Blind Source Separation in Nonlinear Mixtures: Separability and a Basic Algorithm

Bahram Ehsandoust*†, Massoud Babaie-Zadeh*, Bertrand Rivet†, Christian Jutten†
* Sharif University of Technology, Electrical Engineering Department, Tehran, IRAN
ehsandoust@ee.sharif.edu and mbzadeh@sharif.edu
† University Grenoble Alpes, CNRS, GIPSA-lab, Grenoble, France
{bahram.ehsandoust,bertrand.rivet,christian.jutten}@gipsa-lab.grenoble-inp.fr

Abstract—In this paper, a novel approach for performing Blind Source Separation (BSS) in nonlinear mixtures is proposed, and their separability is studied. It is shown that this problem can be solved under a few assumptions, which are satisfied in most practical applications. The main idea can be considered as transforming a time-invariant nonlinear BSS problem to local linear ones varying along the time, using the derivatives of both sources and observations.

Taking into account the proposed idea, numerous algorithms can be developed performing the separation. In this regard, an algorithm, supported by simulation results, is also proposed in this paper. It can be seen that the algorithm well separates the mixed sources, however, as the conventional linear BSS methods, the nonlinear BSS suffers from ambiguities, which are discussed in the paper.

Index Terms—Blind Source Separation, Nonlinear Mixtures, Nonlinear Regression, Independent Component Analysis

I. INTRODUCTION

The Blind Source Separation (BSS) problem was firstly introduced in 1980's [1], [2], and since then, it has been thoroughly studied in the signal processing community. Roughly speaking, in this problem there are a number of source signals that are mixed in some way to make a number (probably not the same number as the sources) of observation signals. The goal is to reconstruct the sources having access only to the observations, i.e. knowing neither the sources nor the mixing model.

BSS problem is formally described as follows. At each time (more generally, sample) t let us consider m observations $x_i(t)$, $i = 1, \ldots, m$, which are unknown time-invariant functions $f_i(\cdot)$ of unknown sources $s_j(t)$, $j = 1, \ldots, n$. For $t = 1, \ldots, T$ measurements, we can write the model as

$$\mathbf{x}(t) = \mathbf{f}(\mathbf{s}(t)), \qquad t = 1, \ldots, T \qquad (1)$$

where $\mathbf{x}(t) = [x_1(t), ..., x_m(t)]^{\mathcal{T}}$ ($\mathcal{T}$ stands for matrix transposition) and $\mathbf{s}(t) = [s_1(t), ..., s_n(t)]^{\mathcal{T}}$ represent the observation and source vectors, respectively, and $\mathbf{f}(\cdot)$ denotes a function from $\mathbb{R}^n$ to $\mathbb{R}^m$.

The problem is generally ill-posed, but it has been shown that assuming some particular structure of $\mathbf{f}$, and/or statistical properties of the sources, it can be solved to some extent and the sources can be reconstructed with ambiguities in their

amplitude and their order. The book [2] provides a comprehensive survey on different structures and proposed algorithms. The key idea to perform separation is trying to recover some characteristics of the sources by estimating a mapping on the observations able to inverse $\mathbf{f}$. Mostly the characteristics are at least one of the "non-properties" (a word borrowed from [3]); e.g. non-dependence (independence), non-Gaussianity, non-stationarity, non-whiteness and non-negativity.

A. Background

The simplest form of the problem is when the mixture model is instantaneous *linear* and the number of the sources is equal to the number of the observations so that (1) becomes $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$ where $\mathbf{A}$ is an unknown mixing matrix. The earliest approach to this case was in [1], [4] which introduced the concept of Independent Component Analysis (ICA). The independence employed in ICA is in the sense of random variables assuming that each source consists of Independent and Identically Distributed (iid) samples, i.e. without taking care of the sample order.

It should be recalled that if two random variables are mutually independent, the joint probability density function (pdf) of them factorizes as the product of their marginal pdf's. On the other hand, two stochastic processes are said to be mutually independent iff they are mutually independent for any sequence of time instants.

Accordingly, the two notions: random variable (RV) independence and stochastic process (SP) independence, should be distinguished. For linear instantaneous mixtures, a very nice result is that signal separation can be achieved if the sources $s_i(t)$ and $s_j(t)$, for any pair $i \neq j$, are mutually independent random variables [4]. It is thus outstanding to note that SP independence is not required in the linear case.

Many algorithms have been designed based on different approximations of RV independence, e.g. CoM2 [4], INFOMAX [5], JADE [6], Normalized EASI [7], HOSVD [8], FastICA [9], and finally AMUSE [10], [11] and SOBI [12] (which exploit the assumption that the source samples are not iid, and consider the statistical independence of delayed samples). Afterwards, taking into account any of the mentioned "non-properties", any combination of them, or even some other characteristics such as sparsity, other separation algorithms have been proposed [2].

Nonetheless, in many applications the mixing model of the sources has to be modeled as nonlinear. Hyperspectral imaging [13], [14], remote sensing data [15], determining the concentration of different ions in a combination via smart chemical sensor arrays [16], and removing show-through in scanned documents [17] are some well-studied examples of such applications. However, in contrast to linear BSS, no general theoretical results on identifiability and separability have been provided for BSS in nonlinear mixtures so far.

B. ICA in Nonlinear Mixtures

Although for linear mixtures, conventional ICA (i.e. based on RV independence) ensures identifiability and separability even for iid sources, it is not sufficient for nonlinear mixtures. In other words, one can find some nonlinear mixtures (with non-diagonal Jacobian) of mutually independent sources which are still mutually independent. In this subsection it is shown by a counter-example why RV-based ICA does not work for nonlinear BSS.

In [18, Section 3.3], it is shown that even for smooth nonlinear mixing functions, source independence (in the sense of random variables) is not a powerful enough criterion for separating the sources. In the following example, at each sample t, the sources are mixed nonlinearly as

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} \cos\alpha(\mathbf{s}(t)) & -\sin\alpha(\mathbf{s}(t)) \\ \sin\alpha(\mathbf{s}(t)) & \cos\alpha(\mathbf{s}(t)) \end{bmatrix} \begin{bmatrix} s_1(t) \\ s_2(t) \end{bmatrix} \quad (2)$$

where $\alpha(\mathbf{s}(t))$ is a differentiable function. In this particular example the determinant of the Jacobian matrix of the nonlinear transformation always equals to one, hence

$$\rho_{X_1,X_2}(x_1, x_2) = \frac{1}{|\det(\mathbf{J_f}(\mathbf{s}))|}\rho_{S_1,S_2}(s_1, s_2)$$
$$= \rho_{S_1,S_2}(s_1, s_2). \quad (3)$$

Particularly, if the source samples are iid and uniformly distributed between -1 and 1, i.e. $\rho_{S_1,S_2}(s_1, s_2) = 0.25$ for $(s_1, s_2) \in [-1, 1] \times [-1, 1]$ and 0 elsewhere and given

$$\alpha(\mathbf{s}(t)) = \begin{cases} \theta_0(1-r)^n & \text{if } 0 \le r \le 1 \\ 0 & \text{if } r \ge 1 \end{cases} \quad (4)$$

where $r \triangleq \sqrt{s_1^2(t) + s_2^2(t)}$ and θ_0 and n are constant real and natural numbers respectively, the observations will also follow a joint uniform distribution as $\rho_{X_1,X_2}(x_1, x_2) = 0.25$ for $(x_1, x_2) \in [-1, 1] \times [-1, 1]$ and 0 elsewhere, which factorizes. Thus the observations are instantaneously mutually independent, even though each of them is a nonlinear mixture of both sources. In other words, this counter-example proves that RV independence is not sufficient for separating nonlinearly mixed signals.

As a consequence, except a few dispersed works (e.g. [19] and [20]), studies in nonlinear BSS were mainly focused on specific mixing models or specific source signals, which were concerned by practical applications and for which RV independence is sufficient for ensuring identifiability and separability. Post-Nonlinear (PNL) [21], [22] and Bi-Linear (or Linear Quadratic) mixtures [17], [23] are known as the two main classes of nonlinear models investigated [24] and for which ICA leads to source separation under mild conditions. In addition, Convolutive Post-Nonlinear mixtures [18], conformal mappings [25], and linear-transformable mappings [26] are some other categories that have been addressed so far and for which RV independence leads to source separation.

C. Our Contribution

However, the above limitations are mainly due to the fact that the temporal information of the sources is not exploited. For example in [27] it is shown that even if for each time instant t_0, $x_1(t_0)$ and $x_2(t_0)$ are independent random variables, stochastic processes $x_1(t)$ and $x_2(t)$ might not be independent stochastic processes, and random variables $x_1(t_0)$ and $x_2(t_0 - 1)$ could be dependent. Taking this fact into account, previous "counter-examples" lose their validity for proving that general nonlinear mixtures are not separable.

Therefore in this work, using a more general definition of independence than RV independence used in ICA, but simpler than SP independence, we address a more general problem without being restricted to any specific mixture or parametric model. We will provide a method, based on which different algorithms can be developed for solving nonlinear BSS problems. We propose a general approach for performing the separation in nonlinear mixtures as well as the necessary conditions on the model. In this work, we also provide a separation algorithm, efficiency of which is proved by simple simulations as a proof of concept.

It should be mentioned that this work, as well as other general nonlinear BSS methods [2], [28], [29], [30], suffers from the ambiguity of a nonlinear transformation that cannot be resolved. However, it is important to differentiate between source *separation* and source *reconstruction*. In fact, once the sources are *separated*, the task of BSS is done. Source reconstruction is a more general task that is out of the scope of this work.

Although source separation can be sufficient and efficient in the cases where BSS is used as a first step before classification, in practical applications of source reconstruction, the proposed method of this paper, as well as most other papers on nonlinear BSS, serves as a first step which *separates* the sources and maybe needs to be followed by a reconstruction method. For this last step, simple and weak priors on a source like sparsity [31], bandwidth [32], zero-crossing [33], etc. can be used for reconstructing it without knowing the nonlinear distortion. This point is more elaborated in the following sections.

Parts of this work have already been presented in the conference paper [34]. The present paper not only extends the example proposed in that one, but also elaborates, details (with mathematical expressions) and discusses more the proposed method, and provides simulations with noticeable results.

The paper is organized as follows. The novel approach for solving the nonlinear BSS problem is introduced in the next section. Then a discussion on the separability and the assumptions on the model is provided. Section III contains the basic algorithms proposed for performing the separation. The algorithms are implemented and tested with examples, the results of which are presented in Section IV. Finally, conclu-
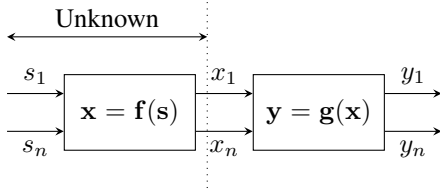
Fig. 1: Nonlinear BSS problem basic model

sions, remained questions and future works are discussed in the last section.

II. THE MAIN IDEA

The problem model, depicted in Fig. 1, considers that the number of the sources is equal to the number of the observations. In this model, we generally expect each of the elements of $\mathbf{y}(t) = \mathbf{g}(\mathbf{x}(t))$ to be a function of only one of the source signals (and each source signal appears in only one entry of $\mathbf{y}(t)$).

Since we are going to exploit *only the statistical independence* of the sources to be retrieved, and since changing the order of source signals and an invertible componentwise nonlinear transformation do not affect the independence condition, one may at most expect to obtain a "nonlinear copy" of the source vector (defined in Section II-A). In other words, for each source to be estimated, a nonlinear function remains as an ambiguity that cannot be resolved. This is discussed in more details in Section V.

The main idea is based on the fact that *the derivatives of the sources are locally mixed linearly even though the mixture model is nonlinear in general*. Indeed, if the nonlinear mapping $\mathbf{f}$ is differentiable at each point, one can derive a local linear approximation of it involving the derivatives of sources and observations. This is easily seen from

$$x_i(t) = f_i(\mathbf{s}(t)) \Rightarrow \frac{dx_i}{dt} = \sum_{j=1}^{n} \frac{\partial f_i}{\partial s_j} \frac{ds_j}{dt} \tag{5}$$

$$\Rightarrow \dot{\mathbf{x}} = \mathbf{J}_{\mathbf{f};t}(\mathbf{s})\dot{\mathbf{s}}, \tag{6}$$

where

$$\mathbf{J}_{\mathbf{f};t}(\mathbf{s}) = \begin{bmatrix} \frac{\partial f_1}{\partial s_1} & \cdots & \frac{\partial f_1}{\partial s_n} \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial s_1} & \cdots & \frac{\partial f_n}{\partial s_n} \end{bmatrix} \tag{7}$$

is the Jacobian matrix of the nonlinear mixing function, and $\dot{\mathbf{x}}$ and $\dot{\mathbf{s}}$ denote the time (or sample) derivatives of $\mathbf{x}$ and $\mathbf{s}$ respectively.

It should be noted that the precise definition of the derivative of a random process $p(t)$ is in the mean square sense, i.e. a random process $\dot{p}(t)$ is the time-derivative of a random process $p(t)$ iff $\lim_{\epsilon \to 0} \mathbf{E}[|\frac{p(t+\epsilon)-p(t)}{\epsilon} - \dot{p}(t)|^2] = 0$ where $\mathbf{E}$ represents the expected value. Nonetheless, in the rest of the paper, for the reason of simplicity, we use the equality symbol "=" for the equality of random processes in the mean squared sense as well.

It is worth noting that $\mathbf{J}_{\mathbf{f};t}(\mathbf{s})$ is the Jacobian of the nonlinear time-invariant function $\mathbf{f}$ and is a function of the sources $\mathbf{s}$, however, since the source vector is a random process and

varies over the time, the elements of $\mathbf{J}_{\mathbf{f};t}(\mathbf{s})$ change over the time as well. This is why t does not directly appear in (7) and is considered as an index of the Jacobian matrix (not an input argument). Thus, (6) is a locally *linear* instantaneous mixture model.

So, one can firstly separate the local linear mixtures of the source derivatives using a linear (but adaptive) BSS technique, and then, use an integration step to reconstruct the source signals themselves. Applying a linear BSS method on *derivatives* of the sources imposes some necessary conditions on them, which will be studied in the following section. Particularly, the DC value of signals is removed in the first step of any classical linear BSS method, hence the derivatives in our framework. Nonetheless, as mentioned earlier, the goal in this work is to reconstruct a "nonlinear copy" of the sources which can still be achieved considering this DC-removal pre-processing.

In the following, the problem of interest is formulated and all the assumptions are mentioned. Then the proposed approach is described and the separability is discussed. The discussion is made from two points of view: mathematical expressions and system analysis.

A. Problem Definition and Assumptions

Definition Let $\mathbf{s}$ be an n-dimensional vector. $\mathbf{y} = \mathbf{c}(\mathbf{s})$ is called a "nonlinear copy" of $\mathbf{s}$ if it has the same dimension as $\mathbf{s}$ and each element y_i of it is an invertible nonlinear function of one and only one of the elements of $\mathbf{s}$. It can be written as

$$\forall \, 1 \leq i \leq n \quad y_i = c_i(s_{\tau_i}) \tag{8}$$

where c_i for $i = 1, \ldots, n$ is an invertible nonlinear function and $(\tau_1, \tau_2, \ldots, \tau_n)$ is a permutation of $(1, 2, \ldots, n)$. $\square$

In this case, the transformation $\mathbf{c}(\cdot)$ which only contains component-wise nonlinear functions and permutations, is called a "nonlinear copy function" or a "trivial nonlinear mapping".

Thus, the problem can be defined as follows. Let an observation vector be an unknown nonlinear mixture of an unknown source vector $\mathbf{s}(t)$ as (1), or equivalently

$$\forall \, i \quad x_i(t) = f_i(\mathbf{s}(t)). \tag{9}$$

Source separation consists of finding a nonlinear mapping $\mathbf{g}$ as

$$\text{find} \quad \mathbf{g} \quad \text{s.t.} \quad \mathbf{g} \circ \mathbf{f} = \mathbf{c} \tag{10}$$

where $\mathbf{c} = \mathbf{g} \circ \mathbf{f}$ is a "nonlinear copy" function.

Note that an ambiguity of a permutation and a nonlinear function in reconstruction of the sources cannot be resolved. It is evident from the definition of a nonlinear copy function and (10). In addition, it can also be understood from another point of view by looking at the Jacobian of the mixing function (see Section II-B).

The above source separation problem is ill-posed without additional assumptions, either on the nonlinear mapping $\mathbf{f}$ or on the sources. In this paper, we consider the following assumptions:

1) The number of the sources is equal to the number of the observations,
2) $\mathbf{f}$ is invertible,
3) $\mathbf{f}$ is memoryless,
4) $\mathbf{f}$ is time-invariant,
5) $\mathbf{f} \in \mathbb{C}^1$ (i.e. it is differentiable with continuous first-order derivative),
6) sources $s_1(t), \ldots, s_n(t)$ are differentiable, hence colored (this assumption implies continuity and smoothness),
7) derivatives of the sources $\{\dot{s}_1(t), \ldots, \dot{s}_n(t)\}$ are mutually independent and
8) at most, one of of the derivatives of the sources follows the Gaussian distribution.

These assumptions are satisfied in most practical applications where the signals and the nonlinear mixing model correspond to real physical phenomena. In fact, the assumptions 1 to 4 are classical assumptions of BSS that are assumed even in linear cases. If the source signals have different origins (i.e. their physical origins are independent), then they will also be mutually independent *stochastic* processes, hence assumptions 6 and 7 hold.

As a consequence, all applications introduced in the Section I-A, including hyperspectral imaging [14] and determining the concentration of different ions in a combination via smart chemical sensor arrays [16] satisfy the mentioned assumptions. Therefore, nonlinear BSS problems which can be treated through the proposed approach in this work do not belong to specific set of functions and are quite general.

The necessity of these assumptions is regarding with the proposed approach which comes in Section II-B. Nevertheless, it is worth adding some remarks about some of them.

The assumption $\mathbf{f} \in \mathbb{C}^1$ imposes the continuity of $\mathbf{J_f}$. Moreover, according to the inverse function theorem [35], if a function $\mathbf{f}$ is invertible on a region in its domain and $\mathbf{f} \in \mathbb{C}^1$, 1) its Jacobian $\mathbf{J_f}$ will be non-singular on that region and 2) the Jacobian of its inverse is equal to the inverse of its Jacobian ($\mathbf{J_f}^{-1} = \mathbf{J_{f^{-1}}}$). Consequently, assumptions 5 and 2 result in continuity and non-singularity of $\mathbf{J_f}$, which makes the local linear BSS problem (6) solvable with ICA.

In addition, $\mathbf{f}$ needs to be memoryless and time-invariant, because otherwise $\mathbf{J_f}$ in (6) would also vary along time, hence the variations of local linear approximation would be too difficult to be followed by a BSS algorithm. This limitation will be better understood after Section III in which we utilize it for amending the initially proposed method.

Moreover, assumption 6, in combination with the differentiability and continuity of $\mathbf{f}$, implies the smoothness of the variations of the nonlinear function, hence its Jacobian $\mathbf{J_f}$, along the time so that it is tractable by adaptive local BSS algorithms. In other words (as it will be elaborated in Section III and simulation results), the performance of the proposed method depends on the speed of the variations of $\mathbf{J_f}$ along the time, which is due to the colorfulness of the sources and the nonlinearity of $\mathbf{f}$ itself.

As mentioned before, the proposed algorithm in this work is based on the statistical independence of the sources. Therefore, as assumed in ICA-based classical BSS methods, mixed signals in (6) need to satisfy certain conditions [4]. This is where the assumptions 7 and 8 come from.

It should be noticed that the assumptions 7 and 8 concern *derivatives* of the sources (because in (6), the mixed signals are the derivatives of the sources). The assumption 7 can be expressed as

$$\rho_S(\dot{\mathbf{s}}) = \prod_{k=1}^{N} \rho_k(\dot{s}_k) \tag{11}$$

where $\rho_S(\dot{\mathbf{s}})$ and $\rho_k(\dot{s}_k)$ correspond to the joint and marginal pdf's of the derivatives of the sources. It should be noted that a more limiting assumption than (11) was proposed as a necessary and sufficient condition for separability of nonlinear mixtures in [20] (but without any proof or explanation), which needed the signals and their derivatives to be jointly statistically independent.

Note that (11) is a completely different condition from RV independence of the source signals and is not a result of that. Generally, a signal and its derivative can be instantaneously independent: for instance, given the position of a particle at a time, one cannot say anything about its speed at that time). However, the derivative of a signal contains some information about the variations of it (which can be translated to the bandwidth or the amount of colorfulness).

To summarize, the proposed approach for nonlinear BSS in this paper is mainly based on local linear approximation of the nonlinear mixture. So, it is applicable to any nonlinear model satisfying the mentioned assumptions. In addition, a discussion is made in Section IV showing how its performance relates with the amount of the nonlinearity of the mixture (supported by simulation results).

It should be finally declared that the mentioned assumptions are not claimed to be necessary for the general separability of nonlinear mixtures. One may suggest other approaches and methods for nonlinear BSS, based on other assumptions. However, in the proposed framework, it is necessary for them to be satisfied and they are sufficient in the sense that if they hold, it is possible to separate the sources based on the proposed approach.

B. Proposed Approach

In order to get $\dot{\mathbf{x}}$, a component-wise derivative operator should be applied on the output of the mixing function $\mathbf{f}(\mathbf{s})$ of Fig. 1. Then, in order to cancel the effect of the differentiation operator (so that the separating function $\mathbf{g}(\cdot)$ in Fig. 1 remains unchanged), an integration operator needs to be added right after the differentiation operator. This will lead to the system which is depicted in Fig. 2.

Therefore, the problem (10) can be equivalently written as

$$\text{find} \quad \mathbf{g} \qquad \text{s.t.} \qquad \mathbf{g} \circ \mathbf{d}^{-1} \circ \mathbf{d} \circ \mathbf{f} = \mathbf{c} \tag{12}$$

where $\mathbf{c}$ is a nonlinear copy function and $\mathbf{d}$ and $\mathbf{d}^{-1}$ are the component-wise differentiation and integration operators respectively. For the reason of homogeneity in expressions, we use the same notation as functions for operators even though it is not mathematically accurate. It should be noted that $\mathbf{d}^{-1} \circ \mathbf{d}$
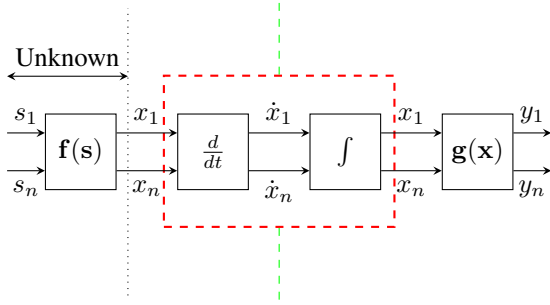
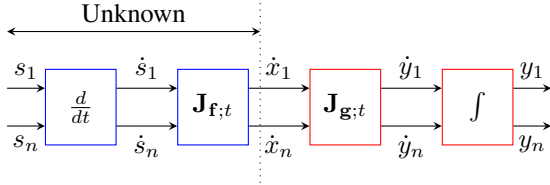Fig. 2: Nonlinear BSS problem alternative model

Fig. 3: Transforming the nonlinear BSS problem model to the linear time-variant one

is not necessarily equal to identity function because the result of integration is not unique and it could be added by any constant (in general, $\mathbf{d}^{-1} \circ \mathbf{d} \circ \mathbf{f} = \mathbf{f} + cte$). However, since $\mathbf{d}$ and $\mathbf{d}^{-1}$ operate component-wise, applying them may just add a constant value to each signal, which does not affect the proposed framework.

However, (6) says that the derivatives of the observations locally are linear mixtures of the derivatives of the sources. It means that they can be achieved by mixing the derivatives of sources via the Jacobian matrix of the nonlinear mixing function. In other words, considering (6), each half of this new model (which is nonlinear), can be replaced by an equivalent one (which is locally linear) shown in Fig. 3.

Mathematically speaking, denoting $\mathbf{J}_{\mathbf{f};t} = \partial \mathbf{f}/\partial \mathbf{s}$ and $\mathbf{J}_{\mathbf{g};t} = \partial \mathbf{g}/\partial \mathbf{x}$ the Jacobian matrices of the mixing function $\mathbf{f}$ and separating function $\mathbf{g}$ respectively, the equivalence of the systems of Fig. 2 and Fig. 3 can be written as

$$\begin{cases} \mathbf{d} \circ \mathbf{f} \equiv \mathbf{J}_{\mathbf{f};t} \circ \mathbf{d} \\ \mathbf{g} \circ \mathbf{d}^{-1} \equiv \mathbf{d}^{-1} \circ \mathbf{J}_{\mathbf{g};t} \end{cases} . \qquad (13)$$

This equation says that instead of taking derivatives of a mixture of sources (i.e. $\mathbf{d} \circ \mathbf{f}$), one can equivalently mix derivatives of the sources via the Jacobian of the mixing function (i.e. $\mathbf{J}_{\mathbf{f};t} \circ \mathbf{d}$).

Then, replacing $\mathbf{d} \circ \mathbf{f}$ and $\mathbf{g} \circ \mathbf{d}^{-1}$ in (12) with their equivalents in (13), the nonlinear BSS problem becomes

$$\forall t, \text{ find } \mathbf{J}_{\mathbf{g};t} \quad \text{s.t.} \quad \mathbf{d}^{-1} \circ \mathbf{J}_{\mathbf{g};t} \circ \mathbf{J}_{\mathbf{f};t} \circ \mathbf{d} = \mathbf{c}. \quad (14)$$

This new model (depicted in Fig. 3) will be used for a discussion on the separability and proposing an algorithm.

Regarding (14) and Fig. 3, the goal is to find a linear time-variant system $\mathbf{J}_{\mathbf{g};t}$ such that each of the output signals $y_1(t), \ldots, y_n(t)$ is a function of only one of the sources, hence $\mathbf{y}$ is a nonlinear copy of the sources.

By left-multiplying both sides of (14) by $\mathbf{d}$, and right-multiplying them by $\mathbf{d}^{-1}$, we will have

$$\mathbf{d} \circ \mathbf{d}^{-1} \circ \mathbf{J}_{\mathbf{g};t} \circ \mathbf{J}_{\mathbf{f};t} \circ \mathbf{d} \circ \mathbf{d}^{-1} = \mathbf{d} \circ \mathbf{c} \circ \mathbf{d}^{-1} \qquad (15)$$

$$\Rightarrow \mathbf{J}_{\mathbf{g};t} \circ \mathbf{J}_{\mathbf{f};t} = \mathbf{d} \circ \mathbf{c} \circ \mathbf{d}^{-1} = \mathbf{c}_2 \quad (16)$$

where the last equation comes from the fact that $\mathbf{c}$ is a nonlinear copy function and, therefore, in combination with $\mathbf{d}$ and $\mathbf{d}^{-1}$ makes another nonlinear copy function named $\mathbf{c}_2$. As a consequence, the basic problem (10) is equivalent to

$$\forall t \quad \text{find} \quad \mathbf{J}_{\mathbf{g};t} \quad \text{s.t.} \quad \mathbf{J}_{\mathbf{g};t} \circ \mathbf{J}_{\mathbf{f};t} = \mathbf{c}_2 \quad (17)$$

where $\mathbf{c}_2$ is a nonlinear copy function. This is a traditional linear BSS problem where the mixing matrix is not constant along the time, and can be solved via existing *adaptive* linear BSS methods (probably, with some modifications). As a conclusion, any nonlinear BSS problem is equivalent to a time-varying linear one and if the linear problem is solved correctly, the nonlinear problem will be solved as well.

It is worth adding two remarks which help better understanding the proposed concept. Firstly, the local linear mixing $\mathbf{J}_{\mathbf{f};t}$ and separating $\mathbf{J}_{\mathbf{g};t}$ matrices are the Jacobian matrices of the nonlinear mixing $\mathbf{f}$ and separating $\mathbf{g}$ functions respectively. Neglecting the indeterminacies in reconstructing the sources, it is obvious from Fig. 3 that the matrix $\mathbf{J}_{\mathbf{g};t}$ should be the inverse of the matrix $\mathbf{J}_{\mathbf{f};t}$. Actually, as mentioned in Section II-A, the inverse of the Jacobian of a function is the Jacobian of the inverse function [35]. This can also be easily shown by writing the equivalency equations of the right half of the systems of Fig. 2 and 3.

Secondly, the proposed approach could also be derived by trying to linearly approximate the nonlinear function via Taylor expansion as

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

$$\forall t \quad \mathbf{x}(t + \epsilon) = \mathbf{x}(t) + \frac{\partial \mathbf{f}}{\partial \mathbf{s}}(\mathbf{s}(t + \epsilon) - \mathbf{s}(t)) + \mathrm{o}(\epsilon) \qquad (18)$$

$$\Rightarrow \mathbf{x}(t + \epsilon) - \mathbf{x}(t) \approx \mathbf{J}_{\mathbf{f};t}(\mathbf{s})\Big|_{\mathbf{s}=\mathbf{s}(t)} (\mathbf{s}(t + \epsilon) - \mathbf{s}(t)) \quad (19)$$

$$\Rightarrow \Delta_{\mathbf{x}}(t) \approx \mathbf{J}_{\mathbf{f};t}(\mathbf{s})\Big|_{\mathbf{s}=\mathbf{s}(t)} \Delta_{\mathbf{s}}(t), \qquad (20)$$

where $\mathrm{o}(\epsilon)$ represents Higher-Order Terms and $\Delta_{\mathbf{x}}(t)$ and $\Delta_{\mathbf{s}}(t)$ are the differences (increments) of the observation and source vectors respectively.

Eq. (20) can also be considered as a discrete-time approximation of (6) using the *difference* instead of the *derivative*. Nevertheless, the proposed framework can also be understood as the (local) linear approximation of the nonlinear mixing function at each point, and trying to separate the sources using adaptive linear BSS methods.

Reconstruction Indeterminacies: Linear BSS methods suffer from ambiguities both in the order of the sources and their scales. On the other hand, as pointed earlier and will be explained in the following, the proposed framework in this paper is based on the local linear approximation of the nonlinear mixture. However, there are two well-known ambiguities in traditional BSS methods: permutation and scaling. So it is important to understand how these *local* ambiguities perform *globally*.

Since the local separating matrix $\mathbf{J}_{\mathbf{g};t}$ is estimated adaptively and continuously, the local permutation matrix should also change continuously. However, a change in permutation cannot be continuous at all. Therefore, *local* permutations in any neighbourhood of observations result in an arbitrary *global* permutation, and do not cause any issue about the alignment of permutations at successive time instants.

Moreover, the time-varying values of the scaling ambiguity on the whole domain of the signals cause a *component-wise* nonlinearity which cannot be resolved by the proposed algorithm, i.e. each output of the algorithm does depend on only one of the sources but with a time-varying scaling factor (i.e. a nonlinear function).

This indeterminacy in reconstructing the sources could also be seen from another point of view. Assume $\mathbf{u}(\cdot)$ is a component-wise nonlinear function as

$$\tilde{\mathbf{y}}(t) = \mathbf{u}(\mathbf{y}(t)) \qquad (21)$$

such that

$$\forall\, 1 \le k \le n \quad \tilde{y}_k(t) = u_k(\mathbf{y}(t)) = \hat{u}_k(y_k(t)) \qquad (22)$$

where $\hat{u}_k(\cdot)$ for $k = 1, \ldots, n$ are 1-dimensional $\mathbb{R} \to \mathbb{R}$ nonlinear functions.

Obviously, the Jacobian of a component-wise function is diagonal. As a consequence, if $\mathbf{J}_{\mathbf{g};t}$ satisfies (17), $\mathbf{J}_{\mathbf{u}\circ\mathbf{g}} = \mathbf{J}_{\mathbf{u}}\mathbf{J}_{\mathbf{g}}$ will satisfy (17) as well. Indeed, if a function $\mathbf{g}$ (resulting in $\mathbf{y}$ as the separated sources) is a separating function, the function $\mathbf{u} \circ \mathbf{g}$ (resulting in $\tilde{\mathbf{y}}(t) = \mathbf{u}(\mathbf{y}(t))$) will also separate the sources. In other words, the proposed approach may result in any component-wise nonlinear function of the sources.

III. Proposed Algorithms

It follows from Fig. 3 that

$$\dot{\mathbf{y}}(t) = \mathbf{J}_{\mathbf{g};t}(\mathbf{x}(t))\dot{\mathbf{x}}(t) = \mathbf{J}_{\mathbf{g};t}(\mathbf{x}(t))\mathbf{J}_{\mathbf{f};t}(\mathbf{s}(t))\dot{\mathbf{s}}(t). \qquad (23)$$

Therefore it is necessary and sufficient for the separation to find a matrix $\mathbf{J}_{\mathbf{g};t}(\mathbf{x}(t))$ such that the off-diagonal elements of $\mathbf{J}_{\mathbf{g};t}(\mathbf{x}(t))\mathbf{J}_{\mathbf{f};t}(\mathbf{s}(t))$ are zero everywhere and its diagonal elements are nonlinear copy functions.

In this section, we are going to propose algorithms in order to perform nonlinear BSS based on the proposed idea. To this end, firstly an adaptive linear BSS method is recalled in subsection III-A, which plays an important role in the proposed algorithms. In this subsection, the necessity of utilizing an adaptive algorithm is highlighted and its exact formulation is provided. Then a basic algorithm is proposed in subsection III-B led by the sequencing steps of the mentioned approach of Section II. Afterwards, in subsection III-C, the main problem of the proposed preliminary algorithm is discussed and addressed by nonlinear regression of the separating function. Finally, in subsection III-D a modified algorithm is proposed employing the "Nonlinear Regression" technique.

A. Adaptive Linear BSS (Normalized EASI)

An adaptive BSS algorithm is an algorithm whose estimation of mixing/separating matrix is on-line, i.e. adjusted at each new sample that is observed. Normalized EASI (Equivariant Adaptive Separation via Independence) [7] is one of the adaptive BSS algorithms that is based on the statistical independence of the sources. This powerful real-time algorithm is used in this work as the adaptive linear BSS method for estimating the $\mathbf{J}_{\mathbf{g};t}$ matrix, which cancels the mixture $\mathbf{J}_{\mathbf{f};t}$. In this purpose, components of $\dot{\mathbf{s}}$ must be statistically independent. In other words, while assumption 7 is necessary because of Normalized EASI, using other algorithms might impose other assumptions on the sources.

Since the mixing matrix in (23) (i.e. $\mathbf{J}_{\mathbf{g};t}$) changes along the time, an *adaptive* technique needs to be utilized to perform the linear BSS (so that it can follows the variations of $\mathbf{J}_{\mathbf{g};t}$). Taking advantage of the equivariancy, good convergence rate and low computational cost of Normalized EASI, it has been used as the adaptive linear BSS algorithm in whole this work.

The update formula of the separating matrix $\mathbf{J}_{\mathbf{g};t}$ according to this algorithm will be as

$$\mathbf{J}_{\mathbf{g};t+1} = \mathbf{J}_{\mathbf{g};t} - \lambda_t \left[\frac{\mathbf{y}(t)\mathbf{y}(t)^{\mathcal{T}} - \mathbf{I}}{1 + \lambda_t \mathbf{y}(t)^{\mathcal{T}}\mathbf{y}(t)} \right.$$
$$\left. + \frac{\mathfrak{h}(\mathbf{y}(t))\mathbf{y}(t)^{\mathcal{T}} - \mathbf{y}(t)\mathfrak{h}(\mathbf{y}(t))^{\mathcal{T}}}{1 + \lambda_t |\mathbf{y}(t)^{\mathcal{T}}\mathfrak{h}(\mathbf{y}(t))|} \right] \mathbf{J}_{\mathbf{g};t} \quad (24)$$

where λ_t is a sequence of positive adaptation steps and $\mathfrak{h}(\cdot)$ is an arbitrary component-wise (n-dimensional) nonlinear function. For a more detailed discussion on the choice of the components $\mathfrak{h}_i(\cdot)$ of $\mathfrak{h}(\cdot)$, the reader is invited to refer to [7].

Plainly, at each iteration, eq. (24) is followed by an update of the output vector as

$$\dot{\mathbf{y}}(t+1) = \mathbf{J}_{\mathbf{g};t+1} \ \dot{\mathbf{x}}(t+1). \qquad (25)$$

B. Preliminary Algorithm

As mentioned earlier, assuming $\mathbf{J}_{\mathbf{g};t}(\mathbf{x}(t))$ in Eq. (23) to vary slowly enough such that it remains almost constant in the temporal neighborhood of each point $\mathbf{x}(t)$, a preliminary algorithm can be suggested simply as locally solving linear BSS problems at all time instants.

Accordingly, the first algorithm, called Adaptive Algorithm for Time-Variant Linear mixtures (AATVL), is sketched in Algorithm 1, where in line (2), EASI or any other *adaptive* linear BSS technique can be employed.

The main problem with this algorithm is the issue of convergence: it always needs to be updated at each new sample of observations. In conventional applications of Normalized EASI, where the mixing matrix is assumed to be constant, after a number of iterations the algorithm (hopefully) converges to the exact separating matrix. However, in our case where $\mathbf{J}_{\mathbf{f};t}$ varies from one sample to another, the algorithm should not only estimate the exact separating matrix $\mathbf{J}_{\mathbf{g};t}$ at each sample, but it should also track the variations of $\mathbf{J}_{\mathbf{f};t}$ along the time.

Algorithm 1 Adaptive Algorithm for Time-Variant Linear mixtures (AATVL)

1: $\dot{\mathbf{x}} \leftarrow$ Derivative (difference) of $\mathbf{x}$
2: **procedure** ADAPTIVE LINEAR BSS METHOD ($\dot{\mathbf{x}}(t)$)
3:      $\mathbf{J}_{\mathbf{g};0} \leftarrow$ Random Initialization
4:      $\dot{\mathbf{y}}(0) = \mathbf{J}_{\mathbf{g};0} \, \dot{\mathbf{x}}(0)$
5:      **for** $t = 0, \ldots, T-1$ **do**
6:          $\mathbf{J}_{\mathbf{g};t+1} \leftarrow$ Update by eq. (24)
7:          $\dot{\mathbf{y}}(t+1) \leftarrow$ Update by eq. (25)
8:      **end for**
9: **end procedure**
10: $\mathbf{y} \leftarrow$ Integral of $\dot{\mathbf{y}}$

So the convergence issue is much more severe than the classic linear problem.

It is worth noting that the variations of $\mathbf{J}_{\mathbf{f};t}(\mathbf{s}(t))$ depend on both the nonlinearity of the mixing model $\mathbf{f}(\cdot)$ and the dynamics of the sources $\mathbf{s}(t)$. Thus, even if the nonlinear mixing function $\mathbf{f}(\cdot)$ is smooth, bursty sources may lead to bursty changes in the mixing values, and consequently, the separating matrix cannot be tracked by the separating algorithm. This is the reason why the proposed approach needs both assumptions 4 (the mixture to be time-invariant) and 6 (the sources to be colored) to impose the smoothness on $\mathbf{J}_{\mathbf{f};t}(\mathbf{s}(t))$ along the time.

Another issue, which makes the convergence problem even more severe, is that the output of this adaptive linear BSS algorithm is going to be *integrated* through a following step to estimate the separated sources. This integration will propagate the estimation error to the other samples as well. As a consequence, the AATVL algorithm (algorithm 1) needs to be modified.

C. Nonlinear Regression

In this subsection, the main problem of the proposed preliminary algorithm (i.e. convergence) is addressed by a nonlinear regression technique. The concept is explained in details providing 2 different methods (subsections III-C1 and III-C2). The second method (which is actually used in the modified algorithm 2) is supported by a simulated preliminary example and a discussion on its performance.

The convergence problem of the algorithm 1 is because it does not exploit the time-invariance and smoothness of the mixing function $\mathbf{f}$. In fact, the original nonlinearity $\mathbf{f}$, and its inverse $\mathbf{g}$, are time-invariant. Therefore the dependence of $\mathbf{J}_{\mathbf{f};t}$ ($\mathbf{J}_{\mathbf{g};t}$) on $\mathbf{s}$ ($\mathbf{x}$) is not time-varying.

In other words, $\mathbf{s}$ and $\mathbf{x}$ themselves are time-varying, and $\mathbf{J}_{\mathbf{f};t}$ and $\mathbf{J}_{\mathbf{g};t}$ are evaluated for sources and observations at successive times as

$$\mathbf{J}_{\mathbf{f};t}(\mathbf{s}(t)) = \left.\frac{\partial \mathbf{f}}{\partial \mathbf{s}}(\mathbf{s})\right|_{\mathbf{s}=\mathbf{s}(t)} \qquad (26)$$

$$\mathbf{J}_{\mathbf{g};t}(\mathbf{x}(t)) = \left.\frac{\partial \mathbf{g}}{\partial \mathbf{x}}(\mathbf{x})\right|_{\mathbf{x}=\mathbf{x}(t)}. \qquad (27)$$

As a result, a modification on the algorithm 1 can be suggested by learning the nonlinear model of $\mathbf{J}_{\mathbf{g};t}(\mathbf{x})$ from

its estimations at different samples (say $\hat{\mathbf{J}}_{\mathbf{g}}(\mathbf{x}(t))$ for $t = 1, \ldots, T$, the outputs of the adaptive linear BSS method). It should be noted that $\mathbf{J}_{\mathbf{g};t}(\mathbf{x})$ is an $n \times n$ matrix and contains n^2 nonlinear functions that should be learned in this approach.

For example, let $[\mathbf{J}_{\mathbf{g};t}(\mathbf{x})]_{i,j}$ denote the $(i,j)^{\text{th}}$ element of the separating matrix. In the "nonlinear regression" stage, we aim at estimating the nonlinear function $[\mathbf{J}_{\mathbf{g};t}(\mathbf{x})]_{i,j}$ from $[\hat{\mathbf{J}}_{\mathbf{g}}(\mathbf{x}(t))]_{i,j}$ for $t = 1, \ldots, T$. In the simplest case, it can be mathematically expressed as for all $1 \le i, j \le n$

$$\underset{[\mathbf{J}_{\mathbf{g};t}(\mathbf{x})]_{i,j}}{\text{minimize}} \sum_{t=1}^{T} \left(\mathrm{d}_w^2([\hat{\mathbf{J}}_{\mathbf{g}}(\mathbf{x}(t))]_{i,j}, [\mathbf{J}_{\mathbf{g};t}(\mathbf{x})]_{i,j}) \right) \qquad (28)$$

where d_w^2 represents a weighted squared distance of a point and a manifold defined as

$$\mathrm{d}_w^2(\cdot, \cdot) = \mathrm{d}^2(\cdot, \cdot) \times w(\mathrm{d}^2(\cdot, \cdot)) \qquad (29)$$

and

$$\mathrm{d}^2([\hat{\mathbf{J}}_{\mathbf{g}}(\mathbf{x}(t))]_{i,j}, [\mathbf{J}_{\mathbf{g};t}(\mathbf{x})]_{i,j}) = \\ \mid [\mathbf{J}_{\mathbf{g};t}(\mathbf{x})]_{i,j}\big|_{\mathbf{x}=\mathbf{x}(t)} - [\hat{\mathbf{J}}_{\mathbf{g}}(\mathbf{x}(t))]_{i,j} \mid^2. \quad (30)$$

Since the error in the estimation $[\hat{\mathbf{J}}_{\mathbf{g}}(\mathbf{x}(t))]_{i,j}$ might be large for some samples (especially due the convergence issue), there might be some *outliers* in the data. Although the outliers are supposed to be rare, due to the power of 2 in (30), they can highly affect the result of the manifold learning process. Consequently, using a *weighted* distance in (28) is essential in order to reduce the effect of the estimations that are too far from the learned manifold.

The weighting function is designed such that it is close to 1 for short distances and it tends to zero as the distance increases. As an example, Gaussian weighting function can be defined as

$$w(\mathrm{d}^2) = e^{-\frac{\mathrm{d}^2}{2\sigma^2}} \qquad (31)$$

where σ is a parameter which can be adjusted according to the data.

The optimization (28), where the cost function should be minimized with respect to a nonlinear manifold, can be performed using either a parametric model (when the nonlinear function is assumed to belong to a specific set of functions, e.g. polynomials) or a non-parametric one (utilizing an interpolation method like smoothing splines). One may also modify a dimension reduction technique (e.g. ISOMAP [36]) in order to solve (28).

1) Parametric Approach: In this approach, a parametric model for each $[\mathbf{J}_{\mathbf{g};t}(\mathbf{x})]_{i,j}$ is assumed and then the minimization of (28) is performed with respect to those parameters. In other words, we assume each manifold to be formulated as

$$[\mathbf{J}_{\mathbf{g};t}(\mathbf{x})]_{i,j} = Q_{i,j}(\mathbf{x}; \boldsymbol{\theta}_{i,j}) \qquad (32)$$

where $\boldsymbol{\theta}_{i,j}$ is a vector of the parameters in nonlinear model of $[\mathbf{J}_{\mathbf{g};t}(\mathbf{x})]_{i,j}$.

As a consequence, with this parametric model, (28) becomes

$$\underset{\boldsymbol{\theta}_{i,j}}{\text{minimize}} \sum_{t=1}^{T} \left(\mathrm{d}_w^2([\hat{\mathbf{J}}_{\mathbf{g}}(\mathbf{x}(t))]_{i,j}, [\mathbf{J}_{\mathbf{g};t}(\mathbf{x})]_{i,j}) \right) \qquad (33)$$
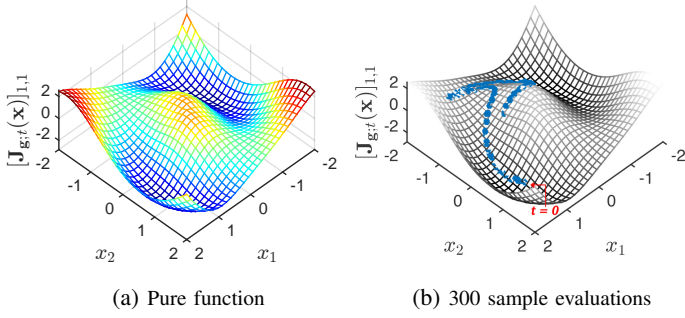
(a) Pure function   (b) 300 sample evaluations

Fig. 4: The nonlinear function of $[\mathbf{J_g}(\mathbf{x})]_{1,1}$ of (36) with respect to the observations

(a) 300 samples   (b) 700 samples

Fig. 5: The estimated (learned) nonlinear model of $[\mathbf{J_g}(\mathbf{x})]_{1,1}$ from 300 (Fig. 5a) and 700 (Fig. 5b) samples of observations. The circles are the outputs of the adaptive linear BSS method $[\hat{\mathbf{J}}_{\mathbf{g}}(\mathbf{x}(t))]_{1,1}$, and hyper-surface is the learned manifold using the introduced smoothing spline technique.

where $\mathrm{d}_w^2([\hat{\mathbf{J}}_{\mathbf{g}}(\mathbf{x}(t))]_{i,j}, [\mathbf{J}_{\mathbf{g};t}(\mathbf{x})]_{i,j})$ can be calculated as a function of the parameters according to (29) and (30). Thus it can be solved, and the optimal parameter vectors $\boldsymbol{\theta}_{i,j}^*$ will let us formulate the $[\mathbf{J}_{\mathbf{g};t}(\mathbf{x})]_{i,j}$'s.

2) Non-Parametric Approach: The other approach proposed for nonlinear regression is non-parametric where no model for the nonlinearity is assumed. To this end, the nonlinear functions are learned by fitting curves using a smoothing method (e.g. smoothing splines [37]) to the estimations $[\hat{\mathbf{J}}_{\mathbf{g}}(\mathbf{x}(t))]_{i,j}$ for $t = 1, \ldots, T$.

In this work, *smoothing spline* [38] is utilized as the smoothing method, for which a penalty function (the second order derivative of $[\mathbf{J}_{\mathbf{g};t}(\mathbf{x})]_{i,j}$) is added to the cost function (28) to impose the smoothness. In this method, there is a smoothing parameter, controlling the trade-off between fidelity to the data and roughness of the function estimate.

This method is explained via studying its performance on an example with a mixing function $\mathbf{f}$ as (2). This model is a rotation with the angle which depends to the norm of the source vector. So the inverse function $\mathbf{g}$ can be easily achieved by another rotation with the negative angle as

$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \begin{bmatrix} \cos\alpha(\mathbf{x}(t)) & \sin\alpha(\mathbf{x}(t)) \\ -\sin\alpha(\mathbf{x}(t)) & \cos\alpha(\mathbf{x}(t)) \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} \quad (34)$$

where

$$\alpha(\mathbf{x}(t)) = \alpha_0 + \gamma \times \sqrt{x_1^2(t) + x_2^2(t)}. \quad (35)$$

Therefore, the exact Jacobian $\mathbf{J_g}(\mathbf{x})$ is calculated as

$$\mathbf{J_g}(\mathbf{x}) = \begin{bmatrix} \cos\alpha(\mathbf{x}) & \sin\alpha(\mathbf{x}) \\ -\sin\alpha(\mathbf{x}) & \cos\alpha(\mathbf{x}) \end{bmatrix}$$
$$\begin{bmatrix} 1 + x_2\frac{\partial\alpha(\mathbf{x})}{\partial x_1} & x_2\frac{\partial\alpha(\mathbf{x})}{\partial x_2} \\ -x_1\frac{\partial\alpha(\mathbf{x})}{\partial x_1} & 1 - x_1\frac{\partial\alpha(\mathbf{x})}{\partial x_2} \end{bmatrix}. \quad (36)$$

Now consider one of the elements of $\mathbf{J_g}(\mathbf{x})$, say $[\mathbf{J_g}(\mathbf{x})]_{1,1}$. In this example, $n = 2$ and the 2-dimensional nonlinear function of the function $[\mathbf{J_g}(\mathbf{x})]_{1,1}$ with respect to x_1 and x_2 (calculated in (36)) is depicted in Fig. 4a.

As an example, suppose that the sources $s_1(t)$ and $s_2(t)$ are integrals of a triangle (with the amplitude of 6 and the primitive period of 200π samples) and a sinusoidal (with the amplitude of 6 and the frequency of $\sqrt{3}/200\pi$ samples) signal respectively. The trajectory of the observation vector along the
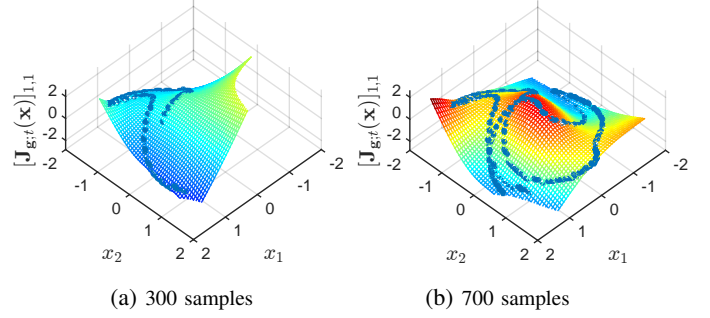
time projected onto the 2-dimensional manifold of $[\mathbf{J_g}(\mathbf{x})]_{1,1}$ for 300 time instants is plotted in Fig. 4b. It illustrates the changes in the value of $[\mathbf{J}_{\mathbf{g};t}(\mathbf{x})]_{1,1}$ along the time. It is nice to see that as time passes, the observation vector takes different values and may span its whole range, such that it will be possible to learn the whole shape of the nonlinear function.

Fig. 5 shows the learned nonlinear model (the hyper-surface) given 300 and 700 samples of $[\hat{\mathbf{J}}_{\mathbf{g}}(\mathbf{x}(t))]_{1,1}$ using the smoothing spline technique. It can be seen that the learned nonlinear model from 700 samples based on smoothing spline is quite accurate in the region of interest, i.e. where samples are available.

Fig. 6 shows the Normalized Root Mean Squared (RMS) error in reconstruction of $[\mathbf{J_g}(\mathbf{x})]_{1,1}$ in (36) with respect to the number of observation samples. The error E_{nrms} is calculated as

$$E_{nrms} = \frac{\left(\iint\limits_{|x_1|,|x_2|\leq M} \left([\mathbf{J_g}(\mathbf{x})]_{1,1} - [\hat{\mathbf{J}}_{\mathbf{g}}(\mathbf{x})]_{1,1} \right)^2 \right)^{\frac{1}{2}}}{\left(\iint\limits_{|x_1|,|x_2|\leq M} \left([\hat{\mathbf{J}}_{\mathbf{g}}(\mathbf{x})]_{1,1} \right)^2 \right)^{\frac{1}{2}}} \quad (37)$$

where $M = \max\left(\max(|x_1(t)|), \max(|x_2(t)|)\right)$ is the maximum range of variations of the observations.

However, a more meaningful definition of the N-RMS error is when it is calculated over the region of interest, as

$$\tilde{E}_{nrms} = \frac{\left(\sum\limits_{t=1,\ldots,T} \left([\mathbf{J_g}(\mathbf{x}(t))]_{1,1} - [\hat{\mathbf{J}}_{\mathbf{g}}(\mathbf{x}(t))]_{1,1} \right)^2 \right)^{\frac{1}{2}}}{\left(\sum\limits_{t=1,\ldots,T} \left([\hat{\mathbf{J}}_{\mathbf{g}}(\mathbf{x}(t))]_{1,1} \right)^2 \right)^{\frac{1}{2}}}. \quad (38)$$

$\tilde{E}_{nrms}$, as well as E_{nrms}, decreases as the number of samples increases (see Fig. 6).

According to Fig. 6, the accuracy of the estimated model improves as the number of input samples grows until a certain number at which the estimation is close enough to the correct model and the error does not decrease anymore.

It should be added that the utilized algorithm in this example (smoothing splines) does not force the model to pass the input points. Nevertheless, depending on the application, other smoothing algorithms with different properties (more robust to
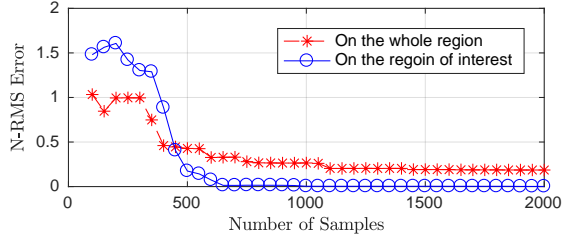
Fig. 6: The N-RMS error of the estimation of the nonlinear model of $[\mathbf{J_g}(\mathbf{x})]_{1,1}$ with respect to the number of samples over 1) an $M \times M$ square (the dashed line) and 2) the region of interest in which the samples exist (the solid line)

Algorithm 2 Batch Algorithm for Time-Invariant Nonlinear mixtures (BATIN)

1: $\dot{\mathbf{x}} \leftarrow$ Derivative (difference) of $\mathbf{x}$
Step Adaptive linear BSS:
2: **procedure** ADAPTIVE LINEAR BSS METHOD ($\dot{\mathbf{x}}(t)$)
3: $\quad\hat{\mathbf{J}}_\mathbf{g}(\mathbf{x}(0)) \leftarrow$ Random Initialization
4: $\quad\dot{\mathbf{y}}(0) = \hat{\mathbf{J}}_\mathbf{g}(\mathbf{x}(0))\, \dot{\mathbf{x}}(0)$
5: $\quad$**for** $t = 0, \ldots, T-1$ **do**
6: $\qquad\hat{\mathbf{J}}_\mathbf{g}(\mathbf{x}(t+1)) \leftarrow$ Update by eq. (24)
7: $\qquad\dot{\mathbf{y}}(t+1) \leftarrow$ Update by eq. (25)
8: $\quad$**end for**
9: **end procedure**
Step Nonlinear Separation:
10: **procedure** NONLINEAR REGRESSION ($\hat{\mathbf{J}}_\mathbf{g}(\mathbf{x}(t)), \mathbf{x}(t)$)
11: $\quad\mathbf{J}_{\mathbf{g};t}(\mathbf{x}) \leftarrow$ Smoothing Spline of $\hat{\mathbf{J}}_\mathbf{g}(\mathbf{x}(t))$
12: **end procedure**
13: **for** $t = 1, \ldots, T$ **do**
14: $\quad\dot{\mathbf{y}}(t) \leftarrow \mathbf{J}_{\mathbf{g};t}(\mathbf{x})\, \dot{\mathbf{x}}(t)$
15: **end for**
16: $\mathbf{y} \leftarrow$ Integral of $\dot{\mathbf{y}}$

---

noise, forcing to pass the points, and so on) may be exploited for estimating the function (e.g. Kalman filter, kernel smoother, Laplacian smoothing, exponential smoothing, etc.).

D. Modified Algorithm

Employing the nonlinear regression idea introduced in Section III-C in combination with algorithm 1 leads to a second algorithm which outperforms the first one. This algorithm includes 2 steps: 1) an "Adaptive linear BSS" algorithm for estimating $\hat{\mathbf{J}}_\mathbf{g}(\mathbf{x}(t))$ for $t = 1, \ldots, T$ and 2) a "Nonlinear Separation" process through which the nonlinear functions $\mathbf{J}_{\mathbf{g};t}(\mathbf{x})$ are learned by the proposed smoothing spline method and are used to separate the sources. Once the nonlinear functions $\mathbf{J}_{\mathbf{g};t}(\mathbf{x})$ are estimated, they are used for separating the derivatives of the sources. The batch Algorithm for Time-Invariant Nonlinear mixtures (BATIN) can thus be proposed as algorithm 2.

It should be finally noted that, the Normalized EASI and the smoothing spline algorithms that are used in algorithm 2, can probably be replaced by other equivalent algorithms depending on the application.
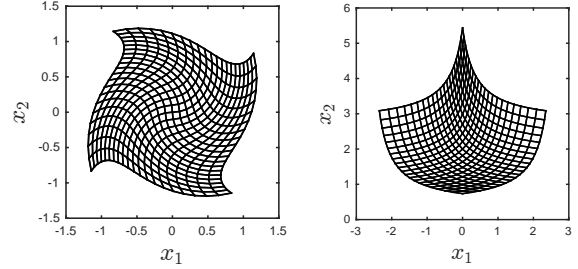
(a) Simulation 1   (b) Simulation 2

Fig. 7: Illustration of the nonlinear mappings. 7a) the mapping follows model (2) and (39) for $\alpha_0 = 0$ and $\gamma = 1$ and 7b) the mapping follows model (40). In both figures, we represent the grid obtained by applying the nonlinear mapping (2) or (40) to the regular grid in the domain $[-1, +1] \times [-1, +1]$, and the input domain is mapped to nonlinear grids in the output domain which are shown.

IV. SIMULATIONS

In this section, simulation results of both proposed algorithms for two different nonlinear functions are shown as proof of concept. The data model, nonlinear functions, the parameters and the details of the simulations come in Section IV-A. Afterwards, the results of the simulations and their performance evaluation are reported in Section IV-B.

A. Simulated Data and Mixture Models

In the first example, consider the two-input two-output system of (2) where instead of (4), $\alpha(\mathbf{s}(t))$ is defined by the parametric model

$$\alpha(\mathbf{s}(t)) = \alpha_0 + \gamma \times \sqrt{s_1^2(t) + s_2^2(t)} \tag{39}$$

where α_0 and γ are some parameters.

In our first simulation, (39) is considered for $\alpha_0 = 0$ and $\gamma = 1$. Secondly, the proposed method is applied to another mixing model defined as

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \mathbf{f}(\mathbf{s}(t)) = \begin{bmatrix} e^{s_1(t)} - e^{s_2(t)} \\ e^{-s_1(t)} + e^{-s_2(t)} \end{bmatrix} \tag{40}$$

which is a nonlinear but invertible mixing model, as well as the first one.

The function mappings of the two simulated models are illustrated in Fig. 7: the figure shows how a regular grid in the input domain is transformed through the functions. As it can be understood from this figure as well as (2) and (40), both models are nonlinear but bijective (one-to-one) in the input range.

In both simulations, the two sources that are mixed are the integrals of a sine wave

$$\dot{s}_1(t) = \sin(\sqrt{3}t/100) \;\Rightarrow\; s_1(t) \propto \int \dot{s}_1(t)\,\mathrm{d}\,t \tag{41}$$

and a triangle wave

$$\dot{s}_2(t) = \mathrm{saw}(t/100) \;\Rightarrow\; s_2(t) \propto \int \dot{s}_2(t)\,\mathrm{d}\,t \tag{42}$$

where $\mathrm{saw}(t)$ is defined as a sawtooth wave with period 2π passing through the points $(0,0)$, $(\pi/2, 1)$, $(3\pi/2, -1)$ and
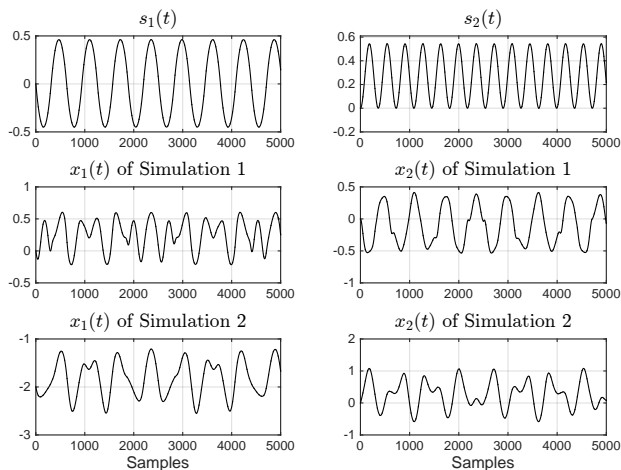
Fig. 8: The sources $s_1(t)$ and $s_2(t)$ (the integral of a sine and a triangle wave) in the top row, and the observations $x_1(t)$ and $x_2(t)$ for the two simulations with the nonlinear model (2) in the middle and with the nonlinear model (40) in the bottom.
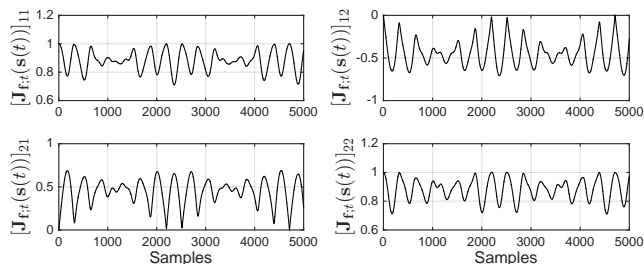

Fig. 9: Variations of the elements of the Jacobian matrix of (2) along the samples

(a) The first source

(b) The second source

Fig. 10: The results of AATVL and BATIN algorithm in the mixture (2)

$(2\pi, 0)$. The sources are chosen well-known simple signals with different frequencies avoiding any coherence, and satisfying assumptions on $\mathbf{s}$, and especially independence of the derivatives (assumption 7).

It should be noted that the integral can be practically approximated by either a recursive summation $s(t) = \dot{s}(t) + s(t-1)$ or a continuous function estimation based on an interpolation method. Simulations (not presented in this paper) show that these two approaches result in almost the same estimation. Thus the summation is used as an approximation of the integral everywhere.

The observations are then calculated by (2) and (40), and are depicted in Fig. 8 as well as the sources themselves.

In order to see the time-variations of the mixing matrix, each of the elements of the Jacobian matrix of the first simulation (2) for $\alpha_0 = 0$ and $\gamma = 0.1$ is plotted separately in Fig. 9. It can be seen that their variations along the time is periodic (because of the dynamics of the source). As mentioned earlier, variations of the value of the Jacobian are due to both time-variations of the sources and nonlinearity of the mixing function (which make the Jacobian dependent on the value of sources).

AATVL and BATIN algorithms are applied on the observations of Fig. 8 to separate the sources. As mentioned earlier, smoothing spline is the algorithm that is utilized for
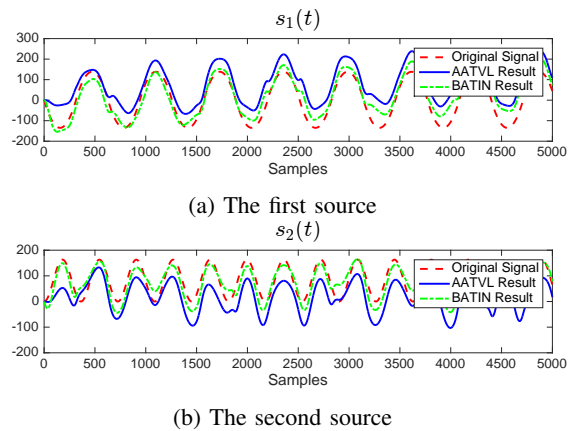
the nonlinear regression step of algorithm 2. Note that the smoothing parameter, which determines the smoothness of the learned manifold in smoothing spline method, is adjusted heuristically in this work. It should be noted that similarly with the integral, the *difference between two successive time samples* is used as an approximation of the time-derivative everywhere in this paper.

In the implementation of Normalized EASI (24) in this work, $\mathfrak{h}(\cdot)$ is chosen as $\mathfrak{h}(\mathbf{y}) = \mathbf{y}^3$. In addition, the adaptation step λ_t in (24) is chosen as

$$\lambda_t = \begin{cases} 1/t, & 1 \le t \le 1000 \\ 1/1000, & 1000 < t \end{cases}. \qquad (43)$$

Even though a decreasing adaptation step (tending to zero as t moves forward) is usually taken in order to stabilize the algorithm after the convergence, in this case it does not go below a threshold. This is because the mixing matrix $\mathbf{J}_{\mathbf{x};t}$ is not constant and should be followed by the algorithm.

B. Simulation Results

Applying AATVL and BATIN algorithms on the observations, we get the results shown in Fig. 10 for the first simulation (mapping of Eq. (2)), and Fig. 11 for the second one (mapping of Eq. (40)). As expected, BATIN surpasses AATVL in estimating the separated sources in both simulations. Especially, the late convergence problem with AATVL has been almost completely resolved by BATIN.

Additionally, in order to see that adaptive linear BSS algorithms are not able to separate the sources (since the mixture is nonlinear), we have also implemented the same algorithm Normalized EASI for separating the mixture (40). It can be seen from Fig. 12 that the nonlinear mixture is not separated at all since EASI never converges.

C. Performance Evaluation

As mentioned earlier in Section II-A, unlike linear BSS where the sources may be estimated up to a scaling (and a permutation), in nonlinear problem, they can be estimated
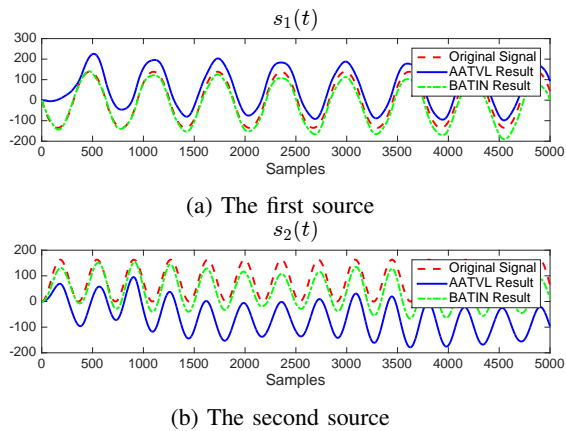
(a) The first source

(b) The second source

Fig. 11: The results of AATVL and BATIN algorithm in the mixture (40)
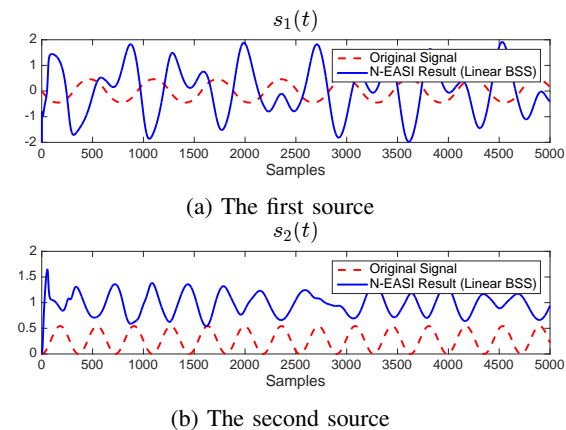

(a) The first source

(b) The second source

Fig. 12: The result of performing adaptive linear BSS (Normalized EASI method) on the sources which are mixed through (40)
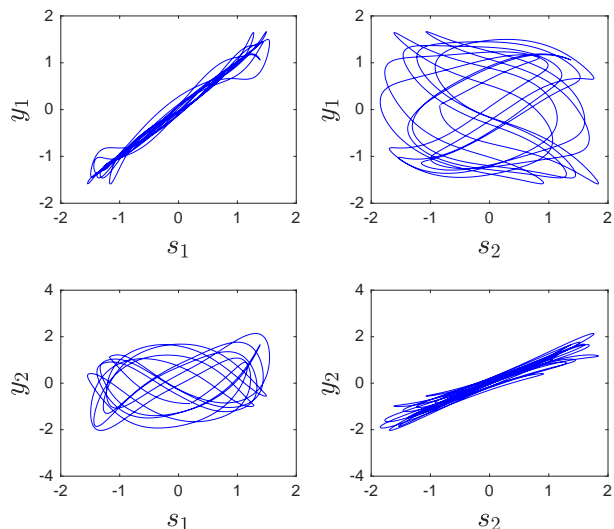

Fig. 13: The estimated sources $y_1(t)$ and $y_2(t)$ against the actual sources $s_1(t)$ and $s_2(t)$, where the thickness of a plot indicates how much the estimated signal (vertical axis) depends on the other source

up to a *nonlinear transformation* (and a permutation). Depending on the application, there should be some known characteristics of the sources (e.g. band-limited, sparse in some domain, bounded amplitude, and so forth) allowing the exact reconstruction of the sources. As a consequence, traditional performance index (e.g. Normalized RMS error) cannot be applied in nonlinear BSS.

Without loosing generality, assume that the sources are separated as $y_i(t) = c_i(s_i(t))$ for $i = 1, \ldots, n$ where c_i's are nonlinear functions. Therefore, the pairs $(s_i(t), y_i(t))$ for $t = 1 \ldots, T$ lie on a 1-dimensional manifold in a 2-dimensional space. However, if y_i depended on another source s_j ($i \neq j$), it would not be a mathematical function of s_i which would make the scatter plot of $(s_i(t), y_i(t))$ thick instead of a 1-dimentional manifold. This fact is also illustrated in Fig. 13. Since the pairs $(s_1(t), y_1(t))$ (similarly $(s_2(t), y_2(t))$) approximately lie on a 1-dimensional manifold, one concludes that y_1 (y_2) is only a function of s_1 (s_2).

If the separation is perfect, y_1 (y_2) will be exactly just a function of s_1 (s_2), hence the pairs $(s_1(t), y_1(t))$ (similarly $(s_2(t), y_2(t))$) exactly make a 1-dimensional manifold. The thicker the plot of the pairs $(s_i(t), y_i(t))$ is, the more separation error we have. So the *thickness* of the scatter plot indicates whether there is a dependence to another signal or not.

We thus propose this error as an index for evaluating the performance of a nonlinear BSS method. It can also be understood by modeling each output $y_i(t)$ as $y_i(t) = h_i(s_i(t)) +$ *interference*. This model highlights that the proposed index approximates the normalized interference to signal ratio of the output.

Although the thickness of data in linear 2-dimensional cases can be easily represented by the second eigenvalue of the autocorrelation matrix, it is not trivial in nonlinear problems. In this work, estimating the index, firstly a nonlinear curve is fitted to the data and then the RMS error of this fitting (similar to (37) but for a 1-dimensional manifold fitting) is introduced as the performance indicator (named as *Normalized Error of Nonlinear Fit (N-ENF)*). Normalized ENF of the ith source separation can be formulated as

$$\tilde{E}_{nenf} = \frac{\left(\sum_{t=1,\ldots,T} \left(\hat{c}_i(s_i(t)) - y_i(t) \right)^2 \right)^{\frac{1}{2}}}{\left(\sum_{t=1,\ldots,T} \left(\hat{c}_i(s_i(t)) \right)^2 \right)^{\frac{1}{2}}} \quad (44)$$

where $\hat{c}_i(s_i(t))$ is a nonlinear curve which is best fitted to the pairs $(s_i(t), y_i(t))$. In this work, the curve is fitted using smoothing splines [37] as

$$\underset{\hat{c}_i}{\text{minimize}} \sum_{t=1}^{T} \left(y_i(t) - \hat{c}_i(s_i(t)) \right)^2 +$$
$$\delta \sum_{t=1,\ldots,T} \left(\hat{c}_i''(s_i(t)) \right)^2 \quad (45)$$

where $\hat{c}_i''(s_i(t))$ is the second-order time-derivative of $\hat{c}_i(s_i(t))$ and δ is a smoothing parameter. It could have also been approximated by splitting the data to small bins and summing up the linear RMS errors over different bins.
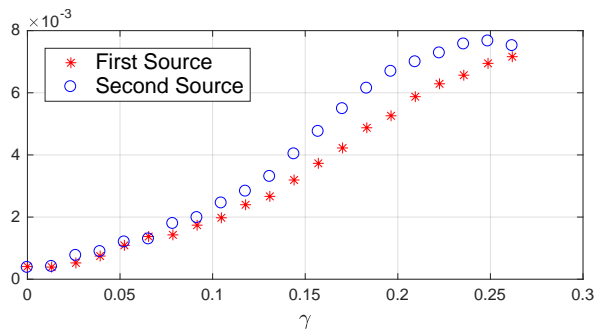
Fig. 14: The normalized ENF error in separating the mixture (2) for different levels of nonlinearity (represented by γ in (39)) using BATIN algorithm

Simulation results of the algorithms are also compared in terms of Normalized ENF error and can be found in table I.

TABLE I: N-ENF Error for AATVL and BATIN in the simulations

	AATVL	BATIN
N-ENF for the **Source 1** in the mixture (2) & (39)	0.0030	0.0019
N-ENF for the **Source 2** in the mixture (2) & (39)	0.0084	0.0031
N-ENF for the **Source 1** in the mixture (40)	0.0025	0.0023
N-ENF for the **Source 2** in the mixture (40)	0.0064	0.0040

These results show that the proposed idea is able to separate the sources that are mixed nonlinearly, which proves the proposed concept. However, as mentioned earlier, the performance of the proposed approach depends on the amount of the nonlinearity of the mixing function, i.e. as the mixing model gets distant from a linear mixture, the performance of the algorithm decreases. In order to show how the performance changes according to the nonlinearity level, a 3rd experiment is provided as follows.

Recall the example (2) with $\alpha(\mathbf{s}(t))$ defined as (39), letting $\alpha_0 = \pi/6$ and the parameter γ vary. In this example, if $\gamma = 0$, the mixture will be linear (a $\pi/6$ rotation). But as γ grows, the mixture will become more nonlinear. Thus γ can be considered as a level of nonlinearity of this parametric model.

Finally, the algorithm BATIN is employed for separating two sources of (41) and (42) mixed by (2), for different values of γ in (39). The normalized ENF error of BATIN for both sources is calculated and plotted in Fig. 14. Evidently, the more the mixture is nonlinear, the less efficient the proposed method is in separating the sources.

V. Conclusions and Perspectives

In this paper, a novel approach for performing nonlinear BSS is proposed. Through this approach, it is shown that the nonlinear mixtures are generally separable under a few assumptions (see subsection II-A). So the counter-examples provided in the literature to show that nonlinear mixtures are not separable, are not valid any more.

The key idea is to regard the time-derivative of the observed signals as a time-varying linear mixture of the (mutually independent) time derivatives of the sources. As a consequence,

the model (6) will be obtained, where the mixing matrix is a function of the sources (not to be confused with a time-variant mixing matrix which is a function of time).

Assuming both sources as functions of the time and non-linear mapping as a function of the sources to be smooth enough yields a sufficiently smooth mixing matrix which can be considered as a time-variant model (AATVL algorithm). However, the model (6) being a function of sources instead of conventional time-variant mixing models, enables performing the nonlinear regression (as explained in Section III-C) and dramatically improves the performance of the separation, which resulted in proposing the second algorithm (BATIN).

Once the sources are *separated*, BSS has been performed. However, aiming at exactly estimating the sources (not only separating them), the problem reduces to compensating an unknown nonlinear distortion. In other words, in order to precisely estimating the source signals (compensating the nonlinear function), each of the separated signals should be considered separately.

Numerous algorithms have been proposed for blind restoration of nonlinearly distorted signals (e.g. [33], [32]). The proposed methods are fundamentally based on retrieving some characteristics of the signal which are affected by nonlinear distortions. For example, nonlinear functions generally widen the bandwidth of signals. Thus, given a distorted band-limited signal, one may recover the original signal by trying to minimize its bandwidth via a nonlinear (compensating) function.

Moreover, assuming that the nonlinearly distorted signal is sparse in some domain, it can be blindly reconstructed [39], [31]. Since nonlinear distortions generally tend to reduce the sparsity, the proposed algorithms compensate the distortion via a sparse recovery procedure.

Nonetheless, depending on the application, there should be some known characteristics of the sources (e.g. band-limited, sparse in some domain, bounded amplitude, and so forth) allowing the exact reconstruction of the sources.

The basic idea proposed in this work is to utilize time-derivatives of the signals. Working with time-derivatives implicitly utilizes temporal information in the signals. This fact also supports the proposition in [27], which says that although we may mix two sources so that the mixtures are instantaneously independent of each other, it is highly probable that their delayed versions are not mutually independent when each of them is temporally correlated. In other words, it is implied in the paper that utilizing the temporal information of the sources may leads to solve nonlinear BSS problems.

It is worth noting that the proposed idea is quite different with respect to the previous works in the literature on nonlinear mixtures; it is more theoretic and general and does not assume any specific mixing model or source signals. Two basic methods, AATVL and BATIN are provided in this work to show how the idea is to be employed. Nevertheless, many different separation algorithms can be suggested based on the proposed approach and they can be optimized to deal with more complex signals/mixtures of practical applications.

However, there are several issues to be considered in the future. Firstly, the statistical characteristics of the derivative of a signal with respect to those of the signal, itself, should be investigated. This might be the key to better understanding of the key feature of derivatives that lets perform the separation, and accordingly, it may lead to new algorithms of nonlinear BSS.

Secondly, the "Nonlinear Regression" used in the proposed algorithm should be improved. The main objective of this step is to accumulate the information of the separation at each sample. For example, if at two different times, the source vector takes the same value, the mixing matrix will remain the same as well.

The problem in this work is considered in the simplest form where there is no noise added to the signals. Since all the signals in practical applications are noisy, and considering the fact that taking the derivatives may dramatically amplifies the noise, new methods should be developed which are more robust to noise. It may also enforce some modifications on "Adaptive Linear BSS" procedure of the algorithms as well.

Last but not least, finding out the relations between autocorrelation functions of the sources (i.e. how much colored they are) and the performance of the proposed approach and trying to quantify it is also an interest for future studies.

REFERENCES

[1] J. Hérault and C. Jutten, "Space or time adaptive signal processing by neural network models," in *AIP Conference Proceedings 151 on Neural Networks for Computing*. Woodbury, NY, USA: American Institute of Physics Inc., 1987, pp. 206–211. [Online]. Available: http://dl.acm.org/citation.cfm?id=24140.24171

[2] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent component analysis and applications*. Academic press, 2010.

[3] T. Mei, F. Yin, and J. Wang, "Blind source separation based on cumulants with time and frequency non-properties," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 6, pp. 1099–1108, 2009.

[4] P. Comon, "Independent component analysis, a new concept?" *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[5] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural computation*, vol. 7, no. 6, pp. 1129–1159, 1995.

[6] J. F. Cardoso and A. Souloumiac, "An efficient technique for the blind separation of complex sources." in *Proceeding of IEEE Signal Processing Workshop on Higher-Order Statistics, Lake Tahoe*, 1993, pp. 275–279.

[7] J. F. Cardoso and B. H. Laheld, "Equivariant adaptive source separation," *IEEE Transactions on Signal Processing*, vol. 44, no. 12, pp. 3017–3030, 1996.

[8] L. De Lathauwer, B. De Moor, and J. Vandewalle, "On the best rank-1 and rank-(r 1, r 2,..., rn) approximation of higher-order tensors," *SIAM Journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1324–1342, 2000.

[9] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.

[10] L. Tong, V. C. Soon, Y. F. Huang, and R. Liu, "AMUSE: a new blind identification algorithm," in *IEEE International Symposium on Circuits and Systems*, May 1990, pp. 1784–1787 vol.3.

[11] L. Tong, R. W. Liu, V. C. Soon, and Y. F. Huang, "Indeterminacy and identifiability of blind identification," *IEEE Transactions on Circuits and Systems*, vol. 38, no. 5, pp. 499–509, 1991.

[12] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 434–444, 1997.

[13] N. Dobigeon, J.-Y. Tourneret, C. Richard, J. C. M. Bermudez, S. McLaughlin, and A. O. Hero, "Nonlinear unmixing of hyperspectral images: Models and algorithms," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 82–94, 2014.

[14] M. Golbabaee, S. Arberet, and P. Vandergheynst, "Compressive source separation: Theory and methods for hyperspectral imaging," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 5096–5110, 2013.

[15] I. Meganem, P. Deliot, X. Briottet, Y. Deville, and S. Hosseini, "Physical modelling and non-linear unmixing method for urban hyperspectral images," in *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 3rd Workshop on*. IEEE, 2011, pp. 1–4.

[16] L. T. Duarte and C. Jutten, "Design of smart ion-selective electrode arrays based on source separation through nonlinear independent component analysis," *Oil & Gas Science and Technology–Revue dIFP Energies nouvelles*, vol. 69, no. 2, pp. 293–306, 2014.

[17] 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 (IJDAR)*, vol. 14, no. 4, pp. 319–333, 2011.

[18] M. Babaie-Zadeh, "On blind source separation in convolutive and nonlinear mixtures," Ph.D. dissertation, Grenoble, INPG, 2002.

[19] T. Blaschke, T. Zito, and L. Wiskott, "Independent slow feature analysis and nonlinear blind source separation," *Neural computation*, vol. 19, no. 4, pp. 994–1021, 2007.

[20] D. N. Levin, "Performing nonlinear blind source separation with signal invariants," *IEEE Transactions on Signal Processing*, vol. 58, no. 4, pp. 2131–2140, 2010.

[21] A. Taleb and C. Jutten, "Source separation in post-nonlinear mixtures," *IEEE Transactions on Signal Processing*, vol. 47, no. 10, pp. 2807–2820, 1999.

[22] Y. Altmann, A. Halimi, N. Dobigeon, and J. Y. Tourneret, "Supervised nonlinear spectral unmixing using a postnonlinear mixing model for hyperspectral imagery," *IEEE Transactions on Image Processing*, vol. 21, no. 6, pp. 3017–3025, 2012.

[23] A. Halimi, Y. Altmann, N. Dobigeon, and J. Y. Tourneret, "Nonlinear unmixing of hyperspectral images using a generalized bilinear model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4153–4162, 2011.

[24] Y. Deville and L. T. Duarte, "An overview of blind source separation methods for linear-quadratic and post-nonlinear mixtures," in *International Conference on Latent Variable Analysis and Signal Separation*. Springer, 2015, pp. 155–167.

[25] A. Hyvärinen and P. Pajunen, "Nonlinear independent component analysis: Existence and uniqueness results," *Neural Networks*, vol. 12, no. 3, pp. 429–439, 1999.

[26] A. M. Kagan, Y. V. Linnik, and C. R. Rao, "Extension of darmois-skitcvic theorem to functions of random variables satisfying an addition theorem," *Communications in Statistics-Theory and Methods*, vol. 1, no. 5, pp. 471–474, 1973.

[27] 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, 2003.

[28] C. Jutten and J. Karhunen, "Advances in nonlinear blind source separation," in *Proceeding of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003)*, 2003, pp. 245–256.

[29] ——, "Advances in blind source separation (BSS) and independent component analysis (ICA) for nonlinear mixtures," *International Journal of Neural Systems*, vol. 14, no. 05, pp. 267–292, 2004.

[30] B. Ehsandoust, B. Rivet, C. Jutten, and M. Babaie-Zadeh, "Nonlinear blind source separation for sparse sources," in *2016 24th European Signal Processing Conference (EUSIPCO)*, Aug 2016, pp. 1583–1587.

[31] L. T. Duarte, R. Suyama, R. Attux, J. M. T. Romano, and C. Jutten, "A sparsity-based method for blind compensation of a memoryless nonlinear distortion: Application to ion-selective electrodes," *IEEE Sensors Journal*, vol. 15, no. 4, pp. 2054–2061, 2015.

[32] K. Dogancay, "Blind compensation of nonlinear distortion for bandlimited signals," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 52, no. 9, pp. 1872–1882, 2005.

[33] F. arvasti and A. K. Jain, "Zero crossings, bandwidth compression, and restoration of nonlinearly distorted band-limited signals," *Journal of Optical Society of America A*, vol. 3, no. 5, pp. 651–654, May 1986.

[34] B. Ehsandoust, M. Babaie-Zadeh, and C. Jutten, *Blind Source Separation in Nonlinear Mixture for Colored Sources Using Signal Derivatives*. Springer International Publishing, 2015, pp. 193–200.

[35] M. Spivak, *Calculus on Manifolds: A Modern Approach to Classical Theorems of Advanced Calculus*, ser. Advanced book program. Avalon Publishing, 1965.

[36] J. B. Tenenbaum, V. d. Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[37] C. H. Reinsch, "Smoothing by spline functions," *Numerische mathematik*, vol. 10, no. 3, pp. 177–183, 1967.

[38] C. De Boor, *A practical guide to splines*. Springer-Verlag New York, 1978, vol. 27.

[39] J. Malek, "Blind compensation of memoryless nonlinear distortions in sparse signals," in *21st European Signal Processing Conference (EUSIPCO 2013)*, Sept 2013, pp. 1–5.