# NON-UNITARY JOINT BLOCK DIAGONALIZATION OF MATRICES USING A LEVENBERG-MARQUARDT ALGORITHM

*O. Cherrak, H. Ghennioui, E.-H. Abarkan*

Université Sidi Mohamed Ben Abdellah
FSTF, LSSC
Faculté des Sciences et Techniques de Fès,
Route Immouzzer, B.P. 2202, Fès-Maroc
{cherrakomar, ghennioui}@gmail.com

*N. Thirion-Moreau*

Aix-Marseille Université, CNRS,
ENSAM, LSIS, UMR 7296,
13397 Marseille-France
Université de Toulon, CNRS, LSIS,
UMR 7296, 83957 La Garde-France
thirion@univ-tln.fr

## ABSTRACT

This communication addresses the problem of the Non-Unitary Joint Block Diagonalization (NU − JBD) of a given set of complex matrices. This problem occurs in various fields of applications, among which is the blind separation of convolutive mixtures of sources. We present a new method for the NU − JBD based on the Levenberg-Marquardt algorithm (LMA). Our algorithm uses a numerical diagram of optimization which requires the calculation of the complex Hessian matrices. The main advantages of the proposed method stem from the LMA properties: it is powerful, stable and more robust. Computer simulations are provided in order to illustrate the good behavior of the proposed method in different contexts. Two cases are studied: in the first scenario, a set of exactly block-diagonal matrices are considered, then these matrices are progressively perturbed by an additive gaussian noise. Finally, this new NU − JBD algorithm is compared to others put forward in the literature: one based on an optimal step-size relative gradient-descent algorithm [1] and one based on a nonlinear conjugate gradient algorithm [2]. This comparison emphasizes the good behavior of the proposed method.

***Index Terms*—** Joint block-diagonalization algorithms, complex Hessian Matrices, Levenberg-Marquardt algorithm.

## 1. INTRODUCTION

In the recent years, the problem of the joint decomposition of matrix (or tensor) sets has often arisen in the signal processing field, especially in blind source separation and array processing applications. One of the first considered problem was the Joint Diagonalization (JD) of a given matrix set under the unitary constraint, leading to the nowadays well-known JADE (Joint Approximate Diagonalization of Eigenmatrices) [3] and SOBI (Second Order Blind Identification) [4] algorithms. The following works have addressed either the problem of the JD of tensors [5][6] or the problem of JD of matrices but discarding the unitary constraint [7][8][9][10]. This first particular type of matrix decomposition is useful both in sources localization and direction finding problems and in blind sources separation of instantaneous mixtures.

A second type of matrix decompositions, namely joint zero-diagonalization, has proven to be useful in blind source separation, telecommunication [11] and cryptography [12]. The first suggested algorithms operated under the unitary constraint [13], since they were applied after a classical pre-whitening stage. But such a preliminary pre-whitening step establishes a bound with regard to the best reachable performances in the context of BSS that is the reason why the unitary constraint was soon discarded, leading to several solutions.

A third type of matrix decompositions, namely the joint block-diagonalization, is encountered both in the wide-band sources localization in the presence of a correlated noise and in the blind separation of convolutive mixtures (or multi-dimensional deconvolution) problems. Several algorithms have been developed, under different assumptions about the considered matrix set (the matrices can be either positive definite or hermitian) and about the block-diagonalizer (it is assumed unitary [14][15] or not [1][2][16] [17]).

In this communication, we present a new approach for the NU − JBD based on a Levenberg-Marquardt optimization scheme. It requires that quantities such as the Hessian matrices are calculated. The most important advantage of this approach is that it is general (the considered matrices are complex and not necessarily hermitian and the block-diagonalizer is not necessarily an unitary matrix) and robust.

## 2. PROBLEM STATEMENT

### 2.1. Non-unitary joint block-diagonalization problem

The problem of the non-unitary joint block diagonalization is stated in the following way [1][16]. We 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,

$$\mathbf{M}_i = \mathbf{A}\mathbf{D}_i\mathbf{A}^H , \quad (1)$$

where $(\cdot)^H$ stands for the transpose conjugate operator and the matrices $\mathbf{D}_i = \begin{pmatrix} \mathbf{D}_{i,11} & \mathbf{0}_{12} & \ldots & \mathbf{0}_{1r} \\ \mathbf{0}_{21} & \mathbf{D}_{i,22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0}_{r-1r} \\ \mathbf{0}_{r1} & \ldots & \mathbf{0}_{rr-1} & \mathbf{D}_{i,rr} \end{pmatrix}$, for all $i \in \{1,\ldots,N_m\}$ are $(N \times N)$ block diagonal matrices with $r$ the number of considered blocks ($r \in \mathbb{N}^*$), $\mathbf{D}_{i,jj}, i \in \{1,\ldots,N_m\}, j \in \{1,\ldots,r\}$ 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. $\mathbf{A}$ is a $M \times N$ ($M \geq N$) full rank matrix and the $N \times M$ 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 by $\mathcal{D}$. The block sizes $n_j$ for all $j = 1,\ldots,r$ are assumed known.

The $\mathsf{NU}-\mathsf{JBD}$ problem consists of estimating the matrix $\mathbf{A}$ and the block-diagonal matrix set $\mathcal{D}$ from only the matrix set $\mathcal{M}$. Several cases were explored for matrix $\mathbf{A}$ and $\mathbf{M}_i$: the case of an unitary matrix $\mathbf{A}$ has been considered in [15] where a first solution is developped. More recently, other solutions have been suggested in [1][2][16][18], for a non-unitary matrix $\mathbf{A}$ and under different assumptions about the matrix set $\mathcal{M}$. To tackle that problem, we propose, here, to consider the following cost function like in [1][16] (in [2] another cost function is used). The reason is that it enables to simplify the calculations:

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

where the matrix operator $\mathsf{OffBdiag}_{(\mathbf{n})}\{.\}$ is defined in the following way:

$$\mathsf{OffBdiag}_{(\mathbf{n})}\{\mathbf{M}\} = \begin{pmatrix} \mathbf{0}_{11} & \mathbf{M}_{12} & \ldots & \mathbf{M}_{1r} \\ \mathbf{M}_{21} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{M}_{r1} & \mathbf{M}_{r2} & \ldots & \mathbf{0}_{rr} \end{pmatrix}. \quad (3)$$

where $\mathbf{M}$ is a $N \times N$ square matrix whose components $\mathbf{M}_{ij}$ for all $i,j = 1,\ldots,r$ are $n_j \times n_j$ matrices (and $n_1 + \ldots + n_r = N$). We denote by $\mathbf{n} = (n_1, n_2, \ldots, n_r)$. From now on, the notation $\mathsf{OffBdiag}_{(\mathbf{n})}\{.\}$ is simplified into $\mathsf{OffBDiag}\{\cdot\}$.

## 3. A NEW JOINT BLOCK-DIAGONALIZATION ALGORITHM

The cost function given in Eq. (2) has to be minimized in order to estimate the matrix $\mathbf{B} \in \mathbb{C}^{N\times M}$. To do this, we suggest a new algorithm based on a Levenberg-Marquardt approach [19]. In the literature, the Levenberg-Marquardt optimization method is well-known for its robustness and its effectiveness. It has been widely used in various fields (e.g. neural networks [20], array processing, blind sources separation and so on).

### 3.1. Principle

We propose, here, to use a Levenberg-Marquardt optimization scheme [19] to minimize the cost function given by Eq. (2) in order to estimate the block-diagonalizer matrix $\mathbf{B}$. It means that $\mathbf{B}$ is re-estimated at each iteration $m$. From now on, it is denoted by $\mathbf{B}^{(m)}$. This matrix can be stored in a vector $\mathbf{b}^{(m)}$ which is obtained by $\mathbf{b}^{(m)} = \mathbf{vec}\left(\mathbf{B}^{(m)}\right)$, where $\mathbf{vec}(.)$ is the vec-operator (applied on a matrix $\mathbf{B}^{(m)} \in \mathbb{C}^{M\times N}$ it stacks its columns into a column vector belonging to $\mathbb{C}^{(M\times N)\times 1}$). This vector is, thus, updated according to the following adaptation rule for all $m = 1, 2, ...,$

$$\mathbf{b}^{(m)} = \mathbf{b}^{(m-1)} - \mu\left[\mathbf{H}^{(m-1)} + \lambda\mathbf{I}_{M^2}\right]^{-1}\mathbf{g}^{(m-1)}, \quad (4)$$

where $[\cdot]^{-1}$ denotes the inverse of a matrix, $\lambda$ is positive a small damping factor. $\mu$ is small positive factor called the step-size, in our case $\mu = 1$, $\mathbf{I}_{M^2}$ is the $M^2 \times M^2$ identity matrix, $\mathbf{H}$ is the Hessian matrix of $\mathcal{C}_{BD}(\mathbf{B})$. The vector $\mathbf{g}$ is defined as $\mathbf{g} = \mathbf{vec}\left(\nabla_a\mathcal{C}_{BD}\left(\mathbf{B}^{(m-1)}\right)\right)$ where $\nabla_a\mathcal{C}_{BD}(\mathbf{B})$ stands for the complex (absolute) gradient matrix. It was previously calculated in [1] and found to be equal to:

$$\begin{aligned}\nabla_a\mathcal{C}_{BD}(\mathbf{B}) &= 2\sum_{i=1}^{N_m}\left(\mathsf{OffBdiag}\{\mathbf{B}\mathbf{M}_i\mathbf{B}^H\}\right)\mathbf{B}\mathbf{M}_i^H \\ &+ 2\sum_{i=1}^{N_m}\left(\mathsf{OffBdiag}\{\mathbf{B}\mathbf{M}_i\mathbf{B}^H\}\right)^H\mathbf{B}\mathbf{M}_i. \end{aligned} \quad (5)$$

If $(\cdot)^T$ stands for the transpose operator and $(\cdot)^*$ for the conjugate operator, we now have to calculate the four complex Hessian matrices that are defined in [21] as:

$$\mathbf{H}_{\mathbf{B},\mathbf{B}^*}(\mathcal{C}_{BD}(\mathbf{B})) = \frac{\mathbf{A}_{0,0} + \mathbf{A}_{1,1}^T}{2} = \left[\mathbf{H}_{\mathbf{B}^*,\mathbf{B}}(\mathcal{C}_{BD}(\mathbf{B}))\right]^T,$$

$$\mathbf{H}_{\mathbf{B}^*,\mathbf{B}^*}(\mathcal{C}_{BD}(\mathbf{B})) = \frac{\mathbf{A}_{0,1} + \mathbf{A}_{0,1}^T}{2}, \mathbf{H}_{\mathbf{B},\mathbf{B}}(\mathcal{C}_{BD}(\mathbf{B})) = \frac{\mathbf{A}_{1,0} + \mathbf{A}_{1,0}^T}{2}, \quad (6)$$

To that aim, we have to derive the expression of $d^2\mathcal{C}_{BD}(\mathbf{B})$ which is written as:

$$d^2\mathcal{C}_{BD}(\mathbf{B}) = \left[d\mathbf{vec}^T(\mathbf{B}^*), d\mathbf{vec}^T(\mathbf{B})\right]\begin{bmatrix}\mathbf{A}_{0,0}, \mathbf{A}_{0,1} \\ \mathbf{A}_{1,0}, \mathbf{A}_{1,1}\end{bmatrix}\begin{bmatrix} d\mathbf{vec}(\mathbf{B}) \\ d\mathbf{vec}(\mathbf{B}^*)\end{bmatrix}. \quad (7)$$

The four complex matrices $\mathbf{A}_{0,0}$, $\mathbf{A}_{0,1}$, $\mathbf{A}_{1,0}$ and $\mathbf{A}_{1,1}$ involved in this expression are shown to be equal to (see the Appendix for the detailed calculations),

$$\begin{aligned}\mathbf{A}_{0,0} &= \left(\mathbf{M}_i^H\mathbf{B}^H \otimes \mathbf{I}_N\right)\mathbf{T}_{\text{Boff}}\left(\mathbf{B}\mathbf{M}_i \otimes \mathbf{I}_N\right) \\ &+ \left(\mathbf{M}_i\mathbf{B}^H \otimes \mathbf{I}_N\right)\mathbf{T}_{\text{Boff}}\left(\mathbf{B}\mathbf{M}_i^H \otimes \mathbf{I}_N\right) \\ &+ \left(\mathbf{M}_i \otimes \left(\mathsf{OffBdiag}\{\mathbf{B}\mathbf{M}_i\mathbf{B}^H\}\right)\right) \\ &+ \left(\mathbf{M}_i^H \otimes \left(\mathsf{OffBdiag}\{\mathbf{B}\mathbf{M}_i^H\mathbf{B}^H\}\right)\right). \end{aligned} \quad (8)$$

$$\begin{aligned}\mathbf{A}_{0,1} &= (\mathbf{T}_{vec})^T\left(\left(\mathbf{I}_N \otimes \mathbf{M}_i^T\mathbf{B}^T\right)\mathbf{T}_{\text{Boff}}\left(\mathbf{B}^*\mathbf{M}_i^T \otimes \mathbf{I}_N\right)\right. \\ &+ \left.\left(\mathbf{I}_N \otimes \mathbf{M}_i^*\mathbf{B}^T\right)\mathbf{T}_{\text{Boff}}\left(\mathbf{B}^*\mathbf{M}_i^* \otimes \mathbf{I}_N\right)\right). \end{aligned} \quad (9)$$

$$\begin{aligned}\mathbf{A}_{1,0} &= (\mathbf{T}_{vec})^T\left(\left(\mathbf{I}_N \otimes \mathbf{M}_i\mathbf{B}^H\right)\mathbf{T}_{\text{Boff}}\left(\mathbf{B}\mathbf{M}_i \otimes \mathbf{I}_N\right)\right. \\ &+ \left.\left(\mathbf{I}_N \otimes \mathbf{M}_i^H\mathbf{B}^H\right)\mathbf{T}_{\text{Boff}}\left(\mathbf{B}\mathbf{M}_i^H \otimes \mathbf{I}_N\right)\right). \end{aligned} \quad (10)$$

$$\begin{aligned}\mathbf{A}_{1,1} &= \left(\mathbf{M}_i^*\mathbf{B}^T \otimes \mathbf{I}_N\right)\mathbf{T}_{\text{Boff}}\left(\mathbf{B}^*\mathbf{M}_i^T \otimes \mathbf{I}_N\right) \\ &+ \left(\mathbf{M}_i^T\mathbf{B}^T \otimes \mathbf{I}_N\right)\mathbf{T}_{\text{Boff}}\left(\mathbf{B}^*\mathbf{M}_i^* \otimes \mathbf{I}_N\right) \\ &+ \left(\mathbf{M}_i^T \otimes \left(\mathsf{OffBdiag}\{\mathbf{B}\mathbf{M}_i\mathbf{B}^H\}\right)\right) \\ &+ \left(\mathbf{M}_i^* \otimes \left(\mathsf{OffBdiag}\{\mathbf{B}\mathbf{M}_i^H\mathbf{B}^H\}\right)^*\right). \end{aligned} \quad (11)$$

where the operator $\otimes$ denotes the Kronecker product, $\mathbf{T}_{vec}$ is a permutation square matrix of size $(MN \times MN)$ and

$\mathbf{T}_{\text{Boff}} = \mathbf{I}_{N^2} - \mathbf{T}_{\text{Diag}}$, is the $N^2 \times N^2$ "transformation" matrix, with $\mathbf{T}_{\text{Diag}} = \text{diag}\{\mathbf{vec}(\text{BDiag}\{\mathbf{1}_N\})\}$, $\mathbf{1}_N$ is the $N \times N$ matrix whose components are all ones, $\mathbf{I}_{N^2} = \text{Diag}\{\mathbf{1}_{N^2}\}$ is the $N^2 \times N^2$ identity matrix, and $\text{Diag}\{\mathbf{a}\}$ is a square diagonal matrix which contains the elements of the vector $\mathbf{a}$.

### 3.2. Summary

The principle of the proposed non-unitary joint block diagonalization algorithm (denoted $\text{JBD}_{\text{LM}}$) is summarized in the following table,

Denote the $N_m$ square matrices as $\mathbf{M}_1, \ldots, \mathbf{M}_{N_m}$.
Given initial estimate $\mathbf{B}^{(0)}$.
Given initial estimate for damping value $\lambda^{(0)}$.
**For** $m = 1, 2, \ldots$

- Compute $\nabla_a \mathcal{C}_{BD}(\mathbf{B})$ whose expression is given by Eq. (5).
- Compute the four complex Hessian matrices whose expressions are given by Eq. (6).
- Compute the error $\mathbf{e}^{(m)}$.
- Compute the updated solution whose expression is given by Eq. (4).
- Evaluate the error at the updated parameter $\mathbf{e}_{lm}$.

  **If** $\mathbf{e}^{(m)} \geq \mathbf{e}_{lm}$,

  $$\lambda^{(m)} = \frac{\lambda^{(m-1)}}{10},\ \mathbf{e}^{(m)} = \mathbf{e}_{lm}.$$

  **Else**,

  $$\lambda^{(m)} = 10\lambda^{(m-1)}.$$

  **EndIf**
- Stop after a fixed number of iterations or when $|\mathbf{B}^{(m)} - \mathbf{B}^{(m-1)}| \leq \varepsilon$ where $\varepsilon$ is a small positive threshold.

**EndFor**

## 4. COMPUTER SIMULATIONS

We present simulations to illustrate the effectiveness of the suggested algorithm. We consider a set $\mathcal{D}$ of $N_m = 2, 10$ and $100$ 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 $\sigma_b^2$ is added. The signal to noise ratio is defined as $\text{SNR} = 10\log(\frac{1}{\sigma_b^2})$. To measure the quality of the estimation, the ensuing performance index (which is an extension of the one introduced in [6]) is used,

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

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 100 Monte-Carlo trials.

First, we compare the results (in terms of performance index) obtained thanks to the three algorithms $\text{JBD}_{\text{ORG}}$ (Optimal Relative Descent-Gradient) [1], $\text{JBD}_{\text{CG}}$ (Non-linear Conjugate Gradient) [2] and the proposed algorithm $\text{JBD}_{\text{LM}}$ versus the number of iterations for number of blocks $r = 2$ for different numbers of matrices ($N_m = 2$, $N_m = 10$ and $N_m = 100$). The obtained results are displayed in the Fig.1 in a noiseless context ($\text{SNR} = 100$ dB) whereas, in the Fig.2 the same study is performed in a noisy context ($\text{SNR} = 40$ dB). All these methods were initialized using the generalized eigenvalue decomposition [2].

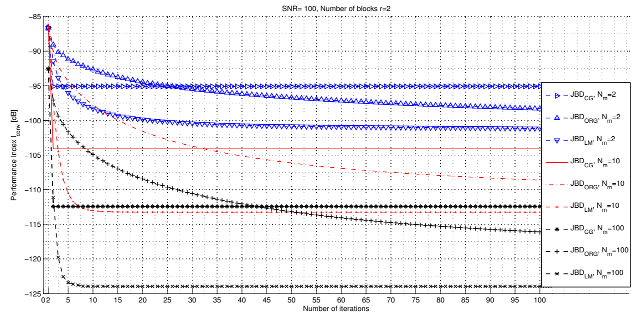


**Fig. 1**. Performance index versus number of iterations for number of matrices : $N_m = 2$, $N_m = 10$ and $N_m = 100$, number of blocks $r = 2$ and $\text{SNR} = 100$ dB.

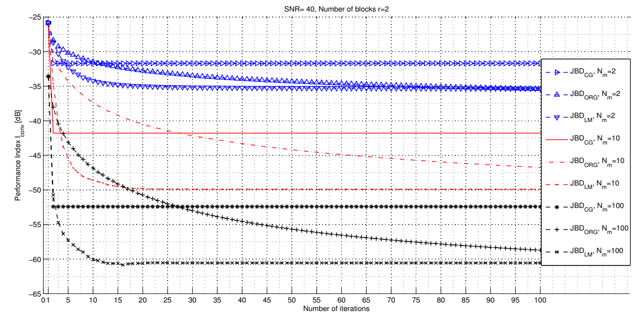


**Fig. 2**. Performance index versus number of iterations for number of matrices : $N_m = 2$, $N_m = 10$ and $N_m = 100$, number of blocks $r = 2$ and $\text{SNR} = 40$ dB.

In the noiseless case, the $\text{JBD}_{\text{LM}}$ algorithm outperforms the two other algorithms ($\text{JBD}_{\text{CG}}$, $\text{JBD}_{\text{ORG}}$) for $N_m = 2$ (resp. 10 and 100), $-101$ dB instead of $-96$ dB and $-98$ dB (resp.$-113$ dB instead of $-104$ dB and $-108$ dB and $-124$ dB instead of $-112$ dB and $-113$ dB). In the noisy context ($\text{SNR} = 40$ dB) for different number of $N_m = 2$ (resp. 10 and 100), $\text{JBD}_{\text{LM}}$ outperforms $\text{JBD}_{\text{CG}}$, $\text{JBD}_{\text{ORG}}$. $\text{JBD}_{\text{LM}}$ and $\text{JBD}_{\text{ORG}}$ reach nearly the same performances which is not surprising since these two algorithms minimize the same cost function. $-35$ dB is reached instead of $-31$ dB for $\text{JBD}_{\text{CG}}$ (resp. -50 dB ($\text{JBD}_{\text{LM}}$) instead of $-42$ ($\text{JBD}_{\text{CG}}$) dB and $-47$ dB ($\text{JBD}_{\text{ORG}}$) and $-61$ dB instead of $-53$ dB and $-58$ dB).

Then, we show the influence of the size $N_m$ of the matrix set to be block-diagonalized. We display the evolution of the performance index versus the number of matrices $N_m$ for the number of blocks $r = 2$ and 4 in a noiseless context in the Fig.3 and in the noisy context in the Fig.4.

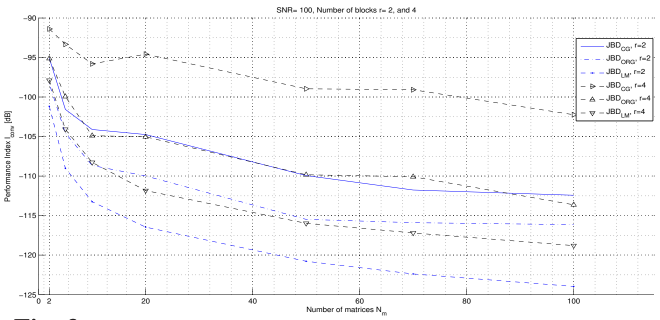


**Fig. 3**. Performance index versus number of matrices for size of blocks: $r = 2$, $r = 4$ and $\mathsf{SNR} = 100$ dB.

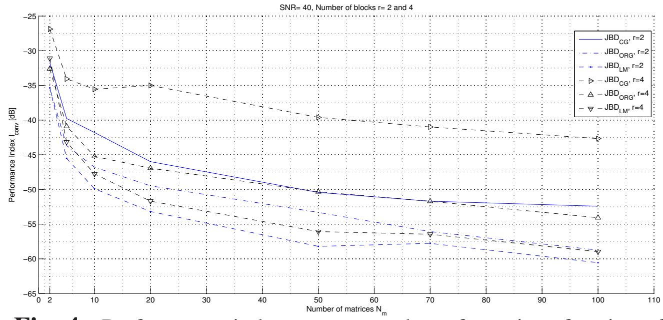


**Fig. 4**. Performance index versus number of matrices for size of blocks: $r = 2$, $r = 4$ and $\mathsf{SNR} = 40$ dB.

These charts illustrate the good behavior of the proposed algorithm. We observe a decrease in the performances when smaller subsets of matrices are used, even though the results remain relatively good.

## 5. CONCLUSION

In this communication, we have suggested a new $\mathsf{NU-JBD}$ algorithm called $\mathsf{JBD_{LM}}$. It relies on a Levenberg-Marquardt optimization scheme which requires the calculation of the complex Hessian matrices. The main advantage of the $\mathsf{JBD_{LM}}$ is that it offers the best performances in complicated situations (noisy cases, and/or few matrices to be joint block-diagonalized). Extensions for future research would be to demonstrate the interest of this algorithm in a true blind multi-dimensional deconvolution context (blind separation of convolutive mixtures of sources).

## APPENDIX

Considering four square matrices $\mathbf{D}_1, \mathbf{D}_2, \mathbf{D}_3$ of size $M \times M$ and $\mathbf{D}_4$ of size $N \times N$. $\mathsf{tr}\{\cdot\}$ denotes the trace operator. Our developments are based on the following properties [21][22],

- $\mathbf{P}_1$. $\|\mathsf{OffBdiag}\{\mathbf{D}_1\}\|_F^2 = \mathsf{tr}\left\{\mathbf{D}_1^H \mathsf{OffBdiag}\{\mathbf{D}_1\}\right\}$.
- $\mathbf{P}_2$. $\mathsf{tr}\{\mathbf{D}_1\} = \mathsf{tr}\left\{\mathbf{D}_1^T\right\}$.
- $\mathbf{P}_3$. $\mathsf{tr}\{\mathbf{D}_1 + \mathbf{D}_2\} = \mathsf{tr}\{\mathbf{D}_1\} + \mathsf{tr}\{\mathbf{D}_2\}$.
- $\mathbf{P}_4$. $\mathsf{tr}\{\mathbf{D}_1\mathbf{D}_2\mathbf{D}_3\} = \mathsf{tr}\{\mathbf{D}_3\mathbf{D}_1\mathbf{D}_2\} = \mathsf{tr}\{\mathbf{D}_2\mathbf{D}_3\mathbf{D}_1\}$.
- $\mathbf{P}_5$. $\mathsf{tr}\left\{\mathbf{D}_1^H\mathbf{D}_2\right\} = (\mathbf{vec}\,(\mathbf{D}_1))^H\,\mathbf{vec}\,(\mathbf{D}_2)$.
- $\mathbf{P}_6$. $d\,(\mathsf{tr}\{\mathbf{D}_1\}) = \mathsf{tr}\{d\,(\mathbf{D}_1)\}$.
- $\mathbf{P}_7$. $\mathbf{vec}\,(\mathbf{D}_1 + \mathbf{D}_2) = \mathbf{vec}\,(\mathbf{D}_1) + \mathbf{vec}\,(\mathbf{D}_2)$.
- $\mathbf{P}_8$. $\mathbf{vec}\,(\mathbf{D}_1\mathbf{D}_2\mathbf{D}_3) = \left(\mathbf{D}_1^H \otimes \mathbf{D}_1\right)\mathbf{vec}\,(\mathbf{D}_2)$.
- $\mathbf{P}_9$. $\mathbf{vec}\,\left(\mathbf{D}_1^T\right) = \mathbf{T}_{vec}\mathbf{vec}\,(\mathbf{D}_1)$.
- $\mathbf{P}_{10}$. $\mathbf{vec}\,(\mathsf{OffBdiag}\{\mathbf{D}_1\}) = \mathbf{T}_{\mathsf{Boff}}\mathbf{vec}\,(\mathbf{D}_1)$.
- $\mathbf{P}_{11}$. $d\,(\mathbf{vec}\,(\mathbf{D}_1\}) = \mathbf{vec}\,(d\,(\mathbf{D}_1))$.
- $\mathbf{P}_{12}$. $(\mathbf{D}_1 \otimes \mathbf{D}_2)^T = \mathbf{D}_1^T \otimes \mathbf{D}_2^T$.
- $\mathbf{P}_{13}$. $(\mathbf{D}_1 \otimes \mathbf{D}_2)^H = \mathbf{D}_1^H \otimes \mathbf{D}_2^H$.
- $\mathbf{P}_{14}$. $(\mathbf{D}_1 \otimes \mathbf{D}_2)(\mathbf{D}_3 \otimes \mathbf{D}_4) = (\mathbf{D}_1\mathbf{D}_3 \otimes \mathbf{D}_2\mathbf{D}_4)$.

The second-order derivative of the cost function (Eq. (2)) is defined like in [21],

$$d^2\mathcal{C}_{BD}(\mathbf{B}) = d\mathcal{D}_B\,(\mathcal{C}_{BD}(\mathbf{B}))\,d\mathbf{vec}(\mathbf{B}) + d\mathcal{D}_B^*\,(\mathcal{C}_{BD}(\mathbf{B}))\,d\mathbf{vec}(\mathbf{B}^*), \tag{12}$$

where $\mathcal{D}_B\,(\mathcal{C}_{BD}(\mathbf{B}))$ and $\mathcal{D}_B^*\,(\mathcal{C}_{BD}(\mathbf{B}^*))$ are the Jacobian matrices of the $\mathcal{C}_{BD}(\mathbf{B})$, defined like in [21],

$$\begin{aligned}\mathcal{D}_B\,(\mathcal{C}_{BD}(\mathbf{B})) &= \mathbf{vec}^T\left(\frac{\partial\mathcal{C}_{BD}(\mathbf{B})}{\partial\mathbf{B}}\right),\\ \mathcal{D}_B^*\,(\mathcal{C}_{BD}(\mathbf{B})) &= \mathbf{vec}^T\left(\frac{\partial\mathcal{C}_{BD}(\mathbf{B})}{\partial\mathbf{B}^*}\right).\end{aligned} \tag{13}$$

The partial derivatives $\frac{\partial\mathcal{C}_{BD}(\mathbf{B})}{\partial\mathbf{B}}$ and $\frac{\partial\mathcal{C}_{BD}(\mathbf{B})}{\partial\mathbf{B}^*}$ of the cost function $\mathcal{C}_{BD}(\mathbf{B})$ (2) are calculated in [1],

$$\begin{aligned}\frac{\partial\mathcal{C}_{BD}(\mathbf{B})}{\partial\mathbf{B}} &= \sum_{i=1}^{N_m}\left(\mathsf{OffBdiag}_{(n)}\{\mathbf{B}\mathbf{M}_i\mathbf{B}^H\}\right)^T\mathbf{B}^*\mathbf{M}_i^*\\ &\quad+\sum_{i=1}^{N_m}\left(\mathsf{OffBdiag}\{\mathbf{B}\mathbf{M}_i\mathbf{B}^H\}\right)^*\mathbf{B}^*\mathbf{M}_i^T,\end{aligned} \tag{14}$$

$$\begin{aligned}\frac{\partial\mathcal{C}_{BD}(\mathbf{B})}{\partial\mathbf{B}^*} &= \sum_{i=1}^{N_m}\left(\mathsf{OffBdiag}\{\mathbf{B}\mathbf{M}_i\mathbf{B}^H\}\right)\mathbf{B}\mathbf{M}_i^H\\ &\quad+\sum_{i=1}^{N_m}\left(\mathsf{OffBdiag}\{\mathbf{B}\mathbf{M}_i\mathbf{B}^H\}\right)^H\mathbf{B}\mathbf{M}_i.\end{aligned} \tag{15}$$

Using the previous properties ($\mathbf{P}_1$, $\mathbf{P}_2$, ..., $\mathbf{P}_{14}$) and the expression of the partial derivatives $\frac{\partial\mathcal{C}_{BD}(\mathbf{B})}{\partial\mathbf{B}}$ (14) and $\frac{\partial\mathcal{C}_{BD}(\mathbf{B})}{\partial\mathbf{B}^*}$ (15), the derivate of the Jacobian matrices can be expressed as,

$$\begin{aligned}&d\mathcal{D}_B\,(\mathcal{C}_{BD}(\mathbf{B})) =\\ &\sum_{i=1}^{N_m}\mathbf{vec}^T\,(d\mathbf{B})\Big[(\mathbf{T}_{vec})^T\left(\mathbf{I}_N\otimes\mathbf{M}_i\mathbf{B}^H\right)\mathbf{T}_{\mathsf{Boff}}\,(\mathbf{B}\mathbf{M}_i\otimes\mathbf{I}_N)\\ &\quad+(\mathbf{T}_{vec})^T\left(\mathbf{I}_N\otimes\mathbf{M}_i^H\mathbf{B}^H\right)\mathbf{T}_{\mathsf{Boff}}\left(\mathbf{B}\mathbf{M}_i^H\otimes\mathbf{I}_N\right)\Big]\\ &\quad+\mathbf{vec}^T\,(d\mathbf{B}^*)\Big[\left(\mathbf{M}_i^H\mathbf{B}^H\otimes\mathbf{I}_N\right)\mathbf{T}_{\mathsf{Boff}}\,(\mathbf{B}\mathbf{M}_i\otimes\mathbf{I}_N)\\ &\quad+\left(\mathbf{M}_i\mathbf{B}^H\otimes\mathbf{I}_N\right)\mathbf{T}_{\mathsf{Boff}}\left(\mathbf{B}\mathbf{M}_i^H\otimes\mathbf{I}_N\right)\\ &\quad+\left(\mathbf{M}_i\otimes\left(\mathsf{OffBdiag}\{\mathbf{B}\mathbf{M}_i\mathbf{B}^H\}\right)\right)\\ &\quad+\left(\mathbf{M}_i^H\otimes\left(\mathsf{OffBdiag}\{\mathbf{B}\mathbf{M}_i^H\mathbf{B}^H\}\right)\right)\Big],\end{aligned} \tag{16}$$

$$\begin{aligned}&d\mathcal{D}_B^*\,(\mathcal{C}_{BD}(\mathbf{B})) =\\ &\sum_{i=1}^{N_m}\mathbf{vec}^T\,(d\mathbf{B}^*)\Big[(\mathbf{T}_{vec})^T\left(\mathbf{I}_N\otimes\mathbf{M}_i^T\mathbf{B}^T\right)\mathbf{T}_{\mathsf{Boff}}\left(\mathbf{B}^*\mathbf{M}_i^T\otimes\mathbf{I}_N\right)\\ &\quad+(\mathbf{T}_{vec})^T\left(\mathbf{I}_N\otimes\mathbf{M}_i^*\mathbf{B}^T\right)\mathbf{T}_{\mathsf{Boff}}\,(\mathbf{B}^*\mathbf{M}_i^*\otimes\mathbf{I}_N)\Big]\\ &\quad+\mathbf{vec}^T\,(d\mathbf{B})\Big[\left(\mathbf{M}_i^*\mathbf{B}^T\otimes\mathbf{I}_N\right)\mathbf{T}_{\mathsf{Boff}}\left(\mathbf{B}^*\mathbf{M}_i^T\otimes\mathbf{I}_N\right)\\ &\quad+\left(\mathbf{M}_i^T\mathbf{B}^T\otimes\mathbf{I}_N\right)\mathbf{T}_{\mathsf{Boff}}\,(\mathbf{B}^*\mathbf{M}_i^*\otimes\mathbf{I}_N)\\ &\quad+\left(\mathbf{M}_i^T\otimes\left(\mathsf{OffBdiag}\{\mathbf{B}\mathbf{M}_i\mathbf{B}^H\}\right)\right)\\ &\quad+\left(\mathbf{M}_i^*\otimes\left(\mathsf{OffBdiag}\{\mathbf{B}\mathbf{M}_i^H\mathbf{B}^H\}\right)^*\right)\Big].\end{aligned} \tag{17}$$

So from Eq. (7) and Eq. (12), it leads to the results stated by Eqs. (8, ..., 11). Finally, we can express the four complex Hessian matrices in the form of Eq. (6).

## 6. REFERENCES


[1] H. Ghennioui, N. Thirion-Moreau, E. Moreau, and D. Aboutajdine, "Gradient-based joint block diagonalization algorithms: Application to blind separation of fir convolutive mixtures," *Signal Proc.*, vol. 90, no. 6, pp. 1836–1849, June 2010.

[2] D. Nion, "A tensor framework for non-unitary joint block diagonalization," *IEEE Trans. Signal Proc.*, vol. 59, no. 10, pp. 4585–4594, Oct. 2011.

[3] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-gaussian signals," *IEEE Proc.-F*, vol. 40, no. 6, pp. 362–370, 1993.

[4] A. Belouchrani, K. Abed-Meraïm, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second order statistics," *IEEE Trans. Signal Proc.*, vol. 45, pp. 434–444, Feb. 1997.

[5] P. Comon, "Independant component analysis, a new concept ?," *Signal Proc.*, vol. 36, pp. 287–314, 1994.

[6] E. Moreau, "A generalization of joint-diagonalization criteria for source separation," *IEEE Trans. Signal Proc.*, vol. 49, no. 3, pp. 530–541, March 2001.

[7] A. Yeredor, A. Ziehe, and K.-R. Müller, "Approximate joint diagonalization using natural gradient approach," in *Lecture Notes in Computer Science, in Proceedings ICA*, Granada, Espagne, May 2004, pp. 89–96.

[8] A. Souloumiac, "Nonorthogonal joint diagonalization by combining givens and hyperbolic rotations," June 2009, vol. 57, pp. 2222–2231.

[9] X.-L. Li and X. D. Zhang, "Nonorthogonal joint diagonalization free of degenerate solution," *IEEE Trans. Signal Proc.*, vol. 55, no. 5, pp. 1803–1814, May 2007.

[10] G. Chabriel and J. Barrère, "A direct algorithm for nonorthogonal approximate joint diagonalization," *IEEE Trans. Signal Proc.*, vol. 60, no. 1, pp. 39–47, 2012.

[11] G. Chabriel and J. Barrère, "Non-symmetrical joint zero-diagonalization and mimo zero-division multiple access," *IEEE Trans. Signal Proc.*, vol. 59, no. 5, pp. 2296–2307, 2011.

[12] J. Walgate, A. J. Short, L. Hardy, and V. Vedral, "Local distinguishability of multipartite orthogonal quantum states," *Physical Review Letter*, vol. 85, pp. 4972–4975, 2000.

[13] A. Belouchrani, K. Abed-Meraïm, M. G. Amin, and A. Zoubir, "Joint anti-diagonalization for blind source separation," in *Proc. IEEE ICASSP*, Salt Lake City, UT, May 2001, pp. 2196–2199.

[14] A. Belouchrani, M. G. Amin, and K. Abed-Meraïm, "Direction finding in correlated noise fields based on joint block-diagonalization of spatio-temporal correlation matrices," *IEEE Signal Lett.*, vol. 4, no. 9, pp. 266–268, Sept. 1997.

[15] L. De Lathauwer, C. Févotte, B. De Moor, and J. Vandewalle, "Jacobi algorithm for block diagonalization in blind identification," in *Proc. Symp. Inform. Theory in the Benelux*, Louvain-la-Neuve, Belgium, May 2002, pp. 155–162.

[16] H. Ghennioui, E.-M. Fadaili, N. Thirion-Moreau, A. Adib, and E. Moreau, "A nonunitary joint block diagonalization algorithm for blind separation of convolutive mixtures of sources," *IEEE Signal Lett.*, vol. 14, no. 11, pp. 860–863, November 2007.

[17] P. Tichavsky and Z. Koldovsky, "Algorithms for nonorthogonal approximate joint block-diagonalization," in *Proceedings EUSIPCO*, Bucharest, Romania, August 2012, pp. 2094–2098.

[18] D.-T. Pham, "Blind separation of cyclostationary sources using joint block approximate diagonalization," in *ICA, Springer-Verlag Berlin Heidelberg LNCS 4666, Mike E. Davies et al. (Eds)*, Sept. 2007, pp. 244–251.

[19] D. G. Luenberger, *Optimization by Vector Space Methods*, John Wiley, New York, 1969.

[20] B. M. Wilamowski and H. Yu, "Improved computation for levenberg-marquardt training," *IEEE trans. on Neural Networks*, vol. 21, no. 6, pp. 25–29, June 2010.

[21] A. Hj$\phi$rungnes, *Complex-Valued Matrix Derivatives With Applications in Signal Processing and Communications*, Cambridge, University Press, February 2011.

[22] M. Joho and K. Rahbar, "Joint diagonalization of correlation matrices by using newton methods with application to blind signal separation," in *SAM Workshop*. IEEE, 2002, pp. 403–407.