



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

Denis G Fantinato, Leonardo T Duarte, Paolo Zanini, Bertrand Rivet, Romis Attux, Christian Jutten

► To cite this version:

Denis G Fantinato, Leonardo T Duarte, Paolo Zanini, Bertrand Rivet, Romis Attux, et al.. A Joint Second-Order Statistics and Density Matching-Based Approach for Separation of Post-Nonlinear Mixtures. LVA/ICA 2017 - 13th International Conference on Latent Variable Analysis and Signal Separation, Olivier Michel; Nadège Thirion-Moreau, Feb 2017, Grenoble, France. pp.499 - 508, 10.1007/978-3-319-53547-0_47 . hal-01479431

HAL Id: hal-01479431

<https://hal.science/hal-01479431v1>

Submitted on 28 Feb 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

Denis G. Fantinato^{1(✉)}, Leonardo T. Duarte², Paolo Zanini³, Bertrand Rivet³,
Romis Attux¹, and Christian Jutten³

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, however, it is usually required the knowledge of the distribution type for all sources, what may be difficult to attend. In view of this, in this work, we propose a less restrictive approach, in which it is required the knowledge of a single source distribution – 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 mixtures samples, this problem is named Blind Source Separation (BSS) [1]. Most part 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 in 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 the evaluation of 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 may depend on certain assumptions over the distribution of all sources in order to exploit its statistical features, which might imply in rather strong assumptions for some practical scenarios [6]. In view of this, 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 from the observed mixtures $\mathbf{x}(n) = \Phi(\mathbf{s}(n))$ the original sources $\mathbf{s}(n)$, 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 $\Phi(\cdot)$ 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 $\mathbf{f}(\cdot)$ is a set of M component-wise functions. The separation system is a mirrored version of that of mixing, 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 in two main classes: the *joint* and the *two-stage* approaches [6].

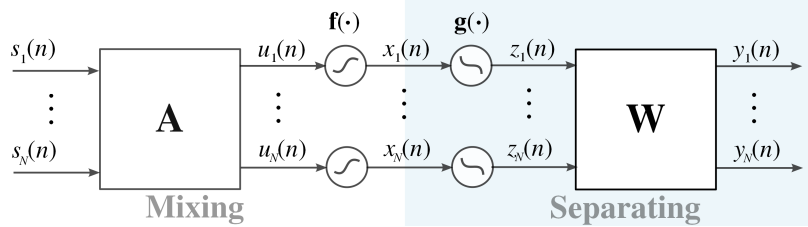


Fig. 1. Mixing and separating systems in the PNL model.

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 –, which are based on some *a priori* information, for instance, sparse and bandlimited [10] signals, 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. In that sense, the most intuitive idea for adapting $\mathbf{g}(\cdot)$ is to turn 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 knowledge of the distribution of $\mathbf{u}(n)$ is required as *a priori* information. 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 [1],[6], which encompasses high-order statistics (HOS). 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. In the work [7], this idea is exploited 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 important because it merges the simplicity of the second-order framework to the complexity of the nonlinear mixtures, which can allow us to improve the perception of this problem.

Although both approaches present its own particular advantages, in this work, we propose the use of a joint approach which is able, to a certain extent, to ally 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 allies the use of SOS and the matching of a (Gaussian) probability density. We start from the assumptions: (i) that at least one source is Gaussian; (ii) that the sources are jointly wide-sense stationary, present a temporal structure (in other words, temporally colored) and are mutually independent; (iii) $\mathbf{f}(\cdot)$ is a set of invertible nonlinear functions; and (iv) all the linear mixing coefficients are non-null.

Since we aim at the joint approach, we seek for a single separation criterion which should be able to jointly adapt $\mathbf{g}(\cdot)$ and $\mathbf{W}$. Notwithstanding, 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:

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

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

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

being $\text{off}(\cdot)$ 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, viz., $\boldsymbol{\theta} = \{\mathbf{g}(\cdot), \mathbf{W}\}$. An additional normalization term $(E[y_i^2(n)] - 1)^2$ for $i = \{1, \dots, N\}$ is considered, since there are no whitening step for the nonlinear case. For separation purposes, it is desired that $J_{SOS}(\boldsymbol{\theta})$ 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 are 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. In literature, there can be found several methods for Gaussianization [6],[7], however, in this work, as we aim at a more general density matching approach to be held in a multi-dimensional context (and encompass the temporal structure), we consider 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 and to present a given temporal correlation (from a covariance matrix). This approach is somehow analogous to the supposition of autoregressive (AR) models for one of the sources, however, in this case, we are not interested in the AR parameters, just in the resultant covariance matrix. This point will be clarified in the following.

In order to use the temporal information, we compose vectors of the type:

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

where d_m is the maximum number of delays considered. Note that we are only considering samples of the first output $y_1(n)$. In this case, the temporal covariance matrix of $\bar{\mathbf{y}}_1(n)$ is $\mathbf{R}_{\bar{\mathbf{y}}_1} = E[\bar{\mathbf{y}}_1(n)\bar{\mathbf{y}}_1^T(n)]$. For AR models, the matrix $\mathbf{R}_{\bar{\mathbf{y}}_1}$ is used for estimating the AR parameters, but in our case, it suffices that $y_1(n)$ is associated to a Gaussian distribution with covariance matrix $\mathbf{R}_{\bar{\mathbf{y}}_1}$.

Based on this premise, 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 $\mathbf{R}_{\bar{\mathbf{y}}_1}$.

$$\begin{aligned} J_{GM}(\boldsymbol{\theta}_1) &= \int_D (f_{\bar{\mathbf{y}}_1}(\mathbf{v}) - G_{\mathbf{R}_{\bar{\mathbf{y}}_1}}(\mathbf{v}))^2 d\mathbf{v} \\ &= \int_D f_{\bar{\mathbf{y}}_1}^2(\mathbf{v}) d\mathbf{v} + \int_D G_{\mathbf{R}_{\bar{\mathbf{y}}_1}}^2(\mathbf{v}) d\mathbf{v} - 2 \int_D f_{\bar{\mathbf{y}}_1}(\mathbf{v}) G_{\mathbf{R}_{\bar{\mathbf{y}}_1}}(\mathbf{v}) d\mathbf{v} \end{aligned} \quad (4)$$

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

For estimation of $f_{\bar{\mathbf{y}}_1}(\mathbf{v})$, we consider a kernel density estimation method [12] allied to the choice of Gaussian functions as kernels, which will lead to further simplifications in our case. Hence, the $f_{\bar{\mathbf{y}}_1}(\mathbf{v})$ can be estimated as:

$$\hat{f}_{\bar{\mathbf{y}}_1}(\mathbf{v}) = \frac{1}{L} \sum_{i=1}^L G_{\Sigma}(\mathbf{v} - \bar{\mathbf{y}}_1(i)), \quad (5)$$

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

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

is the multivariate symmetric Gaussian kernel with covariance matrix $\Sigma = \sigma^2 \mathbf{I}$, being $\mathbf{I}$ the identity matrix of order d_m+1 and σ^2 the kernel size; $|\Sigma|$ is the determinant of Σ . Substituting the estimation $f_{\bar{\mathbf{y}}_1}(\mathbf{v})$ into Eq. (4), it is possible to write, after some algebraic manipulation:

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

Thus, we wish to minimize the cost $\hat{J}_{GM}(\boldsymbol{\theta}_1)$ so that, ideally, $\hat{J}_{GM}(\boldsymbol{\theta}_1) = 0$ and $\hat{f}_{\bar{\mathbf{y}}_1}(\mathbf{v})$ will be the desired Gaussian distribution – in this case, no constraints over $\boldsymbol{\theta}_1$ are necessary. It is also expected that, in the optimization process, $\mathbf{R}_{\bar{\mathbf{y}}_1}$ converges to a scaled version of $\mathbf{R}_{\bar{\mathbf{s}}_k}$, the temporal covariance matrix of a Gaussian source $s_k(n)$ – this point will be explained ahead.

It is worth mentioning that this method requires the adjustment of the kernel size σ , which, since we face Gaussian distributions, the Silverman's rule can be successfully applied [13]. Hence, the optimal value for σ – in a squared error sense – is $\sigma_o = \sigma_{y_1} (4/(L(2(d_m+1)+1)))^{1/(d_m+5)}$, where σ_{y_1} is the standard deviation of $y_1(n)$. The number of delays, d_m , is recommended to be small in order to avoid an elevated computational cost.

Finally, with both costs $\hat{J}_{SOS}(\boldsymbol{\theta})$ and $\hat{J}_{GM}(\boldsymbol{\theta}_1)$ at hands, we can define a combined criterion by minimizing $J(\boldsymbol{\theta}) = \hat{J}_{SOS}(\boldsymbol{\theta}) + \hat{J}_{GM}(\boldsymbol{\theta}_1)$, w.r.t. $\boldsymbol{\theta}$. The other parameters that requires (pre-)adjustment are the number of samples and of time delays for both costs.

4 Some Aspects of the Proposal

Although it might seem at a first glance that the two costs $J_{SOS}(\boldsymbol{\theta})$ and $J_{GM}(\boldsymbol{\theta}_1)$ – Eqs. (2) and (7), respectively – are not closely related, they perform together an essential role for nonlinear BSS separation. In the sequence, we consider some illustrative cases that might be further clarifying, however, 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 assume 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. 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 $\mathbf{x}(n)$ drives even more away 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 this case (i), the minimization of the cost $J_{SOS}(\boldsymbol{\theta})$ must perform an important role to make the Gaussian convert to the correct one. Notwithstanding, in scenario (ii), since the

mixtures of Gaussian distributions remains Gaussian, we have that $\mathbf{u}(n)$ would be jointly Gaussian. The nonlinearity $\mathbf{f}(\cdot)$, again, will push 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 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 $\mathbf{R}_{\bar{\mathbf{y}}_1}$ will be more precise. Undoubtedly, it is not possible to determine which of the Gaussian source will be recovered at $y_1(n)$, but since the BSS problem admits permutation of the solutions this is not an issue.

On the other hand, by considering only the minimization of $J_{SOS}(\boldsymbol{\theta})$, we know it is sufficient for linear mixtures, but this might not hold for nonlinear mixtures [1]. Indeed, if the nonlinearities $\mathbf{f}(\cdot)$ are not compensated by $\mathbf{g}(\cdot)$, possibly an infinite number of independent solutions can be achieved – i.e., depending on the degree of freedom of the combined nonlinearities $\mathbf{f} \circ \mathbf{g}$. However, by constraining one of the sources to be Gaussian, it is expected a single possible solution for $J_{SOS}(\boldsymbol{\theta})$, which is the desired one. In fact, the bond between both criteria happens in the temporal information used in both costs, where there is an important synergy: the minimization of the $\mathbf{R}_{\mathbf{y},d_s}$ diagonals aids the convergence of $\mathbf{R}_{\bar{\mathbf{y}}_1}$ to $\mathbf{R}_{\bar{\mathbf{s}}_k}$ – 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 $\mathbf{R}_{\bar{\mathbf{y}}_1}$ tends to $\mathbf{R}_{\bar{\mathbf{s}}_k}$, it can aid the the second-order independence in $\mathbf{R}_{\mathbf{y},d_s}$. To better illustrate it, in the next section, we present some simulation results.

5 Simulation Results

In order to test our proposed method, we consider two simulation scenarios and a performance comparison along with the methods: SOS (joint approach) and Gaussianization followed by SOS (two-stage approach [7]). For the Gaussianization, it was considered the maximization of the Shannon’s entropy, using (univariate) Gaussian kernel estimators [14]. In the first scenario, 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. 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 parametric functions for $\mathbf{g}(\cdot)$ of the type $z_i(n) = g_{i,1}x_i(n) + g_{i,2}\text{sign}(x_i(n))\sqrt[3]{|x_i(n)|}$, where the operator $\text{sign}(\cdot)$ returns a +1 if $x_i(n) \geq 0$ or a -1 if $x_i(n) < 0$; followed by a 2×2 matrix $\mathbf{W}$.

Common to all considered methods, the SOS cost – defined in Eq. (2) – was chosen to encompass 3 delays with $\mathcal{S} = \{0,1,2\}$ and to be estimated from 500,000 samples of $\mathbf{y}(n)$ (for higher accuracy). Hence, for the entropy and for the density matching – Eq. (4) –, since they are more complex methods, it were considered 500 samples of $\mathbf{z}(n)$ and $y_1(n)$, respectively – note that the density matching requires a lower number of samples.

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, it is necessary two DE runs, 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 (after the sign and variance correction), defined as $SIR = 10 \log (E[y_i(n)^2]/E[(s_i(n) - y_i(n))^2])$, and displayed in Tab. 1. The joint SOS has not performed well and attained low values of SIR, which indicates, as expected, that its sole criterion is not sufficient for separation. For the two-step Gaussianization+SOS method, one of the sources was estimated with SIR of 70.28 dB, but the other source showed a poorer quality with SIR of 19.49 dB. For the proposed method, both sources were recovered with a high SIR value.

Table 1. Performance in terms of SIR [dB]

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

Fig. 2 shows the scatter plot and histogram of outputs for the last two methods (the red scatter plots are the output samples for the entropy and for the density matching). It is possible to see that the Gaussianization step (in the two-stage approach) provided two Gaussian distributions, however, the first Gaussian could not converge to the desired temporal correlation in the linear step (in this case, the cost $J_{SOS}(\theta)$ could not achieve its lowest value), probably due to a small nonlinear residual. For the proposed method, the joint approach helped obtaining the first Gaussian and to turn it uncorrelated to the second source, what contributed to a better performance.

In a second scenario, we consider the sources: one uniformly distributed with no temporal structure and one temporally correlated Gaussian (by the filter $h_1(z)$). The mixture was kept the same. We repeated the procedure and obtained the results shown in Tab. 1. Again, the joint SOS has not performed well, but the other methods reached a reasonable good performance in terms of SIR. As shown in Fig. 3, the Gaussianization+SOS method carried a higher residual nonlinear error due to the presence of the uniform distribution in the mixture, what caused a reduction on the performance. On the other hand, since the proposed method

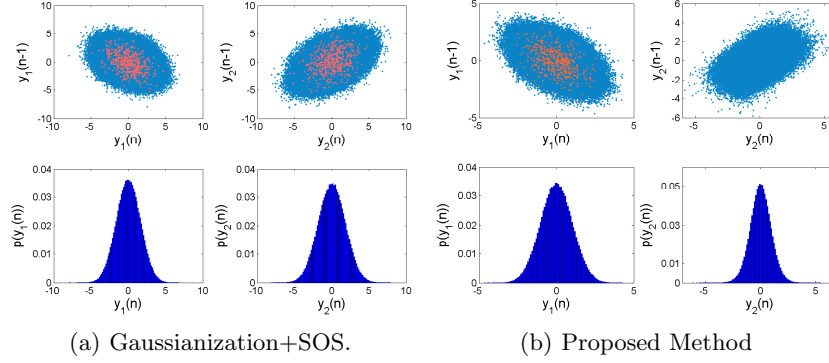


Fig. 2. Scenario 1 - Temporally colored Gaussian sources.

does not encompass any assumption on the distribution shape of the sources different from the one that is Gaussian, it can obtain better results.

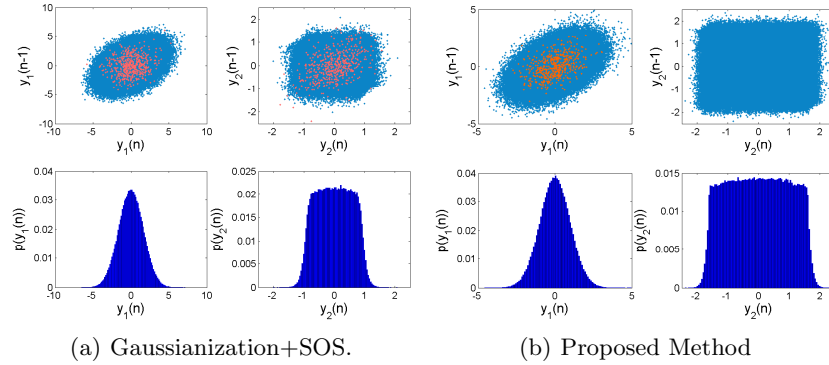


Fig. 3. Scenario 2 - Uniform and Gaussian Sources.

6 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 temporal colored sources and that at least one source is Gaussian distributed, this method is able to perform the separation based on less restrictive requirements than the usual two-stage methods, whose assumptions applies to all sources, and 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 incipient, 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 that one of the sources has a known distribution, but is not Gaussian; and, finally, the proposition of a gradient-based algorithm.

Acknowledgements. This work was supported by FAPESP (2013/14185-2, 2015/23424-6), CNPq and CHES.

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)