The Annals of Statistics
2018, Vol. 46, No. 1, 90–118
https://doi.org/10.1214/17-AOS1543

EXACT FORMULAS FOR THE NORMALIZING CONSTANTS OF WISHART DISTRIBUTIONS FOR GRAPHICAL MODELS

BY CAROLINE UHLER[1], ALEX LENKOSKI[2] AND DONALD RICHARDS[3]

Massachusetts Institute of Technology, Norwegian Computing Center and Penn State University

Gaussian graphical models have received considerable attention during the past four decades from the statistical and machine learning communities. In Bayesian treatments of this model, the G-Wishart distribution serves as the conjugate prior for inverse covariance matrices satisfying graphical constraints. While it is straightforward to posit the unnormalized densities, the normalizing constants of these distributions have been known only for graphs that are chordal, or decomposable. Up until now, it was unknown whether the normalizing constant for a general graph could be represented explicitly, and a considerable body of computational literature emerged that attempted to avoid this apparent intractability. We close this question by providing an explicit representation of the G-Wishart normalizing constant for general graphs.

1. Introduction. Let $G = (V, E)$ be an undirected graph with vertex set $V = \{1, \ldots, p\}$ and edge set E. Let $\mathbb{S}^p$ be the set of symmetric $p \times p$ matrices and $\mathbb{S}^p_{\succ 0}$ the cone of positive definite matrices in $\mathbb{S}^p$. Let

$$\mathbb{S}^p_{\succ 0}(G) = \{M = (M_{ij}) \in \mathbb{S}^p_{\succ 0} \mid M_{ij} = 0 \text{ for all } (i, j) \notin E\} \tag{1.1}$$

denote the cone in $\mathbb{S}^p$ of positive definite matrices with zeros in all entries not corresponding to edges in the graph. Note that the positivity of all diagonal entries M_{ii} follows from the positive-definiteness of the matrices M.

A random vector $X \in \mathbb{R}^p$ is said to *satisfy the Gaussian graphical model (GGM) with graph G* if X has a multivariate normal distribution with mean μ and covariance matrix Σ, denoted $X \sim \mathcal{N}_p(\mu, \Sigma)$, where $\Sigma^{-1} \in \mathbb{S}^p_{\succ 0}(G)$. The inverse co-

Received June 2014; revised January 2017.

[1]Supported by the Austrian Science Fund (FWF) Y 903-N35.

[2]Supported by Statistics for Innovation *sfi*[2] in Oslo.

[3]Supported in part by the U.S. National Science Foundation Grant DMS-13-09808; and by a Romberg Guest Professorship at the Heidelberg University Graduate School for Mathematical and Computational Methods in the Sciences, funded by German Universities Excellence Initiative Grant GSC 220/2.

MSC2010 subject classifications. Primary 62H05, 60E05; secondary 62E15.

Key words and phrases. Bartlett decomposition, bipartite graph, Cholesky decomposition, chordal graph, directed acyclic graph, G-Wishart distribution, Gaussian graphical model, generalized hypergeometric function of matrix argument, moral graph, normalizing constant, Wishart distribution.

variance matrix Σ^{-1} is called the *concentration matrix* and, throughout this paper, we denote Σ^{-1} by K.

Statistical inference for the concentration matrix K constrained to $\mathbb{S}^p_{\succ 0}(G)$ goes back to [6], who proposed an algorithm for determining the maximum likelihood estimator (cf. [33]). A Bayesian framework for this problem was introduced by [5], who proposed the Hyper-Inverse Wishart (HIW) prior distribution for *chordal* (also known as *decomposable* or *triangulated*) graphs G.

Chordal graphs enjoy a rich set of properties that led the HIW distribution to be particularly amenable to Bayesian analysis. Indeed, for nearly a decade after the introduction of GGMs, focus on the Bayesian use of GGMs was placed primarily on chordal graphs (see, e.g., [11]). This tractability stems from two causes: the ability to sample directly from HIWs [30], and the ability to calculate their normalizing constants.

Roverato [31] extended the HIW distribution to general G. Following [23, 24] termed this distribution the *G-Wishart distribution*. Atay-Kayis and Massam [2] developed a Monte Carlo method to compute numerically the normalizing constant of this distribution. For $D \in \mathbb{S}^p_{\succ 0}(G)$ and $\delta \in \mathbb{R}$, the G-Wishart density has the form

$$f_G(K \mid \delta, D) \propto |K|^{\frac{1}{2}(\delta-2)} \exp\biggl(-\frac{1}{2}\operatorname{tr}(KD)\biggr)\mathbf{1}_{K \in \mathbb{S}^p_{\succ 0}(G)}.$$

This distribution is conjugate [31] and proper for $\delta > 1$ [26].

Early work on the G-Wishart distribution was largely computational in nature [2, 4, 7, 8, 18, 23, 26, 35, 36] and was predicated on two assumptions: first, that a direct sampler was unavailable for this class of models and, second, that the normalizing constant could not be calculated explicitly. Lenkoski [22], motivated by the algorithm of [6], developed a direct sampler for G-Wishart variates and thereby resolved the first open question. In this paper, we close the second question by deriving for general graphs G an explicit formula for the G-Wishart normalizing constant

$$C_G(\delta, D) = \int_{\mathbb{S}^p_{\succ 0}(G)} |K|^{\frac{1}{2}(\delta-2)} \exp\biggl(-\frac{1}{2}\operatorname{tr}(KD)\biggr)\, \mathrm{d}K,$$

where

$$\mathrm{d}K = \prod_{i=1}^{p} \mathrm{d}k_{ii} \cdot \prod_{i<j,(i,j)\in E} \mathrm{d}k_{ij}$$

denotes the product of differentials corresponding to all distinct nonzero entries in K.

For notational simplicity, we will consider the integral

$$I_G(\delta, D) = \int_{\mathbb{S}^p_{\succ 0}(G)} |K|^{\delta} \exp\bigl(-\operatorname{tr}(KD)\bigr)\, \mathrm{d}K,$$

which can be expressed in terms of $C_G(\delta, D)$ as follows: Denote by $|E|$ the cardinality of the edge set E; by changing variables, $K \to 2K$, one obtains

$$C_G(\delta, D) = 2^{\frac{1}{2}p\delta+|E|} I_G\left(\frac{1}{2}(\delta-2), D\right).$$

The normalizing constant $I_G(\delta, D)$ is well known for *complete graphs*, in which every pair of vertices is connected by an edge. In such cases,

$$I_{\text{complete}}(\delta, D) = |D|^{-(\delta+\frac{1}{2}(p+1))} \Gamma_p\left(\delta + \frac{1}{2}(p+1)\right), \tag{1.2}$$

where

$$\Gamma_p(\alpha) = \pi^{p(p-1)/4} \prod_{i=1}^{p} \Gamma\left(\alpha - \frac{1}{2}(i-1)\right), \tag{1.3}$$

$\operatorname{Re}(\alpha) > \frac{1}{2}(p-1)$ is the *multivariate gamma function*. The formula (1.2) has a long history, dating back to [16, 37, 38], *Hilfssatz* 37 of [32], [25] and many derivations of a statistical nature; see [29] and [10], page 224.

As noted above, $I_G(\delta, D)$ is also known for *chordal graphs*. Let G be chordal, and let $(T_1, \ldots, T_d)$ denote a *perfect sequence* of cliques (i.e., complete subgraphs) of V. Further, let $S_i = (T_1 \cup \cdots \cup T_i) \cap T_{i+1}$, $i = 1, \ldots, d-1$; then $S_1, \ldots, S_{d-1}$ are called the *separators* of G. Note that the separators S_i are cliques as well. We denote the cardinalities by $t_i = |T_i|$ and $s_i = |S_i|$. For $S \subseteq \{1, \ldots, p\}$, let D_{SS} denote the submatrix of D corresponding to the rows and columns in S. Then

$$\begin{aligned} I_G(\delta, D) &= \frac{\prod_{i=1}^{d} I_{T_i}(\delta, D_{T_iT_i})}{\prod_{j=1}^{d-1} I_{S_j}(\delta, D_{S_jS_j})} \\ &= \frac{\prod_{i=1}^{d}(|D_{T_iT_i}|^{-(\delta+\frac{1}{2}(t_i+1))}\Gamma_{t_i}(\delta+\frac{1}{2}(t_i+1)))}{\prod_{j=1}^{d-1}(|D_{S_jS_j}|^{-(\delta+\frac{1}{2}(s_j+1))}\Gamma_{s_j}(\delta+\frac{1}{2}(s_j+1)))}. \end{aligned} \tag{1.4}$$

This result follows because, for a chordal graph G, the G-Wishart density function can be factored into a product of density functions [5].

For nonchordal graphs, the problem of calculating $I_G(\delta, D)$ has been open for over 20 years, and much of the computational methodology mentioned above was developed with the objective of either approximating $I_G(\delta, D)$ or avoiding its calculation. Our result shows that an explicit representation of this quantity is indeed possible.

In deriving the explicit formula for the normalizing constant $I_G(\delta)$, we utilize methods that are familiar to researchers in this area. These methods include the Cholesky decomposition or the Bartlett decomposition of a positive definite matrix, Schur complements for factorizing determinants and the chordal cover of a graph. Furthermore, we make crucial use of certain formulas from the theory of generalized hypergeometric functions of matrix argument [14, 17], and analytic continuation of differential operators on the cone of positive definite matrices [9].

The article proceeds as follows. In Section 2, we treat the case in which $D = \mathbb{I}_p$, the $p \times p$ identity matrix, deriving a closed-form product formula for the normalizing constant $I_G(\delta, \mathbb{I}_p)$ for various classes of nonchordal graphs. In Section 3, we consider the case of general matrices D; in our main result in Theorem 3.3, we derive an explicit representation of $I_G(\delta, D)$ for general graphs as a closed-form product formula involving differentials of principal minors of D. We end with a brief discussion in Section 4.

2. Computing the normalizing constant $I_G(\delta, \mathbb{I}_p)$. In this section, we compute $I_G(\delta, \mathbb{I}_p)$ for two classes of nonchordal graphs. We begin in Section 2.1 with the class of complete bipartite graphs and use an approach based on Schur complements to attain a closed-form formula. In Section 2.2, we introduce directed Gaussian graphical models and show how these models relate to a Cholesky factor approach to computing $I_G(\delta, \mathbb{I}_p)$. This leads to a formula for computing normalizing constants of graphs with *minimum fill-in* equal to 1, namely graphs that become chordal after the addition of one edge. However, these approaches do not lead to a general formula for the normalizing constant in the case $D = \mathbb{I}_p$. To obtain a formula for any graph G, we found it necessary to calculate the more general case $I_G(\delta, D)$ and then specialize $D = \mathbb{I}_p$, as is done for moment generating functions or Laplace transforms. This is explained in Section 3.

2.1. *Bipartite graphs.* A *complete bipartite graph* on $m + n$ vertices, denoted by $H_{m,n}$, is an undirected graph whose vertices can be divided into disjoint sets $U = \{1, \dots, m\}$ and $V = \{m + 1, \dots, m + n\}$, such that each vertex in U is connected to every vertex in V, but there are no edges within U or V. For the graph $H_{m,n}$, the corresponding matrix K is a block matrix,

$$K = \begin{pmatrix} K_{AA} & K_{AB} \\ K_{AB}^T & K_{BB} \end{pmatrix},$$

where K_{AA}, K_{BB} are diagonal matrices of sizes $m \times m$ and $n \times n$, respectively, and K_{AB} is *unconstrained*, that is, no entry of K_{AB} is constrained to be zero.

PROPOSITION 2.1. *The integral $I_{H_{m,n}}(\delta, \mathbb{I}_{m+n})$ converges absolutely for all $\delta > -1$, and*

$$\begin{aligned} I_{H_{m,n}}(\delta, \mathbb{I}_{m+n}) &= \left[\Gamma\left(\delta + \frac{1}{2}n + 1\right)\right]^m \left[\Gamma\left(\delta + \frac{1}{2}m + 1\right)\right]^n \\ &\quad \times \frac{\Gamma_{m+n}(\delta + \frac{1}{2}(m+n+1))}{\Gamma_m(\delta + \frac{1}{2}(m+n+1))\Gamma_n(\delta + \frac{1}{2}(m+n+1))}. \end{aligned} \tag{2.1}$$

PROOF. Applying the Schur complement determinant formula for block matrices,

$$|K| = |K_{AA}|\left|K_{BB} - K_{AB}^T (K_{AA})^{-1} K_{AB}\right|, \tag{2.2}$$

we obtain

$$\begin{aligned} I_{H_{m,n}}(\delta, \mathbb{I}_{m+n}) &= \int_{\mathbb{S}_{\succ 0}^{m+n}(G)} |K|^{\delta} \exp\bigl(-\operatorname{tr}(K)\bigr)\,\mathrm{d}K \\ &= \int_{\mathbb{S}_{\succ 0}^{m+n}(G)} |K_{AA}|^{\delta}\left|K_{BB} - K_{AB}^T (K_{AA})^{-1} K_{AB}\right|^{\delta} \\ &\quad \cdot \exp\bigl(-\operatorname{tr}(K_{AA}) - \operatorname{tr}(K_{BB})\bigr)\,\mathrm{d}K_{AA}\,\mathrm{d}K_{AB}\,\mathrm{d}K_{BB}. \end{aligned}$$

Since K_{AB} is unconstrained, we can change variables by replacing K_{AB} by $K_{AA}^{1/2} K_{AB} K_{BB}^{1/2}$; then the corresponding Jacobian is $|K_{AA}|^{n/2}|K_{BB}|^{m/2}$. Since

$$\left|K_{BB} - K_{BB}^{1/2} K_{AB}^T K_{AB} K_{BB}^{1/2}\right| = |K_{BB}| \cdot \left|\mathbb{I}_n - K_{AB}^T K_{AB}\right|,$$

we obtain

$$\begin{aligned} I_{H_{m,n}}(\delta, \mathbb{I}_{m+n}) &= \int_{\mathbb{S}_{\succ 0}^{m+n}(G)} |K_{AA}|^{\delta + \frac{1}{2}n} |K_{BB}|^{\delta + \frac{1}{2}m} \left|\mathbb{I}_n - K_{AB}^T K_{AB}\right|^{\delta} \\ &\quad \cdot \exp\bigl(-\operatorname{tr}(K_{AA}) - \operatorname{tr}(K_{BB})\bigr)\,\mathrm{d}K_{AA}\,\mathrm{d}K_{AB}\,\mathrm{d}K_{BB}, \end{aligned}$$

where the range of integration is such that each diagonal entry of K_{AA} and K_{BB} is positive, K_{AB} is unconstrained and $\mathbb{I}_n - K_{AB}^T K_{AB}$ is positive definite. Integrating over each diagonal entry of K_{AA} and K_{BB}, we obtain

$$\begin{aligned} I_{H_{m,n}}(\delta, \mathbb{I}_{m+n}) &= \left[\Gamma\left(\delta + \frac{1}{2}n + 1\right)\right]^m \left[\Gamma\left(\delta + \frac{1}{2}m + 1\right)\right]^n \\ &\quad \times \int_{K_{AB}} \left|\mathbb{I}_n - K_{AB}^T K_{AB}\right|^{\delta} \mathrm{d}K_{AB}. \end{aligned}$$

Finally, since K_{AB} is unconstrained, we deduce from (3.4) the value of the latter integral. □

In this computation, we used the special structure of the graph to decompose the inverse covariance matrix K into a special block matrix. In Section 3, we use a similar approach to show how the normalizing constant changes when removing a clique (i.e., a completely connected subgraph) from a graph. This leads to an algorithm for computing the normalizing constant $I_G(\delta, D)$ for any graph G. In the remainder of this section, we show how an approach based on the Cholesky factorization of K can be used to easily compute the normalizing constant for graphs that have minimum fill-in equal to 1. This requires introducing directed Gaussian graphical models.

2.2. *Directed Gaussian graphical models.* Let $\mathcal{G} = (V, \mathcal{E})$ be a directed acyclic graph (DAG) consisting of vertices $V = \{1, \dots, p\}$ and directed edges $\mathcal{E}$. We assume, without loss of generality, that the vertices in $\mathcal{G}$ are *topologically ordered*, meaning that $i < j$ for all $(i, j) \in \mathcal{E}$. We associate to $\mathcal{G}$ a strictly upper-triangular matrix B of edge weights. So $B = (b_{ij})$ with $b_{ij} \neq 0$ if and only if $(i, j) \in \mathcal{E}$. Then a *directed Gaussian graphical model* on $\mathcal{G}$ for a random variable $X \in \mathbb{R}^p$ is defined by $X \sim \mathcal{N}_p(0, \Sigma)$ with $\Sigma = [(I - B)D(I - B)^T]^{-1}$, where D is a diagonal matrix.

To simplify notation, let $a_{ii} = d_{ii}$ and $a_{ij} = -b_{ij}\sqrt{d_{jj}}$, and let $A = (A_{ij})$ with $A_{ii} = \sqrt{a_{ii}}$ and $A_{ij} = -a_{ij}$ for all $i \neq j$. Then $\Sigma^{-1} = AA^T$, and $a_{ij} \neq 0$ for $i \neq j$ if and only if $(i, j) \in \mathcal{E}$. Note that AA^T is the upper Cholesky decomposition of Σ^{-1}. Such a decomposition exists for any positive definite matrix and is unique.

We will associate to a DAG, $\mathcal{G} = (V, \mathcal{E})$, and its corresponding directed Gaussian graphical model two undirected graphs. We denote by $\mathcal{G}^s = (V, \mathcal{E}^s)$ the *skeleton* of $\mathcal{G}$ obtained by replacing all directed edges in $\mathcal{G}$ by undirected edges. We denote by $\mathcal{G}^m = (V, \mathcal{E}^m)$ the *moral graph* of $\mathcal{G}$, which reflects the conditional independencies in $\mathcal{N}_p(0, \Sigma)$, that is,

$$(i, j) \notin \mathcal{E}^m \qquad \text{if and only if} \qquad X_i \perp\!\!\!\perp X_j \mid X_{V \setminus \{i,j\}}.$$

Since Σ^{-1} also encodes the conditional independence relations of the form $X_i \perp\!\!\!\perp X_j \mid X_{V \setminus \{i,j\}}$, this is equivalent to the criterion

$$(i, j) \notin \mathcal{E}^m \qquad \text{if and only if} \qquad (\Sigma^{-1})_{ij} = 0.$$

So, the moral graph $\mathcal{G}^m$ reflects the zero pattern of Σ^{-1}.

The moral graph of $\mathcal{G}$ can also be defined graph-theoretically. It is formed by connecting all nodes $i, j \in V$ that have a common child in $\mathcal{G}$, that is, for which there exists a node $k \in V \setminus \{i, j\}$ such that $(i, k), (j, k) \in \mathcal{E}$, and then making all edges in the graph undirected. The name stems from the fact that the moral graph is obtained by "marrying" the parents. For a review of basic graph-theoretic concepts see, for example, [21], Chapter 2.

The moral graph is an important concept for our application. Let $G = (V, E)$ be an undirected graph, with $V = \{1, \dots, p\}$, for which we want to compute $I_G(\delta, \mathbb{I}_p)$. Let $G_0 = (V, E_0)$ with $G_0 = G$. Given a labeling of the vertices V, we associate a DAG, $\mathcal{G}_0 = (V, \mathcal{E}_0)$, to G_0 by orienting the edges in E_0 according to the topological ordering, that is, for all $(i, j) \in E_0$ let $(i, j) \in \mathcal{E}_0$ if $i < j$. Note that the skeleton of $\mathcal{G}_0$ is the original undirected graph G_0. Let $G_1 = (V, E_1)$ be the moral graph of $\mathcal{G}_0$, that is, $G_1 = \mathcal{G}_0^m$, and let $\mathcal{G}_1 = (V, \mathcal{E}_1)$ be the corresponding DAG obtained by orienting the edges in E_1 according to the ordering of the vertices V. So $\mathcal{G}_0$ is a subgraph of $\mathcal{G}_1$. We repeat this procedure until $\mathcal{G}_{q+1} = \mathcal{G}_q$. This results in a sequence of DAGs,

$$\mathcal{G}_0 \subsetneq \mathcal{G}_1 \subsetneq \cdots \subsetneq \mathcal{G}_q.$$

In the following, we denote by $\mathcal{G} = (V, \mathcal{E})$ the DAG associated to $G = (V, E)$ obtained by orienting the edges in E according to the ordering of the vertices V. We denote by $\bar{\mathcal{G}} = (V, \bar{\mathcal{E}})$ the DAG associated to $G = (V, E)$ obtained by repeatedly marrying parents in $\mathcal{G}$, that is, $\bar{\mathcal{G}} = \mathcal{G}_q$. We call $\bar{\mathcal{G}}$ the *moral DAG* of G. Note that $\bar{\mathcal{G}}^s$, the skeleton of $\bar{\mathcal{G}}$, is a chordal graph with $G \subset \bar{\mathcal{G}}^s$ ([21], Chapter 2), so $\bar{\mathcal{G}}^s$ is a *chordal cover* of G. A chordal cover in general is not unique; however, $\bar{\mathcal{G}}^s$ is the unique chordal cover obtained by repeatedly marrying parents according to the vertex labeling V. We call this chordal cover the *moral chordal graph* of G and denote it by $\bar{G} = (V, \bar{E})$.

We now show how to deduce from the undirected graph $G = (V, E)$ the normalizing constant $I_G(\delta, \mathbb{I}_p)$ as an integral in terms of the Cholesky factor A. Atay-Kayis and Massam ([2], equation (39)) also give an integral formula for $I_G(\delta, D)$ in terms of the Cholesky factor of the precision matrix K. New in the following result is the use of $\bar{\mathcal{G}}$ for the parametrization, which is crucial in order to determine orderings of the vertices that lead to simpler integral formulas and ultimately allow us to obtain a closed-form formula. Since the proof is the same for general correlation matrices $D \in \mathbb{S}^p_{\succ 0}$, we give the result directly for $I_G(\delta, D)$. In the following, we use the standard graph-theoretic notation $\operatorname{indeg}(i)$ for the *indegree* of node i, representing the number of edges "arriving at" (or "pointing to") node i in a DAG $\mathcal{G}$.

THEOREM 2.2. *Let $G = (V, E)$ be an undirected graph with vertices $V = \{1, \ldots, p\}$. Let $\mathcal{G} = (V, \mathcal{E})$ be the DAG associated to $G = (V, E)$ obtained by orienting the edges in E according to the ordering of the vertices in V. Let $\bar{\mathcal{G}} = (V, \bar{\mathcal{E}})$ denote the moral DAG of G and $\bar{G} = (V, \bar{E})$ its skeleton, the moral chordal graph of G. Let A be an upper-triangular $p \times p$ matrix with diagonal entries $A_{ii} = \sqrt{a_{ii}}$ and off-diagonal entries $A_{ij} = -a_{ij}$ for all $i < j$. Then*

$$I_G(\delta, D) = \int_{A_*} \left(\prod_{i=1}^{p} a_{ii}^{\delta + \frac{1}{2}\operatorname{indeg}(i)} \right) \exp\left[-\sum_{i=1}^{p} \left(a_{ii} + \sum_{j:(i,j)\in\bar{\mathcal{E}}} a_{ij}^2 \right) \right]$$
$$\cdot \exp\left[-2 \sum_{(i,j)\in\mathcal{E}} d_{ij} \left(-a_{ij}\sqrt{a_{jj}} + \sum_{l:(i,l),(j,l)\in\bar{\mathcal{E}}} a_{il}a_{jl} \right) \right] \mathrm{d}A_*,$$

where $D \in \mathbb{S}^p_{\succ 0}$ is a correlation matrix, $A_* = \{a_{ij} : i = j \text{ or } (i, j) \in \mathcal{E}\}$, the range of a_{ii} is $(0, \infty)$, the range of a_{ij} for $(i, j) \in \mathcal{E}$ is $(-\infty, \infty)$, $\operatorname{indeg}(i)$ denotes the indegree of node i in $\mathcal{G}$, and for $a_{ij} \notin A_*$,

$$a_{ij} = \begin{cases} 0 & \text{if } (i, j) \notin \bar{\mathcal{E}}, \\ \dfrac{1}{\sqrt{a_{jj}}} \displaystyle\sum_{\substack{l\in V \\ (i,l),(j,l)\in\bar{\mathcal{E}}}} a_{il}a_{jl} & \text{if } (i, j) \in \bar{\mathcal{E}} \setminus \mathcal{E}. \end{cases}$$

PROOF. Let $K \in \mathbb{S}^p_{\succ 0}(G)$. Since $G \subset \bar{G}$, then $K \in \mathbb{S}^p_{\succ 0}(\bar{G})$ and we can view K as an inverse covariance matrix of a directed Gaussian graphical model on $\bar{\mathcal{G}}$. Let A denote the unique upper Cholesky factor of K as described at the beginning of Section 2.2. Then one can verify that $a_{ij} = 0$ for all $(i, j) \notin \bar{\mathcal{E}}$.

Let (i, j) be an edge that is present in the moral chordal graph $\bar{G}$ but not in G. We can assume that $i < j$. Hence, $(i, j) \in \bar{\mathcal{E}} \setminus \mathcal{E}$ and, therefore,

$$0 = K_{ij} = (AA^T)_{ij} = -a_{ij}\sqrt{a_{jj}} + \sum_{l > \max(i,j)} a_{il}a_{jl}.$$

Thus, for each edge $(i, j) \in \bar{\mathcal{E}} \setminus \mathcal{E}$, we obtain an equation

$$a_{ij} = \frac{1}{\sqrt{a_{jj}}} \sum_{\substack{l \in V \\ (i,l),(j,l) \in \bar{\mathcal{E}}}} a_{il}a_{jl}.$$

To complete the proof, we need to compute the Jacobian J of the change of variables from K to A. We list the a_{ij}'s column-wise, meaning that a_{ij} precedes a_{lm} if $j < m$ or if $j = m$ and $i < l$, omitting a_{ij} for $(i, j) \notin \mathcal{E}$, corresponding to the zeros in K. We list the k_{ij}'s in the same ordering. Let the a_{ij}'s correspond to the columns of the Jacobian, while the k_{ij}'s correspond to the rows. In order to form J, we calculate the partial derivative of each k_{ij} with respect to each a_{lm}. Since $K = AA^T$ and A is upper-triangular, then J also is upper-triangular; therefore, $|J| = |\operatorname{diag}(J)|$. Since

$$k_{ii} = a_{ii} + \sum_{(i,j)\in\bar{\mathcal{E}}} a_{ij}^2 \quad \text{and} \quad k_{ij} = -a_{ij}\sqrt{a_{jj}} + \sum_{\substack{l \in V \\ (i,l),(j,l) \in \bar{\mathcal{E}}}} a_{il}a_{jl},$$

for all $(i, j) \in \mathcal{E}$, then

$$|J| = \prod_{i=1}^{p} a_{ii}^{\operatorname{indeg}(i)/2}.$$

Collecting together these formulas completes the proof. □

The number of edges in $\bar{\mathcal{E}} \setminus \mathcal{E}$ depend on the ordering of the vertices. It is well known (see, e.g., [21], Chapter 2) that one can find an ordering of the vertices such that $\bar{\mathcal{G}} = \mathcal{G}$ if and only if G is chordal. Hence, when G is chordal we can directly derive the normalizing constant of $I_G(\delta, \mathbb{I}_p)$ from Theorem 2.2 by evaluating Gaussian and Gamma integrals. One could also prove the following corollary using equation (1.4).

COROLLARY 2.3. *Let $G = (V, E)$ be a chordal graph, where the vertices $V = \{1, \dots, p\}$ are labeled according to a perfect ordering. Then*

$$I_G(\delta, \mathbb{I}_p) = \pi^{|E|/2} \prod_{i=1}^{p} \Gamma\left(\delta + \frac{1}{2}\operatorname{indeg}(i) + 1\right),$$

where $\operatorname{indeg}(i)$ *denotes the indegree of node i in the corresponding DAG $\mathcal{G}$.*

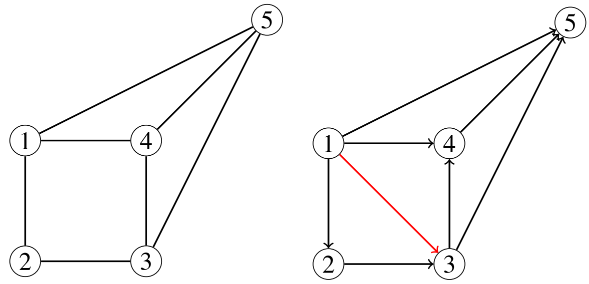

FIG. 1. *Undirected graph G_5 (left) discussed in Example 2.4 and its moral DAG $\bar{\mathcal{G}}_5$ (right).*

EXAMPLE 2.4. We illustrate Theorem 2.2 by studying the nonchordal graph G_5, shown in Figure 1(left). We wish to calculate

$$I_{G_5}(\delta, \mathbb{I}_5) = \int_{K \in \mathbb{S}^5_{\succ 0}(G_5)} |K|^\delta \exp(-\operatorname{tr}(K))\, \mathrm{d}K \tag{2.3}$$

through the change of variables, $K = AA^T$. The moral DAG of G_5 is denoted by $\bar{\mathcal{G}}_5$ and depicted in Figure 1(right). Since the edges $(2,4)$ and $(2,5)$ are missing in $\bar{\mathcal{G}}_5$, we immediately deduce that $a_{24} = a_{25} = 0$. In this example, we chose an ordering where only one edge needed to be added in the process of marrying parents, namely the edge $(1,3)$. This results in one equation for a_{13}, which can be deduced from the *colliders* over the additional edge, that is, nodes $l \in V$ with $(1,l), (3,l) \in \bar{\mathcal{G}}$, and results in

$$a_{13} = \frac{1}{\sqrt{a_{33}}}(a_{14}a_{34} + a_{15}a_{35}).$$

Finally, the Jacobian can be deduced from the indegrees of the nodes in $\mathcal{G}_5$, which corresponds to the moral DAG $\bar{\mathcal{G}}_5$ after omitting the red edge. Therefore, the determinant of the Jacobian is

$$a_{11}^{0/2} a_{22}^{1/2} a_{33}^{1/2} a_{44}^{2/2} a_{55}^{3/2},$$

and we find that the integral (2.3) equals

$$\int_A a_{11}^{\delta} a_{22}^{\delta+1/2} a_{33}^{\delta+1/2} a_{44}^{\delta+1} a_{55}^{\delta+3/2}$$
$$\times \exp\biggl[-\biggl(a_{11} + a_{12}^2 + \biggl(\frac{a_{14}a_{34} + a_{15}a_{35}}{\sqrt{a_{33}}}\biggr)^2 + a_{14}^2 + a_{15}^2$$
$$+ a_{22} + a_{23}^2 + a_{33} + a_{34}^2 + a_{35}^2 + a_{44} + a_{45}^2 + a_{55}\biggr)\biggr]\, \mathrm{d}A,$$

where $a_{ii} > 0$; $a_{ij} \in \mathbb{R}$, $i < j$; and $\mathrm{d}A$ denotes the product of all differentials.

As seen in Example 2.4, the equations corresponding to the additional edges $(i, j) \in \bar{\mathcal{E}} \setminus \mathcal{E}$ complicate the integral significantly. Therefore, given a nonchordal graph G, it is desirable to find an ordering such that $|\bar{\mathcal{E}} \setminus \mathcal{E}|$ is minimized. This ordering is given by a perfect ordering of a minimal chordal cover of G, where minimality is with respect to the number of edges that need to be added in order to make G chordal. Using Corollary 2.3, we can compute the normalizing constant corresponding to a minimal chordal cover of G. The question arises: Can one compute the normalizing constant of G from the normalizing constant of a minimal chordal cover of G? In the following theorem, we show how one can compute the normalizing constant of a graph G that results from removing one edge from a chordal graph. Such graphs are said to have *minimum fill-in equal to* 1.

THEOREM 2.5. *Let $G = (V, E)$ be an undirected graph with minimum fill-in* 1 *and with vertices $V = \{1, \dots, p\}$. Let $G^e = (V, E^e)$ denote the graph G with one additional edge e, that is, $E^e = E \cup \{e\}$, such that G^e is chordal. Let d denote the number of triangles formed by the edge e and two other edges in G^e. Then*

$$I_G(\delta, \mathbb{I}_p) = \pi^{-1/2} \frac{\Gamma(\delta + \frac{1}{2}(d+2))}{\Gamma(\delta + \frac{1}{2}(d+3))} I_{G^e}(\delta, \mathbb{I}_p).$$

PROOF. We begin by defining an ordering of the vertices in such a way that one can directly integrate out the variables corresponding to the end points of e and the variable corresponding to e itself.

Let one of the end points of e be labeled as "1," the other end point as '$d+2$' and label the d vertices involved in triangles over the edge e by $2, \dots, d+1$. Label all remaining vertices by $d+3, \dots, p$. Let $\bar{\mathcal{G}}^e$ denote the moral DAG to G^e with edge set $\bar{\mathcal{E}}^e$. Then the chosen ordering of the vertices guarantees that $\bar{\mathcal{E}}^e = \bar{\mathcal{E}} \cup \{e\}$, and $e \notin \bar{\mathcal{E}}$.

Also, since all vertices $2, \dots, d+1$ are connected to vertex $d+2$, no added edge in $\bar{\mathcal{E}} \setminus \mathcal{E}$ points to vertex $d+2$, and hence $a_{d+2,d+2}$ does not appear in any equation for the edges in $\bar{\mathcal{E}} \setminus \mathcal{E}$. Similar arguments hold for vertex 1, since due to the ordering there can be no edge pointing to node 1.

Let A and A^e denote the Cholesky factors of G and G^e, respectively. Then

$$A_{ij} = \begin{cases} A^e_{ij} & \text{for all } (i, j) \neq (1, d+2), \\ 0 & \text{if } (i, j) = (1, d+2). \end{cases}$$

Let indeg denote the indegree with respect to $\mathcal{G}$ and indeg^e the indegree with respect to the DAG $\mathcal{G}^e$. Let $A_* = ((a_{ii})_{i \notin \{1, d+2\}}, (a_{ij})_{(i,j) \in \mathcal{E}})$. Note that

$$\text{(2.4)} \quad \operatorname{indeg}^e(1) = 0 = \operatorname{indeg}(1), \qquad \operatorname{indeg}^e(d+2) = d+1 = \operatorname{indeg}(d+2) + 1.$$

Then by Theorem 2.2,

$$\begin{aligned} I_{G^e}(\delta,\mathbb{I}_p) &= \int\left(\prod_{i=1}^{p} a_{ii}^{\delta+\frac{1}{2}\operatorname{indeg}^e(i)}\exp(-a_{ii})\right) \\ &\qquad\cdot\exp\left[-\sum_{(i,j)\in\bar{\mathcal{E}}^e} a_{ij}^2\right]\mathrm{d}a_{11}\,\mathrm{d}a_{d+2,d+2}\,\mathrm{d}a_{1,d+2}\,\mathrm{d}A_* \\ &= \int_{-\infty}^{\infty}\exp(-a_{1,d+2}^2)\,\mathrm{d}a_{1,d+2}\cdot\int_0^{\infty} a_{11}^{\delta+\frac{1}{2}\operatorname{indeg}^e(1)}\exp(-a_{11})\,\mathrm{d}a_{11} \\ &\qquad\cdot\int_0^{\infty} a_{d+2,d+2}^{\delta+\frac{1}{2}\operatorname{indeg}^e(d+2)}\exp(-a_{d+2,d+2})\,\mathrm{d}a_{d+2,d+2} \\ &\qquad\cdot\int_{A_*}\left[\prod_{i\notin\{1,d+2\}} a_{ii}^{\delta+\frac{1}{2}\operatorname{indeg}^e(i)}\exp(-a_{ii})\right]\exp\left[-\sum_{(i,j)\in\bar{\mathcal{E}}} a_{ij}^2\right]\mathrm{d}A_*. \end{aligned}$$

The integral with respect to $a_{1,d+2}$ is a Gaussian integral, with value $\sqrt{\pi}$. Also, by (2.4),

$$\int_0^{\infty} a_{11}^{\delta+\frac{1}{2}\operatorname{indeg}^e(1)}\exp(-a_{11})\,\mathrm{d}a_{11} = \int_0^{\infty} a_{11}^{\delta+\frac{1}{2}\operatorname{indeg}(1)}\exp(-a_{11})\,\mathrm{d}a_{11}.$$

Again by (2.4), we have

$$\begin{aligned} &\int_0^{\infty} a_{d+2,d+2}^{\delta+\frac{1}{2}\operatorname{indeg}^e(d+2)}\exp(-a_{d+2,d+2})\,\mathrm{d}a_{d+2,d+2} \\ &\qquad = \frac{\Gamma(\delta+\frac{1}{2}(d+1)+1)}{\Gamma(\delta+\frac{1}{2}d+1)}\int_0^{\infty} a_{d+2,d+2}^{\delta+\frac{1}{2}\operatorname{indeg}(d+2)}\exp(-a_{d+2,d+2})\,\mathrm{d}a_{d+2,d+2}. \end{aligned}$$

Finally, since $\operatorname{indeg}^e(i)=\operatorname{indeg}(i)$ for all $i\notin\{1,d+2\}$, we obtain

$$\begin{aligned} I_{G^e}(\delta,\mathbb{I}_p) &= \sqrt{\pi}\,\frac{\Gamma(\delta+\frac{1}{2}(d+1)+1)}{\Gamma(\delta+\frac{1}{2}d+1)}\int_0^{\infty} a_{11}^{\delta+\operatorname{indeg}(1)/2}\exp(-a_{11})\,\mathrm{d}a_{11} \\ &\qquad\cdot\int_0^{\infty} a_{d+2,d+2}^{\delta+\frac{1}{2}\operatorname{indeg}(d+2)}\exp(-a_{d+2,d+2})\,\mathrm{d}a_{d+2,d+2} \\ &\qquad\cdot\int_{A_*}\left(\prod_{i\notin\{1,d+2\}} a_{ii}^{\delta+\frac{1}{2}\operatorname{indeg}(i)}\exp(-a_{ii})\right)\exp\left[-\sum_{(i,j)\in\bar{\mathcal{E}}} a_{ij}^2\right]\mathrm{d}A_* \\ &= \sqrt{\pi}\,\frac{\Gamma(\delta+\frac{1}{2}(d+3))}{\Gamma(\delta+\frac{1}{2}(d+2))} I_G(\delta,\mathbb{I}_p). \end{aligned}$$

The proof now is complete. □

EXAMPLE 2.6. Since the graph G_5 discussed in Example 2.4 has minimum fill-in equal to 1, we can apply Theorem 2.5 to compute its normalizing constant.

The skeleton of the graph shown in Figure 1(right) is a chordal cover of G_5 and the given vertex labeling is a perfect labeling. By applying Proposition 2.3, we deduce the normalizing constant for the graph G_5 with the additional edge $e = (1, 3)$:

$$I_{G_5^e}(\delta, \mathbb{I}_p) = \pi^4 \Gamma(\delta + 1)\Gamma\left(\delta + \frac{3}{2}\right)[\Gamma(\delta + 2)]^2 \Gamma\left(\delta + \frac{5}{2}\right).$$

Since the number of triangles over the red edge $(1, 3)$ is $d = 3$, we find by Theorem 2.5 that

$$\begin{aligned} I_{G_5}(\delta, \mathbb{I}_p) &= \pi^{-1/2} \frac{\Gamma(\delta + \frac{3}{2} + 1)}{\Gamma(\delta + \frac{4}{2} + 1)} I_{G_5^e}(\delta, \mathbb{I}_p) \\ &= \pi^{7/2} \frac{\Gamma(\delta + \frac{5}{2})}{\Gamma(\delta + 3)} \Gamma(\delta + 1)\Gamma\left(\delta + \frac{3}{2}\right)[\Gamma(\delta + 2)]^2 \Gamma\left(\delta + \frac{5}{2}\right). \end{aligned} \tag{2.5}$$

3. Computing $I_G(\delta, D)$ for general nonchordal graphs. In this section, we study $I_G(\delta, D)$ for general D. In Theorem 3.3, we show how the normalizing constant changes when removing not only an edge, but an entire clique (i.e., a completely connected subgraph) from a graph. This leads to an algorithm for computing the normalizing constant $I_G(\delta, D)$ for any graph G, which can then be specialized to the case in which $D = \mathbb{I}_p$.

3.1. *Some results on a generalized hypergeometric function of matrix argument.* We list in this subsection some results, involving a generalized hypergeometric function of matrix argument, that we will apply repeatedly in this section.

For $a \in \mathbb{C}$ and $k \in \{0, 1, 2, \ldots\}$, we denote the *rising factorial* by

$$(a)_k = \frac{\Gamma(a + k)}{\Gamma(a)} = a(a + 1)(a + 2) \cdots (a + k - 1).$$

For $t \in \mathbb{C}$ and $\rho \notin \{0, -1, -2, \ldots\}$, the *classical generalized hypergeometric function*, ${}_0F_1(\rho, t)$, may be defined by the series expansion

$${}_0F_1(\rho; t) = \sum_{l=0}^{\infty} \frac{t^l}{l!(\rho)_l}. \tag{3.1}$$

We refer to Andrews et al. [1] for many other properties of this function.

The *generalized hypergeometric function of matrix argument*, ${}_0F_1(\rho; Y)$, $Y \in \mathbb{S}^p_{\succ 0}$, is defined by the Laplace transform

$$\frac{1}{\Gamma_p(\rho)} \int_{\mathbb{S}^p_{\succ 0}} |Y|^{\rho - \frac{1}{2}(p+1)} \exp\bigl(-\operatorname{tr}(YD)\bigr)\, {}_0F_1(\rho; Y)\, \mathrm{d}Y = |D|^{-\rho} \exp\bigl(\operatorname{tr}\bigl(D^{-1}\bigr)\bigr),$$

valid for $\operatorname{Re}(\rho) > \frac{1}{2}(p - 1)$ and $D \in \mathbb{S}^p_{\succ 0}$. Herz [14] provided an extensive theory of the analytic properties of the function ${}_0F_1$. In particular, ${}_0F_1(\rho; Y)$ is simultaneously analytic in ρ for $\operatorname{Re}(\rho) > \frac{1}{2}(p - 1)$ and entire in Y; so, as a function of Y, its domain of definition extends to the set $\mathbb{S}^p$ and to the set of of complex symmetric matrices. Other properties of the function ${}_0F_1$, such as zonal polynomial expansions which generalize (3.1), are given by James [17], Muirhead [28] and Gross and Richards [12].

Herz ([14], page 497) proved that the function ${}_0F_1(\rho; Y)$ depends only on the eigenvalues of Y, and moreover that if $\mathrm{Re}(\rho) > \frac{1}{2}(p-1)$, $D \in \mathbb{S}^p_{\succ 0}$, and $C \in \mathbb{S}^p$, then there holds the Laplace transform formula

$$\begin{aligned}\int_{\mathbb{S}^p_{\succ 0}} |Y|^{\rho-\frac{1}{2}(p+1)} \exp\bigl(-\mathrm{tr}(YD)\bigr)\, {}_0F_1(\rho; YC)\, \mathrm{d}Y \\ = \Gamma_p(\rho)|D|^{-\rho} \exp\bigl(\mathrm{tr}\bigl(D^{-1}C\bigr)\bigr),\end{aligned} \tag{3.2}$$

where, by convention, ${}_0F_1(\rho; YC)$ is an abbreviation for ${}_0F_1(\rho; Y^{1/2}CY^{1/2})$ and $Y^{1/2} \in \mathbb{S}^p_{\succ 0}$ is the unique square root of Y. Setting $C = 0$ (the zero matrix) in (3.2), we deduce from the uniqueness of the Laplace transform and (1.2) that ${}_0F_1(\rho; 0) = 1$.

We will apply repeatedly a generalization of the Poisson integral to matrix spaces (see [14], pages 495–496, and [17], equation (151)): If A is a $k \times p$ matrix such that $k \le p$, and $\mathrm{Re}(\rho) > \frac{1}{2}(k+p-1)$, then

$$\begin{aligned}\int_{0<XX^T<\mathbb{I}_k} \bigl|\mathbb{I}_k - XX^T\bigr|^{\rho-\frac{1}{2}(k+p+1)} \exp\bigl(\mathrm{tr}\bigl(AX^T\bigr)\bigr)\, \mathrm{d}X \\ = \frac{\pi^{kp/2}\Gamma_k(\rho-\frac{1}{2}p)}{\Gamma_k(\rho)}\, {}_0F_1\biggl(\rho; \frac{1}{4}AA^T\biggr),\end{aligned} \tag{3.3}$$

where the region of integration is the set of all $k \times p$ matrices X such that $XX^T \in \mathbb{S}^k_{\succ 0}$ and $I - XX^T \in \mathbb{S}^k_{\succ 0}$. In particular, on setting $A = 0$ we obtain

$$\int_{0<XX^T<\mathbb{I}_k} \bigl|\mathbb{I}_k - XX^T\bigr|^{\rho-\frac{1}{2}(k+p+1)}\, \mathrm{d}X = \frac{\pi^{kp/2}\Gamma_k(\rho-\frac{1}{2}p)}{\Gamma_k(\rho)}, \tag{3.4}$$

a result which was used in Proposition 2.1.

For the case in which Y is a 2×2 matrix, Muirhead [27] proved that

$${}_0F_1(\rho; Y) = \sum_{q=0}^{\infty} \frac{1}{q!(\rho)_{2q}(\rho-\frac{1}{2})_q} |Y|^q\, {}_0F_1\bigl(\rho+2q; \mathrm{tr}(Y)\bigr), \tag{3.5}$$

where the ${}_0F_1$ functions on the right-hand side are the classical generalized hypergeometric functions given in (3.1). In the special case in which Y is of rank 1, it follows from Herz ([14], page 497) or directly from (3.5), that

$${}_0F_1(\rho; Y) = {}_0F_1\bigl(\rho; \mathrm{tr}(Y)\bigr). \tag{3.6}$$

3.2. *The normalizing constant for nonchordal graphs.* We want to calculate

$$I_G(\delta, D) = \int_{\mathbb{S}^p_{\succ 0}(G)} |K|^\delta \exp\bigl(-\operatorname{tr}(KD)\bigr)\,\mathrm{d}K,$$

the normalizing constant for G, a general nonchordal graph. By making the change of variables $K \to \operatorname{diag}(D)^{-1/2} K \operatorname{diag}(D)^{-1/2}$, we can assume, without loss of generality, that D has ones on the diagonal and, therefore, is a correlation matrix; this assumption will be maintained explicitly for the remainder of the paper.

In the sequel, we will encounter a $2 \times m$ matrix $C = (C_{ij})$, and then we use the notation $|C_{\{1,2\},\{i,j\}}|$ for the minor corresponding to rows 1 and 2 and to columns i and j, where $i, j \in \{1, \dots, m\}$. We will need $L = (L_{ij})$, a $2 \times m$ matrix of nonnegative integers such that $\sum_{i=1}^2 \sum_{j=1}^m L_{ij} = l$, and we adopt the notation

$$\binom{l}{L} = \frac{l!}{\prod_{i=1}^2 \prod_{j=1}^m L_{ij}!}, \qquad L_{i+} = \sum_{j=1}^m L_{ij}, \quad \text{and} \quad L_{+j} = \sum_{i=1}^2 L_{ij}.$$

We will also have $Q = (Q_{ij})_{1 \le i < j \le m}$, a matrix of nonnegative integers such that $\sum_{1 \le i < j \le m} Q_{ij} = q$, and we set

$$\binom{q}{Q} = \frac{q!}{\prod_{1 \le i < j \le m} Q_{ij}!}, \qquad Q_{i+} = \sum_{j=i+1}^m Q_{ij}, \quad \text{and} \quad Q_{+j} = \sum_{i=1}^{j-1} Q_{ij}.$$

In the following result, we obtain the normalizing constant for $H_{2,m}$, a complete bipartite graph on $2 + m$ vertices.

PROPOSITION 3.1. *The integral $I_{H_{2,m}}(\delta, D)$ converges absolutely for all $\delta > -1$ and $D \in \mathbb{S}^{2+m}_{\succ 0}$. Let $C = (C_{ij})$ denote the $2 \times m$ submatrix of D corresponding to the edges in G; then $I_{H_{2,m}}(\delta, D)$ equals*

$$\begin{aligned} &I_{H_{2,m}}(\delta, \mathbb{I}_{m+2}) \\ &\quad \cdot \sum_{q=0}^\infty \frac{(\delta + \frac{1}{2}(m+2))_q [(\delta+2)_q]^m}{q!(\delta + \frac{1}{2}(m+3))_{2q}} \sum_{l=0}^\infty \frac{1}{l!(\delta + 2q + \frac{1}{2}(m+3))_l} \\ &\quad \cdot \sum_L \binom{l}{L} \Biggl(\prod_{i=1}^2 \prod_{j=1}^m C_{ij}^{L_{ij}}\Biggr) \Biggl(\prod_{i=1}^2 \Bigl(\delta + q + \frac{1}{2}(m+2)\Bigr)_{L_{i+}}\Biggr) \Biggl(\prod_{j=1}^m (\delta+2)_{L_{+j}}\Biggr) \\ &\quad \cdot \sum_Q \binom{q}{Q} \Biggl(\prod_{1 \le i < jm} |C_{\{1,2\},\{i,j\}}|^{2Q_{ij}}\Biggr) \Biggl(\prod_{j=1}^m (\delta + L_{+j} + 2)_{Q_{j+} + Q_{+j}}\Biggr), \end{aligned}$$

with

$$I_{H_{2,m}}(\delta, \mathbb{I}_{m+2}) = \frac{\pi^m \Gamma_2(\delta + \frac{3}{2})}{\Gamma_2(\delta + \frac{1}{2}(m+3))} [\Gamma(\delta+2)]^m \Gamma\Bigl(\delta + \frac{1}{2}(m+2)\Bigr)^2. \tag{3.7}$$

PROOF. We order the vertices such that

$$K = \begin{pmatrix} K_{AA} & K_{AB} \\ K_{BA} & K_{BB} \end{pmatrix},$$

where $K_{AA} = \mathrm{diag}(\kappa_1, \kappa_2)$, $K_{BB} = \mathrm{diag}(k_1, \dots, k_m)$, and K_{AB} is unconstrained. We partition D in a similar way:

$$D = \begin{pmatrix} D_{AA} & D_{AB} \\ D_{BA} & D_{BB} \end{pmatrix},$$

where $\mathrm{diag}(D) = (1, \dots, 1)$ and $D_{AB} = C$. By applying the determinant formula for block matrices, (2.2), and making a change-of-variables to replace K_{AB} by $K_{AA}^{1/2} K_{AB} K_{BB}^{1/2}$, we obtain similarly as in the proof of Proposition 2.1:

$$\begin{aligned} I_{H_{2,m}}(\delta, D) &= \int_{\mathbb{S}^{2+m}_{\succ 0}(G)} |K|^{\delta} \exp\bigl(-\mathrm{tr}(KD)\bigr) \,\mathrm{d}K \\ &= \int_{\mathbb{S}^{2+m}_{\succ 0}(G)} |K_{AA}|^{\delta + \frac{1}{2}m} |K_{BB}|^{\delta+1} |\mathbb{I}_m - K_{AB}^T K_{AB}|^{\delta} \\ &\quad \cdot \exp\bigl(-\mathrm{tr}(K_{AA}) - \mathrm{tr}(K_{BB})\bigr) \\ &\quad \cdot \exp\bigl[-2\,\mathrm{tr}\bigl(K_{AA}^{1/2} K_{AB} K_{BB}^{1/2} C^T\bigr)\bigr] \,\mathrm{d}K_{AA}\,\mathrm{d}K_{AB}\,\mathrm{d}K_{BB}. \end{aligned}$$

Applying (3.3) to integrate over K_{AB} we obtain

$$\begin{aligned} I_{H_{2,m}}(\delta, D) &= \frac{\pi^m \Gamma_2(\delta + \frac{3}{2})}{\Gamma_2(\delta + \frac{1}{2}(m+3))} \int |K_{AA}|^{\delta + \frac{1}{2}m} |K_{BB}|^{\delta+1} \\ &\quad \cdot \exp\bigl(-\mathrm{tr}(K_{AA}) - \mathrm{tr}(K_{BB})\bigr) \\ &\quad \cdot {}_0F_1\left(\delta + \frac{1}{2}(m+3); K_{AA} C K_{BB} C^T\right) \mathrm{d}K_{AA}\,\mathrm{d}K_{BB}. \end{aligned}$$

Applying (3.5) to expand this ${}_0F_1$ function of matrix argument in terms of a classical ${}_0F_1$ function of $\mathrm{tr}(K_{AA} C K_{BB} C^T)$, and applying (3.1), we get

$$\begin{aligned} &{}_0F_1\left(\delta + \frac{1}{2}(m+3); K_{AA} C K_{BB} C^T\right) \\ &\quad = \sum_{q=0}^{\infty} \frac{1}{q!(\delta + \frac{1}{2}(m+3))_{2q} (\delta + \frac{1}{2}(m+2))_q} |K_{AA} C K_{BB} C^T|^q \\ &\quad\quad \cdot {}_0F_1\left(\delta + 2q + \frac{1}{2}(m+3); \mathrm{tr}\bigl(K_{AA} C K_{BB} C^T\bigr)\right) \\ &\quad = \sum_{q=0}^{\infty} \frac{1}{q!(\delta + \frac{1}{2}(m+3))_{2q} (\delta + \frac{1}{2}(m+2))_q} |K_{AA} C K_{BB} C^T|^q \\ &\quad\quad \cdot \sum_{l=0}^{\infty} \frac{1}{l!(\delta + 2q + \frac{1}{2}(m+3))_l} \bigl(\mathrm{tr}\bigl(K_{AA} C K_{BB} C^T\bigr)\bigr)^l. \end{aligned}$$

By the Binet–Cauchy formula (see Karlin [19], page 1),

$$\begin{aligned}|K_{AA}CK_{BB}C^T| &= |K_{AA}|\cdot|CK_{BB}C^T| \\ &= |K_{AA}|\sum_{1\le i<j\le m} k_ik_j|C_{\{1,2\},\{i,j\}}|^2.\end{aligned}$$

Hence, by the multinomial theorem,

$$\begin{aligned}&|K_{AA}CK_{BB}C^T|^q \\ &\quad= |K_{AA}|^q\sum_Q\binom{q}{Q}\prod_{1\le i<j\le m}\bigl(k_ik_j|C_{\{1,2\},\{i,j\}}|^2\bigr)^{Q_{ij}} \\ &\quad= |K_{AA}|^q\sum_Q\binom{q}{Q}\Biggl(\prod_{i=1}^m k_i^{Q_{i+}+Q_{+i}}\Biggr)\Biggl(\prod_{1\le i<j\le m}|C_{\{1,2\},\{i,j\}}|^{2Q_{ij}}\Biggr),\end{aligned}$$

where $Q=(Q_{ij})_{1\le i<j\le m}$ is a matrix of nonnegative integers, as defined earlier. Also,

$$\operatorname{tr}(K_{AA}CK_{BB}C^T)=\sum_{i=1}^2\sum_{j=1}^m\kappa_ik_jC_{ij},$$

and hence, by the multinomial theorem,

$$\begin{aligned}\bigl(\operatorname{tr}(K_{AA}CK_{BB}C^T)\bigr)^l &= \Biggl(\sum_{i=1}^2\sum_{j=1}^m\kappa_ik_jC_{ij}\Biggr)^l \\ &= \sum_L\binom{l}{L}\prod_{i=1}^2\prod_{j=1}^m(\kappa_ik_jC_{ij})^{L_{ij}} \\ &= \sum_L\binom{l}{L}\Biggl(\prod_{i=1}^2(\kappa_i)^{L_{i+}}\Biggr)\Biggl(\prod_{j=1}^m k_j^{L_{+j}}\Biggr)\Biggl(\prod_{i=1}^2\prod_{j=1}^m(C_{ij})^{L_{ij}}\Biggr),\end{aligned}$$

where $L=(L_{ij})$ is a $2\times m$ nonnegative integer matrix defined earlier. Hence,

$$\begin{aligned}I_{H_{2,m}}(\delta,D) &= \frac{\pi^m\Gamma_2(\delta+\frac{3}{2})}{\Gamma_2(\delta+\frac{1}{2}(m+3))}\sum_{q=0}^\infty\frac{1}{q!(\delta+\frac{1}{2}(m+3))_{2q}(\delta+\frac{1}{2}(m+2))_q} \\ &\quad\cdot\sum_{l=0}^\infty\frac{1}{l!(\delta+2q+\frac{1}{2}(m+3))_l} \\ &\quad\cdot\sum_L\binom{l}{L}\Biggl(\prod_{i=1}^2\prod_{j=1}^m {C_{ij}}^{L_{ij}}\Biggr)\Biggl(\prod_{i=1}^2\int_0^\infty\kappa_i^{\delta+q+L_{i+}+\frac{1}{2}m}e^{-\kappa_i}\,\mathrm{d}\kappa_i\Biggr)\end{aligned}$$

$$\cdot \sum_{Q} \binom{q}{Q} \Bigl(\prod_{1 \le i < j \le m} |C_{\{1,2\},\{i,j\}}|^{2Q_{ij}} \Bigr)$$

$$\cdot \left(\prod_{j=1}^{m} \int_0^{\infty} k_j^{\delta + Q_{j+} + Q_{+j} + L_{+j} + 1} e^{-k_j} \, \mathrm{d}k_j \right).$$

Evaluating each gamma integral and simplifying the outcomes, we obtain

$$I_{H_{2,m}}(\delta, D) = \frac{\pi^m \Gamma_2(\delta + \frac{3}{2})}{\Gamma_2(\delta + \frac{1}{2}(m+3))} \bigl[\Gamma(\delta+2)\bigr]^m \left[\Gamma\left(\delta + \frac{1}{2}(m+2)\right)\right]^2$$

$$\cdot \sum_{q=0}^{\infty} \frac{(\delta + \frac{1}{2}(m+2))_q ((\delta+2)_q)^m}{q!(\delta + \frac{1}{2}(m+3))_{2q}} \sum_{l=0}^{\infty} \frac{1}{l!(\delta + 2q + \frac{1}{2}(m+3))_l}$$

$$\cdot \sum_{L} \binom{l}{L} \left(\prod_{i=1}^{2} \prod_{j=1}^{m} (C_{ij})^{L_{ij}} \right) \left(\prod_{i=1}^{2} \left(\delta + q + \frac{1}{2}(m+2)\right)_{L_{i+}} \right)$$

$$\cdot \left(\prod_{j=1}^{m} (\delta+2)_{L_{+j}} \right)$$

$$\cdot \sum_{Q} \binom{q}{Q} \Bigl(\prod_{1 \le i < j \le m} |C_{\{1,2\},\{i,j\}}|^{2Q_{ij}} \Bigr)$$

$$\cdot \left(\prod_{j=1}^{m} (\delta + L_{+j} + 2)_{Q_{j+} + Q_{+j}} \right).$$

Finally, the value of $I_{H_{2,m}}(\delta, \mathbb{I}_{m+2})$ is obtained by applying Proposition 2.1 or Theorem 2.5, so the proof now is complete. □

Note that if we set $D = \mathbb{I}_{m+2}$ in the proof of Proposition 3.1 then $|C_{\{1,2\},\{i,j\}}| = C_{ij} = 0$. Hence, in the infinite series, the only nonzero terms are those for which $l = q = 0$, so the series reduces identically to 1.

The special structure of K was crucial for the proof of Proposition 3.1. We now combine Proposition 3.1 with the approach developed in Theorem 2.2, of representing K by its upper Cholesky decomposition, to describe how the normalizing constant changes when removing an edge from a chordal graph with maximal clique size at most 3. Similarly, as in the proof of Theorem 2.5, the main difficulty lies in defining a good ordering of the nodes. For simplifying notation, we denote the quotient of the normalizing constants for general D and the identity matrix by $\bar{I}_G(\delta, D)$, that is,

$$\bar{I}_G(\delta, D) = \frac{I_G(\delta, D)}{I_G(\delta, \mathbb{I}_p)}.$$

As an example, note that $\bar{I}_{H_{2,m}}(\delta, D)$ is given in Proposition 3.1.

COROLLARY 3.2. *Let $G = (V, E)$ be an undirected graph of minimum fill-in* 1 *with vertices $V = \{1, \dots, p\}$ and maximal clique size at most* 3. *Let $G^e = (V, E^e)$ denote the graph G with one additional edge e, that is, $E^e = E \cup \{e\}$, such that G^e is chordal and its maximal clique size is also at most* 3. *Let d denote the number of triangles formed by the edge e and two other edges in G^e. Then*

$$I_G(\delta, D) = \pi^{-1/2} \frac{\Gamma(\delta + \frac{1}{2}(d+2))}{\Gamma(\delta + \frac{1}{2}(d+3))} \frac{|D_{\{1,d+2\}}|^{d-1}}{\prod_{j=2}^{d+1} |D_{\{1,j,d+2\}}|} \bar{I}_{H_{2,d}}(\delta, D) I_{G^e}(\delta, D),$$

where $D_{\{i_1,\dots,i_k\}}$ denotes the principal submatrix of D corresponding to the rows and columns $i_1, \dots, i_k$.

PROOF. We define an ordering of the vertices in such a way that the integral for the normalizing constant $I_G(\delta, D)$ decomposes into an integral over a bipartite graph and an integral over the remaining variables. Similarly, as in the proof of Theorem 2.5, label one of the end points of e as "1," label the other end point as "$d + 2$," and label the d vertices involved in triangles over the edge e by $2, \dots, d + 1$. Label all remaining vertices by $d + 3, \dots, p$. Let $\bar{\mathcal{G}}$ denote the moral DAG to G with edge set $\bar{\mathcal{E}}$ and similarly for G^e.

By Theorem 2.2, the normalizing constant for G decomposes into an integral over the variables $A = \{a_{ij} \mid (i, j) \in \bar{\mathcal{E}}, i, j \le d + 2\}$ and an integral over the variables $B = \{a_{ij} \mid (i, j) \in \bar{\mathcal{E}}, a_{ij} \notin A\}$. The equivalent statement holds for the graph G^e with $A^e = A \cup \{e\}$ and $B^e = B$. Note that the integral over B is the same for G as for G^e. The integral over A is the normalizing constant for the complete bipartite graph $H_{2,d}$ with $U = \{1, d + 2\}$ and $V = \{2, \dots, d + 1\}$ where every vertex in U is connected to all vertices in V, but there are no edges within U nor within V. The integral over $A^e = A \cup \{e\}$ is the normalizing constant for the complete bipartite graph $H_{2,d}$ with one additional edge connecting the two nodes in U. We denote this graph by $H^e_{2,d}$. So

$$\begin{aligned} I_G(\delta, D) &= I_{G^e}(\delta, D) \frac{I_{H_{2,d}}(\delta, D)}{I_{H^e_{2,d}}(\delta, D)} \\ &= I_{G^e}(\delta, D) \frac{I_{H_{2,d}}(\delta, \mathbb{I}_{d+2}) \bar{I}_{H_{2,d}}(\delta, D)}{I_{H^e_{2,d}}(\delta, D)}, \end{aligned}$$

where $\bar{I}_{H_{2,d}}(\delta, D)$ is given by Proposition 3.1.

The additional edge e makes the graph $H^e_{2,m}$ chordal, and hence the normalizing constant is computed using (1.4):

$$I_{H^e_{2,d}}(\delta, D) = I_{H^e_{2,d}}(\delta, \mathbb{I}_{d+2}) \frac{\prod_{j=2}^{d+1} |D_{\{1,j,d+2\}}|}{|D_{\{1,d+2\}}|^{d-1}}.$$

By Theorem 2.5,

$$\frac{I_{H_{2,d}}(\delta, \mathbb{I}_{d+2})}{I_{H_{2,d}^e}(\delta, \mathbb{I}_{d+2})} = \pi^{-1/2} \frac{\Gamma(\delta + \frac{1}{2}(d+2))}{\Gamma(\delta + \frac{1}{2}(d+3))}.$$

By collecting all terms we find

$$\begin{aligned} I_G(\delta, D) &= I_{G^e}(\delta, D) \frac{I_{H_{2,d}}(\delta, \mathbb{I}_{d+2}) \bar{I}_{H_{2,d}}(\delta, D)}{I_{H_{2,d}^e}(\delta, \mathbb{I}_{d+2})} \frac{|D_{1,d+2}|^{d-1}}{\prod_{j=2}^{d+1} |D_{1,j,d+2}|} \\ &= \pi^{-1/2} \frac{\Gamma(\delta + \frac{1}{2}(d+2))}{\Gamma(\delta + \frac{1}{2}(d+3))} \frac{|D_{\{1,d+2\}}|^{d-1}}{\prod_{j=2}^{d+1} |D_{\{1,j,d+2\}}|} \bar{I}_{H_{2,d}}(\delta, D) I_{G^e}(\delta, D). \end{aligned}$$

The proof now is complete. □

Corollary 3.2 can be generalized to graphs of minimum fill-in 1 and arbitrary tree-width to obtain an extension of Theorem 2.5 to general D. This involves decomposing the normalizing constant for G into a normalizing constant for the chordal graph G^e and the quotient of the normalizing constants for the subgraph induced by the triangles over the edge e. This technical result is given in Theorem (S.3) in the supplementary material [34].

We now prove our main result which can be applied to compute the normalizing constant for any graph. It involves showing how the normalizing constant changes when removing a whole clique from a graph. However, for graphs of minimum fill-in 1 it is advisable for computational reasons to use the specialized result given in Theorem (S.3) in the supplementary material.

In the following, we denote by G_A the subgraph of G induced by the vertices $A \subset V$. In the following theorem, we will encounter a symmetric matrix $T_{AA} = (T_{ij})_{i,j \in A}$. Denoting Kronecker's delta by δ_{ij}, we define the matrix of differential operators,

$$\frac{\partial}{\partial T_{AA}} = \left(\frac{1}{2}(1 + \delta_{ij}) \frac{\partial}{\partial T_{ij}} \right)_{i,j \in A},$$

as in [9, 25]. The corresponding determinant, $\det(\partial/\partial T_{AA})$, and the (r, s)th cofactor, $\mathrm{Cof}_{rs}(\partial/\partial T_{AA})$, are defined in the usual way.

We will also make use of fractional powers of differential operators, a concept which is widely used in some areas of probability theory and mathematical analysis [3, 15], and which has also arisen in research on statistical inference for Wishart distributions [13, 20]. In the simplest formulation of such fractional powers, suppose that a function $f : \mathbb{R} \to \mathbb{R}$ is such that its nth derivative, $(\mathrm{d}/\mathrm{d}x)^n f(x)$, can be analytically continued as a function of n to a domain in $\mathbb{C}$; this allows us to define the αth derivative, $(\mathrm{d}/\mathrm{d}x)^\alpha f(x)$ where α belongs to the domain of analyticity.

Gårding [9] defined fractional powers, $(\det(\partial/\partial T_{AA}))^{\alpha}$, of the determinant $\det(\partial/\partial T_{AA})$ by means of analytic continuation in α. We will apply Gårding's fractional powers of operators to calculate the normalizing constant $I_G(\delta, D)$, and we provide in Example 3.5 an explicit calculation for a case in which the fractional power of the determinant $\det(\partial/\partial T_{AA})$ is $-1/2$.

The following theorem is the main result of the paper. In this result, we express $I_G(\delta, D)$ in terms of a series in which derivatives with respect to the U_{AA} are calculated, then the outcome is evaluated at $U_{AA} = T_{AA}$, then derivatives with respect to the T_{AA} are calculated, and then the resulting expression is evaluated at $T_{AA} = D_{AA}$.

THEOREM 3.3. *Let $G = (V, E)$ be an undirected graph and partition $V = A \cup B$ such that the induced subgraph G_B is a clique. Let $I = \{(i, j) \in A \times B \mid (i, j) \in E\}$ denote the edges connecting A and B, and let I_1 and I_2 denote the projection of I onto the first and second coordinate, respectively. Define*

$$\partial_{I_1,I_2}(D, T_{AA}) = \left(-D_{I_2(r),I_2(s)} \operatorname{Cof}_{I_1(r),I_1(s)}\left(\frac{\partial}{\partial T_{AA}}\right)\right)_{r,s=1}^{|I|}, \tag{3.8}$$

a $|I| \times |I|$ matrix of differential operators. Then

$$\begin{aligned} I_G(\delta, D) = \pi^{|I|/2}\Gamma_{|B|}\left(\delta + \frac{1}{2}(|B|+1)\right)|D_{BB}|^{-(\delta+\frac{1}{2}(|B|+1))} \\ \cdot \left(\det \partial_{I_1,I_2}(D, T_{AA})\right)^{-1/2} \\ \cdot \sum_{\substack{0\le j_{rs}<\infty \\ 1\le r\le s\le|I|}} \cdots \sum \left(\det \partial_{I_1,I_2}(D, U_{AA})\right)^{-j_{\cdot\cdot}} \\ \cdot \left(\prod_{1\le r\le s\le|I|} \frac{(1+\delta_{rs})^{j_{rs}}}{j_{rs}!} D^{j_{rs}}_{I_1(r),I_2(r)} D^{j_{rs}}_{I_1(s),I_2(s)} \right. \\ \left. \cdot \left(\operatorname{Cof}_{rs} \partial_{I_1,I_2}(D, U_{AA})\right)^{j_{rs}}\right) \\ \cdot I_{G_A}\left(\delta + \frac{1}{2}|I| + j_{\cdot\cdot}, U_{AA}\right)\Big|_{U_{AA}=T_{AA}}\Big|_{T_{AA}=D_{AA}}. \end{aligned}$$

As a corollary of this theorem, we obtain an analogous formula for the case in which $D = \mathbb{I}_p$.

COROLLARY 3.4. *Let $G = (V, E)$ be an undirected graph with vertices $V = \{1, \ldots, p\}$. Let V be partitioned such that $V = A \cup B$ and the induced subgraph G_B is a clique. Let $I = \{(i, j) \in A \times B \mid (i, j) \in E\}$ denote the edges connecting A, B and let I_1 and I_2 denote the projection of I onto the first and second coordinate,* *respectively. Then*

$$I_G(\delta, \mathbb{I}_p) = \pi^{|I|/2}\Gamma_{|B|}\Big(\delta + \frac{1}{2}(|B|+1)\Big) \\ \cdot \partial_{I_1,I_2}(D, T_{AA}) I_{G_A}\big(\delta + |I|/2, T_{AA}\big)\big|_{T_{AA}=\mathbb{I}_{|A|}}.$$

Theorem 3.3 and Corollary 3.4 enable calculation of the normalizing constant of the G-Wishart distribution for any graph by removing cliques sequentially until the resulting graph is chordal, in which case the normalizing constant is known. In the following example, we show how to apply Theorem 3.3 in order to compute the normalizing constant for general D for the graph G_5 given in Figure 1.

EXAMPLE 3.5. We wish to calculate

$$I_{G_5}(\delta, D) = \int_{K\in\mathbb{S}^5_{\succ 0}(G_5)} |K|^{\delta} \exp\big(-\operatorname{tr}(KD)\big)\,\mathrm{d}K.$$

We partition the matrix K into blocks,

$$K = \begin{pmatrix} K_{AA} & K_{AB} \\ K_{AB}^T & K_{BB} \end{pmatrix}, \qquad \text{where}$$

$$K_{AA} = \begin{pmatrix} k_{11} & k_{12} \\ k_{12} & k_{22} \end{pmatrix}, \qquad K_{AB} = \begin{pmatrix} 0 & k_{14} & k_{15} \\ k_{23} & 0 & 0 \end{pmatrix},$$

$$K_{BB} = \begin{pmatrix} k_{33} & k_{34} & k_{35} \\ k_{34} & k_{44} & k_{45} \\ k_{35} & k_{45} & k_{55} \end{pmatrix}.$$

Noting that K_{BB} is unconstrained, we now apply Theorem 3.3. In the following, we provide all the ingredients of the calculation, *namely*,

$$I_1 = (2, 1, 1), \qquad I_2 = (3, 4, 5), \qquad \operatorname{vec}\big(D_{AB}^I\big) = \begin{pmatrix} d_{23} \\ d_{14} \\ d_{15} \end{pmatrix}.$$

Further, the matrix of differential operators is

$$\partial_{I_1,I_2}(D, T_{AA}) = \Big(-D_{I_2(r),I_2(s)} \operatorname{Cof}_{I_1(r),I_1(s)}\Big(\frac{\partial}{\partial T_{AA}}\Big)\Big)_{r,s=1}^3$$

$$= \begin{pmatrix} -d_{33}\dfrac{\partial}{\partial T_{11}} & \dfrac{1}{2}d_{34}\dfrac{\partial}{\partial T_{12}} & \dfrac{1}{2}d_{35}\dfrac{\partial}{\partial T_{12}} \\ \dfrac{1}{2}d_{34}\dfrac{\partial}{\partial T_{12}} & -d_{44}\dfrac{\partial}{\partial T_{22}} & -\dfrac{1}{2}d_{45}\dfrac{\partial}{\partial T_{22}} \\ \dfrac{1}{2}d_{35}\dfrac{\partial}{\partial T_{12}} & -\dfrac{1}{2}d_{45}\dfrac{\partial}{\partial T_{22}} & -d_{55}\dfrac{\partial}{\partial T_{22}} \end{pmatrix}$$

and similarly for $\partial_{I_1,I_2}(D, U_{AA})$.

Since K_{AA} is unconstrained, the integral $I_{G_A}(\delta, U_{AA})$ is a standard Wishart normalizing constant, so we have

$$I_{G_A}(\delta, U_{AA}) = \Gamma_2\left(\delta + \frac{3}{2}\right)|U_{AA}|^{-(\delta+\frac{3}{2})}.$$

Then from Theorem 3.3, we obtain

$$\begin{aligned}
I_{G_5}(\delta, D) = \pi^{3/2}&\Gamma_3(\delta+2)|D_{BB}|^{-(\delta+2)}\bigl(\det\partial_{I_1,I_2}(D, T_{AA})\bigr)^{-1/2} \\
&\cdot \sum_{\substack{0\le j_{rs}<\infty \\ 1\le r\le s\le 3}}\cdots\sum \Gamma_2(\delta+3+j_{..})\bigl(\det\partial_{I_1,I_2}(D, U_{AA})\bigr)^{-j_{..}} \\
&\cdot \left(\prod_{1\le r\le s\le 3}\frac{(1+\delta_{rs})^{j_{rs}}}{j_{rs}!}D_{I_1(r),I_2(r)}^{j_{rs}}D_{I_1(s),I_2(s)}^{j_{rs}}\right. \\
&\left.\cdot \bigl(\mathrm{Cof}_{rs}\,\partial_{I_1,I_2}(D, U_{AA})\bigr)^{j_{rs}}\right) \\
&\cdot |U_{AA}|^{-(\delta+3+j_{..})}\Big|_{U_{AA}=T_{AA}}\Big|_{T_{AA}=D_{AA}}.
\end{aligned} \tag{3.9}$$

For the case in which $D = \mathbb{I}_5$, we have $D_{I_1(r),I_2(r)} = 0$ for all $r = 1, 2, 3$ and hence we deduce the result given in Corollary 3.4, namely,

$$\begin{aligned}
I_{G_5}(\delta, \mathbb{I}_5) = \pi^{3/2}&\Gamma_3(\delta+2)\Gamma_2(\delta+3) \\
&\cdot \bigl(\det\partial_{I_1,I_2}(D, T_{AA})\bigr)^{-1/2}|T_{AA}|^{-(\delta+3)}\big|_{T_{AA}=\mathbb{I}_{|A|}}.
\end{aligned}$$

By (3.8),

$$\begin{aligned}
&\bigl(\det\partial_{I_1,I_2}(D, T_{AA})\bigr)^{n}|T_{AA}|^{-(\delta+3)}\big|_{T_{AA}=\mathbb{I}_{|A|}} \\
&\quad = (-1)^n\left(\frac{\partial}{\partial T_{11}}\right)^n\left(\frac{\partial}{\partial T_{22}}\right)^{2n}(T_{11}T_{22})^{-(\delta+3)}\Big|_{T_{11}=T_{22}=1} \\
&\quad = (\delta+3)(\delta+4)\cdots(\delta+2+n)(\delta+3)(\delta+4)\cdots(\delta+2+2n) \\
&\quad = \frac{\Gamma(\delta+3+n)}{\Gamma(\delta+3)}\frac{\Gamma(\delta+3+2n)}{\Gamma(\delta+3)}.
\end{aligned}$$

The latter expression, considered as a function of a complex variable n, is analytic in the complex plane on a region containing the point $n = -\frac{1}{2}$. Therefore, in accordance with Gårding's fractional calculus,

$$\begin{aligned}
\bigl(\det\partial_{I_1,I_2}(D, T_{AA})\bigr)^{-1/2}|T_{AA}|^{-(\delta+3)}\big|_{T_{AA}=\mathbb{I}_{|A|}} &= \frac{\Gamma(\delta+3+n)}{\Gamma(\delta+3)}\frac{\Gamma(\delta+3+2n)}{\Gamma(\delta+3)}\Big|_{n=-\frac{1}{2}} \\
&= \frac{\Gamma(\delta+\frac{5}{2})}{\Gamma(\delta+3)}\frac{\Gamma(\delta+2)}{\Gamma(\delta+3)},
\end{aligned}$$

so we obtain the same result for $I_{G_5}(\delta, \mathbb{I}_5)$ as in (2.5).

To complete this section, we now provide the proofs of Theorem 3.3 and Corollary 3.4.

PROOF OF THEOREM 3.3. The matrix K is of the form

$$K = \begin{pmatrix} K_{AA} & K_{AB} \\ K_{AB}^T & K_{BB} \end{pmatrix} \in \mathbb{S}^p_{\succ 0},$$

where K_{BB} has no zero constraints. By applying the determinant formula for block matrices, (2.2), making a change-of-variables, $K_{BB} \to K_{BB} + K_{AB}^T (K_{AA})^{-1} K_{AB}$, and applying (1.2) to compute the integral over K_{BB}, we obtain

$$\begin{aligned} I_G(\delta, D) = {} & \Gamma_{|B|}\biggl(\delta + \frac{1}{2}(|B|+1)\biggr) |D_{BB}|^{-(\delta + \frac{1}{2}(|B|+1))} \\ & \cdot \int |K_{AA}|^{\delta} \exp\bigl(-\operatorname{tr}(K_{AA} D_{AA})\bigr) \\ & \cdot \int \exp\bigl(-2 \operatorname{tr}(K_{AB} D_{AB})\bigr) \\ & \cdot \exp\bigl(-\operatorname{tr}\bigl(D_{BB} K_{AB}^T (K_{AA})^{-1} K_{AB}\bigr)\bigr) \, \mathrm{d}K_{AB} \, \mathrm{d}K_{AA}. \end{aligned}$$

Denote by $\operatorname{vec}(K_{AB})$ the vectorized matrix K_{AB}, written column-by-column. We apply a formula for the Kronecker product of matrices (see Muirhead [28], page 76) to obtain

$$\operatorname{tr}\bigl(D_{BB} K_{AB}^T (K_{AA})^{-1} K_{AB}\bigr) = \bigl(\operatorname{vec}(K_{AB})\bigr)^T \bigl(D_{BB} \otimes (K_{AA})^{-1}\bigr) \operatorname{vec}(K_{AB}).$$

Let $I = \{(i, j) \in A \times B \mid (K_{AB})_{ij} \neq 0\}$ and let I_1 denote the projection of I onto the first index and I_2 the projection of I onto the second index. Let $\operatorname{vec}(K_{AB}^I)$ denote the column vector containing the nonzero entries of $\operatorname{vec}(K_{AB})$ and let Λ^{-1} be a matrix containing the entries of $D_{BB} \otimes (K_{AA})^{-1}$ corresponding to the components of $\operatorname{vec}(K_{AB}^I)$, that is,

$$\begin{aligned} (\Lambda^{-1})_{rs} &= D_{I_2(r), I_2(s)} \bigl(K_{AA}^{-1}\bigr)_{I_1(r), I_1(s)} \\ &= D_{I_2(r), I_2(s)} \frac{1}{|K_{AA}|} \operatorname{Cof}_{I_1(r), I_1(s)}(K_{AA}), \end{aligned} \tag{3.10}$$

where $\operatorname{Cof}_{ij}(K_{AA})$ denotes the (i, j)th entry of the cofactor matrix of K_{AA}. Then

$$\begin{aligned} \operatorname{tr}(K_{AB} D_{AB}) &= \operatorname{vec}\bigl(K_{AB}^I\bigr)^T \operatorname{vec}\bigl(D_{AB}^I\bigr), \\ \operatorname{tr}\bigl(D_{BB} K_{AB}^T (K_{AA})^{-1} K_{AB}\bigr) &= \operatorname{vec}\bigl(K_{AB}^I\bigr)^T \Lambda^{-1} \operatorname{vec}\bigl(K_{AB}^I\bigr), \end{aligned}$$

and hence we obtain the integral over K_{AB} in the form of a Gaussian integral

$$\begin{aligned}
&\int \exp(-2\operatorname{tr}(K_{AB}D_{AB}))\exp(-\operatorname{tr}(D_{BB}K_{AB}^T(K_{AA})^{-1}K_{AB}))\,\mathrm{d}K_{AB}\\
&\quad= \int \exp(-2\operatorname{vec}(K_{AB}^I)^T\operatorname{vec}(D_{AB}^I))\\
&\qquad\cdot\exp(-\operatorname{vec}(K_{AB}^I)^T\Lambda^{-1}\operatorname{vec}(K_{AB}^I))\,\mathrm{d}K_{AB}^I\\
&\quad= \pi^{|I|/2}|\Lambda|^{1/2}\exp(\operatorname{vec}(D_{AB}^I)^T\Lambda\operatorname{vec}(D_{AB}^I)).
\end{aligned}$$

Therefore,

$$\begin{aligned}
I_G(\delta, D) ={}&\pi^{|I|/2}\Gamma_{|B|}\Big(\delta+\frac{1}{2}(|B|+1)\Big)|D_{BB}|^{-(\delta+\frac{1}{2}(|B|+1))}\\
&\cdot\int|K_{AA}|^{\delta+|I|/2}\exp(-\operatorname{tr}(K_{AA}D_{AA}))\\
&\cdot\det([D_{I_2(r),I_2(s)}\operatorname{Cof}_{I_1(r),I_1(s)}(K_{AA})]_{r,s=1}^{|I|})^{-1/2}\\
&\cdot\exp(\operatorname{vec}(D_{AB}^I)^T\Lambda\operatorname{vec}(D_{AB}^I))\,\mathrm{d}K_{AA}.
\end{aligned}$$

Now note that

$$\begin{aligned}
&\det([D_{I_2(r),I_2(s)}\operatorname{Cof}_{I_1(r),I_1(s)}(K_{AA})]_{r,s=1}^{|I|})\exp(-\operatorname{tr}(K_{AA}T_{AA}))\\
&\quad=\det(\partial_{I_1,I_2}(D,T_{AA}))\exp(-\operatorname{tr}(K_{AA}T_{AA})).
\end{aligned}\tag{3.11}$$

By analytic continuation [9], we obtain

$$\begin{aligned}
I_G(\delta, D) = {}&\pi^{|I|/2}\Gamma_{|B|}\Big(\delta+\frac{1}{2}(|B|+1)\Big)|D_{BB}|^{-(\delta+\frac{1}{2}(|B|+1))}\\
&\cdot\det(\partial_{I_1,I_2}(D,T_{AA}))^{-1/2}\\
&\cdot\int|K_{AA}|^{\delta+|I|/2}\exp(-\operatorname{tr}(K_{AA}T_{AA}))\\
&\cdot\exp(\operatorname{vec}(D_{AB}^I)^T\Lambda\operatorname{vec}(D_{AB}^I))\,\mathrm{d}K_{AA}\Big|_{T_{AA}=D_{AA}}.
\end{aligned}\tag{3.12}$$

Now we write the exponential function as an infinite series and apply the cofactor formula to express Λ in terms of the entries of Λ^{-1}:

$$\begin{aligned}
&\exp(\operatorname{vec}(D_{AB}^I)^T\Lambda\operatorname{vec}(D_{AB}^I))\\
&\quad=\sum_{0\le j_{rs}<\infty}\cdots\sum\prod_{1\le r\le s\le|I|}\frac{(1+\delta_{rs})^{j_{rs}}}{j_{rs}!}D_{I_1(r),I_2(r)}^{j_{rs}}D_{I_1(s),I_2(s)}^{j_{rs}}\Lambda_{rs}^{j_{rs}}\\
&\quad=\sum_{0\le j_{rs}<\infty}\cdots\sum\prod_{1\le r\le s\le|I|}\frac{(1+\delta_{rs})^{j_{rs}}}{j_{rs}!}D_{I_1(r),I_2(r)}^{j_{rs}}D_{I_1(s),I_2(s)}^{j_{rs}}
\end{aligned}$$

$$\cdot |K_{AA}|^{j_{rs}} \operatorname{Cof}_{rs}\bigl(\bigl[D_{I_2(a),I_2(b)} \operatorname{Cof}_{I_1(a),I_1(b)}(K_{AA})\bigr]_{a,b=1}^{|I|}\bigr)^{j_{rs}}$$
$$\cdot \det\bigl(\bigl[D_{I_2(a),I_2(b)} \operatorname{Cof}_{I_1(a),I_1(b)}(K_{AA})\bigr]_{a,b=1}^{|I|}\bigr)^{-j_{rs}}.$$

Inserting the latter series expansion for $\exp(\operatorname{vec}(D_{AB}^I)^T \Lambda \operatorname{vec}(D_{AB}^I))$ in the integral for the normalizing constant, interchanging the integral and the summation, and denoting $\sum\sum_{0\le r<s\le|I|} j_{rs}$ by $j_{\cdot\cdot}$, we obtain

$$\begin{aligned} I_G(\delta, D) &= \pi^{|I|/2}\Gamma_{|B|}\biggl(\delta + \frac{1}{2}(|B|+1)\biggr)|D_{BB}|^{-(\delta+\frac{1}{2}(|B|+1))} \\ &\quad\cdot \det\biggl(\biggl[-D_{I_2(r),I_2(s)} \operatorname{Cof}_{I_1(r),I_1(s)}\biggl(\frac{\partial}{\partial T_{AA}}\biggr)\biggr]_{r,s=1}^{|I|}\biggr)^{-1/2} \\ &\quad\cdot \sum_{0\le j_{rs}<\infty}\cdots\sum \int \det\bigl(\bigl[D_{I_2(a),I_2(b)} \operatorname{Cof}_{I_1(a),I_1(b)}(K_{AA})\bigr]_{a,b=1}^{|I|}\bigr)^{-j_{\cdot\cdot}} \\ &\quad\cdot \biggl(\prod_{1\le r\le s\le |I|} \frac{(1+\delta_{rs})^{j_{rs}}}{j_{rs}!} D_{I_1(r),I_2(r)}^{j_{rs}} D_{I_1(s),I_2(s)}^{j_{rs}} \\ &\quad\cdot \operatorname{Cof}_{rs}\bigl(\bigl[D_{I_2(a),I_2(b)} \operatorname{Cof}_{I_1(a),I_1(b)}(K_{AA})\bigr]_{a,b=1}^{|I|}\bigr)^{j_{rs}}\biggr) \\ &\quad\cdot |K_{AA}|^{\delta+|I|/2+j_{\cdot\cdot}} \exp\bigl(-\operatorname{tr}(K_{AA}T_{AA})\bigr)\Big|_{T_{AA}=D_{AA}}. \end{aligned}$$

To complete the proof, we introduce the differentials

$$\frac{\partial}{\partial U_{AA}} = \biggl(\frac{1}{2}(1+\delta_{ij})\frac{\partial}{\partial U_{ij}}\biggr)_{i,j\in A},$$

and define the operator $\partial_{I_1,I_2}(D, U_{AA})$ analogously to (3.8). Similar to (3.11),

$$\begin{aligned} &\operatorname{Cof}_{rs}\bigl(\partial_{I_1,I_2}(D,U_{AA})\bigr)\det\bigl(\partial_{I_1,I_2}(D,U_{AA})\bigr)\exp\bigl(-\operatorname{tr}(K_{AA}U_{AA})\bigr) \\ &\quad = \operatorname{Cof}_{rs}\bigl(\bigl[D_{I_2(a),I_2(b)} \operatorname{Cof}_{I_1(a),I_1(b)}(K_{AA})\bigr]_{a,b=1}^{|I|}\bigr) \\ &\qquad \cdot \det\bigl(\bigl[D_{I_2(a),I_2(b)} \operatorname{Cof}_{I_1(a),I_1(b)}(K_{AA})\bigr]_{a,b=1}^{|I|}\bigr)\exp\bigl(-\operatorname{tr}(K_{AA}U_{AA})\bigr). \end{aligned}$$

Inserting this result in the latter expression for $I_G(\delta, D)$, and applying the necessary evaluations, we obtain

$$\begin{aligned} I_G(\delta, D) &= \pi^{|I|/2}\Gamma_{|B|}\biggl(\delta + \frac{1}{2}(|B|+1)\biggr)|D_{BB}|^{-(\delta+\frac{1}{2}(|B|+1))} \\ &\quad\cdot \det\biggl(\biggl[-D_{I_2(r),I_2(s)} \operatorname{Cof}_{I_1(r),I_1(s)}\biggl(\frac{\partial}{\partial T_{AA}}\biggr)\biggr]_{r,s=1}^{|I|}\biggr)^{-1/2} \end{aligned}$$

$$
\begin{aligned}
&\cdot \sum_{0\le j_{rs}<\infty}\cdots\sum \det\biggl(\biggl[-D_{I_2(a),I_2(b)}\operatorname{Cof}_{I_1(a),I_1(b)}\biggl(\frac{\partial}{\partial U_{AA}}\biggr)\biggr]_{a,b=1}^{|I|}\biggr)^{-j_{\cdot\cdot}}\\
&\cdot \biggl(\prod_{1\le r\le s\le |I|}\frac{(1+\delta_{rs})^{j_{rs}}}{j_{rs}!}D_{I_1(r),I_2(r)}^{j_{rs}}D_{I_1(s),I_2(s)}^{j_{rs}}\\
&\cdot \operatorname{Cof}_{rs}\biggl(\biggl[-D_{I_2(a),I_2(b)}\operatorname{Cof}_{I_1(a),I_1(b)}\biggl(\frac{\partial}{\partial U_{AA}}\biggr)\biggr]_{a,b=1}^{|I|}\biggr)^{j_{rs}}\biggr)\\
&\cdot I_{G_A}(\delta+|I|/2+j_{\cdot\cdot},U_{AA})\Big|_{U_{AA}=T_{AA}}\Big|_{T_{AA}=D_{AA}},
\end{aligned}
$$

where, in the last line, we used the fact that

$$
I_{G_A}(\delta,U_{AA})=\int |K_{AA}|^{\delta}\exp(-\operatorname{tr}(K_{AA}U_{AA}))\,\mathrm{d}K_{AA}.
$$

This completes the proof. □

PROOF OF COROLLARY 3.4. This follows from Theorem 3.3 by setting $D=\mathbb{I}_p$ in (3.12). □

4. Discussion. In this paper, we provided an explicit representation of the G-Wishart normalizing constant for general graphs. Theorem 3.3 is our main result and it can be applied to compute the normalizing constant of any graph. However, for particular classes of graphs one might be able to obtain simpler formulas using a more specialized approach as can be seen by comparing the two formulas (3.9) and (4.1) for G_5. In Proposition 3.1, we provided a simpler formula for bipartite graphs $H_{2,m}$, and in Corollary 3.2 and in Theorem (S.3), which can be found in the supplementary material, for graphs with minimum fill-in 1. Note that Corollary 3.2 and Theorem (S.3) can be applied to graphs of minimum fill-in 1 and also to graphs which are clique sums of graphs of minimum fill-in 1.

Even in modest dimensions the size of the graph space necessitates iterative methods to address model uncertainty, as exhaustive enumeration is infeasible. Since the graphical model may be just one part of a larger hierarchy, Markov chain Monte Carlo methods are naturally used to perform posterior inference. In such scenarios, the chain moves between graphs in each scan of the parameter set and the transition probability reduces to the evaluation of ratios of G-Wishart normalizing constants. Since direct evaluation of these constants has appeared infeasible, previous work used computationally intensive sampling-based methods to approximate this ratio or sampled auxiliary parameters to avoid evaluating the normalizing constant altogether [4, 22, 36].

Our paper shows that computing the exact normalizing constant of the G-Wishart distribution is possible in principle. The various examples in this paper also make it clear that one can hope to find more computationally efficient procedures than Theorem 3.3 for computing the normalizing constant for particular classes of graphs. Important future work is the development of specialized methods for computing the normalizing constants of different classes of graphs that are important for applications, one example being grids, which are widely used in spatial applications.

SUPPLEMENTARY MATERIAL

Supplement to "Exact formulas for the normalizing constants of Wishart distributions for graphical models" (DOI: 10.1214/17-AOS1543SUPP; .pdf). Exact formulas for the normalizing constants of Wishart distributions for graphical models with minimum fill-in 1.

REFERENCES

[1] ANDREWS, G. E., ASKEY, R. and ROY, R. (1999). *Special Functions*. *Encyclopedia of Mathematics and Its Applications* **71**. Cambridge Univ. Press, Cambridge. MR1688958

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

[3] BOJDECKI, T. and GOROSTIZA, L. G. (1999). Fractional Brownian motion via fractional Laplacian. *Statist*. *Probab*. *Lett*. **44** 107–108. MR1706362

[4] CHENG, Y. and LENKOSKI, A. (2012). Hierarchical Gaussian graphical models: Beyond reversible jump. *Electron*. *J*. *Stat*. **6** 2309–2331. MR3020264

[5] DAWID, A. P. and LAURITZEN, S. L. (1993). Hyper-Markov laws in the statistical analysis of decomposable graphical models. *Ann*. *Statist*. **21** 1272–1317. MR1241267

[6] DEMPSTER, A. P. (1972). Covariance selection. *Biometrics* **28** 157–175.

[7] DOBRA, A. and LENKOSKI, A. (2011). Copula Gaussian graphical models and their application to modeling functional disability data. *Ann*. *Appl*. *Stat*. **5** 969–993. MR2840183

[8] DOBRA, A., LENKOSKI, A. and RODRIGUEZ, A. (2011). Bayesian inference for general Gaussian graphical models with application to multivariate lattice data. *J*. *Amer*. *Statist*. *Assoc*. **106** 1418–1433. MR2896846

[9] GÅRDING, L. (1947). The solution of Cauchy's problem for two totally hyperbolic linear differential equations by means of Riesz integrals. *Ann*. *of Math*. (2) **48** 785–826. MR0022648

[10] GIRI, N. C. (2004). *Multivariate Statistical Analysis*. Dekker, New York. MRMR0468025

[11] GIUDICI, P. and GREEN, P. J. (1999). Decomposable graphical Gaussian model determination. *Biometrika* **86** 785–801. MR1741977

[12] GROSS, K. I. and RICHARDS, D. ST. P. (1987). Special functions of matrix argument. I. Algebraic induction, zonal polynomials, and hypergeometric functions. *Trans*. *Amer. Math*. *Soc*. **301** 781–811. MR0882715

[13] HAFF, L. R., KIM, P. T., KOO, J.-Y. and RICHARDS, D. ST. P. (2011). Minimax estimation for mixtures of Wishart distributions. *Ann*. *Statist*. **39** 3417–3440. MR3012414

[14] HERZ, C. S. (1955). Bessel functions of matrix argument. *Ann*. *of Math*. (2) **61** 474–523. MR0069960

[15] HILLE, E. and PHILLIPS, R. S. (1957). *Functional Analysis and Semi-Groups*, rev. ed. *American Mathematical Society Colloquium Publications* **31**. Amer. Math. Soc., Providence, RI. MR0089373

[16] INGHAM, A. E. (1933). An integral which occurs in statistics. *Math. Proc. Cambridge Philos. Soc.* **29** 271–276.

[17] JAMES, A. T. (1964). Distributions of matrix variates and latent roots derived from normal samples. *Ann. Math. Stat.* **35** 475–501. MR0181057

[18] JONES, B., CARVALHO, C., DOBRA, A., HANS, C., CARTER, C. and WEST, M. (2005). Experiments in stochastic computation for high-dimensional graphical models. *Statist. Sci.* **20** 388–400. MR2210226

[19] KARLIN, S. (1968). *Total Positivity, Vol. I*. Stanford Univ. Press, Stanford, CA. MR0230102

[20] KIM, P. T. and RICHARDS, D. ST. P. (2011). Deconvolution density estimation on the space of positive definite symmetric matrices. In *Nonparametric Statistics and Mixture Models* 147–168. World Sci. Publ., Hackensack, NJ. MR2838725

[21] LAURITZEN, S. L. (1996). *Graphical Models*. *Oxford Statistical Science Series* **17**. Clarendon Press, Oxford. MR1419991

[22] LENKOSKI, A. (2013). A direct sampler for G-Wishart variates. *Stat* **2** 119–128.

[23] LENKOSKI, A. and DOBRA, A. (2011). Computational aspects related to inference in Gaussian graphical models with the G-Wishart prior. *J. Comput. Graph. Statist.* **20** 140–157. Supplementary material available online. MR2816542

[24] LETAC, G. and MASSAM, H. (2007). Wishart distributions for decomposable graphs. *Ann. Statist.* **35** 1278–1323. MR2341706

[25] MAASS, H. (1971). *Siegel's Modular Forms and Dirichlet Series*. *Lecture Notes in Mathematics* **216**. Springer, Berlin. Dedicated to the last great representative of a passing epoch Carl Ludwig Siegel on the occasion of his seventy-fifth birthday. MR0344198

[26] MITSAKAKIS, N., MASSAM, H. and ESCOBAR, M. D. (2011). A Metropolis–Hastings based method for sampling from the G-Wishart distribution in Gaussian graphical models. *Electron. J. Stat.* **5** 18–30. MR2763796

[27] MUIRHEAD, R. J. (1975). Expressions for some hypergeometric functions of matrix argument with applications. *J. Multivariate Anal.* **5** 283–293. MR0381137

[28] MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York. MR0652932

[29] OLKIN, I. (2002). The 70th anniversary of the distribution of random matrices: A survey. *Linear Algebra Appl.* **354** 231–243. Ninth special issue on linear algebra and statistics. MR1927659

[30] PICCIONI, M. (2000). Independence structure of natural conjugate densities to exponential families and the Gibbs' sampler. *Scand. J. Stat.* **27** 111–127. MR1774047

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

[32] SIEGEL, C. L. (1935). Über die analytische Theorie der quadratischen Formen. *Ann. of Math.* (2) **36** 527–606. MR1503238

[33] SPEED, T. P. and KIIVERI, H. T. (1986). Gaussian Markov distributions over finite graphs. *Ann. Statist.* **14** 138–150. MR0829559

[34] UHLER, C., LENKOSKI, A. and RICHARDS, D. (2017). Supplement to "Exact formulas for the normalizing constants of Wishart distributions for graphical models." DOI:10.1214/17-AOS1543SUPP.

[35] WANG, H. and CARVALHO, C. M. (2010). Simulation of hyper-inverse Wishart distributions for non-decomposable graphs. *Electron. J. Stat.* **4** 1470–1475. MR2741209

[36] WANG, H. and LI, S. Z. (2012). Efficient Gaussian graphical model determination under G-Wishart prior distributions. *Electron. J. Stat.* **6** 168–198. MR2879676

[37] WISHART, J. (1928). The generalised product moment distribution in samples from a normal multivariate population. *Biometrika* **20A** 32–52.

[38] WISHART, J. and BARTLETT, M. S. (1933). The generalised product moment distribution in a normal system. *Math. Proc. Cambridge Philos. Soc.* **29** 260–270.

C. UHLER
LABORATORY FOR INFORMATION AND DECISION SYSTEMS
AND INSTITUTE FOR DATA, SYSTEMS AND SOCIETY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS 02139
USA
E-MAIL: cuhler@mit.edu

A. LENKOSKI
NORWEGIAN COMPUTING CENTER
OSLO
NORWAY
E-MAIL: alex@nr.no

D. RICHARDS
DEPARTMENT OF STATISTICS
PENN STATE UNIVERSITY
UNIVERSITY PARK, PENNSYLVANIA 16802
USA
E-MAIL: richards@stat.psu.edu