Supplement for:
A Variational inference framework for inverse problems

BY L. MAESTRINI[1], R.G. AYKROYD[2] AND M.P. WAND[3]

[1]*Australian National University,* [2]*University of Leeds and* [3]*University of Technology Sydney*

S.1. Algorithm derivations

This section provides full justification of Algorithms 1 and 2.

S.1.1. Derivation of Algorithm 1

The optimal approximating densities satisfy expression (9). The full conditional density functions of each hidden node in the directed acyclic graph in Figure 3 allow to derive expressions for the optimal approximating densities. First,

$$\begin{aligned} p\left(\boldsymbol{x}|\text{rest}\right) &\propto p\left(\boldsymbol{y}|\boldsymbol{x},\sigma_\varepsilon^2\right) p\left(\boldsymbol{x}|\boldsymbol{b},\sigma_x^2\right) \\ &\propto \exp\left[-\frac{1}{2}\left\{\boldsymbol{x}^T\boldsymbol{L}^T\left(\frac{1}{\sigma_\varepsilon^2}\boldsymbol{K}^T\boldsymbol{K}+\frac{1}{\sigma_x^2}\,\text{diag}\left(\boldsymbol{b}\right)\right)\boldsymbol{L}\boldsymbol{x}-\frac{2}{\sigma_\varepsilon^2}\boldsymbol{x}^T\boldsymbol{K}^T\boldsymbol{y}\right\}\right], \end{aligned}$$

that is,

$$\boldsymbol{x}|\text{rest} \sim N\left(\left(\frac{1}{\sigma_\varepsilon^2}\boldsymbol{K}^T\boldsymbol{K}+\frac{1}{\sigma_x^2}\boldsymbol{L}^T\,\text{diag}\left(\boldsymbol{b}\right)\boldsymbol{L}\right)^{-1}\frac{1}{\sigma_\varepsilon^2}\boldsymbol{K}^T\boldsymbol{y},\left(\frac{1}{\sigma_\varepsilon^2}\boldsymbol{K}^T\boldsymbol{K}+\frac{1}{\sigma_x^2}\boldsymbol{L}^T\,\text{diag}\left(b\right)\boldsymbol{L}\right)^{-1}\right).$$

Hence $q^*(\boldsymbol{x})$ is $N\big(\boldsymbol{\mu}_{q(\boldsymbol{x})},\boldsymbol{\Sigma}_{q(\boldsymbol{x})}\big)$ with

$$\boldsymbol{\mu}_{q(\boldsymbol{x})} \equiv \mu_{q(1/\sigma_\varepsilon^2)}\boldsymbol{\Sigma}_{q(\boldsymbol{x})}\boldsymbol{K}^T\boldsymbol{y} \;\text{ and }\; \boldsymbol{\Sigma}_{q(\boldsymbol{x})} \equiv \left(\mu_{q(1/\sigma_\varepsilon^2)}\boldsymbol{K}^T\boldsymbol{K}+\mu_{q(1/\sigma_x^2)}\boldsymbol{L}^T\,\text{diag}(\mu_{q(\boldsymbol{b})})\boldsymbol{L}\right)^{-1}.$$

Given that

$$p\left(\boldsymbol{b}\right) \propto \prod_{j=1}^{d} b_j^{-2}\exp\left\{-1/(2b_j)\right\}, \tag{S.1}$$

then, for $1 \le j \le d$,

$$p\left(b_j|\text{rest}\right) \propto p\left(\boldsymbol{x}|b_j,\sigma_x^2\right)p\left(b_j\right) \propto b_j^{-3/2}\exp\left[-\frac{1}{2}\left\{\frac{b_j\left(\boldsymbol{L}\boldsymbol{x}\right)_j^2}{\sigma_x^2}+\frac{1}{b_j}\right\}\right]$$

and so

$$b_j|\text{rest} \sim \text{Inverse-Gaussian}\Big(\sqrt{\sigma_x^2/\left(\boldsymbol{L}\boldsymbol{x}\right)_j^2},1\Big).$$

It follows that $q^*\left(b_j\right)$ is Inverse-Gaussian$\big(\mu_{q(b_j)},\lambda_{q(b_j)}\big)$, with

$$\mu_{q(b_j)} \equiv \left[\mu_{q(1/\sigma_x^2)}\left\{\left(\boldsymbol{L}\boldsymbol{\mu}_{q(\boldsymbol{x})}\right)_j^2+\left(\,\text{diagonal}(\boldsymbol{L}\boldsymbol{\Sigma}_{q(\boldsymbol{x})}\boldsymbol{L}^T)\right)_j\right\}\right]^{-1/2} \;\text{ and }\; \lambda_{q(b_j)} \equiv 1.$$

Next,

$$p\left(\sigma_\varepsilon^2|\text{rest}\right) \propto p\left(\boldsymbol{y}|\boldsymbol{x},\sigma_\varepsilon^2\right) p\left(\sigma_\varepsilon^2|a_\varepsilon\right) \propto \left(\sigma_\varepsilon^2\right)^{-\frac{m+1}{2}-1} \exp\left\{-\frac{1}{2\sigma_\varepsilon^2}\left(\frac{1}{a_\varepsilon}+\|\boldsymbol{y}-\boldsymbol{K}\boldsymbol{x}\|^2\right)\right\},$$

that provides

$$\sigma_\varepsilon^2|\text{rest} \sim \text{Inverse-}\chi^2\left(m+1, 1/a_\varepsilon+\|\boldsymbol{y}-\boldsymbol{K}\boldsymbol{x}\|^2\right).$$

Then $q^*\left(\sigma_\varepsilon^2\right)$ is Inverse-$\chi^2\big(\kappa_{q(\sigma_\varepsilon^2)}, \lambda_{q(\sigma_\varepsilon^2)}\big)$ with

$$\kappa_{q(\sigma_\varepsilon^2)} \equiv m+1 \;\text{ and }\; \lambda_{q(\sigma_\varepsilon^2)} \equiv \mu_{q(1/a_\varepsilon)} + \|\boldsymbol{y}-\boldsymbol{K}\boldsymbol{\mu}_{q(\boldsymbol{x})}\|^2 + \text{tr}\left(\boldsymbol{K}^T\boldsymbol{K}\boldsymbol{\Sigma}_{q(\boldsymbol{x})}\right),$$

which provide $\mu_{q(1/\sigma_\varepsilon^2)} \equiv \kappa_{q(\sigma_\varepsilon^2)}/\lambda_{q(\sigma_\varepsilon^2)}$.

Also,

$$p\left(\sigma_x^2|\text{rest}\right) \propto p\left(\boldsymbol{x}|\boldsymbol{b},\sigma_x^2\right) p\left(\sigma_x^2|a_x\right) \propto \left(\sigma_x\right)^{-\frac{d+1}{2}-1} \exp\left\{-\frac{1}{2\sigma_x^2}\left(\frac{1}{a_x}+\boldsymbol{x}^T\boldsymbol{L}^T\,\text{diag}\left(\boldsymbol{b}\right)\boldsymbol{L}\boldsymbol{x}\right)\right\},$$

implying that

$$\sigma_x^2|\text{rest} \sim \text{Inverse-}\chi^2\left(d+1, 1/a_x+\sum_{j=1}^{d} b_j\left(\boldsymbol{L}\boldsymbol{x}\right)_j^2\right).$$

This leads to $q^*\left(\sigma_x^2\right)$ being Inverse-$\chi^2\left(\kappa_{q(\sigma_x^2)}, \lambda_{q(\sigma_x^2)}\right)$ with

$$\kappa_{q(\sigma_x^2)} \equiv d+1 \;\text{ and }\; \lambda_{q(\sigma_x^2)} \equiv \mu_{q(1/a_x)} + \sum_{j=1}^{d} \mu_{q(b_j)}\left\{\left(\boldsymbol{L}\boldsymbol{\mu}_{q(\boldsymbol{x})}\right)_j^2 + \left(\,\text{diagonal}(\boldsymbol{L}\boldsymbol{\Sigma}_{q(\boldsymbol{x})}\boldsymbol{L}^T)\right)_j\right\}.$$

The moment expression $\mu_{q(1/\sigma_x^2)} \equiv \kappa_{q(\sigma_x^2)}/\lambda_{q(\sigma_x^2)}$ follows.

As regards auxiliary variable a_ε,

$$p(a_\varepsilon|\text{rest}) \propto p\left(\sigma_\varepsilon^2|a_\varepsilon\right) p\left(a_\varepsilon\right) \propto \exp\left\{-2\log(\alpha_\varepsilon) - \frac{1}{2\alpha_\varepsilon}\left(\frac{1}{\sigma_\varepsilon^2}+\frac{1}{A_\varepsilon^2}\right)\right\},$$

which implies

$$a_\varepsilon|\text{rest} \sim \text{Inverse-}\chi^2\left(2, 1/\sigma_\varepsilon^2+1/A_\varepsilon^2\right).$$

It follows that $q^*(a_\varepsilon)$ is Inverse-$\chi^2\left(\kappa_{q(a_\varepsilon)}, \lambda_{q(a_\varepsilon)}\right)$ with

$$\kappa_{q(a_\varepsilon)} \equiv 2 \;\text{ and }\; \lambda_{q(a_\varepsilon)} \equiv \mu_{q(1/\sigma_\varepsilon^2)} + 1/A_\varepsilon^2.$$

This gives $\mu_{q(1/a_\varepsilon)} \equiv 2/\lambda_{q(a_x)}$.

Analogously to a_ε, $q^*(a_x)$ is Inverse-$\chi^2\left(\kappa_{q(a_x)}, \lambda_{q(a_x)}\right)$ with

$$\kappa_{q(a_x)} \equiv 2 \;\text{ and }\; \lambda_{q(a_x)} \equiv \mu_{q(1/\sigma_x^2)} + 1/A_x^2.$$

Expression $\mu_{q(1/a_x)} \equiv 2/\lambda_{q(a_x)}$ follows.

S.1.2. Derivation of Algorithm 2

Consider expression (23) for $\log p\left(\boldsymbol{x}|\boldsymbol{b},\sigma_x^2\right)$ as a function of the single components $\boldsymbol{x}$, σ_x^2 and $\boldsymbol{b}$. As a function of $\boldsymbol{x}$, we get

$$\log p\left(\boldsymbol{x}|\boldsymbol{b},\sigma_x^2\right) = \frac{1}{\sigma_x^2}\begin{bmatrix} \boldsymbol{x} \\ \text{vec}\left(\boldsymbol{x}\boldsymbol{x}^T\right) \end{bmatrix}^T \begin{bmatrix} \boldsymbol{0} \\ -\frac{1}{2}\,\text{vec}\left(\boldsymbol{L}^T\,\text{diag}\left(\boldsymbol{b}\right)\boldsymbol{L}\right) \end{bmatrix} + \text{const.}$$

Hence

$$m_{p(\boldsymbol{x}|\,\boldsymbol{b},\,\sigma_x^2)\,\to\,\boldsymbol{x}}\left(\boldsymbol{x}\right)=\exp\left\{\left[\begin{array}{c}\boldsymbol{x}\\ \mathrm{vec}\left(\boldsymbol{x}\boldsymbol{x}^T\right)\end{array}\right]^T\boldsymbol{\eta}_{p(\boldsymbol{x}|\boldsymbol{b},\,\sigma_x^2)\,\to\,\boldsymbol{x}}\right\},$$

which is within the Multivariate Normal family, with

$$\boldsymbol{\eta}_{p(\boldsymbol{x}|\boldsymbol{b},\,\sigma_x^2)\,\to\,\boldsymbol{x}}\equiv E_{\boxtimes}\left(1/\sigma_x^2\right)\left[\begin{array}{c}\mathbf{0}\\ -\frac{1}{2}\,\mathrm{vec}\left(\boldsymbol{L}^T\,\mathrm{diag}\left\{E_{\oplus}\left(\boldsymbol{b}\right)\right\}\boldsymbol{L}\right)\end{array}\right],$$

where $E_{\boxtimes}$ denotes expectation with respect to the normalization of

$$m_{p(\boldsymbol{x}|\,\boldsymbol{b},\,\sigma_x^2)\,\to\,\sigma_x^2}\left(\sigma_x^2\right)m_{\sigma_x^2\,\to\,p(\boldsymbol{x}|\,\boldsymbol{b},\,\sigma_x^2)}\left(\sigma_x^2\right)$$

and $E_{\oplus}$ denotes expectation with respect to the normalization of

$$m_{p(\boldsymbol{x}|\,\boldsymbol{b},\,\sigma_x^2)\,\to\,\boldsymbol{b}}\left(\boldsymbol{b}\right)m_{\boldsymbol{b}\,\to\,p(\boldsymbol{x}|\,\boldsymbol{b},\,\sigma_x^2)}\left(\boldsymbol{b}\right).$$

As a function of σ_x^2,

$$\log p\left(x|\boldsymbol{b},\sigma_x^2\right)=\left[\begin{array}{c}\log\left(\sigma_x^2\right)\\ 1/\sigma_x^2\end{array}\right]^T\left[\begin{array}{c}-d/2\\ -\frac{1}{2}\boldsymbol{x}^T\boldsymbol{L}^T\,\mathrm{diag}\left(\boldsymbol{b}\right)\boldsymbol{L}\boldsymbol{x}\end{array}\right]+\mathrm{const.}$$

It follows that

$$m_{p(\boldsymbol{x}|\,\boldsymbol{b},\,\sigma_x^2)\,\to\,\sigma_x^2}\left(\sigma_x^2\right)=\exp\left\{\left[\begin{array}{c}\log\left(\sigma_x^2\right)\\ 1/\sigma_x^2\end{array}\right]^T\boldsymbol{\eta}_{p(\boldsymbol{x}|\,\boldsymbol{b},\,\sigma_x^2)\,\to\,\sigma_x^2}\right\},$$

which is within the Inverse Chi-Squared family, where

$$\boldsymbol{\eta}_{p(\boldsymbol{x}|\,\boldsymbol{b},\,\sigma_x^2)\,\to\,\sigma_x^2}\equiv\left[\begin{array}{c}-d/2\\ -\frac{1}{2}E_{\otimes}\left[\boldsymbol{x}^T\boldsymbol{L}^T\,\mathrm{diag}\left\{E_{\oplus}\left(\boldsymbol{b}\right)\right\}\boldsymbol{L}\boldsymbol{x}\right]\end{array}\right],$$

with $E_{\otimes}$ denoting expectation with respect to the normalization of

$$m_{p(\boldsymbol{x}|\,\boldsymbol{b},\,\sigma_x^2)\,\to\,\boldsymbol{x}}\left(\boldsymbol{x}\right)m_{\boldsymbol{x}\,\to\,p(\boldsymbol{x}|\,\boldsymbol{b},\,\sigma_x^2)}\left(\boldsymbol{x}\right).$$

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

$$\log p\left(\boldsymbol{x}|\boldsymbol{b},\sigma_x^2\right)=\frac{1}{2}\sum_{j=1}^{d}\left\{\log\left(b_j\right)-\left(\boldsymbol{L}\boldsymbol{x}\right)_j^2\,b_j/\sigma_x^2\right\}+\mathrm{const.}$$

Therefore

$$m_{p(\boldsymbol{x}|\,\boldsymbol{b},\,\sigma_x^2)\,\to\,\boldsymbol{b}}\left(\boldsymbol{b}\right)\propto\prod_{j=1}^{d}b_j^{1/2}\exp\left[-\frac{1}{2}E_{\boxtimes}\left(1/\sigma_x^2\right)E_{\otimes}\left\{\left(\boldsymbol{L}\boldsymbol{x}\right)_j^2\right\}\right].$$

It is simple to show that

$$m_{p(\boldsymbol{b})\,\to\,\boldsymbol{b}}\left(\boldsymbol{b}\right)\propto p\left(\boldsymbol{b}\right)\propto\prod_{j=1}^{d}b_j^{-2}\exp\left\{-1/(2b_j)\right\}. \tag{S.2}$$

From (16),

$$m_{\boldsymbol{b}\,\to\,p(\boldsymbol{x}|\,\boldsymbol{b},\,\sigma_x^2)}\left(\boldsymbol{b}\right)=m_{p(\boldsymbol{b})\,\to\,\boldsymbol{b}}\left(\boldsymbol{b}\right)$$

and so

$$m_{p(\boldsymbol{x}|\,\boldsymbol{b},\,\sigma_x^2)\,\to\,\boldsymbol{b}}\,(\boldsymbol{b})\,m_{\boldsymbol{b}\,\to\,p(\boldsymbol{x}|\,\boldsymbol{b},\,\sigma_x^2)}\,(\boldsymbol{b})$$
$$= \prod_{j=1}^{d} b_j^{-3/2} \exp\left\{ \begin{bmatrix} b_j \\ 1/b_j \end{bmatrix}^T \begin{bmatrix} -\frac{1}{2} E_{\boxtimes}\left(1/\sigma_x^2\right) E_{\otimes}\left\{(\boldsymbol{L}\boldsymbol{x})_j^2\right\} \\ -1/2 \end{bmatrix} \right\},$$

which is a product of Inverse Gaussian density functions with natural parameter vector

$$\left[-\frac{1}{2} E_{\boxtimes}\left(1/\sigma_x^2\right) E_{\otimes}\left\{(\boldsymbol{L}\boldsymbol{x})_j^2\right\}, \quad -\frac{1}{2}\right]_{1\le j\le d}^T.$$

It follows from Table S.1 of the supplementary material of Wand (2017) that

$$E_{\odot}\left(\boldsymbol{b}\right) = \left[\left[E_{\boxtimes}\left(1/\sigma_x^2\right) E_{\otimes}\left\{(\boldsymbol{L}\boldsymbol{x})_j^2\right\}\right]^{-1/2}\right]_{1\le j\le d}.$$

Again from Table S.1 of of Wand (2017), since $E_{\boxtimes}$ corresponds to the expectation of an Inverse Chi-Squared distribution with natural parameter vector $\boldsymbol{\eta}_{p(\boldsymbol{x}|\,\boldsymbol{b},\,\sigma_x^2)\,\leftrightarrow\,\sigma_x^2}$,

$$E_{\boxtimes}\left(1/\sigma_x^2\right) = \left\{\left(\boldsymbol{\eta}_{p(\boldsymbol{x}|\,\boldsymbol{b},\,\sigma_x^2)\,\leftrightarrow\,\sigma_x^2}\right)_1 + 1\right\} \Big/ \left\{\left(\boldsymbol{\eta}_{p(\boldsymbol{x}|\,\boldsymbol{b},\,\sigma_x^2)\,\leftrightarrow\,\sigma_x^2}\right)_2\right\}.$$

Next,

$$E_{\otimes}\left\{(\boldsymbol{L}\boldsymbol{x})_j^2\right\} = Var_{\otimes}\left\{(\boldsymbol{L}\boldsymbol{x})_j\right\} + \left[E_{\otimes}\left\{(\boldsymbol{L}\boldsymbol{x})_j\right\}\right]^2.$$

Making use of Table S.1 of Wand (2017), we get

$$\left[Var_{\otimes}\left\{(\boldsymbol{L}\boldsymbol{x})_j\right\}\right]_{1\le j\le d} = -\frac{1}{2}\,\text{diagonal}\left[\boldsymbol{L}\left\{\overset{-1}{\text{vec}}\left(\left(\boldsymbol{\eta}_{p(\boldsymbol{x}|\,\boldsymbol{b},\,\sigma_x^2)\,\leftrightarrow\,\boldsymbol{x}}\right)_2\right)\right\}^{-1}\boldsymbol{L}^T\right]$$
$$\left[\left[E_{\otimes}\left\{(\boldsymbol{L}\boldsymbol{x})_j\right\}\right]^2\right]_{1\le j\le d} = \frac{1}{4}\left[\boldsymbol{L}\left\{\overset{-1}{\text{vec}}\left(\left(\boldsymbol{\eta}_{p(\boldsymbol{x}|\,\boldsymbol{b},\,\sigma_x^2)\,\leftrightarrow\,\boldsymbol{x}}\right)_2\right)\right\}^{-1}\left(\boldsymbol{\eta}_{p(\boldsymbol{x}|\,\boldsymbol{b},\,\sigma_x^2)\,\leftrightarrow\,\boldsymbol{x}}\right)_1\right]^2$$

and

$$E_{\otimes}\left\{(\boldsymbol{L}\boldsymbol{x})_j^2\right\} = \boldsymbol{\omega}_3 \odot \boldsymbol{\omega}_3 + \boldsymbol{\omega}_4,$$

where the expressions for $\boldsymbol{\omega}_3$ and $\boldsymbol{\omega}_4$ are provided in Algorithm 2. Similarly,

$$E_{\otimes}\left[\boldsymbol{x}^T\boldsymbol{L}^T\,\text{diag}\left\{E_{\oplus}\left(\boldsymbol{b}\right)\right\}\boldsymbol{L}\boldsymbol{x}\right] = E_{\oplus}\left(\boldsymbol{b}\right)^T\Big[\left\{\boldsymbol{L}E_{\oplus}\left(\boldsymbol{x}\right)\right\} \odot \left\{\boldsymbol{L}E_{\oplus}\left(\boldsymbol{x}\right)\right\}$$
$$+\,\text{diagonal}\left\{\boldsymbol{L}Cov_{\otimes}\left(\boldsymbol{x}\right)\boldsymbol{L}^T\right\}\Big] = \mu_{q(\boldsymbol{b})}^T\left(\boldsymbol{\omega}_3^2 + \boldsymbol{\omega}_4\right).$$

The expressions for $\boldsymbol{\eta}_{p(\boldsymbol{x}|\boldsymbol{b},\,\sigma_x^2)\,\to\,\boldsymbol{x}}$ and $\boldsymbol{\eta}_{p(\boldsymbol{x}|\,\boldsymbol{b},\,\sigma_x^2)\,\to\,\sigma_x^2}$ outputted by Algorithm 2 follow.

S.1.3. Introducing alternative penalizations

One of the continuous distributions for sparse signal shrinkage of Neville et al. (2014) can be used in lieu of the Laplace penalization. This entails replacement of (S.1) for MFVB and (S.2) for VMP with one of

the following:

$$m_{p(\boldsymbol{b})\to\boldsymbol{b}}(\boldsymbol{b}) \propto \prod_{j=1}^{d} b_j^{-1/2}(1+b_j)^{-1} \qquad \text{(Horseshoe)},$$

$$m_{p(\boldsymbol{b})\to\boldsymbol{b}}(\boldsymbol{b}) \propto \prod_{j=1}^{d} b_j^{\lambda-1}(1+b_j)^{-1-\lambda} \qquad \text{(Negative-Exponential-Gamma)},$$

$$m_{p(\boldsymbol{b})\to\boldsymbol{b}}(\boldsymbol{b}) \propto \prod_{j=1}^{d} b_j^{(\lambda-2)/2} e^{\lambda^2 b_j/4}\mathcal{D}_{-\lambda-2}\left(\lambda\sqrt{b_j}\right) \qquad \text{(Generalized-Double-Pareto)}.$$

The adjustments in Algorithms 1 and 2 that allow for inclusion of such penalizations are summarized in Table 1.

S.2. Implementation of variational message passing

This section demonstrates how to fully implement VMP on the base model, making use of Algorithm 2 and other relevant VMP algorithms for fragments that have already been studied in previous works.

The VMP approach to fit model (2) under restriction (8) takes a response vector $\boldsymbol{y}$ of length m and a $\boldsymbol{K}$ matrix of size $(m \times m)$ as data inputs, and $A_\varepsilon, A_x > 0$ as hyperparameter inputs. At convergence, VMP provides the optimal posterior density function approximations (10)–(15).

S.2.1. Initialization

The message natural parameters arising from the factor graph in Figure 4 have to be initialised at feasible points in the parameter space.

The natural parameter vector $\boldsymbol{\eta}_{p(a_x)\to a_x}$ can be initialized through the Inverse G-Wishart Prior Fragment (Maestrini and Wand, 2021, Algorithm 1) with inputs:

$$G_\Theta = G_{\text{diag}},\ \ \xi_\Theta = 1,\ \ \text{and}\ \ \boldsymbol{\Lambda}_\Theta = A_x^{-2}.$$

The algorithm also provides the graph $G_{p(a_x)\to a_x}$ as an output. An analogous call to the same algorithm provides $\boldsymbol{\eta}_{p(a_\varepsilon)\to a_\varepsilon}$ and $G_{p(a_\varepsilon)\to a_\varepsilon}$. The remaining factor to stochastic node message natural parameters can be initialized, for example, as follows:

$$\boldsymbol{\eta}_{p(\sigma_x^2|a_x)\to\sigma_x^2} \longleftarrow \begin{bmatrix} -3/2 \\ -1 \end{bmatrix},\ \ \boldsymbol{\eta}_{p(\sigma_x^2|a_x)\to a_x} \longleftarrow \begin{bmatrix} -3/2 \\ -1 \end{bmatrix},$$

$$\boldsymbol{\eta}_{p(\boldsymbol{x}|\boldsymbol{b},\sigma_x^2)\to\boldsymbol{x}} \longleftarrow \begin{bmatrix} \mathbf{0}_m \\ -\frac{1}{2}\operatorname{vec}(\boldsymbol{I}_m) \end{bmatrix},\ \ \boldsymbol{\eta}_{p(\boldsymbol{x}|\,\boldsymbol{b},\sigma_x^2)\to\sigma_x^2} \longleftarrow \begin{bmatrix} -3/2 \\ -1 \end{bmatrix},$$

$$\boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{x},\sigma_\varepsilon^2)\to\boldsymbol{x}} \longleftarrow \begin{bmatrix} \mathbf{0}_m \\ -\frac{1}{2}\operatorname{vec}(\boldsymbol{I}_m) \end{bmatrix},\ \ \boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{x},\sigma_\varepsilon^2)\to\sigma_\varepsilon^2} \longleftarrow \begin{bmatrix} -3/2 \\ -1 \end{bmatrix},$$

$$\boldsymbol{\eta}_{p(\sigma_\varepsilon^2|a_\varepsilon)\to\sigma_\varepsilon^2} \longleftarrow \begin{bmatrix} -3/2 \\ -1 \end{bmatrix},\ \ \boldsymbol{\eta}_{p(\sigma_\varepsilon^2|a_\varepsilon)\to a_\varepsilon} \longleftarrow \begin{bmatrix} -3/2 \\ -1 \end{bmatrix}.$$

One way to initialize the stochastic node to factor message natural parameters is the following:

$$\boldsymbol{\eta}_{a_x\to p(\sigma_x^2|a_x)} \longleftarrow \boldsymbol{\eta}_{p(a_x)\to a_x},\ \ \boldsymbol{\eta}_{a_\varepsilon\to p(\sigma_\varepsilon^2|a_\varepsilon)} \longleftarrow \boldsymbol{\eta}_{p(a_\varepsilon)\to a_\varepsilon},$$

$$\boldsymbol{\eta}_{\sigma_\varepsilon^2\to p(\boldsymbol{y}|\boldsymbol{x},\sigma_\varepsilon^2)} \longleftarrow \boldsymbol{\eta}_{p(\sigma_\varepsilon^2|a_\varepsilon)\to\sigma_\varepsilon^2},$$

$$\boldsymbol{\eta}_{\sigma_\varepsilon^2\to p(\sigma_\varepsilon^2|a_\varepsilon)} \longleftarrow \begin{bmatrix} -3/2 \\ -1 \end{bmatrix},\ \ \boldsymbol{\eta}_{\sigma_x^2\to p(\sigma_x^2|a_x)} \longleftarrow \begin{bmatrix} -3/2 \\ -1 \end{bmatrix},$$

$$\boldsymbol{\eta}_{\boldsymbol{x}\to p(\boldsymbol{x}|\,\boldsymbol{b},\sigma_x^2)} \longleftarrow \begin{bmatrix} \boldsymbol{K}^T\boldsymbol{y} \\ -\frac{1}{2}\operatorname{vec}(\boldsymbol{K}^T\boldsymbol{K}) \end{bmatrix},\ \ \boldsymbol{\eta}_{\boldsymbol{x}\to p(\boldsymbol{y}|\boldsymbol{x},\sigma_\varepsilon^2)} \longleftarrow \begin{bmatrix} \mathbf{0}_m \\ -\frac{1}{2}\operatorname{vec}(\boldsymbol{I}_m) \end{bmatrix}.$$

S.2.2. Variational message passing iterations

Once the natural parameter vector initializations are carried out, the stochastic node to factor and factor to stochastic node message parameters are updated in cycle until convergence. A possible way to assess convergence is monitoring the relative difference of parameter estimates from subsequent iterations. The updates for factor to stochastic node message parameters are performed via Algorithm 2 and other VMP schemes proposed in the existing literature.

S.2.2.1. Stochastic node to factor message parameter updates

The stochastic node to factor message updates follow from (16). For the factor graph of Figure (4) these updates are:

$$\begin{gathered}
\boldsymbol{\eta}_{\boldsymbol{x}\to p(\boldsymbol{y}|\boldsymbol{x},\sigma_\varepsilon^2)} \longleftarrow \boldsymbol{\eta}_{p(\boldsymbol{x}|\boldsymbol{b},\sigma_x^2)\to \boldsymbol{x}},\ \ \boldsymbol{\eta}_{\boldsymbol{x}\to p(\boldsymbol{x}|\,\boldsymbol{b},\sigma_x^2)} \longleftarrow \boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{x},\sigma_\varepsilon^2)\to \boldsymbol{x}},\\
\boldsymbol{\eta}_{\sigma_\varepsilon^2\to p(\boldsymbol{y}|\boldsymbol{x},\sigma_\varepsilon^2)} \longleftarrow \boldsymbol{\eta}_{p(\sigma_\varepsilon^2|a_\varepsilon)\to \sigma_\varepsilon^2},\ \ \boldsymbol{\eta}_{\sigma_\varepsilon^2\to p(\sigma_\varepsilon^2|a_\varepsilon)} \longleftarrow \boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{x},\sigma_\varepsilon^2)\to \sigma_\varepsilon^2},\\
G_{\sigma_\varepsilon^2\to p(\sigma_\varepsilon^2|a_\varepsilon)} \longleftarrow G_{\text{full}},\ \ \boldsymbol{\eta}_{\sigma_x^2\to p(\boldsymbol{x}|\,\boldsymbol{b},\sigma_x^2)} \longleftarrow \boldsymbol{\eta}_{p(\sigma_x^2|a_x)\to \sigma_x^2},\\
G_{\sigma_x^2\to p(\sigma_x^2|a_x)} \longleftarrow G_{\text{full}},\ \ \boldsymbol{\eta}_{\sigma_x^2\to p(\sigma_x^2|a_x)} \longleftarrow \boldsymbol{\eta}_{p(\boldsymbol{x}|\,\boldsymbol{b},\sigma_x^2)\to \sigma_x^2},\\
G_{a_\varepsilon\to p(\sigma_\varepsilon^2|a_\varepsilon)} \longleftarrow G_{p(a_\varepsilon)\to a_\varepsilon},\ \ \boldsymbol{\eta}_{a_\varepsilon\to p(\sigma_\varepsilon^2|a_\varepsilon)} \longleftarrow \boldsymbol{\eta}_{p(a_\varepsilon)\to a_\varepsilon},\\
G_{a_x\to p(\sigma_x^2|a_x)} \longleftarrow G_{p(a_x)\to a_x},\ \ \boldsymbol{\eta}_{a_x\to p(\sigma_x^2|a_x)} \longleftarrow \boldsymbol{\eta}_{p(a_x)\to a_x},\\
\boldsymbol{\eta}_{a_\varepsilon\to p(a_\varepsilon)} \longleftarrow \boldsymbol{\eta}_{p(\sigma_\varepsilon^2|a_\varepsilon)\to a_\varepsilon},\ \ \boldsymbol{\eta}_{a_x\to p(a_x)} \longleftarrow \boldsymbol{\eta}_{p(\sigma_x^2|a_x)\to a_x}.
\end{gathered}$$

Note that the updates for $\boldsymbol{\eta}_{a_\varepsilon\to p(\sigma_\varepsilon^2|a_\varepsilon)}$ and $\boldsymbol{\eta}_{a_x\to p(\sigma_x^2|a_x)}$ remain constant throughout the iterations.

S.2.2.2. Factor to stochastic node message parameter updates

The updates for the parameters of factor to stochastic node messages require use of the VMP algorithms described in Subsection 3.2. The following is a detailed explanation of their usage to obtain the remaining updates.

Use Algorithm 2 of Maestrini and Wand (2021) for the iterated Inverse G-Wishart Fragment with:

Graph Input: $G = G_{\text{full}}$.

Shape Parameter Input: 1.

Message Graph Input: $G_{a_x\to p(\sigma_x^2|a_x)} = G_{\text{diag}}$.

Natural Parameter Inputs: $\boldsymbol{\eta}_{\sigma_x^2\to p(\sigma_x^2|a_x)}$, $\boldsymbol{\eta}_{p(\sigma_x^2|a_x)\to \sigma_x^2}$, $\boldsymbol{\eta}_{a_x\to p(\sigma_x^2|a_x)}$, $\boldsymbol{\eta}_{p(\sigma_x^2|a_x)\to a_x}$.

Graph Outputs: $G_{p(\sigma_x^2|a_x)\to \sigma_x^2}$, $G_{p(\sigma_x^2|a_x)\to a_x}$.

Natural Parameter Outputs: $\boldsymbol{\eta}_{p(\sigma_x^2|a_x)\to \sigma_x^2}$, $\boldsymbol{\eta}_{p(\sigma_x^2|a_x)\to a_x}$.

Use Algorithm 2 of Maestrini and Wand (2021) for the iterated Inverse G-Wishart Fragment with:

Graph Input: $G = G_{\text{full}}$.

Shape Parameter Input: 1.

Message Graph Input: $G_{a_\varepsilon\to p(\sigma_\varepsilon^2|a_\varepsilon)} = G_{\text{diag}}$.

Natural Parameter Inputs: $\boldsymbol{\eta}_{\sigma_\varepsilon^2\to p(\sigma_\varepsilon^2|a_\varepsilon)}$, $\boldsymbol{\eta}_{p(\sigma_\varepsilon^2|a_\varepsilon)\to \sigma_\varepsilon^2}$, $\boldsymbol{\eta}_{a_\varepsilon\to p(\sigma_\varepsilon^2|a_\varepsilon)}$, $\boldsymbol{\eta}_{p(\sigma_\varepsilon^2|a_\varepsilon)\to a_\varepsilon}$.

Graph Outputs: $G_{p(\sigma_\varepsilon^2|a_\varepsilon)\to \sigma_\varepsilon^2}$, $G_{p(\sigma_\varepsilon^2|a_\varepsilon)\to a_\varepsilon}$.

Natural Parameter Outputs: $\boldsymbol{\eta}_{p(\sigma_\varepsilon^2|a_\varepsilon)\to \sigma_\varepsilon^2}$, $\boldsymbol{\eta}_{p(\sigma_\varepsilon^2|a_\varepsilon)\to a_\varepsilon}$.

Use Algorithm 2 of the current work with:

Data Inputs: $\boldsymbol{y}$, $\boldsymbol{K}$.

Natural Parameter Inputs: $\boldsymbol{\eta}_{\boldsymbol{x} \to p(\boldsymbol{x}|\,\boldsymbol{b},\sigma_x^2)}$, $\boldsymbol{\eta}_{p(\boldsymbol{x}|\boldsymbol{b},\sigma_x^2) \to \boldsymbol{x}}$, $\boldsymbol{\eta}_{\sigma_x^2 \to p(\boldsymbol{x}|\,\boldsymbol{b},\sigma_x^2)}$, $\boldsymbol{\eta}_{p(\boldsymbol{x}|\,\boldsymbol{b},\sigma_x^2) \to \sigma_x^2}$.

Natural Parameter Outputs: $\boldsymbol{\eta}_{p(\boldsymbol{x}|\boldsymbol{b},\sigma_x^2) \to \boldsymbol{x}}$, $\boldsymbol{\eta}_{p(\boldsymbol{x}|\,\boldsymbol{b},\sigma_x^2) \to \sigma_x^2}$.

Use the algorithm of Section 4.1.5 of Wand (2017) for the Gaussian Likelihood Fragment with:

Natural Parameter Inputs: $\boldsymbol{\eta}_{\boldsymbol{x} \to p(\boldsymbol{y}|\boldsymbol{x},\sigma_\varepsilon^2)}$, $\boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{x},\sigma_\varepsilon^2) \to \boldsymbol{x}}$, $\boldsymbol{\eta}_{\sigma_\varepsilon^2 \to p(\boldsymbol{y}|\boldsymbol{x},\sigma_\varepsilon^2)}$, $\boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{x},\sigma_\varepsilon^2) \to \sigma_\varepsilon^2}$.

Natural Parameter Outputs: $\boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{x},\sigma_\varepsilon^2) \to \boldsymbol{x}}$, $\boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{x},\sigma_\varepsilon^2) \to \sigma_\varepsilon^2}$.

S.2.3. Approximating density functions

After reaching convergence, the remaining task is deriving the optimal approximating densities. The densities of main interest are $q^*(\boldsymbol{x})$, $q^*(\sigma_\varepsilon^2)$ and $q^*(\sigma_x^2)$, and have the form expressed in (10), (12) and (13) that come from (19). In particular,

$$q^*(\boldsymbol{x}) \propto \exp\left\{ \begin{bmatrix} \boldsymbol{x} \\ \text{vec}\left(\boldsymbol{x}\boldsymbol{x}^T\right) \end{bmatrix}^T \boldsymbol{\eta}_{q(\boldsymbol{x})} \right\}, \text{ with } \boldsymbol{\eta}_{q(\boldsymbol{x})} \equiv \boldsymbol{\eta}_{p(\boldsymbol{x}|\boldsymbol{b},\sigma_x^2) \to \boldsymbol{x}} + \boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{x},\sigma_\varepsilon^2) \to \boldsymbol{x}},$$

$$q^*(\sigma_\varepsilon^2) \propto \exp\left\{ \begin{bmatrix} \log(\sigma_\varepsilon^2) \\ 1/\sigma_\varepsilon^2 \end{bmatrix}^T \boldsymbol{\eta}_{q(\sigma_\epsilon^2)} \right\}, \text{ with } \boldsymbol{\eta}_{q(\sigma_\epsilon^2)} \equiv \boldsymbol{\eta}_{p(\sigma_\varepsilon^2|a_\varepsilon) \to \sigma_\varepsilon^2} + \boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{x},\sigma_\varepsilon^2) \to \sigma_\varepsilon^2}$$

$$\text{and } q^*(\sigma_x^2) \propto \exp\left\{ \begin{bmatrix} \log(\sigma_x^2) \\ 1/\sigma_x^2 \end{bmatrix}^T \boldsymbol{\eta}_{q(\sigma_x^2)} \right\}, \text{ with } \boldsymbol{\eta}_{q(\sigma_x^2)} \equiv \boldsymbol{\eta}_{p(\sigma_x^2|a_x) \to \sigma_x^2} + \boldsymbol{\eta}_{p(\boldsymbol{x}|\,\boldsymbol{b},\sigma_x^2) \to \sigma_x^2}.$$

The q-density common parameters are then the following:

$$\boldsymbol{\mu}_{q(\boldsymbol{x})} = \boldsymbol{\Sigma}_{q(\boldsymbol{x})} \left(\boldsymbol{\eta}_{q(\boldsymbol{x})}\right)_{1:m}, \quad \boldsymbol{\Sigma}_{q(\boldsymbol{x})} = -\tfrac{1}{2}\,\text{vec}^{-1}\left\{\left(\boldsymbol{\eta}_{q(\boldsymbol{x})}\right)_{(m+1):m^2}\right\},$$

$$\kappa_{q(\sigma_\varepsilon^2)} = -2\left(1 + \left(\boldsymbol{\eta}_{q(\sigma_\epsilon^2)}\right)_1\right), \quad \lambda_{q(\sigma_\varepsilon^2)} = -2\left(\boldsymbol{\eta}_{q(\sigma_\epsilon^2)}\right)_2,$$

$$\kappa_{q(\sigma_x^2)} = -2\left(1 + \left(\boldsymbol{\eta}_{q(\sigma_x^2)}\right)_1\right), \quad \lambda_{q(\sigma_x^2)} = -2\left(\boldsymbol{\eta}_{q(\sigma_x^2)}\right)_2.$$

Expressions for the other optimal approximating densities can be obtained in a similar manner.

S.3. Removing $\boldsymbol{L}$ from the variational algorithms

We provide results that allow to simplify the variational algorithm updates involving the contrast matrix $\boldsymbol{L}$ by means of simpler computational steps. The results are presented separately for one- and two-dimensional problems and make use of the following notation.

Definition 1. *For vectors* $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p$,

$$\underset{i=1,\ldots,p}{\text{stack}}\,(\boldsymbol{v}_i) \equiv \begin{bmatrix} \boldsymbol{v}_1 \\ \vdots \\ \boldsymbol{v}_p \end{bmatrix}.$$

Definition 2. *For vectors $\boldsymbol{v}$ and $\tilde{\boldsymbol{v}}$, respectively of length $d_{\boldsymbol{v}}$ and $d_{\boldsymbol{v}} - 1$,*

$$\text{tridiag}\left(\boldsymbol{v}, \tilde{\boldsymbol{v}}\right) \equiv \begin{bmatrix} v_1 & \tilde{v}_1 & 0 & \cdots & 0 \\ \tilde{v}_1 & v_2 & \tilde{v}_2 & \ddots & \vdots \\ 0 & \tilde{v}_2 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & v_{d_{\boldsymbol{v}}-1} & \tilde{v}_{d_{\boldsymbol{v}}-1} \\ 0 & \cdots & 0 & \tilde{v}_{d_{\boldsymbol{v}}-1} & v_{d_{\boldsymbol{v}}} \end{bmatrix}.$$

Definition 3. *For vectors $\boldsymbol{v}$ and $\tilde{\boldsymbol{v}}$, respectively of length $d_{\boldsymbol{v}}$ and $(d_{\boldsymbol{v}} - c)$, with $d_{\boldsymbol{v}} > c$ and $c \in \mathbb{N}$,*

$$\text{sparsetridiag}\left(\boldsymbol{v}, \tilde{\boldsymbol{v}}, c\right) \equiv \begin{bmatrix} v_1 & \mathbf{0}_{c-1}^T & \tilde{v}_1 & 0 & \cdots & 0 \\ \mathbf{0}_{c-1} & v_2 & & \tilde{v}_2 & \ddots & \vdots \\ \tilde{v}_1 & & & \ddots & \ddots & 0 \\ 0 & \tilde{v}_2 & \ddots & \ddots & & \tilde{v}_{d_{\boldsymbol{v}}-c} \\ \vdots & \ddots & \ddots & & & \mathbf{0}_{c-1}^T \\ 0 & \cdots & 0 & \tilde{v}_{d_{\boldsymbol{v}}-c} & \mathbf{0}_{c-1} & v_{d_{\boldsymbol{v}}} \end{bmatrix}.$$

Note that if $c = 1$, then $\text{sparsetridiag}\left(\boldsymbol{v}, \tilde{\boldsymbol{v}}, 1\right) = \text{tridiag}\left(\boldsymbol{v}, \tilde{\boldsymbol{v}}\right)$.

S.3.1. One-dimensional case

Consider a one-dimensional inverse problem analyzed via model (2) and suppose $\boldsymbol{x}$ has one-to-one correspondence with a sequence of hidden and equispaced locations on a line. Assume first nearest neighbor differences are modeled via the contrast matrix $\boldsymbol{L}_{1D}$ defined in (4). Then matrix $\boldsymbol{L}$ can be removed from (24) by making use of the following lemma.

Lemma 1. *Let $\boldsymbol{v}$ be a vector of length $d_{\boldsymbol{v}}$ and $\boldsymbol{L}_{1D}$ a $(d_{\boldsymbol{v}} - 1) \times d_{\boldsymbol{v}}$ matrix having form* (4)*. Then*

$$\boldsymbol{L}_{1D}\boldsymbol{v} = \begin{bmatrix} v_2 - v_1 \\ v_3 - v_2 \\ \vdots \\ v_{d_{\boldsymbol{v}}} - v_{d_{\boldsymbol{v}}-1} \end{bmatrix},$$

which is a vector of length $d_{\boldsymbol{v}} - 1$.

We can get rid of matrix $\boldsymbol{L}$ from expressions having form (25) through the next lemma.

Lemma 2. *Let $\boldsymbol{M}$ be a symmetric $d_{\boldsymbol{M}} \times d_{\boldsymbol{M}}$ matrix and $\boldsymbol{L}_{1D}$ a $(d_{\boldsymbol{M}} - 1) \times d_{\boldsymbol{M}}$ matrix having form* (4)*. Then for $i = 1, \ldots, d_{\boldsymbol{M}} - 1$,*

$$\left(\boldsymbol{L}_{1D}\boldsymbol{M}\boldsymbol{L}_{1D}^T\right)_{ii} = M_{i+1,i+1} - 2M_{i+1,i} + M_{i,i}$$

or, equivalently,

$$\begin{aligned} \text{diagonal}\left(\boldsymbol{L}_{1D}\boldsymbol{M}\boldsymbol{L}_{1D}^T\right) = {} & \text{diagonal}\left(\boldsymbol{M}\right)_{2:d_{\boldsymbol{M}}} + \text{diagonal}\left(\boldsymbol{M}\right)_{1:(d_{\boldsymbol{M}}-1)} \\ & - 2\ \text{diagonal}\left(\boldsymbol{M}_{2:d_{\boldsymbol{M}},1:(d_{\boldsymbol{M}}-1)}\right), \end{aligned}$$

which is a vector of length $d_{\boldsymbol{M}} - 1$.

The following lemma simplifies the computation of (26).

Notation	Description
m_1	number of rows in $\boldsymbol{X}$ (and $\boldsymbol{Y}$)
m_2	number of columns in $\boldsymbol{X}$ (and $\boldsymbol{Y}$)
$m = m_1 \times m_2$	total number of elements in $\boldsymbol{X}$ (and $\boldsymbol{Y}$)
$d^{\text{H}} = m_1(m_2 - 1)$	number of horizontal differences in $\boldsymbol{X}$
$d^{\text{V}} = (m_1 - 1)m_2$	number of vertical differences in $\boldsymbol{X}$
$d = d^{\text{H}} + d^{\text{V}}$	total number of differences in $\boldsymbol{X}$

Table S.1: *Notation used for two-dimensional inverse problems with first nearest neighbor differences and one-to-one correspondence between $\boldsymbol{X}$ and $\boldsymbol{Y}$.*

Lemma 3. *Let $\boldsymbol{w}$ be a vector of length $d_{\boldsymbol{w}}$ and $\boldsymbol{L}_{1D}$ a $(d_{\boldsymbol{w}} - 1) \times d_{\boldsymbol{w}}$ matrix having form* (4)*. Then*

$$\boldsymbol{L}_{1D}^T \operatorname{diag}(\boldsymbol{w}) \boldsymbol{L}_{1D} = \operatorname{tridiag}\left(\begin{bmatrix} w_1 \\ \boldsymbol{w}_{1:(d_{\boldsymbol{w}}-1)} + \boldsymbol{w}_{2:d_{\boldsymbol{w}}} \\ w_{d_{\boldsymbol{w}}} \end{bmatrix}, -\boldsymbol{w} \right),$$

which is a matrix of size $d_{\boldsymbol{w}} \times d_{\boldsymbol{w}}$.

In the R computing environment Lemmas 1–3 can be implemented with standard base functions. In particular, the function `diff(v)` automatically produces the result stated in Lemma 1. The expressions originated by Lemmas 2 and 3 can be visualized through the examples provided in Section S.3.3.

S.3.2. Two-dimensional case

Consider the study of bidimensional inverse problems through model (2) with $\boldsymbol{x} = \text{vec}(\boldsymbol{X})$ and $\boldsymbol{X}$ being an $m_1 \times m_2$ matrix whose entries correspond to a regular grid of locations. Then a first nearest neighbor contrast matrix, here denoted by $\boldsymbol{L}_{2D}$, can be conveniently defined as

$$\boldsymbol{L}_{2D} \equiv \begin{bmatrix} \boldsymbol{L}^{\text{H}}\boldsymbol{C} \\ \boldsymbol{L}^{\text{V}} \end{bmatrix}, \tag{S.3}$$

which has size $d \times (m_1 m_2)$, with $d = d^{\text{H}} + d^{\text{V}}$, $d^{\text{H}} = m_1(m_2 - 1)$ and $d^{\text{V}} = (m_1 - 1)m_2$. The single components of $\boldsymbol{L}_{2D}$ are defined as follows:

$$\boldsymbol{L}^{\text{H}} \equiv \boldsymbol{I}_{m_1} \otimes \boldsymbol{L}_0^{\text{H}} \quad \text{and} \quad \boldsymbol{L}^{\text{V}} \equiv \boldsymbol{I}_{m_2} \otimes \boldsymbol{L}_0^{\text{V}}, \tag{S.4}$$

where $\boldsymbol{L}^{\text{H}}$ and $\boldsymbol{L}^{\text{V}}$ are matrices of size $d^{\text{H}} \times (m_1 m_2)$ and $d^{\text{V}} \times (m_1 m_2)$, respectively. The matrix $\boldsymbol{C}$ is a $m_1 \times m_2$ *commutation matrix* such that

$$\boldsymbol{C}\,\text{vec}(\boldsymbol{X}) = \text{vec}(\boldsymbol{X}^T).$$

Lastly, matrices $\boldsymbol{L}_0^{\text{H}}$ and $\boldsymbol{L}_0^{\text{V}}$ have the form (4) identified for the one-dimensional case, and dimensions $(m_2 - 1) \times m_2$ and $(m_1 - 1) \times m_1$, respectively. If for instance $\boldsymbol{X}$ is of size 3×4 as in the example of Figure 1, then

$$\boldsymbol{L}_0^{\text{H}} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{L}_0^{\text{V}} = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}. \tag{S.5}$$

Superscripts H and V refer to differences between pairs of locations respectively computed in a horizontal and vertical fashion, given the row-wise and column-wise orientation identified by $\boldsymbol{X}$. The explicit expression of $\boldsymbol{L}_{2D}$ for a 3×4 matrix of parameters $\boldsymbol{X}$ is provided in Section S.3.4. For clarity, the coefficients that define the dimensions of the $\boldsymbol{L}_{2D}$ matrix components are summarized in Table (S.1).

Lemmas 4–6 are extensions of Lemmas 1–3 to the two-dimensional case, provided that $\boldsymbol{L}_{2D}$ has the form identified by (S.3)–(S.5). Vector and matrix dimensions are intentionally emphasized to highlight analogies with one-dimensional problems and set up a framework extendible to higher dimensions.

The next lemma simplifies the computation of expression (24) and avoids calculating and storing matrix $\boldsymbol{L}_{2D}$.

Lemma 4. *Let $\boldsymbol{V}$ be a matrix of size $m_1 \times m_2$, $\boldsymbol{v} = \text{vec}\,(\boldsymbol{V})$ a vector of length $d_v = m_1 m_2$ and $\boldsymbol{L}_{2D}$ a matrix of size $(d^H + d^V) \times d_v$ having the form defined in* (S.3)–(S.5) *with $d^H = m_1(m_2 - 1)$ and $d^V = (m_1 - 1)m_2$. Then*

$$\boldsymbol{L}_{2D}\boldsymbol{v} = \begin{bmatrix} \boldsymbol{t}^H \\ \boldsymbol{t}^V \end{bmatrix},$$

where

$$\boldsymbol{t}^H = \underset{i=1,\ldots,m_1}{\text{stack}} \left(\begin{bmatrix} u_{m_2(i-1)+2} - u_{m_2(i-1)+1} \\ u_{m_2(i-1)+3} - u_{m_2(i-1)+2} \\ \vdots \\ u_{m_2 i} - u_{m_2 i-1} \end{bmatrix} \right)$$

and

$$\boldsymbol{t}^V = \underset{i=1,\ldots,m_2}{\text{stack}} \left(\begin{bmatrix} v_{m_1(i-1)+2} - v_{m_1(i-1)+1} \\ v_{m_1(i-1)+3} - v_{m_1(i-1)+2} \\ \vdots \\ v_{m_1 i} - v_{m_1 i-1} \end{bmatrix} \right),$$

are vectors of length d^H and d^V, respectively, and $\boldsymbol{u} = \text{vec}\left(\boldsymbol{V}^T\right)$.

Expression (25) can be efficiently computed using the following lemma.

Lemma 5. *Let $\boldsymbol{M}$ be a symmetric $d_M \times d_M$ matrix with $d_M = m_1 m_2$ and $\boldsymbol{L}_{2D}$ a matrix of size $(d^H + d^V) \times d_M$ having the form defined in* (S.3)–(S.5) *with $d^H = m_1(m_2 - 1)$ and $d^V = (m_1 - 1)m_2$. Then*

$$\text{diagonal}\left(\boldsymbol{L}_{2D}\boldsymbol{M}\boldsymbol{L}_{2D}^T\right) = \begin{bmatrix} \boldsymbol{s}^H \\ \boldsymbol{s}^V \end{bmatrix},$$

where $\boldsymbol{s}^H$ and $\boldsymbol{s}^V$ are vectors of length d^H and d^V, respectively, such that

$$\boldsymbol{s}^H = \text{vec}\left\{ \left(\text{vec}^{-1}_{m_1,(m_2-1)}\left(\boldsymbol{s}_0^H\right) \right)^T \right\},$$

$$\boldsymbol{s}_0^H = \text{diagonal}\,(\boldsymbol{M})_{(m_1+1):d_M} - 2\ \text{diagonal}\left(\boldsymbol{M}_{(m_1+1):d_M,1:(d_M-m_1)}\right) + \text{diagonal}\,(\boldsymbol{M})_{1:(d_M-m_1)}$$

and

$$\boldsymbol{s}^V = \text{diagonal}\,(\boldsymbol{M})_{(k_1+1):(k_{d^V}+1)} - 2\ \text{diagonal}\left(\boldsymbol{M}_{(k_1+1):(k_{d^V}+1),k_1:k_{d^V}}\right) + \text{diagonal}\,(\boldsymbol{M})_{k_1:k_{d^V}},$$

and $\boldsymbol{k}$ is a vector of d^V indeces obtained as follows:

$$\boldsymbol{k} = \underset{i=1,\ldots,m_2}{\text{stack}} \left(\begin{bmatrix} m_1(i-1)+1 \\ m_1(i-1)+2 \\ \vdots \\ m_1 i - 1 \end{bmatrix} \right).$$

The last lemma reduces the computational effort for implementing (26).

Lemma 6. *Let $\boldsymbol{w}$ be a vector of length $d_{\boldsymbol{w}} = d^H + d^V$, with $d^H = m_1(m_2 - 1)$ and $d^V = (m_1 - 1)m_2$, and $\boldsymbol{L}_{2D}$ a matrix of size $d_{\boldsymbol{w}} \times (m_1 m_2)$ having the form defined in* (S.3)–(S.5)*. Then*

$$\boldsymbol{L}_{2D}^T \operatorname{diag}(\boldsymbol{w}) \boldsymbol{L}_{2D} = \boldsymbol{R} - \operatorname{diag}\left(\left[\sum_{j=1}^{m_1 m_2} R_{i,j} \right]_{i=1,\ldots,m_1 m_2} \right),$$

where

$$\boldsymbol{R} = \text{sparsetridiag}\left(\mathbf{0}_{m_1 m_2}, r^H, m_1\right) + \text{sparsetridiag}\left(\mathbf{0}_{m_1 m_2}, r^V, 1\right)$$

is a matrix of size $(m_1 m_2) \times (m_1 m_2)$, and

$$r^H = -\operatorname{vec}\left\{ \left(\operatorname{vec}^{-1}_{(m_2-1),m_1}\left(\boldsymbol{w}_{1:d^H}\right) \right)^T \right\}$$

$$\text{and} \quad r^V = -\left[\operatorname{vec}\left(\begin{bmatrix} \operatorname{vec}^{-1}_{(m_1-1),m_2}\left(\boldsymbol{w}_{(d^H+1):d_w}\right) \\ \mathbf{0}_{m_2}^T \end{bmatrix} \right) \right]_{1:(m_1 m_2 - 1)}$$

are vectors of length d^H and $m_1 m_2 - 1$, respectively.

The supplement provides heuristic arguments to prove Lemmas 4–6 and demonstrates the implementation of the two-dimensional case lemmas in R.

S.3.3. Visualization of Lemmas 1–3

Consider a simple one-dimensional inverse problem modeled through (2) where $m = 4$ and $d = m - 1 = 3$. For this particular example the contrast matrix has size $d \times m$ and the following explicit expression:

$$\boldsymbol{L}_{1D} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}.$$

S.3.3.1. Visualization of Lemma 1

Vector $\boldsymbol{v}$ has length $d_{\boldsymbol{v}} = m = 4$ and

$$\boldsymbol{L}_{1D}\boldsymbol{v} = \begin{bmatrix} v_2 - v_1 \\ v_3 - v_2 \\ v_4 - v_3 \end{bmatrix}.$$

S.3.3.2. Visualization of Lemma 2

Matrix $\boldsymbol{M}$ is symmetric and has size $d_{\boldsymbol{M}} \times d_{\boldsymbol{M}}$, with $d_{\boldsymbol{M}} = m = 4$, and

$$\operatorname{diagonal}\left(\boldsymbol{L}_{1D} \boldsymbol{M} \boldsymbol{L}_{1D}^T \right) = \begin{bmatrix} M_{2,2} - 2M_{2,1} + M_{1,1} \\ M_{3,3} - 2M_{3,2} + M_{2,2} \\ M_{4,4} - 2M_{4,3} + M_{3,3} \end{bmatrix}.$$

S.3.3.3. Visualization of Lemma 3

Vector $\boldsymbol{w}$ has length $d_{\boldsymbol{w}} = d = 3$ and

$$\boldsymbol{L}_{1D}^T \operatorname{diag}\left(\boldsymbol{w}\right) \boldsymbol{L}_{1D} = \begin{bmatrix} w_1 & -w_1 & 0 & 0 \\ -w_1 & w_1 + w_2 & -w_2 & 0 \\ 0 & -w_2 & w_2 + w_3 & -w_3 \\ 0 & 0 & -w_3 & w_3 \end{bmatrix}.$$

S.3.4. Visualization and R implementation of Lemmas 4–6

This subsection shows the results stated in Lemmas 4–6 through a simple two-dimensional inverse problem model where $\boldsymbol{X}$ is a matrix of parameters of size $m_1 \times m_2$ with $m_1 = 3$ and $m_2 = 4$. Referring to model (2), the length of $\boldsymbol{x} = \text{vec}(\boldsymbol{X})$ is $m = m_1 \times m_2$. For this particular case the contrast matrix $\boldsymbol{L}_{2\text{D}}$ defined in (S.3) is a matrix of size $d \times m$, hence of size 17×12, given that $d = d^{\text{H}} + d^{\text{V}}$, $d^{\text{H}} = m_1(m_2 - 1) = 9$ and $d^{\text{V}} = (m_1 - 1)m_2 = 8$. The sub-components of $\boldsymbol{L}_{2\text{D}}$ are

$$\boldsymbol{L}^{\text{H}}\boldsymbol{C} = \begin{bmatrix} -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 \end{bmatrix}$$

and

$$\boldsymbol{L}^{\text{V}} = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix},$$

where $\boldsymbol{L}^{\text{H}}$ and $\boldsymbol{L}^{\text{V}}$ are defined through (S.4) by making use of matrices $\boldsymbol{L}_0^{\text{H}}$ and $\boldsymbol{L}_0^{\text{V}}$ shown in (S.5), and $\boldsymbol{C}$ is a commutation matrix of appropriate size.

The objective of the following subsubsections is to visualize the results expressed in Lemmas 4–6 for the particular case under examination.

S.3.4.1. Visualization of Lemma 4

Vector $\boldsymbol{v}$ has length $d_v = m_1 m_2 = 12$ and

$$\boldsymbol{L}_{2\text{D}}\boldsymbol{v} = \begin{bmatrix} \boldsymbol{t}^{\text{H}} \\ \boldsymbol{t}^{\text{V}} \end{bmatrix},$$

where

$$\boldsymbol{t}^{\text{H}} = \begin{bmatrix} v_4 - v_1 \\ v_7 - v_4 \\ v_{10} - v_7 \\ v_5 - v_2 \\ v_8 - v_5 \\ v_{11} - v_8 \\ v_6 - v_3 \\ v_9 - v_6 \\ v_{12} - v_9 \end{bmatrix}, \quad \boldsymbol{t}^{\text{V}} = \begin{bmatrix} v_2 - v_1 \\ v_3 - v_2 \\ v_5 - v_4 \\ v_6 - v_5 \\ v_8 - v_7 \\ v_9 - v_8 \\ v_{11} - v_{10} \\ v_{12} - v_{11} \end{bmatrix}.$$

S.3.4.2. Visualization of Lemma 5

Matrix $\boldsymbol{M}$ is symmetric and has size $d_M \times d_M$, with $d_M = m_1 m_2 = 12$, and

$$\text{diagonal}\left(\boldsymbol{L}_{2\text{D}}\boldsymbol{M}\boldsymbol{L}_{2\text{D}}^T\right) = \begin{bmatrix} \boldsymbol{s}^{\text{H}} \\ \boldsymbol{s}^{\text{V}} \end{bmatrix},$$

where

$$\boldsymbol{s}^{\mathrm{H}} = \begin{bmatrix} M_{4,4} - 2M_{4,1} + M_{1,1} \\ M_{7,7} - 2M_{7,4} + M_{4,4} \\ M_{10,10} - 2M_{10,7} + M_{7,7} \\ M_{5,5} - 2M_{5,2} + M_{2,2} \\ M_{8,8} - 2M_{8,5} + M_{5,5} \\ M_{11,11} - 2M_{11,8} + M_{8,8} \\ M_{6,6} - 2M_{6,3} + M_{3,3} \\ M_{9,9} - 2M_{9,6} + M_{6,6} \\ M_{12,12} - 2M_{12,9} + M_{9,9} \end{bmatrix}, \quad \boldsymbol{s}^{\mathrm{V}} = \begin{bmatrix} M_{2,2} - 2M_{2,1} + M_{1,1} \\ M_{3,3} - 2M_{3,2} + M_{2,2} \\ M_{5,5} - 2M_{5,4} + M_{4,4} \\ M_{6,6} - 2M_{6,5} + M_{5,5} \\ M_{8,8} - 2M_{8,7} + M_{7,7} \\ M_{9,9} - 2M_{9,8} + M_{8,8} \\ M_{11,11} - 2M_{11,10} + M_{10,10} \\ M_{12,12} - 2M_{12,11} + M_{11,11} \end{bmatrix}.$$

S.3.4.3. Visualization of Lemma 6

Vector $\boldsymbol{w}$ has length $d_{\boldsymbol{w}} = d = 17$ and

$$\boldsymbol{L}_{2\mathrm{D}}^{T} \operatorname{diag}(\boldsymbol{w}) \boldsymbol{L}_{2\mathrm{D}} = \boldsymbol{R} + \operatorname{diag}\left(\begin{bmatrix} w_1 + w_{10} \\ w_4 + w_{10} + w_{11} \\ w_7 + w_{11} \\ w_1 + w_2 + w_{12} \\ w_4 + w_5 + w_{12} + w_{13} \\ w_7 + w_8 + w_{13} \\ w_2 + w_3 + w_{14} \\ w_5 + w_6 + w_{14} + w_{15} \\ w_8 + w_9 + w_{15} \\ w_3 + w_{16} \\ w_6 + w_{16} + w_{17} \\ w_9 + w_{17} \end{bmatrix}\right)$$

where

$$\boldsymbol{R} = \begin{bmatrix} 0 & -w_{10} & 0 & -w_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -w_{10} & 0 & -w_{11} & 0 & -w_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -w_{11} & 0 & 0 & 0 & -w_7 & 0 & 0 & 0 & 0 & 0 & 0 \\ -w_1 & 0 & 0 & 0 & -w_{12} & 0 & -w_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & -w_4 & 0 & -w_{12} & 0 & -w_{13} & 0 & -w_5 & 0 & 0 & 0 & 0 \\ 0 & 0 & -w_7 & 0 & -w_{13} & 0 & 0 & 0 & -w_8 & 0 & 0 & 0 \\ 0 & 0 & 0 & -w_2 & 0 & 0 & 0 & -w_{14} & 0 & -w_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & -w_5 & 0 & -w_{14} & 0 & -w_{15} & 0 & -w_6 & 0 \\ 0 & 0 & 0 & 0 & 0 & -w_8 & 0 & -w_{15} & 0 & 0 & 0 & -w_9 \\ 0 & 0 & 0 & 0 & 0 & 0 & -w_3 & 0 & 0 & 0 & -w_{16} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -w_6 & 0 & -w_{16} & 0 & -w_{17} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -w_9 & 0 & -w_{17} & 0 \end{bmatrix}.$$

This involves the following two vectors:

$$\boldsymbol{r}^{\mathrm{H}} = -\begin{bmatrix} w_1 \\ w_4 \\ w_7 \\ w_2 \\ w_5 \\ w_8 \\ w_3 \\ w_6 \\ w_9 \end{bmatrix} \text{ and } \boldsymbol{r}^{\mathrm{V}} = -\begin{bmatrix} w_{10} \\ w_{11} \\ 0 \\ w_{12} \\ w_{13} \\ 0 \\ w_{14} \\ w_{15} \\ 0 \\ w_{16} \\ w_{17} \end{bmatrix}.$$

S.3.5. Implementation of Lemmas 4–6 in R

This subsection provides R code to implement Lemmas A, B and C for two-dimensional inverse problems. We make use of function *invvec* from package `ks` (Duong, 2024). Note that this function specifies matrix dimension with the number of columns preceding the number of rows. This is at odds with the definition of vec^{-1} given in Section 1.2.

Consider, for instance, a two-dimensional dataset of size 50×40:

```
# Load required library:

library(ks)

# Obtain dimensions:

m1 <- 50   ;   m2 <- 40
m <- m1*m2
dH <- m1*(m2-1)  ;  dV <- (m1-1)*m2
d <- dH + dV
```

The result expressed in Lemma 4 can be computed in R with the following code:

```
# Create a vector of length m:

v <- rnorm(m)

# Compute the result of Lemma 4:

vecVt <- vec(t(invvec(v,m2,m1)))
indSetH <- seq(from=m2,to=dV,by=m2)
indSetV <- seq(from=m1,to=dH,by=m1)
tH <- diff(vecVt)[-indSetH]
tV <- diff(v)[-indSetV]
lemma4res <- c(tH,tV)
```

The result expressed in Lemma 5 can be computed in R with the following code:

```
# Generate a square symmetric matrix M via vector v

M <- tcrossprod(v)
dM <- dim(M)[1]

# Compute the result of Lemma 5:

sH <- (diag(M)[(m1+1):dM] - 2*diag(M[(m1+1):dM,1:(dM-m1)])
       + diag(M)[1:(dM-m1)])
sH <- vec(t(invvec(sH,m2-1,m1)))
sV <- (diag(M)[setdiff(2:dM,indSetV+1)]
       - 2*diag(M[setdiff(2:dM,indSetV+1),
                  setdiff(1:(dM-1),indSetV)])
       + diag(M)[setdiff(1:(dM-1),indSetV)])
lemma5res <- c(sH,sV)
```

The result expressed in Lemma 6 can be computed in R as follows:

```
# Generate a vector w of lenght d

w <- rnorm(d)

# Compute the result of Lemma 6:

rH <- - vec(t(invvec(w[1:dH],m1,m2-1)))
rV <- - vec(rbind(invvec(w[-(1:dH)],m2,m1-1),rep(0,m2)))[-m]
lemma6res <- matrix(0,m,m)
diag(lemma6res[1:(m-m1),(m1+1):m]) <- rH
diag(lemma6res[(m1+1):m,1:(m-m1)]) <- rH
diag(lemma6res[1:(m-1),2:m]) <- rV
diag(lemma6res[2:m,1:(m-1)]) <- rV
diag(lemma6res) <- - rowSums(lemma6res)
```

S.4. Additional biomedical data details

Here we provide supplementary results to assess the performance of variational inference versus MCMC for the real biomedical data and additional details concerning the simulated biomedical data.

S.4.1. Details on real biomedical data

Figures S.1–S.3 provide a comparison between variational inference and MCMC in terms of marginal and bivariate marginal posterior densities for Location 1 (external to the mouse body), Location 2 (mouse thyroid) and Location 6 (mouse bladder) of Figure 6. The bivariate marginal densities are 95% ellipsoid obtained via the function *geom_density* of the R package `ggplot2` (Wickham, 2016). While variational inference provides satisfactory approximations for Locations 1 and 2, its performance is less satisfactory for Location 6 corresponding to the mouse bladder and therefore to a location corresponding to higher technetium radioisotope emissions.

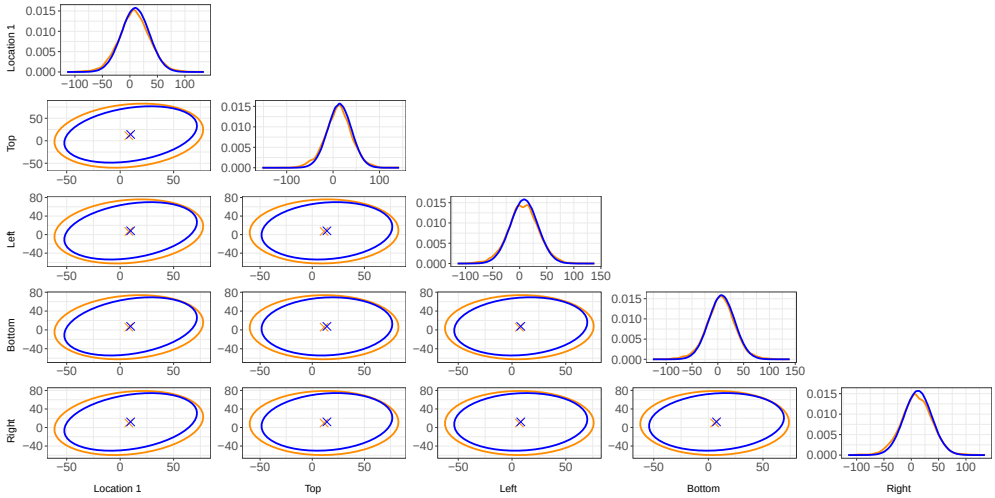

Figure S.1: *Marginal and bivariate marginal posterior densities obtained via variational inference (blue) and MCMC (orange) for Location 1 of Figure 6 and the adjacent pixels on its top, left, bottom and right sides.*

S.4.2. Details on simulated biomedical data

Figure S.4 shows plots of three simulated biomedical datasets, one for each δ value considered in the simulation study. Higher δ values correspond to more blur in the images.

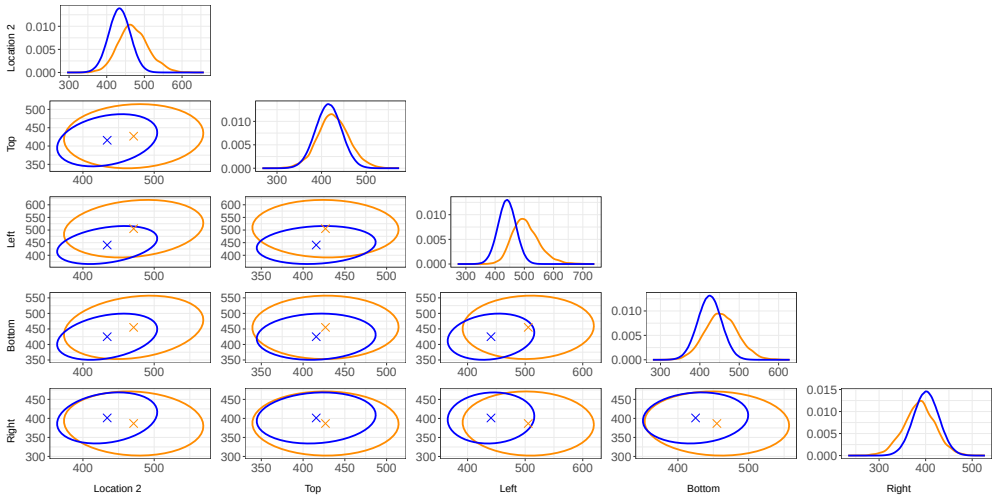

Figure S.2: *Marginal and bivariate marginal posterior densities obtained via variational inference (blue) and MCMC (orange) for Location 2 of Figure 6 and the adjacent pixels on its top, left, bottom and right sides.*

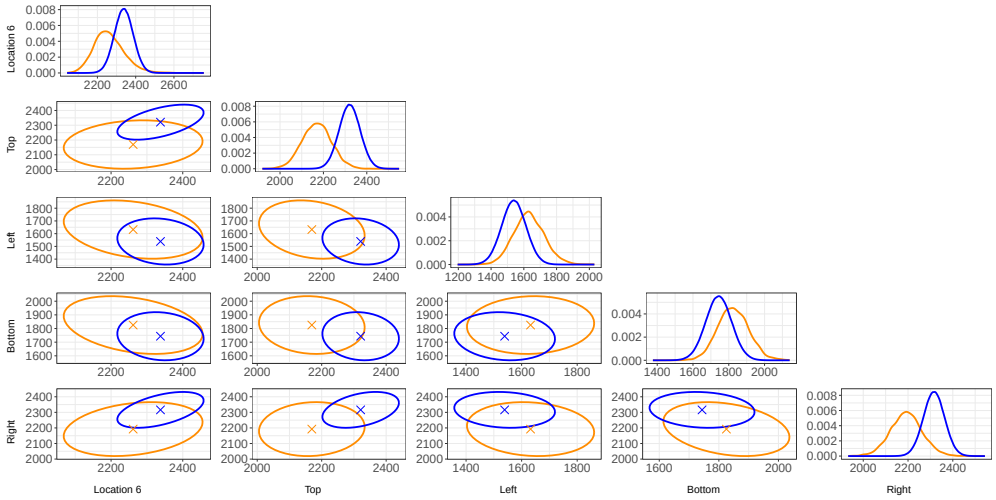

Figure S.3: *Marginal and bivariate marginal posterior densities obtained via variational inference (blue) and MCMC (orange) for Location 1 of Figure 6 and the adjacent pixels on its top, left, bottom and right sides.*

S.5. Illustration for archaeological data

We here show how the message passing on factor graph fragments paradigm can be used to move from an inverse problem model analysis to another without deriving a variational inference algorithm from scratch. This is illustrated through data from archaeological magnetometry. In archaeology it is often required to investigate a potential site via geophysical remote sensing methods before any physical excavation is commenced. The model we consider includes a Skew Normal distribution for the response vector and a Horseshoe penalization in replacement to the Normal response and Laplace penalization of model (2).

The data were collected from a mid Iron Age farmstead known as 'The Park' through an archaeological exploration that took place in 1994 at Guiting Power in Gloucestershire, United Kingdom. After data collection, part of the area was also excavated and archaeologists drew an impression of the remains that were brought to light.

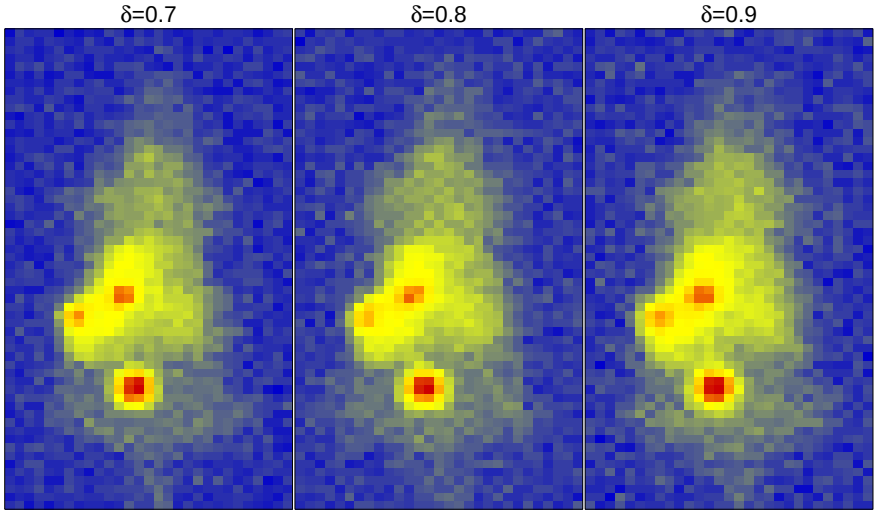

Figure S.4: *Examples of data generated in the biomedical simulation study for different values of δ.*

The archaeological site was partitioned into a grid of 10m$\times$10m squares with the survey axes aligned in the directions of magnetic north and east. A fluxgate gradiometer FM18 with 0.1nT sensitivity was used to collect the data at 0.5m intervals. For each survey square, the gradiometer output was an array of 20×20 magnetic anomaly readings corresponding to the difference between the signals detected by the lower and upper sensors. The gradiometer lower sensor was held at 0.2m above the surface of the site and the upper sensor was fixed a further 0.5m higher. The Earth's magnetic field is detected by both sensors, whereas the magnetic field from the buried feature is mostly detected by the lower sensor. Therefore the difference between the two sensors' readings provides the magnetic anomaly due to the hidden feature, together with small random noise components. The random noise may be due to systematic causes such as machine-rounding errors, or non-systematic factors such as disturbance by small stones in the soil or interference from local magnetic sources (Scollar, 1970).

We model the hidden surface through a single layer of rectangular prisms at a fixed burial depth. The scope is to estimate the prisms' magnetic *susceptibility* in order to separate the constituent epochs of the site and locate relevant artifacts prior to excavation. The subsurface layer is assumed to be fixed at a burial depth of 0.3m, given that the topsoil across the excavated region of Guiting Power was found to be between 0.25 and 0.3m deep. According to the single layer subsurface model, each prism in the layer has the same constant extent, which we set to 0.5m. Although the vertical extent of each of the excavated features vary from 0.45m to 1.6m, the chosen value of prism vertical extent increases the chances of distinguishing low-susceptibility features.

All this information is relevant to design an appropriate matrix $\boldsymbol{K}$ suitable for inverse problems concerning archaeological data.

S.5.1. $\boldsymbol{K}$ matrix for archaeological data

The matrix $\boldsymbol{K}$ we adopt for the illustration on archaeological data has a more complicated definition and relies upon the *spread function* defined in Section 2 of Aykroyd et al. (2001). The spread function models magnetic anomalies and depends on latitude and longitude of the archaeological site on the Earth's surface, the geometry of the gradiometer used to survey the area and the site physical properties.

Suppose all the magnetic features are located at the same depth below the surface, have the same vertical thickness and the susceptibility is constant along any vertical line through the features, but may vary between horizontal locations. We model the subsurface of the archaeological site as an ensemble of volume elements of equal size, called *prisms*, each having uniform susceptibility, fixed vertical depth (0.3m) and extent (0.5m), and a square cross section in the horizontal plane. Let (x_1, y_1, z_1) and (x_2, y_2, z_2) be the

coordinates of opposite vertices of a prism with unit susceptibility, where the x, y and z axes point north, east and vertically downward, respectively. Then the vertical component of the anomaly due to the prism at a point with coordinates (x, y, z) is

$$\Delta Z(x,y,z) = \frac{B}{4\pi}\left\{\left[\Delta Z^{(1)} + \Delta Z^{(2)} + \Delta Z^{(3)}\right]_{\zeta=z-z_2}^{\zeta=z-z_1}\right\},$$

where $B \approx 4.8 \times 10^4$nT (nanoteslas) is the magnitude of the magnetic flux density due to the Earth's field. The three additive components of $\Delta Z(x, y, z)$ are:

$$\begin{aligned}
\Delta Z^{(1)} &= \left[-\sin I \tan^{-1}\left(\frac{\xi\eta}{\zeta(\xi^2+\eta^2+\zeta^2)^{1/2}}\right)\right]_{\xi=x-x_2,\eta=y-y_2}^{\xi=x-x_1,\eta=y-y_1},\\
\Delta Z^{(2)} &= \left[\frac{1}{2}\cos I\cos\theta\log\left(\frac{(\xi^2+\eta^2+\zeta^2)^{1/2}+\eta}{(\xi^2+\eta^2+\zeta^2)^{1/2}-\eta}\right)\right]_{\xi=x-x_2,\eta=y-y_2}^{\xi=x-x_1,\eta=y-y_1},\\
\Delta Z^{(3)} &= \left[\frac{1}{2}\cos I\sin\theta\log\left(\frac{(\xi^2+\eta^2+\zeta^2)^{1/2}+\xi}{(\xi^2+\eta^2+\zeta^2)^{1/2}-\xi}\right)\right]_{\xi=x-x_2,\eta=y-y_2}^{\xi=x-x_1,\eta=y-y_1},
\end{aligned}$$

where I is the inclination of the Earth's magnetic field and θ is the angle between the direction of magnetic north and the x axis. In our application $I = 65°$and $\theta = 0°$.

The difference between two simultaneous readings from two sensors is recorded. One sensor is mounted vertically above the other at a distance of 0.5m. Then the recorded reading originated by a prism identified by coordinates (x, y) is

$$h(x,y) = \Delta Z(x,y,z_{\text{U}}) - \Delta Z(x,y,z_{\text{L}}), \tag{S.6}$$

where z_{U} is the vertical coordinate of the upper sensor and z_{L} is that of the lower sensor, which are held at 0.7m and 0.2m above the surface throughout the survey.

Suppose that a vector $\boldsymbol{y}$ of n readings is recorded over a rectangular site at coordinates $\boldsymbol{s}_j$, $j = 1, \ldots, n$. Assume that the subsurface is divided into a rectangular assemblage of m prisms having coordinates $\boldsymbol{t}_i$, $i = 1, \ldots, m$, and producing susceptibilities that are collected in an $m \times 1$ vector $\boldsymbol{x}$. Then the influence of the prism i at surface location j is

$$K_{ij} = h(\boldsymbol{t}_i - \boldsymbol{s}_j) \tag{S.7}$$

where h is the function defined in (S.6). The linear relationship between $\boldsymbol{y}$ and $\boldsymbol{x}$ is then modeled as $E(\boldsymbol{y}) = \boldsymbol{K}\boldsymbol{x}$, where each element of the $n \times m$ matrix $\boldsymbol{K}$ is given by (S.7).

S.5.2. Model and VMP Implementation

As affirmed in Aykroyd et al. (2001), the assumptions used to model the hidden surface provide a realistic basis for modeling the data but can be quite restrictive. The recorded magnetic anomaly may include both positive and negative values, generate shifts in the apparent location of features, be asymmetric in shape and vary across different sampling regions. For this reason we model the outcome variable $\boldsymbol{y}$, i.e. the anomaly detected by the gradiometer, through a Skew Normal distribution and impose a Horseshoe penalization to the difference between hidden susceptibility values $\boldsymbol{x}_\Delta$. The Horseshoe prior shrinks the small signals and enhances the big ones, highlighting the contrast between buried features and plain soil.

The model we fit is the following:

$$\begin{gathered}
y_i|\boldsymbol{x},\sigma_\varepsilon^2,\lambda,c_i \overset{\text{ind.}}{\sim} N\left((\boldsymbol{K}\boldsymbol{x})_i + \tfrac{\sigma_\varepsilon^2\lambda|c_i|}{\sqrt{1+\lambda^2}}, \tfrac{\sigma_\varepsilon^2}{1+\lambda^2}\right), \quad c_i \overset{\text{ind.}}{\sim} N(0,1), \quad i = 1,\ldots,m,\\
(\boldsymbol{x}_\Delta)_j|b_j,\sigma_x^2 \overset{\text{ind.}}{\sim} N\left(0,\sigma_x^2/b_j\right), \quad b_j \overset{\text{ind.}}{\sim} \pi^{-1}b_j^{-1/2}(1+b_j)^{-1}, \quad j = 1,\ldots,d,\\
\sigma_\varepsilon^2 \sim \text{Inverse-}\chi^2\left(A_\varepsilon, B_\varepsilon\right), \quad \lambda \sim N\left(0, S_\lambda^2\right),\\
\sigma_x^2|a_x \sim \text{Inverse-}\chi^2\left(1, 1/a_x\right), \quad a_x \sim \text{Inverse-}\chi^2\left(1, 1/A_x^2\right).
\end{gathered} \tag{S.8}$$

The first line of this model gives rise to a Skew Normal likelihood and is equivalent to

$$y_i|\boldsymbol{x}, \sigma_\varepsilon^2, \lambda \overset{\text{ind.}}{\sim} \text{Skew-Normal}\left((\boldsymbol{K}\boldsymbol{x})_i, \sigma_\varepsilon^2, \lambda\right).$$

Here the density function of a random variable z having a Skew-Normal $\left(\mu, \sigma^2, \lambda\right)$ distribution is $p(z) = (2/\sigma)\phi\{(z-\mu)/\sigma\}\}\Phi\{\lambda(z-\mu)/\sigma\}$, with scale parameter $\sigma > 0$, skewness parameter λ, and ϕ and Φ denoting the Standard Normal density and cumulative distribution functions. The second line of (S.8) specifies a Horseshoe penalization on the $(\boldsymbol{x}_\Delta)_j$'s, as indicated by Table 1. The priors at line three give rise to conjugate message passing updates for the stochastic nodes of σ_ε and λ. As in model (2), σ_x is assigned a Half-Cauchy(A_x) prior through the auxiliary variable a_x.

The joint density function of model (S.8) is given by

$$\begin{aligned} p\left(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{c}, \boldsymbol{b}, \sigma_\varepsilon^2, \lambda, \sigma_x^2, a_x\right) = {} & p\left(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{c}, \sigma_\varepsilon^2, \lambda\right) p\left(\boldsymbol{x}|\boldsymbol{b}, \sigma_x^2\right) p\left(\boldsymbol{c}\right) p\left(\boldsymbol{b}\right) \\ & \times p\left(\sigma_\varepsilon^2\right) p\left(\lambda\right) p\left(\sigma_x^2|a_x\right) p\left(a_x\right). \end{aligned} \tag{S.9}$$

Steps similar to those presented for the base model can be used to implement VMP for the model with Skew Normal responses and Horseshoe prior. The starting point is again the choice of a factorization for the approximating q-densities. We impose the following mean field approximation to the joint posterior:

$$p\left(\boldsymbol{x}, \boldsymbol{c}, \boldsymbol{b}, \sigma_\varepsilon^2, \lambda, \sigma_x^2, a_x|\boldsymbol{y}\right) \approx q\left(\boldsymbol{x}\right) q\left(\sigma_\varepsilon^2\right) q\left(\lambda\right) q\left(\sigma_x^2\right) q\left(a_x\right) \prod_{i=1}^{m} q_i\left(c_i\right) \prod_{j=1}^{d} q_j\left(b_j\right). \tag{S.10}$$

The right-hand sides of (S.9) gives rise to the factor graph representation displayed as Figure S.5. In this

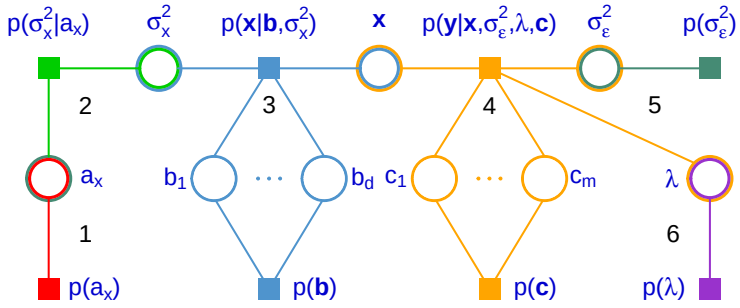

Figure S.5: *Factor graph representation of the Skew Normal response model with Horseshoe penalization in* (S.8)*, where the square nodes correspond to the density functions, or factors, on the right-hand side of* (S.9)*. The circular nodes correspond to stochastic nodes of the q-density factorization in* (S.10)*. Numbers are used to show the distinction between fragments, whereas colors identify different fragments types.*

factor graph, four of the seven fragments arising from the base model (2), those numbered 1–3, are preserved. Fragment 4 is now the *Skew Normal likelihood fragment* studied in Section 3.4 of McLean and Wand (2019). The message passed from this fragment to σ_ε^2 takes the form

$$m_{p(\boldsymbol{y}|\boldsymbol{x},\sigma_\varepsilon^2,\lambda,\boldsymbol{c})\to\sigma_\varepsilon^2}\left(\sigma_\varepsilon^2\right) = \exp\left\{ \begin{bmatrix} \log(\sigma_\varepsilon^2) \\ 1/\sigma_\varepsilon \\ 1/\sigma_\varepsilon^2 \end{bmatrix}^T \boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{x},\sigma_\varepsilon^2,\lambda,\boldsymbol{c})\to\sigma_\varepsilon^2} \right\} \tag{S.11}$$

and is proportional to an *Inverse Square Root Nadarajah* density function, whereas the message to λ is

$$m_{p(\boldsymbol{y}|\boldsymbol{x},\sigma_\varepsilon^2,\lambda,\boldsymbol{c})\to\lambda}(\lambda) = \exp\left\{ \begin{bmatrix} \log(1+\lambda^2) \\ \lambda^2 \\ \lambda\sqrt{1+\lambda^2} \end{bmatrix}^T \boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{x},\sigma_\varepsilon^2,\lambda,\boldsymbol{c})\to\lambda} \right\} \tag{S.12}$$

and is part of the *Sea Sponge* family. These two families of distributions are defined in Sections S.2.3 and S.2.5 of the supplementary material of McLean and Wand (2019). Priors on σ_ε^2 and λ that are conjugate to these messages must have the form

$$m_{p(\sigma_\varepsilon^2)\to\sigma_\varepsilon^2}(\sigma_\varepsilon^2) = \exp\left\{ \begin{bmatrix} \log(\sigma_\varepsilon^2) \\ 1/\sigma_\varepsilon \\ 1/\sigma_\varepsilon^2 \end{bmatrix}^T \boldsymbol{\eta}_{p(\sigma_\varepsilon^2)\to\sigma_\varepsilon^2} \right\} \tag{S.13}$$

$$\text{and } m_{p(\lambda)\to\lambda}(\lambda) = \exp\left\{ \begin{bmatrix} \log(1+\lambda^2) \\ \lambda^2 \\ \lambda\sqrt{1+\lambda^2} \end{bmatrix}^T \boldsymbol{\eta}_{p(\lambda)\to\lambda} \right\}, \tag{S.14}$$

for some 3×1 vectors $\boldsymbol{\eta}_{p(\sigma_\varepsilon^2)\to\sigma_\varepsilon^2}$ and $\boldsymbol{\eta}_{p(\lambda)\to\lambda}$. Priors $\sigma_\varepsilon^2 \sim$ Inverse-$\chi^2\left(A_\varepsilon, B_\varepsilon\right)$ and $\lambda \sim N\left(0, S_\lambda^2\right)$ of model (S.8) are respectively conjugate to (S.11) and (S.12), since for this choice of priors the messages $m_{p(\sigma_\varepsilon^2)}(\sigma_\varepsilon^2)$ and $m_{p(\lambda)\to\lambda}(\lambda)$ can be written as (S.13) and (S.14) with

$$\boldsymbol{\eta}_{p(\sigma_\varepsilon^2)\to\sigma_\varepsilon^2} = [-(A_\varepsilon/2)-1 \;\; 0 \;\; -B_\varepsilon/2]^T \text{ and } \boldsymbol{\eta}_{p(\lambda)\to\lambda} = [0 \;\; -1/(2S_\lambda^2) \;\; 0]^T.$$

We impose diffuse priors using $A_\varepsilon = B_\varepsilon = 10^{-2}$ and $S_\lambda = 10^5$.

Full implementation of VMP is based on iteratively updating, for each factor graph fragment depicted in Figure S.5: (i) the parameter vectors of messages passed from the fragment's neighboring stochastic nodes to the fragment's factors; (ii) the parameter vectors of the messages passed from the fragment's factors to their neighboring stochastic nodes. The first step is very simple and entails application of (7) of Wand (2017). The factor to stochastic node updates of the second step can be performed through various VMP procedures:

- Fragment 1 is an Inverse G-Wishart prior fragment and the factor to stochastic node parameter vector updates can be performed using Algorithm 1 of Maestrini and Wand (2021) with inputs $G_\Theta = G_{\text{full}}$, $\xi_\Theta = 1$ and $\boldsymbol{\Lambda}_\Theta = (A_x^2)^{-1}$.

- Fragment 2 is an iterated Inverse G-Wishart prior fragment and the factor to stochastic node parameter vector updates can be performed using Algorithm 2 of Maestrini and Wand (2021) with inputs $G = G_{\text{full}}$, $\xi = 1$ and $G_{\boldsymbol{A}\to p(\boldsymbol{\Sigma}|\boldsymbol{A})} = G_{\text{diag}}$.

- The factor to stochastic node parameter vector updates of fragment 3 can be performed through Algorithm 2.

- Fragment 4 is the Skew Normal likelihood fragment and the factor to stochastic node parameter vector updates can be performed using Algorithm 4 of McLean and Wand (2019) with $\boldsymbol{y}$ and $\boldsymbol{K}$ as data inputs.

- Fragment 5 corresponds to the imposition of an Inverse Square Root Nadarajah prior distribution on the variance parameter σ_ε^2. The output of VMP applied to this fragment is the natural parameter vector of the prior density function, that is, $\boldsymbol{\eta}_{p(\sigma_\varepsilon^2)\to\sigma_\varepsilon^2}$ from (S.13).

- Fragment 6 corresponds to the imposition of a Sea Sponge prior distribution on the skewness parameter λ. The output of VMP applied to this fragment is the natural parameter vector of the prior density function, that is, $\boldsymbol{\eta}_{p(\lambda)\to\lambda}$ from (S.14).

We restrict our attention to the portion of the archaeological dataset corresponding to the excavated area, which enables a qualitative performance assessment of our fitting method. Figure S.6 displays the data under examination ($\boldsymbol{Y}$) together with the results of the application of VMP to model (S.8) and the impression drawn by archaeologists. The data were handled in a disjoint way by separating the two rectangular areas corresponding to indices J and K of the archaeologists' impression. It is standard practice to examine grids as soon as they are collected and this division of the surveyed area allows to partition the data into two full matrices $\boldsymbol{Y}_J$ for area J and $\boldsymbol{Y}_K$ for area K of size 30×20 and 40×20, respectively. Model (S.8) can be used setting $\boldsymbol{y} = \text{vec}(\boldsymbol{Y}_J)$ or $\boldsymbol{y} = \text{vec}(\boldsymbol{Y}_K)$ and the reconstruction $\widehat{\boldsymbol{X}}$ is simply obtained as the inverse vectorization of the VMP estimate of $\boldsymbol{x}$. We employ the mean of the optimal approximating density $q^*(\boldsymbol{x})$, which is a $N\big(\boldsymbol{\mu}_{q(\boldsymbol{x})}, \boldsymbol{\Sigma}_{q(\boldsymbol{x})}\big)$ density function, to estimate $\boldsymbol{x}$. Expressions for $q^*(\boldsymbol{x})$ and the other q-densities of interest are provided in the supplement.

If compared to the original dataset, the VMP reconstruction of Figure S.6 shows greater contrast between background and features, and some weak features are more evident in the posterior mean reconstruction. Despite the data being treated in a disjoint way, discontinuity in the estimate of $\boldsymbol{X}$ between areas J and K is not very apparent. Careful inspection of the reconstruction shows the locations of some reconstructed features are shifted if compared to their apparent position in the original survey data image. This is important information, considering that each pixel corresponds to a square area of 0.5m side and that archaeological excavations require intensive manual work. The approximate posterior densities of λ seem to indicate that the data from area J is symmetric, whereas that from area K is positively skewed.

S.5.3. Approximating density functions

From (10) of Wand (2017), the densities of main interest, $q^*(\boldsymbol{x})$, $q^*(\sigma_\varepsilon^2)$, $q^*(\lambda)$ and $q^*(\sigma_x^2)$, have the following forms:

$$q^*(\boldsymbol{x}) \propto \exp\left\{ \begin{bmatrix} \boldsymbol{x} \\ \text{vec}\left(\boldsymbol{x}\boldsymbol{x}^T\right) \end{bmatrix}^T \boldsymbol{\eta}_{q(\boldsymbol{x})} \right\}, \text{ with } \boldsymbol{\eta}_{q(\boldsymbol{x})} \equiv \boldsymbol{\eta}_{p(\boldsymbol{x}|\boldsymbol{b},\,\sigma_x^2)\,\to\,\boldsymbol{x}} + \boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{x},\sigma_\varepsilon^2,\lambda,\boldsymbol{c})\to\boldsymbol{x}},$$

$$q^*(\sigma_\varepsilon^2) \propto \exp\left\{ \begin{bmatrix} \log(\sigma_\varepsilon^2) \\ 1/\sigma_\varepsilon \\ 1/\sigma_\varepsilon^2 \end{bmatrix}^T \boldsymbol{\eta}_{q(\sigma_\epsilon^2)} \right\}, \text{ with } \boldsymbol{\eta}_{q(\sigma_\epsilon^2)} \equiv \boldsymbol{\eta}_{p(\sigma_\varepsilon^2)} + \boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{x},\sigma_\varepsilon^2,\lambda,\boldsymbol{c})\to\sigma_\varepsilon^2},$$

$$q^*(\lambda) \propto \exp\left\{ \begin{bmatrix} \log(1+\lambda^2) \\ \lambda^2 \\ \lambda\sqrt{1+\lambda^2} \end{bmatrix}^T \boldsymbol{\eta}_{q(\lambda)} \right\}, \text{ with } \boldsymbol{\eta}_{q(\lambda)} \equiv \boldsymbol{\eta}_{p(\lambda)} + \boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{x},\sigma_\varepsilon^2,\lambda,\boldsymbol{c})\to\lambda}$$

$$\text{and } q^*(\sigma_x^2) \propto \exp\left\{ \begin{bmatrix} \log(\sigma_x^2) \\ 1/\sigma_x^2 \end{bmatrix}^T \boldsymbol{\eta}_{q(\sigma_x^2)} \right\}, \text{ with } \boldsymbol{\eta}_{q(\sigma_x^2)} \equiv \boldsymbol{\eta}_{p(\sigma_x^2|a_x)\,\to\,\sigma_x^2} + \boldsymbol{\eta}_{p(\boldsymbol{x}|\,\boldsymbol{b},\,\sigma_x^2)\,\to\,\sigma_x^2}.$$

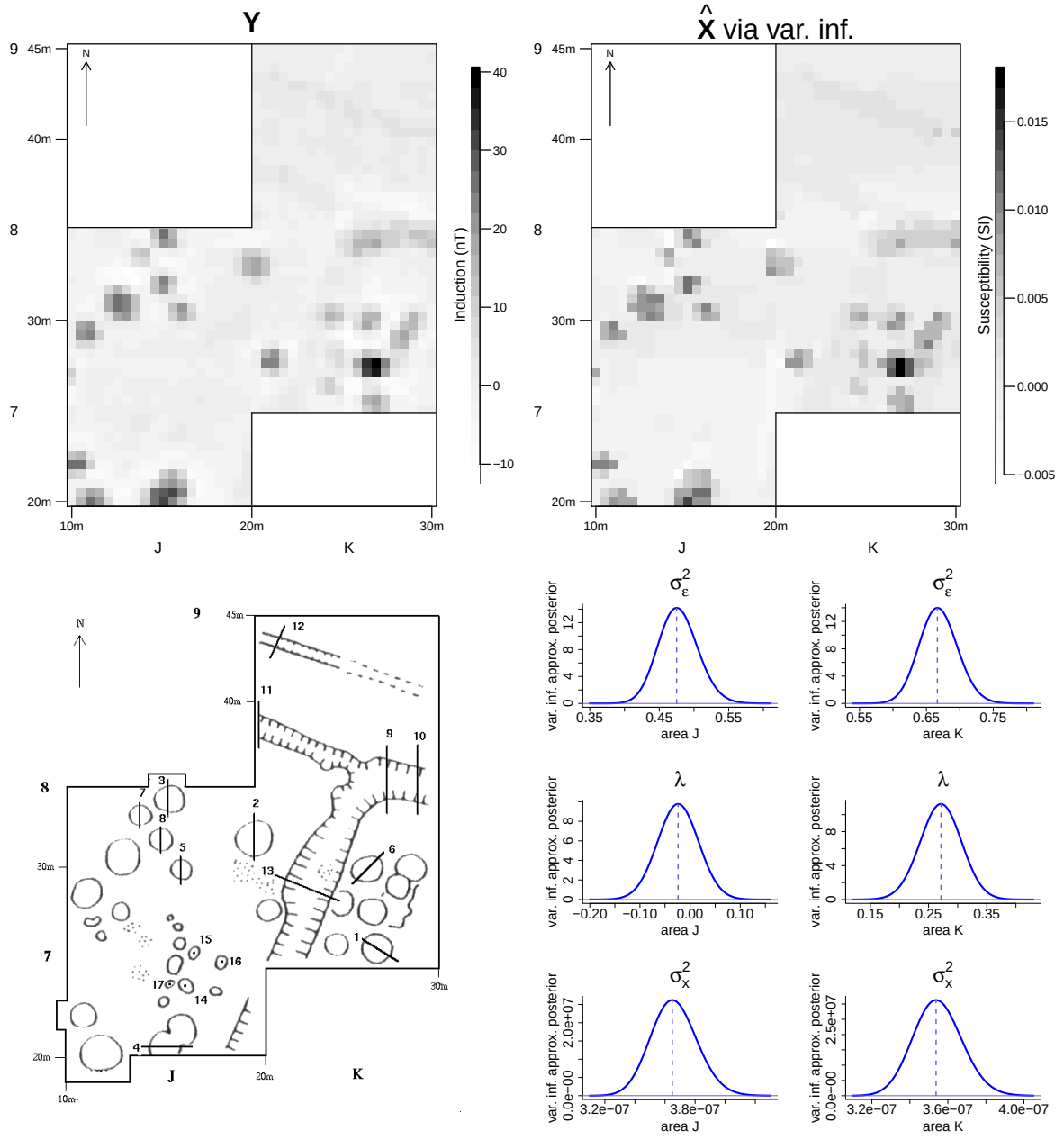

Figure S.6: *Results of the archaeological data study conducted via model* (S.8) *and variational inference. The top-left image displays the archaeological dataset denoted by* $\boldsymbol{Y}$*. Its variational inference reconstruction, denoted by* $\widehat{\boldsymbol{X}}$*, is shown in the top-right image, whereas the bottom-left image corresponds to the archaeologist's impression of the 1994 excavation of 'The Park'. The plots in the bottom-right side show the variational approximate marginal posterior densities of the Skew Normal variance and skewness parameters for areas J and K of the archaeological site.*

Expressions for the other optimal approximating densities can be obtained in a similar manner. The full list of optimal approximating density functions respecting restriction (S.10) is the following:

$$
\begin{aligned}
&q^*(\boldsymbol{x}) \text{ is a } N\big(\boldsymbol{\mu}_{q(\boldsymbol{x})}, \boldsymbol{\Sigma}_{q(\boldsymbol{x})}\big) \text{ density function,} \\
&q^*(c_i) \propto \exp\left\{ \begin{bmatrix} |c_i| \\ c_i^2 \end{bmatrix}^T \boldsymbol{\eta}_{c_i} \right\}, \text{ for a } 2\times 1 \text{ natural paramter vector } \boldsymbol{\eta}_{c_i} \text{ and } i = 1, \ldots, m, \\
&q^*(b_j) \text{ is an Inverse-Gaussian}\big(\mu_{q(b_j)}, \lambda_{q(b_j)}\big) \text{ density function, for } j = 1, \ldots, d, \\
&q^*(\sigma_\varepsilon^2) \text{ is an Inverse-Square-Root-Nadarajah } \big(\alpha_{q(\sigma_\varepsilon^2)}, \beta_{q(\sigma_\varepsilon^2)}, \gamma_{q(\sigma_\varepsilon^2)}\big) \text{ density function,} \\
&q^*(\lambda) \text{ is a Sea-Sponge } \big(\alpha_{q(\lambda)}, \beta_{q(\lambda)}, \gamma_{q(\lambda)}\big) \text{ density function,} \\
&q^*(\sigma_x^2) \text{ is an Inverse-}\chi^2\big(\kappa_{q(\sigma_x^2)}, \lambda_{q(\sigma_x^2)}\big) \text{ density function} \\
\text{and } &q^*(a_x) \text{ is an Inverse-}\chi^2\big(\kappa_{q(a_x)}, \lambda_{q(a_x)}\big) \text{ density function.}
\end{aligned}
$$

The common parameters of the q-densities of interest are:

$$
\begin{gathered}
\boldsymbol{\mu}_{q(\boldsymbol{x})} = \boldsymbol{\Sigma}_{q(\boldsymbol{x})} \left(\boldsymbol{\eta}_{q(\boldsymbol{x})}\right)_{1:m}, \quad \boldsymbol{\Sigma}_{q(\boldsymbol{x})} = -\tfrac{1}{2} \operatorname{vec}^{-1}\left\{ \left(\boldsymbol{\eta}_{q(\boldsymbol{x})}\right)_{(m+1):m^2} \right\}, \\
\alpha_{q(\sigma_\varepsilon^2)} = -2\left(1 + \left(\boldsymbol{\eta}_{q(\sigma_\epsilon^2)}\right)_1\right), \quad \beta_{q(\sigma_\varepsilon^2)} = -\left(\boldsymbol{\eta}_{q(\sigma_\epsilon^2)}\right)_3, \quad \gamma_{q(\sigma_\varepsilon^2)} = -\left(\boldsymbol{\eta}_{q(\sigma_\epsilon^2)}\right)_2, \\
\alpha_{q(\lambda)} = \left(\boldsymbol{\eta}_{q(\lambda)}\right)_1, \quad \beta_{q(\lambda)} = -\left(\boldsymbol{\eta}_{q(\lambda)}\right)_2, \quad \gamma_{q(\lambda)} = \left(\boldsymbol{\eta}_{q(\lambda)}\right)_3, \\
\kappa_{q(\sigma_x^2)} = -2\left(1 + \left(\boldsymbol{\eta}_{q(\sigma_x^2)}\right)_1\right), \quad \lambda_{q(\sigma_x^2)} = -2\left(\boldsymbol{\eta}_{q(\sigma_x^2)}\right)_2.
\end{gathered}
$$

To obtain the common parameters of $q(\sigma_\varepsilon^2)$ and $q(\sigma_\lambda^2)$ from their natural parameters we make use of the results from Sections S.2.3 and S.2.5 of the supplementary material of McLean and Wand (2019) concerning the Inverse Square Root Nadarajah and Sea Sponge distributions.