Nonunitary Joint Diagonalization for Overdetermined Convolutive Blind Signal Separation

Wei-Tao Zhang and Jin-Ling Sun

School of Electronic Engineering, Xidian University

Xi'an 710071, China

Email: zhwt-work@foxmail.com

Abstract—As is known that nonunitary joint diagonalization (JD) has some advantages over the unitary one in terms of system identification accuracy. However, the existing nonunitary JD algorithms are prone to converge to degenerate (even singular) solutions, which result in deteriorated identification performance. Moreover, the existing algorithms usually seek a square diagonalizing matrix, which greatly limits their application in overdetermined system identification scenario. In order to overcome these drawbacks, we reformulate the nonunitary JD as a multicriteria optimization model. The resulting algorithm can converges to a nonsquare well-conditioned diagonalizing matrix.

Index Terms—Joint diagonalization (JD), degenerate solution, convolutive blind source separation (CBSS).

I. INTRODUCTION

In recent years, the joint diagonalization of a set of matrices or higher order tensor has received more and more research interest in a variety of fields [1], such as multiple source localization, multiuser MIMO communication, blind beamforming and blind source separation.

A number of state-of-the-art algorithms were proposed to solve JD problem. The first considered method seeks a unitary diagonalizer, Cardoso used successive Givens rotations [2], while our previous work [3] used the Householder transform. In BSS context, the unitary JD is applied to identify the demixing system after a whitening phase on the raw observations. Whereas Yeredor pointed out that the whitening phase practically distorts the LS criterion, yielding degraded separation performance [4]. To avoid the whitening phase, a number of nonunitary JD algorithms were developed in the literature [5-13]. By minimizing a maximum likelihood criterion, Pham proposed a computationally efficient algorithm for JD [5]. However it can only be applied to positive definite target matrices, which greatly limits its usage in BSS context. The QDIAG algorithm, which minimizes a weighted least squares (WLS) measure of diagonality, was proposed in [6]. By adding a simple constraint on diagonalizer, QDIAG can avoid the trivial solutions for JD. However, it is prone to converge to some undesired degenerate (even singular) solutions, which results in incomplete separation of sources in BSS. To avoid degenerate solutions in nonunitary JD,

the FAJD algorithm minimizes a regularization term based WLS criterion [7], whereas FAJD requires a squared demixing matrix, which limits its use in more generally overdetermined BSS scenario. Besides, the other state-of-the-art algorithms [8-12] for nonunitary JD use a multiplicative update rule to minimize the WLS criterion, the J-Di uses the Givens and Hyperbolic rotations for real JD [8], and its complex version is proposed in [9] by using Shear and Givens rotations. The FFDIAG is proposed in [10]. The trigonometric parameterization method is proposed in [11], which can be regarded as a class of Jacobi-like algorithms for nonunitary JD [12]. Although these multiplicative update algorithms are able to avoid the degenerate solutions by using some proper constraint mechanism, they lead to a squared demixing matrix as usual, and hence limited applications.

In this paper, we focus our attention on the well-conditioned solution and nonsquare demixing matrix to nonunitary JD. To this end, we reformulate the nonunitary JD as a multicriteria optimization problem. The resulting algorithm can be applied in overdetermined mixing scenario, and successfully eliminates the degenerate solutions.

II. PROBLEM STATEMENT

Consider a set $\mathcal{R}$ of Q $M \times M$ complex matrices $\mathbf{R}_q$, built as

$$\mathbf{R}_q = \mathbf{A}\mathbf{D}_q\mathbf{A}^H, \quad q = 1, \ldots, Q \tag{1}$$

where $\mathbf{A}$ is a $M \times N$ ($M \geq N$) mixing matrix with full column rank, $\mathbf{D}_q$ is a $N \times N$ diagonal matrix. The goal of nonunitary JD is to seek a nonsingular diagonalizer $\mathbf{B}$ of dimension $M \times N$ such that the congruent transform $\mathbf{B}^H\mathbf{R}_q\mathbf{B}$, $q = 1, \ldots, Q$ are all diagonal matrices.

JD is actually a blind identification problem since $\mathbf{A}$ and $\mathbf{D}_q$, $q = 1, \ldots, Q$ are all supposed unknown. So the diagonalizer $\mathbf{B}$ can only be estimated up to the scaling and permutation of its columns, i.e., $\hat{\mathbf{B}} = \mathbf{B}\mathbf{\Sigma}\mathbf{\Pi}$ with $\mathbf{\Pi}$ and $\mathbf{\Sigma}$ denote, respectively, a permutation matrix and a nonsingular diagonal matrix. This corresponds to the arbitrary attenuation and order of restored source signals in BSS context.

A straightforward criterion for nonunitary JD is the WLS

measure of diagonality

$$\mathcal{J}(\mathbf{B}) = \sum_{q=1}^{Q} \left\| Off(\mathbf{B}^H \mathbf{R}_q \mathbf{B}) \right\|_F^2 \tag{2}$$

where the operator $Off(\cdot)$ forms a zero diagonal matrix by replacing the diagonal entries of its argument with zeros, and $\|\cdot\|_F$ denotes the Frobenius norm. Direct minimization of cost function (2) generally results in degenerate solutions [7]. In order to avoid degenerate solutions, the FAJD algorithm minimizes the following cost function

$$\mathcal{J}'(\mathbf{B}) = \mathcal{J}(\mathbf{B}) - \log|\det(\mathbf{B})| \tag{3}$$

where $\det(\cdot)$ denotes the matrix determinant. Though eliminating the degenerate solutions, FAJD is limited to square mixing case with $M = N$, and the numerical problem will arise when it is applied in exact diagonalizable data set.

III. MULTICRITERIA OPTIMIZATION FOR JD

A. The Criteria

Instead of eliminating degenerate solutions, we focus on a well-conditioned solution to nonunitary JD, which firstly requires that the columns of $\mathbf{B}$ have uniform norms. Without losing generality, we use the unit norm constraint, i.e., $\mathbf{b}_n^H \mathbf{b}_n = 1$, $n = 1, \ldots, N$, where $\mathbf{b}_n$ denotes the nth column vector of $\mathbf{B}$. Further note that

$$\kappa(\mathbf{B}) \le \kappa(\mathbf{B}^H \mathbf{B}) < \frac{2}{\det(\mathbf{B}^H \mathbf{B})} \left(\frac{\|\mathbf{B}^H \mathbf{B}\|_F^2}{N} \right)^{\frac{N}{2}} \tag{4}$$

where $\kappa(\cdot)$ is the matrix condition number. It is well known that a well conditioned solution is the solution with small condition number. So we propose to consider the following multicriteria model

$$\min_{\mathbf{B}} \mathcal{J}(\mathbf{B}), \quad \max_{\mathbf{B}} \det(\mathbf{B}^H \mathbf{B}),$$
$$s.t. \quad \mathbf{b}_n^H \mathbf{b}_n = 1, \quad n = 1, \ldots, N \tag{5}$$

We see that the diagonalizer $\mathbf{B}$ can be nonsquare in model (5), the minimization of $\mathcal{J}(\mathbf{B})$ guarantees that $\mathbf{B}$ is the diagonalizer of set $\mathcal{R}$, while the maximization of $\det(\mathbf{B}^H \mathbf{B})$ under the unit norm constraint corresponds to the minimization of the upper bound of $\kappa(\mathbf{B})$, which leads to a well-conditioned $\mathbf{B}$. Besides we see that $\mathcal{J}(\mathbf{B})$ is lower bounded by 0, and for $\det(\mathbf{B}^H \mathbf{B})$, by using the Hadamard inequality we have

$$\det(\mathbf{B}^H \mathbf{B}) \le \prod_{n=1}^{N} \tilde{b}_{nn} = 1 \tag{6}$$

where $\tilde{b}_{nn}$ denotes the diagonal entries of $\mathbf{B}^H \mathbf{B}$, one sees $\tilde{b}_{nn} = 1$, $n = 1, \ldots, N$ under the unit norm constraint. The inequality (6) shows that the second criterion in (5) is upper bounded by 1.

B. Optimization Algorithm

Direct optimization of (5) with respect to (w.r.t.) $\mathbf{B}$ is cumbersome, we here divide the overall optimization into N sub-optimization problems, in each sub problem we optimize (5) w.r.t. a selected column of $\mathbf{B}$, say $\mathbf{b}_n$, with the other columns known and fixed. A sweep consists of solving N sub problems, convergence is achieved after several sweeps.

Now we rewrite (5) as the function of $\mathbf{b}_n$, we have

$$\mathcal{J}(\mathbf{b}_n) = tr(\mathbf{b}_n^H \mathbf{Q}_n \mathbf{b}_n) \tag{7}$$

where $tr(\cdot)$ denotes the trace of a matrix, and

$$\mathbf{Q}_n = \sum_{q=1}^{Q} \mathbf{R}_q \mathbf{B}_n \mathbf{B}_n^H \mathbf{R}_q^H + \mathbf{R}_q^H \mathbf{B}_n \mathbf{B}_n^H \mathbf{R}_q \tag{8}$$

with $\mathbf{B}_n$ arises from the deletion of $\mathbf{b}_n$ from $\mathbf{B}$. Let $\bar{\mathbf{B}} = [\mathbf{b}_n, \mathbf{B}_n]$ be the column exchanged matrix of $\mathbf{B}$, then for $\det(\mathbf{B}^H \mathbf{B})$, we have

$$\det(\mathbf{B}^H \mathbf{B}) = \det(\bar{\mathbf{B}}^H \bar{\mathbf{B}})$$
$$= \det \begin{bmatrix} \mathbf{b}_n^H \mathbf{b}_n & \mathbf{b}_n^H \mathbf{B}_n \\ \mathbf{B}_n^H \mathbf{b}_n & \mathbf{B}_n^H \mathbf{B}_n \end{bmatrix}$$
$$= \det(\mathbf{B}_n^H \mathbf{B}_n) \mathbf{b}_n^H \mathbf{P}_n^\perp \mathbf{b}_n \tag{9}$$

where

$$\mathbf{P}_n^\perp = \mathbf{I} - \mathbf{B}_n [\mathbf{B}_n^H \mathbf{B}_n]^{-1} \mathbf{B}_n^H \tag{10}$$

with $\mathbf{I}$ denoting the identity matrix. Since $\mathbf{B}_n$ is assumed known and independent of $\mathbf{b}_n$, then the nth sub problem in one sweep reads

$$\min_{\mathbf{b}_n} \mathbf{b}_n^H \mathbf{Q}_n \mathbf{b}_n, \quad \max_{\mathbf{b}_n} \mathbf{b}_n^H \mathbf{P}_n^\perp \mathbf{b}_n,$$
$$s.t. \quad \mathbf{b}_n^H \mathbf{b}_n = 1, \quad n \in \{1, \ldots, N\} \tag{11}$$

For exact diagonalizable data set $\mathcal{R}$, the solution to (11) should consider two different stages. In the initial stage, $\mathbf{B}$ is far from optimal, then matrix $\mathbf{Q}_n$ is invertible, we can take $\mathbf{b}_n$ as the unit norm generalized eigenvector of matrix pencil $(\mathbf{P}_n^\perp, \mathbf{Q}_n)$ associated with the largest eigenvalue. In the second stage, the algorithm approaches convergence, $\mathbf{B}$ is nearly optimum, then $\mathbf{Q}_n$ will be singular with rank $N - 1$, then the above procedure is numerically unstable. In this stage, let $\mathbf{U}_0$ be the eigenvectors of $\mathbf{Q}_n$ associated with the $M - N + 1$ zero eigenvalues (the $M - N + 1$ smallest eigenvalues in practice), we take

$$\mathbf{b}_n = \mathbf{U}_0 \mathbf{w} \tag{12}$$

where vector $\mathbf{w}$ consists of $M - N + 1$ weight coefficients. Substituting (12) into $\mathbf{b}_n^H \mathbf{P}_n^\perp \mathbf{b}_n$ leads to

$$\mathbf{w}_{opt} = \arg\max_{\mathbf{w}} \mathbf{w}^H \mathbf{U}_0^H \mathbf{P}_n^\perp \mathbf{U}_0 \mathbf{w} \tag{13}$$

The optimal solution to $\mathbf{w}$ is obtained by taking the eigenvector of $\mathbf{U}_0^H \mathbf{P}_n^\perp \mathbf{U}_0$ associated with the largest eigenvalue. One sees that the proposed algorithm (termed as JD-NS) can be applied in more general overdetermined BSS scenario. Moreover, the initialization of the JD-NS is quite simple, the initial value of $\mathbf{B}$ can be identity matrix or randomly generated matrix.

IV. APPLICATION IN CBSS

The objective of CBSS is to recover the multiple unknown sources from their convolved measurements in a reverberant environment, the mixture model reads

$$x_m(t) = \sum_{n=1}^{N} \sum_{l=0}^{L} a_{mn}(l) s_n(t-l) + v_m(t), \quad m = 1, \ldots, M$$
(14)

where $x_m(t)$ denotes the signal observed at the m-th receiver, $s_n(t)$ denotes the n-th source signal, the FIR filter $a_{mn}(l)$ represents the impulse responses of the convolutive channel between the n-th source signal and the m-th receiver, $v_m(t)$ denotes the additive white Gaussian noise at the m-th receiver, M, N, L denote the number of observations, the number of source signals, and the length of the FIR filter. We usually assume $M \geq N$ for an overdetermined scenario.

CBSS can be achieved by either joint block diagonalization (JBD) of correlation matrices of time domain transformed measurements [14]-[16] or JD of covariance matrices of frequency domain measurements [17]. Since the transformed signal model for CBSS in time domain usually results in a high dimensional JBD problem, which is computationally demanding (even prohibitive) in a higher order convolved mixtures (corresponds to a severe reverberant scenario). Here we devote to frequency domain CBSS.

Let $\mathbf{s}(t) = [s_1(t), \ldots, s_N(t)]^T$, $\mathbf{x}(t) = [x_1(t), \ldots, x_M(t)]^T$, $\mathbf{v}(t) = [v_1(t), \ldots, v_M(t)]^T$ denote the source signal, the received signal and the noise vectors respectively, by using the Discrete Fourier Transform (DFT), the mixture model (14) can be written in frequency domain as

$$\mathbf{x}(\omega) = \mathbf{A}(\omega)\mathbf{s}(\omega) + \mathbf{v}(\omega), \quad \omega \in [0, \pi]$$
(15)

where $\mathbf{A}(\omega)$ is the impulse response matrix in frequency domain. The goal of CBSS can be stated in frequency domain as estimating the de-mixing filters $\mathbf{W}(\omega)$ from the observed signals such that

$$\mathbf{W}(\omega)\mathbf{A}(\omega) = \mathbf{\Pi}\mathbf{\Sigma}(\omega), \quad \forall \omega \in [\mathbf{0}, \pi]$$
(16)

where $\mathbf{\Pi}$ is a frequency independent permutation matrix, and $\mathbf{\Sigma}(\omega)$ is frequency dependent diagonal matrix.

Assume that the source signals are piecewise stationary (e.g., speech signals) and independent of each other, then the cross-spectral density matrix of the observed signals at frequency $\omega_k(k = 0, \ldots, K-1)$ and time epoch $q(q = 1, \ldots, Q)$ can be written as

$$\mathbf{R_x}(\omega_k, q) = \mathbf{A}(\omega_k)\mathbf{R_s}(\omega_k, q)\mathbf{A}^H(\omega_k) + \sigma^2\mathbf{I}$$
(17)

where σ^2 denotes the noise variance. Note that $\mathbf{R_s}(\omega_k, q)$ are diagonal for all ω_k and q, estimating the de-mixing filters in frequency domain can be achieved by minimizing the following cost function

$$\mathcal{J}(\mathbf{W}(\omega_k)) = \sum_{k=1}^{K} \sum_{q=1}^{Q} \left\| Off(\mathbf{W}(\omega_k)\mathbf{R}(\omega_k, q)\mathbf{W}^H(\omega_k)) \right\|_F^2$$
(18)

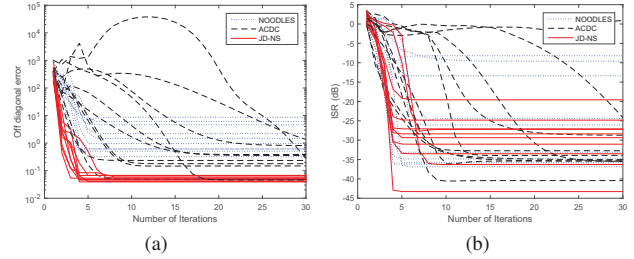
Fig. 1: The convergence patterns of three competitors and their performance in 10 independent trials.

where $\mathbf{R}(\omega_k, q) = \mathbf{R_x}(\omega_k, q) - \sigma^2\mathbf{I}$. From (18), we see that the optimization for each frequency bin ω_k are independent of each other, that is to say, optimization (18) can be solved by K different JD problems. However, the CBSS requires that permutation is consistent for each frequency bins, we always initialize the current JD algorithm by using the diagonalizer corresponding to the previous frequency bin.

V. SIMULATION RESULTS

The performance of the proposed algorithm is investigated using both synthetic data and real speech signals, and the results are compared with two nonunitary JD algorithms (ACDC [4] and NOODLES [13]). The performance of the JD algorithms is measured in terms of interference to signal ratio (ISR), defined as

$$\text{ISR} = \frac{1}{N} \left[\sum_{i=1}^{N} \left(\sum_{j=1}^{N} \frac{|g_{ij}|^2}{\max_l |g_{il}|^2} - 1 \right) \right.$$
$$\left. + \sum_{j=1}^{N} \left(\sum_{i=1}^{N} \frac{|g_{ij}|^2}{\max_l |g_{lj}|^2} - 1 \right) \right]$$
(19)

where $\mathbf{G} = \mathbf{B}^H\mathbf{A} = \{g_{ij}\}$ is the N dimensional global mixing-separating matrix, the lower the ISR, the better the performance.

Example 1: In this example, the JD algorithms are tested and compared using the synthetic data. The target matrices in set $\mathcal{R}$ are generated from model $\mathbf{R}_k = \mathbf{A}(\mathbf{D}_k + \sigma_f^2\mathbf{F}_k\mathbf{F}_k^H)\mathbf{A}^H + \sigma_e^2\mathbf{E}_k\mathbf{E}_k^H$, $k = 1, \ldots, K$, where the entries in $\mathbf{A}$, $\mathbf{D}_k$, $\mathbf{E}_k$ and $\mathbf{F}_k$ are drawn from a complex Normal distribution with zero mean and unit variance, $\mathbf{E}_k$ and $\mathbf{F}_k$ represent the disturbance of the data model, which are used to imitate the effect of sensor noises and statistics estimation error of the source signals respectively. σ_f^2 and σ_e^2 are used to control the level of disturbance, the signal to noise ratio (SNR) is defined as $\text{SNR} = 10\log_{10}(1/\sigma_e^2)$.

We set $M = 5$, $N = 3$, $K = 20$ and SNR = 10dB, $\sigma_f^2 = 0.01$, Fig. 1 plots the typical convergence patterns and the ISR performance of three competitors in 10 independent trials. Observing the evolution of the off diagonal error $\mathcal{J}$, we see that ACDC converges slower than its competitors, this is because that ACDC solves a inverse problem, which
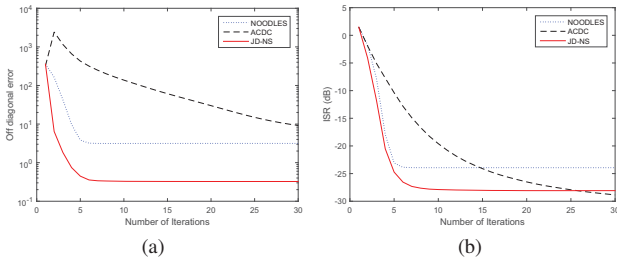
(a)   (b)

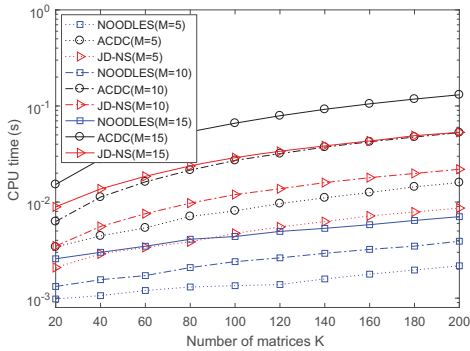Fig. 2: The averaged convergence patterns and ISR performance over 1000 independent trials.

Fig. 3: The elapsed time of each competitor for one single iteration.
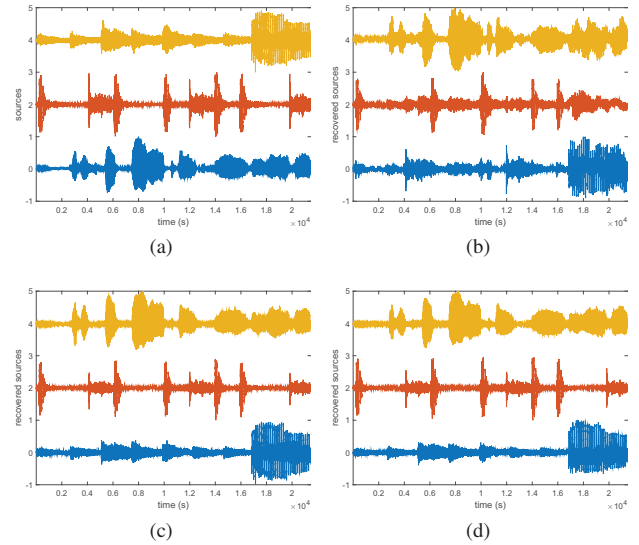

(a)   (b)

(c)   (d)

Fig. 4: Separation performance of JD algorithms in frequency domain CBSS: (a) source signals; (b) ACDC; (c) NOODLES; (d) JD-NS.

TABLE I: The SIR of recovered signals when SNR = 10dB.

sources	SIR (dB)		
	ACDC	NOODLES	JD-NS
1	2.1356	12.5451	16.6666
2	5.0295	14.8377	17.4827
3	7.1117	14.6993	20.2705

indirectly estimate the mixing matrix rather than the diagonalizer. Although NOODLES and the proposed algorithm use the same off diagonal error cost function, the proposed JD-NS algorithm converges faster than NOODLES. Observing the evolution of ISR, we see that ACDC and NOODLES converge with poor performance in some trials. In fact, they converge to degenerate solutions in these trials. The proposed algorithm converges to a well-conditioned solution, and hence uniformly reasonable perfromance. Fig. 2 shows the averaged performance, one sees that the convergence speed and ISR accuracy coincide with that in Fig. 1. In order to assess the complexity of proposed algorithm, we provide the elapsed time per iteration for three competitors under different dimensions of target matrices. Observing Fig. 3, we see that the proposed algorithm has a medium complexity.

Example 2: In this example, we demonstrate the performance of JD algorithms in CBSS application. We use $M = 4$ synthetic mixtures of $N = 3$ audio signals (one speech and two music shown in Fig. 4(a)) each with 21462 samples, each convolutive channel is modeled as an 8-tap FIR filter whose impulse responses are selected randomly from a Normal distribution with zero mean and unit variance. White Gaussian noise was added to the output of the convolutive mixing system at SNR = 10dB. As for the estimation of cross spectral density matrices, we divide the total data into $Q = 20$ epochs. At each epoch, $K = 256$-point FFTs, applied to time segments overlapping by 50%, weighted by Hamming

windows were used. Fig. 4 shows the separation performance of the JD algorithms in CBSS context. Observing the separated waveforms, we see that ACDC is inferior to NOODLES and JD-NS.

In order to quantitatively measure the separation accuracy of each JD algorithm in CBSS, we evaluate the signal to interference ratio (SIR) for three recovered source signals

$$\text{SIR}(n) = \frac{\max_j \tilde{g}_{nj}}{\sum_{j=1}^N \tilde{g}_{nj} - \max_j \tilde{g}_{nj}}, \quad n = 1, \ldots, N \quad (20)$$

where $\tilde{g}_{nj} = \sum_{k=0}^{K-1} |g_{nj}(\omega_k)|^2$, $g_{nj}(\omega_k)$ is the nj-th entry of frequency domain global mixing-separating matrix $\mathbf{G}(\omega_k) = \mathbf{W}(\omega_k)\mathbf{A}(\omega_k)$. Table 1 summarizes the averaged SIR of each recovered source signal for three competitors over 100 independent trials. We see that the proposed algorithm outperforms ACDC and NOODLES in terms of SIR performance for three source signals.

VI. CONCLUSION

We investigated the nonunitary JD with well-conditioned solutions for overdetermined CBSS, a multicriteria optimization model is proposed to solve JD problem. The resulting algorithm can successfully eliminated the degenerate solutions. Simulation results validate the superior ability of the proposed algorithm.

REFERENCES

[1] G. Chabriel, M. Kleinsteuber, E. Moreau, H. Shen, P. Tichavsky, and A. Yeredor, "Joint matrices decompositions and blind source separation," *IEEE Signal Process. Mag.*, vol. 31, no. 3, pp. 34–43, 2014.

[2] J.-F. Cardoso and A. Souloumiac, "Jacobi angles for simultaneous diagonalization," *SIAM J. Matrix Anal. Appl.*, vol. 17, no. 1, pp. 161-164, 1996.

[3] W.-T. Zhang and S.-T. Lou, "Householder transform based joint diagonal zero diagonalization for source separation using time-frequency distributions," *Multidim. Syst. Signal Process.*, vol. 21, no. 2, pp. 161-177, 2010.

[4] A. Yeredor, "Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation," *IEEE Trans. Signal Process.*, vol. 50, no. 7, pp. 1545-1553, 2002.

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

[6] R. Vollgraf and K. Obermayer, "Quadratic optimization for simultaneous matrix diagonalization," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3270-3278, 2006.

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

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

[9] A. Mesloub, K. Abed-Meraim and A. Belouchrani, "A new algorithm for complex non-orthogonal joint diagonalization based on Shear and Givens rotations," *IEEE Trans. Signal Process.* vol. 62, no. 8, pp. 1913-1925, 2014.

[10] A. Ziehe, P. Laskov, G. Nolte and K.-R. Muller, "A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation," *J. Mach. Learn. Res.*, vol. 5, pp. 777-800, 2004.

[11] F. Wang, Z. Liu, and J. Zhang, "Nonorthogonal joint diagonalization algorithm based on trigonometric parameterization," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5299-5308, 2007.

[12] V. Maurandi and E. Moreau, "A decoupled Jacobi-like algorithm for non-unitary joint diagonalization of complex-valued matrices," *IEEE Signal Process. Lett.*, vol. 21, no. 12, pp. 1453-1456, 2014.

[13] T. Trainini and E. Moreau, "A coordinate descent algorithm for complex joint diagonalization under hermitian and transpose congruences," *IEEE Trans. Signal Process.*, vol. 62, no. 19, pp. 4974-4983, 2014.

[14] 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 Process. Lett.*, vol. 14, no. 11, pp. 860-863, 2007.

[15] W.-T. Zhang, S.T. Lou, and H.M. Lu, "Fast nonunitary joint block iagonalization with degenerate solution elimination for convolutive blind source separation," *Digit. Signal Process.*, vol. 22, no. 5, pp. 808-819, 2012.

[16] W.-T. Zhang and S.-T. Lou, "Multicriteria optimization for nonunitary joint block diagonalization," in *ICASSP2016*, Mar. 20-25, Shanghai, China, pp. 2509-2513, 2016.

[17] K. Rahbar and J. P. Reilly, "A frequency domain method for blind source separation of convolutive audio mixtures," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 5, pp. 832-844, 2005.