

A VARIATIONAL BAYES ALGORITHM FOR JOINT-SPARSE ABUNDANCE ESTIMATION

Paris V. Giampouras, Konstantinos E. Themelis, Athanasios A. Rontogiannis, Konstantinos D. Koutroumbas

IAASARS, National Observatory of Athens, GR-15236, Penteli, Greece
{parisg,themelis,tronto,koutroum}@noa.gr

ABSTRACT

This paper presents a variational Bayesian scheme for semi-supervised unmixing on hyperspectral images that exploits the inherent spatial correlation between neighboring pixels. More specifically, a hierarchical Bayesian model that promotes a joint-sparse profile on the abundance vectors of adjacent pixels is developed and a computationally efficient variational Bayes algorithm is incorporated to perform Bayesian inference. The benefits of the proposed joint-sparse model are demonstrated via simulations on both synthetic and real data.

Index Terms— spatial correlation, joint-sparse model, semi-supervised unmixing, variational Bayesian inference

1. INTRODUCTION

Spectral unmixing (SU) has attracted considerable attention in the signal processing literature in recent years. SU can be considered as the process of identifying the spectrally pure signatures (also called endmembers) present in a hyperspectral scene, and the determination of their proportionate abundances in each pixel, [1]. Recently, the exploitation of spatial correlations in hyperspectral data as well as sparse representations of measured pixels have emerged as novel research ideas for the SU problem. These ideas are based on the reasonable assumptions that adjacent pixels in a hyperspectral scene are in all probability mixed using the same materials, and that any given pixel may be composed by only a few of the endmembers present in the scene. In this perspective, both these assumptions have sparked a new interest in the SU problem with the promise of producing better abundance estimation results.

However, sparsity and spatial correlations cognizant SU is a research field still at its infancy, although several methods have been proposed lately that employ either one of the two approaches, [2, 3]. In [2] a Bayesian model is presented that imposes the concept of sparsity as well as the nonnegativity constraint for the abundance estimation problem. A Bayesian

approach is also presented in [3], where a Markov random field is used to capture spatial correlations. Interestingly, a deterministic algorithm that employs both unmixing approaches has been recently presented in [4]. The abundance estimation problem is formulated either as a joint-sparse abundance estimation problem incorporating the sparsity promoting ℓ_1 -norm, or as a low rank representation problem of the abundance matrix.

In this paper we provide a Bayesian approach for the joint-sparse abundance estimation problem. To this end, we propose a hierarchical Bayesian model that a) comprises conjugate prior distributions that facilitate posterior inference, b) imposes the nonnegativity constraint on the abundance coefficients, and c) imposes a joint sparse profile on the abundance vectors that correspond to neighboring pixels. The proposed model can be considered as an extension of the sparsity promoting Bayesian model in [2] for enhancing unmixing with spatial correlations. A variational Bayes algorithm is then presented that performs inference for all model parameters. Besides its fast convergence, the proposed variational Bayes algorithm offers improved estimation performance, as verified by its application in homogeneous regions in both synthetic and real hyperspectral data.

2. PROBLEM FORMULATION

Let \mathbf{y} be a $M \times 1$ dimensional vector containing the spectral signature of a single pixel, where M is the number of spectral bands. In this paper, we consider a sliding square window centered at \mathbf{y} and containing besides \mathbf{y} its $K - 1$ closest neighboring pixels. Let \mathbf{Y} denote the $M \times K$ matrix that has as columns the spectral signatures \mathbf{y}_k 's of the pixels in the window, Φ be an $M \times N$ endmember matrix whose columns correspond to pure materials' spectra, and $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_K]$ be the $N \times K$ matrix consisting of K nonnegative abundance column vectors \mathbf{w}_k 's, where each \mathbf{w}_k is the N -dimensional abundance vector corresponding to the observed pixel \mathbf{y}_k ¹. By assuming the linear mixing model, the window's measured pixels are mixed as

$$\mathbf{Y} = \Phi \mathbf{W} + \mathbf{E}, \quad (1)$$

This research has been co-financed by the European Union (European Social Fund - ESF) and Greek national funds through the Operational Program Education and Lifelong Learning of the National Strategic Reference Framework (NSRF) - Research Funding Program: ARISTEIA- HSI-MARS-1413.

¹Apparently, one of the \mathbf{y}_k 's coincides with \mathbf{y}

where \mathbf{E} is a $M \times K$ matrix of zero mean, independent and identically distributed (i.i.d) Gaussian noise samples, with precision β .

In this paper the semi-supervised SU framework is considered. In this regard we assume that the endmember matrix Φ is known a priori, and that only a few of the N available endmembers are present in each \mathbf{y}_k in \mathbf{Y} . Thus, \mathbf{Y} adopts a *sparse* representation in Φ , expressed by a joint-sparse profile on the abundance vectors \mathbf{w}_k 's (equivalently, \mathbf{W} is a row sparse matrix). Our goal is to estimate the abundance matrix \mathbf{W} , subject to the nonnegativity and sparsity constraints. To this end, we develop a hierarchical Bayesian model that imposes both nonnegativity and sparsity constraints on the abundance matrix \mathbf{W} , and then, we present a fast variational Bayes algorithm to perform posterior inference.

3. BAYESIAN MODELING

The likelihood function of the observations \mathbf{Y} , determined by the assumed i.i.d. Gaussian noise \mathbf{E} in (1), is given by

$$p(\mathbf{Y} | \mathbf{W}, \beta) = \mathcal{N}(\mathbf{Y} | \Phi \mathbf{W}, \beta^{-1} \mathbf{I}_{MK}) \\ = \prod_{k=1}^K \frac{\beta^{M/2}}{(2\pi)^{M/2}} \exp \left[-\frac{\beta}{2} \|\mathbf{y}_k - \Phi \mathbf{w}_k\|_2^2 \right]. \quad (2)$$

In our Bayesian approach, the likelihood is complemented by suitable priors for the model parameters $\{\mathbf{W}, \beta\}$. As a prior for noise precision β a conjugate Gamma distribution is assumed, expressed as

$$p(\beta) = \mathcal{G}(\beta; \rho, \delta), \quad (3)$$

where ρ and δ represent the shape and scale parameters respectively (set to 10^{-6} in our experiments). For the abundance matrix \mathbf{W} , we define a two level hierarchical prior that a) retains the conjugacy among the prior distributions, b) imposes a joint-sparse profile on the abundance vectors \mathbf{w}_k 's and c) imposes nonnegativity. In the first level, a nonnegatively truncated Gaussian prior is used for each \mathbf{w}_k , i.e.,

$$p(\mathbf{W} | \alpha, \beta) = \prod_{k=1}^K \mathcal{N}_{R_+^N}(\mathbf{w}_k | \mathbf{0}, \beta^{-1} \mathbf{A}^{-1}), \quad (4)$$

where $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_N]^\top$ is the precision parameter vector common for all \mathbf{w}_k 's, $\mathbf{A} = \text{diag}(\alpha)$ is the corresponding diagonal matrix, and $\mathcal{N}_{R_+^N}$ signifies the N -variate normal distribution truncated at the nonnegative orthant of R^N , denoted by R_+^N , [2]. Note from (4) that a single precision α_n is used for each *row vector* \mathbf{w}_n^\top of \mathbf{W} . In the second level of hierarchy, the precision parameters α_n follow a Gamma distribution, i.e.,

$$p(\alpha_n) = \mathcal{G}(\alpha_n; \kappa, \theta), \quad (5)$$

where κ and θ are fixed hyperparameters (set to 10^{-6}). Utilizing (4) and (5), and integrating out the precision parameters

α_n 's, a heavy-tailed, nonnegatively truncated Student-t prior distribution arises for \mathbf{W} , expressed as

$$p(\mathbf{W}) = \prod_{n=1}^N \mathcal{St}_{2\kappa}^+(\mathbf{w}_n^\top; \mathbf{0}, \frac{\theta}{\kappa\beta} \mathbf{I}_K), \quad (6)$$

where $\mathcal{St}_\nu^+(\mathbf{w}; \zeta, \Upsilon)$ denotes the nonnegatively truncated *multivariate* Student-t distribution with location ζ , scale matrix Υ and ν degrees of freedom. The heavy-tailed Student-t distribution is widely known to promote sparsity. Notice in (6) that *independent* K -variate nonnegative Student-t priors are assigned to the rows \mathbf{w}_n^\top 's of \mathbf{W} . Hence, our Bayesian model imposes both a common sparsity profile for all abundance column vectors \mathbf{w}_k 's as well as the nonnegativity constraint on the abundance coefficients.

4. VARIATIONAL INFERENCE

The posterior distribution of $\mathbf{W}, \alpha, \beta$ is expressed by the Bayes' rule as

$$p(\mathbf{W}, \alpha, \beta | \mathbf{Y}) = \frac{p(\mathbf{Y} | \mathbf{W}, \beta) p(\mathbf{W} | \alpha, \beta) p(\alpha) p(\beta)}{\int p(\mathbf{Y}, \mathbf{W}, \alpha, \beta) d\mathbf{W} d\alpha d\beta}. \quad (7)$$

However, the integration at the denominator is intractable and, hence, we need to resort to approximate methods. Variational approximation is an efficient optimization tool for this task. Employing the mean-field theory, the posterior $p(\mathbf{W}, \alpha, \beta | \mathbf{Y})$ is approximated by a variational distribution $q(\mathbf{W}, \alpha, \beta)$, which is assumed to be of the factorized form

$$q(\mathbf{W}, \alpha, \beta) = \prod_{k=1}^K q(\mathbf{w}_k) \prod_{n=1}^N q(\alpha_n) q(\beta). \quad (8)$$

The estimation of $q(\mathbf{W}, \alpha, \beta)$ is carried out through the minimization of the Kullback-Leibler (KL) divergence between the true posterior $p(\mathbf{W}, \alpha, \beta | \mathbf{Y})$ and the $q(\mathbf{W}, \alpha, \beta)$ itself. The conjugacy of our Bayesian model in conjunction with the assumed posterior independence of the model parameters provides an inference process that is *tractable*. Straightforward computations (not shown here due to space limitations) yield that the posterior for each \mathbf{w}_k is the truncated Gaussian distribution

$$q(\mathbf{w}_k) = \mathcal{N}_{R_+^N}(\mathbf{w}_k | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}), \quad (9)$$

with

$$\boldsymbol{\mu}_k = \langle \beta \rangle \boldsymbol{\Sigma} \Phi^\top \mathbf{y}_k, \text{ and } \boldsymbol{\Sigma} = \langle \beta \rangle^{-1} \left(\Phi^\top \Phi + \mathbf{A} \right)^{-1}, \quad (10)$$

where the $\langle \cdot \rangle$ operator, denotes expectation of a random quantity with respect to the associated $q(\cdot)$. For each precision parameter α_n we get the following Gamma posterior

$$q(\alpha_n) = \mathcal{G} \left(\alpha_n; \frac{K}{2} + \kappa, \frac{\beta \langle \mathbf{w}_n^\top \mathbf{w}_n \rangle}{2} + \theta \right), \quad (11)$$

where \mathbf{w}_n^\top refers to the row vectors of \mathbf{W} . A conjugate Gamma posterior is also obtained for the noise precision β ,

$$q(\beta) = \mathcal{G}\left(\beta; \frac{K(M+N)}{2} + \rho, \frac{1}{2} \sum_{k=1}^K \langle \|\mathbf{y}_k - \Phi \mathbf{w}_k\|_2^2 \rangle + \frac{1}{2} \sum_{k=1}^K \langle \mathbf{w}_k^\top \mathbf{A} \mathbf{w}_k \rangle + \delta\right). \quad (12)$$

Note that each posterior distribution parameter in (9)-(10), (11), (12), is expressed in terms of the expectations of the remaining ones, that is, all parameters are interrelated. More specifically, the expected values of the above parameters are

$$\langle w_{nk} \rangle = \mu_{nk_{tr}}, \quad (\langle w_{nk}^2 \rangle = \mu_{nk_{tr}}^2 + \sigma_{n_{tr}}^2) \quad (13)$$

$$\langle \alpha_n \rangle = \frac{K + 2\kappa}{\langle \beta \rangle \langle \mathbf{w}_n^\top \mathbf{w}_n \rangle + 2\theta} \quad (14)$$

$$\langle \beta \rangle = \frac{K(M+N) + 2\rho}{\sum_{k=1}^K \langle \|\mathbf{y}_k - \Phi \mathbf{w}_k\|_2^2 \rangle + \sum_{k=1}^K \langle \mathbf{w}_k^\top \mathbf{A} \mathbf{w}_k \rangle + 2\delta}, \quad (15)$$

with

$$\langle \mathbf{w}_n^\top \mathbf{w}_n \rangle = \sum_{k=1}^K \langle w_{nk}^2 \rangle, \quad \langle \mathbf{w}_k^\top \mathbf{A} \mathbf{w}_k \rangle = \sum_{n=1}^N \langle w_{nk}^2 \rangle \langle a_n \rangle \quad (16)$$

$$\langle \|\mathbf{y}_k - \Phi \mathbf{w}_k\|_2^2 \rangle = \|\mathbf{y}_k - \Phi \boldsymbol{\mu}_{k_{tr}}\|_2^2 + \text{Trace}\{\Phi \boldsymbol{\Sigma}_{tr} \Phi^\top\}, \quad (17)$$

where w_{nk} denotes the n -th element of \mathbf{w}_k , $\boldsymbol{\mu}_{k_{tr}}$ and $\boldsymbol{\Sigma}_{tr}$ are the mean and covariance matrix of the truncated Gaussian distribution in (9), respectively, $\mu_{nk_{tr}}$ is the n -element of $\boldsymbol{\mu}_{k_{tr}}$, and $\sigma_{n_{tr}}^2$ is the n -th diagonal element of $\boldsymbol{\Sigma}_{tr}$. Hence, a cyclic iterative procedure can be established for the estimation of the means of the parameters, involving (13), (14), (15).

To avert the computationally demanding task of estimating $\boldsymbol{\Sigma}_{tr}$, the approximations $\langle w_{nk}^2 \rangle \approx \mu_{nk_{tr}}^2$ and $\langle \|\mathbf{y}_k - \Phi \mathbf{w}_k\|_2^2 \rangle \approx \|\mathbf{y}_k - \Phi \boldsymbol{\mu}_{k_{tr}}\|_2^2$, are used in our simulations. Moreover, to obtain a fast approximation on the mean $\boldsymbol{\mu}_{k_{tr}}$ of the multivariate truncated Gaussian in (9), we adopt the fast approximation method proposed in [2, 5]. To this end, an iterative scheme is utilized, that entails the cyclic computation of the one dimensional means of the per-coordinate conditional truncated Gaussian distributions². The proposed joint-sparse variational Bayes algorithm is detailed in Algorithm 1. The convergence of the proposed scheme is ensured by the convexity of the approximating distribution. Its complexity is $\mathcal{O}(KN^2)$ per iteration t . The algorithm needs only a few iterations to converge and produces row sparse estimates for the abundance matrix \mathbf{W} , as expected. It should be further emphasized that in the proposed variational Bayes scheme, all model parameters are directly inferred from the data and thus there is no need for parameter cross validation or fine-tuning.

²More details on this approximation in [5]

Algorithm 1 The proposed variational Bayes algorithm

Inputs \mathbf{Y}, Φ
Initialize β, α
 $\mathbf{T} = \Phi^\top \Phi$ and $\mathbf{Z} = \Phi^\top \mathbf{Y}$
for $t = 1, 2, \dots$ **do**
 $\mathbf{V}(t) = \mathbf{T} + \mathbf{A}(t)$
 for $n = 1, 2, \dots, N$ **do**
 Extract $\mathbf{v}_{-n}(t)$ and $v_{nn}(t)$ from $\mathbf{V}(t)$
 $\sigma_n^{*2} = \langle \beta \rangle^{-1} / v_{nn}$
 for $k = 1, 2, \dots, K$ **do**
 $\mu_{nk}^* = \frac{1}{v_{nn}} (z_{nk} - \mathbf{v}_{-n}^\top \boldsymbol{\mu}_{-nk_{tr}})$
 $\langle w_{nk} \rangle \equiv \mu_{nk_{tr}} = \mu_{nk}^* + \frac{\frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2} \frac{\mu_{nk}^{*2}}{\sigma_n^{*2}})}{1 - \frac{1}{2} \text{erfc}(\frac{\mu_{nk}^*}{\sqrt{2\sigma_n^{*2}}})} \sigma_n^*$
 end for
 $\langle \alpha_n \rangle = \frac{K+2\kappa}{\langle \beta \rangle \langle \mathbf{w}_n^\top \mathbf{w}_n \rangle + 2\theta}$
 end for
 $\langle \beta \rangle = \frac{K(M+N) + 2\rho}{\sum_{k=1}^K \langle \|\mathbf{y}_k - \Phi \mathbf{w}_k\|_2^2 \rangle + \sum_{k=1}^K \langle \mathbf{w}_k^\top \mathbf{A} \mathbf{w}_k \rangle + 2\delta}$
end for

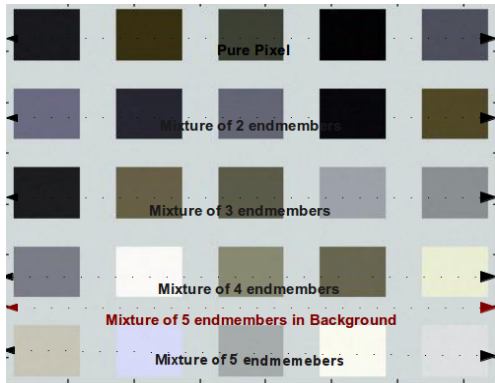
5. EXPERIMENTAL RESULTS

In this section we evaluate the performance of the proposed joint-sparse variational Bayes algorithm using both synthetic and real hyperspectral data. The benefits of taking into account spatial correlations are exposed by comparing the proposed joint-sparse algorithm a) to its single pixel counterpart, where each pixel is unmixed without exploiting the information from its adjacent neighbors (when Algorithm 1 is applied for $K = 1$), and the BiICE algorithm, proposed in [2]. In our experiments we used a 3×3 window (i.e. $K = 9$). In addition, we considered four performance measures. Specifically, root mean square error (RMSE) and signal to reconstruction error (SRE) measurements are used for the synthetic data set. On the other hand, when the true abundance values are not available, reconstruction error (RE) measurements and the spectral angle mapper (SAM) metric are used instead³.

To assess the accuracy of the proposed algorithm on the unmixing of homogeneous regions, a hyperspectral image is simulated according to the experimental settings of [4]. The simulated image was generated using a total of twelve endmember spectra, randomly selected from the USGS library, [6]. It comprises 150×150 pixels, with 25 homogeneous regions of size 20×20 . Each pixel is mixed using up to five endmembers, with the homogeneous regions having no more endmembers than the identically generated background pixels, as shown in Fig. 1(a). The unmixing results are provided in Fig. 1(b). We observe that the proposed method outperforms its single pixel counterpart, both in terms of the RMSE and SRE. Thus, it becomes evident that the exploitation of the spatial information of homogeneous images can result in improved abundance estimation performance.

Next, the Cuprite data set, [2, 3], is used to illustrate the

³Measure definitions can be found in [4].



(a) Synthetic Image

	Single Pixel	BiICE	Joint Sparse
RMSE(10^{-2})	2.5013	2.5585	1.8621
SRE(dB)	13.8106	13.7952	13.8285

(b) Results

Fig. 1. Synthetic Image and its results

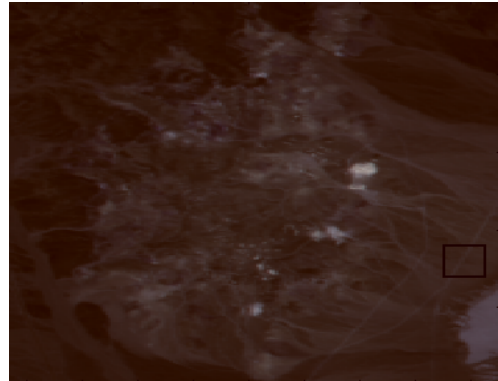
performance of the proposed scheme on real hyperspectral data. In this case, the Vertex Component Algorithm, [7], is first applied to the data set, in order to extract $N = 12$ pure endmembers. Then, the proposed algorithm is applied to a subregion in the Cuprite data set, highlighted by the rectangular area in Fig. 2(a). The unmixing results are provided in Fig. 2(b). Again it is confirmed that the utilization of the existing spatial correlation in the selected area leads to higher estimation accuracy for the joint-sparse model, as compared to its single pixel variant, in terms of both the SAM and RE.

6. CONCLUSIONS

In this paper, a hierarchical Bayesian methodology for the semi-supervised unmixing problem, where the spatial correlation is taken into account, is introduced. As shown from the experimental results, the embodiment of the spatial correlation to our model leads to enhanced performance in terms of the accuracy of the estimations, when working on homogeneous regions of the images.

7. REFERENCES

- [1] W.K. Ma, J.M. Bioucas-Dias, Tsung-Han Chan, N. Gillis, P. Gader, A.J. Plaza, A. Ambikapathi, and Chong-Yung Chi, "A signal processing perspective on hyperspectral unmixing: Insights from remote sensing," *Signal Processing Magazine, IEEE*, vol. 31, no. 1, pp. 67–81, Jan 2014.
- [2] K.E. Themelis, A.A. Rontogiannis, and K.D. Koutroumbas,



(a)Cuprite Image

	Single Pixel	BiICE	Joint Sparse
RE(10^{-3})	6.2712	5.9606	5.8313
SAM	0.1633	0.1555	0.1523

(b) Results

Fig. 2. Cuprite image and its experimental results

bas, "A novel hierarchical Bayesian approach for sparse semisupervised hyperspectral unmixing," *Signal Processing, IEEE Transactions on*, vol. 60, no. 2, pp. 585–599, Feb 2012.

- [3] O. Eches, N. Dobigeon, and J-Y Tourneret, "Enhancing hyperspectral image unmixing with spatial correlations," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 11, pp. 4239–4247, Nov 2011.
- [4] Q. Qu, N.M. Nasrabadi, and T.D. Tran, "Abundance estimation for bilinear mixture models via joint sparse and low-rank representation," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 52, no. 7, pp. 4404–4423, July 2014.
- [5] A.A. Rontogiannis, K.E. Themelis, and K.D. Koutroumbas, "A fast variational Bayes algorithm for sparse semi-supervised unmixing of omega/mars express data," in *5th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (IEEE WHISPERS)*, Florida, June 2013.
- [6] R. N. Clark, G. A. Swayze, R. Wise, K. E. Livo, T. M. Hoefen, R. F. Kokaly, and S. J. Sutley, "USGS digital spectral library," 2007, <http://speclab.cr.usgs.gov/spectral.lib06/ds231/datatable.html>.
- [7] J.M.P. Nascimento and J.M. Bioucas-Dias, "Vertex component analysis: a fast algorithm to unmix hyperspectral data," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 43, no. 4, pp. 898–910, April 2005.