XXXVII SIMPÓSIO BRASILEIRO DE TELECOMUNICAÇÕES E PROCESSAMENTO DE SINAIS - SBrT2019, 29/09/2019–02/10/2019, PETRÓPOLIS, RJ

An Epanechnikov Kernel Based Method for Source Separation in Post-Nonlinear Mixtures

Caroline P. A. Moraes, Denis G. Fantinato, Aline Neves

***Abstract*— In the context of the nonlinear Blind Source Separation problem, Post-Nonlinear mixtures can be separated via Mutual Information minimization. In this case, methods based on score functions can be used and the recovered sources distributions can be estimated by kernel methods. Usually a Gaussian kernel function is used. However, other kernel functions with interesting properties can be used, such as the Epanechnikov kernel. Based on this, we apply the Epanechnikov kernel to estimate the pdf and the relative gradient, in order to recover the sources. Also, we compare a classic Gaussian kernel with the Epanechnikov kernel, showing that the latter performs better than the former.**

***Keywords*— Blind Source Separation, Post-Nonlinear Mixtures, Mutual Information, Epanechnikov Kernel.**

I. Introduction

Blind Source Separation (BSS) is a classic problem within the signal processing area occupying a prominent position in view of its versatility and its wide range of practical applications [1]. Since the 80's, great efforts have been dedicated to this problem, whose objective is to create an artificial way to retrieve the signals of unknown sources through a set of observed mixtures [2], [3]. The problem is called "blind" because the sources and the mixing process are unknown. The assumption of linear and instantaneous mixing models present a solid theoretical framework and counts with applications in several areas, such as biomedical signal processing [4], communication systems [1] and geophysical signal analysis [5]. The Independent Component Analysis (ICA) is a set of techniques widely used for solving the BSS problem [5], which assumes the hypothesis that the sources are mutually statistically independent. In that sense, several criteria are used to retrieve the sources, some of them are based on cumulants [6], maximum likelihood [7] and maximum correlation [8].

In many applications such as hyperspectral imaging [9], intelligent chemical sensors arrays [10] and remote sensing [11], the linear assumption is not enough for retrieving the sources, since the observed signal present a nonlinear distortion. However, a generic nonlinear approach still lacks methodology capable of ensuring source separation, which makes this research field very modern and challenging. There are several models that can be applied, being the efforts dedicated to models where ICA techniques are still valid, under some constraints, such as Linear-Quadratic [11] and Bilinear form [12]. In this paper, we investigate nonlinear mixtures using the Post-Nonlinear model (PNL), which is more complex, but also able to represent a larger set of nonlinear processes.

Caroline P. A. Moraes, e-mail: caroline.moraes@ufabc.edu.br; Denis G. Fantinato, e-mail: denis.fantinato@ufabc.edu.br; Aline Neves, e-mail: aline.neves@ufabc.edu.br. UFABC - Federal University of ABC, Santo André - SP. This work was partially supported by FAPESP (2018/17678-3) and CAPES (DS/1766640).

In Section II, we describe the linear and nonlinear mixing models. Section III shows the criterion based on the mutual information and its associated parameters. Section IV analyzes the Epanechnikov and Gaussian kernels, and their respective gradients. In Section V the simulations results are presented and analysed. Finally, we conclude in Section VI.

II. Blind Source Separation

In the past decades a large number of BSS algorithms have been developed and studied, all of them based on linear and/or nonlinear mixing models. In the following we describe both of them.

A. Linear Mixing Model

For the case of linear mixtures, $\mathbf{x}$ is the M-dimensional observation vector $\mathbf{x}(n) = [x_1(n), x_2(n), \ldots, x_M(n)]^T$, $\mathbf{s}$ is the vector of unknown sources $\mathbf{s}(n) = [s_1(n), s_2(n), \ldots, s_N(n)]^T$, where N is the number of sources and $\mathbf{A}$ is the mixing matrix. It can be modeled as:

$$\mathbf{x}(n) = \mathbf{A}\,\mathbf{s}(n), \tag{1}$$

where n is the time index. In this work, we will restrain our study to the cases where $N = M$.

To separate the sources from $\mathbf{x}$, the objective is to obtain a separating matrix $\mathbf{B}$, ideally equal to the inverse of $\mathbf{A}$, up to a scale and a permutation factors. Sources can be recovered by:

$$\mathbf{y}(n) = \mathbf{B}\,\mathbf{x}(n). \tag{2}$$

where $\mathbf{y}(n) = [y_1(n), y_2(n), \ldots, y_N(n)]^T$ is the vector with the estimated sources.

B. Post-Nonlinear Model

The Post-Nonlinear model assumes that the mixing process occurs through a linear stage followed by a nonlinear function $f(\cdot)$, that is supposed to be invertible and monotonic. Mathematically, the process can be modeled by:

$$x_i(n) = f_i\Big(\sum_{j=1}^{N} a_{ij}\, s_j(n)\Big),\ \forall\, i \in \{1, \ldots, M\}, \tag{3}$$

where the term a_{ij} represents the coefficients of the mixing matrix $\mathbf{A}$.

The aim is to find the matrix $\mathbf{B}$ and the function $g(\cdot)$ that could retrieve the sources, so that:

$$y_i(n) = \sum_{j=1}^{N} b_{ij}(g_i(x_j(n))), \quad (4)$$

in which b_{ij} represents the coefficients of the matrix $\mathbf{B}$ and $y_i(n)$ represents the *i-th* recovered source, admitting scale invariance and/or permutation between the signals.

The system model is as illustrated in Fig. 1.

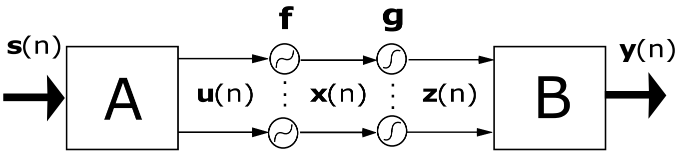

Fig. 1. Mixture Model Post-Nonlinear.

III. Criterion Based on Mutual Information

There are several criteria that allow solving the BSS problem. In [13], the authors proposed a criterion based on the minimization of the mutual information among the estimated sources using an adaptive estimation of the log-derivative of densities.

Mathematically, the accurate mutual information (MI) measure is very difficult to be obtained since it demands knowledge of the joint and marginal signals probability density functions (pdf). The work of Taleb and Jutten [13] proposes a simplified criterion, by expressing the MI only as a function of the marginal entropies, resulting in:

$$C(Y) = \sum_{i=1}^{N} H(Y_i) - \sum_{i=1}^{N} H(Z_i) - \log|\det B|, \quad (5)$$

where Y_i is the random variable associated with $y_i(n)$, Z_i is the random variable associated with $z_i(n)$ and $H(\cdot)$ is Shannon's entropy function. As shown in Fig. 1, $z_i(n) = g(x_i(n))$.

The marginal entropy can be estimated using a sliding window (l) and is expressed as:

$$\hat{H}(D) = -\frac{1}{l}\sum_{n=1}^{l} \log(P_D(D(n))),\ \forall\, n \in \{1, \ldots, l\}, \quad (6)$$

where D is a random variable and P_D is its associated probability density function whose estimation will be discussed in Section IV.

Having defined the cost function given by (5), its optimization may be achieved through an algorithm that iteratively updates the matrix $\mathbf{B}$ and the nonlinear function $g(\cdot)$ using the relative gradient [14]. This method was developed in the context of the EASI (Equivariant Adaptive Source Separation) technique, showing good convergence properties.

For $\mathbf{B}$, the adaptation equation is:

$$\mathbf{B} \leftarrow (\mathbf{I} - \lambda\, \psi(Y_i))\mathbf{B}, \quad (7)$$

where $\mathbf{I}$ is the identity matrix, λ is the adaptation step size and $\psi(Y_i)$ is the score function of the gradient. At each iteration, the rows of B are also normalized. For the nonlinear function $g(\cdot)$,

$$g_i \leftarrow (1 - \mu\, \delta(Y_i, Z_i))g_i, \quad (8)$$

where μ is the adaptation step size, and $\delta(Y_i, Z_i)$ is the score function of the gradient in each iteration.

The score functions $\psi(Y_i)$ and $\delta(Y_i, Z_i)$ depend on P_Y and P_Z. Its estimation will be detailed in the following section.

IV. Kernel Density Estimation

Parzen window is a non-parametric method to estimate the pdf through the use of kernel functions, based on a finite data sample [15]:

$$\hat{P}_D(D(n)) = \frac{1}{l}\sum_{m=1}^{l} K(d(n) - d(m)), \quad (9)$$

where $K(\cdot)$ is the kernel function.

This method consists in summing several kernel functions, centered at each signal sample, for estimating the pdf. The function must present certain characteristics to be used as kernel [16]. The Central Limit Theorem (CLT) [5] establishes that, in distributions of independent random variables, if we combine an infinite number of variables, the distribution will tend to a normal distribution. Due to the CLT and the large occurrence of Gaussian signals in nature, the Gaussian kernel is one of the most used:

$$K_{gau}(x) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(\frac{-x^2}{2\sigma^2}\right), \quad (10)$$

where σ is the bandwidth of the estimator and x is the sample value.

However, in [16] it is shown that, based on the AMISE (Asymptotic Mean Integrated Square Error) the Epanechnikov kernel has a better performance for pdf estimation than the Gaussian kernel. In the analysis developed in [16], Epanechnikov is shown to be an optimal kernel for pdf estimation. It's expression is given by:

$$K_{epa}(x) = \frac{3}{4\sigma}\left(1 - \frac{x^2}{\sigma^2}\right), -\sigma \leq x \leq \sigma. \quad (11)$$

As described in Section III, the MI based criterion (5) can be optimized in two stages: a linear one given by (7) and nonlinear one given by (8). In the following we will discuss each one of them taking into account the kernel function chosen to estimate the $P_X(X)$.

A. Linear Stage

In the linear stage, matrix $\mathbf{B}$ is updated through (7). We may define $\psi(Y_i)$ as:

$$\psi(Y_i) = \sum_{n=1}^{N} \frac{\partial \hat{H}(Y_i)}{\partial b_{ij}}, \quad (12)$$

where, for the Gaussian kernel,

$$\frac{\partial \hat{H}(Y_i)}{\partial b_{ij}} = \frac{1}{l^2}\sum_{n=1}^{l}\frac{1}{\hat{P}_Y(Y_i)}\sum_{m=1}^{l}\frac{K_{gau}(Y_i)\mathcal{Y}_{i_{n,m}}\ \mathcal{Z}_{j_{n,m}}}{\sigma_i}, \quad (13)$$

and for the Epanechnikov kernel

$$\frac{\partial \hat{H}(Y_i)}{\partial b_{ij}} = \frac{1}{l^2}\sum_{n=1}^{l}\frac{1}{\hat{P}_Y(Y_i)}\sum_{m=1}^{l}\frac{6\ \mathcal{Y}_{i_{n,m}}\ \mathcal{Z}_{j_{n,m}}}{4\sigma_i^3}, \quad (14)$$

where $\mathcal{Y}_{i_{n,m}} = y_i(n) - y_i(m)$ is the vector of recovered sources and $\mathcal{Z}_{j_{n,m}} = z_j(n) - z_j(m)$ is the signal obtained after the function $g(\cdot)$.

B. Non-Linear Stage

The nonlinear function $g(\cdot)$ can be approximated by a truncated Taylor series expansion [17], as follows:

$$g_i(x_i(n)) = \mathbf{c}_i^T\ \boldsymbol{\xi}_i(n),\ \forall\ i \in \{1,\dots,M\} \quad (15)$$

where $\mathbf{c}_i = [c_{i1},\ c_{i2}\ \dots\ c_{ib}]^T$ are the coefficients of function $g_i(\cdot)$ and $\boldsymbol{\xi}_i(n) = [x_i^1(n)\ x_i^2(n)\ \dots\ x_i^b(n)]^T$ are the nonlinear mixtures with exponents from 1 to b, being b the highest degree of the truncated series.

An iteration of the algorithm includes the linear update of $\mathbf{B}$ followed by the nonlinear update of $g_i(\cdot)$. This last stage is obtained by taking the derivative of (5) with respect to g_i's coefficients $\mathbf{c}_i$. Thus, in the update equation (8),

$$\delta(Y_i, Z_i) = \sum_{n=1}^{N}\frac{\partial \hat{H}(Y_i)}{\partial c_i} - \sum_{n=1}^{N}\frac{\partial \hat{H}(Z_i)}{\partial c_i}. \quad (16)$$

For the Gaussian kernel, this leads to:

$$\frac{\partial \hat{H}(Y_i)}{\partial c_i} = \frac{1}{l^2}\sum_{n=1}^{l}\frac{1}{\hat{P}_Y(Y_i)}\sum_{m=1}^{l}\frac{K_{gau}(Y_i)\mathcal{Y}_{i_{n,m}}\mathbf{b}_i}{\sigma_i}\frac{\partial \mathcal{Z}_{n,m}}{\partial c_i}, \quad (17)$$

$$\frac{\partial \hat{H}(Z_i)}{\partial c_i} = \frac{1}{l^2}\sum_{n=1}^{l}\frac{1}{\hat{P}_Z(Z_i)}\sum_{m=1}^{l}\frac{K_{gau}(Z_i)\mathcal{Z}_{i_{n,m}}}{\sigma_i}\frac{\partial \mathcal{Z}_{i_{n,m}}}{\partial c_i}, \quad (18)$$

where $\mathbf{b}_i = [b_{i1}\ b_{i2}]$ is the row vector of matrix $\mathbf{B}$.
For the Epanechnikov kernel,

$$\frac{\partial \hat{H}(Y_i)}{\partial c_i} = \frac{1}{l^2}\sum_{n=1}^{l}\frac{1}{\hat{P}_Y(Y_i)}\sum_{m=1}^{l}\frac{6\ \mathcal{Y}_{i_{n,m}}\ \mathbf{b}_i}{4\sigma_i^3}\ \frac{\partial \mathcal{Z}_{n,m}}{\partial c_i}, \quad (19)$$

$$\frac{\partial \hat{H}(Z_i)}{\partial c_i} = \frac{1}{l^2}\sum_{n=1}^{l}\frac{1}{\hat{P}_Z(Z_i)}\sum_{m=1}^{l}\frac{6\ \mathcal{Z}_{i_{n,m}}}{4\sigma_i^3}\frac{\partial \mathcal{Z}_{i_{n,m}}}{\partial c_i}. \quad (20)$$

V. Performance Analysis

A. Simulation Scenario

In the previous section we described a method for PNL mixtures separation based on the use of two different kernel functions, the Gaussian and the Epanhechnikov ones. In order to understand the performance difference between implementing the method with each of these two kernel functions, in this section we present some simulations results. We considered two independent sources both with uniform distribution and with a range from -1 to 1.

Firstly the sources were mixed by matrix $\mathbf{A}$:

$$\mathbf{A} = \begin{bmatrix} 0.65 & 0.23 \\ 0.35 & 0.76 \end{bmatrix}$$

and, in the sequel, a cubic non-linear function $f(\cdot)$, is applied, resulting in:

$$x_i(n) = f_i(u_i(n)) = (u_i(n))^3. \quad (21)$$

As shown in Section II, function $g(\cdot)$ searches the inverse of $f(\cdot)$ to recover the sources:

$$g(x_i) = c_1\ x_i\ +\ c_2\ sgn(x_i)\sqrt[3]{|x_i|}. \quad (22)$$

where $sgn(x_i)$ is the signal function. In this work, we considered the same $f(\cdot)$ for all mixtures, thus $c_{11} = c_{21} = c_1$ and $c_{12} = c_{22} = c_2$.

Ideally, if the coefficient $c_1 = 0$ and $c_2 = 1$, we have that $g(\cdot) = f^{-1}(\cdot)$, i.e., exactly the inverse of the nonlinear function $f(\cdot)$. However, such solution is hardly attained due to the impossibility of obtaining an accurate pdf and to the presence of imprecisions in the estimation process. The algorithm tries to obtain the best solution possible, close to the ideal one.

In order to evaluate the algorithms performance, we will use the Signal-to-Interference Ratio (SIR) that can be obtained by the expression:

$$SIR = 10\ \log\left(\frac{E[y_i(n)^2]}{E[(s_i(n) - y_i(n))^2]}\right). \quad (23)$$

Thereby, high values of SIR mean that sources were retrieved with higher quality.

B. Results and Discussion

In order to study the effects and the performance of the method with the Epanechnikov and the Gaussian kernel, in this section we will analyze the behavior of the algorithm presented for the separation of the PNL mixture.

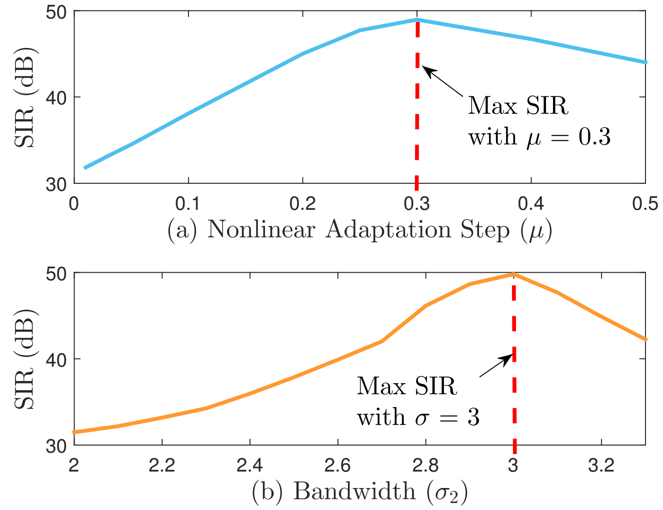

Fig. 2. Sigmas and Steps Analysis.

TABLE I
THE CHOSEN VALUES OF THE BEST STEP SIZES AND BANDWIDTHS.

	K_{epa}	K_{gau}
Linear Step λ	0.1	0.03
Nonlinear Step μ	0.3	0.1
σ_1	3.3	1.3
σ_2	3	1.6

1) Kernel Bandwidth and Adaptation Steps: Firstly, some tests were performed to choose the algorithms parameters that would provide the best results. This is necessary because the parameters σ, l μ and λ have significant influence on the results. Due to their interdependence, it is necessary that they be adjusted jointly, to achieve the maximization of the SIR.

Fig. 2 shows some examples of how the SIR values vary with respect to the step size (figure (a)) and the bandwidth (figure (b)). The displayed results are the mean SIR values obtained after 10 Monte Carlo simulations, sweeping values of μ and σ. For figure (a), we used $\sigma_1 = 3.3$, $\sigma_2 = 3$, $l = 100$ and $\lambda = 0.1$, where σ_1 refers to the first source while σ_2 refers to the second one. For figure (b) we used $\sigma_1 = 3.3$, $l = 100$ $\mu = 0.3$ and $\lambda = 0.1$. In both cases, the Epanechnikov kernel was used. Fig. 2a shows that the best value for the nonlinear stage step size, is $\mu = 0.3$. Fig. 2b shows the performance with respect to the bandwidth, achieving the maximum value of SIR with $\sigma_2 = 3$. The same procedure was executed to select the best parameters for the Gaussian kernel.

Throughout this analysis, it was possible to chose the best parameters for both kernels. Tab. I presents the values used for the step sizes and the bandwidths. In general, the linear stage step sizes are smaller than nonlinear stage ones and the Gaussian kernel presented its best performance with the smallest of them ($\lambda = 0.03$) which may lead the algorithm to converge slower. The Epanechnikov kernel bandwidths are higher than the ones used by Gaussian kernel, but these values depend on the parameters interdependence.

2) Sliding Window Size: Another very important parameter is the sliding window size, l, due to its high relevance in

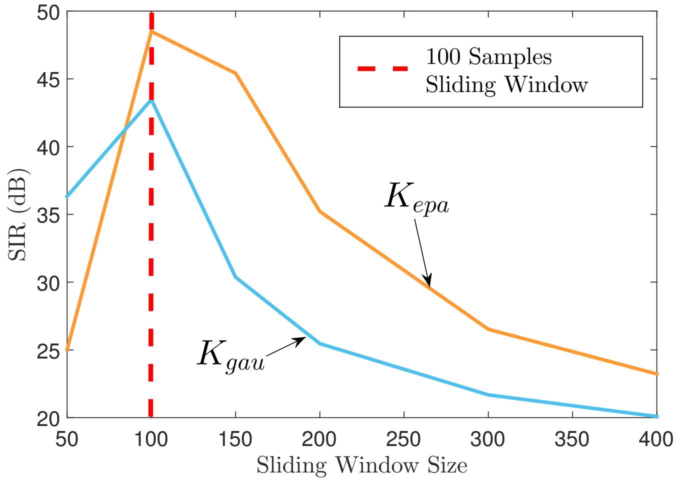

Fig. 3. Variation of Sliding Window Size for Epanechnikov and Gaussian kernel.

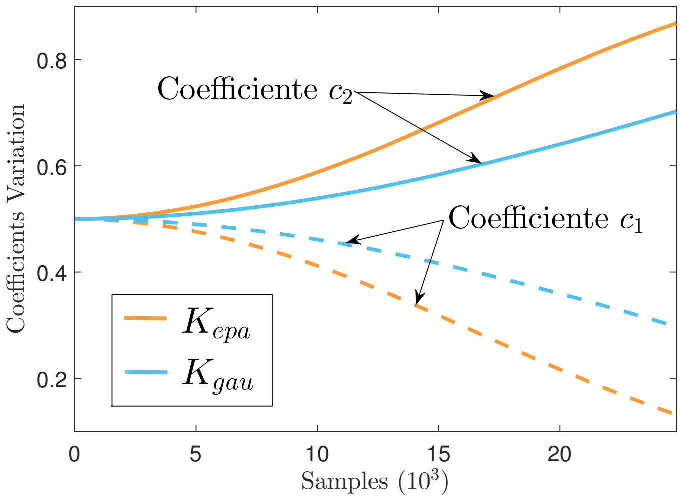

Fig. 4. Comparison of the coefficients convergence for Epanechnikov and Gaussian kernel.

the SIR values, as shown in Fig. 3. For this simulation we used the parameters given in Tab.I. Since the used parameters μ, λ and σ were optimized considering $l = 100$, it was expected the maximum SIR value to happen at this point. Since the parameters were kept fixed during simulation, a loss in performance for other values of l was expected. Due to the interdependence between the parameters, for each value of l it would be possible to obtain the values of μ, λ and σ leading to the highest value of SIR. Comparing the performances of the method with both kernels, we can observe that for l greater than 100 samples, the use of the Epanechnikov kernel always outperforms the use of the Gaussian kernel, presenting a difference that can be higher than 15 dB ($l = 150$) in terms of SIR.

3) Nonlinear Function Coefficients: In Eq. (22), it is shown that the nonlinear function $g(\cdot)$ depends on coefficients c_1 and c_2. Thus, we can analyze the convergence of the nonlinear part of the algorithm by analyzing the behavior of these two coefficients throughout iterations, as shown in Fig. 4.

TABLE II
THE BEST RESULTS OF SIR FOR THE EPANECHNIKOV AND GAUSSIAN KERNEL

	SIR (dB)	
	K_{epa}	K_{gau}
Recovered $\hat{S}_1$	37.93	52.96
Recovered $\hat{S}_2$	61.59	33.98
Average Value	49.76	43.47

In Fig. 4 we note that using the Epanechnikov kernel, the algorithm converges faster than using the Gaussian kernel, keeping the parameters shown in Tab.I. Looking closer, the K_{epa} coefficients final values are: 0.13 and 0.87, while the K_{gau} coefficients are: 0.29 and 0.7. Remembering that the ideal values would be 0 and 1, K_{epa} leads to much better results.

Finally, having chosen the best parameters for each kernel function, we present, in Tab. II, the SIR values for the recovery of the two sources, considering the parameters show in Tab. I, l=100 and an average through 30 simulations. It is possible to observe that, using the Epanechnikov kernel, the SIR results are higher for both sources.

Comparing the SIR value for each source, the Epanechnikov kernel was 3.95 dB and 8.63 dB higher, respectively. The average of the SIR for the recovered signals was 6.29 dB higher than the one obtained with the Gaussian kernel, showing that the Epanechnikov can be better than Gaussian kernel in this case.

VI. CONCLUSIONS

In this work, the BSS problem using the PNL model was studied using the mutual information minimization criterion based on the marginal entropies and on kernel functions to estimate the pdf. We verified, through simulations, that the use of the Epanechnikov kernel for pdf estimation leads to a better result than the Gaussian kernel, leading to an average gain in performance in terms of SIR, of 6.29 dB. We also showed how the choice of the algorithms parameters such as step sizes, kernel bandwidth and sliding window size affect the algorithm. A bad choice of such parameters may largely compromise the methods performance. Furthermore, the Epanechnikov kernel uses a square polynomial function, providing a lower computational cost, while the Gaussian depends on the exponential function. In the future, we consider extending the work using other nonlinearities and applying this method in practical scenarios.

ACKNOWLEDGEMENTS

The authors would like to thank São Paulo Research Foundation (FAPESP) and National Council for the Improvement of Higher Education (CAPES) by the financial support to the work.

REFERENCES

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

[2] J.-F. Cardoso, "Source separation using higher order moments," in *International Conference on Acoustics, Speech, and Signal Processing,*, pp. 2109–2112, IEEE, 1989.

[3] J.-F. Cardoso, "Blind signal separation: statistical principles," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 2009–2025, 1998.

[4] J. M. T. Romano, R. Attux, C. C. Cavalcante, and R. Suyama, *Unsupervised signal processing: channel equalization and source separation*. CRC Press, 2018.

[5] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent component analysis*, vol. 46. John Wiley & Sons, 2004.

[6] A. Hyvarinen, "Blind source separation by nonstationarity of variance: a cumulant-based approach," *IEEE Transactions on Neural Networks*, vol. 12, no. 6, pp. 1471–1474, 2001.

[7] D. T. Pham and P. Garat, "Blind separation of mixture of independent sources through a maximum likelihood approach," in *In Proc. EUSIPCO*, Citeseer, 1997.

[8] F. R. Bach and M. I. Jordan, "Kernel independent component analysis," *Journal of machine learning research*, vol. 3, no. Jul, pp. 1–48, 2002.

[9] I. Meganem, Y. Deville, S. Hosseini, P. Déliot, X. Briottet, and L. T. Duarte, "Linear-quadratic and polynomial non-negative matrix factorization; application to spectral unmixing," in *2011 19th European Signal Processing Conference*, pp. 1859–1863, IEEE, 2011.

[10] L. T. Duarte, C. Jutten, and S. Moussaoui, "A bayesian nonlinear source separation method for smart ion-selective electrode arrays," *IEEE Sensors Journal*, vol. 9, no. 12, pp. 1763–1771, 2009.

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

[12] A. Ypma, A. Leshem, and R. P. Duin, "Blind separation of rotating machine sources: bilinear forms and convolutive mixtures," *Neurocomputing*, vol. 49, no. 1-4, pp. 349–368, 2002.

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

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

[15] D.-Y. Yeung and C. Chow, "Parzen-window network intrusion detectors," in *Object recognition supported by user interaction for service robots*, vol. 4, pp. 385–388, IEEE, 2002.

[16] D. W. Scott, *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons, 2015.

[17] D. G. Fantinato, L. T. Duarte, Y. Deville, C. Jutten, R. Attux, and A. Neves, "Using taylor series expansions and second-order statistics for blind source separation in post-nonlinear mixtures," in *International Conference on Latent Variable Analysis and Signal Separation*, pp. 193–203, Springer, 2018.