



Heriot-Watt University  
Research Gateway

## Fast Hyperspectral Unmixing in Presence of Nonlinearity or Mismodelling Effects

### Citation for published version:

Halimi, A, Bioucas-Dias, JM, Dobigeon, N, Buller, GS & McLaughlin, S 2017, 'Fast Hyperspectral Unmixing in Presence of Nonlinearity or Mismodelling Effects', *IEEE Transactions on Computational Imaging*, vol. 3, no. 2, pp. 146-159. <https://doi.org/10.1109/TCI.2016.2631979>

### Digital Object Identifier (DOI):

[10.1109/TCI.2016.2631979](https://doi.org/10.1109/TCI.2016.2631979)

### Link:

[Link to publication record in Heriot-Watt Research Portal](#)

### Document Version:

Publisher's PDF, also known as Version of record

### Published In:

IEEE Transactions on Computational Imaging

### Publisher Rights Statement:

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see <http://creativecommons.org/licenses/by/3.0/>

### General rights

Copyright for the publications made accessible via Heriot-Watt Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

Heriot-Watt University has made every reasonable effort to ensure that the content in Heriot-Watt Research Portal complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [open.access@hw.ac.uk](mailto:open.access@hw.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

# Fast Hyperspectral Unmixing in Presence of Nonlinearity or Mismodeling Effects

Abderrahim Halimi, *Member, IEEE*, José M. Bioucas-Dias, *Senior Member, IEEE*,  
Nicolas Dobigeon, *Senior Member, IEEE*, Gerald S. Buller, and Stephen McLaughlin, *Fellow, IEEE*

**Abstract**—This paper presents two novel hyperspectral mixture models and associated unmixing algorithms. The two models assume a linear mixing model corrupted by an additive term whose expression can be adapted to account for multiple scattering nonlinearities (NL), or mismodeling effects (ME). The NL model generalizes bilinear models by taking into account higher order interaction terms. The ME model accounts for different effects, such as endmember variability or the presence of outliers. The abundance and residual parameters of these models are estimated by considering a convex formulation suitable for fast estimation algorithms. This formulation accounts for constraints, such as the sum-to-one and nonnegativity of the abundances, the nonnegativity of the nonlinearity coefficients, the spectral smoothness of the ME terms and the spatial sparseness of the residuals. The resulting convex problem is solved using the alternating direction method of multipliers whose convergence is ensured theoretically. The proposed mixture models and their unmixing algorithms are validated on both synthetic and real images showing competitive results regarding the quality of the inference and the computational complexity when compared to the state-of-the-art algorithms.

**Index Terms**—Hyperspectral imagery, collaborative sparse regression, ADMM, nonlinear unmixing, robust unmixing, convex optimization.

## I. INTRODUCTION

**H**YPERSPECTRAL imaging is a remote sensing technology that collects three dimensional data cubes composed of 2D spatial images acquired in numerous contiguous spectral bands. Assuming that each pixel spectrum is a mixture of several pure materials (endmembers), spectral unmixing consists of recovering the spectral signatures (endmembers) of the materials

Manuscript received July 8, 2016; revised October 30, 2016; accepted November 11, 2015. Date of publication November 22, 2016; date of current version May 8, 2017. This work was supported in part by the EPSRC Grant EP/J015180/1, Grant EP/N003446/1, and Grant EP/K015338/1, the MapInvPlnt ERA-NET MED Project no. ANR-15-NMED-0002-02, the HYPANEMA ANR Project under Grant ANR-12-BS03-003, and the Portuguese Fundação para a Ciência e Tecnologia (FCT) under grants UID/EEA/5008/2013 and ERANETMED/0001/2014. The guest editor coordinating the review of this manuscript and approving it for publication was Dr. Jocelyn Chanussot.

A. Halimi, G. S. Buller, and S. McLaughlin are with the School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh EH14 4AS, U.K. (e-mail: a.halimi@hw.ac.uk; g.s.buller@hw.ac.uk; s.mclaughlin@hw.ac.uk).

J. M. Bioucas-Dias is with the Instituto de Telecomunicações and Instituto Superior Técnico, Universidade de Lisboa, Lisboa 1049-001, Portugal (e-mail: bioucas@lx.it.pt).

N. Dobigeon is with the University of Toulouse, Toulouse 31071, France (e-mail: Nicolas.Dobigeon@enseiht.fr).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TCI.2016.2631979

present in the scene, and quantifying their proportions within each hyperspectral image pixel [1]. More precisely, unmixing hyperspectral images consists of three stages: (i) determining the number of endmembers and possibly projecting the data onto a subspace of reduced dimension [2]–[4], (ii) identifying the endmembers using an endmember extraction algorithm (EEA) such as vertex component analysis (VCA) [5], and N-FINDR [6] and (iii) estimating their abundances [7]–[9]. Akin to [7], [9]–[11], this paper considers a supervised unmixing scenario which aims at estimating the abundances while assuming that the two first unmixing steps have been successfully implemented.

As a result of its simplicity, the linear mixing model (LMM) is used by many of the hyperspectral unmixing algorithms presented in the literature [7], [9]. This is generally justified when considering flat scenes without component interactions, and a fixed endmember spectra for all the pixels. However, an inherent limitation of the LMM occurs in presence of volumetric scattering, terrain relief, or intimate mixtures of materials which require the definition of new sophisticated models, to take these effects into account. Nonlinear mixture models are an alternative to better account for those effects [12], [13] and we distinguish between two main families: the first is signal processing based and seeks to construct flexible models that can represent a wide range of nonlinearities. The second is physical based models that include the intimate mixture models [14] and those accounting for bilinear interactions [15]–[20]. This paper considers a physical based nonlinearity which generalizes the bilinear formulation in [10], [21] to account for multiple scattering effects. A second inherent limitation of the LMM appears when the endmember spectra vary spectrally and spatially causing what is known as endmember variability (EV) [22]–[27]. In this case, and under a supervised SU scenario, the endmember fluctuation can not be captured by traditional EEA algorithms which affect the LMM by the presence of an additional spectrally smooth residual component [21]. A third LMM limitation is related to the presence of sparse outliers, e.g. due to the presence of impulse noise, horizontal or vertical line stripes, dead lines, and others types of noise [28], [29]. The latter two LMM limitations can be solved separately by considering specialized algorithms that deal with EV [30]–[32] or outliers [28], [29]. In this paper, we adopt the same strategy as in [21], [33] and propose a robust algorithm that encompasses the first two effects described above.

The first contribution of this paper is the introduction of two mixture models to deal with NL and ME. The models

proposed are based on the residual component principle [34] and are closely related to the RCA-NL and RCA-ME models introduced in [21]. More precisely, the proposed NL generalizes RCA-NL by accounting for multiple scattering effects. Indeed, the residual term is assumed to be a linear combination of high order interaction spectra. Due to the high number of interactions, the non-negative nonlinearity coefficients are assumed sparse so that only a few interactions are active for each pixel. The resulting formulation is then general, and covers many NL models [15]–[20], [35]. In a similar fashion to RCA-ME, the proposed ME model assumes a spectrally smooth residual term. However, in contrast with RCA-ME that adopts a statistical approach to account for this smooth property, the proposed model assumes the residual term to be a sparse linear combination over a dictionary (such as the discrete cosine transform (DCT), or the spline decomposition). For both models, the corrupted pixels are assumed spatially sparse meaning that only a small number of nonlinear or outlier pixels are present, as previously suggested in [35], [36] for NL and in [28], [29] for outliers. This effect has been introduced by considering the well known collaborative sparse regression strategy [28], [36]–[39] since it promotes group-sparsity over the residual terms while using the information of the residuals in all the pixels. Note that the first motivation for these new reformulations is that both models assume a residual term that is written as a linear combination of sparse coefficients, which is suitable for the development of a joint formulation to achieve the unmixing strategy. The second motivation is related to the unmixing problem that is significantly simplified by considering separable variables (between the abundances and the residual coefficients) as well as a linear expression for both the LMM term and the residual term.

The second contribution of this paper is the introduction of a convex formulation for unmixing the proposed observation models. The convexity is obtained thanks to the linearity of the observation models with respect to the unknown parameters, as well as the considered regularization terms. Indeed, the formulation accounts for the known physical constraints on the estimated parameters such as the sum-to-one and non-negativity of the abundances, the non-negativity of the nonlinearity coefficients, the spectral smoothness of the ME terms and the spatial sparseness of the residuals. The resulting convex problem is solved using the alternating direction method of multipliers (ADMM) whose convergence is theoretically ensured. More precisely, we propose two algorithms denoted as NUSAL- $K$  for nonlinear unmixing by variable splitting and augmented Lagrangian with order  $K$ , and RUSAL for robust unmixing by variable splitting and augmented Lagrangian. Note that the ADMM algorithms are well adapted for large scale problems, i.e., with a large number of parameters to be estimated [40], [41]. Moreover, this method offers good performance at a reduced computational cost as already shown in many hyperspectral unmixing works [9], [38], [39]. The proposed mixture models and estimation algorithms are validated using synthetic and real hyperspectral images. The results obtained are very promising and show the potential of the proposed mixture models and associated inference algorithms with respect to the estimation quality and the computational cost.

The paper is structured as follows. Section II presents the proposed NL and ME mixture models considered in this study. Section III introduces the convex unmixing formulations and the ADMM-based optimization algorithms associated with the two mixture models. Section IV analyzes the performance of the proposed algorithms when applied to synthetic images with known ground truth. Results on real hyperspectral images are presented in Section V and conclusions and future work are reported in Section VI.

## II. MIXTURE MODELS

As a result of its simplicity, the LMM is widely used in hyperspectral images. However, the LMM has some limitations in the presence of nonlinearity or outlier effects. This paper deals with these issues by considering the observation model proposed in [21], itself inspired from the residual component analysis model described in [34]. This model introduces a general formulation that is expressed as the sum of a linear model and a residual term that accounts for the remaining effects. The general observation model for the  $(L \times 1)$  pixel spectrum  $\mathbf{y}_n$ , where  $L$  is the number of spectral bands, is given by

$$\begin{aligned} \mathbf{y}_n &= \sum_{r=1}^R a_{r,n} \mathbf{m}_r + \phi_n(\mathbf{M}, \mathbf{a}_n, \mathbf{x}_n) + \mathbf{e}_n \\ &= \mathbf{M} \mathbf{a}_n + \phi_n(\mathbf{M}, \mathbf{a}_n, \mathbf{x}_n) + \mathbf{e}_n, \end{aligned} \quad (1)$$

where  $\mathbf{a}_n = (a_{1,n}, \dots, a_{R,n})^T$  is an  $(R \times 1)$  vector of abundances associated with the  $n$ th pixel,  $\mathbf{x}_n = (x_{1,n}, \dots, x_{D,n})^T$  is a  $(D \times 1)$  vector of residual coefficients associated with the  $n$ th pixel,  $R$  (resp.  $D$ ) is the number of endmembers (resp. residual coefficients),  $\mathbf{e}_n \sim \mathcal{N}(\mathbf{0}, \Sigma)$  is a centered Gaussian noise and  $\phi_n$  is a residual term that might depend on the endmembers, the abundances or residual coefficients to account for the additional mismodeling effect. In model (1), the endmembers matrix  $\mathbf{M}$  is fixed (extracted using an EEA) and endmember variability can be accounted for by the pixel dependent residual term  $\phi_n$ . Moreover, the paper deals with the supervised case in which we assume the endmembers to be known and we only estimate the abundances and the residual terms. Due to physical constraints, the abundance vector  $\mathbf{a}_n$  satisfies the following abundance non-negativity (ANC) and abundance sum-to-one (ASC) constraints

$$a_{r,n} \geq 0, \forall r \in \{1, \dots, R\} \quad \text{and} \quad \sum_{r=1}^R a_{r,n} = 1. \quad (2)$$

The main motivation of this paper is to deal with some limitations of the RCA models and CDA algorithms introduced in [21], i.e., (i) the non-convexity of the resulting NL and ME problems, (ii) the consideration of only second order NL interaction terms and (iii) the convergence limitation of the CDA algorithms (convergence to a local minimum). Eq. (1) shows a general model that can be adapted to account for different physical phenomena. The next sections present in details the considered model variants that will account for NL and ME.

### A. Effects of Nonlinearity (NL)

Nonlinear mixing models provide a powerful tool to deal with the inherent limitations of the LMM. Many nonlinear models have been introduced in the literature and we can divide them into two categories: physical based models (including bilinear and intimate mixture models) and signal processing models (such as the PPNMM [15], [42]). This paper considers a physical based model to deal with the multiple scattering effects. More precisely, the model considered accounts for higher order interactions between the endmembers and reduces to [10], [21] when only the bilinear second order interactions are considered. Note that bilinear models assume that the effect of the interaction terms decreases as the order increases, as suggested in [16], [18], [19]. However, in this paper, we include higher order interaction terms in the proposed model/algorithm to highlight their benefit as recently shown in [43]–[45]. The proposed NL model considering the  $K$ th order of interactions is given by

$$\mathbf{y}_n = \mathbf{M}\mathbf{a}_n + \phi_n^{\text{NL-}K}(\mathbf{M}, \boldsymbol{\gamma}_n) + e_n \quad (3)$$

where the residual component is

$$\phi_n^{\text{NL-}K}(\mathbf{M}, \boldsymbol{\gamma}_n) = \mathbf{Q}^{(K)}(\mathbf{M})\boldsymbol{\gamma}_n, \quad (4)$$

with  $\boldsymbol{\gamma}_n = (\gamma_n^{(1)}, \dots, \gamma_n^{(D_K)})^T$ ,  $\forall n$  is the  $(D_K \times 1)$  vector of non-negative coefficients (i.e.,  $\gamma_n^{(d)} \geq 0, \forall n, d$ ),  $\mathbf{Q}^{(K)}$  is the  $(L \times D_K)$  matrix gathering the interaction spectra of the form  $\mathbf{m}_i \odot \mathbf{m}_j \odot \dots \odot \mathbf{m}_l$ ,  $\odot$  denotes the Hadamard (term-wise) product, and  $D_K$  is the number of coefficients associated with the interaction terms that have an order lower or equal to  $K$ . More details regarding the construction of  $\mathbf{Q}^{(K)}$  are provided in Appendix A. For instance, considering only second order interaction terms (i.e.,  $K = 2$ ) leads to  $D_2 = \frac{R(R+1)}{2}$ ,  $\boldsymbol{\gamma}_n(2) = (\gamma_n^{(1,2)}, \dots, \gamma_n^{(R-1,R)}, \gamma_n^{(1,1)}, \dots, \gamma_n^{(R,R)})^T$ ,  $\forall n$ ,  $\mathbf{Q}^{(2)}(\mathbf{M}) = (\sqrt{2}\mathbf{m}_{12}, \dots, \sqrt{2}\mathbf{m}_{R-1,R}, \mathbf{m}_{11}, \dots, \mathbf{m}_{RR})$ , and a residual term similar to [10] as follows

$$\begin{aligned} \phi_n^{\text{NL-2}}(\mathbf{M}, \boldsymbol{\gamma}_n) &= \mathbf{Q}^{(2)}(\mathbf{M})\boldsymbol{\gamma}_n = \sum_{r=1}^R \gamma_n^{(r,r)} \mathbf{m}_{r,r} \\ &+ \sum_{r=1}^{R-1} \sum_{r'=r+1}^R \gamma_n^{(r,r')} \sqrt{2}\mathbf{m}_{r,r'} \end{aligned} \quad (5)$$

where  $\mathbf{m}_{i,j} = \mathbf{m}_i \odot \mathbf{m}_j$ , and the interaction terms are weighted by the coefficient  $\sqrt{2}$  obtained by comparison with a homogeneous polynomial kernel of the 2nd degree (see Appendix A for more details regarding these coefficients). In what follows, and for brevity, we drop the order index ( $K$ ) for general statements (related to all interaction orders) and only include it when dealing with specific orders. The model proposed in (3) reduces to the LMM for  $\boldsymbol{\gamma}_n = \mathbf{0}$ ,  $\forall n$  and has many links to state-of-the-art models. Indeed, model (3) with  $K = 2$  is similar to [10] and has a close relation to the RCA model [35] (as shown in [10]). Moreover, it generalizes the GBM model [16], [17] by accounting for self-interaction between the endmembers, and also generalizes the PPNMM [15] by considering different weights for the bilinear terms. Overall, model (3) is of a similar polynomial form as

the bilinear models (RCA [21], GBM [16], PPNMM [15], Nascimento [18], Fan [19], and Meganem [20] models) with the main difference due to the introduction of higher order interaction terms, and the non-negativity and sum-to-one constraints associated with each model. In contrast with the model described in [43], which accounts for all the interactions by using only one parameter, the model (3) includes a different coefficient for each interaction term, which enables analysis of the interaction between any specific physical components (i.e., availability of interaction maps).

Note that the nonlinear behavior generally affects some pixels of the image as already exploited in [35], [36], which suggest a spatial sparsity of the nonlinear pixels. Moreover, it makes sense to assume that the elements of the nonlinear vector  $\boldsymbol{\gamma}_n$  will not be active at the same time, meaning that the vector is sparse. This can be explained since the lowest order of interactions have often a higher effect [16], [18], [19] and all the interactions between endmembers are not likely to be active at the same time. These sparsity properties are of great importance and will be exploited when designing the unmixing algorithm associated with model (3) in Section III.

### B. Mismodeling Effects (ME) or Outliers

In recent years, there has been considerable interest in robust hyperspectral unmixing to enable adaptation of the simple LMM to realistic scenes which often present outliers or other unknown effects [46]. This goal can be achieved using different strategies such as adapting the optimization cost function [47] or changing the observation model by introducing a residual term that accounts for the mismodeling effects [21], [33], [36]. The latter strategy is adopted in this paper by considering spectrally smooth residuals as for the ME model introduced in [21]. More precisely, the model is

$$\mathbf{y}_n = \mathbf{M}\mathbf{a}_n + \phi_n^{\text{ME}}(\mathbf{b}_n) + e_n, \quad (6)$$

where the residual component is

$$\phi_n^{\text{ME}}(\mathbf{b}_n) = \mathbf{F}^\top \mathbf{b}_n, \quad (7)$$

with  $\mathbf{F}$  is a  $D \times L$  matrix gathering the first  $D$  rows of the DCT,  $\mathbf{b}_n$  is a vector of DCT coefficients and  $\phi_n^{\text{ME}}$  is a smooth spectral function. In this paper, the smooth property of  $\phi_n^{\text{ME}}$  is obtained by imposing sparsity on the elements of each vector  $\mathbf{b}_n$ ,  $\forall n$ . Model (6) reduces to the LMM for  $\mathbf{b}_n = \mathbf{0}_L$ ,  $\forall n$ . Moreover, the residual terms  $\{\phi_1^{\text{ME}}, \dots, \phi_N^{\text{ME}}\}$  are assumed to be spatially sparse to approximate sparse nonlinear effects, end-member variability effects or other mismodeling effects such as outliers. In the following, we highlight the link between model (6) and each of these phenomena. Consider first the NL model (3) with  $\gamma_n = \gamma_n^{(d)} = \gamma_n^{(d')}, \forall d \neq d'$ . In this special case, the nonlinear term reduces to  $\phi_n^{\text{NL}}(\mathbf{M}) = \gamma_n \sum_{d=1}^D \mathbf{q}_d$ , where  $\mathbf{q}_d$  represents the  $d$ th column of  $\mathbf{Q}$ . Thus, as a result of the smooth spectral property of the interaction spectra  $\mathbf{q}_d$ , the nonlinear term  $\phi_n^{\text{NL}}$  can be approximated by the term  $\phi_n^{\text{ME}}$ . This means that model (6) links to the NL model (3) for this special case. Second, it has been shown in [21], [48] that the EV effect can be approximated by assuming pixel dependent endmembers

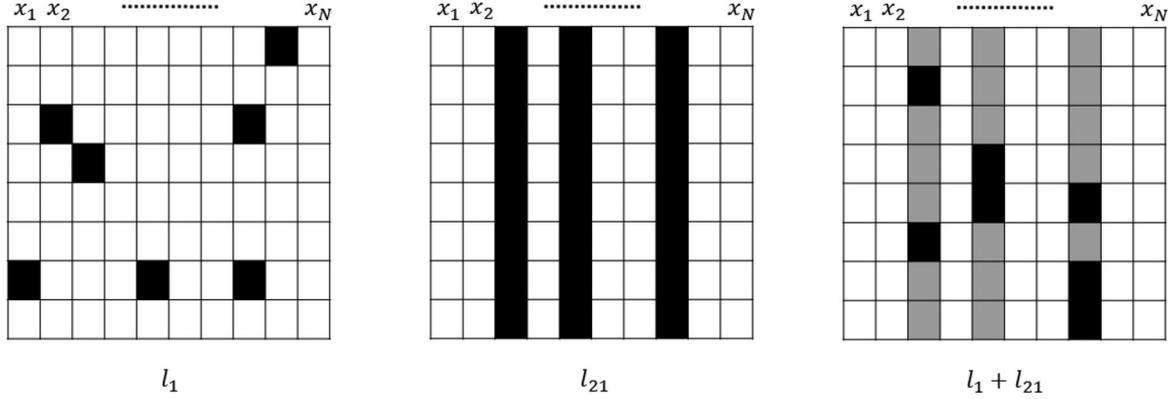


Fig. 1. Graphical illustration of the performance of the regularization terms. Active members of the matrix  $\mathbf{X}$  are represented in black, and non-active members are shown in white. The active blocks when considering the  $\ell_1 + \ell_{21}$  norms are shown in gray. (left)  $\ell_1$  norm, (middle)  $\ell_{21}$  norm, and (right) the considered  $\ell_1 + \ell_{21}$  norms.

$\mathbf{s}_{r,n} = \mathbf{m}_{r,n} + \mathbf{k}_{r,n}$ , where  $\mathbf{k}_{r,n}$  is a spectral vector (assumed smooth in [21]). This model reduces to the proposed model (6) when the same variability affects the different endmembers (i.e.,  $\mathbf{k}_{r,n} = \mathbf{k}_{r',n}, \forall r \neq r'$ ). This provides an intuitive interpretation of the ME model in term of EV. Third, spatially sparse outliers can be present in hyperspectral images as shown in [28], [29], [36], and can also be approximated using  $\phi_n^{\text{ME}}$ . The latter differs from these works by accounting for the smooth spectral behavior of the residuals as in [21]. This illustrates how the model described by (6) can be used to analyze hyperspectral images with a combination of different effects such as NL, EV and/or outliers. Note finally that the ME model estimates smooth residual spectra without any additional information regarding their origins which can be due to physical effects (such as NL, EV), data corruptions or other outliers. The next section introduces the proposed estimation algorithms associated with these NL and ME models.

### III. PROPOSED UNMIXING ALGORITHMS: NUSAL- $K$ , AND RUSAL

This section introduces the unmixing algorithms used to estimate the abundances and the residual coefficients of the proposed models. To this end, we adopt an optimization approach that minimizes a regularized data fidelity cost function. More precisely, considering an independent and identically distributed (i.i.d.) Gaussian noise ( $\Sigma$  proportional to the identity matrix) in model (1) leads to the following negative log-likelihood (referred to as data fidelity term in what follows, and defined up to a multiplicative constant which is the noise variance)

$$\mathcal{L}_P(\mathbf{Z}) = \frac{1}{2} \|\mathbf{Y} - [\mathbf{M}, \mathbf{P}]\mathbf{Z}\|_F^2 \quad (8)$$

where  $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_N]$ ,  $N$  is the total number of pixels,  $\mathbf{Z} = [\mathbf{A}^\top, \mathbf{X}^\top]^\top$  is the  $(R+D) \times N$  matrix gathering the  $(R \times N)$  abundance matrix  $\mathbf{A}$  and the  $(D \times N)$  residual coefficients  $\mathbf{X}$  and  $\|\mathbf{Y}\|_F = \sqrt{\text{trace}(\mathbf{Y}\mathbf{Y}^\top)}$  denotes the Frobenius norm. Note that  $\mathbf{P} = \mathbf{Q}$  and  $\mathbf{x} = \gamma$  (resp.  $\mathbf{P} = \mathbf{F}^\top$  and  $\mathbf{x} = \mathbf{b}$ ) when considering the NL model (resp. the ME model).

Estimating the abundances and the residual coefficients is an ill-posed inverse problem that requires the introduction of prior knowledge (or regularization terms) about these parameters of interest. Therefore, we propose to solve the following regularized optimization problem

$$\begin{aligned} \mathcal{C}(\mathbf{Z}) = & \mathcal{L}_P(\mathbf{Z}) + i_{\mathbb{R}_+}(\mathbf{A}) + i_{\{\mathbf{1}_{(1,R)}\}}(\mathbf{1}_{(1,R)}\mathbf{A}) \\ & + \tau_1 \|\mathbf{X}\|_1 + \tau_2 \|\mathbf{X}\|_{2,1} + \psi(\mathbf{X}) \end{aligned} \quad (9)$$

where  $i_{\mathbb{R}_+}(\mathbf{A}) = \sum_{n=1}^N i_{\mathbb{R}_+}(\mathbf{a}_n)$  is the indicator function that imposes the ANC ( $i_{\mathbb{R}_+}(\mathbf{a}_n) = 0$  if  $\mathbf{a}_n$  belongs to the non-negative orthant and  $+\infty$  otherwise),  $i_{\{\mathbf{1}_{(1,R)}\}}(\mathbf{1}_{(1,R)}\mathbf{A}) = \sum_{n=1}^N i_{\{1\}}(\mathbf{1}_{(1,R)}\mathbf{a}_n)$  is the indicator function that imposes the ASC to each abundance vector  $\mathbf{a}_n$ ,  $\mathbf{1}_{(i,j)}$  denotes the  $i \times j$  vector of 1s,  $\psi(\mathbf{X}) = i_{\mathbb{R}_+}(\mathbf{X})$  when considering the NL model and  $\psi(\mathbf{X}) = 0$  for the ME model. The first line of (9) is a sum of the quadratic data fidelity term associated with the Gaussian noise statistics and two convex terms imposing the abundance constraints. The second line of (9) accounts for the sparsity behavior of the residual coefficients. The first convex term  $\|\mathbf{X}\|_1 = \sum_{n=1}^N \|\mathbf{x}_n\|_1$  is an  $\ell_1$  norm that promotes element-wise sparsity on the  $D \times N$  matrix  $\mathbf{X}$ . This behavior is illustrated in Fig. 1 (left) which shows a point-wise repartition of the active elements of  $\mathbf{X}$ . The second convex term  $\|\mathbf{X}\|_{2,1} = \sum_{n=1}^N \|\mathbf{x}_n\|_2 = \sum_{n=1}^N \sqrt{\mathbf{x}_n^\top \mathbf{x}_n}$  is the  $\ell_{21}$  mixed norm of  $\mathbf{X}$  which promotes sparsity among the columns of  $\mathbf{X}$ , i.e., it promotes solutions of (9) with a small number of nonlinear or outlier pixels. This regularization term has received increasing interest in recent years [28], [36]–[39] and is known as a collaborative regularization since it uses information about the residuals in all the pixels to promote group-sparsity over the columns of  $\mathbf{X}$ . The effect of this mixed norm is illustrated in Fig. 1 (middle). Equation (9) includes a combination of the  $\ell_1$  norm and the  $\ell_{21}$  mixed norm which leads to a slightly different effect as highlighted in Fig. 1 (right). Indeed, this combination allows for sparsity among the elements of the columns of  $\mathbf{X}$ . Finally the cost function (9) is a sum of convex functions that is solved using the ADMM algorithm proposed in [40], [49] and described in the next section.

### A. The ADMM Algorithm

---

**Algorithm 1:** ADMM variant for (10).

---

```

1: Initialization
2: Initialize  $\mathbf{U}_j^{(0)}, \mathbf{D}_j^{(0)}, \forall j, \mu > 0$ . Set  $k \leftarrow 0, \text{conv} \leftarrow 0$ 
3: while conv = 0 do
4:   for j=1:J do
5:      $\xi_j^{(k)} \leftarrow \mathbf{U}_j^{(k)} + \mathbf{D}_j^{(k)}$ ,
6:   end for
7:   Linear system of equations
8:    $\mathbf{Z}^{(k+1)} \leftarrow \mathbf{G}^{-1} \sum_{j=1}^J (\mathbf{H}_j)^\top \xi_j^{(k)}$ ,
9:   Moreau proximity operators
10:  for j=1:J do
11:     $\mathbf{V}_j^{(k)} \leftarrow \mathbf{H}_j \mathbf{Z}^{(k+1)} - \mathbf{D}_j^{(k)}$ ,
12:     $\mathbf{U}_j^{(k+1)} \leftarrow \underset{\mathbf{U}_j}{\text{argmin}} \frac{\mu}{2} \|\mathbf{U}_j - \mathbf{V}_j^{(k)}\|^2 + g_j(\mathbf{U}_j)$ ,
13:  end for
14:  Update Lagrange multipliers
15:  for j=1:J do
16:     $\mathbf{D}_j^{(k+1)} \leftarrow \mathbf{U}_j^{(k+1)} - \mathbf{V}_j^{(k)}$ ,
17:  end for
18:   $k = k + 1$ 
19: end while

```

---

Consider the optimization problem

$$\underset{\mathbf{Z}}{\text{argmin}} \mathcal{C}(\mathbf{Z}) = \underset{\mathbf{Z}}{\text{argmin}} \sum_{j=1}^J g_j(\mathbf{H}_j \mathbf{Z}) \quad (10)$$

where  $\mathbf{Z} \in \mathbb{R}^{(R+D) \times N}$ ,  $g_j: \mathbb{R}^{p_j \times N} \rightarrow \mathbb{R}$  are closed, proper, convex functions, and  $\mathbf{H}_j \in \mathbb{R}^{p_j \times (R+D)}$  are arbitrary matrices. After denoting  $\mathbf{U}_j = \mathbf{H}_j \mathbf{Z} \in \mathbb{R}^{p_j \times N}$  and introducing the auxiliary variable  $\mathbf{D}_j \in \mathbb{R}^{p_j \times N}$ , the authors in [40], [49] introduced the ADMM variant summarized in Algo. 1 to solve (10) using a variable splitting and an Augmented Lagrangian algorithm. This algorithm is designed to solve any sum of an  $\ell_2$  norm with convex functions. Moreover, [50, Theorem 1] states that Algo. 1 converges when the matrix  $\mathbf{G} = \left[ \sum_{j=1}^J (\mathbf{H}_j)^\top \mathbf{H}_j \right]$  has full rank, and the functions  $g_j$  are closed, proper, and convex. Under these conditions, the same theorem states that, for any  $\mu > 0$ , if (10) has a non-empty set of solutions, then the generated sequence  $\mathbf{Z}^{(k)}$  converges to a solution. If (10) does not have a solution, then at least one of the sequences  $\mathbf{U}^{(k)}$  or  $\mathbf{D}^{(k)}$  diverges. These conditions will be studied for each of the proposed optimization problems in the next sections. Note that the main steps of Algo. 1, in each iteration, are the solution of a linear system of equations (line 8), the computation of the Moreau proximity operators (MPOs) [51] (line 12), and the updating of the Lagrange multipliers (line 16). More details regarding these computations are provided in Appendix B. Another important point to note is that the setting of  $\mu$  has a strong impact on the convergence speed of the algorithm. In this paper,  $\mu$  is updated using the adaptive procedure described in [38], [41], whose objective is to keep the ratio between the ADMM primal and dual residual norms within a given positive interval,

as they both converge to zero. Note finally that the algorithms are stopped if the primal or dual residual norms are lower than a given threshold [41]. We refer the reader to [38], [40], [41], [49] for more details regarding the ADMM algorithm.

### B. The NUSAL-K Algorithm

This section presents the optimization problem considered for estimating the parameters of the NL model (3). We first recall the two assumptions: (i) the nonlinearity appears in some pixels of the image, (ii) in a nonlinear pixel, only a few interactions are active. Under these considerations, we propose to solve the following optimization problem

$$\begin{aligned} \mathcal{C}_{\text{NUSAL-K}}(\mathbf{Z}) = & \frac{1}{2} \|\mathbf{Y} - [\mathbf{M}, \mathbf{Q}]\mathbf{Z}\|_F^2 \\ & + \tau_1 \|\mathbf{\Gamma}\|_1 + \tau_2 \|\mathbf{\Gamma}\|_{2,1} \\ & + i_{\mathbb{R}_+}(\mathbf{Z}) + i_{\{\mathbf{1}_{(1,R)}\}}(\mathbf{1}_{(1,R)}\mathbf{A}) \end{aligned} \quad (11)$$

where  $\mathbf{\Gamma} = [\gamma_1, \dots, \gamma_N]$  is a  $(D_K \times N)$  matrix of nonlinear coefficients, and  $\mathbf{Z} = [\mathbf{A}^\top, \mathbf{\Gamma}^\top]^\top$ . The mixed norm  $\ell_{2,1}$  imposes sparsity on the nonlinear pixels, i.e., it imposes sparsity on the columns of  $\mathbf{\Gamma}$  (see Fig. 1). In addition, the  $\ell_1$  norm further enforces sparsity on the nonlinear interactions in the active nonlinear pixels as highlighted in Fig. 1 (right). Using the same notation as in (10), problem (11) can be expressed as the sum of  $J = 5$  convex terms given by

$$\begin{aligned} g_1(\mathbf{U}_1) &= \mathcal{L}_{\mathbf{Q}}(\mathbf{U}_1), & \mathbf{H}_1 &= \mathbb{I}_{(R+D_K)} \\ g_2(\mathbf{U}_2) &= \tau_1 \|\mathbf{U}_2\|_1, & \mathbf{H}_2 &= [\mathbf{0}_{(D_K,R)}, \mathbb{I}_{D_K}] \\ g_3(\mathbf{U}_3) &= \tau_2 \|\mathbf{U}_3\|_{2,1}, & \mathbf{H}_3 &= [\mathbf{0}_{(D_K,R)}, \mathbb{I}_{D_K}] \\ g_4(\mathbf{U}_4) &= i_{\mathbb{R}_+}(\mathbf{U}_4), & \mathbf{H}_4 &= \mathbb{I}_{(R+D_K)} \\ g_5(\mathbf{U}_5) &= i_{\{\mathbf{1}^\top\}}(\mathbf{1}^\top \mathbf{U}_5), & \mathbf{H}_5 &= [\mathbb{I}_R, \mathbf{0}_{(R,D_K)}] \end{aligned} \quad (12)$$

where  $\mathbb{I}_n$  denotes the  $n \times n$  identity matrix and  $\mathbf{0}_{(i,j)}$  denotes the  $i \times j$  matrix of zeros. For this problem, the matrix  $\mathbf{G}$  is given by  $\mathbf{G} = \text{diag}\{[3\mathbf{1}_{(1,R)}, 4\mathbf{1}_{(1,D_K)}]\}$  which is clearly of full rank. This matrix and the properties of  $g_i, i \in \{1, \dots, J\}$  ensures the algorithm convergence.

### C. The RUSAL Algorithm

The optimization problem used to estimate the parameters of the ME model (6) is based on following assumptions: (i) the outliers appear at some pixels of the image, (ii) the residual spectra are smooth (i.e., the DCT coefficients are sparse). Under these considerations, we propose to solve the following optimization problem

$$\begin{aligned} \mathcal{C}_{\text{RUSAL}}(\mathbf{Z}) = & \frac{1}{2} \|\mathbf{Y} - [\mathbf{M}, \mathbf{F}^T]\mathbf{Z}\|_F^2 \\ & + \tau_1 \|\mathbf{B}\|_1 + \tau_2 \|\mathbf{B}\|_{2,1} \\ & + i_{\mathbb{R}_+}(\mathbf{A}) + i_{\{\mathbf{1}_{(1,R)}\}}(\mathbf{1}_{(1,R)}\mathbf{A}) \end{aligned} \quad (13)$$

where  $\mathbf{B} = [\mathbf{b}_1, \dots, \mathbf{b}_N]$ , and  $\mathbf{Z} = [\mathbf{A}^\top, \mathbf{B}^\top]^\top$ . In a similar fashion to NUSAL-K, the mixed norm  $\ell_{2,1}$  ensures spatial

sparsity of the mismodeling coefficients  $\mathbf{B}$ . In addition, the  $\ell_1$  norm further enforces sparsity on the DCT coefficients of each active pixel to impose spectral smoothness of the residuals. Using the same notation as in (10), problem (13) can be expressed as the sum of  $J = 5$  convex terms given by

$$\begin{aligned} g_1(\mathbf{U}_1) &= \mathcal{L}_{\mathbf{F}^\top}(\mathbf{U}_1), & \mathbf{H}_1 &= \mathbb{I}_{(R+D)} \\ g_2(\mathbf{U}_2) &= \tau_1 \|\mathbf{U}_2\|_1, & \mathbf{H}_2 &= [\mathbf{0}_{(D,R)}, \mathbb{I}_D] \\ g_3(\mathbf{U}_3) &= \tau_2 \|\mathbf{U}_3\|_{2,1}, & \mathbf{H}_3 &= [\mathbf{0}_{(D,R)}, \mathbb{I}_D] \\ g_4(\mathbf{U}_4) &= i_{\mathbb{R}_+}(\mathbf{U}_4), & \mathbf{H}_4 &= [\mathbb{I}_R, \mathbf{0}_{(R,D)}] \\ g_5(\mathbf{U}_5) &= i_{\{\mathbf{1}^\top\}}(\mathbf{1}^\top \mathbf{U}_5), & \mathbf{H}_5 &= [\mathbb{I}_R, \mathbf{0}_{(R,D)}]. \end{aligned} \quad (14)$$

For this problem, the full rank matrix  $\mathbf{G}$  is given by  $\mathbf{G} = 3\mathbb{I}_{(R+D)}$  which, in addition to the properties of  $g_i, i \in \{1, \dots, J\}$ , ensures the algorithm convergence.

#### D. Computational Complexity

The ADMM algorithm involves the iterative update of the matrices  $\mathbf{Z} \in \mathbb{R}^{(R+D) \times N}$  (line 8 in algo. 1) and  $\mathbf{U} \in \mathbb{R}^{p_j \times N}$  (line 12 in Algo. 1), where the details of the optimizations with respect to  $\mathbf{U}_j, j \in \{1, \dots, 5\}$  are provided in the Appendix B. The computational complexity of Algo. 1 per iteration is  $\mathcal{O}((R+D)^2 N)$ , which is related to the most expensive step introduced by the calculus of  $\mathbf{U}_1$ . Finally, it is interesting to note that the matrices to inverse involve low complexity since the matrix  $\mathbf{G}$  in line 8 is diagonal, and the matrix to inverse to update  $\mathbf{U}_1$  is fixed and then can be computed beforehand outside the iterative loop.

### IV. SIMULATION RESULTS ON SYNTHETIC DATA

This section evaluates the performance of the proposed algorithms with synthetic data. This enables the performance of the algorithms to be compared on data with a known ground truth. All simulations have been implemented using MATLAB R2015a on a computer with Intel(R) Core(TM) i7-4790 CPU@3.60GHz and 32GB RAM. The section is divided into three parts whose objectives are: 1) introducing the criteria used for the evaluation of the unmixing quality, 2) description of the synthetic images considered in the experiments, and 3) evaluating and comparing the proposed NUSAL- $K$  and RUSAL algorithms with other state-of-the-art algorithms.

#### A. Evaluation Criteria

The performance of the algorithm has been assessed in terms of abundance estimation by comparing the estimated and actual abundances using the average root mean square error (aRMSE) defined by  $\text{aRMSE}(\mathbf{A}) = \sqrt{\frac{1}{NR} \sum_{n=1}^N \|\mathbf{a}_n - \hat{\mathbf{a}}_n\|_2^2}$ . As a measure of fit, we consider the following reconstruction error  $\text{RE} = \sqrt{\frac{1}{NL} \sum_{n=1}^N \|\hat{\mathbf{y}}_n - \mathbf{y}_n\|_2^2}$  and spectral angle mapper  $\text{SAM} = \frac{1}{N} \sum_{n=1}^N \arccos\left(\frac{\hat{\mathbf{y}}_n^\top \mathbf{y}_n}{\|\hat{\mathbf{y}}_n\|_2 \|\mathbf{y}_n\|_2}\right)$  criteria, where  $\arccos$

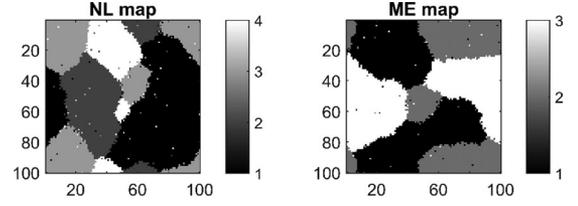


Fig. 2. Label maps associated with (left) the NL synthetic image, (right) the ME synthetic image.

$(\cdot)$  is the inverse cosine operator and  $\mathbf{y}_n, \hat{\mathbf{y}}_n$  denote the  $\#n$ th measured and estimated pixel spectra.

#### B. Description of the Synthetic Images

The proposed unmixing algorithms are evaluated on two images with different parameters. The images of size  $100 \times 100$  pixels and  $L = 207$  spectral bands have been generated using  $R$  endmembers corresponding to spectral signatures available in the ENVI software library [52]. All images have been corrupted by i.i.d. Gaussian noise of variance  $\sigma^2$  whose level is adjusted to obtain  $\text{SNR} = 25$  dB where  $\text{SNR} = 10 \log\left(\frac{\|\mathbf{M}\mathbf{A} + \Phi\|_F^2}{LN\sigma^2}\right)$ . The images have been generated using different mixture models as follows

- 1) Linear+Nonlinear models: image  $I_1$  has been generated with 4 linear/nonlinear models. An image partition into 4 classes has been generated by considering a Potts-Markov random field (with granularity parameter  $\beta = 0.8$ ) as shown in Fig. 2 (left). The four spatial classes are associated with the LMM, NL-3 model (3) (with  $\gamma_n \sim \mathcal{N}_{(\mathbb{R}_+)^p}(\mathbf{0}_{D,1}, 0.1\mathbb{I}_D)$ ), GBM (with random nonlinear coefficients in  $[0.8, 1]$ ) and PPNMM (with  $b = 0.5$ ), respectively. Note that the generated nonlinear coefficients  $\gamma_n$  are not sparse, which is a challenging scenario for the NUSAL- $K$  algorithm.
- 2) Mismodeling effects: image  $I_2$  has been partitioned into 3 classes by considering a Potts-Markov random field as shown in Fig. 2 (right). Pixels of the first class have been generated according to the LMM model, and the pixels of class 2 have been generated while considering EV. This has been achieved by varying the endmembers in each pixel of the image. Indeed, a pixel dependent smooth spectral function  $\mathbf{p}_{rn} \in \mathbb{R}^{L \times 1}$  has been added to each endmember to model EV. As in [21], the smooth functions were generated as follows  $\mathbf{p}_{rn} \sim \mathcal{N}(\mathbf{0}_{L \times 1}, \epsilon^2 \Sigma_p)$ , where  $\Sigma_p$  is an  $(L \times L)$  squared-exponential covariance matrix modeling the spectral correlations and  $\epsilon^2 = 0.001$ . The pixels of class 3 have been generated according to the ME model proposed in [21], since it leads to smooth spectral residuals as in (6). More precisely, the residuals have been generated as follows  $\phi_n^{\text{ME}} \sim \mathcal{N}(\mathbf{0}_{L \times 1}, \epsilon^2 \Sigma_p)$ , with  $\epsilon^2 = 0.002$ .

For both images and in each class, the abundances have been generated uniformly in the simplex defined by the ANC and ASC, i.e., they are distributed according to a Dirichlet distribution with parameters equal to  $\mathbf{1}_{R,1}$ . Note finally that both images

TABLE I  
RESULTS ON THE LMM-NL BASED SYNTHETIC IMAGE  $I_1$  FOR  $R \in \{3, 6\}$  ENDMEMBERS AND SNR= 25 DB

|         | R= 3  |       |       |       |       |      |       |      | R= 6  |       |       |       |       |       |       |      |
|---------|-------|-------|-------|-------|-------|------|-------|------|-------|-------|-------|-------|-------|-------|-------|------|
|         | RMSE  |       |       |       | RMSE  | RE   | SAM   | Time | RMSE  |       |       |       | RMSE  | RE    | SAM   | Time |
|         | $C_1$ | $C_2$ | $C_3$ | $C_4$ |       |      |       |      | $C_1$ | $C_2$ | $C_3$ | $C_4$ |       |       |       |      |
| LMM     | NL-3  | GBM   | PPNMM | LMM   | NL-3  | GBM  | PPNMM |      |       |       |       |       |       |       |       |      |
| FCLS    | 1.4   | 20.3  | 5.8   | 11.8  | 10.82 | 9.66 | 7.62  | 1    | 3.7   | 39.8  | 13.0  | 16.8  | 20.78 | 27.72 | 11.42 | 2    |
| SUNSAL  | 1.4   | 20.3  | 5.8   | 11.9  | 10.82 | 9.66 | 7.62  | 0.1  | 3.7   | 40.6  | 13.0  | 16.8  | 21.11 | 27.26 | 11.42 | 0.2  |
| SKhype  | 2.2   | 11.7  | 3.0   | 3.9   | 6.01  | —    | —     | 466  | 5.0   | 16.7  | 5.4   | 5.7   | 9.09  | —     | —     | 201  |
| CDA-NL  | 1.4   | 4.5   | 2.1   | 4.2   | 2.93  | 2.67 | 5.79  | 182  | 3.7   | 13.2  | 11.5  | 6.1   | 9.00  | 3.50  | 7.64  | 1560 |
| CDA-EV  | 3.3   | 23.3  | 4.5   | 9.6   | 11.84 | 2.94 | 6.18  | 246  | 5.0   | 37.5  | 6.7   | 8.6   | 18.42 | 6.56  | 9.17  | 555  |
| PLMM    | 1.6   | 28.2  | 7.1   | 15.8  | 14.81 | 1.01 | 2.15  | 17   | 10.7  | 82.8  | 12.0  | 14.0  | 40.17 | 2.25  | 5.08  | 234  |
| CDA-ME  | 1.8   | 21.2  | 5.4   | 11.1  | 11.04 | 2.59 | 5.66  | 64   | 4.1   | 30.9  | 7.9   | 9.6   | 15.63 | 3.14  | 7.17  | 66   |
| RNMF    | 1.5   | 12.8  | 2.5   | 5.2   | 6.45  | 8.04 | 6.77  | 110  | 5.3   | 22.2  | 6.7   | 8.1   | 11.79 | 2.63  | 3.85  | 50   |
| RUSAL   | 1.4   | 17.8  | 6.5   | 10.6  | 9.72  | 3.04 | 6.24  | 48   | 5.7   | 38.0  | 11.7  | 14.5  | 19.74 | 3.54  | 7.37  | 35   |
| NUSAL-2 | 1.4   | 3.9   | 2.0   | 5.0   | 2.88  | 2.69 | 5.81  | 7    | 3.7   | 9.1   | 5.7   | 6.2   | 6.04  | 3.42  | 7.47  | 26   |
| NUSAL-3 | 1.4   | 2.9   | 2.0   | 4.9   | 2.59  | 2.65 | 5.75  | 19   | 3.7   | 7.4   | 4.6   | 5.4   | 5.16  | 3.25  | 7.29  | 96   |

have been generated with the number of endmembers varying in the interval  $\{3, 6\}$ .

### C. Performance of the Proposed Algorithms

The proposed RUSAL and NUSAL- $K$  algorithms are compared to state-of-the-art algorithms by processing the generated synthetic images. We consider the two variants NUSAL-2 and NUSAL-3 to study the effect of high order interaction terms. The comparison algorithms are associated with different mixture models as follows

- 1) Linear unmixing: the abundances are estimated using the FCLS algorithm [7] and the SUNSAL algorithm [9].
- 2) Nonlinear unmixing: the abundances are estimated using the CDA-NL algorithm [21] and the SKhype algorithm [11].
- 3) Endmember variability: the abundances are estimated using the CDA-EV algorithm [21] and the PLMM algorithm [48].
- 4) Mismodeling effects (robust algorithms): the abundances are estimated using the CDA-ME algorithm [21] and the RNMF algorithm<sup>1</sup> [36].

For comparison purposes, the endmembers of these algorithms have been fixed to the actual spectra used to generate the data (the endmember update step in RNMF and PLMM has been removed). Moreover, the CDA algorithms have been used while fixing the illumination coefficient to the value #1 to provide a fair comparison with the remaining algorithms. Note also that the PLMM, RNMF, NUSAL- $K$  and RUSAL algorithms require the regularization parameters to be set. In this study, we provide the best performance (in terms of abundance RMSE) of these algorithms when varying the regularization parameters as follows:  $\gamma$  of PLMM varies in  $\{0.3, 0.5, 1\}$ ,  $\lambda$  of RNMF varies in  $\{0.01\lambda_0, 0.1\lambda_0, \lambda_0\}$  (where  $\lambda_0$  has been suggested in [36]), for RUSAL:  $\tau_1$  and  $\tau_2$  vary in  $\{0.001, 0.003, 0.006, 0.01, 0.05, 0.1\}$  and  $D = 20$  in all experiments, and for NUSAL- $K$   $\tau_1$  and  $\tau_2$  vary in  $\{0.01, 0.05, 0.1\}$ . Table I reports the results when processing the first image with  $R \in \{3, 6\}$  endmembers. The

RMSE of each spatial class (associated with different mixture model) are also reported. The proposed NUSAL-2 and NUSAL-3 algorithms provide the best RMSE performance for the LMM, RCA-NL-3 and the GBM pixels. For PPNMM, the best RMSE is obtained with SKhype that is well adapted to this polynomial nonlinearity. The best overall RMSE is obtained by the NUSAL-2 and NUSAL-3 algorithms with a slightly better values for NUSAL-3 since it estimates more parameters than NUSAL-2. Except for the LMM-based algorithms, the data are well fitted by the algorithms as indicated by the values of RE and SAM (especially with PLMM which probably overfits the data). Moreover, it is important to mention the reduced computational time of the proposed NUSAL- $K$  algorithms. Indeed, Table I clearly shows that the NUSAL-2 and NUSAL-3 algorithms are faster than the NL state-of-the-art algorithms, i.e., CDA-NL and SKhype. It is also shown that NUSAL-3 requires more computational time than NUSAL-2, while it performs slightly better. This highlights the effect of the third order nonlinear interaction terms that improve the unmixing at a price of a higher computational time. As expected, the mismodeling-based algorithms CDA-ME, RNMF and RUSAL provide an intermediate performance between the LMM algorithms and the NL algorithms. Indeed, these algorithms are designed to deal with different effects including the NL effect. Moreover, RUSAL is less sensitive to the variation of the endmember number  $R$  than NUSAL- $K$  and CDA-NL. Indeed, the latter algorithms account for interaction terms whose number increases with  $R$ , while RUSAL use a flexible residual formulation that is not related to  $R$  (it simply accounts for the spectral smoothness of the residuals). Table II shows the obtained results when processing the second image that includes pixels with LMM, EV and ME. The best RMSE performance are generally obtained with the CDA-ME algorithm. The proposed RUSAL algorithm provides competitive abundance RMSE with RNMF with the advantage of a reduced computational time. In contrast with SKhype which demonstrates robust behavior, the NL based algorithms are not well adapted to these data and provide a lower unmixing quality than the ME algorithms. These results highlight the benefit of the NUSAL- $K$  and RUSAL algorithms that show competitive results when compared to the other algorithms. Moreover, the proposed algorithms exhibit a reduced computational cost that is suitable for real world applications. While both NUSAL- $K$  and

<sup>1</sup>The RNMF was introduced in [36] as a nonlinear algorithm. In this paper, we consider it as an intermediate model between ME models (since it does not account for multiple scattering) and NL model (since it includes the non-negativity constraint).

TABLE II  
RESULTS ON THE LMM-ME BASED SYNTHETIC IMAGE  $I_2$  FOR  $R \in \{3, 6\}$  ENDMEMBERS AND SNR= 25 DB

|         | R= 3  |       |       |      |     |     |      | R= 6  |       |       |      |     |     |      |
|---------|-------|-------|-------|------|-----|-----|------|-------|-------|-------|------|-----|-----|------|
|         | RMSE  |       |       | RMSE | RE  | SAM | Time | RMSE  |       |       | RMSE | RE  | SAM | Time |
|         | $C_1$ | $C_2$ | $C_3$ |      |     |     |      | $C_1$ | $C_2$ | $C_3$ |      |     |     |      |
| FCLS    | 1.2   | 6.7   | 12.6  | 8.1  | 3.3 | 5.7 | 1    | 2.4   | 6.8   | 13.4  | 8.6  | 2.5 | 5.3 | 1    |
| SUNSAL  | 1.2   | 6.7   | 12.6  | 8.1  | 3.3 | 5.7 | 0.1  | 2.4   | 6.8   | 13.4  | 8.6  | 2.5 | 5.3 | 0.2  |
| SKhype  | 1.9   | 5.2   | 9.2   | 6.0  | —   | —   | 437  | 3.2   | 4.7   | 8.4   | 5.8  | —   | —   | 223  |
| CDA-NL  | 1.2   | 6.9   | 13.2  | 8.4  | 2.7 | 5.7 | 317  | 2.4   | 6.6   | 12.3  | 8.0  | 2.2 | 5.0 | 2278 |
| CDA-EV  | 2.3   | 3.5   | 6.8   | 4.5  | 2.3 | 5.0 | 227  | 3.0   | 4.6   | 9.9   | 6.5  | 2.0 | 4.7 | 391  |
| PLMM    | 1.4   | 7.3   | 14.7  | 9.3  | 1.7 | 3.6 | 20   | 5.0   | 6.4   | 12.4  | 8.4  | 1.5 | 3.6 | 36   |
| CDA-ME  | 1.5   | 3.3   | 3.8   | 2.9  | 2.3 | 5.0 | 49   | 2.7   | 4.6   | 6.0   | 4.5  | 2.0 | 4.7 | 71   |
| RNMF    | 1.2   | 6.8   | 12.9  | 8.3  | 3.4 | 5.7 | 137  | 2.6   | 5.8   | 12.2  | 7.8  | 2.1 | 4.5 | 279  |
| RUSAL   | 1.3   | 5.4   | 8.9   | 5.9  | 2.2 | 4.8 | 31   | 2.5   | 6.1   | 11.1  | 7.2  | 2.0 | 4.6 | 35   |
| NUSAL-2 | 1.2   | 6.8   | 12.8  | 8.1  | 3.0 | 5.7 | 6    | 2.4   | 6.1   | 12.8  | 8.1  | 2.3 | 5.1 | 41   |
| NUSAL-3 | 1.2   | 6.8   | 12.8  | 8.2  | 3.0 | 5.7 | 19   | 2.5   | 6.1   | 12.8  | 8.1  | 2.3 | 5.1 | 138  |

CDA-NL algorithms are sensitive to an increase in the number of endmembers, this effect is more important for CDA-NL while it is reduced for the NUSAL- $K$  algorithms that remain faster than SKhype for  $R = 6$  and  $K = 3$  (i.e., 77 interaction terms as reported in Table V). Note finally that additional experiments, conducted with SNR= 15 dB, show a reduction of the unmixing quality for all algorithms. However, the algorithms relative behavior is similar to the studied case, and the conclusions remain valid. These results are not provided here for brevity.

## V. RESULTS ON REAL DATA

This section illustrates the performance of the proposed algorithms when applied to three real hyperspectral images. The first hyperspectral image has received much attention in the remote sensing community [16], [53]. This image was acquired over Moffett Field, CA, in 1997 by AVIRIS. The dataset contains  $100 \times 100$  pixels,  $L = 152$  spectral bands (after removing water absorption bands) acquired in the interval  $0.4 - 2.5 \mu\text{m}$ , has a spatial resolution of 100m and is mainly composed of three components: water, soil, and vegetation (see Fig. 3(a)). This image is interesting since it is known to include bilinear scattering effects [16], [21], [36] which makes it suitable for the assessment of the NUSAL- $K$  and RUSAL algorithms presented in this paper. The second image, denoted as Madonna, was acquired in 2010 by the Hypspec hyperspectral scanner over Villelongue, France (00 03'W and 4257'N). The dataset contains  $L = 160$  spectral bands recorded from the visible to near infrared (400 – 1000 nm) with a spatial resolution of 0.5m [54]. It has previously been studied in [21], [55], [56] showing NL effects (between the trees and the soil), EV effects (mainly for the vegetation) and shadow effect. The subimage considered contains  $160 \times 200$  pixels and is composed of  $R = 4$  components: tree, grass, soil and shadow (see Fig. 3(b)). For these two images, the VCA algorithm [5] was used to extract the corresponding endmembers, i.e.,  $R = 3$  endmembers for the Moffett image and  $R = 4$  endmembers for the Madonna image. The third image was acquired by the AVIRIS sensor, in 1998, over Salinas Valley, California (see Fig. 3(c)). The dataset contains  $86 \times 83$  pixels, 204 spectral bands with the same spectral resolution and spectral range as the Moffett image (the water absorption bands were removed) and a spatial resolution of

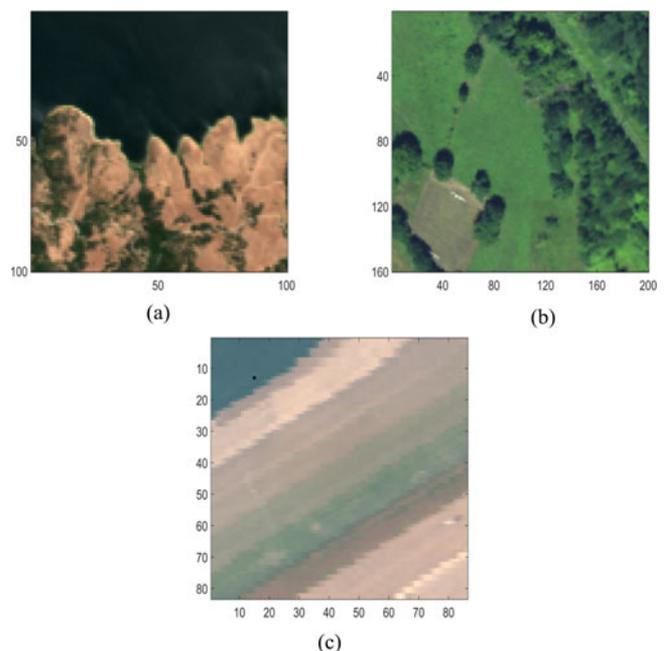


Fig. 3. Real hyperspectral images. (a) Moffett image, (b) Madonna image, (c) Salinas image.

3.7 m. This image is interesting since it includes different species of vegetables showing endmember variability, which makes it suitable for the assessment of the RUSAL algorithm. According to the ground truth information [57], this image contains 6 classes that are: Broccoli, Corn\_senesced\_green\_weeds, lettuce of different ages (4, 5, 6, and 7 weeks). As a result of the similarity between the different spectra and the presence of highly mixed pixels [58], we have manually extracted 4 endmembers associated with these classes: Corn\_senesced\_green\_weeds + lettuce-4-5, Broccoli, lettuce-6, and lettuce-7.<sup>2</sup> Indeed, these endmembers have a different shape (minimum pairwise angle of 9 degrees) while the remaining fluctuations can be associated with the effect of EV.

Table III shows the unmixing performance for the different algorithms. Overall, the NL and robust algorithms provide a

<sup>2</sup>Each endmember is obtained by averaging bundle of spectra belonging to its class.

TABLE III  
RESULTS ON REAL IMAGES. RE (RESP. SAM) SHOULD BE MULTIPLIED BY  $10^{-3}$   
(RESP.  $\times 10^{-2}$ ). THE TIME OF PROCESSING THE WHOLE IMAGE  
IS GIVEN IN SECONDS.

|         | Moffett |      |      | Madonna |     |      | Salinas |     |      |
|---------|---------|------|------|---------|-----|------|---------|-----|------|
|         | RE      | SAM  | Time | RE      | SAM | Time | RE      | SAM | Time |
| FCLS    | 13.6    | 12.7 | 1    | 13.7    | 4.8 | 5    | 8.1     | 3.8 | 1    |
| SUNSAL  | 13.6    | 12.7 | 0.1  | 13.7    | 4.8 | 0.5  | 8.4     | 3.9 | 0.1  |
| SKhype  | -       | -    | 177  | -       | -   | 551  | -       | -   | 136  |
| CDA-NL  | 7.3     | 10.7 | 317  | 12.5    | 4.6 | 2700 | 21      | 6.3 | 206  |
| CDA-EV  | 3.7     | 5.5  | 252  | 8.8     | 3.3 | 2972 | 2.0     | 1.8 | 27   |
| PLMM    | 3.4     | 3.6  | 23   | 4.6     | 1.7 | 158  | 0.5     | 0.3 | 90   |
| CDA-ME  | 2.8     | 3.3  | 17   | 8.6     | 3.1 | 342  | 1.7     | 1.7 | 40   |
| RNMF    | 5.1     | 8.3  | 278  | 6.7     | 1.9 | 703  | 8.3     | 3.2 | 97   |
| RUSAL   | 2.7     | 3.7  | 46   | 7.1     | 2.6 | 94   | 1.8     | 1.6 | 20   |
| NUSAL-2 | 7.0     | 11.0 | 13   | 12.7    | 4.6 | 94   | 8.0     | 3.8 | 5    |
| NUSAL-3 | 6.9     | 10.4 | 29   | 12.7    | 4.6 | 181  | 8.0     | 3.8 | 19   |

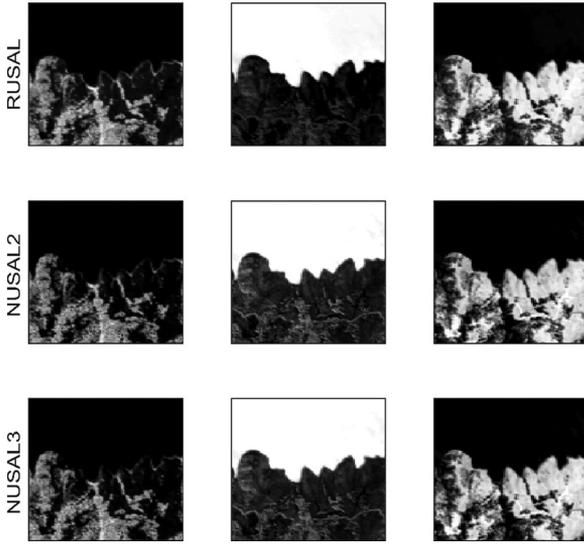


Fig. 4. Estimated abundance maps with different algorithms for the Moffett image. (Left) vegetation, (middle) water, (right) soil.

better fit than the LMM-based ones. Among the sophisticated algorithms, the proposed NUSAL-2, NUSAL-3 and RUSAL algorithms provide the best performance for the computational cost. The algorithms all generated similar abundance maps for the Moffett image and we only show those of NUSAL-2, NUSAL-3 and RUSAL in Fig. 4, for brevity. Fig. 5 presents the residual maps associated with the NL, EV and robust algorithms. This figure highlights good agreement between the NL algorithms (CDA-NL, NUSAL-2, NUSAL-3) that detect nonlinearity in the coastal region (as in [16]). In addition to this region, the robust algorithms (RNMF, ME, RUSAL) detect other mismodeling effects probably due to EV as already reported in [21] and shown by the result of CDA-EV. The NL coefficients estimated by NUSAL-2 and NUSAL-3 are reported in Fig. 6. This figure shows good agreement between the estimated bilinear coefficients when considering NUSAL-2 and NUSAL-3. Moreover, it highlights the sparse behavior of the nonlinear coefficients and clearly shows that they are mainly due to the second order interactions. Indeed, Fig. 6 (middle) shows that half of the averaged coefficient values of each pixel are negligible and Fig. 6 (bottom-right) shows that the average

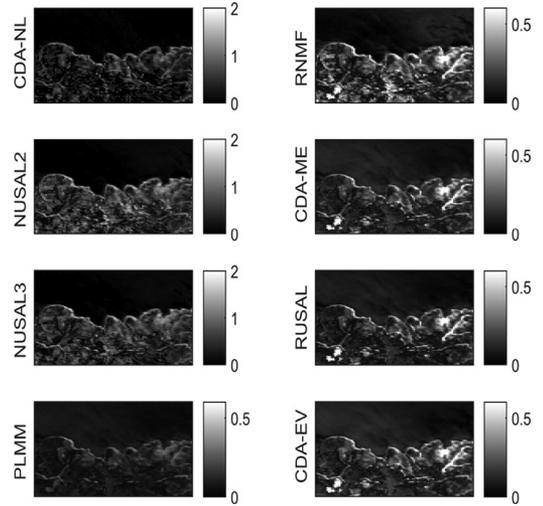


Fig. 5. Residual maps for the Moffett image obtained with  $\|\hat{\mathbf{y}}_{i,j} - \mathbf{M}\hat{\mathbf{a}}_{i,j}\|$ .

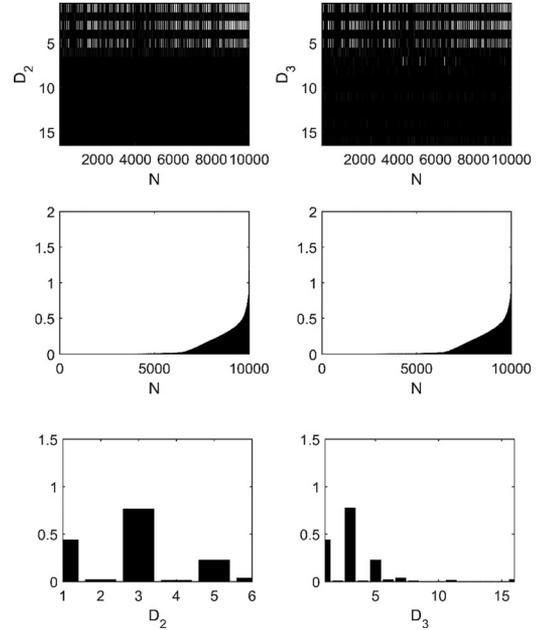


Fig. 6. Nonlinear coefficients obtained with NUSAL-2 and NUSAL-3 for the Moffett image. (Top) matrix ( $D_k \times N$ ) of NL coefficients (the color scale is  $[0,1]$ ), (middle) sorted averaged coefficient values of each pixel ( $1/D \sum_{r,r'} \gamma_n^{(r,r')}, \forall n$ ), (bottom) averaged coefficient values of each nonlinear interaction term ( $1/N \sum_{n=1}^N \gamma_n^{(r,r')}, \forall r, r'$ ).

of the nonlinear coefficients over all the pixels is higher for the first six terms, i.e., the second order terms.

The abundances obtained for the Madonna scene are displayed in Fig. 7 for SKhype, RNMF, RUSAL, NUSAL-2, and NUSAL-3 (the other algorithms provided similar maps to NUSAL/RUSAL and were not displayed for brevity). This figure shows a slight difference between the RNMF soil map and the other algorithms. Similar differences are observed when considering the residual maps in Fig. 8 since RNMF detected a higher residual effect in the soil area (bottom-left corner in the RNMF image) than ME and RUSAL. Apart this, the robust algorithms detected residuals in the shadow areas and in trees. The

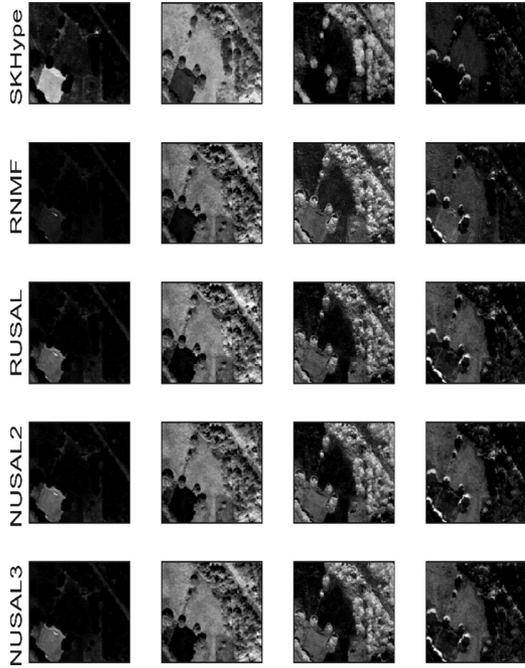


Fig. 7. Estimated abundance maps with different algorithms for the Madonna image. from left to right: soil, grass, tree, shadow.

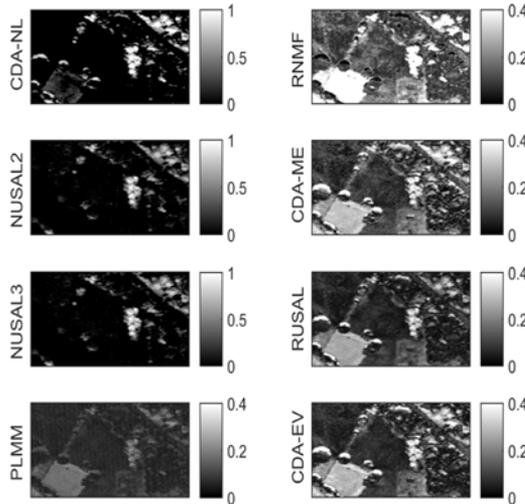


Fig. 8. Residual maps for the Madonna image obtained with  $\|\hat{\mathbf{y}}_{i,j} - \mathbf{M}\hat{\mathbf{a}}_{i,j}\|$ .

latter is mainly due to the presence of multiple scattering effects as highlighted by the NL algorithms that show similar maps (see left figures). In a similar manner to Moffett field, the NUSAL-2 and NUSAL-3 estimated NL coefficients are pixel-wise sparse and mainly due to bilinear interactions as highlighted in Fig. 9. This justifies the good behavior of the bilinear models [15]–[20], [35] that assume that the effect of the interaction terms decreases when increasing the interaction orders.

The abundances obtained for the Salinas scene are displayed in Fig. 10 for PLMM, SKHype, CDA-NL, RUSAL, NUSAL-2, and NUSAL-3 (the other algorithms provided similar maps to NUSAL/RUSAL and were not displayed for brevity). Because of the high EV effect, all of PLMM, CDA-NL and SKHype fail to

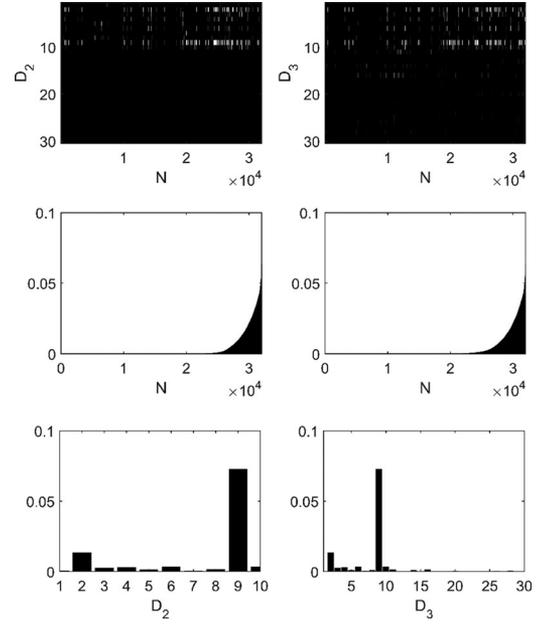


Fig. 9. Nonlinear coefficients obtained with NUSAL-2 and NUSAL-3 for the Madonna image. (Top) matrix ( $D_k \times N$ ) of NL coefficients (the color scale is  $[0,0.2]$ ), (middle) sorted averaged coefficient values of each pixel ( $1/D \sum_{r,r'} \gamma_n^{(r,r')}, \forall n$ ), (bottom) averaged coefficient values of each nonlinear interaction term ( $1/N \sum_{n=1}^N \gamma_n^{(r,r')}, \forall r, r'$ ).

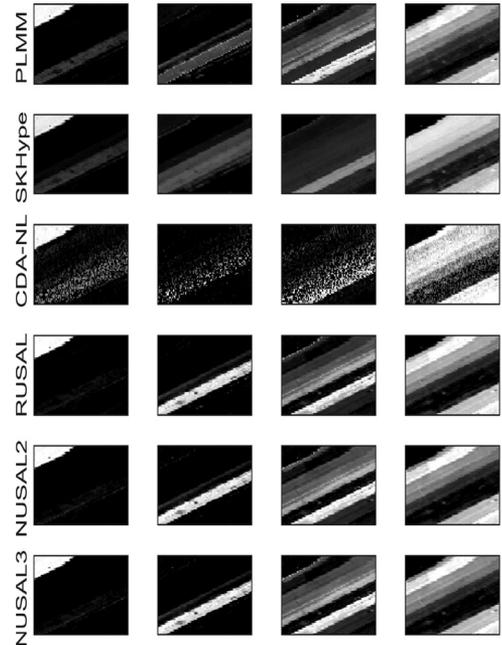


Fig. 10. Estimated abundance maps with different algorithms for the Salinas image. From left to right: Corn\_senesced\_green\_weeds + lettuce-4-5, Broccoli, lettuce-6, and lettuce-7.

extract the abundances of Broccoli, and lettuce-6. The residual maps shown in Fig. 11 confirm this since NL algorithms detect a reduced effect while PLMM, CDA-EV, CDA-ME and RUSAL detect more EV effect especially in the region of the lettuce. These results highlight the ability of CDA-ME and RUSAL to capture EV effects. Fig. 12 shows some randomly selected

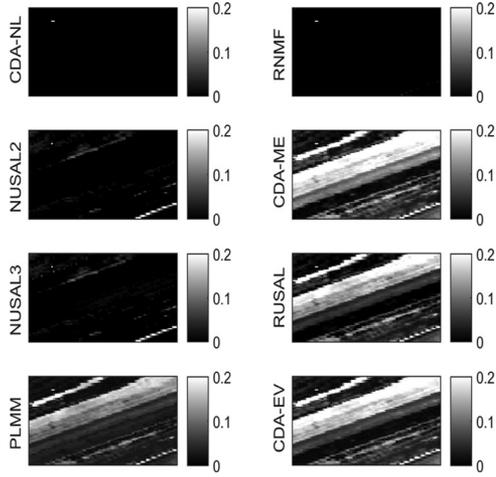


Fig. 11. Residual maps for the Salinas image obtained with  $\|\hat{y}_{i,j} - M\hat{a}_{i,j}\|$ .

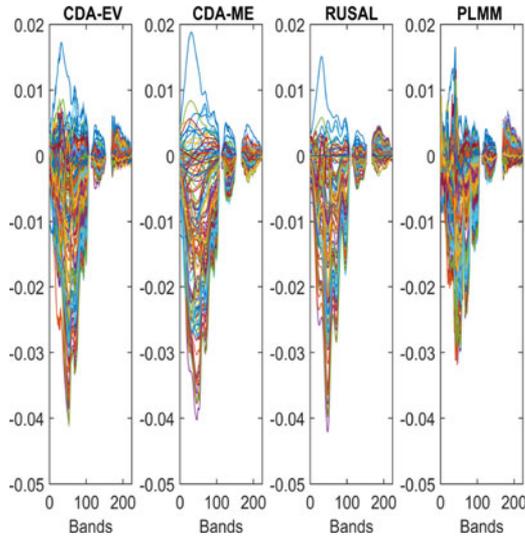


Fig. 12. Example of outlier spectra for the Salinas image obtained with (from left to right) CDA-EV, CDA-ME, RUSAL and PLMM.

outlier spectra obtained with CDA-EV, CDA-ME, RUSAL and PLMM algorithms. These spectra show a similar global shape while they highlight the properties of each algorithm. Indeed, it can be seen that PLMM provides noisy spectra since it does not account for their spectral correlations. In contrast, the CDA-EV, CDA-ME and RUSAL algorithms consider this correlation effect and provide more realistic results. In addition, the RUSAL algorithm allows the absence of outliers (null spectra) thanks to the sparsity promoting property imposed on the outliers. To summarize, the obtained results highlighted the benefit of RUSAL/NUSAL- $K$  that estimate abundance and residual maps which are in good agreement with state-of-the-art algorithms, but at a lower computational cost. NUSAL- $K$  generalizes the common bilinear models and provide NL coefficient maps associated with different interaction orders. This provides a useful tool to better analyze the scattering effect between the physical elements. RUSAL provides a flexible tool to capture different

TABLE IV  
CHARACTERISTICS OF THE STUDIED MODELS/ALGORITHMS. “POS.” STANDS FOR POSITIVITY, “SPAT.” FOR SPATIAL, “SPEC.” FOR SPECTRAL, “ILLUMIN.” FOR ILLUMINATION, “SM” FOR SMOOTH, “SP” FOR SPARSE, (+++) BEST RESULTS, AND (+) GOOD RESULTS.

|            | Effects,<br>LMM+ | Residuals |       |       | Illumin.<br>coeff. c | Robust<br>to R | Time |
|------------|------------------|-----------|-------|-------|----------------------|----------------|------|
|            |                  | Pos.      | Spat. | Spec. |                      |                |      |
| SKhype     | NL               | X         | –     | –     | X                    | ✓              | +    |
| CDANL      | NL-2             | ✓         | SM    | –     | ✓                    | X              | +    |
| CDAEV      | EV               | X         | SM    | SM    | X                    | X              | ++   |
| PLMM       | EV               | X         | –     | –     | X                    | X              | +++  |
| CDAME      | NL + EV          | X         | SM    | SM    | ✓                    | ✓              | +++  |
| RNMF       | NL               | ✓         | SP    | –     | X                    | ✓              | +    |
| RUSAL      | NL + EV          | X         | SP    | SM    | X                    | ✓              | +++  |
| NUSAL- $K$ | NL- $K$          | ✓         | SP    | –     | X                    | X              | +++  |

TABLE V  
EXAMPLE OF  $D_k$  FOR DIFFERENT VALUES OF  $R$  AND  $K$

|          | $K = 2$ | $K = 3$ | $K = 4$ | $K = 5$ |
|----------|---------|---------|---------|---------|
| $R = 3$  | 6       | 16      | 31      | 52      |
| $R = 6$  | 21      | 77      | 203     | 455     |
| $R = 10$ | 55      | 275     | 990     | 2992    |

mismodeling effects due to EV, NL or outliers. It is more robust than NUSAL- $K$  with respect to the variation of the endmember number  $R$ , thus, it is recommended to use RUSAL in presence of a large number of endmembers. However, RUSAL provides less information than NUSAL regarding the interaction terms and the origin of its residual term may be due to physical effects (such as NL, EV), data corruptions or other outliers. Table IV finally summarizes the main characteristics of the nonlinear and robust algorithms considered in this paper.

## VI. CONCLUSION

This paper has introduced two mixture models and their supervised unmixing algorithms. The two models accounted for the presence of nonlinearity or mismodeling effects by considering a residual term in addition to the linear mixture of endmembers. The residual term was expressed as a sparse linear combination of some signals, thus, the proposed models reduced to a linear combination with respect to the abundances and the residual coefficients. The unknown parameters associated with these models were estimated using an optimization approach that included convex regularization terms. More precisely, the non-negativity and sum-to-one constraints were imposed on the abundances and the residual terms were assumed to be spatially sparse by considering a collaborative sparse regression approach. The resulting convex problem was solved using an alternating direction method of multipliers whose convergence was theoretically ensured. The proposed algorithms showed good performance when processing synthetic data generated with the linear model or other more sophisticated models. Results on real data confirmed the good performance of the proposed algorithms and showed their ability to extract different

features in the observed scenes, with a reduced computational cost. These results confirmed that most vegetation nonlinearity can be captured by bilinear interactions and that endmember variability is mainly located in vegetation areas. Future work includes the introduction of spatial correlation on the abundances. Considering endmember variability jointly with nonlinearity is also an interesting issue which would deserve to be investigated.

## APPENDIX DERIVATIONS

### A. Construction of $\mathbf{Q}^{(K)}$

Model (3) requires the definition of the  $(L \times D_K)$  matrix  $\mathbf{Q}^{(K)}$  gathering the interaction spectra of all the orders lower than  $K$ . This section describes this matrix by providing its size and the coefficient of each interaction term. Before providing the full description of this matrix, let consider an example of  $R = 3$  endmembers and the matrix  $\mathbf{Q}^{(K=3)} = [\mathbf{Q}_2^{(3)}, \mathbf{Q}_3^{(3)}]$ . The number of interaction spectra is given by 16 (see Table V) while the corresponding spectra are given by concatenating the two matrices  $\mathbf{Q}_2^{(3)} = (\sqrt{2}\mathbf{m}_{12}, \sqrt{2}\mathbf{m}_{13}, \sqrt{2}\mathbf{m}_{23}, \mathbf{m}_{11}, \mathbf{m}_{22}, \mathbf{m}_{33})$ , and  $\mathbf{Q}_3^{(3)} = (\sqrt{3}\mathbf{m}_{112}, \sqrt{3}\mathbf{m}_{113}, \sqrt{3}\mathbf{m}_{122}, \sqrt{3}\mathbf{m}_{322}, \sqrt{3}\mathbf{m}_{133}, \sqrt{3}\mathbf{m}_{233}, \sqrt{6}\mathbf{m}_{123}, \mathbf{m}_{111}, \mathbf{m}_{222}, \mathbf{m}_{333})$ , with  $\mathbf{m}_{ijk} = \mathbf{m}_i \odot \mathbf{m}_j \odot \mathbf{m}_k$ . For a formal mathematical description, denote  $\mathbf{Q}^{(K)} = [\mathbf{Q}_2^{(K)}, \mathbf{Q}_3^{(K)}, \dots, \mathbf{Q}_K^{(K)}]$ , where  $\mathbf{Q}_i^{(K)}$  gathers the interaction spectra of the  $i$ th order. The size of  $\mathbf{Q}^{(K)}$  is then obtained by summing the size of the interaction spectra  $D_K(i)$  associated with the  $i$ th order, as follows

$$D_K = \sum_{i=2}^K D_K(i) = \sum_{i=2}^K \frac{(R+i-1)!}{i!(R-1)!} \quad (15)$$

where  $x! = 1 \cdots (x-2)(x-1)x$ , denotes the factorial of  $x$ . Table V shows some examples of  $D_K$  for different values of  $R$  and  $K$ . It is clear that increasing the interaction term  $K$  leads to a fast increase of the number of interaction terms included in  $\mathbf{Q}^{(K)}$ . However, it is interesting to note that the sparsity promoting norms ( $\ell_1$  and  $\ell_{21}$ ) are well adapted to deal with large  $D_K$ .

Similarly to [10], [11], [21], each interaction term in  $\mathbf{Q}_i^{(K)}$  is weighted by a coefficient that is obtained by comparison with a homogeneous polynomial kernel of the  $i$ th degree. Straightforward computations show that the  $i$ th order spectra gathered in  $\mathbf{Q}_i^{(K)}$  are given by

$$\sqrt{\frac{i!}{\prod_{r=1}^R k_r!}} \prod_{1 \leq r \leq R} \mathbf{m}_r^{k_r}, \text{ subject to } \sum_{r=1}^R k_r = i. \quad (16)$$

### B. ADMM Algorithm

The list shown below provides details regarding the considered ADMM algorithm for both NUSAL and RUSAL. More precisely, we provide the solutions for the linear system of equations shown in line 8 of Algo. 1 and the MPO optimization

problems shown in line 12. The details of the MPOs can be found, for example, in [51].

1) Linear system of equations:

$$\mathbf{Z}^{(k+1)} \leftarrow \mathbf{G}^{-1} \sum_{j=1}^J (\mathbf{H}_j)^\top \xi_j^{(k)},$$

with  $\mathbf{G} = \text{diag}\{[3\mathbf{1}_{(1,R)}, 4\mathbf{1}_{(1,D_K)}]\}$  for the NL model and  $\mathbf{G} = 3\mathbb{I}_{(R+D)}$  for the ME model

2) MPO for  $g_1(\mathbf{U}_1) = \mathcal{L}_P(\mathbf{U}_1)$ :

$$\mathbf{U}_1^{(k+1)} \leftarrow \{[\mathbf{M}, \mathbf{P}]^\top [\mathbf{M}, \mathbf{P}] + \mu \mathbb{I}_{D+R}\}^{-1} \times \{[\mathbf{M}, \mathbf{P}]^\top \mathbf{Y} + \mu \mathbf{V}_1^{(k)}\}$$

3) MPO for  $g_2(\mathbf{U}_2) = \tau_1 \|\mathbf{U}_2\|_1$ :

$$\mathbf{U}_2^{(k+1)} \leftarrow \text{soft}\left(\mathbf{V}_2^{(k)}, \frac{\tau_1}{\mu}\right)$$

4) MPO for  $g_3(\mathbf{U}_3) = \tau_2 \|\mathbf{U}_3\|_{2,1}$ :

$$\mathbf{u}_{3,n}^{(k+1)} \leftarrow \text{vect-soft}\left(\mathbf{v}_{3,n}^{(k)}, \frac{\tau_2}{\mu}\right), \forall n$$

5) MPO for  $g_4(\mathbf{U}_4) = i_{\mathbb{R}_+}(\mathbf{U}_4)$ :

$$\mathbf{U}_4^{(k+1)} \leftarrow \max\{\mathbf{V}_4^{(k)}, 0\}$$

6) MPO for  $g_5(\mathbf{U}_5) = i_{\{\mathbf{1}^\top\}}(\mathbf{1}^\top \mathbf{U}_5)$ :

$$\mathbf{U}_5^{(k+1)} \leftarrow \left(\mathbb{I}_R - \frac{1}{R} \mathbf{1}_{(R,R)}\right) \mathbf{V}_5^{(k)} + \frac{1}{R} \mathbf{1}_{(R,N)}$$

where  $\text{soft}(\cdot)$  denotes the soft threshold operator given by  $\text{soft}\left(\mathbf{V}, \frac{\tau}{\mu}\right) = \text{sign}(\mathbf{V}) \odot \max\{|\mathbf{V}| - \frac{\tau}{\mu}, 0\}$ ,  $\text{sign}(\cdot)$  denotes the element-wise application of the sign function,  $|\mathbf{V}|$  denotes the matrix of absolute values of the elements of  $\mathbf{V}$ ,  $\max(\cdot)$  is the element-wise maximum operator, and  $\text{vect-soft}(\cdot)$  is the well known vect-soft-threshold operator given by  $\text{vect-soft}\left(\mathbf{v}, \frac{\tau}{\mu}\right) = \mathbf{v} \left(\frac{\max\{\|\mathbf{v}\|_2 - \frac{\tau}{\mu}, 0\}}{\max\{\|\mathbf{v}\|_2 - \frac{\tau}{\mu}, 0\} + \frac{\tau}{\mu}}\right)$ . Note finally that  $\mathbf{P} = \mathbf{Q}$  for NUSAL and  $\mathbf{P} = \mathbf{F}^\top$  for RUSAL.

## REFERENCES

- [1] J. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.
- [2] A. Halimi, P. Honeine, M. Kharouf, C. Richard, and J. Y. Tourneret, "Estimating the intrinsic dimension of hyperspectral images using a noise-whitened eigengap approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 7, pp. 3811–3821, Jul. 2016.
- [3] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.
- [4] C. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 608–619, Mar. 2004.
- [5] J. M. P. Nascimento and J. M. 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.
- [6] M. Winter, "Fast autonomous spectral end-member determination in hyperspectral data," in *Proc. 13th Int. Conf. Appl. Geologic Remote Sens.*, Apr. 1999, vol. 2, pp. 337–344.

- [7] D. C. Heinz and C.-I. Chang, "Fully constrained least-squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 29, no. 3, pp. 529–545, Mar. 2001.
- [8] J. Chen, C. Richard, and P. Honeine, "Nonlinear estimation of material abundances in hyperspectral images with  $\ell_1$ -norm spatial regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2654–2665, May 2014.
- [9] J. Bioucas-Dias and M. Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," in *Proc. IEEE GRSS Workshop Hyperspectral Image Signal Process.: Evol. Remote Sens.*, Jun. 2010, pp. 1–4.
- [10] Y. Altmann, M. Pereyra, and S. McLaughlin, "Bayesian nonlinear hyperspectral unmixing with spatial residual component analysis," *IEEE Trans. Comput. Imag.*, vol. 1, no. 3, pp. 174–185, Sep. 2015.
- [11] J. Chen, C. Richard, and P. Honeine, "Nonlinear unmixing of hyperspectral data based on a linear-mixture/nonlinear-fluctuation model," *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 480–492, Jan. 2013.
- [12] R. Heylen, M. Parente, and P. Gader, "A review of nonlinear hyperspectral unmixing methods," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 1844–1868, Jun. 2014.
- [13] N. Dobigeon, J.-Y. Tourneret, C. Richard, J. Bermudez, S. McLaughlin, and A. Hero, "Nonlinear unmixing of hyperspectral images: Models and algorithms," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 82–94, Jan. 2014.
- [14] B. W. Hapke, "Bidirectional reflectance spectroscopy. I. Theory," *J. Geophys. Res.*, vol. 86, pp. 3039–3054, 1981.
- [15] Y. Altmann, A. Halimi, N. Dobigeon, and J.-Y. Tourneret, "Supervised nonlinear spectral unmixing using a postnonlinear mixing model for hyperspectral imagery," *IEEE Trans. Image Process.*, vol. 21, no. 6, pp. 3017–3025, Jun. 2012.
- [16] A. Halimi, Y. Altmann, N. Dobigeon, and J.-Y. Tourneret, "Nonlinear unmixing of hyperspectral images using a generalized bilinear model," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4153–4162, Nov. 2011.
- [17] A. Halimi, Y. Altmann, N. Dobigeon, and J.-Y. Tourneret, "Unmixing hyperspectral images using the generalized bilinear model," in *Proc. IEEE Int. Conf. Geosci. Remote Sens.*, Jul. 2011, pp. 1886–1889.
- [18] J. M. Bioucas-Dias and J. M. P. Nascimento, "Nonlinear mixture model for hyperspectral unmixing," in *Proc. SPIE Image Signal Process. Remote Sens. XV*, 2009, vol. 7477, no. 1, Art. no. 74770I.
- [19] W. Fan, B. Hu, J. Miller, and M. Li, "Comparative study between a new nonlinear model and common linear model for analysing laboratory simulated-forest hyperspectral data," *Int. J. Remote Sens.*, vol. 30, no. 11, pp. 2951–2962, Jun. 2009.
- [20] I. Meganem, P. Deliot, X. Briottet, Y. Deville, and S. Hosseini, "Linear-quadratic mixing model for reflectances in urban environments," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 544–558, Jan. 2014.
- [21] A. Halimi, P. Honeine, and J. M. Bioucas-Dias, "Hyperspectral unmixing in presence of endmember variability, nonlinearity or mismodeling effects," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4565–4579, Oct. 2016.
- [22] B. Somers, G. P. Asner, L. Tits, and P. Coppin, "Endmember variability in spectral mixture analysis: A review," *Remote Sens. Environ.*, vol. 115, no. 7, pp. 1603–1616, 2011.
- [23] A. Zare and K. Ho, "Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 95–104, Jan. 2014.
- [24] L. Drumetz, M. A. Veganzones, S. Henrot, R. Phlypo, J. Chanussot, and C. Jutten, "Blind hyperspectral unmixing using an extended linear mixing model to address spectral variability," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3890–3905, Aug. 2016.
- [25] S. Henrot, J. Chanussot, and C. Jutten, "Dynamical spectral unmixing of multitemporal hyperspectral images," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3219–3232, Jul. 2016.
- [26] M. A. Veganzones *et al.*, "A new extended linear mixing model to address spectral variability," in *Proc. IEEE Workshop Hyperspectral Image Signal Process.: Evol. Remote Sens.*, Jun. 2014, p. n/c.
- [27] A. Halimi, N. Dobigeon, J. Y. Tourneret, S. McLaughlin, and P. Honeine, "Unmixing multitemporal hyperspectral images accounting for endmember variability," in *Proc. 23rd Eur. Signal Process. Conf.*, Aug. 2015, pp. 1656–1660.
- [28] H. K. Aggarwal and A. Majumdar, "Hyperspectral unmixing in the presence of mixed noise using joint-sparsity and total variation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, 2016, vol. 9, no. 9, pp. 4257–4266, Sep. 2016.
- [29] W. He, H. Zhang, and L. Zhang, "Sparsity-regularized robust non-negative matrix factorization for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4267–4279, Sep. 2016.
- [30] O. Eches, N. Dobigeon, C. Mailhes, and J.-Y. Tourneret, "Bayesian estimation of linear mixtures using the normal compositional model: Application to hyperspectral imagery," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1403–1413, Jun. 2010.
- [31] A. Zare, P. Gader, and G. Casella, "Sampling piecewise convex unmixing and endmember extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 3, pp. 1655–1665, Mar. 2013.
- [32] B. Somers, M. Zortea, A. Plaza, and G. Asner, "Automated extraction of image-based endmember bundles for improved spectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 396–408, Apr. 2012.
- [33] Y. Altmann, S. McLaughlin, and A. Hero, "Robust linear spectral unmixing using anomaly detection," *IEEE Trans. Comput. Imag.*, vol. 1, no. 2, pp. 74–85, Jun. 2015.
- [34] A. A. Kalaitzis and N. D. Lawrence, "Residual components analysis," in *Proc. Int. Conf. Mach. Learn.*, 2012, pp. 1–3.
- [35] Y. Altmann, N. Dobigeon, S. McLaughlin, and J.-Y. Tourneret, "Residual component analysis of hyperspectral images: Application to joint nonlinear unmixing and nonlinearity detection," *IEEE Trans. Image Process.*, vol. 23, no. 5, pp. 2148–2158, May 2014.
- [36] C. Févotte and N. Dobigeon, "Nonlinear hyperspectral unmixing with robust nonnegative matrix factorization," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4810–4819, Jun. 2015.
- [37] P. Sprechmann, I. Ramirez, G. Sapiro, and Y. C. Eldar, "C-Hilasso: A collaborative hierarchical sparse modeling framework," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4183–4198, Sep. 2011.
- [38] M. D. Iordache, J. M. 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.
- [39] M. D. Iordache, J. M. Bioucas-Dias, A. Plaza, and B. Somers, "MUSIC-CSR: Hyperspectral unmixing via multiple signal classification and collaborative sparse regression," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 4364–4382, Jul. 2014.
- [40] M. Afonso, J. Bioucas-Dias, and M. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–695, Mar. 2011.
- [41] 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.
- [42] Y. Altmann, A. Halimi, N. Dobigeon, and J. Y. Tourneret, "Supervised nonlinear spectral unmixing using a polynomial post nonlinear model for hyperspectral imagery," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2011, pp. 1009–1012.
- [43] R. Heylen and P. Scheunders, "A multilinear mixing model for nonlinear spectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 240–251, Jan. 2016.
- [44] A. Marinoni, J. Plaza, A. Plaza, and P. Gamba, "Nonlinear hyperspectral unmixing using nonlinearity order estimation and polytope decomposition," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2644–2654, Jun. 2015.
- [45] A. Marinoni and P. Gamba, "A novel approach for efficient p-linear hyperspectral unmixing," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 6, pp. 1156–1168, Sep. 2015.
- [46] A. Halimi *et al.*, "Robust unmixing algorithms for hyperspectral imagery," in *Proc. Sensor Signal Process. Defence*, Sep. 2016, pp. 1–5.
- [47] F. Zhu, A. Halimi, P. Honeine, B. Chen, and N. Zheng, "Correntropy maximization via ADMM—Application to robust hyperspectral unmixing," in *ArXiv e-prints*, Feb. 2016.
- [48] P. A. Thouvenin, N. Dobigeon, and J. Y. Tourneret, "Hyperspectral unmixing with spectral variability using a perturbed linear mixing model," *IEEE Trans. Signal Process.*, vol. 64, no. 2, pp. 525–538, Jan. 2016.
- [49] M. Figueiredo and J. Bioucas-Dias, "Restoration of poissonian images using alternating direction optimization," *IEEE Trans. Image Process.*, vol. 19, no. 12, pp. 3133–3145, Dec. 2010.
- [50] J. Eckstein and D. P. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Program.*, vol. 55, no. 1, pp. 293–318, 1992.

- [51] P. L. Combettes and J.-C. Pesquet, *Proximal Splitting Methods in Signal Processing*. New York, NY, USA: Springer, 2011, pp. 185–212.
- [52] RSI, *ENVI User's Guide Version 4.0*, Research Systems Inc., Boulder, CO USA, Sep. 2003.
- [53] N. Dobigeon, J.-Y. Tourneret, and C.-I. Chang, "Semi-supervised linear spectral unmixing using a hierarchical Bayesian model for hyperspectral imagery," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2684–2695, Jul. 2008.
- [54] D. Sheeren, M. Fauvel, S. Ladet, A. Jacquin, G. Bertoni, and A. Gibon, "Mapping ash tree colonization in an agricultural mountain landscape: Investigating the potential of hyperspectral imagery," in *Proc. IEEE Int. Conf. Geosci. Remote Sens.*, Jul. 2011, pp. 3672–3675.
- [55] A. Halimi, N. Dobigeon, and J.-Y. Tourneret, "Unsupervised unmixing of hyperspectral images accounting for endmember variability," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4904–4917, 2015.
- [56] Y. Altmann, N. Dobigeon, S. McLaughlin, and J.-Y. Tourneret, "Unsupervised post-nonlinear unmixing of hyperspectral images using a Hamiltonian Monte Carlo algorithm," *IEEE Trans. Image Process.*, vol. 23, no. 6, pp. 2663–2675, Jun. 2014.
- [57] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4085–4098, Nov. 2010.
- [58] A. Plaza, P. Martinez, J. Plaza, and R. Perez, "Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 466–479, Mar. 2005.



**Abderrahim Halimi** (S'11–M'14) received the Engineering degree in electronics from the National Polytechnic School of Algiers, Algeria, in 2009, and the M.Sc. and Ph.D. degrees in signal processing from the Institut National Polytechnique de Toulouse, Toulouse, France, in 2010 and 2013, respectively. From October 2013 to September 2015, he was a Postdoctoral Research Associate with the University of Toulouse and the University of Technology of Troyes, France, under the support of the HYPANEMA ANR Project. Since November 2015, he

has been a Postdoctoral Research Associate in the School of Engineering and Physical Sciences, Heriot-Watt University, Scotland. His research activities focus on statistical signal and image processing, with a particular interest in Bayesian inverse problems with applications to hyperspectral imaging, satellite altimetry, and single photon depth imaging.



**José M. Bioucas-Dias** (SM'15) received the E.E., M.Sc., Ph.D., and "Agregado" degrees from Instituto Superior Técnico (IST), Technical University of Lisbon (TULisbon, now University of Lisbon), Portugal, in 1985, 1991, 1995, and 2007, respectively, all in electrical and computer engineering.

Since 1995, he has been in the Department of Electrical and Computer Engineering, IST, where he was an Assistant Professor from 1995 to 2007 and an Associate Professor since 2007. Since 1993, he is also a Senior Researcher in the Pattern and Image

Analysis group, Instituto de Telecomunicações, which is a private nonprofit research institution. His research interests include inverse problems, signal and image processing, pattern recognition, optimization, and remote sensing. His research work has been highly cited and he is included in Thomson Reuters' Highly Cited Researchers 2015 list.

Dr. Bioucas-Dias was an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS (1997–2000) and IEEE TRANSACTIONS ON IMAGE PROCESSING (2010–2014) and he is a Senior Area Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING and an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He was the General Co-Chair of the Third IEEE GRSS Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote sensing (WHISPERS'2011) and has been a member of program/technical committees of several international conferences.



**Nicolas Dobigeon** (S'05–M'08–SM'13) was born in Angoulême, France, in 1981. He received the Engineering degree in electrical engineering from ENSEEIHT, Toulouse, France, and the M.Sc. degree in signal processing from INP Toulouse, both in 2004, the Ph.D. degree and the Habilitation à Diriger des Recherches in signal processing from INP Toulouse in 2007 and 2012, respectively.

From 2007 to 2008, he was a Postdoctoral Research Associate in the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA. Since 2008, he has been with INP-ENSEEIHT Toulouse, University of Toulouse, where he is currently a Professor. He conducts his research within the Signal and Communications Group, IRIT Laboratory, and he is also an Affiliated Faculty Member of the TeSA Laboratory. His recent research activities have been focused on statistical signal and image processing, with a particular interest in Bayesian inverse problems with applications to remote sensing and biomedical imaging.



**Gerald S. Buller** received the B.Sc.(Hons.) degree in natural philosophy from the University of Glasgow, Scotland, in 1986 and the Ph.D. degree in physics from Heriot-Watt University, Edinburgh, U.K. in 1989. He is currently a Professor of physics at Heriot-Watt University. In 2002, he co-founded Helia Photonics Ltd., based in Livingston, Scotland. His current research interests include single-photon detection methods, particularly at infrared wavelengths, and applications involving the use of single-photons.

These applications include single-photon depth imaging and novel quantum communications protocols. Prof. Buller is a Fellow of the Royal Society of Edinburgh, a Fellow of the UK Institute of Physics, and a Fellow of the Optical Society of America. In 2015, he received the EPSRC Established Career Fellowship in Quantum Technology to research sparse photon imaging.



**Stephen McLaughlin** (F'11) was born in Clydebank, Scotland, in 1960. He received the B.Sc. degree in electronics and electrical engineering from the University of Glasgow, Scotland, in 1981 and the Ph.D. degree from the University of Edinburgh, Edinburgh, Scotland, in 1990. From 1981 to 1984, he was a Development Engineer in industry. In 1986, he joined the University of Edinburgh where he studied the performance of linear adaptive algorithms in high noise and nonstationary environments. In 1988, he joined the academic staff at Edinburgh, and from 1991 until

2001, he held a Royal Society University Research Fellowship to study nonlinear signal processing techniques. In 2002, he received the personal Chair in Electronic Communication Systems at the University of Edinburgh. In October 2011, he joined Heriot-Watt University as a Professor of Signal Processing and Head of the School of Engineering and Physical Sciences. His research interests include fields of adaptive signal processing and nonlinear dynamical systems theory and their applications to biomedical, energy, and communication systems. Prof. McLaughlin is a Fellow of the Royal Academy of Engineering, the Royal Society of Edinburgh, and the Institute of Engineering and Technology.