MULTICRITERIA OPTIMIZATION FOR NONUNITARY JOINT BLOCK DIAGONALIZATION

Wei-Tao Zhang and Shun-Tian Lou

School of Electronic Engineering, Xidian University, Xi'an 710071, China

ABSTRACT

This paper deals with the nonunitary joint block diagonalization (JBD) of a set of given matrices for nonsquare mixing. As is known that the elimination of the degenerate solutions is a crucial problem in nonunitary JBD. Unlike the existing method, which optimizes a penalty term based least squares criterion to eliminate the degenerate solutions. In this paper we seek a well conditioned solution to nonunitary JBD instead. From this point of view, the nonunitary JBD problem is reformulated as a multicriteria optimization model. The resulting algorithm is numerically robust, and can be used in nonsquare mixing scenario. Moreover, it outperforms the existing JBD algorithms in terms of convergence rate and stability, which has been verified by our simulation results.

***Index Terms*—** Joint block diagonalization (JBD), multicriteria optimization, degenerate solution, nonsquare mixing.

1. INTRODUCTION

In recent years, the joint decomposition of a set of matrices or higher order tensor has received more and more research interest in a variety of fields [1]. Among which, the joint block diagonalization (JBD) of multiple target matrices is a powerful tool in many signal processing applications, such as direction finding in array processing [2], multidimensional independent component analysis [3,4], and time domain convolutive blind source separation (CBSS) [5-7], whose common objective is to estimate a matrix, termed as block diagonalizer, which jointly transforms the target matrices into block diagonal matrices.

Belouchrani et al. resolve the JBD problem by using successive Givens rotations [2,5], which corresponds to the unitary JBD since the block diagonalizer is restricted to be a unitary matrix. However, the unitary JBD algorithms result in exact block diagonalization of the one selected matrix at the cost of poor block diagonalization of others, and hence the degraded performance in applications. Nonunitary JBD algorithm has been developed for CBSS in [6] by using quadratic optimization, which turns out to be superior to the unitary one in terms of attainable performance. However, it has been shown in our previous work that this algorithm is prone to converge to some undesired degenerate solutions (singular or ill conditioned solutions). In the context of CBSS, the degenerate solutions will result in incomplete separation of sources, i.e., separating some of source signals with the others unretrieved. In order to eliminate the degenerate solutions, we have proposed a penalty term based least squares criterion, and developed a fast nonunitary JBD algorithm in [7], which exhibits the ability to eliminate the degenerate solutions in nonunitary JBD and provides better separation performance. Unfortunately, our previous work is devoted to square mixing case, where the mixing matrix is limited to be a square matrix. This leads to limited applications.

In this paper, we focus our attention on the well conditioned solution to JBD for nonsquare mixture, and reformulate the nonunitary JBD as a multicriteria optimization problem subject to commonly used constraint. The resulting algorithm can be applied in nonsquare mixing scenario, and again successfully eliminates the degenerate solutions.

This work was supported by the National Natural Science Foundation of China by Grants 61571339 and 61201299.

2. PROBLEM STATEMENT

Let us consider a set $\mathcal{R}$ of $K \quad M \times M$ complex valued matrices $\mathbf{R}_k$, built as

$$\mathbf{R}_k = \mathbf{A}\mathbf{D}_k\mathbf{A}^H, \quad k = 1, \dots, K \tag{1}$$

where $\mathbf{A}$ is a $M \times N$ $(M \geq N)$ dimensional mixing matrix with full column rank, $\mathbf{D}_k$ is a $N \times N$ dimensional block diagonal matrix with r diagonal blocks of square matrices each with size n_i, $i = 1, \dots, r$,($\sum_{i=1}^{r} n_i = N$), and its off-diagonal blocks are all zero matrices of proper size. The goal of nonunitary JBD is to seek a nonsingular block diagonalizer $\mathbf{B}$ of dimension $M \times N$ such that the matrices $\mathbf{B}^H\mathbf{R}_k\mathbf{B}$, $k = 1, \dots, K$ are all block diagonal matrices.

JBD is actually a blind identification problem since $\mathbf{A}$ and $\mathbf{D}_k$, $k = 1, \dots, K$ are all supposed unknown. Let us divide $\mathbf{B}$ into r partitions as $\mathbf{B} = [\mathbf{B}_1, \mathbf{B}_2, ..., \mathbf{B}_r]$, where $\mathbf{B}_i$, $i = 1, \dots, r$ consists of the associated n_i columns of $\mathbf{B}$. So the matrix $\mathbf{B}$ can only be estimated up to the scaling and permutation of its partitions, i.e., $\hat{\mathbf{B}} = \mathbf{B}\Sigma\Pi$ with Π and Σ denote, respectively, a block permutation matrix and a nonsingular block diagonal matrix of proper size [7]. This corresponds to the arbitrary attenuation and order of restored source signals in CBSS context.

The nonunitary JBD algorithm proposed in [6] (termed as JBD-NU) minimizes the following cost function

$$\mathcal{J}_1(\mathbf{B}) = \sum_{k=1}^{K} \left\| OffB(\mathbf{B}^H \mathbf{R}_k \mathbf{B}) \right\|_F^2 \tag{2}$$

where the operator $OffB(\cdot)$ forms a matrix by replacing the diagonal blocks of its argument with zero matrices of proper size, and $\|\cdot\|_F$ denotes the matrix Frobenius norm. The criterion (2) can be regarded as the block diagonalization error. However, the JBD-NU algorithm may converge to some undesired degenerate solutions in many situations, since the minimization of (2) can not guarantee the nonsingularity of $\mathbf{B}$. To this end, we have proposed the following criterion [7]

$$\mathcal{J}_2(\mathbf{B}) = \mathcal{J}_1(\mathbf{B}) - \log|\det(\mathbf{B})| \tag{3}$$

where $\det(\cdot)$ denotes the matrix determinant. One sees that the minimization of (3) results in nonsingular solution to JBD, and hence eliminates the degenerate solutions. However, this algorithm is limited to square mixing case with $M = N$, and the numerical problem will arise when it is applied in exact block diagonalizable data set (see section 3.2).

3. MULTICRITERIA OPTIMIZATION FOR JBD

In order to overcome the above drawbacks, in this section, we firstly present a multicriteria model for nonunitary JBD, then the optimization algorithm is derived, and its detailed implementation is also summarized.

3.1. The Criteria

To eliminate the degenerate solutions, we now focus on a well conditioned solution to nonunitary JBD, which firstly requires that the columns of $\mathbf{B}$ have uniform norms. Without losing generality, we impose the unit norm restriction on the columns of $\mathbf{B}$, i.e., $\mathbf{b}_j^H \mathbf{b}_j = 1, j = 1, \ldots, N$, where $\mathbf{b}_j$ denotes the jth column vector of $\mathbf{B}$. Further note that [8]

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

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 for JBD

$$\begin{aligned} &\min_{\mathbf{B}} \mathcal{J}_1(\mathbf{B}), \quad \max_{\mathbf{B}} \det(\mathbf{B}^H\mathbf{B}), \\ &s.t. \quad \mathbf{b}_i^H \mathbf{b}_i = 1, \quad i = 1, \ldots, N \end{aligned} \tag{5}$$

We can see that the block diagonalizer $\mathbf{B}$ can be nonsquare in model (5), the minimization of $\mathcal{J}_1(\mathbf{B})$ guarantees that $\mathbf{B}$ is the block diagonalizer of the set $\mathcal{R}$, while the maximization of $\det(\mathbf{B}^H\mathbf{B})$ under the unit norm constraint guarantees that $\mathbf{B}$ is well conditioned. Moreover, we see that $\mathcal{J}_1(\mathbf{B})$ is lower bounded by 0, and for $\det(\mathbf{B}^H\mathbf{B})$, by using the Hadamard inequality we have

$$\det(\mathbf{B}^H\mathbf{B}) \leq \prod_{j=1}^{N} \tilde{b}_{jj} = 1 \tag{6}$$

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

3.2. Optimization Algorithm

Direct optimization of (5) with respect to (w.r.t.) $\mathbf{B}$ is analytically cumbersome, we here exploit the cyclic optimization technique [9], whose basic idea is that firstly dividing the parameters remains to be optimized into multiple groups, then optimizing the criterion w.r.t. one group with the others known and fixed. The overall optimization is achieved by alternately optimizing the criterion w.r.t. each group. Then for the lth partition $\mathbf{B}_l$, after some matrix manipulations, we have

$$\mathcal{J}_1(\mathbf{B}_l) = tr(\mathbf{B}_l^H \mathbf{Q}_l \mathbf{B}_l) \tag{7}$$

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

$$\mathbf{Q}_l = \sum_{k=1}^{K} \mathbf{R}_k \mathbf{B}_{(l)} \mathbf{B}_{(l)}^H \mathbf{R}_k^H + \mathbf{R}_k^H \mathbf{B}_{(l)} \mathbf{B}_{(l)}^H \mathbf{R}_k \tag{8}$$

with $\mathbf{B}_{(l)}$ arises from the deletion of partition $\mathbf{B}_l$ from $\mathbf{B}$. Let $\bar{\mathbf{B}} = [\mathbf{b}_{lm}, \mathbf{B}_{(lm)}]$ be the column exchanged matrix of $\mathbf{B}$, where $\mathbf{b}_{lm}$ denotes the mth ($m \in \{1, \ldots, n_l\}$) column vector of partition $\mathbf{B}_l$, and $\mathbf{B}_{(lm)}$ arise from the deletion of vector $\mathbf{b}_{lm}$ from $\mathbf{B}$. Then for $\det(\mathbf{B}^H\mathbf{B})$, we have

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

where

$$\mathbf{P}_{lm}^{\perp} = \mathbf{I} - \mathbf{B}_{(lm)} [\mathbf{B}_{(lm)}^H \mathbf{B}_{(lm)}]^{-1} \mathbf{B}_{(lm)}^H \tag{10}$$

with $\mathbf{I}$ denoting the identity matrix. Since $\mathbf{b}_{lm}$ is independent of $\mathbf{B}_{(lm)}$, we show that the model (5) is equivalent to

$$\begin{aligned} &\min_{\mathbf{b}_{lm}} \mathbf{b}_{lm}^H \mathbf{Q}_l \mathbf{b}_{lm}, \quad \max_{\mathbf{b}_{lm}} \mathbf{b}_{lm}^H \mathbf{P}_{lm}^{\perp} \mathbf{b}_{lm}, \\ &s.t. \quad \mathbf{b}_{lm}^H \mathbf{b}_{lm} = 1, \quad l = 1, \ldots, r, \quad m = 1, \ldots, n_l \end{aligned} \tag{11}$$

For exact block diagonalizable data set $\mathcal{R}$, the solution to (11) should consider two different stages. In the initial stage,

Table 1: The pseudo code of JBD-NS algorithm.

Initialize $\mathbf{B}$, e.g., $\mathbf{B} = [\mathbf{I}, \mathbf{0}]^T$ of size $M \times N$.

Repeat
For $l = 1, ..., r$ do
1) Compute $\mathbf{Q}_l$ by using (8).
2) **For** $m = 1, ..., n_l$ do
a) Compute $\mathbf{P}_{lm}^{\perp}$ by using (10)
b) If $\mathbf{Q}_l$ is invertible, then $\mathbf{b}_{lm} = \mathbf{v}_{\max}$
Else $\mathbf{b}_{lm}$ is solved by (12) and (13)
End
End
Until convergence

the matrix $\mathbf{Q}_l$ is invertible, then we take $\mathbf{b}_{lm} = \mathbf{v}_{\max}$, where $\mathbf{v}_{\max}$ denotes the unit norm generalized eigenvector of matrix pencil $(\mathbf{P}_{lm}^{\perp}, \mathbf{Q}_l)$ associated with the largest generalized eigenvalue. In the second stage, i.e., when the algorithm approaches convergence, $\mathbf{Q}_l$ will be singular with rank $N - n_l$, then the above procedure is numerically unstable (our previous work also used the inverse of $\mathbf{Q}_l$, this is also the reason for numerical problem). In this stage, let $\mathbf{U}_0$ be the eigenvectors of $\mathbf{Q}_l$ associated with the $M - N + n_l$ zero eigenvalues (the $M - N + n_l$ smallest eigenvalues in practice), we take

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

where vector $\mathbf{w}$ consists of $M - N + n_l$ weight coefficients. Substituting (12) into the second criterion of (11), we have

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

The optimal solution to $\mathbf{w}$ is obtained by taking the eigenvector of $\mathbf{U}_0^H \mathbf{P}_{lm}^{\perp} \mathbf{U}_0$ associated with the largest eigenvalue. We see that in the convergence stage, $\min_{\mathbf{b}_{lm}} \mathbf{b}_{lm}^H \mathbf{Q}_l \mathbf{b}_{lm}$ is solved by (12) and $\max_{\mathbf{b}_{lm}} \mathbf{b}_{lm}^H \mathbf{P}_{lm}^{\perp} \mathbf{b}_{lm}$ is solved by (13). Finally we summarize the multicriteria optimization algorithm for nonsquare mixing JBD (termed as JBD-NS) in Table 1.

4. RESULTS

The performance of the proposed algorithm is demonstrated in three simulation examples, and the results are compared with the JBD-NU algorithm in [6] and our previous work for square mixing case in [7] (termed as JBD-S). The interference to signal ratio (ISR), defined as [5-7]

$$\mathrm{ISR} = \frac{1}{r(r-1)} \left[\sum_{i=1}^{r} \left(\sum_{j=1}^{r} \frac{||\mathbf{G}_{ij}||_F^2}{\max_l ||\mathbf{G}_{il}||_F^2} - 1 \right) + \sum_{j=1}^{r} \left(\sum_{i=1}^{r} \frac{||\mathbf{G}_{ij}||_F^2}{\max_l ||\mathbf{G}_{lj}||_F^2} - 1 \right) \right] \tag{14}$$

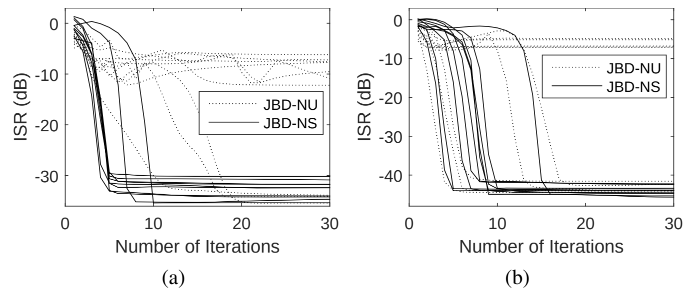

Fig. 1: The ISR versus the number of iterations of 10 independent trials: (a) SNR = 5dB; (b) SNR = 10dB.

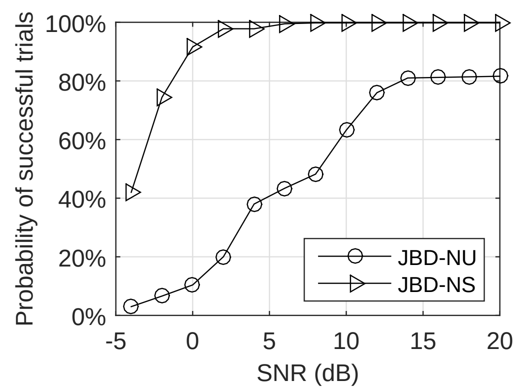

Fig. 2: The probability of successful trials for two competitors.

is used as the performance measure, where $\mathbf{G} = \mathbf{B}^H \mathbf{A}$ is the global mixing-separating matrix. For a perfect separation of multiple sources in CBSS context, matrix $\mathbf{G}$ should be equal to a block permutation matrix multiplied by a nonsingular block diagonal matrix, so the lower the ISR value, the better the performance. Besides the separation performance above, another important issue is the reliability of the nonunitary JBD algorithms, which is measured by the probability of successful trials (PST) in Monte Carlo simulations. A trial is said to be successful if the algorithm converges with a reasonable ISR performance.

In the following simulation examples, The target matrices in set $\mathcal{R}$ are generated from model $\mathbf{R}_k = \mathbf{A}\mathbf{D}_k\mathbf{A}^H + \sigma^2 \mathbf{E}_k \mathbf{E}_k^H$, $k = 1, \ldots, K$, where $\mathbf{D}_k$ is a block diagonal matrix (for the seek of simplicity, the r diagonal blocks of $\mathbf{D}_k$ have the same size n, i.e., $N = nr$), and the entries in $\mathbf{A}$, $\mathbf{D}_k$ and $\mathbf{E}_k$ are drawn from a complex Normal distribution with zero mean and unit variance, then the signal to noise ratio is defined as $\mathrm{SNR} = 10 \log_{10}(1/\sigma^2)$.

Example 1: We set $M = 15$, $N = 12$, $K = 20$ and $r = 3$. Fig. 1 plots the typical convergence patterns of the ISR in 10 independent trials for two nonunitary JBD algorithms. One sees that the JBD-NU algorithm is prone to converge to some solutions with inferior ISR performance, which actu-

Table 2: The PST and ISR under different number of diagonal blocks with $M = 30$, $N = 24$.

SNR	algorithms	r			
		3	4	6	8
5dB	JBD-NU	67.5%, -36.24	0, /	0, /	0, /
	JBD-NS	100%, -36.35	100%, -32.92	94.5%, -28.45	60%, -25.89
10dB	JBD-NU	95%, -47.06	0, /	0, /	0, /
	JBD-NS	100%, -47.07	100%, -45.02	99.5%, -41.62	82%, -38.34

Table 3: The PST and ISR under different matrix size with $M = 24$, $r = 4$.

SNR	algorithms	N			
		12	16	20	24
5dB	JBD-NU	0, /	0, /	13%, -30.54	24%,-23.88
	JBD-S	/	/	/	62%,-20.32
	JBD-NS	100%, -31.26	97%, -33.03	98%, -30.62	98%, -23.83
10dB	JBD-NU	0, /	0, /	18%, -43.19	46%, -36.92
	JBD-S	/	/	/	91%,-32.65
	JBD-NS	100%, -41.89	100%, -43.98	99%, -43.14	99%, -37.74

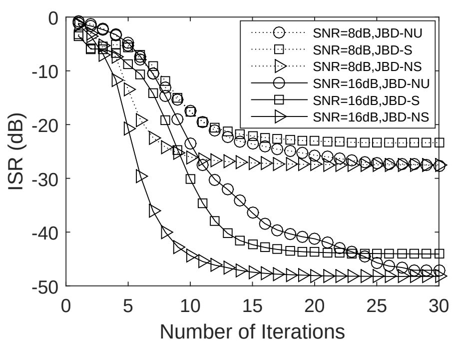

Fig. 3: The averaged convergence behavior of three competitors.

ally corresponds to the degenerate solutions. Even in higher SNR scenario, the JBD-NU algorithm may still converges to degenerate solutions with a certain probability. The proposed algorithm converges to solutions with reasonable ISR in all trials, which corresponds to well conditioned solutions. In order to investigate the reliability of two competitors, we perform 500 independent trials, Fig. 2 shows the PST at the various SNR values. We see that the JBD-NU algorithm is not reliable even in higher SNR cases, while the proposed algorithm is more reliable even in worse SNR conditions.

Example 2: In this example, we investigate the convergence behavior of the nonunitary JBD algorithms by setting $M = N = 12$, $K = 20$, $r = 3$. In this square case, the JBD-S algorithm is also considered. Fig. 3 shows the averaged convergence behavior over those successful trials, where a total of 500 independent trials are performed. We see that the proposed algorithm yields the fastest convergence rate.

Example 3: In this example, we investigate how the number of diagonal blocks r and the block diagonal matrix size N affects the performance of the nonunitary JBD algorithms. By setting $M = 30$, $N = 24$, $K = 40$. Table 2 summarizes the PST and the corresponding final ISR performance in various SNR scenario. We see that when the number of diagonal blocks increases, the PST of the nonunitary JBD algorithms decreases. This is not surprising since the increase of r actually indicates the increase of interference, and hence the degraded performance. One sees that even in strong interference scenario, the proposed algorithm still yields reasonable PST and ISR, while the JBD-NU algorithm completely fails. Table 3 summarizes the PST and the corresponding ISR performance for fixed number of diagonal blocks $r = 4$. One sees that JBD-NU algorithm can not be used in the case that M is far greater than N, whereas the proposed algorithm yields the good ISR performance with higher reliability for all cases.

5. CONCLUSION

In this paper, we investigate the nonunitary JBD with well conditioned solutions for nonsquare mixing case. To this end, the nonunitary JBD is reformulated as a multicriteria optimization problem. The cyclic minimization technique is exploited, yielding a numerically robust JBD algorithm applicable to nonsquare mixtures. Simulation results show that the proposed algorithm is more reliable for nonunitary JBD in nonsquare mixing case, and exhibits faster convergence rate.

6. 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, May 2014.

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

[3] J.-F. Cardoso, "Multidimensional independent component analysis," in *Proc. ICASSP*, Seattle, May 1998, pp. 1941–1944.

[4] F.J. Theis, "Blind signal separation into groups of dependent signals using joint block diagonalization," in *Proc. ISCAS*, Kobe, Japan, 2005, pp. 5878–5881.

[5] H. Bousbia-Salah, A. Belouchrani, and K. Abed-Meraim, "Jacobi-like algorithm for blind signal separation of convolutive mixtures," *Electron. Lett.*, vol. 37, no. 16, pp. 1049–1050, 2001.

[6] 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.

[7] 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.

[8] J.K. Merikoski, U. Urpala, A. Virtanen, T.-Y. Tam, and F. Uhlig, "A best upper bound for the 2-norm condition number of a matrix," *Linear Algebra Appl.*, vol. 254, pp. 355-365, 1997.

[9] P. Stoica and Y. Selen, "Cyclic minimizers, majorization techniques, and the expectation maximization algorithm: a refresher," *IEEE Signal Process. Mag.*, vol. 21, no. 1, pp. 112-114, 2004.