SOURCE SEPARATION: PRINCIPLES, CURRENT ADVANCES AND APPLICATIONS

Christian JUTTEN[1], *Massoud BABAIE-ZADEH*[2]

[1] Laboratoire des Images et des Signaux (LIS), Institut National Polytechnique de Grenoble (INPG), Grenoble, France.
[2] Electrical engineering department, Sharif University of Technology, Tehran, Iran.
`Christian.Jutten@inpg.fr`, `mbzadeh@sharif.edu`

ABSTRACT

This paper is a survey on the source separation problem, and on methods used for solving this problem. In a blind context, *i.e.* without information about the sources but their mutual independence, methods are based on Independent Component Analysis (ICA). On the contrary, using priors on sources, one can developed semi-blind approaches which are very efficient and often much more simpler. Current advances aim to take into account various priors like positivity or sparsity. This paper will finish with sketches of source separation applications which will give practical examples.

1. INTRODUCTION

Source separation consists in retrieving unknown signals, $\mathbf{s} = (s_1(t), \ldots, s_n(t))^T$, which are observed through unknown mixture of them. Denoting the observations $\mathbf{x}(t) = (x_1(t), \ldots, x_p(t))^T$, one can write :

$$\mathbf{x}(t) = \mathcal{A}(\mathbf{s}(t)),\ t =, \ldots, T, \tag{1.1}$$

where $\mathcal{A}(.)$ denotes the unknown mixture, a function from $\mathbb{R}^n$ to $\mathbb{R}^p$. If the number of observations p is greater or equal to the number of sources, n, the main idea for separating the sources is to estimate a transform $\mathcal{B}(.)$ which inverses the mixture $\mathcal{A}(.)$, and, without extra effort, provides estimates of the unknown sources.

Of course, without other assumptions, this problem cannot be solved. Basically, it is necessary to have priors about

- the nature of the mixtures: it is very important to chose a separating transform $\mathcal{B}(.)$ suited to the mixture transform $\mathcal{A}(.)$,
- the sources: sources properties - even weak - are necessary for driving the $\mathcal{B}(.)$ estimation.

This work has been partially funded by French Embassy in Tehran, and by Center for International Research and Collaboration (ISMO) in the framework of the GUNDISHAPUR cooperation program.

Because of the very weak assumptions, the problem is referred as blind source separation (BSS), and the method based on the property of source independence has been called independent component analysis (ICA) [1, 2].

In fact, one often has priors on signals. A natural idea is then to add these priors in the model, for simplifying or improving the separation methods.

This paper is organized as follows. In Section 2, we recall usual assumptions of blind source separation, and principles of ICA. Section 3 is devoted to Gaussian non iid sources. In Section 4, we show that priors like discrete-valued or bounded sources lead to simple geometrical algorithms. In Section 5, we briefly present a few applications, before a short conclusion (Section 6).

2. BLIND SOURCE SEPARATION

Source separation methods have been developed intensively for linear mixtures, instantaneous as well as convolutive, and more recently by a few researchers for nonlinear mixtures. In the most general case, the only assumption done on the sources is that they are statistically independent. From Darmois's result [3], one deduces that this problem has no solution for mutually independent Gaussian sources, with temporally independent and identically distributed (iid) samples. Then, since the Gaussian iid model has no solution, one must add priors, which are threefold [4]:

- Non Gaussian iid,
- Gaussian but non temporally independent (the first *i* of *iid* is relaxed), *i.e.* temporally correlated,
- Gaussian, but non identically distributed (*id* of *iid* is relaxed), *i.e.* non stationary sources.

2.1. Existence of ICA solutions

Initially, even if it was not clearly stated [5], the problem has been related to the non Gaussian iid model, and has been refered to as blind source separation (BSS). The non Gaussian property appears clearly in the Comon's theorem [2] for linear mixtures.

Theorem 2.1 *Let* $\mathbf{x} = \mathbf{A}\mathbf{s}$ *be a* p*-dimension regular mixture of mutually independent random variables, with at most one Gaussian,* $\mathbf{y} = \mathbf{B}\mathbf{x}$ *has mutually independent components iff* $\mathbf{BA} = \mathbf{PD}$*, where* $\mathbf{P}$ *and* $\mathbf{D}$ *are permutation and diagonal matrices, respectively.*

This theorem is only based on the independence of random variables. The independence criterion involves (explicitly or implicitly) higher order (than 2) statistics, but does not take into account the order of samples. It means that the iid assumption is not required, it is just a default assumption: consequently, ICA methods can be applied for iid as well as for not iid sources, but it does not work for Gaussian sources.

Moreover, the theorem points out that the sources cannot be exactly estimated, but only up to a scale and a permutation. These are the typical undeterminacies[1] of source separation in linear mixtures.

2.2. Independence criteria

Assuming that the output signals have probability density functions (pdf), mutual independence of $\mathbf{y}$ means that:

$$p_{\mathbf{y}}(y_1, \dots, y_p) = \prod_{i=1}^{n} p_{y_i}(y_i), \tag{2.1}$$

where $p_{\mathbf{Y}}(y_1, \dots, y_p)$ denotes the joint density of the random vector $\mathbf{Y}$ and $p_{Y_i}(y_i)$ denotes the marginal density of the random variable Y_i. Of course, measuring independence with Eq. (2.2) is not very convenient since it concerns multivariate functions. A classical (in statistics) divergence measure between two distributions, p and q, of the same random variables, $u_1, \dots, u_p$, is the Kullback-Leibler divergence:

$$KL(p||q) = \int p(u_1, \dots, u_p) \frac{p(u_1, \dots, u_p)}{q(u_1, \dots, u_p)} du_1 \dots du_p. \tag{2.2}$$

One can shown that the Kullback-Leibler divergence is positive and equals to zero if and only if $p = q$. Applying this measure the the joint and the marginal density leads to an independence measure:

$$KL(p_{\mathbf{y}}|| \prod_{i=1}^{n} p_{y_i}) = \tag{2.3}$$

$$\int p_{\mathbf{y}}(y_1, \dots, y_p) \frac{p_{\mathbf{y}}(y_1, \dots, y_p)}{\prod_{i=1}^{n} p_{y_i}(y_i)} dy_1 \dots dy_p.$$

[1] it also means that the mixture $\mathcal{A}$ cannot be blindly identified

In that case, the Kullback-Leibler divergence is positive and vanishes if and only if the random vector $\mathbf{Y}$ has statistically independent components. This measure is also related to the mutual information (MI) usual in information theory [6]:

$$KL(p_{\mathbf{y}}|| \prod_{i=1}^{n} p_{y_i}) = I(\mathbf{y}), \tag{2.4}$$

which can be expressed using joint differential and marginal entropies, $H(\mathbf{y}) = -E[\log(p_{\mathbf{y}})]$ and $H(y_i) = -E[\log(p_{y_i})]$ respectively, as:

$$I(\mathbf{y}) = \sum_{i=1}^{p} H(y_i) - H(\mathbf{y}). \tag{2.5}$$

The main drawback of MI is that it requires estimation of joint and marginal densities. However, since $\mathbf{y} = \mathcal{B}(\mathbf{x})$ where $\mathcal{B}$ is supposed invertible, MI can be written as:

$$I(\mathbf{y}) = \sum_{i=1}^{p} H(y_i) - H(\mathbf{x}) + \log |\mathbf{J}_{\mathcal{B}}|, \tag{2.6}$$

where $\mathbf{J}_{\mathcal{B}}$ denotes the Jacobian of the transform $\mathcal{B}$. With this *trick*, since $H(\mathbf{x})$ is a constant with respect of the inverse transform $\mathcal{B}$, one notes that, up to this constant, estimation of $I(\mathbf{y})$ requires only marginal pdf estimations in the terms $H(y_i)$. The direct minimization of the MI, with respect of the parameters of the transform $\mathcal{B}$, equivalent to minimization of $I(\mathbf{y}) - H(\mathbf{x})$ and based on accurate estimations of marginal pdf (for instance, using kernel estimates) has been used by a few authors [2, 7, 8, 9, 10]. This approach may be shown to provide asymptotically a Maximum Likelihood (ML) estimation of the source signals [11]. Moreover, as explained in the next subsections, many simple criteria can be derived from MI, with approximate pdf estimates.

2.3. A few simple criteria derived from MI minimization

2.3.1. Cancelling nonlinear cross-correlations

Indendence can also be expressed as suggested by Papoulis [12], using nonlinear decorrelations [13, 14, 15, 16]. For instance, in [13], source separation is achieved by cancelling the cross-correlations:

$$E[f(y_i)g(y_j)], \forall i \neq j, \tag{2.7}$$

where f and g are different odd functions. In fact, for linear mixtures (*i.e.* if $\mathcal{A}$ and $\mathcal{B}$ are matrices), deriving the MI leads to similar estimation equations:

$$E[\psi_{y_i}(y_i) y_j] = 0, \forall i \neq j, \tag{2.8}$$

where $\psi_{y_i} = -\partial \log p_{y_i}(u)/\partial u$ is the score function. This result gives the optimal nonlinear functions with respect to MI minimization.

2.3.2. *Expansion of pdf*

Simpler estimates of pdf lead to simpler criteria, which, although they only approximate independence, can also lead to source separation. For instance, 4-th order Gram-Charlier or Edgeworth expansions provide criteria involving 4-th order cumulants [17, 18].

2.3.3. *MI and non-Gaussianity*

Finally, for linear mixtures, one can derive other families of algorithms by considering particular factorizations of the inverse transform, which is a matrix $\mathbf{G}$. Due to the scale indeterminacy, one can look for unit variance source and a usual idea is to factorize $\mathbf{G} = \mathbf{U}\mathbf{W}$, where $\mathbf{W}$ is a whitening matrix and $\mathbf{U}$ is an orthogonal matrix. After estimating $\mathbf{W}$ such that $E[(\mathbf{W}\mathbf{x})(\mathbf{W}\mathbf{x})^T] = I$ with second order statistics, one can estimate (with higher order statistics) $\mathbf{U}$ by minimizing MI. Denoting $\mathbf{z} = \mathbf{W}\mathbf{x}$, the MI becomes :

$$I(\mathbf{y}) = \sum_{i=1}^{p} H(y_i) - H(\mathbf{z}) + \log|\det \mathbf{U}|. \qquad (2.9)$$

Since $\mathbf{U}$ is an orthogonal matrix, the last term equals 0, and minimizing MI is equivalent to minimizing the marginal entropy of y_i. For unit variance signals (as the y_i's), the entropy is maximum for Gaussian sources: consequently, minimizing the MI is equivalent to look for sources as non-Gaussian as possible [19, 20].

2.3.4. *Infomax*

Applying nonlinear transform ϕ_j in the output y_j of the separation structure $\mathcal{B}$ so that $z_j = \phi_j(y_j)$ are uniformly distributed[2] in $[0, 1]$, one can write:

$$I(\mathbf{z}) = I(\mathbf{y}) = \sum_j H(z_j) - H(\mathbf{z}) = -H(\mathbf{z}) + \text{cte}, \qquad (2.10)$$

since (1) $\phi_j(y_j)$ are invertible and (2) the entropies of y_j, uniformly distributed in $[0, 1]$, are constant. Then, minimizing $I(\mathbf{y})$ is equivalent to minimize $I(\mathbf{z})$ or to maximize the joint entropy $H(\mathbf{z})$. It is the Infomax principles introduced in [21].

2.4. Contrast functions

The concept of contrast function is another generic approach for designing simple criteria. Inspired by Donoho's work on blind deconvolution [22], constrast functions for source separation have been introduced by Comon [2].

Definition 2.1 *A function $C(\mathbf{y})$ of the random vector $\mathbf{y}$ is a contrast function if it satisfies the following conditions:*

- $C(\mathbf{A}\mathbf{y}) \leq C(\mathbf{y})$,
- $C(\mathbf{A}\mathbf{y}) = C(\mathbf{y})$, *if and only if* $\mathbf{A} = \mathbf{D}\mathbf{P}$ *where* $\mathbf{D}$ *and* $\mathbf{P}$ *are a diagonal matrix and a permutation matrix, respectively.*

As examples, the opposite of MI is a contrast, as well as many criteria derived from MI: $H(\mathbf{z})$, the opposite of non-Gaussianity, *i.e.* $-\sum_j H(y_j)$, the joint entropy $H(\mathbf{z})$ in (2.10) are contrast functions.

2.5. Source separation in other mixtures

More complicated mixing systems have also been studied in the literature.

For example, in (linear) convolutive mixtures, the mixing model is $\mathbf{x}(n) = \mathbf{B}_0\mathbf{x}(n) + \mathbf{B}_1\mathbf{x}(n-1) + \cdots + \mathbf{B}_p\mathbf{x}(n-p) = [\mathbf{B}(z)]\mathbf{x}(n)$, which has been shown [23] to be separable.

Nonlinear mixtures are not in general separable [24]. A practically important case of nonlinear mixtures is Post Nonlinear (PNL) mixtures [25], in which a linear mixture is followed by nonlinear sensors. It has been shown that PNL mixtures are separable using statistical independence, too [25, 26, 27], with the same undeterminacies than linear mixtures.

However, if some weak prior information about the source signals is available, then the performance of the source separation algorithms may be significantly improved. Thus, these methods are not 'Blind' but 'Semi-Blind'. In the next sections of this paper, some of most frequently used priors have been considered.

3. SEPARATION OF NON IID SOURCES

Suppose that we know that the source samples are not iid, *i.e.* that sources are temporally correlated, or non stationary.

3.1. Separation of correlated sources

Several approaches have been proposed for separating correlated sources [28, 29, 30]. Pham and Garat [31] showed that time-correlated Gaussian sources can be separated provided than their spectra are different. In that case, the separation can be achieved by estimating a separation matrix $\mathbf{B}$ which minimizes the criterion

$$C(\mathbf{B}) = \sum_{l=1}^{L} w_l \,\text{off}(\mathbf{B}\hat{\mathbf{R}}(\tau_l)\mathbf{B}^T), \qquad (3.1)$$

where w_l are weighting coefficients, $\text{off}(.)$ is a measure of deviation from diagonality, which is positive and vanishes iff $(.)$ is diagonal and which satisfies:

$$\text{off}(\mathbf{R}) = KL(\mathbf{R} \,||\, \text{diag}\mathbf{R}), \qquad (3.2)$$

[2] $\phi_j(y_j)$ is then the cumulative density function of the random variable y_j. Note also that this function is invertible

where $KL(\mathbf{R}_i \mid\mid \mathbf{R}_j)$ denotes the Kullback-Leibler divergence of two zero mean multivariate normal densities, with variance-covariance matrices $\mathbf{R}_i$ and $\mathbf{R}_j$, and $\mathrm{diag}\mathbf{R}$ is the diagonal matrice composed by diagonal entries of $\mathbf{R}$ and zeros elsewhere.

The criterion (3.1) involves a set of variance-covariance matrices with various delays τ_l : $\hat{\mathbf{R}}(\tau_l) = \hat{E}[\mathbf{y}(t-\tau_l)\mathbf{y}(t)^T]$, where $\hat{E}[.]$ is estimated using an empirical mean. Basically, minimizing this criterion is equivalent to estimate the separation matrix $\mathbf{B}$ which diagonalizes jointly the set of the variance-covariance matrices. This approach has a few advantages:

- it only requires second-order statistics,
- it can then separate Gaussian sources,
- there exist many very fast and efficient algorithms for jointly diagonalizing matrices [32, 33].

3.2. Separation of nonstationary sources

Source nonstationarity has been first used by Matsuoka *et al.* [34]. More recently, Pham et Cardoso developed a rigourous formalization, and proved that nonstationary Gaussian sources can be separated provided than the variance ratios $\sigma_i^2(t)/\sigma_j^2(t)$ are not constant. In that case, the separation can be achieved by estimating a separation matrix $\mathbf{B}$ which minimizes the criterion

$$C(\mathbf{B}) = \sum_{l=1}^{L} w_l \mathrm{off}(\mathbf{B}\hat{\mathbf{R}}_l\mathbf{B}^T), \quad (3.3)$$

where we use the same notations than in the previous subsection. In Eq. (3.3), matrices $\hat{\mathbf{R}}_l$ are variance-covariance matrices estimated by empirical mean on successive sample blocks T_l. Among a few algorithms, the separation matrix $\mathbf{B}$ can be computed as the matrix which jointly diagonalizes the set of the variance-covariance matrices $\mathbf{R}_l$.

The method has the same advantages than the method exploiting the temporal correlation.

Moreover, it can be easily extended to linear convolutive mixtures, considered in the frequency domain. In fact, after Fourier transform, in each frequency band the signal tends to be close to a Gaussian signal, and consequently the method based on non Gaussian iid model are not efficient. Conversely, if the source is non stationary, one can extend the above algorithm in the frequency domain. This idea provides a very efficient method for speech signal [35].

4. GEOMETRICAL METHODS FOR SOURCE SEPARATION

4.1. Bounded sources

Suppose we know that all the sources are bounded. This simple prior leads to simple geometrical interpretations and methods for source separation (firstly introduced in [36]).

Consider, for example, separating two sources from two mixtures. Because of the scale indeterminacy, the mixing matrix may be assumed to be of the form:

$$\mathbf{A} = \begin{pmatrix} 1 & a \\ b & 1 \end{pmatrix} \quad (4.1)$$

where a and b are constants to be estimated from the observed signals. Since the sources are bounded, the Probability Density Function (PDF) of each source has a bounded support, *i.e.* $p_i(s_i)$ (the PDF of the ith source) is non-zero only inside an interval $\alpha_i < s_i < \beta_i$. Then, the joint PDF $p_\mathbf{s}(\mathbf{s}) = p_1(s_1)p_2(s_2)$ is non-zero only in the rectangular region $\{(s_1, s_2) \mid \alpha_1 < s_1 < \beta_1, \alpha_2 < s_2 < \beta_2\}$. Consequently, if we have 'enough samples' $(s_1(n), s_2(n))$ from the sources, they form a rectangular region in the s-plane (see Fig. 1.a). This rectangle will be transformed, by the linear transformation $\mathbf{x} = \mathbf{A}\mathbf{s}$, into a parallelogram and the slopes of the borders of this parallelogram determine a and b (Fig. 1.b).

The above idea may be even generalized for separating PNL mixtures [37]: in a PNL mixture, the parallelogram of Fig. 1.b is again transformed, by 'component-wise' nonlinearities (corresponding to sensor nonlinearities), into a nonlinear region (Fig. 1.c). It has been proved [37] that if this nonlinear region is transformed again into a parallelogram by 'component-wise' nonlinearities, the sensor nonlinearities have been completely compensated. An iterative algorithm is then proposed in [37] for estimating the borders and inverting them.

4.2. Sparse sources

Geometrical ideas are specially useful for separating sparse sources, *i.e.* sources for which the probability of a sample to be large is very close to 0. Consequently, for sparse sources, the joint probability that a sample $(s_1(n), s_2(n))$ is observed at the borders of the rectangular region of Fig. 1.a is very low, and hence we cannot rely on estimating the borders of the parallelogram of Fig. 1.b for source separation. However, for these sources, two 'axes' (parallel to the borders of the parallelogram) are easily visible, and their slopes again determine the mixing matrix (see Fig. 2.a and 2.b, obtained from synthetic sparse signals). Moreover, for sparse sources, two new important advantages may be obtained:

1. Contrary to the traditional geometrical algorithm, it is easy to generalize the above geometric idea to higher

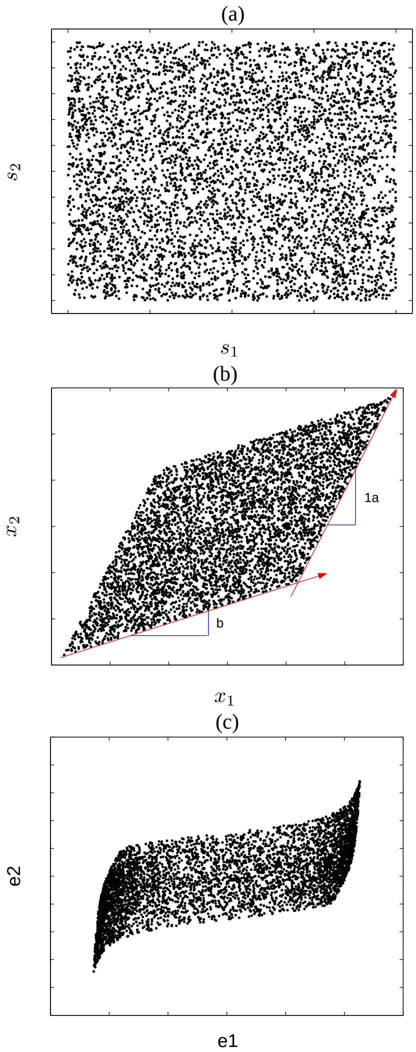

Fig. 1. Distribution of a) source samples, b) observation samples in linear mixtures, observation samples in Post nonlinear (PNL) mixtures.

dimensions (*i.e* separating n sources from n mixtures) [38].

2. Sparsity enables to estimate the mixing matrix (and even recovering the sources) in the underdetermined case, that is, where there is less sensors than sources [39]. Consider, for example the case of 3 sources and 2 sensors (Fig. 2.c). Three 'axes' are visible in this scatter plot, and they correspond to the 3 columns of the mixing matrix. This is because $\mathbf{x} = s_1\mathbf{a}_1 + s_2\mathbf{a}_2 + s_3\mathbf{a}_3$, where $\mathbf{a}_i$'s are the columns of the mixing matrix, and consequently the axes of Fig. 2.c (which correspond to the instances where 2 among the 3 sources are nearly zero) are in the directions of $\mathbf{a}_i$'s. This idea can be directly generalized to more number of sources and sensors.

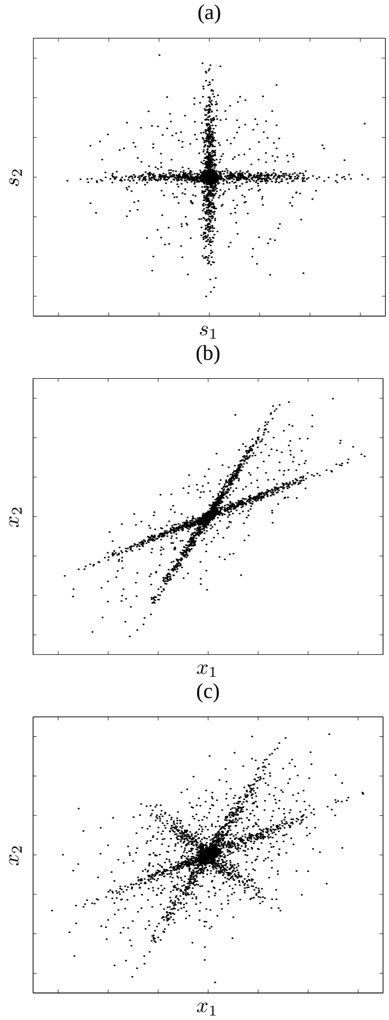

Fig. 2. Distribution of a) two sparse sources, (b) mixture of two sparse sources, (c) mixture of three sparse sources.

In that case, since the mixing model (the matrix $\mathbf{A}$ in linear mixtures) is not invertible, note that source separation requires generally two distinct steps: identification of the mixing matrix $\mathbf{A}$ and source restoration under the constraint of the mixing model. The second step is generally much more complex and basic algorithms are based on linear programming. A main restriction of the above idea for identifying the mixing matrix in underdetermined case, is that it is implicitly assumed that, most of the times, there is just one 'active' (*i.e.* high-energy) source. The expected number of active sources at each instant is nP, where n is the number of sources, and P is the probability of a source being active (small by sparsity assumption). When nP is large (*e.g.* because of a large P) the above idea fails. A solution to this problem has been proposed in [40].

Moreover, from the above geometric ideas, it is visually seen that for separating sparse sources, the independence of source signals is of minor importance. In fact, even this assumption may be dropped, leading to the name Sparse Component Analysis (SCA).

4.3. Discrete-valued sources

Another prior used in some applications, especially in digital communications [41, 42, 43, 44] is that the sources are discrete (*e.g.* binary or k-valued), and the observations are continuous mixtures of them. Since the discrete sources are also bounded, the methods for separating bounded sources may be used for separating these mixtures, too. However, they can be modified to gain more advantages (*e.g.* simplicity, accuracy, or considering noisy mixtures). Moreover, for underdetermined mixtures of discrete sources, it is possible to identify and even recover the sources (much easier than for sparse sources). This can be seen by having in mind a geometrical interpretation like the previous section. Furthermore, for discrete sources, even the independence assumption may be dropped [42].

In [42], a geometrical approach (similar to what is presented in the previous section) is presented for separating discrete (k-valued) sources, in which the independence of the sources is not required. A Maximum Likelihood method for separating these mixtures (which works for underdetermined mixtures, too) has been proposed in [41], in which, it is assumed that the source distribution, too, is known a priori. The case of binary valued sources has been considered in [44] and a method based on creating virtual observations has been proposed. The same authors have proposed a solution based on a polynomial criterion [45] for PSK communication sources. In [43] the underdetermined BSS problem has been considered in a general case, and then a solution has been proposed for the case of discrete sources. An extension to the case of Post-Nonlinear mixtures, where the source alphabet (except its size) is not known a priori, is considered in [46].

4.4. Sparsifying the observations

In the two previous subsections, sparsity is evidently a source property. More generally, and it will be explained in details in [47], we can apply a sparsifying transform $\mathcal{S}$, which preserves the linearity of the mixing model and transforms observations $\mathbf{x}$ in new sparse (or sparser) observations: $\mathcal{S}(\mathbf{x}) = \mathcal{S}(\mathbf{A}\mathbf{s}) = \mathbf{A}\mathcal{S}(\mathbf{s})$, i.e. $\tilde{\mathbf{x}} = \mathbf{A}\tilde{\mathbf{s}}$. Hence, in the transformed domain [3], methods exploiting sparsity can be used for estimating the sources $\tilde{\mathbf{s}}$. Then, applying the inverse of the sparsifying transform, $\mathcal{S}^{-1}$, provides source estimation: $\hat{\mathbf{s}} = \mathcal{S}^{-1}(\tilde{\mathbf{s}})$.

[3] *e.g.* wavelet or time-frequency domains

5. APPLICATIONS

In fact, the main interest of source separation problem is to be relevant in many application domains, providing than we have multi-dimensional observations. In the simplest case, this diversity is spatial and is obtained by using many sensors. Then, for providing efficient solutions, as for any estimation problem, one has to choose carefully the following ingredients:

- the model of mixture, *i.e.* what is the relationship between the observations, $\mathbf{x}$, and the sources, $\mathbf{s}$,
- the criterion: is source independence relevant ? have the sources other properties that could be used: temporal coloration, non-stationarity, sparsity, discree values, etc.?
- the optimization algorithm.

In the following, we only consider the two first ingredients, that we discuss briefly in the framework of a few applications.

5.1. Biomedical applications

In electrocardiogram (ECG), electroencephalogram (EEG) or magnetoencephalogram (MEG) signal processing, one uses a large set of electrodes (from 10 in ECG to more than 100 in EEG and MEG), and the signals received on the electrodes is related to the electric or magnetic fields due to the electrical activity of heart or neurones. The propagation in the biological tissues is very fast and linear instantaneous mixtures are relevant models:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t). \tag{5.1}$$

Independence assumption is generally true, but the nature of signals suggests to use other priors: for instance, ECG are sparse signals, and most of the biological signals are temporally correlated and non-stationary. Results obtained show that source separation is very efficient, for extracting artifacts [48, 49] or sources of interest, like ECG of the foetus [50, 51, 52].

5.2. Communications

In digital mobile communications, received signals are corrupted by multi-path propagation of a unique source, or by sources in a multi-user context. Then, blind equalization or source separation is an essential step in the signal processing. Basically, the mixture model must take into account the propagation, and is a convolutive model:

$$\mathbf{x}(k) = [\mathbf{A}(z)]\mathbf{s}(k), \tag{5.2}$$

where k denotes discrete time and $\mathbf{A}(z)$ denotes the matrix of filters in the z-domain. Signals coming from different users can be assumed to be statistically independent, and the iid assumption is relevant, too. However, it is very efficient to take into account the discrete nature of signals [44, 53, 46], or cyclostationarity [54]: it allows (1) to achieve better performance, even in noisy environments and (2) eventually to separate more sources than sensors.

5.3. Audio and music

Basically, in audio (speech and music) applications, the mixture model is convolutive for taking into account the sound propagation. Generally, methods proposed in the time domain are very intricated [], especially since realistic filters require a large number of taps. The most efficient approach consider the problem in the frequency domain, after applying a short term Fourier transform on the observations $\mathbf{x}(k)$. Hence, the difficult convolutive problem in the time domain is transformed in many simple instantaneous problems, one for each frequency band . The main problem is to cancel the permutation undeterminacy, which exists at each frequency and corrupts the wide band source reconstruction [35].

Source separation has been used for speech enhancement [55, 35] and for music separation, exploiting for instance the music or speech sparsness in the time-frequency domain [56, 57]. The general framework of Bayesian approaches can also be used for music instrument separation even in mono recordings [56]. In this section, we show how visual information can enhanced speech separation.

In the two next subsections, we show tha speech (linear instantaneous or convolutive) mixtures, $\mathbf{x}(t)$ can be completed by the video recording of the speaker (of interest) face, $V(t')$, sampled at 20ms. Moreover, it allows to extract one speech signal[4], the one associated to the visual cue.

5.3.1. Extraction based on audio-video spectrum estimation

The basic idea is to use the simple visual cue, $V(t') = [h(t'), w(t')]^T$ associated to the height, $h(t')$, and the width, $w(t')$, lip opening, for estimating a rough estimation of the speech spectrum of the speaker. Since lip motions are related to sounds but present ambiguities, from a set of audio-visual data, we first estimated (with learning) a probabilistic audio-visual model. Then, by maximizing the audio-video likelihood by the EM algorithm, we can extract the audio source associated to the video. This method have been compared to Jade [32] and is much more efficient. It has mainly two advantages:

- it is very efficient for low SNR,
- it select the source of interest among all the sources.

[4] instead separation of all the sources as usual in source separation

The method can be extended for convolutive mixtures, in the frequency domain. In that case, a similar approach is done in each frequency band. Moreover, the video information is also very efficient for cancelling the permutation indeterminacies. [58, 59]

5.3.2. Extraction based on voice visual activity (VVA) detection

Another idea is to use the video signal for detecting the voice activity. As a simple idea, we claim that, on the frame t', there is voice activity if the lip motion is greater than a threshold, *i.e.* if:

$$\mathrm{vva}(t') = \left|\frac{\partial h(t')}{\partial t'}\right| + \left|\frac{\partial w(t')}{\partial t'}\right|. \qquad (5.3)$$

For avoiding noisy estimations, the actual VVA is decided after smoothing on the T previous frames:

$$\mathrm{VVA}(t') = \sum_{k=0}^{T} a_k \mathrm{vva}(t'-k), \qquad (5.4)$$

where a_k are the coefficients of a truncated first-order IIR low-pass filter. This visual voice activity detector is very efficient for cancelling permutation indeterminacies in frequency domain source separation algorithms for convolutive mixtures [60].

5.4. Sensor arrays

Any set of sensors receiving mixture of signals can be enhanced using source separation methods. This idea have been applied in many domains. For instance, for monitoring dam motion, one can measure the deviation to verticality with plumblines distributed along the wall of the dam. The model of observations assumes that the deviation is a linear mixture of the water level, of the temperature and of other unexpected signals (sismic motions, aging of the dam, etc.), which can be separated using ICA algorithms [61].

Source separation can also be used for enhancing the performance of sensors array. For instance, with Silicon Hall effect sensor array, one can improve the selectivity of the sensor and process simultaneously a few signals [62], even with very close sensors (a few hundredths of micrometers) on integrated circuits.

Source separation can also be applied to chemical ISFET sensor array [63], usefull for environmental applications (water pollution). In that case, the mixture model is much more complex, since the output (drain) current of each

ISFET sensor is a nonlinear mixture of the different chemical species:

$$I_d = A + B \ln \left(a_i + \sum_j k_{ij} a_j^{z_i/z_j} \right), \tag{5.5}$$

where A and B are constant depending of technological and geometrical parameters of the ISFET transistor, k_{ij} is the sensor sensitivity to secondary ions, a_i and z_i are activity and valence of the ion i, respectively. Of course, parameters A, B and k_{ij} vary from a sensor to another one, and source separation methods can exploit this spatial diversity.

The mixture model (5.5) is nothing but a Post-Nonlinear (PNL) mixture, in fact simplified since the nonlinearity is known ($\log$ function with unknwon parameters). With these priors, adaptation of algorithms for PNL mixtures is easy and provides good separation performance [63].

5.5. Sparse decompositions

In many problems, observations are positive mixtures of positive data [64]. It is for instance the case of nuclear magnetic resonance spectroscopy of chemical compounds [65], or of hyperspectral images [66, 67]. Moreover, in these cases, the spectra of the different species are basically non independent It is generally more or less sparse, too. Consequently, using ICA for recovering the spectra generally fails, or provides spectra with spurious peaks. Taking into account the positivity of the mixture matrix entries, improves the solution, but is generally not sufficient (due to the spectrum dependence, ICA can fail). Currently, in such cases, Bayesian methods, able to manage all the priors, especially positivity and sparsity, are the most efficient [68].

Practically, in these examples, it is clear that independence is wrong and ICA will fail. On the contrary, relevant decompositions can be provided using positive and sparsity.

6. CONCLUSION

Now, it must be clear that *blind* source separation does not really exist. First, although this point has not been adressed in this paper, it is important to have priors on the mixture models, and to consider a suitable separation model. Second, priors on sources are essential. From a statistical point of view, since the problem has no solution for Gaussian iid signals, 3 types of statistical priors are possible : sources are non Gaussian iid, sources are Gaussian temporally correlated, sources are Gaussian nonstationary. Remember that, in the 2 former cases, Gaussian means that second order statistics is sufficient, and that it is then possible to consider Gaussian sources, but the methods works for non Gaussian sources too.

Additionaly, other priors can provide original, simple and efficient algorithms. For instance, bounded sources or discrete sources leads to geometrical algorithms. It is also possible to exploit other informations like positivity of sources or to add a visual cue to enhance speech processing.

Two very interesting approaches, which provide a general framework, are the Bayesian ICA which is able to take into account any priors, and the Sparse Component Analysis, which both exploits the data sparsity and looks for sparse representations. These two approaches are explained in details in the survey papers of A. Mohammad-Djafari [69] and Gribonval and Lesage [47].

7. REFERENCES

[1] C. Jutten and J. Hérault. Independent components analysis versus principal components analysis. In *Proc. European Signal Processing Conf. EUSIPCO 88*, pages 643–646, Grenoble, France, September 1988.

[2] P Comon. Independent component analysis, a new concept ? *Signal Processing*, 36(3):287–314, 1994.

[3] G. Darmois. Analyse générale des liaisons stochastiques. *Rev. Inst. Intern. Stat.*, 21:2–8, 1953.

[4] J.-F. Cardoso. The three easy routes to independent component analysis: contrasts and geometry. In *Proceeding of ICA 2001*, San Diego, USA, 2001.

[5] C. Jutten and J. Hérault. Blind separation of sources, Part I: an adaptive algorithm based on a neuromimetic architecture. *Signal Processing*, 24(1):1–10, 1991.

[6] T. Cover and J. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications, 1991.

[7] D. T. Pham. Mutual information approach to blind separation of stationary sources. In *Proceedings of ICA'99*, pages 215–220, Aussois, France, January 1999.

[8] G. A. Darbellay and P. Tichavský. Independent component analysis through direct estimation of the mutual information. In *ICA2000*, pages 69–74, Helsinki, Finland, June 2000.

[9] L. B. Almeida. MISEP – linear and nonlinear ica based on mutual information. *Journal of Machine Learning Research*, 4:1297–1318, December 2003.

[10] M. Babaie-Zadeh and C. Jutten. A general approach for mutual information minimization and its application to blind source separation. *Signal Processing*, 85(5):975–995, May 2005.

[11] J.-F. Cardoso. Blind signal separation: statistical principles. *Proceedings IEEE*, 9:2009–2025, 1998.

[12] A. Papoulis. *Signal Analysis*. McGraw-Hill, 1977.

[13] J. Hérault and C. Jutten. Space or time adaptive signal processing by neural networks models. In *Intern. Conf. on Neural Networks for Computing*, pages 206–211, Snowbird (Utah, USA), 1986.

[14] A. Cichocki, R. Unbehauen, and E. Rummert. Robust learning algorithm for blind separation of signals. *Electronics Letters*, 30(17):1386–1387, 1994.

[15] J.-F. Cardoso and B. Laheld. An information-maximization approach to blind separation and blind deconvolution. *IEEE Trans. on SP*, 44:3017–3030, 1996.

[16] S. I. Amari, A. Cichocki, and Yang H. H. A new learning algorithm for blind source separation. *Advances in neural information processing systems*, 8:757–763, 1996.

[17] J.-L. Lacoume and P. Ruiz. Sources identification: a solution based on cumulants. In *IEEE ASSP Workshop*, Mineapolis, USA, August 1988.

[18] P. Comon. Separation of sources using higher-order cumulants. In *SPIE Vol. 1152 Advanced Algorithms and Architectures for Signal Processing IV*, San Diego (CA), USA, August 8-10 1989.

[19] A. Hyvärinen and E. Oja. A fast fixed point algorithm for independent component analysis. *Neural computation*, 9:1483–1492, 1997.

[20] N. Delfosse and Ph. Loubaton. Adaptive blind separation of independent sources: A deflation approach. *Signal Processing*, 45:59–83, 1995.

[21] T. Bell and T. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Comutation*, 7(6):1004–1034, 1995.

[22] D. L. Donoho. On minimum entropy deconvolution. In *Proc. 2nd Applied Time Series Symp.*, Tulsa, 1980. reprinted in *Applied Time Series Analysis II*, Academic Press, New York, 1981, pp. 565-609.

[23] D Yellin and E. Weinstein. Criteria for multichannel signal separation. *IEEE Trans. Signal Processing*, pages 2158–2168, August 1994.

[24] C. Jutten and J. Karhunen. Advances in blind source separation (BSS) and independent component analysis (ICA) for nonlinear mixtures. *International Journal of Neural Systems*, 14(5):1–26, 2004.

[25] A. Taleb and C. Jutten. Source separation in post nonlinear mixtures. *IEEE Tr. on SP*, 47(10):2807–2820, 1999.

[26] M. Babaie-Zadeh. *On blind source separation in convolutive and nonlinear mixtures*. PhD thesis, INP Grenoble, 2002.

[27] S. Achard and C. Jutten. Identifiability of post nonlinear mixtures. *IEEE Signal Processing Letters*, 12(5):423–426, May 2005.

[28] L. Tong, V. Soon, R. Liu, and Y. Huang. Amuse: a new blind identification algorithm. In *Proc. ISCAS*, New Orleans, USA, 1990.

[29] A. Belouchrani and J.-F. Cardoso. Maximum likelihood source separation by the expectation-maximization technique. In *NOLTA 95*, Las Vegas, USA, 1995.

[30] L. Molgedey and H.G. Schuster. Separation of a mixture of independent signals using time delayed correlations. *Physical Review Letters*, 72(23):3634–3637, June 1994.

[31] D. T. Pham and Ph. Garat. Blind separation of mixture of independent sources through a quasimaximum likelihood approach. *IEEE Trans. on SP*, 45:1712–1725, 1997.

[32] J.-F. Cardoso and A. Souloumiac. An efficient technique for blind separation of complex sources. In *Proc. IEEE Signal Processing Workshop on Higher-Order Statistics*, pages 275–279, South Lac Tahoe, USA (CA), June 1993.

[33] D.T. Pham. Blind separation of instantaneous mixtures via an independent component analysis. *SIAM J. on Matrix Anal. and Appl.*, 22(4):1136–1152, 2001.

[34] K. Matsuoka, M. Ohya, and M. Kawamoto. A neural net for blind separation of nonstationary signals. *Neural Networks*, 8(3):411–419, 1995.

[35] D.T. Pham, C. Servière, and H. Boumaraf. Blind separation of convolutive audio mixtures using nonstationarity. In *Proceedings of Int.Workshop on Independent Component Analysis and Signal Separation (ICA 2003)*, pages 981–986, Nara, Japan, April 2003.

[36] C. Puntonet, A. Mansour, and C. Jutten. A geometrical algorithm for blind separation of sources. In *Actes du XV Colloque GRETSI 95*, Juan-Les-Pins, France, Septembre 1995.

[37] M. Babaie-Zadeh, C. Jutten, and K. Nayebi. A geometric approach for separating Post Non-Linear mixtures. In *EUSIPCO*, volume II, pages 11–14, Toulouse, France, September 2002.

[38] M. Babaie-Zadeh, C. Jutten, and A. Mansour. Sparse ica via cluster-wise pca. *Neurocomputing*, 2006. (to be appeared).

[39] P. Bofill and M. Zibulevsky. Source separation using sparse representations. *Signal Processing*, 81:2353–2362, 2001. (to be appeared).

[40] Z. He and A. Cichocki. K-subspace clustering and its application in sparse component analysis. In *Proc. ESANN 2006*, Bruges, Belgium, April 2006.

[41] A. Belouchrani and J.-F. Cardoso. Maximum likelihood source separation for discrete sources. In *EUSIPCO*, pages 768–771, Edinburgh, Scotland, September 1994.

[42] C. Puntonet, A. Prieto, C. Jutten, M. Rodriguez-Alvarez, and J. Ortega. Separation of sources: A geometry-based procedure for reconstruction of n-valued signals. *Signal Processing*, 46(3):267–284, 1995.

[43] A. Taleb and C. Jutten. On underdetermined source separation. In *ICASSP'99*, Arizona, March 1999.

[44] P. Comon and O. Grellier. Non-linear inversion of under-determined mixtures. In *Proceedings of International Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, pages 461–465, Aussois, France, January 1999.

[45] O. Grellier and P. Comon. Blind separation of discrete sources. *IEEE Signal Processing Letters*, 5(8):212–214, August 1998.

[46] B. Lachover and A. Yeredor. Separation of polynomial post non-linear mixtures of discrete sources. In *Proceedings of The 2005 IEEE Workshop on Statistical Signal Processing (SSP2005)*, Bordeaux, France, July 2005.

[47] R. Gribonval and S. Lesage. A survey of sparse components analysis for blind source separation: principles, perspectives and new challenges. In *Proc. ESANN 2006*, Bruges, Belgium, April 2006.

[48] AK Barros, A. Mansour, and N. Ohnishi. Removing artifacts from ECG signals using independent components analysis. *Neurocomputing*, 22:173–18, 1998.

[49] R. Vigario and E. Ojae. Independence: a new criterion for the analysis of the electromagnetic fields in the global brain. *Neural Networks*, 13:891–907, 2000.

[50] L. de Lathauwer, B. de Moor, and J. Vandewalle. Fetal electrocardiogram extraction by blind source subspace separation. *IEEE Trans. on Biomedical Enginering*, 47:567–572, May 2000.

[51] F. Vrins, V. Vigneron, C. Jutten, and M. Verleysen. Abdominal electrodes analysis by statistical processing for fetal electrocardiogram extraction. In *BioMed 2004, 2nd IASTED Int. Conf. on Biomedical Engineering*, pages 244–249, Innsbruck (Austria), 2004.

[52] R. Sameni, , C. Jutten, and M. B. Shamsollahi. What ICA Provides for ECG Processing: Application to Noninvasive Fetal ECG Extraction. In *Proceedings of the 6th IEEE International Symposium on Signal Processing and Information Technology, Aug. 27-30, 2006, Vancouver, Canada*, 2006.

[53] S. Houcke, A. Chevreuil, and Ph. Loubaton. Blind equalization: case of an unknown symbol period. *IEEE Trans. on Signal Processing*, 51:781–793, 2003.

[54] P. Jallon and A. Chevreuil. Second-order based cyclic frequency estimates : the case of digital communication signals. In *Proceedings of ICASSP 2006*, Toulouse, France, May 2005.

[55] H.L. Nguyen Thi and C. Jutten. Blind sources separation for convolutive mixtures. *Signal Processing*, 45:209–229, 1995.

[56] E. Vincent. Musical source separation using time-frequency source priors. *IEEE Trans. on Audio, Speech and Language Processing*, 14(1):91–98, 2006.

[57] F. Abrard and Y. Deville. A time-frequency blind signal separation method applicable to underdetermined mixtures of dependent sources. *Signal Processing*, 85(7):1389–1403, 2005.

[58] W. Wang, D. Cosker, Y. Hicks, S. Sanei, and J. Chambers. Video assisted speech source separation. In *Proceedings of the International Conference of Acoustic, Speech and Signal Processing (ICASSP2005)*, Philadelphia (USA), March 2005.

[59] B. Rivet, L. Girin, C. Jutten, and J.-L. Schwartz. Solving the indeterminations of blind source separation of convolutive speech mixtures. In *Proceedings of the International Conference of Acoustic, Speech and Signal Processing (ICASSP2005)*, Philadelphia (USA), March 2005.

[60] B. Rivet, C. Servière, L. Girin, D.T. Pham, and C. Jutten. Using a visual voice activity detector to cance the permutation indeterminacy in blind source separation of speech mixtures. In *Proceedings of EUSIPCO 06*, Florence, Italy, September 2006. submitted.

[61] G. D'Urso, P. Prieur, C. Vincent, F. Bonnaire, M. Rouzaud, and L. Lacroix. Modélisation des déplacements de barrages. In *16 ème colloque international sur le traitement du signal et des images (GRETSI 97)*, pages 215–218, Grenoble, France, septembre 1997.

[62] A. Paraschiv-Ionescu, C. Jutten, and G. Bouvier. Source separation based processing for smart sensor arrays. *IEEE Sensors Journal*, 2(6):663–673, December 2002.

[63] G. Bedoya, C. Jutten, S. Bermejo, and J. Cabestany. Improving semiconductor-based chemical sensor arrays using advanced algorithms for blind source separation. In *Proceedings of the ISA/IEEE Sensors for Industry Conference (SiCon04)*, pages 149–154, New Orleans (USA), January, 27-29 2004.

[64] P. O. Hoyer. Non-negative sparse coding. In *Proceedings of IEEE Workshop on Neural Networks for Signal Processing (NNSP'2002)*, pages 557–565, 2002.

[65] S. Moussaoui, D. Brie, and C. Carteret. Non-negative source separation using the maximum likelihood approach. In *Proceedings of the International Workshop of Statistical Signal Processing (SSP2005)*, Bordeaux, France, July 2005.

[66] M. Lennon, G. Mercier, M. Mouchot, and L. Hubert-Moy. Spectral unmixing of hyperspectral images with the independent component analysis and wavelet packets. In *Int. Geoscience Remote Sensing Symposium*, Sydney, Australia, July 2001.

[67] H. Hauksdottir, C. Jutten, F. Schmidt, J. Chanussot, J.A. Benediktsson, and S. Douté. The physical meaning of independent components and artifact removal of hyperspectral data from Mars using ICA. In *7th Nordic Signal Processing Symposium (NORSIG'2006)*, page , Reykjavik, Iceland, 2006.

[68] S. Moussaoui, D. Brie, C. Carteret, and A. Mohammad-Djafari. Application of bayesian non-negative source separation to mixture analysis in spectroscopy. In *Proceedings of the 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MaxEnt'2004)*, Garching, Germany, July 2004.

[69] A. Mohamed-Djafari. Bayesian source separation: beyong PCA and ICA. In *Proc. ESANN 2006*, Bruges, Belgium, April 2006.