Supplement to Bayesian Markov Blanket Estimation

Dinu Kaufmann, Sonali Parbhoo, Aleksander Wieczorek,
Sebastian Keller, David Adametz, Volker Roth
Departement of Mathematics and Computer Science, University of Basel, Switzerland

1 Model

Let $\mathbf{S} \sim \mathcal{W}_{p+q}(n, \mathbf{W})$ be Wishart distributed with n degrees of freedom and inverse covariance $\mathbf{W} = \mathbf{\Sigma}^{-1}$, and define the following partitioning:

$$\mathbf{S} = \begin{array}{c c} & \begin{array}{cc} p & q \end{array} \\ \begin{array}{c} p \\ q \end{array} & \begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{pmatrix} \end{array}, \quad \mathbf{W} = \begin{array}{c c} & \begin{array}{cc} p & q \end{array} \\ \begin{array}{c} p \\ q \end{array} & \begin{pmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{21} & \mathbf{W}_{22} \end{pmatrix} \end{array}. \tag{1}$$

Likelihood

$$\begin{aligned} p(\mathbf{S}|\mathbf{W}) &\propto \det(\mathbf{W})^{\frac{n}{2}} \exp \operatorname{tr}\left(-\frac{1}{2}\mathbf{W}\mathbf{S}\right) \\ &= \det(\mathbf{W}_{11})^{\frac{n}{2}} \det(\mathbf{W}_{22\cdot 1})^{\frac{n}{2}} \\ &\quad \times \exp \operatorname{tr}\left(-\frac{1}{2}\left(\mathbf{W}_{11}\mathbf{S}_{11} + \mathbf{W}_{12}\mathbf{S}_{21} + \mathbf{W}_{21}\mathbf{S}_{12} + \mathbf{W}_{11}^{-1}\mathbf{W}_{12}\mathbf{S}_{22}\mathbf{W}_{21}\right)\right) \\ &\quad \times \exp \operatorname{tr}\left(-\frac{1}{2}\mathbf{W}_{22\cdot 1}\mathbf{S}_{22}\right) \end{aligned} \tag{2}$$

Prior

$$\begin{aligned} p(\mathbf{W}) &\propto \exp \operatorname{tr}\left(-\frac{1}{2}\mathbf{I}\mathbf{W}\right) \prod_{w_{ij}\in \mathbf{W}_{12}} \frac{1}{\sqrt{2\pi t_{ij}}} \exp\left(-\frac{w_{ij}^2}{2t_{ij}}\right) \frac{\gamma^2}{2} \exp\left(-\frac{\gamma^2}{2} t_{ij}\right) \\ &= \exp \operatorname{tr}\left(-\frac{1}{2}\mathbf{I}\mathbf{W}\right) \prod_{i=1}^{p} \exp\left(-\frac{1}{2}\boldsymbol{\beta}_i^\mathsf{T}\mathbf{D}_i\boldsymbol{\beta}_i\right) \prod_{j=p+1}^{p+q} \frac{1}{\sqrt{2\pi t_{ij}}} \frac{\gamma^2}{2} \exp\left(-\frac{\gamma^2}{2} t_{ij}\right) \end{aligned} \tag{3}$$

Here, $\boldsymbol{\beta}_i = (\mathbf{W}_{12})_{i\cdot}$ denotes the ith row of $\mathbf{W}_{12}$, $\mathbf{D}_i = \operatorname{diag}\left((T_{i\cdot})^{-1}\right)$, and $T_{i\cdot} = \operatorname{diag}(t_{i1}, \ldots, t_{ip})$.

Posterior

$$\begin{aligned} p(\mathbf{W}_{11}, \mathbf{W}_{12}|\mathbf{S}) &\propto \det(\mathbf{W}_{11})^{\frac{n}{2}} \\ &\times \exp \operatorname{tr}\left(-\frac{1}{2}\left(\mathbf{W}_{11}(\mathbf{S}_{11}+\mathbf{I}) + \mathbf{W}_{12}\mathbf{S}_{21} + \mathbf{W}_{21}\mathbf{S}_{12} + \mathbf{W}_{11}^{-1}\mathbf{W}_{12}(\mathbf{S}_{22}+\mathbf{I})\mathbf{W}_{21}\right)\right) \\ &\times \prod_{k=1}^{p} \exp\left(-\frac{1}{2}\boldsymbol{\beta}_k^\mathsf{T}\mathbf{D}_k\boldsymbol{\beta}_k\right) \prod_{j=p+1}^{p+q} \frac{1}{\sqrt{2\pi t_{ij}}} \frac{\gamma^2}{2} \exp\left(-\frac{\gamma^2}{2} t_{ij}\right) \end{aligned} \tag{4}$$

2 Wishart Distribution

If $\mathbf{S} \sim \mathcal{W}_{p+q}(n, \mathbf{\Sigma})$, $n > p + q - 1$, then

$$\begin{aligned} p(\mathbf{S}) &\propto \det(\mathbf{\Sigma})^{-\frac{n}{2}} \det(\mathbf{S})^{\frac{n-(p+q)-1}{2}} \exp \operatorname{tr}\left(-\frac{1}{2}\mathbf{\Sigma}^{-1}\mathbf{S}\right) \\ &= \det(\mathbf{W})^{\frac{n}{2}} \det(\mathbf{S})^{\frac{n-(p+q)-1}{2}} \exp \operatorname{tr}\left(-\frac{1}{2}\mathbf{W}\mathbf{S}\right) \end{aligned} \tag{5}$$

where $\mathbf{W} = \mathbf{\Sigma}^{-1}$.

We will prove Lemma 1 from the paper.

Lemma 1 The probability density function of a Wishart distribution factorises in $\mathbf{W}$ as follows:

$$\mathcal{L}_{\mathbf{S}}(\mathbf{W}) \propto \mathcal{L}_1(\mathbf{W}_{11}, \mathbf{W}_{12})\mathcal{L}_2(\mathbf{W}_{22\cdot 1}) \tag{6}$$

Proof The proof is similar to [2]. Factorising the Wishart density according to

$$\begin{aligned} \mathbf{W}\mathbf{S} &= \mathbf{W}_{11}\mathbf{S}_{11} + \mathbf{W}_{12}\mathbf{S}_{21} + \mathbf{W}_{21}\mathbf{S}_{12} + \mathbf{W}_{22}\mathbf{S}_{22} \\ &= \mathbf{W}_{11}\mathbf{S}_{11} + \mathbf{W}_{12}\mathbf{S}_{21} + \mathbf{W}_{21}\mathbf{S}_{12} + (\mathbf{W}_{22} - \mathbf{W}_{21}\mathbf{W}_{11}^{-1}\mathbf{W}_{12} + \mathbf{W}_{21}\mathbf{W}_{11}^{-1}\mathbf{W}_{12})\mathbf{S}_{22} \\ &= \mathbf{W}_{11}\mathbf{S}_{11} + \mathbf{W}_{12}\mathbf{S}_{21} + \mathbf{W}_{21}\mathbf{S}_{12} + \mathbf{W}_{21}\mathbf{W}_{11}^{-1}\mathbf{W}_{12}\mathbf{S}_{22} + \mathbf{W}_{22\cdot 1}\mathbf{S}_{22} \end{aligned} \tag{7}$$

and

$$\det(\mathbf{W})^{\frac{n}{2}} = \det(\mathbf{W}_{11})^{\frac{n}{2}} \det(\mathbf{W}_{22\cdot 1})^{\frac{n}{2}} \tag{8}$$

the factorisation follows immediately with

$$\begin{aligned} &\mathcal{L}_1(\mathbf{W}_{11}, \mathbf{W}_{12}) \\ &= \det(\mathbf{W}_{11})^{\frac{n}{2}} \exp \operatorname{tr}\left(-\frac{1}{2}\left(\mathbf{W}_{11}\mathbf{S}_{11\cdot 2} + \mathbf{W}_{12}\mathbf{S}_{21} + \mathbf{W}_{21}\mathbf{S}_{12} + \mathbf{W}_{21}\mathbf{W}_{11}^{-1}\mathbf{W}_{12}\mathbf{S}_{22}\right)\right) \end{aligned} \tag{9}$$

and

$$\mathcal{L}_2(\mathbf{W}_{22\cdot 1}) = \det(\mathbf{W}_{22\cdot 1})^{\frac{n}{2}} \exp \operatorname{tr}\left(-\frac{1}{2}\left(\mathbf{W}_{22\cdot 1}\mathbf{S}_{22}\right)\right). \tag{10}$$

3 Sampling from $vec(\mathbf{W}_{12}^{\mathsf{T}})|\mathbf{W}_{11}, \mathbf{S}, \{t_{ij}\}$

We will prove Theorem 2, part 1 from the paper.

Theorem 2, Part 1 Vectorised rows of $\mathbf{W}_{12}$ follow a joint Normal distribution

$$vec(\mathbf{W}_{12}^{\mathsf{T}})|\mathbf{W}_{11}, \mathbf{S} \sim \mathcal{N}_{pq}\left(vec(-(\mathbf{S}_{22}+\mathbf{I})^{-\mathsf{T}}\mathbf{S}_{12}^{\mathsf{T}}\mathbf{W}_{11}^{\mathsf{T}}), \mathbf{C}^{-1}\right) \tag{11}$$

where the matrix is $\mathbf{C} = \mathbf{W}_{11}^{-1} \otimes (\mathbf{S}_{22}+\mathbf{I}) + \mathrm{diag}\,(\mathbf{D}_1, \ldots, \mathbf{D}_p)$, and $\mathbf{D}_i = \mathrm{diag}\left((T_{i\cdot})^{-1}\right)$ is a diagonal matrix containing $T_{i\cdot} = (t_{i1}, \ldots, t_{iq})$.

Proof The fist step of the proof is similar to [2]. We factorise the Wishart likelihood according to

$$\begin{aligned}
\mathbf{WS} &= \mathbf{W}_{11}\mathbf{S}_{11} + \mathbf{W}_{12}\mathbf{S}_{21} + \mathbf{W}_{21}\mathbf{S}_{12} + \mathbf{W}_{22}\mathbf{S}_{22} \\
&= \mathbf{W}_{11}(\mathbf{S}_{11} - \mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21} + \mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21}) + \mathbf{W}_{12}\mathbf{S}_{21} + \mathbf{W}_{21}\mathbf{S}_{12} \\
&\quad + (\mathbf{W}_{22} - \mathbf{W}_{21}\mathbf{W}_{11}^{-1}\mathbf{W}_{12} + \mathbf{W}_{21}\mathbf{W}_{11}^{-1}\mathbf{W}_{12})\mathbf{S}_{22} \\
&= \mathbf{W}_{11}\mathbf{S}_{11\cdot 2} + \mathbf{W}_{11}\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21} + \mathbf{W}_{12}\mathbf{S}_{21} + \mathbf{W}_{21}\mathbf{S}_{12} + \mathbf{W}_{21}\mathbf{W}_{11}^{-1}\mathbf{W}_{12}\mathbf{S}_{22} + \mathbf{W}_{22\cdot 1}\mathbf{S}_{22} \\
&= \underbrace{\mathbf{W}_{11}\mathbf{S}_{11\cdot 2}}_{\mathcal{W}} + \underbrace{\mathbf{S}_{22}\left(\mathbf{W}_{12} + \mathbf{W}_{11}\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\right)^{\mathsf{T}}\mathbf{W}_{11}^{-1}\left(\mathbf{W}_{12} + \mathbf{W}_{11}\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\right)}_{\mathcal{MN}} + \underbrace{\mathbf{W}_{22\cdot 1}\mathbf{S}_{22}}_{\mathcal{W}}
\end{aligned} \tag{12}$$

where we changed variables $\mathbf{W}_{22\cdot 1} = \mathbf{W}_{22} - \mathbf{W}_{21}\mathbf{W}_{11}^{-1}\mathbf{W}_{12}$ with Jacobian 1 and integrated over $\mathbf{W}_{22\cdot 1}$. Factorise also the determinants according to

$$\begin{aligned}
\det(\mathbf{W})^{\frac{n}{2}} &= \det(\mathbf{W}_{11})^{\frac{n}{2}} \det(\mathbf{W}_{22\cdot 1})^{\frac{n}{2}} \\
&= \det(\mathbf{W}_{11})^{\frac{n+p}{2}} \underbrace{\det(\mathbf{W}_{11})^{-\frac{p}{2}}}_{\mathcal{MN}} \det(\mathbf{W}_{22\cdot 1})^{\frac{n}{2}}
\end{aligned} \tag{13}$$

and

$$\begin{aligned}
\det(\mathbf{S})^{\frac{n-(p+q)-1}{2}} &= \det\left(\mathbf{S}_{22}\right)^{\frac{n-p-q-1}{2}} \det\left(\mathbf{S}_{11\cdot 2}\right)^{\frac{n-p-q-1}{2}} \\
&= \det(\mathbf{S}_{22})^{\frac{n-p-2q-1}{2}} \underbrace{\det(\mathbf{S}_{22}^{-1})^{-\frac{q}{2}}}_{\mathcal{MN}} \det(\mathbf{S}_{11\cdot 2})^{\frac{n-p-q-1}{2}}.
\end{aligned} \tag{14}$$

To construct the posterior, we include the Wishart part of the prior. This results in

$$\mathbf{W}_{12}|\mathbf{W}_{11}, \mathbf{S} \sim \mathcal{MN}_{p\times q}\left(-\mathbf{W}_{11}\mathbf{S}_{12}(\mathbf{S}_{22}+\mathbf{I})^{-1}, \mathbf{W}_{11}, (\mathbf{S}_{22}+\mathbf{I})^{-1}\right) \tag{15}$$

which is equivalent to

$$vec(\mathbf{W}_{12}^{\mathsf{T}})|\mathbf{W}_{11}, \mathbf{S} \sim \mathcal{N}_{pq}\left(vec(-(\mathbf{S}_{22}+\mathbf{I})^{-\mathsf{T}}\mathbf{S}_{12}^{\mathsf{T}}\mathbf{W}_{11}^{\mathsf{T}}), \mathbf{W}_{11}^{-1} \otimes (\mathbf{S}_{22}+\mathbf{I})\right) \tag{16}$$

where $vec(\mathbf{W}_{12}^{\mathsf{T}})$ are the vectorised rows of matrix $\mathbf{W}_{12}$ and $\otimes$ denotes the Kronecker product. For inclusion of the double exponential prior, it has to be rewritten as

$$\begin{aligned}
\prod_{w_{ij}\in\mathbf{W}_{12}} \frac{1}{\sqrt{2\pi t_{ij}}} \exp\left(-\frac{w_{ij}^2}{2t_{ij}}\right) &= \prod_{i=1}^{p} \exp\left(-\frac{1}{2}(\mathbf{W}_{12})_{i,\cdot}^{\mathsf{T}}\mathbf{D}_i(\mathbf{W}_{12})_{i,\cdot}\right) \\
&= \exp\left(-\frac{1}{2}vec(\mathbf{W}_{12}^{\mathsf{T}})^{\mathsf{T}}\mathbf{D}vec(\mathbf{W}_{12}^{\mathsf{T}})\right)
\end{aligned} \tag{17}$$

where $(\mathbf{W}_{12})_{i,\cdot}$ denotes the ith row of $\mathbf{W}_{12}$, $\mathbf{D}_i = \mathrm{diag}\left((T_{i\cdot})^{-1}\right)$, and $\mathbf{D} = \mathrm{diag}\,(\mathbf{D}_1, \ldots, \mathbf{D}_p)$. The result follows from multiplying the prior by the expanded density in Eq. (16).

4 Matrix Generalised Inverse Gaussian ($\mathcal{MGIG}$) Distribution

$\mathbf{X} \in \mathbb{R}^{p\times p}$ follows a Matrix Generalised Inverse Gaussian (MGIG) distribution, if $\lambda > \frac{p-1}{2}$ and

$$\mathbf{X} \sim \mathcal{MGIG}_{p\times p}(\lambda, \mathbf{A}, \mathbf{B}) \propto \det(\mathbf{X})^{-\lambda-1} \exp\operatorname{tr}\left(-\frac{1}{2}(\mathbf{A}\mathbf{X} + \mathbf{B}\mathbf{X}^{-1})\right). \tag{18}$$

Lemma 2 Let $p(\mathbf{X}) \propto \det(\mathbf{X})^{-\lambda-1}\exp\operatorname{tr}\left(-\frac{1}{2}(\mathbf{A}\mathbf{X} + \mathbf{B}\mathbf{X}^{-1})\right)$ be MGIG distributed, then $\mathbf{W} = \mathbf{X}^{-1}$ is also MGIG distributed:

$$p(\mathbf{W}) \propto \det(\mathbf{W})^{\lambda-p}\exp\operatorname{tr}\left(-\frac{1}{2}(\mathbf{A}\mathbf{W}^{-1} + \mathbf{B}\mathbf{W})\right) \tag{19}$$

Proof Transforming $\mathbf{W} = \mathbf{X}^{-1}$ with $\mathrm{d}\mathbf{X} = -\mathbf{W}^{-1}\mathrm{d}\mathbf{W}\mathbf{W}$

$$\begin{aligned} p(\mathbf{W}) &\propto \det(\mathbf{W})^{-(p+1)}\det(\mathbf{W}^{-1})^{-\lambda-1}\exp\operatorname{tr}\left(-\frac{1}{2}(\mathbf{A}\mathbf{W}^{-1} + \mathbf{B}\mathbf{W})\right) \\ &= \det(\mathbf{W})^{\lambda-p}\exp\operatorname{tr}\left(-\frac{1}{2}(\mathbf{A}\mathbf{W}^{-1} + \mathbf{B}\mathbf{W})\right). \end{aligned} \tag{20}$$

Theorem 2, Part 2 If $\mathbf{S} \sim \mathcal{W}_{p+q}(n, \mathbf{\Sigma})$ is Wishart distributed, then

$$\mathbf{W}_{11}|\mathbf{W}_{12}, \mathbf{S} \sim \mathcal{MGIG}_{p\times p}\left(\frac{n}{2} + p, \mathbf{W}_{12}(\mathbf{S}_{22} + \mathbf{I})\mathbf{W}_{21}, \mathbf{S}_{11} + \mathbf{I}\right) \tag{21}$$

is MGIG distributed.

Proof The proof is similar to [1]. Factorising the Wishart density according to

$$\begin{aligned} &\exp\operatorname{tr}\left(-\frac{1}{2}\mathbf{W}\mathbf{S}\right) = \exp\operatorname{tr}\left(-\frac{1}{2}\left(\mathbf{W}_{11}\mathbf{S}_{11} + \mathbf{W}_{12}\mathbf{S}_{21} + \mathbf{W}_{21}\mathbf{S}_{12} + \mathbf{W}_{22}\mathbf{S}_{22}\right)\right) \\ &= \exp\operatorname{tr}\left(-\frac{1}{2}\left(\mathbf{W}_{11}\mathbf{S}_{11} + \mathbf{W}_{12}\mathbf{S}_{21} + \mathbf{W}_{21}\mathbf{S}_{12} + (\mathbf{W}_{22} - \mathbf{W}_{21}\mathbf{W}_{11}^{-1}\mathbf{W}_{12} + \mathbf{W}_{21}\mathbf{W}_{11}^{-1}\mathbf{W}_{12})\mathbf{S}_{22}\right)\right) \\ &= \exp\operatorname{tr}\left(-\frac{1}{2}\left(\mathbf{W}_{11}\mathbf{S}_{11} + \mathbf{W}_{12}\mathbf{S}_{21} + \mathbf{W}_{21}\mathbf{S}_{12} + (\mathbf{W}_{22\cdot 1} + \mathbf{W}_{21}\mathbf{W}_{11}^{-1}\mathbf{W}_{12})\mathbf{S}_{22}\right)\right) \\ &= \exp\operatorname{tr}\left(-\frac{1}{2}\left(\mathbf{W}_{11}\underbrace{\mathbf{S}_{11}}_{=\mathbf{B}} + \mathbf{W}_{12}\mathbf{S}_{21} + \mathbf{W}_{21}\mathbf{S}_{12} + \underbrace{\mathbf{W}_{12}\mathbf{S}_{22}\mathbf{W}_{21}}_{=\mathbf{A}}\mathbf{W}_{11}^{-1} + \mathbf{W}_{22\cdot 1}\mathbf{S}_{22}\right)\right) \end{aligned} \tag{22}$$

and

$$\det(\mathbf{W})^{\frac{n}{2}} = \det(\mathbf{W}_{11})^{\frac{n}{2}}\det(\mathbf{W}_{22\cdot 1})^{\frac{n}{2}} \tag{23}$$

where we changed variables $\mathbf{W}_{22\cdot 1} = \mathbf{W}_{22} - \mathbf{W}_{21}\mathbf{W}_{11}^{-1}\mathbf{W}_{12}$ with Jacobian 1 and integrated over $\mathbf{W}_{22\cdot 1}$. Comparing Eq. (19) with Eq. (22), we can identify $\mathbf{A}$, $\mathbf{B}$, and λ:

$$\frac{n}{2} \stackrel{!}{=} \lambda - p \quad \Rightarrow \quad \lambda = \frac{n}{2} + p. \tag{24}$$

such that

$$\mathbf{W}_{11}|\mathbf{W}_{12}, \mathbf{S} \sim \mathcal{MGIG}_{p\times p}\left(\frac{n}{2} + p, \mathbf{W}_{12}\mathbf{S}_{22}\mathbf{W}_{21}, \mathbf{S}_{11}\right) \tag{25}$$

Since the double exponential prior does not affect the distribution of $\mathbf{W}_{11}$, we only have to include the Wishart prior. The result follows immediately.

References

[1] Ronald W Butler. Generalized inverse gaussian distributions and their wishart connections. *Scandinavian journal of statistics*, 25(1):69–75, 1998.

[2] Arjun K Gupta and Daya K Nagar. *Matrix variate distributions*, volume 104. CRC Press, 1999.