BLIND EXTRACTION OF NONCIRCULAR COMPLEX SIGNALS USING A WIDELY LINEAR PREDICTOR

Soroush Javidi, Beth Jelfs and Danilo P. Mandic

Department of Electrical and Electronic Engineering, Imperial College, London, UK
{soroush.javidi, beth.jelfs05, d.mandic}@imperial.ac.uk

ABSTRACT

Real valued blind source extraction based on a linear predictor is extended to the complex domain using recent advances in complex domain statistics. It is shown that, in general, the mean square prediction error of the algorithm depends both on the covariance matrix and the pseudo-covariance matrix of the source signals. To fully utilise the available information, it is thus natural to adopt a widely linear predictor to extract the latent sources from the observed mixture. This way, we derive a new algorithm for the extraction of general complex signals and provide simulation results using benchmark complex data.

1. INTRODUCTION

One of the more dynamic topics in signal processing is the blind source separation (BSS) of signals, which has diverse applications in the biomedical, radar and communications fields. In BSS, the aim is to recover latent sources from a mixture of observations, without explicit knowledge of the sources or the mixing process. This is achieved by using various assumptions, such as the independence of the sources, typically used in independent component analysis (ICA). While various algorithms are already established for processing real-valued data, complex-valued variants are only just being introduced. These include frequency domain implementations for separation of electroencephalographic (EEG) signals, the complex FastICA and its extension the noncircular complex FastICA for the separation of both circular and noncircular complex-valued signals using fixed–point updates, and negentropy maximisation [1].

Blind Source Extraction (BSE) is a class of BSS algorithms where the aim is not to recover all latent sources, but to recover a single source of interest based on various criteria [2]. For example, optimisation of the kurtosis [3] or predictability [4] of the signal may be exploited for the extraction of the desired signal. Such algorithms are suitable for real-valued signals only, and it would be desirable to extend their functionality to the complex domain.

Algorithms designed specifically for complex signals have traditionally been straightforward extensions of their real domain counterparts. However, recent studies in complex statistics have revealed several new aspects of calculus in $\mathbb{C}$. The complex *circularity* [5] assumes a rotation-invariant distribution of a complex variable, and a second-order circular variable is termed *proper*. For a complex random vector $\mathbf{x}$ with zero mean this is reflected by a vanishing pseudo-covariance matrix $E\{\mathbf{x}\mathbf{x}^T\} = \mathbf{0}$. The conventional covariance matrix is defined as $E\{\mathbf{x}\mathbf{x}^H\}$, with $(\cdot)^H$ denoting a conjugate transpose, which for a random vector with unit variance gives the indentity matrix $\mathbf{I}$.

The $\mathbb{CR}$ calculus [6], also referred to as Wirtinger calculus, allows us to calculate derivatives of complex functions even for the cases where the Cauchy-Riemann equations do not hold as is the case with most cost functions used in signal processing. This is because typically cost functions $\mathcal{J}(\mathbf{z}) : \mathbb{C}^N \mapsto \mathbb{R}$ are real-valued and the stringent analyticity conditions imposed by the Cauchy Riemann equations do now allow for their evaluation. By using $\mathbb{CR}$ calculus, on the other hand, the cost function is defined as a function of the random vector $\mathbf{z}$ and its conjugate $\mathbf{z}^*$, i.e. $\mathcal{J}(\mathbf{z}, \mathbf{z}^*) : \mathbb{C}^N \times \mathbb{C}^N \mapsto \mathbb{R}$. The derivative of such functions are taken with respect to $\mathbf{z}$ and $\mathbf{z}^*$, while keeping the other variable constant. For such $\mathbb{R}$-analytic functions [6] the calculations are greatly simplified, allowing for direct calculations in $\mathbb{C}$ rather than in the real domain.

Utilising these advances, we propose a complex BSE algorithm based on a widely linear (WL) predictor. While a linear predictor is sufficient for real-valued signals [4], in $\mathbb{C}$ it is appropriate to utilise a WL predictor model so as to make use of the full information available in the complex signal, namely that of the covariance and pseudo-covariance matrices. This will provide the basis for the minimisation of the prediction error and extraction of the desired source signal. Widely linear prediction filters were previously introduced, such as the augmented complex least means square (ACLMS) algorithm [7] which was shown to provide a comparably better prediction of real world complex data, and the widely linear IIR algorithm for filtering of noncircular signals [8].

In the next section, we will introduce the complex BSE problem using a WL predictor, and derive the update equations for its adaptive operation. This is followed by simula-

978-1-4244-2710-9/09/$25.00 © 2009 IEEE

tions using synthesised complex signals and the analysis of the results.

2. THE COMPLEX BSE PROBLEM

2.1. Problem Setup

An observation $\mathbf{x}(k) \in \mathbb{C}^N$ at time index k, can be modelled as

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

where $\mathbf{A} \in \mathbb{C}^{N\times N}$ is a mixing matrix and $\mathbf{s}(k) \in \mathbb{C}^N$ is the vector of latent sources. It is assumed that the sources are stationary and spatially uncorrelated, that is, the covariance and pseudo-covariance matrices are diagonal matrices [9], and no assumptions are made regarding the circularity of the source. This can be written as

$$\begin{aligned} \mathcal{C}_{\mathbf{ss}}(\delta k) &= E\{\mathbf{s}(k)\mathbf{s}^H(k-\delta k)\} \quad (2)\\ &= diag(\sigma_N^2(\delta k), \ldots, \sigma_N^2(\delta k)) \\ \mathcal{P}_{\mathbf{ss}}(\delta k) &= E\{\mathbf{s}(k)\mathbf{s}^T(k-\delta k)\} \quad (3)\\ &= diag(\tau_N(\delta k), \ldots, \tau_N(\delta k)) \end{aligned}$$

where $\mathcal{C}_{\mathbf{ss}}$ and $\mathcal{P}_{\mathbf{ss}}$ denote respectively the covariance and pseudo-covariance matrix, and δk is a lag. As shown in Figure 1, the extracted signal $y(k)$ is formed by applying the demixing vector $\mathbf{w}$ to $\mathbf{x}(k)$, i.e.

$$y(k) = \mathbf{w}^H\mathbf{x}(k) \tag{4}$$

The aim is to find $\mathbf{w}$ such that $\mathbf{w}^H\mathbf{A} = [0, \ldots, e^{\jmath\varphi}, 0, \ldots, 0]$ extracts only a single source, based on a desired criterion. The predictability can be used as a means to extract signals by minimising the mean square prediction error (MSPE) [2] and by using this information to adapt the values of $\mathbf{w}$. As the sources are general complex signals, it is only correct to write the filter output as a WL process

$$y_{WL}(k) = \mathbf{h}^T(k)\mathbf{y}(k) + \mathbf{g}^T(k)\mathbf{y}^*(k) \tag{5}$$

where $\mathbf{h}(k)$ and $\mathbf{g}(k)$ are the coefficient vectors of a filter with length M, $\mathbf{y}(k) = [y(k-1), \ldots, y(k-M)]^T$ are delayed versions of the extracted signal and $(\cdot)^*$ denotes the complex conjugate operator. Note that while it is possible to employ fixed values for the coefficient vectors, we have chosen to update $\mathbf{h}$ and $\mathbf{g}$ adaptively, which will provide the largest relative difference in the MSPE values [4].

From Figure 1 and (5) the prediction error is

$$e(k) = y(k) - y_{WL}(k) \tag{6}$$

and the MSPE can be calculated as

$$\begin{aligned} E\{|e(k)|^2\} &= E\{e(k)e^*(k)\} \quad (7)\\ &= \Re\{\mathbf{w}^H\mathbf{A}\hat{\mathcal{C}}_{\mathbf{ss}}\mathbf{A}^H\mathbf{w} - \mathbf{w}^H\mathbf{A}\hat{\mathcal{P}}_{\mathbf{ss}}\mathbf{A}^T\mathbf{w}^*\} \end{aligned}$$

where

$$\begin{aligned} \hat{\mathcal{C}}_{\mathbf{ss}} &= \mathcal{C}_{\mathbf{ss}}(0) - 2\sum_{m=1}^{M} h_m^*(k)\mathcal{C}_{\mathbf{ss}}(m) \quad (8)\\ &+ 2\sum_{m,l=1}^{M}[h_m(k)h_l^*(k) + g_m^*(k)g_l(k)]\mathcal{C}_{\mathbf{ss}}(l-m) \end{aligned}$$

$$\hat{\mathcal{P}}_{\mathbf{ss}} = -2\sum_{m=1}^{M} g_m^*(k)\mathcal{P}_{\mathbf{ss}}(m) + 2\sum_{m,l=1}^{M} h_m(k)g_l(k)\mathcal{P}_{\mathbf{ss}}(l-m) \tag{9}$$

Due to lack of space, the complete derivation is not included, however the above result can be obtained by expansion of the MSPE terms and using equations (1) and (4). Observe that:-

- The MSPE of complex signals is a function of the covariance and pseudo-covariance matrices of the latent complex sources. In the particular case of circular sources, $\hat{\mathcal{P}}_{\mathbf{ss}} = \mathbf{0}$ and the MSPE is dependent only on the source covariance matrix.
- Given that the sources are uncorrelated, the terms $\hat{\mathcal{C}}_{\mathbf{ss}}$ and $\hat{\mathcal{P}}_{\mathbf{ss}}$ are diagonal matrices, where the prediction error $e_n(k)$ pertaining to a particular source $s_n(k)$ can be represented using the corresponding diagonal elements.

These results can be seen as a generalisation of the real domain equivalent given in [4].

2.2. Cost Function

To circumvent the ambiguity associated with the power levels of the source signals an algorithm based on the normalised MSPE was proposed in [4]. In this manner, the normalised MSPE does not change with variation in the source signal magnitude, and we can use this concept to define the cost function

$$\mathcal{J}(\mathbf{w}, \mathbf{w}^*) = \frac{E\{|e(k)|^2\}}{E\{|y(k)|^2\}} \tag{10}$$

where $\mathcal{J} \in \mathbb{R}$ is given by its conjugate (augmented) coordinates $\mathbf{w}$ and $\mathbf{w}^*$. The BSE optimisation problem is then

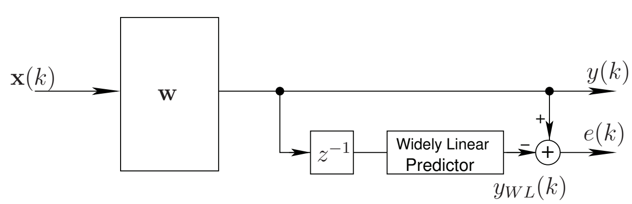

Fig. 1. The complex BSE algorithm using a widely linear predictor

stated as

$$\mathbf{w}_{opt} = \arg \max_{||\mathbf{w}||_2^2=1} \frac{E\{|e(k)|^2\}}{E\{|y(k)|^2\}} \quad (11)$$

where the vector $\mathbf{w}_{opt}$ has a single non-zero value corresponding to the source with the smallest normalised MSPE. The value of $\mathbf{w}$ is thus updated by an online gradient descent equation

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \mu\nabla_{\mathbf{w}^*}\mathcal{J}, \quad \mathbf{w}(k+1) \leftarrow \frac{\mathbf{w}(k+1)}{||\mathbf{w}(k+1)||_2^2} \quad (12)$$

where μ is a step-size and the direction of steepest descent is in the direction of the conjugate vector $\mathbf{w}^*$. The gradient $\nabla_{\mathbf{w}^*}\mathcal{J}$ is derived using $\mathbb{CR}$ Calculus as

$$\nabla\mathcal{J} = \left[e^*(k)\hat{\mathbf{x}}_h(k) - e(k)\hat{\mathbf{x}}_g(k) - \frac{\sigma_e^2(k)}{\sigma_y^2(k)}y^*(k)\mathbf{x}(k)\right]\frac{1}{\sigma_y^2(k)} \quad (13)$$

where

$$\hat{\mathbf{x}}_h(k) \triangleq \mathbf{x}(k) - \sum_{m=1}^{M} h_m(k)\mathbf{x}(k-m) \quad (14)$$

$$\hat{\mathbf{x}}_g(k) \triangleq -\sum_{m=1}^{M} g_m^*(k)\mathbf{x}(k-m)$$

The MSPE, represented as $\sigma_e^2(k)$ and the variance of the extracted signal $\sigma_y^2(k)$ are estimated recursively using a convex combination of the instantaneous power and the previous instance, where β_e and β_y are the corresponding forgetting factors (mixing parameters) of the error and extracted signal power [4], that is

$$\sigma_e^2(k) = \beta_e\sigma_e^2(k-1) + (1-\beta_e)e^2(k) \quad (15)$$

$$\sigma_y^2(k) = \beta_y\sigma_y^2(k-1) + (1-\beta_y)y^2(k)$$

Note that the gradient (13) can easily simplify to the use with a standard complex predictor by assuming the coefficient vector $\mathbf{g}(k) = \mathbf{0}$. In this manner, the complexity of the algorithm is reduced when the input sources are known to be complex circular.

3. PERFORMANCE OF THE COMPLEX BSE

The performance of the complex BSE algorithm using a WL predictor was analysed through simulations using sources with different distributions and degrees of circularity. Figure 2 shows the complex sources generated for these simulations. Signal $s_1(k)$ was a noncircular Gaussian source with a high degree of noncircularity. Source $s_2(k)$ was a circular super-Gaussian distribution with the smallest MSPE. Source $s_3(k)$ was a noncircular sub-Gaussian BPSK signal and $s_4(k)$ was a circular 4-QAM source. These properties

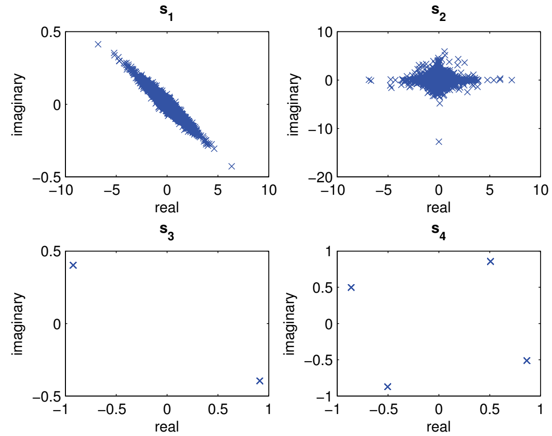

Fig. 2. The complex source signals used in simulations

are summarised in Table 1 which describes the source distribution, degree of circularity and MSPEs.

The number of sources was $N = 4$ with 5000 samples, which were mixed using a complex mixing matrix to form the observation vector. The demixing vector $\mathbf{w}$ was then adapted using the steepest descent update (12) with the gradient vector given in (13). Learning curves were obtained by averaging 100 trials of independent simulations. An adaptive predictor with $M = 20$ taps for each of the coefficient vectors $h(k)$ and $g(k)$ was used with the forgetting factors $\beta_e = \beta_y = 0.025$ and step-size $\mu = 0.0025$. The coefficient vectors were initially randomly generated and then updated using an ACLMS algorithm with step-size $\mu_h = \mu_g = 0.001$.

The performance index (PI) [2] was used to assess the quality of the extraction process. Defining $\mathbf{c} = \mathbf{w}^T\mathbf{A} = [c_1, \ldots, c_m]$, the PI is given as

$$PI = 10\log_{10}\left(\frac{1}{M}\left(\sum_{i=1}^{M}\frac{c_i^2}{\max\{c_1^2,\ldots,c_M^2\}}\right)\right). \quad (16)$$

The absolute value of the sources $|s_n(k)|$, $n = 1, \ldots, 4$ can be seen in Figure 3. For comparison, the absolute value of the extracted source are superimposed on the graphs. It can be seen that in the first iteration, the algorithm was able to extract the second source $s_2(k)$. This can also be seen from Figure 4, where the PI learning curve converge after 500 iterations reaching an average value of $-19\ dB$ in steady-state.

For comparison, a linear prediction filter (i.e. $\mathbf{g}(k) = \mathbf{0}$) was used with the same data to minimise the normalised MSPE cost function. The PI is shown in Figure 4, which convergence after 4000 iterations and reached an average value of $-13.5\ dB$, and the source $s_2(k)$ was not extracted successfully.

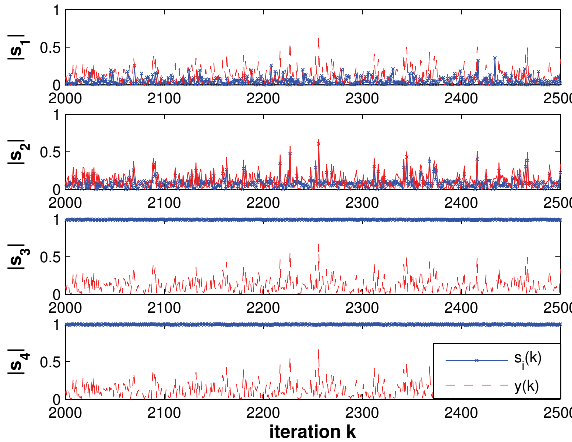

Fig. 3. Absolute values of the sources $s_n(k)$ (*solid line with '$\times$' marks*) and extracted source (*broken line*). Note adequate extraction of source $s_2(k)$. A widely linear algorithm was used to train the prediction filter.

Table 1. The complex sources used in simulations. The degree of noncircularity $\eta = \sqrt{\sigma^2_{\mathbf{z}_r}/\sigma^2_{\mathbf{z}_i}}$, $\eta > 1$

Source	distribution	deg. of circ. (η)	MSPE
$s_1(k)$	Gaussian	214.13	221.90
$s_2(k)$	Super-Gaussian	1.04	**4.47**
$s_3(k)$	Sub-Gaussian	5.30	83.69
$s_4(k)$	Sub-Gaussian	1.00	73.57

4. FUTURE WORK AND CONCLUSIONS

A new method for the blind extraction of general complex valued sources has been introduced. This has been achieved based on the mean square prediction error of a widely linear adaptive predictor. For rigour, the MSPE of general complex sources has been derived and has been shown to be a function of both the covariance and pseudo-covariance matrices, thus justifying the use of a widely linear predictor. The performance of the proposed algorithm has been illustrated on the extraction of general complex signals.

Suitable extensions of this work include the analysis of the effect of noise on the extraction process, specifically on the widely linear model. It is also of interest to extend the BSE algorithm using alternative cost functions such as maximisation of kurtosis or the generalised kurtosis. Finally, the performance of nonlinear adaptive filters using split-complex and fully-complex nonlinearities may also be analysed for the extraction of general complex sources.

5. REFERENCES

[1] M. Novey and T. Adalı, "Complex ICA by negentropy maximization," *IEEE Transactions on Neural Networks*, vol. 19, no. 4, pp. 596–609, 2008.

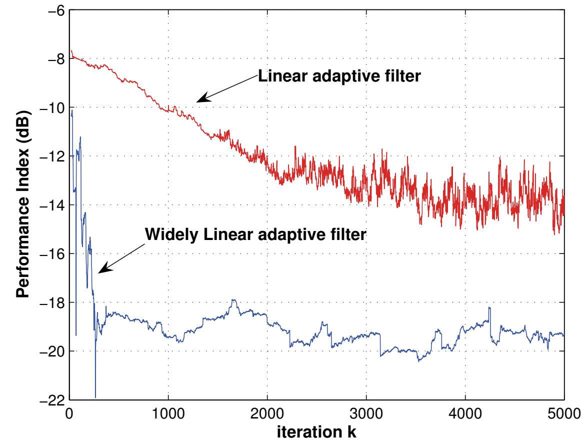

Fig. 4. Performance index (in dB) using widely linear (bottom graph) and linear (top graph) adaptive filters.

[2] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing, Learning Algorithms and Applications*, Wiley, 2002.

[3] W.Y. Leong, W. Liu, and D.P. Mandic, "Blind source extraction: Standard approaches and extensions to noisy and post-nonlinear mixing," *Neurocomputing*, vol. 71, pp. 2344 – 2355, 2008.

[4] W. Liu, D.P. Mandic, and A. Cichocki, "Blind source extraction based on a linear predictor," *IET Signal Processing*, vol. 1, no. 1, pp. 29–34, March 2007.

[5] B. Picinbono, "On circularity," *IEEE Transactions on Signal Processing*, vol. 42, no. 12, pp. 3473–3482, 1994.

[6] K. Kreutz-Delgado, "The complex gradient operator and the $\mathbb{CR}$-calculus," *Dept. of Electrical and Computer Engineering, UC San Diego, Course Lecture Supplement No. ECE275A*, pp. 1–74, 2006.

[7] S. Javidi, M. Pedzisz, S.L. Goh, and D.P. Mandic, "The augmented complex least mean square algorithm with application to adaptive prediction problems," in *Proc. 1st IARP Workshop on Cognitive Information Processing*, 2008, pp. 54–57.

[8] C. Cheong Took and D. Mandic, "Adaptive IIR filtering of noncircular complex signals," *IEEE Transactions of Signal Processing*, 2009 (accepted).

[9] J. Eriksson and V. Koivunen, "Complex random vectors and ICA models: Identifiability, uniqueness, and separability," *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 1017–1029, 2006.