Trace class Markov chains for Bayesian inference with generalized double Pareto shrinkage priors

Subhadip Pal, Kshitij Khare and James P. Hobert

Department of Statistics
102 Griffin-Floyd Hall
University of Florida
Gainesville, FL 32605

Abstract: Bayesian shrinkage methods have generated a lot of interest in recent years, especially in the context of high-dimensional linear regression. Armagan, Dunson and Lee (2013) propose a Bayesian shrinkage approach using generalized double Pareto priors. They establish several useful properties of this approach, including the derivation of a tractable three-block Gibbs sampler to sample from the resulting posterior density. We show that the Markov operator corresponding to this three-block Gibbs sampler is not Hilbert-Schmidt. We propose a simpler two-block Gibbs sampler, and show that the corresponding Markov operator is trace class (and hence Hilbert-Schmidt). Establishing the trace class property for the proposed two-block Gibbs sampler has several useful consequences. Firstly, it implies that the corresponding Markov chain is geometrically ergodic, thereby implying the existence of a Markov chain CLT, which in turn enables computation of asymptotic standard errors for Markov chain based estimates of posterior quantities. Secondly, since the proposed Gibbs sampler uses two-blocks, standard recipes in the literature can be used to construct a sandwich Markov chain (by inserting an appropriate extra step) to gain further efficiency and to achieve faster convergence. The trace class property for the two-block sampler implies that the corresponding sandwich Markov chain is also trace class and thereby geometrically ergodic. Finally, it also guarantees that all eigenvalues of the sandwich chain are dominated by the corresponding eigenvalues of the Gibbs sampling chain (with at least one strict domination). Our results demonstrate that a minor change in the structure of a Markov chain can lead to fundamental changes in its theoretical properties. We illustrate the improvement in efficiency and convergence resulting from our proposed Markov chains using simulated and real examples.

Keywords and phrases: Bayesian shrinkage, double Pareto prior, trace class operator, geometric ergodicity, sandwich algorithm.

1. Introduction

Consider the linear model $\mathbf{y} = X\boldsymbol{\beta} + \sigma\boldsymbol{\epsilon}$, where $\mathbf{y}$ is the $n \times 1$ vector of responses, X is the $n \times p$ design matrix, $\boldsymbol{\beta}$ is the $p \times 1$ vector of unknown regression coefficients, σ is an unknown scale parameter, and the entries of $\boldsymbol{\epsilon}$ are independent standard normal. Classical least squares methods fail when $p > n$, and the lasso (Tibshirani, 1996) was developed to estimate $\boldsymbol{\beta}$ in this case. The well-known Bayesian interpretation of the lasso (involving i.i.d. Laplace priors on the components of $\boldsymbol{\beta}$) has led to a flurry of recent research concerning the development of prior distributions for $(\boldsymbol{\beta}, \sigma)$ that yield posterior distributions with high (posterior) probability around sparse values of $\boldsymbol{\beta}$, i.e., values of $\boldsymbol{\beta}$ that have many entries equal to 0. Such prior distributions are referred to as "continuous shrinkage priors " and the corresponding models are referred to as "Bayesian shrinkage models." For an overview, see Polson and Scott (2010) and Bhattacharya et al. (2015). The posterior distributions associated with these models are highly intractable and are usually explored using MCMC algorithms.

In this paper, we focus on a Bayesian shrinkage model recently introduced by Armagan, Dunson and Lee (2013). The model can be specified as follows

$$\begin{aligned}
\mathbf{y} \mid \boldsymbol{\beta}, \sigma^2, \boldsymbol{\tau}, \boldsymbol{\lambda} &\sim N_n\left(X\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n\right) \\
\boldsymbol{\beta} \mid \sigma^2, \boldsymbol{\tau}, \boldsymbol{\lambda} &\sim N_p\left(\mathbf{0}, \sigma^2 D_{\boldsymbol{\tau}}\right) \\
\tau_i \mid \lambda_i &\sim \text{Exp}\left(\frac{\lambda_i^2}{2}\right) \text{ independently for } i = 1, 2, \ldots, p \\
\lambda_i &\sim \text{Gamma}\left(\zeta, \eta\right) \text{ independently for } i = 1, 2, \ldots, p \text{ with } \zeta > 0, \eta > 0 \\
\sigma^2 &\sim \text{Inverse-Gamma}\left(\alpha, \xi\right) \text{with } \alpha \geq 0, \xi \geq 0, \qquad (1)
\end{aligned}$$

where N_d denotes the d-variate normal density, and $D_{\boldsymbol{\tau}}$ is a diagonal matrix with diagonal entries given by the entries $\{\tau_j\}_{j=1}^p$ of $\boldsymbol{\tau}$. Also, ζ and α denote the shape parameters, and η and ξ denote the *rate* parameters for the Gamma and Inverse-Gamma densities respectively. Hence, the Inverse-Gamma(α, ξ) prior for σ^2 corresponds to the improper Jeffery's prior when $\alpha = 0$ and $\xi = 0$, and its (improper) density is proportional to $\frac{1}{\sigma^2}$. Straightforward calculations show that the posterior density corresponding to the improper prior with $\alpha = 0$ and $\xi = 0$ is a proper probability density.

It can be shown that for the above model, all the entries of $\boldsymbol{\beta}$ (given only σ^2) are mutually independent, and have a generalized double Pareto distribution. As a result, this model is referred to as the generalized double Pareto shrinkage model. The generalized double Pareto distribution has a spike at zero with Student's t-like heavy tails. This property makes it attractive for robust Bayesian shrinkage. Armagan, Dunson and Lee (2013) also show that the Normal-Jeffrey's prior and the Laplace prior (Bayesian lasso) can be obtained as limiting cases of their class of priors. Note that the parameters of interest here are $(\boldsymbol{\beta}, \sigma^2)$, and the τ's and λ's are 'augmented' parameters. As discussed below, these augmented parameters help in the development of a tractable MCMC approach for posterior computation. We refer the reader to Armagan, Dunson and Lee (2013) for a detailed study of other useful properties of the class of generalized double Pareto priors.

The joint posterior density of $(\boldsymbol{\beta}, \sigma^2)$ (the parameters of interest) is intractable in the sense that it is not feasible to generate direct i.i.d. samples from this density. To explore this posterior density, Armagan, Dunson and Lee (2013) propose a three-block Gibbs sampler, denoted here by $\tilde{\Phi} := \{\left(\tilde{\boldsymbol{\beta}}_m, \tilde{\sigma}^2_m\right)\}_{m=0}^{\infty}$ (on the state space $\mathbb{R}^p \times \mathbb{R}_+$), driven by the Markov transition density (Mtd)

$$\tilde{k}\left((\boldsymbol{\beta}, \sigma^2), (\check{\boldsymbol{\beta}}, \check{\sigma}^2)\right) = \int_{\mathbb{R}_+^p} \int_{\mathbb{R}_+^p} \pi\left(\check{\sigma}^2 \mid \check{\boldsymbol{\beta}}, \boldsymbol{\tau}, \boldsymbol{\lambda}, y\right) \pi\left(\check{\boldsymbol{\beta}} \mid \sigma^2, \boldsymbol{\tau}, \boldsymbol{\lambda}, y\right) \pi\left(\boldsymbol{\tau}, \boldsymbol{\lambda} \mid \boldsymbol{\beta}, \sigma^2, y\right)\, d\boldsymbol{\lambda}\, d\boldsymbol{\tau}. \tag{2}$$

Here $\pi(\cdot \mid \cdot)$ denotes the conditional density of the first group of arguments given the second group of arguments. The one-step dynamics of this Markov chain can be described as follows. To move from the current state, $\left(\tilde{\boldsymbol{\beta}}_m, \tilde{\sigma}^2_m\right)$, to the next state, $\left(\tilde{\boldsymbol{\beta}}_{m+1}, \tilde{\sigma}^2_{m+1}\right)$, first a random sample $(\boldsymbol{\tau}, \boldsymbol{\lambda})$ is drawn from the conditional density given $\tilde{\boldsymbol{\beta}}_m, \tilde{\sigma}^2_m, \mathbf{y}$. Then $\tilde{\boldsymbol{\beta}}_{m+1}$ is generated from the conditional density given $\tilde{\sigma}^2_m, \boldsymbol{\tau}, \boldsymbol{\lambda}, \mathbf{y}$, and finally $\tilde{\sigma}^2_{m+1}$ is simulated from the conditional density given $\tilde{\boldsymbol{\beta}}_{m+1}, \boldsymbol{\tau}, \boldsymbol{\lambda}, \mathbf{y}$. We refer to $\tilde{\Phi}$ as a three-block chain because it involves sampling three sets of parameters, namely $\boldsymbol{\beta}$, σ^2 and $(\boldsymbol{\tau}, \boldsymbol{\lambda})$, from their full conditional distributions. It can be shown that the Gibbs sampling chain $\tilde{\Phi}$ driven by the Mtd $\tilde{k}$, is easy to implement, and involves sampling from standard distributions (see Section A). However, crucial convergence and functional theoretic properties of the Gibbs sampling Markov chain $\tilde{\Phi}$ have not yet been investigated. The aim of this paper is to investigate these properties, and to construct alternative Markov chains which have provably better properties than $\tilde{\Phi}$.

It is straightforward to show that the Markov chain $\tilde{\Phi}$ described above is Harris ergodic with the appropriate stationary distribution. Harris ergodicity provides justification for using the Markov chain to construct strongly consistent estimators of intractable posterior expectations. For instance, if h is a real-valued measurable function, then we can conclude that the estimator $\bar{h}_m := \frac{1}{m+1} \sum_{i=0}^{m} h\left(\tilde{\boldsymbol{\beta}}_m, \tilde{\sigma}^2_m\right)$ is strongly consistent for the posterior expectation of h (assuming it exists), irrespective of the starting point of the Markov chain. However, to use the estimator $\bar{h}_m$ for practical purposes, it is also required to estimate the standard error associated with it. All known methods to compute asymptotically consistent estimates of standard errors for $\bar{h}_m$ require the existence of a Markov chain central limit theorem (CLT). In particular, we need to establish that

$$\sqrt{m}\left(\bar{h}_m - E_\pi h\right) \xrightarrow{d} N\left(0, c^2\right),$$

where c^2 is a finite positive constant. Currently, there are two standard methods available in the literature to prove a Markov chain CLT. One of the methods is to prove geometric ergodicity by establishing a geometric drift condition and an associated minorization condition (Jones and Hobert, 2001; Rosenthal, 1995) and the second method is to show that the underlying Markov chain is Hilbert-Schmidt (see Section 2). Establishing drift and minorization conditions to prove geometric ergodicity can be challenging. In fact, despite several attempts, we have been unable to prove that $\tilde{\Phi}$ is geometrically ergodic using the standard drift and minorization argument. On the other hand, there is a simple necessary and sufficient condition for a Markov operator to be Hilbert-Schmidt (see Section 2). We use this condition to prove that the Markov operator associated with Armagan et al.'s (2013) Gibbs sampler ($\tilde{\Phi}$) is never Hilbert-Schmidt (Theorem 2). While this result does not resolve the question of whether $\tilde{\Phi}$ is geometrically ergodic, it does carry pertinent information about this Markov chain. In particular, it shows that the absolute value of the corresponding operator either has (at least some) continuous spectrum, or has a countable set of eigenvalues which are not square-summable.

After studying the joint posterior density of all the parameters, we noticed that a two block Gibbs sampler, simpler than the three block chain $\tilde{\Phi}$, can be used to generate approximate samples from the posterior distribution of $(\boldsymbol{\beta}, \sigma^2)$. In particular, let $\Phi = \{(\boldsymbol{\beta}_m, \sigma_m^2)\}_{m=0}^\infty$ be a Markov chain on the state space $\mathbb{R}^p \times \mathbb{R}_+$, driven by the Markov transition density

$$k\left((\boldsymbol{\beta}, \sigma^2), (\check{\boldsymbol{\beta}}, \check{\sigma}^2)\right) = \int_{\mathbb{R}_+^p} \int_{\mathbb{R}_+^p} \pi\left(\check{\boldsymbol{\beta}}, \check{\sigma}^2 \mid \boldsymbol{\tau}, \boldsymbol{\lambda}, \mathbf{y}\right) \pi\left(\boldsymbol{\tau}, \boldsymbol{\lambda} \mid \boldsymbol{\beta}, \sigma^2, \mathbf{y}\right) \, d\boldsymbol{\lambda} \, d\boldsymbol{\tau}. \tag{3}$$

To move from the current state, $\left(\boldsymbol{\beta}_m, \sigma_m^2\right)$, to the next state, $\left(\boldsymbol{\beta}_{m+1}, \sigma_{m+1}^2\right)$, we first draw $(\boldsymbol{\tau}, \boldsymbol{\lambda})$ from the conditional density given $\boldsymbol{\beta}_m, \sigma_m^2, \mathbf{y}$, and then we draw $\left(\boldsymbol{\beta}_{m+1}, \sigma_{m+1}^2\right)$ from the conditional density given $\boldsymbol{\tau}, \boldsymbol{\lambda}, \mathbf{y}$. We refer to Φ as a two-block chain because it involves sampling $(\boldsymbol{\beta}, \sigma^2)$ and $(\boldsymbol{\tau}, \boldsymbol{\lambda})$ from their full conditional distributions. It can again be shown that the two-block Gibbs sampling chain Φ is easy to implement, and involves sampling from standard distributions (see Section A).

Note that, the only difference between the two-block Gibbs sampler Φ that we propose, and the three-block Gibbs sampler $\tilde{\Phi}$, is the strategy for sampling $\boldsymbol{\beta}$ and σ^2. To be specific, for Φ, we adopt a more efficient scheme to sample $\boldsymbol{\beta}$ and σ^2 jointly as a block from the density $\pi\left(\boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{\lambda}, \boldsymbol{\tau}, \mathbf{y}\right)$, whereas in $\tilde{\Phi}$, each of parameters $\boldsymbol{\beta}$ and σ^2 are drawn separately from their full conditional posterior distribution. Note that, unlike $\tilde{\Phi}$, the Markov chain Φ is reversible, and therefore is simpler to handle in terms of theoretical analysis. Moreover, there is some theory suggesting that blocking can improve the performance of a Gibbs sampler, as it is likely to reduce the correlation between successive iterates of the corresponding Markov chain (Liu, Wong and Kong, 1994).

Indeed, we prove that the Markov operator associated with the Gibbs sampling Markov chain Φ is trace-class, and hence Hilbert-Schmidt, when the design matrix X has full column rank (Theorem 1). Our results imply that the absolute value of the operator corresponding to $\tilde{\Phi}$ either has (at least some) continuous spectrum, or has a countable set of eigenvalues which are not square-summable, while the eigenvalues of the (self-adjoint) operator corresponding to Φ are not only square-summable, but in fact summable (note that all the aforementioned eigenvalues are less than 1 in absolute value). These results indicate that the Markov chain Φ is likely to be more efficient than the three-block chain $\tilde{\Phi}$ from Armagan, Dunson and Lee (2013). The simulation and real data experiments in Section 5 strongly support this assertion. Our results concretely demonstrate that a small change in the structure of a Markov chain can lead to fundamental changes in its theoretical properties.

Another advantage of the proposed two-block Gibbs sampler Φ is that it can be interpreted as a basic data augmentation (DA) algorithm with $(\boldsymbol{\beta}, \sigma^2)$ as the parameter block of interest, and $(\boldsymbol{\tau}, \boldsymbol{\lambda})$ as the augmented parameter block. This enables us to use the Haar PX-DA technique (Liu and Wu, 1999; Hobert and Marchev, 2008) to construct a 'sandwich' Markov chain by exploiting an appropriate group structure in the model (see Section 4). The trace class property for Φ, in conjunction with the results in Khare and Hobert (2011), implies that the operator corresponding to the sandwich Markov chain is also trace-class. Consequently, it follows that both Markov chains are geometrically ergodic. Moreover, for each $i \in \mathbb{N}$, the i^{th} largest eigenvalue of the sandwich operator is less than or equal to the corresponding eigenvalue of the DA operator, with strict inequality for at least one i.

The rest of the paper is organized as follows. In Section 2, we discuss relevant concepts from functional analysis. Investigation of the Hilbert-Schmidt property for the Markov chains $\tilde{\Phi}$ and Φ is undertaken in Sections 3. The construction of the sandwich algorithm is described in Section 4. In Section 5 we provide two simulation examples, one with $n > p$ and one with $n < p$, and a real data example, to demonstrate the efficiency gain obtained by using our proposed Markov chains (as compared to the three-block Gibbs sampler). The supplemental document contains details of the conditional posterior distributions necessary for analyzing the various Markov chains discussed above, along with some proofs and relevant mathematical identities.

2. Hilbert-Schmidt and trace class operators

In this section, we review the definitions of Hilbert-Schmidt and trace-class Markov operators, and provide necessary and sufficient conditions for showing that a given Markov operator is Hilbert-Schmidt or trace-class. Let $(\mathsf{X}, \mathcal{A}, \mu)$ be a measure space equipped with a countably generated σ-field $\mathcal{A}$ and a σ-finite measure μ. Let $\pi(dx) = \pi(x)\mu(dx)$ be an intractable probability measure defined on the above measure space. We assume that $\pi(x)$ is strictly positive almost everywhere on X. Define $L_0^2(\pi) = \{f \in L^2(\pi) : \pi f = 0\}$. Then, it follows that $L_0^2(\pi)$ is a separable Hilbert space (see Proposition 3.4.5 in Cohn (2013)) equipped with the inner product $\langle f, g\rangle_{L_0^2(\pi)} = \int_{\mathsf{X}} f(x)g(x)\,\pi(dx)$. The corresponding norm is given by $\|f\|_{L_0^2(\pi)} = \sqrt{\langle f, f\rangle_{L_0^2(\pi)}}$. Let $P(x, dy) = p(x, y)\mu(dy)$ be a Markov transition density with π as its invariant measure. Then, P defines an operator (also denoted by P for simplicity of notation), that acts on $f \in L_0^2(\pi)$ through

$$(Pf)(x) = \int_{\mathsf{X}} p(x, y)f(y)\mu(dy).$$

The operator P on $L_0^2(\pi)$ is defined to be Hilbert-Schmidt if for every orthonormal sequence $\{f_n\}_{n\geq 0}$ for $L_0^2(\pi)$, we have

$$\sum_{n=0}^{\infty} \|Pf_n\|^2_{L_0^2(\pi)} < \infty.$$

If P is a positive self-adjoint operator, then P is defined to be trace class if for every orthonormal sequence $\{f_n\}_{n\geq 0}$ for $L_0^2(\pi)$, we have

$$\sum_{n=0}^{\infty} \langle Pf_n, f_n\rangle_{L_0^2(\pi)} < \infty.$$

In the current setting, straightforward necessary and sufficient conditions can be derived for the two notions defined above. In particular, it can be shown that (see for example Jorgens (1982)) P is Hilbert-Schmidt if and only if

$$\int_{\mathsf{X}}\int_{\mathsf{X}} \frac{p(x, y)^2\pi(x)}{\pi(y)}\mu(dy)\mu(dx) = \int_{\mathsf{X}}\int_{\mathsf{X}} \left(\frac{p(x, y)}{\pi(y)}\right)^2 \pi(x)\pi(y)\mu(dy)\mu(dx) < \infty. \tag{4}$$

Also, a positive self-adjoint Markov operator P is trace class if and only if

$$\int_{\mathsf{X}} p(x,x)\mu(dx) < \infty. \tag{5}$$

Establishing the Hilbert-Schmidt or trace class property for a Markov operator P has important implications for the associated Markov chain. If P is Hilbert-Schmidt, then it follows that it is compact, and its singular values are square-summable. If P (positive, self-adjoint) is trace class, then it again follows that it is compact, and its singular values are summable (stronger than square-summable). In either case, if the corresponding Markov chain is Harris ergodic, then compactness implies that the spectral radius of the Markov operator is less than 1. It follows that the corresponding Markov chain is geometrically ergodic (see Proposition 2.1 and Remark 2.1 in Roberts and Rosenthal (1997)). On the other hand, if a Markov operator P is not Hilbert-Schmidt, then it follows that either its absolute value operator ($\sqrt{P^*P}$) does not have a countable spectrum, or it has a countable set of eigenvalues which are not square-summable.

3. Properties of the two and three block Gibbs samplers

In this section, we show that the operator associated with the proposed two-block Gibbs sampler Φ, with Markov transition density k specified in (3) is trace class when the design matrix X has full column rank. Since the Markov transition density in (3) is strictly positive, it follows that Φ is Harris ergodic, see (Meyn and Tweedie, 1993, Page 87) and Asmussen and Glynn (2011). Based on the discussion in Section 2, it follows that the corresponding operator is Hilbert-Schmidt, compact, and that the two-block chain is geometrically ergodic. We also show that the operator associated with the three-block Gibbs sampler $\tilde{\Phi}$, with Markov transition density $\tilde{k}$ specified in (2), is not Hilbert-Schmidt. The detailed form of various relevant conditional densities in (2) and (3) is provided the Supplementary Section A.

Let K be the Markov operator corresponding to the two block sampler Φ, and assuming X has full column rank, let $P_X = X(X^TX)^{-1}X^T$ be the projection matrix on the column space of X. We prove the following result.

Theorem 1 *Let $p < n$ and $rank(X) = p$, i.e., X has full column rank. Also let $\mathbf{y}^T(I_n - P_X)\mathbf{y} > 0$. Then, the Markov operator K is trace class.*

Remark 1 *Note that,* $\mathbf{y}^T(I_n - P_X)\mathbf{y} = 0$ *if and only if* $\mathbf{y}$ *lies in the column space of* X*. Since* $p < n$*, the probability (under the model in (1)) that the* $n \times 1$ *data vector* $\mathbf{y}$ *lies in the column space of the* $n \times p$ *matrix* X *is zero. Hence, before the data are observed, the assumption* $\mathbf{y}^T(I_n - P_X)\mathbf{y} > 0$ *holds with probability* 1*.*

Proof: By (5), to prove the required result, we need to show that

$$I := \int_{\mathbb{R}^p} \int_{\mathbb{R}_+} k\left((\boldsymbol{\beta}, \sigma^2), (\boldsymbol{\beta}, \sigma^2)\right) \; d\sigma^2 d\boldsymbol{\beta} < \infty. \tag{6}$$

By Fubini's theorem, we get that

$$I \;\; = \;\; \int_{\mathbb{R}_+^p} \int_{\mathbb{R}_+^p} \int_{\mathbb{R}^p} \int_{\mathbb{R}_+} \pi\left(\boldsymbol{\beta} \mid \boldsymbol{\tau}, \boldsymbol{\lambda}, \sigma^2, \mathbf{y}\right) \pi\left(\sigma^2 \mid \boldsymbol{\tau}, \boldsymbol{\lambda}, \mathbf{y}\right) \pi\left(\boldsymbol{\tau}, \boldsymbol{\lambda} \mid \boldsymbol{\beta}, \sigma^2, \mathbf{y}\right) \; d\sigma^2 \; d\boldsymbol{\beta} \; d\boldsymbol{\tau} \; d\boldsymbol{\lambda}. \tag{7}$$

To show the finiteness of I, we will first break I as a sum of 2^p integrals, and show that each one of them is finite. To achieve this, we will first integrate out σ^2 using the Inverse-Gamma density, and then using the t-density show that the integral with respect to $\boldsymbol{\beta}$ of the resulting function is bounded above by a constant multiple of

$$\prod_{j=1}^{p} \lambda_j^{1+\zeta} \tau_j^{\frac{1}{2}-1} e^{-\frac{\lambda_j^2 \tau_j}{2} - \eta\lambda_j},$$

which has a finite integral on $\mathbb{R}_p^+ \times \mathbb{R}_p^+$.

Before proceeding ahead, recall that

$$\widehat{\boldsymbol{\beta}} = (X^T X + D_{\boldsymbol{\tau}}^{-1})^{-1} X^T \mathbf{y}.$$

Let

$$\Delta_1 = \left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\right)^T (X^T X + D_{\boldsymbol{\tau}}^{-1}) \left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\right) \text{ and } \widehat{\Delta} = (\mathbf{y} - X\widehat{\boldsymbol{\beta}})^T (\mathbf{y} - X\widehat{\boldsymbol{\beta}}) + \widehat{\boldsymbol{\beta}}^T D_{\boldsymbol{\tau}}^{-1} \widehat{\boldsymbol{\beta}} + 2\xi.$$

It follows from (A.2), (A.4), (A.5) and (A.6) that

$$
\begin{aligned}
&\pi\left(\boldsymbol{\beta} \mid \boldsymbol{\tau}, \boldsymbol{\lambda}, \sigma^2, \mathbf{y}\right) \pi\left(\sigma^2 \mid \boldsymbol{\tau}, \boldsymbol{\lambda}, \mathbf{y}\right) \pi\left(\boldsymbol{\tau} \mid \boldsymbol{\beta}, \sigma^2, \boldsymbol{\lambda}, \mathbf{y}\right) \pi\left(\boldsymbol{\lambda} \mid \boldsymbol{\beta}, \sigma^2, \mathbf{y}\right) \\
= \; & \left\{ \frac{|X^T X + D_{\boldsymbol{\tau}}^{-1}|^{\frac{1}{2}}}{(\sqrt{2\pi})^p \sigma^p} e^{-\frac{\Delta_1}{2\sigma^2}} \right\} \left\{ \frac{\widehat{\Delta}^{\frac{n+2\alpha}{2}}}{2^{\frac{n+2\alpha}{2}} \Gamma(\frac{n+2\alpha}{2})} (\sigma^2)^{-\frac{n+2\alpha}{2}-1} e^{-\frac{\widehat{\Delta}}{2\sigma^2}} \right\} \\
& \times \left\{ \prod_{j=1}^p \frac{\lambda_j}{\sqrt{2\pi}} \tau_j^{\frac{1}{2}-1} e^{-\frac{1}{2}\left\{\lambda_j^2 \tau_j + \frac{\beta_j^2}{\sigma^2} \frac{1}{\tau_j}\right\}} e^{\frac{\lambda_j |\beta_j|}{\sigma}} \right\} \times \left\{ \prod_{j=1}^p \frac{\left(\frac{|\beta_j|}{\sigma} + \eta\right)^{\zeta+1} \lambda_j^{\zeta}}{\Gamma(\zeta+1)} \; e^{-\left(\frac{|\beta_j|}{\sigma}+\eta\right)\lambda_j} \right\} \\
= \; & C_2 \left\{ |X^T X + D_{\boldsymbol{\tau}}^{-1}|^{\frac{1}{2}} \widehat{\Delta}^{\frac{n+2\alpha}{2}} \; (\sigma^2)^{-\frac{n+p+2\alpha}{2}-1} e^{-\frac{\Delta_1 + \widehat{\Delta} + \boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1} \boldsymbol{\beta}}{2\sigma^2}} \right\} \\
& \times \left\{ \prod_{j=1}^p \lambda_j \tau_j^{\frac{1}{2}-1} e^{-\frac{\lambda_j^2 \tau_j}{2}} \right\} \times \left\{ \prod_{j=1}^p \left(\frac{|\beta_j|}{\sigma} + \eta \right)^{\zeta+1} \lambda_j^{\zeta} \; e^{-\eta\lambda_j}, \right\} \\
= \; & C_2 \left\{ |X^T X + D_{\boldsymbol{\tau}}^{-1}|^{\frac{1}{2}} \widehat{\Delta}^{\frac{n+2\alpha}{2}} \; (\sigma^2)^{-\frac{n+p+2\alpha}{2}-1} e^{-\frac{\Delta_1 + \widehat{\Delta} + \boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1} \boldsymbol{\beta}}{2\sigma^2}} \right\} \\
& \times \left\{ \sum_{(\delta_1,\ldots,\delta_p) \in \{0,1\}^p} \left[\eta^{(\zeta+1)(p - \sum_{i=1}^p \delta_i)} \prod_{j=1}^p \left(\frac{|\beta_j|}{\sigma} \right)^{\delta_j(\zeta+1)} \right] \right\} \left\{ \prod_{j=1}^p \lambda_j^{1+\zeta} \tau_j^{\frac{1}{2}-1} e^{-\frac{\lambda_j^2 \tau_j}{2} - \eta\lambda_j} \right\},
\end{aligned}
\tag{8}
$$

where $C_2 = \left(2^{\frac{n+2\alpha}{2}} \Gamma(\frac{n+2\alpha}{2}) \left(\Gamma(\zeta+1) 2\pi\right)^p\right)^{-1}$. For arbitrary $\boldsymbol{\delta} := (\delta_1, \ldots, \delta_p) \in \{0,1\}^p$, let

$$
\begin{aligned}
f_{\boldsymbol{\delta}}\left(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\lambda}, \boldsymbol{\tau}\right) \quad := \quad & \left\{ \frac{|X^T X + D_{\boldsymbol{\tau}}^{-1}|^{\frac{1}{2}} \widehat{\Delta}^{\frac{n+2\alpha}{2}}}{\Gamma\left(\frac{n+p+2\alpha+(1+\zeta) \sum_{i=1}^p \delta_i}{2} \right)} \; (\sigma^2)^{-\frac{n+p+2\alpha+(1+\zeta)\sum_{i=1}^p \delta_i}{2} - 1} e^{-\frac{\Delta_1 + \widehat{\Delta} + \boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1} \boldsymbol{\beta}}{2\sigma^2}} \right\} \\
& 2^{-\frac{n+p+2\alpha+(1+\zeta)\sum_{i=1}^p \delta_i}{2}} \times \left\{ \prod_{j=1}^p |\beta_j|^{\delta_j(\zeta+1)} \right\} \left\{ \prod_{j=1}^p \lambda_j^{1+\zeta} \tau_j^{\frac{1}{2}-1} e^{-\frac{\lambda_j^2 \tau_j}{2} - \eta\lambda_j} \right\}.
\end{aligned}
\tag{9}
$$

From (8), it is easy to see that (6) holds if we can prove that

$$
\int_{\mathbb{R}_+^p} \int_{\mathbb{R}_+^p} \int_{\mathbb{R}^p} \int_{\mathbb{R}_+} f_{\boldsymbol{\delta}}\left(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\lambda}, \boldsymbol{\tau}\right) d\sigma^2 \, d\boldsymbol{\beta} \, d\boldsymbol{\tau} \, d\boldsymbol{\lambda} < \infty. \tag{10}
$$

for arbitrary $\boldsymbol{\delta} \in \{0,1\}^p$. From (9) and the form of the Inverse-Gamma density, it follows that

$$\begin{aligned}
&\int_{\mathbb{R}_+} f_{\boldsymbol{\delta}}\left(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\lambda}, \boldsymbol{\tau}\right) d\sigma^2 \\
= \quad & \left\{ \frac{|X^TX + D_{\boldsymbol{\tau}}^{-1}|^{\frac{1}{2}} \widehat{\Delta}^{\frac{n+2\alpha}{2}} \prod_{j=1}^p |\beta_j|^{\delta_j(\zeta+1)}}{\left(\Delta_1 + \widehat{\Delta} + \boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1}\boldsymbol{\beta}\right)^{\frac{n+p+2\alpha+(1+\zeta)\sum\limits_{i=1}^p \delta_i}{2}}} \right\} \times \left\{ \prod_{j=1}^p \lambda_j^{1+\zeta} \tau_j^{\frac{1}{2}-1} e^{-\frac{\lambda_j^2 \tau_j}{2} - \eta\lambda_j} \right\}. \qquad (11)
\end{aligned}$$

Note that

$$\begin{aligned}
(\mathbf{y} - X\widehat{\boldsymbol{\beta}})^T(\mathbf{y} - X\widehat{\boldsymbol{\beta}}) + \widehat{\boldsymbol{\beta}}^T D_{\boldsymbol{\tau}}^{-1}\widehat{\boldsymbol{\beta}} + 2\xi &= \mathbf{y}^T\mathbf{y} - 2\widehat{\boldsymbol{\beta}}^T X^T\mathbf{y} + \widehat{\boldsymbol{\beta}}^T (X^TX + D_{\boldsymbol{\tau}}^{-1})\widehat{\boldsymbol{\beta}} + 2\xi \\
&= \mathbf{y}^T\mathbf{y} - \mathbf{y}^TX(X^TX + D_{\boldsymbol{\tau}}^{-1})^{-1}X^T\mathbf{y} + 2\xi \\
&\leq \mathbf{y}^T\mathbf{y} + 2\xi, \qquad (12)
\end{aligned}$$

and

$$\begin{aligned}
\Delta_1 + \widehat{\Delta} + \boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1}\boldsymbol{\beta} \quad &\geq \quad \left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\right)^T (X^TX + D_{\boldsymbol{\tau}}^{-1}) \left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\right) + (\mathbf{y} - X\widehat{\boldsymbol{\beta}})^T(\mathbf{y} - X\widehat{\boldsymbol{\beta}}) + \widehat{\boldsymbol{\beta}}^T D_{\boldsymbol{\tau}}^{-1}\widehat{\boldsymbol{\beta}} + \\
&\quad\quad 2\xi \\
&= \quad \left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\right)^T (X^TX + D_{\boldsymbol{\tau}}^{-1}) \left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\right) + \mathbf{y}^T\mathbf{y} - 2\widehat{\boldsymbol{\beta}}^T X^T\mathbf{y} + \widehat{\boldsymbol{\beta}}^T (X^TX + D_{\boldsymbol{\tau}}^{-1})\widehat{\boldsymbol{\beta}} + 2\xi \\
&= \quad \left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\right)^T (X^TX + D_{\boldsymbol{\tau}}^{-1}) \left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\right) + \mathbf{y}^T\mathbf{y} - \mathbf{y}^TX(X^TX + D_{\boldsymbol{\tau}}^{-1})^{-1}X^T\mathbf{y} + 2\xi \\
&\geq \quad \left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\right)^T (X^TX + D_{\boldsymbol{\tau}}^{-1}) \left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\right) + \mathbf{y}^T\mathbf{y} - \mathbf{y}^TX(X^TX)^{-1}X^T\mathbf{y} + 2\xi.
\end{aligned}$$

It follows that

$$\begin{aligned}
\Delta_1 + \widehat{\Delta} + \boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1}\boldsymbol{\beta} &= \left[\mathbf{y}^T(I_n - P_X)\mathbf{y} + 2\xi\right] \left\{1 + \left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\right)^T \frac{(X^TX + D_{\boldsymbol{\tau}}^{-1})}{\mathbf{y}^T(I_n - P_X)\mathbf{y} + 2\xi} \left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\right)\right\}, \\
&\geq \left[\mathbf{y}^T(I_n - P_X)\mathbf{y} + 2\xi\right] \left\{1 + \left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\right)^T \frac{(\varpi_{min} I_p + D_{\boldsymbol{\tau}}^{-1})}{\mathbf{y}^T(I_n - P_X)\mathbf{y} + 2\xi} \left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\right)\right\}, \\
&= \left[\mathbf{y}^T(I_n - P_X)\mathbf{y} + 2\xi\right] \left\{1 + \sum_{j=1}^p \frac{(\beta_j - \widehat{\beta}_j)^2}{\nu_j \xi_j^2}\right\}, \qquad (13)
\end{aligned}$$

where ϖ_{min} is the smallest eigenvalue of X^TX, $\xi_j = \sqrt{\frac{\mathbf{y}^T(I_n - P_X)\mathbf{y} + 2\xi}{(\varpi_{min} + \frac{1}{\tau_j})\nu_j}}$, $\nu_j = \frac{n+2\alpha}{p} + (1+\zeta)\delta_j$.

From (11), (12), (13) we get that,

$$
\begin{aligned}
&\int_{\mathbb{R}_+} f_{\boldsymbol{\delta}}\left(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\lambda}, \boldsymbol{\tau}\right) d\sigma^2 \qquad (14)\\
\leq\ & \left\{ \frac{|X^T X + D_{\boldsymbol{\tau}}^{-1}|^{\frac{1}{2}} \left(\mathbf{y}^T\mathbf{y} + 2\xi\right)^{\frac{n+2\alpha}{2}} \prod_{j=1}^p |\beta_j|^{\delta_j(\zeta+1)}}{\left([\mathbf{y}^T(I_n - P_X)\mathbf{y} + 2\xi] \left\{1 + \sum_{j=1}^p \frac{(\beta_j - \widehat{\beta}_j)^2}{\nu_j \xi_j^2}\right\}\right)^{\frac{n+p+2\alpha+(1+\zeta)\sum_{i=1}^p \delta_i}{2}}} \right\} \times \left\{ \prod_{j=1}^p \lambda_j^{1+\zeta} \tau_j^{\frac{1}{2}-1} e^{-\frac{\lambda_j^2 \tau_j}{2} - \eta\lambda_j} \right\}.
\end{aligned}
$$

Now, for $\boldsymbol{\delta} \in \{0,1\}^p$ such that at least one $\delta_j \neq 0$, we get that

$$
\begin{aligned}
& \left[1 + \sum_{j=1}^p \frac{(\beta_j - \widehat{\beta}_j)^2}{\nu_j \xi_j^2}\right]^{\frac{n+p+2\alpha+(1+\zeta)\sum_{j=1}^p \delta_j}{2}} \\
= \ & \left[1 + \sum_{j=1}^p \frac{(\beta_j - \widehat{\beta}_j)^2}{\nu_j \xi_j^2}\right]^{\frac{n+p+2\alpha}{2}} \left[1 + \sum_{j=1}^p \frac{(\beta_j - \widehat{\beta}_j)^2}{\nu_j \xi_j^2}\right]^{\frac{(1+\zeta)\sum_{j=1}^p \delta_j}{2}} \\
= \ & \left[\sum_{j=1}^p \frac{1}{p}\left(1 + \frac{(\beta_j - \widehat{\beta}_j)^2}{\nu_j \xi_j^2}\right)\right]^{\frac{n+p+2\alpha}{2}} \left[1 + \sum_{j=1}^p \frac{(\beta_j - \widehat{\beta}_j)^2}{\nu_j \xi_j^2}\right]^{\frac{(1+\zeta)\sum_{j=1}^p \delta_j}{2}} \\
\overset{(a)}{\geq} \ & \left[\prod_{j=1}^p \left(1 + \frac{(\beta_j - \widehat{\beta}_j)^2}{\nu_j \xi_j^2}\right)\right]^{\frac{n+p+2\alpha}{2p}} \left[\sum_{j=1}^p \left(\frac{\delta_j}{\sum_{i=1}^p \delta_i}\right)\left(1 + \frac{(\beta_j - \widehat{\beta}_j)^2}{\nu_j \xi_j^2}\right)\right]^{\frac{(1+\zeta)\sum_{j=1}^p \delta_j}{2}} \\
\overset{(b)}{\geq} \ & \left[\prod_{j=1}^p \left(1 + \frac{(\beta_j - \widehat{\beta}_j)^2}{\nu_j \xi_j^2}\right)\right]^{\frac{n+p+2\alpha}{2p}} \left[\prod_{j=1}^p \left(1 + \frac{(\beta_j - \widehat{\beta}_j)^2}{\nu_j \xi_j^2}\right)^{\frac{(1+\zeta)\delta_j}{2}}\right] \\
= \ & \left[\prod_{j=1}^p \left(1 + \frac{(\beta_j - \widehat{\beta}_j)^2}{\nu_j \xi_j^2}\right)^{\frac{1+(1+\zeta)\delta_j + \frac{n+2\alpha}{p}}{2}}\right] = \left[\prod_{j=1}^p \left(1 + \frac{(\beta_j - \widehat{\beta}_j)^2}{\nu_j \xi_j^2}\right)^{\frac{1+\nu_j}{2}}\right],
\end{aligned}
$$

where (a) follows from the AM-GM inequality, and (b) follows from a generalized version which says that

$$
\frac{\sum_{i=1}^m a_i p_i}{\sum_{i=1}^m p_i} \geq \prod_{i=1}^m a_i^{p_i / \sum_{j=1}^m p_j}
$$

for non-negative a_i and p_i with at least one positive p_i (Steele, 2004). Also, if $\sum_{j=1}^p \delta_j = 0$, (i.e. $\delta_j = 0$ for all $j \in 1, 2, \ldots, p$) then it again follows by the AM-GM inequality that

$$
\left[1 + \sum_{j=1}^p \frac{(\beta_j - \widehat{\beta}_j)^2}{\nu_j \xi_j^2}\right]^{\frac{n+p+2\alpha+(1+\zeta)\sum_{j=1}^p \delta_j}{2}} \geq \prod_{j=1}^p \left(1 + \frac{(\beta_j - \widehat{\beta}_j)^2}{\nu_j \xi_j^2}\right)^{\frac{1+\nu_j}{2}}. \qquad (15)
$$

Hence (15) holds for all $(\delta_1, \delta_2, \ldots, \delta_p) \in \{0,1\}^p$. From (14) and (15) we get that

$$\begin{aligned}\int_{\mathbb{R}_+} f_{\boldsymbol{\delta}}\left(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\lambda}, \boldsymbol{\tau}\right) d\sigma^2 \quad &\leq \quad \left\{ \frac{|X^TX + D_{\boldsymbol{\tau}}^{-1}|^{\frac{1}{2}} \left(\mathbf{y}^T\mathbf{y} + 2\xi\right)^{\frac{n+2\alpha}{2}}}{[\mathbf{y}^T(I_n - P_X)\mathbf{y} + 2\xi]^{\frac{n+p+2\alpha+(1+\zeta)\sum_{i=1}^{p}\delta_i}{2}}} \quad \prod_{j=1}^{p} \frac{|\beta_j|^{\delta_j(\zeta+1)}}{\left(1 + \frac{(\beta_j - \widehat{\beta}_j)^2}{\nu_j \xi_j^2}\right)^{\frac{1+\nu_j}{2}}} \right\} \times \\ &\left\{ \prod_{j=1}^{p} \lambda_j^{1+\zeta} \tau_j^{\frac{1}{2}-1} e^{-\frac{\lambda_j^2 \tau_j}{2} - \eta\lambda_j} \right\}. \end{aligned} \tag{16}$$

From (16) and Proposition A2 in the appendix, we get that

$$\begin{aligned}&\int_{\mathbb{R}^p} \int_{\mathbb{R}_+} f_{\boldsymbol{\delta}}\left(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\lambda}, \boldsymbol{\tau}\right) d\sigma^2 \, d\boldsymbol{\beta} \\ \leq \quad & \left\{ \frac{|X^TX + D_{\boldsymbol{\tau}}^{-1}|^{\frac{1}{2}} \left(\mathbf{y}^T\mathbf{y} + 2\xi\right)^{\frac{n+2\alpha}{2}}}{[\mathbf{y}^T(I_n - P_X)\mathbf{y} + 2\xi]^{\frac{n+p+2\alpha+(1+\zeta)\sum_{i=1}^{p}\delta_i}{2}}} \quad \prod_{j=1}^{p} \frac{C_{0j}}{\sqrt{\varpi_{min} + \frac{1}{\tau_j}}} \right\} \times \\ & \left\{ \prod_{j=1}^{p} \lambda_j^{1+\zeta} \tau_j^{\frac{1}{2}-1} e^{-\frac{\lambda_j^2 \tau_j}{2} - \eta\lambda_j} \right\}. \\ \leq \quad & \left\{ \frac{\left(\mathbf{y}^T\mathbf{y} + 2\xi\right)^{\frac{n+2\alpha}{2}}}{[\mathbf{y}^T(I_n - P_X)\mathbf{y} + 2\xi]^{\frac{n+p+2\alpha+(1+\zeta)\sum_{i=1}^{p}\delta_i}{2}}} \quad \prod_{j=1}^{p} \frac{C_{0j}\, \sqrt{\varpi_{max} + \frac{1}{\tau_j}}}{\sqrt{\varpi_{min} + \frac{1}{\tau_j}}} \right\} \times \\ & \left\{ \prod_{j=1}^{p} \lambda_j^{1+\zeta} \tau_j^{\frac{1}{2}-1} e^{-\frac{\lambda_j^2 \tau_j}{2} - \eta\lambda_j} \right\} \\ \leq \quad & \left\{ \frac{\left(\mathbf{y}^T\mathbf{y} + 2\xi\right)^{\frac{n+2\alpha}{2}}}{[\mathbf{y}^T(I_n - P_X)\mathbf{y} + 2\xi]^{\frac{n+p+2\alpha+(1+\zeta)\sum_{i=1}^{p}\delta_i}{2}}} \quad \prod_{j=1}^{p} \frac{C_{0j}\, \sqrt{\varpi_{max}}}{\sqrt{\varpi_{min}}} \right\} \times \left\{ \prod_{j=1}^{p} \lambda_j^{1+\zeta} \tau_j^{\frac{1}{2}-1} e^{-\frac{\lambda_j^2 \tau_j}{2} - \eta\lambda_j} \right\}, \end{aligned} \tag{17}$$

where ϖ_{max} denotes the maximum eigenvalue of X^TX. Hence we get that

$$\begin{aligned}\int_{\mathbb{R}_+^p} \int_{\mathbb{R}_+^p} \int_{\mathbb{R}^p} \int_{\mathbb{R}_+} f_{\boldsymbol{\delta}}\left(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\lambda}, \boldsymbol{\tau}\right) d\sigma^2 \, d\boldsymbol{\beta} \, d\boldsymbol{\tau} \, d\boldsymbol{\lambda} \quad &\leq \quad C_3 \left\{ \prod_{j=1}^{p} \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} \lambda_j^{1+\zeta} \tau_j^{\frac{1}{2}-1} e^{-\frac{\lambda_j^2 \tau_j}{2} - \eta\lambda_j} d\tau_j \, d\lambda_j \right\} \\ &\leq \quad C_3 \left\{ \prod_{j=1}^{p} \int_{\mathbb{R}_+} \sqrt{2\pi}\, \lambda_j^{(1+\zeta)-1} e^{-\eta\lambda_j} d\lambda_j \right\} \\ &\leq \quad C_3 \left\{ \frac{\sqrt{2\pi}\, \Gamma(\zeta+1)}{\eta^{\zeta+1}} \right\}^p < \infty, \end{aligned}$$

where

$$C_3 = \left\{ \frac{\left(\mathbf{y}^T\mathbf{y} + 2\xi\right)^{\frac{n+2\alpha}{2}}}{\left[\mathbf{y}^T(I_n - P_X)\mathbf{y} + 2\xi\right]^{\frac{n+p+2\alpha+(1+\zeta)\sum_{i=1}^{p}\delta_i}{2}}} \prod_{j=1}^{p} \frac{C_{0j}\sqrt{\varpi_{max}}}{\sqrt{\varpi_{min}}} \right\}.$$

□

Let $\tilde{K}$ be the Markov operator corresponding to the three-block sampler $\tilde{\Phi}$. A proof of the following result is provided in Supplemental Section B.

Theorem 2 *The Markov operator $\tilde{K}$ is not Hilbert- Schmidt (for all possible values of p and n).*

Despite our substantial efforts, the question of whether Theorem 1 holds in the case $p > n$ (which implies that X does not have full column rank) still remains unresolved. Some *key* steps of our proof need ϖ_{min} (the smallest eigenvalue of X^TX) to be strictly positive. Note from (11) that

$$\int_{\mathbb{R}^p}\int_{\mathbb{R}_+} f_{\boldsymbol{\delta}}\left(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\lambda}, \boldsymbol{\tau}\right) d\sigma^2 d\boldsymbol{\beta}$$

$$= \left(\int_{\mathbb{R}^p} \left\{ \frac{|X^TX + D_{\boldsymbol{\tau}}^{-1}|^{\frac{1}{2}} \widehat{\Delta}^{\frac{n+2\alpha}{2}} \prod_{j=1}^{p} |\beta_j|^{\delta_j(\zeta+1)}}{\left(\Delta_1 + \widehat{\Delta} + \boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1}\boldsymbol{\beta}\right)^{\frac{n+p+2\alpha+(1+\zeta)\sum_{i=1}^{p}\delta_i}{2}}} \right\} d\boldsymbol{\beta} \right) \times \left\{ \prod_{j=1}^{p} \lambda_j^{1+\zeta} \tau_j^{\frac{1}{2}-1} e^{-\frac{\lambda_j^2\tau_j}{2} - \eta\lambda_j} \right\} \tag{18}$$

Using the assumption that $\varpi_{min} > 0$, we are able to show that the $\boldsymbol{\beta}$-integral in (18) is bounded above by a constant (see the last step of (17) and derivation of C_{0j} in the proof of Proposition C2). This is crucial for the proof, as we know that

$$\int_{\mathbb{R}_+^p}\int_{\mathbb{R}_+^p} \left\{ \prod_{j=1}^{p} \lambda_j^{1+\zeta} \tau_j^{\frac{1}{2}-1} e^{-\frac{\lambda_j^2\tau_j}{2} - \eta\lambda_j} \right\} < \infty.$$

However, if $\varpi_{min} = 0$ (which will be the case if $n < p$), then the dominant term in the bound that we can get for this $\boldsymbol{\beta}$-integral is a constant multiple of $\prod_{j=1}^{p} \tau_j^{\delta_j(1+\zeta)/2}$, and

$$\int_{\mathbb{R}_+}\int_{\mathbb{R}_+} \tau_j^{(1+\zeta)/2} \lambda_j^{1+\zeta} \tau_j^{\frac{1}{2}-1} e^{-\frac{\lambda_j^2\tau_j}{2} - \eta\lambda_j} = \infty.$$

Hence, except for the case when all the δ_j's are zero, our proof does not work if X^TX is singular. The above problem still persists even if we assume $\alpha > 0$ or $\xi > 0$. Nevertheless, the experiments in Section 5 suggest that the proposed two-block sampler (and the associated sandwich algorithm introduced in Section 4) is more efficient that the three block sampler in the case $p > n$ as well.

4. Construction of the sandwich Markov chain

As noted earlier, the two-block Gibbs sampler can be regarded as a Data Augmentation (DA) algorithm with $(\boldsymbol{\beta}, \sigma^2)$ as the parameter block of interest, and $(\boldsymbol{\tau}, \boldsymbol{\lambda})$ as the augmented parameter block. The sandwich algorithm is a powerful method for improving the convergence and efficiency of the DA algorithm. This method was introduced independently by Liu and Wu (1999), who call it "PX-DA", and Meng and van Dyk (1999), who call it "Marginal Augmentation" (MA). The basic idea behind the method is to introduce an additional step in the DA algorithm, which is much cheaper computationally than the two conditional draws in the DA algorithm, while preserving the stationary distribution and reversibility. It is often possible to construct a sandwich algorithm that converges much faster than the underlying DA algorithm while requiring roughly the same computational effort per iteration (see Liu and Wu (1999); Meng and van Dyk (1999); Marchev and Hobert (2004); Hobert, Roy and Robert (2011) for examples). The Haar PX-DA algorithm introduced by Liu and Wu (1999), and generalized by Hobert and Marchev (2008), has been shown by these authors to be the best among a class of sandwich algorithms in terms of efficiency and operator norm. In this section, we construct a sandwich Markov chain by adapting the Haar PX-DA algorithm in the current setting.

We start by making appropriate choices for all the necessary ingredients for the Haar PX-DA algorithm in the current context, and then combining these ingredients to construct the sandwich Markov chain. In the context of the DA algorithm (two-block Gibbs sampler) described in Section 3, let us consider $\mathcal{V} := \mathbb{R}_+^p \times \mathbb{R}_+^p$, the space of all possible values of $(\boldsymbol{\tau}, \boldsymbol{\lambda})$, and $\mathcal{U} := \mathbb{R}^p \times \mathbb{R}_+$, the space of all possible values of $(\boldsymbol{\beta}, \sigma^2)$. Let G denote the multiplicative group of positive real numbers with identity element $e = 1$. Note that G is a unimodular group with Haar measure $\mathcal{H}(dg) = \frac{dg}{g}$. Consider a group action (from the left) of G on the set $\mathcal{V}$ given the following function:

$$g \star (\boldsymbol{\tau}, \boldsymbol{\lambda}) = (g\boldsymbol{\tau}, \boldsymbol{\lambda})$$

where $g\boldsymbol{\tau} = (g\tau_1, g\tau_2, \ldots, g\tau_p)$ denotes the scalar multiplication of $\boldsymbol{\tau}$ by the number g. Consider the function $\chi : G \to \mathbb{R}_+$ defined by $\chi(g) = g^p$. Note that

$$\chi(g_1 g_2) = \chi(g_1)\chi(g_2) = g_1^p g_2^p$$

for all $g_1, g_2 \in G$, and

$$\chi\left(g\right) \int_{\mathcal{V}} \phi\left(g \star v\right) dv = \int_{\mathcal{V}} \phi\left(v\right) dv$$

for any $g \in G$ and any real valued integrable function ϕ on $\mathcal{V}$. Hence, χ is a multiplier function, and the Lebesgue measure on $\mathcal{V}$ is relatively left invariant with respect to χ (Edwards, 1995, Page 252). Using the quantities defined above, based on Hobert and Marchev (2008)'s recipe, we now define the density f_G on G (with respect to the Haar measure) by

$$f_G(g)dg = \frac{\pi(g\boldsymbol{\tau}, \boldsymbol{\lambda})\chi(g)}{m\left(\boldsymbol{\tau}, \boldsymbol{\lambda}\right)}\mathcal{H}(dg), \tag{19}$$

where $m\left(\boldsymbol{\tau}, \boldsymbol{\lambda}\right) = \int_G \pi(g\boldsymbol{\tau}, \boldsymbol{\lambda})\chi(g)\mathcal{H}(dg)$ is the normalizing constant. From (A.1), we get that

$$\pi(\boldsymbol{\tau}, \boldsymbol{\lambda}) \propto \frac{\Pi_{j=1}^{p} \lambda_j^{1+\zeta} \tau_j^{\frac{1}{2}-1}\, e^{-\frac{\lambda_j^2 \tau_j}{2} - \eta\lambda_j}}{\left\{\mathbf{y}^T\mathbf{y} - \mathbf{y}^T X^T \left(X^T X + D_{\boldsymbol{\tau}}^{-1}\right)^{-1} X^T\mathbf{y} + 2\xi\right\}^{\frac{n}{2}+\alpha} |X^T X + D_{\boldsymbol{\tau}}^{-1}|^{\frac{1}{2}}}. \tag{20}$$

It follows from (19) and (20) that

$$f_G(g) \quad \propto \quad \frac{g^{p/2-1}\, e^{-g\left(\sum_{j=1}^{p} \frac{\lambda_j^2 \tau_j}{2}\right)}}{\left\{\mathbf{y}^T\mathbf{y} - \mathbf{y}^T X^T \left(X^T X + \frac{1}{g} D_{\boldsymbol{\tau}}^{-1}\right)^{-1} X^T\mathbf{y} + 2\xi\right\}^{\frac{n}{2}+\alpha} |X^T X + \frac{1}{g} D_{\boldsymbol{\tau}}^{-1}|^{\frac{1}{2}}}. \tag{21}$$

Note that even if f_G is not a standard distribution, it is a univariate density. An efficient and straightforward rejection sampler algorithm to sample from f_G has been provided in the appendix. Using f_G, we can now define the sandwich Markov chain, denoted by $\Phi^* = \{(\boldsymbol{\beta}_m, \sigma_m^2)\}_{m=0}^{\infty}$, whose one step transition from $(\boldsymbol{\beta}_m, \sigma_m^2)$ to $(\boldsymbol{\beta}_{m+1}, \sigma_{m+1}^2)$ can be described as follows.

Iteration $(m+1)$ of the Gibbs sampler:

1. *Draw $(\boldsymbol{\tau}, \boldsymbol{\lambda})$ by the following method*
   (a) *Draw $\boldsymbol{\lambda}$ from the distribution $\pi(\cdot \mid \sigma_m^2, \boldsymbol{\beta}_m, \mathbf{y})$*
   (b) *Draw $\boldsymbol{\tau}$ from the distribution $\pi(\cdot \mid \boldsymbol{\lambda}, \sigma_m^2, \boldsymbol{\beta}_m, \mathbf{y})$*
2. *Draw g according to the density f_G.*
3. *Draw $(\sigma_{m+1}^2, \boldsymbol{\beta}_{m+1})$ by the following procedure*
   (a) *Draw σ_{m+1}^2 from $\pi(\cdot \mid g\boldsymbol{\tau}, \boldsymbol{\lambda}, \mathbf{y})$.*
   (b) *Draw $\boldsymbol{\beta}_{m+1}$ from $\pi(\cdot \mid g\boldsymbol{\tau}, \boldsymbol{\lambda}, \sigma_{m+1}^2, \mathbf{y})$.*

Note that the only difference between the DA Markov chain (two-block Gibbs sampler) and the sandwich Markov chain described above is the univariate draw from the density f_G. The rejection sampler for f_G can be inefficient when *both* n and p are large, but in all other cases provides efficient and inexpensive draws from f_G. As we will see in Section 5, adding this inexpensive univariate draw can lead to significant improvement in convergence and efficiency. Also, the following result follows immediately from Theorem 1 and results in Khare and Hobert (2011).

Corollary 1 *If $p < n$, $rank(X) = p$ and $\mathbf{y}^T(I_n - P_X)\mathbf{y} > 0$, then the Markov operator corresponding to the sandwich chain Φ^* is trace class. Moreover, each eigenvalue of this operator is less than or equal to by the corresponding eigenvalue of the DA Markov operator K, with at least one strict domination.*

5. Examples

In this section, we consider two simulated data examples (one each for $n > p$ and $n < p$) and a real data example, to compare the performance (in terms of convergence and efficiency) for all the three Markov chains discussed in this paper, the three-block Gibbs sampler in Armagan, Dunson and Lee (2013), the proposed two-block Gibbs sampler in Section 3, and the sandwich Markov chain derived in Section 4.

5.1. Simulations

We consider a setting with $n = 15 < p = 26$ for the first simulation, and $n = 25 > p = 20$ for the second simulation. For both cases, the respective datasets are generated using a linear model with only three (true) regression coefficients chosen to be non zero. The elements of the design matrix X were chosen by generating i.i.d. $N(0, 1)$ random variables. For both subsequently generated datasets, we fit the generalized double Pareto model in (1) with hyper parameters $\eta = \zeta = 1$ and $\xi = \alpha = 0$. To compare the efficiency performance of the Markov chains, we compute the autocorrelations (up to lag 10) for all the Markov chains for the function $(\mathbf{y} - X\boldsymbol{\beta})^T(\mathbf{y} - X\boldsymbol{\beta}) + \sigma^2$. The results are summarized in Figure 1 for the first simulation, and in Figure 2 for the second simulation. We can clearly see that for both datasets, the two-block Gibbs sampler has significantly lower autocorrelations than the three-block Gibbs sampler, and that the magnitude of the autocorrelations for the sandwich Markov chain is lowest. We also computed the autocorrelations for individual coordinates of $\boldsymbol{\beta}$, and they follow a similar pattern. A related measure of performance for Markov chains is the effective sample size. We use two different methods (from Kass et al. (1998) and Gong and Flegal (2014)) to compute/estimate

	Simulation $n < p$		Simulation $n > p$		Real data	
ESS	Method 1	Method 2	Method 1	Method 2	Method 1	Method 2
Three block	7.75	9.19	35.73	30.56	25.82	19.11
Two block DA	13.55	13.94	58.38	61.31	45.79	31.65
Sandwich	23.98	30.16	73.36	81.89	73.41	71.21

TABLE 1

Effective sample sizes (ESS) out of 100 for the three-block, two-block and sandwich Markov chains

the effective sample sizes for each of the three Markov chains in both simulations. These effective sample sizes for the three Markov chains (out of 100) for both simulations are provided in the first four columns of Table 1. While estimating eigenvalues and trace of Markov chains (if they exist) is not possible/feasible in general, these results provide strong evidence that the two-block Gibbs sampler and the sandwich Markov chain are much more efficient than the three-block Gibbs sampler.

5.2. Real data example

In this section, we consider the wheat data set from Perez and de los Campos (2014). The data was obtained from a study which included numerous international trials across a wide variety of wheat-producing environments. The different environmental conditions specified in these trials were grouped into four basic sets of environmental categories involving four main agro-climatic regions. The phenotypic traits, or responses, considered here were the average grain yield (GY) of the wheat lines evaluated in each of these four categories of environments. The information on the genotypes of the corresponding ”Wheat lines”, i.e. the binary variables regarding the presence of the genotypes are also available in the data set. The data set is available in the R package *BGLR*, and more details can be found in Perez and de los Campos (2014).

For our analysis, we consider the average grain yield for a particular environmental condition (there are four to choose from) as the response variable, and 40 binary variables containing genotypic information as the predictors. We fit the generalized double Pareto model in (1) with hyper parameters $\eta = \zeta = 1$ and $\xi = \alpha = 0$. As with the simulated datasets, we compute the autocorrelations (up to lag 10) for all the three Markov chains for the function $(\mathbf{y} - X\boldsymbol{\beta})^T(\mathbf{y} - X\boldsymbol{\beta}) + \sigma^2$ (see Figure 3). Effective sample sizes are also computed and are reported in the last two columns of Table 1. We see that the two-block Gibbs sampler and the sandwich Markov chain are significantly more efficient than the three block Gibbs sampler.

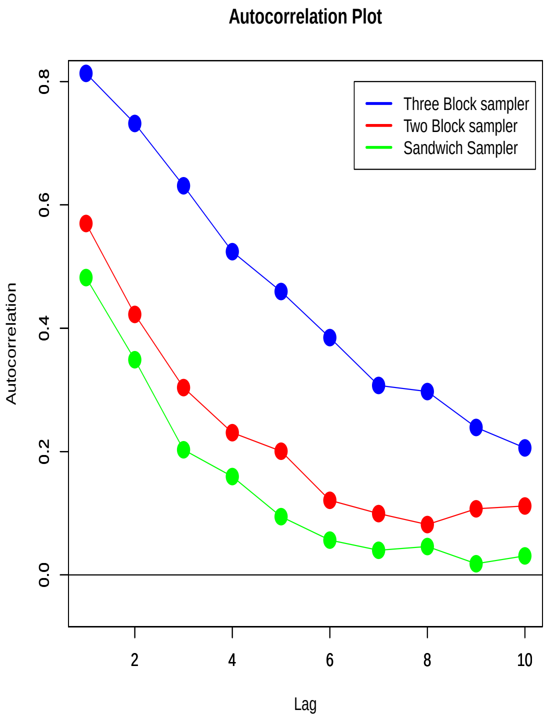

Fig 1. *Autocorrelation plot for the function* $(\mathbf{y} - X\boldsymbol{\beta})^T(\mathbf{y} - X\boldsymbol{\beta}) + \sigma^2$ *for the simulated data set with* $n < p$.

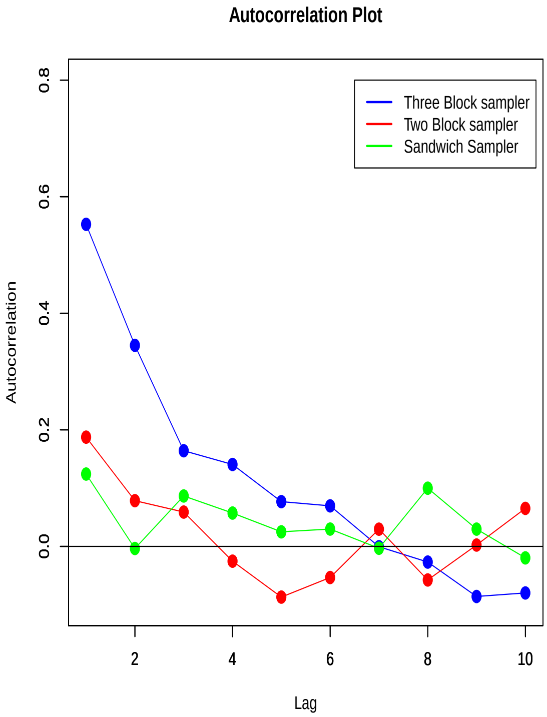

FIG 2. *Autocorrelation plot for the function* $(\mathbf{y} - X\boldsymbol{\beta})^T(\mathbf{y} - X\boldsymbol{\beta}) + \sigma^2$ *for the simulated data set with* $n > p$.

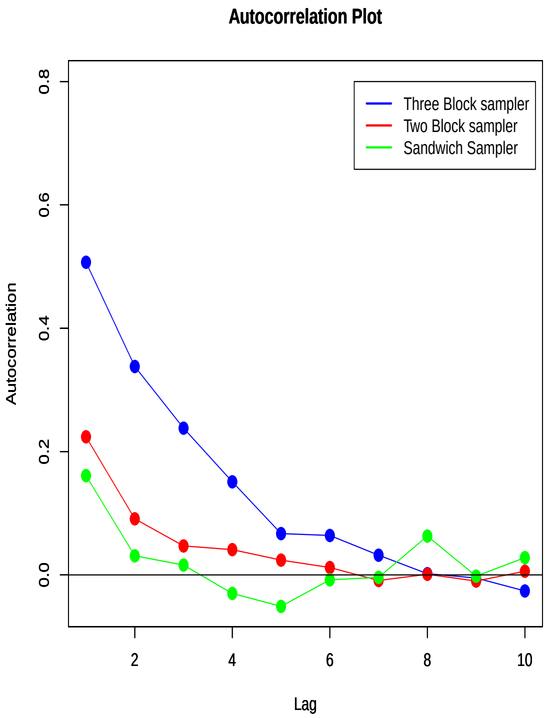

FIG 3. *Autocorrelation plot for the function* $(\mathbf{y} - X\boldsymbol{\beta})^T(\mathbf{y} - X\boldsymbol{\beta}) + \sigma^2$ *for the wheat data.*

References

Armagan, A., Dunson, D. B. and Lee, J. (2013). Generalized double Pareto shrinkage. *Statistica Sinica* **23** 119–143.

Asmussen, S. and Glynn, P. W. (2011). A new proof of convergence of MCMC via the ergodic theorem. *Statistics and Probability Letters* **81** 1482-1485.

Bhattacharya, A., Pati, D., Pillai, N. S. and Dunson, D. B. (2015). Dirichlet-Laplace priors for optimal shrinkage. *J. Amer. Statist. Assoc.* **110** 479-490.

Cohn, H. A. (2013). *Measure Theory.* Springer, New York, NY, USA.

Edwards, R. E. (1995). *Functional Analysis: Theory and Applications.* Dover Publications, New York.

Gong, L. and Flegal, J. M. (2014). A practical sequential stopping rule for high-dimensional MCMC and its application to spatial-temporal Bayesian models. *arxiv.*

Hobert, J. P. and Marchev, D. (2008). A theoretical comparison of the data augmentation, marginal augmentation and PX-DA algorithms. *Annals of Statistics* **36** 532-554.

Hobert, J. P., Roy, V. and Robert, C. P. (2011). Improving the convergence properties of the data augmentation algorithm with an application to Bayesian mixture modeling. *Statistical Science* **26** 332-351.

Jones, G. L. and Hobert, J. P. (2001). Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Statistical Science* **16** 312-334.

Jorgens, K. (1982). *Linear Integral Operators.* Pitman Books, London.

Kass, R. E., Carlin, B. P., Gelman, A. and Neal, R. M. (1998). Markov chain Monte Carlo in practice: A roundtable discussion. *The American Statistician* **52** 93-100.

Khare, K. and Hobert, J. P. (2011). A spectral analytic comparison of trace-class data augmentation algorithms and their sandwich variants. *Annals of Statistics* **39** 2585–2606.

Kotz, S. and Nadarajah, S. (2004). *Multivariate t Distributions and Their Applications. Cambridge University Press.* Cambridge University Press.

Liu, J. S., Wong, W. H. and Kong, A. (1994). Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika* **81** 27-40.

Liu, J. S. and Wu, Y. N. (1999). Parameter expansion for data augmentation. *J. Amer. Statist. Assoc.* **94** 1264-1274.

Marchev, D. and Hobert, J. P. (2004). Geometric ergodicity of van Dyk and Meng's algorithm for the multivariate Student's t model. *Journal of the American Statistical Association* **99** 228-238.

Meng, X. L. and van Dyk, D. A. (1999). Seeking efficient data augmentation schemes via conditional and marginal augmentation. *Biometrika* **86** 301-320.

Meyn, S. P. and Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer-Verlag, London.

Perez, P. and de los Campos, G. (2014). Genome-Wide regression and prediction with the BGLR statistical package. *Genetics* **198** 483-495.

Polson, N. G. and Scott, J. G. (2010). Shrink globally, act locally: Sparse Bayesian regularization and prediction. *Bayesian Statistics* **9** 501-538.

Roberts, G. O. and Rosenthal, J. S. (1997). Geometric ergodicity of hybrid Markov chains. *Electronic Communications in Probability* **2** 13-25.

Rosenthal, J. S. (1995). Minorization conditions and convergence rates for Markov chain Monte Carlo. *Journal of the American Statistical Association* **90** 558-566.

Steele, J. M. (2004). *The Cauchy-Schwarz Master Class: An Introduction to the Art of Mathematical Inequalities*. Cambridge University Press.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B* **58** 267-288.

Supplemental Document for "Trace class Markov chains for Bayesian inference with generalized double Pareto shrinkage priors"

A. Form of relevant densities

In this section, we provide the form of various relevant densities corresponding to the Bayesian shrinkage model in (1). These are required for constructing the Gibbs sampling Markov chains, and for the subsequent analysis. The posterior density of $(\boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\lambda}, \sigma^2)$ conditioned on the observed data $\mathbf{y}$ is given by

$$\pi(\boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\lambda}, \sigma^2 \mid \mathbf{y}) \propto \frac{e^{-\frac{(\mathbf{y}-X\boldsymbol{\beta})^T(\mathbf{y}-X\boldsymbol{\beta})}{2\sigma^2}}}{(\sqrt{2\pi})^n \sigma^n} \frac{e^{-\frac{\boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1} \boldsymbol{\beta}}{2\sigma^2}}}{(\sqrt{2\pi})^p \sigma^p} \left(\prod_{j=1}^{p} \tau_j^{\frac{1}{2}-1} e^{-\frac{\lambda_j^2}{2}\tau_j} \lambda_j^{\zeta+1} e^{-\eta\lambda_j} \right) (\sigma^2)^{-\alpha-1} e^{-\frac{\xi}{\sigma^2}} \tag{A.1}$$

for every $(\boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\lambda}, \sigma^2) \in \mathbb{R}^p \times \mathbb{R}_+^p \times \mathbb{R}_+^p \times \mathbb{R}_+$. Based on the joint density in (A.1), the following conditional distributions can be derived in a straightforward fashion.

- The conditional density of $\boldsymbol{\beta}$ given $\boldsymbol{\tau}, \boldsymbol{\lambda}, \sigma^2, \mathbf{y}$ is the

  $$N_p\left((X^TX + D_{\boldsymbol{\tau}}^{-1})^{-1} X^T \mathbf{y}, \sigma^2 (X^TX + D_{\boldsymbol{\tau}}^{-1})^{-1}\right)$$

  density on $\mathbb{R}^p$. In particular,

  $$\pi(\boldsymbol{\beta} \mid \boldsymbol{\tau}, \boldsymbol{\lambda}, \sigma^2, \mathbf{y}) = \frac{|X^TX + D_{\boldsymbol{\tau}}^{-1}|^{\frac{1}{2}}}{(\sqrt{2\pi})^p \sigma^p} e^{-\frac{\left(\boldsymbol{\beta}-(X^TX+D_{\boldsymbol{\tau}}^{-1})^{-1}X^T\mathbf{y}\right)^T (X^TX+D_{\boldsymbol{\tau}}^{-1}) \left(\boldsymbol{\beta}-(X^TX+D_{\boldsymbol{\tau}}^{-1})^{-1}X^T\mathbf{y}\right)}{2\sigma^2}}, \tag{A.2}$$

  for $\boldsymbol{\beta} \in \mathbb{R}^p$.

- The conditional density of σ^2 given $\boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\lambda}, \mathbf{y}$ is the

  $$\text{Inverse-Gamma}\left(\frac{n+p+2\alpha}{2}, \frac{(\mathbf{y}-X\boldsymbol{\beta})^T(\mathbf{y}-X\boldsymbol{\beta}) + \boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1}\boldsymbol{\beta} + 2\xi}{2}\right)$$

  density. In particular,

  $$\begin{aligned} \pi(\sigma^2 \mid \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\lambda}, \mathbf{y}) \quad = \quad & \frac{\left((\mathbf{y}-X\boldsymbol{\beta})^T(\mathbf{y}-X\boldsymbol{\beta}) + \boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1}\boldsymbol{\beta} + 2\xi\right)^{\frac{n+p+2\alpha}{2}}}{2^{\frac{n+p+2\alpha}{2}} \Gamma(\frac{n+p+2\alpha}{2})} (\sigma^2)^{-\frac{n+p+2\alpha}{2}-1} \\ & \times e^{-\frac{(\mathbf{y}-X\boldsymbol{\beta})^T(\mathbf{y}-X\boldsymbol{\beta}) + \boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1}\boldsymbol{\beta} + 2\xi}{2\sigma^2}}, \end{aligned} \tag{A.3}$$

  for $\sigma^2 \in \mathbb{R}_+$.

- The conditional density of σ^2 given $\boldsymbol{\tau}, \boldsymbol{\lambda}, \mathbf{y}$ is the

  $$\text{Inverse-Gamma}\left(\frac{n+2\alpha}{2}, \frac{(\mathbf{y}-X\widehat{\boldsymbol{\beta}})^T(\mathbf{y}-X\widehat{\boldsymbol{\beta}}) + \widehat{\boldsymbol{\beta}}^T D_{\boldsymbol{\tau}}^{-1}\widehat{\boldsymbol{\beta}} + 2\xi}{2}\right)$$

  density, where $\widehat{\boldsymbol{\beta}} = (X^TX + D_{\boldsymbol{\tau}}^{-1})^{-1}X^T\mathbf{y}$. In particular,

  $$\begin{aligned}\pi(\sigma^2 \mid \boldsymbol{\tau}, \boldsymbol{\lambda}, \mathbf{y}) \quad = \quad & \frac{\left((\mathbf{y}-X\widehat{\boldsymbol{\beta}})^T(\mathbf{y}-X\widehat{\boldsymbol{\beta}}) + \widehat{\boldsymbol{\beta}}^T D_{\boldsymbol{\tau}}^{-1}\widehat{\boldsymbol{\beta}} + 2\xi\right)^{\frac{n+2\alpha}{2}}}{2^{\frac{n+2\alpha}{2}}\Gamma(\frac{n+2\alpha}{2})}(\sigma^2)^{-\frac{n+2\alpha}{2}-1} \\ & \times e^{-\frac{(\mathbf{y}-X\widehat{\boldsymbol{\beta}})^T(\mathbf{y}-X\widehat{\boldsymbol{\beta}}) + \widehat{\boldsymbol{\beta}}^T D_{\boldsymbol{\tau}}^{-1}\widehat{\boldsymbol{\beta}} + 2\xi}{2\sigma^2}},\end{aligned} \tag{A.4}$$

  for $\sigma^2 \in \mathbb{R}_+$.

- Given $\boldsymbol{\beta}, \sigma^2, \boldsymbol{\lambda}$ and $\mathbf{y}$, the variables $\tau_1, \tau_2, \cdots, \tau_p$ are conditionally independent, and the conditional density of τ_j given $\boldsymbol{\beta}, \sigma^2, \boldsymbol{\lambda}$ and $\mathbf{y}$ is Generalized Inverse Gaussian $\left(\frac{1}{2}, \lambda_j^2, \frac{\beta_j^2}{\sigma^2}\right)$. In particular

  $$\begin{aligned}\pi(\boldsymbol{\tau} \mid \boldsymbol{\beta}, \sigma^2, \boldsymbol{\lambda}, \mathbf{y}) \quad &= \quad \prod_{j=1}^p \left(\frac{\lambda_j\sigma}{|\beta_j|}\right)^{\frac{1}{2}} \frac{1}{2K_{\frac{1}{2}}\left(\frac{\lambda_j|\beta_j|}{\sigma}\right)}\tau_j^{\frac{1}{2}-1}e^{-\frac{1}{2}\left\{\lambda_j^2\tau_j + \frac{\beta_j^2}{\sigma^2}\frac{1}{\tau_j}\right\}} \\ &= \quad \prod_{j=1}^p \frac{1}{\sqrt{2\pi}}\lambda_j\tau_j^{\frac{1}{2}-1}e^{-\frac{1}{2}\left\{\lambda_j^2\tau_j + \frac{\beta_j^2}{\sigma^2}\frac{1}{\tau_j}\right\}}e^{\frac{\lambda_j|\beta_j|}{\sigma}}\end{aligned} \tag{A.5}$$

  for $\boldsymbol{\tau} \in \mathbb{R}_+^p$.

- Given $\boldsymbol{\beta}, \sigma^2$ and $\mathbf{y}$, the variables $\lambda_1, \lambda_2, \cdots, \lambda_p$ are conditionally independent, and the conditional density of λ_j given $\boldsymbol{\beta}, \sigma^2$ and $\mathbf{y}$ is Gamma $\left(\zeta + 1, \frac{|\beta_j|}{\sigma} + \eta\right)$. In particular

  $$\pi(\boldsymbol{\lambda} \mid \boldsymbol{\beta}, \sigma^2, \mathbf{y}) \quad = \quad \prod_{j=1}^p \frac{\left(\frac{|\beta_j|}{\sigma} + \eta\right)^{\zeta+1}\lambda_j^\zeta}{\Gamma(\zeta+1)} \; e^{-\left(\frac{|\beta_j|}{\sigma}+\eta\right)\lambda_j} \tag{A.6}$$

  for $\boldsymbol{\lambda} \in \mathbb{R}_+^p$.

Note that samples can be easily generated from all the conditional densities in (A.2), (A.3), (A.4) and (A.5) by using standard statistical software (such as R).

B. Proof of Theorem 2

Proof By (4), to prove the required result, we need to show that

$$\tilde{I} := \int_{\mathbb{R}^p}\int_{\mathbb{R}_+}\int_{\mathbb{R}^p}\int_{\mathbb{R}_+} \tilde{k}^2\left((\boldsymbol{\beta},\sigma^2),(\check{\boldsymbol{\beta}},\check{\sigma}^2)\right)\frac{\pi\left(\boldsymbol{\beta},\sigma^2 \mid \mathbf{y}\right)}{\pi\left(\check{\boldsymbol{\beta}},\check{\sigma}^2 \mid \mathbf{y}\right)} d\boldsymbol{\beta}\, d\sigma^2\, d\check{\boldsymbol{\beta}}\, d\check{\sigma}^2 = \infty. \tag{B.7}$$

From (2) we get that

$$\begin{aligned}
&\tilde{k}^2\left((\boldsymbol{\beta},\sigma^2),(\check{\boldsymbol{\beta}},\check{\sigma}^2)\right)\\
=&\left[\int_{\mathbb{R}_+^p}\int_{\mathbb{R}_+^p}\pi\left(\check{\sigma}^2 \mid \check{\boldsymbol{\beta}},\boldsymbol{\tau},\boldsymbol{\lambda},\mathbf{y}\right)\pi\left(\check{\boldsymbol{\beta}} \mid \boldsymbol{\tau},\boldsymbol{\lambda},\sigma^2,\mathbf{y}\right)\pi\left(\boldsymbol{\tau},\boldsymbol{\lambda} \mid \boldsymbol{\beta},\sigma^2,\mathbf{y}\right)\, d\boldsymbol{\tau}\, d\boldsymbol{\lambda}\right]^2\\
=&\int_{\mathbb{R}_+^p}\int_{\mathbb{R}_+^p}\int_{\mathbb{R}_+^p}\int_{\mathbb{R}_+^p}\pi\left(\check{\sigma}^2 \mid \check{\boldsymbol{\beta}},\boldsymbol{\tau},\boldsymbol{\lambda},\mathbf{y}\right)\pi\left(\check{\boldsymbol{\beta}} \mid \boldsymbol{\tau},\boldsymbol{\lambda},\sigma^2,\mathbf{y}\right)\pi\left(\boldsymbol{\tau},\boldsymbol{\lambda} \mid \boldsymbol{\beta},\sigma^2,\mathbf{y}\right)\\
&\pi\left(\check{\sigma}^2 \mid \check{\boldsymbol{\beta}},\check{\boldsymbol{\tau}},\check{\boldsymbol{\lambda}},\mathbf{y}\right)\pi\left(\check{\boldsymbol{\beta}} \mid \check{\boldsymbol{\tau}},\check{\boldsymbol{\lambda}},\sigma^2,\mathbf{y}\right)\pi\left(\check{\boldsymbol{\tau}},\check{\boldsymbol{\lambda}} \mid \boldsymbol{\beta},\sigma^2,\mathbf{y}\right)\, d\boldsymbol{\tau}\, d\boldsymbol{\lambda} d\check{\boldsymbol{\tau}}\, d\check{\boldsymbol{\lambda}}.
\end{aligned} \tag{B.8}$$

It follows from (B.8) that

$$\begin{aligned}
\tilde{I} =& \int_{\mathbb{R}^p}\int_{\mathbb{R}_+}\int_{\mathbb{R}^p}\int_{\mathbb{R}_+}\int_{\mathbb{R}_+^p}\int_{\mathbb{R}_+^p}\int_{\mathbb{R}_+^p}\int_{\mathbb{R}_+^p}\pi\left(\check{\sigma}^2 \mid \check{\boldsymbol{\beta}},\boldsymbol{\tau},\boldsymbol{\lambda},\mathbf{y}\right)\pi\left(\check{\boldsymbol{\beta}} \mid \boldsymbol{\tau},\boldsymbol{\lambda},\sigma^2,\mathbf{y}\right)\pi\left(\boldsymbol{\tau},\boldsymbol{\lambda} \mid \boldsymbol{\beta},\sigma^2,\mathbf{y}\right)\\
&\pi\left(\check{\sigma}^2 \mid \check{\boldsymbol{\beta}},\check{\boldsymbol{\tau}},\check{\boldsymbol{\lambda}},\mathbf{y}\right)\pi\left(\check{\boldsymbol{\beta}} \mid \check{\boldsymbol{\tau}},\check{\boldsymbol{\lambda}},\sigma^2,\mathbf{y}\right)\pi\left(\check{\boldsymbol{\tau}},\check{\boldsymbol{\lambda}} \mid \boldsymbol{\beta},\sigma^2,\mathbf{y}\right)\frac{\pi\left(\boldsymbol{\beta},\sigma^2 \mid \mathbf{y}\right)}{\pi\left(\check{\boldsymbol{\beta}},\check{\sigma}^2 \mid \mathbf{y}\right)}\\
&d\boldsymbol{\tau}\, d\boldsymbol{\lambda} d\check{\boldsymbol{\tau}}\, d\check{\boldsymbol{\lambda}}\, d\boldsymbol{\beta}\, d\sigma^2\, d\check{\boldsymbol{\beta}}\, d\check{\sigma}^2.
\end{aligned} \tag{B.9}$$

Now, a straightforward rearrangement of conditional densities shows that

$$\begin{aligned}
&\pi\left(\check{\sigma}^2 \mid \check{\boldsymbol{\beta}},\check{\boldsymbol{\tau}},\check{\boldsymbol{\lambda}},\mathbf{y}\right)\pi\left(\check{\boldsymbol{\beta}} \mid \check{\boldsymbol{\tau}},\check{\boldsymbol{\lambda}},\sigma^2,\mathbf{y}\right)\pi\left(\check{\boldsymbol{\tau}},\check{\boldsymbol{\lambda}} \mid \boldsymbol{\beta},\sigma^2,\mathbf{y}\right)\frac{\pi\left(\boldsymbol{\beta},\sigma^2 \mid \mathbf{y}\right)}{\pi\left(\check{\boldsymbol{\beta}},\check{\sigma}^2 \mid \mathbf{y}\right)}\\
=\;&\pi\left(\boldsymbol{\beta} \mid \check{\boldsymbol{\tau}},\check{\boldsymbol{\lambda}},\sigma^2,\mathbf{y}\right)\pi\left(\sigma^2 \mid \check{\boldsymbol{\beta}},\check{\boldsymbol{\tau}},\check{\boldsymbol{\lambda}},\mathbf{y}\right)\pi\left(\check{\boldsymbol{\tau}},\check{\boldsymbol{\lambda}} \mid \check{\boldsymbol{\beta}},\check{\sigma}^2,\mathbf{y}\right).
\end{aligned}$$

It follows from (B.9), and by using Fubini's theorem (for exchanging the order of integration) that

$$\begin{aligned}
\tilde{I} =& \int_{\mathbb{R}_+^p}\int_{\mathbb{R}_+^p}\int_{\mathbb{R}_+^p}\int_{\mathbb{R}_+^p}\int_{\mathbb{R}^p}\int_{\mathbb{R}^p}\int_{\mathbb{R}_+}\int_{\mathbb{R}_+}\pi\left(\check{\sigma}^2 \mid \check{\boldsymbol{\beta}},\boldsymbol{\tau},\boldsymbol{\lambda},\mathbf{y}\right)\pi\left(\check{\boldsymbol{\beta}} \mid \boldsymbol{\tau},\boldsymbol{\lambda},\sigma^2,\mathbf{y}\right)\pi\left(\boldsymbol{\tau},\boldsymbol{\lambda} \mid \boldsymbol{\beta},\sigma^2,\mathbf{y}\right)\\
&\pi\left(\boldsymbol{\beta} \mid \check{\boldsymbol{\tau}},\check{\boldsymbol{\lambda}},\sigma^2,\mathbf{y}\right)\pi\left(\sigma^2 \mid \check{\boldsymbol{\beta}},\check{\boldsymbol{\tau}},\check{\boldsymbol{\lambda}},\mathbf{y}\right)\pi\left(\check{\boldsymbol{\tau}},\check{\boldsymbol{\lambda}} \mid \check{\boldsymbol{\beta}},\check{\sigma}^2,\mathbf{y}\right)d\sigma^2\, d\check{\sigma}^2\, d\boldsymbol{\beta}\, d\check{\boldsymbol{\beta}}\, d\boldsymbol{\tau}\, d\boldsymbol{\lambda}\, d\check{\boldsymbol{\tau}}\, d\check{\boldsymbol{\lambda}}
\end{aligned} \tag{B.10}$$

We will now show that for arbitrarily fixed $\boldsymbol{\tau},\boldsymbol{\lambda},\check{\boldsymbol{\tau}},\check{\boldsymbol{\lambda}}$, the integrand in (B.10) has an infinite integral as a function of $\boldsymbol{\beta},\sigma^2,\check{\boldsymbol{\beta}},\check{\sigma}^2$. This will be done by obtaining an appropriate lower bound, and then integrating out $\sigma^2, \check{\sigma}^2$ and $\boldsymbol{\beta}$ using the Inverse-Gamma and t-densities. Finally, it will be shown, using the properties of the t-density again, that the resulting function of $\check{\boldsymbol{\beta}}$ has infinite integral over $\mathbb{R}^p$. For ease of exposition, we will use the following notations in the subsequent analysis.

$$\begin{aligned}
\widehat{\boldsymbol{\beta}} &= (X^TX + D_{\boldsymbol{\tau}}^{-1})^{-1}X^T\mathbf{y} & \widehat{\boldsymbol{\beta}}_* &= (X^TX + D_{\check{\boldsymbol{\tau}}}^{-1})^{-1}X^T\mathbf{y} \\
\tilde{\Delta}_1 &= \left(\check{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}\right)^T (X^TX + D_{\boldsymbol{\tau}}^{-1}) \left(\check{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}\right) & \Delta_{1*} &= \left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_*\right)^T (X^TX + D_{\check{\boldsymbol{\tau}}}^{-1}) \left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_*\right) \\
\tilde{\Delta} &= (\mathbf{y} - X\check{\boldsymbol{\beta}})^T(\mathbf{y} - X\check{\boldsymbol{\beta}}) + \check{\boldsymbol{\beta}}^T D_{\boldsymbol{\tau}}^{-1}\check{\boldsymbol{\beta}} + 2\xi & \tilde{\Delta}_* &= (\mathbf{y} - X\check{\boldsymbol{\beta}})^T(\mathbf{y} - X\check{\boldsymbol{\beta}}) + \check{\boldsymbol{\beta}}^T D_{\check{\boldsymbol{\tau}}}^{-1}\check{\boldsymbol{\beta}} + 2\xi
\end{aligned} \tag{B.11}$$

From (A.2), (A.3), (A.5), (A.6) and (B.11), we get that

$$\begin{aligned}
& \pi\left(\check{\sigma}^2 \mid \check{\boldsymbol{\beta}}, \boldsymbol{\tau}, \boldsymbol{\lambda}, \mathbf{y}\right) \pi\left(\check{\boldsymbol{\beta}} \mid \boldsymbol{\tau}, \boldsymbol{\lambda}, \sigma^2, \mathbf{y}\right) \pi\left(\boldsymbol{\tau}, \boldsymbol{\lambda} \mid \boldsymbol{\beta}, \sigma^2, \mathbf{y}\right) \\
& \pi\left(\boldsymbol{\beta} \mid \check{\boldsymbol{\tau}}, \check{\boldsymbol{\lambda}}, \sigma^2, \mathbf{y}\right) \pi\left(\sigma^2 \mid \check{\boldsymbol{\beta}}, \check{\boldsymbol{\tau}}, \check{\boldsymbol{\lambda}}, \mathbf{y}\right) \pi\left(\check{\boldsymbol{\tau}}, \check{\boldsymbol{\lambda}} \mid \check{\boldsymbol{\beta}}, \check{\sigma}^2, \mathbf{y}\right) \\
= \quad & C_1 \left\{ \frac{\tilde{\Delta}^{\frac{n+p+2\alpha}{2}} e^{-\frac{\tilde{\Delta}}{2\check{\sigma}^2}}}{(\check{\sigma}^2)^{\frac{n+p+2\alpha}{2}+1}} \right\} \left\{ \frac{|X^TX + D_{\boldsymbol{\tau}}^{-1}|^{\frac{1}{2}} e^{-\frac{\tilde{\Delta}_1}{2\sigma^2}}}{\sigma^p} \right\} \left\{ \prod_{j=1}^{p} \lambda_j^{\zeta+1} \tau_j^{-\frac{1}{2}} e^{-\frac{1}{2}\left(\lambda_j^2\tau_j + \frac{\beta_j^2}{\sigma^2\tau_j}\right) - \eta\lambda_j} \left(\eta + \frac{|\beta_j|}{\sigma}\right)^{\zeta+1} \right\} \\
& \left\{ \frac{|X^TX + D_{\check{\boldsymbol{\tau}}}^{-1}|^{\frac{1}{2}} e^{-\frac{\Delta_{1*}}{2\sigma^2}}}{\sigma^p} \right\} \left\{ \frac{\tilde{\Delta}_*^{\frac{n+p+2\alpha}{2}} e^{-\frac{\tilde{\Delta}_*}{2\sigma^2}}}{(\sigma^2)^{\frac{n+p+2\alpha}{2}+1}} \right\} \left\{ \prod_{j=1}^{p} \check{\lambda}_j^{\zeta+1} \check{\tau}_j^{-\frac{1}{2}} e^{-\frac{1}{2}\left(\check{\lambda}_j^2\check{\tau}_j + \frac{(\check{\beta}_j)^2}{\check{\sigma}^2\check{\tau}_j}\right) - \eta\check{\lambda}_j} \left(\eta + \frac{|\check{\beta}_j|}{\check{\sigma}}\right)^{\zeta+1} \right\}, \\
\geq \quad & C_1\ f_1\left(\boldsymbol{\tau}, \boldsymbol{\lambda}, \check{\boldsymbol{\tau}}, \check{\boldsymbol{\lambda}}\right) \left\{ \frac{\tilde{\Delta}^{\frac{n+p+2\alpha}{2}} e^{-\frac{\tilde{\Delta} + \check{\boldsymbol{\beta}}^T D_{\check{\boldsymbol{\tau}}}^{-1}\check{\boldsymbol{\beta}}}{2\check{\sigma}^2}}}{(\check{\sigma}^2)^{\frac{n+p+2\alpha}{2}+1}} \right\} \times \left\{ \frac{\tilde{\Delta}_*^{\frac{n+p+2\alpha}{2}} e^{-\frac{\tilde{\Delta}_1 + \Delta_{1*} + \tilde{\Delta}_* + \boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1}\boldsymbol{\beta}}{2\sigma^2}}}{(\sigma^2)^{\frac{n+p+2\alpha}{2}+p+1}} \right\}
\end{aligned} \tag{B.12}$$

where

$$C_1 = \left[2^{\frac{n+p+2\alpha}{2}}\Gamma\left(\frac{n+p+2\alpha}{2}\right)\left\{2\pi\Gamma\left(\zeta+1\right)\right\}^p\right]^{-2},$$

and

$$\begin{aligned}
f_1\left(\boldsymbol{\tau}, \boldsymbol{\lambda}, \check{\boldsymbol{\tau}}, \check{\boldsymbol{\lambda}}\right) \quad = \quad & \eta^{2p(\zeta+1)} \left\{ \prod_{j=1}^{p} \lambda_j^{\zeta+1} \tau_j^{-\frac{1}{2}} e^{-\frac{\lambda_j^2\tau_j}{2} - \eta\lambda_j} \right\} \left\{ \prod_{j=1}^{p} \check{\lambda}_j^{\zeta+1} \check{\tau}_j^{-\frac{1}{2}} e^{-\frac{\check{\lambda}_j^2\check{\tau}_j}{2} - \eta\check{\lambda}_j} \right\} \\
& |X^TX + D_{\check{\boldsymbol{\tau}}}^{-1}|^{\frac{1}{2}} |X^TX + D_{\boldsymbol{\tau}}^{-1}|^{\frac{1}{2}}.
\end{aligned}$$

The last inequality is obtained by replacing $\left(\eta + \frac{|\beta_j|}{\sigma}\right)$ and $\left(\eta + \frac{|\check{\beta}_j|}{\check{\sigma}}\right)$ with just η in the appropriate places.

From (B.12), and the form of the Inverse-Gamma density, we get that

$$\begin{aligned}
& \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} \pi\left(\check{\sigma}^2 \mid \check{\boldsymbol{\beta}}, \boldsymbol{\tau}, \boldsymbol{\lambda}, \mathbf{y}\right) \pi\left(\check{\boldsymbol{\beta}} \mid \boldsymbol{\tau}, \boldsymbol{\lambda}, \sigma^2, \mathbf{y}\right) \pi\left(\boldsymbol{\tau}, \boldsymbol{\lambda} \mid \boldsymbol{\beta}, \sigma^2, \mathbf{y}\right) \\
& \pi\left(\boldsymbol{\beta} \mid \check{\boldsymbol{\tau}}, \check{\boldsymbol{\lambda}}, \sigma^2, \mathbf{y}\right) \pi\left(\sigma^2 \mid \check{\boldsymbol{\beta}}, \check{\boldsymbol{\tau}}, \check{\boldsymbol{\lambda}}, \mathbf{y}\right) \pi\left(\check{\boldsymbol{\tau}}, \check{\boldsymbol{\lambda}} \mid \check{\boldsymbol{\beta}}, \check{\sigma}^2, \mathbf{y}\right) \; d\sigma^2 \; d\check{\sigma}^2 \\
\geq \quad & \left[\left\{2\pi\Gamma\left(\zeta+1\right)\right\}^p\right]^{-2} \; f_1\left(\boldsymbol{\tau}, \boldsymbol{\lambda}, \check{\boldsymbol{\tau}}, \check{\boldsymbol{\lambda}}\right) \left\{ \frac{\tilde{\Delta}^{\frac{n+p+2\alpha}{2}}}{\left[\tilde{\Delta} + \check{\boldsymbol{\beta}}^T D_{\check{\boldsymbol{\tau}}}^{-1} \check{\boldsymbol{\beta}}\right]^{\frac{n+p+2\alpha}{2}}} \right\} \times \\
& \left\{ \frac{\tilde{\Delta}_*^{\frac{n+p+2\alpha}{2}}}{\left[\tilde{\Delta}_1 + \Delta_{1*} + \tilde{\Delta}_* + \boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1} \boldsymbol{\beta}\right]^{\frac{n+p+2\alpha}{2}+p}} \right\} \frac{2^p \Gamma\left(\frac{n+p+2\alpha}{2}+p\right)}{\Gamma\left(\frac{n+p+2\alpha}{2}\right)}. \qquad \text{(B.13)}
\end{aligned}$$

Let $\widehat{\boldsymbol{\beta}}_{**} = (X^T X + D_{\check{\boldsymbol{\tau}}}^{-1} + D_{\boldsymbol{\tau}}^{-1})^{-1} X^T \mathbf{y}$. A straightforward computation shows that

$$\begin{aligned}
\Delta_{1*} + \boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1} \boldsymbol{\beta} &= \left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_*\right)^T (X^T X + D_{\check{\boldsymbol{\tau}}}^{-1}) \left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_*\right) + \boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1} \boldsymbol{\beta} \\
&= \left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_{**}\right)^T (X^T X + D_{\check{\boldsymbol{\tau}}}^{-1} + D_{\boldsymbol{\tau}}^{-1}) \left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_{**}\right) + \\
&\quad \left(\widehat{\boldsymbol{\beta}}_{**} - \widehat{\boldsymbol{\beta}}_*\right)^T (X^T X + D_{\check{\boldsymbol{\tau}}}^{-1}) \left(\widehat{\boldsymbol{\beta}}_{**} - \widehat{\boldsymbol{\beta}}_*\right) + \widehat{\boldsymbol{\beta}}_{**}^T D_{\boldsymbol{\tau}}^{-1} \widehat{\boldsymbol{\beta}}_{**} \\
&= f_2\left(\boldsymbol{\tau}, \check{\boldsymbol{\tau}}\right) + \left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_{**}\right)^T (X^T X + D_{\check{\boldsymbol{\tau}}}^{-1} + D_{\boldsymbol{\tau}}^{-1}) \left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_{**}\right), \qquad \text{(B.14)}
\end{aligned}$$

where

$$f_2\left(\boldsymbol{\tau}, \check{\boldsymbol{\tau}}\right) = \left(\widehat{\boldsymbol{\beta}}_{**} - \widehat{\boldsymbol{\beta}}_*\right)^T (X^T X + D_{\check{\boldsymbol{\tau}}}^{-1}) \left(\widehat{\boldsymbol{\beta}}_{**} - \widehat{\boldsymbol{\beta}}_*\right) + \widehat{\boldsymbol{\beta}}_{**}^T D_{\boldsymbol{\tau}}^{-1} \widehat{\boldsymbol{\beta}}_{**}.$$

Hence, by (B.14) and the form of the multivariate-t distribution (see for example Kotz and Nadarajah (2004)), we get that

$$\begin{aligned}
& \int_{\mathbb{R}^p} \frac{1}{\left[\tilde{\Delta}_1 + \Delta_{1*} + \tilde{\Delta}_* + \boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1} \boldsymbol{\beta}\right]^{\frac{n+p+2\alpha}{2}+p}} \, d\boldsymbol{\beta} \\
= \quad & \int_{\mathbb{R}^p} \frac{1}{\left[\tilde{\Delta}_1 + \tilde{\Delta}_* + f_2\left(\boldsymbol{\tau}, \check{\boldsymbol{\tau}}\right) + \left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_{**}\right)^T (X^T X + D_{\check{\boldsymbol{\tau}}}^{-1} + D_{\boldsymbol{\tau}}^{-1}) \left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_{**}\right)\right]^{\frac{n+p+2\alpha}{2}+p}} \, d\boldsymbol{\beta} \\
= \quad & \frac{\Gamma\left(\frac{n+2p+2\alpha}{2}\right) \sqrt{\pi}^p |X^T X + D_{\check{\boldsymbol{\tau}}}^{-1} + D_{\boldsymbol{\tau}}^{-1}|^{-\frac{1}{2}}}{\Gamma\left(\frac{n+p+2\alpha}{2} + p\right) \left[\tilde{\Delta}_1 + \tilde{\Delta}_* + f_2\left(\boldsymbol{\tau}, \check{\boldsymbol{\tau}}\right)\right]^{\frac{n+p+2\alpha}{2}+\frac{p}{2}}}. \qquad \text{(B.15)}
\end{aligned}$$

It follows from (B.13) and (B.15) that

$$\begin{aligned}
&\int_{\mathbb{R}^p}\int_{\mathbb{R}_+}\int_{\mathbb{R}_+} \pi\left(\check{\sigma}^2 \mid \check{\boldsymbol{\beta}}, \boldsymbol{\tau}, \boldsymbol{\lambda}, \mathbf{y}\right) \pi\left(\check{\boldsymbol{\beta}} \mid \boldsymbol{\tau}, \boldsymbol{\lambda}, \sigma^2, \mathbf{y}\right) \pi\left(\boldsymbol{\tau} \mid \boldsymbol{\beta}, \sigma^2, \boldsymbol{\lambda}, \mathbf{y}\right) \pi\left(\boldsymbol{\lambda} \mid \boldsymbol{\beta}, \sigma^2, \mathbf{y}\right) \\
&\pi\left(\boldsymbol{\beta} \mid \check{\boldsymbol{\tau}}, \check{\boldsymbol{\lambda}}, \sigma^2, \mathbf{y}\right) \pi\left(\sigma^2 \mid \check{\boldsymbol{\beta}}, \check{\boldsymbol{\tau}}, \check{\boldsymbol{\lambda}}, \mathbf{y}\right) \pi\left(\check{\boldsymbol{\tau}} \mid \check{\boldsymbol{\beta}}, \check{\sigma}^2, \check{\boldsymbol{\lambda}}, \mathbf{y}\right) \pi\left(\check{\boldsymbol{\lambda}} \mid \check{\boldsymbol{\beta}}, \check{\sigma}^2, \mathbf{y}\right)\; d\sigma^2\, d\check{\sigma}^2\, d\boldsymbol{\beta} \\
\geq\quad & f_1\left(\boldsymbol{\tau}, \boldsymbol{\lambda}, \check{\boldsymbol{\tau}}, \boldsymbol{\lambda}'\right) \left\{ \frac{\tilde{\Delta}^{\frac{n+p+2\alpha}{2}}\ \tilde{\Delta}_*^{\frac{n+p+2\alpha}{2}}}{\left[\tilde{\Delta} + \check{\boldsymbol{\beta}}^T D_{\check{\boldsymbol{\tau}}}^{-1} \check{\boldsymbol{\beta}}\right]^{\frac{n+p+2\alpha}{2}}} \right\} \\
&\left\{ \frac{2^p \sqrt{\pi}^p |X^T X + D_{\check{\boldsymbol{\tau}}}^{-1} + D_{\boldsymbol{\tau}}^{-1}|^{-\frac{1}{2}}}{\left[\tilde{\Delta}_1 + \tilde{\Delta}_* + f_2\left(\boldsymbol{\tau}, \check{\boldsymbol{\tau}}\right)\right]^{\frac{n+p+2\alpha}{2} + \frac{p}{2}}} \right\} \left[\left\{2\pi\Gamma\left(\zeta+1\right)\right\}^p\right]^{-2}
\end{aligned} \tag{B.16}$$

Note that

$$\frac{\check{\boldsymbol{\beta}}^T D_{\check{\boldsymbol{\tau}}}^{-1} \check{\boldsymbol{\beta}}}{\tilde{\Delta}} \quad = \quad \frac{\check{\boldsymbol{\beta}}^T D_{\check{\boldsymbol{\tau}}}^{-1} \check{\boldsymbol{\beta}}}{(\mathbf{y} - X\check{\boldsymbol{\beta}})^T(\mathbf{y} - X\check{\boldsymbol{\beta}}) + \check{\boldsymbol{\beta}}^T D_{\boldsymbol{\tau}}^{-1} \check{\boldsymbol{\beta}} + 2\xi} \leq \frac{\check{\boldsymbol{\beta}}^T D_{\check{\boldsymbol{\tau}}}^{-1} \check{\boldsymbol{\beta}}}{\check{\boldsymbol{\beta}}^T D_{\boldsymbol{\tau}}^{-1} \check{\boldsymbol{\beta}}} \leq \max_{1 \leq j \leq p} \left(\frac{\tau_j}{\check{\tau}_j}\right), \tag{B.17}$$

and

$$\begin{aligned}
\frac{\tilde{\Delta}_1}{\tilde{\Delta}_*} \quad &= \quad \frac{\left(\check{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}\right)^T (X^T X + D_{\boldsymbol{\tau}}^{-1}) \left(\check{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}\right)}{(\mathbf{y} - X\check{\boldsymbol{\beta}})^T(\mathbf{y} - X\check{\boldsymbol{\beta}}) + \check{\boldsymbol{\beta}}^T D_{\check{\boldsymbol{\tau}}}^{-1} \check{\boldsymbol{\beta}} + 2\xi} \\
&\leq \quad \frac{\left(\check{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}\right)^T (X^T X + D_{\boldsymbol{\tau}}^{-1}) \left(\check{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}\right) + \widehat{\boldsymbol{\beta}}^T D_{\boldsymbol{\tau}}^{-1} \widehat{\boldsymbol{\beta}} + (\mathbf{y} - X\widehat{\boldsymbol{\beta}})^T(\mathbf{y} - X\widehat{\boldsymbol{\beta}})}{(\mathbf{y} - X\check{\boldsymbol{\beta}})^T(\mathbf{y} - X\check{\boldsymbol{\beta}}) + \check{\boldsymbol{\beta}}^T D_{\check{\boldsymbol{\tau}}}^{-1} \check{\boldsymbol{\beta}} + 2\xi} \\
&= \quad \frac{(\mathbf{y} - X\check{\boldsymbol{\beta}})^T(\mathbf{y} - X\check{\boldsymbol{\beta}}) + \check{\boldsymbol{\beta}}^T D_{\boldsymbol{\tau}}^{-1} \check{\boldsymbol{\beta}}}{(\mathbf{y} - X\check{\boldsymbol{\beta}})^T(\mathbf{y} - X\check{\boldsymbol{\beta}}) + \check{\boldsymbol{\beta}}^T D_{\check{\boldsymbol{\tau}}}^{-1} \check{\boldsymbol{\beta}} + 2\xi} \\
&\leq \quad 1 + \frac{\check{\boldsymbol{\beta}}^T D_{\boldsymbol{\tau}}^{-1} \check{\boldsymbol{\beta}}}{\check{\boldsymbol{\beta}}^T D_{\check{\boldsymbol{\tau}}}^{-1} \check{\boldsymbol{\beta}}} \\
&\leq \quad 1 + \max_{1 \leq j \leq p} \left(\frac{\check{\tau}_j}{\tau_j}\right).
\end{aligned} \tag{B.18}$$

From (B.16), (B.17), (B.18) and the fact

$$\tilde{\Delta}_* = (\mathbf{y} - X\check{\boldsymbol{\beta}})^T(\mathbf{y} - X\check{\boldsymbol{\beta}}) + \check{\boldsymbol{\beta}}^T D_{\check{\boldsymbol{\tau}}}^{-1} \check{\boldsymbol{\beta}} + 2\xi \geq \widehat{\boldsymbol{\beta}}_*^T D_{\check{\boldsymbol{\tau}}}^{-1} \widehat{\boldsymbol{\beta}}_* + (\mathbf{y} - X\widehat{\boldsymbol{\beta}}_*)^T(\mathbf{y} - X\widehat{\boldsymbol{\beta}}_*) + 2\xi$$

(since $\widehat{\boldsymbol{\beta}}_*$ minimizes $\tilde{\Delta}_*$ as a function of $\check{\boldsymbol{\beta}}$), it follows that

$$\begin{aligned}
&\int_{\mathbb{R}^p}\int_{\mathbb{R}_+}\int_{\mathbb{R}_+} \pi\left(\check{\sigma}^2 \mid \check{\boldsymbol{\beta}}, \boldsymbol{\tau}, \boldsymbol{\lambda}, \mathbf{y}\right) \pi\left(\check{\boldsymbol{\beta}} \mid \boldsymbol{\tau}, \boldsymbol{\lambda}, \sigma^2, \mathbf{y}\right) \pi\left(\boldsymbol{\tau} \mid \boldsymbol{\beta}, \sigma^2, \boldsymbol{\lambda}, \mathbf{y}\right) \pi\left(\boldsymbol{\lambda} \mid \boldsymbol{\beta}, \sigma^2, \mathbf{y}\right) \\
&\pi\left(\boldsymbol{\beta} \mid \check{\boldsymbol{\tau}}, \check{\boldsymbol{\lambda}}, \sigma^2, \mathbf{y}\right) \pi\left(\sigma^2 \mid \check{\boldsymbol{\beta}}, \check{\boldsymbol{\tau}}, \check{\boldsymbol{\lambda}}, \mathbf{y}\right) \pi\left(\check{\boldsymbol{\tau}} \mid \check{\boldsymbol{\beta}}, \check{\sigma}^2, \check{\boldsymbol{\lambda}}, \mathbf{y}\right) \pi\left(\check{\boldsymbol{\lambda}} \mid \check{\boldsymbol{\beta}}, \check{\sigma}^2, \mathbf{y}\right)\ d\sigma^2\ d\check{\sigma}^2\ d\boldsymbol{\beta} \\
\geq\ & f_3\left(\boldsymbol{\tau}, \boldsymbol{\lambda}, \check{\boldsymbol{\tau}}, \check{\boldsymbol{\lambda}}\right) \left\{\frac{1}{\left[\tilde{\Delta}_1 + \tilde{\Delta}_* + f_2\left(\boldsymbol{\tau}, \check{\boldsymbol{\tau}}\right)\right]^{\frac{p}{2}}}\right\}, \qquad \text{(B.19)}
\end{aligned}$$

where

$$\begin{aligned}
& f_3\left(\boldsymbol{\tau}, \boldsymbol{\lambda}, \check{\boldsymbol{\tau}}, \check{\boldsymbol{\lambda}}\right) \\
=\ & \left[\left\{2\pi\Gamma\left(\zeta+1\right)\right\}^p\right]^{-2}\ f_1\left(\boldsymbol{\tau}, \boldsymbol{\lambda}, \check{\boldsymbol{\tau}}, \check{\boldsymbol{\lambda}}\right) \left\{\frac{2^p\sqrt{\pi}^p|X^TX + D_{\check{\boldsymbol{\tau}}}^{-1} + D_{\boldsymbol{\tau}}^{-1}|^{-\frac{1}{2}}}{\left[1 + \max_{1\leq j\leq p}\left(\frac{\tau_j}{\check{\tau}_j}\right)\right]^{\frac{n+p+2\alpha}{2}}}\right\} \\
& \left\{\frac{1}{\left[2 + \max_{1\leq j\leq p}\left(\frac{\check{\tau}_j}{\tau_j}\right) + \frac{f_2(\boldsymbol{\tau}, \check{\boldsymbol{\tau}})}{\widehat{\boldsymbol{\beta}}_*^T D_{\check{\boldsymbol{\tau}}}^{-1}\widehat{\boldsymbol{\beta}}_* + (\mathbf{y} - X\widehat{\boldsymbol{\beta}}_*)^T(\mathbf{y} - X\widehat{\boldsymbol{\beta}}_*) + 2\xi}\right]^{\frac{n+p+2\alpha}{2}}}\right\}.
\end{aligned}$$

Let $\hat{\beta}_{***} = (2X^TX + D_{\tau}^{-1} + D_{\tau'}^{-1})^{-1}2X^Ty$. Note that

$$\begin{aligned}
& \tilde{\Delta}_1 + \tilde{\Delta}_* + f_2\left(\boldsymbol{\tau}, \check{\boldsymbol{\tau}}\right) \\
=\ & \left(\check{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}\right)^T (X^TX + D_{\boldsymbol{\tau}}^{-1}) \left(\check{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}\right) + (\mathbf{y} - X\check{\boldsymbol{\beta}})^T(\mathbf{y} - X\check{\boldsymbol{\beta}}) + \check{\boldsymbol{\beta}}^T D_{\check{\boldsymbol{\tau}}}^{-1}\check{\boldsymbol{\beta}} + 2\xi + f_2\left(\boldsymbol{\tau}, \check{\boldsymbol{\tau}}\right) \\
=\ & \left(\check{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}\right)^T (X^TX + D_{\boldsymbol{\tau}}^{-1}) \left(\check{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}\right) + \left(\check{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_*\right)^T (X^TX + D_{\check{\boldsymbol{\tau}}}^{-1}) \left(\check{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_*\right) + \\
& (\mathbf{y} - X\widehat{\boldsymbol{\beta}}_*)^T(\mathbf{y} - X\widehat{\boldsymbol{\beta}}_*) + \widehat{\boldsymbol{\beta}}_*^T D_{\boldsymbol{\tau}}^{-1}\widehat{\boldsymbol{\beta}}_* + 2\xi + f_2\left(\boldsymbol{\tau}, \check{\boldsymbol{\tau}}\right) \\
=\ & \left(\check{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{***}\right)^T (2X^TX + D_{\boldsymbol{\tau}}^{-1} + D_{\check{\boldsymbol{\tau}}}^{-1}) \left(\check{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{***}\right) \\
& + \left(\widehat{\boldsymbol{\beta}}_{***} - \widehat{\boldsymbol{\beta}}\right)^T (X^TX + D_{\boldsymbol{\tau}}^{-1}) \left(\widehat{\boldsymbol{\beta}}_{***} - \widehat{\boldsymbol{\beta}}\right) + \left(\widehat{\boldsymbol{\beta}}_{***} - \widehat{\boldsymbol{\beta}}_*\right)^T (X^TX + D_{\check{\boldsymbol{\tau}}}^{-1}) \left(\widehat{\boldsymbol{\beta}}_{***} - \widehat{\boldsymbol{\beta}}_*\right) \\
& +(\mathbf{y} - X\widehat{\boldsymbol{\beta}}_*)^T(\mathbf{y} - X\widehat{\boldsymbol{\beta}}_*) + \widehat{\boldsymbol{\beta}}_*^T D_{\boldsymbol{\tau}}^{-1}\widehat{\boldsymbol{\beta}}_* + 2\xi + f_2\left(\boldsymbol{\tau}, \check{\boldsymbol{\tau}}\right), \\
=\ & \left(\check{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{***}\right)^T (2X^TX + D_{\boldsymbol{\tau}}^{-1} + D_{\check{\boldsymbol{\tau}}}^{-1}) \left(\check{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{***}\right) + f_4\left(\boldsymbol{\tau}, \check{\boldsymbol{\tau}}\right), \qquad \text{(B.20)}
\end{aligned}$$

where

$$\begin{aligned}
& f_4\left(\boldsymbol{\tau}, \check{\boldsymbol{\tau}}\right) = \left(\widehat{\boldsymbol{\beta}}_{***} - \widehat{\boldsymbol{\beta}}\right)^T (X^TX + D_{\boldsymbol{\tau}}^{-1}) \left(\widehat{\boldsymbol{\beta}}_{***} - \widehat{\boldsymbol{\beta}}\right) + \left(\widehat{\boldsymbol{\beta}}_{***} - \widehat{\boldsymbol{\beta}}_*\right)^T (X^TX + D_{\check{\boldsymbol{\tau}}}^{-1}) \left(\widehat{\boldsymbol{\beta}}_{***} - \widehat{\boldsymbol{\beta}}_*\right) \\
& +(\mathbf{y} - X\widehat{\boldsymbol{\beta}}_*)^T(\mathbf{y} - X\widehat{\boldsymbol{\beta}}_*) + \widehat{\boldsymbol{\beta}}_*^T D_{\widehat{\boldsymbol{\tau}}}^{-1}\widehat{\boldsymbol{\beta}}_* + 2\xi + f_2\left(\boldsymbol{\tau}, \check{\boldsymbol{\tau}}\right).
\end{aligned}$$

From (B.19) and (B.20), we get that

$$\begin{aligned}
& \int_{\mathbb{R}^p}\int_{\mathbb{R}^p}\int_{\mathbb{R}_+}\int_{\mathbb{R}_+} \pi\left(\check{\sigma}^2 \mid \check{\boldsymbol{\beta}}, \boldsymbol{\tau}, \boldsymbol{\lambda}, \mathbf{y}\right) \pi\left(\check{\boldsymbol{\beta}} \mid \boldsymbol{\tau}, \boldsymbol{\lambda}, \sigma^2, \mathbf{y}\right) \pi\left(\boldsymbol{\tau} \mid \boldsymbol{\beta}, \sigma^2, \boldsymbol{\lambda}, \mathbf{y}\right) \pi\left(\boldsymbol{\lambda} \mid \boldsymbol{\beta}, \sigma^2, \mathbf{y}\right) \\
& \pi\left(\boldsymbol{\beta} \mid \check{\boldsymbol{\tau}}, \check{\boldsymbol{\lambda}}, \sigma^2, \mathbf{y}\right) \pi\left(\sigma^2 \mid \check{\boldsymbol{\beta}}, \check{\boldsymbol{\tau}}, \check{\boldsymbol{\lambda}}, \mathbf{y}\right) \pi\left(\check{\boldsymbol{\tau}} \mid \check{\boldsymbol{\beta}}, \check{\sigma}^2, \check{\boldsymbol{\lambda}}, \mathbf{y}\right) \pi\left(\check{\boldsymbol{\lambda}} \mid \check{\boldsymbol{\beta}}, \check{\sigma}^2, \mathbf{y}\right)\ d\sigma^2\, d\check{\sigma}^2\, d\boldsymbol{\beta}\, d\check{\boldsymbol{\beta}} \\
\geq\ & f_3\left(\boldsymbol{\tau}, \boldsymbol{\lambda}, \check{\boldsymbol{\tau}}, \check{\boldsymbol{\lambda}}\right) \int_{\mathbb{R}^p} \left\{ \frac{1}{\left[f_4\left(\boldsymbol{\tau}, \check{\boldsymbol{\tau}}\right) + \left(\check{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{***}\right)^T (2X^TX + D_{\boldsymbol{\tau}}^{-1} + D_{\check{\boldsymbol{\tau}}}^{-1}) \left(\check{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{***}\right)\right]^{\frac{p}{2}}}\, d\check{\boldsymbol{\beta}} \right\} \\
\geq\ & \frac{f_3\left(\boldsymbol{\tau}, \boldsymbol{\lambda}, \check{\boldsymbol{\tau}}, \check{\boldsymbol{\lambda}}\right)}{\left[f_4\left(\boldsymbol{\tau}, \check{\boldsymbol{\tau}}\right)\right]^{\frac{p}{2}}} \int_{\mathbb{R}^p} \left\{ \frac{1}{\left[1 + \left(\check{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{***}\right)^T \frac{(2X^TX + D_{\boldsymbol{\tau}}^{-1} + D_{\check{\boldsymbol{\tau}}}^{-1})}{f_4(\boldsymbol{\tau}, \check{\boldsymbol{\tau}})} \left(\check{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{***}\right)\right]^{\frac{p}{2}}}\, d\check{\boldsymbol{\beta}} \right\} \\
=\ & \infty \qquad\qquad \text{(B.21)}
\end{aligned}$$

for every $(\boldsymbol{\tau}, \boldsymbol{\lambda}, \check{\boldsymbol{\tau}}, \check{\boldsymbol{\lambda}}) \in \mathbb{R}_+^p \times \mathbb{R}_+^p \times \mathbb{R}_+^p \times \mathbb{R}_+^p$. The fact that the last integral is infinite follows by noting that the multivariate t-distribution with 1 degree of freedom does not have a finite mean (Kotz and Nadarajah, 2004). Hence, it follows from (B.7), (B.10) and (B.21) that the Markov operator $\tilde{K}$ is not Hilbert-Schmidt. □

C. Mathematical identities

Proposition C1 *Suppose the random variable U has a $t-$distribution with scale parameter κ, location parameter ϑ and degrees of freedom ν. Then for $\omega < \nu$,*

$$E\left(|U|^\omega\right) \leq (2|\vartheta|)^\omega + \kappa^\omega \left[(4\nu)^{\frac{\omega}{2}} \frac{\Gamma(\frac{\omega+1}{2})\Gamma(\frac{\nu-\omega}{2})}{\sqrt{2}\Gamma(\frac{\nu}{2})}\right].$$

Proof If the random variable U has a $t-$distribution with scale parameter κ, location parameter ϑ and degrees of freedom ν then $U = \vartheta + \kappa\ T$ where T is a standard t-distribution with ν degrees of freedom. Hence

$$\begin{aligned}
E\left(|U|^\omega\right) = E\left(|\vartheta + \kappa\ T|^\omega\right) &\leq E\left((2|\vartheta|)^\omega 1_{\vartheta \geq \kappa T}\right) + E\left((2\kappa|T|)^\omega 1_{\vartheta \leq \kappa T}\right) \\
&\leq (2|\vartheta|)^\omega + (2\ \kappa)^\omega E\left(|T|^\omega\right) \\
&= (2|\vartheta|)^\omega + (2\ \kappa)^\omega E\left(\frac{|Z|^\omega}{(W/\nu)^{\frac{\omega}{2}}}\right),
\end{aligned}$$

where Z and W are independent with $Z \sim N(0,1)$ and W is χ^2 with ν degrees of freedom. Hence we get that

$$\begin{aligned} E\left(|U|^{\omega}\right) &\leq (2|\vartheta|)^{\omega}+(2\sqrt{\nu}\ \kappa)^{\omega} E\left(|Z|^{\omega}\right) E\left(\frac{1}{W^{\frac{\omega}{2}}}\right) \\ &= (2|\vartheta|)^{\omega}+(2\sqrt{\nu}\ \kappa)^{\omega}\left[\frac{2^{\frac{\omega}{2}}\ \Gamma(\frac{\omega+1}{2})}{\sqrt{\pi}}\right]\left[\frac{\Gamma(\frac{\nu-\omega}{2})}{2^{\frac{\omega}{2}}\Gamma(\frac{\omega}{2})}\right] \\ &= (2|\vartheta|)^{\omega}+\kappa^{\omega}\left[(4\nu)^{\frac{\omega}{2}}\frac{\Gamma(\frac{\omega+1}{2})\Gamma(\frac{\nu-\omega}{2})}{\sqrt{\pi}\Gamma(\frac{\omega}{2})}\right]. \qquad \text{(C.22)} \end{aligned}$$

□

Proposition C2 *Let $\zeta > 0$, $\delta_j \in \{0,1\}$, $\hat{\beta}_j = e_j^T\widehat{\boldsymbol{\beta}}$, i.e. the j^{th} component of $\widehat{\boldsymbol{\beta}}$, $\xi_j = \sqrt{\frac{\mathbf{y}^T(I_n-P_X)\mathbf{y}+2\xi}{(\varpi_{min}+\frac{1}{\tau_j})\nu_j}}$ and $\nu_j = \frac{n+2\alpha}{p}+(1+\zeta)\delta_j$ then there is a finite constant C_{0j} such that,*

$$\int_{\mathbb{R}} \frac{|\beta_j|^{(1+\zeta)\delta_j}}{\left[1+\frac{(\beta_j-\hat{\beta}_j)^2}{\nu_j\xi_j^2}\right]^{\frac{1+\nu_j}{2}}}\, d\beta_j \leq \frac{C_{0j}}{\sqrt{\varpi_{min}+\frac{1}{\tau_j}}}\ ,$$

where ϖ_{min} is the smallest eigenvalue of X^TX.

Proof Note that

$$\int_{\mathbb{R}} \frac{|\beta_j|^{(1+\zeta)\delta_j}}{\left[1+\frac{(\beta_j-\hat{\beta}_j)^2}{\nu_j\xi_j^2}\right]^{\frac{1+\nu_j}{2}}}\, d\beta_j = \xi_j\frac{\Gamma(\frac{\nu_j}{2})\sqrt{\pi\ \nu_j}}{\Gamma(\frac{\nu_j+1}{2})}E\left(|U_j|^{(1+\zeta)\delta_j}\right),$$

where U_j follows a t-distribution with scale ξ_j, location $\hat{\beta}_j$ and degrees of freedom ν_j. Using Proposition C1 and the fact $\delta_j \in \{0,1\}$, we get that

$$\begin{aligned} \int_{\mathbb{R}} \frac{|\beta_j|^{(1+\zeta)\delta_j}}{\left[1+\frac{(\beta_j-\hat{\beta}_j)^2}{\nu_j\xi_j^2}\right]^{\frac{1+\nu_j}{2}}}\, d\beta_j &= \xi_j\frac{\Gamma(\frac{\nu_j}{2})\sqrt{\pi\ \nu_j}}{\Gamma(\frac{\nu_j+1}{2})}\left[E\left(|U_j|^{(1+\zeta)}\right)\right]^{\delta_j} \\ &\leq \xi_j\frac{\Gamma(\frac{\nu_j}{2})\sqrt{\pi\ \nu_j}}{\Gamma(\frac{\nu_j+1}{2})}\left[\left(2|\hat{\beta}_j|\right)^{1+\zeta}+{\xi_j}^{1+\zeta}(4\nu_j)^{\frac{1+\zeta}{2}}\frac{\Gamma(\frac{\zeta+2}{2})\Gamma(\frac{n+2\alpha}{2p})}{\sqrt{\pi}\Gamma(\frac{(1+\zeta)}{2})}\right]^{\delta_j}. \end{aligned}$$

Since we are assuming X^TX is a positive definite matrix, $\varpi_{min} > 0$. Note that $|\hat{\beta}_j| \leq \sqrt{\hat{\boldsymbol{\beta}}^T\hat{\boldsymbol{\beta}}} = \sqrt{\mathbf{y}^TX(X^TX+D_\tau^{-1})^{-2}X^T\mathbf{y}} \leq \sqrt{\mathbf{y}^TX(X^TX)^{-2}X^T\mathbf{y}}$, since the positive definite matrix $(X^TX)(X^TX+D_\tau^{-1})^{-2}(X^TX) = (I_p+(X^TX)^{-1/2}D_\tau^{-1}(X^TX)^{-1/2})^{-1}$ has all eigenvalues bounded by 1. Recall that $\xi_j = \sqrt{\frac{\mathbf{y}^T(I_n-P_X)\mathbf{y}+2\xi}{(\varpi_{min}+\frac{1}{\tau_j})\nu_j}} \leq \sqrt{\frac{\mathbf{y}^T(I_n-P_X)\mathbf{y}+2\xi}{\varpi_{min}\ \nu_j}}$. Hence, we get that

$$\int_{\mathbb{R}} \frac{|\beta_j|^{(1+\zeta)\delta_j}}{\left[1+\frac{(\beta_j-\hat{\beta}_j)^2}{\nu_j\xi_j^2}\right]^{\frac{1+\nu_j}{2}}}\, d\beta_j \leq \frac{C_{0j}}{\sqrt{\varpi_{min}+\frac{1}{\tau_j}}},$$

where

$$
\begin{aligned}
C_{0j} \;\; = \;\; & \sqrt{\frac{\mathbf{y}^T(I_n - P_X)\mathbf{y} + 2\xi}{\nu_j}} \;\; \frac{\Gamma(\frac{\nu_j}{2})\sqrt{\pi\,\nu_j}}{\Gamma(\frac{\nu_j+1}{2})} \\
& \left[\left(2\sqrt{\mathbf{y}^T X(X^T X)^{-2} X^T \mathbf{y}}\right)^{1+\zeta} + \left(\sqrt{\frac{\mathbf{y}^T(I_n - P_X)\mathbf{y} + 2\xi}{\varpi_{min}\;\nu_j}}\right)^{1+\zeta} (4\nu_j)^{\frac{1+\zeta}{2}} \frac{\Gamma(\frac{\zeta+2}{2})\Gamma(\frac{n+2\alpha}{2p})}{\sqrt{\pi}\Gamma(\frac{(1+\zeta)}{2})}\right]^{\delta_j}.
\end{aligned}
$$

□

D. Rejection sampling approach to sample from f_G

Note that the extra step density f_G (with respect to the Lebesgue measure on $\mathbb{R}_+$ is given by

$$
f_G(g) \;\; = \;\; K \frac{g^{p/2-1}\; e^{-g\left(\sum_{j=1}^{p} \frac{\lambda_j^2 \tau_j}{2}\right)}}{\left\{\mathbf{y}^T\mathbf{y} - \mathbf{y}^T X^T \left(X^T X + \frac{1}{g} D_{\boldsymbol{\tau}}^{-1}\right)^{-1} X^T\mathbf{y} + 2\xi\right\}^{\frac{n}{2}+\alpha} |X^T X + \frac{1}{g} D_{\boldsymbol{\tau}}^{-1}|^{\frac{1}{2}}},
$$

where K is an appropriate normalizing constant. In the case when $X^T X$ is a positive definite matrix, we get

$$
f_G(g) \;\; \leq \;\; \frac{K^*}{\left\{\mathbf{y}^T\mathbf{y} - \mathbf{y}^T X^T \left(X^T X\right)^{-1} X^T\mathbf{y} + 2\xi\right\}^{\frac{n}{2}+\alpha} |X^T X|^{\frac{1}{2}}} f_*(g),
$$

where K_* is appropriate constant and f_* is a Gamma density with shape parameter $\frac{p}{2}$ and rate parameter $\frac{1}{2}\sum_{j=1}^{p}\lambda_j^2\tau_j$. Hence a rejection sampler algorithm based on the Gamma distribution can easily be implemented.

References

KOTZ, S. and NADARAJAH, S. (2004). *Multivariate t Distributions and Their Applications.* Cambridge University Press.