Electronic Journal of Statistics
Vol. 13 (2019) 166–207
ISSN: 1935-7524
https://doi.org/10.1214/18-EJS1491

Trace class Markov chains for the Normal-Gamma Bayesian shrinkage model

Liyuan Zhang[*], **Kshitij Khare**[*,†,‡] **and Zeren Xing**[*]

Department of Statistics, University of Florida
e-mail: zhangliyuan@ufl.edu; kdkhare@stat.ufl.edu; zeren.xing@ufl.edu

Abstract: High-dimensional data, where the number of variables exceeds or is comparable to the sample size, is now pervasive in many scientific applications. In recent years, Bayesian shrinkage models have been developed as effective and computationally feasible tools to analyze such data, especially in the context of linear regression. In this paper, we focus on the Normal-Gamma shrinkage model developed by Griffin and Brown [7]. This model subsumes the popular Bayesian lasso model, and a three-block Gibbs sampling algorithm to sample from the resulting intractable posterior distribution has been developed in [7]. We consider an alternative two-block Gibbs sampling algorithm, and rigorously demonstrate its advantage over the three-block sampler by comparing specific spectral properties. In particular, we show that the Markov operator corresponding to the two-block sampler is trace class (and hence Hilbert-Schmidt), whereas the operator corresponding to the three-block sampler is not even Hilbert-Schmidt. The trace class property for the two-block sampler implies geometric convergence for the associated Markov chain, which justifies the use of Markov chain CLT's to obtain practical error bounds for MCMC based estimates. Additionally, it facilitates theoretical comparisons of the two-block sampler with sandwich algorithms which aim to improve performance by inserting inexpensive extra steps in between the two conditional draws of the two-block sampler.

MSC 2010 subject classifications: Primary 60J05, 60J20; secondary 33C10.
Keywords and phrases: Data Augmentation, Markov chain Monte Carlo, normal-Gamma model, Bayesian shrinakge, trace class operators.

Received March 2017.

Contents

[*]Kshitij Khare is Associate Professor, Liyuan Zhang and Zeren Xing are Graduate Students, Department of Statistics, University of Florida.

[†]Kshitij Khare was partially supported by NSF grant DMS-1511945.

[‡]When this article was first made public, on January 16, 2019, Kshitij Khare funding information was left out of the article. The article was corrected on July 30, 2019.

1. Introduction

In recent years, the explosion of data, due to advances in science and information technology, has left almost no field untouched. The availability of high-throughput data from genomic, finance, environmental, marketing (among other) applications has created an urgent need for methodology and tools for analyzing high-dimensional data. In particular, consider the linear model $\mathbf{Y} = X\boldsymbol{\beta} + \sigma\boldsymbol{\epsilon}$, where $\mathbf{Y}$ is an $n \times 1$ real valued response variable, X is a known $n \times p$ matrix, $\boldsymbol{\beta}$ is an unknown $p \times 1$ vector of regression coefficients, σ is an unknown scale parameter and the entries of $\boldsymbol{\epsilon}$ are independent standard normals. In the high-dimensional datasets mentioned above, often $n < p$, and classical least squares methods fail. The lasso [30] was developed to provide sparse estimates of the regression coefficient vector $\boldsymbol{\beta}$ in these sample-starved settings (several adaptations/alternatives have been proposed since then). It was observed in [30] that the lasso estimate is the posterior mode obtained when one puts i.i.d Laplace priors on the elements of $\boldsymbol{\beta}$ (conditional on σ). This observation 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". Bayesian shrinkage methods have gained popularity and have been extensively used in a variety of applications including ecology, finance, image processing and neuroscience (see, for example, [34, 4, 5, 11, 33, 20, 8, 25, 26]).

In this paper, we focus on the well-known Normal-Gamma shrinkage model introduced in Griffin and Brown [7]. The model is specified as follows:

$$\begin{aligned}
&\mathbf{Y} \mid \boldsymbol{\beta}, \boldsymbol{\tau}, \sigma^2 \sim \mathcal{N}_n\left(X\beta, \sigma^2 I_n\right) \\
&\boldsymbol{\beta} \mid \sigma^2, \boldsymbol{\tau} \sim \mathcal{N}_p\left(\mathbf{0}_p, \sigma^2 D_{\boldsymbol{\tau}}\right) \\
&\sigma^2 \sim \text{Inverse} - \text{Gamma}\left(\alpha, \xi\right) \quad \text{(allow for impropriety via } \alpha = 0 \text{ or } \xi = 0\text{)} \\
&\tau_j \overset{i.i.d}{\sim} \text{Gamma}(a, b) \quad \text{for } j = 1, 2, ..., p, \qquad (1)
\end{aligned}$$

where $\mathcal{N}_p$ denotes the p-variate normal density, and $D_{\boldsymbol{\tau}}$ is a diagonal matrix with diagonal entries given by $\{\tau_j\}_{j=1}^p$. Also, Inverse-Gamma(α, ξ) and Gamma(a, b) denote the Inverse-Gamma and Gamma densities with shape parameters α and a, and rate parameters ξ and b respectively. The marginal density of $\boldsymbol{\beta}$ given σ^2 in the above model is given by

$$\pi(\boldsymbol{\beta} \mid \sigma^2) = \prod_{j=1}^{p} \frac{b^a}{\Gamma(a)\sqrt{2\pi}\sigma} \left(\frac{\beta_j^2}{2b\sigma^2}\right)^{a/2} K_a\left(\frac{|\beta_j|\sqrt{2b}}{\sigma}\right),$$

where K_a is the modified Bessel function of the second kind. The popular Bayesian lasso model of Park and Casella [23] is a special case of the Normal-Gamma model above with $a = 1$, where the marginal density of $\boldsymbol{\beta}$ simplifies to

$$\pi(\boldsymbol{\beta} \mid \sigma^2) = \prod_{j=1}^{p} \frac{\sqrt{b}}{\sqrt{2}\sigma} \exp\left(-\frac{|\beta_j|\sqrt{2b}}{\sigma}\right).$$

In this case, the marginal density for each β_j (given σ^2) is the double exponential density. The Normal-Gamma family offers a wider choice for the tail behavior (as a decreases, the marginal distribution becomes more peaked at zero, but has heavier tails), and thereby a more flexible mechanism for model shrinkage.

The posterior density of $(\boldsymbol{\beta}, \sigma^2)$ for the Normal-Gamma model is intractable in the sense that closed form computation or direct sampling is not feasible. Griffin and Brown [7] note that the full conditional densities of $\boldsymbol{\beta}, \sigma^2$ and $\boldsymbol{\tau}^2$ are easy to sample from, and develop a three-block Gibbs sampling Markov chain to generate samples from the desired posterior density. This Markov chain, denoted by $\tilde{\Phi} := \{(\tilde{\boldsymbol{\beta}}_m, \tilde{\sigma}_m^2)\}_{m=0}^{\infty}$ (on the state space $\mathbb{R}^p \times \mathbb{R}_+$), is driven by the Markov transition density (Mtd)

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

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 to move from the current state, $\left(\tilde{\boldsymbol{\beta}}_m, \tilde{\sigma}_m^2\right)$, to the next state, $\left(\tilde{\boldsymbol{\beta}}_{m+1}, \tilde{\sigma}_{m+1}^2\right)$ can be described as follows:

- Draw $\boldsymbol{\tau}$ from $\pi(\cdot \mid \tilde{\boldsymbol{\beta}}_m, \tilde{\sigma}_m^2, \mathbf{Y})$.
- Draw $\tilde{\sigma}_{m+1}^2$ from $\pi(\cdot \mid \tilde{\boldsymbol{\beta}}_m, \boldsymbol{\tau}, \mathbf{Y})$.
- Draw $\tilde{\boldsymbol{\beta}}_{m+1}$ from $\pi(\cdot \mid \tilde{\sigma}_{m+1}^2, \boldsymbol{\tau}, \mathbf{Y})$.

In [21], the authors show that the distribution of the Markov chain $\tilde{\Phi}$ converges to the desired posterior distribution at a geometric rate (as the number of steps converges to ∞).

As mentioned previously, the Bayesian lasso Markov chain of [23] is a special case of the Normal-Gamma Markov chain when $a = 1$. In recent work [12, 29], it was shown that a two-block version of the Bayesian lasso chain (and a variety of other chains for Bayesian regression) can be developed. The authors in [29] then focus their theoretical investigations on the Bayesian lasso, and show that the two-block Bayesian lasso chain has a better behaved spectrum than the original three-block Bayesian lasso chain in the following sense: the Markov operator corresponding to the two-block Bayesian lasso chain is trace class (eigenvalues are countable and summable, and hence in particular square-summable), while the Markov operator corresponding to the original three-block Bayesian lasso chain is not Hilbert-Schmidt (the corresponding absolute value operator either does not have a countable spectrum, or has a countable set of eigenvalues that are not square-summable).

Based on the method outlined in [29], a two-block version of the three-block Normal-Gamma Markov chain $\tilde{\Phi}$ can be constructed as follows. The two-block Markov chain, denoted by $\Phi = \{(\boldsymbol{\beta}_m, \sigma_m^2)\}_{m=0}^{\infty}$ (on the state space $\mathbb{R}^p \times \mathbb{R}_+$), is driven by the Markov transition density (Mtd)

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

The one-step dynamics of this Markov chain to move from the current state, $\left(\boldsymbol{\beta}_m, \sigma_m^2\right)$, to the next state, $\left(\boldsymbol{\beta}_{m+1}, \sigma_{m+1}^2\right)$ can be described as follows:

- Draw $\boldsymbol{\tau}$ from $\pi(\cdot \mid \boldsymbol{\beta}_m, \sigma_m^2, \mathbf{Y})$.
- Draw σ_{m+1}^2 from $\pi(\cdot \mid \boldsymbol{\tau}, \mathbf{Y})$ and draw $\boldsymbol{\beta}_{m+1}$ from $\pi(\cdot \mid \sigma_{m+1}^2, \boldsymbol{\tau}, \mathbf{Y})$.

Note that the Markov chain Φ arises from a two-block Data Augmentation (DA) algorithm, with $(\boldsymbol{\beta}, \sigma^2)$ as the parameter block of interest and $\boldsymbol{\tau}$ as the augmented parameter block. Hence the corresponding Markov operator is a positive, self-adjoint operator (see [9]).

The goal of this paper is to investigate whether the theoretical results for the Bayesian lasso in [29] hold for the more general and complex setting of the Normal-Gamma model. In particular, we establish that the Markov operator corresponding to the two-block chain Φ is trace class when $a > \frac{1}{2}$ (Theorem 1). On the other hand, the Markov operator corresponding to the three-block chain $\tilde{\Phi}$ is not Hilbert-Schmidt for all values of a (Theorem 2). These results hold **for all** values of the sample size n and the number of independent variables p. Since the Bayesian lasso is a special case with $a = 1$, our results subsume the spectral results in [29]. Establishing that the positive self-adjoint operator Φ is trace class implies that it has a discrete spectrum, and that (countably many, non-negative) eigenvalues are summable. The trace class property implies compactness of the corresponding Markov operator, which further implies geometric ergodicity of the underlying Markov chain (see [22, Section 2], for example). Geometric ergodicity, in turn, facilitates use of Markov chain central limit theorems to provide error bounds for Markov chain based estimates of relevant posterior expectations. Geometric ergodicity of the three block chain $\tilde{\Phi}$ has already been established in [21].

Since both chains are geometrically ergodic, a natural question for a practitioner is: do the theoretical results in this paper provide any insight to help him/her choose between the two Markov chains? Note that the two main results (Φ is trace class, $\tilde{\Phi}$ is not Hilbert-Schmidt) give us information about the spectrum of these two Markov chains. The spectrum of a Markov chain plays a crucial role in quantities of practical interest such as distance from stationarity or effective sample size. In particular, if a self-adjoint Markov operator K (with stationary density π) has a countable spectrum $\{\lambda_i(K)\}_{i=0}^{\infty}$ (with $\lambda_0 = 1$), and corresponding sequence of eigenfunctions $\{\phi_i\}_{i=0}^{\infty}$, then or any $h \in L^2(\pi)$, the asymptotic variance of the Markov chain based cumulative averages for estimating $E_\pi[h]$ is given by

$$E_\pi[h^2] + 2\sum_{i=1}^{\infty} \frac{\lambda_i(K)}{1-\lambda_i(K)} (E_\pi[h_i \phi_i])^2, \tag{4}$$

and the χ^2-distance from stationarity after n steps of the Markov chain with initial state x is given by

$$\sum_{i=1}^{\infty} \lambda_i(K)^{2n} \phi_i^2(x). \tag{5}$$

The trace class property for the two block chain Φ is equivalent to $\sum_{i=1}^{\infty} \lambda_i(\Phi) < \infty$, whereas the lack of Hilbert-Schmidt property for the three-block chain $\tilde{\Phi}$ implies either that the corresponding operator is not compact, or it is compact and $\sum_{i=1}^{\infty} \lambda_i(\tilde{\Phi}^* \tilde{\Phi}) = \infty$. Although this does not exactly characterize the difference between the asymptotic variance or distance to stationarity between the two chains, in view of (4) and (5), one would in general expect the asymptotic variance for the two block chain to be lower than the three block chain (same for the distance to stationarity). Hence, the results in this paper rigorously establish one way in which blocking affects the properties of hte Normal-Gamma Markov chain, and indicates why one should expect the blocked chain to have better performance in terms of essential sample size and convergence than the original Normal-Gamma chain.

There are two more consequences of establishing the trace-class property for Φ. Recent work in [27] provides a rigorous approach to estimate the largest eigenvalue of trace class Markov chains. Hence, the trace class property (along with an additional variance condition that we establish in Section 7) enables us to estimate the largest eigenvalue of Φ, and explore its dependence on the underlying parameters such as a, n, p. See Section 7 for an illustration. The DA interpretation of Φ also enables us to use the Haar PX-DA technique from [9] and construct a "sandwich" Markov chain by adding an inexpensive extra step in between the two conditional draws involved in one step of Φ (see Section 5 for details). The trace class property for Φ, along with results in [14], implies that the sandwich chain is also trace class, and that each ordered eigenvalue of the sandwich chain is dominated by the corresponding ordered eigenvalue of Φ (with at least one strict domination). While both these methods (eigenvalue estimation and sandwich) are applicable to the two-block Normal-Gamma chain in theory, in the context of the Normal-Gamma Markov chain, we found that while these methods work smoothly for small p settings, both encounter practical computational difficulties in the large p setting. Even so, obtaining reasonable bounds for eigenvalues of continuous state space Markov chains used in statistical practice is recognized as a really challenging problem, and getting bounds on the largest eigenvalue even in the small p case is a useful step ahead.

The rest of the paper is organized as follows. In Section 2, we provide the form of the relevant conditional densities for the Markov chains Φ and $\tilde{\Phi}$. In Section 3, we establish the trace class property for the two-block Markov chain Φ. In Section 4, we show that the three-block Markov chain is not Hilbert-Schmidt. In Section 5, we derive the Haar PX-DA sandwich chain corresponding to the two-block DA chain. Finally, in Section 6 we compare the performance of the two-block, three-block and the Haar PX-DA based chains on simulated and real datasets. In Section 7, we show that our results can be used along with a recent technique in [27] to estimate the largest eigenvalue of the two-block chain.

2. Form of relevant densities

In this section, we present expressions for various densities corresponding to the Normal-Gamma model in (1). These densities appear in the Mtd for the Markov chains Φ and $\tilde{\Phi}$.

The joint density for the parameter vector $(\boldsymbol{\beta}, \boldsymbol{\tau}, \sigma^2)$ conditioned on the data vector $\mathbf{y}$ is given by the following:

$$\pi(\boldsymbol{\beta}, \boldsymbol{\tau}, \sigma^2|\mathbf{Y}) \propto \frac{\exp\left(-\frac{\left((\mathbf{Y}-X\boldsymbol{\beta})^T(\mathbf{Y}-X\boldsymbol{\beta})\right)}{2\sigma^2}\right)}{(\sqrt{2\pi})^n\sigma^n} \frac{\exp\left(-\frac{\boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1}\boldsymbol{\beta}}{2\sigma^2}\right)}{(\sqrt{2\pi})^p\sigma^p} \times \left(\prod_{j=1}^{p} \tau_j^{a-\frac{1}{2}-1}\exp\left(-b\tau_j\right)\right)(\sigma^2)^{-a-1}\exp\left(-\frac{\xi}{\sigma^2}\right). \quad (6)$$

Based on the joint density in (6), the following conditional distributions can be derived in a straightforward fashion.

- $\boldsymbol{\beta} \mid \sigma^2, \boldsymbol{\tau}, \mathbf{Y} \sim \boldsymbol{\mathcal{N}_p}\left(\left(X^TX + D_{\boldsymbol{\tau}}^{-1}\right)^{-1} X^T\mathbf{y}, \sigma^2\left(X^TX + D_{\boldsymbol{\tau}}^{-1}\right)^{-1}\right)$,

where $\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ denotes the p-variate normal density with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ.

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

Here Inverse-Gamma(α, λ) referes to the Inverse-Gamma density with shape parameter α and rate parameter λ. In particular, the Inverse-Gamma(α, λ) is given by

$$\Gamma(\alpha)^{-1}\lambda^{\alpha}x^{-\alpha-1}\exp(-\lambda/x)$$

for $x > 0$.

- $\sigma^2 \mid \boldsymbol{\tau}, \mathbf{Y} \sim \text{Inverse} - \text{Gamma}\left(\frac{n+2\alpha}{2}, \frac{\mathbf{Y}^T\left(I - X\mathbf{A}_{\boldsymbol{\tau}}^{-1}X^T\right)\mathbf{Y}+2\xi}{2}\right)$,

where $A_{\boldsymbol{\tau}} = X^TX + D_{\boldsymbol{\tau}}$.

- Given $\boldsymbol{\beta}, \sigma^2$ and $\mathbf{y}$, the variables $\tau_1, \tau_2, ..., \tau_p$ are conditionally independent, and the conditional density of τ_j given $\boldsymbol{\beta}, \sigma^2$ and $\mathbf{y}$ is $\mathrm{GIG}(a - \frac{1}{2}, 2b, \frac{\beta_j^2}{\sigma^2})$.

Here GIG refers to the generalized inverse Gaussian density. The $GIG(\nu, r, s)$ density is given by

$$\frac{(r/s)^{\nu/2}}{2K_\nu(\sqrt{rs})} x^{v-1} \exp\left(-rx/2 - s/2x\right),$$

for $x > 0$, with K_ν being a modified Bessel function of the second kind.

2.1. Computational complexity

In this section, we evaluate the computational complexity (per iteration) of the original three-block Gibbs sampler with the proposed two-block Gibbs sampler. To this end, we examine the complexity of one iteration of each sampler starting at $(\boldsymbol{\beta}, \sigma^2)$ and ending at $(\boldsymbol{\beta}', \sigma^{2'})$. Note that starting from $(\boldsymbol{\beta}, \sigma^2)$, both algorithms sample $\boldsymbol{\tau}$ from the conditional posterior density given $\boldsymbol{\beta}, \sigma^2, \mathbf{Y}$ which corresponds to p GIG draws and required $O(p)$ computations. Then, the three block sampler obtains $\sigma^{2'}$ by sampling from an

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

density. This takes $O(np)$ iterations, with computing $X\beta$ being the computationally dominant step. On the other hand, the two block sampler obtains $\sigma^{2'}$ be sampling from an

$$\text{Inverse} - \text{Gamma}\left(\frac{n+2\alpha}{2}, \frac{\mathbf{Y}^T\left(I - X\mathbf{A}_{\boldsymbol{\tau}}^{-1}X^T\right)\mathbf{Y} + 2\xi}{2}\right)$$

density. Using the Sherman-Morrison-Woodbury identity it can be shown that

$$A_{\boldsymbol{\tau}}^{-1}X^T\mathbf{Y} = D_{\boldsymbol{\tau}}X^T(XD_{\boldsymbol{\tau}}X^T + I_n)^{-1}\mathbf{Y}.$$

and hence can be computed in $O(\min(n^2p, p^3))$ iterations. Finally, both samplers obtain $\boldsymbol{\beta}'$ from a p- variate Gaussian density, which requires $O(\min(n^2p, p^3))$ computations. Hence, while the second step of the two block sampler is slightly more expensive than the second step of the three block sampler ($O(n^2p)$ as compared to $O(np)$), the overall computational complexity per iteration for both algorithms is $O(\min(n^2p, p^3))$.

3. Properties of the two-block Gibbs sampler

In this section, we show that the operator associated with the two-block Gibbs sampler Φ, with Markov transition density k specified in (3) is trace class when $a > \frac{1}{2}$ and is not trace class when $0 < a \leq \frac{1}{2}$.

Theorem 1. *For all values of n and p, the Markov operator corresponding to the two-block Markov chain Φ is trace class (and hence Hilbert-Schmidt) when $a > \frac{1}{2}$ and is not trace class when $0 < a \leq \frac{1}{2}$.*

Proof. In the current setting, the trace class property is equivalent to the finiteness of the integral (see [13] and [22, Section 2], for example)

$$\iint\limits_{\mathbb{R}^p \times \mathbb{R}_+} k\left(\left(\boldsymbol{\beta}, \sigma^2\right), \left(\boldsymbol{\beta}, \sigma^2\right)\right) \, d\boldsymbol{\beta} \, d\sigma^2. \tag{7}$$

We will consider five separate cases: $a > 1$, $3/4 \leq a \leq 1$, $1/2 < a < 3/4$, $0 < a < 1/2$ and $a = 1/2$. In the first three cases, we will show that the integral in (7) is finite, and in the last two cases we will show that the integral in (7) is infinite. The proof is a lengthy and intricate algebraic exercise involving careful upper/lower bounds for modified Bessel functions and conditional densities, and we will try to provide a road-map/explanation whenever possible. We will start with the case $a > 1$.

Case 1: $a > 1$
By the definition of k, we have

$$\begin{aligned}
&\iint\limits_{\mathbb{R}^p \times \mathbb{R}_+} k\left(\left(\boldsymbol{\beta}, \sigma^2\right), \left(\boldsymbol{\beta}, \sigma^2\right)\right) \, d\boldsymbol{\beta} \, d\sigma^2 \\
=& \iiint\limits_{\mathbb{R}^p \times \mathbb{R}_+^p \times \mathbb{R}_+} \pi\left(\boldsymbol{\tau} \mid \boldsymbol{\beta}, \sigma^2, \mathbf{Y}\right) \pi\left(\boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{\tau}, \mathbf{Y}\right) \, d\boldsymbol{\beta} \, d\boldsymbol{\tau} \, d\sigma^2 \\
=& \iiint\limits_{\mathbb{R}^p \times \mathbb{R}_+^p \times \mathbb{R}_+} \pi\left(\boldsymbol{\tau} \mid \boldsymbol{\beta}, \sigma^2, \mathbf{Y}\right) \pi\left(\sigma^2 \mid \boldsymbol{\tau}, \mathbf{Y}\right) \pi\left(\boldsymbol{\beta} \mid \sigma^2, \boldsymbol{\tau}, \mathbf{Y}\right) \, d\boldsymbol{\beta} \, d\boldsymbol{\tau} \, d\sigma^2. \quad (8)
\end{aligned}$$

As a first step, we will gather all the terms with $\boldsymbol{\tau}$, and then focus on finding an upper bound for the inner integral with respect to $\boldsymbol{\tau}$. Using the conditional densities in Section 2, we get,

$$\begin{aligned}
&\iint\limits_{\mathbb{R}^p \times \mathbb{R}_+} k\left(\left(\boldsymbol{\beta}, \sigma^2\right), \left(\boldsymbol{\beta}, \sigma^2\right)\right) \, d\boldsymbol{\beta} \, d\sigma^2 \\
=& C_1 \iiint\limits_{\mathbb{R}^p \times \mathbb{R}_+^p \times \mathbb{R}_+} \prod_{j=1}^{p} \frac{\left(2b\sigma^2\right)^{\frac{a-\frac{1}{2}}{2}}}{2\left|\beta_j\right|^{a-\frac{1}{2}} K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)} \tau_j^{\left(a-\frac{1}{2}\right)-1}
\end{aligned}$$

$$\times \exp\left(-\frac{1}{2}\left\{2b\tau_j + \frac{\beta_j^2}{\sigma^2}\frac{1}{\tau_j}\right\}\right)\left(\sigma^2\right)^{-\frac{n+2\alpha}{2}-1}$$

$$\times \exp\left(-\frac{\mathbf{Y}^T\left(I - X\mathbf{A}_{\boldsymbol{\tau}}^{-1}X^T\right)\mathbf{Y} + 2\xi}{2\sigma^2}\right)\left(\mathbf{Y}^T\left(I - X\mathbf{A}_{\boldsymbol{\tau}}^{-1}X^T\right)\mathbf{Y} + 2\xi\right)^{\frac{n+2\alpha}{2}}$$

$$\times (\sigma^2)^{-\frac{p}{2}}|\mathbf{A}_{\boldsymbol{\tau}}|^{\frac{1}{2}}$$

$$\times \exp\left(-\frac{\left(\boldsymbol{\beta} - \mathbf{A}_{\boldsymbol{\tau}}^{-1}X^T\mathbf{Y}\right)^T\mathbf{A}_{\boldsymbol{\tau}}\left(\boldsymbol{\beta} - \mathbf{A}_{\boldsymbol{\tau}}^{-1}X^T\mathbf{Y}\right)}{2\sigma^2}\right)d\boldsymbol{\beta}\,d\boldsymbol{\tau}\,d\sigma^2$$

$$\overset{(a)}{\leq} C_2 \iiint\limits_{\mathbb{R}^p\times\mathbb{R}_+^p\times\mathbb{R}_+} \frac{\exp\left(-\frac{\xi}{\sigma^2}\right)}{(\sigma^2)^{\frac{n+2\alpha}{2}+1}}\frac{|\mathbf{A}_{\boldsymbol{\tau}}|^{\frac{1}{2}}}{(\sigma^2)^{\frac{p}{2}}}\prod_{j=1}^{p}\frac{\left(2b\sigma^2\right)^{\frac{a-\frac{1}{2}}{2}}}{2\,|\beta_j|^{a-\frac{1}{2}}K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)}\tau_j^{\left(a-\frac{1}{2}\right)-1}$$

$$\times \exp\left(-\frac{1}{2}\left\{2b\tau_j + \frac{\beta_j^2}{\sigma^2}\frac{1}{\tau_j}\right\}\right.$$

$$\left. - \frac{1}{2\sigma^2}\left(\boldsymbol{\beta}^T\mathbf{A}_{\boldsymbol{\tau}}\boldsymbol{\beta} - 2\boldsymbol{\beta}^TX^T\mathbf{Y} + \mathbf{Y}^T\mathbf{Y}\right)\right)d\boldsymbol{\beta}\,d\boldsymbol{\tau}\,d\sigma^2$$

$$\overset{(a')}{\leq} C_2 \iiint\limits_{\mathbb{R}^p\times\mathbb{R}_+^p\times\mathbb{R}_+} \frac{\exp\left(-\frac{\xi}{\sigma^2}\right)}{(\sigma^2)^{\frac{n+p+2\alpha}{2}+1}}|\mathbf{A}_{\boldsymbol{\tau}}|^{\frac{1}{2}}\prod_{j=1}^{p}\frac{\left(2b\sigma^2\right)^{\frac{a-\frac{1}{2}}{2}}}{2\,|\beta_j|^{a-\frac{1}{2}}K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)}\tau_j^{\left(a-\frac{1}{2}\right)-1}$$

$$\times \exp\left(-\frac{1}{2}\left\{2b\tau_j + \frac{2\beta_j^2}{\sigma^2}\frac{1}{\tau_j}\right\}\right)d\boldsymbol{\beta}\,d\boldsymbol{\tau}\,d\sigma^2, \tag{9}$$

where $C_1 = \frac{1}{(2\pi)^{\frac{p}{2}}2^{\frac{n+2\alpha}{2}}\Gamma\left(\frac{n+2\alpha}{2}\right)}$ and $C_2 = \left(\mathbf{Y}^T\mathbf{Y} + 2\xi\right)^{\frac{n+2\alpha}{2}}C_1$. Note that (a) follows from

$$\mathbf{Y}^T\left(I - X\mathbf{A}_{\boldsymbol{\tau}}^{-1}X^T\right)\mathbf{Y} + 2\xi \leq \mathbf{Y}^T\mathbf{Y} + 2\xi,$$

and then combining/simplifying terms in the exponent. Similarly, (a') follows from

$$\begin{aligned}
&\exp\left(-\frac{1}{2\sigma^2}\left(\boldsymbol{\beta}^T\mathbf{A}_{\boldsymbol{\tau}}\boldsymbol{\beta} - 2\boldsymbol{\beta}^TX^T\mathbf{Y} + \mathbf{Y}^T\mathbf{Y}\right)\right)\\
&= \exp\left(-\frac{\boldsymbol{\beta}D_{\boldsymbol{\tau}}^{-1}\boldsymbol{\beta}}{2\sigma^2}\right)\exp\left\{-\frac{1}{2\sigma^2}\left(\boldsymbol{\beta}^TX^TX\boldsymbol{\beta} - 2\boldsymbol{\beta}^TX^T\mathbf{Y} + \mathbf{Y}^T\mathbf{Y}\right)\right\}\\
&\leq \exp\left(-\frac{\boldsymbol{\beta}D_{\boldsymbol{\tau}}^{-1}\boldsymbol{\beta}}{2\sigma^2}\right),
\end{aligned}$$

and then combining/simplifying terms in the exponent. We now focus on the inner integral in (9) defined by

$$H\left(\boldsymbol{\beta},\sigma^2\right) \triangleq \int_{\mathbb{R}_+^p} \mid \mathbf{A}_{\boldsymbol{\tau}} \mid^{\frac{1}{2}} \prod_{j=1}^{p} \frac{\left(2b\sigma^2\right)^{\frac{a-\frac{1}{2}}{2}}}{2\left|\beta_j\right|^{a-\frac{1}{2}} K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)} \tau_j^{\left(a-\frac{1}{2}\right)-1}$$

$$\times \exp\left(-\frac{1}{2}\left\{2b\tau_j + \frac{2\beta_j^2}{\sigma^2}\frac{1}{\tau_j}\right\}\right) d\boldsymbol{\tau} \tag{10}$$

Let λ denote the largest eigenvalue of $X^T X$. Using the definition of $\mathbf{A}_{\boldsymbol{\tau}}$, it follows that

$$\begin{aligned}
& \mid \mathbf{A}_{\boldsymbol{\tau}} \mid^{\frac{1}{2}} \prod_{j=1}^{p} \frac{\left(2b\sigma^2\right)^{\frac{a-\frac{1}{2}}{2}}}{2\left|\beta_j\right|^{a-\frac{1}{2}} K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)} \tau_j^{\left(a-\frac{1}{2}\right)-1} \\
& \times \exp\left(-\frac{1}{2}\left\{2b\tau_j + \frac{2\beta_j^2}{\sigma^2}\frac{1}{\tau_j}\right\}\right) \\
\leq \; & \prod_{j=1}^{p}\left(\sqrt{\lambda} + \frac{1}{\sqrt{\tau_j}}\right) \prod_{j=1}^{p} \frac{\left(2b\sigma^2\right)^{\frac{a-\frac{1}{2}}{2}}}{2\left|\beta_j\right|^{a-\frac{1}{2}} K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)} \tau_j^{\left(a-\frac{1}{2}\right)-1} \\
& \times \exp\left(-\frac{1}{2}\left\{2b\tau_j + \frac{2\beta_j^2}{\sigma^2}\frac{1}{\tau_j}\right\}\right) \\
= \; & \left[\lambda^{\frac{p}{2}} + \left(\frac{1}{\sqrt{\tau_1}} + \ldots + \frac{1}{\sqrt{\tau_p}}\right)\lambda^{\frac{p-1}{2}}\right. \\
& \left. + \left(\frac{1}{\sqrt{\tau_1\tau_2}} + \ldots + \frac{1}{\sqrt{\tau_i\tau_j}} + \ldots\right)\lambda^{\frac{p-2}{2}} + \ldots + \frac{1}{\sqrt{\tau_1\tau_2\ldots\tau_p}}\right] \prod_{j=1}^{p} c_j
\end{aligned} \tag{11}$$

where

$$c_j = \frac{\left(2b\sigma^2\right)^{\frac{a-\frac{1}{2}}{2}}}{2\left|\beta_j\right|^{a-\frac{1}{2}} K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)} \tau_j^{\left(a-\frac{1}{2}\right)-1} \exp\left(-\frac{1}{2}\left\{2b\tau_j + \frac{2\beta_j^2}{\sigma^2}\frac{1}{\tau_j}\right\}\right).$$

We now examine a generic term of the sum in (11). Note that c_j and $\frac{c_j}{\sqrt{\tau_j}}$ are both (unnormalized) GIG densities. Hence, for any subset $\mathcal{L} = \{\ell_1, \ell_2, \cdots, \ell_m\}$

of $\{1, 2, \cdots, p\}$, using the form of the GIG density, we get

$$
\begin{aligned}
&\int_{\mathbb{R}_+^p} \frac{1}{\sqrt{\tau_{\ell_1}\tau_{\ell_2}\cdots\tau_{\ell_m}}} \prod_{j=1}^p c_j d\boldsymbol{\tau} \\
= &\left(\prod_{j\notin\mathcal{L}} \int c_j d\tau_j\right) \times \left(\prod_{j\in\mathcal{L}} \int \frac{c_j}{\sqrt{\tau_j}} d\tau_j\right) \\
= &\left(\prod_{j\notin\mathcal{L}} (\sqrt{2})^{a-\frac{1}{2}} \frac{K_{a-\frac{1}{2}}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)}{K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)}\right) \\
&\times \left(\prod_{j\in\mathcal{L}} \left(2b\sigma^2\right)^{\frac{1}{4}} \left(\sqrt{2}\right)^{a-1} |\beta_j|^{-\frac{1}{2}} \; \frac{K_{a-1}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)}{K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)}\right). \qquad (12)
\end{aligned}
$$

First, by [16, Page 266], we get that

$$
\frac{K_{a-\frac{1}{2}}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)}{K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)} < \exp\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}} - \sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right) < \exp\left(-\frac{\sqrt{b}\,|\beta_j|}{2\sigma}\right)
$$

for all $a > \frac{1}{2}$. Next, using the fact that if $x > 0$, then $\nu \to K_\nu(x)$ is an increasing function for $\nu > 0$ (again, see [16, Page 266]), and $a > 1$, we get

$$
\frac{K_{a-1}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)}{K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)} \le \frac{K_{a-\frac{1}{2}}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)}{K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)} < \exp\left(-\frac{\sqrt{b}\,|\beta_j|}{2\sigma}\right). \qquad (13)
$$

Hence, from (12), we get that

$$
\begin{aligned}
&\int_{\mathbb{R}_+^p} \frac{1}{\sqrt{\tau_{\ell_1}\tau_{\ell_2}\cdots\tau_{\ell_m}}} \prod_{j=1}^p c_j d\boldsymbol{\tau} \\
&\quad < \sqrt{2}^{pa} (b\sigma^2)^{\frac{|\mathcal{L}|}{4}} \left(\prod_{j\in\mathcal{L}} |\beta_j|^{-\frac{1}{2}}\right) \exp\left(-\sum_{j=1}^p \frac{\sqrt{b}\,|\beta_j|}{2\sigma}\right). \qquad (14)
\end{aligned}
$$

It follows from (10), (11) and (14) that

$$
H\left(\boldsymbol{\beta}, \sigma^2\right) \le \sum_{\mathcal{L}\subseteq\{1,2,\cdots,p\}} \lambda^{p-|\mathcal{L}|} \sqrt{2}^{pa} (b\sigma^2)^{\frac{|\mathcal{L}|}{4}} \left(\prod_{j\in\mathcal{L}} |\beta_j|^{-\frac{1}{2}}\right) \exp\left(-\sum_{j=1}^p \frac{\sqrt{b}\,|\beta_j|}{2\sigma}\right).
$$

By (9) and (10), the trace class property will be established if we show that for every $\mathcal{L} \subseteq \{1, 2, \cdots, p\}$, the integral

$$\iint\limits_{\mathbb{R}_+ \times \mathbb{R}^p} \frac{\exp\left(-\frac{\xi}{\sigma^2}\right)}{(\sigma^2)^{\frac{n+p+2\alpha}{2} - \frac{|\mathcal{L}|}{4} + 1}} \left(\prod_{j \in \mathcal{L}} |\beta_j|^{-\frac{1}{2}}\right) \exp\left(-\sum_{j=1}^{p} \frac{\sqrt{b}\,|\beta_j|}{2\sigma}\right) d\boldsymbol{\beta}\, d\sigma^2$$

is finite. We proceed to show this by first simplifying the inner integral with respect to $\boldsymbol{\beta}$. Using the form of the Gamma density, we get

$$\begin{aligned}
&\int_{\mathbb{R}^p} \left(\prod_{j \in \mathcal{L}} |\beta_j|^{-\frac{1}{2}}\right) \exp\left(-\sum_{j=1}^{p} \frac{\sqrt{b}\,|\beta_j|}{2\sigma}\right) d\boldsymbol{\beta} \\
= \;& \left(\prod_{j \notin \mathcal{L}} \int_{\mathbb{R}} \exp\left(-\frac{\sqrt{b}\,|\beta_j|}{2\sigma}\right) d\beta_j\right) \times \left(\prod_{j \in \mathcal{L}} \int_{\mathbb{R}} |\beta_j|^{-\frac{1}{2}} \exp\left(-\frac{\sqrt{b}\,|\beta_j|}{2\sigma}\right) d\beta_j\right) \\
= \;& \left(\frac{4\sigma}{\sqrt{b}}\right)^{p-|\mathcal{L}|} \left(2\Gamma\left(\frac{1}{2}\right)\sqrt{\frac{2\sigma}{\sqrt{b}}}\right)^{|\mathcal{L}|} \\
\leq \;& \frac{8^p}{(\sqrt{b})^{p-\frac{|\mathcal{L}|}{2}}} \sigma^{p-\frac{|\mathcal{L}|}{2}}. \qquad (15)
\end{aligned}$$

It follows by (15) that

$$\begin{aligned}
&\iint\limits_{\mathbb{R}_+ \times \mathbb{R}^p} \frac{\exp\left(-\frac{\xi}{\sigma^2}\right)}{(\sigma^2)^{\frac{n+p+2\alpha}{2} - \frac{|\mathcal{L}|}{4} + 1}} \left(\prod_{j \in \mathcal{L}} |\beta_j|^{-\frac{1}{2}}\right) \exp\left(-\sum_{j=1}^{p} \frac{\sqrt{b}\,|\beta_j|}{2\sigma}\right) d\boldsymbol{\beta}\, d\sigma^2 \\
\leq \;& \frac{8^p}{(\sqrt{b})^{p-\frac{|\mathcal{L}|}{2}}} \int_{\mathbb{R}_+} \frac{\exp\left(-\frac{\xi}{\sigma^2}\right)}{(\sigma^2)^{\frac{n+p+2\alpha}{2} - \frac{|\mathcal{L}|}{4} + 1}} \sigma^{p-\frac{|\mathcal{L}|}{2}}\, d\sigma^2 \\
\leq \;& \frac{8^p}{(\sqrt{b})^{p-\frac{|\mathcal{L}|}{2}}} \int_{\mathbb{R}_+} \frac{\exp\left(-\frac{\xi}{\sigma^2}\right)}{(\sigma^2)^{\frac{n}{2}+\alpha+1}}\, d\sigma^2 \\
= \;& \frac{8^p \Gamma\left(\frac{n}{2}+\alpha\right)}{(\sqrt{b})^{p-\frac{|\mathcal{L}|}{2}} \xi^{\frac{n}{2}+\alpha}} < \infty.
\end{aligned}$$

As discussed above, this establishes the trace class property in the case $a > 1$.

Case 2: $3/4 \leq a \leq 1$

In this case, we first note that all arguments in Case 1 go through verbatim until (12). Next, we note that

$$\frac{K_{a-1}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)}{K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)} = \frac{K_{a-\frac{1}{2}}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)}{K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)} \frac{K_{a-1}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)}{K_{a-\frac{1}{2}}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)} \tag{16}$$

If $a \geq \frac{3}{4}$, then $a - \frac{1}{2} > 0$, and by [16, Page 266], we get

$$\frac{K_{a-\frac{1}{2}}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)}{K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)} < \exp\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}} - \sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right) < \exp\left(-\frac{\sqrt{b}\,|\beta_j|}{2\sigma}\right) \tag{17}$$

Using the property that $K_\nu(x) = K_{-\nu}(x)$ (see [1], Page 375), we obtain

$$\frac{K_{a-1}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)}{K_{a-\frac{1}{2}}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)} = \frac{K_{1-a}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)}{K_{a-\frac{1}{2}}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)}$$

If $\frac{3}{4} \leq a < 1$, then $0 < 1 - a \leq a - \frac{1}{2}$. Since $\nu \to K_\nu(x)$ is increasing in $\nu > 0$ for $x > 0$ (see [16, Page 266]), it follows that $K_{1-a}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right) \leq K_{a-\frac{1}{2}}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)$ for $\frac{3}{4} \leq a < 1$. Also, by the integral formula (see [1], Page 376)

$$K_\nu(t) = \int_0^\infty \exp\left(-t\cosh z\right)\cosh(\nu z)\,dz, \nu \in \mathbb{R}.$$

Since $\cosh(\nu z) \geq \cosh(0)$ for any $\nu > 0, z > 0$ ($x \to \cosh(x)$ is increasing on $[0,\infty)$), we get

$$K_\nu(t) \geq K_0(t)$$

for $\nu > 0$. In particular, $K_0(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}) \leq K_{\frac{1}{2}}(\sqrt{4b\frac{\beta_j^2}{\sigma^2}})$. Hence for all $a \in \left[\frac{3}{4}, 1\right]$, we have

$$\frac{K_{a-1}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)}{K_{a-\frac{1}{2}}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)} \leq 1. \tag{18}$$

It follows from (16), (17) and (18) that

$$\frac{K_{a-1}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)}{K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)} < \exp\left(-\frac{\sqrt{b}\,|\beta_j|}{2\sigma}\right)$$

Now, using exactly the same arguments as in the proof of Case 1 (following (12)) the trace class property can be shown the case $\frac{3}{4} \leq a \leq 1$.

Case 3: $1/2 < a < 3/4$
Again, in this case, we first note that all arguments in Case 1 go through verbatim until (12). Also, by [16, Page 266] and $K_\nu(x) = K_{-\nu}(x)$ for $x > 0$, we get

$$\frac{K_{a-1}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)}{K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)} = \frac{K_{a-1}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)}{K_{a-\frac{1}{2}}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)} \frac{K_{a-\frac{1}{2}}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)}{K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)}$$
$$< \exp\left(-\frac{\sqrt{b}\,|\beta_j|}{2\sigma}\right) \frac{K_{1-a}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)}{K_{a-\frac{1}{2}}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)}$$

Note that if $1/2 < a < 3/4$, then $1 - a - (a - \frac{1}{2}) = \frac{3}{2} - 2a \in (0, \frac{1}{2})$. It follows by [21, Page 640]) that

$$\frac{K_{1-a}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)}{K_{a-\frac{1}{2}}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)} \leq \frac{(2a)^{\frac{3}{2}-2a}}{\left(\sqrt{4b}\frac{|\beta_j|}{\sigma}\right)^{\frac{3}{2}-2a}} + 1$$

Hence,

$$\frac{K_{a-1}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)}{K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)} < \exp\left(-\frac{\sqrt{b}\,|\beta_j|}{2\sigma}\right)\left(\frac{(2a)^{\frac{3}{2}-2a}}{\left(\sqrt{4b}\frac{|\beta_j|}{\sigma}\right)^{\frac{3}{2}-2a}} + 1\right).$$

By (12), for any subset $\mathcal{L} = \{\ell_1, \ell_2, \cdots, \ell_m\}$ of $\{1, 2, \cdots, p\}$ we get

$$\int_{\mathbb{R}_+^p} \frac{1}{\sqrt{\tau_{\ell_1}\tau_{\ell_2}\cdots\tau_{\ell_m}}} \prod_{j=1}^{p} c_j d\boldsymbol{\tau}$$
$$= \left(\prod_{j\notin\mathcal{L}} (\sqrt{2})^{a-\frac{1}{2}} \frac{K_{a-\frac{1}{2}}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)}{K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)}\right)$$
$$\times \left(\prod_{j\in\mathcal{L}} \left(2b\sigma^2\right)^{\frac{1}{4}} \left(\sqrt{2}\right)^{a-1} |\beta_j|^{-\frac{1}{2}} \; \frac{K_{a-1}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)}{K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)}\right)$$

$$
\begin{aligned}
&\leq \quad \sqrt{2}^{pa} b^{\frac{|\mathcal{L}|}{4}} (\sigma^2)^{\frac{|\mathcal{L}|}{4}} \left(\prod_{j \in \mathcal{L}} |\beta_j|^{-\frac{1}{2}} \right) \\
&\qquad \times \prod_{j \in \mathcal{L}} \left(\frac{(2a)^{\frac{3}{2}-2a}}{\left(\sqrt{4b \frac{|\beta_j|}{\sigma}} \right)^{\frac{3}{2}-2a}} + 1 \right) \exp\left(-\sum_{j=1}^{p} \frac{\sqrt{b}\,|\beta_j|}{2\sigma} \right). \qquad (19)
\end{aligned}
$$

It follows from (10) that

$$
\begin{aligned}
H\left(\boldsymbol{\beta}, \sigma^2\right) \leq &\sum_{\mathcal{L} \subseteq \{1,2,\cdots,p\}} \lambda^{p-|\mathcal{L}|} \sqrt{2}^{pa} b^{\frac{|\mathcal{L}|}{4}} (\sigma^2)^{\frac{|\mathcal{L}|}{4}} \left(\prod_{j \in \mathcal{L}} |\beta_j|^{-\frac{1}{2}} \right) \\
&\times \prod_{j \in \mathcal{L}} \left(\frac{(2-2a)^{\frac{3}{2}-2a}}{\left(\sqrt{4b \frac{|\beta_j|}{\sigma}} \right)^{\frac{3}{2}-2a}} + 1 \right) \exp\left(-\sum_{j=1}^{p} \frac{\sqrt{b}\,|\beta_j|}{2\sigma} \right)
\end{aligned}
$$

By (9), the trace class property will be established if we show that for every $\mathcal{L} \subseteq \{1, 2, \cdots, p\}$, the integral

$$
\begin{aligned}
&\iint_{\mathbb{R}_+ \times \mathbb{R}^p} \frac{\exp\left(-\frac{\xi}{\sigma^2}\right)}{(\sigma^2)^{\frac{n+p+2\alpha}{2} - \frac{|\mathcal{L}|}{4} + 1}} \left(\prod_{j \in \mathcal{L}} |\beta_j|^{-\frac{1}{2}} \right) \\
&\times \prod_{j \in \mathcal{L}} \left(\frac{(2a)^{\frac{3}{2}-2a}}{\left(\sqrt{4b \frac{|\beta_j|}{\sigma}} \right)^{\frac{3}{2}-2a}} + 1 \right) \exp\left(-\sum_{j=1}^{p} \frac{\sqrt{b}\,|\beta_j|}{2\sigma} \right) d\boldsymbol{\beta}\, d\sigma^2 \qquad (20)
\end{aligned}
$$

is finite. We proceed to show this by first integrating out $\boldsymbol{\beta}$. Using the form of the Gamma density, we get

$$
\begin{aligned}
&\int_{\mathbb{R}^p} \left(\prod_{j \in \mathcal{L}} |\beta_j|^{-\frac{1}{2}} \right) \prod_{j \in \mathcal{L}} \left(\frac{(2a)^{\frac{3}{2}-2a}}{\left(\sqrt{4b \frac{|\beta_j|}{\sigma}} \right)^{\frac{3}{2}-2a}} + 1 \right) \exp\left(-\sum_{j=1}^{p} \frac{\sqrt{b}\,|\beta_j|}{2\sigma} \right) d\boldsymbol{\beta} \\
&= \left(\prod_{j \notin \mathcal{L}} \int_{\mathbb{R}} \exp\left(-\frac{\sqrt{b}\,|\beta_j|}{2\sigma} \right) d\beta_j \right) \\
&\quad \times \left(\prod_{j \in \mathcal{L}} \int_{\mathbb{R}} |\beta_j|^{-\frac{1}{2}} \left(\frac{(2a)^{\frac{3}{2}-2a}}{\left(\sqrt{4b \frac{|\beta_j|}{\sigma}} \right)^{\frac{3}{2}-2a}} + 1 \right) \exp\left(-\frac{\sqrt{b}\,|\beta_j|}{2\sigma} \right) d\beta_j \right) \\
&= \left(\frac{4\sigma}{\sqrt{b}} \right)^{p-|\mathcal{L}|} \prod_{j \in \mathcal{L}} \left[\int_{\mathbb{R}} |\beta_j|^{-\frac{1}{2}} \exp\left(-\frac{\sqrt{b}\,|\beta_j|}{2\sigma} \right) d\beta_j \right.
\end{aligned}
$$

$$
\begin{aligned}
&+\int_{\mathbb{R}} |\beta_j|^{-\frac{1}{2}} \frac{(2a)^{\frac{3}{2}-2a}}{\left(\sqrt{4b}\frac{|\beta_j|}{\sigma}\right)^{\frac{3}{2}-2a}} \exp\left(-\frac{\sqrt{b}\,|\beta_j|}{2\sigma}\right) d\beta_j \Bigg] \\
&= \left(\frac{4\sigma}{\sqrt{b}}\right)^{p-|\mathcal{L}|} \prod_{j\in\mathcal{L}} \left(2\Gamma\left(\frac{1}{2}\right)\sqrt{\frac{2\sigma}{\sqrt{b}}} + a^{\frac{3}{2}-2a}\Gamma(2a-1)2^{2a}\left(\sqrt{b}\right)^{-\frac{1}{2}}\sigma^{\frac{1}{2}}\right) \\
&= \left(\frac{4\sigma}{\sqrt{b}}\right)^{p-|\mathcal{L}|} \sigma^{\frac{|\mathcal{L}|}{2}} \left(\sqrt{b}\right)^{-\frac{|\mathcal{L}|}{2}} C_3(\mathcal{L}) \\
&\leq 4^p \left(\sqrt{b}\right)^{\frac{|\mathcal{L}|}{2}-p} C_3(\mathcal{L}) \sigma^{p-\frac{|\mathcal{L}|}{2}} \qquad (21)
\end{aligned}
$$

where $C_3(\mathcal{L}) = \left(2\sqrt{2}\Gamma(\frac{1}{2}) + a^{\frac{3}{2}-2a}\Gamma(2a-1)2^{2a}\right)^{|\mathcal{L}|}$. It follows by (20) that

$$
\begin{aligned}
&\iint_{\mathbb{R}_+\times\mathbb{R}^p} \frac{\exp\left(-\frac{\xi}{\sigma^2}\right)}{(\sigma^2)^{\frac{n+p+2\alpha}{2}-\frac{|\mathcal{L}|}{4}+1}} \left(\prod_{j\in\mathcal{L}} |\beta_j|^{-\frac{1}{2}}\right) \\
&\times \prod_{j\in\mathcal{L}} \left(\frac{(2-2a)^{\frac{3}{2}-2a}}{\left(\sqrt{4b}\frac{|\beta_j|}{\sigma}\right)^{\frac{3}{2}-2a}} + 1\right) \exp\left(-\sum_{j=1}^{p} \frac{\sqrt{b}\,|\beta_j|}{2\sigma}\right) d\boldsymbol{\beta}\, d\sigma^2 \\
&\leq 4^p \left(\sqrt{b}\right)^{\frac{|\mathcal{L}|}{2}-p} C_3(\mathcal{L}) \int_{\mathbb{R}_+} \frac{\exp\left(-\frac{\xi}{\sigma^2}\right)}{(\sigma^2)^{\frac{n}{2}+\alpha+1}} d\sigma^2 \\
&= \frac{4^p \left(\sqrt{b}\right)^{\frac{|\mathcal{L}|}{2}-p} C_3(\mathcal{L})\Gamma(\frac{n}{2}+\alpha)}{\xi^{\frac{n}{2}+\alpha}} < \infty.
\end{aligned}
$$

As discussed above, this establishes the trace class property in the case $1/2 < a < 3/4$.

Case 4: $0 < a < 1/2$

Now, we'll show that when $a \in \left(0, \frac{1}{2}\right)$,

$$
\iint_{\mathbb{R}_+^p\times\mathbb{R}_+} k\left(\left(\boldsymbol{\beta}, \sigma^2\right), \left(\boldsymbol{\beta}, \sigma^2\right)\right) d\boldsymbol{\beta}\, d\sigma^2 = \infty.
$$

Note that

$$
\frac{K_{a-1}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)}{K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)} = \frac{K_{1-a}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)}{K_{\frac{1}{2}-a}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)}.
$$

By [1, Page 375], if $\nu > 0$, then $\frac{K_\nu(x)}{\frac{2^\nu \Gamma(\nu)}{2x^\nu}} \to 1$ as $x \to 0$. Let $y = \sqrt{\frac{2b\beta_j^2}{\sigma^2}}$. It follows that

$$\frac{K_{1-a}(\sqrt{2}y)}{\frac{2^{1-a}\Gamma(1-a)}{2(\sqrt{2}y)^{1-a}}} \left(\frac{K_{\frac{1}{2}-a}(y)}{\frac{2^{\frac{1}{2}-a}\Gamma(\frac{1}{2}-a)}{2y^{\frac{1}{2}-a}}} \right)^{-1} \to 1$$

as $y \to 0$. Hence there exists $\epsilon_1 > 0$ such that

$$\frac{K_{1-a}(\sqrt{2}y)}{\frac{2^{1-a}\Gamma(1-a)}{2(\sqrt{2}y)^{1-a}}} \left(\frac{K_{\frac{1}{2}-a}(y)}{\frac{2^{\frac{1}{2}-a}\Gamma(\frac{1}{2}-a)}{2y^{\frac{1}{2}-a}}} \right)^{-1} \geq \frac{1}{2}$$

for $0 < y < \epsilon_1$. Thus if $0 < y < \epsilon_1$, we have

$$\frac{K_{1-a}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)}{K_{\frac{1}{2}-a}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)} = \frac{K_{1-a}(\sqrt{2}y)}{K_{\frac{1}{2}-a}(y)} \geq \frac{\left(\sqrt{2}\right)^a \Gamma(1-a)}{2\Gamma(\frac{1}{2}-a)} \frac{1}{\sqrt{y}}$$

Since $K_\nu(x) > 0$ for positive ν and x, we have

$$\frac{K_{1-a}(\sqrt{2}y)}{K_{\frac{1}{2}-a}(y)} \geq \frac{K_{1-a}(\sqrt{2}y)}{K_{\frac{1}{2}-a}(y)} I_{(0<y<\epsilon_1)} \geq \frac{\left(\sqrt{2}\right)^a \Gamma(1-a)}{2\Gamma(\frac{1}{2}-a)} \frac{I_{(0<y<\epsilon_1)}}{\sqrt{y}},$$

Using $y = \sqrt{\frac{2b\beta_j^2}{\sigma^2}}$, we get

$$\frac{K_{a-1}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)}{K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)} \geq \frac{\left(\sqrt{2}\right)^a \Gamma(1-a)}{2\,(2b)^{\frac{1}{4}}\,\Gamma(\frac{1}{2}-a)} \frac{\sigma^{\frac{1}{2}}}{|\beta_j|^{\frac{1}{2}}} I_{(0<|\beta_j|<\frac{\sigma\epsilon_1}{\sqrt{2b}})} \tag{22}$$

It follows from the form of the conditional densities in Section 2 that

$$\begin{aligned}
&\iint\limits_{\mathbb{R}^p\times\mathbb{R}_+} k\left(\left(\boldsymbol{\beta},\sigma^2\right),\left(\boldsymbol{\beta},\sigma^2\right)\right)\, d\boldsymbol{\beta}\, d\sigma^2 \\
&= C_1 \iiint\limits_{\mathbb{R}^p\times\mathbb{R}_+^p\times\mathbb{R}_+} \prod_{j=1}^p \frac{\left(2b\sigma^2\right)^{\frac{a-\frac{1}{2}}{2}}}{2\,|\beta_j|^{a-\frac{1}{2}}\, K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)} \tau_j^{\left(a-\frac{1}{2}\right)-1} \\
&\times \left(-\frac{1}{2}\left\{2b\tau_j + \frac{\beta_j^2}{\sigma^2}\frac{1}{\tau_j}\right\}\right)\left(\sigma^2\right)^{-\frac{n+2\alpha}{2}-1} \\
&\times \exp\left(-\frac{1}{2\sigma^2}\left(\mathbf{Y}^T\left(I - X\mathbf{A}_{\boldsymbol{\tau}}^{-1}X^T\right)\mathbf{Y} + 2\xi\right)\right)
\end{aligned}$$

$$\times \left(\mathbf{Y}^T\left(I - X\mathbf{A}_{\boldsymbol{\tau}}^{-1}X^T\right)\mathbf{Y} + 2\xi\right)^{\frac{n+2\alpha}{2}} (\sigma^2)^{-\frac{p}{2}}$$
$$\times |\mathbf{A}_{\boldsymbol{\tau}}|^{\frac{1}{2}} \times \exp\left(-\frac{\left(\boldsymbol{\beta} - \mathbf{A}_{\boldsymbol{\tau}}^{-1}X^T\mathbf{Y}\right)^T \mathbf{A}_{\boldsymbol{\tau}}\left(\boldsymbol{\beta} - \mathbf{A}_{\boldsymbol{\tau}}^{-1}X^T\mathbf{Y}\right)}{2\sigma^2}\right) d\boldsymbol{\beta}\, d\boldsymbol{\tau}\, d\sigma^2 \quad (23)$$

Furthermore, we have

$$|\mathbf{A}_{\boldsymbol{\tau}}|^{\frac{1}{2}} = |X^TX + D_{\boldsymbol{\tau}}^{-1}|^{\frac{1}{2}} \geq |D_{\boldsymbol{\tau}}^{-1}|^{\frac{1}{2}}, \quad \left(\mathbf{Y}^T\left(I - X\mathbf{A}_{\boldsymbol{\tau}}^{-1}X^T\right)\mathbf{Y}\right) + 2\xi \geq 2\xi, \quad (24)$$

and

$$\begin{aligned}
&\exp\left(-\frac{1}{2\sigma^2}\left(\mathbf{Y}^T\left(I - X\mathbf{A}_{\boldsymbol{\tau}}^{-1}X^T\right)\mathbf{Y} + 2\xi\right)\right)\\
&\times \exp\left(-\frac{\left(\boldsymbol{\beta} - \mathbf{A}_{\boldsymbol{\tau}}^{-1}X^T\mathbf{Y}\right)^T \mathbf{A}_{\boldsymbol{\tau}}\left(\boldsymbol{\beta} - \mathbf{A}_{\boldsymbol{\tau}}^{-1}X^T\mathbf{Y}\right)}{2\sigma^2}\right)\\
&= \exp\left(-\frac{2\xi}{2\sigma^2}\right)\exp\left\{-\frac{1}{2\sigma^2}\left(\mathbf{Y}^T\mathbf{Y} + \boldsymbol{\beta}^T(X^TX + D_{\boldsymbol{\tau}}^{-1})\boldsymbol{\beta} - 2\boldsymbol{\beta}^TX^T\mathbf{Y}\right)\right\}\\
&= \exp\left(-\frac{2\xi + \mathbf{Y}^T\mathbf{Y}}{2\sigma^2}\right)\exp\left(-\frac{\boldsymbol{\beta}D_{\boldsymbol{\tau}}^{-1}\boldsymbol{\beta}}{2\sigma^2}\right)\exp\left\{-\frac{1}{2\sigma^2}\left(\boldsymbol{\beta}^TX^TX\boldsymbol{\beta} - 2\boldsymbol{\beta}^TX^T\mathbf{Y}\right)\right\}
\end{aligned} \quad (25)$$

If we denote the entries of X^TX and $X^T\mathbf{Y}$ by a_{ij}, b_i, respectively, then it is easy to see there is at least i such that $a_{ii} > 0$ (if not, $a_{ii} = 0$ for all i, indicating X is exactly $\mathbf{0}$). Without loss of generality, we assume $a_{11} > 0$, so

$$\begin{aligned}
&\exp\left\{-\frac{1}{2\sigma^2}\left(\boldsymbol{\beta}^TX^TX\boldsymbol{\beta} - 2\boldsymbol{\beta}^TX^T\mathbf{Y}\right)\right\}\\
&\quad = g(\sigma^2, \beta_2, \cdots, \beta_p)\exp\left\{-\frac{a_{11}\left(\beta_1 + c\right)^2}{2\sigma^2}\right\}
\end{aligned} \quad (26)$$

where $g(\sigma^2, \beta_2, \cdots, \beta_p) = \exp\left\{\frac{a_{11}c^2}{2\sigma^2} + 2\sum_{i=2}^{p}\beta_i b_i - \sum_{2\leq i,j\leq p} a_{ij}\beta_i\beta_j\right\}$ and $c = \frac{a_{12}\beta_2 + a_{13}\beta_3 + \cdots + a_{1p}\beta_p - b_1}{a_{11}}$. It follows from (23), (24), (25) and (26) that

$$\begin{aligned}
&\iint\limits_{\mathbb{R}_+^p \times \mathbb{R}_+} k\left(\left(\boldsymbol{\beta}, \sigma^2\right), \left(\boldsymbol{\beta}, \sigma^2\right)\right) d\boldsymbol{\beta}\, d\sigma^2\\
&\geq C_1 \iiint\limits_{\mathbb{R}^p \times \mathbb{R}_+^p \times \mathbb{R}_+} \frac{\exp\left(-\frac{2\xi + \mathbf{Y}^T\mathbf{Y}}{2\sigma^2}\right)}{\left(\sigma^2\right)^{\frac{n+2\alpha}{2}+1}} (\sigma^2)^{-\frac{p}{2}}\, g(\sigma^2, \beta_2, \cdots, \beta_p)\\
&\times \exp\left\{-\frac{a_{11}\left(\beta_1 + c\right)^2}{2\sigma^2}\right\} |D_{\boldsymbol{\tau}}^{-1}|^{\frac{1}{2}} \exp\left(-\frac{\boldsymbol{\beta}D_{\boldsymbol{\tau}}^{-1}\boldsymbol{\beta}}{2\sigma^2}\right)
\end{aligned}$$

$$\times \prod_{j=1}^{p} \frac{\left(2b\sigma^2\right)^{\frac{a-\frac{1}{2}}{2}}}{2\,|\beta_j|^{a-\frac{1}{2}}\, K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)} \tau_j^{\left(a-\frac{1}{2}\right)-1}$$

$$\times \exp\left(-\frac{1}{2}\left\{2b\tau_j + \frac{\beta_j^2}{\sigma^2}\frac{1}{\tau_j}\right\}\right) d\boldsymbol{\beta}\, d\boldsymbol{\tau}\, d\sigma^2$$

$$= C_1 \iint\limits_{\mathbb{R}^p\times\mathbb{R}_+} \frac{\exp\left(-\frac{2\xi+\mathbf{Y}^T\mathbf{Y}}{2\sigma^2}\right)}{\left(\sigma^2\right)^{\frac{n+p+2\alpha}{2}+1}} \quad g(\sigma^2, \beta_2, \cdots, \beta_p)$$

$$\times \exp\left\{-\frac{a_{11}}{2\sigma^2}\left(\beta_1 + c\right)^2\right\} \int\limits_{\mathbb{R}_+^p} |D_{\boldsymbol{\tau}}^{-1}|^{\frac{1}{2}} \ \exp\left(-\frac{\boldsymbol{\beta} D_{\boldsymbol{\tau}}^{-1}\boldsymbol{\beta}}{2\sigma^2}\right)$$

$$\times \prod_{j=1}^{p} \frac{\left(2b\sigma^2\right)^{\frac{a-\frac{1}{2}}{2}}}{2\,|\beta_j|^{a-\frac{1}{2}}\, K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)} \tau_j^{\left(a-\frac{1}{2}\right)-1}$$

$$\times \exp\left(-\frac{1}{2}\left\{2b\tau_j + \frac{\beta_j^2}{\sigma^2}\frac{1}{\tau_j}\right\}\right) d\boldsymbol{\tau}\, d\boldsymbol{\beta}\, d\sigma^2 \tag{27}$$

By (22), the inner integral can be bounded below as

$$\int\limits_{\mathbb{R}_+^p} |D_{\boldsymbol{\tau}}^{-1}|^{\frac{1}{2}} \ \exp\left(-\frac{\boldsymbol{\beta} D_{\boldsymbol{\tau}}^{-1}\boldsymbol{\beta}}{2\sigma^2}\right) \ \prod_{j=1}^{p} \frac{\left(2b\sigma^2\right)^{\frac{a-\frac{1}{2}}{2}}}{2\,|\beta_j|^{a-\frac{1}{2}}\, K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)} \tau_j^{\left(a-\frac{1}{2}\right)-1}$$

$$\times \exp\left(-\frac{1}{2}\left\{2b\tau_j + \frac{\beta_j^2}{\sigma^2}\frac{1}{\tau_j}\right\}\right) d\boldsymbol{\tau}$$

$$= \prod_{j=1}^{p} \int\limits_{\mathbb{R}_+} \frac{\left(2b\sigma^2\right)^{\frac{a-\frac{1}{2}}{2}}}{2\,|\beta_j|^{a-\frac{1}{2}}\, K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)} \tau_j^{(a-1)-1} \exp\left(-\frac{1}{2}\left\{2b\tau_j + \frac{2\beta_j^2}{\sigma^2}\frac{1}{\tau_j}\right\}\right) d\tau_j$$

$$= \prod_{j=1}^{p} \left(2b\sigma^2\right)^{\frac{1}{4}} \left(\sqrt{2}\right)^{a-1} |\beta_j|^{-\frac{1}{2}} \ \frac{K_{a-1}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)}{K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)}$$

$$\geq \prod_{j=1}^{p} \left(2b\sigma^2\right)^{\frac{1}{4}} \left(\sqrt{2}\right)^{a-1} |\beta_j|^{-\frac{1}{2}} \ \frac{\left(\sqrt{2}\right)^a \Gamma(1-a)}{2\,(2b)^{\frac{1}{4}}\,\Gamma(\frac{1}{2}-a)} \frac{\sigma^{\frac{1}{2}}}{|\beta_j|^{\frac{1}{2}}} I_{\left(0<|\beta_j|<\frac{\sigma\epsilon_1}{\sqrt{2b}}\right)}$$

$$= \frac{\left(\sqrt{2}\right)^{2a-3}\Gamma(1-a)}{\Gamma(\frac{1}{2}-a)} (\sigma^2)^{\frac{p}{2}} \prod_{j=1}^{p} \frac{1}{|\beta_j|} I_{\left(0<|\beta_j|<\frac{\sigma\epsilon_1}{\sqrt{2b}}\right)} \tag{28}$$

It follows from (27) and (28) that

$$\begin{aligned}
&\iint\limits_{\mathbb{R}_+^p\times\mathbb{R}_+} k\left(\left(\boldsymbol{\beta},\sigma^2\right),\left(\boldsymbol{\beta},\sigma^2\right)\right)\,d\boldsymbol{\beta}\,d\sigma^2\\
&\geq \frac{\left(\sqrt{2}\right)^{2a-3}\Gamma(1-a)C_1}{\Gamma(\frac{1}{2}-a)}\iint\limits_{\mathbb{R}^p\times\mathbb{R}_+}\frac{\exp\left(-\frac{2\xi+\mathbf{Y}^T\mathbf{Y}}{2\sigma^2}\right)}{\left(\sigma^2\right)^{\frac{n+2\alpha}{2}+1}}\\
&\times g(\sigma^2,\beta_2,\cdots,\beta_p)\exp\left\{-\frac{a_{11}\left(\beta_1+c\right)^2}{2\sigma^2}\right\}\\
&\times\prod_{j=1}^p\frac{1}{|\beta_j|}I_{(0<|\beta_j|<\frac{\sigma\epsilon_1}{\sqrt{2b}})}\,d\boldsymbol{\beta}\,d\sigma^2\\
&= C_1^*\int\limits_{\mathbb{R}_+}\frac{\exp\left(-\frac{2\xi+\mathbf{Y}^T\mathbf{Y}}{2\sigma^2}\right)}{\left(\sigma^2\right)^{\frac{n+2\alpha}{2}+1}}\left\{\int_{-\frac{\sigma\epsilon_1}{\sqrt{2b}}}^{\frac{\sigma\epsilon_1}{\sqrt{2b}}}\frac{1}{|\beta_1|}\exp\left\{-\frac{a_{11}\left(\beta_1+c\right)^2}{2\sigma^2}\right\}d\beta_1\right\}\\
&\times\left\{\int_{\mathbb{R}^{p-1}}g(\sigma^2,\beta_2,\cdots,\beta_p)\prod_{j=2}^p\frac{1}{|\beta_j|}I_{(0<|\beta_j|<\frac{\sigma\epsilon_1}{\sqrt{2b}})}\prod_{j=2}^p d\beta_j\right\}d\sigma^2 \qquad (29)
\end{aligned}$$

where $C_1^* = \frac{\left(\sqrt{2}\right)^{2a-3}\Gamma(1-a)C_1}{\Gamma(\frac{1}{2}-a)}$. However, we note that

$$\int_{-\frac{\sigma\epsilon_1}{\sqrt{2b}}}^{\frac{\sigma\epsilon_1}{\sqrt{2b}}}\frac{1}{|\beta_1|}\exp\left\{-\frac{a_{11}\left(\beta_1+c\right)^2}{2\sigma^2}\right\}d\beta_1 \geq \int\limits_0^{\frac{\sigma\epsilon_1}{\sqrt{2b}}}\frac{1}{\beta_1}\exp\left\{-\frac{a_{11}\left(\beta_1+c\right)^2}{2\sigma^2}\right\}d\beta_1=\infty$$

where the last step follows from Propositon A1. By (29), it follows that the operator corresponding to the Markov transition density k is not trace class when $0<a<1/2$.

Case 5: $a=1/2$
Finally, we show that when $a=\frac{1}{2}$, we have

$$\iint\limits_{\mathbb{R}_+^p\times\mathbb{R}_+} k\left(\left(\boldsymbol{\beta},\sigma^2\right),\left(\boldsymbol{\beta},\sigma^2\right)\right)\,d\boldsymbol{\beta}\,d\sigma^2=\infty.$$

When $a=\frac{1}{2}$, $\frac{K_{a-1}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)}{K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)}=\frac{K_{-\frac{1}{2}}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)}{K_0\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)}=\frac{K_{\frac{1}{2}}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)}{K_0\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)}$. By [1, Page 375], if $z\to 0$, then $K_0(z)\sim-\ln(z)$ and $K_{\frac{1}{2}}(z)\sim\frac{\Gamma(\frac{1}{2})}{2}\left(\frac{2}{z}\right)^{\frac{1}{2}}$. As we did in Case 4, let $y=\sqrt{\frac{2b\beta_j^2}{\sigma^2}}$. It follows that

$$\frac{K_{\frac{1}{2}}(\sqrt{2}y)}{\frac{\Gamma(\frac{1}{2})}{2}\left(\frac{2}{\sqrt{2}y}\right)^{\frac{1}{2}}}\left(\frac{K_0(y)}{-\ln(y)}\right)^{-1}\to 1 \quad \text{as } y\to 0.$$

Hence there exists $\epsilon_2\in(0,1)$ such that $\frac{K_{\frac{1}{2}}(\sqrt{2}y)}{K_0(y)}\geq\frac{\Gamma(\frac{1}{2})2^{\frac{1}{4}}}{4}\frac{1}{-\sqrt{y}\ln(\sqrt{2}y)}$ for $0<y<\epsilon_2$. It follows that

$$\begin{aligned}\frac{K_{\frac{1}{2}}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)}{K_0\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)} = \frac{K_{\frac{1}{2}}(\sqrt{2}y)}{K_0(y)} &\geq \frac{K_{\frac{1}{2}}(\sqrt{2}y)}{K_0(y)}I_{(0<y<\epsilon_2)}\\ &\geq\frac{\Gamma(\frac{1}{2})2^{\frac{1}{4}}}{4}\frac{1}{-\sqrt{y}\ln(\sqrt{2}y)}I_{(0<y<\epsilon_2)}\\ &=\frac{C_5(\sigma^2)^{\frac{1}{4}}}{|\beta_j|^{\frac{1}{2}}\left(-\ln(\sqrt{2b})+\ln\sigma-\ln|\beta_j|\right)}I_{(0<|\beta_j|<\frac{\sigma\epsilon_2}{\sqrt{2b}})},\end{aligned}$$

where $C_5=\frac{\Gamma(\frac{1}{2})}{4b^{\frac{1}{4}}}$. We use this to get a lower bound for the inner integral with respect to $\boldsymbol{\tau}$ in (27). In particular, we note that

$$\begin{aligned}&\int_{\mathbb{R}_+^p}|D_{\boldsymbol{\tau}}^{-1}|^{\frac{1}{2}}\exp\left(-\frac{\boldsymbol{\beta}D_{\boldsymbol{\tau}}^{-1}\boldsymbol{\beta}}{2\sigma^2}\right)\prod_{j=1}^p\frac{\left(2b\sigma^2\right)^{\frac{a-\frac{1}{2}}{2}}}{2\,|\beta_j|^{a-\frac{1}{2}}K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)}\tau_j^{(a-\frac{1}{2})-1}\\ &\times\exp\left(-\frac{1}{2}\left\{2b\tau_j+\frac{\beta_j^2}{\sigma^2}\frac{1}{\tau_j}\right\}\right)d\boldsymbol{\tau}\\ &=\prod_{j=1}^p\left(2b\sigma^2\right)^{\frac{1}{4}}\left(\sqrt{2}\right)^{a-1}|\beta_j|^{-\frac{1}{2}}\;\frac{K_{-\frac{1}{2}}\left(\sqrt{4b\frac{\beta_j^2}{\sigma^2}}\right)}{K_0\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)}\\ &\geq\prod_{j=1}^p\left(2b\sigma^2\right)^{\frac{1}{4}}\left(\sqrt{2}\right)^{a-1}|\beta_j|^{-\frac{1}{2}}\;\frac{C_5(\sigma^2)^{\frac{1}{4}}}{|\beta_j|^{\frac{1}{2}}\left(-\ln(\sqrt{2b})+\ln\sigma-\ln|\beta_j|\right)}I_{(0<|\beta_j|<\frac{\sigma\epsilon_2}{\sqrt{2b}})}\\ &=(2b)^{\frac{p}{4}}(\sqrt{2})^{p(a-1)}C_5^p\;(\sigma^2)^{\frac{p}{2}}\prod_{j=1}^p\frac{I_{(0<|\beta_j|<\frac{\sigma\epsilon_2}{\sqrt{2b}})}}{|\beta_j|\left(-\ln(\sqrt{2b})+\ln\sigma-\ln|\beta_j|\right)}\end{aligned}$$

Using (27), it follows that

$$\iint_{\mathbb{R}_+^p\times\mathbb{R}_+}k\left(\left(\boldsymbol{\beta},\sigma^2\right),\left(\boldsymbol{\beta},\sigma^2\right)\right)d\boldsymbol{\beta}\,d\sigma^2$$

$$
\begin{aligned}
&\geq C_1 \iint\limits_{\mathbb{R}^p \times \mathbb{R}_+} \frac{\exp\left(-\frac{2\xi + \mathbf{Y}^T\mathbf{Y}}{2\sigma^2}\right)}{(\sigma^2)^{\frac{n+p+2\alpha}{2}+1}} \; g(\sigma^2, \beta_2, \cdots, \beta_p) \; \exp\left\{-\frac{a_{11}}{2\sigma^2}(\beta_1 + c)^2\right\} \\
&\quad \times \int\limits_{\mathbb{R}_+^p} |D_{\boldsymbol{\tau}}^{-1}|^{\frac{1}{2}} \; \exp\left(-\frac{\boldsymbol{\beta} D_{\boldsymbol{\tau}}^{-1}\boldsymbol{\beta}}{2\sigma^2}\right) \\
&\quad \times \prod_{j=1}^{p} \frac{\left(2b\sigma^2\right)^{\frac{a-\frac{1}{2}}{2}}}{2\,|\beta_j|^{a-\frac{1}{2}}\, K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)} \tau_j^{\left(a-\frac{1}{2}\right)-1} \\
&\quad \times \exp\left(-\frac{1}{2}\left\{2b\tau_j + \frac{\beta_j^2}{\sigma^2}\frac{1}{\tau_j}\right\}\right) \, d\boldsymbol{\tau} \, d\boldsymbol{\beta} \, d\sigma^2 \\
&\geq (2b)^{\frac{p}{4}} (\sqrt{2})^{p(a-1)} C_1 C_5^p \iint\limits_{\mathbb{R}^{p-1}\times\mathbb{R}_+} \frac{\exp\left(-\frac{2\xi + \mathbf{Y}^T\mathbf{Y}}{2\sigma^2}\right)}{(\sigma^2)^{\frac{n+2\alpha}{2}+1}} g(\sigma^2, \beta_2, \cdots, \beta_p) \\
&\quad \times \prod_{j=2}^{p} \frac{I_{(0<|\beta_j|<\frac{\sigma\epsilon_2}{\sqrt{2b}})}}{|\beta_j|\left(-\ln(\sqrt{2b}) + \ln\sigma - \ln|\beta_j|\right)} \\
&\quad \times \left\{ \int\limits_{R} \exp\left\{-\frac{a_{11}}{2\sigma^2}(\beta_1 + c)^2\right\} \frac{I_{(0<|\beta_1|<\frac{\sigma\epsilon_2}{\sqrt{2b}})}}{|\beta_1|\left(-\ln(\sqrt{2b}) + \ln\sigma - \ln|\beta_1|\right)} \, d\beta_1 \right\} d\boldsymbol{\beta}' \, d\sigma^2,
\end{aligned}
$$

where $\boldsymbol{\beta}' = (\beta_2, \cdots, \beta_p)$. By Proposition A2, we obtain

$$
\begin{aligned}
&\int\limits_{R} \exp\left\{-\frac{a_{11}}{2\sigma^2}(\beta_1 + c)^2\right\} \frac{I_{(0<|\beta_1|<\frac{\sigma\epsilon_2}{\sqrt{2b}})}}{|\beta_1|\left(-\ln(\sqrt{2b}) + \ln\sigma - \ln|\beta_1|\right)} \, d\beta_1 \\
&\geq \int\limits_0^{\frac{\sigma\epsilon_2}{\sqrt{2b}}} \frac{\exp\left\{-\frac{a_{11}}{2\sigma^2}(\beta_1 + c)^2\right\}}{\beta_1\left(-\ln(\sqrt{2b}) + \ln\sigma - \ln\beta_1\right)} \, d\beta_1 \\
&= \infty
\end{aligned}
$$

It follows that the operator corresponding to the Markov transition density k is not trace class when $a = \frac{1}{2}$. □

4. Properties of the three-block Gibbs sampler

In this section, we show that when $a > 0$, the Markov operator corresponding to the three-block Gibbs sampler $\tilde{\Phi}$, with Markov transition density $\tilde{k}$ specified in (1), is not Hilbert-Schmidt. Let $\tilde{K}$ be the Markov operator corresponding to $\tilde{\Phi}$. We prove the following result.

Theorem 2. *For all $a > 0$, the Markov operator $\tilde{K}$ is not Hilbert-Schmidt for all possible values of p and n.*

Proof. Note that the Markov operator $\tilde{K}$ corresponding to the density $\tilde{k}$ is Hilbert-Schmidt if and only if $\tilde{K}^*\tilde{K}$ is trace class (see [13], for example). Here $\tilde{K}^*$ denotes the adjoint of $\tilde{K}$. It follows that $\tilde{K}$ is Hilbert-Schmidt if and only if $I < \infty$, where

$$\begin{aligned} I &:= \int_{\mathbb{R}^p}\int_{\mathbb{R}_+}\int_{\mathbb{R}^p}\int_{\mathbb{R}_+} \tilde{k}\Big((\boldsymbol{\beta},\sigma^2), \big(\tilde{\boldsymbol{\beta}},\tilde{\sigma}^2\big)\Big)\tilde{k}^*\Big(\big(\tilde{\boldsymbol{\beta}},\tilde{\sigma}^2\big), (\boldsymbol{\beta},\sigma^2)\Big) d\boldsymbol{\beta} d\sigma^2 d\tilde{\boldsymbol{\beta}} d\tilde{\sigma^2} \\ &= \int_{\mathbb{R}^p}\int_{\mathbb{R}_+}\int_{\mathbb{R}^p}\int_{\mathbb{R}_+} \tilde{k}^2\Big((\boldsymbol{\beta},\sigma^2) \ , \ \big(\tilde{\boldsymbol{\beta}},\tilde{\sigma}^2\big)\Big) \frac{\pi\left(\boldsymbol{\beta},\sigma^2 \mid \mathbf{Y}\right)}{\pi\left(\tilde{\boldsymbol{\beta}},\tilde{\sigma}^2 \mid \mathbf{Y}\right)} d\boldsymbol{\beta} d\sigma^2 d\tilde{\boldsymbol{\beta}} d\tilde{\sigma^2}. \end{aligned} \tag{30}$$

The last step follows from the detailed balance condition for $\tilde{k}$ and its adjoint $\tilde{k}^*$, i.e.,

$$\pi\left(\boldsymbol{\beta},\sigma^2 \mid \mathbf{Y}\right)\tilde{k}\Big((\boldsymbol{\beta},\sigma^2), \big(\tilde{\boldsymbol{\beta}},\tilde{\sigma}^2\big)\Big) = \pi\left(\tilde{\boldsymbol{\beta}},\tilde{\sigma}^2 \mid \mathbf{Y}\right)\tilde{k}^*\Big(\big(\tilde{\boldsymbol{\beta}},\tilde{\sigma}^2\big), (\boldsymbol{\beta},\sigma^2)\Big).$$

By (2), a straightforward manipulation of conditional densities, and Fubini's theorem, we obtain

$$\begin{aligned} I = &\int_{\mathbb{R}_+}\int_{\mathbb{R}_+}\int_{\mathbb{R}^p}\int_{\mathbb{R}^p}\int_{\mathbb{R}_+^p}\int_{\mathbb{R}_+^p} \pi\left(\tilde{\sigma}^2 \mid \tilde{\boldsymbol{\beta}},\boldsymbol{\tau},\mathbf{Y}\right)\pi\left(\tilde{\boldsymbol{\beta}} \mid \boldsymbol{\tau},\sigma^2,\mathbf{Y}\right)\pi\left(\boldsymbol{\tau} \mid \boldsymbol{\beta},\sigma^2,\mathbf{Y}\right) \times \\ &\pi\left(\boldsymbol{\beta} \mid \tilde{\boldsymbol{\tau}},\sigma^2,\mathbf{Y}\right)\pi\left(\sigma^2 \mid \tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\tau}},\mathbf{Y}\right)\pi\left(\tilde{\boldsymbol{\tau}} \mid \tilde{\boldsymbol{\beta}},\tilde{\sigma}^2,\mathbf{Y}\right) d\sigma^2 d\tilde{\sigma^2} d\boldsymbol{\beta} d\tilde{\boldsymbol{\beta}} d\boldsymbol{\tau} d\tilde{\boldsymbol{\tau}} \end{aligned} \tag{31}$$

For convenience, we introduce and use the following notation in the subsequent proof.

$$\begin{aligned} \widehat{\boldsymbol{\beta}} &= \mathbf{A}_{\boldsymbol{\tau}}^{-1}X^T\mathbf{Y} & \widehat{\boldsymbol{\beta}}_* &= \mathbf{A}_{\tilde{\boldsymbol{\tau}}}^{-1}X^T\mathbf{Y} \\ \Delta_1 &= (\tilde{\boldsymbol{\beta}}-\widehat{\boldsymbol{\beta}})^T\mathbf{A}_{\boldsymbol{\tau}}(\tilde{\boldsymbol{\beta}}-\widehat{\boldsymbol{\beta}}) & \Delta_{1*} &= (\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}_*)^T\mathbf{A}_{\tilde{\boldsymbol{\tau}}}(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}_*) \\ \tilde{\Delta} &= (\mathbf{Y}-X\tilde{\boldsymbol{\beta}})^T(\mathbf{Y}-X\tilde{\boldsymbol{\beta}}) & \tilde{\Delta}_* &= (\mathbf{Y}-X\tilde{\boldsymbol{\beta}})^T(\mathbf{Y}-X\tilde{\boldsymbol{\beta}}) \\ &+\tilde{\boldsymbol{\beta}}^TD_{\boldsymbol{\tau}}^{-1}+\tilde{\boldsymbol{\beta}}+2\xi & &+\tilde{\boldsymbol{\beta}}^TD_{\tilde{\boldsymbol{\tau}}}^{-1}\tilde{\boldsymbol{\beta}}+2\xi. \end{aligned} \tag{32}$$

(33)

We first show $I = \infty$ for the simpler case with $a > \frac{1}{2}$ and then consider the significantly more complicated case $0 < a \leq \frac{1}{2}$.

Case 1: $a > 1/2$

Using $2K_\nu(x) \leq x^{-\nu}\Gamma(\nu)2^\nu$ for $\nu > 0, x > 0$ (Proposition A7 of [21]), we obtain that if $a > \frac{1}{2}$,

$$\frac{\left(2b\sigma^2\right)^{\frac{a-\frac{1}{2}}{2}}}{2\left|\beta_j\right|^{a-\frac{1}{2}}K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)} \geq \frac{(2b)^{a-\frac{1}{2}}}{\Gamma\left(a-\frac{1}{2}\right)2^{a-\frac{1}{2}}}. \tag{34}$$

Similarly

$$\frac{\left(2b\tilde{\sigma}^2\right)^{\frac{a-\frac{1}{2}}{2}}}{2\left|\tilde{\beta}_j\right|^{a-\frac{1}{2}} K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\tilde{\beta}_j^2}{\tilde{\sigma}^2}}\right)} \geq \frac{(2b)^{a-\frac{1}{2}}}{\Gamma(a-\frac{1}{2})2^{a-\frac{1}{2}}}. \tag{35}$$

Using the conditional densities from Section 2, along with (34) and (35), we get

$$\begin{aligned}
& \pi\left(\tilde{\sigma}^2 \mid \tilde{\boldsymbol{\beta}}, \boldsymbol{\tau}, \mathbf{Y}\right) \pi\left(\tilde{\boldsymbol{\beta}} \mid \boldsymbol{\tau}, \sigma^2, \mathbf{Y}\right) \pi\left(\boldsymbol{\tau} \mid \boldsymbol{\beta}, \sigma^2, \mathbf{Y}\right) \\
& \times \pi\left(\boldsymbol{\beta} \mid \tilde{\boldsymbol{\tau}}, \sigma^2, \mathbf{Y}\right) \pi\left(\sigma^2 \mid \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\tau}}, \mathbf{Y}\right) \pi\left(\tilde{\boldsymbol{\tau}} \mid \tilde{\boldsymbol{\beta}}, \tilde{\sigma}^2, \mathbf{Y}\right) \\
= \; & D_1\left\{\frac{\tilde{\Delta}^{\frac{n+p+2\alpha}{2}} \exp\left(-\frac{\tilde{\Delta}}{2\tilde{\sigma}^2}\right)}{(\tilde{\sigma}^2)^{\frac{n+p+2\alpha}{2}+1}}\right\}\left\{\frac{|\mathbf{A}_{\boldsymbol{\tau}}|^{\frac{1}{2}} \exp\left(-\frac{\Delta_1}{2\sigma^2}\right)}{\sigma^p}\right\} \\
& \times\left\{\prod_{j=1}^{p} \frac{\left(2b\sigma^2\right)^{\frac{a-\frac{1}{2}}{2}}}{2\left|\beta_j\right|^{a-\frac{1}{2}} K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)} \tau_j^{\left(a-\frac{1}{2}\right)-1} \exp\left(-\frac{1}{2}\left\{2b\tau_j + \frac{\beta_j^2}{\sigma^2}\frac{1}{\tau_j}\right\}\right)\right\} \\
& \times\left\{\frac{|\mathbf{A}_{\tilde{\boldsymbol{\tau}}}|^{\frac{1}{2}} \exp\left(-\frac{\Delta_{1*}}{2\sigma^2}\right)}{\sigma^p}\right\}\left\{\frac{\tilde{\Delta}_*^{\frac{n+p+2\alpha}{2}} \exp\left(-\frac{\tilde{\Delta}_*}{2\sigma^2}\right)}{(\sigma^2)^{\frac{n+p+2\alpha}{2}+1}}\right\} \\
& \times\left\{\prod_{j=1}^{p} \frac{\left(2b\tilde{\sigma}^2\right)^{\frac{a-\frac{1}{2}}{2}}}{2\left|\tilde{\beta}_j\right|^{a-\frac{1}{2}} K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\tilde{\beta}_j^2}{\tilde{\sigma}^2}}\right)} \tilde{\tau}_j^{\left(a-\frac{1}{2}\right)-1} \exp\left(-\frac{1}{2}\left\{2b\tilde{\tau}_j + \frac{\tilde{\beta}_j^2}{\tilde{\sigma}^2}\frac{1}{\tilde{\tau}_j}\right\}\right)\right\} \\
\geq \; & D_1 f_1(\boldsymbol{\tau}, \tilde{\boldsymbol{\tau}})\left\{\frac{\tilde{\Delta}^{\frac{n+p+2\alpha}{2}} \exp\left(-\frac{\tilde{\Delta}+\tilde{\boldsymbol{\beta}}^T D_{\tilde{\boldsymbol{\tau}}}^{-1}\tilde{\boldsymbol{\beta}}}{2\tilde{\sigma}^2}\right)}{(\tilde{\sigma}^2)^{\frac{n+p+2\alpha}{2}+1}}\right\} \\
& \times\left\{\frac{\tilde{\Delta}_*^{\frac{n+p+2\alpha}{2}} \exp\left(-\frac{\Delta_1+\Delta_{1*}+\tilde{\Delta}_*+\boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1}\boldsymbol{\beta}}{2\sigma^2}\right)}{(\sigma^2)^{\frac{n+p+2\alpha}{2}+p+1}}\right\}
\end{aligned} \tag{36}$$

where

$$D_1 = \frac{1}{\left[(2\pi)^p 2^{\frac{n+p+2\alpha}{2}}\Gamma\left(\frac{n+p+2\alpha}{2}\right)\right]^2}\left(\frac{(2b)^{a-\frac{1}{2}}}{\Gamma(a-\frac{1}{2})2^{a-\frac{1}{2}}}\right)^{2p}$$

and

$$\begin{aligned}
& f_1(\boldsymbol{\tau}, \tilde{\boldsymbol{\tau}}) \\
& \quad = \left\{\prod_{j=1}^{p} \tau_j^{\left(a-\frac{1}{2}\right)-1} \exp\left(-b\tau_j\right)\right\}\left\{\prod_{j=1}^{p} \tilde{\tau}_j^{\left(a-\frac{1}{2}\right)-1} \exp\left(-b\tilde{\tau}_j\right)\right\}|\mathbf{A}_{\boldsymbol{\tau}}|^{\frac{1}{2}}|\mathbf{A}_{\tilde{\boldsymbol{\tau}}}|^{\frac{1}{2}}.
\end{aligned}$$

It follows from (36) that

$$\int_{\mathbb{R}_+}\int_{\mathbb{R}_+}\int_{\mathbb{R}^p}\int_{\mathbb{R}^p} \pi\left(\tilde{\sigma}^2 \mid \tilde{\boldsymbol{\beta}}, \boldsymbol{\tau}, \mathbf{Y}\right) \pi\left(\tilde{\boldsymbol{\beta}} \mid \boldsymbol{\tau}, \sigma^2, \mathbf{Y}\right) \pi\left(\boldsymbol{\tau} \mid \boldsymbol{\beta}, \sigma^2, \mathbf{Y}\right) \times$$

$$
\begin{aligned}
& \pi\left(\boldsymbol{\beta} \mid \tilde{\boldsymbol{\tau}}, \sigma^2, \mathbf{Y}\right) \pi\left(\sigma^2 \mid \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\tau}}, \mathbf{Y}\right) \pi\left(\tilde{\boldsymbol{\tau}} \mid \tilde{\boldsymbol{\beta}}, \tilde{\sigma}^2, \mathbf{Y}\right) d\sigma^2 d\tilde{\sigma}^2 d\boldsymbol{\beta} d\tilde{\boldsymbol{\beta}} \\
\geq \quad & D_1 f_1(\boldsymbol{\tau}, \tilde{\boldsymbol{\tau}}) \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} \left\{ \frac{\tilde{\Delta}^{\frac{n+p+2\alpha}{2}} \exp\left(-\frac{\tilde{\Delta}+\tilde{\boldsymbol{\beta}}^T D_{\tilde{\boldsymbol{\tau}}}^{-1} \tilde{\boldsymbol{\beta}}}{2\tilde{\sigma}^2}\right)}{(\tilde{\sigma}^2)^{\frac{n+p+2\alpha}{2}+1}} \right\} \\
& \times \left\{ \frac{\tilde{\Delta}_*^{\frac{n+p+2\alpha}{2}} \exp\left(-\frac{\Delta_1+\Delta_{1*}+\tilde{\Delta}_*+\boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1} \boldsymbol{\beta}}{2\sigma^2}\right)}{(\sigma^2)^{\frac{n+p+2\alpha}{2}+p+1}} \right\} d\sigma^2 d\tilde{\sigma}^2 d\boldsymbol{\beta} d\tilde{\boldsymbol{\beta}} \\
\overset{(a)}{=} \quad & \infty
\end{aligned}
$$

for every $(\boldsymbol{\tau}, \tilde{\boldsymbol{\tau}}) \in \mathbb{R}_+^p \times \mathbb{R}_+^p$. Here (a) follows by repeating verbatim the arguments between Equations (S4) - (S12) in [29]. We conclude from this fact that the Markov operator $\tilde{K}$ is not Hilbert-Schmidt when $a > \frac{1}{2}$.

Case 2: $0 < a \leq 1/2$

By the integral formula (see [1], Page 376)

$$K_\nu(t) = \int_0^\infty \exp\left(-t \cosh z\right) \cosh(vz)\, dz, \nu \in \mathbb{R}.$$

Since the hyperbolic function cosh is strictly decreasing on interval $(-\infty, 0]$, for every $x > 0$, $K_\nu(x)$ is strictly decreasing as ν increases on the interval $(-\infty, 0]$. Note that when $0 < a \leq \frac{1}{2}$, $-a - \frac{3}{2} < a - \frac{1}{2} \leq 0$. It follows that

$$K_{a-\frac{1}{2}}(x) < K_{-a-\frac{3}{2}}(x)$$

for all $x > 0$. Moreover, when $\nu < 0$ and $x > 0$, $2K_\nu(x) \leq x^\nu \Gamma(-\nu) 2^{-\nu}$ (see Proposition A7 of [21]), which implies

$$2K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right) < 2K_{-a-\frac{3}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right) \leq \left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)^{-a-\frac{3}{2}} \Gamma(a+\frac{3}{2})\; 2^{a+\frac{3}{2}}$$

and

$$\frac{\left(2b\sigma^2\right)^{\frac{a-\frac{1}{2}}{2}}}{2\left|\beta_j\right|^{a-\frac{1}{2}} K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)} \geq \frac{b^{a+\frac{1}{2}}}{2\Gamma(a+\frac{3}{2})} \left(\frac{\beta_j^2}{\sigma^2}\right). \tag{37}$$

Similarly, we get

$$\frac{\left(2b\tilde{\sigma}^2\right)^{\frac{a-\frac{1}{2}}{2}}}{2\left|\tilde{\beta}_j\right|^{a-\frac{1}{2}} K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\tilde{\beta}_j^2}{\tilde{\sigma}^2}}\right)} \geq \frac{b^{a+\frac{1}{2}}}{2\Gamma(a+\frac{3}{2})} \left(\frac{\tilde{\beta}_j^2}{\tilde{\sigma}^2}\right). \tag{38}$$

Using the conditional densities from Section 2, along with (37) and (38), we obtain

$$\pi\left(\tilde{\sigma}^2 \mid \tilde{\boldsymbol{\beta}}, \boldsymbol{\tau}, \mathbf{Y}\right) \pi\left(\tilde{\boldsymbol{\beta}} \mid \boldsymbol{\tau}, \sigma^2, \mathbf{Y}\right) \pi\left(\boldsymbol{\tau} \mid \boldsymbol{\beta}, \sigma^2, \mathbf{Y}\right)$$

$$
\begin{aligned}
&\times \pi\left(\boldsymbol{\beta} \mid \tilde{\boldsymbol{\tau}}, \sigma^2, \mathbf{Y}\right) \pi\left(\sigma^2 \mid \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\tau}}, \mathbf{Y}\right) \pi\left(\tilde{\boldsymbol{\tau}} \mid \tilde{\boldsymbol{\beta}}, \tilde{\sigma}^2, \mathbf{Y}\right) \\
= \quad & D_2 \left\{ \frac{\tilde{\Delta}^{\frac{n+p+2\alpha}{2}} \exp\left(-\frac{\tilde{\Delta}}{2\tilde{\sigma}^2}\right)}{(\tilde{\sigma}^2)^{\frac{n+p+2\alpha}{2}+1}} \right\} \left\{ \frac{|\mathbf{A}_{\boldsymbol{\tau}}|^{\frac{1}{2}} \exp\left(-\frac{\Delta_1}{2\sigma^2}\right)}{\sigma^p} \right\} \\
&\times \left\{ \prod_{j=1}^p \frac{\left(2b\sigma^2\right)^{\frac{a-\frac{1}{2}}{2}}}{2\left|\beta_j\right|^{a-\frac{1}{2}} K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)} \tau_j^{\left(a-\frac{1}{2}\right)-1} \exp\left(-\frac{1}{2}\left\{2b\tau_j + \frac{\beta_j^2}{\sigma^2}\frac{1}{\tau_j}\right\}\right) \right\} \\
&\times \left\{ \frac{|\mathbf{A}_{\tilde{\boldsymbol{\tau}}}|^{\frac{1}{2}} \exp\left(-\frac{\Delta_{1*}}{2\sigma^2}\right)}{\sigma^p} \right\} \times \left\{ \frac{\tilde{\Delta}_*^{\frac{n+p+2\alpha}{2}} \exp\left(-\frac{\tilde{\Delta}_*}{2\sigma^2}\right)}{(\sigma^2)^{\frac{n+p+2\alpha}{2}+1}} \right\} \\
&\times \left\{ \prod_{j=1}^p \frac{\left(2b\tilde{\sigma}^2\right)^{\frac{a-\frac{1}{2}}{2}}}{2\left|\tilde{\beta}_j\right|^{a-\frac{1}{2}} K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\tilde{\beta}_j^2}{\tilde{\sigma}^2}}\right)} \tilde{\tau}_j^{\left(a-\frac{1}{2}\right)-1} \exp\left(-\frac{1}{2}\left\{2b\tilde{\tau}_j + \frac{\tilde{\beta}_j^2}{\tilde{\sigma}^2}\frac{1}{\tilde{\tau}_j}\right\}\right) \right\} \\
\geq \quad & D_2 f_1(\boldsymbol{\tau}, \tilde{\boldsymbol{\tau}}) \prod_{j=1}^p \left(\beta_j^2 \tilde{\beta}_j^2\right) \left\{ \frac{\exp\left(-\frac{\tilde{\Delta} + \tilde{\boldsymbol{\beta}}^T D_{\tilde{\boldsymbol{\tau}}}^{-1} \tilde{\boldsymbol{\beta}}}{2\tilde{\sigma}^2}\right)}{(\tilde{\sigma}^2)^{\frac{n+p+2\alpha}{2}+p+1}} \right\} \\
&\times \left\{ \frac{\exp\left(-\frac{\Delta_1 + \Delta_{1*} + \tilde{\Delta}_* + \boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1} \boldsymbol{\beta}}{2\sigma^2}\right)}{(\sigma^2)^{\frac{n+p+2\alpha}{2}+2p+1}} \right\}, \qquad (39)
\end{aligned}
$$

where

$$
D_2 = \frac{(2\xi)^{n+p+2\alpha}}{\left[(2\pi)^p 2^{\frac{n+p+2\alpha}{2}} \Gamma\left(\frac{n+p+2\alpha}{2}\right)\right]^2} \left(\frac{b^{a+\frac{1}{2}}}{2\Gamma(a+\frac{3}{2})}\right)^{2p},
$$

$$
f_1(\boldsymbol{\tau}, \tilde{\boldsymbol{\tau}}) = \left\{ \prod_{j=1}^p \tau_j^{a-2} \exp\left(-b\tau_j\right) \right\} \left\{ \prod_{j=1}^p \tilde{\tau}_j^{a-2} \exp\left(-b\tilde{\tau}_j\right) \right\},
$$

and the last inequality follows by

$$
\tilde{\Delta}^{\frac{n+p+2\alpha}{2}} \geq (2\xi)^{\frac{n+p+2\alpha}{2}}, \ \tilde{\Delta}_*^{\frac{n+p+2\alpha}{2}} \geq (2\xi)^{\frac{n+p+2\alpha}{2}}, \ |\mathbf{A}_{\boldsymbol{\tau}}|^{\frac{1}{2}} \geq |D_{\boldsymbol{\tau}}|^{-\frac{1}{2}} \text{ and } |\mathbf{A}_{\tilde{\boldsymbol{\tau}}}|^{\frac{1}{2}} \geq |D_{\tilde{\boldsymbol{\tau}}}|^{-\frac{1}{2}}.
$$

It follows by (39) and the form of the Inverse-Gamma density that

$$
\begin{aligned}
&\int_{\mathbb{R}_+} \int_{\mathbb{R}_+} \pi\left(\tilde{\sigma}^2 \mid \tilde{\boldsymbol{\beta}}, \boldsymbol{\tau}, \mathbf{Y}\right) \pi\left(\tilde{\boldsymbol{\beta}} \mid \boldsymbol{\tau}, \sigma^2, \mathbf{Y}\right) \pi\left(\boldsymbol{\tau} \mid \boldsymbol{\beta}, \sigma^2, \mathbf{Y}\right) \\
&\times \pi\left(\boldsymbol{\beta} \mid \tilde{\boldsymbol{\tau}}, \sigma^2, \mathbf{Y}\right) \pi\left(\sigma^2 \mid \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\tau}}, \mathbf{Y}\right) \pi\left(\tilde{\boldsymbol{\tau}} \mid \tilde{\boldsymbol{\beta}}, \tilde{\sigma}^2, \mathbf{Y}\right) d\sigma^2 d\tilde{\sigma^2} \\
\geq \quad & D_3 f_1(\boldsymbol{\tau}, \tilde{\boldsymbol{\tau}}) \prod_{j=1}^p \left(\beta_j^2 \tilde{\beta}_j^2\right) \left\{ \frac{1}{\left[\tilde{\Delta} + \tilde{\boldsymbol{\beta}}^T D_{\tilde{\boldsymbol{\tau}}}^{-1} \tilde{\boldsymbol{\beta}}\right]^{\frac{n+p+2\alpha}{2}+p}} \right\}
\end{aligned}
$$

$$\times \left\{ \frac{1}{\left[\Delta_1 + \Delta_{1*} + \tilde{\Delta}_* + \boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1} \boldsymbol{\beta}\right]^{\frac{n+p+2\alpha}{2}+2p}} \right\} \tag{40}$$

where

$$D_3 = 2^{n+4p+2\alpha}\ \Gamma\left(\frac{n+p+2\alpha}{2}+p\right)\ \Gamma\left(\frac{n+p+2\alpha}{2}+2p\right)\ D_2$$

We now establish some inequalities which will help converting the lower bound in (40) into a simpler form. By (33), it follows that

$$\begin{aligned}\tilde{\Delta} + \tilde{\boldsymbol{\beta}}^T D_{\tilde{\boldsymbol{\tau}}}^{-1} \tilde{\boldsymbol{\beta}} &= \tilde{\boldsymbol{\beta}}^T (X^T X + D_{\boldsymbol{\tau}}^{-1}) \tilde{\boldsymbol{\beta}} - 2\mathbf{Y}^T X \tilde{\boldsymbol{\beta}} + \mathbf{Y}^T \mathbf{Y} + 2\xi + \tilde{\boldsymbol{\beta}}^T D_{\tilde{\boldsymbol{\tau}}}^{-1} \tilde{\boldsymbol{\beta}} \\ &\le \tilde{\boldsymbol{\beta}}^T (X^T X + D_{\boldsymbol{\tau}}^{-1}) \tilde{\boldsymbol{\beta}} - 2\mathbf{Y}^T X \tilde{\boldsymbol{\beta}} + \mathbf{Y}^T \mathbf{Y} + \tilde{\Delta}_* \\ &\le \tilde{\boldsymbol{\beta}}^T (X^T X + D_{\boldsymbol{\tau}}^{-1}) \tilde{\boldsymbol{\beta}} - 2\mathbf{Y}^T X \tilde{\boldsymbol{\beta}} + \mathbf{Y}^T \mathbf{Y} + \tilde{\Delta}_* + \Delta_{1*} \\ &\quad + \boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1} \boldsymbol{\beta},\end{aligned}$$

and

$$\begin{aligned}&\Delta_1 + \Delta_{1*} + \tilde{\Delta}_* + \boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1} \boldsymbol{\beta} \\ &= \tilde{\boldsymbol{\beta}}^T (X^T X + D_{\boldsymbol{\tau}}^{-1}) \tilde{\boldsymbol{\beta}} - 2\mathbf{Y}^T X \tilde{\boldsymbol{\beta}} + \mathbf{Y}^T X (X^T X + D_{\boldsymbol{\tau}}^{-1})^{-1} X^T \mathbf{Y} \\ &\quad + \Delta_{1*} + \tilde{\Delta}_* + \boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1} \boldsymbol{\beta} \\ &\le \tilde{\boldsymbol{\beta}}^T (X^T X + D_{\boldsymbol{\tau}}^{-1}) \tilde{\boldsymbol{\beta}} - 2\mathbf{Y}^T X \tilde{\boldsymbol{\beta}} + \mathbf{Y}^T \mathbf{Y} + \tilde{\Delta}_* + \Delta_{1*} + \boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1} \boldsymbol{\beta}.\end{aligned}$$

Also, note that

$$\begin{aligned}&\tilde{\boldsymbol{\beta}}^T (X^T X + D_{\boldsymbol{\tau}}^{-1}) \tilde{\boldsymbol{\beta}} - 2\mathbf{Y}^T X \tilde{\boldsymbol{\beta}} + \mathbf{Y}^T \mathbf{Y} + \tilde{\Delta}_* + \Delta_{1*} + \boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1} \boldsymbol{\beta} \\ &= \tilde{\boldsymbol{\beta}}^T (X^T X + D_{\boldsymbol{\tau}}^{-1}) \tilde{\boldsymbol{\beta}} - 2\mathbf{Y}^T X \tilde{\boldsymbol{\beta}} + \mathbf{Y}^T \mathbf{Y} \\ &\quad + \tilde{\boldsymbol{\beta}}^T (X^T X + D_{\tilde{\boldsymbol{\tau}}}^{-1}) \tilde{\boldsymbol{\beta}} - 2\mathbf{Y}^T X \tilde{\boldsymbol{\beta}} + \mathbf{Y}^T \mathbf{Y} + 2\xi + \Delta_{1*} + \boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1} \boldsymbol{\beta} \\ &= \tilde{\boldsymbol{\beta}}^T (2X^T X + D_{\boldsymbol{\tau}}^{-1} + D_{\tilde{\boldsymbol{\tau}}}^{-1}) \tilde{\boldsymbol{\beta}} - 4\mathbf{Y}^T X \tilde{\boldsymbol{\beta}} + 2\mathbf{Y}^T \mathbf{Y} + 2\xi + \Delta_{1*} + \boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1} \boldsymbol{\beta} \\ &= (\tilde{\boldsymbol{\beta}} - \mu)^T (2X^T X + D_{\boldsymbol{\tau}}^{-1} + D_{\tilde{\boldsymbol{\tau}}}^{-1}) (\tilde{\boldsymbol{\beta}} - \mu) \\ &\quad - 4\mathbf{Y}^T X (2X^T X + D_{\boldsymbol{\tau}}^{-1} + D_{\tilde{\boldsymbol{\tau}}}^{-1})^{-1} X^T \mathbf{Y} + f_2(\boldsymbol{\beta}, \tilde{\boldsymbol{\tau}}) + \boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1} \boldsymbol{\beta} \\ &\le (\tilde{\boldsymbol{\beta}} - \mu)^T (2X^T X + D_{\boldsymbol{\tau}}^{-1} + D_{\tilde{\boldsymbol{\tau}}}^{-1}) (\tilde{\boldsymbol{\beta}} - \mu) + f_2(\boldsymbol{\beta}, \tilde{\boldsymbol{\tau}}) + \boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1} \boldsymbol{\beta} \\ &\le \left[(\tilde{\boldsymbol{\beta}} - \mu)^T (2X^T X + D_{\boldsymbol{\tau}}^{-1} + D_{\tilde{\boldsymbol{\tau}}}^{-1}) (\tilde{\boldsymbol{\beta}} - \mu) + f_2(\boldsymbol{\beta}, \tilde{\boldsymbol{\tau}}) + 1\right] (\boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1} \boldsymbol{\beta} + 1)\end{aligned}$$

where

$$\mu = (2X^T X + D_{\boldsymbol{\tau}}^{-1} + D_{\tilde{\boldsymbol{\tau}}}^{-1})^{-1} X^T \mathbf{Y} \text{ and } f_2(\boldsymbol{\beta}, \tilde{\boldsymbol{\tau}}) = 2\mathbf{Y}^T \mathbf{Y} + 2\xi + \Delta_{1*}.$$

By (40), we get

$$\int_{\mathbb{R}_+} \int_{\mathbb{R}_+} \pi\left(\tilde{\sigma}^2 \mid \tilde{\boldsymbol{\beta}}, \boldsymbol{\tau}, \mathbf{Y}\right) \pi\left(\tilde{\boldsymbol{\beta}} \mid \boldsymbol{\tau}, \sigma^2, \mathbf{Y}\right) \pi\left(\boldsymbol{\tau} \mid \boldsymbol{\beta}, \sigma^2, \mathbf{Y}\right) \times$$

$$
\begin{aligned}
& \pi\left(\boldsymbol{\beta} \mid \tilde{\boldsymbol{\tau}}, \sigma^2, \mathbf{Y}\right) \pi\left(\sigma^2 \mid \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\tau}}, \mathbf{Y}\right) \pi\left(\tilde{\boldsymbol{\tau}} \mid \tilde{\boldsymbol{\beta}}, \tilde{\sigma}^2, \mathbf{Y}\right) d\sigma^2 d\tilde{\sigma^2} \\
\geq \; & \frac{D_2 f_1(\boldsymbol{\tau}, \tilde{\boldsymbol{\tau}})}{(\boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1} \boldsymbol{\beta} + 1)^{n+4p+2\alpha}} \\
& \times \frac{\prod_{j=1}^{p} \left(\beta_j^2 \tilde{\beta}_j^2\right)}{\left[(\tilde{\boldsymbol{\beta}} - \mu)^T (2X^T X + D_{\boldsymbol{\tau}}^{-1} + D_{\tilde{\boldsymbol{\tau}}}^{-1})(\tilde{\boldsymbol{\beta}} - \mu) + f_2(\boldsymbol{\beta}, \tilde{\boldsymbol{\tau}}) + 1\right]^{n+4p+2\alpha}} \\
= \; & \frac{D_2 f_1(\boldsymbol{\tau}, \tilde{\boldsymbol{\tau}})}{(\boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1} \boldsymbol{\beta} + 1)^{n+4p+2\alpha}} \\
& \times \frac{\prod_{j=1}^{p} \left(\beta_j^2 \tilde{\beta}_j^2\right)}{\left[(f_2(\boldsymbol{\beta}, \tilde{\boldsymbol{\tau}}) + 1)\left(1 + \frac{(\tilde{\boldsymbol{\beta}} - \mu)^T (2X^T X + D_{\boldsymbol{\tau}}^{-1} + D_{\tilde{\boldsymbol{\tau}}}^{-1})(\tilde{\boldsymbol{\beta}} - \mu)}{f_2(\boldsymbol{\beta}, \tilde{\boldsymbol{\tau}}) + 1}\right)\right]^{n+4p+2\alpha}} \\
= \; & \frac{D_2 f_1(\boldsymbol{\tau}, \tilde{\boldsymbol{\tau}})}{(f_2(\boldsymbol{\beta}, \tilde{\boldsymbol{\tau}}) + 1)^{n+4p+2\alpha} (\boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1} \boldsymbol{\beta} + 1)^{n+4p+2\alpha}} \\
& \times \frac{\prod_{j=1}^{p} \left(\beta_j^2 \tilde{\beta}_j^2\right)}{\left(1 + \frac{(\tilde{\boldsymbol{\beta}} - \mu)^T (2X^T X + D_{\boldsymbol{\tau}}^{-1} + D_{\tilde{\boldsymbol{\tau}}}^{-1})(\tilde{\boldsymbol{\beta}} - \mu)}{f_2(\boldsymbol{\beta}, \tilde{\boldsymbol{\tau}}) + 1}\right)^{n+4p+2\alpha}} \\
\geq \; & \frac{D_2 f_1(\boldsymbol{\tau}, \tilde{\boldsymbol{\tau}})}{(f_2(\boldsymbol{\beta}, \tilde{\boldsymbol{\tau}}) + 1)^{n+4p+2\alpha} (\boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1} \boldsymbol{\beta} + 1)^{n+4p+2\alpha}} \\
& \times \frac{\prod_{j=1}^{p} \left(\beta_j^2 \tilde{\beta}_j^2\right)}{\left(1 + \frac{(\tilde{\boldsymbol{\beta}} - \mu)^T (2\lambda I_p + D_{\boldsymbol{\tau}}^{-1} + D_{\tilde{\boldsymbol{\tau}}}^{-1})(\tilde{\boldsymbol{\beta}} - \mu)}{f_2(\boldsymbol{\beta}, \tilde{\boldsymbol{\tau}}) + 1}\right)^{n+4p+2\alpha}} \\
\geq \; & \frac{D_2 f_1(\boldsymbol{\tau}, \tilde{\boldsymbol{\tau}})}{(f_2(\boldsymbol{\beta}, \tilde{\boldsymbol{\tau}}) + 1)^{n+4p+2\alpha} (\boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1} \boldsymbol{\beta} + 1)^{n+4p+2\alpha}} \\
& \times \frac{\prod_{j=1}^{p} \left(\beta_j^2 \tilde{\beta}_j^2\right)}{\prod_{j=1}^{p} \left(1 + \frac{\left(\tilde{\beta}_j - \mu_j\right)^2}{\nu_j \epsilon_j^2}\right)^{n+4p+2\alpha}} \\
= \; & \frac{D_2 f_1(\boldsymbol{\tau}, \tilde{\boldsymbol{\tau}}) \prod_{j=1}^{p} \beta_j^2 \prod_{j=2}^{p} \tilde{\beta}_j^2}{(f_2(\boldsymbol{\beta}, \tilde{\boldsymbol{\tau}}) + 1)^{n+4p+2\alpha} \prod_{j=2}^{p} \left(1 + \frac{\left(\tilde{\beta}_j - \mu_j\right)^2}{\nu_j \epsilon_j^2}\right)^{n+4p+2\alpha} (\boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1} \boldsymbol{\beta} + 1)^{n+4p+2\alpha}} \\
& \times \frac{\tilde{\beta}_1^2}{\left(1 + \frac{\left(\tilde{\beta}_1 - \mu_1\right)^2}{\nu_1 \epsilon_1^2}\right)^{\frac{1+\nu_1}{2}}}
\end{aligned} \tag{41}
$$

where $\mu_j = e_j^T \mu,\ \nu_j = 2n + 8p + 4\alpha - 1,\ \epsilon_j = \sqrt{\frac{f_2(\boldsymbol{\beta},\tilde{\boldsymbol{\tau}})+1}{(2\lambda+\frac{1}{\tau_j}+\frac{1}{\tilde{\tau}_j})\nu_j}}$ and λ is the greatest eigenvalue of matrix $X^T X$. By Proposition A4, we have

$$\int_{\mathbb{R}} \frac{\tilde{\beta}_1^2}{\left(1+\frac{(\tilde{\boldsymbol{\beta}}_1-\mu_1)^2}{\nu_1\epsilon_1^2}\right)^{\frac{1+\nu_1}{2}}} d\tilde{\beta}_1 \geq f_3(\boldsymbol{\beta},\tilde{\boldsymbol{\tau}})\left(2\lambda+\frac{1}{\tau_1}+\frac{1}{\tilde{\tau}_1}\right)^{-\frac{3}{2}} \tag{42}$$

Hence, it follows from (41) and (42) that

$$\begin{aligned}
&\int_{\mathbb{R}_+}\int_{\mathbb{R}_+}\int_{\mathbb{R}} \pi\left(\tilde{\sigma}^2 \mid \tilde{\boldsymbol{\beta}},\boldsymbol{\tau},\mathbf{Y}\right)\pi\left(\tilde{\boldsymbol{\beta}} \mid \boldsymbol{\tau},\sigma^2,\mathbf{Y}\right)\pi\left(\boldsymbol{\tau} \mid \boldsymbol{\beta},\sigma^2,\mathbf{Y}\right)\times \\
&\pi\left(\boldsymbol{\beta} \mid \tilde{\boldsymbol{\tau}},\sigma^2,\mathbf{Y}\right)\pi\left(\sigma^2 \mid \tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\tau}},\mathbf{Y}\right)\pi\left(\tilde{\boldsymbol{\tau}} \mid \tilde{\boldsymbol{\beta}},\tilde{\sigma}^2,\mathbf{Y}\right)d\sigma^2 d\tilde{\sigma^2} d\tilde{\beta}_1 \\
&\geq \frac{D_2\ f_1(\boldsymbol{\tau},\tilde{\boldsymbol{\tau}})\prod_{j=1}^p \beta_j^2 \prod_{j=2}^p \tilde{\beta}_j^2}{(f_2(\boldsymbol{\beta},\tilde{\boldsymbol{\tau}})+1)^{n+4p+2\alpha}\prod_{j=2}^p\left(1+\frac{(\tilde{\beta}_j-\mu_j)^2}{\nu_j\epsilon_j^2}\right)^{n+4p+2\alpha}(\boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1}\boldsymbol{\beta}+1)^{n+4p+2\alpha}} \\
&\quad\times\int_{\mathbb{R}} \frac{\tilde{\beta}_1^2}{\left(1+\frac{(\tilde{\boldsymbol{\beta}}_1-\mu_1)^2}{\nu_1\epsilon_1^2}\right)^{\frac{1+\nu_1}{2}}}\ d\tilde{\beta}_1 \\
&\geq \frac{D_2\ f_1(\boldsymbol{\tau},\tilde{\boldsymbol{\tau}})\prod_{j=1}^p \beta_j^2 \prod_{j=2}^p \tilde{\beta}_j^2\ f_3(\boldsymbol{\beta},\tilde{\boldsymbol{\tau}})\left(2\lambda+\frac{1}{\tau_1}+\frac{1}{\tilde{\tau}_1}\right)^{-\frac{3}{2}}}{(f_2(\boldsymbol{\beta},\tilde{\boldsymbol{\tau}})+1)^{n+4p+2\alpha}\prod_{j=2}^p\left(1+\frac{(\tilde{\beta}_j-\mu_j)^2}{\nu_j\epsilon_j^2}\right)^{n+4p+2\alpha}(\boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1}\boldsymbol{\beta}+1)^{n+4p+2\alpha}} \\
&= \frac{D_2\left\{\prod_{j=2}^p \tau_j^{(a-\frac{1}{2})-1}\exp\left(-b\tau_j\right)\right\}\left\{\prod_{j=1}^p \tilde{\tau}_j^{(a-\frac{1}{2})-1}\exp\left(-b\tilde{\tau}_j\right)\right\}}{(f_2(\boldsymbol{\beta},\tilde{\boldsymbol{\tau}})+1)^{n+4p+2\alpha}\prod_{j=2}^p\left(1+\frac{(\tilde{\beta}_j-\mu_j)^2}{\nu_j\epsilon_j^2}\right)^{n+4p+2\alpha}} \\
&\quad\times\ \prod_{j=1}^p \beta_j^2 \prod_{j=2}^p \tilde{\beta}_j^2\ f_3(\boldsymbol{\beta},\tilde{\boldsymbol{\tau}})\frac{\tau_1^{(a-\frac{1}{2})-1}\exp\left(-b\tau_1\right)\left(2\lambda+\frac{1}{\tau_1}+\frac{1}{\tilde{\tau}_1}\right)^{-\frac{3}{2}}}{\left(\sum_{j=2}^p\frac{\beta_j^2}{\tau_j}+\frac{\beta_1^2}{\tau_1}+1\right)^{n+4p+2\alpha}}
\end{aligned} \tag{43}$$

From (43), we obtain

$$\begin{aligned}
&\int_{\mathbb{R}_+}\int_{\mathbb{R}_+}\int_{\mathbb{R}}\int_{\mathbb{R}_+} \pi\left(\tilde{\sigma}^2 \mid \tilde{\boldsymbol{\beta}},\boldsymbol{\tau},\mathbf{Y}\right)\pi\left(\tilde{\boldsymbol{\beta}} \mid \boldsymbol{\tau},\sigma^2,\mathbf{Y}\right)\pi\left(\boldsymbol{\tau} \mid \boldsymbol{\beta},\sigma^2,\mathbf{Y}\right)\times \\
&\pi\left(\boldsymbol{\beta} \mid \tilde{\boldsymbol{\tau}},\sigma^2,\mathbf{Y}\right)\pi\left(\sigma^2 \mid \tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\tau}},\mathbf{Y}\right)\pi\left(\tilde{\boldsymbol{\tau}} \mid \tilde{\boldsymbol{\beta}},\tilde{\sigma}^2,\mathbf{Y}\right)d\sigma^2 d\tilde{\sigma^2} d\tilde{\beta}_1 d\tau_1
\end{aligned}$$

$$
\begin{aligned}
&\geq \frac{D_2 \left\{ \prod_{j=2}^{p} \tau_j^{\left(a-\frac{1}{2}\right)-1} \exp\left(-b\tau_j\right) \right\} \left\{ \prod_{j=1}^{p} \tilde{\tau}_j^{\left(a-\frac{1}{2}\right)-1} \exp\left(-b\tilde{\tau}_j\right) \right\}}{\left(f_2(\boldsymbol{\beta}, \tilde{\boldsymbol{\tau}})+1\right)^{n+4p+2\alpha} \prod_{j=2}^{p} \left(1+\frac{\left(\tilde{\beta}_j-\mu_j\right)^2}{\nu_j \epsilon_j^2}\right)^{n+4p+2\alpha}} \\
&\quad \times \prod_{j=1}^{p} \beta_j^2 \prod_{j=2}^{p} \tilde{\beta}_j^2 \ f_3(\boldsymbol{\beta}, \tilde{\boldsymbol{\tau}}) \int_{\mathbb{R}_+} \frac{\tau_1^{\left(a-\frac{1}{2}\right)-1} \exp\left(-b\tau_1\right) \left(2\lambda+\frac{1}{\tau_1}+\frac{1}{\tilde{\tau}_1}\right)^{-\frac{3}{2}}}{\left(\sum_{j=2}^{p} \frac{\beta_j^2}{\tau_j}+\frac{\beta_1^2}{\tau_1}+1\right)^{n+4p+2\alpha}} \, d\tau_1 \\
&= \infty
\end{aligned}
$$

for every $\left(\boldsymbol{\beta}, \tilde{\boldsymbol{\tau}}, (\tilde{\beta}_2, ..., \tilde{\beta}_p)^T, (\tau_2, ..., \tau_p)^T\right) \in \mathbb{R}^p \times \mathbb{R}_+^p \times \mathbb{R}^{p-1} \times \mathbb{R}_+^{p-1}$. The above integral diverges because $\int_{\mathbb{R}_+} \tau_1^{\left(a-\frac{1}{2}\right)-1} \exp\left(-b\tau_1\right) d\tau_1 = \infty$ if $a \leq \frac{1}{2}$ and

$$
\begin{aligned}
&\lim_{\tau_1 \to \infty} \frac{\tau_1^{\left(a-\frac{1}{2}\right)-1} \exp\left(-b\tau_1\right) \left(2\lambda+\frac{1}{\tau_1}+\frac{1}{\tilde{\tau}_1}\right)^{-\frac{3}{2}}}{\left(\sum_{j=2}^{p} \frac{\beta_j^2}{\tau_j}+\frac{\beta_1^2}{\tau_1}+1\right)^{n+4p+2\alpha}} \tau_1^{1-\left(a-\frac{1}{2}\right)} \exp\left(b\tau_1\right) \\
&= \lim_{\tau_1 \to \infty} \left(2\lambda+\frac{1}{\tau_1}+\frac{1}{\tilde{\tau}_1}\right)^{-\frac{3}{2}} \left(\sum_{j=2}^{p} \frac{\beta_j^2}{\tau_j}+\frac{\beta_1^2}{\tau_1}+1\right)^{-n-4p-2\alpha} \\
&= \left(2\lambda+\frac{1}{\tilde{\tau}_1}\right)^{-\frac{3}{2}} \left(\sum_{j=2}^{p} \frac{\beta_j^2}{\tau_j}+1\right)^{-n-4p-2\alpha} \in (0, \infty) \qquad \square
\end{aligned}
$$

5. Construction of the Haar PX-DA sandwich Markov chain

The two-block Markov chain Φ can be interpreted as a Data Augmentation (DA) algorithm, with $(\boldsymbol{\beta}, \sigma^2)$ as the parameter block of interest, and $\boldsymbol{\tau}$ as the augmented block. The DA algorithm can suffer from slow convergence (just like the EM algorithm, its analogous version in likelihood maximization). The sandwich algorithm, introduced in [17, 9], aims to improve the convergence and efficiency of the DA algorithm by adding an inexpensive extra step in between the two conditional draws of the DA algorithm. In fact, there are many DA chains (see [17, 19, 18, 10, 22], for example) where sandwich chains have been constructed and shown to be significantly more efficient with roughly the same computational effort per iteration. In this section, we will focus on deriving the Haar PX-DA sandwich algorithm in the normal-gamma prior setting. The Haar PX-DA algorithm has been shown in [9] to be the best among a class of sandwich algorithms in terms of efficiency and operator norm.

A key ingredient in constructing the Haar PX-DA algorithm is a unimodular group which acts on the augmented variable space ($\mathbb{R}^p$ in our case). We consider the multiplicative group G of positive real numbers, which acts on an element of $\mathbb{R}^p$ through scalar multiplication. In particular, if $g \in G$ and $\boldsymbol{\tau} \in \mathbb{R}^p$, then the result of the action of g on $\boldsymbol{\tau}$ is given by $g\boldsymbol{\tau} = (g\tau_1, g\tau_2, \cdots, g\tau_p)$. Another choice that we need to make is the choice of the multiplier function $\chi : G \to \mathbb{R}_+$, which satisfies

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

for any pair $g_1, g_2 \in G$, and

$$\chi(g) \int_{\mathbb{R}^p} \phi(g\boldsymbol{\tau}) d\boldsymbol{\tau} = \int_{\mathbb{R}^p} \phi(\boldsymbol{\tau}) d\boldsymbol{\tau}$$

for any $g \in G$ and any function $\phi : \mathbb{R}^p \to \mathbb{R}$. In this setting, the function $\chi(g) = g^p$ serves as a valid multiplier function. Also, the unimodular group G has a Haar measure $\mathcal{H}(dg) = \frac{dg}{g}$. With these ingredients in hand, we define the density f_G on G (with respect to the Haar measure) by

$$f_{\mathcal{G}}(g) = \frac{\pi(g\boldsymbol{\tau} \mid \mathbf{Y})\chi(g)}{m(\boldsymbol{\tau})} \mathcal{H}(dg),$$

where $m(\boldsymbol{\tau}) = \int_{\mathcal{G}} \pi(g\boldsymbol{\tau})\chi(g)\mathcal{H}(dg)$ is the normalizing constant. From (6), it follows that

$$\pi(\boldsymbol{\tau} \mid \mathbf{Y}) \propto \frac{\prod_{j=1}^p \tau_j^{a-\frac{1}{2}-1} \exp\left(-b\tau_j\right)}{\left\{\mathbf{y}^T\mathbf{y} - \mathbf{y}^T X(X^T X + D_{\boldsymbol{\tau}}^{-1})^{-1} X^T \mathbf{y} + 2\xi\right\}^{\frac{n}{2}+\alpha} |X^T X + D_{\boldsymbol{\tau}}^{-1}|^{\frac{1}{2}}}$$

and

$$f_{\mathcal{G}}(g) \propto \frac{g^{p(a-\frac{1}{2})-1} \exp\left(-g(\sum_{j=1}^p b\tau_j)\right)}{\left\{\mathbf{y}^T\mathbf{y} - \mathbf{y}^T X(X^T X + \frac{1}{g} D_{\boldsymbol{\tau}}^{-1})^{-1} X^T \mathbf{y} + 2\xi\right\}^{\frac{n}{2}+\alpha} |X^T X + \frac{1}{g} D_{\boldsymbol{\tau}}^{-1}|^{\frac{1}{2}}}. \tag{44}$$

Using $f_{\mathcal{G}}$, we can now define the Haar PX-DA 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 follow.

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

The lemma below, regarding spectral properties of the Haar PX-DA chain, follows by combining Theorem 1 and results from [14].

Lemma 5.1. *The operator corresponding to the Haar PX-DA Markov chain* Φ^* *is trace class. Also, if* $\{\lambda_i^*\}_{i=1}^{\infty}$ *and* $\{\lambda_i\}_{i=1}^{\infty}$ *are the ordered eigenvalues corresponding to* Φ^* *and* Φ *respectively, then* $\lambda_i^* \leq \lambda_i$ *for* $i \geq 1$ *with a strict inequality holding for at least one* i*.*

Note that $f_{\mathcal{G}}$ is not a standard density. Samples from this univariate density can be generated using a rejection sampling algorithm with respect to a Gamma proposal density with shape parameter pa and rate parameter $b\sum_{j=1}^{p}\tau_j$. The ratio of the unnormalized density on the right hand side of (44) and this Gamma density can be shown to be bounded by

$$\left(\prod_{j=1}^{p}\sqrt{\tau_j}\right)(2\xi)^{-(n/2+\alpha)}\Gamma(pa)\left(b\sum_{j=1}^{p}\tau_j\right)^{-pa}.$$

We found that for small p, this rejection sampler works fairly well, and the sandwich step does not add much to the overall complexity of the DA chain. However, when p is large, we found that the rejection sampler becomes very inefficient. We tried other approaches such as approximate discretization or Metropolis-Hastings, but these approaches also turn out to be computationally inefficient is the large p setting, i.e., the extra computations for the sandwich step are too burdensome in comparison to any potential improvement in the speed of convergence.

6. Examples

In this section, we consider simulated and real data examples to compare the performance of the three block, two block and Haar PX-DA sandwich chains.

6.1. Simulation I: small p

In this section, we consider two simulation settings with small values of p. We consider a setting with $n = 10 < p = 15$ for the first simulation, and $n = 15 > p = 10$ for the second simulation. For both cases,the elements of the design matrix X and response y were chosen by generating i.i.d. $\mathcal{N}(0,1)$ random variables. We fit the Normal-Gamma model in (1) with hyper parameters $a = 0.75, b = 2, \xi = 100, \alpha = 0$. 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 Table 1 for the first simulation, and in Table 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.

Table 1
First ten autocorrelations for simulated data with $n < p$

Lag	1	2	3	4	5	6	7	8	9	10
Three block	0.502	0.269	0.151	0.079	0.04	0.026	0.01	0.005	0.005	-0.01
Two Block	0.164	0.072	0.055	0.022	0.005	0.009	0.016	0.003	0.017	0.017
Sandwich	0.018	0.001	-0.008	-0.036	-0.011	-0.015	-0.007	0.011	0.017	0.005

Table 2
First ten autocorrelations for simulated data with $n > p$

Lag	1	2	3	4	5	6	7	8	9	10
Three block	0.541	0.35	0.231	0.156	0.103	0.064	0.042	0.023	0.02	0.014
Two Block	0.116	0.054	0.03	0.041	0.034	0.013	0.004	-0.002	0.018	-0.004
Sandwich	-0.02	-0.024	-0.008	0.051	-0.008	0.023	-0.004	-0.001	-0.005	-0.026

6.2. Simulation II: large p

In this section we consider a more extensive simulation setting with large p. In particular, we choose $p = 500$, $n = 100$, consider five different values of a (namely $0.25, 0.5, 0.75, 1, 1.25$), and consider five different correlations (namely $0.1, 0.3, 0.5, 0.7, 0.9$) between the entries of the design matrix X. For a given a and correlation value r, the np elements of the design matrix X are drawn as standard normal random variables with all pairwise correlations equal to r. The response vector $\mathbf{Y}$ was generated as $X\boldsymbol{\beta}_* + \boldsymbol{\epsilon}$, where the first one-fifth elements of $\boldsymbol{\beta}_*$ are independent t_2 random variables, and all other elements are set to zero, and $\boldsymbol{\epsilon}$ has independent t_4 elements. We set $\alpha = 0.1, b = 2$, and $\xi = 10$. The original and fast Bayesian lasso chains were run for each a and r setting (there are 25 such settings, giving a total of 50 Markov chains) [1]. For each Markov chain, we have a burn-in of $50,000$ iterations. For this simulation, based on a referee's suggestion, we use a recently introduced measure of Multivariate Essential Sample Size (MESS) in [31] to measure performance of the Markov chain. MESS uses multivariate batch means to construct an estimator of the effective sample size. Hence, if we have N iterations of the Markov chain after burn-in, then we divide the observations into $a_n = \sqrt{N}$ batches of size $b_n = \sqrt{N}$ each. Determinants of the covariance matrix $\hat{\Lambda}$ of the a_n batch-mean vectors, and the covariance matrix $\hat{\Sigma}$ of all the N Markov chain iterates are then used to construct the MESS estimator. In order to ensure the positivity of the determinant of $\hat{\Lambda}$ we need $\sqrt{N} \geq p + 1$ (the total number of parameters in $(\boldsymbol{\beta}, \sigma^2)$ is 501). Hence $N > 250000$ is needed just for the estimator to be well-defined. We found that even after $N = 500000$ or even $N = 10^6$ iterations the ESS estimator did not stabilize, and was sometimes giving unacceptable values (such

[1]Since the sandwich algorithm is computationally too burdensome for this setting, we do not use it for comparison

as a value greater than the number of iterates N). For computational efficiency, we made the following adjustment. We run the Markov chain for $N = 500000$ iterations after burn-in, then we randomly choose 10 elements of $\boldsymbol{\beta}$ along with σ^2 as a representative set of parameters and compute the MESS based on these 21 parameters (instead of 501 parameters). The corresponding MESS values are reported in Table 3 (for the original chain) and Table 4 (for the fast chain).

Table 3
Essential sample size for the original chain with 500000 *iterations for various values of a and design matrix correlation r*

	$a = 0.25$	$a = 0.5$	$a = 0.75$	$a = 1$	$a = 1.25$
$r = 0.1$	186168	271031	303710	328622	346134
$r = 0.3$	193194	281160	305893	319902	339629
$r = 0.5$	194229	271408	306098	323637	346594
$r = 0.7$	201308	275623	313876	326339	339132
$r = 0.9$	223571	287344	320764	331529	345776

Table 4
Essential sample size for the fast chain with 500000 *iterations for various values of a and design matrix correlation r*

	$a = 0.25$	$a = 0.5$	$a = 0.75$	$a = 1$	$a = 1.25$
$r = 0.1$	220021	326830	364800	384333	420936
$r = 0.3$	219422	330641	362150	381485	413435
$r = 0.5$	226298	339746	371326	401799	410920
$r = 0.7$	234131	325674	364228	404453	411481
$r = 0.9$	267711	339578	381108	402151	411189

6.3. Real data example

We consider the wheat data set from Perez and de los Campos [24], which is available in the R package BGLR. The data was obtained from numerous international trials for $n = 599$ wheat lines across a wide variety of wheat-producing environments. 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 $p = 20$ binary variables containing genotypic information as the predictors. We fit the Normal-Gamma model in (1) with $a = 0.75, b = 0.2, \xi = 1, \alpha = 0$ and compute autocorrelations for the function $(\mathbf{Y} - X\boldsymbol{\beta})^T (\mathbf{Y} - X\boldsymbol{\beta}) + \sigma^2$ for the three block, two block and Haar PX-DA sandwich chains. The results are shown in Table 5. As in the simulated data examples, the two-block chain has lower autocorrelations than the three-block chain, and the Haar PX-DA sandwich chain is the most efficient among all three Markov chains.

Table 5
First ten autocorrelations with wheat data

Lag	1	2	3	4	5	6	7	8	9	10
Three block	0.458	0.225	0.119	0.043	0.021	0.019	0.017	0.011	-0.007	-0.007
Two block	0.08	0.016	0.005	0.006	0.007	0.002	0.002	-0.022	-0.01	-0.013
Sandwich	0.054	0.016	0.001	0.034	0.007	0.031	-0.019	-0.044	-0.01	-0.016

6.4. Discussion of numerical results

For both the simulated and real data settings, the two-block chain clearly has a significantly better performance than the three-block chain. For example, in all the settings the Lag 1 autocorrelation drops by 80% or more in the small p setting, and the essential sample size increases by 20% or more in the large p setting, when we compare the three-block and the two-block chains. These findings support the theoretical results (Theorem 1 and Theorem 2) in the paper. Since the two-block chain and the three-block chain require roughly the same computational effort, our theoretical and experimental results, support the overall conclusion that a practitioner should prefer the two-block chain over the three-block chain.

7. Estimation of the largest eigenvalue of the two-block chain

In recent work [27], the authors provide a Monte Carlo based algorithm to estimate the largest eigenvalue of a trace-class Markov operator. This algorithm provides upper and lower bounds for the largest eigenvalue. Both these bounds are shown to converge to the largest eigenvalue. We briefly describe this algorithm in the context of the Normal-Gamma two block Gibbs sampler.

For $j \geq 1$, let $k^j(\cdot, \cdot)$ denote the j-step transition density of the two-block Markov chain, and let

$$s_j = \int_{\mathbb{R}^p} \int_{\mathbb{R}} k^j((\boldsymbol{\beta}, \sigma^2), (\boldsymbol{\beta}, \sigma^2)) d\boldsymbol{\beta} d\sigma^2.$$

Qin et al. [27] show that the trace-class property of k implies that all s_j's are finite. Furthermore, if λ_1 denotes the largest eigenvalue of k, then

$$u_k := (s_k - 1)^{1/k} \downarrow \lambda_1,$$

and

$$l_k := \frac{s_k - 1}{s_{k-1} - 1} \uparrow \lambda_1$$

as $k \to \infty$ ([27, Proposition 1]).

Let ω be a density on $\mathbb{R}_+^p$ defined by

$$\omega(\boldsymbol{\tau}) \propto \prod_{j=1}^{p} \tau_j^{(a-\frac{1}{2})-1} \exp\left(-\lambda \tau_j\right).$$

Then s_j can be alternatively expressed as

$$s_j = E\left(\frac{\pi(\boldsymbol{\tau}_j^* \mid \boldsymbol{\beta}_j^*, (\sigma^2)_j^*, \mathbf{Y})}{\omega(\boldsymbol{\tau}_j^*)}\right),$$

where $\boldsymbol{\tau}_j^* \sim \omega$, and $\boldsymbol{\beta}_j^*, (\sigma^2)_j^* \,|\, \boldsymbol{\tau}_j^* \sim \int_{\mathbb{R}^p \times \mathbb{R}_+} k^{j-1}((\boldsymbol{\beta}', \sigma^{2\prime}), \cdot) \pi(\boldsymbol{\beta}', \sigma^{2\prime} \,|\, \boldsymbol{\tau}_j^*) d\boldsymbol{\beta}' d\sigma^{2\prime}$. This interpretation of s_j allows us to use classical Monte Carlo to estimate s_j

by generating i.i.d. copies of $(\boldsymbol{\beta}_j^*, (\sigma^2)_j^*, \boldsymbol{\tau}_j^*)$ and computing the average of the corresponding $\frac{\pi(\boldsymbol{\beta}_j^*, (\sigma^2)_j^* | \boldsymbol{\tau}_j^*, \mathbf{Y})}{\omega(\boldsymbol{\beta}_j^*, (\sigma^2)_j^*)}$ values (denote this average by $\hat{s}_j$). Of course, the key to successful implementation of any Monte Carlo method is a finite variance, and hence we need to check that

$$Var\left(\frac{\pi(\boldsymbol{\beta}_j^*, (\sigma^2)_j^* \mid \boldsymbol{\tau}_j^*, \mathbf{Y})}{\omega(\boldsymbol{\beta}_j^*, (\sigma^2)_j^*)}\right) < \infty.$$

The following theorem, combined with [27, Theorem 2] guarantees the finiteness of the variance corresponding to $\hat{s}_j$ for *every* $j \geq 1$.

Theorem 3. *For all values of* n *and* p*, and* $a > \frac{1}{2}$*, the conditional densities for the Normal-Gamma chain satisfy*

$$\int_{\mathbb{R}^p}\int_{\mathbb{R}^+}\int_{\mathbb{R}_+^p} \frac{\pi^3(\boldsymbol{\tau} \mid \boldsymbol{\beta}, \sigma^2, \mathbf{Y})\pi(\boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{\tau}, \mathbf{Y})}{\omega^2(\boldsymbol{\tau})} d\boldsymbol{\beta} d\boldsymbol{\tau} d\sigma^2 < \infty.$$

Proof. Note that

$$\begin{aligned}
\frac{\pi\left(\boldsymbol{\tau}|\boldsymbol{\beta}, \sigma^2, \mathbf{Y}\right)}{\omega(\boldsymbol{\tau})} =& \prod_{j=1}^{p} \frac{\frac{\left(2b\sigma^2\right)^{\frac{a-\frac{1}{2}}{2}}}{2|\beta_j|^{a-\frac{1}{2}} K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)} \tau_j^{(a-\frac{1}{2})-1} \exp\left(-\frac{1}{2}\left\{2b\tau_j + \frac{\beta_j^2}{\sigma^2}\frac{1}{\tau_j}\right\}\right)}{\lambda^{a-\frac{1}{2}}(\Gamma(a-\frac{1}{2}))^{-1}\tau_j^{(a-\frac{1}{2})-1}\exp\left(-\lambda\tau_j\right)} \\
=& C_1 \prod_{j=1}^{p} \exp\left(\lambda\tau_j\right)\left(\frac{\sigma}{|\beta_j|}\right)^{a-\frac{1}{2}} \left(K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)\right)^{-1} \\
& \times \exp\left(-\frac{1}{2}\left\{2b\tau_j + \frac{\beta_j^2}{\sigma^2}\frac{1}{\tau_j}\right\}\right),
\end{aligned} \tag{45}$$

where $C_1 = \left(\frac{(2b)^{\frac{a-\frac{1}{2}}{2}}\Gamma(a=\frac{1}{2})}{2\lambda^{a-\frac{1}{2}}}\right)^p$. When $a > \frac{1}{2}$, by Proposition A.2 in [21, Page 639], there exists $\epsilon > 0$ and $C_\epsilon > 0$ such that

$$\begin{aligned}
\left(K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)\right)^{-1} &\leq C_\epsilon \left(K_{\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)\right)^{-1} \left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)^{(a-\frac{1}{2})-\frac{1}{2}} \\
&= C_2\left(\frac{|\beta_j|}{\sigma}\right)^{a-\frac{1}{2}} \exp\left(\sqrt{2b}\frac{|\beta_j|}{\sigma}\right),
\end{aligned} \tag{46}$$

for $0 < \sqrt{2b\frac{\beta_j^2}{\sigma^2}} < \epsilon$. Here $C_2 = \sqrt{\frac{2}{\pi}}C_\epsilon$. When $\sqrt{2b\frac{\beta_j^2}{\sigma^2}} \geq \epsilon$, by Theorem 7 in [28], for all $x > 0$, we get

$$\begin{aligned}
\left(K_{a-\frac{1}{2}}(x)\right)^{-1} &\leq \frac{a + \sqrt{a^2 + x^2}}{x}\left(K_{a+\frac{1}{2}}(x)\right)^{-1} \\
&\leq \frac{2a + x}{x}\left(K_{a+\frac{1}{2}}(x)\right)^{-1}.
\end{aligned}$$

Next, using the fact that if $x > 0$, then $\nu \to K_\nu(x)$ is an increasing function for $\nu > 0$ (again, see [16, Page 266]), we get

$$\begin{aligned}\left(K_{a-\frac{1}{2}}(x))\right)^{-1} &\leq \frac{2a+x}{x}\left(K_{a+\frac{1}{2}}(x)\right)^{-1}\\ &\leq \frac{2a+x}{x}\left(K_{\frac{1}{2}}(x)\right)^{-1}\\ &= \sqrt{\frac{2}{\pi}}\frac{2a+x}{\sqrt{x}}e^x.\end{aligned}$$

Thus when $x \geq \epsilon > 0$, we get

$$\left(K_{a-\frac{1}{2}}(x))\right)^{-1} \leq C_3(1+x^{\frac{1}{2}})e^x,$$

Here $C_3 = \max\{\sqrt{\frac{2}{\pi}}\frac{2a}{\sqrt{\epsilon}}, \sqrt{\frac{2}{\pi}}\}$. Hence for $a > \frac{1}{2}$ and $\sqrt{2b\frac{\beta_j^2}{\sigma^2}} \geq \epsilon$,

$$\left(K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)\right)^{-1} \leq C_4\left(1+\left(\frac{|\beta_j|}{\sigma}\right)^{\frac{1}{2}}\right)\exp\left(\sqrt{2b}\frac{|\beta_j|}{\sigma}\right), \tag{47}$$

where $C_4 = \max\{1, (2b)^{\frac{1}{4}}\}C_3$. It follows from (45), (46), (47) that

$$\begin{aligned}\frac{\pi\left(\boldsymbol{\tau}|\boldsymbol{\beta},\sigma^2,\mathbf{Y}\right)}{\omega(\boldsymbol{\tau})} \leq & C_5\prod_{j=1}^p \exp\left(\lambda\tau_j\right)\exp\left(-\frac{1}{2}\left(2b\tau_j - 2\sqrt{2b}\frac{|\beta_j|}{\sigma} + \frac{\beta_j^2}{\sigma^2}\frac{1}{\tau_j}\right)\right)\\ &\times\left\{\mathbb{1}_{\left\{\frac{|\beta_j|}{\sigma}<\frac{\epsilon}{\sqrt{2b}}\right\}} + \left(\left(\frac{|\beta_j|}{\sigma}\right)^{\frac{1}{2}-a} + \left(\frac{|\beta_j|}{\sigma}\right)^{1-a}\right)\mathbb{1}_{\left\{\frac{|\beta_j|}{\sigma}\geq\frac{\epsilon}{\sqrt{2b}}\right\}}\right\}\\ \leq & C_6\prod_{j=1}^p \exp\left(\lambda\tau_j\right)\left(1+\left(\frac{|\beta_j|}{\sigma}\right)^{\frac{1}{2}}\right),\end{aligned} \tag{48}$$

where $C_5 = \max\{C_2, C_4\}$ and $C_6 = \max\{1, \left(\frac{\epsilon}{\sqrt{2b}}\right)^{\frac{1}{2}-a}\}C_5$. It follows from (48) that we only need to prove

$$\begin{aligned}\iiint\limits_{\mathbb{R}_+^p\times\mathbb{R}^p\times\mathbb{R}_+} & \prod_{j=1}^p \exp\left(2\lambda\tau_j\right)\left(1+2\left(\frac{|\beta_j|}{\sigma}\right)^{\frac{1}{2}}+\frac{|\beta_j|}{\sigma}\right)\\ &\times\pi\left(\boldsymbol{\tau}|\boldsymbol{\beta},\sigma^2,\mathbf{Y}\right)\pi\left(\boldsymbol{\beta},\sigma^2|\boldsymbol{\tau},\mathbf{Y}\right)d\boldsymbol{\tau}d\boldsymbol{\beta}d\sigma^2 < \infty.\end{aligned} \tag{49}$$

Note that

$$\iiint\limits_{\mathbb{R}_+^p\times\mathbb{R}^p\times\mathbb{R}_+} \pi\left(\boldsymbol{\tau}|\boldsymbol{\beta},\sigma^2,\mathbf{Y}\right)\pi\left(\boldsymbol{\beta},\sigma^2|\boldsymbol{\tau},\mathbf{Y}\right)d\boldsymbol{\tau}d\boldsymbol{\beta}d\sigma^2$$

$$= \iint\limits_{\mathbb{R}^p \times \mathbb{R}_+} k\left(\left(\boldsymbol{\beta}, \sigma^2\right), \left(\boldsymbol{\beta}, \sigma^2\right)\right) \, d\boldsymbol{\beta} \, d\sigma^2,$$

and the finiteness of this integral has been proved in Theorem 1. Hence, we can prove (49) by closely following the proof of Theorem 1. Using $\lambda = \frac{b}{8}$, we can reach a result similar to (11) except that instead of c_j, we have

$$c_j' = \left(1 + 2\left(\frac{|\beta_j|}{\sigma}\right)^{\frac{1}{2}} + \frac{|\beta_j|}{\sigma}\right) \frac{\left(2b\sigma^2\right)^{\frac{a-\frac{1}{2}}{2}}}{2\,|\beta_j|^{a-\frac{1}{2}} K_{a-\frac{1}{2}}\left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}}\right)} \tau_j^{\left(a-\frac{1}{2}\right)-1}$$
$$\times \exp\left(-\frac{1}{2}\left\{\frac{3b}{2}\tau_j + \frac{2\beta_j^2}{\sigma^2}\frac{1}{\tau_j}\right\}\right). \tag{50}$$

With our c_j', the bound of (49) can be established with similar arguments to those following (11). $\square$

We now illustrate the application of the method in [27] to the two-block Normal-Gamma chain. We use $p = 10$ and $n = 10$ and generate datasets using the same method as in Section 6.2 for $a = 0.55, 0.65, 0.75, 0.85$ and the common design matrix correlation $r = 0.1, 0.3, 0.5, 0.7, 0.9$. We used 10^7 Monte Carlo iterations to estimate $\hat{s}_5$ values for each setting. The $\hat{s}_5$ value was then used to compute the lower and upper bounds for the largest eigenvalue. The results are provided in Table 6, and show that the performance of the chain generally improves as a increases.

TABLE 6
Lower and upper bound estimates for the largest eigenvalue of the two-block chain in various settings using the method of [27]

	$a = 0.55$	$a = 0.65$	$a = 0.75$	$a = 0.85$
$r = 0.1$	(0.35, 0.57)	(0.32, 0.42)	(0.26, 0.37)	(0.14, 0.24)
$r = 0.3$	(0.24, 0.50)	(0.28, 0.42)	(0.30, 0.35)	(0.2, 0.29)
$r = 0.5$	(0.30, 0.53)	(0.38, 0.44)	(0.23, 0.34)	(0.24, 0.32)
$r = 0.7$	(0.48, 0.58)	(0.36, 0.45)	(0.31, 0.37)	(0.24, 0.32)
$r = 0.9$	(0.44, 0.61)	(0.28, 0.44)	(0.37, 0.40)	(0.21, 0.34)

We found that for larger p (such as $p = 100$ or $p = 500$), 10^7 or even 10^8 Monte Carlo iterations are not enough. The estimates of s_k are much less than 1, whereas we know that $s_k > 1$. The problem, which is common to importance-sampling based procedures, and is enhanced as p increases, is that most of the support for $\pi(\cdot \mid \boldsymbol{\beta}, \sigma^2)$ lies in a region which is a rare event for ω. While there are methods available to practically address this issue, we were not able to prove finiteness of the Monte Carlo variance for these revised approaches. Hence, applying the method in [27] for the Normal-Gamma chain in large p settings still remains a challenge.

Appendix

Propositions A1, A2 and A3 are rather trivial, and hence we state them without proof.

Proposition A1. *Let* $x \sim N(\mu, \sigma^2)$*, then* $\int_0^c \frac{1}{x} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \infty$ *for any positive constant* $c > 0$.

Proposition A2. *Let* $x \sim N(\mu, \sigma^2)$*, then* $\int_0^{c_1} \frac{1}{x(c_2 - \ln x)} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \infty$ *for any positive constant* c_1 *and* $c_2 > \ln c_1$.

Proposition A3. *Suppose the trandom variable* U *has a t-distribution with scale parameter* κ *, location parameter* ϑ *and degrees of freedom* ν*. Then for* $\nu > 2$,

$$E(U^2) \geq \frac{k^2 \nu}{\nu - 2}$$

.

Proposition A4. *Let* $\mu_1 = e_1^T \mu$*, i.e. the first component of* μ*,* $\nu_1 = 2n + 8p + 4\alpha - 1$*, and* $\epsilon_1 = \sqrt{\frac{f_2(\boldsymbol{\beta}, \tilde{\boldsymbol{\tau}}) + 1}{(2\lambda + \frac{1}{\tau_1} + \frac{1}{\tilde{\tau}_1})\nu_1}}$*. Then there is a finite constant* $f_3(\boldsymbol{\beta}, \tilde{\boldsymbol{\tau}})$ *such that*

$$\int_{\mathbb{R}} \frac{\tilde{\beta}_1^2}{\left(1 + \frac{(\tilde{\boldsymbol{\beta}}_1 - \mu_1)^2}{\nu_1 \epsilon_1^2}\right)^{\frac{1+\nu_1}{2}}} d\tilde{\beta}_1 \geq f_3(\boldsymbol{\beta}, \tilde{\boldsymbol{\tau}}) \left(\frac{1}{2\lambda + \frac{1}{\tau_1} + \frac{1}{\tilde{\tau}_1}}\right)^{\frac{3}{2}}.$$

Proof. Note that

$$\int_{\mathbb{R}} \frac{\tilde{\beta}_1^2}{\left(1 + \frac{(\tilde{\boldsymbol{\beta}}_1 - \mu_1)^2}{\nu_1 \epsilon_1^2}\right)^{\frac{1+\nu_1}{2}}} d\tilde{\beta}_1 = \epsilon_1 \frac{\Gamma(\frac{\nu_1}{2})\sqrt{\pi\nu}}{\Gamma(\frac{\nu+1}{2})} E(U^2),$$

where U follows a t-distribution with scale ϵ_1, location μ_1 and degrees of freedom ν_1. Using Proposition A3, we get that

$$\begin{aligned}
&\int_{\mathbb{R}} \frac{\tilde{\beta}_1^2}{\left(1 + \frac{(\tilde{\boldsymbol{\beta}}_1 - \mu_1)^2}{\nu_1 \epsilon_1^2}\right)^{\frac{1+\nu_1}{2}}} d\tilde{\beta}_1 \\
&= \epsilon_1 \frac{\Gamma(\frac{\nu_1}{2})\sqrt{\pi\nu}}{\Gamma(\frac{\nu+1}{2})} E(U^2) \\
&\quad \geq \frac{\Gamma(\frac{\nu_1}{2})\sqrt{\pi\nu}}{\Gamma(\frac{\nu+1}{2})} \frac{\nu_1}{\nu_1 - 2}\ \epsilon_1^3, \\
&\quad = \frac{\Gamma(\frac{\nu_1}{2})\sqrt{\pi\nu}}{\Gamma(\frac{\nu+1}{2})} \frac{\nu_1}{\nu_1 - 2} \left(\frac{f_2(\boldsymbol{\beta}, \tilde{\boldsymbol{\tau}}) + 1}{\nu_1}\right)^{\frac{3}{2}} \left(\frac{1}{2\lambda + \frac{1}{\tau_1} + \frac{1}{\tilde{\tau}_1}}\right)^{\frac{3}{2}}
\end{aligned}$$

$$= f_3(\boldsymbol{\beta}, \tilde{\boldsymbol{\tau}}) \left(2\lambda + \frac{1}{\tau_1} + \frac{1}{\tilde{\tau}_1}\right)^{-\frac{3}{2}}.$$

where $f_3(\boldsymbol{\beta}, \tilde{\boldsymbol{\tau}}) = \frac{\Gamma(\frac{\nu_1}{2})\sqrt{\pi\nu}}{\Gamma(\frac{\nu+1}{2})} \frac{\nu_1}{\nu_1 - 2} \left(\frac{f_2(\boldsymbol{\beta}, \tilde{\boldsymbol{\tau}})+1}{\nu_1}\right)^{\frac{3}{2}}$. $\square$

References

[1] Abramowitz, M. and Stegun, I. A. (1965). Handbook of Mathematical Functions:With Formulas, Graphs, and Mathematical Tables (Dover Books on Mathematics), 1 ed. *Dover books on mathematics*. Dover Publications. MR1225604

[2] Adamczak, R. and Bednorz, W. (2015). Some remarks on MCMC estimation of spectra of integral operators. *Bernoulli* **21**, 2073-2092. MR3378459

[3] Chakraborty, S. and Khare, K. (2018). Consistent estimation of the spectrum of trace class data augmentation algorithms, *arxiv*.

[4] de los Campos, G., Naya, H., Gianola, D., Crossa, J., Legarra, A., Manfredi, E., Weigel, K. and Cotes, J.M. (2009). Predicting quantitative traits with regression models for dense molecular markers and pedigree, *Genetics* **182**, 375-385.

[5] DeMiguel, V., Garlappi, L., Nogales, F. J. and Uppal, R. (2009). A generalized approach to portfolio optimization: improving performance by constraining portfolio norms, *Management Science* **55**, 798-812.

[6] Fiedler, M.(1971). Bounds for the Determinant of the Sum of Hermitian Matrices, *Proceedings of the American Mathematical Society* **30**, 27-31. MR0286814

[7] Griffin, J. E. and Brown, P. J. (2010). Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis* **5**, 171-188. MR2596440

[8] Gu, X., Yin, G. and Lee, J. J. (2013). Bayesian two-step Lasso strategy for biomarker selection in personalized medicine development for time-to-event endpoints, *Contemporary Clinical Trials* **36**, 642-650.

[9] Hobert, J. P. and Marchev, D. (2008). A theoretical comparison of the data augmentation, marginal augmentation and PX-DA algorithms, *The Annals of Statistics* **36**, 532-554. MR2396806

[10] 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. MR2918006

[11] Jacquemin, S. J. and Doll, J. C. (2014). Body size and geographic range do not explain long term variation in fish populations: a Bayesian phylogenetic approach to testing assembly processes in stream fish assemblages, *PLoS ONE* **9**, 1-7.

[12] Johndrow, J.E., Orenstein, P. and Bhattacharya, A. (2018). Scalable MCMC for Bayes shrinkage priors, *arxiv*.

[13] Jorgens, K. (1982). *Linear Integral Operators*, Pitman Books, London. MR0647629

[14] Khare, K. and Hobert, J. P. (2011). A spectral analytic comparison of trace-class data augmentation algorithms and their sandwich variants. *The Annals of Statistics* **39**, 2585-2606. MR2906879

[15] Kotz, S. and Nadarajah, S. (2004). *Multivariate t Distributions and Their Applications*, Cambridge University Press, Cambridge. MR2038227

[16] Laforgia, A. (1991). Bounds for modified Bessel functions. *Journal of Computational and Applied Mathematics* **34**, 263-267. MR1102583

[17] Liu, J. S. and Wu, Y. N. (1999). Parameter expansion for data augmentation, *J. Amer. Statist. Assoc.* **94**, 1264-1274. MR1731488

[18] 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. MR2054301

[19] Meng, X. L. and van Dyk, D. A. (1999). Seeking efficient data augmentation schemes via conditional and marginal augmentation, *Biometrika* **86**, 301-320. MR1705351

[20] Mishchenko, Y. and Paninski, L. (2012). A Bayesian compressed-sensing approach for reconstructing neural connectivity from subsampled anatomical data, *Journal of Computational Neuroscience* **33**, 371-388. MR2973795

[21] Pal, S and Khare, K. (2014). Geometric ergodicity for Bayesian shrinkage models, *Electronic Journal of Statistics* **8**, 604-645. MR3211026

[22] Pal, S. , Khare, K. and Hobert, J. P. (2016). Trace class Markov chains for Bayesian inference with generalized double Pareto shrinkage priors. *Scandinavian Journal of Statistics*, doi: 10.1111/sjos.12254. MR3658516

[23] Park, T. and Casella, G. (2008). The Bayesian Lasso, *Journal of the American Statistical Association* **103**, 681-686. MR2524001

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

[25] Pong-Wong, R. (2014). Estimation of genomic breeding values using the Horseshoe prior, *BMC Proc.* **8**, Suppl 5.

[26] Pong-Wong, R. and Woolliams, J. (2014). Bayes U: A genomic prediction method based on the Horseshoe prior, *World Congress of Genetics Applied to Livestock Production* **10**, Vancouver, Canada.

[27] Qin, Q., Hobert, J. and Khare, K. (2018). Estimating the spectral gap of a trace-class Markov operator, *arxiv*.

[28] Ruiz-Antolin, D. and Segura, J.(2016). A new type of sharp bounds for ratios of modified Bessel functions, arXiv:1606.02008. MR3514344

[29] Rajaratnam,B., Sparks,D., Khare,K. and Zhang, L. (2016). Scalable Bayesian shrinkage and uncertainty quantification for high-dimensional regression, arXiv:1509.03697, Cornell University Library. MR3453650

[30] Tibshirani, R. (1994). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B* **58** 267-288. MR1379242

[31] Vats, D., Flegal, J. and Jones, G. (2018). Multivariate output analysis for Markov Chain Monte Carlo, to appear in *Biometrika*. MR3653667

[32] Watson, G.N. (1944). *A Treatise on the Theory of Bessel Functions*, Cambridge University Press, Cambridge, 2^{nd} edition. MR0010746

[33] Xing, Z., Zhou, M., Castrodad, A., Sapiro, G. and Carin, L. (2012). Dictio-

nary learning for noisy and incomplete hyperspectral images, *SIAM Journal of Imaging Sciences* **5**, 33-56. MR2902656

[34] Yi, N. and Xu, S. (2008). Bayesian LASSO for quantitative trait loci mapping, *Genetics* **179**, 1045-1055.