

A BAYESIAN MODEL AND GIBBS SAMPLER FOR HYPERSPECTRAL IMAGING

Gabriel A. Rodriguez-Yam,
Department of Statistics, Colorado State University
rodrigue@stat.colostate.edu

Richard A. Davis,
Department of Statistics, Colorado State University
rdavis@stat.colostate.edu

Louis L. Scharf,
Departments of Electrical and Computer Engineering, and Statistics, Colorado State University
scharf@engr.colostate.edu

ABSTRACT

In this ongoing work, we propose a Bayesian model that can be used to detect targets in multispectral images when the signals from the materials in the image mix linearly, the noise is Gaussian, and abundance parameters are *non-negative*. By using efficient implementations of the Gibbs sampler, the expectation of any measurable functional of the abundance parameters, relative to the posterior distribution, can be computed easily. This general approach can be used to include additional constraints.

1. INTRODUCTION

In the linear mixing model, the spectrum of a mixed pixel is represented as a linear combination of component spectra, i.e.,

$$\mathbf{y} = \mathbf{M}\mathbf{a} + \mathbf{n}, \quad (1)$$

where \mathbf{M} is an $N \times r$ matrix whose columns correspond to the spectra of the r materials present in the pixel, \mathbf{a} is an $r \times 1$ vector consisting of the abundances of the materials in a pixel, and \mathbf{n} is an $N \times 1$ vector corresponding to the noise [10], [14]. Due to physical considerations the components of the vector \mathbf{a} are considered to be non-negative. These constraints can be expressed as $\mathbf{a} \in \mathbf{T} := [0, \infty)^r = [0, \infty) \times \dots \times [0, \infty)$. The matrix \mathbf{M} is assumed to be of full rank and the noise in the model is assumed to be Gaussian, i.e. $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{s}^2 \mathbf{I}_N)$.

In section 2 we list standard results for the truncated multivariate distribution that are used in this work and we prove the main result of this work. In section 3 a

Bayesian framework for the model in (1) is described. In section 4 we briefly review the basic idea behind the Gibbs sampler. In section 5 we propose various implementations of this technique for the model and discuss our findings.

2. TRUNCATED MULTIVARIATE NORMAL DISTRIBUTION

Definition 1. (*Truncated Multivariate Normal Distn*). Let $\mathbf{f}(\mathbf{x}; \mathbf{m}, \mathbf{S})$ denote a k -variate normal density with mean \mathbf{m} and covariances \mathbf{S} . Then $X \sim N_{\mathbf{R}}(\mathbf{m}, \mathbf{S})$ denotes a random vector whose density is proportional to $\mathbf{f}(\mathbf{x}; \mathbf{m}, \mathbf{S})\mathbf{I}_{\mathbf{R}}(\mathbf{x})$, where $\mathbf{R} \subset \hat{\mathbf{A}}^k$ has positive Lebesgue measure and $\mathbf{I}_{\mathbf{R}}(\cdot)$ is the indicator function for \mathbf{R} . We call $N_{\mathbf{R}}$ a truncated normal distribution.

Result 1. Let T be a subset of $\hat{\mathbf{A}}^k$ that has positive Lebesgue measure. Suppose $X \sim N_T(\mathbf{m}, \mathbf{S})$. Define $0 < k_1 < k$, and partition \mathbf{m} and \mathbf{S} as

$$\mathbf{m} = \begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{bmatrix},$$

where \mathbf{m}_1 is a vector with k_1 components and \mathbf{S}_{11} is a $k_1 \times k_1$ -matrix. Then the generalized Gauss-Markov theorem for the truncated normal distribution N_T is

$$\text{i) } p_X(x_1, x_2) \propto \mathbf{f}(x_1; \mathbf{m}_1, \mathbf{S}_{11}) \mathbf{f}(x_2; \mathbf{m}_2^*, \mathbf{S}_{22}^*) \mathbf{I}_T(x),$$

where

$$\mathbf{m}_2^* = \mathbf{m}_2 + \mathbf{S}_{21}(\mathbf{S}_{11})^{-1}(x_1 - \mathbf{m}_1),$$

$$\mathbf{S}_{22}^* = \mathbf{S}_{21}(\mathbf{S}_{11})^{-1}\mathbf{S}_{12}.$$

ii) If $T = T_1 \times T_2$, where T_1 and T_2 are subsets of $\hat{\mathbf{A}}^{k_1}$ and $\hat{\mathbf{A}}^{k_2}$, respectively, with positive Lebesgue measure, then

$$X_2 | X_1 \sim N_{T_2}(\mathbf{m}_2^*, \mathbf{S}_{22}^*) \quad \blacksquare$$

This work was supported by Colorado Advanced Software Institute (CASI), Data Fusion Corporation and Consejo Nacional de Ciencia y Tecnologia (CONACYT).

Let \mathbf{Z} be a vector of order k , i.e. $\mathbf{Z} = [Z_1, \dots, Z_k]^T$. We define $\mathbf{Z}_{[j:l]} := [Z_j, \dots, Z_l]^T$. Also, given a square matrix $\mathbf{A} = \{a_{ij}\}$ of order k , $\mathbf{A}_{[j:l, m:n]}$ denotes the submatrix of \mathbf{A} containing the rows j thru l and columns m thru n . When $j=l$ or $m=n$ we can write $\mathbf{A}_{[j, m:n]}$ to denote a subrow and $\mathbf{A}_{[j:l, m]}$ to denote a subcolumn, respectively. When $j=l$ and $m=n$ then $\mathbf{A}_{[j:l, m:n]} = \mathbf{A}_{[j, m]} = a_{jm}$.

The following result establishes the action of linear transformations on truncated normal distributions. This result is crucial to implement an efficient Gibbs sampler for the Bayesian linear mixing model proposed in this work.

Result 2. Suppose $\mathbf{X} \sim N_T(\mathbf{m}, \mathbf{s}^2 \mathbf{S})$, where $T \subseteq \mathfrak{R}^k$ has positive Lebesgue measure, and \mathbf{S} is positive definite. Moreover, let \mathbf{A} be a real $k \times k$ matrix of full rank and define $\mathbf{a} := \mathbf{A}\mathbf{m}$ and $\mathbf{Y} := \mathbf{A}\mathbf{X}$, and $S := \{\mathbf{A}\mathbf{x} : \mathbf{x} \in T\}$. Now, let $S_j(y_1, \dots, y_{j-1}, y_{j+1}, \dots, y_k)$ denote the subset of \mathfrak{R} , defined by

$$S_j(y_1, \dots, y_{j-1}, y_{j+1}, \dots, y_k) := \{y_j : (y_1, \dots, y_k) \in S\}. \quad (2)$$

Then

- (a) $\mathbf{Y} \sim N_S(\mathbf{a}, \mathbf{s}^2 \mathbf{A}\mathbf{S}\mathbf{A}^T)$.
(b) If in particular \mathbf{A} is a lower triangular matrix such that $\mathbf{A}\mathbf{S}\mathbf{A}^T = \mathbf{I}_k$, then writing $\mathbf{a} = [\mathbf{a}_1, \dots, \mathbf{a}_k]^T$,

$$y_j | (y_1, \dots, y_{j-1}, y_{j+1}, \dots, y_k) \sim N_{S_j}(\mathbf{a}_j, \mathbf{s}^2), \quad (3)$$

provided that $S_j := S_j(y_1, \dots, y_{j-1}, y_{j+1}, \dots, y_k)$ has positive Lebesgue measure.

- (c) If \mathbf{A} and S_j are as in (b) and $T := [0, \infty)^k$, then

$$S_j = \{a_j + a_{jz} : \mathbf{b}_{jz} \leq \mathbf{c}_j, z \geq 0\}, \quad (4)$$

where

$$a_j := \mathbf{A}_{[j, 1:j-1]} \mathbf{x}_{[1:j-1]} \quad (5)$$

$$\mathbf{b}_j := (\mathbf{A}_{[j+1:k, j+1:k]})^{-1} \mathbf{A}_{[j+1:k, j]} \quad (6)$$

$$\mathbf{c}_j := \mathbf{r}_j - (\mathbf{A}_{[j+1:k, j+1:k]})^{-1} \mathbf{A}_{[j+1:k, 1:j-1]} \mathbf{x}_{[1:j-1]} \quad (7)$$

$$\mathbf{r}_j := (\mathbf{A}_{[j+1:k, j+1:k]})^{-1} \mathbf{y}_{[j+1:k]} \quad (8)$$

$$\mathbf{x}_{[1:j-1]} := (\mathbf{A}_{[1:j-1, 1:j-1]})^{-1} \mathbf{y}_{[1:j-1]}. \quad (8)$$

For the particular cases $j=1$ and $j=k$ we take

$$a_1 := 0 \quad (9)$$

$$\mathbf{c}_1 := (\mathbf{A}_{[2:k, 2:k]})^{-1} \mathbf{y}_{[2:k]} \quad (10)$$

$$S_k(y_1, \dots, y_{k-1}) := \{a_k + a_{kk}z : z \geq 0\}. \quad (11)$$

Proof. Only (c) needs to be shown. Note that S has the form

$$S = \{\mathbf{y} \in \hat{\mathbf{A}}^k : \mathbf{y} = \mathbf{A}\mathbf{x}, \mathbf{x} \in T\}. \quad (12)$$

For $j \in \{2, \dots, k-1\}$, partition \mathbf{x} and \mathbf{y} , respectively, as

$$[\mathbf{x}_{[1:j-1]}^T, x_j, \mathbf{x}_{[j+1:k]}^T]^T, [\mathbf{y}_{[1:j-1]}^T, y_j, \mathbf{y}_{[j+1:k]}^T]^T,$$

and \mathbf{A} as

$$\begin{bmatrix} \mathbf{A}_{[1:j-1, 1:j-1]} & 0 & 0 \\ \mathbf{A}_{[j, 1:j-1]} & a_{jj} & 0 \\ \mathbf{A}_{[j+1:k, 1:j-1]} & \mathbf{A}_{[j+1:k, j]} & \mathbf{A}_{[j+1:k, j+1:k]} \end{bmatrix}.$$

Then from $\mathbf{y} = \mathbf{A}\mathbf{x}$ we obtain

$$\mathbf{y}_{[1:j-1]} = \mathbf{A}_{[1:j-1, 1:j-1]} \mathbf{x}_{[1:j-1]} \quad (13)$$

$$y_j = \mathbf{A}_{[j, 1:j-1]} \mathbf{x}_{[1:j-1]} + a_{jj} x_j \quad (14)$$

$$\mathbf{x}_{[j+1:k]} = (\mathbf{A}_{[j+1:k, j+1:k]})^{-1} (\mathbf{y}_{[j+1:k]} - \mathbf{A}_{[j+1:k, 1:j-1]} \mathbf{x}_{[1:j-1]} - \mathbf{A}_{[j+1:k, j]} x_j) \quad (15)$$

Then (8) follows from (13). Now, taking a_j , \mathbf{b}_j and \mathbf{c}_j as in (5)-(7), respectively, (14) and (15) become

$$y_j = a_j + a_{jj} x_j \quad (16)$$

$$\mathbf{x}_{[j+1:k]} = \mathbf{c}_j - \mathbf{b}_j x_j \quad (17)$$

Thus, (4) follows using (16), (17) and (12) in (2). When $j=1$ or $j=k$, partitioning appropriately the system $\mathbf{y} = \mathbf{A}\mathbf{x}$, and repeating the previous procedure, produces equations (9) to (11). ■

Example 1. Consider the case $\mathbf{A}\mathbf{S}\mathbf{A}^T = \mathbf{I}$, where

$$\mathbf{S} = \begin{bmatrix} 1 & 4/5 \\ 4/5 & 1 \end{bmatrix} \text{ and } \mathbf{A} = \begin{bmatrix} 1 & 0 \\ -4/3 & 5/3 \end{bmatrix}.$$

Thus, if $\mathbf{X} \sim N_T(\mathbf{m}, \mathbf{s}^2 \mathbf{S})$, and $\mathbf{Y} = \mathbf{A}\mathbf{X}$, where $T = [0, \infty)^2$ then from (a) of Result 2, $\mathbf{Y} \sim N_S(\mathbf{a}, \mathbf{s}^2 \mathbf{I})$ where $S = \{\mathbf{A}\mathbf{x} : \mathbf{x} \in T\}$ and $\mathbf{a} := \mathbf{A}\mathbf{m} = [\mathbf{m}, -4\mathbf{m}/3 + 5\mathbf{m}/3]^T$. Moreover, from (3),

$$y_1 | y_2 \sim N_{S_1}(\mathbf{m}, \mathbf{s}^2),$$

$$y_2 | y_1 \sim N_{S_2}(-\frac{4}{3}\mathbf{m} + \frac{5}{3}\mathbf{m}, \mathbf{s}^2),$$

where $S_1 := S_1(y_2)$ and $S_2 := S_2(y_1)$ can be found using (c) of Result 2 as follows: from (6), (9) and (10) we obtain $b_1 = -4/5$, $a_1 = 0$, and $c_1 = 3y_2/5$. Thus, from (4), $S_1 = [\max\{0, -3y_2/4\}, \infty)$. Now, using (8), (5) and (11) we find that $\mathbf{x}_{[1:1]} = y_1$, $a_2 = (-4/3)y_1$ and $S_2 = [-4y_1/3, \infty)$. The regions T and S for this problem are shown in Figure 1. Although S_1 and S_2 could have been obtained directly from this figure, in general for higher dimensions the calculation of S_j must be coded as in Result 2.

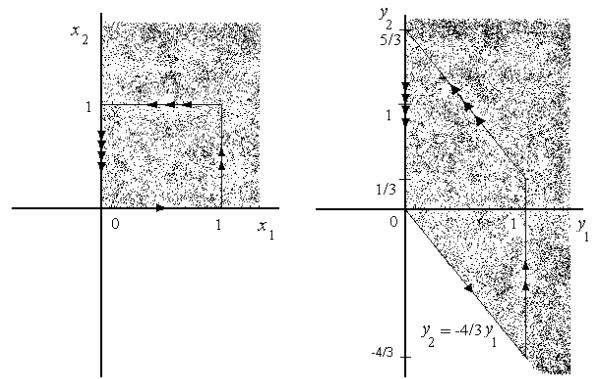


Figure 1. Regions T and $S = \{\mathbf{A}\mathbf{x} : \mathbf{x} \in T\}$. ■

For $T := [0, \infty)^k$, Result 2 suggests how to implement an efficient Gibbs sampler for a truncated multivariate normal distribution. First we shall transform as in (b). Then (3) gives the distributions to be sampled in each cycle of the Gibbs sampler (see Section 4). This

implementation gives a chain that solves the problem of slow mixing that may appear in the standard Gibbs sampler implementation for the truncated multivariate normal distribution [5] with T of the above form. For a general convex region such as $T := \{\mathbf{x} \in \hat{\mathbf{A}}^k: \mathbf{a} \leq \mathbf{B}\mathbf{x} \leq \mathbf{b}\}$, where \mathbf{B} is a full rank matrix of dimension $m \times k$ ($m \leq k$) and the components of \mathbf{a} and \mathbf{b} are allowed to be $-\infty$ and $+\infty$, respectively, a procedure as in (c) of Result 2 can be obtained with only minor adjustments.

3. MODEL

The model in (1) describes the conditional distribution of y given the parameter vector $\mathbf{q} := (\mathbf{a}, \mathbf{s}^2)$ consisting of abundances \mathbf{a} and noise variance \mathbf{s}^2 . Now we assume the prior distribution for \mathbf{q} to be

$$\begin{aligned} p(\mathbf{a}, \mathbf{s}^2) &= p(\mathbf{a}) p(\mathbf{s}^2), \\ \mathbf{a} &\sim N_T(\mathbf{m}, \mathbf{s}_0^2 (\mathbf{M}^T \mathbf{M})^{-1}), \\ \mathbf{s}^2 &\sim IG(\mathbf{n}, \mathbf{I}), \end{aligned} \quad (18)$$

where $T := [0, \infty)^k$ and IG is the inverse gamma distribution (see [15]). We assume \mathbf{s}_0^2 , \mathbf{n} and \mathbf{I} are known positive scalars, and \mathbf{m} is a known vector. Let $p(\mathbf{a}, \mathbf{s}^2 | D)$ denote the posterior distribution of \mathbf{q} given the pixel values D . Then

$$p(\mathbf{a}, \mathbf{s}^2 | D) \propto L(y; \mathbf{a}, \mathbf{s}^2) p(\mathbf{a}) p(\mathbf{s}^2), \quad (19)$$

where L is the likelihood of \mathbf{q} given y from the model (1), when $\mathbf{n} \sim N(\mathbf{0}, \mathbf{s}^2 \mathbf{I}_N)$.

The problem addressed in this work is to obtain a sample $\mathbf{q}_1, \dots, \mathbf{q}_t$ from the posterior density $p(\mathbf{q} | D)$, for which functionals of the posterior can be estimated. Such functionals can be expressed as

$$E[h(\mathbf{q}) | D] = \int h(\mathbf{q}) p(\mathbf{q} | D) d\mathbf{q}, \quad (20)$$

where $h(\cdot)$ is a real-valued function of \mathbf{q} that is integrable relative to $p(\mathbf{q} | D)$. Notice that (20) reduces to

- (a) the posterior mean of \mathbf{q} when $h(\mathbf{q}) = \mathbf{q}$.
- (b) the posterior covariance of \mathbf{q} and \mathbf{q} when $h(\mathbf{q}) = (\mathbf{q} - E(\mathbf{q} | D))(\mathbf{q} - E(\mathbf{q} | D))^T$.
- (c) the posterior probability of a set A if $h(\mathbf{q}) = \mathbf{I}_A(\mathbf{q})$.

4. GIBBS SAMPLER

To draw a sample from the posterior distribution $p(\mathbf{a}, \mathbf{s}^2 | D)$ in (19) we consider the Gibbs sampler, an example of a Markov Chain Monte Carlo (MCMC) simulation. The latter originated with the algorithm of [11] and immediately found widespread use in Physics and Chemistry. This algorithm, known as the Metropolis algorithm, was generalized in [9], but this work did not find extensive use in statistics until the appearance of the Gibbs sampler introduced by [3] in the context of image

restoration. The basic idea is to simulate a stationary ergodic Markov chain, whose stationary distribution has the desired posterior density.

To implement the Gibbs sampler, write the components of the vector \mathbf{q} as $(\mathbf{q}_1, \dots, \mathbf{q}_t)$ where the \mathbf{q}_i 's are either uni- or multidimensional. Assuming that we can simulate a random value from the conditional density functions $p(\mathbf{q}_i | \mathbf{q}_1, \dots, \mathbf{q}_{i-1}, \mathbf{q}_{i+1}, \dots, \mathbf{q}_t, D)$, then the basic scheme of the Gibbs sampler is given as follows

Step 0. Choose an arbitrary starting point $\mathbf{q}_0 = (\mathbf{q}_{1,0}, \dots, \mathbf{q}_{t,0})$ in the support of $p(\mathbf{q} | D)$ and set $t = 0$.

Step 1. Generate $\mathbf{q}_{t+1} = (\mathbf{q}_{1,t+1}, \dots, \mathbf{q}_{t,t+1})$ as follows:

- Generate $\mathbf{q}_{1,t+1}$ from $p(\mathbf{q}_1 | \mathbf{q}_{2,t}, \dots, \mathbf{q}_{t,t}, D)$.
- Generate $\mathbf{q}_{2,t+1}$ from $p(\mathbf{q}_2 | \mathbf{q}_{1,t+1}, \mathbf{q}_{3,t}, \dots, \mathbf{q}_{t,t}, D)$.
- \vdots
- Generate $\mathbf{q}_{t,t+1}$ from $p(\mathbf{q}_t | \mathbf{q}_{1,t+1}, \mathbf{q}_{2,t+1}, \dots, \mathbf{q}_{t-1,t+1}, D)$.

Step 2. Set $t = t + 1$ and go to step 1.

Notice that each component of the vector \mathbf{q} is updated in the natural order, using most recent updates of all other components of \mathbf{q} and that a cycle in the scheme requires the generation of q random quantities. Under certain regularity conditions (see [2] for example), the Markov chain $\{\mathbf{q}_0, \mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \dots\}$ has a stationary distribution which is the posterior density $p(\mathbf{q} | D)$.

5. GIBBS SAMPLER IMPLEMENTATIONS

Using (18) in (19) we find the following posterior conditional densities

$$\mathbf{a} | (\mathbf{s}^2, D) \sim N_T(\mathbf{m}, \mathbf{S}_1), \quad (21)$$

$$\mathbf{s}^2 | (\mathbf{a}, D) \sim (SS(\mathbf{a}) + 2\mathbf{I})^{-1} \mathbf{c}_{N+2\mathbf{n}}^2, \quad (22)$$

where $\mathbf{c}_{N+2\mathbf{n}}^2$ denotes a chi-squared distribution with $N + 2\mathbf{n}$ degrees of freedom, and

$$\mathbf{m} = \frac{\mathbf{s}_0^2}{\mathbf{s}_0^2 + \mathbf{s}^2} \hat{\mathbf{a}} + \frac{\mathbf{s}^2}{\mathbf{s}_0^2 + \mathbf{s}^2} \mathbf{m}_0,$$

$$\mathbf{S}_1 = \frac{\mathbf{s}_0^2 \mathbf{s}^2}{\mathbf{s}_0^2 + \mathbf{s}^2} (\mathbf{M}^T \mathbf{M})^{-1},$$

$$SS(\mathbf{a}) := (\mathbf{y} - \mathbf{M}\mathbf{a})^T (\mathbf{y} - \mathbf{M}\mathbf{a}).$$

Here $\hat{\mathbf{a}}$ is the ordinary least squares estimates of \mathbf{a} from the model in (1). Notice that \mathbf{m} is a convex lineal combination of prior \mathbf{m}_0 and $\hat{\mathbf{a}}$ and $SS(\mathbf{a})$ is the residual sum of squares; also \mathbf{S}_1 is the least squares covariance matrix, scaled by the factor for $\hat{\mathbf{a}}$ in \mathbf{m} .

5.1. Full conditionals (scalar Gibbs)

As a first implementation of the Gibbs sampler for the posterior density $p(\mathbf{a}, \mathbf{s}^2 | D)$, we consider the set of unidimensional conditional distributions. Hence, to

update the current value $(\mathbf{a}_i, \mathbf{s}_i^2) = (\mathbf{a}_{1,i}, \dots, \mathbf{a}_{r,i}, \mathbf{s}_i^2)$ of the i -th iteration we proceed as follows:

- draw $\mathbf{a}_{1,i+1}$ from $p(\mathbf{a}_1 | \mathbf{a}_{2,i}, \dots, \mathbf{a}_{r,i}, \mathbf{s}_i^2, D)$,
- draw $\mathbf{a}_{2,i+1}$ from $p(\mathbf{a}_2 | \mathbf{a}_{1,i+1}, \mathbf{a}_{3,i}, \dots, \mathbf{a}_{r,i}, \mathbf{s}_i^2, D)$,
- \vdots
- draw $\mathbf{a}_{r,i+1}$ from $p(\mathbf{a}_r | \mathbf{a}_{1,i+1}, \mathbf{a}_{2,i+1}, \dots, \mathbf{a}_{r-1,i}, \mathbf{s}_i^2, D)$,
- draw \mathbf{s}_{i+1}^2 from $p(\mathbf{s}^2 | \mathbf{a}_{i+1}, D)$.

The conditional distribution $p(\mathbf{s}^2 | \mathbf{a}_{i+1}, D)$ is given in (22) while $\mathbf{a}_j | (\mathbf{a}_{1,i+1}, \dots, \mathbf{a}_{j-1,i+1}, \mathbf{a}_{j+1,i}, \dots, \mathbf{a}_{r,i}, \mathbf{s}_i^2, D)$ are truncated normal, as (21) and (ii) of Result 1 show.

With this standard implementation of the Gibbs sampler the MC values may not properly mix (i.e., the chain does not move rapidly through the "entire" support of the posterior distribution). This problem is particularly acute when the abundance parameters are highly correlated in the stationary distribution [6], [12].

5.2. Grouping (vector Gibbs)

Blocking highly correlated components into a higher-dimensional components may improve mixing [12]. By grouping the abundance parameters, i.e., write $\mathbf{q} = (\mathbf{q}_1, \mathbf{q}_2) = (\mathbf{a}, \mathbf{s}^2)$, the modified updating process of the Gibbs sampler becomes:

Given $\mathbf{q}_i := (\mathbf{a}_i, \mathbf{s}_i^2) = (\mathbf{a}_{1,i}, \dots, \mathbf{a}_{k,i}, \mathbf{s}_i^2)$, update \mathbf{q}_{i+1} in two steps:

- generate \mathbf{a}_{i+1} from $p(\mathbf{a} | \mathbf{s}_i^2, D)$,
- generate \mathbf{s}_{i+1}^2 from $p(\mathbf{s}^2 | \mathbf{a}_{i+1}, D)$,

where the required distributions are given in (21) and (22), respectively.

5.3. Reparameterization (transformed Gibbs)

A reparameterization may improve the mixing of the chain ([6]). Result 2 allows us to reparameterize the abundance parameters as follows. Let U be an upper triangular matrix for which $U^T U = (M^T M)^{-1}$. Denote $\mathbf{A} := U^{-T}$ and consider the transformation

$$\mathbf{h} := \mathbf{A} \mathbf{a}.$$

Let us suppose that we have already finished the i -th iteration of the Gibbs sampler, i.e., the last term of the current path is given by $(\mathbf{a}_{1,i}, \dots, \mathbf{a}_{r,i}, \mathbf{s}_i^2)$. From (21) we know that

$$\mathbf{a} | (\mathbf{s}_i^2, D) \sim N_T(\mathbf{m}, \mathbf{S}_i),$$

where

$$\begin{aligned} \mathbf{m} &= \mathbf{g} \hat{\mathbf{a}} + (1 - \mathbf{g}) \mathbf{m}, \\ \mathbf{S}_i &= \mathbf{s}_i^2 \mathbf{g} (M^T M)^{-1}, \\ \mathbf{g} &:= \mathbf{s}_0^2 / (\mathbf{s}_0^2 + \mathbf{s}_i^2). \end{aligned}$$

Using this and Result 2, it follows that

$$\mathbf{h} | (\mathbf{s}_i^2, D) \sim N_S(\mathbf{A} \mathbf{m}, \mathbf{s}_i^2 \mathbf{g} \mathbf{I}), \quad S = \{\mathbf{A} \mathbf{x} : \mathbf{x} \in T\} \quad (23)$$

Let $\mathbf{h}_i = (\mathbf{h}_{1,i}, \dots, \mathbf{h}_{r,i}) = \mathbf{A} \mathbf{a}_i$. We update $(\mathbf{a}_{1,i}, \dots, \mathbf{a}_{r,i}, \mathbf{s}_i^2)$ via $(\mathbf{h}_{1,i}, \dots, \mathbf{h}_{r,i}, \mathbf{s}_i^2)$ as follows: For $j = 1, \dots, r$,

- Draw $\mathbf{h}_{j,i+1}$ from $p(\mathbf{h}_j | (\mathbf{h}_{1,i+1}, \dots, \mathbf{h}_{j-1,i+1}, \mathbf{h}_{j+1,i}, \dots, \mathbf{h}_{r,i}, \mathbf{s}_i^2, D))$.
- Obtain $\mathbf{a}_{j,i+1} = (\mathbf{h}_{j,i+1} - a_j) / a_{jj}$ (see (16)).
- Draw \mathbf{s}_{i+1}^2 from $p(\mathbf{s}^2 | \mathbf{a}_{i+1}, D)$.

From Result 2 and (23), the posterior conditionals $p(\mathbf{h}_j | \mathbf{h}_{1,i+1}, \dots, \mathbf{h}_{j-1,i+1}, \mathbf{h}_{j+1,i}, \dots, \mathbf{h}_{r,i}, \mathbf{s}_i^2, Y)$, $j = 1, \dots, r$ are truncated normal. Also, $p(\mathbf{s}^2 | \mathbf{a}_{i+1}, Y)$ is given in (22).

6. NUMERICAL EXAMPLE

In a model with two variables and $n=20$, we set $\mathbf{a}_1=0.98$, $\mathbf{a}_2=0.02$, and $\mathbf{s}^2=1$. With simulated values from model (1), the least squares estimates of \mathbf{a} and \mathbf{s}^2 are $[1.0133, -0.0162]^T$ and 0.894, respectively. The constrained maximum likelihood estimates (CMLE) of these parameters are $[0.9983, 0]^T$ and 0.9951, respectively. Now, to define the prior in (18) we use $\mathbf{s}_0^2=0.9951$, $\mathbf{m} = [0.9983, 0]^T$ (the CMLE), $r=0.001$ and $\mathbf{I}=0.001$. Using this prior, we obtained Gibbs paths of length 1600 for each Gibbs sampler from section 5. The scalar and transformed Gibbs implementations in sections 5.1 and 5.3 need to draw from a univariate truncated normal. We followed the procedures described in [5] and [13]. The implementation of the Gibbs sampler in section 5.2 needs a procedure to draw from a multivariate truncated normal density. We followed the naive procedure, consisting in generating successively from the unrestricted normal until a value in the region of interest is obtained. Alternative procedures are the Accept-Reject method from [7] and the Geweke-Hajivassiliou-Keane simulator (GHK) [4], [8]. In Figure 2, the autocorrelation plots of the first component (\mathbf{a}_1) of the chain are shown for the scalar and transformed Gibbs samplers. The scalar Gibbs sampler was approximately two times faster than the Gibbs with transformation. However its draws are highly correlated. As commented in [1], slow decay in the autocorrelations of monitored functionals suggests slow mixing within a chain and usually slow convergence to the posterior distribution. Thus, following [1] we observe slow mixing within the scalar Gibbs chain and fast mixing within the transformed Gibbs. The corresponding plot obtained with vector Gibbs, which shows better mixing even than the transformed Gibbs is not shown. Even though the chain that results with the vector Gibbs implementation is "ideal", we warn that this implementation can be computationally expensive. The naive method becomes impractical when the conditional probability of the unconstrained abundance parameters, given noise variance, is small. We are unaware of the

existence of an efficient accept-reject procedure to simulate a truncated multivariate normal.

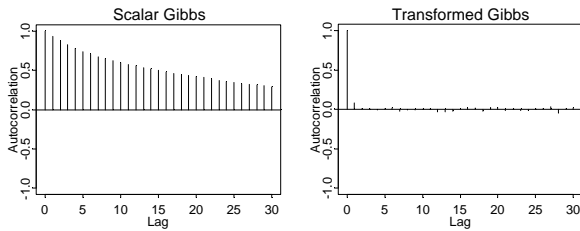


Figure 2. Autocorrelation plots of the monitored first component (\mathbf{a}_1) of the chains for the scalar and transformed Gibbs Samplers.

In Figure 3 we show the posterior density $p(\mathbf{a}_1|\mathbf{D})$ obtained using the transformed implementation of the Gibbs sampler and the prior $p(\alpha_1)$. We include also $p(\mathbf{a}_1|\mathbf{D})$ when no constraints are imposed on the abundance parameters.

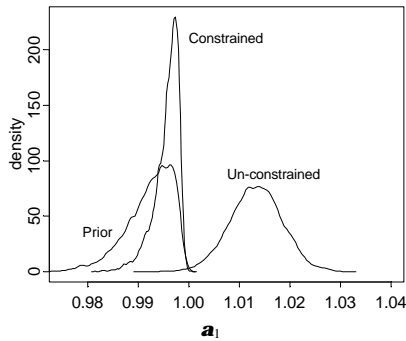


Figure 3. Constrained and unconstrained density $p(\mathbf{a}_1|\mathbf{D})$ and prior $p(\mathbf{a}_1)$.

7. CONCLUSIONS

We have derived three different Gibbs samplers for drawing non-negative multispectral abundances from the posterior distribution density of the abundances. Of these MCMC methods, the scalar Gibbs mixes poorly. While the vector and transformed Gibbs speed up the mixing, the former mixes better than the latter. Computationally, however, the scalar implementation is the cheapest, and the vector Gibbs is the most expensive. The Gibbs with transformation requires only little additional computational cost over that of scalar Gibbs.

8. REFERENCES

[1] Chen, M-H., Q-M. Shao, and J. G. Ibrahim, *Monte Carlo Methods in Bayesian Computation*, Springer, New York, 2000.
 [2] A. E. Gelfand, and A. F. M. Smith, "Sampling-based Approaches to Calculating Marginal Densities,"

Journal of the American Statistical Association, **85**, pp. 398-409, 1990.
 [3] S. Geman, and D. Geman, "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, pp. 721-741, 1984.
 [4] Geweke, J. *Efficient Simulation from The Multivariate Normal Distribution Subject to Linear Inequality Constraints and the Evaluation of Constraint Probabilities* (Mimeo), Duke University, Durham, NC, 1989.
 [5] J. Geweke, "Efficient Simulation From the Multivariate Normal and Student t-Distributions Subject to Linear Constraints," *Computer Sciences and Statistics: Proc. 23d Symp. Interface*, pp. 571-577, 1991.
 [6] W. R. Gilks, and G. O. Roberts, "Strategies for Improving MCMC," In *Markov Chain Monte Carlo in Practice* (eds W. R. Gilks, S. Richardson and D. J. Spiegelhalter), Chapman & Hall, London, pp. 89-114, 1995.
 [7] Hahivassiliou, V., and D. McFadden, *The Method of Simulated Scores for the Estimation of LDV Models*. Discussion Paper No. EM/97/328, School of Economics and Political Science, London, 1997.
 [8] V. Hahivassiliou, D. McFadden, and P. Ruud, "Simulation of Multivariate Normal Rectangle Probabilities and their Derivatives. Theoretical and Computational Results," *Journal of Econometrics*, **72**, pp. 85-134, 1996.
 [9] W. K. Hastings, "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika*, **57**, pp. 97-109, 1980.
 [10] D. Manolakis, and G. Shaw, "Detection Algorithms for Hyperspectral Imaging Applications," *IEEE Signal Processing magazine*, **19**, pp. 29-43, 2002.
 [11] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equations of State Calculations by Fast Computing Machines," *Journal of Chemical Physics*, **21**, pp. 1087-1091, 1953.
 [12] O'Hagan, A., *Kendall's Advanced Theory of Statistics 2B: Bayesian Inference*, Edward Arnold, London, 1994.
 [13] C. P. Robert, "Simulation of Truncated Normal Variables," *Statistics and Computing*, **5**, pp. 121-125, 1994.
 [14] G. Shaw, and D. Manolakis, "Signal Processing for Hyperspectral Image Exploitation," *IEEE Signal Processing magazine*, **19**, pp. 12-16, 2002.
 [15] Tanner, M. A., *Tools for Statistical Inference*, Springer, Berlin, 1996.