On Computation of Approximate Joint Block-Diagonalization Using Ordinary AJD*

Petr Tichavský[1,3], Arie Yeredor[2], and Zbyněk Koldovský[1,3]

[1] Institute of Information Theory and Automation, Pod vodárenskou věží 4, P.O. Box 18, 182 08 Praha 8, Czech Republic
tichavsk@utia.cas.cz
[2] Dept. of Electrical Engineering - Systems, School of Electrical Engineering, Tel-Aviv University, P.O. Box 39040 Tel-Aviv, 69978 Israel
[3] Faculty of Mechatronic and Interdisciplinary Studies
Technical University of Liberec, Studentská 2, 461 17 Liberec, Czech Republic

Abstract. Approximate joint block diagonalization (AJBD) of a set of matrices has applications in blind source separation, e.g., when the signal mixtures contain mutually independent subspaces of dimension higher than one. The main message of this paper is that certain ordinary approximate joint diagonalization (AJD) methods (which were originally derived for "degenerate" subspaces of dimension 1) can also be used successfully for AJBD, but not all are suitable equally well. In particular, we prove that when the set is exactly jointly block-diagonalizable, perfect block-diagonalization is attainable by the recently proposed AJD algorithm "U-WEDGE" (uniformly weighted exhaustive diagonalization with Gaussian iteration) - but this basic consistency property is not shared by some other popular AJD algorithms. In addition, we show using simulation, that in the more general noisy case, the subspace identification accuracy of U-WEDGE compares favorably to competitors.

1 Introduction

Consider a set of square symmetric matrices $\mathbf{M}_i$, $i = 1, \ldots, N$, that are all block diagonal, with K blocks of size $L \times L$ along its main diagonal, $\mathbf{M}_i = \mathrm{Bdiag}(\mathbf{M}_{i1}, \ldots, \mathbf{M}_{iK})$, where $\mathbf{M}_{ik}$ is the $k-$th block of $\mathbf{M}_i$ and the $\mathrm{Bdiag}(\cdot)$ operator constructs a block-diagonal matrix from its argument matrices. It follows that the dimension of the matrices is $LK \times LK$. An example of such matrices is illustrated in Figure 2(a) at the end of the paper. Note that the assumption that all blocks are of the same size is only used here to simplify the exposition, and can be relaxed via straightforward generalization.

Next, assume that (possibly perturbed) congruence transformations of these matrices are given as

$$\mathbf{R}_i = \mathbf{A}\mathbf{M}_i\mathbf{A}^T + \mathbf{N}_i, \qquad i = 1, \ldots, N \tag{1}$$

* This work was supported by Ministry of Education, Youth and Sports of the Czech Republic through the project 1M0572 and by Grant Agency of the Czech Republic through the project 102/09/1278.

F. Theis et al. (Eds.): LVA/ICA 2012, LNCS 7191, pp. 163–171, 2012.

where the superscript T denotes a matrix transposition, $\mathbf{A}$ is an unknown square "mixing matrix", and $\mathbf{N}_i$ is a perturbation (or "noise") matrix. We shall refer to the case where all $\mathbf{N}_i = \mathbf{0}$, $i = 1, \ldots, N$ as the "unperturbed" (or "noiseless") case. The choice of symbol $\mathbf{R}$ reflects the fact that the matrices in the set often play a role of (sample-) covariance matrices of a partitioned data, or time-lagged (sample-) covariance matrices.

The goal in Approximate Joint Block Diagonalization (AJBD) is to find a "demixing" matrix $\mathbf{W}$, such that the matrices

$$\widehat{\mathbf{M}}_i = \mathbf{W}\mathbf{R}_i\mathbf{W}^T, \qquad i = 1, \ldots, N \tag{2}$$

are all approximately block diagonal, having the blocks on the main diagonal of the same size as the original matrices $\mathbf{M}_{ik}$. Ideally, one may wish to estimate $\mathbf{W} = \mathbf{A}^{-1}$ and get $\widehat{\mathbf{M}}_i \approx \mathrm{Bdiag}(\widehat{\mathbf{M}}_{i1}, \ldots, \widehat{\mathbf{M}}_{iK})$, where $\widehat{\mathbf{M}}_{ik} \approx \mathbf{M}_{ik}$.

In general, however, it is impossible to recover the original blocks $\mathbf{M}_i$ (even in the "noiseless" case), because of inherent ambiguities of the problem (e.g., [10]), but it is possible to recover "independent subspaces", as explained below.

Let $\mathbf{W}_0 = \mathbf{A}^{-1}$ be partitioned in K blocks $\mathbf{W}_k$ of size $L \times KL$, $\mathbf{W}_0 = [\mathbf{W}_1^T, \ldots, \mathbf{W}_K^T]^T$. Each block $\mathbf{W}_k$ represents a linear space of all linear combinations of its rows. These linear spaces are in general uniquely identifiable [10,4]. Let $\widehat{\mathbf{W}}$ be an estimated demixing matrix. We say that $\widehat{\mathbf{W}}$ is "essentially equivalent" to $\mathbf{W}_0$ (and therefore represents an ideal joint block diagonalization), if there exists a suitable $LK \times LK$ permutation matrix $\boldsymbol{\Pi}$ such that for each $k = 1, \ldots, K$ the subspaces spanned by $\mathbf{W}_k$ and by the respective k-th block of $\boldsymbol{\Pi}\widehat{\mathbf{W}}$ coincide (two subspaces are said to coincide if their mutual angle[1] is zero).

Some existing AJBD algorithms are restricted to the case where $\mathbf{A}$ (and therefore also $\widehat{\mathbf{W}}$) are orthogonal [5], some other algorithms consider a general matrix $\mathbf{A}$ [6,10]. In this paper, we examine the general case.

It is known that reasonable solutions to AJBD can be obtained using a two steps approach, by first applying an ordinary approximate joint diagonalization (AJD) algorithm, and then clustering the separated components (rows of the demixing matrix) [7,12]. In Section 3 we suggest a method for the clustering operation, followed by the main point of this paper: we show that not all AJD algorithms are equally suitable for such a two-steps AJBD approach. More specifically, we prove that unlike several popular AJD approaches, one recently proposed AJD method (U-WEDGE, Uniformly Weighted Exhaustive Diagonalization with Gauss itErations [14]) features a unique ability to attain ideal separation in the unperturbed ("noiseless") case, for general (not necessarily orthogonal) matrices $\mathbf{A}$. Our theoretical results are corroborated with simulation experiments in Section 4, both for the unperturbed and perturbed cases, showing the empirical advantages of U-WEDGE for the latter. We start, however, with a short overview of the AJD methods considered in this work. Their applicability in solving the block AJD problem is studied later in Section 4.

[1] The mutual angle between two subspaces can be obtained in Matlab® using the `subspace` function.

2 Survey of Main AJD Methods

Several well-known AJD methods are based on minimization of one of the three following criteria, possibly subject to one of the two constraints stated below.

$$C_{\rm LS}(\mathbf{W}) = \sum_{i=1}^{N} \|{\rm Off}(\mathbf{W}\mathbf{R}_i\mathbf{W}^T)\|_F^2 \tag{3}$$

$$C_{\rm LL}(\mathbf{W}) = \sum_{i=1}^{N} \log \frac{\det {\rm Ddiag}(\mathbf{W}\widehat{\mathbf{R}}_i\mathbf{W}^T)}{\det(\mathbf{W}\widehat{\mathbf{R}}_i\mathbf{W}^T)} \tag{4}$$

$$C_{\rm J2}(\mathbf{W}) = \sum_{i=1}^{N} \|\mathbf{R}_i - \mathbf{W}^{-1}{\rm Ddiag}(\mathbf{W}\mathbf{R}_i\mathbf{W}^T)\mathbf{W}^{-T}\|_F^2 \tag{5}$$

where the operator "Off" nullifies the diagonal elements, whereas "Ddiag" nullifies the off-diagonal elements of a square matrix, ${\rm Ddiag}(\mathbf{M}) = \mathbf{M} - {\rm Off}(\mathbf{M})$), and "$\|\cdot\|_F$" stands for the Frobenius norm. The possible associated constraints are

1. Each row of the estimated demixing matrix $\widehat{\mathbf{W}}$ has unit Euclidean norm.
2. $\widehat{\mathbf{W}}\mathbf{R}_1\widehat{\mathbf{W}}^T$ has an all-ones main diagonal.

The latter constraint usually corresponds (in the BSS context) to some scaling constraint on the estimated sources.

In the sequel we shall examine five AJD methods: QAJD [15], FAJD [9], LLAJD [11], QRJ2D [2] and WEDGE [14], especially in its unweighted version U-WEDGE. QAJD is based on minimization of the criterion (3) under the constraint 2. FAJD minimizes (3), penalized by a term proportional to $\log|\det \mathbf{W}|$. LLAJD minimizes (4) and QRJ2D minimizes (5), both under the constraint 1 (which is actually immaterial to the minimization in these cases).

WEDGE and its more simple unweighted (or uniformly-weighted) version U-WEDGE, which we consider in here, are different. U-WEDGE seeks a demixing matrix $\mathbf{W}$ which satisfies

$$\operatorname{argmin}_{\mathbf{A}} \sum_{i=1}^{N} \|\mathbf{W}\mathbf{R}_i\mathbf{W}^T - \mathbf{A}\,{\rm Ddiag}(\mathbf{W}\mathbf{R}_i\mathbf{W}^T)\,\mathbf{A}^T\|_{\rm F}^2 = \mathbf{I} \tag{6}$$

where $\mathbf{I}$ is the $LK \times LK$ identity matrix. Roughly speaking, this implies that the set of matrices $\{\mathbf{W}\mathbf{R}_i\mathbf{W}^T\}$ cannot be jointly-diagonalized any further, since its "residual mixing" matrix, or its "best direct-form diagonalizer" (in the LS sense) is $\bar{\mathbf{A}} = \mathbf{I}$, the identity matrix.

It was shown in [14] that a necessary and sufficient condition for $\mathbf{A} = \mathbf{I}$ to be a stationary point of the criterion in (6) is a simpler set of nonlinear "normal equations",

$${\rm Off}\left[\sum_{i=1}^{N}(\mathbf{W}\mathbf{R}_i\mathbf{W}^T){\rm Ddiag}(\mathbf{W}\mathbf{R}_i\mathbf{W}^T)\right] = \mathbf{0}\ . \tag{7}$$

The more general WEDGE algorithm differs from U-WEDGE by incorporating special weight matrices in the quadratic criterion in (6). Although apparently complicated, both versions are computationally very efficient. Particular forms of WEDGE are used successfully in WASOBI for asymptotically optimal blind separation of stationary sources with spectrum diversity, and in BGSEP for separation of nonstationary sources [14]. Although our analytical proof and experiments in the sequel refer to the more simple version U-WEDGE, our experience shows that WEDGE shares the same ability of U-WEDGE to attain exact joint block-diagonalization in the unperturbed case.

3 AJD Methods in the Block Scenario

A natural extension of AJD methods in the block scenario is to replace the criterion (3) by

$$C_{\mathrm{BLS}}(\mathbf{W}) = \sum_{i=1}^{N} \|\mathrm{Boff}(\mathbf{W}\mathbf{R}_i\mathbf{W}^T)\|_F^2 \tag{8}$$

where the operator "Boff" nullifies the elements of a matrix that lie in the diagonal blocks. This is the main idea in [5].

It is obvious that since the criteria (3) and (8) are generally different, their minima differ as well, in general. If the diagonal blocks' sizes L are small, then one may expect the AJD and AJBD solutions to resemble. It is, however, necessary to permute (namely to properly cluster) the rows in the estimated demixing matrix, because the resulting order of rows is arbitrary in plain AJD algorithms.

3.1 Clustering of AJD Components

In this subsection a simple method of clustering the rows of de-mixing matrix is proposed. It allows to reveal (or at least to enhance) the block structure of the result. We suggest the following greedy algorithm: Given the AJD demixing matrix $\mathbf{W}$, compute an auxiliary matrix $\mathbf{B}$ as

$$\mathbf{B} = \sum_{i=1}^{N} |\mathbf{W}\mathbf{R}_i\mathbf{W}^{\mathbf{T}}| \ . \tag{9}$$

where the absolute value is taken elementwise. If the demixing is perfect, $\mathbf{B}$ should have, after arranging columns and rows, the same block structure as the original matrices $\mathbf{M}_i$. Take the first column of $\mathbf{B}$ and sort its elements decreasingly. Let $i_1, \ldots, i_L$ be the indices of the column elements with the L largest values. Then $\mathbf{W}_1$ is built of the rows of $\mathbf{W}$ with these indices. The rows and columns of $\mathbf{B}$ at the positions $i_1, \ldots, i_L$ are set to zero, and the procedure iterates further sorting of the column of $\mathbf{B}$ with the next nonzero elements, until all subspaces (blocks) $\mathbf{W}_k$, $k = 1, \ldots, K$, have been determined.

3.2 U-WEDGE Provides Perfect Separation of the Blocks

In this subsection we prove that in the unperturbed ("noiseless") case, U-WEDGE provides, upon convergence, perfect separation of the blocks.[2] Let $\mathbf{V}_k$ be the result of the hypothetical operation of applying U-WEDGE to each of the blocks-sets $\mathbf{M}_{1k}, \ldots, \mathbf{M}_{Nk}$ for $k = 1, \ldots, K$, where $\mathbf{M}_{ik}$ is the kth diagonal block of $\mathbf{M}_i$, $i = 1, \ldots, N$. It follows from (7) that each $\mathbf{V}_k$ obeys

$$\mathrm{Off}\left[\sum_{i=1}^{N} (\mathbf{V}_k \mathbf{M}_{ik} \mathbf{V}_k^T)\, \mathrm{Ddiag}(\mathbf{V}_k \mathbf{M}_{ik} \mathbf{V}_k^T)\right] = \mathbf{0}\ . \tag{10}$$

Now, define $\mathbf{W}_U$ as

$$\mathbf{W}_U = \mathrm{Bdiag}(\mathbf{V}_1, \ldots, \mathbf{V}_K)\, \mathbf{A}^{-1}\ . \tag{11}$$

It is straightforward to see that $\mathbf{W}_U$ is a U-WEDGE block diagonalizer of the original matrix set $\mathbf{R}_i = \mathbf{A}\mathbf{M}_i\mathbf{A}^T$, because it obeys the corresponding normal equation

$$\mathrm{Off}\left[\sum_{i=1}^{N} (\mathbf{W}_U \mathbf{R}_i \mathbf{W}_U^T)\, \mathrm{Ddiag}(\mathbf{W}_U \mathbf{R}_i \mathbf{W}_U^T)\right] = \mathbf{0}\ , \tag{12}$$

and on the other hand, that $\mathbf{W}_U\mathbf{R}_i\mathbf{W}_U^T$ has the perfect block-diagonal structure,

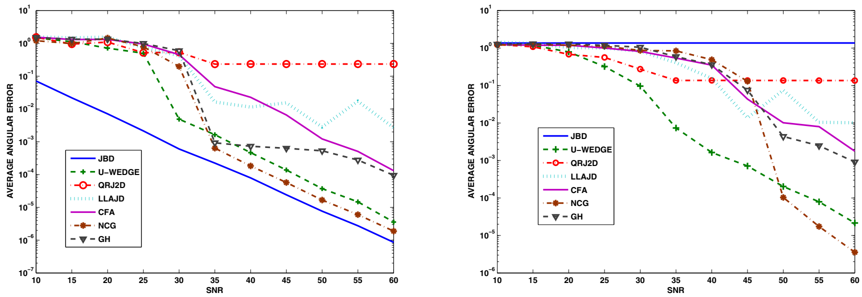
$$\mathbf{W}_U \mathbf{R}_i \mathbf{W}_U^T = \mathrm{Bdiag}(\mathbf{V}_1 \mathbf{M}_{i1} \mathbf{V}_1^T, \ldots, \mathbf{V}_K \mathbf{M}_{iK} \mathbf{V}_K^T), \qquad i = 1, \ldots, N\ . \tag{13}$$

We note in passing, that since, as mentioned in [14], (7) is also a necessary condition for a solution of the FFDiag AJD algorithm [16], this property is shared by the latter as well.

4 Simulation Experiments

We first consider an experiment reflecting the unperturbed case, as shown in Figure 2 at the end of the paper. We generated $N = 3$ block-diagonal matrices $\mathbf{M}_i$, $i = 1, 2, 3$, of dimension 20×20, each containing four symmetric 5×5 blocks $\mathbf{M}_{ik}$ generated as $\mathbf{M}_{ik} = \mathbf{H}_{ik}\mathbf{H}_{ik}^T$, $\mathbf{H}_{ik}$ being random 5×5 matrices with independent standard Gaussian elements. The matrices $\mathbf{M}_i$ are shown in diagram (a). The 20×20 mixing matrix $\mathbf{A}$ was generated as random orthogonal, via QR decomposition of a random matrix. Diagram (b) shows raw results of applying U-WEDGE to the unperturbed set $\mathbf{R}_i = \mathbf{A}\mathbf{M}_i\mathbf{A}^T$, $i = 1, 2, 3$. Obviously, the block-diagonal structure of the results is obscured by residual random permutations in these

[2] Theoretically U-WEDGE can be stacked in a false solution [14], but in practice it is very rare, and the solution is unique up to well known permutation ambiguity.

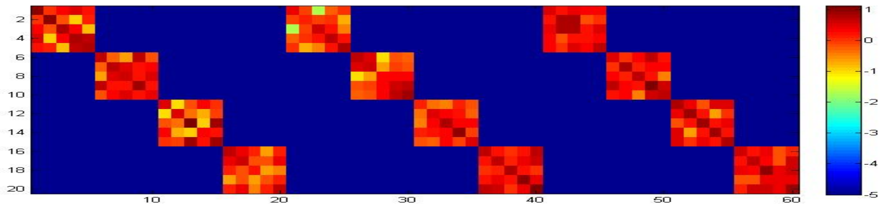


(a) (b)

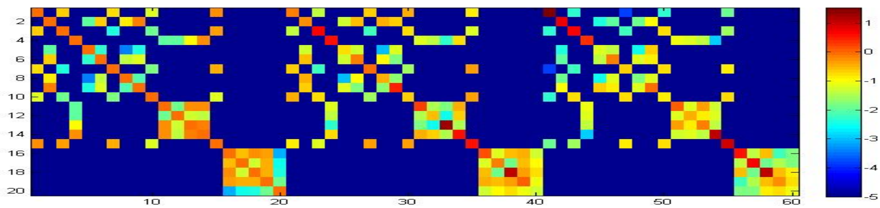
Fig. 1. Average subspace angular error for different AJD techniques versus SNR. (a) orthogonal mixing matrix, (b) random mixing matrix.

matrices. Diagram (c) shows the same matrices after applying the re-ordering procedure described in section 3.1. The angular error between the estimated and original subspaces (blocks of $\mathbf{W}$ and $\mathbf{A}^{-1}$) are zeros. Diagrams (d) and (e) show results obtained using the same procedure with the AJD algorithms QRJ2D and LLAJD. The average angular errors of the estimated subspaces were 4.6x10^{-3} (rad) and 1.3x10^{-4}(rad), respectively. The algorithms QAJD and FAJD were excluded, as they did not converge properly in this experiment.

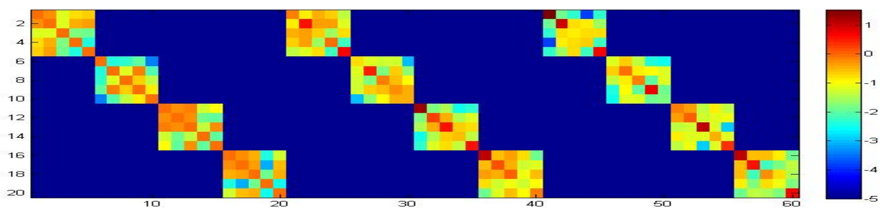
In Figure 1 we proceed to compare the performance in the perturbed (“noisy") case. We plot the average angular subspace errors vs. the Signal-to-Noise Ratio (SNR) for the two-steps method using U-WEDGE, QRJ2D, LLAJD. For reference, we also compare to a unitary JBD algorithm [5], and three non-unitary algorithms: the closed form algorithm, utilizing only the first two matrices, labelled as CFA [10], the algorithm of Ghennioui et al, labelled as GH, [6], and the nonlinear conjugent gradient (NCG) of Nion [10]. The random noise matrices $\mathbf{N}_i$ were taken as symmetric with zero-mean entries, Gaussian-distributed with variance $10^{-\mathrm{SNR}/10}$. The average of the angular error is taken with respect to the four block and over 10 independent trials (with newly generated blocks and the noise, and the same mixing matrix $\mathbf{A}$). We consider both the case of orthogonal (Fig.1(a)) and non-orthogonal (Fig.1(b)) $\mathbf{A}$. We note that JBD (which assumes orthogonality) performs best in the former but fails in the latter. Among the AJD-based methods, U-WEDGE based AJBD usually attains the best results for moderate SNR's. It is outperformed by NCG, when the SNR is high. The worse performance of NCG at low SNR is probably due to getting the algorithm stacked in side local minima. Note a huge difference in computation speed. While one run of NCG takes cca 90 s, one run of U-WEDGE takes about 0.01 s of matlab running time on an ordinary PC with a 3GHz processor.



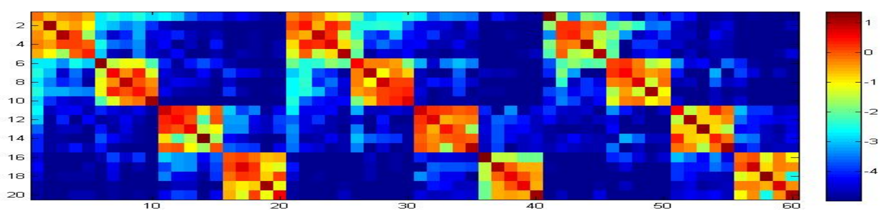
(a)



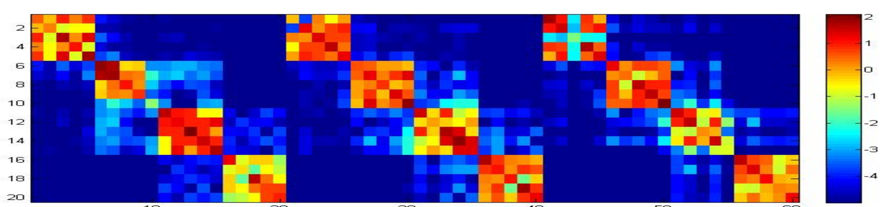
(b)



(c)



(d)



(e)

Fig. 2. Original and demixed matrices, displayed as $\log_{10}(|\mathbf{M}_i| + 10^{-5})$: (a) Original block-diagonal matrices (b) the matrices after mixing and de-mixing by U-WEDGE (c) the matrices after sorting row and columns (d) result for QRJ2D (e) result for LLAJD

5 Conclusions

We have shown theoretically and demonstrated in simulations that in the context of AJD-based AJBD, U-WEDGE attains an exact solution in the unperturbed case (with general mixing matrices), and usually performs better than other AJD algorithms in the perturbed case. The paper gives an explanation why the BG-WEDGE algorithm (which is similar) works so well in the time domain blind audio source separation [8].

Acknowledgements. The authors wish to give thanks to Dr. Dimitri Nion for sending them a matlab code of his algorithm JBD-NCG.

References

1. Afsari, B.: Sensitivity analysis for the problem of matrix joint diagonalization. SIMAX 30(3), 1148–1171 (2008)
2. Afsari, B.: Simple LU and QR Based Non-orthogonal Matrix Joint Diagonalization. In: Rosca, J.P., Erdogmus, D., Príncipe, J.C., Haykin, S. (eds.) ICA 2006. LNCS, vol. 3889, pp. 1–7. Springer, Heidelberg (2006)
3. Abed-Meraim, K., Belouchrani, A.: Algorithms for Joint Block Diagonalization. In: Proc. of EUSIPCO 2004, Vienna, Austria, pp. 209–212 (2004)
4. de Lathauwer, L.: Decomposition of higher-order tensor in block terms - Part II: definitions and uniqueness. SIAM J. Matrix Anal. and Appl. 30(3), 1033–1066 (2008)
5. Févotte, C., Theis, F.J.: Pivot Selection Strategies in Jacobi Joint Block-Diagonalization. In: Davies, M.E., James, C.J., Abdallah, S.A., Plumbley, M.D. (eds.) ICA 2007. LNCS, vol. 4666, pp. 177–184. Springer, Heidelberg (2007)
6. Ghennioui, H., et al.: A Nonunitary Joint Block Diagonalization Algorithm for Blind Separation of Convolutive Mixtures of Sources. IEEE Signal Processing Letters 14(11), 860–863 (2007)
7. Koldovský, Z., Tichavský, P.: A Comparison of Independent Component and Independent Subspace Analysis Algorithms. In: EUSIPCO 2009, Glasgow, Scotland, April 24-28, pp. 1447–1451 (2009)
8. Koldovský, Z., Tichavský, P.: Time-domain blind separation of audio sources based on a complete ICA decomposition of an observation space. IEEE Tr. Audio, Speech, and Language Processing 19(2), 406–416 (2011)
9. Li, X.-L., Zhang, X.D.: Nonorthogonal joint diagonalization free of degenerate solutions. IEEE Tr. Signal Processing 55(5), 1803–1814 (2007)
10. Nion, D.: A Tensor Framework for Nonunitary Joint Block Diagonalization. IEEE Tr. Signal Processing 59(10), 4585–4594 (2011)
11. Pham, D.-T.: Joint approximate diagonalization of positive definite Hermitian matrices. SIAM J. Matrix Anal. and Appl. 22(4), 1136–1152 (2001)
12. Szabó, Z., Póczos, B., Lörincz, A.: Separation theorem for independent subspace analysis and its consequences. Pattern Recognition 45(4), 1782–1791 (2012)

13. Tichavský, P.: Matlab code for U-WEDGE, WEDGE, BG-WEDGE and WASOBI, `http://si.utia.cas.cz/Tichavsky.html`
14. Tichavský, P., Yeredor, A.: Fast Approximate Joint Diagonalization Incorporating Weight Matrices. IEEE Tr. Signal Processing 57(3), 878–891 (2009)
15. Vollgraf, R., Obermayer, K.: Quadratic optimization for simultaneous matrix diagonalization. IEEE Tr. Signal Processing 54(9), 3270–3278 (2006)
16. Ziehe, A., Laskov, P., Nolte, G., Müller, K.-R.: A Fast Algorithm for Joint Diagonalization with Non-orthogonal Transformations and its application to Blind Source Separation. Journal of Machine Learning Research 5, 777–800 (2004)