

Geometric ergodicity of the Bayesian lasso

Kshitij Khare and James P. Hobert

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

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

Abstract: Consider the standard linear model $\mathbf{y} = X\boldsymbol{\beta} + \sigma\epsilon$, where the components of ϵ are iid standard normal errors. Park and Casella [14] consider a Bayesian treatment of this model with a Laplace/Inverse-Gamma prior on $(\boldsymbol{\beta}, \sigma)$. They introduce a Data Augmentation approach that can be used to explore the resulting intractable posterior density, and call it the Bayesian lasso algorithm. In this paper, the Markov chain underlying the Bayesian lasso algorithm is shown to be geometrically ergodic, for *arbitrary* values of the sample size n and the number of variables p . This is important, as geometric ergodicity 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. Kyung et al. [12] provide a proof of geometric ergodicity for the restricted case $n \geq p$, but as we explain in this paper, their proof is incorrect. Our approach is different and more direct, and enables us to establish geometric ergodicity for arbitrary n and p .

MSC 2010 subject classifications: Primary 60J27; secondary 62F15.

Keywords and phrases: Convergence rate, geometric drift condition, Markov chain, Bayesian lasso.

Received March 2013.

1. Introduction

Consider the standard linear model $\mathbf{y} = X\boldsymbol{\beta} + \sigma\epsilon$, 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 ϵ are iid standard normal errors, and σ^2 is the (unknown) variance parameter. The objective is to estimate $(\boldsymbol{\beta}, \sigma^2)$. In a variety of modern datasets from genetics, finance, and environmental sciences, the dimension p of $\boldsymbol{\beta}$ is much larger than the sample size n . An extremely popular approach to handle these kinds of datasets is the *lasso*, which was introduced in Tibshirani [19]. In this approach, the estimate of $\boldsymbol{\beta}$ is obtained as

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

where λ is a user-specified tuning parameter. Note that the estimate in (1.1) can be regarded as the posterior mode of β (conditional on σ^2) if one puts independent Laplace priors on the entries of β . Based on this observation, several authors proposed a Bayesian analysis using a Laplace-like prior for β (see for example [3, 7, 20]). Park and Casella [14] construct the following hierarchical Bayesian model as a Bayesian parallel/alternative to the lasso.

$$y \mid \beta, \sigma^2, \tau^2 \sim \mathcal{N}_n(X\beta, \sigma^2 I_n) \tag{1.2}$$

$$\beta \mid \sigma^2, \tau^2 \sim \mathcal{N}_p(\mathbf{0}_p, \sigma^2 D_\tau), \text{ where } D_\tau = \text{diag}(\tau_1^2, \tau_2^2, \dots, \tau_p^2) \tag{1.3}$$

$$\sigma^2 \sim \text{Inverse-Gamma}(\alpha, \xi) \text{ (allow for impropriety via } \alpha = 0 \text{ or } \xi = 0) \tag{1.4}$$

$$\tau_j^2 \stackrel{i.i.d.}{\sim} \text{Exponential}\left(\frac{\lambda^2}{2}\right) \text{ for } j = 1, 2, \dots, p \tag{1.5}$$

To see the connection with lasso, note that the prior density of β given σ^2 is given by

$$f(\beta \mid \sigma^2) = \prod_{i=1}^p \frac{\lambda}{2\sigma} e^{-\frac{\lambda|\beta_j|}{\sigma}}.$$

Indeed, it is shown in [14] that the full conditional distributions of β, τ^2, σ^2 are given by

$$\beta \mid \sigma^2, \tau^2, y \sim \mathcal{N}_p((X^T X + D_\tau^{-1})^{-1} X^T y, \sigma^2 (X^T X + D_\tau^{-1})^{-1}) \tag{1.6}$$

$$\frac{1}{\tau_j^2} \mid \beta, \sigma^2, y \sim \text{Inverse-Gaussian}\left(\sqrt{\frac{\lambda^2 \sigma^2}{\beta_j^2}}, \lambda^2\right), \text{ independently for } 1 \leq j \leq p \tag{1.7}$$

$$\sigma^2 \mid \beta, \tau^2, y \sim \text{Inverse-Gamma}\left(\alpha', \frac{(y - X\beta)^T (y - X\beta) + \beta^T D_\tau^{-1} \beta + 2\xi}{2}\right), \tag{1.8}$$

where $\alpha' = \frac{n+p+2\alpha}{2}$. Note that the Inverse-Gaussian(μ', λ') density is given by

$$f(x) = \sqrt{\frac{\lambda'}{2\pi}} x^{-\frac{3}{2}} e^{-\frac{\lambda'(x-\mu')^2}{2(\mu')^2 x}}.$$

Using the full conditional distributions specified above, one can construct a systematic scan Gibbs sampling algorithm to sample from the joint posterior density of $(\beta, \tau^2, \sigma^2)$ (and hence to obtain the desired sample from the posterior density of (β, σ^2)). Let $\{(\beta_m, \tau_m^2, \sigma_m^2)\}_{m=0}^\infty$ be a Markov chain (with state space $\mathbb{R}^p \times \mathbb{R}_+^p \times \mathbb{R}_+$) whose dynamics are defined (implicitly) through the following three-step procedure for moving from the current state, $(\beta_n, \tau_n^2, \sigma_n^2)$, to $(\beta_{n+1}, \tau_{n+1}^2, \sigma_{n+1}^2)$.

Iteration $n + 1$ of Park and Casella's Bayesian lasso Gibbs sampler

1. Draw σ_{n+1}^2 from the conditional distribution in (1.8) given (β_n, τ_n^2, y) .

2. Draw τ_{n+1}^2 from the conditional distribution in (1.7) given $(\beta_n, \sigma_{n+1}^2, \mathbf{y})$.
3. Draw β_{n+1} from the conditional distribution in (1.6) given $(\tau_{n+1}^2, \sigma_{n+1}^2, \mathbf{y})$.

In Section 2, the Markov transition density (Mtd) of the resulting Gibbs Markov chain, $\{(\beta_m, \tau_m^2, \sigma_m^2)\}_{m=0}^\infty$, is defined and then used to establish that the chain is well behaved (i.e., Harris ergodic) and converges to the target posterior distribution. Thus, we can use this chain to construct strongly consistent estimators of intractable posterior expectations. To be specific, for $q > 0$, let $L_q(\pi)$ denote the set of functions $g : \mathbb{R}^p \times \mathbb{R}_+^p \times \mathbb{R}_+ \rightarrow \mathbb{R}$ such that

$$E_\pi |g|^q := \int_{\mathbb{R}^p} \int_{\mathbb{R}_+^p} \int_{\mathbb{R}_+} |g(\beta, \tau^2, \sigma^2)|^q \pi(\beta, \tau^2, \sigma^2 | \mathbf{y}) d\sigma^2 d\tau^2 d\beta < \infty,$$

where $\pi(\beta, \tau^2, \sigma^2 | \mathbf{y})$ represents the joint posterior density evaluated at $(\beta, \tau^2, \sigma^2)$. Harris ergodicity implies that, if $g \in L_1(\pi)$, then the estimator

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

is strongly consistent for $E_\pi g$, no matter how the chain is started. Of course, in practice, an estimator is only useful if it is possible to compute an associated standard error. All available methods of computing a valid asymptotic standard error for \bar{g}_m are based on the existence of a central limit theorem (CLT) for \bar{g}_m ; that is, we require that

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

for some positive, finite ϕ^2 . Unfortunately, even if $g \in L_q(\pi)$ for all $q > 0$, Harris ergodicity is not enough to guarantee the existence of such a CLT (see for example [16, 17]). The standard method of establishing the existence of CLTs is to prove that the underlying Markov chain converges at a geometric rate.

Let $\mathcal{B}(\mathbf{X})$ denote the Borel sets in $\mathbf{X} := \mathbb{R}^p \times \mathbb{R}_+^p \times \mathbb{R}_+$, and let $K^m : \mathbf{X} \times \mathcal{B}(\mathbf{X}) \rightarrow [0, 1]$ denote the m -step Markov transition function of the Gibbs Markov chain. That is, $K^m((\beta_0, \tau_0^2, \sigma_0^2), A)$ is the probability that $(\beta_m, \tau_m^2, \sigma_m^2) \in A$, given that the chain is started at $(\beta_0, \tau_0^2, \sigma_0^2)$. Also, let $\Pi(\cdot)$ denote the joint posterior distribution. The chain is called geometrically ergodic if there exist a function $M : \mathbf{X} \rightarrow [0, \infty)$ and a constant $\gamma \in [0, 1)$ such that, for all $(\beta, \tau^2, \sigma^2) \in \mathbf{X}$ and all $m = 0, 1, \dots$, we have

$$\|K^m((\beta, \tau^2, \sigma^2), \cdot) - \Pi(\cdot)\|_{\text{TV}} \leq M(\beta, \tau^2, \sigma^2) \gamma^m,$$

where $\|\cdot\|_{\text{TV}}$ denotes the total variation norm. The relationship between geometric convergence and CLTs is simple: If the chain is geometrically ergodic and $E_\pi |g|^{2+\delta} < \infty$ for some $\delta > 0$, then \bar{g}_m satisfies a CLT. Moreover, because the

Mtd is strictly positive on X (see Section 2), the same $2 + \delta$ moment condition implies that the usual estimators of the asymptotic variance, ϕ^2 , are consistent [2, 8, 9, 11]. Our main result, which is proven in Section 3 using a geometric drift condition and an associated minorization condition, is the following.

Proposition 1. *The Bayesian lasso Gibbs Markov chain is geometrically ergodic for $n \geq 3$ and arbitrary p, X, λ .*

Hence, the Markov chain CLT holds and can be used to obtain asymptotic standard errors of posterior estimates.

In the restricted case when $n \geq p$, Kyung et al. [12] contains, among other results, a proof of geometric ergodicity of the Bayesian lasso Gibbs Markov chain. However, there are serious errors in this proof, which essentially arise from their assertion that the Bayesian lasso model in (1.2)–(1.5) can be obtained as a marginal model of a hierarchical random effects model. A detailed explanation of the problems with the proof are provided in the appendix.

2. The Bayesian lasso Markov chain

We first provide the expressions for the full posterior conditional densities of β , τ^2 and σ^2 respectively. Let $f(\cdot | \tau^2, \sigma^2, \mathbf{y})$ denote the full posterior conditional density of β (on \mathbb{R}^p) given $\tau^2, \sigma^2, \mathbf{y}$. By (1.6),

$$f(\beta | \tau^2, \sigma^2, \mathbf{y}) = \frac{|X^T X + D_{\tau}^{-1}|^{\frac{1}{2}}}{\sqrt{2\pi\sigma^2}^p} e^{-\frac{(\beta - (X^T X + D_{\tau}^{-1})^{-1} X^T \mathbf{y})^T (X^T X + D_{\tau}^{-1}) (\beta - (X^T X + D_{\tau}^{-1})^{-1} X^T \mathbf{y})}{2\sigma^2}}.$$

Let $f(\cdot | \beta, \sigma^2, \mathbf{y})$ denote the full posterior conditional density of τ^2 (on \mathbb{R}_+) given $\beta, \sigma^2, \mathbf{y}$. By (1.7),

$$f(\tau^2 | \beta, \sigma^2, \mathbf{y}) = \prod_{j=1}^p \sqrt{\frac{\lambda^2}{2\pi}} (\tau_j^2)^{-\frac{1}{2}} e^{-\frac{\beta_j^2 \left(1 - \tau_j^2 \sqrt{\frac{\lambda^2 \sigma^2}{\beta_j^2}}\right)^2}{2\sigma^2 \tau_j^2}}.$$

Note that the reciprocal of the entries of τ^2 (and not the entries themselves) have an Inverse-Gaussian distribution. Let $f(\cdot | \beta, \tau^2, \mathbf{y})$ denote the full posterior conditional density of σ^2 (on \mathbb{R}_+) given $\beta, \tau^2, \mathbf{y}$. By (1.8),

$$f(\sigma^2 | \beta, \tau^2, \mathbf{y}) = \frac{\left((\mathbf{y} - X\beta)^T (\mathbf{y} - X\beta) + \beta^T D_{\tau}^{-1} \beta + 2\xi \right)^{\frac{n+p+2\alpha}{2}}}{2^{\frac{n+p+2\alpha}{2}} \Gamma\left(\frac{n+p+2\alpha}{2}\right)} (\sigma^2)^{-\frac{n+p+2\alpha}{2} - 1} \times e^{-\frac{(\mathbf{y} - X\beta)^T (\mathbf{y} - X\beta) + \beta^T D_{\tau}^{-1} \beta + 2\xi}{2\sigma^2}}.$$

Let η denote the Lebesgue measure on $\mathbb{R}^p \times \mathbb{R}_+^p \times \mathbb{R}_+$. The Bayesian lasso Gibbs Markov chain has a Mtd (with respect to η) given by

$$k\left((\boldsymbol{\beta}, \boldsymbol{\tau}^2, \sigma^2) \mid (\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\tau}}^2, \tilde{\sigma}^2)\right) = f(\sigma^2 \mid \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\tau}}^2, \mathbf{y}) f(\boldsymbol{\tau}^2 \mid \tilde{\boldsymbol{\beta}}, \sigma^2, \mathbf{y}) f(\boldsymbol{\beta} \mid \boldsymbol{\tau}^2, \sigma^2, \mathbf{y}). \quad (2.1)$$

It is well known, and can be verified by a straightforward calculation, that the joint posterior density of $(\boldsymbol{\beta}, \boldsymbol{\tau}^2, \sigma^2)$ is invariant for the Gibbs transition density k defined above. The Mtd is strictly positive, which implies that the chain is aperiodic and η -irreducible [13, Page 87]. Moreover, the existence of an invariant probability density together with η -irreducibility imply that the chain is positive Harris recurrent (see for example [1]). Note also that η is equivalent to the maximal irreducibility measure. We now prove geometric ergodicity by establishing a geometric *drift condition* and an associated *minorization condition* for the Gibbs transition density k .

3. Proof of geometric ergodicity

3.1. Drift condition

Consider the function

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

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

Proposition 2. *If $n \geq 3$, there exist constants $0 \leq \gamma < 1$ and $b > 0$ such that*

$$E_k [V(\boldsymbol{\beta}, \boldsymbol{\tau}^2, \sigma^2) \mid \boldsymbol{\beta}_0, \boldsymbol{\tau}_0^2, \sigma_0^2] \leq \gamma V(\boldsymbol{\beta}_0, \boldsymbol{\tau}_0^2, \sigma_0^2) + b, \quad (3.1)$$

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

Proof. It follows by the definition of the Markov transition density k that

$$\begin{aligned} & E_k [V(\boldsymbol{\beta}, \boldsymbol{\tau}^2, \sigma^2) \mid \boldsymbol{\beta}_0, \boldsymbol{\tau}_0^2, \sigma_0^2] \\ &= E [E [E [V(\boldsymbol{\beta}, \boldsymbol{\tau}^2, \sigma^2) \mid \boldsymbol{\tau}^2, \sigma^2] \mid \boldsymbol{\beta}_0, \sigma^2] \mid \boldsymbol{\beta}_0, \boldsymbol{\tau}_0^2]. \end{aligned} \quad (3.2)$$

We evaluate the three conditional expectations one step at a time. We start with the innermost conditional expectation in (3.2).

$$\begin{aligned} & E [V(\boldsymbol{\beta}, \boldsymbol{\tau}^2, \sigma^2) \mid \boldsymbol{\tau}^2, \sigma^2] \\ &= E \left[(\mathbf{y} - X\boldsymbol{\beta})^T (\mathbf{y} - X\boldsymbol{\beta}) + \boldsymbol{\beta}^T D_{\boldsymbol{\tau}}^{-1} \boldsymbol{\beta} \mid \boldsymbol{\tau}^2, \sigma^2 \right] + \sum_{j=1}^p \tau_j^2 \end{aligned}$$

$$\begin{aligned}
 &= \mathbf{y}^T \mathbf{y} + E \left[\boldsymbol{\beta}^T (X^T X + D_{\boldsymbol{\tau}}^{-1}) \boldsymbol{\beta} \mid \boldsymbol{\tau}^2, \sigma^2 \right] - 2\mathbf{y}^T X E \left[\boldsymbol{\beta} \mid \boldsymbol{\tau}^2, \sigma^2 \right] + \sum_{j=1}^p \tau_j^2 \\
 &= \mathbf{y}^T \mathbf{y} + \left\| (X^T X + D_{\boldsymbol{\tau}}^{-1})^{\frac{1}{2}} (X^T X + D_{\boldsymbol{\tau}}^{-1})^{-1} X^T \mathbf{y} \right\|_2^2 \\
 &\quad + \text{tr} \left((X^T X + D_{\boldsymbol{\tau}}^{-1})^{\frac{1}{2}} (X^T X + D_{\boldsymbol{\tau}}^{-1})^{-1} (X^T X + D_{\boldsymbol{\tau}}^{-1})^{\frac{1}{2}} \right) \sigma^2 \\
 &\quad - 2\mathbf{y}^T X (X^T X + D_{\boldsymbol{\tau}}^{-1})^{-1} X^T \mathbf{y} + \sum_{j=1}^p \tau_j^2 \\
 &= \mathbf{y}^T \mathbf{y} - \mathbf{y}^T X (X^T X + D_{\boldsymbol{\tau}}^{-1})^{-1} X^T \mathbf{y} + p\sigma^2 + \sum_{j=1}^p \tau_j^2 \\
 &\leq \mathbf{y}^T \mathbf{y} + \sum_{j=1}^p \tau_j^2 + p\sigma^2. \tag{3.3}
 \end{aligned}$$

For the next step, we analyze terms related to the middle conditional expectation in (3.2). Note that if $Z \sim \text{Inverse-Gaussian}(\mu', \lambda')$, then

$$E \left[\frac{1}{Z} \right] = \frac{1}{\mu'} + \frac{1}{\lambda'}.$$

Hence, it follows from (1.7) that

$$\begin{aligned}
 E \left[\tau_j^2 \mid \boldsymbol{\beta}_0, \sigma^2 \right] &= \sqrt{\frac{\beta_{0j}^2}{\lambda^2 \sigma^2} + \frac{1}{\lambda^2}} \\
 &= \sqrt{\frac{n+p+2\alpha}{\lambda^2} \frac{\beta_{0j}^2}{(n+p+2\alpha)\sigma^2} + \frac{1}{\lambda^2}} \\
 &\leq \frac{n+p+2\alpha}{2\lambda^2} + \frac{\beta_{0j}^2}{2(n+p+2\alpha)\sigma^2} + \frac{1}{\lambda^2}. \tag{3.4}
 \end{aligned}$$

Finally, we analyze terms related to the outermost conditional expectation in (3.2). Note that

$$E \left[\sigma^2 \mid \boldsymbol{\beta}_0, \boldsymbol{\tau}_0^2 \right] = \frac{(\mathbf{y} - X\boldsymbol{\beta}_0)^T (\mathbf{y} - X\boldsymbol{\beta}_0) + \boldsymbol{\beta}_0^T D_{\boldsymbol{\tau}_0}^{-1} \boldsymbol{\beta}_0 + 2\xi}{n+p+2\alpha-2}, \tag{3.5}$$

and

$$E \left[\frac{1}{\sigma^2} \mid \boldsymbol{\beta}_0, \boldsymbol{\tau}_0^2 \right] = \frac{n+p+2\alpha}{(\mathbf{y} - X\boldsymbol{\beta}_0)^T (\mathbf{y} - X\boldsymbol{\beta}_0) + \boldsymbol{\beta}_0^T D_{\boldsymbol{\tau}_0}^{-1} \boldsymbol{\beta}_0 + 2\xi}. \tag{3.6}$$

Combining (3.2)–(3.6), we get that

$$\begin{aligned}
 &E_k \left[V(\boldsymbol{\beta}, \boldsymbol{\tau}^2, \sigma^2) \mid \boldsymbol{\beta}_0, \boldsymbol{\tau}_0^2, \sigma_0^2 \right] \\
 &\leq \mathbf{y}^T \mathbf{y} + \frac{p(n+p+2\alpha)}{2\lambda^2} + \frac{p}{\lambda^2}
 \end{aligned}$$

$$\begin{aligned}
& + \frac{p \left((\mathbf{y} - X\boldsymbol{\beta}_0)^T (\mathbf{y} - X\boldsymbol{\beta}_0) + \boldsymbol{\beta}_0^T D_{\tau_0}^{-1} \boldsymbol{\beta}_0 + 2\xi \right)}{n + p + 2\alpha - 2} \\
& + \frac{(n + p + 2\alpha) \sum_{j=1}^p \beta_{0j}^2}{2(n + p + 2\alpha) \left((\mathbf{y} - X\boldsymbol{\beta}_0)^T (\mathbf{y} - X\boldsymbol{\beta}_0) + \boldsymbol{\beta}_0^T D_{\tau_0}^{-1} \boldsymbol{\beta}_0 + 2\xi \right)} \\
\leq & \mathbf{y}^T \mathbf{y} + \frac{p(n + 2p + 2\alpha)}{2\lambda^2} + \frac{p}{\lambda^2} + \frac{2p\xi}{n + p + 2\alpha - 2} \\
& + \frac{p}{n + p + 2\alpha - 2} \left((\mathbf{y} - X\boldsymbol{\beta}_0)^T (\mathbf{y} - X\boldsymbol{\beta}_0) + \boldsymbol{\beta}_0^T D_{\tau_0}^{-1} \boldsymbol{\beta}_0 \right) \\
& + \frac{\sum_{j=1}^p \beta_{0j}^2}{2 \left((\mathbf{y} - X\boldsymbol{\beta}_0)^T (\mathbf{y} - X\boldsymbol{\beta}_0) + \boldsymbol{\beta}_0^T D_{\tau_0}^{-1} \boldsymbol{\beta}_0 + 2\xi \right)} \\
\leq & \mathbf{y}^T \mathbf{y} + \frac{p(n + 2p + 2\alpha)}{2\lambda^2} + \frac{p}{\lambda^2} + \frac{2p\xi}{n + p + 2\alpha - 2} \\
& + \frac{p}{n + p + 2\alpha - 2} \left((\mathbf{y} - X\boldsymbol{\beta}_0)^T (\mathbf{y} - X\boldsymbol{\beta}_0) + \boldsymbol{\beta}_0^T D_{\tau_0}^{-1} \boldsymbol{\beta}_0 \right) + \frac{1}{2} \frac{\sum_{j=1}^p \beta_{0j}^2}{\boldsymbol{\beta}_0^T D_{\tau_0}^{-1} \boldsymbol{\beta}_0} \\
\leq & \mathbf{y}^T \mathbf{y} + \frac{p(n + 2p + 2\alpha)}{2\lambda^2} + \frac{p}{\lambda^2} + \frac{2p\xi}{n + p + 2\alpha - 2} \\
& + \frac{p}{n + p + 2\alpha - 2} \left((\mathbf{y} - X\boldsymbol{\beta}_0)^T (\mathbf{y} - X\boldsymbol{\beta}_0) + \boldsymbol{\beta}_0^T D_{\tau_0}^{-1} \boldsymbol{\beta}_0 \right) + \frac{1}{2} \sum_{j=1}^p \tau_{0j}^2. \quad (3.7)
\end{aligned}$$

The last inequality follows from the fact that

$$\begin{aligned}
\sum_{j=1}^p \beta_{0j}^2 & \leq \left(\sum_{j=1}^p |\beta_{0j}| \right)^2 \\
& = \left(\sum_{j=1}^p \sqrt{\frac{\beta_{0j}^2}{\tau_{0j}^2}} \sqrt{\tau_{0j}^2} \right)^2 \\
& \leq \left(\sum_{j=1}^p \frac{\beta_{0j}^2}{\tau_{0j}^2} \right) \left(\sum_{j=1}^p \tau_{0j}^2 \right).
\end{aligned}$$

It follows from (3.7) that

$$E_k [V(\boldsymbol{\beta}, \boldsymbol{\tau}^2, \sigma^2) \mid \boldsymbol{\beta}_0, \boldsymbol{\tau}_0^2, \sigma_0^2] \leq \gamma V(\boldsymbol{\beta}_0, \boldsymbol{\tau}_0^2, \sigma_0^2) + b, \quad (3.8)$$

where

$$\gamma = \max \left(\frac{p}{n + p + 2\alpha - 2}, \frac{1}{2} \right), \quad (3.9)$$

and

$$b = \mathbf{y}^T \mathbf{y} + \frac{p(n + 2p + 2\alpha)}{2\lambda^2} + \frac{p}{\lambda^2} + \frac{2p\xi}{n + p + 2\alpha - 2}. \quad (3.10)$$

Hence, the required geometric drift condition is established. \square

3.2. *Minorization condition*

Let us recall that $V(\beta, \tau^2, \sigma^2) = (\mathbf{y} - X\beta)^T(\mathbf{y} - X\beta) + \beta^T D_{\tau}^{-1}\beta + \sum_{j=1}^p \tau_j^2$. For every $d > 0$, let $B_{V,d} = \{(\beta, \tau^2, \sigma^2) : V(\beta, \tau^2, \sigma^2) \leq d\}$. The following proposition establishes an associated *minorization condition* to the geometric drift condition established in Proposition 2.

Proposition 3. *There exists a constant $0 < \epsilon = \epsilon(V, d) \leq 1$ and a probability density function \tilde{f} on $\mathbb{R}^p \times \mathbb{R}_+^p \times \mathbb{R}_+$ such that*

$$k((\beta, \tau^2, \sigma^2) | (\beta_0, \tau_0^2, \sigma_0^2)) \geq \epsilon \tilde{f}(\beta, \tau^2, \sigma^2), \tag{3.11}$$

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

Proof. Note that

$$\begin{aligned} f(\tau_j^2 | \beta_0, \sigma^2, \mathbf{y}) &= \sqrt{\frac{\lambda'}{2\pi}} (\tau_j^2)^{-\frac{1}{2}} e^{-\frac{\lambda'(1-\tau_j^2\mu'_j)^2}{2(\mu'_j)^2\tau_j^2}} \\ &= \sqrt{\frac{\lambda'}{2\pi}} (\tau_j^2)^{-\frac{1}{2}} e^{-\frac{\lambda'\tau_j^2}{2} - \frac{\lambda'}{2(\mu'_j)^2\tau_j^2} + \frac{\lambda'}{\mu'_j}} \\ &\geq \sqrt{\frac{\lambda'}{2\pi}} (\tau_j^2)^{-\frac{1}{2}} e^{-\frac{\lambda'\tau_j^2}{2} - \frac{\lambda'}{2(\mu'_j)^2\tau_j^2}}, \end{aligned}$$

where $\lambda' = \lambda^2$ and $\mu'_j = \sqrt{\frac{\lambda^2\sigma^2}{\beta_{0j}^2}}$. If $(\beta_0, \tau_0^2, \sigma_0^2) \in B_{V,d}$, then

$$\begin{aligned} &\beta_0^T D_{\tau_0}^{-1}\beta_0 + \sum_{j=1}^p \tau_{0j}^2 \leq d \\ \Rightarrow &\left(\beta_0^T D_{\tau_0}^{-1}\beta_0\right) \left(\sum_{j=1}^p \tau_{0j}^2\right) \leq d^2 \\ \Rightarrow &\sum_{j=1}^p \beta_{0j}^2 \leq d^2. \end{aligned}$$

Hence, if $(\beta_0, \tau_0^2, \sigma_0^2) \in B_{V,d}$, then

$$\frac{\lambda'}{(\mu'_j)^2} = \frac{\beta_{0j}^2}{\sigma^2} \leq \frac{d^2}{\sigma^2},$$

and,

$$\begin{aligned} &f(\tau_j^2 | \beta_0, \sigma^2, \mathbf{y}) \\ \geq &\sqrt{\frac{\lambda^2}{2\pi}} (\tau_j^2)^{-\frac{1}{2}} e^{-\frac{\lambda^2\tau_j^2}{2} - \frac{d^2}{2\sigma^2\tau_j^2}} \end{aligned}$$

$$\begin{aligned} &\geq \sqrt{\frac{\lambda^2}{2\pi}}(\tau_j^2)^{-\frac{1}{2}}e^{-\frac{\lambda^2\tau_j^2}{2}-\frac{d^2}{2\sigma^2\tau_j^2}+\sqrt{\frac{\lambda^2d^2}{\sigma^2}}}e^{-\sqrt{\frac{\lambda^2d^2}{\sigma^2}}} \\ &\geq g_1(\tau_j^2 | \sigma^2)e^{-\sqrt{\frac{\lambda^2d^2}{\sigma^2}}}, \end{aligned} \tag{3.12}$$

where $g_1(\tau_j^2 | \sigma^2)$ is the density of the reciprocal of an Inverse-Gaussian random variable with parameters $\sqrt{\frac{\lambda^2\sigma^2}{d^2}}$ and λ^2 .

Let us now consider the full conditional density of σ^2 given $\beta_0, \tau_0^2, \mathbf{y}$. Note that for $(\beta_0, \tau_0^2, \sigma_0^2) \in B_{V,d}$,

$$\frac{(\mathbf{y} - X\beta_0)^T(\mathbf{y} - X\beta_0) + \beta_0^T D_{\tau_0}^{-1}\beta_0 + 2\xi}{2} \leq \frac{d + 2\xi}{2}, \tag{3.13}$$

and

$$\begin{aligned} &(\mathbf{y} - X\beta_0)^T(\mathbf{y} - X\beta_0) + \beta_0^T D_{\tau_0}^{-1}\beta_0 + 2\xi \\ &= \mathbf{y}^T \mathbf{y} - 2\mathbf{y}^T X\beta_0 + \beta_0^T (X^T X + D_{\tau_0}^{-1})\beta_0 + 2\xi \\ &= \mathbf{y}^T \mathbf{y} - \mathbf{y}^T X (X^T X + D_{\tau_0}^{-1})^{-1} X^T \mathbf{y} + 2\xi \\ &\quad + (X^T \mathbf{y} - (X^T X + D_{\tau_0}^{-1})\beta_0)^T (X^T X + D_{\tau_0}^{-1})^{-1} (X^T \mathbf{y} - (X^T X + D_{\tau_0}^{-1})\beta_0) \\ &\geq \mathbf{y}^T \mathbf{y} - \mathbf{y}^T X (X^T X + D_{\tau_0}^{-1})^{-1} X^T \mathbf{y} + 2\xi \\ &\geq \mathbf{y}^T \mathbf{y} - \mathbf{y}^T X \left(X^T X + \frac{1}{d}I_p \right)^{-1} X^T \mathbf{y} + 2\xi. \end{aligned} \tag{3.14}$$

The last inequality follows since $\tau_{0j}^2 \leq d$ for every $1 \leq j \leq p$. Note that $I_n - X(X^T X + \frac{1}{d}I_p)^{-1}X^T$ is a positive definite matrix. Hence

$$\mathbf{y}^T \mathbf{y} - \mathbf{y}^T X \left(X^T X + \frac{1}{d}I_p \right)^{-1} X^T \mathbf{y} > 0.$$

It follows from (3.13) and (3.14) that for $(\beta_0, \tau_0^2, \sigma_0^2) \in B_{V,d}$,

$$\begin{aligned} &e^{-p\sqrt{\frac{\lambda^2d^2}{\sigma^2}}}f(\sigma^2 | \beta_0, \tau_0^2, \mathbf{y}) \\ &\geq \left(\frac{\mathbf{y}^T \mathbf{y} - \mathbf{y}^T X (X^T X + \frac{1}{d}I_p)^{-1} X^T \mathbf{y} + 2\xi}{2} \right)^{\frac{n+p+2\alpha}{2}} \\ &\quad \times \frac{(\sigma^2)^{-\frac{n+p+2\alpha+2}{2}}e^{-\frac{d+2\xi}{2\sigma^2}}e^{-p\sqrt{\frac{\lambda^2d^2}{\sigma^2}}}}{\Gamma\left(\frac{n+p+2\alpha}{2}\right)} \\ &\geq \left(\frac{\mathbf{y}^T \mathbf{y} - \mathbf{y}^T X (X^T X + \frac{1}{d}I_p)^{-1} X^T \mathbf{y} + 2\xi}{2} \right)^{\frac{n+p+2\alpha}{2}} \\ &\quad \times \frac{(\sigma^2)^{-\frac{n+p+2\alpha+2}{2}}e^{-\frac{d+2\xi}{2\sigma^2}}e^{-\frac{1}{2}-\frac{p^2\lambda^2d^2}{2\sigma^2}}}{\Gamma\left(\frac{n+p+2\alpha}{2}\right)} \end{aligned}$$

$$= e^{-\frac{1}{2}} \left(\frac{\mathbf{y}^T \mathbf{y} - \mathbf{y}^T X (X^T X + \frac{1}{d} I_p)^{-1} X^T \mathbf{y} + 2\xi}{d + 2\xi + p^2 \lambda^2 d^2} \right)^{\frac{n+p+2\alpha}{2}} g_2(\sigma^2), \tag{3.15}$$

where $g_2(\sigma^2)$ is the density of the Inverse-Gamma distribution with parameters $\frac{n+p+2\alpha}{2}$ and $\frac{d+2\xi+\lambda^2 d^2}{2}$. It follows from (2.1), (3.12) and (3.15) that, for all $(\beta_0, \tau_0^2, \sigma_0^2) \in B_{V,d}$,

$$\begin{aligned} & k((\beta, \tau^2, \sigma^2) \mid (\beta_0, \tau_0^2, \sigma_0^2)) \\ &= f(\beta \mid \tau^2, \sigma^2, \mathbf{y}) f(\tau^2 \mid \beta_0, \sigma^2, \mathbf{y}) f(\sigma^2 \mid \beta_0, \tau_0^2, \mathbf{y}) \\ &\geq \epsilon \tilde{f}(\beta, \tau^2, \sigma^2), \end{aligned}$$

where

$$\epsilon = e^{-\frac{1}{2}} \left(\frac{\mathbf{y}^T \mathbf{y} - \mathbf{y}^T X (X^T X + \frac{1}{d} I_p)^{-1} X^T \mathbf{y} + 2\xi}{d + 2\xi + p^2 \lambda^2 d^2} \right)^{\frac{n+p+2\alpha}{2}}, \tag{3.16}$$

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

$$\tilde{f}(\beta, \tau^2, \sigma^2) = f(\beta \mid \tau^2, \sigma^2, \mathbf{y}) \left(\prod_{j=1}^p g_1(\tau_j^2 \mid \sigma^2) \right) g_2(\sigma^2).$$

Hence, the required minorization condition for k has been established. □

The drift and minorization conditions in Proposition 2 and Proposition 3 can be combined with Theorem 12 of Rosenthal [18] to establish geometric ergodicity of the Bayesian lasso Gibbs Markov chain.

Proposition 4. *Let $n \geq 3$, and γ, b and ϵ be as defined in (3.9), (3.10), and (3.16) respectively. Let $d > \frac{2b}{1-\gamma}$. Let*

$$A = \frac{1+d}{1+2b+\gamma d} \text{ and } U = 1+2(\gamma d+b).$$

Then for any $(\beta_0, \tau_0^2, \sigma_0^2) \in \mathbb{R}^p \times \mathbb{R}_+^p \times \mathbb{R}_+$, $m \in \mathbb{N}$, and $0 < r < 1$

$$\begin{aligned} & \|K^m((\beta_0, \tau_0^2, \sigma_0^2), \cdot) - \Pi(\cdot)\|_{TV} \\ &\leq (1-\epsilon)^{rm} + \left(\frac{Ur}{A^{1-r}}\right)^m \left(1 + \frac{b}{1-\gamma} + V(\beta_0, \tau_0^2, \sigma_0^2)\right). \end{aligned}$$

Proposition 1 is an immediate corollary of the above result.

4. Discussion

The Bayesian lasso is a popular and widely used algorithm for sparse Bayesian estimation in linear regression. In this paper, we have established geometric ergodicity of the Markov chain corresponding to the Bayesian lasso algorithm.

As discussed previously, this proves the existence of the Markov chain CLT in this context, thereby allowing the user to obtain valid asymptotic standard errors for Markov chain based estimates of posterior quantities.

The Laplace prior for β used in the Bayesian lasso algorithm is an example of the so-called “continuous shrinkage priors” for sparse Bayesian estimation. In recent years, many other continuous shrinkage priors have been introduced and studied in the literature (see [4, 5, 6, 10, 15] and the references therein), and sampling from the resulting posterior distribution is often achieved by using MCMC. To the best of our knowledge, geometric ergodicity of the corresponding Markov chains has not been investigated. Using the insights obtained from the analysis in this paper, we are currently investigating geometric ergodicity for Markov chains corresponding to some of the other continuous shrinkage priors proposed in the literature.

5. Appendix

We point out the errors in the proof of geometric ergodicity of the Bayesian lasso Gibbs Markov chain in Kyung et al. [12] (referred to henceforth as KGGC). The authors in KGGC only consider the case $n \geq p$. The authors assert that the Bayesian lasso model can be obtained by appropriately marginalizing a hierarchical random effects model. They then prove the geometric ergodicity of the Markov chain corresponding to this hierarchical random effects model, and argue that this implies geometric ergodicity of the Bayesian lasso model. We start by reproducing (verbatim) the hierarchical random effects model from KGGC (Page 406, appendix) below.

$$\mathbf{y}_{p \times 1} \sim \mathcal{N}(\boldsymbol{\theta}, \lambda_e^{-1} I) \quad (5.1)$$

$$\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu} \mathbf{1}, \Sigma(\boldsymbol{\tau})) \quad (5.2)$$

$$\boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\mu}_0, \lambda_0^{-1}) \quad (5.3)$$

$$\lambda_e \sim \text{Gamma}(a_2, b_2) \quad (5.4)$$

$$\frac{1}{\tau_j^2} \stackrel{i.i.d.}{\sim} \text{Gamma}(a_1, b_2^*), \quad (5.5)$$

where $\Sigma(\boldsymbol{\tau})$ is of the form

$$\Sigma(\boldsymbol{\tau}) = \text{diag}(\tau_1^2, \tau_2^2, \dots, \tau_p^2) \sim \prod_{i=1}^p \frac{\lambda}{2} e^{-\frac{\lambda \tau_i^2}{2}} d\tau_i^2. \quad (5.6)$$

We refer to this hierarchical random effects model as the ‘RE model’.

- There is a contradiction between (5.5) and (5.6) regarding the prior distribution of τ_j^2 . In (5.5), τ_j^2 is assumed to be Inverse-Gamma, whereas in (5.6), τ_j^2 is assumed to be Exponential. The authors proceed to prove geometric ergodicity of the RE model under the Inverse-Gamma assumption in (5.5). Since the Bayesian lasso model assumes the prior for τ_j^2 to be

Exponential, it clearly cannot be obtained by marginalizing the RE model with the Inverse-Gamma assumption in (5.5). Hence, the proof of geometric ergodicity for the RE model with the Inverse-Gamma assumption has no bearing on the geometric ergodicity for the Bayesian lasso model.

- \mathbf{y} is assumed to be p -dimensional in (5.1). In the Bayesian lasso model, the data vector \mathbf{y} is n -dimensional. Since the authors are considering the case $n > p$, this clearly rules out the possibility of marginalizing the RE model to get the Bayesian lasso model. If this is a typo, and y is n -dimensional, then $\boldsymbol{\theta}$ is n -dimensional. Hence, $\Sigma(\boldsymbol{\tau})$ is an $n \times n$ matrix. However, $\Sigma(\boldsymbol{\tau})$ is assumed to be a $p \times p$ matrix in (5.6).
- There is another error in the marginalizing argument provided in KGGC. On Page 381 (Section 4) the authors consider the model

$$\mathbf{y} \mid \boldsymbol{\theta}, \sigma^2 \sim \mathcal{N}_n(\boldsymbol{\theta}, \sigma^2 I_n) \text{ and } \boldsymbol{\theta} \mid \sigma^2 \sim \mathcal{N}_n(\mu \mathbf{1}_n, \Sigma_{\boldsymbol{\theta}}), \tag{5.7}$$

where

$$\Sigma_{\boldsymbol{\theta}}^{-1} = \lambda_0 J + \frac{\lambda}{\sigma^2} X(X^T X)^{-1} D_{\boldsymbol{\tau}}^{-1} (X^T X)^{-1} X^T + \hat{X} \hat{X}^T,$$

and $\boldsymbol{\theta} = \mu \mathbf{1}_n + X\boldsymbol{\beta} + \hat{X}\boldsymbol{\eta}$ (with $\mathbf{1}_n, X, \hat{X}$ mutually orthogonal). Here $\mathbf{1}_n$ is the n -dimensional vector of all ones, and J is the $n \times n$ matrix of all ones. They claim that marginalizing over $\boldsymbol{\eta}$ in (5.7) gives the marginal model

$$\mathbf{y} \mid \boldsymbol{\beta} \sim \mathcal{N}_n(\mu \mathbf{1}_n + X\boldsymbol{\beta}, \sigma^2 I_n) \text{ and } \boldsymbol{\beta} \sim \mathcal{N}_p\left(\mathbf{0}_p, \frac{\sigma^2}{\lambda} D_{\boldsymbol{\tau}}\right). \tag{5.8}$$

Firstly, there is a small issue in terms of μ . Note that, by the mutual orthogonality of $\mathbf{1}_n, X, \hat{X}$,

$$\frac{\mathbf{1}_n^T \boldsymbol{\theta}}{n} = \frac{\mathbf{1}_n^T (\mu \mathbf{1}_n + X\boldsymbol{\beta} + \hat{X}\boldsymbol{\eta})}{n} = \mu.$$

Since $\boldsymbol{\theta} \sim \mathcal{N}_n(\mu \mathbf{1}_n, \Sigma_{\boldsymbol{\theta}})$, this implies

$$\mu \sim \mathcal{N}\left(\mu, \frac{\mathbf{1}_n^T \Sigma_{\boldsymbol{\theta}} \mathbf{1}_n}{n^2}\right),$$

which is paradoxical, as it implies that μ is a random variable with non-zero variance ($\Sigma_{\boldsymbol{\theta}}$ is a full rank matrix), and also implies that μ is a constant. Let us get rid of this discrepancy by assuming $\mu = 0$, $\Sigma_{\boldsymbol{\theta}}^{-1} = \frac{\lambda}{\sigma^2} X(X^T X)^{-1} D_{\boldsymbol{\tau}}^{-1} (X^T X)^{-1} X^T + \hat{X} \hat{X}^T$, and choosing the transformation

$$\boldsymbol{\theta} \rightarrow X\boldsymbol{\beta} + \hat{X}\boldsymbol{\eta},$$

where $\boldsymbol{\eta}$ is now $(n-p)$ -dimensional, and $X^T \hat{X} = 0$. Let us assume without loss of generality that $\hat{X}^T \hat{X} = I_{n-p}$ (the choice of \hat{X} is completely in our hands anyway). Still, the marginal model after integrating out $\boldsymbol{\eta}$ is not the same as (5.8). An immediate way to check this is that $V(\mathbf{y}) = \sigma^2 I_n + \Sigma_{\boldsymbol{\theta}}$

under (5.7), while $V(\mathbf{y}) = \sigma^2 I_n + \frac{\sigma^2}{\lambda} X D_\tau X^T$ under (5.8). We now derive the correct marginal model. Note that the joint density of $(\mathbf{y}, \boldsymbol{\theta})$ is given by

$$f(\mathbf{y}, \boldsymbol{\theta}) = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{(\mathbf{y}-\boldsymbol{\theta})^T(\mathbf{y}-\boldsymbol{\theta})}{2\sigma^2}} \frac{|\Sigma_{\boldsymbol{\theta}}|^{-\frac{1}{2}}}{(\sqrt{2\pi})^n} e^{-\frac{\boldsymbol{\theta}^T \Sigma_{\boldsymbol{\theta}}^{-1} \boldsymbol{\theta}}{2}}.$$

By making the linear transformation $(\mathbf{y}, \boldsymbol{\theta}) \rightarrow (\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\eta})$, we get that

$$\begin{aligned} & f(\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\eta}) \\ &= \text{const} \times |\Sigma_{\boldsymbol{\theta}}|^{-\frac{1}{2}} e^{-\frac{(\mathbf{y}-X\boldsymbol{\beta}-\hat{X}\boldsymbol{\eta})^T(\mathbf{y}-X\boldsymbol{\beta}-\hat{X}\boldsymbol{\eta})}{2}} \\ & \quad \times e^{-\frac{(X\boldsymbol{\beta}+\hat{X}\boldsymbol{\eta})^T \left(\frac{\lambda}{\sigma^2} X(X^T X)^{-1} D_\tau^{-1} (X^T X)^{-1} X^T + \hat{X}\hat{X}^T \right) (X\boldsymbol{\beta}+\hat{X}\boldsymbol{\eta})}{2}} \\ &= \text{const} \times e^{-\frac{(1+\sigma^2)\boldsymbol{\eta}^T \boldsymbol{\eta} + \boldsymbol{\eta}^T \hat{X}^T \mathbf{y}}{2\sigma^2}} e^{-\frac{(\mathbf{y}-X\boldsymbol{\beta})^T (\mathbf{y}-X\boldsymbol{\beta}) + \lambda \boldsymbol{\beta}^T D_\tau^{-1} \boldsymbol{\beta}}{2\sigma^2}}. \end{aligned}$$

After completing the square, and integrating out $\boldsymbol{\eta}$, the marginal density of $(\mathbf{y}, \boldsymbol{\beta})$ is obtained as

$$\begin{aligned} & f(\mathbf{y}, \boldsymbol{\beta}) \\ &= \text{const} \times e^{\frac{\mathbf{y}^T \hat{X} \hat{X}^T \mathbf{y}}{2\sigma^2(\sigma^2+1)}} e^{-\frac{(\mathbf{y}-X\boldsymbol{\beta})^T (\mathbf{y}-X\boldsymbol{\beta}) + \lambda \boldsymbol{\beta}^T D_\tau^{-1} \boldsymbol{\beta}}{2\sigma^2}}. \end{aligned}$$

By completing the squares again, and using $\hat{X}^T X = 0$, it follows that the marginal model is given by

$$\mathbf{y} \mid \boldsymbol{\beta}, \sigma^2 \sim \mathcal{N}_n(X\boldsymbol{\beta}, \sigma^2 I_n + \hat{X}\hat{X}^T) \text{ and } \boldsymbol{\beta} \mid \sigma^2 \sim \mathcal{N}_p\left(\mathbf{0}_p, \frac{\sigma^2}{\lambda} D_\tau\right).$$

This is different from the Bayesian lasso model, as the conditional variance of \mathbf{y} given $\boldsymbol{\beta}, \sigma^2$ is $\sigma^2 I_n + \hat{X}\hat{X}^T$ (as opposed to $\sigma^2 I_n$ in the Bayesian lasso model).

Acknowledgment

The first author was supported by NSF Grant DMS-11-06084, and the second by NSF Grant DMS-11-06395.

References

- [1] ASMUSSEN, S. and GLYNN, P.W. (2011). A new proof of convergence of MCMC via the ergodic theorem, *Statistics & Probability Letters* **81**, 1482–1485. [MR2818658](#)
- [2] BEDNORZ, W. and LATUSZYNSKI, K. (2007). A few remarks on “Fixed-Width Output Analysis for Markov Chain Monte Carlo” by Jones et al., *JASA* **102**, 1485–1486. [MR2412582](#)
- [3] BAE, K. and MALLICK, B.K. (2004). Gene selection using a two-level hierarchical Bayesian model, *Bioinformatics* **20**, 3423–3430.

- [4] BHATTACHARYA, A., PATI, D., PILLAI, N.S., and DUNSON, D.B. (2012). Bayesian shrinkage, *arXiv preprint*.
- [5] CARVALHO, C.M., POLSON, N.G., and SCOTT, J.G. (2009). Handling sparsity via the horseshoe, *Journal of Machine Learning Research W & CP* **5**, 73–80.
- [6] CARVALHO, C.M., POLSON, N.G., and SCOTT, J.G. (2010). The horseshoe estimator for sparse signals, *Biometrika* **97**, 465–480. [MR2650751](#)
- [7] FIGUEIREDO, M.A.T. (2003). Adaptive sparseness for supervised learning, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25**, 1150–1159.
- [8] FLEGAL, J.M., HARAN, M., and JONES, G.M. (2008). Markov chain Monte Carlo: Can we trust the third significant figure? *Statistical Science* **23**, 250–260. [MR2516823](#)
- [9] FLEGAL, J.M. and JONES, G.L. (2010). Batch means and spectral variance estimators in Markov chain Monte Carlo, *Annals of Statistics* **38**, 1034–1070. [MR2604704](#)
- [10] GRIFFIN, J.E. and BROWN, P.J. (2010). Inference with normal-gamma prior distributions in regression problems, *Bayesian Analysis* **5**, 171–188. [MR2596440](#)
- [11] JONES, G.L., HARAN, M., CAFFO, B.S., and NEATH, R. (2006). Fixed-width output analysis for Markov chain Monte Carlo, *Journal of the American Statistical Association* **101**, 1537–1547. [MR2279478](#)
- [12] KYUNG, M., GILL, J., GHOSH, M., and CASELLA, G. (2010). Penalized regression, standard errors and Bayesian lassos, *Bayesian Analysis* **5**, 369–412. [MR2719657](#)
- [13] MEYN, S.P. and TWEEDIE, R.L. (1993). *Markov Chains and Stochastic Stability*, Springer-Verlag, London. [MR1287609](#)
- [14] PARK, T. and CASELLA, G. (2008). The Bayesian lasso, *Journal of the American Statistical Association* **103**, 681–686. [MR2524001](#)
- [15] POLSON, N.G. and SCOTT, J.G. (2010). Shrink globally, act locally: Sparse Bayesian regularization and prediction, *Bayesian Statistics* **9**, J.M. Bernardo, M.J. Bayarri, J.O. Berger, A.P. Dawid, D. Heckerman, A.F.M. Smith, and M. West, eds., Oxford University Press, New York.
- [16] ROBERTS, G.O. and ROSENTHAL, J.S. (1998). Markov chain Monte Carlo: Some practical implications of theoretical results (with discussion), *Canadian Journal of Statistics* **26**, 5–31. [MR1624414](#)
- [17] ROBERTS, G.O. and ROSENTHAL, J.S. (2004). General state space Markov chains and MCMC algorithms, *Probability Surveys*, 20–71. [MR2095565](#)
- [18] ROSENTHAL, J.S. (1995). Minorization conditions and convergence rates for Markov chain Monte Carlo, *JASA* **90**, 558–566. [MR1340509](#)
- [19] TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society B* **58**, 267–288. [MR1379242](#)
- [20] YUAN, M., and LIN, Y. (2005). Efficient empirical Bayes variable selection and estimation in linear models, *Journal of the American Statistical Association* **100**, 1215–1225. [MR2236436](#)