Collaborative sparse regression using spatially correlated supports – Application to hyperspectral unmixing

Yoann Altmann, Marcelo Pereyra and Jose Bioucas Dias

Abstract

This paper presents a new Bayesian collaborative sparse regression method for linear unmixing of hyperspectral images. Our contribution is twofold; first, we propose a new Bayesian model for structured sparse regression in which the supports of the sparse abundance vectors are a priori spatially correlated across pixels (i.e., materials are spatially organised rather than randomly distributed at a pixel level). This prior information is encoded in the model through a truncated multivariate Ising Markov random field, which also takes into consideration the facts that pixels cannot be empty (i.e., there is at least one material present in each pixel), and that different materials may exhibit different degrees of spatial regularity. Secondly, we propose an advanced Markov chain Monte Carlo algorithm to estimate the posterior probabilities that materials are present or absent in each pixel, and, conditionally to the maximum marginal a posteriori configuration of the support, compute the MMSE estimates of the abundance vectors. A remarkable property of this algorithm is that it self-adjusts the values of the parameters of the Markov random field, thus relieving practitioners from setting regularisation parameters by cross-validation. The performance of the proposed methodology is finally demonstrated through a series of experiments with synthetic and real data and comparisons with other algorithms from the literature.

This study was supported by the Direction Générale de l’armement, French Ministry of Defence, by the SuSTaIN program - EPSRC grant EP/D063485/1 - at the Department of Mathematics, University of Bristol, by the Portuguese Science and Technology Foundation, Projects PEst-OE/EEI/LA0008/2013 and PTDC/EIE-PRO/1470/2012.

Y. Altmann is with the School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh U.K. (email: Y.Altmann@hw.ac.uk).

M. Pereyra is with the School of Mathematics of the University of Bristol, Bristol U.K. (email: marcelo.pereyra@bristol.ac.uk).

J.M. Bioucas-Dias is with the Instituto de Telecomunicações and Instituto Superior Técnico, Universidade de Lisboa, Portugal (email: bioucas@lx.it.pt).
I. INTRODUCTION

Spectral unmixing (SU) of hyperspectral images is a challenging problem that has received a lot of attention over the last few years [1]–[3]. It consists in identifying the materials (endmembers) present in an image and simultaneously quantifying their fractions or proportions within each pixel (abundances). This source separation problem has been widely studied for applications where pixel reflectances are linear combinations of pure component spectra [4]–[8]. The typical SU processing pipeline is then decomposed into three main estimation steps: the estimation of the number of different materials present in the image, the estimation or extraction of their spectral signatures, and finally the quantification of their abundances.

Abundance estimation is known to be a challenging problem, in particular in scenarios involving materials with similar spectral signatures. In these cases, exploiting prior knowledge about the problem can improve estimation performance dramatically. A particularly important form of prior knowledge is that the number of materials within each pixel is typically much smaller than the total number of materials present in the scene (this property is accentuated in modern images that are acquired with high spatial resolution sensors). In other words, the abundance vectors are generally sparse. Sparsity also arises naturally when SU is performed with a dictionary or library containing the spectral signatures of a large number materials, some of which are possibly present in the scene [9].

Once sparsity is taken into consideration, SU can be conveniently formulated as a sparse regression problem whose objective is to jointly identify the materials within each pixel and to quantify their abundance. This regression problem is often solved by penalised maximum likelihood estimation, which can be efficiently computed with state-of-the-art optimisation algorithms (typically an $\ell_1$ penalty is used to promote sparse solutions) [10]. Recently, Iordache et al. [11] proposed a collaborative sparse regression technique (CLSunSAL) based on an $\ell_{2,1}$ penalty function that enforces group-sparsity for the abundances of each material. This method was further improved in [12] by introducing a pre-processing step that identifies the elements from the spectral library that are more likely present in the image. The resulting MUSIC-CSR algorithm solves a sparse regression problem that is collaborative in the sense that all the image pixels are used to identify the active endmembers. Sparse regression for SU
can also be successfully performed within the Bayesian framework. For example, Dobigeon et al. [13] and Eches et al. [14] propose Bayesian models and Monte Carlo algorithms to identify the active endmembers in an HSI from a spectral library while ensuring that the abundances of absent endmembers are zero. Note that library-based methods are not the only strategy to address the absence of pure pixels (see [4], [15]–[17] for more details).

It is widely acknowledged that collaborative sparse regression methods can produce very accurate SU results. Collaboration is key because it improves the estimation of the support of the sparse abundance vectors, thus reducing significantly the number of unknowns. However, most existing collaborative techniques only exploit global information and therefore can only seek to determine if an endmember is present/absent in the entire image (that is, can only estimate the union of the supports of all the abundance vectors in the image). As a result, global collaborative techniques may overestimate significantly the support of the actual abundance vector of each pixel.

This paper presents a new collaborative sparse regression technique that exploits the local spatial correlations in the image to accurately detect the endmembers that are active/inactive in each pixel. Precisely, we present a Bayesian model that enforces simultaneously sparsity on the abundance vectors and strong spatial correlation on their supports (i.e., non-zero elements). This approach differs from the strategies adopted in the previous works [18]–[22] where spatial correlation was introduced by regularising the abundances or the nonlinear effects occurring in the image, and which do not promote directly solutions for the abundances with sparse and correlated supports.

The remainder of the paper is organized as follows. Section II recalls the classical linear mixing model for SU and presents the proposed Bayesian model for sparse regression. In Section IV we propose an original Monte Carlo method to perform Bayesian inference in this model and perform SU. The proposed methodology is demonstrated on synthetic and real HSI in in Sections V and VI. Conclusions are finally reported in Section VII.

II. PROBLEM STATEMENT

Consider a hyperspectral image \( \mathbf{Y} \in \mathbb{R}^{L \times N} \), where \( L \) is the number of spectral bands considered and \( N = N_{\text{row}} \times N_{\text{col}} \) corresponds to the total number of pixels. Under the linear mixing assumption, each image pixel \( \mathbf{y}_n = [Y_{1,n}, \ldots, Y_{L,n}]^T \in \mathbb{R}^L \) can be expressed as a linear combination of \( R \) known spectral signatures \( \mathbf{m}_1, \ldots, \mathbf{m}_R \) corrupted by zero-mean
Gaussian noise with covariance $\sigma^2 I_L$, that is,
\[
y_n = \sum_{r=1}^{R} a_{r,n} m_r + e_n, \quad e_n \sim \mathcal{N}(0, \sigma^2 I_L)\tag{1}
\]
where $a_{r,n}$ is the mixing coefficient associated with the $r$th endmember in the $n$th pixel. By setting $M = [m_1, \ldots, m_R]$ and $a_n = [a_{1,n}, \ldots, a_{R,n}]^T$, Eq. (1) can be conveniently expressed in matrix notation as $y_n = Ma_n + e_n$.

This paper considers the supervised spectral unmixing problem of the hyperspectral image $Y$, i.e., the estimation of the $R \times N$ abundance matrix $A = [a_1, \ldots, a_N]$. More precisely, we consider Bayesian methods for estimating $A$ given $Y$ (and the endmember matrix $M$) subject to the following two sets of physical constraints: first, the abundances are non-negative quantities, i.e., $a_{r,n} \geq 0 \quad \forall r, n$; second, there is at least one material present in each pixel and therefore $\|a_n\|_0 > 0$, where $\|\cdot\|_0$ denotes the $\ell_0$ vector pseudo-norm. We also assume that the value of the noise variance $\sigma^2$ is unknown, though prior knowledge about this value, if available, can be easily integrated into the model.

In a manner akin to [23], we model explicitly the sparsity of $A$ by using the decomposition
\[
A = Z \odot X\tag{2}
\]
where $Z \in \{0, 1\}^{R \times N}$ is a matrix of Bernoulli variables that “labels” each material as present (active) or absent (inactive) in each pixel, $X \in \mathbb{R}^{R \times N}$ is a matrix with positive entries that (jointly with $Z$) quantifies the abundances, and $\odot$ denotes the Hadamard (termwise) matrix product. This decomposition is particularly useful in Bayesian sparse regression problems because it allows eliciting separate statistical models for $a_n$’s support (through modelling $Z$), and for the values of the positive elements of $a_n$ (through $X$).

The next section presents a Bayesian model for estimating $Z$ and $X$ subject to the physical constraints discussed above. A key aspect of this model is that it will capture the fact that the pixels in which a material is active (or inactive) generally form spatial clusters (i.e., exhibit spatial group sparsity). In difficult unmixing scenarios, exploiting this strong prior information can significantly improve estimation results, as will be shown in this paper.

### III. BAYESIAN MODEL

This section presents the proposed the hierarchical Bayesian model for performing sparse source separation with collaborative supports. This model is defined by specifying the likelihood and the prior distribution of the parameters of interested $Z$ and $X$, as well as for
the other unknown parameters in the model (e.g. \( \sigma^2 \)) that will be subsequently removed by marginalisation (i.e., integrated out of the model’s joint posterior distribution).

A. Likelihood

From the observation model (1) and the parametrisation of \( A \) described in (2), the likelihood of the image \( Y \) given the unknown parameters is

\[
f(Y|Z, X, M, \sigma^2) = \prod_n f(y_n|z_n, x_n, M, \sigma^2) = \prod_n p_N(y_n|M(z_n \odot x_n), \sigma^2 I_L) \tag{3}
\]

where \( z_n \) (resp. \( x_n \)) is the \( n \)th column of \( Z \) (resp. \( X \)) and

\[
p_N(y_n|M(z_n \odot x_n), \sigma^2 I_L) = (2\pi\sigma^2)^L \exp \left(-\frac{||y_n - M(z_n \odot x_n)||_2^2}{2\sigma^2}\right) \tag{4}
\]

is the probability density function of a multivariate Gaussian vector with mean vector \( M(z_n \odot x_n) \) and diagonal covariance matrix \( \sigma^2 I_n \).

1) Prior distribution of \( Z \): As explained previously, a key aspect of the proposed Bayesian model is to take into account the fact that the pixels in which a given material is present or absent typically form spatial groups or clusters (as opposed to being randomly distributed in space). From a modelling viewpoint, this can be represented by correlating the Bernoulli variables or labels \( z_{r,n} \) across the spatial dimension indexed by \( n \). This can be achieved, for example, by stating that if a certain material is present (or absent) in a given pixel, this increases the probability of its presence (or absence) in neighbouring pixels. Taking into account that each material can exhibit its own spatial configuration, and the constraint that there must be at least one material present in each pixel, we propose to assign \( Z \) the following \textit{truncated multivariate Ising Markov random field} prior

\[
f(Z|\beta) = \frac{\psi(Z)}{C(\beta)} \exp \left( \sum_{r=1}^R \beta_r \phi_r(Z) \right) \tag{5}
\]

with \( \beta = \{\beta_1, \ldots, \beta_R\} \),

\[
\phi_r(Z) = \sum_n \sum_{n' \in \mathcal{V}(n)} \delta(z_{r,n} - z_{r,n'}), \tag{6}
\]

\[
\psi(Z) = \prod_n \max_r(z_{r,n}), \tag{7}
\]

\[
C(\beta) = \sum_Z \psi(Z) \exp \left[ \sum_{r=1}^R \beta_r \phi_r(Z) \right], \tag{8}
\]

and where \( \delta(\cdot) \) is the Kronecker delta function and \( \mathcal{V}(n) \) denotes the set of neighbours of pixel \( (n) \) (Figure 1 depicts two examples of neighbourhood structures that are commonly used
in image processing, in this paper we have used the 8-pixel neighbourhood of Fig. 1(b)). The hyper-parameters $\beta_1, \ldots, \beta_R$ act as regularization parameters that control the degree of spatial smoothness or regularity associated with each endmember, accounting for the fact that different materials may exhibit different spatial distributions.

To gain intuition about the proposed prior, we note that setting $\beta = 0$ in (5) and considering the prior (10) for the matrix $X$ leads to a Bernoulli-Gaussian type prior for the abundances which is closely related to the $\ell_0-\ell_2$ penalty often used for sparse regression. In the general scenarios where $\beta \neq 0$, Eq. (5) introduces spatial correlations between the components of $Z$ and, together with (10), leads to an $\ell_0-\ell_2$-type penalty promoting spatial group sparsity for the abundances. Prior (5) can also be understood as a collaborative prior, in the sense that, by assigning higher probabilities to configurations in which the supports of the abundance vectors $a_n$ are spatially correlated, it pools or shares information between the sparse regressions that take place at each pixel. Crucially, by capturing the spatial correlations that occur naturally in hyperspectral images, this prior can improve significantly estimation results.

Fig. 1. 4-pixel (left) and 8-pixel (right) neighborhood structures. The considered pixel appears in black and its neighbors in white.
B. Prior distribution of $X$

We assign each element of $X$ an independent conjugate Gaussian prior, truncated to $\mathbb{R}^+$ to satisfy the positivity constraint of these parameters, i.e.,

$$x_{r,n} \mid s_r^2 \sim \mathcal{N}_{\mathbb{R}^+} \left(0, s_r^2\right).$$

(9)

Note that to account for material-dependent variability each endmember $\mathbf{m}_r$ has its own scale hyper-parameter $s_r^2$. Setting $s^2 = (s_1^2, \ldots, s_R^2)^T$ the joint prior of $X$ is

$$f(X \mid s^2) \propto \prod_{n=1}^N \prod_{r=1}^R p_{\mathcal{N}}(x_{r,n} \mid 0, s_r^2)1_{\mathbb{R}^+}(x_{r,n}).$$

(10)

The value of the hyper-parameter vector $s^2$ remains unspecified and will be assigned an appropriate hyper-prior distribution (see Section III-D).

C. Prior distribution of the noise variance $\sigma^2$

In this paper we consider that there is no prior knowledge available about the value of the noise variance and assign $\sigma^2$ its non-informative Jeffreys prior [24]

$$f(\sigma^2) \propto \frac{1}{\sigma^2} 1_{\mathbb{R}^+}(\sigma^2).$$

(11)

Note that in scenarios where prior knowledge about the value of $\sigma^2$ is available, this can be easily integrated into the model by replacing the Jeffreys prior with an inverse gamma conjugate prior with hyper-parameter values reflecting this prior knowledge.

D. Hyperprior of the scale hyper-parameter vector $s^2$

Finally, we assign each scale hyper-parameter an independent conjugate inverse-gamma prior

$$s_r^2 \mid \gamma, \nu \sim \mathcal{IG}(\gamma, \nu)$$

(12)

where the values of $\gamma$ and $\nu$ are fixed to produce a vague or weakly informative prior (we have used $\gamma = 2.1$ and $\nu = 1.1$ in all the experiments reported in this paper). The joint prior of $s^2 = (s_1^2, \ldots, s_R^2)^T$ is thus given by

$$f(s^2 \mid \gamma, \nu) = \prod_{R=1}^R p_{\mathcal{IG}}(s_r^2 \mid \gamma, \nu)$$

(13)

with $p_{\mathcal{IG}}(s_r^2 \mid \gamma, \nu) = s_r^{-2\gamma-2} \exp\left(-\nu/s_r^2\right)\nu^\gamma/\Gamma(\gamma)$. 

September 30, 2014 DRAFT
E. Regularisation parameter $\beta$

A main advantage of Bayesian methods is that they allow estimating the appropriate amount of regularisation from data, thus freeing practitioners from the difficulty of setting regularisation parameters by cross-validation. Indeed, there are several Bayesian strategies for selecting the value of the regularisation parameter $\beta$ in a fully automatic manner (see [25] for a recent detailed survey on this topic). In this paper we use the empirical Bayes technique recently proposed in [26], where the value of $\beta$ is estimated by maximum marginal likelihood.

F. Joint posterior distribution of $Z, X, \sigma^2, s^2$

The structure of the proposed hierarchical Bayesian model is summarised in the directed acyclic graph (DAG) depicted below in Fig. 2.

Fig. 2. Directed acyclic graph (DAG) of the proposed hierarchical Bayesian model (parameters with fixed values are represented using black boxes).

Using Bayes’ theorem, and taking into account the conditional independences of the Bayesian model (see Fig. 2), the joint posterior distribution of $Z, X, \sigma^2$ and $s^2$ given the observations $Y$, the library of spectral signatures $M$ and the model’s fixed parameters $\beta, \gamma$ and $\nu$, is given by

$$f(Z, X, \sigma^2, s^2 | Y, M, \beta, \gamma, \nu) \propto f(Y | Z, X, M, \sigma^2) f(Z | \beta) f(X | s^2) f(s^2 | \gamma, \nu) f(\sigma^2).$$ (14)

The following section presents a Monte Carlo algorithm to perform Bayesian inference using the proposed model.
The Bayesian model defined in Section III specifies a joint posterior distribution for the unknown parameters $Z, X, \sigma^2, s^2$ given the fixed quantities $Y, M, \gamma, \nu$ and the hyperparameter $\beta$ which is unknown but represented as a deterministic parameter (whose value will be tuned during the inference procedure). According to the Bayesian paradigm, this posterior distribution fully describes the information about the unknowns that is provided by the data and by the prior knowledge available. However, for unmixing applications it is necessary to summarise this posterior distribution in the form of point estimates; that is, to assign specific values for the unknown quantities of interest (in our problem the abundance vectors). Here we consider the following coupled Bayesian estimators that are particularly suitable for sparse regression problems: the marginal maximum a posteriori (MMAP) estimator for the support of the abundance vectors or “presence maps”

$$z_{r,n}^{MMAP} = \arg\max_{z_{r,n} \in \{0, 1\}} f(z_{r,n} | Y, M, \beta, \gamma, \nu),$$

(15)

and, conditionally on the estimated supports, the minimum mean square error estimator of the abundances

$$a_{r,n}^{MMSE} = E \left[ x_{r,n} | z_{r,n} = z_{r,n}^{MMAP}, Y, M, \beta, \gamma, \nu \right],$$

(16)

where

$$f(z_{r,n} | Y, M, \beta, \gamma, \nu) = \int f(Z, X, \sigma^2, s^2 | Y, M, \beta, \gamma, \nu) dZ_{\setminus z_{r,n}} dX d\sigma ds,$$

with the matrix $Z_{\setminus z_{r,n}}$ containing the remaining elements of $Z$ once $z_{r,n}$ has been removed and where $E [\cdot]$ denotes the expectation with respect to the conditional marginal density

$$f(x_{r,n} | z_{r,n}, Y, M, \beta, \gamma, \nu) = \frac{\int f(Z, X, \sigma^2, s^2 | Y, M, \beta, \gamma, \nu) dZ_{\setminus z_{r,n}} dX_{\setminus z_{r,n}} d\sigma ds}{f(z_{r,n} | Y, M, \beta, \gamma, \nu)}.$$

Note that the abundance estimator (16) is sparse by construction, i.e.,

$$E [x_{r,n} | z_{r,n} = 0, Y, M, \beta, \gamma, \nu] = 0,$$

and that, by marginalising out the other unknowns, it automatically takes into account the uncertainty about $\sigma^2$ and $s^2$.

Computing (15) and (16) is challenging because it requires having access to the univariate marginal densities of $z_{r,n}$ and the joint marginal densities of $(x_{r,n}, z_{r,n})$, which in
turn require computing the posterior (14) and integrating it over a very high-dimensional space. Fortunately these estimators can be efficiently approximated with arbitrarily large accuracy by Monte Carlo integration. Precisely, it is possible to compute (15) and (16) by first using a Markov Chain Monte Carlo (MCMC) computational method to generate samples asymptotically distributed according to (14), and subsequently using these samples to approximate the required marginal probabilities and expectations.

Here we propose a Gibbs sampler to simulate samples from (14), as this type of MCMC method is particularly suitable for models involving hidden Markov random fields [27, Chap. 10]. The output of this algorithm are two Markov chains of \(N_{MC}\) samples \(\{X^{(1)}, \ldots, X^{(N_{MC})}\}\) and \(\{Z^{(1)}, \ldots, Z^{(N_{MC})}\}\) that are asymptotically distributed according to the posterior distribution \(f(Z, X|Y, M, \beta, \gamma, \nu)\). The first \(N_{bi}\) samples of these chains correspond to the so-called burn-in transient period and should be discarded (the length of this period can be assessed visually from the chain plots or by computing convergence tests). The remaining \(N_{MC} - N_{bi}\) of each chain are used to approximate the Bayesian estimators (15) and (16) as follows

\[
\hat{z}_{r,n}^{MMAP} = \arg\max_{u=\{0,1\}} \sum_{t=N_{bi}+1}^{N_{MC}} \delta \left( z_{r,n}^{(t)} - u \right),
\]

and

\[
\hat{a}_{r,n}^{MMSE} = \frac{\sum_{t=N_{bi}+1}^{N_{MC}} z_{r,n}^{(t)} \delta \left( z_{r,n}^{(t)} - \hat{z}_{r,n}^{MMAP} \right)}{\sum_{t=N_{bi}+1}^{N_{MC}} \delta \left( z_{r,n}^{(t)} - \hat{z}_{r,n}^{MMAP} \right)}.
\]

Note that (17) and (18) converge to the true Bayesian estimators (15) and (16) as \(N_{MC} \to \infty\). The remainder of this sections provides details about the main steps of the proposed Gibbs sampler, termed Collaborative sparse Unmixing (CSU) and summarised in Algo. [I] below. Note that for clarity the dependence of all distributions on the known fixed quantities \(M, \gamma, \nu\) and \(\beta\) is omitted.
Algorithm 1

Collaborative sparse Unmixing (CSU)

1: Fixed input parameters: M, K, γ, ν, number of burn-in iterations \(N_{bi}\), total number of iterations \(N_{MC}\)
2: Initialization \((t = 0)\)
   • Set \(X^{(0)}, Z^{(0)}, \sigma^2(0), s^2(0), \beta(0)\)
3: Iterations \((1 \leq t \leq N_{MC})\)
4: Set \(Z^* = Z^{(t-1)}\)
5: for \(n = 1 : N\) do
6: \(\text{Sample } z_n^* \sim f(z_n^* | Y, Z^{(t-1)}, \sigma^2(t-1), s^2(t-1)) \text{ in (19)}\)
7: end for
8: Set \(Z^{(t)} = Z^*\)
9: Sample \(X^{(t)} \sim f(X | Y, Z^{(t)}, \sigma^2(t-1), s^2(t-1), \text{ in (20)}\)
10: Sample \(\sigma^2(t) \sim f(\sigma^2 | Y, X^{(t)}, Z^{(t)}, s^2(t-1)) \text{ in (22)}\)
11: Sample \(s^2(t) \sim f(s^2 | Y, X^{(t)}, Z^{(t)}, \sigma^2(t)) \text{ in (23)}\)
12: Update \(\beta(t) \leftarrow \beta^{(t-1)}\) using [26].
13: Set \(t = t + 1\).

A. Sampling the label matrix \(Z\)

The label matrix \(Z\) is updated pixel-wise by iteratively simulating from the distribution of the labels at each pixel given the other pixels; that is, the distribution of the vector \(z_n\) given the matrix \(Z_{\setminus z_n}\) containing the remaining elements of \(Z\) once \(z_n\) is removed

\[
P(z_n | Y, Z_{\setminus z_n}, X, \sigma^2, s^2) \propto f(y_n | z_n, x_n, \sigma^2) f(z_n | Z_{\setminus z_n})
\]

\[
\propto \exp \left[ 2 \sum_{r=1}^R \beta_r \sum_{n' \in \mathcal{V}(n)} \delta(z_{r,n'} - z_{r,n}) \right] \times \exp \left[ -\frac{1}{2\sigma^2} \|y_n - M(x_n \odot z_n)\|^2_2 \right] \tag{19}
\]

for \(||z_n||_0 > 0\) and \(P(z_n | Y, Z_{\setminus z_n}, X, \sigma^2, s^2) = 0\) otherwise. Algorithmically, this simulation step can be achieved by indexing the \(2^R - 1\) admissible configurations of \(z_n\) and then randomly selecting a specific one with probability defined in (19).

B. Sampling \(X\)

The conditional distribution of \(X\) given the other unknown parameters can be factorised pixel-wise as a product of \(N\) marginal distributions

\[
f(X | Y, Z, \sigma^2, s^2) = \prod_{n=1}^N f(x_n | y_n, z_n, \sigma^2, s^2), \tag{20}
\]
that can be efficiently sampled independently and in parallel

\[ x_n | y_n, z_n, \sigma^2, s^2 \sim N_\Omega(\bar{x}_n, \Sigma_n), \]  

(21)

where \( \Omega = (\mathbb{R}^+)^R \) is the positive orthant of \( \mathbb{R}^R \) and

\[
\Sigma_n = (\sigma^{-2} D_n M^T M D_n + S^{-1})^{-1},
\]

\[
\bar{x}_n = \sigma^{-2} \Sigma_n D_n M^T y_n,
\]

and where \( S = \text{diag}(s^2) \) and \( D_n = \text{diag}(z_n) \) are diagonal matrices with diagonal elements given by \( s^2 \) and \( z_n \). In this paper we use the method [28] to simulate efficiently from (21).

### C. Sampling the noise variance \( \sigma^2 \)

The conditional distribution associated with \( \sigma^2 \) has a simple closed form expression and is given by

\[ \sigma^2 | Y, X, Z, s^2 \sim IG \left( N/2, E_\sigma \right), \]  

(22)

with \( E_\sigma = \sum_n \|y_n - M(x_n \odot z_n)\|_2^2 / 2 \).

### D. Sampling the hyperparameter vector \( s^2 \)

The conditional distribution of \( s^2 \) can be factorised endmember-wise as a product of \( R \) independent marginal distributions

\[ f(s^2 | Y, X, Z, \sigma^2) = \prod_{r=1}^R f(s^2_r | X) \]  

(23)

that can be easily sampled independently and in a parallel manner

\[ s^2_r | X \sim IG \left( N/2 + \gamma, \frac{\sum_n x^2_{r,n}}{2} + \nu \right) . \]  

(24)

### V. Validation with synthetic data

This section demonstrates the proposed methodology on a series of experiments conducted using synthetic data. An applications to a real hyperspectral image is reported in Section VI.
A. Data sets

The performance of the proposed collaborative sparse unmixing (CSU) method is first evaluated on two synthetic images \(I_1\) and \(I_2\) of size \(100 \times 100\) pixels and \(L = 224\) spectral bands. There are \(R_0 = 5\) endmembers present in the images which correspond to the minerals Dipyre, Spodumene, Clinoptilolite, Mordenite and Olivine 1. Their spectral signatures have been obtained from the USGS spectral library [29] and are depicted in Fig. 3. Note that the angles between the spectral signatures are between \(3.01^\circ\) and \(3.05^\circ\) (i.e., the endmembers are highly correlated with mutual coherence [30] that equals to \(M = 0.9986\)), making the unmixing problem very challenging. The support maps that determine the spatial distribution of each material have been generated by simulating from the prior model (5) with \(\beta = [0.2; 0.275; 0.35; 0.425; 0.5]^T\), and are depicted in the top row of Fig. 4. For both images the matrix \(X\) has been generated by sampling from (10) with \(s_r = 0.3, \forall r\). Finally, the average signal-to-noise ratio (SNR) for the images \(I_1\) and \(I_2\) are approximately 30dB \((\sigma^2 = 8.10^{-4})\)
and 20dB ($\sigma^2 = 8.10^{-3}$), respectively.

![Active support maps of the $R_0 = 5$ endmembers associated with $I_1$ and $I_2$.](image_url)

Fig. 4. Top row: Active support maps of the $R_0 = 5$ endmembers associated with $I_1$ and $I_2$. Active support maps estimated with the proposed CSU algorithm and CLSunSAL ($\rho = 0.01$) for $I_1$ (second and third top rows) and $I_2$ (fourth and bottom rows). White (resp. black) pixels correspond to regions where a component is present (resp. absent).

For comparison we use the state-of-the-art collaborative sparse regression algorithm CLSunSAL [11], that we implemented with regularisation parameter value $\lambda = 3.10^{-4}$. For completeness we also compare with the widely used Non-negatively Constrained Least-Squares algorithm (NCLS) [5], which solves a maximum likelihood problem and does not exploit any prior information about the abundance vectors.

### B. Supervised unmixing

In this first experiment we consider that the materials present in the images $I_1$ and $I_2$ are perfectly known and we estimate the abundance vectors with CSU, CLSunSAL and
NCLS. The CSU algorithm has been implemented using $N_{MC} = 3000$ and $N_{bi} = 1000$, and by allowing the algorithm to self-adjust the regularisation parameter $\beta$ with the technique proposed in [26]. The estimated values of $\beta$ are $\beta = [0.19, 0.28, 0.33, 0.37, 0.44]^T$ and $\beta = [0.20, 0.28, 0.33, 0.33, 0.44]^T$ for $I_1$ and $I_2$, respectively. The associated computation times for a Matlab implementation on a 3.0GHz Intel Xeon quad-core workstation are provided in Table I. For illustration, the estimated supports obtained with CSU and CLSunSAL for each image are depicted in Fig. 4. For CLSunSAL, the detection maps have been obtained by thresholding the estimated abundances with a threshold arbitrarily set to $\rho = 0.01$. We observe that the results obtained with CSU are in very good agreement with the ground truths for $I_1$ and $I_2$. On the other hand, the abundances obtained with CLSunSAL are significantly less accurate, thus confirming that taking into account the spatially correlations between the abundance vectors is key to achieving accurate estimation results. This valuable prior knowledge is all the more important in low SNR scenarios such as the one depicted in two bottom rows of Fig. 4, where CSU exhibits a very good robustness to noise and clearly outperforms CLSunSAL. Finally, note that the detection maps obtained using NCLS (not presented here) are similar to those obtained with CLSunSAL.

For numerical comparison, we computed the root mean square error (RMSE)

$$\text{RMSE}_n = \sqrt{\| \hat{a}_n - a_n \|^2} \quad (25)$$

that quantifies the average accuracy of the estimated abundances $\hat{a}_n$ with respect to the truth $a_n$ at the $n$-th pixel. The first four rows of Table I show the average RMSEs for CSU, CLSunSAL, NCLS, and the oracle NCLS (o-NCLS), which consists of applying NCLS using only the active materials in each pixel (i.e., with perfect knowledge of the support of the abundance vectors). We observe that CSU provides significantly more accurate estimations than CLSunSAL and NCLS, achieving an average RMSE that is close to the oracle. Again, the good performance of CSU results from the collaboration between pixels introduced by the prior model (5). By comparing the first and second columns of Table I we confirm that this prior knowledge becomes all the more important as the noise level increases.

Finally, in order to further highlight the good performance of the proposed prior model we have also computed the reconstruction error (RE) defined as

$$\text{RE}_n = \sqrt{\| MA_n - y_n \|^2}. \quad (26)$$

Note that this error essentially measures the likelihood of the abundance estimates given the observed image $Y$. However, because the SU problem is not well-posed, the capacity of
the likelihood to identify good solutions is severely limited and the additional information provided by the prior model is key to deliver accurate estimation results. In these scenarios Bayesian methods, which combine observed and prior information, can greatly outperform other estimation techniques. The average RE for each method is reported in Table III. We observe that CSU, CLSunSAL and NCLS exhibit very similar REs. However, we know from Table II that CSU outperforms significantly NCLS in terms of estimation accuracy. The contrast between these two figures of merit confirms that the superior performance of CSU is directly related to the proposed prior model, which captures the correlations between the supports of the abundance vectors and effectively introduces a means of collaboration that allows sharing information between pixels and increasing robustness to noise.
C. Semi-supervised unmixing

We now consider that two additional endmembers are incorrectly included in the library M, although they are not present in the scene. To make the SU problem particularly challenging, the two additional endmembers are Olivine 2 and Adularia, whose spectral signatures are highly correlated with the signatures of the other endmembers. For example, note from Fig. 3 that discriminating between Olivine 1 and Olivine 2 is very difficult. We contrast CSU with CLSunSAL and NCLS (CSU is implemented using $N_{MC} = 7000$ and $N_{bi} = 5000$ and the estimated values of $\beta$ are $\beta = [0.20, 0.28, 0.35, 0.40, 0.45, 0.47, 0.47]^T$ and $\beta = [0.20, 0.29, 0.36, 0.37, 0.45, 0.46, 0.42]^T$ for $I_1$ and $I_2$, respectively.).

Fig. 5 shows the support estimations obtained with CSU and CSunSAL for $I_1$ and $I_2$ with the $R = 7$ endmembers (two of which are absent from the scene). Again, we observe that the results obtained with CSU are in very good agreement with the ground truth for both $I_1$ and $I_2$ (SNR = 30dB and 20dB, respectively), whereas the results obtained with CLSunSAL are significantly less accurate. More importantly, we observe that CSU has successfully detected that two endmembers (Olivine 2 and Adularia) are not present in the scene, in spite of the strong similarities between Olivine 1 and Olivine 2.

Finally, the three bottom rows of Table II show the RNMSEs obtained with CSU, CLSunSAL and NCLS for $I_1$ and $I_2$ with $R = 7$. We observe that CSU is extremely robust to the presence of the two redundant endmembers and is able to accurately discriminate between Olivine 1 and Olivine 2. On the other hand, CLSunSAL and NCLS have difficulties detecting the true supports of the abundance vectors and produce estimation results that are significantly less accurate, in particular in low SNR conditions. Again, the superior performance of CSU is directly related to the prior model (5) that captures the spatial correlations in the image.
Fig. 5. Top row: Active support maps of the $R = 7$ endmembers associated with $I_1$ and $I_2$. Active support maps estimated with the proposed CSU algorithm and CLSunSAL ($\rho = 0.01$) for $I_1$ (second and third top rows) and $I_2$ (fourth and bottom rows). White (resp. black) pixels correspond to regions where a component is present (resp. absent).

and regularises the supports of the abundance vectors. For completeness, Fig. 6 presents the abundances estimated with CSU and CLSunSAL for $I_1$. We observe that without the additional information provided by the spatial correlation, CLSunSAL can hardly discriminate between the two kinds of Olivine.

VI. APPLICATION TO REAL A HYPERSONTICAL IMAGE

This section presents an application of the proposed CSU method to the publicly available “TERRAIN” hyperspectral image\(^1\) of size $500 \times 307$. This image was acquired with the HYDICE sensor\(^{31}\) and has $L = 166$ spectral bands recorded from the visible to near infrared spectrum and with a spatial resolution of 0.75m (the bands with low SNR due to water absorption have been removed). This dataset has already been studied in\(^2\) and is known to be mainly composed of soil, trees and grass. The library of endmembers for this image has been estimated by using the HySime algorithm\(^{32}\) followed by VCA algorithm

\(^1\)Data set available at [http://www.agc.army.mil/Missions/Hypercube.aspx](http://www.agc.army.mil/Missions/Hypercube.aspx)
Fig. 6. Left: Abundance maps of the $R = 7$ endmembers for $I_1$. Middle: Abundance maps estimated by the proposed algorithm. Right: Abundance maps estimated by the NCLS algorithm.

[33] and retaining the principal $R = 5$ spectral signatures related to two types of soil, shade, trees and grass. The signatures of these 5 endmembers are depicted in Fig. [7].
The first row of Fig. 8 shows the presence maps for each material in the image, which have been estimated with CSU using $N_{\text{MC}} = 5000$, $N_{\text{bi}} = 3000$ and by allowing the algorithm to self-adjust the regularisation parameter $\beta$ with the technique [26]. The second row of Fig. 8 shows the abundances of each material estimated with CSU. For comparison, the third and forth rows present the abundances obtained with NCLS and CLSunSAL. We observe that the estimates obtained with CSU are more accurate than the ones obtained with NCLS and CLSunSAL (note in particular the horizontal artifacts that are present in the maps obtained with NCLS and CLSunSAL for “shade”, which are not present in the CSU maps). For completeness, Table IV reports the reconstruction errors associated with each algorithms. Similarly to the synthetic experiments, we observe that CSU scores comparably to CLSunSAL and NCLS, despite the fact that the estimation of the abundances seems more accurate (less artifacts).

Fig. 7. Left: TERRAIN HSI (true color). Right: the $R = 5$ endmembers estimated with VCA algorithm [33] for the TERRAIN-HSI.
Fig. 8. Top: Active support maps of estimated by the CSU algorithm for the real image (white (resp. black) pixels correspond to regions where a component is present (resp. absent)). The $R = 5$ abundance maps estimated by the CSU (middle) and NCLS (bottom) algorithms (red pixels correspond to large abundances, contrary to blue pixels).

<table>
<thead>
<tr>
<th>Unmixing algo.</th>
<th>ARE ($\times 10^{-3}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSU</td>
<td>7.2</td>
</tr>
<tr>
<td>CLSunSAL</td>
<td>6.6</td>
</tr>
<tr>
<td>NCLS</td>
<td>6.6</td>
</tr>
</tbody>
</table>
VII. CONCLUSION

This paper presented a new Bayesian method for linear unmixing of hyperspectral image that is based on a collaborative sparse regression formulation. The main novelty is a Bayesian model for sparse regression that takes into account the fact that the supports of the abundance vectors are spatially correlated (i.e., the spatial distributions of the materials in the scene exhibit some degree of regularity). This prior information is encoded in the model by using a truncated multivariate Ising Markov random field model, which also accounts for the facts that pixels cannot be empty (i.e., there is at least one material per pixel) and that each material in the scene may require a different amount of spatial regularisation. The proposed Bayesian model also takes into consideration that material abundances are non-negative quantities and that the level of noise contaminating the image may be fully or partially unknown. Following on from this, we presented a Markov chain Monte Carlo algorithms to perform Bayesian inference with this model and compute the statistical estimators of interest. Precisely, we proposed a Gibbs sampler that allows estimating the probabilities that materials are present or absence in each pixel, and, conditionally on any given configuration (typically the maximum a posteriori), computing the MMSE estimates of the abundance vectors. A remarkable characteristic of the proposed inference algorithm is that it self-adjusts the amount of regularity enforced by the random field, thus relieving practitioners from the difficult task of setting regularisation parameters by cross-validations. Finally, the good performance of the proposed methodology was demonstrated through a series of experiments with synthetic and real data and comparisons with other algorithms from the literature.

Due to the computational complexity of the proposed sampling strategy, the algorithm presented in this paper can only be directly applied to problems with small numbers of endmembers (e.g., $R \leq 25$). For problems with larger libraries it is computationally more efficient to generate samples from a relaxed posterior in which the “non-empty pixel” constraint of the MRF is removed, and then reintroduce this constraint by importance sampling (i.e., by removing from the chain the samples that do not verify the constraint). Another possibility for problems with large libraries is to use the MUSIC-CSR algorithm as a pre-processing step to identify endmembers that are absent from the scene, and then apply our method using a pruned library (note that MUSIC-CSR is based on a convex problem that can be solved efficiently using state-of-the-art optimization algorithms).

It is important to note that some other linear unmixing methods consider that the values
of the abundance vectors are spatially correlated, whereas we consider the correlations between their supports (i.e., presence and absence patterns). Combining both approaches would certainly lead to a more accurate prior model for the abundance vectors and is definitively interesting and worth investigating in future work. For many applications or hyperspectral datasets, it makes sense to consider additional abundance constraints, such as the sum-to-one constraint to further improve unmixing results. Embedding this constraint within $\ell_0$-type sparse regression models is a challenging problem that is currently under investigation. Another perspective for future work is to investigate more sophisticated spatial models that describe hyperspectral images more accurately, in particular in complex scenes with numerous materials and non-linear effects. Finally, we believe that the methodology presented in this paper could be interesting for other regression problems that exhibit structured sparsity and intend to further investigate this is the future.

ACKNOWLEDGMENTS

The authors would like to thank Prof. Jean-Yves Tourneret and Dr Nicolas Dobigeon, from the University of Toulouse, IRIT-ENSEEIHT, France, for interesting discussion regarding this work.

REFERENCES


