Geometric Ergodicity of Gibbs Samplers for Bayesian General Linear Mixed Models with Proper Priors

Jorge Carlos Román
Department of Mathematics
Vanderbilt University

James P. Hobert
Department of Statistics
University of Florida

July 2013

Abstract

When a Bayesian version of the general linear mixed model is created by adopting a conditionally conjugate prior distribution, a simple block Gibbs sampler can be employed to explore the resulting intractable posterior density. In this article it is shown that, under mild conditions that nearly always hold in practice, the block Gibbs Markov chain is geometrically ergodic.

1 Introduction

The general linear mixed model is one of the most frequently applied statistical models. Bayesian versions of this model require a prior for the parameters but, unfortunately, any non-trivial prior leads to an intractable posterior density. However, if a (conditionally) conjugate prior is adopted, then a simple (block) Gibbs sampler can be used to explore the resulting posterior density. In this article, we study the convergence properties of the Markov chains that underlie this Gibbs sampler.

The first stage of the Bayesian hierarchical model is

$$Y|\beta, u, \lambda \sim \mathrm{N}_N\Big(X\beta + \sum_{i=1}^{r} Z_i u_i, \lambda_e^{-1} I\Big) \,, \tag{1}$$

AMS 2010 *subject classifications.* Primary 65C05; secondary 62F15

Key words and phrases. Conditionally conjugate prior; Convergence rate; Geometric drift condition; Markov chain; Matrix inequality; Monte Carlo

where Y is an $N \times 1$ response vector, X and Z_i are known matrices of dimensions $N \times p$ and $N \times q_i$, respectively, β is a $p \times 1$ regression coefficient, u_i is a $q_i \times 1$ vector that represents the ith random factor in the model, $u := \left(u_1^T \; u_2^T \cdots u_r^T\right)^T$, λ_e is the precision parameter associated with Y, λ_{u_i} is the precision parameter associated with u_i, and $\lambda := \left(\lambda_e \; \lambda_{u_1} \; \cdots \; \lambda_{u_r}\right)^T$. Given λ, the random elements β and u are assumed to be mutually independent and the second stage specifies their distribution:

$$\beta|\lambda \sim \mathrm{N}_p(\mu_\beta, \Sigma_\beta) \quad \text{and} \quad u|\lambda \sim \mathrm{N}_q\left(0, \Lambda_u^{-1}\right) , \tag{2}$$

where $\Lambda_u = \oplus_{i=1}^r \lambda_{u_i} I_{q_i}$ and $q = q_1 + \cdots + q_r$. Finally, the third stage of the model specifies the distribution of the precision parameters, which are independent with marginals given by

$$\lambda_e \sim \text{Gamma}(a_e, b_e) \quad \text{and} \quad \lambda_{u_i} \sim \text{Gamma}(a_i, b_i) \text{ , for } i = 1, \ldots, r \; . \tag{3}$$

The hyper-parameters μ_β, Σ_β, $a = (a_e \; a_1 \; \cdots \; a_r)^T$ and $b = (b_e \; b_1 \; \cdots \; b_r)^T$ are all assumed to be known and are restricted to their usual ranges to ensure a proper prior.

We now assemble the (proper) posterior density associated with our Bayesian model. Define $\theta = (\beta^T \; u^T)^T$, $Z = (Z_1 \, Z_2 \cdots Z_r)$, and $W = (X \; Z)$, so that $W\theta = X\beta + Zu = X\beta + \sum_{i=1}^r Z_i u_i$. Also, let y denote the observed response vector and let $f_{\mathrm{N}}(\cdot\,; \mu, \Sigma)$ and $f_{\mathrm{G}}(\cdot\,; c, d)$ denote the probability densities of $N(\mu, \Sigma)$ and Gamma(c, d) random variables, respectively. In Gamma(c, d), d is the rate parameter so

$$f_{\mathrm{G}}(g; c, d) = \frac{d^c}{\Gamma(c)} g^{c-1} e^{-dg} I_{\mathbb{R}_+}(g) \; ,$$

where $c, d > 0$ and $\mathbb{R}_+ := (0, \infty)$. The posterior density is

$$\pi(\lambda, \theta|y) = \frac{\pi^*(\lambda, \theta|y)}{m(y)} \; ,$$

where π^* is an unnormalized density given by

$$\pi^*(\lambda, \theta|y) = f_{\mathrm{N}}(y; W\theta, \lambda_e^{-1} I) \, f_{\mathrm{N}}(\beta; \mu_\beta, \Sigma_\beta) \, f_{\mathrm{N}}\left(u; 0, \Lambda_u^{-1}\right) \, f_{\mathrm{G}}(\lambda_e; a_e, b_e) \left[\prod_{i=1}^r f_{\mathrm{G}}(\lambda_{u_i}; a_i, b_i)\right] \; ,$$

and m is the marginal density of the data, which serves as a normalizing constant and is given by

$$m(y) = \int_{\mathbb{R}_+^{r+1}} \int_{\mathbb{R}^{p+q}} \pi^*(\lambda, \theta|y) \, d\theta d\lambda < \infty \; .$$

It should be noted that π, π^*, and m all depend on X, $Z_1, \ldots, Z_r$, and the hyper-parameters but, as usual, this dependence will be suppressed in the notation.

Bayesian inference requires the computation of expectations with respect to the posterior density. For a π-integrable function of interest g, let $E_\pi g$ denote its posterior expectation; that is,

$$E_\pi g = \int_{\mathbb{R}_+^{r+1}} \int_{\mathbb{R}^{p+q}} g(\lambda, \theta) \pi(\lambda, \theta | y) \, d\theta \, d\lambda \; .$$

A straightforward manipulation of the posterior density shows that $E_\pi g$ can be expressed as a ratio of two integrals with dimensions $p+q+r+1$ and $r+1$. However, for virtually any function g, these integrals (which are often high-dimensional) are intractable. Thus, in general, exact calculation of posterior expectations is impossible.

In this article, we study a Markov chain Monte Carlo (MCMC) algorithm that can be used to effectively approximate the intractable quantity $E_\pi g$. In particular, we analyze a block Gibbs sampler that simulates a Markov chain $\{(\lambda_n, \theta_n)\}_{n=0}^\infty$, that lives on $\mathsf{X} = \mathbb{R}_+^{r+1} \times \mathbb{R}^{p+q}$, and has the posterior density (of λ and θ) as its invariant density. This Gibbs sampler is referred to as a *block* Gibbs sampler because it *blocks* $\lambda_e, \lambda_{u_1}, \ldots, \lambda_{u_r}$ into a single vector λ and blocks β and u into the vector θ; that is, all the variables within a block are updated simultaneously from the joint conditional distribution given the variables in the other block. Specifically, if the current state of the chain is (λ_n, θ_n), then the next state, $(\lambda_{n+1}, \theta_{n+1})$, is simulated in two steps. Indeed, we draw λ_{n+1} from the conditional posterior density of λ given $\theta = \theta_n$, which is a product of $r+1$ univariate gamma densities, and then we draw θ_{n+1} from the conditional posterior density of θ given $\lambda = \lambda_{n+1}$, which is a $(p+q)$-dimensional multivariate normal density. The exact forms of these conditional densities are given in Section 2.

The main theoretical contribution of this article is the following result, which provides easily-checked conditions under which the block Gibbs Markov chain is geometrically ergodic. A Markov chain is said to be geometrically ergodic if it converges to its invariant distribution (using total variation distance) at a geometric rate. For a formal definition of geometric ergodicity, see Meyn and Tweedie (1993, Chapter 15).

Proposition 1. *The block Gibbs Markov chain, $\{(\lambda_n, \theta_n)\}_{n=0}^\infty$, is geometrically ergodic if*

1. *X has full column rank,*
2. *$a_e > \frac{1}{2}(\mathit{rank}(Z) - N + 2)$, and*
3. *$\min\left\{a_1 + \frac{q_1}{2}, \ldots, a_r + \frac{q_r}{2}\right\} > \frac{1}{2}\big(q - \mathit{rank}(Z)\big) + 1$.*

Note that the conditions of Proposition 1 do not involve the hyper-parameters μ_β, Σ_β and b, nor do they involve the observed response vector, y.

There are well known advantages to using an MCMC algorithm that is driven by a geometrically ergodic Markov chain (see, e.g., Roberts and Rosenthal, 1998, 2004). In particular, when the chain is geometric, sample averages satisfy central limit theorems, and these allow for the computation of asymptotically valid standard errors for MCMC-based estimates (Bednorz and Łatuszyński, 2007; Flegal and Jones, 2010; Hobert et al., 2002; Jones et al., 2006). The ability to compute such standard errors is very important from a practical standpoint because it leads to a coherent strategy for deciding how long to run the simulation. In the remainder of this section, we describe how our results relate to the existing literature on convergence rates of Gibbs samplers for Bayesian linear models.

Proposition 1 is a substantial improvement upon the main result in Johnson and Jones (2010), which is applicable only when: (i) $X^T Z = 0$, (ii) $r = 1$, and (iii) Z is of full rank. Note that the condition $X^T Z = 0$ would rarely, if ever, hold in practice. Moreover, this condition implies a certain conditional independence which greatly simplifies the analysis of the Gibbs Markov chain (see Section 2 for details). The key ideas behind our analysis are new and different from the approach taken by Johnson and Jones (2010). In particular, we exploit singular value and spectral decompositions of key matrices, and this allows us to establish a series of matrix inequalities that lead to very weak conditions that imply geometric ergodicity.

An important special case of the general linear mixed model is the one-way random effects model given by

$$Y_{ij} = \beta + \alpha_i + \epsilon_{ij} \ ,$$

where $i = 1, \ldots, c$, $j = 1, \ldots, n_i$, the α_is are iid $\mathrm{N}(0, \lambda_\alpha^{-1})$, the ϵ_{ij}s, which are independent of the α_is, are iid $\mathrm{N}(0, \lambda_e^{-1})$, and the priors for β and the precision parameters are the usual normal and gamma distributions (with known hyper-parameters). This is the "non-centered" parameterization (NCP) of the one-way model. In the "centered" parameterization (CP), the parameter β does not appear in the model equation $-$ it appears as the mean of the α_is. (The CP model is not a special case of our linear model.) An application of our Proposition 1 to the NCP model yields the following result.

Corollary 1. *Assume that $c \geq 2$ and $n_i \geq 2$ for $i = 1, 2, \ldots, c$. Then the block Gibbs sampler associated with the NCP one-way model is geometrically ergodic.*

It should be noted that the conditions of Corollary 1 do not involve the hyper-parameters at all; that is, if the sample size conditions hold, then the block Gibbs sampler for the NCP one-way model is geometrically ergodic for *any* choice of hyper-parameters. This suggests that the conditions on the hyper-parameters in Proposition 1 are weak.

In general, Gibbs samplers derived under different parametrizations of a target density may have different convergence rates (see, e.g., Gelfand et al. (1995); Papaspiliopoulos et al. (2007)). However, a recent result of Román et al. (2013) shows that the block Gibbs samplers for the CP and NCP models described above converge at exactly the same rate. Hence, the conditions of Corollary 1 also imply the geometric ergodicity of the block Gibbs sampler for the CP model. This result was actually established directly using geometric drift conditions in Hobert and Geyer (1998).

Other related work can be found in Román and Hobert (2012) which provides a geometric ergodicity result (analogous to our Proposition 1) for a family of improper priors that includes the priors considered in Hobert and Casella (1996). Aside from the work that has already been cited, the only other articles in which convergence rates of Gibbs samplers for Bayesian linear models are studied are Jones and Hobert (2004), which considered the same model as Hobert and Geyer (1998), Tan and Hobert (2009), which examined a CP one-way model with improper priors, and Papaspiliopoulos and Roberts (2008), which considered simpler linear models with known variance components.

The remainder of this paper is organized as follows. Section 2 contains a formal definition of the block Gibbs Markov chain. The proof of Proposition 1 is presented in Section 3. Finally, in Section 4, we provide a simple extension of Proposition 1 that is applicable when the conditional conjugate priors are replaced by finite mixtures of such.

2 The Gibbs Sampler

We begin with a formal definition of the Markov transition density (Mtd) of the block Gibbs Markov chain. Let $\|A\|$ be the Frobenius norm of matrix A; that is, $\|A\| = \sqrt{\text{tr}(A^T A)}$. It follows from the form of the target posterior density, $\pi(\lambda, \theta|y)$, that the components of λ are conditionally independent given $\theta = (\beta^T u^T)^T$ (and the data y); that is,

$$\pi(\lambda|\theta, y) = \pi(\lambda_e|\theta, y) \times \prod_{i=1}^{r} \pi(\lambda_{u_i}|\theta, y) \; .$$

Moreover, it's easy to show that

$$\lambda_e|\theta, y \sim \text{Gamma}\left(a_e + \frac{N}{2}, b_e + \frac{\|y - W\theta\|^2}{2}\right) ,$$

and, for $i \in \{1, 2, \ldots, r\}$,

$$\lambda_{u_i}|\theta, y \sim \text{Gamma}\left(a_i + \frac{q_i}{2}, b_i + \frac{\|u_i\|^2}{2}\right) ,$$

where $W = (X\ Z)$.

Routine manipulation of $\pi(\lambda, \theta|y)$ shows that $\pi(\theta|\lambda, y)$ is a multivariate normal density. To define the mean and covariance matrix, we need to introduce a bit more notation. Define $T_\lambda = \lambda_e X^T X + \Sigma_\beta^{-1}$, $M_\lambda = I - \lambda_e X T_\lambda^{-1} X^T$, and $Q_\lambda = \lambda_e Z^T M_\lambda Z + \Lambda_u$. The mean of the multivariate normal is

$$E(\theta|\lambda, y) = \begin{bmatrix} T_\lambda^{-1}(\lambda_e X^T y + \Sigma_\beta^{-1}\mu_\beta) - \lambda_e^2 T_\lambda^{-1} X^T Z Q_\lambda^{-1} Z^T (M_\lambda y - X T_\lambda^{-1} \Sigma_\beta^{-1} \mu_\beta) \\ \lambda_e Q_\lambda^{-1} Z^T (M_\lambda y - X T_\lambda^{-1} \Sigma_\beta^{-1} \mu_\beta) \end{bmatrix} ,$$

and the covariance matrix is

$$\text{Var}(\theta|\lambda, y) = \begin{bmatrix} T_\lambda^{-1} + \lambda_e^2 T_\lambda^{-1} X^T Z Q_\lambda^{-1} Z^T X T_\lambda^{-1} & -\lambda_e T_\lambda^{-1} X^T Z Q_\lambda^{-1} \\ -\lambda_e Q_\lambda^{-1} Z^T X T_\lambda^{-1} & Q_\lambda^{-1} \end{bmatrix} .$$

Recall that Johnson and Jones (2010) require that $X^T Z = 0$, and note the massive simplification that occurs when this is the case. In particular, when $X^T Z = 0$, the two components of θ, that is β and u, are conditionally independent given λ and y.

The Mtd of the block Gibbs Markov chain is

$$k(\lambda, \theta|\lambda', \theta') = \pi(\theta|\lambda)\pi(\lambda|\theta') . \tag{4}$$

That is, if the current state is $(\lambda_n, \theta_n) = (\lambda', \theta')$, then the density of the next state, $(\lambda_{n+1}, \theta_{n+1})$, is (4). (Since the data are fixed throughout our convergence analysis, the dependency on y will be suppressed in the notation.)

It is easy to see that the two marginal sequences, $\{\lambda_n\}_{n=0}^\infty$ and $\{\theta_n\}_{n=0}^\infty$, are themselves (reversible) Markov chains and their invariant densities are the marginal posterior distributions of λ and θ, respectively. Furthermore, all three Markov chains are Harris ergodic (Román, 2012), and geometric ergodicity is a solidarity property for these chains (Diaconis et al., 2008; Roberts and Rosenthal, 2001). That is, either all three chains are geometrically ergodic or none of them is. Consequently, we can prove that the block Gibbs chain is geometrically ergodic by proving that either of the marginal chains converges at a geometric rate.

3 Geometric Ergodicity

In this section, we will establish Proposition 1 by showing that the θ-chain, $\{\theta_n\}_{n=0}^{\infty}$, is geometrically ergodic. This is accomplished by establishing a *geometric drift condition* for the θ-chain. In particular, we will prove the following result.

Proposition 2. *Under the three conditions of Proposition 1, there exist a $\rho \in [0,1)$ and a finite constant L such that, for every $\theta' \in \mathbb{R}^{p+q}$,*

$$E\big(v(\theta)|\theta'\big) \leq \rho\, v(\theta') + L\ , \tag{5}$$

where the drift function is defined as

$$v(\theta) = \alpha \left\|y - W\theta\right\|^2 + \left\|u\right\|^2\ .$$

The constant α can be any real number that satisfies

$$\alpha > \frac{tr\big((Z^TZ)^+\big)}{2a_e + N - 2 - rank(Z)} > 0\ ,$$

where A^+ denotes the Moore-Penrose inverse of the matrix A.

For a precise explanation of the reason why the geometric drift condition (5) implies geometric ergodicity of the θ-chain, see Appendix A.

Remark 1. *Johnson and Jones (2010) also used a geometric drift function approach, and their drift function, which is quite similar to ours, is given by $v_{JJ}(\theta) = \|y - W\theta\|^2 + \|u\|^2$. We note, however, that their proof relies on an application of their Lemma A.2., which is actually false as stated. For a counterexample, take (in their notation) $m = 1$, $x = 1$, $A = 2$, and $B = 1$.*

Remark 2. *Recall that Román and Hobert (2012) provide an analogous result for a family of improper priors. The proof of geometric ergodicity under proper and improper priors are similar in that a drift analysis is employed in both cases. However, due to some technical difficulties related to the use of improper priors, Roman and Hobert (2012) were forced to work with the marginal Markov chain associated with the variance components, which is quite different than working with the θ-chain as is done herein.*

It should be noted that the constants ρ and L in (5) depend on the choice of α. Specifically,

$$\rho = \rho(\alpha) = \frac{1}{2}\max\left\{\frac{\text{rank}(Z) + \alpha^{-1}\text{tr}\big((Z^TZ)^+\big)}{a_e + \frac{N}{2} - 1}, \frac{q - \text{rank}(Z)}{\min\big\{a_1 + \frac{q_1}{2}, \ldots, a_r + \frac{q_r}{2}\big\} - 1}\right\}\ , \tag{6}$$

and

$$L = L(\alpha) = \big(\alpha\,\mathrm{rank}(Z) + \mathrm{tr}\big((Z^TZ)^+\big)\big)\,\frac{b_e}{a_e + \frac{N}{2} - 1} + (q - \mathrm{rank}(Z))\sum_{i=1}^{r}\frac{b_i}{a_i + \frac{q_i}{2} - 1} + K(\alpha)\ ,$$

where $K(\alpha) < \infty$. The expression for $K(\alpha)$ is

$$K(\alpha) = \alpha\Big(s_{\max}^2\mathrm{rank}(Z) + \mathrm{tr}(X\Sigma_\beta X^T) + C^2\Big) + s_{\max}^2\mathrm{tr}\big((Z^TZ)^+\big) + \big(K_y + K_{\mu_\beta}\big)^2\ , \quad (7)$$

where $s_{\max}$ denotes the largest singular value of $X\Sigma_\beta^{1/2}$,

$$C = \sqrt{N}\|y\| + \|X\Sigma_\beta X^T\|\|X(X^TX)^{-1}\Sigma_\beta^{-1}\mu_\beta\| + \sqrt{N}\|Z\|(K_y + K_{\mu_\beta})\ ,$$

and the expressions for the constants K_y and K_{μ_β} are given in Appendix B.4.

Before we present our proof of Proposition 2, we provide an outline of it. First, note that the left-hand side of (5) satisfies

$$E(v(\theta)|\theta') = E\big[E(v(\theta)|\lambda)\big|\theta'\big] = E\big[\alpha E(\|y - W\theta\|^2|\lambda)|\theta'\big] + E\big[E\big(\|u\|^2|\lambda\big)\big|\theta'\big]\ . \quad (8)$$

Suppose for a moment that we are able to obtain a functional upper bound on $E(v(\theta)|\lambda)$ of the form

$$B_c(\lambda) = c_e(\alpha)\lambda_e^{-1} + c_u\sum_{i=1}^{r}\lambda_{u_i}^{-1} + K(\alpha)\ , \quad (9)$$

for some $c = (c_e\ c_u)^T$ (c_e might depend on α) and $K(\alpha)$ is as above. Under the conditions of Proposition 1, we have

$$\min\Big\{a_e + \frac{N}{2}, a_1 + \frac{q_1}{2}, \ldots, a_r + \frac{q_r}{2}\Big\} > 1\ ,$$

which implies that $E\big(\lambda_e^{-1}\big|\theta'\big)$ and $E\big(\lambda_{u_i}^{-1}\big|\theta'\big)$, $i \in \{1, 2, \ldots, r\}$ are all finite. Define

$$d_e = \frac{b_e}{a_e + \frac{N}{2} - 1} \quad \text{and} \quad d_i = \frac{b_i}{a_i + \frac{q_i}{2} - 1}\ , \quad i = 1, 2, \ldots, r\ .$$

Then, it is easy to verify that

$$E\big(\lambda_e^{-1}\big|\theta'\big) = \frac{\|y - W\theta'\|^2}{2a_e + N - 2} + d_e\ ,$$

and, for each $i \in \{1, 2, \ldots, r\}$,

$$E\big(\lambda_{u_i}^{-1}\big|\theta'\big) = \frac{\|u_i'\|^2}{2a_i + q_i - 2} + d_i\ .$$

Recall (9) and note that

$$E\big(B_c(\lambda)|\theta'\big) \le C_e(\alpha)\|y - W\theta'\|^2 + C_u\|u'\|^2 + c_e(\alpha)d_e + c_u\sum_{i=1}^{r} d_i + K(\alpha)\ ,$$

where

$$C_e(\alpha) = \frac{c_e(\alpha)}{2a_e + N - 2} \quad \text{and} \quad C_u = \frac{c_u}{2\min\big\{a_1 + \frac{q_1}{2}, \ldots, a_r + \frac{q_r}{2}\big\} - 2}\ .$$

Thus, combining what we have so far, we obtain

$$E(v(\theta)|\theta') = E\big[E(v(\theta)|\lambda)\big|\theta'\big] \le E\big(B_c(\lambda)|\theta'\big) \le \rho(\alpha)v(\theta') + L(\alpha)\ , \tag{10}$$

where $\rho(\alpha) = \max\{C_e(\alpha)\alpha^{-1}, C_u\}$ and

$$L(\alpha) = c_e(\alpha)d_e + c_u\sum_{i=1}^{r} d_i + K(\alpha) < \infty.$$

Hence, establishing the drift condition (5) comes down to a search for functional upper bounds on $E(v(\theta)|\lambda)$ [as in (9)] that lead to $\rho(\alpha) < 1$ for some α.

It follows from (8) that, in order to obtain functional upper bounds on $E(v(\theta)|\lambda)$ of the form $B_c(\lambda)$, we can focus our attention on the terms

$$E\big(\|y - W\theta\|^2\big|\lambda\big) = \text{tr}(W\,\text{Var}(\theta|\lambda)W^T) + \|y - WE(\theta|\lambda)\|^2 \tag{11}$$

and

$$E\big(\|u\|^2|\lambda\big) = \text{tr}\big(Q_\lambda^{-1}\big) + \|E(u|\lambda)\|^2\ , \tag{12}$$

which are complicated functions of λ. Recall from Section 2 that the matrix $\text{Var}(\theta|\lambda)$ involves the matrices $T_\lambda^{-1} = (\lambda_e X^T X + \Sigma_\beta^{-1})^{-1}$ and $Q_\lambda^{-1} = (\lambda_e Z^T M_\lambda Z + \Lambda_u)^{-1}$.

Our strategy for finding functional upper bounds of the form $B_c(\lambda)$ entails two main steps: (i) to find functional upper bounds on the "trace" terms in (11) and (12), and (ii) to show that, as functions of λ, the "norm" terms on the right-hand side of (11) and (12) are bounded. That is, only the bounds for the "trace" terms will be allowed to depend on λ.

We now state two preliminary results that will be used to obtain the bound $B_c(\lambda)$. Appendix B contains the proofs. The first result provides an upper bound on the "trace" terms $\text{tr}\big(Q_\lambda^{-1}\big)$ and

$$\text{tr}(W\,\text{Var}(\theta|\lambda)W^T) = \text{tr}\big(ZQ_\lambda^{-1}Z^T\big) + \text{tr}(XT_\lambda^{-1}X^T) - \text{tr}\big((I - M_\lambda)ZQ_\lambda^{-1}Z^T(I + M_\lambda)\big)\ . \tag{13}$$

Lemma 1. *If rank*$(X) = p$*, then for all* $\lambda \in \mathbb{R}_+^{r+1}$*,*

1. $tr\big(Q_\lambda^{-1}\big) \le \lambda_e^{-1} tr\big((Z^TZ)^+\big) + \big(q - rank(Z)\big)\big(\sum_{i=1}^r \lambda_{u_i}^{-1}\big) + s_{\max}^2 tr\big((Z^TZ)^+\big)$

2. $tr(W\,\mathrm{Var}(\theta|\lambda)W^T) \le rank(Z)\,\lambda_e^{-1} + s_{\max}^2 rank(Z) + tr(X\Sigma_\beta X^T)$.

The second preliminary result will be used to obtain an upper bound for the "norm" terms.

Lemma 2. *If* $rank(X) = p$*, then, for all* $\lambda \in \mathbb{R}_+^{r+1}$*,*

$$\|\lambda_e Q_\lambda^{-1} Z^T M_\lambda\, y\| \le K_y \quad \textit{and} \quad \|\lambda_e Q_\lambda^{-1} Z^T X T_\lambda^{-1} \Sigma_\beta^{-1}\| \le K_{\mu_\beta}\ .$$

We are now in position to prove Proposition 2.

Proof of Proposition 2. We begin by bounding the "norm" terms.

An application of Lemma 2 together with the triangle inequality yields

$$\begin{aligned}\|E(u|\lambda)\| &= \|\lambda_e Q_\lambda^{-1} Z^T (M_\lambda y - X T_\lambda^{-1}\Sigma_\beta^{-1}\mu_\beta)\| \\ &\le \|\lambda_e Q_\lambda^{-1} Z^T M_\lambda y\| + \|\lambda_e Q_\lambda^{-1} Z^T X T_\lambda^{-1}\Sigma_\beta^{-1}\mu_\beta\| \le K_y + K_{\mu_\beta}\ .\end{aligned} \tag{14}$$

Moreover, an application of Lemmas 4 and 5 in Appendix B shows that

$$\|X T_\lambda^{-1} X^T\| \le \|X\Sigma_\beta X^T\| \quad \text{and} \quad \|M_\lambda\| = \sqrt{N}\ . \tag{15}$$

Thus, using (15) and properties of the (Frobenius) matrix norm, we see that

$$\begin{aligned}\|y - WE(\theta|\lambda)\| &= \|M_\lambda y - X T_\lambda^{-1}\Sigma_\beta^{-1}\mu_\beta - M_\lambda Z E(u|\lambda)\| \\ &\le \|M_\lambda y\| + \|X T_\lambda^{-1} X^T X (X^TX)^{-1}\Sigma_\beta^{-1}\mu_\beta\| + \|M_\lambda\|\|Z\|\|E(u|\lambda)\| \\ &\le \|M_\lambda\|\|y\| + \|X T_\lambda^{-1} X^T\|\|X(X^TX)^{-1}\Sigma_\beta^{-1}\mu_\beta\| + \|M_\lambda\|\|Z\|\|E(u|\lambda)\| \\ &\le \sqrt{N}\|y\| + \|X\Sigma_\beta X^T\|\|X(X^TX)^{-1}\Sigma_\beta^{-1}\mu_\beta\| + \sqrt{N}\|Z\|(K_y + K_{\mu_\beta}) \\ &= C\ ,\end{aligned} \tag{16}$$

where C is the constant that appears in (7).

Finally, an application of Lemma 1 together with (14), and (16), gives us the following upper bound on $E(v(\theta)|\lambda)$

$$B_c(\lambda) = c_e(\alpha)\lambda_e^{-1} + c_u \sum_{i=1}^r \lambda_{u_i}^{-1} + K(\alpha) \tag{17}$$

where $K(\alpha)$ is as in (7), $c_e(\alpha) = \alpha\,\mathrm{rank}(Z) + \mathrm{tr}\big((Z^TZ)^+\big)$ and $c_u = q - \mathrm{rank}(Z)$.

To finish the proof, recall from (10) that the resulting ρ takes the form $\rho(\alpha) = \max\{C_e(\alpha)\alpha^{-1}, C_u\}$ which is exactly the ρ that appears in (6). Moreover, under the conditions of Proposition 1, $C_u < 1$ and $\text{rank}(Z) < 2a_e + N - 2$. Thus, for any

$$\alpha > \alpha' := \frac{\text{tr}\big((Z^T Z)^+\big)}{2a_e + N - 2 - \text{rank}(Z)} > 0 \ ,$$

we have

$$C_e(\alpha)\alpha^{-1} = \frac{\text{rank}(Z)}{2a_e + N - 2} + \alpha^{-1}\frac{\text{tr}\big((Z^T Z)^+\big)}{2a_e + N - 2} < 1 \ .$$

Therefore there exists $\alpha > \alpha'$ with $\rho(\alpha) = \max\{C_e(\alpha)\alpha^{-1}, C_u\} < 1$, and this completes the proof.

□

4 An Extension to Mixture Priors

We now consider a generalization of our Bayesian hierarchical model in which the conditionally conjugate priors are replaced by finite mixtures of such. Specifically, we replace the prior for β in (2) with

$$\beta|\lambda \sim \sum_i w_i \text{N}_p(\mu_\beta^{(i)}, \Sigma_\beta^{(i)}) \ ,$$

where the w_is are (known) non-negative weights that sum to one and the $\mu_\beta^{(i)}$s and $\Sigma_\beta^{(i)}$s are known hyper-parameters. Also, we replace the priors for the precision parameters in (3) with

$$\lambda_e \sim \sum_j \omega_j \text{Gamma}(a_e^{(j)}, b_e^{(j)}) \quad \text{and} \quad \lambda_{u_k} \sim \sum_{l_k} \varpi_{kl_k} \text{Gamma}(a_k^{(l_k)}, b_k^{(l_k)}) \ , \quad k = 1, 2, \ldots, r \ ,$$

where the ω_js and the ϖ_{kl_k}s are weights and $a_e^{(j)}$, $b_e^{(j)}$, $a_k^{(l_k)}$, and $b_k^{(l_k)}$ are the known hyper-parameters.

Johnson and Jones (2010) also considered this model (with $r = 1$), but as we explain below, the full conditional distributions given in that paper are incorrect. In the remainder of this section, we derive the correct full conditionals, and sketch the proof of the following extension of Proposition 1.

Proposition 3. *The block Gibbs Markov chain, $\{(\lambda_n, \theta_n)\}_{n=0}^{\infty}$, is geometrically ergodic if*

1. *X has full column rank,*
2. *$\min_j a_e^{(j)} > \frac{1}{2}(\text{rank}(Z) - N + 2)$, and*
3. *$\min_k \min_{l_k} \big\{a_k^{(l_k)} + \frac{q_k}{2}\big\} > \frac{1}{2}\big(q - \text{rank}(Z)\big) + 1$.*

To keep the notation under control, we restrict attention to the case where $r = 1$. The argument for the general case is similar and it only involves more tedious notation.

When $r = 1$, the posterior density is

$$\tilde{\pi}(\lambda, \theta|y) = \frac{\tilde{\pi}^*(\lambda, \theta|y)}{\tilde{m}(y)} \ ,$$

where $\tilde{\pi}^*$ is an unnormalized density given by

$$\tilde{\pi}^*(\lambda, \theta|y) = \sum_i \sum_j \sum_l w_i \omega_j \varpi_l \tilde{\pi}^*_{ijl}(\lambda, \theta|y) \ ,$$

and

$$\tilde{\pi}^*_{ijl}(\lambda, \theta|y) = f_{\mathrm{N}}(y; W\theta, \lambda_e^{-1} I)\, f_{\mathrm{N}}(\beta; \mu_\beta^{(i)}, \Sigma_\beta^{(i)})\, f_{\mathrm{N}}\left(u; 0, \Lambda_u^{-1}\right)\ f_{\mathrm{G}}(\lambda_e; a_e^{(j)}, b_e^{(j)}) f_{\mathrm{G}}(\lambda_{u_1}; a_1^{(l)}, b_1^{(l)}) \ ,$$

and the normalizing constant is given by

$$\tilde{m}(y) = \sum_i \sum_j \sum_l w_i \omega_j \varpi_l \int_{\mathbb{R}^2_+} \int_{\mathbb{R}^{p+q}} \tilde{\pi}^*_{ijl}(\lambda, \theta|y)\, d\theta d\lambda < \infty \ .$$

We now derive the full conditional distributions. As before, λ_e and $\lambda_u := \lambda_{u_1}$ are conditionally independent (given θ and y) but, in this case, their densities are

$$\sum_j \tilde{\omega}_j f_{\mathrm{G}}\left(\lambda_e; a_e^{(j)} + \frac{N}{2}, b_e^{(j)} + \frac{\|y - W\theta\|^2}{2}\right) \quad \text{and} \quad \sum_l \tilde{\varpi}_l f_{\mathrm{G}}\left(\lambda_u; a_1^{(l)} + \frac{q_1}{2}, b_1^{(l)} + \frac{\|u\|^2}{2}\right) \ , \tag{18}$$

where the (new) weights are

$$\tilde{\omega}_j = \tilde{\omega}_j(\theta, y) := \frac{\omega_j \Delta_j^e(\theta, y)}{\sum_{j'} \omega_{j'} \Delta_{j'}^e(\theta, y)} \quad \text{with} \quad \Delta_j^e(\theta, y) := \frac{\Gamma(\frac{N}{2} + a_e^{(j)})(b_e^{(j)})^{a_e^{(j)}}}{\Gamma(a_e^{(j)}) \left(b_e^{(j)} + \frac{\|y - W\theta\|^2}{2}\right)^{a_e^{(j)}}} \ , \tag{19}$$

and

$$\tilde{\varpi}_l = \tilde{\varpi}_l(\theta) := \frac{\varpi_l \Delta_l^u(\theta)}{\sum_{l'} \varpi_{l'} \Delta_{l'}^u(\theta)} \quad \text{with} \quad \Delta_l^u(\theta) := \frac{\Gamma(\frac{q}{2} + a_1^{(l)})(b_1^{(l)})^{a_1^{(l)}}}{\Gamma(a_1^{(l)}) \left(b_1^{(l)} + \frac{\|u\|^2}{2}\right)^{a_1^{(l)}}} \ . \tag{20}$$

So the full conditional distributions of the precision parameters are also finite mixtures. That is, when the prior is taken to be a finite mixture of (conditionally) conjugate priors, the full conditional distribution takes the same form. However, it is important to note that the weights in the full conditional distribution are not the same as the prior weights. This is the error in Johnson and Jones (2010) − the weights in their full conditional distributions are the same as the prior weights.

To establish (18)-(20), note that

$$\begin{aligned}\tilde{\pi}(\lambda|\theta,y) &= \frac{\tilde{\pi}^*(\lambda,\theta|y)}{\int_{\mathbb{R}^2_+}\tilde{\pi}^*(\lambda,\theta|y)d\lambda} \\ &= \sum_i\sum_j\sum_l \frac{w_i\omega_j\varpi_l}{\sum_{i'}\sum_{j'}\sum_{l'}w_{i'}\omega_{j'}\varpi_{l'}\int_{\mathbb{R}^2_+}\tilde{\pi}^*_{i'j'l'}(\lambda,\theta|y)d\lambda}\tilde{\pi}^*_{ijl}(\lambda,\theta|y)\ .\end{aligned}$$

The unnormalized density $\tilde{\pi}^*_{ijl}(\lambda,\theta|y)$ can be written as the product of

$$(2\pi)^{-\frac{N+q}{2}}\Delta^e_j(\theta,y)\Delta^u_l(\theta)f_{\mathrm{N}}(\beta;\mu^{(i)}_\beta,\Sigma^{(i)}_\beta)$$

and

$$f_{\mathrm{G}}\left(\lambda_e;\frac{N}{2}+a^{(j)}_e,b^{(j)}_e+\frac{\|y-W\theta\|^2}{2}\right)f_{\mathrm{G}}\left(\lambda_u;\frac{q}{2}+a^{(l)}_1,b^{(l)}_1+\frac{\|u\|^2}{2}\right)\ ,$$

which makes it easy to see that

$$\int_{\mathbb{R}^2_+}\tilde{\pi}^*_{ijl}(\lambda,\theta|y)d\lambda=(2\pi)^{-\frac{N+q}{2}}\Delta^e_j(\theta,y)\Delta^u_l(\theta)f_{\mathrm{N}}(\beta;\mu^{(i)}_\beta,\Sigma^{(i)}_\beta)\ .$$

Therefore $\tilde{\pi}(\lambda|\theta,y)$ can be expressed as

$$\begin{aligned}&\sum_j\sum_l\sum_i\frac{w_i\omega_j\varpi_l}{\sum_{j'}\sum_{l'}\omega_{j'}\varpi_{l'}(2\pi)^{-\frac{N+q}{2}}\Delta^e_{j'}(\theta,y)\Delta^u_{l'}(\theta)\left[\sum_{i'}w_{i'}f_{\mathrm{N}}(\beta;\mu^{(i')}_\beta,\Sigma^{(i')}_\beta)\right]}\tilde{\pi}^*_{ijl}(\lambda,\theta|y)\\ &=\sum_j\sum_l\tilde{\omega}_j(\theta,y)\tilde{\varpi}_l(\theta)f_{\mathrm{G}}\left(\lambda_e;\frac{N}{2}+a^{(j)}_e,b^{(j)}_e+\frac{\|y-W\theta\|^2}{2}\right)f_{\mathrm{G}}\left(\lambda_u;\frac{q}{2}+a^{(l)}_1,b^{(l)}_1+\frac{\|u\|^2}{2}\right)\ ,\end{aligned}$$

which establishes (18).

We now concentrate on the conditional density $\tilde{\pi}(\theta|\lambda,y)$. A calculation similar to the one given above shows that

$$\tilde{\pi}(\theta|\lambda,y)=\sum_i\tilde{w}_i(\lambda,y)f_{\mathrm{N}}(\theta;m_i,V_i)\ ,$$

where

$$\tilde{w}_i(\lambda,y)=\frac{w_iC^\theta_i(\lambda,y)}{\sum_{i'}w_{i'}C^\theta_{i'}(\lambda,y)}\quad\text{with}\quad C^\theta_i(\lambda,y)=\left(\frac{\det(V_i)}{\det(\Sigma^{(i)}_\beta)}\right)^{\frac{1}{2}}e^{-\frac{1}{2}(\mu^{(i)}_\beta)^T(\Sigma^{(i)}_\beta)^{-1}\mu^{(i)}_\beta}e^{\frac{1}{2}m_i^TV_i^{-1}m_i}\ ,$$

and m_i and V_i are the conditional mean and variance-covariance matrix of $\theta|\lambda,y$, respectively, when the prior for β is $\mathrm{N}_p(\mu^{(i)}_\beta,\Sigma^{(i)}_\beta)$.

The Mtd of the block Gibbs sampler for the mixture case is

$$\tilde{k}(\lambda,\theta|\lambda',\theta')=\tilde{\pi}(\theta|\lambda)\tilde{\pi}(\lambda|\theta')\ ,$$

where the dependence on the data is suppressed in the notation. Our strategy for establishing geometric ergodicity remains the same as before; that is, we will analyze the θ-chain using the same exact drift function that we used in the non-mixture case.

In the mixture case, we have

$$E(v(\theta)|\lambda) = \sum_i \tilde{w}_i(\lambda, y) E_i(v(\theta)|\lambda) \ ,$$

where E_i denotes the (conditional) expectation with respect to $f_{\text{N}}(\theta; m_i, V_i)$. Using the same calculations that led to (17), we see that

$$E_i(v(\theta)|\lambda) \leq c_e(\alpha)\lambda_e^{-1} + c_u\lambda_u^{-1} + K_i(\alpha)$$

where K_i is just like K but uses $\mu_\beta^{(i)}$ and $\Sigma_\beta^{(i)}$ instead of β and Σ_β. It is important to note that $c_e(\alpha) = \alpha\,\text{rank}(Z) + \text{tr}\big((Z^TZ)^+\big)$ and $c_u = q - \text{rank}(Z)$ do not depend on i. Now using the fact that the weights are non-negative and sum to one, we have

$$E(v(\theta)|\lambda) \leq c_e(\alpha)\lambda_e^{-1} + c_u\lambda_u^{-1} + \sum_i \tilde{w}_i(\lambda, y)K_i(\alpha) \leq c_e(\alpha)\lambda_e^{-1} + c_u\lambda_u^{-1} + \sum_i K_i(\alpha) \ ,$$

which leads to

$$\begin{aligned} E(v(\theta)|\theta') &= E\big[E(v(\theta)|\lambda)\big|\theta'\big] \\ &\leq c_e(\alpha)E(\lambda_e^{-1}|\theta') + c_uE(\lambda_u^{-1}|\theta') + \sum_i K_i(\alpha) \\ &= c_e(\alpha)\sum_j \tilde{\omega}_j(\theta, y)E_j(\lambda_e^{-1}|\theta') + c_u\sum_l \tilde{\varpi}_l(\theta)E_l(\lambda_u^{-1}|\theta') + \sum_i K_i(\alpha) \ , \end{aligned}$$

where E_j and E_l are expectations corresponding to

$$f_{\text{G}}\left(\lambda_e; a_e^{(j)} + \frac{N}{2}, b_e^{(j)} + \frac{\|y - W\theta\|^2}{2}\right) \quad \text{and} \quad f_{\text{G}}\left(\lambda_u; a_1^{(l)} + \frac{q_1}{2}, b_1^{(l)} + \frac{\|u\|^2}{2}\right) \ ,$$

respectively. It is easy to see that the hypotheses of Proposition 3 guarantee that

$$E_j\big(\lambda_e^{-1}\big|\theta'\big) = \frac{\|y - W\theta'\|^2}{2a_e^{(j)} + N - 2} + d_e^{(j)} \quad \text{and} \quad E_l\big(\lambda_u^{-1}\big|\theta'\big) = \frac{\|u'\|^2}{2a_1^{(l)} + q_1 - 2} + d_1^{(l)} \ ,$$

where

$$d_e^{(j)} = \frac{b_e^{(j)}}{a_e^{(j)} + \frac{N}{2} - 1} \quad \text{and} \quad d_1^{(l)} = \frac{b_1^{(l)}}{a_1^{(l)} + \frac{q_1}{2} - 1} \ .$$

Using these facts, we obtain

$$E(v(\theta)|\theta') \leq \rho'(\alpha)v(\theta') + L'(\alpha) \ ,$$

where

$$\rho'(\alpha) = \max\left\{\alpha^{-1}c_e(\alpha)\sum_j \frac{\tilde{\omega}_j(\theta,y)}{2a_e^{(j)}+N-2}, c_u\sum_l \frac{\tilde{\varpi}_l(\theta)}{2a_1^{(l)}+q_1-2}\right\}$$

and

$$L'(\alpha) = \sum_i K_i(\alpha) + c_e(\alpha)\sum_j d_e^{(j)} + c_u\sum_l d_1^{(l)} .$$

Moreover, since the weights sum to one, we have

$$\begin{aligned}\rho'(\alpha) &= \max\left\{\alpha^{-1}c_e(\alpha)\sum_j \frac{\tilde{\omega}_j(\theta,y)}{2a_e^{(j)}+N-2}, c_u\sum_l \frac{\tilde{\varpi}_l(\theta)}{2a_1^{(l)}+q_1-2}\right\} \\ &\leq \max\left\{\alpha^{-1}\frac{c_e(\alpha)}{2\min_j a_e^{(j)}+N-2}, \frac{c_u}{2\min_l a_1^{(l)}+q_1-2}\right\} .\end{aligned}$$

An argument similar to the one used in the proof of Proposition 2 shows that the hypotheses of Proposition 3 ensure the existence of an α such that $\rho'(\alpha) < 1$. This establishes the desired drift condition which guarantees the geometric ergodicity of the block Gibbs sampler in the mixture case.

Acknowledgment. The second author was supported by NSF Grant DMS-11-06395.

Appendices

A Why Does the Drift Condition Imply Geometric Convergence?

Recall that our drift function is given by

$$v(\theta) = \alpha\,\|y - W\theta\|^2 + \|u\|^2 ,$$

where $\alpha > 0$ does not depend on θ. Here we will show that if X has full column rank then $v(\theta)$ is *unbounded off compact sets*; that is, for every $d \in \mathbb{R}$, the set

$$S_d = \left\{\theta \in \mathbb{R}^{p+q} : \alpha\|y - W\theta\|^2 + \|u\|^2 \leq d\right\}$$

is compact. There are two cases. If d is such that $S_d = \emptyset$, then S_d is trivially compact. So suppose now that S_d is non-empty and let u_{ij} denote the jth component of u_i, where $i = 1, \ldots, r$ and $j = 1, \ldots, q_i$. Since $v(\theta)$ is continuous, S_d must be closed, so it suffices to show that β_l

and u_{ij} are bounded for all $l = 1, \ldots, p$, all $i = 1, \ldots, r$, and all $j = 1, \ldots, q_i$. Since $\|u\|^2 = \sum_{i=1}^{r} \sum_{j=1}^{q_i} u_{ij}^2 \to \infty$ as $|u_{ij}| \to \infty$, we have all the u_{ij} contained. Now for a given vector u, let $y_u := y - Zu$ and $\hat{\beta}_u := (X^T X)^{-1} X^T y_u$. Note that since all the $|u_{ij}|$ are contained, the elements of y_u and $\hat{\beta}_u$ are also bounded. Finally, since $\alpha \|y - W\theta\|^2 \leq d$ and X has full rank we have

$$\upsilon_{\min}(X^T X)(\beta - \hat{\beta}_u)^T(\beta - \hat{\beta}_u) \leq \|y_u - X\hat{\beta}_u\|^2 + (\beta - \hat{\beta}_u)^T X^T X(\beta - \hat{\beta}_u) = \|y_u - X\beta\|^2 \leq \frac{d}{\alpha} ,$$

where $\upsilon_{\min}(X^T X) > 0$ denotes the the smallest eigenvalue of $X^T X$. Thus, for any $\theta \in S_d$, we have all the $|u_{ij}|$ contained and

$$|\beta_l - (\hat{\beta}_u)_l| \leq \sqrt{\frac{d}{\alpha \upsilon_{\min}(X^T X)}} .$$

This implies that S_d is bounded which establishes that $v(\theta)$ is unbounded off compact sets.

Román (2012) shows that the θ-chain is a Feller chain and that its maximal irreducibility measure is equivalent to Lebesgue measure on $\mathbb{R}^{p+q}$. Hence, Meyn and Tweedie's (1993) Theorem 6.0.1 implies that all compact sets in $\mathbb{R}^{p+q}$ are *petite sets* for the θ-chain. Therefore, the drift function $v(\theta)$ is *unbounded off petite sets* (Meyn and Tweedie, 1993, p.191). It now follows from Meyn and Tweedie's (1993) Lemma 15.2.8 that the geometric drift condition in Proposition 2 implies that the θ-chain is geometrically ergodic.

B Preliminary Results

B.1 A Matrix Result

Here we present a general matrix result that will be used in the proof of Lemma 1. The proof of Lemma 1 appears in Appendix B.3.

Before we state the result we briefly review some facts about non-negative definite matrices. Recall that if C is a non-negative definite matrix then $\text{tr}(C) \geq 0$. If A and B are symmetric matrices (of the same dimension) such that $B - A$ is non-negative definite, we write $A \preceq B$. Also, if $A \preceq B$ then $\text{tr}(A) \leq \text{tr}(B)$. Furthermore, if A and B are positive definite matrices, then $A \preceq B$ if and only if $B^{-1} \preceq A^{-1}$.

Lemma 3. *Suppose Ω is an $n \times n$ matrix of the form*

$$\Omega = A^T A v + \Upsilon ,$$

where υ is a positive constant, A is a non-null $m \times n$ matrix and Υ is an $n \times n$ diagonal matrix with positive diagonal elements, $\{\upsilon_i\}_{i=1}^n$. Let $O^T DO$ be the spectral decomposition of $A^T A$, so O is an n-dimensional orthogonal matrix, and D is a diagonal matrix whose diagonal elements, $\{d_i\}_{i=1}^n$, are the eigenvalues of $A^T A$. Also, let $D^\perp$ denote the n-dimensional diagonal matrix whose diagonal elements, $\{d_i^\perp\}_{i=1}^n$, are given by

$$d_i^\perp = \begin{cases} 1 & d_i = 0 \\ 0 & d_i \neq 0 \, . \end{cases}$$

Then

1. $\Omega^{-1} \preceq (A^T A)^+ \upsilon^{-1} + O^T D^\perp O \upsilon_{\min}^{-1}$
2. $tr\big(\Omega^{-1}\big) \leq tr\big((A^T A)^+\big)\upsilon^{-1} + (n - rank(A))\upsilon_{\min}^{-1}$
3. $tr\big(A\Omega^{-1}A^T\big) \leq rank(A)\upsilon^{-1}$,

where $(A^T A)^+$ denotes the Moore-Penrose inverse of $A^T A$ and $\upsilon_{\min} = \min_{1 \leq i \leq n}\{\upsilon_i\}$.

Remark 3. *Note that $rank(A) = rank(A^T A) = rank(O^T DO) = rank(D)$. Thus $rank(A)$ equals the number of non-zero eigenvalues of $A^T A$.*

Proof. It is clear that

$$A^T A\upsilon + I_n \upsilon_{\min} \preceq A^T A\upsilon + \Upsilon \, .$$

This yields

$$\begin{aligned} \Omega^{-1} &= \Big(A^T A\upsilon + \Upsilon\Big)^{-1} \\ &\preceq \Big(A^T A\upsilon + I_n\upsilon_{\min}\Big)^{-1} \\ &= O^T\Big(D\upsilon + I_n\upsilon_{\min}\Big)^{-1} O \, . \end{aligned} \tag{21}$$

Now let D^+ be a diagonal matrix whose diagonal elements, $\{d_i^+\}_{i=1}^n$, are given by

$$d_i^+ = \begin{cases} d_i^{-1} & d_i \neq 0 \\ 0 & d_i = 0 \, . \end{cases}$$

Note that, for each $i \in \{1, 2, \ldots, n\}$, we have

$$\frac{1}{d_i\upsilon + \upsilon_{\min}} \leq d_i^+ \upsilon^{-1} + I_{\{0\}}(d_i)\upsilon_{\min}^{-1} \, .$$

This shows that

$$\Big(D\upsilon + I_n\upsilon_{\min}\Big)^{-1} \preceq D^+\upsilon^{-1} + D^\perp\upsilon_{\min}^{-1} \,.$$

Together with (21), this leads to

$$\Omega^{-1} \preceq O^T\Big(D\upsilon + I_n\upsilon_{\min}\Big)^{-1}O \preceq O^T\big(D^+\upsilon^{-1} + D^\perp\upsilon_{\min}^{-1}\big)O = (A^TA)^+\upsilon^{-1} + O^TD^\perp O\upsilon_{\min}^{-1} \,,$$

which proves the first statement. The second statement follows by taking traces on both sides and the fact that $\text{tr}(D^\perp)$ equals the number of zero eigenvalues of A^TA. Pre- and post-multiplying the first statement by A and A^T, respectively, and then taking traces yields

$$\text{tr}\big(A\Omega^{-1}A^T\big) \leq \text{tr}\big(A(A^TA)^+A^T\big)\upsilon^{-1} + \text{tr}\big(AO^TD^\perp OA^T\big)\upsilon_{\min}^{-1} \,. \tag{22}$$

Since $(A^TA)(A^TA)^+$ is idempotent, we have

$$\text{tr}\big(A(A^TA)^+A^T\big) = \text{tr}\big(A^TA(A^TA)^+\big) = \text{rank}\big(A^TA(A^TA)^+\big) = \text{rank}(A^TA) = \text{rank}(A) \,.$$

Furthermore,

$$\text{tr}\big(AO^TD^\perp OA^T\big) = \text{tr}\big(O^TD^\perp OA^TA\big) = \text{tr}\big(O^TD^\perp OO^TDO\big) = \text{tr}\big(O^TD^\perp DO\big) = 0 \,,$$

where the last line follows from the fact that $D^\perp D = 0$. It now follows from (22) that

$$\text{tr}\big(A\Omega^{-1}A^T\big) \leq \text{rank}(A)\upsilon^{-1} \,,$$

and the third statement has been established. □

B.2 Matrix Decompositions

Before we provide the matrix decompositions, we introduce some notation. Let $\tilde{X} := X\Sigma_\beta^{1/2}$ and note that since X is assumed to be full rank and Σ_β is positive definite, $\tilde{X}$ also has full rank. Thus $\tilde{X}$ can be written as $\tilde{X} = USV^T$, where U and V are orthogonal matrices of dimension N and p, respectively, and S is the $N \times p$ matrix given by

$$S = \begin{pmatrix} S_* \\ \mathbf{0} \end{pmatrix} \,,$$

where S_* is a $p \times p$ diagonal matrix whose diagonal elements, $\{s_i\}_{i=1}^p$, are the (strictly positive) singular values of $\tilde{X}$. The following results provide decompositions for the matrices $XT_\lambda^{-1}X^T$ and $M_\lambda = I - \lambda_e XT_\lambda^{-1}X^T$ which will be used in the the proofs of Lemma 1 and Lemma 2.

Lemma 4. *If rank*$(X) = p$*, then* $XT_\lambda^{-1}X^T = UR_\lambda U^T$ *where* R_λ *is an* $N \times N$ *diagonal matrix whose diagonal elements,* $\{r_i\}_{i=1}^N$*, are given by*

$$r_i = \begin{cases} \frac{s_i^2}{\lambda_e s_i^2 + 1} & i \in \{1, 2, \ldots, p\} \\ 0 & i \in \{p+1, p+2, \ldots, N\} \end{cases} .$$

Moreover, $\|XT_\lambda^{-1}X^T\| \leq \|X\Sigma_\beta X^T\|$ *for all* $\lambda \in \mathbb{R}_+^{r+1}$.

Proof. Note that

$$\begin{aligned} XT_\lambda^{-1}X^T &= X\Sigma_\beta^{1/2}\Sigma_\beta^{-1/2}T_\lambda^{-1}\Sigma_\beta^{-1/2}\Sigma_\beta^{1/2}X^T \\ &= \tilde{X}\left(\lambda_e \tilde{X}^T\tilde{X} + I\right)^{-1}\tilde{X}^T \\ &= USV^T(\lambda_e VS^TSV^T + VV^T)^{-1}VS^TU^T \\ &= US(\lambda_e S^TS + I)^{-1}S^TU^T \\ &= US(\lambda_e S_*^2 + I)^{-1}S^TU^T \\ &= UR_\lambda U^T . \end{aligned}$$

Using the above decomposition, we see that

$$\|XT_\lambda^{-1}X^T\|^2 = \text{tr}\left((UR_\lambda U^T)^T UR_\lambda U^T\right) = \text{tr}\left(R_\lambda^2\right) = \sum_{i=1}^p \left(\frac{s_i^2}{\lambda_e s_i^2 + 1}\right)^2 \leq \sum_{i=1}^p s_i^4$$

but since $\tilde{X}\tilde{X}^T = USS^TU^T$ is symmetric and $\text{tr}(SS^TSS^T) = \text{tr}(S^TSS^TS) = \text{tr}(S_*^4)$, we have

$$\sum_{i=1}^p s_i^4 = \text{tr}\left(S_*^4\right) = \text{tr}(SS^TSS^T) = \text{tr}\left(USV^TVS^TU^TUSV^TVS^TU^T\right) = \|\tilde{X}\tilde{X}^T\|^2 .$$

Hence, $\|XT_\lambda^{-1}X^T\| \leq \|X\Sigma_\beta X^T\|$. □

The following result provides a decomposition for $M_\lambda = I - \lambda_e XT_\lambda^{-1}X^T$.

Lemma 5. *If rank*$(X) = p$*, then* $M_\lambda = UH_\lambda U^T$ *where* H_λ *is an* $N \times N$ *diagonal matrix whose diagonal elements,* $\{h_i\}_{i=1}^N$*, are defined as*

$$h_i = \begin{cases} \frac{1}{\lambda_e s_i^2 + 1} & i \in \{1, 2, \ldots, p\} \\ 1 & i \in \{p+1, p+2, \ldots, N\} \end{cases} .$$

Furthermore, $(\lambda_e s_{\max}^2 + 1)^{-1} I \preceq M_\lambda \preceq I$ *and* $\|M_\lambda\| \leq \sqrt{N}$ *for all* $\lambda \in \mathbb{R}_+^{r+1}$*, where* $s_{\max}$ *denotes the largest singular value of* $\tilde{X}$.

Proof. The decomposition follows immediately from Lemma 4 since

$$M_\lambda = I - \lambda_e X T_\lambda^{-1} X^T = UU^T - \lambda_e U R_\lambda U^T = U\big(I - \lambda_e R_\lambda\big)U^T \ .$$

Now for each $i = 1, 2, \ldots, N$, $0 < (\lambda_e s_{\max}^2 + 1)^{-1} \leq h_i \leq 1$, and it follows that

$$(\lambda_e s_{\max}^2 + 1)^{-1} I = U(\lambda_e s_{\max}^2 + 1)^{-1} U^T \preceq U H_\lambda U^T \preceq UU^T = I \ .$$

Finally, we have

$$\|M_\lambda\|^2 = \text{tr}\left((UH_\lambda U^T)^T UH_\lambda U^T\right) = \text{tr}\left(H_\lambda^2\right) = \sum_{i=1}^{p} \left(\frac{1}{\lambda_e s_i^2 + 1}\right)^2 + N - p \leq N \ ,$$

and this completes the proof. □

B.3 Proof of Lemma 1

Here we present a proof of Lemma 1 which uses the results given in the previous two sections.

Proof of Lemma 1. Statement 1. in Lemma 1 follows almost immediately from Lemma 3. To see this, note that by Lemma 5, we know that $(\lambda_e s_{\max}^2 + 1)^{-1} I \preceq M_\lambda \preceq I$. So it follows that

$$\lambda_e(\lambda_e s_{\max}^2 + 1)^{-1} Z^T Z + \Lambda_u \preceq \lambda_e Z^T M_\lambda Z + \Lambda_u = Q_\lambda \ ,$$

and thus

$$Q_\lambda^{-1} \preceq \left(\lambda_e(\lambda_e s_{\max}^2 + 1)^{-1} Z^T Z + \Lambda_u\right)^{-1} \ .$$

Now, an application of Lemma 3 shows that

$$\begin{aligned}
\text{tr}(Q_\lambda^{-1}) &\leq \text{tr}\left(\left(\lambda_e(\lambda_e s_{\max}^2 + 1)^{-1} Z^T Z + \Lambda_u\right)^{-1}\right) \\
&\leq \lambda_e^{-1}(\lambda_e s_{\max}^2 + 1)\text{tr}\big((Z^T Z)^+\big) + (q - \text{rank}(Z))\left(\min_{1 \leq i \leq r} \lambda_{u_i}\right)^{-1} \\
&\leq \lambda_e^{-1}\text{tr}\big((Z^T Z)^+\big) + s_{\max}^2 \text{tr}\big((Z^T Z)^+\big) + (q - \text{rank}(Z)) \sum_{i=1}^{r} \lambda_{u_i}^{-1} \ .
\end{aligned}$$

This establishes statement 1. in Lemma 1.

Another application of Lemma 3 shows that

$$\text{tr}\big(ZQ_\lambda^{-1} Z^T\big) \leq \lambda_e^{-1}\text{rank}(Z) + s_{\max}^2 \text{rank}(Z) \ . \tag{23}$$

Also, note that by Lemma 5, we have

$$0 \preceq U(I - H_\lambda^2)U^T = I - M_\lambda^2 = (I + M_\lambda)(I - M_\lambda) \ ,$$

and since $A_\lambda := (I + M_\lambda)(I - M_\lambda)$ and Q_λ both have square roots, we have

$$\begin{aligned} \mathrm{tr}\big((I - M_\lambda)ZQ_\lambda^{-1}Z^T(I + M_\lambda)\big) &= \mathrm{tr}\big(A_\lambda^{1/2}A_\lambda^{1/2}ZQ_\lambda^{-1/2}Q_\lambda^{-1/2}Z^T\big) \\ &= \mathrm{tr}\big(Q_\lambda^{-1/2}Z^TA_\lambda^{1/2}A_\lambda^{1/2}ZQ_\lambda^{-1/2}\big) \\ &\geq 0 \ . \end{aligned}$$

Finally, since $\Sigma_\beta^{-1} \preceq T_\lambda$ we must have $T_\lambda^{-1} \preceq \Sigma_\beta$ which implies that $\mathrm{tr}(XT_\lambda^{-1}X^T) \leq \mathrm{tr}(X\Sigma_\beta X^T)$. Hence,

$$\mathrm{tr}(XT_\lambda^{-1}X^T) - \mathrm{tr}\big((I - M_\lambda)ZQ_\lambda^{-1}Z^T(I + M_\lambda)\big) \leq \mathrm{tr}(X\Sigma_\beta X^T) \ ,$$

and combining this with (13) and (23), we see that statement 2. holds. □

B.4 Proof of Lemma 2

Recall the decompositions for $\tilde{X} = X\Sigma_\beta^{1/2}$ given in Section B.2 and define $\tilde{Z} = U^TZ$, $\tilde{y} = U^Ty$ and $y^* = U^TX(X^TX)^{-1}\Sigma_\beta^{-1}\mu_\beta$. Also, let $\tilde{z}_i$ denote the ith column of $\tilde{Z}^T$, and let $\tilde{y}_i$ and y_i^* denote the ith component of the vectors $\tilde{y}$ and y^*, respectively.

The constants K_y and K_{μ_β} that appear in statement of Lemma 2 (and in the expression for $K(\alpha)$) are

$$K_y := \sum_{i=1}^N |\tilde{y}_i| \left[\sup_{c \in \mathbb{R}_+^{N+q}} t_i^T \left(t_i t_i^T + \sum_{j \in \{1,2,\ldots,N+q\} \setminus \{i\}} c_j t_j t_j^T + c_i I \right)^{-2} t_i \right]^{1/2} ,$$

and

$$K_{\mu_\beta} := \sum_{i=1}^p |y_i^*| \left[\sup_{c \in \mathbb{R}_+^{N+q}} t_i^T \left(s_i^{-2} t_i t_i^T + \sum_{j \in \{1,2,\ldots,N+q\} \setminus \{i\}} c_j t_j t_j^T + c_i I \right)^{-2} t_i \right]^{1/2} .$$

For $j = 1, 2, \ldots, N$, $t_j = \tilde{z}_j$, and for $j \in \{N+1, \ldots, N+q\}$, the t_j are the standard orthonormal basis vectors in $\mathbb{R}^q$; that is, t_{N+l} has a one in the lth position and zeros everywhere else.

The following result from Khare and Hobert (2011) will be used in the proof of Lemma 2.

Lemma 6. *Fix $n \in \{2, 3, \ldots\}$ and $m \in \mathbb{N}$, and let $x_1, \ldots, x_n$ be vectors in $\mathbb{R}^m$. Then*

$$C_{m,n}(x_1; x_2, \ldots, x_n) := \sup_{w \in \mathbb{R}_+^n} x_1^T \Big(x_1 x_1^T + \sum_{i=2}^n w_i x_i x_i^T + w_1 I \Big)^{-2} x_1$$

is finite.

Proof of Lemma 2. We will show that

$$\|\lambda_e Q_\lambda^{-1} Z^T M_\lambda \, y\| \leq K_y < \infty \quad \text{and} \quad \|\lambda_e Q_\lambda^{-1} Z^T X T_\lambda^{-1} \Sigma_\beta^{-1}\| \leq K_{\mu_\beta} < \infty \, .$$

Using Lemma 5, we have

$$\begin{aligned}
\|\lambda_e Q_\lambda^{-1} Z^T M_\lambda \, y\| &= \|\left(Z^T M_\lambda Z + \lambda_e^{-1}\Lambda_u\right)^{-1} Z^T M_\lambda \, y\| \\
&= \|\left(Z^T U H_\lambda U^T Z + \lambda_e^{-1}\Lambda_u\right)^{-1} Z^T U H_\lambda U^T \, y\| \\
&= \left\| \sum_{i=1}^N \left(\tilde{Z}^T H_\lambda \tilde{Z} + \lambda_e^{-1}\Lambda_u\right)^{-1} \tilde{z}_i h_i \tilde{y}_i \right\| \\
&\leq \sum_{i=1}^N \left\| \left(\tilde{Z}^T H_\lambda \tilde{Z} + \lambda_e^{-1}\Lambda_u\right)^{-1} \tilde{z}_i h_i \tilde{y}_i \right\| \\
&= \sum_{i=1}^N \left\| \left(\sum_{j=1}^N \tilde{z}_j \tilde{z}_j^T h_j + \lambda_e^{-1}\Lambda_u \right)^{-1} \tilde{z}_i h_i \tilde{y}_i \right\| \\
&= \sum_{i=1}^N |\tilde{y}_i| \left\| \left(\tilde{z}_i \tilde{z}_i^T + \sum_{j \in \{1,2,\ldots,N\}\setminus\{i\}} \tilde{z}_j \tilde{z}_j^T \frac{h_j}{h_i} + h_i^{-1}\lambda_e^{-1}\Lambda_u \right)^{-1} \tilde{z}_i \right\| .
\end{aligned}$$

We now focus on bounding

$$C_i(\lambda) := \left\| \left(\tilde{z}_i \tilde{z}_i^T + \sum_{j \in \{1,2,\ldots,N\}\setminus\{i\}} \tilde{z}_j \tilde{z}_j^T \frac{h_j}{h_i} + h_i^{-1}\lambda_e^{-1}\Lambda_u \right)^{-1} \tilde{z}_i \right\| .$$

Define $\lambda_\bullet = \sum_{i=1}^r \lambda_{u_i}^{-1}$ and note that

$$\begin{aligned}
C_i^2(\lambda) &= \tilde{z}_i^T \left(\tilde{z}_i \tilde{z}_i^T + \sum_{j \in \{1,2,\ldots,N\}\setminus\{i\}} \tilde{z}_j \tilde{z}_j^T \frac{h_j}{h_i} + h_i^{-1}\lambda_e^{-1}\Lambda_u \right)^{-2} \tilde{z}_i \\
&= \tilde{z}_i^T \left(\tilde{z}_i \tilde{z}_i^T + \sum_{j \in \{1,2,\ldots,N\}\setminus\{i\}} \tilde{z}_j \tilde{z}_j^T \frac{h_j}{h_i} + h_i^{-1}\lambda_e^{-1}\left(\Lambda_u - \lambda_\bullet^{-1} I\right) + \frac{\lambda_\bullet^{-1}}{h_i \lambda_e} I \right)^{-2} \tilde{z}_i \\
&\leq \sup_{c \in \mathbb{R}_+^{N+q}} t_i^T \left(t_i t_i^T + \sum_{j \in \{1,2,\ldots,N\}\setminus\{i\}} c_j t_j t_j^T + \sum_{j=N+1}^{N+q} c_j t_j t_j^T + c_i I \right)^{-2} t_i \, .
\end{aligned}$$

Applications of Lemma 6 show that each $C_i^2(\lambda)$ is finite. Hence, $\|\lambda_e Q_\lambda^{-1} Z^T M_\lambda y\| \leq K_y$ and K_y is finite.

Recall that $y^* = U^T X (X^T X)^{-1} \Sigma_\beta^{-1} \mu_\beta$ and y_i^* denotes the ith component of y^*. Then using

Lemma 4 we have

$$\begin{aligned}\|\lambda_e Q_\lambda^{-1} Z^T X T_\lambda^{-1} \Sigma_\beta^{-1} \mu_\beta\| &= \|\lambda_e Q_\lambda^{-1} Z^T X T_\lambda^{-1} X^T X (X^T X)^{-1} \Sigma_\beta^{-1} \mu_\beta\| \\ &= \|\lambda_e Q_\lambda^{-1} Z^T U R_\lambda U^T X (X^T X)^{-1} \Sigma_\beta^{-1} \mu_\beta\| \\ &= \|\lambda_e Q_\lambda^{-1} \tilde{Z}^T R_\lambda \, y^*\| \, .\end{aligned}$$

Recall Lemmas 4 and 5, and note that when $i \in \{1, 2, \ldots, p\}$, we have $h_i/r_i = s_i^{-2}$ while for $i \in \{p+1, p+2, \ldots, N\}$, $r_i = 0$. Hence,

$$\begin{aligned}\|\lambda_e Q_\lambda^{-1} Z^T X T_\lambda^{-1} \Sigma_\beta^{-1}\| &= \|\lambda_e Q_\lambda^{-1} \tilde{Z}^T R_\lambda \, y^*\| \\ &= \|\left(\tilde{Z}^T H_\lambda \tilde{Z} + \lambda_e^{-1} \Lambda_u\right)^{-1} \tilde{Z}^T R_\lambda \, y^*\| \\ &= \left\| \sum_{i=1}^{N} \left(\tilde{Z}^T H_\lambda \tilde{Z} + \lambda_e^{-1} \Lambda_u\right)^{-1} \tilde{z}_i r_i y_i^* \right\| \\ &= \left\| \sum_{i=1}^{p} \left(\tilde{Z}^T H_\lambda \tilde{Z} + \lambda_e^{-1} \Lambda_u\right)^{-1} \tilde{z}_i r_i y_i^* \right\| \\ &\leq \sum_{i=1}^{p} \left\| \left(\tilde{Z}^T H_\lambda \tilde{Z} + \lambda_e^{-1} \Lambda_u\right)^{-1} \tilde{z}_i r_i y_i^* \right\| \\ &= \sum_{i=1}^{p} \left\| \left(\sum_{j=1}^{N} \tilde{z}_j \tilde{z}_j^T h_j + \lambda_e^{-1} \Lambda_u \right)^{-1} \tilde{z}_i r_i y_i^* \right\| \\ &= \sum_{i=1}^{p} |y_i^*| s_i \left\| \left(\tilde{z}_i \tilde{z}_i^T s_i^{-2} + \sum_{j \in \{1,2,\ldots,N\} \setminus \{i\}} \tilde{z}_j \tilde{z}_j^T \frac{h_j}{r_i} + r_i^{-1} \lambda_e^{-1} \Lambda_u \right)^{-1} \left(s_i^{-1} \tilde{z}_i\right) \right\| .\end{aligned}$$

Applications of Lemma 6 show that, for each $i \in \{1, 2, \ldots, p\}$,

$$\left\| \left(\tilde{z}_i \tilde{z}_i^T s_i^{-2} + \sum_{j \in \{1,2,\ldots,N\} \setminus \{i\}} \tilde{z}_j \tilde{z}_j^T \frac{h_j}{r_i} + r_i^{-1} \lambda_e^{-1} \Lambda_u \right)^{-1} \left(s_i^{-1} \tilde{z}_i\right) \right\|^2$$

is bounded above by the finite constant

$$\sup_{c \in \mathbb{R}_+^{N+q}} (s_i^{-1} t_i)^T \left(s_i^{-2} t_i t_i^T + \sum_{j \in \{1,2,\ldots,N\} \setminus \{i\}} c_j t_j t_j^T + \sum_{j=N+1}^{N+q} c_j t_j t_j^T + c_i I \right)^{-2} (s^{-1} t_i) \, .$$

Hence, $\|\lambda_e Q_\lambda^{-1} Z^T X T_\lambda^{-1} \Sigma_\beta^{-1}\| \leq K_{\mu_\beta}$ and K_{μ_β} is finite. □

References

BEDNORZ, W. and ŁATUSZYŃSKI, K. (2007). A few remarks on "Fixed-width output analysis for Markov chain Monte Carlo" by Jones *et al*. *Journal of the American Statistical Association*, **102** 1485–1486.

DIACONIS, P., KHARE, K. and SALOFF-COSTE, L. (2008). Gibbs sampling, exponential families and orthogonal polynomials (with discussion). *Statistical Science*, **23** 151–200.

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.

GELFAND, A. E., SAHU, S. K. and CARLIN, B. P. (1995). Efficient parametrisations for normal linear mixed models. *Biometrika*, **82** 479–488.

HOBERT, J. P. and CASELLA, G. (1996). The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *Journal of the American Statistical Association*, **91** 1461–1473.

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.

HOBERT, J. P., JONES, G. L., PRESNELL, B. and ROSENTHAL, J. S. (2002). On the applicability of regenerative simulation in Markov chain Monte Carlo. *Biometrika*, **89** 731–743.

JOHNSON, A. A. and JONES, G. L. (2010). Gibbs sampling for a Bayesian hierarchical general linear model. *Electronic Journal of Statistics*, **4** 313–333.

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.

JONES, G. L. and HOBERT, J. P. (2004). Sufficient burn in for Gibbs samplers for a hierarchical random effects model. *Annals of Statistics*, **32** 784–817.

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

MEYN, S. P. and TWEEDIE, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer Verlag, London.

PAPASPILIOPOULOS, O. and ROBERTS, G. O. (2008). Stability of the Gibbs sampler for Bayesian hierarchical models. *Annals of Statistics*, **36** 95–117.

PAPASPILIOPOULOS, O., ROBERTS, G. O. and SKÖLD, M. (2007). A general framework for the parametrization of hierarchical models. *Statistical Science*, **22** 59–73.

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.

ROBERTS, G. O. and ROSENTHAL, J. S. (2001). Markov chains and de-initializing processes. *Scandinavian Journal of Statistics*, **28** 489–504.

ROBERTS, G. O. and ROSENTHAL, J. S. (2004). General state space Markov chains and MCMC algorithms. *Probability Surveys*, **1** 20–71.

ROMÁN, J. C. (2012). *Convergence Analysis of Block Gibbs Samplers for Bayesian General Linear Mixed Models*. Ph.D. thesis, Department of Statistics, University of Florida.

ROMÁN, 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.

ROMÁN, J. C., HOBERT, J. P. and PRESNELL, B. (2013). On reparametrization and the Gibbs sampler. Tech. rep., Vanderbilt University.

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.