Source Separation of Multimodal Data: A Second-Order Approach Based on a Constrained Joint Block Decomposition of Covariance Matrices

Bertrand Rivet

Source separation of multimodal data: a second-order approach based on a constrained joint block decomposition of covariance matrices

Bertrand Rivet

Abstract—**Blind source separation aims at extracting unknown sources from mixtures of them. When multimodal data are considered (i.e. multi-set or multi-kind), some joint analysis are needed, for instance multi-set canonical correlation analysis or independent vector analysis. However, these methods only consider unidimensional sources in each set/modality. In this letter, an approach for dealing with multidimensional sources in each modality is derived. It assumes that the underlying dimensions in each modality for each source are known and it is based on a piecewise second order stationary model. Based on the likelihood, a contrast function is derived for the Gaussian case and is shown to be a constrained joint block decomposition of covariance matrices. Numerical simulations exhibit the merit of using a few number of modalities: it improves the quality of the separation and reduces the variance on the estimates. Finally, the proposed method outperforms the multi-set canonical correlation analysis and the independent component analysis applied to each individual modality followed by a clustering.**

Index Terms—**Blind source separation, multimodal data, multidimensional signals, joint block matrix decomposition**

I. Introduction

Nowadays, it is classical to record a physical phenomenon *at the same time* with several kinds of sensors: for instance, brain activities can be recorded using electroencephalography (EEG), magnetoencephalography (MEG) and functional near-infrared spectroscopy (NIRS). Moreover, the recorded signals can be contaminated by electrocardiograms (ECG) artifacts and/or eye movements artifacts (e.g., blinks). As a consequence, one deals with *multimodal*[1] data, i.e. the algorithms can process *simultaneously* several sets of data recorded by several kinds of sensors to improve the estimation compared to the use of a single modality: for instance, one can also use extra ECG sensors placed on the chest and/or an eye-tracker to improve the quality of artifacts estimation compared to use only EEG or only MEG sensors. It is worth noting that multimodal data (i.e. data recorded from several kinds of sensors) are not similar to multidimensional data (i.e. data recorded from several sensors of the same kind).

Moreover, the recorded signals are often mixtures of several initial sources of interest which must be estimated to analyze the process. This problem can be tackled by the blind source separation framework (BSS) [1]. In particular, independent component analysis (ICA) has been developed to separate I sources $s_i(t) \in \mathbb{R}$, $i \in \{1, \ldots, I\}$ that are statistically mutually independent [2], [1]. More recently two extensions of ICA have been proposed to address two distinct practical problems. Firstly, when considering simultaneously two or more sets of data, the joint blind source separation (joint-BSS) aims to achieve a separation of a multi-set data such that the I recovered sources $s_i^{[k]}(t) \in \mathbb{R}$ are aligned for each dataset $k \in \{1, \ldots, K\}$, with K the total number of datasets. Joint-BSS can be obtained for example by multi-set canonical correlation analysis (MCCA) [3], [4], mCCA+joint-ICA [5] or independent vector analysis (IVA) [6]. Secondly, the separation of I multidimensional sources $\mathbf{s}_i(t) \in \mathbb{R}^{n_i}$, with n_i the dimension of the ith source, assuming that multidimensional sources are mutually independent (i.e. inter-source independence) while allowing dependence between each source components (i.e. intra-source dependence) relies on multidimensional independence component analysis (MICA) [7] or independent subspace analysis [8].

The multimodal separation of multidimensional sources tackled in this letter can be seen, in a very general way, as a merging of these two extensions: processing simultaneously K sets of mixtures of I multidimensional sources $\mathbf{s}_i^{[k]}(t) \in \mathbb{R}^{n_i^{[k]}}$, $k \in \{1, \ldots, K\}$ and $i \in \{1, \ldots, I\}$. The proposed multimodal multidimensional blind source separation (MM-BSS) approach extends the IVA since the sources are not necessarily present in all sets and the dimension of the sources can be different in each modality. The proposed algorithm to separate the sources is based on a constrained joint block decomposition of covariance matrices.

The remainder of this letter is organized as follows. Section II describes the modeling of multimodal source separation and the proposed algorithm to extract the multidimensional sources. Section III presents the numerical results, before the conclusions and perspectives in Section IV.

II. Multimodal source separation

From the proposed modeling of MM-BSS (Section II-A), a likelihood approach is used to express a contrast function assuming that sources are piecewise second-order stationary signals (Section II-B) by generalizing models of unimodal unidimensional sources [9] and of unimodal multidimensional sources [10], [11]. Finally, a constrained gradient based approach is expressed to optimize it (Section II-C).

B. Rivet is with the GIPSA-lab, CNRS UMR 5216, Grenoble Institute of Technology, Grenoble, France. e-mail: bertrand.rivet@gipsa-lab.grenoble-inp.fr (see http://www.gipsa-lab.grenoble-inp.fr).
This work has been partially funded by ERC AdG-2012-320864-CHESS.

[1]Note that in this article, multimodal does not refer to the statistical meaning: i.e., it does not refer to a probability distribution function with several modes, but to several sets of data or several modalities used to record the data.

A. Multimodal multidimensional modeling

Let us consider K modalities and I multimodal multidimensional sources: $\mathbf{s}_i^{[k]}(t) \in \mathbb{R}^{n_i^{[k]}}$ denotes the vector of length $n_i^{[k]}$ related to the ith source in the modality indexed k at time t. In each modality, the recorded signals are considered to be instantaneous linear mixtures of these sources

$$\mathbf{x}^{[k]}(t) = A^{[k]}\mathbf{s}^{[k]}(t), \quad \forall k \in \{1, \ldots, K\}, \tag{1}$$

where $\mathbf{s}^{[k]}(t) = [\mathbf{s}_1^{[k]^\dagger}(t), \ldots, \mathbf{s}_I^{[k]^\dagger}(t)]^\dagger$ and $A^{[k]} \in \mathbb{R}^{n^{[k]} \times n^{[k]}}$ are the source vector and the full-rank mixing matrix of the kth modality, respectively. $n^{[k]} = \sum_{i=1}^{I} n_i^{[k]}$, and $\cdot^\dagger$ denotes the transpose operator. Model (1) is recast into components by

$$\mathbf{x}^{[k]}(t) = \sum_{i=1}^{I} A_i^{[k]}\mathbf{s}_i^{[k]}(t), \quad \forall k \in \{1, \ldots, K\}, \tag{2}$$

where $A_i^{[k]} \in \mathbb{R}^{n^{[k]} \times n_i^{[k]}}$ are column mixing sub matrices related to the ith source of the kth modality so that $A^{[k]} = [A_1^{[k]}, \ldots, A_I^{[k]}]$. Moreover, by denoting the full multimodal source vector $\mathbf{s}(t) = [\mathbf{s}^{[1]^\dagger}(t), \ldots, \mathbf{s}^{[K]^\dagger}(t)]^\dagger$, model (1) is written as

$$\mathbf{x}(t) = A\mathbf{s}(t), \tag{3}$$

where A is a block diagonal matrix whose ith diagonal block is equal to $A^{[k]}$. Moreover, let us introduce the multimodal vector of the ith source $\tilde{\mathbf{s}}_i(t) = [\mathbf{s}_i^{[1]^\dagger}(t), \ldots, \mathbf{s}_i^{[K]^\dagger}(t)]^\dagger$, and $\tilde{\mathbf{s}}(t) = [\tilde{\mathbf{s}}_1(t), \ldots, \tilde{\mathbf{s}}_I(t)]^\dagger$. Therefore, $\mathbf{s}(t)$ and $\tilde{\mathbf{s}}(t)$ are equal up to a permutation matrix L: $\tilde{\mathbf{s}}(t) = L\mathbf{s}(t)$.

Finally, the MM-BSS problem corresponds to the estimation of a set of K demixing matrices $B^{[k]}$ such that the components $\mathbf{y}_i^{[k]}(t)$ of

$$\mathbf{y}^{[k]}(t) = B^{[k]}\mathbf{x}^{[k]}(t), \tag{4}$$

where $\mathbf{y}^{[k]}(t) = [\mathbf{y}_1^{[k]^\dagger}(t), \ldots, \mathbf{y}_I^{[k]^\dagger}(t)]^\dagger$ with $\mathbf{y}_i^{[k]}(t) \in \mathbb{R}^{n_i^{[k]}}$, lie in the same subspaces as $\mathbf{s}_i^{[k]}(t)$, $\forall k \in \{1, \ldots, K\}$. Indeed, this problem suffers from severe indeterminacies as the original unimodal unidimensional or multidimensional blind source separation problems: any set of $n_i^{[k]} \times n_i^{[k]}$ invertible matrices $W_i^{[k]}$ $(1 \leq k \leq K)$ leads to the same problem (2) when right multiplying $A_i^{[k]}$ by $W_i^{[k]}$ and left multiplying $\mathbf{s}_i^{[k]}(t)$ by $W_i^{[k]^{-1}}$. It is worth noting that, in the separation formulation (4), the estimation of the set of matrices $B^{[k]}$ is equivalent to estimate the mixing matrices $A^{[k]}$ since $B^{[k]} = \left(A^{[k]}\right)^{-1}$ up to block permutations corresponding to the arbitrary order of the sources.

B. Likelihood expression and contrast function

Let us consider that sources are piecewise second-order stationary signals and that the partition of the observation interval $[1, T]$ into P domains $\mathcal{D}_p,\ p \in \{1, \ldots, P\}$, is known. Each domain $\mathcal{D}_p$ contains T_p samples such that $T = \sum_{p=1}^{P} T_p$. Finally, the sources $\mathbf{s}_i^{[k]}(t)$ are assumed to be temporally white, mutually decorrelated with a zero mean and wide-sense stationary on $\mathcal{D}_p$:

$$E\Big[\mathbf{s}_i^{[k]}(t)\mathbf{s}_i^{[k]^\dagger}(t')\Big] = 0, \qquad \text{if } t \neq t',\ \forall i,\ \forall k \tag{5}$$

$$E\Big[\mathbf{s}_i^{[k]}(t)\Big] = 0, \qquad \forall t,\ \forall i,\ \forall k \tag{6}$$

$$E\Big[\mathbf{s}_i^{[k]}(t)\mathbf{s}_j^{[l]^\dagger}(t)\Big] = 0, \qquad \text{if } i \neq j,\ \forall t,\ \forall (k,l) \tag{7}$$

$$E\Big[\mathbf{s}_i^{[k]}(t)\mathbf{s}_i^{[l]^\dagger}(t)\Big] = R^{(p)}_{s_i^{[k]}, s_i^{[l]}}, \quad \forall t \in \mathcal{D}_p,\ \forall i,\ \forall (k,l) \tag{8}$$

where $E[\cdot]$ is the expectation operator, $R^{(p)}_{s_i^{[k]}, s_i^{[l]}}$ is the multimodal covariance matrix of source i between modalities k and l on domain $\mathcal{D}_p$. Assumptions (7) and (8) can be recast in a more elegant global form by

$$E\big[\tilde{\mathbf{s}}(t)\tilde{\mathbf{s}}^\dagger(t)\big] = R^{(p)}_{\tilde{\mathbf{s}},\tilde{\mathbf{s}}}, \quad \forall t \in \mathcal{D}_p, \tag{9}$$

where the full multimodal covariance matrix of the sources $R^{(p)}_{\tilde{\mathbf{s}},\tilde{\mathbf{s}}}$ is block diagonal, each block is of size $n_i = \sum_{k=1}^{K} n_i^{[k]}$. Therefore the multimodal covariance matrix of sources $\mathbf{s}(t)$ is expressed by

$$E\big[\mathbf{s}(t)\mathbf{s}^\dagger(t)\big] = R^{(p)}_{\mathbf{s},\mathbf{s}} = L^\dagger R^{(p)}_{\tilde{\mathbf{s}},\tilde{\mathbf{s}}} L, \quad \forall t \in \mathcal{D}_p, \tag{10}$$

and has a structure denoted $\mathrm{Struct}_{\mathrm{R}_{\mathbf{s},\mathbf{s}}} \equiv L^\dagger \mathrm{BDiag}_{R_{\tilde{\mathbf{s}},\tilde{\mathbf{s}}}} L$, where $\mathrm{BDiag}_{.}$ is the structure of a block diagonal matrix defined by the indexed matrix. This means that $R^{(p)}_{\mathbf{s},\mathbf{s}}$ has the structure of a block diagonal matrix left and right multiplied by the permutation matrix L.

From these assumptions and with normally distributed sources $\mathbf{s}_i^{[k]}(t)$, the log-likelihood of the described model is

$$\mathcal{L}(\{\{\mathbf{x}^{[k]}(t)\}_{t=1}^{T}\}_{k=1}^{K}; \{A^{[k]}\}_{k=1}^{K}, \{R^{(p)}_{\mathbf{s},\mathbf{s}}\}_{p=1}^{P}) = -\frac{1}{2}\sum_{p=1}^{P} T_p\Big[\log\det\big(2\pi R^{(p)}_{\mathbf{x},\mathbf{x}}\big) + \mathrm{Tr}\big(\hat{R}^{(p)}_{\mathbf{x},\mathbf{x}} R^{(p)^{-1}}_{\mathbf{x},\mathbf{x}}\big)\Big], \tag{11}$$

where $\mathrm{Tr}(\cdot)$ is the trace operator, $R^{(p)}_{\mathbf{x},\mathbf{x}}$ is the full covariance matrix of the multimodal multidimensional observations $\mathbf{x}(t) = [\mathbf{x}^{[1]^\dagger}(t), \ldots, \mathbf{x}^{[K]^\dagger}(t)]^\dagger$ and

$$\hat{R}^{(p)}_{\mathbf{x},\mathbf{x}} = \frac{1}{T_p}\sum_{t \in \mathcal{D}_p} \mathbf{x}(t)\mathbf{x}^\dagger(t)$$

is its estimate. From linear model (3), the likelihood (11) can be rewritten as

$$\mathcal{L}(\{\{\mathbf{x}^{[k]}(t)\}_{t=1}^{T}\}_{k=1}^{K}; \{A^{[k]}\}_{k=1}^{K}, \{R^{(p)}_{\mathbf{s},\mathbf{s}}\}_{p=1}^{P}) = -T\Big\langle D\big(A^{-1}\hat{R}^{(p)}_{\mathbf{x},\mathbf{x}}A^{-\dagger}, R^{(p)}_{\mathbf{s},\mathbf{s}}\big)\Big\rangle_{p=1}^{P} + \kappa, \tag{12}$$

with $\cdot^{-\dagger}$ the inverse of the transposed matrix and

$$\big\langle M^{(p)}\big\rangle_{p=1}^{P} = \frac{1}{T}\sum_{p=1}^{P} T_p M^{(p)},$$

and where for any $m \times m$ positive definite matrices R_1, R_2

$$D(R_1, R_2) = \frac{1}{2}\Big(\mathrm{Tr}(R_1 R_2^{-1}) - \log\det(R_1 R_2^{-1}) - m\Big),$$

and

$$\kappa = -\frac{1}{2}\Big(nT + \sum_{p=1}^{P} T_p \log\det\big(2\pi\hat{R}^{(p)}_{\mathbf{x},\mathbf{x}}\big)\Big),$$

with $n = \sum_{i=1}^{I} n_i = \sum_{k=1}^{K} n^{[k]} = \sum_{i=1}^{I}\sum_{k=1}^{K} n_i^{[k]}$. The contrast function [2], [1] is then expressed as

$$C(A) = \left\langle D\big(A^{-1}\hat{R}^{(p)}_{\mathbf{x},\mathbf{x}}A^{-\dagger}, \mathrm{Struct}_{R_{\mathbf{s},\mathbf{s}}}\big(A^{-1}\hat{R}^{(p)}_{\mathbf{x},\mathbf{x}}A^{-\dagger}\big)\big)\right\rangle_{p=1}^{P}, \tag{13}$$

where $\mathrm{Struct}_M(\cdot)$ is the operator which orthogonally projects matrix $\cdot$ onto matrix space of structure similar to M (i.e. this operator puts to zero all entries outside the structure of matrix M) and $\hat{R}^{(p)}_{\mathbf{s},\mathbf{s}} = \mathrm{Struct}_{R_{\mathbf{s},\mathbf{s}}}\big(A^{-1}\hat{R}^{(p)}_{\mathbf{x},\mathbf{x}}A^{-\dagger}\big)$. It is worth noting that multimodal mixing matrix A is constrained to be block-diagonal.

C. Constrained relative gradient based approach

The aim of MM-BSS thus leads to the following constrained optimization problem

$$\hat{A} = \arg\min_{A} C(A), \quad \text{s.t. } A \in \mathcal{BD}iag_{\mathbf{n}}, \tag{14}$$

where $\mathcal{BD}iag_{\mathbf{n}}$ is the set of block-diagonal matrices defined by pattern $\mathbf{n} = [n^{[1]}, \ldots, n^{[K]}]^{\dagger}$ (i.e. the size of the kth diagonal block of A is equal to $n^{[k]}$). It is worth noting that an unconstrained joint block diagonalization (JBD) algorithm cannot be used since the estimation of the mixing matrix A will also contain some cross-modalities term (i.e. non null term outside the block diagonal) that are undesired.

Optimizing criterion (14) is achieved by a relative gradient based approach: the iterative estimation $A_{(l)}$ of A at the lth iteration is expressed as

$$A_{(l)} = A_{(l-1)}\Big(\mathbb{I} - \lambda_{(l-1)}\nabla C\big(A_{(l-1)}\big)\Big), \tag{15}$$

where $\lambda_{(l-1)} > 0$ is the step of the gradient method and $\nabla C(A)$ is the relative gradient [9] of (13) derived as

$$\nabla C(A) = -\Big\langle \mathrm{BDiag}_A\Big(\mathrm{Struct}^{-1}_{R_{\mathbf{s},\mathbf{s}}}\big(A^{-1}\hat{R}^{(p)}_{\mathbf{x},\mathbf{x}}A^{-\dagger}\big)\times \big(A^{-1}\hat{R}^{(p)}_{\mathbf{x},\mathbf{x}}A^{-\dagger}\big)\Big)\Big\rangle_{p=1}^{P} + \mathbb{I}, \tag{16}$$

with $\mathbb{I}$ the identity matrix and $\mathrm{Struct}_{\cdot}^{-1}(\cdot)$ is the inverse of the resulting matrix. In this study, the choice of $\lambda_{(l)}$ is set by backtracking using the Wolfe conditions [12].

It is worth noting from (13), that MM-BSS can be interpreted as a joint block decomposition of a set of multimodal covariance matrices, since $D(\cdot,\cdot)$ is a measure of discrepancy between $A^{-1}\hat{R}^{(p)}_{\mathbf{x},\mathbf{x}}A^{-\dagger}$ and $\hat{R}^{(p)}_{\mathbf{s},\mathbf{s}}$.

Moreover, one has to know the dimensions $n_i^{[k]}$ of each source i in each modality k to set the constraints on the structure of the multimodal covariance matrices $R^{(p)}_{\mathbf{s},\mathbf{s}}$. This can be done in a semi-blind approach from the knowledge on the experiment. However, in a fully blind context, these informations are unknown and need to be estimated from the data. This latter point is out of the scope of this letter and will be addressed in later studies.

III. NUMERICAL EXPERIMENTS

First, to assess the quality of the estimation, a performance index is presented (Section III-A). Two numerical simulations are performed (Section III-B): convergence rate of the proposed iterative method and influence of the number K of modalities with comparisons with existing ICA, IVA and M-BSS estimations.

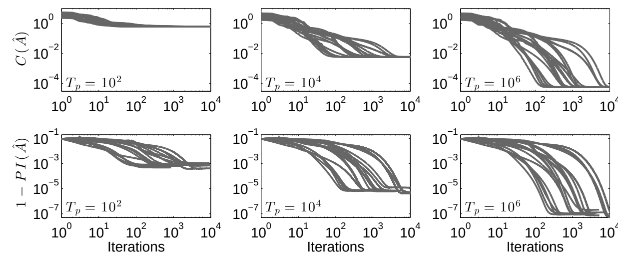


Fig. 1. Convergence rate of the proposed method: performance of the separation against the number of iterations for three values of T_p.

A. Performance measurement and data generation

The multimodal multidimensional model (Section II-A) suffers from indeterminacies due to the multidimensional nature of the sources. Indeed, as pointed out in [10], the multidimensional sources cannot be defined by the values of mixing sub-matrices $A_i^{[k]}$, but rather by the sub-spaces spanned by these sub-matrices. As a consequence, the average performance index (PI) [13] to quantify the quality of the separation is defined as the mean of cosines of the angles between the subspaces [14] spanned by $A_i^{[k]}$ and $\hat{A}_i^{[k]}$:

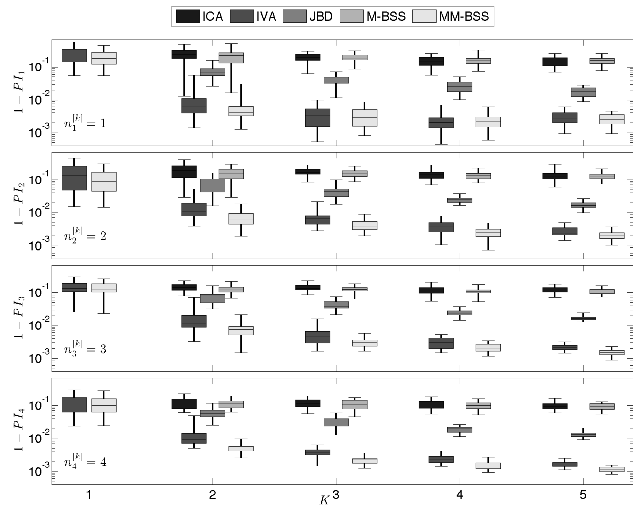
$$PI(\hat{A}) = \frac{1}{I}\sum_{i=1}^{I} PI_i\big(\{\hat{A}_i^{[k]}\}_k\big), \tag{17}$$

with the performance index of the ith source $PI_i\big(\{\hat{A}_i^{[k]}\}_k\big) = \frac{1}{K}\sum_{k=1}^{K}\alpha\Big(A_i^{[k]}, \hat{A}_i^{[k]}\Big)$ where $\alpha(\cdot,\cdot)$ denotes the average cosine between the two subspaces defined by the matrices in arguments, computed as the mean of singular values of pair $(Q_i^{[k]}, \hat{Q}_i^{[k]})$, with $Q_i^{[k]}$ an orthonormalization of $A_i^{[k]}$. The value of PI is in $[0, 1]$: the higher the PI value is, the better the separation is.

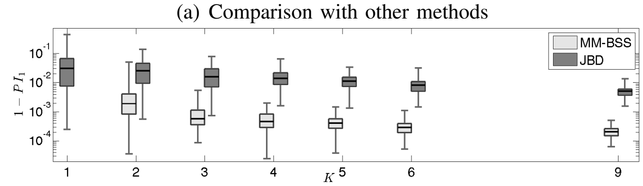
During the simulations, the mixing matrices $A^{[k]}$ are generating as $I + \mathcal{E}$, where $\mathcal{E}$ is a random matrix whose entries are independently drawn from a uniform distribution in $[-.4, .4]$. This perturbation is large enough to show the behavior of the method while assuring a convergence to the global optimum. The covariance matrices $R^{(p)}_{\mathbf{s},\mathbf{s}}$ are drawn as $U^{\dagger}U$ and then vanishing theoretically null entries depending of the structure of $R_{\mathbf{s},\mathbf{s}}$ with entries of U drawn independently from a uniform distribution in $[-.2, .8]$ to ensure a correlation between components and modalities. Finally, estimations of covariance matrices $\hat{R}^{(p)}_{\mathbf{s},\mathbf{s}}$ are drawn from a multivariate Wishart distribution with covariance matrix $R^{(p)}_{\mathbf{s},\mathbf{s}}$ and T_p degrees of freedom. Finally, the observed covariance matrices $\hat{R}^{(p)}_{\mathbf{x},\mathbf{x}}$ are expressed as $\hat{R}^{(p)}_{\mathbf{x},\mathbf{x}} = A\hat{R}^{(p)}_{\mathbf{s},\mathbf{s}}A^{\dagger} + \hat{R}^{(p)}_{\mathbf{b},\mathbf{b}}$, with $\hat{R}^{(p)}_{\mathbf{b},\mathbf{b}}$ a perturbation matrix (i.e. additive noise) drawn randomly as $V^{\dagger}V$ so that its diagonal entries are 10% of the diagonal entries of $A\hat{R}^{(p)}_{\mathbf{s},\mathbf{s}}A^{\dagger}$.

B. Numerical simulations

The convergence rate of the proposed method is shown on Fig. 1. The values of $C(\hat{A})$ and $PI(\hat{A})$ are displayed



(a) Comparison with other methods



(b) Influence of the number of modalities

Fig. 2. Influence of the number of modalities K: performance of the separation against K. The central mark is the median, the edges of the box are the 25th and 75th percentiles. The whiskers extend to the extreme values.

against the number of iterations for three values of T_p. For each plot, 20 random configurations are overlapped. In this simulation, there are four modalities ($K = 4$) and three sources ($I = 3$), with $\mathbf{n}_1 = [2, 2, 1, 0]^\dagger$, $\mathbf{n}_2 = [1, 2, 0, 3]^\dagger$ and $\mathbf{n}_3 = [0, 2, 2, 4]^\dagger$. For each realization, a random multimodal mixing matrix A is generated while keeping fixed the observed multimodal covariance matrices $\hat{R}^{(p)}_{\mathbf{s},\mathbf{s}}$. The iterative algorithm is initialized with the identity matrix $\mathbb{I}$. It is worth noting that each realization converges to the same value of $C(\hat{A})$, for a fixed T_p, highlighting the equivariance [15] of the criterion (14). Furthermore, increasing the value of T_p improves the estimation accuracy of $\hat{R}^{(p)}_{\mathbf{s},\mathbf{s}}$ leading to a better estimation of the multimodal mixing matrix A since $1-PI(\hat{A})$ decreases when T_p increases. Finally, there is a high correlation between minimizing the contrast function $C(\hat{A})$ (13) and maximizing the average PI (17).

In the second numerical experiment, the proposed method is compared to ICA applied on each modality followed by a clustering to reorder the estimated components, to IVA followed by a clustering and to M-BSS applied on each modality (Figure 2(a)). For both ICA and IVA, the clustering method minimizes the angles between the subspaces spanned by the columns of $A^{[k]}$ and of $\hat{A}^{[k]}$. ICA, IVA and M-BSS have been performed by a joint diagonalization algorithm [16], by the algorithms described in [17] and in [10], respectively. An additional method (JBD) is used for the comparison: it consists of minimizing the proposed criterion (14) without any constraints on A (i.e. joint block diagonalizing the set of multimodal covariance matrices) and then projecting the estimate $\hat{A}$ on $\mathcal{BD}iag_{\mathbf{n}}$. Thirty configurations are drawn randomly with $P = 2$ and $T_p = 10000$. To compare with IVA method (which is special case of the proposed method), the dimension of the sources must be the same in all modalities: we choose $\forall i \in \{1, \ldots, 4\},\ n_i^{[k]} = i, \forall k$. Note that with one modality ($K = 1$), IVA and ICA are equivalent and JBD, M-BSS and MM-BSS are equivalent. The results highlight that the proposed method outperforms all the other ones. These results show that modeling the multimodal nature of the data increases the estimation of the mixing matrix: IVA is better than ICA and MM-BSS is better than M-BSS. The same positive behavior is observed by considering the multidimensionality of the sources: M-BSS is better than ICA and MM-BSS is better than IVA. Finally, the constraints on the mixing matrix A for the JBD are necessary to achieve a good estimate of it since MM-BSS outperforms JBD. In practice, the initialization $A_{(0)}$ can be obtained by a JBD estimate and projecting it onto $\mathcal{BD}iag_{\mathbf{n}}$ or by IVA if the dimensions of the sources are reliable.

Finally, the influence of the number of modalities K is shown in Figure 2(b) for general configurations. It consists of extracting a bidimensional ($n_1^{[k]} = 2,\ \forall k$) source (say $\mathbf{s}_1^{[k]}(t)$) from several modalities. In this experiment, for each number of modalities K, 250 trial configurations are drawn randomly with $P = 2$ and $T_p = 10000$ and different A, $R^{(p)}_{\mathbf{s},\mathbf{s}}$ and $\hat{R}^{(p)}_{\mathbf{s},\mathbf{s}}$. The number of the other sources in each modality $(I_k - 1)$ is uniformly sampled from $\{1, 2, 3\}$ and their dimensions $n_i^{[k]}$ are in $\{0, 1, 2\}$. The results are reported in Figure 2(b). It is worth noting that increasing the number of modalities can improve the quality of the extraction. Indeed, the median value of PI_1 increases with K. Moreover, a visual inspection leads to observe that the variance decreases with the number of modalities used. Even if the proposed MM-BSS is closely related to a JBD of a set of matrices, one can see the positive impact of explicitly taking into account the constraints on the mixing matrix A (MM-BSS) instead of a global extraction without embedded the multimodal constraints (JBD).

IV. Conclusions and perspectives

In this letter, the problem of source separation of multimodal multidimensional signals has been considered. A contrast function (13) has been derived based on the likelihood of the multimodal mixing matrix that shares the equivariance property. It has been shown that minimizing this contrast function leads to a constrained joint decomposition of covariance matrices assuming a second order model for the sources. In practice, an iterative relative gradient descent algorithm is used which ensures a local convergence. The proposed method is a generalization of both the unimodal multidimensional source separation and of the multimodal unidimensional source separation. The numerical simulations show that modeling the multi-modality nature of the signal outperforms separate extraction followed by a clustering or a global extraction without embedded the multimodal constraints. Furthermore, using a few number of modalities can improve the quality of the extraction. However increasing it too much will not further improve it, but will only reduce the variance of the estimates. Future studies will address this later remark from a theoretical point of view using some information theory considerations, as well as a theoretical performance analysis.

REFERENCES

[1] P. Comon and C. Jutten, Eds., *Handbook of Blind Source Separation Independent Component Analysis and Applications*. Academic Press, 2010.

[2] P. Comon, "Independent component analysis, a new concept?" *Signal Processing*, vol. 36, no. 3, pp. 287–314, April 1994.

[3] J. R. Kettenring, "Canonical analysis of several sets of variables," *Biometrika*, vol. 58, no. 3, pp. 433–451, December 1971.

[4] L.-O. Li, T. Adali, W. Wang, and V. D. Calhoun, "Joint blind source separation by multiset canonical correlation analysis," *IEEE Transactions on Signal Processing*, vol. 57, no. 10, pp. 3918–3929, October 2009.

[5] J. Sui, G. Pearlson, A. Caprihan, T. Adali, K. A. Kiehl, J. Liu, J. Yamamoto, and V. D. Calhoun, "Discriminating schizophrenia and bipolar disorder by fusing fMRI and DTI in a multimodal CCA+joint ICA model," *NeuroImage*, vol. 57, no. 3, pp. 839 – 855, 2011.

[6] D. Kim, I. Lee, and T.-W. Lee, "Independent vector analysis: Definition and algorithms," in *Signals, Systems and Computers, 2006. ACSSC '06. Fortieth Asilomar Conference on*, Oct 2006, pp. 1393–1396.

[7] J. Cardoso, "Multidimensional independent component analysis," in *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, vol. 4, May 1998, pp. 1941–1944 vol.4.

[8] A. Hyvärinen and P. Hoyer, "Emergence of phase- and shift-invariant features by decomposition of natural images into independent feature subspaces," *Neural Computation*, vol. 12, no. 7, pp. 1705–1720, Jul. 2000.

[9] D.-T. Pham and J.-F. Cardoso, "Blind separation of instantaneous mixtures of nonstationary sources," *IEEE Transactions on Signal Processing*, vol. 49, no. 9, pp. 1837–1848, September 2001.

[10] D. Lahat, J.-F. Cardoso, and H. Messer, "Second-order multidimensional ica: performance analysis," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4598–4610, September 2012.

[11] ——, "Blind separation of multi-dimensional components via subspace decomposition: Performance analysis," *IEEE Transactions on Signal Processing*, vol. 62, no. 11, pp. 2894–2905, June 2014.

[12] P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization*. Academic Press, 1981.

[13] P. Tichavský, A. Yeredor, and Z. Koldovský, "On computation of approximate joint block-diagonalization using ordinary AJD," in *Proc. of Latent Variable Analysis and Signal Separation (LVA/ICA)*, ser. Lecture Notes in Computer Science, vol. 7191, 2012, pp. 163–171.

[14] G. H. Golub and C. F. Van Loan, *Matrix Computation*, 3rd ed. Johns Hopkins University Press, 1996.

[15] J.-F. Cardoso, "Blind signal separation: statistical principles," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 2009–2025, October 1998.

[16] D.-T. Pham, "Joint approximate diagonalization of positive definite matrices," *SIAM Journal on Matrix Analysis And Applications*, vol. 22, no. 4, pp. 1136–1152, 2001.

[17] J. Via, M. Anderson, X.-L. Li, and T. Adali, "A maximum likelihood approach for independent vector analysis of gaussian data sets," in *IEEE International Workshop on Machine Learning for Signal Processing (MLSP),*, Sept 2011, pp. 1–6.