = \{6\}$ and $S_4 = \{4\}$. Recall the notation $fa(j) = pa(j) \cup \{j\}$. Using the relationships between the vertices in $\mathcal{D}$ we have $C_1 = fa(3)$, $C_2 = fa(4)$, $C_3 = fa(2)$, $C_4 = fa(1)$, $S_2 = pa(3)$,

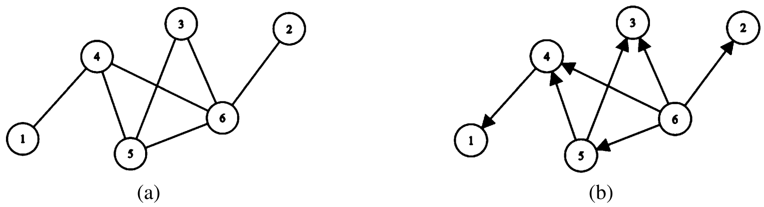


FIG 2. *A counterexample to the LM conjecture II.*

$S_3 = pa(2)$ and $S_4 = pa(1)$. Now we can write the Markov ratio as

$$\frac{\prod_{t=1}^{4} \det \Sigma_{C_t}^{-\frac{c_t+1}{2}}}{\prod_{t=2}^{4} \det \Sigma_{S_t}^{-\frac{s_t+1}{2}}} = \frac{\det \Sigma_{fa(3)}^{-2} \det \Sigma_{fa(4)}^{-2} \det \Sigma_{fa(2)}^{-3/2} \det \Sigma_{fa(1)}^{-3/2}}{\det \Sigma_{pa(3)}^{-3/2} \det \Sigma_{pa(2)}^{-1} \det \Sigma_{pa(1)}^{-1}}$$

$$= \frac{\Sigma_{33|pa(3)}^{-2} \det \Sigma_{pa(3)}^{-2} \Sigma_{44|pa(4)}^{-2} \det \Sigma_{pa(4)}^{-2} \Sigma_{22|pa(2)}^{-3/2} \det \Sigma_{pa(2)}^{-3/2} \Sigma_{11|pa(1)}^{-3/2} \det \Sigma_{pa(1)}^{-3/2}}{\det \Sigma_{pa(3)}^{-3/2} \det \Sigma_{pa(2)}^{-1} \det \Sigma_{pa(1)}^{-1}}$$

$$= \Sigma_{33|pa(3)}^{-2} \det \Sigma_{pa(3)}^{-1/2} \Sigma_{44|pa(4)}^{-2} \det \Sigma_{pa(4)}^{-1/2} \Sigma_{55|pa(5)}^{-3/2} \det \Sigma_{pa(5)}^{-1/2} \Sigma_{22|pa(2)}^{-3/2}$$

$$\times \det \Sigma_{pa(2)}^{-1/2} \Sigma_{11|pa(1)}^{-3/2} \det \Sigma_{pa(1)}^{-1/2} \Sigma_{66}^{-1}$$

$$= \prod_{j=1}^{6} \Sigma_{jj|pa(j)}^{-\frac{pa_j+2}{2}} \det \Sigma_{pa(j)}^{-1/2}.$$

The second key lemma we require in the proof of Theorem 4.1 is as follows.

Lemma 4.3. *Let $\mathcal{D}$ be a DAG version of the decomposable graph $\mathcal{G}$ induced by the perfect order $\mathcal{P}$. Then for each $j = 2, \ldots, r$ we have* $\det\left(\Sigma_{R_t|S_t}\right) = \prod_{i \in R_t} \Sigma_{ii|\mathrm{pa}(i)}$.

Proof. Let $K_t = H_t \setminus C_t$ for each $j = 2, \ldots, r$. Consider partitioning of Σ_{H_t} as

$$\Sigma_{H_t} = \begin{pmatrix} \Sigma_{R_t} & \Sigma_{R_t S_t} & \Sigma_{R_t K_t} \\ \Sigma_{S_t R_t} & \Sigma_{S_t} & \Sigma_{S_t K_t} \\ \Sigma_{K_t R_t} & \Sigma_{K_t S_t} & \Sigma_{K_t} \end{pmatrix}.$$

Now for each $t = 2, \ldots, r$, by Corollary A.1 in Appendix A.1, we have $\Sigma_{H_t} \in \mathrm{PD}_{\mathcal{G}_{H_t}}$, and S_t separates R_t from K_t. By Lemma 5.5 in [18] we have

$$\det\left(\Sigma_{H_t}\right) = \frac{\det\left(\Sigma_{C_t}\right) \det\left(\Sigma_{H_{t-1}}\right)}{\det\left(\Sigma_{S_t}\right)}.$$

By rewriting this and using Lemma A.4 we obtain

$$\det\left(\Sigma_{R_t|S_t}\right) = \det\left(\Sigma_{H_t}\right) \det\left(\Sigma_{H_{t-1}}\right)^{-1} = \prod_{k \in R_t} D_{kk},$$

where, for an index j, D_{jj} denotes $\Sigma_{jj|pa(j)}$ (we refer the reader to Lemma A.4 in Appendix for another usage of this notation). $\square$

4.4. Proof of Theorem 4.1

Let $\mathcal{P}$ be a perfect order of $\mathcal{G}$. By using either Proposition 4.1 or Remark 4.2, there exists a perfect DAG version $\mathcal{D}$ of $\mathcal{G}$ induced by $\mathcal{P}$ such that S_2 is ancestral in $\mathcal{D}$ and $r_{\mathcal{D}} \geq 2$. We proceed to prove that the dimension of $\mathcal{A}$ and $\mathcal{B}$ is greater than or equal to $r + r_{\mathcal{D}}$.

A) We show that the dimension of $\mathcal{A}$ is greater than or equal to $r + r_{\mathcal{D}}$ as follows. For this, first we use Lemma 4.2 to rewrite the density of the Type I Wishart distribution.

$$\begin{aligned}
H_{\mathcal{G}}\left(\alpha, \beta, \Sigma_E\right) &= \frac{\prod_{t=1}^{r} \det\left(\Sigma_{C_t}\right)^{\alpha_t}}{\prod_{t=2}^{r} \det\left(\Sigma_{S_t}\right)^{\beta_t}} \\
&= \det\left(\Sigma_{R_1|S_2}\right)^{\alpha_1} \det\left(\Sigma_{S_2}\right)^{\alpha_1} \prod_{t=2}^{r} \det\left(\Sigma_{R_t|S_t}\right)^{\alpha_t} \prod_{t=2}^{r} \det\left(\Sigma_{S_t}\right)^{\alpha_t - \beta_t} \\
&= \det\left(\Sigma_{R_1|S_2}\right)^{\alpha_1} \prod_{t=2}^{r} \det\left(\Sigma_{R_t|S_t}\right)^{\alpha_t} \prod_{S \in \mathscr{S}^{\mathcal{D}}} \det(\Sigma_S)^{\eta_S} \\
&\quad \times \prod_{S \notin \mathscr{S}^{\mathcal{D}}}^{r} \det\left(\Sigma_S\right)^{\sum\left(\alpha_t : t \in J(\mathcal{P}, S)\right) - \nu(S)\beta(S)},
\end{aligned} \tag{4.2}$$

where η_S is a number determined by S, α and β. By Lemma 4.3 and Part (ii) of Lemma A.4 in Appendix A.1, each term such as $\det\left(\Sigma_{R_t|S_t}\right)$ and $\det\left(\Sigma_S\right)$, for each $S \in \mathscr{S}^{\mathcal{D}}$, is products of $\Sigma_{jj|pa(j)}$ for some $j \in V$. Let

$$\tilde{A}_{\mathcal{P}} = \Big\{ (\alpha, \beta) \in \mathbb{R}^r \times \mathbb{R}^{r-1} : \sum \left(\alpha_t : \ t \in J\left(\mathcal{P}, S\right)\right) - \nu\left(S\right) \beta\left(S\right) = 0, \\ \forall S \notin \mathscr{S}^{\mathcal{D}} \Big\}.$$

Then for every $(\alpha, \beta) \in \tilde{A}_{\mathcal{P}}$, Equation (4.2) is written as

$$H_{\mathcal{G}}\left(\alpha, \beta, \Sigma_E\right) = \left(\Sigma_{jj|pa(j)}\right)^{\lambda_j}, \tag{4.3}$$

where $\lambda_j = \lambda_j\left(\alpha, \beta\right)$ is a linear combination of the components of α and β. Therefore, if (α, β) is restricted to $\tilde{A}_{\mathcal{P}}$, then Equation (4.1) and Equation (4.3) together imply that

$$\begin{aligned}
&\omega_{Q_G}\left(\alpha, \beta, U_E, d\Sigma_E\right) \\
&= \exp\left\{-\operatorname{tr}\left(\Sigma U^{-1}\right)\right\} \prod_{j=1}^{p} \Sigma_{jj|pa(j)}^{\lambda_j} \prod_{j=1}^{p} \Sigma_{jj|pa(j)}^{-\frac{pa_j+2}{2}} \det \Sigma_{pa(j)}^{-\frac{1}{2}} d\Sigma_E.
\end{aligned} \tag{4.4}$$

Now the integral of the right-hand side expression in Equation (4.4) is finite, over $\mathrm{Q}_{\mathcal{G}}$, if and only if $\lambda_j > pa_j/2$ for each $j = 1, \ldots, p$ (see [1, section 7]). Furthermore, if this condition is satisfied, then we have

$$\int_{\mathrm{Q}_{\mathcal{G}}} \exp\left\{-\operatorname{tr}\left(\Sigma U^{-1}\right)\right\} \prod_{j=1}^{p} \Sigma_{jj|pa(j)}^{\lambda_j} \prod_{j=1}^{p} \Sigma_{jj|pa(j)}^{-\frac{pa_j+2}{2}} \det \Sigma_{pa(j)}^{-\frac{1}{2}} d\Sigma_E \propto \prod_{j=1}^{p} U_{jj|pa(j)}^{\lambda_j}$$

for $\lambda_j > pa_j/2$.

Therefore, for each $(\alpha,\beta)\in\tilde{A}_{\mathcal{P}}$ and $\lambda_j > pa_j/2$, Equation (A1) is satisfied. It is clear that Equation (A2) is also satisfied, because for each $(\alpha,\beta)\in\tilde{A}_{\mathcal{P}}$ by Equation (4.3) we have $H_{\mathcal{G}}\left(\alpha,\beta,U_E\right)=\prod_{j=1}^{p}\left(U_{jj|pa(j)}\right)^{\lambda_j}$, which implies that the proportionality constant is not a function of the hyper-parameter U. This shows that both (A1) and (A2) are satisfied on a manifold of dimension $\geq r+r_{\mathcal{D}}$, the dimension of $\tilde{A}_{\mathcal{P}}$.

B) We proceed to show that the dimension of $\mathcal{B}$ is also larger than or equal to $r+r_{\mathcal{D}}$. By using Equation (4.1) and Equation (4.2), we write the the Markov ratio that appears in the density of $W_{\mathrm{P}_{\mathcal{G}}}$ as

$$\begin{aligned}
&H\left(\alpha+\frac{c+1}{2},\beta+\frac{s+1}{2}\right)=\frac{\prod_{t=1}^{r}\det\left(\Sigma_{C_t}\right)^{\alpha_t+\frac{c_t+1}{2}}}{\prod_{t=2}^{r}\det\left(\Sigma_{S_t}\right)^{\beta_t+\frac{s_t+1}{2}}}\\
&=\det\left(\Sigma_{R_1|S_2}\right)^{\alpha_1+\frac{c_1+1}{2}}\prod_{t=2}^{r}\det\left(\Sigma_{R_t|S_t}\right)^{\alpha_t+\frac{c_t+1}{2}}\prod_{S\in\mathscr{S}^{\mathcal{D}}}\det(\Sigma_S)^{\eta_S+\frac{|S|+1}{2}}\\
&\times\prod_{S\notin\mathscr{S}^{\mathcal{D}}}^{r}\det\left(\Sigma_S\right)^{\sum\left(\alpha_t+\frac{c_t-|S|}{2}:t\in J(\mathcal{P},S)\right)-\nu(S)\beta(S)}\\
&=\prod_{j=1}^{p}D_{jj}^{\gamma_j}\prod_{S\notin\mathscr{S}^{\mathcal{D}}}^{r}\det\left(\Sigma_S\right)^{\sum\left(\alpha_t+\frac{c_t-|S|}{2}:t\in J(\mathcal{P},S)\right)-\nu(S)\beta(S)},
\end{aligned}\tag{4.5}$$

where each γ_j is an affine combination of the components of α and β. Consequently, if (α,β) is restricted to the set

$$\tilde{B}_{\mathcal{P}}=\{(\alpha,\beta)\in\mathbb{R}^r\times\mathbb{R}^{r-1}\ :\sum_{t\in J(\mathcal{P},S)}(\alpha_t+\frac{c_t-|S|}{2})-\nu\left(s\right)\beta(S)=0,\quad\forall S\notin\mathscr{S}^{\mathcal{D}}\}.$$

Then, $\omega_{\mathrm{P}_{\mathcal{G}}}$, the non-normalized version of the Type II Wishart distribution, is written as

$$\omega_{\mathrm{P}_{\mathcal{G}}}\left(\alpha,\beta,U_E,d\Omega\right)=\exp\left\{-\mathrm{tr}\left(\Omega U\right)\right\}\prod_{j=1}^{p}D_{jj}^{\gamma_j}d\Omega\qquad\text{for every }(\alpha,\beta)\in\tilde{B}_{\mathcal{P}}.\tag{4.6}$$

Now in Appendix A.5 we have shown that if we set

$$\widehat{\pi}_{\mathcal{D}}(\alpha,\beta,U,d\Omega)=\exp\left\{-\frac{1}{2}\mathrm{tr}\left(\Omega U\right)\right\}\prod_{i=1}^{p}D_{jj}^{-\frac{1}{2}\eta_j+pa_j+2}d\Omega,\tag{4.7}$$

then

$$\int_{\mathrm{P}_{\mathcal{G}}}\widehat{\pi}_{\mathrm{P}_{\mathcal{D}}}\left(\eta,U,d\Omega\right)=\prod_{j=1}^{p}\frac{\Gamma\left(\frac{\eta_j}{2}-\frac{pa_j}{2}-1\right)2^{\frac{\eta_j}{2}-1}\left(\sqrt{\pi}\right)^{pa_j}\det\left(U_{pa(j)}\right)^{\frac{\eta_j}{2}-\frac{pa_j}{2}-\frac{3}{2}}}{\det\left(U_{fa(j)}\right)^{\frac{\eta_j}{2}-\frac{pa_j}{2}-1}},$$

for $\eta_j > pa_j + 2$. If the exponents $-\frac{1}{2}\eta_j + pa_j + 2$ in Equation (4.7) are replaced by γ_j, and U is replaced by $2U$, then we conclude that

$$\int_{\mathrm{P}_{\mathcal{G}}} \exp\left\{-\mathrm{tr}\left(\Omega U\right)\right\} \prod_{i=1}^{p} D_{jj}^{\gamma_j} d\Omega = \prod_{j=1}^{p} \frac{\Gamma\left(-\gamma_j + \frac{pa_j}{2} + 1\right)\left(\sqrt{\pi}\right)^{pa_j}}{U_{jj|pa(j)}^{-\gamma_j + \frac{pa_j}{2} + 1} \det\left(U_{pa(j)}\right)^{\frac{1}{2}}}, \tag{4.8}$$

for $\gamma_j < pa_j/2+1$. Using Equation (4.6) and Equation (4.1) for each $(\alpha, \beta) \in \tilde{B}_{\mathcal{P}}$ and $\gamma_j < pa_j/2 + 1$, we have

$$\begin{aligned}
&\int_{\mathrm{P}_{\mathcal{G}}} \omega_{\mathrm{P}_{\mathcal{G}}}\left(\alpha, \beta, U_E, d\Omega\right) \propto U_{jj|pa(j)}^{\gamma_j - \frac{pa_j}{2} - 1} \det\left(U_{pa(j)}\right)^{-\frac{1}{2}} \\
&= U_{jj|pa(j)}^{-\frac{pa_j}{2} - 1} \det\left(U_{pa(j)}\right)^{-\frac{1}{2}} \prod_{j=1}^{p} U_{jj|pa(j)}^{\gamma_j} \\
&\quad \times \prod_{S \notin \mathscr{S}^{\mathcal{D}}}^{r} \det\left(\Sigma_S\right)^{\sum\left(\alpha_t + \frac{c_t - |S|}{2} : j \in J(\mathcal{P}, S)\right) - \nu(S)\beta(S)} \\
&= H\left(-\frac{c+1}{2}, -\frac{s+1}{2}, U_E\right) H\left(\alpha + \frac{c+1}{2}, \beta + \frac{s+1}{2}, U_E\right) \\
&= H\left(\alpha, \beta, U_E\right).
\end{aligned}$$

From these we now conclude that Equation (B1) and Equation (B2) are satisfied on a set of dimension larger than or equal to $r + r_{\mathcal{D}}$. The rest of the proof now follows from Remark 3.1 since we have shown that $\mathcal{A}$ and $\mathcal{B}$ contain a set of dimension $\geq r + r_{\mathcal{D}} > r + 1$. □

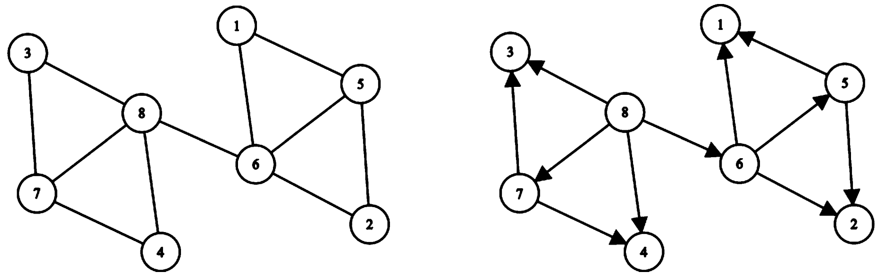
Remark 4.4. The proof of Theorem 4.1 shows that we can modify the definition of $A_{\mathcal{P}}$ and $B_{\mathcal{P}}$ by reducing the number of equality constraints in (A1) and (B1) by removing constraints with $S \in \mathscr{S}^{\mathcal{D}}$. This identifies larger subsets of $\mathcal{A}$ and $\mathcal{B}$ of dimensions $r + r_{\mathcal{D}}$. However, the following example shows that this modification is not enough to fully describe $\mathcal{A}$ and $\mathcal{B}$ since the dimension of the latter sets can be even larger than $r + r_{\mathcal{D}}$.

Example 4.2. Consider the graph $\mathcal{G}$ given in Figure 1(a) and its DAG version given in Figure 3(b). Note that $\mathcal{D}$ is induced by the perfect order

$$\mathcal{P} = \left\{C_1 = \{4, 7, 8\}, C_2 = \{3, 7, 8\}, C_3 = \{6, 8\}, C_4 = \{2, 5, 6\}, C_5 = \{1, 5, 6\}\right\}.$$

First note that the only ancestral separators in this DAG version of $\mathcal{G}$ are $S_2 = \{7, 8\}$ and $S_3 = \{8\}$, thus Theorem 4.1 guarantees that the dimension of $\mathcal{A}$ and that of $\mathcal{B}$ are larger than or equal to 7. However, a similar proof as that of Theorem 4.1 shows that these dimensions are in fact greater than or equal to 8. For this, let us rewrite the Markov ratio as:

$$\frac{\prod_{j=1}^{5} \det\left(\Sigma_{C_j}\right)^{\alpha_j - \frac{c_j+1}{2}}}{\prod_{j=2}^{5} \det\left(\Sigma_{S_j}\right)^{\beta_j - \frac{s_j+1}{2}}}$$



(a) (b)

FIG 3. *Another counterexample to the LM conjecture. In this graph, the dimension of $\mathcal{A}$ and $\mathcal{B}$ is ≥ 8.*

$$
\begin{aligned}
&= \frac{\det(\Sigma_{\preceq 4 \succeq})^{\alpha_1} \det(\Sigma_{\preceq 3 \succeq})^{\alpha_2} \det(\Sigma_{\preceq 6 \succeq})^{\alpha_3} \det(\Sigma_{\preceq 2 \succeq})^{\alpha_4} \det(\Sigma_{\preceq 1 \succeq})^{\alpha_5}}{\det(\Sigma_{\prec 3 \succ})^{\beta_2} \det(\Sigma_{\prec 6 \succ})^{\beta_3} \det(\Sigma_{\prec 5 \succ})^{\beta_4} \det(\Sigma_{\prec 1 \succ})^{\beta_5}} \\
&\times \frac{\det(\Sigma_{\preceq 4 \succeq})^{-2} \det(\Sigma_{\preceq 3 \succeq})^{-2} \det(\Sigma_{\preceq 6 \succeq})^{-\frac{3}{2}} \det(\Sigma_{\preceq 2 \succeq})^{-2} \det(\Sigma_{\preceq 1 \succeq})^{-2}}{\det(\Sigma_{\prec 3 \succ})^{-\frac{3}{2}} \det(\Sigma_{\prec 6 \succ})^{-1} \det(\Sigma_{\prec 5 \succ})^{-1} \det(\Sigma_{\prec 1 \succ})^{-\frac{3}{2}}} \\
&= \frac{D_{44}^{\alpha_1} D_{77}^{\alpha_1} D_{88}^{\alpha_1} D_{33}^{\alpha_2} D_{77}^{\alpha_2} D_{88}^{\alpha_2} D_{66}^{\alpha_3} D_{88}^{\alpha_3} D_{22}^{\alpha_4} D_{55}^{\alpha_4} \Sigma_{66}^{\alpha_4} D_{11}^{\alpha_5} D_{55}^{\alpha_5} \Sigma_{66}^{\alpha_5}}{D_{77}^{\beta_2} D_{88}^{\beta_2} D_{88}^{\beta_3} \Sigma_{66}^{\beta_4} D_{55}^{\beta_5} \Sigma_{66}^{\beta_5}} \\
&\times \prod_{j=1}^{8} \left(\Sigma_{jj|\prec j \succ}\right)^{-\frac{pa_j+2}{2}} \det(\Sigma_{\prec j \succ})^{-\frac{1}{2}} \\
&= D_{11}^{\alpha_5} D_{22}^{\alpha_4} D_{33}^{\alpha_2} D_{44}^{\alpha_1} D_{55}^{\alpha_4+\alpha_5-\beta_5} D_{66}^{\alpha_3} D_{77}^{\alpha_1+\alpha_2-\beta_2} D_{88}^{\alpha_1+\alpha_2+\alpha_3-\beta_2-\beta_3} \Sigma_{66}^{\alpha_4+\alpha_5-\beta_4-\beta_5} \\
&\times \prod_{j=1}^{8} \Sigma_{jj|\prec j \succ}^{-\frac{pa_j+2}{2}} \det \Sigma_{\prec j \succ}^{-\frac{1}{2}}, \qquad (4.9)
\end{aligned}
$$

where

$$
pa_j = \begin{cases} 2 & \text{for } j = 1,2,3,4, \\ 1 & \text{for } j = 5,6,7, \\ 0 & \text{for } j = 8. \end{cases}
$$

Let λ_j be the exponent of D_{jj} in Equation (4.9) for each $j = 1, \ldots, 8$. If we set $\alpha_4 + \alpha_5 - \beta_4 - \beta_5 = 0$, then we obtain

$$
\begin{aligned}
&\int_{\mathrm{Q}_{\mathcal{G}}} \omega_{\mathrm{Q}\mathcal{G}}(\alpha, \beta, U, d\Sigma_E) \\
&= \int_{\mathrm{Q}_{\mathcal{G}}} \exp\{-\mathrm{tr}(\Sigma U)\} \prod_{j=1}^{8} D_{jj}^{\lambda_j} \prod_{j=1}^{8} \left(\Sigma_{jj|pa(j)}\right)^{-\frac{pa_j+2}{2}} \det\left(\Sigma_{pa(j)}\right)^{-\frac{1}{2}} d\Sigma_E. \qquad (4.10)
\end{aligned}
$$

The integrand in the right-hand-side of Equation (4.10) corresponds to the non-normalized density of the generalized Riesz distribution on $\mathrm{Q}_{\mathcal{D}}$ and has a finite integral if and only if $\lambda_j > pa_j/2$ for each $j = 1, \ldots, 8$. Furthermore, under these conditions, we have

$$\int_{\mathrm{Q}_{\mathcal{G}}} \exp\left\{-\mathrm{tr}\left(\Sigma U\right)\right\} \prod_{j=1}^{8} D_{jj}^{\lambda_j} \prod_{j=1}^{8} \left(\Sigma_{jj|pa(j)}\right)^{-\frac{pa_j+2}{2}} \det\left(\Sigma_{pa(j)}\right)^{-\frac{1}{2}} d\Sigma_E$$
$$\propto \prod_{j=1}^{p} \left(U_{jj|pa(j)}\right)^{\lambda_j}. \tag{4.11}$$

See [1, §6] for details. By very similar calculations as those who led to Equation (4.9) we can show that

$$H_{\mathcal{G}}\left(\alpha, \beta, U\right) = \prod_{j=1}^{p} \left(U_{jj|pa(j)}\right)^{\lambda_j} U_{66}^{\alpha_4+\alpha_5-\beta_4-\beta_5}.$$

Therefore from Equation (4.11), for any (α, β) satisfying $\lambda_j\left(\alpha, \beta\right) > pa_j/2$ and $\alpha_4 + \alpha_5 - \beta_4 - \beta_5 = 0$ we conclude that

$$\int_{\mathrm{Q}_{\mathcal{G}}} \omega_{\mathrm{Q}\mathcal{G}}\left(\alpha, \beta, U, d\Sigma_E\right)/H_{\mathcal{G}}\left(\alpha, \beta, U\right) < \infty,$$

and the proportionality constant is not a function of the hyper-parameter U. Consequently, $\mathcal{A}$ contains a manifold of dimension greater than or equal to 8. We leave it to the reader to show that the dimension of $\mathcal{B}$ is also greater than or equal to 8.

In this example, the main difficulties in fully identifying $\mathcal{A}$, and similarly $\mathcal{B}$, arise in integrating (4.9) in the presence of the term $\Sigma_{66}^{\alpha_4+\alpha_5-\beta_4-\beta_5}$ (associated with the non-ancestral separator $S = \{6\}$) and then checking whether conditions (A2) and (B2) are satisfied. Here, we simply eliminate this term by equating $\alpha_4+\alpha_5-\beta_4-\beta_5 = 0$. This may be necessary, but more involved calculations are required to prove it. When $\mathcal{G}$ is an arbitrary decomposable graph we encounter a similar problem in Equation (4.2) and Equation (4.5) in the presence of the terms like $\prod_{S\notin\mathscr{S}^{\mathcal{D}}} \Sigma_S^{\nu_S}$.

Remark 4.5. In Theorem 4.1 we have proved that a necessary condition for the LM conjecture to hold for a decomposable graph is that the graph should not have nested separators. An open question is that whether this condition is also sufficient. More generally, giving sufficient and necessary conditions for decomposable graphs on which the LM conjecture holds remains an open problem.

Acknowledgment

We thank the anonymous reviewers for their careful reading of our manuscript and their many constructive comments and suggestions for improving the paper.

References

[1] Steen A. Andersson and Thomas Klein. On Riesz and Wishart distributions associated with decomposable undirected graphs. *J. Multivariate Anal.*, 101(4):789–810, 2010. MR2584900

[2] Steen A. Andersson and Michael D. Perlman. Normal linear regression models with recursive graphical Markov structure. *J. Multivariate Anal.*, 66(2):133–187, 1998. MR1642461

[3] Emanuel Ben-David and Bala Rajaratnam. Positive definite completion problems for Bayesian networks. *SIAM J. Matrix Anal. Appl.*, 33(2):617–638, 2012. MR2970222

[4] C. M. Carvalho and J. G. Scott. Objective Bayesian model selection in Gaussian graphical models. *Biometrika*, 96(3):497–512, 2009. MR2538753

[5] D. Corneil, Y. Perl, and L. Stewart. A linear recognition algorithm for cographs. *SIAM J. Comput.*, 14(4):926–934, 1985. MR0807891

[6] Elias Dahlhaus. Efficient parallel recognition algorithms of cographs and distance hereditary graphs. *Discrete Appl. Math.*, 57(1):29–44, 1995. MR1317192

[7] Philip A. Dawid and Steffen L. Lauritzen. Hyper-Markov laws in the statistical analysis of decomposable graphical models. *Ann. Statist.*, 21(3):1272–1317, 1993. MR1241267

[8] Persi Diaconis, Kshitij Khare, and Laurent Saloff-Coste. Gibbs sampling, exponential families and orthogonal polynomials. *Statist. Sci.*, 23(2):151–178, 05 2008. MR2446500

[9] Jacques Faraut and Adam Korányi. *Analysis on Symmetric Cones*. Oxford mathematical monographs. Clarendon Press, 1994. MR1446489

[10] Simon G. Gindikin. Invariant generalized functions in homogeneous domains. *Funct. Anal. Appl.*, 9, 1975. MR0377423

[11] Piotr Graczyk, Hideyuki ISHI, Salha Mamane, and Hiroyuki Ochiai. On the Letac-Massam conjecture on cones Q_{A_n}. *Proc. Japan Acad. Ser. A Math. Sci.*, 93(3):16–21, 03 2017. MR3619768

[12] Robert Grone, Charles R. Johnson, Eduardo M. de Sá, and Henry Wolkowicz. Positive definite completions of partial Hermitian matrices. *Linear Algebra Appl.*, 58:109–124, 1984. MR0739282

[13] Hideyuki ISHI. Positive Riesz distributions on homogeneous cones. *J. Math. Soc. Japan*, 52(1):161–186, 01 2000. MR1727195

[14] Hideyuki ISHI. On a class of homogeneous cones consisting of real symmetric matrices. *Josai Mathematical Monographs*, (6):71–80, 2013.

[15] Hideyuki ISHI. Homogeneous cones and their applications to statistics. *Modern methods of multivariate statistics*, Travaux en Cours (82):135–154, Hermann, 2014.

[16] Kshitij Khare and Bala Rajaratnam. Wishart distributions for decomposable covariance graph models. *Ann. Statist.*, 39(1):514–555, 2011. MR2797855

[17] Kshitij Khare and Bala Rajaratnam. Sparse matrix decompositions and graph characterizations. *Linear Algebra Appl.*, 437(3):932–947, 2012.

MR2921746
[18] Steffen L. Lauritzen. *Graphical models*, volume 17 of *Oxford Statistical Science Series*. The Clarendon Press, Oxford University Press, New York, 1996. Oxford Science Publications. MR1419991
[19] Gérard Letac and Hélène Massam. Wishart distributions for decomposable graphs. *Ann. Statist.*, 35(3):1278–1323, 2007. MR2341706
[20] Editors Marloes Maathuis, Mathias Drton, Steffen Lauritzen, and Martin Wainwright. *Handbook of graphical models*. CRC Press, Boca Raton, Florida, 2019. Wiley Series in Probability and Mathematical Statistics. MR3889064
[21] Bala Rajaratnam, Hélène Massam, and Carlos M. Carvalho. Flexible covariance estimation in graphical Gaussian models. *Ann. Statist.*, 36(6):2818–2849, 2008. MR2485014
[22] Alberto Roverato. Cholesky decomposition of a hyper inverse Wishart matrix. *Biometrika*, 87(1):99–112, 2000. MR1766831
[23] Nanny Wermuth. Linear recursive equations, covariance selection, and path analysis. *J. Amer. Statist. Assoc.*, 75(372):963–972, 1980. MR0600984

Appendix

A.1. Gaussian DAG models

Let $D = (V, F)$ be DAG. Without loss of generality, we can assume that the vertices are parent ordered, that is, for each $i, j \in V$ if $i \to j$, then $i > j$. The Gaussian DAG model over $\mathcal{D}$, denoted by $\mathscr{N}(\mathcal{D})$, is the family of Gaussian distributions $N_p(0, \Sigma)$ which are (directed) Markov with respect to $\mathcal{D}$. A simple observation in [2] shows that a random vector $\mathbf{X} = N_p(0, \Sigma)$ is directed Markov with respect to $\mathcal{D}$ if and only if $\Sigma \succ 0$ and

$$\Sigma_{pr(j),j} = \Sigma_{pr(j),pa(j)} \left(\Sigma_{pa(j)}\right)^{-1} \Sigma_{pa(j),j} \quad \text{for every } j \in V, \tag{A.12}$$

where $pr(j) = \{i : \ i \in \operatorname{nd}(j) \text{ and } i > j\}$. Therefore, the space of covariance matrices for $\mathscr{N}(\mathcal{D})$ is

$$\mathrm{PD}_{\mathcal{D}} = \left\{\Sigma \succ 0 : \Sigma_{pr(j),j} = \Sigma_{pr(j),pa(j)} \left(\Sigma_{pa(j)}\right)^{-1} \Sigma_{pa(j),j},\ \forall j \in V\right\}. \tag{A.13}$$

For a DAG $\mathcal{D}$, we define $\mathrm{Q}_{\mathcal{D}} = \mathrm{Q}_{\mathcal{D}^u}$. It is clear that if $\mathcal{D}$ is a perfect DAG version of a decomposable graph G, then $Q_{\mathcal{D}} = \mathrm{Q}_{\mathcal{G}}$.

Here we state and prove the following lemma and its corollary, which we will need in Subsection 4.3.

Lemma A.4. *Let $\mathcal{D}$ be a DAG and let $\mathbf{X} \sim N_p(0, \Sigma) \in \mathscr{N}(\mathcal{D})$. Let $\Sigma^{-1} = \Omega = LD^{-1}L^\top$ be the modified Cholesky decomposition of Σ^{-1}, which means D is a diagonal matrix and L is a unit lower triangular matrix.*

Part (i) *For each $i, j \in V$ if $i \in \operatorname{pa}(j)$, then $L_{ij} = -\beta_{ji}$, where β_{ji} is the partial regression coefficient of X_i in the linear regression of X_j on $\mathbf{X}_{pa(i)}$ and $D_{kk} = \Sigma_{jj|pa(k)}$.*

Part (ii) If A is an ancestral subset of V, then $(\Sigma_A)^{-1} = L_A D_A^{-1} L_A^\top$. In particular, $\det(\Sigma_A) = \prod_{k \in A} D_{kk}$ *(see [17] for a related result).*

Proof. Part (i) follows from Equation (A.12) (see [3, 23] for details). We proceed to prove Part (ii). Since A is ancestral in $\mathcal{D}$, by using Equation (A.13), one can easily show that $\mathbf{X}_A \sim N_{|A|}(0, \Sigma_A) \in \mathscr{N}(\mathcal{D}_A)$. Now let $(\Sigma_A)^{-1} = K\Lambda^{-1}K^\top$ be the modified Cholesky decomposition of $(\Sigma_A)^{-1}$. Part (i) and the fact that A is ancestral imply that for each $i \in \mathrm{pa}(j)$, $K_{ij} = -\beta_{ij} = L_{ji}$ and $\Lambda_{jj} = \Sigma_{jj|pa(j)} = D_{jj}$. This implies that $K = L_A$ and $\Lambda = D_A$ □

Corollary A.1. *Let $\mathcal{D}$ be a perfect DAG version of $\mathcal{G}$ and let $\Sigma \in \mathrm{PD}_{\mathcal{G}}$. If A is an ancestral subset of V (in $\mathcal{D}$), then $\Sigma_A \in \mathrm{PD}_{\mathcal{G}_A}$.*

Proof. Let $\Sigma = LD^{-1}L^\top$ be the modified Cholesky decomposition of Σ. If A is an ancestral subset of V, then $(\Sigma_A)^{-1} = L_A D_A^{-1} L_A^\top$ by Part (ii) of Lemma A.4. Using this fact one can easily check that $(\Sigma_A)^{-1}_{ij} = 0$ whenever $\{i, j\} \notin A$. □

A.2. Hyper Markov properties

Let $\{P_\theta : \theta \in \Theta\}$ be a family of identifiable distributions that are Markov with respect to a decomposable graph $\mathcal{G} = (V, E)$. Suppose $X \sim P_\theta$. For $A \subset V$, let θ_A denote the parameter involved in the marginal distribution of X_A. Similarly, for $A, B \subset V$, let $\theta_{B|A}$ be the parameter involved in the conditional distribution of $X_B|X_A$ (at any value $X_A = x_A$). Note that the mapping $\theta \mapsto (\theta_A, \theta_{A|B})$ is a bijection [7, Lemma 3.1]. Recall that (A, B) is said to be a decomposition of $\mathcal{G}$ if $A \cup B = V$ and $A \cap B$ separates A from B.

Definition A.2. Let $\mathcal{L}$ denote the law of a fixed probability distribution on Θ. The law $\mathcal{L}$ is said to be weak hyper Markov with respect to $\mathcal{G}$ if for any decomposition (A, B) of $\mathcal{G}$, under the mapping $\theta \mapsto (\theta_A, \theta_{A|B})$,

$$\theta_A \perp\!\!\!\perp \theta_B | \theta_{A\cap B}.$$

The law $\mathcal{L}$ is said to be strong hyper Markov with respect to $\mathcal{G}$ if for any decomposition (A, B) of $\mathcal{G}$,

$$\theta_A \perp\!\!\!\perp \theta_{B|A}.$$

A.3. The domain of parameters for Type I and Type II Wishart distributions on homogeneous graphs

We consider the domain of parameters $\mathcal{A}$ and $\mathcal{B}$ for the Type I and II Wishart distributions when the graph is homogeneous. For any homogeneous graph $\mathcal{H}$, let $\mathcal{T}_{\mathcal{H}} = (T, \mathscr{V}_{\mathcal{H}}, \preceq)$ be the Hasse tree of $\mathcal{H}$ (see [19, §2.2] for greater detail). Note that a node in T is indeed an equivalent class $\bar{i} \subseteq V$ for some node $i \in V$, where $j \in \bar{i}$ if and only if $\mathrm{cl}(j) = \mathrm{cl}(i)$. The relation $\bar{j} \preceq \bar{i}$ holds if and only if $\mathrm{cl}(j) \subseteq \mathrm{cl}(i)$ and defines a partial order on T. If $t \in T$ is an internal node, that is, a node that has a child, then $C_t := \bigcup\{u \in T : u \preceq t\}$ is a clique of $\mathcal{H}$. Also if $q \in T$ is a leaf node, that is, has no child, then $S_q := \bigcup \{u \in T :\ u \preceq q\}$ is a separator of $\mathcal{H}$. For each $u \in T$ define

$$\rho_u(\alpha,\beta) = \sum_{u \preceq t} \alpha(C_t) - \sum_{u \preceq q} \nu(S_q)\,\beta(S_q),$$
$$m_u := \sum_{t \preceq u} n_t,$$

where n_t is the number of the elements in the equivalent class t. By Theorem 3.1 and Theorem 3.2 in [19],

$$\mathcal{A} = \left\{(\alpha,\beta)\ :\ \rho_u(\alpha,\beta) > \frac{1}{2}\sum_{u \preceq t} n_t - \frac{1}{2},\ \ \forall u \in T\right\} \tag{A.14}$$

$$\mathcal{B} = \left\{(\alpha,\beta)\ :\ \rho_u(\alpha,\beta) < \frac{1}{2} - \frac{1}{2}\sum_{u \preceq t} n_t,\ \ \forall u \in T\right\} \tag{A.15}$$

A.4. Constructing a decomposable graph with nested separators from a homogenous graph

Let $\mathcal{H}$ be a homogeneous graph with two or more separators. Proposition 2.2 in [19] guarantees that $\mathcal{H}$ has at least two nested separators. Since $\mathcal{H}$ is not complete it contains an induced path of length 2, say $v_1 - v_2 - v_3$. Let us augment $\mathcal{H}$ to obtain a graph $\mathcal{G}$ by adding a vertex v_0 and an edge between v_0 and v_1. It is clear that $\mathcal{G}$ is decomposable since this construction creates no cycles. However, this addition creates the induced path $v_0 - v_1 - v_2 - v_3$, thus $\mathcal{G}$ is not homogeneous. Moreover, $\mathcal{G}$ contains nested separators since two nested separators in $\mathcal{H}$ remain nested separators in $\mathcal{G}$.

A.5. Evaluating the integral in Equation (4.7)

Let $\mathcal{D}$ be a DAG and let $\Theta_{\mathcal{D}}$ denote the image of $\mathrm{P}_{\mathcal{D}}$ under the mapping $\Omega \mapsto (D, L)$, where (D, L) denotes the Cholesky factor of Ω, as defined in Appendix A.1.

Proposition A.2. *Let $dL = \prod_{(i,j)\in E, i>j} dL_{ij}$ and $dD = \prod_{i=1}^{p} dD_{ii}$. Then for any $\alpha = (\alpha_1, \dots, \alpha_p)$ and $U \succ 0$,*

$$z_{\mathcal{D}}(U,\alpha) = \int_{\Theta_{\mathcal{D}}} \exp\left\{-\frac{1}{2}\mathrm{tr}(LD^{-1}L^{\top}U)\right\} \prod_{i=1}^{p} D_{ii}^{-\alpha_i/2} dL dD < \infty,$$

if and only if $\alpha_i > pa_i + 2$ for each $i = 1, \dots, p$. Furthermore, in this case

$$z_{\mathcal{D}}(U,\alpha) = \prod_{i=1}^{p} \frac{\Gamma\left(\frac{\alpha_i}{2} - \frac{pa_i}{2} - 1\right) 2^{\alpha_i/2-1} (\sqrt{\pi})^{pa_i} \det(U_{\mathrm{pa}(i)})^{\alpha_i/2 - pa_i/2 - 3/2}}{\det(U_{\mathrm{fa}(i)})^{\alpha_i/2 - pa_i/2 - 1}}. \tag{A.16}$$

Proof. First, we integrate out the terms involving D_{ii}'s.

$$\int \exp\left[-\frac{1}{2}\mathrm{tr}\left\{\left(LD^{-1}L^\top\right)U\right\}\right]\prod_{i=1}^p D_{ii}^{-\alpha_i/2}dLdD$$
$$= \int \exp\left[-\frac{1}{2}\mathrm{tr}\left\{D^{-1}\left(L^\top UL\right)\right\}\right]\prod_{i=1}^p D_{ii}^{-\alpha_i/2}dLdD$$
$$= \int \exp\left\{-\frac{1}{2}\sum_{i=1}^p D_{ii}^{-1}(L^\top UL)_{ii}\right\}\prod_{i=1}^p D_{ii}^{-\alpha_i/2}dDdL$$
$$= \int \left[\prod_{i=1}^p \int \exp\left\{-\frac{1}{2}D_{ii}^{-1}(L^\top UL)_{ii}\right\} D_{ii}^{-\alpha_i/2}dD_{ii}\right]dL$$
$$= \int \prod_{i=1}^p \frac{\Gamma\left(\frac{\alpha_i}{2}-1\right)2^{\alpha_i/2-1}}{\left((L^\top UL)_{ii}\right)^{\alpha_i/2-1}}dL \qquad \text{(if and only if } \alpha_i > 2 \text{ for each } i = 1,2,\ldots,p)$$
$$= \int \prod_{i=1}^p \frac{\Gamma\left(\frac{\alpha_i}{2}-1\right)2^{\alpha_i/2-1}}{\left((L_{\cdot i})^\top UL_{\cdot i}\right)^{\alpha_i/2-1}}dL$$
$$= \int \prod_{i=1}^p \frac{\Gamma\left(\frac{\alpha_i}{2}-1\right)2^{\alpha_i/2-1}}{\left\{\begin{pmatrix}1 & L_{\mathrm{pa}(i),i}^\top\end{pmatrix}\begin{pmatrix}U_{ii} & U_{i,\mathrm{pa}(i)}\\ U_{\mathrm{pa}(i),i} & U_{\mathrm{pa}(i)}\end{pmatrix}\begin{pmatrix}1\\ L_{\mathrm{pa}(i),i}\end{pmatrix}\right\}^{\alpha_i/2-1}}dL$$
$$= \prod_{i=1}^p \int_{\mathbb{R}^{pa_i}} \frac{\Gamma\left(\frac{\alpha_i}{2}-1\right)2^{\alpha_i/2-1}}{\left\{\begin{pmatrix}1 & L_{\mathrm{pa}(i),i}^\top\end{pmatrix}\begin{pmatrix}U_{ii} & U_{i,\mathrm{pa}(i)}\\ U_{\mathrm{pa}(i),i} & U_{\mathrm{pa}(i)}\end{pmatrix}\begin{pmatrix}1\\ L_{\mathrm{pa}(i),i}\end{pmatrix}\right\}^{\alpha_i/2-1}}dL_{\mathrm{pa}(i),i}. \tag{A.17}$$

We evaluate this integral by considering a more general form:

$$\int_{\mathbb{R}^d} \frac{dx}{\left\{\begin{pmatrix}1 & x^\top\end{pmatrix}\begin{pmatrix}\lambda & b^\top\\ b & A\end{pmatrix}\begin{pmatrix}1\\ x\end{pmatrix}\right\}^\gamma},$$

where the block partitioned matrix, formed by $\lambda \in \mathbb{R}$, $b \in \mathbb{R}^d$ and the $(d-1)\times(d-1)$ matrix A, is positive definite. To simplify the integral above we proceed in two steps.

1) When $x \in \mathbb{R}$, by the formula provided on [8, page 16] we have the one dimensional integral

$$\int_{\mathbb{R}} \frac{dx}{(1+x^2)^\gamma} = \begin{cases}\frac{\sqrt{\pi}\Gamma\left(\gamma-\frac{1}{2}\right)}{\Gamma(\gamma)} & \gamma > \frac{1}{2},\\ \infty & \text{otherwise.}\end{cases}$$

The d-dimensional version of this integral by repeated application of the right-hand-side formula is computed as

$$\int_{\mathbb{R}^d} \frac{dx}{(1+x^\top x)^\gamma} = \begin{cases}\frac{(\sqrt{\pi})^d\Gamma\left(\gamma-\frac{d}{2}\right)}{\Gamma(\gamma)} & \gamma > \frac{d}{2},\\ \infty & \text{otherwise.}\end{cases}$$

2) Consider the general integral

$$\int_{\mathbb{R}^d} \frac{dx}{\left\{ \begin{pmatrix} 1 & x^\top \end{pmatrix} \begin{pmatrix} \lambda & b^\top \\ b & A \end{pmatrix} \begin{pmatrix} 1 \\ x \end{pmatrix} \right\}^\gamma}.$$

Under the linear transformation $y = A^{1/2}x + A^{-1/2}b$, for $\gamma > \frac{d}{2}$ we have

$$\begin{aligned} \int_{\mathbb{R}^d} \frac{dx}{\left\{ \begin{pmatrix} 1 & x^\top \end{pmatrix} \begin{pmatrix} \lambda & b^\top \\ b & A \end{pmatrix} \begin{pmatrix} 1 \\ x \end{pmatrix} \right\}^\gamma} &= \frac{1}{\det(A)^{1/2}} \int_{\mathbb{R}^d} \frac{1}{(y^\top y + a - b^\top A^{-1} b)^\gamma} dy \\ &= \frac{(\sqrt{\pi})^d \Gamma\left(\gamma - \frac{d}{2}\right)}{\Gamma(\gamma) \det(A)^{1/2} \left(a - b^\top A^{-1} b\right)^{\gamma - d/2}}. \end{aligned} \tag{A.18}$$

By applying (A.18) to the integral in (A.17) we obtain

$$\begin{aligned} & z_{\mathcal{D}}(U, \alpha) \\ &= \prod_{i=1}^{p} \int_{\mathbb{R}^{pa_i}} \frac{\Gamma\left(\frac{\alpha_i}{2} - 1\right) 2^{\alpha_i/2-1}}{\left\{ \begin{pmatrix} 1 & L_{\mathrm{pa}(i),i}^\top \end{pmatrix} \begin{pmatrix} U_{ii} & U_{i,\mathrm{pa}(i)} \\ U_{\mathrm{pa}(i),i} & U_{\mathrm{pa}(i)} \end{pmatrix} \begin{pmatrix} 1 \\ L_{\mathrm{pa}(i),i} \end{pmatrix} \right\}^{\alpha_i/2-1}} dL_{\mathrm{pa}(i),i} \\ &= \prod_{i=1}^{p} \frac{\Gamma\left(\frac{\alpha_i}{2} - \frac{pa_i}{2} - 1\right) 2^{\alpha_i/2-1} (\sqrt{\pi})^{pa_i} \det(U_{\mathrm{pa}(i)})^{\alpha_i/2 - pa_i/2 - 3/2}}{\det(U_{\mathrm{fa}(i)})^{\alpha_i/2 - pa_i/2 - 1}}, \end{aligned}$$

where $\det(U_{\mathrm{pa}(i)}) = 1$ whenever $pa(i) = \emptyset$. Thus $z_{\mathcal{D}}(U, \alpha)$ is finite if and only if $\alpha_i > pa_i + 2$ for each $i = 1, \ldots, p$. □

Lemma A.5 ([22, 16])**.** *The Jacobian of the mapping* $\psi : (D, L) \mapsto \left(LD^{-1}L^\top\right)$ *is* $\prod_{j=1}^{p} D_{jj}^{-(pa_j+2)}$.

Corollary A.2. *Let*

$$\widehat{\pi}_{\mathcal{D}}(\alpha, \beta, U, d\Omega) = \exp\left\{ -\frac{1}{2} \mathrm{tr}\left(\Omega U\right) \right\} \prod_{i=1}^{p} D_{jj}^{-\frac{1}{2}\eta_j + pa_j + 2} d\Omega,$$

then

$$\int_{\mathrm{P}_{\mathcal{D}}} \widehat{\pi}_{\mathrm{P}_{\mathcal{D}}}\left(\eta, U, d\Omega\right) = \prod_{j=1}^{p} \frac{\Gamma\left(\frac{\eta_j}{2} - \frac{pa_j}{2} - 1\right) 2^{\frac{\eta_j}{2} - 1} \left(\sqrt{\pi}\right)^{pa_j} \det\left(U_{pa(j)}\right)^{\frac{\eta_j}{2} - \frac{pa_j}{2} - \frac{3}{2}}}{\det\left(U_{fa(j)}\right)^{\frac{\eta_j}{2} - \frac{pa_j}{2} - 1}},$$

for $\eta_j > pa_j + 2$.

Proof. The result is immediate by change of variables using Lemma A.5 and then Proposition A.2. □