Fast Monte Carlo Markov chains
for Bayesian shrinkage models with random effects

Tavis Abrahamsen
Department of Statistical Science
Duke University
tavis.abrahamsen@duke.edu

James P. Hobert
Department of Statistics
University of Florida
jhobert@stat.ufl.edu

November 17, 2017

Abstract

When performing Bayesian data analysis using a general linear mixed model, the resulting posterior density is almost always analytically intractable. However, if proper conditionally conjugate priors are used, there is a simple two-block Gibbs sampler that is geometrically ergodic in nearly all practical settings, including situations where $p > n$ (Abrahamsen and Hobert, 2017). Unfortunately, the (conditionally conjugate) multivariate normal prior on β does not perform well in the high-dimensional setting where $p \gg n$. In this paper, we consider an alternative model in which the multivariate normal prior is replaced by the normal-gamma shrinkage prior developed by Griffin and Brown (2010). This change leads to a much more complex posterior density, and we develop a simple MCMC algorithm for exploring it. This algorithm, which has both deterministic and random scan components, is easier to analyze than the more obvious three-step Gibbs sampler. Indeed, we prove that the new algorithm is geometrically ergodic in most practical settings.

Key words and phrases. Bayesian shrinkage prior; geometric drift condition; geometric ergodicity; high dimensional inference; large p - small n; Markov chain Monte Carlo

1 Introduction

The general linear mixed model (or variance components model) is one of the most frequently applied statistical models. It takes the form

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

where Y is an observable $n \times 1$ data vector, X and $\{Z_i\}_{i=1}^m$ are known matrices, β is an unknown $p \times 1$ vector of regression coefficients, $\{u_i\}_{i=1}^m$ are independent random vectors whose elements represent the various levels of the random factors in the model, and $e \sim \mathrm{N}_n(0, \lambda_0^{-1} I)$. The random vectors e and $u := \left(u_1^T \ u_2^T \cdots u_m^T\right)^T$ are independent, and $u \sim \mathrm{N}_q(0, \Lambda^{-1})$, where u_i is $q_i \times 1$, $q = q_1 + \cdots + q_m$, and $\Lambda = \oplus_{i=1}^m \lambda_i I_{q_i}$. (We assume throughout that $n \geqslant 2$, and that $q_i \geqslant 2$ for each $i = 1, 2, \ldots, m$.) For a book-length treatment of this model and its many applications, see McCulloch et al. (2008).

In the Bayesian setting, prior distributions are assigned to β and $\lambda := (\lambda_0 \ \lambda_1 \ \cdots \ \lambda_m)^T$. Unfortunately, any non-trivial prior leads to an intractable posterior density. However, if β and λ are assigned conditionally conjugate priors, then a simple two-block Gibbs sampler can be used to explore the resulting posterior density. In particular, if we assign a multivariate normal prior to β, and independent gamma priors to the precision parameters, then, letting $\theta = (\beta^T \ u^T)^T$, it is easily shown that given observed data y, $\theta | \lambda, y$ is multivariate normal, and $\lambda | \theta, y$ is a product of independent gammas. (Since u is unobservable, it is treated like a parameter.) Convergence rate results for this block Gibbs sampler can be found in Abrahamsen and Hobert (2017).

Now consider this Bayesian mixed model in the high-dimensional setting where $p \gg n$. This situation can arise, e.g., in genetics and neuroscience where variability between subjects is most appropriately handled with random effects (see, e.g., Fazli et al., 2011; Rohart et al., 2014). While the model described above could certainly be used in this setting, the multivariate normal prior on β is really not suitable. Indeed, when $p \gg n$, it is often assumed that β is sparse, i.e., that many components of β are zero. Unfortunately, the multivariate normal prior for β will *shrink* the estimated coefficients towards zero, but not enough to produce an (approximately) sparse estimate of β. Additionally, when the components of β have varying magnitudes, the estimates of the "large" components will be shrunk disproportionately compared to the estimates of the "small" components. Below we propose an alternative prior for β that is tailored to the high-dimensional setting.

The well-known Bayesian interpretation of the lasso (involving iid Laplace priors for the regression parameters) has led to a flurry of recent research concerning the development of prior distributions for regression parameters (in linear models *without* random effects) that yield posterior distributions with high posterior probability around sparse values of β. These prior distributions are called *continuous shrinkage priors* and the corresponding statistical models are referred to as *Bayesian shrinkage models* (see, e.g., Bhattacharya et al. (2013, 2015), Griffin and Brown (2010), Polson and Scott (2010), and Park and Casella (2008)). One such Bayesian shrinkage model is the so-called normal-gamma model of Griffin and Brown (2010), which is given by

$$\begin{aligned} Y|\beta,\tau,\lambda_0 &\sim \mathrm{N}_n(X\beta, \lambda_0^{-1} I_n) \\ \beta|\tau,\lambda_0 &\sim \mathrm{N}_p(0, \lambda_0^{-1} D_\tau)\ , \end{aligned} \tag{1}$$

where $\tau := (\tau_1 \cdots \tau_p)^T$ and D_τ is a diagonal matrix with the τ_is on the diagonal. The precision parameter, λ_0, and the components of τ are assumed to be *a priori* independent with $\lambda_0 \sim \mathrm{Gamma}(a,b)$ and $\tau_i \overset{\text{iid}}{\sim} \mathrm{Gamma}(c,d)$ for $i = 1,\ldots,p$. When $c = 1$, this model becomes the Bayesian lasso model introduced by Park and Casella (2008). We note that Bhattacharya et al. (2013, 2015) show that, in terms of frequentist optimality, the Bayesian lasso has sub-optimal prior concentration rates in that it does not place sufficient mass around sparse values of β. Alternatively, shrinkage priors that have singularities at zero and robust tails (such as in the normal-gamma model with $c < 1/2$), have been shown to perform well in empirical studies.

In this paper, we propose and analyze an MCMC algorithm for a new Bayesian general linear mixed model in which the standard multivariate normal prior on β is replaced with the continuous shrinkage prior from the normal-gamma model. Our high-dimensional Bayesian general linear mixed model is defined as follows

$$\begin{aligned} Y|\beta,u,\tau,\lambda &\sim \mathrm{N}_n\left(X\beta + \sum_{i=1}^m Z_i u_i, \lambda_0^{-1} I_n\right) \\ \beta|u,\tau,\lambda &\sim \mathrm{N}_p(0, \lambda_0^{-1} D_\tau) \\ u|\tau,\lambda &\sim \mathrm{N}_q(0, \Lambda^{-1})\ , \end{aligned} \tag{2}$$

where λ and τ are a priori independent with $\lambda_i \overset{\text{ind}}{\sim} \mathrm{Gamma}(a_i,b_i)$, for $i = 0,1,\ldots,m$, and $\tau_i \overset{\text{iid}}{\sim} \mathrm{Gamma}(c,d)$ for $i = 1,\ldots,p$. This model can be considered a Bayesian analog of the frequentist, high dimensional mixed model developed by Schelldorfer et al. (2011). (Of course, it can also be viewed as a mixed version of the normal-gamma shrinkage model.) A similar sparse Bayesian linear mixed model has been proposed by Zhou et al. (2013) for polygenic modeling. They assume a "spike and slab" prior consisting of a mixture of a point mass at 0 and a normal distribution for the components of β. However, it is well-known that spike and slab priors lead to MCMC algorithms that have convergence problems, especially when p is large (Bhattacharya et al. (2015); Polson and Scott (2010)).

Recall that $\theta = (\beta^T \; u^T)^T$, and let $\pi(\theta, \lambda, \tau|y)$ denote the posterior density associated with model (2). This density is highly intractable and Bayesian inference requires MCMC, which should, of course, be based on a geometrically ergodic Monte Carlo Markov chain (see, e.g. Flegal et al., 2008; Jones and Hobert, 2001; Roberts and Rosenthal, 1998). As we show in Section 2, the full conditional densities $\pi_1(\theta|\lambda, \tau, y)$, $\pi_2(\lambda|\theta, \tau, y)$, and $\pi_3(\tau|\theta, \lambda, y)$ all have standard forms, which means that there is a simple three-block Gibbs sampler available. Unfortunately, we have been unable to establish a convergence rate for this Gibbs sampler (in either deterministic or random scan form). However, we have been able to prove that a related *hybrid* algorithm does converge at a geometric rate. The invariant density of our Markov chain is $\pi(\theta, \lambda|y) := \int_{\mathbb{R}_+^p} \pi(\theta, \lambda, \tau|y)\, d\tau$. Let $r \in (0, 1)$ be fixed, and denote the Markov chain by $\{(\theta_k, \lambda_k)\}_{k=0}^{\infty}$. If the current state is $(\theta_k, \lambda_k) = (\theta, \lambda)$, then we simulate the new state, $(\theta_{k+1}, \lambda_{k+1})$, using the following three-step procedure.

Iteration $k + 1$ of the hybrid algorithm:

1. Draw $\tau \sim \pi_3(\cdot|\theta, \lambda, y)$, and, independently, $U \sim \text{Uniform}(0, 1)$.

2a. If $U < r$, set $(\theta_{k+1}, \lambda_{k+1}) = (\theta', \lambda)$ where $\theta' \sim \pi_1(\cdot|\lambda, \tau, y)$.

2b. Otherwise, set $(\theta_{k+1}, \lambda_{k+1}) = (\theta, \lambda')$ where $\lambda' \sim \pi_2(\cdot|\theta, \tau, y)$.

At each iteration, this sampler first performs a deterministic update of τ from its full conditional distribution. Next, a random scan update is performed, which updates either θ or λ with probability r and $(1-r)$, respectively. This sampler is no more difficult to implement than the three-block Gibbs sampler. Moreover, it is straightforward to show that the Markov chain driving this algorithm is reversible with respect to $\pi(\theta, \lambda|y)$, and that it is Harris ergodic. This algorithm is actually a special case of a more general MCMC algorithm for Bayesian latent data models that was recently developed by Jung (2015).

Our main result provides a set of conditions under which the hybrid Markov chain defined above is geometrically ergodic (as defined by Jones and Hobert (2001, p.319)). Here is the result.

Proposition 1. *The Markov chain,* $\{(\theta_k, \lambda_k)\}_{k=0}^{\infty}$, *is geometrically ergodic for all* $r \in (0, 1)$ *if*

1. $Z := (Z_1 \; Z_2 \cdots Z_m)$ *has full column rank,*

2. $a_0 > \frac{1}{2}\left(\text{rank}(X) - n + (2c+1)p + 2\right)$, *and*

3. $a_i > 1$ *for each* $i \in \{1, 2, \ldots, m\}$.

Note that the conditions of Proposition 1 are quite simple to check. We do require Z to be full column rank, which holds for most basic designs, but there is no such restriction on X, so the result is applicable when $p > n$. While the second condition may become restrictive when $p \gg n$, this can be mitigated to some extent by the fact that the user if free to choose any positive value for the hyperparameter b_0. Indeed, when a large value of a_0 is required to satisfy condition (2), b_0 can be chosen such that the prior mean and variance of λ_0 (which are given by $a_0 b_0^{-1}$ and $a_0 b_0^{-2}$, respectively) have reasonable values.

Abrahamsen and Hobert (2017) established simple, easily-checked sufficient conditions for geometric ergodicity of the two-block Gibbs sampler for the standard Bayesian general linear mixed model that assigns a multivariate normal prior to β. Note, however, that unlike the multivariate normal prior, the continuous shrinkage prior that we use here is hierarchical, which means that an additional set of latent variables (the τ_is) must be managed, and this leads to more complicated MCMC algorithms that are more difficult to analyze. On a related note, Pal and Khare (2014) obtained geometric ergodicity results for the Gibbs sampler developed for the original normal-gamma model (without random effects). But again, adding random effects to the normal-gamma model leads to new MCMC algorithms that are more complex and harder to handle.

The remainder of this paper is organized as follows. Section 2 contains a formal definition of the Markov chain that drives our hybrid sampler. A proof of Proposition 1 is given in Section 3. Finally, a good deal of technical material, such as proofs of the lemmas that are used in the main proof, has been relegated to the Appendices.

2 The Hybrid Sampler

In this section, we formally define the Markov transition function (Mtf) of the hybrid algorithm. We begin with a brief derivation of the conditional densities, π_i, $i = 1, 2, 3$. Let $W = [X \;\; Z]$, $\mathbb{R}_+ = (0, \infty)$, and recall that $\theta = (\beta^T \; u^T)^T$. The posterior density can be expressed up to a constant of proportionality as

$$
\begin{aligned}
\pi(\theta,\tau,\lambda|y) &\propto \pi(y|\beta,u,\tau,\lambda)\pi(\beta|\tau,\lambda)\pi(u|\tau,\lambda)\pi(\tau)\pi(\lambda) \\
&\propto \lambda_0^{n/2} \exp\left\{-\frac{\lambda_0}{2}(y-W\theta)^T(y-W\theta)\right\} \\
&\quad \times \lambda_0^{p/2}\left[\prod_{j=1}^{p}\tau_j^{-1/2}\right]\exp\left\{-\frac{\lambda_0}{2}\beta^T D_\tau^{-1}\beta\right\} \\
&\quad \times \left[\prod_{i=1}^{m}\lambda_i^{q_i/2}\right]\exp\left\{-\frac{1}{2}u^T\Lambda u\right\} \\
&\quad \times \left[\prod_{j=1}^{p}\tau_j^{c-1}e^{-d\tau_j}\, I_{\mathbb{R}_+}(\tau_j)\right]\left[\prod_{i=0}^{m}\lambda_i^{a_i-1}e^{-b_i\lambda_i}\, I_{\mathbb{R}_+}(\lambda_i)\right].
\end{aligned} \tag{3}
$$

We will use (3) to derive the full conditional distributions of θ, τ and λ. First, it is shown in the Appendix that the full conditional distribution of θ is multivariate normal with

$$
\mathrm{E}[\theta|\tau,\lambda,y] = \begin{bmatrix} \lambda_0 T_{\lambda,\tau}^{-1}X^Ty - \lambda_0^2 T_{\lambda,\tau}^{-1}X^TZQ_{\lambda,\tau}^{-1}Z^TM_{\lambda,\tau}y \\ \lambda_0 Q_{\lambda,\tau}^{-1}Z^TM_{\lambda,\tau}y \end{bmatrix}, \tag{4}
$$

and

$$
\mathrm{Var}[\theta|\tau,\lambda,y] = \begin{bmatrix} T_{\lambda,\tau}^{-1} + \lambda_0^2 T_{\lambda,\tau}^{-1}X^TZQ_{\lambda,\tau}^{-1}Z^TXT_{\lambda,\tau}^{-1} & -\lambda_0 T_{\lambda,\tau}^{-1}X^TZQ_{\lambda,\tau}^{-1} \\ -\lambda_0 Q_{\lambda,\tau}^{-1}Z^TXT_{\lambda,\tau}^{-1} & Q_{\lambda,\tau}^{-1} \end{bmatrix}, \tag{5}
$$

where $T_{\lambda,\tau} = \lambda_0(X^TX + D_\tau^{-1})$, $M_{\lambda,\tau} = I - \lambda_0 X T_{\lambda,\tau}^{-1}X^T$, and $Q_{\lambda,\tau} = \lambda_0 Z^T M_{\lambda,\tau} Z + \Lambda$.

Next, it's clear from (3) that the components of λ are conditionally independent, and that each has a gamma distribution. Indeed,

$$
\pi_2(\lambda_0|\theta,\tau,y) \propto \lambda_0^{n/2+p/2+a_0-1} e^{-\lambda_0\left(\frac{||y-W\theta||^2}{2}+\frac{\beta^TD_\tau^{-1}\beta}{2}+b_0\right)}\, I_{\mathbb{R}_+}(\lambda_0)\;, \tag{6}
$$

and, for $i = 1, 2, \ldots, m$,

$$
\pi_2(\lambda_i|\theta,\tau,y) \propto \lambda_i^{q_i/2+a_i-1} e^{-\lambda_i\left(\frac{||u_i||^2}{2}+b_i\right)}\, I_{\mathbb{R}_+}(\lambda_i)\;. \tag{7}
$$

Lastly,

$$
\begin{aligned}
\pi_3(\tau|\theta,\lambda,y) &\propto \left[\prod_{j=1}^{p}\tau_j^{-1/2}\right]\exp\left\{-\frac{\lambda_0}{2}\sum_{j=1}^{p}\frac{\beta_j^2}{\tau_j}\right\}\left[\prod_{j=1}^{p}\tau_j^{c-1}e^{-d\tau_j}\, I_{\mathbb{R}_+}(\tau_j)\right] \\
&= \prod_{j=1}^{p}\tau_j^{(c-1/2)-1}\exp\left\{-\frac{1}{2}\left(\frac{\lambda_0\beta_j^2}{\tau_j}+2d\tau_j\right)\right\}\, I_{\mathbb{R}_+}(\tau_j)\;.
\end{aligned}
$$

Thus, the τ_js are conditionally independent, and

$$\pi_3(\tau_j|\theta,\lambda,y) \propto \tau_j^{(c-1/2)-1} e^{-\frac{1}{2}\left(\lambda_0 \frac{\beta_j^2}{\tau_j} + 2d\tau_j\right)} I_{\mathbb{R}_+}(\tau_j) \; . \tag{8}$$

This brings us to a subtle technical problem that turns out to be very important in our convergence rate analysis. Note that, if $c \leqslant 1/2$ and $\beta_j = 0$, then the right-hand side of (8) is not integrable. Moreover, it is precisely these values of c that yield effective shrinkage priors. Of course, from a simulation standpoint, this technical problem is a non-issue because we will never observe an exact zero from the normal distribution. However, in order to perform a theoretical analysis of the Markov chain, we are obliged to define the Mtf for *all* points in the state space. Our solution is to simply delete the offending points from the state space. (Alternatively, we could make a special definition of the Mtf at the offending points, but this leads to a Markov chain that lacks the *Feller* property (Meyn and Tweedie, 2009, p.124), and this prevents us from employing Meyn and Tweedie's (2009) Lemma 15.2.8.) Thus, we define the state space of our Markov chain to be

$$\mathsf{X} = \left\{ (\theta,\lambda) \in \mathbb{R}^{p+q} \times \mathbb{R}_+^{m+1} : |\beta_j| > 0 \;\text{ for } j = 1,\ldots,p \right\} .$$

Taking the state space to be X instead of $\mathbb{R}^{p+q} \times \mathbb{R}_+^{m+1}$ has no effect on posterior inference because the difference between these two sets is a set of measure zero. However, as will become clear in Section 3, the deleted points do create some complications in the drift analaysis. We note that this particular technical issue has surfaced before (see, e.g., Román and Hobert, 2012), although, in contrast with the current situation, the culprit is typically improper priors.

It's clear from (8) that conditional on θ, λ and y, the distribution of τ_j is GIG$(c-1/2, 2d, \lambda_0\beta_j^2)$, where GIG$(\zeta,\xi,\psi)$ denotes the Generalized Inverse Gaussian distribution with parameters $\zeta \in \mathbb{R}$, $\xi > 0$, and $\psi > 0$. The density is given by

$$f_{\text{GIG}}(x;\zeta,\xi,\psi) = \frac{\xi^{\zeta/2}}{2\psi^{\zeta/2} K_\zeta(\sqrt{\xi\psi})} x^{\zeta-1} e^{-\frac{1}{2}(\xi x + \frac{\psi}{x})} I_{\mathbb{R}_+}(x) \; , \tag{9}$$

where $K_\zeta(\cdot)$ denotes the modified Bessel function of the second kind.

Fix $r \in (0,1)$, and let A denote a measurable set in X. The Mtf of the Markov chain that drives our hybrid algorithm is given by

$$\begin{aligned} P((\theta,\lambda),A) = r \int_{\mathbb{R}^{p+q}} I_A(\tilde\theta,\lambda) \left[\int_{\mathbb{R}_+^p} \pi(\tilde\theta|\tau,\lambda)\, \pi(\tau|\theta,\lambda)\, d\tau \right] d\tilde\theta \\ + (1-r) \int_{\mathbb{R}_+^m} I_A(\theta,\tilde\lambda) \left[\int_{\mathbb{R}_+^p} \pi(\tilde\lambda|\theta,\tau)\, \pi(\tau|\theta,\lambda)\, d\tau \right] d\tilde\lambda \; , \end{aligned}$$

where we (henceforth) omit the dependence on y from the conditional densities for notational convenience. The Appendix contains a proof that the Markov chain defined by P is a Feller chain. In the next section, we prove Proposition 1.

3 Proof of Proposition 1

This section contains a proof of Proposition 1. In particular, we establish a geometric drift condition for our Markov chain, $\{(\theta_k, \lambda_k)\}_{k=0}^{\infty}$, using a drift function, $v : \mathsf{X} \to [0, \infty)$, that is unbounded off compact sets. Since our chain is Feller, geometric ergodicity then follows immediately from Meyn and Tweedie's (2009) Theorem 6.0.1 and Lemma 15.2.8.

3.1 The drift function

Recall that in our model, $\tau_i \overset{\text{iid}}{\sim} \text{Gamma}(c, d)$. The hyperparameter c will play a crucial role in our drift function. Define $\nu : \mathbb{R}_+ \to (0, 1/2]$ as

$$\nu(c) = c\, I_{(0,\frac{1}{2}]}(c) + \min\{1/2, 2c - 1\}\, I_{(\frac{1}{2},\infty)}(c) \; .$$

Now let $\delta = (\delta_1 \cdots \delta_m)^T$, $\eta = (\eta_1 \cdots \eta_m)^T$, and define the drift function as follows

$$\begin{aligned} v(\theta, \lambda) = \; & \alpha_1 \|y - W\theta\|^2 + \alpha_2 \|\beta\|^2 \\ & + \sum_{j=1}^{p} \frac{1}{|\beta_j|^{\nu(c)}} + \sum_{i=1}^{m} \delta_i \|u_i\|^2 \\ & + \alpha_3 \lambda_0 + \alpha_4 \lambda_0^{\nu(c)/2} + \lambda_0^{-1} + \sum_{i=1}^{m} \lambda_i + \sum_{i=1}^{m} \eta_i \lambda_i^{-1} \; , \end{aligned} \tag{10}$$

where $\alpha_1, \alpha_2, \alpha_3, \alpha_4 \in \mathbb{R}_+$ and $\delta, \eta \in \mathbb{R}_+^m$. The values of these constants are to be determined. The Appendix contains a proof that $v(\theta, \lambda)$ is unbounded off compact sets.

Remark 1. *Using X instead of $\mathbb{R}^{p+q} \times \mathbb{R}_+^{m+1}$ as the state space has implications for the construction of the drift function. In particular, since the hyper-planes where $\beta_j = 0$ are not part of X, and we need the drift function to be unbounded off compact sets, we must have a term in the drift function that diverges as $|\beta_j| \to 0$. In fact, this is the only reason why the term $\sum_{j=1}^{p} \frac{1}{|\beta_j|^{\nu(c)}}$ is part of v. Interestingly, Pal and Khare (2014) established their convergence rate results using a drift/minorization argument, which does not require the drift function to be unbounded off compact sets, yet they still require this same term in their drift function. Fortunately, we are able to reuse some of the bounds that they developed.*

Our goal is to demonstrate that

$$\mathrm{E}[v(\tilde{\theta},\tilde{\lambda})|\theta,\lambda] \leqslant \rho\, v(\theta,\lambda) + L\ , \tag{11}$$

for some $\rho \in [0,1)$ and some finite constant L. First, note that

$$\begin{aligned}
\mathrm{E}[v(\tilde{\theta},\tilde{\lambda})|\theta,\lambda] &= r\int_{\mathbb{R}^{p+q}} v(\tilde{\theta},\lambda)\left[\int_{\mathbb{R}_+^p} \pi(\tilde{\theta}|\tau,\lambda)\pi(\tau|\theta,\lambda)d\tau\right] d\tilde{\theta} \\
&\quad + (1-r)\int_{\mathbb{R}_+^m} v(\theta,\tilde{\lambda})\left[\int_{\mathbb{R}_+^p} \pi(\tilde{\lambda}|\theta,\tau)\pi(\tau|\theta,\lambda)\, d\tau\right] d\tilde{\lambda} \\
&= r\mathrm{E}\left[\mathrm{E}[v(\tilde{\theta},\lambda)|\tau,\lambda]\Big|\theta,\lambda\right] + (1-r)\mathrm{E}\left[\mathrm{E}[v(\theta,\tilde{\lambda})|\tau,\theta]\Big|\theta,\lambda\right]\ .
\end{aligned}$$

It follows that

$$\begin{aligned}
\mathrm{E}[v(\tilde{\theta},\tilde{\lambda})|\theta,\lambda] &= r\mathrm{E}\Bigg[\mathrm{E}\Bigg[\alpha_1\|y-W\tilde{\theta}\|^2 + \alpha_2\|\tilde{\beta}\|^2 \\
&\qquad + \sum_{j=1}^p \frac{1}{|\tilde{\beta}_j|^{\nu(c)}} + \sum_{i=1}^m \delta_i\|\tilde{u}_i\|^2\Big|\tau,\lambda\Bigg]\Bigg|\theta,\lambda\Bigg] \\
&\quad + r\left(\alpha_3\lambda_0 + \alpha_4\lambda_0^{\nu(c)/2} + \lambda_0^{-1} + \sum_{i=1}^m \lambda_i + \sum_{i=1}^m \eta_i\lambda_i^{-1}\right) \\
&\quad + (1-r)\left(\alpha_1\|y-W\theta\|^2 + \alpha_2\|\beta\|^2 + \sum_{j=1}^p \frac{1}{|\beta_j|^{\nu(c)}} + \sum_{i=1}^m \delta_i\|u_i\|^2\right) \\
&\quad + (1-r)\mathrm{E}\Bigg[\mathrm{E}\Bigg[\alpha_3\tilde{\lambda}_0 + \alpha_4\tilde{\lambda}_0^{\nu(c)/2} + \tilde{\lambda}_0^{-1} + \sum_{i=1}^m \tilde{\lambda}_i + \sum_{i=1}^m \eta_i\tilde{\lambda}_i^{-1}\Big|\tau,\theta\Bigg]\Bigg|\theta,\lambda\Bigg]\ .
\end{aligned} \tag{12}$$

Remark 2. *The hybrid sampler employs three full conditional densities, yet the expression above contains only two nested expectations. This is due to the random scan step in the hybrid sampler. In contrast, a drift analysis of the deterministic scan Gibbs sampler for this problem involves similar equations having three nested expectations, which greatly complicates the calculations. On the other hand, a drift analysis of the random scan Gibbs sampler involves no nested expectations, which suggests that such an analysis would be relatively easy. However, it turns out to be more difficult than the analysis of the hybrid algorithm.*

Now define

$$h(\tilde{\theta}) = \alpha_1\|y-W\tilde{\theta}\|^2 + \alpha_2\|\tilde{\beta}\|^2 + \sum_{j=1}^p \frac{1}{|\tilde{\beta}_j|^{\nu(c)}} + \sum_{i=1}^m \delta_i\|\tilde{u}_i\|^2\ ,$$

and

$$g(\tilde{\lambda}) = \alpha_3\tilde{\lambda}_0 + \alpha_4\tilde{\lambda}_0^{\nu(c)/2} + \tilde{\lambda}_0^{-1} + \sum_{i=1}^m \tilde{\lambda}_i + \sum_{i=1}^m \eta_i\tilde{\lambda}_i^{-1}\ .$$

In order to bound (12), we need to develop bounds for $\mathrm{E}\Big[\mathrm{E}\big[h(\tilde{\theta})|\tau,\lambda\big]\Big|\theta,\lambda\Big]$ and $\mathrm{E}\Big[\mathrm{E}\big[g(\tilde{\lambda})|\tau,\theta\big]\Big|\theta,\lambda\Big]$. These will be handled separately in the next two subsections.

3.2 A bound on $\mathrm{E}\Big[\mathrm{E}\big[h(\tilde{\theta})|\tau,\lambda\big]\Big|\theta,\lambda\Big]$

Let

$$R_i = [0_{q_i\times q_1}\ldots 0_{q_i\times q_{i-1}}\; I_{q_i\times q_i}\; 0_{q_i\times q_{i+1}}\ldots 0_{q_i\times q_r}],$$

so that $u_i = R_i u$ for $i = 1,\ldots,m$. It's easy to see that

$$\begin{aligned}
\mathrm{E}\left[\|y - W\tilde{\theta}\|^2\Big|\tau,\lambda\right] &= \mathrm{tr}(W\mathrm{Var}[\tilde{\theta}|\tau,\lambda]W^T) + \|y - W\mathrm{E}[\tilde{\theta}|\tau,\lambda]\|^2 \;, \\
\mathrm{E}\left[\|\tilde{\beta}\|^2\Big|\tau,\lambda\right] &= \mathrm{tr}(\mathrm{Var}[\tilde{\beta}|\tau,\lambda]) + \|\mathrm{E}[\tilde{\beta}|\tau,\lambda]\|^2 \;, \text{ and} \\
\mathrm{E}\left[\|\tilde{u}_i\|^2\Big|\tau,\lambda\right] &= \mathrm{tr}(R_i Q_{\lambda,\tau}^{-1} R_i^T) + \|\mathrm{E}[\tilde{u}_i|\tau,\lambda]\|^2 \;.
\end{aligned} \tag{13}$$

Hence,

$$\begin{aligned}
\mathrm{E}\left[\mathrm{E}\left[h(\tilde{\theta})\Big|\tau,\lambda\right]\Big|\theta,\lambda\right] &= \alpha_1 \mathrm{E}\Big[\mathrm{tr}(W\mathrm{Var}[\tilde{\theta}|\tau,\lambda]W^T) + \|y - W\mathrm{E}[\tilde{\theta}|\tau,\lambda]\|^2|\theta,\lambda\Big] \\
&\quad + \alpha_2 \mathrm{E}\Big[\mathrm{tr}(\mathrm{Var}[\tilde{\beta}|\tau,\lambda]) + \|\mathrm{E}[\tilde{\beta}|\tau,\lambda]\|^2|\theta,\lambda\Big] \\
&\quad + \mathrm{E}\left[\mathrm{E}\left[\sum_{j=1}^{p}\frac{1}{|\tilde{\beta}_j|^{\nu(c)}}\Big|\tau,\lambda\right]\Big|\theta,\lambda\right] \\
&\quad + \sum_{i=1}^{m}\delta_i \mathrm{E}\big[\mathrm{tr}(R_i Q_{\lambda,\tau}^{-1} R_i^T) + \|\mathrm{E}[\tilde{u}_i|\tau,\lambda]\|^2|\theta,\lambda\big] \;.
\end{aligned} \tag{14}$$

We will bound (14) using the following lemmas, which are proven in the Appendix.

Lemma 1. *For all $\tau \in \mathbb{R}_+^p$ and $\lambda \in \mathbb{R}_+^{m+1}$,*

(1) $\mathrm{tr}(W\mathrm{Var}[\theta|\tau,\lambda]W^T) \leqslant \mathrm{tr}(XT_{\lambda,\tau}^{-1}X^T) + \mathrm{tr}(ZQ_{\lambda,\tau}^{-1}Z^T)\;,$ *and*

(2) $\mathrm{tr}(XT_{\lambda,\tau}^{-1}X^T) \leqslant \mathrm{rank}(X)\lambda_0^{-1}\;,$

(3) $\mathrm{tr}(ZQ_{\lambda,\tau}^{-1}Z^T) \leqslant \mathrm{tr}(ZZ^T)\sum_{i=1}^m \lambda_i^{-1}\;.$

Lemma 2. *For all $\tau \in \mathbb{R}_+^p$ and $\lambda \in \mathbb{R}_+^{m+1}$,*

$$\|y - W\mathrm{E}[\theta|\tau,\lambda]\|^2 \leqslant 2n\|y\|^2 + 2n^3\|y\|^2 \;.$$

Lemma 3. *For all $\tau \in \mathbb{R}_+^p$ and $\lambda \in \mathbb{R}_+^{m+1}$,* *(1)* $\mathrm{tr}(\mathrm{Var}[\beta|\tau,\lambda]) \leqslant \lambda_0^{-1}\sum_{j=1}^p \tau_j + c^*\,\mathrm{tr}(ZZ^T)\sum_{i=1}^m \lambda_i^{-1}$ *, and*

(2) $\|\mathrm{E}[\beta|\tau,\lambda]\|^2 \leqslant c^* n^2 \|y\|^2 \left(s_{\max}^2 \sum_{j=1}^p \tau_j + 1\right)$ *.*

where $s_{\max}$ is the largest singular value of X and c^* is a finite positive constant.

Lemma 4. *For all $\tau \in \mathbb{R}_+^p$ and $\lambda \in \mathbb{R}_+^{m+1}$,*

$$\mathrm{E}\left[\sum_{j=1}^p \frac{1}{|\tilde{\beta}_j|^{\nu(c)}}\Big|\tau,\lambda\right] \leqslant p\kappa(c) s_{\max}^{\nu(c)}\,\lambda_0^{\nu(c)/2} + \kappa(c)\lambda_0^{\nu(c)/2}\sum_{j=1}^p \frac{1}{\tau_j^{\nu(c)/2}} \ ,$$

where

$$\kappa(c) := \frac{\Gamma\left(\frac{1-\nu(c)}{2}\right) 2^{\frac{1-\nu(c)}{2}}}{\sqrt{2\pi}} \ ,$$

and $s_{\max}$ is the largest singular value of X.

Lemma 5. *Assume that Z has full column rank. For all $\tau \in \mathbb{R}_+^p$, $\lambda \in \mathbb{R}_+^{m+1}$ and $i = 1,\ldots,m$,*

(1) $\mathrm{tr}(R_i Q_{\lambda,\tau}^{-1} R_i^T) \leqslant q_i \lambda_i^{-1}$ *, and*

(2) $||\mathrm{E}[u_i|\tau,\lambda]||^2 \leqslant q_i \mathrm{tr}[(Z^TZ)^{-1}] n^3 \|y\|^2 \left(s_{\max}^2 \sum_{j=1}^p \tau_j + 1\right)$ *.*

Substituting into (14) gives

$$\begin{aligned}
\mathrm{E}\left[\mathrm{E}\left[h(\tilde{\theta})\Big|\tau,\lambda\right]\Big|\ \theta,\lambda\right] \leqslant\ & \alpha_1 \mathrm{rank}(X)\lambda_0^{-1} + \alpha_1 \mathrm{tr}(ZZ^T)\sum_{i=1}^m \lambda_i^{-1} \\
& + \alpha_2\lambda_0^{-1}\sum_{j=1}^p \mathrm{E}\left[\tau_j|\theta,\lambda\right] + \alpha_2 c^* \mathrm{tr}(ZZ^T)\sum_{i=1}^m \lambda_i^{-1} \\
& + \alpha_2 c^* n^2 \|y\|^2 s_{\max}^2 \sum_{j=1}^p \mathrm{E}\left[\tau_j|\theta,\lambda\right] \\
& + p\kappa(c) s_{\max}^{\nu(c)}\,\lambda_0^{\nu(c)/2} \\
& + \kappa(c)\lambda_0^{\nu(c)/2}\sum_{j=1}^p \mathrm{E}\left[\frac{1}{\tau_j^{\nu(c)/2}}\Big|\theta,\lambda\right] \\
& + \sum_{i=1}^m \delta_i q_i \lambda_i^{-1} \\
& + \max_i\{\delta_i\} q\,\mathrm{tr}[(Z^TZ)^{-1}]\|y\|^2 n^3 s_{\max}^2 \sum_{j=1}^p \mathrm{E}\left[\tau_j|\theta,\lambda\right] \\
& + K_0(\alpha_1,\alpha_2,\delta) \ ,
\end{aligned} \tag{15}$$

where

$$K_0(\alpha_1, \alpha_2, \delta) := 2\alpha_1(n\|y\|^2 + n^3\|y\|^2) + \alpha_2 c^* n^2 \|y\|^2 + \mathrm{tr}[(Z^T Z)^{-1}]\|y\|^2 n^3 \sum_{i=1}^{m} \delta_i q_i \ .$$

3.3 A bound on $\mathrm{E}\Big[\mathrm{E}\big[g(\tilde{\lambda})\big|\tau, \theta\big]\Big|\theta, \lambda\Big]$

From (6), we have

$$\begin{aligned} \mathrm{E}[\lambda_0|\theta, \tau] &= \frac{\Gamma\left(\frac{n+p+2a_0+2}{2}\right)}{\Gamma\left(\frac{n+p+2a_0}{2}\right)} \left(\frac{||y - W\theta||^2 + \beta^T D_\tau^{-1}\beta + 2b_0}{2}\right)^{-1} \\ &\leqslant (n + p + 2a_0)\, b_0^{-1} \ , \end{aligned} \tag{16}$$

and

$$\begin{aligned} \mathrm{E}[\lambda_0^{-1}|\theta, \tau] &= \frac{\Gamma\left(\frac{n+p+2a_0-2}{2}\right)}{\Gamma\left(\frac{n+p+2a_0}{2}\right)} \left(\frac{\|y - W\theta\|^2 + \beta^T D_\tau^{-1}\beta + 2b_0}{2}\right) \\ &= \frac{1}{n + p + 2a_0 - 2} \left(\|y - W\theta\|^2 + \beta^T D_\tau^{-1}\beta + 2b_0\right) \ . \end{aligned} \tag{17}$$

Also, from Jensen's inequality and (16),

$$\mathrm{E}\left[\lambda_0^{\nu(c)/2}|\tau, \theta\right] \leqslant \mathrm{E}\left[\lambda_0|\tau, \theta\right]^{\nu(c)/2} \leqslant [(n + p + 2a_0)\, b_0^{-1}]^{\nu(c)} \ . \tag{18}$$

Similarly, from (7), for $i = 1, 2, \dots, m$ we have

$$\mathrm{E}[\lambda_i|\theta, \tau] = \frac{\Gamma\left(\frac{q_i+2a_i+2}{2}\right)}{\Gamma\left(\frac{q_i+2a_i}{2}\right)} \left(\frac{\|u_i\|^2 + 2b_i}{2}\right)^{-1} \leqslant (q_i + 2a_i)\, b_i^{-1} \ , \tag{19}$$

and

$$\begin{aligned} \mathrm{E}[\lambda_i^{-1}|\theta, \tau] &= \frac{\Gamma\left(\frac{q_i+2a_i-2}{2}\right)}{\Gamma\left(\frac{q_i+2a_i}{2}\right)} \left(\frac{\|u_i\|^2 + 2b_i}{2}\right) \\ &= \frac{1}{q_i + 2a_i - 2} \left(\|u_i\|^2 + 2b_i\right) \ . \end{aligned} \tag{20}$$

Conditions (2) and (3) of Proposition 1 imply that all of the above Gamma functions have positive arguments.

From (16)-(20), we have

$$\begin{aligned} \mathrm{E}\left[\mathrm{E}\left[g(\tilde{\lambda})\Big|\tau, \theta\right]\Big|\theta, \lambda\right] \leqslant\ & \frac{1}{n + p + 2a_0 - 2}\|y - W\theta\|^2 \\ & + \frac{1}{n + p + 2a_0 - 2} \sum_{j=1}^{p} \beta_j^2\, \mathrm{E}\left[\tau_j^{-1}|\theta, \lambda\right] \\ & + \sum_{i=1}^{m} \frac{\eta_i}{q_i + 2a_i - 2}\|u_i\|^2 + K_1(\alpha_3, \alpha_4, \eta) \ , \end{aligned} \tag{21}$$

where

$$K_1(\alpha_3, \alpha_4, \eta) := \alpha_3(n + p + 2a_0 + 1)\, b_0^{-1} + \alpha_4[(n + p + 2a_0 + 1)\, b_0^{-1}]^{\nu(c)} + \frac{2b_0}{n + p + 2a_0 - 2} + \sum_{i=1}^{m}(q_i + 2a_i + 1)\, b_i^{-1} + \sum_{i=1}^{m} \frac{2\eta_i b_i}{q_i + 2a_i - 2}.$$

In the next subsection, we bound the right-hand sides of (15) and (21), and then combine these new bounds to get a bound on $\mathrm{E}[v(\tilde{\theta}, \tilde{\lambda})|\theta, \lambda]$.

3.4 A bound on $\mathrm{E}[v(\tilde{\theta}, \tilde{\lambda})|\theta, \lambda]$

If $X \sim \mathrm{GIG}(\zeta, \xi, \psi)$, then

$$\mathrm{E}[X^p] = \frac{\psi^{p/2} K_{\zeta+p}(\sqrt{\xi\psi})}{\xi^{p/2} K_{\zeta}(\sqrt{\xi\psi})} \,. \tag{22}$$

Hence, for all $(\theta, \lambda) \in \mathsf{X}$,

$$\mathrm{E}[\tau_j|\theta, \lambda] = \sqrt{\frac{\lambda_0 \beta_j^2}{2d}}\, \frac{K_{c+\frac{1}{2}}\left(\sqrt{2d\lambda_0\beta_j^2}\right)}{K_{c-\frac{1}{2}}\left(\sqrt{2d\lambda_0\beta_j^2}\right)} \,, \tag{23}$$

$$\mathrm{E}[\tau_j^{-1}|\theta, \lambda] = \sqrt{\frac{2d}{\lambda_0 \beta_j^2}}\, \frac{K_{c-\frac{3}{2}}\left(\sqrt{2d\lambda_0\beta_j^2}\right)}{K_{c-\frac{1}{2}}\left(\sqrt{2d\lambda_0\beta_j^2}\right)} \,, \tag{24}$$

and

$$\mathrm{E}\left[\frac{1}{\tau_j^{\nu(c)/2}}\middle|\theta, \lambda\right] = \left(\frac{2d}{\lambda_0\beta_j^2}\right)^{\nu(c)/4} \frac{K_{c-\frac{1}{2}-\frac{\nu(c)}{2}}\left(\sqrt{2d\lambda_0\beta_j^2}\right)}{K_{c-\frac{1}{2}}\left(\sqrt{2d\lambda_0\beta_j^2}\right)} \,, \tag{25}$$

for $j = 1, 2, \ldots, p$.

Next, we will make use of the following lemmas, which are proved in the Appendix.

Lemma 6. *For all* $(\theta, \lambda) \in \mathsf{X}$*,*

1. $\mathrm{E}[\tau_j|\theta, \lambda] \leqslant \frac{4c+1}{4d} + \frac{\lambda_0\beta_j^2}{2}$ *, and*
2. $\mathrm{E}[\tau_j|\theta, \lambda] \leqslant \frac{c}{d} + \frac{\beta_j^2}{2C} + \frac{\lambda_0 C}{4d}$ *,*

for every $C > 0$*.*

Lemma 7. *For all* $(\theta, \lambda) \in \mathsf{X}$*,*

$$\mathrm{E}[\tau_j^{-1}|\theta, \lambda] \leqslant d + \frac{3}{2\lambda_0\beta_j^2} \,.$$

Lemma 8. *For all* $(\theta, \lambda) \in \mathsf{X}$*,*

$$\mathrm{E}\left[\frac{1}{\tau_j^{\nu(c)/2}}\Big|\theta, \lambda\right] \leqslant M_1 \frac{1}{\lambda_0^{\nu(c)/2}|\beta_j|^{\nu(c)}} + M_2 \ ,$$

where M_1 *is a positive constant such that* $M_1\kappa(c) < 1$*, and* M_2 *is a positive finite constant.*

Applying Lemma 6, Lemma 8 and (15), we have

$$\begin{aligned}
\mathrm{E}\left[\mathrm{E}\left[h(\tilde{\theta})\Big|\tau, \lambda\right]\Big|\theta, \lambda\right] \leqslant\ & \alpha_1 \mathrm{rank}(X)\lambda_0^{-1} + \alpha_1 \mathrm{tr}(ZZ^T)\sum_{i=1}^{m}\lambda_i^{-1} \\
& + \alpha_2\lambda_0^{-1}\sum_{j=1}^{p}\left(\frac{4c+1}{4d} + \frac{\lambda_0\beta_j^2}{2}\right) + \alpha_2 c^* \mathrm{tr}(ZZ^T)\sum_{i=1}^{m}\lambda_i^{-1} \\
& + \alpha_2 c^* n^2 \|y\|^2 s_{\max}^2 \sum_{j=1}^{p}\left(\frac{c}{d} + \frac{\beta_j^2}{2C_1} + \frac{\lambda_0 C_1}{4d}\right) \\
& + p\kappa(c) s_{\max}^{\nu(c)}\lambda_0^{\nu(c)/2} \\
& + \kappa(c)\lambda_0^{\nu(c)/2}\sum_{j=1}^{p}\left[M_1\frac{1}{\lambda_0^{\nu(c)/2}|\beta_j|^{\nu(c)}} + M_2\right] \\
& + \sum_{i=1}^{m}\delta_i q_i \lambda_i^{-1} \\
& + \max_i\{\delta_i\} C_3 \sum_{j=1}^{p}\left(\frac{c}{d} + \frac{\beta_j^2}{2C_2} + \frac{\lambda_0 C_2}{4d}\right) \\
& + K_0(\alpha_1, \alpha_2, \delta),
\end{aligned}$$

where C_1 and C_2 are positive constants and $C_3 = q\,\mathrm{tr}[(Z^TZ)^{-1}]\|y\|^2 n^3 s_{\max}^2$. Hence,

$$
\begin{aligned}
\mathrm{E}\left[\mathrm{E}\left[h(\tilde{\theta})\middle|\tau,\lambda\right]\middle|\theta,\lambda\right] \leqslant\ & \alpha_1 \mathrm{rank}(X)\lambda_0^{-1} + \alpha_1 \mathrm{tr}(ZZ^T)\sum_{i=1}^m \lambda_i^{-1} \\
& + \alpha_2 \frac{p(4c+1)}{4d}\lambda_0^{-1} + \alpha_2 \frac{\|\beta\|^2}{2} \\
& + \alpha_2 c^* \mathrm{tr}(ZZ^T)\sum_{i=1}^m \lambda_i^{-1} \\
& + \alpha_2 c^* n^2 \|y\|^2 s_{\max}^2 \left(\frac{\|\beta\|^2}{2C_1} + \frac{pC_1}{4d}\lambda_0\right) \\
& + p\kappa(c)\left(s_{\max}^{\nu(c)} + M_2\right)\lambda_0^{\nu(c)/2} \\
& + \kappa(c) M_1 \sum_{j=1}^p \frac{1}{|\beta_j|^{\nu(c)}} + \sum_{i=1}^m \delta_i q_i \lambda_i^{-1} \\
& + \max_i\{\delta_i\} C_3 \left(\frac{\|\beta\|^2}{2C_2} + \frac{pC_2}{4d}\lambda_0\right) \\
& + K_0'(\alpha_1, \alpha_2, \delta),
\end{aligned}
\tag{26}
$$

where

$$
K_0'(\alpha_1,\alpha_2,\delta) := K_0(\alpha_1,\alpha_2,\delta) + \frac{pc}{d}\left(\alpha_2 c^* n^2 \|y\|^2 s_{\max}^2 + C_3 \max_i\{\delta_i\}\right).
$$

Next, using Lemma 7 and (21), we have

$$
\begin{aligned}
\mathrm{E}\left[\mathrm{E}\left[g(\tilde{\lambda})\middle|\tau,\theta\right]\middle|\theta,\lambda\right] \leqslant\ & \frac{1}{n+p+2a_0-2}\|y - W\theta\|^2 \\
& + \frac{1}{n+p+2a_0-2}\left(d\|\beta\|^2 + \frac{3p}{2\lambda_0}\right) \\
& + \sum_{i=1}^m \frac{\eta_i}{q_i + 2a_i - 2}\|u_i\|^2 + K_1(\alpha_3,\alpha_4,\eta).
\end{aligned}
\tag{27}
$$

Substituting (26) and (27) into (12) and rearranging gives

$$
\begin{aligned}
\mathrm{E}[v(\tilde{\theta},\tilde{\lambda})|\theta,\lambda] \leqslant \alpha_1(1-r)&\left[1+\frac{1}{\alpha_1(n+p+2a_0-2)}\right]\|y-W\theta\|^2 \\
&+\alpha_2\left[\frac{r}{2}+\frac{rc^* n^2\|y\|^2 s_{\max}^2}{2C_1}+\frac{r\max_i\{\delta_i\}C_3}{2\alpha_2 C_2}\right. \\
&\qquad\qquad\left.+\ (1-r)+\frac{(1-r)d}{\alpha_2(n+p+2a_0-2)}\right]\|\beta\|^2 \\
&+[1-r(1-\kappa(c)M_1)]\sum_{j=1}^{p}\frac{1}{|\beta_j|^{\nu(c)}} \\
&+\sum_{i=1}^{m}\delta_i(1-r)\left[1+\frac{\eta_i}{\delta_i(q_i+2a_i-2)}\right]\|u_i\|^2 \\
&+\alpha_3\left[\frac{r\alpha_2 c^* n^2\|y\|^2 s_{\max}^2 pC_1}{4\alpha_3 d}+\frac{r\max_i\{\delta_i\}pC_2C_3}{4\alpha_3 d}+r\right]\lambda_0 \\
&+\alpha_4 r\left[1+\frac{p\kappa(c)\left(s_{\max}^{\nu(c)}+M_2\right)}{\alpha_4}\right]\lambda_0^{\nu(c)/2} \\
&+\left[r\alpha_1\mathrm{rank}(X)+\frac{r\alpha_2 p(4c+1)}{4d}+r\right. \\
&\qquad\qquad\left.+\frac{(1-r)\,3p}{2(n+p+2a_0-2)}\right]\lambda_0^{-1} \\
&+r\sum_{i=1}^{m}\lambda_i \\
&+\sum_{i=1}^{m}\eta_i\left[\frac{r\alpha_1\mathrm{tr}(ZZ^T)}{\eta_i}+\frac{r\alpha_2 c^*\mathrm{tr}(ZZ^T)}{\eta_i}+\frac{r\delta_i q_i}{\eta_i}+r\right]\lambda_i^{-1} \\
&+L(\alpha_1,\alpha_2,\alpha_3,\alpha_4,\delta,\eta,r)\ ,
\end{aligned}
\tag{28}
$$

where

$$
L(\alpha_1,\alpha_2,\alpha_3,\alpha_4,\delta,\eta,r) := rK_0'(\alpha_1,\alpha_2,\delta)+(1-r)K_1(\alpha_3,\alpha_4,\eta)\ .
$$

3.5 The final step

Fix $r \in (0,1)$ and note that (aside from L) the terms of (28) agree with the terms of (10), except that each term in (28) has an extra *constant* factor (coefficient). Therefore, we can establish that (11) holds by demonstrating the existence of $\delta, \eta \in \mathbb{R}_+^m$ and $\alpha_1, \alpha_2, \alpha_3, \alpha_4, C_1, C_2 \in \mathbb{R}_+$ such that all of these coefficients are simultaneously less than 1. Moreover, if the chain is geometrically ergodic for at least one $r \in (0,1)$ then it is geometrically ergodic for all $r \in (0,1)$ (Jones et al., 2014; Jung, 2015). Thus, we can treat r as another free parameter. (A similar analysis was performed in Johnson and Jones (2015).)

We begin by noting that two of the coefficients are always less than 1. Indeed, the coefficient of $\sum_{i=1}^m \lambda_i$ is just r, and the coefficient of $\sum_{j=1}^p \frac{1}{|\beta_j|^{\nu(c)}}$ is $[1 - r(1 - \kappa(c)M_1)]$, which is less than 1 since, by Lemma 8, $0 < \kappa(c)M_1 < 1$. Therefore, it suffices to show that we can identify $\delta, \eta \in \mathbb{R}_+^m$ and α_1, α_2, α_3, α_4, C_1, $C_2 \in \mathbb{R}_+$ such that the following seven inequalities all hold simultaneously:

$$\rho_1(\alpha_1, r) := (1 - r)\left[1 + \frac{1}{\alpha_1(n + p + 2a_0 - 2)}\right] < 1\ , \tag{29}$$

$$\begin{aligned}\rho_2(\alpha_2, \delta, C_1, C_2, r) := \frac{r}{2} + \frac{rc^* n^2 \|y\|^2 s_{\max}^2}{2C_1} + \frac{r \max_i\{\delta_i\} C_3}{2\alpha_2 C_2} \\ + (1 - r) + \frac{(1 - r)d}{\alpha_2(n + p + 2a_0 - 2)} < 1\ ,\end{aligned} \tag{30}$$

$$\rho_{3i}(\delta_i, \eta_i, r) := (1 - r)\left[1 + \frac{\eta_i}{\delta_i(q_i + 2a_i - 2)}\right] < 1\ , \quad \text{for } i = 1, \dots, m\ , \tag{31}$$

$$\begin{aligned}\rho_4(\alpha_2, \alpha_3, \delta, C_1, C_2, r) := \frac{r\alpha_2 c^* n^2 \|y\|^2 s_{\max}^2 p C_1}{4\alpha_3 d} \\ + \frac{r \max_i\{\delta_i\} p C_2 C_3}{4\alpha_3 d} + r < 1\ ,\end{aligned} \tag{32}$$

$$\rho_5(\alpha_4, r) := r\left[1 + \frac{p\kappa(c)\left(s_{\max}^{\nu(c)} + M_2\right)}{\alpha_4}\right] < 1\ , \tag{33}$$

$$\begin{aligned}\rho_6(\alpha_1, \alpha_2, r) := r\alpha_1 \text{rank}(X) + \frac{r\alpha_2 p(4c + 1)}{4d} \\ + r + \frac{(1 - r)\,3p}{2(n + p + 2a_0 - 2)} < 1\ ,\end{aligned} \tag{34}$$

and

$$\begin{aligned}\rho_{7i}(\alpha_1, \alpha_2, \delta_i, \eta_i, r) := \frac{r\alpha_1 \text{tr}(ZZ^T)}{\eta_i} + \frac{r\alpha_2 c^* \text{tr}(ZZ^T)}{\eta_i} \\ + \frac{r\delta_i q_i}{\eta_i} + r < 1\ , \quad \text{for } i = 1, \dots, m\ .\end{aligned} \tag{35}$$

We now derive a solution. Solving (29) for α_1 gives

$$\alpha_1 > \frac{(1 - r)}{r(n + p + 2a_0 - 2)}\ .$$

Next define

$$\alpha_1^* = \frac{1}{r(n + p + 2a_0 - 2)}\ , \tag{36}$$

so that $\rho_1(\alpha_1^*, r) < 1$ for all $0 < r < 1$. Next, let

$$\alpha_2^* := \frac{2d}{r(n + p + 2a_0 - 2)} \ . \tag{37}$$

Substituting into (30) gives

$$\begin{aligned} \rho_2(\alpha_2^*, \delta, C_1, C_2, r) &= \frac{r}{2} + \frac{rc^* n^2 \|y\|^2 s_{\max}^2}{2C_1} + \frac{r \max_i\{\delta_i\} C_3}{2\alpha_2^* C_2} + (1 - r) + \frac{(1 - r)d}{\alpha_2^*(n + p + 2a_0 - 2)} \\ &= \frac{r}{2} + \frac{rc^* n^2 \|y\|^2 s_{\max}^2}{2C_1} + \frac{r^2 \max_i\{\delta_i\} C_3 (n + p + 2a_0 - 2)}{4dC_2} \\ &\quad + (1 - r) + (1 - r)\frac{r}{2} \\ &= \frac{rc^* n^2 \|y\|^2 s_{\max}^2}{2C_1} + \frac{r^2 \max_i\{\delta_i\} C_3 (n + p + 2a_0 - 2)}{4dC_2} + 1 - \frac{r^2}{2} \ . \end{aligned} \tag{38}$$

Thus, choosing

$$C_1^* > \frac{2c^* n^2 \|y\|^2 s_{\max}^2}{r} \quad \text{and} \quad C_2^* > \frac{\max_i\{\delta_i\} C_3 (n + p + 2a_0 - 2)}{d} \ ,$$

we get

$$\rho_2(\alpha_2^*, \delta, C_1^*, C_2^*, r) < \frac{r^2}{4} + \frac{r^2}{4} + 1 - \frac{r^2}{2} = 1 \ , \tag{39}$$

for all $\delta \in \mathbb{R}_+^m$ and $0 < r < 1$.

Next, using (34), (36) and (37), we get

$$\begin{aligned} \rho_6(\alpha_1^*, \alpha_2^*, r) &= r\alpha_1^* \operatorname{rank}(X) + \frac{r\alpha_2^* \, p(4c + 1)}{4d} + r + \frac{(1 - r)3p}{2(n + p + 2a_0 - 2)} \\ &= \frac{\operatorname{rank}(X)}{n + p + 2a_0 - 2} + \frac{p(4c + 1)}{2(n + p + 2a_0 - 2)} + r + \frac{(1 - r)3p}{2(n + p + 2a_0 - 2)} \\ &< \frac{\operatorname{rank}(X) + 2p(c + 1)}{n + p + 2a_0 - 2} + r \ , \end{aligned} \tag{40}$$

and from condition (2) of Proposition 1,

$$0 < \frac{\operatorname{rank}(X) + 2p(c + 1)}{n + p + 2a_0 - 2} < \frac{\operatorname{rank}(X) + 2p(c + 1)}{n + p + 2\left(\frac{\operatorname{rank}(X) - n + (2c+1)p) + 2}{2}\right) - 2} = 1 \ .$$

Thus, for

$$r < 1 - \frac{\operatorname{rank}(X) + 2p(c + 1)}{n + p + 2a_0 - 2} \ , \tag{41}$$

$\rho_6(\alpha_1^*, \alpha_2^*, r) < 1$.

Next, solving (31) for δ_i gives

$$\delta_i > \frac{(1 - r)\eta_i}{r(q_i + 2a_i - 2)} \quad \text{for } i = 1, \ldots, m \ . \tag{42}$$

Hence, by defining

$$\delta_i^* = \frac{\eta_i}{r(q_i + 2a_i - 1)}, \quad i = 1, \ldots, m \; , \tag{43}$$

it follows that $\rho_{3i}(\delta_i^*, \eta_i, r) < 1$ for all $\eta_i > 0$ and $r \in (0,1)$, $i = 1, \ldots, m$.

Using equations (35), (36), (37) and (43) we get

$$\begin{aligned} \rho_{7i}(\alpha_1^*, \alpha_2^*, \delta_i^*, \eta_i, r) &= \frac{r\alpha_1^* \, \mathrm{tr}(ZZ^T)}{\eta_i} + \frac{r\alpha_2^* \, c^* \mathrm{tr}(ZZ^T)}{\eta_i} + \frac{r\delta_i^* \, q_i}{\eta_i} + r \\ &= \frac{\mathrm{tr}(ZZ^T)}{\eta_i(n + p + 2a_0 - 2)} + \frac{2dc^* \mathrm{tr}(ZZ^T)}{\eta_i(n + p + 2a_0 - 2)} + \frac{q_i}{q_i + 2a_i - 2} + r \; , \end{aligned} \tag{44}$$

for $i = 1, \ldots, m$, and from condition (3) of Proposition 1,

$$1 - \frac{q_i}{q_i + 2a_i - 2} = \frac{2a_i - 2}{q_i + 2a_i - 2} > 0 \; .$$

Thus, for

$$\eta_i^* > \frac{q_i + 2a_i - 2}{2a_i - 2} \left[\frac{\mathrm{tr}(ZZ^T)}{(n + p + 2a_0 - 2)} + \frac{2dc^* \mathrm{tr}(ZZ^T)}{(n + p + 2a_0 - 2)} \right] \; , \tag{45}$$

it follows that

$$0 < \frac{\mathrm{tr}(ZZ^T)}{\eta_i^* \, (n + p + 2a_0 - 2)} + \frac{2dc^* \mathrm{tr}(ZZ^T)}{\eta_i^* \, (n + p + 2a_0 - 2)} + \frac{q_i}{q_i + 2a_i - 2} < 1 \; ,$$

for $i = 1, \ldots, m$. Hence, $\rho_{7i}(\alpha_1^*, \alpha_2^*, \delta_i^*, \eta_i^*, r) < 1$ when

$$r < 1 - \frac{\mathrm{tr}(ZZ^T)}{\eta_i^*(n + p + 2a_0 - 2)} + \frac{2dc^* \mathrm{tr}(ZZ^T)}{\eta_i^*(n + p + 2a_0 - 2)} + \frac{q_i}{q_i + 2a_i - 2} \; , \quad i = 1, \ldots, m \; .$$

Next, solving (32) for α_3 and (33) for α_4 gives

$$\alpha_3 > \frac{1}{1 - r} \left[\frac{r\alpha_2 c^* n^2 \|y\|^2 s_{\max}^2 p C_1}{4d} + \frac{r \max_i\{\delta_i\} p C_2 C_3}{4d} \right], \tag{46}$$

and

$$\alpha_4 > \frac{r}{1 - r} \left[p\kappa(c) \left(s_{\max}^{\nu(c)} + M_2 \right) \right], \tag{47}$$

respectively. Let α_3^* satisfy (46), then $\rho_4(\alpha_2, \alpha_3^*, \delta, C_1, C_2, r) < 1$ for all $\alpha_2, C_1, C_2 > 0$, $\delta \in \mathbb{R}_+^p$ and $r \in (0,1)$. Now, let α_4^* satisfy (47), then $\rho_5(\alpha_4^*, r) < 1$ for all $r \in (0,1)$.

Lastly, choose r^* such that

$$\begin{aligned} r^* < 1 - \max \Bigg\{ & \frac{\mathrm{rank}(X) + 2p(c + 1)}{n + p + 2a_0 - 2}, \\ & \max_{1 \leqslant i \leqslant m} \left\{ \frac{\mathrm{tr}(ZZ^T)}{\eta_i^*(n + p + 2a_0 - 2)} + \frac{2dc^* \mathrm{tr}(ZZ^T)}{\eta_i^*(n + p + 2a_0 - 2)} + \frac{q_i}{q_i + 2a_i - 2} \right\} \Bigg\} \; . \end{aligned}$$

Let $\delta^* := (\delta_1^* \; \delta_2^* \; \ldots \; \delta_m^*)^T$ and $\eta^* := (\eta_1^* \; \eta_2^* \; \ldots \; \eta_m^*)^T$. The inequalities (29) - (35) are then satisfied for δ^*, η^*, α_1^*, α_2^*, α_3^*, α_4^*, C_1^*, C_2^* and r^*. Therefore,

$$\begin{aligned}
\mathrm{E}[v(\tilde{\theta}, \tilde{\lambda})|\theta, \lambda] \leqslant\; & \rho_1(\alpha_1^*, r^*)\, \alpha_1^* \|y - W\theta\|^2 \\
& + \rho_2(\alpha_2^*, \delta^*, C_1^*, C_2^*, r^*)\, \alpha_2^* \|\beta\|^2 \\
& + [1 - r^*(1 - \kappa(c) M_1)] \sum_{j=1}^{p} \frac{1}{|\beta_j|^{\nu(c)}} \\
& + \sum_{i=1}^{r} \rho_{3i}(\delta_i^*, \eta_i^*, r^*)\, \delta_i^* \|u_i\|^2 \\
& + \rho_4(\alpha_2^*, \alpha_3^*, \delta^*, C_1^*, C_2^*, r^*)\, \alpha_3^* \lambda_0 \\
& + \rho_5(\alpha_4^*, r^*)\, \alpha_4^* \lambda_0^{\nu(c)/2} \\
& + \rho_6(\alpha_1^*, \alpha_2^*, r^*) \lambda_0^{-1} \\
& + r^* \sum_{i=1}^{r} \lambda_i + \sum_{i=1}^{r} \rho_{7i}(\alpha_1^*, \alpha_2^*, \delta_i^*, \eta_i^*)\, \eta_i^* \lambda_i^{-1} \\
& + L(\alpha_1^*, \alpha_2^*, \alpha_3^*, \alpha_4^*, \delta^*, \eta^*, r^*) \; .
\end{aligned} \tag{48}$$

To formally complete the argument, let ρ^* denote the maximum of all of the coefficients. Then,

$$\begin{aligned}
\mathrm{E}[v(\tilde{\theta}, \tilde{\lambda})|\theta, \lambda] & \leqslant \rho^* \Bigg(\alpha_1^* \|y - W\theta\|^2 + \alpha_2^* \|\beta\|^2 + \sum_{j=1}^{p} \frac{1}{|\beta_j|^{\nu(c)}} + \sum_{i=1}^{r} \delta_i^* \|u_i\|^2 \\
& \qquad + \alpha_3^* \lambda_0 + \alpha_4^* \lambda_0^{\nu(c)/2} + \lambda_0^{-1} + \sum_{i=1}^{r} \lambda_i + \sum_{i=1}^{r} \eta_i^* \lambda_i^{-1} \Bigg) \\
& \quad + L(\alpha_1^*, \alpha_2^*, \alpha_3^*, \alpha_4^*, \delta^*, \eta^*, r^*) \\
& = \rho^* v(\theta, \lambda) + L(\alpha_1^*, \alpha_2^*, \alpha_3^*, \delta^*, \eta^*, r^*) \; ,
\end{aligned}$$

where $\rho^* < 1$ and $L(\alpha_1^*, \alpha_2^*, \alpha_3^*, \delta^*, \eta^*, r^*) < \infty$. Therefore, the chain is geometrically ergodic for $r = r^*$, which implies that it is geometrically ergodic for all $r \in (0, 1)$. This proves Proposition 1.

Acknowledgment. The second author was supported by NSF Grant DMS-15-11945.

Appendices

A Derivation of $\pi(\theta|\tau,\lambda)$

From (3),

$$\begin{aligned}\pi(\theta|\tau,\lambda,y) &\propto \exp\left\{-\frac{\lambda_0}{2}(y-W\theta)^T(y-W\theta)\right\}\exp\left\{-\frac{\lambda_0}{2}\beta^T D_\tau^{-1}\beta\right\}\exp\left\{-\frac{1}{2}u^T\Lambda u\right\}\\ &\propto \exp\left\{-\frac{\lambda_0}{2}(y-W\theta)^T(y-W\theta)\right\}\exp\left\{-\frac{1}{2}\theta^T C\theta\right\},\end{aligned}$$

where

$$C=\left[\begin{array}{cc}\lambda_0 D_\tau^{-1} & 0\\ 0 & \Lambda\end{array}\right].$$

Thus,

$$\begin{aligned}\pi(\theta|\tau,\lambda,y) &\propto \exp\left\{-\frac{1}{2}\left[\theta^T(\lambda_0 W^T W+C)\theta-2\theta^T(\lambda_0 W^T y)\right]\right\}\\ &\propto \exp\left\{-\frac{1}{2}\left[\theta^T(\lambda_0 W^T W+C)\theta-2\theta^T(\lambda_0 W^T W+C)(\lambda_0 W^T W+C)^{-1}(\lambda_0 W^T y)\right]\right\}.\end{aligned}$$

Therefore, conditional on τ, λ and y, θ is multivariate normal with mean $(\lambda_0 W^T W+C)^{-1}(\lambda_0 W^T y)$ and covariance matrix $(\lambda_0 W^T W+C)^{-1}$. It is now left for us to compute these two values. From the definition of W and C,

$$(\lambda_0 W^T W+C)^{-1}=\left[\begin{array}{cc}\lambda_0 X^T X+\lambda_0 D_\tau^{-1} & \lambda_0 X^T Z\\ \lambda_0 Z^T X & \lambda_0 Z^T Z+\Lambda\end{array}\right]^{-1}.$$

We will make use of the following inverse formula for block matrices

$$\left[\begin{array}{cc}A & B\\ C & D\end{array}\right]^{-1}=\left[\begin{array}{cc}A^{-1}+A^{-1}B(D-CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D-CA^{-1}B)^{-1}\\ -(D-CA^{-1}B)^{-1}CA^{-1} & (D-CA^{-1}B)^{-1}\end{array}\right]. \quad \text{(A.1)}$$

Let $T_{\lambda,\tau}=\lambda_0(X^T X+D_\tau^{-1})$, $M_{\lambda,\tau}=I-\lambda_0 X T_{\lambda,\tau}^{-1}X^T$, and $Q_{\lambda,\tau}=\lambda_0 Z^T M_{\lambda,\tau}Z+\Lambda$. Then,

$$(\lambda_0 W^T W+C)^{-1}=\left[\begin{array}{cc}T_{\lambda,\tau} & \lambda_0 X^T Z\\ \lambda_0 Z^T X & \lambda_0 Z^T Z+\Lambda\end{array}\right]^{-1}:=\left[\begin{array}{cc}\Omega_{11} & \Omega_{12}\\ \Omega_{21} & \Omega_{22}\end{array}\right]:=\Omega\,.$$

From (A.1),

$$\begin{aligned}\Omega_{22} &= \Big[\lambda_0 Z^T Z + \Lambda - (\lambda_0 Z^T X) T_{\lambda,\tau}^{-1} (\lambda_0 X^T Z)\Big]^{-1} \\ &= \Big[\lambda_0 Z^T (I - \lambda_0 X T_{\lambda,\tau}^{-1} X^T) Z + \Lambda\Big] \\ &= \big[\lambda_0 Z^T M_{\lambda,\tau} Z + \Lambda\big]^{-1} \\ &= Q_{\lambda,\tau}^{-1} \ .\end{aligned}$$

Next,

$$\begin{aligned}\Omega_{11} &= T_{\lambda,\tau}^{-1} + T_{\lambda,\tau}^{-1}(\lambda_0 X^T Z)[\lambda_0 Z^T Z + \Lambda - (\lambda_0 Z^T X) T_{\lambda,\tau}^{-1}(\lambda_0 X^T Z)]^{-1}(\lambda_0 Z^T X) T_{\lambda,\tau}^{-1}, \\ &= T_{\lambda,\tau}^{-1} + \lambda_0^2 T_{\lambda,\tau}^{-1} X^T Z Q_{\lambda,\tau}^{-1} Z^T X T_{\lambda,\tau}^{-1} \ .\end{aligned}$$

Similarly,

$$\Omega_{12} = -T_{\lambda,\tau}^{-1}(\lambda_0 X^T Z)[\lambda_0 Z^T Z + \Lambda - (\lambda_0 Z^T X) T_{\lambda,\tau}^{-1}(\lambda_0 X^T Z)]^{-1} = -\lambda_0 T_{\lambda,\tau}^{-1} X^T Z Q_{\lambda,\tau}^{-1} \ ,$$

and

$$\Omega_{21} = -\lambda_0 Q_{\lambda,\tau}^{-1} Z^T X T_{\lambda,\tau}^{-1} \ .$$

Hence,

$$\mathrm{Var}[\theta|\tau,\lambda,y] = \begin{bmatrix} T_{\lambda,\tau}^{-1} + \lambda_0^2 T_{\lambda,\tau}^{-1} X^T Z Q_{\lambda,\tau}^{-1} Z^T X T_{\lambda,\tau}^{-1} & -\lambda_0 T_{\lambda,\tau}^{-1} X^T Z Q_{\lambda,\tau}^{-1} \\ -\lambda_0 Q_{\lambda,\tau}^{-1} Z^T X T_{\lambda,\tau}^{-1} & Q_{\lambda,\tau}^{-1} \end{bmatrix}. \tag{A.2}$$

Now we just need to compute the conditional mean of θ given τ, λ and y. Notice that

$$\mathrm{E}[\theta|\tau,\lambda,y] = \Omega(\lambda_0 W^T y) = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix} \begin{bmatrix} \lambda_0 X^T y \\ \lambda_0 Z^T y \end{bmatrix},$$

and

$$\begin{aligned}\Omega_{11}(\lambda_0 X^T y) + \Omega_{12}(\lambda_0 Z^T y) &= \lambda_0 (T_{\lambda,\tau}^{-1} + \lambda_0^2 T_{\lambda,\tau}^{-1} X^T Z Q_{\lambda,\tau}^{-1} Z^T X T_{\lambda,\tau}^{-1}) X^T y - \lambda_0(\lambda_0 T_{\lambda,\tau}^{-1} X^T Z Q_{\lambda,\tau}^{-1}) Z^T y \\ &= \lambda_0 T_{\lambda,\tau}^{-1} X^T y + \lambda_0^3 T_{\lambda,\tau}^{-1} X^T Z Q_{\lambda,\tau}^{-1} Z^T X T_{\lambda,\tau}^{-1} X^T y - \lambda_0^2 T_{\lambda,\tau}^{-1} X^T Z Q_{\lambda,\tau}^{-1} Z^T y \\ &= \lambda_0 T_{\lambda,\tau}^{-1} X^T y - \lambda_0^2 T_{\lambda,\tau}^{-1} X^T Z Q_{\lambda,\tau}^{-1} Z^T (I - \lambda_0 X T_{\lambda,\tau}^{-1} X^T) y \\ &= \lambda_0 T_{\lambda,\tau}^{-1} X^T y - \lambda_0^2 T_{\lambda,\tau}^{-1} X^T Z Q_{\lambda,\tau}^{-1} Z^T M_{\lambda,\tau} y \ .\end{aligned}$$

Additionally, we have

$$\begin{aligned}\Omega_{21}(\lambda_0 X^T y) + \Omega_{22}(\lambda_0 Z^T y) &= \lambda_0(-\lambda_0 Q_{\lambda,\tau}^{-1} Z^T X T_{\lambda,\tau}^{-1}) X^T y + \lambda_0 Q_{\lambda,\tau}^{-1} Z^T y \\ &= \lambda_0 Q_{\lambda,\tau}^{-1} Z^T (I - \lambda_0 X T_{\lambda,\tau}^{-1} X^T) y \\ &= \lambda_0 Q_{\lambda,\tau}^{-1} Z^T M_{\lambda,\tau} y \; .\end{aligned}$$

Hence

$$\mathrm{E}[\theta|\tau,\lambda,y] = \begin{bmatrix} \lambda_0 T_{\lambda,\tau}^{-1} X^T y - \lambda_0^2 T_{\lambda,\tau}^{-1} X^T Z Q_{\lambda,\tau}^{-1} Z^T M_{\lambda,\tau} y \\ \lambda_0 Q_{\lambda,\tau}^{-1} Z^T M_{\lambda,\tau} y \end{bmatrix} . \tag{A.3}$$

Thus, the full conditional distribution of θ is multivariate normal with mean and covariance matrix given by (4) and (5), respectively.

B Proof that the Markov chain $\{(\theta_k, \lambda_k)\}_{k=0}^{\infty}$ is Feller

To prove that the Markov chain generated by the hybrid sampler is a Feller chain, we must show that for each open set O, $P(\cdot, O)$ is a lower semi-continuous function on X. Let (θ_l, λ_l) be a sequence in X converging to $(\theta, \lambda) \in \mathsf{X}$. Then,

$$\begin{aligned}\liminf_{l\to\infty} P((\theta_l,\lambda_l),O) &\geqslant \liminf_{l\to\infty} \; r \int_{\mathbb{R}^{p+q}} I_O(\tilde{\theta},\lambda_l) \left[\int_{\mathbb{R}_+^p} \pi(\tilde{\theta}|\tau,\lambda_l)\pi(\tau|\theta_l,\lambda_l) d\tau \right] d\tilde{\theta} \\ &\qquad + \liminf_{l\to\infty} (1-r) \int_{\mathbb{R}_+^m} I_O(\theta_l,\tilde{\lambda}) \left[\int_{\mathbb{R}_+^p} \pi(\tilde{\lambda}|\theta_l,\tau,)\pi(\tau|\theta_l,\lambda_l)\, d\tau \right] d\tilde{\lambda} \\ &\geqslant r \int_{\mathbb{R}^{p+q}} I_O(\tilde{\theta},\lambda) \left[\int_{\mathbb{R}_+^p} \liminf_{l\to\infty} \; \pi(\tilde{\theta}|\tau,\lambda_l)\pi(\tau|\theta_l,\lambda_l) d\tau \right] d\tilde{\theta} \\ &\qquad + (1-r) \int_{\mathbb{R}_+^m} I_O(\theta,\tilde{\lambda}) \left[\int_{\mathbb{R}_+^p} \liminf_{l\to\infty} \; \pi(\tilde{\lambda}|\theta_l,\tau)\pi(\tau|\theta_l,\lambda_l)\, d\tau \right] d\tilde{\lambda} \\ &= r \int_{\mathbb{R}^{p+q}} I_O(\tilde{\theta},\lambda) \left[\int_{\mathbb{R}_+^p} \pi(\tilde{\theta}|\tau,\lambda)\pi(\tau|\theta,\lambda)\, d\tau \right] d\tilde{\theta} \\ &\qquad + (1-r) \int_{\mathbb{R}_+^m} I_O(\theta,\tilde{\lambda}) \left[\int_{\mathbb{R}_+^p} \pi(\tilde{\lambda}|\theta,\tau)\pi(\tau|\theta,\lambda)\, d\tau \right] d\tilde{\lambda} \\ &= P((\theta,\lambda),O) \; ,\end{aligned}$$

where the penultimate equality follows from the fact that all three conditional densities are continuous in the conditioning variables (Abrahamsen, 2016).

C Proof that $v(\theta, \lambda)$ is unbounded off compact sets

Recall that the drift function $v(\theta, \lambda)$ is given by

$$v(\theta, \lambda) = \alpha_1 \|y - W\theta\|^2 + \alpha_2 \|\beta\|^2 + \sum_{j=1}^{p} \frac{1}{|\beta_j|^{\nu(c)}} + \sum_{i=1}^{m} \delta_i \|u_i\|^2$$
$$+ \alpha_3 \lambda_0 + \alpha_4 \lambda_0^{\nu(c)/2} + \lambda_0^{-1} + \sum_{i=1}^{m} \lambda_i + \sum_{i=1}^{m} \eta_i \lambda_i^{-1} .$$

We need to show that this function is unbounded off compact sets; that is, we must demonstrate that, for every $d \in \mathbb{R}$, the set

$$S_d := \{(\theta, \lambda) \in \mathsf{X} : v(\theta, \lambda) \leqslant d\} ,$$

is compact. Let d be such that S_d is nonempty (otherwise S_d is trivially compact). Since $v(\theta, \lambda)$ is continuous on X, S_d is a closed set. Now define

$$A_j = \left\{ \beta_j \in \mathbb{R}\backslash\{0\} : \alpha_2 \beta_j^2 + \frac{1}{|\beta_j|^{\nu(c)}} \leqslant d \right\}, \quad j = 1, \ldots, p ,$$
$$B_i = \{u_i \in \mathbb{R} : \delta_i u_i^2 \leqslant d\} , \quad i = 1, \ldots, m ,$$
$$C_0 = \{\lambda_0 \in \mathbb{R}_+ : \alpha_3 \lambda_0 + \lambda_0^{-1} + \alpha_4 \lambda_0^{\nu(c)/2} \leqslant d\} ,$$
$$C_i = \{\lambda_i \in \mathbb{R}_+ : \lambda_i + \eta_i \lambda_i^{-1} \leqslant d\}, \quad i = 1, \ldots, m .$$

All of the above sets are closed and bounded, and thus the set

$$T_d := \bigtimes_{j=1}^{p} A_i \times \bigtimes_{j=1}^{m} B_j \times \bigtimes_{i=0}^{m} C_i ,$$

is a compact set in X. Since, S_d is a closed set and $S_d \subseteq T_d$, it follows that S_d is a compact set in X.

D Proofs of the lemmas

D.1 Preliminary results

We begin by introducing some notation and stating a few necessary facts about non-negative definite matrices. Note that if C is a non-negative definite matrix then $\mathrm{tr}(C)$ is non-negative. If $A, B \in \mathbb{R}^{n\times n}$ are symmetric matrices such that $B - A$ is non-negative definite, we write $A \preceq B$. Similarly, if $B - A$ is positive definite, we write $A \prec B$. Additionally, if $A \preceq B$, then $\mathrm{tr}(A) \leqslant \mathrm{tr}(B)$. Furthermore, if A and B

are positive definite matrices then $A \preceq B \Leftrightarrow B^{-1} \preceq A^{-1}$. Lastly, for a matrix D we let $\|D\|$ represent the Frobenius norm of the matrix $\|D\| := \sqrt{\text{tr}(D^T D)}$.

We will also require the singular value decomposition of several matrices in our proofs, thus it will be helpful to establish some common notation. For a matrix $A \in \mathbb{R}^{n\times m}$, let $k_A = \text{rank}(A) \leqslant \min\{n, m\}$ and denote the singular value decomposition of A by $U_A \Gamma_A V_A^T$, where U_A and V_A are orthogonal matrices of dimension n and m, respectively, and

$$\Gamma_A := \begin{bmatrix} \Gamma_A^* & 0_{k_A,\, m-k_A} \\ 0_{n-k_A,\, k_A} & 0_{n-k_A,\, m-k_A} \end{bmatrix},$$

where $\Gamma_A^* := \text{diag}\{\gamma_{A\,1}, \ldots, \gamma_{A\,k}\}$. The values $\gamma_{A\,1}, \ldots, \gamma_{A\,k}$ are the singular values of A, which are strictly positive. We denote $\gamma_{A\max}$ as the largest singular value of A. Lastly, in an abuse of notation, $\gamma_{A\,i}^2 := 0$ whenever $i > k_A$.

In order to prove Lemmas 1 - 5, we will need the following results.

Lemma 9. *For all $\tau \in \mathbb{R}_+^p$ and $\lambda \in \mathbb{R}_+^p$,*

(1) $M_{\lambda,\tau} = U R_{\lambda,\tau} U^T$ where $U \in \mathbb{R}^{n\times n}$ is an orthogonal matrix and $R_{\lambda,\tau} :=$ diag$\{r_1, r_2, \ldots, r_n\}$ where $0 < r_i \leqslant 1$ for $i = 1, 2, \ldots, n$.

(2) $0 \prec (\tau_{\max} s_{\max}^2 + 1)^{-1} I \preceq M_{\lambda,\tau} \preceq I$, where $\tau_{\max} := \max\{\tau_1, \ldots, \tau_p\}$ and $s_{\max}$ is the largest singular value of the matrix X.

(3) $||M_{\lambda,\tau}|| \leqslant \sqrt{n}$.

Proof of Lemma 9. This proof is similar to the proof of Lemmas 4 and 5 of Román and Hobert (2015). Recall,

$$\begin{aligned} M_{\lambda,\tau} &= I - \lambda_0 X T_{\lambda,\tau}^{-1} X^T \\ &= I - \lambda_0 X[\lambda_0(X^T X + D_\tau^{-1})]^{-1} X^T \\ &= I - X(X^T X + D_\tau^{-1})^{-1} X^T \\ &= I - X D_\tau^{1/2}(D_\tau^{1/2} X^T X D_\tau^{1/2} + I_p)^{-1} D_\tau^{1/2} X^T. \end{aligned}$$

Let $B := XD_\tau^{1/2}$ and let $U_B\Gamma_B V_B^T$ be the singular value decomposition of B. Then

$$\begin{aligned}
M_{\lambda,\tau} &= I - U_B\Gamma_B V_B^T (V_B\Gamma_B^T U_B^T U_B \Gamma_B V_B^T + I_p)^{-1} V_B \Gamma_B^T U_B^T \\
&= I - U_B\Gamma_B(\Gamma_B^T\Gamma_B + I_p)^{-1}\Gamma_B^T U_B^T \\
&= U_B(I - \Gamma_B(\Gamma_B^T\Gamma_B + I_p)^{-1}\Gamma_B^T)U_B^T \\
&= U_B R_{\lambda,\tau} U_B^T \ ,
\end{aligned}$$

where $R_{\lambda,\tau} :=$ diag$\{r_1, r_2, \ldots, r_n\}$, with

$$r_i = 1 - \frac{\gamma_{B\,i}^2}{\gamma_{B\,i}^2 + 1} = \frac{1}{\gamma_{B\,i}^2 + 1} \ ,$$

where $\gamma_{B\,i}^2 = 0$ for $i > k_B$, and $0 < r_i \leqslant 1$ for $i = 1, \ldots, n$. This proves (1).

Next, let $\tau_{\max} := \max\{\tau_1, \tau_2, \ldots, \tau_p\}$, and notice that

$$\begin{aligned}
X(X^TX + D_\tau^{-1})^{-1}X^T &\preceq X(X^TX + \tau_{\max}^{-1}I_p)^{-1}X^T \\
&= U_X\Gamma_X V_X^T(V_X\Gamma_X^T U_X^T U_X\Gamma_X V_X^T + \tau_{\max}^{-1}I_p)^{-1}V_X\Gamma_X^T U_X^T \\
&= U_X\Gamma_X(\Gamma_X^T\Gamma_X + \tau_{\max}^{-1}I_p)^{-1}\Gamma_X^T U_X^T \ .
\end{aligned}$$

Thus,

$$\begin{aligned}
M_{\lambda,\tau} &= I - \lambda_0 X T_{\lambda,\tau}^{-1} X^T \\
&= I - X(X^TX + D_\tau^{-1})^{-1}X^T \\
&\succeq I - U_X\Gamma_X(\Gamma_X^T\Gamma_X + \tau_{\max}^{-1}I_p)^{-1}\Gamma_X^T U_X^T \\
&= U_X[I - \Gamma_X(\Gamma_X^T\Gamma_X + \tau_{\max}^{-1}I_p)^{-1}\Gamma_X^T]U_X^T \ .
\end{aligned}$$

Notice that $[I - \Gamma_X(\Gamma_X^T\Gamma_X + \tau_{\max}^{-1}I_p)^{-1}\Gamma_X^T]$ is a diagonal matrix whose entries are given by

$$1 - \frac{\gamma_{X\,i}^2}{\gamma_{X\,i}^2 + \tau_{\max}^{-1}} = \frac{\tau_{\max}^{-1}}{\gamma_{X\,i}^2 + \tau_{\max}^{-1}} = \frac{1}{\gamma_{X\,i}^2\,\tau_{\max} + 1} \geqslant \frac{1}{s_{\max}^2\tau_{\max} + 1} \ ,$$

for all $i = 1, \ldots, n$ where $s_{\max}$ is the largest singular value of X.

Thus,

$$0 \prec \left(\frac{1}{\tau_{\max}s_{\max}^2 + 1}\right) I = U_X\left(\frac{1}{\tau_{\max}s_{\max}^2 + 1}\right)U_X^T \preceq M_{\lambda,\tau} = U_B R_{\lambda,\tau}\,U_B^T \preceq U_B U_B^T = I,$$

which proves (2).

Lastly,

$$\|M_{\lambda,\tau}\|^2 = \mathrm{tr}(M_{\lambda,\tau}^2) = \mathrm{tr}(U_B R_{\lambda,\tau} U_B^T U_B R_{\lambda,\tau} U_B^T) = \mathrm{tr}(U_B R_{\lambda,\tau}^2 U_B^T) \leqslant \mathrm{tr}(U_B U_B^T) = n \ ,$$

and therefore $||M_{\lambda,\tau}|| \leqslant \sqrt{n}$. □

Lemma 10. *For all $\lambda \in \mathbb{R}_+^{m+1}$ and all $\tau \in \mathbb{R}_+^p$,*

(1) $\lambda_0 M_{\lambda,\tau}^{1/2} Z Q_{\lambda,\tau}^{-1} Z^T M_{\lambda,\tau}^{1/2} \preceq I_n$.

(2) $\|\lambda_0 M_{\lambda,\tau}^{1/2} Z Q_{\lambda,\tau}^{-1} Z^T M_{\lambda,\tau}^{1/2}\|^2 \leqslant n$.

Proof of Lemma 10. Notice that

$$\begin{aligned}
\lambda_0 M_{\lambda,\tau}^{1/2} Z Q_{\lambda,\tau}^{-1} Z^T M_{\lambda,\tau}^{1/2} &= \lambda_0 M_{\lambda,\tau}^{1/2} Z(\lambda_0 Z^T M_{\lambda,\tau} Z + \Lambda)^{-1} Z^T M_{\lambda,\tau}^{1/2} \\
&= M_{\lambda,\tau}^{1/2} Z(Z^T M_{\lambda,\tau} Z + \lambda_0^{-1}\Lambda)^{-1} Z^T M_{\lambda,\tau}^{1/2} \\
&= M_{\lambda,\tau}^{1/2} Z\Lambda^{-1/2}(\Lambda^{-1/2} Z^T M_{\lambda,\tau}^{1/2} M_{\lambda,\tau}^{1/2} Z\Lambda^{-1/2} + \lambda_0^{-1} I)^{-1}\Lambda^{-1/2} Z^T M_{\lambda,\tau}^{1/2} \ ,
\end{aligned} \tag{D.1}$$

Let $A := M_{\lambda,\tau}^{1/2} Z\Lambda^{-1/2}$ and let $U_A\Gamma_A V_A^T$ represent the singular value decomposition of A. Then, from (D.1),

$$\begin{aligned}
\lambda_0 M_{\lambda,\tau}^{1/2} Z Q_{\lambda,\tau}^{-1} Z^T M_{\lambda,\tau}^{1/2} &= U_A\Gamma_A V_A^T (V_A\Gamma_A^T U_A^T U_A \Gamma_A V_A^T + \lambda_0^{-1} I)^{-1} V_A \Gamma_A^T U_A^T \\
&= U_A\Gamma_A(\Gamma_A^T\Gamma_A + \lambda_0^{-1} I)^{-1}\Gamma_A^T U_A^T \ ,
\end{aligned} \tag{D.2}$$

where $\Gamma_A(\Gamma_A^T\Gamma_A + \lambda_0^{-1} I)^{-1}\Gamma_A^T$ is a diagonal matrix whose elements are given by

$$\frac{\gamma_{A\,i}^2}{\gamma_{A\,i}^2 + \lambda_0^{-1}} \leqslant 1 \ \text{ for } i = 1, \ldots, n \ . \tag{D.3}$$

Thus,

$$\lambda_0 M_{\lambda,\tau}^{1/2} Z Q_{\lambda,\tau}^{-1} Z^T M_{\lambda,\tau}^{1/2} = U_A\Gamma_A(\Gamma_A^T\Gamma_A + \lambda_0^{-1} I)^{-1}\Gamma_A^T U_A^T \preceq U_A U_A^T = I_n \ , \tag{D.4}$$

which proves (1).

Next, from (D.2) and (D.3)

$$\begin{aligned}
\|\lambda_0 M_{\lambda,\tau}^{1/2} Z Q_{\lambda,\tau}^{-1} Z^T M_{\lambda,\tau}^{1/2}\|^2 &= \mathrm{tr}[(\lambda_0 M_{\lambda,\tau}^{1/2} Z Q_{\lambda,\tau}^{-1} Z^T M_{\lambda,\tau}^{1/2})^T (\lambda_0 M_{\lambda,\tau}^{1/2} Z Q_{\lambda,\tau}^{-1} Z^T M_{\lambda,\tau}^{1/2})] \\
&= \mathrm{tr}[(U_A \Gamma_A (\Gamma_A^T \Gamma_A + \lambda_0^{-1} I)^{-1} \Gamma_A^T U_A^T)^T U_A \Gamma_A (\Gamma_A^T \Gamma_A + \lambda_0^{-1} I)^{-1} \Gamma_A^T U_A^T] \\
&= \mathrm{tr}[U_A \Gamma_A (\Gamma_A^T \Gamma_A + \lambda_0^{-1} I)^{-1} \Gamma_A^T U_A^T U_A \Gamma_A (\Gamma_A^T \Gamma_A + \lambda_0^{-1} I)^{-1} \Gamma_A^T U_A^T] \\
&= \mathrm{tr}[U_A \Gamma_A (\Gamma_A^T \Gamma_A + \lambda_0^{-1} I)^{-1} \Gamma_A^T \Gamma_A (\Gamma_A^T \Gamma_A + \lambda_0^{-1} I)^{-1} \Gamma_A^T U_A^T] \\
&= \mathrm{tr}[\Gamma_A (\Gamma_A^T \Gamma_A + \lambda_0^{-1} I)^{-1} \Gamma_A^T \Gamma_A (\Gamma_A^T \Gamma_A + \lambda_0^{-1} I)^{-1} \Gamma_A^T U_A^T U_A] \\
&= \mathrm{tr}[\Gamma_A (\Gamma_A^T \Gamma_A + \lambda_0^{-1} I)^{-1} \Gamma_A^T \Gamma_A (\Gamma_A^T \Gamma_A + \lambda_0^{-1} I)^{-1} \Gamma_A^T] \\
&= \sum_{i=1}^{n} \left(\frac{\gamma_{A\,i}^2}{\gamma_{A\,i}^2 + \lambda_0^{-1}} \right)^2 \\
&\leqslant n \ ,
\end{aligned} \tag{D.5}$$

which proves (2). □

Lemma 11. *For all* $\lambda \in \mathbb{R}_+^{m+1}$ *and* $\tau \in \mathbb{R}_+^p$

(1) $\|\lambda_0 T_{\lambda,\tau}^{-1} X^T\|^2 < \infty$.

(2) $\|I - \lambda_0 Z Q_{\lambda,\tau}^{-1} Z^T M_{\lambda,\tau}\|^2 \leqslant n^2 \left(s_{\max}^2 \sum_{j=1}^p \tau_j + 1 \right)$.

Proof of Lemma 11.

$$\begin{aligned}
\|\lambda_0 T_{\lambda,\tau}^{-1} X^T\|^2 &= \|(X^T X + D_\tau^{-1})^{-1} X^T\|^2 \\
&= \mathrm{tr}(X (X^T X + D_\tau^{-1})^{-2} X^T) \\
&= \sum_{i=1}^{n} e_i^T X (X^T X + D_\tau^{-1})^{-2} X^T e_i \\
&= \sum_{i=1}^{n} \|(X^T X + D_\tau^{-1})^{-1} X^T e_i\|^2 \ ,
\end{aligned} \tag{D.6}$$

where $e_i \in \mathbb{R}^n$, $i = 1, \ldots, n$, are the standard unit vectors. Let x_i represent the ith column vector of X^T.

For a given i,

$$
\begin{aligned}
\|(X^TX + D_\tau^{-1})^{-1}X^Te_i\|^2 &= \|(X^TX + D_\tau^{-1})^{-1}x_i\|^2 \\
&= \|(\sum_{j=1}^{n} x_jx_j^T + D_\tau^{-1})^{-1}x_i\|^2 \\
&= \|(x_ix_i^T + \sum_{i\neq j}^{n} x_jx_j^T + D_\tau^{-1})^{-1}x_i\|^2 \\
&= \|(x_ix_i^T + \sum_{i\neq j}^{n} x_jx_j^T + (D_\tau^{-1} - \tau_\bullet^{-1}I_p) + \tau_\bullet^{-1}I_p)^{-1}x_i\|^2 \ ,
\end{aligned}
\tag{D.7}
$$

where $\tau_\bullet^{-1} = (\tau_1 + \ldots + \tau_p)^{-1}$. Define the vectors $t_1, t_2, \ldots, t_{n+p} \in \mathbb{R}^p$ such that $t_j = x_j$ for $j = 1, \ldots, n$, and $t_{n+k} = e_k$, $k = 1, \ldots, p$, where e_k are the standard unit vectors in $\mathbb{R}^p$. Next, define the positive constants $w_1, w_2, \ldots, w_{n+p} \in \mathbb{R}_+$ as follows

$$
w_l = \begin{cases}
\tau_\bullet^{-1} & l = i \ , \\
1 & l \neq i, 1 \leqslant l \leqslant p \ , \\
\tau_1^{-1} - \tau_\bullet^{-1} & l = n+1 \ , \\
\vdots & \vdots \\
\tau_p^{-1} - \tau_\bullet^{-1} & l = n+p \ .
\end{cases}
$$

Thus,

$$
\begin{aligned}
\|(X^TX + D_\tau^{-1})^{-1}X^Te_i\|^2 &= \|(t_it_i^t + \sum_{l\in\{1,2,\ldots,n\}\setminus\{i\}} w_lt_lt_l^T + \sum_{l=n+1}^{n+p} w_lt_lt_l^T + w_iI_p)^{-1}t_i\|^2 \\
&= t_i^T \left(t_it_i^t + \sum_{l\in\{1,2,\ldots,n\}\setminus\{i\}} w_lt_lt_l^T + \sum_{l=n+1}^{n+p} w_lt_lt_l^T + w_iI_p \right)^{-2} t_i \\
&\leqslant \sup_{w\in\mathbb{R}^{n+p}} t_i^T \left(t_it_i^t + \sum_{l\in\{1,2,\ldots,n\}\setminus\{i\}} w_lt_lt_l^T + \sum_{l=n+1}^{n+p} w_lt_lt_l^T + w_iI_p \right)^{-2} t_i \\
&:= C_i^* < \infty \ , \quad i = 1, \ldots, n \ ,
\end{aligned}
\tag{D.8}
$$

where the last inequality follows from Khare and Hobert (2011). Thus

$$
\|\lambda_0 T_{\lambda,\tau}^{-1} X^T\|^2 \leqslant \sum_{i=1}^{n} C_i^* < \infty \ , \tag{D.9}
$$

which proves (1).

Next,

$$
\begin{aligned}
\|I-\lambda_0 ZQ_{\lambda,\tau}^{-1}Z^TM_{\lambda,\tau}\|^2 &= \mathrm{tr}[(I-\lambda_0 ZQ_{\lambda,\tau}^{-1}Z^TM_{\lambda,\tau})^T(I-\lambda_0 ZQ_{\lambda,\tau}^{-1}Z^TM_{\lambda,\tau})] \\
&= \mathrm{tr}[I-\lambda_0 ZQ_{\lambda,\tau}^{-1}Z^TM_{\lambda,\tau}-\lambda_0 M_{\lambda,\tau}ZQ_{\lambda,\tau}^{-1}Z^T+\lambda_0^2 M_{\lambda,\tau}ZQ_{\lambda,\tau}^{-1}Z^TZQ_{\lambda,\tau}^{-1}Z^TM_{\lambda,\tau}] \\
&= \mathrm{tr}(I)-2\lambda_0\mathrm{tr}(ZQ_{\lambda,\tau}^{-1}Z^TM_{\lambda,\tau})+\lambda_0^2\mathrm{tr}(M_{\lambda,\tau}ZQ_{\lambda,\tau}^{-1}Z^TZQ_{\lambda,\tau}^{-1}Z^TM_{\lambda,\tau}) \\
&= n-2\lambda_0\mathrm{tr}(M_{\lambda,\tau}^{1/2}ZQ_{\lambda,\tau}^{-1}Z^TM_{\lambda,\tau}^{1/2})+\lambda_0^2\mathrm{tr}(M_{\lambda,\tau}ZQ_{\lambda,\tau}^{-1}Z^TZQ_{\lambda,\tau}^{-1}Z^TM_{\lambda,\tau}) \\
&\leqslant n+\lambda_0^2\mathrm{tr}(M_{\lambda,\tau}ZQ_{\lambda,\tau}^{-1}Z^TZQ_{\lambda,\tau}^{-1}Z^TM_{\lambda,\tau})\ .
\end{aligned}
\tag{D.10}
$$

From (2) of Lemma 9,

$$
\begin{aligned}
M_{\lambda,\tau}ZQ_{\lambda,\tau}^{-1}Z^TZQ_{\lambda,\tau}^{-1}Z^TM_{\lambda,\tau} &= M_{\lambda,\tau}ZQ_{\lambda,\tau}^{-1}Z^TM_{\lambda,\tau}^{1/2}M_{\lambda,\tau}^{-1}M_{\lambda,\tau}^{1/2}ZQ_{\lambda,\tau}^{-1}Z^TM_{\lambda,\tau} \\
&= M_{\lambda,\tau}^{1/2}(M_{\lambda,\tau}^{1/2}ZQ_{\lambda,\tau}^{-1}Z^TM_{\lambda,\tau}^{1/2})M_{\lambda,\tau}^{-1}(M_{\lambda,\tau}^{1/2}ZQ_{\lambda,\tau}^{-1}Z^TM_{\lambda,\tau}^{1/2})M_{\lambda,\tau}^{1/2} \\
&\preceq (\tau_{\max}s_{\max}^2+1)M_{\lambda,\tau}^{1/2}(M_{\lambda,\tau}^{1/2}ZQ_{\lambda,\tau}^{-1}Z^TM_{\lambda,\tau}^{1/2})^2M_{\lambda,\tau}^{1/2}\ .
\end{aligned}
\tag{D.11}
$$

From (2) of Lemma 10 and (2) of Lemma 9,

$$
\begin{aligned}
\lambda_0^2\mathrm{tr}(M_{\lambda,\tau}ZQ_{\lambda,\tau}^{-1}Z^TZQ_{\lambda,\tau}^{-1}Z^TM_{\lambda,\tau}) &\leqslant (\tau_{\max}s_{\max}^2+1)\mathrm{tr}[\lambda_0^2M_{\lambda,\tau}^{1/2}(M_{\lambda,\tau}^{1/2}ZQ_{\lambda,\tau}^{-1}Z^TM_{\lambda,\tau}^{1/2})^2M_{\lambda,\tau}^{1/2}] \\
&= (\tau_{\max}s_{\max}^2+1)\|\lambda_0(M_{\lambda,\tau}^{1/2}ZQ_{\lambda,\tau}^{-1}Z^TM_{\lambda,\tau}^{1/2})M_{\lambda,\tau}^{1/2}\|^2 \\
&\leqslant (\tau_{\max}s_{\max}^2+1)\|\lambda_0M_{\lambda,\tau}^{1/2}ZQ_{\lambda,\tau}^{-1}Z^TM_{\lambda,\tau}^{1/2}\|^2\|M_{\lambda,\tau}^{1/2}\|^2 \\
&\leqslant (\tau_{\max}s_{\max}^2+1)n\,\|M_{\lambda,\tau}^{1/2}\|^2 \\
&= (\tau_{\max}s_{\max}^2+1)n\,\mathrm{tr}(M_{\lambda,\tau}) \\
&\leqslant (\tau_{\max}s_{\max}^2+1)n\,\mathrm{tr}(I) \\
&= n^2(s_{\max}^2\tau_{\max}+1) \\
&\leqslant n^2\left(s_{\max}^2\sum_{j=1}^p\tau_j+1\right)\ ,
\end{aligned}
\tag{D.12}
$$

which proves (2). □

D.2 Proof of Lemma 1

Lemma 1. *For all* $\tau \in \mathbb{R}_+^p$ *and* $\lambda \in \mathbb{R}_+^{m+1}$,

(1) $\mathrm{tr}(W\mathrm{Var}[\tilde{\theta}|\tau,\lambda]W^T) \leqslant \mathrm{tr}(XT_{\lambda,\tau}^{-1}X^T)+\mathrm{tr}(ZQ_{\lambda,\tau}^{-1}Z^T)$,

(1) $\mathrm{tr}(XT_{\lambda,\tau}^{-1}X^T) \leqslant \mathrm{rank}(X)\lambda_0^{-1}$, and

(2) $\mathrm{tr}(ZQ_{\lambda,\tau}^{-1}Z^T) \leqslant \mathrm{tr}(ZZ^T)\sum_{i=1}^m \lambda_i^{-1}$.

Proof of Lemma 1. First,

$$
\begin{aligned}
W\mathrm{Var}[\tilde{\theta}|\tau,\lambda]W^T &= [X\ Z]\begin{bmatrix} T_{\lambda,\tau}^{-1} + \lambda_0^2 T_{\lambda,\tau}^{-1}X^TZQ_{\lambda,\tau}^{-1}Z^TXT_{\lambda,\tau}^{-1} & -\lambda_0 T_{\lambda,\tau}^{-1}X^TZQ_{\lambda,\tau}^{-1} \\ -\lambda_0 Q_{\lambda,\tau}^{-1}Z^TXT_{\lambda,\tau}^{-1} & Q_{\lambda,\tau}^{-1} \end{bmatrix}\begin{bmatrix} X^T \\ Z^T \end{bmatrix} \\
&= XT_{\lambda,\tau}^{-1}X^T + \lambda_0^2 XT_{\lambda,\tau}^{-1}X^TZQ_{\lambda,\tau}^{-1}Z^TXT_{\lambda,\tau}^{-1}X^T - \lambda_0 XT_{\lambda,\tau}^{-1}X^TZQ_{\lambda,\tau}^{-1}Z^T \\
&\quad - \lambda_0 ZQ_{\lambda,\tau}^{-1}Z^TXT_{\lambda,\tau}^{-1}X^T + ZQ_{\lambda,\tau}^{-1}Z^T.
\end{aligned}
$$

Notice that $I - M_{\lambda,\tau} = \lambda_0 XT_{\lambda,\tau}^{-1}X^T$, and therefore

$$
\begin{aligned}
W\mathrm{Var}[\tilde{\theta}|\tau,\lambda]W^T &= XT_{\lambda,\tau}^{-1}X^T + (I - M_{\lambda,\tau})ZQ_{\lambda,\tau}^{-1}Z^T(I - M_{\lambda,\tau}) - (I - M_{\lambda,\tau})ZQ_{\lambda,\tau}^{-1}Z^T \\
&\quad - ZQ_{\lambda,\tau}^{-1}Z^T(I - M_{\lambda,\tau}) + ZQ_{\lambda,\tau}^{-1}Z^T \\
&= XT_{\lambda,\tau}^{-1}X^T + (I - M_{\lambda,\tau})ZQ_{\lambda,\tau}^{-1}Z^T(I - M_{\lambda,\tau} - I) \\
&\quad - ZQ_{\lambda,\tau}^{-1}Z^T(I - M_{\lambda,\tau}) + ZQ_{\lambda,\tau}^{-1}Z^T \\
&= XT_{\lambda,\tau}^{-1}X^T - (I - M_{\lambda,\tau})ZQ_{\lambda,\tau}^{-1}Z^T(I + M_{\lambda,\tau}) + (I - M_{\lambda,\tau})ZQ_{\lambda,\tau}^{-1}Z^T \\
&\quad - ZQ_{\lambda,\tau}^{-1}Z^T(I - M_{\lambda,\tau}) + ZQ_{\lambda,\tau}^{-1}Z^T.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\mathrm{tr}(W\mathrm{Var}[\tilde{\theta}|\tau,\lambda]W^T) &= \mathrm{tr}(XT_{\lambda,\tau}^{-1}X^T) - \mathrm{tr}((I - M_{\lambda,\tau})ZQ_{\lambda,\tau}^{-1}Z^T(I + M_{\lambda,\tau})) \\
&\quad + \mathrm{tr}((I - M_{\lambda,\tau})ZQ_{\lambda,\tau}^{-1}Z^T) - \mathrm{tr}(ZQ_{\lambda,\tau}^{-1}Z^T(I - M_{\lambda,\tau})) + \mathrm{tr}(ZQ_{\lambda,\tau}^{-1}Z^T) \\
&= \mathrm{tr}(XT_{\lambda,\tau}^{-1}X^T) - \mathrm{tr}((I - M_{\lambda,\tau})ZQ_{\lambda,\tau}^{-1}Z^T(I + M_{\lambda,\tau})) + \mathrm{tr}(ZQ_{\lambda,\tau}^{-1}Z^T) \\
&= \mathrm{tr}(XT_{\lambda,\tau}^{-1}X^T) - \mathrm{tr}(Q_{\lambda,\tau}^{-1/2}Z^T(I - M_{\lambda,\tau}^2)ZQ_{\lambda,\tau}^{-1/2}) + \mathrm{tr}(ZQ_{\lambda,\tau}^{-1}Z^T)\ .
\end{aligned}
$$

Applying (1) of Lemma 9 and using the fact that $R_{\lambda,\tau}^2 \preceq I$, we get

$$
Q_{\lambda,\tau}^{-1/2}Z^T(I - M_{\lambda,\tau}^2)ZQ_{\lambda,\tau}^{-1/2} = Q_{\lambda,\tau}^{-1/2}Z^TU(I - R_{\lambda,\tau}^2)U^TZQ_{\lambda,\tau}^{-1/2} \succeq 0\ .
$$

Hence $\mathrm{tr}(Q_{\lambda,\tau}^{-1/2}Z^T(I - M_{\lambda,\tau}^2)ZQ_{\lambda,\tau}^{-1/2}) \geqslant 0$, and therefore

$$
\mathrm{tr}(W\mathrm{Var}[\tilde{\theta}|\tau,\lambda]W^T) \leqslant \mathrm{tr}(XT_{\lambda,\tau}^{-1}X^T) + \mathrm{tr}(ZQ_{\lambda,\tau}^{-1}Z^T)\ , \qquad \text{(D.13)}
$$

which proves (1),

Next, notice that

$$XT_{\lambda,\tau}^{-1}X^T = X[\lambda_0(X^TX + D_\tau^{-1})]^{-1}X^T = \lambda_0^{-1}X[X^TX + D_\tau^{-1}]^{-1}X^T \preceq \lambda_0^{-1}X[X^TX + \tau_{\max}^{-1}I_p]^{-1}X^T.$$

Then,

$$\begin{aligned}
\mathrm{tr}(XT_{\lambda,\tau}^{-1}X^T) &\leqslant \lambda_0^{-1}\mathrm{tr}[U_X\Gamma_XV_X^T(V_X\Gamma_X^TU_X^TU_X\Gamma_XV_X^T + \tau_{\max}^{-1}I_p)^{-1}V_X\Gamma_X^TU_X^T] \\
&= \lambda_0^{-1}\mathrm{tr}[U_X\Gamma_X(\Gamma_X^T\Gamma_X + \tau_{\max}^{-1}I_p)^{-1}\Gamma_X^TU_X^T] \\
&= \lambda_0^{-1}\mathrm{tr}[\Gamma_X(\Gamma_X^T\Gamma_X + \tau_{\max}^{-1}I_p)^{-1}\Gamma_X^TU_X^TU_X] \\
&= \lambda_0^{-1}\mathrm{tr}[\Gamma_X(\Gamma_X^T\Gamma_X + \tau_{\max}^{-1}I_p)^{-1}\Gamma_X^T] \\
&= \lambda_0^{-1}\sum_{i=1}^{k_X}\frac{\gamma_i^2}{\gamma_i^2 + \tau_{\max}^{-1}} \\
&\leqslant \lambda_0^{-1}k_X = \lambda_0^{-1}\text{rank}(X)\ ,
\end{aligned}$$

which proves (2).

Finally,

$$ZQ_{\lambda,\tau}^{-1}Z^T = Z(\lambda_0Z^TM_{\lambda,\tau}Z + \Lambda)^{-1}Z^T \preceq Z\Lambda^{-1}Z^T \preceq \lambda_{\min}^{-1}ZZ^T,$$

where $\lambda_{\min} := \min\{\lambda_1, \lambda_2, \ldots, \lambda_r\}$. Thus,

$$\mathrm{tr}(ZQ_{\lambda,\tau}^{-1}Z^T) \leqslant \mathrm{tr}(\lambda_{\min}^{-1}ZZ^T) \leqslant \lambda_{\min}^{-1}\mathrm{tr}(ZZ^T) \leqslant \mathrm{tr}(ZZ^T)\sum_{i=1}^{m}\lambda_i^{-1},$$

which proves (3). □

D.3 Proof of Lemma 2

Lemma 2. *For all* $\tau \in \mathbb{R}_+^p$ *and* $\lambda \in \mathbb{R}_+^{m+1}$

$$\|y - W\mathrm{E}[\theta|\tau,\lambda]\|^2 \leqslant 2n\|y\|^2 + 2n^3\|y\|^2\ .$$

Proof of Lemma 2. First,

$$\begin{aligned}
W\mathrm{E}[\theta|\tau,\lambda] &= [X\ Z]\begin{bmatrix} \lambda_0T_{\lambda,\tau}^{-1}X^Ty - \lambda_0^2T_{\lambda,\tau}^{-1}X^TZQ_{\lambda,\tau}^{-1}Z^TM_{\lambda,\tau}y \\ \lambda_0Q_{\lambda,\tau}^{-1}Z^TM_{\lambda,\tau}y \end{bmatrix}, \\
&= \lambda_0XT_{\lambda,\tau}^{-1}X^Ty - \lambda_0^2XT_{\lambda,\tau}^{-1}X^TZQ_{\lambda,\tau}^{-1}Z^TM_{\lambda,\tau}y + \lambda_0ZQ_{\lambda,\tau}^{-1}Z^TM_{\lambda,\tau}y\ .
\end{aligned}$$

Thus,

$$
\begin{aligned}
\|y - W\mathrm{E}[\theta|\tau,\lambda]\|^2 &= \|y - \lambda_0 X T_{\lambda,\tau}^{-1} X^T y + \lambda_0^2 X T_{\lambda,\tau}^{-1} X^T Z Q_{\lambda,\tau}^{-1} Z^T M_{\lambda,\tau} y - \lambda_0 Z Q_{\lambda,\tau}^{-1} Z^T M_{\lambda,\tau} y\|^2 \\
&= \|(I - \lambda_0 X T_{\lambda,\tau}^{-1} X^T) y - (I - \lambda_0 X T_{\lambda,\tau}^{-1} X^T)\lambda_0 Z Q_{\lambda,\tau}^{-1} Z^T M_{\lambda,\tau} y\|^2 \\
&= \|M_{\lambda,\tau} y - \lambda_0 M_{\lambda,\tau} Z Q_{\lambda,\tau}^{-1} Z^T M_{\lambda,\tau} y\|^2 \\
&\leqslant 2\|M_{\lambda,\tau} y\|^2 + \|\lambda_0 M_{\lambda,\tau} Z Q_{\lambda,\tau}^{-1} Z^T M_{\lambda,\tau} y\|^2 \\
&\leqslant 2\|M_{\lambda,\tau}\|^2 \|y\|^2 + 2\|\lambda_0 M_{\lambda,\tau} Z Q_{\lambda,\tau}^{-1} Z^T M_{\lambda,\tau} y\|^2 \\
&\leqslant 2n\|y\|^2 + 2\|\lambda_0 M_{\lambda,\tau} Z Q_{\lambda,\tau}^{-1} Z^T M_{\lambda,\tau} y\|^2 \ ,
\end{aligned} \tag{D.14}
$$

where the last inequality follows from (3) of Lemma 9.

Indeed,

$$
\begin{aligned}
\|\lambda_0 M_{\lambda,\tau} Z Q_{\lambda,\tau}^{-1} Z^T M_{\lambda,\tau} y\|^2 &= \|\lambda_0 M_{\lambda,\tau}^{1/2} M_{\lambda,\tau}^{1/2} Z Q_{\lambda,\tau}^{-1} Z^T M_{\lambda,\tau}^{1/2} M_{\lambda,\tau}^{1/2} y\|^2, \\
&\leqslant \|M_{\lambda,\tau}^{1/2}\|^2 \|\lambda_0 M_{\lambda,\tau}^{1/2} Z Q_{\lambda,\tau}^{-1} Z^T M_{\lambda,\tau}^{1/2}\|^2 \|M_{\lambda,\tau}^{1/2} y\|^2 \\
&\leqslant \|M_{\lambda,\tau}^{1/2}\|^2 \|\lambda_0 M_{\lambda,\tau}^{1/2} Z Q_{\lambda,\tau}^{-1} Z^T M_{\lambda,\tau}^{1/2}\|^2 \|M_{\lambda,\tau}^{1/2}\|^2 \|y\|^2 \\
&\leqslant n^3 \|y\|^2 \ ,
\end{aligned} \tag{D.15}
$$

where the last inequality follows from (2) of Lemma 10, and (2) of Lemma 9 since

$$
\|M_{\lambda,\tau}^{1/2}\|^2 = \mathrm{tr}(M_{\lambda,\tau}) \leqslant \mathrm{tr}(I_n) = n \ .
$$

□

Thus,

$$
\|y - W\mathrm{E}[\theta|\tau,\lambda]\|^2 \leqslant 2n\|y\|^2 + 2n^3\|y\|^2.
$$

D.4 Proof of Lemma 3

Lemma 3. *For all* $\tau \in \mathbb{R}_+^p$ *and* $\lambda \in \mathbb{R}_+^{m+1}$*,*

(1) $\mathrm{tr}(\mathrm{Var}[\beta|\tau,\lambda]) \leqslant \lambda_0^{-1} \sum_{j=1}^p \tau_j + c^* \, \mathrm{tr}(ZZ^T) \sum_{i=1}^m \lambda_i^{-1}$, and

(2) $\|\mathrm{E}[\beta|\tau,\lambda]\|^2 \leqslant c^* \|y\|^2 n^2 \left(s_{\max}^2 \sum_{j=1}^p \tau_j + 1 \right)$,

where $s_{\max}$ *is the largest singular value of* X *and* c^* *is a finite positive constant.*

Proof of Lemma 3. From (5)

$$\operatorname{tr}(\operatorname{Var}[\beta|\tau,\lambda]) = \operatorname{tr}(T_{\lambda,\tau}^{-1}) + \operatorname{tr}(\lambda_0^2 T_{\lambda,\tau}^{-1} X^T Z Q_{\lambda,\tau}^{-1} Z^T X T_{\lambda,\tau}^{-1}) \ . \tag{D.16}$$

Notice that that

$$T_{\lambda,\tau}^{-1} = [\lambda_0(X^T X + D_\tau^{-1})]^{-1} \preceq \lambda_0^{-1} \left(D_\tau^{-1}\right)^{-1} = \lambda_0^{-1} D_\tau \ ,$$

and thus

$$\operatorname{tr}(T_{\lambda,\tau}^{-1}) \leqslant \lambda_0^{-1} \operatorname{tr}(D_\tau) = \lambda_0^{-1} \sum_{j=1}^{p} \tau_j \ . \tag{D.17}$$

Next,

$$\begin{aligned} \operatorname{tr}(\lambda_0^2 T_{\lambda,\tau}^{-1} X^T Z Q_{\lambda,\tau}^{-1} Z^T X T_{\lambda,\tau}^{-1}) &= \|Q_{\lambda,\tau}^{-1/2} Z^T (\lambda_0 X T_{\lambda,\tau}^{-1})\|^2 \\ &\leqslant \|Q_{\lambda,\tau}^{-1/2} Z^T\|^2 \|\lambda_0 X T_{\lambda,\tau}^{-1}\|^2 \\ &= \operatorname{tr}(Z Q_{\lambda,\tau}^{-1} Z^T) \operatorname{tr}(\lambda_0^2 T_{\lambda,\tau}^{-1} X^T X T_{\lambda,\tau}^{-1}) \\ &= \operatorname{tr}\left(Z(\lambda_0 Z^T M_{\lambda,\tau} Z^T + \Lambda)^{-1} Z^T\right) \operatorname{tr}(\lambda_0^2 X T_{\lambda,\tau}^{-2} X^T) \\ &\leqslant \lambda_{\min}^{-1} \operatorname{tr}(Z Z^T) \|\lambda_0 T_{\lambda,\tau}^{-1} X^T\|^2 \\ &\leqslant c^* \operatorname{tr}(Z Z^T) \sum_{i=1}^{m} \lambda_i^{-1} \ , \end{aligned} \tag{D.18}$$

where the last inequality follows from (1) of Lemma 11 since there exists some finite c^* such that $\|\lambda_0 T_{\lambda,\tau}^{-1} X^T\|^2 \leqslant c^*$. Thus, from (D.17) and (D.18)

$$\operatorname{tr}(\operatorname{Var}[\theta|\tau,\lambda]) \leqslant \lambda_0^{-1} \sum_{j=1}^{p} \tau_j + c^* \operatorname{tr}(Z Z^T) \sum_{i=1}^{m} \lambda_i^{-1} \ ,$$

which proves (1).

To prove (2), it follows from (4)

$$\begin{aligned} \|\mathrm{E}[\beta|\tau,\lambda]\|^2 &= \|\lambda_0 T_{\lambda,\tau}^{-1} X^T y - \lambda_0^2 T_{\lambda,\tau}^{-1} X^T Z Q_{\lambda,\tau}^{-1} Z^T M_{\lambda,\tau} y\|^2 \\ &= \|\lambda_0 T_{\lambda,\tau}^{-1} X^T (I - \lambda_0 Z Q_{\lambda,\tau}^{-1} Z^T M_{\lambda,\tau}) y\|^2 \\ &\leqslant \|\lambda_0 T_{\lambda,\tau}^{-1} X^T\|^2 \|I - \lambda_0 Z Q_{\lambda,\tau}^{-1} Z^T M_{\lambda,\tau}\|^2 \|y\|^2 \ , \end{aligned} \tag{D.19}$$

and from (1) and (2) of Lemma 11,

$$\|\mathrm{E}[\beta|\tau,\lambda]\|^2 \leqslant c^* \|y\|^2 n^2 \left(s_{\max}^2 \sum_{j=1}^{p} \tau_j + 1 \right) \ . \tag{D.20}$$

□

D.5 Proof of Lemma 4

Lemma 4. *For all* $\tau \in \mathbb{R}_+^p$ *and* $\lambda \in \mathbb{R}_+^{m+1}$,

$$\mathrm{E}\left[\sum_{j=1}^{p} \frac{1}{|\beta_j|^{\nu(c)}} \middle| \tau, \lambda\right] \leqslant p\kappa(c) s_{\max}^{\nu(c)} \lambda_0^{\nu(c)/2} + \kappa(c)\lambda_0^{\nu(c)/2} \sum_{j=1}^{p} \frac{1}{\tau_j^{\nu(c)/2}} ,$$

where

$$\kappa(c) := \frac{\Gamma\left(\frac{1-\nu(c)}{2}\right) 2^{\frac{1-\nu(c)}{2}}}{\sqrt{2\pi}} ,$$

and $s_{\max}$ *is the largest sinular value of the matrix* X.

Proof of Lemma 4. Recall that given τ and λ, $\beta \sim N(\mu, \Sigma)$ where μ and Σ are given by (4) and (5), respectively. Thus, $\beta_j \sim N(\mu_j, \sigma_j^2)$ where $\mu_j = e_j^T \mu$ and $\sigma_j^2 = e_j^T \Sigma e_j$ for $j = 1, \ldots, p$. As in Lemma 11, $e_1, \ldots, e_p$ represent the standard unit vectors for $\mathbb{R}^p$. From Proposition A1 of Pal and Khare (2014), it follows that

$$\mathrm{E}\left[\frac{1}{|\beta_j|^{\nu(c)}} \middle| \tau, \lambda\right] \leqslant \frac{\kappa(c)}{\sigma_j^{\nu(c)}}, \text{ for } j = 1, \ldots, p , \tag{D.21}$$

where

$$\kappa(c) := \frac{\Gamma\left(\frac{1-\nu(c)}{2}\right) 2^{\frac{1-\nu(c)}{2}}}{\sqrt{2\pi}} .$$

From (5),

$$
\begin{aligned}
\frac{1}{\sigma_j^{\nu(c)}} &= \left(\frac{1}{e_j^T[T_{\lambda,\tau}^{-1} + \lambda_0^2 T_{\lambda,\tau}^{-1} X^T Z Q_{\lambda,\tau}^{-1} Z^T X T_{\lambda,\tau}^{-1}]e_j}\right)^{\nu(c)/2} \\
&\leqslant \left(\frac{1}{e_j^T T_{\lambda,\tau}^{-1} e_j}\right)^{\nu(c)/2} \\
&= \left(\frac{1}{e_j^T[\lambda_0(X^T X + D_\tau^{-1})]^{-1} e_j}\right)^{\nu(c)/2} \\
&= \left(\frac{1}{e_j^T[\lambda_0(V_X \Gamma_X^T U_X^T U_X \Gamma_X V_X^T + D_\tau^{-1})]^{-1} e_j}\right)^{\nu(c)/2} \\
&= \left(\frac{1}{e_j^T[\lambda_0(V_X \Gamma_X^T \Gamma_X V_X^T + D_\tau^{-1})]^{-1} e_j}\right)^{\nu(c)/2} \\
&\leqslant \left(\frac{1}{e_j^T[\lambda_0(s_{\max}^2 I + D_\tau^{-1})]^{-1} e_j}\right)^{\nu(c)/2} \\
&= \lambda_0^{\nu(c)/2}\left(s_{\max}^2 + \frac{1}{\tau_j}\right)^{\nu(c)/2} \\
&\leqslant (s_{\max}^2)^{\nu(c)/2}\lambda_0^{\nu(c)/2} + \lambda_0^{\nu(c)/2}\frac{1}{\tau_j^{\nu(c)/2}} \ ,
\end{aligned}
$$

where $s_{\max}$ is the largest singular value of X, and the last inequality follows from the fact that $(x+y)^\delta \leqslant x^\delta + y^\delta$ for $\delta \in (0,1)$. Thus, from (D.21)

$$
\mathrm{E}\left[\sum_{j=1}^{p} \frac{1}{|\tilde{\beta}_j|^c}\middle|\tau,\lambda\right] \leqslant \sum_{j=1}^{p} \frac{\kappa(c)}{\sigma_j^{\nu(c)}} \leqslant p\kappa(c) s_{\max}^{\nu(c)}\lambda_0^{\nu(c)/2} + \kappa(c)\lambda_0^{\nu(c)/2}\sum_{j=1}^{p}\frac{1}{\tau_j^{\nu(c)/2}} \ . \tag{D.22}
$$

□

D.6 Proof of Lemma 5

Lemma 5. *Suppose that Z has full column rank. For all $\tau \in \mathbb{R}_+^p$ and $\lambda \in \mathbb{R}_+^{m+1}$,*

(1) $\mathrm{tr}(R_i Q_{\lambda,\tau}^{-1} R_i^T) \leqslant q_i \lambda_i^{-1}$, and

(2) $||\mathrm{E}[u_i|\tau,\lambda]||^2 \leqslant q_i \mathrm{tr}[(Z^T Z)^{-1}] n^3 \|y\|^2 \left(s_{\max}^2 \sum_{j=1}^p \tau_j + 1\right)$,

for $i = 1, \dots, m$.

Proof of Lemma 5. Note that

$$Q_{\lambda,\tau}^{-1} = (\lambda_0 Z^T M_{\lambda,\tau} Z + \Lambda)^{-1} \preceq \Lambda^{-1} \ , \tag{D.23}$$

thus

$$\mathrm{tr}(R_i Q_{\lambda,\tau}^{-1} R_i^T) \leqslant \mathrm{tr}(R_i \Lambda^{-1} R_i^T) = q_i \lambda_i^{-1}, \ \ i = 1, \ldots, m \ , \tag{D.24}$$

which proves the first result.

Next, from (4),

$$\begin{aligned}
\|\mathrm{E}[u_i|\tau,\lambda]\|^2 &= \|\mathrm{E}[R_i u|\tau,\lambda]\|^2 \\
&\leqslant \|R_i\|^2 \|\mathrm{E}[u|\tau,\lambda]\|^2 \\
&= \mathrm{tr}(I_{q_i}) \|\lambda_0 Q_{\lambda,\tau}^{-1} Z^T M_{\lambda,\tau} y\|^2 \\
&= q_i \|\lambda_0 Q_{\lambda,\tau}^{-1} Z^T M_{\lambda,\tau} y\|^2 \\
&= q_i \|\lambda_0 (Z^T Z)^{-1} Z^T Z Q_{\lambda,\tau}^{-1} Z^T M_{\lambda,\tau} y\|^2 \\
&\leqslant q_i \|(Z^T Z)^{-1} Z^T\|^2 \|\lambda_0 Z Q_{\lambda,\tau}^{-1} Z^T M_{\lambda,\tau} y\|^2 \\
&= q_i \|(Z^T Z)^{-1} Z^T\|^2 \|\lambda_0 M_{\lambda,\tau}^{-1/2} M_{\lambda,\tau}^{1/2} Z Q_{\lambda,\tau}^{-1} Z^T M_{\lambda,\tau} y\|^2 \\
&\leqslant q_i \|(Z^T Z)^{-1} Z^T\|^2 \|M_{\lambda,\tau}^{-1/2}\|^2 \|\lambda_0 M_{\lambda,\tau}^{1/2} Z Q_{\lambda,\tau}^{-1} Z^T M_{\lambda,\tau}^{1/2}\|^2 \|M_{\lambda,\tau}^{1/2}\|^2 \|y\|^2 \\
&= q_i \mathrm{tr}((Z^T Z)^{-1} Z^T Z (Z^T Z)^{-1}) \mathrm{tr}(M_{\lambda,\tau}^{-1}) \|\lambda_0 M_{\lambda,\tau}^{1/2} Z Q_{\lambda,\tau}^{-1} Z^T M_{\lambda,\tau}^{1/2}\|^2 \mathrm{tr}(M_{\lambda,\tau}) \|y\|^2 \\
&= q_i \mathrm{tr}((Z^T Z)^{-1}) \mathrm{tr}(M_{\lambda,\tau}^{-1}) \|\lambda_0 M_{\lambda,\tau}^{1/2} Z Q_{\lambda,\tau}^{-1} Z^T M_{\lambda,\tau}^{1/2}\|^2 \mathrm{tr}(M_{\lambda,\tau}) \|y\|^2 \ .
\end{aligned} \tag{D.25}$$

Thus, from (2) of Lemma 9 and (2) of Lemma 10,

$$\begin{aligned}
\|\mathrm{E}[u_i|\tau,\lambda]\|^2 &\leqslant q_i \mathrm{tr}[(Z^T Z)^{-1}] (\tau_{\max} s_{\max}^2 + 1) \mathrm{tr}(I_n) n \mathrm{tr}(I_n) \|y\|^2 \\
&= q_i \mathrm{tr}[(Z^T Z)^{-1}] n^3 \|y\|^2 (\tau_{\max} s_{\max}^2 + 1) \\
&= q_i \mathrm{tr}[(Z^T Z)^{-1}] n^3 \|y\|^2 \left(s_{\max}^2 \sum_{j=1}^{p} \tau_j + 1 \right),
\end{aligned} \tag{D.26}$$

which proves the second result. □

D.7 Proof of Lemma 6

Lemma 6. *For all* $(\theta, \lambda) \in \mathsf{X}$*,*

1. $\mathrm{E}[\tau_j|\theta,\lambda] \leqslant \frac{4c+1}{4d} + \frac{\lambda_0 \beta_j^2}{2}$, and

2. $\mathrm{E}[\tau_j|\theta,\lambda] \leqslant \frac{c}{d} + \frac{\beta_j^2}{2C} + \frac{\lambda_0 C}{4d}$.

for every $C > 0$.

Proof of Lemma 6. From (23),

$$\mathrm{E}[\tau_j|\theta,\lambda] = \sqrt{\frac{\lambda_0\beta_j^2}{2d}}\,\frac{K_{c+\frac{1}{2}}\left(\sqrt{2d\lambda_0\beta_j^2}\right)}{K_{c-\frac{1}{2}}\left(\sqrt{2d\lambda_0\beta_j^2}\right)} \,. \tag{D.27}$$

Theorem 2 of Segura (2011) states that

$$\frac{K_{\nu-\frac{1}{2}}(x)}{K_{\nu+\frac{1}{2}}(x)} > \frac{x}{\sqrt{x^2+\nu^2}+\nu} \,,$$

for $\nu > 0$ and $x > 0$.

Thus,

$$\begin{aligned}
\mathrm{E}[\tau_j|\theta,\lambda] &< \sqrt{\frac{\lambda_0\beta_j^2}{2d}} \times \frac{\sqrt{c^2+2d\lambda_0\beta_j^2}+c}{\sqrt{2d\lambda_0\beta_j^2}} \\
&= \frac{\sqrt{c^2+2d\lambda_0\beta_j^2}+c}{2d} \\
&\leqslant \frac{c}{2d} + \frac{c}{2d} + \sqrt{\frac{\lambda_0\beta_j^2}{2d}} \\
&= \frac{c}{d} + \sqrt{\frac{\lambda_0\beta_j^2}{2d}} \,,
\end{aligned} \tag{D.28}$$

where we make use of the fact that $\sqrt{x^2+y^2} \leqslant |x| + |y|$.

To get the first inequality, we use the fact that $|xy| \leqslant (x^2+y^2)/2$. Thus

$$\begin{aligned}
\mathrm{E}[\tau_j|\theta,\lambda] &\leqslant \frac{c}{d} + \frac{\lambda_0\beta_j^2}{2} + \frac{1}{4d} \\
&= \frac{4c+1}{4d} + \frac{\lambda_0\beta_j^2}{2} \,.
\end{aligned} \tag{D.29}$$

Similarly, for any constant $C > 0$,

$$\begin{aligned}
\mathrm{E}[\tau_j|\theta,\lambda] &\leqslant \frac{c}{d} + \sqrt{\frac{\lambda_0\beta_j^2}{2d}} \\
&= \frac{c}{d} + \sqrt{\frac{\lambda_0\beta_j^2 C}{2dC}} \\
&\leqslant \frac{c}{d} + \frac{\beta_j^2}{2C} + \frac{\lambda_0 C}{4d} \,.
\end{aligned} \tag{D.30}$$

□

D.8 Proof of Lemma 7

Lemma 7. *For all* $(\theta, \lambda) \in \mathsf{X}$,

$$\mathrm{E}[\tau_j^{-1}|\theta, \lambda] \leqslant d + \frac{3}{2\lambda_0\beta_j^2} \ .$$

Proof of Lemma 7. From (24),

$$\mathrm{E}[\tau_j^{-1}|\theta, \lambda] = \sqrt{\frac{2d}{\lambda_0\beta_j^2}}\, \frac{K_{c-\frac{3}{2}}\left(\sqrt{2d\lambda_0\beta_j^2}\right)}{K_{c-\frac{1}{2}}\left(\sqrt{2d\lambda_0\beta_j^2}\right)} \ .$$

From Lemma 2.2 of Ismail and Muldoon (1978), for each $x > 0$, $s > 0$ and $s_1 \in \mathbb{R}$, the function $K_{s_1+s}(x)/K_{s_1}(x)$ is increasing in s_1. Thus, for $s_1 < s_2$, $s > 0$ and $x > 0$

$$\frac{K_{s_1+s}(x)}{K_{s_1}(x)} \leqslant \frac{K_{s_2+s}(x)}{K_{s_2}(x)} \ .$$

Thus, taking $s_1 = -\frac{3}{2}$, $s_2 = -\frac{1}{2}$ and $s = c$, we have

$$\frac{K_{c-\frac{3}{2}}(x)}{K_{-\frac{3}{2}}(x)} \leqslant \frac{K_{c-\frac{1}{2}}(x)}{K_{-\frac{1}{2}}(x)} \ .$$

Since $K_s(x) > 0$ for $s \in \mathbb{R}$ and $x > 0$ (Abramowitz and Stegun (1965), page 374), it follows that

$$\frac{K_{c-\frac{3}{2}}(x)}{K_{c-\frac{1}{2}}(x)} \leqslant \frac{K_{-\frac{3}{2}}(x)}{K_{-\frac{1}{2}}(x)} \ .$$

Next, using the fact that

$$K_{-\frac{1}{2}}(x) = e^{-x}\sqrt{\frac{\pi}{2x}} \ ,$$

and

$$K_{-\frac{3}{2}}(x) = e^{-x}\sqrt{\frac{\pi}{2x}}\left(1 + \frac{1}{x}\right),$$

hence

$$\frac{K_{-\frac{3}{2}}(x)}{K_{-\frac{1}{2}}(x)} = \left(1 + \frac{1}{x}\right) \ ,$$

for all $x \in \mathbb{R}$.

Thus,

$$\begin{aligned}
\mathrm{E}[\tau_j^{-1}|\theta,\lambda] &\leqslant \sqrt{\frac{2d}{\lambda_0\beta_j^2}}\left(1+\frac{1}{\sqrt{2d\lambda_0\beta_j^2}}\right)\\
&= \sqrt{\frac{2d}{\lambda_0\beta_j^2}}+\frac{1}{\lambda_0\beta_j^2}\\
&\leqslant d+\frac{1}{2\lambda_0\beta_j^2}+\frac{1}{\lambda_0\beta_j^2}\\
&= d+\frac{3}{2\lambda_0\beta_j^2}\ ,
\end{aligned}$$

where the second inequality follows from the fact that $|xy| \leqslant (x^2 + y^2)/2$. □

E Proof of Lemma 8

Lemma 8. *For all* $(\theta, \lambda) \in \mathsf{X}$,

$$\mathrm{E}\left[\frac{1}{\tau_j^{\nu(c)/2}}\Big|\theta,\lambda\right] \leqslant M_1\frac{1}{\lambda_0^{\nu(c)/2}|\beta_j|^{\nu(c)}}+M_2\ ,$$

where M_1 *is a positive constant such that* $M_1\kappa(c) < 1$*, and* M_2 *is a positive finite constant.*

Proof of Lemma 8. This follows directly from the arguments on pp. 613-616 and p. 618 of Pal and Khare (2014). □

References

ABRAHAMSEN, T. (2016). *Convergence analysis of MCMC samplers for Bayesian linear mixed models for large p, small n problems.* Ph.D. thesis, Department of Statistics, University of Florida.

ABRAHAMSEN, T. and HOBERT, J. P. (2017). Convergence analysis of block Gibbs samplers for Bayesian linear mixed models with $p > N$. *Bernoulli*, **23** 459–478.

ABRAMOWITZ, M. and STEGUN, I. A. (1965). *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables (Dover Books on Mathematics).* 1st ed. Dover Publications.

BHATTACHARYA, A., PATI, D., PILLAI, N. S. and DUNSON, D. B. (2013). Bayesian shrinkage. *arXiv preprint arXiv:1212.6088.*

BHATTACHARYA, A., PATI, D., PILLAI, N. S. and DUNSON, D. B. (2015). Dirichlet-Laplace priors for optimal shrinkage. *Journal of the American Statistical Association*, **110** 1479–1490.

FAZLI, S., DANÓCZY, M., SCHELLDORFER, J. and MÜLLER, K.-R. (2011). *l*1-penalized linear mixed-effects models for high dimensional data with application to BCI. *NeuroImage*, **56** 2100–2108.

FLEGAL, J. M., HARAN, M. and JONES, G. L. (2008). Markov chain Monte Carlo: Can we trust the third significant figure? *Statistical Science*, **23** 250–260.

GRIFFIN, J. E. and BROWN, P. J. (2010). Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis*, **5** 171–188.

ISMAIL, M. E. and MULDOON, M. E. (1978). Monotonicity of the zeros of a cross-product of Bessel functions. *SIAM Journal on Mathematical Analysis*, **9** 759–767.

JOHNSON, A. A. and JONES, G. L. (2015). Geometric ergodicity of random scan Gibbs samplers for hierarchical one-way random effects models. *Journal of Multivariate Analysis*, **140** 325–342.

JONES, G. L. and HOBERT, J. P. (2001). Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Statistical Science*, **16** 312–334.

JONES, G. L., ROBERTS, G. O. and ROSENTHAL, J. S. (2014). Convergence of conditional Metropolis-Hastings samplers. *Advances in Applied Probability*, **46** 422–445.

JUNG, Y. J. (2015). *Convergence analysis of Markov chain Monte Carlo algorithms for Bayesian regression models with non-Gaussian errors.* Ph.D. thesis, Department of Statistics, University of Florida.

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

MCCULLOCH, C. E., SEARLE, S. R. and NEUHAUS, J. M. (2008). *Generalized, Linear, and Mixed Models*. 2nd ed. John Wiley & Sons, New York.

MEYN, S. P. and TWEEDIE, R. L. (2009). *Markov Chains and Stochastic Stability*. 2nd ed. Cambridge University Press, Cambridge.

PAL, S. and KHARE, K. (2014). Geometric ergodicity for Bayesian shrinkage models. *Electronic Journal of Statistics*, **8** 604–645.

PARK, T. and CASELLA, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association*, **103** 681–686.

POLSON, N. G. and SCOTT, J. G. (2010). Shrink globally, act locally: Sparse Bayesian regularization and prediction. *Bayesian Statistics*, **9** 501–538.

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.

ROHART, F., SAN-CRISTOBAL, M. and LAURENT, B. (2014). Selection of fixed effects in high dimensional linear mixed models using a multicycle ECM algorithm. *Computational Statistics and Data Analysis*, **80** 209–222.

ROMÁN, J. C. and HOBERT, J. P. (2012). Convergence analysis of the Gibbs sampler for Bayesian general linear mixed models with improper priors. *The Annals of Statistics*, **40** 2823–2849.

ROMÁN, J. C. and HOBERT, J. P. (2015). Geometric ergodicity of Gibbs samplers for Bayesian general linear mixed models with proper priors. *Linear Algebra and its Applications*, **473** 54–77.

SCHELLDORFER, J., BÜHLMANN, P. and VAN DE GEER, S. (2011). Estimation for high-dimensional linear mixed-effects models using $l1$-penalization. *Scandinavian Journal of Statistics*, **38** 197–214.

SEGURA, J. (2011). Bounds for ratios of modified Bessel functions and associated Turán-type inequalities. *Journal of Mathematical Analysis and Applications*, **374** 516–528.

ZHOU, X., CARBONETTO, P. and STEPHENS, M. (2013). Polygenic modeling with Bayesian sparse linear mixed models. *PLoS Genet*, **9** e1003264.