Bernoulli **23**(1), 2017, 459–478
DOI: 10.3150/15-BEJ749

Convergence analysis of block Gibbs samplers for Bayesian linear mixed models with $p > N$

TAVIS ABRAHAMSEN* and JAMES P. HOBERT**

Department of Statistics, University of Florida, Gainesville, Florida 32611, USA.
E-mail: *tabrahamsen@stat.ufl.edu; **jhobert@stat.ufl.edu

Exploration of the intractable posterior distributions associated with Bayesian versions of the general linear mixed model is often performed using Markov chain Monte Carlo. In particular, if a conditionally conjugate prior is used, then there is a simple two-block Gibbs sampler available. Román and Hobert [*Linear Algebra Appl.* **473** (2015) 54–77] showed that, when the priors are proper and the X matrix has full column rank, the Markov chains underlying these Gibbs samplers are nearly always geometrically ergodic. In this paper, Román and Hobert's (2015) result is extended by allowing improper priors on the variance components, and, more importantly, by removing all assumptions on the X matrix. So, not only is X allowed to be (column) rank deficient, which provides additional flexibility in parameterizing the fixed effects, it is also allowed to have more columns than rows, which is necessary in the increasingly important situation where $p > N$. The full rank assumption on X is at the heart of Román and Hobert's (2015) proof. Consequently, the extension to unrestricted X requires a substantially different analysis.

Keywords: conditionally conjugate prior; convergence rate; geometric drift condition; Markov chain; matrix inequality; Monte Carlo

1. Introduction

The general linear mixed model (GLMM) is one of the most frequently applied statistical models. A GLMM with r random factors takes the form

$$Y = X\beta + \sum_{i=1}^{r} Z_i u_i + e,$$

where Y is an observable $N \times 1$ data vector, X and $\{Z_i\}_{i=1}^r$ are known matrices, β is an unknown $p \times 1$ vector of regression coefficients, $\{u_i\}_{i=1}^r$ are independent random vectors whose elements represent the various levels of the random factors in the model, and $e \sim \mathrm{N}_N(0, \lambda_e^{-1} I)$. Assume that e and $u := (u_1^T \;\; u_2^T \;\; \cdots \;\; u_r^T)^T$ are independent, and that $u \sim \mathrm{N}_q(0, \Lambda^{-1})$, where u_i is $q_i \times 1$, $q = q_1 + \cdots + q_r$, and $\Lambda = \bigoplus_{i=1}^r \lambda_{u_i} I_{q_i}$. Letting $Z = (Z_1 \;\; Z_2 \;\; \cdots \;\; Z_r)$, we can write $\sum_{i=1}^r Z_i u_i = Zu$. Let λ denote the vector of precision parameters, i.e., $\lambda = (\lambda_e \;\; \lambda_{u_1} \;\; \cdots \;\; \lambda_{u_r})^T$. To rule out degenerate cases, we assume throughout that $N \geq 2$, and that $q_i \geq 2$ for each $i = 1, 2, \ldots, r$. For a book-length treatment of the GLMM, which is sometimes called the *variance components model*, see [15].

In the Bayesian setting, prior distributions are assigned to the unknown parameters, β and λ. Unfortunately, the Bayes estimators associated with any non-trivial prior cannot be obtained in closed form. This is because such estimators take the form of ratios of high-dimensional, intractable integrals. The dimensionality also precludes the use of classical Monte Carlo methods that require the ability to draw samples directly from the posterior distribution. Instead, the parameter estimates are typically obtained using Markov chain Monte Carlo (MCMC) methods. In particular, when (proper or improper) conditionally conjugate priors are adopted for β and λ, there is a simple block Gibbs sampler that can be used to explore the intractable posterior density. Let $\theta = (\beta^T \;\; u^T)^T$, and denote the posterior density as $\pi(\theta, \lambda|y)$, where y denotes the observed data vector. (Since u is unobservable, it is treated like a parameter.) When the conditionally conjugate priors are adopted, it is straightforward to simulate from $\theta|\lambda, y$, and from $\lambda|\theta, y$. Indeed, $\theta|\lambda, y$ is multivariate normal and, given (θ, y), the components of λ are independent gamma variates. Hence, it is straightforward to simulate a Markov chain, $\{(\theta_n, \lambda_n)\}_{n=0}^{\infty}$, that has $\pi(\theta, \lambda|y)$ as its invariant density. Our main results concern the convergence properties of this block Gibbs sampler. We now provide some background about Markov chains on $\mathbb{R}^d$, which will allow us to describe our results and their practical importance.

Let $V = \{V_m\}_{m=0}^{\infty}$ denote a Markov chain with state space $\mathsf{V} \subset \mathbb{R}^d$ and assume the chain is Harris ergodic; that is, ψ-irreducible, aperiodic and positive Harris recurrent (see [8] for definitions). Assume further that the chain has a Markov transition density (with respect to Lebesgue measure), $k : \mathsf{V} \times \mathsf{V} \to [0, \infty)$. Then, for any measurable set A, we have

$$\Pr(V_{m+1} \in A | V_m = v) = \int_A k(v'|v)\, dv'.$$

For $m \in \{2, 3, 4, \ldots\}$, the m-step Markov transition density (Mtd) is defined inductively as follows

$$k^m(v'|v) = \int_{\mathsf{V}} k^{m-1}(v'|u) k(u|v)\, du.$$

Of course, $k^1 \equiv k$, and $k^m(\cdot|v)$ is the density of V_m conditional on $V_0 = v$. Suppose that the invariant probability distribution also has a density (with respect to Lebesgue measure), $\kappa : \mathsf{V} \to [0, \infty)$. The chain V is *geometrically ergodic* if there exist $M : \mathsf{V} \to [0, \infty)$ and $\gamma \in [0, 1)$ such that, for all $m \in \mathbb{N}$,

$$\int_{\mathsf{V}} \left| k^m(v'|v) - \kappa(v') \right| dv' \le M(v) \gamma^m. \tag{1}$$

Of course, the quantity on the left-hand side of (1) is the total variation distance between the invariant distribution and the distribution of V_m given $V_0 = v$.

There are many important practical and theoretical benefits of using a geometrically ergodic Markov chain as the basis of one's MCMC algorithm (see, e.g., [3,6,9]). Perhaps the most important of these is the ability to construct valid asymptotic standard errors for MCMC-based estimators. Let $h : \mathsf{V} \to \mathbb{R}$ be a function such that $\int_{\mathsf{V}} |h(v)| \kappa(v)\, dv < \infty$, and suppose that the chain V is to serve as the basis of an MCMC algorithm for estimating $\omega := \int_{\mathsf{V}} h(v) \kappa(v)\, dv < \infty$. Harris ergodicity guarantees that the standard estimator of ω, $\overline{h}_m := \frac{1}{m} \sum_{i=0}^{m-1} h(V_i)$, is strongly consistent. However, Harris ergodicity is *not* enough to ensure that $\overline{h}_m$ satisfies a central limit theorem (CLT). On the other hand, if V is geometrically ergodic and there exists an $\varepsilon > 0$ such that $\int_{\mathsf{V}} |h(v)|^{2+\varepsilon} \kappa(v)\, dv < \infty$, then $\overline{h}_m$ does indeed satisfy a $\sqrt{m}$-CLT; that is, under these conditions, there exists a positive, finite σ^2 such that, as $m \to \infty$, $\sqrt{m}(\overline{h}_m - \omega) \xrightarrow{d} \mathrm{N}(0, \sigma^2)$. This is extremely important from a practical standpoint because all of the standard methods of calculating valid asymptotic standard errors for $\overline{h}_m$ are based on the existence of this CLT (Flegal, Haran and Jones [3]).

There have been several studies of the convergence properties of the block Gibbs sampler for the GLMM (Johnson and Jones [5], Román and Hobert [12,13]). These have resulted in easily-checked sufficient conditions for geometric ergodicity of the underlying Markov chain. However, in all of the studies to date, the matrix X has been assumed to have full column rank. In this paper, we extend the results to the case where X is completely unrestricted. So, not only do we allow for a rank deficient X, which provides additional flexibility in parameterizing the fixed effects, we also allow for X with $p > N$, which is necessary in the increasingly important situation where there are more predictors than data points. Two different families of conditionally conjugate priors are considered, one proper and one improper. We now describe our results, beginning with the results for proper priors.

Assume that β and the components of λ are all *a priori* independent, and that $\beta \sim \mathrm{N}_p(\mu_\beta, \Sigma_\beta)$, $\lambda_e \sim \mathrm{Gamma}(a_0, b_0)$ and, for $i = 1, \ldots, r$, $\lambda_{u_i} \sim \mathrm{Gamma}(a_i, b_i)$. Our result for proper priors, which is a corollary of Proposition 1 from Section 3, is as follows.

Corollary 1. *Under a proper prior, the block Gibbs Markov chain*, $\{(\lambda_n, \theta_n)\}_{n=0}^{\infty}$, *is geometrically ergodic if*:

1. $a_0 > \frac{1}{2}(\mathrm{rank}(Z) - N + 2)$, *and*
2. $\min\{a_1 + \frac{q_1}{2}, \ldots, a_r + \frac{q_r}{2}\} > \frac{1}{2}(q - \mathrm{rank}(Z)) + 1$.

The conditions of Corollary 1 are quite weak in the sense that they would nearly always be satisfied in practice. Indeed, it would typically be the case that $\mathrm{rank}(Z) - N < -2$ (making the first condition vacuous) and $q - \mathrm{rank}(Z)$ is close to zero (making the second condition easily satisfied). In fact, if $q = \mathrm{rank}(Z)$, which is the case for many standard designs, then the second condition is also vacuous.

Román and Hobert [13] (hereafter R&H15) proved this same result under the restrictive assumption that X has full column rank. Moreover, the rank assumption is at the very heart of their proof. Indeed, these authors established a geometric drift condition for the marginal chain, $\{\theta_n\}_{n=0}^{\infty}$, but their drift (Lyapunov) function is *only valid when X has full column rank*. Our proof is significantly different. We analyze the other marginal chain, $\{\lambda_n\}_{n=0}^{\infty}$, using a drift function that does not involve the matrix X. Generally speaking, minor changes in a drift function often lead to significant differences in what one is able to prove. Thus, it is somewhat surprising that we are able to recover *exactly* the conditions of R&H15. To be fair, we are able to use several of their matrix bounds, but only *after* extending them to the case where X is unrestricted.

When X does not have full column rank, a flat prior on β leads to an improper posterior. Thus, the improper priors that we consider are actually partially proper. In particular, assume again that β and the components of λ are all *a priori* independent, and that $\beta \sim \mathrm{N}_p(\mu_\beta, \Sigma_\beta)$. But now take the prior on λ_e to be (proportional to) $\lambda_e^{a_0-1} e^{-b_0\lambda_e} I_{(0,\infty)}(\lambda_e)$, and for $i = 1, \ldots, r$, take the prior on λ_{u_i} to be $\lambda_{u_i}^{a_i-1} e^{-b_i \lambda_{u_i}} I_{(0,\infty)}(\lambda_{u_i})$. Assume that $\min\{a_i, b_i\} \le 0$ for at least one $i = 0, 1, \dots, r$; otherwise, we are back to the proper priors described above. (See [4] for a comprehensive discussion about improper priors for variance components.) Let $W = (X \;\; Z)$, so that $W\theta = X\beta + Zu$, and define $\text{SSE} = \|y - W\hat{\theta}\|^2$, where $\hat{\theta} = (W^T W)^+ W^T y$ and superscript "+" on a matrix denotes Moore–Penrose inverse. Our result for improper priors, which is another corollary of Proposition 1, is as follows.

Corollary 2. *Under an improper prior, the block Gibbs Markov chain, $\{(\lambda_n, \theta_n)\}_{n=0}^{\infty}$, is geometrically ergodic if:*

1. $2b_0 + \text{SSE} > 0$,
2. *For each $i \in \{1, 2, \dots, r\}$, either $b_i > 0$ or $a_i < b_i = 0$,*
3. $a_0 > \frac{1}{2}(\text{rank}(Z) - N + 2)$, *and*
4. $\min\{a_1 + \frac{q_1}{2}, \dots, a_r + \frac{q_r}{2}\} > \frac{1}{2}(q - \text{rank}(Z)) + 1$.

Note that the two conditions of Corollary 1 are exactly the same as the third and fourth conditions of Corollary 2. Furthermore, the first two conditions of Corollary 2 are *necessary* for posterior propriety [16], and hence for geometric ergodicity. Consequently, the commentary above regarding the weakness of the conditions of Corollary 1 applies here as well.

Corollary 2 is the first convergence rate result for the block Gibbs sampler for this set of partially proper priors. Román and Hobert [12] (hereafter R&H12) proved a similar result (see their Corollary 1) for a different family of improper priors in which our proper multivariate normal prior on β is replaced by a flat prior. Of course, because they used a flat prior on β, their results are only relevant in the case where X has full column rank.

The remainder of this paper is organized as follows. A formal definition of the block Gibbs Markov chain is given in Section 2. Section 3 contains our convergence rate analysis of the block Gibbs sampler under proper and improper priors. A short discussion concerning an alternative result for proper priors appears in Section 4. Some technical details are relegated to an Appendix.

2. The block Gibbs sampler

The block Gibbs sampler is driven by the Markov chain $\{(\theta_n, \lambda_n)\}_{n=0}^{\infty}$, which lives on the space $\mathbb{R}^{p+q} \times \mathbb{R}_+^{r+1}$, where $\mathbb{R}_+ := (0, \infty)$. The Markov transition density (of the version that updates θ first) is given by

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

We will often suppress dependence on y, as we have in the Markov transition density. The conditional densities, $\pi(\lambda|\theta, y)$ and $\pi(\theta|\lambda, y)$, are now described. The following formulas hold for both sets of priors (proper and improper). The components of λ are conditionally independent given θ, and we have

$$\lambda_e | \theta \sim \text{Gamma}\left(a_0 + \frac{N}{2}, b_0 + \frac{\|y - W\theta\|^2}{2}\right), \tag{2}$$

and, for $i = 1, \dots, r$,

$$\lambda_{u_i} | \theta \sim \text{Gamma}\left(a_i + \frac{q_i}{2}, b_i + \frac{\|u_i\|^2}{2} \right). \tag{3}$$

When considering improper priors, we assume that these conditional distributions are all well defined. In other words, we assume that $\{a_i\}_{i=0}^r$ and $\{b_i\}_{i=0}^r$ are such that all of the shape and rate parameters in the gamma distributions above are strictly positive. Of course, this is not enough to guarantee posterior propriety. However, the drift technique that we employ is equally applicable to positive recurrent (proper posterior) and non-positive recurrent (improper posterior) Markov chains [11]. Furthermore, geometrically ergodic chains are necessarily positive recurrent, so any Gibbs Markov chain that we conclude is geometrically ergodic, necessarily corresponds to a proper posterior. Consequently, there is no need to check for posterior propriety before proceeding with the convergence analysis.

Now define $T_\lambda = \lambda_e X^T X + \Sigma_\beta^{-1}$, $M_\lambda = I - \lambda_e X T_\lambda^{-1} X^T$, and $Q_\lambda = \lambda_e Z^T M_\lambda Z + \Lambda$. Conditional on λ, θ is multivariate normal with mean

$$\text{E}[\theta|\lambda] = \begin{bmatrix} T_\lambda^{-1}\left(\lambda_e X^T y + \Sigma_\beta^{-1}\mu_\beta\right) - \lambda_e^2 T_\lambda^{-1} X^T Z Q_\lambda^{-1} Z^T \left(M_\lambda y - X T_\lambda^{-1} \Sigma_\beta^{-1} \mu_\beta\right) \\ \lambda_e Q_\lambda^{-1} Z^T \left(M_\lambda y - X T_\lambda^{-1} \Sigma_\beta^{-1} \mu_\beta\right) \end{bmatrix}, \tag{4}$$

and covariance matrix

$$\text{Var}[\theta|\lambda] = \begin{bmatrix} T_\lambda^{-1} + \lambda_e^2 T_\lambda^{-1} X^T Z Q_\lambda^{-1} Z^T X T_\lambda^{-1} & -\lambda_e T_\lambda^{-1} X^T Z Q_\lambda^{-1} \\ -\lambda_e Q_\lambda^{-1} Z^T X T_\lambda^{-1} & Q_\lambda^{-1} \end{bmatrix}. \tag{5}$$

(A derivation of these conditionals can be found in [1].)

The two marginal sequences, $\{\theta_n\}_{n=0}^\infty$ and $\{\lambda_n\}_{n=0}^\infty$, are themselves Markov chains, and it is easy to establish that (when the posterior is proper) all three chains are Harris ergodic. Moreover, geometric ergodicity is a solidarity property for these three chains, that is, either all three chains are geometrically ergodic, or none of them is (see, e.g., [2,10,14]). Again, in contrast with R&H15, who analyzed the θ-chain, $\{\theta_n\}_{n=0}^\infty$, we establish our results by analyzing the λ-chain, $\{\lambda_n\}_{n=0}^\infty$. The Mtd of the λ-chain is given by

$$k_l(\tilde{\lambda}|\lambda) = \int_{\mathbb{R}^{p+q}} \pi(\tilde{\lambda}|\theta, y)\pi(\theta|\lambda, y)\, d\theta.$$

R&H12 also analyzed the λ-chain, and their analysis serves as a road map for ours. In fact, we use the same drift function as R&H12.

3. Convergence analysis of the block Gibbs sampler

In order to state our main result, we require a couple of definitions. For $i = 1, 2, \dots, r$, let R_i be the $q_i \times q$ matrix defined as $R_i = [0_{q_i \times q_1} \cdots 0_{q_i \times q_{i-1}} \;\; I_{q_i \times q_i} \;\; 0_{q_i \times q_{i+1}} \cdots 0_{q_i \times q_r}]$. Note that

$u_i = R_i u$. Let $P_{Z^T Z}$ denote the orthogonal projection onto the column space of $Z^T Z$. Finally, define

$$\tilde{s} = \min\left\{a_0 + \frac{N}{2}, a_1 + \frac{q_1}{2}, \dots, a_r + \frac{q_r}{2}\right\}.$$

The following result holds for both sets of priors (proper and improper).

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

1. $\tilde{s} > 0$;
2. $2b_0 + SSE > 0$;
3. *For each $i \in \{1, 2, \dots, r\}$, either $b_i > 0$ or $a_i < b_i = 0$; and*
4. *There exists $s \in (0, 1] \cap (0, \tilde{s})$ such that*

$$\max\left\{\frac{\Gamma(a_0 + N/2 - s)}{\Gamma(a_0 + N/2)}\left(\frac{\operatorname{rank}(Z)}{2}\right)^s, \sum_{i=1}^{r} \frac{\Gamma(a_i + q_i/2 - s)}{\Gamma(a_i + q_i/2)}\left(\frac{\operatorname{tr}(R_i(I - P_{Z^T Z})R_i^T)}{2}\right)^s\right\} < 1. \tag{6}$$

Remark 1. When the prior is proper, that is, when $a_i > 0$ and $b_i > 0$ for all $i \in \{0, 1, \dots, r\}$, the first three conditions are automatically satisfied, and $\tilde{s} > 1$. On the other hand, when the prior is improper, these three conditions ensure that $\pi(\lambda|\theta, y)$ is well defined.

Before embarking on our proof of Proposition 1, we quickly demonstrate that Corollaries 1 and 2 follow immediately from it.

Proof of Corollary 1. Since the prior is proper, it is enough to show that the conditions of Corollary 1 imply that (6) is satisfied for some $s \in (0, 1]$. We show that this is indeed the case, with $s = 1$. First,

$$\frac{\Gamma(a_0 + N/2 - 1)}{\Gamma(a_0 + N/2)} \frac{\operatorname{rank}(Z)}{2} = \frac{\operatorname{rank}(Z)}{2a_0 + N - 2},$$

which yields the first half of (6). Now note that

$$\sum_{i=1}^{r} \operatorname{tr}\big(R_i(I - P_{Z^T Z})R_i^T\big) = \operatorname{tr}\left[(I - P_{Z^T Z})\sum_{i=1}^{r} R_i^T R_i\right] = \operatorname{tr}(I - P_{Z^T Z}) = q - \operatorname{rank}(Z).$$

Thus,

$$\sum_{i=1}^{r} \frac{\Gamma(a_i + q_i/2 - 1)}{\Gamma(a_i + q_i/2)} \frac{\operatorname{tr}(R_i(I - P_{Z^T Z})R_i^T)}{2} = \sum_{i=1}^{r} \frac{\operatorname{tr}(R_i(I - P_{Z^T Z})R_i^T)}{2a_i + q_i - 2}$$
$$\leq \frac{q - \operatorname{rank}(Z)}{\min_{i=1,\dots,r}(2a_i + q_i - 2)},$$

which yields the second half of (6). □

Proof of Corollary 2. First note that conditions 3 and 4 of Corollary 2 imply that $\tilde{s} > 1$. The rest of the proof is the same as the proof of Corollary 1. □

Our proof of Proposition 1 is based on four lemmas, which are proven in the Appendix. Let $d_{\max}$ denote the largest singular value of the matrix $\tilde{X} = X\Sigma_\beta^{1/2}$.

Lemma 1. *For each $i \in \{1, 2, \dots, r\}$, we have*

$$\operatorname{tr}\big(R_i Q_\lambda^{-1} R_i^T\big) \le \big(d_{\max}^2 + \lambda_e^{-1}\big)\operatorname{tr}\big(R_i\big(Z^T Z\big)^+ R_i^T\big) + \operatorname{tr}\big(R_i(I - P_{Z^T Z})R_i^T\big)\sum_{j=1}^r \lambda_{u_j}^{-1}.$$

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

Lemma 3. *There exist finite constants K_1 and K_2, not depending on λ, such that $\|\mathrm{E}[R_i u|\lambda]\| \le \sqrt{q_i}K_1$ for $i = 1, \dots, r$, and $\|y - W\mathrm{E}[\theta|\lambda]\| \le K_2$.*

Remark 2. The constants K_1 and K_2 are defined in the Appendix. They do not have a closed form.

We will write $A \preceq B$ to mean that $B - A$ is nonnegative definite. Let $\psi_{\max}$ denote the largest eigenvalue of $Z^T Z$.

Lemma 4. *For each $i \in \{1, 2, \dots, r\}$, we have $(\psi_{\max}\lambda_e + \lambda_{u_i})^{-1} I_{q_i} \preceq R_i Q_\lambda^{-1} R_i^T$.*

Proof of Proposition 1. Define the drift function as follows

$$v(\lambda) = \alpha\lambda_e^c + \alpha\lambda_e^{-s} + \sum_{i=1}^r \lambda_{u_i}^c + \sum_{i=1}^r \lambda_{u_i}^{-s},$$

where α and c are positive constants (that are explicitly constructed in the proof), and s is from the fourth condition in Proposition 1. We will show that there exist $\rho \in [0, 1)$ and a finite constant L such that

$$\mathrm{E}\big[v(\tilde{\lambda})|\lambda\big] = \int_{\mathbb{R}_+^{r+1}} v(\tilde{\lambda}) k_l(\tilde{\lambda}|\lambda)\, d\tilde{\lambda} \le \rho v(\lambda) + L. \tag{7}$$

Then because the λ-chain is a Feller chain [1] and the function $v(\cdot)$ is unbounded off compact sets (R&H12), by Meyn and Tweedie's [8], Lemma 15.2.8, the geometric drift condition (7) implies that the λ-chain is geometrically ergodic. We now establish (7).

First, note that

$$\begin{aligned}\mathrm{E}\big[v(\tilde{\lambda})|\lambda\big] &= \int_{\mathbb{R}^{p+q}} \bigg[\int_{\mathbb{R}_+^{r+1}} v(\tilde{\lambda})\pi(\tilde{\lambda}|\theta, y)\, d\tilde{\lambda}\bigg]\pi(\theta|\lambda, y)\, d\theta = \mathrm{E}\big[\mathrm{E}\big[v(\tilde{\lambda})|\theta\big]|\lambda\big] \\ &= \alpha \mathrm{E}\big[\mathrm{E}\big[\tilde{\lambda}_e^c|\theta\big]|\lambda\big] + \alpha \mathrm{E}\big[\mathrm{E}\big[\tilde{\lambda}_e^{-s}|\theta\big]|\lambda\big] \\ &\quad + \sum_{i=1}^r \mathrm{E}\big[\mathrm{E}\big[\tilde{\lambda}_{u_i}^c|\theta\big]|\lambda\big] + \sum_{i=1}^r \mathrm{E}\big[\mathrm{E}\big[\tilde{\lambda}_{u_i}^{-s}|\theta\big]|\lambda\big].\end{aligned}$$

Using (2) and the fact that $0 < b_0 + \mathrm{SSE}/2 < b_0 + \|y - W\theta\|^2/2$, we have

$$\begin{aligned}\mathrm{E}\big[\tilde{\lambda}_e^c|\theta\big] &= \frac{\Gamma(a_0 + N/2 + c)}{\Gamma(a_0 + N/2)}\bigg(b_0 + \frac{\|y - W\theta\|^2}{2}\bigg)^{-c} \\ &\le \frac{\Gamma(a_0 + N/2 + c)}{\Gamma(a_0 + N/2)}\bigg(b_0 + \frac{\mathrm{SSE}}{2}\bigg)^{-c}.\end{aligned} \tag{8}$$

As we shall see, since this upper bound does not depend on θ, it can be absorbed into the constant term, L, and we will no longer have to deal with this piece of the drift function. Now,

$$\begin{aligned}\mathrm{E}\big[\tilde{\lambda}_e^{-s}|\theta\big] &= \frac{\Gamma(a_0 + N/2 - s)}{\Gamma(a_0 + N/2)}\bigg(b_0 + \frac{\|y - W\theta\|^2}{2}\bigg)^{s} \\ &\le \frac{\Gamma(a_0 + N/2 - s)}{\Gamma(a_0 + N/2)}\bigg(|b_0|^s + \bigg[\frac{\|y - W\theta\|^2}{2}\bigg]^s\bigg),\end{aligned} \tag{9}$$

where the inequality follows from the fact that $(x_1 + x_2)^\xi \le x_1^\xi + x_2^\xi$ for $x_1, x_2 \ge 0$ and $\xi \in (0, 1]$. Similarly, using (3), for each $i \in \{1, \dots, r\}$ we have

$$\begin{aligned}\mathrm{E}\big[\tilde{\lambda}_{u_i}^{-s}|\theta\big] &= \frac{\Gamma(a_i + q_i/2 - s)}{\Gamma(a_i + q_i/2)}\bigg(b_i + \frac{\|u_i\|^2}{2}\bigg)^{s} \\ &\le \frac{\Gamma(a_i + q_i/2 - s)}{\Gamma(a_i + q_i/2)}\bigg(b_i^s + \bigg[\frac{\|u_i\|^2}{2}\bigg]^s\bigg).\end{aligned} \tag{10}$$

Now, for each $i \in \{1, \dots, r\}$, we have

$$\begin{aligned}\mathrm{E}\big[\tilde{\lambda}_{u_i}^{c}|\theta\big] &= \frac{\Gamma(a_i + q_i/2 + c)}{\Gamma(a_i + q_i/2)}\bigg(b_i + \frac{\|u_i\|^2}{2}\bigg)^{-c} \\ &\le \frac{\Gamma(a_i + q_i/2 + c)}{\Gamma(a_i + q_i/2)}\bigg[\bigg(\frac{\|u_i\|^2}{2}\bigg)^{-c} I_{\{0\}}(b_i) + b_i^{-c} I_{\mathbb{R}_+}(b_i)\bigg].\end{aligned} \tag{11}$$

Note that when $b_i > 0$ there is a simple upper bound for this term that does not depend on θ. Therefore, we will first consider the case in which $\min\{b_1, \dots, b_r\} > 0$, and we will return to the other (more complicated) case later.

Assume for the time being that $\min\{b_1, \dots, b_r\} > 0$. Then combining (8), (9), (10) and (11), and applying Jensen's inequality twice yields

$$\begin{aligned} \mathrm{E}\big[v(\tilde{\lambda})|\lambda\big] \le{}& \frac{\alpha}{2^s} \frac{\Gamma(a_0 + N/2 - s)}{\Gamma(a_0 + N/2)} \mathrm{E}\big[\|y - W\theta\|^2|\lambda\big]^s \\ &+ 2^{-s} \sum_{i=1}^r \frac{\Gamma(a_i + q_i/2 - s)}{\Gamma(a_i + q_i/2)} \mathrm{E}\big[\|u_i\|^2|\lambda\big]^s + K_0, \end{aligned} \tag{12}$$

where

$$\begin{aligned} K_0 = {}& \alpha \frac{\Gamma(a_0 + N/2 + c)}{\Gamma(a_0 + N/2)} \left(b_0 + \frac{\mathrm{SSE}}{2}\right)^{-c} + \alpha \frac{\Gamma(a_0 + N/2 - s)}{\Gamma(a_0 + N/2)} |b_0|^s \\ &+ \sum_{i=1}^r \left[\frac{\Gamma(a_i + q_i/2 + c)}{\Gamma(a_i + q_i/2)} b_i^{-c} + \frac{\Gamma(a_i + q_i/2 - s)}{\Gamma(a_i + q_i/2)} b_i^s \right]. \end{aligned}$$

It follows from (5) that

$$\mathrm{E}\big[\|y - W\theta\|^2|\lambda\big] = \mathrm{tr}\big(W \,\mathrm{Var}(\theta|\lambda) W^T\big) + \big\|y - W\mathrm{E}[\theta|\lambda]\big\|^2.$$

Similarly, since $u_i = R_i u$, we also have

$$\mathrm{E}\big[\|u_i\|^2|\lambda\big] = \mathrm{E}\big[\|R_i u\|^2|\lambda\big] = \mathrm{tr}\big(R_i Q_\lambda^{-1} R_i^T\big) + \big\|\mathrm{E}[R_i u|\lambda]\big\|^2.$$

Now, using Lemmas 2 and 3, we have

$$\begin{aligned} \mathrm{E}\big[\|y - W\theta\|^2|\lambda\big]^s &\le \big[\lambda_e^{-1} \,\mathrm{rank}(Z) + d_{\max}^2 \,\mathrm{rank}(Z) + \mathrm{tr}\big(X\Sigma_\beta X^T\big) + K_2^2\big]^s \\ &\le \lambda_e^{-s}\big(\mathrm{rank}(Z)\big)^s + \big[d_{\max}^2 \,\mathrm{rank}(Z) + \mathrm{tr}\big(X\Sigma_\beta X^T\big) + K_2^2\big]^s. \end{aligned} \tag{13}$$

Similarly, using Lemmas 1 and 3, we have

$$\begin{aligned} \mathrm{E}\big[\|u_i\|^2|\lambda\big]^s \le{}& \Bigg[\big(d_{\max}^2 + \lambda_e^{-1}\big) \mathrm{tr}\big(R_i (Z^T Z)^+ R_i^T\big) \\ &+ \mathrm{tr}\big(R_i (I - P_{Z^T Z}) R_i^T\big) \sum_{j=1}^r \lambda_{u_j}^{-1} + q_i K_1^2 \Bigg]^s \\ \le{}& \lambda_e^{-s} \big(\mathrm{tr}\big(R_i (Z^T Z)^+ R_i^T\big)\big)^s + \sum_{j=1}^r \lambda_{u_j}^{-s} \big(\mathrm{tr}\big(R_i (I - P_{Z^T Z}) R_i^T\big)\big)^s \\ &+ \big[d_{\max}^2 \,\mathrm{tr}\big(R_i (Z^T Z)^+ R_i^T\big) + q_i K_1^2\big]^s. \end{aligned} \tag{14}$$

Define a function $\delta(\cdot)$ as follows:

$$\delta(\alpha) = \frac{\Gamma(a_0 + N/2 - s)}{\Gamma(a_0 + N/2)} \left(\frac{\operatorname{rank}(Z)}{2}\right)^s + \frac{1}{\alpha} \sum_{i=1}^{r} \frac{\Gamma(a_i + q_i/2 - s)}{\Gamma(a_i + q_i/2)} \left(\frac{\operatorname{tr}(R_i (Z^T Z)^+ R_i^T)}{2}\right)^s.$$

Combining (12), (13) and (14) yields

$$\begin{aligned} \mathrm{E}\big[v(\tilde{\lambda})|\lambda\big] &\leq \alpha\delta(\alpha)\lambda_e^{-s} \\ &\quad + \left[\sum_{i=1}^{r} \frac{\Gamma(a_i + q_i/2 - s)}{\Gamma(a_i + q_i/2)} \left(\frac{\operatorname{tr}(R_i(I - P_{Z^T Z})R_i^T)}{2}\right)^s\right] \sum_{j=1}^{r} \lambda_{u_j}^{-s} + L, \end{aligned} \tag{15}$$

where

$$\begin{aligned} L = K_0 &+ \frac{\alpha}{2^s} \frac{\Gamma(a_0 + N/2 - s)}{\Gamma(a_0 + N/2)} \big[d_{\max}^2 \operatorname{rank}(Z) + \operatorname{tr}\big(X \Sigma_\beta X^T\big) + K_2^2\big]^s \\ &+ 2^{-s} \sum_{i=1}^{r} \frac{\Gamma(a_i + q_i/2 - s)}{\Gamma(a_i + q_i/2)} \big[d_{\max}^2 \operatorname{tr}\big(R_i \big(Z^T Z\big)^+ R_i^T\big) + q_i K_1^2\big]^s. \end{aligned}$$

Next, defining

$$\rho(\alpha) := \max\left\{\delta(\alpha), \sum_{i=1}^{r} \frac{\Gamma(a_i + q_i/2 - s)}{\Gamma(a_i + q_i/2)} \left(\frac{\operatorname{tr}(R_i(I - P_{Z^T Z})R_i^T)}{2}\right)^s\right\},$$

we have from (15) that

$$\begin{aligned} \mathrm{E}\big[v(\tilde{\lambda})|\lambda\big] &\leq \alpha\rho(\alpha)\lambda_e^{-s} + \rho(\alpha) \sum_{j=1}^{r} \lambda_{u_j}^{-s} + L \\ &\leq \rho(\alpha)\left(\alpha\lambda_e^c + \alpha\lambda_e^{-s} + \sum_{j=1}^{r} \lambda_{u_j}^c + \sum_{j=1}^{r} \lambda_{u_j}^{-s}\right) + L \\ &= \rho(\alpha) v(\lambda) + L. \end{aligned}$$

Hence, all that is left is to demonstrate the existence of an $\alpha \in (0, \infty)$ such that $\rho(\alpha) \in [0, 1)$. By (6), we know that

$$\sum_{i=1}^{r} \frac{\Gamma(a_i + q_i/2 - s)}{\Gamma(a_i + q_i/2)} \left(\frac{\operatorname{tr}(R_i(I - P_{Z^T Z})R_i^T)}{2}\right)^s < 1.$$

Therefore, it suffices to show that there exists an $\alpha \in (0, \infty)$ such that $\delta(\alpha) < 1$. But $\delta(\alpha) < 1$ as long as

$$\alpha > \frac{\sum_{i=1}^{r} \frac{\Gamma(a_i + q_i/2 - s)}{\Gamma(a_i + q_i/2)} (\operatorname{tr}(R_i(I - P_{Z^T Z})R_i^T)/2)^s}{1 - \frac{\Gamma(a_0 + N/2 - s)}{\Gamma(a_0 + N/2)} (\operatorname{rank}(Z)/2)^s}, \tag{16}$$

which is a well-defined positive number by (6). The result has now been proven for the case in which $\min\{b_1,\dots,b_r\} > 0$.

Remark 3. Note that the two terms in the drift function involving c were both absorbed into the constant in the first step of the iterated expectation. It follows that, at least in the case where $\min\{b_1,\dots,b_r\} > 0$, any $c > 0$ can be used in the drift function.

We now proceed to the case in which there is at least one $b_i = 0$. Let $B = \{i \in \{1,\dots,r\} : b_i = 0\}$. It follows from the development above that the following holds for any $c > 0$:

$$\alpha \mathrm{E}\big[\tilde{\lambda}_e^c|\lambda\big] + \alpha \mathrm{E}\big[\tilde{\lambda}_e^{-s}|\lambda\big] + \sum_{i\notin B} \mathrm{E}\big[\tilde{\lambda}_{u_i}^c|\lambda\big] + \sum_{i=1}^{r} \mathrm{E}\big[\tilde{\lambda}_{u_i}^{-s}|\lambda\big] \le \rho(\alpha)\left(\alpha\lambda_e^{-s} + \sum_{j=1}^{r}\lambda_{u_j}^{-s}\right) + L. \quad (17)$$

Of course, if α satisfies (16), then $\rho(\alpha) \in [0,1)$. Now suppose we can find $c > 0$, α satisfying (16), and $\rho'(\alpha) \in [0,1)$ such that

$$\sum_{i\in B} \mathrm{E}\big[\tilde{\lambda}_{u_i}^c|\lambda\big] \le \rho'(\alpha)\left(\alpha\lambda_e^c + \sum_{i\in B}\lambda_{u_i}^c\right). \quad (18)$$

Then combining (17) and (18), we would have

$$\begin{aligned}\mathrm{E}\big[v(\tilde{\lambda})|\lambda\big] &\le \rho(\alpha)\left(\alpha\lambda_e^{-s} + \sum_{j=1}^{r}\lambda_{u_j}^{-s}\right) + L + \rho'(\alpha)\left(\alpha\lambda_e^c + \sum_{i\in B}\lambda_{u_i}^c\right)\\ &\le \max\big\{\rho(\alpha),\rho'(\alpha)\big\}v(\lambda) + L,\end{aligned}$$

which establishes the drift condition. Therefore, to prove the result when $\min\{b_1,\dots,b_r\} = 0$, it suffices to establish (18). If $i \in B$, then

$$\mathrm{E}\big[\tilde{\lambda}_i^c|\theta\big] = \frac{\Gamma(a_i + q_i/2 + c)}{\Gamma(a_i + q_i/2)}\left(\frac{\|u_i\|^2}{2}\right)^{-c}.$$

It follows from (5) that the conditional distribution of $(R_i Q_\lambda^{-1} R_i^T)^{-1/2}u_i$ given λ is multivariate normal with identity covariance matrix. Thus, $u_i^T(R_i Q_\lambda^{-1} R_i^T)^{-1}u_i$ has a non-central chi-squared distribution with q_i degrees of freedom. An application of Lemma 4 from R&H12 shows that, if $c \in (0,1/2)$, then

$$\mathrm{E}\Big[\big(u_i^T\big(R_i Q_\lambda^{-1} R_i^T\big)^{-1}u_i\big)^{-c}|\lambda\Big] \le 2^{-c}\frac{\Gamma(q_i/2 - c)}{\Gamma(q_i/2)}.$$

Putting this together with Lemma 4, we have that, if $i \in B$ and $c \in (0,1/2)$, then

$$\begin{aligned}\mathrm{E}\big[\big(\|u_i\|^2\big)^{-c}|\lambda\big] &= (\psi_{\max}\lambda_e + \lambda_{u_i})^c\,\mathrm{E}\Big[\big(u_i^T(\psi_{\max}\lambda_e + \lambda_{u_i})I_{q_i}u_i\big)^{-c}|\lambda\Big]\\ &\le (\psi_{\max}\lambda_e + \lambda_{u_i})^c\,\mathrm{E}\Big[\big(u_i^T\big(R_i Q_\lambda^{-1} R_i^T\big)^{-1}u_i\big)^{-c}|\lambda\Big]\end{aligned}$$

$$\leq 2^{-c} \frac{\Gamma(q_i/2 - c)}{\Gamma(q_i/2)} (\psi_{\max}\lambda_e + \lambda_{u_i})^c$$
$$\leq 2^{-c} \frac{\Gamma(q_i/2 - c)}{\Gamma(q_i/2)} \big(\psi_{\max}^c \lambda_e^c + \lambda_{u_i}^c\big).$$

Define $\delta'(\cdot)$ as follows:

$$\delta'(\alpha) = \frac{\psi_{\max}^c}{\alpha} \sum_{i \in B} \frac{\Gamma(a_i + q_i/2 + c)}{\Gamma(a_i + q_i/2)} \frac{\Gamma(q_i/2 - c)}{\Gamma(q_i/2)}.$$

Now we have

$$\sum_{i\in B} \mathrm{E}\big[\tilde{\lambda}_{u_i}^c|\lambda\big] \leq \sum_{i \in B} \frac{\Gamma(a_i + q_i/2 + c)}{\Gamma(a_i + q_i/2)} \frac{\Gamma(q_i/2 - c)}{\Gamma(q_i/2)} \big(\psi_{\max}^c \lambda_e^c + \lambda_{u_i}^c\big)$$
$$= \alpha\delta'(\alpha)\lambda_e^c + \sum_{i \in B} \frac{\Gamma(a_i + q_i/2 + c)}{\Gamma(a_i + q_i/2)} \frac{\Gamma(q_i/2 - c)}{\Gamma(q_i/2)} \lambda_{u_i}^c.$$

Next, defining

$$\rho'(\alpha) = \max\left\{\delta'(\alpha), \max_{i\in B}\left\{\frac{\Gamma(a_i + q_i/2 + c)}{\Gamma(a_i + q_i/2)} \frac{\Gamma(q_i/2 - c)}{\Gamma(q_i/2)}\right\}\right\},$$

we have

$$\sum_{i\in B} \mathrm{E}\big[\tilde{\lambda}_{u_i}^c|\lambda\big] \leq \rho'(\alpha)\left(\alpha\lambda_e^c + \sum_{i\in B}\lambda_{u_i}^c\right).$$

Hence, all we have left to do is to prove that there exist $c \in (0, 1/2)$ and α satisfying (16) such that $\rho'(\alpha) \in [0, 1)$. First, define $\tilde{a} = -\max_{i\in B} a_i$, and note that this quantity is positive. R&H12 show that, if $c \in (0, 1/2) \cap (0, \tilde{a})$, then

$$\max_{i\in B}\left\{\frac{\Gamma(a_i + q_i/2 + c)}{\Gamma(a_i + q_i/2)} \frac{\Gamma(q_i/2 - c)}{\Gamma(q_i/2)}\right\} < 1.$$

Fix $c \in (0, 1/2) \cap (0, \tilde{a})$. Now it suffices to show that there exists an α satisfying (16) such that $\delta'(\alpha) < 1$. But $\delta'(\alpha) < 1$ as long as

$$\alpha > \psi_{\max}^c \sum_{i \in B} \frac{\Gamma(a_i + q_i/2 + c)}{\Gamma(a_i + sq_i/2)} \frac{\Gamma(q_i/2 - c)}{\Gamma(q_i/2)}.$$

So, (18) is satisfied for $c \in (0, 1/2) \cap (0, \tilde{a})$ and

$$\alpha > \max\left\{\frac{\sum_{i=1}^r \frac{\Gamma(a_i+q_i/2-s)}{\Gamma(a_i+q_i/2)}(\mathrm{tr}(R_i(I - P_{Z^TZ})R_i^T)/2)^s}{1 - \frac{\Gamma(a_0+N/2-s)}{\Gamma(a_0+N/2)}(\mathrm{rank}(Z)/2)^s},\right.$$
$$\left.\psi_{\max}^c \sum_{i \in B} \frac{\Gamma(a_i + q_i/2 + c)}{\Gamma(a_i + q_i/2)} \frac{\Gamma(q_i/2 - c)}{\Gamma(q_i/2)}\right\}. \qquad \square$$

4. Discussion

Our Corollary 1 is a direct generalization of Román and Hobert's [13] Proposition 1 where we have removed all restrictions on the matrix X. We now present a related result from [1] that is established using a different drift function.

Proposition 2. *Under a proper prior, the block Gibbs Markov chain,* $\{(\lambda_n, \theta_n)\}_{n=0}^{\infty}$, *is geometrically ergodic if* $\min\{a_0, a_1, \dots, a_r\} > 1$.

Like Corollary 1, this result holds for any X. Neither result is uniformly better than the other. That is, there are situations where the conditions of Corollary 1 hold, but those of Proposition 2 do not, and vice versa. However, the condition $\min\{a_0, a_1, \dots, a_r\} > 1$ appears to be more restrictive than the conditions of Corollary 1 in nearly all *practical* settings. In fact, the only examples we could find where Proposition 2 is better than Corollary 1 involve models that have more random effects than observations. On the other hand, we do feel that Proposition 2 is worth mentioning because its simple form may render it useful to practitioners. For example, in an exploratory phase where a number of different models are being considered for a given set of data, one could avoid having to recheck the conditions of Corollary 1 each time the model is changed simply by taking $a_0 = a_1 = \dots = a_r = a > 1$ for all models under consideration.

Appendix A: Preliminary results

Let $k = \text{rank}(\tilde{X}) = \text{rank}(X) \le \min\{N, p\}$, and consider a singular value decomposition of $\tilde{X}$ given by UDV^T, where U and V are orthogonal matrices of dimension N and p, respectively, and

$$D := \begin{bmatrix} D_* & 0_{k,p-k} \\ 0_{N-k,k} & 0_{N-k,p-k} \end{bmatrix},$$

where $D_* := \text{diag}\{d_1, \dots, d_k\}$. The values $d_1, \dots, d_k$ are the singular values of $\tilde{X}$, which are strictly positive. Again, $d_{\max}$ denotes the largest singular value. The following result is an extension of Lemmas 4 and 5 in R&H15.

Lemma 5. *The matrix* M_λ *can be represented as* $UH_\lambda U^T$ *where* H_λ *is an* $N \times N$ *diagonal matrix,* $H_\lambda = \text{diag}\{h_1, \dots, h_N\}$, *where*

$$h_i = \begin{cases} \dfrac{1}{\lambda_e d_i^2 + 1}, & i \in \{1, \dots, k\}, \\ 1, & i \in \{k+1, \dots, N\}. \end{cases}$$

Furthermore, $(\lambda_e d_{\max}^2 + 1)^{-1} I \preceq M_\lambda \preceq I$.

Proof. Using the definitions of T_λ^{-1} and $\tilde{X}$, we have

$$M_\lambda = I - \lambda_e X T_\lambda^{-1} X^T = I - \lambda_e \tilde{X}\big(\lambda_e \tilde{X}^T \tilde{X} + I\big)^{-1} \tilde{X}^T.$$

Now using $\tilde{X} = UDV^T$ leads to

$$M_\lambda = U\big(I - \lambda_e D\big(\lambda_e D^T D + I\big)^{-1} D^T\big)U^T.$$

The matrix $\lambda_e D(\lambda_e D^T D + I)^{-1} D^T$ is an $N \times N$ diagonal matrix whose jth diagonal element is given by

$$\frac{\lambda_e d_j^2}{\lambda_e d_j^2 + 1} I_{\{1,2,\dots,k\}}(j).$$

Hence, $I - \lambda_e D(\lambda_e D^T D + I)^{-1} D^T = H_\lambda$, and $M_\lambda = U H_\lambda U^T$. To prove the second part, note that, for $j = 1, \dots, N$, $0 < (\lambda_e d_{\max}^2 + 1)^{-1} \le h_i \le 1$. Thus,

$$\big(\lambda_e d_{\max}^2 + 1\big)^{-1} I = U\big(\lambda_e d_{\max}^2 + 1\big)^{-1} U^T \preceq U H_\lambda U^T \preceq U U^T = I. \qquad \square$$

Next, we develop an extension of Lemma 2 in R&H15. Define $\tilde{Z} = U^T Z$, $\tilde{y} = U^T y$ and $\eta = V^T \Sigma_\beta^{-1/2} \mu_\beta$. Also, let $\tilde{z}_i$ denote the ith column of $\tilde{Z}^T$, and let $\tilde{y}_i$ and η_i represent the ith components of the vectors $\tilde{y}$ and η, respectively. Let $t_1, t_2, \dots, t_{N+q}$ be a set of q-vectors defined as follows. For $j = 1, \dots, N$, let $t_j = \tilde{z}_j$, and let $t_{N+1}, \dots, t_{N+q}$ be the standard basis vectors in $\mathbb{R}^q$. For $i = 1, \dots, N$, define

$$C_i^* = \left[\sup_{a \in \mathbb{R}_+^{N+q}} t_i^T \left(t_i t_i^T + \sum_{j \in \{1,2,\dots,N\} \setminus \{i\}} a_j t_j t_j^T + \sum_{j=N+1}^{N+q} a_j t_j t_j^T + a_i I\right)^{-2} t_i\right]^{1/2}.$$

The C_i^*s are finite by [7], Lemma 3.

Lemma 6. *For all* $\lambda \in \mathbb{R}^{r+1}$,

$$\big\|\lambda_e Q_\lambda^{-1} Z^T M_\lambda y\big\| \le \sum_{j=1}^N |\tilde{y}_j| C_j^* < \infty$$

and

$$\big\|\lambda_e Q_\lambda^{-1} Z^T X T_\lambda^{-1} \Sigma_\beta^{-1} \mu_\beta\big\| \le \sum_{j=1}^k d_j |\eta_j| C_j^* < \infty.$$

Proof. Even though R&H15 assume X to be full column rank, their argument still works to establish the first inequality, so we omit this argument. We now establish the second inequality. First,

$$U^T X T_\lambda^{-1} = U^T \tilde{X}\big(\lambda_e \tilde{X}^T \tilde{X} + I\big)^{-1} \Sigma_\beta^{1/2} = D\big(\lambda_e D^T D + I\big)^{-1} V^T \Sigma_\beta^{1/2}.$$

Define $R_\lambda = D(\lambda_e D^T D + I)^{-1}$. This is an $N \times p$ diagonal matrix, with diagonal elements $r_1, r_2, \ldots, r_{\min\{N,p\}}$. These take the form

$$r_j = \frac{d_j}{\lambda_e d_j^2 + 1} I_{\{1,2,\ldots,k\}}(j).$$

Now

$$\begin{aligned}
\left\|\lambda_e Q_\lambda^{-1} Z^T X T_\lambda^{-1} \Sigma_\beta^{-1} \mu_\beta\right\| &= \left\|\lambda_e Q_\lambda^{-1} \tilde{Z}^T R_\lambda V^T \Sigma_\beta^{-1/2} \mu_\beta\right\| \\
&= \left\|\lambda_e \big(\lambda_e Z^T M_\lambda Z + \Lambda\big)^{-1} \tilde{Z}^T R_\lambda V^T \Sigma_\beta^{-1/2} \mu_\beta\right\| \\
&= \left\|\big(\tilde{Z}^T H_\lambda \tilde{Z} + \lambda_e^{-1}\Lambda\big)^{-1} \tilde{Z}^T R_\lambda \eta\right\| \\
&= \left\|\sum_{i=1}^k \big(\tilde{Z}^T H_\lambda \tilde{Z} + \lambda_e^{-1}\Lambda\big)^{-1} \tilde{z}_i r_i \eta_i\right\| \\
&\le \sum_{i=1}^k \left\|\big(\tilde{Z}^T H_\lambda \tilde{Z} + \lambda_e^{-1}\Lambda\big)^{-1} \tilde{z}_i r_i \eta_i\right\| \\
&= \sum_{i=1}^k \left\|\left(\sum_{j=1}^N \tilde{z}_j \tilde{z}_j^T h_j + \lambda_e^{-1}\Lambda\right)^{-1} \tilde{z}_i r_i \eta_i\right\| \\
&= \sum_{i=1}^k \left\|\left(\tilde{z}_i \tilde{z}_i^T + \sum_{j\neq i} \tilde{z}_j \tilde{z}_j^T \frac{h_j}{h_i} + h_i^{-1}\lambda_e^{-1}\Lambda\right)^{-1} \tilde{z}_i \frac{r_i}{h_i} \eta_i\right\| \\
&= \sum_{i=1}^k d_i |\eta_i| \left\|\left(\tilde{z}_i \tilde{z}_i^T + \sum_{j\neq i} \tilde{z}_j \tilde{z}_j^T \frac{h_j}{h_i} + h_i^{-1}\lambda_e^{-1}\Lambda\right)^{-1} \tilde{z}_i\right\|,
\end{aligned}$$

where, in the last step, we have used the fact that $h_i d_i = r_i$ for $i = 1, \ldots, k$. For $i = 1, 2, \ldots, k$, define

$$C_i(\lambda) = \left\|\left(\tilde{z}_i \tilde{z}_i^T + \sum_{j\neq i} \tilde{z}_j \tilde{z}_j^T \frac{h_j}{h_i} + h_i^{-1}\lambda_e^{-1}\Lambda\right)^{-1} \tilde{z}_i\right\|.$$

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

$$\begin{aligned}
C_i^2(\lambda) &= \tilde{z}_i^T \left(\tilde{z}_i \tilde{z}_i^T + \sum_{j\neq i} \tilde{z}_j \tilde{z}_j^T \frac{h_j}{h_i} + h_i^{-1}\lambda_e^{-1}\Lambda\right)^{-2} \tilde{z}_i \\
&= \tilde{z}_i^T \left(\tilde{z}_i \tilde{z}_i^T + \sum_{j\neq i} \tilde{z}_j \tilde{z}_j^T \frac{h_j}{h_i} + h_i^{-1}\lambda_e^{-1}\big(\Lambda - \lambda_\bullet^{-1} I\big) + \frac{1}{h_i \lambda_e \lambda_\bullet} I\right)^{-2} \tilde{z}_i.
\end{aligned}$$

Define $\{w_j\}_{j=1}^{N+q}$ as follows:

$$w_j = \begin{cases} \dfrac{h_j}{h_i}, & j = 1, \dots, i-1, i+1, \dots, N, \\ \dfrac{1}{h_i \lambda_e \lambda_\bullet}, & j = i, \\ \dfrac{\lambda_{u_1} - \lambda_\bullet^{-1}}{h_i \lambda_e}, & j = N+1, \dots, N+q_1, \\ \dfrac{\lambda_{u_2} - \lambda_\bullet^{-1}}{h_i \lambda_e}, & j = N+q_1+1, \dots, N+q_1+q_2, \\ \vdots & \vdots \\ \dfrac{\lambda_{u_r} - \lambda_\bullet^{-1}}{h_i \lambda_e}, & j = N+q_1+\cdots+q_{r-1}+1, \dots, N+q. \end{cases} \tag{A.1}$$

Then

$$C_i^2(\lambda) = t_i^T \left(t_i t_i^T + \sum_{j \in \{1,2,\dots,N\} \setminus \{i\}} w_j t_j t_j^T + \sum_{j=N+1}^{N+q} w_j t_j t_j^T + w_i I \right)^{-2} t_i .$$

Clearly, $w_j > 0$ for all $j = 1, \dots, N+q$. It follows that

$$C_i^2(\lambda) \le \sup_{a \in \mathbb{R}_+^{N+q}} t_i^T \left(t_i t_i^T + \sum_{j \in \{1,2,\dots,N\} \setminus \{i\}} a_j t_j t_j^T + \sum_{j=N+1}^{N+q} a_j t_j t_j^T + a_i I \right)^{-2} t_i = \left(C_i^*\right)^2 .$$

Hence,

$$\left\| \lambda_e Q_\lambda^{-1} Z^T X T_\lambda^{-1} \Sigma_\beta^{-1} \mu_\beta \right\| \le \sum_{i=1}^{k} d_i |\eta_i| C_i^* .$$

□

Appendix B: Proof of Lemma 1

Lemma 1. *For each $i \in \{1, 2, \dots, r\}$, we have*

$$\operatorname{tr}\left(R_i Q_\lambda^{-1} R_i^T\right) \le \left(d_{\max}^2 + \lambda_e^{-1}\right) \operatorname{tr}\left(R_i \left(Z^T Z\right)^+ R_i^T\right) + \operatorname{tr}\left(R_i (I - P_{Z^T Z}) R_i^T\right) \sum_{j=1}^{r} \lambda_{u_j}^{-1} .$$

Proof. From Lemma 5 we have

$$Q_\lambda = \lambda_e Z^T M_\lambda Z + \Lambda \succeq \frac{\lambda_e}{\lambda_e d_{\max}^2 + 1} Z^T Z + \Lambda \succeq \frac{\lambda_e}{\lambda_e d_{\max}^2 + 1} Z^T Z + \lambda_{\min} I,$$

where $\lambda_{\min} = \min\{\lambda_{u_1}, \dots, \lambda_{u_r}\}$. Letting $O\Psi O^T$ be the spectral decomposition of $Z^T Z$, we have

$$Q_\lambda^{-1} \preceq \left(\frac{\lambda_e}{\lambda_e d_{\max}^2 + 1} Z^T Z + \lambda_{\min} I \right)^{-1} = O\left(\frac{1}{d_{\max}^2 + \lambda_e^{-1}} \Psi + \lambda_{\min} I \right)^{-1} O^T. \tag{B.1}$$

Next, let Ψ^+ be a $q \times q$ diagonal matrix whose ith diagonal element is

$$\psi_i^+ = \psi_i^{-1}\big(1 - I_{\{0\}}(\psi_i)\big).$$

Now note that, for $i = 1, \dots, q$, we have

$$\left(\frac{\psi_i}{d_{\max}^2 + \lambda_e^{-1}} + \lambda_{\min} \right)^{-1} \le \big(d_{\max}^2 + \lambda_e^{-1}\big)\psi_i^+ + \lambda_{\min}^{-1} I_{\{0\}}\big(\psi_i^+\big).$$

Hence,

$$\left(\frac{1}{d_{\max}^2 + \lambda_e^{-1}} \Psi + \lambda_{\min} I \right)^{-1} \preceq \big(d_{\max}^2 + \lambda_e^{-1}\big)\Psi^+ + \lambda_{\min}^{-1}(I - P_\Psi), \tag{B.2}$$

where P_Ψ is a $q \times q$ diagonal matrix whose ith diagonal entry is $1 - I_{\{0\}}(\psi_i)$. Combining (B.1) and (B.2) yields

$$\begin{aligned} Q_\lambda^{-1} &\preceq \big(d_{\max}^2 + \lambda_e^{-1}\big) O \Psi^+ O^T + \lambda_{\min}^{-1} O(I - P_\Psi) O^T \\ &= \big(d_{\max}^2 + \lambda_e^{-1}\big)\big(Z^T Z\big)^+ + \lambda_{\min}^{-1} O(I - P_\Psi) O^T. \end{aligned}$$

Let $\mathcal{I} = \{i \in \{1, \dots, q\} : \psi_i > 0\}$, and let $\tilde{O}$ be the sub-matrix of O consisting of the column vectors o_i where $i \in \mathcal{I}$. Then

$$O P_\Psi O^T = \sum_{i \in \mathcal{I}} o_i o_i^T = \tilde{O}\tilde{O}^T.$$

Since $\{o_i\}_{i \in \mathcal{I}}$ forms an orthonormal basis for the column space of $Z^T Z$, it follows that $\tilde{O}\tilde{O}^T$ is the orthogonal projection onto $Z^T Z$. Consequently,

$$O(I - P_\Psi)O^T = O O^T - O P_\Psi O^T = I - \tilde{O}\tilde{O}^T = I - P_{Z^T Z}.$$

Thus,

$$Q_\lambda^{-1} \preceq \big(d_{\max}^2 + \lambda_e^{-1}\big)\big(Z^T Z\big)^+ + (I - P_{Z^T Z}) \sum_{i=1}^r \lambda_{u_i},$$

and finally,

$$\operatorname{tr}\big(R_i Q_\lambda^{-1} R_i^T\big) \le \big(d_{\max}^2 + \lambda_e^{-1}\big) \operatorname{tr}\big(R_i \big(Z^T Z\big)^+ R_i^T\big) + \operatorname{tr}\big(R_i (I - P_{Z^T Z}) R_i^T\big) \sum_{j=1}^r \lambda_{u_j}^{-1}.$$

□

Appendix C: Proof of Lemma 2

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

Proof. R&H15 show that

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

and that $\mathrm{tr}((I-M_\lambda)ZQ_\lambda^{-1}Z^T(I+M_\lambda)) \ge 0$. Hence,

$$\mathrm{tr}\big(W\,\mathrm{Var}(\theta|\lambda)W^T\big) \le \mathrm{tr}\big(ZQ_\lambda^{-1}Z^T\big) + \mathrm{tr}\big(XT_\lambda^{-1}X^T\big),$$

Next, note that $\Sigma_\beta^{-1} \preceq \lambda_e X^T X + \Sigma_\beta^{-1} = T_\lambda$. Hence, $\Sigma_\beta \succeq T_\lambda^{-1}$, and

$$\mathrm{tr}\big(XT_\lambda^{-1}X^T\big) \le \mathrm{tr}\big(X\Sigma_\beta X^T\big).$$

Now, from Lemma 5, we have

$$\frac{\lambda_e}{\lambda_e d_{\max}^2 + 1} Z^T Z + \Lambda \preceq \lambda_e Z^T M_\lambda Z + \Lambda = Q_\lambda,$$

and it follows that

$$\mathrm{tr}\big(ZQ_\lambda^{-1}Z^T\big) \le \mathrm{tr}\left(Z\left(\frac{\lambda_e}{\lambda_e d_{\max}^2 + 1} Z^T Z + \Lambda\right)^{-1} Z^T\right).$$

Finally, using Lemma 3 from R&H15, we have

$$\begin{aligned}\mathrm{tr}\left(Z\left(\frac{\lambda_e}{\lambda_e d_{\max}^2 + 1} Z^T Z + \Lambda\right)^{-1} Z^T\right) &\le \left(\frac{\lambda_e}{\lambda_e d_{\max}^2 + 1}\right)^{-1} \mathrm{rank}(Z) \\ &= \lambda_e^{-1}\,\mathrm{rank}(Z) + d_{\max}^2\,\mathrm{rank}(Z).\end{aligned}$$ □

Appendix D: Proof of Lemma 3

Lemma 3. *There exist finite constants K_1 and K_2, not depending on λ, such that $\|\mathrm{E}[R_i u|\lambda]\| \le \sqrt{q_i}K_1$ for $i = 1,\dots,r$, and $\|y - W\mathrm{E}[\theta|\lambda]\| \le K_2$.*

Proof. From (4) and Lemma 6, we have

$$\begin{aligned}\big\|\mathrm{E}[u|\lambda]\big\| &= \big\|\lambda_e Q_\lambda^{-1}Z^T\big(M_\lambda y - XT_\lambda^{-1}\Sigma_\beta^{-1}\mu_\beta\big)\big\| \\ &\le \big(\big\|\lambda_e Q_\lambda^{-1}Z^T M_\lambda y\big\| + \big\|\lambda_e Q_\lambda^{-1}Z^T X T_\lambda^{-1}\Sigma_\beta^{-1}\mu_\beta\big\|\big) \\ &\le \left(\sum_{j=1}^N |y_j| C_j^* + \sum_{j=1}^k d_j|\eta_j| C_j^*\right) := K_1.\end{aligned}$$

Now, for each $i \in \{1, \dots, q\}$, we have

$$\left\| \mathrm{E}[R_i u|\lambda] \right\| \le \|R_i\| K_1 = \sqrt{\mathrm{tr}\big(R_i^T R_i\big)} K_1 = \sqrt{q_i} K_1.$$

This proves the first part. Now, it follows from page 10 of R&H15 that

$$\left\| y - W\mathrm{E}[\theta|\lambda] \right\| \le \|M_\lambda\| \|y\| + \left\| X T_\lambda^{-1} \Sigma_\beta^{-1} \mu_\beta \right\| + \|M_\lambda\| \|Z\| \left\| \mathrm{E}[u|\lambda] \right\|.$$

Now, using Lemma 5, and the fact that $h_i \le 1$, for $i = 1, \dots, N$, we have

$$\|M_\lambda\|^2 = \mathrm{tr}\big(M_\lambda^T M_\lambda\big) = \sum_{j=1}^{N} h_i^2 \le N.$$

Recall from the proof of Lemma 6 that $U^T X T_\lambda^{-1} = R_\lambda V^T \Sigma_\beta^{1/2}$, and note that

$$\|R_\lambda\|^2 = \mathrm{tr}\big(R_\lambda^T R_\lambda\big) = \sum_{j=1}^{k} r_i^2 \le k d_{\max}^2.$$

Therefore,

$$\begin{aligned} \left\| X T_\lambda^{-1} \Sigma_\beta^{-1} \mu_\beta \right\| &= \left\| U U^T X T_\lambda^{-1} \Sigma_\beta^{-1} \mu_\beta \right\| \\ &= \left\| U R_\lambda V^T \Sigma_\beta^{-1/2} \mu_\beta \right\| \\ &\le \|U\| \|R_\lambda\| \left\| V^T \Sigma_\beta^{-1/2} \mu_\beta \right\| \\ &\le \sqrt{N} \sqrt{k} d_{\max} \left\| V^T \Sigma_\beta^{-1/2} \mu_\beta \right\|. \end{aligned}$$

Putting all of this together, we have

$$\begin{aligned} \left\| y - W\mathrm{E}[\theta|\lambda] \right\| &\le \|M_\lambda\| \|y\| + \left\| X T_\lambda^{-1} \Sigma_\beta^{-1} \mu_\beta \right\| + \|M_\lambda\| \|Z\| \left\| \mathrm{E}[u|\lambda] \right\| \\ &\le \sqrt{N} \|y\| + \sqrt{N} \sqrt{k} d_{\max} \left\| V^T \Sigma_\beta^{-1/2} \mu_\beta \right\| + \sqrt{N} \|Z\| K_1. \end{aligned}$$

□

Appendix E: Proof of Lemma 4

Lemma 4. *For each $i \in \{1, 2, \dots, r\}$, we have $(\psi_{\max} \lambda_e + \lambda_{u_i})^{-1} I_{q_i} \preceq R_i Q_\lambda^{-1} R_i^T$.*

Proof. Lemma 5 implies that $Z^T M_\lambda Z \preceq Z^T Z$. It follows that

$$Q_\lambda = \lambda_e Z^T M_\lambda Z + \Lambda \preceq \lambda_e Z^T Z + \Lambda \preceq \lambda_e \psi_{\max} I + \Lambda.$$

Thus,

$$(\lambda_e \psi_{\max} + \lambda_{u_i})^{-1} I = R_i (\lambda_e \psi_{\max} I + \Lambda)^{-1} R_i^T \preceq R_i Q_\lambda^{-1} R_i^T.$$

□

Acknowledgment

The authors thank three anonymous reviewers for helpful comments and suggestions that led to a substantially improved version of the paper.

Tavis Abrahamsen supported by NSF Grant 08-01544 (in the Quantitative Spatial Ecology, Evolution and Environment Program at the University of Florida). James P. Hobert supported by NSF Grant DMS-11-06395.

References

[1] Abrahamsen, T. (2015). Convergence analysis of Gibbs samplers for Bayesian linear mixed models for large p, small n problems. Ph.D. thesis, Dept. Statistics, Univ. Florida.

[2] Diaconis, P., Khare, K. and Saloff-Coste, L. (2008). Gibbs sampling, exponential families and orthogonal polynomials. *Statist. Sci.* **23** 151–178. MR2446500

[3] Flegal, J.M., Haran, M. and Jones, G.L. (2008). Markov chain Monte Carlo: Can we trust the third significant figure? *Statist. Sci.* **23** 250–260. MR2516823

[4] Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Anal.* **1** 515–534.

[5] Johnson, A.A. and Jones, G.L. (2010). Gibbs sampling for a Bayesian hierarchical general linear model. *Electron. J. Stat.* **4** 313–333. MR2645487

[6] Jones, G.L. and Hobert, J.P. (2001). Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Statist. Sci.* **16** 312–334. MR1888447

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

[8] Meyn, S.P. and Tweedie, R.L. (1993). *Markov Chains and Stochastic Stability*. *Communications and Control Engineering Series*. London: Springer. MR1287609

[9] Roberts, G.O. and Rosenthal, J.S. (1998). Markov-chain Monte Carlo: Some practical implications of theoretical results. *Canad. J. Statist.* **26** 5–31. MR1624414

[10] Roberts, G.O. and Rosenthal, J.S. (2001). Markov chains and de-initializing processes. *Scand. J. Stat.* **28** 489–504. MR1858413

[11] Román, J.C. (2012). Convergence analysis of block Gibbs samplers for Bayesian general linear mixed models. Ph.D. thesis, Dept. Statistics, Univ. Florida.

[12] Román, J.C. and Hobert, J.P. (2012). Convergence analysis of the Gibbs sampler for Bayesian general linear mixed models with improper priors. *Ann. Statist.* **40** 2823–2849. MR3097961

[13] Román, J.C. and Hobert, J.P. (2015). Geometric ergodicity of Gibbs samplers for Bayesian general linear mixed models with proper priors. *Linear Algebra Appl.* **473** 54–77. MR3338325

[14] Román, J.C., Hobert, J.P. and Presnell, B. (2014). On reparametrization and the Gibbs sampler. *Statist. Probab. Lett.* **91** 110–116. MR3208124

[15] Searle, S.R., Casella, G. and McCulloch, C.E. (1992). *Variance Components*. *Wiley Series in Probability and Mathematical Statistics*: *Applied Probability and Statistics*. New York: Wiley. MR1190470

[16] Sun, D., Tsutakawa, R.K. and He, Z. (2001). Propriety of posteriors with improper priors in hierarchical linear mixed models. *Statist. Sinica* **11** 77–95. MR1820002

Received February 2015