

# Nonlinear Blind Source Separation for Sparse Sources

Bahram Ehsandoust<sup>\*†</sup>, Bertrand Rivet<sup>\*</sup>, Christian Jutten<sup>\*</sup>, Massoud Babaie-Zadeh<sup>†</sup>

<sup>\*</sup> University Grenoble Alpes, CNRS, GIPSA-lab, Grenoble, France

{bahram.ehsandoust@gipsa-lab, bertrand.rivet@phelma, christian.jutten@gipsa-lab}.grenoble-inp.fr

<sup>†</sup> Sharif University of Technology, Electrical Engineering Department, Tehran, Iran

ehsandoust@ee.sharif.edu and mbzadeh@sharif.edu

**Abstract**—Blind Source Separation (BSS) is the problem of separating signals which are mixed through an unknown function from a number of observations, without any information about the mixing model. Although it has been mathematically proven that the separation can be done when the mixture is linear, there is not any proof for the separability of nonlinearly mixed signals.

Our contribution in this paper is performing nonlinear BSS for sparse sources. It is shown in this case, sources are separable even if the problem is under-determined (the number of observations is less than the number of source signals). However in the most general case (when the nonlinear mixing model can be of any kind and there is no side-information about that), an unknown nonlinear transformation of each source is reconstructed. It is shown why the problem reconstructing the exact sources is severely ill-posed and impossible to do without any other information.

**Index Terms**—Blind Source Separation, Independent Component Analysis, Sparse Signals, Manifold Learning

## I. INTRODUCTION

In signal processing applications, it is usual to have a number of signals measured by some sensors, while each of them might be a mixture of a number of source signals. BSS is the problem of separating the original sources from the observed mixture, without any information about either the source signals or the mixing model. The basic model of the mixing model is depicted in Fig. 1, and can be expressed as

$$\mathbf{x}(t) = \mathbf{f}(s(t)) \quad (1)$$

where  $\mathbf{f} : \mathbb{R}^N \rightarrow \mathbb{R}^M$  is the unknown mixing function,  $N$  and  $M$  are the number of source and observation signals respectively and  $\mathbf{s}(t) = [s_1(t), s_2(t), \dots, s_N(t)]^T$  and  $\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_M(t)]^T$  ( $T$  denotes the transpose of the vector) are the source and observation vectors respectively. The BSS problem can also be written as

$$\text{find } \mathbf{g} \quad \text{s.t.} \quad \mathbf{y}(t) = \mathbf{g}(\mathbf{x}(t)) = \mathbf{g}(\mathbf{f}(s(t))) = \hat{\mathbf{s}}(t) \quad (2)$$

where  $\mathbf{g} : \mathbb{R}^M \rightarrow \mathbb{R}^N$  is the separating function which is to be estimated and  $\mathbf{y}(t) = [y_1(t), y_2(t), \dots, y_N(t)]^T$  is the vector of reconstructed sources (the output of the BSS algorithm).

It should be noted that the goal of BSS is to “separate” the sources and not to “reconstruct” them. Depending on the mixing model, there are some ambiguities in reconstructing the sources that can only be resolved using some prior information about the source signals, which is out of the scope of BSS

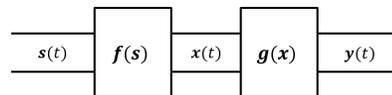


Fig. 1. BSS problem basic model

(for example in case of linear mixtures, separated sources would be equal to the original ones up to a scaling factor and reordering).

BSS problem has been studied since the early works in 80s [1], and noticeable results have been achieved for some cases. The general idea of performing BSS is to design a separating function ( $\mathbf{g}$ ) such that each entry of the output signals  $\mathbf{y}(t)$  only depends on a single original source. For example, it has been shown [2] that if source signals are mutually independent, the mixing function ( $\mathbf{f}$ ) is linear and the number of the sources is equal to the number of the observations ( $\mathbf{x}(t)$ ), one can perform the separation by making the outputs mutually independent as well as the sources (which is called ICA<sup>1</sup>). In this case, the “independence” is the property of the sources, which is tried to be recovered for the separation. There are some other works exploiting other properties to perform the separation (a comprehensive survey can be found in [2]), such as non-negativity [3], non-Gaussianity [4], non-stationarity [5] and non-whiteness [6] (they are named as the “non-properties” [7]).

Although there are many works and proposed algorithms on BSS for linear mixtures, nonlinear BSS has not been well developed ([8] provides a short survey on nonlinear BSS). It is shown through counter examples that ICA does not separate the sources which are non-linearly mixed [9], [10]. Therefore, except a few general approaches which try to investigate the separability of any kind of nonlinear mixture models [11], [12], studies on nonlinear BSS are focused on specific mixture model and source signals, out of which we can name Post Nonlinear (PNL) [13] and Bi-Linear (or Linear Quadratic) mixtures [14], Convolutional Post Nonlinear mixtures [10] and conformal mappings [15].

In this paper, we consider the problem where the mixing model is nonlinear and the sources are sparse (the case of

<sup>1</sup>Independent Component Analysis

sparse sources has already been studied for linear mixtures in many works like [16]). So the mixing model that is studied in this work is not limited to a specific kind and can be any invertible function. We show that in this case, the sources can be separated up to a nonlinear function.

The rest of the paper is organized as follows. In the next section, the main idea is introduced and the approach and the algorithm for performing the separation are proposed. Then some simulation results are shown in the section III. A comprehensive discussion on the performance of the proposed approach and how to develop it for the future works is finally made in section IV.

## II. THE PROPOSED ALGORITHM

Sparse signals in the time domain are the signals which most of the time take the value zero. In other words, the probability of their value being equal to zero is high. Therefore, if all of the sources are sparse, it will be even more rare that they take non-zero values at the same time. This is the main feature that is exploited in this approach to perform the separation. The proposed algorithm comprises two main steps of 1) Clustering the Observations and 2) Unfolding the Manifold, which are described as follows.

### A. Step 1 - Clustering the Observations

Let us assume a  $2 \times 2$  mixing system (2 source signals and 2 observations) as an example and

$$x_1(t) = e^{s_1(t)} - e^{s_2(t)} \quad (3)$$

$$x_2(t) = e^{-s_1(t)} + e^{-s_2(t)} \quad (4)$$

as the nonlinear mixing function ( $x_1(t)$  and  $x_2(t)$  are subtracted by their mean in the plot so as to be centered). For investigating what would happen in the case of sparse sources from a geometric point of view, the scatter plots of the observations for the two cases whether the sources are sparse or not, are plotted in Fig. 2. As it is already mentioned, when the sources are sparse, the samples of the source vector are mainly concentrated around the axes because it is quite rare that both of the sources take a non-zero value at the same time. So in this case, the scatter plot of the observations (Fig. 2) contains two manifolds (each of which is the result of the transformation of one of the axes in the source space).

As a consequence, if we consider the points where two sources are active (i.e. nonzero) at the same time as outliers, we can apparently distinguish the two manifolds in the scatter plot of the observation vector. This phenomenon can be also expressed mathematically as follows. let us rewrite (1) as

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \mathbf{f}\left(\begin{bmatrix} s_1(t) \\ s_2(t) \end{bmatrix}\right) = \begin{bmatrix} f_1(s_1(t), s_2(t)) \\ f_2(s_1(t), s_2(t)) \end{bmatrix} \quad (5)$$

where  $f_1(\cdot)$  and  $f_2(\cdot)$  are the components of the mixing matrix. For the times when only  $s_1$  is active we have

$$\begin{aligned} x_1(t) &= f_1(s_1(t), 0) = h_{11}(s_1(t)) \\ x_2(t) &= f_2(s_1(t), 0) = h_{12}(s_1(t)) \end{aligned} \quad (6)$$

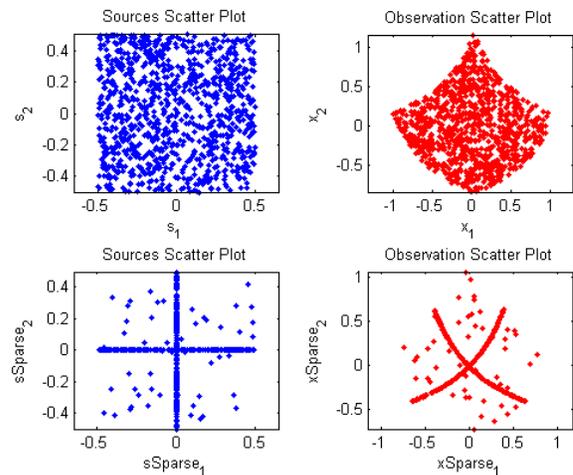


Fig. 2. Comparing scatter plots of the source and observation vectors whether the sources are sparse or not

and when only  $s_2$  is active we have

$$\begin{aligned} x_1(t) &= f_1(0, s_2(t)) = h_{21}(s_2(t)) \\ x_2(t) &= f_2(0, s_2(t)) = h_{22}(s_2(t)) \end{aligned} \quad (7)$$

where  $h_{11}(\cdot)$  and  $h_{21}(\cdot)$  are  $\mathbb{R} \rightarrow \mathbb{R}$  marginal functions of  $f_1(\cdot, \cdot)$ , and  $h_{12}(\cdot)$  and  $h_{22}(\cdot)$  are  $\mathbb{R} \rightarrow \mathbb{R}$  marginal functions of  $f_2(\cdot, \cdot)$ . Since when  $s_1(t)$  is active, both  $x_1(t)$  and  $x_2(t)$  are functions of only one source ( $s_1(t)$ ), they would lie on a 1-dimensional manifold in two dimensional space, and similarly for the samples when only  $s_2(t)$  is active. In other words, when only  $s_1(t)$  is active and at least one of  $h_{11}$  and  $h_{12}$  is invertible (we assume  $h_{11}$  is invertible), we would have

$$x_2(t) = h_{12}(s_1(t)) = h_{12}(h_{11}^{-1}(x_1(t))) = q_1(x_1(t)) \quad (8)$$

and similarly when only  $s_2(t)$  is active and assuming  $h_{21}$  is invertible), we would have

$$x_1(t) = h_{21}(s_2(t)) = h_{21}(h_{21}^{-1}(x_1(t))) = q_2(x_1(t)). \quad (9)$$

Thus  $x_2(t) = q_1(x_1(t))$  and  $x_1(t) = q_2(x_1(t))$  would be the two manifolds that can be clustered in the observation space via some clustering (manifold learning) algorithms. The samples which lie on the manifold of  $q_1$ , are assigned to  $s_1(t)$  and the ones that lie on the manifold of  $q_2$ , are assigned to  $s_2(t)$ .

It should be noted that although the above explanation was based on a simple  $2 \times 2$  mixing system, the argument is still valid for higher dimensions as well. Because for any number of sources and observations, for sparse sources, only one of the sources is active in most of the times. Therefore, each axis in the N-dimensional source space transforms to a 1-dimensional manifold in the M-dimensional observation space.

### B. Step 2 - Unfolding the Manifold

As mentioned before, each manifold in the observation space is a nonlinear transformation of one of the axes in the

source space. The problem of estimating the inverse of this transformation is severely ill-posed. This can also be shown through mathematical expressions. From (6) and (7) we have

$$s_1(t) = \begin{cases} h_{11}^{-1}(x_1(t)) = h_{12}^{-1}(x_2(t)) & \mathbf{x}(t) \in q_1 \\ 0 & \mathbf{x}(t) \in q_2 \\ \text{Unknown} & \text{o.w. (outlier)} \end{cases} \quad (10)$$

$$s_2(t) = \begin{cases} 0 & \mathbf{x}(t) \in q_1 \\ h_{21}^{-1}(x_1(t)) = h_{22}^{-1}(x_2(t)) & \mathbf{x}(t) \in q_2 \\ \text{Unknown} & \text{o.w. (outlier)} \end{cases} \quad (11)$$

while we have information about none of  $h_{ij}(\cdot)$  functions. The “unknown” values mentioned in (10) and (11) are because, generally, we do not have enough information to separate the outliers (when more than one source are simultaneously active) and they actually cause the error in separation. More discussion on this can be found in the following.

In other words, it let  $u_1(\cdot)$  and  $u_2(\cdot)$  be two arbitrary  $\mathbb{R} \rightarrow \mathbb{R}$  invertible functions. So, if instead of  $s_1(t)$  and  $s_2(t)$ ,  $u_1(s_1(t))$  and  $u_2(s_2(t))$  are mixed through the function  $\mathbf{f}(\cdot)$ , (6) and (7) would become

$$\begin{aligned} x_1(t) &= h_{11}(u_1(s_1(t))) = \hat{h}_{11}(s_1(t)) \\ x_2(t) &= h_{12}(u_1(s_1(t))) = \hat{h}_{12}(s_1(t)) \end{aligned} \quad (12)$$

(when only  $s_1$  is active) and

$$\begin{aligned} x_1(t) &= h_{21}(u_2(s_2(t))) = \hat{h}_{21}(s_2(t)) \\ x_2(t) &= h_{22}(u_2(s_2(t))) = \hat{h}_{22}(s_2(t)) \end{aligned} \quad (13)$$

(when only  $s_2$  is active) where  $\hat{h}_{ij} = h_{ij} \circ u_i$  for  $i, j \in \{1, 2\}$  are some other nonlinear functions of the original sources. Consequently, considering (8) and (9) we can say

$$\begin{aligned} x_2(t) &= \hat{h}_{12}(s_1(t)) = \hat{h}_{12} \circ \hat{h}_{11}^{-1}(x_1(t)) \\ &= h_{12} \circ u_1 \circ u_1^{-1} \circ h_{11}^{-1}(x_1(t)) \\ &= h_{12} \circ h_{11}^{-1}(x_1(t)) = q_1(x_1(t)) \end{aligned} \quad (14)$$

(when only  $s_1$  is active) and

$$\begin{aligned} x_2(t) &= \hat{h}_{22}(s_2(t)) = \hat{h}_{22} \circ \hat{h}_{21}^{-1}(x_1(t)) \\ &= h_{22} \circ u_2 \circ u_2^{-1} \circ h_{21}^{-1}(x_1(t)) \\ &= h_{22} \circ h_{21}^{-1}(x_1(t)) = q_2(x_1(t)) \end{aligned} \quad (15)$$

(when only  $s_2$  is active), which means that the  $q_1$  and  $q_2$  manifolds in the observation space remain the same. That is, the scatter plot of the observation vector does not contain any information about the probable nonlinear transformation of the sources. Therefore, the ambiguity of a nonlinear function in reconstructing the sources cannot be resolved.

It is worth noting again that although the above description is made based on a simple  $2 \times 2$  example, all arguments are still valid for the general case. In this case, each of the  $N$  sparse sources (i.e. each of the  $N$  axes in the source space) transforms to a 1-dimensional manifold in the  $M$ -dimensional observation space (which are called as  $q_1, q_2, \dots$  and  $q_N$ ).

It should also be emphasized that the proposed algorithm works even for under-determined cases (where the number of observations is less than the number of source signals). This outstanding capability comes from the fact that the separation of sparse sources is based on separating  $N$  1-dimensional manifolds, and even in a 2-dimensional space we can have any number of different 1-dimensional manifolds. Therefore, the minimum required number of the observation signals, despite of the number of sources, is always 2.

As a result, each of the observations can be taken as the separated sources. So the separated sources in the general form can be reconstructed as

$$\hat{y}_i(t) = \begin{cases} x_j(t) & \mathbf{x}(t) \in q_i \\ 0 & \mathbf{x}(t) \in q_k \ (k \neq i) \\ \text{Unknown} & \text{o.w. (outlier)} \end{cases} \quad (16)$$

where  $\hat{\mathbf{y}}(t) = [\hat{y}_1(t), \hat{y}_2(t), \dots, \hat{y}_N(t)]^T$  is the vector of the separated sources (here we use  $\hat{(\cdot)}$  because it is not the final result; after proposing a method for solving the problem of outliers, the final output is formulated in the following),  $i$  is the index of the separated source which is between 1 and  $N$ ,  $k$  is any index between 1 and  $N$ , and  $1 \leq j \leq M$  is the index of one of the observations (which is chosen arbitrarily). As mentioned earlier, values of the separated signals for outliers (when more than one source is active), is “unknown”, but considering some additional information about either the sources or the mixing model, leads to an estimation for them.

As explained before, the proposed approach is based on the clustering of the manifolds in the observation space (each of which corresponds to one of the sources). Thus, the observation vector lies on a manifold when only one of the sources is active. When two or more sources are active, the observation sample is assumed to be outlier (in the manifold learning process) and is not assigned to any of the sources. This is why the sources can not be estimated in these cases and their values are said to be “unknown” in (10), (11) and (16). Different manners can be suggested to estimate the value of each source for outliers depending on probable prior information about either the sources or the mixing function. The manners can be classified in the two following categories.

The first category of methods for estimating the sources in case of outlier observations, is based on using the sources themselves. In this case, we face the problem of estimating some missing values of a signal while we know it for most of the time. Designing the best restoration technique, one has to employ some known features of the signal (e.g. its bandwidth) if there is any. So the separated sources in this approach would be as

$$y_i(t) = \begin{cases} x_j(t) & \mathbf{x}(t) \in q_i \\ 0 & \mathbf{x}(t) \in q_k \ (k \neq i) \\ \phi(\hat{y}_i(t)) & \text{o.w. (outlier)} \end{cases} \quad (17)$$

where  $\phi(\cdot)$  is the chosen restoration function and  $\hat{y}_i(t)$  comes from (16).

The other class of methods for doing the estimation is based on the mixing function and the separated manifolds.

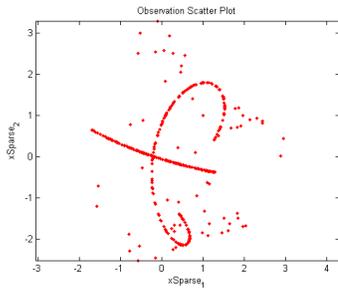


Fig. 3. A manifold whose projections on the axes are not invertible

Using this approach, the value of each source can be estimated through nonlinearly projecting the outliers on the corresponding manifold in the observation space. However, the nonlinear projection is not unique, and would be more accurate if there is some side-information about the mixing model. Nevertheless, the separated sources in this approach can be written as

$$y_i(t) = \begin{cases} x_j(t) & \mathbf{x}(t) \in q_i \\ 0 & \mathbf{x}(t) \in q_k \ (k \neq i) \\ x'_j(t) & \text{o.w. (outlier)} \end{cases} \quad (18)$$

where  $x'_j(t)$  is the chosen nonlinear projection of  $\mathbf{x}(t)$  onto the  $j^{\text{th}}$  coordinate (axis of  $e_j$ ).

However, if there is prior information about the sources and/or the mixing model, it can be utilized to improve the performance of the separating algorithm and trying to reconstruct the sources (resolving the ambiguity of the nonlinear function on each source). For example if we know that the transformation preserves the local angles, if its time-derivative is continuous, if it has a parametric model, and so forth, this may lead to a much better estimation of the sources signals (which is left for the future works).

Finally, it is worth noting that although choosing one of the observations for reconstructing the separated sources is the easiest way, one may find some other wiser ways to make the separated signals. For example if the projection of a manifold ( $q_i$ ) on one of the axes ( $x_j$ ) is not an invertible transformation, that axis ( $x_j$ ) is not a good choice for reconstructing the separated signals (e.g. in Fig. 3, projecting the more complicated manifold on none of the axes would be invertible and would lose information). Indeed in this case, the unresolved nonlinear transformation of the source in the reconstructed signal is not invertible and loses some information of the source.

Therefore, it is better to take an unfolded version of each manifold as the separated source which is made by a dimension reduction algorithm (e.g. ISOMAP [17]). Moreover, one may suggest other methods for unfolding the manifold considering the nature of the sources and the mixing model. For example assuming that the nonlinear mixing function preserves distances or angles, leads to specific unfolding (nonlinear dimension reduction) techniques.

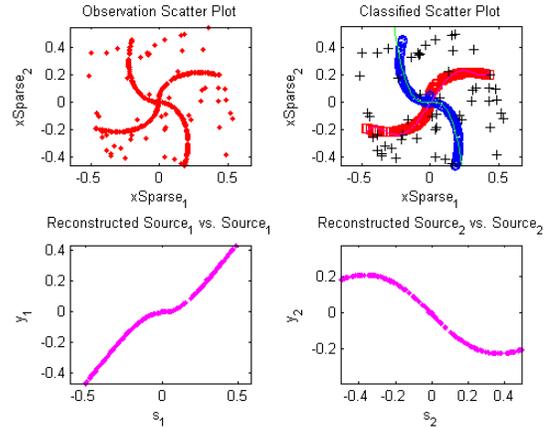


Fig. 4. The result of the algorithm for the model  $x_1(t) = \cos(\alpha(t))s_1(t) - \sin(\alpha(t))s_2(t)$  and  $x_2(t) = \sin(\alpha(t))s_1(t) + \cos(\alpha(t))s_2(t)$  where  $\alpha(t) = \frac{\pi}{2}(1 - \sqrt{s_1^2(t) + s_2^2(t)})^2$

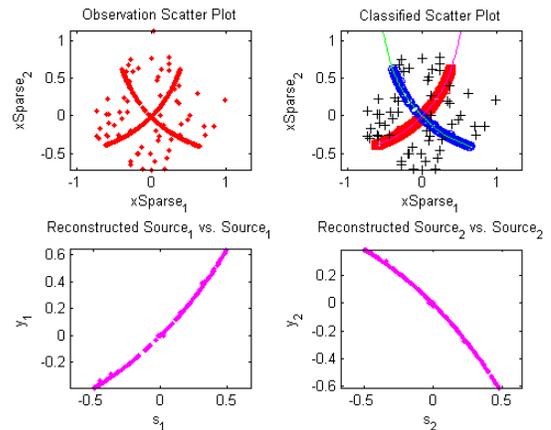


Fig. 5. The result of the algorithm for the model of 3 and 4

### III. SIMULATION RESULTS

In order to simulate the proposed algorithm, we have used several nonlinear mixing models and we have applied them on two sparse sources (with the sparsity of 75%) that are uniformly distributed in  $[-0.5, 0.5]$  when they are active.

In the simulation, firstly outliers are omitted (they are detected by computing their average distance from their few neighbors). Then a simple curve fitting algorithm is applied to perform clustering. It is obvious that in more complicated examples (like Fig. 3), one should employ more powerful clustering techniques and might use kernels.

In each of the figures from Fig. 4 to 6, the scatter plot of the observations and the clustering result are shown in the above plots (where each cluster is indicated with a different color and the outliers are marked as "+"). The two plots below of each figure represent the relation between each of outputs and the corresponding source signal (which is, as expected,

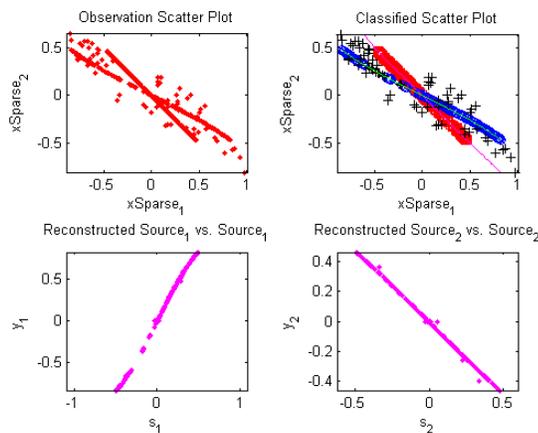


Fig. 6. The result of the algorithm for the model  $x_1(t) = \sin(2s_1(t) - s_2(t))$  and  $x_2(t) = \sin(s_2(t) - s_1(t))$

a nonlinear function that can not be resolved without further information).

#### IV. DISCUSSION AND FUTURE WORKS

In this work, we proposed a nonlinear BSS algorithm for sparse sources. The algorithm works even if the problem is under-determined hence the number of observations is less than the number of sources (it only needs more than one observation). We also showed that the sources can be separated but they cannot be exactly reconstructed in the most general case without further information.

The proposed algorithm is also useful when the number of source signals is unknown. It is due to two facts: 1) despite of the number of sources, if they are spatially sparse, the separation can be performed as soon as we have more than one observation and 2) if the manifold learning technique which is used for clustering the observation is powerful enough (and does not need the number of clusters in advance), any number of manifolds can be separated in the observation space and hence the generalization of the proposed method would be straightforward.

In addition, the sparsity feature used in the proposed algorithm is satisfied even if each source signal takes a constant value for most of the time (not necessarily zero) and has sparse variations. In this case, all mentioned arguments remain valid but in (6) and (7),  $f_j(s_1(t), 0)$  and  $f_j(0, s_2(t))$  need to be replaced with  $f_j(s_1(t), c_2)$  and  $f_j(c_1, s_2(t))$  for  $j \in \{1, 2\}$ , where  $c_1$  and  $c_2$  are respectively the constant values that  $s_1(t)$  and  $s_2(t)$  are assumed to take most of the time.

For the future work, it would be interesting to develop the proposed framework for the sources that are not sparse in time domain, but in some other domain like frequency domain. In these cases, one has to firstly model the transformation to the sparse domain for both sources and observations in order to be able to cluster the observation from the manifolds in the sparse domain and then separate the observations using the classification which is done.

Moreover, it would be useful to apply the proposed approach to practical applications and to utilize the prior information (related to the real case) for better unfolding the separated manifolds and more precisely estimating the values corresponding to the outliers. It might also need a discussion on different choices of  $\phi(\cdot)$  in (17) or the nonlinear projection method in (18).

#### ACKNOWLEDGEMENT

This work is partly funded by the European project 2012-ERC-Adv-320684 CHESS.

#### REFERENCES

- [1] J. Héroult and C. Jutten, "Space or time adaptive signal processing by neural networks models," in *Intern. Conf. on Neural Networks for Computing*, Snowbird (Utah, USA), 1986, pp. 206–211.
- [2] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent component analysis and applications*. Academic press, 2010.
- [3] A. Cichocki, R. Zdunek, and S.-i. Amari, "New algorithms for non-negative matrix factorization in applications to blind source separation," in *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, vol. 5. IEEE, 2006, pp. V–V.
- [4] P. Comon, "Independent component analysis, a new concept?" *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.
- [5] L. Parra and C. Spence, "Convolutional blind separation of non-stationary sources," *Speech and Audio Processing, IEEE Transactions on*, vol. 8, no. 3, pp. 320–327, 2000.
- [6] H. Buchner, R. Aichner, and W. Kellermann, "Blind source separation for convolutional mixtures exploiting nongaussianity, nonwhiteness, and nonstationarity," in *Conf. Rec. of the Seventh International Workshop on Acoustic Echo and Noise Control (IWAENC 03)*. Citeseer, 2003.
- [7] T. Mei, F. Yin, and J. Wang, "Blind source separation based on cumulants with time and frequency non-properties," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 17, no. 6, pp. 1099–1108, 2009.
- [8] Y. Deville and L. T. Duarte, "An overview of blind source separation methods for linear-quadratic and post-nonlinear mixtures," in *Latent Variable Analysis and Signal Separation*. Springer, 2015, pp. 155–167.
- [9] S. Hosseini and C. Jutten, "On the separability of nonlinear mixtures of temporally correlated sources," *IEEE signal processing letters*, vol. 10, no. 2, pp. 43–46, February 2003.
- [10] M. Babaie-Zadeh, "On blind source separation in convolutional and nonlinear mixtures," Ph.D. dissertation, INP Grenoble, 2002.
- [11] B. Ehsandoust, M. Babaie-Zadeh, and C. Jutten, "Blind source separation in nonlinear mixture for colored sources using signal derivatives," in *Latent Variable Analysis and Signal Separation*. Springer, 2015, pp. 193–200.
- [12] D. N. Levin, "Performing nonlinear blind source separation with signal invariants," *Signal Processing, IEEE Transactions on*, vol. 58, no. 4, pp. 2131–2140, 2010.
- [13] A. Taleb and C. Jutten, "Source separation in post nonlinear mixtures," *IEEE Transactions on Signal Processing*, vol. 47, no. 10, pp. 2807–2820, 1999.
- [14] F. Merrikh-Bayat, M. Babaie-Zadeh, and C. Jutten, "Linear-quadratic blind source separating structure for removing show-through in scanned documents," *International Journal on Document Analysis and Recognition (IJ DAR)*, vol. 14, no. 4, pp. 319–333, 2011.
- [15] A. Hyvärinen and P. Pajunen, "Nonlinear independent component analysis: Existence and uniqueness results," *Neural Networks*, vol. 12, pp. 429–439, 1999.
- [16] P. Bofill and M. Zibulevsky, "Underdetermined blind source separation using sparse representations," *Signal Processing*, vol. 81, pp. 2353–2362, 2001. [Online]. Available: cite-seer.ist.psu.edu/bofill01underdetermined.html
- [17] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *science*, vol. 290, no. 5500, pp. 2319–2323, 2000.