From Separability/Identifiability Properties of Bilinear and Linear-Quadratic Mixture Matrix Factorization to Factorization Algorithms

Yannick Deville

IRAP (Institut de Recherche en Astrophysique et Planétologie), Université de Toulouse, UPS, CNRS, CNES, 14 Av. Edouard Belin, F-31400 Toulouse, France.
E-mail: yannick.deville@irap.omp.eu .

Abstract

Blind source separation (BSS) and Blind Mixture Identification (BMI) methods typically concern unknown source signals, transferred through a given class of functions with unknown parameter values, which yields mixed observations. Using only these observations, BSS/BMI aims at estimating the source signals and/or mixing parameters. Most investigations concern linear instantaneous mixing functions. They contain two aspects. The first one consists in *proposing* general BSS/BMI principles, e.g. Independent Component Analysis, Sparse Component Analysis or Nonnegative Matrix Factorization (NMF), and/or deriving associated practical algorithms. The second aspect consists in *analyzing* the properties resulting from these principles. This is of utmost importance, to determine if the proposed BSS/BMI principles are guaranteed to separate the source signals and to identify the considered mixing model up to acceptable indeterminacies. These separability/identifiability analyses are even more important for *nonlinear* mixtures, that were shown to potentially yield higher indeterminacies. Among them, bilinear and linear-quadratic mixtures are receiving increasing attention, e.g. due to their application to remote sensing. Especially, extensions of NMF were recently proposed for them, but the resulting separability/identifiability properties were not analyzed. We here address this topic, moreover proceeding further by investigating Bilinear and Linear-Quadratic Mixture Matrix Factorization (BMMF and LQMMF) approaches without nonnegativity constraints. We especially show that, whereas nonlinearity is often considered to be a burden, it yields an essentially unique decomposition under mild conditions for BMMF. On the contrary, full LQMMF is shown to yield spurious solutions, which increases the usefulness of combining it with nonnegativity constraints in applications where data meet these constraints. Algorithms based on this framework are also defined in this paper and their performance is reported.

Keywords: nonlinear blind source separation, blind nonlinear mixture identification, bilinear and linear-quadratic mixtures, matrix factorization, separability and identifiability, uniqueness of decomposition

1. Introduction

Blind source separation (BSS) methods aim at restoring a set of unknown source signals from a set of observed signals which are mixtures of these source signals [2], [4], [6], [10]. The considered mixing model, i.e. the class of mixing functions, is usually predefined, but the values of its parameters are initially unknown. BSS is thus closely linked to the estimation of these parameters, i.e. to Blind Mixture Identification (BMI) [7], [10]. In the most studied mixing model, the observations are linear instantaneous (i.e. memoryless) combinations of the source signals. Several classes of BSS/BMI methods were developed for this model, especially including Independent Component Analysis (ICA), Sparse Component Analysis (SCA) and Nonnegative Matrix Factorization (NMF) [4], [7], [10].

The standard procedure [6], [10] for developing a complete BSS/BMI method consists in defining five items: the considered mixing model, separating structure, separation principle (e.g. enforcing mutual statistical independence of separating system outputs in ICA), separation criterion (e.g. minimization of a cost function) and separation algorithm (e.g. gradient descent for minimizing a given cost function). Besides, the above steps allow one to *propose* a BSS/BMI method or a class of methods, but one should in addition *analyze* if that/these methods have acceptable separability/identifiability properties. At the most general level, the latter task consists in considering all the class of BSS/BMI methods associated with the considered mixing model, separating structure and separation principle, and in deriving all the solutions of that configuration, in terms of estimated sources (separability properties) and/or estimated mixing model parameters (identifiability properties). This defines the indeterminacies of that configuration, i.e. the residual transforms up to which the source signals may be separated and/or the mixing model may be identified. Whereas many investigations were devoted to *proposing* new linear instantaneous BSS/BMI methods since the 1990s, far fewer *analyses* of separability/identifiability properties were reported. These analyses are defined in Appendix A.

Beyond the above linear (instantaneous) mixtures, BSS/BMI methods were also developed for nonlinear mixing models, especially for post-nonlinear models and for linear-quadratic ones, including their bilinear version: see e.g. the surveys of these methods and of their applications in [8] or [10]. In particular, extensions of NMF intended for bilinear and linear-quadratic mixtures were especially reported in [13], [17], [18], [19], [20], [21].

In this paper, we first address bilinear mixtures, defined in Section 2. More specifically, we consider the class of BSS/BMI methods called Bilinear Mixture Matrix Factorization, or BMMF, methods (this terminology is justified in Appendix B). We stress that these methods do not assume nonnegative sources and mixing coefficients (although they allow them), unlike the above-mentioned NMF-based methods. In the very first stage of this investigation reported in the short conference paper [9], we "proposed but almost did not analyze" this class of BMMF methods, in the above sense: considering bilinear mixtures, that paper [9] introduced associated separating structure and separation principle (summarized hereafter in Section 3), then focused on specific separation criteria and algorithms, but did not analyze the indeterminacies thus obtained, except by performing direct calculations for the very specific case of two sources.

In the present paper, our first contribution dealing with bilinear mixtures is therefore fully complementary to the above article: we address the complete class of BSS/BMI methods based on the BMMF separation principle defined in Section 3 (i.e. without depending on any separation criterion or algorithm) and we provide an analytical derivation of its indeterminacies for an arbitrary number of sources, which requests us to use a completely different approach as compared with [9]. This separability/identifiability analysis is presented in Section 4. Whereas that analysis applies to any number M of sources, it may be better understood by focusing on the simplest cases, that is $M = 2$ and $M = 3$. Additional explanations are therefore provided for these two cases. They are gathered in Appendix C, for the sake of readability. Once the attractive separability/identifiability properties of the proposed BMMF approach have thus been established, corresponding separation criteria and algorithms should be derived, in the framework of the above-defined procedure. The range of options for these criteria and algorithms is quite wide, so that they will be analyzed in more detail in a future paper. However, some of them are here presented in Sections 5 and 6 to completely illustrate the proposed procedure. Moreover, numerical tests performed with data corresponding to major applications of the considered mixing model are then reported in Section 6. They first prove the relevance of the assumptions used in our investigation of BMMF separability/identifiability, and they then show the performance of the considered practical BMMF algorithms.

An extended version of the above model, namely the full linear-quadratic model, is then considered in Section 7, and the resulting separability/identifiability issues are illustrated. Finally, conclusions are drawn from this investigation in Section 8.

2. Bilinear mixing model

Considering real-valued signals which depend on a discrete variable n, the scalar form of the bilinear (instantaneous) and noiseless mixing model reads

$$x_i(n) = \sum_{j=1}^{M} a_{ij} s_j(n) + \sum_{j=1}^{M-1} \sum_{k=j+1}^{M} b_{ijk} s_j(n) s_k(n) \qquad \forall\ i \in \{1, \ldots, P\} \tag{1}$$

where $x_i(n)$ are the values of the P observed mixed signals for the sample index n and $s_j(n)$ are the values of the M unknown source signals which yield these observations, with $M \geq 2$ hereafter, whereas a_{ij} and b_{ijk} are respectively the linear and quadratic real-valued mixing coefficients (with unknown values in the blind case) which define the considered source-to-observation transform.

A first matrix form of that model (1) reads

$$x(n) = As(n) + Bp(n) \tag{2}$$

where the source and observation vectors are

$$s(n) = [s_1(n), \dots, s_M(n)]^T \tag{3}$$

$$x(n) = [x_1(n), \dots, x_P(n)]^T, \tag{4}$$

where ^{T} stands for transpose and matrix A consists of the mixing coefficients a_{ij}. The column vector $p(n)$ is composed of all source products $s_j(n)s_k(n)$ of (1), i.e. with $1 \leq j < k \leq M$, arranged in a fixed, arbitrarily selected, order [9]. The matrix B is composed of all entries b_{ijk} arranged so that i is the row index of B and the columns of B are indexed by (j, k) and arranged in the same order as the source products $s_j(n)s_k(n)$ in $p(n)$.

An even more compact form of this model may be derived by stacking row-wise the vectors $s(n)$ and $p(n)$ of sources and source products in an extended vector

$$\widetilde{s}(n) = \begin{bmatrix} s(n) \\ p(n) \end{bmatrix} \tag{5}$$

whereas the corresponding matrices A and B are stacked column-wise in an extended matrix

$$\widetilde{A} = [A \;\; B]\,. \tag{6}$$

The bilinear mixing model (2) then yields

$$x(n) = \widetilde{A}\widetilde{s}(n). \tag{7}$$

A third matrix-form model may eventually be derived by stacking column-wise all available signal samples, which correspond to n ranging from 1 to N, in the matrices

$$\widetilde{S} = [\widetilde{s}(1), \dots, \widetilde{s}(N)] \tag{8}$$

$$X = [x(1), \dots, x(N)]. \tag{9}$$

The single-sample model (7) thus yields its overall matrix version

$$X = \widetilde{A}\widetilde{S}. \tag{10}$$

3. Separating structure and separation principle

Generally speaking, a separating system aims at providing estimates of source signals, by using adequately tuned parameters. Within this framework, a standard approach uses systems which receive the observations as their inputs and which combine them according to a model which implements a class of functions equal to the *inverse* of the class of functions corresponding to the mixing model. The parameter values of such a system define one single function within this class and should be selected so as to match those of the single function corresponding to the considered mixture. The outputs of this separating system thus yield estimates of the source signals.

Within the above overall framework, we here consider a different approach, defined in [9], which consists in building a system which aims at modeling the *direct*, i.e. mixing, function. It thus does not require the analytical form of the inverse model to be known. Since the direct function is defined by (10), the variables involved in the considered separating structure consist of two matrices, $\check{A}$ and $\check{S}$, which respectively aim at estimating $\widetilde{A}$ and $\widetilde{S}$ (possibly up to some indeterminacies). The rows of $\widetilde{S}$ and thus $\check{S}$ may be seen as vectors used to decompose the row vectors of X, whereas $\widetilde{A}$ and thus $\check{A}$ contain the coefficients of this decomposition. Moreover, matrix $\widetilde{S}$ is guaranteed to meet a constraint: as shown by (5) and (8), only its top M rows are free, i.e. they contain the source values, whereas all subsequent rows are element-wise products of two of the above rows. Therefore, the same constraint is here set on the adaptive variable $\check{S}$ of the separating structure. This means that the top M rows of $\check{S}$ are master, i.e. freely tuned, variables. These M row vectors are respectively denoted as $\check{s}_1$ to $\check{s}_M$. On the contrary, all subsequent rows of $\check{S}$ are slave variables, which are updated together with the above top M rows, so as to contain element-wise products $\check{s}_j \odot \check{s}_k$ of those top M rows. These $\check{s}_j \odot \check{s}_k$ products are only stored for $1 \leq j < k \leq M$ and arranged in a fixed, arbitrarily selected, order [9].

The separation principle used hereafter for adapting matrices $\check{A}$ and $\check{S}$ of the above separating structure consists in updating these variables associated with the direct model so that their product $\check{A}\check{S}$ fits the observation matrix X, in order to ideally achieve $\check{A}\check{S} = X$. This class of methods and their separation principle are therefore called Bilinear Mixture Matrix Factorization, or BMMF. Hereafter, we first determine all values of $\check{A}$ and $\check{S}$ which yield an exact decomposition $\check{A}\check{S} = X$ of the observed data.

4. Uniqueness of decomposition

4.1. Notations

The vectors in the top M rows of $\widetilde{S}$ are respectively denoted as s_1 to s_M. The matrix composed of these row vectors, i.e. the "linear part" of $\widetilde{S}$, is denoted as S:

$$S = \begin{bmatrix} s_1 \\ \ldots \\ s_M \end{bmatrix}. \tag{11}$$

The matrix composed of the top M rows of $\check{S}$, i.e. the "linear part" of $\check{S}$, is denoted as $\check{S}_L$:

$$\check{S}_L = \begin{bmatrix} \check{s}_1 \\ \ldots \\ \check{s}_M \end{bmatrix}. \tag{12}$$

We denote as $\widetilde{M}$ the overall number of "extended sources", which correspond to the above original source vectors s_1 to s_M and their products $s_j \odot s_k$ involved in (1). This yields

$$\widetilde{M} = \frac{M(M+1)}{2}. \tag{13}$$

The subspace spanned by the row vectors of $\widetilde{S}$ is denoted as $\int_{\widetilde{S}}$. The subspace spanned by the row vectors of X is denoted as $\int_X$.

4.2. Assumptions

The following assumptions are used as from Section 4.3. The number of samples of all source vectors s_j is supposed to be at least equal to $\widetilde{M}$. The rank of $\widetilde{S}$ is then limited by its *row* rank: it is at most equal to $\widetilde{M}$. Moreover, we hereafter consider the following case:

Assumption 1. *$\widetilde{S}$ has full row rank:* $\mathrm{rowrank}(\widetilde{S}) = \widetilde{M}$.

In other words, all $\widetilde{M}$ vectors s_j and $s_j \odot s_k$ corresponding to (1) are assumed to be linearly independent (as explained above, $\odot$ denotes element-wise vector product).

The considered mixture is assumed to be determined or over-determined with respect to the extended set of sources, i.e. the number P of rows of X

is assumed to be at least equal to $\widetilde{M}$. Besides, all row vectors of X are linear combinations of the $\widetilde{M}$ extended source vectors, as shown by (10). Therefore, the row rank of X is at most equal to $\widetilde{M}$. Moreover, we hereafter consider the following case:

Assumption 2. $\mathrm{rowrank}(X) = \widetilde{M}$.

Note that Assumption 2 implicitly implies that Assumption 1 is met. Besides, we introduce the following assumption, used as from Section 4.4:

Assumption 3. *All vectors s_1 to s_M and all different vectors involved in the right-hand part of Eq. (15) introduced below (these vectors are element-wise products of two to four vectors s_j) are linearly independent.*

When counting the number of different vectors of (15) involved in Assumption 3, one should not take into account the total number of *terms* in the right-hand part of (15), but only the number of *different element-wise vector products* that (15) yields, which is therefore lower. More precisely, let us first consider the case when $M = 2$. In that case, the first line of the right-hand part of (15) contains four terms, respectively proportional to $s_1 \odot s_1$ and $s_1 \odot s_2$ for $j = 1$, and then $s_2 \odot s_1$ and $s_2 \odot s_2$ for $j = 2$. But $s_1 \odot s_2 = s_2 \odot s_1$ due to the commutativity of the element-wise product $\odot$, so that these four terms only yield *three different products $s_j \odot s_{j'}$*. The same principle applies to the terms in the subsequent lines of (15). More generally speaking, for any M with $M \geq 4$, the number of different vector products contained in (15) may be shown to be as follows:

- The number of different products of two source vectors, which correspond to the first line of (15), is equal to $\dfrac{M(M+1)}{2}$.
- The number of different products of three source vectors, which correspond to the second and third lines of (15), is equal to $\dfrac{M(M-1)(M+4)}{6}$.
- The number of different products of four source vectors, which correspond to the fourth line of (15), is equal to $\dfrac{M(M-1)[(M-2)(M+9)+12]}{24}$.

For instance, for $M = 4$, Eq. (15) thus yields $10 + 16 + 19 = 45$ different source products. When adding this number to the number of plain source vectors s_1 to

s_M with $M = 4$ also involved in Assumption 3, the overall number of different vectors involved in Assumption 3 for $M = 4$ is equal to 49.

Note that Assumption 3 implies that Assumption 1 is met. Also note that this requires us to assume that the number of samples of all source vectors s_j is at least equal to the number of different vectors involved in Assumption 3.

4.3. Structure of the rows of $\check{S}$

Under the assumptions of Section 4.2, we derive the following properties, focused on the general structure of the rows of $\check{S}$.

Lemma 1. *A basis of $\int_X$ consists of all row vectors of $\widetilde{S}$.*

Proof Due to (10), all rows of X are linear combinations of the $\widetilde{M}$ rows of $\widetilde{S}$. Besides, due to Assumptions 1 and 2, the rows of X actually span all $\int_{\widetilde{S}}$. $\blacksquare$

Lemma 2. *If $X = \check{A}\check{S}$, then a basis of $\int_X$ consists of all row vectors of $\check{S}$.*

Proof Due to Assumption 2, $\int_X$ has dimension $\widetilde{M}$. If $X = \check{A}\check{S}$, then all directions of $\int_X$ are actually spanned by the $\widetilde{M}$ row vectors of $\check{S}$, which are therefore linearly independent. Therefore, the latter vectors span a subspace with dimension $\widetilde{M}$ and form a basis of $\int_X$. $\blacksquare$

Lemma 3. *If $X = \check{A}\check{S}$, then*

$$\check{s}_\ell = \sum_{j=1}^{M} e_{\ell j} s_j + \sum_{j=1}^{M-1} \sum_{k=j+1}^{M} e_{\ell j k} s_j \odot s_k \qquad \forall\ \ell \in \{1, \ldots, M\} \tag{14}$$

where $e_{\ell j}$ and $e_{\ell j k}$ are coefficients.

Proof Combine Lemmas 1 and 2. Thus, if $X = \check{A}\check{S}$, then each row vector of $\check{S}$ is a linear combination of all row vectors of $\widetilde{S}$. In particular, this applies to any of the top M rows of $\check{S}$. $\blacksquare$

Lemma 4. *If $X = \check{A}\check{S}$, then*

$$\begin{aligned}
\check{s}_\ell \odot \check{s}_m \quad = \quad & \sum_{j=1}^{M}\sum_{j'=1}^{M} e_{\ell j} e_{mj'} s_j \odot s_{j'} \\
& + \sum_{j=1}^{M}\sum_{j'=1}^{M-1}\sum_{k'=j'+1}^{M} e_{\ell j} e_{mj'k'} s_j \odot s_{j'} \odot s_{k'} \\
& + \sum_{j=1}^{M-1}\sum_{k=j+1}^{M}\sum_{j'=1}^{M} e_{\ell jk} e_{mj'} s_j \odot s_k \odot s_{j'} \\
& + \sum_{j=1}^{M-1}\sum_{k=j+1}^{M}\sum_{j'=1}^{M-1}\sum_{k'=j'+1}^{M} e_{\ell jk} e_{mj'k'} s_j \odot s_k \odot s_{j'} \odot s_{k'} \\
& \forall\ \ell \in \{1, \dots, M\},\ \forall\ m \in \{1, \dots, M\}. \qquad (15)
\end{aligned}$$

Proof Apply Lemma 3 to $\check{s}_\ell$ and $\check{s}_m$, and take their element-wise product. ∎

For more details about the terms of (15) when $M = 2$ or $M = 3$, see Appendix C.

4.4. Simplification of linear part of rows of $\check{S}$

Under the assumptions of Section 4.2, we hereafter derive properties of the coefficients $e_{\ell j}$ of (14).

Lemma 5. *If $X = \check{A}\check{S}$, then for each j with $j \in \{1, \dots, M\}$, at least one of the coefficients $e_{\ell j}$ in (14) is non-zero.*

Proof For a given, arbitrary, value of j, with $j \in \{1, \dots, M\}$, let us assume that all coefficients $e_{\ell j}$, with $\ell \in \{1, \dots, M\}$, are equal to zero, i.e. that s_j does not appear in any of the linear parts of the vectors $\check{s}_\ell$ defined by (14). Then, whatever $\check{A}$, the row rank of $\check{A}\check{S}$ is strictly lower than $\widetilde{M}$. But this contradicts the joint assumption composed of condition $X = \check{A}\check{S}$ and Assumption 2. Therefore, if $X = \check{A}\check{S}$ and using Assumption 2, one gets the result of Lemma 5. ∎

Lemma 6. *If $X = \check{A}\check{S}$, then*

$$e_{\ell j} e_{mj} = 0 \quad \forall\ j, \ell, m \in \{1, \dots, M\} \text{ with } \ell \neq m. \qquad (16)$$

Proof Let us first consider the case when $\ell < m$. As already shown in the proof of Lemma 3, if $X = \check{A}\check{S}$, then each row vector of $\check{S}$ is a linear combination of all row vectors of $\widetilde{S}$. We here apply this to any bottom row of $\check{S}$, which is defined by (15), here with $1 \leq \ell < m \leq M$. Then, by combining the property provided above in the current proof and Assumption 3, all components $s_j \odot s_j$ of (15) are equal to zero. Considering the terms with $j = j'$ of the first sum (i.e. first line) of (15), this yields (16).

Let us now consider the case when $\ell > m$. Then, by permuting the roles of ℓ and m, i.e. setting $\ell' = m$ and $m' = \ell$, rewriting (16) as $e_{mj}e_{\ell j} = 0$ and thus $e_{\ell' j}e_{m' j} = 0$, and using the above proof for ℓ' and m', with $\ell' < m'$, the desired result is obtained. ∎

Lemma 7. *If $X = \check{A}\check{S}$, then for each j with $j \in \{1, \ldots, M\}$, at most one of the coefficients $e_{\ell j}$ in (14) is non-zero.*

Proof Use Lemma 6: for two different indices ℓ and m, the product of $e_{\ell j}$ and e_{mj} is equal to zero, so at least one of these coefficients is equal to zero. ∎

Lemma 8. *If $X = \check{A}\check{S}$, then for each j with $j \in \{1, \ldots, M\}$, exactly one of the coefficients $e_{\ell j}$ in (14) is non-zero.*

Proof Combine Lemma 5 and Lemma 7. ∎

Lemma 9. *If $X = \check{A}\check{S}$, then each row vector $\check{s}_\ell$ in (14) has at least one non-zero coefficient $e_{\ell j}$ with $j \in \{1, \ldots, M\}$.*

Proof Let us assume that $K > 0$ of the vectors $\check{s}_\ell$, with $\ell \in \{1, \ldots, M\}$, have only zero-valued coefficients $e_{\ell j}$ in the first term of their decomposition (14). Then, let us consider all possible linear combinations of all vectors $\check{s}_1$ to $\check{s}_M$ (and of $\check{s}_\ell \odot \check{s}_m$ with $1 \leq \ell < m \leq M$) and let us focus on the "linear parts" of these combinations, i.e. their terms which are linear combinations of s_1 to s_M. Due to (14), these linear parts are combinations of $(M - K)$ vectors which are linear combinations of s_1 to s_M, so that these linear parts span a subspace with dimension (at most) $M - K < M$. Therefore, if $X = \check{A}\check{S}$, the dimension of $\int_X$ is strictly lower than $\widetilde{M}$. But this contradicts Assumption 2. Therefore, assumption $K > 0$ is false, i.e. $K = 0$, which yields Lemma 9. ∎

Lemma 10. *If $X = \check{A}\check{S}$, then each row vector $\check{s}_\ell$ in (14) has exactly one non-zero coefficient $e_{\ell j}$ with $j \in \{1, \ldots, M\}$.*

Proof We here consider the case when $X = \check{A}\check{S}$. Let us stack row-wise the vectors $\check{s}_1$ to $\check{s}_M$, consider their "linear part", which corresponds to the first sum in (14), and let us consider column-wise the contributions respectively associated with s_1 to s_M in this linear part. First, due to Lemma 8, in each column j (corresponding to s_j), with $j \in \{1, \ldots, M\}$, exactly one coefficient $e_{\ell j}$ with $\ell \in \{1, \ldots, M\}$, is non-zero. Second, due to Lemma 9, in each row ℓ (corresponding to $\check{s}_\ell$), with $\ell \in \{1, \ldots, M\}$, at least one coefficient $e_{\ell j}$, with $j \in \{1, \ldots, M\}$, is non-zero. The first of these two properties shows that, when successively considering rows 1 to M, for each row ℓ, whenever a coefficient $e_{\ell j}$ in column j is non-zero, that column is thus "used" and cannot be "used again" for other rows, i.e. all other $e_{\ell j}$ with the same j as above are equal to zero. Besides, the second above property shows that each row ℓ thus "uses" at least one column (since it has at least one non-zero coefficient $e_{\ell j}$). We now add the following extension of that result: each row ℓ thus uses *exactly* one column. This may be shown as follows. If at least one row ℓ uses more than one column (i.e. if more that one coeffcient $e_{\ell j}$, with $j \in \{1, \ldots, M\}$, is non-zero), then these columns cannot be used again by other rows, i.e. at most $(M-2)$ columns are still available for the other $(M-1)$ rows $\ell' \neq \ell$. This number of available columns is too low with respect to the above need to have one different column available for each of the remaining $(M-1)$ rows. ∎

Lemma 11. *If $X = \check{A}\check{S}$, then the $M \times M$ matrix composed of the coefficients $e_{\ell j}$ defined in (14) and denoted as E reads*

$$E = \Lambda \Pi \tag{17}$$

where Λ is a diagonal matrix with non-zero elements on its diagonal, and Π is a permutation matrix.

Proof Combine Lemmas 8 and 10 (and see the organization in rows and columns in the proof of Lemma 10). Thus, if $X = \check{A}\check{S}$, then the above-defined matrix E contains exactly one non-zero coefficient per row and per column. This may be expressed as

$$E = \Lambda\Pi \tag{18}$$

where Λ is a diagonal matrix with non-zero elements on its diagonal, and Π is a permutation matrix. $\blacksquare$

In other words, the linear parts of the vectors $\check{s}_\ell$, corresponding to the first sum in (14), are equal to the vectors s_ℓ, up to scale and permutation indeterminacies.

4.5. Simplification of quadratic part of rows of $\check{S}$

Under the assumptions of Section 4.2, we now analyze the second-order terms of the vectors $\check{s}_\ell$, corresponding to the coefficients $e_{\ell jk}$ of the second sum in (14). To this end, we consider the third-order terms of the vectors products $\check{s}_\ell \odot \check{s}_m$, corresponding to the second and third sums (and lines) of (15).

Lemma 12. *If $X = \check{A}\check{S}$, then for any given integers ℓ and m with $1 \leq \ell < m \leq M$, a given vector $s_{i_1} \odot s_{i_2} \odot s_{i_3}$, i.e. with a given* unordered *set of indices $\{i_1, i_2, i_3\}$, appears for at most one set of values of j, j' and k' in the second line of (15).*

Proof For a direct proof of Lemma 12 when $M = 2$ or $M = 3$, see Appendix C. Instead, we hereafter provide a general proof applicable to any value of M.

When $X = \check{A}\check{S}$, and when ℓ and therefore $\check{s}_\ell$ are fixed, that vector $\check{s}_\ell$ contains exactly one linear term in the first sum of (14), due to Lemma 10. We denote as j_0 the index of that term (i.e., the linear term of $\check{s}_\ell$ is $e_{\ell j_0} s_{j_0}$, with $e_{\ell j_0} \neq 0$). The value of j_0 depends on ℓ. Since all other coefficients $e_{\ell j}$ with $j \neq j_0$ in the second sum of (15) are equal to zero, that sum reduces to

$$\sum_{j'=1}^{M-1} \sum_{k'=j'+1}^{M} e_{\ell j_0} e_{m j' k'} s_{j_0} \odot s_{j'} \odot s_{k'}. \tag{19}$$

Any given term in this sum, i.e. with given values of j' and k' in addition to j_0, yields a component associated with the vector $s_{j_0} \odot s_{j'} \odot s_{k'}$, and this vector is entirely defined by the unordered set of indices $\{j_0, j', k'\}$. The question we investigate here is: given such a set of indices $\{j_0, j', k'\}$, with

$$1 \leq j' < k' \leq M, \tag{20}$$

is there another term in the second sum of (15) which corresponds to the same vector ? In other words, is there another value of the couple of indices j' and k', denoted as j'' and k'', with

$$1 \leq j'' < k'' \leq M \tag{21}$$

and with $j'' \neq j'$ or $k'' \neq k'$, such that the associated unordered set $\{j_0, j'', k''\}$ is equal to the above fixed unordered set $\{j_0, j', k'\}$? We hereafter show that no such other couple exists, by successively considering all possible cases for the other potential *ordered* set (j'', k''), i.e. we consider all possible correspondences between each of the indices j'' and k'' and an index among j' and k'. We show that none of these cases can occur, because they are not compatible with the above-defined constraints. These cases are as follows.

Case 1: $j'' = j'$ and $k'' = k'$: these conditions are not compatible with the fact that the ordered set (j'', k'') is assumed to be different from the ordered set (j', k').

Case 2: $j'' = k'$ and $k'' = j'$: then (20) yields $j'' > k''$, which is not compatible with (21). $\blacksquare$

Theorem 1. *If $X = \check{A}\check{S}$, then in (14) the coefficients $e_{\ell j}$ meet the conditions of Lemma 11 and all coefficients $e_{\ell jk}$ are equal to zero. This reads as follows in matrix form, using (11) and (12):*

$$\check{S}_L = \Lambda \Pi S \tag{22}$$

where Λ is a diagonal matrix with non-zero elements on its diagonal and Π is a permutation matrix.

Proof The proof of this theorem is significantly longer than those of the lemmas and other theorems of this paper. It is therefore provided in an appendix: see Appendix D. $\blacksquare$

Theorem 1 is the main result of this section. In the above-defined conditions, it guarantees that, if $X = \check{A}\check{S}$, then the estimated source vectors $\check{s}_1$ to $\check{s}_M$ are equal to the actual source vectors s_1 to s_M up to scale and permutation indeterminacies[1]. Therefore, the indeterminacies resulting from the separation principle considered here are restricted to the same as those encountered with usual linear instantaneous BSS methods, although the mixing model is more complex here. Moreover, we stress that this identifiability property only requires the mild conditions of Section 4.2, that is essentially the linear independence of a set of vectors derived from the source vectors. If disregarding the fact that this set contains element-wise products of the original source vectors in addition to the latter vectors, the above identifiability conditions may be considered to be much less restrictive than those set by various usual classes of linear instantaneous BSS methods: briefly, the latter methods not only require linear independence but also (multi-lag) uncorrelatedness (related to orthogonality) or even statistical independence of the sources, or different additional types of properties, such as sparsity and/or nonnegativity. This may be interpreted as follows: the existence of a specific class of nonlinear terms in the mixing model considered here is an additional constraint, that we also impose on the associated "separating structure" and that allows us to reduce the other constraints set on the sources as compared with usual classes of linear instantaneous BSS methods, while preserving the same indeterminacies. In this sense, although mixture nonlinearity is usually considered to be an additional burden as compared with the linear case, it here turns out to be quite helpful when adequately used.

Moreover, the estimated mixing coefficients meet the following property:

Theorem 2. *If $X = \check{A}\check{S}$, then for any such matrix $\check{S}$ (which meets Theorem 1, as shown above), the corresponding matrix $\check{A}$ is unique.*

Proof Let us consider the case when $X = \check{A}\check{S}$, with a given matrix $\check{S}$. Then, due to Lemma 2, each row vector of X has a unique decomposition over the row vectors contained in $\check{S}$. For the overall matrix X, this yields a unique decomposition matrix $\check{A}$ such that $X = \check{A}\check{S}$. ∎

[1] In this paper, we investigate the analytical solution of the exact factorization problem $X = \check{A}\check{S}$. Numerical conditioning is a quite different problem, which would deserve a separate investigation if it were to be studied.

Theorem 1 yields a necessary condition for achieving $X = \check{A}\check{S}$. It may then be proved as follows that this condition is also sufficient:

Theorem 3. *If*

$$\check{S}_L = \Lambda \Pi S, \tag{23}$$

where Λ *is a diagonal matrix with non-zero elements on its diagonal, and* Π *is a permutation matrix, then there exists a unique matrix* $\check{A}$ *such that* $X = \check{A}\check{S}$.

Proof If (23) is met, then the (unordered) set of row vectors in matrix $\check{S}_L$ is equal to the set of row vectors in S, up to non-zero scale factors. The overall set of row vectors contained in $\check{S}$ is then equal to the overall set of row vectors in $\widetilde{S}$, up to non-zero scale factors. The row vectors of $\check{S}$ therefore span the same subspace as the row vectors of $\widetilde{S}$. Moreover, this subspace is $\int_X$, due to Lemma 1. Therefore, each row vector of X has a unique decomposition over the row vectors of $\check{S}$, i.e. there exists a unique matrix $\check{A}$ such that $X = \check{A}\check{S}$. $\blacksquare$

This completes the proof of the existence of a unique factorization (up to the above-defined scale factors and permutation) for the above bilinearly mixed sources. As the relevance of BMMF has thus been established, we hereafter show how corresponding practical algorithms may be derived.

5. Separation criteria and algorithms based on BMMF

5.1. Methods based on the above source-constrained separating structure

We here first consider the separating structure defined in Section 3 and we aim at defining practical methods for adapting its matrices $\check{A}$ and $\check{S}$. These methods are based on the separation principle introduced in Section 3, which consists in adapting $\check{A}$ and $\check{S}$ so that their product $\check{A}\check{S}$ fits the observation matrix X. Several criteria may be used to this end. The most natural one consists in minimizing the cost function

$$J_1 = ||X - \check{A}\check{S}||_F \tag{24}$$

(or its square), where $||.||_F$ stands for Frobenius norm. Moreover, a modified version of this BMMF method may be derived as follows.

5.2. Methods based on a doubly-constrained separating structure

In the above version of our methods, both $\check{A}$ and the top M rows of $\check{S}$ are master, i.e. independently updated, variables. However, since this adaptation aims at minimizing $J_1 = ||X - \check{A}\check{S}||_F$, a different adaptation scheme may be used. In this scheme, only the top M rows of $\check{S}$ are considered as master variables. In each occurence of the loop for updating $\check{S}$, the slave variable $\check{A}$ is set to its optimum value, i.e. to its value which minimizes $||X - \check{A}\check{S}||_F$ with respect to $\check{A}$ for the considered value of $\check{S}$. This optimum is nothing but the least squares (LS) solution, i.e. (assuming $\check{S}$ has full row rank) [23]

$$\check{A}_{opt} = X\check{S}^T(\check{S}\check{S}^T)^{-1}. \tag{25}$$

Setting $\check{A} = \check{A}_{opt}$ in (24), the cost function to be optimized (only with respect to the top M rows of $\check{S}$) becomes

$$J_2 = ||X(I - \check{S}^T(\check{S}\check{S}^T)^{-1}\check{S})||_F \tag{26}$$

(or its square). Using $\check{A} = \check{A}_{opt}$ is attractive, first because the number of master variables adapted when using $\check{A}_{opt}$ and therefore J_2 is much lower than when using J_1, so that the searched space has a much lower dimension, which may decrease computational time and improve convergence properties. Moreover, $\check{A}$ and J_2 are thus defined by a closed-form expression, which allows one to derive the gradient of J_2 with respect to the master part of $\check{S}$. This gradient may then be used in gradient-based optimization algorithms.

The last step of the development of BMMF methods consists in defining the considered separation algorithm(s). Various algorithms may be derived for minimizing the above cost functions J_1 or preferably J_2. As suggested above, this e.g. includes standard gradient descent and extended gradient-based minimization methods, that we will report elsewhere. Derivative-free optimization algorithms may also be used. In particular, the algorithm used hereafter to minimize J_2 is the Nelder-Mead (NM) method, as implemented in the fminsearch() Matlab function. The resulting version of our BMMF methods is therefore called BMMF-LS-NM. Moreover, several runs of this method may be combined to improve performance, as detailed in Section 6.2.

6. Numerical tests

6.1. Numerical validation of the assumptions of the separability analysis

The theoretical results of Section 4 were obtained under some assumptions concerning the considered data. We hereafter show that these assumptions are indeed met for actual data. To this end, we consider synthetic but realistic bilinear mixtures, corresponding to remote sensing applications such as those described in [18], [19], [20], [21]. We first use $M = 2$ source vectors s_1 and s_2, which are reflectance spectra derived from the USGS hyperspectral database. The considered two original USGS spectra are shown in Fig. 1 (each of their samples corresponds to one narrow spectral band). The resulting source vectors s_1 and s_2 here contain 10 samples, which corresponds to a typical situation for multispectral images. Each of these samples of s_1 and s_2 is derived as the average of 200 adjacent samples of an original USGS spectrum (we use the 2000 samples of the USGS spectra which have the lowest indices). These source spectra s_1 and s_2, and their product $s_1 \odot s_2$ are shown in Fig. 2 ($s_1 \odot s_2$ has a lower magnitude than s_1 and s_2, because the latter vectors contain reflectance values, which are therefore lower than 1). Using only 10 samples per source vector at this stage allows us to investigate numerical performance in the difficult case when limited information is available about the sources. Similarly, only 10 mixtures of these sources are used (this models 10 pixels of an observed image). Based on the physical mixing model derived in [20], the linear mixing coefficients a_{ij} are randomly, uniformly, drawn over $[0, 1]$ and then rescaled so as to sum to one in each observed vector, whereas the second-order coefficients b_{ijk} are uniformly drawn over $[0, 0.2]$.

In these conditions, matrix $\widetilde{S}$ consists of three rows, which respectively contain s_1, s_2 and $s_1 \odot s_2$. *Its rank is therefore expected to be equal to 3. This is confirmed by computing this rank with the Matlab rank() function.* Assumption 1 is therefore actually met for these data.

Besides, the set of vectors involved in Assumption 3 here consists of the 8 vectors $\{s_1, s_2, s_1 \odot s_1, s_1 \odot s_2, s_2 \odot s_2, s_1 \odot s_1 \odot s_2, s_1 \odot s_2 \odot s_2, s_1 \odot s_1 \odot s_2 \odot s_2\}$. *The rank of the matrix composed of these vectors is expected to be equal to 8. Here again, this is confirmed by the Matlab rank() function.* Assumption 3 is therefore actually met for these data.

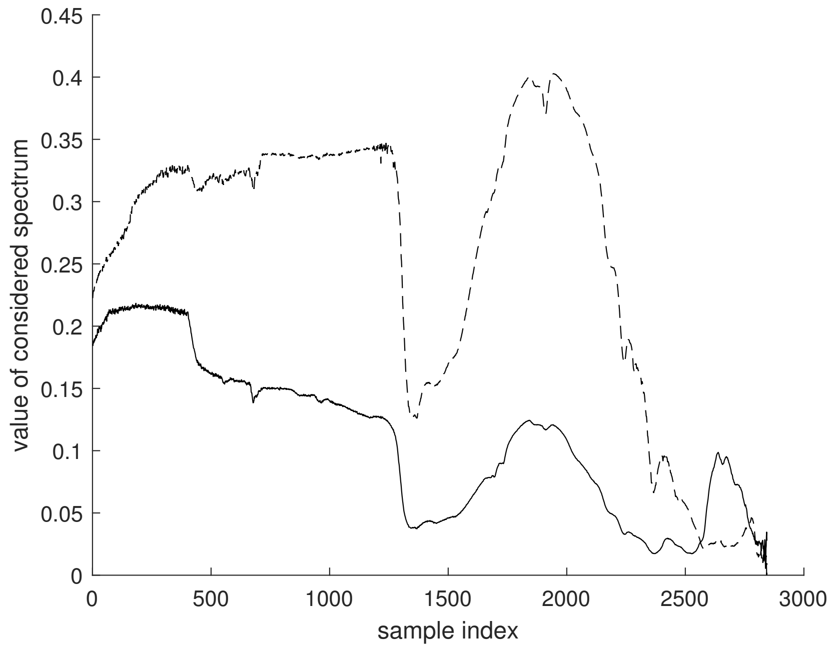

Figure 1: Two original spectra from the USGS database.

Finally, we created various matrices X of spectra mixed with random coefficients that meet the above-defined constraints. *In all cases, the rank of X computed by Matlab was equal to 3, as expected.* So, these data meet Assumption 2.

We then performed another series of tests, using $M = 4$ source vectors, with each of their 50 samples obtained as the average of 40 adjacent samples of an original spectrum from the USGS database. 50 mixtures of these sources are here used. *The ranks computed with Matlab for matrix $\widetilde{S}$, for the matrix composed of all the vectors involved in Assumption 3 and for matrix X with various sets of mixing coefficients were here respectively equal to 10, 49 and 10, which confirms that all assumptions of Section 4.2 are here again met* (see (13) and see Section 4.2 concerning the fact that the rank of the matrix composed of all the vectors involved in Assumption 3 is here equal to 49).

Finally, we performed another series of tests, again with $M = 2$ source vectors s_1 and s_2, but now with 100 samples in each of these vectors, which

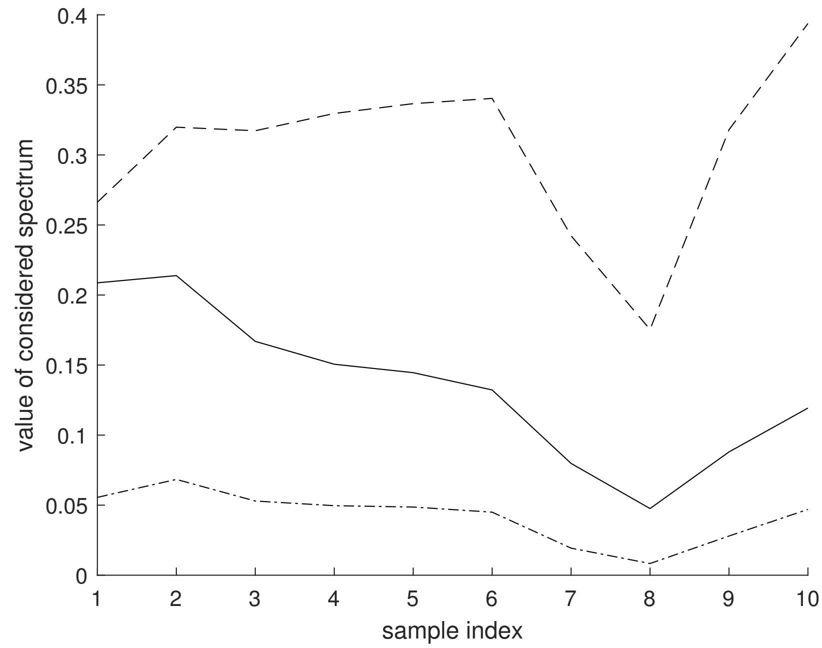

Figure 2: 10-sample source spectra s_1 (solid line) and s_2 (dashed line), and their element-wise product $s_1 \odot s_2$ (dash-dotted line).

corresponds to a typical situation for hyperspectral images and which allows us to investigate the numerical performance of our approach for high-dimensional data. Each of the samples of s_1 and s_2 is here derived as the average of 20 adjacent samples of the leftmost part of an USGS spectrum of Fig. 1. These source spectra s_1 and s_2, and their product $s_1 \odot s_2$ are shown in Fig. 3. Here again, 10 mixtures of these sources are created. *The ranks computed with Matlab for matrix $\widetilde{S}$, for the matrix composed of all the vectors involved in Assumption 3 and for matrix X with various sets of mixing coefficients were here respectively equal to 3, 8 and 3, which confirms that all assumptions of Section 4.2 are here again met.*

6.2. Numerical validation of separation algorithms with multispectral data

We then tested the performance of the BMMF-LS-NM separation method defined in Section 5.2. To this end, we first performed 100 Monte-Carlo tests

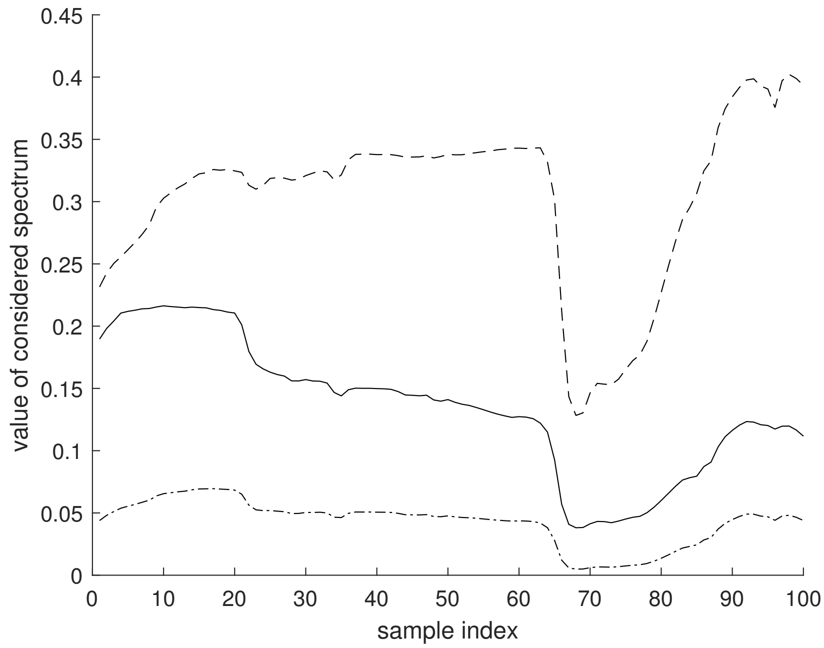

Figure 3: 100-sample source spectra s_1 (solid line) and s_2 (dashed line), and their element-wise product $s_1 \odot s_2$ (dash-dotted line).

with 10 mixtures of the two multispectral (i.e., 10-sample) sources s_1 and s_2 defined in the first part of Section 6.1. The master variables $\check{s}_1$ and $\check{s}_2$ of the BMMF-LS-NM method were initialized with values respectively equal to s_1 and s_2 plus random noise.

Performance is analyzed by computing two error parameters involving the value of $\check{S}$ obtained after our BMMF-LS-NM method converged and the associated value $\check{A}_{opt}$ of $\check{A}$ defined by (25). First, the normalized root-mean-square error for sources is defined as

$$E_{src} = \frac{\sqrt{\min\limits_{i \neq j \in \{1,2\}} (F_{ij})}}{\sqrt{||s_1||^2 + ||s_2||^2}} \tag{27}$$

where F_{ij} is equal to

$$\min_{\epsilon_1 = \pm 1} \left(||s_1 + \epsilon_1 \frac{||s_1||}{||\check{s}_i||} \check{s}_i||^2 \right) + \min_{\epsilon_2 = \pm 1} \left(||s_2 + \epsilon_2 \frac{||s_2||}{||\check{s}_j||} \check{s}_j||^2 \right).$$

Then, the normalized reconstruction error is defined as

$$E_{recons} = \frac{||X - \check{A}_{opt}\check{S}||_F}{||X||_F}. \tag{28}$$

The corresponding scatter plot in the (E_{src}, E_{recons}) plane, for all 100 Monte-Carlo tests, is shown in Fig. 4.

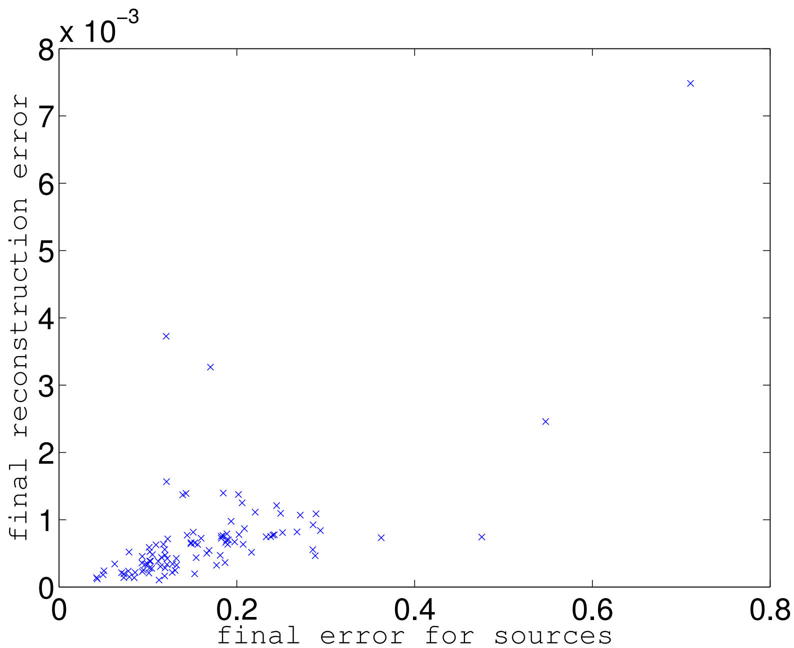

Figure 4: Scatter plot in (E_{src}, E_{recons}) plane, after convergence of the BMMF-LS-NM separation algorithm, for 10-sample spectra.

This figure shows that the BMMF-LS-NM method yields low errors E_{src} and E_{recons} in a large number of runs. In a few runs, however, E_{src} is higher. One may expect that, if one would instead use an algorithm which is able to converge to lower values of the cost function J_2 defined in (26), and hence of its normalized version E_{recons} defined in (28), then the error E_{src} for the sources would be lower than in the above tests (because, in the limit-case when the cost function J_2 is made equal to zero, our analytical analysis of Section 4 shows that the error E_{src} for the sources becomes equal to zero, too). Therefore, beyond the simple separation method obtained here by just applying a Matlab built-in optimization algorithm, one may aim at developing more advanced methods for the proposed BMMF framework, in order to achieve lower errors. This will require a detailed investigation, which is beyond the scope of the current paper. Yet, the potential directions of investigation for these future practical methods are outlined hereafter. These investigations will first consist in developing optimization algorithms associated with the considered separation criteria, thus especially requiring one to derive the analytical expressions of the gradients of the considered cost functions. Moreover, numerical investigations may be performed e.g. to investigate the sensitivity of the proposed iterative algorithms to initialization or the conditioning properties of the considered methods. Finally, applying metaheuristic methods to a set of runs of the elementary algorithms suggested here may be a way to make them insensitive to the lower performance obtained in rare runs which was observed above.

Still considering this idea of combining the results of elementary runs in order to improve performance, a much simpler approach than the above-mentioned one based on metaheuristics consists of the following steps:

- First perform a set of elementary runs.
- Then keep the "best" runs in the sense of a performance criterion that may be measured in a blind framework, i.e. when only knowing the observed data matrix X and the output of the separation algorithm. The normalized reconstruction error E_{recons} defined in (28), or equivalently the value of the cost function defined by (24) or (26) at the end of the iterations of the separation algorithm, are such performance criteria (on the contrary, the criterion E_{src} defined in (27) cannot be used in this blind framework, because the source signals are unknown). Keeping the best runs may be achieved either by ordering them according to the values of the considered performance criterion or by keeping all the runs for which the values of that criterion are lower than a given threshold.
- Eventually gather the results of all runs kept above, by computing, separately for each source vector to be estimated, the mean of all its estimates obtained in the runs which were kept above.

We applied this approach to the results of the 100 runs defined at the beginning of this Section 6.2, keeping the 10 best runs. The resulting average estimated source spectra are shown in Fig. 5, together with the actual source spectra. This shows that these estimated spectra are accurate enough to e.g. allow a remote sensing expert to visually derive the nature of the pure materials whose mixtures are observed in a given application, or to allow an automatic classifier to perform that pure material identification.

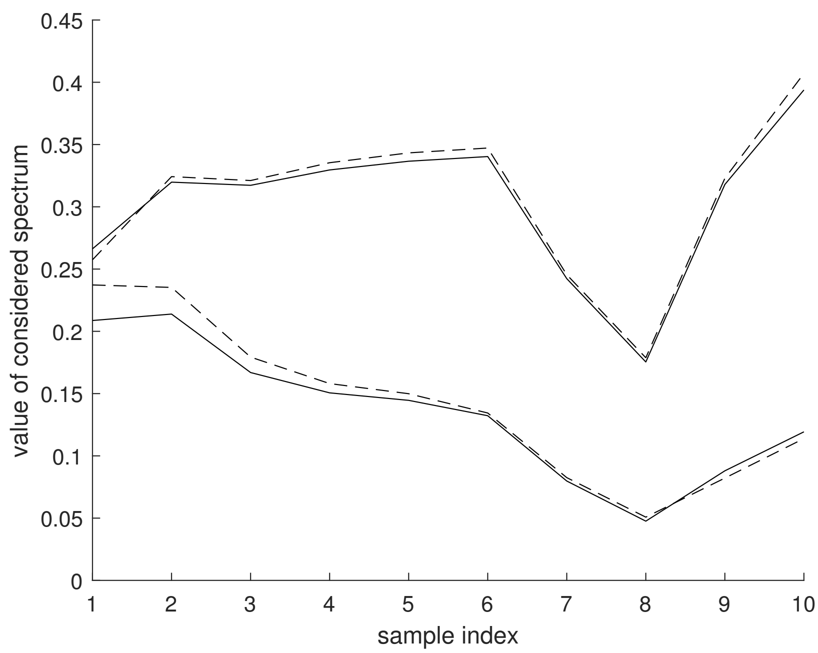

Figure 5: 10-sample source spectra s_1 and s_2 (solid lines) and their estimates obtained as the means of the outputs of selected runs of the BMMF-LS-NM method (dashed lines).

6.3. Numerical validation of separation algorithms with hyperspectral data

We here again consider 10 mixtures of the two hyperspectral (i.e., 100-sample) sources s_1 and s_2 defined in the last part of Section 6.1. Using the same protocol as in Section 6.2, we first performed 100 independent runs of the BMMF-LS-NM separation method with these data. The resulting performance is shown in Fig. 6. This yields the same type of comments as in Section 6.2, except that the phenomenon of rare cases with higher final errors E_{src} for sources which was observed in Section 6.2 is reduced here, possibly thanks to the richer information about these sources which is available here. We then applied the extended separation method of Section 6.2 to the results of the above 100 runs, keeping the 9 best runs. The resulting average estimated source spectra are shown in Fig. 7, together with the actual source spectra. In this high-dimensional configuration, these estimated spectra somewhat differ from the actual ones for the samples which have the lowest indices. However, the overall shapes of these estimated spectra are here again sufficiently similar to the shapes of the actual spectra to e.g. allow one to identify the corresponding pure materials.

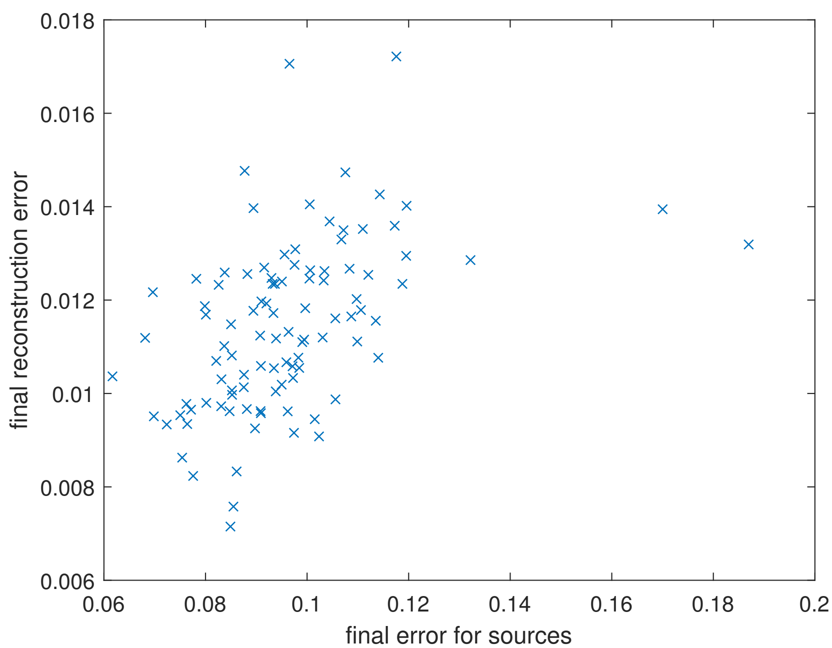

Figure 6: Scatter plot in (E_{src}, E_{recons}) plane, after convergence of the BMMF-LS-NM separation algorithm, for 100-sample spectra.

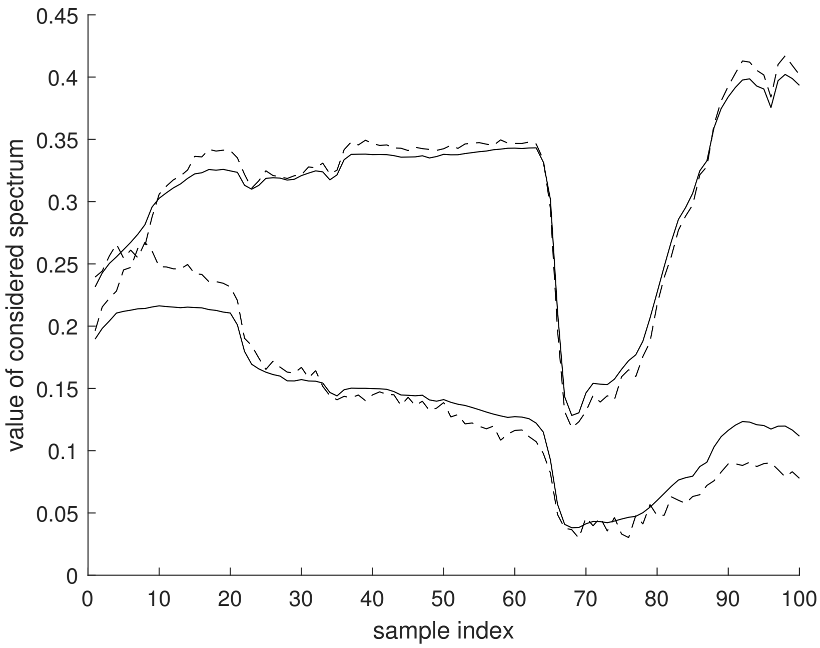

Figure 7: 100-sample source spectra s_1 and s_2 (solid lines) and their estimates obtained as the means of the outputs of selected runs of the BMMF-LS-NM method (dashed lines).

7. Extension to full linear-quadratic mixtures

An extension of the above bilinear mixing model (1) is the linear-quadratic one, which reads

$$x_i(n) = \sum_{j=1}^{M} a_{ij} s_j(n) + \sum_{j=1}^{M} \sum_{k=j}^{M} b_{ijk} s_j(n) s_k(n) \qquad \forall\ i \in \{1, \ldots, P\}. \tag{29}$$

This model thus contains additional terms, involving the squares of the source signals $s_j(n)$: these are the second-order auto-terms corresponding to $k = j$ in the second sum of (29). This model is thus nonlinear with respect to each source signal. On the contrary, the bilinear model (1) is linear with respect to each source signal separately (hence its name), despite its second-order cross-terms involving products of two *different* sources, that is $s_j(n)s_k(n)$ with $k \neq j$.

Starting from the scalar form (29) of the linear-quadratic model, associated matrix forms may be derived in the same way as in Section 2. This again eventually yields (10), but with an extended version of vector $p(n)$ and matrix B, which here also include additional entries for the above-defined second-order auto-terms.

In the same way as in Section 3, one may then introduce adaptive matrices $\check{A}$ and $\check{S}$ which respectively have the same structure as the extended version of $\widetilde{A}$ and $\widetilde{S}$ considered here. Hence, one may propose the separation principle which consists in adapting these extended matrices $\check{A}$ and $\check{S}$ so as to ideally achieve $\check{A}\check{S} = X$, thus introducing Linear-Quadratic Mixture Matrix Factorization, or LQMMF.

In Section 4, we showed that BMMF yields no spurious solutions. On the contrary, the following spurious solutions may be exhibited for LQMMF. For the sake of simplicity, let us consider the case when $M = 2$ and when the row vectors s_1, s_2, $s_1 \odot s_1$, $s_1 \odot s_2$ and $s_2 \odot s_2$ which here form $\widetilde{S}$ are linearly independent. Besides, let us consider the case when the two estimated source row vectors $\check{s}_1$ and $\check{s}_2$ are linear combinations of the actual source vectors s_1 and s_2, with non-zero combination coefficients which are arbitrary but such that the 2×2 matrix composed of these four coefficients is non-singular. It may then be shown that the row vectors $\check{s}_1$, $\check{s}_2$, $\check{s}_1 \odot \check{s}_1$, $\check{s}_1 \odot \check{s}_2$ and $\check{s}_2 \odot \check{s}_2$ which here form $\check{S}$ are linearly independent. Since they are here linear combinations of s_1, s_2, $s_1 \odot s_1$, $s_1 \odot s_2$ and $s_2 \odot s_2$, they thus form a basis of $\int_{\widetilde{S}}$. Then, each row of X may be decomposed over this basis. Denoting $\check{A}$ the matrix of coefficients of this decomposition, we thus get $\check{A}\check{S} = X$. This shows that, for LQMMF, the condition which defines the considered separation principle yields spurious solutions for $\check{s}_1$ and $\check{s}_2$, namely solutions which are not equal to the actual sources (up to the indeterminacies of BMMF), but which are *mixtures* of these sources.

The above result may be interpreted as follows with respect to the approach developed in Section 4 for bilinear mixtures. BMMF is more constraining than LQMMF is the sense that, to achieve $\check{A}\check{S} = X$, it does not allow the estimated source row vector products $\check{s}_\ell \odot \check{s}_m$ to contain terms proportional to second-order *auto-terms* $s_j \odot s_j$ with respect to the source vectors s_j, as shown in the proof of Lemma 6. On the contrary, when considering LQMMF, such terms are allowed by the considered mixing model itself, and we showed above that this entails spurious solutions.

8. Conclusion and future work

In this paper, we considered bilinear mixtures (this also applies to their subclass associated with [13]) and complete linear-quadratic mixtures, without nonnegativity constraints. We showed that they may be reformulated as products of two matrices, where the extended matrix related to sources contains values of sources and source products. We proved that the associated Bilinear Mixture Matrix Factorization (BMMF) separation principle yields a unique solution (up to the usual scale and permutation indeterminacies) under mild conditions, whereas Linear-Quadratic Mixture Matrix Factorization (LQMMF) yields spurious points. Although uniqueness is thus achieved only when the second-order terms of the mixing model are restricted to products of *different* sources, this result is of high interest, because of its practical applicability: second-order mixtures are especially faced in urban scenes in remote sensing applications [18, 19], [20], [21], where second-order terms most often consist of products of *different* reflectances (sources), since they correspond to different materials.

The nonlinearity of the mixing model may thus help constraining the solutions of matrix factorization and hence reducing required constraints on sources, as compared with (possibly constrained) linear NMF. Moreover, mixture nonlinearity may be combined with nonnegativity conditions to further constrain solutions or optimization algorithms in applications where the data meet such nonnegativity properties. This paper therefore contributes to proving the relevance of matrix factorization methods (with or without nonnegativity constraints) for second-order mixtures. We here started to define practical algorithms associated with the proposed BMMF framework and to check their performance. Various detailed investigations may then be performed to develop and test more powerful BMMF algorithms. We will report them in a future paper.

Appendix A. Separability/identifiability analyses available from the literature

The following separability/identifiability analyses were reported in the literature for the main classes of BSS/BMI methods intended for the standard (that is, linear instantaneous) mixing model, which were defined in Section 1. For ICA, Ref. [3] introduced specific BSS/BMI methods and proved that the indeterminacies entailed by the ICA separation principle are restricted to scale factors and permutation (for at most one Gaussian source). Within the framework of SCA, two major subclasses of methods were studied, depending on the considered type of sparsity properties. The first subclass is typically based on the minimization of L0 pseudo-norm or L1 norm. Related properties about the uniqueness of sparse decompositions were especially addressed in [12], [15] and references therein. The second subclass takes advantage of zones in (possibly transformed) signals where only one source is active, i.e. non-zero (see especially [1], [5], [7]). Such methods were shown to estimate the columns of the mixing matrix, and hence the source signals, up to scale factors and permutation, for determined mixtures [1], [5], [7]. Finally, the standard NMF separation principle yields high indeterminacies. Various investigations were devoted (i) to their analysis, (ii) to the determination of conditions in which they are reduced so that NMF factorization yields a unique solution up to standard, acceptable, underdeterminacies or (iii) to the development of extended methods aiming at reducing these indeterminacies (see e.g. [2], [11], [14], [16], [22]).

Appendix B. About the terminology

The terminology "Bilinear Mixture Matrix Factorization" used in this paper refers to the fact that the considered BSS/BMI approach consists in developing source and mixture estimators eventually based on a matrix product, for a mixing model which is *bilinear with respect to the source signals*: it contains products $s_j(n)s_k(n)$ of two (different) source signals, as detailed in the second (double) sum of terms in (1). The above source products are moreover multiplied by mixing coefficients b_{ijk} thus eventually yielding products $b_{ijk}s_j(n)s_k(n)$ of three terms, as shown in (1).

Similarly, the terminology “NMF” or “linear NMF”, used for standard NMF involving linear mixtures, refers to the fact that the considered mixing model is *linear with respect to the source signals*, i.e. it only consists of linear combinations $a_{ij}s_j(n)$ of the source signals (which also appear in the first sum of terms in (1)).

The above terminology is therefore coherent. It should be noted that a different terminology might be preferred if one would not focus on the dependence of the mixing model with respect to the sources, as was done above, but on the dependence of this model with respect to the overall set of parameters composed of both the sources and mixing coefficients. Then, when considering the above-mentioned standard NMF, in which each observed signal is the sum of terms $a_{ij}s_j(n)$ (as in the first sum of terms in (1)), one might define a terminology focused on the fact that these terms are *bilinear* with respect to the *overall* set of parameters composed of both the sources and mixing coefficients, as opposed to the emphasis on *linearity* with respect to the sources only (related to our focus on BSS), for this linear NMF, that we put in the terminology used in this paper.

Appendix C. Illustrations of the separation analysis

The analysis provided in Section 4 may be better understood by illustrating it with examples, focusing on low values of the number M of source signals. We therefore here consider the cases $M = 2$ and $M = 3$.

We first detail the “third-order source vector products” which appear in the second and third lines of (15). For its second line, each source vector product $s_j \odot s_{j'} \odot s_{k'}$ is defined by the corresponding indices j (which corresponds to expanding $\check{s}_\ell$ according to (14)), and j' and k' (which both correspond to expanding $\check{s}_m$ according to (14)). The complete set of their possible values is provided in Table C.1 (respectively C.2) for $M = 2$ (respectively $M = 3$). Similarly, for the third line of (15), each source vector product $s_j \odot s_k \odot s_{j'}$ is defined by the corresponding indices j and k (which both correspond to expanding $\check{s}_\ell$ according to (14)), and j' (which corresponds to expanding $\check{s}_m$ according to (14)). The complete set of their possible values is provided in Table C.3 (respectively C.4) for $M = 2$ (respectively $M = 3$).

j	(j', k')
	(1,2)
1	$\{1, 1, 2\}$
2	$\{1, 2, 2\}$

Table C.1: Indices of source vectors in second line of (15), for $M = 2$. Each row of this table corresponds to a possible value of j, and each column to a possible value of the ordered couple (j', k'). Each cell of the table contains the corresponding unordered set $\{j, j', k'\}$, rearranged according to increasing values.

j	(j', k')		
	(1,2)	(1,3)	(2,3)
1	$\{1, 1, 2\}$	$\{1, 1, 3\}$	$\{1, 2, 3\}$
2	$\{1, 2, 2\}$	$\{1, 2, 3\}$	$\{2, 2, 3\}$
3	$\{1, 2, 3\}$	$\{1, 3, 3\}$	$\{2, 3, 3\}$

Table C.2: Same as Table C.1, for $M = 3$.

Tables C.1 and C.2 yield a direct proof of Lemma 12, respectively for $M = 2$ and $M = 3$:

- For $M = 2$, Table C.1 shows that a given unordered set of indices $\{i_1, i_2, i_3\}$ (as defined in Lemma 12) appears at most once in the overall table, and therefore at most once for any given integers ℓ and m.

- For $M = 3$, the reader may have noted that some values of the unordered set of indices $\{i_1, i_2, i_3\}$ appear several times in Table C.2 (see the set $\{1, 2, 3\}$ on the antidiagonal). However, Lemma 12 does not concern the overall set of triplets $\{i_1, i_2, i_3\}$ obtained in the complete table, but the subset of such triplets corresponding to given integers ℓ and m. In other words, Lemma 12 concerns the triplets $\{j, j', k'\}$ of Table C.2 which correspond to a single (arbitrary) value of j (this is detailed in the general proof of Lemma 12 provided in Section 4.5, where this value of j is denoted as j_0;

(j,k)	j'	
	1	2
(1,2)	$\{1,1,2\}$	$\{1,2,2\}$

Table C.3: Indices of source vectors in third line of (15), for $M=2$. Each row of this table corresponds to a possible value of the ordered couple (j,k), and each column to a possible value of j'. Each cell of the table contains the corresponding unordered set $\{j,k,j'\}$, rearranged according to increasing values.

(j,k)	j'		
	1	2	3
(1,2)	$\{1,1,2\}$	$\{1,2,2\}$	$\{1,2,3\}$
(1,3)	$\{1,1,3\}$	$\{1,2,3\}$	$\{1,3,3\}$
(2,3)	$\{1,2,3\}$	$\{2,2,3\}$	$\{2,3,3\}$

Table C.4: Same as Table C.3, for $M=3$.

it depends on ℓ). The analysis should therefore be performed separately for each row of Table C.2. Then, that table actually shows that, in any of its rows, a given unordered set of indices $\{i_1, i_2, i_3\}$ appears at most once.

Finally, for $M=2$ and $M=3$, Tables C.1 to C.4 yield an alternative, direct, proof of the main property which was established for any M in the proof of Theorem 1 provided in Section 4.5, and which may here be expressed as follows. We analyze the case when $X = \check{A}\check{S}$. We consider a coefficient $e_{mj'k'}$ involved in a term of the second sum in (15), with given indices m, j' and k', and therefore with a given value of the index j'_0 that we associated with m in the proof of Theorem 1 in Section 4.5. The vector which corresponds to the coefficient $e_{mj'k'}$ in the second sum in (15) is $s_{j_0} \odot s_{j'} \odot s_{k'}$, where j_0 (corresponding to the index ℓ of $\check{s}_\ell$ in the proof of Theorem 1 in Section 4.5) is free at this stage, but different from j'_0. The unordered set of indices $\{j_0, j', k'\}$ associated with the above vector $s_{j_0} \odot s_{j'} \odot s_{k'}$ has the following relationship with the above tables: for any above-defined value of j_0 and for the considered fixed value of (j', k'), the unordered set of indices $\{j_0, j', k'\}$ is the content of the cell corresponding to row $j = j_0$ and column (j', k') in Table C.1 for $M = 2$ or Table C.2 for $M = 3$. The property of interest, that we established for any M in the proof of Theorem 1 in Section 4.5 and that we here aim at independently proving or illustrating in more detail for $M = 2$ and $M = 3$, may first be summarized as follows: in the above conditions, there exists at least one value j_0, with $j_0 \neq j'_0$, such that the above vector $s_{j_0} \odot s_{j'} \odot s_{k'}$ does not appear in the third sum of (15). Moreover, each term of the latter sum corresponds to the vector $s_j \odot s_k \odot s_{j'_0}$ and therefore to row (j, k) and column $j' = j'_0$ of Table C.3 for $M = 2$ (or Table C.4 for $M = 3$). Therefore, the above property may be restated as: there exists at least one value j_0, with $j_0 \neq j'_0$, such that the above unordered set of indices $\{j_0, j', k'\}$ does not appear in any of the cells corresponding to any row (j, k) and to the fixed column $j' = j'_0$ of Table C.3 for $M = 2$ (or Table C.4 for $M = 3$).

The validity of the above property for $M = 2$ may be directly checked as follows from Tables C.1 and C.3. We consider an arbitrary coefficient $e_{mj'k'}$, with $1 \leq j' < k' \leq M$, which only yields one possible case for the ordered couple (j', k'), namely $(j', k') = (1, 2)$ (this corresponds to the single column of Table C.1). Besides, depending on index m, its associated index $j'_0 \in \{1, \ldots, M\}$ here has two possible values. We successively consider each of them:

- If $j'_0 = 1$, then $j_0 = 2$, because $j_0 \in \{1, \ldots, M\}$ and $j_0 \neq j'_0$. In that case, the content of the cell corresponding to row $j = j_0$ and column (j', k') in Table C.1 is $\{1, 2, 2\}$. Moreover, the complete set of cells corresponding to any row (j, k) and to the fixed column $j' = j'_0$ of Table C.3 here similarly reduces to the single row and to column $j'_0 = 1$ of that table, which contains $\{1, 1, 2\}$. Therefore, the triplets of indices in "these cells" (that is, only one cell, here) of Table C.3 are indeed different from the triplets of indices in the considered single cell of Table C.1.

- The case when $j'_0 = 2$ is analyzed in the same way. Briefly, then $j_0 = 1$, the content of the cell corresponding to row $j = j_0$ and column (j', k') in Table C.1 is $\{1, 1, 2\}$, whereas the complete set of cells corresponding to any row (j, k) and to the fixed column $j' = j'_0$ of Table C.3 reduces to the single row and to column $j_0' = 2$ of that table, which contains $\{1, 2, 2\}$. Therefore, the triplets of indices in "these cells" of Table C.3 are indeed different from the triplet of indices in the considered single cell of Table C.1.

The above discussion also illustrates a property for $M = 2$ that was mentioned in the proof of Theorem 1 in Section 4.5: the index j_0' is necessarily equal to one of the indices j' and k' (whereas j_0 is different from j_0' and therefore equal to the other index among j' and k'). Therefore, for $M = 2$, whatever the considered coefficient $e_{mj'k'}$, we are in the second case defined in the proof of Theorem 1 provided in Section 4.5.

Similarly, the validity of the above property for $M = 3$ may be directly checked as follows from Tables C.2 and C.4. We consider an arbitrary coefficient $e_{mj'k'}$, with $1 \leq j' < k' \leq M$, which yields three possible values for the ordered couple (j', k'), namely $(1, 2)$, $(1, 3)$ and $(2, 3)$ (this corresponds to the three columns of Table C.2). Besides, depending on index m, its associated index $j_0' \in \{1, \ldots, M\}$ here has three possible values, namely 1, 2 and 3. For each given value of j_0', the index j_0 has two possible values, because $j_0 \in \{1, \ldots, M\}$ and $j_0 \neq j_0'$ (for instance, if $j_0' = 1$, then $j_0 = 2$ or $j_0 = 3$). For any such value of j_0, we consider the content of the cell corresponding to row $j = j_0$ and column (j', k') in Table C.2. Besides, we consider the complete set of cells corresponding to any row (j, k) and to the fixed column $j' = j_0'$ of Table C.4. Two cases then exist, depending on the values of the considered indices:

1. The first case considered here is the same as the first case introduced in the proof of Theorem 1 provided in Section 4.5. That case is defined by (D.3). In that case, the reader may check from the above tables that, whatever j_0 with $j_0 \neq j_0'$, the triplets of indices in the all above-defined cells of Table C.4 are indeed different from the triplet of indices in the considered single cell of Table C.2. For instance, when selecting $j_0' = 1, j' = 2, k' = 3$, one should consider the following two subcases:
   (a) First, $j_0 = 2$. The considered cell with row $j = j_0$ and column (j', k') in Table C.2 then contains $\{2, 2, 3\}$, whereas the considered cells corresponding to any row (j, k) and to column $j' = j_0'$ in Table

C.4 contain $\{1,1,2\}$, $\{1,1,3\}$, $\{1,2,3\}$ and are thus different from the above cell of Table C.2.

   (b) Then, $j_0 = 3$. The considered cell with row $j = j_0$ and column (j', k') in Table C.2 then contains $\{2,3,3\}$, whereas the considered cells corresponding to any row (j,k) and to column $j' = j'_0$ in Table C.4 are the same as above and thus contain $\{1,1,2\}$, $\{1,1,3\}$, $\{1,2,3\}$, so that they are different from the above cell of Table C.2.

2. The second case considered here is the same as the second case introduced in the proof of Theorem 1 provided in Section 4.5. That case is defined by (D.4). It deserves two comments:
   (a) If, as proposed in the proof of Theorem 1 provided in Section 4.5, one sets j_0 to the value, among j' and k', which is not equal to j'_0, then, for $M = 3$, the reader may check from the above tables that, *for this value of* j_0, the triplets of indices in the all above-defined cells of Table C.4 are indeed different from the triplet of indices in the considered single cell of Table C.2. For instance, if $j'_0 = 1$, $j' = 1$, $k' = 2$, one selects $j_0 = 2$. Then, the considered single cell of Table C.2 contains $\{1,2,2\}$, whereas the above-defined cells of Table C.4 contain $\{1,1,2\}$, $\{1,1,3\}$, $\{1,2,3\}$.
   (b) If, as opposed to the proof of Theorem 1 in Section 4.5, one would *not* set j_0 to the value, among j' and k', which is not equal to j'_0, then, for $M = 3$, the reader may check from the above tables that, for some of these values j_0, some of the triplets of indices in the all above-defined cells of Table C.4 would be equal to the triplet of indices in the considered single cell of Table C.2. For instance, let us keep $j'_0 = 1$, $j' = 1$, $k' = 2$ as in the above Subcase 2a) of the current section, but let us now select $j_0 = 3$. Then, the considered single cell of Table C.2 contains $\{1,2,3\}$, whereas the above-defined cells of Table C.4 still contain $\{1,1,2\}$, $\{1,1,3\}$, $\{1,2,3\}$, so that the content of the third of these cells is equal to the content of the considered cell of Table C.2.

   The present subcase 2b) shows that, in the framework of Case 2 of the present alternative proof of Theorem 1, not all values of j_0 yield the desired property. This is the reason why we had to prove that there exists at least one value of j_0 which does yield it. The value of j_0 that we exhibited to this end is the one considered in Subcase 2a) of Case 2 of the present alternative proof of Theorem 1.

Let us state again that we put the emphasis on the above property in the proof of Theorem 1 for the following reason. For any coefficient $e_{mj'k'}$, this property allows us to find at least one vector $s_{j_0} \odot s_{j'} \odot s_{k'}$ such that a single term is associated with this vector, and the corresponding factor is proportional to $e_{mj'k'}$ (and the other factor is non-zero). Since the contribution due to this vector must be zero in the considered conditions, this allows us to prove that $e_{mj'k'} = 0$.

Appendix D. Proof of Theorem 1

We analyze the case when $X = \check{A}\check{S}$. Let us consider a term of the second sum in (15), corresponding to a coefficient $e_{mj'k'}$, with given indices m, j' and k'. The index m and therefore the vector $\check{s}_m$ are fixed, and that vector $\check{s}_m$ contains exactly one linear term in the corresponding first sum of (14), due to Lemma 10. We denote as j'_0 the index of that term (i.e., the linear term of $\check{s}_m$ is $e_{mj'_0} s_{j'_0}$, with $e_{mj'_0} \neq 0$). The value of j'_0 depends on m.

Besides, the considered term in (15) is obtained for a certain vector $\check{s}_\ell$. Again due to Lemma 10, $\check{s}_\ell$ contains a single linear term, whose index is denoted as j_0 hereafter, i.e. the linear term of $\check{s}_\ell$ is $e_{\ell j_0} s_{j_0}$, with

$$e_{\ell j_0} \neq 0. \tag{D.1}$$

The value of j_0 thus depends on ℓ.

The considered third-order term in (15) thus corresponds to the vector $s_{j_0} \odot s_{j'} \odot s_{k'}$, which is defined by the *unordered* set of indices $\{j_0, j', k'\}$, with $1 \leq j_0 \leq M$ and $1 \leq j' < k' \leq M$. Moreover, we consider the case when $\ell \neq m$. Therefore,

$$j_0 \neq j'_0, \tag{D.2}$$

due to Lemma 11.

The indices that were initially fixed in this analysis are those of $e_{mj'k'}$, that is m, j' and k', or equivalently j'_0, j' and k'. At least two of these indices are different, namely j' and k'. Two cases may therefore exist and are successively considered hereafter (for a detailed analysis of these cases when $M = 2$ or $M = 3$, see Appendix C).

The first case is when all indices are different[2], i.e.

$$j'_0 \neq j' \qquad \text{and} \qquad j'_0 \neq k'. \tag{D.3}$$

We consider an arbitrary term in the second set of third-order terms, corresponding to the third sum in (15). For the vector $\check{s}_m$ (and $\check{s}_\ell$) defined above, this term of (15) corresponds to the vector $s_j \odot s_k \odot s_{j'_0}$ and therefore to the unordered set of indices $\{j, k, j'_0\}$. This set cannot be equal to the unordered set $\{j_0, j', k'\}$ considered above. This is due to the fact that the element j'_0 of the first set cannot belong to the second set, since it meets conditions (D.3) and (D.2). Let us note that this result applies whatever j_0 with $j_0 \neq j'_0$ (and whatever j, k).

The second case is when

$$j'_0 = j' \qquad \text{or} \qquad j'_0 = k' \tag{D.4}$$

and not both, because $j' \neq k'$. Then, let us denote as v_1 the value which is shared by j'_0 and j' or k', and v_2 the other value among j' and k'. Let us consider then case when $j_0 = v_2$ (ℓ can indeed be selected so as to meet this condition, since this is compatible with (D.2)). As in the first case, let us now consider the set of indices associated with the vector $s_{j_0} \odot s_{j'} \odot s_{k'}$ corresponding to a given coefficient $e_{mj'k'}$, that is $\{j_0, j', k'\}$, with $j' < k'$ and therefore $j' \neq k'$. Due to the above analysis, this set is here equal to the unordered set $\{v_2, v_1, v_2\}$ (because $j' = v_1$ and $k' = v_2$, or $j' = v_2$ and $k' = v_1$). Moreover, this set cannot be equal to any above-defined unordered set $\{j, k, j'_0\}$, i.e. for any j and k with $j < k$ and therefore

$$j \neq k, \tag{D.5}$$

[2] For $M = 2$ (only), this case cannot occur, because $j'_0 = 1$ or 2 and the ordered couple (j', k') can only take the value $(1, 2)$.

where the latter set is here equal to the set $\{j, k, v_1\}$. This results from the fact that, if these two sets were equal, then we would have $\{v_2, v_1, v_2\} = \{j, k, v_1\}$, and therefore $\{v_2, v_2\} = \{j, k\}$, which contradicts (D.5).

So, at this stage, we showed that the following result applies in all cases (i.e. in the above two cases, which are the only possible ones). Starting from an arbitrary coefficient $e_{mj'k'}$ corresponding to the second sum in (15), we can associate it with at least one value j_0 and therefore at least one vector $s_{j_0} \odot s_{j'} \odot s_{k'}$ so that this vector (with an arbitrary order for its three factors) appears only once in the second sum of (15), due to Lemma 12, and does not appear in the third sum of (15). Therefore, the component corresponding to this vector is restricted to $e_{\ell j_0} e_{mj'k'}$. As already shown in the proof of Lemma 3, since $X = \check{A}\check{S}$ here, each row vector of $\check{S}$ is a linear combination of all row vectors of $\widetilde{S}$. We here apply it to the vector $\check{s}_\ell \odot \check{s}_m$ with $\ell \neq m$ considered above. Then, due to Assumption 3, the above component $e_{\ell j_0} e_{mj'k'}$ of $s_{j_0} \odot s_{j'} \odot s_{k'}$ is equal to zero. Due to (D.1), this yields $e_{mj'k'} = 0$. Since this applies to any values of m, j' and k' (with $j' < k'$), this shows that, if $X = \check{A}\check{S}$, then the vectors $\check{s}_m$ defined by (14) with ℓ replaced by m contain no second-order terms, i.e. are restricted to their linear part obtained in Lemma 11.

Acknowledgment

This work was partly supported by the French ANR project "HYEP ANR 14-CE22-0016-01".

References

[1] F. Abrard, Y. Deville, "A time-frequency blind signal separation method applicable to underdetermined mixtures of dependent sources", Signal Processing, vol. 85, no. 7, pp. 1389-1403, July 2005.

[2] A. Cichocki, R. Zdunek, A. H. Phan, S.-I. Amari, "Nonnegative matrix and tensor factorizations. Applications to exploratory multi-way data analysis and blind source separation", Wiley, Chichester, UK, 2009.

[3] P. Comon, "Independent Component Analysis, a new concept ?", Signal Processing, vol. 36, no. 3, pp. 287-314, April 1994.

[4] P. Comon and C. Jutten Eds, “Handbook of blind source separation. Independent component analysis and applications”, Academic Press, Oxford, UK, 2010.

[5] Y. Deville, M. Puigt, ”Temporal and time-frequency correlation-based blind source separation methods. Part I: determined and underdetermined linear instantaneous mixtures”, Signal Processing, vol. 87, no. 3, pp. 374-407, March 2007.

[6] Y. Deville, “Traitement du signal : signaux temporels et spatiotemporels - ... séparation aveugle de sources”, Ellipses Editions Marketing, Paris, France, 2011.

[7] Y. Deville, ”Chapter 6. Sparse component analysis: a general framework for linear and nonlinear blind source separation and mixture identification”, in ”Blind Source Separation: Advances in Theory, Algorithms and Applications”, G. Naik and W. Wang Eds, pp. 151-196, Springer, Berlin, Germany, 2014.

[8] Y. Deville, L. T. Duarte, “An overview of blind source separation methods for linear-quadratic and post-nonlinear mixtures”, Proc. of the 12th Int. Conference on Latent Variable Analysis and Signal Separation (LVA/ICA 2015), Springer International Publishing Switzerland, vol. LNCS 9237, pp. 155-167, Liberec, Czech Republic, Aug. 25-28, 2015.

[9] Y. Deville, ”Matrix factorization for bilinear blind source separation: methods, separability and conditioning”, Proceedings of the 23rd European Signal Processing Conference (EUSIPCO 2015), pp. 1945-1949, Nice, France, Aug. 31 - Sept. 4, 2015.

[10] Y. Deville, “Blind source separation and blind mixture identification methods”, Wiley Encyclopedia of Electrical and Electronics Engineering, pp. 1-33, J. Webster (ed.), Wiley, 2016. DOI: 10.1002/047134608X.W8300

[11] D. Donoho, V. Stodden, “When does Non-Negative Matrix Factorization give a correct decomposition into parts?” Advances in Neural Information Processing Systems 16 (NIPS 2003), Vancouver and Whistler, Canada, December 8-13, 2003.

[12] D. L. Donoho, “For most large underdetermined systems of equations, the minimal l^1-norm near-solution approximates the sparsest near-Solution”, *Technical Report*, Stanford University, 2004. Available at http://statweb.stanford.edu/∼donoho/Reports/2004/l1l0approx.pdf .

[13] O. Eches, M. Guillaume, “A bilinear-bilinear nonnegative matrix factorization method for hyperspectral unmixing”, IEEE Geoscience and Remote Sensing Letters, vol. 11, no. 4, pp. 778-782, April 2014.

[14] N. Gillis, “Sparse and Unique Nonnegative Matrix Factorization Through Data Preprocessing”, Journal of Machine Learning Research, vol. 13, pp. 3349-3386, 2012.

[15] R. Gribonval, M. Nielsen, “Sparse representations in unions of bases”, IEEE Transactions on Information Theory vol. 49, no. 12, pp. 3320-3325, Dec. 2003.

[16] K. Huang, N. D. Sidiropoulos, A. Swami, “Non-Negative Matrix Factorization Revisited: Uniqueness and Algorithm for Symmetric Decomposition”, IEEE Transactions on Signal Processing, vol. 62, no. 1, pp. 211-224, Jan. 1, 2014.

[17] Q. Liu, W. Wang, “Show-through removal for scanned images using nonlinear NMF with adaptive smoothing”, Proceedings of 2013 IEEE China Summit & International Conference on Signal and Information Processing (ChinaSIP), pp. 650-654, Beijing, China, July 6-10, 2013.

[18] I. Meganem, P. Déliot, X. Briottet, Y. Deville, S. Hosseini, “Physical modelling and non-linear unmixing method for urban hyperspectral images”, Proc. of the Third Workshop on Hyperspectral Image and Signal Processing (WHISPERS 2011), Lisbon, Portugal, June 6-9, 2011.

[19] I. Meganem, Y. Deville, S. Hosseini, P. Déliot, X. Briottet, L. T. Duarte, “Linear-quadratic and polynomial Non-negative Matrix Factorization; application to spectral unmixing”, Proceedings of the 19th European Signal Processing Conference (EUSIPCO 2011), Barcelona, Spain, Aug. 29 - Sept. 2, 2011.

[20] I. Meganem, P. Déliot, X. Briottet, Y. Deville, S. Hosseini, "Linear-quadratic mixing model for reflectances in urban environments", IEEE Transactions on Geoscience and Remote Sensing, vol. 52, no. 1, pp. 544-558, Jan. 2014.

[21] I. Meganem, Y. Deville, S. Hosseini, P. Déliot, X. Briottet, "Linear-quadratic blind source separation using NMF to unmix urban hyperspectral images", IEEE Transactions on Signal Processing, vol. 62, no. 7, pp. 1822-1833, April 1, 2014.

[22] S. Moussaoui, D. Brie, J. Idier, "Non-negative source separation: range of admissible solutions and conditions for the uniqueness of the solution", Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05), pp. V-289 - V-292, Philadelphia, USA, March 19-23, 2005.

[23] A. Smilde, R. Bro, P. Geladi, "Multi-way analysis with applications in the chemical sciences", Wiley, Chichester, England, 2004.