JACOBI LIKE ALGORITHM FOR NON-ORTHOGONAL JOINT DIAGONALIZATION OF HERMITIAN MATRICES

Victor MAURANDI(1,2), *Eric MOREAU*(1,2) *and Christophe DE LUIGI*(1,2)

(1) Aix Marseille Université, CNRS, ENSAM, LSIS, UMR 7296
13397 Marseille, France.

(2) Université de Toulon, CNRS, LSIS, UMR 7296
83957 La Garde, France.

Email : {maurandi, moreau, deluigi}@univ-tln.fr

ABSTRACT

In this paper, we consider the problem of non-orthogonal joint diagonalization of a set of hermitian matrices. This appears in many blind signal processing problems as source separation and independent component analysis. We propose a new Jacobi like algorithm based on a LU decomposition. The main point consists of the analytical derivation of the elementary two by two matrix. In order to determine the diagonalizing matrix parameters, we propose a useful approximation. Numerical simulations illustrate the overall good performances of the proposed algorithm in comparison to two other Jacobi like algorithms existing in the literature.

***Index Terms*—** Blind source separation, Independent component analysis, Jacobi algorithm, Joint diagonalization

1. INTRODUCTION

Joint diagonalization of sets of matrices is an important issue in blind signal processing and more particularly in source separation [1]-[16]. The purpose of joint diagonalization is to estimate an unknown matrix which jointly diagonalizes a set of matrices. For this, different approaches have been proposed in the literature. Some of them are directly based on the estimation of the diagonalizing matrix [1, 2, 3] and some others use a decomposition allowing to implement a Jacobi like iterative scheme. We mention that Jacobi like methods are really simple to implement and often allows a computational parallelism. This is particularly the case when one can derive an analytical solution in the basic 2×2 matrix case. In this paper we focus on such approach.

We notice that the unitary case has been first considered yielding the two popular algorithms JADE [4] and SOBI [5] in the hermitian case. The JADE algorithm was generalized in [6]. The other interesting case, the complex symmetric one, was considered in [7]. Nowadays the non-unitary case has become important mainly because it allows to skip a first processing step (whitening of the observations in source separation) which limits the performances in practical cases. Recently non-unitary Jacobi like algorithms have been proposed for the real case in [8, 9], for the complex symmetric case in [10] and for the hermitian case in [11] and [12]. In this last paper, Pham develops an algorithm in the same spirit that the one proposed here but only in the framework of the positive definite hermitian matrices. As will be shown hereafter, the hermitian case is specific and is not directly related to the complex symmetric one.

In this paper, we thus develop a new Jacobi like algorithm for the non-orthogonal joint diagonalization of hermitian matrices. The main point of the paper is the derivation of an analytical solution in the basic 2×2 case leading to a very simple implementation. By computer simulations, we illustrate the good behavior of the proposed algorithm and we compare it to the ones suggested in [11] and in [9]. The latter one is generalized to the hermitian case for the simulation purposes.

2. PROBLEM FORMULATION

We consider K ($K \geq 2$) hermitian matrices, $\mathbf{M}_k$, $k = 1, \dots, K$, defined as:

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

where $(\cdot)^H$ is the conjugate transpose operator. Throughout the paper, we consider the case of square matrices, all of them of size $N \times N$. The matrices $\mathbf{D}_k$ are diagonal and real, the matrix $\mathbf{A}$ is complex invertible. Here $\mathbf{A}$ is the so-called mixing matrix.

From the set $\{\mathbf{M}_k\}$, the objective is to estimate the diagonalizing matrix $\mathbf{B}$ (ideally equal to the inverse of the mixing matrix $\mathbf{A}$ up to the product of a diagonal matrix and a permutation matrix) such that the matrices $\mathbf{B}\mathbf{M}_k\mathbf{B}^H$ are (approximately) jointly diagonal. The cost function used to jointly diagonalize the $\mathbf{M}_k$ matrices is the inverse criterion which is

The authors thank the *Direction Générale de l'Armement (DGA)* for its financial support.

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

defined as:

$$J(\mathbf{B}) = \sum_{k=1}^{K} ||\mathsf{Zdiag}\{\mathbf{B}\mathbf{M}_k\mathbf{B}^H\}||_F^2, \tag{2}$$

where $\mathsf{Zdiag}\{\mathbf{X}\}$ is the matrix defined as the matrix $\mathbf{X}$ with zeros on its main diagonal and where $||\cdot||_F$ is the Frobenius norm. In order to jointly diagonalize the set $\{\mathbf{M}_k\}$, we have to estimate an invertible matrix $\mathbf{B}$ minimizing $J(\mathbf{B})$. For this purpose, we propose a Jacobi like algorithm based on the LU decomposition.

3. PROPOSED ALGORITHMS

All square matrices can be decomposed as $\mathbf{DPLU}$, where $\mathbf{D}$ is a diagonal matrix, $\mathbf{P}$ is a permutation matrix and $\mathbf{L}$ and $\mathbf{U}$ are, respectively lower and upper triangular matrices with diagonal coefficients equal to 1. In the joint diagonalization problem, matrices $\mathbf{D}$ and $\mathbf{P}$ correspond to classical indeterminacies and can thus be dropped. Then it remains to determine the diagonalizing matrix as $\mathbf{B} = \mathbf{LU}$. A great advantage of this parameterization is that $\mathbf{B}$ will be invertible.

In order to estimate the matrix $\mathbf{B}$, we consider a Jacobi like procedure. This procedure consists of decomposing a problem of size $N \times N$ as $\frac{N(N-1)}{2}$ sub-problems of "size" 2×2 considering all couples with the same indexes of rows and columns. For example, in case $N = 3$, this can be written as:

$$\begin{aligned}\mathbf{B} &= \prod_{i=1}^{N-1} \prod_{j=i+1}^{N} \mathbf{B}^{ij} \\ &= \begin{pmatrix} B_{11}^1 & B_{12}^1 & 0 \\ B_{21}^1 & B_{22}^1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} B_{11}^2 & 0 & B_{13}^2 \\ 0 & 1 & 0 \\ B_{31}^2 & 0 & B_{33}^2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & B_{22}^3 & B_{23}^3 \\ 0 & B_{32}^3 & B_{33}^3 \end{pmatrix},\end{aligned} \tag{3}$$

where, for a given N, the uppervalue of the $\mathbf{B}^{ij}$ components, noted x, is a bijective function of i and j defined as

$$x = (i-1)N + j - \frac{i(i+1)}{2} \tag{4}$$

with $(i, j) \in \{1, \dots, N\}^2$ and $i < j$. So x varies between 1 and $\frac{N(N-1)}{2}$.

Notice that the elementary matrices $\mathbf{B}^{ij}$ only depend on (at most) four parameters which each corresponds to a 2×2 submatrix. The derivation of each of these 2×2 submatrices is detailed in the next section. When all useful indexes (i, j) have been considered, this corresponds to a sweep. Notice that the matrix set is updated as

$$\mathbf{M}_k \longleftarrow \mathbf{B}^{ij}\mathbf{M}_k\mathbf{B}^{ij\,H}, \tag{5}$$

after each elementary matrix is derived. The sweep is iterated until convergence. Now we focus on the 2×2 matrix case and we derive an analytical solution.

3.1. The analytical resolution of the 2×2 case

The 2×2 diagonalizing matrix $\mathbf{B}$ is defined as the following $\mathbf{L}$ and $\mathbf{U}$ matrix product

$$\mathbf{B} = \begin{pmatrix} 1 & 0 \\ \ell & 1 \end{pmatrix} \begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & u \\ \ell & 1 + \ell u \end{pmatrix}, \tag{6}$$

where ℓ and u are the sought complex parameters.

Let $\mathbf{M}'_k$, $k = 1, \dots, K$, the matrices of size 2×2 defined as

$$\mathbf{M}'_k = \begin{pmatrix} M'_{k,11} & M'_{k,12} \\ M'_{k,21} & M'_{k,22} \end{pmatrix} = \mathbf{B}\mathbf{M}_k\mathbf{B}^H. \tag{7}$$

The update (5) keeps the $\mathbf{M}_k$ hermitian property, so $\mathbf{M}'_k$ matrices are hermitian too. Thus $M'_{k,11}$ and $M'_{k,22}$ are real and $M'_{k,21} = M'^{*}_{k,12}$, where $(\cdot)^*$ is the conjugate operator. Hence, the cost function in (2), in the 2×2 case, named $J_2(\mathbf{B})$ can be rewritten as:

$$J_2(\mathbf{B}) = 2 \sum_{k=1}^{K} |M'^{*}_{k,12}|^2. \tag{8}$$

In using (7) and (6), we have $M'^{*}_{k,12} = \mathbf{c}_k^T \mathbf{p}$ with

$$\mathbf{p} = \begin{pmatrix} \ell \\ 1 + \ell u \\ \ell u^* \\ u^*(1 + \ell u) \end{pmatrix} \text{ and } \mathbf{c}_k = \begin{pmatrix} M_{k,11} \\ M^*_{k,12} \\ M_{k,12} \\ M_{k,22} \end{pmatrix}. \tag{9}$$

Now the cost function (8) can be written as the quadratic form:

$$J_2(\mathbf{B}) = 2\,\mathbf{p}^H \sum_{k=1}^{K} \mathbf{c}_k^* \mathbf{c}_k^T \mathbf{p} = 2\mathbf{p}^H \mathbf{C}\,\mathbf{p}. \tag{10}$$

Our goal consists of the analytical derivation of the two unknown parameters ℓ and u in the least squares sense. For that, since $J_2(\mathbf{B})$ has to be minimized, we have to solve a minor eigenvalue problem. To be well defined, the minor eigenvector, say $\mathbf{e}$, has to be unique (only up to the product by a constant coefficient), *i.e.* the corresponding subspace has to be of dimension 1. Thus the minor eigenvalue has to be of multiplicity 1. Even if it is the case, it is easily shown that the system of equations $\mathbf{e} = \alpha\mathbf{p}$ for the derivation of ℓ and u does not have a solution. An additional problem is that the matrix $\mathbf{C}$ can have zero eigenvalue with multiplicity up to two. We now propose an approximation of $M'^{*}_{k,12}$ in (8) allowing to overcome the two above problems. From (7), we directly have

$$\begin{aligned} M'^{*}_{k,12} &= M_{k,11}\ell + M_{k,12}\ell u^* + M^*_{k,12}(1 + \ell u) \\ &\quad + M_{k,22}u^*(1 + \ell u). \end{aligned} \tag{11}$$

For the approximation, we assume that we are close to a diagonalizing solution. In this case, we have $|M_{k,12}| \ll 1$, $|\ell| \ll 1$ and $|u| \ll 1$. Thus the term $M_{k,12}\ell u^*$ in (11) can be clearly neglected in comparison to the three other terms, leading to

$$M'^{*}_{k,12} \approx \mathbf{c}_{1,k}^T\,\mathbf{p}_1, \tag{12}$$

where

$$\mathbf{p}_1 = \begin{pmatrix} \ell \\ 1+\ell u \\ u^*(1+\ell u) \end{pmatrix} \text{ and } \mathbf{c}_{1,k} = \begin{pmatrix} M_{k,11} \\ M^*_{k,12} \\ M_{k,22} \end{pmatrix}. \tag{13}$$

Finally the criterion (10) is also approximated as

$$J_2(\mathbf{B}) \approx 2\,\mathbf{p}_1^H \sum_{k=1}^{K} \mathbf{c}^*_{1,k}\mathbf{c}^T_{1,k}\,\mathbf{p}_1 = 2\,\mathbf{p}_1^H \mathbf{C}_1\,\mathbf{p}_1. \tag{14}$$

The minimization of $J_2(\mathbf{B})$ is solved by finding the unit-norm minor eigenvector of $\mathbf{C}_1$, denoted by $\mathbf{e}'$. If the minor eigenvalue is of multiplicity 1, we are going to show that the parameters ℓ and u can be determined analytically. We have to solve the following non-linear system of three equations

$$\mathbf{e}' = \begin{pmatrix} e'_1 & e'_2 & e'_3 \end{pmatrix}^T = \beta\,\mathbf{p}_1, \tag{15}$$

with the three unknowns ℓ, u and the additional complex parameter β. After straightforward calculations, we obtain the following analytical solution

$$\begin{cases} u = \left(\dfrac{e'_3}{e'_2}\right)^* \\ \beta = e'_2 - u\,e'_1 \\ \ell = \dfrac{e'_1}{\beta}. \end{cases} \tag{16}$$

It has to be noticed that the above solution happens to be very simple and that the determination of the minor eigenvector can also be done analytically since a 3×3 matrix is involved.

3.2. Balancing phase

In order to improve the robustness of the algorithm, we have to pay attention that the norm of $\mathbf{B}$ does not increase too much within the iterations. This is mainly due to the fact that in (16) the values of ℓ and u result from a division. That is why we propose to normalize $\mathbf{B}$ in using a left-multiplication by a diagonal weighted matrix defined as:

$$\mathbf{W} = \begin{pmatrix} w & 0 \\ 0 & w^{-1} \end{pmatrix}, \tag{17}$$

where w is a real parameter. This parameter is determined by minimizing the norm of $\mathbf{WB}$. Straightforward derivations yield to

$$w = \frac{(|\ell|^2 + |1+\ell u|^2)^{1/4}}{(1+|u|^2)^{1/4}}. \tag{18}$$

The resulting overall algorithm is denoted HCLU.

4. SIMULATION RESULTS

4.1. Algorithm performances

In order to evaluate the algorithm performances, we use the performance index proposed in [6][13][14]. It compares the global matrix $\mathbf{S} = \mathbf{BA} = (S_{ij})$ to the product of a permutation matrix and a diagonal matrix as follow:

$$I(\mathbf{S}) = \frac{1}{2N(N-1)} \left(\sum_{i=1}^{N} \left(\sum_{j=1}^{N} \frac{|S_{ij}|^2}{\max_\ell |S_{i\ell}|^2} - 1 \right) + \sum_{j=1}^{N} \left(\sum_{i=1}^{N} \frac{|S_{ij}|^2}{\max_\ell |S_{\ell j}|^2} - 1 \right) \right). \tag{19}$$

This non-negative index is zero if $\mathbf{S}$ satisfies $\mathbf{B} = \mathbf{DPA}^{-1}$. We consider 25 hermitian matrices of size 15×15 with an additive noise as $\mathbf{M}_k + t\,\mathbf{N}_k$ where $\mathbf{N}_k$ ($k = 1, \ldots, 25$) are hermitian matrices of size 15×15 and t is an real scaling factor.

The real diagonal matrices $\mathbf{D}_k$ follows a zero mean unit variance normal distribution. The mixing matrix $\mathbf{A}$ and the matrices of noise $\mathbf{N}_k$ are also drawn following a zero mean unit variance normal distribution for both their real part and their imaginary part. Finally, we display the mean value of the performance index $I(\mathbf{S})$ (cf fig.1 - fig.4) *w.r.t.* the number of sweeps over one hundred independent draws.

In figure (1), we exhibit the performances of HCLU algorithm with an initial matrix close to a diagonalizing solution. Hence, we initialize $\mathbf{B}$ as the matrix which jointly diagonalizes two matrices of the set $\{\mathbf{M}_k\}$. One can see that in this context, HCLU quickly converges as expected with a good performance. This first simulation confirms the validity of the approximation.

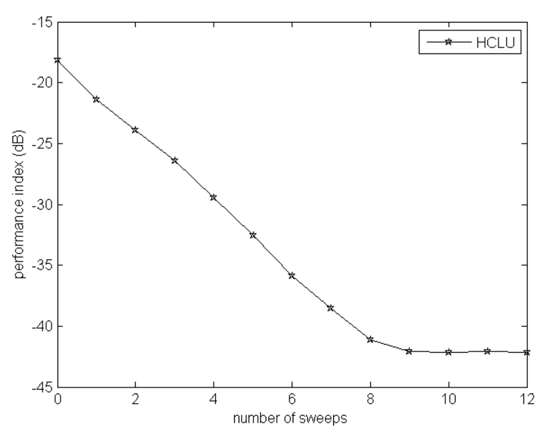

Fig. 1. Algorithms performances for 25 matrices of size 15×15 with a good initialization of $\mathbf{B}$ and with $t = 10^{-2}$.

Now far from a diagonalizing solution, the initial matrix being the identity one, the results are shown in figure (2).

This shows that the HCLU algorithm is really robust when the approximation does not hold and good performances are obtained.

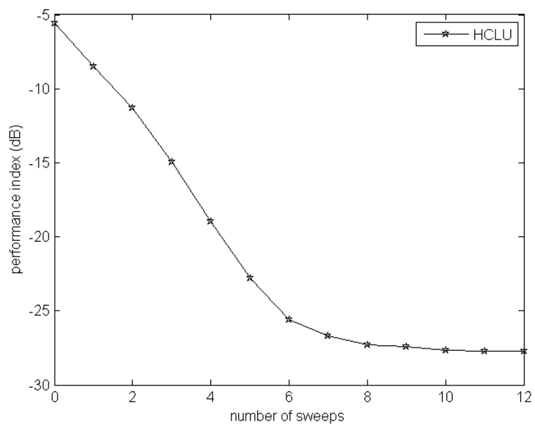

Fig. 2. Algorithms performances for 25 matrices of size 15×15 with $\mathbf{B}$ initialized as the identity matrix and with $t = 10^{-2}$.

4.2. Algorithm comparisons

Now we propose to compare the proposed HCLU algorithm to two Jacobi like algorithms developed in the non-orthogonal case. The first one is the JTJD algorithm suggested in [11] and the second one is the LUJ1D algorithm suggested in [9] in the real case and generalized here to the complex case. In the same simulation context as described in section 4.1, the considered initial matrix being the identity matrix, we compare these three algorithms by first taking $t = 10^{-2}$ (see fig.3) and then by taking $t = 10^{-1}$ (see fig.4).

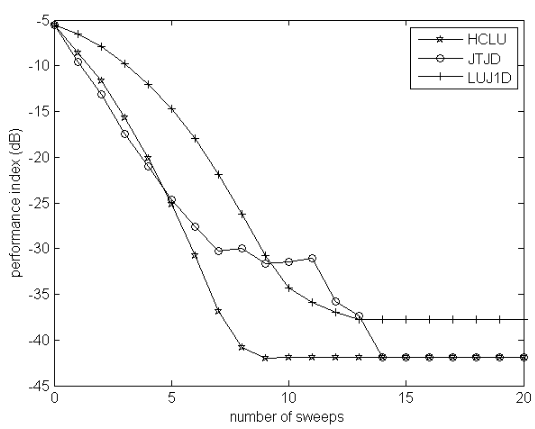

Fig. 3. Performances of the HCLU algorithm compared to the JTJD and LUJ1D algorithms when $t = 10^{-2}$.

The figure 3 shows us that HCLU algorithm outperforms LUJ1D concerning the convergence speed and after convergence LUJ1D is around $4dB$ less accurate than HCLU. Even if the convergence speed of JTJD is better on the very first sweeps, HCLU converges to $-42dB$ in only 9 sweeps whereas JTJD join the same performance level but in 14 sweeps.

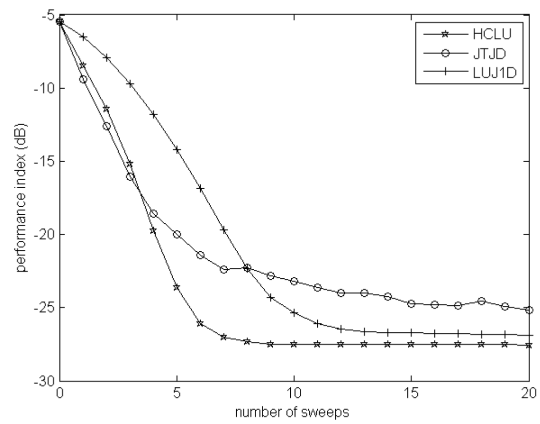

Fig. 4. Performances of the HCLU algorithm compared to the JTJD and LUJ1D algorithms when $t = 10^{-1}$.

In the figure 4, with a higher level of noise, the results are similar. One can notice that after convergence, the HCLU algorithm presents a better performance level ($-27.5dB$) than LUJ1D ($-26.9dB$) and JTJD ($-25.2dB$).

5. CONCLUSION

For the non-orthogonal joint diagonalization of hermitian matrices, we have proposed a new Jacobi like algorithm based on a LU decomposition of the diagonalizing matrix. In order to determine analytically the parameters, we have suggested an approximation of the criterion in the 2×2 case. The computer simulations illustrate the overall good behavior of the HCLU algorithm and show that it compares favorably to the JTJD and LUJ1D algorithms.

6. REFERENCES

[1] A. Yeredor, "Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation", *IEEE Trans. Signal Processing*, Vol. 50, No. 7, pp 1545-1553, July 2002.

[2] P. Tichavsky and A. Yeredor "Fast approximate joint diagonalization incorporating weight matrices", emphIEEE Trans. Signal Processing, Vol. 57, No. 3, pp 878-891, March 2009.

[3] X.-L. Li and X.-D. Zhang, "Nonorthogonal joint diagonalization free of degenerate solution," *IEEE Trans. on Sig. Pro.*, Vol. 55, No. 5, May 2007

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

[5] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso and E. Moulines, "A blind source separation technique using second order statistics", *IEEE Trans. Signal Processing*, vol. 45, no. 2, pp 434-444, February 1997.

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

[7] L. De Lathauwer, B. De Moor and J. Vandewalle, "ICA techniques for more sources than sensors", in *Proc. of the IEEE Workshop on Higher-Order Statistics (HOS'99)*, Caesarea, Israel, pp. 121-124, June 1999.

[8] A. Souloumiac, "Nonorthogonal joint diagonalization by combining Givens and hyperbolic rotations," *IEEE Trans. Signal Processing*, vol. 57, no. 6, pp. 2222-2231, June 2009.

[9] B. Afsari, "Simple LU and QR based non-orthogonal matrix joint diagonalization", in *Proc. ICA'2006, Springer LNCS series*, pp 1-7, 2006.

[10] V. Maurandi, C. De Luigi and E. Moreau, "Fast Jacobi like algorithms for joint diagonalization of complex symmetric matrices", in *Proc. EUSIPCO'2013*, Marrakech, Morocco, September 2013.

[11] X. Guo, S. Zhu, S. Miron and D. Brie, "Approximate joint diagonalization by nonorthogonal nonparametric Jacobi transformations," in *Proc. ICASSP'2010*, Dallas, USA, March 2010.

[12] D. T. Pham, "Joint approximate diagonalization of positive definite hermitian matrices," *SIAM J. Matrix Anal. Appl.*, vol. 22, no. 4, pp. 1136-1152, 2001.

[13] E. Moreau and O. Macchi, "A one stage self-adaptive algorithm for source separation", in Proc. ICASSP, vol. 3, pp. 49-52, April 1994.

[14] E. Moreau and O. Macchi, "High order contrasts for self-adaptive source separation", International Journal of Adaptive Control and Signal Processing, vol. 10, no. 1, pp. 19-46, January 1996.

[15] 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 Process. Letters*, Vol. 14, No. 11, pp. 860-863, November 2007.

[16] T. Trainini and E. Moreau, "Variations around gradient like algorithms for joint diagonalization of Hermitian matrices," in *Proc. EUSIPCO'2012*, Bucharest, Romania, August 2012.