Reproducing Kernel Hilbert Space Method for Blind Source Extraction

Eder Santana[1], Ewaldo Santana[2], R. C. S. Freire[3], Allan Kardec Barros[4]

[1]Federal University of Maranhão, São Luís, Brazil, eeder@live.com
[2] Federal University of Campina Grande, Campina Grande, Brazil, ewaldo@fama.br
[3] Federal University of Campina Grande, Campina Grande, Brazil, freire@dee.ufcg.edu.br
[4]Federal University of Maranhão, São Luís, Brazil, allan@elo.com.br

Abstract: We present a method for extracting a specific signal from a blind instantaneous mixture. The method is developed using a priori information about the temporal structure of the desired signal in the Reproducing Kernel Hilbert Space. The approach here carried yields better results than methods present in the literature for Blind Source Extraction using temporal structure.

Keywords: Reproducing kernel Hilbert space, blind source extraction, nonlinear filtering.

1. INTRODUCTION

In Blind Source Extraction (BSE) [1] one wants to recover a desired signal $s_i(t)$ that is linearly mixtured with a finite set of sources. Representing the sources as $\mathbf{s}(t)=[s_1(t),...,s_i(t),...,s_n(t)]^T$, the problem is called blind because we suppose an unknown mixing matrix $\mathbf{A}$ yielding the mixture $\mathbf{x}=\mathbf{As}$. For recovering $s_i(t)$ from the observations $\mathbf{x}$, further suppositions need to be made on the structures of the sources $\mathbf{s}$. In Independent Component Analysis (ICA) [2], for example, one supposes that the signals $s_k(t)$ are statistically independent. This is a strong statistical supposition, but has been proved to be useful in practical problems. Unfortunately, the methods based on ICA simultaneously separate all the sources in $\mathbf{s}$ and, when we are only interested in one signal, it can be a problem. For example, in large array records such as Magnetocardiography (MCG) several channels are available and only feel signals are interesting [1]. In this case, separate all the sources and then find the desired ones can be infadoneous.

This problem can be solved using more specific information about the signals in the mixture, such as their kurtosis [3] or temporal structures [4-7]. The later is what is used in this paper.

Furthermore, we use a kernel method [8] for filtering the desired signal. Reproducing Kernel Hilbert Space (RKHS) methods has recently been introduced in the measurement and signal processing literature with large success because it provides an adequate mathematical tool for solving several problems [9]. In RKHS we suppose that a positive definite symmetric function, κ, called kernel, reproduces an inner product in a high dimensional Hilbert space H . In words we mean that for two vectors $\mathbf{x}$ and $\mathbf{y}$ in the sample space V , there is a nonlinear map, $\phi: V\rightarrow H$, such that

$$\kappa(\mathbf{x},\mathbf{y})=\left\langle \phi(\mathbf{x}),\phi(\mathbf{y})\right\rangle_H \tag{1}$$

The equality in (1) is often referred as the “kernel trick”. With the kernel trick we can evaluate the inner product in H, indirectly by κ, which yields nonlinear methods that can be manipulated by linear algebra. There are several advantages in extending linear methods to a nonlinear RKHS. In Kernel Support Vector Machine (KSVM) [10], one can guarantee the successfulness of the classification due to Cover’s theorem [11] for classification in high dimensional spaces. The recently introduced Kernel Least Mean Square (KLMS) [12] algorithm also surpasses the traditional LMS algorithm by achieving smaller errors and has faster convergence.

As far as the authors know, this paper is the first attempt to develop a RKHS method for extracting specific signals using their temporal structure.

The reminder of the paper is organized as follows. In Section 2 we present our RKHS method. Section 3 is devoted to experiments for illustrating extractions of specific signals in blind mixtures. In Section 4 we conclude the paper.

2. METHOD

For developing the method we will suppose a Hilbert space H, in a fairly informal, but intuitive way. For more precise definitions of RKHS we refer to [13]. Let $\mathbf{x}(t)=[x_1(t),x_2(t),..\ x_n(t)]^T$ be a set of observations, with $t\in[1,2,...,m]$. Now, using a nonlinear map, ϕ, we define the feature space H by the set of all finite linear combinations of the vectors $\phi(\mathbf{x}(t))$. Thus, each vector $\xi\in H$ can be expressed by

$$\xi=\sum_{j=1}^{m}\alpha_j\phi(\mathbf{x}(j)) \tag{2}$$

For practical purposes let us write $\boldsymbol{\xi}(t) = \phi(\mathbf{x}(t))$, where $\boldsymbol{\xi}(t) = [\xi_1(t), \xi_2(t), .. \xi_k(t)]^T$, reminding that k can be possibly infinite depending on the choice of ϕ [8]. Supposing that the vectors $\boldsymbol{\xi}(t)$ have zero mean and were generated by an unknown instantaneous mixture, we have

$$\boldsymbol{\xi}(t) = \mathbf{A}\mathbf{s}(t) \tag{3}$$

where $\mathbf{A}$ is an unknown $k \times k$ mixing matrix and $\mathbf{s}(t)$ are the source vectors. As stated in Section I, our problem is to extract the signal $s_i(t)$. Thence, we must find a vector $\boldsymbol{\omega} \in H$ satisfying

$$y(t) = \left\langle \boldsymbol{\omega}, \boldsymbol{\xi}(t) \right\rangle_H \tag{4}$$

where $y(t)$ is the best estimation, in mean square sense, of $s_i(t)$. Assuming the existence of a positive definite function κ that makes (1) true, we can rewrite (4) as

$$\begin{aligned} y(t) &= \sum_{j=1}^{m} \alpha_j \left\langle \boldsymbol{\xi}(j), \boldsymbol{\xi}(t) \right\rangle_H \\ &= \sum_{j=1}^{m} \alpha_j \kappa(\mathbf{x}(j), \mathbf{x}(t)) \end{aligned} \tag{5}$$

where we assumed the coefficients α_j to satisfy $\boldsymbol{\omega} = \sum_{j=1}^{m} \alpha_j \boldsymbol{\xi}(j)$ because $\boldsymbol{\omega} \in H$. For a proper definition of an objective function to estimate $\boldsymbol{\omega}$, note that if we assume ϕ as the identity mapping and $\kappa(\mathbf{x}(j), \mathbf{x}(t)) = \mathbf{x}(j)^T \mathbf{x}(t)$, our approach turns into a linear filter for extracting $s_i(t)$. Such filter was studded by Barros and Cichocki in [6]. They proposed that the maximization of the autocorrelation $E[y(t)y(t-\tau)]$ is sufficient to approximate $s_i(t)$ by an equation as the one in (5). Further the value of τ is assumed to be a priori known and carries specific information about the desired signal. This a priori information can be resumed as

$$\begin{aligned} &E[s_i(t)s_i(t-\tau)] \neq 0 \\ &E[s_i(t)s_j(t)] = 0 \quad \forall i \neq j \\ &E[s_i(t)s_j(t-\tau)] = 0 \quad \forall i \neq j \end{aligned} \tag{6}$$

With $\mathbf{s}(t)$ satisfying (6) we generalize Barros and Cichocki's approach by assuming κ as any positive definite symmetric function. In this generalization, we propose the following sample estimation of the objective function

$$\begin{aligned} J(\boldsymbol{\alpha}) &= \frac{1}{m} \sum_{j=1}^{m} y(t) y(t-\tau) \\ &= \frac{1}{m} \sum_{j=1}^{m} \sum_{k=1}^{m} \sum_{l=1}^{m} \alpha_j \alpha_k \kappa(\mathbf{x}(j), \mathbf{x}(l)) \kappa(\mathbf{x}(l), \mathbf{x}(k)) \end{aligned} \tag{7}$$

where $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, ..., \alpha_m]^T$. In matrix notation we have

$$J(\boldsymbol{\alpha}) = \boldsymbol{\alpha}^T \mathbf{K}\mathbf{K}_\tau \boldsymbol{\alpha} \tag{8}$$

where $\mathbf{K}(i, j) = \kappa(\mathbf{x}(i), \mathbf{x}(j))$ is the Gram matrix of the observations in the feature space. Note that this matrix is symmetric. The matrix $\mathbf{K}_\tau$ is constructed by circularly shifting the rows of $\mathbf{K}$, mathematically we can write

$$\mathbf{K}_\tau(t, i) = \begin{cases} \mathbf{K}(t-\tau, i), & if \ t > \tau \\ \mathbf{K}(n-\tau+t, i), & if \ t \leq \tau \end{cases} \tag{9}$$

Fixing the norm of $\boldsymbol{\alpha}$, the extremization of (8) converges to the following eigenvalue problem

$$\boldsymbol{\alpha} = eig(\mathbf{K}\mathbf{K}_\tau + \mathbf{K}_\tau^T \mathbf{K}) \tag{10}$$

where $eig(\cdot)$ returns the normalized eigenvector corresponding to the maximum eigenvalue of the matrix in its argument. Using the theory here developed we propose Algorithm 1

Algorithm 1	
1 - Input:	whiten data $\mathbf{x}(t)$, delay τ
2 - Define:	the kernel function $\kappa(\cdot,\cdot)$
3 – Do	1 - calculate $\mathbf{K}$ 2- $\boldsymbol{\alpha}$ receives the second eigenvector of $\bar{\mathbf{K}}^T \bar{\mathbf{K}}_\tau + \bar{\mathbf{K}}_\tau^T \bar{\mathbf{K}}$
4-Return	$\mathbf{y}\boldsymbol{\alpha}= \mathbf{K}^T$

There is one last point to be noted before we test the proposed algorithm. The fact that we developed our theory supposing the vectors $\boldsymbol{\xi}(t)$. Despite of the simplicity, this cannot be always guaranteed in practice. For this reason the first eigenvector of $\bar{\mathbf{K}}^T \bar{\mathbf{K}}_\tau + \bar{\mathbf{K}}_\tau^T \bar{\mathbf{K}}$ captures the dc-component of the signal [15] and our desired signal source is extracted when $\boldsymbol{\alpha}$ is its second eigenvector.

3. EXPERIMENTS

The validity of our method can be simply proofed by generating a source $\mathbf{s}(t) = [s_1(t), s_2(t), s_3(t), s_4(t)]$ with $t = 1, 2, ..., 10^3$. Taking $s_1(t)$ to be a sine function with period of 13 samples. Assuming that $s_2(t)$ and $s_3(t)$ are random Gaussian signals and $s_4(t)$ is a random signal with uniform distribution. One can prove that $\mathbf{s}(t)$ satisfies the conditions (6) for every delay τ. This way, it is expected that algorithms for BSE extracts the sine function from a

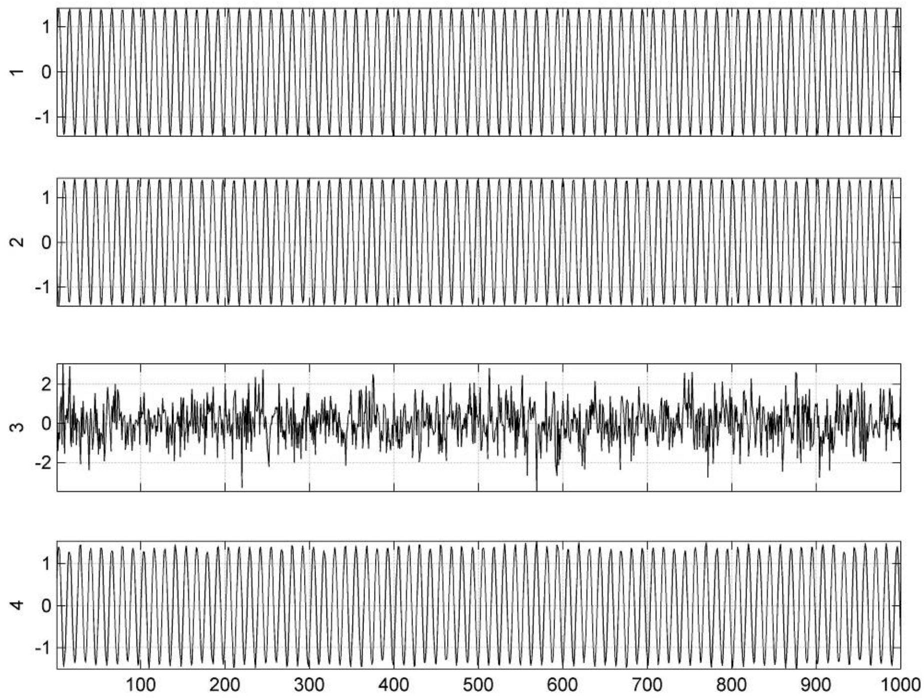

Fig. 1. Extracted signals for miscalculated $\tau = 30$ (the optimum would be $\tau = 13$). 1) Original sine-function signal. 2) Extracted signal with Algorithm 1. 3) Extracted Signal with Zhang's algorithm. 4) Extracted signal with Barros' Algorithm.

blind linear mixture, regardless to the value of τ. But better results should be considered more likely for $\tau = 13$ or multiples, since we are dealing with finite sample signals. We mixed the sources $\mathbf{s}(t)$ with a random matrix $\mathbf{A}$, giving $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$. The signal $\mathbf{x}(t)$ is used as input to Algorithm 1 with the kernel function chosen to be the radial basis function [8]

$$\kappa_\sigma(\mathbf{v},\mathbf{u}) = \exp\left[-\frac{(\mathbf{v}-\mathbf{u})^T(\mathbf{v}-\mathbf{u})}{2\sigma^2}\right] \tag{11}$$

with size $\sigma = 10$. Using the kernel (11), the nonlinear map ϕ that satisfies (1), transform the vectors $\mathbf{x}(t)$ to functions in an infinite dimensional sphere in the Hilbert space [13]. This is why we expect the proposed algorithm exploiting more deeply the temporal structures of the sources. For comparison, we also used $\mathbf{x}(t)$ as input to the algorithms proposed by Barros and Cichocki in [6] and to the algorithm of Zhang and Yi [7]. Both algorithms uses information about the autocorrelation function for extracting specific signals.

The results of that experiment were fairly favorable to our proposed algorithm. For $\tau = 13$ the three compared algorithms extracted the sine function, but when get far from that optimum value the methods in [6] and [7] extracted with distortion or failed to extract the sinusoid, while our method don't. In Fig. 1 we plot the extracted signals for $\tau = 30$. A more conclusive experiment can be performed by making τ vary in a range and analyze the error between the extracted signal and the original one in $s_1(t)$ for several trials. We varied τ from 0 to 50, mixing the sources with different randomly generated matrices, $\mathbf{A}$. For each delay τ we repeated the extraction procedure 100 times and calculated the Root-Mean-Square Error (RMSE) between the extracted signals and the original one. The averaged value of the RMSE in function of τ for each compared algorithm is shown in Fig. 2. One can see that our algorithm is the most robust with respect the choice of τ between the compared algorithms for extraction of specific signals.

4. CONCLUSION

In this work we introduced a kernel method for extraction of specific signals using a priori information about their temporal structures. We ran simulations that shown that the present method is more robust, to error in the a priori information, than the presents in the literature [6][7]. The method is based only in second order statistics of the data in the Reproducing Kernel Hilbert Space and does not need the independence assumption for separation.

We believe that the robustness of the method is due to the relation of kernel methods with information processing techniques using Parzen estimation [14]. In future works we intend to formalize this relation and test our proposed method for extraction of real word signals.

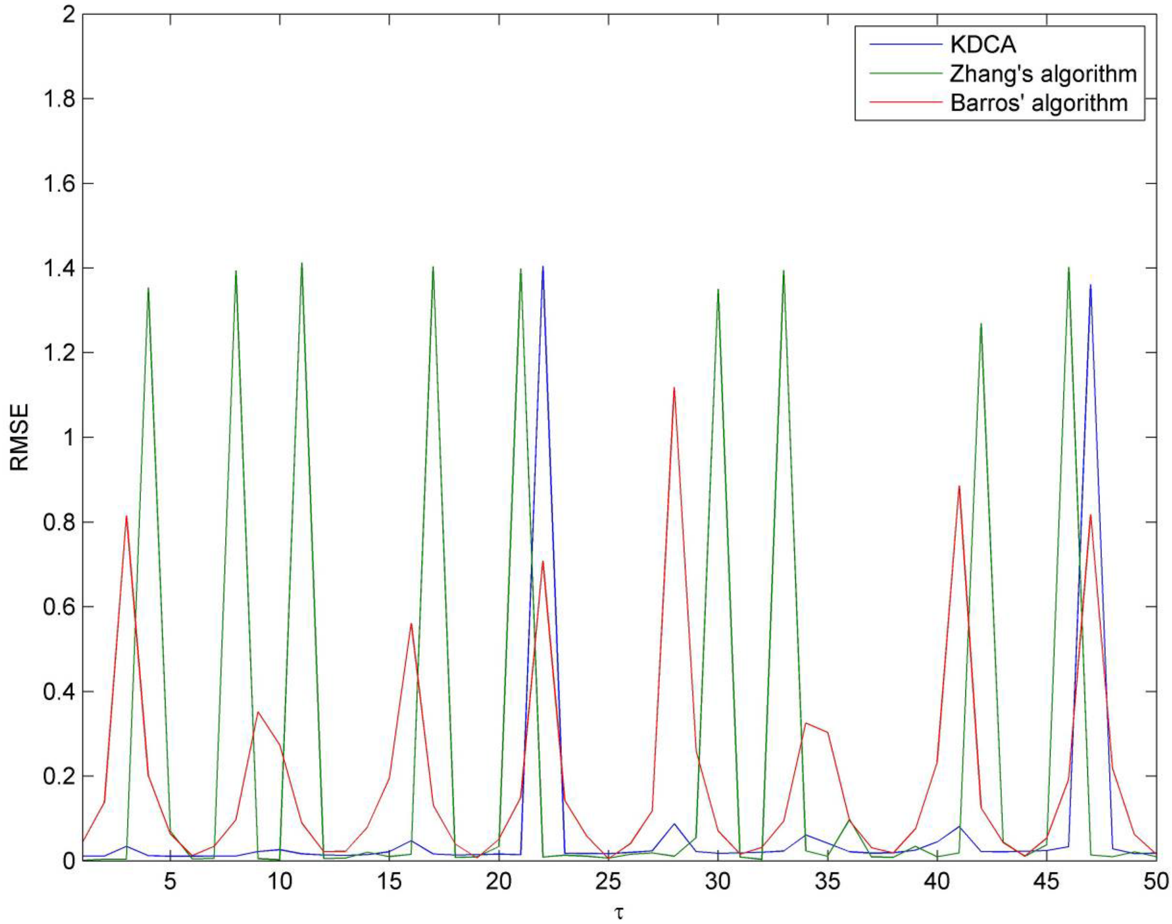

Fig. 2. RMSE between extracted and original signal in function of time delay for extraction of sine function in noise.

AKNOWLEDGMENTS

This work was partially supported by the Brazilian agency FAPEMA. The authors thank to the PIB team, especially to Enio Aguiar, for the useful discussions.

REFERENCES

[1] A. Cichocki and S.-I. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. New York: John Wiley & Sons, Inc. 2002.

[2] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. Wiley-Interscience, May 2001.

[3] W. L., D. P. Mandic, "A normalised kurtosis-based algorithm for blind source extraction from noisy measurements", *Signal Processing*, Volume 86, Issue 7, July 2006, Pages 1580-1585

[4] Cichocki, A., Barros, K. Robust batch algorithm for sequential blind extraction of noisy biomedical signals. In 1999 Proc. ISSPA'99. Australia.

[5] A. Cichocki, R. Thawonmas and S. Amari, "Sequential blind signal extraction in order specified by stochastic properties," Electronics Letters, vol.33, no. 1, pp. 64-65. Jan. 1997,

[6] A. K. Barros and A. Cichocki, "Extraction of Specific Signal with Temporal Structure," Neural Computation. Vol. 13, no. 1, Sep. 2001.

[7] Z.-L. Zhang and Z. Yi, "Robust extraction of specific signals with temporal structure", Neurocomputing, vol. 69, Issues 7-9, pp. 888-893, March 2006

[8] B. Scholkopf and A. J. Smola, Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. Cambridge: MIT Press, 2001.

[9] T. Siraya, "Reproducing Kernel Hilbert Space Methods in Measurement Science", XVII World Congress Metrology in the 3rd Millennium. Jun, 2003, Dubrovink, Croatia

[10] C. Cortes and V. Vapnik, "Support-Vector Networks", Machine Learning, vol.20, Issue 3, pp.273-297, Sep. 1995.

[11] J. C. Principe, N. R. Euliano and W. C. Lefebvre, Neural and Adaptive Systems: Fundamentals through Simulations with CD-ROM, 1st edition. John Wiley and Sons, Inc. New York, 1999.

[12] Weifeng Liu; Pokharel, P.P.; Principe, J.C., "The Kernel Least-Mean-Square Algorithm," Signal Processing, IEEE Transactions on, vol.56, no.2, pp.543-554, Feb. 2008

[13] N. Aronszajn, "Theory of Reproducing Kernels", Transactions of the American Mathematical Society, vol. 68, number 3, pages 337–404, 1950.

[14] R. Jenssen, T. Eltoft, D. Erdogmus and J. C. Principe, "Some equivalences between kernel methods and information theoretic methods," *The Journal of VLSI Signal Processing*, vol. 45, no. 1-2, pp. 49-65, November 2006.

[15] Weifeng Liu; Pokharel, P.P.; Principe, J.C., "Correntropy: Properties and Applications in Non-Gaussian Signal Processing," Signal Processing, IEEE Transactions on , vol.55, no.11, pp.5286-5298, Nov. 2007.