
Supplement to *Efficient methods for Gaussian Markov random fields under sparse linear constraints*

David Bolin
 King Abdullah University of
 Science and Technology
 david.bolin@kaust.edu.sa

Jonas Wallin
 Department of Statistics,
 Lund University
 jonas.wallin@stat.lu.se

In this document we provide further details of the constraint basis construction as well as additional simulation results are given in Section A. We then provide the proofs of the four main theorems of the paper in Section B. Finally, Section C provides details about the conditional constrained distribution $\pi(\mathbf{x}|\mathbf{A}\mathbf{x} = \mathbf{b})$.

A Further details of the basis construction and simulations

In Algorithm 2 it was not explained how the reordering of the \mathbf{A} matrix is done. This is illustrated in Algorithm 6 where we show how to build the sub-matrices $\{\tilde{\mathbf{A}}\}_{k=1}^m$.

Algorithm 6 Find all non-overlapping sub-matrices

Require: \mathbf{A} (a $k \times n$ matrix)
 1: $\{\tilde{\mathbf{A}}_1, \mathbf{B}\} \leftarrow \text{overlap}(\mathbf{A})$
 2: $m \leftarrow 1$
 3: **while** $\mathbf{B} \neq \emptyset$ **do**
 4: $m \leftarrow m + 1$
 5: $\{\tilde{\mathbf{A}}_m, \mathbf{B}\} \leftarrow \text{overlap}(\mathbf{B})$
 6: **end while**
 7: Return $\{\tilde{\mathbf{A}}\}_{k=1}^m$

Algorithm 7 $\text{overlap}(\mathbf{A})$ Find first sub-matrix

Require: \mathbf{A} (a $k \times n$ matrix)
 1: $U \leftarrow \{1\}$
 2: $d \leftarrow 0$
 3: $D \leftarrow \text{id}(\mathbf{A}_{U,\cdot})$
 4: **while** 1 **do**
 5: $D \leftarrow \text{id}(\mathbf{A}_{U,\cdot})$
 6: $U \leftarrow \text{id}\left(\left(\mathbf{A}_{\cdot,D}\right)^\top\right)$
 7: **if** $d = |U|$ **then**
 8: **break**
 9: **end if**
 10: $d \leftarrow |U|$
 11: **end while**
 12: $\tilde{\mathbf{A}} = \mathbf{A}_{\cdot,U}$
 13: $\tilde{\mathbf{A}}^c = \mathbf{A}_{\cdot,U^c}$
 14: Return $\{\tilde{\mathbf{A}}, \tilde{\mathbf{A}}^c\}$

In Section 6.1 in the main article, we provided results for $\alpha = 2$ and $\alpha = 4$. In Figure 1 the corresponding results are provided for $\alpha = 1$ and $\alpha = 3$.

B Proofs

In this section we prove the four main theorems of the paper.

Proof of Theorem 1. We first transform the density of $\mathbf{A}\mathbf{X}$ to the basis represented by \mathbf{T} ,

$$\pi_{\mathbf{A}\mathbf{X}}(\mathbf{b}) = \pi_{\mathbf{A}\mathbf{T}^\top\mathbf{X}^*}(\mathbf{b}) = \pi_{\mathbf{X}_{\mathcal{C}}^*}(\mathbf{H}^{-1}\mathbf{b}) \left| |\mathbf{H}|^{-1} \right| = \pi_{\mathbf{X}_{\mathcal{C}}^*}(\mathbf{H}^{-1}\mathbf{b}) |\mathbf{A}\mathbf{A}^\top|^{-1/2}.$$

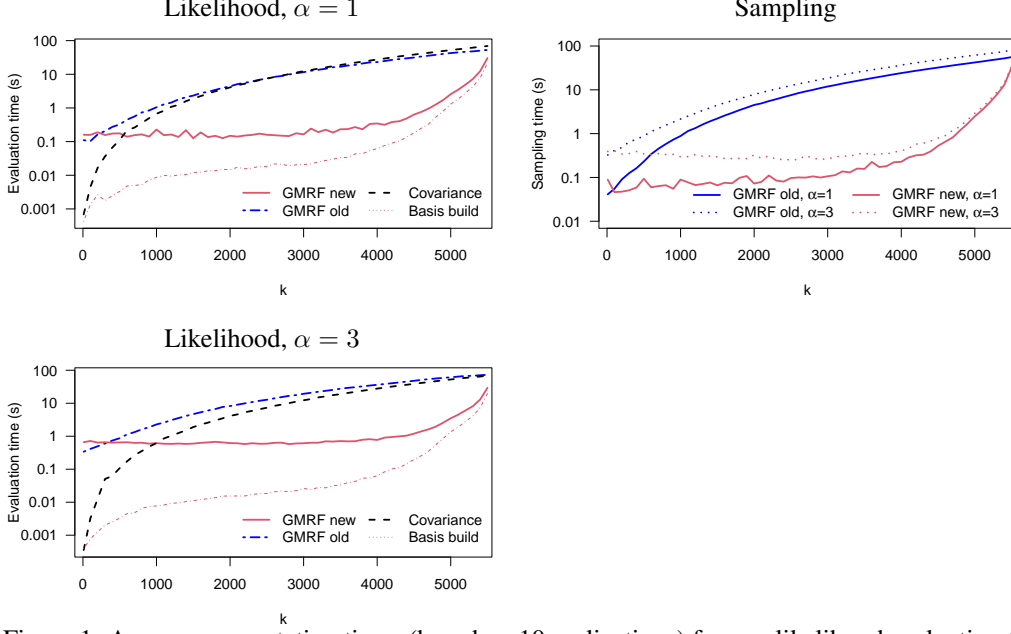


Figure 1: Average computation times (based on 10 replications) for one likelihood evaluation (left) and one sample from $\mathbf{X}|\mathbf{A}_Y\mathbf{X} = \mathbf{y}$ (top right) of the Matérn model with $\alpha = 1$ and $\alpha = 3$ as functions of the number of observations k . The computation times for the new GMRF method includes the time needed to construct the basis matrix \mathbf{T} , which also is shown separately.

In order to derive the density $\mathbf{X}_{\mathcal{L}}^*$, note that the density of \mathbf{X}^* is

$$\pi_{\mathbf{X}_{\mathcal{L}}^*}(\mathbf{x}^*) = \frac{|\mathbf{Q}|^{\dagger/2}}{(2\pi)^{\frac{n-s}{2}}} \exp\left(-\frac{1}{2}Q(\mathbf{x}^*)\right),$$

where the quadratic form $Q(\mathbf{x}^*)$ is

$$\begin{aligned} Q(\mathbf{x}^*) &= \begin{bmatrix} \mathbf{x}_{\mathcal{U}}^* - \boldsymbol{\mu}_{\mathcal{U}}^* \\ \mathbf{x}_{\mathcal{L}}^* - \boldsymbol{\mu}_{\mathcal{L}}^* \end{bmatrix}^{\top} \begin{bmatrix} \mathbf{Q}_{\mathcal{U}\mathcal{U}}^* & \mathbf{Q}_{\mathcal{L}\mathcal{U}}^* \\ \mathbf{Q}_{\mathcal{U}\mathcal{L}}^* & \mathbf{Q}_{\mathcal{L}\mathcal{L}}^* \end{bmatrix} \begin{bmatrix} \mathbf{x}_{\mathcal{U}}^* - \boldsymbol{\mu}_{\mathcal{U}}^* \\ \mathbf{x}_{\mathcal{L}}^* - \boldsymbol{\mu}_{\mathcal{L}}^* \end{bmatrix} \\ &= (\mathbf{x}_{\mathcal{L}}^* - \boldsymbol{\mu}_{\mathcal{L}}^*)^{\top} \mathbf{Q}_{\mathcal{L}\mathcal{L}}^* (\mathbf{x}_{\mathcal{L}}^* - \boldsymbol{\mu}_{\mathcal{L}}^*) + (\mathbf{x}_{\mathcal{L}}^* - \boldsymbol{\mu}_{\mathcal{L}}^*)^{\top} \mathbf{Q}_{\mathcal{L}\mathcal{U}}^* (\mathbf{x}_{\mathcal{U}}^* - \boldsymbol{\mu}_{\mathcal{U}}^*) \\ &\quad + (\mathbf{x}_{\mathcal{U}}^* - \boldsymbol{\mu}_{\mathcal{U}}^*)^{\top} \mathbf{Q}_{\mathcal{U}\mathcal{L}}^* (\mathbf{x}_{\mathcal{L}}^* - \boldsymbol{\mu}_{\mathcal{L}}^*) + (\mathbf{x}_{\mathcal{U}}^* - \boldsymbol{\mu}_{\mathcal{U}}^*)^{\top} \mathbf{Q}_{\mathcal{U}\mathcal{U}}^* (\mathbf{x}_{\mathcal{U}}^* - \boldsymbol{\mu}_{\mathcal{U}}^*) \\ &= (\mathbf{x}_{\mathcal{L}}^* - \boldsymbol{\mu}_{\mathcal{L}}^*)^{\top} \mathbf{Q}_{\mathcal{L}\mathcal{L}}^* (\mathbf{x}_{\mathcal{L}}^* - \boldsymbol{\mu}_{\mathcal{L}}^*) - (\mathbf{x}_{\mathcal{L}}^* - \boldsymbol{\mu}_{\mathcal{L}}^*)^{\top} \mathbf{Q}_{\mathcal{L}\mathcal{U}}^* \mathbf{Q}_{\mathcal{U}\mathcal{U}}^{*\dagger} \mathbf{Q}_{\mathcal{U}\mathcal{L}}^* (\mathbf{x}_{\mathcal{L}}^* - \boldsymbol{\mu}_{\mathcal{L}}^*) + \\ &\quad + \left(\mathbf{x}_{\mathcal{U}}^* - \boldsymbol{\mu}_{\mathcal{U}}^* + \mathbf{Q}_{\mathcal{U}\mathcal{U}}^{*\dagger} \mathbf{Q}_{\mathcal{U}\mathcal{L}}^* (\mathbf{x}_{\mathcal{L}}^* - \boldsymbol{\mu}_{\mathcal{L}}^*) \right)^{\top} \mathbf{Q}_{\mathcal{U}\mathcal{U}}^* \left(\mathbf{x}_{\mathcal{U}}^* - \boldsymbol{\mu}_{\mathcal{U}}^* + \mathbf{Q}_{\mathcal{U}\mathcal{U}}^{*\dagger} \mathbf{Q}_{\mathcal{U}\mathcal{L}}^* (\mathbf{x}_{\mathcal{L}}^* - \boldsymbol{\mu}_{\mathcal{L}}^*) \right). \end{aligned}$$

Here we in the last step wrote the expression so that we easily can integrate out $\mathbf{X}_{\mathcal{U}}^*$ on the complement to the null space of $\mathbf{Q}_{\mathcal{U}\mathcal{U}}^*$. Doing so yields the desired result,

$$\pi(\mathbf{x}_{\mathcal{L}}^*) = \int \pi_{\mathbf{X}_{\mathcal{L}}^*}(\mathbf{x}^*) d\mathbf{x}_{\mathcal{U}}^* \propto \frac{|\mathbf{Q}|^{\dagger/2}}{|\mathbf{Q}_{\mathcal{U}\mathcal{U}}^*|^{\dagger/2}} \exp\left(-\frac{1}{2}\hat{Q}(\mathbf{x}_{\mathcal{L}}^*)\right),$$

where

$$\hat{Q}(\mathbf{x}_{\mathcal{L}}^*) = (\mathbf{x}_{\mathcal{L}}^* - \boldsymbol{\mu}_{\mathcal{L}}^*)^{\top} \left(\mathbf{Q}_{\mathcal{L}\mathcal{L}}^* - \mathbf{Q}_{\mathcal{L}\mathcal{U}}^* \mathbf{Q}_{\mathcal{U}\mathcal{U}}^{*\dagger} \mathbf{Q}_{\mathcal{U}\mathcal{L}}^* \right) (\mathbf{x}_{\mathcal{L}}^* - \boldsymbol{\mu}_{\mathcal{L}}^*).$$

□

To prove Theorem 2 we need the following lemma.

Lemma 1. Under Assumption 1 one has $\text{rank}(\mathbf{Q}_{\mathcal{L}\mathcal{L}}^*) = k - k_0$ and $\text{rank}(\mathbf{Q}_{\mathcal{U}\mathcal{U}}^*) = n - s - (k - k_0)$.

Proof. We have $\text{rank}(\mathbf{Q}^*) = \text{rank}(\mathbf{Q}) = n - s$ since \mathbf{T} is orthonormal matrix. Further, using the eigen-decomposition of \mathbf{Q} we can express \mathbf{Q}^* as

$$\mathbf{Q}^* = \begin{bmatrix} \mathbf{T}_{\mathcal{L}} \\ \mathbf{T}_{\mathcal{U}} \end{bmatrix} \begin{bmatrix} \mathbf{E}_{0^c} \\ \mathbf{E}_0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{E}_{0^c} \\ \mathbf{E}_0 \end{bmatrix}^{\top} \begin{bmatrix} \mathbf{T}_{\mathcal{L}} \\ \mathbf{T}_{\mathcal{U}} \end{bmatrix}^{\top},$$

where Λ is a diagonal matrix with the non-zero eigenvalues of \mathbf{Q} . Since $\text{rank}(\mathbf{A}\mathbf{E}_0) = k_0$ it follows that also $\text{rank}(\mathbf{T}_\ell\mathbf{E}_0) = k_0$ and $\text{rank}(\mathbf{T}_{\mathcal{U}}\mathbf{E}_0) = s - k_0$. By Theorem 4.3.28 of [HJ13] there exists an eigenvector, \mathbf{e} , of $\mathbf{Q}_{\mathcal{U}\mathcal{U}}^*$ that has a corresponding eigenvalue 0 if and only if $\mathbf{Q}_{\mathcal{U}\mathcal{U}}^*\mathbf{e} = \mathbf{0}$ and $\mathbf{Q}_{\ell\mathcal{U}}^*\mathbf{e} = \mathbf{0}$. By construction, any vector constructed by the linear span of $\mathbf{T}_{\mathcal{U}}\mathbf{E}_0$ satisfies this requirement, and no other vector does. Hence, the rank of $\mathbf{Q}_{\mathcal{U}\mathcal{U}}^*$ is $n - k - (s - k_0)$ and the rank of $\mathbf{Q}_{\ell\mathcal{U}}^*$ is $k - k_0$. \square

Proof of Theorem 2. To derive the distribution we note that the conditional distribution of $\mathbf{X}_{\mathcal{U}}^*|\mathbf{X}_\ell^*$ is proportional to $\exp(-\frac{1}{2}Q(\mathbf{x}^*))$, where

$$\begin{aligned} Q(\mathbf{x}^*) &= \begin{bmatrix} \mathbf{x}_{\mathcal{U}}^* - \boldsymbol{\mu}_{\mathcal{U}}^* \\ \mathbf{x}_\ell^* - \boldsymbol{\mu}_\ell^* \end{bmatrix}^\top \begin{bmatrix} \mathbf{Q}_{\mathcal{U}\mathcal{U}}^* & \mathbf{Q}_{\ell\mathcal{U}}^* \\ \mathbf{Q}_{\ell\mathcal{U}}^{*\top} & \mathbf{Q}_{\ell\ell}^* \end{bmatrix} \begin{bmatrix} \mathbf{x}_{\mathcal{U}}^* - \boldsymbol{\mu}_{\mathcal{U}}^* \\ \mathbf{x}_\ell^* - \boldsymbol{\mu}_\ell^* \end{bmatrix} \\ &= (\mathbf{x}_{\mathcal{U}}^* - \boldsymbol{\mu}_{\mathcal{U}}^*)^\top \mathbf{Q}_{\mathcal{U}\mathcal{U}}^* (\mathbf{x}_{\mathcal{U}}^* - \boldsymbol{\mu}_{\mathcal{U}}^*) + 2(\mathbf{x}_{\mathcal{U}}^* - \boldsymbol{\mu}_{\mathcal{U}}^*)^\top \mathbf{Q}_{\ell\mathcal{U}}^* (\mathbf{x}_\ell^* - \boldsymbol{\mu}_\ell^*) + C, \end{aligned}$$

where C is a constant independent of $\mathbf{x}_{\mathcal{U}}^*$. Now, since $\mathbf{Q}_{\mathcal{U}\mathcal{U}}^* (\mathbf{Q}_{\ell\mathcal{U}}^*)^\dagger \mathbf{Q}_{\ell\mathcal{U}}^* = \mathbf{Q}_{\ell\mathcal{U}\ell}^*$, the quadratic form $Q(\mathbf{x}^*)$ can be written as a constant plus $\mathbf{v}^\top \mathbf{Q}_{\ell\mathcal{U}\ell}^* \mathbf{v}$, where

$$\mathbf{v} = \mathbf{x}_{\mathcal{U}}^* - \boldsymbol{\mu}_{\mathcal{U}}^* + \mathbf{Q}_{\ell\mathcal{U}\ell}^{*\dagger} \mathbf{Q}_{\ell\mathcal{U}}^* (\mathbf{x}_\ell^* - \boldsymbol{\mu}_\ell^*).$$

Hence

$$\mathbf{X}_{\mathcal{U}}^*|\mathbf{X}_\ell^* = \mathbf{b}^* \sim \mathcal{N}_C \left(\mathbf{Q}_{\ell\mathcal{U}\ell}^* \left(\boldsymbol{\mu}_{\mathcal{U}}^* - \mathbf{Q}_{\ell\mathcal{U}\ell}^{*\dagger} \mathbf{Q}_{\ell\mathcal{U}}^* (\mathbf{b}^* - \boldsymbol{\mu}_\ell^*) \right), \mathbf{Q}_{\ell\mathcal{U}\ell}^* \right). \quad (10)$$

Combining this with the fact that $\mathbf{X} = \mathbf{T}^\top \mathbf{X}^*$ gives the desired expression for the distribution of $\mathbf{X}|\mathbf{A}\mathbf{X} = \mathbf{b}$. Finally, since \mathbf{T} is orthonormal, the rank of $\mathbf{Q}_{\mathbf{X}|\mathbf{b}}$ is the same as the rank of $\mathbf{Q}_{\mathcal{U}\mathcal{U}}$, and the result follows from Lemma 1. \square

We start by proving Theorem 4 as we will use result from that proof in the proof of Theorem 3.

Proof of Theorem 4. First, note that the density

$$\pi_{\mathbf{Y}|\mathbf{X}_\ell^*, \mathbf{X}_{\mathcal{U}}^*}(\mathbf{y}|\mathbf{x}_{\mathcal{U}}^*, \mathbf{b}^*) = \frac{1}{(2\pi)^{\frac{m}{2}} \sigma_Y^m} \exp \left(-\frac{1}{2\sigma_Y^2} \left(\mathbf{y} - \mathbf{B}^* \begin{bmatrix} \mathbf{b}^* \\ \mathbf{x}_{\mathcal{U}}^* \end{bmatrix} \right) \left(\mathbf{y} - \mathbf{B}^* \begin{bmatrix} \mathbf{b}^* \\ \mathbf{x}_{\mathcal{U}}^* \end{bmatrix} \right)^\top \right), \quad (11)$$

can, as a function of $\mathbf{x}_{\mathcal{U}}^*$, be written as

$$\pi_{\mathbf{Y}|\mathbf{X}_\ell^*, \mathbf{X}_{\mathcal{U}}^*}(\mathbf{y}|\mathbf{b}^*, \mathbf{x}_{\mathcal{U}}^*) \propto \exp \left(-\frac{\mathbf{x}_{\mathcal{U}}^{*\top} \mathbf{B}_{\mathcal{U}\mathcal{U}}^{*\top} \mathbf{B}_{\mathcal{U}\mathcal{U}}^* \mathbf{x}_{\mathcal{U}}^*}{2\sigma_Y^2} + \frac{\mathbf{y}^{*\top} \mathbf{B}_{\mathcal{U}\mathcal{U}}^* \mathbf{x}_{\mathcal{U}}^*}{\sigma_Y^2} \right).$$

Further, from (10), we have that, as a function of $\mathbf{x}_{\mathcal{U}}^*$,

$$\pi_{\mathbf{X}_{\mathcal{U}}^*|\mathbf{X}_\ell^*}(\mathbf{x}_{\mathcal{U}}^*|\mathbf{b}^*) \propto \exp \left(-\frac{1}{2} (\mathbf{x}_{\mathcal{U}}^* - \tilde{\boldsymbol{\mu}}_{\mathcal{U}}^*)^\top \mathbf{Q}_{\mathcal{U}\mathcal{U}}^* (\mathbf{x}_{\mathcal{U}}^* - \tilde{\boldsymbol{\mu}}_{\mathcal{U}}^*) \right),$$

where $\tilde{\boldsymbol{\mu}}_{\mathcal{U}}^* = \boldsymbol{\mu}_{\mathcal{U}}^* - \mathbf{Q}_{\ell\mathcal{U}\ell}^{*\dagger} \mathbf{Q}_{\ell\mathcal{U}}^* (\mathbf{b}^* - \boldsymbol{\mu}_\ell^*)$. Since $\pi_{\mathbf{X}_{\mathcal{U}}^*|\mathbf{Y}, \mathbf{X}_\ell^*}(\mathbf{x}_{\mathcal{U}}^*|\mathbf{y}, \mathbf{b}^*)$ is proportional to $\pi_{\mathbf{Y}|\mathbf{X}_\ell^*, \mathbf{X}_{\mathcal{U}}^*}(\mathbf{y}|\mathbf{b}^*, \mathbf{x}_{\mathcal{U}}^*)\pi(\mathbf{x}_{\mathcal{U}}^*|\mathbf{b}^*)$, it follows that

$$\begin{aligned} \pi_{\mathbf{X}_{\mathcal{U}}^*|\mathbf{Y}, \mathbf{X}_\ell^*}(\mathbf{x}_{\mathcal{U}}^*|\mathbf{y}, \mathbf{b}^*) &\propto \exp \left(-\frac{1}{2} \mathbf{x}_{\mathcal{U}}^{*\top} \frac{\mathbf{B}_{\mathcal{U}\mathcal{U}}^{*\top} \mathbf{B}_{\mathcal{U}\mathcal{U}}^*}{\sigma_Y^2} \mathbf{x}_{\mathcal{U}}^* + \left(\frac{\mathbf{B}_{\mathcal{U}\mathcal{U}}^{*\top} \mathbf{y}^*}{\sigma_Y^2} \right)^\top \mathbf{x}_{\mathcal{U}}^* \right) \\ &\quad \exp \left(-\frac{1}{2} \mathbf{x}_{\mathcal{U}}^{*\top} \mathbf{Q}_{\mathcal{U}\mathcal{U}}^* \mathbf{x}_{\mathcal{U}}^* + (\mathbf{Q}_{\mathcal{U}\mathcal{U}}^* \tilde{\boldsymbol{\mu}}_{\mathcal{U}}^*)^\top \mathbf{x}_{\mathcal{U}}^* \right) \\ &\propto \exp \left(-\frac{1}{2} (\mathbf{x}_{\mathcal{U}}^* - \hat{\boldsymbol{\mu}}_{\mathcal{U}}^*)^\top \hat{\mathbf{Q}}_{\mathcal{U}\mathcal{U}}^* (\mathbf{x}_{\mathcal{U}}^* - \hat{\boldsymbol{\mu}}_{\mathcal{U}}^*) \right). \end{aligned}$$

Finally, using the relation $\mathbf{X} = \mathbf{T}^\top \mathbf{X}^*$ completes the proof. \square

Proof of Theorem 3. First note that $\pi_{\mathbf{Y}|\mathbf{A}\mathbf{X}}(\mathbf{y}|\mathbf{b}) = \pi_{\mathbf{Y}|\mathbf{X}_\epsilon^*}(\mathbf{y}|\mathbf{b}^*)$ and

$$\begin{aligned}\pi_{\mathbf{Y}|\mathbf{X}_\epsilon^*}(\mathbf{y}|\mathbf{b}^*) &= \int \pi_{\mathbf{X}_\mathcal{U}^*, \mathbf{Y}|\mathbf{X}_\epsilon^*}(\mathbf{x}_\mathcal{U}^*, \mathbf{y}|\mathbf{b}^*) d\mathbf{x}_\mathcal{U}^* \\ &= \int \pi_{\mathbf{Y}|\mathbf{X}_\mathcal{U}^*, \mathbf{X}_\epsilon^*}(\mathbf{y}|\mathbf{x}_\mathcal{U}^*, \mathbf{b}^*) \pi_{\mathbf{X}_\mathcal{U}^*|\mathbf{X}_\epsilon^*}(\mathbf{x}_\mathcal{U}^*|\mathbf{b}^*) d\mathbf{x}_\mathcal{U}^*.\end{aligned}\quad (12)$$

The goal is now to derive an explicit form of the density by evaluating the integral in (12). By the expressions in the proof of Theorem 4 we have

$$\begin{aligned}\pi_{\mathbf{Y}|\mathbf{X}_\mathcal{U}^*, \mathbf{X}_\epsilon^*}(\mathbf{y}|\mathbf{x}_\mathcal{U}^*, \mathbf{b}^*) \pi_{\mathbf{X}_\mathcal{U}^*|\mathbf{X}_\epsilon^*}(\mathbf{x}_\mathcal{U}^*|\mathbf{b}^*) &= \exp\left(-\frac{1}{2} \mathbf{x}_\mathcal{U}^{*\top} \frac{\mathbf{B}_\mathcal{U}^{*\top} \mathbf{B}_\mathcal{U}^*}{\sigma_Y^2} \mathbf{x}_\mathcal{U}^* + \left(\frac{\mathbf{B}_\mathcal{U}^{*\top} \mathbf{y}^*}{\sigma_Y^2}\right)^\top \mathbf{x}_\mathcal{U}^*\right) \\ &\quad \exp\left(-\frac{1}{2} \mathbf{x}_\mathcal{U}^{*\top} \mathbf{Q}_{\mathcal{U}\mathcal{U}}^* \mathbf{x}_\mathcal{U}^* + (\mathbf{Q}_{\mathcal{U}\mathcal{U}}^* \tilde{\boldsymbol{\mu}}_\mathcal{U}^*)^\top \mathbf{x}_\mathcal{U}^*\right) \\ &\quad \frac{|\mathbf{Q}_{\mathcal{U}\mathcal{U}}^*|^{\dagger/2}}{(2\pi)^{c_0} \sigma_Y^m} \exp\left(-\frac{1}{2} \left[\frac{\mathbf{y}^{*\top} \mathbf{y}^*}{\sigma_Y^2} + \tilde{\boldsymbol{\mu}}_\mathcal{U}^{*\top} \mathbf{Q}_{\mathcal{U}\mathcal{U}}^* \tilde{\boldsymbol{\mu}}_\mathcal{U}^*\right]\right) \\ &= \pi_{\mathbf{X}_\mathcal{U}^*|\mathbf{Y}, \mathbf{X}_\epsilon^*}(\mathbf{x}_\mathcal{U}^*|\mathbf{y}, \mathbf{b}^*) \frac{\exp\left(\frac{1}{2} \hat{\boldsymbol{\mu}}_\mathcal{U}^{*\top} \hat{\mathbf{Q}}_{\mathcal{U}\mathcal{U}}^* \hat{\boldsymbol{\mu}}_\mathcal{U}^*\right)}{|\hat{\mathbf{Q}}_{\mathcal{U}\mathcal{U}}^*|^{\dagger/2}} \\ &\quad \frac{|\mathbf{Q}_{\mathcal{U}\mathcal{U}}^*|^{\dagger/2}}{(2\pi)^{c_1} \sigma_Y^m} \exp\left(-\frac{1}{2} \left[\frac{\mathbf{y}^{*\top} \mathbf{y}^*}{\sigma_Y^2} + \boldsymbol{\mu}_\mathcal{U}^{*\top} \mathbf{Q}_{\mathcal{U}\mathcal{U}}^* \boldsymbol{\mu}_\mathcal{U}^*\right]\right),\end{aligned}$$

where c_0 and c_1 are positive constants. Inserting this expression in (12) and evaluating the integral, where one notes that $\pi_{\mathbf{X}_\mathcal{U}^*|\mathbf{Y}, \mathbf{X}_\epsilon^*}(\mathbf{x}_\mathcal{U}^*|\mathbf{y}, \mathbf{b}^*)$ integrates to one, gives the desired result. \square

C Conditional constrained distribution

In order to derive the conditional density $\pi(\mathbf{x}|\mathbf{A}\mathbf{x} = \mathbf{b})$ in (3) we will use what is known as the disintegration technique. The proof is built on the results in [CP97], which has the following definition.

Definition 1. Let $(\mathcal{X}, \mathcal{A}, \lambda)$ and $(\mathcal{T}, \mathcal{B}, \mu)$ be two measure spaces with σ -finite measures λ and μ . The measure λ has a disintegration $\{\lambda_b\}$ with respect to the measurable map $A : (\mathcal{X}, \mathcal{A}) \rightarrow (\mathcal{T}, \mathcal{B})$ and the measure μ , or a $(A(x), \mu)$ -disintegration if:

- (i) λ_b is a σ -finite measure on \mathcal{A} such that $\lambda_b(A(x) \neq b) = 0$, for μ -almost all b ,

and, for each non-negative measurable function f on \mathcal{X} :

- (ii) $b \rightarrow \int f d\lambda_b$ is measurable.

- (iii) $\int f d\lambda = \int \int f d\lambda_b d\mu$.

In the following theorem, we use the notation from Appendix B and let λ_n denote the Lebesgue measure on \mathbb{R}^n . Further, we define $\lambda_\mathcal{U}$ as the image measure of the projection onto the image of \mathbf{A} (which is not σ -finite), and $\lambda_\mathcal{E}$ as the image measure of the projection onto the null-space of \mathbf{A} .

Theorem 6. Let \mathbf{X} be a multivariate random variable with distribution \mathbb{P} on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, where \mathbb{P} has density $\pi(\mathbf{x})$ with respect to λ_n . Then the random variable $\mathbf{X}|\mathbf{A}\mathbf{X} = \mathbf{b}$ has density

$$\pi(\mathbf{x}|\mathbf{A}\mathbf{x} = \mathbf{b}) = \frac{\mathbb{I}(\mathbf{A}\mathbf{x} = \mathbf{b}) |\mathbf{A}\mathbf{A}^\top|^{-1/2} \pi(\mathbf{x})}{\pi_{\mathbf{A}\mathbf{X}}(\mathbf{b})},$$

with respect to the measure $\mathcal{L}_\mathbf{b}(\cdot) = \lambda_\mathcal{U}(\cdot \cap \{x : \mathbf{A}\mathbf{x} = \mathbf{b}\})$ on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$.

The proof is based on the following lemma.

Lemma 2. The measure $\mathcal{L}_\mathbf{b}(\cdot)$ is the $(\mathbf{A}, \mathcal{L}^k)$ -disintegration of the Lebesgue measure λ_n .

Proof. Thus we need to show that (i), (ii), and (iii) of Definition 1 holds. Clearly, (i) follows immediate from $\cdot \cap \{x : \mathbf{Ax} = \mathbf{b}\}$. To show (ii), note that

$$\begin{aligned} \int f d\mathcal{L}_{\mathbf{b}} &= \int f \left(\mathbf{T}^\top \mathbf{x}^* \right) \mathbb{I}_{\mathbf{AT}^\top \mathbf{x}^* = \mathbf{b}} (d\mathbf{x}^*) d\mathbf{x}_{\mathcal{U}}^* \\ &= \|\mathbf{H}\| \int_{\{\mathbf{x}^* : \mathbf{x}_{\mathcal{C}}^* = \mathbf{H}^{-1}\mathbf{b}\}} f^*(\mathbf{x}_{\mathcal{C}}^*, \mathbf{x}_{\mathcal{U}}^*) d\mathbf{x}_{\mathcal{U}}^* = \|\mathbf{H}\| \int f^*(\mathbf{H}^{-1}\mathbf{b}, \mathbf{x}_{\mathcal{U}}^*) d\mathbf{x}_{\mathcal{U}}^*, \end{aligned}$$

where $\mathbf{H} = (\mathbf{AT}^\top)_{\mathcal{C}\mathcal{C}}$ as defined in Section 3.2, $\|\mathbf{H}\|$ denotes the absolute value of the determinant of \mathbf{H} , and $f^*(\mathbf{x}^*) = f(\mathbf{T}^\top \mathbf{x})$. Since f is a measurable function it follows by Tonelli Theorem [Pol02] that above partial integral is measurable. Finally, to show (iii), we continue from the equation above and get

$$\begin{aligned} \iint f d\mathcal{L}_{\mathbf{b}} d\mathbf{b} &= \int \|\mathbf{H}\| \int f^*(\mathbf{H}^{-1}\mathbf{b}, \mathbf{x}_{\mathcal{U}}^*) d\mathbf{x}_{\mathcal{U}}^* d\mathbf{b} \\ &= \|\mathbf{H}\| |\mathbf{H}^{-1}| \iint f^*(\mathbf{b}^*, \mathbf{x}_{\mathcal{U}}^*) d\mathbf{x}_{\mathcal{U}}^* d\mathbf{b}^* = \int f d\lambda_n. \end{aligned}$$

□

Proof of Theorem 6. By Lemma 2 above and Theorem 3 (v) in [CP97] it follows that the random variable has density

$$\pi(\mathbf{x} | \mathbf{AX} = \mathbf{b}) = \frac{\pi(\mathbf{x})}{\mathcal{L}_{\mathbf{b}}\pi(\mathbf{x})} = \frac{\pi(\mathbf{x})}{\pi_{\mathbf{AX}}(\mathbf{b}) \|\mathbf{H}\|} = \frac{|\mathbf{AA}^\top|^{-1/2} \pi(\mathbf{x})}{\pi_{\mathbf{AX}}(\mathbf{b})},$$

a.e. with respect to $\mathcal{L}_{\mathbf{b}}$. Finally, it holds that $\pi(\mathbf{x} | \mathbf{AX} = \mathbf{b}) = \mathbb{I}(\mathbf{Ax} = \mathbf{b}) \pi(\mathbf{x} | \mathbf{AX} = \mathbf{b})$ a.e. since $\mathcal{L}_{\mathbf{b}}(\cdot) = \lambda_{\mathcal{U}}(\cdot \cap \{\mathbf{x} : \mathbf{Ax} = \mathbf{b}\})$. □

References

- [CP97] Joseph T Chang and David Pollard. “Conditioning as disintegration”. In: *Statist. Neerlandica* 51.3 (1997), pp. 287–317. ISSN: 0039-0402. DOI: 10.1111/1467-9574.00056.
- [HJ13] Roger A. Horn and Charles R. Johnson. *Matrix analysis*. Second. Cambridge University Press, Cambridge, 2013, pp. xviii+643. ISBN: 978-0-521-54823-6.
- [Pol02] David Pollard. *A user’s guide to measure theoretic probability*. Vol. 8. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2002, pp. xiv+351. ISBN: 0-521-80242-3; 0-521-00289-3.