

Blind and fully constrained unmixing of hyperspectral images

Rita Ammanouil, André Ferrari, *Member, IEEE*, Cédric Richard, *Senior Member, IEEE*,
David Mary, *Member, IEEE*

Abstract—This paper addresses the problem of blind and fully constrained unmixing of hyperspectral images. Unmixing is performed without the use of any dictionary, and assumes that the number of constituent materials in the scene and their spectral signatures are unknown. The estimated abundances satisfy the desired sum-to-one and nonnegativity constraints. Two models with increasing complexity are developed to achieve this challenging task, depending on how noise interacts with hyperspectral data. The first one leads to a convex optimization problem, and is solved with the Alternating Direction Method of Multipliers. The second one accounts for signal-dependent noise, and is addressed with a Reweighted Least Squares algorithm. Experiments on synthetic and real data demonstrate the effectiveness of our approach.

I. INTRODUCTION

Hyperspectral imaging is a continuously growing area of remote sensing, which has received considerable attention in the last decade. Hyperspectral data provide spectral images over hundreds of narrow and adjacent bands, coupled with a high spectral resolution. These characteristics are suitable for detection and classification of surfaces and chemical elements in the observed images. Applications include land use analysis, pollution monitoring, wide-area reconnaissance, and field surveillance, to cite a few. When unmixing hyperspectral images [1], two types of pixels can be distinguished: the pure pixels and the mixed ones. Each pure pixel, also called endmember, contains the spectral signature of a constituent material in the scene, whereas a mixed pixel consists of a mixture of the endmembers. The fraction of each endmember in a mixed pixel is called abundance. Three consecutive tasks are usually required for unmixing: determining the number of endmembers, extracting the spectral signature of the endmembers, and estimating their abundances for every pixel in the scene. Several algorithms have been proposed to perform each stage separately. Virtual Dimensionality (VD) [2], followed by N-FINDR [3] and FCLS [4] is among the most widely used processing pipeline. Alternative methods jointly perform (part of) these tasks in order to solve the blind source separation problem [5]–[8].

In order to introduce our approach, we shall now describe the noise-free case first. Consider the linear mixing model where a mixed pixel is expressed as a linear combination of

the endmembers weighted by their fractional abundances. In matrix form, we simply have

$$\tilde{S} = RA \quad (1)$$

where $\tilde{S} = (\tilde{s}_1, \dots, \tilde{s}_N)$, $R = (r_1, \dots, r_M)$, $A = (a_1, \dots, a_M)^\top$, and \tilde{s}_j is the L -dimensional spectrum of the j -th pixel, L is the number of frequency bands, r_i is L -dimensional spectrum of the i -th endmember, M is the number of endmembers, a_i is the N -dimensional abundance map of the i -th endmember, and N is the number of pixels in the image. Model (1) means that the (i, j) -th entry A_{ij} of matrix A represents the abundance of the endmember r_i in pixel \tilde{s}_j . The abundances obey the nonnegativity and sum-to-one constraints: $A_{ij} \geq 0$ for all i and j , and $\sum_{i=1}^M A_{ij} = 1$ for all j . Note that the tilde over symbols refers to noise-free data and all vectors are column vectors.

In this study, we shall assume that the endmembers are unknown but present in the scene, possibly corrupted by noise. Let ω be a subset of N' indexes in $\{1, \dots, N\}$ that contains at least the column index of each endmember. Under these assumptions, and without loss of generality, we observe that the mixing model (1) can be reformulated as follows

$$\tilde{S} = \tilde{S}_\omega X \quad (2)$$

where $\tilde{S}_\omega = (\tilde{s}_{\omega_1}, \dots, \tilde{s}_{\omega_{N'}})$ denotes the restriction of \tilde{S} to its columns indexed by ω , and $X = (x_1, \dots, x_{N'})^\top$ is the abundance matrix. Similarly as above, X_{ij} is the abundance of \tilde{s}_{ω_i} in \tilde{s}_j . On the one hand, if \tilde{s}_{ω_i} is an endmember, x_i has non-zero entries and represents the corresponding abundance map. On the other hand, if \tilde{s}_{ω_i} is a mixed pixel, x_i has all its elements equal to zero. As a consequence, X admits $N' - M$ rows of zeros, the other rows being equal to rows of A . This means that X allows to identify the endmembers in \tilde{S} through its non-zero rows, which is an interesting property to be exploited in the case where the endmembers are unknown. Let us now turn to the more realistic situation where some noise corrupts the observations. In this case, model (2) becomes

$$S = \tilde{S} + E = \tilde{S}_\omega X + E \quad (3)$$

where S denotes the available data, and E is the noise supposed to be additive.

The aim of this paper is to derive two unmixing approaches with increasing complexity, depending on how noise is to be handled. These methods are blind in the sense that the endmembers and their cardinality are unknown. The first one considers the approximate model

$$S \approx S_\omega X + E \quad (4)$$

R. Ammanouil, A. Ferrari, C. Richard and D. Mary are with the Lagrange Laboratory, University of Nice Sophia-Antipolis, CNRS, Côte d'Azur Observatory, 06108 Nice cedex 2, France (e-mail: rita.ammanouil@unice.fr; andre.ferrari@unice.fr; cedric.richard@unice.fr; david.mary@unice.fr).

This work was partly supported by the Agence Nationale pour la Recherche, France, (Hypanema project, ANR-12-BS03-003), and the regional council of Provence-Alpes-Côte d'Azur.

compared to (3), we thus assume that noise does not dramatically affect the factorization of the mixing process, which is valid for very high signal-to-noise ratio (SNR). With this approach, we shall look for a few columns of \mathbf{S}_ω that can effectively represent the whole scene. This strategy subserves a blind and self-dependent framework. It departs from methods based on a preselected dictionary of endmembers estimated from other experimental conditions, and thus do not accurately represent the endmembers in \mathbf{R} . In order to estimate the abundance matrix \mathbf{X} , we use prior information. First, we impose that the estimated abundances obey the non-negativity and sum-to-one constraints, namely, $\mathbf{X}_{ij} \geq 0$ for all (i, j) , and $\sum_{i=1}^N \mathbf{X}_{ij} = 1$ for all j . In addition, as discussed above, the algorithm has to force rows of \mathbf{X} to be zero vectors in order to identify the endmembers. Because the locations and the cardinality of the endmembers are unknown, the set of candidates has to be sufficiently large, that is, $N' \gg M$. We thus expect many rows in \mathbf{X} to be equal to zero. To promote this effect, the so-called Group Lasso $\ell_{2,1}$ -norm regularization can be employed [9]. Because model (4) is a poor approximation of model (3) as the noise power increases, we shall also propose an alternative strategy to solve the unmixing problem based on the exact model (3). The first approach leads to a convex optimization problem that can be solved with the Alternating Direction Method of Multipliers (ADMM) [10]. The second one takes the noise in \mathbf{S}_ω into account, which results in a non-convex and heteroscedastic optimization problem. The latter will be solved with an Iterative Reweighted Least Squares (IRLS) algorithm.

To the best of our knowledge, this work is the first that proposes to solve the noisy problem (3). Few models similar to the approximate model (4) have been studied in the literature [11]–[14]. These last four works assume that \mathbf{S}_ω is noise-free. Moreover, in [11], the authors use an $\ell_{1,\infty}$ -norm instead of the $\ell_{2,1}$ -norm regularization, and incorporate an additional ℓ_1 -norm instead of the unit-sum constraint considered here. In [12], the authors derive a Matching Pursuit approach [15] in order to estimate the endmembers. With this greedy approach, neither the positivity, nor the sum-to-one constraints, are taken into account. A similar technique is considered in [13], but the authors do not assume that the endmembers are present in the scene and use a predefined dictionary. In their recent work [14], the authors of [13] apply model (4) in order to extract the endmembers from the observations.

The rest of this paper is organized as follows. Sections II and III respectively describe the unmixing models (4) and (3), and the corresponding estimation methods. Section IV provides experimental results on synthetic and real data. Finally, Section V concludes this paper.

II. GROUP LASSO WITH UNIT SUM AND POSITIVITY CONSTRAINTS (GLUP)

A. Model description

The aim of this section is to derive the estimation method for model (4), and finally define each step of the ADMM that is performed to get the solution. With this approximate model, we assume that the noise \mathbf{E} is Gaussian independent and

identically distributed, with zero mean and possibly unknown variance σ^2 , that is, $\mathbf{E}_{ki} \sim \mathcal{N}(0, \sigma^2)$. The negative log-likelihood for model (4) is given by

$$\mathcal{L}(\mathbf{X}) = \frac{NL}{2} \log(2\pi) + \frac{NL}{2} \log(\sigma^2) + \frac{1}{2\sigma^2} \|\mathbf{S} - \mathbf{S}_\omega \mathbf{X}\|_F^2 \quad (5)$$

The Maximum Likelihood (ML) estimate, namely, the minimizer of $\mathcal{L}(\mathbf{X})$, is the solution of the Least Squares (LS) approximation problem $\min_{\mathbf{X}} \|\mathbf{S} - \mathbf{S}_\omega \mathbf{X}\|_F^2$. Since model (4) follows from an approximation of model (3), the relevance of this LS fidelity term is essentially to ensure that $\mathbf{S}_\omega \mathbf{X}$ matches \mathbf{S} . The unmixing problem under investigation, however, requires that \mathbf{X} only has a few rows different from zero, in addition to the non-negativity and sum-to-one constraints. This leads to following convex optimization problem

$$\begin{aligned} \min_{\mathbf{X}} \quad & \frac{1}{2} \|\mathbf{S} - \mathbf{S}_\omega \mathbf{X}\|_F^2 + \mu \sum_{k=1}^N \|\mathbf{x}_k\|_2 \\ \text{subject to} \quad & \mathbf{X}_{ij} \geq 0 \quad \forall i, j \\ & \sum_{i=1}^N \mathbf{X}_{ij} = 1 \quad \forall j \end{aligned} \quad (6)$$

with $\mu \geq 0$ a regularization parameter and \mathbf{x}_k the k -th row of \mathbf{X} . The Group Lasso regularization term induces sparsity in the estimated abundance matrix at the group level [9], by possibly driving several rows \mathbf{x}_k of \mathbf{X} to zero. It is worth noting that when $\mu = 0$ and $\mathbf{S}_\omega = \mathbf{S}$, the identity matrix is a solution of problem (6). This solution may not be unique depending on \mathbf{S} . It follows that the efficiency of our approach relies on the $\ell_{2,1}$ -norm regularization function.

B. ADMM algorithm

The solution of problem (6) can be obtained in a simple and flexible manner using the ADMM algorithm [10]. We consider the canonical form

$$\begin{aligned} \min_{\mathbf{X}, \mathbf{Z}} \quad & \frac{1}{2} \|\mathbf{S} - \mathbf{S}_\omega \mathbf{X}\|_F^2 + \mu \sum_{k=1}^N \|\mathbf{z}_k\|_2 + \mathcal{I}(\mathbf{Z}) \\ \text{subject to} \quad & \mathbf{A}\mathbf{X} + \mathbf{B}\mathbf{Z} = \mathbf{C} \end{aligned} \quad (7)$$

with

$$\mathbf{A} = \begin{pmatrix} \mathbf{I} \\ \mathbf{1}^\top \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} -\mathbf{I} \\ \mathbf{0}^\top \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} \mathbf{0} \\ \mathbf{1}^\top \end{pmatrix},$$

where \mathcal{I} is the indicator of the positive orthant guarantying the positivity constraint, that is, $\mathcal{I}(\mathbf{Z}) = 0$ if $\mathbf{Z} \succeq \mathbf{0}$ and $+\infty$ otherwise. The equality constraint $\mathbf{A}\mathbf{X} + \mathbf{B}\mathbf{Z} = \mathbf{C}$ imposes the consensus $\mathbf{X} = \mathbf{Z}$ and the sum-to-one constraint. Note that defining the constraint matrices differently, in particular setting $\mathbf{A} = \mathbf{I}$, $\mathbf{B} = -\mathbf{I}$, and $\mathbf{C} = \mathbf{0}$ allows to relax the sum-to-one constraint. In matrix form, the augmented Lagrangian for problem (7) is given by [16]

$$\begin{aligned} \mathcal{L}_\rho(\mathbf{X}, \mathbf{Z}, \mathbf{\Lambda}) = & \frac{1}{2} \|\mathbf{S} - \mathbf{S}_\omega \mathbf{X}\|_F^2 + \mu \sum_{k=1}^N \|\mathbf{z}_k\|_2 + \mathcal{I}(\mathbf{Z}) \\ & + \text{trace}(\mathbf{\Lambda}^\top (\mathbf{A}\mathbf{X} + \mathbf{B}\mathbf{Z} - \mathbf{C})) + \frac{\rho}{2} \|\mathbf{A}\mathbf{X} + \mathbf{B}\mathbf{Z} - \mathbf{C}\|_F^2 \end{aligned}$$

where $\mathbf{\Lambda}$ is the matrix of Lagrange multipliers, μ and ρ are positive regularization and penalty parameters, respectively. The flexibility of the ADMM lies in the fact that it splits the initial variable \mathbf{X} into two variables, \mathbf{X} and \mathbf{Z} , and equivalently the initial problem into two subproblems. At iteration $k + 1$, the ADMM algorithm is outlined by three sequential steps.

1) *Minimization of $\mathcal{L}_\rho(\mathbf{X}, \mathbf{Z}^k, \mathbf{\Lambda}^k)$ with respect to \mathbf{X} :*

This step takes into account the previous estimates of \mathbf{Z} and $\mathbf{\Lambda}$. The augmented Lagrangian is quadratic in terms of \mathbf{X} . As a result, the solution has an analytical expression that is obtained by setting the gradient of $\mathcal{L}_\rho(\mathbf{X}, \mathbf{Z}^k, \mathbf{\Lambda}^k)$ to zero:

$$\mathbf{X}^{k+1} = (\mathbf{S}_\omega^\top \mathbf{S}_\omega + \rho \mathbf{A}^\top \mathbf{A})^{-1} (\mathbf{S}_\omega^\top \mathbf{S} - \mathbf{A}^\top [\mathbf{\Lambda}^k + \rho (\mathbf{BZ}^k - \mathbf{C})]). \quad (8)$$

2) *Minimization of $\mathcal{L}_\rho(\mathbf{X}^{k+1}, \mathbf{Z}, \mathbf{\Lambda}^k)$ with respect to \mathbf{Z} :*

After discarding the terms that are independent of \mathbf{Z} , the minimization of $\mathcal{L}_\rho(\mathbf{X}^{k+1}, \mathbf{Z}, \mathbf{\Lambda}^k)$ with respect to \mathbf{Z} reduces to solving the following problem:

$$\begin{aligned} \min_{\mathbf{Z}} \quad & \mu \sum_{k=1}^N \|\mathbf{z}_k\|_2 + \text{trace}(\mathbf{\Lambda}^\top \mathbf{BZ}) \\ & + \frac{\rho}{2} \|\mathbf{AX} + \mathbf{BZ} - \mathbf{C}\|_F^2 \\ \text{subject to} \quad & \mathbf{Z} \succeq \mathbf{0}. \end{aligned} \quad (9)$$

This minimization step can be split into N problems given the structure of matrices \mathbf{A} and \mathbf{B} , one for each row of \mathbf{Z} , that is,

$$\min_{\mathbf{z}} \quad \frac{1}{2} \|\mathbf{z} - \mathbf{v}\|_2^2 + \alpha \|\mathbf{z}\|_2 + \mathcal{I}(\mathbf{z}) \quad (10)$$

where $\mathbf{v} = \mathbf{x} + \rho^{-1} \mathbf{\lambda}$, $\alpha = \rho^{-1} \mu$. Vectors $\mathbf{\lambda}$, \mathbf{x} and \mathbf{z} correspond to a given row of $\mathbf{\Lambda}$, \mathbf{X} and \mathbf{Z} , respectively. The minimization problem (10) admits a unique solution given by the proximity operator [17] of $f(\mathbf{z}) = \alpha \|\mathbf{z}\|_2 + \mathcal{I}(\mathbf{z})$:

$$\begin{cases} \mathbf{z}^* = \mathbf{0} & \text{if } \|(v)_+\|_2 < \alpha \\ \mathbf{z}^* = \left(1 - \frac{\alpha}{\|(v)_+\|_2}\right) (v)_+ & \text{otherwise} \end{cases} \quad (11)$$

where $(\cdot)_+ = \max(\mathbf{0}, \cdot)$. On the one hand, the proximity operator of $f_1(\mathbf{z}) = \alpha \|\mathbf{z}\|_2$ is the *Multidimensional Shrinkage Thresholding Operator* (MiSTO) [18]. On the other hand, the proximity operator of the indicator function $f_2(\mathbf{z}) = \mathcal{I}(\mathbf{z})$ is the projection onto the positive orthant. The proximity operator of $f(\mathbf{z})$ in (11), that we refer to as Positively constrained MiSTO, is an extension of both previous operators. The solution is of the form $\text{prox}_f = \text{prox}_{f_1} \circ \text{prox}_{f_2}$, that is, the thresholding of the projection. Operator (11) was recently used in [19]. The derivation of this operator can be found in the Appendix.

3) *Update of the Lagrange multipliers $\mathbf{\Lambda}$:* Update of the Lagrange multipliers is carried out at the end of each iteration. $\mathbf{\Lambda}^{k+1}$ represents the running sum of residuals. It gives an insight on the convergence of the algorithm. As k tends to infinity, the primal residual tends to zero and $\mathbf{\Lambda}^{k+1}$ converges to the dual optimal point.

$$\mathbf{\Lambda}^{k+1} = \mathbf{\Lambda}^k + \rho (\mathbf{AX}^{k+1} + \mathbf{BZ}^{k+1} - \mathbf{C}). \quad (12)$$

As suggested in [10], a reasonable stopping criteria is that the primal and dual residuals must be smaller than some tolerance thresholds, namely,

$$\begin{aligned} \|\mathbf{AX}^{k+1} + \mathbf{BZ}^{k+1} - \mathbf{C}\|_2 &\leq \epsilon_{\text{pri}} \\ \|\rho \mathbf{A}^\top \mathbf{B}(\mathbf{Z}^{k+1} - \mathbf{Z}^k)\|_2 &\leq \epsilon_{\text{dual}}. \end{aligned} \quad (13)$$

The pseudocode for the so-called GLUP method is provided by Algorithm 1. It is worth emphasizing that the main difference between the ADMM steps developed in GLUP and those in [13] arises in the ADMM variable splitting. The global

problem in [13] is decomposed into three subproblems: the least squares minimization, the Group Lasso regularization, and projection on the positive orthant. A consequence is that three ADMM variables are used instead of two, which leads to additional steps. In addition, the sum-to-one constraint is not considered in [13].

Algorithm 1 : $\mathbf{X} = \text{GLUP}(\mathbf{S}, \mathbf{S}_\omega, \rho, \mu)$

```

1: Precompute  $\mathbf{A}$ ,  $\mathbf{B}$ , and  $\mathbf{C}$ 
2: Initialize  $\mathbf{Z} = \mathbf{0}$  and  $\mathbf{\Lambda} = \mathbf{0}$ 
3:  $\mathbf{Q} = (\mathbf{S}_\omega^\top \mathbf{S}_\omega + \rho \mathbf{A}^\top \mathbf{A})^{-1}$ 
4: while  $\|\mathbf{R}\|_2 \geq \epsilon_{\text{pri}}$  or  $\|\mathbf{P}\|_2 \geq \epsilon_{\text{dual}}$  do
5:    $\mathbf{X} = \mathbf{Q}(\mathbf{S}_\omega^\top \mathbf{S} - \mathbf{A}^\top (\mathbf{\Lambda} + \rho [\mathbf{BZ} - \mathbf{C}]))$ 
6:    $\mathbf{Z}^{\text{old}} = \mathbf{Z}$ 
7:   for  $i = 1 \dots N'$  do
8:      $\mathbf{v}_i = ((\mathbf{x}_i)^\top + \rho^{-1} \mathbf{\lambda}_i)_+$ 
9:     if  $\|\mathbf{v}_i\|_2 < \rho^{-1} \mu$  then
10:       $\mathbf{z}_i = \mathbf{0}$ 
11:     else
12:       $\mathbf{z}_i = \left(1 - \frac{\mu}{\rho \|\mathbf{v}_i\|_2}\right) \mathbf{v}_i$ 
13:     end if
14:   end for
15:    $\mathbf{R} = \mathbf{AX} + \mathbf{BZ} - \mathbf{C}$ 
16:    $\mathbf{P} = \rho \mathbf{AB}(\mathbf{Z} - \mathbf{Z}^{\text{old}})$ 
17:    $\mathbf{\Lambda} = \mathbf{\Lambda} + \rho (\mathbf{AX} + \mathbf{BZ} - \mathbf{C})$ 
18: end while

```

III. REDUCED NOISE FOR GROUP LASSO WITH UNIT SUM AND POSITIVITY CONSTRAINTS (NGLUP)

A. Model description

We now turn to the more realistic model (3). Let \mathbf{E}_ω and \mathbf{I}_ω be the L -by- N' and N -by- N' restrictions of \mathbf{E} and \mathbf{I} to the columns indexed by ω , respectively. The noisy mixing model (3) is given by

$$\mathbf{S} = (\mathbf{S}_\omega - \mathbf{E}_\omega) \mathbf{X} + \mathbf{E} = \mathbf{S}_\omega \mathbf{X} + \mathbf{E}(\mathbf{I} - \mathbf{I}_\omega \mathbf{X}). \quad (14)$$

This model belongs to the family of heteroscedastic regression [20], where the variance of the additive noise depends on \mathbf{X} . Let us define the matrix $\mathbf{C}(\mathbf{X})$ as

$$\mathbf{C}(\mathbf{X}) = (\mathbf{I} - \mathbf{I}_\omega \mathbf{X})^\top (\mathbf{I} - \mathbf{I}_\omega \mathbf{X}). \quad (15)$$

It follows that

$$\text{vec}(\mathbf{E}(\mathbf{I} - \mathbf{I}_\omega \mathbf{X})) \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{C}(\mathbf{X}) \otimes \mathbf{I}) \quad (16)$$

where \otimes represents the Kronecker product of matrices, and $\text{vec}(\cdot)$ is the operator that stacks the columns of a matrix on top of each other. The presence of \mathbf{X} in the expression of the noise variance expression has consequences on the negative log-likelihood of model (14), which no longer leads to the LS

approximation problem

$$\begin{aligned}
\mathcal{L}(\mathbf{X}, \sigma^2) &= \frac{1}{2} \log |\sigma^2 \mathbf{C}(\mathbf{X}) \otimes \mathbf{I}| \\
&+ \frac{1}{2} \text{vec}(\mathbf{S} - \mathbf{S}_\omega \mathbf{X})^\top (\sigma^2 \mathbf{C}(\mathbf{X}) \otimes \mathbf{I})^{-1} \text{vec}(\mathbf{S} - \mathbf{S}_\omega \mathbf{X}) \\
&= \frac{L}{2} \log |\sigma^2 \mathbf{C}(\mathbf{X})| \\
&+ \frac{1}{2} \text{trace}((\mathbf{S} - \mathbf{S}_\omega \mathbf{X})(\sigma^2 \mathbf{C}(\mathbf{X}))^{-1}(\mathbf{S} - \mathbf{S}_\omega \mathbf{X})^\top) \\
&= \frac{L}{2} \log |\sigma^2 \mathbf{C}(\mathbf{X})| + \frac{1}{2} \|\mathbf{S} - \mathbf{S}_\omega \mathbf{X}\|_{(\sigma^2 \mathbf{C}(\mathbf{X}))^{-1}}^2
\end{aligned} \tag{17}$$

The ML estimate for problem (17) with the Group Lasso regularization, nonnegativity and sum-to-one constraints yields the following constrained optimization problem

$$\begin{aligned}
\min_{\mathbf{X}, \sigma^2} \quad & \frac{L}{2} \log |\sigma^2 \mathbf{C}(\mathbf{X})| + \frac{1}{2} \|\mathbf{S} - \mathbf{S}_\omega \mathbf{X}\|_{(\sigma^2 \mathbf{C}(\mathbf{X}))^{-1}}^2 \\
& + \mu \sum_{k=1}^N \|\mathbf{x}_k\|_2 \\
\text{subject to} \quad & \mathbf{X}_{ij} \geq 0 \quad \forall i, j \\
& \sum_{i=1}^N \mathbf{X}_{ij} = 1 \quad \forall j
\end{aligned} \tag{18}$$

B. Alternating ADMM algorithm

Problem (18) is not convex and requires the estimation of σ^2 . The second term in the objective function is closely related to Iteratively Reweighted Least Squares (IRLS) algorithms used as a solution in heteroscedastic models [21]. Note that, in IRLS algorithms, $(\mathbf{S} - \mathbf{S}_\omega \mathbf{X})$ in equation (17) is usually substituted by $(\mathbf{S} - \mathbf{S}_\omega \mathbf{X})^\top$. This has consequences on the \mathbf{X} minimization step. In IRLS, the estimation process is carried out in two steps. The first step consists of updating weights, which are usually set to be inversely proportional to variances. The second step is the calculation of the LS estimator using the updated weights. Many strategies can be used to estimate the variances for the weight matrix, see for example [20], [22], [23].

The resolution of problem (18) with respect to σ^2 for fixed \mathbf{X} gives

$$\sigma^2(\mathbf{X}) = \frac{1}{NL} \text{trace}((\mathbf{S} - \mathbf{S}_\omega \mathbf{X}) \mathbf{C}(\mathbf{X})^{-1} (\mathbf{S} - \mathbf{S}_\omega \mathbf{X})^\top). \tag{19}$$

Let $\mathbf{W}(\mathbf{X}) = \sigma^2(\mathbf{X}) \mathbf{C}(\mathbf{X})$ denote the weight matrix of the least squares term in (18). To solve problem (18) with respect to σ^2 and \mathbf{X} , we propose to proceed iteratively. Let \mathbf{X}^k be the solution of the previous iteration. The first step consists of calculating $\mathbf{W}(\mathbf{X}^k)$ using equations (15) and (19). In the second step, this updated weight matrix is used to estimate \mathbf{X}^{k+1} as follows

$$\begin{aligned}
\min_{\mathbf{X}} \quad & \frac{1}{2} \|\mathbf{S} - \mathbf{S}_\omega \mathbf{X}\|_{(\mathbf{W}^k)^{-1}}^2 + \mu \sum_{k=1}^N \|\mathbf{x}_k\|_2 \\
\text{subject to} \quad & \mathbf{X}_{ij} \geq 0 \quad \forall i, j \\
& \sum_{i=1}^N \mathbf{X}_{ij} = 1 \quad \forall j
\end{aligned} \tag{20}$$

where $\mathbf{W}^k = \mathbf{W}(\mathbf{X}^k)$. Given \mathbf{W}^k , problem (20) reduces to a weighted version of GLUP (6) due to the weighted norm in the first term. The ADMM solution developed in section II can be adapted to solve the optimization problem (20). Minimizing the augmented Lagrangian with respect to \mathbf{Z} , and

updating the Lagrange multipliers, can be carried out exactly as in Section II. For concision, only the \mathbf{X} -minimization step is described hereafter.

Minimization of $\mathcal{L}_\rho(\mathbf{X}, \mathbf{Z}^k, \mathbf{\Lambda}^k)$ with respect to \mathbf{X} : Omitting the terms that do not depend on \mathbf{X} , the minimization of the augmented Lagrangian $\mathcal{L}_\rho(\mathbf{X}, \mathbf{Z}^k, \mathbf{\Lambda}^k)$ with respect to \mathbf{X} leads to

$$\begin{aligned}
\min_{\mathbf{X}} \quad & \frac{1}{2} \|\mathbf{S} - \mathbf{S}_\omega \mathbf{X}\|_{(\mathbf{W}^k)^{-1}}^2 \\
& + \text{trace}(\mathbf{\Lambda}^\top (\mathbf{A}\mathbf{X})) + \frac{\rho}{2} \|\mathbf{A}\mathbf{X} + \mathbf{B}\mathbf{Z} - \mathbf{C}\|_{\mathbf{F}}^2.
\end{aligned} \tag{21}$$

Problem (21) is quadratic in \mathbf{X} and admits an analytical solution obtained by setting the gradient to zero. This amounts to solving the Sylvester equation [24], which has an analytic solution

$$\begin{aligned}
& \mathbf{S}_\omega^\top \mathbf{S}_\omega \mathbf{X} (\mathbf{W}^k)^{-1} + \rho \mathbf{A}^\top \mathbf{A} \mathbf{X} \\
& = \mathbf{S}_\omega^\top \mathbf{S} (\mathbf{W}^k)^{-1} - \rho \mathbf{A}^\top \left(\mathbf{B}\mathbf{Z}^k - \mathbf{C} + \frac{\mathbf{\Lambda}^k}{\rho} \right).
\end{aligned} \tag{22}$$

Problem (18) is not convex. An alternating optimization algorithm is more likely to converge to local minima with worse accuracy than the convex version. For this reason, we suggest, as a warm start, to initialize NGLUP with GLUP estimate. Algorithm 2 provides the pseudocode for NGLUP. The algorithm contains two main loops. The inner loop aims at finding the solution of problem (20), whereas the outer loop updates the least-square weight matrix.

Algorithm 2 : $\mathbf{X} = \text{NGLUP}(\mathbf{S}, \mathbf{S}_\omega, \rho^\circ, \mu^\circ, \rho, \mu)$

- 1: Precompute \mathbf{A} , \mathbf{B} , and \mathbf{C}
 - 2: Initialize $\mathbf{X} = \text{GLUP}(\mathbf{S}, \mathbf{S}_\omega, \rho^\circ, \mu^\circ)$, $\mathbf{Z} = \mathbf{X}$, $\mathbf{\Lambda} = \mathbf{0}$
 - 3: **while** $\|\mathbf{X} - \mathbf{X}_{\text{old}}\|_2 \geq \epsilon_{\text{tol}}$ **do**
 - 4: $\mathbf{C}(\mathbf{X}) = (\mathbf{I} - \mathbf{I}_\omega \mathbf{X})^\top (\mathbf{I} - \mathbf{I}_\omega \mathbf{X})$
 - 5: $\sigma^2(\mathbf{X}) = \frac{1}{NL} \text{trace}((\mathbf{S} - \mathbf{S}_\omega \mathbf{X}) \mathbf{C}(\mathbf{X})^{-1} (\mathbf{S} - \mathbf{S}_\omega \mathbf{X})^\top)$
 - 6: $\mathbf{W}(\mathbf{X}) = \sigma^2(\mathbf{X}) \mathbf{C}(\mathbf{X})$
 - 7: $\mathbf{X}^{\text{old}} = \mathbf{X}$, $J = 1$
 - 8: **while** $(\|\mathbf{R}\|_2 \geq \epsilon_{\text{pri}} \text{ or } \|\mathbf{P}\|_2 \geq \epsilon_{\text{dual}})$ and $(J \leq J_{\text{max}})$ **do**
 - 9: $\mathbf{X} = \text{solution of Sylvester equation (22)}$
 - 10: $\mathbf{Z}_{\text{old}} = \mathbf{Z}$
 - 11: **for** $i = 1 \dots N'$ **do**
 - 12: $\mathbf{v}_i = ((\mathbf{x}_i)^\top + \rho^{-1} \mathbf{\Lambda}_i)_+$
 - 13: **if** $\|\mathbf{v}_i\|_2 < \rho^{-1} \mu$ **then**
 - 14: $\mathbf{z}_i = \mathbf{0}$
 - 15: **else**
 - 16: $\mathbf{z}_i = \left(1 - \frac{\mu}{\rho \|\mathbf{v}_i\|_2}\right) \mathbf{v}_i$
 - 17: **end if**
 - 18: **end for**
 - 19: $\mathbf{R} = \mathbf{A}\mathbf{X} + \mathbf{B}\mathbf{Z} - \mathbf{C}$
 - 20: $\mathbf{P} = \rho \mathbf{A} \mathbf{B} (\mathbf{Z} - \mathbf{Z}_{\text{old}})$
 - 21: $\mathbf{\Lambda} = \mathbf{\Lambda} + \rho (\mathbf{A}\mathbf{X} + \mathbf{B}\mathbf{Z} - \mathbf{C})$
 - 22: $J = J + 1$
 - 23: **end while**
 - 24: **end while**
-

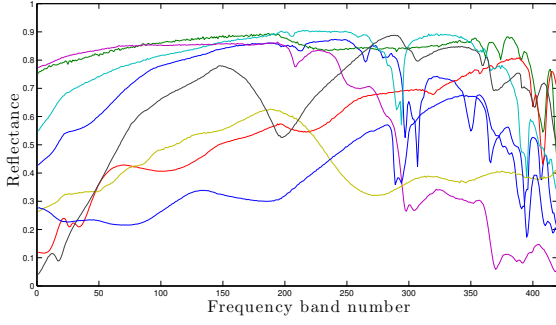


Fig. 1. Reflectance of selected endmembers from the USGS Library.

TABLE I
PROBABILITY OF DETECTING \hat{M} ENDMEMBERS, USING SYNTHETIC DATA
GENERATED WITH $M = 7$ ENDMEMBERS.

M	2	3	6	7	8	9
GLUP (30 dB)	0	0	0	1	0	0
NGLUP (30 dB)	0	0	0	0.98	0.02	0
VD (30 dB)	0.79	0.21	0	0	0	0
GLUP (20 dB)	0	0	0	0.71	0.02	0.27
NGLUP (20 dB)	0	0	0.01	0.96	0.03	0
VD (20 dB)	1	0	0	0	0	0

IV. EXPERIMENTAL RESULTS

A. Synthetic Data

The performances of GLUP and NGLUP were evaluated using synthetic data. We used eight endmembers with 420 spectral samples extracted from the USGS library. Figure 1 shows the reflectance of the endmembers. The spectral mutual coherence between two spectra is defined as $\theta_{ij} = \frac{\langle \mathbf{s}_i, \mathbf{s}_j \rangle}{\|\mathbf{s}_i\| \|\mathbf{s}_j\|}$. The maximum mutual coherence of the eight endmembers was $\theta_{\max} = 0.9940$. The abundances were generated based on a Dirichlet distribution with unit parameter, as a consequence of which the resulting abundances obeyed the non-negativity and sum-to-one constraint, and were uniformly distributed over this simplex. This model is widely used to generate synthetic abundance maps [25]–[27].

First, we used three endmembers to generate an hyperspectral data set containing $N = 100$ (resp. 500) pixels with a SNR of 50 dB. The pure pixels were indexed by integers 1–3 for simplicity, the mixed pixels being indexed by integers 4–100 (resp. 4–500). We ran GLUP algorithm using all the observations ($\mathbf{S}_\omega = \mathbf{S}$) with $\mu = 10$ and $\rho = 100$. The primal and dual tolerances were set to 10^{-5} . The first row of Figure 2 shows the mean of each row \mathbf{x}_k of the estimated abundance matrices $\hat{\mathbf{X}}$. We observe that the first three pixels in Figures 2 (a) and (b) can be identified as the endmembers since the mean values of the first three rows are clearly different from zero. The second row of Figure 2 shows the projection of the data onto the space spanned by the first two PCA axes. Blue stars indicate the data points, and red squares indicate the points that had a non-zero row in $\hat{\mathbf{X}}$, namely, those that were identified

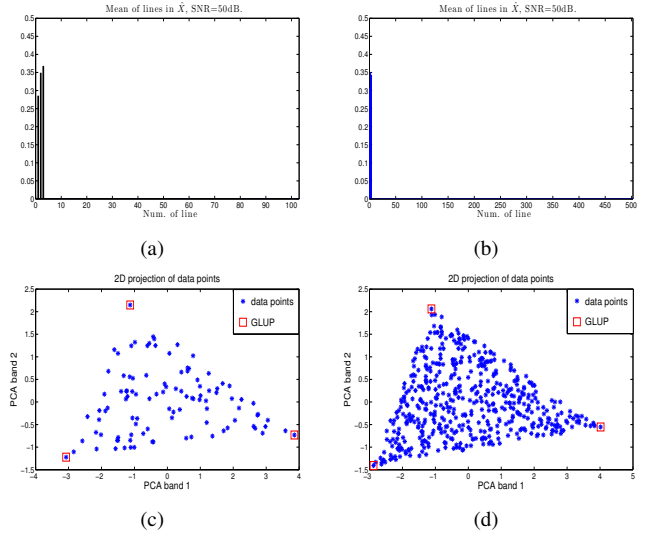


Fig. 2. First row: Mean value of each row of $\hat{\mathbf{X}}$ estimated with GLUP, obtained with (a) $N = 100$ and (b) $N = 500$ pixels with SNR = 50 dB. Second row: 2D data projection and GLUP estimated endmembers obtained with (c) $N = 100$ and (d) $N = 500$ pixels.

as the endmembers. We can see from Figures 2 (c) and (d) that the red squares correspond to the vertices of the simplex enclosing all the data points. GLUP provided the results in 4.73 (resp. 138.59) seconds¹ with a Root Mean Square Error (RMSE), defined as $\frac{1}{N^2} \|\hat{\mathbf{X}} - \mathbf{X}\|_F^2$, equal to 0.0049 (resp., 0.0097) for $N = 100$ (resp., $N = 500$).

We tested NGLUP in less favorable conditions by increasing the number of endmembers and decreasing the SNR. To this end, 7 endmembers were used to generate 93 (resp. 493) mixed pixels. The pure pixels were indexed by integers 1–7. Data points were corrupted with an additive Gaussian noise, corresponding to a SNR of 20 dB. We tested the algorithm for a maximum number of inner iterations $J_{\max} = 1, 10$ and 100. We found that NGLUP converged to the same solution even when the number of inner iterations J was equal to 1. For this reason, only one inner iteration per outer iteration was used for the rest of the experiments. The running time of the algorithm was 77 seconds (resp. 45 min). The first row of Figure 3 shows the mean value of each row \mathbf{x}_k of the abundance matrix $\hat{\mathbf{X}}$ estimated by GLUP for (a) $N = 100$ and (b) $N = 500$ pixels. The second row of Figure 3 shows the mean value of each row \mathbf{x}_k of the abundance matrix $\hat{\mathbf{X}}$ estimated by NGLUP for (a) $N = 100$ and (b) $N = 500$ pixels. The 7 largest mean values correspond to the 7 endmembers. As expected, NGLUP converged to a sparser and more accurate solution than GLUP. We observed that similar results can be obtained with approximately sparse mixtures of the endmembers. This scenario can be simulated by setting the scale parameter of the Dirichlet distribution to some positive value smaller than one.

We repeated the previous simulation with $N = 100$ pixels 100 times. For each realization, we examined the number of mean values of the rows of $\hat{\mathbf{X}}$ that were larger than a predefined threshold equal to 0.01. We considered this value

¹Machine specifications: 2.2 GHz Intel Core i7 processor and 8 GB RAM

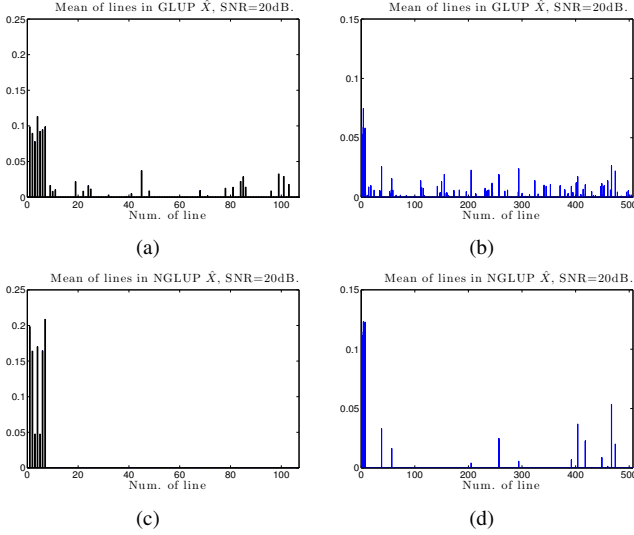


Fig. 3. First row: Mean value of each row of $\hat{\mathbf{X}}$ estimated with GLUP, obtained with (a) $N = 100$ and (b) $N = 500$ pixels with $\text{SNR} = 20$ dB. Second row: Mean value of each row of $\hat{\mathbf{X}}$ estimated with NGLUP, obtained with (a) $N = 100$ and (b) $N = 500$ pixels with $\text{SNR} = 20$ dB.

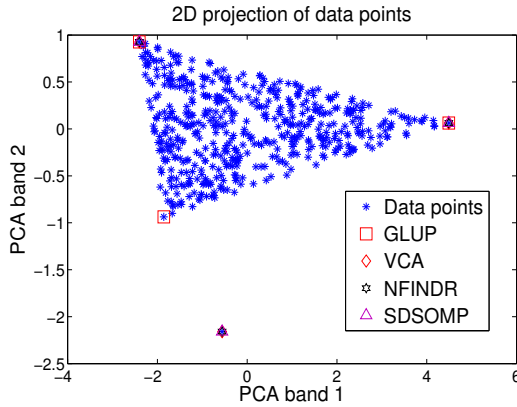


Fig. 4. Endmembers estimated by GLUP, VCA, NFINDR, and SDSOMP when the data contains an outlier.

\hat{M} as the estimated number of endmembers in the scene. Table I provides the probability of detecting \hat{M} endmembers with our two approaches, given synthetic data generated with $M = 7$ endmembers. The same task was performed using Virtual Dimensionality (VD) [2]. We compared the results of NGLUP with those of VD, the probability of false alarm of VD being set to 10^{-3} . Table I shows that NGLUP was able to identify the presence of 7 endmembers in 98% (resp., 96%) of the cases with an SNR of 30 dB (resp. 20 dB). VD only identified 2 endmembers in most cases. Even with higher values of the SNR, VD did not identify the correct number of endmembers. This is due to the fact that VD has asymptotic convergence, and thus requires a very large number of observations in order to converge. This explains the poor performance of VD compared to NGLUP.

Finally, we compared the performance of the proposed approach with 3 endmember extraction algorithms, namely, NFINDR [3], VCA [25], and SDSOMP [12]. In particular, we considered the case where an outlier is present among the

observations. In real data, an outlier usually corresponds to bad sensor measurements. To this end, 3 endmembers were used to generate 500 mixed pixels. The 3 endmembers as well as an additional spectra, the outlier, were inserted among the observations. Figure 4 shows the 2D data projection after performing 2-dimensional PCA. The outlier can be determined by visual inspection as it is the only point outside the simplex. GLUP found 3 endmembers denoted by green stars. We can see that they correspond to the 3 vertices of the simplex, thus to the true endmembers. On the other hand, with VCA, NFINDR and SDSOMP, the number of endmembers to find was explicitly set to 3. The three algorithms detected 2 endmembers over 3, and the outlier instead of the third one. This advantage over geometrical and greedy approaches is related to the formulation of endmember extraction as a penalized optimization problem.

B. Real data: Cuprite

In this section, we shall evaluate the performance of NGLUP using real hyperspectral data. The tests were performed on the so-called images of Cuprite, provided by NASA's sensor AVIRIS. The scene has been captured over a mining district in southern Nevada. The relatively low spatial resolution of the measurements makes this data particularly interesting to test unmixing algorithms. The spatial resolution is about 17 meters. Originally, the images were collected with 224 spectral bands over the wavelength interval 400–2500 nm. After removing the water absorption bands (1–2, 105–115, 150–170, and 223–224), 188 bands were left for the analysis. The image we use for the experiments is a subset of 250×191 pixels, that is, a total of 475000 pixels.

Typically, \mathbf{S} should contain all the available observations, that is, $N = 47750$. In order to alleviate the computational burden, we selected a subset of samples from the original scene. The sampling strategy must guarantee the presence of the endmembers among the selected candidates. To perform this task, we initialized \mathbf{S} with the whole observations. Next, we computed the mutual coherence between all pairs of candidates. First, the pair with the largest mutual coherence was identified, and one of the two spectra was randomly discarded. The process was repeated until 300 spectra were left in \mathbf{S} . As shown by the experimental results provided hereafter, this led to a subset of samples sufficiently representative of the original data to identify the endmembers. Following this strategy, the mutual coherence in the case of Cuprite was reduced from 1 (with $N = 47750$) to 0.9996 (with $N = 300$). Another advantage of this strategy is that reducing the coherence of the dictionary, which is the set of available data in our case, improves the performance of the ℓ_1 penalized algorithm [28], [29]. Other algorithms should be used to perform this task, for example K-means clustering with an angle constraint as in [11]. However, we found this sampling strategy efficient in our experiments.

We applied GLUP and NGLUP (with $\mathbf{S} = \mathbf{S}_\omega$) successively using the subset of pixels, the penalty parameter μ being set to 1 and 10000 respectively. With this setting, we obtained 11 non-zero means in the estimated abundance matrix, that

TABLE II
RECONSTRUCTION QUALITY OF CUPRITE USING NGLUP, N-FINDR,
AND VCA WITH FCLS.

Algorithm	RMSE	avg angle	max angle
NGLUP ($N=300$)	0.0075	1.115°	4.679°
N-FINDR ($N=300$)	0.0115	1.583°	9.286°
N-FINDR ($N=47750$)	0.0095	1.400°	7.626°
VCA ($N=300$)	0.0120	1.289°	7.450°
VCA ($N=47750$)	0.0062	0.855°	8.020°

is, 11 endmembers. We also extracted the endmembers using NFINDR and VCA, the number of endmembers being set to 11. Given that VCA and NFINDR have lower computational complexity, we applied them on the subset of pixels and on the whole image. Figure 5 shows the identified endmembers in each case. It is worth noting that NGLUP was able to identify the endmembers without any prior knowledge on their number. In Figure 6, we compare 6 of the spectra estimated by NGLUP with those estimated by NFINDR when the latter is applied on the whole scene. It can be observed that the 6 spectra correspond to some of the major minerals present in the scene: Sphene, Kaolinite, Muscovite, Alunite, Dumortierite [25]. Figure 7 shows the corresponding abundance maps estimated by FCLS, the endmembers being those estimated by NGLUP. Finally, Table II shows the reconstruction error of the original image. In particular, we report the RMSE, average and maximum spectral angles. Let θ_i be the spectral angle between the i -th original spectrum s_i and its reconstructed version \hat{s}_i defined as $\theta_i = \arccos(\frac{\langle s_i, \hat{s}_i \rangle}{\|s_i\| \|\hat{s}_i\|})$. The maximum and average reconstruction angles are given by $\theta_{\max} = \max_{i=1 \dots N} (\theta_i)$, and $\theta_{\text{avg}} = \frac{1}{N} \sum_{i=1}^N \theta_i$ respectively.

When the three algorithms were applied with the sampled subset, NGLUP always had better scores. When VCA and NFINDR were applied over the whole scene, NGLUP slightly outperformed NFINDR and had comparable performance to VCA.

V. CONCLUSION AND PERSPECTIVES

In this work, we presented two approaches for blind and fully constrained unmixing. Both methods are based on mixing models with increasing complexity, and allow to simultaneously determine the endmembers and estimate their local abundance in the scene. Compared to the first model called GLUP, the second model NGLUP explicitly considers that endmembers present in the scene are corrupted by noise. Experiments on synthetic and real data demonstrated the excellent performance of both approaches. Future work includes their extension to an online framework, which should allow to reduce their complexity and to make them adaptive to changing environmental conditions.

APPENDIX

Proof: Since problem (11) is convex, we simply have to check the validity of the solution in the two cases $\|(v)_+\|_2 > \alpha$ and $\|(v)_+\|_2 < \alpha$. Let $f_0(z) = \frac{1}{2}\|z - v\|_2^2 + \alpha\|z\|_2$. For $\|(v)_+\|_2 > \alpha$, the gradient of f_0 is given by

$$\nabla f_0(z^*) = \left(1 + \frac{\alpha}{\|z^*\|_2}\right) z^* - v. \quad (23)$$

Replacing by the appropriate expression from (11) yields

$$\nabla f_0(z^*) = (v)_+ - v \geq 0 \quad (24)$$

$$z_i^* \cdot \nabla f_0(z^*)_i \propto ((v)_+)_i \cdot ((v)_+ - v)_i = 0. \quad (25)$$

These two conditions correspond the optimality conditions, which means that $z \succeq 0$ is a solution for the constrained problem. For more details, refer to section 4.2.3 in [30].

For the second case, note that for every $z \succeq 0$, we have

$$\sum_i z_i v_i \leq \sum_i z_i (v_i)_+ \leq \|z\|_2 \cdot \|(v)_+\|_2. \quad (26)$$

It follows that

$$\begin{aligned} f_0(z) - f_0(0) &= \frac{1}{2} \sum_i z_i^2 - \sum_i z_i v_i + \alpha\|z\|_2 \\ &\geq \frac{1}{2} \|z\|_2^2 - \|z\|_2 \cdot \|(v)_+\|_2 + \alpha\|z\|_2 \\ &\geq \frac{1}{2} \|z\|_2^2 + \|z\|_2 (\alpha - \|(v)_+\|_2). \end{aligned} \quad (27)$$

This proves that for $\|(v)_+\|_2 \leq \alpha$, the minimum is reached for $z^* = 0$. ■

REFERENCES

- [1] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 44–57, 2002.
- [2] C. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 3, pp. 608–619, March 2004.
- [3] M. E. Winter, "N-FINDR: an algorithm for fast autonomous spectral endmember determination in hyperspectral data," in *Proc. SPIE Imaging Spectrometry*, October 1999.
- [4] D. C. Heinz and C. I. Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 3, pp. 529–545, March 2001.
- [5] O. Eches, N. Dobigeon, C. Mailhes, and J.-Y. Tourneret, "Bayesian estimation of linear mixtures using the normal compositional model," *IEEE Transaction on Image Processing*, vol. 19, no. 6, pp. 1403–1413, June 2010.
- [6] P. Honeine and C. Richard, "Geometric unmixing of large hyperspectral images: A barycentric coordinate approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 6, pp. 2185–2195, June 2012.
- [7] L. Miao, H. Qi, and H. Szu, "A maximum entropy approach to unsupervised mixed-pixel decomposition," *IEEE Transaction on Image Processing*, vol. 16, no. 4, pp. 1008–1021, April 2007.
- [8] Z. Yang, G. Zhou, S. Xie, S. Ding, J.-M. Yang, and J. Zhang, "Blind spectral unmixing based on sparse nonnegative matrix factorization," *IEEE Transaction on Image Processing*, vol. 20, no. 4, pp. 1112–1125, April 2011.
- [9] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B (Statistical methodology)*, vol. 68, no. 1, pp. 49–67, February 2006.
- [10] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

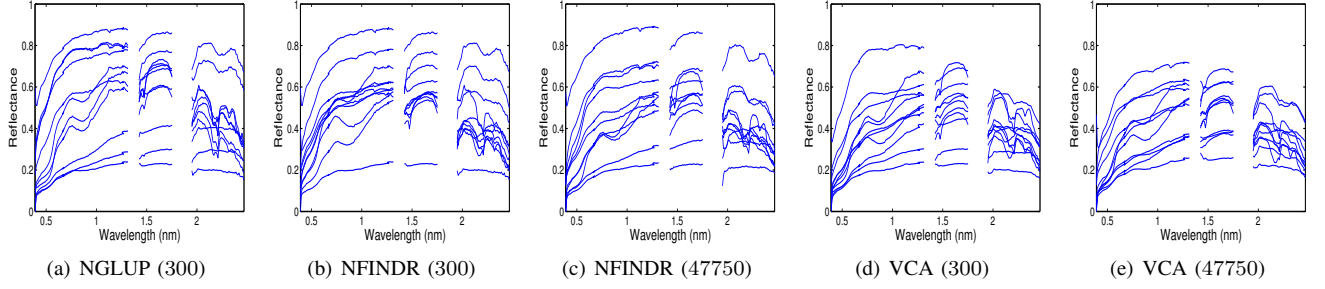


Fig. 5. Estimated endmembers obtained with (a) NGLUP with 300 samples (b) NFINDR with 300 samples (c) NFINDR with all samples (d) VCA with 300 samples (e) VCA with all samples.

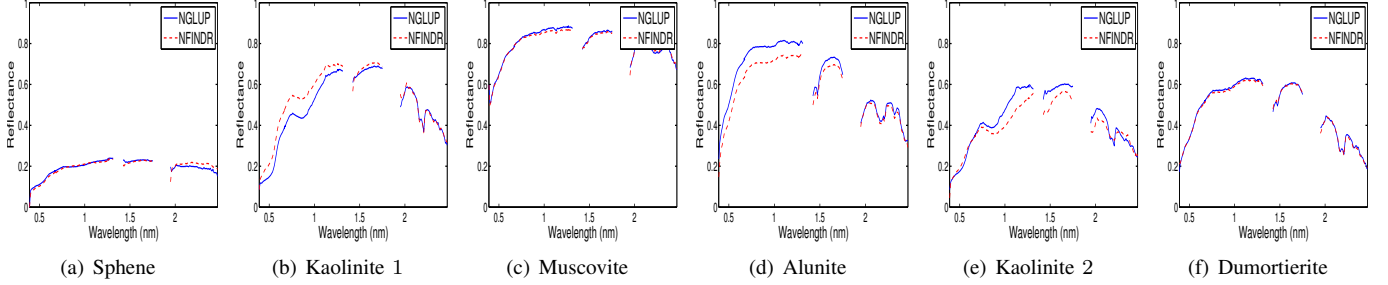


Fig. 6. Comparison between six endmembers' spectra estimated by NGLUP and by NFINDR when applied on the whole AVIRIS scene of Cuprite.

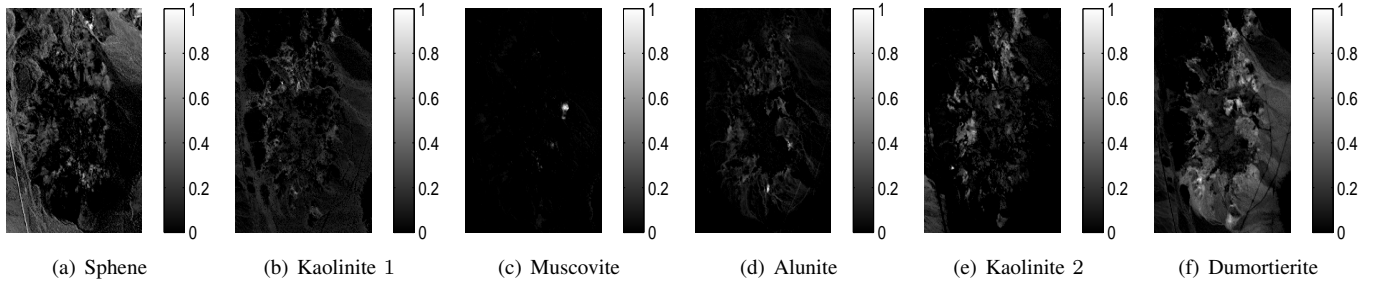


Fig. 7. Abundance maps estimated by FCLS for some of NGLUP endmembers: Sphene, Kaolinite, Muscovite, Alunite, Kaolinite 2, and Dumortierite.

- [11] E. Esser, M. Moller, S. Osher, G. Sapiro, and J. Xin, "A convex model for nonnegative matrix factorization and dimensionality reduction on physical space," *IEEE Transaction on Image Processing*, vol. 21, no. 7, pp. 3239–3252, July 2012.
- [12] X. Fu, W. Ma, T. Chan, J. Bioucas-Dias, and M.-D. Iordache, "Greedy algorithms for pure pixels identification in hyperspectral unmixing: A multiple-measurement vector viewpoint," in *Proc. EUSIPCO*, 2013.
- [13] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Collaborative sparse regression for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 1, pp. 341–354, February 2013.
- [14] M.-D. Iordache, A. Okujeni, S. van der Linden, J. Bioucas-Dias, A. Plaza, and B. Somers, "A multi-measurement vector approach for endmember extraction in urban environments," in *Image Information Mining Conference: The Sentinels Era*, Bucharest, Romania, 2014.
- [15] Z. Zhang, S. G. Mallat, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, december 1993.
- [16] J. Eckstein and D. P. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Mathematical Programming*, vol. 55, no. 1, pp. 293–318, 1992.
- [17] P. L. Combettes and J. C. Pesquet, "Proximal splitting methods in signal processing," in *arXiv:0912.3522*, 2009.
- [18] A. T. Puig, A. Wiesel, G. Fleury, and A. O. Hero, "Multidimensional shrinkage-thresholding operator and group lasso penalties," *IEEE Signal Processing Letters*, vol. 18, no. 6, pp. 363–366, April 2011.
- [19] E. Thiebaut, F. Soulez, and L. Denis, "Exploiting spatial sparsity for multiwavelength imaging in optical interferometry," *Journal of the Optical Society of America A*, vol. 30, no. 2, pp. 160–170, February 2013.
- [20] P. M. Hooper, "Iterative weighted least squares estimation in heteroscedastic linear models," *Journal of the American Statistical Association*, vol. 88, no. 421, pp. 179–184, 1993.
- [21] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, "Iteratively reweighted least squares minimization for sparse recovery," *Communications on Pure and Applied Mathematics*, vol. 63, no. 1, pp. 1–38, 2010.
- [22] W. A. Fuller and J. N. K. Rao, "Estimation for a linear regression model with unknown diagonal covariance matrix," *The Annals of Statistics*, vol. 6, no. 5, pp. 1149–1158, september 1978.
- [23] R. J. Carroll and D. B. H. Cline, "An asymptotic theory for weighted least-squares with weights estimated by replication," *Biometrika*, vol. 75, no. 1, pp. 35–43, March 1988.
- [24] R. H. Bartels and G. W. Stewart, "Solution of the matrix equation $AX + XB = C$," *Communications of the ACM*, vol. 15, no. 9, pp. 820–826, 1972.
- [25] J. M. P. Nascimento and J. Bioucas-Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898–910, 2005.
- [26] Y. Altmann, N. Dobigeon, S. McLaughlin, and J.-Y. Tournet, "Nonlinear spectral unmixing of hyperspectral images using gaussian processes," *IEEE Transactions on Signal Processing*, vol. 61, no. 10, pp. 242–245, 2013.
- [27] M.-D. Iordache, J. Bioucas-Dias, A. Plaza, and B. Somers, "MUSIC-CSR: Hyperspectral unmixing via multiple signal classification and

collaborative sparse regression,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, pp. 4364–4382, 2014.

- [28] A. Bruckstein, D. Donoho, and M. Elad, “From sparse solutions of systems of equations to sparse modeling of signals and images,” *SIAM Review*, vol. 51, pp. 34–81, 2009.
- [29] E. Candes, J. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements,” *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207, 2006.
- [30] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2008.



Rita Ammanouil was born in Beirut, Lebanon in January 1991. She received the bachelor in Computer and Communication Engineering from the Holly Spirit University of Kaslik, Lebanon, in 2013. Since October 2013 she is working towards the Ph.D. degree at the Lagrange Laboratory (University of Nice Sophia Antipolis, CNRS, Observatoire de la Côte d’Azur). Her research activity is focused on Hyperspectral image unmixing.



André Ferrari (SM’91-M’93) received the Ingénieur degree from École Centrale de Lyon, Lyon, France, in 1988 and the M.Sc. and Ph.D. degrees from the University of Nice Sophia Antipolis (UNS), France, in 1989 and 1992, respectively, all in electrical and computer engineering. He is currently a Professor at UNS.

He is currently a Professor at UNS. He is a member of the Joseph-Louis Lagrange Laboratory (CNRS, OCA), where his research activity is centered around statistical signal processing and modeling, with a particular interest in applications to astrophysics.



Cédric Richard (S’98–M’01–SM’07) was born January 24, 1970 in Sarrebourg, France. He received the Dipl.-Ing. and the M.S. degrees in 1994 and the Ph.D. degree in 1998 from the University of Technology of Compiègne, France, all in electrical and computer engineering. From 1999 to 2003, he was an Associate Professor at the University of Technology of Troyes (UTT), France. From 2003 to 2009, he was a Full Professor at UTT. Since September 2009, he is a Full Professor in the Lagrange Laboratory (University of Nice Sophia Antipolis, CNRS, Observatoire de la Côte d’Azur). In Winter 2009 and 2014, and Fall 2010, 2011 and 2013, he was a Visiting Researcher with the Department of Electrical Engineering, Federal University of Santa Catarina (UFSC), Florianopolis, Brazil. He is a junior member of the Institut Universitaire de France since October 2010. His current research interests include statistical signal processing and machine learning.

Cédric Richard is the author of over 220 papers. He was the General Chair of the XXIth Francophone conference GRETSI on Signal and Image Processing that was held in Troyes, France, in 2007, and of the IEEE Statistical Signal Processing Workshop (IEEE SSP’11) that was held in Nice, France, in 2011. He will be the Technical Chair of EUSIPCO 2015. Since 2005, he is a member of GRETSI association board and of the EURASIP society, and Senior Member of the IEEE. In 2006–2010, he served as an associate editor of the IEEE Transactions on Signal Processing. Currently, he serves as an Associate Editor of Signal Processing Elsevier, and of the IEEE Signal Processing Letters. He is an Eurasp liaison local officer. He is a member of the Signal Processing Theory and Methods (SPTM TC) Technical Committee, and of the Machine Learning for Signal Processing (MLSP TC) Technical Committee, of the IEEE Signal Processing Society.

Paul Honeine and Cédric Richard won the Best Paper Award for “Solving the preimage problem in kernel machines: a direct method” at the 2009 IEEE International Workshop on Machine Learning for Signal Processing.



David Mary David Mary received in 2003 a Ph.D. degree in Signal Processing from the Ecole Nationale Supérieure des Télécommunications, Paris, France. In 2004, he joined the Aryabhatta Research Institute of Observational Sciences, Nainital, India. In 2006, he joined the Astronomisches Rechen Institut, Heidelberg, Germany. In 2007, he joined the Lagrange Laboratory (University of Nice and Observatory of the Côte d’Azur) Nice, France, where he now serves as a Professor. D. Mary’s research interests include statistical estimation, detection and

learning, and their applications to Astronomy.