2015 3rd International Workshop on Compressed Sensing Theory and its Applications to Radar, Sonar, and Remote Sensing (CoSeRa)

# A Sparse Reduced-Rank Regression Approach for Hyperspectral Image Unmixing

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

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

***Abstract*—In this paper we propose a semi-supervised method for hyperspectral image unmixing. Given a set of endmembers present in the image, we assume that (a) each pixel is composed of a subset of the available endmembers and (b) adjacent pixels are, in all possibility, correlated. Then, we define an inverse problem, where the abundance matrix to be estimated is assumed to be simultaneously *sparse* and *low-rank*. These assumptions give rise to a regularized linear regression problem, where a mixed penalty is enforced, comprising the weighted $\ell_1$ norm and an upper bound of the nuclear matrix norm. The resulting optimization problem is efficiently solved using a novel coordinate descend type unmixing algorithm. The estimation performance of the proposed scheme is illustrated in experiments conducted on both simulated and real data.**

## I. Introduction

Spectral unmixing is, undoubtedly, one of the most prominent and challenging problems in hyperspectral image processing, [1]. Spectral unmixing is the process of indentifying (a) the pure material spectra (called endmembers) that are present in the scene, and (b) their corresponding fractional proportions in each pixel (called abundances). The steps of this process are commonly handled as two individual tasks in the unmixing literature giving rise to endmember extraction and abundance estimation techniques, respectively.

In this paper we assume that a set of endmembers is given a priori and that the measured image pixel spectra are linear combinations of these endmembers. We focus on the inverse problem of abundance estimation, subject to the physical constraint of nonnegativity. Besides nonnegativity, recent novel ideas in the subject have put forth the exploitation of sparse representations, and the spatial correlation in hyperspectral images. Sparsity with respect to a given endmembers' dictionary has been explored in several works lately, e.g., [2]–[4]. Such a property is reasonable, since usually only a few endmembers contribute to each pixel's spectrum. As a result, abundance vectors shall have only a few non-zero entries. On the other hand, it has recently been shown that the estimation performance of algorithms can be improved by taking advantage of the spatial correlation that exists naturally among pixels lying in homogeneous regions of hyperspectral images. In this spirit, it is logically assumed that neighboring pixels share the same support set and thus their corresponding abundance vectors form a matrix that possesses a joint-sparse or low-rank structure, [5], [6].

Motivated by the above, in this work we present a new unmixing algorithm that imposes concurrently the two aforementioned constraints, i.e. sparsity and spatial correlation. To this end, for the unmixing of each single pixel a square window is employed that includes the pixel and its neighboring ones. Hence, abundance estimation is expressed as a matrix regression problem, with the corresponding abundance matrix constrained to be simultaneously sparse and low-rank. These two constraints are imposed by regularizing an initial least squares cost function with a sparsity inducing weighted $\ell_1$-norm and an upper bound of the low-rank promoting trace norm, [7], thus, giving rise to a *sparse reduced-rank regression* scheme, [8]. Then, a low complexity alternating coordinate descent type algorithm is developed, after splitting the original nonconvex cost function into mutually dependent convex subproblems. The proposed algorithm is robust and converges after a few iterations. Its performance is validated through experiments on both synthetic and real data.

## II. System Model and Problem Formulation

Let $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_K]$ denote the $L \times K$ observation matrix, consisting of the $L$-band spectral signatures of $K = \kappa \times \kappa$ neighboring pixels enclosed in a sliding square window of size $\kappa$. Assuming that the observations are produced via a linear regression process i.e., $\mathbf{Y}$ is modelled via the linear mixing model (LMM), we have

$$\mathbf{Y} = \mathbf{\Phi}\mathbf{W} + \mathbf{E}. \tag{1}$$

Matrix $\mathbf{\Phi} = [\phi_1, \phi_2, \ldots, \phi_N] \in \mathcal{R}_+^{L \times N}$ is the dictionary with the spectra of $N$ endmembers as its columns. The $N \times K$ matrix $\mathbf{W}$ is composed of the abundance column vectors of the $K$ pixels contained in the window. Finally, $\mathbf{E}$ stands for the additive stochastic noise matrix that corrupts the data, the elements of which are assumed to be zero mean Gaussian independent and identically distributed (i.i.d).

Regarding the abundance matrix $\mathbf{W}$, two additional constraints are imposed, termed in hyperspectral jargon as *nonnegativity* and *sum-to-one* constraints i.e.,

$$\mathbf{W} \geq 0, \text{ and } \mathbf{1}^T\mathbf{W} = \mathbf{1}^T. \tag{2}$$

where $\mathbf{1}$ is the all ones vector. In this paper, the controversial sum-to-one constraint is purposely relaxed (see [9] for a detailed explanation of this argument). Then given the pixels' spectra matrix $\mathbf{Y}$ and the endmembers' dictionary $\mathbf{\Phi}$, our

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.

978-1-4799-7420-7/15/$31.00 ©2015 IEEE

objective is to estimate the non-negative abundance matrix $\mathbf{W}$. Hence, based on the above, spectral unmixing is formulated as a constrained linear matrix regression problem.

Several methods recently proposed in the signal processing literature, advocate for parameterizing further the unmixing problem by adding more constraints on the abundance matrix $\mathbf{W}$. In this regard, *sparsity* constitutes a widespread hypothesis effectively incorporated in numerous unmixing algorithms, e.g. [4], [9]. This assumption implies that only a small subset of the $N$ endmembers contribute to the formation of the spectral signature of each pixel. In addition, recently proposed window-based unmixing algorithms e.g. [6], [10], have brought *spatial correlation* into play that doubtlessly exists in homogeneous regions of hyperspectral images. In that framework, spatial correlation can enter the scene via the abundance matrice $\mathbf{W}$, by imposing joint-sparsity and/or low-rankness on its structure.

In the sequel, we depart from the usual paradigm by taking *simultaneously* into account spatial correlation and sparsity. More specifically, the unmixing problem is now formulated as a sparse reduced-rank regression problem. This formulation entails the pursuit of an abundance matrix $\mathbf{W}$ that a) fits well to the data matrix $\mathbf{Y}$ with respect to a given dictionary $\mathbf{\Phi}$, b) has low rank and c) has only a few nonzero, positive elements. Accounting for all the above-mentioned requirements, $\mathbf{W}$ derives as the solution of the following convex optimization problem,

$$(\text{P1}): \hat{\mathbf{W}} = \underset{\mathbf{W}\in\mathcal{R}_{+}^{N\times K}}{\operatorname{argmin}} \frac{1}{2}\Big\{\|\mathbf{Y}-\mathbf{\Phi}\mathbf{W}\|_F^2 + \lambda_*\|\mathbf{W}\|_* + \lambda_1\|\mathbf{W}\|_1\Big\},$$

where $\|\cdot\|_F$ stands for the Frobenius norm. In (P1), both the nuclear (denoted as $\|\cdot\|_*$) and the $\ell_1$ ($\|\cdot\|_1$) norms, are utilized as convex surrogates of their nonconvex relevants (namely $\operatorname{rank}$ and $\ell_0$, [11]) for enforcing the low-rankness and sparsity on $\mathbf{W}$, respectively. Additionally, $\lambda_*$ and $\lambda_1$ are the reguralization parameters that control the importance of low-rankness and sparsity of the related terms.

A common issue that arises in optimization problems that involve rank minimization utilizing the nuclear norm, is the high computational complexity. This is a consequence of the fact that the computation of the nuclear norm requires a singular value decomposition (SVD) step, whose complexity when applied on a matrix of size $N\times K$ is $\mathcal{O}(N^2K+NK^2+K^3)$. This can be a serious impediment, especially in cases where the handling of large scale data is needed. Herein, we follow an alternative way for minimizing the rank of $\mathbf{W}$. Specifically, we adopt an explicit low-rank parameterization for $\mathbf{W}$, [7], which assumes that an upper bound $r$ for the rank of $\mathbf{W}$ ($r \geq rank(\mathbf{W})$) is a priori available. According to this parameterization, $\mathbf{W}$ is written as the product of two matrices $\mathbf{P}\in\mathcal{R}^{N\times r}$ and $\mathbf{Q}\in\mathcal{R}^{K\times r}$, i.e., $\mathbf{W}=\mathbf{P}\mathbf{Q}^T$. The gain from this parameterization is that now it holds,

$$\|\mathbf{W}\|_* \leq \frac{1}{2}\left(\|\mathbf{P}\|_F^2 + \mathbf{Q}\|_F^2\right), \tag{3}$$

that is, $\|\mathbf{W}\|_*$ is a tight lower bound of a sum of Frobenious norms. Incorporating this parameterization to (P1), we end up with the following optimization problem (P2), which is approximate to (P1)

$$(\text{P2}): \left\{\hat{\mathbf{W}},\hat{\mathbf{P}},\hat{\mathbf{Q}}\right\} = \underset{\mathbf{P},\mathbf{Q},\mathbf{W}}{\operatorname{argmin}}\mathcal{L}(\mathbf{P},\mathbf{Q},\mathbf{W}), \tag{4}$$

where,

$$\mathcal{L}(\mathbf{P},\mathbf{Q},\mathbf{W}) = \Big\{\frac{1}{2}\|\mathbf{Y}-\mathbf{\Phi}\mathbf{P}\mathbf{Q}^T\|_F^2 + \frac{\lambda_*}{2}\left(\|\mathbf{P}\|_F^2+\|\mathbf{Q}\|_F^2\right) + \lambda_1\|\mathbf{W}\|_1 + \frac{\mu}{2}\|\mathbf{W}-\mathbf{P}\mathbf{Q}^T\|_F^2\Big\}. \tag{5}$$

In (5), the term $\|\mathbf{W}-\mathbf{P}\mathbf{Q}^T\|_F^2$ imposes the fitting between $\mathbf{W}$ and its low-rank representation $\mathbf{P}\mathbf{Q}^T$, with $\mu$ being the weighting parameter of this term. (P2) is a nonconvex optimization problem with respect to matrices $\mathbf{W},\mathbf{P}$ and $\mathbf{Q}$. Moreover, the presence of the $\ell_1$ norm induces a nonsmooth behaviour of the objective function that must be suitably handled. In the following, a novel algorithm is presented that solves (P2) efficiently and provides an estimate of the abundance matrix $\mathbf{W}$.

## III. The Proposed Algorithm

To address the nonconvexity and nonsmoothness of (P2), the minimization problem is split into three distinct subproblems and an alternating coordinate descend type optimization strategy is utilized. To be more specific, we define the following

$$(\text{P2a}): \hat{\mathbf{P}} = \underset{\mathbf{P}}{\operatorname{argmin}}\mathcal{L}(\mathbf{P},\mathbf{Q},\mathbf{W}),$$
$$(\text{P2b}): \hat{\mathbf{Q}} = \underset{\mathbf{Q}}{\operatorname{argmin}}\mathcal{L}(\mathbf{P},\mathbf{Q},\mathbf{W}),$$
$$(\text{P2c}): \hat{\mathbf{W}} = \underset{\mathbf{W}}{\operatorname{argmin}}\mathcal{L}(\mathbf{P},\mathbf{Q},\mathbf{W}).$$

The above minimization problems are now convex and can be solved independently, as explained below. In addition, it turns out naturally that the solutions of these problems are mutually dependent, thus giving us the possibility to define a cyclic iterative scheme that provides a solution to the initial problem (P2), after convergence.

### A. Solution of (P2a)

Since $\mathcal{L}(\mathbf{P},\mathbf{Q},\mathbf{W})$ is differentiable with respect to $\mathbf{P}$, $\mathbf{P}$ can be obtained as the solution of the following equation

$$\hat{\mathbf{P}}: \frac{\partial\mathcal{L}(\mathbf{P},\mathbf{Q},\mathbf{W})}{\partial\mathbf{P}} = \mathbf{0}. \tag{6}$$

Calculating the derivative in (6) yields

$$\left(\mathbf{\Phi}^T\mathbf{\Phi}+\mu\mathbf{I}_N\right)\mathbf{P}\mathbf{Q}^T\mathbf{Q} + \lambda_*\mathbf{P} = \left(\mathbf{\Phi}^T\mathbf{Y}+\mu\mathbf{W}\right)\mathbf{Q}, \tag{7}$$

where $\mathbf{I}_N$ denotes identity matrix of size $N$. Setting $\mathbf{A} = \mathbf{\Phi}^T\mathbf{\Phi}+\mu\mathbf{I}_N$, $\mathbf{B}=\mathbf{Q}^T\mathbf{Q}$ and $\mathbf{C}=\left(\mathbf{\Phi}^T\mathbf{Y}+\mu\mathbf{W}\right)\mathbf{Q}$, eq. (7) is compactly written as

$$\mathbf{A}\mathbf{P}\mathbf{B}+\lambda_*\mathbf{P}=\mathbf{C}. \tag{8}$$

Eq. (8) is a form of the Stein matrix equation, that has been widely used in the field of control, [12]. To solve the Stein equation (8), we adopt the robust algorithm proposed in [13]. In this algorithm, matrix $\mathbf{A}$ is reduced to its Hessenberg form $\mathbf{H}=\mathbf{U}\mathbf{A}\mathbf{U}^T$ and matrix $\mathbf{B}$ is suitably replaced by its Schur

representation $\mathbf{S} = \mathbf{V}\mathbf{B}\mathbf{V}^T$, where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices. It should be noted that since, in our problem, $\mathbf{A}$ and $\mathbf{B}$ are symmetric, $\mathbf{H}$ and $\mathbf{S}$ turn out to be tri-diagonal and diagonal matrices, respectively. If we now multiply both sides of (8) from the left and the right by $\mathbf{U}^T$ and $\mathbf{V}$ respectively, and define $\mathbf{X} = \mathbf{U}\mathbf{P}\mathbf{V}^T$, (8) is rewritten as:

$$\mathbf{H}\mathbf{X}\mathbf{S} + \lambda_* \mathbf{X} = \mathbf{F}, \tag{9}$$

where $\mathbf{F} = \mathbf{U}^T\mathbf{C}\mathbf{V}$. Let us denote by $\mathbf{x}_i$, $\mathbf{f}_i$ the $i$th columns of $\mathbf{X}, \mathbf{F}$ and by $s_{ii}$ the $i$th diagonal element of $\mathbf{S}$. Then, we get from (9) the following system of equations

$$\left(s_{ii}\mathbf{H} + \lambda_* \mathbf{I}_N\right)\mathbf{x}_i = \mathbf{f}_i, \tag{10}$$

which due to the tri-diagonal form of $\mathbf{H}$, can be solved for $\mathbf{x}_i$ with only $\mathcal{O}(N)$ operations, [14]. After estimating $\mathbf{X}$, column by column, matrix $\mathbf{P}$ is obtained by the inverse transform

$$\hat{\mathbf{P}} = \mathbf{U}^T\mathbf{X}\mathbf{V}. \tag{11}$$

### *B. Solution of (P2b)*

Similarly to (P2a), the minimization problem (P2b) can be solved as

$$\hat{\mathbf{Q}} : \frac{\partial \mathcal{L}(\mathbf{P}, \mathbf{Q}, \mathbf{W})}{\partial \mathbf{Q}} = \mathbf{0}. \tag{12}$$

Utilizing $\mathbf{A}, \mathbf{C}$ defined above, (12) results to the closed form expression

$$\hat{\mathbf{Q}} = \mathbf{C}^T\mathbf{P}\left(\mathbf{P}^T\mathbf{A}\mathbf{P} + \lambda_*\mathbf{I}_r\right)^{-1}. \tag{13}$$

### *C. Solution of (P2c)*

The optimization problem (P2c) is employed in order to estimate matrix $\mathbf{W}$. Considering $\mathbf{P}$ and $\mathbf{Q}$ as constants, minimization of $\mathcal{L}(\mathbf{P}, \mathbf{Q}, \mathbf{W})$ with respect to $\mathbf{W}$ leads to

$$\hat{\mathbf{W}} = \underset{\mathbf{W}}{\operatorname{argmin}} \|\mathbf{W}\|_1 + \frac{\mu}{2\lambda_1}\|\mathbf{W} - \mathbf{P}\mathbf{Q}^T\|_F^2 \tag{14}$$

which is the proximal operator of the $\ell_1$ norm on $\mathbf{P}\mathbf{Q}^T$. Writing (14) as

$$\hat{\mathbf{W}} = \underset{\mathbf{W}}{\operatorname{argmin}} \sum_{n=1}^{N}\sum_{k=1}^{K}\left(|w_{nk}| + \frac{\mu}{2\lambda_1}\left(w_{nk} - \mathbf{p}_n^T\mathbf{q}_k\right)\right), \tag{15}$$

where $\mathbf{p}_n^T$ denotes the $n$th row of matrix $\mathbf{P}$ and $\mathbf{q}_k^T$ the $k$th row of $\mathbf{Q}$, $\hat{\mathbf{W}}$ can be determined via elementwise soft-thresholding, [15]. Thus,

$$\hat{w}_{nk} = \mathrm{SHR}_{\lambda_1/\mu}(\mathbf{p}_n^T\mathbf{q}_k) \tag{16}$$

where $\mathrm{SHR}_\lambda(x) = \mathrm{sign}(x)\max(0, |x| - \lambda)$. In this paper, in an attempt to enhance the imposition of sparsity, we use the weighted version of the $\ell_1$ norm thus matrix $\mathbf{D}$ is employed, whose elements are given by

$$d_{nk} = \frac{1}{|\hat{w}_{nk}| + \epsilon}. \tag{17}$$

$\hat{w}_{nk}$ in (17) is the estimate of $w_{nk}$ at the previous iteration of the algorithm and $\epsilon$ is a very small constant. The solution of (14) can be written in a more compact form as

$$\hat{\mathbf{W}} = \mathrm{SHR}_{\mathbf{D}(\lambda_1/\mu)}(\mathbf{P}\mathbf{Q}^T) \tag{18}$$

The previous analysis summarizes to the new ALternating Minimization Sparse Low-Rank Unmixing (ALMSpLRU) algorithm shown in Algorithm 1, below. It should be noted that the aforementioned nonnegativity constraint is imposed by projecting $\mathbf{W}$ onto the nonnegative orthant of $\mathcal{R}^{N\times K}$, denoted as $\mathcal{R}_+^{N\times K}$ (step 9). In Algorithm 1, the computational complexity of each step is also included. We observe that the most computationally demanding steps, are steps 4, 6 and 7 requiring $\mathcal{O}(N^2 r)$ operations per iteration. However, this computational cost would be lower compared to that of a SVD-based scheme, especially as the size $\kappa = \sqrt{K}$ of the square window increases. Moreover, as verified by extensive simulations, the proposed algorithm is robust, and converges in a small number of iterations.

**Algorithm 1** ALMSpLRU

Inputs $\mathbf{Y}$, $\mathbf{\Phi}$
Initialize parameters $\lambda_1, \lambda_*, \mu$
Initialize $\mathbf{P}, \mathbf{Q}, \mathbf{W} = \mathbf{P}\mathbf{Q}^T$
Set $\mathbf{A} = \mathbf{\Phi}^T\mathbf{\Phi} + \mu\mathbf{I}_N$
Set $[\mathbf{U}, \mathbf{H}] = \mathrm{hess}(\mathbf{A})$ ▷ Hessenberg form of $\mathbf{A}$
Set $\mathbf{T} = \mathbf{\Phi}^T\mathbf{Y}$
**repeat**
1: $\mathbf{B} = \mathbf{Q}^T\mathbf{Q}$, $\mathcal{O}(Nr^2)$
2: $[\mathbf{V}, \mathbf{S}] = \mathrm{schur}(\mathbf{B})$, ▷ Schur form of $\mathbf{B}$ $\mathcal{O}(r^3)$
3: $\mathbf{C} = (\mathbf{T} + \mu\mathbf{W})\,\mathbf{Q}$, $\mathcal{O}(NKr)$
4: $\mathbf{F} = \mathbf{U}^T\mathbf{C}\mathbf{V}$, $\mathcal{O}(N^2 r)$
5: $(s_{ii}\mathbf{H} + \lambda_*\mathbf{I}_N)\,\mathbf{x}_i = \mathbf{f}_i$, ▷ $i = 1, 2, \ldots, r$ $\mathcal{O}(rN)$
6: $\mathbf{P} = \mathbf{U}\mathbf{X}\mathbf{V}^T$, $\mathcal{O}(N^2 r)$
7: $\mathbf{Q} = \mathbf{C}^T\mathbf{P}\left(\mathbf{P}^T\mathbf{A}\mathbf{P} + \lambda_*\mathbf{I}_r\right)^{-1}$, $\mathcal{O}(N^2 r)$
8: $\mathbf{W} = \mathrm{SHR}_{\mathbf{D}(\lambda_1/\mu)}(\mathbf{P}\mathbf{Q}^T)$, $\mathcal{O}(NK)$
9: Project $\mathbf{W}$ onto $\mathcal{R}_+^{N\times K}$ $\mathcal{O}(NK)$
**until** convergence
Output $\hat{\mathbf{W}}$

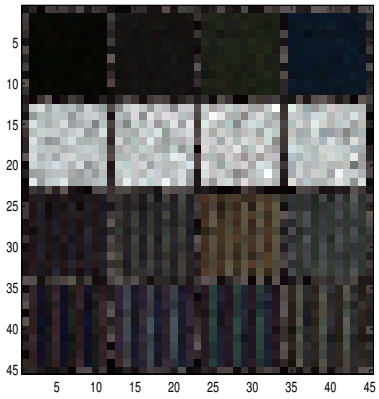
## IV. Experimental Results

In this section, the performance of ALMSpLRU is evaluated using both simulated and real data experiments. The proposed algorithm is compared with two state-of-the-art unmixing algorithms, namely CSUnSAL, [3], and MMV-ADMM, [6]. CSUnSAL is a single-pixel based sparse unmixing algorithm that allows for the non-negativity constraint. On the other hand, MMV-ADMM, similarly to ALMSpLRU, is a window-based method, that aims at exploiting the presumed spatial correlation of homogeneous regions of hyperspectral images. To this end, MMV-ADMM seeks abundance matrices characterized by a joint-sparse structure.

For the experiment conducted on synthetic data, the mean square error (MSE) and the signal to reconstruction error (SRE) were employed as objective metrics for evaluating the performance of the three algorithms. MSE is defined as,

$$\mathrm{MSE} = \frac{1}{NP}\sum_{i=1}^{P}\|\hat{\mathbf{w}}_i - \mathbf{w}_i\|_2, \tag{19}$$

where $N, P$ are the total number of the endmembers and the pixels in the image, respectively. $\hat{\mathbf{w}}_i$ stands for the estimated




(a) Synthetic Image, 16 blocks of size $10 \times 10$ pixels each.

| row | column | | | |
|---|---|---|---|---|
| | $1^{st}$ | $2^{nd}$ | $3^{rd}$ | $4^{th}$ |
| joint sparse - $1^{st}$ | $(4,1)$ | $(8,2)$ | $(12,3)$ | $(16,4)$ |
| low-rank - $2^{nd}$ | $(100,1)$ | $(100,2)$ | $(100,3)$ | $(100,4)$ |
| sparse & low-rank - $3^{rd}$ | $(4,2)$ | $(8,2)$ | $(12,2)$ | $(16,2)$ |
| sparse & low-rank - $4^{th}$ | $(4,3)$ | $(8,3)$ | $(12,3)$ | $(16,3)$ |

(b) Structure of $\mathbf{W}$ in each block of the synthetic image, each cell contains the pair : $\left(\frac{|\text{supp}(\mathbf{W})|}{NK}\%, \text{rank}(\mathbf{W})\right)$.

| **Algorithm** | $1^{st}$ row | | $2^{nd}$ row | | $3^{rd}$ row | | $4^{th}$ row | |
|---|---|---|---|---|---|---|---|---|
| | MSE | SRE | MSE | SRE | MSE | SRE | MSE | SRE |
| CSunSAL | 0.18 | 15.25 | 1.27 | 12.36 | 0.39 | 15.76 | 0.54 | 14.23 |
| MMV-ADMM | 0.17 | 15.78 | 1.47 | 12.97 | 0.42 | 13.84 | 0.55 | 13.50 |
| ALMSpLRU | 0.09 | 21.17 | 0.46 | 18.87 | 0.05 | 24.29 | 0.15 | 20.60 |

(c) MSE ($10^{-2}$) and SRE (dB) results on synthetic image for each row.

Fig. 1: Structure of the synthetic image and results.

abundance vector of the $i$th pixel (which for MMV-ADMM and ALMSpLRU coincides with the central column vector of the estimated abundance matrix $\mathbf{W}$), and $\mathbf{w}_i$ represents the actual abundance vector. SRE is known to reflect the ratio of the power of the estimate versus the power of the estimation error, and is given by

$$\text{SRE} = 10\log_{10}\left(\frac{\frac{1}{P}\sum_{i=1}^{P}\|\hat{\mathbf{w}}_i\|_2^2}{\frac{1}{P}\sum_{i=1}^{P}\|\hat{\mathbf{w}}_i - \mathbf{w}_i\|_2^2}\right). \qquad (20)$$

In this work, for the experiments conducted both on synthetic and real data, the size $\kappa$ of the square window and the upper bound of the rank $r$ are set to 3 and 5, respectively.

### A. Synthetic data

In an attempt to explore the estimation performance of the proposed ADSpLRU algorithm when either or both the assumptions of sparsity and low-rankness are met for the abundance matrix $\mathbf{W}$, a simulated hyperspectral datacube is generated as described below. First, the spectral signatures of $N = 25$ endmembers are randomly selected from the USGS library, [16], to form our dictionary $\mathbf{\Phi}$. The reflectances are observed in $L = 224$ spectral bands uniformly distributed in the range $0.4$ to $2.5\mu$m. Next, a hyperspectral image of size $45 \times 45$ is produced according to the linear mixing model described in (1). Finally, all the pixels of the image are corrupted with i.i.d zero mean Gaussian noise of SNR = 30dB.

As shown in Fig. 1a, the synthetic image consists of 4 "block" rows, with each such row corresponding to a distinct structure of the abundance matrices $\mathbf{W}$'s. To be more specific, the first row is generated by joint-sparse $\mathbf{W}$'s, the second row by only low-rank $\mathbf{W}$'s and rows 3 and 4 are produced by simultaneously sparse and low-rank $\mathbf{W}$'s. Additionally, each row consists of 4 blocks of pixels of size $10 \times 10$. Each block is generated by abundance matrices of different sparsity-level and rank. Further details on the specific parameterization of the abundance matrices in each block are given in the table of Fig. 1b.

The sparsity imposing parameter $\lambda_1$ used in all the tested algorithms as well as parameter $\lambda_*$ of ALMSpLRU that imposes low-rankness, were fine tuned with respect to MSE. The regularization parameter $\mu$ is set to $10^{-2}$ for all the algorithms. The table in Fig. 1c summarizes the results obtained for all the rows of the synthetic image. As it is shown in this table, ALMSpLRU outperforms CSUnSAL and MMV-ADMM both in terms of MSE and SRE. Interestingly, this happens not only for the sparse and low-rank case, but also when the abundance matrix that produces the data is either sparse or low-rank only.

### B. Real data

This section illustrates the performance of the proposed algorithm on a real hyperspectral image. The image under consideration is a $150 \times 150$ pixels sub-region of the Salinas vegetation scene acquired by AVIRIS sensor over Salinas Valley in California. Salinas hyperspectral image consists of $L = 204$ spectral bands and its spatial resolution is 3.7 meters. The endmembers dictionary $\mathbf{\Phi}$ is composed of 17, manually selected, pure pixel's spectral signatures.

Fig. 2 shows the abundance maps of four endmembers that correspond to different plant species, i.e., grapes, broccoli 1, broccoli 2 and lettuce 3, as estimated by ALMSpLRU, CSUnSAL and MMV-ADMM. By a careful inspection, we see that all the algorithms provide abundance maps with similar patterns. Additionally, the proposed algorithm seemingly produces maps with more detailed information especially concerning the endmembers broccoli 1 and broccoli 2.

## V. Conclusion

This paper presented a novel semi-supervised hyperspectral unmixing algorithm. The proposed algorithm aims at exploiting both the sparse representation of the pixels spectra, and the spatial correlation of homogeneous regions of hyperspectral images. Thus, we proposed to seek abundance matrices that are simultaneously sparse and of low-rank. A weighted $\ell_1$ norm and an upper bound of the trace norm were utilized for imposing sparsity and low-rankenss respectively. The altenating minimaztion strategy that was developed for solving

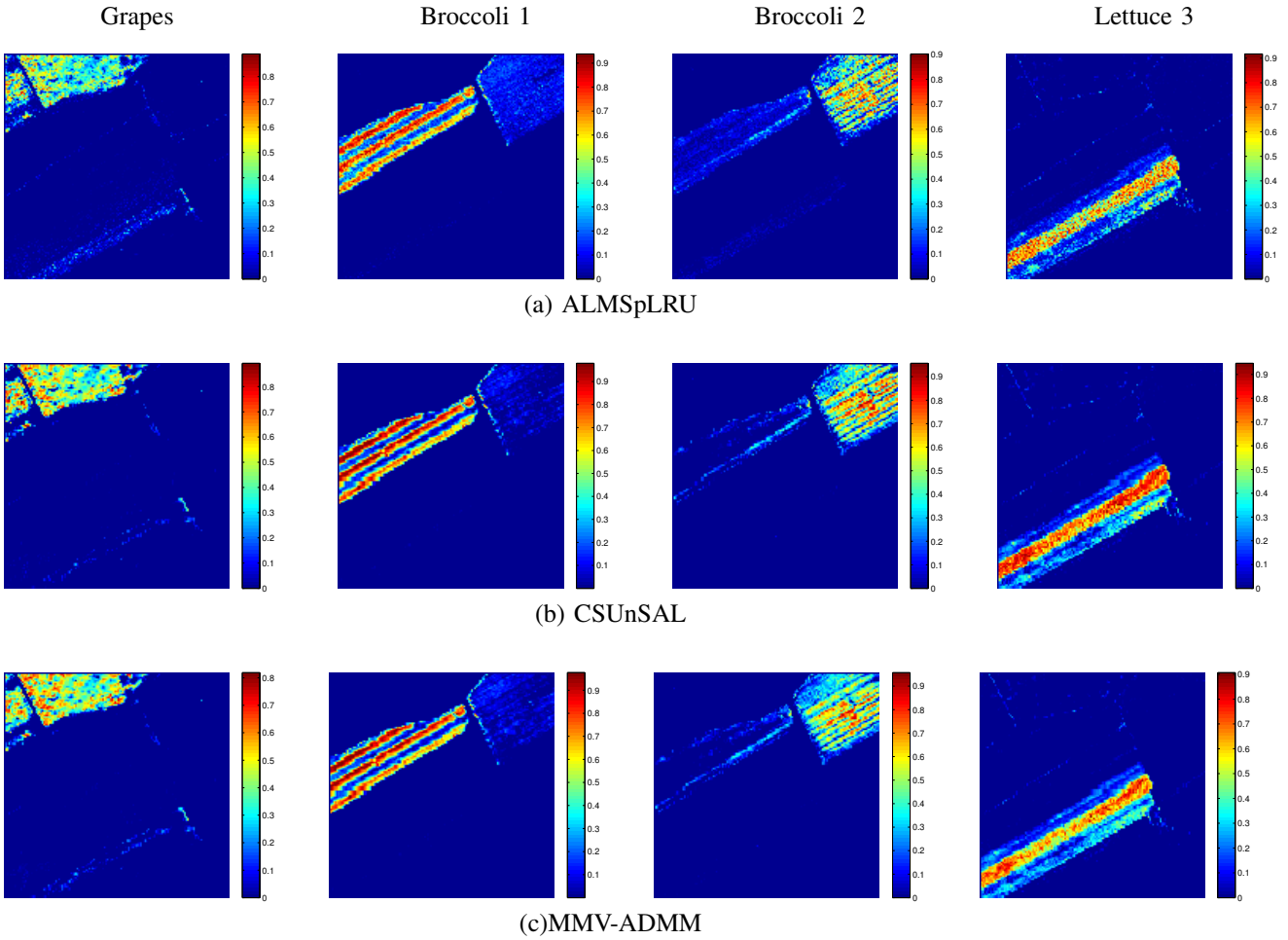



Fig. 2: Abundance maps of a subregion of the Salinas hyperspectral image.

the novel optimization problem, was shown to offer promising hyperspectral unmixing results on experiments conducted both on synthetic and real data.

## 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, "Semi-supervised hyperspectral unmixing via the weighted lasso," in *ICASSP*, 2010, pp. 1194–1197.

[3] J. M. Bioucas-Dias and M. A.T. Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," in *WHISPERS, 2010 2nd Workshop on*, 2010, pp. 1–4.

[4] K. E. Themelis, A. A. Rontogiannis, and K. D. Koutroumbas, "A novel hierarchical Bayesian approach for sparse semisupervised hyperspectral unmixing," *Signal Processing, IEEE Transactions on*, vol. 60, no. 2, pp. 585–599, Feb. 2012.

[5] M-D Iordache, J. M. Bioucas-Dias, and A. Plaza, "Collaborative sparse regression for hyperspectral unmixing," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 52, no. 1, pp. 341–354, 2014.

[6] 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.

[7] N. Srebro and A. Shraibman, "Rank, trace-norm and max-norm," in *Learning Theory*, pp. 545–560. Springer, 2005.

[8] L. Chen and J. Z. Huang, "Sparse reduced-rank regression for simultaneous dimension reduction and variable selection," *Journal of the American Statistical Association*, vol. 107, no. 500, pp. 1533–1545, 2012.

[9] M-D Iordache, J. M. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 6, pp. 2014–2039, 2011.

[10] P. V. Giampouras, K. E. Themelis, A. A. Rontogiannis, and K. D. Koutroumbas, "A variational Bayes algorithm for joint-sparse abundance estimation," in *WHISPERS, 2014 6th Workshop on*, 2014.

[11] F. Bach, R. Jenatton, J. Mairal, G. Obozinski, et al., "Convex optimization with sparsity-inducing norms," *Optimization for Machine Learning*, pp. 19–53, 2011.

[12] H. Abou-Kandil, *Matrix Riccati equations: in control and systems theory*, Springer Science & Business Media, 2003.

[13] G. Golub, S. Nash, and C. Van Loan, "A hessenberg-schur method for the problem ax + xb= c," *Automatic Control, IEEE Transactions on*, vol. 24, no. 6, pp. 909–913, Dec. 1979.

[14] C. de Boor, *Elementary numerical analysis*, McGraw-Hill, 1972.

[15] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 123–231, 2013.

[16] 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.