

A Second-Order Statistics Method for Blind Source Separation in Post-Nonlinear Mixtures

Denis G. Fantinato^{a,*}, Leonardo T. Duarte^b, Yannick Deville^c, Romis Attux^d,
Christian Jutten^e, Aline Neves^f

^a*Mathematics, Computation and Cognition Center (CMCC), Federal University of ABC,
Santo André/SP, Brazil*

^b*School of Applied Sciences, University of Campinas, Limeira/SP, Brazil*

^c*Université de Toulouse, UPS, CNRS, CNES, IRAP, 31400 Toulouse, France*

^d*School of Electrical and Computer Engineering (FEEC), University of Campinas,
Campinas/SP, Brazil*

^e*GIPSA-Lab, Grenoble INP, CNRS, Grenoble, France*

^f*Engineering, Modeling and Applied Social Science Center (CECS), Federal University of
ABC, Santo André/SP, Brazil*

Abstract

In the context of nonlinear Blind Source Separation (BSS), the Post-Nonlinear (PNL) model is of great importance due to its suitability for practical nonlinear problems. Under certain mild constraints on the model, Independent Component Analysis (ICA) methods are valid for performing source separation, but requires use of Higher-Order Statistics (HOS). Conversely, regarding the sole use of the Second-Order Statistics (SOS), their study is still in an initial stage. In that sense, in this work, the conditions and the constraints on the PNL model for SOS-based separation are investigated. The study encompasses a time-extended formulation of the PNL problem with the objective of extracting the temporal structure of the data in a more extensive manner, considering SOS-based methods for separation, including the proposition of a new one. Based on this, it is shown that, under some constraints on the nonlinearities and if a

[☆]This work was partially supported by CNPq, FAPESP (2017/11488-5,2015/23424-6) and ERC project 2012-ERC-AdG-320684 CHESS.

*Corresponding author

Email addresses: denis.fantinato@ufabc.edu.br (Denis G. Fantinato),
leonardo.duarte@fca.unicamp.br (Leonardo T. Duarte), yannick.deville@irap.omp.eu
(Yannick Deville), attux@dca.fee.unicamp.br (Romis Attux),
christian.jutten@gipsa-lab.grenoble-inp.fr (Christian Jutten),
aline.neves@ufabc.edu.br (Aline Neves)

given number of time delays is considered, source separation can be successfully achieved, at least for polynomial nonlinearities. With the aid of metaheuristics called Differential Evolution and Clonal Selection Algorithm for optimization, the performances of the SOS-based methods are compared in a set of simulation scenarios, in which the proposed method shows to be a promising approach.

Keywords: Blind Source Separation, Post-Nonlinear, Second-Order Statistics

1. Introduction

In many practical applications, retrieving a set of source signals from some observations that are actually mixtures of these sources can be of great relevance. If there is no (or incomplete) prior knowledge about the sources and mixing
5 transform, the problem is referred to as Blind Source Separation (BSS) [1, 2, 3]. Throughout three decades of existence, this problem has been subject of great attention from the academic community, where the initial efforts were mainly aimed at the standard linear and instantaneous mixing model, with the assumption that the sources are statistically mutually independent. Indeed, the study
10 of this topic contributed to a solid theoretical framework known as Independent Component Analysis (ICA) [1], which provided methods that explicitly use Higher-Order Statistics (HOS), such as FastICA, JADE and Infomax [1, 4].

Although the mutual independence assumption is sufficient for linear separation, the use of the HOS may result in some estimation difficulties. However,
15 in certain cases, additional information about the sources can be explored to lead to simpler methods [5]. For instance, the sources may exhibit temporal structures (i.e., temporal dependence among samples), which can be statistically characterized e.g. by temporal correlations, being thus possible to perform source separation based only on Second-Order Statistics (SOS), if there is sufficient temporal diversity [1]. In this case, even Gaussian sources, which are not
20 separable using HOS methods, can be separated. This perspective motivated the proposal of diverse SOS-based methods, such as SOBI [6], WASOBI [7], AMUSE, TDSEP, among others [1, 4]

Regarding the nonlinear BSS problem, there has been a considerable effort
25 for extending the ICA framework and other BSS methods to nonlinear mixing
models, in view of promising applications like smart chemical sensor arrays [8],
hyperspectral imaging [9, 10] and quantum bit uncoupling [11]. However, from
a general standpoint, mutual independence may not be sufficient for performing
nonlinear separation [1]. Thus, except recent works [12] which propose a new
30 general framework for nonlinear source separation, the studies on this topic were
focused on a constrained set of nonlinear models in which the ICA methods are
still valid [13]. For instance, the Post-Nonlinear (PNL) mixing model is com-
posed of an instantaneous linear stage followed by a set of univariate nonlinear
functions, which must obey some constraints (e.g., monotonicity) to enable the
35 use ICA-based methods for separation [14].

However, when dealing with temporally structured sources in nonlinear mod-
els, there is still no parallel with the linear case [5, 13], especially because the
study based on the sole use of the SOS is very incipient, raising fundamental
issues like conditions for identifiability and uniqueness of solutions. In view of
40 this, in this work, we focus on the analysis of SOS-based methods in the PNL
mixing problem, prompting to outline the set of constraints and conditions to
ensure the validity of the approach. This initial step is made by considering the
case in which the nonlinearities belong to a class of cubic polynomials, revealing
promising perspectives – including the nonlinear separation of Gaussian sources.

45 The SOS-based investigation to be followed is performed considering block-
structured correlation matrices with a set of time delayed samples, being able
to encompass the temporal structure of data more organically [15]. However,
the potential nonlinear statistical dependencies it carries demands a suitable
treatment of the statistical information, which will be attained through the
50 proposition of a novel SOS-based criterion. Besides these contributions, we con-
sider an extended formulation of the PNL problem, for which the manipulation
of the block-structured matrices is more straightforward. This will allow the
analytical computation of the considered SOS-based cost functions, providing a
rich theoretical analysis.

55 Although SOS-based methods are usually associated with mathematical simplicity, the nonlinear context may lead to inherent complex multimodal SOS-based cost functions and hence, high computational cost. In that sense, a proper exploration of the search space can be performed by metaheuristics. In this work, the metaheuristics Differential Evolution (DE) [16] and Clonal Selection
60 Algorithm (CLONALG) [17] are considered for parameter optimization in order to avoid local convergence.

This work is organized as follows. In Section 2, the PNL model is presented considering the temporal-extended formulation and the particular nonlinear case. Section 3 describes the SOS-based criteria to be used, introducing
65 the proposed method; the analytical computation of the covariance matrices is presented as well. The identifiability conditions and bounds are analyzed in Section 4 and the performance results are shown in Section 5. Finally, Section 6 concludes the work.

2. The Post-Nonlinear Mixing Model

70 In the problem of Blind Source Separation (BSS), the main objective is to recover the original sources $\mathbf{s}(n)$ from observed mixtures $\mathbf{x}(n) = \Phi(\mathbf{s}(n))$, with $\mathbf{x}(n) = [x_1(n), \dots, x_M(n)]^T$ the observation vector with M mixtures, $\mathbf{s}(n) = [s_1(n), \dots, s_N(n)]^T$ the vector with N source signals at time instant n and $\Phi(\cdot)$ the mixing mapping [1]. Generally, $\Phi(\cdot)$ is assumed to be linear and instantaneous,
75 however, this approach may not be adequate for certain applications and a nonlinear model must be considered, such as the Post-Nonlinear (PNL) [14].

The PNL structure is particularly interesting because it sequentially combines linear and nonlinear stages, as shown in Fig. 1. Due to their relative simplicity, the mixtures can be mathematically written as $\mathbf{x}(n) = \mathbf{f}(\mathbf{A}\mathbf{s}(n))$,
80 where \mathbf{A} is an $M \times N$ matrix and $\mathbf{f}(\cdot)$ is a set of M component-wise functions. The separation system is the mirrored version of the mixing system, with output given by $\mathbf{y}(n) = \mathbf{W}\mathbf{g}(\mathbf{x}(n))$, where \mathbf{W} is an $N \times M$ matrix and $\mathbf{g}(\cdot)$ is a set of M component-wise functions [1]. In this work, we focus on the case in which

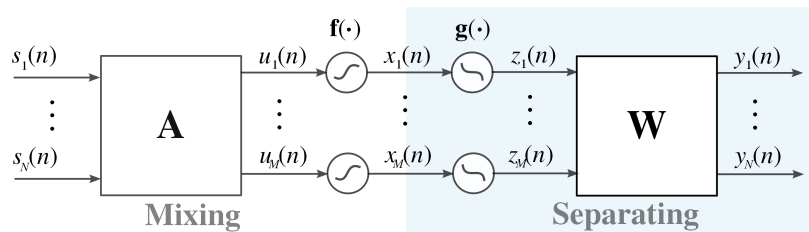


Figure 1: Mixing and separating systems in the PNL model.

$N = M$ (determined case), i.e., when the number of sources and of mixtures
 85 are equal (for a single source, the PNL model can establish important analogies
 with Wiener Hammerstein systems, due to their similarity [18]).

Under certain constraints on the nonlinear functions $\mathbf{f}(\cdot)$ and $\mathbf{g}(\cdot)$, it is known
 that the (independent) sources can be separated by ICA methods, which requires
 the use of HOS [1, 19]. However, under the additional assumption that the
 90 sources are temporally colored, the viability and the conditions for separation
 of PNL mixtures using only Second-Order Statistics (SOS) still remain open
 questions. In that sense, in this work, we search for these answers by considering
 a class of PNL mixtures to be separated by SOS-based methods. However, the
 temporal information will be extracted by block-structured correlation matrices
 95 with an arbitrary number of time delays. To facilitate the manipulation of
 these matrices, we formulate a temporal-extended version of the PNL problem,
 presented in the following.

2.1. Time-Dependent Sources in the PNL Model

The temporal structure in the sources is usually seen as the inherent result
 100 of the system which generates them. However, in some applications, it can be
 modeled as the result of independent and identically distributed (*i.i.d.*) signals
 processed by linear or nonlinear systems, whose signature is the temporal struc-
 ture imprinted on the signals. For simplicity, in this work, to investigate the
 (SOS) features in the PNL models, we restrain ourselves to the case in which
 105 the temporal structure is obtained by means of a linear filtering system.

To suitably describe the temporal structure in the sources, we consider vectors with the N sources at time instant n concatenated with d delayed versions of them in the following form (similarly to [15]):

$$\begin{aligned}\underline{\mathbf{s}}(n) &= [s_1(n), \dots, s_1(n-d), s_2(n), \dots, s_2(n-d), \dots, s_N(n), \dots, s_N(n-d)]^T \\ &= [\underline{\mathbf{s}}_1^T(n), \underline{\mathbf{s}}_2^T(n), \dots, \underline{\mathbf{s}}_N^T(n)]^T,\end{aligned}\quad (1)$$

where d is the maximum considered time delay and $\underline{\mathbf{s}}_i(n) = [s_i(n), \dots, s_i(n-d)]^T$, for $i = \{1, \dots, N\}$. We wish to express the time-extended sources $\underline{\mathbf{s}}(n)$ as functions of *i.i.d* signals $\underline{\mathbf{r}}(n)$ (note that all the underlined variables are time-extended versions of the classical formulation, similarly to $\underline{\mathbf{s}}(n)$ and $\mathbf{s}(n)$). In order to do so, we consider a set of N Finite Impulse Response (FIR) filters that are responsible for introducing correlation in the signals $\underline{\mathbf{s}}(n)$ (i.e., $\underline{\mathbf{s}}(n)$ are composed of moving average (MA) processes [20]). The coefficients of each FIR filter are arranged in vectors \mathbf{h}_i , for $i = \{1, \dots, N\}$. Hence, for instance, an FIR filter with transfer function $H_i(z) = h_{i,0} + h_{i,1}z^{-1} + \dots + h_{i,L_{h_i}}z^{-L_{h_i}}$ is represented by the vector $\mathbf{h}_i = [h_{i,0}, h_{i,1}, \dots, h_{i,L_{h_i}}]$. Based on this, we define

$$\mathbf{H}_i = \mathbf{h}_i \triangleright \mathbf{I}_{d+1} = \begin{bmatrix} h_{i,0} & \cdots & h_{i,L_{h_i}} & 0 & 0 & \cdots & 0 \\ 0 & h_{i,0} & \cdots & h_{i,L_{h_i}} & 0 & \cdots & 0 \\ \vdots & \ddots & & \ddots & & \ddots & \vdots \\ 0 & \cdots & 0 & h_{i,0} & \cdots & h_{i,L_{h_i}} & 0 \\ 0 & \cdots & 0 & 0 & h_{i,0} & \cdots & h_{i,L_{h_i}} \end{bmatrix}, \quad (2)$$

in which $\mathbf{h}_i \triangleright \mathbf{I}_{d+1}$ is the diagonal replication of vector \mathbf{h}_i , being the resulting matrix \mathbf{H}_i of dimension $(d+1) \times (L_{h_i} + d + 1)$, and \mathbf{I}_{d+1} the identity matrix of size $d+1$.

For the sake of simplicity, we assume henceforth $N = 2$ sources without loss of generality. In this case, the sources $\underline{\mathbf{s}}(n)$ can be written as functions of $\underline{\mathbf{r}}(n)$:

$$\underline{\mathbf{s}}(n) = \underline{\mathcal{H}}\underline{\mathbf{r}}(n) = \begin{bmatrix} \mathbf{H}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_2 \end{bmatrix} \underline{\mathbf{r}}(n), \quad (3)$$

where $\underline{\mathcal{H}}$ is a block-diagonal matrix with dimensions $N(d+1) \times (\sum_{i=1}^N (L_{h_i} + d + 1))$
110 and $\underline{\mathbf{r}}(n) = [r_1(n), r_1(n-1), \dots, r_1(n-L_{h_1}-d), r_2(n), r_2(n-1), \dots, r_2(n-L_{h_2}-d)]^T$

is the temporal-extended *i.i.d.* vector with the original signals $r_1(n)$, $r_2(n)$ and their delayed versions. Note that the samples of each *i.i.d.* signal $r_i(n)$ considered separately will be combined by $\underline{\mathcal{H}}$, but there is no mixing between $r_1(n)$ and $r_2(n)$. Eq. (3) shall be useful for analytically computing certain statistical
115 moments, since it express the sources $s_i(n)$ as functions of *i.i.d.* signals.

Proceeding with the PNL temporal-extended formulation, we can write the linear mixtures as $\underline{\mathbf{u}}(n) = \underline{\mathcal{A}}\underline{\mathbf{s}}(n)$, with the extended linear mixing matrix $\underline{\mathcal{A}}$:

$$\underline{\mathcal{A}} = \begin{bmatrix} a_{11}\mathbf{I}_{d+1} & a_{12}\mathbf{I}_{d+1} \\ a_{21}\mathbf{I}_{d+1} & a_{22}\mathbf{I}_{d+1} \end{bmatrix}, \quad (4)$$

in which each element is replicated along a diagonal of a (sub)matrix of size $d+1$. The time-extended observations (mixtures) $\underline{\mathbf{x}}(n)$ can be written as

$$\underline{\mathbf{x}}(n) = \underline{\mathcal{F}}(\underline{\mathbf{u}}(n)) = \underline{\mathcal{F}}(\underline{\mathcal{A}}\underline{\mathbf{s}}(n)), \quad (5)$$

where $\underline{\mathcal{F}}(\cdot)$ is a set of functions diagonally positioned as

$$\underline{\mathcal{F}}(\cdot) = \begin{bmatrix} f_1(\cdot) \odot \mathbf{I}_{d+1} & \mathbf{0} \\ \mathbf{0} & f_2(\cdot) \odot \mathbf{I}_{d+1} \end{bmatrix} = \begin{bmatrix} f_1(\cdot) & 0 & 0 & & & \\ 0 & \ddots & 0 & & \mathbf{0} & \\ 0 & 0 & f_1(\cdot) & & & \\ & & & f_2(\cdot) & 0 & 0 \\ & \mathbf{0} & & 0 & \ddots & 0 \\ & & & 0 & 0 & f_2(\cdot) \end{bmatrix}, \quad (6)$$

in which $f(\cdot) \odot \mathbf{I}_{d+1}$ is the diagonal replication of function $f(\cdot)$.

The separating system is a mirrored version of the mixing one, with output

$$\underline{\mathbf{y}}(n) = \underline{\mathcal{W}}\underline{\mathcal{G}}(\underline{\mathbf{x}}(n)), \quad (7)$$

where $\underline{\mathcal{G}}(\cdot)$ and $\underline{\mathcal{W}}$ have structures similar to $\underline{\mathcal{F}}(\cdot)$ and $\underline{\mathcal{A}}$, respectively.

By combining Eqs. (7), (5) and (3), we are able to directly express the separated sources $\underline{\mathbf{y}}(n)$ as functions of $\underline{\mathbf{r}}(n)$ as

$$\underline{\mathbf{y}}(n) = \underline{\mathcal{W}}\underline{\mathcal{G}}(\underline{\mathcal{F}}(\underline{\mathcal{A}}\underline{\mathcal{H}}\underline{\mathbf{r}}(n))). \quad (8)$$

Undoubtedly, in practical scenarios, the elements $\underline{\mathcal{F}}$, $\underline{\mathcal{A}}$, $\underline{\mathcal{H}}$ and $\underline{\mathbf{r}}(n)$ are considered unknown and the separation task may be performed relying on, for instance, some statistical properties of the sources $\underline{\mathbf{s}}(n)$, like the mutual independence. Notwithstanding, Eq. (8) is of great theoretical importance, since it exposes a direct relation to *i.i.d.* signals and opens the way for the analytical computation of the statistics involved in the separation process, as we intend to show. It is important to note that some additional assumptions may be necessary, for example, the definition of the type of the nonlinearities $\underline{\mathcal{F}}(\cdot)$ and $\underline{\mathcal{G}}(\cdot)$.

2.2. A Special Case: The Cubic Nonlinearity

In order to find a subset of constrained PNL models in which the SOS-based methods are sufficient for separation, we start from a simple hypothesis that the combined nonlinear function $\underline{\mathcal{G}} \circ \underline{\mathcal{F}}$ yields as output

$$\underline{\mathbf{z}}(n) = \underline{\mathcal{A}}\underline{\mathbf{s}}(n) + \underline{\mathbf{\Gamma}}(\underline{\mathcal{A}}\underline{\mathbf{s}}(n))^{\odot 3}, \quad (9)$$

where

$$\underline{\mathbf{\Gamma}} = \begin{bmatrix} \gamma_1 \mathbf{I}_{d+1} & \mathbf{0} \\ \mathbf{0} & \gamma_2 \mathbf{I}_{d+1} \end{bmatrix} \quad (10)$$

and $(\cdot)^{\odot 3}$ is the Hadamard power of 3 (i.e., an element-wise cubic operator). This can be viewed as, for instance, the combination of a cubic nonlinearity $\underline{\mathcal{F}}(\underline{\mathbf{u}}(n)) = \underline{\mathbf{u}}^{\odot 3}(n)$ and a $\underline{\mathcal{G}}(\underline{\mathbf{x}}(n)) = \text{sgn}(\underline{\mathbf{x}}(n)) \odot (|\underline{\mathbf{x}}(n)|)^{\odot 1/3} + \underline{\mathbf{\Gamma}}\underline{\mathbf{x}}(n)$. Based on Eq. (8), the system output can now be written as

$$\underline{\mathbf{y}}(n) = \underline{\mathcal{W}}\underline{\mathbf{z}}(n) = \underline{\mathcal{W}}\underline{\mathcal{A}}\underline{\mathcal{H}}\underline{\mathbf{r}}(n) + \underline{\mathcal{W}}\underline{\mathbf{\Gamma}}(\underline{\mathcal{A}}\underline{\mathcal{H}}\underline{\mathbf{r}}(n))^{\odot 3}. \quad (11)$$

Hence, according to this model, two observations can be outlined: (i) the separated sources can be viewed as the combination of a linear mixing/demixing term and a nonlinear mixing/demixing term; (ii) if $\underline{\mathbf{\Gamma}} = \mathbf{0}$, then the second (nonlinear) term vanishes and the problem is reduced to the linear one.

Very interestingly, each element of $(\underline{\mathcal{A}}\underline{\mathcal{H}}\underline{\mathbf{r}}(n))^{\odot 3}$ can be viewed as a polynomial raised to the power of 3, which can be expanded and rearranged in a matrix form. For instance, in a hypothetical simple case with constants c_1 and

c_2 , we intend to perform the following rearrangement: $(c_1 r_1(n) + c_2 r_1(n-1))^3 = c_1^3 r_1^3(n) + c_2^3 r_1^3(n-1) + 3c_1^2 c_2 r_1^2(n) r_1(n-1) + 3c_1 c_2^2 r_1(n) r_1^2(n-1) = [c_1^3, c_2^3, 3c_1^2 c_2, 3c_1 c_2^2] [r_1^3(n), r_1^3(n-1), r_1^2(n) r_1(n-1), r_1(n) r_1^2(n-1)]^T$, where the constant terms and the signals $r_1(n)$ and $r_1(n-1)$ are separated in different vectors. This procedure can be extended to all elements of $(\underline{\mathcal{A}}\mathbf{H}\mathbf{r}(n))^{\odot 3}$, but, all considered signals in $\mathbf{r}(n)$ (i.e., $r_1(n), \dots, r_1(n-L_{h_1}-d), r_2(n), \dots, r_2(n-L_{h_2}-d)$) must be taken into account, resulting

$$\begin{aligned}
(\underline{\mathcal{A}}\mathbf{H}\mathbf{r}(n))^{\odot 3} &= \begin{bmatrix} (a_{11}h_{1,0})^3 r_1^3(n) + (a_{11}h_{1,1})^3 r_1^3(n-1) + \dots \\ \vdots \\ (a_{11}h_{1,0})^3 r_1^3(n-d) + (a_{11}h_{1,1})^3 r_1^3(n-d-1) + \dots \\ (a_{21}h_{1,0})^3 r_1^3(n) + (a_{21}h_{1,1})^3 r_1^3(n-1) + \dots \\ \vdots \\ (a_{21}h_{1,0})^3 r_1^3(n-d) + (a_{21}h_{1,1})^3 r_1^3(n-d-1) + \dots \end{bmatrix} \\
&= \begin{bmatrix} [\xi_{11}, \xi_{12}, \dots] [r_1^3(n), r_1^3(n-1), \dots]^T \\ \vdots \\ [\xi_{11}, \xi_{12}, \dots] [r_1^3(n-d), r_1^3(n-d-1), \dots]^T \\ [\xi_{21}, \xi_{22}, \dots] [r_1^3(n), r_1^3(n-1), \dots]^T \\ \vdots \\ [\xi_{21}, \xi_{22}, \dots] [r_1^3(n-d), r_1^3(n-d-1), \dots]^T \end{bmatrix}, \tag{12}
\end{aligned}$$

where ξ_{ij} corresponds to the factor that multiplies the term involving the original signals $r_1(n)$, $r_2(n)$ and/or their delayed versions, for $i = 1, \dots, N$ (recall that it is assumed $N = 2$) and for $j = 1, \dots, L_v$, with L_v equal to the resulting number of terms after the cubic expansion. The number of elements L_v can be combinatorially obtained: assuming that all FIR filters have the same maximum length L_h+1 , without loss of generality, we have that $L_v = (N(L_h+1)+2)!/(3!(N(L_h+1)-1)!)$. Due to the excessive length of the vectors, we have shown only the initial terms.

Assuming now that $\boldsymbol{\theta}_i = [\xi_{i1}, \xi_{i2}, \dots, \xi_{iL_v}]$ is a row vector with L_v elements, and that $\boldsymbol{\rho}(n) = [r_1^3(n), r_1^3(n-1), r_2^3(n), r_2^3(n-1), \dots, r_1^2(n)r_2(n-1), \dots, r_1(n)r_2(n)r_2(n-1), \dots]^T$ is the column vector with L_v terms involving the sig-

nals $r_1(n)$, $r_2(n)$ and their delayed versions, Eq. (12) can be rewritten as

$$(\underline{\mathcal{A}\mathcal{H}\mathbf{r}}(n))^{\odot 3} = \begin{bmatrix} \boldsymbol{\theta}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\theta}_1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\theta}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{\theta}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\rho}(n) \\ \boldsymbol{\rho}(n-1) \\ \vdots \\ \boldsymbol{\rho}(n-d) \end{bmatrix} = \underline{\boldsymbol{\Theta}}\underline{\boldsymbol{\rho}}(n), \quad (13)$$

where $\underline{\boldsymbol{\rho}}(n)$ is the column vector with $\boldsymbol{\rho}(n)$ and its d delayed versions – so that
140 $\underline{\boldsymbol{\rho}}(n)$ has length $(d+1)(N(L_h+1)+2)!/(3!(N(L_h+1)-1)!)$ – and $\underline{\boldsymbol{\Theta}}$ is a matrix with dimensions $N(d+1) \times (d+1)(N(L_h+1)+2)!/(3!(N(L_h+1)-1)!)$. Note that the computational complexity can increase drastically, depending on the values of N (here chosen to be 2), d and L_h .

It is interesting to note in Eq. (13) that $\underline{\boldsymbol{\rho}}(n)$ encompasses elements of $\mathbf{r}(n)$ up
145 to the power of 3 (due to the assumed cubic nonlinearity) but we have expressed it by means of a linear matrix multiplication, i.e., $(\underline{\mathcal{A}\mathcal{H}\mathbf{r}}(n))^{\odot 3} = \underline{\boldsymbol{\Theta}}\underline{\boldsymbol{\rho}}(n)$.

As an example, we consider two linear FIR filters $\mathbf{h}_1 = [h_{1,0}, h_{1,1}]$ and $\mathbf{h}_2 = [h_{2,0}, h_{2,1}]$ and the mixtures $\mathbf{x}(n) = (\mathbf{A}\mathbf{s}(n))^{\odot 3}$, with $\mathbf{A} = [a_{11}, a_{12}; a_{21}, a_{22}]$. If we consider $d = 1$, the mixtures can be expressed as

$$\begin{aligned} \underline{\mathbf{x}}(n) &= \mathcal{F}(\underline{\mathbf{A}\mathbf{s}}(n)) = (\underline{\mathcal{A}\mathcal{H}\mathbf{r}}(n))^{\odot 3} \\ &= \left(\begin{bmatrix} a_{11} & 0 & a_{12} & 0 \\ 0 & a_{11} & 0 & a_{12} \\ a_{21} & 0 & a_{22} & 0 \\ 0 & a_{21} & 0 & a_{22} \end{bmatrix} \begin{bmatrix} h_{1,0} & h_{1,1} & 0 & 0 & 0 & 0 \\ 0 & h_{1,0} & h_{1,1} & 0 & 0 & 0 \\ 0 & 0 & 0 & h_{2,0} & h_{2,1} & 0 \\ 0 & 0 & 0 & 0 & h_{2,0} & h_{2,1} \end{bmatrix} \cdot \left[r_1(n), r_1(n-1), r_1(n-2), r_2(n), r_2(n-1), r_2(n-2) \right]^T \right)^{\odot 3}. \end{aligned} \quad (14)$$

By proceeding with the cubic (or Volterra) expansion, the vector $\boldsymbol{\rho}(n)$ has $L_v = (N(L_h+1)+2)!/(3!(N(L_h+1)-1)!)$ = 20 elements, which are all possible triplets

among $r_1(n)$, $r_1(n-1)$, $r_2(n)$ and $r_2(n-1)$, i.e.,

$$\begin{aligned} \boldsymbol{\rho}(n) = [& r_1^3(n), & r_1^3(n-1), & r_2^3(n), & r_2^3(n-1), \\ & r_1^2(n)r_1(n-1), & r_1^2(n)r_2(n), & r_1^2(n)r_2(n-1), & r_1(n)r_1^2(n-1), \\ & r_1(n)r_2^2(n), & r_1(n)r_2^2(n-1), & r_1^2(n-1)r_2(n), & r_1^2(n-1)r_2(n-1), \\ & r_1(n-1)r_2^2(n), & r_1(n-1)r_2^2(n-1), & r_2^2(n)r_2(n-1), & r_2(n)r_2^2(n-1), \\ & r_1(n)r_1(n-1)r_2(n), & r_1(n)r_1(n-1)r_2(n-1), & r_1(n)r_2(n)r_2(n-1), \\ & r_1(n-1)r_2(n)r_2(n-1) &]^T. \end{aligned} \tag{15}$$

Returning to the (cubic) case with generic FIR filters and arbitrary d , it is possible to write a linear-like expression for the separated sources:

$$\underline{\mathbf{y}}(n) = \underline{\mathcal{W}}\underline{\mathcal{A}}\underline{\mathcal{H}}\mathbf{r}(n) + \underline{\mathcal{W}}\underline{\Gamma}\underline{\Theta}\boldsymbol{\rho}(n). \tag{16}$$

Henceforth, we adopt the cubic nonlinearity for our analysis, but it is important to emphasize that the idea of using the Volterra expansion is also valid for other polynomial functions (although this topic will be left for future works). In the following, we present the SOS-based criteria to perform separation and the analytical computation of the covariance matrices involved.

3. Joint Diagonalization of Correlation Matrices

The use of SOS in the linear instantaneous BSS problem is known to be effective when it involves sources that present temporal structure [1, 6]. In such a case, the main idea is to jointly diagonalize the correlation matrices between retrieved sources for a given number of delays; in other words, the objective is to mutually decorrelate the outputs $y_i(n)$, for $i = 1, \dots, N$, but considering different time delays. There are several methods that perform second-order separation, among which we cite the algorithms SOBI [6], WASOBI [7], AMUSE and TDSEP [1, 4]. However, for PNL mixtures, the exclusive use of the SOS for separation has not been addressed yet. In that sense, the temporal-extended formulation of the PNL model may give us important elements to help clarify

certain theoretical aspects in this approach. Furthermore, given the complexity of the PNL model, we opt for diversified uses of the SOS, which includes the classical SOBI criterion for separation [6] and an alternative SOS-based separation
165 measure that combines the temporal formulation with the mutual information measure [1], as will be described in the following.

3.1. The Block-Diagonalization

Based on the classical SOBI criterion, the following statement can be written: given a number of time delays d , it is desired that the correlation matrix

$$\underline{\mathbf{R}}_{\mathbf{y}\mathbf{y}} = E [\underline{\mathbf{y}}(n)\underline{\mathbf{y}}^T(n)] = \begin{bmatrix} \underline{\mathbf{R}}_{y_1y_1} & \underline{\mathbf{R}}_{y_1y_2} \\ \underline{\mathbf{R}}_{y_2y_1} & \underline{\mathbf{R}}_{y_2y_2} \end{bmatrix}. \quad (17)$$

be block-diagonal, i.e., that the cross correlation matrices between outputs (diagonal blocks in Eq. (17)) are all null, where $\underline{\mathbf{R}}_{y_iy_j} = E [\underline{\mathbf{y}}_i(n)\underline{\mathbf{y}}_j^T(n)]$, with $\underline{\mathbf{y}}_i(n) = [y_i(n), y_i(n-1), \dots, y_i(n-d)]^T$ and $\underline{\mathbf{y}}(n) = [\underline{\mathbf{y}}_1^T(n), \underline{\mathbf{y}}_2^T(n)]^T$. Hence, we can write the block-diagonalization (BD) criterion as

$$J_{BD} = \min_{\underline{\mathcal{W}}, \underline{\mathcal{G}}} \text{blkoff}(\underline{\mathbf{R}}_{\mathbf{y}\mathbf{y}}), \quad (18)$$

where $\text{blkoff}(\cdot)$ is the sum of squared elements in the off-block-diagonal of a square matrix. Note that, as the number of delays d increases, the larger the correlation matrices get and more information can be considered by the criterion. Additionally, a norm constraint can be applied to force the main diagonal of $\underline{\mathbf{R}}_{\mathbf{y}\mathbf{y}}$ to be unitary (assuming that $y_i(n)$ is stationary, for $i = 1, \dots, N$), e.g., by summing to the cost (18) the following term:

$$J_c = \min \sum_{i=1}^N (E [y_i^2(n)] - 1)^2, \quad (19)$$

This constraint is necessary in order to avoid – trivial – null solutions (as a con-
170 sequence, it is also useful for identifying an unique solution). Different weights can be applied to J_{BD} and J_c , causing changes on local minima position – gradient-based optimization methods may be more sensitive to these changes. However, since we are using metaheuristics for parameters optimization (which

are intended to be more robust against local minima convergence), we assume
175 equal (unitary) weights for J_{BD} and J_c .

3.2. The Second-Order Mutual Information Measure

The mutual information allows an efficient manner of measuring the sta-
tistical independence. In order to extract the temporal information in a more
complete fashion and also to restrain our analysis to the SOS, we make the
180 following assumption: the sources are mutually independent and (zero-mean)
jointly Gaussian distributed for all considered delays and also present a time
structure. This will lead to an alternative cost function, when compared to the
BD cost, as described bellow.

We start with the definition of mutual independence encompassing the tem-
poral structure of data, i.e.,

$$f_{\underline{\mathbf{y}}}(\underline{\mathbf{v}}) = \prod_{i=1}^N f_{\underline{\mathbf{y}}_i}(\underline{\mathbf{v}}_i), \quad (20)$$

where $f_{\underline{\mathbf{y}}}(\underline{\mathbf{v}})$ and $f_{\underline{\mathbf{y}}_i}(\underline{\mathbf{v}}_i)$ are the multivariate probability density functions asso-
185 ciated with $\underline{\mathbf{y}}(n)$ and $\underline{\mathbf{y}}_i(n)$, respectively. The temporal structure, in this case,
is inherently taken into account by the multivariate densities.

To measure the independence, one can use the mutual information [1] :
 $I(\mathbf{y}(n), \dots, \mathbf{y}(n-d)) = -H(\underline{\mathbf{y}}) + \sum_{i=1}^N H(\underline{\mathbf{y}}_i)$, where $H(\cdot)$ is Shannon's entropy,
defined as $H(\underline{\mathbf{y}}) = -\int_D p(\underline{\mathbf{y}}) \log(p(\underline{\mathbf{y}})) d\underline{\mathbf{y}}$, with $D \subseteq \mathbb{R}^{N(d+1)}$, and $H(\underline{\mathbf{y}}_i) =$
 $-\int_{D_i} p(\underline{\mathbf{y}}_i) \log(p(\underline{\mathbf{y}}_i)) d\underline{\mathbf{y}}_i$, with $D_i \subseteq \mathbb{R}^{d+1}$, for the marginal entropies. The
mutual information is always non-negative and, when independence is reached
for all delays, $I(\mathbf{y}(n), \dots, \mathbf{y}(n-d)) = 0$. In its strict form, $I(\mathbf{y}(n), \dots, \mathbf{y}(n-d))$ is
difficult to be calculated since it demands the estimation of the densities (which
is critical in our case, where all densities are multivariate). However, under the
supposition of a successful separation, it is expected that the recovered sources
be jointly Gaussian as well, i.e., $f_{\underline{\mathbf{y}}}(\underline{\mathbf{v}}) \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_{\underline{\mathbf{y}\mathbf{y}}})$ or

$$f_{\underline{\mathbf{y}}}(\underline{\mathbf{v}}) = \frac{1}{\sqrt{|2\pi\mathbf{R}_{\underline{\mathbf{y}\mathbf{y}}}|}} \exp\left(\frac{-1}{2}\underline{\mathbf{v}}^T \mathbf{R}_{\underline{\mathbf{y}\mathbf{y}}}^{-1} \underline{\mathbf{v}}\right), \quad (21)$$

where $\underline{\mathbf{R}}_{\mathbf{y}\mathbf{y}}$ is as defined by Eq. (17) and $|\cdot|$ is the determinant operator.

It can be shown that, by combining Eq. (21) with $I(\mathbf{y}(n), \dots, \mathbf{y}(n-d))$, the separation criterion associated with the mutual information reduces to [21]

$$J_{SOMI} = \min_{\underline{\mathcal{W}}, \underline{\mathcal{G}}} \frac{1}{2} \log \left(\frac{\prod_{i=1}^N |\underline{\mathbf{R}}_{\underline{y}_i \underline{y}_i}|}{|\underline{\mathbf{R}}_{\mathbf{y}\mathbf{y}}|} \right). \quad (22)$$

It is important to mention that a similar expression was already obtained through the spectral density of Gaussian sources, being named Gaussian Mutual Information (GMI) [1, 21, 22], and the temporal-extended covariance matrices were used in the convolutive mixing problem [15], but its application to the PNL problem is novel. Hence, we refer to Eq. (22) as the Second-Order Mutual Information (SOMI). As the BD cost, Eq. (22) uses only the SOS information, but instead of using summation of quadratic terms (Eq. (18)), the determinants of matrices $\underline{\mathbf{R}}_{\underline{y}_i \underline{y}_i}$ and $\underline{\mathbf{R}}_{\mathbf{y}\mathbf{y}}$ are considered (note, however, that the computational complexity is increased by $O(n^3)$ due to the determinant operator).

The objective of the SOMI criterion is to minimize the cost J_{SOMI} so that $J_{SOMI} = 0$. However, the norm constraint given by Eq. (19) is necessary to avoid null (trivial) solutions.

3.3. The Quadratic SOMI Cost

A closer observation on Eq. (22) reveals that, in fact, a matching between the determinant terms $\prod_{i=1}^N |\underline{\mathbf{R}}_{\underline{y}_i \underline{y}_i}|$ and $|\underline{\mathbf{R}}_{\mathbf{y}\mathbf{y}}|$ would lead the cost J_{SOMI} to be equal to zero (i.e., the mutual information is null). In that sense, a similar cost can be written without relying on the logarithm properties, but on the simplicity of a quadratic difference:

$$J_{SOMIq} = \min_{\underline{\mathcal{W}}, \underline{\mathcal{G}}} \left(\prod_{i=1}^N |\underline{\mathbf{R}}_{\underline{y}_i \underline{y}_i}| - |\underline{\mathbf{R}}_{\mathbf{y}\mathbf{y}}| \right)^2, \quad (23)$$

where the minimal (and desired) cost value is zero. Note that the norm constraint (Eq. (19)) is also necessary to avoid null solutions. This cost is named SOMIq due to its quadratic term.

For SOMI and SOMIq, when the correlation matrix $\underline{\mathbf{R}}_{\mathbf{y}\mathbf{y}}$ is block-diagonal, we have that $|\underline{\mathbf{R}}_{\mathbf{y}\mathbf{y}}| = \prod_{i=1}^N |\underline{\mathbf{R}}_{\mathbf{y}_i\mathbf{y}_i}|$ and the costs are both null. In that sense, we expect that the solutions for SOMI and SOMIq be the same. However, the quadratic relation in SOMIq may be able to provide a desirable cost shape in the optimization process – we will discuss this point in more detail ahead.

3.4. The Analytical Calculation of the Cost Functions

As usual in SOS-based approaches, the main entity is the correlation matrix and, if one considers the mixing and separating model given by Eq. (16), the expanded correlation matrix $\underline{\mathbf{R}}_{\mathbf{y}\mathbf{y}}$ can be computed analytically:

$$\begin{aligned}
\underline{\mathbf{R}}_{\mathbf{y}\mathbf{y}} &= E [\underline{\mathbf{y}}(n)\underline{\mathbf{y}}^T(n)] \\
&= E [\underline{\mathcal{W}}\underline{\mathcal{A}}\underline{\mathcal{H}}\underline{\mathbf{r}}(n)\underline{\mathbf{r}}^T(n)\underline{\mathcal{H}}^T\underline{\mathcal{A}}^T\underline{\mathcal{W}}^T] + E [\underline{\mathcal{W}}\underline{\mathcal{A}}\underline{\mathcal{H}}\underline{\mathbf{r}}(n)\underline{\boldsymbol{\rho}}^T(n)\underline{\boldsymbol{\Theta}}^T\underline{\boldsymbol{\Gamma}}^T\underline{\mathcal{W}}^T] \\
&\quad + E [\underline{\mathcal{W}}\underline{\boldsymbol{\Gamma}}\underline{\boldsymbol{\Theta}}\underline{\boldsymbol{\rho}}(n)\underline{\mathbf{r}}^T(n)\underline{\mathcal{H}}^T\underline{\mathcal{A}}^T\underline{\mathcal{W}}^T] + E [\underline{\mathcal{W}}\underline{\boldsymbol{\Gamma}}\underline{\boldsymbol{\Theta}}\underline{\boldsymbol{\rho}}(n)\underline{\boldsymbol{\rho}}^T(n)\underline{\boldsymbol{\Theta}}^T\underline{\boldsymbol{\Gamma}}^T\underline{\mathcal{W}}^T] \\
&= \underline{\mathcal{W}}\underline{\mathcal{A}}\underline{\mathcal{H}}\underline{\mathbf{R}}_{rr}\underline{\mathcal{H}}^T\underline{\mathcal{A}}^T\underline{\mathcal{W}}^T + \underline{\mathcal{W}}\underline{\mathcal{A}}\underline{\mathcal{H}}\underline{\mathbf{R}}_{r\rho}\underline{\boldsymbol{\Theta}}^T\underline{\boldsymbol{\Gamma}}^T\underline{\mathcal{W}}^T + \underline{\mathcal{W}}\underline{\boldsymbol{\Gamma}}\underline{\boldsymbol{\Theta}}\underline{\mathbf{R}}_{\rho r}\underline{\mathcal{H}}^T\underline{\mathcal{A}}^T\underline{\mathcal{W}}^T \\
&\quad + \underline{\mathcal{W}}\underline{\boldsymbol{\Gamma}}\underline{\boldsymbol{\Theta}}\underline{\mathbf{R}}_{\rho\rho}\underline{\boldsymbol{\Theta}}^T\underline{\boldsymbol{\Gamma}}^T\underline{\mathcal{W}}^T,
\end{aligned} \tag{24}$$

where $\underline{\mathbf{R}}_{rr} = E [\underline{\mathbf{r}}(n)\underline{\mathbf{r}}^T(n)]$, $\underline{\mathbf{R}}_{r\rho} = E [\underline{\mathbf{r}}(n)\underline{\boldsymbol{\rho}}^T(n)]$, $\underline{\mathbf{R}}_{\rho r} = E [\underline{\boldsymbol{\rho}}(n)\underline{\mathbf{r}}^T(n)]$ and $\underline{\mathbf{R}}_{\rho\rho} = E [\underline{\boldsymbol{\rho}}(n)\underline{\boldsymbol{\rho}}^T(n)]$ are the correlation matrices as a function of $\underline{\mathbf{r}}(n)$ and $\underline{\boldsymbol{\rho}}(n)$ – in which $\underline{\boldsymbol{\rho}}(n)$ is the Volterra expansion of $\underline{\mathbf{r}}(n)$. Since $\underline{\mathbf{r}}(n)$ is an *i.i.d.* vector, these covariance matrices have, as non-null elements, only the terms involving $E [r_i^2(n)]$, $E [r_i^4(n)]$ and $E [r_i^6(n)]$, which can be easily obtained. This reveals that some HOS are directly encompassed by the correlation matrices, which might be essential to the nonlinear separation process.

In the BD cost function, only the off-block-diagonal elements are considered, so that the matrices $\underline{\mathbf{R}}_{\mathbf{y}_i\mathbf{y}_j}$, for $i \neq j$, are the ones effectively used. For the SOMI and SOMIq costs, the block-diagonal elements $\underline{\mathbf{R}}_{\mathbf{y}_i\mathbf{y}_i}$ are considered. Based on Eq. (24), we can write:

$$\begin{aligned}
\underline{\mathbf{R}}_{\mathbf{y}_i\mathbf{y}_j} &= \mathcal{W}_i \left(\underline{\mathcal{A}}\underline{\mathcal{H}}\underline{\mathbf{R}}_{rr}\underline{\mathcal{H}}^T\underline{\mathcal{A}}^T + \underline{\mathcal{A}}\underline{\mathcal{H}}\underline{\mathbf{R}}_{r\rho}\underline{\boldsymbol{\Theta}}^T\underline{\boldsymbol{\Gamma}}^T \right. \\
&\quad \left. + \underline{\boldsymbol{\Gamma}}\underline{\boldsymbol{\Theta}}\underline{\mathbf{R}}_{\rho r}\underline{\mathcal{H}}^T\underline{\mathcal{A}}^T + \underline{\boldsymbol{\Gamma}}\underline{\boldsymbol{\Theta}}\underline{\mathbf{R}}_{\rho\rho}\underline{\boldsymbol{\Theta}}^T\underline{\boldsymbol{\Gamma}}^T \right) \mathcal{W}_j^T,
\end{aligned} \tag{25}$$

where $\underline{\mathcal{W}}_i$ is the i th block with $d+1$ rows of $\underline{\mathcal{W}}$. It is possible to note that each of the $(d+1)^2$ elements of $\underline{\mathbf{R}}_{\underline{y}_i \underline{y}_j}$ are quadratic polynomials in function of $\underline{\Gamma}$ and $\underline{\mathcal{W}}_i$ – the separation coefficients – and can contribute with additional information
220 for solving the system. However, there might be redundant equations, since, under the assumption of stationary discrete-time stochastic processes, $\underline{\mathbf{R}}_{\underline{y}_i \underline{y}_i}$ is Toeplitz, and $\underline{\mathbf{R}}_{\underline{y}_i \underline{y}_j} = \underline{\mathbf{R}}_{\underline{y}_j \underline{y}_i}^T$ by definition.

Using the relations (24) and (25), the costs BD, SOMI and SOMIq can be analytically obtained, i.e., the cost functions can be exactly evaluated without
225 any estimation errors. Note that this approach requires the knowledge of the matrices $\underline{\mathcal{H}}$, $\underline{\mathcal{A}}$, $\underline{\Theta}$ and $\underline{\Gamma}$, which, in practice, are not known. However, under a theoretical perspective, it may contribute to a better understanding of the PNL mixtures behavior, as shown next.

4. Identifiability and Bounds on Number of Delays

230 The three aforementioned criteria share a common feature when a solution is found: the extended correlation matrix of the output signals $\underline{\mathbf{R}}_{\mathbf{y}\mathbf{y}}$ is precisely a block-diagonal matrix, i.e., all the off-block-diagonal elements are null. This observation allows us to point out some general aspects involving the SOS-based costs in the context of the particular PNL mixture case considered.

235 4.1. Blind Identifiability

In the linear BSS problem, the study of the blind identification conditions for the SOS-based approaches is a well studied topic [6]: it is known that the linear mixing matrix \mathbf{A} can be identified, up to permutation and scale factors, if the source signals have different spectral shapes. Generally, the demonstration
240 is done by ensuring that the diagonalization process of the correlation matrix for different delays yields eigenvalues that are distinct [4, 1]. In the temporal-extended formulation, this means that $\underline{\mathbf{R}}_{\mathbf{y}\mathbf{y}}$ should be block-diagonal and that each block of the main diagonal – i.e., $\underline{\mathbf{R}}_{\underline{y}_i \underline{y}_i}$, for $i = 1, \dots, N$ – present, at least, two distinct eigenvalues.

245 The extension of this idea to the general PNL problem becomes more complex, since the identification conditions must be valid for \mathbf{A} and $\mathbf{f}(\cdot)$. In the studied case, the problem may be posed in a simpler manner, since we are able to express the nonlinear mixing functions by means of a linear matrix multiplication (Eq. (16)). In that sense, similarly to the linear case, it is possible to
 250 verify the the conditions for identification [1] through the block-diagonalization of $\mathbf{R}_{\mathbf{y}\mathbf{y}}$ in function of \mathcal{W} and $\mathbf{\Gamma}$. To solve this problem, we consider two cases: (i) $\mathbf{\Gamma}$ is null and (ii) $\mathbf{\Gamma}$ has a non-null value.

In case (i), when $\mathbf{\Gamma}$ is null, i.e., the case in which the nonlinear part is solved, \mathcal{W} must block-diagonalize only $\mathcal{A}\mathcal{H}\mathbf{R}_{rr}\mathcal{H}^T\mathcal{A}^T$ and the conditions are the same
 255 as for the linear case [1]: the source signals must have different spectral shapes.

For case (ii), $\mathbf{\Gamma}$ is a non-null diagonal matrix (i.e., there remains a nonlinear residual error), the matrices $\mathbf{R}_{r\rho}$, $\mathbf{R}_{\rho r}$ and $\mathbf{R}_{\rho\rho}$ are not block-diagonal (due to inherent HOS encompassed in the process), and \mathcal{W} is unable to compensate $\mathbf{\Theta}$ by itself. Since the matrices $\mathbf{R}_{r\rho}$, $\mathbf{R}_{\rho r}$ and $\mathbf{R}_{\rho\rho}$ are not block-diagonal, the
 260 diagonalization of $\mathbf{R}_{\mathbf{y}\mathbf{y}}$ would result in non-orthogonal $\mathcal{W}\mathcal{A}\mathcal{H}$ or $\mathcal{W}\mathbf{\Gamma}\mathbf{\Theta}$, with eigenvalues not unique, which is not a desired solution [1].

Regarding the impossibility of the compensation of $\mathbf{\Theta}$ by \mathcal{W} , we have that \mathcal{W} will not be able to jointly diagonalize the linear term as well as the nonlinear ones. For instance, if the linear term ($\mathcal{A}\mathcal{H}\mathbf{R}_{rr}\mathcal{H}^T\mathcal{A}^T$) is block-diagonalized and $\mathcal{W}\mathcal{A}\mathcal{H}\mathbf{R}_{rr}\mathcal{H}^T\mathcal{A}^T\mathcal{W}^T$ is orthogonal (i.e., $\mathcal{W} = \mathcal{A}^{-1}$), then,

$$\mathbf{R}_{\mathbf{y}\mathbf{y}} = \mathcal{H}\mathbf{R}_{rr}\mathcal{H}^T + \mathcal{H}\mathbf{R}_{r\rho}\mathbf{\Theta}^T\mathbf{\Gamma}^T\mathcal{W}^T + \mathcal{W}\mathbf{\Gamma}\mathbf{\Theta}\mathbf{R}_{\rho r}\mathcal{H}^T + \mathcal{W}\mathbf{\Gamma}\mathbf{\Theta}\mathbf{R}_{\rho\rho}\mathbf{\Theta}^T\mathbf{\Gamma}^T\mathcal{W}^T, \quad (26)$$

and the nonlinear terms are not block-diagonal for non-null $\mathbf{\Gamma}$. On the opposite, if the nonlinear terms are made block-diagonal, the linear term will not be block-diagonal. Hence, the possible orthogonal solution (with distinct eigenvalues) is
 265 that with null $\mathbf{\Gamma}$. The only exception happens when $\mathcal{A} = \mathbf{I}$, i.e., when the linear mixing part is reduced to identity (and all the terms will be block-diagonal). However, this case is not considered, since there is no mixture [19].

Thus, in short, the SOS identifiability conditions are that the signals must

present different spectral shapes and that the linear mixing part of the PNL
 270 model must effectively occur. However, there is a crucial condition in the separating model: the SOS-based approach requires that $\underline{\mathbf{R}}_{\mathbf{y}\mathbf{y}}$ encompasses a linear part, which is equivalent to requiring that the combined nonlinear function $\underline{\mathcal{G}} \circ \underline{\mathcal{F}}$ yields a $\underline{\mathbf{z}}(n)$ with at least one linear term (similarly to Eq. (9)). In other words, $\underline{\mathcal{G}}$ must admit at least one term that compensates the nonlinearities $\underline{\mathcal{F}}$ –
 275 yielding the linear term – but can also present a nonlinear residual. The linear term must not vanish, even during the coefficients adaptation. This may be a strong constraint on the PNL separating model, since the choice of $\underline{\mathcal{G}}$ must have a fixed term that compensates $\underline{\mathcal{F}}$; but note that, when it is known that $\underline{\mathcal{F}}$ is composed of polynomial functions, $\underline{\mathcal{G}}$ can be easily constructed by composing
 280 several compensating polynomial terms and fixing the promising ones. This general polynomial case, however, will be treated in future works, since, for the moment, the cubic case will be sufficient to provide insightful perspectives.

4.2. Bounds on the Number of Delays

As previously mentioned, each element in the off-block-diagonal of $\underline{\mathbf{R}}_{\mathbf{y}\mathbf{y}}$ forms
 285 a quadratic polynomial as a function of $\underline{\mathbf{\Gamma}}$ and $\underline{\mathcal{W}}$, which may compose a system of quadratic equations. The number of unknown variables, k , in our studied case, is $k = N(N+1)$, which are the coefficients of $\underline{\mathbf{\Gamma}}$ and $\underline{\mathcal{W}}$.

Some elements (or equations) of $\underline{\mathbf{R}}_{\mathbf{y}\mathbf{y}}$, however, are redundant. For instance, the elements of the main diagonal of $\underline{\mathbf{R}}_{y_i y_j}$ form the same equation in function
 290 of the unknown variables and, hence, they only contribute as a single equation (new information) to the system. In addition, we have that $\underline{\mathbf{R}}_{y_i y_j} = \underline{\mathbf{R}}_{y_j y_i}^T$, i.e., the sub-diagonals are equivalent, which also reduces the effective number of equations in the system. In that sense, by removing the redundancy, we have $N(N-1)/2$ matrices $\underline{\mathbf{R}}_{y_i y_j}$, in which the number of effective equations are
 295 $d(d+1)+1$ each one, resulting in a total of $N(N-1)(d(d+1)+1)/2$ equations in the system. Besides, the normalization given by Eq. (19) also performs a role as a constraint, and can contribute to the system with N equations. Finally, it is possible to state that the number of effective equations in the system is

$N(N-1)(d(d+1)+1)/2 + N$. Hence, to obtain valid solutions, it is necessary
 300 that d be chosen so that $N(N-1)(d(d+1)+1)/2+N \geq k$.

Notwithstanding, it is also possible that some of the off-diagonal elements
 of $\mathbf{R}_{\underline{y}_i \underline{y}_j}$ be equivalent, depending on the temporal structure of the mixtures
 (i.e., $y_i(n)$ and $y_j(n)$ have similar or equal spectral densities) and, in that case,
 the number of valid equations might be reduced. In that sense, the expression
 305 $N(N-1)(d(d+1)+1)/2+N \geq k$ is only a lower bound for choosing d .

In order to illustrate the system of equations, we consider a 2-source and
 2-mixture case in which the linear mixing part of the PNL model is a rotation
 matrix, i.e., $\mathbf{A} = [\cos(\phi_a), -\sin(\phi_a); \sin(\phi_a), \cos(\phi_a)]$. For the separation, based
 on Eq. (16), we have 2 unknown variables for the joint nonlinear part, γ_1 and
 310 γ_2 , and 1 unknown variable, ϕ_w , for the linear separating matrix \mathbf{W} (which is
 a rotation matrix, similar to \mathbf{A}). Thus, we have that, for $N=2$ and $d=1$, the
 number of equations is, at most, $N(N-1)(d(d+1)+1)/2 + N = 5$. Fig. 2(a)
 shows the surface of each equation for given temporally colored sources, with
 $\phi_a = 1.02$ rad. In this case, the off-diagonal elements of $\mathbf{R}_{\underline{y}_1 \underline{y}_2}$ are coincident
 315 and we only have 4 valid equations, resulting in 4 surfaces (plotted with different
 colors) in Fig. 2(a). The intersection points of the surfaces will determine the
 regions where all equations are satisfied (their intersection occurs for $\gamma_1 = \gamma_2 = 0$
 and $\phi_w = k\pi - \phi_a$, $k = 0, \pm 1, \pm 2, \dots$ – not observable in Fig. 2(a)). Indeed, any
 of these points will be a valid solution for the BD, SOMI and SOMIq criteria.

Although the SOS-based criteria are intended to present the same solution,
 320 their cost shapes may differ. Consider the previous example but assume that
 the linear part is has already been solved, leaving just γ_1 and γ_2 to be adjusted.
 In Fig. 2(b), we show the contours as functions of γ_1 and γ_2 and for d equal to
 1, 2 and 4. In all cases, the global solution is $\gamma_1 = \gamma_2 = 0$ (denoted by an “X”
 325 in Fig. 2(b)), which is the desired solution, but local minima also exist. Very
 interestingly, as d increases, the “weight” of the local optima is reduced, being
 the global solution more evident (mainly for the SOMIq cost).

Next, we consider the case where $\mathbf{\Gamma}$ and \mathcal{W} have no constraints – increasing
 the space of candidate solutions – to evaluate the criteria performance.

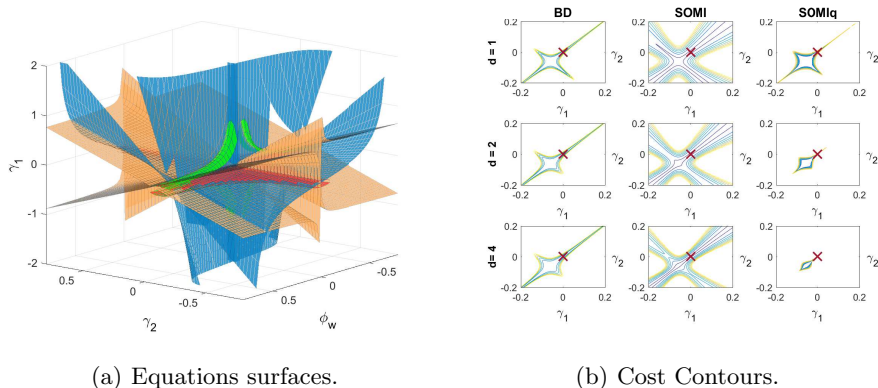


Figure 2: Number of delays: equations and solutions.

330 5. Performance Analysis

So far, we have verified that the SOS-based criteria share some features, but they might differ in their cost shapes. In fact, when an optimization task is performed, the cost shape might cause a significant impact on the performance. In order to test this effect, we consider simulation scenarios with $N=M=2$ and $N=M=3$ and with cubic nonlinear mixing functions, i.e., $f_i(u_i(n)) = u_i^3(n)$, for $i = 1, \dots, N$, or, for the extended-temporal version, $\mathcal{F}(\mathbf{u}(n)) = \mathbf{u}^{\odot 3}(n)$.

As usual in BSS problems, $\mathbf{s}(n)$, \mathbf{A} and $\mathbf{f}(\cdot)$ are not known. Hence, it could be difficult to define a separation structure. For our simulation tests, we assume that the nonlinear compensating function was chosen to be of the form $g_i(x_i(n)) = g_{i,0}\text{sgn}(x_i(n))(|x_i(n)|)^{1/3} + g_{i,1}x_i(n)$, for $i = 1, \dots, M$. With the objective of obtaining a linear term, which is a separation condition, it is necessary to keep fixed one of the coefficients ($g_{i,0}$ or $g_{i,1}$). Thus, we assume that $g_{i,0} = 1$ remains fixed and $g_{i,1} = \gamma_i$ is allowed to vary (the suitable choice of the fixed coefficients might require some heuristic tests). For the extended-temporal notation, this is equivalent to $\underline{\mathcal{G}}(\mathbf{x}(n)) = \text{sgn}(\mathbf{x}(n)) \odot (|\mathbf{x}(n)|)^{\odot 1/3} + \underline{\mathbf{\Gamma}}\mathbf{x}(n)$. Conveniently, this model is exactly as proposed in Eq. (11), where $\underline{\mathbf{\Gamma}}$ and $\underline{\mathcal{W}}$ must be adapted for performing separation.

For the optimization of the weights (linear and nonlinear), we adopt the metaheuristics called Differential Evolution (DE) and Clonal Selection Algo-

350 rithm (CLONALG), which are efficient techniques to explore the search space
 and to avoid convergence to local optima [16, 17]. The DE metaheuristic relies
 on probabilistic vector operations, while CLONALG is based on an artificial
 immune system inspired by the clonal selection theory and antigen-antibody
 interactions [17]. Their main difference is the fact that, in DE, the candidate
 355 solutions are adapted by mechanisms that exploit the information about the
 search space that is available in the current population, while, in CLONALG,
 conventional operators based on random perturbations are used (for more de-
 tails, please refer to [16, 17]). In simulations, the metaheuristics parameters
 were adjusted to: the DE parameters were chosen to be $F = 0.5$ (adaptation
 360 step) and $CR = 0.9$ (crossover constant), and the CLONALG parameters were
 $N_c = 10$ (number of clones), $\beta = 5$ (decay of mutation), 15% of new random
 cells, $T = 50$ (period of cells insertion). For the $N=M=2$ case (6 weights: 2
 for $\underline{\mathbf{I}}$ and 4 for $\underline{\mathbf{W}}$), we set $N_P = 500$ (population size/number of cells) and
 5000 iterations for both DE and CLONALG, whereas, for the $N=M=3$ case
 365 (12 weights: 3 for $\underline{\mathbf{I}}$ and 9 for $\underline{\mathbf{W}}$), we used $N_P = 700$ (the other parameters
 were kept the same). These parameters were constant for all simulation cases
 (however, for scenarios with more sources and/or more complex nonlinearities,
 it is recommended that the search power of the optimization strategy be in-
 creased for a higher global convergence rate). In the end of the adaptation, the
 370 individual with the best fitness (lower SOS-based cost) is selected to provide the
 solution. The general steps of the optimization method are as shown in Alg. 1

The performance of the found solutions can be measured in terms of the
 Signal-to-Interference Ratio (SIR) (after permutation, sign and variance correc-
 tion), which is defined as $\text{SIR} = 10 \log (E[y_i(n)^2]/E[(s_i(n) - y_i(n))^2])$. In that
 375 sense, higher SIR values mean better performance solutions.

5.1. Performance Using the Analytical Covariance Matrices

In the first scenario, we wish to investigate the effect of the number of delays
 d in the SOS-based separation criteria using the analytical covariance matrices,
 i.e., by obtaining $\underline{\mathbf{R}}_{\mathbf{y}\mathbf{y}}$ directly from Eq. (24), without estimation errors. For

Algorithm 1 SOS-based method for optimization using DE/CLONALG

Initialization of DE/CLONALG parameters;

Randomly initialize all N_P individuals in the $\underline{\mathbf{I}}$ and $\underline{\mathcal{W}}$ search space;

while Maximum number of iterations is not reached **do**

for Each individual $i \in N_P$ **do**

if Using DE [16]: **then**

 Generate mutated vector by randomly picking 3 different individuals;

 Combination with the original individual (F, CR);

end if

if Using CLONALG [17]: **then**

 Generate N_c clones;

 Mutation of the clones (β);

end if

 Selection:

 Obtain $\underline{\mathbf{R}}_{yy}$ analytically (Eq. (25)) or via sample estimation;

 Performance Evaluation, according to

 BD (Eq. (18)), SOMI (Eq. (22)) or SOMIq (Eq. (23))

 Keep the best (original or combination/mutated) individual;

if Using CLONALG: **then**

 Insertion of new individuals at period T ;

end if

end for

Pick the best individual of the population and present as best found candidate for solution at this stage

end while

380 $N=2$ sources, two *i.i.d.* Gaussian signals ($r_1(n)$ and $r_2(n)$) are generated and temporally colored by the FIR filters $\mathbf{h}_1 = [1, 0.6, -0.3, 0.1, 0.4, 0.3, -0.22, 0.18, 0.5]$ and $\mathbf{h}_2 = [1, -0.2, -0.8, 0.2, 0.1, -0.41, 0.5, 0.1]$, separately. Note that the temporal structures provided by \mathbf{h}_1 and \mathbf{h}_2 are of finite length and, hence, there is a limited amount of temporal information to be extracted. The linear mixing

385 matrix is $\mathbf{A} = [0.25, 0.86; -0.86, 0.25]$ and we wish to adapt $\underline{\Gamma}$ and $\underline{\mathcal{W}}$. Supposing that $r_1(n)$ and $r_2(n)$ are zero-mean and unit variance Gaussian processes and that \mathbf{h}_1 and \mathbf{h}_2 as well as the mixing coefficients are known, the covariance matrices can be analytically computed with (24).

We considered that the number of delays d can vary from 1 to 7 and, for

390 each value of d , we performed 50 independent runs of the DE and CLONALG with the aim of minimizing the BD, SOMI and SOMIq costs (separately). The found solutions were evaluated in terms of SIR for 700,000 test samples (used only for evaluation). Fig. 3(a) shows the mean SIR values for each considered delay, while Fig. 3(b) shows the SIR values for the best solution found by DE and CLONALG throughout the 50 runs. From the mean SIR values, one can

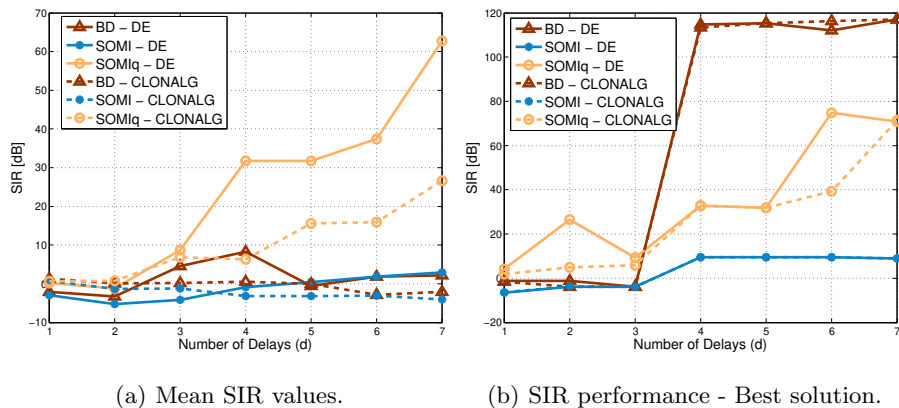


Figure 3: Mean and Best Performance - Analytical Costs.

395

note that the found solutions for the BD and SOMI criteria led to a low value of SIR for all considered delays, for both DE and CLONALG, indicating that their solutions might not be adequate for performing BSS. However, in terms of the best found solution, the BD criterion shows an intriguing result: for $d \geq 4$, the

400 solutions for the BD are able to separate the sources with the highest level of SIR. This indicates that, although we have employed a huge search resource for DE and CLONALG, they have presented difficulties in finding the best solution for BD. For the SOMI criterion, the best found solution shows an improved performance for $d \geq 4$, but it remains around the 10 dB level. On the other hand, for the SOMIq criterion, the DE and CLONALG found good solutions more easily, without great discrepancies between Figs. 3(a) and 3(b), and, from a general perspective, it is possible to say that the solutions SIR level tends to increase with d , being the sources successfully separated. Note that, for $d < 2$, the SOMIq best solutions found by DE are not able to separate the sources, 405 which is in accordance with the bound on the number of delays (Section 4.2).

In order to clarify the obtained results, we compare in Fig. 4(a) the lowest attained costs values of BD, SOMI and SOMIq for a ‘regular’ solution (a randomly picked solution) and the best solution throughout 50 runs of the DE, all for $d = 4$. It is possible to note that, for the BD cost, the difference between

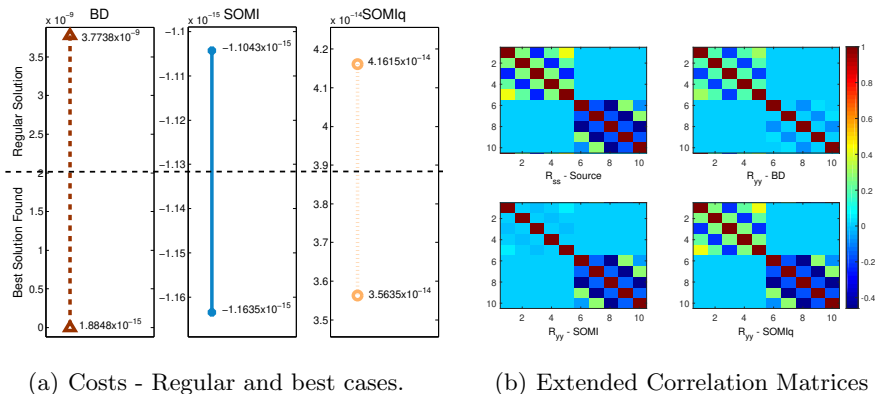


Figure 4: Cost Comparison and Correlation matrices.

415 the two cases is larger, being clear that DE is presenting difficulties to find the global optima, differently from the other costs, where the differences were relatively small. This indicates that the BD cost shape may impose some difficulties in finding the global optima, being necessary to increase the search power of the DE metaheuristic. For SOMI, the minimization of Eq. (22) led to small values

420 of the cost, but, due to precision issues, the cost values were negative (ideally positive), which contributed to its poorer performance in comparison with the other criteria. The SOMIq cost, on the other hand, solves the SOMI drawback and converges to close and positive small values.

A more intuitive comparison can be obtained from Fig. 4(b), where we illustrate a colored version of the extended correlation matrix of the sources $\mathbf{R}_{\mathbf{ss}}$ and of the outputs $\mathbf{R}_{\mathbf{yy}}$ for the BD, SOMI and SOMIq solutions in one of the executions (the same solution picked as ‘regular’) of the DE metaheuristic, all for $d=4$. It is possible to note that, for $\mathbf{R}_{\mathbf{ss}}$, the main diagonal blocks are colored in different patterns, which reveals the temporal structure of the sources, whereas the off-diagonal blocks correspond to uncorrelated values and present a single color. Ideally, the objective is to obtain $\mathbf{R}_{\mathbf{yy}}$ as close as possible to $\mathbf{R}_{\mathbf{ss}}$. For the BD solution found by DE, the temporal structure of only one of the sources was preserved, while the other source presented small temporal correlation. A similar result also applies to SOMI, whereas, for the SOMIq output, the found solution was the desired one, whose sources are mutually uncorrelated and with their temporal structure preserved. The observations outlined here were similar to the results found by the CLONALG, hence, we decided to omit them. These results reveal that, although encompassing different searching mechanisms, both metaheuristics were able to identify better solutions with SOMIq, indicating that its cost is able to express the required SOS more efficiently.

5.2. Performance Using Estimated Covariance Matrices

For real-world problems, the covariance matrices are generally estimated from samples (via sample mean), which certainly leads to approximated values with reduced accuracy. This could be of major importance for the algorithms performance with tendency of converging to local solutions, which, as indicated in the previous analysis, is the case of the BD cost. This issue will be investigated now along with the cases in which the sources are Gaussian and non-Gaussian.

We consider the $N=M=2$ and $N=M=3$ cases. For $N=M=2$, the mixing matrix is $\mathbf{A} = [0.55, -0.92; -0.82, 0.38]$, while, for $N=M=3$, $\mathbf{A} = [0.55, -0.92, 0.20;$

450 $[-0.82, 0.38, -0.24; -0.52, -0.27, 0.79]$. For both, the nonlinearity is the cubic function, given by Eq. (9), as adopted throughout this paper. The temporal correlation is obtained through FIR filters, chosen to be $\mathbf{h}_1 = [1, 0.6, -0.3, 0.1, 0.4]$, $\mathbf{h}_2 = [1, -0.2, -0.8, 0.2, 0.1]$ and $\mathbf{h}_3 = [1, 0.4, -0.7, 1.3, 0.2]$ (for $N=2$, only \mathbf{h}_1 and \mathbf{h}_2 are used). Now, we assume two types of distributions for $r_i(n)$: in the first case, for all $i=1, \dots, N$, $r_i(n)$ is an *i.i.d.* Gaussian signal with zero mean and variance equal to 2 and, in the second case, for all $i=1, \dots, N$, $r_i(n)$ is an *i.i.d.* signal uniformly distributed between -1 and $+1$. We consider that the number of samples of $y_i(n)$, may vary from 250 up to 700,000 for the covariance matrices estimation, being used, in the sequel, for obtaining BD, SOMI and SOMIq costs. A test set with 700,000 samples will be used for SIR estimation. 465
The number of considered delays is $d = 4$.

To adapt the coefficients $\underline{\Gamma}$ and $\underline{\mathcal{W}}$, we use the DE metaheuristic with the same previously defined parameters and perform 100 independent runs (CLONALG was not considered in this case, since its performance is similar to DE). 465
For each number of samples considered in the covariance matrices estimation, the resulting mean SIR performance for the BD, SOMI and SOMIq solutions found by the DE are exhibited in Fig. 5 for the Gaussian and uniform sources (solid lines for the $N=2$ case and dotted lines for the $N=3$ case). It is possible

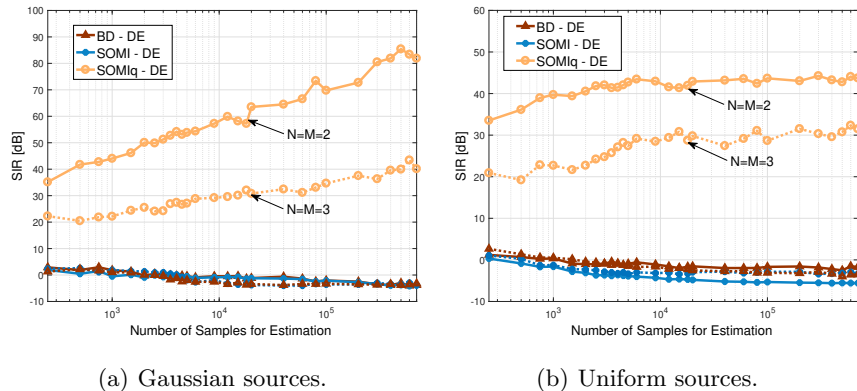


Figure 5: Mean SIR [dB] vs. Number of Samples (log).

to note that the BD and SOMI solutions found by the DE presented similar but

470 lower values of SIR in all cases, while the SOMIq solutions achieved the best re-
sults: for the Gaussian sources, the higher the number of samples, the higher the
SIR level obtained; however, for uniform sources, the number of samples causes
less impact on the SIR performance (this limitation might be a consequence
of the Gaussianity assumption in the SOMIq cost derivation). In addition, in
475 comparison with the $N=2$ case, SOMIq is still able to perform separation for
 $N=3$ but with lower SIR values – since, with 12 adjustable coefficients, the DE
convergence to local solutions is higher.

6. Conclusion

In this work, the problem of BSS was investigated in the context of PNL mix-
480 tures from an SOS-based perspective. In order to identify the constraints and
conditions for the SOS-based approach, a temporal-extended formulation and a
PNL model with cubic polynomial nonlinear functions were considered. Within
this context, the classical SOS-based SOBI and GMI criteria were written under
this temporal-extended standpoint, being named BD and SOMI, respectively.
485 Moreover, to reduce the mathematical complexity of SOMI, a quadratic-like
expression was proposed and named SOMIq.

Due to the simplicity of the SOS-based methods and the assumed cubic
nonlinear functions (and also their Volterra expansion), the covariance matrices
could be analytically computed. Based on this, a theoretical analysis on the
490 costs defined the identifiability conditions and a lower bound on the number of
delays that must be considered for separation: the number of delays will depend
on the degrees of freedom of the separation system, from which a resulting linear
term must always exist. Interestingly, the analysis might be extended to any
polynomial functions. To evaluate the performance of the criteria in the consid-
495 ered PNL model, some simulations were held in scenarios using analytical and
estimated versions of the covariance matrices, being the optimization process
made by the DE and CLONALG metaheuristics. In the analytical case, the
results indicated that the BD cost shape caused some difficulties for the meta-

heuristics to find the global optima; the SOMI criterion presented some issues
500 in its cost minimization, which led to solutions that were not able to establish
mutual independent sources with the desired precision; on the other hand, the
SOMIq criterion presented higher global convergence than BD, with solutions
that preserved the mutual independence (from an SOS point of view) and the
temporal structure of data – also, by increasing the number of delays above
505 the lower bound, we observed an improvement of the performance in terms
of SIR. For the case with estimated covariance matrices, the SOMIq criterion
again presented the best SIR performance for scenarios with Gaussian and uni-
formly distributed sources. However, a better performance can be achieved in
the Gaussian case, since the SOMIq assumes Gaussian sources.

510 Although the present analysis focuses on a specific case of the PNL mixtures,
it can be viewed as a relevant step towards the use of the SOS framework in
the nonlinear BSS problem. In that sense, for future works, we consider the
extension of this analysis to other polynomial nonlinearities, other nonlinear
mixing models and the investigation of possible computational improvement
515 based on the covariance matrices structure.

References

- [1] P. Comon, C. Jutten, Handbook of Blind Source Separation: Independent
Component Analysis and Applications, Academic Press, 2010.
- [2] J. M. T. Romano, R. R. F. Attux, C. C. Cavalcante, R. Suyama, Unsu-
520 pervised Signal Processing: Channel Equalization and Source Separation,
CRC Press, 2010.
- [3] Y. Deville, Blind Source Separation and Blind Mixture Identification Meth-
ods, in: J. G. Webster (Ed.), Wiley Encyclopedia of Electrical and Elec-
tronics Engineering, American Cancer Society, 2016, pp. 1–33.
- 525 [4] A. Hyvärinen, J. Karhunen, E. Oja, Independent Component Analysis,
Wiley, 2001.

- [5] C. Jutten, J. Karhunen, Advances in Blind Source Separation (BSS) and Independent Component Analysis (ICA) for Nonlinear Mixtures, *International Journal of Neural Systems* 14 (05) (2004) 267–292.
- 530 [6] A. Belouchrani, K. Abed-Meraim, J. Cardoso, E. Moulines, A Blind Source Separation Technique Using Second-Order Statistics, *IEEE Transactions on Signal Processing* 45 (2) (1997) 434–444.
- [7] A. Yeredor, Blind Separation of Gaussian Sources via Second-Order Statistics with Asymptotically Optimal Weighting, *IEEE Signal Processing Letters* 7 (7) (2000) 197–200.
- 535 [8] L. T. Duarte, C. Jutten, S. Moussaoui, A Bayesian Nonlinear Source Separation Method for Smart Ion-Selective Electrode Arrays, *IEEE Sensors Journal* 9 (12) (2009) 1763–1771.
- [9] I. Meganem, Y. Deville, S. Hosseini, P. Déliot, X. Briottet, L. T. Duarte, Linear-Quadratic and Polynomial Non-Negative Matrix Factorization; Application to Spectral Unmixing, in: 19th European Signal Processing Conference, IEEE, 2011, pp. 1859–1863.
- 540 [10] I. Meganem, Y. Deville, S. Hosseini, P. Déliot, X. Briottet, Linear-Quadratic Blind Source Separation Using NMF to Unmix Urban Hyperspectral Images, *IEEE Transactions on Signal Processing* 62 (7) (2014) 1822–1833.
- 545 [11] Y. Deville, A. Deville, Quantum-Source Independent Component Analysis and Related Statistical Blind Qubit Uncoupling Methods, in: G. R. Naik, W. Wang (Eds.), *Blind Source Separation*, Springer, 2014, pp. 3–37.
- 550 [12] B. Ehsandoust, M. Babaie-Zadeh, B. Rivet, C. Jutten, Blind Source Separation in Nonlinear Mixtures: Separability and a Basic Algorithm, *IEEE Transactions on Signal Processing* 65 (16) (2017) 4339–4352.

- [13] S. Hosseini, C. Jutten, On the Separability of Nonlinear Mixtures of Temporally Correlated Sources, *IEEE Signal Processing Letters* 10 (2) (2003) 43–46.
- 555
- [14] A. Taleb, C. Jutten, Source Separation in Post-Nonlinear Mixtures, *IEEE Transactions on Signal Processing* 47 (10) (1999) 2807–2820.
- [15] H. Buchner, R. Aichner, W. Kellermann, A Generalization of Blind Source Separation Algorithms for Convolutive Mixtures Based on Second-Order Statistics, *IEEE Transactions on Speech and Audio Processing* 13 (1) (2005) 120–134.
- 560
- [16] K. Price, R. M. Storn, J. A. Lampinen, *Differential Evolution: A Practical Approach to Global Optimization*, Springer, 2005.
- [17] L. N. De Castro, F. J. Von Zuben, *Recent Developments in Biologically Inspired Computing*, Igi Global, 2005.
- 565
- [18] A. Taleb, J. Solé, C. Jutten, Quasi-Nonparametric Blind Inversion of Wiener Systems, *IEEE Transactions on Signal Processing* 49 (5) (2001) 917–924.
- [19] S. Achard, C. Jutten, Identifiability of Post-Nonlinear Mixtures, *IEEE Signal Processing Letters* 12 (5) (2005) 423–426.
- 570
- [20] J. Thomas, Y. Deville, S. Hosseini, Time-Domain Fast Fixed-Point Algorithms for Convolutive ICA, *IEEE Signal Processing Letters* 13 (4) (2006) 228–231.
- [21] D.-T. Pham, Blind Separation of Instantaneous Mixture of Sources Via the Gaussian Mutual Information Criterion, *IEEE Transactions on Signal Processing* 81 (4) (2001) 855–870.
- 575
- [22] D.-T. Pham, J.-F. Cardoso, Blind Separation of Instantaneous Mixtures of Non-Stationary Sources, *IEEE Transactions on Signal Processing* 49 (9) (2001) 1837–1848.