Web-supplement for:

The Inverse G-Wishart Distribution and Variational Message Passing

BY L. MAESTRINI AND M.P. WAND

University of Technology Sydney

S.1 Natural Parameter Versions and Mappings

Throughout this article we use the "vech" versions of the natural parameter forms of the Multivariate Normal and Inverse G-Wishart distributions. However, Wand (2017) and McLean & Wand (2019) used "vec" versions of these distributions. The "vech" version has the attraction of being more compact since entries of symmetric matrices are not duplicated. However, adoption of the "vech" version entails use of duplication matrices. For implementation in the R language (R Core Team, 2020) we note that the function `duplication.matrix()` in the package matrixcalc (Novomestky, 2012) returns the duplication matrix of a given order.

First we explain the two versions for the Multivariate Normal distribution. Suppose that the $d \times 1$ random vector $\boldsymbol{v}$ has a $N(\boldsymbol{u}, \boldsymbol{\Sigma})$ distribution. Then the density function of $\boldsymbol{v}$ is

$$p(\boldsymbol{v}) \propto \exp\left\{\begin{bmatrix} \boldsymbol{v} \\ \text{vec}(\boldsymbol{v}\boldsymbol{v}^T) \end{bmatrix}^T \boldsymbol{\eta}_v^{\text{vec}}\right\} = \exp\left\{\begin{bmatrix} \boldsymbol{v} \\ \text{vech}(\boldsymbol{v}\boldsymbol{v}^T) \end{bmatrix}^T \boldsymbol{\eta}_v^{\text{vech}}\right\}$$

where

$$\boldsymbol{\eta}_v^{\text{vec}} \equiv \begin{bmatrix} \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \\ -\frac{1}{2}\text{vec}(\boldsymbol{\Sigma}^{-1}) \end{bmatrix} \quad \text{and} \quad \boldsymbol{\eta}_v^{\text{vech}} \equiv \begin{bmatrix} \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \\ -\frac{1}{2}\boldsymbol{D}_d^T\text{vec}(\boldsymbol{\Sigma}^{-1}) \end{bmatrix}.$$

The two natural parameter vectors can be mapped between each other using

$$\boldsymbol{\eta}_v^{\text{vech}} = \text{blockdiag}(\boldsymbol{I}_d, \boldsymbol{D}_d^T)\boldsymbol{\eta}_v^{\text{vec}} \quad \text{and} \quad \boldsymbol{\eta}_v^{\text{vec}} = \text{blockdiag}(\boldsymbol{I}_d, \boldsymbol{D}_d^{+T})\boldsymbol{\eta}_v^{\text{vech}}. \tag{S.1}$$

Now we explain the interplay between the "vec" and "vech" forms of the Inverse G-Wishart distribution. Let the $d \times d$ matrix $\boldsymbol{V}$ have an Inverse-G-Wishart$(G, \xi, \boldsymbol{\Lambda})$ distribution. Then the density function of $\boldsymbol{V}$ is

$$p(\boldsymbol{V}) \propto \exp\left\{\begin{bmatrix} \log|\boldsymbol{V}| \\ \text{vec}(\boldsymbol{V}^{-1}) \end{bmatrix}^T \boldsymbol{\eta}_{\boldsymbol{V}}^{\text{vec}}\right\} = \exp\left\{\begin{bmatrix} \log|\boldsymbol{V}| \\ \text{vech}(\boldsymbol{V}^{-1}) \end{bmatrix}^T \boldsymbol{\eta}_{\boldsymbol{V}}^{\text{vech}}\right\}$$

where

$$\boldsymbol{\eta}_{\boldsymbol{V}}^{\text{vec}} \equiv \begin{bmatrix} -\frac{1}{2}(\xi+1) \\ -\frac{1}{2}\text{vec}(\boldsymbol{\Lambda}) \end{bmatrix} \quad \text{and} \quad \boldsymbol{\eta}_{\boldsymbol{V}}^{\text{vech}} \equiv \begin{bmatrix} -\frac{1}{2}(\xi+1) \\ -\frac{1}{2}\boldsymbol{D}_d^T\text{vec}(\boldsymbol{\Lambda}) \end{bmatrix}.$$

Mappings between the two natural parameter vectors are as follows:

$$\boldsymbol{\eta}_{\boldsymbol{V}}^{\text{vech}} = \text{blockdiag}(1, \boldsymbol{D}_d^T)\,\boldsymbol{\eta}_{\boldsymbol{V}}^{\text{vec}} \quad \text{and} \quad \boldsymbol{\eta}_{\boldsymbol{V}}^{\text{vec}} = \text{blockdiag}(1, \boldsymbol{D}_d^{+T})\,\boldsymbol{\eta}_{\boldsymbol{V}}^{\text{vech}}. \tag{S.2}$$

S.2 Justification of Algorithm 2

We now provide justification for Algorithm 2, which is concerned with the graph and natural parameter updates for the iterated Inverse G-Wishart fragment.

S.2.1 The Updates for $m_{p(\mathbf{\Sigma}|\boldsymbol{A})\,\rightarrow\,\mathbf{\Sigma}}(\mathbf{\Sigma})$

As a function of $\mathbf{\Sigma}$,

$$\log\,p(\mathbf{\Sigma}|\boldsymbol{A})=\left[\begin{array}{c}\log|\mathbf{\Sigma}|\\ \mathrm{vech}(\mathbf{\Sigma}^{-1})\end{array}\right]^T\left[\begin{array}{c}-\frac{1}{2}\,(\xi+2)\\ -\frac{1}{2}\boldsymbol{D}_d^T\mathrm{vec}(\boldsymbol{A}^{-1})\end{array}\right]+\mathrm{const}$$

where 'const' denotes terms that do not depend on $\mathbf{\Sigma}$. Hence

$$m_{p(\mathbf{\Sigma}|\boldsymbol{A})\,\rightarrow\,\mathbf{\Sigma}}(\mathbf{\Sigma})=\exp\left\{\left[\begin{array}{c}\log|\mathbf{\Sigma}|\\ \mathrm{vech}(\mathbf{\Sigma}^{-1})\end{array}\right]^T\boldsymbol{\eta}_{p(\mathbf{\Sigma}|\boldsymbol{A})\,\rightarrow\,\mathbf{\Sigma}}\right\}$$

where

$$\boldsymbol{\eta}_{p(\mathbf{\Sigma}|\boldsymbol{A})\,\rightarrow\,\mathbf{\Sigma}}=\left[\begin{array}{c}-\frac{1}{2}\,(\xi+2)\\ -\frac{1}{2}\boldsymbol{D}_d^T\mathrm{vec}\Big(E_q(\boldsymbol{A}^{-1})\Big)\end{array}\right]\qquad\text{(S.3)}$$

and E_q denotes expectation with respect to the normalization of

$$m_{p(\mathbf{\Sigma}|\boldsymbol{A})\,\rightarrow\,\boldsymbol{A}}(\boldsymbol{A})\,m_{\boldsymbol{A}\,\rightarrow\,p(\mathbf{\Sigma}|\boldsymbol{A})}(\boldsymbol{A}).$$

Let $q(\boldsymbol{A})$ denote this normalized density function. Then $q(\boldsymbol{A})$ is an Inverse-G-Wishart distribution with graph $G_{p(\mathbf{\Sigma}|\boldsymbol{A})\,\rightarrow\,\boldsymbol{A}}\in\{G_{\text{full}},G_{\text{diag}}\}$ and natural parameter vector $\boldsymbol{\eta}_{p(\mathbf{\Sigma}|\boldsymbol{A})\,\leftrightarrow\,\boldsymbol{A}}$. From Result 6,

$$E_q(\boldsymbol{A}^{-1})=\begin{cases}\Big\{\big(\boldsymbol{\eta}_{p(\mathbf{\Sigma}|\boldsymbol{A})\,\leftrightarrow\,\boldsymbol{A}}\big)_1+\frac{1}{2}(d+1)\Big\}\Big\{\mathrm{vec}^{-1}\Big(\boldsymbol{D}_d^{+T}\big(\boldsymbol{\eta}_{p(\mathbf{\Sigma}|\boldsymbol{A})\,\leftrightarrow\,\boldsymbol{A}}\big)_2\Big)\Big\}^{-1}\\ \qquad\qquad\text{if } G_{p(\mathbf{\Sigma}|\boldsymbol{A})\,\rightarrow\,\boldsymbol{A}}=G_{\text{full}},\\ \Big\{\big(\boldsymbol{\eta}_{p(\mathbf{\Sigma}|\boldsymbol{A})\,\leftrightarrow\,\boldsymbol{A}}\big)_1+1\Big\}\Big\{\mathrm{vec}^{-1}\Big(\boldsymbol{D}_d^{+T}\big(\boldsymbol{\eta}_{p(\mathbf{\Sigma}|\boldsymbol{A})\,\leftrightarrow\,\boldsymbol{A}}\big)_2\Big)\Big\}^{-1}\\ \qquad\qquad\text{if } G_{p(\mathbf{\Sigma}|\boldsymbol{A})\,\rightarrow\,\boldsymbol{A}}=G_{\text{diag}}.\end{cases}$$

Noting that the first factor of $E_q(\boldsymbol{A}^{-1})$ is $(\boldsymbol{\eta}_{p(\mathbf{\Sigma}|\boldsymbol{A})\,\leftrightarrow\,\boldsymbol{A}})_1+\omega_1$, where

$$\omega_1=\omega_1(d,G)=\begin{cases}(d+1)/2 & \text{if } G=G_{\text{full}}\\ 1 & \text{if } G=G_{\text{diag}},\end{cases}$$

the first update of $E_q(\boldsymbol{A}^{-1})$ in Algorithm 2 is justified. Lastly, we need to possibly adjust for the fact that $m_{p(\mathbf{\Sigma}|\boldsymbol{A})\,\rightarrow\,\mathbf{\Sigma}}(\mathbf{\Sigma})$ is proportional to an Inverse G-Wishart density function with $G=G_{\text{diag}}$. This is achieved by the conditional step:

$$\text{If } G_{p(\mathbf{\Sigma}|\boldsymbol{A})\,\rightarrow\,\mathbf{\Sigma}}=G_{\text{diag}}\text{ then } E_q(\boldsymbol{A}^{-1})\longleftarrow\mathrm{diag}\Big\{\mathrm{diagonal}\Big(E_q(\boldsymbol{A}^{-1})\Big)\Big\}.$$

S.2.2 The Updates for $m_{p(\mathbf{\Sigma}|\boldsymbol{A})\,\rightarrow\,\boldsymbol{A}}(\boldsymbol{A})$

As a function of $\boldsymbol{A}$,

$$\log\,p(\mathbf{\Sigma}|\boldsymbol{A})=\left[\begin{array}{c}\log|\boldsymbol{A}|\\ \mathrm{vech}(\boldsymbol{A}^{-1})\end{array}\right]^T\left[\begin{array}{c}-(\xi+2-2\omega_2)/2\\ -\frac{1}{2}\boldsymbol{D}_d^T\mathrm{vec}(\mathbf{\Sigma}^{-1})\end{array}\right]+\mathrm{const}$$

where

$$\omega_2=\omega_2(d,G)=\begin{cases}(d+1)/2 & \text{if } G=G_{\text{full}},\\ 1 & \text{if } G=G_{\text{diag}}\end{cases}\qquad\text{(S.4)}$$

and 'const' denotes terms that do not depend on $\boldsymbol{A}$. Hence

$$m_{p(\boldsymbol{\Sigma}|\boldsymbol{A})\,\to\,\boldsymbol{A}}(\boldsymbol{A}) = \exp\left\{\begin{bmatrix}\log|\boldsymbol{A}|\\ \text{vech}(\boldsymbol{A}^{-1})\end{bmatrix}^T \boldsymbol{\eta}_{p(\boldsymbol{\Sigma}|\boldsymbol{A})\,\to\,\boldsymbol{A}}\right\}$$

where

$$\boldsymbol{\eta}_{p(\boldsymbol{\Sigma}|\boldsymbol{A})\,\to\,\boldsymbol{A}} = \begin{bmatrix}-(\xi+2-2\omega_2)/2\\ -\frac{1}{2}\boldsymbol{D}_d^T\text{vec}\Big(E_q(\boldsymbol{\Sigma}^{-1})\Big)\end{bmatrix} \tag{S.5}$$

and E_q denotes expectation with respect to the normalization of

$$m_{p(\boldsymbol{\Sigma}|\boldsymbol{A})\,\to\,\boldsymbol{\Sigma}}(\boldsymbol{\Sigma})\, m_{\boldsymbol{\Sigma}\,\to\,p(\boldsymbol{\Sigma}|\boldsymbol{A})}(\boldsymbol{\Sigma}).$$

Let $q(\boldsymbol{\Sigma})$ denote this normalized density function. Then $q(\boldsymbol{\Sigma})$ is an Inverse-G-Wishart distribution with graph $G_{p(\boldsymbol{\Sigma}|\boldsymbol{A})\,\to\,\boldsymbol{\Sigma}} \in \{G_{\text{full}}, G_{\text{diag}}\}$ and natural parameter vector $\boldsymbol{\eta}_{p(\boldsymbol{\Sigma}|\boldsymbol{A})\,\leftrightarrow\,\boldsymbol{\Sigma}}$. From Result 6,

$$E_q(\boldsymbol{\Sigma}^{-1}) = \begin{cases}\Big\{\big(\boldsymbol{\eta}_{p(\boldsymbol{\Sigma}|\boldsymbol{A})\,\leftrightarrow\,\boldsymbol{\Sigma}}\big)_1 + \frac{1}{2}(d+1)\Big\}\Big\{\text{vec}^{-1}\Big(\boldsymbol{D}_d^{+T}\big(\boldsymbol{\eta}_{p(\boldsymbol{\Sigma}|\boldsymbol{A})\,\leftrightarrow\,\boldsymbol{\Sigma}}\big)_2\Big)\Big\}^{-1} & \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } G_{p(\boldsymbol{\Sigma}|\boldsymbol{A})\,\to\,\boldsymbol{\Sigma}} = G_{\text{full}}, & \\ \Big\{\big(\boldsymbol{\eta}_{p(\boldsymbol{\Sigma}|\boldsymbol{A})\,\leftrightarrow\,\boldsymbol{\Sigma}}\big)_1 + 1\Big\}\Big\{\text{vec}^{-1}\Big(\boldsymbol{D}_d^{+T}\big(\boldsymbol{\eta}_{p(\boldsymbol{\Sigma}|\boldsymbol{A})\,\leftrightarrow\,\boldsymbol{\Sigma}}\big)_2\Big)\Big\}^{-1} & \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } G_{p(\boldsymbol{\Sigma}|\boldsymbol{A})\,\to\,\boldsymbol{\Sigma}} = G_{\text{diag}}. & \end{cases}$$

Noting that the first factor of $E_q(\boldsymbol{\Sigma}^{-1})$ is $(\boldsymbol{\eta}_{p(\boldsymbol{\Sigma}|\boldsymbol{A})\,\leftrightarrow\,\boldsymbol{\Sigma}})_1 + \omega_2$, where ω_2 is given by (S.4), the first update of $E_q(\boldsymbol{\Sigma}^{-1})$ in Algorithm 2 is justified. Finally, there is the possible need to adjust for the fact that $m_{p(\boldsymbol{\Sigma}|\boldsymbol{A})\,\to\,\boldsymbol{A}}(\boldsymbol{A})$ is proportional to an Inverse G-Wishart density function with $G = G_{\text{diag}}$. This is achieved by the conditional step:

$$\text{If } G_{p(\boldsymbol{\Sigma}|\boldsymbol{A})\,\to\,\boldsymbol{A}} = G_{\text{diag}} \text{ then } E_q(\boldsymbol{\Sigma}^{-1}) \longleftarrow \text{diag}\Big\{\text{diagonal}\Big(E_q(\boldsymbol{\Sigma}^{-1})\Big)\Big\}.$$

S.3 Illustrative Example Variational Message Passing Details

The variational message passing approach to fitting and approximate inference for statistical models is still quite a new concept. In this section we provide details on the approach for the illustrative example involving the t response linear mixed model described in Section 8.

S.3.1 Data and Hyperparameter Inputs

Let $\boldsymbol{y}$ be the vector of responses as defined in (21). Also, let

$$\boldsymbol{C} = [\boldsymbol{X}\ \boldsymbol{Z}]$$

be the full design matrix, where the matrices $\boldsymbol{X}$ and $\boldsymbol{Z}$ are as defined in (21). The data inputs are $\boldsymbol{y}$ and $\boldsymbol{C}$.

The hyperparameter inputs are

$$\sigma_{\boldsymbol{\beta}}, s_{\sigma^2}, \lambda_\nu, s_{\boldsymbol{\Sigma},1}, \ldots s_{\boldsymbol{\Sigma},q} > 0.$$

S.3.2 Factor to Stochastic Node Parameter Initialisations

Initialize $G_{p(\boldsymbol{A})\,\rightarrow\,\boldsymbol{A}}$ and $\boldsymbol{\eta}_{p(\boldsymbol{A})\,\rightarrow\,\boldsymbol{A}}$ via a call to Algorithm 1 with hyperparameter inputs:

$$G_{\boldsymbol{\Theta}}=G_{\text{diag}},\quad \xi_{\boldsymbol{\Theta}}=1\quad\text{and}\quad \boldsymbol{\Lambda}_{\boldsymbol{\Theta}}=\{2\text{diag}(s^2_{\boldsymbol{\Sigma},1},\ldots s^2_{\boldsymbol{\Sigma},q})\}^{-1}.$$

Initialize $G_{p(a)\,\rightarrow\,a}$ and $\boldsymbol{\eta}_{p(a)\,\rightarrow\,a}$ via a call to Algorithm 1 with hyperparameter inputs:

$$G_{\boldsymbol{\Theta}}=G_{\text{diag}},\quad \xi_{\boldsymbol{\Theta}}=1\quad\text{and}\quad \boldsymbol{\Lambda}_{\boldsymbol{\Theta}}=(s^2_{\sigma})^{-1}.$$

Note that the initialisations of $G_{p(\boldsymbol{A})\,\rightarrow\,\boldsymbol{A}}$, $\boldsymbol{\eta}_{p(\boldsymbol{A})\,\rightarrow\,\boldsymbol{A}}$, $G_{p(a)\,\rightarrow\,a}$ and $\boldsymbol{\eta}_{p(a)\,\rightarrow\,a}$ are part of the prior impositions for $\boldsymbol{\Sigma}$ and σ^2. These four factor to stochastic node parameters remain constant throughout the variational message passing iterations.

Initialize

$$\boldsymbol{\eta}_{p(v)\,\rightarrow\,v}\longleftarrow\begin{bmatrix}0\\-\lambda_{\nu}\end{bmatrix}.$$

This initialization of $\boldsymbol{\eta}_{p(v)\,\rightarrow\,v}$ corresponds to the prior imposition for v. This factor to stochastic node natural parameter remains constant throughout the variational message passing iterations.

The remaining factor to stochastic node natural parameters in the Figure 5 factor graph are updated in the variational message passing iterations, but require initial values. In theory, they can be set to any legal value according to the relevant exponential family. The following initialisations, which are used in the code that produced Figure 6, are simple legal natural parameter vectors:

$$G_{p(\boldsymbol{\Sigma}|\boldsymbol{A})\,\rightarrow\,\boldsymbol{A}}\longleftarrow G_{\text{diag}},\quad \boldsymbol{\eta}_{p(\boldsymbol{\Sigma}|\boldsymbol{A})\,\rightarrow\,\boldsymbol{A}}\longleftarrow\begin{bmatrix}-\frac{1}{2}\\-\frac{1}{2}\boldsymbol{D}_q^T\text{vec}(\boldsymbol{I}_q)\end{bmatrix},$$

$$G_{p(\boldsymbol{\Sigma}|\boldsymbol{A})\,\rightarrow\,\boldsymbol{\Sigma}}\longleftarrow G_{\text{full}},\quad \boldsymbol{\eta}_{p(\boldsymbol{\Sigma}|\boldsymbol{A})\,\rightarrow\,\boldsymbol{\Sigma}}\longleftarrow\begin{bmatrix}-\frac{1}{2}\\-\frac{1}{2}\boldsymbol{D}_q^T\text{vec}(\boldsymbol{I}_q)\end{bmatrix},$$

$$G_{p(\sigma^2|\,a)\,\rightarrow\,a}\longleftarrow G_{\text{diag}},\quad \boldsymbol{\eta}_{p(\sigma^2|a)\,\rightarrow\,a}\longleftarrow\begin{bmatrix}-2\\-1\end{bmatrix},$$

$$G_{p(\sigma^2|\,a)\,\rightarrow\,\sigma^2}\longleftarrow G_{\text{full}},\quad \boldsymbol{\eta}_{p(\sigma^2|a)\,\rightarrow\,\sigma^2}\longleftarrow\begin{bmatrix}-2\\-1\end{bmatrix},$$

$$\boldsymbol{\eta}_{p(\boldsymbol{\beta},\boldsymbol{u}|\boldsymbol{\Sigma})\,\rightarrow\,\boldsymbol{\Sigma}}\longleftarrow\begin{bmatrix}-\frac{1}{2}\\-\frac{1}{2}\text{vec}(\boldsymbol{I}_q)\end{bmatrix},\quad \boldsymbol{\eta}_{p(\boldsymbol{\beta},\boldsymbol{u}|\boldsymbol{\Sigma})\,\rightarrow\,(\boldsymbol{\beta},\boldsymbol{u})}\longleftarrow\begin{bmatrix}\boldsymbol{0}_{p+mq}\\-\frac{1}{2}\text{vec}(\boldsymbol{I}_{p+mq})\end{bmatrix},$$

$$\boldsymbol{\eta}_{p(\boldsymbol{y}|\,\boldsymbol{\beta},\boldsymbol{u},\sigma^2,\boldsymbol{b})\,\rightarrow\,(\boldsymbol{\beta},\boldsymbol{u})}\longleftarrow\begin{bmatrix}\boldsymbol{0}_{p+mq}\\-\frac{1}{2}\text{vec}(\boldsymbol{I}_{p+mq})\end{bmatrix},\quad \boldsymbol{\eta}_{p(\boldsymbol{y}|\,\boldsymbol{\beta},\boldsymbol{u},\sigma^2,\boldsymbol{b})\,\rightarrow\,\sigma^2}\longleftarrow\begin{bmatrix}-2\\-1\end{bmatrix}$$

and

$$\boldsymbol{\eta}_{p(\boldsymbol{b}|\,v)\,\rightarrow\,v}\longleftarrow\begin{bmatrix}1\\-1.1\end{bmatrix}.$$

The messages involving the b_ℓ, $1\le\ell\le N$, nodes do not need to be included here since there messages are subsumed in the calculations used for the natural parameter updates for the model parameters in Algorithm 2 of McLean & Wand (2019).

S.3.3 Variational Message Passing Iterations

With all factor to stochastic node initialisations accomplished, now we describe the iterative updates inside the variational message passing cycle loop. Each iteration involves:

- updating the stochastic node to factor message parameters.
- updating the factor to stochastic node message parameters.

S.3.3.1 Stochastic Node to Factor Message Parameter Updates

The stochastic node to factor message updates are quite simple and follow from, e.g., equation (7) of Wand (2017). For the Figure 5 factor graph the updates are:

$$
\begin{gathered}
G_{\boldsymbol{A}\to p(\boldsymbol{\Sigma}|\boldsymbol{A})} \longleftarrow G_{p(\boldsymbol{A})\to\boldsymbol{A}},\quad \boldsymbol{\eta}_{\boldsymbol{A}\to p(\boldsymbol{\Sigma}|\boldsymbol{A})} \longleftarrow \boldsymbol{\eta}_{p(\boldsymbol{A})\to\boldsymbol{A}},\\
G_{\boldsymbol{\Sigma}\to p(\boldsymbol{\Sigma}|\boldsymbol{A})} \longleftarrow G_{\text{full}},\quad \boldsymbol{\eta}_{\boldsymbol{\Sigma}\to p(\boldsymbol{\Sigma}|\boldsymbol{A})} \longleftarrow \boldsymbol{\eta}_{p(\boldsymbol{\beta},\boldsymbol{u}|\boldsymbol{\Sigma})\to\boldsymbol{\Sigma}},\\
\boldsymbol{\eta}_{\boldsymbol{\Sigma}\to p(\boldsymbol{\beta},\boldsymbol{u}|\boldsymbol{\Sigma})} \longleftarrow \boldsymbol{\eta}_{p(\boldsymbol{\Sigma}|\boldsymbol{A})\to\boldsymbol{\Sigma}},\quad \boldsymbol{\eta}_{(\boldsymbol{\beta},\boldsymbol{u})\to p(\boldsymbol{\beta},\boldsymbol{u}|\boldsymbol{\Sigma})} \longleftarrow \boldsymbol{\eta}_{p(\boldsymbol{y}|\,\boldsymbol{\beta},\boldsymbol{u},\sigma^2,\boldsymbol{b})\to(\boldsymbol{\beta},\boldsymbol{u})},\\
\boldsymbol{\eta}_{(\boldsymbol{\beta},\boldsymbol{u})\to p(\boldsymbol{y}|\,\boldsymbol{\beta},\boldsymbol{u},\sigma^2,\boldsymbol{b})} \longleftarrow \boldsymbol{\eta}_{p(\boldsymbol{\beta},\boldsymbol{u}|\boldsymbol{\Sigma})\to(\boldsymbol{\beta},\boldsymbol{u})},\quad G_{a\to p(\sigma^2|a)} \longleftarrow G_{p(a)\to a},\\
\boldsymbol{\eta}_{a\to p(\sigma^2|a)} \longleftarrow \boldsymbol{\eta}_{p(a)\to a},\quad G_{\sigma^2\to p(\sigma^2|a)} \longleftarrow G_{\text{full}},\\
\boldsymbol{\eta}_{\sigma^2\to p(\sigma^2|a)} \longleftarrow \boldsymbol{\eta}_{p(\boldsymbol{y}|\,\boldsymbol{\beta},\boldsymbol{u},\sigma^2,\boldsymbol{b})\to\sigma^2},\quad \boldsymbol{\eta}_{\sigma^2\to p(\boldsymbol{y}|\,\boldsymbol{\beta},\boldsymbol{u},\sigma^2,\boldsymbol{b})} \longleftarrow \boldsymbol{\eta}_{p(\sigma^2|a)\to\sigma^2}
\end{gathered}
$$

and

$$\boldsymbol{\eta}_{\upsilon\to p(\boldsymbol{b}|\,\upsilon)} \longleftarrow \boldsymbol{\eta}_{p(\upsilon)\to\upsilon}.$$

Some additional remarks concerning stochastic node to factor updates are:

- The stochastic node to factor messages corresponding to the extremities of the Figure 5 factor graph, such as the message from $\boldsymbol{A}$ to $p(\boldsymbol{A})$, are not required in the variational message passing iterations. Therefore, updates for these messages can be omitted.
- Some of the stochastic node to factor message parameter updates, such as that for $\boldsymbol{\eta}_{\boldsymbol{A}\to p(\boldsymbol{\Sigma}|\boldsymbol{A})}$, remain constant throughout the iterations. However, for simplicity of exposition, we list all of the updates together.

S.3.3.2 Factor to Stochastic Node Message Parameter Updates

The updates for the parameters of factor to stochastic node messages are a good deal more complicated than the reverse messages. For the illustrative example, these updates are encapsulated in three algorithms across three different articles. Algorithm 2 plays an important role for the variance and covariance matrix parameter parts of the factor graph.

Use Algorithm 2 with:

> **Shape Parameter Input**: 1.
> **Graph Inputs**: $G_{\sigma^2\to p(\sigma^2|a)}$, $G_{a\to p(\sigma^2|a)}$.
> **Natural Parameter Inputs:** $\boldsymbol{\eta}_{\sigma^2\to p(\sigma^2|a)}$, $\boldsymbol{\eta}_{p(\sigma^2|a)\to\sigma^2}$ $\boldsymbol{\eta}_{a\to p(\sigma^2|a)}$, $\boldsymbol{\eta}_{p(\sigma^2|a)\to a}$
> **Outputs:** $G_{p(\sigma^2|\,a)\to\sigma^2}$, $\boldsymbol{\eta}_{p(\sigma^2|a)\to\sigma^2}$, $G_{p(\sigma^2|\,a)\to a}$, $\boldsymbol{\eta}_{p(\sigma^2|a)\to a}$

Use Algorithm 2 with:

> **Shape Parameter Input**: $2q$
> **Graph Inputs**: $G_{\boldsymbol{\Sigma}\to p(\boldsymbol{\Sigma}|\boldsymbol{A})}$, $G_{\boldsymbol{A}\to p(\boldsymbol{\Sigma}|\boldsymbol{A})}$

Natural Parameter Inputs: $\boldsymbol{\eta}_{\boldsymbol{\Sigma}\,\to\, p(\boldsymbol{\Sigma}|\boldsymbol{A})}$, $\boldsymbol{\eta}_{p(\boldsymbol{\Sigma}|\boldsymbol{A})\,\to\,\boldsymbol{\Sigma}}$, $\boldsymbol{\eta}_{\boldsymbol{A}\,\to\, p(\boldsymbol{\Sigma}|\boldsymbol{A})}$, $\boldsymbol{\eta}_{p(\boldsymbol{\Sigma}|\boldsymbol{A})\,\to\,\boldsymbol{A}}$
Outputs: $G_{p(\boldsymbol{\Sigma}|\boldsymbol{A})\,\to\,\boldsymbol{\Sigma}}$, $\boldsymbol{\eta}_{p(\boldsymbol{\Sigma}|\boldsymbol{A})\,\to\,\boldsymbol{\Sigma}}$, $G_{p(\boldsymbol{\Sigma}|\boldsymbol{A})\,\to\,\boldsymbol{A}}$, $\boldsymbol{\eta}_{p(\boldsymbol{\Sigma}|\boldsymbol{A})\,\to\,\boldsymbol{A}}$

Use the Gaussian Penalisation Fragment of Wand (2017, Section 4.1.4):

Hyperparameter Input: $\sigma^2_{\boldsymbol{\beta}}$
Natural Parameter Inputs: $\boldsymbol{\eta}_{(\boldsymbol{\beta},\boldsymbol{u})\,\to\, p(\boldsymbol{\beta},\boldsymbol{u}|\boldsymbol{\Sigma})}$, $\boldsymbol{\eta}_{p(\boldsymbol{\beta},\boldsymbol{u}|\boldsymbol{\Sigma})\,\to\,(\boldsymbol{\beta},\boldsymbol{u})}$,
$\boldsymbol{\eta}_{\boldsymbol{\Sigma}\,\to\, p(\boldsymbol{\beta},\boldsymbol{u}|\boldsymbol{\Sigma})}$, $\boldsymbol{\eta}_{p(\boldsymbol{\beta},\boldsymbol{u}|\boldsymbol{\Sigma})\,\to\,\boldsymbol{\Sigma}}$
Outputs: $\boldsymbol{\eta}_{p(\boldsymbol{\beta},\boldsymbol{u}|\boldsymbol{\Sigma})\,\to\,(\boldsymbol{\beta},\boldsymbol{u})}$, $\boldsymbol{\eta}_{p(\boldsymbol{\beta},\boldsymbol{u}|\boldsymbol{\Sigma})\,\to\,\boldsymbol{\Sigma}}$

Use the t Likelihood Fragment of McLean & Wand (2019, Algorithm 2):

Data Inputs: $\boldsymbol{y}$, $\boldsymbol{C}$
Natural Parameter Inputs: $\boldsymbol{\eta}_{(\boldsymbol{\beta},\boldsymbol{u})\,\to\, p(\boldsymbol{y}|\,\boldsymbol{\beta},\boldsymbol{u},\sigma^2,\boldsymbol{b})}$, $\boldsymbol{\eta}_{p(\boldsymbol{y}|\,\boldsymbol{\beta},\boldsymbol{u},\sigma^2,\boldsymbol{b})\,\to\,(\boldsymbol{\beta},\boldsymbol{u})}$,
$\boldsymbol{\eta}_{\sigma^2\,\to\, p(\boldsymbol{y}|\,\boldsymbol{\beta},\boldsymbol{u},\sigma^2,\boldsymbol{b})}$, $\boldsymbol{\eta}_{p(\boldsymbol{y}|\,\boldsymbol{\beta},\boldsymbol{u},\sigma^2,\boldsymbol{b})\,\to\,\sigma^2}$,
$\boldsymbol{\eta}_{\upsilon\,\to\, p(\boldsymbol{b}|\,\upsilon)}$, $\boldsymbol{\eta}_{p(\boldsymbol{b}|\,\upsilon)\,\to\,\upsilon}$
Outputs: $\boldsymbol{\eta}_{p(\boldsymbol{y}|\,\boldsymbol{\beta},\boldsymbol{u},\sigma^2,\boldsymbol{b})\,\to\,(\boldsymbol{\beta},\boldsymbol{u})}$, $\boldsymbol{\eta}_{p(\boldsymbol{y}|\,\boldsymbol{\beta},\boldsymbol{u},\sigma^2,\boldsymbol{b})\,\to\,\sigma^2}$, $\boldsymbol{\eta}_{p(\boldsymbol{b}|\,\upsilon)\,\to\,\upsilon}$

Regarding, the last two fragment updates it should be noted that Wand (2017) and McLean & Wand (2019) work with the "vec" versions of Multivariate Normal and Inverse G-Wishart natural parameter vectors. To match the "vech" natural parameter forms used in Algorithms 1 and 2 of the current article conversions given by (S.1) and (S.2) are required.

S.3.4 Determination of Posterior Density Function Approximations

After convergence of the variational message passing iterations, the optimal q^*-densities for each stochastic node are obtained by multiplying each of the messages that pass messages to that node. See, for example, (10) of Wand (2017). We now give details for the model parameters $\boldsymbol{\Sigma}$, σ^2, $(\boldsymbol{\beta},\boldsymbol{u})$ and υ.

S.3.4.1 Determination of $q^*(\boldsymbol{\Sigma})$

From (10) of Wand (2017):

$$q^*(\boldsymbol{\Sigma}) \propto \exp\left\{\left[\begin{array}{c}\log|\boldsymbol{\Sigma}|\\ \text{vech}(\boldsymbol{\Sigma}^{-1})\end{array}\right]^T\left(\boldsymbol{\eta}_{p(\boldsymbol{\Sigma}|\boldsymbol{A})\,\to\,\boldsymbol{\Sigma}}+\boldsymbol{\eta}_{p(\boldsymbol{\beta},\boldsymbol{u}|\boldsymbol{\Sigma})\,\to\,\boldsymbol{\Sigma}}\right)\right\}.$$

It is apparent that $q^*(\boldsymbol{\Sigma})$ is an Inverse Wishart density function with natural parameter vector

$$\boldsymbol{\eta}_{q(\boldsymbol{\Sigma})} \equiv \boldsymbol{\eta}_{p(\boldsymbol{\Sigma}|\boldsymbol{A})\,\to\,\boldsymbol{\Sigma}}+\boldsymbol{\eta}_{p(\boldsymbol{\beta},\boldsymbol{u}|\boldsymbol{\Sigma})\,\to\,\boldsymbol{\Sigma}}.$$

S.3.4.2 Determination of $q^*(\sigma^2)$

Using (10) of Wand (2017):

$$q^*(\sigma^2) \propto \exp\left\{\left[\begin{array}{c}\log(\sigma^2)\\ 1/\sigma^2\end{array}\right]^T\left(\boldsymbol{\eta}_{p(\sigma^2|a)\,\to\,\sigma^2}+\boldsymbol{\eta}_{p(\boldsymbol{y}|\,\boldsymbol{\beta},\boldsymbol{u},\sigma^2,\boldsymbol{b})\,\to\,\sigma^2}\right)\right\}.$$

We see that $q^*(\sigma^2)$ is an Inverse Chi-Squared density function with natural parameter vector

$$\boldsymbol{\eta}_{q(\sigma^2)} \equiv \boldsymbol{\eta}_{p(\sigma^2|a)\,\to\,\sigma^2}+\boldsymbol{\eta}_{p(\boldsymbol{y}|\,\boldsymbol{\beta},\boldsymbol{u},\sigma^2,\boldsymbol{b})\,\to\,\sigma^2}.$$

S.3.4.3 Determination of $q^*(\boldsymbol{\beta},\boldsymbol{u})$

Another application of (10) of Wand (2017) leads to:

$$q^*(\boldsymbol{\beta},\boldsymbol{u})\propto\exp\left\{\left[\begin{array}{c}\boldsymbol{\beta}\\ \boldsymbol{u}\\ \text{vech}\left(\left[\begin{array}{c}\boldsymbol{\beta}\\ \boldsymbol{u}\end{array}\right]\left[\begin{array}{c}\boldsymbol{\beta}\\ \boldsymbol{u}\end{array}\right]^T\right)\end{array}\right]^T\left(\boldsymbol{\eta}_{p(\boldsymbol{\beta},\boldsymbol{u}|\boldsymbol{\Sigma})\,\rightarrow\,(\boldsymbol{\beta},\boldsymbol{u})}+\boldsymbol{\eta}_{p(\boldsymbol{y}|\,\boldsymbol{\beta},\boldsymbol{u},\sigma^2,\boldsymbol{b})\,\rightarrow\,(\boldsymbol{\beta},\boldsymbol{u})}\right)\right\}.$$

We then have $q^*(\boldsymbol{\beta},\boldsymbol{u})$ having a Multivariate Normal density function with natural parameter vector

$$\boldsymbol{\eta}_{q(\boldsymbol{\beta},\boldsymbol{u})}\equiv\boldsymbol{\eta}_{p(\boldsymbol{\beta},\boldsymbol{u}|\boldsymbol{\Sigma})\,\rightarrow\,(\boldsymbol{\beta},\boldsymbol{u})}+\boldsymbol{\eta}_{p(\boldsymbol{y}|\,\boldsymbol{\beta},\boldsymbol{u},\sigma^2,\boldsymbol{b})\,\rightarrow\,(\boldsymbol{\beta},\boldsymbol{u})}.$$

S.3.4.4 Determination of $q^*(\upsilon)$

One last application of (10) of Wand (2017) gives:

$$q^*(\upsilon)\propto\exp\left\{\left[\begin{array}{c}\upsilon\log(\upsilon)-\log\{\Gamma(\upsilon)\}\\ \upsilon\end{array}\right]^T\left(\boldsymbol{\eta}_{p(\boldsymbol{b}|\,\upsilon)\,\rightarrow\,\upsilon}+\boldsymbol{\eta}_{p(\upsilon)\,\rightarrow\,\upsilon}\right)\right\}.$$

Therefore, $q^*(\upsilon)$ is a Moon Rock density function with natural parameter vector

$$\boldsymbol{\eta}_{q(\upsilon)}\equiv\boldsymbol{\eta}_{p(\boldsymbol{b}|\,\upsilon)\,\rightarrow\,\upsilon}+\boldsymbol{\eta}_{p(\upsilon)\,\rightarrow\,\upsilon}.$$

S.3.5 Conversion from Natural Parameters to Common Parameters

A final set of steps involves conversion of the q^*-densities to common parameter forms.

S.3.5.1 Conversion of $q^*(\boldsymbol{\Sigma})$ to Common Parameter Form

The common parameter form of $q^*(\boldsymbol{\Sigma})$ is the Inverse-G-Wishart$(G_{\text{full}},\xi_{q(\boldsymbol{\Sigma})},\boldsymbol{\Lambda}_{q(\boldsymbol{\Sigma})})$ density function where

$$\xi_{q(\boldsymbol{\Sigma})}=-2\big(\boldsymbol{\eta}_{q(\boldsymbol{\Sigma})}\big)_1-2\quad\text{and}\quad\boldsymbol{\Lambda}_{q(\boldsymbol{\Sigma})}=-2\text{vec}^{-1}\Big(\boldsymbol{D}_q^{+T}\big(\boldsymbol{\eta}_{q(\boldsymbol{\Sigma})}\big)_2\Big).$$

Alternatively, $q^*(\boldsymbol{\Sigma})$ is the Inverse-Wishart$(\kappa_{q(\boldsymbol{\Sigma})},\boldsymbol{\Lambda}_{q(\boldsymbol{\Sigma})})$ density function, as defined by (16), where

$$\kappa_{q(\boldsymbol{\Sigma})}=\xi_{q(\boldsymbol{\Sigma})}-q+1.$$

S.3.5.2 Conversion of $q^*(\sigma^2)$ to Common Parameter Form

The common parameter form of $q^*(\sigma^2)$ is the Inverse-$\chi^2(\delta_{q(\sigma^2)},\lambda_{q(\sigma^2)})$ density function where

$$\delta_{q(\sigma^2)}=-2\big(\boldsymbol{\eta}_{q(\sigma^2)}\big)_1-2\quad\text{and}\quad\lambda_{q(\sigma^2)}=-2\big(\boldsymbol{\eta}_{q(\sigma^2)}\big)_2.$$

S.3.5.3 Conversion of $q^*(\boldsymbol{\beta},\boldsymbol{u})$ to Common Parameter Form

The common parameter form of $q^*(\boldsymbol{\beta},\boldsymbol{u})$ is the $N\big(\boldsymbol{\mu}_{q(\boldsymbol{\beta},\boldsymbol{u})},\boldsymbol{\Sigma}_{q(\boldsymbol{\beta},\boldsymbol{u})}\big)$ density function where

$$\boldsymbol{\mu}_{q(\boldsymbol{\beta},\boldsymbol{u})}=-\tfrac{1}{2}\left\{\text{vec}^{-1}\Big(\boldsymbol{D}_{p+mq}^{+T}\big(\boldsymbol{\eta}_{q(\boldsymbol{\beta},\boldsymbol{u})}\big)_2\Big)\right\}^{-1}\big(\boldsymbol{\eta}_{q(\boldsymbol{\beta},\boldsymbol{u})}\big)_1$$

and

$$\boldsymbol{\Sigma}_{q(\boldsymbol{\beta},\boldsymbol{u})}=-\tfrac{1}{2}\left\{\text{vec}^{-1}\Big(\boldsymbol{D}_{p+mq}^{+T}\big(\boldsymbol{\eta}_{q(\boldsymbol{\beta},\boldsymbol{u})}\big)_2\Big)\right\}^{-1}.$$

Here $\big(\boldsymbol{\eta}_{q(\boldsymbol{\beta},\boldsymbol{u})}\big)_1$ denotes the first $p+mq$ entries of $\boldsymbol{\eta}_{q(\boldsymbol{\beta},\boldsymbol{u})}$ and $\big(\boldsymbol{\eta}_{q(\boldsymbol{\beta},\boldsymbol{u})}\big)_2$ denotes the remaining entries of the same vector.

S.3.5.4 Conversion of $q^*(\upsilon)$ to Common Parameter Form and Conversion to $q^*(\nu)$

Recall that $q^*(\upsilon)$ is a Moon Rock density function. The Moon Rock distribution is not as established as the other distributions appearing in this subsection. Nevertheless, the web-supplement of McLean & Wand (2019) defines a random variable x to have a Moon Rock distribution with parameters $\alpha > 0$ and $\beta > \alpha$, written $x \sim \text{Moon-Rock}(\alpha, \beta)$, if the density function of x is

$$p(x) = \left[\int_0^\infty \{t^t/\Gamma(t)\}^\alpha \exp(-\beta\, t)\, dt\right]^{-1} \{x^x/\Gamma(x)\}^\alpha \exp(-\beta\, x), \quad x > 0.$$

Therefore, $q^*(\upsilon)$ has a Moon-Rock$(\alpha_{q(\upsilon)}, \beta_{q(\upsilon)})$ density function where

$$\alpha_{q(\upsilon)} = \big(\boldsymbol{\eta}_{q(\upsilon)}\big)_1 \quad \text{and} \quad \beta_{q(\upsilon)} = -\big(\boldsymbol{\eta}_{q(\upsilon)}\big)_2.$$

Explicitly,

$$q^*(\upsilon) = \left[\int_0^\infty \{t^t/\Gamma(t)\}^{\alpha_{q(\upsilon)}} \exp(-\beta_{q(\upsilon)}\, t)\, dt\right]^{-1} \{\upsilon^\upsilon/\Gamma(\upsilon)\}^{\alpha_{q(\upsilon)}} \exp(-\beta_{q(\upsilon)}\, \upsilon), \quad \upsilon > 0.$$

Lastly, we note that since $\nu = 2\upsilon$ the q^*-density function of ν is

$$q^*(\nu) = \tfrac{1}{2}\left[\int_0^\infty \{t^t/\Gamma(t)\}^{\alpha_{q(\upsilon)}} \exp(-\beta_{q(\upsilon)}\, t)\, dt\right]^{-1}$$
$$\times\{(\nu/2)^{\nu/2}/\Gamma(\nu/2)\}^{\alpha_{q(\upsilon)}} \exp\big(-\tfrac{1}{2}\beta_{q(\upsilon)}\nu\big), \quad \nu > 0.$$

Reference

Novomestky, F. (2012). matrixcalc: Collection of functions for matrix calculations. R package. https://CRAN.R-project.org/package=matrixcalc