IMPROVING JOINT SPARSE HYPERSPECTRAL UNMIXING BY SIMULTANEOUSLY CLUSTERING PIXELS ACCORDING TO THEIR MIXTURES

Seyyede Fatemeh Seyyedsalehi, Hamid R. Rabiee

Department of Computer Engineering, Sharif University of Technology, Iran

ABSTRACT

In this paper we propose a novel hierarchical Bayesian model for sparse regression problem to use in semi-supervised hyperspectral unmixing which assumes the signal recorded in each hyperspectral pixel is a linear combination of members of the spectral library contaminated by an additive Gaussian noise. To effectively utilizing the spatial correlation between neighboring pixels during the unmixing process, we exploit a Markov random field to simultaneously group pixels to clusters which are associated to regions with homogeneous mixtures in a natural scene. We assume Sparse fractional abundances of members of a cluster to be generated from an exponential distribution with the same rate parameter. We show that our method is able to detect unconnected regions which have similar mixtures. Experiments on synthetic and real hyperspectral images confirm the superiority of the proposed method compared to alternatives.

***Index Terms*—** Semi-supervised hyperspectral unmixing, joint sparse regression, spatial correlations, Markov random field.

1 Introduction

Hyperspectral imaging collects the electromagnetic spectra with hundreds of narrow wavelength bands from both of visible and infrared regions. Due to the limiting spatial resolution of imaging devices, a hyperspectral pixel represents different spectral signatures mixed together [1]. These spectral signatures are the function of pure material spectra existing in the scene. The goal of hyperspectral unmixing is to find the proportion of each pure material spectrum in the mixture of each pixel. Most of unmixing algorithms are based on a mathematical model that describes the combination of pure spectra. The most well-known one is the linear mixture model (LMM) which represents a satisfying estimation of nonlinear processes, in many hyperspectral applications [1].

Unsupervised hyperspectral unmixing directly extracts pure material signatures and their proportions from the observed pixels. Initially introduced algorithms for endmember extraction methods assume a pure pixel for each endmember spectrum exists in the hyperspectral image or observed signals lie in a convex region called simplex which are not satisfied in many natural hyperspectral images. Motivated by the success of deep learning methods, works in [2, 3] exploit a deep autoencoder for unsupervised unmixing. To incorporate the spatial correlation between neighboring pixels, works in [4, 5] utilize convolutional layers in their deep networks. Two main disadvantages of these deep networks are dependency of their performance on the availability of a large training set and the lack of interpretation.

Semi-supervised unmixing algorithms use hyperspectral libraries to tackle the issues related to the endmember extraction methods. Hyperspectral libraries contain the spectral signature of pure materials. Most of semi-supervised algorithms consider a fractional abundance vector for whole of the library members for each pixel and attempt to estimate this vector. However, the number of pure material spectra in the mixture of a pixel is much fewer than the size of the spectral library and the fractional abundance vector is sparse. Convex ℓ_1 minimizer and the weighted ℓ_1 are used for hyperspectral data in [6, 7], respectively. The work in [8] tries the well-known greedy algorithm orthogonal matching pursuit (OMP), for hyperspectral data. The work of [9, 10, 11, 12, 13] introduce convex optimizations in which the spatial correlation between pixels is incorporated in the objective function to enhance the performance.

Bayesian models are another trend. The work of Dobigeon [14] proposes a Gibbs sampler to estimate fractional abundances which adds new steps to the sampler to select pure endmembers. The model in [15] omits the endmember selection step and chooses appropriate prior distributions to impose sparsity on fractional abundance vectors. Both of [15, 14] solve the problem for each pixel independently. Motivated by the similarity of neighboring pixels, Altmann in [16] use a Markov random field (MRF) to model spatial correlations. It considers a matrix of Bernoulli variables that determines the existence of each pure material in the mixture of each pixel which leads to a large number of hidden variables.

In this paper we introduce a semi-supervised unmixer which is able to simultaneously cluster pixels to groups with similar mixtures and improve the unmixing result. We exploit Potts Markov random field to model spatial correlations between pixels in homogeneous regions. The work in [17] adopts this MRF for supervised unmixing in which endmembers in the mixture of pixels is known. Compared to the work by Altmann in [16] which has a large number of variables

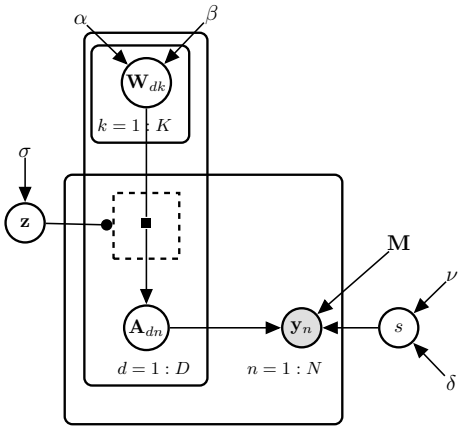


Fig. 1. The proposed hierarchical Bayesian model.

should be inferred, ours leads to a limited number of random variables and a simpler inference step. To incorporate spatial correlations the work in [10] groups all pixels in an image in one cluster and [18] segments it to rectangular patches. Obviously these patterns do not match to the exact borders of homogeneous regions of a natural scene which have arbitrary shapes. In the presented work borders of homogeneous regions can be explored automatically. To encourage the similarity between pixels in each cluster, we assume their sparse fractional abundances are generated from an exponential distribution with the same rate parameter which is more appropriate for heterogeneous areas compared to works in [9, 10, 11] which explicitly enforce the similarity. Compared to our prior work [18] the proposed method is able to form larger groups of similar pixels with fewer errors, so rate parameters are inferred more precisely which leads to a more accurate unmixer.

2 Proposed Method

Our proposed method is based on the linear mixture model in which each observed spectral vector is a linear combination of endmember spectra existing in the scene,

$$\mathbf{y}_n = \mathbf{M}\mathbf{a}_n + \varepsilon \tag{1}$$

where $\mathbf{y}_n \in \mathbb{R}^L$ is the observed spectral vector of a pixel recorded in L channels. The matrix $\mathbf{M} = [\mathbf{m}_1, \ldots, \mathbf{m}_D] \in \mathbb{R}^{L\times D}$ generally shows the matrix of endmember spectra but in semi-supervised linear unmixing, it is the spectral library. The $\mathbf{a}_n \in \mathbb{R}^D$ is the fractional abundance vector with non-negativity and sum-to-one constraints [1]. Finally, $\varepsilon \in \mathbb{R}^L$ is the zero-mean i.i.d. Gaussian noise with the variance s.

2.1 The proposed hierarchical Bayesian model

In a natural hyperspectral image, each pixel lies in a physical region of the scene. Motivated by the similarity of the mixture of pixels of one region, we classify pixels into K groups. K should be set manually for each hyperspectral image. We consider a discrete random variable z_n as a label for nth pixel of the image which shows its group. A Markov random field is a common joint distribution in many image segmentation methods. The set $\mathbf{z} = \{z_n | n = 1 \cdots N\}$ is a Potts random field with the following joint distribution,

$$\mathbf{z} \sim \frac{1}{G(\sigma)}\, exp\left(\sum_{n=1}^{N}\sum_{n'\in\vartheta(n)} \sigma\delta(z_n - z_{n'})\right) \tag{2}$$

where σ is the granularity coefficient which shows the homogeneity degree of the model and we set it manually, $G(\sigma)$ is the partition function, N is the number of pixels in the image, $\vartheta(n)$ is the set of neighbors of nth pixel and $\delta(.)$ is the Kronecker function. To define the neighboring relation, we use the $1\text{–}order\ neighborhood$ in which each pixel has at most four neighbors.

To impose sparsity on fractional abundance vectors, we assume each element of matrix $\mathbf{A} = \{\mathbf{a}_1, \ldots \mathbf{a}_N\}$ is from an Exponential distribution that is peaked at zero with an adjustable rate and its domain is limited on positive real numbers which satisfies the non-negativity constraint. The rate parameter of dth fractional abundance of nth pixel $\mathbf{A}_{dn}$ is determined by its group. For the kth group, we consider random variables $\mathbf{W}_{dk},\ d = 1 \ldots D$ which show the rate parameter of the group's members. So the conditional distribution of $\mathbf{A}_{dn}$ is,

$$\mathbf{A}_{dn}|z_n = k,\ \mathbf{W}_{dk} \sim exp(\mathbf{A}_{dn}|\mathbf{W}_{dk}) \tag{3}$$

Fig. 1 shows the graphical model of the proposed model. Labels z_n are attached to the model by a gate structure. This structure assigns $\mathbf{A}_{dn}$ to the $\mathbf{W}_{dz_n}$. We consider a conjugate Gamma distribution for $\mathbf{W}_{dk}$ as follows:

$$\mathbf{W}_{dk} \sim Gamma(\mathbf{W}_{dk}|\alpha, \beta) \tag{4}$$

where α and β are scale and shape parameters.

Finally, considering the linear mixture model and an i.i.d. Gaussian noise, the nth observed spectral vector $\mathbf{y}_n$ is a Normal random variable conditioned on the fractional abundance vector $\mathbf{a}_n$ and the noise variance s,

$$\mathbf{y}_n|\mathbf{a}_n, s \sim \mathcal{N}(\mathbf{M}\mathbf{a}_n, sI_L) \tag{5}$$

A conjugate Inverse-Gamma distribution with parameters ν and δ is considered for the noise variance s,

$$s \sim Inverse\text{–}Gamma(s|\nu, \delta) \tag{6}$$

2.2 The joint posterior distribution

Considering the independence of variables according to Fig. 1, the posterior distribution of hidden variables condi-

Algorithm 1 The proposed algorithm

function BHUPC UNMIXER
  $\mathbf{A}^0 \leftarrow \arg\min_{\mathbf{A}} \|\mathbf{Y} - \mathbf{MA}\|_F^2 + \|\mathbf{A}\|_1$
  $\mathbf{z}^0 \leftarrow$ Sample Discrete-Uniform$(1, 2, \dots K)$
  $\mathbf{W}^0 \leftarrow \frac{\alpha}{\beta}$
  $s^0 \leftarrow 0.01$
  for $t = 1, 2, \dots T$ **do**
    $\mathbf{A}^t \leftarrow$ Sample $f(\mathbf{A}|\mathbf{W}^{t-1}, \mathbf{z}^{t-1}, s^{t-1})$
    $\mathbf{W}^t \leftarrow$ Sample $f(\mathbf{W}|\mathbf{A}^t, \mathbf{z}^{t-1}, s^{t-1})$
    $s^{t-1} \leftarrow$ Sample $f(\mathbf{s}|\mathbf{W}^t, \mathbf{z}^{t-1}, A^t)$
    for $k = 1, 2, \dots K$ **do**
      $p_k = f(\mathbf{z} = k|\mathbf{W}^t, \mathbf{A}^t, s^t)$
    end for
    $\mathbf{z}^t \leftarrow$ Sample Multinomial$(p_1,\ p_2, \dots, p_K)$
  end for
  $\mathbf{z}_{est} \leftarrow$ MAP estimation from $\mathbf{z}^{1,2,\dots T}$
  $\mathbf{A}_{est} \leftarrow$ MMSE estimation from $\mathbf{A}^{\{t|\mathbf{z}^t = \mathbf{z}_{est}\}}$
end function

tioned on observed signals can be expressed as follows:

$$f(\mathbf{A},\ \mathbf{W},\ \mathbf{z},\ s|\mathbf{Y}) \propto \prod_{n=1}^{N} f(\mathbf{y}_n|\mathbf{a}_n, s)\times \prod_{n=1}^{N}\prod_{d=1}^{D} f(\mathbf{A}_{dn}|\mathbf{W}_{dz_n}) \times \prod_{k=1}^{K}\prod_{d=1}^{D} f(\mathbf{W}_{dk}) \times f(\mathbf{z}) \times f(s) \tag{7}$$

which is too complex to obtain a closed-form expression for Bayesian estimators. Therefore, we use the Gibbs sampler that is an approximate inference technique which attempts to asymptotically generate samples from a joint distribution.

2.3 The Gibbs sampler

The Gibbs sampler produces a chain of samples from a joint desired distribution by sampling from the marginal distribution over each variable conditioned on observing other ones. Algorithm 1 shows the general approach. First, variables are initialized. Then, a complete sample from variables is produced in each iteration. After a sufficient number of iterations, samples can be considered from the desired distribution. Finally, we use maximum a posteriori (MAP) and minimum mean square error (MMSE) estimators to infer variables. Obtained conditional distributions for the Gibbs sampler, $f_G(.)$, are provided at the remainder of this section.

The conditional distribution of fractional abundance variables $\mathbf{A}_{dn}$ is as follows,

$$f_G(\mathbf{A}_{dn}) = N(\mathbf{A}_{dn}|\mu,\ \Sigma) \times I_{(0,\ \infty)}(\mathbf{A}_{dn}) \tag{8}$$

where,

$$\begin{cases} \mu = -\frac{\sum_{j=1;j\neq d}^{D} \mathbf{A}_{jn}\left(\left(\mathbf{M}^T\mathbf{M}\right)_{jd} + \left(\mathbf{M}^T\mathbf{M}\right)_{dj}\right) + 2\mathbf{W}_{dz_n} s - 2\mathbf{y}_n^T\mathbf{M}_d}{2\left(\mathbf{M}^T\mathbf{M}\right)_{dd}} \\ \Sigma = \frac{s}{\left(\mathbf{M}^T\mathbf{M}\right)_{dd}} \end{cases} \tag{9}$$

which is a truncated normal distribution. The group number, z_n, is a discrete variable. To obtained a sample we first calculate the conditional probability of $f_G(z_n = k)$ for all K possible values of z_n as follows,

$$\exp\left[\sum_{n=1}^{N}\sum_{n'\in\vartheta(n)} \sigma\ \delta(z_n - z_{n'})\right] \times \prod_{d=1}^{D} exp(\mathbf{A}_{dn}|\mathbf{W}_{dk}) \tag{10}$$

Then we use a Multinomial distribution to draw a sample for z_n. The conditional distribution of rate parameters $\mathbf{W}_{dk}$ is a Gamma distribution $f_G(\mathbf{W}_{dk})$ with following parameters,

$$Gamma\left(\mathbf{W}_{dk}|\alpha + n_k,\ \frac{\beta}{1 + \beta\sum_{p\in N_k}\mathbf{A}_{dp}}\right) \tag{11}$$

where N_k is the set of all pixels with the group number k, $N_K = \{n|\ z_n = k\}$, with the cardinality of n_k. The noise variance s is sampled from the following Inverse-Gamma distribution $f_G(s)$,

$$IG\left(s \mid \nu + \frac{LN}{2},\ \ \delta + \frac{\sum_{n=1}^{N}||\mathbf{y}_n - \mathbf{Ma}_n||_2^2}{2}\right) \tag{12}$$

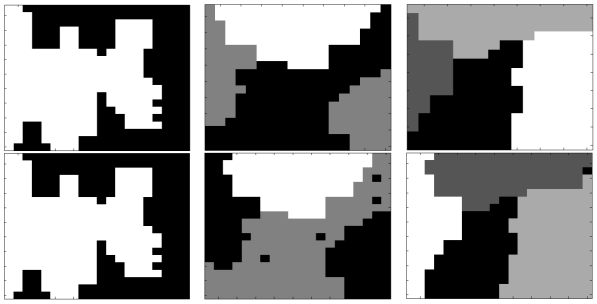
The final inference is performed according to produced samples. To have samples from the marginal distribution over fractional abundances, we can simply omit other variables of obtained samples from the Gibbs sampler. We call the proposed method BHUPC because it is a Bayesian method for simultaneously Hyperspectral Unmixing and Pixel Clustering

3 Experiments

In this section, we compare our proposed unmixer, BHUPC, with NCLS as a classic least square optimization with non-negativity constraint, SUnSAL [6] which applies the ℓ_1 minimizer for each pixel independently, CLSUnSAL [10] which applies ℓ_1 norm simultaneously to all pixels, SUnSAL-TV (TV) [9] which forces homogeneity among neighboring pixels, and our prior work LWSL [18] which applies a probabilistic joint spare model to each patch of an image. We evaluate methods on both synthetic, section 3.1 and real data, section 3.2. As common performance measures for unmixers [10, 9, 18], we use the Signal to Reconstruction Error (SRE) calculated as SRE $= E[\|\mathbf{x}\|_2^2]/E[\|\mathbf{x} - \mathbf{x}'\|_2^2]$ and Probability of Success (Ps) calculated as Ps $= P(\|\mathbf{x} - \mathbf{x}'\|_2^2/\|\mathbf{x}\|_2^2) \leq \theta$ where $\mathbf{x}$ and $\mathbf{x}'$ are the main and estimated signals respectively and θ is a threshold which is set to 3.16 (5 dB) [10, 9, 18]. SRE measure is reported in dB, SRE(dB) $= 10\log_{10}(\text{SRE})$.

3.1 Synthetic Data

Our spectral library is a subset of the standard USGS library, "Splib06" with 240 members which is recorded in 224



(a) (b) (c)

Fig. 2. Simulated images with (a) 2, (b) 3 unconnected and (c) 4 regions. Top maps show the ground-truth and down ones are the BHUPC estimations.

frequency bands and uniformly scattered between 0.4 and 2.5 μm. We simulated three 20×20 hyperspectral images with $\{2, 3, 4\}$ pysical regions. The map of regions are generated randomly. Pixels of each region use a same set of minerals in their mixtures which are randomly selected from the spectral library. To satisfy the non-negativity and sum-to-one constraints, the fractional abundances of each pixel are drawn from a Dirichlet distribution. Parameters of Dirichlet distribution are set to be uniform. As the noise of the real hyperspectral images is highly band correlated [8], simulated images are corrupted with 30 dB correlated Gaussian noise achieved by low-pass filtering the white Gaussian noise with a normalized cutoff frequency $5\pi/L$ [8]. For each scenario, the average of results over ten randomly generated images is reported.

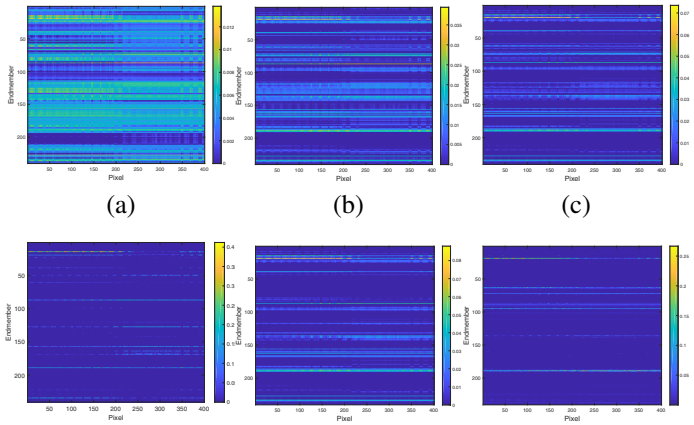
Table 1. SRE obtained for simulated images with 2 (R2), 3 (R3), and 4 (R4) regions.

Data	NCLS	SUnSAL	CLSUnSAL	TV	LWSL	BHUPC
R2	9.75	10.01	9.97	23.58	22.88	32.42
R3	4.21	5.24	5.18	7.19	7.70	10.26
R4	9.89	12.83	10.95	21.67	22.29	22.34

Fig. 2 shows the ground-truth partitioning maps and ones obtained by BHUPC, which have satisfying similar patterns. However by increasing the number of regions the accuracy is decreased which is because of decreasing the number of pixels in each region. The map in Fig. 2-(b) includes two unconnected regions with same mixtures. As shown BHUPC is able to detect the similarity between the mixture of pixels in these two regions and assign a same group number to them. Table 1 reports the SRE measure obtained by each method for simulated images in Fig. 2. In all cases the proposed BHUPC has the best performance. The obtained Ps measure for BHUPC in all simulated images is 1, which shows its success to achieve acceptable quality.

3.2 Real Data

We selected a 20×20 subset of Cuprite95 dataset recorded by Airborne Visible Infrared Imaging Spectrometer (AVIRIS) in 1995 which is used frequently for evaluation of unmixing



(a) (b) (c) (d) (d) (e)

Fig. 3. Fractional abundance maps obtained by (a) NCLS, (b) SUnSAL, (c) CLSUnSAL, (d) TV, (e) LWSL and (f) BHUPC which is the least noisy and the most sparse one.

algorithms [10, 18, 15]. It is available from Tetracorder 4.4 dataset. The spectral library is the same as the synthetic data. Because of the low signal to noise ratio and water absorption effects, we omitted the signals recorded in the frequency bands 1-2, 105-115, 150-170 and 223-224 in the real hyperspectral image and spectral library [10]. Fig. 3 shows the fractional abundance maps obtained by each method. As shown BHUPC has the most sparse matrix. In table 2 we report the SRE measure between the observed image and the one which is constructed by the inferred fractional abundances by each method according to the linear mixture model. BHUPC has the largest SRE. It shows that BHUPC can reconstruct the observed image with the highest accuracy by the most sparse solution. So it is more reliable than other algorithms. Finally the computational time of algorithms for each pixel of the real dataset is shown in table 3.

Table 2. SRE of reconstructed real images by each method.

NCLS	SUnSAL	CLSUnSAL	TV	LWSL	BHUPC
18.6030	23.0769	25.7528	25.7673	26.5021	27.1022

Table 3. Time needed for unmixing one pixel in second.

NCLS	SUnSAL	CLSUnSAL	TV	LWSL	BHUPC
0.005	0.024	0.019	0.104	0.01	2.42

4 Conclusion

In this work we introduced a joint sparse regression semi-supervised unmixer which simultaneously detect groups of pixels with similar mixtures. This intelligent grouping helps to groups pixels to regions which are more similar to homogeneous areas in a natural scene which leads to estimate hidden variables more precisely. For future work, we are interested to infer the number of regions automatically during the unmixing process. It is specifically useful for heterogeneous and unfamiliar areas

5 References

[1] José M Bioucas-Dias, Antonio Plaza, Nicolas Dobigeon, Mario Parente, Qian Du, Paul Gader, and Jocelyn Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 5, no. 2, pp. 354–379, 2012.

[2] Yuanchao Su, Jun Li, Antonio Plaza, Andrea Marinoni, Paolo Gamba, and Somdatta Chakravortty, "Daen: Deep autoencoder networks for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 7, pp. 4309–4321, 2019.

[3] Burkni Palsson, Johannes R Sveinsson, and Magnus O Ulfarsson, "Spectral-spatial hyperspectral unmixing using multitask learning," *IEEE Access*, vol. 7, pp. 148861–148872, 2019.

[4] Savas Ozkan and Gozde Bozdagi Akar, "Deep spectral convolution network for hyperspectral unmixing," in *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2018, pp. 3313–3317.

[5] Burkni Palsson, Magnus O Ulfarsson, and Johannes R Sveinsson, "Convolutional autoencoder for spectral-spatial hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, 2020.

[6] José M Bioucas-Dias and Mário AT Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," in *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2010 2nd Workshop on*. IEEE, 2010, pp. 1–4.

[7] Konstantinos Themelis, Athanasios A Rontogiannis, and Konstantinos Koutroumbas, "Semi-supervised hyperspectral unmixing via the weighted lasso.," in *ICASSP*, 2010, pp. 1194–1197.

[8] Marian-Daniel Iordache, José M Bioucas-Dias, and Antonio Plaza, "Sparse unmixing of hyperspectral data," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 6, pp. 2014–2039, 2011.

[9] Marian-Daniel Iordache, José M Bioucas-Dias, and Antonio Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 50, no. 11, pp. 4484–4502, 2012.

[10] Marian-Daniel Iordache, José M Bioucas-Dias, and Antonio Plaza, "Collaborative sparse regression for hyperspectral unmixing," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 52, no. 1, pp. 341–354, 2014.

[11] Rui Wang and Heng-Chao Li, "Nonlocal similarity regularization for sparse hyperspectral unmixing," in *Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International*. IEEE, 2014, pp. 2926–2929.

[12] Marian-Daniel Iordache, José M Bioucas-Dias, Antonio Plaza, and Ben Somers, "Music-csr: Hyperspectral unmixing via multiple signal classification and collaborative sparse regression," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 52, no. 7, pp. 4364–4382, 2014.

[13] Xiao Fu, Wing-Kin Ma, José M Bioucas-Dias, and Tsung-Han Chan, "Semiblind hyperspectral unmixing in the presence of spectral library mismatches," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 9, pp. 5171–5184, 2016.

[14] Nicolas Dobigeon, Jean-Yves Tourneret, and Chein-I Chang, "Semi-supervised linear spectral unmixing using a hierarchical bayesian model for hyperspectral imagery," *Signal Processing, IEEE Transactions on*, vol. 56, no. 7, pp. 2684–2695, 2008.

[15] Konstantinos E Themelis, Athanasios Rontogiannis, Konstantinos D Koutroumbas, et al., "A novel hierarchical bayesian approach for sparse semisupervised hyperspectral unmixing," *Signal Processing, IEEE Transactions on*, vol. 60, no. 2, pp. 585–599, 2012.

[16] Yoann Altmann, Marcelo Pereyra, and Jose Bioucas-Dias, "Collaborative sparse regression using spatially correlated supports-application to hyperspectral unmixing," *arXiv preprint arXiv:1409.8129*, 2014.

[17] Olivier Eches, Nicolas Dobigeon, and Jean-Yves Tourneret, "Enhancing hyperspectral image unmixing with spatial correlations," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 11, pp. 4239–4247, 2011.

[18] Seyyede Fatemeh Seyyedsalehi, Hamid R Rabiee, Ali Soltani-Farani, and Ali Zarezade, "A probabilistic joint sparse regression model for semisupervised hyperspectral unmixing," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 5, pp. 592–596, 2017.