Nonlinear Hyperspectral Unmixing With Robust Nonnegative Matrix Factorization

Cédric Févotte and Nicolas Dobigeon

Abstract—We introduce a robust mixing model to describe hyperspectral data resulting from the mixture of several pure spectral signatures. The new model extends the commonly used linear mixing model by introducing an additional term accounting for possible nonlinear effects, that are treated as sparsely distributed additive outliers. With the standard nonnegativity and sum-to-one constraints inherent to spectral unmixing, our model leads to a new form of robust nonnegative matrix factorization with a group-sparse outlier term. The factorization is posed as an optimization problem, which is addressed with a block-coordinate descent algorithm involving majorization–minimization updates. Simulation results obtained on synthetic and real data show that the proposed strategy competes with the state-of-the-art linear and nonlinear unmixing methods.

Index Terms—Hyperspectral imagery, nonlinear unmixing, robust nonnegative matrix factorization, group-sparsity.

I. INTRODUCTION

SPECTRAL unmixing (SU) is an issue of prime interest when analyzing hyperspectral data since it provides a comprehensive and meaningful description of the collected measurements in various application fields including remote sensing [1], planetology [2], food monitoring [3] or spectro-microscopy [4]. SU consists in decomposing P multi-band observations $\mathbf{Y} = \left[\mathbf{y}_1, \ldots, \mathbf{y}_P\right]$ into a collection of K individual spectra $\mathbf{M} = [\mathbf{m}_1, \ldots, \mathbf{m}_K]$, called *endmembers*, and estimating their relative proportions (or *abundances*) $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_P]$ for each observation [5], [6]. Most of the hyperspectral unmixing algorithms proposed in the signal & image processing and geoscience literatures rely on the commonly admitted linear mixing model (LMM), $\mathbf{Y} \approx \mathbf{MA}$. Indeed, LMM provides a good approximation of the physical process underlying the observations and has resulted in useful results for many applications. However, for some other specific applications, LMM may be inaccurate and other models, referred to as "nonlinear", need to be considered. For

C. Févotte is with the Laboratoire Lagrange, Centre National de la Recherche Scientifique, Observatoire de la Côte d'Azur, Université Nice Sophia Antipolis, Nice 06108, France (e-mail: cfevotte@unice.fr).

N. Dobigeon is with Institut de Recherche en Informatique de Toulouse/ INP-École nationale supérieure d'électrotechnique, d'électronique, d'informatique, d'hydraulique et des télécommunications, University of Toulouse, Toulouse 31071, France (e-mail: nicolas.dobigeon@enseeiht.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/TIP.2015.2468177

instance, when considering sand-like scenes, light is subjected to multiple scattering and absorption phenomena, which result in highly nonlinear effects. Proper analysis of such mixtures involves very complex optical modeling and one typically needs to resort to approximate models to make the problem tractable. As such, the Hapke model [7] has been widely advocated, in particular for unmixing purposes [8]–[11]. Conversely, in remotely sensed images composed of vegetation (e.g., trees), interactions of photons with multiple components of the scene lead to nonlinear effects that can be taken into account using bilinear models [12]–[17]. As explained in [18], many of these models only differ by the constraints imposed on the bilinear term. Additionally, to approximate a large range of second-order nonlinearities, Altmann *et al.* [19] introduced a polynomial post-nonlinear model that has demonstrated its ability to describe many nonlinear effects, in particular in vegetated areas [20]. A common feature of these models is that they all incorporate a supplementary additive term to the standard LMM, that accounts for the nonlinearities. One major drawback of these models, however, is that they require to choose a specific form of nonlinearity, and this can be limiting in practice. References [21] and [22] offer comprehensive and recent overviews of nonlinear models and associated unmixing algorithms.

In this paper, a new so-called robust LMM (rLMM) is proposed. In a similar fashion to the nonlinear models detailed above, it is built on the standard LMM, with the inclusion of a supplementary additive term that accounts for nonlinear effects. As a consequence, the proposed method can be considered to analyze any mixture that results from the combination of a linear contribution $\mathbf{MA}$ and a residual term. Bilinear models [18] and the multi-mixture pixel model (MMP) [23], which combines macroscopic and intimate mixtures, are archetypal examples of this class of models that are encompassed by our proposed approach. However, contrary to these models, rLMM does not require to specify a given analytical form of the nonlinearity. Instead, nonlinearities are merely treated as *outliers*. The main motivation for introducing rLMM is to introduce a flexible unmixing method able to analyze a large variety of remotely sensed scenes. In particular, it is well admitted that, for most of these images, the linear mixing model (LMM) is a relevant description of the data. For such datasets, resorting to nonlinear mixing models is not always necessary and may in fact result in inaccurate estimation (due to, e.g., overfitting). Thus the LMM can be considered as a valid model to describe the majority of the pixels in a remotely sensed image. Conversely, as illustrated in previous works [21],

the LMM assumption does not hold for specific and localized areas, mainly situated at the interface of heterogeneous regions. For this limited number of pixels, LMM-based unmixing algorithms fail in properly recovering the materials and their abundances. From this premise, our goal is to propose a new model (as well as the corresponding unmixing algorithm) appropriate for *both* scenarios. As such, we propose to decompose the $L \times P$ matrix of the multi-band observations as $\mathbf{Y} \approx \mathbf{MA} + \mathbf{R}$, where $\mathbf{R}$ is a sparse (and nonnegative) residual term accounting for outliers (i.e., nonlinear effects). To reflect the assumption that the LMM holds for most of the observed pixels, the sparsity constraint is imposed at the group-level, i.e., a column of $\mathbf{R}$ will be assumed to be either entirely zero or not. Finally, it is also worthy to note that the proposed decomposition relates to *robust nonnegative matrix factorization* (rNMF) as will be explained in more details in the sequel of the paper.

The article is organized as follows. The rLMM is detailed in Section II. Section III describes a block-coordinate descent algorithm for rLMM estimation. Experimental results obtained on synthetic data are reported in Section IV. Two real hyperspectral images are investigated in Section V and Section VI concludes. This article extends our preliminary conference paper [24] in a significant way. We show how some of the multiplicative updates obtained heuristically in [24] can be rigorously obtained via majorization-minimization. While [24] only considers the case where the measure of fit is the squared Euclidean distance, this article additionally considers the Kullback-Leibler divergence, a common measure of fit for nonnegative data. In this article, we also describe an efficient rule of thumb for the critical task of choosing the value of the penalty weight. Finally, we provide extensive experimental results on synthetic and real data.

II. ROBUST LINEAR MIXING MODEL

A. Model Design

The proposed rLMM is described by

$$\mathbf{y}_p \approx \sum_{k=1}^{K} a_{kp}\mathbf{m}_k + \mathbf{r}_p, \qquad (1)$$

where $\mathbf{y}_p = [y_{1p}, \ldots, y_{Lp}]^T$ denotes the pth pixel spectrum observed in L spectral bands, $\mathbf{m}_k = [m_{1k}, \ldots, m_{Lk}]^T$ denotes the kth endmember spectrum, $\mathbf{a}_p = [a_{1p}, \ldots, a_{Kp}]^T$ denotes the abundances associated with the pth pixel and $\mathbf{r}_p = [r_{1p}, \ldots, r_{Lp}]^T$ denotes the outlier term (accounting for nonlinearities). The matrix formulation of Eq. (1) is given by

$$\mathbf{Y} \approx \mathbf{MA} + \mathbf{R}. \qquad (2)$$

Per se, this robust LMM generalizes most of the bilinear mixing models [18] that have been widely used to analyze hyperspectral scenes acquired over multi-layered areas by remote sensors. This specific context has been extensively studied in the literature, for planetary and Earth science [25], [26], in particular to characterize vegetation [13], [27]–[30] or urban [17] canopies. Note however that the proposed nonlinear model cannot be used to

analyze mineralogical data, that result from intimate mixtures of sand-like materials.

The approximation symbol in Eqs. (1) and (2) underlies the minimization of a measure of dissimilarity $D(\mathbf{Y}|\mathbf{MA} + \mathbf{R})$. The measure of dissimilarity is such that $D(\mathbf{A}|\mathbf{B}) = \sum_{ij} d(a_{ij}|b_{ij})$, where $d(x|y)$ is either the squared Euclidean distance (SED), given by $d_{\text{SED}}(x|y) = \frac{1}{2}(x - y)^2$, or the Kullback-Leibler divergence (KLD), given by $d_{\text{KLD}}(x|y) = x \log \frac{x}{y} - x + y$. We address these two measures of dissimilarity because they are the most commonly used in NMF. The SED is the more common one in hyperspectral unmixing, but recent papers such as [31] have also pointed the benefits of using the KLD. As will be discussed in Section VI, the methodology presented in the paper could also accommodate other measures of fit such as the more general β-divergence [32].

The matrices $\mathbf{Y}$, $\mathbf{M}$ and $\mathbf{A}$ are nonnegative by nature and we assume the abundance coefficients to sum to one, i.e.,

$$\mathbf{a}_p \in \mathbb{S}^K \overset{\text{def}}{=} \left\{ \boldsymbol{a} \in \mathbb{R}^K \,\middle|\, a_k \geq 0, \ \sum_{k=1}^{K} a_k = 1 \right\}, \qquad (3)$$

as commonly assumed in most hyperspectral data models. In this work, we assume the nonlinear component $\mathbf{r}_p$ to be nonnegative as well, like in the bilinear models of [12], [13], and [15] and the polynomial model with constructive interferences of [19]. This assumption allows a fair comparison with the latter works, which inspired us the proposed methodology, and is physically well-motivated for multi-layered models (scenes with multiple reflections). One could however envisage relaxing this assumption and the pros and cons of doing so are discussed in Section VI.

As discussed in the introduction, we expect $\mathbf{r}_p$ to be often zero, i.e., pixels to follow the standard LMM in general. For pixels where the LMM assumption fails, nonlinearities will become "active" and $\mathbf{r}_p$ will become nonzero. This amounts to say that the energy vector

$$\mathbf{e} = \left[\|\mathbf{r}_1\|_2, \ldots, \|\mathbf{r}_P\|_2 \right] \qquad (4)$$

is sparse. In Eq. (4), $\|\cdot\|_2$ denotes the Euclidean norm defined by $\|\mathbf{x}\|_2 = \sqrt{\sum_k x_k^2}$. Sparsity can routinely be enforced by ℓ_1-regularization, as done next.

B. Objective Function

In light of previous section, our objective is to solve the minimization problem defined by

$$\min_{\mathbf{M}, \mathbf{A}, \mathbf{R}} \quad J(\mathbf{M}, \mathbf{A}, \mathbf{R}) = D(\mathbf{Y}|\mathbf{MA} + \mathbf{R}) + \lambda \|\mathbf{R}\|_{2,1}$$

$$\text{s.t.} \quad \mathbf{M} \geq 0, \ \mathbf{A} \geq 0, \ \mathbf{R} \geq 0 \ \text{ and } \ \|\mathbf{a}_p\|_1 = 1, \qquad (5)$$

where λ is a nonnegative penalty weight, $\mathbf{A} \geq 0$ denotes nonnegativity of the coefficients of $\mathbf{A}$, $\|\mathbf{x}\|_1 = \sum_k |x_k|$ and $\|\cdot\|_{2,1}$ is the so-called $\ell_{2,1}$-norm defined by

$$\|\mathbf{R}\|_{2,1} = \|\mathbf{e}\|_1 = \sum_{p=1}^{P} \|\mathbf{r}_p\|_2. \qquad (6)$$

Eq. (5) defines a robust NMF problem. Robust NMF is a nonnegative variant of robust PCA [33] which has appeared in different forms in the literature. In [34], the outlier term $\mathbf{R}$ is nonnegative and penalized by the ℓ_1 norm. In [35] and [36], $\mathbf{R}$ is real-valued and penalized by ℓ_1 and $\ell_{1,2}$ norms, respectively. In [37], the $\ell_{2,1}$ norm of $(\mathbf{Y} - \mathbf{MA})$ is minimized (noise-free scenario). A so-called robust nonnegative matrix factorization approach was introduced for the reconstruction of reflectance spectra in [38]; however the term "robust" there refers to a different feature, namely the use of a data-fitting term (the hypersurface cost function) that is less sensitive to outlier observations than the traditional SED, for the computation of a regular NMF $\mathbf{Y} \approx \mathbf{MA}$. Note finally that other articles have addressed hyperspectral unmixing with regular NMF (i.e., in the standard linear model), e.g., [39]–[42].

To the best of our knowledge, the formulation of robust NMF described by Eq. (5), where $\mathbf{R}$ is nonnegative and penalized by the $\ell_{2,1}$ norm (and where the abundances sum to one), is entirely novel. Furthermore, previous works [34]–[37] have only considered robust NMF with the SED, i.e., $D(\mathbf{Y}|\mathbf{MA} + \mathbf{R}) = \|\mathbf{Y} - \mathbf{MA} - \mathbf{R}\|_2^2$, while we here also address the case of the KLD.

III. BLOCK-COORDINATE DESCENT ALGORITHM

In order to solve the rNMF minimization problem defined at Eq. (5), we present an iterative block-coordinate descent algorithm that updates each of the parameters $\mathbf{M}$, $\mathbf{A}$ and $\mathbf{R}$ in turn. Each parameter is updated conditionally upon the current value of the other parameters and such that the objective function is decreased. This is the updating scheme employed by virtually all NMF algorithms. Unfortunately, given the non-convexity of the objective function $J(\mathbf{M}, \mathbf{A}, \mathbf{R})$, this strategy can return local solutions and proper initialization is required. This will be addressed in Section IV. The updates of the parameters are described next. In short, the parameters $\mathbf{M}$ and $\mathbf{R}$ are updated via majorization-minimization (MM). Generally speaking, MM consists in optimizing an easier-to-minimize tight upper-bound of the original objective function [43]. The parameter $\mathbf{A}$ is updated using a heuristic scheme that has proven to work well in the literature. All the updates turn out to be "multiplicative", i.e., such that the new update is obtained by term-to-term multiplying the previous update by a nonnegative matrix, hence automatically preserving the nonnegativity of the estimates through iterations. The resulting algorithm has linear complexity $\mathcal{O}(LKP)$ (in flops) per iteration.

A. Update of the Endmember Spectra M

Updating $\mathbf{M}$ given the current values of $\mathbf{A}$ and $\mathbf{R}$ involves solving the following minimization problem

$$\min_{\mathbf{M}} \; C(\mathbf{M}) = D(\mathbf{Y}|\mathbf{MA} + \mathbf{R}) \quad \text{s.t.} \quad \mathbf{M} \geq 0, \qquad (7)$$

where $D(\cdot|\cdot)$ is either the SED or the KLD. When $\mathbf{R} = \mathbf{0}$, this problem boils down to updating the dictionary/endmembers matrix in NMF. MM algorithms have been designed for that purpose in [32], [44], and [45]. MM algorithms have been among the very first to be considered for NMF and they are still a prime choice today, thanks to their low complexity and

ease of implementation. In this section, we extend the MM approach to the case where $\mathbf{R} \geq \mathbf{0}$. Denote by $\tilde{\mathbf{M}}$ the estimate of $\mathbf{M}$ at current iteration. The first step of MM consists in building an upper bound $G(\mathbf{M}|\tilde{\mathbf{M}})$ of $C(\mathbf{M})$ which is tight for $\mathbf{M} = \tilde{\mathbf{M}}$, i.e., $C(\mathbf{M}) \leq G(\mathbf{M}|\tilde{\mathbf{M}})$ for all $\mathbf{M}$ and $C(\tilde{\mathbf{M}}) = G(\tilde{\mathbf{M}}|\tilde{\mathbf{M}})$. The second step consists in minimizing the bound with respect to (w.r.t) $\mathbf{M}$, producing a valid descent algorithm. Indeed, at iteration $i + 1$, it holds by construction that $C(\mathbf{M}^{(i+1)}) \leq G(\mathbf{M}^{(i+1)}|\mathbf{M}^{(i)}) \leq G(\mathbf{M}^{(i)}|\mathbf{M}^{(i)}) = C(\mathbf{M}^{(i)})$. The bound $G(\mathbf{M}|\tilde{\mathbf{M}})$ will be referred to as *auxiliary function*.

Using convexity of $D(\cdot|\cdot)$ w.r.t its second argument, $C(\mathbf{M})$ can be majorized using Jensen's inequality, as follows. Let us denote $\tilde{y}_{lp} = \sum_k \tilde{m}_{lk} a_{kp} + r_{lp}$ the data approximation formed with the current iterate $\tilde{\mathbf{M}}$ (and recall that $\mathbf{A}$ and $\mathbf{R}$ are here treated as constants). Then, define for $k = 1, \ldots, K$, $\tilde{\lambda}_{lkp} = \tilde{m}_{lk} a_{kp} / \tilde{y}_{lp}$ and for $k = K + 1$, $\tilde{\lambda}_{lkp} = r_{lp} / \tilde{y}_{lp}$. By construction, we have $\sum_{k=1}^{K+1} \tilde{\lambda}_{lkp} = 1$. Then, using definition of convexity, we have

$$
\begin{aligned}
C(\mathbf{M}) &= \sum_{lp} d\left(y_{lp} | \sum_k m_{lk} a_{kp} + r_{lp}\right) \\
&= \sum_{lp} d\left(y_{lp} | \sum_k \tilde{\lambda}_{lkp} \frac{m_{lk} a_{kp}}{\tilde{\lambda}_{lkp}} + \tilde{\lambda}_{l(K+1)p} \frac{r_{lp}}{\tilde{\lambda}_{l(K+1)p}}\right) \\
&\leq \sum_{lp} \left[\sum_{k=1}^{K} \tilde{\lambda}_{lkp} d\left(y_{lp} | \frac{m_{lk} a_{kp}}{\tilde{\lambda}_{lkp}}\right) \right. \\
&\qquad\qquad \left. + \tilde{\lambda}_{l(K+1)p} d\left(y_{lp} | \frac{r_{lp}}{\tilde{\lambda}_{l(K+1)p}}\right) \right] \\
&= \sum_{lp} \left[\sum_{k=1}^{K} \frac{\tilde{m}_{lk} a_{kp}}{\tilde{y}_{lp}} d\left(y_{lp} | \tilde{y}_{lp} \frac{m_{lk}}{\tilde{m}_{lk}}\right) + \frac{r_{lp}}{\tilde{y}_{lp}} d(y_{lp}|\tilde{y}_{lp}) \right] \\
&\overset{\text{def}}{=} G(\mathbf{M}|\tilde{\mathbf{M}}) \qquad\qquad (8)
\end{aligned}
$$

The auxiliary function essentially "takes out" the sum over k from within the measure of similarity in the expression of $C(\mathbf{M})$ (first line of Eq. (8)) to make the optimization over $\mathbf{M}$ separable w.r.t its entries m_{lk}. Skipping details for brevity, the resulting function can be minimized in closed-form w.r.t $\tilde{\mathbf{M}}$, resulting in the following updates

$$m_{lk}^{\text{SED}} = \tilde{m}_{lk} \frac{\sum_p a_{kp} y_{lp}}{\sum_p a_{kp} \tilde{y}_{lp}}, \qquad\qquad (9)$$

$$m_{lk}^{\text{KLD}} = \tilde{m}_{lk} \sum_p \frac{a_{kq}}{\sum_q a_{kq}} \frac{y_{lp}}{\tilde{y}_{lp}} \qquad\qquad (10)$$

where we recall that $\tilde{y}_{lp} = \sum_k \tilde{m}_{lk} a_{kp} + r_{lp}$ is the data approximation at current iteration. Note that the two updates can be expressed in the same form

$$m_{lk} = \tilde{m}_{lk} \frac{\sum_p a_{kp} y_{lp} \tilde{y}_{lp}^{\beta-2}}{\sum_p a_{kp} \tilde{y}_{lp}^{\beta-1}} \qquad\qquad (11)$$

where $\beta = 2$ for the SED and $\beta = 1$ for the KLD.

With the SED, the minimization problem defined at Eq. (7) could alternatively be cast as a nonnegative quadratic problem described by

$$\min_{\mathbf{M}} \; C(\mathbf{M}) = \frac{1}{2} \|(\mathbf{Y} - \mathbf{R}) - \mathbf{MA}\|_2^2 \quad \text{s.t.} \quad \mathbf{M} \geq 0. \qquad (12)$$

In this formulation the problem can be solved using the MM algorithm of Sha et al. [46], based on a different auxiliary function which does not require $\mathbf{R}$ to be nonnegative.[1] The algorithm results in the following multiplicative update

$$m_{lk} = \tilde{m}_{lk} \frac{|\sum_p a_{kp}(y_{lp} - r_{lp})| + \sum_p a_{kp}(y_{lp} - r_{lp})}{2 \sum_p a_{kp} \tilde{s}_{lp}}, \quad (13)$$

where $\tilde{s}_{lp} = \sum_k \tilde{m}_{lk} a_{kp}$. We compared the updates defined by Eqs. (11) & (13) experimentally and found no significant difference in convergence behavior, with however Eq. (13) being on average marginally faster. One interest of our approach, as opposed to using the algorithm of [46], is that it is not specific to the SED, and can readily be adapted to other measures of fit as the derivations in Eq. (8) hold for any divergence $d(x|y)$ which is convex w.r.t y.

B. Update of the Outlier Term $\mathbf{R}$

Updating $\mathbf{R}$ given the current values of $\mathbf{M}$ and $\mathbf{A}$ involves solving the following minimization problem

$$\min_{\mathbf{R}} C(\mathbf{R}) = D(\mathbf{Y}|\mathbf{MA} + \mathbf{R}) + \lambda \|\mathbf{R}\|_{2,1} \quad \text{s.t. } \mathbf{R} \geq 0. \quad (14)$$

The problem has an analytical solution in the squared Euclidean real-case [47], but does not to appear to have one in the nonnegative case, and nor with the KLD. We resort again to MM. The data-fitting term may be majorized exactly as we did in Section III-A. Denote by $\tilde{\mathbf{R}}$ the current update of $\mathbf{R}$, $s_{lp} = \sum_k m_{lk} a_{kp} = [\mathbf{MA}]_{lp}$ the low-rank component and $\tilde{y}_{lp} = s_{lp} + \tilde{r}_{lp}$ the current data approximation.[2] Then, applying the Jensen inequality we obtain

$$D(\mathbf{Y}|\mathbf{MA} + \mathbf{R}) \leq \sum_{lp} \left[\frac{\tilde{r}_{lp}}{\tilde{y}_{lp}} d(y_{lp}|\tilde{y}_{lp} \frac{r_{lp}}{\tilde{r}_{lp}}) + \frac{s_{lp}}{\tilde{y}_{lp}} d(y_{lp}|\tilde{y}_{lp}) \right]. \quad (15)$$

Denote by $F(\mathbf{R}|\tilde{\mathbf{R}})$ the right-hand side of Eq. (15). An auxiliary function for $C(\mathbf{R})$ may simply be obtained as $G(\mathbf{R}|\tilde{\mathbf{R}}) = F(\mathbf{R}|\tilde{\mathbf{R}}) + \lambda \|\mathbf{R}\|_{2,1}$. However, this specific auxiliary function is yet not amenable to optimization w.r.t $\mathbf{R}$ in our case (no closed-form solution). Hence, the first step of our strategy is to majorize the penalty function $\|\mathbf{R}\|_{2,1}$ as well. By concavity of the square-root function, we may write

$$\|\mathbf{R}\|_{2,1} \leq \frac{1}{2} \sum_p \left(\frac{\|\mathbf{r}_p\|_2^2}{\|\tilde{\mathbf{r}}_p\|_2} + \|\tilde{\mathbf{r}}_p\|_2 \right). \quad (16)$$

Eq. (16) essentially replaces $\sqrt{\sum_l r_{lp}^2}$ by a quadratic tight upper-bound that involves $\sum_l r_{lp}^2$, with the effect of decoupling the spectral bands from within the square root. This extra majorization returns an auxiliary function which

is now amenable to optimization, leading to the following updates

$$r_{lp}^{\text{SED}} = \tilde{r}_{lp} \frac{y_{lp}}{\tilde{y}_{lp} + \lambda \frac{\tilde{r}_{lp}}{\|\tilde{\mathbf{r}}_p\|_2}}, \quad (17)$$

$$r_{lp}^{\text{KLD}} = \frac{\|\tilde{\mathbf{r}}_p\|_2}{2\lambda} \left(\sqrt{1 + 4\lambda \frac{\tilde{r}_{lp}}{\|\tilde{\mathbf{r}}_p\|_2} \frac{y_{lp}}{\tilde{y}_{lp}}} - 1 \right). \quad (18)$$

Update (18) becomes numerically unstable when $\lambda \to 0$. As such, a better-behaved update may be obtained by further majorizing the linear term r_{lp} that appears in the expression of $G(\mathbf{R}|\tilde{\mathbf{R}})$ for the KLD by a monomial of degree 2, matching in this way the quadratic upper bound of the penalty function – such a trick is used in other settings in [48] and [49]. This results in the alternative multiplicative update given by

$$r_{lp}^{\text{KLD}} = \tilde{r}_{lp} \left(\frac{y_{lp}/\tilde{y}_{lp}}{1 + \lambda \frac{\tilde{r}_{lp}}{\|\tilde{\mathbf{r}}_p\|_2}} \right)^{\frac{1}{2}}. \quad (19)$$

In practice, ignoring the exponent $\frac{1}{2}$ in Eq. (19) still reduces the objective function at each iteration and produces faster convergence. This may be interpreted as over-relaxation of the MM update, see [32] for further discussion on this subject. Again, the final updates for $\mathbf{R}$ can be expressed in the same form

$$r_{lp} = \tilde{r}_{lp} \left(\frac{y_{lp} \tilde{y}_{lp}^{\beta-2}}{\tilde{y}_{lp}^{\beta-1} + \lambda \frac{\tilde{r}_{lp}}{\|\tilde{\mathbf{r}}_p\|_2}} \right), \quad (20)$$

where $\beta = 2$ for the SED and $\beta = 1$ for the KLD.

C. Update of the Abundances $\mathbf{A}$

Updating $\mathbf{A}$ given the current values of $\mathbf{M}$ and $\mathbf{R}$ involves solving the following minimization problem

$$\min_{\mathbf{A}} \quad C(\mathbf{A}) = D(\mathbf{Y}|\mathbf{MA} + \mathbf{R})$$
$$\text{s.t.} \quad \mathbf{A} \geq 0 \quad \text{and} \quad \forall p, \|\mathbf{a}_p\|_1 = 1. \quad (21)$$

The sum-to-one constraint on the abundances induces an extra difficulty as compared to the optimization problems involved by the updates of $\mathbf{M}$ and $\mathbf{R}$. In some cases such a constraint can be handled using Lagrange multipliers, but this approach does not succeed in our setting. We hence resort to another common approach based on a change of variable. We introduce the variable $\mathbf{U}$ to be a nonnegative matrix of dimension $K \times P$ and set

$$a_{kp} = \frac{u_{kp}}{\sum_k u_{kp}}. \quad (22)$$

The optimization problem of Eq. (21) is turned into the new optimization problem

$$\min_{\mathbf{U}} \quad C(\mathbf{U}) = D \left(\mathbf{Y} \mid \mathbf{M} \left[\frac{\mathbf{u}_1}{\|\mathbf{u}_1\|_1}, \dots, \frac{\mathbf{u}_P}{\|\mathbf{u}_P\|_1} \right] + \mathbf{R} \right)$$
$$\text{s.t.} \quad \mathbf{U} \geq 0 \quad (23)$$

which is free from the sum-to-one constraint. This approach has been used for NMF in [50]. Unfortunately, we were not

---

[1] In our own derivations, the nonnegativity of $\mathbf{R}$ is used in the definition of $\lambda_{l(K+1)p}$.

[2] The same notation $\tilde{y}_{lp}$ is used for $\tilde{y}_{lp} = \sum_k \tilde{m}_{lk} a_{kp} + r_{lp}$ in Section III-A and for $\tilde{y}_{lp} = \sum_k m_{lk} a_{kp} + \tilde{r}_{lp}$ in Section III-B. Our intent is too avoid the use of too many notations and the definition of $\tilde{y}_{lp}$ should be clear from context (i.e., which parameter update is considered).

able to produce an auxiliary function for the new objective function in (23) – in particular because it is no longer convex w.r.t the variable to be optimized. Instead, we resort to a heuristic commonly used in NMF, see, e.g., [51], [52], as follows. As it appears, the gradient of $C(\mathbf{U})$ can be expressed as the difference of two nonnegative functions such that

$$\nabla_{u_{kp}} C(\mathbf{U}) = \nabla_{u_{kp}}^+ C(\mathbf{U}) - \nabla_{u_{kp}}^- C(\mathbf{U}). \tag{24}$$

The heuristic algorithm simply writes

$$u_{kp} = \tilde{u}_{kp} \frac{\nabla_{u_{kp}}^- C(\tilde{\mathbf{U}})}{\nabla_{u_{kp}}^+ C(\tilde{\mathbf{U}})}. \tag{25}$$

It ensures the nonnegativity of the parameter updates provided initialization with a nonnegative value, and produces a descent algorithm in the sense that u_{kp} is updated towards left (resp., right) when the gradient is positive (resp., negative). The algorithm was found experimentally to decrease the value of the objective function at each iteration. Denoting $\tilde{s}_{lp} = \sum_k m_{lk} \tilde{a}_{kp}$ and $\tilde{y}_{lp} = \tilde{s}_{lp} + r_{lp}$, the update is found to be

$$u_{kp} = \tilde{u}_{kp} \frac{\sum_l (m_{lk} y_{lp} \tilde{y}_{lp}^{\beta-2} + \tilde{s}_{lp} \tilde{y}_{lp}^{\beta-1})}{\sum_l (m_{lk} \tilde{y}_{lp}^{\beta-1} + \tilde{s}_{lp} y_{lp} \tilde{y}_{lp}^{\beta-2})}. \tag{26}$$

where $\beta = 2$ for the SED and $\beta = 1$ for the KLD. The update for $\mathbf{A}$ is then simply $a_{kp} = u_{kp}/\sum_k u_{kp}$.

As it turns out, the updates (11), (20) and (26) can be implemented in matrix form, as shown in Algorithm 1, which recapitulates the overall procedure. In Algorithm 1, all operators preceded by a dot '·' are entrywise MATLAB-like operations and fraction bars shall be taken term-to-term as well. Additionally, $\mathbf{1}_{M,N}$ denotes the $M \times N$ matrix with coefficients equal to 1. A MATLAB implementation of the algorithm is available from the authors' webpages.

D. Setting the Value of λ

The hyperparameter λ controls the trade-off between the data-fitting term $D(\mathbf{Y}|\mathbf{MA} + \mathbf{R})$ and the penalty term $\|\mathbf{R}\|_{2,1}$. Setting the "right" value of λ is a difficult task, like in any other so-called variational approach that involves a regularization term. We describe in this paragraph a rule of thumb for choosing λ in a plausible range of values. Our approach is based on the method of moments. It consists in interpreting the objective function (5) as a joint likelihood and in matching the empirical mean of the data with its prior expectation in the statistical model.

The SED and the KLD are "pseudo-likelihood" for probabilistic models (Gaussian and Poisson, respectively) such that $\mathrm{E}[\mathbf{Y}|\mathbf{MA} + \mathbf{R}] = \mathbf{MA} + \mathbf{R}$. In the same analogy, the term $\lambda \|\mathbf{R}\|_{2,1}$ can be interpreted as a log-prior term. Using some results from [53], the corresponding prior distribution $p(\mathbf{r}_p|\lambda)$ for each column of $\mathbf{R}$ can be obtained as a scale mixture of conditionally independent half-Normal distributions, with a Gamma distribution assigned to the scale parameter. In particular, the expectation of r_{lp} under this prior can be found to be

$$\mathrm{E}[r_{lp}|\lambda] = \frac{2}{\sqrt{\pi}} \frac{\Gamma(K/2+1)}{\Gamma(K/2+1/2)} \frac{1}{\lambda} \stackrel{\mathrm{def}}{=} \frac{C}{\lambda}. \tag{27}$$

Algorithm 1 Group Robust NMF

Set $\beta = 2$ (SED) or 1 (KLD)
Set value of λ
Set convergence tolerance parameter 'tol'
Initialize $\mathbf{M}$, $\mathbf{A}$ and $\mathbf{R}$
$\mathbf{S} = \mathbf{MA}$
$\hat{\mathbf{Y}} = \mathbf{S} + \mathbf{R}$
while err $\geq$ tol **do**
  % Update outlier term $\mathbf{R}$

$$\mathbf{R} \leftarrow \mathbf{R}. \left[\frac{\mathbf{Y}.\hat{\mathbf{Y}}^{.(\beta-2)}}{\hat{\mathbf{Y}}^{.(\beta-1)} + \lambda \mathbf{R} \mathrm{diag}[\|\mathbf{r}_1\|_1, \dots, \|\mathbf{r}_P\|_1]^{-1}} \right]$$

$$\hat{\mathbf{Y}} \leftarrow \mathbf{S} + \mathbf{R}$$

  % Update abundances $\mathbf{A}$

$$\mathbf{A} \leftarrow \mathbf{A}. \frac{\mathbf{M}^T(\mathbf{Y}.\hat{\mathbf{Y}}^{.(\beta-2)}) + \mathbf{1}_{K,L}(\mathbf{S}.\hat{\mathbf{Y}}^{.(\beta-1)})}{\mathbf{M}^T(\hat{\mathbf{Y}}^{.(\beta-1)}) + \mathbf{1}_{K,L}(\mathbf{S}.\mathbf{Y}.\hat{\mathbf{Y}}^{.(\beta-2)})}$$

$$\mathbf{A} \leftarrow \mathbf{A} \mathrm{diag}[\|\mathbf{a}_1\|_1, \dots, \|\mathbf{a}_P\|_1]^{-1}$$

$$\mathbf{S} \leftarrow \mathbf{MA}$$

$$\hat{\mathbf{Y}} \leftarrow \mathbf{S} + \mathbf{R}$$

  % Update endmembers $\mathbf{M}$

$$\mathbf{M} \leftarrow \mathbf{M}. \left[\frac{(\mathbf{Y}.\hat{\mathbf{Y}}^{.(\beta-2)})\mathbf{A}^T}{(\hat{\mathbf{Y}}^{.(\beta-1)})\mathbf{A}^T} \right]$$

$$\mathbf{S} \leftarrow \mathbf{MA}$$

$$\hat{\mathbf{Y}} \leftarrow \mathbf{S} + \mathbf{R}$$

  Compute the objective function relative decrease 'err' (or any other convergence criterion).
end while

Let us now assume an unspecified independent prior model for $\mathbf{MA}$ but such that $\mathrm{E}[[\mathbf{MA}]_{lp}] = \rho$. Denoting by $\hat{\mu} = (LP)^{-1} \sum y_{lp}$ the empirical data expectation, our approach consists in matching $\hat{\mu}$ with $\mathrm{E}[[\mathbf{MA}]_{lp}] + \mathrm{E}[r_{lp}|\lambda]$, leading to

$$\hat{\lambda} = \frac{C}{\hat{\mu} - \rho}. \tag{28}$$

We insist that the latter expression only provides a handy gross estimate of λ that comes with no statistical guarantee (and that is sensitive to parametrization). In particular the estimate of λ is very dependent on ρ, the prior expectation of $[\mathbf{MA}]_{lp}$. However, because ρ is lower bounded by 0, the estimate of λ is lower bounded by $\lambda_0 = C/\hat{\mu}$, corresponding to a plausible minimum degree of sparsity. We used $\lambda = \lambda_0$ in the evaluations below and this was found to provide satisfactory results.

IV. EXPERIMENTS WITH SYNTHETIC DATA

In this section we evaluate the relevance of the rLMM proposed in Section II and the accuracy of the corresponding rNMF algorithm described in Section III using synthetic data. Since the main motivation for introducing the rLMM

is to introduce a flexible unmixing method able to analyze a large variety of remotely sensed scenes, these datasets have been generated following the main assumption underlying this model, i.e., assuming that most of the pixels result from linear combinations of the endmembers, as detailed below. Note that the proposed rLMM-based unmixing strategy may not be suitable for unmixing scenes which consists of only nonlinearly mixed data. However, such data is unusual and can generally be identified before any analysis. In such case, specific fully nonlinear unmixing strategies need to be invoked, which is out of the scope of the present paper.

A. Data Generation

Four 64×64-pixel images composed of $K = 3$ or 6 pure spectral components have been generated according to four different linear and nonlinear models. The endmember spectra have been extracted from the spectral library provided with the ENVI software [54]. The first image, denoted as $\mathcal{I}_{\mathrm{LMM}}$, is composed of pixels following the standard LMM (no nonlinear component)

$$\mathbf{y}_p = \sum_{k=1}^{K} a_{kp} \mathbf{m}_k + \mathbf{n}_p, \qquad (29)$$

where $\mathbf{n}_p$ denotes white additive Gaussian noise and $\mathbf{a}_p \in \mathbb{S}^K$. The three other images, denoted $\mathcal{I}_{\mathrm{FM}}$, $\mathcal{I}_{\mathrm{GBM}}$ and $\mathcal{I}_{\mathrm{MMP}}$ are generated as follows. 75% of the image pixels are generated according to the LMM in (29) and the remaining 25% is generated according to a model that features nonlinear component interactions. More precisely, the latter pixels are generated according to the following models (in which $\mathbf{m}_i \odot \mathbf{m}_j$ stands for the termwise Hadamard product):

- the Fan bilinear model (FM) [13]

$$\mathbf{y}_p = \sum_{k=1}^{K} a_{kp} \mathbf{m}_k + \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} a_{ip} a_{jp} \, \mathbf{m}_i \odot \mathbf{m}_j + \mathbf{n}_p,$$

  with $\mathbf{a}_p \in \mathbb{S}^K$,

- the generalized bilinear model (GBM) [19]

$$\mathbf{y}_p = \sum_{k=1}^{K} a_{kp} \mathbf{m}_k + \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} \gamma_{ijp} \, a_{ip} a_{jp} \, \mathbf{m}_i \odot \mathbf{m}_j + \mathbf{n}_p,$$

  with $\mathbf{a}_p \in \mathbb{S}^K$ and where the nonlinear coefficient $\gamma_{ijp} \in (0, 1)$ adjust the bilinear interaction between the ith and jth endmembers in the pth pixel,

- the multi-mixture pixel (MMP) model [23], which combines macroscopic and microscopic (i.e., intimate) mixtures

$$\mathbf{y}_p = \sum_{k=1}^{K} a_{kp} \mathbf{m}_k + a_{(K+1)p} \mathcal{R} \left(\sum_{k=1}^{K} f_{kp} \mathbf{w}_k \right) + \mathbf{n}_p,$$

  with $\mathbf{a}_p \in \mathbb{S}^{K+1}$ and where $\mathbf{f}_p = \left[f_{1p}, \ldots, f_{Kp} \right]^T$ are the microscopic proportions, $\{\mathbf{w}_k\}$ are the albedo-domain endmember signatures and $\mathcal{R}(\cdot)$ is the mapping function from the albedo domain to the reflectance domain chosen as in [23].

In a first experiment, the four images $\mathcal{I}_{\mathrm{LMM}}$, $\mathcal{I}_{\mathrm{FM}}$, $\mathcal{I}_{\mathrm{GBM}}$ and $\mathcal{I}_{\mathrm{MMP}}$ have been generated by sampling the abundance coefficients $\mathbf{a}_p$ randomly and uniformly over their admissible set $\mathbb{S}^K$ while the macroscopic abundances $\mathbf{a}_p$ and microscopic abundances $\mathbf{f}_p$ of the MMP model have been randomly and uniformly sampled over $\mathbb{S}^{K+1}$ and $\mathbb{S}^K$, respectively. In a second experiment, we wanted to evaluate the robustness of the algorithms w.r.t the absence of pure pixels in the images to be unmixed. To do so, we imposed a cutoff to the abundance coefficients that removes pure pixels from the observations. As such, in this case the abundances have been uniformly drawn over a truncated version of the set defined by (3), namely

$$\mathbb{S}_{0.9}^K = \left\{ \boldsymbol{a} \in \mathbb{R}^K \,\middle|\, 0 \le a_k \le 0.9, \ \sum_{k=1}^{K} a_k = 1 \right\}. \qquad (30)$$

Finally, in the two experiments the interaction coefficients γ_{ijp} appearing in the GBM have been uniformly drawn over the set $(0, 1)$ and the signal-to-noise ratio is set to SNR $= 40$dB, which is an admissible value for most of the real imaging spectrometers.

B. Compared Methods

The four images have been unmixed using rNMF and state-of-the-art algorithms specially designed for the considered models. For these models, the state-of-the-art algorithms are two-steps; the endmember matrix $\mathbf{M}$ is estimated in a first step, and then the abundance matrix $\mathbf{A}$ is estimated in a second step, given the endmember estimates (in a so-called "inversion" step). In contrast, rNMF performs a joint estimation of $\mathbf{M}$ and $\mathbf{A}$ (and $\mathbf{R}$).

We considered vertex component analysis (VCA) [55] coupled with fully constrained least squares (FCLS) [56]. VCA and FCLS are standard endmember extraction and inversion methods designed for the LMM. Besides, we considered the nonlinear endmember extraction technique proposed in [57], denoted as Heylen's algorithm in what follows, coupled with four different inversion methods designed for various nonlinear models, namely the FM, GBM and the very flexible polynomial post-nonlinear mixing model (PPNM) [19]. FM inversion is achieved with the algorithm detailed in [13], which exploits a first-order Taylor series expansion of the nonlinear term. GBM inversion is achieved with the gradient descent algorithm from [58]. PPNM inversion is achieved with the subgradient-based optimization scheme from [19]. Finally, the sparsity promoting bilinear unmixing algorithm, denoted BISPICE [59], has been also considered, initialized with either VCA or Heylen's endmember extraction.

The proposed rNMF has been applied with the SED, reflecting the Gaussian additive noise used in the data generation and for fair comparison with the other methods that rely on this assumption as well. We set $\lambda = \lambda_0$ and considered initializations by either VCA or Heylen's algorithm. Convergence was stopped when the relative difference between two successive values of the objective function fell under 10^{-5}.

TABLE I

ESTIMATION PERFORMANCE IN TERM OF aSAM (**M**) ($\times 10^{-3}$) AND GMSE (**A**) ($\times 10^{-3}$). BEST SCORES APPEAR IN BOLDFACE AND SECOND BEST SCORES APPEAR IN ITALIC. rNMF AND BISPICE ARE INITIALIZED BY EITHER VCA OR HEYLEN METHOD, AS STATED BETWEEN BRACKETS. REFER TO TEXT FOR OTHER DETAILS

			aSAM (**M**)						GMSE² (**A**)							
			VCA	Heylen	BISPICE		rNMF		VCA	Heylen			BISPICE		rNMF	
					(VCA)	(Heylen)	(VCA)	(Heylen)	+FCLS	+FM	+GBM	+PPNM	(VCA)	(Heylen)	(VCA)	(Heylen)
w/o pure pixels	K=3	$\mathcal{I}_{LMM}$	51.47	55.43	**13.16**	*13.16*	27.15	28.26	2.32	10.07	1.98	1.91	*0.20*	**0.16**	0.87	0.59
		$\mathcal{I}_{FM}$	48.75	78.35	67.43	67.43	**28.80**	*46.66*	3.21	14.36	3.51	3.52	20.50	20.50	**1.60**	*2.26*
		$\mathcal{I}_{GBM}$	47.45	55.95	33.26	33.26	**26.93**	*30.85*	2.52	8.70	1.87	1.75	10.00	10.02	*1.03*	**0.89**
		$\mathcal{I}_{MMP}$	48.35	41.29	**9.79**	*9.79*	26.13	20.07	7.62	26.00	5.86	5.90	*3.81*	**3.76**	5.55	4.79
	K=6	$\mathcal{I}_{LMM}$	84.89	90.80	*36.73*	36.73	**33.68**	83.70	4.47	35.77	8.60	7.49	**0.69**	*1.05*	1.13	2.64
		$\mathcal{I}_{FM}$	*80.06*	156.91	*115.20*	115.20	154.30	242.84	*10.11*	48.12	35.95	32.98	**8.04**	27.50	13.74	17.98
		$\mathcal{I}_{GBM}$	60.66	96.29	**45.58**	*45.58*	47.40	92.60	4.90	33.77	9.95	10.00	**3.12**	*3.40*	4.70	16.35
		$\mathcal{I}_{MMP}$	113.34	68.08	58.33	58.33	*46.70*	**43.75**	8.13	26.66	2.68	2.62	*1.61*	**0.82**	1.97	1.65
with pure pixels	K=3	$\mathcal{I}_{LMM}$	**6.19**	13.48	8.92	8.92	*6.19*	6.59	*0.03*	4.56	0.06	0.06	0.06	0.04	**0.03**	0.04
		$\mathcal{I}_{FM}$	10.46	14.78	61.96	61.96	**8.21**	*8.23*	0.78	4.24	*0.18*	**0.14**	20.45	20.45	0.69	0.66
		$\mathcal{I}_{GBM}$	8.79	14.48	26.15	26.15	**7.76**	*8.08*	0.26	3.33	**0.08**	*0.08*	10.46	10.47	0.22	0.25
		$\mathcal{I}_{MMP}$	6.96	13.80	7.24	7.24	*6.73*	**6.47**	4.16	20.42	4.22	4.20	*3.78*	**3.69**	4.17	4.15
	K=6	$\mathcal{I}_{LMM}$	54.69	70.04	**28.92**	*28.92*	34.67	34.86	1.54	28.08	2.80	2.64	**0.39**	*0.46*	0.72	0.93
		$\mathcal{I}_{FM}$	**76.70**	169.60	*118.47*	118.47	123.29	174.62	**7.17**	49.88	32.65	31.82	*7.98*	21.90	9.28	17.31
		$\mathcal{I}_{GBM}$	55.12	104.75	**41.14**	*41.14*	66.09	130.41	*3.13*	31.96	8.14	7.24	**2.62**	3.54	4.64	14.60
		$\mathcal{I}_{MMP}$	112.49	72.34	107.70	107.70	**26.48**	*32.81*	8.77	33.69	3.55	3.09	6.36	**0.69**	2.02	*1.55*

C. Performance Measures

The performance of the unmixing algorithms has been evaluated in terms of endmember estimation accuracy using the average spectral angle mapper (aSAM)

$$\mathrm{aSAM}\,(\mathbf{M}) = \frac{1}{K}\sum_{k=1}^{K}\mathrm{acos}\left(\frac{\langle \mathbf{m}_k, \hat{\mathbf{m}}_k \rangle}{\|\mathbf{m}_k\| \, \|\hat{\mathbf{m}}_k\|}\right)$$

and of abundance estimation accuracy using the global mean square error (GMSE)

$$\mathrm{GMSE}\,(\mathbf{A}) = \frac{1}{KP}\sum_{p=1}^{P}\left\|\mathbf{a}_p - \hat{\mathbf{a}}_p\right\|^2 .$$

D. Results and Discussion

The performance measures returned by the unmixing methods are reported in Table I. First, the aSAM values show that the proposed rNMF algorithm competes favorably with the two considered state-of-the-art endmember extraction algorithms, namely VCA and Heylen's algorithm, as well as with the MMP-based BISPICE unmixing algorithm. Initialized by these algorithms, the rLMM algorithm almost always improves the endmember estimation accuracy, with or without pure pixels. In particular, the proposed algorithm provides noticeable improvements when analyzing data resulting from the mixtures of $K = 3$ endmembers, while the expected gain seems less significant in case of 6-endmember mixtures. Similarly, when analyzing the GMSE related to abundance estimation, these results demonstrate the flexibility of the rLMM to model observations coming from various scenarios. More generally, these results demonstrate the ability of the rLMM-based unmixing technique to mitigate several kinds of nonlinear effects while preserving good estimation performance when analyzing only linear mixtures. Note that these results have been obtained with

a single realization of each dataset. However, no significant statistical difference in the unmixing performance has been observed with more realizations of the simulated datasets.

V. EXPERIMENTS WITH REAL DATA

In this section we apply rNMF to real hyperspectral datasets and discuss the results.

A. Description of the Datasets

We consider two real hyperspectral images that have been chosen because of availability of partial ground truth. The first image was acquired over Moffett Field, CA, in 1997, by the the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) [60]. Water absorption bands have been removed from the 224 spectral bands, leading to $L = 189$ spectral bands ranging from 0.4μm to 2.5μm with a nominal bandwidth of 10nm. The scene of interest, of size of 50×50 pixels, consists of a part of lake and a coastal area composed of soil and vegetation. This dataset has been previously studied in [15] and [61] and, thus, the unmixing results obtained in the current work can be compared to those reported in these later references. This dataset will be referred to as the "Moffett" image in the following and is depicted in Fig. 1(a).

The second considered dataset was acquired by the Hyspex hyperspectral scanner over Villelongue, France, in 2010. The sensed spectral domain consists of $L = 160$ spectral bands ranging from 0.4μm to 1.0μm. This image consists of a forested area where 12 vegetation species have been identified, during the Madonna project [62]. The sub-image of interest, of size of 50×50 pixels, is known to be mainly composed of oak and chestnut trees, with an additional unknown non-planted-tree endmember (referred to as Endm. #3 in what follows). This dataset will be referred to as the "Madonna" image in the following and is depicted in Fig. 1(b).
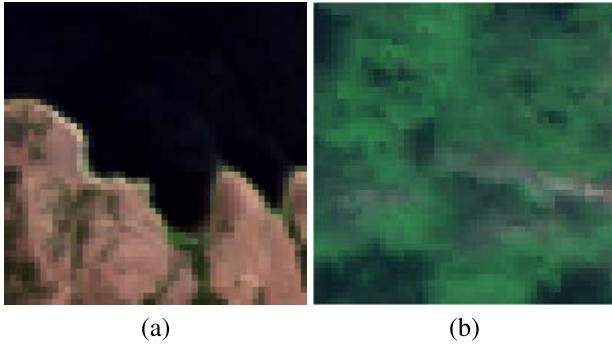
Fig. 1. Red-green-blue compositions of the real hyperspectral images. (a) Moffett. (b) Madonna.
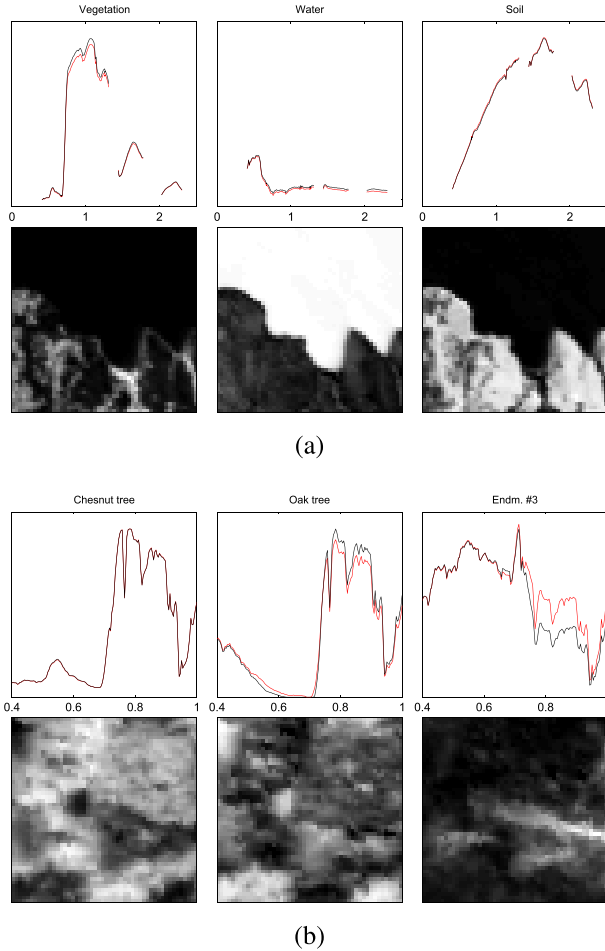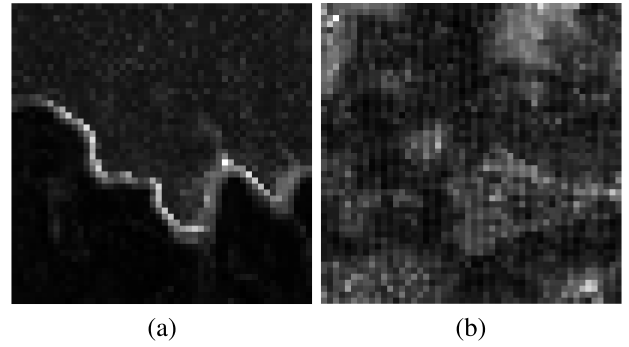

Fig. 3. Energy of the nonlinear components returned by rNMF under the KLD. Dark (resp. light) pixels correspond to small (resp. large) values. (a) Moffett. (b) Madonna.

The unmixing results are in good agreement with previous results [61], [63]. However, in addition to the standard description of the data by linearly mixed endmembers, the proposed model also provides information regarding the pixels that cannot be explained with the standard LMM. As such, Fig. 3 displays the energy $\mathbf{e} = \left[\|\mathbf{r}_1\|_2, \ldots, \|\mathbf{r}_P\|_2\right]$ of the residual outlier component estimated by rNMF. Regarding the Moffett image, the maps demonstrate that most of the pixels of this scene can be accurately described using the LMM. However, a few pixels, mainly located in the lake shore, appear at outliers. These pixels probably correspond to areas where some interactions between several endmembers occur (e.g., water/vegetation, water/soil). Similar results have been already observed in [15] and [64], which confirms the relevance of the proposed method. For the Madonna image, the energy map exhibits outlier terms that are mainly located in the area occupied by the oak trees and the unknown 3rd endmember. Furthermore, the image shows regular vertical patterns that are almost surely due to a sensor defect or miscalibration during the data post-processing.

Fig. 2. Unmixing results of two real hyperspectral images. Top of each image: endmembers estimated by the proposed rNMF-based unmixing algorithm under the KLD (red lines) and under the SED (black lines). Bottom of each image: estimated abundance maps obtained with the KLD; black (resp. white) pixels correspond to absence (resp. presence) of the associated endmembers. (a) Moffett. (b) Madonna.

B. Robust Unmixing Results

The proposed rLMM-unmixing technique has been applied on the real Moffett and Madonna images, with $K = 3$, with both the SED and the KLD. The endmember spectra and abundance maps estimated by rNMF are depicted in Fig. 2. For conciseness, only the abundance maps obtained with the KLD are displayed as the results for the SED are visually very similar.

VI. CONCLUSIONS AND FUTURE WORK

In this paper we have presented a new mixing model to describe hyperspectral data. This model, denoted as rLMM, extends the standard LMM by including an outlier residual term that can capture so-called nonlinear effects. These nonlinear effects are treated as additive and sparsely active outliers. In contrast with state-of-the-art literature on nonlinear hyperspectral unmixing, our approach does not require the specification of a particular model of nonlinearity.

The resulting unmixing problem was formulated as a new form of robust NMF problem, for which we developed a simple and effective block-coordinate descent algorithm that involves multiplicative updates. We provided an effective rule of thumb for setting the value of the penalty weight, which leaves our algorithm virtually free of parameters (only the number of endmembers needs to be specified). Simulations conducted on synthetic and real data have illustrated the relevance of rLMM, which outperformed many unmixing methods designed for various linear and nonlinear models.

We conclude this paper by sketching avenues for future work. Let us first emphasize that the methodology described in this paper is not specific to hyperspectral unmixing *per*

se and may be relevant to other fields where NMF is used. For instance, NMF with the KLD is popular in text retrieval, where the columns of $\mathbf{Y}$ are word counts from documents. The proposed model, which accounts for outliers, can be useful to such a setting. Additionally, the proposed model and methodology can be extended to other data-fidelity terms, such as the more general β-divergence [32], [65] (which takes the SED, the KLD and the Itakura-Saito divergence as special cases). Algorithmically, this only results in choosing β to be any real value instead of being either 2 or 1 in Eqs. (11), (20) & (26) – we explain this in details in the technical report [66]. The more general model has for example potential applications in music signal processing where the β-divergence is often used for spectrogram factorization [52], [67].

Secondly, another line of investigation concerns the relevance of lifting the nonnegative constraint on the outlier term $\mathbf{R}$. We made such an assumption in this paper because our primary motivation was to generalize existing nonlinear unmixing models, the majority of which assume the nonlinearity term to be nonnegative. Assuming $\mathbf{R}$ to be real-valued may model shadow effects that induce a decrease of reflectance. Whether such a model is efficient or not for hyperspectral unmixing requires substantial experimental validation that is left for future work. On the algorithmic side, it has to be noted that assuming $\mathbf{R}$ to be real-valued comes with some changes (for the worse or the better). In particular, for the divergences that are only defined for nonnegative numbers (such as the KLD), the approximate $\mathbf{MA} + \mathbf{R}$ needs to remain nonnegative and this is a challenge in itself. In the specific case of the SED, for which negative values can be tolerated, the update described by Eq. (13) can be used for $\mathbf{M}$ when $\mathbf{R}$ is real-valued. The update of $\mathbf{R}$ itself becomes easier as problem (14) has a closed-form solution in the real case [47]. The update of the abundances $\mathbf{A}$ requires more study as the positive-negative decomposition of the gradient used in Section III-C may not hold strictly anymore.

ACKNOWLEDGEMENTS

REFERENCES

[1] G. P. Asner and K. B. Heidebrecht, "Spectral unmixing of vegetation, soil and dry carbon cover in arid regions: Comparing multispectral and hyperspectral observations," *Int. J. Remote Sens.*, vol. 23, no. 19, pp. 3939–3958, Oct. 2002.

[2] K. E. Themelis, F. Schmidt, O. Sykioti, A. A. Rontogiannis, K. D. Koutroumbas, and I. A. Daglis, "On the unmixing of MEx/OMEGA hyperspectral data," *Planetary Space Sci.*, vol. 68, no. 1, pp. 34–41, 2012.

[3] A. A. Gowen, C. P. O'Donnell, P. J. Cullen, G. Downey, and J. M. Frias, "Hyperspectral imaging—An emerging process analytical tool for food quality and safety control," *Trends Food Sci. Technol.*, vol. 18, no. 12, pp. 590–598, 2007.

[4] N. Dobigeon and N. Brun, "Spectral mixture analysis of EELS spectrum-images," *Ultramicroscopy*, vol. 120, pp. 25–34, Sep. 2012.

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

[6] J. M. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observat. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[7] B. W. Hapke, "Bidirectional reflectance spectroscopy. I. Theory," *J. Geophys. Res.*, vol. 86, no. B4, pp. 3039–3054, Apr. 1981.

[8] J. F. Mustard and C. M. Pieters, "Photometric phase functions of common geologic minerals and applications to quantitative analysis of mineral mixture reflectance spectra," *J. Geophys. Res.*, vol. 94, no. B10, pp. 13619–13634, Oct. 1989.

[9] K. J. Guilfoyle, M. L. Althouse, and C.-I. Chang, "A quantitative and comparative analysis of linear and nonlinear spectral mixture models using radial basis function neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 10, pp. 2314–2318, Oct. 2001.

[10] J. Broadwater, R. Chellappa, A. Banerjee, and P. Burlina, "Kernel fully constrained least squares abundance estimates," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, Jul. 2007, pp. 4041–4044.

[11] J. Broadwater and A. Banerjee, "A comparison of kernel functions for intimate mixture models," in *Proc. 1st Workshop Hyperspectral Image Signal Process., Evol. Remote Sens. (WHISPERS)*, Aug. 2009, pp. 1–4.

[12] J. M. P. Nascimento and J. M. Bioucas-Dias, "Nonlinear mixture model for hyperspectral unmixing," *Proc. SPIE*, vol. 7477, p. 74770I, Sep. 2009.

[13] 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.

[14] B. Somers *et al.*, "Nonlinear hyperspectral mixture analysis for tree cover estimates in orchards," *Remote Sens. Environ.*, vol. 113, no. 6, pp. 1183–1193, Feb. 2009.

[15] 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.

[16] B. Somers, L. Tits, and P. Coppin, "Quantifying nonlinear spectral mixing in vegetated areas: Computer simulation model validation and first results," *IEEE J. Sel. Topics Appl. Earth Observat. Remote Sens.*, vol. 7, no. 6, pp. 1956–1965, Jun. 2014.

[17] I. Meganem, P. Déliot, 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.

[18] Y. Altmann, N. Dobigeon, and J.-Y. Tourneret, "Bilinear models for nonlinear unmixing of hyperspectral images," in *Proc. 3rd Workshop Hyperspectral Image Signal Process., Evol. Remote Sens. (WHISPERS)*, Lisbon, Portugal, Jun. 2011, pp. 1–4.

[19] 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.

[20] N. Dobigeon, L. Tits, B. Somers, Y. Altmann, and P. Coppin, "A comparison of nonlinear mixing models for vegetated areas using simulated and real hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observat. Remote Sens.*, vol. 7, no. 6, pp. 1869–1878, Jun. 2014.

[21] N. Dobigeon, J.-Y. Tourneret, C. Richard, J. C. M. Bermudez, S. McLaughlin, and A. O. Hero, "Nonlinear unmixing of hyperspectral images: Models and algorithms," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 82–94, Jan. 2014.

[22] R. Heylen, M. Parente, and P. Gader, "A review of nonlinear hyperspectral unmixing methods," *IEEE J. Sel. Topics Appl. Earth Observat. Remote Sens.*, vol. 7, no. 6, pp. 1844–1868, Jun. 2014.

[23] R. Close, P. Gader, J. Wilson, and A. Zare, "Using physics-based macroscopic and microscopic mixture models for hyperspectral pixel unmixing," *Proc. SPIE*, vol. 8390, p. 83901L, May 2012.

[24] N. Dobigeon and C. Févotte, "Robust nonnegative matrix factorization for nonlinear unmixing of hyperspectral images," in *Proc. IEEE Workshop Hyperspectral Image Signal Process., Evol. Remote Sens. (WHISPERS)*, Gainesville, FL, USA, Jun. 2013, pp. 1–4.

[25] R. B. Singer and T. B. McCord, "Mars: Large scale mixing of bright and dark surface materials and implications for analysis of spectral reflectance," in *Proc. 10th Lunar Planetary Sci. Conf.*, Mar. 1979, pp. 1835–1848.

[26] J. B. Adams, M. O. Smith, and P. E. Johnson, "Spectral mixture modeling: A new analysis of rock and soil types at the Viking Lander 1 site," *J. Geophys.*, vol. 91, no. B8, pp. 8098–8112, 1986.

[27] A. R. Huete, R. D. Jackson, and D. F. Post, "Spectral response of a plant canopy with different soil backgrounds," *Remote Sens. Environ.*, vol. 17, no. 1, pp. 37–53, Feb. 1985.

[28] T. M. Ray and B. C. Murray, "Nonlinear spectral mixing in desert vegetation," *Remote Sens. Environ.*, vol. 55, no. 1, pp. 59–64, 1996.

[29] L. Zhang, D. Li, Q. Tong, and L. Zheng, "Study of the spectral mixture model of soil and vegetation in Poyang Lake area, China," *Int. J. Remote Sens.*, vol. 19, no. 11, pp. 2077–2084, 1998.

[30] X. Chen and L. Vierling, "Spectral mixture analyses of hyperspectral data acquired using a tethered balloon," *Remote Sens. Environ.*, vol. 103, no. 3, pp. 338–350, Aug. 2006.

[31] S. Wang, N. Wang, D. Tao, L. Zhang, and B. Du, "A K-L divergence constrained sparse NMF for hyperspectral signal unmixing," in *Proc. IEEE Int. Conf. Secur., Pattern Anal., Cybern. (SPAC)*, Oct. 2014, pp. 223–228.

[32] C. Févotte and J. Idier, "Algorithms for nonnegative matrix factorization with the β-divergence," *Neural Comput.*, vol. 23, no. 9, pp. 2421–2456, Sep. 2011.

[33] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, pp. 1–37, 2009.

[34] P. Sprechmann, A. Bronstein, and G. Sapiro, "Real-time online singing voice separation from monaural recordings using robust low-rank modeling," in *Proc. Int. Soc. Music Inf. Retr. Conf. (ISMIR)*, Porto, Portugal, Oct. 2012, pp. 1–6.

[35] L. Zhang, Z. Chen, M. Zheng, and X. He, "Robust non-negative matrix factorization," *Frontiers Elect. Electron. Eng. China*, vol. 6, no. 2, pp. 192–200, 2011.

[36] B. Shen, L. Si, R. Ji, and B. Liu. (2012). "Robust nonnegative matrix factorization via ℓ₁ norm regularization." [Online]. Available: http://arxiv.org/abs/1204.2311/

[37] D. Kong, C. Ding, and H. Huang, "Robust nonnegative matrix factorization using ℓ₂₁-norm," in *Proc. 20th ACM Int. Conf. Inf. Knowl. Manage.*, 2011, pp. 673–682.

[38] A. B. Hamza and D. J. Brady, "Reconstruction of reflectance spectra using robust nonnegative matrix factorizations," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3637–3642, Sep. 2006.

[39] V. P. Pauca, J. Piper, and R. J. Plemmons, "Nonnegative matrix factorization for spectral data analysis," *Linear Algebra Appl.*, vol. 416, no. 1, pp. 29–47, 2006.

[40] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 3, pp. 765–777, Mar. 2007.

[41] Z. Yang, G. Zhou, S. Xie, S. Ding, J.-M. Yang, and J. Zhang, "Blind spectral unmixing based on sparse nonnegative matrix factorization," *IEEE Trans. Image Process.*, vol. 20, no. 4, pp. 1112–1125, Apr. 2011.

[42] 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 Trans. Image Process.*, vol. 21, no. 7, pp. 3239–3252, Jul. 2012.

[43] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *Amer. Statist.*, vol. 58, no. 1, pp. 30–37, 2004.

[44] A. R. De Pierro, "On the relation between the ISRA and the EM algorithm for positron emission tomography," *IEEE Trans. Med. Imag.*, vol. 12, no. 2, pp. 328–333, Jun. 1993.

[45] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2001, pp. 556–562.

[46] F. Sha, Y. Lin, L. K. Saul, and D. D. Lee, "Multiplicative updates for nonnegative quadratic programming," *Neural Comput.*, vol. 19, no. 8, pp. 2004–2031, 2007.

[47] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Roy. Statist. Soc. B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2006.

[48] Z. Yang and E. Oja, "Unified development of multiplicative algorithms for linear and quadratic nonnegative matrix factorization," *IEEE Trans. Neural Netw.*, vol. 22, no. 12, pp. 1878–1891, Dec. 2011.

[49] V. Y. F. Tan and C. Févotte, "Automatic relevance determination in nonnegative matrix factorization with the β-divergence," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 7, pp. 1592–1605, Jul. 2013.

[50] J. Eggert and E. Körner, "Sparse coding and NMF," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Jul. 2004, pp. 2529–2533.

[51] T. Virtanen, "Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria," *IEEE Trans. Audio, Speech, Language Process.*, vol. 15, no. 3, pp. 1066–1074, Mar. 2007.

[52] C. Févotte, N. Bertin, and J.-L. Durrieu, "Nonnegative matrix factorization with the Itakura–Saito divergence: With application to music analysis," *Neural Comput.*, vol. 21, no. 3, pp. 793–830, Mar. 2009.

[53] A. Lee, F. Caron, A. Doucet, and C. Holmes. (Sep. 2010). "A hierarchical Bayesian framework for constructing sparsity-inducing priors." [Online]. Available: http://arxiv.org/abs/1009.1914

[54] *ENVI User's Guide Version 4.0*, Research Systems Inc., Boulder, CO, USA, Sep. 2003.

[55] 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.

[56] 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. 39, no. 3, pp. 529–545, Mar. 2001.

[57] R. Heylen, D. Burazerovic, and P. Scheunders, "Non-linear spectral unmixing by geodesic simplex volume maximization," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 534–542, Jun. 2011.

[58] 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. (IGARSS)*, Vancouver, BC, Canada, Jul. 2011, pp. 1886–1889.

[59] P. Gader, D. Dranishnikov, A. Zare, and J. Chanussot, "A sparsity promoting bilinear unmixing model," in *Proc. 4th Workshop Hyperspectral Image Signal Process., Evol. Remote Sens. (WHISPERS)*, Shanghai, China, Jun. 2012, pp. 1–4.

[60] Jet Propulsion Lab., California Institute of Technology, Pasadena, CA, USA. (2006). *AVIRIS Free Data*. [Online]. Available: http://aviris.jpl.nasa.gov/html/aviris.freedata.html

[61] 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.

[62] 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. (IGARSS)*, Vancouver, BC, Canada, Jul. 2011, pp. 3672–3675.

[63] Y. Altmann, N. Dobigeon, S. McLaughlin, and J.-Y. Tourneret, "Nonlinear spectral unmixing of hyperspectral images using Gaussian processes," *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2442–2453, May 2013.

[64] O. Besson, N. Dobigeon, and J.-Y. Tourneret, "Minimum mean square distance estimation of a subspace," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 5709–5720, Dec. 2011.

[65] A. Cichocki and S. Amari, "Families of alpha- beta- and gamma-divergences: Flexible and robust measures of similarities," *Entropy*, vol. 12, no. 6, pp. 1532–1568, Jun. 2010.

[66] C. Févotte and N. Dobigeon. (Jan. 2014). "Nonlinear hyperspectral unmixing with robust nonnegative matrix factorization." [Online]. Available: http://arxiv.org/abs/1401.5649

[67] E. Vincent, N. Bertin, and R. Badeau, "Adaptive harmonic spectral decomposition for multiple pitch estimation," *IEEE Trans. Audio, Speech, Language Process.*, vol. 18, no. 3, pp. 528–537, Mar. 2010.

Cédric Févotte received the Ph.D. degree from the Ecole Centrale de Nantes. He held a post-doctoral position with the University of Cambridge from 2003 to 2006. He was a CNRS Researcher with Télécom ParisTech from 2007 to 2013, and a Research Engineer with Mist-Technologies/Audionamix in 2006. He is currently a CNRS Senior Researcher with the Laboratoire Lagrange, Nice, France. His research interests concern statistical signal processing and machine learning for inverse problems and source separation. In 2014, he was a co-recipient of the IEEE Signal Processing Society Best Paper Award. He serves on the Editorial Board of the IEEE TRANSACTIONS ON SIGNAL PROCESSING and on the IEEE Machine Learning for Signal Processing Technical Committee.

Nicolas Dobigeon received the Ph.D. degree in signal processing from the INP Toulouse, in 2007. He was a Post-Doctoral Research Associate with the University of Michigan, Ann Arbor, MI, USA, from 2007 to 2008. Since 2008, he has been with the National Polytechnic Institute of Toulouse, University of Toulouse, where he is currently an Associate Professor. His current research interests include statistical signal and image processing, with a particular interest in Bayesian inverse problems with applications to remote sensing, biomedical imaging, and genomics.