BLIND IDENTIFICATION OF UNDERDETERMINED MIXTURES BASED ON THE HEXACOVARIANCE

Laurent Albera[1,2], Pierre Comon[1], Pascal Chevalier[2], Anne Ferréol[2]

(1) I3S, Algorithmes-Euclide-B, BP 121, F-06903 Sophia-Antipolis, France
(2) THALES Communications, 146 Boulevard de Valmy, BP 82, F-92704 Colombes, France
comon@i3s.unice.fr

ABSTRACT

Static linear mixtures with more sources than sensors are considered. Blind Identification (BI) of underdetermined mixtures is addressed by taking advantage of Sixth Order (SixO) statistics and the Virtual Array (VA) concept. Surprisingly, identification methods solely based on the hexacovariance well succeed, despite their expected high estimation variance; this is due to the inherently good conditioning of the problem. A computationally simple but efficient algorithm, named BIRTH, is proposed and enables to identify the steering vectors of up to $P = N^2 - N + 1$ sources for arrays of N sensors with space diversity only, and up to $P = N^2$ for those with angular and polarization diversities. Five numerical algorithms are compared.

1. INTRODUCTION

Linear mixtures of independent random sources are referred to as *underdetermined* if the number of sources, P, always exceeds the number of sensors, N. In other words, underdetermined mixtures do not enjoy sparsity properties such as disjoint source spectra, or sources non permanently present (property often exploited in Speech applications [1]).

Blind Source separation (BSS) algorithms performing a Second Order (SO) prewhitening in order to identify a unitary separator in a second stage, or extracting sources linearly by deflation, are not able to identify underdetermined mixtures [2] [3] [4] [5]. With our terminology, these techniques address the *overdetermined* case, that is, when the number of sensors is larger than or equal to the number of sources.

Identifiability of linear mixtures received on a single sensor requires source distributions to have an indecomposable characteristic function [6] [7]; for instance in digital communications, BPSK sources are indecomposable but QPSK are not. This condition can be deflated for underdetermined mixtures received on 2 sensors [8]. On the other hand, for overdetermined mixtures, the only pathological distributions are Gaussian [9] [10] [7] [3]. In the sequel, it is assumed that an underdetermined mixture is available on more than one sensor, viz $1 < N < P$.

Blind source extraction from underdetermined mixtures is a difficult problem since these mixtures cannot be linearly inverted [1] [11]. On the other hand, Blind Identification (BI) of the mixture matrix can be performed without extracting the sources (at least in a first stage), as in [12] [13] [14] [1] [15] [16]. More precisely, the methods proposed in [12] [13] [14] [16] only use the data FO statistics, whereas the one proposed in [15] exploits the information contained in the second characteristic function of observations.

However, all these methods have drawbacks in operational contexts. Indeed, the method [12] is still very difficult to implement and does not ensure the BI of the source steering vectors when the sources have the same kurtosis. The BI methods [13] [14] assume FO non-circularity and thus fail in separating FO circular sources. Besides, the theory developed in [13] only confines itself to three sources and two sensors. The method [1] requires sparsity conditions, and thus does not address the general case when all sources are always present. In addition, the method [15] has been developed only for real mixtures of real-valued sources, and the issue of robustness with respect to an over estimation of the number of sources remains open. Eventually, although the FOBIUM algorithm [16] performs BI of up to $P = N^2$ sources for arrays of N different sensors, it requires sources with different FO spectral densities.

In this paper, five algorithms using solely Sixth Order (SixO) statistics are compared; two are entirely new, while the others have been already presented in [17] [18] under the name of BIRTH (Blind Identification of mixtures of sources using Redundancies in the daTa Hexacovariance matrix), but not all implemented however. It is assumed throughout that sources have nonzero marginal 6th order cumulants with the same sign (verified in most radio communications contexts). We have also recently shown [18] that the principles theoretically extend to statistics of arbitrary order $2q$, $q > 2$, but computer simulations still remain to be completed for $q > 3$.

2. ASSUMPTIONS AND NOTATION

Assume that for any fixed time index k, N complex outputs $x_n(k)$ of a noisy mixture of P statistically independent sources $s_p(k)$ are available. The vector $\boldsymbol{x}(k)$ of the mea-

sured array outputs is modeled by

$$\boldsymbol{x}(k) = \boldsymbol{A}\,\boldsymbol{s}(k) + \boldsymbol{\nu}(k) \tag{1}$$

where $\boldsymbol{A}$, $\boldsymbol{s}(k)$, $\boldsymbol{\nu}(k)$ are the $N \times P$ constant mixing matrix, and the source and noise random vectors, respectively. In addition, for any fixed time index k, $\boldsymbol{s}(k)$ and $\boldsymbol{\nu}(k)$ are statistically independent. Vectors and matrices are denoted in bold.

For the sake of convenience we need to define, for any k, the entries of the SixO cumulant tensor $\boldsymbol{C_x}$ of a random vector, $\boldsymbol{x}(k)$, stationary and ergodic up to order 6:

$$C_{i_1,i_2,i_3,\boldsymbol{x}}^{i_4,i_5,i_6} = \mathsf{Cum}\{x_{i_1}(k), x_{i_2}(k), x_{i_3}(k), x_{i_4}(k)^*, x_{i_5}(k)^*, x_{i_6}(k)^*\} \tag{2}$$

Such components may be ordered in the $\boldsymbol{H_x}$ hexacovariance matrix as follows:

$$H_{\boldsymbol{x}}(j_1, j_2) = C_{i_1, i_2, i_3, \boldsymbol{x}}^{i_4, i_5, i_6} \tag{3}$$

$$j_1 = i_6 + N(i_2 + N(i_1 - 1)), \quad j_2 = i_3 + N(i_5 + N(i_4 - 1)). \tag{4}$$

Expressions of SixO cumulants, in the complex case, as a function of moments, can be found in [18, appendix D] for zero-mean variables which are distributed symmetrically with respect to the origin. Note that the impact of the chosen way to arrange them in the hexacovariance matrix is analysed in [18, section 5]. More generally, it is shown in [19] that there exists an optimal matrix arrangement of the SixO cumulants with respect to the maximal number of statistically independent sources to be processed by a method exploiting the algebraic structure of the hexacovariance. Especially for the BIRTH method, this optimal arrangement is shown [18, section 5] to correspond to the one described by (3). Now, we further assume the following hypotheses:

- **A1**. Vector $\boldsymbol{s}(k)$ is stationary, ergodic (or *cyclostationary* and *cycloergodic*, respectively), with components a priori in the complex field and mutually uncorrelated at order 6 (the *cyclostationarity* case has been addressed in [18, section 3.3]);

- **A2**. Noise vector $\boldsymbol{\nu}(k)$ is stationary, ergodic and Gaussian with components a priori in the complex field too;

- **A3**. SixO marginal source cumulants, $C_{p,p,p,\boldsymbol{s}}^{p,p,p}$, are not null and have all the same sign;

- **A4**. Column vectors $\boldsymbol{a}_p$ of $\boldsymbol{A}$, also called steering vectors, are not collinear and have not any null component;

- **A5**. The $N^2 \times P$ matrix $\boldsymbol{\mathcal{A}}_2 = \boldsymbol{A} \oslash \boldsymbol{A}^*$, where $\oslash$ denotes the Khatri-Rao product [20] [18, section 3.2.1], is full column rank (this implies that $P \leq N^2$);

The goal of BI consists of determining an estimate of the mixture matrix $\boldsymbol{A}$.

3. ALGEBRAIC PRELIMINARY

Consider the following mathematical problem:

Problem: *Given M matrices $\boldsymbol{\Xi}_m$, $1 \leq m \leq M$, each of size $N \times P$, find a $N \times P$ matrix $\boldsymbol{A}$, and $P \times P$ diagonal matrices $\boldsymbol{D}_m$ such that $\|\boldsymbol{D}_m\| = 1$ and*

$$\boldsymbol{\Xi}_m \boldsymbol{D}_m \approx \boldsymbol{A} \tag{5}$$

Matrices $\boldsymbol{A}$ and $\boldsymbol{D}_m$ can be obtained as stationary values of the Least Squares (LS) criterion below:

$$\varepsilon = \sum_{m=1}^{M} \|\boldsymbol{\Xi}_m \boldsymbol{D}_m - \boldsymbol{A}\|_F^2 \tag{6}$$

where $\|\boldsymbol{B}\|_F$ is the Frobenius norm of matrix $\boldsymbol{B}$. As a consequence, they satisfy the following system of equations, obtained by cancelling the gradient of ε with respect to $\boldsymbol{D}_m$ and $\boldsymbol{A}$:

$$\begin{cases} \forall m, \ \forall p, \ \{\boldsymbol{\Xi}_m^{\mathsf{H}}(\boldsymbol{\Xi}_m \boldsymbol{D}_m - \boldsymbol{A})\}(p,p) = 0 \\ \forall\, (n,p), \ \sum_{m=1}^{M} \{\boldsymbol{\Xi}_m \boldsymbol{D}_m - \boldsymbol{A}\}(n,p) = 0 \end{cases} \tag{7}$$

where $\boldsymbol{B}(n,p)$ is the (n,p)-th component of matrix $\boldsymbol{B}$. It is then not hard to obtain the closed form expression for $\boldsymbol{A}$:

$$\boldsymbol{A} = \frac{1}{M} \sum_{m=1}^{M} \boldsymbol{\Xi}_m \boldsymbol{D}_m \tag{8}$$

By plugging back this solution in system (7), one gets after some manipulations:

$$\forall p, 1 \leq p \leq P, \qquad \boldsymbol{F}_p\, \boldsymbol{d}_p = 0 \tag{9}$$

where

$$\boldsymbol{F}_p(m_1, m_2) = \begin{cases} (M-1)\{\boldsymbol{\Xi}_{m_1}^{\mathsf{H}} \boldsymbol{\Xi}_{m_1}\}(p,p) \ \text{if } m_1 = m_2 \\ -\{\boldsymbol{\Xi}_{m_1}^{\mathsf{H}} \boldsymbol{\Xi}_{m_2}\}(p,p) \ \text{otherwise} \end{cases} \tag{10}$$

and where $\boldsymbol{d}_p = \begin{bmatrix} \boldsymbol{D}_1(p,p) & \boldsymbol{D}_2(p,p) & \cdots & \boldsymbol{D}_M(p,p) \end{bmatrix}^{\mathsf{T}}$. In other words, the solution to the LS problem under the constraint that, for any fixed index p, $\sum_m |\boldsymbol{D}_m(p,p)|^2 = 1$ is obtained when the vector $\boldsymbol{d}_p$ is the right singular vector of matrix $\boldsymbol{F}_p$ associated with the minimal singular value. Once every entry $\boldsymbol{D}_m(p,p)$ is obtained, matrix $\boldsymbol{A}$ can be calculated thanks to (8). This solution is thus not iterative (though we could possibly run alternate iterations).

4. APPLICATION TO BLIND IDENTIFICATION OF UNDERDETERMINED MIXTURES

Under assumptions (**A1**)-(**A5**), the BIRTH algorithm succeeds [17, section 3.3] [18, section 4.2.1] in identifying matrix $\boldsymbol{\mathcal{A}}_3$, defined by the triple columnwise product

$$\boldsymbol{\mathcal{A}}_3 = \boldsymbol{A} \oslash \boldsymbol{A} \oslash \boldsymbol{A}^* = [\boldsymbol{a}_1 \otimes \boldsymbol{a}_1 \otimes \boldsymbol{a}_1^* \ \cdots \ \boldsymbol{a}_P \otimes \boldsymbol{a}_P \otimes \boldsymbol{a}_P^*] \tag{11}$$

up to a multiplicative trivial matrix $\mathcal{T}$ (a trivial matrix is of the form $\Lambda\Pi$ where Λ is invertible diagonal and Π a permutation), where $\otimes$ denotes the Kronecker product. The five methods described below allow to estimate matrix $\boldsymbol{A}$ from the estimate $\mathcal{A}_3$ output by the core of the BIRTH algorithm. We do not describe this core step here for reasons of space. Anyway, our contribution lies mainly in methods 4 and 5, and the comparisons, not in the core itself.

Three methods have been proposed in [18, section 4.2.2] in order to extract mixture $\boldsymbol{A}$ from the estimate $\widehat{\mathcal{A}}_3$ of matrix $\mathcal{A}_3$:

Method 1: *Take as estimate the matrix block made up of the N first rows of the conjugate of matrix $\widehat{\mathcal{A}}_3$.*

Method 2: *Define the N matrix blocks $\boldsymbol{\Xi}_n$, of size $N \times P$, made up of the successive rows of the conjugate of matrix $\widehat{\mathcal{A}}_3$; take as estimate the average of these N blocks.*

Method 3: *Fully exploit each column vector $\widehat{\boldsymbol{b}}_p$ of $\widehat{\mathcal{A}}_3$. In order to do this, first extract, from vector $\widehat{\boldsymbol{b}}_p$, the N vectors $\widehat{\boldsymbol{b}}_p(n)$ of size $N^2 \times 1$, then remodel them into N matrices $\widehat{\boldsymbol{B}}_p(n)$ of size $N \times N$, and finally build the matrix whose p-th column vector is the eigenvector (approximately) in common within the N matrices $\widehat{\boldsymbol{B}}_p(n)^*$ ($1 \leq n \leq N$) and associated with the largest eigenvalue using the Joint Approximate Diagonalization (JAD) algorithm described in [21].*

Methods 1 and 2 ensue immediately from the structure of matrix $\widehat{\mathcal{A}}_3$. In fact, it can be shown [18, equation 34] that matrix $\widehat{\mathcal{A}}_3$ may be written as:

$$\widehat{\mathcal{A}}_3 = \mathcal{A}_3 \, \mathcal{T} = [[\boldsymbol{A}^*\mathcal{T}_1]^\mathsf{T} \, [\boldsymbol{A}^*\mathcal{T}_2]^\mathsf{T} \cdots [\boldsymbol{A}^*\mathcal{T}_{N^2}]^\mathsf{T}]^\mathsf{T} \quad (12)$$

where the N^2 matrices $\mathcal{T}_n$ ($1 \leq n \leq N^2$) of size $P \times P$ are trivial. As for method 3, it is shown in [18, appendix C] that

$$\forall n, \; 1 \leq n \leq N, \quad \widehat{\boldsymbol{b}}_p(n) \propto \left[\boldsymbol{a}_{\xi(p)} \otimes \boldsymbol{a}^*_{\xi(p)} \right] \quad (13)$$

where $\xi(\cdot)$ is a bijective function of $\{1, 2, \ldots, P\}$ into itself (i.e. a permutation function). Then it is straightforward to show that

$$\forall n, \; 1 \leq n \leq N, \quad \widehat{\boldsymbol{B}}_p(n) \propto \left[\boldsymbol{a}_{\xi(p)} \, \boldsymbol{a}_{\xi(p)}^{\mathsf{H}} \right]^* \quad (14)$$

and hence the method 3 result. Note that, although the JAD algorithm [21] is restricted to unitary joint diagonalizers, it can be used in method 3 since matrices $\widehat{\boldsymbol{B}}_p(n)^*$ are of rank 1, from (14).

Method 4: *The fourth method we consider performs an unrestricted (non-unitary) LS joint diagonalization scheme, as for instance the one described by Yeredor in [20], yielding a better LS fit.*

Method 5: *Follow the approach given in section 3 in order to extract mixture $\widehat{\boldsymbol{A}}$ from $\widehat{\mathcal{A}}_3$, with matrices $\boldsymbol{\Xi}_n$ defined as in Method 2.*

The latter algorithm does not take into account the fact that diagonal matrices $\boldsymbol{D}_m$ should contain products of entries of $\boldsymbol{A}$, and is therefore expected to yield less accurate results. However, subsequent simulations demonstrate that the loss in performance is little compared to the gain in computational complexity.

5. COMPUTER RESULTS

We proceed in this section to two types of simulations. First, in order to test the five BI methods previously described independently of the BIRTH algorithm, we have generated P vectors $\widehat{\boldsymbol{b}}_p$ such that

$$\widehat{\boldsymbol{b}}_p = \boldsymbol{b}_p + \boldsymbol{\nu}_p \quad (15)$$

where $\boldsymbol{b}_p$ is the p-th column vector of matrix $\mathcal{A}_3$ and where the $N^3 \times 1$ noise random vectors $\boldsymbol{\nu}_p$ are chosen to be Gaussian spatially and temporally white such that their covariance matrices $\boldsymbol{R}_{\boldsymbol{\nu}_p}$ verify $\boldsymbol{R}_{\boldsymbol{\nu}_p} = \sigma^2 \, \boldsymbol{I}_{N^3}$. We took a uniformly spaced circular array of $N = 3$ identical sensors, of radius R such that $R/\lambda = 0.55$, and $P = 12$ directional vectors. The chosen BI performance criterion is as follows.
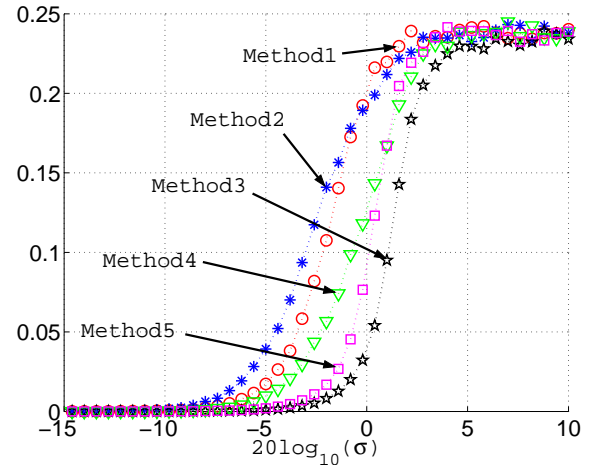

Fig. 1. Mean of the $P = 12$ gaps α_p

As in [17], for each estimated column $\hat{\boldsymbol{a}}_p$, one computes the gap α_p:

$$\alpha_p = \min_{1 \leq i \leq P} d(\hat{\boldsymbol{a}}_p, \boldsymbol{a}_i)$$

where $d(\cdot, \, \cdot)$ denotes the pseudo-distance $d(\boldsymbol{u}, \, \boldsymbol{v}) = 1 - |\langle \boldsymbol{u}, \boldsymbol{v} \rangle|^2 / ||\boldsymbol{u}||^2 ||\boldsymbol{v}||^2$. We report the average of the P gaps obtained by the five methods in figure 1, as a function of the noise level. It can be seen that method 5 is almost as good as the most complex one, namely method 3.

Second, we now incorporate the BIRTH core step in the comparison. Sources are BPSK modulated, with a raised cosine pulse shape of roll-off equal to 0.25, and assumed synchronized. Figure 2 shows BI results obtained when 7 BPSK sources are received by the same array as above.
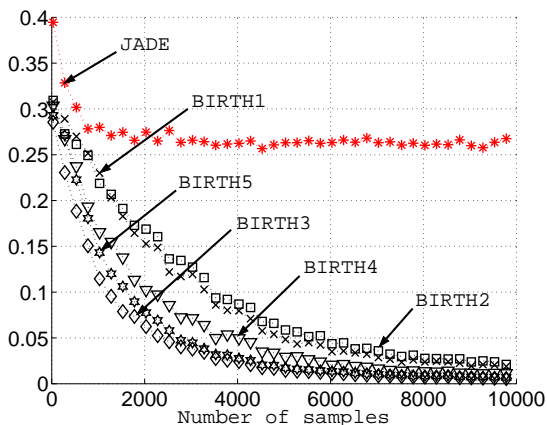
Fig. 2. Mean of the $P=7$ gaps α_p

Their symbol periods are equal to twice the sample period and their carrier residuals are all null. In this figure, the label "BIRTHm" corresponds to the BIRTH algorithm followed by method m of section 4. Again, it can be seen that the five methods can be sorted in the same way: method 3, the most complex, is followed by method 5. The latter thus appears to exhibit the best trade-off between performance and computational complexity.

6. CONCLUSION

As surveyed in introduction, there are few algorithms able to identify blindly underdetermined mixtures (*i.e.* in the absence of sparsity). The algorithm BIRTH3 has been recently proposed by the authors, but had not been implemented nor tested yet. Next, the alternate estimation algorithm of Yeredor has also been implemented (BIRTH4). Last, we have devised a simpler algorithm, BIRTH5, by relaxing the structure of the problem, and compared its performances with the former methods; according to the computer results reported above, BIRTH5 appears to be the most attractive.

These algorithms, and in particular BIRTH3 and BIRTH5, can be used for blind beamforming. Yet, there exist techniques based on the array manifold knowledge that can handle underdetermined mixtures, such as the so-called $4-$MUSIC. It could be interesting to compare its performances with the above as well, which could yield a performance bound. On the other hand, identifiability issues remain to be addressed for the most general cases of underdetermined mixtures.

7. REFERENCES

[1] T. W. LEE, M. S. LEWICKI, M. GIROLAMI, and T. J. SEJNOWSKI, "Blind source separation of more sources than mixtures using overcomplete representations," *IEEE Signal Processing Letters*, vol. 4, no. 4, pp. 87–90, 1999.

[2] J.-F. CARDOSO and A. SOULOUMIAC, "Blind beamforming for non-gaussian signals," *IEE Proceedings-F*, vol. 140, no. 6, pp. 362–370, December 1993.

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

[4] E. BINGHAM and H. HYVARINEN, "A fast fixed-point algorithm for independent component analysis of complex valued signals," *International Journal of Neural Systems*, vol. 10, no. 1, pp. 1–8, 2000.

[5] A. BELOUCHRANI, K. ABED-MERAIM, J.-F. CARDOSO, and E. MOULINES, "A blind source separation technique using second-order statistics," *IEEE Trans. Sig. Proc.*, vol. 45, no. 2, pp. 434–444, Feb. 1997.

[6] E. LUKACS, *Characteristic functions*, Griffin, 1970.

[7] A. M. KAGAN, Y. V. LINNIK, and C.R. RAO, *Characterization Problems in Mathematical Statistics*, Wiley, 1973.

[8] A. TALEB and C. JUTTEN, "On underdetermined source separation," in *ICASSP99*, Phoenix, Arizona, March, 1999.

[9] W. FELLER, *An Introduction to Probability Theory and its Applications*, vol. II, Wiley, 1966.

[10] G. DARMOIS, "Analyse générale des liaisons stochastiques," *Rev. Inst. Internat. Stoch.*, vol. 21, pp. 2–8, 1953.

[11] P. COMON and O. GRELLIER, "Non-linear inversion of underdetermined mixtures," in *ICA 99, Second Internat. Symp. Indep. Comp. Ana. Blind Sig. Sep.*, Aussois, France, January 11-15 1999, pp. 461–465.

[12] J.-F. CARDOSO, "Super-symetric decomposition of the fourth-order cumulant tensor. Blind identification of more sources than sensors," in *ICASSP'91*, Toronto, Canada, May 1991, pp. 3109–3112.

[13] P. COMON, "Blind channel identification and extraction of more sources than sensors," in *SPIE Conference*, San Diego, US, July 19-24 1998, pp. 2–13.

[14] L. De LATHAUWER, P. COMON, B. De MOOR, and J. VANDEWALLE, "ICA algorithms for 3 sources and 2 sensors," in *HOS 99, IEEE Sig. Proc. Workshop on Higher-Order Stat.*, Caesarea, Israel, June 14-16 1999, pp. 116–120.

[15] A. TALEB, "An algorithm for the blind identification of N independent signal with 2 sensors," in *ISSPA 01, 16th Symp. Sig. Proc. Appl.*, Kuala-Lumpur, Malaysia, August 13-16 2001, pp. 5–8.

[16] A. FERREOL, L. ALBERA, and P. CHEVALIER, "Fourth order blind identification of underdetermined mixtures of sources (FOBIUM)," in *ICASSP 03*, Hong Kong, China, April 6-10 2003, pp. 41–44.

[17] L. ALBERA, A. FERREOL, P. COMON, and P. CHEVALIER, "Sixth order blind identification of underdetermined mixtures (BIRTH) of sources," in *ICA 03, 4th Int. Symp. Indep. Comp. Anal. Blind Sig. Sep.*, Nara, Japan, April 1-4 2003, pp. 909–914.

[18] L. ALBERA, A. FERREOL, P. COMON, and P. CHEVALIER, "Blind Identification of Overdetermined and underdetermined Mixtures of sources (BIOME)," *Linear Algebra Applications*, april 2003, submitted.

[19] P. CHEVALIER, L. ALBERA, A. FERREOL, and P. COMON, "On the virtual array concept for higher order array processing," *IEEE Trans. Sig. Proc.*, 2003, submitted.

[20] A. YEREDOR, "Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation," *IEEE Transactions On Signal Processing*, vol. 50, no. 7, pp. 1545–1553, July 2002.

[21] J.-F. CARDOSO and A. SOULOUMIAC, "Jacobi angles for simultaneous diagonalization," *SIAM Journal Matrix Analysis and Applications*, vol. 17, no. 1, pp. 161–164, 1996.