4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003), April 2003, Nara, Japan

A UNIFIED PRESENTATION OF BLIND SEPARATION METHODS FOR CONVOLUTIVE MIXTURES USING BLOCK-DIAGONALIZATION

Cédric Févotte and Christian Doncarli

Institut de Recherche en Communications et Cybernétique de Nantes (IRCCyN)
UMR CNRS 6597, 1 rue de la Noë, BP 92101, 44321 Nantes Cedex 03, France
{fevotte,doncarli}@irccyn.ec-nantes.fr

ABSTRACT

We address in this paper a unified presentation of previous works by A. Belouchrani, K. Abed-Meraim and co-authors about separation of FIR convolutive mixtures using (joint) block-diagonalization. We first present general equations in the stochastic context. Then the implementation of the general method is studied in practice and linked with previous works for stationary and non-stationary sources. The non-stationary case is especially studied within a time-frequency framework: we introduce Spatial Wigner-Ville Spectrum and propose a criterion based on single *block* auto-terms identification to select efficiently, in practice, the matrices to be joint block-diagonalized.

1. INTRODUCTION

Recently a very interesting approach was proposed by A. Belouchrani, K. Abed-Meraim and co-authors to tackle the problem of blind separation of FIR convolutive mixtures. Their work relies on formulating the FIR convolutive mixture as an instantaneous mixture by introducing some appropriate variables. Then the problem comes down to separation of an instantaneous mixture of some new sources but some of them are now dependent. To solve this challenging problem, the latter authors have generalized some standard BSS methods for the instantaneous case by using a joint block-diagonalization scheme. The standard algorithm SOBI [1] for stationary ergodic sources is generalized in [2] while in the non-stationary case, [3] and [4] propose extensions of [5] and [6].

In this paper, we propose a unified presentation of these algorithms and we show that papers [2–4] can be interpreted as practical implementations of some theoretical Blind Source Separation (BSS) equations that we present in a *stochastic* context. In Section 2 we state BSS aim and assumptions for our problem. In Section 3 we introduce several variables to rearrange the convolutive mixture into an instantaneous mixture. Then, the overall BSS problem is formulated in the time-lag plane in Section 4. Some general equations are then derived in Section 5. They show, without approximation, that the sources can be retrieved up to unknown filters. Then, some practical implementations of these equations are considered in Sections 6 and 7. The non-stationary case is especially studied in Section 7.3: we introduce Spatial Wigner-Ville Spectrum and propose a criterion based on single *block* auto-terms identification to select efficiently, in practice, the matrices to be joint block-diagonalized.

2. AIM AND ASSUMPTIONS

We consider the following discrete-time FIR MIMO model:

$$\begin{aligned}\mathbf{x}[t] &= \mathbf{H}[0]\,\mathbf{s}[t] + \mathbf{H}[1]\,\mathbf{s}[t-1] + \ldots + \mathbf{H}[L]\,\mathbf{s}[t-L] \\ &\quad + \mathbf{n}[t] \qquad (1)\end{aligned}$$

where $\mathbf{x}[t] = [x_1[t], \ldots, x_m[t]]^T$ is the vector of size m containing the observations, $\mathbf{s}[t] = [s_1[t], \ldots, s_n[t]]^T$ is the vector of size n containing the stochastic sources (assumed zero-mean and mutually independent at every time instant), $\mathbf{H}[k] = \{h_{ij}[k]\}$, $k = 0 \ldots L$, are $m \times n$ matrices with $m > n$ and $\mathbf{n}[t]$ is a i.i.d noise vector, independent of the sources with:

$$\mathrm{E}\,\{\mathbf{n}[t]\,\mathbf{n}^H[t+\tau]\} = \delta[\tau]\,\sigma^2 \mathbf{I}_m \qquad (2)$$

where $\mathbf{I}_m$ denotes the identity matrix of size m, $\delta[\tau]$ is the Kronecker δ function, $\cdot^H$ denotes “conjugate transpose” and σ^2 denotes the unknown variance of the noise, assumed identical for all observations.

The overall objective of BSS in the convolutive context is to obtain estimates of the mixing filters and/or estimates of the sources up to standard BSS indeterminacies on ordering, scale and phase. In this paper we aim at retrieving estimates of the sources up to unknown FIR filters.

3. FROM CONVOLUTIVE MIXING BACK TO INSTANTANEOUS MIXING

We recall from [2–4] how the convolutive mixing (1) can be rearranged into an instantaneous mixing.

3.1. Notations

Let L' be an integer such that $mL' \geq n(L+L')$ (L' exists when $m > n$). We note, for $i = 1, \ldots, n$:

$$\mathbf{S}_i[t] = [s_i[t], \ldots, s_i[t-(L+L')+1]]^T$$

and for $j = 1, \ldots, m$

$$\begin{aligned}\mathbf{X}_j[t] &= [x_j[t], \ldots, x_j[t-L'+1]]^T \\ \mathbf{N}_j[t] &= [n_j[t], \ldots, n_j[t-L'+1]]^T\end{aligned}$$

where $\cdot^T$ denotes "transpose". Then we introduce:

$$\begin{aligned}\mathbf{S}[t] &= \left[\mathbf{S}_1[t]^T, \ldots, \mathbf{S}_n[t]^T\right]^T \\ \mathbf{X}[t] &= \left[\mathbf{X}_1[t]^T, \ldots, \mathbf{X}_m[t]^T\right]^T \\ \mathbf{N}[t] &= \left[\mathbf{N}_1[t]^T, \ldots, \mathbf{N}_m[t]^T\right]^T\end{aligned}$$

$\forall t$, $\mathbf{S}[t]$ is a column vector of size $n(L+L')$, $\mathbf{X}[t]$ and $\mathbf{N}[t]$ are column vectors of size mL'. For simplicity we note $N = n(L+L')$ and $M = mL'$.

For $i = 1, \ldots, n$ and $j = 1, \ldots, m$ we note $\mathbf{A}_{ij}$ the following $L' \times (L+L')$ matrix:

$$\mathbf{A}_{ij} = \begin{bmatrix} h_{ij}[0] & \ldots & h_{ij}[L] & 0 & \ldots & 0 \\ & \ddots & & \ddots & & \\ & & \ddots & & \ddots & \\ 0 & \ldots & 0 & h_{ij}[0] & \ldots & h_{ij}[L] \end{bmatrix}$$

Finally, we note:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \ldots & \mathbf{A}_{1n} \\ \vdots & & \vdots \\ \mathbf{A}_{1m} & \ldots & \mathbf{A}_{mn} \end{bmatrix}$$

$\mathbf{A}$ is a $M \times N$ matrix which satisfies:

$$\mathbf{X}[t] = \mathbf{A}\,\mathbf{S}[t] + \mathbf{N}[t] \tag{3}$$

In the following we assume that $\mathbf{A}$ is full rank.

3.2. Discussion

Eq. (3) shows that the convolutive mixing (1) can be written as a noisy instantaneous mixing. Such mixtures (3) have been widely studied in BSS/ICA literature. However the big difference here is that the components of $\mathbf{S}[t]$ are not *all* mutually independent: when the sources are not white, for $i = 1, \ldots, n$, the components of $\mathbf{S}_i[t]$ are dependent. Moreover the noise vector $\mathbf{N}[t]$ is stationary but not white (see Section 4.2). However, the variables $\mathbf{S}_1[t], \ldots, \mathbf{S}_n[t]$ are mutually independent. Hence, the separation problem described by Eq. (3) enters the scope of Multidimensional Independent Component Analysis (MICA) introduced in [7], which is the generalization of ICA when some sources are mutually dependent. Then, as explained in [7], we can already state that the components $\mathbf{S}_1[t], \ldots, \mathbf{S}_n[t]$ will be estimated only up to an invertible matrix which models indeterminacies inherent to MICA (see Section 5.3).

4. TIME-LAG REPRESENTATION

In this section, we formulate the overall problem described by Eq. (3) in the time-lag plane.

4.1. Covariance relations

For $(t,\tau) \in \mathbb{Z}^2$ we note $\mathcal{R}_{\mathbf{SS}}[t,\tau]$ the covariance matrix of $\mathbf{S}[t]$ defined by:

$$\mathcal{R}_{\mathbf{SS}}[t,\tau] \stackrel{\text{def}}{=} \mathrm{E}\{\mathbf{S}[t]\,\mathbf{S}[t+\tau]^H\}$$

The vector signals $\mathbf{S}_1[t], \ldots, \mathbf{S}_n[t]$ being mutually independent, the $N \times N$ covariance matrix $\mathcal{R}_{\mathbf{SS}}[t,\tau]$ is block-diagonal with n blocks of dimensions $(L+L')$, such that:

$$\mathcal{R}_{\mathbf{SS}}[t,\tau] = \begin{bmatrix} \mathcal{R}_{\mathbf{S}_1\mathbf{S}_1}[t,\tau] & & \\ & \ddots & \\ & & \mathcal{R}_{\mathbf{S}_n\mathbf{S}_n}[t,\tau] \end{bmatrix}$$

With Eq. (3) we have:

$$\mathcal{R}_{\mathbf{XX}}[t,\tau] = \mathbf{A}\,\mathcal{R}_{\mathbf{SS}}[t,\tau]\,\mathbf{A}^H + \mathcal{R}_{\mathbf{NN}}[t,\tau] \tag{4}$$

4.2. About the noise

The expression of $\mathcal{R}_{\mathbf{NN}}[t,\tau]$ can be further developed. First, the noise being assumed stationary, we have:

$$\mathcal{R}_{\mathbf{NN}}[t,\tau] = \mathcal{R}_{\mathbf{NN}}[\tau]$$

Under the noise assumption (2), the vector signals $\mathbf{N}_1[t], \ldots, \mathbf{N}_m[t]$ are mutually independent. Hence we have:

$$\mathcal{R}_{\mathbf{NN}}[\tau] = \begin{bmatrix} \mathcal{R}_{\mathbf{N}_1\mathbf{N}_1}[\tau] & & \\ & \ddots & \\ & & \mathcal{R}_{\mathbf{N}_m\mathbf{N}_m}[\tau] \end{bmatrix} \tag{5}$$

And for $i = 1, \ldots, m$ we have:

$$\mathcal{R}_{\mathbf{N}_i\mathbf{N}_i}[\tau] = \sigma^2\,\tilde{\mathbf{I}}_{L'}[\tau]$$

where $\tilde{\mathbf{I}}_{L'}[\tau]$ is the $L' \times L'$ matrix which contains ones on the τ^{th} superdiagonal if $0 \le \tau < L'$ or on the $|\tau|^{th}$ subdiagonal if $-L' < \tau \le 0$ and zeros elsewhere [1]. We see that $\mathcal{R}_{\mathbf{NN}}[\tau]$ merely depends of σ^2.

5. GENERAL BSS EQUATIONS

A two-steps BSS method (whitening and rotation) can be classically derived from Eq. (3).

5.1. Whitening

First step consists in finding a whitening matrix. Ideally, this is a $N \times M$ matrix $\mathbf{W}$ such that $\mathbf{W}(\mathbf{A}\mathbf{A}^H)\mathbf{W}^H = \mathbf{I}_N$. In the instantaneous case ($L = 0, L' = 1$), $\mathbf{W}$ can be classically estimated from the eigenelements of the correlation matrix of the observations [1, 8].

In the convolutive case, it appears that such a matrix $\mathbf{W}$ cannot be computed in practice and we rather look for $\mathbf{W}$ as a $N \times M$ matrix satisfying:

$$\mathbf{W}(\mathbf{A}\;\mathbf{B}\;\mathbf{A}^H)\mathbf{W}^H = \mathbf{I}_N \tag{6}$$

[1] We use by convention $\tilde{\mathbf{I}}_{L'}[0] = \mathbf{I}_{L'}$ and $\tilde{\mathbf{I}}_{L'}[\tau] = \mathbf{0}_{L'}$ if $|\tau| \ge L'$.

where $\mathbf{B}$ is a $N \times N$ positive definite block-diagonal matrix with n blocks of dimensions $(L + L')$. We assume for the moment that the quantity $\mathbf{ABA}^H$ can be retrieved from the observations, this assumption is relaxed in Sections 6 and 7.

$\mathbf{W}$ can be computed from $\mathbf{ABA}^H$ by standard subspace analysis like in [1]. We note $[\lambda_1 \dots \lambda_M]$ the eigenvalues of $\mathbf{ABA}^H$, sorted in decreasing order, and $[\mathbf{h}_1 \dots \mathbf{h}_M]$ the corresponding eigenvectors. With $\mathbf{B}$ being a definite positive $N \times N$ matrix and $\mathbf{A}$ being a full rank $M \times N$ matrix with $M \geq N$, $\mathbf{ABA}^H$ is positive semidefinite and has *exactly* M-N zero eigenvalues $[\lambda_{N+1} \dots \lambda_M]$. Then, for $i = 1, \dots, N$ we can write: $\lambda_i{}^{-\frac{1}{2}} \mathbf{h}_i^H \, \mathbf{ABA}^H \, \lambda_i{}^{-\frac{1}{2}} \mathbf{h}_i = 1$ and $\mathbf{W}$ writes:

$$\mathbf{W} = \left[\frac{\mathbf{h}_1}{\sqrt{\lambda_1}}, \dots, \frac{\mathbf{h}_N}{\sqrt{\lambda_N}} \right]^H \tag{7}$$

5.2. Rotation

Let $\mathbf{B}^{\frac{1}{2}}$ denote an arbitrary square root matrix of $\mathbf{B}$ ($\mathbf{B}^{\frac{1}{2}}$ exists because $\mathbf{B}$ is positive). We have $\mathbf{B} = \mathbf{B}^{\frac{1}{2}} (\mathbf{B}^{\frac{H}{2}})$, where $\cdot^{\frac{H}{2}}$ denotes the conjugate transpose of the square root matrix. We note $\mathbf{U}$ the $N \times N$ matrix defined by:

$$\mathbf{U} = \mathbf{W}\,\mathbf{A}\,\mathbf{B}^{\frac{1}{2}} \tag{8}$$

It is shown in Section 5.3 that the sources may be retrieved (up to a filter) from the knowledge of $\mathbf{U}$ and $\mathbf{W}$. We now show a method to retrieve $\mathbf{U}$.

By definition of $\mathbf{W}$, $\mathbf{U}$ is unitary ($\mathbf{U}\,\mathbf{U}^H = \mathbf{I}_N$). Furthermore, $\mathbf{B}$ being definite we have:

$$\mathbf{W}\,\mathbf{A} = \mathbf{U}\,\mathbf{B}^{-\frac{1}{2}} \tag{9}$$

We define "whitened and noise-compensated" covariance matrices $\underline{\mathcal{R}}_{\mathbf{XX}}[t,\tau]$ such that:

$$\underline{\mathcal{R}}_{\mathbf{XX}}[t,\tau] = \mathbf{W} \; [\mathcal{R}_{\mathbf{XX}}[t,\tau] - \mathcal{R}_{\mathbf{NN}}[t,\tau]] \; \mathbf{W}^H \tag{10}$$

With Eq.'s (4) and (9) we have:

$$\begin{aligned} \underline{\mathcal{R}}_{\mathbf{XX}}[t,\tau] &= \mathbf{W}\,\mathbf{A}\,\mathcal{R}_{\mathbf{SS}}[t,\tau]\,\mathbf{A}^H\,\mathbf{W}^H \\ &= \mathbf{U}\,(\mathbf{B}^{-\frac{1}{2}}\,\mathcal{R}_{\mathbf{SS}}[t,\tau]\,\mathbf{B}^{-\frac{H}{2}})\,\mathbf{U}^H \end{aligned} \tag{11}$$

$\mathbf{B}^{-\frac{1}{2}}$, $\mathbf{B}^{-\frac{H}{2}}$ and $\mathcal{R}_{\mathbf{SS}}[t,\tau]$ are $N \times N$ block-diagonal matrices with n blocks of dimension $(L+L') \times (L+L')$. Hence, Eq. (11) shows that $\mathbf{U}$ block-diagonalizes $\underline{\mathcal{R}}_{\mathbf{XX}}[t,\tau]$ for all $(t,\tau) \in \mathbb{Z}^2$.

Thus, $\mathbf{U}$ can be retrieved in theory from the block-diagonalization of any matrix $\underline{\mathcal{R}}_{\mathbf{XX}}[t,\tau]$, with the condition that $\underline{\mathcal{R}}_{\mathbf{XX}}[t,\tau]$ has distinct block eigenvalues (up to a unitary matrix) to prevent from indeterminacy in the columns $\mathbf{U}$. As first introduced in [9], in practice an estimation of $\mathbf{U}$ should rather be computed from the joint block-diagonalization (JBD) of a set of K matrices $\{\underline{\mathcal{R}}_{\mathbf{XX}}[t_i,\tau_i],$ $i = 1...K\}$. JBD provides a more robust estimate of $\mathbf{U}$ with respect to estimation errors on $\underline{\mathcal{R}}_{\mathbf{XX}}[t,\tau]$ (see Sections 6 and 7) and reduces indeterminacies in the same way joint-diagonalization does [1]. JBD provides a matrix $\mathbf{U}_{JBD}$ such that:

$$\mathbf{U}_{JBD} = \mathbf{U}\,\mathbf{P} \tag{12}$$

where $\mathbf{P}$ is a $N \times N$ unitary matrix that models JBD indeterminacies. $\mathbf{P}$ is the product of a block-diagonal unitary matrix with n blocks of dimension $(L+L') \times (L+L')$ with a permutation matrix of these blocks. A Jacobi-like JBD algorithm is presented in [10].

5.3. Retrieving the sources

In this Section we compute estimates of the sources (up to unknown filters) from $\mathbf{U}_{JBD}$ and $\mathbf{W}$. We define the following column vector $\hat{\mathbf{S}}[t]$ of dimension N:

$$\hat{\mathbf{S}}[t] = \mathbf{U}_{JBD}^H\,\mathbf{W}\,\mathbf{X}[t] \tag{13}$$

Eq.'s (12), (9) and (3) yield:

$$\hat{\mathbf{S}}[t] = \mathbf{P}^H\,\mathbf{B}^{-\frac{1}{2}}\,\mathbf{S}[t] + \mathbf{P}^H\,\mathbf{U}^H\,\mathbf{W}\,\mathbf{N}[t] \tag{14}$$

Thus, if we denoise $\hat{\mathbf{S}}[t]$ or neglect the noise contribution, with the notation $\mathbf{F} = \mathbf{P}^H\,\mathbf{B}^{-\frac{1}{2}}$ we have:

$$\hat{\mathbf{S}}[t] \approx \mathbf{F}\,\mathbf{S}[t] \tag{15}$$

$\mathbf{F}$ is a $N \times N$ block-diagonal matrix with n blocks $\mathbf{F}_1, \dots, \mathbf{F}_n$ of dimensions $(L+L') \times (L+L')$. We decompose $\hat{\mathbf{S}}[t]$ into n subvectors of dimension $(L+L')$ such that

$$\hat{\mathbf{S}}[t] = \left[\hat{\mathbf{S}}_1[t]^T, \dots, \hat{\mathbf{S}}_n[t]^T \right]^T$$

and then with Eq. (15), for $i = 1, \dots, n$, we have:

$$\hat{\mathbf{S}}_i[t] = \mathbf{F}_i\,\mathbf{S}_i[t] \tag{16}$$

We recall that $\mathbf{S}_i[t] = [s_i[t], \dots, s_i[t-(L+L')+1]]^T$. Hence, Eq. (16) means that each component of $\hat{\mathbf{S}}_i[t]$ is a FIR filtered version of the i^{th} source $s_i[t]$. The coefficients of the filters are contained in corresponding rows of $\mathbf{F}_i$. Then, for each source $s_i[t]$, we retrieve $(L+L')$ filtered versions of $s_i[t]$. Thus, a further blind SIMO system identification [11] step is required to estimate the original sources instead of filtered versions of them. However we will not deal with this problem in this paper. Note that this filtering arises from the indeterminacy of $\mathbf{B}$ as defined in Eq. (6) and the from JBD indeterminacies modeled by $\mathbf{P}$.

We have shown in this Section that provided the statistics $\mathcal{R}_{\mathbf{XX}}[t,\tau]$, σ^2 and $\mathbf{W}$ (computed from some matrix of the form $\mathbf{A}\,\mathbf{B}\,\mathbf{A}^H$ with $\mathbf{B}$ block-diagonal positive definite), we are able to recover some estimates of the sources up to unknown FIR filters. We now show that previous works [2–4] can be interpreted as practical implementations of the general equations of this section in the stationary ergodic case (Section 6) and in the non-stationary case (Section 7).

6. PRACTICAL IMPLEMENTATION FOR STATIONARY ERGODIC SOURCES

The stationary case is presented in [2], in the noiseless case. In this section the noise is taken into account. When the sources are assumed stationary, we have $\mathcal{R}_{\mathbf{SS}}[t,\tau] = \mathcal{R}_{\mathbf{SS}}[\tau]$ and $\mathcal{R}_{\mathbf{XX}}[t,\tau] = \mathcal{R}_{\mathbf{XX}}[\tau]$. Furthermore when the signals are also ergodic with length T, the covariance matrix of $\mathbf{X}[t]$ is classically estimated by:

$$\mathbf{R}_{\mathbf{XX}}[\tau] = \frac{1}{T-\tau} \sum_{t=1}^{T} \mathbf{X}[t]\,\mathbf{X}[t+\tau]^H$$

(We use bold capital letters when dealing with estimates.) Eq. (4) becomes, for $\tau = 0$:

$$\mathcal{R}_{\mathbf{XX}}[0] = \mathbf{A}\,\mathcal{R}_{\mathbf{SS}}[0]\,\mathbf{A}^H + \sigma^2\,\mathbf{I}_M \tag{17}$$

Thus, with similar subspace analysis as in Section 5.1, the $M-N$ smallest eigenvalues of $\mathcal{R}_{\mathbf{XX}}[0]$ equal σ^2. Then, in practice an estimate $\hat{\sigma}^2$ of σ^2 can be computed as the average value of the $M-N$ smallest eigenvalues of $\mathbf{R}_{\mathbf{XX}}[0]$.

$\mathbf{W}$ is computed like in Section 5.1, using $\mathbf{B} = \mathcal{R}_{\mathbf{SS}}[0]$. Indeed, with Eq. (17) we have then:

$$\mathbf{A}\,\mathbf{B}\,\mathbf{A}^H \approx \mathbf{R}_{\mathbf{XX}}[0] - \hat{\sigma}^2\,\mathbf{I}_M$$

Finally, like for SOBI [1], an estimate of $\mathbf{U}$ is then computed from the JBD of a set of whitened and noise compensated matrices $\underline{\mathbf{R}}_{\mathbf{XX}}(\tau)$ corresponding to several arbitrary lags.

7. PRACTICAL IMPLEMENTATION FOR NON-STATIONARY SOURCES

We now consider the non-stationary case.

7.1. Whitening

We have $\forall t \in \mathbb{Z}$:

$$\mathcal{R}_{\mathbf{XX}}[t,0] = \mathbf{A}\,\mathcal{R}_{\mathbf{SS}}[t,0]\,\mathbf{A}^H + \sigma^2\,\mathbf{I}_M \tag{18}$$

Eq. (18) is similar to Eq. (17) and $\mathbf{W}$ could be estimated like in Section 6, provided an estimate of $\mathcal{R}_{\mathbf{XX}}[t,0]$. However, because of the non-stationarity, $\mathcal{R}_{\mathbf{XX}}[t,0]$ varies with time and it cannot be estimated by ergodic formula. Nonetheless, with Eq. (3) and with the noise being stationary and independent of the sources, we have:

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}[t]\,\mathbf{X}[t]^H \approx \mathbf{A}\,\frac{1}{T}\sum_{t=1}^{T}\mathbf{S}[t]\,\mathbf{S}[t]^H\,\mathbf{A}^H + \sigma^2\,\mathbf{I}_M$$

In practice, if the *realizations* of the source signals are decorrelated, in the sense that $\frac{1}{T}\sum_{t=1}^{T}\mathbf{s}[t]\,\mathbf{s}[t]^H$ is close to diagonality [2], $\frac{1}{T}\sum_{t=1}^{T}\mathbf{S}[t]\,\mathbf{S}[t]^H$ is close to block-diagonality. In that case, we can choose $\mathbf{B} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{S}[t]\,\mathbf{S}[t]^H$, estimate σ^2 like in Section 6 and then $\mathbf{W}$ like in Section 5.1.

At this step, estimates of σ^2 and $\mathbf{W}$ are available and we now focus on the estimation of $\mathcal{R}_{\mathbf{XX}}[t,\tau]$ in the next two subsections.

[2] For example this is often the case for audio signals with a large number of samples.

7.2. Locally stationary sources

When the sources are supposed to be non-stationary but locally stationary and ergodic, the covariance $\mathcal{R}_{\mathbf{XX}}[t,\tau]$ can be estimated locally on subintervals of the observations, if necessary with the use of a smoothing window. This is the approach of [3] which is the generalization of [5] to convolutive mixtures. However, these papers use a non-orthogonal BSS algorithm in the sense that $\mathbf{A}$ is estimated without prewhitening.

7.3. Time-frequency approach

We now consider the case when the assumption of local stationarity is not compliant with the sources. In this context, time-frequency (t-f) signal representations are very useful tools to deal with non-stationary signals. We show in this subsection that the key Eq. (4) has a strictly equivalent formulation in the t-f plane, introducing Spatial Wigner-Ville Spectrum (SWVS). In practice, the SWVS is approximated by Cohen's class Spatial Time-Frequency Distributions (STFDs) and then the separation problem comes down to the work in [4] (in the noiseless case).

7.3.1. Spatial Wigner-Ville Spectrum

The discrete-time/continuous-frequency auto-Wigner-Ville Spectrum (WVS) of a discrete-time scalar signal $x[t]$ is defined for $t \in \mathbb{Z}$ and for $f \in [-\frac{1}{2}, \frac{1}{2}]$ by [12]:

$$\mathcal{WVS}_{xx}[t,f] = 2\sum_{\tau=-\infty}^{+\infty} \mathrm{E}\{x[t+\tau]\,x^\star[t-\tau]\}\,e^{-j4\pi f\tau}$$

where $\cdot^\star$ denotes "conjugate". We define the cross-WVS of two scalar signals $x[t]$ and $y[t]$ by:

$$\mathcal{WVS}_{xy}[t,f] = 2\sum_{\tau=-\infty}^{+\infty} \mathrm{E}\{x[t+\tau]\,y^\star[t-\tau]\}\,e^{-j4\pi f\tau}$$

Hence, the SWVS of the vector signal $\mathbf{x}[t]$ is defined by:

$$\begin{aligned}\mathcal{SWVS}_{\mathbf{xx}}[t,f] &= 2\sum_{\tau=-\infty}^{+\infty} \mathrm{E}\{\mathbf{x}[t+\tau]\,\mathbf{x}^H[t-\tau]\}\,e^{-j4\pi f\tau}\\ &= 2\sum_{\tau=-\infty}^{+\infty} \mathcal{R}_{\mathbf{xx}}[t+\tau,-2\,\tau]\,e^{-j4\pi f\tau} \end{aligned}\tag{19}$$

7.3.2. Time-frequency formulation of the problem

With Eq's. (4) and (19) we have then:

$$\mathcal{SWVS}_{\mathbf{XX}}[t,f] = \mathbf{A}\,\mathcal{SWVS}_{\mathbf{SS}}[t,f]\,\mathbf{A}^H + \mathcal{SWVS}_{\mathbf{NN}}[t,f] \tag{20}$$

Similarly to Eq. (10), we define $\underline{\mathcal{SWVS}}_{\mathbf{XX}}[t,f]$ as:

$$\underline{\mathcal{SWVS}}_{\mathbf{XX}}[t,f] = \mathbf{W}\,\left[\mathcal{SWVS}_{\mathbf{XX}}[t,f] - \mathcal{SWVS}_{\mathbf{NN}}[t,f]\right]\,\mathbf{W}^H \tag{21}$$

And similarly to Eq. (11), we have:

$$\underline{\mathcal{SWVS}}_{\mathbf{XX}}[t,f) = \mathbf{U}\,(\mathbf{B}^{-\frac{1}{2}}\,\underline{\mathcal{SWVS}}_{\mathbf{SS}}[t,f)\,\mathbf{B}^{-\frac{H}{2}})\,\mathbf{U}^H \tag{22}$$

Thus, $\mathbf{U}$ can be estimated by JBD of a set of SWVS matrices $\{\underline{\mathcal{SWVS}}_{\mathbf{XX}}[t_i,f_i), i=1...K\}$ instead of a set of covariance matrices $\{\underline{\mathcal{R}}_{\mathbf{XX}}[t_i,\tau_i], i=1...K\}$.

7.3.3. Estimation of the Spatial Wigner-Ville Spectrum

We now deal with the estimation of $\mathcal{SWVS}_{\mathbf{XX}}[t,f)$ based on one realization of $\mathbf{X}[t]$ only. The SWVS can be interpreted as:

$$\mathcal{SWVS}_{\mathbf{XX}}[t,f) = \mathrm{E}\{\mathbf{D}^{WV}_{\mathbf{XX}}[t,f)\}$$

where $\mathbf{D}^{WV}_{\mathbf{XX}}[t,f)$ is the Spatial Wigner-Ville Distribution, defined for a discrete-time signal $\mathbf{x}[t]$ by:

$$\mathbf{D}^{WV}_{\mathbf{xx}}[t,f) = 2\sum_{\tau=-\infty}^{+\infty}\mathbf{x}[t+\tau]\mathbf{x}[t-\tau]^H\,e^{-j4\pi f\tau}$$

Thus, the Spatial Wigner-Ville Distribution is a rough approximation of the SWVS based on one realization of the signals only. It is shown in [12, 13] that smoothed Wigner-Ville Distributions, i.e., Cohen's class Time-Frequency Distributions (TFDs), yield better estimators of the WVS. For a given smoothing kernel $\phi[t,f)$ we denote the STFD of $\mathbf{x}(t)$:

$$\mathbf{D}^{\phi}_{\mathbf{xx}}[t,f) \overset{\text{def}}{=} \sum_{u=-\infty}^{+\infty}\int_{v=-\frac{1}{2}}^{\frac{1}{2}}\phi[u-t,v-f)\,\mathbf{D}^{WV}_{\mathbf{xx}}[u,v)\,dv$$

$\phi[t,f)$ is chosen of unit energy such that the SWVS and STFDs have same magnitude. Thus, given the estimates $\hat{\mathbf{W}}$ and $\hat{\sigma}^2$, in practice Eq.'s (22) and (21) become:

$$\underline{\mathbf{D}}^{\phi}_{\mathbf{XX}}[t,f) \approx \mathbf{U}(\mathbf{B}^{-\frac{1}{2}}\,\mathbf{D}^{\phi}_{\mathbf{ss}}[t,f)\,\mathbf{B}^{-\frac{H}{2}})\mathbf{U}^H \tag{23}$$

where

$$\underline{\mathbf{D}}^{\phi}_{\mathbf{XX}}[t,f) \overset{\text{def}}{=} \hat{\mathbf{W}}\left(\mathbf{D}^{\phi}_{\mathbf{XX}}[t,f) - \mathbf{D}^{\hat{\sigma}}_{\mathbf{NN}}[t,f)\right)\hat{\mathbf{W}}^H$$

$\mathbf{D}^{\hat{\sigma}}_{\mathbf{NN}}[t,f)$ denotes the estimation of $\mathcal{SWVS}_{\mathbf{NN}}[t,f)$, deduced from $\mathcal{R}_{\mathbf{NN}}[\tau]$ with Eq. (19) but where $\mathcal{R}_{\mathbf{NN}}[\tau]$ is replaced by its estimate given by Eq. (5) where σ is replaced by $\hat{\sigma}$. Eq. (23) is the equation [4] starts from, in the noiseless case.

However, $\mathbf{D}^{\phi}_{\mathbf{SS}}[t,f)$ is only an estimate of $\mathcal{SWVS}_{\mathbf{SS}}[t,f)$ and thus it appears that it is not block-diagonal for *every* t-f location (as opposed to $\mathcal{SWVS}_{\mathbf{SS}}[t,f)$). Hence, we cannot block-diagonalize *any* matrix $\underline{\mathbf{D}}^{\phi}_{\mathbf{XX}}[t,f)$ corresponding to *any* t-f location. Prior to the JBD of the set of matrices $\{\underline{\mathbf{D}}^{\phi}_{\mathbf{XX}}[t_i,f_i), i=1\ldots K\}$, we need to find blindly (i.e., from the observations only) the set of locations $\{(t_i,f_i), i=1\ldots K\}$, for which $\mathbf{D}^{\phi}_{\mathbf{SS}}[t,f)$ is block-diagonal.

7.3.4. Single block auto-terms selection

In [8], in the instantaneous case, it is proposed to search for sources single auto-terms t-f locations. Indeed, it is shown that diagonal matrices $\mathbf{D}^{\phi}_{\mathbf{SS}}[t,f)$ are very likely to have only one non-zero diagonal entry and a criterion is proposed to select corresponding t-f locations from the observations. The interest of sources STFD matrices with only one non-zero diagonal entry is also emphasized in [14]. We propose to generalize the criterion in [8] to the convolutive case by searching for single *block* auto-terms.

We then look for t-f locations (t,f) such that $\mathbf{D}^{\phi}_{\mathbf{SS}}[t,f)$ is block-diagonal with only one non-zero diagonal block, relying on the observations only. These t-f locations are referred to as single block auto-terms locations. Let $[t_0,f_0)$ be a single block auto-term location. One diagonal block only is non-zero in $\mathbf{D}^{\phi}_{\mathbf{SS}}[t,f)$ and thus $[t_0,f_0)$ satisfies:

$$\max_i \|\mathbf{D}^{\phi}_{\mathbf{S}_i\mathbf{S}_i}[t_0,f_0)\|_F = \|\mathbf{D}^{\phi}_{\mathbf{SS}}[t_0,f_0)\|_F$$

where $\|\cdot\|_F$ denotes Frobenius norm.

Furthermore, $\mathbf{B}^{-\frac{1}{2}}$ being block-diagonal, $[t_0,f_0)$ satisfies:

$$C[t_0,f_0) \overset{\text{def}}{=} \frac{\max_i \|\mathbf{B}_i^{-\frac{1}{2}}\,\mathbf{D}^{\phi}_{\mathbf{S}_i\mathbf{S}_i}[t_0,f_0)\,\mathbf{B}_i^{-\frac{H}{2}}\|_F}{\|\mathbf{B}^{-\frac{1}{2}}\mathbf{D}^{\phi}_{\mathbf{SS}}[t_0,f_0)\,\mathbf{B}^{-\frac{H}{2}}\|_F} = 1 \tag{24}$$

We show now how $C[t_0,f_0)$ can be computed from the observations only. Let us define an arbitrary block eigen decomposition of $\underline{\mathbf{D}}^{\phi}_{\mathbf{XX}}[t_0,f_0)$ such that:

$$\underline{\mathbf{D}}^{\phi}_{\mathbf{XX}}[t_0,f_0) = \mathbf{V}_{[t_0,f_0)}\,\mathbf{E}_{[t_0,f_0)}\,\mathbf{V}^H_{[t_0,f_0)}$$

where $\mathbf{V}_{[t_0,f_0)}$ is a $N\times N$ unitary matrix and $\mathbf{E}_{[t_0,f_0)}$ is a block-diagonal matrix with n blocks of dimension $(L+L')\times(L+L')$. With Eq. (23) we have then:

$$\mathbf{B}^{-\frac{1}{2}}\mathbf{D}^{\phi}_{\mathbf{SS}}[t_0,f_0)\,\mathbf{B}^{-\frac{H}{2}} = \mathbf{E}_{[t_0,f_0)}\,\mathbf{Q}_{[t_0,f_0)}$$

where $\mathbf{Q}_{[t_0,f_0)}$ is a unitary $N\times N$ matrix which models block-diagonalization indeterminacies. $\mathbf{Q}_{[t_0,f_0)}$ is the product of a block-diagonal unitary matrix with a permutation matrix of these blocks. Thus, since Frobenius norm is unchanged under unitary matrix multiplication, $\forall(t,f)$ Eq. (24) writes:

$$C[t,f) = \frac{\max_i \|\,[\mathbf{E}_{[t,f)}]_i\,\|_F}{\|\mathbf{E}_{[t,f)}\|_F} \tag{25}$$

where $\mathbf{E}_{[t,f)}$ is the block-diagonal matrix provided by any arbitrary block eigen decomposition of $\underline{\mathbf{D}}^{\phi}_{\mathbf{XX}}[t,f)$ and $[\mathbf{E}_{[t,f)}]_i$ denotes the i^{th} $(L+L')\times(L+L')$ diagonal block of $\mathbf{E}_{[t,f)}$.

Ideally single block auto-terms t-f locations satisfy $C[t,f)=1$. However in practice matrices $\mathbf{D}^{\phi}_{\mathbf{SS}}[t,f)$ can never be strictly block-diagonal. Hence, single block auto-terms t-f locations should be selected as $C[t,f)\geq 1-\varepsilon$, with ε close to zero. [8] propose an optimal selection which consist in selecting $C[t,f)$ local maxima. $\mathbf{U}_{JBD}$ is then computed by JBD of the matrices $\mathbf{D}^{\phi}_{\mathbf{XX}}[t,f)$ corresponding to the selected t-f locations.

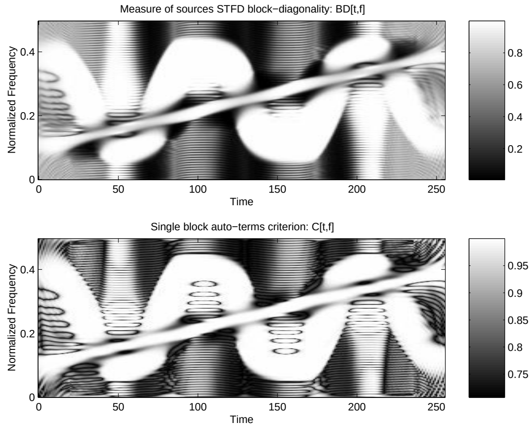

Fig. 1. Evaluation of the selection criterion $C[t, f)$

7.3.5. Selection performance

The performance of our selection criterion is evaluated on a mixture of $n = 2$ analytic sources of length $T = 256$ with $m = 3$ observations. One source is a linear chirp, the other is a sine frequency modulated signal. The convolutive mixing is arbitrarily chosen as:

$$\mathbf{H}[z] = \begin{bmatrix} 1 + 0.8z^{-1} + 0.5z^{-2} & 0.8 + 0.7z^{-1} + 0.4z^{-2} \\ 0.9 + 0.4z^{-1} + 0.6z^{-2} & 1 + 0.9z^{-1} + 0.3z^{-2} \\ 0.7 + 0.6z^{-1} + 0.5z^{-2} & 0.8 + 0.3z^{-1} + 0.6z^{-2} \end{bmatrix}$$

TFDs are Reduced Interference Distributions with Bessel kernel [12]. On first plot of Fig. 1, we draw for $t = 0, \ldots, T-1$ and $f = 0, \frac{1}{T}, \ldots, 0.5 - \frac{1}{T}$ the following measure of "block-diagonality" of $\mathbf{D}^{\phi}_{\mathbf{SS}}[t, f)$:

$$BD[t, f) = \frac{\sum_{i=1}^{n} \|\mathbf{D}^{\phi}_{\mathbf{S}_i\mathbf{S}_i}[t, f)\|_F}{\|\mathbf{D}^{\phi}_{\mathbf{SS}}[t, f)\|_F}$$

On second plot of Fig. 1 we draw criterion $C[t, f)$. We see that maxima of $C[t, f)$ match with maxima of $BD[t, f)$. This means that 1) most block-diagonal matrices $\mathbf{D}^{\phi}_{\mathbf{SS}}[t, f)$ have indeed only one non-zero diagonal block 2) the criterion $C[t, f)$ identifies efficiently these matrices, relying on the observations only.

8. CONCLUSIONS

We have shown that previous works [2–4] admit a unified presentation in the sense that these papers consider different implementations of the general equations we provide in the stochastic context. In particular, in the non-stationary case, we showed that the STFDs used in [4] can be interpreted as estimates of the SWVS. Due to this approximation the STFD of the vector signal $\mathbf{S}[t]$ is not block-diagonal for any t-f location. Hence, we proposed an efficient criterion to select t-f locations where the sources STFDs are block-diagonal, relying on the observations only. The matrices $\underline{\mathbf{D}}^{\phi}_{\mathbf{XX}}[t, f)$ corresponding to the selected locations enter JBD. However, the sources can only be estimated up to unknown filters because of the indeterminacies in the whitening and the JBD. However this estimation may be sufficient for many audio applications such as speech recognition or indexation.

9. REFERENCES

[1] A. Belouchrani, K. Abed-Meraim, J. F. Cardoso, and É. Moulines, "A blind source separation technique based on second order statistics," *IEEE Trans. Signal Processing*, vol. 45, no. 2, pp. 434–444, Feb 1997.

[2] H. Bousbiah-Salah, A. Belouchrani, and K. Abed-Meraim, "Jacobi-like algorithm for blind signal separation of convolutive mixtures," *Electronics Letters*, vol. 37, no. 16, pp. 1049–1050, Aug 2001.

[3] H. Bousbiah-Salah, A. Belouchrani, and K. Abed-Meraim, "Blind separation of non stationary sources using joint block diagonalization," in *Proc. IEEE Workshop on Statistical Signal Processing*, 2001, pp. 448–451.

[4] K. Abed-Meraim, A. Belouchrani, and R. Leyman, *Time-Frequency Signal Analysis and Processing*, chapter Blind Source Separation Using Time Frequency Distributions, Prentice-Hall, Editor: B. Boashash, Jan. 2003.

[5] D. T. Pham and J. F. Cardoso, "Blind separation of instantaneous mixtures of non stationary sources," *IEEE Trans. Signal Processing*, vol. 49, no. 9, pp. 1837–1848, Sep 2001.

[6] A. Belouchrani and M. G. Amin, "Blind source separation based on time-frequency signal representations," *IEEE Trans. Signal Processing*, vol. 46, no. 11, pp. 2888–2897, Nov 1998.

[7] J. F. Cardoso, "Multidimensional independent component analysis," in *Proc. ICASSP*, 1998.

[8] A. Holobar, C. Févotte, C. Doncarli, and D. Zazula, "Single autoterms selection for blind source separation in time-frequency plane," in *Proc. EUSIPCO*, 2002, vol. 1, pp. 565–568.

[9] 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 Letters*, vol. 4, no. 9, pp. 266–268, Sept 1997.

[10] A. Belouchrani, K. Abed-Meraim, and Y. Hua, "Jacobi-like algorithms for joint block diagonalization: Application to source localization," in *Proc. ISPACS*, Nov. 1998.

[11] K. Abed-Meraim, W. Qiu, and Y. Hua, "Blind system identification," *Proceedings of the IEEE*, vol. 85, no. 8, pp. 1310–1322, Aug. 1997.

[12] P. Flandrin, *Time-Frequency/Time-Scale Analysis*, Academic Press, San Diego, 1999.

[13] A. M. Sayeed and D. L. Jones, "Optimal kernels for nonstationary spectral estimation," *IEEE Trans. Signal Processing*, vol. 43, no. 2, pp. 478–491, Feb 1995.

[14] L.-T. Nguyen, A. Belouchrani, K. Abed-Meraim, and B. Boashash, "Separating more sources than sensors using time-frequency distributions," in *Proc. ISSPA*, 2001.