THE MODES OF POSTERIOR DISTRIBUTIONS FOR MIXED LINEAR MODELS

ALICIA L. CARRIQUIRY
IOWA STATE UNIVERSITY, U. S. A.
PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE, CHILE
and
WOLFGANG KLIEMANN
IOWA STATE UNIVERSITY, U. S. A.
PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE, CHILE

Abstract

Mixed linear models, also known as two-level hierarchical models, are commonly used in many applications. In this paper, we consider the marginal distribution that arises within a Bayesian framework, when the components of variance are integrated out of the joint posterior distribution. We provide analytical tools for describing the surface of the distribution of interest. The main theorem and its proof show how to determine the number of local maxima, and their approximate location and relative size. This information can be used by practitioners to assess the performance of Laplace-type integral approximations, to compute possibly disconnected highest posterior density regions, and to custom-design numerical algorithms.

1. INTRODUCTION

Mixed linear models (sometimes known as two-level hierarchical models or as components of variance models) are used in a wide variety of fields, including genetics, econometrics, biomedical and environmental applications, and many others (e.g., Fuller and Harter, 1987; Harvey, 1989; Smith et al., 1990; Harville and Carriquiry, 1992; Wang et al., 1994; Hobert and Casella, 1996; Van Dyk, 1999). The two-level linear hierarchical model can be formulated as follows:

$$(1.1) \quad \begin{aligned} \mathbf{y}|\boldsymbol{\beta},\mathbf{s},\tau_e &\sim \quad \mathrm{N}(\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\mathbf{s},\mathbf{I}\tau_e^{-1}) \\ \mathbf{s}|\mathbf{A},\tau_s &\sim \quad \mathrm{N}(\mathbf{0},\mathbf{A}\tau_s^{-1}), \end{aligned}$$

where $\mathbf{y} \in I\!\!R^n$ is the observed data vector, $\boldsymbol{\beta} \in I\!\!R^p$ is a vector of unknown fixed (or covariate) effects, $\mathbf{s} \in I\!\!R^q$ is a vector of random effects, and $\mathbf{X} \in I\!\!R^{n \cdot p}$ and $\mathbf{Z} \in I\!\!R^{n \cdot q}$ are the model matrices corresponding to the fixed and random effects, respectively. Here, the matrix $\mathbf{A}$ is assumed to be known; in animal genetics applications, for example, the elements of $\mathbf{A}$ reflect the familial relations of animals in the dataset. Practitioners are often interested in estimating $\boldsymbol{\beta}$, $\mathbf{s}$, or a combination of both. The variance components $(\tau_s^{-1}, \ \tau_e^{-1})$ are considered nuisance parameters that must, however, be accounted for in estimation.

The recent introduction of Markov chain Monte Carlo (MCMC) methods (e.g., Gelfand and Smith, 1990; Besag and Green, 1993) has facilitated estimation in mixed linear models from a Bayesian perspective. In the Bayesian framework, prior distributions are chosen for the vector $\boldsymbol{\beta}$ and the scalar-valued variance components, and the marginal posterior distributions of all parameters in the model are sought. These marginal posterior distributions can be approximated simultaneously, by generating Markov chains with stationary distributions equal to the marginal distributions of interest. Except in trivial cases, analytical derivation of all marginal distributions of interest is not possible.

In some applications, however, the dimensions p and q of the vectors $\boldsymbol{\beta}$ and $\mathbf{s}$, respectively, can be large enough as to make the use of MCMC methods impractical. If practitioners are interested only in the first and second order moments of the marginal posterior distributions of parameters in the model, then Laplace-type approximations can be used very effectively. Several approximations to the posterior mean of a parameter vector have

been proposed. Lindley (1980) was first to suggest the use of asymptotic expansions for ratios of integrals to approximate posterior moments. He derived Taylor expansions around the maximum likelihood (ML) estimate of the parameter of interest, as far as terms of order n^{-1} . Tierney et al. (1989) and Kass and Steffey (1989) proposed the application of Laplace's method to expand the ratios of integrals, with modifications which enable the evaluation of nonpositive functions. Morris (1988) proposed a more general version of Laplace-type approximations, where the kernel is a distribution chosen from the Pearson family, which includes as a special case the normal family of distributions. While authors suggested that expansions be carried out around the MLE of the parameter of interest, they mentioned that more satisfactory results may be obtained by expanding, instead, around the posterior mode. Laplace's method (and its variations) work only in the case of unimodal surfaces, and fails when the surface of interest has more than one mode. Robert and Casella (1999, Chapter 3) stress once again the importance of unimodality of the underlying distribution.

In this paper, we address the problem of analytically finding the mode(s) of the posterior distribution $p(\boldsymbol{\beta}, \mathbf{s}|\mathbf{y})$ that arises from model (1) after integration of the joint posterior distribution with respect to the variance components (τ_s, τ_e). We consider a rather general class of prior distributions for the variance components, but concentrate on a family of distributions that guarantees integrability of the joint posterior distribution (see, e.g., Hobart and Casella, 1996). Specifically, we investigate the surface of the posterior distribution of interest, and attempt to answer the following questions:

- How many (local) maxima does the distribution have?

- If multimodal, what is the relative size of each mode?

- Where are the modes located?

Answering these questions is interesting from an academic viewpoint, given the widespread use of two-level hierarchical models. A more pragmatic motivation for this work was presented above, as we argued that the performance of Laplace approximations depends to a large extent on whether $p(\boldsymbol{\beta}, \mathbf{s}|\mathbf{y})$ is unimodal or multimodal. But even beyond the issue of behavior of integral approximations, in some applications posterior modes (or maximum a posteriori (MAP)) are used as point estimates of location parameters. This is because maxima are typically easier to compute than expectations. In these cases, the existence of more than one mode requires

careful consideration, in particular when no single mode is associated to a significantly larger mass than the others. Similarly, identification of highest posterior density regions for $(\boldsymbol{\beta}, \mathbf{s})$ requires knowledge of the number of modes of $p(\boldsymbol{\beta}, \mathbf{s}|\mathbf{y})$, as in the multimodal case the region may well be disconnected (Berger, 1985).

This paper is organized as follows. Additional notation, as well as prior and posterior distributions of model parameters, are presented in Section 2. In the same section, we discuss the relation between our work and that of Bauwens, Drèze, Lubrano, Richard, and Tompa (Drèze, 1977; Richard and Tompa, 1980; Drèze and Richard, 1983; Lubrano et al., 1984; Bauwens and Richard, 1985), who have proposed numerical algorithms to draw values from a distribution very similar to $p(\boldsymbol{\beta}, \mathbf{s}|\mathbf{y})$ that arises in a specific econometric application. Section 3 presents the main theorem (its proof is given in the Appendix) and several corollaries and remarks to the theorem. Some explicit criteria for determining uni- or multimodality of the posterior distribution $p(\boldsymbol{\beta}, \mathbf{s}|\mathbf{y})$ are also given in this section. Section 4 contains an illustrative example. Finally, some discussion and conclusions are presented in Section 5.

2. PRIOR AND POSTERIOR DISTRIBUTIONS

2.1. The Mixed Linear Model

Consider the hierarchical model (1), and write it as a linear mixed model (e.g., Goldberger, 1962; Henderson, 1963):

(2.1) $$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{s} + \mathbf{e}.$$

Let $\mathbf{s} \sim \mathrm{N}\left(\mathbf{0}, \tau_s^{-1}\mathbf{I}\right), \tau_s \in I\!R, \tau_s > 0$, and $\mathbf{e} \in I\!R^n \sim \mathrm{N}\left(\mathbf{0}, \tau_e^{-1}\mathbf{I}\right), \tau_e \in I\!R$, $\tau_e > 0$. Further, let $\mathrm{Cov}\left(\mathbf{e}, \mathbf{s}\right) = \mathbf{0}, rank\left(\mathbf{X}\right) = p < n$, and $rank\left(\mathbf{Z}\right) \leq q < n$

The first assumption seems unnecessarily restrictive. More general models, however, can be brought into this form by a linear transformation. Let the random effects satisfy $\mathbf{s}^* \sim \mathrm{N}\left(\mathbf{0}, \tau_s^{-1}\mathbf{A}\right)$, where $\mathbf{A} \in I\!R^{q \cdot q}$ is a known, positive semi-definite matrix and denote by $\mathbf{A}^{1/2}$ its (unique) positive semi-definite square root. Then, $\mathbf{s} = \mathbf{A}^{1/2}\mathbf{s}^*$ has a distribution that is $\mathrm{N}\left(\mathbf{0}, \tau_s^{-1}\mathbf{I}\right)$ for the nonzero components, and the results below apply to $\mathbf{s}^* = \mathbf{A}^{1/2}\mathbf{s}$. Furthermore, if $\ddot{\mathbf{X}}$ is a model matrix with rank $\left(\ddot{\mathbf{X}}\right) < p$, then the model can be reparameterized to obtain a matrix $\mathbf{X}$ with full (column) rank and a new

parameter vector $\widehat{\boldsymbol{\beta}}$. The new vector $\widehat{\boldsymbol{\beta}}$ contains linear combinations of the components of $\boldsymbol{\beta}$, and only these linear combinations are estimable (e.g., Searle, 1981).

Using the notation $\delta = \tau_s/\tau_e$, we obtain $\mathbf{y} \sim \mathrm{N}\left(\mathbf{X}\boldsymbol{\beta}, \tau_e^{-1}\left(\mathbf{I} + \delta^{-1}\mathbf{Z}\mathbf{Z}'\right)\right)$, where $'$ denotes the transpose of a matrix. The parameter space is then defined as

$$(2.2) \qquad (\boldsymbol{\beta}, \delta, \tau_e) \in \Omega = \left\{(\boldsymbol{\beta}, \tau_e, \delta) \in I\!\!R \times (0, \infty) \times (0, \infty)\right\}.$$

Let $\boldsymbol{\varphi}(\boldsymbol{\beta}, \mathbf{s}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{s}$. The likelihood function of $\mathbf{y}$, conditional on $\boldsymbol{\beta}, \mathbf{s}$, and τ_e is then given by

$$(2.3) \qquad l\left(\mathbf{y}|\boldsymbol{\beta}, \mathbf{s}, \tau_e\right) = (2\pi)^{-n/2}\, \tau_e^{n/2} \exp\left[-\frac{1}{2}\tau_e\left(\mathbf{y} - \boldsymbol{\varphi}\right)'\left(\mathbf{y} - \boldsymbol{\varphi}\right)\right].$$

2.2. Prior Distributions

Within the Bayesian framework, prior distributions must be chosen for the unknown parameters in model (2). In this case, prior distributions for $\boldsymbol{\beta}$ and for the variance components for $\mathbf{s}$ and $\mathbf{e}$ (or for δ and τ_e) must be chosen. Assume that a priori, $\boldsymbol{\beta}$ has a multivariate normal distribution,

$$(2.4) \qquad\qquad \boldsymbol{\beta} \sim MVN\left(\boldsymbol{\alpha}, \epsilon\mathbf{I}\right),$$

with $\boldsymbol{\alpha} \in I\!\!R^p$ and $\epsilon > 0$ known, and denote its density by $\pi_1\left(\boldsymbol{\beta}\right)$. For the vector (δ, τ_e) consider the following family of prior distributions

$$(2.5) \qquad \pi_2\left(\delta, \tau_e\right) = G_1\left(\delta\right)\left(\tau_e\right)^{G_2(\delta)} \exp\left[-\frac{1}{2}\tau_e G_3\left(\delta\right)\right],$$

where $G_1\left(\delta\right)$, $G_2\left(\delta\right)$, and $G_3\left(\delta\right)$ are functions of δ which satisfy the conditions $G_1\left(\delta\right) > 0$, $G_2\left(\delta\right) < \frac{1}{2}\left(n - p - 2\right)$, and $G_3\left(\delta\right) \geq 0$. This family includes commonly used prior distributions for scale parameters, such as the uniform noninformative prior (Box and Tiao, 1973, Section 1.2.1), Jeffreys' (1961) priors, and the conjugate (Gamma) family (Raiffa and Schlaiffer, 1971).

2.3. Posterior Distribution

The joint posterior density of $(\boldsymbol{\beta}, \mathbf{s}, \delta, \tau_e)$ given $\mathbf{y}$ can now be written in the form

$$
\begin{aligned}
p\left(\boldsymbol{\beta},\ \mathbf{s},\ \delta,\ \tau_e | \mathbf{y}\right) \ &=\ k_1 l\left(\mathbf{y},\ \boldsymbol{\beta},\ \mathbf{s},\ \tau_e\right) \pi_1\left(\boldsymbol{\beta}\right) \pi_2\left(\delta,\ \tau_e\right) \\
&=\ k_2 \tau_e^{n'(\delta)} \delta^{1/2} G_1\left(\delta\right) \epsilon^{-p/2} \\
&\quad \exp\left\{-\tfrac{1}{2}\tau_e \left[(\mathbf{y}-\boldsymbol{\varphi})'(\mathbf{y}-\boldsymbol{\varphi}) + \delta \mathbf{s}'\mathbf{s} + 2G_3\left(\delta\right)\right]\right. \\
&\quad \left. -(2\epsilon)^{-1}(\boldsymbol{\beta}-\boldsymbol{\alpha})'(\boldsymbol{\beta}-\boldsymbol{\alpha})\right\},
\end{aligned}
$$

(2.6)

where k_1 is the normalizing constant, $k_2 = k_1 (2\pi)^{-1/2(n+q+p)}$, and $n'(\delta) = .\tfrac{1}{2}(n+q+2G_2(\delta)) + 1$. Consider the case where no prior information on $\boldsymbol{\beta}$ is available. That is, consider the limiting case where $\epsilon \to \infty$, which yields (with the appropriate scaling factor $\epsilon^{p/2}$)

$$
\lim_{\epsilon \to \infty} \epsilon^{p/2} p\left(\boldsymbol{\beta}, \mathbf{s}, \delta, \tau_e | \mathbf{y}\right) \propto
$$

(2.7)
$$
k_2 \tau_e^{n'(\delta)} \delta^{q/2} G_1\left(\delta\right) \exp\left\{-\tfrac{1}{2}\tau_e\left[(\mathbf{y}-\boldsymbol{\varphi})'(\mathbf{y}-\boldsymbol{\varphi}) + \delta\mathbf{s}'\mathbf{s} + 2G_3\left(\delta\right)\right]\right\},
$$

where $\propto$ means "proportional to." Upon integrating (8) analytically with respect to τ_e, the marginal posterior density of $\boldsymbol{\beta}$, $\mathbf{s}$, and δ, given $\mathbf{y}$ can be written in the form

$$
\begin{aligned}
p\left(\boldsymbol{\beta},\ \mathbf{s},\ \delta | \mathbf{y}\right) \ =\ & k_2 \delta^{1/2} G_1\left(\delta\right) \times \Gamma\left\{\tfrac{1}{2}\left[n+q+2G_2\left(\delta\right)\right]\right\} \\
& \left\{\tfrac{1}{2}(\mathbf{y}-\boldsymbol{\varphi})'(\mathbf{y}-\boldsymbol{\varphi}) + \delta\mathbf{s}'\mathbf{s} + 2G_3\left(\delta\right)\right\}^{-[n+q+2G_2(\delta)]/2}.
\end{aligned}
$$

(2.8)

Inferences about the fixed and the random effects $(\boldsymbol{\beta}, \mathbf{s})$ should be based on $p(\boldsymbol{\beta}, \mathbf{s}|\mathbf{y})$. This marginal posterior density can be written in closed form (up to a normalizing constant), if the functions $G_1(\delta)$, $G_2(\delta)$, and $G_3(\delta)$ are specified more precisely. Let

(2.9) $\qquad G_1\left(\delta\right) = \dfrac{b_s^{a_s} b_e^{a_e}}{\Gamma\left(a_s\right)\Gamma\left(a_e\right)} \times \delta^{a_s-1},\ b_s,\ b_e > 0,\ a_s,\ a_e > 0,$

(2.10) $\qquad\qquad\qquad G_2\left(\delta\right) = \left(a_s + a_e - 1\right),$

(2.11) $\qquad\qquad\qquad G_3\left(\delta\right) = b_s\delta + b_e.$

These choices for $G_1(\delta)$, $G_2(\delta)$, and $G_3(\delta)$ correspond to the case where a Gamma (a_s, b_s) prior is assumed for τ_s, a Gamma(a_e, b_e) prior is assumed for τ_e, and τ_s and τ_e are independent a priori. We have, therefore, restricted $\pi_2(\delta, \tau_e)$ to any prior distribution obtained by considering a conjugate prior family for the independent precision components τ_s and τ_e.

Let

$$
\begin{aligned}
p_1 &= \frac{1}{2}(n + q + 2a_s + 2a_e), \\
k_3 &= \frac{k_2 b_s^{a_s} b_e^{a_e}}{\Gamma(a_s)\Gamma(a_e)} 2^{p_1} \Gamma(p_1),
\end{aligned}
$$

and

$$
p_2 = \frac{1}{2}(1 + 2a_s - 2).
$$

Then, substituting (10)-(12) into (9) yields

$$(2.12) \quad p(\boldsymbol{\beta}, \mathbf{s}, \delta | \mathbf{y}) = k_3 \delta^{p_2} \left[(\mathbf{y} - \boldsymbol{\varphi})'(\mathbf{y} - \boldsymbol{\varphi}) + 2b_e + \delta (\mathbf{s}'\mathbf{s} + 2b_s) \right]^{-p_1}.$$

Define, further, $q_1 = (\mathbf{y} - \boldsymbol{\varphi})'(\mathbf{y} - \boldsymbol{\varphi}) + 2b_e$ and $q_2 = \mathbf{s}'\mathbf{s} + 2b_s$, and rewrite (13) in the form

$$(2.13) \quad p(\boldsymbol{\beta}, \mathbf{s}, \delta | \mathbf{y}) = k_3 \delta^{p_2} [q_1 + \delta q_2]^{-p_1} = k_3 \delta^{p_2} q_1^{-p_1} \left[1 + \delta \frac{q_2}{q_1} \right]^{-p_1}.$$

By using the appropriate change of variable (Stroud, 1987), (14) can be written in the form of a Beta function, and integration with respect to δ yields

$$
\begin{aligned}
p(\boldsymbol{\beta}, \mathbf{s} | y) &= k_3 q_1^{-p_1} \int_0^1 q_1^{p_2} q_2^{-p_2} x^{p_2} (1-x)^{p_1} q_1 q_2^{-1} (1-x)^{-2} \, dx \\
&= k_3 q_1^{(-p_1 + p_2 + 1)} q_2^{-(p_2 + 1)} \int_0^1 x^{p_2} (1-x)^{(p_1 - p_2 - 2)} \, dx \\
&= k_4 \left[(\mathbf{y} - \boldsymbol{\varphi})'(\mathbf{y} - \boldsymbol{\varphi}) + 2b_e \right]^{-\frac{1}{2}(n + 2a_e)} [\mathbf{s}'\mathbf{s} + 2b_s]^{-\frac{1}{2}(q + a_s)},
\end{aligned}
$$

(2.14)

where $\boldsymbol{\varphi} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{s}$. In the remainder of this paper, we study the behavior of (15) in $I\!\!R^p \times I\!\!R^q$. We are interested in determining whether (15) is uni- or multimodal, and in establishing the location and relative size of

the mode(s). Inferences about $\boldsymbol{\beta}$ and $\mathbf{s}$ can be based on the joint marginal posterior distribution of $\boldsymbol{\beta}$ and $\mathbf{s}$, as given in (15).

Note that $p(\boldsymbol{\beta}, \mathbf{s}|\mathbf{y})$ is proportional to the product of two multivariate t-density kernels. This distribution is known as a product form poly-t, and was studied by Drèze (1977), Richard and Tompa (1980), Drèze and Richard (1983), Lubrano et al. (1984), Richard (1984) and Bauwens and Richard (1985). The poly-t distribution arises as the posterior density function of regression coefficients of a single structural equation under various non-informative prior specifications, or in instrumental variable settings within a Bayesian framework. In a series of papers, the authors above characterize the poly-t distribution, both in product and ratio form, and propose algorithms and software that can be used to evaluate normalizing constants, first and second order moments of the distributions, and various fractiles. Results presented in this paper complement and expand the work of Drèze, Richard, and collaborators. The special structure of the product form poly-t distribution will be useful when trying to describe $p(\boldsymbol{\beta}, \mathbf{s}|\mathbf{y})$ in $I\!\!R^p \times I\!\!R^q$.

Two special cases of model (2) may be of interest. Consider first the random effects model with a common mean μ. For this model,

$$(2.15) \qquad \mathbf{y} = \mathbf{1}\mu + \mathbf{Z}\mathbf{s} + \mathbf{e},$$

where $\mathbf{1} \in I\!\!R^n$ is a vector of 1's. Under the set of preceding assumptions, and by substituting $\mathbf{1}\mu$ for $\mathbf{X}\boldsymbol{\beta}$, the form of the marginal posterior density of μ and $\mathbf{s}$ is

$$
p(\mu, \mathbf{s}|\mathbf{y}) = k_5 \left[(\mathbf{y} - \mathbf{1}\mu - \mathbf{Z}\mathbf{s})'(\mathbf{y} - \mathbf{1}\mu - \mathbf{Z}\mathbf{s}) + 2b_e \right]^{-\frac{1}{2}(n+2a_e)}
$$
(2.16)
$$
\left[\mathbf{s}'\mathbf{s} + 2b_s \right]^{-\frac{1}{2}(q+a_s)}.
$$

The second special case is the completely random model

$$(2.17) \qquad \mathbf{y} = \mathbf{Z}\mathbf{s} + \mathbf{e}$$

which leads in a similar fashion to

$$
p(\mathbf{s}|\mathbf{y}) = k_6 \left[(\mathbf{y} - \mathbf{Z}\mathbf{s})'(\mathbf{y} - \mathbf{Z}\mathbf{s}) + 2b_e \right]^{-\frac{1}{2}(n+2a_e)} \left[\mathbf{s}'\mathbf{s} + 2b_s \right]^{-\frac{1}{2}(q+a_s)}.
$$
(2.18)

In Section 3, we analyze the local maxima, i.e. the modes, of the posterior density (15) The function (17) is just a special case of (15), and the form (17) is discussed at the end of the next section.

3. THE MAXIMA OF POSTERIOR DENSITIES

The main result of this paper is the following.

Theorem 1 *Consider the mixed linear model (2.1), with prior distributions given in (2.4) - (2.5), (2.9)-(2.11), for the limiting case (2.7). Then the posterior density $p(\beta, \mathbf{s}|\mathbf{y})$ in (2.14) attains all its (local) maxima on a one-dimensional curve $\Phi(\kappa)$, $\kappa \in [0, \bar{\kappa}]$, in $(\beta, \mathbf{s})$-space.*

The proof of Theorem 1 is given in the Appendix. Here we describe how to compute the curve $\Phi(\kappa)$ in general, and for an important special case. The (local) maxima of the posterior density $p(\beta, \mathbf{s}|\mathbf{y})$ can then be found by maximizing a real function defined on a bounded interval $[0, \bar{\kappa}]$, using for example, Newton-Raphson type algorithms.

We first introduce some notation to simplify the expressions in $p(\beta, \mathbf{s}|\mathbf{y})$ from (2.14). Let

$$c_1 = 2b_e > 0,\ c_2 = 2b_s > 0,\ m = \frac{1}{2}(n + 2a_e) \geq 1,\ l = \frac{1}{2}(q + a_s) \geq 1.$$

Then the function $h : I\!\!R^p \times I\!\!R^q \to I\!\!R$, defined by

$$(3.1)\quad h(\beta, \mathbf{s}) = \left[(\mathbf{y} - \mathbf{X}\beta - \mathbf{Z}\mathbf{s})'(\mathbf{y} - \mathbf{X}\beta - \mathbf{Z}\mathbf{s}) + c_1\right]^{-m}\left[\mathbf{s}'\mathbf{s} + c_2\right]^{-l}$$

attains its extreme values at the same points as $p(\beta, \mathbf{s}|\mathbf{y})$, and these values differ only by multiplication with the positive constant k_4.

Let $\mathbf{M} = (\mathbf{X}, \mathbf{Z})$ be the matrix made up of the columns of $\mathbf{X}$ and $\mathbf{Z}$ and let $p + r$, with $0 \leq r \leq q$ be the rank of $\mathbf{M}$. Let $\mathbf{N}$ be a submatrix of $\mathbf{M}$ with $p + r$ linearly independent columns, and let $V \subset I\!\!R^p \times I\!\!R^q$ be the corresponding subspace. We first define a curve $\sim\Phi(\kappa)$, $\kappa \in [0, \bar{\kappa}]$ in V.

Let $\mathbf{v}_1 = (\mathbf{N}'\mathbf{N})^{-1}\mathbf{N}'\mathbf{y}$ with components $\mathbf{v}_1 = \begin{pmatrix} \beta_1 \\ \mathbf{s}_1 \end{pmatrix} \in V$, and

$\mathbf{v}_0 = \begin{pmatrix} \beta_0 \\ \mathbf{0} \end{pmatrix} \in V$ with $\beta_0 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. Set $\bar{\kappa} = \mathbf{s}_1'\mathbf{s}_1$. The curve $\sim\Phi(\kappa)$ is defined by

$$(3.2) \qquad \sim\Phi(0) \;=\; \mathbf{v}_1,$$

$$(3.3) \qquad \sim\Phi(\bar{\kappa}) \;=\; \mathbf{v}_0,$$

$$(3.4) \qquad \sim\Phi(\kappa) \;=\; \begin{pmatrix} \beta(\kappa) \\ \mathbf{s}(\kappa) \end{pmatrix}, \; 0 < \kappa < \bar{\kappa}.$$

For example, the curve $\sim\Phi$ evaluated at 0, is a point in $(\beta, \mathbf{s})$-space whose coordinates are the components of the vector $\mathbf{v}_1$. Here, $\mathbf{s}(\kappa)$ is the solution to the system of equations

$$(3.5) \qquad \mathbf{s} \;=\; -\lambda\left[\mathbf{Z}_s'(\mathbf{I}-\mathbf{P})\mathbf{Z}_s(\mathbf{s}_1 - \mathbf{s})\right]$$

$$\kappa = (\mathbf{s}_1 - \mathbf{s})'(\mathbf{I}-\mathbf{P})'(\mathbf{I}-\mathbf{P})\mathbf{Z}_s(\mathbf{s}_1 - \mathbf{s})$$

$$(3.6) \qquad\qquad =\; (\mathbf{s}_1 - \mathbf{s})'(\mathbf{I}-\mathbf{P})\mathbf{Z}_s(\mathbf{s}_1 - \mathbf{s}),$$

with $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, and $\mathbf{N} = (\mathbf{X}, \mathbf{Z}_s)$, and $\beta(\kappa)$ is given by

$$(3.7) \qquad \mathbf{0} \;=\; \mathbf{X}'\mathbf{X}(\beta_1 - \beta(\kappa)) + \mathbf{X}'\mathbf{Z}_s(\mathbf{s}_1 - \mathbf{s}(\kappa))$$

$$(3.8) \qquad \mathbf{s}(\kappa) \;=\; -\lambda\left[\mathbf{Z}_s'\mathbf{X}(\beta_1 - \beta(\kappa)) + \mathbf{Z}_s'\mathbf{Z}_s(\mathbf{s}_1 - \mathbf{s}(\kappa))\right].$$

Note that $\beta(\kappa)$ is a straight line in $\mathbb{R}^p$.

To obtain the values of h in (3.1) on the maximum curve $\Phi(\kappa)$, $\kappa \in [0, \bar{\kappa}]$, define

$$\tilde{f} \;:\; V \to \mathbb{R}, \; \tilde{f}(\beta, \mathbf{s}) = (\mathbf{s}'\mathbf{s} + c_2)^l$$

$$\tilde{g} \;:\; V \to \mathbb{R}, \; \tilde{g}(\beta, \mathbf{s}) = \left[\left(\mathbf{y} - \mathbf{N}\begin{pmatrix} \beta \\ \mathbf{s} \end{pmatrix}\right)'\left(\mathbf{y} - \mathbf{N}\begin{pmatrix} \beta \\ \mathbf{s} \end{pmatrix}\right) + c_1\right]^m$$

$$\xi^2 \;=\; \Phi(0)_s'\,\boldsymbol{\Phi}(0)_s\,/\!\sim\!\Phi(0)_s'\,\sim\!\Phi(0)_s.$$

(3.9)

Then

$$(3.10) \qquad h(\Phi(\kappa)) = \tilde{g}(\sim\Phi(\kappa))^{-1} \cdot \tilde{f}(\xi \cdot \sim\Phi(\kappa))^{-1}$$

and $\Phi\left(\kappa\right) = \begin{pmatrix} \Phi\left(\kappa\right)_\beta \\ \Phi\left(\kappa\right)_s \end{pmatrix} \in \mathbb{R}^p \times \mathbb{R}^q$ is of the form

$$(3.11) \qquad \Phi\left(\kappa\right)_\beta = \sim\Phi\left(\kappa\right)_\beta$$

$$(3.12) \qquad \Phi\left(\kappa\right)_s = \left(\hat{\mathbf{Z}}'\hat{\mathbf{Z}}\right)^{-1}\hat{\mathbf{Z}}'\sim\Phi\left(\kappa\right)_s$$

with $\hat{\mathbf{Z}}$ a $q \times r$ matrix whose columns are r linearly independent rows of $\mathbf{Z}$. Note that $\Phi\left(\bar{\kappa}\right) = \begin{pmatrix} \beta_0 \\ \mathbf{0} \end{pmatrix} = \mathbf{v}_0$, and $\Phi\left(0\right) = \begin{pmatrix} \beta_1 \\ \mathbf{s}_0 \end{pmatrix} = \mathbf{w}_1$ with

$$\mathbf{s}_0 = \left(\hat{\mathbf{Z}}'\hat{\mathbf{Z}}\right)^{-1}\hat{\mathbf{Z}}'\mathbf{s}_1.$$

Maximization of the function h in (3.10) is done over the interval $[0, \bar{\kappa}]$, and the points where the (local) maxima are attained, can be computed according to (3.11)-(3.12).

Corollary 2

1. *If $r = 0$, then the posterior distribution $p\left(\boldsymbol{\beta}, \mathbf{s}|\mathbf{y}\right)$ has exactly one maximum, attained at $\mathbf{v}_1 \in \mathbb{R}^p$, which is given by the normal equations $\mathbf{X}'\mathbf{X}\mathbf{v}_1 = \mathbf{X}'\mathbf{y}$, or $\mathbf{v}_1 = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$.*

2. *If $\mathbf{v}_1 = \left(\mathbf{N}'\mathbf{N}\right)^{-1}\mathbf{N}\mathbf{y} \in \mathbb{R}^p$, then $p\left(\beta, \mathbf{s}|\mathbf{y}\right)$ has exactly one maximum, attained at $\mathbf{v}_1$.*

3. *In both cases, $\mathbf{v}_1 = \mathbb{E}\left\{\beta, \mathbf{s}|\mathbf{y}\right\}$, i.e., $\mathbf{v}_1$ is the joint posterior mean of β and $\mathbf{s}$.*

Proof: 1. and 2. follow directly from the computations in the proof of the theorem. The fact that $p\left(\beta, \mathbf{s}|\mathbf{y}\right)$ is symmetric on each line through $\mathbf{v}_1$ under conditions 1. or 2. yields the result 3.

We now consider an important special case, for which the (local) maxima can be computed explicitly. Recall from the discussion above that the β-component of $\sim\boldsymbol{\Phi}\left(\kappa\right)$ (and hence, of $\Phi\left(\kappa\right)$) is always a straight line. Furthermore, it follows from the proof of Theorem 1 in the Appendix, that $\sim\Phi\left(\kappa\right)_s = \mathbf{s}\left(\kappa\right)$ is a line if and only if one of the (generalized) eigendirections of $\mathbf{Z}'_s\left(\mathbf{I} - \mathbf{P}\right)\mathbf{Z}_s$ is $\mathbf{s}_1$. This is, in particular, the case if the level surfaces of this quadratic form are spheres, or if $r = 1$. If $\sim\Phi\left(\kappa\right)$ is a straight line in

V, so is $\Phi(\kappa)$ in $\mathbb{R}^p \times \mathbb{R}^q$. For this case, we obtain expressions, as follows, for the modes of the posterior density in explicit form. In order to simplify notation, we reparameterize the line segment $\{\Phi(\kappa), \kappa \in [0, \bar{\kappa}]\}$ as

$\quad\quad \Phi: \mathbb{R} \to \mathbb{R}^p \times \mathbb{R}^q$,

$$(3.13) \quad\quad\quad \Phi(\alpha) = (1 - \alpha)\mathbf{v}_0 + \alpha\mathbf{w}_1, \quad \alpha \in [0, 1].$$

The mode(s) of $p(\beta, \mathbf{s}|\mathbf{y})$ can now be computed as the maxima of the function $h(\Phi(\kappa))$, with h defined as in (3.1). With $\mathbf{v}_0$ and $\mathbf{w}_1$ as above, let

$$(3.14) \quad\quad a_0 = \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{M}\mathbf{v}_0 + \mathbf{v}_0'\mathbf{M}'\mathbf{M}\mathbf{v}_0 + c_1,$$

$$(3.15) \quad\quad a_1 = -\mathbf{y}'\mathbf{M}\mathbf{w}_1 + \mathbf{y}'\mathbf{M}\mathbf{v}_0 + \mathbf{w}_1'\mathbf{M}'\mathbf{M}\mathbf{v}_0 - \mathbf{v}_0'\mathbf{M}'\mathbf{M}\mathbf{v}_0$$

$$(3.16) \quad\quad a_2 = \mathbf{w}_1'\mathbf{M}'\mathbf{M}\mathbf{w}_1 - 2\mathbf{w}_1'\mathbf{M}'\mathbf{M}\mathbf{v}_0 + \mathbf{v}_0'\mathbf{M}'\mathbf{M}\mathbf{v}_0.$$

Then $g(\Phi(\alpha)) = \left[\left(\mathbf{y} - \mathbf{M}\begin{pmatrix}\beta\\\mathbf{s}\end{pmatrix}\right)'\left(\mathbf{y} - \mathbf{M}\begin{pmatrix}\beta\\\mathbf{s}\end{pmatrix}\right) + c_1\right]^m$ reads

$$(3.17) \quad\quad\quad g(\Phi(\alpha)) = \left(a_0 + 2a_1\alpha + a_2\alpha^2\right)^m.$$

For $\mathbf{x} \in \mathbb{R}^p \times \mathbb{R}^q$ denote by $\mathbf{x}_s$ the components in the $\mathbf{s}$-space $\mathbb{R}^q$. Set

$$(3.18) \quad\quad\quad b_0 = \mathbf{v}_{0,s}'\mathbf{v}_{0,s} + c_2$$

$$(3.19) \quad\quad\quad b_1 = \mathbf{v}_{0,s}'\mathbf{w}_{1,s} - \mathbf{v}_{0,s}'\mathbf{v}_{0,s},$$

$$(3.20) \quad\quad\quad b_2 = \mathbf{w}_{1,s}'\mathbf{w}_{1,s} - 2\mathbf{v}_{0,s}'\mathbf{w}_{1,s} + \mathbf{v}_{0,s}'\mathbf{v}_{0,s}.$$

Then $f(\Phi(\alpha)) = (\mathbf{s}'\mathbf{s} + c_2)^l$ can be written as

$$(3.21) \quad\quad\quad f(\Phi(\alpha)) = \left(b_0 + 2b_1\alpha + b_2\alpha^2\right)^l.$$

Furthermore, let $A = 2ma_1b_0 + 2la_0b_1$, $B = 4ma_1b_1 + 2ma_2b_0 + 4la_1b_1 + 2la_0b_2$, $C = 2ma_1b_2 + 4ma_2b_1 + 2la_2b_1 + 4la_1b_2$, and $D = 2ma_2b_2 + 2la_2b_2$. Then, differentiating the product $f(\Phi(\alpha)) \cdot g(\Phi(\alpha))$ with respect to α, and equating to zero yields,

$$(3.22) \quad\quad\quad P(\alpha) = D\alpha^3 + C\alpha^2 + B\alpha + A = 0,$$

a third-degree polynomial in α. The explicit solution of (3.22) can be obtained via Cardano's rule. Let

$$a = \frac{B}{D} - \frac{C^2}{3D^2}, \quad b = \frac{2}{27}\frac{C^3}{D^3} - \frac{BC}{3D^2} + \frac{A}{D},$$

and let F denote the discriminant of the reduced equation (with $\gamma = \alpha + \frac{C}{3D}$)

(3.23) $$\gamma^3 + a\gamma + b = 0,$$

(3.24) $$F = \left(\frac{a}{3}\right)^3 + \left(\frac{b}{2}\right)^2.$$

We then obtain, with $u = \left(-\frac{b}{2} + \sqrt{F}\right)^{1/3}$, $t = \left(-\frac{b}{2} - \sqrt{F}\right)^{1/3}$, the following criterion:

1. If $F > 0$, then (3.22) has one real root, namely $\gamma_1 = u + t$.

2. If $F = 0$, then (3.22) has three real roots, one of which is double, namely $\gamma_1 = -\frac{1}{2}(u + t)$, $\gamma_2 = \frac{1}{2}(u + t)$, where γ_2 is the double root.

3. If $F < 0$, then (3.22) has three distinct real roots, namely $\gamma_1 = u + t$, $\gamma_2 = -\frac{1}{2}(u + t) + j(u - t)\sqrt{3}$, $\gamma_3 = -\frac{1}{2}(u + t) - j\frac{1}{2}(u - t)\sqrt{3}$, where j is the imaginary unit.

Resubstitution via $\alpha_i = \gamma_i - \frac{C}{3D}$ yields:

1. The posterior density $p(\beta, \mathbf{s}|\mathbf{y})$ is unimodal, iff $F \geq 0$, and in this case, $\nu = \alpha_1 \mathbf{w}_1 + (1 - \alpha_1)\mathbf{v}_0$ is the mode.

2. The posterior density $p(\beta, \mathbf{s}|\mathbf{y})$ is bimodal, iff $F < 0$, and in this case, $\nu_i = \alpha_i \mathbf{w}_1 + (1 - \alpha_i)\mathbf{v}_0$, $i = 1, 2$, are the modes.

These computations show that in the setup of this special case, no iterative procedure is necessary for determining the modes. It is only necessary to invert the matrices $\mathbf{X}'\mathbf{X}$ and $\mathbf{N}'\mathbf{N}$, to compute $\left(\hat{\mathbf{Z}}'\hat{\mathbf{Z}}\right)^{-1}$, and to solve (3.12) for $\mathbf{w}_1$. These computations are straightforward.

Results from Theorem 1 and the special case above can be used to custom-tailor MCMC algorithms:

- If the mode ν is unique, use information about the location of ν on the curve $\{\Phi(\kappa), \kappa \in [0, \bar{\kappa}]\}$ to "blast" into the area of $p(\beta, \mathbf{s}|\mathbf{y})$ of highest mass.

- If the stationary distribution $p(\beta, \mathbf{s}|\mathbf{y})$ is multi-modal, i.e., has multiple local maxima $(\Phi(\kappa_i), \nu_i)$, $i \geq 2$, then design the algorithm so that it visits all the modes, and set the number of visits to be proportional to the relative size ν_i of each mode.

We conclude this section with some remarks on the application of our main result. Consider the completely random model (2.17) $\mathbf{y} = \mathbf{Z}\mathbf{s} + \mathbf{e}$. Define $W := l.s. \{Z_1, ..., Z_q\} \subset I\!\!R^n$, and reorder the columns of $\mathbf{Z}$ such that $\{Z_1, ..., Z_r\}$ is a basis of W. Let $V = \mathbf{Z}^{-1}[W]$, the inverse image of W under $\mathbf{Z}$ with basis $\{\hat{e}_1, ..., \hat{e}_r\}$, and denote the kernel of $\mathbf{Z}$ by K. Then the point $\mathbf{v}_1$ is given by the normal equations $\mathbf{v}_1 = \left(\mathbf{Z}|_V' \mathbf{Z}|_V\right)^{-1} \mathbf{Z}|_V' \mathbf{y} \in I\!\!R^q$, and we have $\mathbf{v}_0 = \mathbf{s}_0 = \mathbf{0}$. With this notation, the proof of Theorem 1 goes through, yielding one or two (local) maxima for the posterior density $p(\mathbf{s}|\mathbf{y})$.

If $p(\beta, \mathbf{s}|\mathbf{y})$ has a unique maximum, the point where this maximum is attained, need not be the posterior mean (see the example in Section 4). However, each component of the curve $\sim\Phi(\kappa)$, $\kappa \in [0, \bar{\kappa}]$ is monotone in V, and hence, the posterior mean in V is always contained in the multidimensional rectangle described by $\sim\Phi(0)$ and $\sim\Phi(\bar{\kappa})$. Of course, the same holds true in $I\!\!R^p \times I\!\!R^q$, using the corners $\Phi(0)$ and $\Phi(\bar{\kappa})$, since Φ is obtained from $\sim\Phi$ by projection along a subspace. This gives a rough estimate for $I\!\!E(\beta, \mathbf{s}|\mathbf{y})$.

The proof of Theorem 1 shows that if $\mathbf{v}_1 \in I\!\!R^p$, then the posterior mode(s) and the posterior mean cannot be attained at $(\beta_0, \mathbf{0})$ or $\mathbf{w}_1$ since $p(\beta, \mathbf{s}|\mathbf{y})$ is strictly increasing at $\mathbf{w}_1 = \Phi(0)$, and strictly decreasing at $(\beta_0, \mathbf{0}) = \Phi(\bar{\kappa})$. For numerical computations of the mode(s) as maxima of $p(\beta, \mathbf{s}|\mathbf{y})$ via Newton-Raphson type methods in $I\!\!R^p \times I\!\!R^q$, however, $(\beta_0, 0)$ and $\mathbf{w}_1$ are reasonable starting values.

4. AN INSTRUCTIVE EXAMPLE

The proof of Theorem 1 is constructive, and therefore explicit criteria for unimodality can be given, once the function Φ, defined in (6.1) in the Appendix, is known. Actually, the equations (**??**)-(**??**) show that it suffices to know $\sim\Phi$ and ξ^2. In general, one can proceed as follows: Solve the normal equations $\mathbf{N}'\mathbf{N}\mathbf{v}_1 = \mathbf{N}'\mathbf{y}$ and $\mathbf{X}'\mathbf{X}\beta_0 = \mathbf{X}'\mathbf{y}$ given in Section 3 to obtain the points $\mathbf{v}_1$ and

$$\mathbf{v}_0 = \begin{pmatrix} \beta_0 \\ \mathbf{0} \end{pmatrix}$$

in V. Compute $\bar{\kappa} = \mathbf{s}_1'\mathbf{s}_1$, and solve the system of equations

$$\mathbf{s} = -\lambda \left[\mathbf{Z}_s' (\mathbf{I} - \mathbf{P}) \mathbf{Z}_s (\mathbf{s}_1 - \mathbf{s})\right]$$

$$(4.1) \qquad\qquad \kappa \;=\; (\mathbf{s}_1 - \mathbf{s})' \, (\mathbf{I} - \mathbf{P}) \, \mathbf{Z} \, (\mathbf{s}_1 - \mathbf{s}),$$

for $\kappa \in [0, \bar{\kappa}]$, to obtain $\mathbf{s}(\kappa)$ in V_s. (One has to set $\mathbf{s}(0) = \mathbf{s}_1$, since the Lagrange multiplier is singular at this point.) Compute $\beta(\kappa)$ via (6.5) - (6.8), to obtain $\sim\!\Phi(\kappa) = (\beta(\kappa), \mathbf{s}(\kappa))$ in V. For $\kappa = 0$, (6.17) is used to get $\Phi(0)_s$, and now ξ^2 in (6.20) can be computed. This allows maximization of $h = \tilde{f} \cdot \tilde{g}$ on $\sim\!\Phi(\kappa)$ to obtain the value(s) κ_i, where the maxima are attained. Another application of (6.18) yields the **s**-components of $\Phi(\kappa_i)$, and finally, we get the desired point(s) $\Phi(\kappa_i)$ from (6.18). The entire procedure requires the computation of three inverse matrices of $(\mathbf{N}'\mathbf{N}, \mathbf{X}'\mathbf{X}, \hat{\mathbf{Z}}'\hat{\mathbf{Z}})$ of dimension $p + r$, p, and $q - r$, respectively, the solution of five systems of linear equations (in $p + r$, p, $q - r$, $q - r$ variables), and the solution of (4.1) (in $r + 1$ variables), which is quadratic in one equation, for all values $\kappa \in (0, \bar{\kappa})$. In specific cases, explicit formulas for $\mathbf{s}(\kappa)$ in (4.1) can be obtained via symbolic manipulation software. In the previous section, we derived an explicit criterion for uni- and bimodality in the case where $\sim\!\Phi$ is a straight line in V. Here, we present a simple example that demonstrates, how uni- and bimodality depend on the data.

4.1 Example

A simple example illustrates the steps needed for the computation of the mode(s) of $p(\beta, \mathbf{s}|\mathbf{y})$, and shows how uni- or multimodality of the posterior distribution depend on the observations. Let $n = 4$, $p = 1$, $q = 2$, and consider the matrices

$$\mathbf{X} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad \mathbf{Z} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \text{and } \mathbf{M} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}.$$

Then, rank$(\mathbf{M}) = 2$, and hence, $r = 1$. Let $a_e = 2$, and $a_s = 1$. Then, $m = \frac{1}{2}(n + 2a_e) = 4$, and $l = \frac{1}{2}(q + a_s) = 1.5$. Set, furthermore, $2b_e = c_1 = 2$, and $2b_s = c_2 = 1.0$. For this setup we have

$$W \;=\; l.s. \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix} \right\} \subset \mathbb{R}^4, \; V = l.s. \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\} \subset \mathbb{R}^3,$$

$$K = \ker(\mathbf{M}) = l.s.\left\{\begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}\right\}.$$

Solutions to normal equations $\mathbf{N'N v_1} = \mathbf{N'y}$ and $\mathbf{X'X}\beta_0 = \mathbf{X'y}$ are the vectors $\mathbf{v_1}$ and $\mathbf{v_0}$, where

$$\mathbf{v_1} = \tfrac{1}{2}\begin{pmatrix} y_1 + y_3 \\ -y_1 + y_2 - y_3 + y_4 \\ 0 \end{pmatrix}, \mathbf{v_0} = \tfrac{1}{4}\begin{pmatrix} y_1 + y_2 + y_3 + y_4 \\ 0 \\ 0 \end{pmatrix}.$$

Equation (6.17) gives, as the component in the kernel K,

$$\mathbf{k} = \frac{\alpha}{4}(y_1 - y_2 + y_3 - y_4)\begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix},$$

and hence,

$$\Phi(\alpha) = \tfrac{1}{4}\begin{bmatrix} \sum_{i=1}^4 y_i \\ 0 \\ 0 \end{bmatrix} + \tfrac{\alpha}{4}\begin{pmatrix} 0 \\ -y_1 + y_2 - y_3 + y_4 \\ y_1 - y_2 + y_3 - y_4 \end{pmatrix}.$$

Consider the following data sets:

Data Set	y_1	y_2	y_3	y_4
1	-1.0	5.0	-2.0	5.0
2	-1.0	5.0	-4.34	5.0
3	-1.0	5.0	-6.0	5.0

The graphs of $h(\Phi(\alpha))$ for $\alpha \in [-0.5, 1.5]$ are shown in Figures 1.a -1.b, for data sets 1 and 3. The corresponding posterior densities are shown in Figures 2 and 3. Solutions for each of the data sets are:

For data set 1: There is a unique mode at $\alpha \approx 0.978$, and hence, the mode is

$$\nu \approx \begin{pmatrix} 1.75 \\ 3.178 \\ -3.178 \end{pmatrix}.$$

For data set 2: There is a unique mode at $\alpha_1 \approx 0.95$, yielding

$$\nu_1 \approx \begin{pmatrix} 1.165 \\ 3.643 \\ -3.643 \end{pmatrix},$$

and a double root of (3.22). Hence, there is an inflection point of $h\left(\Phi\left(\alpha\right)\right)$ at $\alpha_2 \approx 0.161$, which corresponds to the point $[1.165, 0.617, -0.617]'$.

For data set 3: There are two modes, at $\alpha_1 \approx 0.918$ and $\alpha_2 \approx 0.08$, i.e.

$$\nu_1 \approx \begin{pmatrix} 0.750 \\ 3.901 \\ -3.901 \end{pmatrix}, \nu_2 \approx \begin{pmatrix} 0.75 \\ 0.34 \\ -0.34 \end{pmatrix}.$$

In the third data set, the first mode ν_1 is approximately 10 times more significant than the second mode ν_2. Therefore, an algorithm designed to visit both modes should spend about 10 times longer around ν_1.
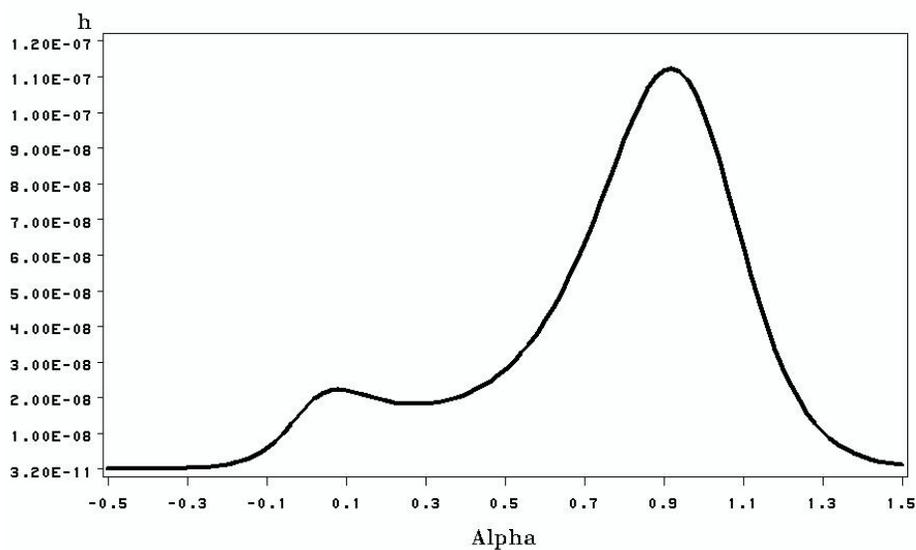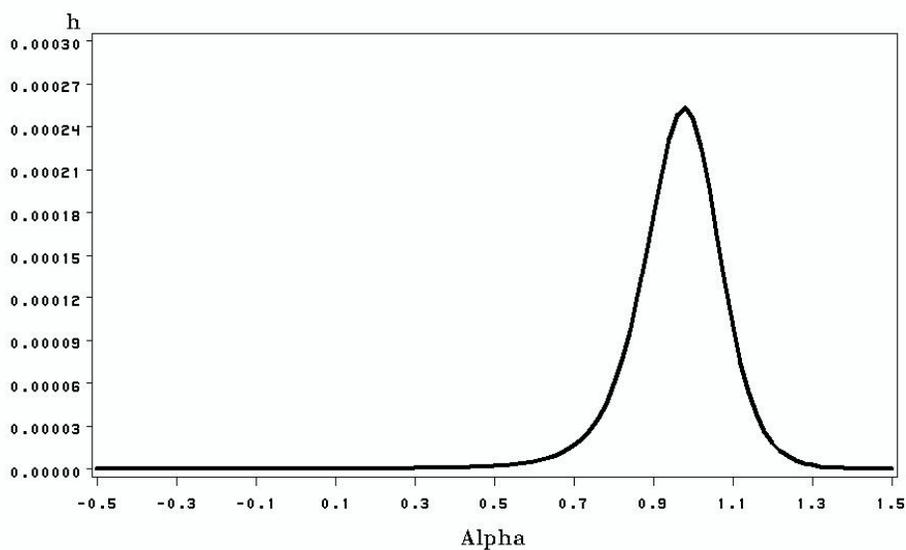
Figure 1a, 1b Plot of $h(\beta, \mathbf{s})$ on the line $\boldsymbol{\Phi}(\alpha)$, against values of $\alpha \in [-0.5, 1.5]$. Plots correspond to data set 1 (Fig. 1a), and data set 3 (Fig. 1b), respectively.
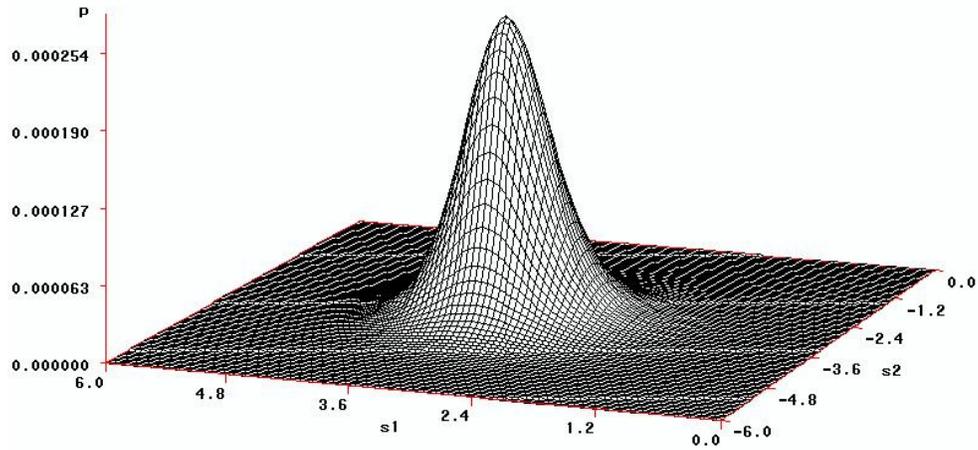
Figure 2 Posterior density $p(\beta, \mathbf{s}|\mathbf{y})$ of $\mathbf{s}$, with β fixed at $\beta = 1.75$. Distribution corresponds to data set 1.
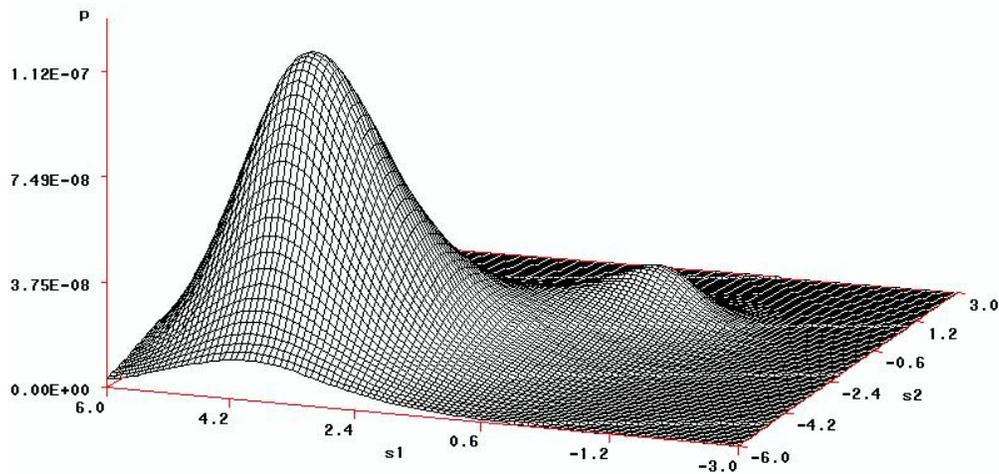
Figure 3 Posterior density $p(\beta, \mathbf{s}|\mathbf{y})$ of $\mathbf{s}$, with β fixed at $\beta = 0.75$. Distribution corresponds to data set 3.

5. CONCLUSIONS

For a mixed linear model like (2.1), with location and dispersion assumptions as those in (2.4)-(2.5), the estimation of the fixed and random effects β and $\mathbf{s}$ may be of interest. If the variance components $(\tau_s)^{-1}$ and $(\tau_e)^{-1}$ are known, the optimum solution to the prediction problem is to obtain the best linear unbiased estimate (BLUE) of β and the best linear unbiased predictor BLUP of $\mathbf{s}$. If the variance components are unknown, β and $\mathbf{s}$ can be estimated within a Bayesian framework, and MCMC methods provide a convenient means to do so. Alternatively, Laplace methods can be implemented to approximate moments of the posterior distribution of interest.

In either case, knowledge about the shape of the marginal posterior distribution of the parameters of interest is important. The number, approximate location, and relative size of each (local) maxima of the distribution provide critical information to assess the performance of Laplace methods, custom-designed MCMC algorithms, and to construct highest posterior density regions for the parameters under study. The type of problems we have in mind are high dimensional, where the vector of parameters might have hundreds of thousands of dimensions, as is the case in animal breeding applications. In those cases, some idea about the characteristics of the very high-dimensional surface can be an important aid in analysis.

We have shown that for the class of mixed linear models with any number of fixed effects and one vector of random effects (possibly very high-dimensional), and for prior assumptions as in (2.4), (2.5), (2.9)-(2.11), the maxima of the posterior distribution $p(\beta, \mathbf{s}|\mathbf{y})$ occur on a curve segment $\Phi(\kappa)$, $\kappa \in [0, \bar{\kappa}]$. If the curve Φ is a straight line, uni- or multimodality of the stationary distribution depends critically on the data, and can be determined by inspection of the roots of a third-degree polynomial.

Information about the mode(s) of the distribution $p(\boldsymbol{\beta}, \mathbf{s}|\mathbf{y})$, that has a poly-t form, can be used to either start a sampling algorithm (see, e.g., Lubrano et al., 1984; Bauwens and Richard, 1985), or to tailor the algorithm so that it will visit all modes. Consider the problem of drawing an initial value $\left(\beta^0, \mathbf{s}^0\right)$ from a starting distribution $p\left(\beta^0, \mathbf{s}^0\right)$ so that $p\left(\beta^0, \mathbf{s}^0|\mathbf{y}\right) > 0$. The value $\left(\beta^0, \mathbf{s}^0\right)$ may result, for example, from importance sampling from an approximation to $p\left(\beta^0, \mathbf{s}^0\right)$ built around the mode(s) of $p(\beta, \mathbf{s}|\mathbf{y})$. The procedure presented in this paper for exploration of a specific stationary distribution is easily implemented, since it does not involve sophisticated numerical methods. In fact, all that is required is the solution of sets of nor-

mal equations, and of systems of linear equations, with one quadratic term. In this regard, it provides a convenient analytical tool for exploratory analysis, and may improve the performance of numerical methods to approximate poly-t distributions.

References

[1] BAUWENS, L., and RICHARD, J-F. A 1-1 Poly-t random variable generator with application to Monte Carlo integration. *J. of Econometrics* **29**, pp. 19–46, (1985).

[2] BERGER, J. O. *Statistical Decision Theory and Bayesian Analysis*, 2nd ed., New York, NY: Springer-Verlag, (1985).

[3] BESAG, J. and GREEN, P. J. Spatial statistics and Bayesian computation. *JRSS (B)* **55**, pp. 25–37, (1993).

[4] BOX, G. E. P., and TIAO, G. C. *Bayesian Inference in Statistical Analysis*. Reading, MA: Addison-Wesley, (1973).

[5] DRÈZE, J. H. Bayesian regression analysis using poly-t densities. *J. of Econometrics* **6**, pp. 329–354, (1977).

[6] DRÈZE, J. H., and RICHARD, J-F. Bayesian analysis of simultaneous equation systems. In: GRILICHES, Z., and INTRILIGATOR, M., (eds) *Handbook of Econometrics*. North Holland, Amsterdam, pp. 369–377, (1983).

[7] FULLER, W. A., and HARTER, R. M. The multivariate components of variance model for small area estimation. In: PLATEK, R., RAO, J. N. K., SARNDAL, C. E., and SINGH, M. P., (eds.) *Small Area Statistics*, pp. 103–123. Wiley, New York, (1987).

[8] GELFAND, A. E., and SMITH, A. F. M. Sampling based approaches to calculating marginal densities. *J. of Amer. Statis. Assoc.* **85**, pp. 398–409, (1990).

[9] GOLDBERGER, A. S. Best linear unbiased prediction in the generalized linear regression. *J. Amer. Statist. Assoc.* **57**, pp. 369–375, (1962).

[10] HARVEY, A. C. *Forecasting, Structural Time Series, and the Kalman Filter*. Cambridge University Press, Cambridge, MA, (1989).

[11] HARVILLE, D. A, and CARRIQUIRY, A. L. Classical and Bayesian prediction as applied to an unbalanced mixed linear model. *Biometrics* **48**: pp. 987–1003, (1992).

[12] HENDERSON, C. R. Selection index and expected genetic advance. In: HANSON, W. D., and ROBINSON, H. F. (eds.) *Statistical Genetics and Plant Breeding*, pp. 141–163. NAS-NRC Publ 982, (1963).

[13] HOBERT, J. P. and CASELLA, G. The effect of improper priors on Gibbs sampling in hierarchical linear models. *J. of Amer. Statis. Assoc.* **91**, pp. 1461–1473, (1996).

[14] JEFFREYS, H. *Theory of Probability*, 3rd ed., Oxford University Press, London, (1961).

[15] KASS, R. E., and STEFFEY, D. Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models). *J. Amer. Statist. Assoc.* **84**, pp. 717–726, (1989).

[16] LINDLEY, D. V. Approximate Bayesian methods. In: BERNARDO, J. M., DEGROOT, M. H., LINDLEY, D. V., and SMITH, A. F. M. (eds.) *Bayesian Statistics 1*, pp. 223–245. University of Valencia, Spain, (1980).

[17] LUBRANO, M., PIERSE, R. G., and RICHARD, J-F. Stability of a UK money demand equation: a Bayesian approach to testing exogeneity. Technical report, CORE, Louvain-la-Neuve, (1984).

[18] MORRIS, C. N. Approximate posterior distributions and posterior moments. In: BERNARDO, J. M., DEGROOT, M. H., LINDLEY, D. V., and SMITH, A. F. M. (eds.) *Bayesian Statistics 3*. Oxford University Press, Oxford, pp. 327–344, (1988).

[19] PROTTER, M. H., and MORREY, C. B. *A First Course in Real Analysis*. Springer-Verlag, New York, NY, (1977).

[20] RAIFFA, H., and SCHLAIFFER, R. *Applied Statistical Decision Theory*. Harvard University Press, Boston, MA, (1971).

[21] RICHARD, J-F. Classical and Bayesian inference in incomplete simultaneous equation models. In: HENDRY, D. F., and WALLIS, K. F. (eds.) *Econometric and Quantitative Economics*. Blackwell, Oxford, (1984).

[22] RICHARD, J-F., and TOMPA, H. On the evaluation of poly-t density functions. *J. of Econometrics* **12**, pp. 335–351, (1980).

[23] ROBERT, C. P., and CASELLA, G. *Monte Carlo Statistical Methods*. Springer-Verlag, New York, NY, (1999).

[24] SEARLE, S. *Linear Models*. Wiley, New York, NY, (1981).

[25] SMITH, A. F. M., GELFAND, A., and RACINE-POON, A. (1990).

[26] STROUD, T. W. F. Bayes and empirical Bayes approaches to small area estimation. In: PLATEK, R., RAO, J. N. K., SARNDAL, C. E., and SINGH, M. P. (eds.) *Small Area Statistics*. Wiley, New York, NY, pp. 124–140, (1987).

[27] TIERNEY, L., KASS, R. E., and KADANE, J. B. Fully exponential Laplace approximations to expectations and variances of nonpositive functions. *J. Amer. Statist. Assoc.* **84** pp. 710–716, (1989).

[28] VAN DYK, D. A. Fitting mixed-effects models using efficient EM-type algorithms. *J. of Graph. and Comput. Statis.*. In press, (1999).

[29] WANG, C. S., RUTLEDGE, J. J., and GIANOLA, D. Bayesian analysis of mixed linear models via Gibbs sampling with an application to litter size of Iberian pigs. *Genetique, Selection, Evolution* **26**, pp. 1–25 (1994).

Alicia L. Carriquiry
Department of Statistics
Iowa State University
Ames, IA 50011
U. S. A.
e-mail : alicia@iastate.edu

Alicia L. Carriquiry
Facultad de Matemáticas
Pontificia Universidad Católica de Chile
Santiago
Chile

and

Wolfgang Kliemann
Department of Mathematics
Iowa State University
Ames, IA 50011
U. S. A.
e-mail : kliemann@iastate.edu

Wolfgang Kliemann
Facultad de Matemáticas
Pontificia Universidad Católica de Chile
Santiago
Chile

6. APPENDIX

We prove the following result:

Theorem 1. Consider the mixed linear model (2.1) with prior distributions given in (2.4) - (2.5), (2.9)-(2.11), for the limiting case (2.7). Then the posterior density $p(\beta, \mathbf{s}|\mathbf{y})$ in (2.14) attains all its (local) maxima on a one-dimensional curve $\Phi(\kappa)$, $\kappa \in [0, \bar{\kappa}]$, in $(\beta, \mathbf{s})$-space.

Proof Let us first introduce some notation to simplify the expressions in $p(\beta, \mathbf{s}|\mathbf{y})$ from (2.14). Let

$c_1 = 2b_e > 0$, $c_2 = 2b_s > 0$, $m = \frac{1}{2}(n + 2a_e) \geq 1$, $l = \frac{1}{2}(q + a_s) \geq 1$.

Then the function $h : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$, defined by

$$(6.1) \quad h(\beta, \mathbf{s}) = \left[(\mathbf{y} - \mathbf{X}\beta - \mathbf{Z}\mathbf{s})'(\mathbf{y} - \mathbf{X}\beta - \mathbf{Z}\mathbf{s}) + c_1 \right]^{-m} \left[\mathbf{s}'\mathbf{s} + c_2 \right]^{-l}$$

attains its extreme values at the same points as $p(\beta, \mathbf{s}|\mathbf{y})$, and these values differ only by multiplication with the positive constant k_4.

Next, let $\mathbf{M} = (\mathbf{X}, \mathbf{Z})$ be the matrix made up of the columns of $\mathbf{X}$ and $\mathbf{Z}$. The matrix $\mathbf{M}$ has rank $p + r$, with $0 \leq r \leq q$. Denote by W the linear span of the columns of $(\mathbf{X}, \mathbf{Z}) = \mathbf{M}$, that is, $W := l.s. \{X_1, ..., X_p, Z_1, ..., Z_q\} \subset \mathbb{R}^n$. Reorder the columns of $\mathbf{Z}$ such that $\{X_1, ..., X_p, Z_1, ..., Z_r\}$ forms a basis of W. Define V to be the inverse image of W under $\mathbf{M}$, that is, the linear span formed by the set of $p + r$ standard basis vectors, i.e., $V := \mathbf{M}^{-1}[W] = l.s. \{e_1, ..., e_p, \hat{e}_1, ..., \hat{e}_r\}$, with the corresponding reordering of the random effects $\mathbf{s} \in \mathbb{R}^q$. Then $\mathbf{M}$ restricted to V, written as $\mathbf{N}$, is a linear isomorphism from V to W. Denote by $K \subset \mathbb{R}^p \times \mathbb{R}^q$, the kernel of $\mathbf{M} : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}^n$. Then, each $(\beta, \mathbf{s}) \in \mathbb{R}^p \times \mathbb{R}^q$ has a unique canonical decomposition $(\beta, \mathbf{s}) = \sum_{i=1}^p \beta_i e_i + \sum_{j=1}^r s_j \hat{e}_j + k$, where $\{e_1, ..., e_p, \hat{e}_1, ..., \hat{e}_r\}$ is the standard basis in V, and $k \in K$. Denote $l.s. \{\hat{e}_1, ..., \hat{e}_r\}$ by V_s, and the corresponding part of $\mathbf{Z}$ by $\mathbf{Z}_s$. Next, we solve the problem in V (the "nonsingular problem") by considering the minima and the level surfaces of the functions

$$(6.2) \qquad\qquad \tilde{f} : V \to \mathbb{R}, \ \tilde{f}(\beta, \mathbf{s}) = (\mathbf{s}'\mathbf{s} + c_2)^l$$

and

$$\tilde{g}: V \to I\!R, \ \tilde{g}(\beta, \mathbf{s}) = \left[\left(\mathbf{y} - \mathbf{N} \begin{pmatrix} \beta \\ \mathbf{s} \end{pmatrix} \right)' \left(\mathbf{y} - \mathbf{N} \begin{pmatrix} \beta \\ \mathbf{s} \end{pmatrix} \right) + c_1 \right]^m.$$

(6.3)

The product $\tilde{h} = \tilde{f} \cdot \tilde{g}$ has its minima exactly at the points, where $h_{|V}$, (h restricted to V) has its maxima.

Regarding $\tilde{f}$, the following results follow from its structure. In V_s, the quadratic form $\mathbf{s}'\mathbf{s}+c_2$ has a unique minimum at $\mathbf{s} = \mathbf{0}$. Hence, the level surfaces of $\tilde{f}$ in V are of the form $I\!R^p \times B(\mathbf{0}, \rho)$ where $B(\mathbf{0}, \rho) = \{ \mathbf{s} \in V_s, \mathbf{s}'\mathbf{s} = \rho; \rho \geq 0 \}$ is the Euclidean ball centered at zero with radius $\sqrt{\rho}$, and the levels of $\tilde{f}$ do not depend on β.

The behavior of the function $\tilde{g}$ can be investigated in a similar way. The quadratic form

$$\left[\left(\mathbf{y} - \mathbf{N} \begin{pmatrix} \beta \\ \mathbf{s} \end{pmatrix} \right)' \left(\mathbf{y} - \mathbf{N} \begin{pmatrix} \beta \\ \mathbf{s} \end{pmatrix} \right) + c_1 \right]$$

has a unique minimum at $\mathbf{v}_1 \in V$, where $\mathbf{v}_1$ is given by the normal equations

(6.4) $$\mathbf{N}'\mathbf{N}\mathbf{v}_1 = \mathbf{N}'\mathbf{y} \ or \ \mathbf{v}_1 = (\mathbf{N}'\mathbf{N})^{-1} \mathbf{N}'\mathbf{y}.$$

The level surfaces of the quadratic form in V are $L(\mathbf{v}_1, \eta) = \{ \mathbf{v} \in V, (\mathbf{v}_1 - \mathbf{v})' \mathbf{N}'\mathbf{N} (\mathbf{v}_1 - \mathbf{v}) = \eta \}$ for $\eta \geq 0$. Hence, the level surfaces of $\tilde{g}$ are of the form $L(\mathbf{v}_1, \eta)$, with levels in $[(\omega + c_1)^m, \infty)$, where ω is the minimal distance from $\mathbf{y}$ to W, i.e., $\omega^2 = (\mathbf{y} - \mathbf{N}\mathbf{v}_1)' (\mathbf{y} - \mathbf{N}\mathbf{v}_1)$.

In order to find the points where the minima of $\tilde{h} = \tilde{f} \cdot \tilde{g}$ occur, we proceed in the following way: $\tilde{f}$ is strictly increasing on the level surfaces $I\!R^p \times B(\mathbf{0}, \rho)$ as ρ increases, and $\tilde{g}$ is strictly increasing on $L(\mathbf{v}_1, \eta)$ for increasing η. Hence, for $\eta \geq 0$ fixed, the minimum of $\tilde{h}$ occurs at the level $\rho(\eta) = \min \{ \rho \geq 0, \ I\!R^p \times B(\mathbf{0}, \rho) \cap L(\mathbf{v}_1, \eta) \neq \emptyset \}$. Note that $L(\mathbf{v}_1, \eta)$ determines a strictly convex body, and that $I\!R \times B(0, \rho(\eta)) \cap L(\mathbf{v}_1, \eta)$ consists of exactly one point. The minima of $\tilde{h}$ occur on the curve $\sim \Phi$ in V, which is given by these points for varying η.

We proceed to compute this curve using Lagrange multipliers (e.g., Protter and Murrey, 1977, Chapter 14.4). The problem is the following: minimize

$\mathbf{s}'\mathbf{s}$ subject to $(\mathbf{v}_1 - \mathbf{v})'\mathbf{N}'\mathbf{N}(\mathbf{v}_1 - \mathbf{v}) = \eta$, where $\mathbf{v} = \begin{pmatrix} \beta \\ \mathbf{s} \end{pmatrix} \in V$. Denote $\mathbf{v}_1 = \begin{pmatrix} \beta_1 \\ \mathbf{s}_1 \end{pmatrix}$.

Necessary conditions are

$$
\begin{aligned}
(6.5) \qquad \mathbf{0} &= \mathbf{X}'\mathbf{X}(\beta_1 - \beta) + \mathbf{X}'\mathbf{Z}_s(\mathbf{s}_1 - \mathbf{s}) \\
(6.6) \qquad \mathbf{s} &= -\lambda\left[\mathbf{Z}_s'\mathbf{X}(\beta_1 - \beta) + \mathbf{Z}_s'\mathbf{Z}_s(\mathbf{s}_1 - \mathbf{s})\right], \\
(6.7) \qquad \eta &= (\mathbf{v}_1 - \mathbf{v})'\mathbf{N}'\mathbf{N}(\mathbf{v}_1 - \mathbf{v}), \\
(6.8) \qquad \rho &= \mathbf{s}'\mathbf{s}.
\end{aligned}
$$

This system of equations has exactly 2 solutions, and a unique one, when considering $\rho(\eta)$.

Let us first compute the point where $\mathbf{s}'\mathbf{s} = 0$. It follows from (6.5)-(6.8) that in this case, we obtain

$$(6.9) \qquad \beta_0 = \beta_1 + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}_s\mathbf{s}_1 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

(Note that $\beta_0 = \beta_1$ iff $\mathbf{X}'\mathbf{Z}_s\mathbf{s}_1 = \mathbf{0}$ iff $\mathbf{X} \perp \mathbf{Z}\mathbf{s}_1$.) We have to compute the curve $\sim\boldsymbol{\Phi}$ between

$$(6.10) \qquad \mathbf{v}_1 = \begin{pmatrix} \beta_1 \\ \mathbf{s}_1 \end{pmatrix} \text{ and } \mathbf{v}_0 = \begin{pmatrix} \beta_0 \\ \mathbf{s}_0 \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \mathbf{0} \end{pmatrix},$$

where the solution to (6.5)-(6.8) is unique. Plugging (6.5) into (6.6) yields

$$\mathbf{s} = -\lambda\left[\mathbf{Z}_s'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}_s(\mathbf{s}_1 - \mathbf{s}) + \mathbf{Z}_s'\mathbf{Z}_s(\mathbf{s}_1 - \mathbf{s})\right]$$

$$(6.11) \qquad = -\lambda\left[\mathbf{Z}_s'\left(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)\mathbf{Z}_s(\mathbf{s}_1 - \mathbf{s})\right].$$

Here, $\mathbf{I} - \mathbf{P} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is the projection matrix that projects the solution in V onto that in V_s. Note that $\mathbf{I} - \mathbf{P}$ is symmetric and idempotent (i.e. $\mathbf{I} - \mathbf{P}$ is its square root), and therefore, (6.11) says that $\mathbf{s}$ is the solution component in V_s if $\mathbf{s}$ is perpendicular to the tangent hyperplane of $K(\mathbf{s}_1, \kappa) = \left\{\mathbf{s} \in \mathbf{V}_s; (\mathbf{s}_1 - \mathbf{s})'\mathbf{Z}_s'(\mathbf{I} - \mathbf{P})\mathbf{Z}_s(\mathbf{s}_1 - \mathbf{s}) = \kappa\right\}$ at the point $\mathbf{s}_1 - \mathbf{s}$. Hence, for $\kappa \in [\mathbf{0}, \bar{\kappa}]$, ρ is a convex function of κ, where $\bar{\kappa} = (\mathbf{s}_1'\mathbf{s}_1)$. (Note that ρ is a linear function of κ iff one of the (generalized) eigendirections of $\mathbf{Z}_s'(\mathbf{I} - \mathbf{P})'(\mathbf{I} - \mathbf{P})\mathbf{Z}_s$ is $\mathbf{s}_1$.)

Now we obtain from (6.7),

$$
\begin{aligned}
\eta &= (\mathbf{v}_1 - \mathbf{v})' \, \mathbf{N}' \mathbf{N} \, (\mathbf{v}_1 - \mathbf{v}) \\
&= (\mathbf{s}_1 - \mathbf{s})' \, \mathbf{Z}'_s \mathbf{X} \, (\mathbf{X}' \mathbf{X})^{-1} \, \mathbf{X}' \mathbf{Z}_s \, (\mathbf{s}_1 - \mathbf{s}) + (\mathbf{s}_1 - \mathbf{s})' \, \mathbf{Z}'_s \mathbf{Z}_s \, (\mathbf{s}_1 - \mathbf{s}) \\
&= (\mathbf{s}_1 - \mathbf{s})' \, \mathbf{Z}'_s \left(\mathbf{I} + \mathbf{X} \, (\mathbf{X}' \mathbf{X})^{-1} \, \mathbf{X}' \right) \mathbf{Z}_s \, (\mathbf{s}_1 - \mathbf{s}) \\
&= (\mathbf{s}_1 - \mathbf{s})' \, \mathbf{Z}'_s \left(\mathbf{I} - \mathbf{X} \, (\mathbf{X}' \mathbf{X})^{-1} \, \mathbf{X}' \right)^2 \mathbf{Z}_s \, (\mathbf{s}_1 - \mathbf{s}) + 2 \, (\mathbf{s}_1 - \mathbf{s})' \\
&\qquad \mathbf{Z}'_s \mathbf{X} \, (\mathbf{X}' \mathbf{X})^{-1} \, \mathbf{X}' \mathbf{Z}_s \, (\mathbf{s}_1 - \mathbf{s})
\end{aligned}
$$

$$
\tag{6.12} = \kappa + 2 \, (\beta_1 - \beta)' \, \mathbf{X}' \mathbf{X} \, (\boldsymbol{\beta}_1 - \beta),
$$

and, therefore, η is a convex function of κ.

Consider the solution $\sim\Phi$ of (6.5)-(6.8) in V, parameterized by $\kappa \in [0, \bar{\kappa}]$. On this curve, ρ and η are convex functions of κ, and so are $\tilde{f}$ and $\tilde{g}$, because $l, m \geq 1$. Note that, by the arguments above, the β-component of $\sim\Phi$ is a straight line between β_1 and β_0. Hence $\sim\Phi$ is a straight line in V iff one of the (generalized) eigendirections of $\mathbf{Z}'_s \, (\mathbf{I} - \mathbf{P})' \, (\mathbf{I} - \mathbf{P}) \, \mathbf{Z}_s$ is $\mathbf{s}_1$: This case has been considered in more detail in the second part of Section 3. We have completed the proof for the case where $\mathbf{N} = \mathbf{M}$, i.e., all columns of $\mathbf{X}$ and $\mathbf{Z}$ are linearly independent.

We now extend the result to $I\!\!R^p \times I\!\!R^q$. Define the functions
$$
f \;:\; I\!\!R^p \times I\!\!R^q \to I\!\!R, \; f(\boldsymbol{\beta}, \mathbf{s}) = (\mathbf{s}' \mathbf{s} + \mathbf{c}_2)^l
$$

$$
g : I\!\!R^p \times I\!\!R^q \to I\!\!R, \; g(\boldsymbol{\beta}, \mathbf{s}) = \left[\left(\mathbf{y} - \mathbf{M} \begin{pmatrix} \beta \\ \mathbf{s} \end{pmatrix} \right)' \left(\mathbf{y} - \mathbf{M} \begin{pmatrix} \beta \\ \mathbf{s} \end{pmatrix} \right) + c_1 \right]^m .
$$

(6.13)

For $(\beta, \mathbf{s}) \in I\!\!R^p \times I\!\!R^q$, we have

$$
\tag{6.14} (\beta, \mathbf{s}) = \sum_{i=1}^{p} \beta_i e_i + \sum_{j=1}^{r} s_j \hat{e}_j + k = \beta + \mathbf{s}_v + k.
$$

For $\kappa \in [0, \bar{\kappa}]$, consider the affine subspace $\sim\Phi(\kappa) + K \subset I\!\!R^p \times I\!\!R^q$, where Φ is the solution curve of (6.5)-(6.8), and K is the kernel of $\mathbf{M}$. We have for all $k \in K$

$$
\tag{6.15} g(\beta, \mathbf{s}_v, \mathbf{k}) = \tilde{g}(\beta, \mathbf{s}_v).
$$

Hence, we need to find the points $\Phi(\kappa) \in \sim\Phi(\kappa) + K$, where f attains its minimum on $\Phi(\kappa) + K$. Since $f(\beta, \mathbf{s}) = f(\mathbf{0}, \mathbf{s})$ for all $\mathbf{s} \in \mathbb{R}^q$, we can restrict our attention to $\mathbb{R}^q$, i.e., to the $\mathbf{s}$-component of $\sim\Phi(\kappa) + K$, denoted by $(\sim\Phi(\kappa) + K)_s$. Since $\mathbf{s}'\mathbf{s}$ is the Euclidean norm in $\mathbb{R}^q$, the unique minimum is attained at $\Phi(\kappa)_s$ with the property that $\Phi(\kappa)_s \perp (\sim\Phi(\kappa) + K)_s$. Hence, $\Phi(\kappa)_s$ lies in $K^\perp$, the orthogonal complement of K, and is the projection of $\sim\Phi(\kappa)_s$ onto $K^\perp$ along K. But $K^\perp$ is spanned by the rows of $\mathbf{M}$, i.e., in $\mathbb{R}^q$ by the rows of $\mathbf{Z}$. In order to arrive at the corresponding normal equations, denote by $\hat{\mathbf{Z}}$ a $q \times r$ matrix, whose columns are r linearly independent rows of $\mathbf{Z}$. Then $\boldsymbol{\Phi}(\kappa)_s$ is given by

$$(6.16) \qquad \hat{\mathbf{Z}}'\hat{\mathbf{Z}}\Phi(\kappa)_s = \hat{\mathbf{Z}}'\sim\Phi(\kappa)_s,$$

or

$$(6.17) \qquad \Phi(\kappa)_s = \left(\hat{\mathbf{Z}}'\hat{\mathbf{Z}}\right)^{-1}\hat{\mathbf{Z}}'\sim\Phi(\kappa)_s.$$

Now, let $\{k_j, j = 1, ..., q - r\}$ be a basis in K. Since $\mathbf{X}$ has full column rank, none of the k_j is contained in $\mathbb{R}^p$. The point $\Phi(\kappa)$ satisfies the system of linear equations

$$(6.18) \qquad \Phi(\kappa) = \begin{pmatrix} \Phi(\kappa)_\beta \\ \Phi(\kappa)_s \end{pmatrix} = \sim\Phi(\kappa) + \sum_{j=1}^{q-r} \alpha_j k_j$$

for some coefficients $\alpha_j \in \mathbb{R}$. This system has a unique solution $(\alpha_1, ..., \alpha_{q-r})$, because the k_j are linearly independent. In fact, $\Phi(\kappa)$ is a curve in $\mathbb{R}^p \times \mathbb{R}^q$, with $\Phi(\bar{\kappa}) = (\boldsymbol{\beta}_0, \mathbf{0})$ and $\Phi(0) = \mathbf{w}_1$, the point corresponding to $\mathbf{v}_1 \in V$.

On the curve $\Phi(\kappa)$, the functions f and g have the following values

$$(6.19) \qquad g(\Phi(\kappa)) = \tilde{g}(\Phi(\kappa)),$$

$$(6.20) \qquad f(\Phi(\kappa)) = \left(\xi^2 \sim\Phi(\kappa)_s' \sim\Phi(\kappa)_s + c_2\right)^l,$$

where $\xi^2 > 0$ is a constant, given e.g. by $\boldsymbol{\Phi}(0)_s' \Phi(0)_s / \sim\Phi(0)_s' \sim\Phi(0)_s$. Therefore, in the same parameterization as above, f and g are convex on Φ, and all minima occur on the segment $\{\Phi(\kappa), \kappa \in [0, \bar{\kappa}]\}$. This proves the theorem for $h(\beta, \mathbf{s})$ as defined in (6.1) on $\mathbb{R}^p \times \mathbb{R}^q$.