RESEARCH

Variance, Eigenvalue, and Correlation Structure Inequalities with Applicability to the Convergence of Two-Block Gibbs samplers

Kjell Nygren

Correspondence:
kjell.a.nygren@gmail.com
Symphony Health Solutions, 550 Blair Mill Road, Horsham, PA, 18940 US
Full list of author information is available at the end of the article

Abstract

A variance inequality showing that the posterior variance-covariance matrix for models with multivariate normal priors and log-concave likelihood functions is bounded from above by the prior variance-covariance matrix is established. A related variance inequality and two eigenvalue inequalities are also established. The results are leveraged in order to bound variance and precision matrices as well as correlation structures associated with two-block Gibbs samplers for Hierarchical Bayesian models with Multivariate Normal prior structures and log-concave likelihood functions. The role the latter results play in the establishment of Rosenthal type minorizations and drift conditions (and hence geometric ergodicity) are discussed.

Keywords: sample; article; author

Introduction

In this paper, we are concerned with properties of hierarchical Bayesian models with log-concave likelihood functions and multivariate normal prior structure. We are particularly interested in the properties of two-block Gibbs samplers [1, 2] applied to such models. To understand their properties, we examine the relationship between the prior and posterior variance-covariance matrices for models with log-concave likelihood functions and multivariate normal priors. We also examine properties of eigenvalues associated with correlation structures for the sampler.

One of our main results shows that posterior variance-covariance matrices in models with log-concave likelihood functions and multivariate normal priors are bounded from above by the corresponding prior variance-covariance matrices (i.e., the difference between the prior and posterior variance-covariance matrix is positive semi-definite). This result is stronger than the well known result [3] that the expected value of the posterior variance is lower than the prior variance. The variance inequality allows us to bound variance and precision matrices associated with the transition Kernel for the two-block Gibbs sampler alluded to above. In particular, it implies that the precision matrix for the marginal transition Kernel is bounded from below by a matrix P_{LB}^{D} and that the variance of the same is bounded from above by $Var_{UB}^{\mu} = (P_{LB}^{D})^{-1}$. These two properties play a role in providing bounds on intercept terms for minorization and drift conditions respectively.

The first of two eigenvalue inequalities establishes conditions under which the eigenvalues for a matrix of the form BAB is less than or equal to the corresponding eigenvalues for the matrix A. This property is shown to hold if A is hermitian positive definite and B is hermitian positive semi-definite with maximal eigenvalue less than or equal to 1. We apply this result together with the variance inequality in order to bound the spectral norm for correlation structures associated with posterior densities associated with the two-block Gibbs sampler alluded to above. In particular, the spectral norms are bounded from above by the spectral norm for the correlation structure associated with the prior. In the univariate case, this is equivalent to saying that the squared correlation under the posterior density is bounded above by the squared correlation under the prior.

A second eigenvalue inequality establishes conditions under which the maximal eigenvalue for a matrix of the form $A(A+B)^{-1}(A+B)^{-1}A$ is non-negative and strictly less than 1. This property is shown to hold if A is hermitian positive semi-definite, B hermitian positive definite, and $(A+B)^{-1}A$ Hermitian. We apply this theorem in order to show that spectral norms for correlation structures arising from positive definite precision matrices are strictly bounded from above by 1. In the univariate case, this is equivalent to saying that the squared correlation is strictly bounded from above by 1. The two results regarding the correlation structures derived from the eigenvalue inequalities together imply that the slope parameter in the Rosenthal [4] type drift condition exists and is less than 1. In particular, it can be taken to be the squared spectral norm for the correlation structure associated with a linearly transformed version of the prior.

The rest of this paper is organized as follows. In the next section, we state the variance and eigenvalue inequalities that drive the rest of the results in our paper. We then introduces the notation for the Hierarchical Bayesian models with multivariate normal prior stuctures and apply the variance inequality in order to provide bounds for precision and variance-covariance matrices associated with the transiton kernel. Correlation structure inequalities are then covered before we conclude the discussion. The appendicies contain details related to some of the proofs. Two supplements contain additional supporting material for some of the lengthier arguments.

Main Theorems

This section contains the main theorems that drive the rest of our results. Separate subsections covers variance inequalities and eigenvalue inequalities respectively.

Variance Inequalities

Here we are interested in the relationship between the prior and the posterior variance. A well known result due to Raiffa and Schlaifer [3] states that the expected value of the posterior variance is less than or equal to the prior variance. Here we are interested in conditions under which the posterior variance-covariance matrix is bounded from above by the prior variance-covariance matrix. Our main theorem shows that this indeed holds for models with log-concave likelihood functions and multivariate normal priors.

Theorem 1 *Suppose β has a multivariate normal prior with mean vector $X\mu$, variance-covariance matrix Σ and a log-concave likelihood function $f(\mathbf{y},.)$. Then the posterior variance satisfies the property that*

$$\Sigma - Var(\beta|X\mu,\Sigma,\mathbf{y}) \tag{1}$$

is positive semidefinite.

Proof See Appendix A. □

The proof of the above in the univariate case relies on the abilty to find an appropriately positioned normal density, with the same variance as the prior, that also has the property that the interval where it falls below the posterior is centered at its mean. The variance inequality is straightforward to verify for such a density. The variance bound above can be further improved when the data precision is bounded from below as made clear by the following corollary.

Corollary 1 *Suppose β has a multivariate normal prior with precision matrix P and a log-concave likelihood function $f(\mathbf{y},.)$. Let P_{LB}^{D} be as in definition 2. Then the matrix*

$$(P+P_{LB}^{D})^{-1} - Var(\beta|X\mu,P,\mathbf{y}) \tag{2}$$

is positive semidefinite.

Proof Construct a fictious model with the same posterior density as the original model by subtracting a quadratic term $\frac{1}{2}\beta^T P_{LB}^{D}\beta$ from the log of the prior and adding it to the log-likelihood function. This fictious model is also a model with a multivariate normal prior and a log-concave likelihood function. Applying theorem 1 to this fictious model yields the desired result. □

The below example shows that the restriction to a log-concave likelihood function in general can not be removed. In particular, the prior and posterior densities in this example fail to satisfy the assumptions of Claim 3.

Example 1 *Suppose that β has a normal prior with mean $\mu = 0$ and precision P. Suppose that the likelihood function is given by*

$$f(y|\beta) \quad = \quad \frac{1}{2}\frac{P_1}{\sqrt{2\pi}}e^{-0.5P_1(\beta-1)^2} + \frac{1}{2}\frac{P_1}{\sqrt{2\pi}}e^{-0.5P_1(\beta+1)^2}.$$

It is straighforward to verify that

$$\begin{aligned} Var(\beta|\mu,y) &= \frac{1}{2}\frac{1}{P+P_1} + \frac{1}{2}\frac{1}{P+P_1} + (\tfrac{1}{2})(\tfrac{1}{2})[\tfrac{P_1}{P+P_1} - [\tfrac{-P_1}{P+P_1}]^2 \\ &= \frac{1}{P+P_1} + [\tfrac{P_1}{P+P_1}]^2 \end{aligned}$$

Now, consider the special case where $P = 6$ and $P_1 = 2$. Then
$Var(\beta|\mu,y) = \frac{1}{8} + [\frac{2}{8}]^2 = \frac{8}{64} + [\frac{4}{64}] = \frac{3}{16} > \frac{3}{18} = \frac{1}{6} = Var(\beta|\mu)$.

When the data precision is bounded from above, the variance for the posterior density can also be bounded from below. The proof of theorem 2 uses similar arguments to those leveraged to establish Theorem 1.

Theorem 2 *Suppose β has a multivariate normal prior with variance-covariance matrix Σ and a log-concave likelihood function. Assume furthermore that there exists a positive semi-definite matrix P^D_{UB} such that for every β, the precision matrix $P^D(\beta)$ for the likelihood function satistifes the property that $P^D_{UB} - P^D(\beta)$ is a positive semi-definite matrix. Then the matrix*

$$Var(\beta|X\mu, \Sigma, \mathbf{y}) - (\Sigma^{-1} + P^D_{UB})^{-1} \tag{3}$$

is positive semidefinite.

Proof See Appendix A. □

Theorem 3 *Let P_1 be a positive definite Hermitian matrix and let P_2 and P_3 be positive semi-definite matrices of the same dimension. Then*

$$(P_1)^{-1} - (P_1 + P_3)^{-1} \geq (P_1 + P_2)^{-1} - (P_1 + P_2 + P_3)^{-1}$$

where $\geq$ means that the difference between the matrices on the left hand side and the right hand side is positive semi-definite.

[NOTE: CHECK PARENTHSES ARE IN CORRECT LOCATIONS BELOW. ALSO COMMENT ON REASONS EQUALITIES/INEQUALITIES HOLD].

Proof From Theorem 1 and the present assumptions, it follows that the matrices on both sides are hermitian and positive semi-definite. It suffices to show that the eigenvalues of the matrix on the left hand side exceed the corresponding eigenvalues on the right hand side. To simplify notation, let $A = P_1$, $B = P_1 + P_3$ and $C = P_2$.

Then we have:

$$
\begin{aligned}
\lambda_i[A^{-1}-B^{-1}] &= \sqrt{\lambda_i[(A^{-1}-B^{-1})(A^{-1}-B^{-1})]} \\
&= \sqrt{\lambda_i[(A^{-1}BB^{-1}-A^{-1}AB^{-1})(B^{-1}BA^{-1}-B^{-1}AA^{-1})]} \\
&= \sqrt{\lambda_i[(A^{-1}(B-A)B^{-1}B^{-1}(B-A)A^{-1})]} \\
&\geq \sqrt{\lambda_i[(A^{-1}(B-A)(B+C)^{-1}(B+C)^{-1}(B-A)A^{-1})]} \\
&= [\lambda_i[(A^{-1}[(B-A)(B+C)^{-1}(B+C)^{-1}(B-A)]^{0.5} \\
&\quad [(B-A)(B+C)^{-1}(B+C)^{-1}(B-A)]^{0.5}A^{-1})]]^{0.5} \\
&= [\lambda_i[([(B-A)(B+C)^{-1}(B+C)^{-1}(B-A)]^{0.5}A^{-1} \\
&\quad A^{-1}[(B-A)(B+C)^{-1}(B+C)^{-1}(B-A)]^{0.5})]]^{0.5} \\
&\geq [\lambda_i[([(B-A)(B+C)^{-1}(B+C)^{-1}(B-A)]^{0.5}(A+C)^{-1} \\
&\quad (A+C)^{-1}[(B-A)(B+C)^{-1}(B+C)^{-1}(B-A)]^{0.5})]]^{0.5} \\
&= [\lambda_i[((A+C)^{-1}[(B-A)(B+C)^{-1}(B+C)^{-1}(B-A)]^{0.5} \\
&\quad [(B-A)(B+C)^{-1}(B+C)^{-1}(B-A)]^{0.5}(A+C)^{-1})]]^{0.5} \\
&= [\lambda_i[((A+C)^{-1}[(B-A)(B+C)^{-1} \\
&\quad (B+C)^{-1}(B-A)](A+C)^{-1})]]^{0.5} \\
&= [\lambda_i[((A+C)^{-1}[((B+C)-(A+C))(B+C)^{-1} \\
&\quad (B+C)^{-1}((B+C)-(A+C))](A+C)^{-1})]]^{0.5} \\
&= \sqrt{\lambda_i[((A+C)^{-1}-(B+C)^{-1})((A+C)^{-1}-(B+C)^{-1})} \\
&= \lambda_i[(A+C)^{-1}-(B+C)^{-1}].
\end{aligned}
$$

□

Lemma 1 *Let $f_1(.)$ and $f_2(.)$ be two twice differentiable log-concave densities. Then the function $g(.)$ defined by $g(x) = (f_1(x) + (f_2(x))/(f_1(x) * f_2(x))$ is log-convex.*

Proof We first note that the gradient for the log of the function $g(.)$ is given by:

$$
\begin{aligned}
\nabla log[g(x)] &= \tfrac{f_1(x)\nabla log[f_1(x)]}{f_1(x)+f_2(x)} + \tfrac{f_2(x)\nabla log[f_2(x)]}{f_1(x)+f_2(x)} \\
&\quad -\nabla log[f_1(x)] - \nabla log[f_2(x)].
\end{aligned}
$$

The second order derivative is then given by

$$
\begin{aligned}
\tfrac{\partial^2 log[g(x)]}{\partial x \partial x^T} &= -\tfrac{f_1(x)}{f_1(x)+f_2(x)}\tfrac{\partial^2 log[f_2(x)]}{\partial x \partial x^T} - \tfrac{f_2(x)}{f_1(x)+f_2(x)}\tfrac{\partial^2 log[f_1(x)]}{\partial x \partial x^T} \\
&\quad +\tfrac{f_1(x)}{f_1(x)+f_2(x)}(\nabla log[f_1(x)])(\nabla log[f_1(x)])^T \\
&\quad +\tfrac{f_2(x)}{f_1(x)+f_2(x)}(\nabla log[f_2(x)])(\nabla log[f_2(x)])^T \\
&\quad -[\tfrac{f_1(x)}{f_1(x)+f_2(x)}\tfrac{f_1(x)}{f_1(x)+f_2(x)}(\nabla log[f_1(x)])(\nabla log[f_1(x)])^T \\
&\quad +\tfrac{f_1(x)}{f_1(x)+f_2(x)}\tfrac{f_2(x)}{f_1(x)+f_2(x)}(\nabla log[f_1(x)])(\nabla log[f_2(x)])^T \\
&\quad +\tfrac{f_2(x)}{f_1(x)+f_2(x)}\tfrac{f_1(x)}{f_1(x)+f_2(x)}(\nabla log[f_2(x)])(\nabla log[f_1(x)])^T \\
&\quad +\tfrac{f_2(x)}{f_1(x)+f_2(x)}\tfrac{f_2(x)}{f_1(x)+f_2(x)}(\nabla log[f_2(x)])(\nabla log[f_2(x)])^T] \\
&= -\tfrac{f_1(x)}{f_1(x)+f_2(x)}\tfrac{\partial^2 log[f_2(x)]}{\partial x \partial x^T} - \tfrac{f_2(x)}{f_1(x)+f_2(x)}\tfrac{\partial^2 log[f_1(x)]}{\partial x \partial x^T} \\
&\quad +\tfrac{f_1(x)}{f_1(x)+f_2(x)}\tfrac{f_2(x)}{f_1(x)+f_2(x)}[(\nabla log[f_1(x)] - \nabla log[f_2(x)]) \\
&\quad (\nabla log[f_1(x)] - \nabla log[f_2(x)])^T]
\end{aligned}
$$

which is easily seen to be positive definite.

□

Eigenvalue Inequalities

Our first eigenvalue inequality relates to matrices that arise when a hermitian positive definite matrix is left and right multiplied by positive semi-definite matrices with eigenvalues less than or equal to 1. It shows that the eigenvalues for the resulting matrix is no larger than the eigenvalues for the original matrix. It plays a key role in establishing an correlation structure inequality relating posterior and prior correlation structures to each other (see our section on correlation structure inequalities).

Theorem 4 *Let A be a Hermitian positive definite matrix and let B be a Hermitian positive semi-definite matrix such that all its eigenvalues are less than or equal to 1. Then the eigenvalues satisfies the properties that for every i, $\lambda_i(BAB) \leq \lambda_i(A)$.*

Proof See Appendix B. $\square$

Our second eigenvalue inequality also relates to the multiplication of matrices. It shows that when Hermitian matrices that are "smaller" than the identity matrix are multiplied by their own transposes, the maximal eigenvalue of the resulting matrix is strictly bounded from above by 1. The result is leveraged in our section on correlation structure inequalities in order to establish properties of correlation structures and the result also plays a key role in establishing a Rosenthal [4] type drift condition.

Theorem 5 *Let A be a Hermitian positive semi-definite matrix and B a Hermitian positive definite matrix. Assume furthermore that $(A+B)^{-1}A$ is Hermitian. Then all eigenvalues for the matrix $T = A(A+B)^{-1}(A+B)^{-1}A$ are non-negative and less than 1.*

Proof See Appendix B. $\square$

Given that the matrix T is Hermitian even if $(A+B)^{-1}A$ is not, one may wonder whether the above properties hold even if $(A+B)^{-1}A$ fails to be Hermitian. The below counterexample shows that this is not generally the case.

Example 2 *Consider the following matrices*

$$A = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix}, B = \begin{pmatrix} 5 & -2 \\ -2 & 1 \end{pmatrix}.$$

Then we have

$$A + B = \begin{pmatrix} 10 & 0 \\ 0 & 2 \end{pmatrix}, (A+B)^{-1} = \begin{pmatrix} 0.1 & 0 \\ 0 & 0.5 \end{pmatrix}$$

and

$$(A+B)^{-1}A = \begin{pmatrix} 0.5 & 0.2 \\ 1 & 0.5 \end{pmatrix}, T = \begin{pmatrix} 1.25 & 0.6 \\ 0.6 & 0.29 \end{pmatrix}.$$

It is straightforward to verify that the eigenvalues for $(A+B)^{-1}A$ *are given approximately by* 0.94721 *and* 0.05278 *respectively, while the eigenvalues for* T *are given by approximately by* 1.53837 *and* 0.00162*. Hence the maximal eigenvalue for* T *is not below* 1*.*

A Model and A Sampler

We now introduce the model and the sampler. Throughout, we will use the following notation

1 β is a column vector of dimension m_1.
2 μ a prior mean column vector of dimension m_2.
3 X a design matrix of dimension $m_1 \times m_2$.
4 P a prior Hermitian precision positive definite matrix for β of dimension $m_1 \times m_1$.
5 ν a prior mean column vector of dimension m_2.
6 P_0 a prior Hermitian precision positive definite matrix for μ of dimension $m_2 \times m_2$

Using this notation, we can now introduce the models under consideration. To simplify expressions and arguments, we parameterize multivariate normal densities using their precision matrices instead of their variance-covariance matrices.

Definition 1 *A Hierarchical Bayesian Model With Multivariate Normal Prior is a model with the following specification:*

1 A likelihood function $f(\mathbf{y}|.)$*,*
2 A multivariate normal prior specification for β *given by* $\pi_\beta(.|X\mu, P)$*,*
3 A multivariate normal prior specification for μ *given by* $\pi_\mu(.|\nu, P_0)$*.*

In applications, it is common for the likelihood function to be of the form $\log(f(\mathbf{y}, \beta)) = \sum_{i=1}^{N} \log(f_i(\mathbf{y}_i, \beta_i))$ and for the matrix P to be block diagonal (i.e., one block associated with each subcomponent i in the sum). It is also common to give P its own prior specifications. Convergence rates for samplers associated with such models are not considered here but will be explored further under separate cover.

Throughout the paper, we will make use of log-concavity assumption 1 below. Indeed, the assumption appears to be critical to the results established. It is worth noting that most commonly used Hierarchical Bayesian generalized linear models (such as logistic regression, poisson regression, and normal data models) all satisfy this assumption. While not required to establish our main result, assumption 2 below (data precision bounded from above) does enable improved bounds on convergence rates for our samplers.

Assumption 1 *The likelihood function $f(\mathbf{y}|.)$ is log-concave.*

Assumption 2 *There exists a Hermitian positive definite upper bound matrix P_{UB}^D of dimension $m_1 \times m_1$ such that for every β, the function $h_1(.)$ defined by*

$$h_1(\beta) = log(f(\mathbf{y}|\beta)) + 0.5\beta^T P_{UB}^D \beta$$

is a (strictly) convex function. We will refer to any such matrix as a (strict) upper bound matrix for the data precision.

No assumptions will be made in the paper regarding properties of lower bound matrices for the data precision (other than the implicit one that the null matrix is one such matrix as long as assumption 1 is satisfied. To clarify the role they can play in impoving bounds on convergence, however, we include a formal definition of lower bound matrices below and include them explicitly in our bounding formulas.

Definition 2 *A Hermitian positive semi-definite definite matrix of dimension $m_1 \times m_1$ is a (strict) lower bound matrix for the data precision if for every β, the function $h_2(.)$ defined by*

$$h_2(\beta) = log(f(\mathbf{y}|\beta)) + 0.5\beta^T P_{LB}^D \beta$$

is a (strictly) concave function.

The following full rank assumption will also be leveraged in order to establish geometric convergence rates for the sampler below. It plays a role in ensuring that certain matrices leveraged in the proof are positive definite and hence invertible. In applications there may be ways of getting around violations of this assumption (e.g., by redefining dimensions so one subset can be simulated independently from the two-block Gibbs sampler while a subset satisfies the full rank assumption).

Assumption 3 *For every hermitian positive definite matrix V of dimension $m_1 \times m_1$ the matrix $X^T PVPX$ is positive definite.*

To clarify notation, we introduce the following additional definitions.

Definition 3 *We will denote by $\pi_\mu(.|\nu, P_0, \beta)$ the conditional density for μ given β, by $\pi_\mu(.|\nu, P_0, \mathbf{y})$ the posterior density for μ and by $\pi_\beta(.|X\mu, P, \mathbf{y})$ the conditional posterior density for β given μ and $\mathbf{y}$.*

Definition 4 *We will denote by $\pi_{\mu,\beta}(.|\nu, P_0, P)$ the joint prior density for μ and β defined by $\pi_{\mu,\beta}(\mu, \beta|\nu, P_0, P) = \pi_\mu(\mu|\nu, P_0)\pi_\beta(\beta|X\mu, P)$ and by $\pi_{\mu,\beta}(.|\nu, P_0, P, \mathbf{y})$ the joint posterior density for μ and β defined by*

$$\pi_{\mu,\beta}(\mu, \beta|\nu, P_0, P, \mathbf{y}) = \pi_\mu(\mu|\nu, P_0, \mathbf{y})\pi_\beta(\beta|X\mu, P, \mathbf{y}).$$

It is worth noting that the conditional density $\pi_{\mu}(.|\nu, P_0, \beta)$ for μ given β is multivariate normal and that *iid* samples can be produced using standard approaches. The conditional density $\pi_{\beta}(.|X\mu, P, \mathbf{y})$ for β given μ and $\mathbf{y}$ is a density resulting from a Multivariate normal prior and (if assumption 1 is satisfied) a log-concave likelihood function. If the latter assumption is satisfied, *iid* samples from this conditional density can be produced using the Likelihood subgradient density approach in Nygren and Nygren [5]. In many applications, this update can be performed using low dimensional updates for components of the vector β as they frequently are conditionally independent. An implementation of the likelihood subgradient approach is available in the glmbayes R package available on Github.

We are now ready to introduce the two-block Gibbs sampler that will be of main interest in the paper.

Definition 5 *A two-block Gibbs sampler for the joint posterior density of the Hierarchical Bayesian model with normal prior in definition 1 is a sampler that generates discrete time, time homogoneus Markov Chains $\mathcal{M} = \{\mathcal{M}^k, k = 0, 1, \ldots\}$ and $\mathcal{B} = \{\mathcal{B}^k, k = 0, 1, \ldots\}$ for which the joint transition Kernel $Q(.,.)$ is given by*

$$Q((\beta^{k-1}, \mu^{k-1}), (\beta^k, \mu^k)) = \pi_{\beta}(\beta^k|X\mu^{k-1}, P, \mathbf{y})\pi_{\mu}(\mu^k|\nu, P_0, \beta^k).$$

As the densities on the right above do not depend on β^{k-1}, we will simply write $Q(\mu^{k-1}, (\beta^k, \mu^k))$ instead of $Q((\beta^{k-1}, \mu^{k-1}), (\beta^k, \mu^k))$.

In many of the arguments used to establish our main result, we will leverage properties of the marginal and conditional transition Kernels defined below. In the statement of our main result, we will also leverage the notation for the k-step transition kernel for $\mathcal{M}$ also intoduced below.

Definition 6 *We will denote by $Q_{\mathcal{M}}(.,.)$ the marginal transition Kernel defined by $Q_{\mathcal{M}}(\mu^{k-1}, \mu^k) = \int Q(\mu^{k-1}, (\beta^k, \mu^k))d\beta^k$ and we will use the notation $Q(\mu^{k-1}, .|\mu^k)$ to denote the conditional transition Kernel for β defined by $Q(\mu^{k-1}, \beta^k|\mu^k) = \frac{Q(\mu^{k-1}, (\beta^k, \mu^k))}{\int Q(\mu^{k-1}, (\beta^k, \mu^k))d\beta^k}$.*

Definition 7 *For $k = 1, 2, 3, \ldots$, we will let $Q^k_{\mathcal{M}}(.,.)$ denote the k-step transition Kernel for $\mathcal{M}$.*

We will be concerned with *geometric ergodicity* for the Markov Chain $\mathcal{M}$. In particular, conditions under which there exists a function $g(.)$ and a constant $0 < t < 1$ such that for any μ,

$$\left\|Q^k_{\mathcal{M}}(\mu, .) - \pi_{\mu}(.|\nu, P_0, \mathbf{y})\right\| \leq g(\mu)t^k \tag{4}$$

for $k = 1, 2, 3, \ldots$ (where the distance is the total variation distance). In order to study these convergence rates, we will find it useful to understand properties of the transition kernel. In particular, we will seek bounds for Variance and Precision matrices associated with the sampler.

The bounds established in this paper are leveraged in a work in progress in order to establish conditions under which Rosenthal [4] type minorization and drift conditions are satisfied. Together these imply geometric ergodicity for the two-block Gibbs sampler with the following explicit bound (where details will be made available in the work in progress).

$$\begin{aligned} ||Q^k_{\mathcal{M}}(\mu^0,.) - \pi_\mu(.|y)|| \quad \leq \quad & (1 - (\tfrac{|P^K_{LB}|}{|P^K_{UB}|})^{0.5} e^{-\frac{1}{4}(\beta^{**}-\beta^*)^T P X W_2^{-1} X^T P(\beta^{**}-\beta^*)-\gamma})^{rk} \\ & +(U^r/\alpha^{1-r})^k [1 + \tfrac{2\,\mathrm{Tr}[W_1^{1/2} Var^{\mu}_{UB} W_1^{1/2}]}{(1-\lambda^*_{W1})} \\ & +(\mu^0 - \mu^*)^T W_1 (\mu^0 - \mu^*)] \end{aligned} \tag{5}$$

It is worth noting that results in both [4] and [6] have shown that Minorization and Drift conditions together imply Geometric Ergodicity (i.e., that the total variation distance from the target density decreases geometrically fast with the number of iterations). A number of authors [7–10] have leveraged these results in order to study the convergence rates for special cases of Gibbs samplers (typically models with normal data as opposed to the log-concave models considered here).

Bounds for Variance and Precision Matrices associated with the transition Kernel

In this section, we apply the variance inequalities from our Main Theorems section in order to establish properties for variance and precision matrices associated with the transition kernel for the sampler in the previous section. First, we introduce some additional notation and definitions used to state the inequalities of interest. Our next subsection then notes the equivalence of the conditional, join, and marginal transition kernels to a specific set of densities with multivariate normal priors and log-concave likelihood functions. Using these equivalence results, we then note properties that follow either from general properties of variance-covariance matrices and/or from our theorems.

Additional Definitions

Our interest here will be in the precision matrix associated with the marginal transition kernel and variance-covarianc matrices associated with the marginal and conditional transition kernels. The following gives the formal notation.

Definition 8 *We will denote by $P^K(\mu^{k-1}, \mu^k)$ the precision matrix for the marginal transition kernel $Q_{\mathcal{M}}(\mu^{k-1}, .)$. Likewise, we will denote by $Var_\mu(\mu^k|\mu^{k-1})$ the variance-covariance matrix for μ^k under the marginal transition Kernel $Q_{\mathcal{M}}(\mu^{k-1}, .)$ and by $Var_\beta(\beta^k|\mu^k, \mu^{k-1})$ the variance-covariance matrix for β^k under the conditional transition kernel $Q(\mu^{k-1}, .|\mu^k)$. For purposes of bounding, we will think of these as matrix valued function of μ^k and μ^{k-1}.*

In order to state the inequalities, we now introduce notation for how matrices are to be compared. The inequalities will be based on the extent to which differences between matrices are positive semi-definite and/or positive definite.

Definition 9 *Let $A(\mu^k, \mu^{k-1})$ be any hermitian matrix valued function of μ^k and μ^{k-1}. Then we will say that*

1. *A matrix A^{UB} is an upper bound matrix for $A(.,.)$ if for every pair (μ^k, μ^{k-1}), the matrix $A^{UB} - A(\mu^k, \mu^{k-1})$ is positive semi-definite.*
2. *A matrix A^{UB} is an strict upper bound matrix for $A(.,.)$ if for every pair (μ^k, μ^{k-1}), the matrix $A^{UB} - A(\mu^k, \mu^{k-1})$ is positive-definite.*
3. *A matrix A^{UB} is an super strict upper bound matrix for $A(.,.)$ if there exists a positive definite matrix B such that for every pair (μ^k, μ^{k-1}), the matrix $(A^{UB} - B) - A(\mu^k, \mu^{k-1})$ is positive-definite.*
4. *A matrix A^{LB} is a lower bound matrix for $A(.,.)$ if for every pair (μ^k, μ^{k-1}), the matrix $A(\mu^k, \mu^{k-1}) - A^{LB}$ is positive semi-definite.*
5. *A matrix A^{LB} is a strict lower bound matrix for $A(.,.)$ if for every pair (μ^k, μ^{k-1}), the matrix $A(\mu^k, \mu^{k-1}) - A^{LB}$ is positive-definite.*

Basic Properties of Variance and Precision Matrices

In order to enable tight bounds for the precision and variance-covariance matrices, we establish equivalence results for the conditional, joint and marginal transition kernel's that allow us to leverage associated precision matrices in the bounding expressions. These equivalence results follow because the transition kernel in a certain sense is bounded by a corresponding transition kernel associated with a two-block Gibbs sampler applied to the prior.

The Conditional Transition Kernel

The following equivalence result plays a key role in the establishment of geometric ergodicity for the sampler. The fact that the mean of the equivalent density has a particularly simple form is leveraged in our companion paper in order to rewrite a bounding function related to minorization.The expression for the precision matrix helps to bound the variance of the conditional transition kernel, which in turn allows us to bound the precision of the marginal transition kernel.

Lemma 2 *Let P_{LB}^D be any lower bound matrix for the data precision as in definition 2 and define a function $g(.)$ by $g(\beta) = exp(-0.5\beta^T P_{LB}^D \beta)$. Let $\tilde{P}_0 = P_0 + X^T P X$ and $\tilde{\nu} = (P_0 + X^T P X)^{-1} P_0 \nu$ and*

$$\tilde{\tilde{\nu}}(\beta^k) = \tilde{\nu} + (P_0 + X^T P X)^{-1} X^T P \beta^k$$

Then the conditional transition Kernel $Q(\mu^{k-1}, .|\mu^k)$ is equivalent to a posterior density arising from a multivariate normal prior density $\pi_\beta(.|\tilde{X}(\mu^k + \mu^{k-1} - \tilde{\nu}), \tilde{P})$ and likelihood function $\tilde{f}_\beta(\mathbf{y}|.)$ defined by

$$\tilde{f}_\beta(\mathbf{y}|\beta) = f(\mathbf{y}|\beta)/g(\beta)$$

where

$$\tilde{P} = P + PX(P_0 + X^T P X)^{-1} X^T P + P_{LB}^D$$

and $\tilde{X} = \tilde{P}^{-1}PX$ *(which implies that* $\tilde{P}\tilde{X} = PX$*). Moreover, if assumption 2 is satisfied with* P_{UB}^D *as an upper bound matrix for the data precision, then* $\tilde{P}_{UB}^D = P_{UB}^D - P_{LB}^D$ *is an upper bound for the precision of the likelihood function* $\tilde{f}_\beta(\mathbf{y}|.)$*.*

Proof See supplement file "Equivalence of Densities and an Expression for the Precision of the Marginal Transition Kernel.pdf". □

The Joint Transition Kernel

The following equivalence result serves as an import interim result that enables the marginal transition kernel result below.

Lemma 3 *The joint transition Kernel* $Q(\mu^{k-1}, .)$ *is equivalent to a posterior density arising from a multivariate normal prior density* $\tilde{Q}(\mu^{k-1}, .)$ *defined by*

$$\tilde{Q}(\mu^{k-1}, (\beta, \mu^k)) = \frac{g(\beta)\pi_\mu(\mu^k|\tilde{\tilde{\nu}}(\beta^k), \tilde{P}_0)\pi_\beta(\beta^k|X\mu^{k-1}, P)}{\int\int[g(\beta)\pi_\mu(\mu^k|\tilde{\tilde{\nu}}(\beta^k), \tilde{P}_0)\pi_\beta(\beta^k|X\mu^{k-1}, P)]d\beta d\mu^k}$$

and a likelihood function $\tilde{f}_\beta(\mathbf{y}|.)$ *defined by*

$$\tilde{f}_\beta(\mathbf{y}|\beta) = f(\mathbf{y}|\beta)/g(\beta).$$

Moreover, the multivariate normal density has precision matrix P^* *given by*

$$P^* = \begin{pmatrix} P^*_{\beta,\beta} & P^*_{\beta,\mu} \\ P^*_{\mu,\beta} & P^*_{\mu,\mu} \end{pmatrix} = \begin{pmatrix} P + PX(P_0 + X^TPX)^{-1}X^TP + P_{LB}^D & -PX \\ -X^TP & P_0 + X^TPX \end{pmatrix}$$

Proof See supplement file "Equivalence of Densities and an Expression for the Precision of the Marginal Transition Kernel.pdf". □

The Marginal Transition Kernel

This result regarding the equivalence for the transition kernel allows us to bound the variance for the marginal transition kernel from above.

Lemma 4 *The marginal transition Kernel* $Q_{\mathcal{M}}(\mu^{k-1}, .)$ *is equivalent to a posterior density arising from a multivariate normal prior density* $\tilde{Q}_{\mathcal{M}}(\mu^{k-1}, .)$ *with precision matrix*

$$P_0 + X^TPX - X^TP(P + PX(P_0 + X^TPX)^{-1}X^TP + P_{LB}^D)^{-1}PX$$

and log-concave likelihood function $\tilde{f}_\mu(\mathbf{y}|.)$ *defined by*

$$\tilde{f}_\mu(\mathbf{y}|\mu) = \int \pi_\beta(.|\tilde{X}(\mu^k + \mu^{k-1} - \tilde{\nu}), \tilde{P})\tilde{f}_\beta(\mathbf{y}|\beta)d\beta$$

where

1 $\pi_\beta(.|\tilde{X}(\mu^k + \mu^{k-1} - \tilde{\nu}), \tilde{P})$ *and* $\tilde{f}_\beta(\mathbf{y}|.)$ *are as in lemma 2, and*
2 $\tilde{Q}_{\mathcal{M}}(\mu^{k-1}, .)$ *is the marginal density for the density* $\tilde{Q}(\mu^{k-1}, .)$ *in lemma 3.*

Proof See supplement file "Equivalence of Densities and an Expression for the Precision of the Marginal Transition Kernel.pdf". □

The following espression for the precision of the marginal transition kernel allows us to bound the precision from above and below by leveraging bounds on the variance for β under the conditional transition kernel.

Lemma 5 *The precision matrix for the marginal transition kernel can explicitly be expressed as*

$$P^K(\mu^{k-1}, \mu^k) = P_0 + X^T PX - X^T PVar_\beta(\beta|X\mu^{k-1}, \mu^k, P, \mathbf{y})PX$$

Proof See supplement file "Equivalence of Densities and an Expression for the Precision of the Marginal Transition Kernel.pdf". □

Properties implied by General Properties of Variances

The variance-covariance matrix is always bounded below by the null matrix $\mathbf{0}$. If we let $P^K_{UB} = P_0 + X^T PX$, then it follows from this basic property and Lemma 5 that P^K_{UB} is an upper bound matrix for $P^K(.,.)$. If the full rank assumption 3 is satisfied, then this upper bound is also a strict upper bound.

Properties Following from Theorem 1 and Corollary 1

Suppose a Hierarchical Bayesian Model satisfies the log-concavity assumption 1. Given lemmas 2 and 4, items (1) and (3) below are then straightforward applications of Theorem 1. It is also straightforward to verify that Lemmas 2 and 5 together with Theorem 1 imply item (2).

1 Let $Var^\beta_{UB} = (P + PX(P_0 + X^T PX)^{-1}X^T P + P^D_{LB})^{-1}$. Then Var^β_{UB} is an upper bound matrix for the function $Var(\beta|., ., P, \mathbf{y})$.
2 Let P^K_{LB} be the matrix defined by

$$P^K_{LB} = P_0 + X^T PX - X^T P[P + PX[P_0 + X^T PX]^{-1}X^T P + P^D_{LB}]^{-1}PX.$$

Then P^K_{LB} is a lower bound matrix for the function $P^K(.,.)$.
3 Let $Var^\mu_{UB} = (P^K_{LB})^{-1}$. Then the matrix Var^μ_{UB} is an upper bound matrix for $Var_\mu(.|.)$.

Theorem 1 and items (2) and (3) all play key roles in the estblishment of Rosenthal [4] type minorization and drift conditions. Theorem 1 allows for the establishment of Lemma 8 which helps provide the slope parameter in Rosenthal's drift condition. Item (2) plays a key role in the establishment of the minorization condition as it allows the establishment of lemma 7. Item (3) finally helps bound a variance term associated with the drift condition and indirectly provides the intercept term in Rosenthal's drift condition.

Properties Following from Theorem 2

Suppose a Hierarchical Bayesian Model satisfies both the log-concavity assumption 1 and the data precision upper bound assumption 2. Given Lemma 2, item (1) below is then a straightforward application of Theorem 2. Item (1) and Lemma 5 then together imply item (2).

1 The matrix $Var_{LB}^{\beta} = (P + PX(P_0 + X^T PX)^{-1} X^T P + P_{UB}^D)^{-1}$ is well defined and is a lower bound matrix for the function $Var(\beta|., ., P, \mathbf{y})$.

2 The matrix P_{UB2}^K defined by

$$P_{UB2}^K = P_0 + X^T PX - X^T P[P + PX[P_0 + X^T PX]^{-1} X^T P + P_{UB}^D]^{-1} PX.$$

is well defined and is an upper bound matrix for the function $P^K(., .)$.

While not required in order to establish drift and minorization conditions, item (2) can help provide improved bounds. This can particularly be seen from the role played by the upper bound for the precision of the marginal transition kernel in Lemma 7 below.

Correlation Structure Inequalities

In this section, we apply the variance and eigenvalue inequalities in order to establish a set of correlation structure inequalities. In the first subsection, we establish correlation structure inequalities that play a role in establishing a minorization condition, while the second covers correlation structure inequalities that play a role in establishing a Rosenthal [4] type drift condition.

Correlation Structure Inequality For Minorization Condition

The establishment of Minorization for the sampler under consideration relies on inequalities involving correlation structures associated with the transition Kernel. In order to clarify this, we first introduce a multivariate generalization of the squared correlation distance. The definition of a squared correlation distance matrix is made using variance-covariance matrices. Our lemma 6 below shows that the determinant of the squared correlation distance matrix is a natural generalization of the squared correlation distance.

Definition 10 *For a Variance Covariance Matrix Σ, the Squared Correlation Distance Matrix $D_{\mathcal{P}_{AB}}(\Sigma)$ between blocks A and B will be defined by*

$$D_{\mathcal{P}_{AB}}(\Sigma) = I - \Sigma_{AA}^{-1} \Sigma_{AB} \Sigma_{BB}^{-1} \Sigma_{BA}$$

Lemma 6 *Let Σ be a hermitian positive definite matrix and partition the columns of Σ into two blocks A and B. Then*

$$0 < |D_{\mathcal{P}_{AB}}(\Sigma)| \leq 1. \tag{6}$$

Proof See Appendix C. □

If A and B form a partition of the columns of the matrix Σ into two blocks, then the marginal precision for A will be denoted $P^M_{AA} = (\Sigma_{AA})^{-1}$ and the conditional precision for A by $P^C_{AA} = (\Sigma_{AA} - \Sigma_{AB}\Sigma^{-1}_{BB}\Sigma_{BA})^{-1}$. We then have

$$\begin{aligned} |D_{\mathcal{P}_{AB}}(\Sigma)| &= |I - \Sigma^{-1}_{AA}\Sigma_{AB}\Sigma^{-1}_{BB}\Sigma_{BA}| \\ &= |\Sigma^{-1}_{AA}||\Sigma_{AA} - \Sigma_{AB}\Sigma^{-1}_{BB}\Sigma_{BA}| \\ &= \frac{|P^M_{AA}|}{|P^C_{AA}|} \end{aligned}$$

and hence the ratio $\frac{|P^M_{AA}|}{|P^C_{AA}|}$ can be viewed as a generalization of the squared correlation distance. A key inequality enabling the establishment of minorization is provided by lemma 7 below. Before stating the lemma, we introduce the concept of a fixed point.

Definition 11 *A vector μ^* is a fixed point for the marginal transition Kernel if the expectation for μ under the transition kernel $Q_{\mathcal{M}}(\mu^*, .)$ equals μ^*.*

We now state our second correlation structure inequality.

Lemma 7 (Transition Kernel:Correlation Distance Inequality) *Let μ^* be a fixed point for the Marginal transition kernel, let μ^{**} be the mode for the transition Kernel $Q(\mu^*, .)$, and let P^K_{UB} and P^K_{LB} be positive definite upper and lower bound matrices for the precision matrix of the marginal transition Kernel. Let $S(.|P^K_{UB})$ be the multivariate normal density with precision matrix P^K_{UB} and mean μ^{**}. Then for every μ^k,*

$$\log[\frac{Q(\mu^*, \mu^k)}{S(\mu^{**}|P^K_{UB})}] \geq 0.5 \log[\frac{|P^K_{LB}|}{|P^K_{UB}|}]. \tag{7}$$

Proof See Appendix C. □

To see that this inequality indeed can be viewed as a correlation distance inequality, it is useful to consider the case where we set $P^K_{UB} = P_0 + X^T PX$, $P^K_{LB} = P_0 + X^T PX - X^T P[P + PX[P_0 + X^T PX]^{-1}X^T P]^{-1}PX$, and where the data is multivariate normal with precision P^D. In this case, the inequality becomes $0.5 \log[\frac{|P^K|}{|P^K_{UB}|}] \geq 0.5 \log[\frac{|P^K_{LB}|}{|P^K_{UB}|}]$. If we denote by $\Sigma^{Post,TK}$ the variance-covariance matrix associated with the transition kernel for the two block Gibbs sampler applied to the posterior density and by $\Sigma^{Prior,TK}$ the variance-covariance matrix associated with the transition kernel for a two-block Gibbs sampler applied to the prior, then the inequality can be re-expressed as

$$\begin{aligned} \log[\frac{Q(\mu^*, \mu^{**})}{S(\mu^{**}|P^K_{UB})}] &= 0.5|I - (\Sigma^{Post,TK}_{AA})^{-1}\Sigma^{Post,TK}_{AB}(\Sigma^{Post,TK}_{BB})^{-1}\Sigma^{Post,TK}_{BA}| \\ &\geq 0.5|I - (\Sigma^{Prior,TK}_{AA})^{-1}\Sigma^{Prior,TK}_{AB}(\Sigma^{Prior,TK}_{BB})^{-1}\Sigma^{Prior,TK}_{BA}| \\ &= 0.5 \log[\frac{|P^K|}{|P^K_{UB}|}] \end{aligned}$$

which says that the squared correlation matrix distance for the transition kernel associated with the posterior two-block Gibbs sampler exceeds the squared correlation matrix distance for the transition kernel associated with the prior two-block Gibbs sampler.

Correlation Structure Inequality for Geometric Drift Condition

The establishment of geometric drift for the sampler under consideration relies on correlation structure inequalities involving the posterior density. In order to clarify this, we will introduce a multivariate generalization of the squared correlation. The formulation will make use of eigenvalues so we first introduce notation for eigenvalues and singular values.

Consider a square matrix A of dimension $m_2 \times m_2$. Then we will denote by $\lambda_1(A) \geq \lambda_2(A) \ldots \geq \lambda_{m_2}(A)$ the ordered eigenvalues for A and by $\sigma_1(A) \geq \sigma_2(A) \ldots \geq \sigma_{m_2}(A)$ the ordered singular values for A.

1 The *spectral norm* of a square matrix A will be defined by

$$\|A\|_{SP} = \sqrt{\lambda_1(A^T A)}.$$

2 The *nuclear norm* of a square matrix A will be defined by

$$\|A\|_{NU} = \mathrm{Tr}\,((A^T A)^{1/2}).$$

3 For a precision matrix P, the *Squared Correlation Matrix* $\mathcal{P}_{AB}(P)$ between blocks A and B will be defined by

$$\mathcal{P}_{AB}(P) = P_{AA}^{-1} P_{AB} P_{BB}^{-1} P_{BA}$$

When blocks A and B are both one-dimensional, the Squared Correlation Matrix $\mathcal{P}_{AB}(P)$ reduces to ρ^2, the correlation between the two blocks, and hence the Squared Correlation Matrix itself as well as its spectral and nuclear norms can both be viewed as multivariate generalizations of the squared correlation. Similarly, the Squared Correlation Distance Matrix reduces to the correlation distance and hence the Squared Correlation Distance Matrix and its determinant can both be viewed as multivariate generalizations of the correlation distance.

We now introduce some additional notation related to the model and sampler in order to be able to compare correlation structures associated with the posterior density to correlation structures associated with a bounding density. It is worth noting that the bounds on the Variance-covariance matrix from Corollary 1 also applies to $Var_{\mu}^{\beta}(\mu^{(k-1)})$ below.

1 Let μ be any point. Then we will denote by $Var_{\mu}^{\beta}(\mu^{(k-1)})$ the matrix with element ij given by $\int_0^1 Cov[\beta_i, \beta_j | X(\mu + t(\mu^{(k-1)} - \mu)), P, \mathbf{y}]dt$.

2 Let $\tilde{R}$ be any Hermitian positive definite matrix of dimension $m_2 \times m_2$. We will then denote by $\tilde{P}^*_{\tilde{R}}(Var_{\mu^*}^{\beta}(\mu^{(k-1)}))$ the *transformed average posterior precision matrix* defined by

$$\begin{aligned}
\tilde{P}^*_{\tilde{R},AA}(Var_{\mu^*}^{\beta}(\mu^{(k-1)})) &= \tilde{R}^{-1/2}(P_0 + X^T P X)\tilde{R}^{-1/2} \\
\tilde{P}^*_{\tilde{R},AB}(Var_{\mu^*}^{\beta}(\mu^{(k-1)})) &= -\tilde{R}^{-1/2} X^T P \\
\tilde{P}^*_{\tilde{R},BA}(Var_{\mu^*}^{\beta}(\mu^{(k-1)})) &= -P X \tilde{R}^{-1/2} \\
\tilde{P}^*_{\tilde{R},BB}(Var_{\mu^*}^{\beta}(\mu^{(k-1)})) &= (Var_{\mu^*}^{\beta}(\mu^{(k-1)}))^{-1}.
\end{aligned}$$

3 Let $\tilde{R}$ be any Hermitian positive definite matrix of dimension $m_2 \times m_2$. We will then denote by $P^*_{\tilde{R}}$ the *transformed posterior precision bounding matrix* defined by

$$\begin{aligned}
P^*_{\tilde{R},AA} &= \tilde{R}^{-1/2}(P_0 + X^T P X)\tilde{R}^{-1/2} \\
P^*_{\tilde{R},AB} &= -\tilde{R}^{-1/2}X^T P \\
P^*_{\tilde{R},BA} &= -PX\tilde{R}^{-1/2} \\
P^*_{\tilde{R},BB} &= P + P^D_{LB}.
\end{aligned}$$

To understand the above two matrices (and the terminology) it is useful to consider the case where the data is normal and $\tilde{R}$ is equal to the identity matrix. In that case, $(Var^{\beta}_{\mu^*}(\mu^{(k-1)}))^{-1} = P + P^D$ and the matrix $\tilde{P}^*_{\tilde{R}}(V_{\mu^*}(\mu^{(k-1)}))$ is just the precision matrix for the posterior density. In the above formula, the precision $P^*_{\tilde{R},BB}(Var^{\beta}_{\mu^*}(\mu^{(k-1)}))$ is the inverse of an average of the conditional variance-covariance matrix for β over a range of values for μ as μ moves between the fixed point μ^* and μ^{k-1} (hence the teminology *average posterior precision matrix*). The interpretation of the matrix $P^*_{\tilde{R}}$ depends on the value for P^D_{LB}. If $P^D_{LB} = \mathbf{0}$, then $P^*_{\tilde{R}}$ is just the precision matrix for the prior density. If $P^D_{LB} = P^D$, then it is also equal to the precision for the posterior density. In all cases, the latter matrix bounds the former from below (i.e,, the difference between the former and the latter is positive semi-definite). We now show that when $\tilde{R}$ is selected carefully two key correlation structure inequalities hold.

Lemma 8 *Suppose the log-concavity assumption 1 and the full rank assumption 3 are both satisfied. Let* $\tilde{R} = tX^T P(P + P^D_{LB})^{-1}PX$ *for some positive constant* t *and let* $P^*_{\tilde{R}}$ *and*

$$\tilde{P}^*_{\tilde{R}}(Var^{\beta}_{\mu^*}(\mu^{(k-1)}))$$

be as above. Then

1 The eigenvalues satisfies the properties that for every i

$$\begin{aligned}
\lambda_i((\mathcal{P}_{AB}(\tilde{P}^*_{\tilde{R}}(Var^{\beta}_{\mu^*}(\mu^{(k-1)}))))^T \\
(\mathcal{P}_{AB}(\tilde{P}^*_{\tilde{R}}(Var^{\beta}_{\mu^*}(\mu^{(k-1)}))))) \quad &\leq \quad \lambda_i((\mathcal{P}_{AB}(P^*_{\tilde{R}}))^T(\mathcal{P}_{AB}(P^*_{\tilde{R}})))
\end{aligned} \tag{8}$$

2 and the spectral norms satisfy the property that

$$\left\|\mathcal{P}_{AB}(\tilde{P}^*_{\tilde{R}}(Var^{\beta}_{\mu^*}(\mu^{(k-1)})))\right\|_{SP} \leq \left\|\mathcal{P}_{AB}(P^*_{\tilde{R}})\right\|_{SP}. \tag{9}$$

Proof See Appendix C. □

When $P^D_{LB} = \mathbf{0}$, this inequality essentially states that the correlation between blocks under the posterior density is bounded above by the correlation between blocks under the prior density.

The following result now shows that under the same restriction on $\tilde{R}$, the spectral norm for the Squared Correlation Matrix is a natural extension of its univariate counterparts. The spectral norm and the nuclear norm are also related in that $\|A\|_{NU} \leq m_2 \|A\|_{SP}$.

Lemma 9 *Suppose the full rank assumption 3 is satisfied and let* $\tilde{R} = tX^T P(P + P_{LB}^D)^{-1} PX$ *for some positive constant* t. *Then*

1. $\mathcal{P}_{AB}(P^*_{\tilde{R}})$ *is a Hermitian positive definite matrix*
2. $\left\| \mathcal{P}_{AB}(P^*_{\tilde{R}}) \right\|_{SP} = \lambda_1(\mathcal{P}_{AB}(P^*_{\tilde{R}}))$*, and*
3. $0 < \left\| \mathcal{P}_{AB}(P^*_{\tilde{R}}) \right\|_{SP} < 1$.

Proof See Appendix C. □

In an article in progress, we leverage this result in order to establish a Rosenthal [4] type drift condition. To understand the role played by the above inequality in the proof, let $\lambda^*_{\tilde{R}} = [\left\| \mathcal{P}_{AB}(P^*_{\tilde{R}}) \right\|_{SP}]^2$,

$$\lambda^*_{\tilde{R}}(Var^{\beta}_{\mu^*}(\mu^{(k-1)})) = [\left\| \mathcal{P}_{AB}(\tilde{P}^*_{\tilde{R}}(Var^{\beta}_{\mu^*}(\mu^{(k-1)}))) \right\|_{SP}]^2$$

and

$$T^*_{\tilde{R}}(Var^{\beta}_{\mu^*}(\mu^{k-1})) = [\mathcal{P}_{AB}(\tilde{P}^*_{\tilde{R}}(Var^{\beta}_{\mu^*}(\mu^{(k-1)})))]^T \mathcal{P}_{AB}(\tilde{P}^*_{\tilde{R}}(Var^{\beta}_{\mu^*}(\mu^{(k-1)}))).$$

If μ^* is a fixed point, then it can be shown that lemma 8 implies the following

$$\begin{aligned}
(E[\mu^{(k)}|\mu^{(k-1)}] - \mu^*)^T \tilde{R} \\
(E[\mu^{(k)}|\mu^{(k-1)}] - \mu^*) &= (\mu^{(k-1)} - \mu^*))^T \tilde{R}^{1/2} T^*_{\tilde{R}}(Var^{\beta}_{\mu^*}(\mu^{k-1})) \\
&\quad \tilde{R}^{1/2}(\mu^{(k-1)} - \mu^*) \\
&\leq \lambda^*_{\tilde{R}}(Var^{\beta}_{\mu^*}(\mu^{(k-1)})(\mu^{(k-1)} - \mu^*)^T \tilde{R}(\mu^{(k-1)} - \mu^*) \\
&\leq \lambda^*_{\tilde{R}}(\mu^{(k-1)} - \mu^*)^T \tilde{R}(\mu^{(k-1)} - \mu^*)
\end{aligned} \tag{10}$$

where $0 \leq \lambda^*_{\tilde{R}} < 1$ follows from lemma 9.

Concluding Discussion

In this paper, we established Variance and Eigenvalue inequalities and applied them in order to establish properties of a two-block Gibbs sampler applied to a general class of Hierarchical Bayesian models with log-concave likelihood functions and Multivariate normal prior structures. We also discussed some of the connections of the results to minorization and drift conditions for the two-block Gibbs sampler. In a companion paper, we leverage the properties established here in order to establish conditions under which the two-block Gibbs sampler indeed is Geometrically ergodic. It is also possible that the results and approaches here can be applied to the convergence rates of other samplers as well.

The variance and eigenvalue inequalities established may also have broader applicability beyond the two-block Gibbs sampler considered here. Our main Variance inequality, for instance, can be viewed as a strengthening of the Variance-inequality of Raiffa and Schlaifer [3] for models with multivariate normal priors and log-concave likelihood function. This may have applicability in the decision theoretic area as the benefits of additional data are particularly strong in this case. Indeed, for every outcome of an experiment with a log-concave likelihood function, the posterior variance is lower than the prior variance. Likewise, our eigenvalue inequalities relates to the convergence of matrices generally. As the convergence of matrices plays a role in many areas, the two step-approach taken here may have broad applications outside of the application considered here.

Appendix A: Variance Inequalities

In this appendix, we state several variance inequalities that together build towards our main theorems. The first result establishes a variance inequality for univariate strictly log-concave likelihood functions, while the second and third extend this to univariate log-concave and multivariate log-concave likelihood functions. We then extend the result to show that the prior variance-covariance matrix bounds the posterior variance-covariance matrix.

Claim 1 *Let $f(\mathbf{y}|.)$ be a strictly log-concave likelihood function with a normal prior $\pi(.)$. Then*

$$\int (\beta - E^{\pi(.|y)}[\beta])^2 \pi(\beta|y) d\beta \leq \int (\beta - E^{\pi(.)}[\beta])^2 \pi(\beta) d\beta. \tag{11}$$

Proof Follows from claims 4 and 5. □

Theorem 6 *Let $f(\mathbf{y}|.)$ be a log-concave likelihood function with a normal prior $\pi(.)$. Then*

$$\int (\beta - E^{\pi(.|y)}[\beta])^2 \pi(\beta|y) d\beta \leq \int (\beta - E^{\pi(.)}[\beta])^2 \pi(\beta) d\beta. \tag{12}$$

Proof We note that the set of concave functions is the closure of the set of strictly concave functions. Hence for any log-concave likelihood function $f(\mathbf{y}|.)$, there exists a sequence $\{f^k(.)\}_{k=1}^{\infty}$ of strictly log-concave likelihood functions such that $\lim_{k\to\infty} f^k(\beta) = f(\beta)$ for every β. From claim 1 we know that the desired inequality holds for every $f^k(.)$. Hence it follows that the inequality also must hold for the limiting function $f(\mathbf{y}|.)$. □

Theorem 7 *Let $f(\mathbf{y}|.)$ be a log-concave likelihood function with a multivariate normal prior $\pi(.)$. Then for every dimension i,*

$$\int (\beta_i - E^{\pi(.|y)}[\beta_i])^2 \pi(\beta|y) d\beta \leq \int (\beta_i - E^{\pi(.)}[\beta_i])^2 \pi(\beta) d\beta. \tag{13}$$

Proof For multivariate normal priors, it is straightforward to verify, using the result from Prekopa [11], that the marginal likelihood for each i is well defined and itself log-concave. Our result then follows from Theorem 6. □

We now further extend these results. Let us first start with a couple of remarks.

Remark 1 *Suppose β has a multivariate normal prior and a log-concave likelihood function. Consider a one-to-one linear transformation of β, $\tilde{\beta} = A\beta$, for which $\beta = A^{-1}\tilde{\beta}$. Then the prior density for $\tilde{\beta}$ is also multivariate normal with a log-concave likelihood function. Moreover, each of the components of $\tilde{\beta}$ satisfies Theorem 7.*

Remark 2 *Let x be a vector (not all components zero), and let $z = x^T\beta$. Then the prior density for z is a normal distribution, the likelihood function log-concave, and the posterior variance less than or equal to the prior variance.*

Proof Construct a matrix A as in our above remark, with x^T as the first row. The result then follows immediately from our earlier remark. □

Fact 1 *Let $x \in \mathbf{R}^m$ be any vector. Then $x^T Var(\beta)x = Var(x^T\beta)$.*

Proof We note that

$$\begin{aligned} x^T Var(\beta)x &= \textstyle\sum_{i=1}^m \sum_{j=1}^m x_i Cov(\beta_i, \beta_j)x_j \\ &= \textstyle\sum_{i=1}^m \sum_{j=1}^m x_i(E[\beta_i\beta_j] - E[\beta_i]E[\beta_j])x_j \\ &= \textstyle\sum_{i=1}^m \sum_{j=1}^m (E[x_i\beta_i x_j\beta_j] - E[x_i\beta_i]E[x_j\beta_j]) \\ &= \textstyle\sum_{i=1}^m \sum_{j=1}^m (E[x_i\beta_i x_j\beta_j]) - \sum_{i=1}^m \sum_{j=1}^m (E[x_i\beta_i]E[x_j\beta_j]) \\ &= E[(\textstyle\sum_{i=1}^m x_i\beta_i) * (\sum_{j=1}^m x_j\beta_j)] - (\sum_{i=1}^m E[x_i\beta_i]) * (\sum_{j=1}^m E[x_j\beta_j]) \\ &= E[(x^T\beta) * (x^T\beta)] - (E[x^T\beta]) * (E[x^T\beta]) \\ &= Var(x^T\beta). \end{aligned}$$

□

Proof of Theorem 2 Consider any vector $x \in \mathbf{R}^m$. Then

$$\begin{aligned} x^T(\Sigma - Var(\beta|X\mu, \mathbf{y}))x &= x^T\Sigma x - x^T Var(\beta|X\mu, \mathbf{y})x \\ &= Var(x^T\beta) - Var(x^T\beta|X\mu, \mathbf{y}) \\ &\geq 0 \end{aligned}$$

where the equality follows from fact 1, and the inequality from remark 2. □

The rest of this appendix contains most of the proofs for our results. The main arguments are contained in the next subsection, while some of the lengthier details are contained in a supplemental file to the present article: "Details Related to Variance Inequalites.pdf".

Main Arguments

Throughout this section, we will denote the mean of densities $f(.)$ and $q(.)$ by $\bar{x}_f$ and $\bar{x}_q$ respectively. We start by noting the following basic property of the variance of any density $f(.)$.

Remark 3

$$\int (\bar{x}_f - x)^2 f(x)dx \leq \int (\bar{x}_q - x)^2 f(x)dx \tag{14}$$

Proof Well known property of the mean. □

We also note the following properties for two densities $f(.)$ and $q(.)$.

Claim 2 *Suppose densities $f(.)$ and $q(.)$ satisfies the properties that there exists r such that*
(i)

$$[q(x) < f(x)] \Rightarrow [|x - \bar{x}_q| < r]$$

and (ii)

$$[q(x) > f(x)] \Rightarrow [|x - \bar{x}_q| > r]$$

then

$$\int (x - \bar{x}_q)^2 f(x)dx \leq \int (x - \bar{x}_q)^2 q(x)dx. \tag{15}$$

Proof Let $A := \{x \in \mathbf{R}|f(x) > q(x)\}$ and $B := \{x \in \mathbf{R}|f(x) < q(x)\}$. Note that $\int_{x\in A}(f(x)-q(x))dx = \int_{x\in B}(q(x)-f(x))dx$. Under the present assumptions, it then follows that

$$\begin{array}{rcl}\int_{x\in A}(x - \bar{x}_q)^2(f(x) - q(x))dx & \leq & \int_{x\in A} r^2(f(x) - q(x))dx \\ & = & \int_{x\in B} r^2(q(x) - f(x))dx \\ & \leq & \int_{x\in B}(x - \bar{x}_q)^2(q(x) - f(x))dx\end{array}$$

Rearranging terms, we have

$$\begin{array}{rcl}\int_{x\in A\bigcup B}(x - \bar{x}_q)^2 f(x)dx & \leq & \int_{x\in A\bigcup B}(x - \bar{x}_q)^2 q(x)dx \\ & \Updownarrow & \\ \int_{x\in \mathbf{R}}(x - \bar{x}_q)^2 f(x)dx & \leq & \int_{x\in \mathbf{R}}(x - \bar{x}_q)^2 q(x)dx\end{array}$$

which completes the proof. □

Combining Remark 3 with Claim 2 allows us to state the following result relating the variances of two densities $f(.)$ and $q(.)$.

Claim 3 *Suppose densities $f(.)$ and $q(.)$ satisfies the properties that there exists r such that*

(i)

$$[q(x) < f(x)] \Rightarrow [|x - \bar{x}_q| < r]$$

and (ii)

$$[q(x) > f(x)] \Rightarrow [|x - \bar{x}_q| > r]$$

then

$$\int (x - \bar{x}_f)^2 f(x)dx \leq \int (x - \bar{x}_q)^2 q(x)dx. \tag{16}$$

Proof Follows from Remark 3 and Claim 2. □

Our next result relates the variance of posterior densities with normal priors to normal densities with the same variance as the prior.

Claim 4 *Let $f(.)$ be a posterior density with a normal prior, and let q be a normal density with the same variance as the prior. Assume that the likelihood function is log-concave and that there exists $r > 0$ such that*

$$q(\bar{x}_q - r) = f(\bar{x}_q - r)$$

and

$$q(\bar{x}_q + r) = f(\bar{x}_q + r).$$

then

$$\int (x - \bar{x}_f)^2 f(x)dx \leq \int (x - \bar{x}_q)^2 q(x)dx. \tag{17}$$

Proof We will show that the two assumptions of claim 3 are satisfied. First note that $log(f(x)) - log(q(x))$ is a concave function (in fact it is a linear transformation of the log-likelihood function). It follows that for any a, the set $\{x \in \mathbf{R} | log(f(x)) - log(q(x)) \geq a\}$ is a convex set. In particular, this holds when $a = 0$, which implies that the set $\{x \in \mathbf{R} | f(x) \geq q(x)\}$ is convex. Hence $f(x)$ must be greater than or equal to $q(x)$ for every $x \in [\bar{x}_q - r, \bar{x}_q - r]$. Hence property (ii) in claim 3 must hold.

To see that $f(x) <= q(x)$ outside of the interval, first note that $log(f(x)) - log(q(x))$ is a concave function and that $log(f(x)) - log(q(x)) >= 0$ for any x' in the interior of the interval $[\bar{x}_q - r, \bar{x}_q - r]$. Now, suppose there existed a point x'' outside of the interval where $log(f(x'')) - log(q(x'')) > 0$. Since $\bar{x}_q - r$ or $\bar{x}_q + r$ is a strict convex combination of x' and x'', it follows from the concavity that either $log(f(\bar{x}_q - r)) - log(q(\bar{x}_q - r)) > 0$ or $log(f(\bar{x}_q + r)) - log(q(\bar{x}_q + r)) > 0$, a contradiction. Hence property (i) of claim 3 must also hold. Our result hence follows from Claim 3. □

Our next claim shows the the required $r > 0$ indeed do exist for strictly log-concave likelihood functions. It relies on rather extensive arguments in a supplemental file to the present article: "Details Related to Variance Inequalites.pdf".

Claim 5 *Assume that the likelihood function is strictly log-concave and the prior is normal. Then there exists a normal density $q^*()$ with the same variance as the prior, and a constant $r > 0$ such that*

$$q^*(\bar{x}_q - r) = f(\bar{x}_q - r)$$

and

$$q^*(\bar{x}_q + r) = f(\bar{x}_q + r).$$

Proof See supplemental file "Details Related to Variance Inequalites.pdf". □

Theorem 2

Claim 6 *Let $f(.)$ be a posterior density with a normal prior and log-concave likelihood function. Assume that the precision is given by $(1/\sigma^2) + P(\beta)$ where $P(\beta)$ is bounded above by P^*_{UB}. Let $q()$ be a normal density with precision $(1/\sigma^2) + P^*_{UB}$ and assume that there exists $r > 0$ such that*

$$q(\bar{x}_q - r) = f(\bar{x}_q - r)$$

and

$$q(\bar{x}_q + r) = f(\bar{x}_q + r).$$

then

$$\int (x - \bar{x}_f)^2 f(x)dx \geq \int (x - \bar{x}_q)^2 q(x)dx. \tag{18}$$

Proof We note that the function $log(q(x)) - log(f(x))$ is a concave function (in fact, the second order derivative is given by $P(\beta) - P^*_{UB}$). The theorem then follows from a symmetric argument to that leveraged to establish Claim 4. □

Our next claim shows the the required $r > 0$ indeed do exist when the bound for the precision is strict. We outline the modificiation to the arguments in the supplemental article "Details Related to Variance Inequalites.pdf" needed in order to establish this results in much the same way we established Claim 5.

Claim 7 *Let $f(.)$ be a posterior density with a normal prior and log-concave likelihood function. Assume that the precision is given by $(1/\sigma^2) + P(\beta)$ where $P(\beta)$*

is strictly bounded above by P^*_{UB} *(i.e,*$P^*_{UB} - P(\beta) > 0$ *for all* β. *Then there exists a normal density* $q^*()$ *with variance equal to* $(1/\sigma^2) + P^*_{UB}$ *and a constant* $r > 0$ *such that*

$$q^*(\bar{x}_q - r) = f(\bar{x}_q - r)$$

and

$$q^*(\bar{x}_q + r) = f(\bar{x}_q + r).$$

Proof The proof of this follows nearly identical arguments to those required to establish Claim 5. Instead of the set M in that proof, we now define a set

$$M1 := \left\{ \mu \in \mathbf{R} \,\middle|\, \exists r1, r2 \in \mathbf{R}_{++} : \begin{array}{l} (i)\pi(\mu|y) < q^\mu(\mu) \\ (ii)\pi(\mu - r1|y) = q^\mu(\mu - r1) \\ (iii)\pi(\mu + r2|y) = q^\mu(\mu + r2) \end{array} \right\}$$

where $q^\mu(.)$ is a normal density with mean μ and precision given by $(1/\sigma^2) + P^*_{UB}$. We also define $g^\mu(\beta) := log(q^\mu(\beta)) - log(\pi(\beta|y))$. The same arguments as applied to M in the earlier proof can now be established using the set $M1$ and the parameterized densities $q^\mu(.)$. The result then follows. □

Proof of Theorem 2 Claims 6 and 7 implies an analogous result to that in Claim 1. Applying the same set of arguments as in the first part of this section, analogous results to there can be established. These in turn imply Theorem 2 through the same type of argument as for theorem 1. □

Appendix B: Eigenvalue Inequalities

We first note a fact used in the proof of the below. This is a straightforward applications of Weyl [12] eigenvalue inequality. For details see e.g., Fisk [13].

Fact 2 *If* B *is a positive definite matrix, then* $\lambda(A)_i < \lambda(A + B)_i$.

Proof of Theorem 4 It is straightforward to verify that all eigenvalues for BAB, A, and B are non-negative. Moreover, we have that $\lambda_i(BAB) = (\sigma_i(A^{1/2}B))^2$, $\lambda_i(A) = (\sigma_i(A^{1/2}))^2$ and $\lambda_i(B) = (\sigma_i(B^{1/2}))^2$ (where σ_i is used to denote the ith singular value). From Theorem 9 in Merikoski and Kumar [14] and the present assumptions, we have $\sigma_i(A^{1/2}B) \leq \sigma_i(A^{1/2})\sigma_1(B) \leq \sigma_i(A^{1/2})$. Squaring both sides now yields the desired inequality for the eigenvalues. □

Proof of Theorem 5 We first note that $(A + B)^{-1}A$ Hermitian implies that $T = [(A+B)^{-1}A]^2$ and that hence the eigenvalues for T are the squares of the eigenvalues for $(A+B)^{-1}A$. It folllows (see e.g., Theorem 7 in Merikoski and Kumar [14]) that $\lambda_i((A + B)^{-1}A) \geq \lambda_i(A)\lambda_m(A + B) >= 0$ which in turn implies that $\lambda_i(T) \geq 0$. Now, consider the matrix $I - (A+B)^{-1}A = (A+B)^{-1}B$ (which also is Hermitian).

Since B is positive definite, it follows that $\lambda_i((A+B)^{-1}B) \geq \lambda_i(B)\lambda_m(A+B) > 0$. From the Weyl [12] eigenvalue inequality, it now follows that

$$\begin{aligned} 1 &= \lambda_i(I) \\ &= \lambda_i([(A+B)^{-1}A] + [(A+B)^{-1}B]) \\ &= \lambda_i([(A+B)^{-1}A]) + \lambda_m([(A+B)^{-1}B]). \end{aligned}$$

Rearranging terms, we have $\lambda_i([(A+B)^{-1}A]) = 1 - \lambda_m([(A+B)^{-1}B]) < 1$. It then in turns follows that $\lambda_i(T) < 1$. □

Appendix C: Proofs Related to Correlation Structures

Proof of Lemma 6

Proof of Lemma 6 We first note that

$$\begin{aligned} |D_{\mathcal{P}_{AB}}(\Sigma)| &= |I - \Sigma_{AA}^{-1}\Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA}| \\ &= |\Sigma_{AA}^{-1}||\Sigma_{AA} - \Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA}| \\ &> 0. \end{aligned}$$

(where the inequality follows since both individual determinants are known to be positive). We also have,

$$\begin{aligned} |D_{\mathcal{P}_{AB}}(\Sigma)| &= |I - \Sigma_{AA}^{-1}\Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA}| \\ &= |\Sigma_{AA}^{-1}||\Sigma_{AA} - \Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA}| \\ &\leq |\Sigma_{AA}^{-1}|[|\Sigma_{AA} - \Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA}| + |\Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA}|] \\ &\leq |\Sigma_{AA}^{-1}||[\Sigma_{AA} - \Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA}] + \Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA}| \\ &= |\Sigma_{AA}^{-1}||\Sigma_{AA}| \\ &= 1. \end{aligned}$$

where the first inequality follows from properties of positive semi-definite matrices and the second from the Minkowski determinant theorem. □

Proof of Lemma 7

Proof of Lemma 7 Define a function $h_3(.|\mu^*)$ by

$$h_3(\mu^k|\mu^*) = \log[Q(\mu^*, \mu^k)] - log[S(\mu^k|P_{LB}^K)].$$

Because P_{LB}^K is a lower bound matrix for the precision matrix associated with $Q(\mu^*, .)$, the function $h_3(.|\mu^*)$ is a concave function for which the supergradient $c(\mu^k)$ vanishes (i.e., equals zero) at μ^{**}. Suppose

$$\log[Q(\mu^*, \mu^{**})] < log[S(\mu^{**}|P_{LB}^K)].$$

Then it follows from properties of concave functions that for every μ^k,

$$\begin{aligned}\log[Q(\mu^*,\mu^k)] - log[S(\mu^k|P_{LB}^K)] &\leq \log[Q(\mu^*,\mu^{**})] - log[S(\mu^{**}|P_{LB}^K)] \\ &\quad + c(\mu^{**})^T(\mu^k-\mu^{**}) \\ &= \log[Q(\mu^*,\mu^{**})] - log[S(\mu^{**}|P_{LB}^K)] < 0.\end{aligned}$$

But then $Q(\mu^*,.)$ and $S(.|P_{LB}^K)$ can't both be proper densities, a contradiction. Hence

$$\log[Q(\mu^*,\mu^{**})] \geq log[S(\mu^{**}|P_{LB}^K)].$$

Since $S(.|P_{LB}^K)$ and $S(.|P_{UB}^K)$ are both Multivariate normal densities obtaining their maximums at μ^{**}, it then follows from the formulas for the two densities at μ^{**} that

$$\log[Q(\mu^*,\mu^{**})] \geq log[S(\mu^{**}|P_{LB}^K)] = 0.5\log[\frac{|P_{LB}^K|}{|P_{UB}^K|}] + log[S(\mu^{**}|P_{UB}^K)].$$

which is equivalent to $\log[\frac{Q(\mu^*,\mu^{**})}{S(\mu^{**}|P_{UB}^K)}] \geq 0.5\log[\frac{|P_{LB}^K|}{|P_{UB}^K|}]$. Define a function $h_1(.|\mu^*)$ by $h_1(\mu^k|\mu^*) = \log(\frac{Q_{\mathcal{M}}(\mu^{k-1},\mu^k)}{S(\mu^k|P_{UB}^K)})$. Because P_{UB}^K is an upper bound matrix for the precision matrix associated with $Q(\mu^*,.)$, the function $h_1(.|\mu^*)$ is a convex function for which the subgradient $\tilde{c}(\mu^k)$ vanishes at μ^{**}. Hence μ^{**} minimizes the function $h_1(.|\mu^*)$. We then have

$$\begin{aligned}\log[\tfrac{Q(\mu^*,\mu^k)}{S(\mu^k|P_{UB}^K)}] &\geq \log[\tfrac{Q(\mu^*,\mu^{**})}{S(\mu^{**}|P_{UB}^K)}] \\ &\geq 0.5\log[\tfrac{|P_{LB}^K|}{|P_{UB}^K|}]\end{aligned}$$

for every μ^k. □

Proof of Lemma 8

Lemma 10 *Suppose the log-concavity assumption 1 and the full rank assumption 3 are both satisfied. Let* $\tilde{R} = tX^TP(P+P_{LB}^D)^{-1}PX$ *for some positive constant* t *and let* A *and* $B(\mu^{k-1})$ *be the Hermitian matrices satisfying*

$$A^{1/2} = (X^TP(P+P_{LB}^D)^{-1}PX)^{1/2}(P_0+X^TPX)^{-1}(X^TP(P+P_{LB}^D)^{-1}PX)^{1/2},$$

$$\begin{aligned}B(\mu^{k-1}) &= (X^TP(P+P_{LB}^D)^{-1}PX)^{-1/2}X^TP \\ &\quad Var_{\mu^*}^{\beta}(\mu^{(k-1)})PX(X^TP(P+P_{LB}^D)^{-1}PX)^{-1/2}\end{aligned}$$

and

$$\begin{aligned}B_2(\mu^{k-1}) &= (X^TP(P+P_{LB}^D)^{-1}PX)^{-1/2}X^TP \\ &\quad ((P+P_{LB}^D)^{-1} - Var_{\mu^*}^{\beta}(\mu^{(k-1)}))PX(X^TP(P+P_{LB}^D)^{-1}PX)^{-1/2}\end{aligned}$$

Then

1 $P^*_{\tilde{R}} = A^{1/2}$
2 $\tilde{P}^*_{\tilde{R}}(Var^{\beta}_{\mu^*}(\mu^{(k-1)})) = A^{1/2}B$
3 *A is positive definite*
4 *$B(\mu^{k-1})$ is positive definite and $B_2(\mu^{k-1})$ is positive semi-definite.*
5 *The maximal eigenvalue for $B(\mu^{k-1})$ is less than or equal to 1.*

Proof To see that (a) holds, note that under the present assumptions $\tilde{R}$ is Hermitian positive definite and hence $\tilde{R}^{-1/2}$ is well defined and also Hermitian positive definite.We then have

$$\begin{aligned}
\mathcal{P}_{AB}(P^*_{\tilde{R}}) &= (P^*_{\tilde{R},AA})^{-1}P^*_{\tilde{R},AB}(P^*_{\tilde{R},BB})^{-1}P^*_{\tilde{R},BA} \\
&= (\tilde{R}^{-1/2}(P_0 + X^TPX)\tilde{R}^{-1/2})^{-1}\tilde{R}^{-1/2}X^TP(P+P^D_{LB})^{-1}PX\tilde{R}^{-1/2} \\
&= \tilde{R}^{1/2}(P_0 + X^TPX)^{-1}X^TP(P+P^D_{LB})^{-1}PX\tilde{R}^{-1/2} \\
&= (tX^TP(P+P^D_{LB})^{-1}PX)^{1/2}(P_0 + X^TPX)^{-1}X^TP \\
&\quad (P+P^D_{LB})^{-1}PX(tX^TP(P+P^D_{LB})^{-1}PX)^{-1/2} \\
&= (X^TP(P+P^D_{LB})^{-1}PX)^{1/2}(P_0 + X^TPX)^{-1} \\
&\quad (X^TP(P+P^D_{LB})^{-1}PX)^{1/2} \\
&= A^{1/2}.
\end{aligned}$$

To see that (b) holds, note that we have

$$\begin{aligned}
\mathcal{P}_{AB}(\tilde{P}^*_{\tilde{R}}(Var^{\beta}_{\mu^*}(\mu^{(k-1)}))) &= (\tilde{R}^{-1/2}(P_0 + X^TPX)\tilde{R}^{-1/2})^{-1} \\
&\quad \tilde{R}^{-1/2}X^TPVar^{\beta}_{\mu^*}(\mu^{(k-1)})PX\tilde{R}^{-1/2} \\
&= \tilde{R}^{1/2}(P_0 + X^TPX)^{-1}\tilde{R}^{1/2} \\
&\quad \tilde{R}^{-1/2}X^TPVar^{\beta}_{\mu^*}(\mu^{(k-1)})PX\tilde{R}^{-1/2} \\
&= (tX^TP(P+P^D_{LB})^{-1}PX)^{1/2}(P_0 + X^TPX)^{-1} \\
&\quad (tX^TP(P+P^D_{LB})^{-1}PX)^{1/2} \\
&\quad (tX^TP(P+P^D_{LB})^{-1}PX)^{-1/2} \\
&\quad X^TPVar^{\beta}_{\mu^*}(\mu^{(k-1)})PX \\
&\quad (tX^TP(P+P^D_{LB})^{-1}PX)^{-1/2} \\
&= (X^TP(P+P^D_{LB})^{-1}PX)^{1/2}(P_0 + X^TPX)^{-1} \\
&\quad (X^TP(P+P^D_{LB})^{-1}PX)^{1/2} \\
&\quad (X^TP(P+P^D_{LB})^{-1}PX)^{-1/2} \\
&\quad X^TPVar^{\beta}_{\mu^*}(\mu^{(k-1)})PX \\
&\quad (X^TP(P+P^D_{LB})^{-1}PX)^{-1/2} \\
&= A^{1/2}B.
\end{aligned}$$

To see that property (c) holds, let $C = (P_0 + X^TPX)^{-1}$ and $D = X^TP(P + P^D_{LB})^{-1}PX$. Under the present assumptions, both are positive definite matrices and $A^{1/2} = C^{1/2}DC^{1/2}$. We have for every eigenvalue associated with $A^{1/2}$ that

$$\lambda_i(A^{1/2}) = \lambda_i(C^{1/2}DC^{1/2}) = \lambda_i(DC) \geq \lambda_i(D)\lambda_{m_2}(C) > 0$$

which implies that A is positive definite.

Property (d) follows using the same type of arguments as for property (c) since $(X^TP(P + P^D_{LB})^{-1}PX)^{-1}$ is positive definite, $X^TPVar^{\beta}_{\mu^*}(\mu^{(k-1)})PX$ is positive definite, and

$$X^T P((P + P_{LB}^D)^{-1} - Var_{\mu^*}^{\beta}(\mu^{(k-1)}))PX$$

is positive semi-definite (the latter due to our variance inequality).

To see that property (e) holds, first note that $B(\mu^{k-1}) + B_2(\mu^{k-1}) = I$. Since $B(\mu^{k-1})$ and $B_2(\mu^{k-1})$ are both positive semi-definite, they both have non-negative eigenvalues while the eigenvalues of I all equal 1. Weyl's eigenvalue inequality then implies:

$$\begin{array}{rcl} \lambda_i(B(\mu^{k-1})) + \lambda_{m_1}(B_2(\mu^{k-1})) & \leq & 1 \\ & \Downarrow & \\ \lambda_i(B(\mu^{k-1})) & \leq & 1. \end{array}$$

□

Proof of Lemma 8 Let A and B be as in lemma 10. Lemma 10 then implies that

$$(\mathcal{P}_{AB}(P_{\tilde{R}}^*))^T(\mathcal{P}_{AB}(P_{\tilde{R}}^*)) = A$$

and

$$(\mathcal{P}_{AB}(\tilde{P}_{\tilde{R}}^*(Var_{\mu^*}^{\beta}(\mu^{(k-1)}))))^T(\mathcal{P}_{AB}(\tilde{P}_{\tilde{R}}^*(Var_{\mu^*}^{\beta}(\mu^{(k-1)})))) \quad = \quad B(\mu^{k-1})AB(\mu^{k-1}).$$

Using the properties in lemma 10 it is straightforward to verify that all the assumptions of lemma 4 are satisfied and hence for every i, we have $\lambda_i(BAB) \leq \lambda_i(A)$ establishing property (a). Taking the square root of the right and left sides of the above for the maximal eigenvalues now yields property (b). □

Proof of Lemma 9

Proof of Lemma 9 To see that (a) holds, note that under the present assumptions $\tilde{R}$ is Hermitian positive definite and hence $\tilde{R}^{-1/2}$ is well defined and also Hermitian positive definite. We then have

$$\begin{array}{rcl} \mathcal{P}_{AB}(P_{\tilde{R}}^*) & = & (P_{\tilde{R},AA}^*)^{-1}P_{\tilde{R},AB}^*(P_{\tilde{R},BB}^*)^{-1}P_{\tilde{R},BA}^* \\ & = & (\tilde{R}^{-1/2}(P_0 + X^T PX)\tilde{R}^{-1/2})^{-1}\tilde{R}^{-1/2}X^T P(P + P_{LB}^D)^{-1}PX\tilde{R}^{-1/2} \\ & = & \tilde{R}^{1/2}(P_0 + X^T PX)^{-1}X^T P(P + P_{LB}^D)^{-1}PX\tilde{R}^{-1/2} \\ & = & (tX^T P(P + P_{LB}^D)^{-1}PX)^{1/2}(P_0 + X^T PX)^{-1}X^T P \\ & & (P + P_{LB}^D)^{-1}PX(tX^T P(P + P_{LB}^D)^{-1}PX)^{-1/2} \\ & = & (X^T P(P + P_{LB}^D)^{-1}PX)^{1/2}(P_0 + X^T PX)^{-1} \\ & & (X^T P(P + P_{LB}^D)^{-1}PX)^{1/2} \end{array}$$

and hence $\mathcal{P}_{AB}(P_{\tilde{R}}^*)$ is a Hermitian positive semi-definite matrix. As

$$(X^T P(P + P_{LB}^D)^{-1}PX)^{1/2}$$

and $(P_0+X^TPX)^{-1}$ are both positive definite, we actually have (see Merikoski and Kumar [14])

$$\begin{aligned}
\lambda_i(\mathcal{P}_{AB}(P^*_{\tilde{R}})) &= \lambda_i((X^TP(P+P^D_{LB})^{-1}PX)^{1/2}(P_0+X^TPX)^{-1} \\
&\quad (X^TP(P+P^D_{LB})^{-1}PX)^{1/2}) \\
&= \lambda_i((P_0+X^TPX)^{-1}(X^TP(P+P^D_{LB})^{-1}PX)^{1/2}) \\
&\geq \lambda_m((P_0+X^TPX)^{-1}) \\
&\quad \lambda_i((P_0+X^TPX)^{-1}(X^TP(P+P^D_{LB})^{-1}PX)^{1/2}) \\
&> 0
\end{aligned}$$

and hence $\mathcal{P}_{AB}(P^*_{\tilde{R}})$ is a Hermitian positive definite matrix as required.

It now follows that $[\mathcal{P}_{AB}(P^*_{\tilde{R}})]^T\mathcal{P}_{AB}(P^*_{\tilde{R}}) = [\mathcal{P}_{AB}(P^*_{\tilde{R}})]^2$ and hence the maximal eigenvalue of the matrix $[\mathcal{P}_{AB}(P^*_{\tilde{R}})]^T\mathcal{P}_{AB}(P^*_{\tilde{R}})$ is equal to $[\lambda_1(\mathcal{P}_{AB}(P^*_{\tilde{R}}))]^2$, which in turn implies (b). To see property (c), let $\Sigma = (P^*_{\tilde{R}})^{-1}$. Using properties of partioned matrices (see e.g., p. 824 of Greene [15]) , we can write

$$\begin{array}{lrcl}
(i) & P^*_{\tilde{R},AA} &=& \Sigma^{-1}_{AA}(I+\Sigma_{AB}(\Sigma_{BB}-\Sigma_{BA}\Sigma^{-1}_{AA}\Sigma_{AB})^{-1} \\
& & & \Sigma_{BA}\Sigma^{-1}_{AA}) \\
(ii) & P^*_{\tilde{R},AB} &=& -\Sigma^{-1}_{AA}\Sigma_{AB}(\Sigma_{BB}-\Sigma_{BA}\Sigma^{-1}_{AA}\Sigma_{AB})^{-1} \\
(iii) & P^*_{\tilde{R},BA} &=& -(\Sigma_{BB}-\Sigma_{BA}\Sigma^{-1}_{AA}\Sigma_{AB})^{-1}\Sigma_{BA}\Sigma^{-1}_{AA} \\
(iv) & P^*_{\tilde{R},BB} &=& (\Sigma_{BB}-\Sigma_{BA}\Sigma^{-1}_{AA}\Sigma_{AB})^{-1} \\
(v) & P^*_{\tilde{R},AB}(P^*_{\tilde{R},BB})^{-1}P^*_{\tilde{R},BA} &=& \Sigma^{-1}_{AA}\Sigma_{AB}(\Sigma_{BB}-\Sigma_{BA}\Sigma^{-1}_{AA}\Sigma_{AB})^{-1}\Sigma_{BA}\Sigma^{-1}_{AA} \\
(vi) & P^*_{\tilde{R},AA} &=& P^*_{\tilde{R},AB}(P^*_{\tilde{R},BB})^{-1}P^*_{\tilde{R},BA}+\Sigma^{-1}_{AA}
\end{array}$$

Now, letting $\tilde{B} = P^*_{\tilde{R},AB}(P^*_{\tilde{R},BB})^{-1}P^*_{\tilde{R},BA}$ and $\tilde{A} = \Sigma^{-1}_{AA}$ (which is positive definite), we have $P^*_{\tilde{R},AA} = \tilde{A}+\tilde{B}$. Hence it follows that

$$\mathcal{P}_{AB}(P^*_{\tilde{R}}) = (\tilde{A}+\tilde{B})^{-1}\tilde{B}.$$

Since $\mathcal{P}_{AB}(P^*_{\tilde{R}})$ is a Hermitian matrix, it follows from the definition of the spectral norm and Theorem 5 that $0 < \left\|\mathcal{P}_{AB}(P^*_{\tilde{R}})\right\|_{SP} < 1$.

□

Competing interests
The author declares that he have no competing interests.

References

1. S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:6:721–741, 1984.
2. A. Gelfand and A. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:398–409, 1990.
3. H. Raiffa and R Schlaifer. *Applied Statistical Decision Theory*. The M. I. T. Press, Cambridge, MA, 1968.
4. J.S. Rosenthal. Minorization conditions and convergence rates for markov chain monte carlo. *Journal of the American Statistical Association*, 90:558–566. Correction, p. 1136, 1995.
5. K Nygren and L. Nygren. Likelihood subgradient densities. *Journal of the American Statistical Association*, 101:1144–1156, 2006.
6. G.O. Roberts and R.L. Tweedie. Bounds on regeneration times and convergence rates for markov chains. *Stochastic Process. Appl*, 80:211–229, 1999.
7. G.O Roberts. Rates of convergence for gibbs sampling for variance component models. *The Annals of Statistics*, 23:740–761, 1995.

8. J.P. Hobert and C.J. Geyer. Geometric ergodicity of gibbs and block gibbs samplers for a hierarchical random effects model. *Journal of Multivariate Analysis*, 67(2):414–430, 1998.
9. G.L. Jones and J.P. Hobert. Sufficient burn-in for gibbs samplers for a hierarchical random effects model. *The Annals of Statistics*, 32(2):784–817, 2004.
10. G. O. Roberts and S. K. Sahu. Approximate pre-determined convergence properties of the gibbs sampler. *Journal of Computational and Graphical Statistics*, 10:216–229, 2001.
11. A Prekopa. On logarithmic concave functions and measures. *Acta Scientarium Mathematicarum*, 34:335–343, 1973.
12. H Weyl. Das asymptotische verteilungsgesetz der eigenwerte linearer partieller differentialgleichungen. *Math. Ann.*,, 71:441–479, 1912.
13. S. Fisk. A note of weyl's inequality. *American Mathematical Monthly*, 104(3):257–258, 1997.
14. J Merikoski and R. Kumar. Inequalities for matrix spreads and products. *Applied Mathematical E-Notes*, 4: 150–159, 2004.
15. W.H Greene. *Econometric Analysis*. Prentice Hall, Upper Saddle River, NJ, 07458, 5th edition, 2002. ISBN 0-13-066189-9.

Additional Files

Details Related to Variance Inequalites.pdf
This supplement contains detailed arguments related to the establishment of Claim 5.

Equivalence of Densities and an Expression for the Precision of the Marginal Transition Kernel.pdf
This supplement contains details of the arguments demonstrating the equivalence of densities in Lemmas 1 through 3 and the expression for the marginal transition kernel in Lemma 4.