



**HAL**  
open science

# Blind Signal Separation for Digital Communication Data

Antoine Chevreuil, Philippe Loubaton

► **To cite this version:**

Antoine Chevreuil, Philippe Loubaton. Blind Signal Separation for Digital Communication Data. Rama Chellappa, Sergios Theodoridis. Academic Press Library in Signal Processing, Academic Press, pp.135-186, 2014, 978-0123965004. 10.1016/B978-0-12-396500-4.00004-1 . hal-00790078

**HAL Id: hal-00790078**

**<https://hal.science/hal-00790078v1>**

Submitted on 17 Oct 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# BLIND SIGNAL SEPARATION FOR DIGITAL COMMUNICATION DATA

A. CHEVREUIL AND P. LOUBATON

ABSTRACT. Blind source separation, often called independent component analysis, is a main field of research in signal processing since the eighties. It consists in retrieving the components, up to certain indeterminacies, of a mixture involving statistically independent signals. Solid theoretical results are known; besides, they have given rise to performant algorithms. There are numerous applications of blind source separation. In this contribution, we particularize the separation of telecommunication sources. In this context, the sources stem from telecommunication devices transmitting at the same time in a given band of frequencies. The received data is a mixed version of all these sources. The aim of the receiver is to isolate (separate) the different contributions prior to estimating the unknown parameters associated with a transmitter. The context of telecommunication signals has the particularity that the sources are not stationary but cyclo-stationary. Now, in general, the standard methods of blind source separation assume the stationarity of the sources. In this contribution, we hence make a survey of the well-known methods and show how the results extend to cyclo-stationary sources.

## 1. INTRODUCTION

**1.1. Generalities on Blind Source Separation.** The goal of blind source separation is to retrieve the components of a mixture of independent signals when no a priori information is available on the mixing matrix. This question was introduced in the eighties in the pioneering works of C. Jutten and J. Herault [32]. Since then, Blind Source Separation (BSS), also called independent component analysis, was developed by many research teams in the context of various applicative contexts. The purpose of this chapter is to present BSS methods that have been developed in the past in the context of digital communications. In this case,  $K$  digital communication devices sharing the same band of frequencies transmit simultaneously  $K$  signals. The receiver is equipped with  $M \geq K$  antennas, and has to retrieve a part of (or even all) the transmitted signals. The use of BSS techniques appears to be relevant when the receiver has no a priori knowledge on the channels between the transmitters and the receiver. As many digital communication systems use training sequences which allow one to estimate the channels at the receiver side, blind source separation is in general not a very useful tool. However, it appears to be of particular interest in contexts such as spectrum monitoring or passive listening in which it is necessary to characterize unknown transmitters (estimation of technical parameters such as the carrier frequency, symbol rate, symbol constellation,...) interfering in a certain bandwidth. For this, it is reasonable to try to firstly retrieve the transmitted signals, and then to analyse each of them in order to characterize the system it has been generated by. In this chapter, we provide a comprehensive introduction to the blind separation techniques that can be used to achieve the first

step.

In order to explain the specificity of the problems we address in the following, we first recall what are the most classical BSS methodologies. The observation is a discrete-time  $M$ -variate signal  $\mathbf{y}(n)$  defined as  $\mathbf{y}(n) = \mathbf{H}\mathbf{s}(n)$  where the components of the  $K$ -dimensional ( $K \leq M$ ) time series  $\mathbf{s}(n) = (s_1(n), \dots, s_K(n))^T$  represent  $K$  signals which are statistically independent. The signal  $\mathbf{y}$  is thus an *instantaneous* mixture of the  $K$  independent source signals  $(s_k)_{k=1, \dots, K}$  in the sense that  $\mathbf{y}(n)$  only depends on the value of  $\mathbf{s}$  at time  $n$ . The signal  $\mathbf{y}$  is said to be a convolutive mixture of the  $K$  independent source signals  $(s_k)_{k=1, \dots, K}$  if  $\mathbf{y}(n) = \sum_{\ell} \mathbf{H}_{\ell} \mathbf{s}(n-\ell) = [\mathbf{H}(z)]\mathbf{s}(n)$  where  $\mathbf{H}(z)$  represents the transfer function  $\mathbf{H}(z) = \sum_{\ell} \mathbf{H}_{\ell} z^{-\ell}$ . For the sake of simplicity, we just consider the context of instantaneous mixtures *in this introductory section*. The goal of blind source separation is to retrieve the signals  $(s_k)_{k=1, \dots, K}$  from the sole knowledge of the observations. Fundamental results of Darmois (see e.g. [12]) show that if the source signals are non Gaussian, then it is possible to achieve the separation of the sources by adapting a matrix  $\mathbf{G}$  in such a way that the components of  $\mathbf{r}(n) = \mathbf{G}\mathbf{y}(n)$  are statistically independent. For this, it has been shown that it is sufficient to optimize over  $\mathbf{G}$  a function, usually called a contrast function, that can be expressed in terms of certain moments of the joint probability distribution of  $\mathbf{r}(n)$ . A number of successful contrast functions have been derived in the case where the signal  $(s_k)_{k=1, \dots, K}$  are stationary sequences [11, 12, 17, 7]. However, it will be explained below that in the context of digital communications, the signals  $(s_k)_{k=1, \dots, K}$  are *not stationary, but cyclostationary*, in the sense that their statistical properties are almost periodic function of the time index. For example, for each  $k$ , the sequence  $n \mapsto \mathbb{E} [|s_k(n)|^2]$  appears to be a superposition of sinusoids whose frequencies, called cyclic frequencies, depend on the symbol rate of transmitter  $k$ , and are therefore unknown at the receiver side. The cyclostationarity of the  $(s_k)_{k=1, \dots, K}$  induces specific methodological difficulties that are not relevant in other applications of blind source separation.

**1.2. Illustration of the potential of BSS techniques for communication signals.** The example we provide is purely academical. We consider the transmission of two BSPK sequences modulated with a Nyquist raised-cosine filter (see section 2.1) whose symbol period is  $T = 1$  and roll-off factor is fixed to 50%. The energy per symbol equals  $\mathcal{E}_1$  for the first source and  $\mathcal{E}_2$  for the second source. The receiver has  $M = 2$  antennas and the channel between the source no  $i$  and the antenna no  $j$  is a delay times a real constant  $h_{i,j}$ . After sampling at  $T/3$ , the noiseless model of the received data is  $\mathbf{Y}(n) = \mathbf{H}\mathbf{s}(n)$  as specified in the introduction, where the component  $(i, j)$  of  $\mathbf{H}$  is  $h_{i,j}$  (more details are provided in section 2.3). An additive white noise Gaussian corrupts the model whose variance is  $\frac{N_0}{T_e}$  where  $\frac{\mathcal{E}_1}{N_0}$  is fixed to 100dB (we purposely fixed the noise level to a low value in order to show results that can be graphically interpreted). Moreover,  $\frac{\mathcal{E}_1}{\mathcal{E}_2} = 3dB$ .

Suppose in a first step that the channel is ideal such that the mixing matrix  $\mathbf{H}$  is the identity matrix. We may have a look at the eye diagrams of the two components of the received data. We obtain the figure 1.1. This is almost a perfectly opened eye since the noise is negligible. We may also have a look at a 2D-histogram of the data. Notice that the components of  $\mathbf{X}(n)$  are not stationary. We hence down-sample these data by a factor 3 in order to have stationary data. We plot the 2D-histogram: see Figure 1.2. As the two components are independent,

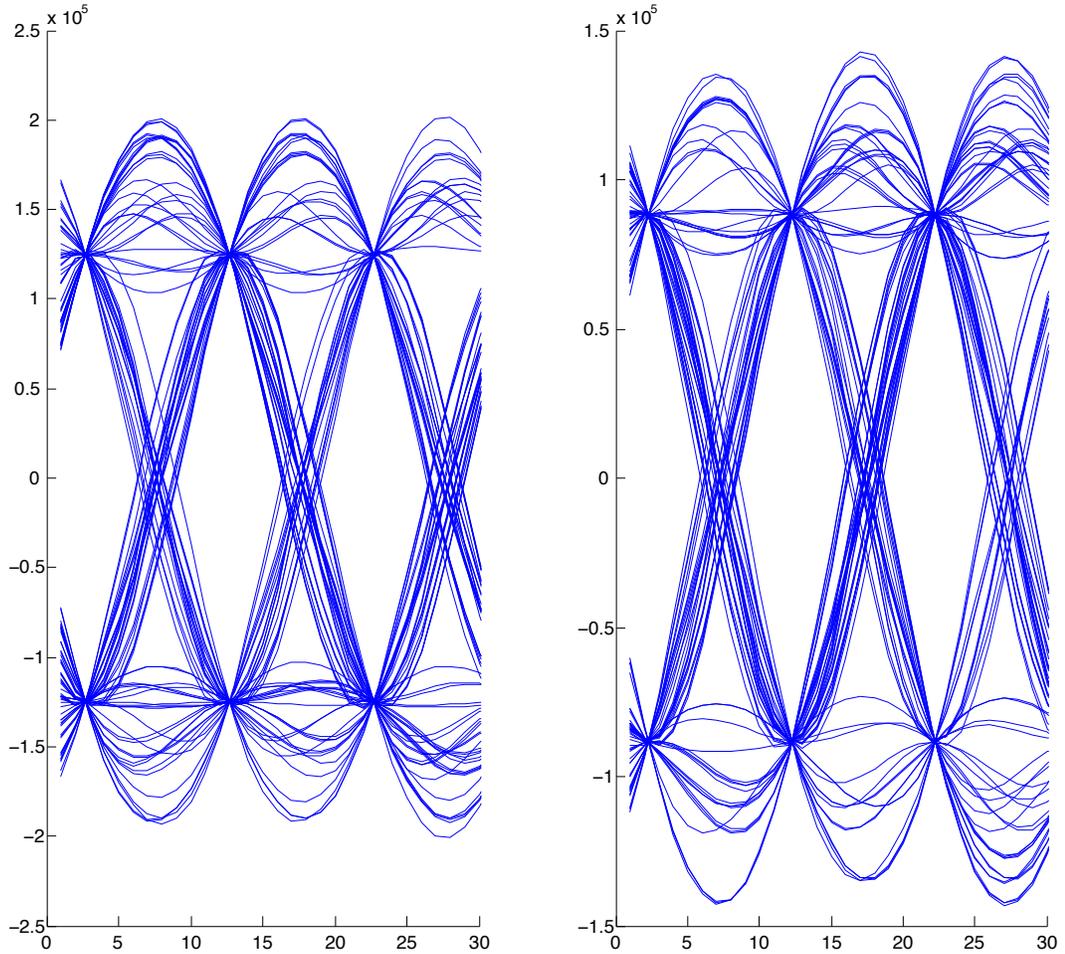


FIGURE 1.1. eye diagrams.  $\frac{\mathcal{E}_s}{N_0} = 100dB$ . Ideal channels. Left: first component. Right: second component.

their joint probability density function (pdf) is separable which seems to be the case in view of the figure.

Let us now consider the case of the channel matrix:

$$\mathbf{H} = \begin{pmatrix} 1 & 0.7 \\ 0.5 & 1 \end{pmatrix}.$$

We obtain the figures 1.3 and 1.4 respectively for the eye diagrams and the 2D-histograms. Clearly the channels are severe and close the eyes. Moreover, the pdf is obviously not separable, which attests to the non independency of the two components of  $\mathbf{X}(n)$ .

We run the JADE algorithm (see sections 3.5 and 3.8) on the data (the observation duration is fixed to 1000 symbols): we obtain a  $2 \times 2$  matrix  $\mathbf{G}$  such that, theoretically at least,  $\mathbf{GH}$  should be diagonal. We form the data  $\mathbf{Y}(n) = \mathbf{GX}(n)$

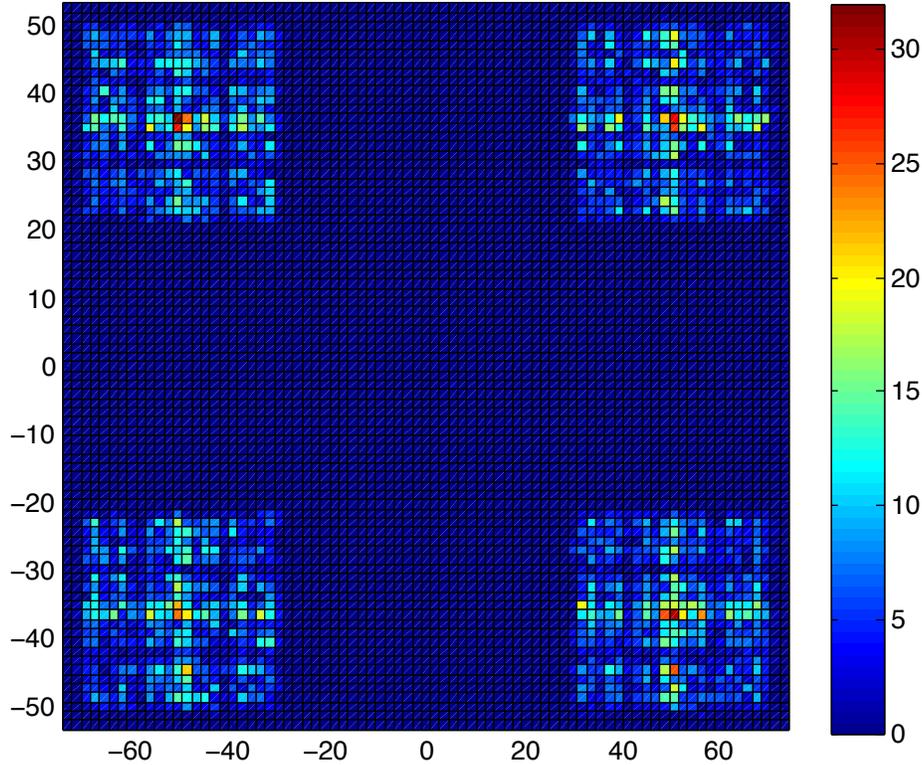


FIGURE 1.2. 2D-histogram.  $\frac{\epsilon_1}{N_0} = 100dB$ . Ideal channels.

and plot the figures 1.5 and 1.6. The eyes have been opened and the joint pdf is hence separable. This is not a surprise since we have computed the resulting matrix  $\mathbf{GH}$  :

$$\mathbf{GH} \approx \begin{pmatrix} 0.8439 & -0.0129 \\ -0.0039 & 1.1829 \end{pmatrix}$$

which is close to a diagonal matrix. We need to explain why BSS has been successfully achieved in this simple example and why it can also be achieved in much more difficult contexts.

**1.3. Organisation of the paper.** This chapter is organized as follows. In Section 2, we provide the model of the signals which are supposed to be linear modulations of symbols (Section 2.1). We discuss the statistics of the sampled versions of the transmitted sources in Section 2.2: in general, a sampled version is cyclo-stationary and we provide the basic tools and notation used along the paper. The model of the received data is specified in Section 2.3: 1) If the propagation channel between each transmitter and the receiver is a single path channel, the received signal is an instantaneous mixture of the transmitted signals 2) if at least one of the propagation channel is a multipath channel, the mixture appears to be convolutive. Besides, we

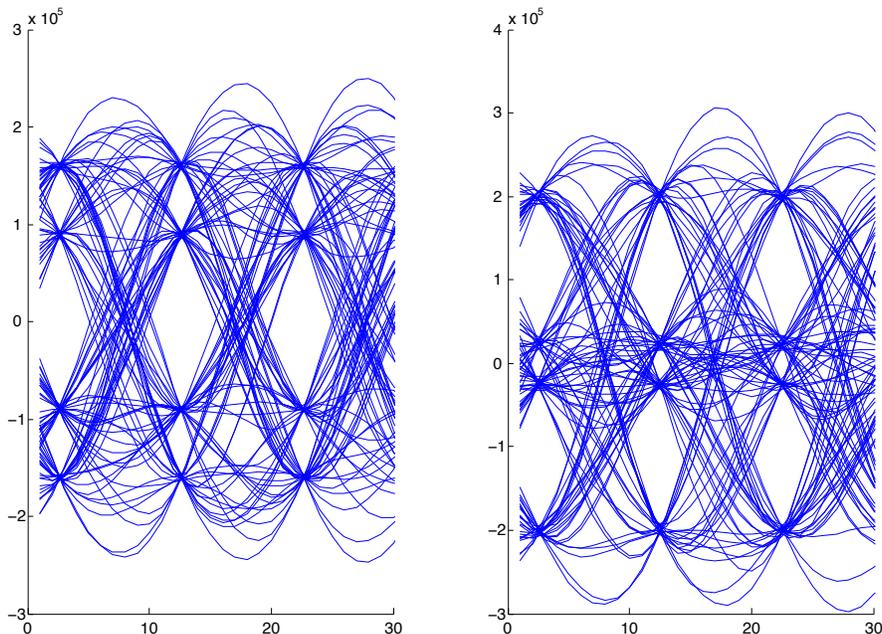


FIGURE 1.3. eye diagrams.  $\frac{\mathcal{E}_1}{N_0} = 100dB$ . Channel  $\mathbf{H}$  . Left: first component. Right: second component.

discuss the assumptions under which the received data are stationary . In general, however, the data are cyclo-stationary with unknown cyclic frequencies.

The case of instantaneous mixtures is addressed in Section 3. When the sources are independent and identically distributed (i.i.d.) (this case is discussed in section 2.3), and that strong a priori information on the constellations are known, it is possible to provide algebraic solutions to the BSS problem, e.g. the Iterative Least Squares Projections (ILSP) algorithm or the Algebraic Constant Modulus Algorithms (ACMA): these methods are explained in section 3.2. In Section 3.3, we consider the case of second-order methods (one of the advantages of these latter is that they are robust to the cyclo-stationarities, hence can be applied to general scenarios): the outlines of one of the most popular approach, the Second-Order Blind Identification (SOBI) algorithm, which consists in estimating the mixing matrix from the autocorrelation function of the received signal. This approach is conceptually simple, and the corresponding scheme allows one to identify the mixture and hence, to separate the source signals. That SOBI is rarely considered for BSS of digital communication signals is explained. The subsections that follow cope with BSS methods based on fourth-order cumulants. They are called «direct» BSS methods since they provide estimates of the sources with no prior estimation of the unknown channel matrix. For pedagogical and historical reasons, we firstly cope with the very particular case of stationary signals. One-by-one methods based are explained (section 3.4) and are shown to be convergent; the associated deflation procedure is introduced and an improvement is presented. Global methods (also called *joint separating methods*) aim at separating jointly the  $K$  sources: they

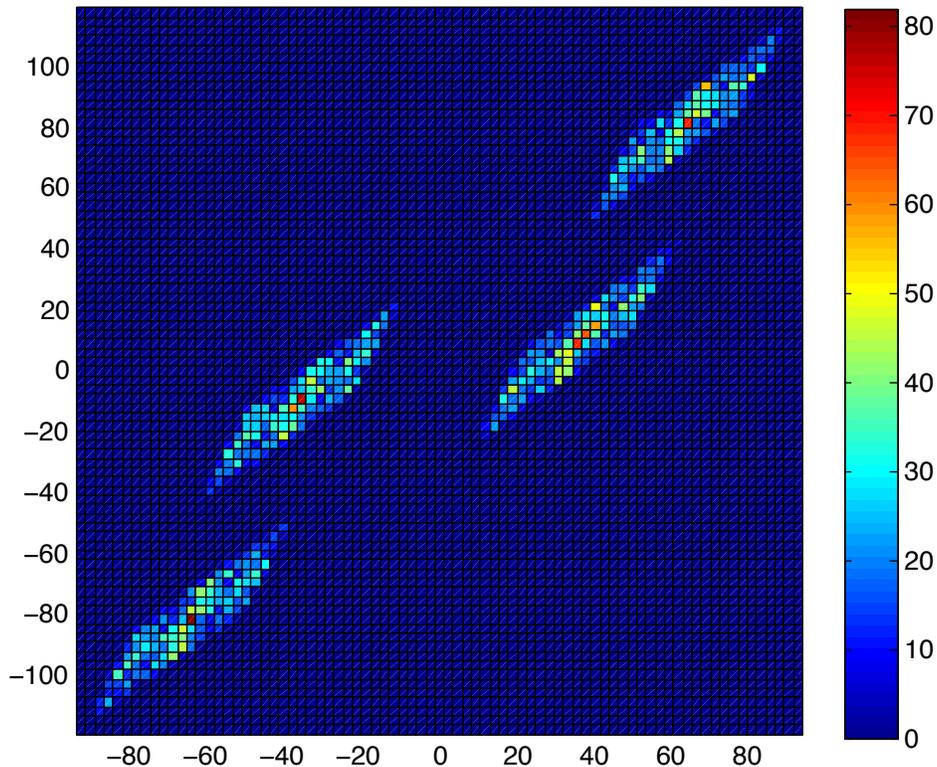


FIGURE 1.4. 2D-histogram.  $\frac{\mathcal{E}_1}{N_0} = 100dB$ . Channel **H**.

are depicted in Section 3.5; these approaches are based on the minimization of well chosen contrast functions over the set of  $K \times K$  unitary matrices: the famous Joint Approximate Diagonalization of Eigenmatrices (JADE) algorithm is presented, since it represents a touchstone in the domain of BSS. When the sources are cyclo-stationary, which is really the interesting point for the context of this paper, the preceding «stationary» methods (one-by-one and global) are again considered. The following problem is addressed: do the convergence result still hold when the algorithms are fed by cyclo-stationary data instead of stationary ones? Sufficient conditions are shown to assure the convergence: semi-analytical computations (Section 3.9) prove that the conditions in question hold true.

In Sections 4 and 5, the case of convolutive mixtures is addressed. In certain particular scenarios, e.g. sparse channels, the gap between the instantaneous case and the convolutive one can be bridged quite directly (section 4). More precisely, if the delays of the various multipaths are sufficiently spread out on the one hand and if, on the other hand, the number of antennas of the receiver is large enough, it is still possible to formulate the source separation problem as the separation of a certain instantaneous mixture. If these conditions do not hold, we face a real

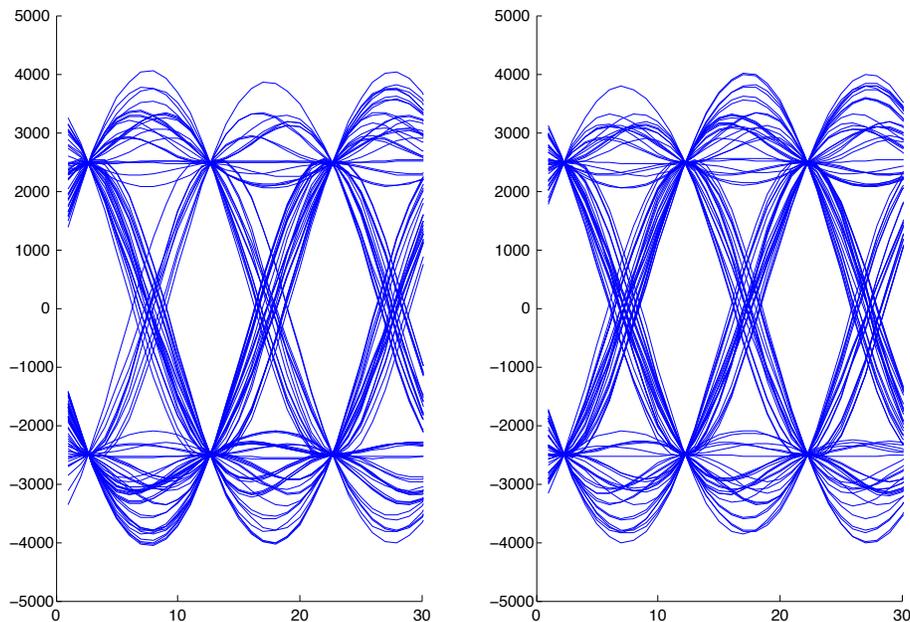


FIGURE 1.5. eye diagramm.  $\frac{\xi_1}{N_0} = 100dB$ . Channel  $\mathbf{H}$ . Left: first component. Right: second component.

convolutive mixture, i.e. the received data are the output of a Multi-Input/Multi-Output (MIMO) unknown filter driven by jointly independent (cyclo-)stationary sources. Due to their historical and theoretical importance, we present algebraic methods (section 5.1) when the data are stationary. Under this latter assumption, the identification of the unknown transfer function can be achieved using standard methods using the Moving Average (MA) or Auto-Regressive (AR) properties: see section 5.2. The famous subspace method, introduced in section 5.3, is based on second-order moments and can be used for general cyclo-stationary data; its inherent numerical problems are discussed. In section 5.4, global direct methods are evoked (temporal domain and frequency domain) for stationary data. In section 5.5, the case of one-to-one methods previously introduced in section 3.4 is extended to the convolutive case and positive results for BSS are provided. The results are further extended for the cyclo-stationary case in section 5.6 where convergence results are shown.

In section 7, we discuss several points that have not been developed in the core of the paper. Further bibliographic entries are provided.

## 2. SIGNALS

We have specified in the Introduction that the domain of source separation is not restricted to the context of telecommunication signals. In the following, however, most of the results apply specifically to digital telecommunication signals.

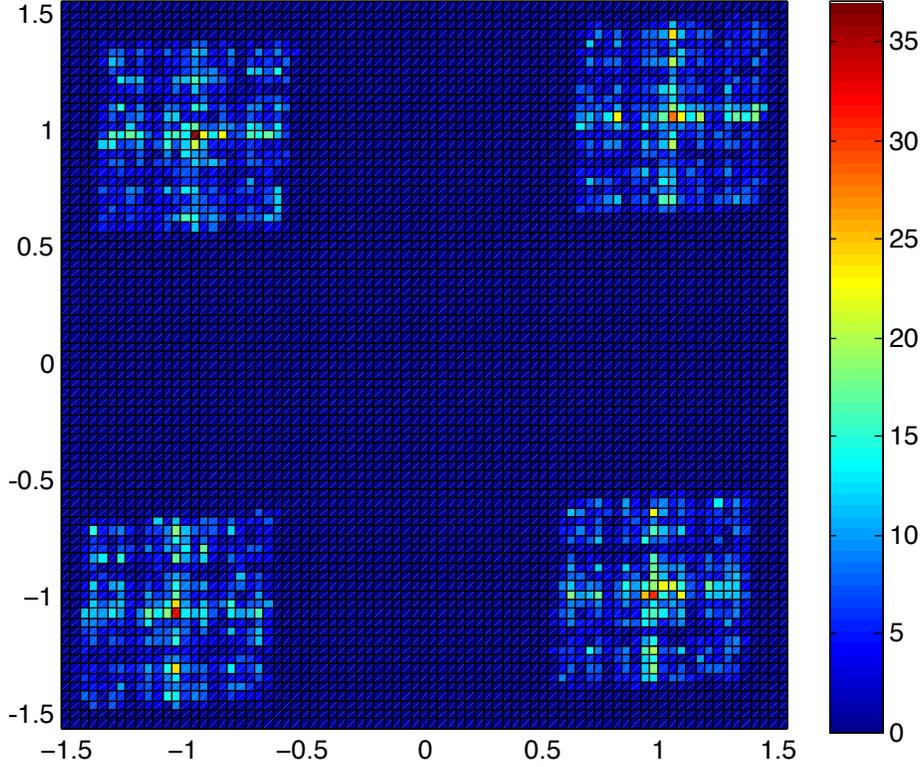


FIGURE 1.6. 2D-histogram.  $\frac{\xi_1}{N_0} = 100dB$ . After the JADE algorithm.

**2.1. Source signals. Basic assumptions .** We assume that  $K$  digital telecommunication devices simultaneously transmit information in the same band of frequencies. For  $t \in \mathbb{R}$ , we denote by  $s_{a,k}(t)$  the complex envelope of the  $k$ -th transmitted signal («the  $k$ -th source»). The subscript «a» in  $s_{a,k}(t)$  underlines that the signal is «analog». Throughout this contribution,  $s_{a,k}(t)$  is supposed to stem from a linear modulation of a sequence of symbols. The model is hence:

$$(2.1) \quad \forall t \in \mathbb{R} \quad s_{a,k}(t) = \sum_n d_k(n) c_{a,k}(t - nT_k).$$

In this latter equation,  $d_k(n)$  is a sequence of symbols belonging to a certain constellation. The function  $c_{a,k}$  is a shaping function and  $T_k$  is the duration of a symbol  $d_k(n)$ . We denote by  $f_k$  the carrier frequency associated with the  $k$ -th source. Along this contribution, the following assumptions and notation are adopted.

ASSUMPTIONS ON THE SOURCE SIGNALS.

For a given index  $k$ , the sequence  $d_k(n)$  is assumed to be independent and identically distributed (i.i.d.). We assume that it has zero mean  $E[d_k(n)] = 0$ . With no

restriction at all, the normalization  $E \left[ |d_k(n)|^2 \right] = 1$  holds. We also suppose that it is second-order complex-circular in the sense that

$$(2.2) \quad E \left[ d_k(n)^2 \right] = 0.$$

It is undoubtedly a restriction to impose the condition (2.2) especially in the telecommunication context of this paper; indeed, the BPSK modulation for instance does not verify (2.2). Some points on the extensions to general non circular mixtures are provided in Section 7.1.

The Kurtosis of the symbol  $d_k(n)$ , is defined as

$$\kappa(d_k) = \text{cum}(d_k(n), d_k(n)^*, d_k(n), d_k(n)^*)$$

where the fourth-order cumulant of a complex-valued random variable  $X$  is defined when it makes sense, as  $\text{cum}(X, X^*, X, X^*) = E \left[ |X|^4 \right] - 2 \left( E \left[ |X|^2 \right] \right)^2 - |E \left[ X^2 \right]|^2$ : see for instance [44]. Here, by the circularity assumption 2.2, we have

$$\kappa(s_k) = E \left[ |d_k(n)|^4 \right] - 2 \left( E \left[ |d_k(n)|^2 \right] \right)^2.$$

We assume that we have, for any index  $k$ :

$$(2.3) \quad \kappa(d_k) < 0.$$

This inequality is given as an assumption; it has more the flavour of a result, since we do not know complex-circular constellations such that (2.3) is not satisfied.

We may now come to the key assumption: the sources  $d_1, d_2, \dots, d_K$  are mutually independent.

Concerning the shaping filter,  $c_{a,k}$  we suppose that  $c_{a,k}$  is a square-root raised cosine with excess bandwidth (also called *roll-off factor*)  $0 \leq \gamma_k \leq 1$ .

**2.2. Cyclo-stationarity of a source.** In this short paragraph, we drop the index of the source and  $s_a, T, c_a, d(n)$  refer to respectively  $s_{a,k}, T_k, c_{a,k}, d_k(n)$ . Thanks to Equation (2.1), it is quite obvious that  $s_a(t)$  and  $s_a(t+T)$  are similarly distributed since  $d_k(n)$  is i.i.d. This simple reasoning applies to any vector  $(s_a(t_1), \dots, s_a(t_m))$  whose distribution equals this of  $(s_a(t_1+T), \dots, s_a(t_m+T))$ . This shows that the process  $s_a$  is cyclo-stationary in the strict sense with period  $T$ . In particular, its second and fourth-order moments evolve as  $T$ -periodic functions of the time. Let us focus on the second-order moments:  $t \mapsto E \left[ s_a(t+\tau)s_a(t)^* \right]$  is hence a periodic function with period  $T$ . We let its Fourier expansion be

$$(2.4) \quad E \left[ s_a(t+\tau)s_a(t)^* \right] = \sum_{m \in \mathbb{Z}} R_{s_a}^{(m/T)}(\tau) e^{i2\pi m t/T}$$

where  $R_{s_a}^{(m/T)}(\tau)$  is called «cyclo-correlation» of  $s_a$  at cyclic frequency  $m/T$  and time lag  $\tau$ . We have the reverse formula:

$$R_{s_a}^{(m/T)}(\tau) = \frac{1}{T} \int_0^T E \left[ s_a(t+\tau)s_a(t)^* \right] e^{-i2\pi m t/T} dt$$

or

$$R_{s_a}^{(m/T)}(\tau) = \lim_{\Delta \rightarrow \infty} \frac{1}{\Delta} \int_{-\Delta/2}^{\Delta/2} E \left[ s_a(t+\tau)s_a(t)^* \right] e^{-i2\pi m t/T} dt.$$

Generally, we may introduce the cyclo-correlation at any cyclic frequency  $\alpha$ :

$$R_{s_a}^{(\alpha)}(\tau) = \lim_{\Delta \rightarrow \infty} \frac{1}{\Delta} \int_{-\Delta/2}^{\Delta/2} E[s_a(t+\tau)s_a(t)^*] e^{-i2\pi\alpha t} dt.$$

In the case of  $s_a$  given by Equation (2.1),  $R_{s_a}^{(\alpha)}(\tau)$  is identically zero for cyclic frequencies  $\alpha$  that are not multiples of  $1/T$ . In passing, we have the following symmetry:

$$(2.5) \quad R_{s_a}^{(\alpha)}(\tau) = R_{s_a}^{(-\alpha)}(-\tau)^*.$$

Let us inspect a bit further a main specificity of linear modulations on the Fourier expansion of Equation (2.4). In this respect, denote by  $\mathbf{f} \mapsto S_{s_a}^{(\alpha)}(\mathbf{f})$  the Fourier transform of the cyclo-correlation function  $\tau \mapsto R_{s_a}^{(\alpha)}(\tau)$ . We have, after elementary calculus (see [21, 41]):

$$S_{s_a}^{(m/T)}(\mathbf{f}) = \frac{1}{T} \hat{c}_a(\mathbf{f}) \hat{c}_a(\mathbf{f} - m/T)^*,$$

where  $\hat{c}_a(\mathbf{f}) = \int_{\mathbb{R}} c_a(t) e^{-i2\pi\mathbf{f}t} dt$  is the Fourier transform of  $c_a$ . This formula is visibly a generalization of the so-called Bennett Equality (see [45] Section 4.4.1) that gives the power spectral density of  $s_a$ : indeed, in the above equation, if one takes  $m = 0$ , we obtain  $S_{s_a}^{(0)}(\mathbf{f}) = \frac{1}{T} |\hat{c}_a(\mathbf{f})|^2$  which is the power spectral density. An important consequence was underlined in [41]:

**Lemma 1.** *For any excess bandwidth factor  $\gamma$  such that  $\gamma \leq 1$ , we have:*

$$\forall |m| > 1 \quad R_{s_a}^{(m/T)}(\tau) \text{ is uniformly zero.}$$

*In other words, the cyclic frequencies of  $s_a(t)$  given by Equation (2.1) belong to the set  $\{0, \pm \frac{1}{T}\}$ .*

The proof is obvious since the support of  $\hat{c}_a(\mathbf{f})$  is  $[-\frac{1+\gamma}{2T}, \frac{1+\gamma}{2T}]$  with  $\gamma < 1$  hence the supports of  $\hat{c}_a(\mathbf{f})$  and  $\hat{c}_a(\mathbf{f} - m/T)^*$  do not overlap except if  $m = 0, \pm 1$ .  $\square$

We deduce from what precedes some consequences on the second-order statistics of a *sampled version* of a source. In this respect, we denote by  $T_e$  any sampling period; the discrete-time signal associated with  $s_a(t)$  is hence  $s(n) = s_a(nT_e)$  for  $n \in \mathbb{Z}$ . Thanks to Lemma 1, the expansion (2.4) may be re-written as:

$$(2.6) \quad \forall \ell \in \mathbb{Z} \quad E[s(n+\ell)s(n)^*] = \sum_{m \in \{-1, 0, 1\}} R_{s_a}^{(m/T)}(\ell T_e) e^{i2\pi n m \alpha}$$

where we let  $\alpha$  be  $\alpha = T_e/T$ . We distinguish between three cases:

- (1) If  $\alpha = T_e/T$  is a integer, the three terms of the r.h.s. of (2.6) all aggregate in a single term, making the function  $E[s(n+\ell)s(n)^*]$  not depend on the time index  $n$ . This is not surprising since the condition  $T_e = pT$  where  $p$  is a non-null integer corresponds to a strict-sense stationary signal (see the polyphase decomposition in [60]). In particular, if  $T_e = T$ , we have:

$$(2.7) \quad s(n) = [c_0(z)] d(n)$$

where  $c_0(z) = \sum_n c_a(nT) z^{-n}$ . In the following, we will not study the case  $T_e = pT$  with  $p \geq 2$ .

- (2) If  $\alpha \neq 0$  modulo  $1/2$ : unless  $\alpha$  is rational, it cannot be said that, as a function of  $n$ ,  $E[s(n+\ell)s(n)^*]$  is periodic: it is called an almost periodic function [16] and  $s(n)$  is hence called almost periodically correlated, having  $0, \pm\alpha$  as cyclic-frequencies. We introduce

$$(2.8) \quad R_s^{(\alpha)}(\ell) = \langle E[s(n+\ell)s(n)^*] e^{-i2\pi\alpha n} \rangle_n$$

where the operator  $\langle \cdot \rangle$  is the time-averaging one, i.e. for a complex-valued (deterministic) series  $u(n)$

$$\langle u(n) \rangle_n = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} u(n)$$

when the limit makes sense. As the three cyclic-frequencies in the expansion (2.6) are distinct modulo 1 then we obtain

$$\forall m \in \{0, \pm 1\} \quad R_s^{(mT_e/T)}(\ell) = R_{s_a}^{(m/T)}(\ell T_e).$$

which reminds us of the Shannon sampling theorem.

- (3) If  $\alpha = \frac{1}{2}$  it turns out that, similarly to the previous case, the discrete-time source is cyclo-stationary, having  $0, \pm\frac{1}{2}$  as cyclic-frequencies. Moreover  $R_s^{(0)}(\ell) = R_{s_a}^{(0)}(\ell T_e)$  and  $R_s^{(1/2)}(\ell) = R_s^{(-1/2)}(\ell) = R_{s_a}^{(1/T)}(\ell T_e) + R_{s_a}^{(-1/T)}(\ell T_e)$ .

**2.3. Received signals.** The receiver is equipped with  $M$  antennas, the number of antennas being as big as the number of sources, i.e.  $M \geq K$  (see section 2.3.2). We denote by

$$\mathbf{y}_a(t) = \begin{pmatrix} y_a^{(1)}(t) \\ \vdots \\ y_a^{(m)}(t) \end{pmatrix}$$

the complex envelope of the received  $M \times 1$  vector computed at a frequency of demodulation denoted by  $f_0$ . We consider that  $\mathbf{y}_a(t)$  obeys the linear model

$$\mathbf{y}_a(t) = \sum_{k=1}^K \mathbf{y}_{a,k}(t)$$

where  $\mathbf{y}_{a,k}(t)$  is the contribution of the  $k$ -th source to the observation. We further assume that  $\mathbf{y}_{a,k}(t)$  stems from delayed/attenuated versions of  $s_{a,k}(t)$ . In this respect, we may write  $y_{a,k}^{(m)}(t)$ , the component of  $\mathbf{y}_{a,k}(t)$  associated with the  $m$ -th sensor, as:

$$(2.9) \quad y_{a,k}^{(m)}(t) = \sum_{\ell=1}^{L_k} \lambda_{k,m,\ell} e^{-i2\pi f_0 \tau_{k,m,\ell}} \tilde{s}_{a,k}(t - \tau_{k,m,\ell})$$

where the index  $\ell$  represents the path index,  $L_k$  the number of paths associated with the source no  $k$ ,  $\lambda_{k,m,\ell}$  an attenuation factor and  $\tau_{k,m,\ell}$  the delay of the propagation along the path no  $\ell$  between the source no  $k$  and the sensor no  $m$ . In this latter equation,  $\tilde{s}_{a,k}(t)$  is the complex envelope of the modulated signal  $s_{a,k}(t)e^{i2\pi f_k t}$  at the demodulation frequency, i.e.  $\tilde{s}_{a,k}(t) = s_{a,k}(t)e^{i2\pi(f_k - f_0)t}$ .

2.3.1. *Models of the sampled data.* We distinguish between two cases.

- (1) **Instantaneous mixture.** This scenario holds when the signal  $\mathbf{y}_{a,k}(t)$  evolves in a linear space of dimension 1, that is when the components  $\tilde{s}_{a,k}(t - \tau_{k,m,\ell})$  in (2.9) do not depend neither on  $\ell$  nor on  $m$ . This holds when there exists  $\tau_k$  such that  $|\tau_{k,m,\ell} - \tau_k| \ll T_k$  for all indices  $m, \ell$ . This happens, for instance, when there is a single path ( $L_k = 1$ ) and the transmitted signal is narrow-band ( $f_k \gg 1/T_k$ ). In this case we have

$$\mathbf{y}_a(t) = \sum_{k=1}^K \mathbf{h}_{1,k} s_{a,k}(t - \tau_k) e^{i2\pi(f_k - f_0)t}.$$

More compactly, this gives

$$\forall t \in \mathbb{R} \quad \mathbf{y}_a(t) = \mathbf{H} \mathbf{s}_a(t)$$

where  $\mathbf{H} = (\mathbf{h}_{1,1}, \dots, \mathbf{h}_{1,K})$  is a  $M \times K$  mixing matrix, and

$$(2.10) \quad \mathbf{s}_a(t) = \left( s_{a,1}(t - \tau_1) e^{i2\pi(f_1 - f_0)t}, \dots, s_{a,K}(t - \tau_K) e^{i2\pi(f_K - f_0)t} \right)^T.$$

If  $T_e$  is the sampling period of the receiver, it is supposed that all the components of the data are low-pass filtered in the sampling band (the matched-filter cannot be considered since the shaping filters are not supposed to be known to the receiver). Finally, the (noiseless model) of the data is

$$(2.11) \quad \text{General instantaneous model: } \forall n \in \mathbb{Z} \quad \mathbf{y}(n) = \mathbf{H} \mathbf{s}(n)$$

where

$$(2.12) \quad \mathbf{s}(n) = \left( s_{a,1}(nT_e - \tau_1) e^{i2\pi(f_1 - f_0)nT_e}, \dots, s_{a,K}(nT_e - \tau_K) e^{i2\pi(f_K - f_0)nT_e} \right)^T.$$

Generally speaking, any of the components of the source vector  $\mathbf{s}(n)$  is cyclo-stationary (see Section 2.2) hence the model given by (2.11) is a cyclo-stationary one. For simplification, let us suppose in the following that that  $f_k = f_0$  for all the indices  $k$  (this point is discussed in section 7). As the original theory of BSS assumed stationary data, we inspect under which conditions the above model can be stationary. the methods A necessary and sufficient condition is that all the components of  $\mathbf{s}(n)$  be stationary. As discussed previously, this can happen when all the symbol periods are equal to, say,  $T$  and if the sampling period  $T_e = T$ . Under these conditions, we even have:

$$s_{a,k}(t - \tau_k) = [c_{\tau_k,k}(z)] d_k(n)$$

where we have set for any delay  $\tau$

$$(2.13) \quad c_{\tau,k}(z) = \sum_{\ell} c_{a,k}(\ell T_k - \tau) z^{-\ell}.$$

The stationary model can be written as:

$$(2.14) \quad \text{Stationary instantaneous model: } \forall n \in \mathbb{Z} \quad \mathbf{y}(n) = \mathbf{H} \mathbf{s}(n)$$

where we have set

$$(2.15) \quad \mathbf{s}(n) = \begin{pmatrix} [c_{\tau_1,1}(z)] d_1(n) \\ \vdots \\ [c_{\tau_K,K}(z)] d_K(n) \end{pmatrix}.$$

(note: the notation might be confusing since  $\mathbf{s}(n)$  was already defined in (2.12): in the following sections, the context is always specified which prevents the confusion) In the literature, it is sometimes required that the sources be i.i.d. In the context of this paper, this i.i.d. condition is fulfilled when the filters  $c_{\tau_k,k}(z)$  all have the form of a constant times a delay: in short, this happens when 1) all the transmitted symbols are synchronized 2) the receiver runs a matched filter (square-root Nyquist) and 3) the symbol synchronization is performed at the receiver. In this case we have:

$$(2.16) \quad \text{i.i.d. instantaneous model: } \forall n \in \mathbb{Z} \quad \mathbf{y}(n) = \mathbf{H} \begin{pmatrix} d_1(n) \\ \vdots \\ d_K(n) \end{pmatrix}.$$

The reader may find this set of condition very restrictive in real scenarios. It is indeed; however, the developements of BSS are based on the stationary assumption. Moreover, many interesting methods exploit the i.i.d. condition.

- (2) **Convulsive mixture.** This is the general case when multi-paths affect the propagation. We provide the discrete-time version of Equation (2.9). Let us begin by the general case. In this respect, we assume that the sampling period  $T_e$  verifies the Shannon sampling condition, i.e.

$$\frac{1}{T_e} > \max_k \left( \frac{1 + \gamma_k}{T_k} + |f_k - f_0| \right).$$

This is a non-restrictive condition whatever the scenario: a crude prior spectral analysis of the data is simply needed. Provided this condition, the discrete-time signal  $s_{a,k}(nT_e - \tau_{k,m,\ell})$ , for any indices  $k, m, \ell$ , is a filtered version of  $(s_{a,k}(nT_e))_{n \in \mathbb{Z}}$ . It is hence easy to deduce that the sampled data  $\mathbf{y}(n)$  follows the equation:

$$(2.17) \quad \text{General convolutive model: } \forall n \in \mathbb{Z} \quad \mathbf{y}(n) = [\mathbf{H}(z)] \mathbf{s}(n)$$

where  $\mathbf{s}(n) = (s_{a,1}(nT_e)e^{i2\pi(f_1-f_0)nT_e}, \dots, s_{a,K}(nT_e)e^{i2\pi(f_K-f_0)nT_e})$  is a vector of mutually independent sources and  $\mathbf{H}(z)$  is certain the  $M \times K$  transfer function whose  $k$ -th column is the digital channel between the  $k$ -th source and the receiver: it depends on the parameters  $(\lambda_{k,m,\ell}, \tau_{k,m,\ell})_{\ell=1,\dots,L_k, m=1,\dots,M}$ . The above general model is, in general, cyclo-stationary. For simplification, we assume in the following that  $f_k = f_0$  for all indices  $k$ . Similarly to the case of instantaneous mixtures, it is instructive to find conditions under which the data are stationary. This occurs when the symbol periods  $T_k$  all coincide with a certain  $T$ , and when the sampling period  $T_e$  equals  $T$ . Under all these conditions,  $\mathbf{y}_k(n)$ , the contribution of the  $k$ -th source to

the mixture, can be written as

$$\mathbf{y}_k(n) = \begin{pmatrix} \sum_{\ell} \lambda_{k,1,\ell} [c_{\tau_{k,1,\ell}, k}(z)] d_k(n) \\ \vdots \\ \sum_{\ell} \lambda_{k,M,\ell} [c_{\tau_{k,M,\ell}, k}(z)] d_k(n) \end{pmatrix} = [\mathbf{E}_k(z)] d_k(n),$$

hence, setting  $\mathbf{E}(z) = (\mathbf{E}_1(z), \dots, \mathbf{E}_K(z))$ , it yields

$$(2.18) \quad \textbf{Stationary convolutive model: } \forall n \in \mathbb{Z} \quad \mathbf{y}(n) = [\mathbf{E}(z)] \mathbf{d}(n)$$

where  $\mathbf{d}(n) = (d_1(n), \dots, d_K(n))^T$  and  $\mathbf{E}(z)$  is a certain  $M \times K$  unknown filter matrix. This shows that  $\mathbf{E}(z)$  depends on the shaping filters, the steering vectors associated with the paths and their corresponding delays.

**2.3.2. Assumptions on the channels .** In this paper, we consider over-determined mixtures, that is: mixtures such that the number of sensors exceeds the number of sources ( $M \geq K$ ). This condition is necessary in order to retrieve the vector  $\mathbf{s}(n)$  - see model (2.11) (respectively (2.17)) - from the data  $\mathbf{y}(n)$  by means of a  $K \times M$  constant matrix (respectively a  $K \times M$  filter). This has to be specified.

For instantaneous mixtures, the following condition holds:

$$(2.19) \quad \textbf{Assumption (instantaneous mixtures): } \text{rank}(\mathbf{H}) = K.$$

Under this assumption, there exist  $K \times M$  matrices  $\mathbf{G}$  such that  $\mathbf{G}\mathbf{y}(n) = \mathbf{s}(n)$ .

For convolutive mixtures, it is conventional to assume that the components of  $\mathbf{H}(z)$  are polynomials in  $z^{-1}$  (this is an approximation that is justified since the shaping filters  $c_{a,k}(t)$  are rapidly vanishing when  $|t| \rightarrow \infty$ ). We further assume that

$$(2.20) \quad \textbf{Assumption (convolutive cyclostationary mixtures): } \forall z \neq 0 \quad \text{rank}(\mathbf{H}(z))=K.$$

Under this condition, there exist polynomial matrices  $\mathbf{G}(z)$  such that  $[\mathbf{G}(z)] \mathbf{y}(n) = \mathbf{s}(n)$ : see for instance [24, 29]. The same kind of assumption holds in the stationary case - see the model (2.18): namely,

$$(2.21) \quad \textbf{Assumption (convolutive stationary mixtures): } \forall z \neq 0 \quad \text{rank}(\mathbf{E}(z))=K.$$

At this level, we would like to point out a curiosity. In this respect, we assume further that the excess bandwidth factors of one source - say the first one - equals zero. As the choice  $T_e = T_1$  satisfies the Shannon sampling condition, we may write  $s_{a,1}(nT_1 - \tau) = [\phi_{\tau}(z)] s_{a,1}(nT_1)$  where

$$\phi_{\tau}(z) = \sum_k \frac{T_1}{\pi(kT_1 - \tau)} \sin\left(\frac{\pi(kT - \tau)}{T_1}\right) z^{-k}.$$

As  $s_{a,1}(nT_1) = [c_{0,1}(z)] d_1(n)$ , the first column  $\mathbf{E}_1(z)$  of  $\mathbf{E}(z)$  can be factored as  $c_{0,1}(z) \tilde{\mathbf{E}}_1(z)$ . In particular, after the standard FIR approximations, it yields that the condition given in (2.21) is not fulfilled.

### 3. INSTANTANEOUS MIXTURES

The model of the data  $\mathbf{y}(n)$  is given by (2.11). The mixing matrix  $\mathbf{H}$  is unknown. BSS can be achieved either by estimating  $\mathbf{H}$  - this is the point of Section 3.3 - or by computing directly estimates of the sources (up to indeterminacies).

**3.1. Indeterminacies.** It is always possible to consider that the sources have equal and normalized power. Indeed, as  $s_1(n) = \rho_1 \frac{s_1(n)}{\rho_1}$  where  $\rho_1 = \sqrt{\langle E [ |s_1(n)|^2 ] \rangle}$  is the the square-root of the power of the first source, we suggest to scale the first column of  $\mathbf{H}$  by  $\rho_1$ . Repeating this process for all the sources, we have constructed a new matrix  $\tilde{\mathbf{H}}$ . Eventually denoting  $\tilde{\mathbf{s}}(n) = \left( \frac{s_1(n)}{\rho_1}, \frac{s_2(n)}{\rho_2}, \dots, \frac{s_K(n)}{\rho_K} \right)^T$ , we obviously show that the data alternatively writes

$$\forall n \in \mathbb{Z} \quad \mathbf{y}(n) = \tilde{\mathbf{H}} \tilde{\mathbf{s}}(n).$$

Though apparently innocent, this remark gives precious a priori indications. First of all, it says that the model (3.1) is not uniquely defined. As a consequence, it is always possible to consider, without restricting the model, that the sources have equal power equal to one - this precisely corresponds to the above defined  $\tilde{\mathbf{s}}(n)$ ; specifically, we will assume in the following that

$$(3.1) \quad \langle E [\mathbf{s}(n)\mathbf{s}(n)^H] \rangle = \mathbf{I}_K$$

This shows it is beyond a reasonable expectation to retrieve the sources with no scaling ambiguities. Similarly, if  $\mathbf{P}$  is a permutation matrix,  $\mathbf{y}(n) = (\mathbf{H}\mathbf{P}^{-1}) (\mathbf{P}\mathbf{s}(n))$ , underlining the non-unicity of the model.

With no further assumptions on the sources, the ultimate result that can be achieved is : retrieve the sources up to unknown *complex scaling factors* (scaling and phase ambiguities) and a *permutation*.

**3.2. Algebraic methods (i.i.d. scenario).** The model of the data is given by (2.16). We may collect the  $N$  available data in a  $M \times N$  matrix  $\mathcal{Y}$ , we have:  $\mathcal{Y} = \mathbf{H}\mathcal{D}$  where  $\mathcal{D} = (\mathbf{d}(0), \dots, \mathbf{d}(N-1))$ . As any entry of  $\mathcal{D}$  corresponds to a symbol, associated specificities (e.g. finite alphabet constellations or modulus one symbols) are a priori relations the receiver can make use of. As far as the identifiability is concerned, it is proven in [61] (Lemma 1) that the above factorization is essentially unique for modulus one symbols, at least if the number of snapshots  $N$  verifies  $N \geq 2K$  (which is the case in practical contexts). By essentially unique, we mean that the rows of  $\mathcal{D}$  may be permuted and/or multiplied by modulus one constants.

Talwar et al. [53, 54] propose iterative algorithms that assume known the alphabets of the symbols. Call  $\hat{\mathbf{H}}^{(\ell)}$  an estimate of  $\mathbf{H}$  at the iteration no  $\ell$ . The Iterative Least Square with Projection (ILSP) is:

- (1) Take any full rank  $\hat{\mathbf{H}}^{(0)}$  for iteration  $\ell = 0$
- (2)  $\ell \leftarrow \ell + 1$ 
  - $\hat{\mathcal{D}}^{(\ell)} = \left( \hat{\mathbf{H}}^{(\ell-1)} \right)^\sharp \mathcal{Y}$  where  $(\cdot)^\sharp$  denotes the pseudo-inverse
  - $\hat{\mathcal{D}}^{(\ell)} \leftarrow$  projection of each component of  $\hat{\mathcal{D}}^{(\ell)}$  on the corresponding alphabet
  - $\hat{\mathbf{H}}^{(\ell)} = \mathcal{Y} \left( \hat{\mathcal{D}}^{(\ell)} \right)^\sharp$ .

Similar projection-based algorithms that rather take into account the constant modulus property of the entries of  $\mathcal{D}$  have been considered [23, 3]: similarly to the IMSP algorithm, no results on the convergence can be given (how many samples are required? are there local minima the algorithm could be trapped in?). Van der Veen et al. [61] proposes a non-iterative algorithm, called the Algebraic CMA (ACMA):

the ACMA provides exactly «the» solution (up to the above mentioned ambiguities) of the factorization of  $\mathcal{Y}$  - at least if the number of data  $N$  exceeds  $K^2$ . It is based on a joint diagonalization of a pencil of  $K$  matrices.

Certain BSS methods for convolutive mixtures need, as a final step, to run such algorithms (see e.g. section 5.1).

### 3.3. Second-order based identification (general cyclo-stationary case) .

In this section, we address the «indirect» BSS; by this terminology, we mean that the BSS is achieved in two steps. The first step consists in estimating the unknown mixing matrix  $\mathbf{H}$  by, say,  $\hat{\mathbf{H}}$ . In a second step, the proper separation is carried out. If  $\hat{\mathbf{H}}$  is an accurate estimate, then  $\hat{\mathbf{s}}(n) = \hat{\mathbf{H}}^\# \mathbf{y}(n)$  is the natural estimate of the source vector. In general, however, noise is present (estimation noise and additive noise in the observed signals) and other strategies have to be considered: this aspect is not addressed in this paper.

The first point to be addressed in this section is the *pre-whitening* of the data. We suppose that  $M = K$  (notice: in the non-square case, a principal component analysis is processed). In this respect, we consider the auto-correlation matrix of the data  $\mathbf{R}_y^{(0)}(0) = \langle E [\mathbf{y}(n)\mathbf{y}(n)^H] \rangle$  can be written (we recall that the sources are assumed to have equal normalized powers as discussed previously):

$$\mathbf{R}_y^{(0)}(0) = \mathbf{H}\mathbf{H}^H.$$

Since  $\mathbf{H}$  is full rank, the above matrix is positive definite and we form the new data

$$\mathbf{x}(n) = \left( \mathbf{R}_y^{(0)}(0) \right)^{-1/2} \mathbf{y}(n).$$

We have:

$$\mathbf{x}(n) = \mathbf{U}\mathbf{s}(n)$$

where  $\mathbf{U} = (\mathbf{H}\mathbf{H}^H)^{-1/2} \mathbf{H}$  is a unitary matrix.

The second point concerns the estimation of the unitary matrix  $\mathbf{U}$ . The data  $\mathbf{x}(n)$  is cyclo-stationary. As the cyclic-frequencies are not always directly accessible, the identification of the unknown mixing matrix  $\mathbf{U}$  is done by solely considering the statistics

$$R_x^{(0)}(\ell) = \langle \mathbf{x}(n+\ell)\mathbf{x}(n)^H \rangle$$

which can be expressed as

$$R_x^{(0)}(\ell) = \mathbf{U}R_s^{(0)}(\ell)\mathbf{U}^H.$$

This says that the normal matrix  $R_x^{(0)}(\ell)$ , for any index  $\ell$ , is diagonalized in the orthonormal basis formed by the columns of  $\mathbf{U}$ . For  $\ell = 0$ , this gives  $R_x^{(0)}(0) = \mathbf{U}\mathbf{U}^H = \mathbf{I}$  and this is clearly not sufficient to identify  $\mathbf{U}$ ! On the contrary, consider that the spectra of the sources  $S_{s_k}^{(0)}(\nu) = \sum_{\ell} R_{s_k}^{(0)}(\ell)e^{-i2\pi\ell\nu}$  are all different at least for a frequency  $\nu$ . For any unitary matrix  $\mathbf{V}$ , the matrix  $\mathbf{V}^H R_x^{(0)}(\ell)\mathbf{V}$  is diagonal for every indices  $\ell$  if and only if the columns of  $\mathbf{V}$  equal these of  $\mathbf{U}$  up to a modulus one factor and a permutation. This remark was done in [4] and an algorithm (SOBI) was deduced based on a joint diagonalization technique [9].

The reader has noticed the suboptimality of the above method when the mixture is cyclo-stationary. The exploited statistics are only the  $R_x^{(0)}(\ell)$  for certain indices

$\ell$ . In [2], it is suggested to take advantage of the cyclic-statistics of the mixture. In this respect, notice that for any  $\alpha$  cyclic frequency of the mixture, we have

$$R_{\mathbf{x}}^{(\alpha)}(\ell) = \mathbf{U}R_{\mathbf{s}}^{(\alpha)}(\ell)\mathbf{U}^H$$

hence these «new» statistics could be added in the pencil of matrices to be jointly diagonalized. This theoretical appeal is attenuated by the fact that the non null-components of  $R_{\mathbf{s}}^{(\alpha)}(\ell)$  are numerically inconsistent.

At this level, we should emphasize that the statistics  $R_{\mathbf{x}}^{(\alpha)}(\ell)$  for any  $\alpha$  (zero or not) are not accessible to the receiver and should be replaced by the empirical estimate denoted by  $\hat{R}_{\mathbf{x}}^{(\alpha)}(\ell)$  and defined for  $\ell \geq 0$  as

$$(3.2) \quad \hat{R}_{\mathbf{x}}^{(\alpha)}(\ell) = \frac{1}{N} \sum_{n=0}^{N-\ell-1} \mathbf{x}(n+\ell)\mathbf{x}(n)^H e^{-i2\pi\alpha n}$$

where  $N$  is the number of snapshots. This estimate is a consistent estimate of the matrix  $R_{\mathbf{x}}^{(\alpha)}(\ell)$ . In an ideal scenario where the model  $\mathbf{x}(n) = \mathbf{U}\mathbf{s}(n)$  holds true, it is remarkable that  $\hat{R}_{\mathbf{x}}^{(\alpha)}(\ell) = \mathbf{U}\hat{R}_{\mathbf{s}}^{(\alpha)}(\ell)\mathbf{U}^H$  and the joint diagonalization of the estimated statistics should provide the exact mixing matrix: the algorithm is called *deterministic*. In a realistic context, however, the data are perturbed by an additive noise term: in this case, the above factorization does not hold true anymore and the joint diagonalization is an *approximate* joint diagonalization.

In practice, despite its attractivity, SOBI is seldom used to achieve BSS of digital communication data. Indeed, the condition that there are no two sources whose spectra are identical (up to a multiplicative constant) does not make sense most of the time. Indeed, the transmitted symbols are generally white sequences whose shaping functions are close from to one another. As the spectra are numerically similar, the joint diagonalization approach is bound to suffer from numerical problems.

**3.4. Iterative BSS (stationary case).** As was specified, the stationary scenario assumes that for all indices  $k$ :  $T_k = T$ , i.e. all the baud-rates are equal, and  $T_e = T$ . Under these very specific circumstances, the model (2.14) involves a source vector  $\mathbf{s}(n)$  whose components are **stationary** and **mutually independent**. We insist on the fact that the components of the source vector are not the i.i.d. symbol sequences but linear processes generated by these symbol sequences as indicated by Equation (2.15). BSS aims at estimating the sources, not the symbol sequences. Hence, BSS may be seen as a preliminary step before the estimation of the symbols.

Contrary to other methods, no pre-processing of the data is necessary (PCA, pre-whitening).

In this section, we firstly design methods able to recover one of the sources (or a scaled version). In a second step, we present the so-called deflation that allows one to run the extraction of another source from a deflated mixture where the contribution of the first estimated source has been removed. The convergence is established: after  $K$  such steps, the  $K$  sources are expected to be estimated. Convergence properties are discussed.

**3.4.1. Estimation of one source: theoretical considerations.** Thanks to the mixing matrix  $\mathbf{H}$  having full-rank - see condition (2.19) - we know that, for any source

index  $k$ , there exist column vectors  $\mathbf{g}_k$  such that

$$\mathbf{g}_k^T \mathbf{y}(n) = s_k(n).$$

Denoting  $r_k(n) = \mathbf{g}_k^T \mathbf{y}(n)$ , we may call this new signal as the reconstructed source since it involves only one source. This new signal is obtained after a so-called spatial filtering of the data. Of course, it is not possible to compute  $\mathbf{g}_k$  since  $\mathbf{H}$  is not accessible. A possible approach hence consists in adapting a spatial filter  $\mathbf{g}$  that makes

$$(3.3) \quad r(n) = \mathbf{g}^T \mathbf{y}(n)$$

resemble one of the sources. This will be done by considering particular statistics of the signal  $r(n)$ . We may write this signal under this form

$$(3.4) \quad r(n) = \sum_{k=1}^K f_k s_k(n)$$

where the taps  $f_k$  are the components of the vector

$$(3.5) \quad \mathbf{f} = \mathbf{g}^T \mathbf{H}.$$

The term  $f_k s_k(n)$  in  $r(n)$  represents the contribution of the  $k$ -th source to the reconstructed signal  $r(n)$ . As may be easily understood, we aim at finding a «good»  $\mathbf{g}$ , i.e. such that  $\mathbf{f}$  is a vector having a single non-null component.

**Definition 2.** A vector  $\mathbf{f}$  is said to be separating if all its components are null except one.

Evidently, the signal  $r(n)$  involves a single source if and only if the composite vector  $\mathbf{f}$  is separating.

We may inspect higher-order statistics and particularly the fourth-order ones. It has been proposed to consider the fourth-order cumulant (see Section 3.6 for theoretical justifications):

$$\kappa(r) = \text{cum}(r(n), r(n)^*, r(n), r(n)^*).$$

In this respect, we may introduce the following function, called *normalized (fourth-order) cumulant*:

$$(3.6) \quad \Upsilon(r) = \frac{\kappa(r)}{\left(E[|r(n)|^2]\right)^2}.$$

Thanks to the definition of the cumulants, we have:  $\kappa(r) = E[|r(n)|^4] - 2\left(E[|r(n)|^2]\right)^2 - |E[r(n)^2]|^2$ . Now, the circularity assumption of the symbol sequences (2.2) implies the circularity of the sources, hence

$$E[r(n)^2] = 0.$$

We re-express  $\Upsilon(r)$  as a function of the moments of  $r(n)$ :

$$(3.7) \quad \Upsilon(r) = \frac{E[|r(n)|^4]}{\left(E[|r(n)|^2]\right)^2} - 2.$$

We have the result:

**Proposition 3.** *As, by assumption, the Kurtosis of the sources,  $\kappa(d_k)$ , are strictly negative, the function  $\Upsilon(r)$  achieves its minimum at a separating vector. Moreover, the separating vector in question has its single non-null element located at an index  $k_0$  such that  $\Upsilon(s_{k_0}) = \min_{k=1, \dots, K} \Upsilon(s_k)$ .*

Proof: on the one hand, the mixture  $r(n)$  is a linear mixture of independent random variables. The multi-linearity of the cumulants [44] gives:

$$\kappa(r) = \sum_{k=1}^K |f_k|^4 \kappa(s_k).$$

After noticing that  $E[|r(n)|^2] = \sum_k |f_k|^2$  we arrive at the expansion:

$$(3.8) \quad \Upsilon(r) = \frac{\sum_{k=1}^K \kappa(s_k) |f_k|^4}{\left(\sum_{k=1}^K |f_k|^2\right)^2}$$

On the other hand,  $s_k(n)$  is a linear process generated by the i.i.d. symbol sequence  $d_k(n)$ : see Equation (2.7). As was supposed (or noticed) in the Introduction,  $\kappa(d_k) < 0$  hence

$$\kappa(s_k) < 0$$

for any index  $k$ . The sources are sometimes referred to as *platykurtic sources*. Denote by  $\kappa_{\min} = \min(\kappa(s_1), \dots, \kappa(s_K))$ . Thanks to the above result,  $\kappa_{\min} < 0$ . Hence  $\kappa(r) \geq -|\kappa_{\min}| \sum_k |f_k|^4$ . Besides, we recall that  $\left(\sum_k |f_k|^2\right)^2 \geq \sum_k |f_k|^4$  with equality if and only if the coefficients  $|f_k|$  are all null except one, i.e. if and only if the vector  $\mathbf{f}$  is separating.

We insist on the fact that the assumption that one of the  $\kappa(s_k)$  is strictly negative is fundamental. Imagine on the contrary that, for all the indices  $k$ ,  $\kappa(d_k) > 0$ . Then  $\sum_{k=1}^K \kappa(s_k) |f_k|^4 \geq \min_k \kappa(s_k) \left(\sum_{k=1}^K |f_k|^4\right)$ . As

$$\left(\frac{1}{K} \sum_{k=1}^K |f_k|^4\right) \geq \left(\frac{1}{K} \sum_{k=1}^K |f_k|^2\right)^2$$

with equality iff all the  $|f_k|$  are equal, this implies that the argument minima of  $\Upsilon(r)$  are not separating (on the contrary, the coefficients  $f_k$  equally weigh the sources).  $\square$

As a remark, it is instructive, though superfluous in this paper since the digital communication symbols have negative Kurtosis, to address the optimization of  $\Upsilon(r)$  for general distributions of the  $\kappa(s_k)$ : the reader may find the details in [17].

One may inspect the minimum of  $\Upsilon(r)$  over all the possible constellations. The Jensen inequality (see [6] page 80) gives:  $E\left[\left(\varphi|r(n)|^2\right)\right] \geq \varphi\left(E\left[|r(n)|^2\right]\right)$  for a convex mapping  $\varphi$ ; the equality is achieved when  $|r(n)| = 1$ . Taking  $\varphi(x) = x^2$ , we obtain

$$\Upsilon(r) \geq -1$$

and the equality is achieved when  $r(n)$  has unit modulus. Of course, this can only happen if one of the sources has a modulus equal to one or, there exists an index  $k$  such that  $\Upsilon(s_k) = -1$ . This does not happen in general, but this remark shows

that the minimization of  $\Upsilon(r)$  tends to make  $r(n)$  resemble as much as possible a constant modulus sequence. We inspect this point a bit further. A way to measure the distance of  $r(n)$  to the modulus one is simply to consider

$$(3.9) \quad \Upsilon_{CM}(r) = E \left[ \left( |r(n)|^2 - 1 \right)^2 \right].$$

This function was originally considered for deconvolution problems [48, 22] and then for source separation problems ([36, 56, 55] for instance).

We may bridge the gap between  $\Upsilon(r)$  and  $\Upsilon_{CM}(r)$  :

**Proposition 4.** *Define  $\Upsilon_{CM,\text{inf}} = \inf_r \Upsilon_{CM}(r)$  and  $\Upsilon_{\text{inf}} = \inf_r \Upsilon(r)$ . The minimization of  $\Upsilon_{CM}(r)$  is linked to the minimization of  $\Upsilon(r)$  in the sense that:*

$$\Upsilon_{CM,\text{inf}} = 1 - \frac{1}{2 + \Upsilon_{\text{inf}}}$$

*if  $\mathbf{f}$  achieves to minimize  $\Upsilon$  then  $\sqrt{\frac{1}{2 + \Upsilon_{\text{inf}} \Upsilon}} \frac{1}{\|\mathbf{f}\|} \mathbf{f}$  is a minimizer of  $\Upsilon_{CM}$ . Conversely, if  $\mathbf{f}$  achieves to minimize  $\Upsilon_{CM}$ , then  $\rho \mathbf{f}$  minimizes  $\Upsilon$  for any  $\rho \neq 0$ .*

Proof: for any  $\mathbf{f}$ , we have:  $\Upsilon_{CM}(r) = E \left[ |r(n)|^4 \right] - 2E \left[ |r(n)|^2 \right] + 1$ . Thanks to the expression of  $\Upsilon(r)$  in (3.7), it is always true that:  $E \left[ |r(n)|^4 \right] \geq (2 + \Upsilon_{\text{inf}}) \left( E \left[ |r(n)|^2 \right] \right)^2$ . We hence have:

$$\Upsilon_{CM}(r) \geq (2 + \Upsilon_{\text{inf}}) \left( E \left[ |r(n)|^2 \right] \right)^2 - 2E \left[ |r(n)|^2 \right] + 1.$$

We set  $\rho = E \left[ |r(n)|^2 \right]$ . The second-order polynomial  $\rho \mapsto (2 + \Upsilon_{\text{inf}})\rho^2 - 2\rho + 1$  has minimal value  $1 - \frac{1}{2 + \Upsilon_{\text{inf}}}$  for  $\rho = \frac{1}{2 + \Upsilon_{\text{inf}}}$ . We deduce the inequality:  $\Upsilon_{CM}(r) \geq 1 - \frac{1}{2 + \Upsilon_{\text{inf}}}$ . If  $\mathbf{f}$  reaches the infimum of  $\Upsilon(r)$  then, evidently, the choice  $\sqrt{\frac{1}{2 + \Upsilon_{\text{inf}}}} \frac{1}{\|\mathbf{f}\|} \mathbf{f}$  makes  $\Upsilon_{CM}(r) = 1 - \frac{1}{2 + \Upsilon_{\text{inf}}}$ . Hence  $1 - \frac{1}{2 + \Upsilon_{\text{inf}}}$  is the minimum of  $\Upsilon_{CM}(r)$ . Conversely, for any  $r(n)$ , by definition, we have:  $E \left[ |r(n)|^4 \right] - 2E \left[ |r(n)|^2 \right] + 1 \geq \Upsilon_{CM,\text{inf}}$ . In this inequality, substitute  $\sqrt{\rho} r(n)$  for any positive  $\rho$ . We have:

$$\rho^2 E \left[ |r(n)|^4 \right] - 2\rho E \left[ |r(n)|^2 \right] + 1 \geq \Upsilon_{CM,\text{inf}}.$$

This is in particular true for  $\rho = \frac{E \left[ |r(n)|^2 \right]}{E \left[ |r(n)|^4 \right]}$  hence showing that  $-\frac{(E \left[ |r(n)|^2 \right])^2}{E \left[ |r(n)|^4 \right]} + 1 \geq \Upsilon_{CM,\text{inf}}$  or

$$\Upsilon(r) \geq 2 + \frac{1}{1 - \Upsilon_{CM,\text{inf}}}.$$

The case of equality in the latter equation occurs when  $\mathbf{f}$  is any non-null scaled version of a minimizer of  $\Upsilon_{CM}$   $\square$

As is explained in the next section, the search of a global minimum of  $\Upsilon(r)$  or  $\Upsilon_{CM}(r)$  is done according to a gradient method. It is well known that such an algorithm may be stuck in a local minimum of the function to be minimized. We have the result (see [17], Lemma 2.1) :

**Lemma 5.** *Fix  $K$  real constants  $\beta_k < 0$ . The local minima of the function  $\mathbf{f} \mapsto \sum_{k=1}^K \beta_k |f_k|^4$  over the unit sphere  $\sum_{k=1}^K |f_k|^2 = 1$  are the separating vectors (of unit norm).*

Thanks to the expansion (3.8), and the fact that for all the sources  $\beta_k = \kappa(s_k) < 0$ , the local minima of  $\Upsilon(r)$  over the unit sphere are the separating vectors (of unit norm). After simple topological considerations, it can even be deduced that:

**Proposition 6.** *Any local minimum of  $\Upsilon$  is separating.*

As far as the function  $\Upsilon_{CM}$  is concerned, we have:

**Proposition 7.** *Any local minimum of  $\Upsilon_{CM}$  is separating.*

Proof: we consider the arguments given in [46]. The idea consists in writing  $\mathbf{f}$  in its polar form  $\mathbf{f} = \rho \tilde{\mathbf{f}}$  where  $\rho = \|\mathbf{f}\|$  and  $\|\tilde{\mathbf{f}}\| = 1$ . After setting  $\tilde{r}(n) = \frac{1}{\|\tilde{\mathbf{f}}\|} r(n)$  the normalized version of  $r(n)$ , we have:  $\Upsilon_{CM}(r) = \rho^4 E \left[ |\tilde{r}(n)|^4 \right] - 2\rho^2 E \left[ |\tilde{r}(n)|^2 \right] + 1$ . Write this function  $\psi(\rho, \tilde{\mathbf{f}})$ . Necessarily, for a stationary point of  $\Upsilon_{CM}$  the derivative of  $\psi$  w.r.t.  $\rho$  is zero. This gives:  $\rho = 0$  or  $\rho(\tilde{\mathbf{f}}) = \frac{E[|\tilde{r}(n)|^2]}{E[|\tilde{r}(n)|^4]}$ . The case  $\rho = 0$  can be shown to correspond to a local maximum [55]. This says that a local minimum of  $\Upsilon_{CM}$  is a local minimum of  $\psi(\rho(\tilde{\mathbf{f}}), \tilde{\mathbf{f}})$ . Now, this latter function is:

$$\psi(\rho(\tilde{\mathbf{f}}), \tilde{\mathbf{f}}) = 1 - \frac{1}{\Upsilon(\tilde{r})}.$$

We deduce that such a local minimum is also a local minimum of  $\Upsilon(r)$  on the unit sphere. Thanks to Lemma 5 we deduce that the local minimum in question is separating.  $\square$

#### 3.4.2. Estimation of one source: practical aspects.

Basic algorithms. Two problems arise when one focuses on the implementation of the results presented so forth: the first one concerns the estimation of the cost functions  $\Upsilon(r)$  or  $\Upsilon_{CM}(r)$ , the second one is to choose a method able to find the argument minima of these estimated functions.

The two functions we have considered involve second and fourth-order moments of the signal  $r(n)$ . As the number of available data is finite - say, we observe  $\mathbf{y}(n)$  for  $n = 0, \dots, N-1$  - it is not possible to compute any of the moments of  $r(n)$ . However, a version of the law of large numbers allows one to consider estimates of the moments:

**Lemma 8.** *For  $p = 2, 4$ , we have, with probability one:*

$$\frac{1}{N} \sum_{n=0}^{N-1} |r(n)|^p \longrightarrow E[|r(n)|^p].$$

We are in position to estimate both functions  $\Upsilon(r)$  and  $\Upsilon_{CM}(r)$  respectively by

$$(3.10) \quad \hat{\Upsilon}(r) = \frac{\frac{1}{N} \sum_{n=0}^{N-1} |r(n)|^4}{\left( \frac{1}{N} \sum_{n=0}^{N-1} |r(n)|^2 \right)^2} - 2$$

$$(3.11) \quad \hat{\Upsilon}_{CM}(r) = \frac{1}{N} \sum_{n=0}^{N-1} \left( |r(n)|^2 - 1 \right)^2.$$

Indeed, we have the result:

**Proposition 9.**  $\hat{\Upsilon}_{CM}(r) \longrightarrow \Upsilon_{CM}(r)$  and  $\hat{\Upsilon}(r) \longrightarrow \Upsilon(r)$  with probability one.

The functions  $\hat{\Upsilon}_{CM}(r)$  and  $\hat{\Upsilon}(r)$  to be minimized are non-convex, and the associated machinery cannot be considered. The functions, however, are regular w.r.t. the parameter  $\mathbf{g}$ . Hence, we choose to seek the argument minima by means of a gradient method. For instance, consider the minimization of  $\Upsilon_{CM}$ . The notation for the gradient of  $\hat{\Upsilon}_{CM}$  calculated at the point  $\mathbf{g}$  being  $\nabla \hat{\Upsilon}_{CM}(r)$ , the gradient algorithm, for a fixed  $\mu > 0$ , can be written as :

- (1) choose an initial vector  $\mathbf{g}^{(0)}$  and compute  $r^{(0)}(n) = \mathbf{g}^{(0)T} \mathbf{y}(n)$  for all the available data;
- (2) at the  $m$ -th step: compute  $\mathbf{g}^{(m)} = \mathbf{g}^{(m-1)} + \mu \nabla \hat{\Upsilon}_{CM}(r^{(m-1)})$  and the associated updated signal  $r^{(m)}(n) = \mathbf{g}^{(m)T} \mathbf{y}(n)$  ;
- (3) redo the above step until the convergence is reached.

The same algorithm could be written for the minimization of  $\hat{\Upsilon}$ . However, the fact this latter function is homogeneous may involve numerical problems (the vector  $\mathbf{g}$  is not bounded). This is why the projected gradient algorithm is preferred: it consists in normalizing at each iteration of the algorithm the updated signal  $r^{(m)}(n)$ , i.e. projecting the current parameter  $\mathbf{g}^{(m)}$  on the set

$$\left\{ \mathbf{g} \mid \mathbf{g}^T \left( \frac{1}{N} \sum_{n=0}^{N-1} E [\mathbf{y}(n)\mathbf{y}(n)^H] \right) \mathbf{g} = 1 \right\}.$$

Whatever the considered cost function, the parameter  $\mu$  controls the performance. The next section faces the problem of choosing  $\mu$ .

Refinement: choosing a locally optimal  $\mu$ . For simplicity, the minimization of  $\Upsilon_{CM}$  is addressed. The same idea may be considered for the minimization of  $\Upsilon$  by means of the projected gradient. In order to boost the speed of convergence, it has been proposed to change  $\mu$  at each step of the algorithm: the parameter  $\mu$  is chosen such that the value of the function evaluated at the point  $\mathbf{g}^{(m)}$  is minimum. It is easily seen that the function  $\mu \mapsto \hat{\Upsilon}_{CM}(r^{(m)})$  is a polynomial of degree four. The minimum is hence easily (numerically) computed.

Robustness of the algorithms to the presence of local minima. It is well-known that such a gradient algorithm may be trapped in a local minimum: this, in general, is a clear limitation to the use of such an algorithm. Of course, it is not possible to say much on the local minima of the estimated functions  $\hat{\Upsilon}(r)$  and  $\hat{\Upsilon}_{CM}(r)$ . However, Propositions 6 and 7 indicate that, asymptotically, if the algorithms are trapped in a local minimum, this does not impact the performance since this local minimum is precisely separating. This remark certainly explains why the algorithms show very good performance.

3.4.3. *The deflation step.* The algorithms depicted above provide a way to retrieve one of the sources. Of course, we aim at estimating all the sources. An idea hence consists in running again the previous algorithm. However, it is not possible to guarantee that the second extracted source is not the first extracted one. In the literature, three methods have been presented that overcome this major problem.

In the first one [19] it is proposed to penalize the cost function  $\Upsilon(r)$  or  $\Upsilon_{CM}(r)$  by adding to them a positive term that gives a measure of decorrelation between the current signal  $r(n)$  and the previously extracted source. It is simple to show that, indeed, the global minimum is achieved if and only if the  $r(n)$  is an other

source. However, this approach has been noticed to show poor performance. The reason is that the extended cost function, contrary to the original, has many local minima that do not correspond to separating solutions. The algorithm is known to be trapped in such local minima and, in this case, the provided solution is not an estimate of one of the remaining sources.

The second one is algebraic: the idea is to estimate the subspace associated with the first estimated source and to run the minimization of  $\Upsilon(r)$  or  $\Upsilon_{CM}(r)$  on the orthogonal complement of the subspace in question: see [17].

The third is the most popular for source separation [58, 50]: it consists in deflating the mixture by subtracting an estimation of the contribution of the extracted source and then to redo the minimization of  $\Upsilon(r)$  or  $\Upsilon_{CM}(r)$ . Ideally, the «new» mixture should not involve the source that has been extracted and the minimization hence allows one to estimate another source. We provide some details.

Thanks to the previous results, we may suppose that we have  $r^{(1)}(n) \approx \alpha s_1(n)$  where  $\alpha$  is an unknown scaling. We have arbitrarily considered that the extracted source was the one numbered «1»: this has of course no impact on the generality. The contribution of the first source in the mixture  $\mathbf{y}(n)$  has the form  $\mathbf{h}_1 s_1(n)$  where  $\mathbf{h}_1$  is the first column of the mixing matrix. We adopt a least square approach: the contribution of the first source is estimated as  $\hat{\mathbf{h}}^{(1)}$  where this vector is defined as the minimizer of

$$\mathbf{h} \mapsto \sum_{n=0}^{N-1} \left\| \mathbf{y}(n) - \mathbf{h} r^{(1)}(n) \right\|^2.$$

Then the «deflated mixture» to be considered is

$$\mathbf{y}^{(2)}(n) = \mathbf{y}(n) - \hat{\mathbf{h}}^{(1)} r^{(1)}(n).$$

Ideally, the deflated mixture should not involve the first source. Hence running the Constant Modulus algorithm on this mixture should provide an estimate  $r^{(2)}(n)$  of another source - say  $s_2(n)$ . The deflation is done again: this time  $\hat{\mathbf{h}}^{(2)} r^{(2)}(n)$  is an estimation of the contribution of the second source. The deflated mixture is

$$\mathbf{y}^{(3)}(n) = \mathbf{y}^{(2)}(n) - \hat{\mathbf{h}}^{(2)} r^{(2)}(n).$$

And so forth until all the source are estimated. Notice that, asymptotically (when  $N \rightarrow \infty$ ) the deflation procedure is convergent: in  $K$  steps the  $K$  sources are estimated.

**3.4.4. Improving the deflation.** Though its inherent advantages (simplicity, convergence of the algorithm of extraction), the above approach is supposed to suffer from the  $K$  deflation steps. Indeed, the deflation is expected to increase, step after step, the noise level, impinging dramatically the extraction of the «last» source. This aspect has already been addressed and partially got round: we shortly address the re-initialization procedure introduced in [59].

Consider the extraction of the «second» source and apply the deflation technique. The source extraction algorithm is run on the deflated mixture and is likely to provide a spatial filter  $\mathbf{g}^{(2)}$ . We have, up to a scaling factor:

$$\mathbf{g}^{(2)T} \left( \underbrace{\mathbf{y}(n) - \hat{\mathbf{h}}^{(1)} \mathbf{g}^{(1)T} \mathbf{y}(n)}_{\text{deflated mixture}} \right) \approx s_2(n)$$

which provides the approximation:

$$\mathbf{g}^T \mathbf{y}(n) \approx s_2(n)$$

where  $\mathbf{g}^T = \mathbf{g}^{(2)T} (\mathbf{I}_M - \hat{\mathbf{h}}^{(1)} \mathbf{g}^{(1)T})$ . We hence have computed a spatial filter  $\mathbf{g}$  that is close to a separating filter w.r.t. the initial mixture. The idea is hence the following: run the algorithm of minimization on the initial mixture, taking  $\mathbf{g}$  as an initial point. As  $\mathbf{g}$  is close to a filter that is a local minimum of the function to minimize (see Propositions 6,7), the computed spatial filter hence obtained after convergence is likely to separate  $s_2$  from the initial mixture. This procedure can be iterated: at each step, the separation is processed on the initial  $\mathbf{y}(n)$  and not on a deflated mixtures. Though simple, this procedure considerably enhances the performance.

3.4.5. *Extensions.* Many contributions in BSS consider such functions as

$$(3.12) \quad \varphi(\kappa(r))$$

over the unit sphere  $E[|r(n)|^2] = 1$  where  $\varphi: \mathbb{R} \rightarrow \mathbb{R}$  is any continuous function on  $\mathbb{R}$  such that  $\varphi(0) = 0$  and  $\varphi$  is strictly monotone over  $\mathbb{R}^+$  and  $\mathbb{R}^-$ . The most common choices are  $\varphi_1(x) = x^2$  and  $\varphi_2(x) = |x|$  and  $\varphi_3(x) = -x$ . It quite simple to prove the following result:

**Proposition 10.** *Provided that  $\varphi(\kappa(r_k)) > 0$  for at least one source, then the local maxima of (3.12) are separating.*

3.5. **Global BSS (stationary mixture).** In this section, we present global methods, i.e. methods that «invert» the system in one shot. Assuming that  $K = M$ , it is possible to linearly transform the data such as in Section 3.3. The «new» data can be written as

$$\mathbf{x}(n) = \mathbf{U} \mathbf{s}(n)$$

where  $\mathbf{U}$  is a unitary  $K \times K$  matrix. A global BSS method hence aims at determining a *unitary* matrix  $\mathbf{V}$  such that the components of

$$\mathbf{r}(n) = \mathbf{V} \mathbf{x}(n)$$

correspond to the sources up to modulus one scalings and a permutation. In this respect, we suggest to take profit of certain results of Section 3.4.

3.5.1. *First result.* Denote by  $r_k(n)$  the  $k$ -th component of  $\mathbf{r}(n)$ . Due to the pre-whitening, we have:  $E[|r_k|^2] = 1$ , hence  $\Upsilon(r_k) = \kappa(r_k)$ . This later can be seen as a function of the  $k$ -th row of the matrix  $\mathbf{V}$ . We have shown that  $\Upsilon(r_k)$  is minimum if  $r_k$  corresponds to one of the sources up to a modulus one scaling, i.e. if the  $k$ -th row of  $\mathbf{F} = \mathbf{V} \mathbf{U}$  is separating, its non-zero component being located at an index corresponding to the sources that have the smallest Kurtosis. The idea is hence to form the function

$$(3.13) \quad \Psi(\mathbf{r}) = \sum_{k=1}^K \kappa(r_k).$$

Obviously, we have  $\Psi(\mathbf{r}) \geq K \min_{k=1, \dots, K} \kappa(s_k)$ . Conversely, this lower bound cannot be achieved in general: assume for instance that  $\min_{k=1, \dots, K} \kappa(s_k)$  is reached *once*: say, the first source. The above lower bound is achieved only if  $\mathbf{F}$  is the matrix

having non-zero components on the first column only. This of course violates the constraint that  $\mathbf{F}$  is unitary. We have the following tight result:

**Proposition 11.** *As  $\kappa(s_k) < 0$  for all the sources, we have, for any unitary  $\mathbf{F}$  :*

$$(3.14) \quad \Psi(\mathbf{r}) \geq \Psi(\mathbf{s}) = \sum_{k=1}^K \kappa(s_k).$$

Moreover, the inequality is an equality if and only if  $\mathbf{F} = \mathbf{V}\mathbf{U}$  essentially equals the identity matrix.

*Proof:* by «essentially equal to», we mean that  $\mathbf{F}$  is a diagonal matrix with modulus entries, whose columns are permuted. The proof of this result follows the proof of Proposition 3 and was suggested by Comon in [12]. We indeed, express  $\Psi(\mathbf{r})$  as

$$\Psi(\mathbf{r}) = \sum_{i=1}^K \sum_{k=1}^K \kappa(s_k) |f_{i,k}|^4 = - \sum_{i=1}^K \sum_{k=1}^K \left( |\kappa(s_k)|^{1/4} |f_{i,k}| \right)^4.$$

On the other hand, we have the inequality:

$$\sum_{k=1}^K \left( |\kappa(s_k)|^{1/4} |f_{i,k}| \right)^4 \leq \left( \sum_{k=1}^K |\kappa(s_k)|^{1/2} |f_{i,k}|^2 \right)^2$$

implying that

$$(3.15) \quad -\Psi(\mathbf{r}) \leq \sum_{i=1}^K \left( \sum_{k=1}^K |\kappa(s_k)|^{1/2} |f_{i,k}|^2 \right)^2$$

We now denote by  $\mathbb{F}$  the matrix whose component  $(i, k)$  is  $|f_{i,k}|^2$  and by  $\mathbf{b}^{(1/2)} = \left( |\kappa(s_1)|^{1/2}, \dots, |\kappa(s_K)|^{1/2} \right)^T$  so that the r.h.s. of (3.15) is simply  $\|\mathbb{F} \mathbf{b}^{(1/2)}\|_2^2$ . As  $\mathbf{F}\mathbf{F}^H = \mathbf{I}_K$  we deduce that the sum of the elements of any row/column of  $\mathbb{F}$  is equal to one:  $\mathbb{F}$  is called doubly stochastic. As a consequence of the Birkhoff theorem (see [38] chapter 2),  $\mathbb{F}$  can be seen as the convex sum of permutation matrices, i.e.  $\mathbb{F} = \sum_{j=1}^K \lambda_j \mathbf{P}_j$  where, for any index  $j$ ,  $\lambda_j \geq 0$  and  $\sum_{j=1}^K \lambda_j = 1$ . We deduce that

$$\begin{aligned} \|\mathbb{F} \mathbf{b}^{(1/2)}\|_2^2 &\leq \left( \sum_{j=1}^K \lambda_j \|\mathbf{P}_j \mathbf{b}^{(1/2)}\|_2 \right)^2 = \left( \sum_{j=1}^K \lambda_j \|\mathbf{b}^{(1/2)}\|_2 \right)^2 \\ &= \sum_{j=1}^K |\kappa(s_k)|. \end{aligned}$$

This proves Equation (3.14). Let us inspect the case of equality. If equality occurs, the inequality (3.15) is necessarily an equality. Hence we have, for any indices  $i$  and  $k_1 \neq k_2$

$$\kappa(s_{k_1}) \kappa(s_{k_2}) f_{i,k_1} f_{i,k_2} = 0.$$

If all the numbers  $\kappa(s_k)$  are strictly negative, hence non-null, the above conditions implies that, for any index  $i$ , the vector  $(f_{i,1}, \dots, f_{i,K})$ , i.e. the  $i$ -th row of the matrix  $\mathbf{F}$ , has at most one non zero component. On the other hand, the matrix  $\mathbf{F}$  is unitary. This imposes that  $\mathbf{F}$  essentially equals the identity matrix  $\square$

In practice, the function  $\Psi(\mathbf{r})$  given by Equation (3.13) cannot be computed. As the data are supposed to be complex-circular at the second-order, we have

$$\kappa(r_k) = E \left[ |r_k(n)|^4 \right] - 2.$$

A consistent estimate of  $\Psi(\mathbf{r})$  is hence

$$(3.16) \quad \hat{\Psi}(\mathbf{r}) = \sum_{k=1}^K \left( \frac{1}{N} \sum_{n=1}^N |r_k(n)|^4 - 2 \right).$$

The minimization of *this* function is carried out over the space of unitary matrices. This can be done by using a Jacobi-like algorithm: unitary matrices are parametrized by means of the Givens angles. The reader may find details in [12]. Notice that no results concerning the convergence of the algorithm can be said, since it has not been shown that the local minima of the «true» function  $\Psi(\mathbf{r})$  are «good ones», i.e. achieve the BSS.

3.5.2. *Generalization: notion of contrast function.* We introduce the function

$$(3.17) \quad \Psi_\varphi(\mathbf{r}) = \sum_{k=1}^K \varphi(\kappa(r_k))$$

where  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  is a function to be specified. Rather than minimizing  $\Psi_\varphi$ , we address its maximization. If the maximum is attained when (and only when) BSS is achieved,  $\Psi_\varphi$  is called a contrast function. As seen previously, the choice  $\varphi = -\text{Id}$  makes  $\Psi_\varphi$  be a contrast function. We have, more generally:

**Proposition 12.** *If  $\varphi$  is a convex function on  $\mathbb{R}$  such that  $\varphi(0) = 0$  and if, for any source index  $k$  the condition  $\varphi(\kappa(s_k)) > 0$  is fulfilled, then  $\Psi_\varphi$  is a contrast function.*

Proof: for any index  $i$ , we have:  $\sum_k |f_{i,k}|^4 \leq \left( \sum_k |f_{i,k}|^2 \right)^2$ . As  $\mathbf{F}$  is unitary, we deduce that  $\sum_k |f_{i,k}|^2 = 1$ . Hence  $\sum_k |f_{i,k}|^4 \leq 1$ . We set  $\rho = 1 - \sum_k |f_{i,k}|^4$ . The Jensen inequality gives here:

$$\varphi \left( \sum_k \kappa(s_k) |f_{i,k}|^4 + 0 (1 - \rho) \right) \leq \sum_k \varphi(\kappa(s_k)) |f_{i,k}|^4 + \varphi(0) (1 - \rho).$$

As  $\varphi(0) = 1$ , we deduce that

$$\Psi_\varphi(\mathbf{r}) \leq \sum_i \sum_k \varphi(\kappa(s_k)) |f_{i,k}|^4.$$

By assumption all the  $\varphi(\kappa(s_k)) > 0$  hence the same argument as the one given in the proof of Proposition 11 shows that the maximum is  $\sum_k \varphi(\kappa(s_k))$  and that this maximum is reached if and only if the matrix  $\mathbf{F}$  is essentially the identity.  $\square$

The standard choice for  $\varphi$  is  $\varphi(x) = x^2$ . In this case, it is enlightening to notice that maximizing  $\Psi_\varphi$  is equivalent to minimizing:

$$(3.18) \quad \sum_{i,j,k,\ell \text{ non all equal}} \left( \text{cum} \left( r_i, r_j^*, r_k, r_\ell^* \right) \right)^2$$

which is clearly a measure of independence (up to the fourth-order).

3.5.3. *A popular algorithm: JADE.* If the indices  $j, k$  are fixed, it may be noticed that the matrix  $\mathbf{M}_{j,k}$  whose entry  $i, \ell$  is given by  $\text{cum}(x_i, x_j^*, x_k, x_\ell^*)$  admits the factorization

$$\mathbf{M}_{j,k} = \mathbf{U}\mathbf{D}_{j,k}\mathbf{U}^*$$

where  $\mathbf{D}_{j,k}$  is diagonal; the entry  $(\ell, \ell)$  is  $u_{j,\ell}^* u_{k,\ell} \kappa(s_\ell)$ . Otherwise stated,  $\mathbf{U}$  diagonalizes the normal matrix  $\mathbf{M}_{j,k}$  whatever the indices  $j, k$ . Introducing the  $\text{off}(\cdot)$  operator that sums all the entries of a matrix except the ones located on the main diagonal, this says that

$$(3.19) \quad \sum_{j,k} \text{off}(\mathbf{V}\mathbf{M}_{j,k}\mathbf{V}^*)^2$$

is minimum, equal to zero, when  $\mathbf{F} = \mathbf{V}\mathbf{U}$  essentially equals the identity matrix. Conversely, it can be shown that (3.19) can be written as (3.18). Hence a minimizer of the function given in (3.19) is a maximizer of  $\Psi_\phi(\mathbf{r})$  with  $\phi(x) = x^2$ . Now, Proposition 12 with the fact that, for any index  $k$ ,  $(\kappa(s_k))^2 > 0$  proves that the maximizers of  $\Psi_\phi(\mathbf{r})$  are such that  $\mathbf{F}$  is essentially equal to the identity matrix. This trick of algebra allows one to achieve the maximization of  $\Psi_\phi$  thanks to a joint diagonalization of a set of normal matrices. It has been proposed by Cardoso [11]. The algorithm associated with this approach is called JADE. It is very popular since efficient algorithms of joint diagonalization are known [9].

Notice that the pencil of matrices is a pencil of normal matrices hence the matrix  $\mathbf{U}$  is searched in the set of unitary matrices. Recently, Yeredor et al. [65] suggest to relax the unitary constraint. These authors even suggest to skip the pre-whitening of the data: they argue that the pre-whitening may limit the attainable performance as Cardoso pointed it out in [10]. With no whitening of the data, the matrices of cumulants are still jointly diagonalized; if the channel matrix  $\mathbf{H}$  is square, the columns of this latter form a basis for the diagonalization. The converse is not clear on the one hand and, on the other hand, the case of tall matrices  $\mathbf{H}$  remains to be addressed.

**3.6. Generalizations.** We have presented *ad hoc* BSS methods, whose theoretical foundations are solid and whose good performance is well-known. In the literature, however, many other methods can be found. They stem from considerations of information theory. We provide some key ideas and related bibliographical references.

After pre-whitening, we recall that the received data is  $\mathbf{x} = \mathbf{U}\mathbf{s}$  (we have dropped the time index): on the one hand,  $\mathbf{U}$  is unitary and on the other hand, the component of the random variables in  $\mathbf{s}$  are mutually independent. The idea of independent Component Analysis (ICA) is hence to exhibit matrices  $\mathbf{V}$  such that the components of  $\mathbf{r} = \mathbf{V}\mathbf{x}$  are «as much independent as possible». This independency may be measured by the Kullback-Leibler divergence between the distribution of  $\mathbf{r}$  and this of the product of the marginals: this is the mutual information  $\mathcal{I}(\mathbf{r})$ . It is well-known that  $\mathcal{I}(\mathbf{r}) \geq 0$  with equality iff the components of  $\mathbf{r}$  are independent.

Besides, it has been underlined by Comon in [12] that the Darmois theorem states the fact: as the sources are non-Gaussian, the fact that the components of  $\mathbf{r}$  are pair-wise independent (which is naturally the case if they are mutually independent) implies that  $\mathbf{F} = \mathbf{V}\mathbf{U}$  is essentially equal to the identity matrix. Hence, the minimization of the mutual information of  $\mathcal{I}(\mathbf{r})$  is legitimate. This

induces of course no BSS algorithm, since the mutual information cannot be simply estimated.

Now, Comon underlines in the seminal paper [12] that  $\mathcal{I}(\mathbf{r}) = \mathcal{J}(\mathbf{r}) - \sum_{k=1}^K \mathcal{J}(r_k)$  where  $\mathcal{J}(\mathbf{r})$  is the negentropy of the vector  $\mathbf{r}$ , i.e.  $\mathcal{J}(\mathbf{r}) = \mathcal{H}(\mathbf{r}_{\text{gaussian}}) - \mathcal{H}(\mathbf{r})$ : here,  $\mathcal{H}(\mathbf{r})$  is the differential entropy of  $\mathbf{r}$  and  $\mathcal{H}(\mathbf{r}_{\text{gaussian}})$  the differential entropy of the Gaussian vector whose mean and covariance matrix are those of  $\mathbf{r}$ . The negentropy shows the nice property of invariance w.r.t. any invertible change of variables: hence  $\mathcal{J}(\mathbf{r})$  does not depend on  $\mathbf{F}$ . The independence is then obtained by maximizing

$$(3.20) \quad \sum_{k=1}^K \mathcal{J}(r_k).$$

Notice that  $\mathcal{J}(r_k) \geq 0$  with equality when  $r_k$  is Gaussian. Hence the maximization in question tends to maximize the distance of the reconstructed source  $\mathbf{r}$  to the Gaussian case.

On the other hand, this shows that, in order to achieve independency, it suffices to consider the maximization of a sum of functions, each of which simply depends on a component of the reconstructed source vector. Evidently, the maximization of the function  $\mathcal{J}(r_1)$  under the constraint that the first row of  $\mathbf{F}$  has norm one, is achieved when  $r_1$  coincides with one of the sources up to a modulus one factor. In this respect, this remark provides a justification of the iterative methods proposed in Section 3.4, even if the function considered are not the negentropy. As far as the function  $\Psi_\phi(\mathbf{r})$  is concerned, notice that it has the form (3.20).

Comon proposes an approximation of the negentropy, based on the Edgeworth expansion of the probability density functions of the random variables  $r_k$ . Thanks to the circularity assumption, this approximation is  $\mathcal{J}(r_k) \approx \frac{1}{48} \kappa(r_k)^2$ . This calls for an important remark: the function  $\Psi_\phi(\mathbf{r})$  given in Equation (3.17) with  $\phi(x) = x^2$  is hence closely connected to a measure of independence - this confirms a remark previously done.

Hyvarinen departs from functions based on the cumulants. In [28], it is suggested to consider a wider class of functions which are not directly related to cumulants. In short, the new functions to be maximized are more robust to outliers. The price to be paid is the weaker results concerning the separation: for instance, the precious results concerning the separability of the local maxima do not hold. An efficient algorithm has given rise to the popular method called FastICA. In the original paper [28], the sources are real valued which is not the case in this paper. An extension to complex-valued sources can be found in [7].

**3.7. Iterative BSS (general cyclo-stationary case).** We specified that the assumption of stationarity of the sources, as required in the previous sections, is somewhat restrictive in a realistic scenario of telecommunication. Indeed, the stationarity implicitly assumes that all the sources have the same symbol period and that the data are sampled at a period equal to the symbol period. In general - think for instance of a passive listening context - the sources have different baud-rates. We denote the symbol periods of the  $K$  sources by  $T_1, \dots, T_K$ . If  $T_e$  is the sampling period - a priori different of any of the symbol periods - we have deduce from Section 2 that, for any index  $k$ , the source  $s_k(n)$  is cyclostationary. In particular, the second moment  $E[|s_k(n)|^2]$  varies with the time-index  $n$ . More specifically, we have

$$E \left[ |s_k(n)|^2 \right] = \sum_{\alpha \in \mathcal{I}_k} R_{s_k}^{(\alpha)} e^{i2\pi\alpha_k n}$$

where

$$\mathcal{I}_k = \left\{ 0, \pm \frac{T_e}{T_k} \right\}$$

is the set of the second-order cyclic frequencies of  $s_k$  and the Fourier coefficients  $R_{s_k}^{(\alpha)}$  are given by Equation (2.8) (we have dropped the time-lag in  $R_{s_k}^{(\alpha)}(\ell)$  since, in the sequel, no other time delay than  $\ell = 0$  is considered). In this section, the channel is supposed to be memoryless, as in all the Section 3. Hence, the model given by Equation (2.11) still holds: the components of the source vector  $\mathbf{s}(n)$  are effectively mutually independent, but are not individually stationary. As far as the normalization of the sources is concerned, it may also be assumed: in the cyclo-stationary context of this section, this means that

$$(3.21) \quad \langle E [\mathbf{s}(n)\mathbf{s}(n)^H] \rangle_n = \mathbf{I}_K$$

In the following, we provide assumptions on the sources that guarantee that the algorithms of source separation depicted in Section 3.4 still converge to desirable solutions, i.e. allow one to separate the sources. In other words, the algorithms previously considered are run as if the data *were* stationary: this means that nothing has to be changed in any part of the stationary BSS algorithms. Surprisingly, the fact that the data are not stationary is *shown* not to impact the convergence to a good (separating) solution.

The algorithms encountered for stationary data (see Section 3.4) are designed to minimize either the function  $\hat{\Upsilon}(r)$  given by Equation (3.10) or the Godard function  $\hat{\Upsilon}_{CM}(r)$  given by Equation (3.11). Let us recall that the signal  $r(n)$  is the output of the variable spatial filter  $\mathbf{g}$ , i.e.

$$\begin{aligned} r(n) &= \mathbf{g}^T \mathbf{y}(n) \\ &= \mathbf{f}^T \mathbf{s}(n) \end{aligned}$$

where  $\mathbf{f}^T = \mathbf{g}^T \mathbf{H}$ . As in the stationary case, we analyze the argument minima of the theoretical associated function. The first point to be addressed is hence: to which functions  $\hat{\Upsilon}(r)$  and  $\hat{\Upsilon}_{CM}(r)$  converge? Due to the non-stationarity of the model, the function  $\hat{\Upsilon}(r)$  for instance does not converge to  $\Upsilon(r)$  as given in Equation (3.9): this cannot be the case since the latter function depends on the time-lag  $n$ .

There is a version of the law of the large numbers for non-stationary data. Lemma 8 can be written in this context under this form:

**Lemma 13.** *For  $p = 2, 4$ , we have, with probability one:*

$$\frac{1}{N} \sum_{n=0}^{N-1} |r(n)|^p \longrightarrow \langle E [|r(n)|^p] \rangle.$$

We deduce that  $\hat{\Upsilon}_{CM}(r) \longrightarrow \left\langle E \left[ \left( |r(n)|^2 - 1 \right)^2 \right] \right\rangle$ . We define this limit as (the superscript «c» means «cyclo-stationary»):

$$(3.22) \quad \Upsilon_{CM}^{(c)}(r) = \left\langle E \left[ \left( |r(n)|^2 - 1 \right)^2 \right] \right\rangle_n .$$

For the same reason, we have  $\hat{\Upsilon}(r) \longrightarrow \Upsilon^{(c)}(r)$  where

$$(3.23) \quad \Upsilon^{(c)}(r) = \frac{\left\langle E \left[ |r(n)|^4 \right] \right\rangle_n}{\left( \left\langle E \left[ |r(n)|^2 \right] \right\rangle_n \right)^2} - 2$$

Notice that, in the case of stationary data,  $\Upsilon^{(c)}(r)$  (respectively  $\Upsilon_{CM}^{(c)}(r)$ ) equals  $\Upsilon(r)$  (respectively  $\Upsilon_{CM}(r)$ ).

We first address the minimization of  $\Upsilon^{(c)}(r)$ . Once done, we deduce results concerning the minimization of  $\Upsilon_{CM}^{(c)}(r)$ .

We consider a prior expansion of the moments involved in  $\Upsilon^{(c)}(r)$ . The following equality always holds true (we recall that the signals are all complex-circular at the second-order):

$$(3.24) \quad E \left[ |r(n)|^4 \right] = \kappa(r(n)) + 2 \left( E \left[ |r(n)|^2 \right] \right)^2 .$$

The multi-linearity of the cumulant gives:

$$\langle \kappa(r(n)) \rangle_n = \sum_{k=1}^K \eta(s_k) |f_k|^4$$

where we let for any source  $s = s_k$  the number  $\eta(s)$  be

$$(3.25) \quad \eta(s) = \langle \kappa(s(n)) \rangle_n .$$

The output of the spatial filter,  $r(n)$ , is obviously a linear combination of the sources  $s_1, \dots, s_K$ :  $r(n) = \sum_{k=1}^K f_k s_k(n)$ . As  $s_k$  admits  $\mathcal{I}_k = \{0, \pm\alpha_k\}$  as the set of its second-order cyclic frequencies, we deduce that  $r(n)$  is cyclo-stationary and its second-order cyclo-frequencies are in the set

$$\mathcal{I} = \bigcup_{k=1}^K \mathcal{I}_k .$$

This means that  $E \left[ |r(n)|^2 \right] = \sum_{\alpha \in \mathcal{I}} R_r^{(\alpha)} e^{i2\pi\alpha n}$ . As a consequence, the term  $\left\langle \left( E \left[ |r(n)|^2 \right] \right)^2 \right\rangle_n$ , may be computed thanks to the Parseval equality. We have indeed

$$\left\langle \left| E \left[ |r(n)|^2 \right] \right|^2 \right\rangle_n = \sum_{\alpha \in \mathcal{I}} \left| R_r^{(\alpha)} \right|^2 .$$

On the other hand, the sources are mutually decorrelated hence

$$\begin{aligned} R_r^{(\alpha)} &= \sum_{k=1}^K R_{f_k s_k}^{(\alpha)} \\ &= \sum_{k=1}^K |f_k|^2 R_{s_k}^{(\alpha)}. \end{aligned}$$

By definition,  $\mathcal{I}^* = \mathcal{I} \setminus \{0\}$  is the set of all the non-null cyclic frequencies of the mixture. After isolating the term associated with the cyclic frequency  $\alpha = 0$ , we may hence write that

$$\begin{aligned} \left\langle \left( E \left[ |r(n)|^2 \right] \right)^2 \right\rangle_n &= \left( R_r^{(0)} \right)^2 + \sum_{\alpha \in \mathcal{I}^*} \left| \sum_{k=1}^K |f_k|^2 R_{s_k}^{(\alpha)} \right|^2 \\ &= \left( R_r^{(0)} \right)^2 + \sum_{k=1}^K |f_k|^4 \left( \sum_{\alpha \in \mathcal{I}_k^*} |R_{s_k}^{(\alpha)}|^2 \right) + \frac{1}{2} \sum_{k_1 \neq k_2} |f_{k_1}|^2 |f_{k_2}|^2 \varepsilon(s_{k_1}, s_{k_2}) \\ &= \left( R_r^{(0)} \right)^2 + 2 \sum_{k=1}^K |f_k|^4 |R_{s_k}^{(\alpha_k)}|^2 + \frac{1}{2} \sum_{k_1 \neq k_2} |f_{k_1}|^2 |f_{k_2}|^2 \varepsilon(s_{k_1}, s_{k_2}) \end{aligned}$$

where

$$(3.26) \quad \varepsilon(s_{k_1}, s_{k_2}) = 2 \sum_{\alpha \in \mathcal{I}^*} R_{s_{k_1}}^{(\alpha)} R_{s_{k_2}}^{(\alpha)*}.$$

We have used the symmetry:

$$(3.27) \quad R_{s_k}^{(-\alpha_k)} = R_{s_k}^{(\alpha_k)*}.$$

The set  $\mathcal{I}^*$  is simply  $\{\pm\alpha_k \mid k = 1, \dots, K\}$ . Denoting by  $\mathcal{I}_+^*$  the set  $\{\alpha_k \mid k = 1, \dots, K\}$  we have, due to (3.27), the alternative expression of  $\varepsilon$  that underlines that  $\varepsilon$  is real:

$$(3.28) \quad \varepsilon(s_{k_1}, s_{k_2}) = 4 \operatorname{Re} \left( \sum_{\alpha \in \mathcal{I}_+^*} R_{s_{k_1}}^{(\alpha)} R_{s_{k_2}}^{(\alpha)*} \right).$$

We hence have the following expression:

$$(3.29) \quad \Upsilon^{(c)}(r) = \frac{\sum_{k=1}^K \zeta(s_k) |f_k|^4 + \sum_{k_1 \neq k_2} |f_{k_1}|^2 |f_{k_2}|^2 \varepsilon(s_{k_1}, s_{k_2})}{\left( \sum_{k=1}^K |f_k|^2 \right)^2}.$$

where we have let, for any of the sources  $s = s_k$  whose positive cyclic frequency is  $\alpha$  the number  $\zeta(s)$  be

$$(3.30) \quad \zeta(s) = \eta(s) + 4 |R_s^{(\alpha)}|^2.$$

It is to be noticed that  $\zeta(s)$  can also be expressed as

$$(3.31) \quad \zeta(s) = \left\langle E \left[ |s(n)|^4 \right] \right\rangle_n - 2.$$

We are in position to discuss the minimization of this function. Two cases have to be distinguished.

3.7.1. *Case of different baud-rates.* If the cyclic frequencies  $\pm \frac{T_e}{T_k}$  are all different modulo 1, the coefficients  $\varepsilon(s_{k_1}, s_{k_2})$  all verify:

$$(3.32) \quad \forall k_1 \neq k_2 \quad \varepsilon(s_{k_1}, s_{k_2}) = 0.$$

Indeed, it suffices to consider the expression of Equation (3.26): let  $\alpha_k$  be the positive cyclo-frequency of the source no  $k$ :  $\alpha_k = \frac{T_e}{T_k}$ . In Equation (3.26), there are at most two terms:  $\alpha$  is either  $\alpha_{k_1}$  or  $-\alpha_{k_1} = 1 - \alpha_{k_1}$  modulo 1. Take for instance  $\alpha = \alpha_{k_1}$ . The associated term is non null if  $\alpha = \alpha_{k_2}$  or  $\alpha = -\alpha_{k_2} = 1 - \alpha_{k_2}$ : if  $\pm\alpha_{k_1}$  and  $\pm\alpha_{k_2}$  are different modulo 1, this cannot occur. The same reasoning applies for  $\alpha = -\alpha_{k_1}$ . In other words,  $\varepsilon(s_{k_1}, s_{k_2}) = 0$ .

*Remark.* It is desirable to bridge the gap between this condition on the disparity of the cyclo-frequencies and practical considerations. We insist on the fact that the above condition is not equivalent to condition that the symbol frequencies are different. Indeed, consider  $\alpha_1 = 1/8$  and  $\alpha_2 = 7/8$ . Then  $-\alpha_2 = \alpha_1$  modulo 1 and the condition is not fulfilled whereas the symbol periods are different. However, if the sampling frequency is big enough such that for all indices  $k$ :  $\alpha_k < 1/2$  then the condition on the disparity of the cyclic frequencies simply means: all the symbol periods are different. In the following, we systematically assume the condition

$$(3.33) \quad \alpha_k < \frac{1}{2}.$$

Proposition 3 can be written in the non-stationary context:

**Proposition 14.** *If the cyclic frequencies  $\pm \frac{T_e}{T_k}$  are all different modulo 1, then the cost function  $\Upsilon^{(c)}(r)$  achieves its minimum at a separating vector if and only if one of the  $\zeta(s_k)$  is strictly negative. Moreover, the separating vector in question has its single non-null element located at an index  $k_0$  such that  $\zeta(s_{k_0}) = \min_{k=1, \dots, K} \zeta(s_k)$ .*

We hence consider the following assumption

$$(3.34) \quad \mathcal{A}_1 \quad \text{for any of source } s \text{ of the mixture} \quad \zeta(s) < 0.$$

that is analyzed in Section 3.9.

Together with Proposition 14, this assumption  $\mathcal{A}_1$  allows one to claim that the minimization of the estimated function  $\hat{\Upsilon}(r)$  is legitimate as far as the extraction of a source is concerned. Moreover, the result of Lemma 5 still holds: indeed, the expression of  $\Upsilon^{(c)}(r)$  given by (3.29) simplifies as

$$\Upsilon^{(c)}(r) = \frac{\sum_{k=1}^K \zeta(s_k) |f_k|^4}{\left(\sum_{k=1}^K |f_k|^2\right)^2}$$

which is formally the same as the function  $\Upsilon(r)$  considered in the stationary case - see Equation (3.8). The local minima of  $\Upsilon^{(c)}(r)$  are separating, which makes the minimization algorithm based on a gradient method robust to the presence of local minima.

We now face the minimization of the function  $\Upsilon_{CM}^{(c)}(r)$ . The question is whether the non-stationarity of the data changes such a result as the one given in Proposition 7 or not. The careful reader will be assured that indeed Proposition 7 is true in the non-stationary case. Moreover, the result of Proposition 4 is unchanged in the cyclo-stationary case. We can claim:

**Proposition 15.** *If the cyclic frequencies  $\pm \frac{T_e}{T_k}$  are all different modulo 1, and if Assumption  $\mathcal{A}_1$  in (3.34) holds, then the local minima of the cost functions  $\Upsilon^{(c)}(r)$  and  $\Upsilon_{CM}^{(c)}$  are separating.*

3.7.2. *General case.* Contrary to the case where the cyclic-frequencies are different, the Equation (3.32) does not hold in general, and the results given in the stationary context concerning the minimization of  $\Upsilon^{(c)}(r)$  cannot be directly recast. However, numerical considerations may help to give similar results. In this respect, the reader should understand that the quantities  $\varepsilon(s_{k_1}, s_{k_2})$  are either zero or «small» since they involve cyclo-correlation at non-null frequencies. This is specific of telecommunication signals; the reason is due to the fact that the excess bandwidth of a transmitted signal is small (see [39]). We will discuss this fact further on, but for the moment, we simply consider the following assumption for any of the sources, denoted by  $s$ :

$$(3.35) \quad \mathcal{A}_2 \quad \text{for any source } s \text{ of the mixture} \quad |R_s^{(\alpha)}| < \frac{1}{2} \sqrt{|\zeta(s)|}$$

We discuss this assumption further in section 3.9. For the moment, we suppose that Assumptions  $\mathcal{A}_1$  and  $\mathcal{A}_2$  both hold.

Consider two distinct indices  $k_1$  and  $k_2$ ; for sake of simplicity, let them be  $k_1 = 1$  and  $k_2 = 2$ . Notice that if the sources numbered 1 and 2 are such that their associated (single) cyclic frequencies are different modulo 1 then  $\varepsilon(s_1, s_2) = 0$ . Suppose on the contrary now that the cyclic frequencies are equal modulo 1 (thanks to (3.33) this happens if and only if the two baud-rates are the same). In this case, the summation (3.28) has only one term and  $\varepsilon(s_1, s_2) = 4\mathcal{R}(R_{s_1}^{(\alpha)} R_{s_2}^{(\alpha)*})$ . Assumption  $\mathcal{A}_2$  given in (3.35) holds, it follows *in both cases* that  $|\varepsilon(s_1, s_2)| < \sqrt{|\zeta(s_1)||\zeta(s_2)|}$ . Whatever the indices  $k_1 \neq k_2$  may be, Assumption  $\mathcal{A}_2$  implies

$$(3.36) \quad \forall k_1 \neq k_2 \quad |\varepsilon(s_{k_1}, s_{k_2})| < \sqrt{|\zeta(s_{k_1})\zeta(s_{k_2})|}.$$

We may consider the expression of  $\Upsilon^{(c)}(r)$  given in Equation (3.29). We have, thanks to (3.36):

$$(3.37) \quad \sum_{k=1}^K \zeta(s_k) |f_k|^4 + \sum_{k_1 \neq k_2}^K |f_{k_1}|^2 |f_{k_2}|^2 \varepsilon(s_{k_1}, s_{k_2}) \geq - \sum_{k=1}^K |\zeta(s_k)| |f_k|^4 - \sum_{k_1 \neq k_2} |f_{k_1}|^2 |f_{k_2}|^2 \sqrt{|\zeta(s_{k_1})\zeta(s_{k_2})|}.$$

As the r.h.s of the latter equation is simply  $-\left(\sum_k \sqrt{|\zeta(s_k)|} |f_k|^2\right)^2$ , this shows that

$$\Upsilon^{(c)}(r) \geq - \left( \frac{\sum_k \sqrt{|\zeta(s_k)|} |f_k|^2}{\sum_k |f_k|^2} \right)^2 \geq - \max_k |\zeta(s_k)|.$$

If Assumption  $\mathcal{A}_1$  holds, this lower bound is evidently reached. It remains to inspect the cases of equality for this lower-bound. In this respect, we suppose that  $\Upsilon^{(c)}(r) = \zeta_{\min}$  where  $\zeta_{\min} = \min_k \zeta(s_k)$ . For convenience, we assume that  $\sum_{k=1}^p |f_k|^2 = 1$ . The case of equality implies that (3.37) is an equality. It implies that

$$\sum_{k_1 \neq k_2} |f_{k_1}|^2 |f_{k_2}|^2 \left( \varepsilon(s_{k_1}, s_{k_2}) + \sqrt{|\zeta(s_{k_1})\zeta(s_{k_2})|} \right) = 0.$$

Now, the inequality  $\mathcal{A}_2$  implies that  $\varepsilon(s_{k_1}, s_{k_2}) + \sqrt{|\zeta(s_{k_1})\zeta(s_{k_2})|} > 0$ . Necessarily, we must have that, for every couples  $(k_1, k_2)$  such that  $k_1 \neq k_2$ ,  $f_{k_1}f_{k_2} = 0$ . Otherwise stated: the vector  $\mathbf{f}$  has a single non-null component, hence is separating.

We have shown the result:

**Proposition 16.** *If the Assumptions  $\mathcal{A}_1$  and  $\mathcal{A}_2$  hold, then the cost function  $\Upsilon^{(c)}(r)$  and  $\Upsilon_{CM}^{(c)}$  achieve their minimum at a separating vector.*

**3.8. Global BSS (general cyclo-stationary case)** . Again in this section, the global source separation method depicted in Section 3.5 is considered. Its convergence was specifically shown for stationary sources. We show that the convergence of the algorithm to a separating matrix is not affected when the data are cyclo-stationary. This key-result was first provided in [18]; a condition of separability of JADE was given, which was rather difficult to interpret especially when the number of sources is greater than 3. We show the result quite differently, following [30].

Notice that the pre-whitening of the observed data is still possible even when the data are cyclo-stationary: the algorithm is even not changed at all. Hence, we begin by considering the estimate  $\hat{\Psi}(\mathbf{r})$  given by Equation (3.16). Obviously, this estimate can not converge to the function of the cumulants given by Equation (3.13) since the terms in this equation depend on the time lag. Nevertheless, the limit as the number of snapshots grows can be expressed as

$$\Psi^{(c)}(\mathbf{r}) = \sum_{i=1}^K \left( \left\langle E \left[ |r_i(n)|^4 \right] \right\rangle_n - 2 \right).$$

Thanks to the algebra already done along Section 3.7, we may directly write

$$\Psi^{(c)}(\mathbf{r}) = \sum_{i=1}^K \left( \sum_{k=1}^K \zeta(s_k) |f_{i,k}|^4 + \sum_{k_1 \neq k_2} |f_{i,k_1}|^2 |f_{i,k_2}|^2 \varepsilon(s_{k_1}, s_{k_2}) \right).$$

It remains little to do as soon as Assumptions  $\mathcal{A}_1$  and  $\mathcal{A}_2$  hold, since the following inequality holds:

$$\Psi^{(c)}(\mathbf{r}) \geq - \sum_{i=1}^K \left( \sum_{k=1}^K \sqrt{|\zeta(s_k)|} |f_{i,k}|^2 \right)^2 = \left\| \mathbf{F} \mathbf{z}^{(1/2)} \right\|_2^2$$

where  $\mathbf{z}^{(1/2)} = (|\zeta(s_1)|, \dots, |\zeta(s_K)|)^T$ . The same argument as the one given in Section 3.5 (Birkhoff theorem) proves that  $\Psi^{(c)}(\mathbf{r}) \geq - \sum_{k=1}^K |\zeta(s_k)|$ . As the numbers  $\zeta(s_k)$ , thanks to Assumption  $\mathcal{A}_1$ , are strictly negative, we directly show that  $-\sum_{k=1}^K |\zeta(s_k)|$  is the infimum of  $\Psi^{(c)}(\mathbf{r})$  and this infimum is reached for unitary matrices  $\mathbf{F}$  that are essentially equal to the identity matrix. This proves the following proposition, that is the sister of Proposition 11:

**Proposition 17.** *If Assumptions  $\mathcal{A}_1$  and  $\mathcal{A}_2$  hold, we have, for any unitary  $\mathbf{F}$  :*

$$\Psi^{(c)}(\mathbf{r}) \geq \Psi^{(c)}(\mathbf{s}) = \sum_{k=1}^K \zeta(s_k).$$

*Moreover, the inequality is an equality if and only if  $\mathbf{F}$  essentially equals the identity matrix.*

### 3.9. Validity of Assumptions $\mathcal{A}_1$ and $\mathcal{A}_2$ : semi-analytical considerations.

In essence, we have shown that the cyclo-stationarity of the data does not affect neither the one-by-one methods of Section 3.7, nor the global method depicted in Section 3.5. This positive answer to our question however requires to inspect the validity of Assumptions  $\mathcal{A}_1$  and  $\mathcal{A}_2$ . We first discuss  $\mathcal{A}_1$ . We recall that, in a stationary environment,  $\zeta(s)$  is simply the fourth-order cumulant of the (normalized) source. As this latter is a filtered version of an i.i.d. sequence of symbols having a strictly negative Kurtosis, it is straight-forward that  $\zeta(s) < 0$ . In a cyclo-stationary environment, the mentioned argument does not hold anymore. In this case, indeed,  $\zeta(s)$  is given by (3.30) and it is not possible to conclude directly that  $\zeta(s) < 0$ . However, these statistics can be shown to express as integrals of the shaping filter  $c_a(t)$ , at least for a generic sampling frequency (see [31]). More precisely, except for four sampling frequencies that are irrelevant to consider, it can be shown quite simply that

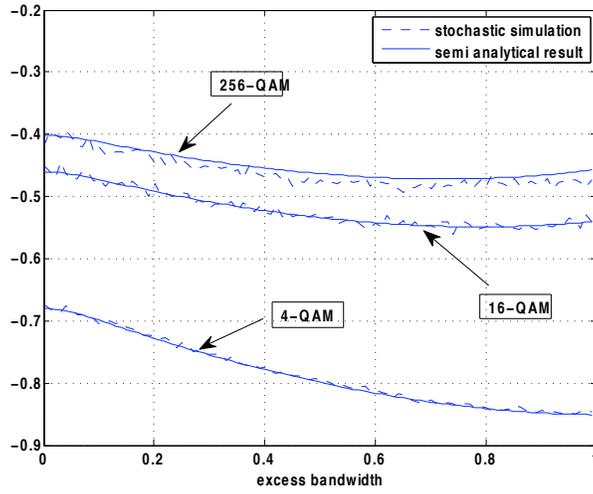
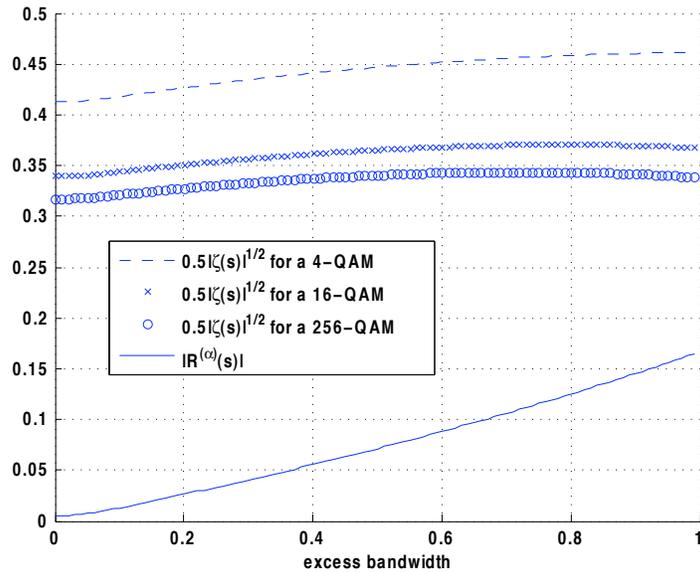
$$R_s^{(\alpha)} = \frac{1}{T} \int |c_a(t)|^2 e^{-i2\pi t/T} dt$$

and

$$\zeta(s) = \kappa(d) \frac{1}{T} \int |c_a(t)|^4 dt + 4 \left| R_s^{(\alpha)} \right|^2$$

where  $c_a$ , as specified in the introduction, is a normalized square-root raised-cosine filter. In [31], it is shown rigorously that  $R_s^{(\alpha)}$  and  $\zeta(s)$  do not depend on the symbol period  $T$ . It is hence possible to compute numerically  $\zeta(s)$  as a function of the excess bandwidth, while considering a few values of  $\kappa(d)$  corresponding to typical modulations. The reader may find details on the computation of these numbers in the above reference. The results are reported in Figure 3.1. In order to validate that the above formulas are correct, we also have plotted the estimate of  $\zeta(s)$  obtained by stochastic simulations with respect to Equation (3.31) (we have generated sequences of 10000 symbols according to the modulation). For all the values of the excess bandwidth, the numerical results let us claim that  $\mathcal{A}_1$  is true.

At a first sight, Assumption  $\mathcal{A}_2$  looks quite audacious; indeed it is. As far as we know, it is not possible, for the same reasons as those mentioned above, to prove this result analytically. Resorting again to semi-analytical considerations, we compute the numbers  $\frac{1}{2} \sqrt{|\zeta(s)|}$  and  $\left| R_s^{(\alpha)} \right|$  for different modulations and excess-bandwidth factors. We have obtained the results that can be seen in Figure 3.2. These computations indicate that  $\mathcal{A}_2$  can be claimed to hold.

FIGURE 3.1. validity of  $\mathcal{A}_1$ FIGURE 3.2. validity of  $\mathcal{A}_2$ 

#### 4. CONVOLUTIVE MIXTURES: CASE OF SPARSE CHANNELS

We now face the problem of blind source separation when the channel is not memoryless. The contribution of a given source to received data is not simply the delayed source up to an unknown constant, but a filtered version of the source. In this section, we specify the model and we explain how it is possible to connect quite

directly the previous results (on instantaneous mixtures) to this model: this requires a strong assumption on the delays as will be explained. We refer to [15](chapter 17) for more details.

The convolutive effect stems from the presence of multiple paths as specified in Section 2.1. Consider a source  $s_{a,k}(t)$ ; the contribution of this source on the received signal ( $m$ -th sensor) is given by (2.9). In most of the high-rate digital communication systems, the narrow-band assumption may be taken for granted. By narrow-band signal, we mean that the carrier frequency is big enough to consider the signal as monochromatic. More specifically, this means that the bandwidth of the signal in baseband is much smaller than the carrier frequency, i.e.

$$\frac{1}{T_k} \ll f_k.$$

On the other hand, for any delay  $\tau$  such that  $\tau \ll T_k$ , we have the approximation:  $s_{a,k}(t - \tau) \approx s_{a,k}(t)$ . In order to take the full benefit of the antenna array, the distance between two consecutive antennas should be of the order of half the wavelength. This says that the delay of propagation between two consecutive antennas is of the order of  $\frac{1}{2f_k} \ll T_k$ . As a consequence, the contribution of the  $k$ -th source and  $\ell$ -th path to the mixture is a rank one signal and can be written as  $\mathbf{h}_{k,\ell} s_a(t - \tau_{k,\ell})$  where  $\mathbf{h}_{k,\ell}$  is called *steering vector*.

We assume that the associated delays are sorted such that  $\tau_{1,k} < \tau_{2,k} < \dots$ . We consider the case when the delays are «sufficiently spread out». In this respect, the reader should notice that it is possible to use the fact that the shaping function  $c_{a,k}$  numerically vanishes:  $c_{a,k}(t)$  can be numerically neglected if  $|t| > P_k T_k$  where the integer  $P_k$  depends on the roll-off. This implies that

$$s_{a,k}(t) \approx \sum_{\ell = \lfloor \frac{t}{T_k} \rfloor - P_k}^{\lfloor \frac{t}{T_k} \rfloor + P_k} d_k(\ell) c_{a,k}(t - \ell T_k).$$

Consider that the above approximation is an equality. As a consequence, if  $\tau_{2,k} - \tau_{1,k} > 2P_k T_k$  then the random variables  $s_{a,k}(t - \tau_{1,k})$  and  $s_{a,k}(t - \tau_{2,k})$  do not involve the same symbols hence they are independent. Two consecutive paths can hence be treated as *independent sources*. Suppose that all the successive delays are separated by more than  $2P_k T_k$ , this property holding true for all the sources. As  $L_k$  denotes the number of paths associated with the  $k$ -th source, we may write the observed data according to Equation (2.11). This time, the «source vector»  $\mathbf{s}_a(t)$  is defined as

$$\left( \underbrace{s_{a,1}(t - \tau_{1,1}), \dots, s_{a,1}(t - \tau_{1,L_1})}_{L_1 \text{ "sources"}}, \dots, \underbrace{s_{a,K}(t - \tau_{1,K}), \dots, s_{a,K}(t - \tau_{L_K,K})}_{L_K \text{ "sources"}} \right)^T.$$

The  $\tilde{K} = L_1 + L_2 + \dots + L_K \geq K$  «sources» in the vector  $\mathbf{s}_a(t)$  are (approximately) mutually independent. The advantage of this formulation is that all the results given in Section 3.7 remain true. Nevertheless, we want to stress the drawbacks of this approach.

- (1) A fundamental requirement is that the number of sensors,  $M$ , is greater than the number of sources. We recall that this necessary condition comes

from the requirement that the mixing matrix  $\mathbf{H}$  should have full -column rank. This gives here

$$M \geq \tilde{K}.$$

This condition may be limiting as soon as the number of paths per communication source is big.

- (2) The source separation methods all require that the components of  $\mathbf{s}_a(t)$  be mutually independent. Now, the delays are spread out in very specific conditions; this particular condition occurs in long-range communication channels (ionospheric transmissions): we refer to the reference [15](chapter 17) for details. On the contrary, this assumption on the distribution of the delays associated with a given source is scarcely fulfilled in many cases (urban GSM channels for instance).
- (3) The complexity of the instantaneous (one-by-one or global) methods directly increases with the number of sources.
- (4) The algorithms of source separation, ideally, should provide, up to constants, all the components of  $\mathbf{s}_a(t)$ . For instance,  $L_1$  of these «sources» among the  $\tilde{K}$  components involve the sequence of symbols  $d_1(n)$ . As the ultimate goal is to eventually estimate these transmitted symbols, a recombination of the  $L_1$  «sources» has to be computed. Notice that the sources are generally not ordered in any way. The association of the reconstructed paths can be processed after lagged correlation between the reconstructed sources. This is clearly not a child's play.

For practical considerations, we refer to Chapter 17 of reference [15] where a comparison between instantaneous (described in this section) and convolutive approaches described in the next section are done.

## 5. CONVOLUTIVE MIXTURES

We now face the case of the general multi-path channels; no such condition as the sparsity of the channels is assumed to hold.

**5.1. Identifying the symbols: algebraic methods (stationary data).** In this section, the model of the data is stationary and is given by (2.18). We have justified in the introduction that it is legitimate to approximate the MIMO filter  $\mathbf{E}(z)$  by a polynomial. We denote by  $L$  its order, i.e.

$\mathbf{E}(z) = \sum_{\ell=0}^L \mathbf{E}_\ell z^{-\ell}$ . The approaches that we want to introduce exploit algebraic properties of the model (convolution) and of the source signal  $\mathbf{d}(n)$ . Van der Veen et al. [62] suggest the following method.

Consider the vector

$$(5.1) \quad \mathbf{y}_P(n) = (\mathbf{y}(n)^T, \mathbf{y}(n-1)^T, \dots, \mathbf{y}(n-(P-1))^T)^T$$

for  $P \geq L$ . We remark that

$$\mathbf{y}_P(n) = \mathcal{T}_P(\mathbf{E})\mathbf{d}_{P+L-1}(n)$$

where  $\mathbf{d}_{P+L-1}(n)$  is defined in the same manner as  $\mathbf{y}_P(n)$  and where matrix  $\mathcal{T}_P(\mathbf{E})$  is the  $MP \times K(L+P)$  block-Toeplitz matrix given by

$$(5.2) \quad \mathcal{T}_P(\mathbf{E}) = \begin{pmatrix} \mathbf{E}_0 & \dots & \mathbf{E}_L & 0 & \dots & 0 \\ 0 & \mathbf{E}_0 & \dots & \mathbf{E}_L & 0 & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \mathbf{E}_0 & \dots & \mathbf{E}_L \end{pmatrix},$$

where we have set  $\mathbf{E} = (\mathbf{E}_0, \dots, \mathbf{E}_L)$ . Denote by  $\mathcal{Y}_P$  the  $MP \times (N-P-L)$  matrix ( $N$  is the number of snapshots) in which all the data  $\mathbf{y}(n)$  from  $n = L+P$  to  $n = N$  are collected ( $P \geq L$ ): the  $n$ -th column of  $\mathcal{Y}_P$  is  $\mathbf{y}_P(n)$  - see (5.1). It yields  $\mathcal{Y}_P = \mathcal{T}_P(\mathbf{E}) \mathcal{D}_{P+L,N}$  where  $\mathcal{D}_{P+L,N}$  is the block-Toeplitz matrix defined by

$$\mathcal{D}_{P+L,N} = \begin{pmatrix} \mathbf{d}(P+L-1) & \mathbf{d}(P+L) & & \mathbf{d}(N) \\ \mathbf{d}(P+L-2) & \mathbf{d}(P+L-1) & \ddots & \\ \vdots & & \ddots & \\ \mathbf{d}(0) & \mathbf{d}(1) & & \mathbf{d}(N-P-L+1) \end{pmatrix}.$$

It is known (see [29, 1]) that the assumption given in (2.21) involves that  $\mathcal{T}_P(\mathbf{E})$  is a full column rank matrix. We deduce that the row space of  $\mathcal{Y}_P$  equals the row space of  $\mathcal{D}_{P+L,N}$ . An idea consists in finding the matrices  $\hat{\mathcal{D}}_{P+L,N}$  having the Toeplitz structure of  $\mathcal{D}_{P+L,N}$  such that their row space is prescribed. Notice that an ambiguity arises since the estimated  $\hat{\mathbf{d}}(n)$  of  $\mathbf{d}(n)$  coincide up to an invertible matrix. This latter can be removed by considering one of the algebraic methods (instantaneous mixtures) evoked in Section 3.2 in which a priori information on the symbols is exploited (finite alphabets or constant modulus modulations).

### 5.2. Estimation of the channels: MA/AR structures (stationary data).

Again in this section, the data are stationary and follow the model (2.18), so that  $\mathbf{y}(n)$  appears to be a moving average model driven by non-Gaussian i.i.d. sequences. A number of blind identification methods of MA models using higher statistics have been derived, and could be used in the present context (see e.g. [52]). However, the corresponding algorithms show poor performance.

If  $\mathbf{E}(z)$  is irreducible - see (2.21) - there exists a left polynomial inverse of  $\mathbf{E}(z)$  [1] - say  $\mathbf{G}(z)$ . This implies that  $[\mathbf{G}(z)]\mathbf{y}(n) = \mathbf{d}(n)$  where  $\mathbf{d}(n)$  is i.i.d., which says that  $\mathbf{y}(n)$  is an AR model. This can be used in order to identify the matrix  $\mathbf{E}(z)$  thanks to a linear prediction approach, and hence to retrieve the symbol sequences [25] up to a constant  $K \times K$  matrix. We note however that the irreducibility of  $\mathbf{E}(z)$  does not hold when the excess band-width are all zero due to the factorization evoked at the end of section 2.3.2. This tends to indicate a certain lack of robustness of the linear prediction method when the excess bandwidth factors are small.

### 5.3. Estimation of the channels: subspace methods (cyclo-stationary data).

In the previous sections, the main assumption is that the data are stationary. Here, we rather consider the general model (2.17). We consider here the general cyclo-stationary model.

In order to achieve BSS, it is suggested here to identify the transfer function  $\mathbf{H}(z)$ , and then to evaluate one of its left inverse in order to retrieve the source signals (this kind of approaches is sometimes called *indirect*). In this respect,  $\mathbf{H}(z)$  is usually modelled as a FIR causal filter, i.e.  $\mathbf{H}(z) = \sum_{\ell=0}^L \mathbf{H}_\ell z^{-\ell}$ . It has been

shown in [1] that if  $M > K$  and  $\mathcal{T}_P(\mathbf{H})$  is full column rank, then  $\mathbf{H}(z)$  can be identified from the column space of  $\mathcal{T}_P(\mathbf{H})$  up to a constant  $K \times K$  matrix which can itself be estimated using any instantaneous mixture blind source separation method. In practice, the column space of  $\mathcal{T}_P(\mathbf{H})$  is usually estimated by means of the eigenvalue/ eigenvector decomposition of an estimate of the covariance matrix of vector  $\mathbf{y}_P(n)$  given in (5.1). We should perhaps specify this point. Indeed, we have

$$\begin{aligned}\hat{R}_{\mathbf{y}_P}(0) &= \frac{1}{N} \sum_{n=L+P}^N \mathbf{y}_P(n) \mathbf{y}_P(n)^H \\ &= \mathcal{T}_P(\mathbf{H}) \hat{\Gamma} \mathcal{T}_P(\mathbf{H})^H\end{aligned}$$

where  $\hat{\Gamma} = \frac{1}{N} \sum_{n=L+P}^N \mathbf{s}_{P+L-1}(n) \mathbf{s}_{P+L-1}(n)$ . If this latter is full rank, the column space of  $\hat{R}_{\mathbf{y}_P}(0)$  coincides with the column space of  $\mathcal{T}_P(\mathbf{H})$ . However, it should be emphasized that it is not always legitimate to assume that  $\hat{\Gamma}$  is full rank except when all the sources occupy all the band of frequencies  $\left[-\frac{1}{2T_e}, \frac{1}{2T_e}\right]$ . On the contrary, the rank of  $\hat{\Gamma}$  is expected to fall. We do not provide the details: the reader should compute the limit of  $\hat{\Gamma}$  as  $N \rightarrow \infty$  and show that the limit  $\Gamma$  is a block Toeplitz matrix whose block  $(p, q)$  has the expression:

$$\Gamma_{p,q} = \int_0^1 e^{-i2\pi(p-q)\nu} S_s^{(0)}(e^{i2\pi\nu}) d\nu$$

where  $S_s^{(0)}(e^{i2\pi\nu})$  is the (diagonal) power spectral density of the vector  $\mathbf{s}(n)$ . As the sampling period verifies the Shannon sampling condition, some of the entries of  $S_s^{(0)}(e^{i2\pi\nu})$  are band-limited, which prevents the rank of  $\Gamma$  from being full (see the works of Slepian and Pollak on the prolate spheroidal wave functions).

Further refinements are proposed in [24]. Notice that this subspace method, although apparently quite appealing, performs poorly as soon as the matrix  $\mathcal{T}_P(\mathbf{H})$  is ill-conditioned, which, in practice, is quite often the case.

#### 5.4. Global BSS approaches.

5.4.1. *Temporal approaches: extensions of the Comon contrast function (stationary data).* We now focus on the direct and global BSS methods, i.e. methods that allow one to compute a global separator in one shot (up to indeterminacies that will be specified). In this respect, we intend to provide extensions of the approaches given for the instantaneous context - see Section 3.5. For simplification, assume the (restrictive) stationary case. As specified in the previous paragraph, the model of the data is not stricto sensu  $[\mathbf{H}(z)] \mathbf{s}(n)$  as in Equation (2.17) since the Shannon sampling condition does not hold, but rather (2.18).

On the one hand, a direct extension of a contrast as defined by Comon can be done (see section 8.4.2 of [15]). The first step to be considered is the decorrelation of the data (the prewhitening), i.e. a filter matrix  $\mathbf{Q}(z)$  is computed such that  $\mathbf{x}(n) = [\mathbf{Q}(z)] \mathbf{y}(n)$  is decorrelated both spatially and temporally. This is equivalent to computing  $\mathbf{Q}(z)$  such that  $\mathbf{U}(z) = \mathbf{Q}(z) \mathbf{E}(z)$  is para-unitary, i.e.  $\mathbf{U}(z)$  verifies for any frequency  $\nu \in \left[-\frac{1}{2}, \frac{1}{2}\right]$ :  $\mathbf{U}(e^{i2\pi\nu}) \mathbf{U}(e^{i2\pi\nu})^H = \mathbf{I}_K$ . The reader may find details of this procedure in [8, 5]. The reconstructed vector of the sources may

hence be searched as  $\mathbf{r}(n) = [\mathbf{V}(z)] \mathbf{x}(n)$  where  $\mathbf{V}(z)$  is a para-unitary matrix. It is possible to consider, as in the instantaneous case, the function

$$\Psi(\mathbf{r}) = \sum_{k=1}^N \varphi(\kappa(r_k))$$

where  $\varphi$  shows the same properties as in Proposition 12. It can be simply shown that  $\Psi$ , as a function of the para-unitary matrix  $\mathbf{V}(z)$ , is a contrast, i.e. achieves its maximum for separating matrices  $\mathbf{U}(z)$ : see [13]. Theoretically appealing (at least in a stationary environment), this solution calls for implicit prerequisites, namely the pre-whitening and the optimization over the set of para-unitary matrices. Concerning this latter point, solutions have been given (see [60][40]); Comon et al. also suggested to consider a subset of the set of the para-unitary matrices [14]. The solutions are not simple. Besides, the algorithms might be trapped in local maxima.

**5.4.2. Frequency-domain approaches.** In the case of stationary data (same restrictive context as in the above subsection), it is possible to recast the results of the instantaneous case after processing the discrete Fourier transform of the data: the separation is processed at each frequency. The difficulty is that the indeterminacies change from a frequency to the other which makes the approach quite difficult. The reader may find references in [64].

Assuming now the general cyclo-stationary model given by (2.17), one may focus on second-order methods. The power spectrum of the data, defined as the discrete Fourier transform of the correlation function at the null cyclic frequency  $(R_{\mathbf{y}}^{(0)}(\ell))_{\ell \in \mathbb{Z}}$  can be written as

$$(5.3) \quad S_{\mathbf{y}}^{(0)}(e^{i2\pi\nu}) = \mathbf{H}(e^{i2\pi\nu}) S_{\mathbf{s}}^{(0)}(e^{i2\pi\nu}) \mathbf{H}(e^{i2\pi\nu})^H$$

where  $S_{\mathbf{s}}^{(0)}(e^{i2\pi\nu})$  is the  $K \times K$  power spectrum of the source vector given by (2.10). The components of  $\mathbf{s}$  being jointly independent, hence decorrelated, the matrices  $S_{\mathbf{s}}^{(0)}(e^{i2\pi\nu})$  are diagonal. As far as the identifiability of the unknown  $\mathbf{H}(z)$  based on the relation (5.3) is addressed, it is shown in [27] that the conditions required are that (2.20) holds on the one hand and, on the other hand, that spectral diversity occurs namely that the entries of  $S_{\mathbf{s}}^{(0)}(e^{i2\pi\nu})$  are all distinct for every frequency. In [33, 34] it is shown that the matrices  $\mathbf{G}(e^{i2\pi\nu})$  such that  $\mathbf{G}(e^{i2\pi\nu})^H S_{\mathbf{y}}^{(0)}(e^{i2\pi\nu}) \mathbf{G}(e^{i2\pi\nu})$  is diagonal for all the frequencies  $\nu$  are separating matrices. Criteria measuring the closeness to diagonal matrices have been proposed. See also [47] and the references therein. Again, the diversity of the spectra is not always pertinent for digital communication contexts.

**5.5. Iterative BSS (stationary case).** Instead of considering a simple spatial filtering of the data as indicated in Equation (3.3) we rather process a spatio-temporal filtering as depicted below:

$$(5.4) \quad r(n) = [\mathbf{g}(z)^T] \mathbf{y}(n).$$

The «reconstructed source»  $r(n)$  can be expanded as

$$(5.5) \quad r(n) = \sum_{k=1}^K \underbrace{[f_k(z)] d_k(n)}_{r_k(n)}$$

where the  $f_k(z)$  are components of the global filter

$$(5.6) \quad \mathbf{f}(z) = \mathbf{g}(z)^T \mathbf{E}(z).$$

A key trick for the following results is the normalisation step: we might write  $r_k(n) = \|f_k\| \tilde{r}_k$  where  $E[|\tilde{r}_k|^2] = 1$  and

$$(5.7) \quad \|f_k\|^2 = \int_0^1 |f_k(e^{i2\pi\nu})|^2 d\nu.$$

Notice that the separation is achieved if and only if the real-valued vector  $(\|f_1\|, \dots, \|f_K\|)^T$  is separating in the sense given in Section 3.4. As a consequence, when the separation is achieved, the «reconstructed source» is one of the sources up to a filter with unit norm.

With no modification as compared to the method given for the instantaneous case in 3.4, we consider the optimization of the function  $\Upsilon(r)$  as given in Equation (3.6) - or equivalently (3.7). For sake of simplicity, we keep the notation  $\Upsilon_{CM}^{(c)}(r)$  and  $\Upsilon^{(c)}(r)$  even if it should be understood that the functions in question **depend on the filters**  $f_1(z), \dots, f_K(z)$ . No extra computation is needed as compared to the instantaneous case: indeed, it suffices to substitute the «new source»  $\tilde{r}_k(n)$  to the actual source  $s_k(n)$ : one arrives at the expression:

$$\Upsilon(r) = \frac{\sum_{k=1}^K \kappa(\tilde{r}_k) \|f_k\|^4}{\left(\sum_{k=1}^K \|f_k\|^2\right)^2}.$$

This time, the cumulants  $\kappa(\tilde{r}_k)$  are not constant but depend on the norm-one filter  $\tilde{f}_k(z)$ . Anyway,  $\tilde{r}_k(n)$  is a linear process generated by the symbol sequence  $d_k(n)$ : indeed,  $\tilde{r}_k(n) = [\tilde{f}_k(z)] s_k(n)$  and  $s_k(n)$  has the form given by Equation (2.7). As such,  $\kappa(\tilde{r}_k) < 0$  since by assumption  $\kappa(d_k) < 0$ . We let  $\kappa_{\min,k}$  be:

$$\kappa_{\min,k} = \inf_{\|\tilde{f}_k\|=1} \kappa([\tilde{f}_k(z)] s_k(n))$$

and  $\kappa_{\min} = \min_k \kappa_{\min,k}$ . We obviously have:  $\kappa_{\min,k} < 0$  and  $\kappa_{\min} < 0$ . As a consequence, the following inequality holds:

$$\Upsilon(r) \geq \kappa_{\min} \frac{\sum_{k=1}^K \|f_k\|^4}{\left(\sum_{k=1}^K \|f_k\|^2\right)^2} \geq \kappa_{\min}.$$

Moreover, the equality holds if and only if

- (1)  $\|f_k\| = \|f_{k_0}\| \delta(k - k_0)$
- (2)  $\tilde{f}_{k_0}(z)$  is such that  $\kappa_{\min} = \kappa_{\min,k_0} = \kappa(\tilde{r}_{k_0}(n))$  where  $\tilde{r}_{k_0}(n) = [\tilde{f}_{k_0}(z)] s_{k_0}(n)$ .

We suggest a qualitative interesting remark as far as the reconstructed signal  $r(n) = \|f_{k_0}\| \tilde{r}_{k_0}(n)$ . Indeed, this filter minimizes the Kurtosis of  $\tilde{r}_{k_0}(n)$ . As  $\tilde{r}_{k_0}(n)$  is a filtered version of the non-Gaussian i.i.d. sequence  $(d_{k_0}(n))_{n \in \mathbb{Z}}$ , it is shown in [49] that its minimum is reached when  $\tilde{r}_{k_0}(n)$  is  $e^{i\theta} d_{k_0}(n - n_0)$  where  $\theta$  is an unknown phase and  $n_0$  an uncontrolled delay. The reader might show that such a result also holds for the function  $\Upsilon_{CM}(r) = E\left[\left(|r(n)|^2 - 1\right)^2\right]$ .

Ultimately, any local minimum of  $\Upsilon(r)$  or  $\Upsilon_{CM}(r)$  can be shown to be separating [37, 58].

**5.6. Iterative BSS (general cyclo-stationary).** Similarly to the instantaneous case (Section 3.7), we recast the approach of the previous section for cyclo-stationary data. The received data is  $\mathbf{y}(n) = [\mathbf{H}(z)] \mathbf{s}(n)$ , the reconstructed source  $r(n)$  is still given by (5.4). We may expand this signal as

$$(5.8) \quad r(n) = \sum_{k=1}^K \underbrace{[f_k(z)] s_k(n)}_{r_k(n)}.$$

This time the global filter  $\mathbf{f}(z)$  is

$$\mathbf{f}(z) = \mathbf{g}^T(z) \mathbf{H}(z).$$

As far as the normalization of  $r(n)$  is concerned, we still have  $r(n) = \sum_k \|f_k\| \tilde{r}_k(n)$  but  $\|f_k\|$  does not have the expression (5.7) since  $s_k(n)$  is not i.i.d. with unit power but has a non-constant power spectral density  $S_k^{(0)}(e^{i2\pi\nu})$ :

$$(5.9) \quad \|f_k\|^2 = \int_0^1 |f_k(e^{i2\pi\nu})|^2 S_k^{(0)}(e^{i2\pi\nu}) d\nu.$$

As a consequence, we consider the minimization of the functions  $\Upsilon_{CM}^{(c)}(r)$  and  $\Upsilon^{(c)}(r)$  whose definition is given by Equations (3.22) and 3.23. Similarly to the instantaneous case, the minimization of  $\Upsilon_{CM}^{(c)}(r)$  and  $\Upsilon^{(c)}(r)$  are equivalent problems in the sense of Proposition 4. Here, the functions in question depend not only on the positive coefficients  $\|f_1\|, \dots, \|f_K\|$  but also on the norm-one filters  $\tilde{f}_1(z), \dots, \tilde{f}_K(z)$ . We recall that  $\tilde{r}_k(n) = [\tilde{f}_k(z)] s_k(n)$  where  $\|\tilde{f}_k\| = 1$ ; then, with practically no effort, we deduce from Section 3.7 that

$$(5.10) \quad \Upsilon^{(c)}(r) = \frac{\sum_{k=1}^K \zeta(\tilde{r}_k) \|f_k\|^4 + \sum_{k_1 \neq k_2} \|f_{k_1}\|^2 \|f_{k_2}\|^2 \varepsilon(\tilde{r}_{k_1}, \tilde{r}_{k_2})}{\left(\sum_{k=1}^K \|f_k\|^2\right)^2}.$$

where

$$(5.11) \quad \varepsilon(\tilde{r}_{k_1}, \tilde{r}_{k_2}) = 4\text{Re} \left( \sum_{\alpha \in \mathcal{I}_+^*} R_{\tilde{r}_{k_1}}^{(\alpha)} R_{\tilde{r}_{k_2}}^{(\alpha)*} \right),$$

$$(5.12) \quad \zeta(\tilde{r}_k) = \eta(\tilde{r}_k) + 4 \left| R_{\tilde{r}_k}^{(\alpha)} \right|^2.$$

$$\eta(\tilde{r}_k) = \langle \kappa(\tilde{r}_k(n)) \rangle$$

It is to be noticed that  $\zeta(\tilde{r}_k)$  can also be expressed as

$$(5.13) \quad \zeta(\tilde{r}_k) = \left\langle E \left[ |\tilde{r}_k(n)|^4 \right] \right\rangle_n - 2.$$

Similarly to the instantaneous case, where the filters  $\tilde{f}_k(z)$  are reduced to being 1, the minimization of  $\Upsilon^{(c)}(r)$  is considerably easier when the cross-terms in (5.10) vanish, i.e. when for all the couples  $(k_1, k_2)$  with  $k_1 \neq k_2$  we have  $\varepsilon(\tilde{r}_{k_1}, \tilde{r}_{k_2}) = 0$ . Indeed, in this case,

$$\Upsilon^{(c)}(r) = \frac{\sum_{k=1}^K \zeta(\tilde{r}_k) \|f_k\|^4}{\left(\sum_{k=1}^K \|f_k\|^2\right)^2}$$

which resembles the expression (3.8) except that here, the numbers  $\zeta(\tilde{r}_k)$  depend on the parameters the minimization is run over. In the following we hence restrict the further analysis to this case. We recall that  $\varepsilon(\tilde{r}_{k_1}, \tilde{r}_{k_2}) = 0$  as soon as the cyclic frequencies  $\alpha_k = \pm \frac{T_e}{T_k}$  are all different modulo one. This condition is fulfilled when all the baudrates are different and the sampling frequency is high enough, i.e.  $\frac{1}{T_e} > \max_k \frac{2}{T_k}$ .

We define  $\zeta_{\min,k} = \inf_{\|\tilde{f}_k\|=1} \zeta([\tilde{f}_k(z)] s_k(n))$ .

**Proposition 18.** *If the cyclic frequencies  $\pm \frac{T_e}{T_k}$  are all different modulo 1 then the cost function  $\Upsilon^{(c)}(r)$  achieves its minimum at a separating vector  $\mathbf{f}_*(z)$  if and only if at least one of the  $\zeta_{\min,k}$  is strictly negative. Moreover, the  $\mathbf{f}_*$  in question has its single non-null element located at an index  $k_0$  such that  $\zeta_{\min,k_0} = \min_{k=1,\dots,K} \zeta_{\min,k}$ .*

Proof: if  $\zeta_{\min,k} < 0$  for a certain index  $k$ , then  $\zeta_{\min} < 0$  where we have defined  $\zeta_{\min} = \min_{\ell} \zeta_{\min,\ell}$ . Evidently,  $\sum_{\ell=1}^K \zeta(\tilde{r}_{\ell}) \|f_{\ell}\|^4 \geq \sum_{\ell=1}^K \zeta_{\min,\ell} \|f_{\ell}\|^4 \geq \zeta_{\min} \sum_{\ell=1}^K \|f_{\ell}\|^4$ . This shows that  $\Upsilon^{(c)}(r) \geq \zeta_{\min}$  and this lower bound is attained for any vector  $\mathbf{f}_*(z)$  whose components are all zero except at an index  $k_0$  such that  $\zeta_{\min,k_0} = \zeta_{\min}$  and  $\zeta([\tilde{f}_{k_0}(z)] s_{k_0}(n)) = \zeta_{\min,k_0}$ . Conversely, if  $\zeta_{\min,k} \geq 0$  whatever  $k$ , the lower bound is attained for non separating filters (see the proof of Proposition 3).  $\square$

Contrary to the stationary case, it is not possible to say much about the minimizing filter  $\tilde{f}_{k_0}(z)$  and hence about the reconstructed signal  $r(n)$ . However, the residual filter  $\tilde{f}_{k_0}(z)$  minimizes the  $\zeta(\tilde{r}_{k_0})$  hence tends to make the modulus of  $\tilde{r}_k(n)$  the most constant as possible.

At this point, we have to analyse the condition:  $\zeta_{\min,k} < 0$ . Recall that this condition is the adaptation to the convolutive case of the assumption  $\mathcal{A}_1$ . This latter was proven to hold true by means of semi-analytical considerations. Curiously, it is possible to prove that the condition  $\zeta_{\min,k} < 0$  holds; no numerical computation is needed. We have:

**Proposition 19.** *As the Kurtosis of the symbols  $\kappa(d_k)$  are all strictly negative, we have*

$$\zeta_{\min,k} = \inf_{\|\tilde{f}_k\|=1} \zeta([\tilde{f}_k(z)] s_k(n)) < 0.$$

Proof: more solid arguments than the ones we give here are provided in [31]. The band of frequencies of the sampled version  $s_k(n)$  of the source number  $k$  is the interval  $[-\frac{1+\gamma_k}{2}\alpha_k, \frac{1+\gamma_k}{2}\alpha_k] \subset [-\frac{1}{2}, \frac{1}{2}]$ . The infimum to be computed is hence over the set of unit-norm filters  $\tilde{f}_k(z)$  belonging to  $\mathcal{F}(\gamma_k)$  where  $\mathcal{F}(\gamma_k)$  is the set of digital filters whose transfer function is limited to the band  $[-\frac{1+\gamma_k}{2}\alpha_k, \frac{1+\gamma_k}{2}\alpha_k]$ . Naturally,  $\gamma'_k < \gamma_k$  implies that  $\mathcal{F}(\gamma'_k) \subset \mathcal{F}(\gamma_k)$ . Taking  $\gamma'_k = 0$ , we deduce the following inequality:

$$\zeta_{\min,k} = \inf_{\|\tilde{f}_k\|=1, \tilde{f}_k(z) \in \mathcal{F}(\gamma_k)} \zeta([\tilde{f}_k(z)] s_k(n)) \leq \inf_{\|\tilde{f}_k\|=1, \tilde{f}_k(z) \in \mathcal{F}(0)} \zeta([\tilde{f}_k(z)] s_k(n)).$$

On the other hand, we recall that  $\zeta([\tilde{f}_k(z)] s_k(n)) = \zeta(\tilde{r}_k) = \langle \kappa(\tilde{r}_k(n)) \rangle + 4 \left| R_{\tilde{r}_k}^{(\alpha_k)} \right|^2$ . Thanks to Section 2.2, we know that  $R_{\tilde{r}_k}^{(\alpha_k)} = 0$  when the excess bandwidth factor

is zero. Hence, for any unit norm filter  $\tilde{f}_k(z) \in \mathcal{F}(0)$ ,  $\zeta(\tilde{r}_k) = \langle \kappa(\tilde{r}_k(n)) \rangle$ . We deduce that, for such a filter,  $\zeta(\tilde{r}_k) < 0$ .  $\square$

When all the baud-rates are equal, the minimization of  $\Upsilon_{CM}^{(c)}(r)$  is much more difficult and sufficient conditions on the sources have been set forth that assure that the minimizers of  $\Upsilon_{CM}^{(c)}(r)$  are separating. We refer to [31].

However, the general case remains to be addressed (i.e. for any distribution of the baud-rates): we conjecture that the global minimum is separating and even that any local minimum is also separating. These conjectures come from intensive simulation experiments.

## 6. SIMULATION

This section does not aim at making a benchmark of all the previous methods. It rather intends to show the pertinence of BSS in digital communication contexts.

We first present the environment. We have considered a mixture of  $K = 3$  sources; the modulations are QPSK (two sources) and 16-QAM (one source). The symbol periods are all equal to  $T$  and  $1/T = 277\text{kHz}$  (which is the rate of GSM); the carrier frequency is 1GHz. This later is assumed known to the receiver so that there are no frequency offset in the source vectors. The excess bandwidth factors all equal  $\gamma = 0.5$ . As far as the antenna array is concerned, we have simulated a circular array of  $M = 5$  sensorsdistanced from one another by half a wavelength, i.e. 30cm. The sampling period  $T_e$  is fixed to  $T_e = T/1.6$  so that the Shannon sampling condition is fulfilled.

The propagation channels are multi path and affected by a Rayleigh fading. An arbitrary path - say number  $\ell$  is characterized by its delay (propagation between the source and a sensor of reference), its elevation, azimuth and attenuation. We consider the ETSI channels BUx, TUx, HTx, RAx. For each experiment, the arrival angles of a path are randomly chosen in  $[-\pi/2, \pi/2]$  for the elevation and  $[-\pi, \pi]$  for the azimuth. The different complex amplitudes on each path are also randomly chosen for each experiment.

The received signal is corrupted by a white, additive complex gaussian noise with power spectral density  $N_0$ . The received signals are low-pass filtered in the band  $[-\frac{1}{2T_e}, \frac{1}{2T_e}]$ . The three sources have the same energy per symbol  $\mathcal{E}_s$ . This latter is chosen so that the  $\mathcal{E}_s/N_0$  of the 16-QAM source would be associated with a bit error rate of  $10^{-3}$  if the channel were a single-path channel and if there were only one antenna.

We have considered three algorithms for the separation. In order to have a reference, we have computed the Wiener solution (the MMSE separator): this is naturally a non-blind algorithm. In this respect, we have supposed that the receiver has access to the transmitted symbols. The two BSS algorithms we have considered are the iterative CMA (see section 5.6) and the global method JADE. Once the separation is (supposedly) achieved, we need to equalize each source in order to compute the bit error rate. This extra-step is done after re-sampling the three «source estimates» at the true sampling rate (supposed at this level to be known). The re-sampled signals are expected to be filtered versions of the symbols. A blind equalization algorithm is run (the CMA) in order to compute estimates of the transmitted symbols. The scaling ambiguity is removed non-blindly.

The performance index for the tested algorithms and for a given source is the following: for each experiment, we inspect if the bit error rate (for the transmitted

TABLE 1. Some simulation results

	2000			1000			500		
<b>channel BUx</b>	QAM-4	QAM-16	QAM-4	QAM-4	QAM-16	QAM-4	QAM-4	QAM-16	QAM-4
MMSE	100%	68,8%	100%	100%	55,9%	100%	100%	23%	100%
CMA	100%	70,5%	100%	100%	20,6%	99,8%	95,9%	0,1 %	97,6%
JADE	58,4	1,3%	60,7%	55,5%	0,7%	57,7%	48,8%	0,1%	51,7%
<b>channel TUX</b>	QAM-4	QAM-16	QAM-4	QAM-4	QAM-16	QAM-4	QAM-4	QAM-16	QAM-4
MMSE	100%	80,2%	100 %	100%	69,8%	100%	100%	35,3%	100%
CMA	99,7%	84,3%	99,3%	99,3%	47,5%	99,2%	98,5%	0,7%	98,9%
JADE	94,9%	29,9%	95,1%	93,9%	18,1%	94,4%	93,1%	3%	92,9%
<b>channel RAX</b>	QAM-4	QAM-16	QAM-4	QAM-4	QAM-16	QAM-4	QAM-4	QAM-16	QAM-4
MMSE	99,8%	91,9%	99,6%	99,8%	88,8%	99,6%	99,7%	70,4%	99,6%
CMA	99,1%	84,7%	99,2%	99%	50,2%	99,3%	98,1%	0,6%	98,6%
JADE	99,7%	89,2%	99,6%	99,6%	79,7%	99,6%	99,6%	24,5%	99,5%
<b>channel HTx</b>	QAM-4	QAM-16	QAM-4	QAM-4	QAM-16	QAM-4	QAM-4	QAM-16	QAM-4
MMSE	99,4%	36,2%	99,6%	99,4%	26,8%	99,7%	99,4%	9%	99,5%
CMA	92%	13%	93,2%	90,8%	1,8%	91,6%	83,5%	0%	82,1%
JADE	86,7%	13,1%	87,4%	85,7%	6,3%	86,3%	83,6%	1,1%	84,1%

symbols) is less than  $10^{-2}$ . Averaging the process on 1000 independent trials allows us to provide a percentage of success.

Three observation durations were considered:  $500T$ ,  $1000T$  and  $2000T$ .

The performance of the algorithms can be seen in Table 1. The performance of the MMSE (non-blind) provides, in some sense, an ultimate bound. In this respect, one may notice that the channels BUx and HTx are by far the “most difficult” channels since they are associated with the weakest performance of the MMSE. For these two channels, the multipath effects are severe and this explains why the JADE algorithm, designed for instantaneous mixtures, performs poorly. On the contrary, the CMA shows good performance as far as the extraction of the two modulus-one sources is concerned, even if the observation duration is small (500 symbols): the performance index is above 80%. However, the 16-QAM source is associated with a miserable performance: the CMA as a BSS algorithm is not to be incriminated, since the two other sources are correctly equalized, hence the 16-QAM is itself correctly separated. Hence the bad performance is due to the equalization algorithm (again the CMA run this time on the re-sampled extracted signal): this is a well-known fact that the CMA equalizes the non-modulus one modulation with difficulties.

As far as the TUX and RAX channels are concerned, they are associated with less severe multi-path effects: this explains why the JADE algorithm performs well - even better than the CMA for the RAX channel.

## 7. EXTENSIONS AND FURTHER READINGS

**7.1. Case of non-circular sources .** Along this paper, the data we considered as circular. This assumption is not crucial for second-order or algebraic methods. However, the presence of a non-circular source in the mixture considerably affects the higher-than-second-order methods. For instance, the fourth-order cumulant

$\kappa(r(n))$  if  $r(n)$  is the output of a separator, does not have the same expression as when all the sources are circular. It can be even been shown that the separation is not always achieved when two sources are non-circular.

The interested reader might find results and references in the following works: [18, 20, 42][66].

**7.2. Exploiting the non-stationarity.** Cyclo-stationarity is a main statistical feature of the mixture that has long been thought of as a benefit for source separation. For instantaneous BSS using second-order moments, see [57, 51]: the idea is that the mixing matrix is constant during the observation, while the second-order statistics of the sources vary. An idea is hence to cut the observation interval in subintervals. Recalling the SOBI approach (see section 3.3), the receiver may compute the correlation matrices for the  $u$ -th interval:  $R_{\mathbf{y}}^{(u)}(\ell) = \mathbf{H}R_{\mathbf{s}}^{(u)}(\ell)\mathbf{H}^H$ . As the sources are non-stationary, the diagonal matrices  $R_{\mathbf{s}}^{(u)}(\ell)$  vary with  $u$  hence the pencil of matrices to be jointly diagonalized has more elements and the conditions of identifiability are weaker, hence the algorithm is more robust. We refer to the work of Pham [43] for the Maximum Likelihood approach. The reader might be interested by a work of Wang et al. [63].

In the case of digital communication signals, the cyclo-stationarity is not strong enough in order to consider such approaches: we have pointed out that the power of a source  $E[s_a(t + \tau)s_a(t)^*]$  has very small variations since the cyclo-spectra at the cyclic-frequencies  $\pm\frac{1}{T}$  are numerically small due to the spectral limitation of the shaping functions. On the one hand, the strength of cyclo-stationarity is too weak to be exploited. On the other hand, it cannot be neglected in the computations (for instance in the expression of the fourth-order cumulants).

**7.3. Presence of additive noise.** In this paper, we have considered a noise-free model. Many references may be found where the impact of the noise on the BSS methods is analyzed. Among others, we would like to cite the work of Cardoso concerning the performance analysis a class of BSS algorithms who have a teh so-called «invariance» feature [10]. Concerning the CMA when noise is present: the reader may have a look at the work of Fijalkow et al. [19] for the use of the CMA as an equalizer algorithm and the proximity of a solution in the presence of noise to a Minimum Mean Square Error (MMSE) equalizer ; the case of BSS is analyzed in [36][26] where it is shown that thee local minima of the CM cost function are «not far» from MMSE separator. A systematic analysis is provided in Leshem et al. [35] where both the presnece of noise and the effect of a finite number of samples are considered.

## 8. CONCLUSION

In this paper, we have given some methods for achieving the BSS in the context of telecommunication. We have focused our attention on contrast functions (joint or deflation-based approaches), and particularly the contrast functions depending on fourth-order statistics of the data. These approaches fit the blind problems evoked in the introduction (spectrum monitoring) since they are associated with convergent and performant algorithms. When the channels involve multi-path effects and no a priori information on the distribution of the delays is available, the deflation-based algorithms such as the CMA or the minimisation of the normalized Kurtosis are good candidates for the BSS.

## REFERENCES

- [1] K. Abed-Meraim, P. Loubaton, and E. Moulines. A subspace algorithm for certain blind identification problems. *IEEE Trans. on Information Theory*, 43(2):499–511, March 1997.
- [2] K. Abed-Meraim, Y. Xiang, H. Manton, and Y. Hua. Blind source separation using second-order cyclostationary statistics. *IEEE Trans. on Signal Processing*, 49(4):694–701, 2001.
- [3] B. G. Agee. The least-squares cma: a new technique for rapid correction of constant modulus signals. In *Proc. ICASSP*, pages 953–956, Tokyo, 1986.
- [4] A. Belouchrani, K. Abed-Meraim, J-F. Cardoso, and E. Moulines. A blind source separation technique using second-order statistics. *IEEE Trans. on Signal Processing*, 45(2):434–444, february 1997.
- [5] A. Belouchrani and Andrzej Cichocki. Robust whitening procedure in blind separation context. *Electronic Letters*, 36(24):2050–2051, 2000.
- [6] Patrick Billingsley. *Probability and Measure, 3rd Edition*. Wiley-Interscience, 3 edition, April 1995.
- [7] E. Bingham and H. Hyvarinen. A fast fixed-point algorithm for independent component analysis of complex valued signals. *Int. Journ. on Neural Systems*, 10(1):1–8, 2000.
- [8] Peter J. Brockwell and Richard A. Davis. *Time Series: Theory and Methods*. Springer, 2nd edition, 1991.
- [9] A. Bunse-Gerstner, R. Byers, and V. Mehrmann. Numerical methods for simultaneous diagonalization. *SIAM Journal on Matrix Analysis and Applications*, 14(4):927–949, 1993.
- [10] J.-F. Cardoso. On the performance of orthogonal source separation algorithms. In *EUSIPCO*, pages 776–779, Edinburgh, 1994.
- [11] J.F. Cardoso and A. Souloumiac. Blind beamforming for non gaussian signals. *IEE Proc-F*, 140(6):362–370, december 1993.
- [12] P. Comon. Independent component analysis, a new concept ? *Signal Processing, special issue on higher-order statistics*, 36(3):287–314, april 1994.
- [13] P. Comon. Contrasts for multichannel blind deconvolution. *IEEE Signal Processing Letters* 3, (209–211), 1996.
- [14] P. Comon and L. Rota. Blind separation of independant sources from convolutive mixtures. *IEICE Trans. on fundamentals of electronics, communications, and computer sciences*, E86-A(3):542–549, 2003.
- [15] Pierre Comon and Christian Jutten. *Handbook of blind source separation: independent component analysis and applications*. Academic Press, February 2010.
- [16] C. Corduneanu. *Almost Periodic Functions*. American Mathematical Society, 2 sub edition, January 1989.
- [17] N. Delfosse and P. Loubaton. Adaptative blind separation of independant sources: a deflation approach. *Signal Processing*, 45:59–83, 1995.
- [18] A. Ferréol and P. Chevalier. On the behavior of current second and higher order blind source separation methods for cyclostationary sources. *IEEE Trans. on Signal Processing*, 48, pp. 990, April 2002.(6 (Erratum : Vol 50, N°4):1712–1725, june 2002.
- [19] I. Fijalkow, A. Touzni, and J.R. Treichler. Fractionally spaced equalization using CMA: robustness to channel noise and lack of disparity. *IEEE Trans. on Signal Processing*, 46(1):227–231, january 1998.
- [20] E. Florian, A. Chevreuil, and P. Loubaton. Blind source separation of convolutive mixtures of non circular linearly modulated signals with unknown baud rates. *Signal Processing*, 92(715–726), 2012.
- [21] W.A. Gardner. Spectral correlation of modulated signals: Part I - analog modulations. *IEEE Trans. on Communications*, 35:584–594, june 1987.
- [22] D N Godard. Self-recovering equalization and carrier tracking in two-dimensionnal data communication systems. *IEEE Trans. on Communications*, 1980.
- [23] R.P. Gooch and J.P. Lundell. The CM array: An adaptive beamformer for constant modulus signals. In *Proc. ICASSP*, pages 2523–2526, Tokyo (Japan), april 1986.
- [24] A. Gorokhov and P. Loubaton. Subspace-based techniques for blind separation of mixtures with temporally correlated sources. *IEEE Trans. on Circuits and Systems*, 44(9):813–820, September 1997.
- [25] A. Gorokhov and Ph. Loubaton. Blind identification of MIMO-FIR systems : a generalized linear prediction approach. *Signal Processing*, 73:104–124, 1999.

- [26] L. Tong H. Zeng and C. Johnson. An analysis of constant modulus receivers. *IEEE Transactions on Signal Processing*, 47:2990–2999, november 1999.
- [27] Yingbo Hua and Jitendra K. Tugnait. Blind identifiability of FIR-MIMO systems with colored input using second order statistics. 7(12):348–350, December 2000.
- [28] A. Hyvarinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634, May 1999.
- [29] Yujiro Inouye and Ruey-Wen Liu. A system-theoretic foundation for blind equalization of an fir mimo channel systemchannel system. *IEEE Trans. on Circuits and Systems—I*, 49(4):425–436, April 2002.
- [30] P. Jallon and A. Chevreuil. Separation of instantaneous mixtures of cyclostationary sources. *Signal Processing*, 87:2718–2732, 2007.
- [31] Pierre Jallon, Antoine Chevreuil, and Philippe Loubaton. Separation of digital communication mixtures with the cma: Case of unknown symbol rates. *Signal Processing*, pages 2633–2647, 2010.
- [32] C. Jutten and J. Héroult. Blind separation of sources, part I: an adaptative algorithm based on neuromimetic architecture. *Signal Processing*, 24:1–10, 1991.
- [33] Mitsuru Kawamoto and Yujiro Inouye. Blind deconvolution of MIMO-FIR systems with colored inputs using second-order statistics. *IEICE Trans. on Fundamentals of Electronics, Communications, and Computer Sciences*, E86-A(3):597–604, March 2003.
- [34] Mitsuru Kawamoto and Yujiro Inouye. Blind separation of multiple convolved colored signals using second-order statistics. In *Proc. of ICA'03*, Nara, Japan, April 2003.
- [35] A. Leshem and A.-J. van der Veen. Blind source separation: the location of local minima in the case of finitely many samples. *IEEE Transactions on Signal Processing*, 59:4340 – 4353, 2008.
- [36] D. Liu and L. Tong. An analysis of constant modulus algorithm for array signal processing. *Signal Processing*, 73:81–104, 1999.
- [37] P. Loubaton and P. A. Regalia. Blind deconvolution of multivariate signals: A deflation approach. In *Int. Conf. Communications (ICC)*, Geneva (Switzerland), 1993.
- [38] Albert W. Marshall, Ingram Olkin, and Barry Arnold. *Inequalities: Theory of Majorization and Its Applications*. Springer, September 2010.
- [39] L. Mazet and P. Loubaton. Cyclic correlation based symbol rate estimation. In *proc. Asilomar Conference on Signals, Systems, and Computers*, pages 1008–1012, october 1999.
- [40] John G. McWhirter, Paul D. Baxter, Tom Cooper, Soydan Redif, and Joanne Foster. An EVD algorithm for Para-Hermitian polynomial matrices. *IEEE Transactions on Signal Processing*, 55(5):2158–2169, May 2007.
- [41] Antonio Napolitano. Cyclic higher-order statistics: Input/output relations for discrete- and continuous-time mimo linear almost-periodically time-variant systems. *Signal Processing*, 42(2):147–166, 1995.
- [42] M. Novey and T. Adali. On extending the complex fastica algorithm to non-circular sources. *IEEE Transactions on Signal Processing*, 56(5):2148–2153, 2008.
- [43] D-T. Pham and J-F. Cardoso. Blind separation of instantaneous mixtures of nonstationary sources. *IEEE Transactions on Signal Processing*, 49(9):1837–1848, 2001.
- [44] Boaz Porat. *A Course in Digital Signal Processing*. Wiley, 1 edition, October 1996.
- [45] John Proakis. *Digital Communications*. McGraw-Hill Science/Engineering/Math, 4 edition, August 2000.
- [46] P.A. Regalia. On the equivalence between the Godard and Shalvi-Weinstein schemes of blind equalization. *Signal Processing*, 73:185–190, 1999.
- [47] K. Sabri, M. El Badaoui, F. Guillet, A. Adib, and D. Aboutajdine. A frequency domain-based approach for blind MIMO system identification using second-order cyclic statistics. *Signal Process.*, 89(1):77–86, January 2009.
- [48] Y Sato. A method of self-recovering equalization for multilevel amplitude-modulation systems. *IEEE Trans. on Communications*, 23(6):679–682, 1975.
- [49] O. Shalvi and E. Weinstein. New criteria for blind deconvolution of non-minimum phase systems. 36:312–321, March 1990.
- [50] C Simon, Ph Loubaton, and C Jutten. Separation of a class of convolutive mixtures: a contrast function approach. *Signal Processing*, 81:883–887, 2001.
- [51] A. Souloumiac. Blind source detection and separation using second-order nonstationarity. In *ICASSP*, pages 1912–1915, 1995.

- [52] A. Swami, G. B. Giannakis, and S. Shamsunder. Multichannel ARMA processes. *Trans. on Signal Processing*, 42(4):898–913, april 1994.
- [53] S. Talwar, M. Viberg, and A. Paulraj. Blind estimation of multiple co-channel digital signals using an antenna array. *Signal Processing Letters*, 1(2):29–31, february 1994.
- [54] S. Talwar, M. Viberg, and A. Paulraj. Blind separation of synchronous co-channel digital signals using an antenna array – part I: Algorithms. *IEEE Trans. on Signal Processing*, 44(5):1184–1197, may 1996.
- [55] J. R. Treichler and B. G. Agee. A new approach to multipath correction of constant modulus signals. In *IEEE Trans. Acoust., Speech, Signal Processing*, volume 31, pages 459–472, april 1983.
- [56] J.R. Treichler and M.G. Larimore. New processing techniques based on constant modulus adaptive algorithm. *IEEE Trans. On Acoustics, Speech and Signal Processing*, 33(8):420–431, april 1985.
- [57] M. K. Tsatsanis and C. Kweon. Source separation using second-order statistics: Identifiability conditions and algorithms. In *Asilomar Conf. Signals, Syst., Comput.*, pages 1574–1578, 1998.
- [58] J.K. Tugnait. Identification and deconvolution of multi-channel non-gaussian processes using higher-order statistics and inverse filter criteria. *IEEE Trans. on Signal Processing*, 45(3):658–672, march 1997.
- [59] J.K. Tugnait. Adaptive blind separation of convolutive mixtures of independent linear signals. *Signal Processing*, 73:139–152, February 1999.
- [60] P. P. Vaidyanathan. *Multirate Systems And Filter Banks*. Prentice Hall, 1 edition, October 1992.
- [61] A.J. Van Der Veen and A. Paulraj. An analytical constant modulus algorithm. *IEEE Trans. on Signal Processing*, 5(44):1136–1155, 1996.
- [62] A.J. Van Der Veen, S. Talwar, and A. Paulraj. A subspace approach to blind space-time signal processing for wireless communication systems. *IEEE Trans. Signal Processing*, 45(1):173–190, january 1997.
- [63] W. Wang, M. G. Jafari, S. Sanei, and J. A. Chambers. Blind separation of convolutive mixtures of cyclo-stationary signals. *International Journal of Adaptive Control and Signal Processing*, 18(3):279–298, 2004.
- [64] W. Wang, S. Sanei, and J.A. Chambers. Penalty function-based joint diagonalization approach for convolutive blind separation of non-stationary sources. *IEEE Transactions on Signal Processing*, 53(5):1654–1669, 2005.
- [65] A. Yeredor. Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation. *IEEE Transactions on Signal Processing*, 50(7):1545 –1553, July 2002.
- [66] H. Zhang, L. Li, and W. Li. Independent vector analysis for convolutive blind non-circular source separation. *Signal Processing*, 92(9):2275–2283, 2012.