

Geometric ergodicity for Bayesian shrinkage models

Subhadip Pal and Kshitij Khare

Department of Statistics

102 Griffin-Floyd Hall

Gainesville, FL 32611

e-mail: subhadippal@stat.ufl.edu; kdkhare@stat.ufl.edu

Abstract: In recent years, a large variety of continuous shrinkage priors have been developed for a Bayesian analysis of the standard linear regression model in high dimensional settings. We consider two such priors, the Dirichlet-Laplace prior (developed in Bhattacharya et al. (2013)), and the Normal-Gamma prior (developed in (Griffin and Brown, 2010)). For both Dirichlet-Laplace and Normal-Gamma priors, Gibbs sampling Markov chains have been developed to generate approximate samples from the corresponding posterior distributions. We show by using a drift and minorization based analysis that the Gibbs sampling Markov chains corresponding to the aforementioned models are geometrically ergodic. Establishing geometric ergodicity of these Markov chains is crucial, as it provides theoretical justification for the use of Markov chain CLT, which can then be used to obtain asymptotic standard errors for Markov chain based estimates of posterior quantities. Both Gibbs samplers in the paper use the Generalized Inverse Gaussian (GIG) distribution, as one of the conditional distributions. A novel contribution of our convergence analysis is the use of drift functions which include terms that are negative fractional powers of normal random variables, to tackle the presence of the GIG distribution.

MSC 2010 subject classifications: Primary 60J05, 60J20; secondary 33C10.

Keywords and phrases: Bayesian shrinkage, Bessel functions, geometric ergodicity, Markov chain Monte Carlo.

Received December 2013.

1. Introduction

In recent years, there has been a huge influx of high-dimensional datasets from various fields such as genomics, environmental sciences, finance and the social sciences. Classical statistical methods are inadequate to analyze these high-dimensional datasets. Hence, development of new techniques for analyzing these datasets has been a major focus of statistical research in the last decade, and continues to generate much interest. In this context, analysis of the standard linear regression model in a high-dimensional setting, is a well-studied and popular research topic.

In particular, consider the model $\mathbf{y} = X\boldsymbol{\beta} + \sigma\boldsymbol{\epsilon}$, where $\mathbf{y} = (y_i)_{i=1}^n \in \mathbb{R}^n$ is the vector of observations, X is the (known) design matrix, $\boldsymbol{\beta} \in \mathbb{R}^p$ is the (unknown) vector of regression coefficients, the components of $\boldsymbol{\epsilon}$ are iid standard

normal errors, and σ^2 is the (unknown) variance parameter. The objective is to estimate the unknown parameters β and σ^2 . For many modern datasets, the number of regression parameters (p , the dimension of β) is much larger than the number of observations (n). Classical statistical methods are not applicable for analyzing such datasets. A natural approach in these situations is to look for sparse estimates of β , i.e., estimates where many entries of β are exactly equal to zero. The *lasso* (introduced in Tibshirani (1994)) is a popular approach to deal with this problem. In this approach, the estimate of β is given by

$$\hat{\beta}_{\text{lasso}} = \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} \left\{ (\mathbf{y} - X\beta)^T (\mathbf{y} - X\beta) + \lambda \sum_{j=1}^p |\beta_j| \right\}, \quad (1.1)$$

where λ is a user-specified tuning parameter. The estimate $\hat{\beta}_{\text{lasso}}$ also has a Bayesian statistical interpretation. It is well known that $\hat{\beta}_{\text{lasso}}$ is the mode of the conditional density of β given σ^2 and \mathbf{y} , assuming that the components of β (conditioned on σ^2) are independently distributed, with each component having a Laplace prior distribution. Based on this observation, several authors have proposed a Bayesian analysis of the linear regression model in high-dimensional settings by using various adaptations/generalizations of the Laplace prior distributions for the entries of β (see Bhattacharya et al. (2013); Carvalho et al. (2010); Figueiredo (2003); Griffin and Brown (2005, 2010); Johnstone and Silverman (2002); Park and Casella (2008); Polson and Scott (2010); Tipping and Smola (2001)). These prior distributions are referred to as “continuous shrinkage priors” and the corresponding statistical models are referred to as “Bayesian shrinkage models”.

In recent work, Bhattacharya et al. (2013) unify many of the standard Bayesian shrinkage models under a common framework, and study their theoretical properties. They show that frequentist optimality of these Bayesian shrinkage methods depends heavily on prior concentration around sparse vectors, i.e., whether the corresponding priors place sufficient mass around sparse β values (especially in high-dimensional settings). They further demonstrate that some popular choices of shrinkage models, including the Bayesian lasso (introduced in Park and Casella (2008)), have sub-optimal prior concentration properties, and develop a new class of continuous shrinkage priors called the Dirichlet-Laplace priors, which have optimal prior concentration properties.

The Normal-Gamma model developed by Griffin and Brown (2010) is another commonly used Bayesian shrinkage model, and contains the Bayesian lasso as a special case. In fact, the Bayesian lasso model corresponds to a specific choice of a hyperparameter in the Normal-Gamma model. Bhattacharya et al. (2013) argue that for an appropriate range of hyperparameter values (not including the Bayesian lasso case) the Normal-Gamma model shares the optimal prior concentration properties of the Dirichlet-Laplace model.

For both the Dirichlet-Laplace and Normal-Gamma model, the posterior density is intractable in the sense that desired expected values cannot be computed in closed form, and it is not possible to directly sample from the posterior density.

As a result, for both these models, Markov chains have been devised to generate approximate samples from the corresponding posterior density. A crucial, and often challenging, task in this context is to investigate whether these Markov chains are geometrically ergodic, i.e., whether the distribution of each of these Markov chains converges to the corresponding posterior distribution at a geometric rate. As we discuss below, establishing geometric ergodicity is important for a meaningful statistical analysis of the corresponding models. A proof of geometric ergodicity of the Markov chain corresponding to the Bayesian Lasso model is provided in Khare and Hobert (2013). However, an investigation of geometric ergodicity for the Markov chain corresponding to the general Normal-Gamma model and the Dirichlet-Laplace model has not been undertaken. Indeed, the general Normal-Gamma model requires a much more complicated and nuanced analysis as compared to the Bayesian lasso model (see Remark 1).

To clarify the importance of establishing geometric ergodicity, consider the details of the Normal-Gamma model introduced in Griffin and Brown (2010). The model is specified as follows:

$$\mathbf{y} \mid \boldsymbol{\beta}, \sigma^2 \sim N_n(X\boldsymbol{\beta}, \sigma^2 I_n) \quad (1.2)$$

$$\boldsymbol{\beta} \mid \sigma^2, \boldsymbol{\tau} \sim N_p(0_p, \sigma^2 D_{\boldsymbol{\tau}}), \quad (1.3)$$

$$\sigma^2 \sim \text{Inverse-Gamma}(\alpha, \xi) \text{ with } \alpha, \xi > 0 \text{ known} \quad (1.4)$$

$$\tau_j \sim \text{Gamma}(a, b) \text{ for } j = 1, 2, \dots, p, \text{ with } a, b > 0 \text{ known,} \quad (1.5)$$

where N_p denotes the p -variate normal density, and $D_{\boldsymbol{\tau}}$ is a diagonal matrix with diagonal entries given by $\{\tau_j\}_{j=1}^p$. The joint density for the parameter vector $(\boldsymbol{\beta}, \boldsymbol{\tau}, \sigma^2)$ conditioned on the data vector \mathbf{y} is given by the following:

$$\begin{aligned} \pi(\boldsymbol{\beta}, \boldsymbol{\tau}, \sigma^2 \mid \mathbf{y}) &\propto \frac{e^{-\frac{(\mathbf{y}-X\boldsymbol{\beta})^T(\mathbf{y}-X\boldsymbol{\beta})}{2\sigma^2}} e^{-\frac{\boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1} \boldsymbol{\beta}}{2\sigma^2}}}{(\sqrt{2\pi})^n \sigma^n} \frac{1}{(\sqrt{2\pi})^p \sigma^p} \times \\ &\quad \left(\prod_{j=1}^p \tau_j^{a-\frac{1}{2}-1} e^{-b\tau_j} \right) (\sigma^2)^{-\alpha-1} e^{-\frac{\xi}{\sigma^2}}. \end{aligned} \quad (1.6)$$

There is no direct method to sample from this intractable posterior density. However, it turns out that $\boldsymbol{\beta} \mid \boldsymbol{\tau}, \sigma^2, \mathbf{y}$ is multivariate normal, $\sigma^2 \mid \boldsymbol{\beta}, \boldsymbol{\tau}, \mathbf{y}$ is Inverse Gamma, and the entries of $\boldsymbol{\tau} \mid \boldsymbol{\beta}, \sigma^2, \mathbf{y}$ are independently distributed with a generalized inverse Gaussian distribution (see Section 3 for more details). Using these standard densities, Griffin and Brown (2010) construct a Gibbs sampling Markov chain $\{(\boldsymbol{\beta}_m, \boldsymbol{\tau}_m, \sigma_m^2)\}_{m \geq 0}$. The transition of this Markov chain from $(\boldsymbol{\beta}_m, \boldsymbol{\tau}_m, \sigma_m^2)$ to $(\boldsymbol{\beta}_{m+1}, \boldsymbol{\tau}_{m+1}, \sigma_{m+1}^2)$ can be described by the following procedure:

Iteration (m + 1) of the Gibbs sampler:

1. Draw σ_{m+1}^2 from $\pi(\cdot \mid \boldsymbol{\beta}_m, \boldsymbol{\tau}_m, \mathbf{y})$, i.e., from the conditional density given $(\boldsymbol{\beta}_m, \boldsymbol{\tau}_m, \mathbf{y})$.
2. Draw $\boldsymbol{\tau}_{m+1}$ from $\pi(\cdot \mid \sigma_{m+1}^2, \boldsymbol{\beta}_m, \mathbf{y})$, i.e., from the conditional density given $(\sigma_{m+1}^2, \boldsymbol{\beta}_m, \mathbf{y})$.

- 3. Draw β_{m+1} from $\pi(\cdot \mid \tau_{m+1}, \sigma_{m+1}^2, \mathbf{y})$, i.e., from the conditional density given $(\tau_{m+1}, \sigma_{m+1}^2, \mathbf{y})$.

As shown in Section 3, it can be established in a straightforward way that the Markov chain described above is Harris ergodic, and its stationary density is given by the density in (1.6). Harris ergodicity can be used to construct strongly consistent estimators of intractable posterior expectations. In particular, if a real-valued measurable function h satisfies

$$E_\pi|h| := \int \int \int |h(\beta, \tau, \sigma^2)| \pi(\beta, \tau, \sigma^2 \mid \mathbf{y}) d\beta d\tau d\sigma^2 < \infty,$$

then irrespective of how the chain is started, the estimator

$$\bar{h}_m := \frac{1}{m+1} \sum_{i=0}^m h(\beta_m, \tau_m, \sigma_m^2)$$

is strongly consistent for $E_\pi h$. However, this estimator is useful only if an estimate of the associated standard error can be provided. All known methods to compute asymptotically consistent estimates of standard errors for \bar{h}_m require the existence of a Markov chain central limit theorem (CLT). In particular, we need to establish that

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

where c^2 is a finite positive constant. In general, the only standard method available to prove a Markov chain CLT, and obtain consistent estimates of the asymptotic variance c^2 , is to prove that the underlying Markov chain is geometrically ergodic (see Chan and Geyer (1994); Flegal and Jones (2010); Mykland, Tierney and Yu (1995); Robert (1995)). In the current context, the Markov chain $\{(\beta_m, \tau_m, \sigma_m^2)\}_{m \geq 0}$ is defined to be geometrically ergodic if there exists a positive real-valued function M , and a constant $\gamma \in [0, 1)$ such that, for every $(\beta_0, \tau_0, \sigma_0^2)$ and $r \in \mathbb{N}$,

$$\|K_{(\beta_0, \tau_0, \sigma_0^2)}^r - \Pi\|_{TV} \leq M(\beta_0, \tau_0, \sigma_0^2) \gamma^r$$

where $K_{(\beta_0, \tau_0, \sigma_0^2)}^r$ denotes the distribution of the Markov chain started at $(\beta_0, \tau_0, \sigma_0^2)$ after r steps, Π denotes the stationary distribution (corresponding to the density in (1.6)), and $\|\cdot\|_{TV}$ denotes the total variation norm.

If the state space of an ergodic Markov chain is finite, then geometric ergodicity is immediate. However, establishing geometric ergodicity is a much more challenging task for a Markov chain with an infinite state space. A significant proportion of Markov chains arising in statistical applications (including the Markov chains considered in this paper) have continuous state spaces. Despite some success stories, for a large majority of these Markov chains, a proof of geometric ergodicity is not available.

For the Normal-Gamma Markov chain described above, we prove the following.

Theorem 1.1. *The Normal-Gamma Markov chain $\{(\beta_m, \tau_m, \sigma_m^2)\}_{m \geq 0}$ is geometrically ergodic for every $n \geq 3, p, X, \alpha, \xi, a$ and b .*

A similar result (Theorem 4.1) is established for the Dirichlet-Laplace Markov chain in Section 4. The analysis is based on Harris recurrence techniques developed in Meyn and Tweedie (2009), Rosenthal (1995) and Roberts and Tweedie (1996) (see Jones and Hobert (2001) for an extensive review). This method of establishing geometric ergodicity of general state space Markov chains has been previously used by several authors, including Hobert (2001); Hobert and Geyer (1998); Jarner and Hansen (2000); Jones and Hobert (2001, 2004); Marchev and Hobert (2004); Mengersen and Tweedie (1996); Mira and Tierney (1997); Roberts and Rosenthal (1998b, 1999); Roman and Hobert (2012); Rosenthal (1996); Roy and Hobert (2007); Tan and Hobert (2009); Tan, Jones and Hobert (2012). In order to use this technique, one needs to construct a *drift function*, and then use it to establish appropriate *drift* and *minorization* conditions (see for example Proposition 3.1 and Proposition 3.2). However, there is no general recipe or guidelines to construct the drift function and the corresponding drift and minorization conditions. Constructing these in any specific application is currently a matter of art.

Both the Gibbs sampling Markov chains analyzed in this paper use the Generalized Inverse Gaussian (GIG) distribution as one of the conditional distributions. To the best of the authors' knowledge, a drift and minorization based analysis for Gibbs samplers involving the GIG distribution has not been undertaken in previous literature. Gibbs sampling Markov chains in Khare and Hobert (2012), Khare and Hobert (2013) and Choi and Hobert (2013) involve the Inverse Gaussian distribution (which is a special case of the GIG distribution), and require a *much simpler* analysis as compared to the analysis in Section 3 and Section 4 (see Remark 1). We construct a new class of drift functions, which include terms that are negative fractional powers of normal random variables, to tackle the presence of the GIG distribution. The details can be found at the beginning of Section 3.1.

The paper is organized as follows. In Section 2, we establish notation for standard densities that will be used in our analysis. Section 3 provides drift and minorization conditions for the Normal-Gamma model. Section 4 provides drift and minorization conditions for the Dirichlet-Laplace model. Section 5 contains a discussion of future research directions. The [appendix](#) contains mathematical results, including some new identities related to modified Bessel functions of the second kind. These mathematical results play an important role in the analysis in Section 3 and Section 4.

2. Notation

We establish notation for some standard densities that will be used in the analysis.

- $N_p(\boldsymbol{\mu}, \Sigma)$ denotes the p -variate normal density with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ .
- $\text{Gamma}(\alpha, \xi)$ denotes the gamma density with shape parameter $\alpha > 0$ and *rate* parameter $\xi > 0$.

- Inverse-Gamma(α, ξ) denotes the inverse-gamma density with shape parameter $\alpha > 0$ and rate parameter $\xi > 0$.
- GIG(ν, α, ξ) denotes the generalized inverse Gaussian density with parameters $\nu \in \mathbb{R}, \alpha > 0, \xi > 0$, and corresponds to the density function

$$f_{\text{GIG}}(x) = \frac{\alpha^{\frac{\nu}{2}}}{2\xi^{\frac{\nu}{2}}K_{\nu}(\sqrt{\alpha\xi})}x^{\nu-1}e^{-\frac{1}{2}(\alpha x + \frac{\xi}{x})}, \text{ for } x > 0.$$

Here $K_{\nu}(\cdot)$ denotes the modified Bessel function of the second kind. Since f_{GIG} is a density function, it follows immediately that

$$K_{\nu}(\sqrt{\alpha\xi}) = \int_0^{\infty} \frac{\alpha^{\frac{\nu}{2}}}{2\xi^{\frac{\nu}{2}}}x^{\nu-1}e^{-\frac{1}{2}(\alpha x + \frac{\xi}{x})}dx. \tag{2.1}$$

Note that for any $x > 0$, $K_{\nu}(x)$ can be obtained from (2.1) by setting $\alpha = x$ and $\xi = x$.

Using (2.1), if $X \sim \text{GIG}(\nu, \alpha, \xi)$, then

$$E[X] = \frac{\sqrt{\xi}K_{\nu+1}(\sqrt{\alpha\xi})}{\sqrt{\alpha}K_{\nu}(\sqrt{\alpha\xi})}. \tag{2.2}$$

The [appendix](#) contains several new mathematical identities (along with proofs) for modified Bessel functions of the second kind. These results are useful in the subsequent analysis.

3. The Normal-Gamma model

In this section, we prove that the Gibbs sampling Markov chain in Griffin and Brown (2010) for drawing approximate samples from the posterior density in (1.6) is geometrically ergodic. We first resolve a minor technical issue, and then proceed to provide the details of the Gibbs sampling algorithm in Griffin and Brown (2010). Consider the Normal-Gamma model specified in (1.2)–(1.5). It can be shown that if a is chosen to be less than $\frac{1}{2}$, then the marginal density of β converges to infinity if one or more entries of β converge to zero. Such a choice is intentional, and the objective is that the posterior density of (β, τ, σ^2) will concentrate on neighborhoods of sparse values of β . To avoid technical complications in defining the conditional densities for the Gibbs sampling algorithm in Griffin and Brown (2010), it is assumed that the parameter β takes values in $\tilde{\mathbb{R}}^p$, where $\tilde{\mathbb{R}} := \mathbb{R} \setminus \{0\}$. Since the set $\mathbb{R}^p \setminus \tilde{\mathbb{R}}^p$ has Lebesgue measure (on \mathbb{R}^p) equal to zero, expectations with respect to the density $\pi(\beta, \tau, \sigma^2 \mid \mathbf{y})$ remain the same, whether we restrict β to $\tilde{\mathbb{R}}^p$ or not. Hence, it is enough to be able to generate samples from the density $\pi(\beta, \tau, \sigma^2 \mid \mathbf{y})$ restricted to the space $\tilde{\mathbb{R}}^p \times \mathbb{R}_+^p \times \mathbb{R}_+$.

Griffin and Brown (2010) show that the conditional densities corresponding to the density in (1.6) are given as follows.

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

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

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

$$\pi(\boldsymbol{\beta} \mid \boldsymbol{\tau}, \sigma^2, \mathbf{y}) = \frac{|X^T X + D_{\boldsymbol{\tau}}^{-1}|^{\frac{1}{2}}}{(\sqrt{2\pi})^p \sigma^p} \times e^{-\frac{(\boldsymbol{\beta} - (X^T X + D_{\boldsymbol{\tau}}^{-1})^{-1} X^T \mathbf{y})^T (X^T X + D_{\boldsymbol{\tau}}^{-1}) (\boldsymbol{\beta} - (X^T X + D_{\boldsymbol{\tau}}^{-1})^{-1} X^T \mathbf{y})}{2\sigma^2}}, \tag{3.1}$$

for $\boldsymbol{\beta} \in \tilde{\mathbb{R}}^p$. Note that a sample from any multivariate normal density of \mathbb{R}^p is with probability 1 also a sample from the same density restricted to $\tilde{\mathbb{R}}^p$. Hence it is easy to generate a sample from the above conditional density.

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

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

density. In particular,

$$\pi(\sigma^2 \mid \boldsymbol{\beta}, \boldsymbol{\tau}, \mathbf{y}) = \frac{\left((\mathbf{y} - X\boldsymbol{\beta})^T (\mathbf{y} - X\boldsymbol{\beta}) + \boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1} \boldsymbol{\beta} + 2\xi \right)^{\frac{n+p+2\alpha}{2}}}{2^{\frac{n+p+2\alpha}{2}} \Gamma(\frac{n+p+2\alpha}{2})} \times (\sigma^2)^{-\frac{n+p+2\alpha}{2} - 1} e^{-\frac{(\mathbf{y} - X\boldsymbol{\beta})^T (\mathbf{y} - X\boldsymbol{\beta}) + \boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1} \boldsymbol{\beta} + 2\xi}{2\sigma^2}}, \tag{3.2}$$

for $\sigma^2 \in \mathbb{R}_+$. Here $\mathbb{R}_+ := (0, \infty)$.

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

$$\pi(\boldsymbol{\tau} \mid \boldsymbol{\beta}, \sigma^2, \mathbf{y}) = \prod_{j=1}^p \frac{(2b\sigma^2)^{\frac{a-\frac{1}{2}}{2}}}{2|\beta_j|^{a-\frac{1}{2}} K_{a-\frac{1}{2}} \left(\sqrt{2b\frac{\beta_j^2}{\sigma^2}} \right)} \tau_j^{(a-\frac{1}{2})-1} e^{-\frac{1}{2} \left\{ 2b\tau_j + \frac{\beta_j^2}{\sigma^2} \frac{1}{\tau_j} \right\}}, \tag{3.3}$$

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

Samples can be generated easily from all the conditional densities in (3.1), (3.2) and (3.3) by using standard statistical software (such as R). As described in Section 1, the conditional densities described above can be used to construct a block Gibbs sampling Markov chain $\{(\boldsymbol{\beta}_m, \boldsymbol{\tau}_m, \sigma_m^2)\}_{m=0}^\infty$. The Markov transition density corresponding to this Markov chain (with respect to the Lebesgue measure on $\tilde{\mathbb{R}}^p \times \mathbb{R}_+^p \times \mathbb{R}_+$) is given by

$$k_{NG} \left((\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\tau}}, \tilde{\sigma}^2), (\boldsymbol{\beta}, \boldsymbol{\tau}, \sigma^2) \right) = \pi(\boldsymbol{\beta} \mid \boldsymbol{\tau}, \sigma^2, \mathbf{y}) \pi(\boldsymbol{\tau} \mid \tilde{\boldsymbol{\beta}}, \sigma^2, \mathbf{y}) \pi(\sigma^2 \mid \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\tau}}, \mathbf{y}).$$

It is well known, and can be easily verified, that the joint density of $(\boldsymbol{\beta}, \boldsymbol{\tau}, \sigma^2)$ (conditioned on \mathbf{y}) defined in (1.6), is invariant for the Gibbs transition density

k_{NG} defined above. Since k_{NG} is strictly positive, it follows that the corresponding Markov chain is irreducible with respect to the Lebesgue measure on $\tilde{\mathbb{R}}^p \times \mathbb{R}_+^p \times \mathbb{R}_+$ (see (Meyn and Tweedie, 2009, Page 87)), and aperiodic. Results in Asmussen and Glynn (2011) can now be invoked to conclude that the Markov chain is positive Harris recurrent. We now proceed to prove geometric ergodicity by establishing a *geometric drift condition* and an *associated minorization condition* for the Gibbs transition density k_{NG} . Using results from Rosenthal (1995), this will immediately establish geometric ergodicity of the corresponding Markov chain.

3.1. Drift condition

Consider the following function

$$V(\boldsymbol{\beta}, \boldsymbol{\tau}, \sigma^2) = (\mathbf{y} - X\boldsymbol{\beta})^T(\mathbf{y} - X\boldsymbol{\beta}) + \boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1} \boldsymbol{\beta} + \sum_{j=1}^p \frac{1}{|\beta_j|^{\delta(a)}} + \sum_{j=1}^p \tau_j,$$

where $D_{\boldsymbol{\tau}}$ is the diagonal matrix with the diagonal elements $\{\tau_j\}_{j=1}^p$, and

$$\delta(a) = aI_{[0 < a \leq \frac{1}{2}]} + \min \left\{ \frac{1}{2}, 2a - 1 \right\} I_{[a > \frac{1}{2}]}.$$

This function will be used to establish a geometric drift condition for the Normal-Gamma Markov chain. The first two terms in $V(\boldsymbol{\beta}, \boldsymbol{\tau}, \sigma^2)$ are quadratic forms in $\boldsymbol{\beta}$. The choice of quadratic terms like these is pretty standard. However, the third term $\sum_{j=1}^p \frac{1}{|\beta_j|^{\delta}}$ is a sum of negative fractional powers of the entries of $\boldsymbol{\beta}$ (whose conditional distribution given $\boldsymbol{\tau}, \sigma^2$ and \mathbf{y} is multivariate normal with a non-zero mean vector in general), and is the new feature of this drift function. We would like to differentiate this choice with the drift function in Tan and Hobert (2009). The drift function in Tan and Hobert (2009) consists of positive fractional powers of quadratic forms of variables whose conditional distribution given the other parameters and the data is multivariate normal.

Let $E_{k_{NG}}[\cdot \mid \boldsymbol{\beta}_0, \boldsymbol{\tau}_0, \sigma_0^2]$ denote the expectation with respect to one step of the Markov chain with transition density k_{NG} , starting at $(\boldsymbol{\beta}_0, \boldsymbol{\tau}_0, \sigma_0^2)$. The following proposition establishes a geometric drift condition for the transition density k_{NG} .

Proposition 3.1. *If $n \geq 3$, there exists constants $0 \leq \gamma < 1$ and $\varphi \geq 0$ such that*

$$E_{k_{NG}}[V(\boldsymbol{\beta}, \boldsymbol{\tau}, \sigma^2) \mid (\boldsymbol{\beta}_0, \boldsymbol{\tau}_0, \sigma_0^2)] \leq \gamma V(\boldsymbol{\beta}_0, \boldsymbol{\tau}_0, \sigma_0^2) + \varphi, \tag{3.4}$$

for every $(\boldsymbol{\beta}_0, \boldsymbol{\tau}_0, \sigma_0^2) \in \tilde{\mathbb{R}}^p \times \mathbb{R}_+^p \times \mathbb{R}_+$.

Proof. It follows by the definition of the Markov transition density k_{NG} that,

$$\begin{aligned} & E_{k_{NG}}[V(\boldsymbol{\beta}, \boldsymbol{\tau}, \sigma^2) \mid (\boldsymbol{\beta}_0, \boldsymbol{\tau}_0, \sigma_0^2)] \\ &= E[E[E[V(\boldsymbol{\beta}, \boldsymbol{\tau}, \sigma^2) \mid \boldsymbol{\tau}, \sigma^2, \mathbf{y}] \mid \boldsymbol{\beta}_0, \sigma^2, \mathbf{y}] \mid \boldsymbol{\beta}_0, \boldsymbol{\tau}_0, \mathbf{y}], \end{aligned} \tag{3.5}$$

where $E[\cdot \mid \boldsymbol{\tau}, \sigma^2, \mathbf{y}]$ denotes the expectation with respect to $\pi(\cdot \mid \boldsymbol{\tau}, \sigma^2, \mathbf{y})$ (specified in (3.1)), $E[\cdot \mid \boldsymbol{\beta}_0, \sigma^2, \mathbf{y}]$ denotes the expectation with respect to $\pi(\cdot \mid \boldsymbol{\beta}_0, \sigma^2, \mathbf{y})$ (specified in (3.2)), and $E[\cdot \mid \boldsymbol{\beta}_0, \boldsymbol{\tau}_0, \mathbf{y}]$ denotes the expectation with respect to $\pi(\cdot \mid \boldsymbol{\beta}_0, \boldsymbol{\tau}_0, \mathbf{y})$ (specified in (3.3)).

We evaluate the three conditional expectations one step at a time. Note that

$$\begin{aligned}
 & E \left[(\mathbf{y} - X\boldsymbol{\beta})^T (\mathbf{y} - X\boldsymbol{\beta}) + \boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1} \boldsymbol{\beta} \mid \boldsymbol{\tau}, \sigma^2, \mathbf{y} \right] \\
 &= \mathbf{y}^T \mathbf{y} + E[\boldsymbol{\beta}^T (X^T X + D_{\boldsymbol{\tau}}^{-1}) \boldsymbol{\beta} \mid \boldsymbol{\tau}, \sigma^2, \mathbf{y}] - 2\mathbf{y}^T X E[\boldsymbol{\beta} \mid \boldsymbol{\tau}, \sigma^2, \mathbf{y}] \\
 &= \mathbf{y}^T \mathbf{y} + \|(X^T X + D_{\boldsymbol{\tau}}^{-1})^{\frac{1}{2}} (X^T X + D_{\boldsymbol{\tau}}^{-1})^{-1} X^T \mathbf{y}\|_2^2 + \\
 &\quad \text{tr} \left((X^T X + D_{\boldsymbol{\tau}}^{-1})^{\frac{1}{2}} (X^T X + D_{\boldsymbol{\tau}}^{-1})^{-1} (X^T X + D_{\boldsymbol{\tau}}^{-1})^{\frac{1}{2}} \right) \sigma^2 \\
 &\quad \quad \quad - 2\mathbf{y}^T X (X^T X + D_{\boldsymbol{\tau}}^{-1})^{-1} X^T \mathbf{y} \\
 &= \mathbf{y}^T \mathbf{y} - \mathbf{y}^T X (X^T X + D_{\boldsymbol{\tau}}^{-1})^{-1} X^T \mathbf{y} + p\sigma^2. \tag{3.6}
 \end{aligned}$$

Let \mathbf{e}_j denote the j^{th} unit vector in \mathbb{R}^p . By (3.1), we get that the conditional density of $\beta_j = \mathbf{e}_j^T \boldsymbol{\beta}$ given $\boldsymbol{\tau}, \sigma^2, \mathbf{y}$ is $N(\mu_j, \sigma_j^2)$, where $\mu_j = \mathbf{e}_j^T (X^T X + D_{\boldsymbol{\tau}}^{-1})^{-1} X^T \mathbf{y}$ and $\sigma_j^2 = \sigma^2 \mathbf{e}_j^T (X^T X + D_{\boldsymbol{\tau}}^{-1})^{-1} \mathbf{e}_j$. Let λ_X be largest eigenvalue of $X^T X$. It follows that for every $j = 1, 2, \dots, p$,

$$\begin{aligned}
 \frac{1}{\sigma_j^{\delta(a)}} &= \left(\frac{1}{\sigma^2 \mathbf{e}_j^T (X^T X + D_{\boldsymbol{\tau}}^{-1})^{-1} \mathbf{e}_j} \right)^{\frac{\delta(a)}{2}} \\
 &\stackrel{(a)}{\leq} \frac{1}{\sigma^{\delta(a)}} \left(\frac{1}{\mathbf{e}_j^T (\lambda_X I_p + D_{\boldsymbol{\tau}}^{-1})^{-1} \mathbf{e}_j} \right)^{\frac{\delta(a)}{2}} \\
 &= \frac{1}{\sigma^{\delta(a)}} \left(\lambda_X + \frac{1}{\tau_j} \right)^{\frac{\delta(a)}{2}} \\
 &\stackrel{(a')}{\leq} \left(\frac{\lambda_X^{\frac{\delta(a)}{2}}}{\sigma^{\delta(a)}} + \frac{1}{\sigma^{\delta(a)} \tau_j^{\frac{\delta(a)}{2}}} \right). \tag{3.7}
 \end{aligned}$$

Here (a) follows from the fact that $\lambda_X I_p - X^T X$ is a non-negative definite matrix and (a') follows from the fact that if $u, v \geq 0$ and $0 < \delta < 1$ then $(u + v)^\delta \leq u^\delta + v^\delta$. Let $\kappa(a) := \frac{\Gamma(\frac{1-\delta(a)}{2}) 2^{\frac{1-\delta(a)}{2}}}{\sqrt{2\pi}}$ for $a > 0$. Using Proposition A1 (see appendix) with β_j playing the role of V , and utilizing the inequality in (3.7), we obtain

$$E \left[\frac{1}{|\beta_j|^{\delta(a)}} \mid \boldsymbol{\tau}, \sigma^2, \mathbf{y} \right] \leq \frac{\kappa(a)}{\sigma_j^{\delta(a)}} \leq \kappa(a) \left(\frac{\lambda_X^{\frac{\delta(a)}{2}}}{\sigma^{\delta(a)}} + \frac{1}{\sigma^{\delta(a)} \tau_j^{\frac{\delta(a)}{2}}} \right) \tag{3.8}$$

for every $j = 1, 2, \dots, p$. It follows by summing over all $j = 1, 2, \dots, p$ that

$$E \left[\sum_{j=1}^p \frac{1}{|\beta_j|^{\delta(a)}} \mid \tau, \sigma^2, \mathbf{y} \right] \leq \frac{p\kappa(a)\lambda_X^{\frac{\delta(a)}{2}}}{\sigma^{\delta(a)}} + \frac{\kappa(a)}{\sigma^{\delta(a)}} \sum_{j=1}^p \frac{1}{\tau_j^{\frac{\delta(a)}{2}}}. \quad (3.9)$$

It follows by (3.6) and (3.9) that the innermost expectation in (3.5) can be bounded as follows

$$E[V(\boldsymbol{\beta}, \boldsymbol{\tau}, \sigma^2) \mid \boldsymbol{\tau}, \sigma^2, \mathbf{y}] \leq \mathbf{y}^T \mathbf{y} + p\sigma^2 + \frac{p\kappa(a)\lambda_X^{\frac{\delta(a)}{2}}}{\sigma^{\delta(a)}} + \frac{\kappa(a)}{\sigma^{\delta(a)}} \sum_{j=1}^p \frac{1}{(\tau_j)^{\frac{\delta(a)}{2}}} + \sum_{j=1}^p \tau_j. \quad (3.10)$$

For the next step, we consider the middle conditional expectation in (3.5). By (3.3), we have $\tau_j \mid \beta_0, \sigma^2, \mathbf{y} \sim \text{GIG}(a - \frac{1}{2}, 2b, \frac{|\beta_{0j}|^2}{\sigma^2})$. It follows by (2.2) that

$$\begin{aligned} E[\tau_j \mid \beta_0, \sigma^2, \mathbf{y}] &= \left(\sqrt{\frac{|\beta_{0j}^2|}{2b\sigma^2}} \right) \frac{K_{a+\frac{1}{2}} \left(\sqrt{2b \frac{|\beta_{0j}|^2}{\sigma^2}} \right)}{K_{a-\frac{1}{2}} \left(\sqrt{2b \frac{|\beta_{0j}|^2}{\sigma^2}} \right)} \\ &\stackrel{(b)}{\leq} \left(\frac{|\beta_{0j}|}{\sqrt{2b\sigma}} \right) \left(\frac{a + \sqrt{a^2 + 2b \left(\frac{|\beta_{0j}|}{\sigma} \right)^2}}{\frac{\sqrt{2b}|\beta_{0j}|}{\sigma}} \right) \\ &\stackrel{(b')}{\leq} \frac{a}{b} + \frac{|\beta_{0j}|}{\sqrt{2b\sigma}} \\ &\stackrel{(b'')}{\leq} \frac{a}{b} + C_1 \frac{|\beta_{0j}|^2}{\sigma^2} + \frac{1}{2bC_1}, \end{aligned} \quad (3.11)$$

where C_1 is an **arbitrary** positive constant (we will make an appropriate choice of C_1 later). Here (b) follows from the fact that $\frac{K_{\nu+\frac{1}{2}}(x)}{K_{\nu-\frac{1}{2}}(x)} \leq \frac{\nu+\sqrt{\nu^2+x^2}}{x}$ for $x, \nu > 0$ (see (Segura, 2011, Theorem 2)), (b') follows from the fact that $\sqrt{x^2 + y^2} \leq |x| + |y|$, and (b'') follows from the fact that $|xy| \leq \frac{x^2+y^2}{2} \leq x^2 + y^2$. By (2.1), we get that

$$\begin{aligned} E \left[\frac{1}{(\tau_j)^{\frac{\delta(a)}{2}}} \right] &= \int_0^\infty \frac{1}{x^{\frac{\delta(a)}{2}}} \left(\frac{\sqrt{2b\sigma}}{|\beta_{0j}|} \right)^{a-\frac{1}{2}} \frac{x^{(a-\frac{1}{2})-1} e^{-\frac{1}{2} \left(2bx + \frac{\beta_{0j}^2}{\sigma^2 x} \right)}}{2K_{a-\frac{1}{2}} \left(\frac{\sqrt{2b}|\beta_{0j}|}{\sigma} \right)} dx \\ &= \left(\frac{\sqrt{2b\sigma}}{|\beta_{0j}|} \right)^{\frac{\delta(a)}{2}} \frac{K_{a-\frac{1}{2}-\frac{\delta(a)}{2}} \left(\frac{\sqrt{2b}|\beta_{0j}|}{\sigma} \right)}{K_{a-\frac{1}{2}} \left(\frac{\sqrt{2b}|\beta_{0j}|}{\sigma} \right)}. \end{aligned} \quad (3.12)$$

Now we separately consider three cases when $a < \frac{1}{2}$, $a > \frac{1}{2}$ and $a = \frac{1}{2}$.

Case 1: $a \in (0, \frac{1}{2})$.

By the property that $K_\nu(x) = K_{-\nu}(x)$ (see (Abramowitz and Stegun, 1965, Page 375)), we obtain

$$E \left[\frac{1}{(\tau_j)^{\frac{\delta(a)}{2}}} \right] = \left(\frac{\sqrt{2b}\sigma}{|\beta_{0j}|} \right)^{\frac{\delta(a)}{2}} \frac{K_{\frac{1}{2} + \frac{\delta(a)}{2} - a} \left(\frac{\sqrt{2b}|\beta_{0j}|}{\sigma} \right)}{K_{\frac{1}{2} - a} \left(\frac{\sqrt{2b}|\beta_{0j}|}{\sigma} \right)}. \quad (3.13)$$

By Proposition A2 (see appendix), with $\nu_1 = \frac{1}{2} + \frac{\delta(a)}{2} - a$, $\nu_2 = \frac{1}{2} - a$ and $x = \frac{\sqrt{2b}|\beta_{0j}|}{\sigma}$, it follows that for arbitrary ϵ^* (we will make an appropriate choice of ϵ^* later), there exists an ϵ depending on ϵ^* and a such that, for $\frac{\sqrt{2b}|\beta_{0j}|}{\sigma} < \epsilon$,

$$\frac{K_{\frac{1}{2} + \frac{\delta(a)}{2} - a} \left(\frac{\sqrt{2b}|\beta_{0j}|}{\sigma} \right)}{K_{\frac{1}{2} - a} \left(\frac{\sqrt{2b}|\beta_{0j}|}{\sigma} \right)} \leq (1 + \epsilon^*) \frac{\Gamma(\frac{1}{2} + \frac{\delta(a)}{2} - a) 2^{\frac{\delta(a)}{2}} \sigma^{\frac{\delta(a)}{2}}}{\Gamma(\frac{1}{2} - a) \left(\sqrt{2b}|\beta_{0j}| \right)^{\frac{\delta(a)}{2}}}. \quad (3.14)$$

By (3.13), we obtain

$$\begin{aligned} & E \left[\frac{1}{(\tau_j)^{\frac{\delta(a)}{2}}} \right] \\ &= \left(\frac{\sqrt{2b}\sigma}{|\beta_{0j}|} \right)^{\frac{\delta(a)}{2}} \frac{K_{\frac{1}{2} + \frac{\delta(a)}{2} - a} \left(\frac{\sqrt{2b}|\beta_{0j}|}{\sigma} \right)}{K_{\frac{1}{2} - a} \left(\frac{\sqrt{2b}|\beta_{0j}|}{\sigma} \right)} I_{\left[\frac{\sqrt{2b}|\beta_{0j}|}{\sigma} < \epsilon \right]} \\ &\quad + \left(\frac{\sqrt{2b}\sigma}{|\beta_{0j}|} \right)^{\frac{\delta(a)}{2}} \frac{K_{\frac{1}{2} + \frac{\delta(a)}{2} - a} \left(\frac{\sqrt{2b}|\beta_{0j}|}{\sigma} \right)}{K_{\frac{1}{2} - a} \left(\frac{\sqrt{2b}|\beta_{0j}|}{\sigma} \right)} I_{\left[\frac{\sqrt{2b}|\beta_{0j}|}{\sigma} \geq \epsilon \right]} \\ &\stackrel{(c)}{\leq} (1 + \epsilon^*) \left(\frac{\sigma}{|\beta_{0j}|} \right)^{\delta(a)} \frac{\Gamma(\frac{1}{2} + \frac{\delta(a)}{2} - a) 2^{\frac{\delta(a)}{2}}}{\Gamma(\frac{1}{2} - a)} I_{\left[\frac{\sqrt{2b}|\beta_{0j}|}{\sigma} < \epsilon \right]} \\ &\quad + \left(\frac{\sqrt{2b}\sigma}{|\beta_{0j}|} \right)^{\frac{\delta(a)}{2}} \left(\frac{(2(1-a))^{\frac{\delta(a)}{2}}}{\left(\frac{\sqrt{2b}|\beta_{0j}|}{\sigma} \right)^{\frac{\delta(a)}{2}}} + 1 \right) I_{\left[\frac{\sqrt{2b}|\beta_{0j}|}{\sigma} \geq \epsilon \right]} \\ &\leq (1 + \epsilon^*) \left(\frac{\sigma}{|\beta_{0j}|} \right)^{\delta(a)} \frac{\Gamma(\frac{1}{2} + \frac{\delta(a)}{2} - a) 2^{\frac{\delta(a)}{2}}}{\Gamma(\frac{1}{2} - a)} + \\ &\quad + (2b)^{\frac{\delta(a)}{2}} \frac{(2(1-a))^{\frac{\delta(a)}{2}}}{\epsilon^{\delta(a)}} + \frac{(2b)^{\frac{\delta(a)}{2}}}{\epsilon^{\frac{\delta(a)}{2}}}, \quad (3.15) \end{aligned}$$

where (c) follows by (3.14) and Proposition A4 (see appendix) with $\nu := \frac{1}{2} - a$ and $x := \frac{\sqrt{2b}|\beta_{0j}|}{\sigma}$.

Case 2: $a > \frac{1}{2}$.

Note that if $x > 0$, then $\nu \mapsto K_\nu(x)$ is an increasing function for $\nu > 0$ (see (Laforgia, 1991, Page 266)). Furthermore, by (Baricz, 2008, Page 69), it follows that if $\nu_1 < \nu_2$, $\nu > 0$, and $x > 0$, then

$$\frac{K_{\nu_1+\nu}(x)}{K_{\nu_1}(x)} \leq \frac{K_{\nu_2+\nu}(x)}{K_{\nu_2}(x)}.$$

By choosing $\nu_1 = -\nu$ and $\nu_2 = 0$, it follows that $K_0^2(x) \leq K_\nu^2(x)$ for $\nu > 0$ and $x > 0$. Since $K_\nu(x)$ is positive for $\nu \geq 0$ and $x > 0$, it follows that $\nu \mapsto K_\nu(x)$ is an increasing function for $\nu \geq 0$. Using this along with (3.12) and the fact $uv \leq \frac{u^2+v^2}{2}$ for $u, v > 0$, we get that

$$E \left[\frac{1}{(\tau_j)^{\frac{\delta(a)}{2}}} \right] \leq \left(\frac{\sqrt{2b}\sigma}{|\beta_{0j}|} \right)^{\frac{\delta(a)}{2}} \leq \frac{1}{2\kappa(a)} \left(\frac{\sigma}{|\beta_{0j}|} \right)^{\delta(a)} + \frac{(2b)^{\frac{\delta(a)}{2}} \kappa(a)}{2}. \quad (3.16)$$

Case 3: $a = \frac{1}{2}$.

From (3.12) it follows that

$$E \left[\frac{1}{(\tau_j)^{\frac{\delta(a)}{2}}} \right] = \left(\frac{\sqrt{2b}\sigma}{|\beta_{0j}|} \right)^{\frac{\delta(a)}{2}} \frac{K_{\frac{\delta(a)}{2}} \left(\frac{\sqrt{2b}|\beta_{0j}|}{\sigma} \right)}{K_0 \left(\frac{\sqrt{2b}|\beta_{0j}|}{\sigma} \right)}. \quad (3.17)$$

By Proposition A3 (see appendix) with $\nu := \frac{\delta(a)}{2}$ and $x := \frac{\sqrt{2b}|\beta_{0j}|}{\sigma}$, there exists and ϵ' such that

$$\frac{K_{\frac{\delta(a)}{2}} \left(\frac{\sqrt{2b}|\beta_{0j}|}{\sigma} \right)}{K_0 \left(\frac{\sqrt{2b}|\beta_{0j}|}{\sigma} \right)} \leq \frac{1}{2\kappa \left(\frac{1}{2} \right) \left(\frac{\sqrt{2b}|\beta_{0j}|}{\sigma} \right)^{\frac{\delta(a)}{2}}} \quad \text{for } 0 < x < \epsilon'. \quad (3.18)$$

Hence by (3.17) and (3.18) we obtain,

$$\begin{aligned} & E \left[\frac{1}{(\tau_j)^{\frac{\delta(a)}{2}}} \right] \\ &= \left(\frac{\sqrt{2b}\sigma}{|\beta_{0j}|} \right)^{\frac{\delta(a)}{2}} \frac{K_{\frac{\delta(a)}{2}} \left(\frac{\sqrt{2b}|\beta_{0j}|}{\sigma} \right)}{K_0 \left(\frac{\sqrt{2b}|\beta_{0j}|}{\sigma} \right)} I_{\left[\frac{\sqrt{2b}|\beta_{0j}|}{\sigma} < \epsilon' \right]} \\ &\quad + \left(\frac{\sqrt{2b}\sigma}{|\beta_{0j}|} \right)^{\frac{\delta(a)}{2}} \frac{K_{\frac{\delta(a)}{2}} \left(\frac{\sqrt{2b}|\beta_{0j}|}{\sigma} \right)}{K_0 \left(\frac{\sqrt{2b}|\beta_{0j}|}{\sigma} \right)} I_{\left[\frac{\sqrt{2b}|\beta_{0j}|}{\sigma} \geq \epsilon' \right]} \\ &\stackrel{(c')}{\leq} \frac{1}{2\kappa \left(\frac{1}{2} \right)} \left(\frac{\sigma}{|\beta_{0j}|} \right)^{\delta(a)} I_{\left[\frac{\sqrt{2b}|\beta_{0j}|}{\sigma} < \epsilon' \right]} \end{aligned}$$

$$\begin{aligned}
& + \left(\frac{\sqrt{2b}\sigma}{|\beta_{0j}|} \right)^{\frac{\delta(a)}{2}} \left(\frac{1}{\left(\frac{\sqrt{2b}|\beta_{0j}|}{\sigma} \right)^{\frac{\delta(a)}{2}}} + 1 \right) I_{\left[\frac{\sqrt{2b}|\beta_{0j}|}{\sigma} \geq \epsilon' \right]} \\
& \leq \frac{1}{2\kappa\left(\frac{1}{2}\right)} \left(\frac{\sigma}{|\beta_{0j}|} \right)^{\delta(a)} I_{\left[\frac{\sqrt{2b}|\beta_{0j}|}{\sigma} < \epsilon' \right]} + \frac{(2b)^{\frac{\delta(a)}{2}}}{\epsilon'^{\delta(a)}} + \frac{(2b)^{\frac{\delta(a)}{2}}}{\epsilon'^{\frac{\delta(a)}{2}}}, \quad (3.19)
\end{aligned}$$

where (c') follows from Proposition A4 (see appendix) with $\nu = 0$ and $x = \frac{\sqrt{2b}|\beta_{0j}|}{\sigma}$.

Let us denote $C_2 = \frac{1}{2\kappa(a)} I_{[a \geq \frac{1}{2}]} + (1 + \epsilon^*) \frac{\Gamma(\frac{1}{2} + \frac{\delta(a)}{2} - a) 2^{\frac{\delta(a)}{2}}}{\Gamma(\frac{1}{2} - a)} I_{[0 < a < \frac{1}{2}]}$ and

$$\begin{aligned}
C_3 & = \frac{(2b)^{\frac{\delta(a)}{2}} \kappa(a)}{2} I_{[a > \frac{1}{2}]} + (2b)^{\frac{\delta(a)}{2}} \left[\frac{(2(1-a))^{\frac{\delta(a)}{2}}}{\epsilon^{\delta(a)}} + \frac{1}{\epsilon^{\frac{\delta(a)}{2}}} \right] I_{[0 < a < \frac{1}{2}]} \\
& + \frac{(2b)^{\frac{\delta(a)}{2}}}{\epsilon'^{\delta(a)}} \left(1 + \epsilon'^{\frac{\delta(a)}{2}} \right) I_{[a = \frac{1}{2}]}.
\end{aligned}$$

Hence, combining the three cases, and using (3.15), (3.16) and (3.19), we have for every $a > 0$,

$$E \left[\frac{1}{(\tau_j)^{\frac{\delta(a)}{2}}} \right] \leq C_2 \left(\frac{\sigma}{|\beta_{0j}|} \right)^{\delta(a)} + C_3. \quad (3.20)$$

By (3.10), (3.11) and (3.20), it follows that

$$\begin{aligned}
& E [E[V(\boldsymbol{\beta}, \boldsymbol{\tau}, \sigma^2) \mid \boldsymbol{\tau}, \sigma^2, \mathbf{y}] \mid \boldsymbol{\beta}_0, \sigma^2, \mathbf{y}] \\
& \leq \mathbf{y}^T \mathbf{y} + p\sigma^2 + \frac{p\kappa(a)\lambda_{X^2}^{\frac{\delta(a)}{2}}}{\sigma^{\delta(a)}} + E \left[\frac{\kappa(a)}{\sigma^{\delta(a)}} \sum_{j=1}^p \frac{1}{(\tau_j)^{\frac{\delta(a)}{2}}} + \sum_{j=1}^p \tau_j \mid \boldsymbol{\beta}_0, \sigma^2, \mathbf{y} \right] \\
& \leq \mathbf{y}^T \mathbf{y} + p\sigma^2 + \frac{p\kappa(a)(\lambda_{X^2}^{\frac{\delta(a)}{2}} + C_3)}{\sigma^{\delta(a)}} + \kappa(a)C_2 \sum_{j=1}^p \frac{1}{|\beta_{0j}|^{\delta(a)}} \\
& \quad + C_1 \sum_{j=1}^p \frac{\beta_{0j}^2}{\sigma^2} + \frac{ap}{b} + \frac{p}{2bC_1}. \quad (3.21)
\end{aligned}$$

To analyze terms related to the outermost conditional expectation in (3.5), note that by (3.2)

$$E[\sigma^2 \mid \boldsymbol{\beta}_0, \boldsymbol{\tau}_0, \mathbf{y}] = \frac{(\mathbf{y} - X\boldsymbol{\beta}_0)^T (\mathbf{y} - X\boldsymbol{\beta}_0) + \boldsymbol{\beta}_0^T D_{\boldsymbol{\tau}_0}^{-1} \boldsymbol{\beta}_0 + 2\xi}{n + p + 2\alpha - 2}, \quad (3.22)$$

and

$$\begin{aligned} E\left[\frac{1}{\sigma^2} \mid \beta_0, \tau_0, \mathbf{y}\right] &= \frac{n+p+2\alpha}{(\mathbf{y}-X\beta_0)^T(\mathbf{y}-X\beta_0)+\beta_0^T D_{\tau_0}^{-1} \beta_0+2\xi} \\ &\leq \frac{n+p+2\alpha}{\beta_0^T D_{\tau_0}^{-1} \beta_0}. \end{aligned} \quad (3.23)$$

Since $\delta(a) \leq \frac{1}{2}$, it follows that $g(x) = x^{\frac{\delta(a)}{2}}$ is a concave function for $x > 0$. Hence, by Jensen's inequality for concave functions, we obtain

$$\begin{aligned} E\left[\frac{1}{\sigma^{\delta(a)}} \mid \beta_0, \tau_0, \mathbf{y}\right] &\leq \left(E\left[\frac{1}{\sigma^2} \mid \beta_0, \tau_0, \mathbf{y}\right]\right)^{\frac{\delta(a)}{2}} \\ &\leq \left[\frac{n+p+2\alpha}{(\mathbf{y}-X\beta_0)^T(\mathbf{y}-X\beta_0)+\beta_0^T D_{\tau_0}^{-1} \beta_0+2\xi}\right]^{\frac{\delta(a)}{2}} \\ &\leq \left[\frac{n+p+2\alpha}{2\xi}\right]^{\frac{\delta(a)}{2}}. \end{aligned} \quad (3.24)$$

It follows by (3.5), (3.21), (3.22), (3.23) and (3.24) that

$$\begin{aligned} &E_{k_{NG}}[V(\beta, \tau, \sigma^2) \mid (\beta_0, \tau_0, \sigma_0^2)] \\ &\leq \mathbf{y}^T \mathbf{y} + \kappa(a)C_2 \sum_{j=1}^p \frac{1}{|\beta_{0j}|^{\delta(a)}} + pE[\sigma^2 \mid \beta_0, \tau_0, \mathbf{y}] + \frac{ap}{b} + \frac{p}{2bC_1} \\ &\quad + E\left[\frac{p\kappa(a)(\lambda_X^{\frac{\delta(a)}{2}} + C_3)}{\sigma^{\delta(a)}} \mid \beta_0, \tau_0, \mathbf{y}\right] + C_1 \sum_{j=1}^p E\left[\frac{\beta_{0j}^2}{\sigma^2} \mid \beta_0, \tau_0, \mathbf{y}\right] \\ &\leq \mathbf{y}^T \mathbf{y} + \kappa(a)C_2 \sum_{j=1}^p \frac{1}{|\beta_{0j}|^{\delta(a)}} + p\kappa(a)(\lambda_X^{\frac{\delta(a)}{2}} + C_3) \left[\frac{n+p+2\alpha}{2\xi}\right]^{\frac{\delta(a)}{2}} \\ &\quad + p \frac{(\mathbf{y}-X\beta_0)^T(\mathbf{y}-X\beta_0)+\beta_0^T D_{\tau_0}^{-1} \beta_0+2\xi}{n+p+2\alpha-2} + \frac{ap}{b} \\ &\quad + C_1(n+p+2\alpha) \sum_{j=1}^p \frac{\beta_{0j}^2}{\beta_0^T D_{\tau_0}^{-1} \beta_0} + \frac{p}{2bC_1}. \end{aligned} \quad (3.25)$$

Note that by the Cauchy-Schwarz inequality,

$$\begin{aligned} \sum_{j=1}^p \beta_{0j}^2 &\leq \left(\sum_{j=1}^p |\beta_{0j}|\right)^2 = \left(\sum_{j=1}^p \sqrt{\frac{\beta_{0j}^2}{\tau_{0j}}} \sqrt{\tau_{0j}}\right)^2 \leq \left(\sum_{j=1}^p \frac{\beta_{0j}^2}{\tau_{0j}}\right) \left(\sum_{j=1}^p \tau_{0j}\right) \\ &\leq (\beta_0^T D_{\tau_0}^{-1} \beta_0) \left(\sum_{j=1}^p \tau_{0j}\right). \end{aligned} \quad (3.26)$$

It follows from (3.25) and (3.26) that

$$\begin{aligned}
& E_{k_{NG}}[V(\boldsymbol{\beta}, \boldsymbol{\tau}, \sigma^2) \mid (\boldsymbol{\beta}_0, \boldsymbol{\tau}_0, \sigma_0^2)] \\
& \leq \kappa(a)C_2 \sum_{j=1}^p \frac{1}{|\beta_{0j}|^{\delta(a)}} + C_1(n+p+2\alpha) \sum_{j=1}^p \tau_{0j} \\
& \quad + \frac{p}{n+p+2\alpha-2} \left[(\mathbf{y} - X\boldsymbol{\beta}_0)^T (\mathbf{y} - X\boldsymbol{\beta}_0) + \boldsymbol{\beta}_0^T D_{\boldsymbol{\tau}_0}^{-1} \boldsymbol{\beta}_0 \right] \\
& \quad + \mathbf{y}^T \mathbf{y} + \frac{2\xi p}{n+p+2\alpha-2} + p\kappa(a)(\lambda_X^{\frac{\delta(a)}{2}} + C_3) \left[\frac{n+p+2\alpha}{2\xi} \right]^{\frac{\delta(a)}{2}} \\
& \quad + \frac{ap}{b} + \frac{p}{2bC_1}.
\end{aligned}$$

Hence, we obtain

$$E_{k_{NG}}[V(\boldsymbol{\beta}, \boldsymbol{\tau}, \sigma^2) \mid (\boldsymbol{\beta}_0, \boldsymbol{\tau}_0, \sigma_0^2)] \leq \gamma V(\boldsymbol{\beta}_0, \boldsymbol{\tau}_0, \sigma_0^2) + \varphi, \quad (3.27)$$

where

$$\begin{aligned}
\varphi &= \mathbf{y}^T \mathbf{y} + \frac{2\xi p}{n+p+2\alpha-2} + p\kappa(a)(\lambda_X^{\frac{\delta(a)}{2}} + C_3) \left[\frac{n+p+2\alpha}{2\xi} \right]^{\frac{\delta(a)}{2}} \\
& \quad + \frac{ap}{b} + \frac{p}{2bC_1},
\end{aligned} \quad (3.28)$$

$$\gamma = \text{Max} \left\{ \kappa(a)C_2, C_1(n+p+2\alpha), \frac{p}{n+p+2\alpha-2} \right\}. \quad (3.29)$$

Recall that $\delta(a) = aI_{[0 < a \leq \frac{1}{2}]} + \min\{\frac{1}{2}, 2a-1\}I_{[a > \frac{1}{2}]}$, $\kappa(a) = \frac{\Gamma(\frac{1-\delta(a)}{2})2^{\frac{1-\delta(a)}{2}}}{\sqrt{2\pi}}$ and $C_2 = \frac{1}{2\kappa(a)}I_{[a \geq \frac{1}{2}]} + (1 + \epsilon^*) \frac{\Gamma(\frac{1}{2} + \frac{\delta(a)}{2} - a)2^{\frac{\delta(a)}{2}}}{\Gamma(\frac{1}{2} - a)}I_{[0 < a < \frac{1}{2}]}$. Hence,

$$\kappa(a)C_2 = \frac{1}{2}I_{[a \geq \frac{1}{2}]} + (1 + \epsilon^*) \frac{[\Gamma(\frac{1}{2} - \frac{a}{2})]^2}{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2} - a)}I_{[0 < a < \frac{1}{2}]}. \quad (3.30)$$

Note that $\frac{[\Gamma(\frac{\vartheta+\zeta}{2})]^2}{\Gamma(\frac{\vartheta}{2})\Gamma(\frac{\vartheta+2\zeta}{2})} \leq \frac{\vartheta}{\vartheta+\zeta^2}$ for $\vartheta > 0$ and $\zeta > 0$ (see Gurland (1956)). Let $0 < a < \frac{1}{2}$. Taking $\zeta = \frac{a}{2}$, $\vartheta = \frac{1}{2} - a$ we get $\frac{[\Gamma(\frac{1}{2} - \frac{a}{2})]^2}{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2} - a)} \leq \frac{1-2a}{(1-2a) + \frac{a^2}{2}}$. If we choose $\epsilon^* < \frac{a^2}{2(1-2a)}$ (remember ϵ^* is an arbitrary positive constant of our choice) then clearly $(1 + \epsilon^*) \frac{[\Gamma(\frac{1}{2} - \frac{a}{2})]^2}{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2} - a)} < 1$. Hence $\kappa(a)C_2 < 1$.

The assumption $n \geq 3$ will automatically lead to the fact that $\frac{p}{n+p+2\alpha-2} < 1$, and if we choose $C_1 = \frac{1}{2(n+p+2\alpha)}$ (recall that C_1 is an arbitrary positive constant of our choice), then $C_1(n+p+2\alpha) = \frac{1}{2}$. Hence $0 \leq \gamma < 1$, which establishes the required result. \square

Remark 1. The Bayesian lasso model in Park and Casella (2008) arises as a special case of the Normal-Gamma model, when the hyperparameter a in the Normal-Gamma model is chosen to be 1. In this case, the conditional distribution of $\frac{1}{\tau_j}$ given $\beta, \sigma^2, \mathbf{y}$ reduces to an Inverse Gaussian distribution for each $1 \leq j \leq p$. In Khare and Hobert (2013), a proof of geometric ergodicity of the Bayesian lasso Markov chain is provided by using a drift function of the form

$$(\mathbf{y} - X\beta)^T(\mathbf{y} - X\beta) + \beta^T D_{\tau}^{-1} \beta + \sum_{j=1}^p \tau_j.$$

Hence, the term of the form $\sum_{j=1}^p \frac{1}{|\beta_j|^{\delta}}$ is not required. This significantly simplifies the drift and minorization analysis in the Bayesian lasso case.

3.2. Minorization condition

For every $d > 0$, let

$$B_{V,d} = \left\{ (\beta_0, \tau_0, \sigma_0^2) \in \tilde{\mathbb{R}}^p \times \mathbb{R}_+^p \times \mathbb{R}_+ : V(\beta_0, \tau_0, \sigma_0^2) \leq d \right\}.$$

The following proposition establishes an associated minorization condition to the geometric drift condition in Proposition 3.1.

Proposition 3.2. *There exist a constant $0 < \tilde{\epsilon} = \tilde{\epsilon}(V, d) \leq 1$ and a probability density function \tilde{f} on $\tilde{\mathbb{R}}^p \times \mathbb{R}_+^p \times \mathbb{R}_+$ such that*

$$k_{NG}((\beta_0, \tau_0, \sigma_0^2), (\beta, \tau, \sigma^2)) \geq \tilde{\epsilon} \tilde{f}(\beta, \tau, \sigma^2) \tag{3.31}$$

for every $(\beta_0, \tau_0, \sigma_0^2) \in B_{V,d}$.

Proof. Recall that

$$\begin{aligned} & k_{NG}((\beta_0, \tau_0, \sigma_0^2), (\beta, \tau, \sigma^2)) \\ &= \pi(\beta \mid \tau, \sigma^2, \mathbf{y}) \pi(\tau \mid \beta_0, \sigma_0^2, \mathbf{y}) \pi(\sigma^2 \mid \beta_0, \tau_0, \mathbf{y}). \end{aligned} \tag{3.32}$$

Note that, if $(\beta_0, \tau_0, \sigma_0^2) \in B_{V,d}$, then $\beta_0^T D_{\tau_0}^{-1} \beta_0 + \sum_{j=1}^p \tau_{0j} \leq d$, hence

$$\left(\beta_0^T D_{\tau_0}^{-1} \beta_0 \right) \left(\sum_{j=1}^p \tau_{0j} \right) \leq d^2.$$

It follows by (3.26) that

$$\sum_{j=1}^p \beta_{0j}^2 \leq d^2. \tag{3.33}$$

By (3.3)

$$\pi(\boldsymbol{\tau} \mid \boldsymbol{\beta}_0, \sigma^2, \mathbf{y}) = \prod_{j=1}^p g_{0j}(\tau_j \mid \boldsymbol{\beta}_0, \sigma^2), \quad (3.34)$$

where $g_{0j}(\cdot \mid \boldsymbol{\beta}_0, \sigma^2)$ is the GIG($a - \frac{1}{2}, 2b, \frac{\beta_{0j}^2}{\sigma^2}$) density. Hence

$$\begin{aligned} & g_{0j}(\tau_j \mid \boldsymbol{\beta}_0, \sigma^2) \\ &= \frac{(\tau_j)^{a-\frac{1}{2}-1} e^{-\frac{1}{2}\left(2b\tau_j + \frac{\beta_{0j}^2}{\tau_j\sigma^2}\right)}}{\left[\frac{\beta_{0j}^2}{2b\sigma^2}\right]^{\frac{(a-\frac{1}{2})}{2}} 2K_{a-\frac{1}{2}}\left(\frac{\sqrt{2b}|\beta_{0j}|}{\sigma}\right)} \\ &\stackrel{(e)}{\geq} \frac{(\tau_j)^{a-\frac{1}{2}-1} e^{-\frac{1}{2}\left(2b\tau_j + \frac{d^2}{\tau_j\sigma^2}\right)}}{\left[\frac{|\beta_{0j}|}{\sigma\sqrt{2b}}\right]^{(a-\frac{1}{2})} 2K_{a-\frac{1}{2}}\left(\frac{\sqrt{2b}|\beta_{0j}|}{\sigma}\right)} \\ &\geq \left\{ \frac{\left[\frac{d}{\sigma\sqrt{2b}}\right]^{(a-\frac{1}{2})} 2K_{a-\frac{1}{2}}\left(\frac{\sqrt{2bd}}{\sigma}\right)}{\left[\frac{|\beta_{0j}|}{\sigma\sqrt{2b}}\right]^{(a-\frac{1}{2})} 2K_{a-\frac{1}{2}}\left(\frac{\sqrt{2b}|\beta_{0j}|}{\sigma}\right)} \right\} \left\{ \frac{(\tau_j)^{a-\frac{1}{2}-1} e^{-\frac{1}{2}\left(2b\tau_j + \frac{d^2}{\tau_j\sigma^2}\right)}}{\left[\frac{d}{\sigma\sqrt{2b}}\right]^{(a-\frac{1}{2})} 2K_{a-\frac{1}{2}}\left(\frac{\sqrt{2bd}}{\sigma}\right)} \right\} \\ &\geq \left\{ \left[\frac{d}{|\beta_{0j}|}\right]^{a-\frac{1}{2}} \frac{2K_{a-\frac{1}{2}}\left(\frac{\sqrt{2bd}}{\sigma}\right)}{2K_{a-\frac{1}{2}}\left(\frac{\sqrt{2b}|\beta_{0j}|}{\sigma}\right)} \right\} \tilde{g}_j(\tau_j \mid \sigma^2), \quad (3.35) \end{aligned}$$

where (e) follows by (3.33), and $\tilde{g}_j(\tau_j \mid \sigma^2)$ is the GIG($a - \frac{1}{2}, 2b, \frac{d^2}{\sigma^2}$) density.

Let us consider the function $x \mapsto \frac{K_0(x)}{K_1(x)}$ for $x > 0$. By (Segura, 2011, Page 526) we get that $-x \frac{d}{dx} \frac{K_0(x)}{K_0(x)} < \frac{1}{2} + \sqrt{\frac{1}{4} + x^2}$ and $-x \frac{d}{dx} \frac{K_1(x)}{K_1(x)} > \frac{1}{2} + \sqrt{\frac{1}{4} + x^2}$. Hence $\frac{d}{dx}(\log(\frac{K_0(x)}{K_1(x)})) = \frac{d}{dx} \frac{K_0(x)}{K_0(x)} - \frac{d}{dx} \frac{K_1(x)}{K_1(x)} > 0$. So $x \mapsto \frac{K_0(x)}{K_1(x)}$ is an increasing function for $x > 0$. As $\frac{\sqrt{2bd}}{\sigma} \geq \frac{\sqrt{2b}|\beta_{0j}|}{\sigma}$, we get that

$$\frac{K_0\left(\frac{\sqrt{2bd}}{\sigma}\right)}{K_1\left(\frac{\sqrt{2bd}}{\sigma}\right)} \geq \frac{K_0\left(\frac{\sqrt{2b}|\beta_{0j}|}{\sigma}\right)}{K_1\left(\frac{\sqrt{2b}|\beta_{0j}|}{\sigma}\right)}. \quad (3.36)$$

Let $\tilde{a} = a + I_{[a=\frac{1}{2}]}$. Hence, by (3.35) and (3.36), we obtain

$$\begin{aligned} g_{0j}(\tau_j \mid \boldsymbol{\beta}_0, \sigma^2, \mathbf{y}) &\geq \left\{ \left[\frac{d}{|\beta_{0j}|}\right]^{a-\frac{1}{2}} \frac{2K_{\tilde{a}-\frac{1}{2}}\left(\frac{\sqrt{2bd}}{\sigma}\right)}{2K_{\tilde{a}-\frac{1}{2}}\left(\frac{\sqrt{2b}|\beta_{0j}|}{\sigma}\right)} \right\} \tilde{g}_j(\tau_j \mid \sigma^2) \\ &\stackrel{(e')}{\geq} \left\{ \left[\frac{d}{|\beta_{0j}|}\right]^{a-\frac{1}{2}} \frac{K_{\frac{2\tilde{a}-1}{4}}(bd^2) e^{-\frac{1}{2\sigma^2}}}{2K_{\tilde{a}-\frac{1}{2}}\left(\frac{\sqrt{2b}|\beta_{0j}|}{\sigma}\right)} \right\} \tilde{g}_j(\tau_j \mid \sigma^2), \quad (3.37) \end{aligned}$$

where (e') follows from the Proposition A5 (see appendix) with $s := \sqrt{2bd}$ and $t := \sigma$ and $\nu := \tilde{a} - \frac{1}{2}$.

Note that $\tilde{a} - \frac{1}{2} \neq 0$ for every $a > 0$. Using Proposition A7 (see appendix) with $x = \frac{\sqrt{2b|\beta_{0j}|}}{\sigma}$ and $\nu = \tilde{a} - \frac{1}{2}$ we have

$$\begin{aligned} & g_{0j}(\tau_j \mid \beta_0, \sigma^2, \mathbf{y}) \\ & \geq \left\{ \left[\frac{(2b)^{\frac{1}{2}|\tilde{a}-\frac{1}{2}|} d^{a-\frac{1}{2}}}{\sigma^{|\tilde{a}-\frac{1}{2}|}} \right] \frac{|\beta_{0j}|^{|\tilde{a}-\frac{1}{2}|-(a-\frac{1}{2})} K_{\frac{2\tilde{a}-1}{4}}(bd^2) e^{-\frac{1}{2\sigma^2}}}{2^{|\tilde{a}-\frac{1}{2}|} \Gamma(|\tilde{a}-\frac{1}{2}|)} \right\} \tilde{g}_j(\tau_j \mid \sigma^2). \end{aligned} \tag{3.38}$$

As $(\beta_0, \tau_0, \sigma_0^2) \in B_{V,d}$ and $\delta(a) > 0$, we have $\frac{1}{|\beta_{0j}|^{\delta(a)}} \leq d$ and hence $|\beta_{0j}| \geq \frac{1}{d^{\delta(a)}}$. It follows that

$$\begin{aligned} & g_{0j}(\tau_j \mid \beta_0, \sigma^2, \mathbf{y}) \\ & \geq \left\{ \frac{e^{-\frac{1}{2\sigma^2}}}{\sigma^{|\tilde{a}-\frac{1}{2}|}} \right\} \left\{ \left[(2b)^{\frac{1}{2}|\tilde{a}-\frac{1}{2}|} d^{(a-\frac{1}{2})-\frac{|\tilde{a}-\frac{1}{2}|-(a-\frac{1}{2})}{\delta(a)}} \right] \frac{K_{\frac{2\tilde{a}-1}{4}}(bd^2)}{2^{|\tilde{a}-\frac{1}{2}|} \Gamma(|\tilde{a}-\frac{1}{2}|)} \right\} \tilde{g}_j(\tau_j \mid \sigma^2). \end{aligned} \tag{3.39}$$

Now using (3.34) and the lower bound specified in (3.39) for each $j = 1, 2, \dots, p$, we get that

$$\begin{aligned} & \pi(\tau \mid \beta_0, \sigma^2, \mathbf{y}) \\ & \geq \left\{ \frac{e^{-\frac{p}{2\sigma^2}}}{\sigma^p |\tilde{a}-\frac{1}{2}|} \right\} \left\{ \left[(2b)^{\frac{1}{2}|\tilde{a}-\frac{1}{2}|} d^{(a-\frac{1}{2})-\frac{|\tilde{a}-\frac{1}{2}|-(a-\frac{1}{2})}{\delta(a)}} \right] \frac{K_{\frac{2\tilde{a}-1}{4}}(bd^2)}{2^{|\tilde{a}-\frac{1}{2}|} \Gamma(|\tilde{a}-\frac{1}{2}|)} \right\}^p \\ & \quad \times \left\{ \prod_{j=1}^p \tilde{g}_j(\tau_j \mid \sigma^2) \right\}. \end{aligned} \tag{3.40}$$

We now bound $\pi(\sigma^2 \mid \beta_0, \tau_0, \mathbf{y})$ and the term depending on σ^2 in (3.40). Note that for $(\beta_0, \tau_0, \sigma_0^2) \in B_{V,d}$, we have

$$\frac{(\mathbf{y} - X\beta_0)^T(\mathbf{y} - X\beta_0) + \beta_0^T D_{\tau_0}^{-1} \beta_0 + 2\xi}{2} \leq \frac{d + 2\xi}{2}. \tag{3.41}$$

Note that

$$\begin{aligned} & (\mathbf{y} - X\beta_0)^T(\mathbf{y} - X\beta_0) + \beta_0^T D_{\tau_0}^{-1} \beta_0 + 2\xi \\ & = \mathbf{y}^T \mathbf{y} - 2\mathbf{y}^T X\beta_0 + \beta_0^T (X^T X + D_{\tau_0}^{-1}) \beta_0 + 2\xi \\ & = \mathbf{y}^T \mathbf{y} - \mathbf{y}^T X (X^T X + D_{\tau_0}^{-1})^{-1} X^T \mathbf{y} + 2\xi + \\ & \quad + \left[\left\| (X^T X + D_{\tau_0}^{-1})^{-\frac{1}{2}} (X^T \mathbf{y} - (X^T X + D_{\tau_0}^{-1}) \beta_0) \right\|_2 \right]^2 \end{aligned}$$

$$\begin{aligned}
&\geq \mathbf{y}^T \mathbf{y} - \mathbf{y}^T X (X^T X + D_{\tau_0}^{-1})^{-1} X^T \mathbf{y} + 2\xi \\
&\geq \mathbf{y}^T \mathbf{y} - \mathbf{y}^T X (X^T X + \frac{1}{d} I_p)^{-1} X^T \mathbf{y} + 2\xi.
\end{aligned} \tag{3.42}$$

The last inequality follows since $0 < \tau_{0j} \leq d$ for every $1 \leq j \leq p$. Note that $I_p - X(X^T X + \frac{1}{d} I_p)^{-1} X^T$ is a positive definite matrix. Hence $\mathbf{y}^T \mathbf{y} - \mathbf{y}^T X (X^T X + \frac{1}{d} I_p)^{-1} X^T \mathbf{y} > 0$. It follows by (3.41) and (3.42) that for $(\beta_0, \tau_0, \sigma_0^2) \in B_{V,d}$,

$$\begin{aligned}
&\frac{e^{-\frac{p}{2\sigma^2}}}{\sigma^{p|\bar{a}-\frac{1}{2}|}} \pi(\sigma^2 \mid \beta_0, \tau_0, \mathbf{y}) \\
&= \frac{e^{-\frac{p}{2\sigma^2}}}{\sigma^{p|\bar{a}-\frac{1}{2}|}} \left[(\sigma^2)^{-\frac{n+p+2\alpha}{2}-1} e^{-\frac{(\mathbf{y}-X\beta_0)^T (\mathbf{y}-X\beta_0) + \beta_0^T D_{\tau_0}^{-1} \beta_0 + 2\xi}{2\sigma^2}} \right] \\
&\quad \times \left[\frac{\left((\mathbf{y}-X\beta_0)^T (\mathbf{y}-X\beta_0) + \beta_0^T D_{\tau_0}^{-1} \beta_0 + 2\xi \right)^{\frac{n+p+2\alpha}{2}}}{2^{\frac{n+p+2\alpha}{2}} \Gamma\left(\frac{n+p+2\alpha}{2}\right)} \right] \\
&\geq \left[\frac{\mathbf{y}^T \mathbf{y} - \mathbf{y}^T X (X^T X + \frac{1}{d} I_p)^{-1} X^T \mathbf{y} + 2\xi}{2} \right]^{\frac{n+p+2\alpha}{2}} \\
&\quad \times \left\{ \frac{(\sigma^2)^{-\frac{n+p+2\alpha+p|2\bar{a}-1|}{2}-1} e^{-\frac{d+2\xi+p}{2\sigma^2}}}{\Gamma\left(\frac{n+p+2\alpha}{2}\right)} \right\} \\
&= \left\{ \frac{\left[\frac{\mathbf{y}^T \mathbf{y} - \mathbf{y}^T X (X^T X + \frac{1}{d} I_p)^{-1} X^T \mathbf{y} + 2\xi}{2} \right]^{\frac{n+p+2\alpha}{2}} \Gamma\left(\frac{n+p+2\alpha+p|2\bar{a}-1|}{2}\right)}{\left[\frac{d+2\xi+p}{2} \right]^{\frac{n+p+2\alpha+p|2\bar{a}-1|}{2}} \Gamma\left(\frac{n+p+2\alpha}{2}\right)} \right\} \tilde{h}(\sigma^2),
\end{aligned} \tag{3.43}$$

where $\tilde{h}(\sigma^2)$ is the Inverse-Gamma $\left(\frac{n+p+2\alpha+p|2\bar{a}-1|}{2}, \frac{d+2\xi+p}{2}\right)$ density. It follows from (3.32), (3.40) and (3.43) that for all $(\beta_0, \tau_0, \sigma_0^2) \in B_{V,d}$,

$$k_{NG}((\beta_0, \tau_0, \sigma_0^2), (\beta, \tau, \sigma^2)) \geq \tilde{\epsilon} \tilde{f}(\beta, \tau, \sigma^2), \tag{3.44}$$

where

$$\begin{aligned}
\tilde{\epsilon} &= \left\{ \left[(2b)^{\frac{1}{2}|\bar{a}-\frac{1}{2}|} d^{(a-\frac{1}{2})-\frac{|\bar{a}-\frac{1}{2}|-(a-\frac{1}{2})}{\delta(a)}} \right] \frac{K_{\frac{2\bar{a}-1}{4}}(bd^2)}{2^{|\bar{a}-\frac{1}{2}|} \Gamma(|\bar{a}-\frac{1}{2}|)} \right\}^p \times \\
&\quad \left\{ \frac{\left[\frac{\mathbf{y}^T \mathbf{y} - \mathbf{y}^T X (X^T X + \frac{1}{d} I_p)^{-1} X^T \mathbf{y} + 2\xi}{2} \right]^{\frac{n+p+2\alpha}{2}} \Gamma\left(\frac{n+p+2\alpha+p|2\bar{a}-1|}{2}\right)}{\left[\frac{d+2\xi+p}{2} \right]^{\frac{n+p+2\alpha+p|2\bar{a}-1|}{2}} \Gamma\left(\frac{n+p+2\alpha}{2}\right)} \right\},
\end{aligned} \tag{3.45}$$

and $\tilde{f}(\cdot)$ is a probability density on $\mathbb{R}^p \times \mathbb{R}_+^p \times \mathbb{R}_+$ given by

$$\tilde{f}(\beta, \tau, \sigma^2) = \pi(\beta \mid \tau, \sigma^2, \mathbf{y}) \left\{ \prod_{j=1}^p \tilde{g}_j(\tau_j \mid \sigma^2) \right\} \tilde{h}(\sigma^2).$$

Since $k_{NG}((\boldsymbol{\beta}_0, \boldsymbol{\tau}_0, \sigma_0^2), \cdot)$ and $\tilde{f}(\cdot)$ are both probability density functions, it follows by (3.44) that $\tilde{\epsilon} \leq 1$. This proves the required result. \square

The drift and minorization conditions in Proposition 3.1 and Proposition 4.2 can be combined with (Rosenthal, 1995, Theorem 12) to establish geometric ergodicity of the Normal-Gamma Markov chain with transition density k_{NG} . Theorem 1.1 is a straightforward corollary of Theorem 3.1. Note that the following theorem not only provides a proof of geometric ergodicity, but also provides quantitative convergence bounds which may be used to compute the number of iterations required for the Markov chain distribution to get sufficiently close to the stationary distribution.

Theorem 3.1. *Let p, X, α, ξ, a, b be arbitrary, $n \geq 3$, and let γ, φ and $\tilde{\epsilon}$ be as defined in (3.29), (3.28), and (3.45) respectively. Let $d > \frac{2\varphi}{1-\gamma}$, $A = \frac{1+d}{1+2\varphi+\gamma d}$ and $U = 1 + 2(\gamma d + \varphi)$. Then for any $(\boldsymbol{\beta}_0, \boldsymbol{\tau}_0, \sigma_0^2) \in \tilde{\mathbb{R}}^p \times \mathbb{R}_+^p \times \mathbb{R}_+$, $r \in \mathbb{N}$ and $0 < s < 1$,*

$$\left\| K_{NG, (\boldsymbol{\beta}_0, \boldsymbol{\tau}_0, \sigma_0^2)}^r - \Pi_{NG} \right\|_{TV} \leq (1-\tilde{\epsilon})^{rs} + \left(\frac{U^s}{A^{1-s}} \right)^r \left(1 + \frac{\varphi}{1-\gamma} + V(\boldsymbol{\beta}_0, \boldsymbol{\tau}_0, \sigma_0^2) \right).$$

where $K_{NG, (\boldsymbol{\beta}_0, \boldsymbol{\tau}_0, \sigma_0^2)}^r$ denotes the distribution of the Normal-Gamma Markov chain started at $(\boldsymbol{\beta}_0, \boldsymbol{\tau}_0, \sigma_0^2)$ after r steps, and Π_{NG} denotes the joint distribution of the $(\boldsymbol{\beta}, \boldsymbol{\tau}, \sigma^2)$ (conditioned on \mathbf{y}) in the Normal-Gamma model.

4. The Dirichlet-Laplace model

In this section we analyze the Dirichlet-Laplace Bayesian shrinkage model introduced by Bhattacharya et al. (2013). As in Section 3, let $\mathbf{y} = (y_i)_{i=1}^n$ denote the data vector and X denote the known design matrix. The Dirichlet-Laplace model in Bhattacharya et al. (2013) can be specified as follows

$$\begin{aligned} \mathbf{y} \mid \boldsymbol{\beta}, \sigma^2 &\sim N(X\boldsymbol{\beta}, \sigma^2 I_n) \\ \boldsymbol{\beta} \mid \sigma^2, \boldsymbol{\psi}, \boldsymbol{\phi}, \theta &\sim N(0, \sigma^2 D_{\boldsymbol{\eta}}) \\ \sigma^2 &\sim \text{Inverse-Gamma}(\alpha, \xi) \quad \alpha, \xi > 0 \text{ fixed} \\ \psi_1, \psi_2, \dots, \psi_p &\stackrel{i.i.d.}{\sim} \text{Exp}\left(\frac{1}{2}\right) \\ (\phi_1, \phi_2, \dots, \phi_p) &\sim \text{Dir}(a, a, \dots, a), \theta \sim \text{Gamma}\left(pa, \frac{1}{2}\right), \end{aligned} \quad (4.1)$$

where $\boldsymbol{\eta} = (\boldsymbol{\psi}, \boldsymbol{\phi}, \theta)$, $D_{\boldsymbol{\eta}}$ denotes a diagonal matrix with diagonal elements $(\psi_j \phi_j^2 \theta^2)_{j=1}^p$, and a is a known positive constant. Let

$$\mathcal{S}_p := \left\{ (\phi_j)_{j=1}^p : \phi_j \geq 0 \text{ and } \sum_{j=1}^p \phi_j = 1 \right\}.$$

Then, the parameter $\boldsymbol{\eta} = (\boldsymbol{\psi}, \phi, \theta)$ takes values in the space $\Delta_p := \mathbb{R}_+^p \times \mathcal{S}_p \times \mathbb{R}_+$. It can be shown if a is chosen to be less than 1 in the Dirichlet-Laplace model, then the marginal density of $\boldsymbol{\beta}$ converges to infinity as one or more entries of $\boldsymbol{\beta}$ converge to zero. For the same reasons as those discussed at the beginning of Section 3, we assume that the parameter $\boldsymbol{\beta}$ takes values in $\widetilde{\mathbb{R}}^p$.

The joint density of $(\boldsymbol{\beta}, \boldsymbol{\eta}, \sigma^2)$ (conditioned on \mathbf{y}) with respect to the Lebesgue measure on $\widetilde{\mathbb{R}}^p \times \Delta_p \times \mathbb{R}_+$ is given as follows

$$\begin{aligned} & \pi(\boldsymbol{\beta}, \boldsymbol{\eta}, \sigma^2 \mid \mathbf{y}) \\ \propto & \frac{e^{-\frac{(\mathbf{y}-X\boldsymbol{\beta})^T(\mathbf{y}-X\boldsymbol{\beta})}{2\sigma^2}}}{(\sqrt{2\pi})^n \sigma^n} \frac{e^{-\frac{\boldsymbol{\beta}^T D_{\boldsymbol{\eta}}^{-1} \boldsymbol{\beta}}{2\sigma^2}}}{(\sqrt{2\pi})^p \sigma^p \theta^p} \prod_{j=1}^p \left(\frac{e^{-\frac{\psi_j}{2}}}{\sqrt{\psi_j}} \phi_j^{a-2} \right) \theta^{pa-1} e^{-\frac{\theta}{2}} (\sigma^2)^{-\alpha-1} e^{-\frac{\xi}{\sigma^2}}. \end{aligned} \quad (4.2)$$

To perform Bayesian statistical inference for this model, we need to compute expectations with respect to the above posterior density. However, these expectations are not available in closed form, and there is no direct method to generate samples from the density in (4.2). Bhattacharya et al. (2013) derive the following expressions for the conditional densities associated with the density in (4.2).

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

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

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

$$\begin{aligned} & \pi(\boldsymbol{\beta} \mid \boldsymbol{\eta}, \sigma^2, \mathbf{y}) \\ = & \frac{|X^T X + D_{\boldsymbol{\eta}}^{-1}|^{\frac{1}{2}}}{(\sqrt{2\pi})^p \sigma^p} e^{-\frac{(\boldsymbol{\beta} - (X^T X + D_{\boldsymbol{\eta}}^{-1})^{-1} X^T \mathbf{y})^T (X^T X + D_{\boldsymbol{\eta}}^{-1}) (\boldsymbol{\beta} - (X^T X + D_{\boldsymbol{\eta}}^{-1})^{-1} X^T \mathbf{y})}{2\sigma^2}}, \end{aligned} \quad (4.3)$$

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

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

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

In particular,

$$\begin{aligned} \pi(\sigma^2 \mid \boldsymbol{\beta}, \boldsymbol{\eta}, \mathbf{y}) &= \frac{\left((\mathbf{y}-X\boldsymbol{\beta})^T(\mathbf{y}-X\boldsymbol{\beta}) + \boldsymbol{\beta}^T D_{\boldsymbol{\eta}}^{-1} \boldsymbol{\beta} + 2\xi \right)^{\frac{n+p+2\alpha}{2}}}{2^{\frac{n+p+2\alpha}{2}} \Gamma\left(\frac{n+p+2\alpha}{2}\right)} \\ &\times (\sigma^2)^{-\frac{n+p+2\alpha}{2}-1} e^{-\frac{(\mathbf{y}-X\boldsymbol{\beta})^T(\mathbf{y}-X\boldsymbol{\beta}) + \boldsymbol{\beta}^T D_{\boldsymbol{\eta}}^{-1} \boldsymbol{\beta} + 2\xi}{2\sigma^2}}, \end{aligned} \quad (4.4)$$

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

- For $j = 1, \dots, p$, conditioned on $\phi, \theta, \beta, \sigma^2$ and \mathbf{y} , the random variables $\psi_j, j = 1, 2, \dots, p$ are independently distributed. The conditional density of ψ_j given $\phi, \theta, \beta, \sigma^2, \mathbf{y}$ is $\text{GIG}(\frac{1}{2}, 1, \frac{\beta_j^2}{(\phi_j \theta \sigma)^2})$. In particular

$$\pi(\psi_j \mid \theta, \phi, \beta, \sigma^2, \mathbf{y}) = \left(\frac{1}{2\pi\psi_j}\right)^{\frac{1}{2}} e^{-\frac{\left(1 - \frac{\phi_j \theta \sigma}{|\beta_j|}\right) \psi_j}{2\psi_j \left(\frac{\phi_j \theta \sigma}{|\beta_j|}\right)^2}}, \tag{4.5}$$

for $\psi_j \in \mathbb{R}_+$. For the $\text{GIG}(\nu, \alpha, \xi)$ distribution, in the special case when $\nu = \frac{1}{2}$ or $\nu = -\frac{1}{2}$, the expected value can be obtained in closed form. In fact,

$$E(\psi_j \mid \beta, \sigma^2, \phi, \theta, \mathbf{y}) = 1 + \frac{|\beta_j|}{\phi_j \theta \sigma}, \tag{4.6}$$

and

$$E\left(\frac{1}{\psi_j} \mid \beta, \sigma^2, \phi, \theta, \mathbf{y}\right) = \frac{\phi_j \theta \sigma}{|\beta_j|}. \tag{4.7}$$

- The conditional density of (ϕ, θ) given $\beta, \sigma^2, \mathbf{y}$ is given by

$$\begin{aligned} \pi(\phi, \theta \mid \beta, \sigma^2, \mathbf{y}) &\propto \frac{1}{\theta^p} \left\{ \prod_{j=1}^p e^{-\frac{|\beta_j|}{\sigma \phi_j \theta}} \phi_j^{a-2} \right\} \theta^{pa-1} e^{-\frac{\theta}{\sigma}} \\ &\propto \theta^{p-1} \left\{ \prod_{j=1}^p e^{-\frac{|\beta_j|}{\sigma \phi_j \theta} - \frac{\theta \phi_j}{2}} (\theta \phi_j)^{a-2} \right\}. \end{aligned} \tag{4.8}$$

Bhattacharya et al. (2013) show that a sample from this density can be generated as follows: Draw random variables T_1, T_2, \dots, T_p with

$$T_j \mid \beta, \sigma^2, \mathbf{y} \sim \text{GIG}\left(a-1, 1, 2\frac{|\beta_j|}{\sigma}\right), \tag{4.9}$$

and set

$$\theta = \sum_{j=1}^p T_j \text{ and } \phi_j = \frac{T_j}{\theta} \text{ for } j = 1, 2, \dots, p. \tag{4.10}$$

Using the conditional densities described above, a Gibbs sampling Markov chain $(\beta_m, \eta_m, \sigma_m^2)_{m=0}^\infty$ can be constructed to sample from the density in (4.2). The state space of this Markov chain is $\tilde{\mathbb{R}}^p \times \Delta_p \times \mathbb{R}_+$, and the transition from $(\beta_m, \eta_m, \sigma_m^2)$ to $(\beta_{m+1}, \eta_{m+1}, \sigma_{m+1}^2)$ can be described as follows

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

1. Draw σ_{m+1}^2 from $\pi(\cdot | \boldsymbol{\beta}_m, \boldsymbol{\eta}_m, \mathbf{y})$ (as defined in (4.4)).
2. Draw $\boldsymbol{\eta}_{m+1} = (\boldsymbol{\psi}_{m+1}, \boldsymbol{\phi}_{m+1}, \theta_{m+1})$ in the following way:
 - (a) Draw $(\boldsymbol{\phi}_{m+1}, \theta_{m+1})$ from $\pi(\cdot | \boldsymbol{\beta}_m, \sigma_{m+1}^2, \mathbf{y})$ (as described in (4.8), (4.9), and (4.10)).
 - (b) Draw $\psi_{m+1,j}$ from $\pi(\cdot | \theta_{m+1}, \boldsymbol{\phi}_{m+1}, \boldsymbol{\beta}_{m+1}, \sigma_{m+1}^2, \mathbf{y})$ (as described in (4.5)) independently for $j = 1, \dots, p$.
3. Draw $\boldsymbol{\beta}_{m+1}$ from $\pi(\cdot | \boldsymbol{\eta}_{m+1}, \sigma_{m+1}^2, \mathbf{y})$ (as described in (4.3)).

The transition density of the Markov chain defined above (with respect to the Lebesgue measure on $\tilde{\mathbb{R}}^p \times \Delta_p \times \mathbb{R}_+$) is given by

$$k_{DL} \left((\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\eta}}, \tilde{\sigma}^2), (\boldsymbol{\beta}, \boldsymbol{\eta}, \sigma^2) \right) = \pi(\boldsymbol{\beta} | \boldsymbol{\eta}, \sigma^2, \mathbf{y}) \left\{ \prod_{j=1}^p \pi(\psi_j | \theta, \boldsymbol{\phi}, \tilde{\boldsymbol{\beta}}, \sigma^2, \mathbf{y}) \right\} \\ \times \pi(\boldsymbol{\phi}, \theta | \tilde{\boldsymbol{\beta}}, \sigma^2, \mathbf{y}) \pi(\sigma^2 | \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\eta}}, \mathbf{y}). \quad (4.11)$$

It is well known, and can be easily verified, that the joint density of $(\boldsymbol{\beta}, \boldsymbol{\eta}, \sigma^2)$ (conditioned on \mathbf{y}) defined in (4.2), is invariant for the Gibbs transition density k_{DL} defined above. Since the Markov transition density k_{DL} is strictly positive, by exactly the same argument as in Section 3, it can be shown that the corresponding Markov chain is Harris recurrent. We now prove geometric ergodicity by establishing a geometric drift condition and an associated minorization condition for the Gibbs transition density k_{DL} .

Remark 2. Note that the conditional densities of $\boldsymbol{\beta}$ and σ^2 (in (4.3) and (4.4)) are very similar to the conditional densities of $\boldsymbol{\beta}$ and σ^2 in the Normal-Gamma model (see (3.1) and (3.2)). Hence, for the steps in the analysis involving the conditional densities of $\boldsymbol{\beta}$ and σ^2 , we heavily use the results established for analogous steps in Section 3. However, the conditional density of the parameter $\boldsymbol{\eta}$ in the Dirichlet-Laplace model is significantly different and much more complicated than the conditional density of the parameter $\boldsymbol{\tau}$ in the Normal-Gamma model. Hence, steps involving components of $\boldsymbol{\eta}$ require a different and more involved analysis than in Section 3.

4.1. Drift condition

Let $\delta^*(a) = \operatorname{argmin}_{\delta \in (0,1)} \frac{\Gamma(\frac{1-\delta}{2})}{2^{\frac{\delta}{2}} \sqrt{\pi}} \frac{\Gamma(1+\frac{\delta}{2}-a)}{\Gamma(1-a)}$. Consider the following function

$$\tilde{V}(\boldsymbol{\beta}, \boldsymbol{\eta}, \sigma^2) = (\mathbf{y} - X\boldsymbol{\beta})^T (\mathbf{y} - X\boldsymbol{\beta}) + \boldsymbol{\beta}^T D_{\boldsymbol{\eta}}^{-1} \boldsymbol{\beta} + \sum_{j=1}^p \frac{1}{|\beta_j|^{\tilde{\delta}(a)}} + \sum_{j=1}^p \psi_j \theta^2 \phi_j^2,$$

where $\tilde{\delta}(a) = \delta^*(a)I_{[a < 1]} + \frac{1}{2}I_{[a=1]} + \min\{2(a-1), \frac{1}{2}\}I_{[a > 1]}$, and $D_{\boldsymbol{\eta}}$ is the diagonal matrix with diagonal elements $\{\psi_j \theta^2 \phi_j^2\}_{j=1}^p$. Note that $0 < \tilde{\delta}(a) < 1$.

Let $E_{k_{DL}}[\cdot \mid \beta_0, \boldsymbol{\eta}_0, \sigma_0^2]$ denote the expectation with respect to one step of the Markov chain with transition density k_{DL} , starting at $(\beta_0, \boldsymbol{\eta}_0, \sigma_0^2)$. The following proposition establishes a geometric *drift condition* for the transition density k_{DL} .

Proposition 4.1. *If $n \geq 3$ and $a > a^*$ (as defined in Proposition A8 (see appendix)), there exists constants $0 \leq \gamma < 1$ and $\varphi \geq 0$ such that*

$$E_{k_{DL}}[\tilde{V}(\boldsymbol{\beta}, \boldsymbol{\eta}, \sigma^2) \mid (\beta_0, \boldsymbol{\eta}_0, \sigma_0^2)] \leq \gamma \tilde{V}(\beta_0, \boldsymbol{\eta}_0, \sigma_0^2) + \varphi, \tag{4.12}$$

for every $(\beta_0, \boldsymbol{\eta}_0, \sigma_0^2) \in \tilde{\mathbb{R}}^p \times \Delta_p \times \mathbb{R}_+$.

Proof It follows by the definition of the Markov transition density k_{DL} that

$$\begin{aligned} & E_{k_{DL}}[\tilde{V}(\boldsymbol{\beta}, \boldsymbol{\eta}, \sigma^2) \mid (\beta_0, \boldsymbol{\eta}_0, \sigma_0^2)] \\ &= E[E[E[\tilde{V}(\boldsymbol{\beta}, \boldsymbol{\eta}, \sigma^2) \mid \boldsymbol{\eta}, \sigma^2, \mathbf{y}] \mid \beta_0, \sigma^2, \mathbf{y}] \mid \beta_0, \boldsymbol{\eta}_0, \mathbf{y}]. \end{aligned} \tag{4.13}$$

As in the proof of Proposition 3.1, we evaluate the three conditional expectations one step at a time. Note that the innermost conditional expectation is with respect to the conditional distribution of $\boldsymbol{\beta}$, and by (4.3),

$$\boldsymbol{\beta} \mid \boldsymbol{\eta}, \sigma^2, \mathbf{y} \sim N_p((X^T X + D_\eta^{-1})^{-1} X^T \mathbf{y}, \sigma^2 (X^T X + D_\eta^{-1})^{-1}).$$

Hence, the method for simplifying the innermost expectation here will be very similar to the method for simplifying the innermost expectation in the proof of Proposition 3.1. By exactly the same argument as the one leading to (3.6), it follows that

$$\begin{aligned} & E\left[(\mathbf{y} - X\boldsymbol{\beta})^T (\mathbf{y} - X\boldsymbol{\beta}) + \boldsymbol{\beta}^T D_\eta^{-1} \boldsymbol{\beta} \mid \boldsymbol{\eta}, \sigma^2, \mathbf{y}\right] \\ &= \mathbf{y}^T \mathbf{y} - \mathbf{y}^T X (X^T X + D_\eta^{-1})^{-1} X^T \mathbf{y} + p\sigma^2. \end{aligned} \tag{4.14}$$

Let $\tilde{\kappa}(a) := \frac{\Gamma(\frac{1-\tilde{\delta}(a)}{2}) 2^{\frac{1-\tilde{\delta}(a)}{2}}}{\sqrt{2\pi}}$ for $a > 0$. By a similar argument as the one leading to (3.7) and (3.8), it follows that for every $j = 1, 2, \dots, p$, $\beta_j \sim N(\mu_j, \sigma_j^2)$, where $\mu_j = \mathbf{e}_j^T (X^T X + D_\eta^{-1})^{-1} X^T \mathbf{y}$ and $\sigma_j^2 = \sigma^2 \mathbf{e}_j^T (X^T X + D_\eta^{-1})^{-1} \mathbf{e}_j$, and

$$E\left[\frac{1}{|\beta_j|^{\tilde{\delta}(a)}} \mid \boldsymbol{\eta}, \sigma^2, \mathbf{y}\right] \leq \frac{\tilde{\kappa}(a)}{\sigma_j^{\tilde{\delta}(a)}} \leq \tilde{\kappa}(a) \left(\frac{\lambda_X^{\frac{\tilde{\delta}(a)}{2}}}{\sigma^{\tilde{\delta}(a)}} + \frac{1}{\sigma^{\tilde{\delta}(a)} \psi_j^{\frac{\tilde{\delta}(a)}{2}} \theta^{\tilde{\delta}(a)} \phi_j^{\tilde{\delta}(a)}} \right). \tag{4.15}$$

It follows from (4.14) and (4.15) that

$$\begin{aligned} & E[\tilde{V}(\boldsymbol{\beta}, \boldsymbol{\eta}, \sigma^2) \mid \boldsymbol{\eta}, \sigma^2, \mathbf{y}] \\ & \leq \mathbf{y}^T \mathbf{y} + p\sigma^2 + \frac{p\tilde{\kappa}(a)\lambda_X^{\frac{\tilde{\delta}(a)}{2}}}{\sigma^{\tilde{\delta}(a)}} + \frac{\tilde{\kappa}(a)}{\sigma^{\tilde{\delta}(a)}} \sum_{j=1}^p \frac{1}{\psi_j^{\frac{\tilde{\delta}(a)}{2}} \theta^{\tilde{\delta}(a)} \phi_j^{\tilde{\delta}(a)}} + \sum_{j=1}^p \psi_j \theta^2 \phi_j^2. \end{aligned} \tag{4.16}$$

We now analyze the middle conditional expectation in (4.13), which requires getting an upper bound for $E\left[\frac{\tilde{\kappa}(a)}{\sigma^{\tilde{\delta}(a)}} \sum_{j=1}^p \frac{1}{\psi_j^{\frac{\tilde{\delta}(a)}{2}} \theta^{\tilde{\delta}(a)} \phi_j^{\tilde{\delta}(a)}} \mid \beta_0, \sigma^2, \mathbf{y}\right]$ and

$$E\left[\sum_{j=1}^p \psi_j \theta^2 \phi_j^2 \mid \beta_0, \sigma^2, \mathbf{y}\right].$$

We first consider $E\left[\sum_{j=1}^p \psi_j \theta^2 \phi_j^2 \mid \beta_0, \sigma^2, \mathbf{y}\right]$. Note by (4.9) and (4.10), $\phi_j \theta \mid \beta_0, \sigma^2, \mathbf{y} \sim \text{GIG}(a-1, 1, 2\frac{|\beta_{0j}|}{\sigma})$.

Hence

$$\begin{aligned} E[\phi_j \theta \mid \beta_0, \sigma^2, \mathbf{y}] &= \left(\sqrt{2\frac{|\beta_{0j}|}{\sigma}}\right) \frac{K_a\left(\sqrt{2\frac{|\beta_{0j}|}{\sigma}}\right)}{K_{a-1}\left(\sqrt{2\frac{|\beta_{0j}|}{\sigma}}\right)} \\ &\stackrel{(g)}{\leq} \left(\sqrt{2\frac{|\beta_{0j}|}{\sigma}}\right) \left(\frac{a + \sqrt{a^2 + 2\frac{|\beta_{0j}|}{\sigma}}}{\sqrt{2\frac{|\beta_{0j}|}{\sigma}}}\right) \\ &\leq 2a + \sqrt{2}\sqrt{\frac{|\beta_{0j}|}{\sigma}}, \end{aligned} \quad (4.17)$$

where (g) follows from (Laforgia and Natalini, 2009, Theorem 1.2). Furthermore,

$$\begin{aligned} E[(\phi_j \theta)^2 \mid \beta_0, \sigma^2, \mathbf{y}] &= \left(2\frac{|\beta_{0j}|}{\sigma}\right) \frac{K_{a+1}\left(\sqrt{2\frac{|\beta_{0j}|}{\sigma}}\right)}{K_{a-1}\left(\sqrt{2\frac{|\beta_{0j}|}{\sigma}}\right)} \\ &\stackrel{(h)}{\leq} \left(2\frac{|\beta_{0j}|}{\sigma}\right) \left\{\frac{K_{a+1}\left(\sqrt{2\frac{|\beta_{0j}|}{\sigma}}\right)}{K_a\left(\sqrt{2\frac{|\beta_{0j}|}{\sigma}}\right)}\right\}^2 \\ &\stackrel{(i)}{\leq} \left(2\frac{|\beta_{0j}|}{\sigma}\right) \left(\frac{(a+1) + \sqrt{(a+1)^2 + 2\frac{|\beta_{0j}|}{\sigma}}}{\sqrt{2\frac{|\beta_{0j}|}{\sigma}}}\right)^2 \\ &\leq 2\left((a+1)^2 + (a+1)^2 + 2\frac{|\beta_{0j}|}{\sigma}\right) \\ &\leq 4(a+1)^2 + 4\frac{|\beta_{0j}|}{\sigma}, \end{aligned} \quad (4.18)$$

where (h) follows from (Laforgia and Natalini, 2009, eq. (1.9)) and (i) follows from (Laforgia and Natalini, 2009, Theorem 1.2). By (4.6), we get that

$$\begin{aligned} &E[\psi_j \phi_j^2 \theta^2 \mid \beta_0, \sigma^2, \mathbf{y}] \\ &= E\left[\phi_j^2 \theta^2 E[\psi_j \mid \phi_j, \theta, \beta_0, \sigma^2, \mathbf{y}] \mid \beta_0, \sigma^2, \mathbf{y}\right] \\ &= E\left[\phi_j^2 \theta^2 \left(1 + \frac{|\beta_{0j}|}{\phi_j \theta \sigma}\right) \mid \beta_0, \sigma^2, \mathbf{y}\right] \end{aligned}$$

$$= E[\phi_j^2 \theta^2 \mid \beta_0, \sigma^2, \mathbf{y}] + \frac{|\beta_{0j}|}{\sigma} E[\phi_j \theta \mid \beta_0, \sigma^2, \mathbf{y}]. \tag{4.19}$$

Let $C_1, C_2 > 0$ be arbitrary (we will make appropriate choices for C_1 and C_2 at the end of the proof). It follows by (4.17), (4.18) and (4.19) that for every $j = 1, 2, \dots, p$,

$$\begin{aligned} & E[\psi_j \phi_j^2 \theta^2 \mid \beta_0, \sigma^2, \mathbf{y}] \\ \leq & 4(a+1)^2 + 4 \frac{|\beta_{0j}|}{\sigma} + \frac{|\beta_{0j}|}{\sigma} \left(2a + \sqrt{2} \sqrt{\frac{|\beta_{0j}|}{\sigma}} \right) \\ \leq & 4(a+1)^2 + 2(a+2) \frac{|\beta_{0j}|}{\sigma} + \sqrt{2} \frac{|\beta_{0j}|}{\sigma} \sqrt{\frac{|\beta_{0j}|}{\sigma}} \\ \stackrel{(j)}{\leq} & 4(a+1)^2 + 2(a+2) \left\{ \frac{C_1}{2} \left(\frac{|\beta_{0j}|}{\sigma} \right)^2 + \frac{1}{2C_1} \right\} + \sqrt{2} \left(\frac{|\beta_{0j}|}{\sigma} \right)^{\frac{3}{2}} \\ \stackrel{(j')}{\leq} & 4(a+1)^2 + \frac{a+2}{C_1} + C_1(a+2) \left(\frac{|\beta_{0j}|}{\sigma} \right)^2 + \left\{ C_2 \left(\frac{|\beta_{0j}|}{\sigma} \right)^2 + \frac{27}{64C_2^3} \right\} \\ \leq & 4(a+1)^2 + \frac{a+2}{C_1} + \frac{1}{2C_2^3} + [C_1(a+2) + C_2] \left(\frac{|\beta_{0j}|}{\sigma} \right)^2, \tag{4.20} \end{aligned}$$

where (j) follows from the fact that $uv \leq \frac{u^2}{2} + \frac{v^2}{2}$ for $u, v > 0$ (with $u = \sqrt{C_1} \frac{|\beta_{0j}|}{\sigma}, v = \frac{1}{\sqrt{C_1}}$) and (j') follows from the fact that $u^{\frac{3}{4}} v^{\frac{1}{4}} \leq \frac{3}{4}u + \frac{1}{4}v$ for $u, v > 0$ (with $u = \frac{1}{\frac{3\sqrt{2}}{4C_2}} \left(\frac{|\beta_{0j}|}{\sigma} \right)^2, v = \left(\frac{3\sqrt{2}}{4C_2} \right)^3$).

Next, we consider $E\left[\frac{\tilde{\kappa}(a)}{\sigma^{\tilde{\delta}(a)} \psi_j^{\frac{\tilde{\delta}(a)}{2}} \theta^{\tilde{\delta}(a)} \phi_j^{\tilde{\delta}(a)}} \mid \beta_0, \sigma^2, \mathbf{y}\right]$. Since $0 < \tilde{\delta}(a) < 1$, it follows that $x \mapsto x^{\tilde{\delta}(a)}$ is a concave function for $x > 0$. By Jensen's inequality and (4.7), we obtain

$$\begin{aligned} & E\left[\frac{1}{(\sigma \phi_j \theta)^{\tilde{\delta}(a)} \psi_j^{\frac{\tilde{\delta}(a)}{2}}} \mid \beta_0, \sigma^2, \mathbf{y}\right] \\ \leq & E\left[\frac{1}{(\sigma \phi_j \theta)^{\tilde{\delta}(a)}} \left\{ E\left[\frac{1}{\psi_j} \mid \beta_0, \sigma, \theta, \phi, \mathbf{y}\right] \right\}^{\frac{\tilde{\delta}(a)}{2}} \mid \beta_0, \sigma, \mathbf{y}\right] \\ = & E\left[\frac{1}{(\sigma \phi_j \theta)^{\tilde{\delta}(a)}} \left\{ \frac{\sigma \theta \phi_j}{|\beta_{0j}|} \right\}^{\frac{\tilde{\delta}(a)}{2}} \mid \beta_0, \sigma, \mathbf{y}\right] \\ = & \frac{1}{(\sigma |\beta_{0j}|)^{\frac{\tilde{\delta}(a)}{2}}} E\left[\frac{1}{(\phi_j \theta)^{\frac{\tilde{\delta}(a)}{2}}} \mid \beta_0, \sigma, \mathbf{y}\right]. \end{aligned}$$

$$\stackrel{(k)}{=} \frac{1}{(2\sigma|\beta_{0j}|^3)^{\frac{\tilde{\delta}(a)}{4}}} \frac{K_{a-\frac{\tilde{\delta}(a)}{2}-1}\left(\sqrt{2\frac{|\beta_{0j}|}{\sigma}}\right)}{K_{a-1}\left(\sqrt{2\frac{|\beta_{0j}|}{\sigma}}\right)}, \quad (4.21)$$

where (k) follows from (4.9), (4.10), and (2.1). Now consider the three cases, $a^* < a < 1$, $a > 1$ and $a = 1$, where a^* is as defined in Proposition A8 (see appendix).

Case 1: $a^* < a < 1$.

Let $\epsilon^* = \frac{1}{2} \left[\frac{2^{\frac{\delta^*(a)}{2}} \sqrt{\pi}}{\Gamma(\frac{1-\delta^*(a)}{2})} \frac{\Gamma(1-a)}{\Gamma(1+\frac{\delta^*(a)}{2}-a)} - 1 \right]$. Proposition A9 (see appendix) ensures that $\epsilon^* > 0$. Using Proposition A2 (see appendix) with $\nu_1 = 1 + \frac{\tilde{\delta}(a)}{2} - a$, $\nu_2 = 1 - a$ and $x = \sqrt{2\frac{|\beta_{0j}|}{\sigma}}$, there exists an ϵ depending on ϵ^*, a such that if $\sqrt{2\frac{|\beta_{0j}|}{\sigma}} < \epsilon$ then

$$\frac{K_{1+\frac{\tilde{\delta}(a)}{2}-a}\left(\sqrt{2\frac{|\beta_{0j}|}{\sigma}}\right)}{K_{1-a}\left(\sqrt{2\frac{|\beta_{0j}|}{\sigma}}\right)} \leq (1 + \epsilon^*) \frac{\Gamma(1 + \frac{\tilde{\delta}(a)}{2} - a) 2^{\frac{\tilde{\delta}(a)}{2}} \sigma^{\frac{\tilde{\delta}(a)}{4}}}{\Gamma(1 - a) (2|\beta_{0j}|)^{\frac{\tilde{\delta}(a)}{4}}}. \quad (4.22)$$

Since $a^* < a < 1$ and $K_\nu(x) = K_{-\nu}(x)$ (Abramowitz and Stegun, 1965, page 375), it follows by (4.21) and (4.22) that

$$\begin{aligned} & E \left[\frac{1}{(\sigma\phi_j\theta)^{\tilde{\delta}(a)} \psi_j^{\frac{\tilde{\delta}(a)}{2}}} \mid \beta_0, \sigma^2, \mathbf{y} \right] \\ & \leq \frac{(1 + \epsilon^*)}{(2\sigma|\beta_{0j}|^3)^{\frac{\tilde{\delta}(a)}{4}}} \frac{\Gamma(1 + \frac{\tilde{\delta}(a)}{2} - a) 2^{\frac{\tilde{\delta}(a)}{2}} \sigma^{\frac{\tilde{\delta}(a)}{4}}}{\Gamma(1 - a) (2|\beta_{0j}|)^{\frac{\tilde{\delta}(a)}{4}}} I_{\left[\sqrt{2\frac{|\beta_{0j}|}{\sigma}} < \epsilon\right]} + \\ & \quad \frac{1}{(2\sigma|\beta_{0j}|^3)^{\frac{\tilde{\delta}(a)}{4}}} \frac{K_{1+\frac{\tilde{\delta}(a)}{2}-a}\left(\sqrt{2\frac{|\beta_{0j}|}{\sigma}}\right)}{K_{1-a}\left(\sqrt{2\frac{|\beta_{0j}|}{\sigma}}\right)} I_{\left[\sqrt{2\frac{|\beta_{0j}|}{\sigma}} \geq \epsilon\right]}. \quad (4.23) \end{aligned}$$

Using 4.23 and Proposition A4 (see appendix) with $\nu = 1 - a$, $\delta = \tilde{\delta}(a)$, and $x = \sqrt{2\frac{|\beta_{0j}|}{\sigma}}$, we obtain

$$\begin{aligned} & E \left[\frac{1}{(\sigma\phi_j\theta)^{\tilde{\delta}(a)} \psi_j^{\frac{\tilde{\delta}(a)}{2}}} \mid \beta_0, \sigma^2, \mathbf{y} \right] \\ & \leq \frac{(1 + \epsilon^*)}{|\beta_{0j}|^{\tilde{\delta}(a)}} \frac{\Gamma(1 + \frac{\tilde{\delta}(a)}{2} - a)}{\Gamma(1 - a)} I_{\left[\sqrt{2\frac{|\beta_{0j}|}{\sigma}} < \epsilon\right]} + \end{aligned}$$

$$\begin{aligned} & \frac{1}{(2\sigma|\beta_{0j}|^3)^{\frac{\tilde{\delta}(a)}{4}}} \left[\frac{(3-2a)^{\frac{\tilde{\delta}(a)}{2}}}{\left[\frac{2|\beta_{0j}|}{\sigma}\right]^{\frac{\tilde{\delta}(a)}{4}}} + 1 \right] I\left[\sqrt{\frac{2|\beta_{0j}|}{\sigma}} \geq \epsilon\right] \\ \leq & \frac{1}{|\beta_{0j}|^{\tilde{\delta}(a)}} (1 + \epsilon^*) \frac{\Gamma(1 + \frac{\tilde{\delta}(a)}{2} - a)}{\Gamma(1 - a)} + \frac{1}{\sigma^{\tilde{\delta}(a)}} \frac{2^{\frac{\tilde{\delta}(a)}{2}}}{\epsilon^{\frac{3\tilde{\delta}(a)}{2}}} \left[\frac{(3-2a)^{\frac{\tilde{\delta}(a)}{2}}}{\epsilon^{\frac{\tilde{\delta}(a)}{2}}} + 1 \right]. \end{aligned} \tag{4.24}$$

Case 2: $a > 1$.

In the argument leading to (3.16), it was established that $\nu \mapsto K_\nu(x)$ is an increasing function for $\nu \geq 0$. It follows from (4.21), and the fact $v^{\frac{3}{4}}u^{\frac{1}{4}} \leq \frac{3}{4}v + \frac{1}{4}u \leq v + \frac{1}{4}u$ for $u, v > 0$, (with $u = \frac{[2\tilde{\kappa}(a)]^3}{(2\sigma)^{\tilde{\delta}(a)}}$ and $v = \frac{1}{2\tilde{\kappa}(a)|\beta_{0j}|^{\tilde{\delta}(a)}}$) that

$$\begin{aligned} E \left[\frac{1}{(\sigma\phi_j\theta)^{\tilde{\delta}(a)} \psi_j^{\frac{\tilde{\delta}(a)}{2}}} \mid \beta_0, \sigma^2, \mathbf{y} \right] & \leq \frac{1}{2\tilde{\kappa}(a)|\beta_{0j}|^{\tilde{\delta}(a)}} + \frac{2^{1-\tilde{\delta}(a)} (\tilde{\kappa}(a))^3}{\sigma^{\tilde{\delta}(a)}} \\ & \leq \frac{1}{2\tilde{\kappa}(a)|\beta_{0j}|^{\tilde{\delta}(a)}} + \frac{2 (\tilde{\kappa}(a))^3}{\sigma^{\tilde{\delta}(a)}}. \end{aligned} \tag{4.25}$$

Case 3: $a = 1$.

By Proposition A3 (see appendix) with $\nu = \frac{\tilde{\delta}(1)}{2}, x = \sqrt{2\frac{|\beta_{0j}|}{\sigma}}$, there exists $\epsilon' > 0$, such that for $\sqrt{2\frac{|\beta_{0j}|}{\sigma}} < \epsilon'$

$$\frac{K_{\frac{\tilde{\delta}(1)}{2}} \left(\sqrt{2\frac{|\beta_{0j}|}{\sigma}} \right)}{K_0 \left(\sqrt{2\frac{|\beta_{0j}|}{\sigma}} \right)} \leq \frac{\sigma^{\frac{\tilde{\delta}(1)}{4}}}{2\tilde{\kappa}(1) [2|\beta_{0j}|]^{\frac{\tilde{\delta}(1)}{4}}}. \tag{4.26}$$

Hence, using (4.21), we obtain

$$\begin{aligned} & E \left[\frac{1}{(\sigma\phi_j\theta)^{\tilde{\delta}(1)} \psi_j^{\frac{\tilde{\delta}(1)}{2}}} \mid \beta_0, \sigma^2, \mathbf{y} \right] \\ \leq & \frac{1}{(2\sigma|\beta_{0j}|^3)^{\frac{\tilde{\delta}(1)}{4}}} \frac{\sigma^{\frac{\tilde{\delta}(1)}{4}}}{2\tilde{\kappa}(1) [2|\beta_{0j}|]^{\frac{\tilde{\delta}(1)}{4}}} I\left[\sqrt{\frac{2|\beta_{0j}|}{\sigma}} < \epsilon'\right] \\ & + \frac{1}{(2\sigma|\beta_{0j}|^3)^{\frac{\tilde{\delta}(1)}{4}}} \frac{K_{\frac{\tilde{\delta}(1)}{2}} \left(\sqrt{2\frac{|\beta_{0j}|}{\sigma}} \right)}{K_0 \left(\sqrt{2\frac{|\beta_{0j}|}{\sigma}} \right)} I\left[\sqrt{\frac{2|\beta_{0j}|}{\sigma}} \geq \epsilon'\right] \end{aligned}$$

$$\begin{aligned}
& \stackrel{(k'')}{\leq} \frac{1}{2^{1+\frac{\delta(1)}{2}} \tilde{\kappa}(1) |\beta_{0j}|^{\tilde{\delta}(1)}} I \left[\sqrt{\frac{2|\beta_{0j}|}{\sigma}} < \epsilon' \right] \\
& \quad + \frac{1}{\sigma^{\tilde{\delta}(1)}} \frac{2^{\frac{\delta(1)}{2}}}{\epsilon'^{\frac{3\tilde{\delta}(1)}{2}}} \left[\frac{1}{\epsilon'^{\frac{\tilde{\delta}(1)}{2}}} + 1 \right] I \left[\sqrt{\frac{2|\beta_{0j}|}{\sigma}} \geq \epsilon' \right] \\
& \leq \frac{1}{2\tilde{\kappa}(1) |\beta_{0j}|^{\tilde{\delta}(1)}} + \frac{1}{\sigma^{\tilde{\delta}(1)}} \frac{2^{\frac{\delta(1)}{2}}}{\epsilon'^{\frac{3\tilde{\delta}(1)}{2}}} \left[\frac{1}{\epsilon'^{\frac{\tilde{\delta}(1)}{2}}} + 1 \right],
\end{aligned} \tag{4.27}$$

where (k'') follows from Proposition A4 (see appendix), with $\nu = 0$ and $x = \sqrt{2\frac{|\beta_{0j}|}{\sigma}}$.

Let us denote $D_3 = \frac{1}{2\tilde{\kappa}(a)} I_{[a \geq 1]} + (1 + \epsilon^*) \frac{\Gamma(1+\frac{\delta(a)}{2}-a)}{\Gamma(1-a)} I_{[a < 1]}$ and

$$\begin{aligned}
D_4 &= 2(\tilde{\kappa}(a))^3 I_{[a > 1]} + \frac{2^{\frac{\delta(a)}{2}}}{\epsilon'^{\frac{3\tilde{\delta}(a)}{2}}} \left[\frac{(3-2a)^{\frac{\delta(a)}{2}}}{\epsilon'^{\frac{\tilde{\delta}(a)}{2}}} + 1 \right] I_{[a < 1]} \\
& \quad + \frac{2^{\frac{\delta(1)}{2}}}{\epsilon'^{\frac{3\tilde{\delta}(1)}{2}}} \left[\frac{1}{\epsilon'^{\frac{\tilde{\delta}(1)}{2}}} + 1 \right] I_{[a=1]}.
\end{aligned}$$

Hence, combining the three cases, and using (4.24), (4.25) and (4.27), we have for every $a > a^*$,

$$E \left[\frac{1}{(\sigma \phi_j \theta)^{\tilde{\delta}(a)} \psi_j^{\frac{\tilde{\delta}(a)}{2}}} \mid \beta_0, \sigma^2, \mathbf{y} \right] \leq \frac{D_3}{|\beta_{0j}|^{\tilde{\delta}(a)}} + \frac{D_4}{\sigma^{\tilde{\delta}(a)}}. \tag{4.28}$$

It follows by (4.16), (4.20) and (4.28) that

$$\begin{aligned}
& E \left[E[\tilde{V}(\beta, \eta, \sigma^2) \mid \eta, \sigma^2, \mathbf{y}] \mid \beta_0, \sigma^2, \mathbf{y} \right] \\
& \leq \mathbf{y}^T \mathbf{y} + p\sigma^2 + \frac{p\tilde{\kappa}(a)\lambda_X^{\frac{\tilde{\delta}(a)}{2}}}{\sigma^{\tilde{\delta}(a)}} + \tilde{\kappa}(a) D_3 \sum_{j=1}^p \frac{1}{|\beta_{0j}|^{\tilde{\delta}(a)}} + \frac{p\tilde{\kappa}(a)D_4}{\sigma^{\tilde{\delta}(a)}} + \\
& \quad p \left(4(a+1)^2 + \frac{a+2}{C_1} + \frac{1}{2C_2^3} \right) + [C_1(a+2) + C_2] \sum_{j=1}^p \frac{|\beta_{0j}|^2}{\sigma^2}.
\end{aligned} \tag{4.29}$$

Finally, we consider the outermost conditional expectation in (4.13). It follows by (4.4) that

$$E[\sigma^2 \mid \beta_0, \eta_0, \mathbf{y}] = \frac{(\mathbf{y} - X\beta_0)^T (\mathbf{y} - X\beta_0) + \beta_0^T D_{\eta_0}^{-1} \beta_0 + 2\xi}{n + p + 2\alpha - 2}, \tag{4.30}$$

$$E \left[\frac{1}{\sigma^2} \mid \beta_0, \eta_0, \mathbf{y} \right] = \frac{n + p + 2\alpha}{(\mathbf{y} - X\beta_0)^T (\mathbf{y} - X\beta_0) + \beta_0^T D_{\eta_0}^{-1} \beta_0 + 2\xi}$$

$$\leq \frac{n+p+2\alpha}{\boldsymbol{\beta}_0^T D_{\boldsymbol{\eta}_0}^{-1} \boldsymbol{\beta}_0}. \quad (4.31)$$

Since $0 < \tilde{\delta}(a) < 1$, it follows that $g(x) = x^{\tilde{\delta}(a)}$ is a concave function for $x > 0$. By Jensen's inequality, we obtain

$$\begin{aligned} E \left[\frac{1}{\sigma^{\tilde{\delta}(a)}} \mid \boldsymbol{\beta}_0, \boldsymbol{\eta}_0, \mathbf{y} \right] &\leq \left[\frac{n+p+2\alpha}{(\mathbf{y} - X\boldsymbol{\beta}_0)^T (\mathbf{y} - X\boldsymbol{\beta}_0) + \boldsymbol{\beta}_0^T D_{\boldsymbol{\eta}_0}^{-1} \boldsymbol{\beta}_0 + 2\xi} \right]^{\frac{\tilde{\delta}(a)}{2}} \\ &\leq \left[\frac{n+p+2\alpha}{2\xi} \right]^{\frac{\tilde{\delta}(a)}{2}}. \end{aligned} \quad (4.32)$$

By (4.29), (4.30), (4.31) and (4.32), we obtain

$$\begin{aligned} &E_k[\tilde{V}(\boldsymbol{\beta}, \boldsymbol{\eta}, \sigma^2) \mid (\boldsymbol{\beta}_0, \boldsymbol{\eta}_0, \sigma_0^2)] \\ &= E[E[E[\tilde{V}(\boldsymbol{\beta}, \boldsymbol{\eta}, \sigma^2, \mathbf{y}) \mid \boldsymbol{\eta}, \sigma^2, \mathbf{y}] \mid \boldsymbol{\beta}_0, \sigma^2, \mathbf{y}] \mid \boldsymbol{\beta}_0, \boldsymbol{\eta}_0, \mathbf{y}]] \\ &\leq \mathbf{y}^T \mathbf{y} + p \left[\frac{(\mathbf{y} - X\boldsymbol{\beta}_0)^T (\mathbf{y} - X\boldsymbol{\beta}_0) + \boldsymbol{\beta}_0^T D_{\boldsymbol{\eta}_0}^{-1} \boldsymbol{\beta}_0 + 2\xi}{n+p+2\alpha-2} \right] \\ &\quad + \tilde{\kappa}(a) D_3 \sum_{j=1}^p \frac{1}{|\beta_{0j}|^{\tilde{\delta}(a)}} + \left[\frac{n+p+2\alpha}{2\xi} \right]^{\frac{\tilde{\delta}(a)}{2}} p \tilde{\kappa}(a) \left\{ \lambda_X^{\frac{\tilde{\delta}(a)}{2}} + D_4 \right\} \\ &\quad + p \left(4(a+1)^2 + \frac{a+2}{C_1} + \frac{1}{2C_2^3} \right) \\ &\quad + [C_1(a+2) + C_2] \sum_{j=1}^p (n+p+2\alpha) \frac{|\beta_{0j}|^2}{\boldsymbol{\beta}_0^T D_{\boldsymbol{\eta}_0}^{-1} \boldsymbol{\beta}_0}. \end{aligned} \quad (4.33)$$

By exactly the same argument as the one leading to (3.26), we obtain $\sum_{j=1}^p \frac{|\beta_{0j}|^2}{\boldsymbol{\beta}_0^T D_{\boldsymbol{\eta}_0}^{-1} \boldsymbol{\beta}_0} \leq \sum_{j=1}^p \psi_{0j} \theta_{0j}^2 \phi_{0j}^2$. It follows from (4.33) that

$$E_{k_{DL}}[\tilde{V}(\boldsymbol{\beta}, \boldsymbol{\eta}, \sigma^2) \mid (\boldsymbol{\beta}_0, \boldsymbol{\eta}_0, \sigma_0^2)] \leq \gamma \tilde{V}(\boldsymbol{\beta}_0, \boldsymbol{\eta}_0, \sigma_0^2) + \varphi, \quad (4.34)$$

where

$$\gamma = \text{Max} \left\{ \frac{p}{n+p+2\alpha-2}, [C_1(a+2) + C_2] (n+p+2\alpha), \tilde{\kappa}(a) D_3 \right\}$$

and

$$\begin{aligned} \varphi &= p \left(4(a+1)^2 + \frac{a+2}{C_1} + \frac{1}{2C_2^3} \right) \\ &\quad + \left[\frac{n+p+2\alpha}{2\xi} \right]^{\frac{\tilde{\delta}(a)}{2}} p \tilde{\kappa}(a) \left\{ \lambda_X^{\frac{\tilde{\delta}(a)}{2}} + D_4 \right\} + \mathbf{y}^T \mathbf{y} + \left[\frac{2p\xi}{n+p+2\alpha-2} \right]. \end{aligned}$$

The assumption $n \geq 3$ automatically leads to $\frac{p}{n+p+2\alpha-2} < 1$. Recall that C_1, C_2 are arbitrary positive numbers of our choice. If we choose the constants $C_1 =$

$\frac{1}{4(a+2)(n+p+2\alpha)}$ and $C_2 = \frac{1}{4(n+p+2\alpha)}$ then clearly $[C_1(a+2) + C_2](n+p+2\alpha) = \frac{1}{2}$. Note that $\tilde{\kappa}(a)D_3 = \frac{1}{2}I_{[a \geq 1]} + (1 + \epsilon^*)\tilde{\kappa}(a)\frac{\Gamma(1+\frac{\tilde{\delta}(a)}{2}-a)}{\Gamma(1-a)}I_{[a < 1]}$. If $a^* < a < 1$, then $\tilde{\delta}(a) = \delta^*(a)$, and it follows by the definition of ϵ^* and Proposition A9 (see appendix) that $\tilde{\kappa}(a)(1 + \epsilon^*)\frac{\Gamma(1+\frac{\tilde{\delta}(a)}{2}-a)}{\Gamma(1-a)} = \frac{1}{2} + \frac{1}{2}\frac{\Gamma(\frac{1-\delta^*(a)}{2})}{2^{\frac{\delta^*(a)}{2}}\sqrt{\pi}}\frac{\Gamma(1+\frac{\delta^*(a)}{2}-a)}{\Gamma(1-a)} < 1$. Hence if $a > a^*$, then $0 \leq \gamma < 1$, which establishes the required geometric drift condition.

4.2. Minorization condition

For every $d > 0$, let

$$B_{\tilde{V},d} = \left\{ (\beta_0, \boldsymbol{\eta}_0, \sigma_0^2) \in \tilde{\mathbb{R}}^p \times \Delta_p \times \mathbb{R}_+ : \tilde{V}(\beta_0, \boldsymbol{\eta}_0, \sigma_0^2) \leq d \right\}.$$

The following proposition establishes an associated *minorization* condition to the geometric drift condition in Proposition 4.1.

Proposition 4.2. *There exist a constant $0 < \epsilon^* = \epsilon^*(\tilde{V}, d) \leq 1$ and a probability density function f^* on $\tilde{\mathbb{R}}^p \times \Delta_p \times \mathbb{R}_+$ such that*

$$k_{DL}((\beta_0, \boldsymbol{\eta}_0, \sigma_0^2), (\boldsymbol{\beta}, \boldsymbol{\eta}, \sigma^2)) \geq \epsilon^* f^*(\boldsymbol{\beta}, \boldsymbol{\eta}, \sigma^2) \quad (4.35)$$

for every $(\beta_0, \boldsymbol{\eta}_0, \sigma_0^2) \in B_{\tilde{V},d}$.

Proof. Throughout this proof we will assume that $(\beta_0, \boldsymbol{\eta}_0, \sigma_0^2) \in B_{\tilde{V},d}$. To establish the minorization condition, we will construct lower bounds for appropriate conditional densities appearing in (4.11). If $(\beta_0, \boldsymbol{\eta}_0, \sigma_0^2) \in B_{\tilde{V},d}$, then by a similar argument as the one leading to (3.33), we get $\sum_{j=1}^p \beta_{0j}^2 \leq d^2$. Let $m_j = \frac{\phi_j \theta \sigma}{|\beta_{0j}|}$ and $m_{*j} = \frac{\phi_j \theta \sigma}{d}$. Since $|\beta_{0j}| \leq d$, it follows that $m_j \geq m_{*j}$. Hence

$$\begin{aligned} \frac{(1 - m_j \psi_j)^2}{2\psi_j m_j^2} &= \frac{(1 + m_j^2 \psi_j^2)}{2\psi_j m_j^2} - \frac{1}{m_j} \leq \frac{1}{2\psi_j m_j^2} + \frac{\psi_j}{2} \\ &\leq \frac{1}{2\psi_j m_{*j}^2} + \frac{\psi_j}{2} \\ &\leq \frac{(1 - m_{*j} \psi_j)^2}{2\psi_j m_{*j}^2} + \frac{1}{m_{*j}}. \end{aligned} \quad (4.36)$$

It follows by (4.5) and (4.36) that,

$$\begin{aligned} \pi(\psi_j \mid \theta, \phi, \beta_0, \sigma^2, \mathbf{y}) &\geq e^{-\frac{1}{m_{*j}}} \left(\frac{1}{2\pi\psi_j} \right)^{\frac{1}{2}} e^{-\frac{(1 - m_{*j} \psi_j)^2}{2\psi_j m_{*j}^2}} \\ &= \left(e^{-\frac{d}{\phi_j \theta \sigma}} \right) g_j(\psi_j \mid \theta, \phi, \sigma^2), \end{aligned} \quad (4.37)$$

where $g_j(\psi_j \mid \theta, \phi, \sigma^2)$ denotes the $\text{GIG}(\frac{1}{2}, 1, \frac{d^2}{\phi_j^2 \theta^2 \sigma^2})$ density. By (4.8) and (Bhattacharya et al., 2013, Page 17), we get that

$$\pi(\phi, \theta \mid \beta_0, \sigma^2, \mathbf{y}) = \theta^{p-1} \prod_{j=1}^p g_{0j}^*(\phi_j \theta \mid \beta_0, \sigma^2), \tag{4.38}$$

where $g_{0j}^*(\cdot \mid \beta_0, \sigma^2)$ denotes the $\text{GIG}(a-1, 1, 2\frac{|\beta_{0j}|}{\sigma})$ density. Note that

$$\begin{aligned} & e^{-\frac{d}{\phi_j \theta \sigma}} g_{0j}^*(\phi_j \theta \mid \beta_0, \sigma^2) \\ &= \frac{(\phi_j \theta)^{a-2} e^{-\frac{d}{\phi_j \theta \sigma}} e^{-\frac{1}{2}(\phi_j \theta + \frac{2|\beta_{0j}|}{\phi_j \theta \sigma})}}{2 \left(\frac{2|\beta_{0j}|}{\sigma}\right)^{\frac{a-1}{2}} K_{a-1}\left(\sqrt{\frac{2|\beta_{0j}|}{\sigma}}\right)} \\ &\geq \frac{(\phi_j \theta)^{a-2} e^{-\frac{1}{2}(\phi_j \theta + \frac{4d}{\phi_j \theta \sigma})}}{2 \left(\frac{2|\beta_{0j}|}{\sigma}\right)^{\frac{a-1}{2}} K_{a-1}\left(\sqrt{\frac{2|\beta_{0j}|}{\sigma}}\right)} \\ &= \left[\frac{2 \left(\frac{4d}{\sigma}\right)^{\frac{a-1}{2}} K_{a-1}\left(\sqrt{\frac{4d}{\sigma}}\right)}{2 \left(\frac{2|\beta_{0j}|}{\sigma}\right)^{\frac{a-1}{2}} K_{a-1}\left(\sqrt{\frac{2|\beta_{0j}|}{\sigma}}\right)} \right] \left\{ \frac{(\phi_j \theta)^{a-2} e^{-\frac{1}{2}(\phi_j \theta + \frac{4d}{\phi_j \theta \sigma})}}{2 \left(\frac{4d}{\sigma}\right)^{\frac{a-1}{2}} K_{a-1}\left(\sqrt{\frac{4d}{\sigma}}\right)} \right\} \\ &= \left[\frac{2 \left(\frac{4d}{\sigma}\right)^{\frac{a-1}{2}} K_{a-1}\left(\sqrt{\frac{4d}{\sigma}}\right)}{2 \left(\frac{2|\beta_{0j}|}{\sigma}\right)^{\frac{a-1}{2}} K_{a-1}\left(\sqrt{\frac{2|\beta_{0j}|}{\sigma}}\right)} \right] g_j^*(\phi_j \theta \mid \sigma^2), \tag{4.39} \end{aligned}$$

where $g_j^*(\cdot \mid \sigma^2)$ denotes the $\text{GIG}(a-1, 1, \frac{4d}{\sigma})$ density. Furthermore by Proposition A6 (see appendix) with $\nu = a-1$, $s = 4d$ and $t = \sigma$ we get

$$2 \left(\frac{4d}{\sigma}\right)^{\frac{a-1}{2}} K_{a-1}\left(\sqrt{\frac{4d}{\sigma}}\right) \geq e^{-\frac{d}{\sigma^2}} \left[e^{-\frac{1}{4}} K_{\frac{a-1}{2}}(\sqrt{2d}) (2d)^{\frac{a-1}{4}} \right]. \tag{4.40}$$

In the argument leading to (3.16), it was established that $\nu \mapsto K_\nu(x)$ is a increasing function for $\nu \geq 0$. Since $K_{-\nu}(x) = K_\nu(x)$ (see (Abramowitz and Stegun, 1965, Page 375)) and $|a-1| \leq a+1$ for $a \geq 0$, it follows that $K_{a-1}(\sqrt{\frac{2|\beta_{0j}|}{\sigma}}) \leq K_{a+1}(\sqrt{\frac{2|\beta_{0j}|}{\sigma}})$ for $a \geq 0$. Therefore it follows by (4.39), (4.40) and Proposition A7 (see appendix) that

$$\begin{aligned} & e^{-\frac{d}{\phi_j \theta \sigma}} g_{0j}^*(\phi_j \theta \mid \beta_0, \sigma^2) \\ &\geq \left[\frac{2 \left(\frac{4d}{\sigma}\right)^{\frac{a-1}{2}} K_{a-1}\left(\sqrt{\frac{4d}{\sigma}}\right)}{2 \left(\frac{2|\beta_{0j}|}{\sigma}\right)^{\frac{a-1}{2}} K_{a+1}\left(\sqrt{\frac{2|\beta_{0j}|}{\sigma}}\right)} \right] g_j^*(\phi_j \theta \mid \sigma^2) \end{aligned}$$

$$\begin{aligned}
&\geq \left[\frac{e^{-\frac{d}{\sigma^2}} \left[e^{-\frac{1}{4}} K_{\frac{a-1}{2}}(\sqrt{2d}) (2d)^{\frac{a-1}{4}} \right]}{\left(\frac{\sigma}{2|\beta_{0j}|} \right) \Gamma(a+1) 2^{a+1}} \right] g_j^*(\phi_j \theta \mid \sigma^2) \\
&\stackrel{(m)}{\geq} \left[\frac{e^{-\frac{1}{4}} K_{\frac{a-1}{2}}(\sqrt{2d}) (2d)^{\frac{a-1}{4}}}{d^{\frac{1}{\delta(a)}} \Gamma(a+1) 2^a} \right] \frac{e^{-\frac{d}{\sigma^2}}}{\sigma} g_j^*(\phi_j \theta \mid \sigma^2), \quad (4.41)
\end{aligned}$$

where (m) follows by the fact that $|\beta_{0j}| \geq \frac{1}{d^{\frac{1}{\delta(a)}}}$ for $(\beta_0, \eta_0, \sigma_0^2) \in B_{\tilde{V},d}$. Finally we bound the conditional density of σ^2 given β_0, η_0 and \mathbf{y} . Note that since $(\beta_0, \eta_0, \sigma_0^2) \in B_{\tilde{V},d}$, it follows that

$$\frac{(\mathbf{y} - X\beta_0)^T(\mathbf{y} - X\beta_0) + \beta_0^T D_{\eta_0}^{-1} \beta_0 + 2\xi}{2} \leq \frac{d + 2\xi}{2}. \quad (4.42)$$

Furthermore by exactly the same argument as the one leading to (3.42) it follows that

$$\begin{aligned}
&(\mathbf{y} - X\beta_0)^T(\mathbf{y} - X\beta_0) + \beta_0^T D_{\eta_0}^{-1} \beta_0 + 2\xi \\
&\geq \mathbf{y}^T \mathbf{y} - \mathbf{y}^T X \left(X^T X + \frac{1}{d} I_p \right)^{-1} X^T \mathbf{y} + 2\xi. \quad (4.43)
\end{aligned}$$

It follows from (4.4), (4.42) and (4.43) that for $(\beta_0, \eta_0, \sigma_0^2) \in B_{\tilde{V},d}$

$$\begin{aligned}
&\frac{e^{-\frac{pd}{\sigma^2}}}{\sigma^p} \pi(\sigma^2 \mid \beta_0, \eta_0, \mathbf{y}) \\
&= \frac{e^{-\frac{pd}{\sigma^2}} (\sigma^2)^{-\frac{n+p+2\alpha}{2}-1} \left((\mathbf{y} - X\beta_0)^T(\mathbf{y} - X\beta_0) + \beta_0^T D_{\eta_0}^{-1} \beta_0 + 2\xi \right)^{\frac{n+p+2\alpha}{2}}}{\sigma^p \frac{2^{\frac{n+p+2\alpha}{2}} \Gamma\left(\frac{n+p+2\alpha}{2}\right)}{\times \left\{ e^{-\frac{(\mathbf{y} - X\beta_0)^T(\mathbf{y} - X\beta_0) + \beta_0^T D_{\eta_0}^{-1} \beta_0 + 2\xi}{2\sigma^2}} \right\}}} \\
&\geq \left[\frac{\mathbf{y}^T \mathbf{y} - \mathbf{y}^T X \left(X^T X + \frac{1}{d} I_p \right)^{-1} X^T \mathbf{y} + 2\xi}{2} \right]^{\frac{n+p+2\alpha}{2}} \\
&\quad \times \left\{ \frac{(\sigma^2)^{-\frac{n+p+2\alpha}{2}-1} e^{-\frac{d+2\xi+2pd}{2\sigma^2}}}{\Gamma\left(\frac{n+p+2\alpha}{2}\right)} \right\} \\
&= \frac{\left[\frac{\mathbf{y}^T \mathbf{y} - \mathbf{y}^T X \left(X^T X + \frac{1}{d} I_p \right)^{-1} X^T \mathbf{y} + 2\xi}{2} \right]^{\frac{n+p+2\alpha}{2}} \Gamma\left(\frac{n+p+2\alpha}{2}\right)}{\left[\frac{d+2\xi+2pd}{2} \right]^{\frac{n+p+2\alpha}{2}} \Gamma\left(\frac{n+p+2\alpha}{2}\right)} h^*(\sigma^2), \quad (4.44)
\end{aligned}$$

where $h^*(\sigma^2)$ denotes the Inverse-Gamma($\frac{n+2p+2\alpha}{2}, \frac{d+2\xi+2pd}{2}$) density. Let

$$\begin{aligned}
 & f^*(\boldsymbol{\beta}, \boldsymbol{\eta}, \sigma^2) \\
 = & \pi(\boldsymbol{\beta} \mid \boldsymbol{\eta}, \sigma^2, \mathbf{y}) \left\{ \prod_{j=1}^p g_j(\psi_j \mid \theta, \phi, \sigma^2) \right\} \left\{ \theta^{p-1} \prod_{j=1}^p g_j^*(\phi_j \theta \mid \sigma^2) \right\} h^*(\sigma^2).
 \end{aligned} \tag{4.45}$$

It follows by (4.37), (4.41) and (4.44) that

$$k_{DL}((\boldsymbol{\beta}_0, \boldsymbol{\eta}_0, \sigma_0^2), (\boldsymbol{\beta}, \boldsymbol{\eta}, \sigma^2)) \geq \epsilon^* f^*(\boldsymbol{\beta}, \boldsymbol{\eta}, \sigma^2), \tag{4.46}$$

where

$$\begin{aligned}
 \epsilon^* = & \left[\frac{e^{-\frac{1}{4}} K_{\frac{a-1}{2}}(\sqrt{2d}) (2d)^{\frac{a-1}{4}}}{d^{\frac{1}{\delta(a)}} \Gamma(a+1) 2^a} \right]^p \\
 & \times \frac{\left[\frac{\mathbf{y}^T \mathbf{y} - \mathbf{y}^T X (X^T X + \frac{1}{d} I_p)^{-1} X^T \mathbf{y} + 2\xi}{2} \right]^{\frac{n+p+2\alpha}{2}} \Gamma\left(\frac{n+2p+2\alpha}{2}\right)}{\left[\frac{d+2\xi+2pd}{2} \right]^{\frac{n+2p+2\alpha}{2}} \Gamma\left(\frac{n+p+2\alpha}{2}\right)}.
 \end{aligned} \tag{4.47}$$

Note that $k_{DL}((\boldsymbol{\beta}_0, \boldsymbol{\eta}_0, \sigma_0^2), \cdot)$ and $f^*(\cdot)$ are both probability density functions. Hence, by (4.46), it follows that $\epsilon^* \leq 1$. \square

The drift and minorization conditions in Proposition 4.1 and Proposition 4.2 can be combined with (Rosenthal, 1995, Theorem 12) to establish geometric ergodicity of the Dirichlet-Laplace Markov chain with transition density k_{DL} .

Theorem 4.1. *Let p, X, α, ξ be arbitrary, $n \geq 3$, $a > a^*$ and $\gamma, \varphi, \epsilon$ be as defined in (4.34), (4.47). Let $d > \frac{2\varphi}{1-\gamma}$, $A = \frac{1+d}{1+2\varphi+\gamma d}$ and $U = 1 + 2(\gamma d + \varphi)$. Then for any $(\boldsymbol{\beta}_0, \boldsymbol{\eta}_0, \sigma_0^2) \in \tilde{\mathbb{R}}^p \times \Delta_p \times \mathbb{R}_+$, $r \in \mathbb{N}$ and $0 < s < 1$,*

$$\left\| K_{DL, (\boldsymbol{\beta}_0, \boldsymbol{\eta}_0, \sigma_0^2)}^r - \Pi_{DL} \right\|_{TV} \leq (1-\epsilon^*)^{rs} + \left(\frac{U^s}{A^{1-s}} \right)^r \left(1 + \frac{\varphi}{1-\gamma} + \tilde{V}(\boldsymbol{\beta}_0, \boldsymbol{\eta}_0, \sigma_0^2) \right)$$

where $K_{DL, (\boldsymbol{\beta}_0, \boldsymbol{\eta}_0, \sigma_0^2)}^r$ denotes the distribution of the Dirichlet-Laplace Markov chain started at $(\boldsymbol{\beta}_0, \boldsymbol{\eta}_0, \sigma_0^2)$ after r steps, and Π_{DL} denotes the joint distribution of $(\boldsymbol{\beta}, \boldsymbol{\eta}, \sigma^2)$ (conditioned on \mathbf{y}) in the Dirichlet-Laplace model.

5. Discussion

In recent years, a number of Bayesian shrinkage models have been developed to tackle high-dimensional data sets. As mentioned earlier, most of these models use Markov chains to sample from the intractable posterior distributions of the parameters. However, important theoretical properties such as geometric

ergodicity of these Markov chains have not been previously investigated. In this paper, geometric ergodicity of Gibbs sampling Markov chains corresponding to the Bayesian shrinkage models in Griffin and Brown (2010) and Bhattacharya et al. (2013) has been established. The new and interesting feature of these Markov chains, especially from the point of view of establishing geometric ergodicity, is the presence of the Generalized Inverse Gaussian distribution as one of the conditional distributions in the Gibbs sampler. To tackle this, we introduced novel drift functions which include the term $\sum_{i=1}^p \frac{1}{|\beta_j|^\delta}$ (recall that for both models, the conditional distribution of β given other parameters and the data is multivariate normal). The GIG distribution involves modified Bessel functions of the second kind, and the properties of these functions are invoked frequently in the intricate drift and minorization analysis presented in Section 3 and Section 4.

Note that the results in Section 3 (for the Normal-Gamma model) hold for every p, X, α, ξ, a and b . The results in Section 4 (for the Dirichlet-Laplace model) hold for all values of p, X, α, ξ , but require that $a > a^*$. As mentioned earlier, the structure of the conditional distribution of the parameter η in the Dirichlet-Laplace model is more complicated than the conditional distribution of the parameter τ in the Normal-Gamma model. Hence, it is not entirely surprising that we were able to prove a stronger result for the Normal-Gamma model. Using the insights obtained from the analysis in this paper, in future research we plan to investigate geometric ergodicity (or the lack of it) for the Dirichlet-Laplace models when $a \leq a^*$, and for Markov chains corresponding to some of the other Bayesian shrinkage models proposed in the literature.

Appendix: Mathematical identities

Proposition A1. *Let V be a normal random variable with mean μ and variance γ^2 . Then for any $\delta \in (0, 1)$, $E[\frac{1}{|V|^\delta}] \leq \frac{\Gamma(\frac{1-\delta}{2})2^{\frac{1-\delta}{2}}}{\sqrt{2\pi}\gamma^\delta}$.*

Proof. Let $f(x) = \frac{1}{\sqrt{2\pi}\gamma} e^{-\frac{(x-\mu)^2}{2\gamma^2}}$. For every $t > 0$, define

$$\begin{aligned} A_t &= \{x : f(x) > t\} \\ &= \left\{x : \frac{1}{\sqrt{2\pi}\gamma} e^{-\frac{(x-\mu)^2}{2\gamma^2}} > t\right\} = \left\{x : |x - \mu| < \sqrt{-2\gamma^2 \log(\sqrt{2\pi}\gamma t)}\right\}. \end{aligned} \tag{A.1}$$

Let A_t^* be the “symmetric rearrangement” of A_t defined by

$$\begin{aligned} A_t^* = \{x : 2|x| < \text{length}(A_t)\} &= \left\{x : |x| < \sqrt{-2\gamma^2 \log(\sqrt{2\pi}\gamma t)}\right\} \\ &= \left\{x : \frac{1}{\sqrt{2\pi}\gamma} e^{-\frac{x^2}{2\gamma^2}} > t\right\}. \end{aligned} \tag{A.2}$$

Let f^* be the “symmetric decreasing rearrangement” of f defined by

$$f^*(x) = \int_0^\infty I_{\{x \in A_t^*\}} dt = \int_0^\infty I_{\left\{ \frac{1}{\sqrt{2\pi}\gamma} e^{-\frac{x^2}{2\gamma^2}} > t \right\}} dt = \frac{1}{\sqrt{2\pi}\gamma} e^{-\frac{x^2}{2\gamma^2}}. \tag{A.3}$$

Note that the function $g(x) = \frac{1}{|x|^\delta}$ where $\delta \in (0, 1)$ is decreasing function of $|x|$ and symmetric around 0. It can be easily proved that the “symmetric decreasing rearrangement” of g is given by $g^*(x) = \frac{1}{|x|^\delta}$. By the Hardy-Littlewood inequality (Lieb and Loss, 2001, Page 82), we get that

$$\int_{\mathbb{R}} f(x)g(x)dx \leq \int_{\mathbb{R}} f^*(x)g^*(x)dx. \tag{A.4}$$

If $V \sim N(\mu, \gamma^2)$, It follows that

$$\begin{aligned} E \left[\frac{1}{|V|^\delta} \right] &= \int_{\mathbb{R}} \frac{1}{|x|^\delta} \frac{e^{-\frac{(x-\mu)^2}{2\gamma^2}}}{\sqrt{2\pi}\gamma} dx = \int_{\mathbb{R}} f(x)g(x)dx \leq \int_{\mathbb{R}} f^*(x)g^*(x)dx \\ &= 2 \int_{\mathbb{R}_+} \frac{1}{|x|^\delta} \frac{e^{-\frac{x^2}{2\gamma^2}}}{\sqrt{2\pi}\gamma} dx. \end{aligned}$$

By a change of variable with $u = \frac{x^2}{\gamma^2}$ we get that

$$E \left[\frac{1}{|V|^\delta} \right] \leq \frac{1}{\gamma^\delta} \int_{\mathbb{R}_+} u^{\frac{1-\delta}{2}-1} \frac{e^{-\frac{u}{2}}}{\sqrt{2\pi}} du = \frac{1}{\gamma^\delta} \left(\frac{2^{\frac{1-\delta}{2}} \Gamma(\frac{1-\delta}{2})}{\sqrt{2\pi}} \right). \tag{A.5}$$

□

Proposition A2. *Let $\nu_1, \nu_2 > 0$. Then for arbitrary $\epsilon^* > 0$ there exists $\epsilon > 0$ such that*

$$\frac{K_{\nu_1}(x)}{K_{\nu_2}(x)} \leq (1 + \epsilon^*) \frac{\Gamma(\nu_1)2^{\nu_1-\nu_2}}{\Gamma(\nu_2)x^{\nu_1-\nu_2}}$$

for all $0 < x < \epsilon$ (obviously ϵ will depend on ϵ^*, ν_1, ν_2).

Proof. By (Abramowitz and Stegun, 1965, Page 375), if $\nu > 0$, then $\frac{K_\nu(x)}{\left[\frac{2^{\nu}\Gamma(\nu)}{2x^\nu} \right]} \rightarrow$

1 as $x \rightarrow 0$. It follows that $\frac{\frac{K_{\nu_1}(x)}{\left[\frac{2^{\nu_1}\Gamma(\nu_1)}{2x^{\nu_1}} \right]}}{\frac{K_{\nu_2}(x)}{\left[\frac{2^{\nu_2}\Gamma(\nu_2)}{2x^{\nu_2}} \right]}} \rightarrow 1$ as $x \rightarrow 0$. Hence for arbitrary

positive ϵ^* there exists $\epsilon > 0$ such that $\frac{\frac{K_{\nu_1}(x)}{\left[\frac{2^{\nu_1}\Gamma(\nu_1)}{2x^{\nu_1}} \right]}}{\frac{K_{\nu_2}(x)}{\left[\frac{2^{\nu_2}\Gamma(\nu_2)}{2x^{\nu_2}} \right]}} \leq 1 + \epsilon^*$ for $0 < x < \epsilon$.

Therefore $\frac{K_{\nu_1}(x)}{K_{\nu_2}(x)} \leq (1 + \epsilon^*) \frac{\Gamma(\nu_1)2^{\nu_1-\nu_2}}{\Gamma(\nu_2)x^{\nu_1-\nu_2}}$ for $0 < x < \epsilon$. □

Proposition A3. *Let $\nu > 0$. Then there exists $\epsilon' > 0$ such that*

$$\frac{K_\nu(x)}{K_0(x)} \leq \frac{2^{\frac{1}{4}}\sqrt{\pi}}{2\Gamma\left(\frac{1}{4}\right)x^\nu},$$

for all $0 < x < \epsilon'$ (obviously ϵ' will depend on ν).

Proof. By (Abramowitz and Stegun, 1965, Page 375), if $\nu > 0$, then $\frac{K_\nu(x)}{\left[\frac{2^\nu\Gamma(\nu)}{2x^\nu}\right]} \rightarrow 1$ as $x \rightarrow 0$, and $\frac{K_0(x)}{[-\log(x)]} \rightarrow 1$ as $x \rightarrow 0$. Hence there exists $\epsilon > 0$ such that $\frac{\frac{K_\nu(x)}{\left[\frac{2^\nu\Gamma(\nu)}{2x^\nu}\right]}}{\frac{K_0(x)}{[-\log(x)]}} \leq 2$ for $0 < x < \epsilon$. Let $\epsilon' := \min\{\epsilon, e^{-\Gamma(\nu)2^{\nu+1}\kappa}\}$ where $\kappa = \frac{\Gamma(\frac{1}{4})}{2^{\frac{1}{4}}\sqrt{\pi}}$. Note that $0 < \epsilon' < 1$. As a result $x \mapsto -\log(x)$ is positive and strictly decreasing function on $x \in (0, \epsilon']$. Hence $\frac{K_\nu(x)}{K_0(x)} \leq \frac{2^\nu\Gamma(\nu)}{x^\nu[-\log(\epsilon')]}$ for $0 < x < \epsilon'$. As $0 < \epsilon' \leq e^{-\Gamma(\nu)2^{\nu+1}\kappa}$, it follows that

$$\frac{K_\nu(x)}{K_0(x)} \leq \frac{2^\nu\Gamma(\nu)}{x^\nu[-\log(e^{-\Gamma(\nu)2^{\nu+1}\kappa})]} = \frac{2^{\frac{1}{4}}\sqrt{\pi}}{2\Gamma\left(\frac{1}{4}\right)x^\nu} \quad \text{for } 0 < x < \epsilon'.$$

□

Proposition A4. *Let $\nu \geq 0$, $\delta \in (0, 1)$ and $x > 0$. Then*

$$\frac{K_{\nu+\frac{\delta}{2}}(x)}{K_\nu(x)} \leq \frac{(2\nu+1)^{\frac{\delta}{2}}}{x^{\frac{\delta}{2}}} + 1.$$

Proof. Let $\nu_1, \nu_2 \in \mathbb{R}, x > 0$ and $0 \leq \omega \leq 1$. By the integral form of the modified Bessel function of the second kind (Watson, 1995, page 181) we obtain,

$$K_{\omega\nu_1+(1-\omega)\nu_2}(x) = \int_0^\infty e^{-x.\cosh(t)} \cosh((\omega\nu_1+(1-\omega)\nu_2)t) dt.$$

Since the hyperbolic cosine function is a log-convex function, we obtain

$$\begin{aligned} & K_{\omega\nu_1+(1-\omega)\nu_2}(x) \\ & \leq \int_0^\infty e^{-x.\cosh(t)} (\cosh(\nu_1t))^\omega (\cosh(\nu_2t))^{1-\omega} dt \\ & = \int_0^\infty \left(e^{-x.\cosh(t)} \cosh(\nu_1t)\right)^\omega \left(e^{-x.\cosh(t)} \cosh(\nu_2t)\right)^{1-\omega} dt \\ & \stackrel{(m)}{\leq} \left(\int_0^\infty e^{-x.\cosh(t)} \cosh(\nu_1t) dt\right)^\omega \left(\int_0^\infty e^{-x.\cosh(t)} \cosh(\nu_2t) dt\right)^{1-\omega} \\ & = [K_{\nu_1}(x)]^\omega [K_{\nu_2}(x)]^{1-\omega}, \end{aligned} \tag{A.6}$$

where (m) follows from Holder's inequality. Hence, for $\nu \geq 0$ and $\delta \in (0, 1)$, we obtain

$$\frac{K_{\nu+\frac{\delta}{2}}(x)}{K_\nu(x)} = \frac{K_{\nu(1-\frac{\delta}{2})+(\nu+1)\frac{\delta}{2}}(x)}{K_\nu(x)}$$

$$\begin{aligned}
 &\leq \frac{[K_\nu(x)]^{1-\frac{\delta}{2}} [K_{(\nu+1)}(x)]^{\frac{\delta}{2}}}{K_\nu(x)} \\
 &= \left[\frac{K_{\nu+1}(x)}{K_\nu(x)} \right]^{\frac{\delta}{2}} \\
 &\stackrel{(q)}{\leq} \left[\frac{(\nu + \frac{1}{2}) + \sqrt{(\nu + \frac{1}{2})^2 + x^2}}{x} \right]^{\frac{\delta}{2}} \\
 &\leq \left[\frac{2(\nu + \frac{1}{2}) + x}{x} \right]^{\frac{\delta}{2}} \\
 &\leq \frac{(2\nu + 1)^{\frac{\delta}{2}}}{x^{\frac{\delta}{2}}} + 1. \tag{A.7}
 \end{aligned}$$

The last inequality follows from the fact that $(y + 1)^{\frac{\delta}{2}} \leq y^{\frac{\delta}{2}} + 1$ if $\delta \in (0, 1)$ and $y > 0$, and (q) follows by (Segura, 2011, Theorem 2). \square

Proposition A5. *Let $\nu \in \mathbb{R}$ and $s, t > 0$. Then $2K_\nu(\frac{s}{t}) \geq K_{\frac{\nu}{2}}(\frac{s^2}{2})e^{-\frac{1}{2t^2}}$.*

Proof. By choosing $\alpha = \frac{s}{t}$ and $\xi = \frac{s}{t}$ in (2.1), and the fact that $uv \leq \frac{u^2}{2} + \frac{v^2}{2}$ for $u, v \geq 0$, we obtain

$$\begin{aligned}
 2K_\nu\left(\frac{s}{t}\right) &= \int_0^\infty x^{\nu-1} e^{-\frac{1}{2}\left(x\frac{s}{t} + \frac{s}{tx}\right)} dx \\
 &\geq \int_0^\infty x^{\nu-1} e^{-\frac{1}{2}\left(\frac{x^2s^2}{2} + \frac{1}{2t^2} + \frac{s^2}{2x^2} + \frac{1}{2t^2}\right)} dx \\
 &= e^{-\frac{1}{2t^2}} \int_0^\infty x^{\nu-1} e^{-\frac{1}{2}\left(\frac{x^2s^2}{2} + \frac{s^2}{2x^2}\right)} dx \\
 &= e^{-\frac{1}{2t^2}} \int_0^\infty y^{\frac{\nu-1}{2}} e^{-\frac{1}{2}\left(\frac{ys^2}{2} + \frac{s^2}{2y}\right)} \frac{1}{2\sqrt{y}} dy \\
 &= e^{-\frac{1}{2t^2}} \frac{1}{2} \int_0^\infty y^{\frac{\nu}{2}-1} e^{-\frac{1}{2}\left(\frac{ys^2}{2} + \frac{s^2}{2y}\right)} dy \\
 &= e^{-\frac{1}{2t^2}} K_{\frac{\nu}{2}}\left(\frac{s^2}{2}\right). \tag{A.8}
 \end{aligned}$$

The last equality follows by choosing $\alpha = \frac{s^2}{2}$ and $\xi = \frac{s^2}{2}$ in (2.1). \square

Proposition A6. *Let $\nu \in \mathbb{R}$ and $s, t > 0$. Then*

$$2\left(\frac{s}{t}\right)^{\frac{\nu}{2}} K_\nu\left(\sqrt{\frac{s}{t}}\right) \geq e^{-\frac{s}{4t^2}} \left[e^{-\frac{1}{4}} K_{\frac{\nu}{2}}\left(\sqrt{\frac{s}{2}}\right) \left(\frac{s}{2}\right)^{\frac{\nu}{4}} \right].$$

Proof. By choosing $\alpha = 1$ and $\xi = \frac{s}{t}$ in (2.1), and the fact that $uv \leq \frac{u^2}{2} + \frac{v^2}{2}$ for $u, v \geq 0$, we obtain

$$2\left(\frac{s}{t}\right)^{\frac{\nu}{2}} K_\nu\left(\sqrt{\frac{s}{t}}\right) = \int_0^\infty x^{\nu-1} e^{-\frac{1}{2}\left(x + \frac{s}{xt}\right)} dx$$

$$\begin{aligned}
 &\geq \int_0^\infty x^{\nu-1} e^{-\frac{1}{2}\left(\frac{x^2+1}{2} + s\frac{\left(\frac{1}{x^2} + \frac{1}{t^2}\right)}{2}\right)} dx \\
 &= e^{-\frac{1}{2}\left(\frac{1}{2} + \frac{s}{2t^2}\right)} \int_0^\infty x^{\nu-1} e^{-\frac{1}{2}\left(x^2 + \frac{s}{2x^2}\right)} e^{\frac{x^2}{4}} dx \\
 &\geq e^{-\frac{1}{2}\left(\frac{1}{2} + \frac{s}{2t^2}\right)} \int_0^\infty \frac{1}{2} y^{\frac{\nu}{2}-1} e^{-\frac{1}{2}\left(y + \frac{s}{2y}\right)} dy \\
 &= e^{-\frac{s}{4t^2}} \left[e^{-\frac{1}{4}} K_{\frac{\nu}{2}} \left(\sqrt{\frac{s}{2}} \right) \left(\frac{s}{2} \right)^{\frac{\nu}{4}} \right]. \tag{A.9}
 \end{aligned}$$

The last equality follows by choosing $\alpha = 1$ and $\xi = \frac{s}{2}$ in (2.1). □

Proposition A7. Let $\nu \neq 0$ and $x > 0$. Then $2K_\nu(x) \leq x^{-|\nu|}\Gamma(|\nu|)2^{|\nu|}$.

Proof. If $\nu > 0$, then for any $x > 0$, by choosing $\alpha = 1$ and $\xi = x^2$ in (2.1), we obtain

$$2K_\nu(x) = x^{-\nu} \int_0^\infty y^{\nu-1} e^{-\frac{1}{2}\left(y + \frac{x^2}{y}\right)} dy \leq x^{-\nu} \int_0^\infty y^{\nu-1} e^{-\frac{y}{2}} dy \leq x^{-\nu}\Gamma(\nu)2^\nu. \tag{A.10}$$

If $\nu < 0$ then for any $x > 0$ we get from (Abramowitz and Stegun, 1965, Page 375) and (A.10) that

$$2K_\nu(x) = 2K_{-\nu}(x) \leq x^{-(-\nu)}\Gamma(-\nu)2^{-\nu} \leq x^\nu\Gamma(-\nu)2^{-\nu}. \tag{A.11}$$

□

Proposition A8. Let $\Psi(\alpha) = \frac{d}{d\alpha}\Gamma(\alpha)$ denote the digamma function. Then, there exists a unique $a^* \in (0, 1)$ such that $\frac{1}{2}\Psi(1 - a^*) - \frac{1}{2}\Psi\left(\frac{1}{2}\right) - \frac{1}{2}\log(2) = 0$.

Proof. Let $g(a) := \frac{1}{2}\Psi(1 - a) - \frac{1}{2}\Psi\left(\frac{1}{2}\right) - \frac{1}{2}\log(2)$. Note that $g'(a) = -\frac{1}{2}\Psi'(1 - a) < 0$ for $a \in (0, 1)$ as $\Psi'(\cdot)$ is always positive on the positive part of real line. Now $g(0) = 0.346 > 0$ and $g\left(\frac{1}{2}\right) = -\frac{1}{2}\log(2) < 0$. Since g is continuous and strictly decreasing on the interval $(0, 1)$, it follows that there exists a unique $a^* \in (0, 1)$ such that $g(a^*) = 0$.

Although there is no closed form expression for a^* , standard numerical methods reveal that $a^* \approx \frac{1}{\pi}$. □

Proposition A9. Let $a \in (a^*, 1)$. Then $\delta^*(a) = \operatorname{argmin}_{\delta \in (0, 1)} \frac{\Gamma\left(\frac{1-\delta}{2}\right)\Gamma\left(1 + \frac{\delta}{2} - a\right)}{2^{\frac{\delta}{2}}\sqrt{\pi}\Gamma(1-a)}$ is well defined, and

$$\frac{\Gamma\left(\frac{1-\delta^*(a)}{2}\right)\Gamma\left(1 + \frac{\delta^*(a)}{2} - a\right)}{2^{\frac{\delta^*(a)}{2}}\sqrt{\pi}\Gamma(1-a)} < 1.$$

Proof. Fix $a \in (a^*, 1)$ arbitrarily. Let $f_a(\delta) := \frac{\Gamma\left(\frac{1-\delta}{2}\right)\Gamma\left(1 + \frac{\delta}{2} - a\right)}{2^{\frac{\delta}{2}}\sqrt{\pi}\Gamma(1-a)}$. Let $g_a(\delta) := \frac{d}{d\delta}(\log(f_a(\delta))) = \frac{1}{2}\Psi\left(1 + \frac{\delta}{2} - a\right) - \frac{1}{2}\Psi\left(\frac{1}{2} - \frac{\delta}{2}\right) - \frac{1}{2}\log(2)$. It follows that $g'_a(\delta) = \frac{1}{4}\Psi'\left(1 + \frac{\delta}{2} - a\right) + \frac{1}{4}\Psi'\left(\frac{1}{2} - \frac{\delta}{2}\right) > 0$, as Ψ' is always positive on $(0, 1)$. As a result

$g_a(\delta)$ is strictly increasing on $(0, 1)$. Since $\Psi(\cdot)$ is an increasing function on $(0, 1)$, we obtain

$$\begin{aligned} g_a(0) &= \frac{1}{2} \left(\Psi(1-a) - \Psi\left(\frac{1}{2}\right) - \log(2) \right) \\ &< \frac{1}{2} \left(\Psi(1-a^*) - \Psi\left(\frac{1}{2}\right) - \log(2) \right) = 0. \end{aligned} \quad (\text{A.12})$$

for $a \in (a^*, 1)$. Note that for a fixed $\delta \in (0, 1)$, the function $a \mapsto g_a(\delta)$ is decreasing in $a \in (0, 1)$, hence for $a \in (a^*, 1)$

$$g_a(.8) > g_1(.8) = \frac{1}{2} (\Psi(.4) - \Psi(.1) - \log(2)) > 0. \quad (\text{A.13})$$

Since the function $\delta \mapsto g_a(\delta)$ is strictly increasing and differentiable on $(0, 1)$ with $g_a(0) < 0$ and $g_a(.8) > 0$, it follows that there exist a number $\delta^*(a) \in (0, 1)$ such that $g_a(\delta^*(a)) = 0$. Since $g_a(\delta) < 0$ for $\delta < \delta^*(a)$ and $g_a(\delta) > 0$ for $\delta > \delta^*(a)$, it follows that $\delta = \delta^*(a)$ is the point of global minima for the function $\delta \mapsto f_a(\delta)$ on $(0, 1)$. Note that $g_a(\delta) = \frac{d}{d\delta} \log(f_a(\delta)) < 0$ for $\delta < \delta^*(a)$. Therefore $f_a(\delta^*(a)) < f_a(0) = 1$. \square

References

- ABRAMOWITZ, M. and STEGUN, I. A. (1965). *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables (Dover Books on Mathematics)*, 1 ed. *Dover books on mathematics*. Dover Publications.
- ASMUSSEN, S. and GLYNN, P. W. (2011). A new proof of convergence of MCMC via the ergodic theorem. *Statistics and Probability Letters* **81** 1482–1485. [MR2818658](#)
- BARICZ, A. (2008). Turan type inequalities for some special functions. *Ph.D. Thesis, Institute of Mathematics, University of Debrecen*.
- BHATTACHARYA, A., PATI, D., PILLAI, N. S. and DUNSON, D. B. (2013). Bayesian shrinkage. *arXiv*.
- CARVALHO, C. M., POLSON, N. G., SCOTT, J. G. and JAMES, G. (2010). The horseshoe estimator for sparse signals. *Biometrika* **97** 465–480. [MR2650751](#)
- CHAN, K. S. and GEYER, C. J. (1994). Discussion of the paper “Markov chains for exploring posterior distributions” by Luke Tierney. *Leipzig und Berlin* **22** 1747–1758. [MR1329166](#)
- CHOI, H. M. and HOBERT, J. P. (2013). Analysis of MCMC algorithms for Bayesian linear regression with Laplace errors. *Journal of Multivariate Analysis* **117** 32–40. [MR3053533](#)
- FIGUEIREDO, M. A. T. (2003). Adaptive sparseness for supervised learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25** 1150–1159.
- FLEGAL, J. M. and JONES, G. L. (2010). Convergence Batch means and spectral variance estimators in Markov chain Monte Carlo. *Annals of Statistics* **38** 1034–1070. [MR2604704](#)

- GRIFFIN, J. E. and BROWN, P. J. (2005). Alternative prior distributions for variable selection with very many more variables than observations. *Technical report, University of Warwick* **32** 1594–1649.
- GRIFFIN, J. E. and BROWN, P. J. (2010). Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis* **5** 171–188. [MR2596440](#)
- GURLAND, J. (1956). An inequality satisfied by the gamma function. *Skand. Aktuarietidskr* **39** 171–172. [MR0095293](#)
- HOBERT, J. P. (2001). Relationships among the Gibbs sampler and its sub chains Stability. *Journal of Computational and Graphical Statistics* **10** 185–205. [MR1939696](#)
- HOBERT, J. P. and GEYER, C. J. (1998). Geometric ergodicity of Gibbs and block Gibbs samplers for a hierarchical random effects model. *Journal of Multivariate Analysis* **67** 414–430. [MR1659196](#)
- JARNER, S. F. and HANSEN, E. (2000). Geometric ergodicity of Metropolis algorithms. *Stochastic Processes and Their Applications* **85** 341–361. [MR1731030](#)
- JOHNSTONE, I. M. and SILVERMAN, B. W. (2002). Needles and straw in haystacks: Empirical Bayes estimates of possibly sparse sequences. *Annals of Statistics* **32** 1594–1649. [MR2089135](#)
- JONES, G. L. and HOBERT, J. P. (2001). Honest Exploration of Intractable Probability Distributions Via Markov Chain Monte Carlo. *Statistical Science* **16** 312–334. [MR1888447](#)
- JONES, G. L. and HOBERT, J. P. (2004). Sufficient burn-in for Gibbs samplers for a hierarchical random effects model. *Annals of Statistics* **3** 784–817. [MR2060178](#)
- KHARE, K. and HOBERT, J. P. (2012). Geometric ergodicity of the Gibbs sampler for Bayesian quantile regression. *Journal of Multivariate Analysis* **112** 108–116. [MR2957289](#)
- KHARE, K. and HOBERT, J. P. (2013). Geometric ergodicity of the Gibbs sampler for Bayesian lasso. *Electronic Journal of Statistics* **7** 2150–2163. [MR3104915](#)
- LAFORGIA, A. (1991). Bounds for modified Bessel functions. *Journal of Computational and Applied Mathematics* **34** 263–267. [MR1102583](#)
- LAFORGIA, A. and NATALINI, P. (2009). Some inequalities for modified Bessel functions. *Journal of Inequalities and Applications*. [MR2592860](#)
- LIEB, E. H. and LOSS, M. (2001). *Analysis*, 2 ed. American Mathematical Society. [MR1817225](#)
- MARCHEV, D. and HOBERT, J. P. (2004). Geometric ergodicity of van Dyk and Meng’s algorithm for the multivariate Student’s t model. *Journal of the American Statistical Association* **99** 228–238. [MR2054301](#)
- MENGERSEN, K. and TWEEDIE, R. L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Annals of Statistics* **24** 101–121. [MR1389882](#)
- MEYN, S. and TWEEDIE, R. L. (2009). *Markov Chains and Stochastic Stability*, 2 ed. Cambridge University Press, New York, NY, USA. [MR2509253](#)

- MIRA, A. and TIERNEY, L. (1997). On the use of auxiliary variables in Markov chain Monte Carlo sampling. *Scandinavian Journal of Statistics*.
- MYKLAND, P., TIERNEY, L. and YU, B. (1995). Discussion regeneration in Markov chain samplers. *Journal of the American Statistical Association* **90** 233–241. [MR1325131](#)
- PARK, T. and CASELLA, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association* **103** 681–686. [MR2524001](#)
- POLSON, N. G. and SCOTT, J. G. (2010). Shrink Globally, Act Locally: Sparse Bayesian Regularization and Prediction. **5** 73–80.
- ROBERT, C. P. (1995). Convergence control methods for Markov chain Monte Carlo algorithms. *Statistical Science* **10** 231–253. [MR1390517](#)
- ROBERTS, G. O. and ROSENTHAL, J. S. (1998b). Convergence rates of Gibbs samplers for uniform distributions. *Annals of Applied Probability* **8** 1291–1302. [MR1661176](#)
- ROBERTS, G. O. and ROSENTHAL, J. S. (1999). Convergence of slice sampler Markov chains. *Journal of the Royal Statistical Society, Series B* **61** 643–660. [MR1707866](#)
- ROBERTS, G. O. and TWEEDIE, R. L. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika* **83** 95–110. [MR1399158](#)
- ROMAN, J. C. and HOBERT, J. P. (2012). Convergence analysis of the Gibbs sampler for Bayesian general linear mixed models with improper priors. *Annals of Statistics* **40** 2823–2849. [MR3097961](#)
- ROSENTHAL, J. S. (1995). Minorization conditions and convergence rates for Markov chain Monte Carlo. *Journal of the American Statistical Association* **90** 558–566. [MR1340509](#)
- ROSENTHAL, J. S. (1996). Analysis of the Gibbs sampler for a model related to James-Stein estimators. *Statistics and Computing* **6** 269–275.
- ROY, V. and HOBERT, J. P. (2007). Convergence rates and asymptotic standard errors for MCMC algorithms for Bayesian probit regression. *Journal of the Royal Statistical Society, Series B* **69** 607–623. [MR2370071](#)
- SEGURA, J. (2011). Bounds for ratios of modified Bessel functions and associated Turan-type inequalities. *Journal of Mathematical Analysis and Applications* **374** 516–528. [MR2729238](#)
- TAN, A. and HOBERT, J. P. (2009). Block Gibbs sampling for Bayesian random effects models with improper priors: Convergence and regeneration. *Journal of Computational and Graphical Statistics* **18** 861–878. [MR2598033](#)
- TAN, A., JONES, G. L. and HOBERT, J. P. (2012). On the geometric ergodicity of two-variable Gibbs samplers. *Advances in Modern Statistical Theory and Applications: A Festschrift in Honor of Morris L. Eaton*.
- TIBSHIRANI, R. (1994). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B* **58** 267–288. [MR1379242](#)
- TIPPING, M. E. and SMOLA, A. (2001). *Sparse Bayesian Learning and the Relevance Vector Machine*.
- WATSON, G. N. (1995). *A Treatise on the Theory of Bessel Functions*, 2 ed. Cambridge University Press. [MR1349110](#)