A Joint Second-Order Statistics and Density Matching-Based Approach for Separation of Post-Nonlinear Mixtures

Denis G. Fantinato[1 (✉)], Leonardo T. Duarte[2], Paolo Zanini[3], Bertrand Rivet[3], Romis Attux[1], and Christian Jutten[3]

[1] School of Electrical and Computer Engineering - University of Campinas, Campinas/SP, Brazil,
[2] School of Applied Sciences - University of Campinas, Campinas/SP, Brazil,
[3] GIPSA-Lab, Grenoble INP, CNRS, Grenoble, France,
{denisgf,attux}@dca.fee.unicamp.br,leonardo.duarte@fca.unicamp.br,{paolo.zanini,bertrand.rivet,christian.jutten}@gipsa-lab.grenoble-inp.fr

Abstract. In the context of Post-Nonlinear (PNL) mixtures, source separation can be performed in a two-stage approach, which encompasses a nonlinear and a linear compensation part. In the former part, it is usually required the knowledge of all the source distributions. In this work, we propose a less restrictive approach, where only one source distribution is needed to be known – here, chosen to be a colored Gaussian. The other sources are only required to present a time structure. The method combines, in a joint-based approach, the use of the second-order statistics (SOS) and the matching of distributions, which shows to be less costly than the classical method of computing the marginal entropy for all sources. The simulation results are favorable to the proposal.

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

1 Introduction

In the area of signal processing, the problem of retrieving a set of source signals from their mixtures has been intensively studied for three decades. Since this task is performed with only the knowledge of some samples of the mixtures, this problem is named Blind Source Separation (BSS) [1]. The majority of the initial efforts were aimed at the standard linear and instantaneous mixture problem, with the assumption that the sources are mutually independent. These studies resulted in a well-founded and solid theoretical framework known as Independent Component Analysis (ICA) [1]. Although it can count with a vast number of practical applications, there are certain cases in which the linear assumption is insufficient – e.g., smart chemical sensor arrays [2] and hyperspectral imaging [3] – and nonlinear mixing models must be considered. Notwithstanding, from a general nonlinear standpoint, the ICA framework may not provide the sufficient information for performing source separation. Thus, the studies on this topic were focused on a constrained set of nonlinear models in which the ICA methods are still valid [4], like the so-called Post-Nonlinear (PNL) models [5].

The approaches for solving the PNL mixing problem can be roughly divided into the *joint* and the *two-stage* approaches [6]. In the former case, an ICA-based method is usually employed [5]. In the second case, the nonlinear part is solved in a first step – e.g., via a Gaussianization method [7] – and, for the subsequent step, there remains a linear BSS problem, which is a well studied issue [1]. Additionally, if the sources present a temporal structure, a second-order statistics (SOS)-based approach can be employed in the second stage [7]. Notwithstanding, these approaches may suffer some drawbacks: in the joint approach, it is usually necessary to estimate the mutual information, which may be computationally costly and also be susceptible to local minima convergence. In the two-stage approach, the nonlinear compensation methods, for achieving accurate enough results, may require prior assumptions, which can not be available in certain scenarios [6]. In this work, we consider a less restrictive approach by assuming the knowledge of the distribution shape of a single source, e.g., a Gaussian distribution, and that the sources present temporal structure. In this case, we propose a joint approach which allies a SOS-based cost function to a density (Gaussian) matching which can be simply performed via kernel estimators [8]. We also consider a robust metaheuristic known as Differential Evolution (DE) [9] to avoid suboptimal convergence.

2 The Post-Nonlinear Mixtures

In the blind source separation (BSS) problem, the main objective is to retrieve the original sources $\mathbf{s}(n)$ from the observed mixtures $\mathbf{x}(n) = \boldsymbol{\Phi}(\mathbf{s}(n))$, where $\mathbf{x}(n) = [x_1(n) \ \cdots \ x_M(n)]^T$ is the observation vector of length M, $\mathbf{s}(n) = [s_1(n) \ \cdots \ s_N(n)]^T$ is the source vector with N elements and $\boldsymbol{\Phi}(\cdot)$ is the mixing function [1]. Classically, it is assumed that the mixing function can be described as linear and instantaneous system of the type $\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n)$, where $\mathbf{A}$ is a $M \times N$ matrix. However, this model is not sufficient for certain applications. In that sense, the *Post-Nonlinear* (PNL) model rises as an emblematic and significant step in nonlinear BSS [1],[5].

The PNL system comprises two stages of mixing: the linear and the nonlinear stages. As illustrated in Fig. 1, the mixtures can be written as $\mathbf{x}(n) = \mathbf{f}(\mathbf{A}\mathbf{s}(n))$, being $\mathbf{f}(\cdot)$ a set of M component-wise functions. The separation system is a mirrored version of the mixing system, being its output given by $\mathbf{y}(n) = \mathbf{W}\mathbf{g}(\mathbf{x}(n))$, where $\mathbf{W}$ is a $N \times M$ matrix and $\mathbf{g}(\cdot)$ is a set of M component-wise functions, ideally the inverse of $\mathbf{f}(\cdot)$ [1].

2.1 Separation Techniques for PNL Mixtures

In the context of PNL mixtures, it is possible to classify the separation techniques into two main classes: the *joint* and the *two-stage* approaches [6].

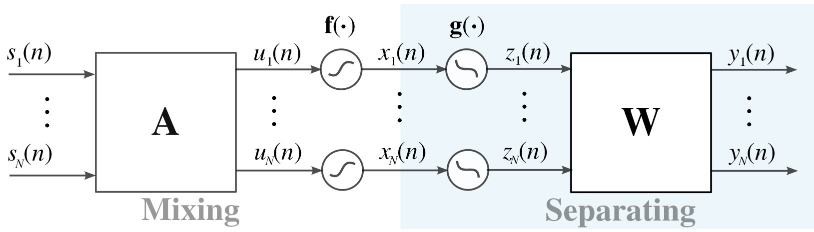


Fig. 1. Mixing and separating systems in the PNL model.

In the former, the main idea is to jointly adjust $\mathbf{g}(\cdot)$ and $\mathbf{W}$ by minimizing a given statistical dependence measure; generally, the use of the ICA framework represents an efficient methodology for performing separation, but issues like local convergence and constrained adaptation of the nonlinearities require special attention – e.g., it is necessary that $\mathbf{f}(\cdot)$ and $\mathbf{g}(\cdot)$ be bijective pairs [6].

On the other hand, for the two-stage approach, the linear and the nonlinear mixing stages are addressed separately, i.e., two different but "simpler" problems need to be solved: $\mathbf{g}(\cdot)$ is adapted so that the nonlinear part of the mixtures are completely suppressed and, then, $\mathbf{W}$ is adjusted to solve the classic linear BSS problem. There are a number of methods for adapting $\mathbf{g}(\cdot)$ – the first stage –, as those based on some *a priori* information [10], but the most common approach is that based on Gaussianization: from the perspective of the central limit theorem, the resultant random variables after the linear mixing stage will tend to be "more" Gaussian. Thus, the most intuitive idea for adapting $\mathbf{g}(\cdot)$ is to make its output $\mathbf{z}(n)$ Gaussian again [7]. This strategy reveals to be more effective when the number of sources N is large – according to the central limit theorem – or when the sources are Gaussian distributed. One can also include among these ideas the notion of the matching of probability distributions, which was one of the first methods in the PNL two stage approaches [11]. In this case, the nonlinearity compensation is accomplished when the distributions associated with $\mathbf{u}(n)$ and with $\mathbf{z}(n)$ are matched – note, however, that the *a priori* knowledge of the distribution of $\mathbf{u}(n)$ is required. This idea will also be relevant for the present work.

The second stage – i.e., the adaptation of the linear term $\mathbf{W}$ – is usually solved with classical ICA methods, which encompass higher-order statistics (HOS) [1], [6]. However, when the sources are temporally colored, methods based on second-order statistics (SOS) can be applied, since they are known for its robustness and reliable simplicity. This idea is exploited in [7] by using a Gaussianization method in the first stage followed by a temporal decorrelation separation (TDSEP) method [1] in the second stage. In fact, this approach is interesting because it merges the simplicity of the second-order framework with simple source priors, for solving the complex nonlinear mixtures.

Although each approach presents its own particular advantages, in this work, we propose the use of a joint approach which is able, to a certain extent, to mix the benefits of a Gaussianization method – by means of a probability density matching – with the simplicity of the separation techniques based on SOS. The method will be described in the next section.

3 Proposed Separation Method

The separation method for PNL mixtures proposed in this work is based on a criterion that mixes the use of SOS and the matching of a (Gaussian) probability density. We start with the following assumptions: (*i*) there is at least one source is Gaussian; (*ii*) the sources are jointly wide-sense stationary, present a temporal structure with different autocorrelation functions and are mutually independent; (*iii*) $\mathbf{f}(\cdot)$ is a set of invertible nonlinear functions; and (*iv*) the linear mixing matrix have, at least, two nonzero entries per row and per column.

Since we aim at the joint approach, we seek a single separation criterion which should be able to jointly adapt $\mathbf{g}(\cdot)$ and $\mathbf{W}$. However, this criterion will be composed of two parts, whose concepts can be understood separately – as we intend to show – but not its *modus operandi*.

3.1 Second-Order Statistics for Blind Separation

The first part of the criterion is based on the temporal structure of the sources. More precisely, we make use of the classical second-order joint diagonalization methods for linear BSS, which were the starting points for approaches and algorithms like SOBI, AMUSE, TDSEP and modified versions [1].

In this case, the SOS are exploited through time lagged covariance matrices:

$$\mathbf{R}_{\mathbf{y},d_s} = E\left[\mathbf{y}(n)\mathbf{y}^T(n-d_s)\right], \tag{1}$$

being d_s a constant lag. The main idea is to simultaneously (approximately) diagonalize the lagged covariance matrices for different values of d_s previously chosen, which can be summarized in the following cost [1]:

$$J_{SOS}(\boldsymbol{\theta}) = \sum_{d_s \in \mathcal{S}} \text{off}\left(\mathbf{R}_{\mathbf{y},d_s}\right) = \sum_{d_s \in \mathcal{S}} \sum_{i \neq j} \left(E\left[y_i(n) y_j(n-d_s)\right]\right)^2, \tag{2}$$

where off$(\cdot)$ the sum of the squares of the off-diagonal elements of a given matrix; $\mathcal{S}$ the set of chosen delays and $\boldsymbol{\theta}$ the set of parameters to be adjusted, i.e., $\boldsymbol{\theta} = \{\mathbf{g}(\cdot), \mathbf{W}\}$. An additional normalization term $(E\left[y_i^2(n)\right] - 1)^2$ for $i = \{1, \ldots, N\}$ is considered, since there are no whitening step for the nonlinear case. To solve the problem, $J_{SOS}(\boldsymbol{\theta})$ has to be minimized under a constraint over the linear separating matrix $\mathbf{W}$, in order to avoid convergence to the trivial solution.

Source separation based only on SOS is known to provide sufficient statistical information in the linear mixing case. However, in the nonlinear problem, additional statistics might be necessary.

3.2 Matching of Gaussian Distributions

Since we consider that at least one of the sources is Gaussian, this statistical information can be used in the second part of the criterion. Instead of using Gaussianization methods [6, 7], a multidimensional density matching approach can be employed, such as the quadratic divergence between densities via kernel density estimators [8].

Basically, the idea is to force one of the recovered sources, say $y_1(n)$, to be Gaussian with a given temporal correlation (from a covariance matrix). In order to use the temporal information, we consider the following vector, related to the first output $y_1(n)$:

$$\overline{\mathbf{y}}_1(n) = [y_1(n)\ y_1(n{-}1)\ \dots\ y_1(n{-}d_m)]^T, \tag{3}$$

where d_m is the maximum number of delays considered. In this case, the temporal covariance matrix of $\overline{\mathbf{y}}_1(n)$ is $\mathbf{R}_{\overline{\mathbf{y}}_1} = E\left[\overline{\mathbf{y}}_1(n)\overline{\mathbf{y}}_1^T(n)\right]$. Hence, we can formulate a criterion that aims at the match of an estimated multivariate density to a multivariate Gaussian distribution with zero mean and covariance matrix $\mathbf{R}_{\overline{\mathbf{y}}_1}$.

$$\begin{aligned} J_{GM}(\boldsymbol{\theta}_1) &= \int_D \left(f_{\overline{Y}_1}(\mathbf{v}) - G_{\mathbf{R}_{\overline{\mathbf{y}}_1}}(\mathbf{v})\right)^2 d\mathbf{v} \\ &= \int_D f^2_{\overline{Y}_1}(\mathbf{v})d\mathbf{v} + \int_D G^2_{\mathbf{R}_{\overline{\mathbf{y}}_1}}(\mathbf{v})d\mathbf{v} - 2\int_D f_{\overline{Y}_1}(\mathbf{v})G_{\mathbf{R}_{\overline{\mathbf{y}}_1}}(\mathbf{v})d\mathbf{v} \end{aligned} \tag{4}$$

where $f_{\overline{Y}_1}(\mathbf{v})$ is the multivariate density associated with the vector $\overline{\mathbf{y}}_1(n)$ at point $\mathbf{v}$; $G_{\mathbf{R}_{\overline{\mathbf{y}}_1}}(\mathbf{v})$ is a Gaussian distribution with covariance matrix $\mathbf{R}_{\overline{\mathbf{y}}_1}$, $D \in \mathbb{R}^{d_m+1}$ and $\boldsymbol{\theta}_1 = \{\mathbf{g}(\cdot), \mathbf{w}_1\}$, being $\mathbf{w}_1$ the vector corresponding to the first row of $\mathbf{W}$.

To estimate $f_{\overline{Y}_1}(\mathbf{v})$, we consider a kernel density estimation method [12] using Gaussian kernels, which will lead to further simplifications in our case. Hence, the kernel estimate of $f_{\overline{Y}_1}(\mathbf{v})$ is:

$$\hat{f}_{\overline{Y}_1}(\mathbf{v}) = \frac{1}{L}\sum_{i=1}^{L} G_{\Sigma}\left(\mathbf{v} - \overline{\mathbf{y}}_1(i)\right), \tag{5}$$

where L is the number of vector samples of $\overline{\mathbf{y}}_1(n)$ and

$$G_{\Sigma}\left(\mathbf{v} - \overline{\mathbf{y}}_1(i)\right) = \frac{1}{\sqrt{(2\pi)^{d_m+1}|\Sigma|}} exp\left[\frac{-1}{2}(\mathbf{v} - \overline{\mathbf{y}}_1(i))^T\Sigma^{-1}(\mathbf{v} - \overline{\mathbf{y}}_1(i))\right], \tag{6}$$

is the multivariate symmetric Gaussian kernel with covariance matrix $\Sigma = \sigma^2\mathbf{I}$, where $\mathbf{I}$ is the identity matrix of order d_m+1 and σ^2 the kernel size; $|\Sigma|$ is the determinant of Σ. Replacing the estimate $f_{\overline{Y}_1}(\mathbf{v})$ into Eq. (4), it is possible to write:

$$\hat{J}_{GM}(\boldsymbol{\theta}_1) = \frac{1}{L^2}\sum_{i=1}^{L}\sum_{j=1}^{L} G_{2\Sigma}\left(\overline{\mathbf{y}}_1(i) - \overline{\mathbf{y}}_1(j)\right) + G_{2\mathbf{R}_{\overline{\mathbf{y}}_1}}(\mathbf{0}) - \frac{2}{L}\sum_{i=1}^{L} G_{\Sigma+\mathbf{R}_{\overline{\mathbf{y}}_1}}\left(\overline{\mathbf{y}}_1(i)\right) \tag{7}$$

where the following relation was used [8]:

$$\int_D G_\Sigma(\mathbf{v} - \overline{\mathbf{y}}_1(i))\, G_\Sigma(\mathbf{v} - \overline{\mathbf{y}}_1(j))\, d\mathbf{v} = G_{2\Sigma}(\overline{\mathbf{y}}_1(i) - \overline{\mathbf{y}}_1(j))\,. \tag{8}$$

The goal is to minimize the cost $\hat{J}_{GM}(\boldsymbol{\theta}_1)$. It is expected that, in the optimization process, $\mathbf{R}_{\overline{\mathbf{y}}_1}$ converges to a scaled version of $\mathbf{R}_{\overline{\mathbf{s}}_k}$, the temporal covariance matrix of a Gaussian source $s_k(n)$, as will be explained ahead.

It is worth mentioning that this method requires the adjustment of the kernel size σ, which, for Gaussian distributions, can be done using the Silverman's rule [13], i.e., $\sigma_o = \sigma_{y_1}\left(4/(L\,(2(d_m{+}1)+1))\right)^{1/(d_m+5)}$, where σ_{y_1} is the standard deviation of $y_1(n)$. The number of delays, d_m, should be a trade-off between the amount of temporal information used and the computational cost.

3.3 The Combined Approach

With both costs $J_{SOS}(\boldsymbol{\theta})$ and $J_{GM}(\boldsymbol{\theta}_1)$ at hand, we are able to analyze some illustrative cases that might be further clarifying. Nonetheless, for the sake of briefness, we appeal to certain intuitive properties within the BSS problem.

We start by considering the sole minimization of $J_{GM}(\boldsymbol{\theta}_1)$ and, for simplicity, we suppose the $N{=}2$ sources case with the following possible types of sources: (i) only one of the sources is Gaussian distributed and (ii) both sources are Gaussian and temporally colored (with different autocorrelation functions). In the scenario (i), we know that, at the end of the linear mixing problem (with all linear coefficients non-null), $\mathbf{u}(n)$ will tend to have a joint Gaussian distribution, but not exactly Gaussian due to one of the sources being not Gaussian. After the nonlinearities $\mathbf{f}(\cdot)$, it is expected that $\mathbf{x}(n)$ will be even farther from the Gaussian distribution. By forcing $y_1(n)$ to be Gaussian via minimization of $J_{GM}(\boldsymbol{\theta}_1)$, it is expected that the nonlinear separating functions $\mathbf{g}(\cdot)$ are able to produce a Gaussian-like distribution for $\mathbf{z}(n)$, so that the linear separating structure $\mathbf{W}_1$ will be able to extract a Gaussian source, but not necessarily the desired one. Hence, in the case (i), if considered the additional minimization of the cost $J_{SOS}(\boldsymbol{\theta})$, it might be able to recover the correct Gaussian and, consequently, the other source, since their lagged covariance matrices will be jointly diagonalized. In case (ii), since the linear mixtures of Gaussian distributions remains Gaussian, we have that $\mathbf{u}(n)$ would be jointly Gaussian. The nonlinearity $\mathbf{f}(\cdot)$, again, will drive the distribution of $\mathbf{x}(n)$ away from Gaussianity. By minimizing $J_{GM}(\boldsymbol{\theta}_1)$ in this case, it is expected that nonlinearities be compensated, but the linear part will be unable to separate between the two Gaussian sources. Now, if we also consider the minimization of $J_{SOS}(\boldsymbol{\theta})$, we know from the linear BSS theory that Gaussian distributions can be separated and the estimation of the temporal covariance matrix $\mathbf{R}_{\overline{\mathbf{y}}_1}$ will be more precise. Undoubtedly, it is not possible to determine which of the Gaussian sources will be recovered at $y_1(n)$, but, since the BSS problem admits permutation of the solutions, this is not an issue.

In fact, a bond between both SOS and GM criteria emerges in the temporal information used by both costs, where there is an important synergy: the diagonalization of $\mathbf{R}_{\mathbf{y},d_s}$ aids the convergence of $\mathbf{R}_{\overline{\mathbf{y}}_1}$ to $\mathbf{R}_{\overline{\mathbf{s}}_k}$ – the temporal covariance matrix of a Gaussian source – in addition, the information that the source $y_1(n)$ is Gaussian can also contribute to it; in turn, when $\mathbf{R}_{\overline{\mathbf{y}}_1}$ tends to $\mathbf{R}_{\overline{\mathbf{s}}_k}$, it can aid with the separation of the other sources when diagonalizing $\mathbf{R}_{\mathbf{y},d_s}$.

These illustrative cases reveal how the joint minimization of the SOS and GM costs might aid the separation task. Hence, we propose the following combined cost:

$$J_{SOS+GM}(\boldsymbol{\theta}) = J_{SOS}(\boldsymbol{\theta}) + \alpha J_{GM}(\boldsymbol{\theta}_1), \tag{9}$$

where α is a trade-off parameter between costs. The other parameters that require (pre-)adjustment are the number of samples and of time delays.

In the following, we present some performance analysis of the proposed SOS+GM criterion in simulation scenarios.

4 Simulation Results

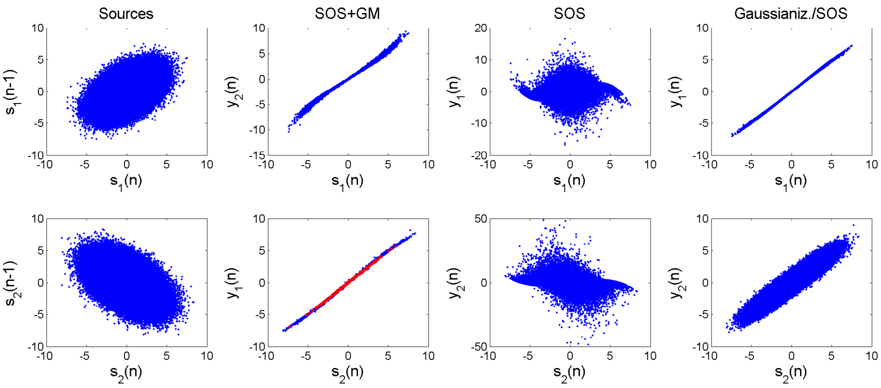
In this section, we analyze the performance of the SOS+GM criterion and compare it with two other methods: the minimization of only the SOS cost (joint approach) and the Gaussianization process followed by the minimization of the SOS cost (the two-stage approach proposed in [7]). For the Gaussianization method, the maximization of Shannon's entropy was considered, using (univariate) Gaussian kernel estimators [14].

The analyses were conducted in two scenarios. In the first one, we consider two Gaussian sources that are temporally colored by the finite impulse response (FIR) filters $h_1(z) = 1{+}0.5z^{-1} + 0.2z^{-2}$ and $h_2(z) = 1{-}0.8z^{-1}$, one for each source. For the second scenario, one of the sources is a temporally correlated Gaussian (by the filter $h_1(z)$) and the other is a uniformly distributed signal (from -1 to $+1$) with no temporal structure. In both scenarios, the mixtures were the result of $\mathbf{x}(n) = (\mathbf{A}\mathbf{s}(n))^3$, being $\mathbf{A} = [0.25\ 0.86; -0.86\ 0.25]$. For the separating structure, we considered, in place of $\mathbf{g}(\cdot)$, parametric functions of the type $z_i(n) = g_{i,1}x_i(n) + g_{i,2}\mathrm{sign}(x_i(n))\sqrt[3]{|x_i(n)|}$, where the operator sign$(\cdot)$ returns a $+1$ if $x_i(n) \geq 0$ or a -1 if $x_i(n) < 0$; followed by a 2×2 matrix $\mathbf{W}$.

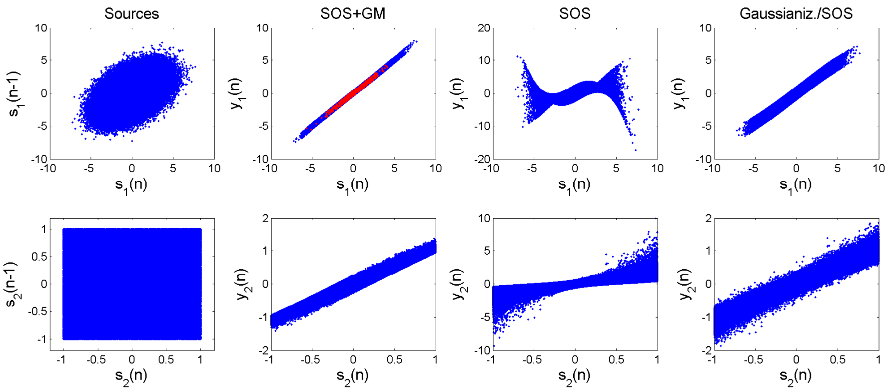
In all cases, the number of delays and the number of samples considered remained fixed. For the SOS cost, common to all considered methods, we adopted 3 delays with $\mathcal{S} = \{0,1,2\}$ and 500,000 samples of $\mathbf{y}(n)$ (the SOS cost demanded a higher accuracy in its estimation, hence the large number of samples). For the Gaussianization method, 500 samples of the vector $\mathbf{z}(n)$ were used to estimate the marginal entropies. For the GM cost (part of the SOS+GM criterion), we considered $d_m = 1$ (delays 0 and 1), 500 samples of the vector $\overline{\mathbf{y}}_1(n)$ and $\alpha = 1$.

To perform the optimization of the weights (nonlinear and linear), we adopted the metaheuristic known as Differential Evolution (DE) [9]. The DE parameters were chosen to be $N_P = 300$ (population size), $F = 0.7$, $CR = 0.7$ and 100 iterations – for more details, please refer to [9]. For the joint approaches, a single run of the DE adapts all coefficients, while, for the two-stage approach, two DE runs are necessary, one for the nonlinear and other for the linear part. After training, the performance of the best individual in the population was measured in terms of SIR, defined as $\text{SIR} = 10 \log \left(E[y_i(n)^2]/E[(s_i(n) - y_i(n))^2] \right)$, after sign and variance correction.

Fig. 2 shows, for both scenarios, the scatter plot $s_i(n) \times s_i(n-1)$ of each source and the outputs of the SOS+GM, the SOS and the Gaussianization/SOS methods through the plots $s(n) \times y(n)$, where a diagonal line means that a perfect separation was achieved (the red dots are the output samples used to estimate the GM cost). The measured SIR values for each case are displayed in Tab. 1.



(a) Scenario 1.



(b) Scenario 2.

Fig. 2. Scatter plots of the sources and of the outputs for each method.

In scenario 1, the proposed SOS+GM method was able to recover both sources with high SIR values. The output $y_1(n)$, in this case, recovered the source $s_2(n)$ and preserved its temporal structure with higher precision, being, consequently, associated with a higher SIR level. The second method (sole SOS criterion) has not performed well, being its plots $s_i(n) \times y_i(n)$ in Fig. 2(a) far from a diagonal line and with outputs associated with low values of SIR. For the two-stage Gaussianization/SOS method, Gaussian signals were recovered, but just one of them preserved the temporal structure of the source ($s_1(n)$). In fact, this result comes from a drawback of the two-stage approach: in the Gaussianization step, the outputs $\mathbf{z}(n)$ can be very close to Gaussian distributions, but may carry a small nonlinear residue (due to precision issues on estimation, for example); then, in the linear separation step, this residue can not be treated. Indeed, in the simulations, the optimization of the SOS cost in the second-stage was not able to achieve its lowest value, since the nonlinear residue could not be treated by a linear structure. Even though, the SIR value of 19.49 dB obtained in the estimation of $s_2(n)$ can be considered acceptable in nonlinear scenarios (note that, in Fig. 2(a), the deviation of the output $y_2(n)$ from $s_2(n)$ are not severe).

Table 1. Performance in terms of SIR [dB]

	Sources	GM+SOS	SOS	Gauss./SOS
Scenario 1	Gaussian $h_1(z)$ - Source 1	61.5618	10.3715	70.2770
	Gaussian $h_2(z)$ - Source 2	75.0141	-3.3308	19.4858
Scenario 2	Gaussian $h_1(z)$ - Source 1	38.4945	-0.5136	21.3580
	Uniform - Source 2	39.6601	5.7545	34.0684

In the second scenario, our SOS+GM method performed as expected and was able to recover the Gaussian source and its temporal structure at output $y_1(n)$, as indicated by the plot $s_1(n) \times y_1(n)$ of Fig. 2(b). Also, the second source was recovered with an SIR value of 39.66 dB, being a reliable estimate. For the SOS method, again, the performance measures were far from the desired, indicating that the sole minimization of SOS cost is not sufficient for nonlinear separation. Finally, the Gaussianization/SOS method could provide reasonable estimates of the sources, as shown in Fig. 2(b). However, due to the presence of the uniform distribution in the mixture, the Gaussianization step was not able to completely compensate the nonlinearities, causing a reduction on the performance. Indeed, in this scenario, since the proposed method does not encompass any assumption on the distribution shape of the sources different from the one that is Gaussian, it can obtain better results.

5 Conclusions

In this work, we have proposed a joint approach for source separation in the PNL model. The method allies the use of the second-order statistics to the density matching approach. By only assuming temporally colored sources and, at least, one Gaussian source, this method is able to perform the separation based on less restrictive requirements than the usual two-stage methods, whose assumptions apply to all sources. Also, it can be computationally simpler than estimating mutual independence in the classical ICA framework. Along with the use of the DE metaheuristic, the simulations indicated that the proposed method is more robust than the Gaussianization method in the case of two Gaussian sources and in the case of one Gaussian and one uniformly distributed source.

Since this work is still in its initial stage, there are plenty of possibilities for future works. We consider, for instance, the analysis of the conditions for the extension to a higher number of sources; the assumption of a known source distribution which is not Gaussian; and, finally, the proposition of a gradient-based algorithm.

Acknowledgements. This work was partly supported by FAPESP (2013/14185-2, 2015/23424-6), CNPq and ERC project 2012-ERC-AdG-320684 CHESS.

References

1. Comon, P., Jutten, C.: Handbook of Blind Source Separation: Independent Component Analysis and Applications. Academic Press (2010)
2. Duarte, L. T., Jutten, C., Moussaoui, S.: A Bayesian Nonlinear Source Separation Method for Smart Ion-Selective Electrode Arrays. IEEE Sensors Journal 9(12), 1763–1771 (2009)
3. Meganem, I., Deville, Y., Hosseini, S., Déliot, P., Briottet, X., Duarte, L. T.: Linear-Quadratic and Polynomial Non-Negative Matrix Factorization; Application to Spectral Unmixing. 19th IEEE Euro. Signal Processing Conf., 1859–1863 (2011)
4. Hosseini, S., Jutten, C.:On the Separability of Nonlinear Mixtures of Temporally Correlated Sources. IEEE Signal Processing Let. 10(2), 43–46 (2003)
5. Taleb, A., Jutten, C.:Source Separation in Post-Nonlinear Mixtures. IEEE Trans. Signal Process. 47(10), 2807–2820, (1999)
6. Deville, Y., Duarte, L. T.:An Overview of Blind Source Separation Methods for Linear-Quadratic and Post-Nonlinear Mixtures. International Conference on Latent Variable Analysis and Signal Separation, 155–167 (2015)
7. Ziehe, A., Kawanabe, M., Harmeling, S., Müller, K.R.:Blind Separation of Post-Nonlinear Mixtures Using Gaussianizing Transformations and Temporal Decorrelation. J. Mach. Learn. Res. 4, 1319–1338 (2003)
8. Fantinato, D., Boccato, L., Neves, A., Attux, R.:Multivariate PDF Matching via Kernel Density Estimation. Proc. of the Symp. Series on Comp. Int. (2014)
9. Price, K., Storn, R., Lampinen, J.:Differential Evolution: A Practical Approach to Global Optimization. Springer (2005)
10. Duarte, L.T., Suyama, R., Attux, R., Rivet, B., Jutten, C., Romano, J.M.T.:Blind Compensation of Nonlinear Distortions: Application to Source Separation of Post-Nonlinear Mixtures. IEEE Trans. Sig. Process. 60, 5832–5844 (2012)
11. White, S.A.:Restoration of Nonlinearly Distorted Audio by Histogram Equalization. J. Audio Eng. Soc. 30(11), 828–832 (1982)
12. Parzen, E.: On the Estimation of a Probability Density Function and the Mode. Ann. Math. Statist 33, 1065–1067 (1962)
13. Silverman, B.:Density Estimation for Statistics and Data Analysis. Chapman and Hall (1986)
14. Silva, L, de Sá, J., Alexandre, L.: Neural Network Classification Using Shannon's Entropy. Europ. Symp. on Art. Neural Net., 217–222 (2005)