ON APPROXIMATE JOINT "BIAGONALIZATION" - A TOOL FOR NOISY BLIND SOURCE SEPARATION

Arie Yeredor

School of Electrical Engineering, Tel-Aviv University

ABSTRACT

Quite a few algorithms for Blind Source Separation (BSS) rely of approximate joint diagonalization (AJD) of a set of matrices. These matrices are usually estimates of some underlying matrices which admit exact joint diagonalization (EJD) in a noiseless scenario. When additive noise is present, the underlying set no longer admits EJD, since an unknown noise-related matrix is usually added to the diagonalizable form. Often this noise-related matrix is known to be diagonal. Hence, we define the "approximate joint biagonalization" (AJB) problem, aimed at fitting the noisy model to the estimated set of matrices. AJB differs from AJD in the presence of an additional unknown diagonal matrix in the model. We provide an iterative algorithm for minimizing the AJB Least-Squares (LS) criterion, based on an extension of an existing AJD algorithm. In addition, we provide some analytical results on exact and approximate biagonalization, applicable only to the special cases of two- and three-dimensional BSS problems.

1. INTRODUCTION

We consider the noisy Blind Source Separation (BSS) model

$$\boldsymbol{x}[t] = \boldsymbol{A}\boldsymbol{s}[t] + \boldsymbol{v}[t] \quad t = 1, 2, ..., T \tag{1}$$

where $\boldsymbol{s}[t]$ is an $N \times 1$ vector of statistically independent source signals, $\boldsymbol{A}$ is an unknown $N \times N$ mixing matrix, $\boldsymbol{v}[t]$ is some additive noise vector and $\boldsymbol{x}[t]$ is the $N \times 1$ vector of observed noisy mixtures (we assume the square model for simplicity).

Quite a few well-known BSS algorithms, originally designed for the noiseless case, rely, for the purpose of estimating $\boldsymbol{A}$, on the joint eigen-structure shared (in the noiseless case) by some set of K $N \times N$ observations-related "property matrices" $\{\boldsymbol{M}_k^x\}_{k=1}^K$ (to be specified immediately),

$$\boldsymbol{M}_k^x = \boldsymbol{A}\boldsymbol{M}_k^s\boldsymbol{A}^T \quad k = 1, 2, ..., K \tag{2}$$

where $\{\boldsymbol{M}_k^s\}_{k=1}^K$ is a similar set of sources-related "property matrices", which, by virtue of the sources' independence, are known to be diagonal (but are otherwise unknown). The observations-related "property-matrices" $\boldsymbol{M}_k^x$ can be estimated from the observed data, but usually their respective estimates $\widehat{\boldsymbol{M}}_k^x$ no longer share the joint eigen-structure (2) (with any diagonal matrices $\widehat{\boldsymbol{M}}_k^s$ replacing the true $\boldsymbol{M}_k^s$). Thus, $\boldsymbol{A}$ (and possibly $\{\widehat{\boldsymbol{M}}_k^s\}_{k=1}^K$ as nuisance parameters) are estimated from $\{\widehat{\boldsymbol{M}}_k^x\}_{k=1}^K$ using approximate joint diagonalization (AJD). Quite a few iterative algorithms for AJD have been proposed in recent years (e.g., [1, 2, 3, 4, 5, 6, 7, 8]).

Examples of such "property matrices" are, e.g., correlation matrices at different lags (SOBI, [9]), cumulant matrices (JADE, [10]), zero-lag correlation matrices over different segments (for nonstationary sources, [11]), Hessians of the second characteristic function [12], spectral density matrices at different frequencies [13].

However, in the case of noisy observations (1), the joint eigen-structure form (2) no longer holds, even for the true set $\{\boldsymbol{M}_k^s\}_{k=1}^K$. In many (but not all) cases, this relation can be substituted by

$$\boldsymbol{M}_k^x = \boldsymbol{A}\boldsymbol{M}_k^s\boldsymbol{A}^T + \boldsymbol{D}_k \quad k = 1, 2, ..., K \tag{3}$$

where $\{\boldsymbol{D}_k\}_{k=1}^K$ are related to statistical properties of the noise. This relation usually holds when the noise is statistically independent of the sources. Naturally, if statistical knowledge about the noise is available to the extent that all $\boldsymbol{D}_k$ are known, then usually by subtracting the known $\boldsymbol{D}_k$ from the estimated $\widehat{\boldsymbol{M}}_k^x$ it is possible to resort to a standard AJD problem.

However, often the statistical properties of the noise (and therefore $\{\boldsymbol{D}_k\}_{k=1}^K$) are unknown, and have to be estimated (as nuisance parameters) as well. Luckily, in the context of some of the above-mentioned cases, further (quite reasonable) assumptions ensure that all $\boldsymbol{D}_k$ are equal, and may therefore be denoted $\boldsymbol{D}$ (independent of k), which in the following cases equals the (constant) zero-lag correlation of the noise signals:

- Zero-lag correlations over different segments: When the noise signals are assumed to be stationary (as opposed to the sources' non-stationarity);
- Hessians of the second characteristic function: When the noise signals are assumed to be Gaussian;
- Spectral density matrices: When the noise signals are assumed to be spectrally white.

Under the additional reasonable assumption that the noise signals are mutually uncorrelated, $\boldsymbol{D}$ is known to be diagonal. Thus, we loosely refer to this relation as *Joint Biagonalization* (as opposed to Joint Diagonalization), where the *Bi-* prefix insinuates that *two* diagonal matrices are involved (for each k).

Therefore, when the estimated set $\{\widehat{\boldsymbol{M}}_k^s\}_{k=1}^K$ is given in the case of noisy BSS, it is often desired to find respective estimates of $\boldsymbol{A}$, $\{\boldsymbol{M}_k^s\}_{k=1}^K$ and $\boldsymbol{D}$ such that the joint biagonalization (JB) model (3) is fit "as closely as possible". One possible measure of fit is the squared norm of differences, giving rise to the following definition of the *Approximate Joint Biagonalization* (AJB) problem (in the Least Squares sense):

Given a set of K $N \times N$ "target matrices" $\{\boldsymbol{M}_k\}_{k=1}^K$, find an $N \times N$ matrix $\boldsymbol{A}$, K $N \times N$ diagonal matrices $\{\boldsymbol{\Lambda}_k\}_{k=1}^K$ and an additional diagonal matrix $\boldsymbol{D}$, such that the following criterion is minimized:

$$C_{LS}\left(\boldsymbol{A}, \{\boldsymbol{\Lambda}_k\}_{k=1}^K, \boldsymbol{D}\right) \triangleq \sum_{k=1}^{K} \|\boldsymbol{M}_k - \boldsymbol{A}\boldsymbol{\Lambda}_k\boldsymbol{A}^T - \boldsymbol{D}\|_F^2 \tag{4}$$

where $\|\cdot\|_F^2$ denotes the squared Frobenius norm.

In addition, we differentiate between two possible "flavors" for this problem: In some applications all of the noise variances can be assumed to be identical (reflecting spatially-white noise), namely $\boldsymbol{D}$ is known to be a multiple of $\boldsymbol{I}$ (the Identity matrix). We refer to such special cases as the *Uniform* AJB problem. In general, however, $\boldsymbol{D}$ would have free diagonal elements, implying the *Non-Uniform* AJB, or simply AJB. Naturally, Uniform AJB is a particular case of AJB, yet it implies different constraints on the solution, an hence would be treated differently.

While, as mentioned above, the classical AJD problem has been studied quite extensively in recent years, the AJB problem has seen very little treatment, if any - despite its relative importance. For example, Cardoso *et al.* [14] and Cardoso and Pham [15] consider the case of noisy mixtures in the context of Gaussian sources and noise, and apply the Estimate-Maximize (EM) algorithm to the data in order to circumvent the associated AJB problem (posed under a different matching criterion). Alternatively, an iterative optimization methods (e.g., a Newton-like optimization) is applied in [15].

In the following section we briefly sketch an extension of the "Alternating Columns - Diagonal Centers" (AC-DC) AJD algorithm [5] to AJB problems, and then move on, in Section 3, to more interesting results, exploring some basic properties of the AJB problem. We outline possible *non-iterative* (closed-form) solutions, which are true minimizers of C_{LS} in the 2×2 ($N = 2$) case (both for the Uniform and non-Uniform cases), and can serve as good initial guesses for iterative algorithms in other cases. These solutions can be viewed as extensions of similar solutions obtained for the noiseless case in [16].

2. AC-DC FOR APPROXIMATE JOINT BIAONALIZATION

In this section we provide a brief extension of the AC-DC algorithm [5], enabling to accommodate the noisy case. AC-DC essentially iterates between two phases: the so-called "Alternating Columns" (AC) phase, in which the LS criterion is minimized with respect to (w.r.t.) each column of $\boldsymbol{A}$ while maintaining all other columns (and other parameters) fixed; and the so-called "Diagonal Centers" (DC) phase, in which the LS criterion is minimized w.r.t. all diagonal matrices $\{\boldsymbol{\Lambda}_k\}_{k=1}^K$ with all other parameters fixed. A natural extension of the AJB problem is to augment the DC phase with the parameter(s) of $\boldsymbol{D}$.

Thus, the extended AC-DC algorithm for the minimization of the AJB criterion C_{LS} (4) would alternate between application of the AC phase (for $\ell = 1, 2, ...N$), and the DC phase, whose extended versions take the following forms:

2.1. AC phase

Minimization of C_{LS} w.r.t. $\boldsymbol{a}_\ell$, the ℓ-th column of $\boldsymbol{A}$: Calculate

$$\boldsymbol{P} \stackrel{\triangle}{=} \sum_{k=1}^{K} \lambda_\ell^{[k]} \left[\boldsymbol{M}_k - \sum_{\substack{n=1\\n\neq\ell}}^{N} \lambda_n^{[k]} \boldsymbol{a}_n \boldsymbol{a}_n^T - \boldsymbol{D} \right];$$

Find the largest eigenvalue μ and the associated unit-norm eigenvector $\boldsymbol{\alpha}$ of $\boldsymbol{P}$; If $\mu < 0$, set $\boldsymbol{a}_\ell = \boldsymbol{0}$, otherwise set

$$\boldsymbol{a}_\ell = \frac{\boldsymbol{\alpha}\sqrt{\mu}}{\sqrt{\sum_{k=1}^{K}(\lambda_\ell^{[k]})^2}}.$$

Here $\lambda_n^{[k]}$ denotes the (n, n)-th element of $\boldsymbol{\Lambda}_k$.

2.2. DC phase

The addition of the unknown parameter(s) in $\boldsymbol{D}$ induces coupling between minimization operations w.r.t. the diagonal matrices $\{\boldsymbol{\Lambda}_k\}_{k=1}^K$ - which renders the DC phase considerably more complicated than in the noiseless case. Nevertheless, the formulation of the associate linear LS problem aimed at minimizing C_{LS} w.r.t. $\{\boldsymbol{\Lambda}_k\}_{k=1}^K$ and $\boldsymbol{D}$ (simultaneously) is quite straightforward, and the solution is tractable. Denoting $\boldsymbol{m}_k \stackrel{\triangle}{=} \mathrm{vec}\{\boldsymbol{M}_k\}$, $\boldsymbol{\lambda}_k \stackrel{\triangle}{=} \mathrm{diag}\{\boldsymbol{\Lambda}_k\}$ and $\boldsymbol{d} \stackrel{\triangle}{=} \mathrm{diag}\{\boldsymbol{D}\}$, we can rewrite C_{LS} (4) as:

$$C_{LS} = \sum_{k=1}^{K} \|\boldsymbol{m}_k - (\boldsymbol{A}^* \odot \boldsymbol{A})\boldsymbol{\lambda}_k - \boldsymbol{G}\boldsymbol{d}\|^2 \quad (5)$$

where $\boldsymbol{A}^* \odot \boldsymbol{A}$ denotes the Khatri-Rao product of $\boldsymbol{A}^*$ and $\boldsymbol{A}$ (essentially a column-wise Kronecker product - see, e.g., [5]), and $\boldsymbol{G}$ depends on whether the noise variances are "Uniform" or "Non-Uniform": In the Non-Uniform case $\boldsymbol{G}$ is a $N^2 \times N$ matrix, whose n-th column is all-zeros except for a 1 at the $((n-1)N+n)$-th location; In the Uniform case $\boldsymbol{G}$ is an $N^2 \times 1$ column vector, essentially given by the sum of all columns of the Non-Uniform $\boldsymbol{G}$. Note also that in the Uniform case the vector $\boldsymbol{d}$ would be replaced by a scalar d (reflecting the single uniform variance). For notation purposes, let us denote $\mathcal{A} \stackrel{\triangle}{=} (\boldsymbol{A}^* \odot \boldsymbol{A})$. It is now straightforward to observe, that minimization w.r.t. each $\boldsymbol{\lambda}_k$ yields

$$\hat{\boldsymbol{\lambda}}_k(\boldsymbol{d}) = (\mathcal{A}^T\mathcal{A})^{-1}\mathcal{A}^T(\boldsymbol{m}_k - \boldsymbol{G}\boldsymbol{d}) \quad k = 1, 2, ..., K \quad (6)$$

(hence they all depend on the unknown $\boldsymbol{d}$, in contrast to the noiseless case). In order to extract the minimizing $\boldsymbol{d}$, we now need to minimize

$$\begin{aligned} C_{LS} &= \sum_{k=1}^{K} \|\boldsymbol{m}_k - \mathcal{A}\hat{\boldsymbol{\lambda}}_k(\boldsymbol{d}) - \boldsymbol{G}\boldsymbol{d}\|^2 = \\ &\sum_{k=1}^{K} \|\boldsymbol{m}_k - \boldsymbol{P}_{\mathcal{A}}(\boldsymbol{m}_k - \boldsymbol{G}\boldsymbol{d}) - \boldsymbol{G}\boldsymbol{d}\|^2 = \\ &\sum_{k=1}^{K} \|\boldsymbol{P}_{\mathcal{A}}^{\perp}\boldsymbol{m}_k - \boldsymbol{P}_{\mathcal{A}}^{\perp}\boldsymbol{G}\boldsymbol{d}\|^2 = \\ &\sum_{k=1}^{K} \|\boldsymbol{P}_{\mathcal{A}}^{\perp}(\boldsymbol{m}_k - \boldsymbol{G}\boldsymbol{d}\|^2 \end{aligned} \quad (7)$$

where $\boldsymbol{P}_{\mathcal{A}} \stackrel{\triangle}{=} \mathcal{A}(\mathcal{A}^T\mathcal{A})^{-1}\mathcal{A}^T$ and $\boldsymbol{P}_{\mathcal{A}}^{\perp} \stackrel{\triangle}{=} \boldsymbol{I} - \boldsymbol{P}_{\mathcal{A}}$ denote projection matrices onto the subspace spanned by the columns of $\mathcal{A}$ and the complementary subspace (resp.). It can now be easily observed that the minimizing $\boldsymbol{d}$ is given by

$$\hat{\boldsymbol{d}} = (\boldsymbol{G}^T\boldsymbol{P}_{\mathcal{A}}^{\perp}\boldsymbol{G})^{-1}\boldsymbol{G}^T\boldsymbol{P}_{\mathcal{A}}^{\perp}\bar{\boldsymbol{m}} \quad (8)$$

where $\bar{\boldsymbol{m}} \stackrel{\triangle}{=} \frac{1}{K}\sum_{k=1}^{K}\boldsymbol{m}_k$ is actually the $\mathrm{vec}\{\cdot\}$ form of the average of all K target matrices. Once computed, $\hat{\boldsymbol{d}}$ can be plugged into $\hat{\boldsymbol{\lambda}}_k(\boldsymbol{d})$ (6) to produce the minimizing $\{\hat{\boldsymbol{\lambda}}_k\}_{k=1}^K$.

3. SUBSPACE ANALYSIS

Next, we proceed with characterizing the solutions in a derivation which extends a similar subspace analysis presented for the noiseless case in [16]. The basic goal in this analysis is to eliminate $\{\boldsymbol{\lambda}_k\}_{k=1}^K$ and $\boldsymbol{d}$ from C_{LS}, expressing C_{LS} in terms of $\boldsymbol{A}$ only, and exploring the minimizing solution. However, in order to avoid over-parameterization, we first slightly modify the definitions of some of the quantities defined above, by essentially replacing the $\mathrm{vec}\{\cdot\}$ operator with the $\mathrm{svec}\{\cdot\}$ operator as follows.

The well-known $\mathrm{vec}\{\cdot\} : \mathbb{R}^{N\times M} \rightarrow \mathbb{R}^{NM}$ operator is often used for transforming an $N \times M$ matrix into an $NM \times 1$ vector by concatenating all columns. For square symmetric matrices, it may be desired to define a more succinct transformation, free of redundancies, by transforming $N\times N$ matrices into $N(N+1)/2\times 1$ vectors. Further, in order to maintain proper weighting of elements in the calculation of norms, all off-diagonal elements should be multiplied by $\sqrt{2}$. To this end, we define the $\mathrm{svec}\{\cdot\} : \mathbb{R}^{N\times N} \rightarrow \mathbb{R}^{N(N+1)/2}$ invertible transformation as follows: For any $N \times N$ symmetric matrix $\boldsymbol{Y}$, the $N(N+1)/2 \times 1$ vector $\boldsymbol{y} = \mathrm{svec}\{\boldsymbol{Y}\}$ is constructed using:

$$y\left(\frac{n(n-1)}{2}+m\right) = \begin{cases} \sqrt{2}Y(m,n) & \text{if } m<n, \\ Y(n,n) & \text{if } m=n, \end{cases}$$
$$\text{for} \quad n=1,2,...,N \quad \text{and} \quad m=1,2,...,n, \quad (9)$$

where $y(i)$ and $Y(i,j)$ denote the i-th and (i,j)-th elements of $\boldsymbol{y}$ and $\boldsymbol{Y}$, respectively. Thus, $\boldsymbol{y}$ contains all the (weighted) elements of the upper right triangular section of $\boldsymbol{Y}$. Conversely, the $\mathrm{unsvec}\{\cdot\} : \mathbb{R}^{N(N+1)/2} \rightarrow \mathbb{R}^{N\times N}$ transformation $\boldsymbol{Y} = \mathrm{unsvec}\{\boldsymbol{y}\}$ would be defined, for any $N(N+1)/2 \times 1$ vector $\boldsymbol{y}$, as the (only) $N \times N$ symmetric matrix $\boldsymbol{Y}$ such that $\boldsymbol{y} = \mathrm{svec}\{\boldsymbol{Y}\}$.

For ease of notation, we shall, from now on, denote the number of "non-redundant elements" as $N_2 \triangleq N(N+1)/2$.

Note that for any symmetric matrix $\boldsymbol{Y}$ we have $\|\boldsymbol{Y}\|_F^2 = \|\mathrm{vec}\{\boldsymbol{Y}\}\|^2 = \|\mathrm{svec}\{\boldsymbol{Y}\}\|^2$. Note further, that there exists an implied $N_2 \times N^2$ constant matrix $\boldsymbol{F}$ (with a single non-zero element in each row, either 1 or $\sqrt{2}$), such that $\mathrm{svec}\{\boldsymbol{Y}\} = \boldsymbol{F}\,\mathrm{vec}\{\boldsymbol{Y}\}$ for all $\boldsymbol{Y} \in \mathbb{R}^{N\times N}$. We shall use $\boldsymbol{F}$ for notation purposes only, without explicitly specifying its structure.

We now redefine accordingly, $\boldsymbol{m}_k \triangleq \mathrm{svec}\{\boldsymbol{M}_k\}$ and $\mathcal{A} \triangleq \boldsymbol{F}(\boldsymbol{A}^* \odot \boldsymbol{A})$, so that C_{LS} maintains the form of (7). Substituting $\hat{\boldsymbol{d}}$ (8) into (7) we further obtain

$$C_{LS} = \sum_{k=1}^{K} \|\boldsymbol{P}_{\mathcal{A}}^{\perp}\boldsymbol{m}_k - \boldsymbol{P}_{\mathcal{A}}^{\perp}\boldsymbol{G}(\boldsymbol{G}^T\boldsymbol{P}_{\mathcal{A}}^{\perp}\boldsymbol{G})^{-1}\boldsymbol{G}^T\boldsymbol{P}_{\mathcal{A}}^{\perp}\bar{\boldsymbol{m}}\|^2$$
$$\triangleq \sum_{k=1}^{K} \|\boldsymbol{P}_{\mathcal{A}}^{\perp}\boldsymbol{m}_k - \boldsymbol{P}_{PG}\bar{\boldsymbol{m}}\|^2, \quad (10)$$

where $\boldsymbol{P}_{PG}$ is a projection matrix onto the subspace spanned by the columns of $\boldsymbol{P}_{\mathcal{A}}^{\perp}\boldsymbol{G}$. With further straightforward manipulations, exploiting the property $\boldsymbol{P}_{\mathcal{A}}^{\perp}\boldsymbol{P}_{PG} = \boldsymbol{P}_{PG}$, we get

$$C_{LS} = Trace\{\boldsymbol{P}_{\mathcal{A}}^{\perp}\tfrac{1}{K}\sum_{k=1}^{K}\boldsymbol{m}_k\boldsymbol{m}_k^T - \boldsymbol{P}_{PG}\bar{\boldsymbol{m}}\bar{\boldsymbol{m}}^T\} \quad (11)$$

which has to be minimized w.r.t $\boldsymbol{A}$ in order to obtain the minimal C_{LS}. We can now make the following observations for the $N=2$ and $N=3$ cases.

3.1. The $N=2$ case

When $N=2$ we have $N_2=3$. Thus, in such cases the ranks of the 3×3 matrices $\boldsymbol{P}_{\mathcal{A}}$ and $\boldsymbol{P}_{\mathcal{A}}^{\perp}$ are 2 and 1 (resp.). In the Uniform case we have $\boldsymbol{G} = [1\ 0\ 1]^T$ (also rank 1), thus $\boldsymbol{P}_{PG} = \boldsymbol{P}_{\mathcal{A}}^{\perp}$ (unless $\boldsymbol{P}_{\mathcal{A}}^{\perp}\boldsymbol{G} = \boldsymbol{0}$, in which case $\boldsymbol{P}_{PG}$ is undetermined, but this only happens on a set of measure zero in terms of $\boldsymbol{A}$). We therefore need to minimize $Trace\{\boldsymbol{P}_{\mathcal{A}}^{\perp}\boldsymbol{R}\}$, where

$$\boldsymbol{R} \triangleq \tfrac{1}{K}\sum_{k=1}^{K}(\boldsymbol{m}_k - \bar{\boldsymbol{m}})(\boldsymbol{m}_k - \bar{\boldsymbol{m}})^T. \quad (12)$$

Now, since $\boldsymbol{P}_{\mathcal{A}}^{\perp}$ has rank $N=2$ (for any choice of a nonsingular $\boldsymbol{A}$, and lower if $\boldsymbol{A}$ is singular), $Trace\{\boldsymbol{P}_{\mathcal{A}}^{\perp}\boldsymbol{R}\}$ cannot be decreased below the smallest eigenvalue of $\boldsymbol{R}$ (by any choice of $\boldsymbol{A}$) - see, e.g., [16]. Moreover, a matrix $\widehat{\boldsymbol{A}}$ which attains this lower bound can almost always be found, as it is given by the exact joint diagonalizer of the two matrices $\mathrm{unsvec}\{\boldsymbol{u}_1\}$ and $\mathrm{unsvec}\{\boldsymbol{u}_2\}$, where $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ are the eigenvectors of $\boldsymbol{R}$ corresponding to its two largest eigenvalues (again, see [16] for a rigorous proof).

Note further, that the rank of $\boldsymbol{R}$ is at most $K-1$ (since the sum of the K vectors $(\boldsymbol{m}_k - \bar{\boldsymbol{m}})$ is zero, hence each one is given by a linear combination of the others). Therefore when $K=2$ or $K=3$, the smallest eigenvalue of $\boldsymbol{R}$ is zero, so *exact joint biagonalization* (EJB) of two or three matrices can almost always be attained. For $K = 2$ this is quite trivial, since it is well-known that exact joint diagonalization (EJD) is almost always possible for two symmetric matrices, and EJD is merely a particular case of EJB (with $\boldsymbol{D} = \boldsymbol{0}$). However, for $K = 3$ this result is new.

In the Non-Uniform case the rank of $\boldsymbol{G}$ is 2, yet the rank of $\boldsymbol{P}_{\mathcal{A}}^{\perp}\boldsymbol{G}$ is still 1 - namely, its two columns are collinear - therefore $\boldsymbol{d}$ cannot be uniquely determined (see (8)), and so neither can $\{\boldsymbol{\lambda}_k\}_{k=1}^{K}$ (see (6)). Nevertheless, we still have $\boldsymbol{P}_{\mathcal{A}}^{\perp} = \boldsymbol{P}_{PG}$, and the conclusions regarding the minimizing $\boldsymbol{A}$ remain valid.

3.2. The $N = 3$ case

When $N = 3$ we have $N_2 = 6$. Thus, the ranks of the 6×6 matrices $\boldsymbol{P}_{\mathcal{A}}$ and $\boldsymbol{P}_{\mathcal{A}}^{\perp}$ are both 3. In the Non-Uniform case the rank of $\boldsymbol{G}$ is also 3, therefore we still have $\boldsymbol{P}_{PG} = \boldsymbol{P}_{\mathcal{A}}^{\perp}$ (except for degenerate cases, which can only happen on a set of measure zero in terms of $\boldsymbol{A}$), and again $C_{LS} = Trace\{\boldsymbol{P}_{\mathcal{A}}^{\perp}\boldsymbol{R}\}$. Therefore, the lower bound on C_{LS} is now the sum of the three smallest eigenvalues of $\boldsymbol{R}$, but in this case we do not have a proposed minimizer $\widehat{\boldsymbol{A}}$ guaranteed to attain that bound, since in general only the two largest eigenvalues of $\boldsymbol{R}$ can be zeroed out by the EJD of $\mathrm{unsvec}\{\boldsymbol{u}_1\}$ and $\mathrm{unsvec}\{\boldsymbol{u}_2\}$. However, denoting the exact joint diagonalizer of $\mathrm{unsvec}\{\boldsymbol{u}_1\}$ and $\mathrm{unsvec}\{\boldsymbol{u}_2\}$ by $\widehat{\boldsymbol{A}}$, we can make the following observations:

- If $K = 2$ or $K = 3$ the third through sixth eigenvalues of $\boldsymbol{R}$ are zero, and EJB of the entire set is attained by $\boldsymbol{A}$, just as in the $N = 2$ case above.

- If $K > 3$, $\widehat{\boldsymbol{A}}$ does not (in general) attain the (generally unattainable) EJB of the entire set, nor does it minimize C_{LS} (in contrast to the $N = 2$ case). Nevertheless, it can serve as a very good "initial guess" for an iterative algorithm (such as the extended AC-DC above), since it eliminates the two largest eigenvalues of $\boldsymbol{R}$ from $Trace\{\boldsymbol{P}_{\mathcal{A}}^{\perp}\boldsymbol{R}\}$ (see [16] for more details).

In the Uniform case, however, $\boldsymbol{G}$ has rank one, so we do not have a simple relation between $\boldsymbol{P}_{PG}$ and $\boldsymbol{P}_{\mathcal{A}}^{\perp}$; Therefore C_{LS} can no longer be expressed as $Trace\{\boldsymbol{P}_{\mathcal{A}}^{\perp}\boldsymbol{R}\}$ (with any $\boldsymbol{R}$), and analytic minimization of C_{LS} w.r.t. $\boldsymbol{A}$ is currently unknown, so an iterative algorithm such as the extended AC-DC may be used.

3.3. The $N > 3$ case

When $N > 3$ the rank of $\boldsymbol{P}_{\mathcal{A}}^{\perp}$, which is $N_2 - N = N(N-1)/2$ is (unfortunately) always larger than the rank of $\boldsymbol{G}$ (be it 1 in the Uniform case or N in the Non-Uniform case), so just like in the $N = 3$ Uniform case - the absence of a simple relation between $\boldsymbol{P}_{PG}$ and $\boldsymbol{P}_{\mathcal{A}}^{\perp}$ prevents a simple analytic solution of the minimization problem.

4. CONCLUSION

We defined the AJB problem, naturally encountered in AJD-based BSS algorithms when additive noise is present. We provided an extended version of the AC-DC algorithm for solving the general EJB problem. In addition, we used subspace analysis of the LS criterion to obtain additional results as follows.

Regarding EJB, the following results were derived (on top of the somewhat trivial result stating that EJB is possible for almost any *two* symmetric matrices):

- In the 2×2 case ($N = 2$), EJB is possible for almost every *triplet* of symmetric matrices, in both the Uniform and Non-Uniform cases. The mixing $\boldsymbol{A}$ is uniquely identifiable (up to the trivial scaling and permutation ambiguities) in both cases. However, in the non-uniform case, $\boldsymbol{\Lambda}_1, \boldsymbol{\Lambda}_2, \boldsymbol{\Lambda}_3$ and $\boldsymbol{D}$ are not uniquely identifiable (hence EJB is possible, but the solution is not unique in terms of these (usually nuisance) matrices). Note that this result is in agreement with general identifiability results on noisy Independent Component Analysis (ICA) [17], developed in a different context. In the uniform case the solution is almost always unique.

- In the 3×3 case ($N = 3$), *Non-Uniform* EJB is possible for almost every *triplet* of symmetric matrices, and the solution is almost always unique; However, *Uniform* EJB is almost always impossible (for any $K > 2$).

- For $N > 3$, EJB (Uniform or Non-Uniform) is almost always impossible for $K > 2$.

In addition, we derived the following results concerning the general AJB (rather than EJB) problem:

- In the 2×2 case ($N = 2$), we provided the true global minimizer of C_{LS} for the EJB problem *in closed form* (i.e., no iterative algorithm is required). This minimizer is not unique in terms of $\boldsymbol{\Lambda}_1, \boldsymbol{\Lambda}_2, \boldsymbol{\Lambda}_3$ and $\boldsymbol{D}$ in the Non-Uniform case, but it is always unique in terms of $\boldsymbol{A}$ (up to the trivial ambiguities).

- In the 3×3 case ($N = 3$) we provided a closed-form approximate solution for the Non-Uniform case. Although this solution is not the true minimizer of C_{LS}, it can serve as a good initial guess for a subsequent iterative algorithm. In addition, we provided a non-trivial lower bound on the attainable value of C_{LS}, useful in gauging the distance of a prospective solution from optimality.

It has to be noted that although the main focus in this paper is on the estimation of $\boldsymbol{A}$, in noisy BSS $\widehat{\boldsymbol{A}}$ is often insufficient for optimal reconstruction of the sources, and the so-called "nuisance parameters" may be of interest as well. The minimization algorithms may yield arbitrary values for these parameters whenever they are non-identifiable (see, e.g., [17] for general identifiability conditions in noisy ICA).

5. REFERENCES

[1] Jean-François Cardoso and Antoine Souloumiac, "Jacobi angles for simultaneous diagonalization," *SIAM Journal on Matrix Analysis and Applications*, vol. 17, no. 1, pp. 161–164, 1996.

[2] Alle-Jan van der Veen and Arogyaswami Paulraj, "An analytical constant modulus algorithm," *IEEE Trans. Signal Processing*, vol. 44, no. 5, pp. 1136–1155, 1996.

[3] Alle-Jan van der Veen, "Joint diagonalization via subspace fitting techniques," *Proceedings ICASSP 2001*, vol. 5, 2001, Salt Lake City.

[4] Dinh-Tuan Pham, "Joint approximate diagonalization of positive definite matrices," *SIAM J. on Matrix Anal. and Appl.*, vol. 22, pp. 1136–1152, 2001.

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

[6] Andreas Ziehe, Motoaki Kawanabe, and Klaus-Robert Müller, "A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation," *Journal of Machine Learning Research*, vol. 5, pp. 801–818, 2004.

[7] Marcel Joho and Heinz Mathis, "Joint diagonalization of correlation matrices by using gradient methods with application to blind signal separation," in *Proc. of IEEE Sensor Array and Multichannel Signal Processing Workshop SAM*, 2002, pp. 273–277.

[8] Arie Yeredor, Andreas Ziehe, and Klaus-Robert Müller, "Approximate joint diagonalization using a natural-gradient approach," *Lecture Notes in Computer Science (LNCS 3195): Proceedings ICA 2004*, pp. 89–96, 2004.

[9] Adel Belouchrani, Karim Abed-Meraim, Jean-François Cardoso, and Eric Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Processing*, vol. 45, no. 2, pp. 434–444, 1997.

[10] Jean-François Cardoso and Antoine Souloumiac, "Blind beamforming for non Gaussian signals," *IEE - Proceedings -F*, vol. 140, no. 6, pp. 362–370, 1993.

[11] Dinh-Tuan Pham and Jean-François Cardoso, "Blind separation of instantaneous mixtures of non-stationary sources," *Proceedings ICA2000*, pp. 187–192, 2000.

[12] Arie Yeredor, "Blind source separation via the second characteristic function," *Signal Processing*, vol. 80, no. 5, pp. 897–902, 2000.

[13] Dinh-Tuan Pham, "Mutual information approach to blind separation of stationary sources," *IEEE Trans. Information Theory*, vol. 48, pp. 1935–1946, 2002.

[14] Jean-François Cardoso, Hichem Snoussi, Jacques Delabrouille, and Guillaume Patanchon, "Blind separation of noisy Gaussian stationary sources. application to cosmic microwave background imaging," *Proc. of EUSIPCO2002*, 2002.

[15] Jean-François Cardoso and Dinh-Tuan Pham, "Optimization issues in noisy Gaussian ICA," *Lecture Notes in Computer Science (LNCS 3195): Proceedings ICA 2004*, pp. 41–48, 2004.

[16] Arie Yeredor, "On using exact joint diagonalization for non-iterative approximate joint diagonalization," *to appear in IEEE Signal Processing Letters*, 2005.

[17] Mike Davies, "Identifiability isuues in noisy ICA," *IEEE Signal Processing Letters*, vol. 11, no. 5, pp. 470–473, 2004.