Electronic Journal of Statistics
Vol. 16 (2022) 1–57
ISSN: 1935-7524
https://doi.org/10.1214/21-EJS1932

Geometric ergodicity of Gibbs samplers for the Horseshoe and its regularized variants

Suman Bhattacharya and Kshitij Khare

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

Subhadip Pal

Department of Bioinformatics and Biostatistics,
University of Louisville
e-mail: subhadip.pal@louisville.edu

Abstract: The Horseshoe is a widely used and popular continuous shrinkage prior for high-dimensional Bayesian linear regression. Recently, regularized versions of the Horseshoe prior have also been introduced in the literature. Various Gibbs sampling Markov chains have been developed in the literature to generate approximate samples from the corresponding intractable posterior densities. Establishing geometric ergodicity of these Markov chains provides crucial technical justification for the accuracy of asymptotic standard errors for Markov chain based estimates of posterior quantities. In this paper, we establish geometric ergodicity for various Gibbs samplers corresponding to the Horseshoe prior and its regularized variants in the context of linear regression. First, we establish geometric ergodicity of a Gibbs sampler for the original Horseshoe posterior under strictly weaker conditions than existing analyses in the literature. Second, we consider the regularized Horseshoe prior introduced in [18], and prove geometric ergodicity for a Gibbs sampling Markov chain to sample from the corresponding posterior without any truncation constraint on the global and local shrinkage parameters. Finally, we consider a variant of this regularized Horseshoe prior introduced in [15], and again establish geometric ergodicity for a Gibbs sampling Markov chain to sample from the corresponding posterior.

MSC2020 subject classifications: Primary 60J05; secondary 60J20.
Keywords and phrases: Markov chain Monte Carlo, geometric ergodicity, high-dimensional linear regression, Horseshoe prior.

Received January 2021.

1. Introduction

Consider the linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \sigma\boldsymbol{\varepsilon}$, where $\mathbf{y} \in \mathbb{R}^n$ is the response vector, $\mathbf{X}$ is the $n \times p$ design matrix, $\boldsymbol{\beta} \in \mathbb{R}^p$ is the vector of regression coefficients, $\boldsymbol{\varepsilon}$ is the error vector with i.i.d. standard normal components, and σ^2 is the error variance. The goal is to estimate the unknown parameters $(\boldsymbol{\beta}, \sigma^2)$. In modern applications, datasets where the number of predictors p is much larger than the sample size n are commonly encountered. A standard approach for meaningful statistical estimation in these over-parametrized settings is to assume that only a few of the signals are prominent (the others are small/insignificant). This is mathematically formalized by assuming that the underlying regression coefficient vector is sparse. In the Bayesian paradigm, this assumption of sparsity is accommodated either by choosing spike-and-slab priors (mixture of point mass at zero and an absolutely continuous density) or absolutely continuous shrinkage priors which selectively shrink the small/insignificant signals.

A variety of useful shrinkage priors have been proposed in the literature (see [2, 4, 19] and the references therein), and the Horseshoe prior ([4]) is a widely used and highly popular choice. The Horseshoe prior for linear regression is specified as follows.

$$\begin{aligned} \boldsymbol{\beta} \mid \boldsymbol{\lambda}, \sigma^2, \tau^2 &\sim \mathcal{N}_p(0, \sigma^2\tau^2\boldsymbol{\Lambda}) \\ \lambda_i &\sim C^+(0,1) \text{ independently for } i = 1, 2, \cdots, p \\ \tau^2 \sim \pi_\tau(\cdot) \quad & \sigma^2 \sim \text{Inverse-Gamma}(a, b) \end{aligned} \tag{1.1}$$

where $\mathcal{N}_d$ denotes the d-variate normal density, $\boldsymbol{\Lambda}$ is a diagonal matrix with diagonal entries given by the entries $\left\{\lambda_j^2\right\}_{j=1}^p$, Inverse-Gamma$(a, b)$ denotes the Inverse-Gamma density with shape parameter a and rate parameter b, $C^+(0,1)$ is the half-Cauchy density on $\mathbb{R}_+$, and π_τ refers to the marginal prior density of τ. The vector $\boldsymbol{\lambda} = (\lambda_j^2)_{j=1}^p$ is referred to as the vector of local (component-wise) shrinkage parameters, while τ^2 is referred to as the global shrinkage parameter.

The resulting posterior distribution for $(\boldsymbol{\beta}, \sigma^2)$ is intractable in the sense that closed form computations or i.i.d. sampling from this distribution are not feasible. Several Gibbs sampling Markov chains have been proposed in the literature to generate approximate samples from the Horseshoe posterior, see for example ([1, 7, 13]).

The fact that parameter values which are far away from zero are not regularized at all due to the heavy tails is considered to be a key strength of the Horseshoe prior. However, as pointed out in [18], this can be undesirable when the parameters are only weakly identified. To address this issue, [18] introduced the regularized Horseshoe prior, given by

$$\begin{aligned} \beta_i \mid \boldsymbol{\lambda}, \sigma^2, \tau^2 \sim \mathcal{N}_p\left(0, \left(\frac{1}{c^2} + \frac{1}{\lambda_i^2\tau^2}\right)^{-1}\sigma^2\right) &\text{ independently for } i = 1, 2, \cdots, p \\ \lambda_i \sim C^+(0,1) &\text{independently for } i = 1, 2, \cdots, p \\ \tau^2 \sim \pi_\tau(\cdot) \quad \sigma^2 &\sim \text{Inverse-Gamma}(a, b) \end{aligned}$$

Here c is a finite constant which controls additional regularization of all regression parameters (large and small). The original Horseshoe prior can be recovered by letting $c \to \infty$. [18] use a Hamiltonian Monte Carlo (HMC) based approach to generate approximate samples from the corresponding regularized Horseshoe posterior distribution. Also, any Gibbs sampler for the Horseshoe posterior can be suitably adapted in the regularized setting.

For any practitioner using Markov chain Monte Carlo, it is crucial to understand the accuracy of the resulting MCMC based estimates by obtaining valid standard errors for these estimates. The notion of geometric ergodicity plays an important role in this endeavor, as explained below. Let $(\boldsymbol{\beta}_m, \sigma_m^2)_{m \geq 0}$ denote a Harris ergodic Markov chain with the Horseshoe or regularized Horseshoe posterior density, denoted by $\pi_H(\cdot \mid \mathbf{y})$, as its stationary density. The Markov chain is said to be geometrically ergodic if

$$\left\| K^m_{\boldsymbol{\beta}_0, \sigma_0^2} - \Pi_H \right\|_{\mathrm{TV}} \leq C\left(\boldsymbol{\beta}_0, \sigma_0^2\right) \gamma^m$$

for positive constants $C\left(\boldsymbol{\beta}_0, \sigma_0^2\right)$ and $\gamma \in [0,1)$. Here $K^m_{\boldsymbol{\beta}_0, \sigma_0^2}$ denotes the distribution of the Markov chain started at $(\boldsymbol{\beta}_0, \sigma_0^2)$ after m steps, Π_H denotes the stationary distribution, and $\|\cdot\|_{\mathrm{TV}}$ denotes the total variation norm. Suppose we wish to evaluate the posterior expectation

$$E_{\pi_H(\cdot|\mathbf{y})} g = \int \int g\left(\boldsymbol{\beta}, \sigma^2\right) \pi_H\left(\boldsymbol{\beta}, \sigma^2 \mid \mathbf{y}\right) d\boldsymbol{\beta} d\sigma^2$$

for a real-valued measurable function g of interest. Harris ergodicity guarantees that the Markov chain based estimator

$$\bar{g}_m := \frac{1}{m+1} \sum_{i=0}^{m} g\left(\boldsymbol{\beta}_i, \sigma_i^2\right)$$

is strongly consistent for $E_{\pi(\cdot|\mathbf{y})} g$. An estimate by itself, however, is not quite useful without an associated standard error. All known methods to compute consistent estimates (see for example [8, 6]) of the standard error for $\bar{g}_m$ require the existence of a Markov chain Central Limit Theorem (CLT) which establishes

$$\sqrt{m}\left(\bar{g}_m - E_{\pi_H(\cdot|\mathbf{y})} g\right) \to \mathcal{N}(0, \sigma_g^2),$$

for $\sigma_g^2 \in (0, \infty)$. In turn, the standard approach for establishing a Markov chain CLT requires proving geometric ergodicity of the underlying Markov chain. To summarize, proving geometric ergodicity helps rigorously establish the asymptotic validity of CLT based standard error estimates used by MCMC practitioners.

Establishing geometric ergodicity for continuous state space Markov chains encountered in most statistical applications is in general a very challenging task. For a significant majority of Markov chains in statistical applications, the question of whether they are geometrically ergodic or not has not been resolved, although there have been some success stories. In the context of Markov chains arising in Bayesian shrinkage, geometric ergodicity of Gibbs samplers corresponding to various shrinkage priors such as the Bayesian lasso, Normal-Gamma, Dirichlet-Laplace and double Pareto priors has been recently established in ([10, 17, 16]).

Results for the Horseshoe prior remained elusive until very recently. The marginal Horseshoe prior on entries of $\boldsymbol{\beta}$ (integrating out $\boldsymbol{\lambda}$, given τ^2) has an infinite spike near zero and significantly heavier tails than the shrinkage priors mentioned above. This structure, while making it very attractive for sparsity selection, implicitly creates a lot of complications and challenges in the geometric ergodicity analysis using drift and minorization techniques. Recently, the authors in [7] derived a two-block Gibbs sampler for the Horseshoe posterior (the 'exact algorithm' in [7, Section 2.1], henceforth referred to as the JOB Gibbs sampler), and established geometric ergodicity ([7, Theorem 14]). However, the truncation assumptions needed for this result are rather restrictive, requiring all the local shrinkage parameters λ_i^2 to be bounded above by a finite constant, and also requiring the global shrinkage parameter τ^2 to be bounded above and below by finite positive constants. In parallel work ([3], uploaded on arXiv less than a month prior to our submission) geometric ergodicity for the JOB Gibbs sampler has now been established without requiring truncation of the local shrinkage parameters. However, the requirement of the global shrinkage parameter τ^2 to be bounded above *and* below remains.

Contribution #1: The first contribution of this paper is the proof of geometric ergodicity for a Horseshoe Gibbs sampler (see Theorem 2.1) with no truncation assumptions on the local shrinkage parameters, and *with the global shrinkage parameter only required to be truncated below by a finite positive constant and to have a finite δ^{th} prior moment for some $\delta > 0.00081$*. Hence, the conditions required for our geometric ergodicity result are strictly weaker than those in [7] and [3]. In fact, as discussed in Remark 2.1, the assumption of truncation below by a positive constant can be further relaxed to existence of the negative $(p+\delta)/2^{th}$ prior moment for some $\delta > 0.00162$.

The Gibbs sampler analyzed in Theorem 2.1 is a slight modification of the JOB Gibbs sampler with latent variables introduced to simplify conditional sampling of the local shrinkage parameters in the Markov chain (see Section 2 for more details). There are also important differences in the technical arguments compared to [7, 3]. We focus on the $\boldsymbol{\lambda}$-block of the Gibbs sampler and establish a drift condition (Lemma 2.1) using a drift function which is 'unbounded off compact sets', and that directly leads to geometric ergodicity. On the other hand, the approaches in [7, 3] use other drift functions (using all the parameters or a different parameter block than $\boldsymbol{\lambda}$) which are not unbounded off compact sets, and hence need an additional minorization argument.

Next we move to the regularized Horseshoe setting of [18]. As mentioned previously, [18] use a Hamiltonian Monte Carlo (HMC) based approach to generate approximate samples from the corresponding regularized Horseshoe posterior distribution, but do not investigate geometric ergodicity of the proposed Markov chain. It is not clear whether the intricate sufficient conditions needed for geometric ergodicity of HMC chains in [11] apply to the HMC chain in [18]. Given the variety of efficient Gibbs samplers available for the original Horseshoe posterior, it is natural to consider an appropriately adapted version of any of these samplers for the regularized Horseshoe posterior.

Contribution #2: As the second main contribution of this paper, we establish geometric ergodicity for one such Gibbs sampler for the regularized Horseshoe posterior (see Theorem 3.1) *with no truncation assumptions on the global* *and local shrinkage parameters at all.* The seemingly minor change in the prior structure (compared to the original Horseshoe), leads to crucial changes in our convergence analysis. For example, we need a different drift function for this analysis (Lemma 3.1) compared to the original Horseshoe analysis. This drift function is not 'unbounded off compact sets', and hence we need an additional minorization condition (Lemma 3.2) to establish geometric ergodicity.

Recently, [15] construct a further variant of the regularized Horseshoe prior of [18] by changing the algebraic form of the conditional prior density of $\boldsymbol{\beta}$ for computational simplicity. Their prior specification is as follows.

$$
\begin{aligned}
\pi\left(\beta_j, \lambda_j^2 \mid \tau^2, \sigma^2\right) \quad &\propto \quad \frac{1}{\sqrt{\tau^2 \lambda_j^2}} \exp\left[-\frac{\beta_j^2}{2\sigma^2}\left(\frac{1}{c^2} + \frac{1}{\tau^2 \lambda_j^2}\right)\right] \pi_\ell\left(\lambda_j\right) \\
&\qquad \text{independently for } j = 1, 2, \cdots, p \\
\tau^2 \sim \pi_\tau(\cdot) \qquad &\qquad \sigma^2 \sim \text{Inverse-Gamma}(a, b)
\end{aligned}
$$

The algebraic modification, in particular removal of the $(c^{-2} + (\lambda_i \tau)^{-2})^{1/2}$ in the conditional prior for β_i simplifies posterior computation (see Section 3.3 for more details). [15] prove geometric ergodicity for the related but structurally different setting of Polya-Gamma logistic regression assuming that the global shrinkage parameter τ^2 is bounded above and below by finite positive constants. However, as discussed in Remark 3.1, several details of this analysis break down in the linear regression setting.

Contribution #3: We focus on the linear regression setting, and leverage our analysis in the original Horseshoe setting to prove geometric ergodicity of a Gibbs sampler corresponding to [15]'s regularized variant with the global shrinkage parameter only required to be bounded below by a finite positive constant and to have a finite $(p+\delta)/2^{th}$ moment for some $\delta > 0.00162$.

The rest of the paper is structured as follows. We introduce the modified version of the JOB Gibbs sampler in Section 2.1. Geometric ergodicity of this Gibbs sampler is established in Section 2.2. The simulation study in Section 2.3 compares the computational time and other metrics for the JOB Gibbs sampler and the proposed modification in a variety of settings. An adaptation of the Horseshoe Gibbs sampler for the regularized Horseshoe posterior is developed in Section 3.1. The geometric ergodicity of this regularized Horseshoe Gibbs sampler is established in Section 3.2. A related Gibbs sampler for the regularized Horseshoe variant of [15] is discussed and analyzed in Section 3.3. Another simulation study in Section 3.4 examinies the computational feasibility/scalability of the Gibbs samplers analyzed in Sections 3.1 and 3.3. The proofs of several technical results used in the analysis are contained in an Appendix.

2. Geometric ergodicity of a Horseshoe Gibbs sampler

2.1. A modified version of the JOB Gibbs sampler

In this section, we describe in detail the Horseshoe Gibbs sampler that will be analyzed in subsequent sections. As pointed out in [13], if

$$\lambda_j^2 \mid \nu_j \sim \text{Inverse-Gamma}(1/2, 1/\nu_j)$$

and $\nu_j \sim$ Inverse-Gamma$(1/2, 1)$, then $\lambda_j \sim C^+(0,1)$. Using this fact, with $\boldsymbol{\nu} = (\nu_1, \nu_2, \cdots, \nu_p)$, the Horseshoe prior in (1.1) can be alternatively written as

$$\begin{aligned} \boldsymbol{\beta} \mid \boldsymbol{\lambda}, \sigma^2, \tau^2 &\sim \mathcal{N}_p(0, \sigma^2\tau^2\boldsymbol{\Lambda}) \\ \lambda_i^2 \mid \boldsymbol{\nu} &\sim \text{Inverse-Gamma}(1/2, 1/\nu_i) \text{ independently for } i = 1, 2, \cdots, p \\ \nu_i &\sim \text{Inverse-Gamma}(1/2, 1) \text{ independently for } i = 1, 2, \cdots, p \\ \tau^2 \sim \pi_\tau(\cdot), \quad \sigma^2 &\sim \text{Inverse-Gamma}(a, b) \end{aligned} \tag{2.1}$$

Using the prior above and after straightforward calculations, various conditional posterior distributions can be derived as follows.

$$\begin{aligned} \boldsymbol{\beta} \big| \, \sigma^2, \tau^2, \boldsymbol{\lambda}, \boldsymbol{\nu}, \boldsymbol{y} &\sim \mathcal{N}_p(A^{-1}\mathbf{X}^T\mathbf{y}, \sigma^2 A^{-1}) \\ \sigma^2 \big| \, \tau^2, \boldsymbol{\lambda}, \boldsymbol{\nu}, \boldsymbol{y} &\sim \text{Inverse-Gamma}\left(a + \frac{n}{2}, \frac{\mathbf{y}^T\left(I_n - \tilde{P}_{\mathbf{X}}\right)\mathbf{y}}{2} + b\right) \\ \lambda_j^2 \big| \, \boldsymbol{\nu}, \sigma^2, \tau^2, \boldsymbol{\beta}, \mathbf{y} &\sim \text{Inverse-Gamma}\left(1, \frac{1}{\nu_j} + \frac{\beta_j^2}{2\sigma^2\tau^2}\right) \text{ indep. for } j = 1, 2, \cdots, p \\ \nu_j \big| \, \boldsymbol{\lambda}, \tau^2, \mathbf{y} &\sim \text{Inverse-Gamma}\left(1, 1 + \frac{1}{\lambda_j^2}\right) \text{ indep. for } j = 1, 2, \cdots, p \\ \tau^2 \big| \, \boldsymbol{\lambda}, \boldsymbol{y} \sim \pi\left(\tau^2 \big| \, \boldsymbol{\lambda}, \boldsymbol{y}\right) &\propto \frac{\left(\frac{\mathbf{y}^T\left(I_n - \tilde{P}_{\mathbf{X}}\right)\mathbf{y}}{2} + b\right)^{-\left(a + \frac{n}{2}\right)}}{\sqrt{|I_p + \mathbf{X}^T\mathbf{X}.\boldsymbol{\Lambda}_*|}} \cdot \pi_\tau(\tau^2) \end{aligned} \tag{2.2}$$

where $\boldsymbol{\Lambda}_* = \tau^2\boldsymbol{\Lambda}$; $A = \mathbf{X}^T\mathbf{X} + \boldsymbol{\Lambda}_*^{-1}$ and $\tilde{P}_{\mathbf{X}} = \mathbf{X}A^{-1}\mathbf{X}^T$.

Consider a two-block Gibbs sampling Markov chain with transition kernel K_{aug} (with blocks $(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\nu}, \tau^2)$ and $\boldsymbol{\lambda}$) whose one-step transition from $(\boldsymbol{\beta}_0, \sigma_0^2, \boldsymbol{\nu}_0, \tau_0^2, \boldsymbol{\lambda}_0)$ to $(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\nu}, \tau^2, \boldsymbol{\lambda})$ is given as follows.

1. Draw $(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\nu}, \tau^2)$ from $\pi(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\nu}, \tau^2 \mid \boldsymbol{\lambda}_0, \mathbf{y})$. This can be done by sequentially drawing $\boldsymbol{\beta}$, then σ^2, then $\boldsymbol{\nu}$, and then τ^2 from appropriate conditional posterior densities in (2.2).
2. Draw $\boldsymbol{\lambda}$ from $\pi(\boldsymbol{\lambda} \mid \boldsymbol{\beta}, \boldsymbol{\nu}, \sigma^2, \tau^2, \mathbf{y})$. This can be done by independently drawing the components of $\boldsymbol{\lambda}$ from the appropriate full conditional posterior density in (2.2).

The JOB Gibbs sampler from [7] is very similar to the above two-block Gibbs sampler K_{aug}. The difference is that the latent variables $\boldsymbol{\nu}$ are not used, and the two blocks used in the JOB Gibbs sampler are $(\boldsymbol{\beta}, \sigma^2, \tau^2)$ and $\boldsymbol{\lambda}$. While the sampling steps for $\boldsymbol{\beta}, \sigma^2, \tau^2$ are exactly the same as above, the components of $\boldsymbol{\lambda}$ are sampled differently. In particular, each λ_j is sampled from the conditional density given $\beta_j, \sigma^2, \tau^2, \mathbf{y}$ (no conditioning on ν_j). This conditional density is not a standard density, and draws are made using a rejection sampler. To summarize, by considering the latent variables $\boldsymbol{\nu}$, we replace the p rejection sampler based draws from a non-standard density in the JOB Gibbs sampler (for components of $\boldsymbol{\lambda}$) with $2p$ draws from standard Inverse-Gamma densities (for components of $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}$).

The Gibbs sampler K_{aug} can essentially be considered a hybrid of the JOB Gibbs sampler and the Gibbs sampler in [13], which uses a latent variable ξ (in addition to $\boldsymbol{\nu}$) to replace the draws from the non-standard $\pi(\tau^2 \mid \boldsymbol{\lambda}, \mathbf{y})$ density with two draws from standard Inverse-Gamma densities. As mentioned in the introduction, the geometric ergodicity result for the JOB Gibbs sampler in [7, Theorem 14] has been established by assuming that the local shrinkage parameters in $\boldsymbol{\lambda}$ are all bounded above, and the global shrinakge parameter τ^2 is bounded above and below. In very recent follow-up work [3], the authors establish geometric ergodcity for a class of Half-t Gibbs samplers of which the JOB Gibbs sampler is a member. In this work, the truncation assumption on the local shrinkage parameters has been removed, but the global shrinkage parameter is still assumed to be truncated above *and* below. However, we show below that geometric ergodicity for the hybrid Gibbs sampler K_{aug} can be established with no truncation at all on the local shrinkage parameters in $\boldsymbol{\lambda}$, and only assuming that the global shrinkage parameter τ^2 is truncated below.

The reasons for this improved analysis of the hybrid chain K_{aug} lie in the intricacies of drift and minorization approach ([21]), which is the state of the art technique for proving geometric ergodicity for general state space Markov chains. The introduction of the latent variables $\boldsymbol{\nu}$, the resulting Inverse-Gamma posterior conditionals for entries of $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}$, and avoiding the latent variable ξ for the global shrinkage parameter τ^2 provide just the right ingredients for establishing a geometric drift condition in Section 2.2 which is then leveraged to establish geometric ergodicity. Even a minor deviation in the structure of the Markov chain (such as in the JOB Gibbs sampler or the Gibbs sampler of [13]) leads to a breakdown of the intricate argument.

Before proceeding further, we note that geometric ergodicity of a two-block Gibbs sampler can be established by showing that any of its two marginal chains is geometrically ergodic (see for example [20]). Hence, we focus on the marginal $\boldsymbol{\lambda}$-chain corresponding to K_{aug}. The one-step transition dynamics of this Markov chain from $\boldsymbol{\lambda}_m$ to $\boldsymbol{\lambda}_{m+1}$ is given as follows:

1. Draw τ^2 from $\pi\left(\tau^2 \middle| \boldsymbol{\lambda}_m, \mathbf{y}\right)$
2. Draw $\boldsymbol{\nu}$ from $\pi\left(\boldsymbol{\nu} \middle| \boldsymbol{\lambda}_m, \tau^2, \mathbf{y}\right) = \prod_{j=1}^{p} \text{Inverse-Gamma}\left(1, 1 + \frac{1}{\lambda_{j;m}^2}\right)$
3. Draw σ^2 from $\pi\left(\sigma^2 \middle| \tau^2, \boldsymbol{\lambda}_m, \boldsymbol{\nu}, \mathbf{y}\right) = \text{Inverse-Gamma}\Bigg(a + \frac{n}{2}, \frac{\mathbf{y}^T\left(I_n - \tilde{P}_{\mathbf{X}}\right)\mathbf{y}}{2}$

$$+ b\Big)$$

4. Draw $\boldsymbol{\beta}$ from $\pi\left(\boldsymbol{\beta}|\,\sigma^2,\tau^2,\boldsymbol{\lambda}_m,\boldsymbol{\nu},\mathbf{y}\right) = \mathcal{N}_p(A^{-1}\mathbf{X}^T\mathbf{y},\sigma^2A^{-1})$
5. Finally draw $\boldsymbol{\lambda}_{m+1}$ from

$$\pi\left(\boldsymbol{\lambda}|\,\boldsymbol{\beta},\boldsymbol{\nu},\sigma^2,\tau^2,\mathbf{y}\right) = \prod_{j=1}^{p} \text{Inverse-Gamma}\left(1, \frac{1}{\nu_j} + \frac{\beta_j^2}{2\sigma^2\tau^2}\right)$$

The Markov transition density (MTD) corresponding to the marginal $\boldsymbol{\lambda}$-chain is given by

$$\begin{aligned} k\left(\boldsymbol{\lambda}_0,\boldsymbol{\lambda}\right) &= \int_{\mathbb{R}_+}\int_{\mathbb{R}_+}\int_{\mathbb{R}^p}\int_{\mathbb{R}_+^p} \pi\big(\boldsymbol{\lambda}|\,\boldsymbol{\beta},\boldsymbol{\nu},\sigma^2,\tau^2,\mathbf{y}\big)\pi\big(\boldsymbol{\beta},\boldsymbol{\nu},\sigma^2,\tau^2\big|\,\boldsymbol{\lambda}_0,\mathbf{y}\big)\,d\boldsymbol{\nu}d\boldsymbol{\beta}d\sigma^2d\tau^2 \\ &= \int_{\mathbb{R}_+}\int_{\mathbb{R}_+}\int_{\mathbb{R}^p}\int_{\mathbb{R}_+^p} \pi\big(\boldsymbol{\lambda}|\,\boldsymbol{\beta},\boldsymbol{\nu},\sigma^2,\tau^2,\mathbf{y}\big)\pi\big(\boldsymbol{\beta}|\,\sigma^2,\tau^2,\boldsymbol{\lambda}_0,\boldsymbol{\nu},\mathbf{y}\big) \\ &\quad \times \pi\left(\sigma^2\big|\,\tau^2,\boldsymbol{\lambda}_0,\boldsymbol{\nu},\mathbf{y}\right)\pi\left(\boldsymbol{\nu}|\,\boldsymbol{\lambda}_0,\tau^2,\mathbf{y}\right)\pi\left(\tau^2\big|\,\boldsymbol{\lambda}_0,\mathbf{y}\right)d\boldsymbol{\nu}d\boldsymbol{\beta}d\sigma^2d\tau^2 \end{aligned} \tag{2.3}$$

We now establish a drift condition for the marginal $\boldsymbol{\lambda}$-chain, which will then be used to establish geometric ergodicity for the two-block Horseshoe Gibbs sampler K_{aug}.

2.2. A drift condition for the $\boldsymbol{\lambda}$-chain

Consider the function $V : \mathbb{R}_+^p \mapsto [0,\infty)$ given by

$$V\left(\boldsymbol{\lambda}\right) = \sum_{j=1}^{p}\left(\lambda_j^2\right)^{\frac{\delta_0}{2}} + \sum_{j=1}^{p}\left(\lambda_j^2\right)^{-\frac{\delta_1}{2}}, \tag{2.4}$$

where $\delta_0,\delta_1 \in (0,1)$ are some constants. The next result establishes a geometric drift condition for the marginal $\boldsymbol{\lambda}$-chain using the function V with appropriately small values of δ_0 and δ_1.

Lemma 2.1. *Suppose the prior density π_τ for the global shrinkage parameter is truncated below i.e., $\pi_\tau(u) = 0$ for $u < T$ for some $T > 0$ and satisfies*

$$\int_T^\infty u^{\delta/2}\pi_\tau(u)du < \infty$$

for some $\delta \in (0.00162, 0.22176)$. Then, there exist $\delta_0,\delta_1 \in (0,1)$ such that for every $\boldsymbol{\lambda}_0 \in \mathbb{R}_+^p$ we have

$$\boldsymbol{E}\left[V\left(\boldsymbol{\lambda}\right)|\,\boldsymbol{\lambda}_0\right] = \int_{\mathbb{R}_+^p} k\left(\boldsymbol{\lambda}_0,\boldsymbol{\lambda}\right)V(\boldsymbol{\lambda})d\boldsymbol{\lambda} \le \gamma^* V\left(\boldsymbol{\lambda}_0\right) + b^* \tag{2.5}$$

with $0 < \gamma^* = \gamma^*\left(\delta_0,\delta_1\right) < 1$ and $b^* = b^*\left(\delta_0,\delta_1\right) < \infty$.

Proof. Note that by linearity

$$\boldsymbol{E}\left[V(\boldsymbol{\lambda})\middle|\boldsymbol{\lambda}_0\right]=\sum_{j=1}^{p}\boldsymbol{E}\left[\left(\lambda_j^2\right)^{\frac{\delta_0}{2}}\middle|\boldsymbol{\lambda}_0\right]+\sum_{j=1}^{p}\boldsymbol{E}\left[\left(\lambda_j^2\right)^{-\frac{\delta_1}{2}}\middle|\boldsymbol{\lambda}_0\right] \tag{2.6}$$

We first consider terms in the second sum in (2.6). Fix $j\in\{1,2,\cdots,p\}$ arbitrarily. It follows from the definition of the MTD (2.3) that

$$\begin{aligned}\boldsymbol{E}\left[\left(\lambda_j^2\right)^{-\frac{\delta_1}{2}}\middle|\boldsymbol{\lambda}_0\right]=\boldsymbol{E}[\boldsymbol{E}[\boldsymbol{E}[\boldsymbol{E}[\boldsymbol{E}[\left(\lambda_j^2\right)^{-\frac{\delta_1}{2}}\mid\boldsymbol{\beta},\boldsymbol{\nu},\sigma^2,\tau^2,\mathbf{y}]\mid\sigma^2,\tau^2,\boldsymbol{\lambda}_0,\boldsymbol{\nu},\mathbf{y}]\\ \mid\tau^2,\boldsymbol{\lambda}_0,\boldsymbol{\nu},\mathbf{y}]\mid\boldsymbol{\lambda}_0,\tau^2,\mathbf{y}]\mid\boldsymbol{\lambda}_0,\mathbf{y}].\end{aligned} \tag{2.7}$$

The five iterated expectations correspond to the five conditional densities in (2.3). Starting with the innermost expectation, and using the fact that $1/\lambda_j^2$ (conditioned on $\boldsymbol{\beta},\boldsymbol{\nu},\sigma^2,\tau^2,\mathbf{y}$) follows a Gamma distribution with shape parameter 1 and rate parameter $1/\nu_j+\beta_j^2/(2\sigma^2\tau^2)$, we obtain

$$\begin{aligned}\boldsymbol{E}\left[\left(\lambda_j^2\right)^{-\frac{\delta_1}{2}}\middle|\boldsymbol{\beta},\boldsymbol{\nu},\sigma^2,\tau^2,\mathbf{y}\right] &= \Gamma\left(1+\frac{\delta_1}{2}\right)\left(\frac{1}{\nu_j}+\frac{\beta_j^2}{2\sigma^2\tau^2}\right)^{-\frac{\delta_1}{2}}\\ &= \Gamma\left(1+\frac{\delta_1}{2}\right)\left(\frac{1}{\nu_j}+\frac{1}{\left\{\frac{(2\sigma^2\tau^2)^{\frac{\delta_1}{2}}}{|\beta_j|^{\delta_1}}\right\}^{\frac{2}{\delta_1}}}\right)^{-\frac{\delta_1}{2}}\end{aligned}$$

Note that the function $y\mapsto\left(c+y^{-\frac{2}{\delta_1}}\right)^{-\delta_1/2}$ on $(0,\infty)$ is concave for $c>0,\delta_1\in(0,1)$. Applying the second iterated expectation, and using Jensen's inequality, it follows that

$$\begin{aligned}&\boldsymbol{E}\left[\boldsymbol{E}\left[\left(\lambda_j^2\right)^{-\frac{\delta_1}{2}}\middle|\boldsymbol{\beta},\boldsymbol{\nu},\sigma^2,\tau^2,\mathbf{y}\right]\middle|\sigma^2,\tau^2,\boldsymbol{\lambda}_0,\mathbf{y}\right]\\ \leq\;&\Gamma\left(1+\frac{\delta_1}{2}\right)\left(\frac{1}{\nu_j}+\frac{1}{\left\{E\left[\frac{(2\sigma^2\tau^2)^{\frac{\delta_1}{2}}}{|\beta_j|^{\delta_1}}\middle|\sigma^2,\tau^2,\lambda_0,\boldsymbol{\nu},\mathbf{y}\right]\right\}^{\frac{2}{\delta_1}}}\right)^{-\frac{\delta_1}{2}}\end{aligned} \tag{2.8}$$

Note that the conditional distribution of β_j given $\sigma^2,\tau^2,\boldsymbol{\lambda}_0,\boldsymbol{\nu},\mathbf{y}$ is a Gaussian distribution with variance $\sigma_j^2\overset{\text{def}}{=}\sigma^2\boldsymbol{e}_j^TA^{-1}\boldsymbol{e}_j\geq\sigma^2\left(\bar{\omega}+\frac{1}{\tau^2\lambda_{j;0}^2}\right)^{-1}$. Here $\bar{\omega}$ is the maximum eigenvalue of $\mathbf{X}^T\mathbf{X}$ and $\boldsymbol{e}_j$ is the $p\times1$ vector with j^{th} entry 1 and other entries equal to 0. Using Proposition A1 from [16] regarding the negative moments of a Gaussian random variable and choosing $\delta_1\in(0,1)$, it follows that

$$E\left[\frac{\left(2\sigma^2\tau^2\right)^{\frac{\delta_1}{2}}}{|\beta_j|^{\delta_1}}\middle|\sigma^2,\tau^2,\boldsymbol{\lambda}_0,\boldsymbol{\nu},\mathbf{y}\right]\quad\leq\quad\left(2\sigma^2\tau^2\right)^{\frac{\delta_1}{2}}\frac{\Gamma\left(\frac{1-\delta_1}{2}\right)2^{\frac{1-\delta_1}{2}}}{\sqrt{2\pi}\sigma_j^{\delta_1}}$$

$$\leq \quad \frac{\Gamma\left(\frac{1-\delta_1}{2}\right)}{\sqrt{\pi}} \left(\bar{\omega}\tau^2 + \frac{1}{\lambda_{j;0}^2}\right)^{\frac{\delta_1}{2}} \tag{2.9}$$

Combining (2.8) and (2.9), we obtain

$$\begin{aligned} &\boldsymbol{E}\left[\boldsymbol{E}\left[\left(\lambda_j^2\right)^{-\frac{\delta_1}{2}}\middle|\boldsymbol{\beta},\boldsymbol{\nu},\sigma^2,\tau^2,\mathbf{y}\right]\middle|\sigma^2,\tau^2,\boldsymbol{\lambda}_0,\mathbf{y}\right] \\ &\leq \quad \Gamma\left(1+\frac{\delta_1}{2}\right)\left(\frac{1}{\nu_j} + \frac{1}{\left\{\frac{\Gamma\left(\frac{1-\delta_1}{2}\right)}{\sqrt{\pi}}\left(\bar{\omega}\tau^2 + \frac{1}{\lambda_{j;0}^2}\right)^{\frac{\delta_1}{2}}\right\}^{\frac{2}{\delta_1}}}\right)^{-\frac{\delta_1}{2}}. \end{aligned}$$

Using the fact $(u+v)^\delta \leq u^\delta + v^\delta$ for $\delta \in (0,1)$ and $u, v \geq 0$, it follows that

$$\begin{aligned} &\boldsymbol{E}\left[\boldsymbol{E}\left[\left(\lambda_j^2\right)^{-\frac{\delta_1}{2}}\middle|\boldsymbol{\beta},\boldsymbol{\nu},\sigma^2,\tau^2,\mathbf{y}\right]\middle|\sigma^2,\tau^2,\boldsymbol{\lambda}_0,\mathbf{y}\right] \\ &\leq \quad \Gamma\left(1+\frac{\delta_1}{2}\right)\left(\frac{1}{\nu_j} + \frac{1}{\left\{\frac{\Gamma\left(\frac{1-\delta_1}{2}\right)}{\sqrt{\pi}}\left(\bar{\omega}^{\frac{\delta_1}{2}}\left(\tau^2\right)^{\frac{\delta_1}{2}} + \left(\lambda_{j;0}^2\right)^{-\frac{\delta_1}{2}}\right)\right\}^{\frac{2}{\delta_1}}}\right)^{-\frac{\delta_1}{2}} \end{aligned} \tag{2.10}$$

Note that the bound in (2.10) does not depend on σ^2. Again, using the fact that $y \mapsto \left(c + y^{-\frac{2}{\delta_1}}\right)^{-\delta_1/2}$ on $(0,\infty)$ is concave for $c > 0, \delta_1 \in (0,1)$, along with Jensen's inequality, we get

$$\begin{aligned} &\boldsymbol{E}\left[\left(\frac{1}{\nu_j} + \frac{1}{\left\{\frac{\Gamma\left(\frac{1-\delta_1}{2}\right)}{\sqrt{\pi}}\left(\bar{\omega}^{\frac{\delta_1}{2}}\left(\tau^2\right)^{\frac{\delta_1}{2}} + \left(\lambda_{j;0}^2\right)^{-\frac{\delta_1}{2}}\right)\right\}^{\frac{2}{\delta_1}}}\right)^{-\frac{\delta_1}{2}}\middle|\tau^2,\boldsymbol{\lambda}_0,\mathbf{y}\right] \\ &\leq \left(\frac{1}{\left\{E\left[\nu_j^{\frac{\delta_1}{2}}\middle|\tau^2,\lambda_0,\mathbf{y}\right]\right\}^{\frac{2}{\delta_1}}} + \frac{1}{\left\{\frac{\Gamma\left(\frac{1-\delta_1}{2}\right)}{\sqrt{\pi}}\left(\bar{\omega}^{\frac{\delta_1}{2}}\left(\tau^2\right)^{\frac{\delta_1}{2}} + \left(\lambda_{j;0}^2\right)^{-\frac{\delta_1}{2}}\right)\right\}^{\frac{2}{\delta_1}}}\right)^{-\frac{\delta_1}{2}} \end{aligned}$$

Since ν_j (given $\tau^2, \lambda_0, \mathbf{y}$) has an Inverse-Gamma distribution with shape parameter 1 and rate parameter $1 + 1/\lambda_{j;0}^2$, it follows that

$$\boldsymbol{E}\left[\left(\frac{1}{\nu_j}+\frac{1}{\left\{\frac{\Gamma\left(\frac{1-\delta_1}{2}\right)}{\sqrt{\pi}}\left(\bar{\omega}^{\frac{\delta_1}{2}}\left(\tau^2\right)^{\frac{\delta_1}{2}}+\left(\lambda_{j;0}^2\right)^{-\frac{\delta_1}{2}}\right)\right\}^{\frac{2}{\delta_1}}}\right)^{-\frac{\delta_1}{2}}\middle|\,\tau^2,\boldsymbol{\lambda}_0,\mathbf{y}\right]=$$
$$\left(\frac{1}{\left\{\Gamma\left(1-\frac{\delta_1}{2}\right)\left(1+\frac{1}{\lambda_{j;0}^2}\right)^{\frac{\delta_1}{2}}\right\}^{\frac{2}{\delta_1}}}+\frac{1}{\left\{\frac{\Gamma\left(\frac{1-\delta_1}{2}\right)}{\sqrt{\pi}}\left(\bar{\omega}^{\frac{\delta_1}{2}}\left(\tau^2\right)^{\frac{\delta_1}{2}}+\left(\lambda_{j;0}^2\right)^{-\frac{\delta_1}{2}}\right)\right\}^{\frac{2}{\delta_1}}}\right)^{-\frac{\delta_1}{2}}. \tag{2.11}$$

Let us now take the expectation of the expression in (2.11) with respect to the conditional distribution of τ^2 given $\boldsymbol{\lambda}_0,\mathbf{y}$. Using for a third time the fact that $y\mapsto\left(c+y^{-\frac{2}{\delta_1}}\right)^{-\delta_1/2}$ on $(0,\infty)$ is concave for $c>0,\delta_1\in(0,1)$, along with Jensen's inequality, we get

$$\begin{aligned}
&\boldsymbol{E}\left[\left(\frac{1}{\left\{\Gamma\left(1-\frac{\delta_1}{2}\right)\left(1+\frac{1}{\lambda_{j;0}^{\delta_1}}\right)\right\}^{\frac{2}{\delta_1}}}+\frac{1}{\left\{\frac{\Gamma\left(\frac{1-\delta_1}{2}\right)}{\sqrt{\pi}}\left(\bar{\omega}^{\frac{\delta_1}{2}}\left(\tau^2\right)^{\frac{\delta_1}{2}}+\left(\lambda_{j;0}^2\right)^{-\frac{\delta_1}{2}}\right)\right\}^{\frac{2}{\delta_1}}}\right)^{-\frac{\delta_1}{2}}\middle|\,\boldsymbol{\lambda}_0,\mathbf{y}\right]\\
&\le\left(\frac{1}{\left\{\Gamma\left(1-\frac{\delta_1}{2}\right)\left(1+\frac{1}{\lambda_{j;0}^{\delta_1}}\right)\right\}^{\frac{2}{\delta_1}}}+\frac{1}{\left\{\frac{\Gamma\left(\frac{1-\delta_1}{2}\right)}{\sqrt{\pi}}\left(\bar{\omega}^{\frac{\delta_1}{2}}\boldsymbol{E}\left[\left(\tau^2\right)^{\frac{\delta_1}{2}}\middle|\,\boldsymbol{\lambda}_0,\mathbf{y}\right]+\left(\lambda_{j;0}^2\right)^{-\frac{\delta_1}{2}}\right)\right\}^{\frac{2}{\delta_1}}}\right)^{-\frac{\delta_1}{2}}\\
&\quad\overset{(\bigstar)}{\le}\left(\frac{1}{\left\{\Gamma\left(1-\frac{\delta_1}{2}\right)\left(1+\frac{1}{\lambda_{j;0}^{\delta_1}}\right)\right\}^{\frac{2}{\delta_1}}}+\frac{1}{\left\{\frac{\Gamma\left(\frac{1-\delta_1}{2}\right)}{\sqrt{\pi}}\left(\bar{\omega}^{\frac{\delta_1}{2}}C_1+\frac{1}{\lambda_{j;0}^{\delta_1}}\right)\right\}^{\frac{2}{\delta_1}}}\right)^{-\frac{\delta_1}{2}}\\
&\le\left(C_0+\frac{1}{\lambda_{j;0}^{\delta_1}}\right)\left(\frac{1}{\left\{\Gamma\left(1-\frac{\delta_1}{2}\right)\right\}^{\frac{2}{\delta_1}}}+\frac{\sqrt{\pi}^{\frac{2}{\delta_1}}}{\left\{\Gamma\left(\frac{1-\delta_1}{2}\right)\right\}^{\frac{2}{\delta_1}}}\right)^{-\frac{\delta_1}{2}}\ ;\quad C_0=\max\left\{1,\bar{\omega}^{\frac{\delta_1}{2}}C_1\right\}
\end{aligned} \tag{2.12}$$

where $(\bigstar)$ follows from Proposition B.1, and C_1 is as defined in Proposition B.1. Combining (2.7), (2.8), (2.10), (2.11) and (2.12), we get

$$\boldsymbol{E}\left[\sum_{j=1}^{p}\left(\lambda_j^2\right)^{-\frac{\delta_1}{2}}\middle|\,\boldsymbol{\lambda}_0\right]\le\gamma\left(\delta_1\right)\sum_{j=1}^{p}\left(\lambda_{j;0}^2\right)^{-\frac{\delta_1}{2}}+b_1 \tag{2.13}$$

where

$$\gamma\left(\delta_1\right)=\Gamma\left(1+\frac{\delta_1}{2}\right)\left(\frac{1}{\left\{\Gamma\left(1-\frac{\delta_1}{2}\right)\right\}^{\frac{2}{\delta_1}}}+\frac{\sqrt{\pi}^{\frac{2}{\delta_1}}}{\left\{\Gamma\left(\frac{1-\delta_1}{2}\right)\right\}^{\frac{2}{\delta_1}}}\right)^{-\frac{\delta_1}{2}}$$

and

$$b_1=p\cdot C_0\cdot\gamma\left(\delta_1\right).$$

Next consider $\boldsymbol{E}\left[\sum_{j=1}^p\left(\lambda_j^2\right)^{\frac{\delta_0}{2}}\middle|\boldsymbol{\lambda}_0\right]$. Fix a $j\in\{1,2,\cdots,p\}$ arbitrarily. Since $\delta_0\in(0,1)$, using the fact that $(u+v)^{\delta_0}\leq u^{\delta_0}+v^{\delta_0}$ for $u,v\geq 0$ we get

$$\begin{aligned}\boldsymbol{E}\left[\left(\lambda_j^2\right)^{\frac{\delta_0}{2}}\middle|\boldsymbol{\beta},\boldsymbol{\nu},\sigma^2,\tau^2,\mathbf{y}\right]&=\Gamma\left(1-\frac{\delta_0}{2}\right)\left(\frac{1}{\nu_j}+\frac{\beta_j^2}{2\sigma^2\tau^2}\right)^{\frac{\delta_0}{2}}\\&\leq\Gamma\left(1-\frac{\delta_0}{2}\right)\left(\frac{1}{\nu_j^{\frac{\delta_0}{2}}}+\frac{|\beta_j|^{\delta_0}}{(2\sigma^2\tau^2)^{\frac{\delta_0}{2}}}\right).\end{aligned}$$

For $j=1,2,\cdots,p$, we denote

$$\mu_j=\boldsymbol{e}_j^TA_0^{-1}\mathbf{X}^T\mathbf{y}\tag{2.14}$$

where $A_0=\mathbf{X}^T\mathbf{X}+(\tau^2\boldsymbol{\Lambda}_0)^{-1}$.

It follows that

$$\begin{aligned}&\boldsymbol{E}\left[\boldsymbol{E}\left[\left(\lambda_j^2\right)^{\frac{\delta_0}{2}}\middle|\boldsymbol{\beta},\boldsymbol{\nu},\sigma^2,\tau^2,\mathbf{y}\right]\middle|\sigma^2,\tau^2,\boldsymbol{\lambda}_0,\mathbf{y}\right]\\&\leq\Gamma\left(1-\frac{\delta_0}{2}\right)\boldsymbol{E}\left[\left(\frac{1}{\nu_j^{\frac{\delta_0}{2}}}+\frac{|\beta_j|^{\delta_0}}{(2\sigma^2\tau^2)^{\frac{\delta_0}{2}}}\right)\middle|\sigma^2,\tau^2,\boldsymbol{\lambda}_0,\mathbf{y}\right]\\&=\Gamma\left(1-\frac{\delta_0}{2}\right)\left(\boldsymbol{E}\left[\frac{1}{\nu_j^{\frac{\delta_0}{2}}}\middle|\sigma^2,\tau^2,\boldsymbol{\lambda}_0,\mathbf{y}\right]+\boldsymbol{E}\left[\frac{|\beta_j|^{\delta_0}}{(2\sigma^2\tau^2)^{\frac{\delta_0}{2}}}\middle|\sigma^2,\tau^2,\boldsymbol{\lambda}_0,\mathbf{y}\right]\right)\\&\leq\Gamma\left(1-\frac{\delta_0}{2}\right)\left(\Gamma\left(1+\frac{\delta_0}{2}\right)+\boldsymbol{E}\left[\frac{|\beta_j-\mu_j|^{\delta_0}}{(2\sigma^2\tau^2)^{\frac{\delta_0}{2}}}\middle|\sigma^2,\tau^2,\boldsymbol{\lambda}_0,\mathbf{y}\right]+\frac{|\mu_j|^{\delta_0}}{(2\sigma^2\tau^2)^{\frac{\delta_0}{2}}}\right);\\&\leq\Gamma\left(1-\frac{\delta_0}{2}\right)\left(\Gamma\left(1+\frac{\delta_0}{2}\right)+\frac{\Gamma\left(\frac{1+\delta_0}{2}\right)}{\sqrt{\pi}}\lambda_{j;0}^{\delta_0}+\frac{|\mu_j|^{\delta_0}}{(2\sigma^2\tau^2)^{\frac{\delta_0}{2}}}\right)\\&\overset{(\bigstar\bigstar)}{\leq}\Gamma\left(1-\frac{\delta_0}{2}\right)\left(\Gamma\left(1+\frac{\delta_0}{2}\right)+\frac{\Gamma\left(\frac{1+\delta_0}{2}\right)}{\sqrt{\pi}}\lambda_{j;0}^{\delta_0}+\frac{T^*}{(2\sigma^2)^{\frac{\delta_0}{2}}}\right),\end{aligned}$$

for some $T^*>0$. Here $(\bigstar\bigstar)$ follows from Proposition A.5 (see Appendix A) and the fact that τ^2 is supported on $[T,\infty)$. Hence,

$$\boldsymbol{E}\left[\left(\lambda_j^2\right)^{\frac{\delta_0}{2}}\middle|\boldsymbol{\lambda}_0\right]=\boldsymbol{E}\left[\boldsymbol{E}\left[\boldsymbol{E}\left[\left(\lambda_j^2\right)^{\frac{\delta_0}{2}}\middle|\boldsymbol{\beta},\boldsymbol{\nu},\sigma^2,\tau^2,\mathbf{y}\right]\middle|\sigma^2,\tau^2,\boldsymbol{\lambda}_0,\mathbf{y}\right]\middle|\boldsymbol{\lambda}_0,\mathbf{y}\right]$$

$$\begin{aligned}
&\leq \Gamma\left(1-\frac{\delta_0}{2}\right)\left(\Gamma\left(1+\frac{\delta_0}{2}\right)+\frac{\Gamma\left(\frac{1+\delta_0}{2}\right)}{\sqrt{\pi}}\lambda_{j;0}^{\delta_0}+\boldsymbol{E}\left[\frac{T^*}{\left(2\sigma^2\right)^{\frac{\delta_0}{2}}}\middle|\boldsymbol{\lambda}_0,\mathbf{y}\right]\right)\\
&=\Gamma\left(1-\frac{\delta_0}{2}\right)\times\\
&\quad\left(\Gamma\left(1+\frac{\delta_0}{2}\right)+\frac{\Gamma\left(\frac{1+\delta_0}{2}\right)}{\sqrt{\pi}}\lambda_{j;0}^{\delta_0}+\boldsymbol{E}\left[\boldsymbol{E}\left[\frac{T^*}{\left(2\sigma^2\right)^{\frac{\delta_0}{2}}}\middle|\tau^2,\boldsymbol{\lambda}_0,\mathbf{y}\right]\middle|\boldsymbol{\lambda}_0,\mathbf{y}\right]\right)\\
&\leq \Gamma\left(1-\frac{\delta_0}{2}\right)\left(\Gamma\left(1+\frac{\delta_0}{2}\right)+\frac{\Gamma\left(\frac{1+\delta_0}{2}\right)}{\sqrt{\pi}}\lambda_{j;0}^{\delta_0}+\frac{T^*}{(2b)^{\frac{\delta_0}{2}}}\cdot\frac{\Gamma\left(a+\frac{n+\delta_0}{2}\right)}{\Gamma\left(a+\frac{n}{2}\right)}\right)
\end{aligned}$$

It follows that

$$\boldsymbol{E}\left[\sum_{j=1}^{p}\left(\lambda_j^2\right)^{\frac{\delta_0}{2}}\middle|\boldsymbol{\lambda}_0\right]\leq\gamma\left(\delta_0\right)\sum_{j=1}^{p}\left(\lambda_{j;0}^2\right)^{\frac{\delta_0}{2}}+b_2 \tag{2.15}$$

where

$$\gamma\left(\delta_0\right)=\Gamma\left(1-\frac{\delta_0}{2}\right)\frac{\Gamma\left(\frac{1+\delta_0}{2}\right)}{\sqrt{\pi}}$$

and

$$b_2=p\cdot\Gamma\left(1-\frac{\delta_0}{2}\right)\Gamma\left(1+\frac{\delta_0}{2}\right)\frac{T^*}{(2b)^{\frac{\delta_0}{2}}}\cdot\frac{\Gamma\left(a+\frac{n+\delta_0}{2}\right)}{\Gamma\left(a+\frac{n}{2}\right)}.$$

The result follows by combining (2.13) and (2.15) with

$$\gamma^*=\max\left\{\gamma\left(\delta_0\right),\gamma\left(\delta_1\right)\right\}$$

and

$$b^*=b_1+b_2.$$

Note that $\gamma^*=\max\left\{\gamma\left(\delta_0\right),\gamma\left(\delta_1\right)\right\}<1$ for small enough choices of δ_0 and δ_1, for example $\delta_0,\delta_1\in(0.00162,0.22176)$. ☐

Remark 2.1. *Note that the only place in the proof of Lemma 2.1 where we need τ^2 to be truncated below is to show that $\mathbf{E}\left[\left(\tau^2\right)^{-\frac{\delta_0}{2}}\mid\boldsymbol{\lambda}_0,\mathbf{y}\right]$ is uniformly bounded in $\boldsymbol{\lambda}_0$. In Proposition B.2, we show this follows by assuming the weaker condition that the prior negative $(p+\delta_0)/2^{th}$ moment for τ^2 is finite.*

We now explain why the geometric drift condition established in Theorem 2.1 for the marginal $\boldsymbol{\lambda}$-chain implies geometric ergodicity of the two-block Horseshoe Gibbs sampler K_{aug}. Note that for every $d\in\mathbb{R}$, the set

$$B\left(V,d\right)=\left\{\boldsymbol{\lambda}\in\mathbb{R}_+^p:V(\boldsymbol{\lambda})=\sum_{j=1}^{p}\left(\lambda_j^2\right)^{\frac{\delta_0}{2}}+\sum_{j=1}^{p}\left(\lambda_j^2\right)^{-\frac{\delta_1}{2}}\leq d\right\}$$

is a compact set. Since $k(\boldsymbol{\lambda}_0, \boldsymbol{\lambda})$ is continuous in $\boldsymbol{\lambda}_0$, a standard argument using Fatou's lemma along with Theorem 6.0.1 of [14] can be used to establish that the marginal $\boldsymbol{\lambda}$-chain is *unbounded off petite sets*. Lemma 15.2.8 of [14] then implies geometric ergodicity of the marginal λ-chain. Using Lemma 2.4 in [5] now gives the following result.

Theorem 2.1. *Suppose the prior density π_τ for the global shrinkage parameter is truncated below i.e., $\pi_\tau(u) = 0$ for $u < T$ for some $T > 0$ and satisfies*

$$\int_0^\infty u^{\delta/2} \pi_\tau(u)\, du < \infty$$

for some $\delta \in (0.00162, 0.22176)$. Then the two-block Horseshoe Gibbs sampler with transition kernel K_{aug} is geometrically ergodic. The assumption of truncation below (i.e., $T > 0$) can be replaced by the weaker assumption that $T = 0$ and that the prior negative $(p+\delta)/2^{th}$ moment for τ^2 is finite for some $\delta > 0.00162$.

Note that the above result establishes geometric ergodicity, which as described earlier, helps rigorously establish the asymptotic validity of Markov chain CLT based standard error estimates. However, if quantitative bounds on the distance to stationarity are needed, then an additional *minorization condition* needs to be established. For the sake of completeness, we derive such a condition in Appendix C (see Lemma C.1).

Note that for K_{aug}, an exact sample from the conditional posterior density of τ^2 given $\boldsymbol{\lambda}$ needs to be drawn. However, $\pi(\tau^2 \mid \boldsymbol{\lambda}, \mathbf{y})$ is not a standard density. Since we are looking at just a univariate draw, and it can be easily shown using (2.2) that $\pi(\tau^2 \mid \boldsymbol{\lambda}, \mathbf{y}) \leq C\pi_\tau(\tau^2)$, one straightforward option is to use a rejection sampler. In high-dimensional settings, however, the rejection probability might become too high. Another alternative is to use a simple discretization based sample by transforming to a bounded range such as $[0, 1]$ (for example by using $\tau^2 \to \frac{\tau^2}{\tau^2+1}$), but this again might become computationally expensive.

In their experiments, [7] consider neither of the above two alternatives for the conditional sampling of τ^2. Instead, they use a version where a Metropolis sampler is used for the sampling of τ^2. The geometric ergodicity of their JOB sampler (using this Metropolis update), however, is established under the restrictive assumption that *each* element of $\boldsymbol{\lambda}$ is bounded away from zero in addition to τ^2 being bounded above and below. More recently in [3], geometric ergodicity for the JOB sampler is established without the lower bound constraint on elements of $\boldsymbol{\lambda}$, but with *exact sampling of τ^2 instead of the Metropolis step* (and τ^2 bounded both above and below). Since changing from an exact to a Metropolis draw for τ^2 changes the structure of the underlying transition density, there is a gap between the geometric ergodicity results and the numerical illustrations.

In our high-dimensional numerical illustrations (Section 2.3) we also use the Metropolis update for τ^2 for a direct comparison with the approach in [7] and for computational simplicity. Hence, it is important to understand if the variant of K_{aug} with a Metropolis update for τ^2 is also geometrically ergodic.

In particular, consider the two-block Metropolis-within-Gibbs chain with transition kernel $K_{MG,q}$ with proposal density q whose transition from $(\boldsymbol{\beta}_0, \sigma_0^2, \boldsymbol{\nu}_0, \tau_0^2, \boldsymbol{\lambda}_0)$ to $(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\nu}, \tau^2, \boldsymbol{\lambda})$ is given as follows.

1. Draw $\boldsymbol{\lambda}$ from $\pi(\boldsymbol{\lambda} \mid \boldsymbol{\beta}_0, \boldsymbol{\nu}_0, \sigma_0^2, \tau_0^2, \mathbf{y})$. This can be done by independently drawing the components of $\boldsymbol{\lambda}$ from the appropriate full conditional posterior density in (2.2).
2.
   - (Metropolis draw for τ^2) Draw ${\tau^2}' \sim q(\cdot \mid \tau_0^2)$ and $U \sim \text{Uniform}[0,1]$ independently. If
   $$U \leq \frac{\pi({\tau^2}' \mid \boldsymbol{\lambda}, \mathbf{y}) q(\tau_0^2 \mid {\tau^2}')}{\pi(\tau_0^2 \mid \boldsymbol{\lambda}, \mathbf{y}) q({\tau^2}' \mid \tau_0^2)}$$
   set $\tau^2 = {\tau^2}'$, else set $\tau^2 = \tau_0^2$.
   - Draw $(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\nu})$ from $\pi(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\nu}, \mid \tau^2, \boldsymbol{\lambda}_0, \mathbf{y})$. This can be done by sequentially drawing $\boldsymbol{\nu}$, then σ^2, and then $\boldsymbol{\beta}$ from appropriate conditional posterior densities in (2.2).

The next theorem first establishes a condition on the Metropolis proposal density q under which the Metropolis-within-Gibbs chain with kernel $K_{MG,q}$ is geometrically ergodic. The next part of the theorem shows this condition is satisfied for three natural choices of the proposal density, including the one used by [7] in their experiments.

Theorem 2.2. *Assume that the conditions in Theorem 2.1 hold, i.e.,* $\pi_\tau(u) = 0$ *for* $u < T$ *for some* $T > 0$ *and satisfies*

$$\int_0^\infty u^{\delta/2} \pi_\tau(u)\, du < \infty$$

for some $\delta \in (0.00162, 0.22176)$.

(a) The Metropolis-within-Gibbs sampler with transition kernel $K_{MG,q}$ *is geometrically ergodic if*

$$\sup_{\tau_{curr}^2, {\tau^2}' \in [T,\infty)} \frac{\pi_\tau({\tau^2}')}{q({\tau^2}' \mid \tau_{curr}^2)} < \infty.$$

(b) The condition in part (a) is satisfied if

(i) $q({\tau^2}' \mid \tau^2) = \pi_\tau({\tau^2}')$ *(independence Metropolis).*

(ii) Random walk Metropolis for $\xi = \frac{1}{\tau^2}$, *where the proposal* $\xi' = 1/{\tau^2}'$ *is drawn from* $N(\xi_{curr}, v^2)$. *Here* $\xi_{curr} = 1/\tau_{curr}^2$, τ_{curr}^2 *is the current value of* τ^2, v^2 *is an arbitrary positive constant, and* π_ξ *(prior density for* $\xi = 1/\tau^2$*) is chosen to be the* $N(0, v^2)$ *density truncated to* $(0, 1/T]$ *(so that* π_τ *is truncated to* $[T, \infty)$*).*

(iii) Random walk Metropolis for $\zeta = \log(\tau^2)$, *where the proposal* $\zeta' = \log({\tau^2}')$ *is drawn from the* $N\left(\zeta_{curr}^*, v^2\right)$ *distribution (where* $\zeta_{curr}^* =$

$\log\left((\tau^2_{curr} \vee 1) \wedge (1/T)\right)$). *Here* τ^2_{curr} *is the current value of* τ^2, v^2 *is an arbitrary positive constant, and* π_τ *is chosen to be the Lognormal*(0, v^2) *density (truncated to* $[T, \infty)$*).*

Proof:

(a) We start by adapting the arguments in [9] to the current more general setting, and showing that geometric ergodicity of $K_{MG,q}$ follows if an (intermediate) supremum is shown to be finite. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2, \boldsymbol{\nu}, \tau^2)$ and Θ denote the corresponding parameter space. Consider the marginal transition kernel $K^{\boldsymbol{\theta}}_{MG,q}$ whose transition from $(\boldsymbol{\beta}_0, \sigma^2_0, \boldsymbol{\nu}_0, \tau^2_0)$ to $(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\nu}, \tau^2)$ is obtained by drawing $\boldsymbol{\lambda}$ as in Step 1 of the $K_{MG,q}$ transition, and then drawing $(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\nu}, \tau^2)$ as in Step 2 of the $K_{MG,q}$ transition. It follows that $K^{\boldsymbol{\theta}}_{MG,q}$ is de-initializing for $K_{MG,q}$ (see [20]). Hence, it is enough to show that $K^{\boldsymbol{\theta}}_{MG,q}$ is geometrically ergodic. It follows from the definition of $K_{MG,q}$ (in particular the Metropolis step for τ^2) that

$$K^{\boldsymbol{\theta}}_{MG,q}(\boldsymbol{\theta}_0, A) \geq \int_{\Theta} h(\boldsymbol{\theta}_0, \boldsymbol{\theta}) d\boldsymbol{\theta}$$

where

$$h(\boldsymbol{\theta}_0, \boldsymbol{\theta}) = \int_{\mathbb{R}^p_+} \pi(\boldsymbol{\lambda} \mid \boldsymbol{\theta}_0, \mathbf{y}) \min\left(\frac{\pi(\tau^2 \mid \boldsymbol{\lambda}, \mathbf{y}) q(\tau^2_0 \mid \tau^2)}{\pi(\tau^2_0 \mid \boldsymbol{\lambda}, \mathbf{y}) q(\tau^2 \mid \tau^2_0)}, 1\right) q(\tau^2 \mid \tau^2_0) \times \pi(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\nu} \mid \boldsymbol{\lambda}, \tau^2, \mathbf{y}) d\boldsymbol{\lambda}.$$

Denote $w\left(\tau^2, \tau^2_0, \boldsymbol{\lambda}\right) = \frac{\pi(\tau^2 \mid \boldsymbol{\lambda}, \mathbf{y})}{q(\tau^2 \mid \tau^2_0)}$. Using the definition of h, we get that

$$\begin{aligned}
& h(\boldsymbol{\theta}_0, \boldsymbol{\theta}) \\
= \; & \int_{\mathbb{R}^p_+} \pi(\boldsymbol{\lambda} \mid \boldsymbol{\theta}_0, \mathbf{y}) \min\left(\frac{1}{w\left({\tau^2}_0, \tau^2, \boldsymbol{\lambda}\right)}, \frac{1}{w\left(\tau^2, {\tau^2}_0, \boldsymbol{\lambda}\right)}\right) \\
& \times \pi(\tau^2 \mid \boldsymbol{\lambda}, \mathbf{y}) \pi(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\nu} \mid \boldsymbol{\lambda}, \tau^2, \mathbf{y}) d\boldsymbol{\lambda} \\
\geq \; & \frac{1}{\sup\limits_{\tau^2_0, \tau^2 \in [T, \infty), \boldsymbol{\lambda} \in \mathbb{R}^p_+} w\left(\tau^2, \tau^2_0, \boldsymbol{\lambda}\right)} \times \\
& \int_{\mathbb{R}^p_+} \pi(\boldsymbol{\lambda} \mid \boldsymbol{\theta}_0, \mathbf{y}) \pi(\tau^2 \mid \boldsymbol{\lambda}, \mathbf{y}) \pi(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\nu} \mid \boldsymbol{\lambda}, \tau^2, \mathbf{y}) d\boldsymbol{\lambda}
\end{aligned}$$

for all $\boldsymbol{\theta}_0, \boldsymbol{\theta} \in \Theta$. The second factor in the last equation is the transition density of the $\boldsymbol{\theta}$-marginal kernel for K_{aug} (which is de-initializing for K_{aug}). By Theorem 2.1, it follows that the $\boldsymbol{\theta}$-marginal kernel for K_{aug} is geometrically ergodic. Finally, since the $\boldsymbol{\theta}$-marginal kernels for K_{aug} and $K_{MG,q}$ are both reversible, by Theorem 1 in [9], it is enough to show

$$\sup_{\tau^2, \tau^2_0 \in [T, \infty), \boldsymbol{\lambda} \in \mathbb{R}^p_+} w\left(\tau^2, \tau^2_0, \boldsymbol{\lambda}\right) < \infty$$

to establish geometric ergodicity of $K_{MG,q}$. Note that for arbitrary $\tau^2_{curr}, \tau^{2'} \geq T$ and $\boldsymbol{\lambda} \in \mathbb{R}^p_+$, we have

$$
\begin{aligned}
&\frac{q(\tau^{2'} \mid \tau^2_{curr}) w\left(\tau^{2'}, \tau^2_{curr}, \boldsymbol{\lambda}\right)}{\pi_\tau(\tau^{2'})} \\
&\leq \left(\frac{b}{\mathbf{y}^T\mathbf{y}+b}\right)^{-\left(a+\frac{n}{2}\right)} \frac{\frac{1}{\sqrt{|I_p+\mathbf{X}^T\mathbf{X}\cdot\tau^{2'}\boldsymbol{\Lambda}|}} I_{[\tau^{2'}\geq T]}}{\int\limits_T^\infty \frac{1}{\sqrt{|I_p+\mathbf{X}^T\mathbf{X}\cdot\tau^2\boldsymbol{\Lambda}|}}\pi_\tau\left(\tau^2\right) d\tau^2} \\
&\leq \left(\frac{b}{\mathbf{y}^T\mathbf{y}+b}\right)^{-\left(a+\frac{n}{2}\right)} \frac{\frac{1}{\sqrt{|\boldsymbol{\Lambda}^{-1}+\tau^{2'}\cdot\mathbf{X}^T\mathbf{X}|}} I_{[\tau^{2'}\geq T]}}{\int\limits_T^{T'} \frac{1}{\sqrt{|\boldsymbol{\Lambda}^{-1}+\tau^2\cdot\mathbf{X}^T\mathbf{X}|}}\pi_\tau\left(\tau^2\right) d\tau^2}; \text{for any } T' \in (T, \infty) \\
&\leq \left(\frac{b}{\mathbf{y}^T\mathbf{y}+b}\right)^{-\left(a+\frac{n}{2}\right)} \frac{\frac{1}{\sqrt{|\boldsymbol{\Lambda}^{-1}+T\cdot\mathbf{X}^T\mathbf{X}|}}}{\int\limits_T^{T'} \frac{1}{\sqrt{|\boldsymbol{\Lambda}^{-1}+T'\cdot\mathbf{X}^T\mathbf{X}|}}\pi_\tau\left(\tau^2\right) d\tau^2} \\
&= \tilde{K}\sqrt{\frac{|\boldsymbol{\Lambda}^{-1}+T'\cdot\mathbf{X}^T\mathbf{X}|}{|\boldsymbol{\Lambda}^{-1}+T\cdot\mathbf{X}^T\mathbf{X}|}}; \ \tilde{K} = \frac{\left(\frac{b}{\mathbf{y}^T\mathbf{y}+b}\right)^{-\left(a+\frac{n}{2}\right)}}{\int\limits_T^{T'} \pi_\tau(u)du} \\
&= \tilde{K}\sqrt{\frac{|I_n+T'\cdot\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^T|}{|I_n+T\cdot\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^T|}}; \text{ by the Matrix determinant formula} \\
&= \tilde{K}\sqrt{\prod_{j=1}^n\left(\frac{1+T'\cdot w_j}{1+T\cdot w_j}\right)}; \ w_j's \text{ are eigenvalues of } \mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^T \\
&\leq \tilde{K}\sqrt{\left(\frac{T'}{T}\right)^n} < \infty;
\end{aligned}
$$

since the function $w \mapsto \frac{1+T'\cdot w}{1+T\cdot w}$ is increasing on $(0,\infty)$ for $T < T'$ and $\lim_{w\to\infty}\frac{1+T'\cdot w}{1+T\cdot w} = \frac{T'}{T}$. This establishes part (a) of the result.

(b) (i) For independent Metropolis the ratio in (a) becomes 1 and the condition is trivially satisfied.

(ii) In this setting, straightforward calculations show that

$$q(\tau^{2'} \mid \tau^2_{curr}) = \frac{1}{(\tau^{2'})^2\sqrt{2\pi v^2}}\exp\left(-\frac{1}{2v^2}\left(\frac{1}{(\tau^{2'})^2}-\frac{1}{({\tau^2}_{curr})^2}\right)^2\right)$$

and

$$\pi_\tau(\tau^{2'}) = \frac{1}{(\tau^{2'})^2\sqrt{2\pi v^2}}\exp\left(-\frac{1}{2v^2}\left(\frac{1}{(\tau^{2'})^2}\right)^2\right)T^*1_{\{\tau^{2'}\geq T\}}$$

where $T^* = (\Phi(1/T) - 1/2)^{-1}$ and Φ is the standard normal cdf. It follows that

$$\begin{aligned}
& \sup_{\tau^2_{curr}, \tau^{2\prime} \in [T,\infty)} \frac{\pi_\tau(\tau^{2\prime})}{q(\tau^{2\prime} \mid \tau^2_{curr})} \\
= \;& T^* \sup_{\tau^2_{curr}, \tau^{2\prime} \in [T,\infty)} \exp\left(\frac{1}{2v^2(\tau^2_{curr})^2} - \frac{1}{v^2 \tau^{2\prime} \tau^2_{curr}} \right) \\
\leq \;& T^* \exp\left(\frac{1}{2v^2 T^2} \right) < \infty.
\end{aligned}$$

(iii) In this setting, straightforward calculations show that

$$q(\tau^{2\prime} \mid \tau^2_{curr}) = \frac{1}{\tau^{2\prime}\sqrt{2\pi v^2}} \exp\left(- \frac{(\log(\tau^{2\prime}) - \log\left((\tau^2_{curr} \vee 1) \wedge T\right))^2}{2v^2} \right)$$

and

$$\pi_\tau(\tau^{2\prime}) = \frac{1}{\tau^{2\prime}\sqrt{2\pi v^2}} \exp\left(- \frac{(\log(\tau^{2\prime}))^2}{2v^2} \right) T^{**} 1_{\{\tau^{2\prime} \geq T\}};$$

where $T^{**} = (1 - \Phi(\log T))^{-1}$. It follows that

$$\begin{aligned}
& \sup_{\tau^2_{curr}, \tau^{2\prime} \in [T,\infty)} \frac{\pi_\tau(\tau^{2\prime})}{q(\tau^{2\prime} \mid \tau^2_{curr})} \\
= \;& T^{**} \sup_{\tau^2_{curr}, \tau^{2\prime} \in [T,\infty)} \exp\left(\frac{(\log\left((\tau^2_{curr} \vee 1) \wedge (1/T)\right))^2}{2v^2} \right) \times \\
& \exp\left(- \frac{\log(\tau^{2\prime}) \log\left((\tau^2_{curr} \vee 1) \wedge (1/T)\right)}{v^2} \right) \\
\leq \;& T^{**} \exp\left(\frac{3(\log T)^2}{2v^2} \right) < \infty. \quad \square
\end{aligned}$$

2.3. A simulation study

The objective of this study is to examine the practical feasibility/scalability of the Gibbs sampler described and analyzed in Sections 2.1 and 2.2 by comparing its computational performance with the JOB Gibbs sampler. Following [7], we will use the Metropolis-within-Gibbs version of both chains. We consider two simulation settings for our numerical illustration/study. In Setting 1, we use a random walk Metropolis step (as specified in Theorem 2.2 part (b) (iii)) for $\log(\tau^2)$ for both the proposed sampler and the JOB sampler. The sample size n is set to be 100 and the number of predictors p is set to be 500. In Setting 2, we use an independence Metropolis step (as specified in Theorem 2.2 part (b) (i)) for τ^2 for both the proposed sampler and the JOB sampler. The sample size n is set to be 100 and the number of predictors p is set to be 750.

For both settings, the first 10 entries of the "true" regression coefficient vector $\boldsymbol{\beta}^0 := (\beta_1^0, \ldots, \beta_p^0)$ are specified as $\beta_j^0 = 2^{s_j}$ where $s_j's$ are a sequence of equally spaced values in the interval $(-1, 3)$, and the other entries are set to zero. The entries of the design matrix $\mathbf{X}$ are generated independently from $\mathcal{N}(0, 1)$. Then, we generate the response vector $\mathbf{y}$ from the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}^0 + \boldsymbol{\epsilon}$ where the error vector $\boldsymbol{\epsilon}$ has i.i.d. normal entries with mean 0 and standard deviation 0.1.

We generate 10 data sets from each of the above simulation settings (so 20 datasets in all), and run both the Gibbs samplers on each of these 10 data sets. For a fair comparison, both algorithms were implemented in *R*. The simulations were run on a machine with a 64 bit macOS Catalina operating system, 8 GB RAM and a 1.6 GHz processor. Note that from a computational point of view, the only difference between the proposed Gibbs sampler and the JOB sampler is that the former uses $2p$ draws from a standard (Gamma) distribution (for $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}$), whereas the latter uses p rejection sampling based draws from a non-standard distribution (for $\boldsymbol{\lambda}$). Hence, one would expect the proposed Gibbs sampler to be faster than the JOB sampler (per iteration). For an R implementation in particular, the $2p$ Gamma draws can be performed really efficiently using just two *rgamma* commands (one for the p components of $\boldsymbol{\lambda}$ and another for the p components of $\boldsymbol{\nu}$). This is indeed borne out by our simulation results. To complete 6000 iterations (for Setting 1), the proposed sampler needed 541 seconds wall clock time on average (over the 10 data sets), while the JOB sampler needed 721 seconds wall clock time on average. For Setting 2, the average wall clock time for 6000 iterations were roughly 810 seconds (proposed sampler) and 1025 seconds (JOB sampler).

The faster speed of the proposed sampler, however, might come at the cost of higher autocorrelation due to the introduction of the augmented parameter $\boldsymbol{\nu}$. Hence, for a balanced comparison we looked at the performance of these two algorithms (in terms of the essential sample size and MCMC standard error) under the *same computational budget*. For Setting 1, both algorithms are evaluated over a wall clock time window of 540 seconds each for all 10 data sets. In the allotted time, the proposed Gibbs sampler was able to complete roughly 6000 iterations, and the JOB sampler is able to complete roughly 4600 iterations for each of the 10 datasets. In the given time window, the essential sample sizes and MCMC standard errors for the first ten components of $\boldsymbol{\beta}$ (which are the only non-zero components in the true $\boldsymbol{\beta}_0$) for both algorithms are provided in Table 1 and Table 2 respectively. For Setting 2, both algorithms are alloted a wall clock time of 810 seconds, and corresponding essential sample size and MCMC standard error values are provided in Table 3 and Table 4 respectively.

Note that both samplers aim to generate samples from the same posterior distribution. It can be seen that in a large majority of the situations, the proposed sampler leads to a higher effective sample size and lower standard error than the JOB sampler with the same computational budget. To conclude, the proposed sampler has asymptotic convergence/standard error guarantees under

Table 1
EFFECTIVE SAMPLE SIZE COMPARISON

	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
Dataset 1 (Prop.)	423	763	605	545	502	1078	280	255	73	473
Dataset 1 (JOB)	298	626	515	204	525	521	223	205	121	674
Dataset 2 (Prop.)	1067	1093	954	323	677	737	143	1178	151	817
Dataset 2 (JOB)	738	682	826	795	708	943	124	1041	175	1258
Dataset 3 (Prop.)	467	398	209	713	224	324	164	105	176	956
Dataset 3 (JOB)	336	167	233	376	217	224	159	116	220	482
Dataset 4 (Prop.)	989	1119	1323	1195	1008	1371	343	950	189	1071
Dataset 4 (JOB)	1012	748	790	801	473	740	678	799	186	967
Dataset 5 (Prop.)	261	175	618	652	402	581	312	119	188	218
Dataset 5 (JOB)	243	234	413	346	356	364	217	114	123	256
Dataset 6 (Prop.)	2101	986	259	565	786	606	859	129	2364	1928
Dataset 6 (JOB)	786	935	386	555	456	530	448	186	1637	1528
Dataset 7 (Prop.)	648	561	820	479	725	514	651	153	906	137
Dataset 7 (JOB)	459	531	564	448	358	486	171	147	966	138
Dataset 8 (Prop.)	861	382	662	1049	691	753	812	532	575	1193
Dataset 8 (JOB)	578	455	347	334	507	427	491	301	82	1155
Dataset 9 (Prop.)	533	384	494	290	413	853	670	609	481	1108
Dataset 9 (JOB)	445	244	502	320	224	823	410	788	393	1002
Dataset 10 (Prop.)	203	818	613	611	324	712	651	621	425	1169
Dataset 10 (JOB)	260	373	650	396	250	566	302	537	116	637

Effective sample size for the first 10 regression coefficients for the proposed sampler (Prop.) and the JOB sampler (JOB) for 10 simulated datasets with $n = 100$ and $p = 500$. Both algorithms are provided the same computational budget of 540 seconds (wall clock time) each. All datasets are simulated from a 'true' linear regression model where only the first 10 regression coefficients are non-zero, and the entries of the design matrix X and the error vector are all independent Gaussian (both the design matrix and the error vector are generated separately for each dataset).

weaker conditions than the JOB sampler, and at the same time can provide computationally competitive or superior performance as compared to the JOB sampler.

3. Geometric ergodicity for regularized Horseshoe Gibbs samplers

3.1. A Gibbs sampler for the regularized Horseshoe

Recall from the introduction that the regularized Horseshoe prior developed in [18] is given by

$$\begin{aligned} \beta_i \mid \lambda_i^2, \sigma^2, \tau^2 \sim \mathcal{N}_p\left(0, \left(\frac{1}{c^2} + \frac{1}{\lambda_i^2\tau^2}\right)^{-1} \sigma^2\right) &\text{independently for } i = 1, 2, \cdots, p \\ \lambda_i \sim C^+(0,1) &\text{independently for } i = 1, 2, \cdots, p \\ \tau^2 \sim \pi_\tau(\cdot) \qquad \sigma^2 \sim \text{Inverse-Gamma}(a, b) \end{aligned} \tag{3.1}$$

The only difference between this prior and the original Horseshoe prior in (1.1) is the additional regularization introduced in the prior conditional variance of

TABLE 2
MCMC STANDARD ERROR COMPARISON

	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
Dataset 1 (Prop.)	.012	.0083	.0094	.0125	.0103	.0067	.0180	.0147	.0482	.0047
Dataset 1 (JOB)	.014	.009	.0101	.0211	.0098	.0098	.0217	.0185	.0363	.0036
Dataset 2 (Prop.)	.0072	.0073	.0083	.0175	.0101	.0110	.0297	.0074	.0164	.0024
Dataset 2 (JOB)	.0082	.0093	.0089	.0106	.0094	.0094	.0327	.0079	.0167	.0017
Dataset 3 (Prop.)	.0106	.0137	.0181	.0097	.0166	.0173	.0207	.0340	.0208	.0033
Dataset 3 (JOB)	.0121	.0208	.0171	.0131	.0167	.02	.0205	.0313	.0173	0.0053
Dataset 4 (Prop.)	.0068	.0063	.0064	.0076	.0066	.0060	.0154	.0073	.0197	.0023
Dataset 4 (JOB)	.0066	.0077	.0085	.0094	.0103	.0084	.0101	.0079	.0192	.0024
Dataset 5 (Prop.)	.0174	.0198	.0104	.0099	.0131	.0104	.015	.0316	.0234	.0096
Dataset 5 (JOB)	.0183	.0167	.0128	.0135	.0140	.0137	.0187	.0317	.0270	.0093
Dataset 6 (Prop.)	.005	.008	.0176	.0113	.0092	.0108	.003	.0298	.0013	0.0014
Dataset 6 (JOB)	.0083	.0082	.0141	.0110	.0119	.0116	.004	.0223	.0015	.0015
Dataset 7 (Prop.)	.0087	.0099	.008	.0113	.0079	.0124	.0105	.0256	.0027	.0246
Dataset 7 (JOB)	.0103	.0102	.0096	.0118	.0112	.0126	.0210	.0265	.0023	.0240
Dataset 8 (Prop.)	.0066	.0101	.0085	.0058	.0073	.0087	.0072	.0101	.0106	.0016
Dataset 8 (JOB)	.0084	.0094	.0121	.0106	.0088	.0115	.0096	.0141	.0352	.0019
Dataset 9 (Prop.)	.0100	.0124	.0101	.0150	.0122	.0092	.0090	.0102	.0041	.0024
Dataset 9 (JOB)	.0113	.0156	.0101	.0141	.0167	.0094	.0112	.0086	.0056	.0022
Dataset 10 (Prop.)	.0152	.0071	.0075	.0075	.0109	.006	.0073	.0073	.0103	.0021
Dataset 10 (JOB)	.0137	.0106	.0071	.0092	.0126	.0066	.0108	.0080	.0214	.0029

MCMC standard error for the first 10 regression coefficients for the proposed sampler (Prop.) and the JOB sampler (JOB) for 10 simulated datasets with $n = 100$ and $p = 500$. Both algorithms are provided the same computational budget of 540 seconds (wall clock time) each. All datasets are simulated from a 'true' linear regression model where only the first 10 regression coefficients are non-zero, and the entries of the design matrix X and the error vector are all independent Gaussian (both the design matrix and the error vector are generated separately for each dataset).

the β_is through the constant c. As $c \to \infty$ in (3.1), then one reverts back to the original Horseshoe specification in (1.1).

Note that one of the salient features of the Horseshoe prior is the lack of shrinkage/regularization of parameter values that are far away from zero. The authors in [18] argue that while this feature is one of the key strengths of the Horseshoe prior in many situations, it can be a drawback in settings where the parameters are weakly identified. We refer the reader to [18] for a thorough motivation and discussion of the properties and performance of this prior vis-a-vis the Horseshoe prior. Our focus in this paper is to look at Markov chains to sample from the resulting intractable regularized Horseshoe posterior, and investigate properties such as geometric ergodicity.

The authors in [18] use Hamiltonian Monte Carlo (HMC) to generate samples from the posterior distribution. Geometric ergodicity of this HMC chain, however, is not established. In recent work [11], sufficient conditions for geometric ergodicity (or lack thereof) for general HMC chains have been provided. However, these conditions, namely Assumptions A1, A2, A3 in [11], are rather complex and intricate, and at least to the best of our understanding it is unclear and hard to verify if these conditions are satisfied by the HMC chain in [18].

Given the host of Gibbs samplers available in the literature for the original

TABLE 3
EFFECTIVE SAMPLE SIZE COMPARISON

	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
Dataset 1 (Prop.)	727	521	164	116	329	729	95	1567	1131	2358
Dataset 1 (JOB)	544	418	693	662	202	467	158	926	590	2063
Dataset 2 (Prop.)	967	470	639	233	150	160	343	218	2538	2316
Dataset 2 (JOB)	267	293	285	240	115	96	728	126	1784	1143
Dataset 3 (Prop.)	1091	149	4043	783	370	429	487	145	3235	538
Dataset 3 (JOB)	347	487	525	507	402	376	634	99	2683	345
Dataset 4 (Prop.)	217	241	160	266	104	146	37	398	462	2400
Dataset 4 (JOB)	204	170	255	321	153	90	66	490	2295	2074
Dataset 5 (Prop.)	933	758	390	185	368	120	88	1787	660	639
Dataset 5 (JOB)	587	207	411	387	92	150	103	1198	571	393
Dataset 6 (Prop.)	316	645	591	265	197	150	462	2876	1121	1294
Dataset 6 (JOB)	204	440	264	160	210	138	286	694	613	1446
Dataset 7 (Prop.)	297	463	285	230	189	74	347	1446	1862	1142
Dataset 7 (JOB)	216	390	188	178	108	72	288	2684	3165	490
Dataset 8 (Prop.)	2471	1669	208	440	117	242	1565	2881	2392	1060
Dataset 8 (JOB)	688	895	247	278	69	136	309	2130	2832	2289
Dataset 9 (Prop.)	732	353	360	429	90	835	64	1376	3030	2347
Dataset 9 (JOB)	597	445	302	481	89	507	84	2231	2425	1471
Dataset 10 (Prop.)	783	595	620	44	129	143	1504	1130	1621	627
Dataset 10 (JOB)	384	257	251	197	144	63	721	1361	1283	601

Effective sample size for the first 10 regression coefficients for the proposed sampler (Prop.) and the JOB sampler (JOB) for 10 simulated datasets with $n = 100$ and $p = 750$. Both algorithms are provided the same computational budget of 810 seconds (wall clock time) each. All datasets are simulated from a 'true' linear regression model where only the first 10 regression coefficients are non-zero, and the entries of the design matrix X and the error vector are all independent Gaussian (both the design matrix and the error vector are generated separately for each dataset).

Horseshoe posterior, it is natural to consider a Gibbs sampler to sample from the regularized Horseshoe posterior as well. In fact, after introducing the augmented variables $\{\nu_j\}_{j=1}^p$, the following conditional posterior distributions can be obtained after straightforward computations:

$$\begin{aligned}
\boldsymbol{\beta} \mid \sigma^2, \tau^2, \boldsymbol{\lambda}, \mathbf{y} &\sim \mathcal{N}\left(A_c^{-1}\mathbf{X}^T\mathbf{y}, \sigma^2 A_c^{-1}\right) \\
\sigma^2 \mid \tau^2, \boldsymbol{\lambda}, \mathbf{y} &\sim \text{Inverse-Gamma}\left(a + \frac{n}{2}, \frac{\mathbf{y}^T\left(I_n - \mathbf{X}A_c^{-1}\mathbf{X}^T\right)\mathbf{y}}{2} + b\right) \\
\nu_j \mid \lambda_j^2, \mathbf{y} &\sim \text{Inverse-Gamma}\left(1, 1 + \frac{1}{\lambda_j^2}\right), \text{ independently for } j = 1, 2, \cdots, p \\
\pi\left(\boldsymbol{\lambda} \mid \boldsymbol{\beta}, \boldsymbol{\nu}, \sigma^2, \tau^2, \mathbf{y}\right) &= \prod_{j=1}^p g\left(\lambda_j^2 \mid \nu_j, \beta_j, \sigma^2, \tau^2, \mathbf{y}\right) \\
\tau^2 \mid \boldsymbol{\lambda}, \mathbf{y} &\sim \pi\left(\tau^2 \mid \boldsymbol{\lambda}, \mathbf{y}\right) \qquad (3.2)
\end{aligned}$$

where

$$g\left(\lambda_j^2 \mid \nu_j, \beta_j, \sigma^2, \tau^2, \mathbf{y}\right) \propto \left(\frac{1}{c^2} + \frac{1}{\tau^2\lambda_j^2}\right)^{\frac{1}{2}} \left(\lambda_j^2\right)^{-\frac{3}{2}} \exp\left[-\frac{1}{\lambda_j^2}\left(\frac{1}{\nu_j} + \frac{\beta_j^2}{2\sigma^2\tau^2}\right)\right]$$

Table 4
MCMC STANDARD ERROR COMPARISON

	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
Dataset 1 (Prop.)	0.0252	0.0148	0.0164	0.0496	0.0157	0.0171	0.0017	0.0053	0.0017	0.0019
Dataset 1 (JOB)	0.0112	0.0137	0.0093	0.0093	0.0207	0.0039	0.022	0.0019	0.0029	0.0012
Dataset 2 (Prop.)	0.0252	0.0148	0.0164	0.0496	0.0157	0.0171	0.0017	0.0053	0.0017	0.0019
Dataset 2 (JOB)	0.0132	0.0135	0.0135	0.014	0.0242	0.0316	0.0029	0.0152	0.0012	0.0014
Dataset 3 (Prop.)	0.0252	0.0148	0.0164	0.0496	0.0157	0.0171	0.0017	0.0053	0.0017	0.0019
Dataset 3 (JOB)	0.0153	0.0129	0.0105	0.0119	0.0119	0.0113	0.0027	0.0281	8e-04	0.0038
Dataset 4 (Prop.)	0.0252	0.0148	0.0164	0.0496	0.0157	0.0171	0.0017	0.0053	0.0017	0.0019
Dataset 4 (JOB)	0.0193	0.0199	0.0173	0.0142	0.0273	0.0219	0.0356	0.0028	9e-04	9e-04
Dataset 5 (Prop.)	0.0252	0.0148	0.0164	0.0496	0.0157	0.0171	0.0017	0.0053	0.0017	0.0019
Dataset 5 (JOB)	0.0103	0.018	0.0128	0.0114	0.0286	0.0168	0.0337	0.0016	0.0032	0.0039
Dataset 6 (Prop.)	0.0252	0.0148	0.0164	0.0496	0.0157	0.0171	0.0017	0.0053	0.0017	0.0019
Dataset 6 (JOB)	0.016	0.0096	0.0143	0.0201	0.0189	0.0147	0.0071	0.002	0.0031	0.0014
Dataset 7 (Prop.)	0.0252	0.0148	0.0164	0.0496	0.0157	0.0171	0.0017	0.0053	0.0017	0.0019
Dataset 7 (JOB)	0.0175	0.013	0.0176	0.0217	0.0255	0.0334	0.0095	8e-04	7e-04	0.003
Dataset 8 (Prop.)	0.0252	0.0148	0.0164	0.0496	0.0157	0.0171	0.0017	0.0053	0.0017	0.0019
Dataset 8 (JOB)	0.0088	0.0081	0.019	0.0148	0.042	0.0219	0.0047	0.001	8e-04	0.001
Dataset 9 (Prop.)	0.0252	0.0148	0.0164	0.0496	0.0157	0.0171	0.0017	0.0053	0.0017	0.0019
Dataset 9 (JOB)	0.0107	0.0125	0.0134	0.0104	0.0336	0.004	0.0365	0.0012	0.001	0.0012
Dataset 10 (Prop.)	0.0252	0.0148	0.0164	0.0496	0.0157	0.0171	0.0017	0.0053	0.0017	0.0019
Dataset 10 (JOB)	0.0145	0.0167	0.0177	0.0222	0.0204	0.0448	0.0029	0.0016	0.0017	0.0037

MCMC standard error for the first 10 regression coefficients for the proposed sampler (Prop.) and the JOB sampler (JOB) for 10 simulated datasets with $n = 100$ and $p = 750$. Both algorithms are provided the same computational budget of 810 seconds (wall clock time) each. All datasets are simulated from a 'true' linear regression model where only the first 10 regression coefficients are non-zero, and the entries of the design matrix X and the error vector are all independent Gaussian (both the design matrix and the error vector are generated separately for each dataset).

for $j = 1, 2, \cdots, p$,

$$\pi\left(\tau^2 \middle| \boldsymbol{\lambda}, \mathbf{y}\right)$$
$$\propto |A_c|^{-\frac{1}{2}} \prod_{j=1}^{p} \left\{ \left(\frac{1}{c^2} + \frac{1}{\tau^2 \lambda_j^2} \right)^{\frac{1}{2}} \right\} \left(\frac{\mathbf{y}^T \left(I_n - \mathbf{X} A_c^{-1} \mathbf{X}^T\right) \mathbf{y}}{2} + b \right)^{-\left(a + \frac{n}{2}\right)} \pi_\tau\left(\tau^2\right)$$

and $A_c = \mathbf{X}^T\mathbf{X} + \left(\tau^2 \boldsymbol{\Lambda}\right)^{-1} + c^{-2} I_p$. Most of the above densities are standard and can be easily sampled from. Very efficient rejection samplers based on mixtures of two Inverse Gamma densities can be used to sample from the one-dimensional non-standard densities $g\left(\lambda_j^2 \middle| \nu_j, \beta_j, \sigma^2, \tau^2, \mathbf{y}\right)$ (see Appendix D). The density $\pi\left(\tau^2 \middle| \boldsymbol{\lambda}, \mathbf{y}\right)$ is algebraically more complicated compared to the conditional λ_j^2 density, and obtaining a Metropolis draw is computationally more attractive and feasible than a direct rejection sampling based draw. Similar to the analysis in the previous section, we will first analyze the setting where an exact sample is generated from $\pi\left(\tau^2 \middle| \boldsymbol{\lambda}, \mathbf{y}\right)$. We will then leverage this analysis to analyze the Metropolis draw based setting.

Hence, we consider a two-block Gibbs sampler, whose one step-transition from $(\boldsymbol{\beta}_0, \sigma_0^2, \boldsymbol{\nu}_0, \tau_0^2, \boldsymbol{\lambda}_0)$ to $(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\nu}, \tau^2, \boldsymbol{\lambda})$ is given by sampling sequentially from $\pi(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\nu}, \tau^2 \mid \boldsymbol{\lambda}_0, \mathbf{y})$ and $\pi(\boldsymbol{\lambda} \mid \boldsymbol{\beta}, \sigma^2, \boldsymbol{\nu}, \tau^2, \mathbf{y})$, can be used to generate approximate samples from the regularized Horseshoe posterior. We will denote the transition kernel of this two-block Gibbs sampler by $K_{aug,reg}$ (analogous to K_{aug} in the original Horseshoe setting).

Our goal now is to establish geometric ergodicity for $K_{aug,reg}$. We will achieve this by focusing on the marginal $\boldsymbol{\lambda}$-chain corresponding to $K_{aug,reg}$. The one-step transition dynamics of this Markov chain from $\boldsymbol{\lambda}_m$ to $\boldsymbol{\lambda}_{m+1}$ is given as follows:

1. Draw τ^2 from $\pi\left(\tau^2\middle|\,\boldsymbol{\lambda}_m, \mathbf{y}\right)$
2. Draw $\boldsymbol{\nu}$ from $\pi\left(\boldsymbol{\nu}\middle|\,\boldsymbol{\lambda}_m, \mathbf{y}\right) = \prod_{j=1}^{p} \text{Inverse-Gamma}\left(1, 1 + \frac{1}{\lambda_{j;m}^2}\right)$
3. Draw σ^2 from
$$\pi\left(\sigma^2\middle|\,\tau^2, \boldsymbol{\lambda}_m, \mathbf{y}\right) = \text{Inverse-Gamma}\left(a + \frac{n}{2}, \frac{\mathbf{y}^T\left(I_n - \mathbf{X}A_c^{-1}X^T\right)\mathbf{y}}{2} + b\right)$$
4. Draw $\boldsymbol{\beta}$ from $\pi\left(\boldsymbol{\beta}\middle|\,\sigma^2, \tau^2, \boldsymbol{\lambda}_m, \mathbf{y}\right) = \mathcal{N}_p(A_c^{-1}\mathbf{X}^T\mathbf{y}, \sigma^2 A_c^{-1})$
5. Finally draw $\boldsymbol{\lambda}_{m+1}$ from $\pi\left(\boldsymbol{\lambda}\middle|\,\boldsymbol{\beta}, \boldsymbol{\nu}, \sigma^2, \tau^2, \mathbf{y}\right) = \prod_{j=1}^{p} g\left(\lambda_j^2\middle|\,\nu_j, \beta_j, \sigma^2, \tau^2, \mathbf{y}\right)$.

The Markov transition density (MTD) corresponding to the marginal $\boldsymbol{\lambda}$-chain is given by

$$\begin{aligned}
& k\left(\boldsymbol{\lambda}_0, \boldsymbol{\lambda}\right) \\
= & \int_{\mathbb{R}_+}\int_{\mathbb{R}_+}\int_{\mathbb{R}^p}\int_{\mathbb{R}_+^p} \pi\left(\boldsymbol{\lambda}\middle|\,\boldsymbol{\beta}, \boldsymbol{\nu}, \sigma^2, \tau^2, \mathbf{y}\right)\pi\left(\boldsymbol{\beta}, \boldsymbol{\nu}, \sigma^2, \tau^2\middle|\,\boldsymbol{\lambda}_0, \mathbf{y}\right) d\boldsymbol{\nu}\, d\boldsymbol{\beta}\, d\sigma^2 d\tau^2 \\
= & \int_{\mathbb{R}_+}\int_{\mathbb{R}_+}\int_{\mathbb{R}^p}\int_{\mathbb{R}_+^p} \pi\left(\boldsymbol{\lambda}\middle|\,\boldsymbol{\beta}, \boldsymbol{\nu}, \sigma^2, \tau^2, \mathbf{y}\right)\pi\left(\boldsymbol{\beta}\middle|\,\sigma^2, \tau^2, \boldsymbol{\lambda}_0, \nu, \mathbf{y}\right) \times \\
& \pi\left(\sigma^2\middle|\,\tau^2, \boldsymbol{\lambda}_0, \nu, \mathbf{y}\right)\pi\left(\boldsymbol{\nu}\middle|\,\boldsymbol{\lambda}_0, \tau^2, \mathbf{y}\right)\pi\left(\tau^2\middle|\,\boldsymbol{\lambda}_0, \mathbf{y}\right) d\boldsymbol{\nu}\, d\boldsymbol{\beta}\, d\sigma^2 d\tau^2 \qquad (3.3)
\end{aligned}$$

3.2. *Drift and minorization analysis for the regularized Horseshoe $\boldsymbol{\lambda}$-chain*

The geometric ergodicity of the $\boldsymbol{\lambda}$-chain will be established using a drift and minorization analysis. However, given the modifications in the regularized Horseshoe posterior, the drift function $V(\boldsymbol{\lambda})$ (see (2.4)) used for the original Horseshoe does not work in this case. We will instead use another drift function $\tilde{V}(\boldsymbol{\lambda})$ defined by

$$\tilde{V}\left(\boldsymbol{\lambda}\right) = \sum_{j=1}^{p}\left(\lambda_j^2\right)^{-\frac{\delta}{2}};\ \text{for some constant } \delta \in (0,1). \qquad (3.4)$$

As discussed previously, the function $V(\boldsymbol{\lambda})$ is unbounded off petite sets and the V-based drift condition in Lemma 2.1 is enough to guarantee geometric ergodicity for the original Horseshoe Gibbs sampler K_{aug}. A minorization condition is only needed if one also wants to get quantitative convergence bounds for distance to stationarity. The function $\tilde{V}$ however, is *not* unbounded off petite sets since

$$B\left(\tilde{V}, d\right) = \left\{\boldsymbol{\lambda} \in \mathbb{R}_+^p : \tilde{V}(\boldsymbol{\lambda}) \le d\right\}$$

is not a compact subset of $\mathbb{R}_+^p$ for $d > 0$. Hence, a drift condition with $\tilde{V}$ needs to be complemented with a minorization condition in order to establish geometric ergodicity (Theorem 3.1). We establish these two conditions respectively in Sections 3.2.1 and 3.2.2 below. As opposed to the original Horseshoe setting, we do not require that the prior density π_τ is truncated below away from zero. Only the existence of the $\delta/2^{th}$-moment is assumed for some $\delta \in (0.00162, 0.22176)$: a very mild condition, satisfied for example by the commonly used half-Cauchy density.

3.2.1. Drift condition

Lemma 3.1. *Suppose $\int_{\mathbb{R}_+} u^{\delta/2}\pi_\tau(u)du < \infty$ for some $\delta \in (0.00162, 0.22176)$. Then, there exist constants $0 < \gamma^* = \gamma^*(\delta) < 1$ and $b^* < \infty$ such that*

$$\boldsymbol{E}\left[\tilde{V}(\boldsymbol{\lambda})\Big|\boldsymbol{\lambda}_0\right] \leq \gamma^* \tilde{V}(\boldsymbol{\lambda}_0) + b^* \tag{3.5}$$

for every $\boldsymbol{\lambda}_0 \in \mathbb{R}_+^p$.

Proof. Note that by linearity

$$\boldsymbol{E}\left[\tilde{V}(\boldsymbol{\lambda})\Big|\boldsymbol{\lambda}_0\right] = \sum_{j=1}^{p} \boldsymbol{E}\left[\left(\lambda_j^2\right)^{-\frac{\delta}{2}}\Big|\boldsymbol{\lambda}_0\right] \tag{3.6}$$

Fix $j \in \{1, 2, \cdots, p\}$ arbitrarily. It follows from the definition of the MTD (3.3) that

$$\begin{aligned}\boldsymbol{E}\left[\left(\lambda_j^2\right)^{-\frac{\delta}{2}}\Big|\boldsymbol{\lambda}_0\right] = \boldsymbol{E}[\boldsymbol{E}[\boldsymbol{E}[\boldsymbol{E}[\boldsymbol{E}[\left(\lambda_j^2\right)^{-\frac{\delta}{2}} \mid \boldsymbol{\beta}, \boldsymbol{\nu}, \sigma^2, \tau^2, \mathbf{y}] \mid \sigma^2, \tau^2, \boldsymbol{\lambda}_0, \boldsymbol{\nu}, \mathbf{y}] \\ \mid \tau^2, \boldsymbol{\lambda}_0, \boldsymbol{\nu}, \mathbf{y}] \mid \boldsymbol{\lambda}_0, \tau^2, \mathbf{y}] \mid \boldsymbol{\lambda}_0, \mathbf{y}].\end{aligned} \tag{3.7}$$

We begin by evaluating the innermost expectation. It follows by using $\sqrt{u+v} \leq \sqrt{u} + \sqrt{v}$ for $u, v \geq 0$ that

$$\begin{aligned}&\boldsymbol{E}\left[\left(\lambda_j^2\right)^{-\frac{\delta}{2}}\Big|\boldsymbol{\beta}, \boldsymbol{\nu}, \sigma^2, \tau^2, \mathbf{y}\right] \\ &= \frac{\int_0^\infty \left(\lambda_j^2\right)^{-\frac{\delta}{2}} \left(\frac{1}{c^2} + \frac{1}{\tau^2\lambda_j^2}\right)^{\frac{1}{2}} \left(\lambda_j^2\right)^{-\frac{3}{2}} \exp\left[-\frac{1}{\lambda_j^2}\left(\frac{1}{\nu_j} + \frac{\beta_j^2}{2\sigma^2\tau^2}\right)\right] d\lambda_j^2}{\int_0^\infty \left(\frac{1}{c^2} + \frac{1}{\tau^2\lambda_j^2}\right)^{\frac{1}{2}} \left(\lambda_j^2\right)^{-\frac{3}{2}} \exp\left[-\frac{1}{\lambda_j^2}\left(\frac{1}{\nu_j} + \frac{\beta_j^2}{2\sigma^2\tau^2}\right)\right] d\lambda_j^2} \\ &\leq \frac{\int_0^\infty \left(\lambda_j^2\right)^{-\frac{\delta}{2}} \left(\frac{1}{|c|} + \frac{1}{\sqrt{\tau^2\lambda_j^2}}\right) \left(\lambda_j^2\right)^{-\frac{3}{2}} \exp\left[-\frac{1}{\lambda_j^2}\left(\frac{1}{\nu_j} + \frac{\beta_j^2}{2\sigma^2\tau^2}\right)\right] d\lambda_j^2}{\int_0^\infty \left(\frac{1}{c^2} + \frac{1}{\tau^2\lambda_j^2}\right)^{\frac{1}{2}} \left(\lambda_j^2\right)^{-\frac{3}{2}} \exp\left[-\frac{1}{\lambda_j^2}\left(\frac{1}{\nu_j} + \frac{\beta_j^2}{2\sigma^2\tau^2}\right)\right] d\lambda_j^2}\end{aligned}$$

$$\leq \frac{\frac{\left(\tau^2\right)^{\frac{\delta}{2}}}{|c|}\int\limits_0^\infty \left(\tau^2\lambda_j^2\right)^{-\frac{\delta}{2}}\left(\lambda_j^2\right)^{-\frac{3}{2}}\exp\left[-\frac{1}{\lambda_j^2}\left(\frac{1}{\nu_j}+\frac{\beta_j^2}{2\sigma^2\tau^2}\right)\right]d\lambda_j^2}{\int\limits_0^\infty\left(\frac{1}{c^2}+\frac{1}{\tau^2\lambda_j^2}\right)^{\frac{1}{2}}\left(\lambda_j^2\right)^{-\frac{3}{2}}\exp\left[-\frac{1}{\lambda_j^2}\left(\frac{1}{\nu_j}+\frac{\beta_j^2}{2\sigma^2\tau^2}\right)\right]d\lambda_j^2}$$
$$+\frac{\int\limits_0^\infty\left(\lambda_j^2\right)^{-\frac{\delta}{2}}\left(\lambda_j^2\right)^{-2}\exp\left[-\frac{1}{\lambda_j^2}\left(\frac{1}{\nu_j}+\frac{\beta_j^2}{2\sigma^2\tau^2}\right)\right]d\lambda_j^2}{\int\limits_0^\infty\left(\lambda_j^2\right)^{-2}\exp\left[-\frac{1}{\lambda_j^2}\left(\frac{1}{\nu_j}+\frac{\beta_j^2}{2\sigma^2\tau^2}\right)\right]d\lambda_j^2}. \tag{3.8}$$

The first term in the last inequality of (3.8) can be expressed as

$$\frac{\left(\tau^2\right)^{\frac{\delta}{2}}}{|c|}\frac{\boldsymbol{E}\left[X^\delta\right]}{\boldsymbol{E}\left[\sqrt{\frac{1}{c^2}+X^2}\right]}$$

where $\tau^2X^2 \sim \text{Gamma}\left(\frac{1}{2},\frac{1}{\nu_j}+\frac{\beta_j^2}{2\sigma^2\tau^2}\right)$. Using Young's inequality, it follows that the first term is bounded above by

$$\frac{\max\{1,|c|\}}{|c|}\sqrt{\delta}\left(\tau^2\right)^{\frac{\delta}{2}}.$$

The second term in the last inequality of (3.8) is basically an Inverse-Gamma expectation, and is exactly equal to

$$\frac{\Gamma\left(1+\frac{\delta}{2}\right)}{\left(\frac{1}{\nu_j}+\frac{\beta_j^2}{2\sigma^2\tau^2}\right)^{\frac{\delta}{2}}}.$$

Hence, we get

$$\boldsymbol{E}\left[\left(\lambda_j^2\right)^{-\frac{\delta}{2}}\middle|\boldsymbol{\beta},\boldsymbol{\nu},\sigma^2,\tau^2,\mathbf{y}\right]\leq\frac{\max\{1,|c|\}}{|c|}\sqrt{\delta}\left(\tau^2\right)^{\frac{\delta}{2}}+\frac{\Gamma\left(1+\frac{\delta}{2}\right)}{\left(\frac{1}{\nu_j}+\frac{\beta_j^2}{2\sigma^2\tau^2}\right)^{\frac{\delta}{2}}}.$$

Note that the conditional distribution of β_j given $\sigma^2,\tau^2,\boldsymbol{\lambda}_0,\boldsymbol{\nu},\mathbf{y}$ is a Gaussian distribution with variance $\sigma_j^2\overset{\text{def}}{=}\sigma^2\boldsymbol{e}_j^TA_c^{-1}\boldsymbol{e}_j\geq\sigma^2\left(\bar{\omega}+\frac{1}{c^2}+\frac{1}{\tau^2\lambda_{j;0}^2}\right)^{-1}$. Here $\bar{\omega}$ is the maximum eigenvalue of $\mathbf{X}^T\mathbf{X}$. Now, proceeding exactly with the analysis from (2.9) to (2.13) in the proof of Lemma 2.1 with $\bar{\omega}$ replaced by $\bar{\omega}+c^{-2}$ and using Proposition B.3 instead of Proposition B.1 yields

$$\boldsymbol{E}\left[\sum_{j=1}^p\left(\lambda_j^2\right)^{-\frac{\delta_1}{2}}\middle|\boldsymbol{\lambda}_0\right]\leq\gamma^*(\delta)\sum_{j=1}^p\left(\lambda_{j;0}^2\right)^{-\frac{\delta_1}{2}}+b^*$$

with

$$\gamma^*(\delta)=\Gamma\left(1+\frac{\delta}{2}\right)\left(\frac{1}{\left\{\Gamma\left(1-\frac{\delta}{2}\right)\right\}^{\frac{2}{\delta}}}+\frac{\sqrt{\pi}^{\frac{2}{\delta}}}{\left\{\Gamma\left(\frac{1-\delta}{2}\right)\right\}^{\frac{2}{\delta}}}\right)^{-\frac{\delta}{2}}$$

and

$$b^* = p\frac{\max\{1,|c|\}}{|c|}\sqrt{\delta}C_2 + p\cdot\max\left\{1,(\bar{\omega}+c^{-2})^{\frac{\delta}{2}}C_2\right\}\cdot\gamma^*(\delta).$$

Here C_2 is as in Proposition B.3. It can be shown that $\gamma^*(\delta) < 1$ for $\delta \in (0.00162, 0.22176)$.

Hence, the required geometric drift condition has been established. □

3.2.2. Minorization condition

As discussed previously, the drift function $\tilde{V}$ is not unbounded off compact sets, and the drift condition in Lemma 3.1 needs to be complemented by an associated minorization condition to establish geometric ergodicity. Fix a $d > 0$. Define

$$B\left(\tilde{V}, d\right) = \left\{\boldsymbol{\lambda} \in \mathbb{R}_+^p : \tilde{V}(\boldsymbol{\lambda}) \leq d\right\} \tag{3.9}$$

We now establish the following minorization condition associated to the geometric drift condition in Lemma 3.1.

Lemma 3.2. *There exists a constant $\epsilon^* = \epsilon^*\left(\tilde{V}, d\right) > 0$ and a density function h on $\mathbb{R}_+^p$ such that*

$$k(\boldsymbol{\lambda}_0, \boldsymbol{\lambda}) \geq \epsilon^* h(\boldsymbol{\lambda}) \tag{3.10}$$

for every $\boldsymbol{\lambda}_0 \in B\left(\tilde{V}, d\right)$.

Proof. Fix a $\boldsymbol{\lambda}_0 \in B\left(\tilde{V}, d\right)$ arbitrarily. In order to prove (3.10) we will demonstrate appropriate lower bounds for the conditional densities appearing in (3.3). From (3.2) we have the following:

$$\pi\left(\tau^2 \middle| \boldsymbol{\lambda}_0, \mathbf{y}\right) \geq b^{a+\frac{n}{2}}\omega_*^{-\frac{p}{2}}\left(1+\frac{1}{\tau^2}\right)^{-\frac{p}{2}}|c|^{-p}\left(\mathbf{y}^T\mathbf{y}+b\right)^{-\left(a+\frac{n}{2}\right)}\pi_\tau\left(\tau^2\right);$$

where $\omega_* = \max\left\{\bar{\omega}+c^{-2}, d^{\frac{2}{\delta}}\right\}$; (recall that $\bar{\omega}$ denotes the maximum eigenvalue of $\mathbf{X}^T\mathbf{X}$),

$$\begin{aligned} & \pi\left(\boldsymbol{\beta}\middle|\sigma^2, \tau^2, \boldsymbol{\lambda}_0, \mathbf{y}\right) \\ \geq\ & \left(2\pi\sigma^2\right)^{-\frac{p}{2}}|c|^{-p} \\ \times\ & \exp\left[-\frac{\left(\boldsymbol{\beta}-\Omega^{-1}\mathbf{X}^T\mathbf{y}\right)^T\Omega\left(\boldsymbol{\beta}-\Omega^{-1}\mathbf{X}^T\mathbf{y}\right)+\mathbf{y}^T\mathbf{X}\left(c^2I_p-\Omega^{-1}\right)\mathbf{X}^T\mathbf{y}}{2\sigma^2}\right]; \end{aligned}$$

where $\Omega = \omega_*\left(1+\frac{1}{\tau^2}\right)I_p$,

$$\pi\left(\boldsymbol{\lambda}\middle|\boldsymbol{\beta}, \boldsymbol{\nu}, \sigma^2, \tau^2, \mathbf{y}\right)$$

$$
\begin{aligned}
&= \prod_{j=1}^{p}\left\{\frac{\left(\frac{1}{c^2}+\frac{1}{\tau^2\lambda_j^2}\right)^{\frac{1}{2}}\left(\lambda_j^2\right)^{-\frac{3}{2}}\exp\left[-\frac{1}{\lambda_j^2}\left(\frac{1}{\nu_j}+\frac{\beta_j^2}{2\sigma^2\tau^2}\right)\right]}{\int\limits_0^\infty\left(\frac{1}{c^2}+\frac{1}{\tau^2\lambda_j^2}\right)^{\frac{1}{2}}\left(\lambda_j^2\right)^{-\frac{3}{2}}\exp\left[-\frac{1}{\lambda_j^2}\left(\frac{1}{\nu_j}+\frac{\beta_j^2}{2\sigma^2\tau^2}\right)\right]d\lambda_j^2}\right\}\\
&\geq \prod_{j=1}^{p}\left\{\frac{\left(\tau^2\right)^{-\frac{1}{2}}\left(\lambda_j^2\right)^{-2}\exp\left[-\frac{1}{\lambda_j^2}\left(\frac{1}{\nu_j}+\frac{\beta_j^2}{2\sigma^2\tau^2}\right)\right]}{k^*\left(\sqrt{\nu_j}+\frac{\sigma^2\sqrt{\tau^2}}{\beta_j^2}\right)}\right\};
\end{aligned}
$$

where $k^*=\max\left\{\sqrt{\pi}|c|^{-1},2\right\}$, and

$$
\begin{aligned}
\pi\left(\boldsymbol{\nu}\middle|\,\boldsymbol{\lambda}_0,\mathbf{y}\right) &\geq \prod_{j=1}^{p}\left\{\nu_j^{-2}\exp\left[-\frac{1}{\nu_j}\left(1+d^{\frac{2}{\delta}}\right)\right]\right\}\\
\pi\left(\sigma^2\middle|\,\tau^2,\boldsymbol{\lambda}_0,\mathbf{y}\right) &\geq \frac{b^{a+\frac{n}{2}}}{\Gamma\left(a+\frac{n}{2}\right)}\left(\sigma^2\right)^{-\left(a+\frac{n}{2}\right)-1}\exp\left[-\frac{1}{\sigma^2}\left(\frac{\mathbf{y}^T\mathbf{y}}{2}+b\right)\right].
\end{aligned}
\tag{3.11}
$$

Combining all the lower bounds provided above, it follows from (3.3) that

$$
\begin{aligned}
&k\left(\boldsymbol{\lambda}_0,\boldsymbol{\lambda}\right)\\
&\geq \frac{(2\pi)^{-\frac{p}{2}}b^{2\left(a+\frac{n}{2}\right)}\left(\sqrt{\omega_*}k^*c^2\right)^{-p}}{\left(\mathbf{y}^T\mathbf{y}+b\right)^{a+\frac{n}{2}}\Gamma\left(a+\frac{n}{2}\right)}\\
&\quad\int\limits_{\mathbb{R}_+}\int\limits_{\mathbb{R}_+}\int\limits_{\mathbb{R}^p}\int\limits_{\mathbb{R}_+^p}\left(\sigma^2\right)^{-\left(a+\frac{n+p}{2}\right)-1}\prod_{j=1}^{p}\left\{\frac{\nu_j^{-2}\exp\left[-\frac{1}{\nu_j}\left(1+d^{\frac{2}{\delta}}+\frac{1}{\lambda_j^2}\right)\right]}{\sqrt{\nu_j}+\frac{\sigma^2\sqrt{\tau^2}}{\beta_j^2}}\right\}\\
&\qquad\exp\left[-\frac{\left(\boldsymbol{\beta}-\Omega^{-1}\mathbf{X}^T\mathbf{y}\right)^T\Omega\left(\boldsymbol{\beta}-\Omega^{-1}\mathbf{X}^T\mathbf{y}\right)+\boldsymbol{\beta}^T\left(\tau^2\boldsymbol{\Lambda}\right)^{-1}\boldsymbol{\beta}}{2\sigma^2}\right]\\
&\quad\prod_{j=1}^{p}\left\{\left(\lambda_j^2\right)^{-2}\right\}\exp\left[-\frac{1}{\sigma^2}\left(\frac{\mathbf{y}^T\mathbf{y}+\mathbf{y}^T\mathbf{X}\left(c^2I_p-\Omega^{-1}\right)\mathbf{X}^T\mathbf{y}}{2}+b\right)\right]\\
&\quad\left(1+\tau^2\right)^{-\frac{p}{2}}\pi_\tau\left(\tau^2\right)d\boldsymbol{\nu}d\boldsymbol{\beta}d\sigma^2d\tau^2
\end{aligned}
$$

Now for the inner most integral wrt $\boldsymbol{\nu}$, substituting the lower bounds given in Proposition B.6, induce the following lower bound on $k\left(\boldsymbol{\lambda}_0,\boldsymbol{\lambda}\right)$:

$$
\begin{aligned}
&k\left(\boldsymbol{\lambda}_0,\boldsymbol{\lambda}\right)\\
&\geq \frac{(2\pi)^{-\frac{p}{2}}b^{2\left(a+\frac{n}{2}\right)}\alpha^p\left(\sqrt{\omega_*}k^*c^2\right)^{-p}}{\left(\mathbf{y}^T\mathbf{y}+b\right)^{a+\frac{n}{2}}\Gamma\left(a+\frac{n}{2}\right)}\\
&\quad\int\limits_{\mathbb{R}_+}\int\limits_{\mathbb{R}_+}\int\limits_{\mathbb{R}^p}\left(\sigma^2\right)^{-\left(a+\frac{n+p}{2}\right)-1}\prod_{j=1}^{p}\left\{\left(\lambda_j^2\right)^{-2}\right\}
\end{aligned}
$$

$$\exp\left[-\frac{\left(\boldsymbol{\beta}-\Omega^{-1}\mathbf{X}^T\mathbf{y}\right)^T\Omega\left(\boldsymbol{\beta}-\Omega^{-1}\mathbf{X}^T\mathbf{y}\right)+\boldsymbol{\beta}^T\left(\tau^2\boldsymbol{\Lambda}\right)^{-1}\boldsymbol{\beta}}{2\sigma^2}\right]$$
$$\prod_{j=1}^{p}\left\{\frac{\left(1+\frac{1}{\lambda_j^2}\right)^{-2}}{1+\frac{\sigma^2\sqrt{\tau^2}}{\beta_j^2}}\right\}$$
$$\exp\left[-\frac{1}{\sigma^2}\left(\frac{\mathbf{y}^T\mathbf{y}+\mathbf{y}^T\mathbf{X}\left(c^2I_p-\Omega^{-1}\right)\mathbf{X}^T\mathbf{y}}{2}+b\right)\right]\left(1+\tau^2\right)^{-\frac{p}{2}}$$
$$\pi_\tau\left(\tau^2\right)d\boldsymbol{\beta}d\sigma^2d\tau^2$$

where α is some positive constant (see Proposition B.6). For the inner most integral wrt $\boldsymbol{\beta}$ we use the lower bound in Proposition B.7 and get the following:

$$\begin{aligned}
&k\left(\boldsymbol{\lambda}_0,\boldsymbol{\lambda}\right)\\
&\geq \frac{b^{2\left(a+\frac{n}{2}\right)}\alpha^p\left(\sqrt{\omega_*}k^*|c|^3\right)^{-p}}{\left(\mathbf{y}^T\mathbf{y}+b\right)^{a+\frac{n}{2}}\Gamma\left(a+\frac{n}{2}\right)}\\
&\quad\int\limits_{\mathbb{R}_+}\int\limits_{\mathbb{R}_+}\left(\sigma^2\right)^{-\left(a+\frac{n}{2}\right)-1}\exp\left[-\frac{1}{\sigma^2}\left(\frac{\mathbf{y}^T\mathbf{y}+2\mathbf{y}^T\mathbf{X}\left(c^2I_p-M_{\tau^2}^{-1}\right)\mathbf{X}^T\mathbf{y}}{2}+b\right)\right]\\
&\quad\prod_{j=1}^{p}\left(1+\lambda_j^2\right)^{-2}\times\left|M_{\tau^2}\right|^{-1}\left(1+\frac{\sqrt{\tau^2}}{c^2}\right)^{-p}\left(1+\tau^2\right)^{-\frac{p}{2}}\pi_\tau\left(\tau^2\right)d\sigma^2d\tau^2;
\end{aligned}$$

where M_{τ^2} is as in Proposition B.7. It follows that

$$\begin{aligned}
&k\left(\boldsymbol{\lambda}_0,\boldsymbol{\lambda}\right)\\
&\geq \frac{b^{2\left(a+\frac{n}{2}\right)}\alpha^p\left(\sqrt{\omega_*}k^*|c|^3\right)^{-p}}{\left(\mathbf{y}^T\mathbf{y}+b\right)^{a+\frac{n}{2}}\Gamma\left(a+\frac{n}{2}\right)}\int\limits_{\mathbb{R}_+}\int\limits_{\mathbb{R}_+}\left(\sigma^2\right)^{-\left(a+\frac{n}{2}\right)-1}\\
&\qquad\exp\left[-\frac{1}{\sigma^2}\left(\frac{\mathbf{y}^T\mathbf{y}+2c^2\mathbf{y}^T\mathbf{X}\mathbf{X}^T\mathbf{y}}{2}+b\right)\right]\\
&\qquad\prod_{j=1}^{p}\left(1+\lambda_j^2\right)^{-2}\times\left|M_{\tau^2}\right|^{-1}\left(1+\frac{\sqrt{\tau^2}}{c^2}\right)^{-p}\left(1+\tau^2\right)^{-\frac{p}{2}}\pi_\tau\left(\tau^2\right)d\sigma^2d\tau^2,
\end{aligned}$$

since $\mathbf{y}^T\mathbf{X}M_{\tau^2}^{-1}\mathbf{X}^T\mathbf{y}/\sigma^2\geq 0$. Next by virtue of the inverse-gamma integral, we have

$$\begin{aligned}
&\int\limits_{\mathbb{R}_+}\left(\sigma^2\right)^{-\left(a+\frac{n}{2}\right)-1}\exp\left[-\frac{1}{\sigma^2}\left(\frac{\mathbf{y}^T\mathbf{y}+2c^2\mathbf{y}^T\mathbf{X}\mathbf{X}^T\mathbf{y}}{2}+b\right)\right]d\sigma^2\\
=\;&\frac{\Gamma\left(a+\frac{n}{2}\right)}{\left(\frac{\mathbf{y}^T\mathbf{y}+2c^2\mathbf{y}^T\mathbf{X}\mathbf{X}^T\mathbf{y}}{2}+b\right)^{a+\frac{n}{2}}}.
\end{aligned}$$

This together with the fact that $\mid M_{\tau^2} \mid \leq \left(1+\frac{1}{\tau^2}\right)^p \prod_{j=1}^p \left(\omega_* + \frac{1}{\lambda_j^2}\right)$ gives the following lower bound:

$$
\begin{aligned}
& k\left(\boldsymbol{\lambda}_0, \boldsymbol{\lambda}\right) \\
\geq & \frac{b^{2\left(a+\frac{n}{2}\right)} \alpha^p \left(\sqrt{\omega_*} k^* |c|^3\right)^{-p}}{\left(\mathbf{y}^T\mathbf{y}+b\right)^{a+\frac{n}{2}} \left(\frac{\mathbf{y}^T\mathbf{y}+2c^2\mathbf{y}^T\mathbf{X}\mathbf{X}^T\mathbf{y}}{2}+b\right)^{\left(a+\frac{n}{2}\right)}} \prod_{j=1}^p \left\{\left(1+\lambda_j^2\right)^{-2} \left(\omega_* + \frac{1}{\lambda_j^2}\right)^{-1}\right\} \\
& \int_0^\infty \left(1+\frac{1}{\tau^2}\right)^{-p} \left(1+\frac{\sqrt{\tau^2}}{c^2}\right)^{-p} \left(1+\tau^2\right)^{-\frac{p}{2}} \pi_\tau\left(\tau^2\right) d\tau^2
\end{aligned}
$$

Further denoting $\eta = \max\left\{1, \omega_*\right\}$ we get

$$
k\left(\boldsymbol{\lambda}_0, \boldsymbol{\lambda}\right) \geq \epsilon^* h\left(\boldsymbol{\lambda}\right)
$$

where

$$
\begin{aligned}
\epsilon^* = & \frac{b^{2\left(a+\frac{n}{2}\right)} \alpha^p \left(2\eta^2\sqrt{\omega_*} k^* |c|^3\right)^{-p}}{\left(\mathbf{y}^T\mathbf{y}+b\right)^{a+\frac{n}{2}} \left(\frac{\mathbf{y}^T\mathbf{y}+2c^2\mathbf{y}^T\mathbf{X}\mathbf{X}^T\mathbf{y}}{2}+b\right)^{\left(a+\frac{n}{2}\right)}} \\
& \boldsymbol{E}_{\pi_\tau(\tau^2)} \left[\left(1+\frac{1}{\tau^2}\right)^{-p} \left(1+\frac{\sqrt{\tau^2}}{c^2}\right)^{-p} \left(1+\tau^2\right)^{-\frac{p}{2}}\right],
\end{aligned}
$$

and h is a probability density on $\mathbb{R}_+^p$ given by

$$
h\left(\boldsymbol{\lambda}\right) = \prod_{j=1}^p \left\{\frac{2\eta^2\lambda_j^2}{\left(1+\eta\lambda_j^2\right)^3} I_{(0,\infty)}\left(\lambda_j^2\right)\right\},
$$

and this completes the proof of minorization condition for the MTD k corresponding to the regularized Horseshoe $\boldsymbol{\lambda}$-chain. $\square$

The drift and minorization conditions in Lemma 3.1 and Lemma 3.2 can be combined with Theorem 12 of [21] to establish geometric ergodicity of the regularized Horseshoe Gibbs sampler which is stated as follows:

Theorem 3.1. *Suppose the prior density $\pi_\tau(\cdot)$ for the global shrinkage parameter satisfies*

$$
\int_{\mathbb{R}_+} u^{\delta/2} \pi_\tau(u) du < \infty
$$

for some $\delta \in (0.00162, 0.22176)$. Then, the regularized Horseshoe Gibbs sampler with transition kernel $K_{aug,reg}$ is geometrically ergodic.

Similar to the original Horseshoe setting, it might be desirable in high-dimensional settings to generate the conditional τ^2 draws using a Metropolis-Hastings step instead of a direct rejection sampling step. Hence, one might be interested in understanding the properties of the transition kernel $K_{MGR,q}$ which can be defined by replacing the exact τ^2 draw for $K_{aug,reg}$ with a Metropolis draw based on a proposal density q (analogous to the definition of the Metropolis kernel $K_{MG,q}$ for the original Horseshoe). The next theorem (similar to Theorem 2.2) first establishes a condition on the Metropolis proposal density q under which the Metropolis-within-Gibbs chain with kernel $K_{MGR,q}$ is geometrically ergodic. The next part of the theorem shows this condition is satisfied for three natural choices of the proposal density q.

Theorem 3.2. *Assume that the conditions in Theorem 3.1 hold.*

(a) The Metropolis-within-Gibbs sampler with transition kernel $K_{MGR,q}$ is geometrically ergodic if

$$\sup_{\tau^2_{curr},\tau^{2'}\in(0,\infty)} \frac{\pi_\tau(\tau^{2'})}{q(\tau^{2'} \mid \tau^2_{curr})} < \infty.$$

(b) The condition in part (a) is satisfied if

(i) $q(\tau^{2'} \mid \tau^2) = \pi_\tau(\tau^{2'})$ (independence Metropolis).

(ii) Random walk Metropolis for $\xi = \frac{1}{\tau^2}$, where the proposal $\xi' = 1/\tau^{2'}$ is drawn from $N(\xi^*_{curr}, v^2)$. Here $\xi^*_{curr} = 1/\tau^2_{curr} \vee T$ for any $T > 0$, τ^2_{curr} is the current value of τ^2, v^2 is an arbitrary positive constant, and π_ξ is chosen to be the $N(0, v^2)$ density (truncated to $(0,\infty)$).

(iii) Random walk Metropolis for $\zeta = \log(\tau^2)$, where the proposal $\zeta' = \log(\tau^{2'})$ is drawn from the $N\left(\zeta^*_{curr}, v^2\right)$ distribution. Here $\zeta^*_{curr} = \log\left((\tau^2_{curr} \vee 1) \wedge (1/T)\right)$ for any $T > 0$, τ^2_{curr} is the current value of τ^2, v^2 is an arbitrary positive constant, and π_τ is chosen to be the Lognormal$(0, v^2)$ density (truncated to $[T,\infty)$).

Note that unlike the original Horseshoe chain with kernel K_{aug}, the regularized Horseshoe chain with kernel $K_{aug,reg}$ does not require π_τ to be truncated away from 0 for proving geometric ergodicity. Theorem 3.2 shows that such a truncation is still not required for establishing geometric ergodicity of $K_{MGR,q}$ when using a π_τ based independence Metropolis draw or the random walk Metropolis for $1/\tau^2$. However, truncation of π_τ away from zero is required to prove geometric ergodicity of $K_{MGR,q}$ for the log-scale random walk Metropolis proposal considered in Theorem 3.2.

Proof of Theorem 3.2: By exactly following the steps in the proof of part (a) of Theorem 2.2, it is enough to show that

$$\sup_{\tau^{2'},\tau^2_{curr}\in(0,\infty),\boldsymbol{\lambda}\in\mathbb{R}^p_+} w\left(\tau^{2'}, \tau^2_{curr}, \boldsymbol{\lambda}\right) < \infty.$$

where $w\left(\tau^{2'}, \tau^2_{curr}, \boldsymbol{\lambda}\right) = \frac{\pi(\tau^{2'}|\boldsymbol{\lambda},\mathbf{y})}{q(\tau^{2'}|\tau^2_{curr})}$. Let $D(\tau^2, \boldsymbol{\lambda})$ denote a diagonal matrix whose j^{th} diagonal entry is given by $c^{-2} + (\tau^2\lambda_j)^{-1}$. It follows that

$$\left(\frac{c^{-2}}{c^{-2} + \text{eig}_{\max}(\mathbf{X}^T\mathbf{X})}\right)^{p/2} \leq \frac{\sqrt{|\ D(\tau^2, \boldsymbol{\lambda})\ |}}{\sqrt{|\ D(\tau^2, \boldsymbol{\lambda}) + \mathbf{X}^T\mathbf{X}\ |}} \leq 1$$

for all $\tau^2 \in (0, \infty)$ and $\boldsymbol{\lambda} \in \mathbb{R}_+^p$, where $\text{eig}_{\max}(\mathbf{X}^T\mathbf{X})$ denotes the maximum eigenvalue of $\mathbf{X}^T\mathbf{X}$. Hence, for arbitrary $\tau^2_{curr}, \tau^{2\prime} \in (0, \infty)$ and $\boldsymbol{\lambda} \in \mathbb{R}_+^p$ we get

$$\begin{aligned}
\frac{q(\tau^{2\prime} \mid \tau^2_{curr}) w\left(\tau^{2\prime}, \tau^2_{curr}, \boldsymbol{\lambda}\right)}{\pi_\tau(\tau^{2\prime})} &\leq \left(\frac{b}{\mathbf{y}^T\mathbf{y} + b}\right)^{-\left(a+\frac{n}{2}\right)} \frac{\frac{\sqrt{|D(\tau^{2\prime}, \boldsymbol{\lambda})|}}{\sqrt{|D(\tau^{2\prime}, \boldsymbol{\lambda}) + \mathbf{X}^T\mathbf{X}|}}}{\int_0^\infty \frac{\sqrt{|D(\tau^2 \boldsymbol{\lambda})|}}{\sqrt{|D(\tau^2, \boldsymbol{\lambda}) + \mathbf{X}^T\mathbf{X}|}} \pi_\tau\left(\tau^2\right) d\tau^2} \\
&\leq \left(\frac{b}{\mathbf{y}^T\mathbf{y} + b}\right)^{-\left(a+\frac{n}{2}\right)} \left(\frac{c^{-2} + \text{eig}_{\max}(\mathbf{X}^T\mathbf{X})}{c^{-2}}\right)^{p/2} \\
&< \infty.
\end{aligned}$$

The proof of part (b) follows is almost identical to that of Theorem 2.2 part (b) with trivial adjustments. □

3.3. *Geometric ergodicity of a Gibbs sampler for the regularized Horseshoe variant in [15]*

The following variant of regularized Horseshoe shrinkage prior has been introduced in [15].

$$\begin{aligned}
\pi\left(\beta_j, \lambda_j \mid \tau^2, \sigma^2\right) \propto \frac{1}{\sqrt{\tau^2\lambda_j^2}} \exp\left[-\frac{\beta_j^2}{2\sigma^2}\left(\frac{1}{c^2} + \frac{1}{\tau^2\lambda_j^2}\right)\right] \pi_\ell\left(\lambda_j\right); \\
\text{independently for } j = 1, 2, \cdots, p \\
\sigma^2 \sim \text{Inverse-Gamma}\left(a, b\right);\ \tau^2 \sim \pi_\tau\left(\cdot\right) \qquad (3.12)
\end{aligned}$$

where π_ℓ and π_τ are probability densities. Note that based on the above specification

$$\beta_j \mid \lambda_j^2, \tau^2, \sigma^2 \sim N\left(0, \left(\frac{1}{c^2} + \frac{1}{\tau^2\lambda_j^2}\right)\right)$$

identical to the specification in [18]. The difference is that instead $\boldsymbol{\lambda}, \tau^2$ and σ^2 having independent priors, we now have

$$\pi(\lambda_j \mid \tau^2, \sigma^2) = c(\tau^2)\left(1 + \frac{\tau^2\lambda_j^2}{c^2}\right)^{-1/2} \pi_\ell(\lambda_j), \qquad (3.13)$$

where

$$1/c(\tau^2) = \int_0^\infty \left(1 + \frac{\tau^2\lambda_j^2}{c^2}\right)^{-1/2} \pi_\ell(\lambda_j) d\lambda_j. \qquad (3.14)$$

The principal motivation for the algebraic modification of the prior as compared to that of [18] is the resulting simplification of the posterior computation, although an alternative interpretation using fictitious data is also discussed in [15]. In fact, using π_ℓ to be the half-Cauchy density and using its representation in terms of a mixture of Inverse-Gamma densities ([13]) the following conditional posterior distributions can be obtained from straightforward computations after augmenting the latent variables $\{\nu_j\}_{j=1}^p$.

$$
\begin{aligned}
&\boldsymbol{\beta}|\sigma^2,\tau^2,\boldsymbol{\lambda},\mathbf{y} \sim \mathcal{N}\left(A_c^{-1}\mathbf{X}^T\mathbf{y},\sigma^2 A_c^{-1}\right)\\
&\sigma^2|\tau^2,\boldsymbol{\lambda},\mathbf{y} \sim \text{Inverse-Gamma}\left(a+\frac{n}{2},\frac{\mathbf{y}^T\left(I_n-\mathbf{X}A_c^{-1}\mathbf{X}^T\right)\mathbf{y}}{2}+b\right)\\
&\nu_j|\lambda_j^2,\mathbf{y} \sim \text{Inverse-Gamma}\left(1,1+\frac{1}{\lambda_j^2}\right),\ \text{independently for } j=1,2,\cdots,p\\
&\tau^2|\boldsymbol{\lambda},\mathbf{y} \sim \pi\left(\tau^2\middle|\,\boldsymbol{\lambda},\mathbf{y}\right)\\
&\lambda_j^2|\nu_j,\beta_j,\sigma^2,\tau^2,\mathbf{y} \sim \text{Inverse-Gamma}\left(1,\frac{1}{\nu_j}+\frac{\beta_j^2}{2\sigma^2\tau^2}\right),\\
&\text{independently for } j=1,2,\cdots,p
\end{aligned}
\tag{3.15}
$$

where

$$
\pi\left(\tau^2\middle|\,\boldsymbol{\lambda},\mathbf{y}\right) \propto \left|\tau^2 A_c\right|^{-\frac{1}{2}}\left(\frac{\mathbf{y}^T\left(I_n-\mathbf{X}A_c^{-1}\mathbf{X}^T\right)\mathbf{y}}{2}+b\right)^{-\left(a+\frac{n}{2}\right)}\pi_\tau\left(\tau^2\right)c(\tau^2)^p
$$

and $A_c=\mathbf{X}^T\mathbf{X}+\left(\tau^2\boldsymbol{\Lambda}\right)^{-1}+c^{-2}I_p$. Most of the above conditional posterior densities, including that for the local shrinkage parameters $\left\{\lambda_j^2\right\}_{j=1}^p$ are standard probability distributions (as opposed to the non-standard ones in the regularized Horseshoe posterior in (3.2)) and can be easily sampled from. An efficient Metropolis sampler for the non-standard (one-dimensional) density $\pi\left(\tau^2\middle|\,\boldsymbol{\lambda},\mathbf{y}\right)$ can be constructed similar to the one provided in Appendix D. Hence, a two-block Gibbs sampler, whose one step-transition from $(\boldsymbol{\beta}_0,\sigma_0^2,\boldsymbol{\nu}_0,\tau_0^2,\boldsymbol{\lambda}_0)$ to $(\boldsymbol{\beta},\sigma^2,\boldsymbol{\nu},\tau^2,\boldsymbol{\lambda})$ is given by sampling sequentially from $\pi(\boldsymbol{\beta},\sigma^2,\boldsymbol{\nu},\tau^2 \mid \boldsymbol{\lambda}_0,\mathbf{y})$ and $\pi(\boldsymbol{\lambda} \mid \boldsymbol{\beta},\sigma^2,\boldsymbol{\nu},\tau^2,\mathbf{y})$, can be used to generate approximate samples from the regularized Horseshoe posterior in (3.15). We will denote the Markov transition kernel of this two-block Gibbs sampler by $\tilde{K}_{aug,reg}$ (analogous to $K_{aug,reg}$ in the regularized Horseshoe setting). The transition density can be obtained by substituting the appropriate conditional posterior densities in the expression (3.3).

Note that the above conditional posterior distributions are very similar to that for the original Horseshoe Gibbs sampler K_{aug} given by (2.2) in Section 2. The only differences are

1. the matrix A appearing in (2.2) has been replaced by A_c (which is the matrix A plus the added regularization introduced in the prior conditional variance of $\boldsymbol{\beta}$ through the constant c) in (3.15), and
2. the form of the posterior conditional density of the global shrinkage parameter, namely, $\pi\left(\tau^2 \middle| \boldsymbol{\lambda}, \mathbf{y}\right)$ is different due to the additional term $c(\tau^2)^p$.

Theorem 3.3. *Suppose the prior density of the global shrinkage parameter is truncated below away from zero; that is, $\pi_\tau(u) = 0$ for $u < T$ for some $T > 0$ and satisfies*

$$\int_T^\infty u^{\frac{p+\delta}{2}} \pi_\tau(u) du < \infty$$

for some $\delta \in (0.00162, 0.22176)$. Then, the regularized Horseshoe Gibbs sampler corresponding to the transition kernel $\tilde{K}_{aug,reg}$ is geometrically ergodic.

The above theorem can be proved by essentially following verbatim the proof of Lemma 2.1 (which establishes geometric ergodicity for K_{aug}) with the same geometric drift function as in Lemma 2.1, and replacing the matrix A by the matrix A_c at relevant places. However, appropriate modifications are needed using the following two facts.

1. In the original Horseshoe setting, a uniform upper bound for the conditional posterior means of β_j's (see (2.14) for definition) was established in Proposition A.5 in Appendix A. However, in the current context, the added regularization of $c^{-2}I_p$ in A_c immediately provides the uniform upper bound without need for additional analysis.
2. The conditional posterior density $\pi\left(\tau^2 \middle| \boldsymbol{\lambda}_0, \mathbf{y}\right)$ is different from the original Horseshoe setting. Hence, the upper bound for the $\delta_0/2^{\text{th}}$ moment of this density for some $\delta_0 \in (0.00162, 0.22176)$ (see (2.12)) needs to be independently established. We have provided this bound in Proposition B.4 of Appendix B. Due to the presence of the additional term $c(\tau^2)^p$ in the conditional density, a stronger assumption of the existence of $(p+\delta_0)/2^{th}$ moment is required (as compared to the $\delta_0/2^{th}$ moment in Theorem 2.1 and Theorem 3.1).

Remark 3.1. *In [15], the authors focus on Bayesian logistic regression for their geometric ergodicity analysis. They use the regularized Horseshoe prior in (3.12) without the parameter σ^2 as their is no need for an error variance parameter for the Binomial likelihood. However, for computational purposes, additional parameters $\boldsymbol{\omega} = \{\omega_j\}_{j=1}^p$ with Polya-Gamma prior distributions are introduced. A two-block Gibbs sampler with blocks $(\boldsymbol{\beta}, \boldsymbol{\lambda})$ and $(\boldsymbol{\omega}, \tau^2)$ is then constructed and its geometric ergodicity is then established assuming that the global shrinkage parameter τ^2 is bounded away from zero and infinity [15, Theorem 4.6].*

Many details of this analysis break down when translating to the Bayesian linear regression framework considered in our paper. The parameters $\boldsymbol{\omega}$ are now replaced by the error variance parameter σ^2. One can still construct a two-block Gibbs sampler with blocks $(\boldsymbol{\beta}, \boldsymbol{\lambda})$ and (σ^2, τ^2), but many conditional independence and other algebraic niceties involving $\boldsymbol{\omega}$ which are crucial in establishing *the minorization condition in the logistic regression context, do not hold analogously with σ^2 in the linear regression context. The structural differences also imply that the drift condition with the function $\sum_{j=1}^{p} |\beta_j|^{-\delta}$ does not work out in the linear regression setting.*

Remark 3.2. *The geometric ergodicity result (Theorem 3.3) corresponding to the regularized Horseshoe variant in [15] requires truncation of the global shrinkage parameter τ^2 below away from zero. Such an assumption is not required for the geometric ergodicity result (Theorem 3.1) corresponding to the regularized Horseshoe of [18]. Also, due to the presence of the additional term $(c(\tau^2))^p$ in $\pi\left(\tau^2 \middle| \boldsymbol{\lambda}_0, \mathbf{y}\right)$, a stronger moment assumption is required for Theorem 3.3 as compared to Theorem 3.1.*

Remark 3.3. ***(Variant with Metropolis step for τ^2)****: The transition kernel $K_{MGN,q}$, where the exact sampling step for τ^2 in $\tilde{K}_{aug,reg}$ is replaced by a Metropolis-Hastings step with proposal density q, is of interest in high-dimensional settings. Geometric ergodicity of the Markov chain corresponding to $\tilde{K}_{MGN,q}$ under the conditions in Theorem 3.3 and those on q in Theorem 3.2 can be established by essentially following the arguments in the proof of Theorem 3.2 with minor adjustments. The only additional complication is the presence of the $c(\tau^2)^p$ term in $\pi\left(\tau^2 \middle| \boldsymbol{\lambda}, \mathbf{y}\right)$. This can be dealt with by observing that $1 \leq c(\tau^2) \leq C \max(1, \tau)$ for an appropriate constant C, and using either (symmetrized) Generalized Inverse Gaussian choices for π_ξ, π_ζ in Part (b)(ii,iii) of Theorem 3.2 or by truncating π_τ above.*

3.4. A simulation study

The primary objective of this study is to examine the practical feasibility/scalability of the two regularized Horseshoe Gibbs samplers described in Sections 3.1 and 3.3. We consider a simulation setting with $n = 100$ samples and $p = 750$ variables. We generate 10 replicated datasets following exactly the same procedure as outlined in Section 2.3. For each of these 10 datasets, we run four Gibbs samplers each: the Gibbs sampler for the regularized Horseshoe in [18] with $c = 1$ and $c = 100$, and the Gibbs sample for the regularized Horseshoe variant in [15] with $c = 1$ and $c = 100$. For all samplers, for computational convenience, we used a Metropolis step for sampling τ^2 as described in Part (b)(iii) of Theorem 3.1 (see also Remark 3.3).

All Markov chains were run for 5000 iterations. Cumulative average plots and trace plots were used to monitor and confirm sufficient mixing of all the Markov chains. In all the settings, and across all the replications, the Gibbs samplers roughly needed 1000 seconds to complete the required 5000 iterations. The essential sample sizes and MCMC standard errors based on 5000 iterations for the first ten components of $\boldsymbol{\beta}$ (which are the only non-zero components in the true $\boldsymbol{\beta}_0$) for various Gibbs samplers are provided in Tables 5, 6, 7, 8 and Tables 9, 10, 11, 12 respectively. Note that all the four Gibbs samplers considered here correspond to different targeted posterior distributions.

Table 5
EFFECTIVE SAMPLE SIZES: REGULARIZED HORSESHOE WITH c=1

	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
Dataset 1	2008	2121	1368	1399	465	934	227	3824	2049	5618
Dataset 2	1813	215	687	208	61	88	306	320	2063	1414
Dataset 3	1211	750	1726	1760	625	55	502	177	3390	1631
Dataset 4	1337	2352	2497	1582	116	754	296	818	5015	3209
Dataset 5	1627	1600	1771	395	47	418	105	4368	1591	704
Dataset 6	441	1452	1075	159	118	50	742	545	2342	2007
Dataset 7	1084	1877	1588	2126	167	205	237	2848	5970	1555
Dataset 8	2399	3828	1232	1696	91	222	941	6634	1779	4801
Dataset 9	1975	1697	1369	1713	131	1909	133	3119	2530	4980
Dataset 10	3462	2532	2739	2294	373	264	1264	4109	1770	2347

Effective sample size for the first 10 regression coefficients for the proposed regularized sampler with $c = 1$ for 10 simulated datasets with $n = 100$ and $p = 750$. A burn in of 200 was applied before calculating the effective sample sizes. All datasets are simulated from a 'true' linear regression model where only the first 10 regression coefficients are non-zero, and the entries of the design matrix X and the error vector are all independent Gaussian (both the design matrix and the error vector are generated separately for each dataset).

Table 6
EFFECTIVE SAMPLE SIZES: REGULARIZED HORSESHOE WITH c=100

	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
Dataset 1	296	526	314	355	90	295	106	944	1560	1879
Dataset 2	560	303	253	272	237	131	343	147	2566	1314
Dataset 3	610	716	534	496	505	223	539	152	2899	1301
Dataset 4	250	104	152	314	66	144	46	303	1506	2211
Dataset 5	681	464	424	649	271	98	58	2232	747	386
Dataset 6	229	381	379	562	239	50	299	117	543	1954
Dataset 7	529	530	517	377	92	128	322	2832	3626	1332
Dataset 8	950	733	322	476	73	211	824	1638	990	945
Dataset 9	807	374	235	335	53	699	66	1167	1015	2574
Dataset 10	501	932	1064	508	56	106	1066	1213	1480	612

Effective sample size for the first 10 regression coefficients for the proposed regularized sampler with $c = 100$ for 10 simulated datasets with $n = 100$ and $p = 750$. A burn in of 200 was applied before calculating the effective sample sizes. All datasets are simulated from a 'true' linear regression model where only the first 10 regression coefficients are non-zero, and the entries of the design matrix X and the error vector are all independent Gaussian (both the design matrix and the error vector are generated separately for each dataset).

We also tried to use the Hamiltonian Monte Carlo based algorithm for the regularized Horseshoe in [18], as implemented in the *R* package *hsstan*. However, the maximum treedepth (set to 10) is exceeded in *all* of the 5000 iterations. This issue persists even after warming up for up to 7000 iterations, and then running for 5000 more iterations. As we understand, this indicates poor adaptation, and raises questions about adequate posterior exploration and mixing of the Markov chain. A proposed remedy in this setting (Chapter 15.2 of the Stan reference manual on *mc-stan.org*) is to increase the tree depth. The *hsstan* function, however, did not allow us to pass the *max_treedepth* or *max_depth* as a parameter and change its value. Anyway, from the point of view of scalability, the time taken per iteration with maximum treedepth 10 was roughly one-third as compared to the various Gibbs samplers. When the maximum treedepth is

TABLE 7
EFFECTIVE SAMPLE SIZES: REGULARIZED HORSESHOE (NISHIMURA) WITH c=1

	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
Dataset 1	73	106	182	144	174	95	884	1034	2602	1061
Dataset 2	82	86	139	365	119	161	2767	227	1073	538
Dataset 3	65	75	61	288	41	82	342	403	1463	292
Dataset 4	91	95	137	247	104	97	2178	1519	2521	5093
Dataset 5	119	132	176	101	114	420	1452	926	990	2265
Dataset 6	102	87	124	198	111	481	249	2262	1453	489
Dataset 7	68	98	154	93	94	385	2965	721	1314	1770
Dataset 8	58	69	195	124	63	578	688	1767	188	12578
Dataset 9	103	141	268	257	196	93	1077	503	623	4821
Dataset 10	71	110	117	176	112	102	297	13000	4239	384

Effective sample size for the first 10 regression coefficients for the for the regularized sampler proposed by Nishimura et.el. with $c = 1$ for 10 simulated datasets with $n = 100$ and $p = 750$. All datasets are simulated from a 'true' linear regression model where only the first 10 regression coefficients are non-zero, and the entries of the design matrix X and the error vector are all independent Gaussian (both the design matrix and the error vector are generated separately for each dataset).

TABLE 8
EFFECTIVE SAMPLE SIZES: REGULARIZED HORSESHOE (NISHIMURA) WITH c=100

	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
Dataset 1	476	547	204	546	85	371	110	240	3320	4372
Dataset 2	455	217	347	263	68	97	84	1461	224	1025
Dataset 3	1890	338	708	719	392	149	990	571	2570	974
Dataset 4	855	737	442	686	338	671	455	2843	1655	2270
Dataset 5	570	468	883	563	316	372	1143	728	1282	801
Dataset 6	372	175	244	218	64	52	530	337	1732	901
Dataset 7	495	440	226	329	105	120	77	972	1848	422
Dataset 8	631	1011	613	325	227	126	2665	4139	2605	1293
Dataset 9	1176	447	497	329	113	140	784	629	3614	662
Dataset 10	246	130	154	224	687	96	768	268	2693	318

Effective sample size for the first 10 regression coefficients for the for the regularized sampler proposed by Nishimura et.el. with $c = 100$ for 10 simulated datasets with $n = 100$ and $p = 750$. All datasets are simulated from a 'true' linear regression model where only the first 10 regression coefficients are non-zero, and the entries of the design matrix X and the error vector are all independent Gaussian (both the design matrix and the error vector are generated separately for each dataset).

increased appropriately to resolve the issue pointed out above, it is very likely that the time taken per iteration will be more than those of the Gibbs samplers (increasing the tree-depth by 1 in the No U-turn HMC sampler effectively doubles the computation time).

To conclude, the Gibbs samplers described in Sections 3.1 and 3.3 provide practically feasible approaches which are computationally competitive with the HMC based approach. The geometric ergodicity results in Theorems 3.1 and 3.3 help provide the practitioner with asymptotically valid standard error estimates for corresponding MCMC based approximations to posterior quantities of interest.

Table 9
MCMC STANDARD ERRORS: REGULARIZED HORSESHOE WITH c=1

	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
Dataset 1	0.0058	0.0056	0.0071	0.0066	0.0145	0.002	0.0106	4e-04	8e-04	5e-04
Dataset 2	0.0054	0.0187	0.0092	0.0184	0.0525	0.0398	0.0046	0.0061	0.0011	0.001
Dataset 3	0.0078	0.011	0.0059	0.0067	0.0118	0.048	0.0032	0.024	7e-04	0.0011
Dataset 4	0.0074	0.005	0.0055	0.0071	0.0274	0.0023	0.0084	0.0023	5e-04	7e-04
Dataset 5	0.0068	0.0068	0.007	0.0128	0.0706	0.0044	0.0243	6e-04	0.0013	0.0027
Dataset 6	0.0128	0.0066	0.0078	0.0246	0.0281	0.0538	0.0023	0.0036	8e-04	7e-04
Dataset 7	0.0072	0.0056	0.0066	0.0058	0.0253	0.0174	0.008	7e-04	5e-04	0.0011
Dataset 8	0.0049	0.0041	0.0089	0.006	0.03	0.0132	0.0015	4e-04	9e-04	4e-04
Dataset 9	0.0058	0.0064	0.0067	0.006	0.0169	0.0011	0.021	6e-04	7e-04	5e-04
Dataset 10	0.0049	0.0052	0.0052	0.0064	0.0064	0.0122	0.0016	6e-04	0.0011	9e-04

MCMC standard error for the first 10 regression coefficients for the proposed regularized sampler with $c = 1$ for 10 simulated datasets with $n = 100$ and $p = 750$. A burn in of 200 was applied before calculating the effective sample sizes. All datasets are simulated from a 'true' linear regression model where only the first 10 regression coefficients are non-zero, and the entries of the design matrix X and the error vector are all independent Gaussian (both the design matrix and the error vector are generated separately for each dataset).

Table 10
MCMC STANDARD ERRORS: REGULARIZED HORSESHOE WITH c=100

	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
Dataset 1	0.0135	0.01	0.0134	0.0121	0.032	0.0064	0.0264	0.0017	0.0015	0.0011
Dataset 2	0.0083	0.0129	0.0127	0.0131	0.0163	0.0283	0.005	0.0148	8e-04	0.0013
Dataset 3	0.0095	0.0091	0.0091	0.0111	0.0101	0.0147	0.0037	0.0208	8e-04	0.0014
Dataset 4	0.015	0.0218	0.0191	0.0139	0.0489	0.0199	0.044	0.0039	0.0011	9e-04
Dataset 5	0.0094	0.0108	0.0128	0.0085	0.015	0.021	0.0436	0.0011	0.0025	0.004
Dataset 6	0.0155	0.0119	0.0122	0.0093	0.0136	0.0454	0.005	0.0125	0.0027	8e-04
Dataset 7	0.0093	0.0094	0.0101	0.0127	0.0326	0.0238	0.0078	7e-04	7e-04	0.0014
Dataset 8	0.0068	0.0083	0.0155	0.0102	0.0415	0.0155	0.0024	9e-04	0.0015	0.0015
Dataset 9	0.0076	0.0114	0.015	0.0119	0.0457	0.003	0.04	0.0013	0.0014	8e-04
Dataset 10	0.0122	0.008	0.0075	0.0125	0.0325	0.0268	0.002	0.0015	0.0014	0.003

MCMC standard error for the first 10 regression coefficients for the proposed regularized sampler (Prop.) with $c = 100$ for 10 simulated datasets with $n = 100$ and $p = 750$. A burn in of 200 was applied before calculating the effective sample sizes. All datasets are simulated from a 'true' linear regression model where only the first 10 regression coefficients are non-zero, and the entries of the design matrix X and the error vector are all independent Gaussian (both the design matrix and the error vector are generated separately for each dataset).

Appendix A: Uniform bound on $\boldsymbol{\mu}_j$

The goal of this subsection is to show that $\mu_j = \boldsymbol{e}_j^T A_0^{-1} \mathbf{X}^T \mathbf{y}$ defined in (2.14) is uniformly bounded in $\boldsymbol{\lambda}_0$ (even when $n < p$). This result will be established through a sequence of five propositions.

Proposition A.1. *Let $\Lambda \in \mathbb{R}^{p \times p}$ be any diagonal matrix with positive diagonal elements $\lambda_1, \lambda_2, \ldots, \lambda_p$. Let $X \in \mathbb{R}^{n \times p}$ be any matrix with rank r. Let the singular value decomposition of X is $X = UDV^T$ where $D \in \mathbb{R}^{r \times r}$ is diagonal matrix with positive diagonal elements $d_1, \ldots d_r$ while $U \in \mathbb{R}^{n \times r}$ and $V \in \mathbb{R}^{p \times r}$ are such that $U^T U = I$ and $V^T V = I$. If $\lambda \leq \min\{\lambda_1, \ldots, \lambda_p\}$ be any positive*

Table 11
MCMC STANDARD ERRORS: REGULARIZED HORSESHOE (NISHIMURA) WITH c=1

	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
Dataset 1	0.0665	0.0374	0.0277	0.0332	0.0325	0.0427	0.0035	0.0025	0.001	0.0019
Dataset 2	0.0567	0.0474	0.0301	0.0161	0.039	0.019	9e-04	0.0051	0.0013	0.0043
Dataset 3	0.0685	0.0519	0.0573	0.0169	0.0682	0.0379	0.0037	0.0032	0.0012	0.0078
Dataset 4	0.0573	0.0495	0.0305	0.0223	0.0491	0.026	0.0011	0.0014	0.0011	7e-04
Dataset 5	0.0454	0.0347	0.0278	0.0414	0.0469	0.0049	0.0017	0.0021	0.0021	0.001
Dataset 6	0.0473	0.0421	0.0323	0.0215	0.0376	0.0038	0.0052	0.0012	0.0012	0.0025
Dataset 7	0.0591	0.0432	0.0316	0.0347	0.0416	0.0079	7e-04	0.0027	0.0014	9e-04
Dataset 8	0.0813	0.049	0.0219	0.0287	0.0351	0.0025	0.0019	0.001	0.0054	4e-04
Dataset 9	0.0496	0.0339	0.0208	0.0188	0.025	0.0307	0.0019	0.0043	0.0042	6e-04
Dataset 10	0.0614	0.0394	0.0338	0.0241	0.0348	0.0286	0.0102	0.0003	0.0009	0.003

MCMC standard error for the first 10 regression coefficients for the regularized sampler proposed by Nishimura et. al. [15] with $c = 1$ for 10 simulated datasets with $n = 100$ and $p = 750$. All datasets are simulated from a 'true' linear regression model where only the first 10 regression coefficients are non-zero, and the entries of the design matrix X and the error vector are all independent Gaussian (both the design matrix and the error vector are generated separately for each dataset).

Table 12
MCMC STANDARD ERRORS: REGULARIZED HORSESHOE WITH c=100

	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
Dataset 1	0.0173	0.011	0.0208	0.0099	0.0349	0.0041	0.0257	0.0089	7e-04	7e-04
Dataset 2	0.0138	0.0219	0.0131	0.0153	0.0427	0.0341	0.0365	0.0017	0.009	0.0015
Dataset 3	0.0065	0.0164	0.0079	0.0076	0.0115	0.0296	0.0016	0.0034	8e-04	0.0012
Dataset 4	0.0094	0.0101	0.0112	0.0078	0.0126	0.0027	0.0042	7e-04	0.0013	9e-04
Dataset 5	0.0117	0.0149	0.0091	0.011	0.0164	0.0051	0.0016	0.0026	0.0017	0.0017
Dataset 6	0.0165	0.0227	0.018	0.0174	0.0361	0.0453	0.0038	0.0066	0.0012	0.0016
Dataset 7	0.0146	0.013	0.017	0.0142	0.0285	0.0259	0.0457	0.002	0.0013	0.0046
Dataset 8	0.0163	0.0079	0.0101	0.0135	0.018	0.0247	7e-04	7e-04	0.001	0.0013
Dataset 9	0.0096	0.0162	0.0131	0.0179	0.0288	0.0188	0.0028	0.003	9e-04	0.0032
Dataset 10	0.0248	0.0386	0.0295	0.0233	0.009	0.0298	0.0027	0.0056	9e-04	0.0063

MCMC standard error for the first 10 regression coefficients for the regularized sampler proposed by Nishimura et. al. [15] with $c = 100$ for 10 simulated datasets with $n = 100$ and $p = 750$. All datasets are simulated from a 'true' linear regression model where only the first 10 regression coefficients are non-zero, and the entries of the design matrix X and the error vector are all independent Gaussian (both the design matrix and the error vector are generated separately for each dataset).

number then for arbitrary $y \in \mathbb{R}^n$

$$y^T X(X^T X + \Lambda)^{-1} X^T y \leq y^T y - \|P_{U^\perp} y\|^2 - \sum_{i=1}^{r} \frac{\lambda \tilde{u}_i^2}{d_i^2 + \lambda},$$

where $\tilde{u}_i$ is the i^{th} component of the vector $\tilde{u} = U^T y$ and $P_{U^\perp}$ is the orthogonal projection matrix for the orthogonal complement of the column space of U.

Proof. Without loss of generality we assume that the matrix Λ is diagonal matrix with diagonal elements $\lambda_1, \lambda_2, \ldots, \lambda_p$ where $0 < \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_p$. According to the condition of the result $\lambda \leq \lambda_1$. Now we define a set of diagonal matrices $\{\Lambda_{(1)}, \ldots, \Lambda_{(j)}, \ldots, \Lambda_{(p)}\}$ in the following manner. For $j = 1, \ldots (p-1)$, the matrix $\Lambda_{(j)}$ has first j diagonal elements to be identical and equal to λ while rest of the $(p-j)$ diagonal elements are identical as that of the matrix Λ. Also let $\Lambda_{(p)} = \lambda I_{p\times p}$ and $\Lambda_{(0)} = \Lambda$. The above set of matrices satisfy the following relation

$$\Lambda_{(j-1)} = \Lambda_{(j)} + (\lambda_j - \lambda)e_j e_j^T \text{ for } j = 1, \ldots, p$$

where $e_j \in \mathbb{R}^p$ denotes the j^{th} elementary vector. Now using the Sherman-Woodbury formula for inverting matrices we get that

$$\begin{aligned} & \left(X^T X + \Lambda_{(j-1)}\right)^{-1} \\ = \; & \left(X^T X + \Lambda_{(j)} + (\lambda_j - \lambda)e_j e_j^T\right)^{-1} \\ = \; & \left(X^T X + \Lambda_{(j)}\right)^{-1} - \frac{(\lambda_j - \lambda)\left(X^T X + \Lambda_{(j)}\right)^{-1} e_j e_j^T \left(X^T X + \Lambda_{(j)}\right)^{-1}}{1 - (\lambda_j - \lambda)e_j^T \left(X^T X + \Lambda_{(j)}\right)^{-1} e_j}. \end{aligned}$$

Consequently,

$$\begin{aligned} & \left(X^T X + \Lambda_{(j-1)}\right)^{-1} - \left(X^T X + \Lambda_{(j)}\right)^{-1} \\ & = -\frac{(\lambda_j - \lambda)\left(X^T X + \Lambda_{(j)}\right)^{-1} e_j e_j^T \left(X^T X + \Lambda_{(j)}\right)^{-1}}{1 - (\lambda_j - \lambda)e_j^T \left(X^T X + \Lambda_{(j)}\right)^{-1} e_j}, \end{aligned} \tag{A.1}$$

Aggregating the equations (A.1) over $j = 1, \ldots p$, we get that

$$\begin{aligned} & \left(X^T X + \Lambda\right)^{-1} \\ & = \left(X^T X + \lambda I\right)^{-1} - \sum_{j=1}^{p} \frac{(\lambda_j - \lambda)\left(X^T X + \Lambda_{(j)}\right)^{-1} e_j e_j^T \left(X^T X + \Lambda_{(j)}\right)^{-1}}{1 - (\lambda_j - \lambda)e_j^T \left(X^T X + \Lambda_{(j)}\right)^{-1} e_j}, \end{aligned} \tag{A.2}$$

where we have used the fact that $\Lambda_{(0)} = \Lambda$ and $\Lambda_{(p)} = \lambda I$. If $y \in \mathbb{R}^n$ be arbitrary vector then it follows from (A.2) that

$$\begin{aligned} & y^T X \left(X^T X + \Lambda\right)^{-1} X^T y \\ & = y^T X \left(X^T X + \lambda I\right)^{-1} X^T y \\ & \quad - \sum_{j=1}^{p} \frac{(\lambda_j - \lambda) y^T X \left(X^T X + \Lambda_{(j)}\right)^{-1} e_j e_j^T \left(X^T X + \Lambda_{(j)}\right)^{-1} X^T y}{1 - (\lambda_j - \lambda)e_j^T \left(X^T X + \Lambda_{(j)}\right)^{-1} e_j} \\ & = y^T X \left(X^T X + \lambda I\right)^{-1} X^T y - \sum_{j=1}^{p} \frac{(\lambda_j - \lambda)\|e_j^T \left(X^T X + \Lambda_{(j)}\right)^{-1} X^T y\|^2}{1 - (\lambda_j - \lambda)e_j^T \left(X^T X + \Lambda_{(j)}\right)^{-1} e_j} \\ & \leq y^T X \left(X^T X + \lambda I\right)^{-1} X^T y, \end{aligned} \tag{A.3}$$

because $\lambda_j \geq \lambda$. Now consider the singular value decomposition of the matrix $X = UDV^T$ where D is a diagonal matrix with positive diagonal elements

$d_1, \ldots, d_r$, $U \in \mathbb{R}^{n\times r}, V \in R^{p\times r}$ such that $U^TU = V^TV = I$. Also let $V^\perp \in \mathbb{R}^{p\times(p-r)}$ be such that the matrix $\left[V, V^\perp\right] \in \mathbb{R}^{p\times p}$ is a orthogonal matrix, i.e. the columns of $V^\perp$ constitutes a orthonormal basis for the orthogonal complement of the column space of V. Now consider the fact that

$$\begin{aligned}(X^TX + \lambda I)^{-1} &= \left[VDU^TUDV^T + \lambda I\right]^{-1} \\ &= \left[VD^2V^T + \lambda VV^T + \lambda V^\perp (V^\perp)^T\right]^{-1} \\ &= \left[V(D^2 + \lambda I)V^T + \lambda V^\perp (V^\perp)^T\right]^{-1} \\ &= V(D^2 + \lambda I)^{-1}V^T + \frac{1}{\lambda} V^\perp (V^\perp)^T.\end{aligned}$$

Note that $(V^\perp)^T V = 0$. Thus

$$\begin{aligned}& y^TX\left(X^TX + \lambda I\right)^{-1} X^Ty \qquad \text{(A.4)} \\ &= y^TX\left[V(D^2 + \lambda I)^{-1}V^T + \frac{1}{\lambda} V^\perp (V^\perp)^T\right] VDU^Ty \\ &= y^T(UDV^T)V(D^2 + \lambda I)^{-1}V^TVDU^Ty \\ &= y^TUD(D^2 + \lambda I)^{-1}DU^Ty \\ &= \sum_{i=1}^r \frac{d_i^2 \tilde{u}_i^2}{d_i^2 + \lambda}, \qquad \text{(A.5)}\end{aligned}$$

where $d_1, \ldots d_r > 0$ are the diagonal elements of the matrix D and $\tilde{u}_i$ is the i^{th} entry of the vector U^Ty. Let $U^\perp$ refers to the orthogonal completion of the matrix U then

$$\sum_{i=1}^r \tilde{u}_i^2 = y^TUU^Ty = y^Ty - y^TU^\perp (U^\perp)^Ty = y^Ty - \|P_{U^\perp} y\|^2, \qquad \text{(A.6)}$$

where $P_{U^\perp}$ denotes the orthogonal projection for the column space of $U^\perp$. Finally, it follows from (A.3), (A.4) and (A.6) that

$$y^TX\left(X^TX + \Lambda\right)^{-1} X^Ty \le \sum_{i=1}^r \tilde{u}_i^2 - \sum_{i=1}^r \frac{\lambda \tilde{u}_i^2}{d_i^2 + \lambda} = y^Ty - \|P_{U^\perp} y\|^2 - \sum_{i=1}^r \frac{\lambda \tilde{u}_i^2}{d_i^2 + \lambda}.$$

Note that $\|P_{U^\perp} y\|^2 + \sum_{i=1}^r \frac{\lambda \tilde{u}_i^2}{d_i^2+\lambda} > 0$ beacuse $\sum_{i=1}^r \tilde{u}_i^2 + \|P_{U^\perp} y\|^2 = y^Ty > 0$. □

Proposition A.2. *Let $X = [\mathbf{x}_1, \ldots, \mathbf{x}_p] \in \mathbb{R}^{n\times p}$ and $X_2 = [\mathbf{x}_2, \ldots, \mathbf{x}_p]$. Assume $X_2 = UDV^T$ be the singular value decomposition where $d_1, \ldots d_r > 0$ are the diagonal elements of D. Let $\delta_1 > 0$ and Δ_{p-1} be any diagonal matrix with* *positive diagonal elements* $\delta_2, \ldots, \delta_p$*. If* $T := (\|\mathbf{x}_1\|^2 + \delta_1) - \mathbf{x}_1^T X_2 (X_2^T X_2 + \Delta_{p-1})^{-1} X_2 \mathbf{x}_1$ *then for any* $0 < \delta \leq \min\{\delta_1, \ldots, \delta_p\}$*,*

$$T \geq \delta_1 + \|P_{U^\perp}\mathbf{x}_1\|^2 + \sum_{i=1}^{r} \frac{\delta \tilde{u}_i^2}{d_i^2 + \delta},$$

where $\tilde{u}_i$ *is the* i^{th} *component of the vector* $\tilde{u} = U^T \mathbf{x}_1$ *and* $P_{U^\perp}$ *is the orthogonal projection matrix for the orthogonal complement of the column space of* U*.*

Proof. Using Result A.1, for any $\delta \leq \min\{\delta_1, \ldots, \delta_p\}$, we get that

$$\mathbf{x}_1^T X_2 (X_2^T X_2 + \Delta_{p-1})^{-1} X_2 \mathbf{x}_1 \leq \mathbf{x}_1^T \mathbf{x}_1 - \|P_{U^\perp}\mathbf{x}_1\|^2 - \sum_{i=1}^{r} \frac{\delta \tilde{u}_i^2}{d_i^2 + \delta}$$

Consequently $T = (\mathbf{x}_1^T \mathbf{x}_1 + \delta_1) - \mathbf{x}_1^T X_2 (X_2^T X_2 + \Delta_{p-1})^{-1} X_2 \mathbf{x}_1 \geq \delta_1 + \|P_{U^\perp}\mathbf{x}_1\|^2 + \sum_{i=1}^{r} \frac{\delta \tilde{u}_i^2}{d_i^2+\delta}$. ☐

.

Proposition A.3. *Let* $X \in \mathbb{R}^{n \times p}$ *be arbitrary matrix and* $\Delta_p \in \mathbb{R}^{p \times p}$ *be any diagonal matrix with positive diagonal elements* $\delta_1, \ldots, \delta_p$*. Consider the following partition of the matrix*

$$X^T X + \Delta_p = \left[\begin{array}{c|c} \|\mathbf{x}_1\|^2 + \delta_1 & \mathbf{x}_1^T X_2 \\ \hline X_2^T \mathbf{x}_1 & X_2^T X_2 + \Delta_{p-1} \end{array}\right],$$

where Δ_{p-1} *is the diagonal matrix with diagonal elements* $\delta_2, \ldots, \delta_p$*. If* $(X_2^T X_2 + \Delta_{p-1})^{-1} \Delta_{p-1}$ *is uniformly bounded, then the first column of the matrix*

$$\left(X^T X + \Delta_p\right)^{-1} \Delta_p$$

is uniformly bounded. The notations $\mathbf{x}_1$ *and* X_2 *are as they are defined in the Result A.2.*

Proof. If we consider the partition of

$$X^T X + \Delta_p = \left[\begin{array}{c|c} \|\mathbf{x}_1\|^2 + \delta_1 & \mathbf{x}_1^T X_2 \\ \hline X_2^T \mathbf{x}_1 & X_2^T X_2 + \Delta_{p-1} \end{array}\right],$$

then the Schur complement of the first block of the matrix is given as

$$T = (\|\mathbf{x}_1\|^2 + \delta_1) - \mathbf{x}_1^T X_2 (X_2^T X_2 + \Delta_{p-1})^{-1} X_2^T \mathbf{x}_1.$$

Employing the inversion formula of the block matrices [12], we get that

$(X^T X + \Delta_p)^{-1} =$

$$\left[\begin{array}{c|c} \frac{1}{T} & \frac{\mathbf{x}_1^T X_2 (X_2^T X_2 + \Delta_{p-1})^{-1}}{T} \\ \hline \frac{(X_2^T X_2 + \Delta_{p-1})^{-1} X_2^T \mathbf{x}_1}{T} & (X_2^T X_2 + \Delta_{p-1})^{-1} + \frac{(X_2^T X_2 + \Delta_{p-1})^{-1} X_2^T \mathbf{x}_1 \mathbf{x}_1^T X_2 (X_2^T X_2 + \Delta_{p-1})^{-1}}{T} \end{array}\right].$$

In the next two bullet points, we are going to show if $(X_2^T X_2 + \Delta_{p-1})^{-1}\Delta_{p-1}$ is uniformly bounded then so is all the entries of the vector

$$\begin{bmatrix} \frac{\delta_1}{T} & \frac{\delta_1 (X_2^T X_2 + \Delta_{p-1})^{-1} X_2^T \mathbf{x}_1}{T} \end{bmatrix}^T,$$

which is the first column of the matrix $(X^T X + \Delta_p)^{-1}\Delta_p$.

- **To show** $0 < \frac{\delta_1}{T} \leq 1$ **:**
  It is evident from the Result A.2 that $T > 0$. Therefore $\frac{\delta_1}{T} > 0$ as $\delta_1 > 0$ as well. On the other hand, a direct implication of Result A.2 is that for $0 < \delta \leq \min\{\delta_1, \ldots, \delta_p\}$,

$$\frac{\delta_1}{T} \leq \frac{\delta_1}{\delta_1 + \|P_{U^\perp}\mathbf{x}_1\|^2 + \sum_{i=1}^r \frac{\delta \tilde{u}_i^2}{d_i^2 + \delta}} \leq 1. \tag{A.7}$$

  where the details about the notations $\tilde{u}_i$, $P_{U^\perp}, d_i$ can be found in Result A.2.

- **To show** $\frac{\delta_1 (X_2^T X_2 + \Delta_{p-1})^{-1} X_2^T \mathbf{x}_1}{T}$ **uniformly bounded:**
  Let $\mathbf{v}_1, \mathbf{v}_2$ be such that $\mathbf{x}_1 = \mathbf{v}_1 + \mathbf{v}_2$ where $\mathbf{v}_1$ belongs to the column space of X_2 and $\mathbf{v}_2$ belongs to the orthogonal complement of the column space of X_2. Therefore $\mathbf{v}_1 = X_2 l$ for some vector $l \in \mathbb{R}^{p-1}$. Consequently,

$$\begin{aligned} (X_2^T X_2 + \Delta_{p-1})^{-1} X_2^T \mathbf{x}_1 &= (X_2^T X_2 + \Delta_{p-1})^{-1} X_2^T (X_2 l + \mathbf{v}_2) \\ &= (X_2^T X_2 + \Delta_{p-1})^{-1} X_2^T X_2 l \\ &= \left[I - (X_2^T X_2 + \Delta_{p-1})^{-1}\Delta_{p-1}\right] l, \end{aligned}$$

  is uniformly bounded as we are assuming that the matrix $(X_2^T X_2 + \Delta_{p-1})^{-1}\Delta_{p-1}$ is uniformly bounded. Combining this fact along with (A.7), we conclude that all the entries of the vector $\frac{\delta_1 (X_2^T X_2 + \Delta_{p-1})^{-1} X_2^T \mathbf{x}_1}{T}$ are also uniformly bounded. ☐

Proposition A.4. *Let $X \in \mathbb{R}^{n\times p}$ be arbitrary matrix and $\Delta_p \in \mathbb{R}^{p\times p}$ be any diagonal matrix with positive diagonal elements $\delta_1, \ldots, \delta_p$. Then for arbitrary p and n the matrix $\left(X^T X + \Delta_p\right)^{-1}\Delta_p$ is uniformly bounded. Specifically*

$$\sup_{\delta_1,\ldots,\delta_p > 0} \left| e_i^T \left(X^T X + \Delta_p\right)^{-1} \Delta_p e_j \right| < C$$

where C is a finite constant that does not depend on $\delta_1, \ldots, \delta_p$.

Proof. We will show the result by induction on the integer k where the hypothesis of induction is as follows,

$\mathcal{H}(k)$: Let n be arbitrary positive integer. Then for any positive integer k, the matrix $\left(X^T X + \Delta_k\right)^{-1}\Delta_k$ is uniformly bounded for all $X \in \mathbb{R}^{n\times k}$ and arbitrary diagonal matrix with positive diagonal elements $\Delta_k \in \mathbb{R}^{k\times k}$.

Initial step: The hypothesis trivially holds for $k = 1$. We will show that $\mathcal{H}(k)$ is true for the case $k = 2$. Let $X = [\mathbf{x}_1, \mathbf{x}_2] \in \mathbb{R}^{n\times 2}$ and $\Delta_2 = \begin{bmatrix}\delta_1 & 0\\ 0 & \delta_2\end{bmatrix}$, $\delta_1, \delta_2 > 0$ be arbitrary. Define $A := \begin{bmatrix}a_{1,1} & a_{1,2}\\ a_{2,1} & a_{2,2}\end{bmatrix} := X^T X$ and then

$$\begin{aligned}&(A+\Delta_2)^{-1}\Delta_2\\ &\quad = \frac{1}{(a_{1,1}a_{2,2}-a_{1,2}a_{2,1})+\delta_1 a_{2,2}+\delta_2 a_{1,1}+\delta_1\delta_2}\begin{bmatrix}\delta_1(a_{2,2}+\delta_2) & -\delta_2 a_{2,1}\\ -\delta_1 a_{1,2} & \delta_2(a_{1,1}+\delta_1)\end{bmatrix}\end{aligned}$$

- Note that

  $$\begin{aligned}&\sup_{\delta_1,\delta_2>0}\left|e_1^T (A+\Delta_2)^{-1}\Delta_2 e_1\right|\\ &\qquad = \sup_{\delta_1,\delta_2>0}\frac{\delta_1(a_{2,2}+\delta_2)}{(a_{1,1}a_{2,2}-a_{1,2}a_{2,1})+\delta_1(a_{2,2}+\delta_2)+\delta_2 a_{1,1}} \le 1\end{aligned}$$

  because $(a_{1,1}a_{2,2}-a_{1,2}a_{2,1}) \ge 0$ as $X^T X$ is nonnegative definte matrix. Additionally $a_{1,1} = \|\mathbf{x}_1\|^2 \ge 0$ and $a_{2,2} = \|\mathbf{x}_2\|^2 \ge 0$, where $X = [\mathbf{x}_1, \mathbf{x}_2]$.

- $$\begin{aligned}&\sup_{\delta_1,\delta_2>0}\left|e_1^T (A+\Delta_2)^{-1}\Delta_2 e_2\right|\\ =&\sup_{\delta_1,\delta_2>0}\frac{|a_{2,1}\delta_2|}{(a_{1,1}a_{2,2}-a_{1,2}a_{2,1})+\delta_1(a_{2,2}+\delta_2)+\delta_2 a_{1,1}} \qquad (A.8)\\ \le&\ \frac{|a_{2,1}|}{a_{1,1}}\end{aligned}$$

  for the case when $a_{1,1} \neq 0$. On the contrary, if $a_{1,1} = 0$ then it follows from (A.8) that

  $$\sup_{\delta_1,\delta_2>0}\left|e_1^T (A+\Delta_2)^{-1}\Delta_2 e_2\right| = 0, \qquad (A.9)$$

  because $a_{1,1} = \|\mathbf{x}_1\|^2 = 0$ implies that $a_{2,1} = \mathbf{x}_2^T\mathbf{x}_1 = 0$.

In a similar fashion we can show that the absolute value of the other two entries of the matrix $\left(X^T X + \Delta_2\right)^{-1}\Delta_2$ can be bounded above by numbers that does not depend on δ_1, δ_2. Consequently $\mathcal{H}(k)$ holds for $k = 2$.

Induction step: Let $\mathcal{H}(k)$ holds for $k = 1, 2, \ldots, (p-1)$. We will show that the result holds for $k = p$ as well. Let $X \in \mathbb{R}^{n\times p}$ be arbitrary matrix and Δ_p for diagonal matrix with positive diagonal elements $\delta_1, \ldots, \delta_p$. Consider the partition of the matrices $X^T X + \Delta_p$ as follows

$$X^T X + \Delta_p = \left[\begin{array}{c|c}\|\mathbf{x}_1\|^2+\delta_1 & \mathbf{x}_1^T X_2\\ \hline X_2^T\mathbf{x}_1 & X_2^T X_2 + \Delta_{p-1}\end{array}\right], \qquad (A.10)$$

where $X_2, \mathbf{x}_1, \Delta_{p-1}$ are as it is in Result A.2. As it satisfies the conditions of the induction hypothesis $\mathcal{H}(p-1)$, the matrix $(X_2^T X_2 + \Delta_{p-1})^{-1}\Delta_{p-1}$ is uniformly bounded. Therefore using Result A.3, the first column of the matrix $(X^T X + \Delta_p)^{-1}\Delta_p$ is uniformly bounded.

In remaining of the proof, we show that the m^{th} column of $(X^T X + \Delta_p)^{-1}\Delta_p$ is uniformly bounded for any $m > 1$. Consider the permutation matrix $P_{1,m} = [e_m, e_2 \ldots, e_{m-1}, e_1 \ldots, e_p]$. Note that $P_{1,m}$ can be generated by exchanging the 1^{st} and m^{th} columns of an identity matrix. $P_{1,m}$ is a symmetric and orthogonal matrix, i.e. $P_{1,m}^T = P_{1,m}$ and $P_{1,m}^T P_{1,m} = P_{1,m}^2 = I$. Now consider

$$\begin{aligned} P_{1,m}(X^T X + \Delta_p)^{-1}\Delta_p P_{1,m} &= P_{1,m}(X^T X + \Delta_p)^{-1} P_{1,m} P_{1,m} \Delta_p P_{1,m} \\ &= (P_{1,m}^T X^T X P_{1,m} + P_{1,m}\Delta_p P_{1,m})^{-1} P_{1,m}\Delta_p P_{1,m} \\ &= ({X^*}^T X^* + \Delta_p^*)^{-1}\Delta_p^*, \end{aligned} \tag{A.11}$$

where the $X^* := X P_{1,m}$ is obtained by exchanging the first and m^{th} columns of X while $\Delta_p^* := P_{1,m}\Delta_p P_{1,m}$ is the diagonal matrix where the first and the m^{th} diagonal elements of Δ_p are exchanged. We can represent ${X^*}^T X^* + \Delta_p^*$ as

$$\left[\begin{array}{c|c} \|\mathbf{x}_1^*\|^2 + \delta_1^* & {\mathbf{x}_1^*}^T X_2^* \\ \hline {X_2^*}^T \mathbf{x}_1^* & {X_2^*}^T X_2 + \Delta_{p-1}^* \end{array}\right],$$

where the notations are equivalent to that of the ones in (A.10). The matrix $({X_2^*}^T X_2^* + \Delta_{p-1}^*)^{-1}\Delta_{p-1}^*$ satisfies the conditions of the induction hypothesis $\mathcal{H}(p-1)$, thus it is uniformly bounded. Therefore using Result A.3, the first column of the matrix $({X^*}^T X^* + \Delta_p^*)^{-1}\Delta_p^*$ is uniformly bounded as well. It follows from (A.11) that the permuted version of the first column of $({X^*}^T X^* + \Delta_p^*)^{-1}\Delta_p^*$ is

$$\begin{aligned} P_{1,m}\left[({X^*}^T X^* + \Delta_p^*)^{-1}\Delta_p^* e_1\right] &= P_{1,m}P_{1,m}(X^T X + \Delta_p)^{-1}\Delta_p P_{1,m} e_1 \\ &= \left[(X^T X + \Delta_p)^{-1}\Delta_p\right] e_m, \end{aligned}$$

which is the m^{th} column of $(X^T X + \Delta_p)^{-1}\Delta_p$. Therefore, we infer that all the columns of the matrix $(X^T X + \Delta_p)^{-1}\Delta_p$ are uniformly bounded and conclude that $\mathcal{H}(k)$ holds for the case $k = p$. □

Proposition A.5. *Let $X \in \mathbb{R}^{n\times p}$ be arbitrary matrix and $\Delta_p \in \mathbb{R}^{p\times p}$ be any diagonal matrix with positive diagonal elements $\delta_1, \ldots, \delta_p$. Then for arbitrary p and n*

1. *The matrix $\left(X^T X + \Delta_p\right)^{-1} X^T X$ is uniformly bounded. Specifically*

$$\sup_{\delta_1,\ldots,\delta_p>0} \left| e_i^T \left(X^T X + \Delta_p\right)^{-1} X^T X e_j \right| < C$$

where C is a finite constant that does not depend on $\delta_1, \ldots, \delta_p$.

2. The vector $\left(X^TX+\Delta_p\right)^{-1}X^Ty$ *is uniformly bounded.*

Proof. part(1): Note that $\left(X^TX+\Delta_p\right)^{-1}X^TX = I - \left(X^TX+\Delta_p\right)^{-1}\Delta_p$. Using Result A.4 we know that the matrix $\left(X^TX+\Delta_p\right)^{-1}\Delta_p$ is uniformly bounded. Consequently $\left(X^TX+\Delta_p\right)^{-1}X^TX$ is also uniformly bounded.

part(2):

Let $y=\mathbf{v}_1+\mathbf{v}_2$ where $\mathbf{v}_1$ belongs to the column space of X and $\mathbf{v}_2$ belongs to the perpendicular to the column space of X. Therefore $\mathbf{v}_1 = Xl$ for some vector $l\in\mathbb{R}^{p-1}$. Consequently,

$$\begin{aligned}(X^TX+\Delta_p)^{-1}X^Ty &= (X^TX+\Delta_p)^{-1}X^T(\mathbf{v}_1+\mathbf{v}_2)\\ &= (X^TX+\Delta_p)^{-1}X^T(Xl+\mathbf{v}_2)\\ &= \left[(X^TX+\Delta_p)^{-1}X^TX\right]l.\end{aligned}$$

Therefore part(a) of the result ensures that the $(X^TX+\Delta_p)^{-1}X^Ty$ is uniformly bounded. ☐

Appendix B: Other technical results

Proposition B.1. *Let* δ *be chosen as in Lemma 2.1. Then for any* $\epsilon>0$ *there exists* $C_1>0$ *(not depending on* $\boldsymbol{\lambda}_0$*) such that*

$$\boldsymbol{E}\left[\left(\tau^2\right)^{\frac{\delta}{2}}\middle|\boldsymbol{\lambda}_0,\mathbf{y}\right]\le C_1$$

.

Proof. For any $\epsilon>0$, note that

$$\begin{aligned}\boldsymbol{E}\left[\left(\tau^2\right)^{\frac{\delta}{2}}\middle|\boldsymbol{\lambda}_0,\mathbf{y}\right] &= \boldsymbol{E}\left[\left(\tau^2\right)^{\frac{\delta}{2}}I_{\left[\tau^2<\epsilon\right]}\middle|\boldsymbol{\lambda}_0,\mathbf{y}\right]+\boldsymbol{E}\left[\left(\tau^2\right)^{\frac{\delta}{2}}I_{\left[\tau^2\ge\epsilon\right]}\middle|\boldsymbol{\lambda}_0,\mathbf{y}\right]\\ &\le \epsilon^{\frac{\delta}{2}}+\boldsymbol{E}\left[\left(\tau^2\right)^{\frac{\delta}{2}}I_{\left[\tau^2\ge\epsilon\right]}\middle|\boldsymbol{\lambda}_0,\mathbf{y}\right] \qquad \text{(B.1)}\end{aligned}$$

Next we demonstrate an upper bound to the second term in (B.1).

$$\begin{aligned}&\boldsymbol{E}\left[\left(\tau^2\right)^{\frac{\delta}{2}}I_{\left[\tau^2\ge\epsilon\right]}\middle|\boldsymbol{\lambda}_0,\mathbf{y}\right]\\ &= \int_{\epsilon}^{\infty}\left(\tau^2\right)^{\frac{\delta}{2}}\pi\left(\tau^2\middle|\boldsymbol{\lambda}_0,\mathbf{y}\right)d\tau^2\\ &= \frac{\int_{\epsilon}^{\infty}\left(\tau^2\right)^{\frac{\delta}{2}}\left(\frac{\mathbf{y}^T\left(I_n-\mathbf{X}A_0^{-1}\mathbf{X}^T\right)\mathbf{y}}{2}+b\right)^{-\left(a+\frac{n}{2}\right)}\frac{\pi_\tau\left(\tau^2\right)}{\left|I_p+\tau^2\mathbf{X}^T\mathbf{X}\cdot\boldsymbol{\Lambda}_0\right|^{\frac{1}{2}}}d\tau^2}{\int_{0}^{\infty}\left(\frac{\mathbf{y}^T\left(I_n-\mathbf{X}A_0^{-1}\mathbf{X}^T\right)\mathbf{y}}{2}+b\right)^{-\left(a+\frac{n}{2}\right)}\frac{\pi_\tau\left(\tau^2\right)}{\left|I_p+\tau^2\mathbf{X}^T\mathbf{X}\cdot\boldsymbol{\Lambda}_0\right|^{\frac{1}{2}}}d\tau^2}\end{aligned}$$

$$\leq \left(1+\frac{\mathbf{y}^T\mathbf{y}}{b}\right)^{a+\frac{n}{2}} \frac{\int\limits_{\epsilon}^{\infty} \left(\tau^2\right)^{\frac{\delta}{2}} \frac{\pi_\tau\left(\tau^2\right)}{\left|I_p+\tau^2\mathbf{X}^T\mathbf{X}\cdot\mathbf{\Lambda}_0\right|^{\frac{1}{2}}} d\tau^2}{\int\limits_{0}^{\epsilon} \frac{\pi_\tau\left(\tau^2\right)}{\left|I_p+\tau^2\mathbf{X}^T\mathbf{X}\cdot\mathbf{\Lambda}_0\right|^{\frac{1}{2}}} d\tau^2}$$

$$\leq \left(1+\frac{\mathbf{y}^T\mathbf{y}}{b}\right)^{a+\frac{n}{2}} \frac{\int\limits_{\epsilon}^{\infty} \left(\tau^2\right)^{\frac{\delta}{2}} \frac{\pi_\tau\left(\tau^2\right)}{\left|I_p+\epsilon\mathbf{X}^T\mathbf{X}\cdot\mathbf{\Lambda}_0\right|^{\frac{1}{2}}} d\tau^2}{\int\limits_{0}^{\epsilon} \frac{\pi_\tau\left(\tau^2\right)}{\left|I_p+\epsilon\mathbf{X}^T\mathbf{X}\cdot\mathbf{\Lambda}_0\right|^{\frac{1}{2}}} d\tau^2}$$

$$\leq \frac{\left(1+\frac{\mathbf{y}^T\mathbf{y}}{b}\right)^{a+\frac{n}{2}}}{\int\limits_0^\epsilon \pi_\tau\left(\tau^2\right)d\tau^2} \int\limits_{\epsilon}^{\infty} \left(\tau^2\right)^{\frac{\delta}{2}} \pi_\tau\left(\tau^2\right) d\tau^2$$

$$< \infty. \tag{B.2}$$

This completes the proof with $C_1 = \epsilon^{\frac{\delta}{2}} + \frac{\left(1+\frac{\mathbf{y}^T\mathbf{y}}{b}\right)^{a+\frac{n}{2}}}{\int_0^\epsilon \pi_\tau(\tau^2)d\tau^2} \int_\epsilon^\infty \left(\tau^2\right)^{\frac{\delta}{2}} \pi_\tau\left(\tau^2\right) d\tau^2$. $\square$

Proposition B.2. *Suppose there exists a* $\delta > 0.00162$ *such that*

$$\int_0^\infty (\tau^2)^{-\frac{p+\delta}{2}} \pi_\tau(\tau^2) d\tau^2 < \infty.$$

Then for any $\epsilon > 0$ *there exists* $\tilde{C}_1 > 0$ *(not depending on* $\boldsymbol{\lambda}_0$*) such that*

$$\boldsymbol{E}\left[\left(\tau^2\right)^{\frac{-\delta}{2}} \middle| \boldsymbol{\lambda}_0, \mathbf{y}\right] \leq \tilde{C}_1.$$

Proof. For any $\epsilon > 0$, note that

$$\begin{aligned} \boldsymbol{E}\left[\left(\tau^2\right)^{-\frac{\delta}{2}} \middle| \boldsymbol{\lambda}_0, \mathbf{y}\right] &= \boldsymbol{E}\left[\left(\tau^2\right)^{-\frac{\delta}{2}} I_{\left[\tau^2<\epsilon\right]} \middle| \boldsymbol{\lambda}_0, \mathbf{y}\right] + \boldsymbol{E}\left[\left(\tau^2\right)^{\frac{\delta}{2}} I_{\left[\tau^2\geq\epsilon\right]} \middle| \boldsymbol{\lambda}_0, \mathbf{y}\right] \\ &\leq \epsilon^{-\frac{\delta}{2}} + \boldsymbol{E}\left[\left(\tau^2\right)^{-\frac{\delta}{2}} I_{\left[\tau^2\leq\epsilon\right]} \middle| \boldsymbol{\lambda}_0, \mathbf{y}\right] \end{aligned} \tag{B.3}$$

Next we demonstrate an upper bound to the second term in (B.3).

$$\begin{aligned} &\boldsymbol{E}\left[\left(\tau^2\right)^{-\frac{\delta}{2}} I_{[\tau^2\leq\epsilon]} \middle| \boldsymbol{\lambda}_0, \mathbf{y}\right] \\ =& \int\limits_0^\epsilon \left(\tau^2\right)^{-\frac{\delta}{2}} \pi\left(\tau^2 \middle| \boldsymbol{\lambda}_0, \mathbf{y}\right) d\tau^2 \\ =& \frac{\int\limits_0^\epsilon \left(\tau^2\right)^{-\frac{\delta}{2}} \left(\frac{\mathbf{y}^T\left(I_n - \mathbf{X}A_0^{-1}\mathbf{X}^T\right)\mathbf{y}}{2} + b\right)^{-\left(a+\frac{n}{2}\right)} \frac{\pi_\tau\left(\tau^2\right)}{|I_p+\tau^2\mathbf{X}^T\mathbf{X}\cdot\mathbf{\Lambda}_0|^{\frac{1}{2}}} d\tau^2}{\int\limits_0^\infty \left(\frac{\mathbf{y}^T\left(I_n - \mathbf{X}A_0^{-1}\mathbf{X}^T\right)\mathbf{y}}{2} + b\right)^{-\left(a+\frac{n}{2}\right)} \frac{\pi_\tau(\tau^2)}{|I_p+\tau^2\mathbf{X}^T\mathbf{X}\cdot\mathbf{\Lambda}_0|^{\frac{1}{2}}} d\tau^2} \end{aligned}$$

$$
\begin{aligned}
&\leq \left(1+\frac{\mathbf{y}^T\mathbf{y}}{b}\right)^{a+\frac{n}{2}} \frac{\int_0^{\epsilon} \left(\tau^2\right)^{-\frac{\delta}{2}} \frac{\pi_\tau\left(\tau^2\right)}{|I_p+\tau^2\mathbf{X}^T\mathbf{X}\cdot\mathbf{\Lambda}_0|^{\frac{1}{2}}} d\tau^2}{\int_\epsilon^{\infty} \frac{\pi_\tau(\tau^2)}{|I_p+\tau^2\mathbf{X}^T\mathbf{X}\cdot\mathbf{\Lambda}_0|^{\frac{1}{2}}} d\tau^2} \\
&= \left(1+\frac{\mathbf{y}^T\mathbf{y}}{b}\right)^{a+\frac{n}{2}} \frac{\int_0^{\epsilon} \left(\tau^2\right)^{-\frac{p+\delta}{2}} \frac{\pi_\tau\left(\tau^2\right)}{|\tau^{-2}I_p+\mathbf{X}^T\mathbf{X}\cdot\mathbf{\Lambda}_0|^{\frac{1}{2}}} d\tau^2}{\int_\epsilon^{\infty} (\tau^2)^{-\frac{p}{2}} \frac{\pi_\tau(\tau^2)}{|\tau^{-2}I_p+\mathbf{X}^T\mathbf{X}\cdot\mathbf{\Lambda}_0|^{\frac{1}{2}}} d\tau^2} \\
&\leq \left(1+\frac{\mathbf{y}^T\mathbf{y}}{b}\right)^{a+\frac{n}{2}} \frac{\int_0^{\epsilon} \left(\tau^2\right)^{-\frac{p+\delta}{2}} \frac{\pi_\tau\left(\tau^2\right)}{|\epsilon^{-1}I_p+\mathbf{X}^T\mathbf{X}\cdot\mathbf{\Lambda}_0|^{\frac{1}{2}}} d\tau^2}{\int_\epsilon^{\infty} (\tau^2)^{-\frac{p}{2}} \frac{\pi_\tau(\tau^2)}{|\epsilon^{-1}I_p+\mathbf{X}^T\mathbf{X}\cdot\mathbf{\Lambda}_0|^{\frac{1}{2}}} d\tau^2} \\
&\leq \left(1+\frac{\mathbf{y}^T\mathbf{y}}{b}\right)^{a+\frac{n}{2}} \frac{\int_0^{\epsilon} \left(\tau^2\right)^{-\frac{p+\delta}{2}} \pi_\tau\left(\tau^2\right)}{\int_\epsilon^{\infty} (\tau^2)^{-\frac{p}{2}} \pi_\tau(\tau^2)\, d\tau^2} \\
&< \infty. \quad \square
\end{aligned}
\tag{B.4}
$$

Proposition B.3. *Refer to (3.2) and Lemma 3.1. Then for any $\epsilon > 0$ and for any $\delta > 0$, there exists $C_2 > 0$ such that*

$$
\boldsymbol{E}\left[\left(\tau^2\right)^{\frac{\delta}{2}}\Big|\,\boldsymbol{\lambda}_0,\mathbf{y}\right] \leq C_2
$$

.

Proof. Fix an $\epsilon > 0$ and a $\delta > 0$.

$$
\begin{aligned}
\boldsymbol{E}\left[\left(\tau^2\right)^{\frac{\delta}{2}}\Big|\,\boldsymbol{\lambda}_0,\mathbf{y}\right] &= \boldsymbol{E}\left[\left(\tau^2\right)^{\frac{\delta}{2}} I_{[\tau^2<\epsilon]}\Big|\,\boldsymbol{\lambda}_0,\mathbf{y}\right] + \boldsymbol{E}\left[\left(\tau^2\right)^{\frac{\delta}{2}} I_{[\tau^2\geq\epsilon]}\Big|\,\boldsymbol{\lambda}_0,\mathbf{y}\right] \\
&\leq \epsilon^{\frac{\delta}{2}} + \boldsymbol{E}\left[\left(\tau^2\right)^{\frac{\delta}{2}} I_{[\tau^2\geq\epsilon]}\Big|\,\boldsymbol{\lambda}_0,\mathbf{y}\right]
\end{aligned}
\tag{B.5}
$$

Next we demonstrate an upper bound to the second term in (B.5).

$$
\begin{aligned}
&\boldsymbol{E}\left[\left(\tau^2\right)^{\frac{\delta}{2}} I_{[\tau^2\geq\epsilon]}\Big|\,\boldsymbol{\lambda}_0,\mathbf{y}\right] \\
&= \int_\epsilon^{\infty} \left(\tau^2\right)^{\frac{\delta}{2}} \pi\left(\tau^2\big|\,\boldsymbol{\lambda}_0,\mathbf{y}\right) d\tau^2 \\
&= \frac{\int_\epsilon^{\infty} \left(\tau^2\right)^{\frac{\delta}{2}} \left(\frac{\mathbf{y}^T\left(I_n-\mathbf{X}A_0^{-1}\mathbf{X}^T\right)\mathbf{y}}{2}+b\right)^{-\left(a+\frac{n}{2}\right)} \frac{\left|c^{-2}I_p+\left(\tau^2\mathbf{\Lambda}_0\right)^{-1}\right|^{1/2}}{\left|\mathbf{X}^T\mathbf{X}+c^{-2}I_p+(\tau^2\mathbf{\Lambda}_0)^{-1}\right|^{1/2}} \pi_\tau\left(\tau^2\right) d\tau^2}{\int_0^{\infty} \left(\frac{\mathbf{y}^T\left(I_n-\mathbf{X}A_0^{-1}\mathbf{X}^T\right)\mathbf{y}}{2}+b\right)^{-\left(a+\frac{n}{2}\right)} \frac{\left|c^{-2}I_p+(\tau^2\mathbf{\Lambda}_0)^{-1}\right|^{1/2}}{\left|\mathbf{X}^T\mathbf{X}+c^{-2}I_p+(\tau^2\mathbf{\Lambda}_0)^{-1}\right|^{1/2}} \pi_\tau\left(\tau^2\right) d\tau^2}
\end{aligned}
$$

$$\leq \left(1+\frac{\mathbf{y}^T\mathbf{y}}{b}\right)^{a+\frac{n}{2}} \frac{\int\limits_{\epsilon}^{\infty} \left(\tau^2\right)^{\frac{\delta}{2}} \frac{\left|c^{-2}I_p+\left(\tau^2\mathbf{\Lambda}_0\right)^{-1}\right|^{1/2}}{\left|\mathbf{X}^T\mathbf{X}+c^{-2}I_p+(\tau^2\mathbf{\Lambda}_0)^{-1}\right|^{1/2}} \pi_\tau\left(\tau^2\right) d\tau^2}{\int\limits_0^{\epsilon} \frac{\left|c^{-2}I_p+(\tau^2\mathbf{\Lambda}_0)^{-1}\right|^{1/2}}{\left|\mathbf{X}^T\mathbf{X}+c^{-2}I_p+(\tau^2\mathbf{\Lambda}_0)^{-1}\right|^{1/2}} \pi_\tau\left(\tau^2\right) d\tau^2} \tag{B.6}$$

$$\leq \frac{\left(1+\frac{\mathbf{y}^T\mathbf{y}}{b}\right)^{a+\frac{n}{2}}}{\int\limits_0^{\epsilon} \pi_\tau\left(\tau^2\right) d\tau^2} \int\limits_0^{\infty} \left(\tau^2\right)^{\frac{\delta}{2}} \pi_\tau\left(\tau^2\right) d\tau^2$$

$$<\infty,$$

Note that the ratio of two determinants inside the integral in the numerator and denominator in (B.6) can be represented in the form

$$\frac{|B_1+\tau^{-2}I_p|}{|B_1+B_2+\tau^{-2}I_p|} = \frac{\prod_{k=1}^p (s_k(B_1)+\tau^{-2})}{\prod_{k=1}^p (s_k(B_1+B_2)+\tau^{-2})}$$

for appropriate symmetric non-negative definite matrices B_1 and B_2, and their respective eigenvalues denoted by $s_k(\cdot)$. Since every eigenvalue of B_1 is bounded above by the corresponding eigenvalue of B_1+B_2, it follows that the ratio of determinants is a decreasing function of τ^2, and can be replaced by the value at $\tau^2=\epsilon$ in both places with the inequality going in the right direction. This completes the proof with

$$C_2 = \epsilon^{\frac{\delta}{2}} + \frac{\left(1+\frac{\mathbf{y}^T\mathbf{y}}{b}\right)^{a+\frac{n}{2}}}{\int\limits_0^{\epsilon} \pi_\tau\left(\tau^2\right) d\tau^2} \int\limits_0^{\infty} \left(\tau^2\right)^{\frac{\delta}{2}} \pi_\tau\left(\tau^2\right) d\tau^2$$

. ∎

Proposition B.4. *Let δ be chosen as in Theorem 3.3. Then for any $\epsilon>0$ there exists $C_3>0$ (not depending on $\boldsymbol{\lambda}_0$) such that*

$$\boldsymbol{E}\left[\left(\tau^2\right)^{\frac{\delta}{2}} \middle| \boldsymbol{\lambda}_0, \mathbf{y}\right] \leq C_3.$$

Proof. For any $\epsilon>0$ note that

$$\begin{aligned} \boldsymbol{E}\left[\left(\tau^2\right)^{\frac{\delta}{2}} \middle| \boldsymbol{\lambda}_0, \mathbf{y}\right] &= \boldsymbol{E}\left[\left(\tau^2\right)^{\frac{\delta}{2}} I_{[\tau^2<\epsilon]} \middle| \boldsymbol{\lambda}_0, \mathbf{y}\right] + \boldsymbol{E}\left[\left(\tau^2\right)^{\frac{\delta}{2}} I_{[\tau^2\geq\epsilon]} \middle| \boldsymbol{\lambda}_0, \mathbf{y}\right] \\ &\leq \epsilon^{\frac{\delta}{2}} + \boldsymbol{E}\left[\left(\tau^2\right)^{\frac{\delta}{2}} I_{[\tau^2\geq\epsilon]} \middle| \boldsymbol{\lambda}_0, \mathbf{y}\right]. \end{aligned} \tag{B.7}$$

Next we demonstrate an upper bound to the second term in (B.7).

$$\begin{aligned} &\boldsymbol{E}\left[\left(\tau^2\right)^{\frac{\delta}{2}} I_{[\tau^2\geq\epsilon]} \middle| \boldsymbol{\lambda}_0, \mathbf{y}\right] \\ = &\int\limits_{\epsilon}^{\infty} \left(\tau^2\right)^{\frac{\delta}{2}} \pi\left(\tau^2 \middle| \boldsymbol{\lambda}_0, \mathbf{y}\right) d\tau^2 \end{aligned}$$

$$
\begin{aligned}
&= \frac{\int\limits_{\epsilon}^{\infty} \left(\tau^2\right)^{\frac{\delta}{2}} \left(\frac{\mathbf{y}^T\left(I_n - \mathbf{X}A_0^{-1}\mathbf{X}^T\right)\mathbf{y}}{2} + b\right)^{-\left(a+\frac{n}{2}\right)} \frac{\pi_\tau\left(\tau^2\right)c(\tau^2)^p}{\left|\tau^2(\mathbf{X}^T\mathbf{X}+c^{-2}I_p)+\mathbf{\Lambda}_0^{-1}\right|^{\frac{1}{2}}} d\tau^2}{\int\limits_{T}^{\infty} \left(\frac{\mathbf{y}^T\left(I_n - \mathbf{X}A_0^{-1}\mathbf{X}^T\right)\mathbf{y}}{2} + b\right)^{-\left(a+\frac{n}{2}\right)} \frac{\pi_\tau(\tau^2)c(\tau^2)^p}{\left|\tau^2(\mathbf{X}^T\mathbf{X}+c^{-2}I_p)+\mathbf{\Lambda}_0^{-1}\right|^{\frac{1}{2}}} d\tau^2} \\
&\leq \left(1 + \frac{\mathbf{y}^T\mathbf{y}}{b}\right)^{a+\frac{n}{2}} \frac{\int\limits_{\epsilon}^{\infty} \left(\tau^2\right)^{\frac{\delta}{2}} \frac{\pi_\tau\left(\tau^2\right)c(\tau^2)^p}{\left|\tau^2(\mathbf{X}^T\mathbf{X}+c^{-2}I_p)+\mathbf{\Lambda}_0^{-1}\right|^{\frac{1}{2}}} d\tau^2}{\int\limits_{T}^{\epsilon} \frac{\pi_\tau(\tau^2)c(\tau^2)^p}{\left|\tau^2(\mathbf{X}^T\mathbf{X}+c^{-2}I_p)+\mathbf{\Lambda}_0^{-1}\right|^{\frac{1}{2}}} d\tau^2} \\
&\leq \left(1 + \frac{\mathbf{y}^T\mathbf{y}}{b}\right)^{a+\frac{n}{2}} \frac{\int\limits_{\epsilon}^{\infty} \left(\tau^2\right)^{\frac{\delta}{2}} \frac{\pi_\tau\left(\tau^2\right)c(\tau^2)^p}{\left|\epsilon(\mathbf{X}^T\mathbf{X}+c^{-2}I_p)+\mathbf{\Lambda}_0^{-1}\right|^{\frac{1}{2}}} d\tau^2}{\int\limits_{T}^{\epsilon} \frac{\pi_\tau(\tau^2)c(\tau^2)^p}{\left|\epsilon(\mathbf{X}^T\mathbf{X}+c^{-2}I_p)+\mathbf{\Lambda}_0^{-1}\right|^{\frac{1}{2}}} d\tau^2} \\
&\leq \tilde{C}^{\frac{p}{2}} \frac{\left(1 + \frac{\mathbf{y}^T\mathbf{y}}{b}\right)^{a+\frac{n}{2}}}{\int\limits_{T}^{\epsilon} \pi_\tau\left(\tau^2\right) d\tau^2} \int\limits_{T}^{\infty} \left(\tau^2\right)^{\frac{\delta}{2}} \left(1+\tau^2\right)^{\frac{p}{2}} \pi_\tau\left(\tau^2\right) d\tau^2. \qquad \text{(B.8)}
\end{aligned}
$$

The last inequality follows from the fact that

$$1 \leq c(\tau^2) \leq \tilde{C}\sqrt{1+\tau^2}$$

for an appropriate constant $\tilde{C}$. □

Proposition B.5. *$f : \mathbb{R}^p \mapsto [0, \infty)$ and $g : \mathbb{R}^p \mapsto (0, \infty)$ be two functions such that $\int_{\mathbb{R}^p} f(\boldsymbol{x})d\boldsymbol{x} < \infty$ and $0 < \int_{\mathbb{R}^p} f(\boldsymbol{x})g(\boldsymbol{x})d\boldsymbol{x} < \infty$. Then $\int_{\mathbb{R}^p} \frac{f(\boldsymbol{x})}{g(\boldsymbol{x})}d\boldsymbol{x} \geq \frac{\left(\int_{\mathbb{R}^p} f(\boldsymbol{x})d\boldsymbol{x}\right)^2}{\int_{\mathbb{R}^p} f(\boldsymbol{x})g(\boldsymbol{x})d\boldsymbol{x}}$.*

Proof. Follows from Cauchy-Schwarz inequality. □

Proposition B.6. *For any $j \in \{1, 2, \cdots, p\}$ and any $d > 0$, there exists some $\alpha > 0$ such that*

$$\int\limits_0^{\infty} \frac{\nu_j^{-2} \exp\left[-\frac{1}{\nu_j}\left(1 + d^{\frac{2}{\delta}} + \frac{1}{\lambda_j^2}\right)\right]}{\sqrt{\nu_j} + \frac{\sigma^2\sqrt{\tau^2}}{\beta_j^2}} d\nu_j \geq \alpha \frac{\left(1 + \frac{1}{\lambda_j^2}\right)^{-2}}{\left(1 + \frac{\sigma^2\sqrt{\tau^2}}{\beta_j^2}\right)}.$$

Proof. Follows from Proposition B.5 with

$f\left(\nu_j\right) = \nu_j^{-2} \exp\left[-\frac{1}{\nu_j}\left(1 + d^{\frac{2}{\delta}} + \frac{1}{\lambda_j^2}\right)\right]$,
$g\left(\nu_j\right) = \sqrt{\nu_j} + \frac{\sigma^2\sqrt{\tau^2}}{\beta_j^2}$,
and $\alpha = \left(1 + d^{2/\delta}\right)^{-2}$ □

Proposition B.7. *There exists a positive definite matrix M_{τ^2} such that*

$$\int\limits_{\mathbb{R}^p} \frac{\exp\left[-\frac{\left(\boldsymbol{\beta}-\Omega^{-1}\mathbf{X}^T\mathbf{y}\right)^T\Omega\left(\boldsymbol{\beta}-\Omega^{-1}\mathbf{X}^T\mathbf{y}\right)+\boldsymbol{\beta}^T\left(\tau^2\boldsymbol{\Lambda}\right)^{-1}\boldsymbol{\beta}}{2\sigma^2}\right]}{\prod\limits_{j=1}^{p}\left(1+\frac{\sigma^2\sqrt{\tau^2}}{\beta_j^2}\right)} d\boldsymbol{\beta}$$

$$\geq \left(2\pi\sigma^2\right)^{\frac{p}{2}} |c|^{-p} \mid M_{\tau^2} \mid^{-1} \left(1+\frac{\sqrt{\tau^2}}{c^2}\right)^{-p}$$

$$\times \exp\left[-\frac{\mathbf{y}^T\mathbf{X}\left(c^2 I_p+\Omega^{-1}-2M_{\tau^2}^{-1}\right)\mathbf{X}^T\mathbf{y}}{2\sigma^2}\right]$$

Proof. Follows from Proposition B.5 with

$$f(\boldsymbol{\beta}) = \exp\left[-\frac{\left(\boldsymbol{\beta}-\Omega^{-1}\mathbf{X}^T\mathbf{y}\right)^T\Omega\left(\boldsymbol{\beta}-\Omega^{-1}\mathbf{X}^T\mathbf{y}\right)+\boldsymbol{\beta}^T\left(\tau^2\boldsymbol{\Lambda}\right)^{-1}\boldsymbol{\beta}}{2\sigma^2}\right],$$

$g(\boldsymbol{\beta}) = \prod_{j=1}^{p}\left(1+\frac{\sigma^2\sqrt{\tau^2}}{\beta_j^2}\right)$,
and $M_{\tau^2} = \Omega + \left(\tau^2\boldsymbol{\Lambda}\right)^{-1}$ □

Appendix C: Minorization condition for Horseshoe Gibbs sampler

Lemma C.1. *For every $d > 0$, there exists a constant $\epsilon^* = \epsilon^*(V, d) > 0$ and a density function h on $\mathbb{R}_+^p$ such that*

$$k(\boldsymbol{\lambda}_0, \boldsymbol{\lambda}) \geq \epsilon^* h(\boldsymbol{\lambda}) \tag{C.1}$$

for every $\boldsymbol{\lambda}_0 \in B(V, d)$ (see Section 2.2 for definition).

Proof: Fix a $\boldsymbol{\lambda}_0 \in B(V, d)$. In order to prove (C.1) we will demonstrate appropriate lower bounds to the conditional densities appearing in (2.3). From (2.2) we have the following:

$$\pi\left(\tau^2\middle|\,\boldsymbol{\lambda}_0, \mathbf{y}\right) \geq \left(\frac{b}{\mathbf{y}^T\mathbf{y}/2+b}\right)^{a+\frac{n}{2}} \omega_*^{-p/2}\left(1+\tau^2\right)^{-p/2}\pi_\tau\left(\tau^2\right)$$

where $\omega_* = \max\left\{1, \bar{\omega}\cdot d^{2/\delta_0}\right\}$ (recall that $\bar{\omega}$ is the maximum eigenvalue of $\mathbf{X}^T\mathbf{X}$ and that the prior density π_τ is truncated below at some $T > 0$).

$$\pi\left(\boldsymbol{\nu}\middle|\,\boldsymbol{\lambda}_0, \mathbf{y}\right) \geq \prod_{j=1}^{p}\left\{\nu_j^{-2}\exp\left[-\frac{1}{\nu_j}\left(1+d^{\frac{2}{\delta_1}}\right)\right]\right\}$$

$$\pi\left(\sigma^2\middle|\,\tau^2, \boldsymbol{\lambda}_0, \mathbf{y}\right) \geq \frac{b^{a+\frac{n}{2}}}{\Gamma\left(a+\frac{n}{2}\right)}\left(\sigma^2\right)^{-\left(a+\frac{n}{2}\right)-1}\exp\left[-\frac{1}{\sigma^2}\left(\frac{\mathbf{y}^T\mathbf{y}}{2}+b\right)\right]$$

$$\pi\left(\boldsymbol{\beta}\middle|\,\sigma^2, \tau^2, \boldsymbol{\lambda}_0, \mathbf{y}\right) \geq \left(2\pi\sigma^2\right)^{-\frac{p}{2}} d^{-p/\delta_0}\left(\tau^2\right)^{-p/2}$$

$$\times \exp\left[-\frac{\left(\boldsymbol{\beta}-M^{-1}\mathbf{X}^T\mathbf{y}\right)^T M\left(\boldsymbol{\beta}-M^{-1}\mathbf{X}^T\mathbf{y}\right)+\mathbf{y}^T\left(I-\mathbf{X}M^{-1}\mathbf{X}\right)\mathbf{y}}{2\sigma^2}\right]$$

since,

$$\begin{aligned}
& \left(\boldsymbol{\beta}-A_0^{-1}\mathbf{X}^T\mathbf{y}\right)^T A_0\left(\boldsymbol{\beta}-A_0^{-1}\mathbf{X}^T\mathbf{y}\right) \\
= \;& \boldsymbol{\beta}^T A_0\boldsymbol{\beta}-2\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{y}+\mathbf{y}^T\mathbf{X}A_0^{-1}\mathbf{X}^T\mathbf{y} \\
\leq \;& \boldsymbol{\beta}^T M\boldsymbol{\beta}-2\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{y}+\mathbf{y}^T\mathbf{y} \\
= \;& \left(\boldsymbol{\beta}-M^{-1}\mathbf{X}^T\mathbf{y}\right)^T M\left(\boldsymbol{\beta}-M^{-1}\mathbf{X}^T\mathbf{y}\right)+\mathbf{y}^T\left(I-\mathbf{X}M^{-1}\mathbf{X}^T\right)\mathbf{y}
\end{aligned}$$

where $M=\omega^*\left(1+\frac{1}{\tau^2}\right)I_p$ and $\omega^*=\max\left\{\bar{\omega},d^{2/\delta_1}\right\}$. Finally,

$$\pi\left(\boldsymbol{\lambda}\mid\boldsymbol{\beta},\boldsymbol{\nu},\sigma^2,\tau^2,\mathbf{y}\right)\geq\prod_{j=1}^{p}\left\{\frac{\beta_j^2}{2\sigma^2\tau^2}\left(\lambda_j^2\right)^{-2}\exp\left[-\frac{1}{\lambda_j^2}\left(\frac{1}{\nu_j}+\frac{\beta_j^2}{2\sigma^2\tau^2}\right)\right]\right\} \tag{C.2}$$

Putting all lower bounds in (C.2) in the equation of MTD (2.3) we have:

$$\begin{aligned}
& k\left(\boldsymbol{\lambda}_0,\boldsymbol{\lambda}\right) \\
& \geq (2\pi)^{-\frac{p}{2}}\left(\omega_*\right)^{-p/2}d^{-p/\delta_0}\left(\frac{b}{\mathbf{y}^T\mathbf{y}/2+b}\right)^{a+\frac{n}{2}}\frac{b^{a+\frac{n}{2}}}{\Gamma\left(a+\frac{n}{2}\right)} \\
& \times\int_{[T,\infty)}\int_{\mathbb{R}_+}\int_{\mathbb{R}^p}\int_{\mathbb{R}_+^p}\prod_{j=1}^{p}\left\{\nu_j^{-2}\exp\left[-\frac{1}{\nu_j}\left(1+d^{\frac{2}{\delta_1}}+\frac{1}{\lambda_j^2}\right)\right]\right\} \\
& \times\exp\left[-\frac{\left(\boldsymbol{\beta}-M^{-1}\mathbf{X}^T\mathbf{y}\right)^T M\left(\boldsymbol{\beta}-M^{-1}\mathbf{X}^T\mathbf{y}\right)+\boldsymbol{\beta}^T\left(\tau^2\boldsymbol{\Lambda}\right)^{-1}\boldsymbol{\beta}}{2\sigma^2}\right] \\
& \times\exp\left[-\frac{\mathbf{y}^T\left(I-\mathbf{X}M^{-1}\mathbf{X}^T\right)\mathbf{y}+\mathbf{y}^T\mathbf{y}+2b}{2\sigma^2}\right] \\
& \times\left(\sigma^2\right)^{-\left(a+\frac{n+p}{2}\right)-1}\prod_{j=1}^{p}\left\{\frac{\beta_j^2}{2\sigma^2\tau^2}\left(\lambda_j^2\right)^{-2}\right\}\left(1+\tau^2\right)^{-p/2}\left(\tau^2\right)^{-p/2}\pi_\tau\left(\tau^2\right)d\boldsymbol{\nu}d\boldsymbol{\beta}d\sigma^2d\tau^2
\end{aligned}$$

Next we perform the inner integral wrt $\boldsymbol{\nu}$ and noting that $1+d^{\frac{2}{\delta_1}}+\frac{1}{\lambda_j^2}\leq\left(1+d^{\frac{2}{\delta_1}}\right)\left(1+\frac{1}{\lambda_j^2}\right)$ we have:

$$\begin{aligned}
& k\left(\boldsymbol{\lambda}_0,\boldsymbol{\lambda}\right) \\
& \quad\geq(2\pi)^{-\frac{p}{2}}\left(\omega_*\right)^{-p/2}d^{-p/\delta_0}\left(1+d^{\frac{2}{\delta_1}}\right)^{-p}\left(\frac{b}{\mathbf{y}^T\mathbf{y}/2+b}\right)^{a+\frac{n}{2}}\frac{b^{a+\frac{n}{2}}}{\Gamma\left(a+\frac{n}{2}\right)} \\
& \quad\times\int_{[T,\infty)}\int_{\mathbb{R}_+}\int_{\mathbb{R}^p}\prod_{j=1}^{p}\left\{\frac{\beta_j^2}{2\sigma^2}\right\} \\
& \quad\times\exp\left[-\frac{\left(\boldsymbol{\beta}-M^{-1}\mathbf{X}^T\mathbf{y}\right)^T M\left(\boldsymbol{\beta}-M^{-1}\mathbf{X}^T\mathbf{y}\right)+\boldsymbol{\beta}^T\left(\tau^2\boldsymbol{\Lambda}\right)^{-1}\boldsymbol{\beta}}{2\sigma^2}\right]
\end{aligned}$$

$$\times \left(\sigma^2\right)^{-\left(a+\frac{n+p}{2}\right)-1} \exp\left[-\frac{\mathbf{y}^T\left(I-\mathbf{X}M^{-1}\mathbf{X}^T\right)\mathbf{y}+\mathbf{y}^T\mathbf{y}+2b}{2\sigma^2}\right]$$
$$\times \prod_{j=1}^{p}\left\{\left(1+\frac{1}{\lambda_j^2}\right)^{-1}\left(\lambda_j^2\right)^{-2}\right\}\left(1+\tau^2\right)^{-p/2}\left(\tau^2\right)^{-3p/2}\pi_\tau\left(\tau^2\right)d\boldsymbol{\beta}d\sigma^2d\tau^2$$

Now recall that $M=\omega^*\left(1+\frac{1}{\tau^2}\right)I_p$. Hence

$$\begin{aligned}
&\left(\boldsymbol{\beta}-M^{-1}\mathbf{X}^T\mathbf{y}\right)^T M\left(\boldsymbol{\beta}-M^{-1}\mathbf{X}^T\mathbf{y}\right)+\boldsymbol{\beta}^T\left(\tau^2\boldsymbol{\Lambda}\right)^{-1}\boldsymbol{\beta}\\
&=\boldsymbol{\beta}^T\left(M+\left(\tau^2\boldsymbol{\Lambda}\right)^{-1}\right)\boldsymbol{\beta}-2\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{y}+\mathbf{y}^T\mathbf{X}M^{-1}\mathbf{X}^T\mathbf{y}\\
&\leq \boldsymbol{\beta}^T Q\boldsymbol{\beta}-2\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{y}+\mathbf{y}^T\mathbf{X}M^{-1}\mathbf{X}^T\mathbf{y}\\
&=\left(\boldsymbol{\beta}-Q^{-1}\mathbf{X}^T\mathbf{y}\right)^T Q\left(\boldsymbol{\beta}-Q^{-1}\mathbf{X}^T\mathbf{y}\right)+\mathbf{y}^T\mathbf{X}\left(M^{-1}-Q^{-1}\right)\mathbf{X}^T\mathbf{y}
\end{aligned}$$

where $Q=\omega^*\left(1+\frac{1}{\tau^2}\right)\left(I_p+\boldsymbol{\Lambda}^{-1}\right)$. Hence it follows that

$$k\left(\boldsymbol{\lambda}_0,\boldsymbol{\lambda}\right)\geq\left(\omega_*\right)^{-p/2}d^{-p/\delta_0}\left(1+d^{\frac{2}{\delta_1}}\right)^{-p}\left(\frac{b}{\mathbf{y}^T\mathbf{y}/2+b}\right)^{a+\frac{n}{2}}\frac{b^{a+\frac{n}{2}}}{\Gamma\left(a+\frac{n}{2}\right)}$$
$$\times\int_{[T,\infty)}\int_{\mathbb{R}_+}\int_{\mathbb{R}^p}\prod_{j=1}^{p}\left\{\frac{\beta_j^2}{2\sigma^2}\right\}\left(2\pi\sigma^2\right)^{-\frac{p}{2}}\mid Q\mid^{1/2}$$
$$\times\exp\left[-\frac{\left(\boldsymbol{\beta}-Q^{-1}\mathbf{X}^T\mathbf{y}\right)^T Q\left(\boldsymbol{\beta}-Q^{-1}\mathbf{X}^T\mathbf{y}\right)}{2\sigma^2}\right]$$
$$\times\mid Q\mid^{-1/2}\left(\sigma^2\right)^{-\left(a+\frac{n}{2}\right)-1}\exp\left[-\frac{\mathbf{y}^T\left(I-\mathbf{X}Q^{-1}\mathbf{X}^T\right)\mathbf{y}+\mathbf{y}^T\mathbf{y}+2b}{2\sigma^2}\right]$$
$$\times\prod_{j=1}^{p}\left\{\left(1+\frac{1}{\lambda_j^2}\right)^{-1}\left(\lambda_j^2\right)^{-2}\right\}\left(1+\tau^2\right)^{-p/2}\left(\tau^2\right)^{-3p/2}\pi_\tau\left(\tau^2\right)d\boldsymbol{\beta}d\sigma^2d\tau^2$$

Note that if $\boldsymbol{\beta}\sim\mathcal{N}\left(Q^{-1}\mathbf{X}^T\mathbf{y},\sigma^2Q^{-1}\right)$ then the inner most integral wrt $\boldsymbol{\beta}$ is equal to

$$\begin{aligned}
\boldsymbol{E}\left[\prod_{j=1}^{p}\left\{\frac{\beta_j^2}{2\sigma^2}\right\}\right] &= \left(2\sigma^2\right)^{-p}\prod_{j=1}^{p}\left\{E\left[\beta_j^2\right]\right\};\ \text{since } Q \text{ is a diagonal matrix, } \beta_j'\text{s are indep.}\\
&\geq \left(2\sigma^2\right)^{-p}\prod_{j=1}^{p}\left\{\mathrm{Var}\left[\beta_j\right]\right\}\\
&= \left(2\omega^*\right)^{-p}\left(1+\frac{1}{\tau^2}\right)^{-p}\prod_{j=1}^{p}\left\{\left(1+\frac{1}{\lambda_j^2}\right)^{-1}\right\}
\end{aligned}$$

Also, noting that $\mid Q\mid=(\omega^*)^p\left(1+\frac{1}{\tau^2}\right)^p\prod_{j=1}^{p}\left(1+\frac{1}{\lambda_j^2}\right)$ we have the following lower bound:

$$k\left(\boldsymbol{\lambda}_0,\boldsymbol{\lambda}\right)\geq 2^{-p}\left(\omega_*\right)^{-p/2}\left(\omega^*\right)^{-3p/2}d^{-p/\delta_0}\left(1+d^{\frac{2}{\delta_1}}\right)^{-p}\left(\frac{b}{\mathbf{y}^T\mathbf{y}/2+b}\right)^{a+\frac{n}{2}}\frac{b^{a+\frac{n}{2}}}{\Gamma\left(a+\frac{n}{2}\right)}$$

$$\times \int_{[T,\infty)} \int_{\mathbb{R}_+} \left(\sigma^2\right)^{-\left(a+\frac{n}{2}\right)-1} \exp\left[-\frac{\mathbf{y}^T\left(I-\mathbf{X}Q^{-1}\mathbf{X}^T\right)\mathbf{y}+\mathbf{y}^T\mathbf{y}+2b}{2\sigma^2}\right]$$
$$\times \prod_{j=1}^{p}\left\{\left(1+\frac{1}{\lambda_j^2}\right)^{-5/2}\left(\lambda_j^2\right)^{-2}\right\}\left(1+\tau^2\right)^{-2p}\pi_\tau\left(\tau^2\right)d\sigma^2 d\tau^2$$

Further noting that $\mathbf{y}^T\left(I-\mathbf{X}Q^{-1}\mathbf{X}^T\right)\mathbf{y} \le \mathbf{y}^T\mathbf{y}$ we have:

$k\left(\boldsymbol{\lambda}_0,\boldsymbol{\lambda}\right)$

$$\ge 2^{-p}\left(\omega_*\right)^{-p/2}\left(\omega^*\right)^{-3p/2} d^{-p/\delta_0}\left(1+d^{\frac{2}{\delta_1}}\right)^{-p}\left(\frac{b}{\mathbf{y}^T\mathbf{y}/2+b}\right)^{a+\frac{n}{2}}\frac{b^{a+\frac{n}{2}}}{\Gamma\left(a+\frac{n}{2}\right)}$$
$$\times \int_{[T,\infty)}\int_{\mathbb{R}_+}\left(\sigma^2\right)^{-\left(a+\frac{n}{2}\right)-1}\exp\left[-\frac{1}{\sigma^2}(\mathbf{y}^T\mathbf{y}+b)\right]$$
$$\times \prod_{j=1}^{p}\left\{\left(1+\frac{1}{\lambda_j^2}\right)^{-5/2}\left(\lambda_j^2\right)^{-2}\right\}\left(1+\tau^2\right)^{-2p} d\sigma^2 d\tau^2$$

Integrating wrt σ^2 we have:

$$\begin{aligned} k\left(\boldsymbol{\lambda}_0,\boldsymbol{\lambda}\right) &\ge 2^{-p}\left(\omega_*\right)^{-p/2}\left(\omega^*\right)^{-3p/2} d^{-p/\delta_0}\left(1+d^{\frac{2}{\delta_1}}\right)^{-p}\left(\frac{b}{\mathbf{y}^T\mathbf{y}+b}\right)^{2a+n} \\ &\times \prod_{j=1}^{p}\left\{\frac{\sqrt{\lambda_j^2}}{\left(1+\lambda_j^2\right)^{5/2}}\right\}\int_T^\infty \left(1+\tau^2\right)^{-2p}\pi_\tau\left(\tau^2\right)d\tau^2 \\ &= \epsilon^* h\left(\boldsymbol{\lambda}\right) \end{aligned}$$

where

$$\epsilon^* = 3^{-p}\left(\omega_*\right)^{-p/2}\left(\omega^*\right)^{-3p/2} d^{-p/\delta_0}\left(1+d^{\frac{2}{\delta_1}}\right)^{-p}\left(\frac{b}{\mathbf{y}^T\mathbf{y}+b}\right)^{2a+n}$$
$$\times \int_T^\infty \left(1+\tau^2\right)^{-2p}\pi_\tau\left(\tau^2\right)d\tau^2$$

and h is a probability density on $\mathbb{R}_+^p$ given by

$$h\left(\boldsymbol{\lambda}\right) = \prod_{j=1}^{p}\left\{\frac{3}{2}\cdot\frac{\sqrt{\lambda_j^2}}{\left(1+\lambda_j^2\right)^{5/2}}\cdot I_{(0,\infty)}\left(\lambda_j^2\right)\right\}$$

Hence, the minorization condition for the MTD (2.3) is established. □

Appendix D: Samplers for conditional posterior distributions of $\boldsymbol{\lambda}$ and $\boldsymbol{\tau^2}$ for $\boldsymbol{K_{aug,reg}}$

D.1. Rejection sampler for $\boldsymbol{\lambda}$

Recall that the target distribution $g\left(\cdot\,|\,\nu,\beta,\sigma^2,\tau^2,\mathbf{y}\right)$ has density proportion to the function $\phi(\cdot)$ where

$$\phi(x)=\left(\frac{1}{c^2}+\frac{1}{\tau^2 x}\right)^{\frac{1}{2}} x^{-\frac{3}{2}} \exp\left[-\frac{1}{x}\left(\frac{1}{\nu}+\frac{\beta_j^2}{2\sigma^2\tau^2}\right)\right]$$

Consider a probability density function ψ on $\mathbb{R}_+$ as follows:

$$\begin{aligned}\psi(x)=&\frac{\sqrt{\pi/u_j}c^{-1}}{\sqrt{\pi/u_j}c^{-1}+(\tau u_j)^{-1}}\sqrt{\frac{u_j}{\pi}}x^{-3/2}\exp\left[-\frac{u_j}{x}\right]\\&+\frac{(\tau u_j)^{-1}}{\sqrt{\pi/u_j}c^{-1}+(\tau u_j)^{-1}}u_j x^{-2}\exp\left[-\frac{u_j}{x}\right],\end{aligned}$$

where $u_j=1/\nu_j+\beta_j^2/(2\sigma^2\tau^2)$. Note that the above is a convex combination of two Inverse-Gamma densities and is easy to sample from. After simple algebraic manipulations based on the inequality

$$\frac{1}{2}\left(c^{-1}+(\tau^2 x)^{-1/2}\right)\le\sqrt{c^{-2}+(\tau^2 x)^{-1}}\le\left(c^{-1}+(\tau^2 x)^{-1/2}\right),$$

it can be established that

$$\frac{1}{2}\le\inf_{x\in(0,\infty)}\frac{\phi(x)}{M\psi(x)}\sup_{x\in(0,\infty)}\frac{\phi(x)}{\psi(x)}\le M,$$

where

$$M=\sqrt{\pi/u_j}c^{-1}+(\tau u_j)^{-1}.$$

We apply the following Accept-Reject algorithm to generate samples from (normalized version of) the density ϕ. For $i=1,2,\cdots$

1. sample X_i from $\psi(\cdot)$
2. sample U_i from the uniform distribution over $(0,1)$
3. Accept X_i if
$$U_i\le\frac{\phi(X_i)}{M\psi(X_i)}$$
for all i; otherwise, we reject X_i.
4. Repeat the above three steps until we reach a predetermined maximum number of iterations.

Since the acceptance probability $\phi(x)/M\psi(x)$ is always greater than 1/2, the above rejection sampler is very efficient.

D.2. Metropolis sampler for τ^2

Recall that the target distribution $\pi(\cdot|\boldsymbol{\lambda},\mathbf{y})$ has density proportion to the function $\phi(\cdot)$ where

$$\phi(x)=|A_c|^{-\frac{1}{2}}\prod_{j=1}^{p}\left\{\left(\frac{1}{c^2}+\frac{1}{x\lambda_j^2}\right)^{\frac{1}{2}}\right\}\left(\frac{\mathbf{y}^T\left(I_n-\mathbf{X}A_c^{-1}\mathbf{X}^T\right)\mathbf{y}}{2}+b\right)^{-\left(a+\frac{n}{2}\right)}\pi_\tau(x),$$

$A_c = \mathbf{X}^T\mathbf{X} + (x\mathbf{\Lambda})^{-1} + c^{-2}I_p$ and $\pi_\tau(\cdot)$ is a probability density function supported on $\mathbb{R}_+$. We will also need to pick what is called a "proposal distribution" that changes location at each iteration in the algorithm. We will call this $q\,(u \mid x)$. Then the algorithm is:

1. Choose some initial value x_0.
2. For $i = 1, \cdots, p$
   (a) sample x_i^* from $q\,(u \mid x_{i-1})$.
   (b) Set $x_i = x_i^*$ with probalility

$$\alpha = \min\left(\frac{\phi(x_i^*)q\,(x_{i-1} \mid x_i^*)}{\phi(x_{i-1})q\,(x_i^* \mid x_{i-1})}, 1\right)$$

   otherwise set $x_i = x_{i-1}$.

Often times we choose q to be a $\mathcal{N}(x, 1)$ distribution. This has the convenient property of symmetry. Which means that $q\,(u \mid x) = q\,(x \mid u)$, so the quantity α can be simplified to

$$\alpha = \min\left(\frac{\phi(x_i^*)}{\phi(x_{i-1})}, 1\right)$$

which is much easier to calculate.

References

[1] Bhattacharya, A., Chakraborty, A. and Mallick, B. K. (2016). Fast sampling with Gaussian scale mixture priors in high-dimensional regression. *Biometrika* **103** 985-991. MR3620452

[2] Bhattacharya, A., Pati, D., Pillai, N. and Dunson, D. (2015). Dirichlet–Laplace Priors for Optimal Shrinkage. *Journal of the American Statistical Association* **110**. MR3449048

[3] Biswas, N., Bhattacharya, A., Jacob, P. and Johndrow, J. (2020). Coupled Markov chain Monte Carlo for high-dimensional regression with Half-t priors. *arXiv*.

[4] Carvalho, C. M., Polson, N. G. and Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika* **97** 465–480. MR2650751

[5] Diaconis, P., Khare, K. and Saloff-Coste, L. (2008). Gibbs Sampling, Exponential Families and Orthogonal Polynomials. *Statistical Science* **23** 151–178. MR2446500

[6] Flegal, J. M. and Jones, G. L. (2010). Batch means and spectral variance estimators in Markov chain Monte Carlo. *Ann. Statist.* **38** 1034–1070. MR2604704

[7] Johndrow, J. E., Orenstein, P. and Bhattacharya, A. (2020). Scalable Approximate MCMC Algorithms for the Horseshoe Prior. *Journal of Machine Learning Research* **21** 1-61. MR4095352

[8] Jones, G. L., Haran, M., Caffo, B. S. and Neath, R. (2006). Fixed-Width Output Analysis for Markov Chain Monte Carlo. *Journal of the American Statistical Association* **101** 1537–1547. MR2279478

[9] Jones, G. L., Roberts, G. O. and Rosenthal, J. S. (2014). Convergence of conditional Metropolis-Hastings samplers. *Advances in Applied Probability* **46** 422 – 445. MR3215540

[10] Khare, K. and Hobert, J. P. (2013). Geometric ergodicity of the Bayesian lasso. *Electron. J. Statist.* **7** 2150–2163. MR3104915

[11] Livingstone, S., Betancourt, M., Byrne, S. and Girolami, M. (2019). On the geometric ergodicity of Hamiltonian Monte Carlo. *Bernoulli* **25** 3109–3138. MR4003576

[12] Lu, T. T. and Shiou, S. H. (2002). Inverses of 2× 2 block matrices. *Computers & Mathematics with Applications* **43** 119–129. MR1873248

[13] Makalic, E. and Schmidt, D. F. (2016). A Simple Sampler for the Horseshoe Estimator. *IEEE Signal Processing Letters* **23** 179–182.

[14] Meyn, S. P. and Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer-Verlag, London. MR1287609

[15] Nishimura, A. and Suchard, M. A. (2020). Shrinkage with shrunken shoulders: inference via geometrically / uniformly ergodic Gibbs sampler. *arXiv*.

[16] Pal, S. and Khare, K. (2014). Geometric ergodicity for Bayesian shrinkage models. *Electron. J. Statist.* **8** 604–645. MR3211026

[17] Pal, S., Khare, K. and Hobert, J. P. (2017). Trace Class Markov Chains for Bayesian Inference with Generalized Double Pareto Shrinkage Priors. *Scandinavian Journal of Statistics* **44** 307-323. MR3658516

[18] Piironen, J. and Vehtari, A. (2017). Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electron. J. Statist.* **11** 5018–5051. MR3738204

[19] Polson, N. G. and Scott, J. G. (2012). On the Half-Cauchy Prior for a Global Scale Parameter. *Bayesian Analysis* **7** 887—902. MR3000018

[20] Roberts, G. and Rosenthal, J. (2001). Markov chains and de-initializing processes. *Scandinavian Journal of Statistics* **28** 489-504. MR1858413

[21] Rosenthal, J. S. (1995). Minorization Conditions and Convergence Rates for Markov Chain Monte Carlo. *Journal of the American Statistical Association* **90** 558–566. MR1340509