TWO NEW GRADIENT BASED NON-UNITARY JOINT BLOCK-DIAGONALIZATION ALGORITHMS

H. Ghennioui[1,2], N. Thirion-Moreau[1], E. Moreau[1], D. Aboutajdine[2], A. Adib[2,3]

[1] LSEET-LEPI, UMR CNRS 6017,
Université du Sud Toulon Var
F-83957, La Garde Cédex, France
email: {thirion,moreau}@univ.tln.fr

[2] GSCM-LRIT, FSR
Av. Ibn Battouta, BP 1014
Rabat, Maroc
email: ghennioui@gmail.com

[3] DPG, IS, Av. Ibn Battouta
BP 703, Rabat, Maroc
email: adib@israbat.ac.ma
aboutaj@ieee.org

ABSTRACT

This paper addresses the problem of the non-unitary joint block diagonalization ($\mathsf{NU-JBD}$) of a given set of matrices. Such a problem arises in various fields of applications among which blind separation of convolutive mixtures of sources and array processing for wide-band signals. We present two new algorithms based respectively on (absolute) gradient and relative gradient descendent approaches. The main advantage of the proposed algorithms is that they are more general (the real, positive definite or hermitian assumptions about the matrices belonging to the considered set are no more necessary and the found joint block diagonalizer can be either a unitary or non-unitary matrix). These algorithms also outperform the JBD algorithm based on an optimal step size but "approximate gradient" approach that we had previously suggested in [12]. In fact, here, the exact calculus of the complex gradient matrix is performed whereas it was approximated in [12]. Finally, by ensuring the invertibility of the estimated matrix, the relative gradient approach makes the proposed $\mathsf{NU-JBD}$ algorithm more stable and consequently more robust. Computer simulations are provided in order to illustrate the effectiveness of the proposed approaches in two cases: when exact block-diagonal matrices are considered and when they are perturbed by an additive Gaussian noise. A comparison with the method presented in [12] is also performed, emphasizing the good behavior of the proposed algorithms.

1. INTRODUCTION

The problem of joint decompositions of matrices or tensors sets arises in many multivariate signal processing applications.

One of the first considered problem was the joint-diagonalization of matrices under the unitary constraint, leading to the nowadays well-known JADE [4] and SOBI [1] algorithms. The following works have addressed either the problem of the joint-diagonalization of tensors [7][10][17] or the problem of the joint-diagonalization of matrices but discarding the unitary constraint [8][11][18][21][22]. A second type of matrices decomposition has proven to be useful in blind source separation, telecommunications and cryptography. It consists of joint zero-diagonalizing a set of matrices either under the unitary constraint [2] or not [6][11]. Most of the proposed (unitary) joint-diagonalization (JD) and/or zero-diagonalization (JZD) algorithms have been applied to the problem of the blind separation of instantaneous mixtures of sources. Finally, a third particular type of matrices decomposition arises in both the wide-band sources localization and the blind separation of convolutive mixtures of sources problems. It is called joint block-diagonalization since the considered matrices are block diagonal ones: a block diagonal matrix is a square diagonal matrix in which the diagonal elements are square matrices of any size (possibly even), and the off-diagonal blocks are zero matrices. Such a problem has been considered in [3][9] where the block-diagonal matrices under consideration have to be positive definite and hermitian matrices and the joint-block diagonalizer is required to be a unitary matrix. When the joint-block diagonalizer is no more necessarily a unitary matrix, alternative solutions have been proposed. [13][19] address the problem of the $\mathsf{NU-JBD}$ of a set of positive definite matrices whereas this assumption is discarded in [12] and the considered matrices simply have to be hermitian ones. Yet, let us remark that this latter solution relies upon an approximation in the calculation of the complex gradient matrix of the considered cost function.

In this communication, we present two new algorithms based respectively on a (absolute) gradient and on a relative gradient-descendent approach. The main advantage of these algorithms is that they are more general (the considered matrices can be complex, non-positive definite, non-hermitian and the joint block-diagonalizer is not necessarily a unitary matrix). Besides, the relative gradient based algorithm ensures the invertibility of the joint block-diagonalizer (when the step size is sufficiently small) and hence a greater numerical stability and robustness of the resulting algorithm.

2. PROBLEM STATEMENT

The non-unitary joint block-diagonalization ($\mathsf{NU-JBD}$) problem is stated in the following way: let us consider a set $\mathcal{M}$ of N_m ($N_m \in \mathbb{N}^*$) square matrices $\mathbf{M}_i \in \mathbb{C}^{M\times M}$, for all $i \in \{1,\ldots,N_m\}$ which all admit the following decomposition:

$$\mathbf{M}_i = \mathbf{A}\mathbf{D}_i\mathbf{A}^H \ , \tag{1}$$

where $\mathbf{D}_i = \begin{pmatrix} \mathbf{D}_{i,11} & \mathbf{0}_{12} & \ldots & \mathbf{0}_{1r} \\ \mathbf{0}_{21} & \mathbf{D}_{i,22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0}_{r-1r} \\ \mathbf{0}_{r1} & \ldots & \mathbf{0}_{rr-1} & \mathbf{D}_{i,rr} \end{pmatrix}$, for all $i \in \{1,\ldots,N_m\}$ are $N\times N$ block diagonal matrices with $r \in \mathbb{N}^*$, $\mathbf{D}_{i,jj}, i \in \{1,\ldots,N_m\}$, $j \in \{1,\ldots,r\}$ are $n_j \times n_j$ square matrices so that $n_1+\ldots+n_r = N$ where $\mathbf{0}_{ij}$ denotes the $n_i \times n_j$ null matrix and $(\cdot)^H$ the transpose conjugate operator. $\mathbf{A}$ is a $M\times N$ ($M \geq N$) full rank matrix and the $N \geq M$ matrix $\mathbf{B}$ is its pseudo-inverse (or generalized Moore-Penrose inverse). The set of the N_m square matrices $\mathbf{D}_i \in \mathbb{C}^{N\times N}$ is denoted $\mathcal{D}$. The $\mathsf{NU-JBD}$ problem consists of estimating the matrix $\mathbf{A}$ or $\mathbf{B}$ and eventually the block-diagonal matrices set $\mathcal{D}$ from only the matrices set $\mathcal{M}$.

The case of a unitary matrix $\mathbf{A}$ has been considered in [9] where a first solution is suggested. Recently, two solutions have been proposed in [13][19] for a non-unitary matrix $\mathbf{A}$ and for a set $\mathcal{M}$ of positive definite matrices. Finally in [12], the matrices in $\mathcal{M}$ simply have to be hermitian. However, the calculus of the complex gradient matrix is an approximated one like in [21] for the JD problem.

3. JOINT BLOCK-DIAGONALIZATION BASED ON GRADIENT APPROACHES

3.1 Principle

Our aim is to present a new algorithm to solve the problem of the $\mathsf{NU-JBD}$. The following cost function is considered as suggested in [12]:

$$\mathcal{C}_{BD}(\mathbf{B}) = \sum_{i=1}^{N_m} \|\mathsf{OffBdiag}_{(\mathbf{n})}\{\mathbf{B}\mathbf{M}_i\mathbf{B}^H\}\|_F^2, \tag{2}$$

where $\|\cdot\|_F$ is the Frobenius norm and the vector $\mathbf{n}$ is defined as $\mathbf{n} = (n_1, \ldots, n_r)$. Considering a square $N \times N$ matrix $\mathbf{M} = (M_{ij})$, such that:

$$\mathbf{M} = \begin{pmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} & \mathbf{M}_{1r} \\ \mathbf{M}_{21} & \ddots & \\ \mathbf{M}_{r1} & \mathbf{M}_{r2} & \mathbf{M}_{rr} \end{pmatrix}, \tag{3}$$

where $\mathbf{M}_{ij}$ for all $i,j = 1,\ldots,r$ are $n_i \times n_j$ matrices. The matrix operator $\mathsf{OffBdiag}_{(\mathbf{n})}$ [1] is then defined as:

$$\mathsf{OffBdiag}_{(\mathbf{n})}\{\mathbf{M}\} = \begin{pmatrix} \mathbf{0}_{11} & \mathbf{M}_{12} & \mathbf{M}_{1r} \\ \mathbf{M}_{21} & \ddots & \\ \mathbf{M}_{r1} & \mathbf{M}_{r2} & \mathbf{0}_{rr} \end{pmatrix}. \tag{4}$$

The $\mathsf{NU-JBD}$ problem is linked to the non-unitary joint diagonalization $(\mathsf{NU-JD})$ one. In this latter, $\mathbf{D}_{i,jj}, i \in \{1,\ldots,N_m\}$, $j \in \{1,\ldots,r\}$ are no more matrices but scalars ($n_j = 1$ for all $j \in \{1,\ldots,r\}$ and $N = r$). The zero block-digonality operator can then be simplified as follows:

$$\mathsf{OffBdiag}_{(\mathbf{1})}\{\mathbf{M}\} = ((1-\delta_{ij})M_{ij})\mathbf{1}_N = \mathsf{OffDiag}\{\mathbf{M}\},$$

where $\mathbf{1}_N$ is a $N \times N$ matrix whose components are all equal to 1 which leads to the minimization of the following cost function:

$$\mathcal{C}_D(\mathbf{B}) = \sum_{k=1}^{N_m} \|\mathsf{OffDiag}\{\mathbf{B}\mathbf{M}_k\mathbf{B}^H\}\|_F^2. \tag{5}$$

It has been used in [4] under the unitary constraint (*i.e.* $M = N$ and $\mathbf{B}\mathbf{B}^H = \mathbf{B}^H\mathbf{B} = \mathbf{I}_N$, with $\mathbf{I}_N$ the $N \times N$ identity matrix). When $\mathbf{B}$ is not necessarily a unitary matrix, this cost function $\mathcal{C}_D(\mathbf{B})$ has also been used in [11][16][21][22].
We propose, here, to use a gradient-descent algorithm to minimize the cost function given by Eq. (2) to estimate the matrix $\mathbf{B} \in \mathbb{C}^{N\times M}$. It means that $\mathbf{B}$ is re-estimated at each iteration m and from now on denoted $\mathbf{B}^{(m)}$. It is updated according to the following adaptation rule for all $m = 1, 2, \ldots$

$$\mathbf{B}^{(m)} = \mathbf{B}^{(m-1)} - \mu_a \nabla_a \mathcal{C}_{BD}(\mathbf{B}^{(m-1)}), \tag{6}$$

where μ_a is positive a small enough number called the step size or adaptation coefficient and where $\nabla_a \mathcal{C}_{BD}(\mathbf{B})$ stands for the complex (absolute) gradient matrix defined by [20]:

$$\nabla_a \mathcal{C}_{BD}(\mathbf{B}) = 2\frac{\partial \mathcal{C}_{BD}(\mathbf{B})}{\partial \mathbf{B}^*}. \tag{7}$$

$(\cdot)^*$ stands for the complex conjugate operator.

[1] When there is no ambiguity, this notation $\mathsf{OffBdiag}_{(\mathbf{n})}\{\cdot\}$ will be simplified into $\mathsf{OffBdiag}\{\cdot\}$

To ensure the invertibility of the matrix $\mathbf{B}$ and hence the stability of the algorithm, one can use another optimization scheme called relative gradient method [5]:

$$\begin{aligned}\mathbf{B}^{(m)} &= \mathbf{B}^{(m-1)} - \mu_r \nabla_r \mathcal{C}_{BD}(\mathbf{B}^{(m-1)})\mathbf{B}^{(m-1)}, \\ &= \left(\mathbf{I}_N - \mu_r \nabla_r \mathcal{C}_{BD}(\mathbf{B}^{(m-1)})\right)\mathbf{B}^{(m-1)},\end{aligned} \tag{8}$$

where

$$\nabla_r \mathcal{C}_{BD}(\mathbf{B}) = 2\frac{\partial \mathcal{C}_{BD}(\mathbf{B})}{\partial \mathbf{B}^*}\mathbf{B}^H = \nabla_a \mathcal{C}_{BD}(\mathbf{B})(\mathbf{B})^H. \tag{9}$$

The complex absolute gradient matrix $\nabla_a \mathcal{C}_{BD}(\mathbf{B})$ (given by Eq. (7)) of the cost function $\mathcal{C}_{BD}(\mathbf{B})$ has to be calculated to derive the algorithm.

3.2 Gradient matrix of the cost function $\mathcal{C}_{BD}(\mathbf{B})$

Let us introduce some notations: $\mathsf{tr}\{\cdot\}$ denotes the trace operator, $d(\cdot)$ stands for the differential operator, $\mathbf{vec}(\cdot)$ is the vec-operator (applied on a matrix $\mathbf{M} \in \mathbb{C}^{N\times N}$ it stacks its columns into a column vector belonging to $\mathbb{C}^{N^2\times 1}$) and $\mathbf{T}_{\text{Boff}}$ is the $N^2 \times N^2$ "transformation" matrix defined as:

$$\mathbf{T}_{\text{Diag}} = \mathsf{diag}\{\mathbf{vec}(\mathsf{Bdiag}_{(\mathbf{n})}\{\mathbf{1}_N\})\}, \tag{10}$$

$$\mathbf{T}_{\text{Boff}} = \mathbf{I}_{N^2} - \mathbf{T}_{\text{Diag}}, \tag{11}$$

where $\mathbf{I}_{N^2} = \mathsf{Diag}\{\mathbf{1}_{N^2}\}$ is the $N^2 \times N^2$ identity matrix, $\mathsf{diag}\{\mathbf{a}\}$ is a square diagonal matrix which contains the elements of the vector $\mathbf{a}$ on its diagonal and $\mathsf{Bdiag}_{(\mathbf{n})}\{\mathbf{M}\} = \mathbf{M} - \mathsf{OffBdiag}_{(\mathbf{n})}\{\mathbf{M}\}$ (generally $\mathsf{Bdiag}_{(\mathbf{1})}\{\cdot\}$ is simply denoted by $\mathsf{Diag}\{\cdot\}$). Considering three $N \times N$ square matrices $\mathbf{D}_1$, $\mathbf{D}_2$ and $\mathbf{D}_3$ and two rectangular matrices $\mathbf{D}_4$ $(M \times N)$ and $\mathbf{D}_5$ $(N \times M)$ and a square $N \times N$ matrix $\mathbf{D}_6$, our developments are based on the following properties [15][16]:

- $\mathbf{P}_1$. $\|\mathsf{OffBdiag}\{\mathbf{D}_1\}\|_F^2 = \mathsf{tr}\left\{(\mathsf{OffBdiag}\{\mathbf{D}_1\})^H\, \mathsf{OffBdiag}\{\mathbf{D}_1\}\right\}$ $= \mathsf{tr}\left\{\mathbf{D}_1^H \mathsf{OffBdiag}\{\mathbf{D}_1\}\right\}$.
- $\mathbf{P}_2$. $\mathsf{tr}\{\mathbf{D}_1\} = \mathsf{tr}\{\mathbf{D}_1^T\}$.
- $\mathbf{P}_3$. $\mathsf{tr}\{\mathbf{D}_1 + \mathbf{D}_2\} = \mathsf{tr}\{\mathbf{D}_1\} + \mathsf{tr}\{\mathbf{D}_2\}$.
- $\mathbf{P}_4$. $\mathsf{tr}\{\mathbf{D}_1\mathbf{D}_2\mathbf{D}_3\} = \mathsf{tr}\{\mathbf{D}_3\mathbf{D}_1\mathbf{D}_2\} = \mathsf{tr}\{\mathbf{D}_2\mathbf{D}_3\mathbf{D}_1\} \Rightarrow \mathsf{tr}\{\mathbf{D}_1\mathbf{D}_2\} = \mathsf{tr}\{\mathbf{D}_2\mathbf{D}_1\}$.
- $\mathbf{P}_4'$. $\mathsf{tr}\{\mathbf{D}_4\mathbf{D}_5\} = \mathsf{tr}\{\mathbf{D}_5\mathbf{D}_4\}$.
- $\mathbf{P}_5$. $\mathsf{tr}\{\mathbf{D}_1^H\mathbf{D}_2\} = (\mathbf{vec}(\mathbf{D}_1))^H\, \mathbf{vec}(\mathbf{D}_2)$.
- $\mathbf{P}_6$. $\mathbf{vec}(\mathsf{OffBdiag}\{\mathbf{D}_6\}) = \mathbf{T}_{\text{Boff}}\mathbf{vec}(\mathbf{D}_6)$.
- $\mathbf{P}_7$. $d(\mathbf{D}_1^H) = d(\mathbf{D}_1)^H$.
- $\mathbf{P}_8$. $d(\mathbf{D}_1^*) = d(\mathbf{D}_1)^*$.
- $\mathbf{P}_9$. $d(\mathbf{D}_1\mathbf{D}_2) = d(\mathbf{D}_1)\mathbf{D}_2 + \mathbf{D}_1 d(\mathbf{D}_2)$.
- $\mathbf{P}_{10}$. $d(\mathbf{D}_1 + \mathbf{D}_2) = d(\mathbf{D}_1) + d(\mathbf{D}_2)$.
- $\mathbf{P}_{11}$. $d(\mathsf{tr}\{\mathbf{D}_1\}) = \mathsf{tr}\{d(\mathbf{D}_1)\}$.
- $\mathbf{P}_{12}$. $d(\mathbf{vec}(\mathbf{D}_1\}) = \mathbf{vec}(d(\mathbf{D}_1))$.
- $\mathbf{P}_{13}$. $d\left(f(\mathbf{Z},\mathbf{Z}^*) = \mathsf{tr}\{\mathbf{D}_1^T\mathbf{Z} + \mathbf{Z}^H\mathbf{D}_2\}\right)$ $= \mathsf{tr}\left\{\mathbf{D}_1^T d\mathbf{Z} + \mathbf{D}_2^T d\mathbf{Z}^*\right\} \Rightarrow \frac{\partial f}{\partial \mathbf{Z}} = \mathbf{D}_1$ and $\frac{\partial f}{\partial \mathbf{Z}^*} = \mathbf{D}_2$.
- $\mathbf{P}_{14}$. $\mathbf{vec}(\mathbf{D}_1\mathbf{D}_2\mathbf{D}_3) = \mathbf{D}_3^T \otimes \mathbf{D}_1\mathbf{vec}(\mathbf{D}_2)$ where $\otimes$ denotes the Kronecker product.
- $\mathbf{P}_{15}$. $(\mathbf{D}_1 \otimes \mathbf{D}_2)^H = \mathbf{D}_1^H \otimes \mathbf{D}_2^H$.

Using $\mathbf{P}_1$, $\mathcal{C}_{BD}(\mathbf{B})$ can be expressed as:

$$\mathcal{C}_{BD}(\mathbf{B}) = \sum_{i=1}^{N_m} \mathsf{tr}\left\{(\mathbf{B}\mathbf{M}_i\mathbf{B}^H)^H \mathsf{OffBdiag}\{\mathbf{B}\mathbf{M}_i\mathbf{B}^H\}\right\}.$$

From the properties $\mathbf{P}_3$, $\mathbf{P}_9$, $\mathbf{P}_{10}$ and $\mathbf{P}_{11}$, the differential of the cost function can be obtained as:

$$
\begin{aligned}
d\mathcal{C}_{BD}(\mathbf{B}) &= \sum_{i=1}^{N_m} \mathsf{tr}\left\{ d\left((\mathbf{B}\mathbf{M}_i\mathbf{B}^H)^H \mathsf{OffBdiag}\{\mathbf{B}\mathbf{M}_i\mathbf{B}^H\}\right)\right\} \\
&= \underbrace{\sum_{i=1}^{N_m} \mathsf{tr}\left\{ d\left((\mathbf{B}\mathbf{M}_i\mathbf{B}^H)^H\right) \mathsf{OffBdiag}\{\mathbf{B}\mathbf{M}_i\mathbf{B}^H\}\right\}}_{\mathcal{F}(\mathbf{B})} \\
&+ \underbrace{\sum_{i=1}^{N_m} \mathsf{tr}\left\{ (\mathbf{B}\mathbf{M}_i\mathbf{B}^H)^H d\left(\mathsf{OffBdiag}\{\mathbf{B}\mathbf{M}_i\mathbf{B}^H\}\right)\right\}}_{\mathcal{G}(\mathbf{B})}.
\end{aligned} \tag{12}
$$

The properties $\mathbf{P}_2$, $\mathbf{P}_3$, $\mathbf{P}_4$, $\mathbf{P}'_4$, $\mathbf{P}_7$, $\mathbf{P}_9$, $\mathbf{P}_8$ and $\mathbf{P}_{12}$ imply that:

$$
\begin{aligned}
\mathcal{F}(\mathbf{B}) &= \sum_{i=1}^{N_m} \mathsf{tr}\left\{ \mathbf{B}\mathbf{M}_i^H d\mathbf{B}^H \mathsf{OffBdiag}\{\mathbf{B}\mathbf{M}_i\mathbf{B}^H\}\right\} \\
&+ \sum_{i=1}^{N_m} \mathsf{tr}\left\{ d\mathbf{B}\mathbf{M}_i^H \mathbf{B}^H \mathsf{OffBdiag}\{\mathbf{B}\mathbf{M}_i\mathbf{B}^H\}\right\} \\
&= \sum_{i=1}^{N_m} \mathsf{tr}\left\{ \left(\mathsf{OffBdiag}\{\mathbf{B}\mathbf{M}_i\mathbf{B}^H\}\mathbf{B}\mathbf{M}_i^H\right)^T d\mathbf{B}^*\right\} \\
&+ \sum_{i=1}^{N_m} \mathsf{tr}\left\{ \left(\left(\mathsf{OffBdiag}\{\mathbf{B}\mathbf{M}_i\mathbf{B}^H\}\right)^T \mathbf{B}^*\mathbf{M}_i^*\right)^T d\mathbf{B}\right\}.
\end{aligned}
$$

While properties $\mathbf{P}_2$, $\mathbf{P}_3$, $\mathbf{P}_4$, $\mathbf{P}'_4$, $\mathbf{P}_5$, $\mathbf{P}_6$, $\mathbf{P}_8$ and $\mathbf{P}_9$ involve

$$
\begin{aligned}
\mathcal{G}(\mathbf{B}) &= \sum_{i=1}^{N_m} (\mathbf{vec}(\mathbf{B}\mathbf{M}_i\mathbf{B}^H))^H \mathbf{vec}(d(\mathsf{OffBdiag}\{\mathbf{B}\mathbf{M}_i\mathbf{B}^H\})) \\
&= \sum_{i=1}^{N_m} \left(\mathbf{T}_{\mathrm{Boff}}\mathbf{vec}\left(\mathbf{B}\mathbf{M}_i\mathbf{B}^H\right)\right)^H \mathbf{vec}\left(d\left(\mathbf{B}\mathbf{M}_i\mathbf{B}^H\right)\right) \\
&= \sum_{i=1}^{N_m} (\mathbf{vec}(\mathsf{OffBdiag}\{\mathbf{B}\mathbf{M}_i\mathbf{B}^H\}))^H \mathbf{vec}(d(\mathbf{B}\mathbf{M}_i\mathbf{B}^H)) \\
&= \sum_{i=1}^{N_m} \mathsf{tr}\left\{ \left(\mathsf{OffBdiag}\{\mathbf{B}\mathbf{M}_i\mathbf{B}^H\}\right)^H d\left(\mathbf{B}\mathbf{M}_i\mathbf{B}^H\right)\right\} \\
&= \sum_{i=1}^{N_m} \mathsf{tr}\left\{ \left(\left(\mathsf{OffBdiag}\{\mathbf{B}\mathbf{M}_i\mathbf{B}^H\}\right)^* \mathbf{B}^*\mathbf{M}_i^T\right)^T d\mathbf{B}\right\} \\
&+ \mathsf{tr}\left\{ \left(\left(\mathsf{OffBdiag}\{\mathbf{B}\mathbf{M}_i\mathbf{B}^H\}\right)^H \mathbf{B}\mathbf{M}_i\right)^T d\mathbf{B}^*\right\}.
\end{aligned}
$$

We replace $\mathcal{F}(\mathbf{B})$ and $\mathcal{G}(\mathbf{B})$ in Eq. (12) to finally find that:

$$
\begin{aligned}
d\mathcal{C}_{BD}(\mathbf{B}) &= \sum_{i=1}^{N_m} \mathsf{tr}\left\{ \left(\left(\left(\mathsf{OffBdiag}\{\mathbf{B}\mathbf{M}_i\mathbf{B}^H\}\right)^T \mathbf{B}^*\mathbf{M}_i^*\right)^T \right.\right. \\
&+ \left.\left. \left(\left(\mathsf{OffBdiag}\{\mathbf{B}\mathbf{M}_i\mathbf{B}^H\}\right)^* \mathbf{B}^*\mathbf{M}_i^T\right)^T \right) d\mathbf{B}\right\} \\
&+ \mathsf{tr}\left\{ \left(\left(\left(\mathsf{OffBdiag}\{\mathbf{B}\mathbf{M}_i\mathbf{B}^H\}\right) \mathbf{B}\mathbf{M}_i^H\right)^T \right.\right. \\
&+ \left.\left. \left(\left(\mathsf{OffBdiag}\{\mathbf{B}\mathbf{M}_i\mathbf{B}^H\}\right)^H \mathbf{B}\mathbf{M}_i\right)^T \right) d\mathbf{B}^*\right\}.
\end{aligned}
$$

Using the property $\mathbf{P}_{13}$ we obtain:

$$
\begin{aligned}
\frac{\partial \mathcal{C}_{BD}(\mathbf{B})}{\partial \mathbf{B}} &= \sum_{i=1}^{N_m} \left(\mathsf{OffBdiag}\{\mathbf{B}\mathbf{M}_i\mathbf{B}^H\}\right)^T \mathbf{B}^*\mathbf{M}_i^* \\
&+ \sum_{i=1}^{N_m} \left(\mathsf{OffBdiag}\{\mathbf{B}\mathbf{M}_i\mathbf{B}^H\}\right)^* \mathbf{B}^*\mathbf{M}_i^T. \\
\frac{\partial \mathcal{C}_{BD}(\mathbf{B})}{\partial \mathbf{B}^*} &= \sum_{i=1}^{N_m} \left(\mathsf{OffBdiag}\{\mathbf{B}\mathbf{M}_i\mathbf{B}^H\}\right) \mathbf{B}\mathbf{M}_i^H \\
&+ \sum_{i=1}^{N_m} \left(\mathsf{OffBdiag}\{\mathbf{B}\mathbf{M}_i\mathbf{B}^H\}\right)^H \mathbf{B}\mathbf{M}_i.
\end{aligned}
$$

Leading to the following result:

$$
\begin{aligned}
\nabla_a \mathcal{C}_{BD}(\mathbf{B}) = 2 \Bigg[&\sum_{i=1}^{N_m} \mathsf{OffBdiag}\{\mathbf{B}\mathbf{M}_i\mathbf{B}^H\}\mathbf{B}\mathbf{M}_i^H \\
&+ \left(\mathsf{OffBdiag}\{\mathbf{B}\mathbf{M}_i\mathbf{B}^H\}\right)^H \mathbf{B}\mathbf{M}_i \Bigg]
\end{aligned} \tag{13}
$$

Eq. (13) is then used in the absolute gradient descent algorithm given by Eq. (6).

3.3 Summary of the proposed algorithms

The proposed non-unitary joint block-diagonalization algorithms based on absolute gradient and relative gradient-descent approaches are respectively denoted $\mathsf{JBD}_{\mathsf{NU},\mathsf{G}_{\mathsf{F},\mathsf{A}}}$ and $\mathsf{JBD}_{\mathsf{NU},\mathsf{G}_{\mathsf{F},\mathsf{R}}}$. Their principle is summed up below:

$\mathsf{NU-JBD}$ algorithms

A1. $\mathsf{JBD}_{\mathsf{NU},\mathsf{G}_{\mathsf{F},\mathsf{A}}}$ based on a (absolute) gradient approach,
A2. $\mathsf{JBD}_{\mathsf{NU},\mathsf{G}_{\mathsf{F},\mathsf{R}}}$ based on a relative gradient approach.

The N_m square matrices of $\mathcal{M}$ are denoted by $\mathbf{M}_1, \mathbf{M}_2, \ldots, \mathbf{M}_{N_m}$.
Given an initial estimate $\mathbf{B}^{(0)}$ (for example, in the square case ($N = M$), one can choose $\mathbf{B}^{(0)} = \mathbf{I}_M$).
For $m = 1, 2, \ldots$

- Compute $\nabla_a \mathcal{C}_{BD}(\mathbf{B}^{(m)})$ given in Eq. (13).
- (**A1**) Set $\mathbf{B}^{(m)} = \mathbf{B}^{(m-1)} - \mu_a \nabla_a \mathcal{C}_{BD}(\mathbf{B}^{(m-1)})$.
- (**A2**) Compute $\nabla_r \mathcal{C}_{BD}(\mathbf{B})$ given in Eq. (9).
- (**A2**) Set $\mathbf{B}^{(m)} = (\mathbf{I}_N - \mu_r \nabla_r \mathcal{C}_{BD}(\mathbf{B}^{(m-1)}))\,\mathbf{B}^{(m-1)}$.
- Eventually normalize $\mathbf{B} = \mathbf{B}/\|\mathbf{B}\|_F$.
- Stop after a fixed number of iterations or when $\|\mathbf{B}^{(m)} - \mathbf{B}^{(m-1)}\|_F \leq \varepsilon$ where ε is a small positive threshold.

EndFor

4. COMPUTER SIMULATIONS

We present simulations to illustrate the effectiveness of the proposed algorithms. We consider a set $\mathcal{D}$ of $N_m = 20$ (resp. 100) matrices, randomly chosen (according to a Gaussian law) of mean 0 and variance 1. Initially these matrices are exactly block-diagonal, then a random noise matrix of mean 0 and variance σ_b^2 is added. The signal to noise ratio is defined as $\mathsf{SNR} = 10\log(\frac{1}{\sigma_b^2})$ (in this case $\sigma_b^2 = 0.01$ implying $\mathsf{SNR} = 20$ dB).
To measure the quality of the estimation, the following performance index (which is an extension of the one introduced

in [17]) is used:

$$\begin{aligned} I_{conv}(\mathbf{G}) &= \frac{1}{r(r-1)} \left[\sum_{i=1}^{r} \left(\sum_{j=1}^{r} \frac{\|(\mathbf{G})_{i,j}\|_F^2}{\max_{\ell} \|(\mathbf{G})_{i,\ell}\|_F^2} - 1 \right) \right. \\ &+ \left. \sum_{j=1}^{r} \left(\sum_{i=1}^{r} \frac{\|(\mathbf{G})_{i,j}\|_F^2}{\max_{\ell} \|(\mathbf{G})_{\ell,j}\|_F^2} - 1 \right) \right], \end{aligned}$$

where $(\mathbf{G})_{i,j}$ for all $i, j \in \{1, \ldots, r\}$ is the (i, j)-th block matrix of $\mathbf{G} = \hat{\mathbf{B}}\mathbf{A}$. The better results are obtained when the index performance $I_{conv}(\cdot)$ is found to be close to 0 in linear scale ($-\infty$ in logarithmic scale). All the displayed results have been averaged over 10 Monte-Carlo trials. The non-unitary joint block diagonalization presented in [12] and based on an optimal step size but approximated gradient is denoted by $\mathsf{JBD}_{\mathsf{NU},\mathsf{G_O}}$. We use a mixture matrix $\mathbf{A}$ whose components are randomly generated according to a uniform law in $[-1, 1]$. It remains unchanged through the Monte-Carlo runs. We consider $M = N = 9$, $r = 3$ and $n_j = 3$ for all $j = 1, \ldots, 3$. Here, the initial matrix $\mathbf{B}^{(0)}$ has been chosen equal to $\mathbf{I}_M$. Let us, however notice, that a good initial estimate remains important to ensure the convergence to the true solution. One possible way to initialize is to consider the solution given by the orthogonal joint block-diagonalisation [9] to start in the neighborhood of the solution.

First, we compare the performance index obtained thanks to the three algorithms ($\mathsf{JBD}_{\mathsf{NU},\mathsf{G_{F,A}}}$, $\mathsf{JBD}_{\mathsf{NU},\mathsf{G_{F,R}}}$ and $\mathsf{JBD}_{\mathsf{NU},\mathsf{G_O}}$) versus the number of iterations for $N_m = 100$ matrices in a quasi noiseless context ($\mathsf{SNR} = 100$ dB) on the Fig. 1 and then in a noisy context ($\mathsf{SNR} = 20$ dB) on the Fig. 2. While the three algorithms behave quite similarly in the nearly noiseless case since they reach nearly the same performance (≈ -120 dB), the algorithm based on a relative gradient approach outperforms the two other algorithms in a noisy context (-43 dB instead of -38 dB and -36 dB). Let us also notice that the convergence is quicker with the $\mathsf{JBD}_{\mathsf{NU},\mathsf{G_O}}$ algorithm since we use the "optimal" step size version of the algorithm which is not the case with the two other algorithms. With regard to $\mathsf{JBD}_{\mathsf{NU},\mathsf{G_{F,A}}}$, $\mathsf{JBD}_{\mathsf{NU},\mathsf{G_{F,R}}}$, we have plotted the evolution of the performance index versus the value of the step size: the convergence speed increases when the step size increases.

Then, we show the influence of the size N_m of $\mathcal{M}$. We have displayed the performance index versus the number of iterations for $N_m = 20$ matrices in a nearly noiseless context ($\mathsf{SNR} = 100$ dB) on the Fig. 3 then in a noisy context ($\mathsf{SNR} = 20$ dB) on the Fig. 4. These charts illustrate the good behavior of the two proposed algorithms. One can effectively notice a decrease of the $\mathsf{JBD}_{\mathsf{NU},\mathsf{G_O}}$ algorithm performance in a noisy context especially when very few matrices are simultaneously joint block-diagonalized. In a rather difficult context (noisy case + few matrices to be joint block diagonalized), the $\mathsf{JBD}_{\mathsf{NU},\mathsf{G_{F,R}}}$ algorithm seems to be numerically more stable and exhibits better performances than those obtained thanks to the other algorithms.

5. CONCLUSION AND DISCUSSION

In this communication, we have proposed two new algorithms (namely $\mathsf{JBD}_{\mathsf{NU},\mathsf{G_{F,A}}}$ and $\mathsf{JBD}_{\mathsf{NU},\mathsf{G_{F,R}}}$). The first one is based on an absolute gradient descent approach while the second one relies upon a relative gradient-descendent approach. They both perform the non-unitary joint block-diagonalization of a given set of complex non necessarily hermitian matrices. One of the main advantages of these algorithms is that they are more general. The algorithm $\mathsf{JBD}_{\mathsf{NU},\mathsf{G_{F,R}}}$ based on a relative gradient approach exhibits the best performances in a difficult context (noisy case and very few matrices to be simultaneously joint block-diagonalized). These algorithms find applications in blind separation of convolutive mixtures of sources and in array processing (see for example [14]). In the blind sources separation context, they should enable to achieve better performances since the unitary constraint is discarded. Extensions for futur researches would be to study optimal step size versions.

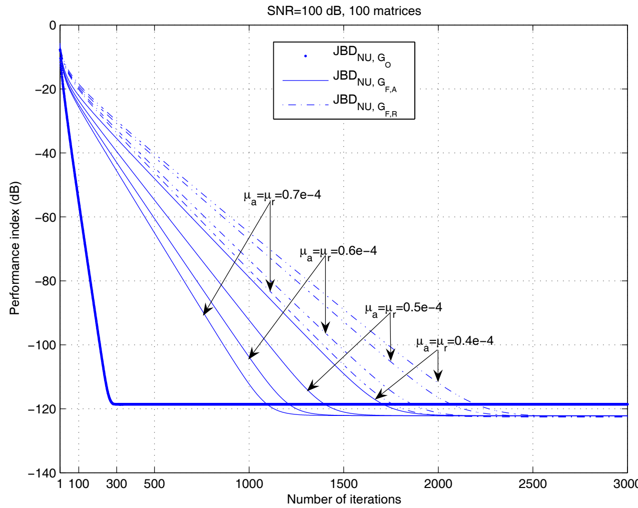

Figure 1: Performance index versus number of iterations: $N_m = 100$ matrices and $\mathsf{SNR} = 100$ dB.

REFERENCES

[1] A. Belouchrani, K. Abed-Meraïm, J.-F. Cardoso and E. Moulines,"A blind source separation technique using second order statistics", IEEE Transactions on Signal Processing, Vol. 45, pp. 434–444, February 1997.

[2] A. Belouchrani, K. Abed-Meraïm, M. Amin and A. Zoubir, "Joint anti-diagonalization for blind source separation", in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'2001), Salt Lake City, Utah, May 2001.

[3] H. Bousbiah-Salah, A. Belouchrani and K. Abed-Meraïm, "Blind separation of non stationary sources using joint block diagonalization", in Proc. IEEE Workshop on Statistical Signal Processing, pp. 448–451, 2001.

[4] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non Gaussian signals", IEEE Proceedings-F, Vol. 40, pp. 362–370, 1993.

[5] J.-F. Cardoso and B. Laheld, "Equivariant adaptative source separation", IEEE Transactions on Signal Processing, Vol. 44, No. 12, pp. 3017–3030, December 1996.

[6] G. Chabriel, J. Barrère, N. Thirion-Moreau and E. Moreau, "Algebraic joint zero-diagonalization and blind source separation", *IEEE Transactions on Signal Processing*, Vol. 56, n°3, pp. 980-989, March 2008.

[7] P. Comon, "Independant component analysis, a new concept ?", Sig. Processing, Vol. 36, pp. 287–314, 1994.

[8] S. Dégerine, "Sur la diagonalisation conjointe approchée par un critère des moindres carrés", in Proc. 18ème Colloque GRETSI, pp. 311–314, Toulouse, Septembre 2001.

[9] L. DeLathauwer, C. Févotte, B. De Moor and J. Vandewalle, "Jacobi algorithm for joint block diagonalization in blind identification", in Proc. *23rd Symposium on*

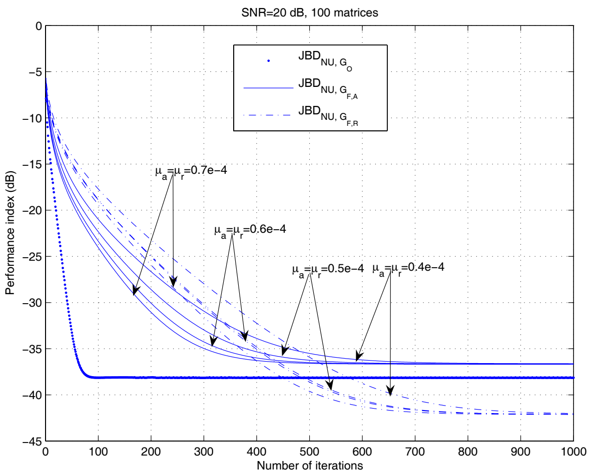

Figure 2: Performance index versus number of iterations for the three algorithms: $N_m = 100$ matrices and $\mathsf{SNR} = 20$ dB.

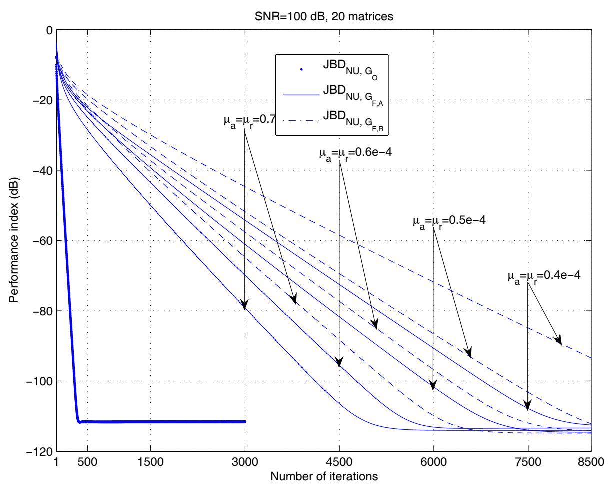

Figure 3: Performance index versus number of iterations for the three algorithms: $N_m = 20$ matrices and $\mathsf{SNR} = 100$ dB.

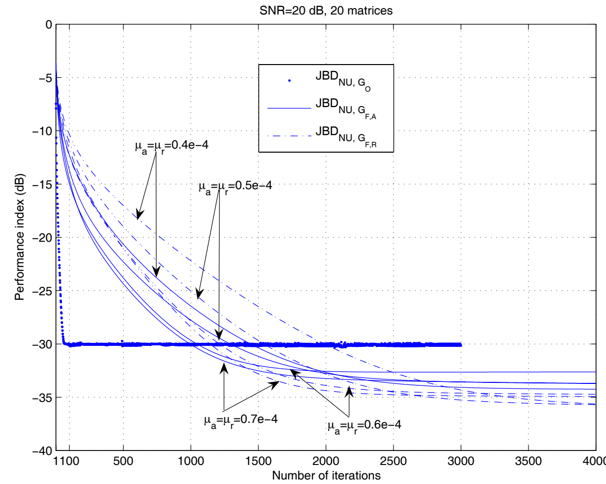

Figure 4: Performance index versus number of iterations for the three algorithms: $N_m = 20$ matrices and $\mathsf{SNR} = 20$ dB.

Information Theory in the Benelux, Louvain-la-Neuve, Belgium, May 2002.

[10] L. DeLathauwer, "Signal processing based on multilinear algebra", PhD Thesis, Université Catholique de Leuven, Belgique, September 1997.

[11] E. M. Fadaili, N. Thirion-Moreau and E. Moreau, "Non orthogonal joint diagonalization/zero-diagonalization for source separation based on time-frequency distributions", IEEE Trans. on Signal Processing, Vol. 55, No 5, pp. 1673-1687, May 2007.

[12] H. Ghennioui, N. Thirion-Moreau, E. Moreau, A. Adib and D. Aboutajdine, "Non unitary joint-block diagonalization of complex matrices using a gradient approach", in Proc. *7th International Conference on Independent Component Analysis and Signal Separation*, LNCS 4666, Springer-Verlag Berlin Heidelberg, LNCS 4666, Mike E. Davies et al. (Eds), pp. 201–208, London, UK, September 2007.

[13] H. Ghennioui, E.-M. Fadaili, N. Thirion-Moreau, A. Adib and E. Moreau, "A non-unitary joint block diagonalization algorithm for blind separation of convolutive mixtures of sources", *IEEE Signal Processing Letters*, pp. 860–863, November 2007.

[14] H. Ghennioui, N. Thirion-Moreau, E. Moreau, A. Adib and D. Aboutajdine, "A novel approach based on non-unitary joint block-diagonalization for the blind MIMO equalization of cyclo-stationary signals", *16th European Signal Processing Conference* (EUSIPCO 2008), Lausanne, Switzerland, 25-29 August 2008.

[15] A. Hjϕrungnes and D. Gesbert, "Complex-valued matrix differentiation: techniques and key results", *IEEE Trans. on Signal Processing*, Vol. 55, No. 6, pp. 2740-2746, June 2007.

[16] M. Joho and K. Rahbar, "Joint diagonalization of correlation matrices by using Newton methods with application to blind signal separation", IEEE Sensor Array and Multichannel Signal Processing Workshop SAM, pp. 403-407, 2002.

[17] E. Moreau, "A generalization of joint-diagonalization criteria for source separation", *IEEE Trans. Signal Processing*, Vol. 49, No. 3, pp. 530–541, March 2001.

[18] D.-T. Pham, "Joint approximate diagonalization of positive definite matrices", in SIAM Journal on Matrix Analysis and Applications, Vol. 22, No. 4, pp. 1136–1152, 2001.

[19] D.-T. Pham, "Blind separation of cyclostationary sources using joint block approximate diagonalization", in Proc. *7th International Conference on Independent Component Analysis and Signal Separation*, Springer-Verlag Berlin Heidelberg, LNCS 4666, Mike E. Davies et al. (Eds), pp. 244-251, London, September 2007.

[20] K. B. Petersen and M. S. Pedersen, "The matrix cookbook", 5 January 2005.

[21] A. Yeredor, A. Ziehe and K.R. Muller, "Approximate joint diagonalization using a natural gradient approach", ICA 2004, LNCS 3195, pp. 89–96, 2004.

[22] A. Ziehe, P. Laskov, G. Nolte, G. and K.-R. Müller, "A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation", Journal of Machine Learning Research, No. 5, pp. 801–818, July 2004.