

Nonlinear Blind Mixture Identification Using Local Source Sparsity and Functional Data Clustering

Matthieu Puigt*, Anthony Griffin*, and Athanasios Mouchtaris*[†]

*FORTH-ICS, Heraklion, Crete, Greece, GR-70013

Email: {mpuigt,agriffin,mouchtar}@ics.forth.gr

[†]University of Crete, Department of Computer Science, Heraklion, Crete, Greece, GR-71409

Abstract—In this paper we propose several methods, using the same structure but with different criteria, for estimating the nonlinearities in nonlinear source separation. In particular and contrary to the state-of-art methods, our proposed approach uses a weak joint-sparsity sources assumption: we look for tiny temporal zones where only one source is active. This method is well suited to non-stationary signals such as speech. We extend our previous work to a more general class of nonlinear mixtures, proposing several nonlinear single-source confidence measures and several functional clustering techniques. Such approaches may be seen as extensions of linear instantaneous sparse component analysis to nonlinear mixtures. Experiments demonstrate the effectiveness and relevancy of this approach.

I. INTRODUCTION

Blind Source Separation (BSS) consists of estimating a set of N unknown source signals s_j from a set of P observations x_i resulting from mixtures of these sources through unknown propagation channels [1]. Among all the proposed approaches, the ones based on source sparsity, known under the name of Sparse Component Analysis (SCA) methods, have met with great interest from the community in the last decade (see e.g. [1, Ch. 10]). Indeed, they are naturally adapted to stationary, non-stationary and/or dependent signals and are thus an alternative to classical Independent Component Analysis (ICA) approaches which assume source mutual independence. Moreover, they allow processing of the underdetermined case where $N > P$.

Most of the SCA approaches have been proposed for linear mixtures, i.e. linear instantaneous (LI) or convolutive mixtures. While many methods assume the sources to be (approximately) W-disjoint orthogonal (WDO) in an analysis domain¹ [2], several other ones highly relax this assumption, by looking for “single-source zones” (i.e. zones where one source is dominant over the others) [3], [4]. Interestingly, while many SCA methods have been proposed for linear mixtures, only a few sparsity-based approaches process nonlinear (NL) configurations [5]–[9]. The work in [5] and [6] required the sources to be approximately WDO² and processed post-nonlinear (PNL) mixtures (i.e. a special configuration where linear mixes of sources are distorted by a componentwise

function which models data acquisition/sensor nonlinearities, such as saturation). [7]–[9] extended the measures for finding single-source zones to particular NL mixtures (respectively bilinear, following the Nicolsky-Eseinman model, and PNL).

In this paper we propose a general approach, using single-source zones as in [7]–[9], which estimates the nonlinear functions in mixtures and which possibly processes the underdetermined case. It may be seen as a generalization of [9] to a wider class of NL mixtures. Our proposed method keeps the same structure as [9], but we additionally introduce several single-source confidence measures and different functional data clustering approaches, thus yielding a way to extend linear SCA [3], [4] to NL-SCA.

II. PROBLEM STATEMENT

In this paper, we assume that N real source signals $\underline{s}(t) = [s_1(t), \dots, s_N(t)]^T$ are mixed by an unknown instantaneous nonlinear mapping \mathcal{A} from $\mathbb{R}^N \rightarrow \mathbb{R}^P$. Observed signals then read:

$$\underline{x}(t) = \mathcal{A}(\underline{s}(t)). \quad (1)$$

This mixing model is extremely general and it is well known [1, Ch. 14, Sect. 14.2.3] that it cannot be solved by *only* assuming source mutual independence³, and additional definitions and assumptions are needed.

Definition 1: A “temporal analysis zone” is a subset \mathcal{T} of the time domain⁴.

From a theoretical point of view, each temporal analysis zone may be set to any kind of subset of the full time domain. However, in practice, we set these zones to intervals.

Definition 2: A source is said to be “isolated” in a temporal analysis zone \mathcal{T} if only this source (among all considered mixed sources) has a nonzero variance in this zone. We then say that this zone is “single-source”.

This definition corresponds to the theoretical point of view. From a practical point of view, this means that the variances of all other sources are negligible with respect to the variance of the source that is isolated.

¹The WDO assumption means that in each atom of an analysis domain (e.g. time, time-frequency, time-scale domain), at most one source is non-zero.

²Actually, in [5], the authors assume the sources to be $(P - 1)$ -sparse, which is equivalent to WDO if $P = 2$. In [6], the approximate WDO is not explicitly assumed but is needed by authors and satisfied in their tests.

³ [1, Ch. 14, Sect. 14.2.3] even provides a way to design an invertible NL operator \mathcal{B} which preserves the mutual independence of the output signals $\underline{y}(t) = \mathcal{B}(\underline{x}(t)) = \mathcal{B} \circ \mathcal{A}(\underline{s}(t))$ while still being mixing.

⁴Actually, our approach is not limited to audio signals, but we refer to time as many LI-SCA approaches usually sparsify the observations with a time-frequency transform.

Definition 3: A source is said to be “accessible” in the time domain if there exist at least one temporal analysis zone where it is isolated.

Assumption 1: (i) The nonlinear mapping \mathcal{A} is smooth, (ii) we assume we know the value of \mathcal{A} for one value $\underline{u}_0 \in \mathbb{R}^N$, and in particular, without loss of generality, we assume that $\underline{u}_0 = \underline{0}$ and that

$$\mathcal{A}(\underline{0}) = 0. \quad (2)$$

Lastly, (iii) \mathcal{A} may be completely estimated by its values in single-source zones \mathcal{T} :

$$x_i(t) = \mathcal{A}_i(\underline{s}(t)) \triangleq \mathcal{A}_{ik}(s_k(t)), \forall i \in \{1, \dots, P\}, \quad (3)$$

where \mathcal{A}_{ik} is an invertible nonlinear function from $\mathbb{R} \rightarrow \mathbb{R}$. Assumptions 1.(i) and 1.(iii) are needed in order to interpolate \mathcal{A} from \mathcal{A}_{ik} . Assumption 1.(ii) is needed to suppress the ambiguities that may appear in the selection of single-source zones, as we will see in Section III-A. Note that Assumption 1.(iii) allows us to tackle many NL configurations, as we will now see. The PNL mixture model e.g. satisfies this assumption: the nonlinear mapping \mathcal{A} may be rewritten as the composition

$$\mathcal{A} = \underline{f} \circ A, \quad (4)$$

where \underline{f} and A model each part of the PNL mixture, i.e. \underline{f} is a NL mapping from $\mathbb{R}^P \rightarrow \mathbb{R}^P$ and A is a $N \times P$ mixing matrix, and where \circ is the composition sign. Observations then read

$$\underline{x}(t) = \underline{f}(A\underline{s}(t)), \quad (5)$$

and the values of $\underline{x}(t)$ at times t when sources are isolated allow the estimation of both \underline{f} and A [5], [6], [9]. Assumption 1.(iii) also allows us to process the situation when each NL function \mathcal{A}_i defined in Eq. (3) is written as a linear combination of NL functions \mathcal{A}_{ij} defined from $\mathbb{R} \rightarrow \mathbb{R}$:

$$\mathcal{A}_i(\underline{s}(t)) = \sum_{j=1}^N \mathcal{A}_{ij}(s_j(t)). \quad (6)$$

In a general way, Assumption 1.(iii) allows us to estimate mappings \mathcal{A}_i that can be inferred from the functions \mathcal{A}_{ik} defined in Eq. (3).

Assumption 2: (i) Source signals are mutually independent and (ii) by considering *several* single-source analysis zones associated with the same source, the amplitude of the observations spans a “wide” range allowing the estimation of the NL functions \mathcal{A}_{ik} .

Note that, contrary to linear SCA methods which can process correlated sources [1, ch. 10], here we need source mutual independence. This is due to the more complex mixing model, as we will see in Section III-A. Assumption 2.(ii) is needed because we want to estimate the nonlinear mappings \mathcal{A}_i on their whole domains. In the case where we should be able to estimate the functions \mathcal{A}_i on a subset of their domain, the whole estimation might be coarse, thus yielding a poor separation quality.

In this paper, our main contribution is dedicated to the estimation of the NL mapping. Recovering the sources is not

straightforward in general and needs extra-assumptions on the mixture model or on the sources (e.g. the sources take a finite set of possible values) [1, Ch. 14]. The only case for which the separation is quite easy is that of the PNL mixture, for which a solution consists of linearizing the observations [5], [6] and of then applying LI-SCA to the newly obtained observations.

The proposed mixture identification structure may be summarized as follows:

- 1) We first look for single-source analysis zones (See Section III-A).
- 2) We then estimate the nonlinear mappings \mathcal{A}_i (See Section III-B).

III. PROPOSED APPROACH

A. Nonlinear single-source confidence measures

We now detail how to find single-source zones in NL mixtures. If a source, say s_k is isolated, then Eq. (3) holds and we then obtain, assuming that \mathcal{A}_{ik} is invertible,

$$s_k(t) = \mathcal{A}_{ik}^{-1}(x_i(t)) \quad \forall i \in \{1, \dots, P\}, \forall t \in \mathcal{T}. \quad (7)$$

We thus have the following relationship between observations x_1 and x_i , for all $t \in \mathcal{T}$:

$$x_i(n) = \mathcal{A}_{ik}(\mathcal{A}_{1k}^{-1}(x_1(t))) = \phi_{ik}(x_1(t)), \quad (8)$$

where the functions ϕ_{ik} are defined as:

$$\phi_{ik}(u) = \mathcal{A}_{ik}(\mathcal{A}_{1k}^{-1}(u)). \quad (9)$$

In an LI problem, the relationships (8) between observations are much simpler. Let us first recall that in that case, the mappings \mathcal{A}_{ik} in Eq. (3) are set to scale coefficients a_{ik} . In the case when a source, say s_k is isolated, then observations read:

$$x_i(t) = a_{ik}s_k(t) \quad \forall i \in \{1, \dots, P\}, \forall t \in \mathcal{T} \quad (10)$$

and we get the following relationship between observations x_1 and x_i :

$$a_{ik}x_1(t) - a_{1k}x_i(t) = 0 \quad \forall i \in \{1, \dots, P\}, \forall t \in \mathcal{T}. \quad (11)$$

In [3], [4], authors proposed finding these single-source zones by means of a “single-source confidence measure” (resp. based on correlation between pairs of observations [3], and local PCA [4]). In this paper, we introduce several methods to extend the confidence measures proposed in [3], [4] to NL mixtures.

The work of [3] proposed estimating the correlation between pairs of observations to find single-source zones. Indeed, this coefficient is equal to ± 1 when one source is isolated and is much lower otherwise. In the considered NL mixture (1), we need to measure a *nonlinear* correlation between observations. This is e.g. provided by the normalized mutual information $\mathcal{I}_{\text{norm}}(\underline{x})$, defined as [8], [9]:

$$\mathcal{I}_{\text{norm}}(\underline{x}) = \sqrt{1 - e^{-2\mathcal{I}(\underline{x})}}, \quad (12)$$

where $\mathcal{I}(\underline{x})$ stands for the mutual information between observations. When normalized, its behavior is similar to that

of a correlation and we consider the analysis zones \mathcal{T} which maximize Eq. (12) as single-source.

Alternatively, as we assume the NL mappings \mathcal{A}_{ik} to be smooth, the resulting functions ϕ_{ik} are also smooth and one may *locally* consider them as linear. Such an idea is quite classical in manifold learning [10] and allows us to extend linear single-source approaches to NL configurations. Several approaches, like LTSA [10, Sect. 3.2.4], learn the manifold by constructing the local tangent space of each observed data point. We propose using such an idea to extend linear single-source confidence measures to NL mixtures. Our approach consists of successively considering each sample $\underline{x}(t_i)$ of the analysis zone \mathcal{T} , of defining its neighborhood (estimated by means of a K nearest neighbor (K -NN) technique), and of applying in this neighborhood a linear single-source confidence measure. As in [4], we e.g. realize an eigendecomposition of the correlation matrix of the data. If one source is isolated, then the rank of the observations is set to 1 and the highest eigenvalue $\lambda_1(t_i)$ is non-negligible while the $K-1$ other ones $\lambda_j(t_i)$ are close to zero. Authors of [4] proposed computing the ratio between the highest eigenvalue and the sum of the others to find single-source zones. In this paper, in order to keep an analogy with the behavior of the correlation, we propose computing the ratio:

$$R(t_i) = \frac{\lambda_1(t_i)}{\sum_{j=1}^K \lambda_j(t_i)}, \quad (13)$$

which is close to 1 in single-source zones and much lower otherwise. Once we compute the ratios $R(t_i)$ for all the data points in \mathcal{T} , we derive the actual *global* single-source confidence measure, denoted $\mathcal{R}(\underline{x})$ hereafter, as the geometric mean of all these ratios. Note that applying other single-source measures like [7] is straightforward and we will denote by $\mathcal{C}(\underline{x})$ the *global* correlation coefficient computed as the geometrical mean of local correlation coefficients estimated in tangent subspaces of the data points $\underline{x}(t)$.

Lastly, there is an issue with both of the above NL confidence measures if in a zone \mathcal{T} , unactive sources $s_j(t)$ are constant but non-null⁵. Such a scenario is not a problem in LI-SCA: observations can be locally centered in each analysis zone, thus zeroing the constant signals [3]. The same solution is possible for some special forms of \mathcal{A} , as e.g. in Eq. (6). But without additional a priori information about the mappings, we need to provide a general solution to this problem: due to Assumption 1.(ii), the value of each nonlinear function \mathcal{A}_i is zero at zero and we can estimate $\widehat{\phi}_{ik}$, the nonlinear relationship between observations defined in Eq. (9) (see Section III-B) and discard the zones where $\widehat{\phi}_{ik}(0) \neq 0$.

Finally, we look for analysis zones which (i) maximize the chosen single-source confidence measure and (ii) which satisfy⁶ $\widehat{\phi}_{ik}(0) = 0$.

⁵Note that contrary to our proposed method, the authors of [8] are looking for zones where all the unactive sources are strictly positive constant.

⁶In practice, we look for zones where the considered single-source confidence measure is above a value $1 - \epsilon_1$ and we only keep the zones such that $|\widehat{\phi}_{ik}(0)| < \epsilon_2$, where ϵ_1 and ϵ_2 are user-defined thresholds.

B. Functional data clustering

Once we have estimated all the single-source analysis zones, we get a subset of the original observations for which, in each zone, a scattered representation of a function ϕ_{ik} is provided. As several zones may lead to the same source, it then makes sense to cluster these data in order to obtain quite dense representations of the functions ϕ_{ik} to be inferred.

Many approaches for clustering functional data belong to one of the two following families. (i) The *regularization* approaches consist of successively interpolating each observed scattered function, of discretizing all of them on the same time grid and lastly clustering them while considered as a high dimensional vectors. However, they are often highly correlated and may lead to unstable solutions, because of the *curse of dimensionality*. (ii) The *filtering* methods consist of approximating each curve with respect to a common finite dimensional basis (B_1, \dots, B_K), e.g. a B-spline basis, and then of clustering the resulting basis coefficients with e.g. classical clustering techniques such as K-means [11] or its median-based extension K-medians [9]. The main difficulty then consists of choosing the “good” basis, with good parameters (e.g. in the case of a B-spline basis, choosing the degree of the spline and the knot locations) [9], [11].

Alternatively, we propose a clustering approach based on the locally linear data approximation. In particular, we estimate this approximation around 0. Indeed, the first order of the Taylor expansion of the functions ϕ_{ik} reads

$$\phi_{ik}(t) = \phi_{ik}(0) + \phi'_{ik}(0) \cdot t + \mathcal{O}(t^2). \quad (14)$$

Eq. (14) thus reveals that the scattered curve is approximated by the slope of its tangent at zero and we can use this slope as a way to cluster the estimated functions. Estimating $\phi'_{ik}(0)$ may be done by LI-SCA, as Eq. (14) is equivalent to Eq. (11). As in Section III-A, we propose using the formalism of manifold learning, and estimating the neighborhood of 0 with the K -NN method, before applying one of the methods in [4], [7] to estimate the slope of $\phi'_{ik}(0)$. We lastly cluster the curves by applying K-medians on these slopes.

IV. PERFORMANCE OF THE PROPOSED APPROACH

In this section, we test the performance of our proposed approaches. As they share the same stages but with different criteria, we propose illustrating separately the influence of each of these criteria. We first investigate the accuracy of the single-source confidence measure in one toy example. We generate two source signals of 20000 samples: one is a white noise realization while the other is another white noise realization multiplied by a gain coefficient ranging from 0 to 1 along the signal length. We then mix them in two observations according to Eq. (6) with $\mathcal{A}_{11}(t) = \mathcal{A}_{22}(t) = \tanh(t)$, $\mathcal{A}_{12}(t) = 0.5t$, and $\mathcal{A}_{21}(t) = 3t$. We thus have one unique isolated source in the first zones and the measures should quickly decrease with the zone index. As the locally-linear approaches use a K -NN method for estimating the linear subspace, we vary K to $\{5, 10, 15, 20\}$. Figure 1 shows the belief of the proposed single-source confidence measures, i.e. $\mathcal{I}_{\text{norm}}(\underline{x})$, estimated

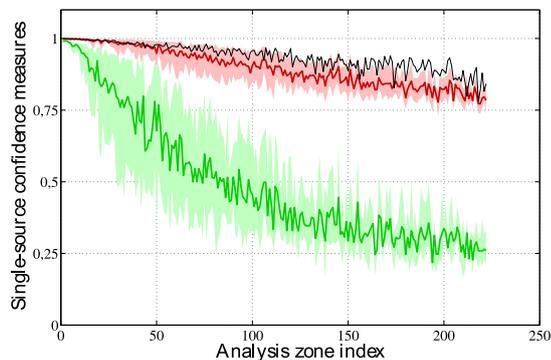


Fig. 1. Values of the different single-source confidence measures for a pair of observations. Black: Normalized mutual information. Red: Ratio of eigenvalues in tangent spaces. Green: Correlation coefficient in tangent spaces.

as in [9] (black line), the mean and the envelope (obtained from the experiments with different values of K) of $\mathcal{R}(\underline{x})$ (red curves) and $\mathcal{C}(\underline{x})$ (green curves). All the proposed single-source confidence measure seem to be well suited to the considered problem. In the single-source zones, they are all extremely close to 1 and both $\mathcal{R}(\underline{x})$ and $\mathcal{C}(\underline{x})$ do not seem sensitive to the value of K . However, in non single-source zones, we notice different behaviours: both $\mathcal{I}_{\text{norm}}(\underline{x})$ and $\mathcal{R}(\underline{x})$ provide quite high values, i.e. not too far from 1, while $\mathcal{C}(\underline{x})$ provide highly decreasing values, thus indicating that it seems the most discriminant measure.

We now test the different functional clustering approaches we discussed in Section III-B. We perform the same series of tests as in [9] for PNL mixtures of audio sources. The nonlinearities of the model simulate the saturation in audio recordings due to small size of microphones e.g. contained in mobile devices. We consider 28 pairs of $N = 2$ speech sources that we mix with the 2×2 matrix $A = [a_{ij}] = [(-1)^j \gamma^{|i-j|}]$ and to which the following NL mappings $f_1(t) = \tanh(t) + t$ and $f_2(t) = \tanh(10t)$ are applied. In all the tests, we detect the single-source zones by using the normalized mutual information $\mathcal{I}_{\text{norm}}(\underline{x})$ as was performed in [9]. We then apply one of the above functional clustering approaches and we provide the mean square error (MSE) between the interpolations of the clustered curves and the theoretical ones in Table I. For the same reasons as in the previous experiment, we defined the K -neighborhood around 0 with several values of K : $\{5, 10, 15, 20\}$. The values inside brackets in Table I give the performance when $K = 5$ while the experiments with the other tested values of K yielded the same MSEs and are shown outside the brackets. All the methods provide very accurate and similar results (when $K \geq 10$), hence showing their relevance. However, the filtering clustering method contains heuristics difficult to set in practice: we must define some common knot locations for all the scattered curves, with some constraints on the number of data points $\underline{x}(t)$ between knots, and a spline degree. Choosing a “good” number of neighbors with the tangent space-based approaches is much simpler. Moreover, the K -NN method was not sensitive to the choice

TABLE I
PERFORMANCE OF THE FUNCTIONAL CLUSTERING METHODS

Method	γ		
	0.1	0.5	0.9
Filtering (MSE) clust. [9] (Std.)	6.7e-2 2.3e-1	4.5e-5 5.9e-5	4.6e-5 1.0e-4
linear approx. (MSE) using [4] (Std.)	4.8e-2 (6.9e-2) 2.4e-1 (2.8e-1)	4.1e-5 (9.0e-3) 5.6e-5 (3.1e-2)	4.6e-5 (1.6e-2) 1.0e-4 (3.6e-2)
linear approx. (MSE) using [7] (Std.)	4.8e-2 (4.6e-2) 2.4e-1 (2.3e-1)	4.1e-5 (8.2e-3) 5.6e-5 (2.9e-2)	4.6e-5 (1.6e-2) 1.0e-4 (3.6e-2)

of K , provided it was “big enough”.

V. CONCLUSION

In this paper we proposed a new method to estimate mappings which hold in non-linear source separation using a weak sparsity assumption (finding zones where only one source is active) and functional clustering. The inversion of the estimated mappings is usually not straightforward, except in some special configurations that we discussed. We investigated several single-source confidence measures, and specifically some measures inspired by manifold learning. We also studied several functional clustering methods and validated their performance with simulations. In the future, we will give a more detailed analysis of the performance of each of the presented criteria. We will also investigate the NL inversion and the estimation of NL mappings with memory effect.

ACKNOWLEDGMENT

This work is funded by the Marie Curie IAPP “AVID MODE” grant within the European Commission’s FP7.

REFERENCES

- [1] P. Comon and C. Jutten, *Handbook of blind source separation. Independent component analysis and applications*. Academic press, 2010
- [2] O. Yilmaz and S. Rickard, “Blind separation of speech mixtures via time-frequency masking”, *IEEE Trans. on Signal Processing*, Vol. 52, No. 7, pp. 1830–1847, 2004.
- [3] Y. Deville and M. Puigt, “Temporal and time-frequency correlation-based blind source separation methods. Part I: determined and underdetermined linear instantaneous mixtures”, *Signal Processing*, Vol. 87, No. 3, pp. 374–407, Mar. 2007.
- [4] S. Arberet, R. Gribonval, and F. Bimbot, “A robust method to count and locate audio sources in a multichannel underdetermined mixture”, *IEEE Trans. on Signal Processing*, Vol. 58, No. 1, Jan. 2010.
- [5] F. Theis and S. Amari, “Postnonlinear overcomplete blind source separation using sparse sources,” in *Proc. of ICA, LNCS*, vol. 3195, pp. 718–725, Granada, Spain, 2004.
- [6] S. van Vaerenbergh and I. Santamaría, “A spectral clustering approach to underdetermined postnonlinear blind source separation of sparse sources”, *IEEE Trans. on Neural Networks*, vol. 17, No. 3, pp. 811–814, 2006.
- [7] Y. Deville and S. Hosseini, “Blind identification and separation methods for linear-quadratic mixtures and/or linearly independent non-stationary signals,” in *Proc. of ISSPA*, Sharjah, United Arab Emirates, 2007.
- [8] L. T. Duarte and C. Jutten, “A nonlinear source separation approach for the Nicolsky-Eisenman model,” in *Proc. of EUSIPCO*, Lausanne, Switzerland, 2008.
- [9] M. Puigt, A. Griffin, and A. Mouchtaris, “Post-nonlinear speech mixture identification using single-source temporal zones & curve clustering,” in *Proc. of EUSIPCO*, Barcelona, Spain, 2011.
- [10] L.J.P. van der Maaten, E.O. Postma, and H.J. van den Herik, “Dimensionality reduction: a comparative review”, *Tilburg University Technical Report*, 2009.
- [11] C. Abraham, P. A. Cornillon, E. Matzner-Løber, and N. Molinari, “Unsupervised curve clustering using B-splines”, *Scandinavian Journal of Statistics*, Vol. 30, No. 3, pp. 581–595, Sep. 2003.