Supplement for:

Factor Graph Fragmentization of Expectation Propagation

BY WILSON Y. CHEN AND MATT P. WAND

University of Technology Sydney

S.1 Function Definitions

Expectation propagation algorithms that are based on the fragments in Section 3 have straightforward implementation once a few key functions are identified. Many of the functions are simple but long-winded. However, they only have to be implemented once and after that all fragment updates are simple.

The functions can be divided into three types:

1. functions defined via non-analytic integral families.

2. a function defined via inversion of an established function

3. functions that are explicit given function types 1. and 2.

We now give details of each of these types in turn.

S.1.1 Functions Defined via Non-Analytic Integral Families

Two fundamental families of integrals for expectation propagation in linear model contexts are:

$$\begin{aligned}\mathcal{A}(p,q,r,s,t,u) &\equiv \int_{-\infty}^{\infty}\frac{x^p\,\exp(qx-rx^2)\,dx}{(x^2+sx+t)^u},\\ &\qquad p\geq 0,\ q\in\mathbb{R}, r>0, s\in\mathbb{R}, t>\tfrac{1}{4}\,s^2, u>0\\ \text{and}\quad \mathcal{B}(p,q,r,s,t,u) &\equiv \int_{-\infty}^{\infty}\frac{x^p\,\exp\{qx-re^x-se^x/(t+e^x)\}\,dx}{(t+e^x)^u},\\ &\qquad p\geq 0,\ q\in\mathbb{R}, r>0, s\geq 0, t>0, u>0.\end{aligned} \tag{S.1}$$

An additional family of non-analytic functions that we need is:

$$\mathcal{C}_b(p,q,r)\equiv\int_{-\infty}^{\infty}x^p\,\exp\{qx-rx^2-b(x)\}\,dx, \tag{S.2}$$

where $q\in\mathbb{R}$, $r>0$ and $b:\mathbb{R}\to\mathbb{R}$ is any function for which $\mathcal{C}_b(p,q,r)$ exists.

To avoid underflow and overflow working with logarithms and suitably modified integrands is recommended. For example,

$$\mathcal{C}(0,q,r)=e^M\int_{-\infty}^{\infty}\exp\{qx-rx^2-b(x)-M\}\,dx$$

where $M\equiv\sup\{x\in\mathbb{R}:qx-rx^2-b(x)\}$ implies that

$$\log\{\mathcal{C}(0,q,r)\}=M+\log\left[\int_{-\infty}^{\infty}\exp\{qx-rx^2-b(x)-M\}\,dx\right]$$

Sine the last-written integrand has a maximum of 1, its values are not overly large or small.

S.1.2 Function Defined via Inversion

Theorem 1 of Kim & Wand (2016) asserts that the function $\log - \text{digamma}$ is a bijective mapping from $\mathbb{R}_+$ onto $\mathbb{R}_+$. Therefore its inverse

$$(\log - \text{digamma})^{-1} : \mathbb{R}_+ \to \mathbb{R}_+$$

is well-defined. Bounds given in Guo & Qi (2013) imply that

$$\frac{1}{2x} < (\log - \text{digamma})^{-1}(x) < \frac{1}{x} \quad \text{for all } x > 0.$$

Therefore,

$$(\log - \text{digamma})^{-1}(x) \approx \text{geometric mean of } \frac{1}{2x} \text{ and } \frac{1}{x} = \frac{1}{x\sqrt{2}}$$

and provides useful starting values for iterative inversion of $\log - \text{digamma}$.

In addition, care is required for evaluation of $(\log - \text{digamma})(x)$ for large x since direct computation round-off error can lead to an erroneous answer of zero. Software such as the function `logmdigamma()` in the R package statmod (Smyth, 2015) rectifies this problem.

S.1.3 Explicit Functions

The $N(0,1)$ density function and cumulative distribution functions are denoted by

$$\phi(x) \equiv (2\pi)^{-1} e^{-x^2/2} \quad \text{and} \quad \Phi(x) \equiv \int_{-\infty}^{x} \phi(t)\, dt.$$

We also define

$$\zeta(x) \equiv \log\{2\Phi(x)\} \quad \text{so that} \quad \zeta'(x) \equiv \phi(x)/\Phi(x).$$

Stable computation of ζ' is available from, for example, the function `zeta()` in the R package sn (Azzalini, 2017).

The functions G^{N} and G^{IG1} defined in Kim & Wand (2016) are also needed here. The function G^{IG2} from Kim & Wand (2016) requires generalization to handle Half-t priors on standard deviation parameters with arbitrary degrees of freedom. The generalization is denoted by G^{IG3}. Each of G^{N}, G^{IG1} and G^{IG3} depend on the functions defined in Appendices S.1.1 and S.1.2 but otherwise ther are simple, albeit long-winded, functions with multiple vector arguments. Their definitions are given in Section A.4 of Kim & Wand (2016) and are repeated here for convenience. We also define the explicit functions H_{probit}, H_{logistic} and H_{Poisson}. First set:

$$\alpha\left(k, \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}\right) = \mathcal{A}\left(k, a_1, -a_2, \frac{-2c_2}{c_1}, \frac{c_3 - 2b_2}{c_1}, \frac{c_1 - 2b_1 - 2}{2}\right)$$

and

$$\beta\left(k, \ell, v, w, \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}\right) = \mathcal{B}\left(k, \frac{\ell + c_1 - 1}{2} - a_1, \frac{c_1 c_3 - c_2^2}{2c_1} - a_2, -b_2\left(\frac{c_2}{c_1} + \frac{b_1}{2b_2}\right)^2, v, w\right).$$

Next, define

$$g(\ell, v, w, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = (\log - \text{digamma})^{-1}\left(\log\left\{\frac{\beta(0, \ell+1, v, w, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})}{\beta(0, \ell-1, v, w, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})}\right\} - \frac{\beta(1, \ell-1, v, w, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})}{\beta(0, \ell-1, v, w, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})}\right).$$

Now we are ready to give the expressions for G^{N}, G^{IG1}, G^{IG2}, and G^{IG3}:

$$G^N(\boldsymbol{a},\boldsymbol{b};\boldsymbol{c}) = \left[\frac{\alpha(2,\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})}{\alpha(0,\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})} - \left\{\frac{\alpha(1,\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})}{\alpha(0,\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})}\right\}^2\right]^{-1}\left[\begin{array}{c}\alpha(1,\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})/\alpha(0,\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})\\ -1/2\end{array}\right] - \boldsymbol{a},$$

$$G^{\mathrm{IG1}}\left(\boldsymbol{a},\left[\begin{array}{c}b_1\\ b_2\end{array}\right];\left[\begin{array}{c}c_1\\ c_2\\ c_3\end{array}\right]\right) = \left[\begin{array}{c}-1-g(0,-2b_2/c_1,\frac{1}{2},\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})\\ \dfrac{-g(0,-2b_2/c_1,\frac{1}{2},\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})\,\beta(0,-1,-2b_2/c_1,\frac{1}{2},\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})}{\beta(0,1,-2b_2/c_1,\frac{1}{2},\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})}\end{array}\right] - \boldsymbol{a}$$

and

$$G^{\mathrm{IG3}}\left(\boldsymbol{a},\left[\begin{array}{c}b_1\\ b_2\end{array}\right];k,\ell\right) =$$

$$\left[\begin{array}{c}-1-g\left(k-2,-b_2/\ell,\ell-k/2-b_1,\boldsymbol{a},\left[\begin{array}{c}0\\ b_2\end{array}\right],\left[\begin{array}{c}2\\ 0\\ 0\end{array}\right]\right)\\ \\ \dfrac{\left\{\begin{array}{l}-g\left(k-2,-b_2/\ell,\ell-k/2-b_1,\boldsymbol{a},\left[\begin{array}{c}0\\ b_2\end{array}\right],\left[\begin{array}{c}2\\ 0\\ 0\end{array}\right]\right)\\ \times\beta\left(0,k-3,-b_2/\ell,\ell-k/2-b_1,\boldsymbol{a},\left[\begin{array}{c}0\\ b_2\end{array}\right],\left[\begin{array}{c}2\\ 0\\ 0\end{array}\right]\right)\end{array}\right\}}{\beta\left(0,k-1,-b_2/\ell,\ell-k/2-b_1,\boldsymbol{a},\left[\begin{array}{c}0\\ b_2\end{array}\right],\left[\begin{array}{c}2\\ 0\\ 0\end{array}\right]\right)}\end{array}\right] - \boldsymbol{a}.$$

This definition is a generalization of the function G^{IG2} given in Kim & Wand (2016, 2017) and is such that

$$G^{\mathrm{IG3}}\left(\boldsymbol{a},\left[\begin{array}{c}b_1\\ b_2\end{array}\right];k,1\right) = G^{\mathrm{IG2}}\left(\boldsymbol{a},\left[\begin{array}{c}b_1\\ b_2\end{array}\right];k\right).$$

Put

$$H_b\left(\left[\begin{array}{c}a_1\\ a_2\end{array}\right];y\right) = \left[\begin{array}{c}\dfrac{\mathcal{C}_b(1,a_1+y,-a_2)/\mathcal{C}_b(0,a_1+y,-a_2)}{\dfrac{\mathcal{C}_b(2,a_1+y,-a_2)}{\mathcal{C}_b(0,a_1+y,-a_2)} - \dfrac{\mathcal{C}_b(1,a_1+y,-a_2)^2}{\mathcal{C}_b(0,a_1+y,-a_2)^2}}\\ \\ \dfrac{-1/2}{\dfrac{\mathcal{C}_b(2,a_1+y,-a_2)}{\mathcal{C}_b(0,a_1+y,-a_2)} - \dfrac{\mathcal{C}_b(1,a_1+y,-a_2)^2}{\mathcal{C}_b(0,a_1+y,-a_2)^2}}\end{array}\right] - \left[\begin{array}{c}a_1\\ a_2\end{array}\right]$$

for any $a_1 \in \mathbb{R}$, $a_2 < 0$ and $y \in \mathbb{R}$ and then let

$$\begin{aligned} &H_{\text{logistic}}\left(\begin{bmatrix}a_1\\ a_2\end{bmatrix};y\right) \equiv H_b\left(\begin{bmatrix}a_1\\ a_2\end{bmatrix};y\right) \quad \text{for} \quad b(x) = \log(1+e^x)\\ \text{and}\quad &H_{\text{Poisson}}\left(\begin{bmatrix}a_1\\ a_2\end{bmatrix};y\right) \equiv H_b\left(\begin{bmatrix}a_1\\ a_2\end{bmatrix};y\right) \quad \text{for} \quad b(x) = e^x.\end{aligned} \qquad \text{(S.3)}$$

Lastly,

$$H_{\text{probit}}\left(\left[\begin{array}{c}a_1\\ a_2\end{array}\right];y\right) \equiv \frac{1}{1-2a_2-\zeta'(r)\{r+\zeta'(r)\}}\begin{bmatrix}a_1(1-2a_2)\\ +(2y-1)\zeta'(r)\sqrt{2a_2(2a_2-1)}\\ \\ a_2(1-2a_2)\end{bmatrix} - \begin{bmatrix}a_1\\ \\ \\ a_2\end{bmatrix}$$

where $r \equiv (2y-1)a_1/\sqrt{2a_2(2a_2-1)}$.

S.2 Derivations

We now provide derivations of each of algorithms given in Section 3. Definitions used in the derivations are:

$$A_{\rm N}(\boldsymbol{\eta}) \equiv -\tfrac{1}{4}(\eta_1^2/\eta_2) - \tfrac{1}{2}\log(-2\eta_2) \quad \text{and} \quad A_{{\rm I}\chi^2}(\boldsymbol{\eta}) \equiv (\eta_1+1)\log(-\eta_2) + \log\Gamma(-\eta_1-1)$$

for the log-partition functions of the Normal and Inverse Chi-Squared families respectively. In a similar vein, $\text{proj}_{\rm N}[\,\cdot\,]$ denotes Kullback-Leibler projection onto the (possibly Multivariate) Normal family of density functions and $\text{proj}_{{\rm I}\chi^2}[\,\cdot\,]$ denotes Kullback-Leibler projection onto the Inverse Chi-Squared family of density functions. We use Z to denote the normalizing factor of the function inside a Kullback-Leibler projection operator.

S.2.1 Derivation of Algorithm 1

Plugging into (9) we get

$$m_{p(\boldsymbol{\theta})\,\rightarrow\,\boldsymbol{\theta}}(\boldsymbol{\theta}) = \frac{\text{proj}_{\rm N}\Big[m_{\theta\,\rightarrow\,p(\boldsymbol{\theta})}(\boldsymbol{\theta})\,p(\boldsymbol{\theta})/Z\Big]}{m_{\theta\,\rightarrow\,p(\boldsymbol{\theta})}(\boldsymbol{\theta})} = \frac{m_{\theta\,\rightarrow\,p(\boldsymbol{\theta})}(\boldsymbol{\theta})\,p(\boldsymbol{\theta})}{m_{\theta\,\rightarrow\,p(\boldsymbol{\theta})}(\boldsymbol{\theta})} = p(\theta).$$

The second equality follows from the fact that $m_{\theta\,\rightarrow\,p(\boldsymbol{\theta})}(\boldsymbol{\theta})$ is a Multivariate Normal density function, which is a consequence of (11). Since

$$p(\boldsymbol{\theta}) \propto \exp\big\{-\tfrac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\mu}_{\boldsymbol{\theta}})^T\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta}-\boldsymbol{\mu}_{\boldsymbol{\theta}})\big\} \propto \exp\left\{\begin{bmatrix}\boldsymbol{\theta}\\ \text{vec}(\boldsymbol{\theta}\boldsymbol{\theta}^T)\end{bmatrix}^T\begin{bmatrix}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\theta}}\\ -\tfrac{1}{2}\text{vec}(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1})\end{bmatrix}\right\}$$

we have

$$m_{p(\boldsymbol{\theta})\,\rightarrow\,\boldsymbol{\theta}}(\boldsymbol{\theta}) = \exp\left\{\begin{bmatrix}\boldsymbol{\theta}\\ \text{vec}(\boldsymbol{\theta}\boldsymbol{\theta}^T)\end{bmatrix}^T\boldsymbol{\eta}_{p(\boldsymbol{\theta})\,\rightarrow\,\boldsymbol{\theta}}\right\}$$

where

$$\boldsymbol{\eta}_{p(\boldsymbol{\theta})\,\rightarrow\,\boldsymbol{\theta}} = \begin{bmatrix}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\theta}}\\ -\tfrac{1}{2}\text{vec}(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1})\end{bmatrix}.$$

S.2.2 Derivation of Algorithm 2

Using arguments similar to those given in Section S.2.1 for the Gaussian prior fragment lead to

$$\begin{aligned} m_{p(\boldsymbol{\Theta})\,\rightarrow\,\boldsymbol{\Theta}}(\boldsymbol{\Theta}) &\propto p(\boldsymbol{\Theta}) \propto |\boldsymbol{\Theta}|^{-(\kappa_{\boldsymbol{\Theta}}+d^{\Theta}+1)/2}\exp\{-\tfrac{1}{2}\text{tr}(\boldsymbol{\Lambda}_{\boldsymbol{\Theta}}\boldsymbol{\Theta}^{-1})\}\\ &= \exp\left\{\begin{bmatrix}\log|\boldsymbol{\Theta}|\\ \text{vec}(\boldsymbol{\Theta}^{-1})\end{bmatrix}^T\begin{bmatrix}-\tfrac{1}{2}(\kappa_{\boldsymbol{\Theta}}+d^{\Theta}+1)\\ -\tfrac{1}{2}\text{vec}(\boldsymbol{\Lambda}_{\boldsymbol{\Theta}}^{-1})\end{bmatrix}\right\}. \end{aligned}$$

Hence

$$m_{p(\boldsymbol{\Theta})\,\rightarrow\,\boldsymbol{\Theta}}(\boldsymbol{\Theta}) = \exp\left\{\begin{bmatrix}\log|\boldsymbol{\Theta}|\\ \text{vec}(\boldsymbol{\Theta}^{-1})\end{bmatrix}^T\boldsymbol{\eta}_{p(\boldsymbol{\Theta})\,\rightarrow\,\boldsymbol{\Theta}}\right\}$$

where

$$\boldsymbol{\eta}_{p(\boldsymbol{\Theta})\,\rightarrow\,\boldsymbol{\Theta}} = \begin{bmatrix}-\tfrac{1}{2}(\kappa_{\boldsymbol{\Theta}}+d^{\Theta}+1)\\ -\tfrac{1}{2}\text{vec}(\boldsymbol{\Lambda}_{\boldsymbol{\Theta}}^{-1})\end{bmatrix}.$$

S.2.3 Derivation of Algorithm 3

The message from $p(\sigma^2|a)$ to σ^2 is

$$m_{p(\sigma^2|a)\,\rightarrow\,\sigma^2}(\sigma^2) \longleftarrow \frac{\text{proj}_{\text{I}\chi^2}\left[m_{\sigma^2\,\rightarrow\,p(\sigma^2|a)}(\sigma^2)\int_0^\infty p(\sigma^2|a)\,m_{a\,\rightarrow\,p(\sigma^2|a)}(a)\,da/Z\right]}{m_{\sigma^2\,\rightarrow\,p(\sigma^2|a)}(\sigma^2)}. \quad \text{(S.4)}$$

Assumption (13) implies that

$$m_{a\,\rightarrow\,p(\sigma^2|a)}(a) = \exp\left\{\begin{bmatrix}\log(a)\\ 1/a\end{bmatrix}^T \boldsymbol{\eta}_{a\,\rightarrow\,p(\sigma^2|a)}\right\} = a^{\eta_1^{\clubsuit}}\exp(\eta_2^{\clubsuit}/a)$$

and

$$m_{\sigma^2\,\rightarrow\,p(\sigma^2|a)}(\sigma^2) = \exp\left\{\begin{bmatrix}\log(\sigma^2)\\ 1/\sigma^2\end{bmatrix}^T \boldsymbol{\eta}_{\sigma^2\,\rightarrow\,p(\sigma^2|a)}\right\} = (\sigma^2)^{\eta_1^{\spadesuit}}\exp(\eta_2^{\spadesuit}/\sigma^2)$$

where

$$\boldsymbol{\eta}^{\clubsuit} \equiv \boldsymbol{\eta}_{a\,\rightarrow\,p(\sigma^2|a)} \quad \text{and} \quad \boldsymbol{\eta}^{\spadesuit} \equiv \boldsymbol{\eta}_{\sigma^2\,\rightarrow\,p(\sigma^2|a)}.$$

As a function of σ^2, the integral in (S.4) is

$$\begin{aligned}
&\int_0^\infty \frac{(\nu/a)^{\nu/2}}{\Gamma(\nu/2)}(\sigma^2)^{-(\nu/2)-1}\exp\{-\nu/(\sigma^2 a)\}\,a^{\eta_1^{\clubsuit}}\exp(\eta_2^{\clubsuit}/a)\,da\\
&\qquad\propto (\sigma^2)^{-(\nu/2)-1}\int_0^\infty a^{-\{-\eta_1^{\clubsuit}+(\nu/2)-1\}-1}\exp[-\{(\nu/\sigma^2)-\eta_2^{\clubsuit}\}/a]\,da\\
&\qquad\propto (\sigma^2)^{-(\nu/2)-1}\{(\nu/\sigma^2)-\eta_2^{\clubsuit}\}^{\eta_1^{\clubsuit}-(\nu/2)+1}.
\end{aligned}$$

Therefore, the density function inside the $\text{proj}_{\text{I}\chi^2}[\,\cdot\,]$ is proportional to

$$(\sigma^2)^{\eta_1^{\spadesuit}-(\nu/2)-1}\{(\nu/\sigma^2)-\eta_2^{\clubsuit}\}^{\eta_1^{\clubsuit}-(\nu/2)+1}\exp(\eta_2^{\spadesuit}/\sigma^2)$$

and the numerator of (S.4) is proportional to

$$\exp\left\{\begin{bmatrix}\log(\sigma^2)\\ 1/\sigma^2\end{bmatrix}^T \boldsymbol{\eta}_{\text{numer}}\right\}$$

where

$$\boldsymbol{\eta}_{\text{numer}} \equiv (\nabla A_{\text{I}\chi^2})^{-1}\left(\begin{bmatrix}\dfrac{\int_0^\infty \log(\sigma^2)(\sigma^2)^{\eta_1^{\spadesuit}-(\nu/2)-1}\{(\nu/\sigma^2)-\eta_2^{\clubsuit}\}^{\eta_1^{\clubsuit}-(\nu/2)+1}\exp(\eta_2^{\spadesuit}/\sigma^2)\,d\sigma^2}{\int_0^\infty (\sigma^2)^{\eta_1^{\spadesuit}-(\nu/2)-1}\{(\nu/\sigma^2)-\eta_2^{\clubsuit}\}^{\eta_1^{\clubsuit}-(\nu/2)+1}\exp(\eta_2^{\spadesuit}/\sigma^2)\,d\sigma^2}\\[2ex] \dfrac{\int_0^\infty (1/\sigma^2)(\sigma^2)^{\eta_1^{\spadesuit}-(\nu/2)-1}\{(\nu/\sigma^2)-\eta_2^{\clubsuit}\}^{\eta_1^{\clubsuit}-(\nu/2)+1}\exp(\eta_2^{\spadesuit}/\sigma^2)\,d\sigma^2}{\int_0^\infty (\sigma^2)^{\eta_1^{\spadesuit}-(\nu/2)-1}\{(\nu/\sigma^2)-\eta_2^{\clubsuit}\}^{\eta_1^{\clubsuit}-(\nu/2)+1}\exp(\eta_2^{\spadesuit}/\sigma^2)\,d\sigma^2}\end{bmatrix}\right).$$

Steps analogous to those given in Appendix A.5.4 of Kim & Wand (2016) can then be used to derive the $\boldsymbol{\eta}_{p(\sigma^2|a)\,\rightarrow\,\sigma^2}$ update. However, note that Algorithm 3 supports general Half-t(A,ν) prior distributions and Kim & Wand (2016) only deal with the $\nu = 1$ (Half-Cauchy) special case. Hence, the function G^{IG2} in Kim & Wand (2016) has to be generalized to the G^{IG3} function.

Also, note that the $\boldsymbol{\eta}_{p(\sigma^2|a)\,\rightarrow\,\sigma^2}$ update has $\overset{\varepsilon}{\longleftarrow}$ rather than $\longleftarrow$ to allow for the damping adjustment defined by (10). The same adjustment applies to the remainder of the derivations given in Section S.2.

The message from $p(\sigma^2|a)$ to a has a similar form and arguments analogous to those just given for the message from $p(\sigma^2|a)$ to σ^2 lead to the update for $\boldsymbol{\eta}_{p(\sigma^2|a)\,\rightarrow\,a}$ given in Algorithm 3.

S.2.4 Derivation of Algorithm 4

Algorithm 4 is the $d^{\boldsymbol{\theta}}=1$ special case of Algorithm 5 and therefore its derivation is covered by the general $d^{\boldsymbol{\theta}}$ case given next in Section S.2.5.

S.2.5 Derivation of Algorithm 5

Algorithm 5 depends on Result 2 below, which extends Theorem 1 of Kim & Wand (2017) to multivariate linear combination derived variables.

Let

$$p_{N_d}(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma})=(2\pi)^{-d/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\{-\tfrac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\}$$

denote the density function of a d-variate $\mathrm{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$ random vector. Then we have:

Result 2. *For all $d\times d'$ matrices $\boldsymbol{L}$ such $\boldsymbol{L}^T\boldsymbol{\Sigma}\boldsymbol{L}$ is positive definite and $d'\times 1$ vectors $\boldsymbol{v}$:*

$$\int_{\mathbb{R}^d} p_{N_d}(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma})\,\delta\big(\boldsymbol{v}-\boldsymbol{L}^T\boldsymbol{x}\big)\,d\boldsymbol{x}=p_{N_{d'}}(\boldsymbol{v};\boldsymbol{L}^T\boldsymbol{\mu},\boldsymbol{L}^T\boldsymbol{\Sigma}\boldsymbol{L}). \tag{S.5}$$

Derivation of Result 2.

Via the change of variable $\boldsymbol{z}=\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{x}-\boldsymbol{\mu})$, the left-hand side of (S.5) is

$$\int_{\mathbb{R}^d} p_{N_d}(\boldsymbol{z};\boldsymbol{0},\boldsymbol{I})\,\delta\big(\boldsymbol{v}-\boldsymbol{L}^T\boldsymbol{\mu}-\boldsymbol{L}^T\boldsymbol{\Sigma}^{1/2}\boldsymbol{z}\big)\,d\boldsymbol{z}=\int_{\mathbb{R}^d} p_{N_d}(\boldsymbol{z};\boldsymbol{0},\boldsymbol{I})\,\delta\big(\boldsymbol{v}^\dagger-(\boldsymbol{L}^\dagger)^T\boldsymbol{z}\big)\,d\boldsymbol{z}$$

where

$$\boldsymbol{v}^\dagger\equiv\boldsymbol{v}-\boldsymbol{L}^T\boldsymbol{\mu}\quad\text{and}\quad\boldsymbol{L}^\dagger\equiv\boldsymbol{\Sigma}^{1/2}\boldsymbol{L}.$$

Next note that

$$\begin{aligned}\int_{\mathbb{R}^d} p_{N_d}(\boldsymbol{x};\boldsymbol{0},\boldsymbol{I})\,\delta\big(\boldsymbol{v}^\dagger-(\boldsymbol{L}^\dagger)^T\boldsymbol{z}\big)\,d\boldsymbol{z}&=\int_{\mathbb{R}^d}\lim_{\varepsilon\to 0} p_{N_d}(\boldsymbol{z};\boldsymbol{0},\boldsymbol{I})\,p_{N_{d'}}\big(\boldsymbol{v}^\dagger-(\boldsymbol{L}^\dagger)^T\boldsymbol{z};\boldsymbol{0},\varepsilon\,\boldsymbol{I}\big)\,d\boldsymbol{z}\\&=(2\pi)^{-d/2}\int_{\mathbb{R}^d}\lim_{\varepsilon\to 0}\big[(2\pi\varepsilon)^{-d'/2}\exp\{r(\boldsymbol{z},\boldsymbol{v}^\dagger,\boldsymbol{L}^\dagger,\varepsilon)\}\big]\,d\boldsymbol{z}\end{aligned}$$

where

$$\begin{aligned}r(\boldsymbol{z},\boldsymbol{v}^\dagger,\boldsymbol{L}^\dagger,\varepsilon)&=-\tfrac{1}{2}\boldsymbol{z}^T\boldsymbol{z}-\tfrac{1}{2\varepsilon}\{\boldsymbol{v}^\dagger-(\boldsymbol{L}^\dagger)^T\boldsymbol{z}\}^T\{\boldsymbol{v}^\dagger-(\boldsymbol{L}^\dagger)^T\boldsymbol{z}\}\\&=-\tfrac{1}{2}(\boldsymbol{z}-\boldsymbol{m})^T\{\boldsymbol{I}+\tfrac{1}{\varepsilon}\boldsymbol{L}^\dagger(\boldsymbol{L}^\dagger)^T\}(\boldsymbol{z}-\boldsymbol{m})-\tfrac{1}{2\varepsilon}(\boldsymbol{v}^\dagger)^T\boldsymbol{v}^\dagger\\&\quad+\tfrac{1}{2\varepsilon^2}(\boldsymbol{L}^\dagger\boldsymbol{v}^\dagger)^T\{\boldsymbol{I}+\tfrac{1}{\varepsilon}\boldsymbol{L}^\dagger(\boldsymbol{L}^\dagger)^T\}^{-1}\boldsymbol{L}^\dagger\boldsymbol{v}^\dagger\end{aligned}$$

with $\boldsymbol{m}=\boldsymbol{m}(\boldsymbol{v}^\dagger,\boldsymbol{L}^\dagger,\varepsilon)\equiv\frac{1}{\varepsilon}(\boldsymbol{I}+\frac{1}{\varepsilon}\boldsymbol{L}\boldsymbol{L}^T)^{-1}(\boldsymbol{L}^\dagger)^T\boldsymbol{v}^\dagger$. We then have

$$\begin{aligned}\textstyle\int_{\mathbb{R}^d} p_{N_d}(\boldsymbol{z};\boldsymbol{0},\boldsymbol{I})\,\delta\big(\boldsymbol{v}^\dagger-(\boldsymbol{L}^\dagger)^T\boldsymbol{z}\big)\,d\boldsymbol{z}&=\lim_{\varepsilon\to 0}\Big[(2\pi\varepsilon)^{-d'/2}|\boldsymbol{I}+\tfrac{1}{\varepsilon}\boldsymbol{L}^\dagger(\boldsymbol{L}^\dagger)^T|^{-1/2}\\&\quad\times\exp\big\{\tfrac{1}{2\varepsilon^2}(\boldsymbol{L}^\dagger\boldsymbol{v}^\dagger)^T\{\boldsymbol{I}+\tfrac{1}{\varepsilon}\boldsymbol{L}^\dagger(\boldsymbol{L}^\dagger)^T\}^{-1}\boldsymbol{L}^\dagger\boldsymbol{v}^\dagger-\tfrac{1}{2\varepsilon}(\boldsymbol{v}^\dagger)^T\boldsymbol{v}^\dagger\big\}\Big]\end{aligned} \tag{S.6}$$

where we have used the fact

$$\int_{\mathbb{R}^d}(2\pi)^{-d/2}|\boldsymbol{I}+\tfrac{1}{\varepsilon}\boldsymbol{L}^\dagger(\boldsymbol{L}^\dagger)^T|^{1/2}\exp\Big\{-\tfrac{1}{2}(\boldsymbol{z}-\boldsymbol{m})^T\big\{\boldsymbol{I}+\tfrac{1}{\varepsilon}\boldsymbol{L}^\dagger(\boldsymbol{L}^\dagger)^T\big\}(\boldsymbol{z}-\boldsymbol{m})\Big\}\,d\boldsymbol{z}=1$$

and assumed that the integrand possesses properties that are sufficient to justify interchanging the limit as $\varepsilon\to 0$ and the integral over $\mathbb{R}^d$. Using Theorem 18.1.1 of Harville (2008), the determinant in (S.6) is

$$|\boldsymbol{I}+\tfrac{1}{\varepsilon}\boldsymbol{L}^\dagger(\boldsymbol{L}^\dagger)^T|=\varepsilon^{-d'}|\varepsilon\,\boldsymbol{I}+(\boldsymbol{L}^\dagger)^T\boldsymbol{L}^\dagger|=\varepsilon^{-d'}|\varepsilon\,\boldsymbol{I}+\boldsymbol{L}^T\boldsymbol{\Sigma}\boldsymbol{L}|. \tag{S.7}$$

Next, application of Theorem 18.2.8 of Harville (2008) gives

$$\{\boldsymbol{I}+\tfrac{1}{\varepsilon}\boldsymbol{L}^{\dagger}(\boldsymbol{L}^{\dagger})^{T}\}^{-1}=\boldsymbol{I}-\boldsymbol{L}^{\dagger}\big\{\varepsilon\boldsymbol{I}+(\boldsymbol{L}^{\dagger})^{T}\boldsymbol{L}^{\dagger}\big\}^{-1}(\boldsymbol{L}^{\dagger})^{T} \tag{S.8}$$

which implies that

$$\begin{aligned}(\boldsymbol{L}^{\dagger}\boldsymbol{v}^{\dagger})^{T}\{\boldsymbol{I}+\tfrac{1}{\varepsilon}\boldsymbol{L}^{\dagger}(\boldsymbol{L}^{\dagger})^{T}\}^{-1}\boldsymbol{L}^{\dagger}\boldsymbol{v}^{\dagger}&=(\boldsymbol{v}^{\dagger})^{T}(\boldsymbol{L}^{\dagger})^{T}\boldsymbol{L}^{\dagger}\boldsymbol{v}^{\dagger}-(\boldsymbol{L}^{\dagger}\boldsymbol{v}^{\dagger})^{T}\boldsymbol{L}^{\dagger}\big\{\varepsilon\boldsymbol{I}+(\boldsymbol{L}^{\dagger})^{T}\boldsymbol{L}^{\dagger}\big\}^{-1}(\boldsymbol{L}^{\dagger})^{T}\boldsymbol{L}^{\dagger}\boldsymbol{v}^{\dagger}\\&=(\boldsymbol{v}^{\dagger})^{T}\boldsymbol{L}^{T}\boldsymbol{\Sigma}\boldsymbol{L}\,\boldsymbol{v}^{\dagger}-(\boldsymbol{v}^{\dagger})^{T}\boldsymbol{L}^{T}\boldsymbol{\Sigma}\boldsymbol{L}\big(\varepsilon\boldsymbol{I}+\boldsymbol{L}^{T}\boldsymbol{\Sigma}\boldsymbol{L}\big)^{-1}\boldsymbol{L}^{T}\boldsymbol{\Sigma}\boldsymbol{L}\boldsymbol{v}^{\dagger}.\end{aligned}$$

The exponent in the expression on the right-hand side of (S.6) is then

$$\begin{aligned}&\tfrac{1}{2\varepsilon^{2}}\Big\{(\boldsymbol{v}^{\dagger})^{T}\boldsymbol{L}^{T}\boldsymbol{\Sigma}\boldsymbol{L}\,\boldsymbol{v}^{\dagger}-(\boldsymbol{v}^{\dagger})^{T}\boldsymbol{L}^{T}\boldsymbol{\Sigma}\boldsymbol{L}\big(\varepsilon\boldsymbol{I}+\boldsymbol{L}^{T}\boldsymbol{\Sigma}\boldsymbol{L}\big)^{-1}\boldsymbol{L}^{T}\boldsymbol{\Sigma}\boldsymbol{L}\boldsymbol{v}^{\dagger}-\varepsilon(\boldsymbol{v}^{\dagger})^{T}\boldsymbol{v}^{\dagger}\Big\}\\&\quad=-\tfrac{1}{2}(\boldsymbol{v}^{\dagger})^{T}\big(\varepsilon\boldsymbol{I}+\boldsymbol{L}^{T}\boldsymbol{\Sigma}\boldsymbol{L}\big)^{-1}\boldsymbol{v}^{\dagger}=-\tfrac{1}{2}(\boldsymbol{v}-\boldsymbol{L}^{T}\boldsymbol{\mu})^{T}\big(\varepsilon\boldsymbol{I}+\boldsymbol{L}^{T}\boldsymbol{\Sigma}\boldsymbol{L}\big)^{-1}(\boldsymbol{v}-\boldsymbol{L}^{T}\boldsymbol{\mu}).\end{aligned}$$

Substitution of this result and (S.7) into (S.6) then gives

$$\begin{aligned}&\int_{\mathbb{R}^{d}}p_{N_d}(\boldsymbol{z};\boldsymbol{0},\boldsymbol{I})\,\delta\big(\boldsymbol{v}^{\dagger}-(\boldsymbol{L}^{\dagger})^{T}\boldsymbol{z}\big)\,d\boldsymbol{z}\\&\quad=\lim_{\varepsilon\to 0}\big[(2\pi)^{-d'/2}|\varepsilon\boldsymbol{I}+\boldsymbol{L}^{T}\boldsymbol{\Sigma}\boldsymbol{L}|^{-1/2}\exp\{-\tfrac{1}{2}(\boldsymbol{v}-\boldsymbol{L}^{T}\boldsymbol{\mu})^{-1}(\varepsilon\boldsymbol{I}+\boldsymbol{L}^{T}\boldsymbol{\Sigma}\boldsymbol{L})^{-1}(\boldsymbol{v}-\boldsymbol{L}^{T}\boldsymbol{\mu})\}\big].\\&\quad=p_{N_{d'}}(\boldsymbol{v};\boldsymbol{L}^{T}\boldsymbol{\mu},\boldsymbol{L}^{T}\boldsymbol{\Sigma}\boldsymbol{L})\end{aligned}$$

and the result follows.

∎

From (9) we have

$$\boldsymbol{\eta}_{\delta(\boldsymbol{\alpha}-\boldsymbol{A}^{T}\boldsymbol{\theta})\to\boldsymbol{\alpha}}(\boldsymbol{\alpha})\longleftarrow\frac{\text{proj}_{\text{N}}\Big[m_{\boldsymbol{\alpha}\to\delta(\boldsymbol{\alpha}-\boldsymbol{A}^{T}\boldsymbol{\theta})}(\boldsymbol{\alpha})\int_{\mathbb{R}^{d^{\boldsymbol{\theta}}}}\delta(\boldsymbol{\alpha}-\boldsymbol{A}^{T}\boldsymbol{\theta})m_{\boldsymbol{\theta}\to\delta(\boldsymbol{\alpha}-\boldsymbol{A}^{T}\boldsymbol{\theta})}(\boldsymbol{\theta})\,d\boldsymbol{\theta}/Z\Big]}{m_{\boldsymbol{\alpha}\to\delta(\boldsymbol{\alpha}-\boldsymbol{A}^{T}\boldsymbol{\theta})}(\boldsymbol{\alpha})} \tag{S.9}$$

where $d^{\boldsymbol{\theta}}$ is the dimension of $\boldsymbol{\theta}$. It follows from assumption (15) and (8) that

$$m_{\boldsymbol{\alpha}\to\delta(\boldsymbol{\alpha}-\boldsymbol{A}^{T}\boldsymbol{\theta})}(\boldsymbol{\alpha})=\exp\left\{\begin{bmatrix}\boldsymbol{\alpha}\\\text{vec}(\boldsymbol{\alpha}\boldsymbol{\alpha}^{T})\end{bmatrix}^{T}\boldsymbol{\eta}_{\boldsymbol{\alpha}\to\delta(\boldsymbol{\alpha}-\boldsymbol{A}^{T}\boldsymbol{\theta})}\right\} \tag{S.10}$$

and

$$\begin{aligned}m_{\boldsymbol{\theta}\to\delta(\boldsymbol{\alpha}-\boldsymbol{A}^{T}\boldsymbol{\theta})}(\boldsymbol{\theta})&=\exp\left\{\begin{bmatrix}\boldsymbol{\theta}\\\text{vec}(\boldsymbol{\theta}\boldsymbol{\theta}^{T})\end{bmatrix}^{T}\boldsymbol{\eta}_{\boldsymbol{\theta}\to\delta(\boldsymbol{\alpha}-\boldsymbol{A}^{T}\boldsymbol{\theta})}\right\}\\&\propto(2\pi)^{-d^{\boldsymbol{\theta}}/2}|\boldsymbol{\Sigma}_{\odot}|^{-1/2}\exp\{-\tfrac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\mu}_{\odot})^{T}\boldsymbol{\Sigma}_{\odot}^{-1}(\boldsymbol{\theta}-\boldsymbol{\mu}_{\odot})\}\end{aligned}$$

where

$$\boldsymbol{\mu}_{\odot}\equiv-\tfrac{1}{2}\Big\{\text{vec}^{-1}\big(\boldsymbol{\eta}_{\boldsymbol{\theta}\to\delta(\boldsymbol{\alpha}-\boldsymbol{A}^{T}\boldsymbol{\theta})}\big)_{2}\Big\}^{-1}\Big(\boldsymbol{\eta}_{\boldsymbol{\theta}\to\delta(\boldsymbol{\alpha}-\boldsymbol{A}^{T}\boldsymbol{\theta})}\Big)_{1}$$

and

$$\boldsymbol{\Sigma}_{\odot}\equiv-\tfrac{1}{2}\Big\{\text{vec}^{-1}\big(\boldsymbol{\eta}_{\boldsymbol{\theta}\to\delta(\boldsymbol{\alpha}-\boldsymbol{A}^{T}\boldsymbol{\theta})}\big)_{2}\Big\}^{-1}$$

are the common parameters matching $\boldsymbol{\eta}_{\boldsymbol{\theta}\to\delta(\boldsymbol{\alpha}-\boldsymbol{A}^{T}\boldsymbol{\theta})}$. From Result 2 in Appendix S.2.5,

$$\begin{aligned}&\int_{\mathbb{R}^{d^{\boldsymbol{\theta}}}}\delta(\boldsymbol{\alpha}-\boldsymbol{A}^{T}\boldsymbol{\theta})(2\pi)^{-d^{\boldsymbol{\theta}}/2}|\boldsymbol{\Sigma}_{\odot}|^{-1/2}\exp\{-\tfrac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\mu}_{\odot})^{T}\boldsymbol{\Sigma}_{\odot}^{-1}(\boldsymbol{\theta}-\boldsymbol{\mu}_{\odot})\}\,d\boldsymbol{\theta}\\&\quad=|2\pi\,\boldsymbol{A}^{T}\boldsymbol{\Sigma}_{\odot}\boldsymbol{A}|^{-1/2}\exp\Big\{-\tfrac{1}{2}(\boldsymbol{\alpha}-\boldsymbol{A}^{T}\boldsymbol{\mu}_{\odot})^{T}(\boldsymbol{A}^{T}\boldsymbol{\Sigma}_{\odot}\boldsymbol{A})^{-1}(\boldsymbol{\alpha}-\boldsymbol{A}^{T}\boldsymbol{\mu}_{\odot})\Big\}.\end{aligned}$$

Substitution into (S.9) then leads to

$$\boldsymbol{\eta}_{\delta(\boldsymbol{\alpha}-\boldsymbol{A}^T\boldsymbol{\theta})\to\boldsymbol{\alpha}}(\boldsymbol{\alpha}) =$$

$$\frac{\text{proj}_{\text{N}}\left[\exp\left\{\begin{bmatrix}\boldsymbol{\alpha}\\ \text{vec}(\boldsymbol{\alpha}\boldsymbol{\alpha}^T)\end{bmatrix}\boldsymbol{\eta}_{\boldsymbol{\alpha}\to\delta(\boldsymbol{\alpha}-\boldsymbol{A}^T\boldsymbol{\theta})}\right\}\exp\left\{\begin{bmatrix}\boldsymbol{\alpha}\\ \text{vec}(\boldsymbol{\alpha}\boldsymbol{\alpha}^T)\end{bmatrix}^T\begin{bmatrix}(\boldsymbol{A}^T\boldsymbol{\Sigma}_{\odot}\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{\mu}_{\odot}\\ -\frac{1}{2}\text{vec}\big((\boldsymbol{A}^T\boldsymbol{\Sigma}_{\odot}\boldsymbol{A})^{-1}\big)\end{bmatrix}\right\}\Big/Z\right]}{\exp\left\{\begin{bmatrix}\boldsymbol{\alpha}\\ \text{vec}(\boldsymbol{\alpha}\boldsymbol{\alpha}^T)\end{bmatrix}\boldsymbol{\eta}_{\boldsymbol{\alpha}\to\delta(\boldsymbol{\alpha}-\boldsymbol{A}^T\boldsymbol{\theta})}\right\}}$$

$$= \exp\left\{\begin{bmatrix}\boldsymbol{\alpha}\\ \text{vec}(\boldsymbol{\alpha}\boldsymbol{\alpha}^T)\end{bmatrix}^T\begin{bmatrix}(\boldsymbol{A}^T\boldsymbol{\Sigma}_{\odot}\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{\mu}_{\odot}\\ -\frac{1}{2}\text{vec}\big((\boldsymbol{A}^T\boldsymbol{\Sigma}_{\odot}\boldsymbol{A})^{-1}\big)\end{bmatrix}\right\}.$$

Therefore, setting $\boldsymbol{\Omega} \longleftarrow -2\boldsymbol{\Sigma}_{\odot}\boldsymbol{A}$, we get

$$\boldsymbol{\eta}_{\delta(\boldsymbol{\alpha}-\boldsymbol{A}^T\boldsymbol{\theta})\to\boldsymbol{\alpha}} \longleftarrow \begin{bmatrix}(\boldsymbol{\Omega}^T\boldsymbol{A})^{-1}\boldsymbol{\Omega}^T\Big(\boldsymbol{\eta}_{\boldsymbol{\theta}\to\delta(\boldsymbol{\alpha}-\boldsymbol{A}^T\boldsymbol{\theta})}\Big)_1\\ \text{vec}\big((\boldsymbol{\Omega}^T\boldsymbol{A})^{-1}\big)\end{bmatrix}$$

which are the first two updates of Algorithm 5.

For the third update, note that (9) gives

$$m_{\delta(\boldsymbol{\alpha}-\boldsymbol{A}^T\boldsymbol{\theta})\to\boldsymbol{\theta}}(\boldsymbol{\theta})$$

$$\longleftarrow \frac{\text{proj}_{\text{N}}\Big[m_{\boldsymbol{\theta}\to\delta(\boldsymbol{\alpha}-\boldsymbol{A}^T\boldsymbol{\theta})}(\boldsymbol{\theta})\int_{\mathbb{R}^{d\boldsymbol{\alpha}}}\delta(\boldsymbol{\alpha}-\boldsymbol{A}^T\boldsymbol{\theta})m_{\boldsymbol{\alpha}\to\delta(\boldsymbol{\alpha}-\boldsymbol{A}^T\boldsymbol{\theta})}(\boldsymbol{\alpha})\,d\boldsymbol{\alpha}/Z\Big]}{m_{\boldsymbol{\theta}\to\delta(\boldsymbol{\alpha}-\boldsymbol{A}^T\boldsymbol{\theta})}(\boldsymbol{\theta})}$$

$$= \frac{\text{proj}_{\text{N}}\Big[m_{\boldsymbol{\theta}\to\delta(\alpha-\boldsymbol{a}^T\boldsymbol{\theta})}(\boldsymbol{\theta})\,m_{\boldsymbol{\alpha}\to\delta(\boldsymbol{\alpha}-\boldsymbol{A}^T\boldsymbol{\theta})}(\boldsymbol{A}^T\boldsymbol{\theta})/Z\Big]}{m_{\boldsymbol{\theta}\to\delta(\alpha-\boldsymbol{a}^T\boldsymbol{\theta})}(\boldsymbol{\theta})}$$

$$= \frac{\text{proj}_{\text{N}}\left[\exp\left\{\begin{bmatrix}\boldsymbol{\theta}\\ \text{vec}(\boldsymbol{\theta}\boldsymbol{\theta}^T)\end{bmatrix}^T\boldsymbol{\eta}_{\boldsymbol{\theta}\to\delta(\boldsymbol{\alpha}-\boldsymbol{A}^T\boldsymbol{\theta})}\right\}\exp\left\{\begin{bmatrix}\boldsymbol{A}^T\boldsymbol{\theta}\\ \text{vec}\big((\boldsymbol{A}^T\boldsymbol{\theta})(\boldsymbol{A}^T\boldsymbol{\theta})^T\big)\end{bmatrix}^T\boldsymbol{\eta}_{\boldsymbol{\alpha}\to\delta(\boldsymbol{\alpha}-\boldsymbol{A}^T\boldsymbol{\theta})}\right\}\Big/Z\right]}{\exp\left\{\begin{bmatrix}\boldsymbol{\theta}\\ \text{vec}(\boldsymbol{\theta}\boldsymbol{\theta}^T)\end{bmatrix}^T\boldsymbol{\eta}_{\boldsymbol{\theta}\to\delta(\boldsymbol{\alpha}-\boldsymbol{A}^T\boldsymbol{\theta})}\right\}}$$

$$= \exp\left\{\begin{bmatrix}\boldsymbol{\theta}\\ \text{vec}(\boldsymbol{\theta}\boldsymbol{\theta}^T)\end{bmatrix}^T\begin{bmatrix}\boldsymbol{A}\Big(\boldsymbol{\eta}_{\alpha\to\delta(\alpha-\boldsymbol{a}^T\boldsymbol{\theta})}\Big)_1\\ (\boldsymbol{A}\otimes\boldsymbol{A})\Big(\boldsymbol{\eta}_{\alpha\to\delta(\alpha-\boldsymbol{a}^T\boldsymbol{\theta})}\Big)_2\end{bmatrix}\right\}.$$

The last step uses the identity $\text{vec}(\boldsymbol{BCD}) = (\boldsymbol{D}^T\otimes\boldsymbol{B})\text{vec}(\boldsymbol{C})$ for any compatible matrices $\boldsymbol{B}$, $\boldsymbol{C}$ and $\boldsymbol{D}$. The third update follows immediately.

S.2.6 Derivation of Algorithm 6

The message from $p(y|\alpha,\sigma^2)$ to α is, from (9),

$$m_{p(y|\alpha,\sigma^2)\to\alpha}(\alpha) = \frac{\text{proj}_{\text{N}}\big[m_{\alpha\to p(y|\alpha,\sigma^2)}(\alpha)\,\int_0^\infty p(y|\alpha,\sigma^2)\,m_{\sigma^2\to p(y|\alpha,\sigma^2)}(\sigma^2)\,d\sigma^2\big]}{m_{\alpha\to p(y|\alpha,\sigma^2)}(\alpha)}. \tag{S.11}$$

It follows from (8) and (18) that

$$m_{\sigma^2\to p(y|\alpha,\sigma^2)}(\sigma^2) = \exp\left\{\begin{bmatrix}\log(\sigma^2)\\ 1/\sigma^2\end{bmatrix}^T\boldsymbol{\eta}_{\sigma^2\to p(y|\alpha,\sigma^2)}\right\} = (\sigma^2)^{\eta_1^{\boxplus}}\exp(\eta_2^{\boxplus}/\sigma^2).$$

where $\eta^{\boxplus} \equiv \boldsymbol{\eta}_{\sigma^2 \,\rightarrow\, p(y|\alpha,\sigma^2)}$. Hence

$$\int_0^\infty p(y|\alpha,\sigma^2)\, m_{\sigma^2 \,\rightarrow\, p(y|\alpha,\sigma^2)}(\sigma^2)\, d\sigma^2 \;=\; (2\pi)^{-1/2} \int_0^\infty (\sigma^2)^{\eta_1^{\boxplus}-1/2} \exp[\{\eta_2^{\boxplus} - \tfrac{1}{2}(y-\alpha)^2\}/\sigma^2)]\, d\sigma^2$$

$$\propto \left\{\tfrac{1}{2}(y-\alpha)^2 - \eta_2^{\boxplus}\right\}^{\eta_1^{\boxplus}+\frac{1}{2}}.$$

Therefore, letting $\eta^{\text{✡}} \equiv \boldsymbol{\eta}_{\alpha \,\rightarrow\, p(y|\alpha,\sigma^2)}$, the numerator of the right-hand side of (S.11) is

$$\text{proj}_{\text{N}}\left[\frac{\exp(\alpha\eta_1^{\text{✡}} + \alpha^2\eta_1^{\text{✡}})}{\left\{(y-\alpha)^2 - 2\eta_2^{\boxplus}\right\}^{-\eta_1^{\boxplus}-\frac{1}{2}}}\right]$$

$$= \exp\left\{\begin{bmatrix}\alpha\\ \alpha^2\end{bmatrix}^T (\nabla A_{\text{N}})^{-1}\left(\begin{bmatrix}\dfrac{\int_{-\infty}^{\infty} \alpha\exp(\alpha\eta_1^{\text{✡}}+\alpha^2\eta_2^{\text{✡}})\big/\left\{(y-\alpha)^2-2\eta_2^{\boxplus}\right\}^{-\eta_1^{\boxplus}-\frac{1}{2}}\, d\alpha}{\int_{-\infty}^{\infty} \exp(\alpha\eta_1^{\text{✡}}+\alpha^2\eta_2^{\text{✡}})\big/\left\{(y-\alpha)^2-2\eta_2^{\boxplus}\right\}^{-\eta_1^{\boxplus}-\frac{1}{2}}\, d\alpha}\\[2ex] \dfrac{\int_{-\infty}^{\infty} \alpha^2\exp(\alpha\eta_2^{\text{✡}}+\alpha^2\eta_2^{\text{✡}})\big/\left\{(y-\alpha)^2-2\eta_2^{\boxplus}\right\}^{-\eta_1^{\boxplus}-\frac{1}{2}}\, d\alpha}{\int_{-\infty}^{\infty} \exp(\alpha\eta_1^{\text{✡}}+\alpha^2\eta_2^{\text{✡}})\big/\left\{(y-\alpha)^2-2\eta_2^{\boxplus}\right\}^{-\eta_1^{\boxplus}-\frac{1}{2}}\, d\alpha}\end{bmatrix}\right)\right\}$$

$$= \exp\left\{\begin{bmatrix}\alpha\\ \alpha^2\end{bmatrix}^T (\nabla A_N)^{-1}\left(\begin{bmatrix}\dfrac{\mathcal{A}(1,\eta_1^{\text{✡}},-\eta_1^{\text{✡}},-2y,y^2-2\eta_2^{\boxplus},-\eta_1^{\boxplus}-\frac{1}{2})}{\mathcal{A}(0,\eta_1^{\text{✡}},-\eta_2^{\text{✡}},-2y,y^2-2\eta_2^{\boxplus},-\eta_1^{\boxplus}-\frac{1}{2})}\\ \dfrac{\mathcal{A}(2,\eta_1^{\text{✡}},-\eta_1^{\text{✡}},-2y,y^2-2\eta_2^{\boxplus},-\eta_1^{\boxplus}-\frac{1}{2})}{\mathcal{A}(0,\eta_1^{\text{✡}},-\eta_2^{\text{✡}},-2y,y^2-2\eta_2^{\boxplus},-\eta_1^{\boxplus}-\frac{1}{2})}\end{bmatrix}\right)\right\}.$$

Hence

$$\boldsymbol{\eta}_{p(y|\alpha,\sigma^2)\,\rightarrow\,\alpha} \longleftarrow (\nabla A_N)^{-1}\left(\begin{bmatrix}\dfrac{\mathcal{A}(1,\eta_1^{\text{✡}},-\eta_1^{\text{✡}},-2y,y^2-2\eta_2^{\boxplus},-\eta_1^{\boxplus}-\frac{1}{2})}{\mathcal{A}(0,\eta_1^{\text{✡}},-\eta_2^{\text{✡}},-2y,y^2-2\eta_2^{\boxplus},-\eta_1^{\boxplus}-\frac{1}{2})}\\ \dfrac{\mathcal{A}(2,\eta_1^{\text{✡}},-\eta_1^{\text{✡}},-2y,y^2-2\eta_2^{\boxplus},-\eta_1^{\boxplus}-\frac{1}{2})}{\mathcal{A}(0,\eta_1^{\text{✡}},-\eta_2^{\text{✡}},-2y,y^2-2\eta_2^{\boxplus},-\eta_1^{\boxplus}-\frac{1}{2})}\end{bmatrix}\right) - \boldsymbol{\eta}_{\alpha\,\rightarrow\,p(y|\alpha,\sigma^2)}$$

and the first update in Algorithm 6 follows from the definition of G^{N} given in Section S.1.3.

The message $p(y|\alpha,\sigma^2)$ to σ^2 is

$$m_{p(by|\alpha,\sigma^2)\,\rightarrow\,\sigma^2}(\sigma^2) = \frac{\text{proj}_{\text{I}\chi^2}\left[m_{\sigma^2\,\rightarrow\,p(y|\alpha,\sigma^2)}(\sigma^2)\;\int_{-\infty}^{\infty} p(y|\alpha,\sigma^2)\, m_{\alpha\,\rightarrow\,p(y|\alpha,\sigma^2)}(\alpha)\, d\alpha\right]}{m_{\sigma^2\,\rightarrow\,p(y|\alpha,\sigma^2)}(\sigma^2)}. \tag{S.12}$$

It follows from (8) and (18) that

$$m_{\alpha\,\rightarrow\,p(y|\alpha,\sigma^2)}(\alpha) = \exp\left\{\begin{bmatrix}\alpha\\ \alpha^2\end{bmatrix}^T \boldsymbol{\eta}_{\alpha\,\rightarrow\,p(y|\alpha,\sigma^2)}\right\} = \exp\left(\alpha\,\eta_1^{\oplus} + \alpha^2\eta_2^{\oplus}\right)$$

where $\boldsymbol{\eta}^{\oplus} \equiv \boldsymbol{\eta}_{\alpha\,\rightarrow\,p(y|\alpha,\sigma^2)}$. The integral in (S.12) is

$$\int_{-\infty}^{\infty} (2\pi\sigma^2)^{-1/2} \exp\left\{\frac{-(y-\alpha)^2}{2\sigma^2}\right\} \exp\left(\alpha\,\eta_1^{\oplus} + \alpha^2\eta_2^{\oplus}\right) d\alpha$$

$$\propto \int_{-\infty}^{\infty} (2\pi\sigma^2)^{-1/2} \exp\left\{\frac{-(y-\alpha)^2}{2\sigma^2}\right\} \{2\pi(\sigma^{\oplus})^2\}^{-1/2} \exp\left\{\frac{-(y-\mu^{\oplus})^2}{2(\sigma^{\oplus})^2}\right\}\, d\alpha$$

where

$$\begin{bmatrix}\mu^{\oplus}\\ (\sigma^{\oplus})^2\end{bmatrix} \equiv \begin{bmatrix}-\eta_1^{\oplus}/(2\eta_2^{\oplus})\\ -1/(2\eta_2^{\oplus})\end{bmatrix}$$

is the common parameter vector corresponding to $\boldsymbol{\eta}^{\oplus}$. Using (A.2) of Wand and Jones (1993), the last-written integral is

$$\{2\pi(\sigma^2+(\sigma^{\oplus})^2\}^{-1/2}\exp\left\{\frac{-(y-\mu^{\oplus})^2}{2\{\sigma^2+(\sigma^{\oplus})^2\}}\right\}.$$

Therefore, the function inside the $\text{proj}_{\text{I}\chi^2}[\,\cdot\,]$ in (S.12) is

$$(\sigma^2)^{\eta_1^{\boxplus}}\exp(\eta_2^{\boxplus}/\sigma^2)[2\pi\{\sigma^2-1/(2\eta_2^{\oplus})\}]^{-1/2}\exp\left[\frac{-\{y+\eta_1^{\oplus}/(2\eta_2^{\oplus})\}^2}{2\{\sigma^2-1/(2\eta_2^{\oplus})\}}\right].$$

Plugging into (S.12) we then have

$$m_{p(by|\alpha,\sigma^2)\,\rightarrow\,\sigma^2}(\sigma^2)=\exp\left\{\begin{bmatrix}\log(\sigma^2)\\1/\sigma^2\end{bmatrix}^T\boldsymbol{\eta}_{p(y|\alpha,\sigma^2)\,\rightarrow\,\sigma^2}\right\}$$

where

$$\boldsymbol{\eta}_{p(y|\alpha,\sigma^2)\,\rightarrow\,\sigma^2}\longleftarrow(\nabla A_{\text{I}\chi^2})^{-1}\left(\begin{bmatrix}\dfrac{\int_0^\infty\log(\sigma^2)(\sigma^2)^{\eta_1^{\boxplus}}\{\sigma^2-1/(2\eta_2^{\oplus})\}^{-1/2}\exp\left[\frac{\eta_2^{\boxplus}}{\sigma^2}-\frac{\{y+\eta_1^{\oplus}/(2\eta_2^{\oplus})\}^2}{2\{\sigma^2-1/(2\eta_2^{\oplus})\}}\right]d\sigma^2}{\int_0^\infty(\sigma^2)^{\eta_1^{\boxplus}}\{\sigma^2-1/(2\eta_2^{\oplus})\}^{-1/2}\exp\left[\frac{\eta_2^{\boxplus}}{\sigma^2}-\frac{\{y+\eta_1^{\oplus}/(2\eta_2^{\oplus})\}^2}{2\{\sigma^2-1/(2\eta_2^{\oplus})\}}\right]d\sigma^2}\\[2ex]\dfrac{\int_0^\infty(1/\sigma^2)(\sigma^2)^{\eta_1^{\boxplus}}\{\sigma^2-1/(2\eta_2^{\oplus})\}^{-1/2}\exp\left[\frac{\eta_2^{\boxplus}}{\sigma^2}-\frac{\{y+\eta_1^{\oplus}/(2\eta_2^{\oplus})\}^2}{2\{\sigma^2-1/(2\eta_2^{\oplus})\}}\right]d\sigma^2}{\int_0^\infty(\sigma^2)^{\eta_1^{\boxplus}}\{\sigma^2-1/(2\eta_2^{\oplus})\}^{-1/2}\exp\left[\frac{\eta_2^{\boxplus}}{\sigma^2}-\frac{\{y+\eta_1^{\oplus}/(2\eta_2^{\oplus})\}^2}{2\{\sigma^2-1/(2\eta_2^{\oplus})\}}\right]d\sigma^2}\end{bmatrix}\right)-\boldsymbol{\eta}^{\boxplus}.$$

The change of variable $\sigma^2=e^{-x}$ and some simple algebra then leads to

$$\boldsymbol{\eta}_{p(y|\alpha,\sigma^2)\,\rightarrow\,\sigma^2}\longleftarrow(\nabla A_{\text{I}\chi^2})^{-1}\left(\begin{bmatrix}\dfrac{-\mathcal{B}(1,-\eta_1^{\boxplus},-\eta_2^{\boxplus},-2\eta_2^{\oplus}\{y+\eta_1^{\oplus}/(2\eta_2^{\oplus})\}^2,-2\eta_2^{\oplus},\frac{1}{2})}{\mathcal{B}(0,-\eta_1^{\boxplus},-\eta_2^{\boxplus},-2\eta_2^{\oplus}\{y+\eta_1^{\oplus}/(2\eta_2^{\oplus})\}^2,-2\eta_2^{\oplus},\frac{1}{2})}\\[2ex]\dfrac{\mathcal{B}(0,1-\eta_1^{\boxplus},-\eta_2^{\boxplus},-2\eta_2^{\oplus}\{y+\eta_1^{\oplus}/(2\eta_2^{\oplus})\}^2,-2\eta_2^{\oplus},\frac{1}{2})}{\mathcal{B}(0,-\eta_1^{\boxplus},-\eta_2^{\boxplus},-2\eta_2^{\oplus}\{y+\eta_1^{\oplus}/(2\eta_2^{\oplus})\}^2,-2\eta_2^{\oplus},\frac{1}{2})}\end{bmatrix}\right)-\boldsymbol{\eta}_{\sigma^2\,\rightarrow\,p(y|\alpha,\sigma^2)}.$$

The second update in Algorithm 6 follows from the definition of G^{IG1} given in Section S.1.3.

S.2.7 Derivation of Algorithm 7

The only factor to stochastic node message for the logistic fragment is, from (9):

$$m_{p(y|\alpha)\,\rightarrow\,\alpha}(\alpha)=\frac{\text{proj}[m_{\alpha\,\rightarrow\,p(y|\alpha)}(\alpha)\,p(y|\alpha)/Z]}{m_{\alpha\,\rightarrow\,p(y|\alpha)}(\alpha)}$$

with projection onto an appropriate exponential family. Assumption (19) and conjugacy considerations implies projection onto the Univariate Normal family. Setting

$$\boldsymbol{\eta}^{\#}\equiv\boldsymbol{\eta}_{\alpha\,\rightarrow\,p(y|\alpha)},$$

we then have

$$m_{p(y|\alpha)\,\rightarrow\,\alpha}(\alpha)=\frac{\text{proj}_{\text{N}}[\exp(\eta_1^{\#}\alpha+\eta_2^{\#}\alpha^2)\,\exp\{y\alpha-\log(1+e^{\alpha})\}]}{\exp(\eta_1^{\#}\alpha+\eta_2^{\#}\alpha^2)}.$$

The numerator is

$$\begin{aligned}&\mathrm{proj}_{\mathrm{N}}\big[\exp\{(\eta_1^{\#}+y)\alpha+\eta_2^{\#}\alpha^2-\log(1+e^{\alpha})\}\big]\\&\qquad\propto\exp\left\{\left[\begin{array}{c}\alpha\\\alpha^2\end{array}\right]^T(\nabla A_N)^{-1}\left(\left[\begin{array}{c}\frac{\int_{-\infty}^{\infty}\alpha\exp\{(\eta_1^{\#}+y)\alpha+\eta_2^{\#}\alpha^2-\log(1+e^{\alpha})\}\,d\alpha}{\int_{-\infty}^{\infty}\exp\{(\eta_1^{\#}+y)\alpha+\eta_2^{\#}\alpha^2-\log(1+e^{\alpha})\}\,d\alpha}\\\frac{\int_{-\infty}^{\infty}\alpha^2\exp\{(\eta_1^{\#}+y)\alpha+\eta_2^{\#}\alpha^2-\log(1+e^{\alpha})\}\,d\alpha}{\int_{-\infty}^{\infty}\exp\{(\eta_1^{\#}+y)\alpha+\eta_2^{\#}\alpha^2-\log(1+e^{\alpha})\}\,d\alpha}\end{array}\right]\right)\right\}\\&\qquad\propto\exp\left\{\left[\begin{array}{c}\alpha\\\alpha^2\end{array}\right]^T(\nabla A_N)^{-1}\left(\left[\begin{array}{c}\frac{\mathcal{C}(1,\eta_1^{\#}+y,\eta_1^{\#})}{\mathcal{C}(0,\eta_1^{\#}+y,\eta_1^{\#})}\\\frac{\mathcal{C}(2,\eta_1^{\#}+y,\eta_1^{\#})}{\mathcal{C}(0,\eta_1^{\#}+y,\eta_1^{\#})}\end{array}\right]\right)\right\}.\end{aligned}$$

Hence

$$\boldsymbol{\eta}_{p(y|\alpha)\,\rightarrow\,\alpha}\longleftarrow(\nabla A_N)^{-1}\left(\left[\begin{array}{c}\frac{\mathcal{C}(1,\eta_1^{\#}+y,\eta_1^{\#})}{\mathcal{C}(0,\eta_1^{\#}+y,\eta_1^{\#})}\\\frac{\mathcal{C}(2,\eta_1^{\#}+y,\eta_1^{\#})}{\mathcal{C}(0,\eta_1^{\#}+y,\eta_1^{\#})}\end{array}\right]\right)-\boldsymbol{\eta}_{\alpha\,\rightarrow\,p(y|\alpha)}$$

and the update in Algorithm 7 follows from the definition of H_{logistic} given in Section S.1.3.

S.2.8 Derivation of Algorithm 8

Via arguments analogous to those given in Section S.2.7, the factor to stochastic node natural parameter update is

$$\boldsymbol{\eta}_{p(y|\alpha)\,\rightarrow\,\alpha}\longleftarrow(\nabla A_N)^{-1}\left(\left[\begin{array}{c}\frac{\int_{-\infty}^{\infty}\alpha\Phi((2y-1)\alpha)\exp\{\eta_1^{\#}\alpha+\eta_2^{\#}\alpha^2\}\,d\alpha}{\int_{-\infty}^{\infty}\Phi((2y-1)\alpha)\exp\{\eta_1^{\#}\alpha+\eta_2^{\#}\alpha^2\}\,d\alpha}\\\frac{\int_{-\infty}^{\infty}\alpha^2\Phi((2y-1)\alpha)\exp\{\eta_1^{\#}\alpha+\eta_2^{\#}\alpha^2\}\,d\alpha}{\int_{-\infty}^{\infty}\Phi((2y-1)\alpha)\exp\{\eta_1^{\#}\alpha+\eta_2^{\#}\alpha^2\}\,d\alpha}\end{array}\right]\right)-\boldsymbol{\eta}_{\alpha\,\rightarrow\,p(y|\alpha)}$$

where, as before, $\boldsymbol{\eta}^{\#}\equiv\boldsymbol{\eta}_{\alpha\,\rightarrow\,p(y|\alpha)}$. The integral results

$$\begin{aligned}\int_{-\infty}^{\infty}\Phi(a+bx)\phi(x)\,dx&=\Phi\left(\frac{a}{\sqrt{b^2+1}}\right),\\\int_{-\infty}^{\infty}x\,\Phi(a+bx)\phi(x)\,dx&=\frac{b}{\sqrt{b^2+1}}\,\phi\left(\frac{a}{\sqrt{b^2+1}}\right)\\\text{and}\quad\int_{-\infty}^{\infty}x^2\,\Phi(a+bx)\phi(x)\,dx&=\Phi\left(\frac{a}{\sqrt{b^2+1}}\right)-\frac{ab^2}{\sqrt{(b^2+1)^3}}\,\phi\left(\frac{a}{\sqrt{b^2+1}}\right),\end{aligned}\tag{S.13}$$

and standard algebraic manipulations lead to

$$\boldsymbol{\eta}_{p(y|\alpha)\,\rightarrow\,\alpha}\longleftarrow H_{\text{probit}}\left(\boldsymbol{\eta}_{\alpha\,\rightarrow\,p(y|\alpha)};y\right)$$

where H_{probit} is defined in Section S.1.3.

S.2.9 Derivation of Algorithm 9

The Poisson likelihood fragment derivation is very similar to that given in Section S.2.7. The only change is that

$$\exp\{y\alpha-\log(1+e^{\alpha})\}\quad\text{is replaced by}\quad\exp(y\alpha-e^{\alpha})$$

which leads to the H_{Poisson} function appearing in the factor to stochastic node update rather than the H_{logistic} function. Section S.1.3 provides the definition of H_{Poisson}.