Chapter 6

Linear and Nonlinear Unmixing in Hyperspectral Imaging

N. Dobigeon†*, Y. Altmann†, N. Brun‡ and S. Moussaoui§

† University of Toulouse, IRIT/INP-ENSEEIHT, 31071 Toulouse Cedex 7, France
‡ School of Engineering and Physical Sciences, Heriot-Watt University, Riccarton, Edinburgh, EH14 4AS, United Kingdom
§ Laboratoire de Physique des Solides, CNRS, Univ. Paris-Sud, Univ. Paris-Saclay, 91405 Orsay Cedex, France

1 Corresponding author: e-mail: Nicolas.Dobigeon@enseeiht.fr

Chapter Outline

1 Spectral Mixture Analysis 185
2 Linear Unmixing 190
   2.1 The Linear Mixing Model (LMM) 190
   2.2 Endmember Extraction Algorithms 192
   2.3 Inversion Algorithms 192
   2.4 Unsupervised SU: Joint Extraction and Inversion 194
3 Nonlinear Unmixing 195
   3.1 Intimate Mixture Models 197
4 Experiments: Unmixing EELS Data 207
5 Conclusion 212
References 214

1 SPECTRAL MIXTURE ANALYSIS

For several decades, development of spectral imaging has significantly contributed to numerous advances in various applicative fields, not only remote sensing and earth observation [1] but also planetology [2–4], food monitoring [5–7] and biomedical engineering [8, 9]. It consists of observing a scene of interest at various wavelengths and, contrary to conventional red-green-blue imaging, it generally offers several hundreds of spectral bands for each spatial measurement. This high spectral resolution confers hyperspectral imaging a powerful ability to discriminate distinct materials that are present in the observed area. The general concept of hyperspectral imaging as a remote sensing tool is illustrated in Fig. 1. In this specific application context, such
measurement technique has shown to be a valuable modality to solve challenging issues, such as material quantification [10], anomaly detection [11, 12] and pixel classification [13–15], avoiding tedious on-ground sampling campaigns during mineralogic surveys.

At a completely different scale, hyperspectral images can be obtained by microscopy imaging techniques. Two examples are Raman microspectroscopy with a pixel size down to about 1μm and electron energy loss spectroscopy (EELS) performed in a scanning transmission electron microscope (STEM) with a sub-nanometer resolution. Measurements collected by these techniques may also be in the form of 3d-datacubes, characterized by a 2d-spatial sampling and a 1d-spectral sampling. As illustrated in Fig. 2, per se, spectroscopic imaging also provides hyperspectral data that share numerous similarities with remotely sensed images, e.g., resulting from a push-broom scanning system. Notably, it has demonstrated its value to address various issues, e.g., for material science and chemical analysis [16–18].

Unfortunately, due to an intrinsic trade-off between spatial and spectral sensitivities of the sensors, conventional hyperspectral images generally suffer from limited spatial resolution when compared to multispectral or
In particular, when they are acquired at a spatial resolution that is higher than the resolution of the sensor (i.e., scanner, microscope), these images are corrupted by blurring effects related to the optical response of the instrument (point spread function, PSF). To deal with this super-resolution problem, deconvolution techniques can be advocated to restore the observed data [20]. Such methods require a previous processing step for PSF estimation. When a complementary image of higher spatial resolution but lower spectral resolution is acquired simultaneously, reconstructing a high spatial and high spectral resolution image can also overcome these limitations. Various multi-band reconstruction techniques, known as hyperspectral pansharpening [21] or multi-band image fusion [22, 23], have shown promising results. However, when such complementary data of high spatial resolution is not available, alternative processing approaches should be considered. Besides, even if these blurring effects are neglected, several macroscopic materials (e.g., vegetation, minerals, and human made construction) usually contribute to the spectrum measured at each given single pixel of the hyperspectral image. This more general restriction may question the operational use of these images when an accurate mapping of the scene components is required.

Specifically, spectral unmixing (SU), or spectral mixture analysis, aims at providing comprehensive descriptions of the hyperspectral measurements. It consists of extracting the spectral signatures that are characteristics of the main materials present in the scene and quantifying their respective spatial distribution (abundance) over the image [24, 25]. More precisely, SU decomposes the \( P \) spectra \( y_1, \ldots, y_P \) related to the hyperspectral image pixels with \( y_p = [y_{1,p}, \ldots, y_{L,p}]^T \in \mathbb{R}^L \) (where \( L \) stands for the number of spectral bands) into a collection of \( R \) elementary spectra \( \mathbf{m}_1, \ldots, \mathbf{m}_R \), and a set of \( P \) abundance vectors \( a_1, \ldots, a_P \), with \( a_p = [a_{1,p}, \ldots, a_{R,p}]^T \) [24, 25]. In the most common...
acception of SU, the spectral signatures $\mathbf{m}_r (r = 1, \ldots, R)$, usually referred to as endmembers, are assumed to represent the pure materials present in the observed scene and the abundance coefficient $a_{r,p}$ quantifies the fractional proportion of the $r$th material in the $p$th pixel.

At this stage of the SU description, one should highlight the fact that the concept of endmember and its definition highly depend on the target application or, at least, the considered problem (see the Chapter by Somers [26] of this book). For instance, in a remote sensing scenario, when observing a vegetated area composed of two distinct tree species, characterizing these species with two endmember signatures may be sufficient to assess the presence of these plants over the observed image. However, when the foliage composition is of prime interest, e.g., for ecological purpose, endmember definition needs to be refined, i.e., by considering a set of two spectral signatures associated with leaves and branches, respectively, for each tree species.

Another difficulty appears when properly defining the abundance coefficients. Generally, they are expected to quantify the presence of the endmembers in the observed pixels. In the most favorable case, i.e., when the mixing model is assumed to be linear (see Section 2), the abundances represent the relative areas occupied by the endmembers in the pixels [27]. Governed by this finding that leads to easier physical interpretations of SU, most unmixing methods incorporate non-negativity and additivity (i.e., sum-to-one) constraints imposed on the abundance coefficients

$$a_{r,p} \geq 0 \quad \forall r, \forall p$$

$$\sum_{r=1}^{R} a_{r,p} = 1 \quad \forall p.$$  

Under such modeling, abundance coefficients are associated with concentration, proportion or even probability of occurrence. In particular these constraints aim at providing a full (i.e., complete) and additive part-based representation of the measurements [24, 25]. However, in more complex acquisition scenario, i.e., when the mixing model departs from the linear one, other interpretations of these abundance coefficients can be advocated, such as fractional masses of the material in the considered mixture. Depending on the observation conditions (e.g., illumination) and the nature of the underlying scene, the sum-to-one constraint might not be appropriate and may be relaxed.

After properly defining the concepts of endmembers and abundances for the considered study framework, and neglecting any side-effects, e.g., induced by the spatial convolution due to the instrument PSF, relating the observed measurement $\mathbf{y}_p$ to the explanatory variables, namely the endmember matrix $\mathbf{M} \equiv [\mathbf{m}_1, \ldots, \mathbf{m}_R]$ and the abundance vector $\mathbf{a}_p$, can be pixel-wise stated as
In Eq. (3), the approximation symbol is due to measurement noises and to errors resulting from the model underlying the function \( f(\cdot, \cdot) \). Thus, setting SU in equation straightforwardly consists of solving the following minimization problem

\[
\min_{\mathbf{M}, \mathbf{A}} \sum_{p=1}^{P} D_p[\mathbf{y}_p|f(\mathbf{M}, \mathbf{a}_p)] \quad \text{s.t. (1)}
\]

where \( \mathbf{A} = [\mathbf{a}_p, \ldots, \mathbf{a}_P] \) and \( D_p(\cdot) \) measures the discrepancy between the observed measurement \( \mathbf{y}_p \) and the predicted spectrum model \( f(\mathbf{M}, \mathbf{a}_p) \). In most research works dealing with SU, the approximation in Eq. (3) takes the form of an additive noise, generally assumed to be pixel independent, centered and normally distributed with covariance matrix \( \Sigma \). Under this widely admitted hypotheses, adopting an optimization point of view leads to the solving of a weighted least squares problem. The discrepancy measure becomes quadratic

\[
D_p[\mathbf{u}, \mathbf{v} | f(\cdot, \cdot)] = ||\mathbf{u} - \mathbf{v}||^2_{\Sigma}, \quad \forall p
\]

where \( ||\mathbf{u}||^2_{\Sigma} = \mathbf{u}^T \Sigma^{-1} \mathbf{u} \) is the Mahalanobis norm induced by \( \Sigma \), which reduces to the Euclidean norm

\[
D_p[\mathbf{u}, \mathbf{v}] = ||\mathbf{u} - \mathbf{v}||^2, \quad \forall p
\]

when the noise sequence is additionally assumed to be band-independent. As stated by Févotte and Dobigeon [28], the Euclidean distance underlying the presence of additive Gaussian noise is the more commonly used discrepancy measure advocated in hyperspectral unmixing. However, it is worthy to note that other measures of divergence can be considered [29]. More generally, within a statistical estimation framework, the discrepancy measure \( D_p(\cdot) \) could derive from the definition of a statistical risk and, in this case, it would be defined as an expectation over random processes [30].

For a long time, SU has been decomposed into two successive processing steps. In the first step, the endmember spectra are identified. This identification can result from \textit{a priori} knowledge, by choosing these signatures from a spectral library or by picking up the spectral signatures associated with specific image pixels which are assumed to be \textit{pure}, i.e., composed of only one endmember. When this prior knowledge is not available, one should resort to a so-called \textit{endmember extraction algorithm} (EEA) able to identify the endmember signatures from the measurements in an unsupervised manner [31]. Subsequently, once these endmembers have been identified, the problem in Eq. (4) reduces to the so-called \textit{inversion} step which consists of estimating the abundance vectors in each pixel by solving for \( p = 1, \ldots, P \)

\[
\min_{\mathbf{a}_p} D_p[\mathbf{y}_p|f(\mathbf{M}, \mathbf{a}_p)] \quad \text{s.t. (1)}.
\]
Whatever the deterministic or statistical framework adopted to tackle the unmixing problem, a crucial issue results from the choice of the modeling function \( f(\cdot, \cdot) \) that relates the observed measurement \( y_p \) to the parameters of interest, namely \( M \) and \( a_p \). The choice for this function generally results from approximate descriptions of the physical processes underlying the acquisition process. The linear mixing model (LMM) has been certainly the most widely used when unmixing hyperspectral data, not only because of its simplicity but also since it provides an acceptable first-order approximation of the mixing process. To address scenarios where the LMM shows to be not sufficient to properly describe the measurements, researchers have developed other nonlinear models able to take various nuisance effects into account. Besides, when accurate approximate physics-based models are unmanageable, i.e., when they lead to a problem formulated by Eq. (4) too difficult to solve, more versatile analytical models, yet not motivated by physical considerations, can be derived. This chapter draws a wide albeit partial panorama of these models, as well as the main corresponding unmixing algorithms available in the literature. It is worth noting that most of the MATLAB codes associated with these algorithms are available online on the first author’s webpage.

2 LINEAR UNMIXING

2.1 The Linear Mixing Model (LMM)

As noticed in Bioucas-Dias et al. [25] and Keshava and Mustard [24], most unmixing algorithms rely on the assumption of a linear mixing process to explain the observed spectra. More precisely, under this linear mixing model (LMM), each observed spectrum in each pixel of a given image is assumed to result from the linear combination of the \( R \) endmember spectra

\[
y_p^{(\text{LMM})} \approx \sum_{r=1}^{R} a_{r,p} m_r \approx M a_p
\]  

(7)

It is worthy to note that, in some sense, following the interpretation proposed by Kent and Mardia [32], SU can be cast as a soft multi-class classification since each measurement \( y_p \) is quantitatively and linearly associated with the endmembers through the pairs \( \{m_1, a_{1,p}\}, \ldots, \{m_R, a_{R,p}\} \).

This LMM has received considerable attention since it generally consists of an acceptable first-order approximation of the physical processes involved in most of the scenes of interest [24]. Consequently, it has motivated a lot of research works, mainly in the remote sensing and image processing literature, that aim at developing efficient EEAs, able to recover pure component

1. http://dobigeon.perso.enseeiht.fr/
signatures in the image, and inversion techniques to estimate the abundance coefficients for a given (estimated or *a priori* known) set of endmembers. Comprehensive overviews of these EEAs and inversion methods can be found in [25], [24], [33], and in the Chapter by Somers [26] of this book.

Interestingly, the earlier studies dedicated to linear SU were founded on geometrical interpretation of the problem. Indeed, under the widely admitted non-negativity and sum-to-one constraints in Eq. (1) imposed on the abundance vectors $a_1, \ldots, a_P$, in the free-noise case, i.e., when perfect equality holds in Eq. (7), the observed measurements lie in the $R$-polytope whose convex hull is formed by the endmember spectral signatures. More precisely, all the pixels $y_p \in \mathbb{R}^L (p = 1, \ldots, P)$ belong to the $(R - 1)$-simplex whose vertices are the endmembers. Fig. 3 illustrates this geometrical representation of the LMM when considering mixtures of $R = 3$ endmembers. When the sum-to-one constraint does not hold, the observed measurements belong to the positive cone defined by the endmember spectral signatures and pre-processing can be used to project the data onto a particular simplex within this cone [25].

This finding clearly shows that the number of endmembers is intrinsically related to the dimension of the subspace actually occupied by the observed measurements. Thus, identifying this signal subspace of interest appears as a relevant pre-processing step. It allows not only the number of endmembers actually involved in the mixtures to be estimated, but also opens the door to straightforward denoising techniques by projecting these measurements onto this subspace of interest. This signal subspace identification problem can be tackled by extracting a subset of the bands chosen for their *a priori* known high signal-to-noise ratios, as in [34]. Conversely, more sophisticated methods attempt to identify the most appropriate subspace to represent the measurements and mainly differ by the criterion designed to measure this appropriateness. Principal component analysis (PCA) [35], minimum noise

![FIG. 3 Schematic representation (in $\mathbb{R}^3$) of the LMM for $R = 3$ endmembers. All the observed pixels (gray points) belong to a 2-simplex (red (gray in the print version) lines) whose vertices are the endmembers (blue (gray in the print version) points).](image-url)
fraction (MNF) [36] and hyperspectral signal identification by minimum error (HySime) [37] are the most widely used subspace identification techniques.

2.2 Endmember Extraction Algorithms

In a first step of SU, the macroscopic materials that are present in the image are estimated using an EEA. The very first automated EEAs have exploited the geometrical interpretation of SU introduced in the previous section. Since the measurements are known to belong to the simplex formed by the endmembers, these pure spectral signatures can be equivalently recovered by identifying this simplex and, more precisely, its vertices. The most popular EEAs include pixel purity index (PPI) [38], N-FINDR [39], and more recently the vertex component analysis [40]. They propose to recover the vertices of the simplex of maximum volume where image pixels are potential candidate for these vertices. An assumption common to these EEAs is that they require the presence of pure pixels in the observed image, since the closer to the simplex the candidates, the better endmember estimates. As noticed by Ma et al. [41], a similar technique of pure pixel pursuit has been independently proposed by Araújo et al. [42] to analyze spectroscopic multicomponents in chemometrics. Other simplex volume maximization algorithms include the schemes detailed by Chan et al. [43] and Wu et al. [44]. Conversely, a reciprocal class of EEAs consists of recovering the smallest simplex that contains all the dataset [45–47]. The major advantage of these strategies is that they do not require the prior assumption of pure pixels in the measurements, in contrast to most of the simplex volume maximization techniques. Robust counterparts of these simplex volume minimization or maximization techniques include [48–50], e.g., to face with noise and/or anomaly corruption.

2.3 Inversion Algorithms

2.3.1 Supervised SU

Once the endmember signatures gathered in the matrix $\mathbf{M}$ have been identified, the second step of SU, referred to as supervised SU or inversion, consists of estimating the set of abundances (or abundance vectors) associated with each measurement. Basically, the initial formulation of SU in Eq. (4) turns out to the following $P$ minimization problems

$$\min_{a_p} D_p(\mathbf{y}_p|\mathbf{Ma}_p) \quad \text{s.t. (1)}.$$  \hspace{1cm} (8)

When the discrepancy measure $D(\cdot|\cdot)$ is chosen as the standard Euclidean distance, the inversion reduces to the constrained least-square (LS) regression

$$\min_{a_p} ||\mathbf{y}_p - \mathbf{Ma}_p||_2^2 \quad \text{s.t. (1)}. $$  \hspace{1cm} (9)

Data Handling in Science and Technology
One difficulty arising from these optimization problems results from the constraints that the abundances should satisfy, which makes impossible any derivation of closed-form solutions. As a consequence, various algorithmic strategies have been proposed to solve the constrained optimization problems in Eq. (9). One may cite the fully constrained least-square (FCLS) algorithm proposed by Heinz and Chang [51] or the detection-oriented technique by Tu et al. [52]. This problem still motivated recent research works, e.g., to ensure the abundance constraints are properly satisfied [53], to reach better convergence guarantees of the algorithm [54, 55] or to cope with the possible high dimension of the problem when unmixing large hyperspectral images [56]. Statistical counterparts of these inversion methods consider the regression task as an estimation problem [57]. Such approach has been for instance considered to unmix hyperspectral data acquired by the MEx/OMEGA instrument over the Mars surface [58]. One major advantage of Bayesian techniques lies in the possibility of including additional nuisance parameters (e.g., noise variance) into the statistical model and estimating them jointly with the abundance vectors [59]. Moreover, in such framework, inference can generally be solved using Markov chain Monte Carlo (MCMC) techniques [60, 61], which allows various Bayesian estimators of the abundance vectors to be computed, e.g., the maximum a posteriori (MAP) and minimum mean square error (MMSE) estimators.

2.3.2 Semi-supervised SU

The inversion step has been also considered to deal with the specific scenario where the endmember spectra \( \mathbf{m}_1, \ldots, \mathbf{m}_R \) are not perfectly identified but known to belong to an a priori available spectral library. This spectral library, also referred to as a dictionary and denoted as \( \mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_K] \), is composed of \( K \) spectral signatures assumed to be acquired in the same conditions as the measurements to be unmixed. Note that the number \( K \) of spectral signatures in the library is much greater than the number \( R \) of endmembers actually involved in the observed mixtures. Thus, in this semi-supervised framework, SU consists of extracting from \( \mathbf{S} \) a subset of signatures that can explain the measurements according to the model in Eq. (7). To cope with the large dimension of the dictionary \( \mathbf{S} (K \gg R) \), the problem in Eq. (8) is slightly modified to include a sparsity constraint or penalization, while at the same time the sum-to-one constraint is relaxed. Under the hypothesis of quadratic cost function, semi-supervised SU consists of solving the following sparse LS regression problems

\[
\min_{\mathbf{a}_p} \| \mathbf{y}_p - \mathbf{Sa}_p \|_2^2 \quad \text{s.t.} \quad \| \mathbf{a}_p \|_0 \leq d \quad \text{and} \quad a_{r,p} \geq 0 \quad \forall r
\]

or alternatively

\[
\min_{\mathbf{a}_p} \| \mathbf{a}_p \|_0 \quad \text{s.t.} \quad \| \mathbf{y}_p - \mathbf{Sa}_p \|_2^2 \leq \epsilon \quad \text{and} \quad a_{r,p} \geq 0 \quad \forall r.
\]
In Eq. (10), \( d \) represents the maximum number of non-zero coefficients in \( a_p \) quantified by the \( \ell_0 \)-semi norm (or counting function) \( \| a_p \|_0 \) and thus the maximum number of endmembers allowed to be involved in the considered mixture. Conversely, in Eq. (11), \( \epsilon \) quantifies the tolerated reconstruction error resulting from any measurement noise or other modeling approximation. Solving such NP-hard problems requires the use of greedy algorithms, such as matching pursuit [62] and orthogonal matching pursuit [63]. Several variational formulations based on a convex relaxation of the counting function \( \| \cdot \|_0 \), which is merely replaced by the sparsity-inducing \( \ell_1 \)-norm, have been proposed in the literature, such as the famous SUNSAL algorithm proposed by Bioucas-Dias and Figueiredo [64]. Comprehensive analysis of ability of such algorithms to solve semi-supervised SU has been conducted by Iordache et al. [65]. Again, Bayesian counterparts of these semi-supervised unmixing techniques have been proposed by Themelis et al. [66] and Dobigeon et al. [61].

### 2.4 Unsupervised SU: Joint Extraction and Inversion

Addressing the SU problem by considering two successive steps may appear as suboptimal, e.g., as the accuracy of the inversion step highly depends on the quality of the endmembers identified by the EEA. To overcome this limitation, some approaches address the SU problem by estimating the endmember signatures \( m_1, \ldots, m_K \) and all the abundance vectors \( a_1, \ldots, a_P \) jointly, i.e., in a single step. In this so-called unsupervised framework, under the LMM assumption, SU in Eq. (3) formulated for all the pixels consists of estimating the endmember \( M \) and abundance \( A \triangleq [a_1, \ldots, a_P] \) matrices given the observation matrix \( Y = [y_1, \ldots, y_P] \) such that

\[
Y \approx MA
\]  

under the set of abundance constraints in Eq. (1). Since the endmember spectra generally correspond to, e.g., radiance, reflectance, absorbance or other energy measures, they can be assumed to be non-negative and thus an additional set of non-negativity constraints is included into the model, namely

\[
m_{l,r} \geq 0, \ \forall l, \ \forall r.
\]  

As a consequence, SU can be cast as a constrained blind source separation (BSS). BSS is a well-known problem, mainly addressed in the signal processing literature for various applications, for instance audio (i.e., music and speech) analysis [67, 68]. The most popular latent variable model-based separation techniques to perform BSS include principal component analysis (PCA) and independent component analysis (ICA). For instance, PCA has been conducted to extract material signatures from EELS spectrum-image by Bonnet et al. [69] or Bosman et al. [70]. However, as noticed by Arenal et al. [71], PCA extracts component signatures that are difficult to be interpreted physically, mainly due to the inherent orthogonality constraint imposed...
by PCA to the components. Naturally, ICA has been considered as an alternative to analysis hyperspectral images acquired from, e.g., remote sensors [72, 73] or STEM-EELS [74, 75]. However, these ICA-based approaches can fail to provide meaningful interpretation of the identified signatures, especially in applications where both these signatures, as well as the abundance maps, are highly correlated, breaking the required independence assumption [76]. Besides, in their standard formulations, both PCA and ICA do not allow the constraints (1) to be satisfied. As a consequence, the end-users, from geoscience or experimental physics communities, should seize other advanced methodological tools able to offer a comprehensive and interpretable description of the hyperspectral data.

Other promising pathways can be followed once the problem of Eq. (12) has been rather interpreted as a non-negative matrix factorization (NMF). Such latent variable models only impose non-negativity constraints on the elements of the two model factors \( A \) and \( M \), an assumption that is widely satisfied by hyperspectral data for most modalities. Since the pioneering works of Lee and Seung [77] and Paatero and Tapper [78], NMF has demonstrated its flexibility and its efficiency to infer relevant factors in various applications. Its potential was materialized by Miao and Qi [79] who combined such approach with geometrical concepts to develop, up to authors’ knowledge, the first fully unsupervised NMF-based unmixing algorithm. Capitalizing on this advance, numerous works have been achieved, mainly by the remote sensing community, to design more efficient NMF-based unmixing algorithms [80–83]. As for the supervised and semi-supervised frameworks, even more flexibility can be reached by tackling the NMF task as a statistical inference problem. Efficient Bayesian source separation algorithms accounting for non-negativity constraints of the source signals and the mixing coefficients have been proposed to analyze spectral mixtures [84] and, in particular, hyperspectral images in various applicative fields, for instance remote sensing [85], food monitoring [86], planetology [87, 88], Raman [89], and near-infrared (NIR) spectroscopy [90]. To cope with the somehow high computational complexity of the MCMC-based inference algorithms, specific algorithmic improvements may be required [93]. Interested readers are invited to refer to as the Chapter by Brie [94] of this book for a comprehensive description of Bayesian NMF.

3 NONLINEAR UNMIXING

According to Bioucas-Dias et al. [25], LMM consists of a reasonable model once two assumptions are satisfied. Firstly, the mixing process should occur at a macroscopic scale [95]. Secondly, the photons that reach the hyperspectral sensor must interact with only one material, as is the case in checkerboard type scenes [96]. A schematic illustration of this model is depicted in Fig. 4 (left) for a scene composed of two materials. When one of these two
FIG. 4  *Left*, linear mixing model (LMM): the imaged pixel is composed of two materials. *Middle*, intimate mixture model (IMM): the imaged pixel is composed of a microscopic mixture of several constituents. *Right*, bilinear mixing model (BMM): the imaged pixel is composed of two endmembers, namely tree and soil. In addition to the individual contribution of each material, nonlinear (e.g., bilinear) interactions between the tree and the soil reach the sensor (from Dobigeon N, Tourneret FY, Richard C, Bermudez JCM, McLaughlin S, Hero AO, *Nonlinear unmixing of hyperspectral images: Models and algorithms*. IEEE Signal Process Mag 2014;31(1):89-94).
assumptions is not satisfied, the LMM may be not valid since various nonlinear effects may occur. To cope with these nonlinear interactions, more complex mixing models have been proposed, mainly in the spectroscopy and remote sensing literature. These nonlinear models have shown interesting properties to unveil meaningful information that would have been unattainable for the standard LMM [97–104].

To describe the light scattering by the materials in an observed scene, an accurate model based on radiative transfer (RT) theory has been advocated by Hapke [105]. Unfortunately, deriving SU algorithms exploiting this model would require to solve too complex problems. In particular, this model assumes linear spectral mixtures (subject to non-negativity and sum-to-one constraints) in the radiance domain instead of in the reflectance domain. To overcome these difficulties, several approximations have been proposed, mainly in the spectroscopy literature, such as the famous Hapke’s bidirectional model [27]. However, these models still require highly non-linear and integral formulations that preclude any real-world implementations of SU techniques. Thus, no unsupervised nonlinear unmixing algorithm able to jointly extract the endmembers from the data and estimate their relative proportions in the pixels have been proposed to handle these models.

More recently, alternate approximate but workable nonlinear mixing models have been recently proposed in the remote sensing and image processing literature. Some of them are motivated by physical findings while others exploit a more flexible nonlinear mathematical formulation to improve unmixing performance when analyzing various types of datasets. Once a particular nonlinear model has been identified to analyze a specific dataset, preprocessing could be conducted before unmixing to mitigate the nonlinearities and to restore the linear mixing assumption. Although possibly fast, such strategy may suffer from some limitations. First, it requires the inversion of the nonlinearity mapping, which is not always straightforward. Additionally, any error produced by this preprocessing step is propagated into the unmixing step, which may drastically impact the unmixing accuracy. Conversely, unified approaches able to tackle the nonlinear unmixing problem directly are expected to be more robust.

The main classes of nonlinear models and corresponding nonlinear SU algorithms are described in the sequel.

3.1 Intimate Mixture Models

As stated before, a key assumption ensuring the validity of LMM is that the mixture processes should occur at a macroscopic scale. However, in some specific situations, interactions occur at a smaller scales, e.g., when the spatial scales involved in the mixtures are smaller than the path length followed by the photons. In this context, the materials are said to be intimately mixed [27]. An illustration of these interactions is represented in Fig. 4 (middle).
Such intimate mixture models (IMM) have been considered, e.g., for analyzing mixtures composed of sand or minerals [106], sometimes in laboratory [107]. Theoretical frameworks based on the RT theory have been built to accurately describe the intimate interactions. Probably the most popular approaches devoted to such mixtures are those introduced by Hapke [27]. Indeed, these models are mainly based on meaningful and interpretable parameters with physical significance, and can relate the spectral measurements to physical characteristics of the endmembers and corresponding abundances. Note that in this case, these abundances are associated with the relative mass fractions of the materials. For instance, Hapke [105] derives a nonlinear model to express the measured reflectances as a function of various parameters of interest, such as the mass fractions, the characteristics of the individual particles (density, size) and the single-scattering albedo. Conversely, Draine [108] and Shkuratov et al. [109] have proposed alternate and relevant approximating nonlinear mixing models. However, again, one limitation inherent to these models is that they also strongly depend on external parameters related to the acquisition, e.g., the perfect knowledge of the geometric positioning of the sensor. As a consequence, performing SU in such conditions can be difficult and the expected results can be irrelevant if such parameters are approximately known, in particular when tackling the SU problem in the challenging unsupervised scenario, i.e., when the spectral signatures of the materials are unknown and need to be jointly estimated with the fractional abundances. Some SU techniques dedicated to analyze remotely sensed images based on IMM or related alternatives have been exploited in the hyperspectral literature [110, 111]. In particular, a common approach to avoid the painful inversion of the reflectance measurements into the single scattering albedo consists of using neural-networks (NN) to learn this nonlinear function. This is the strategy followed by Guilfoyle et al. [110], for which several improvements have been proposed by Altmann et al. [112] to reduce the computationally intensive learning step. In these NN-based approaches, the endmembers should be assumed to be a priori known, and are required to train the NN. Other NN-based algorithms have been studied by Plaza et al. [113–116].

Besides, Broadwater and Banerjee [117–119] derived various kernel-based unmixing techniques that implicitly relied on the Hapke model. Even more recently, Close et al. [120] introduce a the multi-mixture pixel (MMP) model, which combines macroscopic and microscopic (i.e., intimate) mixtures

$$y_p^{(MMP)} \approx \sum_{r=1}^{R} a_{r,p} m_r + a_{R+1,p} R \left( \sum_{r=1}^{R} f_{r,p} w_r \right).$$

The first term is similar to the contribution encountered in LMM in (7) and comes from the macroscopic mixing process. The second one, considered as an additional endmember with abundance $a_{R+1,p}$, corresponds to an intimate mixture described as the average single-scattering albedo [105] expressed in
the reflective domain, where \( \mathbf{f}_p = [f_{1p}, \ldots, f_{R_p}]^T \) correspond to the microscopic proportions, \( \{ \mathbf{w}_r \} \) are the albedo-domain endmember signatures and \( \mathbf{R}(\cdot) \) is the mapping function from the albedo domain to the reflectance domain. The same authors have introduced as well an unsupervised algorithm to perform MMP-based SU [120, 121].

### 3.2 Bilinear Mixing Models

Another assumption required to ensure that LMM is a relevant model to describe mixtures is related to the paths followed by the photons. Specifically, before reaching the hyperspectral sensor they must interact with only one material. However, when analyzing scenes remotely sensed over vegetated or urban areas are known to be subjected to more complex interactions that can not be properly taken into account by a simple LMM [97, 101–103, 122–125]. In these specific scenarios, due to differences in elevation between different materials (e.g., vegetation canopies vs the relatively flat soil surfaces), photons are submitted to multipath and scattering effects. An archetypal example of this kind of scene is shown in Fig. 4 (right). Similar interaction effects have been also encountered when analyzing urban scenes acquired over buildings areas, sometimes referred to as urban canyon [126–128]. Therefore, various attempts have been conducted to overcome the intrinsic limitations of the LMM. A large family of nonlinear models that have been proposed to analyze these so-called multilayered configurations consists of including powers of products of reflectance into the standard LMM, generally neglecting interactions of orders greater than two. The resulting models are known as the family of the bilinear mixing models (BMM) defined as [129]

\[
y_p^{(BMM)} \approx \sum_{r=1}^{R} \alpha_{r,p} \mathbf{m}_r + \sum_{i=1}^{R-1} \sum_{j=i+1}^{R} b_{i,j,p} \mathbf{m}_i \odot \mathbf{m}_j
\]  

(14)

where \( \odot \) stands for the termwise (Hadamard) product

\[
\mathbf{m}_i \odot \mathbf{m}_j = \begin{pmatrix} m_{1,i} \\ \vdots \\ m_{L,i} \end{pmatrix} \odot \begin{pmatrix} m_{1,j} \\ \vdots \\ m_{L,j} \end{pmatrix} = \begin{pmatrix} m_{1,i} m_{1,j} \\ \vdots \\ m_{L,i} m_{L,j} \end{pmatrix}
\]

In (14), the observed pixel is composed of a linear contribution similar to the LMM and an additive nonlinear term

\[
\sum_{i=1}^{R-1} \sum_{j=i+1}^{R} b_{i,j,p} \mathbf{m}_i \odot \mathbf{m}_j
\]  

(15)

that models nonlinear interactions between the materials. The nonlinearity coefficients \( \{ b_{i,j,p} \}_{i,j} \) allow the amount of nonlinearity in the \( p \)th pixel to be adjusted between each pair of materials \( \mathbf{m}_i \) and \( \mathbf{m}_j \). Most of the various
bilinear models of the literature mainly differ by the definition of these coefficients \( b_{i,j,p} \) and the associated constraints they should satisfied. The most common models are described below.

Somers et al. \[102\] and Nascimento and Bioucas-Dias \[130\] propose to include the nonlinearity coefficients \( b_{i,j,p} \) within the set of constraints (1), leading to

\[
y_p^{\text{NM}} \approx \sum_{i=1}^{R} a_{r,p} \mathbf{m}_i + \sum_{i=1}^{R-1} \sum_{j=i+1}^{R} b_{i,j,p} \mathbf{m}_i \otimes \mathbf{m}_j \tag{16}
\]

with

\[
\begin{align*}
& a_{r,p} \geq 0, \quad \forall r, \quad \forall p \\
& b_{i,j,p} \geq 0, \quad \forall r, \quad \forall i \neq j \\
& \sum_{r=1}^{R} a_{r,p} + \sum_{i=1}^{R-1} \sum_{j=i+1}^{R} b_{i,j,p} = 1, \quad \forall p.
\end{align*}
\]

It is worthy to note that this so-called Nascimento’s model (NM), also used by Raksuntorn and Du \[131\], reduces to the LMM when \( b_{i,j,p} = 0, \forall i \neq j \). This is an interesting property since, as stated before, LMM is known to be an admissible first approximation of the actually involved physical processes. However, in a more general setting (i.e., \( b_{i,j,p} \neq 0 \)), the linear abundance coefficients \( a_{r,p} \) are not subject to the sum-to-one constraints defined in (1). Moreover, NM can be also interpreted as a specific LMM with additional virtual endmembers. Indeed, considering \( \mathbf{m}_i \otimes \mathbf{m}_j \) as a pure component spectral signature with corresponding abundance \( b_{i,j,p} \), the model in (16) can be rewritten

\[
y_p \approx \sum_{i=1}^{\tilde{R}} \tilde{a}_{s,p} \tilde{\mathbf{m}}_s
\]

with the non-negativity and sum-to-one constraints in (1) imposed on the set of new abundance coefficients \( \tilde{a}_{s,p} \) and

\[
\begin{align*}
& \tilde{a}_{s,p} \triangleq a_{r,p}, \quad \tilde{\mathbf{m}}_s \triangleq \mathbf{m}_s, \quad s = 1, \ldots, \tilde{R} \\
& \tilde{a}_{s,p} \triangleq b_{i,j,p}, \quad \tilde{\mathbf{m}}_s \triangleq \mathbf{m}_i \otimes \mathbf{m}_j, \quad s = R + 1, \ldots, \tilde{R}
\end{align*}
\]

and \( \tilde{R} = \frac{1}{2} R (R + 1) \). Again, this formulation shows that NM reduces to a standard LMM when \( \tilde{a}_{s,p} = 0 \) for \( s = R + 1, \ldots, \tilde{R} \). Besides, Eq. (3.2) shows that, in the case where the endmember signatures \( \mathbf{m}_1, \ldots, \mathbf{m}_R \) are known, supervised SU, i.e., estimation of the abundance coefficients, can be conducted by an LMM-based inversion. More precisely, assuming an Euclidean discrepancy measure as in Eq. (9), the linear abundances \( a_s \) and nonlinear coefficients \( b_{i,j,p} \) can be recovered by solving the constrained LS regression.
i.e., given the endmember signature matrix and the LMM, which can restrict its applicability. In a supervised fashion, nonlinear interactions may occur. However, this bilinear model does not extend in that pixel: the more a given material is present in the pixel, the more the interaction coefficient of real and virtual endmembers needs to be jointly estimated with the abundance and nonlinear coefficients. Specifically Gader et al. [132] introduced an unsupervised NM-based SU algorithm by implementing updating rules that generalize the SPICE algorithm introduced by Zare and Gader [133] for the LMM.

Conversely, Fan et al. [101] have defined the nonlinearity coefficients $b_{i,j,p}$ as the product of the abundances, $b_{i,j,p} = \alpha_{i,p} \alpha_{j,p}$, under the standard abundance constraints in (1), leading to the so-called Fan’s Model (FM)

$$\mathbf{y}^{(FM)}_p \approx \sum_{r=1}^{R} \alpha_{r,p} \mathbf{m}_r + \sum_{i=1}^{R-1} \sum_{j=i+1}^{R} \gamma_{i,j,p} \alpha_{i,p} \alpha_{j,p} \mathbf{m}_i \otimes \mathbf{m}_j. \quad (18)$$

The main argument for relating the amount of nonlinear interactions (governed by $b_{i,j,p}$) to the amount of linear contribution (governed by $\alpha_{i,p}$ and $\alpha_{j,p}$) is straightforward: if the $i$th endmember is absent in the $p$th pixel, then $\alpha_{i,p} = 0$ and there are no interactions between $\mathbf{m}_i$ and the other materials $\mathbf{m}_j$ ($j \neq i$). More generally, the quantity of nonlinear interactions in a given pixel between two materials is directly related to the quantity of each material present in that pixel: the more a given material is present in the pixel, the more nonlinear interactions may occur. However, this bilinear model does not extend the LMM, which can restrict its applicability. In a supervised fashion, i.e., given the endmember signature matrix $\mathbf{M}$, the FM abundances can be estimated with the algorithm proposed by Fan et al. [101], which is based on a first-order Taylor series expansion of the nonlinearity. More recently, Echensang Guillaume [134] have proposed an unsupervised FM-based SU algorithm, based on an extension of the NMF framework introduced in Section 2.4.

To cope with the issue raised by the fact FM does not generalize the LMM, the generalized bilinear model (GBM) proposed by Halimi et al. [135] weights the products of abundances $\alpha_{i,p} \alpha_{j,p}$ by additional free parameters $\gamma_{i,j,p} \in (0, 1)$ that tune the amount of nonlinear interactions, leading to $b_{i,j,p} = \gamma_{i,j,p} \alpha_{i,p} \alpha_{j,p}$ and

$$\mathbf{y}^{(GBM)}_p \approx \sum_{r=1}^{R} \alpha_{r,p} \mathbf{m}_r + \sum_{i=1}^{R-1} \sum_{j=i+1}^{R} \gamma_{i,j,p} \alpha_{i,p} \alpha_{j,p} \mathbf{m}_i \otimes \mathbf{m}_j. \quad (19)$$

The interaction coefficient $\gamma_{i,j,p} \in (0, 1)$ quantifies the nonlinear interaction between the spectral components $\mathbf{m}_i$ and $\mathbf{m}_j$. Having $\gamma_{i,j,p} > 0$ indicates that

$$\min_{\mathbf{a}} \left\| \mathbf{y}_p - \tilde{\mathbf{M}} \mathbf{a}_p \right\|_2^2 \quad \text{s.t. (1)}$$

where $\mathbf{a}_p \triangleq \left[ \tilde{\alpha}_{1,p}, \ldots, \tilde{\alpha}_{R,p} \right]^T$ and $\tilde{\mathbf{M}} \triangleq \left[ \tilde{m}_1, \ldots, \tilde{m}_R \right]$. As a consequence, any LMM-based inversion algorithm introduced in Section 2.3.1, such as FCLS [51], can be employed to recover the abundances of the NM, given the matrix of real and virtual endmembers $\tilde{\mathbf{M}}$. When NM-based SU should be conducted in an unsupervised framework, the endmembers signatures are unknown and need to be jointly estimated with the abundance and nonlinear coefficients. Specifically Gader et al. [132] introduced an unsupervised NM-based SU algorithm by implementing updating rules that generalize the SPICE algorithm introduced by Zare and Gader [133] for the LMM.
only constructive interactions are considered. This model has the same interesting characteristic as the FM: the amount of nonlinear interactions is governed by the presence of the endmembers that linearly interact. In particular, if an endmember is absent in a pixel, there is no nonlinear interaction coming from this endmember. However, it also has the nice properties of i) generalizing the LMM by enforcing $\gamma_{i,j,p} = 0 \, (\forall i,j)$, similarly to NM but contrary to FM and ii) having the amount of nonlinear interactions to be proportional to the material abundances, similarly to FM but contrary to NM. Various algorithmic strategies have been proposed to solved the GBM-based SU in a supervised framework, i.e., when the endmember signatures $\mathbf{M}$ are known. Halimi et al. [135] have initially introduced a MCMC-based algorithm to approximate the Bayesian estimators associated with the GBM abundance and nonlinear parameters. To overcome the computational complexity imposed by MCMC, optimization techniques have been proposed. Following the strategy proposed by Fan et al. [101], Halimi et al. [136] propose to linearize the objective function via a first-order Taylor series expansion of the nonlinearity. Then, the FCLS algorithm proposed by Heinz and Chang [51] can be used to estimate the new parameter vector. An alternate scheme proposed by Halimi et al. [136] consists of resorting to a gradient descent method coupled with constrained line search procedure enforcing the constraints inherent to the mixing model to adjust the step-size parameter. Yokoya et al. [137] have proposed a semi-NMF framework to perform supervised GBM-based SU. When the endmember signatures also need to be recovered (i.e., in an unsupervised SU framework) Heylen and Scheunders [138] have adopted a geometrical point-of-view and conducted an integral analysis to compute geodesic distances on the nonlinear manifold induced by the GBM. This formulation opens the door to an EEA able to identify the simplex of maximum volume contained in the manifold defined by the GBM-mixed pixels, following the geometrical ideas detailed in Section 2.2.

### 3.3 Polynomial Mixing Models

All the bilinear mixing models introduced above only include between-component interactions $\mathbf{m}_i \odot \mathbf{m}_j$ with $i \neq j$ but no within-component interactions $\mathbf{m}_i \odot \mathbf{m}_i$. Meganem et al. [128] have derived a nonlinear mixing model based on RT theory to analyze a simple canyon-like urban scene, leading, after successive approximations and simplifying assumptions, to the following linear-quadratic mixing model (LQM)

$$
\mathbf{y}_p \approx \sum_{i=1}^{R} \alpha_{i,p} \mathbf{m}_i + \sum_{i=1}^{R} \sum_{j=1}^{R} b_{i,j,p} \mathbf{m}_i \odot \mathbf{m}_j
$$

(20)

with the non-negativity and additivity constraints in (1) and $b_{i,j,p} \in (0,1)$. This model, also encountered in the blind source separation community [139], is
similar to the general formulation of the bilinear models in Eq. (14), with the noticeable difference that the nonlinear contribution includes quadratic terms $m_i \odot m_i$. An NMF-like algorithm has been introduced by Meganem et al. [128] to perform unsupervised LQM-based SU.

Inspired by pioneering works in blind source separation by Taleb and Jutten [140], Altmann et al. [141] have proposed an approximating model able to describe a wide class of nonlinearities. More precisely, the $p$th observed pixel spectrum is defined as a nonlinear transformation $g_p(\cdot)$ of a linear mixture of the endmember spectra

$$y_p \approx g_p \left( \sum_{r=1}^{R} a_{r,p} m_r \right)$$

where the nonlinear function $g_p$ is defined as a second order polynomial governed by a unique parameter $b_p$

$$g_p : (0,1)^L \rightarrow \mathbb{R}^L$$

$$x \mapsto [x_1 + b_p x_1^2, \ldots, x_L + b_p x_L^2]^T$$

This polynomial post-nonlinear mixing model (PPNM) can be rewritten as

$$y_p^{(PPNM)} \approx M a_p + b_p (M a_p) \odot (M a_p)$$

or

$$y_p^{(PPNM)} \approx M a_p + b_p \sum_{j=1}^{R} \sum_{i=1}^{R} a_{i,p} a_{j,p} m_i \odot m_j.$$

Eq. (24) shows that PPNM includes bilinear terms $m_i \odot m_j$ ($j \neq i$) similar to those involved in the wide class of BMM (i.e., NM, FM, and GBM), including quadratic terms $m_i \odot m_i$ encountered in the LQM (20), which may account for interactions between similar materials. The PPNM has demonstrated a great flexibility to model various nonlinearities not only for unmixing purposes [104, 141] but also to detect nonlinear mixtures in the analyzed scene [142]. This model has also the great advantage of having the amount of nonlinearity to be governed by a unique parameter $b_p$ in each pixel, contrary to other BMM such as NM or GBM. When the endmember spectral signatures $M$ are known, estimating the abundance vector $a_p$ and the nonlinearity parameter $b_p$ defining the PPNM can be performed following the various Bayesian and optimization algorithms proposed by Altmann et al. [141]. When these signatures are unknown, Altmann et al. [143] derives a unsupervised PPNM-based SU algorithm based on an MCMC method.

Note finally that other nonlinear models of higher order polynomials have been recently considered by Marinoni et al. [144] and Heylen and Scheunders [145].
3.4 Robust Linear Models

Although nonlinear mixing models can significantly improve the unmixing procedure in the presence of complex mixing processes, linear models often stand as a satisfactory trade-off between algorithm complexity and performance for most observed mixed spectra. Thus, recent methods have investigated robust LMMs generally defined as

\[ y_p^{(\text{LMM})} \approx M_a p + r_p \]

where the additional term \( r_p \) represents any potential deviation from the classical LMM. This class of models for robust linear SU allows for a wide variety of additional deviations from the LMM to be handled, such as nonlinear effects, outliers or possible endmember variability (see Section 5). As illustrated in Sections 3.2 and 3.3, the term \( r_p \) can be specified to explicitly account for its dependency on the endmember spectra and/or abundances. This is the case when considering polynomial models (e.g., bilinear models and PPNM) or the residual component analysis (RCA) based model proposed by Altmann et al. [146], where \( r_p \) has been considered as a random vector described by Gaussian processes whose covariance matrices involve the endmember matrix \( M \). Such models will be further discussed in Section 3.5.

In this section, contrary to the previously mentioned models, the term \( r_p \) is assumed to be not explicitly parametrized by the endmember spectra or their respective proportions. Conversely, the set of terms \( \{r_p\}_p \) presents sparsity structures, i.e., contains many zero entries. Such assumption can be easily motivated by the fact that LMM is an admissible first approximation of the physical processes underlying the mixtures in both the spectral and the spatial domains. Févote and Dobigeon [28] have proposed an NMF method to jointly estimate the endmembers and abundances, assuming that the potential non-zeros entries in \( R = [r_1, \ldots, r_P] \) appear only in a reduced number of spatial measurements. This is for instance the case when nonlinear mixtures or other anomalies occur in a few specific spatial locations (i.e., pixels) and produce nonlinear contributions in most corresponding wavelengths. Mathematically speaking, this robust LMM assumes that the energy of the residual terms \( \|r_p\|_2 \) \((p = 1, \ldots, P)\) is null for most pixels, i.e., the LMM holds, while some of these residual energies are non-zeros. In other words, most of the columns of the residual component matrix \( R \) are assumed to be null vectors (for spatial indexes \( p \) corresponding to pixels resulting from LMM). In such cases solving the unsupervised SU problem can be formulated as

\[
\min_{M, A, R} \sum_{p=1}^{P} D_p[y_p|M_a p + r_p] + \lambda \|R\|_{2,1} \quad \text{s.t. (1)}
\]

where \( \|R\|_{2,1} = \sum_{p=1}^{P} \|r_p\|_2 \) is the so-called \( \ell_2 - \ell_1 \) norm and \( \lambda \) is a regularization parameter which controls the structured sparsity of the solution.
Féotte and Dobigeon [28] have shown that the proposed approach seems to be sufficiently flexible to analyze hyperspectral data for which some measured pixel spectra result from various mixtures, e.g., BMM or MMP model described in Sections 3.1 and 3.2.

Similarly, Newstadt et al. [147] have considered spatial and spectral sparsity structures (i.e., blocks) in the anomaly/outlier matrix term $R$, which allows deviations from the LMM to be accounted for when these deviations occur in specific spatial regions or spectral bands of the hyperspectral data. This is typically the case when actual outliers are present, but also when material spectral signatures present have significant variations in particular spectral ranges (e.g., due to natural variability of the materials). This approach has been extended by Altmann et al. [148] who have modeled the spectral and spatial dependencies through more flexible Markovian models. Both Newsstadt et al. [147] and Altmann et al. [148] have proposed to solve the resulting (NP-hard) BSS problems involving $l_0$-type penalizations using MCMC methods.

### 3.5 Nonparametric SU Techniques

SU appears even more challenging when no prior knowledge regarding the nonlinear process is available. To face with this issue, in a supervised scenario, Plaza et al. [149] and Li et al. [150] develop a classification-like scheme to solve the inversion step and estimate the abundance coefficients using support vector machines. In an unsupervised setting, when the endmember signatures $m_1, \ldots, m_R$ are not known, a geometrical-oriented SU can be conducted, based on graph-based approximate geodesic distances [151], or manifold learning techniques [152, 153]. Specifically, as stated by Dobigeon et al. [154], when the nonlinear effects are not too severe and the pure pixel assumption discussed in Section 2.2 is fulfilled, geometrical-based EEA still provide reasonable estimation of the endmember signatures.

Other relevant approaches lie on nonparametric methods based on reproducing kernels as developed by Chen et al. [155, 156], Li et al. [157], and Nguyen et al. [158]. More precisely, operating in reproducing kernel Hilbert spaces (RKHS) is a well-admitted framework in the machine learning and signal processing community. For instance, kernel-based methods have been widely considered for detection and classification in hyperspectral images [159]. Conversely, when tackling the SU problem nonlinear unmixing approaches exploiting RKHS formalism have received much less interest. Broadwater et al. [160] and Broadwater and Banerjee [117] have merely proposed to replace each inner product between endmember signatures involved in the target cost functions with a kernel function. This strategy has been motivated by its high efficiency when solving classification problems since it is well known that an appropriate nonlinear mapping generally increases the class separability [161]. It can be straightforwardly interpreted as a
nonlinear mapping operating on each the endmember signature, while neglecting interactions between the materials. Unfortunately, such strategy is more difficult to be advocated when conducting SU. Indeed, nonlinear interactions involved in the mixing process are rarely uniquely dominated by material-wise distortions, but rather by nonlinear interactions between the components themselves, as discussed in the previous section. In a supervised setting, Chen et al. [155] have proposed to take the nonlinear interactions of the endmembers into account by exploiting a new kernel-based paradigm. More precisely, exploiting the mixing relation in Eq. (3), the authors have formulated the SU problem in Eq. (4) as its functional counterpart

$$\min_{f_b \in \mathcal{H}} \| y_p - f_b(M, a_p) \|^2_2 + \lambda \| f_b \|^2_{\mathcal{H}},$$

(27)

where $\mathcal{H}$ stands for a given functional space, the function $f_b$ in Eq. (3) is now parametrized by additional (nonlinear) parameters $b_p$ and $\lambda$ adjusts the trade-off between regularity of the function $f_b$ and the data-fitting term. The RKHS $\mathcal{H}$ may drastically impact the quality of the SU task and should be carefully chosen to address various nonlinear interactions. Chen et al. [155] have proposed to define the function $f_b$ as a combination of a linear contribution similar to LMM and a nonparametric term accounting for any nonlinear interactions

$$f_b(M, a_p) = Ma_p + f_{NLMM}(M)$$

(28)

where $f_{NLMM}(\cdot)$ is a band-wise and material-wise real-valued function of an appropriate RKHS. In particular, when this latter is chosen as the space of 2nd order polynomials, the proposed approach generalizes the PPNM introduced in Section 3.3. Moreover, due to the strong connections between RKHS and Gaussian process formalisms, the RCA-based model proposed by Altmann et al. [146] and further generalized by Altmann et al. [162] to account for the spatial dependencies between nonlinearity effects, can also be associated to a polynomial kernel. Other nonlinear interactions can be considered by choosing an alternate reproducing kernel, inducing a specific RKHS.

Another promising nonparametric strategy has been introduced by Altmann et al. [163] who considers a kernel-based unsupervised SU approach. It relies on a nonlinear dimensionality reduction through a Gaussian process latent variable model (GPLVM). In this work, a particular form of kernel have been considered to extend the generalized bilinear model in (19). The resulting Bayesian algorithm is fully unsupervised since the endmember signatures can be jointly estimated with the abundance coefficients, as a complementary result of a prediction step [163]. Another advantage of GPLVMs is their ability to accurately describe various nonlinearities. This strategy has been also considered by Nguyen et al. [158] to solve the so-called pre-image problem [164]. When addressing the SU context, GPLVMs assume that spectrally
different pixels result from different abundance vectors and thus can be represented by different latent variables. Conversely, spectrally close pixels are expected to have similar abundance vectors and thus similar latent variables. Several approaches have been proposed to preserve similarities, including back-constraints [165], dynamical models [166] and locally linear embedding (LLE) [167].

4 EXPERIMENTS: UNMIXING EELS DATA

By far, SU has been mainly considered by the geoscience community to analyze hyperspectral images acquired over the Earth [19]. In a more limited way, unmixing algorithms have been also advocated to analyze planetological data, e.g., acquired over Mars [58, 87]. More recently, since the first attempt by Dobigeon and Brun [18], SU has been considered to analyze data acquired by scanning transmission electron microscope [168–170], in particular EELS data. Following this recent trending, this section demonstrates that SU can extract meaningful and relevant characteristics from EELS spectrum-images.

The hyperspectral dataset has been acquired on a multilayer consisting of altered layers of carbon and boron nitride deposited by ion beam assisted deposition (IBAD) [171]. The cross-sectional specimen has been prepared using tripod polishing and ion milling. The data have been acquired on an aberration-corrected Nion UltraSTEM microscope operated at 100 kV. The 250 × 39 pixels (0.078 nm/pixel) spectrum-image has been acquired using 0.05 s/pixel acquisition time, 0.3 eV/channel energy dispersion (1340 channels).

The data have been analyzed using a set of six representative SU methods, i.e., two linear and four nonlinear SU methods presented in Sections 2 and 3 of this chapter. More precisely, the first linear method, referred to as BLU, is the Bayesian unsupervised unmixing method proposed by Dobigeon et al. [85] and relying on an additive Gaussian noise model whose variance is constant across the observed pixels and wavelengths. The second linear method, referred to as BLU+, is also a Bayesian method relying on a Gaussian noise assumption and generalizes BLU by allowing the noise variance to change across wavelengths. This method is a special case of the robust linear SU proposed by Altmann et al. [148]. The first two nonlinear methods are based on the PPNM presented in Section 3. The first algorithm (referred to as SPPNM for supervised PPNM [141]) uses the spectral signatures estimated by BLU+ as endmembers and thus solves a supervised nonlinear SU problem. The second PPNM-based method, referred to as UPPNM (unsupervised PPNM), is the unsupervised SU method proposed by Altmann et al. [143]. The third method is the RCA-based algorithm from Altmann et al. [162] (GRCA for generalized RCA), which estimates the nonlinearities and abundances associated with the endmembers estimated by BLU+. One of the main advantages of these three nonlinear methods is the fact that the deviations from the LMM
are characterized by these models using a single parameter ($b_p$ for SPPNM and UPPNM and $s^2_p$ for GRCA) which tends to 0 when the nonlinear effects decrease and which can thus be used to quantify the levels of nonlinearity. The last three nonlinear methods rely on robust LMMs. Precisely, the fourth (RNMF) and fifth (RNMF-KL) methods are algorithms proposed by Févotte and Dobigeon [28] for unsupervised nonlinear unmixing and rely on Gaussian (squared-Euclidean distance) and Poisson noise (generalized Kullback-Leibler divergence) assumptions, respectively. Finally, the last method, referred to as RBLU for robust BLU, is the Bayesian algorithm proposed by Altmann et al. [148] based on a Markovian spatial/spectral anomaly model. The arbitrary parameters (if any) of each method have been tuned by cross-validation.

For the considered dataset, the number of endmembers has been chosen as $R = 6$ from preliminary runs. The corresponding endmember spectra estimated by the different algorithms are depicted in Fig. 5. The corresponding estimated abundance maps are depicted in Fig. 6. These two figures show that all methods generally estimate similar endmembers and abundances, although local differences can be observed.

The endmember #1 corresponds to the substrate. The other endmembers spectra characterize the constituents of the layers: the identified endmembers #2 and #6 correspond to a graphite-like carbon K-edge, the endmember #3 to a boron-like K-edge, and the endmembers #4 and #5 to interlayer compounds resulting from the complex apparatus geometry combining evaporation with ion bombardment. Thus endmembers #4 and #5 appear at the interface between nominal layers and could be related to intermediate compounds as non-stoichiometric BN$_x$ and BC$_x$, respectively, as endmembers #4 and #5 exhibit changes in the fine structure of the B K edge at 180 eV related to different bonding of B compounds [71]. Indeed, a variety of bonding structures can be obtained by IBAD within the B-C-N system. The interest of SU in this case is clearly to reveal the existence of such compounds at a nanometer scale.

Fig. 7 depicts the average noise variance in each channel, estimated by the different methods. While the methods assuming the same noise level in all channels (i.e., BLU and RNMF) provide a single estimated variance, the other methods provide more informative results. Indeed they all show that the noise variance generally decreases when the energy dispersion increases. They also highlight specific channels of high data variability (two peaks around 200 and 300 eV).

Finally, Fig. 8 depicts the level of nonlinearity estimated by the nonlinear SU methods in the $250 \times 39$ pixels. Precisely, this level has been represented by $|b_p|$ for the PPNM-based methods, by the logarithm of $s^2_p$ for GRCA and by the norm of the deviation term $||r_p||^2$ for the robust LMM-based methods. Although the results depend on the underlying mixing model, spatial structures in agreement with the abundances depicted in Fig. 6 can be identified. In particular, nonlinear effects seem to be identified at the boundaries between
FIG. 5 Endmember spectral signatures extracted by various algorithms.
FIG. 6 Abundance maps estimated by various algorithms.

FIG. 7 Noise variances estimated by various algorithms.
FIG. 8 Nonlinearity levels estimated by various algorithms.
the substrate and the graphite-like carbon K-edge and between the graphite-like carbon-and boron K-edges where complex interactions are expected to occur.

5 CONCLUSION

This chapter introduced spectral unmixing as a powerful analysis tool able to reveal latent and unobserved spectral and spatial structures in hyperspectral images acquired through various modalities, from long-range remote sensors to microscopy imagers. By identifying the spectral signatures of the main components present in the imaged scene while quantifying their respective spatial distributions over the scene, SU provides a compact, comprehensive and meaningful (i.e., physically interpretable) description of the whole set of measurements. In an unsupervised scenario, i.e., when both the endmember spectra and abundance vectors are unknown, SU can be formulated as a blind source separation problem. However, contrary to other popular methods of multivariate data analysis, such as principal or independent component analysis, the mathematical formulation underlying SU is driven by an acute will to model the physical interactions involved in the mixing process. This explains why SU provides latent parameters whose denominations (component spectral signatures and fractional abundances) directly refers to physical quantities of interest.

As a consequence, one key step when formulating SU consists of properly defining the mixture model that could explain the hyperspectral measurements. By far, the linear mixing model is the most commonly used since it is an admissible first approximation of the actual material interactions and generally provides relevant results. In this case, solving SU can be casted as a nonnegative matrix factorization problem, complemented by additional specific constraints. For more complex scenes, e.g., composed of multi-layered structures (vegetated or urban areas) or sand-like materials, other nonlinear mixing models need to be considered, at the price of a higher computational burden. To avoid the use of these nonlinear models when the linear mixing model is sufficient, detection-oriented techniques can be implemented to distinguish the linearly from the nonlinearly mixed pixels [142, 172, 173].

Akin to a large majority of BSS or NMF techniques, SU algorithms generally operates in a pixel-wise fashion, i.e., the mixing and unmixing processes are defined for each pixel independently. However some noticeable contributions have shown that exploiting contextual (e.g., spatial) information can improve the accuracy of the mixing model as well as the unmixing results. Indeed, firstly, when defining the model that governs the material interactions that occur in a given pixel, only components that are assumed to be present in this pixel are generally considered. However taking adjacency effects, i.e., additional interactions coming from spectral interference caused by
atmospheric scattering, into account has been addressed in an unmixing context in Burazerovic et al. [174].

Moreover, when dealing with unmixing, most of the inversion or NMF-based SU algorithms work pixel-by-pixel. It is worthy to note that the Bayesian framework initially adopted by Dobigeon et al. [61, 85] to design SU algorithms offers the possibility of incorporating spatial dependencies between neighboring pixels through the use of Markov random fields. This strategy has been considered by Nishii and Ozaki [175] and Eches et al. [176, 177] to promote similar statistical characteristics of the abundance vectors associated with neighboring pixels. Conversely, such Markovian models have been also resorted by Altmann et al. [178] to improve the identification of the endmembers when performing semi-supervised SU. Deterministic counterparts of these Bayesian approaches have also shown interesting results when incorporating additional spatial information through an appropriate regularization. For instance Chen et al. [156] and Iordache et al. [179] have proposed to consider an \( \ell_1 \)-type regularizer form to promote piecewise-constant transitions in the fractional abundances. Legendre et al. [180] have defined an alternate regularizer to preserve the image edges.

Another major assumption commonly admitted when performing SU is the spectral invariance of the materials. More precisely, in common SU approaches, each component is assumed to be uniquely represented by a single spectral signature. However, spectral variability is known to be one of the major source of errors when conducting unmixing [181]. Various techniques have been proposed in the literature to mitigate this spectral variability. As stated by Somers et al. [33], two classes of SU methods have been mainly developed. The first one represents each endmember as a collection of spectral signatures, or bundles, representative of the corresponding materials [182]. Conversely, another class of methods represent the endmembers as statistical distributions [177, 183]. More recently, Thouvenin et al. [184] and Veganzones et al. [185] have proposed to derive parametric models to explicitly account for endmember variability over the pixels of a given image.

Finally, it is worthy to note that preliminary works have shown that SU can be implemented for a collection of hyperspectral images acquired over time, which opens the door to multitemporal SU [186, 187].

ACKNOWLEDGMENTS

N.D. and Y.A. would like to thank their close collaborators, Jean-Yves Tourneret, University of Toulouse; Steve McLaughlin and Abderrahim Halimi, Heriot-Watt University; Alfred O. Hero, University of Michigan; Cédric Févotte, University of Nice; and Jose M. Bioucas Dias, Universidade de Lisboa; for stimulating and fruitful discussions during the past years while facing with unmixing problems. N.B. would like to thank Virginie Serin (CEMES, Toulouse) and Ignacio Jimenez (ICMM-CSIC Madrid) for providing the C/BN sample and for comments on the results.
REFERENCES


