Simultaneously Sparse and Low-Rank Abundance Matrix Estimation for Hyperspectral Image Unmixing

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

Abstract—In a plethora of applications dealing with inverse problems, e.g., image processing, social networks, compressive sensing, and biological data processing, the signal of interest is known to be structured in several ways at the same time. This premise has recently guided research into the innovative and meaningful idea of imposing multiple constraints on the unknown parameters involved in the problem under study. For instance, when dealing with problems whose unknown parameters form sparse and low-rank matrices, the adoption of suitably combined constraints imposing sparsity and low rankness is expected to yield substantially enhanced estimation results. In this paper, we address the spectral unmixing problem in hyperspectral images. Specifically, two novel unmixing algorithms are introduced in an attempt to exploit both spatial correlation and sparse representation of pixels lying in the homogeneous regions of hyperspectral images. To this end, a novel mixed penalty term is first defined consisting of the sum of the weighted ℓ_1 and the weighted nuclear norm of the abundance matrix corresponding to a small area of the image determined by a sliding square window. This penalty term is then used to regularize a conventional quadratic cost function and impose simultaneous sparsity and low rankness on the abundance matrix. The resulting regularized cost function is minimized by: 1) an *incremental proximal sparse and low-rank* unmixing algorithm; and 2) an algorithm based on the *alternating direction method of multipliers*. The effectiveness of the proposed algorithms is illustrated in experiments conducted both on simulated and real data.

Index Terms—Abundance estimation, alternating direction method of multipliers (ADMM), hyperspectral images (HSIs), proximal methods, semisupervised spectral unmixing (SU), simultaneously sparse and low-rank matrices.

P. V. Giampouras is with the Department of Informatics and Telecommunications, University of Athens, 157 84 Athens, Greece, and also with the Institute for Astronomy, Astrophysics, Space Applications and Remote Sensing (IAASARS), National Observatory of Athens, 15236 Athens, Greece (e-mail: parisg@noa.gr).

K. E. Themelis, A. A. Rontogiannis, and K. D. Koutroumbas are with the Institute for Astronomy, Astrophysics, Space Applications and Remote Sensing (IAASARS), National Observatory of Athens, 15236 Athens, Greece (e-mail: themelis@noa.gr; tronto@noa.gr; koutroum@noa.gr).

I. INTRODUCTION

SPECTRAL unmixing (SU) of hyperspectral images (HSIs) has attracted considerable attention in recent years both in research and applications. SU can be considered as the process of 1) identifying the spectral signatures of the materials (*endmembers*) whose mixing generates the (so-called) *mixed* pixels of an HSI and 2) deriving their corresponding fractions (*abundances*) in the formation of each HSI pixel [1]. The latter constitutes the so-called *abundance vector* of the pixel. Those two tasks have given rise to a plethora of methods tackling either one or both of them. Diverse statistical and geometrical approaches have been lately put forward in the literature addressing the first step, which is commonly known as endmembers' extraction (e.g., [2] and [3]). On the other hand, there have been many research works that assume that the spectral signatures of the endmembers are available and focus on the abundance estimation task. Algorithms that fall into this class need to make a fundamental assumption concerning the inherent mixing process that generates the spectral signatures of the HSI pixels.

In view of the latter, the linear mixing model (LMM) holds a dominant position being widely adopted in numerous state-of-the-art unmixing algorithms (see, e.g., [1] and the references therein). More specifically, these algorithms are based on the premise that the pixels' spectral signatures are generated by a linear combination of endmembers' spectra contained in a predefined set, which is usually termed as *endmembers' dictionary*. Abundance estimation is henceforth treated as a linear regression problem. The LMM has prevailed over other models, due to its simplicity and mathematical tractability. Physical considerations that naturally arise impose various constraints on the unmixing problem. In this context, the so-called *abundance nonnegativity* and the *abundance sum-to-one* constraints are usually adopted. That said, unmixing can be viewed as a constrained linear regression problem.

In an attempt to achieve better abundance estimation results, recent novel ideas promote the incorporation of further prior knowledge in the unmixing problem. In light of this, several methods bring into play the *sparsity* assumption [4]–[7]. Its adoption is justified by the fact that (in practice) only a few of the available endmembers participate in the formation of a given pixel, particularly in the case of large-size endmembers' dictionaries. In other words, it is envisaged that pixels' spectral signatures accept sparse representations with respect to a given endmembers' dictionary; that is, they are deemed to have only

a few nonzero values. Practically, sparsity is imposed on abundances via the ℓ_1-norm regularization [4]–[6] when a deterministic approach is followed. On the other hand, in Bayesian schemes, appropriate sparsity inducing priors are adopted for the abundance vectors [7], [8]. *Spatial correlation* is another constraint that has been recently incorporated in the unmixing process, offering stimulating results [9]–[11]. In that vein, the additional information that exists in the homogeneous regions of HSIs is subject to exploitation. In fact, in such regions, there is a high degree of correlation among the spectral signatures of neighboring pixels. It is hence anticipated that there should also be correlation among the abundance vectors corresponding to these pixels. This has led to the development of novel unmixing schemes, whereby the information provided by the neighboring pixels is taken into account in the abundance estimation of each single pixel.

In this spirit, a collaborative deterministic scheme, which is termed CLSUnSAL, was recently proposed in [11], which uses a wealth of information stemming from all the pixels of the examined HSI. CLSUnSAL adopts dictionaries consisting of a large number of endmembers. Then, it assumes that spatial correlation translates into abundance vectors sharing the same support set, i.e., presenting a similar sparsity pattern. Thus, the matrix whose columns are the abundance vectors of all HSI pixels (called *abundance matrix*) should meaningfully be of a joint-sparse structure.[1] To impose joint sparsity, CLSUnSAL applies a $\ell_{2,1}$-norm on the sought abundance matrix, which is then used to penalize a suitably defined quadratic cost function. Minimization of the resulting regularized cost function is performed by an alternating direction method of multipliers (ADMM) [13]. A similar perspective is followed in [10], but in a "localized" fashion. Specifically, Qu *et al.* [10] proposed the use of a 3×3 square window that slides in the spatial dimensions of the image. The abundance vector of the central pixel is then inferred by taking into account the spectral signatures of the adjacent pixels contained in the window. Based on this idea, two algorithms are derived: the MMV-ADMM, which, in a similar fashion as CLSUnSAL, seeks joint-sparse abundance matrices utilizing the $\ell_{2,1}$-norm, and the LRR algorithm, which promotes a low-rank structure on the abundance matrix. In fact, the LRR algorithm presents an alternative way of modeling the spatial correlation among neighboring pixels. That is, it assumes that the correlation among pixels' spectral signatures is reflected as linear dependence among their corresponding abundance vectors. Apparently, the matrix formed by these abundance vectors should be of low rank. That said, a nuclear norm is imposed on the abundance matrix, and a properly adapted augmented Lagrangian cost function is minimized in an alternating minimization fashion.

In this paper, we introduce a novel idea for performing abundance estimation in HSIs under the LMM, which simultaneously takes spatial correlation and sparsity into consideration. As in [10], we also utilize a $\kappa \times \kappa$ square sliding window with κ odd, and we consider the spectral signatures of adjacent pixels lying in it. Departing from the usual paradigm, we

propose to seek for κ^2-column abundance matrices that are *simultaneously sparse and of low rank*. SU is thus formulated as a *sparse reduced-rank regression* problem [14]. As previously stated, low rankness naturally arises in abundance matrices corresponding to relatively homogeneous regions, due to the linear dependence of the respective abundance vectors. At the same time, sparsity is a reasonable hypothesis that still holds independently, as previously explained, within each individual abundance vector. Broadly speaking, imposing multiple structures on the same mathematical object when dealing with inverse problems is a strategy still in its very infancy in the signal processing and machine learning literature [15]–[18]. The aforementioned sparsity and low-rank constraints give rise to a mixed penalty term that regularizes a least squares fitting function through the weighted ℓ_1-norm and the weighted trace norm of the abundance matrices, respectively. To minimize the cost function, two novel iterative algorithms are proposed, namely, the *incremental proximal sparse low-rank* unmixing algorithm (IPSpLRU) and the *alternating direction sparse and low-rank* unmixing algorithm (ADSpLRU). As implied by its name, IPSpLRU is based on an *incremental* application of proximal operators on the individual terms that compose the cost function. Such a newly introduced approach [19] is of particular interest among others when solving optimization problems characterized by multiple constraints. Optimization problems of this type can usually be expressed by cost functions consisting of multiple component functions of the unknown parameter, as is also the case with the problem studied in this paper. Motivated by this, we have adopted the incremental proximal strategy in [19] that, due to its nature, leads to a computationally efficient iterative estimation scheme. On the other hand, ADSpLRU is an ADMM-based approach properly adapted to our problem formulation. ADSpLRU offers, in general, better estimation performance than IPSpLRU, although at a higher computational cost. The proposed algorithms are compared with state-of-the-art unmixing techniques, and their effectiveness is demonstrated via extensive simulated and real-data experiments.

Notation: Matrices are represented as boldface uppercase letters, e.g., $\mathbf{X}$, and column vectors are represented as boldface lowercase letters, e.g., $\mathbf{x}$, whereas the ith component of vector $\mathbf{x}$ is denoted by x_i, and the ijth element of matrix $\mathbf{X}$ is denoted by x_{ij}. Moreover, ^{T} denotes transposition, $\mathbf{I}_N$ is the $N \times N$ identity matrix, $\mathbf{0}$ is a zero matrix with respective dimensions, $\mathbf{1}$ denotes the all-1's vector, $\mathrm{rank}(\mathbf{X})$ is the rank of $\mathbf{X}$, $\mathrm{tr}[\mathbf{X}]$ denotes the trace of matrix $\mathbf{X}$, $\mathrm{diag}(\mathbf{x})$ is a diagonal matrix with the elements of vector $\mathbf{x}$ on its diagonal, $\sigma_i(\mathbf{X})$ is the ith largest singular value of $\mathbf{X}$, $\| \cdot \|_2$ is the standard ℓ_2 (Euclidean) vector norm, $\|\mathbf{X}\|_* = \mathrm{Tr}(\sqrt{\mathbf{X}^T\mathbf{X}}) = \sum_{k=1}^{\mathrm{rank}(\mathbf{X})} \sigma_i(\mathbf{X})$ denotes the nuclear norm (or trace norm), $\|\mathbf{X}\|_1 = \sum_i \sum_j |x_{ij}|$ is the sum of the absolute values of all entries of $\mathbf{X}$ (called the ℓ_1-norm), and $\|\mathbf{X}\|_F = \sqrt{\sum_i \sum_j x_{ij}^2}$ stands for the Frobenius norm. $\mathcal{N}(\cdot)$ denotes the Gaussian distribution. Moreover, $\mathbb{R}^k$ stands for the k-dimensional Euclidean space, and $\mathbb{R}^k_+$ denotes the k-dimensional nonnegative orthant. The matrix inequality $\mathbf{X} \geq \mathbf{Y}$ declares element-wise comparison, and $\odot$ stands for component-wise multiplication between matrices of the same size.

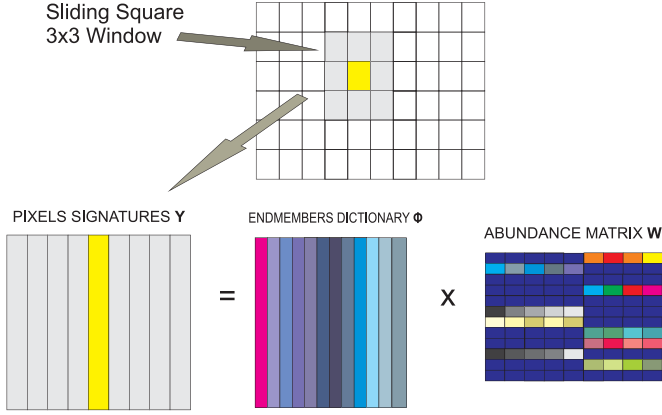[1]A joint-sparse Bayesian unmixing scheme was also presented in [12].

Fig. 1. Graphical illustration of the sliding window approach of our unmixing algorithms. The abundance matrix is considered sparse and of low rank (in this example, rank = 2). Blue cells in matrix $\mathbf{W}$ represent zero values.

II. PROBLEM FORMULATION

We consider an L-spectral band HSI, whose pixels are composed of N endmembers. Let $\mathbf{\Phi} = [\phi_1, \phi_2, \dots, \phi_N]$ stand for the $L \times N$ endmembers' dictionary, where $\phi_i \in \mathbb{R}_+^L$, $i = 1, 2, \dots, N$, is the spectral signature of the ith endmember. Consider also a small sliding square window that contains K adjacent pixels ($K = \kappa \times \kappa$), with the measurement spectra $\mathbf{y}_k$, $k = 1, 2, \dots, K$, that are assumed to share the same endmember matrix $\mathbf{\Phi}$, as graphically shown in Fig. 1 for $K = 9$. In matrix notation, let $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_K]$ be the $L \times K$ matrix containing the spectra of the K pixels in the window as its columns. Utilizing the LMM, the mixing process can be described by the following equation:

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

where $\mathbf{W} \in \mathbb{R}_+^{N \times K}$ is the abundance matrix whose columns are the N-dimensional abundance vectors of the corresponding K pixels, and $\mathbf{E} \in \mathbb{R}^{L \times K}$ is an independent and identically distributed (i.i.d.) zero-mean Gaussian noise matrix. Due to physical considerations, the abundance coefficients in $\mathbf{W}$ should satisfy two constraints, namely, the *abundance nonnegativity* and the *abundance sum-to-one* [20], i.e.,

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

Nevertheless, in the following, we relax the sum-to-one constraint based on the reasoning presented in [6]. That said, the general problem considered in this paper is as follows: "Given the spectral measurements $\mathbf{Y}$ and the endmember matrix $\mathbf{\Phi}$, estimate the abundance matrix $\mathbf{W}$ subject to the nonnegativity constraint." This is a typical inverse problem, which has been addressed via many methods in the signal processing literature. However, the efficacy of the proposed approach lies in the exploitation of the intrinsic structural characteristics of $\mathbf{W}$, i.e., sparsity and low rankness. To this end, we concurrently impose two naturally justified *structural constraints* on the abundance matrix $\mathbf{W}$, which promote *low rankness* and *sparsity*.

Low-Rankness Property: A logical consideration is that all pixels belonging to the same window are correlated, i.e., they are composed of the same materials, although maybe in different proportions. This property suggests that the abundance matrix $\mathbf{W}$ to be estimated has linearly dependent columns and, thus, is either of low rank or can be well approximated by a low-rank matrix. In the bibliography, low-rank matrix estimation techniques have recently emerged as powerful estimation tools, e.g., [10] and [21]–[23]. These estimators are mainly based on regularization by the nuclear norm of $\mathbf{W}$. A similar regularization is also adopted in this paper to impose the low-rank constraint.

Sparsity Property: Another typical assumption is that only a small portion of the N endmembers will be present in the spatial area marked by the $\kappa \times \kappa$ shifting window. In other words, it is safe to assume that the abundance matrix $\mathbf{W}$ has a *sparse representation* in terms of the endmember matrix dictionary $\mathbf{\Phi}$. This motivates the use of a sparsity-cognizant estimator for abundance matrix $\mathbf{W}$, which is envisaged to produce more robust unmixing results. It should be noted that sparsity has already been successfully exploited in many SU algorithms, e.g., [5]–[8], [11], and [24].

It is worth mentioning that the sparsity of $\mathbf{W}$ does by no means invalidate its low rankness. On the contrary, both structural hypotheses are assumed to hold *simultaneously* on $\mathbf{W}$, although low rankness implicitly imposes some kind of structure on sparsity. So far, reports in the SU literature explore either the sparsity, e.g., [7] and [24], or the low-rankness property of $\mathbf{W}$, e.g., [10]. To the best of our knowledge, this is the first time that SU is formulated as a simultaneously sparse and low-rank matrix estimation problem. That is, we seek a matrix $\mathbf{W} \geq \mathbf{0}$ that, apart from fitting the data well in the least squares sense, has minimum rank and only a few positive elements. To achieve this, we define the following optimization problem:

$$(\text{P1}): \hat{\mathbf{W}} = \underset{\mathbf{W} \in \mathbb{R}_+^{N \times K}}{\arg\min} \left\{ \frac{1}{2}\|\mathbf{Y} - \mathbf{\Phi}\mathbf{W}\|_F^2 + \gamma\|\mathbf{W}\|_1 + \tau\|\mathbf{W}\|_* \right\}$$
$$\tag{3}$$

where $\gamma, \tau \geq 0$ are parameters that control the tradeoff between the sparsity and rank regularization terms and the data fidelity term. Being parameterized, (P1) becomes flexible enough to impose either one of the two structures on $\mathbf{W}$. For example, by setting $\gamma = 0$, (P1) results in searching for a matrix that is of low-rank structure. Accordingly, setting $\tau = 0$ is tantamount to searching for a sparse matrix. The flexibility of the proposed model certainly provides an advantage over either low-rank or sparse estimation methods, as will be demonstrated later in Section IV.

It is also worth pointing out that (P1) involves the convex surrogates of the zero norm $\|\mathbf{W}\|_0$ and rank($\mathbf{W}$), i.e., the ℓ_1-norm and the nuclear norm, respectively. In an attempt to further promote the robustness and consistency of the proposed estimator, we propose to use *weighted* ℓ_1 and nuclear norms in (P1). Such an approach is expected to enhance the sparsity on the individual elements w_{ij} and the singular values $\sigma_i(\mathbf{W})$, e.g., [25]–[28]. These weighted norms are defined as

$$\|\mathbf{A} \odot \mathbf{W}\|_1 = \sum_{i=1}^{N} \sum_{j=1}^{K} a_{ij}|w_{ij}| \tag{4}$$

$$\|\mathbf{W}\|_{\mathbf{b},*} = \sum_{i=1}^{\text{rank}(\mathbf{W})} b_i \sigma_i(\mathbf{W}) \tag{5}$$

where a_{ij} and b_i are nonnegative weighting coefficients (i.e., $a_{ij} \geq 0$ and $b_i \geq 0$). Utilizing (4) and (5), the proposed optimization problem is rewritten as

$$(P2): \quad \hat{\mathbf{W}} = \arg\min_{\mathbf{W} \in \mathbb{R}_+^{N \times K}} \left\{ \frac{1}{2} \|\mathbf{Y} - \mathbf{\Phi}\mathbf{W}\|_F^2 + \gamma \|\mathbf{A} \odot \mathbf{W}\|_1 \right.$$
$$\left. + \tau \|\mathbf{W}\|_{\mathbf{b},*} \right\}. \quad (6)$$

To the best of our knowledge, such a formulation has not been used before as a regularizer for promoting simultaneous sparsity and low rankness. In the following, problem (P2) will be studied under the following assumption.

Assumption 1: For the weighting coefficients b_i of the nuclear norm, it holds that $b_i = b$, $i = 1, 2, \ldots, K$.

Under Assumption 1, the nuclear norm is convex [27], [29], whereas the weighted ℓ_1-norm is always convex for nonnegative $\mathbf{A}$. Thus, the overall cost function of (P2) is convex. Alternative options for the selection of parameters $\mathbf{A}$ and $\mathbf{b}$ are discussed in Section III-C. Although convex, (P2) under Assumption 1 is a nontrivial problem to solve, due to the nondifferentiable form of the ℓ_1-norm and nuclear norm regularizers [30]. In the following, we suitably explore two standard convex optimization tools to tackle this problem: an incremental proximal method and an ADMM-based technique.

III. Proposed Algorithms

In this section, we present two algorithms to address the nonsmooth, constrained, and convex optimization problem in (P2). The first algorithm comes from the family of incremental proximal algorithms, which was recently presented and analyzed in [19], and makes use of the proximal operators of all the terms appearing in (P2), whereas the second algorithm exploits the splitting strategy of the ADMM philosophy [13].

A. Incremental Proximal Sparse and Low-Rank Unmixing Algorithm

Let us first recall that the proximal operator of a function $f(\cdot)$ is defined as [31], [32]

$$\text{prox}_{\lambda f(\cdot)}(\mathbf{U}) = \arg\min_{\mathbf{W}} \left(f(\mathbf{W}) + \frac{1}{2\lambda} \|\mathbf{W} - \mathbf{U}\|_F^2 \right) \quad (7)$$

where $\mathbf{U} \in \mathbb{R}^{n \times k}$ and $\mathbf{W} \in \text{dom}(f)$, the domain of f. In [19], the following minimization problem is considered:

$$\min_{\mathbf{W} \in \mathcal{W}} \sum_{i=1}^{m} f_i(\mathbf{W}) \quad (8)$$

where $f_i(\mathbf{W})$, $i = 1, 2, \ldots, m$ are convex functions, and $\mathcal{W} \subseteq \mathbb{R}^{n \times k}$ is a closed convex set. One version of the algorithm proposed in [19] to solve this problem is the following. The proximal operators of all f_i's are first derived, and then, a sequential scheme is defined, in which the proximal operator of $f_i(\mathbf{W})$ is evaluated at the point provided by its predecessor (the proximal operator of $f_{i-1}(\mathbf{W})$), for $i = 2, 3, \ldots, m$. This procedure is repeated in a cyclic manner at each iteration of

the algorithm.[2] A convergence and rate of convergence analysis of this incremental proximal scheme is also given in [19]. It is worth noting that, as advocated in [19], this incremental approach favorably suits the minimization problems consisting of multiple constraints and may offer significant advantages over conventional nonincremental methods, such as reduced computational complexity, possibility of distributed processing, etc.

After this short introduction, we may observe that (P2) in (6) with $\mathbf{b} = b\mathbf{1}$ has exactly the same form with the minimization problem in (8), with respect to $\mathbf{W}$. Embedding the nonnegativity to the cost function in (6), we obtain the following regularized quadratic loss function:

$$\mathcal{L}_1(\mathbf{W}) = \frac{1}{2}\|\mathbf{Y} - \mathbf{\Phi}\mathbf{W}\|_F^2 + \gamma\|\mathbf{A}\odot\mathbf{W}\|_1 + \tau\|\mathbf{W}\|_{\mathbf{b},*} + \mathcal{I}_{\mathbb{R}_+}(\mathbf{W}) \quad (9)$$

where the nonnegativity constraint is now replaced by the (convex) indicator function $\mathcal{I}_{\mathbb{R}_+}(\mathbf{W})$, which is zero when all $w_{ij} \geq 0$, $i = 1, 2, \ldots, N$, $j = 1, 2, \ldots, K$, and $+\infty$ if at least one w_{ij} is negative. Typically, we wish to minimize $\mathcal{L}_1(\mathbf{W})$ with respect to $\mathbf{W}$. Notice that $\mathcal{L}_1(\mathbf{W})$ is the sum of four convex functions, and the incremental proximal algorithm in [19] can be directly applied to our problem. Next, the proximal operators of all four convex functions are obtained. Starting with the least squares fitting term, we readily get

$$\text{prox}_{\lambda\frac{1}{2}\|\mathbf{Y}-\mathbf{\Phi}\cdot\|_F^2}(\mathbf{W}) = (\mathbf{\Phi}^T\mathbf{\Phi} + \lambda^{-1}\mathbf{I}_N)^{-1}(\mathbf{\Phi}^T\mathbf{Y} + \lambda^{-1}\mathbf{W}). \quad (10)$$

Before we give the proximal operators for the next three terms, some necessary definitions are in order. First, we define the soft-thresholding operator on matrix $\mathbf{W} = [w_{ij}]$ as

$$\text{SHR}_{\mathbf{\Delta}}(\mathbf{W}) = \text{sign}(\mathbf{W}) \max(\mathbf{0}, |\mathbf{W}| - \mathbf{\Delta}) \quad (11)$$

where $\mathbf{\Delta} = [\delta]_{ij}$ is the matrix that contains thresholding parameters. Note that soft thresholding in (11) is performed in an element-wise manner, i.e., $\text{SHR}_{\delta_{ij}}(w_{ij}) = \text{sign}(w_{ij}) \max(0, |w_{ij}| - \delta_{ij})$. Notably, when we apply the soft-thresholding operator on a diagonal matrix, we shrink only the elements belonging to its diagonal. These elements are assumed to be shrinked by thresholding parameters contained in a vector. With this in mind, we define the singular value thresholding operation by

$$\text{SVT}_{\boldsymbol{\delta}}(\mathbf{W}) = \mathbf{U}\,\text{SHR}_{\boldsymbol{\delta}}(\mathbf{\Sigma})\,\mathbf{V}^T$$

where $\mathbf{W} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ is the singular value decomposition (SVD) of $\mathbf{W}$, and $\boldsymbol{\delta}$ is the vector whose entries are the thresholding parameters that reduce the corresponding diagonal elements of matrix $\mathbf{\Sigma}$. Finally, we define the projection operator on the set of nonnegative real numbers, i.e.,

$$\Pi_{\mathbb{R}_+}(v) = \arg\min_{x \in \mathbb{R}_+} |x - v| = \begin{cases} 0, & v < 0 \\ v, & v \geq 0 \end{cases} \quad (12)$$

which can also be applied to matrices in an element-wise manner.

---

[2]Instead of sequential, a randomized evaluation of the proximals of f_i's could be also employed [19].

Utilizing the given definitions, we can compute the proximal operators for all regularizing convex functions in (9). Specifically, $\text{prox}_{\gamma\|\mathbf{A}\odot\cdot\|_1}(\mathbf{W})$ is computed by soft-thresholding matrix $\mathbf{W}$ with $\gamma\mathbf{A}$ as follows:

$$\text{prox}_{\gamma\|\mathbf{A}\odot\cdot\|_1}(\mathbf{W}) = \text{SHR}_{\gamma\mathbf{A}}(\mathbf{W}). \tag{13}$$

Similarly, the proximal operator of the nuclear norm can be expressed via a soft-thresholding operation on the singular values of $\mathbf{W}$, i.e.,

$$\text{prox}_{\tau\|\cdot\|_{\mathbf{b},*}}(\mathbf{W}) = \text{SVT}_{\tau\mathbf{b}}(\mathbf{W}). \tag{14}$$

Moreover, the computation of $\text{prox}_{\mathcal{I}_{\mathbb{R}_+}(\cdot)}(\mathbf{W})$ reduces to a projection operation, i.e.,

$$\text{prox}_{\mathcal{I}_{\mathbb{R}_+}(\cdot)}(\mathbf{W}) = \Pi_{\mathbb{R}_+}(\mathbf{W}). \tag{15}$$

The proposed incremental proximal sparse and low-rank unmixing algorithm (IPSpLRU) iterates among the proximal operators (10) and (13)–(15) in a cyclic order until convergence [19]. IPSpLRU is summarized in Algorithm 1. Note that to retain the convexity of the composite functions, the weighting parameters $\mathbf{A}$ and $\mathbf{b}$ are initialized and kept fixed during the execution of the algorithm. The issue of dynamic selection of these parameters is discussed in Section III-C.

The incremental proximal approach employed above for deriving IPSpLRU is closely related to the incremental subgadient method [19], and parameters λ, γ, and τ can be seen as the step sizes of the corresponding subgradient steps. By invoking [19, Proposition 3.2], it arises that for fixed values of these parameters, the incremental proximal algorithm converges to a neighborhood of the optimum, which shrinks to zero as their values are closer to zero. On the other hand, exact convergence to the optimal solution of the cost function is achieved when the values of these step sizes diminish over iterations, while they additionally satisfy certain conditions described in [19]. Herein, parameters λ, γ, and τ are selected to be fixed to positive constants during the execution of the algorithm. In doing so, we sacrifice the accuracy of the estimations in favor of a faster convergence rate.

Algorithm 1 The proposed IPSpLRU algorithm

Inputs $\mathbf{Y}$, $\boldsymbol{\Phi}$
Select parameters $\mathbf{A}, \mathbf{b}, \lambda, \tau, \gamma$
Set $\mathbf{R} = \left(\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \lambda^{-1}\mathbf{I}_N\right)^{-1}$, $\mathbf{P} = \boldsymbol{\Phi}^T\mathbf{Y}$, $\mathbf{Q} = \mathbf{RP}$
Initialize $\mathbf{W}^0$ and set $t = 1$
repeat
    $\mathbf{W}^t = \mathbf{Q} + \lambda^{-1}\mathbf{RW}^{t-1}$
    $\mathbf{W}^t = \text{prox}_{\gamma\|\mathbf{A}\odot\cdot\|_1}(\mathbf{W}^t)$
    $\mathbf{W}^t = \text{prox}_{\tau\|\cdot\|_{\mathbf{b},*}}(\mathbf{W}^t)$
    $\mathbf{W}^t = \text{prox}_{\mathcal{I}_{\mathbb{R}_+}(\cdot)}(\mathbf{W}^t)$
until convergence
Output: Abundance matrix $\hat{\mathbf{W}} = \mathbf{W}^t$

Concerning the computational complexity of IPSpLRU, the most complex step is the SVD of abundance matrix $\mathbf{W}^t$, which

takes place at each iteration and is on the order of $\mathcal{O}(KN^2 + K^3)$, [33]. Note that matrices $\mathbf{R} = \left(\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \lambda^{-1}\mathbf{I}_N\right)^{-1}$, $\mathbf{P} = \boldsymbol{\Phi}^T\mathbf{Y}$, and $\mathbf{Q} = \mathbf{RP}$ are computed only once at the initialization stage, and thus, the first step of the algorithm just requires a fast matrix-by-matrix multiplication. The algorithm rapidly converges and terminates when either the following stopping criterion is satisfied:

$$\frac{\|\mathbf{W}^t - \mathbf{W}^{t-1}\|_F^2}{\|\mathbf{W}^{t-1}\|_F^2} < \delta \tag{16}$$

where δ is a predefined threshold value, or a preset maximum number of iterations is reached. In the following section, we present an alternative approach to solve the same problem by employing a primal–dual ADMM-type technique.

B. ADMM for Sparse and Low-Rank Unmixing

In this section, we develop an instance of the ADMM that solves the abundance matrix estimation problem (P2). To proceed, we utilize the auxiliary matrix variables $\boldsymbol{\Omega}_1$, $\boldsymbol{\Omega}_2$, $\boldsymbol{\Omega}_3$, and $\boldsymbol{\Omega}_4$ of proper dimensions (similar to [11] and [24]) and reformulate the original problem (P2) into its equivalent ADMM form [13], i.e.,

$$(\text{P3}): \min_{\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \boldsymbol{\Omega}_3, \boldsymbol{\Omega}_4} \left\{ \frac{1}{2}\|\boldsymbol{\Omega}_1 - \mathbf{Y}\|_F^2 + \gamma\|\mathbf{A}\odot\boldsymbol{\Omega}_2\|_1 \right.$$
$$\left. + \tau\|\boldsymbol{\Omega}_3\|_{\mathbf{b},*} + \mathcal{I}_{\mathbb{R}_+}(\boldsymbol{\Omega}_4) \right\} \tag{17}$$
$$\text{s.t. } \boldsymbol{\Omega}_1 - \boldsymbol{\Phi}\mathbf{W} = 0, \ \boldsymbol{\Omega}_2 - \mathbf{W} = 0, \ \boldsymbol{\Omega}_3 - \mathbf{W} = 0$$
$$\boldsymbol{\Omega}_4 - \mathbf{W} = 0.$$

Based on (P3), the following augmented Lagrangian function is optimized with respect to $\mathbf{W}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \boldsymbol{\Omega}_3,$ and $\boldsymbol{\Omega}_4$:

$$\mathcal{L}_2(\mathbf{W}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \boldsymbol{\Omega}_3, \boldsymbol{\Omega}_4)$$
$$= \frac{1}{2}\|\boldsymbol{\Omega}_1 - \mathbf{Y}\|_F^2 + \gamma\|\mathbf{A}\odot\boldsymbol{\Omega}_2\|_1$$
$$+ \tau\|\boldsymbol{\Omega}_3\|_{\mathbf{b},*} + \mathcal{I}_{\mathbb{R}_+}(\boldsymbol{\Omega}_4) + \text{tr}\left[\mathbf{D}_1^T(\boldsymbol{\Omega}_1 - \boldsymbol{\Phi}\mathbf{W})\right]$$
$$+ \text{tr}\left[\mathbf{D}_2^T(\boldsymbol{\Omega}_2 - \mathbf{W})\right] + \text{tr}\left[\mathbf{D}_3^T(\boldsymbol{\Omega}_3 - \mathbf{W})\right]$$
$$+ \text{tr}\left[\mathbf{D}_4^T(\boldsymbol{\Omega}_4 - \mathbf{W})\right]$$
$$+ \frac{\mu}{2}\left(\|\boldsymbol{\Phi}\mathbf{W} - \boldsymbol{\Omega}_1\|_F^2 + \|\mathbf{W} - \boldsymbol{\Omega}_2\|_F^2\right.$$
$$\left. + \|\mathbf{W} - \boldsymbol{\Omega}_3\|_F^2 + \|\mathbf{W} - \boldsymbol{\Omega}_4\|_F^2\right) \tag{18}$$

where the $L \times K$ matrix $\mathbf{D}_1$ and the $N \times K$ matrices $\mathbf{D}_2$, $\mathbf{D}_3, \mathbf{D}_4$ are the Lagrange multipliers, and $\mu > 0$ is a positive penalty parameter. Note that, again, the nonnegative weights $\mathbf{A}$ and $\mathbf{b}$ are considered to be constant, and Assumption 1 also holds here. Let

$$\boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\Omega}_1 \\ \boldsymbol{\Omega}_2 \\ \boldsymbol{\Omega}_3 \\ \boldsymbol{\Omega}_4 \end{bmatrix}, \ \mathbf{G} = \begin{bmatrix} \boldsymbol{\Phi} \\ \mathbf{I}_N \\ \mathbf{I}_N \\ \mathbf{I}_N \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} -\mathbf{I}_L & 0 & 0 & 0 \\ 0 & -\mathbf{I}_N & 0 & 0 \\ 0 & 0 & -\mathbf{I}_N & 0 \\ 0 & 0 & 0 & -\mathbf{I}_N \end{bmatrix}. \tag{19}$$

Then, (18) can be written in an equivalent form as

$$\mathcal{L}_3(\mathbf{W}, \boldsymbol{\Omega}, \boldsymbol{\Lambda}) = \frac{1}{2}\|\boldsymbol{\Omega}_1 - \mathbf{Y}\|_F^2 + \gamma\|\mathbf{A} \odot \boldsymbol{\Omega}_2\|_1$$
$$+ \tau\|\boldsymbol{\Omega}_3\|_{\mathbf{b},*} + \mathcal{I}_{\mathbb{R}_+}(\boldsymbol{\Omega}_4) + \frac{\mu}{2}\|\mathbf{G}\mathbf{W} + \mathbf{B}\boldsymbol{\Omega} - \boldsymbol{\Lambda}\|_F^2 \quad (20)$$

where $\boldsymbol{\Lambda} = [\boldsymbol{\Lambda}_1^T \ \boldsymbol{\Lambda}_2^T \ \boldsymbol{\Lambda}_3^T \ \boldsymbol{\Lambda}_4^T]^T$, $\boldsymbol{\Lambda}_i = (1/\mu)\mathbf{D}_i$, $i = 1, \ldots, 4$, contains the scaled Lagrange multipliers. Having expressed the augmented Lagrangian function as in (20), the ADMM proceeds by minimizing $\mathcal{L}_3(\mathbf{W}, \boldsymbol{\Omega}, \boldsymbol{\Lambda})$ sequentially, each time with respect to a single matrix variable, keeping the remaining variables at their latest values. The dual variables (Lagrange multipliers) are also updated via a gradient ascend step at the end of each alternating minimization cycle.

To further elaborate on the steps of the ADMM, the optimization with respect to $\mathbf{W}$ gives

$$\mathbf{W}^t = \arg\min_{\mathbf{W}} \mathcal{L}_3(\mathbf{W}, \boldsymbol{\Omega}^{t-1}, \boldsymbol{\Lambda}^{t-1})$$
$$= (\boldsymbol{\Phi}^T\boldsymbol{\Phi} + 3\mathbf{I}_N)^{-1} \big[\boldsymbol{\Phi}^T (\boldsymbol{\Omega}_1^{t-1} + \boldsymbol{\Lambda}_1^{t-1}) + \boldsymbol{\Omega}_2^{t-1} + \boldsymbol{\Lambda}_2^{t-1}$$
$$+ \boldsymbol{\Omega}_3^{t-1} + \boldsymbol{\Lambda}_3^{t-1} + \boldsymbol{\Omega}_4^{t-1} + \boldsymbol{\Lambda}_4^{t-1} \big].$$
$$(21)$$

Next, the optimization with respect to $\boldsymbol{\Omega}_1$ is performed as

$$\boldsymbol{\Omega}_1^t = \arg\min_{\boldsymbol{\Omega}_1} \mathcal{L}_3(\mathbf{W}^t, \boldsymbol{\Omega}, \boldsymbol{\Lambda}^{t-1})$$
$$= \frac{1}{1+\mu} \big(\mathbf{Y} + \mu \big(\boldsymbol{\Phi}\mathbf{W}^t - \boldsymbol{\Lambda}_1^{t-1}\big) \big). \quad (22)$$

The remaining auxiliary variables $\boldsymbol{\Omega}_2$, $\boldsymbol{\Omega}_3$, and $\boldsymbol{\Omega}_4$ are involved in nondifferentiable norms, namely, the weighted ℓ_1-norm, the weighted nuclear norm, and the indicator function, respectively. In this regard, the minimization task with respect to these variables resolves to computing some of the proximity operators that we introduced in the previous section. Minimizing (20) with respect to $\boldsymbol{\Omega}_2$ yields

$$\boldsymbol{\Omega}_2^t = \arg\min_{\boldsymbol{\Omega}_2} \mathcal{L}_3(\mathbf{W}^t, \boldsymbol{\Omega}, \boldsymbol{\Lambda}^{t-1})$$
$$= \text{SHR}_{\gamma\mathbf{A}} \big(\mathbf{W}^t - \boldsymbol{\Lambda}_2^{t-1} \big). \quad (23)$$

In the same vein, $\boldsymbol{\Omega}_3$ is computed by a shrinkage operation, i.e.,

$$\boldsymbol{\Omega}_3^t = \arg\min_{\boldsymbol{\Omega}_3} \mathcal{L}_3(\mathbf{W}^t, \boldsymbol{\Omega}, \boldsymbol{\Lambda}^{t-1})$$
$$= \text{SVT}_{\tau\mathbf{b}} \big(\mathbf{W}^t - \boldsymbol{\Lambda}_3^{t-1} \big). \quad (24)$$

Next, for the auxiliary variable $\boldsymbol{\Omega}_4$, a projection onto the nonnegative orthant is required, i.e.,

$$\boldsymbol{\Omega}_4^t = \arg\min_{\boldsymbol{\Omega}_4} \mathcal{L}_3(\mathbf{W}^t, \boldsymbol{\Omega}, \boldsymbol{\Lambda}^{t-1})$$
$$= \Pi_{\mathbb{R}_+} \big(\mathbf{W}^t - \boldsymbol{\Lambda}_4^{t-1} \big). \quad (25)$$

At the final step of the proposed method, the scaled Lagrange multipliers in $\boldsymbol{\Lambda}$ are sequentially updated by performing gradient ascent on the dual problem [13], as follows:

$$\boldsymbol{\Lambda}_1^t = \boldsymbol{\Lambda}_1^{t-1} - \boldsymbol{\Phi}\mathbf{W}^t + \boldsymbol{\Omega}_1^t$$
$$\boldsymbol{\Lambda}_i^t = \boldsymbol{\Lambda}_i^{t-1} - \mathbf{W}^t + \boldsymbol{\Omega}_i^t, \ i = 2, 3, 4. \quad (26)$$

The proposed algorithm, which is termed as the alternating direction sparse and low-rank unmixing algorithm (ADSpLRU), is summarized in Algorithm 2. An iteration of ADSpLRU consists of the update steps given in (21)–(26). Its computational complexity is $\mathcal{O}(KLN + KN^2)$ per iteration, which is slightly higher than that of IPSpLRU, since it usually holds $L > N$. However, as verified by the simulations in the following section, ADSpLRU requires fewer iterations than IPSpLRU to converge,[3] while its convergence is also guaranteed as explained in [34]. Moreover, it achieves a slightly lower steady-state error. Note that all functions that form the objective function $\mathcal{L}_1(\mathbf{W})$ in (9) are closed, proper, and convex. Since matrix $\mathbf{G}$ has full column rank, the convergence conditions defined in [34] are met, and if an optimal solution exists, ADSpLRU converges, for any $\mu > 0$. This, in turn, implies that for the primal and dual residuals $\mathbf{r}^t$ and $\mathbf{d}^t$ given by

$$\mathbf{r}^t = \mathbf{G}\mathbf{W}^t + \mathbf{B}\boldsymbol{\Omega}^t$$
$$\mathbf{d}^t = \mu\mathbf{G}^\top\mathbf{B}(\boldsymbol{\Omega}^t - \boldsymbol{\Omega}^{t-1})$$

it holds that $\mathbf{r}^t \to 0$ and $\mathbf{d}^t \to 0$, respectively, as $t \to \infty$. In this paper, ADSpLRU stops when either the following termination criterion:

$$\|\mathbf{r}^t\|_2 \leq \zeta \text{ and } \|\mathbf{d}^t\|_2 \leq \zeta \quad (27)$$

holds for the primal and dual residuals, where $\zeta = \sqrt{(3N + L)K}\zeta^{\text{rel}}$ [13] (the relative tolerance $\zeta^{\text{rel}} > 0$ takes its value depending on the application and, in our experimental study, has been empirically determined to 10^{-4}), or the maximum number of iterations is reached.

Algorithm 2 The proposed ADSpLRU algorithm

Inputs $\mathbf{Y}, \boldsymbol{\Phi}$
Select parameters $\mathbf{A}, \mathbf{b}, \mu, \tau, \gamma$
Set $\mathbf{R} = (\boldsymbol{\Phi}^T\boldsymbol{\Phi} + 3\mathbf{I}_N)^{-1}$
Initialize $\mathbf{W}^0, \boldsymbol{\Omega}^0, \boldsymbol{\Lambda}^0$ and set $t = 1$
repeat
    $\mathbf{W}^t = \mathbf{R}[\boldsymbol{\Phi}^T(\boldsymbol{\Omega}_1^{t-1} + \boldsymbol{\Lambda}_1^{t-1}) + \boldsymbol{\Omega}_2^{t-1} + \boldsymbol{\Lambda}_2^{t-1} + \boldsymbol{\Omega}_3^{t-1} + \boldsymbol{\Lambda}_3^{t-1} + \boldsymbol{\Omega}_4^{t-1} + \boldsymbol{\Lambda}_4^{t-1}]$
    $\boldsymbol{\Omega}_1^t = 1/(1+\mu)(\mathbf{Y} + \mu(\boldsymbol{\Phi}\mathbf{W}^t - \boldsymbol{\Lambda}_1^{t-1}))$
    $\boldsymbol{\Omega}_2^t = \text{SHR}_{\gamma\mathbf{A}}(\mathbf{W}^t - \boldsymbol{\Lambda}_2^{t-1})$
    $\boldsymbol{\Omega}_3^t = \text{SVT}_{\tau\mathbf{b}}(\mathbf{W}^t - \boldsymbol{\Lambda}_3^{t-1})$
    $\boldsymbol{\Omega}_4^t = \Pi_{\mathbb{R}_+}(\mathbf{W}^t - \boldsymbol{\Lambda}_4^{t-1})$
    $\boldsymbol{\Lambda}_1^t = \boldsymbol{\Lambda}_1^{t-1} - \boldsymbol{\Phi}\mathbf{W}^t + \boldsymbol{\Omega}_1^t$
    $\boldsymbol{\Lambda}_i^t = \boldsymbol{\Lambda}_i^{t-1} - \mathbf{W}^t + \boldsymbol{\Omega}_i^t, i = 2, 3, 4$
until convergence
Output: Abundance matrix $\hat{\mathbf{W}} = \mathbf{W}^t$

C. Selection of Weighting Coefficients and Regularization Parameters

As previously mentioned, in both IPSpLRU and ADSpLRU, the weighting coefficients $\mathbf{A}$ and $\mathbf{b}$ are predetermined, satisfy certain constraints, and remain constant during the execution

---

[3]The reason for this may be that ADSpLRU manipulates the whole cost function at each step, whereas IPSpLRU splits the cost function in a number of convex terms and treats each term individually at every step of the algorithm.

TABLE I
COMPUTATIONAL COMPLEXITY PER PIXEL AND ITERATION

Algorithm	IPSpLRU	ADSpLRU	CSUnSAl [5]	MMV-ADMM [10]	BiICE [8]
Computational complexity	$\mathcal{O}(KN^2 + K^3)$	$\mathcal{O}(KN^2 + KLN)$	$\mathcal{O}(N^2)$	$\mathcal{O}(KN^2 + KLN)$	$\mathcal{O}(N^2)$

of the algorithms. As is widely known [25], [26], [29], proper selection of these parameters is quite crucial as for the accuracy of the estimations. In view of this, two potential choices are:

a) to select the weighting coefficients based on the least squares estimate $\mathbf{W}^{\mathrm{LS}}$ of $\mathbf{W}$, i.e.,

$$a_{ij} = \left(\frac{1}{w_{ij}^{\mathrm{LS}} + \epsilon}\right) \qquad b_i = \left(\frac{1}{\sigma_i(\mathbf{W}^{\mathrm{LS}}) + \epsilon}\right) \quad (28)$$

where $\epsilon = 10^{-16}$ is a small constant added to avoid singularities; or

b) to update them at each iteration t of the algorithms based on the current estimate $\mathbf{W}^t$ of $\mathbf{W}$, i.e.,

$$a_{ij}^t = \left(\frac{1}{w_{ij}^t + \epsilon}\right) \qquad b_i^t = \left(\frac{1}{\sigma_i(\mathbf{W}^t) + \epsilon}\right). \quad (29)$$

It should be noted that both these two options render the minimization problem (P2) nonconvex, since the weighted nuclear norm is known to be convex only if the weights b_i, $i = 1, 2, \ldots, K$ are nonnegative and nonascending [29], [35]. Additionally, the reweighting norm minimization problem is known to be inherently nonconvex [26], whereas its theoretical convergence analysis for these cases is difficult to be established.[4] Nevertheless, numerous research works advocate the positive impact of these nonconvex weighted norms on the performance of general constrained estimation tasks [26], [27], [29], [35] as well as in hyperspectral unmixing [36], [37]. Along this line of thought, the algorithms presented in the previous section are modified by adopting the reweighting scheme given by (29). As verified in our empirical study presented in the following section, such an option enhances to a large degree the effectiveness of the proposed algorithms, while no numerical issues have been encountered in our experiments.

As far as the remaining parameters λ and μ are concerned, which control the convergence behavior of IPSpLRU and ADSpLRU, respectively, they take positive values, with μ close to zero and λ on the order of 1. In all our experiments, we fixed $\mu = 0.01$ and $\lambda = 0.5$. On the other hand, the low-rank-and sparsity-promoting parameters τ and γ are chosen via fine tuning, as is commonly done in relevant deterministic schemes. This is so because the optimal set of these parameters depends on the unknown in advance particular structure of the sought abundance matrix, which is an issue that is further explained in the following section.

IV. EXPERIMENTAL RESULTS

This section unravels the performance characteristics of the proposed IPSpLRU and ADSpLRU algorithms via experiments conducted both on simulated and real data. We compare our

techniques with three well-known state-of-the-art unmixing algorithms, namely, the nonnegative constraint sparse unmixing by variable splitting and augmented Lagrangian algorithm (CSUnSAL) [5], the recently reported nonnegative constraint joint-sparse method (MMV-ADMM) [10], and, finally, the fast Bayesian inference iterative conditional expectations (BiICE) unmixing algorithm [8]. The computational complexity (in terms of the number of multiplications) of all tested algorithms is given in Table I. As shown in the table, the spatial-correlation-aware algorithms, namely, IPSpLRU, ADSpLRU, and MMV-ADMM, present higher complexity since the information from K pixels is used for the unmixng of a single pixel. Moreover, it is noticed that among the two proposed algorithms, IPSpLRU has lower computational complexity than ADSpLRU per iteration, resulting from its more simplistic incremental approach.

In what follows, we first refer to the parameters' setting established for all the involved algorithms and the performance evaluation metrics that are utilized in the experimental procedure. To corroborate the effectiveness and robustness of the proposed algorithms, we execute six different types of synthetic-data experiments whose detailed description is given below. Finally, we empirically compare the abundance maps as revealed by all examined algorithms, when applied on a real HSI.[5]

A. Setting of Parameters and Performance Evaluation Criteria

For simplicity reasons, we use γ for the sparsity-imposing parameter in all tested algorithms (except BiICE that has no regularization parameters [7]), μ for the Lagrange multiplier regularization parameter of the ADMM-type techniques, and λ for the relevant to μ regularization parameter of IPSpLRU. Additionally, the low-rank-promoting parameter of the proposed algorithms is denoted by τ. Parameters τ and γ are fine tuned with ten different values, as shown in Table II. On the other hand, the Lagrange multiplier regularization parameter μ and the regularization parameter λ of IPSpLRU, which influence to a lesser extent the efficiency of the corresponding algorithms, are set to a fixed value. To assess the performance of the proposed algorithms and the competing algorithms, we consider two metrics for the experiments conducted on synthetic data. First, the root mean square error (RMSE), i.e.,

$$\mathrm{RMSE} = \sqrt{\frac{1}{Nn} \sum_{i=1}^{n} \|\hat{\mathbf{w}}_i - \mathbf{w}_i\|_2} \quad (30)$$

where $\hat{\mathbf{w}}_i$ and $\mathbf{w}_i$ represent the estimated and actual abundance vectors of the ith pixel, respectively; n is the total number of the pixels in the image under study; and N, as mentioned in previous sections, stands for the number of endmembers. The second metric is the signal-to-reconstruction error (SRE) [6], which

---

[4]It should be noted that the convergence results for the incremental proximal algorithms provided in [19] do not hold for nonconvex f_i's.

[5]The MATLAB code of the proposed algorithms is provided at http://members.noa.gr/parisg/demo_splr_unmixing.zip.

TABLE II
PARAMETER SETTING

Algorithm	τ (rank regularization parameter)	γ (sparsity regularization parameter)	μ	λ
IPSpLRU	$\{0, 10^{-10}, 10^{-9}, \ldots, 10^{-1}\}$	$\{0, 10^{-10}, 10^{-9}, \ldots, 10^{-1}\}$	Not applicable	0.5
ADSpLRU	$\{0, 10^{-10}, 10^{-9}, \ldots, 10^{-1}\}$	$\{0, 10^{-10}, 10^{-9}, \ldots, 10^{-1}\}$	10^{-2}	Not applicable
CSUnSAL	Not applicable	$\{0, 10^{-10}, 10^{-9}, \ldots, 10^{-1}\}$	10^{-2}	Not applicable
MMV-ADMM	Not applicable	$\{0, 10^{-10}, 10^{-9}, \ldots, 10^{-1}\}$	10^{-2}	Not applicable

reflects the ratio between the power of the signal and the power of the estimation error, and is given by the following formula:

$$\text{SRE} = 10 \log_{10} \left(\frac{\frac{1}{n} \sum_{i=1}^{n} \|\hat{\mathbf{w}}_i\|_2^2}{\frac{1}{n} \sum_{i=1}^{n} \|\hat{\mathbf{w}}_i - \mathbf{w}_i\|_2^2} \right). \qquad (31)$$

Of great importance is to notice that for the sliding-window-based algorithms, the abundance vectors $\mathbf{w}_i$'s and their estimates $\hat{\mathbf{w}}_i$'s coincide with the central column vectors of the corresponding abundance matrices $\mathbf{W}_i$'s and their estimates $\hat{\mathbf{W}}_i$'s, as becomes clear in Fig. 1.

B. Experiments on Simulated Datacubes

In the sequel, N endmembers are randomly selected from the U.S. Geological Survey (USGS) library $\mathbf{Q} \in \mathbb{R}_+^{224 \times 498}$ [38], so as to form our endmembers' dictionary $\boldsymbol{\Phi}$. Their reflectance values correspond to $L = 224$ spectral bands, uniformly distributed in the interval 0.4–2.5 μm. The LMM of (1) is then utilized for generating spectral signatures subject to given, different in each experiment, abundance matrices $\mathbf{W}$'s.

1) Reweighting Coefficient Efficiency and Convergence Behavior of IPSpLRU and ADSpLRU: Herein, we aspire to demonstrate the merits emerging from the utilization of the reweighting of $\mathbf{A}$ and $\mathbf{b}$ from (29), on the estimation performance of the proposed algorithms. In light of this, we consider a rank-3 and sparsity level 10% (i.e., 10% of its entries are nonzero) abundance matrix corresponding to $N = 50$ endmembers, $K = 9$ pixels. Then, $K = 9$ spectral signatures are generated according to the LMM and contaminated by Gaussian noise such that SNR = 30 dB.

For $p = 100$ realizations, Fig. 2 depicts the normalized mean square estimation error (NMSE; defined as $\text{NMSE}(t) = (1/p) \sum_{i=1}^{p} (\|\hat{\mathbf{W}}_i^t - \mathbf{W}_i\|_F^2 / \|\mathbf{W}_i\|_F^2)$, where $\hat{\mathbf{W}}_i^t$ and $\mathbf{W}_i$ are the estimated matrices at the tth iteration and the true matrices of the ith realization, respectively) as it evolves over 2000 iterations. Three different cases are investigated, corresponding to 1) updating weighting coefficients from (29), 2) keeping fixed the weighting coefficients based on (28), and 3) no weighting coefficients, i.e., the weighted norms degenerate to their non-weighted versions by setting $\mathbf{b} = \mathbf{1}$ and $\mathbf{A} = [\mathbf{1}, \mathbf{1}, \ldots \mathbf{1}]$. As it is clearly evident in Fig. 2, both IPSpLRU and ADSpLRU achieve remarkably higher estimation accuracy in terms of NMSE, when using reweighting as compared with the case where fixed or no weights are employed. It is thus empirically verified that the enhanced efficiency of the reweighted ℓ_1 and nuclear norms, emphatically advocated in [25]–[27], is retained when using the sum of these two norms. The price to be paid is that such an option might increase numerical risks, since the problem is rendered nonconvex and (yet) no theoretical
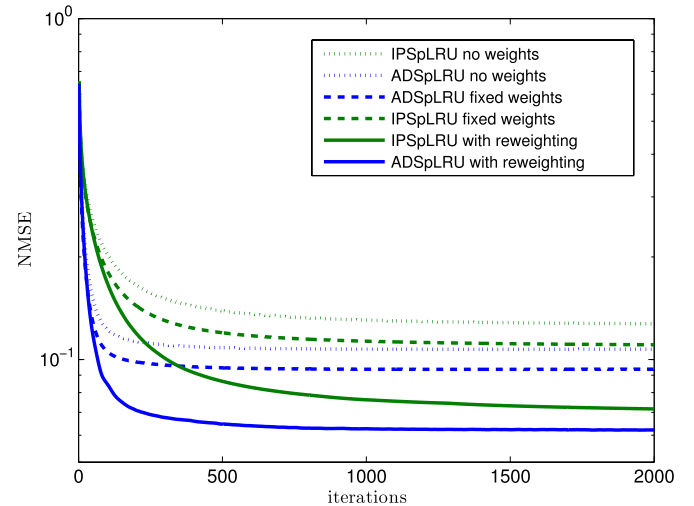

Fig. 2. Convergence curves of IPSpLRU and ADSpLRU for 1) updating weighting coefficients, 2) fixed weighting coefficients, and 3) no weighting coefficients.

convergence analysis has been established. Nevertheless, it is worth mentioning that, despite the fact that convergence is not theoretically guaranteed, in all our experiments, both IPSpLRU and ADSpLRU exhibited a very robust convergence behavior.

It is also noticed that ADSpLRU needs less iterations to converge as compared with IPSpLRU, and it converges to a slightly lower NMSE. This results from the inherent nature of the two proposed algorithms, as previously explained. Interestingly, the faster convergence rate of ADSpLRU with reweighting comes at the price of its higher per-iteration computational complexity as compared with that of IPSpLRU.

2) Window Size Selection and Runtime Comparison Between IPSpLRU and ADSpLRU: In this experiment, we explore the performance behavior of IPSpLRU and ADSpLRU as a function of the number of pixels K in the sliding window. Recall that HSIs are, in general, of low spatial resolution; thus, spatial correlation is mostly found in small regions of the images, corresponding to a limited number of pixels. That said, the low-rankness level (LRL) of the corresponding abundance matrix (which we define as the ratio $\min(K, N)$/true rank of $\mathbf{W}$) usually decreases as the size K of the window increases.[6] Keeping that in mind, we simulate abundance matrices setting the number of the endmembers N to 50, while varying the size of the window. To account for the decreasing LRL as size K grows, the rank of the generated abundance matrices was appropriately set for each K. Moreover, the sparsity level of each matrix is set to 0.5. Spectral signatures are then produced

---

[6] In line with this, in this experiment, the true rank of $\mathbf{W}$ grows more rapidly than $\min(K, N)$; hence, the ratio $\min(K, N)$/true rank of $\mathbf{W}$ takes smaller values, as K increases, also shown in Table III.

TABLE III
PERFORMANCE AND RUNTIME COMPARISON BETWEEN IPSpLRU AND ADSpLRU FOR
DIFFERENT NUMBERS OF PIXELS K OF THE SLIDING WINDOW AND LRLs

Algorithm	$K = 1$, LRL $= 1$		$K = 9$, LRL $= 4.5$		$K = 25$, LRL $= 2.5$		$K = 49$, LRL $= 2.2$		$K = 81$, LRL $= 2.04$	
	RMSE	runtime (sec)	RMSE	runtime (sec)	RMSE	runtime (sec)	RMSE	runtime (sec)	RMSE	runtime (sec)
IPSpLRU	0.0601	0.0602	0.0450	0.0673	0.0467	0.1034	0.0503	0.1814	0.0523	0.2037
ADSpLRU	0.0519	0.1187	0.0368	0.1363	0.0384	0.1811	0.0423	0.2734	0.0432	0.3335

via the LMM with noise of SNR $= 35$ dB corrupting the data, for 100 different realizations of the experiment. As shown in Table III, the performance of both IPSpLRU and ADSpLRU is optimized in terms of the RMSE for $K = 9$. This is so since, for $K = 9$, the generated abundance matrices are characterized by the maximum LRL, which is efficiently exploited by both the proposed algorithms.

As far as the accuracy of the estimated abundance matrices is concerned, ADSpLRU outweighs IPSpLRU for all different K's. However, ADSpLRU demands more runtime compared with IPSpLRU. This fact is in good agreement with the computational complexity of the two proposed algorithms given in Table I. Therein, it is shown that due to its novel incremental approach, IPSpLRU's computational complexity scales with $KN^2 + K^3$ while $KN^2 + KLN$ arises correspondingly for ADSpLRU. As a result, when the size of the selected window K is smaller than the number of the spectral bands L, as well as the number of the endmembers N contained in the endmember's library (a hypothesis that shall hold to aptly exploit spatial correlation, as previously advocated), IPSpLRU offers a more computationally efficient solution compared with ADSpLRU.

3) Toy Example: In this experiment, our goal is to highlight the significance of the approach followed in this work, i.e., the simultaneous incorporation of sparsity and low rankness on the abundance estimation problem. To this end, we initially derive the single prior counterparts of our algorithms. We first focus on the low-rankness assumption; thus, the sparsity-imposing norm is ignored ($\gamma = 0$). IPSpLRU and ADSpLRU are then reduced to their modified versions, namely, IPLRU and ADLRU, respectively. As implied by their names, the aforementioned methods exclusively allow for the low-rank assumption. Similarly, IPSpU and ADSpU are formed by solely accounting for sparsity. That said, IPSpU and ADSpU emerge after dropping the low-rank prior constraint ($\tau = 0$). Next, we generate an $N \times K$ (where $N = 50$ and $K = 9$) simultaneously sparse and low-rank abundance matrix $\mathbf{W}$ of rank 2 with a sparsity level of 20%, which is graphically illustrated in Fig. 3(a). Using this $\mathbf{W}$, we generate the $L \times K$ observation matrix $\mathbf{Y}$ via the LMM in (1), where the noise matrix $\mathbf{E}$ is Gaussian i.i.d., and SNR $= 35$ dB.

Fig. 3 shows the merits of the proposed IPSpLRU and ADSpLRU algorithms. Specifically, it appears that the concurrent exploitation of sparsity and low rankness leads to significantly more accurate abundance matrix estimates, as compared with their single-constraint counterparts, namely, IPLRU, IPSpU and ADLRU, ADSpU, respectively. This is clearly seen in terms of the RMSE, as well as from a careful visual inspection of both the recovered abundance matrices and their residuals with the true abundace matrix (i.e., $|\hat{\mathbf{W}} - \mathbf{W}|$), depicted in pair in Fig. 3(b)–(m).

4) Key Role of the Parameters γ, τ: As previously explained, parameters γ and τ control the imposition of sparsity and low rankness, respectively, on abundance matrix $\mathbf{W}$. Herein, we unveil the dependence of the optimal (with respect to RMSE minimization) set of these parameters on the inherent structure of the sought abundance matrix. In this vein, five different types of abundance matrices are generated, each reflecting a specific combination of rank and sparsity level. Next, $K = 9$ linearly mixed pixels are produced, corrupted with Gaussian i.i.d. noise, and SNR $= 35$ dB. For each of the five experiments, 100 independent realizations are run, and the average RMSE is demonstrated as a function of τ and γ. As shown in Fig. 4, in the first case [see Fig. 4(a) and (e)], which corresponds to solely low-rank abundance matrices (without any presence of sparsity), the sparsity-promoting parameter γ does not affect the estimation accuracy. In a similar manner, in the fourth experiment [see Fig. 4(d) and (h)], where the abundance matrix is considered of full rank and sparse, the low-rank-promoting parameter has no impact on the estimation performance. Notably, in the other two cases (columns 2 and 3) where both sparse and low-rank abundance matrices are considered, RMSE is minimized for nonzero values of both τ and γ. Such a result is consistent with the fundamental premise of our algorithms, which is the improvement in the abundance matrix estimation by simultaneously exploiting sparsity and low rankness.

Moreover, the given results indicate that the optimal choice of τ, γ depends on the particular structure (sparse and/or low-rank) of the abundance matrix. Thus, proper selection of these parameters shall involve fine-tuning schemes, which are commonplace when it comes to algorithms dealing with regularized inverse problems.

5) Performance in the Presence of Noise: In this experiment, we aim at exhibiting the performance of the proposed algorithms in the presence of white and correlated noise corruption. To this end, we stick with a specific simultaneously sparse and low-rank abundance matrix $\mathbf{W}$ of sparsity level 20% and rank-3. Based on this $\mathbf{W}$, $K = 9$ linearly mixed pixels are generated, in the same way as previously described. Then, depending on the case, white or colored Gaussian noise contaminates the data. Sixteen SNR values are considered ranging from 10 to 40 dB, whereas 100 realizations are run for each SNR value, and the mean of the RMSE and SRE metrics is calculated.

- *White Gaussian Noise*: Fig. 5 shows the RMSE and SRE curves obtained for the proposed IPSpLRU and ADSpLRU and the three competing algorithms, namely, CSUnSAL, MMV-ADMM, and BiICE. It is easily seen that both IPSpLRU and ADSpLRU attain remarkably better results comparing with CSUnSAL, MMV-ADMM, and BiICE in all the examined SNR values. Additionally,
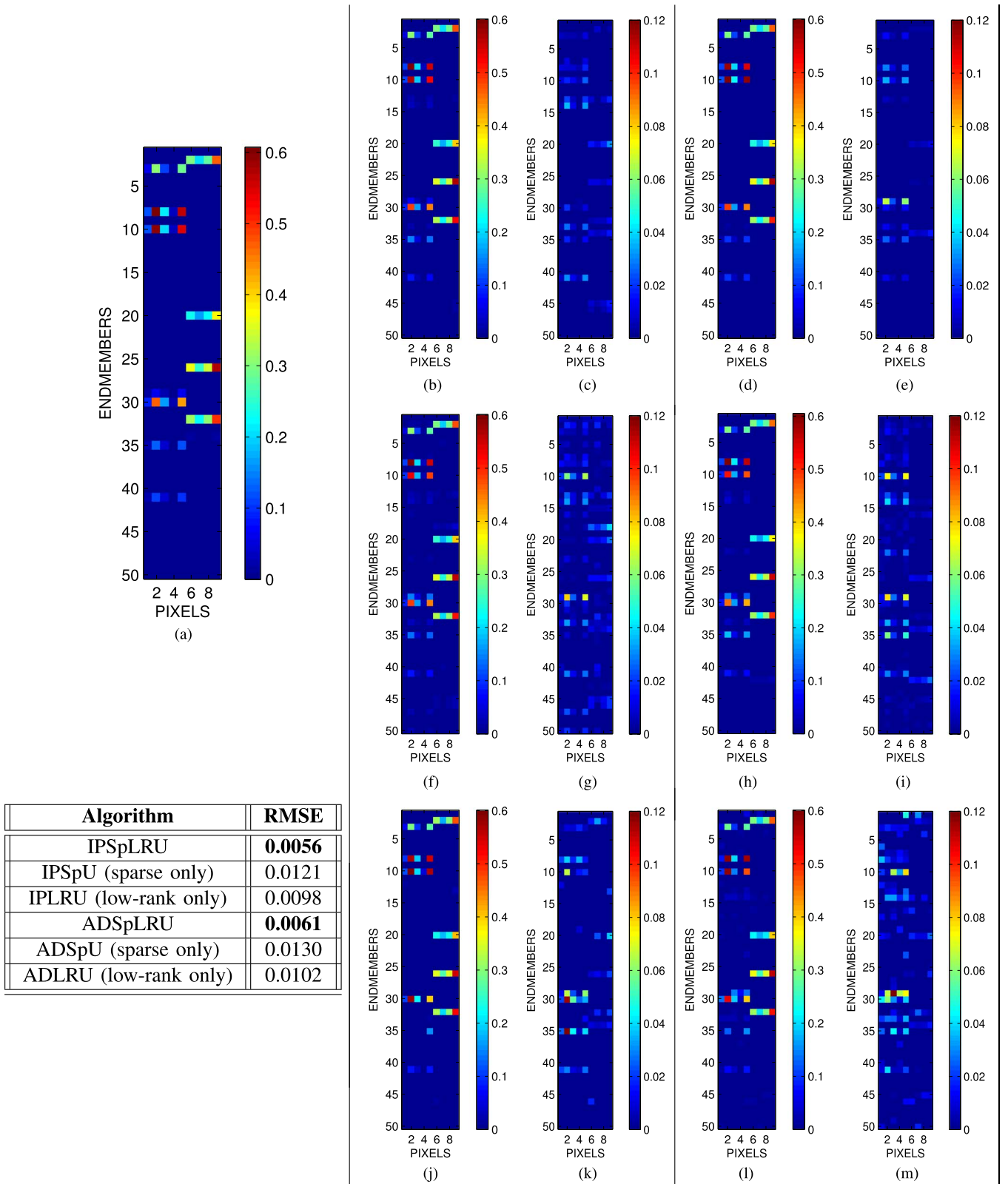
Algorithm	RMSE
IPSpLRU	**0.0056**
IPSpU (sparse only)	0.0121
IPLRU (low-rank only)	0.0098
ADSpLRU	**0.0061**
ADSpU (sparse only)	0.0130
ADLRU (low-rank only)	0.0102

Fig. 3. Proposed sparse and low-rank algorithms versus their sparse-only and low-rank-only counterparts. (a) $\mathbf{W}$, Ground truth. (b) $\hat{\mathbf{W}}$, IPSpLRU. (c) residual, IPSpLRU. (d) $\hat{\mathbf{W}}$, ADSpLRU. (e) residual, ADSpLRU. (f) $\hat{\mathbf{W}}$, IPLRU. (g) residual, IPLRU. (h) $\hat{\mathbf{W}}$, ADLRU. (i) residual, ADLRU. (j) $\hat{\mathbf{W}}$, IPSpU. (k) residual, IPSpU. (l) $\hat{\mathbf{W}}$, ADSpU. (m) residual, ADSpU.

we note that ADSpLRU performs slightly better as compared with IPSpLRU, particularly for SNR values greater than 32 dB. The price to be paid is that the computational complexity per iteration of ADSpLRU is higher than that of IPSpLRU. It is hence shown that sparse and low-rank methods are robust to different levels of white noise. At
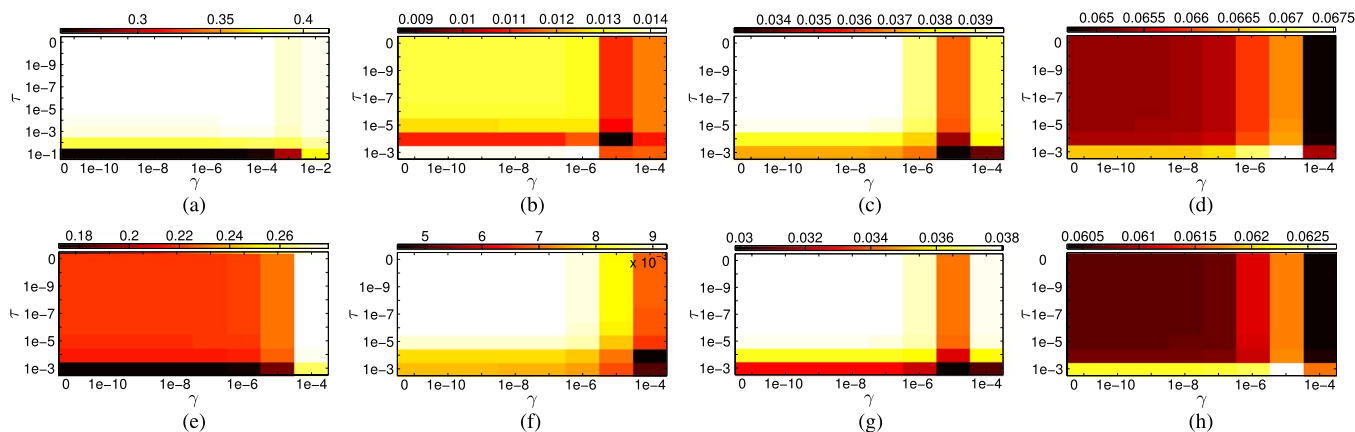
Fig. 4. RMSE as a function of the low rankness and the sparsity regularization parameters τ and γ, respectively. (a) Sparsity level = 100%, rank = 1, IPSpLRU. (b) Sparsity level = 10%, rank = 4, IPSpLRU. (c) Sparsity level = 20%, rank = 5, IPSpLRU. (d) Sparsity level = 10%, rank = 9, IPSpLRU. (e) Sparsity level = 100%, rank = 1, ADSpLRU. (f) Sparsity level = 10%, rank = 4, ADSpLRU. (g) Sparsity level = 20%, rank = 5, ADSpLRU. (h) Sparsity level = 10%, rank = 9, ADSpLRU.
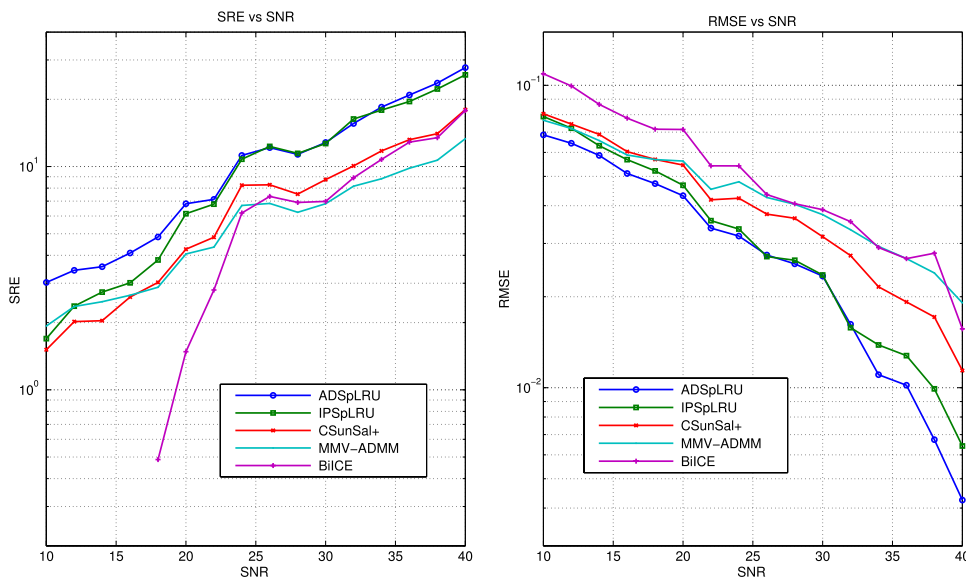

Fig. 5. Performance in the presence of white noise (SRE and RMSE).

the same time, IPSpLRU and ADSpLRU outperform the sparse-only CSUnSAL and BiICE algorithms as well as the joint-sparse MMV-ADMM algorithm, provided that both sparsity and low rankness characterize the abundance matrix.

- *Colored Gaussian Noise*: In real HSIs, the noise that corrupts the data is rather structured than white. Thus, to assess the behavior of the proposed methods in such realistic conditions, we simulate correlated Gaussian noise that adds up to the linearly mixed pixels. Fig. 6 illustrates the effectiveness of the tested algorithms in terms of RMSE and SRE, for different SNR values. Therein as well, we can see that IPSpLRU and ADSpLRU achieve better results than their competing algorithms in the whole range of the examined SNRs. Furthermore, ADSpLRU performs better for high SNR values (> 32 dB), as compared with IPSpLRU. As a result, the robustness of our proposed methods is also corroborated in the presence of correlated noise with different magnitude levels.

6) Synthetic Image: This experiment highlights the effectiveness of the proposed methods in estimating sparse, low-rank, or both sparse and low-rank abundance matrices. Focused on this purpose, we form a simulated HSI using the LMM (1) and the same aforementioned endmembers' dictionary $\mathbf{\Phi}$. As shown in Fig. 7(a), the simulated HSI consists of four rows, each consisting of four 10×10 blocks of pixels. Each of the "block rows" is generated by abundance matrices of a distinct structure. To be more specific, the first row is generated by joint-sparse $\mathbf{W}$'s, the second by solely low-rank $\mathbf{W}$'s, whereas rows 3 and 4 are produced by simultaneously sparse and low-rank abundance matrices. The pixels in each block correspond to abundance matrices of a particular combination of sparsity level and rank. The detailed description of these structures is depicted in the table of Fig. 7(b). The linearly mixed pixels are corrupted by white Gaussian i.i.d. noise such that SNR = 30 dB. The table in Fig. 7(c) contains the obtained RMSE and SRE for all algorithms tested. It is worth pointing out that our introduced IPSpLRU and ADSpLRU algorithms outperform
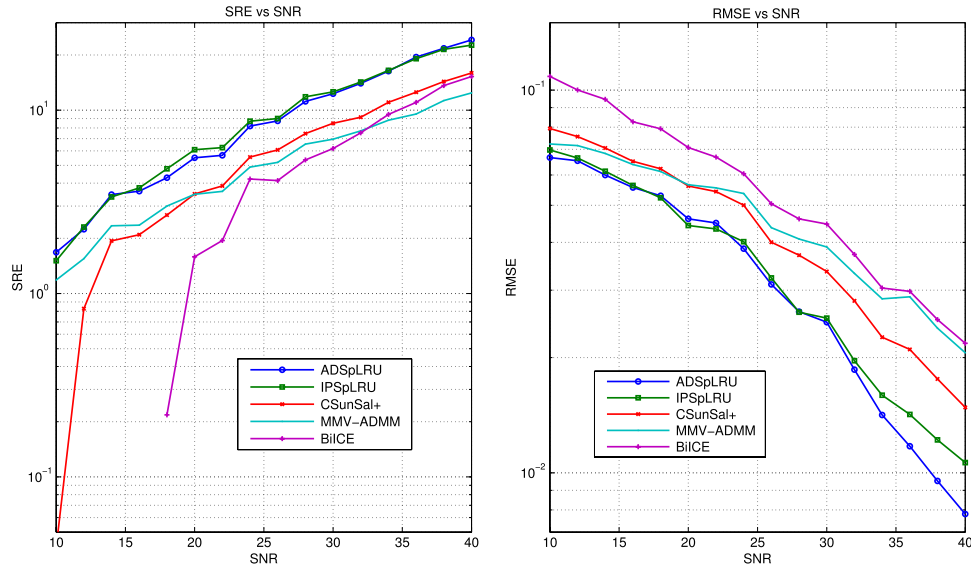
Fig. 6. Performance in the presence of colored noise (SRE and RMSE).

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

(b)

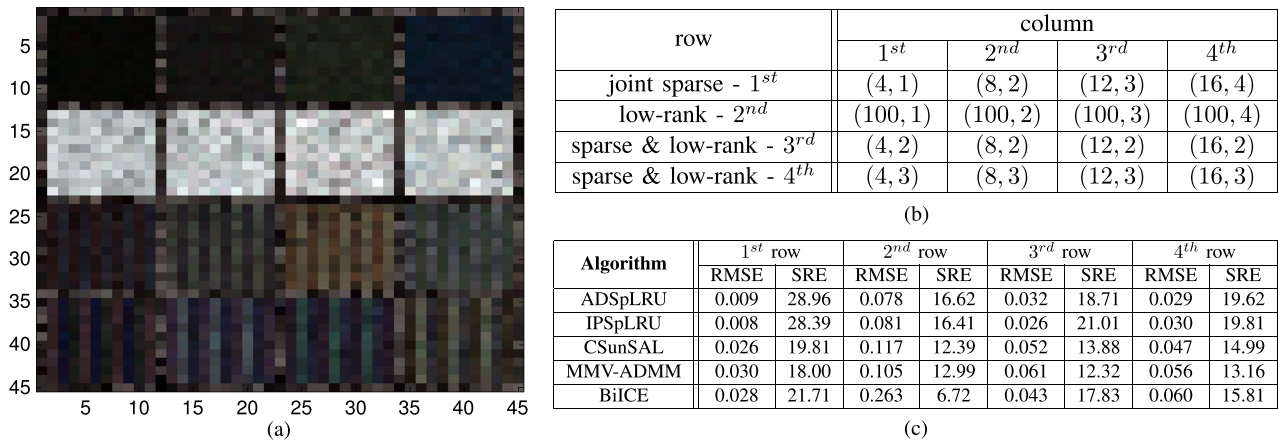| Algorithm	1st row		2nd row		3rd row		4th row	
	RMSE	SRE	RMSE	SRE	RMSE	SRE	RMSE	SRE
ADSpLRU	0.009	28.96	0.078	16.62	0.032	18.71	0.029	19.62
IPSpLRU	0.008	28.39	0.081	16.41	0.026	21.01	0.030	19.81
CSunSAL	0.026	19.81	0.117	12.39	0.052	13.88	0.047	14.99
MMV-ADMM	0.030	18.00	0.105	12.99	0.061	12.32	0.056	13.16
BiICE	0.028	21.71	0.263	6.72	0.043	17.83	0.060	15.81

(c)

Fig. 7. Structure of the synthetic image and results. (a) Synthetic image, 16 blocks of size 10×10 pixels each. (b) Structure of $\mathbf{W}$ in each block of the synthetic image, each cell contains the pair: [sparsity-level%, rank($\mathbf{W}$)]. (c) RMSE and SRE (dB) results on synthetic image for each row.
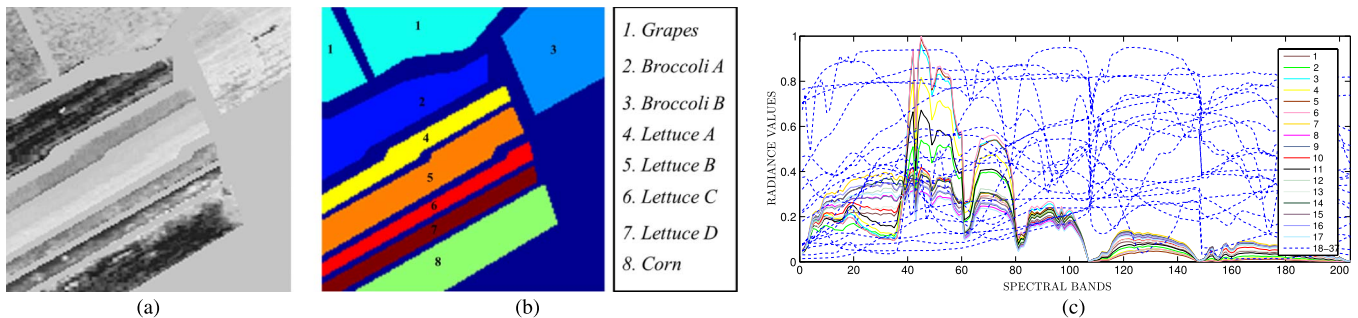

Fig. 8. Salinas Valley image and endmembers' dictionary. (a) Fifth PC of the Salinas Valley scene. (b) Rough ground truth information for a part of the Salinas Valley scene under study. (c) Spectral signatures of the 37 endmembers, 17 of them manually selected from the scene as pure pixels and 20 (dashed curves) randomly chosen from the USGS library [38].

their rivals, not only in "both sparse and low-rank" rows 3 and 4 but also in rows 1 and 2 that correspond to either sparse-only or low-rank-only $\mathbf{W}$'s.

C. Experiment on Real Data

This section illustrates the performance of the proposed algorithms when applied on a real HSI. The hyperspectral scene under examination is a portion of the widely used Salinas vegetation scene acquired by the AVIRIS sensor over Salinas Valley in California. This scene contains eight different vegetation species, namely, grapes, brocolli_A, brocolli_B, lettuce_a, lettuce_b, lettuce_c, lettuce_d, and corn, as shown in Fig. 8(b). Salinas HSI consists of $L = 204$ spectral bands (after excluding 20 noisy bands), and its spatial resolution is 3.7 m. Taking the principal components (PCs) of the image, it can be seen
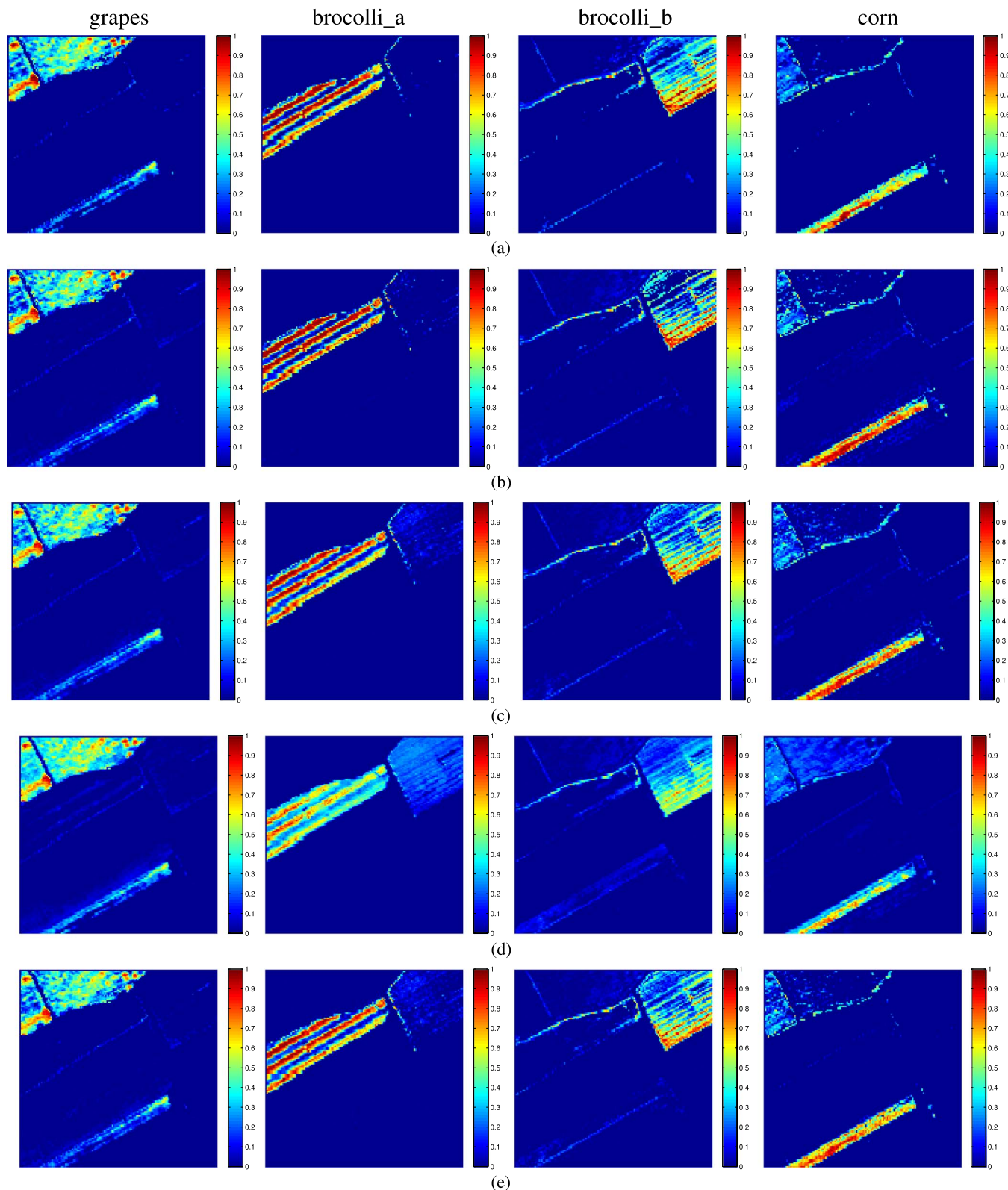
Fig. 9. Abundance maps of Salinas HSI. (a) IPSpLRU. (b) ADSpLRU. (c) CSUnSAL. (d) MMV-ADMM. (e) BiICE.

that only the first six of them contain meaningful information. Focusing on them, we can see that the first PCs give more rough information on the formation of the vegetation, whereas less significant PCs give more refined information on the vegetation formation [39]. Fig. 8(a) shows the fifth PC of the scene under study, where most of the vegetation is depicted. To make things

more interesting, the endmembers' dictionary Φ is composed of 37 spectral signatures, 17 of them manually selected from the scene, as in [40], and 20 randomly chosen from the USGS library [38]. As depicted in Fig. 8(c), the 20 USGS endmembers (blue dashed curves) significantly differ from the other 17 pure pixel spectral signatures. This is so, since those signatures

correspond to materials (minerals, organic and volatile compounds, etc.) nonexisting in the region under study, whereas the rest of the 17 endmembers correspond to the various vegetation types existing in the scene. However, the USGS endmembers were purposely included in the dictionary for investigating the competence of the proposed algorithms in distinguishing the present endmembers over the nonpresent endmembers, by exploiting sparsity in the abundance matrices.

Fig. 9 shows abundance maps corresponding to the region of interest, as obtained by the proposed IPSpLRU and ADSpLRU and the three state-of-the-art competing algorithms, namely, CSUnSAL, MMV-ADMM, and BiICE for $\gamma = 10^{-3}$, $\tau = 10^{-4}$, $\lambda = 0.5$, and $\mu = 10^{-2}$. Specifically, four different maps are depicted for each algorithm, corresponding to four vegetation species, namely, grapes, brocolli_a, brocolli_b, and corn. It is worth pointing out that since detailed ground truth information is not available, the evaluation is done in qualitative terms. From a careful visual inspection of the generated maps, we can see that the abundances obtained by IPSpLRU and ADSpLRU present patterns that are closer to those revealed by the first five PCs of the HSI provided in [40]. This is particularly clear for the maps corresponding to brocolli_a and brocolli_b. More specifically, it is shown that the presence of these two species, which is mainly located in two distinct regions, is better emphasized by the proposed algorithms. Remarkably, the erroneous detection of these vegetation types is eliminated more effectively by IPSpLRU and ADSpLRU, as also verified by comparing Figs. 8(a) and 9. It should be also noted that both IPSpLRU and ADSpLRU are proven competent in discerning the 20 USGS (nonexisting in the scene) endmembers contained in the dictionary, by efficiently exploiting their sparsity characteristic for imposing zero values on the respective abundances. Hence, it is corroborated that the exploitation of both sparsity and the inherent spatial correlation existing in HSIs can lead to qualitatively better results, thus verifying the significance of our approach.

V. Conclusion and Future Directions

In this paper, we have presented a novel approach for performing HSI unmixing, simultaneously exploiting sparsity and spatial correlation. A novel cost function was first introduced comprising a least squares proximity component regularized by a linear combination of the weighted ℓ_1-norm and the weighted nuclear norm of the latent abundance matrix. The unmixing problem was thus treated as a sparse reduced-rank regression problem. Two different algorithms were then developed for solving it, namely, an incremental-proximal-type algorithm called IPSpLRU and an ADMM-based strategy called ADSpLRU. Extensive simulations on both synthetic and real data corroborate the effectiveness of the proposed approach and algorithms, compared with other related state-of-the-art unmixing schemes. The derivation of more computationally efficient schemes alleviating the need for SVD is under current investigation. Another relevant future research direction is the exploitation of the specific structure or pattern of sparsity in the abundance matrices implicitly imposed by the low-rankness

property, which could further improve estimation performance. This is also a topic of interest in the framework of a future work.

References

[1] W. Ma *et al.*, "A signal processing perspective on hyperspectral unmixing: Insights from remote sensing," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 67–81, Jan. 2014.

[2] J. Nascimento and J. Bioucas-Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.

[3] J. Li and J. Bioucas-Dias, "Minimum volume simplex analysis: A fast algorithm to unmix hyperspectral data," in *Proc. IEEE IGARSS*, Jul. 2008, vol. 3, pp. III-250–III-253.

[4] K. E. Themelis, A. A. Rontogiannis, and K. D. Koutroumbas, "Semi-supervised hyperspectral unmixing via the weighted lasso," in *Proc. IEEE ICASSP*, Mar. 2010, pp. 1194–1197.

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

[6] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, Jun. 2011.

[7] K. E. Themelis, A. A. Rontogiannis, and K. D. Koutroumbas, "A novel hierarchical Bayesian approach for sparse semi-supervised hyperspectral unmixing," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 585–599, Feb. 2012.

[8] 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 *Proc. 5th WHISPERS*, Jun. 2013, pp. 974–978.

[9] O. Eches, N. Dobigeon, and J.-Y. Tourneret, "Enhancing hyperspectral image unmixing with spatial correlations," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4239–4247, Nov. 2011.

[10] Q. Qu, N. Nasrabadi, and T. Tran, "Abundance estimation for bilinear mixture models via joint sparse and low-rank representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 4404–4423, Jul. 2014.

[11] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Collaborative sparse regression for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 341–354, Jan. 2014.

[12] P. V. Giampouras, K. E. Themelis, A. A. Rontogiannis, and K. D. Koutroumbas, "A variational Bayes algorithm for joint-sparse abundance estimation," in *Proc. 6th WHISPERS*, Laussane, Switzerland, Jun. 2014.

[13] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[14] L. Chen and J. Z. Huang, "Sparse reduced-rank regression for simultaneous dimension reduction and variable selection," *J. Amer. Stat. Assoc.*, vol. 107, no. 500, pp. 1533–1545, Dec. 2012.

[15] P.-A. Savalle, E. Richard, and N. Vayatis, "Estimation of simultaneously sparse and low rank matrices," in *Proc. ICML*, Jun. 2012, pp. 1–8.

[16] S. Oymak, A. Jalali, M. Fazel, Y. Eldar, and B. Hassibi, "Simultaneously structured models with application to sparse and low-rank matrices," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2886–2908, May 2015.

[17] M. Golbabaee and P. Vandergheynst, "Compressed sensing of simultaneous low-rank and joint-sparse matrices," *arXiv preprint arXiv:1211.5058*, 2012.

[18] E. Richard, G. R. Obozinski, and J.-P. Vert, "Tight convex relaxations for sparse matrix factorization," in *Advances in Neural Information Processing Systems 27*. Red Hook, NY, USA: Curran Associates, Inc., 2014, pp. 3284–3292.

[19] D. P. Bertsekas, "Incremental proximal methods for large scale convex optimization," *Math. Programm.*, vol. 129, no. 2, pp. 163–195, Oct. 2011.

[20] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.

[21] S. Negahban and M. J. Wainwright, "Estimation of (near) low-rank matrices with noise and high-dimensional scaling," *Ann. Stat.*, vol. 39, no. 2, pp. 1069–1097, Apr. 2011.

[22] S. Babacan, M. Luessi, R. Molina, and A. Katsaggelos, "Sparse Bayesian methods for low-rank matrix estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 3964–3977, Aug. 2012.

[23] F. R. Bach, "Consistency of trace norm minimization," *J. Mach. Learn. Res.*, vol. 9, pp. 1019–1048, Jun. 2008.

[24] W. Tang, Z. Shi, Y. Wu, and C. Zhang, "Sparse unmixing of hyperspectral data using spectral a priori information," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 770–783, Feb. 2015.

[25] H. Zou, "The adaptive lasso and its oracle properties," *J. Amer. Stat. Assoc.*, vol. 101, no. 476, pp. 1418–1429, Dec. 2006.

[26] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ1 minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5/6, pp. 877–905, Nov. 2008.

[27] C. Lu, J. Tang, S. Yan, and Z. Lin, "Generalized nonconvex nonsmooth low-rank minimization," in *Proc. IEEE CVPR*, Jun. 2014, pp. 4130–4137.

[28] Y.-D. Kim and S. Choi, "Variational Bayesian view of weighted trace norm regularization for matrix factorization," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 261–264, Mar. 2013.

[29] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *Proc. IEEE Conf. Comput. Visi. Pattern Recog.*, 2014, pp. 2862–2869.

[30] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, "Convex optimization with sparsity-inducing norms," in *Optimization for Machine Learning*. Cambridge, MA, USA: MIT Press, 2011, pp. 19–53.

[31] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*. New York, NY, USA: Springer-Verlag, 2011, pp. 185–212.

[32] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 123–231, 2013.

[33] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD, USA: The Johns Hopkins Univ. Press, 2012.

[34] J. Eckstein and D. P. Bertsekas, "On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Programm.*, vol. 55, no. 1–3, pp. 293–318, Apr. 1992.

[35] W. Dong, G. Shi, X. Li, Y. Ma, and F. Huang, "Compressive sensing via nonlocal low-rank regularization," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3618–3632, Aug. 2014.

[36] F. Chen and Y. Zhang, "Sparse hyperspectral unmixing based on constrained lp–l2 optimization," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 5, pp. 1142–1146, Sep. 2013.

[37] X. Fu, W.-K. Ma, J. Bioucas-Dias, and T.-H. Chan, "Semiblind hyperspectral unmixing in the presence of spectral library mismatches," *arXiv preprint arXiv:1507.01661*, 2015.

[38] R. N. Clark *et al.*, USGS Digital Spectral Library, 2007. [Online]. Available: http://speclab.cr.usgs.gov/spectral.lib06/ds231/datatable.html

[39] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 4th ed. San Francisco, CA, USA: Academic, 2008.

[40] E. Mylona, O. Sykioti, K. Koutroumbas, and A. A. Rontogiannis, "Joint spectral unmixing and clustering for identifying homogeneous regions in hyperspectral images," in *Proc. IEEE IGARSS*, Jul. 2015, pp. 2409–1412.

Paris V. Giampouras was born in Athens, Greece, in 1986. He received the Diploma degree in electrical and computer engineering from the National Technical University of Athens, Athens, in 2011 and the M.Sc. degree in information technologies in medicine and biology from the University of Athens, Athens, in 2014, where he is currently working toward the Ph.D. degree.

His main research interests include signal processing and machine learning with applications to the processing of hyperspectral images and other kinds of large-scale data sets.

Konstantinos E. Themelis was born in Piraeus, Greece, in 1981. He received the Diploma degree in computer engineering and informatics from the University of Patras, Patras, Greece, in 2005 and the Ph.D. degree in signal processing from the University of Athens, Athens, Greece, in 2012.

Since 2012, he has been a Postdoctoral Research Associate with the Institute for Astronomy, Astrophysics, Space Applications and Remote Sensing, National Observatory of Athens, Athens. His research interests include statistical signal processing and probabilistic machine learning with application to image processing.

Dr. Themelis is a member of the Technical Chamber of Greece.

Athanasios A. Rontogiannis (M'97) was born in Lefkada Island, Greece, in 1968. He received the Diploma degree in electrical engineering from the National Technical University of Athens, Athens, Greece, in 1991; the M.A.Sc. degree in electrical and computer engineering from the University of Victoria, Victoria, BC, Canada, in 1993; and the Ph.D. degree in communications and signal processing from the University of Athens, Athens, in 1997.

From 1998 to 2003, he was with the University of Ioannina, Ionnina, Greece. In 2003, he joined the Institute for Astronomy, Astrophysics, Space Applications and Remote Sensing, National Observatory of Athens, Athens, where he has been a Senior Researcher since 2011. His research interests are in the general areas of statistical signal processing and wireless communications, with emphasis on adaptive estimation, hyperspectral image processing, Bayesian compressive sensing, channel estimation/equalization, and cooperative communications.

Dr. Rontogiannis has been on the Editorial Board of Springer's EURASIP *Journal on Advances in Signal Processing* since 2008 and Elsevier's EURASIP *Signal Processing Journal* since 2011. He is a member of the IEEE Signal Processing and Communication Societies and the Technical Chamber of Greece.

Konstantinos D. Koutroumbas received the Diploma degree from the University of Patras, Patras, Greece, in 1989; the M.Sc. degree in advanced methods in computer science from the Queen Mary College of the University of London, London, U.K., in 1990; and the Ph.D. degree from the University of Athens, Athens, Greece, in 1995.

Since 2001, he has been with the Institute for Astronomy, Astrophysics, Space Applications and Remote Sensing, National Observatory of Athens, Athens, where he is currently a Senior Researcher. He is a coauthor of the books *Pattern Recognition* (1st, 2nd, 3rd, and 4th editions) and *Introduction to Pattern Recognition: A MATLAB Approach*. He has over 2500 citations in his work. His research interests include pattern recognition, time-series estimation, and their applications to remote sensing and the estimation of characteristic quantities of the upper atmosphere.