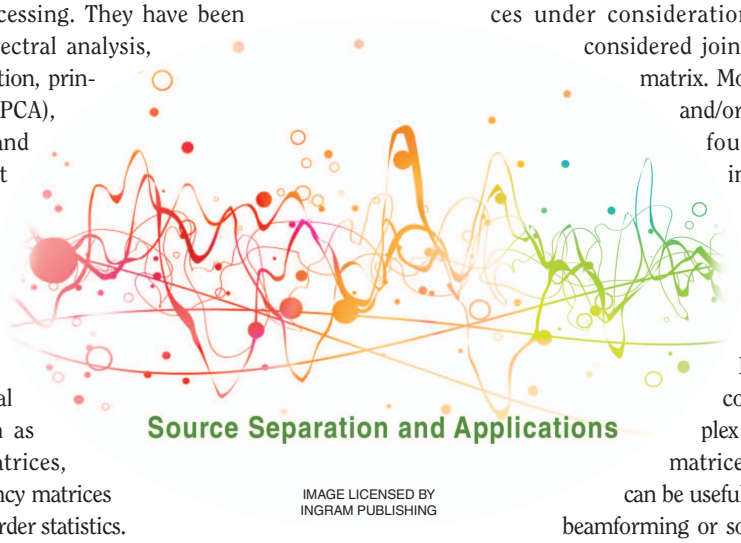
[Gilles Chabriel, Martin Kleinsteuber, Eric Moreau,
Hao Shen, Petr Tichavsky, and Arie Yeredor]

Joint Matrices Decompositions and Blind Source Separation

[A survey of methods, identification, and applications]

Matrix decompositions such as the eigenvalue decomposition (EVD) or the singular value decomposition (SVD) have a long history in signal processing. They have been used in spectral analysis, signal/noise subspace estimation, principal component analysis (PCA), dimensionality reduction, and whitening in independent component analysis (ICA). Very often, the matrix under consideration is the covariance matrix of some observation signals. However, many other kinds of matrices can be encountered in signal processing problems, such as time-lagged covariance matrices, quadratic spatial time-frequency matrices [21], and matrices of higher-order statistics.



In concert with this diversity, the joint diagonalization (JD) or approximate JD (AJD) of a set of matrices has been recently recognized to be instrumental in signal processing, mainly because of its importance in practical signal processing problems such as source separation, blind beamforming, image denoising, blind channel identification for multiple-input, multiple-output (MIMO) telecommunication system, Doppler-shifted echo extraction in radar, and ICA. Perhaps one of the first such algorithms is the joint approximate diagonalization of eigenmatrices (JADE) algorithm proposed in [8]. In this algorithm, the matrices under consideration are Hermitian and the considered joint diagonalizer is a unitary matrix. More recently, generalizations and/or new decompositions were found to be of considerable interest. They concern new sets of matrices, a nonunitary joint diagonalizer, and new decompositions.

INTRODUCTION

In the context of noncircular complex-valued signals, complex symmetric (non-Hermitian) matrices provide information that can be useful and even sufficient for blind beamforming or source separation. One example is the complementary covariance matrix, also called the *pseudocovariance matrix*. With such complex symmetric matrices, one ends up with jointly diagonalizing a set of matrices via either the transpose congruence transform or Hermitian congruence transform. For the special two-matrix case with one Hermitian and one complex symmetric matrix, there are particularly fast JD algorithms based on EVD and SVD.

This article provides a comprehensive survey of matrix joint decomposition techniques in the context of source separation. More precisely, we first intend to elaborate upon the signal models leading to different useful sets of matrices and their

Digital Object Identifier 10.1109/MSP.2014.2298045
Date of publication: 7 April 2014

1053-5888/14/$31.00©2014IEEE

joint decompositions. Second, we present recent identifiability results and algorithms in distinguishing important classes.

SIGNAL MODEL, MATRICES, AND DECOMPOSITIONS

To motivate the JD problem from the perspective of blind source separation (BSS), let us consider the classical linear memoryless source mixing model with additive noise described by

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t), \tag{1}$$

where $\mathbf{x}(t) \in \mathbb{C}^M$ is the observation vector, $\mathbf{s}(t) \in \mathbb{C}^N$ the source (component) vector, $\mathbf{n}(t) \in \mathbb{C}^M$ the noise vector, and $\mathbf{A} \in \mathbb{C}^{M\times N}$ the mixing matrix assumed full column rank (we shall not address the underdetermined case of fewer rows than columns until the section on direct fit methods). Since we suppose that we have more than one source, this leads to $M \geq N \geq 2$. The index t characterizes the variability of the signals. It is very often the time index but it can be the frequency index or the position index for an image, or any physical variables describing the considered signals. For convenience, it is considered in the sequel as the time index.

IN CONCERT WITH THIS DIVERSITY, THE JOINT DIAGONALIZATION OR APPROXIMATE JOINT DIAGONALIZATION OF A SET OF MATRICES HAS BEEN RECENTLY RECOGNIZED TO BE INSTRUMENTAL IN SIGNAL PROCESSING.

In BSS, the mixing matrix is assumed unknown and the sources not observable. The problem is then the estimation of the sources given only the observations.

When statistical (or other) information is available regarding the noise, such information can be accounted for in the estimation of the unknown mixing matrix, as well as in the estimation of the sources (even when the mixing matrix is known). However, to capture the essence of the problem and of its links to JD, we shall ignore the noise in here and assume $\mathbf{n}(t) = \mathbf{0}$.

Since the mixing model is not unique, it is well known that estimation of the sources is possible only up to some indeterminacies about the sources' scaling and ordering (see the section "Identifiability Issues for the Symmetric Case"). Among other things, this can be done by estimating a (left) pseudo-inverse (or simply the inverse in the square case) of $\mathbf{A}$ denoted (generically) by $\mathbf{B}$. Basically there are two ways for that: the first one consists of estimating $\mathbf{A}$, followed by the calculation of its pseudo-inverse whereas the second one consists of estimating $\mathbf{B}$ directly. Notice that the estimation of $\mathbf{A}$ corresponds to the so-called blind identification problem in signal processing while the direct estimation of $\mathbf{B}$ corresponds to the classical BSS.

The estimation of $\mathbf{A}$ or $\mathbf{B}$ can be formulated as a joint decomposition of a set of well-chosen matrices, to which we shall refer as *target-matrices*. Hence the first step is to choose target-matrices admitting a specific decomposition with respect to (w.r.t.) the matrix for which we are looking. The choice of useful matrices depends on a source model.

Quite commonly, the target-matrices are constructed from statistics of the observation. It is common practice to assume that the sources have zero mean, hence first-order statistics are of no interest. Thus second-order statistics (SOS) is considered. For a complex-valued random observation vector, one can define two kinds of SOS matrices,

$$\mathbf{R}_x(t,\tau) = \mathrm{E}\{\mathbf{x}(t)\mathbf{x}^H(t-\tau)\}, \quad \widetilde{\mathbf{R}}_x(t,\tau) = \mathrm{E}\{\mathbf{x}(t)\mathbf{x}^T(t-\tau)\},$$

where $(\cdot)^T$ and $(\cdot)^H$ are the transpose operator and the transpose conjugate operator, respectively, and $\mathrm{E}\{\cdot\}$ denotes the expectation operator. The first matrix $\mathbf{R}_x(t,\tau)$ is the classical correlation matrix, whereas the second one $\widetilde{\mathbf{R}}_x(t,\tau)$ is the so-called complementary correlation matrix. The usefulness of the complementary correlation matrix is directly related to a noncircularity property of the sources since for circular sources this matrix would be null.

One can also consider higher-order statistics (HOS) described by cumulants. Since third-order statistics are not so useful in practice mainly because the probability density function (PDF) of the sources is often close to symmetric, fourth-order statistics are often considered. In a very general way, they are defined as

$$C_{x,\mathrm{ijkl}}(t,\{\tau\}) = \mathrm{Cum}\{x_i(t), x_j^{(*)_1}(t-\tau_1), x_k(t-\tau_2), x_l^{(*)_2}(t-\tau_3)\},$$

where $(*)_1$ and $(*)_2$ denote optional complex conjugates and $\{\tau\} \equiv \{\tau_1, \tau_2, \tau_3\}$. One way to construct matrices from cumulants consists of considering a linear combination of the above cumulants while keeping free the first two indices that will be used as row and column indices for the constructed matrix. This is written as

$$(\mathbf{C}_x(t,\{\tau\}))_{\mathrm{ij}} = \sum_{k,l=1}^{M} G_{\mathrm{kl}} C_{x,\mathrm{ijkl}}(t,\{\tau\}),$$

where $\mathbf{G} = (G_{kl})$ is a fixed coefficients matrix. All of the above statistics generally depend on the time index t. In such a case the sources are called *nonstationary*. In the special case where the dependence w.r.t. t is periodic, the sources are called *cyclostationary*. When the statistics do not depend on t, the sources are called *stationary*.

In the noiseless case, using (1), the matrices $\mathbf{R}_x(t,\tau)$ and $\mathbf{C}_x(t,\{\tau\})$ with $(*)_1 \equiv *$, denoted generically as $\mathbf{M}_x$, all admit the factorization $\mathbf{M}_x = \mathbf{A}\mathbf{M}_s\mathbf{A}^H$, whereas the matrices $\widetilde{\mathbf{R}}_x(t,\tau)$ and $\mathbf{C}_x(t,\{\tau\})$ with $(*)_1 \equiv 1$, denoted generically as $\widetilde{\mathbf{M}}_x$, all admit the factorization $\widetilde{\mathbf{M}}_x = \mathbf{A}\widetilde{\mathbf{M}}_s\mathbf{A}^T$. With no further assumptions regarding the sources, the matrices $\mathbf{M}_s$ and $\widetilde{\mathbf{M}}_s$ do not possess any special algebraical structures compared to $\mathbf{M}_x$ and $\widetilde{\mathbf{M}}_x$, so these decompositions are noninformative. However, quite often some plausible assumptions regarding certain properties of the sources imply a special and "simplified" structure (diagonal or other) of $\mathbf{M}_s$ and $\widetilde{\mathbf{M}}_s$. This is directly linked to an identifiability property that has to be considered to be able to separate the sources. Basically, for stochastic sources, the classical identifiability assumption is their statistical independence, leading to the ICA problem. For independent

sources, the matrices $\mathbf{M}_s$ and $\widetilde{\mathbf{M}}_s$ are always diagonal [5], [8], [15], [20], [29], giving rise to the concept of "JD" of the selected target-matrices, as the result of representing each of these matrices as the corresponding transformation of a respective diagonal matrix.

In practice, however, the set of "true" target-matrices (specifically, the respective true SOS or HOS of the observations) is not available. Only sample-estimates of these matrices may be available, and these estimated matrices may no longer admit an exact JD transformation. In such cases one must resort to AJD, in an attempt to find a transformation being "as close as possible" to JD, with various measures for the quality of the approximation.

Simplifying the notations, we can always consider a set of K complex matrices $\mathbf{M}_k$, to be decomposed as

$$\mathbf{M}_k = \mathbf{A}\mathbf{D}_k\mathbf{A}^{\ddagger} + \mathbf{\Pi}_k \qquad k = 1,\ldots,K, \tag{2}$$

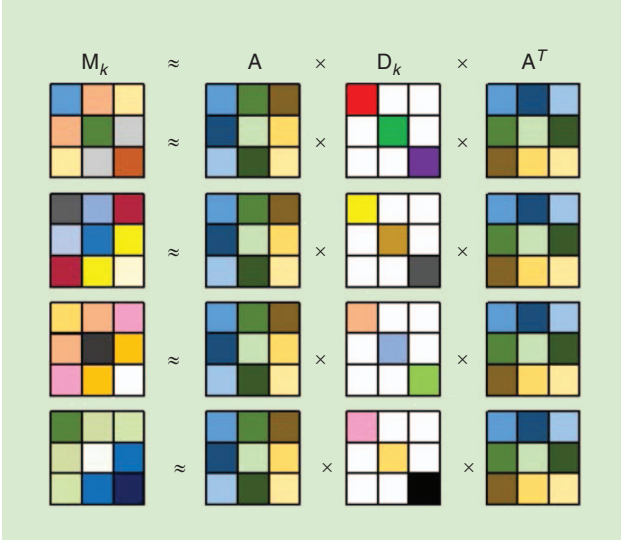
where $(\cdot)^{\ddagger}$ corresponds to either the transpose or the conjugate transpose of the matrix argument and matrices $\mathbf{D}_k$ all share some prescribed common structure. Depending on the signal model, the matrices $\mathbf{D}_k$ can be either all diagonal, all block diagonal, or all zero diagonal, as we shall explain in the sequel. The residual matrices $\mathbf{\Pi}_k$ are perturbation matrices which are linked to estimation errors and/or to modeling errors. This is referred to as the *symmetric case*; see Figure 1. Note that another model $\mathbf{M}_k = \mathbf{A}\mathbf{D}_k\mathbf{A}^{-1} + \mathbf{\Pi}_k$ has been studied as well [24] but is less popular in applications. A more general formulation, which is sometimes found to be more useful, reads

$$\mathbf{M}_k = \mathbf{A}_L\mathbf{D}_k\mathbf{A}_R + \mathbf{\Pi}_k, \tag{3}$$

where the matrices $\mathbf{A}_L$ and $\mathbf{A}_R$ are a priori arbitrary; see, e.g., [12]. This is referred to as the *nonsymmetric case* since $\mathbf{A}_R$ is not directly (or explicitly) linked to $\mathbf{A}_L$.

The main problem consists of estimating $\mathbf{A}$ (or $\mathbf{A}_L$) or its left inverse up to acceptable indeterminacies. In practice, these indeterminacies correspond to the estimation of all columns of $\mathbf{A}$ up to a scaling factor and up to ordering. This is the concept of essential uniqueness, which will be discussed in the section "Identifiability Issues for the Symmetric Case."

In all of the following sections, we denote $\mathbf{B} = \mathbf{A}^{\dagger}$, $\mathbf{B}_R = \mathbf{A}_R^{\dagger}$, and $\mathbf{B}_L = \mathbf{A}_L^{\dagger}$ where $(\cdot)^{\dagger}$ stands for the pseudo-inverse of the matrix argument or directly the inverse in the square case.



[FIG1] An illustration of AJD of four 3×3 symmetric target-matrices $\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3, \mathbf{M}_4$ ($N = 3, K = 4$).

IDENTIFIABILITY ISSUES FOR THE SYMMETRIC CASE

One fundamental question in the context of the BSS problem is: "Under what conditions on the sources can the mixing process be uniquely identified up to ordering and scaling?" This is evidently a question of general identifiability conditions, which are independent of a particular separation approach, and have been derived, e.g., in [15], [30], and [41]. However, in the context of JD-based BSS, the identifiability issue is closely related to the uniqueness (up to the trivial ambiguities) of the JD solution, which in turn relies on properties of the target-matrices. An underlying assumption is that under asymptotic conditions the estimated target-matrices can become arbitrarily close to the true target-matrices, and therefore the uniqueness of the joint diagonalizer has to be explored w.r.t. the true target-matrices, in the context of exact, rather than approximate JD. When the mixing matrix is invertible, identifiability of the mixing matrix implies the ability to separate the sources and is therefore often associated with separability. However, even when it is not invertible, the mixing matrix may still be identifiable (even by AJD), but such identifiability would not imply separability of the sources in such cases. Additionally, in some scenarios that are beyond the scope of this article, some sources may be separable from the mixture based on their special key properties (e.g., sparsity) but still without the need for identifiability of the full mixing matrix. In this section, we only focus on the symmetric case with an invertible mixing matrix.

Identifiability conditions for some specific scenarios have been provided, e.g., for the unitary case [5] and for the nonorthogonal real-valued case [2]. In this section, we summarize the necessary and sufficient conditions for the joint diagonalizer to be unique up to permutation and scaling for the noiseless, symmetric JD case (2). While general identifiability conditions for the nonsymmetric case (3) are still an open question, for particular nonsymmetric algorithms [12], [13] a necessary and sufficient condition can be provided (see the sections "Nonunitary Joint Diagnalization" and "Nonunitary Joint Zero Diagonalization," respectively, for the diagonal and zero-diagonal cases).

In the noiseless case, and under the assumption of full column rank of $\mathbf{A}$, whenever $M > N$ one can easily find N of the M observed mixtures that would be linearly independent and ignore the other observed mixtures without loss of information. Therefore, without loss of generality, we consider the square (or "determined") BSS problem, i.e., $M = N$. Given $\mathbf{B}, \mathbf{B}' \in \mathbb{C}^{N\times N}$, $\mathbf{B}$ is said to be essentially equivalent to $\mathbf{B}'$, and vice versa, if $\mathbf{B}$ is only different from $\mathbf{B}'$ by at most a row-wise ordering and scaling. Moreover, we say that the solution of a JD problem is essentially unique, if all solutions are essentially equivalent.

HOMOGENEOUS MODELS AND THEIR UNIQUENESS RESULTS

The uniqueness of JD under transpose congruence transform has been considered for real matrices in [2]. Recently, the extension to complex matrices has been studied in [3] and [25]. We first discuss the case where all matrices are to be diagonalized with the same transformation [i.e., ‡ is exclusively either T or H in (2)] and refer to this as the *homogeneous case*. For that, we require a measure of collinearity for diagonal matrices, which is obtained by means of the complex angle between the vectors formed by stacking the entries at corresponding positions together. The relation in (4) illustrates an example with (2×2) diagonal matrices. Let $\mathbf{D}_k := \operatorname{diag}(d_{k1},...,d_{kN}) \in \mathbb{C}^{N\times N}$ for $k = 1,\ldots,K$. For a fixed diagonal position i, we denote by $\mathbf{d}_i := [d_{1i},\ldots,d_{Ki}]^T \in \mathbb{C}^K$ the vector consisting of the ith diagonal element of each matrix, respectively,

$$\underbrace{\begin{bmatrix} d_{11} & 0 \\ 0 & d_{12} \end{bmatrix}}_{\mathbf{D}_1}, \underbrace{\begin{bmatrix} d_{21} & 0 \\ 0 & d_{22} \end{bmatrix}}_{\mathbf{D}_2}, \ldots, \underbrace{\begin{bmatrix} d_{K1} & 0 \\ 0 & d_{K2} \end{bmatrix}}_{\mathbf{D}_K} \Rightarrow \begin{matrix} [d_{11}, d_{21}, \ldots, d_{K1}]^T =: \mathbf{d}_1 \\ [d_{12}, d_{22}, \ldots, d_{K2}]^T =: \mathbf{d}_2 \end{matrix}. \tag{4}$$

Recall that the cosine of the complex angle between two nonzero vectors $\mathbf{v}, \mathbf{w} \in \mathbb{C}^K$ computed as $c(\mathbf{v},\mathbf{w}) := (\mathbf{v}^H\mathbf{w})/(\|\mathbf{v}\|\,\|\mathbf{w}\|)$, where $\|\mathbf{v}\|$ denotes the Euclidean norm of a vector $\mathbf{v}$. (If one of the two vectors is zero, the cosine is defined to be one by convention.) The uniqueness result states that, for a given set of matrices $\mathbf{M}_k$, the joint diagonalizer $\mathbf{B}$ is essentially unique, if and only if $|c(\mathbf{d}_i,\mathbf{d}_j)| \neq 1$ for all pairs (i, j) with $i \neq j$. In particular, for $K = 2$, this condition allows to uniquely solve the JD problem simply via a generalized EVD approach, i.e., $\mathbf{M}_1\mathbf{B} = \mathbf{M}_2\mathbf{B}\mathbf{\Lambda}$, where $\mathbf{\Lambda}$ is diagonal [33].

A HYBRID MODEL AND ITS UNIQUENESS RESULTS

The uniqueness results above state that, when there exists one pair of collinear concatenated vectors $(\mathbf{d}_i,\mathbf{d}_j)$, the solutions under homogeneous transforms are not essentially unique. However, it is known that signals with distinct second-order circularity coefficients are uniquely identifiable via a nonhomogeneous JD of only one covariance matrix (using the conjugate transpose operator $(\cdot)^H$) and one pseudo-covariance matrix (using the transpose operator $(\cdot)^T$). The corresponding method is known as strong uncorrelating transform (SUT) [20].

Recent works in [47] and [39] generalize the SUT approach to jointly diagonalize both Hermitian and complex symmetric matrices. The following statement provides a necessary and sufficient condition for the JD problem with a mixture of Hermitian congruence and transpose congruence. For given matrices $\mathbf{M}_k := \mathbf{A}\mathbf{D}_k\mathbf{A}^H$ with $k = 1,\ldots,K$ and $\widetilde{\mathbf{M}}_l := \mathbf{A}\widetilde{\mathbf{D}}_l\mathbf{A}^T$ with $l = 1,\ldots,L$, the common joint diagonalizer $\mathbf{B}$ is essentially unique, if and only if there exists no pair (i, j) with $i \neq j$, such that the following two conditions hold:

$$1)\ |c(\mathbf{d}_i,\mathbf{d}_j)| = |c(\widetilde{\mathbf{d}}_i,\widetilde{\mathbf{d}}_j)| = 1; \qquad 2)\ \|\mathbf{d}_i\|\,\|\mathbf{d}_j\| = \|\widetilde{\mathbf{d}}_i\|\,\|\widetilde{\mathbf{d}}_j\|.$$

In other words, when there is at least one pair of collinear concatenated vectors $(\mathbf{d}_i,\mathbf{d}_j)$, then the essential uniqueness implies that the respective norms are not proportional.

In the simplest case, where only one Hermitian and one complex symmetric matrix are considered, $\mathbf{d}_i$, $\mathbf{d}_j$, $\widetilde{\mathbf{d}}_i$, and $\widetilde{\mathbf{d}}_j$ are all scalars, so all pairs are trivially collinear. Then the previous result boils down to the following. Given two matrices $\mathbf{M} := \mathbf{A}\mathbf{D}\mathbf{A}^H$ and $\widetilde{\mathbf{M}} := \mathbf{A}\widetilde{\mathbf{D}}\mathbf{A}^T$ with $\mathbf{D} = \operatorname{diag}(d_1,\ldots,d_M)$ and $\widetilde{\mathbf{D}} = \operatorname{diag}(\tilde{d}_1,\ldots,\tilde{d}_M)$, the joint diagonalizer $\mathbf{B}$ is essentially unique if and only if the condition $|d_i||d_j| \neq |\tilde{d}_i||\tilde{d}_j|$ holds for all pairs (i, j) with $i \neq j$. This result simply recovers the uniqueness condition for SUT, where the matrix $\mathbf{M}$ is Hermitian and positive definite. We refer to [25] for a study of a further generalization of SUT, known as the pseudo-uncorrelating transform (PUT) and to [52] for the separation performance of the SUT for specific signal models.

The identifiability results yield a sufficient theoretical condition on the properties of the sources, such that the BSS problem is uniquely solvable, independent of any JD algorithms. Meanwhile, depending on the properties of the sources, it allows to determine a set of matrices, such that an exact JD solution yields the correct demixing matrix. In the presence of noise, AJD algorithms are used to find a matrix that minimizes some diagonality measures.

MATRIX NORMALIZATION FOR JOINT DIAGONALIZATION

For simplicity, let us only consider the symmetric case. A normalizing linear transformation $\mathbf{B}_\mathrm{n}$ can be applied to the observations as $\mathbf{x}_\mathrm{n}(t) = \mathbf{B}_\mathrm{n}\mathbf{x}(t)$ or directly onto the set of target-matrices as

$$\mathbf{M}_{\mathrm{n},k} = \mathbf{B}_\mathrm{n}\mathbf{M}_k\mathbf{B}_\mathrm{n}^H \ \text{ and/or } \ \widetilde{\mathbf{M}}_{\mathrm{n},k} = \mathbf{B}_\mathrm{n}\widetilde{\mathbf{M}}_k\mathbf{B}_\mathrm{n}^T, \quad k = 1,\ldots,K,$$

(here the subscript $(\cdot)_\mathrm{n}$ denotes "normalization") in such a way that the overall problem is normalized or simplified. The normalizing matrix $\mathbf{B}_\mathrm{n}$ is usually determined by selecting a particular Hermitian matrix denoted $\mathbf{M}_0$, which would be exactly diagonalized by this transformation, and may (or may not) coincide with one of the target-matrices $\mathbf{M}_1,\ldots,\mathbf{M}_K$. It is well known [23] that any such matrix admits diagonalization as $\mathbf{M}_0 = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, where $\mathbf{\Lambda}$ is a real-valued diagonal matrix of eigenvalues and $\mathbf{U}$ is a unitary matrix of orthonormal eigenvectors in its columns, i.e., $\mathbf{U}\mathbf{U}^H = \mathbf{U}^H\mathbf{U} = \mathbf{I}$ where $\mathbf{I}$ is the identity matrix. In the context of BSS, and especially when $\mathbf{M}_0$ coincides with one of the target-matrices, the matrix $\mathbf{U}$ can often serve as a reasonable initial guess for the approximate joint diagonalizer of the entire set, or can at least serve to "simplify" the matrix set by considering $\mathbf{B}_\mathrm{n} = \mathbf{U}^H$.

Now, if the number of sources N is known (or well estimated), then one can do a little more. If N eigenvalues of $\mathbf{M}_0$ are nonzero and all the $M-N$ others are zero, then we denote $\mathbf{\Lambda}_s$ the diagonal matrix corresponding to these N nonzero eigenvalues and $\mathbf{U}_s$ the matrix of corresponding eigenvectors (spanning the so-called signal subspace). Then we directly have $\mathbf{M}_0 = \mathbf{U}_s\mathbf{\Lambda}_s\mathbf{U}_s^H$. Now we can consider $\mathbf{B}_\mathrm{n} = \mathbf{U}_s^H$, which corresponds to a projection of the observations onto the signal subspace. Hence all new matrices

$\mathbf{M}_{n,k}$ and/or $\widetilde{\mathbf{M}}_{n,k}$ are of size $N \times N$. This is essentially a PCA operation, which corresponds to a useful dimension reduction when $M > N$.

Finally, we point out that $\mathbf{M}_0$ is often positive semidefinite, that is all nonzero components of $\mathbf{\Lambda}$ (specifically, the diagonal components of $\mathbf{\Lambda}_s$) are positive. This is usually the case when $\mathbf{M}_0$ is selected as the (zero lag) sample-covariance matrix of the observations. Then we can set $\mathbf{B}_n = \mathbf{V}\mathbf{\Lambda}_s^{-(1/2)}\mathbf{U}_s^H$, where $\mathbf{V}$ is any $N \times N$ unitary matrix. This operation is known as *whitening*, and it can be shown (as evident from the above definition using a nondetermined unitary matrix $\mathbf{V}$) that following such a whitening step, any unitary diagonalizer of the normalized ("whitened") set $\mathbf{M}_{n,1}, \ldots, \mathbf{M}_{n,K}$ would maintain the whiteness of the transformed $\mathbf{M}_0$. Therefore, when a whitening stage is used, the diagonalizer of the whitened set is usually constrained to be unitary, which simplifies the search.

IN MANY APPLICATIONS, PERFORMANCE OF THE NONUNITARY AJD ALGORITHMS CAN BE SIGNIFICANTLY ENHANCED BY APPROPRIATE WEIGHTING, INTRODUCED IN THE OPTIMIZATION CRITERION.

NONUNITARY JOINT DIAGONALIZATION

In this section, we address the nonunitary AJD problem as the most important and common case. Following the alternating columns, diagonal centers (ACDC) algorithm [50], many AJD algorithms have been proposed over the last decade; see, e.g., [13], [21], [38], [43], and [49]. These papers only consider the symmetric version with $\mathbf{B}_L = \mathbf{B}_R^T = \mathbf{B}$. Since, however, an extension to the nonsymmetric version is possible (straightforward for some of these algorithms), we present the problem in the latter form. In the existing literature, we can distinguish four groups of nonunitary AJD algorithms.

1) Minimizing the so-called indirect least-squares criterion, which may be a possibly weighted square norm of off-diagonal elements of the transformed matrices $\mathbf{B}_L\mathbf{M}_k\mathbf{B}_R$. To use this criterion the matrices $\mathbf{B}_L$ and $\mathbf{B}_R$ must be properly constrained so as to avoid the trivial zero solution and/or degenerate solutions.

2) Minimizing the direct least-square criterion (which can also be weighted), measuring the squared difference between the matrices and their representations, specifically the norms of the residual matrices Π_k in (2) or (3).

3) A combination of these two criteria. Here, one seeks matrices $\mathbf{B}_L$ and $\mathbf{B}_R$ that transform the given set of matrices into a set of nearly diagonal matrices, which cannot be diagonalized any further in the direct-fit sense, specifically such that the best direct-fit diagonalizer of the transformed set is the identity matrix.

4) Minimizers of an approximate log-likelihood criterion. So far the log-likelihood criterion was derived only for the case where the given matrices reflect second-order statistics of a mixture of Gaussian vector processes.

In many applications, performance of the nonunitary AJD algorithms can be significantly enhanced by appropriate weighting, introduced in the optimization criterion. When a statistical model for the sources is fully known, the optimal weighting may be deduced in advance (e.g., [51]). Usually, however, the proper weights are not known in advance but may be estimated from the observed data, e.g., when a statistical model for the sources is only known up to some parameters, which nonetheless can be estimated as a by-product of the diagonalization process (e.g., [42]), or when multiple snapshots of the data are available for nonparametric estimation of the weights [53].

MINIMIZING THE INDIRECT-FIT CRITERION

Historically, the first, natural choice of an indirect-fit criterion is

$$C_{J1}(\mathbf{B}_L, \mathbf{B}_R) = \sum_{k=1}^{K} \| \operatorname{Zdiag}\{\mathbf{B}_L\mathbf{M}_k\mathbf{B}_R\} \|_F^2, \tag{5}$$

where $\operatorname{Zdiag}\{\cdot\}$ sets the diagonal entries of the matrix argument to zero. Since, however, trivial minimization by downscaling towards $\mathbf{B}_L = \mathbf{B}_R = \mathbf{0}$ is clearly not interesting, one has to consider some constraint or barrier function to evade this. The following options have been proposed in the literature, together with appropriate minimization procedures, derived for the symmetric case where $\mathbf{B}_L = \mathbf{B}$ and $\mathbf{B}_R = \mathbf{B}^T$:

1) $\mathbf{B}$ is unitary. This choice has already been discussed in the previous section.

2) The rows of $\mathbf{B}$ have unit norm. This constraint is weaker than the former one and was used in [19].

3) $\mathbf{B}\mathbf{M}_0\mathbf{B}^T$ must have an all-ones main diagonal [48], where $\mathbf{M}_0$ may or may not be included among $\{\mathbf{M}_k\}$. In the BSS context, if $\mathbf{M}_0$ is the zero-lag covariance matrix of the observations, this constraint corresponds to the constraint on the separated sources, that they should have equal (unit) power. The proposed method of optimization uses iterative generalized matrix eigenvector computation.

4) $\mathbf{B}$ has a unit determinant. The optimization can be attained through Givens and hyperbolic rotations [40].

5) In [28], a penalty term (proportional to $\log|\det\mathbf{B}|$) is added to (5). The optimization proceeds by alternating between optimization w.r.t. individual rows of matrix $\mathbf{B}$.

6) Another suitable AJD criterion, which is scale-invariant in $\mathbf{B}$, was proposed in [1] and [2],

$$C_{J2}(\mathbf{B}) = \sum_{k=1}^{K} \| \mathbf{M}_k - \mathbf{B}^{-1}\operatorname{Diag}\{\mathbf{B}\mathbf{M}_k\mathbf{B}^T\}\mathbf{B}^{-T} \|_F^2 . \tag{6}$$

The scale invariance means that the criterion is not affected by changing scale of any rows of $\mathbf{B}$. The optimization was achieved by combination of triangular Jacobi matrices and Givens rotations.

MINIMIZING THE DIRECT-FIT CRITERION

The direct-fit criterion is a measure of difference between the given matrices $\mathbf{M}_k$ and their assumed model in terms of the

estimating the left and right mixing matrix $\mathbf{A}_L$ and $\mathbf{A}_R$ and diagonal matrices $\mathbf{D}_k$, $k = 1,\ldots,K$,

$$C_{\mathrm{J3}}(\mathbf{A}_L,\mathbf{A}_R,\{\mathbf{D}_k\}) = \sum_{k=1}^{K} \|\mathbf{M}_k - \mathbf{A}_L\mathbf{D}_k\mathbf{A}_R\|_F^2. \tag{7}$$

Minimization of this criterion is directly linked to tensor decompositions, as we explain shortly. The noiseless part $\mathbf{A}_L\mathbf{D}_k\mathbf{A}_R$ of the target-matrices $\mathbf{M}_k$ together represents a third-order tensor $\mathcal{T}$ of dimensions $M\times M\times K$, with elements $\mathcal{T}_{\mathrm{ijk}}$, $i, j = 1,\ldots,M$ and $k = 1,\ldots,K$ such that its kth slice $\mathcal{T}_{:,:,k}$ equals $\mathbf{A}_L\mathbf{D}_k\mathbf{A}_R$, i.e.,

$$\mathcal{T}_{\mathrm{ijk}} = \sum_{r=1}^{M} (\mathbf{A}_L)_{\mathrm{ir}}(\mathbf{A}_R)_{\mathrm{rj}}(\mathbf{D}_k)_{\mathrm{rr}}. \tag{8}$$

In the tensor terminology, $\mathcal{T}$ is a tensor of rank at most M, because it can be written as a sum of M rank-1 tensors, each of them being an outer product of three vectors, specifically the rth column of $\mathbf{A}_L$, rth row of $\mathbf{A}_R$, and a vector composed of the (r,r)th elements of $\mathbf{D}_k$, $k = 1,\ldots,k$. The decomposition of this kind is called *canonical polyadic* or *CANDECOMP-PARAFAC* (CP) decomposition [9], [22]. The special case when two or more factor matrices coincide (in our case, the coinciding factor matrices might be $\mathbf{A}_L$ and $\mathbf{A}_R^T$) is called *individual differences in scaling* (*INDSCAL*) [10] (see Figure 2 for an illustration).

The direct-fit criterion, or the CP decomposition, offers more flexibility than the indirect fit: it allows for treating situations where the number of separated sources is not necessarily equal to the dimension of the mixture. If the number of sources is smaller than M, it is still possible to use an indirect-fit criterion and identify the sources among spurious (noisy) ones. A less trivial task is to separate underdetermined mixtures, where the number of sources exceeds the number of mixtures. CP decomposition allows such a separation [16], [45].

The area of CP tensor decompositions is a rapidly growing field, and many techniques have been proposed. A traditional and still the most popular technique is the alternating least squares. Other methods include enhanced line search, damped Gauss–Newton method (also called Levenberg–Marquardt), and others; see, e.g., references in [36]. A link between the CP decomposition and AJD (even in the underdetermined case, rank greater than the dimension) was shown in [17], and more recently was exploited in [37]. It was shown that CP decomposition can be attained through approximate JD of certain sets of matrices.

Beside the CP decomposition approach, a suboptimum direct-fit optimization of (7) was proposed in [13] (called DIEM for "diagonalization using equivalent matrices"), offering a closed-form (noniterative) solution. Moreover, DIEM can deal with the nonsymmetric case since the matrices $\mathbf{A}_L$ and $\mathbf{A}_R$ are not constrained or linked in the derivations. A necessary and sufficient condition for the uniqueness of the DIEM solution is that the set of K underlying diagonal matrices $\mathbf{D}_1,\ldots,\mathbf{D}_K$ spans the N-dimensional subspace of diagonal matrices in $\mathbb{C}^{N\times N}$. New BSS applications using nonsymmetric JD are discussed in the last section.

COMBINATION OF THE DIRECT AND INDIRECT-FIT CRITERIA

Combining the indirect and direct fit of JD is conceptually simple and computationally efficient. A generic algorithm of this type works with a partially diagonalized set of matrices $\mathbf{B}_L^{[i]}\mathbf{M}_k\mathbf{B}_R^{[i]}$, $k = 1,\ldots,K$, i is the iteration index. Initially, one can start with $\mathbf{B}_L^{[0]} = \mathbf{B}_R^{[0]} = \mathbf{I}$. Each step consists of one iteration of a direct-fit procedure, which may or may not use weighting. In the unweighted (or uniformly weighted) version, we have

$$\{\mathbf{A}_L^{[i]}, \mathbf{A}_R^{[i]}\} = \underset{\mathbf{A}_L,\mathbf{A}_R}{\operatorname{argmin}} \sum_{k=1}^{K} \|\mathbf{B}_L^{[i]}\,\mathbf{M}_k\,\mathbf{B}_R^{[i]} - \mathbf{A}_L\,\mathbf{D}_{k,\mathrm{B}}\,\mathbf{A}_R\|_{\mathrm{F}}^2, \tag{9}$$

where $\mathbf{D}_{k,\mathrm{B}} \triangleq \mathrm{Diag}\{\mathbf{B}_L^{[i]}\mathbf{M}_k\mathbf{B}_R^{[i]}\}$. The direct-fit procedure can be of Gauss–Newton type, for fast convergence in a neighborhood of the true local minimum and is sought close to $\mathbf{A}_L = \mathbf{A}_R = \mathbf{I}$. Only one iteration of the Gauss–Newton procedure is applied in each step because, at the initial point $\mathbf{A}_L = \mathbf{A}_R = \mathbf{I}$, the Hessian matrix has an attractive decoupled form that enables its inversion through solving distinct sets of 2×2 linear equations.

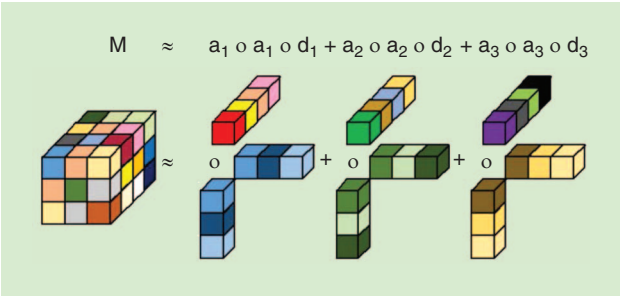
Once an approximation of the best fitting mixing matrices $\mathbf{A}_L^{[i]}$ and $\mathbf{A}_R^{[i]}$ is found, the estimated demixing matrices are updated as $\mathbf{B}_L^{[i+1]} = (\mathbf{A}_L^{[i]})^{-1}\mathbf{B}_L^{[i]}$ and $\mathbf{B}_R^{[i+1]} = \mathbf{B}_R^{[i]}\left(\mathbf{A}_R^{[i]}\right)^{-1}$. This algorithm was named WEDGE (for "weighted exhaustive diagonalization with Gauss iteration"), or U-WEDGE in its uniformly weighted version, in [43].

MAXIMIZATION OF A LOG-LIKELIHOOD CRITERION

The last principle of AJD is a maximum likelihood (ML) approach. It was developed by Pham [35] for JD of a set of sample covariance matrices taken from distinct signal-blocks, where the statistical model assumes independent Gaussian distributed sources with variances that are constant within each block but varying between blocks. For real-valued signals and mixtures, the ML method with $\mathbf{B} = \mathbf{B}_L = \mathbf{B}_R^T$ leads to the criterion

$$C_{\mathrm{LL}}(\mathbf{B}) = \sum_{k=1}^{K} \log\frac{\det\,\mathrm{Diag}\{\mathbf{B}\mathbf{M}_k\mathbf{B}^T\}}{\det(\mathbf{B}\mathbf{M}_k\mathbf{B}^T)}, \tag{10}$$

which is scale-invariant in $\mathbf{B}$ and does not require any constraints. This criterion may also be used as a generic AJD criterion (outside the ML framework), however it is meaningful only for positive definite target-matrices $\{\mathbf{M}_k\}$.



[FIG2] The AJD of the target-matrices of Figure 1 viewed as a partially symmetric CP decomposition (INDSCAL).

COMPUTATIONAL COMPLEXITY

The fastest nonorthogonal AJD algorithms such as U-WEDGE/WEDGE [43], fast AJD (FAJD) [28], and Pham's [35] have asymptotic computational complexity of $O(KM^2)$ operations per iteration. This is the lower bound for any algorithm that should have access to all elements of the target-matrices. Some other algorithms have slightly higher complexity, $O(KM^3)$ operations per iteration, such as quadratic AJD (QAJD) [48], Souloumiac's [40] or Afsari's [1]. The number of iterations is varying. Among the algorithms, U-WEDGE/WEDGE, Pham's, Souloumiac's, and Afsari's algorithm exhibit a quadratic convergence, as inherited from the approximate Gauss–Newton methods, and usually only need a few dozens of iterations to converge; FAJD and QAJD are based on alternating minimization, exhibit only linear convergence, and usually require hundreds of iterations.

For fixed and moderate K and M, a very fast AJD algorithm is the noniterative DIEM algorithm of Chabriel and Barrère [13] that, however, only attains an approximate optimum of the direct fit, and works with matrices of the size $M^2 \times M^2$, so that its complexity is at least $O(KM^6)$.

All direct-fit algorithms have complexity of at least $O(KM^3)$ per iteration because this is the complexity of one least-squares solution step (fixing two factor matrices and minimizing w.r.t. the third one). Indeed, more complex algorithms require a higher number of operations per iteration. For example, the fastest available implementation of the Levenberg–Marquardt algorithm has complexity $O(KM^3 + M^6)$ operations per iteration.

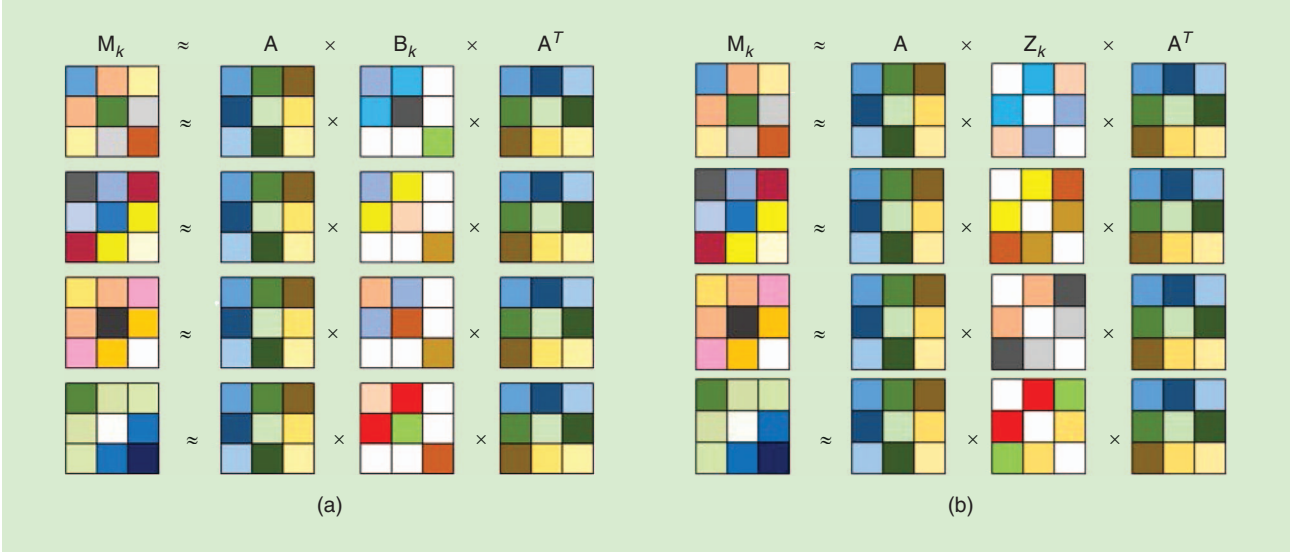
APPROXIMATE JOINT BLOCK DIAGONALIZATION

In this subsection, we briefly mention the concept of approximate joint block diagonalization that was first introduced in [54]. Indeed it might happen that for some given sets of target-matrices it is not possible to find mixing or demixing matrices such that the indirect or direct-fit error is satisfactorily small, but it is possible to fit them well by a block diagonal model. The latter model resembles (3), but the matrices $\mathbf{D}_k$ are block diagonal, with diagonal blocks of appropriate size; see Figure 3(a). Such a model is usually relevant in cases where not all sources are independent, but several groups of sources exist, with intragroup dependence but with intergroup independence. As in the ordinary diagonalization task, the block diagonalization can be either unitary or nonunitary. The first block-diagonalization algorithms were unitary, [7]. Later, nonunitary algorithms were proposed as well: direct-fit methods by Nion [31], indirect methods by Tichavsky et al. [44], and ML methods by Lahat et al. [27].

NONUNITARY JOINT ZERO DIAGONALIZATION

In this section, we consider the case where, in (2) or (3), the matrices $\mathbf{D}_k$ are zero diagonal for all k and where the searched matrix is a priori nonunitary; see Figure 3(b). This problem is termed *approximate zero diagonalization* (*AZD*). The matrices $\mathbf{D}_k$ for all k are denoted $\mathbf{Z}_k$ here for a direct interpretation. We consider both the symmetric and the nonsymmetric cases when all matrices are square $N \times N$. This can always be considered in using a first-dimension reduction operation.

In the symmetric case, the problem can be addressed by considering the indirect least squares criterion [21] $C(\mathbf{B}) = \sum_{k=1}^{K} \| \mathrm{Diag}\{\mathbf{B}\mathbf{M}_k\mathbf{B}^H\} \|^2$ that has to be minimized. As initially proposed in [21], the optimization of $C(\mathbf{B})$ can be performed row by row in searching iteratively for eigenvectors associated to matrices built from the target-matrices. Even if the optimization scheme is rather simple, it can lead to nonuseful solutions (certainly corresponding to local minima). However since JD algorithms are more robust, very interestingly, it can be shown that the above problem can be cast as an ordinary nonunitary JD problem [11]. This is possible when the set of K underlying zero-diagonal matrices $\mathbf{Z}_1, \ldots, \mathbf{Z}_K$ spans the $N^2 - N$ dimensional subspace of zero-diagonal matrices in $\mathbb{C}^{N \times N}$. In fact, this



[FIG3] (a) The approximate joint block diagonalization of four 3×3 symmetric target-matrices $\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3, \mathbf{M}_4$, with one 2×2 block and one 1×1 "block." (b) The approximate joint zero diagonalization of four 3×3 symmetric target-matrices $\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3, \mathbf{M}_4$.

condition is also a sufficient for essential uniqueness of the joint zero diagonalizer. Note that the number of target-matrices has to be relatively large, $K \geq N^2 - N$ for this condition to hold.

In the nonsymmetric case, matrices $\mathbf{A}_R$ and $\mathbf{A}_L$ are a priori not linked. Following [11], a nonsymmetric algorithm [12] can be derived in turning the problem into a nonsymmetric JD one that can be solved using, e.g., the nonsymmetric version of DIEM mentioned earlier.

EXAMPLES OF BSS APPLICATIONS

In this section, we mention examples of two applications of AJD-based BSS techniques, one for symmetric AJD and the other one for symmetric or nonsymmetric AJD. As an application of joint zero diagonalization, we can mention the zero-division multiple access wireless telecommunications system [12] where all the signals to be sent use the same bandwidth.

BLIND AUDIO SOURCE SEPARATION

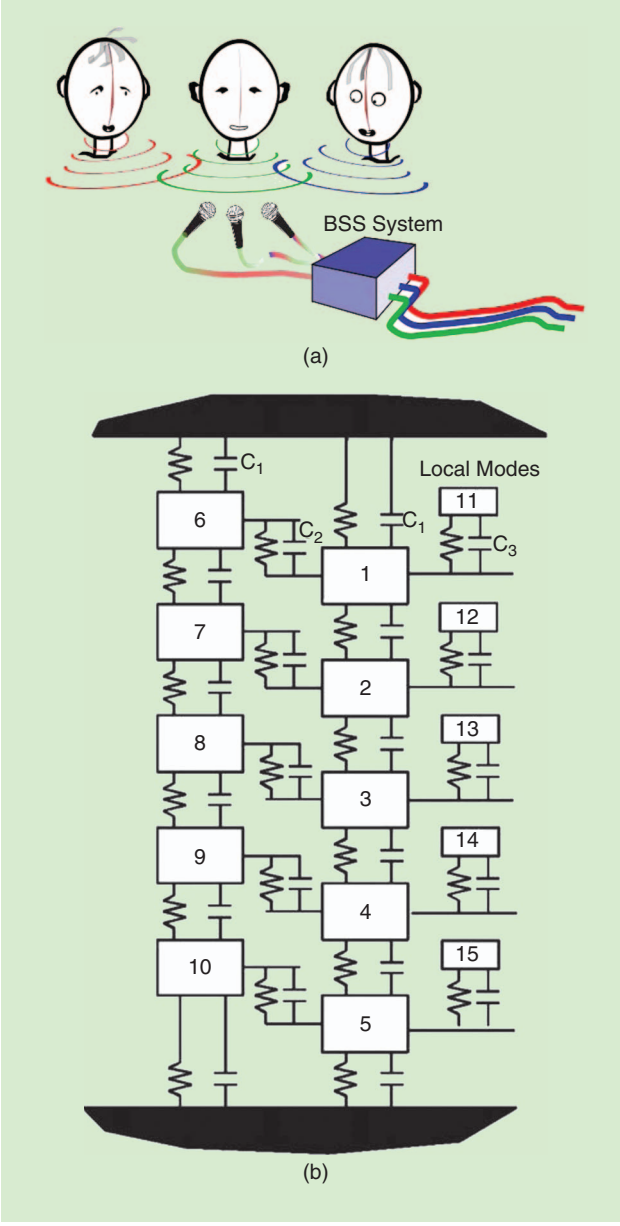
Since modern AJD algorithms allow the JD of sets of large matrices with dimensions such as 100×100 or 500×500 within time of order fraction of seconds or few seconds, they enable us to solve blind audio source separation (BASS), also known as the "cocktail-party problem"; see Figure 4(a), in the time domain [26].

Most of the existing BASS algorithms work in the frequency domain by transforming the convolutive mixture model into an instantaneous mixture model using the short-time Fourier transform. The individual sources were separated in each frequency bin independently. Since, however, the order of sources obtained in each bin is arbitrary, it is necessary to resolve the permutation ambiguity simultaneously in all the bins. Random errors in the estimated order of the components in different bins, which are inevitable in practice, lead to nonlinear distortion of the estimated signals.

Time-domain BSS methods do not produce nonlinear distortions in the data but estimate linear MIMO filters that separate the sources. In short, the input signals measured by the microphones are augmented by their time-shifted replicas to become a multidimensional input of an BSS algorithm. The number of the time-shifted replicas should be large enough to cover mutual time delays of arrival of the individual source signals at the microphones and their reflections. An insufficient number of the time replicas would lead to poor performance of the whole system.

In principle, it is possible to use any ICA algorithm to transform the input data set (microphone outputs with their time-shifted replicas), in "pseudo-independent" components; the "pseudo-independent" components are further grouped and used to reconstruct the source images (contributions of all sources at all microphones) [26]. A successful ICA algorithm in this application was the block Gaussian separation algorithm, which consists in applying a nonunitary AJD algorithm UWEDGE/WEDGE to the set of covariance matrices of the input signals at nonoverlapping time windows.

In [14], the cocktail-party problem is addressed differently, using a compact array of microphones. It is shown that if the distance between any pair of microphones is sufficiently small (relatively to the coherence time of each source), a linear instantaneous mixing model, as the one in (1), holds, but w.r.t. an extended mixture model, where the temporal derivatives of the sources are added as "additional" (pseudo-independent) sources.



[FIG4] (a) An Illustration of the cocktail-party problem with three speakers and three microphones. (b) An analytical 15 degrees-of-freedom system. (Figure reprinted from [4], with permission from Elsevier.)

OPERATIONAL MODAL ANALYSIS

Operational modal analysis (OMA) is concerned with the analysis of a mechanical or electrical vibration system in terms of individual vibration modes; see Figure 4(b). The analysis is based on the system output, assuming white input noise. It was shown in [4] that SOS-based BSS methods are able to separate a

set of system responses into modal coordinates from which the system poles can be extracted by single-degree-of-freedom techniques. In addition, these methods return a mixing matrix whose columns are the estimates of the system mode shapes.

The method is based on JD of the set of time-lagged covariance matrices of the observations. The authors of [4] considered the algorithm for multiple unknown signals extraction (AMUSE) (based on generalized eigendecomposition of a pair of the covariance matrices with lags 0 and $\tau \neq 0$) [46] and second-order blind identification [5] algorithms in forming the AJD problem. Next, they proposed their own nonsymmetric AJD algorithm, which was shown to be more adequate for this problem.

CONCLUSIONS

We presented a survey of AJD methods and related joint matrix decomposition methods that can be used in various BSS applications, together with conditions for uniqueness of the solutions. In addition, we pointed out the option of weighted AJD methods, which might yield optimized performance through proper selection of the weights. The selection of the most suitable AJD/AZD method will always depend on the target application because criteria of success might be quite different. Finally, we mentioned joint nonsymmetric matrix decompositions that should lead to new (promising) BSS or array processing applications.

ACKNOWLEDGMENT

The work of Martin Kleinsteuber and Hao Shen has been supported by the German DFG-funded Cluster of Excellence CoTeSys—Cognition for Technical Systems. The work of Petr Tichavsky was supported by the Czech Science Foundation through project 102/09/1278.

AUTHORS

Gilles Chabriel (gilles.chabriel@im2np.fr) received the microelectronic and automatic engineer diploma from the Institut des Sciences de l'Ing'enieur de Montpellier (currently Polytech Montpellier), Montpellier, France, in 1992 and the Ph.D. degree in physics from the Université de Toulon (UTLN), La Garde, France, in 1997. He is currently an assistant professor in the Sciences and Applied Sciences Department of UTLN, where he teaches courses in electronics and signal processing. In the Signals and Systems Team of the Institut Mat'eriaux, Micro'electronique, Nanosciences de Provence (IM2NP, UMR CNRS 7334), his research is in the areas of signal processing, telecommunications, and electronic systems.

Martin Kleinsteuber (kleinsteuber@tum.de) received the Ph.D. degree in mathematics from the University of Würzburg, Germany, in 2006. After postdoctoral positions at National ICT Australia Ltd, the Australian National University, Canberra, and the University of Würzburg, he was appointed assistant professor for geometric optimization and machine learning with the Department of Electrical Engineering and Information Technology, Technische Universität München, Germany, in 2009. He won the SIAM Student Paper Prize in 2004 and the Robert-Sauer-Award of the Bavarian Academy of Science in 2008 for his works on Jacobi-type methods on Lie algebras. His research focuses on latent variable analysis, pattern recognition, and feature extraction in signals together with their applications to audio and image processing.

Eric Moreau (moreau@univ-tln.fr) received the Ph.D. degree in signal processing from Université Paris 11 Orsay, France, in 1995 and the habilitation degree from the Université de Toulon, France, in 2000. He is currently a full professor with the Engineering School ISITV at Université de Toulon and head of the Telecommunications Department. He is also head of the research group SIIM (Signal and Images) with LSIS UMR CNRS 7296. His main research interests are in statistical signal processing and matrix/tensor decompositions with an emphasis on source separation.

Hao Shen (hao.shen@tum.de) received the bachelor's degree in mechanical engineering and applied mathematics from Xi'an Jiaotong University, China, in 2000, the master's degrees in computer studies and computer science from the University of Wollongong, Australia, in 2002 and 2004, respectively, and the Ph.D. degree in engineering from the Australian National University in 2008. He is a postdoctoral researcher at the Institute for Data Processing, Technische Universität München, Germany. His research interests focus on machine learning for signal processing, in particular, BSS, sparse representation, geometric optimization, and reinforcement learning.

Petr Tichavsky (tichavsk@utia.cas.cz) received the Ph.D. degree in theoretical cybernetics from the Czechoslovak Academy of Sciences in 1992. Since then, he has been with the Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic in Prague. He was an associate editor of *IEEE Signal Processing Letters* from 2002 to 2004 and an associate editor of *IEEE Transactions on Signal Processing* from 2005 to 2009 and from 2011–present. He has also served as a general cochair of the 2011 36th IEEE International Conference on Acoustics, Speech, and Signal Processing in Prague, Czech Republic. He is a Senior Member of the IEEE.

Arie Yeredor (arie@eng.tau.ac.il) received the B.Sc. (summa cum laude) and Ph.D. degrees in electrical engineering from Tel-Aviv University (TAU), Tel-Aviv, Israel, in 1984 and 1997, respectively. He is currently an associate professor with the School of Electrical Engineering at TAU, where his research and teaching areas are in statistical and digital signal processing. He was previously an associate editor of *IEEE Signal Processing Letters*, *IEEE Transactions on Circuits and Systems-Part II*, and *IEEE Transactions on Signal Processing*. He recently served as general cochair of the 2012 10th International Conference on Latent Variables Analysis and Signal Separation. He is a Senior Member of the IEEE.

REFERENCES

[1] B. Afsari, "Simple LU and QR based non-orthogonal matrix joint diagonalization," in *Proc. 6th Int. Conf. Independent Component Analysis and Blind Source Separation*, ICA 2006, LNCS 3889, Charleston, J. P. Rosca, D. Erdogmus, J. C. Prncipe, and S. Haykin, Eds. Springer Verlag: Berlin, Heidelberg, 2006, pp. 17–24.

[2] B. Afsari, "Sensitivity analysis for the problem of matrix joint diagonalization," *SIAM J. Matrix Anal. Applicat.*, vol. 30, no. 3, pp. 1148–1171, 2008.

[3] A. Aïssa-El-Bey, K. Abed-Meraim, Y. Grenier, and Y. Hua, "A general framework for second-order blind separation of stationary colored sources," *Signal Process.*, vol. 88, no. 9, pp. 2123–2137, Sept. 2008.

[4] J. Antoni and S. Chauhan, "A study and extension of second-order blind source separation to operational modal analysis," *J. Sound Vib.*, vol. 332, no. 4, pp. 1079–1106, Feb. 2013.

[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, Feb. 1997.

[6] A. Belouchrani, K. Abed-Meraim, M. G. Amin, and A. M. Zoubir, "Blind separation of non-stationary sources," *IEEE Signal Processing Lett.*, vol. 11, no. 7, pp. 605–608, July 2004.

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

[8] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *IEE Proc. Radar and Signal Processing F*, vol. 140, no. 6, pp. 362–370, Dec. 1993.

[9] J. D. Carroll, "Generalization of canonical correlation analysis to three or more sets of variables," in *Proc. 76th Annu. Conf. American Psychological Association*, vol. 3, pp. 227–228, 1968.

[10] J. D. Carroll and J.-J. Chang, "Analysis of individual differences in multidimensional scaling via an N-way generalization of Eckart-Young decomposition," *Psychometrika*, vol. 35, no. 3, pp. 283–319, Sept. 1970.

[11] G. Chabriel, J. Barrère, N. Thirion-Moreau, and E. Moreau, "Algebraic joint zero-diagonalization and blind sources separation," *IEEE Trans. Signal Processing*, vol. 56, no. 3, pp. 980–989, Mar. 2008.

[12] G. Chabriel and J. Barrère, "Non-symmetrical joint zero-diagonalization and MIMO zero division multiple access," *IEEE Trans. Signal Processing*, vol. 59, no. 5, pp. 2296–2307, May 2011.

[13] G. Chabriel and J. Barrère, "A direct algorithm for nonorthogonal approximate joint diagonalization," *IEEE Trans. Signal Processing*, vol. 60, no. 1, pp. 39–47, Jan. 2012.

[14] G. Chabriel and J. Barrère, "An instantaneous formulation of mixtures for blind separation of propagating waves," *IEEE Trans. Signal Processing*, vol. 54, no. 1, pp. 49–58, 2006.

[15] P. Comon, "Independent component analysis, a new concept?," *Signal Process.*, vol. 36, no. 3, pp. 287–314, 1994.

[16] L. De Lathauwer and J. Castaing, "Blind identification of underdetermined mixtures by simultaneous matrix diagonalization," *IEEE Trans. Signal Processing*, vol. 56, no. 3, pp. 1096–1105, 2008.

[17] L. De Lathauwer, "A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization," *SIAM J. Matrix Anal. Applicat.*, vol. 28, no. 3, pp. 642–666, July 2006.

[18] L. De Lathauwer, B. De Moor, and J. Vandewalle, "Computation of the canonical decomposition by means of a simultaneous generalized Schur decompositions," *SIAM J. Matrix Anal. Applicat.*, vol. 26, no. 2, pp. 295–327, 2004.

[19] S. Dégerine and E. Kane, "A comparative study of approximate joint diagonalization algorithms for blind source separation in presence of additive noise," *IEEE Trans. Signal Processing*, vol. 55, no. 6 (part 2), pp. 3022–3031, 2007.

[20] J. Eriksson and V. Koivunen, "Complex random vectors and ICA models: Identifiability, uniqueness, and separability," *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 1017–1029, 2006.

[21] E.-M. Fadaili, N. Thirion-Moreau, and E. Moreau, "Non-orthogonal joint diagonalization/zero-diagonalization for source separation based on time-frequency distributions," *IEEE Trans. Signal Processing*, vol. 55, no. 5, pp. 1673–1687, May 2007.

[22] R. A. Harshman, "Foundations of the PARAFAC procedure: Model and conditions for an "explanatory" multimode factor analysis," *UCLA Working Papers Phonet.*, vol. 16, pp. 1–84, 1970.

[23] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[24] R. Iferroudjene, K. Abed-Meraim, and A. Belouchrani, "Joint diagonalization of non defective matrices using generalized Jacobi rotations," in *Proc. 10th Int. Conf. Information Science, Signal Processing and Their Applications (ISSPA 2010)*, pp. 345–348.

[25] M. Kleinsteuber and H. Shen, "Uniqueness analysis of non-unitary matrix joint diagonalization," *IEEE Trans. Signal Processing*, vol. 61, no. 7, pp. 1786–1796, Apr. 2013.

[26] Z. Koldovsky and P. Tichavsky, "Time-domain blind separation of audio sources based on a complete ICA decomposition of an observation space." *IEEE Trans. Audio, Speech Lang. Process*ing, vol. 19, no. 2, pp. 406–416, 2011.

[27] D. Lahat, J. Cardoso, and H. Messer, "Second-order multidimensional ICA: Performance analysis," *IEEE Trans. Signal Processing*, vol. 60, no. 9, pp. 4598–4610, 2012.

[28] X.-L. Li and X. D. Zhang, "Nonorthogonal joint diagonalization free of degenerate solutions," *IEEE Trans. Signal Processing*, vol. 55, no. 5, pp. 1803–1814, May 2007.

[29] E. Moreau, "A generalization of joint-diagonalization criteria for source separation," *IEEE Trans. Signal Processing*, vol. 49, no. 3, pp. 530–541, Mar. 2001.

[30] E. Moreau and N. Thirion-Moreau, "Non symmetrical contrasts for sources separation," *IEEE Trans. Signal Processing*, vol. 47, no. 8, pp. 2241–2252, Aug. 1999.

[31] D. Nion, "A tensor framework for nonunitary joint block diagonalization," *IEEE Trans. Signal Processing*, vol. 59, no. 10, pp. 4585–4594, 2011.

[32] E. Ollila and V. Koivunen, "Complex ICA using generalized uncorrelating transform," *Signal Process.*, vol. 89, no. 4, pp. 365–377, 2009.

[33] L. Parra and P. Sajda, "Blind source separation via generalized eigenvalue decomposition," *J. Mach. Learn. Res.*, vol. 4, nos. 7–8, pp. 1261–1269, 2008.

[34] D.-T. Pham and Ph. Garat, "Blind separation of mixtures of independent sources through a quasi maximum likelihood approach," *IEEE Trans. Signal Processing*, vol. 45, no. 7, pp. 1712–1725, July 1997.

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

[36] A. H. Phan, P. Tichavsky, and A. Cichocki, "Low complexity damped Gauss-Newton algorithms for parallel factor analysis," *SIAM J. Linear Algebra Applicat.*, vol. 34, no. 1, pp. 126–147, Jan. 2013.

[37] F. Roemer and M. Haardt, "A semi-algebraic framework for approximate CP decomposition via simultaneous matrix diagonalization (SECSI)," *Signal Process.*, vol. 93, no. 9, pp. 2722–2738, Sept. 2013.

[38] H. Shen and M. Kleinsteuber, "Non-unitary matrix joint diagonalization for complex independent vector analysis," *EURASIP J. Adv. Signal Process.*, vol. 241, pp. 1–10, Nov. 2012.

[39] H. Shen and M. Kleinsteuber, "Complex blind source separation via simultaneous strong uncorrelating transform," in *Proc. 9th Int. Conf. Latent Variable Analysis and Signal Separation*, vol. 6365. Berlin/Heidelberg: Springer-Verlag, 2010, pp. 287–294.

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

[41] F. J. Theis, "A new concept for separability problems in blind source separation," *Neural Computat.*, vol. 16, no. 9, pp. 1827–1850, 2004.

[42] P. Tichavsky, E. Doron, A. Yeredor, and J. Nielsen, "A computationally affordable implementation of an asymptotically optimal BSS algorithm for AR sources," in *Proc. EUSIPCO* 2006, Florence, Italy, Sept. 2006, pp. 1–5.

[43] P. Tichavsky and A. Yeredor, "Fast approximate joint diagonalization incorporating weight matrices," *IEEE Trans. Signal Processing*, vol. 57, no. 3, pp. 878–891, Mar. 2009.

[44] P. Tichavsky and Z. Koldovsky, "Algorithms for nonorthogonal approximate joint block-diagonalization," in *Proc. 20th European Signal Processing Conf. (EUSIPCO)*, Bucharest, Romania, Aug. 27–31, 2012, pp. 2094–2098.

[45] P. Tichavsky, and Z. Koldovsky, "Weight adjusted tensor method for blind separation of underdetermined mixtures of nonstationary sources," *IEEE Trans. Signal Processing*, vol. 59, no. 3, pp. 1037–1047, Mar. 2011.

[46] L. Tong, V. C. Soon, Y. Huang, and R. Liu, "AMUSE: A new blind identification algorithm," in *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS)*, New Orleans, LA, 1990, vol. 3, pp. 1784–1787.

[47] T. Trainini, X.-L. Li, E. Moreau, and T. Adali, "A relative gradient algorithm for joint decompositions of complex matrices," in *Proc. 18th European Signal Processing Conf.*, 2010, pp. 1073–1076.

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

[49] A. Yeredor, "On using exact joint-diagonalization for non-iterative approximate joint-diagonalization," *IEEE Signal Processing Lett.*, vol. 12, no. 9, pp. 645–648, Sept. 2005.

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

[51] A. Yeredor, "Blind separation of Gaussian sources with general covariance structures: Bounds and optimal estimation," *IEEE Trans. Signal Processing*, vol. 58, no. 10, pp. 5057–5068, Oct. 2010.

[52] A. Yeredor, "Performance analysis of the strong uncorrelating transformation in blind separation of complex-valued sources," *IEEE Trans. Signal Processing*, vol. 60, no. 1, pp. 478–453, Jan. 2012.

[53] A. Yeredor, "Multiple-snapshots BSS with general covariance structures: A partial maximum likelihood approach involving weighted joint diagonalization," *Signal Process.*, vol. 92, no. 8, pp. 1832–1843, Aug. 2012.

[54] 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 Processing Lett.*, vol. 4, no. 9, pp. 266–268, Sept. 1997.

[55] A. Ziehe, P. Laskov, G. Nolte, and K.-R. Müller, "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, July 2004.

[SP]