A Necessary and Sufficient Condition for the Blind Extraction of the Sparsest Source in Convolutive Mixtures

Yves-Marie Batany*†, Daniela Donno*, Leonardo Tomazeli Duarte‡, Hervé Chauris*, Yannick Deville§ and João Marcos Travassos Romano†
* Centre de Géosciences, MINES ParisTech, PSL Research University
35 rue Saint Honoré, 77305 Fontainebleau, France
Email: yves-marie.batany@mines-paristech.fr
† School of Electrical and Computing Engineering, University of Campinas (UNICAMP)
Avenida Albert Einstein, 400, CEP 13083-852, Campinas-SP Brazil
‡ School of Applied Sciences, University of Campinas (UNICAMP)
Rua Pedro Zaccaria, 1300, CEP 13484-350, Limeira-SP, Brazil
§ Institut de Recherche en Astrophysique et Planétologie (IRAP), Université de Toulouse, UPS-CNRS-OMP
14 Avenue Edouard Belin, 31400 Toulouse, France

***Abstract*—This paper addresses sparse component analysis, a powerful framework for blind source separation and extraction that is built upon the assumption that the sources of interest are sparse in a known domain. We propose and discuss a necessary and sufficient condition under which the ℓ_0 pseudo-norm can be used as a contrast function in the blind source extraction problem in both instantaneous and convolutive mixing models, when the number of observations is at least equal to the number of sources. The obtained conditions allow us to relax the sparsity constraint of the sources to its maximum limit, with possibly overlapping sources. In particular, the W-disjoint orthogonality assumption of the sources can be discarded. Moreover, no assumption is done on the mixing system except invertibility. A differential evolution algorithm based on a smooth approximation of the ℓ_0 pseudo-norm is used to illustrate the benefits brought by our contribution.**

***Keywords*—Blind Source Separation, Blind Source Extraction, Convolutive Mixture, Sparse Component Analysis, ℓ_0 pseudo-norm, MIMO.**

I. Introduction

Blind source separation (BSS) and blind source extraction (BSE) problems arise in many applications such as speech processing, medical imaging or geophysics [1]. Independent component analysis (ICA) has been developed as a powerful tool for solving BSS and BSE problems when the original sources can be considered as statistically independent random variables [1]. In some cases, this assumption fails and other priors must be considered such as sparsity, for which the ℓ_0 pseudo-norm is the most common measure. Sparse component analysis (SCA) has emerged as another powerful tool for solving BSS and BSE problems [2].

Since earliest works [3], SCA has shown to be efficient in the blind identification of the mixing system, especially in under-determined problems when there are less observations than sources. Some techniques make the strong assumption that the original sources are W-disjoint orthogonal [4], allowing only one source to be active at each point of the considered signal representation. Other techniques assume that the sources and the mixing system are non-negative. For this assumption, the necessary and sufficient conditions have been discussed in [5].

For instantaneous mixtures with at least as many observations than sources, a sufficient condition for the ℓ_0 pseudo-norm to be a contrast function for BSE has been proposed in [6]. But, to our knowledge, necessary and sufficient conditions, on the sources only, have not been discussed yet. In this sense, the present work extend some results proposed in [6]. The paper is organised as follows. BSS and BSE problems are presented in section II. Section III gives the definitions of the key concepts used hereafter. Section IV presents our results on necessary and sufficient conditions. Finally, section V describes a numerical example with a differential evolution algorithm.

II. Blind Source Separation and Extraction

A BSS problem consists of recovering a set of original signals $\boldsymbol{s}[n] \in \mathbb{R}^Q$, $n = 1, \ldots, N$, through R linear combinations of these sources. For an instantaneous mixture, the observations are given by the mixing equation

$$\boldsymbol{x}[n] = \boldsymbol{A}\boldsymbol{s}[n], \tag{1}$$

where the mixing matrix $\boldsymbol{A} \in \mathbb{R}^{R\times Q}$ is unknown. For a finite impulse response multiple-input multiple-output (FIR-MIMO) convolutive mixture [7], the mixing equation is given by

$$\boldsymbol{x}[n] = \sum_{k=0}^{K} \boldsymbol{A}_k \boldsymbol{s}[n-k], \tag{2}$$

where K is the memory length of the mixing system and all matrices $\boldsymbol{A}_k \in \mathbb{R}^{R\times Q}$ are unknown. We consider $\boldsymbol{s}[n-k] = \boldsymbol{0}$ if $n-k < 1$ in Equation 2, i.e. that the sources are zero-padded.

If $R < Q$, the problem is called under-determined. If $R = Q$ or $R > Q$ the problem is called determined or over-determined, respectively. When there is only one source, i.e. $Q = 1$, the problem is known as blind deconvolution (BD). In this work, we consider $R \geq Q$, i.e. the number of observations is at least equal to the number of sources. As shown in Figure 1, $\boldsymbol{s}[n] \in \mathbb{R}^Q$ denotes the vector containing all the source values at time n and $\boldsymbol{s}_q \in \mathbb{R}^N$ gathers all the values of one source at indices $n = 1, 2, \ldots, N$. A non-bold character $s_q[n]$ indicates one single value.

The BSE problem is closely related to BSS but it aims at recovering a single source. In SCA-based BSE problems, the sparsest source must be recovered. The sparsity of a vector $\boldsymbol{y}$ can be measured by its ℓ_0 pseudo-norm defined as

$$\|\boldsymbol{y}\|_0 = \#\{y[n] \ : \ y[n] \neq 0\}. \tag{3}$$

The ℓ_0 pseudo-norm gives the size of the active support of a vector. The size of the inactive support is given by $N - \|\boldsymbol{y}\|_0$. Previous works on SCA-based BSS mainly consider the "column sparsity" $\|\boldsymbol{s}[n]\|_0$ of the signals [5]. In our work, we consider the "row sparsity" $\|\boldsymbol{s}_q\|_0$ of the signals. A determined instantaneous BSE problem based on the ℓ_0 pseudo-norm can be solved up to an amplitude ambiguity[1] because $\|\alpha\boldsymbol{y}\|_0 = \|\boldsymbol{y}\|_0, \ \forall \alpha \in \mathbb{R}\backslash\{0\}$. For a determined convolutive BSE problem, an additional shift ambiguity appears after the zero padding of $\boldsymbol{y}$ because $\|\delta_k(\boldsymbol{y})\|_0 = \|\boldsymbol{y}\|_0$ for any time shifting operator δ_k defined as $\delta_k(\boldsymbol{y}) = \boldsymbol{y}[n-k]$.

The extraction of a source in a determined convolutive BSE problem can be achieved by finding a multiple-input single-output (MISO) separating system of $L+1$ extraction vectors $\boldsymbol{w}_l \in \mathbb{R}^R$, $l = 0, 1, \ldots, L$ with $L \geq K$, such that

$$y[n] = \sum_{l=0}^{L} \boldsymbol{w}_l^T \boldsymbol{x}[n-l] \tag{4}$$

$$= \sum_{l=0}^{L} \boldsymbol{w}_l^T \sum_{k=0}^{K} \boldsymbol{A}_k \boldsymbol{s}[n-l-k] \tag{5}$$

$$= \sum_{j=0}^{J} \boldsymbol{h}_j^T \boldsymbol{s}[n-j], \tag{6}$$

where we defined $J = K + L$ and the vectors $\boldsymbol{h}_j$ are the mapping vectors between the original sources and the extracted signal defined such that $\boldsymbol{h}_j^T = \sum_{l=0}^{L} \boldsymbol{w}_l^T \boldsymbol{A}_{j-l}$ with $0 < j-l \leq K$. We consider $\boldsymbol{x}[n-l] = \boldsymbol{0}$ if $n-l < 1$ in Equation 4, i.e. that the observations are zero-padded. For instantaneous problems, we have $K = L = 0$ and a single extraction vector $\boldsymbol{w}_0$ needs to be recovered.

III. Definitions

Considering the amplitude and shift ambiguities in BSE problems based on the ℓ_0 pseudo-norm, we call $\mathcal{S}_q$ the set of solutions in the extracted vector space corresponding to the extraction of $\boldsymbol{s}_q$ such that

$$\mathcal{S}_q = \{\boldsymbol{y} \ : \ \boldsymbol{y} = \alpha\delta_k(\boldsymbol{s}_q), \ \forall \alpha \in \mathbb{R}\backslash\{0\}, \ \forall k \in \mathbb{Z}\}, \tag{7}$$

where δ_k denotes the time shift operator. We call $\mathcal{G}_q$ the set of all mapping vectors such that $\{\boldsymbol{h}_j\}_{j=0}^{J} \in \mathcal{G}_q \ \Leftrightarrow \ \boldsymbol{y} \in \mathcal{S}_q$ (see Equation (6)). The set $\mathcal{G}_q$ denotes the solution set of global mapping corresponding to the correct extraction of the source

[1]In BSS problems, a permutation ambiguity also appears.

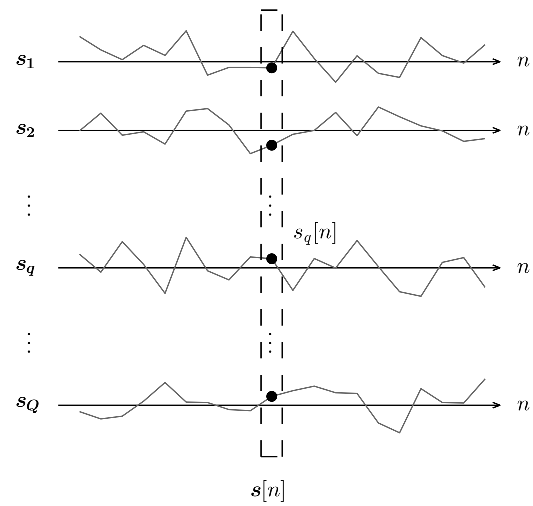

Fig. 1. Clarification of the notations: $\boldsymbol{s}_q$ contains all the values of one source and $\boldsymbol{s}[n]$ contains the values of all sources for index n. A single sample of a source q for index n is denoted $s_q[n]$. The same notations are used for the observations.

$\boldsymbol{s}_q$. The BSE problem is equivalently solved when $\boldsymbol{y} \in \mathcal{S}_q$ or when $\{\boldsymbol{h}_j\}_{j=0}^{J} \in \mathcal{G}_q$.

Definition 1: A function f is said to be a contrast function for the extraction of the source $\boldsymbol{s}_q$ if $f(\boldsymbol{y}) > f(\boldsymbol{s}_q), \ \forall \boldsymbol{y} \notin \mathcal{S}_q$ [1].

The properties describing a set of source signals $\{\boldsymbol{s}_q\}_{q=1}^{Q}$ can be divided in two categories. A first category contains the properties of each single signal $\boldsymbol{s}_q$ taken independently from the others. A second category contains the properties linked to the relations between several signals, i.e. $\boldsymbol{s}[n]$. For instance, the kurtosis is defined for a single source, independently from the others. On the other hand, the covariance structure or the statistical dependence are properties defining the way all the sources interact between them.

Auto-regressive processes refer to the first category of signal properties. A signal $\boldsymbol{s}_q$ is said to be an auto-regressive process of order D if there exists a set of $D+1$ parameters c_d such that [8]

$$\sum_{d=0}^{D} c_d s_q[n-d] = 0, \tag{8}$$

where at least two parameters c_d are non-null. From a geometric point of view, any set of $D+1$ consecutive coefficients extracted from an auto-regressive process is located in a hyperplane in $\mathbb{R}^{D+1}$ defined by its normal vector $\boldsymbol{c} = \{c_d\}_{d=0}^{D}$.

We propose to extend the concept of *auto-regressive* processes to the second category of properties by introducing the concept of *inter-regressive* processes. A set of Q signals is said to be an inter-regressive process of order D if there exists a set of $Q \times (D+1)$ parameters c_{qd} such that

$$\sum_{q=1}^{Q} \sum_{d=0}^{D} c_{qd} s_q[n-d] = 0, \tag{9}$$

where at least two parameters c_{qd} are non-null. From a geometric point of view, any set of $D+1$ consecutive source

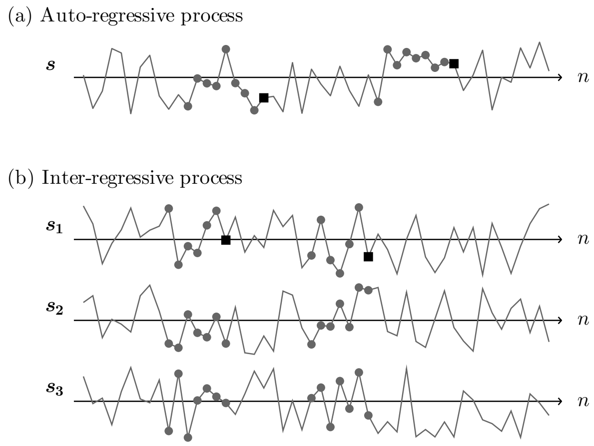

Fig. 2. Examples of construction of (a) an auto-regressive process of order $D = 8$ and length $E = 2$ and (b) an inter-regressive process of order $D = 6$ and length $E = 2$ for $Q = 3$ signals. See detailed comments in the text below.

vectors extracted from an inter-regressive process and forming a set of $Q(D+1)$ coefficients is located in a hyperplane in $\mathbb{R}^{Q(D+1)}$. For an inter-regressive process of order $D = 0$, Equation (9) can be written as $\boldsymbol{c}^T \boldsymbol{s}[n] = 0$. When there is only one source, $Q = 1$, an inter-regressive process becomes an auto-regressive process.

Both the above definitions generally consider that Equations (8) and (9) must be true for all indices n or for a closed support of consecutive indices. For our purpose, we propose to expand these definitions. A signal is said to yield an auto-regressive process of order D and length $E \in \mathbb{N}$ if Equation (8) is true for E indices n, possibly not consecutive. Equivalently, signals are said to yield an inter-regressive process of order D and length $E \in \mathbb{N}$ if Equation (9) is true for E indices n, possibly not consecutive.

Figure 2 shows the construction of both auto-regressive and inter-regressive processes of hypothetical length $E = 2$. For the auto-regressive process, each black square is fully determined by the value of all antecedent grey dots, with a unique set of parameters $\{c_d\}$. For the inter-regressive process, each black square is fully determined by the value of all antecedent and concomitant grey dots, with a unique set of parameters $\{c_{qd}\}$. Figure 3 shows an example of an inter-regressive process of order $D = 0$ and length $E = 6$ for three sources. The position of each samples is shown in the source space.

IV. Necessary and Sufficient Conditions

In this section, we give necessary and sufficient conditions, on the sources only, to use the ℓ_0 pseudo norm as a contrast function in linear BSE problems, for both instantaneous and convolutive mixtures. In other words, we discuss the conditions under which the solution of the ℓ_0 pseudo-norm minimisation problem

$$\{\boldsymbol{w}_l\}^* = \min_{\{\boldsymbol{w}_l\}} \left\| y[n] = \sum_{l=0}^{L} \boldsymbol{w}_l^T \boldsymbol{x}[n-l] \right\|_0 , \qquad (10)$$

extracts a signal $\boldsymbol{y}^* \in \mathcal{S}_1$ corresponding to the recovery of the sparsest source. Without loss of generality, we consider that the sources are sorted in order of decreasing sparsity such that $\|\boldsymbol{s}_1\|_0 < \|\boldsymbol{s}_2\|_0 < \cdots < \|\boldsymbol{s}_Q\|_0$. We also assume that $\|\boldsymbol{s}_1\|_0 < \|\boldsymbol{s}_q\|_0, \ \forall q \neq 1$, to avoid any competition between the extraction of the sparsest source $\boldsymbol{s}_1$ and another source. For the sake of clarity, the instantaneous case is treated first and then the generalisation to the convolutive case is presented. Both proofs are similar.

Theorem 1 (Extraction from instantaneous mixture): The ℓ_0 pseudo-norm is a contrast function for the extraction of the sparsest source $\boldsymbol{s}_1$ if and only if the sources do not have any inter-regressive process of order 0 with a length higher than or equal to the size of the inactive support of $\boldsymbol{s}_1$.

Proof: From Definition 1, the ℓ_0 pseudo-norm is a contrast function for the extraction of $\boldsymbol{s}_1$ if and only if

$$\|\boldsymbol{s}_1\|_0 < \|\boldsymbol{y}\|_0 \qquad \forall \boldsymbol{y} \notin \mathcal{S}_1,$$

i.e. if and only if

$$\begin{aligned}
&\|\boldsymbol{s}_1\|_0 < \#\{y[n] = \boldsymbol{h}^T \boldsymbol{s}[n] : y[n] \neq 0\} && \forall \boldsymbol{h} \notin \mathcal{G}_1,\\
&\|\boldsymbol{s}_1\|_0 < N - \#\{y[n] = \boldsymbol{h}^T \boldsymbol{s}[n] : y[n] = 0\} && \forall \boldsymbol{h} \notin \mathcal{G}_1,\\
&N - \|\boldsymbol{s}_1\|_0 > \#\{y[n] = \boldsymbol{h}^T \boldsymbol{s}[n] : y[n] = 0\} && \forall \boldsymbol{h} \notin \mathcal{G}_1,\\
&N - \|\boldsymbol{s}_1\|_0 > E^*,
\end{aligned}$$

where we defined

$$E^* = \max \left[\#\{y[n] = \boldsymbol{h}^T \boldsymbol{s}[n] : y[n] = 0\}, \ \boldsymbol{h} \notin \mathcal{G}_1 \right].$$

$N - \|\boldsymbol{s}_1\|_0$ is the number of null values of $\boldsymbol{s}_1$. E^* is the maximum length of an inter-regressive process of order 0 among the sources. ■

Theorem 2 (Extraction from convolutive mixture): The ℓ_0 pseudo-norm is a contrast function for the extraction of the sparsest source $\boldsymbol{s}_1$ if and only if the sources do not have any inter-regressive process of order $J = K + L$ with a length higher than or equal to the size of the inactive support of $\boldsymbol{s}_1$.

Proof: From Definition 1, the ℓ_0 pseudo-norm is a contrast function for the extraction of $\boldsymbol{s}_1$ if and only if

$$\|\boldsymbol{s}_1\|_0 < \|\boldsymbol{y}\|_0 \qquad \forall \boldsymbol{y} \notin \mathcal{S}_1,$$

i.e. if and only if

$$\begin{aligned}
&\|\boldsymbol{s}_1\|_0 < \#\{y[n] = \sum_{j=0}^{J} \boldsymbol{h}_j^T \boldsymbol{s}[n-j] : y[n] \neq 0\} && \forall \boldsymbol{h} \notin \mathcal{G}_1,\\
&\|\boldsymbol{s}_1\|_0 < N - \#\{y[n] = \sum_{j=0}^{J} \boldsymbol{h}_j^T \boldsymbol{s}[n-j] : y[n] = 0\} && \forall \boldsymbol{h} \notin \mathcal{G}_1,\\
&N - \|\boldsymbol{s}_1\|_0 > \#\{y[n] = \sum_{j=0}^{J} \boldsymbol{h}_j^T \boldsymbol{s}[n-j] : y[n] = 0\} && \forall \boldsymbol{h} \notin \mathcal{G}_1,\\
&N - \|\boldsymbol{s}_1\|_0 > E^*,
\end{aligned}$$

where we defined

$$E^* = \max \left[\#\{y[n] = \sum_{j=0}^{J} \boldsymbol{h}_j^T \boldsymbol{s}[n-j] : y[n] = 0\}, \ \boldsymbol{h} \notin \mathcal{G}_1 \right].$$

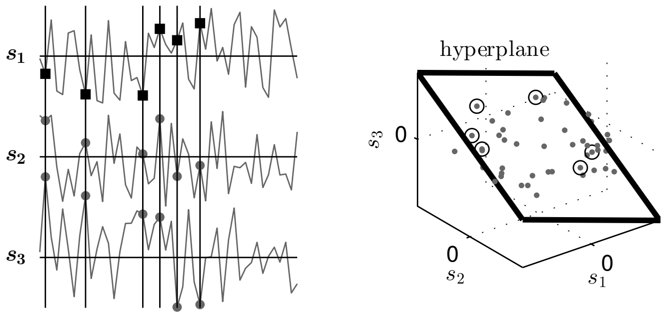

Fig. 3. Example of three sources having an inter-regressive process of order $D = 0$ and length $E = 6$. In the right hand figure, the grey dots circled in black in the source space belong to this inter-regressive process and are inside the same hyperplane.

$N - \|\boldsymbol{s}_1\|_0$ is the number of null values of $\boldsymbol{s}_1$. E^* is the maximum length of an inter-regressive process of order $J = K + L$ among the sources. ■

In both Theorems 1 and 2, the assumption of W-disjoint orthogonality of the sources is not necessary and the sources can overlap. This will be shown in the next section. SCA is often presented as a method able to separate signals violating the independence assumption. We emphasis here that this assertion is true below the limit defined by Theorem 1 and 2: the limit for SCA-based BSE is one kind of strong linear dependency among the sources, named an inter-regressive process.

V. Numerical Examples

We propose to illustrate Theorem 2 using a synthetic example solved with a differential evolution (DE) algorithm [9], combined with a gradient approach. The ℓ_0 pseudo-norm is approximated by a smooth version (SL0) proposed by [10] with a Gaussian kernel such that

$$\|\boldsymbol{y}\|_{0,\sigma} = N - \sum_{n=1}^{N} \mathrm{e}^{-y[n]^2/2\sigma^2}, \tag{11}$$

where σ is a shaping parameter and the vector $\boldsymbol{y}$ is normalised by its ℓ_2 norm. We fixed $2\sigma^2 = 10^{-3}$ as we have $\lim_{\sigma\to 0} \|\boldsymbol{y}\|_{0,\sigma} = \|\boldsymbol{y}\|_0$. The search of the extraction vectors $\{\boldsymbol{w}_l\}_{l=0}^{L}$ is performed in the time domain. The scale and shift ambiguities are not avoided by a specific parametrisation but an extracted vector $\boldsymbol{y}$ is projected back on the ℓ_2-ball before the computation of its smooth ℓ_0 pseudo-norm.

We adopt a DE/rand/1 strategy [9]. The number of individuals in the population is chosen to be ten times the number of parameters, i.e. $N_{pop} = 10 \times R(L+1)$. A single individual in the population is denoted $\underline{\boldsymbol{w}}^{i_0}$ where the superscript $i_0 \in \{1, 2, \ldots, N_{pop}\}$ indicates its position in the population. An individual $\underline{\boldsymbol{w}}^{i_0}$ contains a set of extracting vector $\{\boldsymbol{w}_l^{i_0}\}_{l=0}^{L}$. At each generation and for each target individual $\underline{\boldsymbol{w}}^{i_0}$, a mutant individual $\underline{\boldsymbol{v}}^{i_0}$ is created such that

$$\underline{\boldsymbol{v}}^{i_0} = \underline{\boldsymbol{w}}^{i_1} + F \times (\underline{\boldsymbol{w}}^{i_2} - \underline{\boldsymbol{w}}^{i_3}) \tag{12}$$

where F is the DE scale parameter and the superscripts i_1, i_2, i_3 denote individuals different from i_0. A trial individual $\underline{\boldsymbol{u}}^{i_0}$ is created by crossing the target and the mutant individuals in such a way that each parameter $\underline{u}^{i_0}[r]$ has a probability Cr to equal $\underline{v}^{i_0}[r]$ and a probability $(1 - Cr)$ to equal $\underline{w}^{i_0}[r]$. The best individual between the target and the trial individual is kept in the next generation. We set $F = 1$ and $Cr = 1/2$.

After a sufficient convergence of the population around a solution by using the DE algorithm, a gradient approach is used to update the best individual and obtain the final solution. The gradient of the objective function with respect to a parameter is given by

$$\frac{\partial \|\boldsymbol{y}\|_{0,\sigma}}{\partial w_l[r]} = \frac{1}{\sigma^2} \sum_{n=1}^{N} x_r[n-l] y[n] \mathrm{e}^{-y[n]^2/2\sigma^2}. \tag{13}$$

The best individual is updated in opposite direction of the gradient until convergence.

Sparse signals of length $N = 128$ are mixed according to Equation (2) by a FIR-MIMO system of length $K = 4$ with $Q = 2$ inputs and $R = 3$ outputs. This configuration guarantees that the global mixing-separating system exists, i.e. that the z-transform polynomial matrix representing the mixing system is left invertible [11]. The extraction is performed after zero-padding the observations by considering the FIR-MISO system of Equation (4) with length $L = 7$. Figure 4 shows the recovery of the sparse signal $\boldsymbol{s}_1$ satisfying the conditions required by Theorem 2. Figure 5 shows an example in which the conditions required by Theorem 2 are violated by constructing an inter-regressive process among the sources. In this case, the desired sparsest signal $\boldsymbol{s}_1$ cannot be recovered.

VI. Conclusion

Our work focused on the problem of blind extraction of the sparsest source in both instantaneous and convolutive models. By considering the case in which the number of mixtures is greater or equal to the number of sources, we provide a necessary and sufficient condition for the extraction of the sparsest source when minimizing the ℓ_0 pseudo-norm. Future work will include the generalisation to the blind source separation problem and complex-valued sources. Also, only FIR systems have been considered. A more general framework should include systems with infinite impulse response as well. Finally, the robustness to noise should be investigated. These aspects will be treated in a future expanded article.

References

[1] P. Comon and C. Jutten, *Handbook of Blind Source Separation*. Oxford: Academic Press, 2010.

[2] Y. Deville, "Sparse component analysis: A general framework for linear and nonlinear blind source separation and mixture identification," in *Blind Source Separation*. Springer, 2014, pp. 151–196.

[3] P. Bofill and M. Zibulevsky, "Underdetermined blind source separation using sparse representations," *Signal Processing*, vol. 81, no. 11, pp. 2353–2362, 2001.

[4] S. Rickard, "The duet blind source separation algorithm," in *Blind Speech Separation*. Springer, 2007, pp. 217–241.

[5] Y. Sun and J. Xin, "Underdetermined sparse blind source separation of nonnegative and partially overlapped data," *SIAM Journal on Scientific Computing*, vol. 33, no. 4, pp. 2063–2094, 2011.

[6] L. T. Duarte, R. Suyama, R. Attux, J. M. T. Romano, and C. Jutten, "Blind extraction of sparse components based on ℓ_0-norm minimization," in *IEEE Statistical Signal Processing (SSP) Workshop*, Nice, France, June 2011, pp. 617–620.

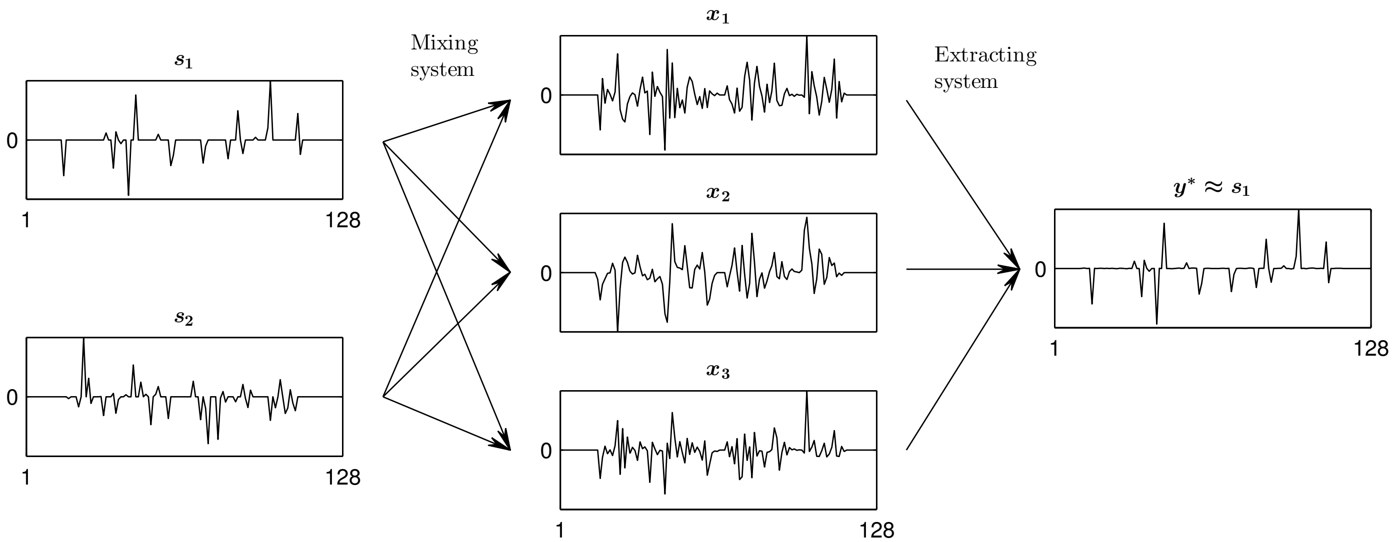

Fig. 4. Example of a successful blind extraction of a sparse source from convolutive mixture with $Q = 2$ sources, $R = 3$ observations and $N = 128$ samples. The observations are mixed by filters of size $K = 4$. The extraction of the signal $\boldsymbol{y}^*$ is achieved by finding 3 filters of size $L = 7$. We observe perfect recovery in this case.

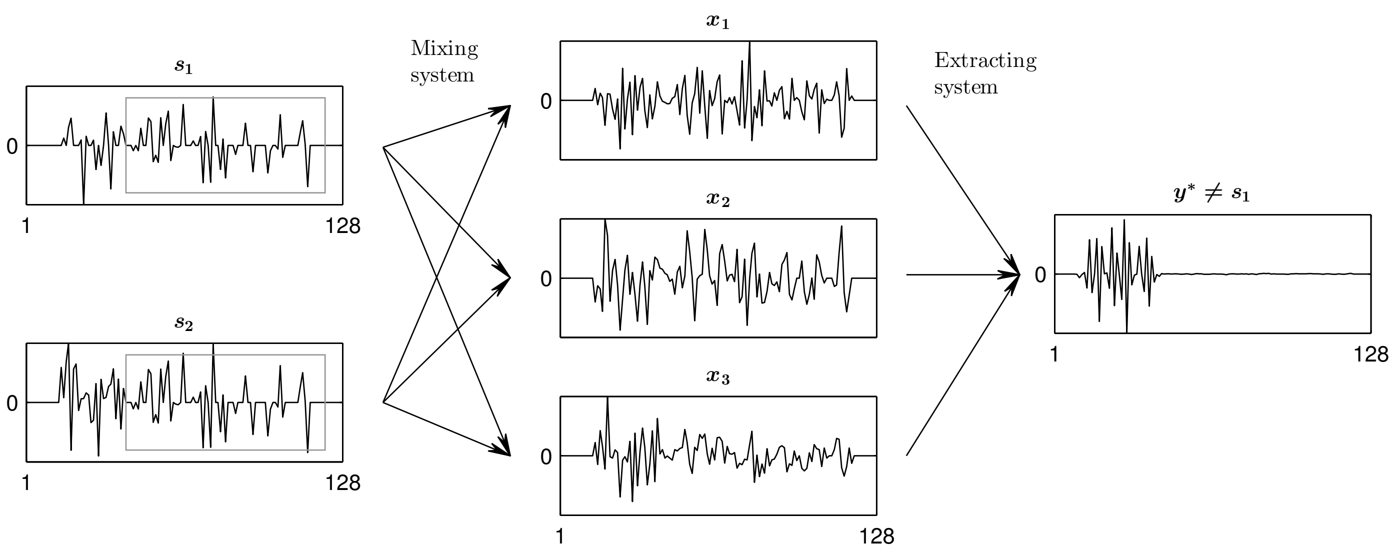

Fig. 5. Example of an unsuccessful blind extraction due to the violation of the conditions of theorem 2. The mixing model is the same as in figure 4. However, the sources are different: there exists an inter-regressive process of order $D = 0$ among the sources. This inter-regressive process is indicated by grey rectangles. The presence of an inter-regressive process among the sources creates a parasitic minimum in the objective function leading to an incorrect source extraction.

[7] M. S. Pedersen, J. Larsen, U. Kjems, and L. C. Parra, "Convolutive blind source separation methods," in *Springer Handbook of Speech Processing*. Springer, 2008, pp. 1065–1094.

[8] S. K. Mitra and J. F. Kaiser, *Handbook for Digital Signal Processing*. J. Wiley & Sons, 1993.

[9] R. Storn and K. Price, "Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces," *Journal of Global Optimization*, vol. 11, no. 4, pp. 341–359, 1997.

[10] F. Naini, G. Mohimani, M. Babaie-Zadeh, and C. Jutten, "Estimating the mixing matrix in sparse component analysis (SCA) based on multidimensional subspace clustering," in *IEEE International Conference on Telecommunications and Malaysia International Conference on Communications (ICT-MICC)*, May 2007, pp. 670–675.

[11] K. Law, R. Fossum, and M. Do, "Generic invertibility of multidimensional fir filter banks and mimo systems," *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4282–4291, 2009.