Pages 1–18

Kurtosis based blind source extraction of complex noncircular signals with application in EEG artifact removal in real-time

S. Javidi [1], D. P. Mandic [1], C. Cheong Took [1], and A. Cichocki [2]

[1]Communications and Signal Processing Research Group, Department of Electrical and Electronic Engineering, Imperial College London, London, SW7 2AZ, U.K.
[2]Laboratory for Advanced Brain Signal Processing, Brain Science Institute, RIKEN, Saitama 351-0198, Japan

ABSTRACT

A new class of complex domain blind source extraction (BSE) algorithms suitable for the extraction of both circular and noncircular complex signals is proposed. This is achieved through sequential extraction based on the degree of kurtosis and in the presence of noncircular measurement noise. The existence and uniqueness analysis of the solution is followed by a study of fast converging variants of the algorithm. The performance is first assessed through simulations on well understood benchmark signals, followed by a case study on real-time artifact removal from EEG signals, verified using both qualitative and quantitative metrics. The results illustrate the power of the proposed approach in real-time blind extraction of general complex-valued sources.

Keywords: Blind source extraction, complex noncircularity, noisy mixtures, noncircular noise, complex kurtosis, EEG artifact removal

1 INTRODUCTION

The aim of blind source separation (BSS) is to reconstruct the original sources by identifying the inverse of the mixing system, without having explicit knowledge of the mixing parameters or sources (Cichocki and Amari, 2002), and has found application in diverse areas including biomedical engineering, communications, sonar and radar (Cichocki and Amari, 2002; Anemüller et al., 2003). Standard BSS methods use cost functions based on second- and higher-order statistics, together with the maximisation of likelihood and entropy (Amari et al., 1997; Bell and Sejnowski, 1995; Hyvärinen et al., 2001). In addition, to facilitate the modelling of real-world systems, noisy environments and post-nonlinear mixtures have been recently studied in real domain algorithms (Leong and Mandic, 2008; Särelä and Valpola, 2005).

Within the BSS methodology, the latent sources are separated in a random order through either a deflationary or symmetric orthogonalisation procedure, that is, one by one or simultaneously. A class of BSS algorithms, termed blind source extraction (BSE), aims to retrieve the sources one by one, based on a fundamental signal property (nonlinearity, sparsity), effectively inducing an order into the separation process. The benefit of BSE becomes apparent in large-scale problems where only a small subset of the sources are of interest, making it possible to extract such sources at a dramatically reduced computational complexity and in real-time. This also relaxes the requirement for pre- or post-processing of the mixture or separated sources, that may be necessary if parallel BSS techniques were employed.

Real domain algorithms performing BSE based on the temporal structure (predictability) of signals are well established (Barros and Cichocki, 2001; Cichocki and Amari, 2002; Mandic and Cichocki, 2003) and modifications to the cost function were proposed to cater for noisy mixtures (Liu et al., 2006b,a). The algorithm in (Cichocki et al., 1997) demonstrated the feasibility of extraction of real-valued sources based on the degree of kurtosis, while (Liu and Mandic, 2006) proposed a modified cost function for the extraction in noisy environments. An overview and discussion on this class of algorithms is also provided in (Leong et al., 2008).

Recent developments in complex statistics (Picinbono, 1994; Neeser and Massey, 1993) have made it possible to introduce a new class of complex domain signal processing algorithms, capable of catering for the generality of complex signals (Mandic and Goh, 2009). This is achieved through the consideration of the circular symmetry of the probability distributions, whereby rotational invariance of the distribution indicates a complex *circular* random variable. However, most complex-valued signals encountered in signal processing application are *noncircular*.

The so called augmented complex statistics (Schreier and Scharf, 2003), enables us to utilise the complete second-order information available in a complex-valued random variable. This way, the second-order statistics are not only based the standard covariance matrix $E\{\mathbf{x}\mathbf{x}^H\}$, but also the pseudo-covariance $E\{\mathbf{x}\mathbf{x}^T\}$. A complex-valued random vector with a vanishing pseudo-covariance is termed *proper* or second-order circular, and is otherwise called *improper* (Picinbono and Bondon, 1997; Schreier and Scharf, 2003). Likewise, widely linear models (Picinbono and Chevalier, 1995) allow for the design of linear mean square error estimation algorithms capable of processing both complex proper and improper signals.

In complex-valued blind source separation research, recent complex-valued algorithms typically use augmented statistics, so as to cater for the generality of complex signals (Erdogan, 2009; Douglas, 2005), with applications in fMRI modelling (Novey and Adalı, 2008) and communications (Ollila and Koivunen, 2009b). In comparison to standard complex BSS methodology (Anemüller

et al., 2003; Bingham and Hyvärinen, 2000), which assumes complex circular sources, these algorithms have been shown to exhibit enhanced performance for noncircular sources and similar performance for circular sources. In the same spirit, the feasibility of blind source extraction of complex sources based on the temporal structure of the latent sources was studied in (Javidi et al., 2009), exploring a widely linear predictor to extract both proper and improper sources. A class of linear predictability based algorithms for blind extraction from noisy complex-valued mixtures has also been recently proposed (Javidi et al., 2010).

In this paper, we introduce an online blind source extraction algorithm suitable for the generality of complex-valued signals, both circular and noncircular. This is achieved based on higher order statistics of latent sources, and using the deflation approach. Further, the cost function based on an extension of the methodology presented in (Liu and Mandic, 2006) is designed so as to be robust to both circular and noncircular second order additive noise. The analysis is supported by benchmark simulations in both noise-free and noisy scenarios, followed by studies of conditioning of EEG signals for the automatic removal of biological and power line artifacts.

The paper is organised as follows. Section 2 provides an overview of complex statistics, complex-valued noise and $\mathbb{CR}$ calculus. The cost function for both noise-free and noisy cases is then presented, together with the derivation and convergence analysis of a real-time adaptive BSE algorithm. In Section 3, after analysing the performance in blind extraction of synthetic sources, EEG signal conditioning for brain computer interfacing is studied. Conclusions are presented in Section 5.

2 MODELS AND METHODS

2.1 Complex statistics: second-order circularity

Second-order circularity is a property of probability density functions, whereby the distribution of a complex random variable $\mathbf{z}$ and its rotation $e^{\jmath\varphi}\mathbf{z}$ are equal for any angle φ (Picinbono, 1994). Within the domain of second-order statistics, to account for complex random variables with noncircular probability density functions (pdf), we need to use both the covariance $\mathcal{C}_{\mathbf{zz}}$ and pseudo-covariance $\mathcal{P}_{\mathbf{zz}}$ matrices (Picinbono and Bondon, 1997)

$$\mathcal{C}_{\mathbf{zz}} = E\{\mathbf{z}\mathbf{z}^H\}, \quad \mathcal{P}_{\mathbf{zz}} = E\{\mathbf{z}\mathbf{z}^T\}. \tag{1}$$

For second-order circular (proper) random variables, the pseudo-covariance matrix vanishes, that is $\mathcal{P}_{\mathbf{zz}} = \mathbf{0}$, whereas for second-order noncircular (improper) random variables the pseudo-covariance matrix is non-zero, $\mathcal{P}_{\mathbf{zz}} \neq \mathbf{0}$, and is generally complex-valued. The pseudo-covariance matrix $\mathcal{P}_{\mathbf{zz}}$ can be written in terms of the covariances of its real and imaginary components

$$\mathcal{P}_{\mathbf{zz}} = E\{\mathbf{z}_r\mathbf{z}_r^T\} - E\{\mathbf{z}_i\mathbf{z}_i^T\} + \jmath\big(E\{\mathbf{z}_i\mathbf{z}_r^T\} + E\{\mathbf{z}_r\mathbf{z}_i^T\}\big)$$

illustrating that for proper signals, the vanishing pseudo-covariance is due to equal powers in the real and imaginary channels, while the cross-covariance is skew-symmetric (Neeser and Massey, 1993).

For an uncorrelated random vector, both the covariance and pseudo-covariance matrices are diagonal (Eriksson and Koivunen, 2006). Examples of complex circular signals in signal processing research are QPSK and BPSK signals in communications, while most complex signals made complex by convenience of representation are noncircular. Examples include EEG signals, and complex-valued wind models (Mandic and Goh, 2009).

Consider a second-order stationary 'augmented' complex random signal $\mathbf{z}^a(k) = [\mathbf{z}^T(k), \mathbf{z}^H(k)]^T$ and its augmented covariance matrix,

$$\begin{aligned}\mathcal{C}_{\mathbf{z}^a\mathbf{z}^a}(\delta) &= E\left[\begin{array}{c}\mathbf{z}(k)\\ \mathbf{z}^*(k)\end{array}\right]\left[\mathbf{z}^H(k-\delta),\, \mathbf{z}^T(k-\delta)\right]\\ &= \left[\begin{array}{cc}\mathcal{C}_{\mathbf{zz}}(\delta) & \mathcal{P}_{\mathbf{zz}}(\delta)\\ \mathcal{P}^*_{\mathbf{zz}}(\delta) & \mathcal{C}^*_{\mathbf{zz}}(\delta)\end{array}\right].\end{aligned} \tag{2}$$

This matrix provides a complete description of the second-order statistics of $\mathbf{z}(k)$. The transformation of this matrix to the frequency domain gives the augmented spectral matrix (Picinbono and Bondon, 1997; Schreier and Scharf, 2003)

$$\mathcal{S}_{\mathbf{z}^a}(\omega) = \left[\begin{array}{cc}\mathcal{S}_{\mathbf{z}}(\omega) & \breve{\mathcal{S}}_{\mathbf{z}}(\omega)\\ \breve{\mathcal{S}}^*_{\mathbf{z}}(-\omega) & \mathcal{S}_{\mathbf{z}}(-\omega)\end{array}\right], \tag{3}$$

with the Fourier transforms of the covariance and pseudo-covariance matrices defined respectively as $\mathcal{S}_{\mathbf{z}}(\omega)$ and $\breve{\mathcal{S}}_{\mathbf{z}}(\omega)$, that is

$$\begin{aligned}\mathcal{S}_{\mathbf{z}}(\omega) &= \mathcal{F}\big(\mathcal{C}_{\mathbf{zz}}(\delta)\big) = \mathcal{F}\big(E\{\mathbf{z}(k)\mathbf{z}^H(k-\delta)\}\big)\\ \breve{\mathcal{S}}_{\mathbf{z}}(\omega) &= \mathcal{F}\big(\mathcal{P}_{\mathbf{zz}}(\delta)\big) = \mathcal{F}\big(E\{\mathbf{z}(k)\mathbf{z}^T(k-\delta)\}\big)\end{aligned} \tag{4}$$

where the symbol δ denotes a discrete time lag and $\mathcal{F}(\cdot)$ the Fourier transform operator.

While the power spectrum provides information on the distribution of signal power over a frequency range, the magnitude of the pseudo-spectrum characterises the second-order circularity of the random variable in the frequency domain. The augmented spectral matrix in (3) is positive semidefinite which results in the condition (Picinbono and Bondon, 1997)

$$|\breve{\mathcal{S}}_{\mathbf{z}}(\omega)|^2 \leq \mathcal{S}_{\mathbf{z}}(\omega)\cdot\mathcal{S}_{\mathbf{z}}(-\omega). \tag{5}$$

2.2 Complex statistics: Kurtosis

Kurtosis has been used routinely to design contrast functions in BSS (Hyvärinen and Oja, 1997) and BSE algorithms (Cichocki et al., 1997). It is common to use the normalised kurtosis $K_R(\cdot)$ instead of the standard kurtosis $\mathrm{kurt}_R(\cdot)$, as it allows for the comparison of the degree of non-Gaussianity of random variables, irrespective of the range of amplitudes. In (Ollila and Koivunen, 2009a), the relevance of this concept in the complex domain, together with as the relation between the kurtosis of the real and imaginary components of a complex random variable, $\mathrm{kurt}_R(z_r)$ and $\mathrm{kurt}_R(z_i)$ and the kurtosis of the complex random variable $\mathrm{kurt}_c(z)$ has been discussed.

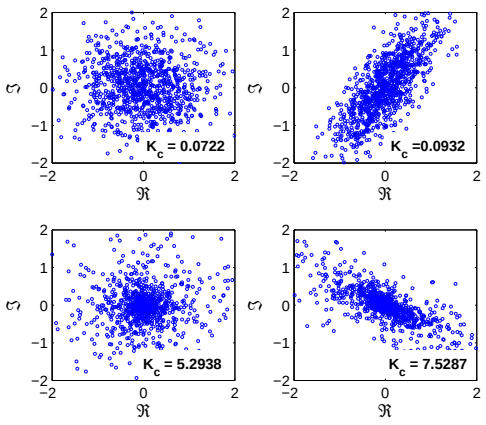
The normalised kurtosis of a complex random variable (real-valued) can be defined as

$$\begin{aligned}K_c(z) &= \frac{\mathrm{kurt}_c(z)}{(E\{|z|^2\})^2}\\ &= \frac{E\{|z|^4\}}{(E\{|z|^2\})^2} - \frac{|E\{z^2\}|^2}{(E\{|z|^2\})^2} - 2\end{aligned} \tag{6}$$

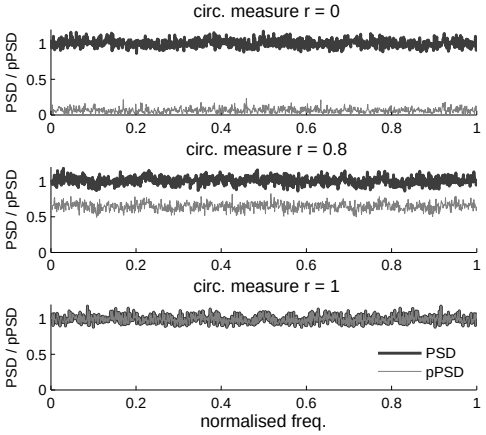
with

$$\mathrm{kurt}_c(z) = E\{|z|^4\} - |E\{z^2\}|^2 - 2(E\{|z|^2\})^2. \tag{7}$$

The first term in (6) is the normalised fourth order moment, the second term is the square of the circularity coefficient (Ollila, 2008),



(a) Scatter plots of complex white noise realisations. *Top row:* circular Gaussian noise (left) and noncircular Gaussian noise ($r = 0.81$) (right). *Bottom row:* circular Laplacian noise (left) and noncircular Laplacian noise ($r = 0.81$) (right). The circularity measure r is defined in (8). The kurtosis values K_c are given for each case.



(b) Power spectra (thick gray line) and pseudo-power spectra (thin gray line) of complex Gaussian noises with varying degrees of noncircularity $r = \{0, 0.8, 1\}$

Fig. 1. Illustration of white circular and doubly white complex-valued noises.

whereas $\mathrm{kurt}_c(z)$ in (7) is the real-valued kurtosis of the complex random variable z. Similar to the kurtosis of a real-valued Gaussian random variable, the value of K_c is zero for both circular and noncircular complex Gaussian random variables. Furthermore, for continuity, this measure makes kurtosis values of a sub-Gaussian complex random variable negative and that of a super-Gaussian complex random variable positive, irrespective of the degree of circularity/noncircularity.

2.3 Complex-valued Noise

The degree of noncircularity can be quantified by the circularity measure r, defined in (Ollila, 2008) as the magnitude of the circularity quotient $\rho(z) = re^{\jmath\theta} \triangleq \tau_z^2/\sigma_z^2$, where

$$r = |\rho(z)| = \frac{|\tau_z^2|}{\sigma_z^2}, \quad r \in [0, 1] \tag{8}$$

measures the degree of noncircularity in the complex signal, with τ_z^2 the pseudo-variance of the signal and the circularity angle $\theta = \arg(\rho(z))$ indicating orientation of the distribution. Note that for a purely circular signal, $r = 0$, with θ not providing additional information about the distribution.

This circularity measure can also be graphically interpreted using an ellipse (centred in the complex plane) of eccentricity ϵ and orientation α, such that $r = \epsilon^2$ and $\theta = 2\alpha$ (Ollila, 2008, Theorem1). For $\epsilon = 0$, the shape becomes a circle, which also indicates a circular signal with $r = 0$, while for the extreme case of $\epsilon = 1$, corresponding to a highly noncircular signal with $r = 1$, the ellipse becomes elongated with a maximal major axis and minor axis of length zero. Note that the pseudo-variance of a general complex Gaussian distribution is then related to the elliptic shape by $\tau^2 = \epsilon^2 e^{\jmath 2\theta}$ (Ollila and Koivunen, 2009b).

It is important to notice that the treatment of a noise vector $\mathbf{v}(k)$ in $\mathbb{C}$ is different to that in the real domain (Picinbono and Bondon, 1997). While in $\mathbb{R}$ only the variance σ_v^2 of the noise signal is of concern, in $\mathbb{C}$ it is necessary to also consider the pseudo-variance τ_v^2, in order to completely model the noise. We therefore differentiate between the following cases of white noise.

1. *Circular white noise*, is considered white in terms of its diagonal covariance matrix, whereas the pseudo-covariance matrix vanishes, that is

$$\mathcal{C}_{\mathbf{vv}}(\delta) = \sigma_v^2 \mathbf{I}, \quad \mathcal{P}_{\mathbf{vv}}(\delta) = \mathbf{0}, \quad \delta = \mathbf{0}$$

   where $\mathbf{I}$ denotes the identity matrix.
   In the frequency domain, the covariance spectrum $\mathcal{S}_{\mathbf{v}}(\omega)$ (also power spectrum, or PSD) of the circular white noise is flat, while the pseudo-covariance spectrum $\breve{\mathcal{S}}_{\mathbf{v}}(\omega)$ (or pPSD) is zero.

2. *Noncircular doubly white noise*, is assumed white for both the real and imaginary components, however, the corresponding distributions and power levels may be different, such that

$$\mathcal{C}_{\mathbf{vv}}(\delta) = \sigma_v^2 \mathbf{I}, \quad \mathcal{P}_{\mathbf{vv}}(\delta) = \tau_v^2 \mathbf{I}, \quad \delta = 0,\ \sigma_v^2 \neq \tau_v^2.$$

   In this case, the power spectrum is flat across all frequencies, while the pseudo-spectrum is non-zero. As the noise becomes more noncircular ($r \to 1$), the pseudo-spectrum approaches its upper-bound defined in (5), where for highly noncircular noise ($r \approx 1$), the magnitudes of the pPSD and PSD are similar. For a scalar complex white noise signal $v(k)$, the relations between the correlation and pseudo-correlation and the respective spectra are given by

$$\mathcal{C}(\delta) = E\{v(k)v^*(k-\delta)\} = \delta(0)\sigma_v^2 \ \xrightarrow{\mathcal{F}}\ \mathcal{S}(\omega) = |\sigma_v^2|$$
$$\mathcal{P}(\delta) = E\{v(k)v(k-\delta)\} = \delta(0)\tau_v^2 \ \xrightarrow{\mathcal{F}}\ \breve{\mathcal{S}}(\omega) = |\tau_v^2|.$$

Examples of circular white Gaussian and Laplacian noise with unit variance are illustrated in the left hand column of Fig. 1(a), whereas

the right hand column demonstrates two examples of noncircular white noise, with the top-right plot showing a noncircular Gaussian noise signal with circularity measure $r = 0.81$ with unit variance and pseudo-variance $\tau_v^2 = -0.38 + \jmath 0.71$, and the bottom-right plot illustrating the scatter plot of noncircular Laplacian noise with circularity measure $r = 0.81$ with unit variance and pseudo-variance of $0.45 - \jmath 0.66$. Also note that in Fig. 1(a) the value of the kurtosis is approximately zero for both the circular and noncircular Gaussian noise signals, whereas the kurtosis values for the circular and noncircular super-Gaussian noise signals follow the real-valued convention.

Fig. 1(b) depicts the PSD and pPSD of circular ($r = 0$) white and noncircular doubly white Gaussian noise for the respective circularity measures $r = \{0.8, 1\}$. Observe that the pseudo-spectrum is zero for the circular noise, while it has a magnitude of 0.64 for the noise with $r = 0.8$, and reaches it upper-bound of 1 in the third realisation where the noise is highly noncircular ($r = 1$). For the Gaussian noise, the spectrum $\mathcal{S}(\omega) = 1$ and the pseudo-spectrum $\breve{\mathcal{S}}(\omega) = |\tau_v^2| = |\epsilon^2 e^{\jmath 2\theta}| = |\epsilon^2| = r$, across all frequencies, thus indicating that by increasing the eccentricity of the ellipse (degree of noncircularity), the magnitude of the pPSD approaches its maximum value of 1.

2.4 $\mathbb{CR}$ calculus: Brief overview

The $\mathbb{CR}$ calculus[1] (Kreutz-Delgado, 2006) allows for the analysis of functions that do not normally satisfy the stringent Cauchy-Riemann conditions of analyticity, such as real-valued cost functions of complex variables. Consider a typical cost function $F(\mathbf{z}) : \mathbb{C}^N \mapsto \mathbb{R}$, a real function of complex variables, which does not satisfy the Cauchy-Riemann properties, required for gradient calculations. However, using the $\mathbb{CR}$ calculus framework, it is possible to calculate the gradients of such functions directly in $\mathbb{C}$, and without the need to obtain derivatives of the real and imaginary components separately.

In the framework of $\mathbb{CR}$ calculus, F is taken as a function of the complex input vector $\mathbf{z}$ and its conjugate $\mathbf{z}^*$, collectively termed the conjugate coordinates, that is

$$F(\mathbf{z}, \mathbf{z}^*) : \mathbb{C}^N \times \mathbb{C}^N \mapsto \mathbb{R}. \tag{9}$$

Note that although $\mathbf{z}$ and $\mathbf{z}^*$ are not statistically independent, this does not affect the calculation of derivatives, defined as (Kreutz-Delgado, 2006)

$$\begin{aligned} \mathbb{R}\text{–derivative}: \quad & \frac{\partial F}{\partial \mathbf{z}}\bigg|_{\mathbf{z}^*=const} = \frac{1}{2}\left(\frac{\partial F}{\partial \mathbf{z}_r} - \jmath\frac{\partial F}{\partial \mathbf{z}_i}\right) \\ \mathbb{R}^*\text{–derivative}: \quad & \frac{\partial F}{\partial \mathbf{z}^*}\bigg|_{\mathbf{z}\ =const} = \frac{1}{2}\left(\frac{\partial F}{\partial \mathbf{z}_r} + \jmath\frac{\partial F}{\partial \mathbf{z}_i}\right). \end{aligned} \tag{10}$$

Also note that the direction of steepest descent is given by the derivative with respect to $\mathbf{z}^*$, the $\mathbb{R}^*$–derivative. This can be shown by using the first order Taylor Series Expansion (TSE) of F (van den Bos, 1994); the magnitude of a small change in the function F is given by

$$|\delta F| = 2\left|\Re\left\{\left(\frac{\partial F}{\partial \mathbf{z}^*}\right)^H \delta\mathbf{z}\right\}\right|$$

and the Cauchy-Schwarz Inequality shows that

$$|\delta F| \leq 2\left\|\frac{\partial F}{\partial \mathbf{z}^*}\right\| \cdot \|\delta\mathbf{z}\|$$

and so $|\delta F|$ is maximised when $\arccos \frac{\langle \frac{\partial F}{\partial \mathbf{z}^*}, \delta\mathbf{z}\rangle}{\|\frac{\partial F}{\partial \mathbf{z}^*}\| \, \|\delta\mathbf{z}\|} = 0$, or in other words the maximum change of the gradient is in the direction of the conjugate of the weight vector (Brandwood, 1983; Kreutz-Delgado, 2006). The operators $\Re\{\cdot\}$ and $\Im\{\cdot\}$, where used, denote the real and imaginary part of a complex quantity, while $\langle\cdot,\cdot\rangle$ is the inner product operator.

Furthermore, in calculating derivatives of analytic functions, as defined, the $\mathbb{R}^*$–derivative vanishes and the derivative is equivalent to the Cauchy-Riemann derivative, demonstrating the flexibility of the framework. This can be illustrated through a simple example. Consider the non-analytic squared error cost function $\mathcal{G} = \|\mathbf{z}\|_2^2 = \mathbf{z}\mathbf{z}^*$. Then, $\frac{\partial \mathcal{G}}{\partial \mathbf{z}} = \mathbf{z}^*$ and $\frac{\partial \mathcal{G}}{\partial \mathbf{z}^*} = \mathbf{z}$. In contrast, for the analytic function $\mathcal{H}(\mathbf{z}) = \mathbf{z}^2$, $\frac{\partial \mathcal{H}}{\partial \mathbf{z}} = 2\mathbf{z}$ and $\frac{\partial \mathcal{H}}{\partial \mathbf{z}^*} = \mathbf{0}$. For further insight into $\mathbb{CR}$ calculus, we refer to the material in (Kreutz-Delgado, 2006; Mandic and Goh, 2009).

2.5 BSE of Complex Noisy Mixtures

The diagram in Fig. 2 shows the complex BSE architecture, where at time instant k, the observed signal $\mathbf{x}(k) \in \mathbb{C}^N$ is given by a linear mixture

$$\mathbf{x}(k) = \mathbf{A}\mathbf{s}(k) + \mathbf{v}(k) \tag{11}$$

where $\mathbf{s}(k) \in \mathbb{C}^{N_s}$ is the vector of latent sources, $\mathbf{A} \in C^{N \times N_s}$ is the mixing matrix[2], and $\mathbf{v}(k) \in \mathbb{C}^N$ is the vector of additive doubly white Gaussian noise (noncircular). The model (11) has been widely used in EEG signal processing, for instance see (Cichocki and Amari, 2002; Sanei and Chambers, 2007). The sources are assumed to be with zero mean and distinct kurtosis values, while no assumptions are made about the circularity. The number of mixtures is assumed to be equal to that of the sources, however, in the case of noisy mixtures, an overdetermined mixture is necessary so as to estimate the second-order statistics of noise parameters.

The adaptive gradient descent algorithm at the extraction stage adapts the parameters of the demixing vector $\mathbf{w}$ such that the source

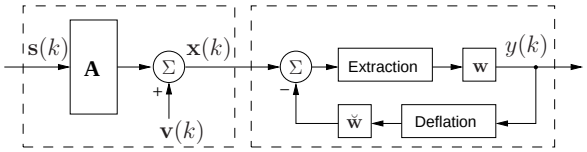


Fig. 2. The noisy mixture model, and BSE architecture.

[1] Also known as Wirtinger calculus.

[2] If a mixing process is considered for the noise given by $\mathbf{B}$, the vector $\mathbf{w}^H$ in the subsequent equations can be replaced by $\boldsymbol{\upsilon}^H = \mathbf{w}^H\mathbf{B}$. This does not affect our algorithm, for the normalised kurtosis of Gaussian noise is unconditionally zero.

signal with the largest (smallest) kurtosis,

$$\begin{aligned} y(k) &= \mathbf{w}^H\mathbf{x}(k) \\ &= \underbrace{\mathbf{w}^H\mathbf{A}}_{\triangleq \mathbf{u}^H}\mathbf{s}(k) + \mathbf{w}^H\mathbf{v}(k) \end{aligned} \tag{12}$$

is first extracted. The variance of $y(k)$ can be written in an expanded form as

$$\begin{aligned} E\{|y(k)|^2\} &= \mathbf{u}^H\mathcal{C}_{\mathbf{ss}}(0)\mathbf{u} + \mathbf{w}^H\mathcal{C}_{\mathbf{vv}}(0)\mathbf{w} \\ &= \mathbf{u}^H\mathbf{u} + \sigma_v^2\mathbf{w}^H\mathbf{w} \end{aligned} \tag{13}$$

where the difference in $\mathcal{C}_{\mathbf{ss}}(0)$ are absorbed into the mixing matrix $\mathbf{A}$ and the noise covariance matrix $\mathcal{C}_{\mathbf{vv}}(0) = \sigma_v^2\mathbf{I}$ (due to the whiteness assumption).
In the same spirit, the normalised kurtosis of the extracted signal $y(k)$ can be written as

$$\begin{aligned} K_c(y) &= \sum_{n=1}^{N_s} K_c(u_n^* s_n) + \sum_{n=1}^{N} \underbrace{K_c(w_n^* v_n)}_{=0} \\ &= \sum_{n=1}^{N_s} |u_n|^4 E\{|s_n|^4\} - 2|u_n|^4 (E\{|s_n|^2\})^2 - |u_n|^4 |E\{s_n^2\}|^2 \\ &= \sum_{n=1}^{N_s} |u_n|^4 K_c(s_n) \end{aligned} \tag{14}$$

thus having zero value for Gaussian noise. In a vectorised form, this is equivalent to

$$K_c(y) = \hat{\mathbf{u}}^H\mathbf{K}_c(\mathbf{s})\hat{\mathbf{u}} \tag{15}$$

where

$$\begin{aligned} \hat{\mathbf{u}} &= [u_1^2, \ldots, u_{N_s}^2] \\ \mathbf{K}_c(\mathbf{s}) &= diag\big(K_c(s_1), \ldots, K_c(s_{N_s})\big). \end{aligned} \tag{16}$$

The next stage within the proposed BSE scheme is the deflation process which aims to remove the extracted source $y(k)$ from the mixture $\mathbf{x}(k)$, such that

$$\mathbf{x}(k) \leftarrow \mathbf{x}(k) - \breve{\mathbf{w}}y(k)$$

where the deflation weight coefficient vector $\breve{\mathbf{w}}$ is updated using an adaptive gradient descent algorithm detailed later in this section. In principle, for $y(k)$ being an estimate of one of the original sources, say $s_n(k)$, the ideal deflation weight vector should be equal to the nth column of the mixing matrix $\mathbf{A}$, such that the effect of this particular source is removed from the mixture. Finally, a threshold can be set on the deflation process, so that extraction is continued until some or all the required sources have been successfully extracted (Thawonmas et al., 1998).

2.6 Cost function

The cost function we employed for the extraction of general complex sources from noisy mixtures is given by

$$\mathcal{J}(\mathbf{w}) = -\beta\frac{\text{kurt}_c\big(y(k)\big)}{\big(E\{|y(k)|^2\} - \mathbf{w}^H\mathcal{C}_{\mathbf{vv}}(0)\mathbf{w}\big)^2}. \tag{17}$$

Note that $\mathcal{J} \in \mathbb{R}$, represents a modified version of the normalised kurtosis defined in (6) and is a generalisation of the methodology presented in (Liu and Mandic, 2006). As illustrated in (13), the variance of $y(k)$ contains the noise variance σ_v^2, thus allowing us to remove the effect of noise from (17) such that only contributions from the latent sources are accounted for. Also note that while the noise variance σ_v^2 is present in the cost function (17), its pseudo-covariance τ_v^2 is not present, suggesting that the complex domain BSE based on kurtosis is unaffected by the pseudo-spectral effects of the additive noise; this is further elaborated in Section 3.
In the cost function (17), the parameter β dictates the order of extraction where for

1. $\beta = 1$, the order of extraction is from the high to low degree of non-Gaussianity (super-Gaussian sources are extracted first),
2. $\beta = -1$, the order of extraction is from low to high degree of non-Gaussianity (sub-Gaussian sources are extracted first).

The optimisation of $\mathcal{J}$ with respect to $\mathbf{w}$ can thus be stated as

$$\mathbf{w}_{opt} = \arg\max_{\|\mathbf{w}\|_2^2=1} \mathcal{J}(\mathbf{w}) \tag{18}$$

where the norm of the demixing vector is constrained to unity to avoid very small coefficient values.
Rewriting and simplifying (17) in terms of (13) and (16) results in

$$\mathcal{J}(\mathbf{w}) = -\frac{\hat{\mathbf{u}}^H|\mathbf{K}_c(\mathbf{s})|\hat{\mathbf{u}}}{(\mathbf{u}^H\mathbf{u})^2} = -\acute{\mathbf{u}}^H|\mathbf{K}_c(\mathbf{s})|\acute{\mathbf{u}} \tag{19}$$

where

$$\acute{\mathbf{u}}^H \triangleq \frac{\hat{\mathbf{u}}^H}{\mathbf{u}^H\mathbf{u}} = \frac{\hat{\mathbf{u}}^H}{\|\mathbf{u}\|_2^2}. \tag{20}$$

Notice that $\|\acute{\mathbf{u}}\|_2^2 = \frac{\|\hat{\mathbf{u}}\|_2^2}{(\|\mathbf{u}\|_2^2)^2} \leq 1$ and is equal to unity only if one of the components in the vector $\mathbf{u}$ is non-zero. Given the constraint on $\|\acute{\mathbf{u}}\|$, the solution to the optimisation of (19) is a vector $\acute{\mathbf{u}}_{opt}$ of unit norm such that $\mathbf{u}_{opt}$ has a single non-zero component at a position corresponding to the diagonal element in $\mathbf{K}_c(\mathbf{s})$ having the largest magnitude. For this to be valid, a demixing vector assumes the form $\mathbf{w}_{opt} = \mathbf{A}^{H\#}\mathbf{u}_{opt}$, where the symbol $(\cdot)^{\#}$ denotes the matrix pseudo-inverse operator (Liu and Mandic, 2006).

2.7 Adaptive algorithm for extraction

Optimisation of (17) is performed using an adaptive gradient descent algorithm which updates the values of $\mathbf{w}$ so as to maximise the modified normalised kurtosis and thus minimise the cost function $\mathcal{J}(\mathbf{w})$. Based on Section 2.4, the gradient[3] is thus expressed as

$$\begin{aligned} \nabla_{\mathbf{w}^*}\mathcal{J} &= \frac{\beta\,\mathbf{x}(k)}{(m_2(y) - \sigma_v^2)^3}\Big[y^*(k)\big(m_4(y) - 2m_2^2(y) - |p_2(y)|^2\big) \\ &+ \big(m_2(y) - \sigma_v^2\big)\big(-y(k)y^{*2}(k) + 2m_2(y)y^*(k) + p_2^*(y)y(k)\big)\Big] \\ &= \phi\big(y(k)\big)\mathbf{x}(k) \end{aligned}$$

where $\phi\big(y(k)\big)$ is used for simplification and $m_\ell(y)$ and $p_\ell(y)$ are respectively the ℓ-th moment and pseudo-moment at time instant

[3] Since the normalised kurtosis $\mathcal{J}$ is real valued, the conjugate gradient $\frac{\partial\mathcal{J}}{\partial\mathbf{w}^*}$ corresponds to the maximum change of the gradient.

k (the time index dropped), estimated using the moving average estimators

$$m_\ell\big(y(k)\big) = (1-\alpha)m_\ell\big(y(k-1)\big) + \alpha|y(k)|^\ell,\ \ell = \{2,4\}$$
$$p_\ell\big(y(k)\big) = (1-\alpha)p_\ell\big(y(k-1)\big) + \alpha\big(y(k)\big)^\ell,\ \ell = 2 \quad (21)$$

where $\alpha \in [0,1]$ is the forgetting factor.
The kurtosis based BSE update algorithm (K-cBSE) for the demixing vector is thus given by

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \mu\phi\big(y(k)\big)\mathbf{x}(k)$$

and its expanded version is given in (22) ,where μ is the small positive step-size. To preserve the unit norm property, the demixing vector is normalised at each iteration, that is

$$\mathbf{w}(k+1) \leftarrow \frac{\mathbf{w}(k+1)}{\|\mathbf{w}(k+1)\|_2}$$

Notice that in extracting circular sources, the moment p_ℓ vanishes, further simplifying the algorithm. Moreover, as mentioned earlier, the cost function and thus the gradient descent algorithm are not dependent on the pseudo-variance of the noise, τ_v^2. The estimation of the noise variance can be performed using a subspace method, as described in (Hayes, 1996). It is thus essential that the number of observations is larger than the number of sources, $N > N_s$, so as to allow for the estimation of the noise variance via eigenvalues of the observation covariance matrix $\mathcal{C}_{\mathbf{xx}}$, that is

$$\mathcal{C}_{\mathbf{xx}} = \mathbf{A}\mathcal{C}_{\mathbf{ss}}\mathbf{A}^H + \mathcal{C}_{\mathbf{vv}} = \mathbf{\Upsilon} + \sigma_v^2\mathbf{I}. \quad (23)$$

The subspace method can be briefly summarised as follows. We can assume Rank($\mathbf{\Upsilon}$) $= N_s$ if $\mathbf{A}$ is of full rank and $\mathcal{C}_{\mathbf{ss}}$ is non-singular. Then, the $(N - N_s)$ eigenvalues of $\mathbf{\Upsilon}$ are zero and hence the $(N - N_s)$ eigenvalues of $\mathcal{C}_{\mathbf{xx}}$ are equal to σ_v^2.

2.8 Modifications to the update algorithm

In order to enhance the performance of the online gradient descent algorithm, adaptive step-size update algorithms are considered. We consider the complex-valued Farhang-Ang type variable step-size (VSS) algorithm (Ang and Farhang-Boroujeny, 2001) and the complex-valued generalised normalised gradient descent (GNGD) type algorithm (Mandic, 2004).
At each iteration k, the VSS algorithm minimises the cost function $\mathcal{J}$ in (17) with respect to $\mu(k-1)$ to provide the update of the step-size, given as

$$\begin{aligned}\mu(k) &= \mu(k-1) - \eta\nabla_\mu\mathcal{J}\big|_{\mu=\mu(k-1)}\\ \nabla_\mu\mathcal{J} &= \nabla_{\mathbf{w}^*}\mathcal{J}\cdot\frac{\partial\mathbf{w}^*(k)}{\partial\mu(k-1)}\\ \psi(k) &= \gamma\psi(k-1) - \nabla_{\mathbf{w}^*}\mathcal{J}\big|_{\mathbf{w}^*=\mathbf{w}^*(k-1)}\end{aligned} \quad (24)$$

where $\psi(k) \triangleq \frac{\partial\mathbf{w}^*(k)}{\partial\mu(k-1)} \approx \frac{\partial\mathbf{w}^*(k)}{\partial\mu(k)}$ and η and γ are step-sizes.
The GNGD-type algorithm is based on a normalised version of (22), given by

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \frac{\mu}{|\phi\big(y(k)\big)|^2\cdot\|\mathbf{x}(k)\|_2^2 + \epsilon(k)}\phi\big(y(k)\big)\mathbf{x}(k) \quad (25)$$

where $\epsilon(k)$ is an adaptive regularisation parameter. The gradient adaptive regularisation parameter is then given by

$$\epsilon(k+1) = \epsilon(k) - \rho\mu\frac{\Re\{\phi\big(y(k)\big)\mathbf{x}^T(k)\phi^*\big(y(k-1)\big)\mathbf{x}^*(k-1)\}}{\big[|\phi\big(y(k-1)\big)\cdot\|\mathbf{x}(k-1)\|_2^2 + \epsilon(k-1)\big]^2} \quad (26)$$

where ρ is a step-size. The derivation of the algorithm is given in the Appendix.

2.9 Adaptive algorithm for deflation

The deflation procedure insures that after each extraction stage, the estimated source is removed from all the mixture vectors, so that the next source with maximum (minimum) kurtosis can be extracted. This can be achieved based on the cost function (Thawonmas et al., 1998)

$$\mathcal{J}_d(\breve{\mathbf{w}}) = \|\mathbf{x}_{n+1}(k)\|^2 = \mathbf{x}_{n+1}^H(k)\mathbf{x}_{n+1}(k) \quad (27)$$

which is minimised with respect to the deflation weight coefficient $\breve{\mathbf{w}}$. We use $\mathbf{x}_n(k)$ to denote the mixture at the nth extraction stage, which is given by vectors

$$\mathbf{x}_{n+1}(k) = \mathbf{x}_n(k) - \breve{\mathbf{w}}(k)y_n(k). \quad (28)$$

Given an invertible mixing matrix $\mathbf{A}$, the vector $\breve{\mathbf{w}}$ is ideally equal to a column of $\mathbf{A}^{-1}$, which corresponds to the nth extracted source $y_n(k)$. The gradient can thus be calculated as

$$\nabla_{\breve{\mathbf{w}}^*}\mathcal{J}_d = \frac{\partial\mathcal{J}_d}{\partial\mathbf{x}_{n+1}^*}\cdot\frac{\partial\mathbf{x}_{n+1}^*}{\partial\breve{\mathbf{w}}^*} = -y_n^*(k)\mathbf{x}_{n+1}(k) \quad (29)$$

and the online algorithm for BSE then becomes

$$\breve{\mathbf{w}}(k+1) = \breve{\mathbf{w}}(k) + \mu_d y_n^*(k)\mathbf{x}_{n+1}(k) \quad (30)$$

with μ_d a step-size. The drawback of this method is that any errors in the deflation process will propagate and affect the extraction and deflation of subsequent stages. It is therefore important that the step-size parameter is set appropriately for each nth deflation stage to ensure successful removal of the extracted source $y_n(k)$.
In the design of complex adaptive algorithms, it is common to utilise a widely linear model to ensure that the algorithm is capable of processing the generality of complex signals (Mandic and Goh, 2009). In the case of the update for the deflation weight coefficient (30), however, a linear model is considered as the original BSS mixing model (11) is strictly linear and thus a widely linear deflation model is not required. For more detail on BSE based on widely linear predictability, see (Javidi et al., 2010).

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \frac{\mu\beta\Big[y^*(k)\big(m_4(y) - 2m_2^2(y) - |p_2(y)|^2\big) + \big(m_2(y) - \sigma_v^2\big)\big(-y(k)y^{*2}(k) + 2m_2(y)y^*(k) + p_2^*(y)y(k)\big)\Big]}{(m_2(y) - \sigma_v^2)^3}\mathbf{x}(k) \quad (22)$$

3 RESULTS AND DISCUSSIONS

We shall consider extraction of both synthetic and real-world sources from noise-free and noisy mixtures, with various degrees of complex noncircular noise levels. The performance for the synthetic data were measured using the Performance Index (PI) (Cichocki and Amari, 2002) given by

$$PI = 10\log_{10}\left(\frac{1}{N}\Big(\sum_{i=1}^{N}\frac{|b_i|^2}{\max\{|b_1|^2,\ldots,|b_N|^2\}}-1\Big)\right). \quad (31)$$

where $\mathbf{u} = \mathbf{A}^H\mathbf{w} = [u_1,\ldots,u_M]$.

For each synthetic experiment, the results were produced through averaging 100 independent trials. The mixing matrix $\mathbf{A}$ was generated randomly as a full rank complex matrix. The values of the extraction and deflation step-size μ were set empirically, and the forgetting factor α in (21) was set as 0.975. The complex additive Gaussian noise was both of circular white with circularity measure $r = 0$ and noncircular doubly white with $r = 0.93$. The real-world sources were the electroencephalogram data corrupted by power line noise and electrooculogram artifacts.

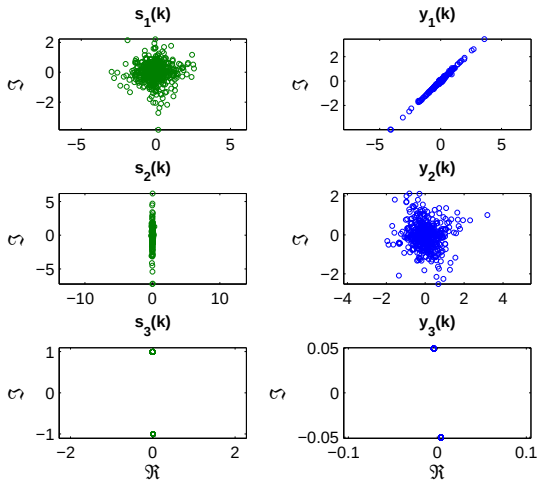


Fig. 3. Scatter plot of the complex-valued sources $s_1(k)$, $s_2(k)$ and $s_3(k)$, with the signal properties described in Table 1 (*left hand column*). Scatter plot of estimated sources $y_1(k)$, $y_2(k)$ and $y_3(k)$, extracted according to a decreasing order of kurtosis ($\beta = 1$) (*right hand column*).

3.1 Benchmark Simulation 1: Synthetic sources

In the first set of simulations, a noise-free mixture of 3 complex sources with various degrees of circularity and $N = 5000$ samples were generated and mixed using a 3×3 mixing matrix. These signals are illustrated in Fig. 3 and their properties listed in Table 1. Extraction was performed in order from highest to lowest kurtosis, hence the value of $\beta = 1$ in (17).

In the first experiment, the performance of the algorithm (22) using the adaptive step-size methods was compared in the extraction of the first source with the value of μ set to 0.01. It can be seen from the performance curves in Fig. 4 that the best performance was achieved using the GNGD method with a PI of around -45dB at the steady-state. The performance curve resulting from the normalised method indicates successful extraction with a PI of around -25dB. The performance of the algorithm using the standard step-size and VSS were comparable, with a PI of around -20dB. In the following simulations, the GNGD based K-cBSE algorithm is utilised.

In the next set of simulations, we considered the extraction of all the three sources (Fig. 3). The value of μ was set respectively to $0.01, 0.008$ and 10^{-5} for the consecutive extraction stages. As shown in Fig. 5, the algorithm successfully extracted all the three sources, as indicated by a PI of less than -20dB at the steady-state for the extraction iteration $i = \{1,2,3\}$, converging to steady-state after 2500 samples in the first extraction stage ($i = 1$) and around 1000 samples in the second and third extraction stage ($i = \{2,3\}$). The decreasing PI value at each consecutive extraction stage can be attributed to the unavoidable errors accumulated in the deflation. The scatter plot of the three estimated sources $y_1(k), y_2(k)$ and $y_3(k)$ are illustrated in Fig. 3. The normalised kurtosis of the estimated sources were respectively calculated as $K_c(y_1) = 11.8425, K_c(y_2) = 1.3599$ and $K_c(y_3) = -1.9956$ corresponding to those of the original sources, given in Table 1; the scale and rotation ambiguities of the source estimates are also visible.

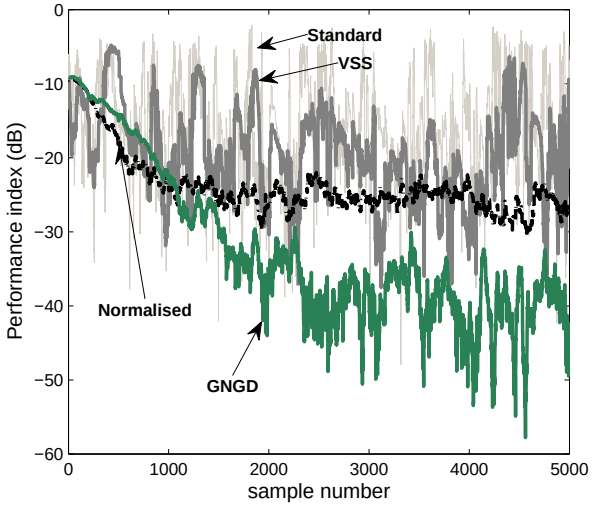


Fig. 4. Comparison of the effect of step-size adaptation on the performance of algorithm (22) for the extraction of a single source.

3.2 Benchmark Simulation 3: Noisy mixture

In the next experiment, we considered the extraction of a complex-valued source from a noisy mixture. Three sources of $N = 5000$ samples were considered (see Table 2, Fig. 6) and mixed using a randomly generated 4×3 mixing matrix $\mathbf{A}$. The additive noise was doubly white Gaussian noise with variance $\sigma_v^2 = 0.1$ and pseudo-variance $\tau_v^2 = 0.0924 + \jmath 0.0011$, estimated using the subspace method described in Section 2.5. The sources were extracted in an increasing order of kurtosis ($\beta = -1$) with the step-size $\mu = 0.5$. The scatter plot of the first estimated source with the smallest kurtosis, $y_1(k)$ is illustrated in Fig. 6 with a calculated normalised

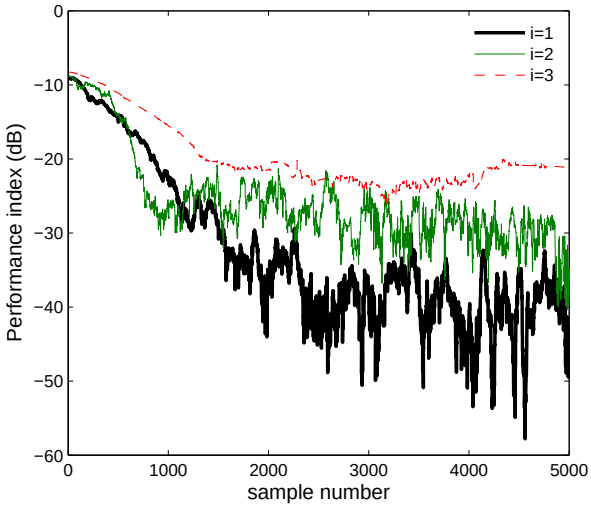


Fig. 5. Extraction of complex circular and noncircular sources from a noise-free mixture based on kurtosis.

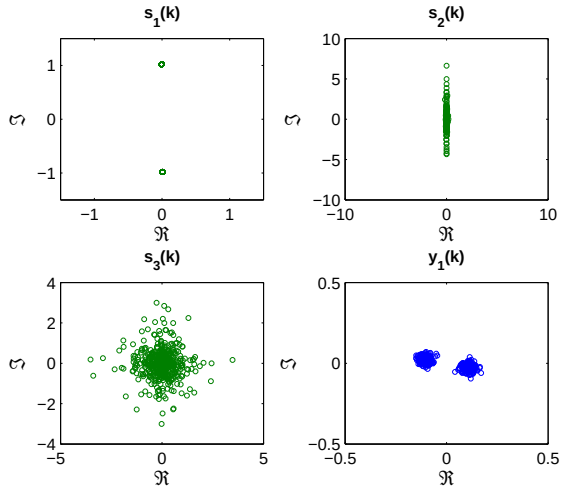


Fig. 6. Scatter plots of the original sources $s_1(k), s_2(k)$ and $s_3(k)$. The first estimated source $y_1(k)$ is shown in the bottom-right plot.

kurtosis of $K_c(y_1) = -1.8002$, which is within a 10% range of the true value, given in Table 2. The Performance Index, shown in Fig. 7, demonstrates a fast convergence to a value of around -40dB in approximately 1000 samples, and continuing a steady convergence to -50dB by 5000 samples.

It was shown in Section 2.5 that the performance of the algorithm (22) is not affected by the degree of circularity of the additive noise, such that doubly white noise is treated in a similar manner to circular white noise, where the pseudo-covariance vanishes. This was explored experimentally by systematically analysing the effect of various noise levels on the BSE algorithm (22). The circularity measure r was varied from a value of $r = 0$ (circular) to a value of $r = 0.9998$ (highly noncircular),

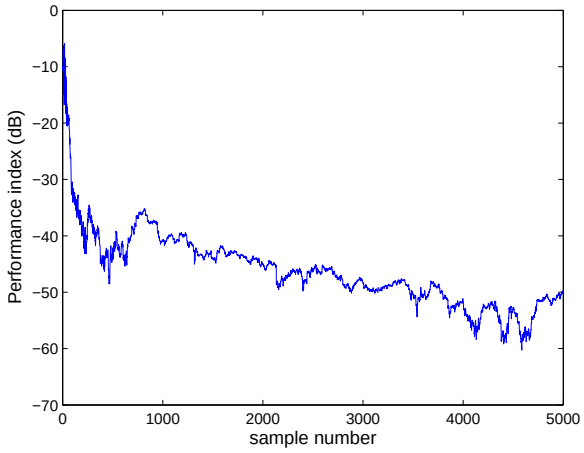


Fig. 7. Extraction of a complex-valued source from a noisy mixture, with the source properties given in Table 2.

Table 1. Source properties for noise-free extraction Benchmark Simulation 1

Source	Distribution	Kurtosis	circ. measure (r)
$s_1(k)$	Super-Gaussian	1.3587	0.0386
$s_2(k)$	Super-Gaussian	11.8890	0.9955
$s_3(k)$	Sub-Gaussian	-1.9999	1.0000

Table 2. Source properties for noisy extraction in Benchmark Simulation 3

Source	Distribution	Kurtosis	circ. measure (r)
$s_1(k)$	Sub-Gaussian	-1.9985	1.0000
$s_2(k)$	Super-Gaussian	19.1167	0.9988
$s_3(k)$	Super-Gaussian	1.5426	0.0147

while the signal-to-noise ratio (SNR) was adjusted from a near-zero noise SNR of 50dB to a high noise environment with SNR value of -10dB. The initial values were generated randomly and PI was averaged over 100 trials. Fig. 8 illustrates the performance curve for the different variations in the noise properties, and confirms that while the performance is dependent on the SNR value, it does not vary with changes in the degree of noise noncircularity. In addition, the maximum effective range of the algorithm in extracting sources (PI $<$ -20dB) can be estimated as an SNR of 1 dB.

3.3 EEG artifact extraction

In order to obtain useful information from EEG data in real-time, it is often necessary to perform post-processing to remove artifacts such as line noise and biological artifacts including those pertaining to eye movement, captured in the form of electrooculogram (EOG) and facial muscle activity represented as electromyogram (EMG). Removal of the effect of such signals from the contaminated EEG has been subject of study in previous years, with several methodologies introduced that attempt to accomplish this utilising both online and offline algorithms (Vigário, 1997; Jung et al., 2000; Delorme et al., 2001; Barbati et al., 2004; Greco et al.,

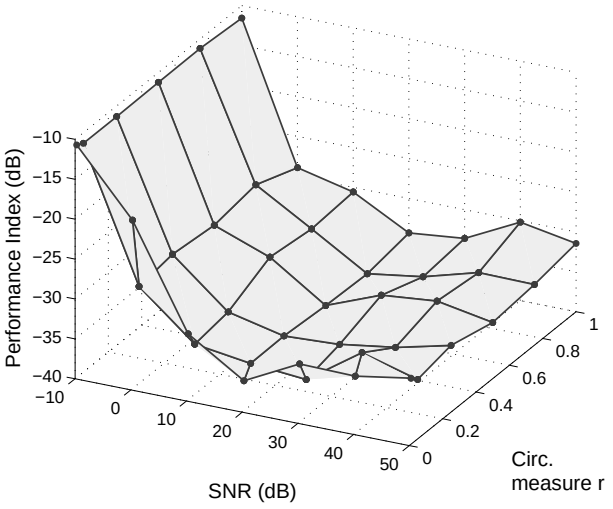


Fig. 8. Comparison of the performance of algorithm (22) with respect to changes in the SNR and the degree of noise circularity.

2005; Delorme et al., 2007; Kumar et al., 2009). While offline algorithms are suitable for processing the recorded EEG data in clinical applications, it is necessary to utilise online algorithms for real-time applications such as those encountered in brain computer interface (BCI) scenarios.

In (Kumar et al., 2009) the authors propose an online algorithm whereby the recorded EEG signals are transformed to the wavelet domain and the EOG contaminants are removed using an adaptive recursive least squares (RLS) algorithm, before transforming the signal back to the time domain. Simulations demonstrate good performance from the algorithm, however, it would be advantageous to perform all the necessary processing in the time domain, as this way the signals are retained in their original form and less computation is required.

In its basic form, ICA can be applied to the contaminated EEG recording and the artifacts removed through visual inspection. As detailed in (Vigário, 1997), an ICA algorithm separates the recorded EEG mixture into its original sources as independent components (ICs), with artifact sources identified and removed. In semi-automatic (Greco et al., 2005) and automatic (Delorme et al., 2001) artifact removal methodologies, several classifications (markers) based on the statistical characteristics of the ICs are considered that allow for the detection of artifacts in the contaminated EEG, which are then compared against thresholds that determine the rejection of particular components. In these methods, both the kurtosis and entropy of independent components have been utilised to identify and remove the artifacts. While the EEG mixtures typically have near-zero kurtosis values, artifacts such as EOG exhibit peaky distributions with highly positive kurtosis values (Delorme et al., 2001), while periodic power line noise has a highly negative kurtosis value. This has been used as the main discrimination in defining classifications based on the the fourth order moment.

3.3.1 Data acquisition and method Our aim is to remove artifacts as independent sources extracted from the recorded EEG mixture directly in the time domain. To this end, the contaminated EEG signals were paired as the real and imaginary components of a complex signal and processed using the architecture described in Section 2.5. In this manner, the phase-amplitude relationship and the full cross-statistical information between symmetric electrode pairs are contained in the corresponding complex-valued EEG signal, and allow for the simultaneous processing of both channels. Further iterations of the extraction process can then be used to obtain the individual pure EEG signals, or even, pipelined to a further post-processing stage, which would then extract the EEG signals based on a desired fundamental property, such as predictability.

The electrodes were placed according to the 10-20 system (Fig. 9), and sampled at 256Hz for 30 seconds. The EEG activity was recorded from electrodes placed at positions Fp1, Fp2, C3, C4, O1, O2 with the ground placed at Cz, while the EOG activity was recorded from the vEOG and hEOG channels with electrodes placed above and on the side of the left eye socket.

Three studies were performed with the aim to remove the artifacts simultaneously. While the rejection of the power line noise artifact is feasible by passing the recorded EEG signals through a notch filter, this solution also leads to the removal of useful information around the 50Hz range pertaining to the EEG signals, in particular those within the gamma band (25Hz-100Hz). It would therefore be desirable to automatically extract the line noise artifact along with the biological artifact from the corrupted EEG signals. In the first study we consider the removal of EOG artifacts ('EYEBLINK' set), the second study focused on eye muscle artifacts from rolling the eyes ('EYEROLL' set), whereas the third study addressed the removal of muscle activity from raising the eyebrow ('EYEBROW' set).

In all the studies, the temporal signals from each channel pair were combined to form three complex EEG channels, given by

$$\begin{aligned} x_1(k) &= \text{Fp1}(k) + \jmath\text{Fp2}(k) \\ x_2(k) &= \text{C3}(k) + \jmath\text{C4}(k) \\ x_3(k) &= \text{O1}(k) + \jmath\text{O2}(k). \end{aligned} \tag{32}$$

This construction of the complex EEG signals allows for the simultaneous processing of the amplitude and phase information using the K-cBSE algorithm (22). Note that the EOG channels were not part of the mixtures considered. They are only used to assess the performance of the proposed BSE algorithm in the extraction of the EOG artifacts.

3.3.2 Performance measures As we have no knowledge of the mixing process, the Performance Index (31) is not applicable for this case and thus several alternative quantitative and qualitative measures were used for the evaluation of the algorithm performance. These are briefly discussed below.

1. Quantitative metrics

   a. *Kurtosis*: The kurtosis values K_c of the complex extracted signals indicate the success of the algorithm in extracting super-Gaussian or sub-Gaussian artifact in a specified order. In addition, the magnitude of the kurtosis K_R of the real and imaginary components of the extracted sources are used to automatically select desired components.

b. *Power spectra Correlation*: In a similar manner to (Barbati et al., 2004), the correlation coefficient between the power spectra of the complex-valued recorded artifact (e.g. EOG) and extracted sources, and likewise, the correlation coefficient between the pseudo-power spectra of the complex-valued recorded artifact and the extracted sources is calculated. This measure indicates the degree of similarity between the extracted and originally recorded artifact, and can be used to automatically select the extracted source pertaining to the biological artifact, while also quantifying the degree of performance of the extraction algorithm.

2. Qualitative metrics

a. *Hilbert-Huang Time-Frequency Analysis*: By employing time-frequency (T-F) analysis using the Hilbert-Huang (H-H) transform (Huang et al., 1998; Huang and Shen, 2005), we can qualitatively assess the extraction performance through comparison of the frequency components of the mixture and extracted source during the recording session. Also, the T-F analysis of the extracted artifacts will demonstrate the corresponding frequency components and their changes over time, making it possible to assess the quality of the extraction procedure over the recording time. In comparison to Fourier transform based T-F analysis, such as the Short-Time Fourier Transform, the H-H transform results in much more detailed spectrogram for a given resolution. The intrinsic mode functions (IMFs) required by the H-H transform were obtained using a multivariate empirical mode decomposition (MEMD) algorithm (Rehman and Mandic, 2010), where the real and imaginary component of the complex-valued signals were taken as a single multivariate signal and processed simultaneously. It was observed that this resulted in a spectrogram with better resolution than those obtained through the separate processing of the individual components using the standard EMD algorithm.

b. *Power Spectral Distribution*: The power and pseudo-power spectra of the complex-valued extracted artifacts were compared to those belonging to the complex-valued recorded artifact. In addition, the pseudo-spectrum demonstrates the quality of the proposed method in extracting noncircular sources. It is also possible to consider the cross-spectrum of the recorded and extracted sources (Palmer et al., 2009).

3.3.3 Case Study 1 – EOG extraction The 'EYEBLINK' dataset contained the EEG recordings contaminated with eye blink artifact as well as line noise. The recorded EEG and EOG signals are plotted in Fig. 10(a), where the effect of the EOG activity is pronounced in the frontal lobe (Fp1 and Fp2 channels), with the effect diminishing with an increase in the distance of the electrodes to the eyes. The effect of the line noise is also visible on the occipital O1 and O2 channels. The H-H T-F spectrogram (Fig 10(b)) describes the frequency changes of the ensemble average of the 6 EEG channels over the recording period. In correspondence with the time plot, the EOG artifacts are visible (with a duration of around 1 seconds); constant frequency components are seen around the 50Hz range due to the line noise. Note that due to the low sampling rate of the recording device, the 50Hz frequency component is not well defined in the T-F analysis and results in scattering of frequency components between 40Hz-60Hz.

The complex EEG signals formed using (32) were processed using the K-cBSE algorithm with the value of $\mu = \{5, 0.09\}$ and $\beta = \{-1, 1\}$ for the consecutive iterations and $\alpha = 0.975$. The choice of value for β ensures that the line noise is initially extracted, followed by the EOG components in the second iteration. The normalised kurtosis values of the original real-valued EEG signals and the extracted EEG signals are given in Tables 3 and 4.

The order of the extracted complex signals were as expected, with the first extracted source $y_1(k)$ (line noise) being sub-Gaussian and $y_2(k)$ (EOG) super-Gaussian. The imaginary component of $y_1(k)$ had the smallest kurtosis, and was automatically chosen as the extracted line noise source, while the near zero kurtosis of the real component $\Re\{y_1(k)\}$ indicates an EEG source. Also, both components of the second extracted source, having a high kurtosis value, were considered as the extracted EOG sources. Fig. 10(c) shows the T-F plots of the imaginary components of the first extracted signal $y_1(k)$ where the presence of the power line artifact is seen, while in Fig. 10(d) the T-F plot of the real and imaginary components of $y_2(k)$ is shown where the frequency components of the EOG artifacts are seen.

We next concentrate on the power spectrum and pseudo-power spectrum of the complex EOG signal, constructed in a similar manner to that in (32); the extracted sources $y_1(k)$ and $y_2(k)$ are depicted in Fig. 10(e). Notice that the distribution of power $\mathcal{S}_{EOG}$ and pseudo-power $\breve{\mathcal{S}}_{EOG}$ is concentrated respectively in the frequency range (0-5)Hz and 50Hz. The spectrum $\mathcal{S}_{y_1}$ and pseudo-spectrum $\breve{\mathcal{S}}_{y_2}$ of the first extracted source can be seen to contain around 0dB of power for a frequency of 50Hz, while having an average power of -40dB in the (0-5)Hz frequency range. These results can also be seen by comparing the frequency components of the recorded EEG mixture and extracted artifactual sources around the 50Hz range, shown in Fig. 10(f). While the presence of the power line artifact is evident in all recorded channels, after the extraction procedure the 50Hz frequency component is only

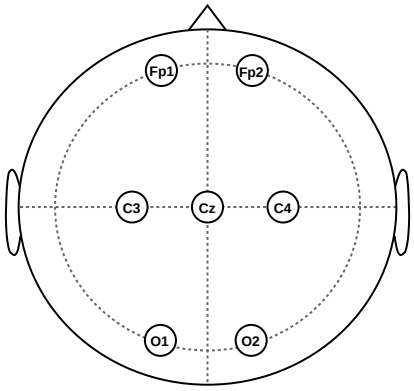


Fig. 9. Placement of the EEG electrodes on the scalp according to the recording 10-20 system.

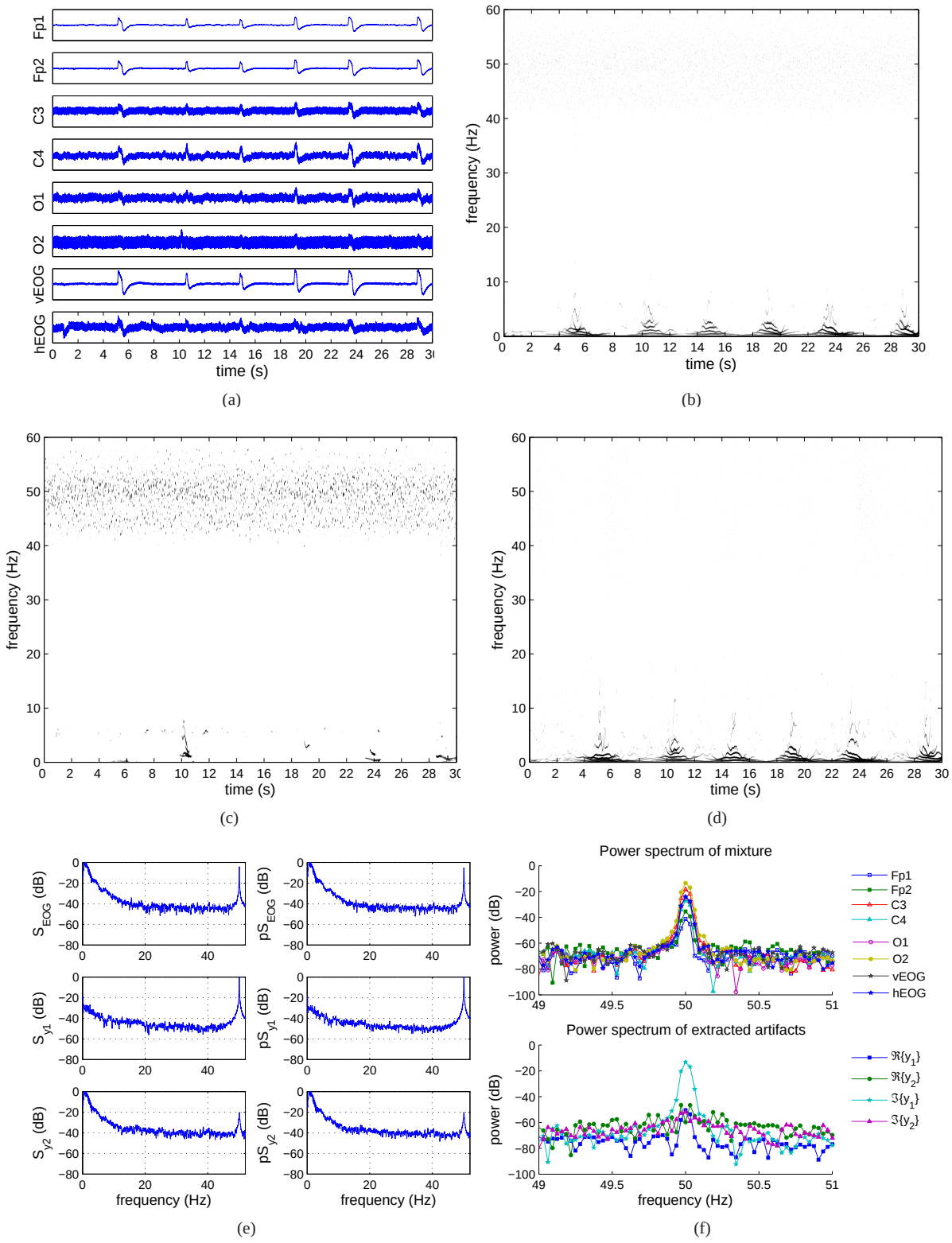


Fig. 10. Recorded and extracted artifacts from the 'EYEBLINK' set. (a) Recorded EEG signals from the 'EYEBLINK' set. (b) The Hilbert-Huang time-frequency plot of the recorded EEG signals. (c) The Hilbert-Huang time-frequency plot of the extracted line noise $\Im\{y_1(k)\}$. (d) The Hilbert-Huang time-frequency plot of the extracted EOG $\Re\{y_2(k)\}, \Im\{y_2(k)\}$. (e) The power spectra (S) and pseudo-spectra (pS) of the recorded EOG, and the extracted signals $y_1(k)$ and $y_2(k)$. (e) The power spectra (S) and pseudo-spectra (pS) of the recorded EOG, and the extracted signals $y_1(k)$ and $y_2(k)$. (f) Frequency components of the recorded EEG signals and the extracted artifacts at the 50Hz frequency range. After extraction, the power line noise is contained in $\Im\{y_1\}$.

Table 3. Normalised kurtosis values of the recorded EEG/EOG signals in real- and complex-valued form

	Set		
Electrode	'EYEBLINK'	'EYEROLL'	'EYEBROW'
Fp1	7.7452	3.3601	7.4152
Fp2	6.4793	2.2608	7.5034
C3	-0.2922	-0.0938	-0.4951
C4	1.1548	1.2469	1.5348
O1	-0.2550	0.8303	-0.5989
O2	-0.9574	-0.6782	-0.9526
vEOG	7.7541	4.8385	10.8653
hEOG	-0.1475	2.3883	-0.3264
$x_1(k)$	7.0318	2.6390	6.1156
$x_2(k)$	0.1006	0.4501	-0.0146
$x_3(k)$	-0.9164	-0.4601	-0.9285

present in $\Im\{y_1(k)\}$. Likewise, the spectra of $y_2(k)$ illustrate the diminished effect of the line noise source with a power of -20dB, while retaining the frequency components of the EOG in the low frequency range. To quantify the observed results, the correlation coefficient between the recorded EOG's PSD and pPSD and those of the extracted sources were calculated (Barbati et al., 2004) and presented in Table 4. For the extracted source $y_1(k)$ these values were respectively 0.2313 and 0.2847, whereas for the source $y_2(k)$ they were 0.9698 and 0.9822. The correspondence of the results between the power and pseudo-power spectra demonstrates the effectiveness of the methodology in extracting artifacts in the complex domain.

3.3.4 Case Study 2 – Eye muscle artifact extraction The 'EYEROLL' dataset had contained artifacts from round movement of the eye during the recording session with EOG activity from eye blinks, shown in Fig. 11(a) and kurtosis values given in Table 3.

The resultant electrical activity from the artifacts were recorded using the vEOG and hEOG channels, with EOG activity seen on the vEOG channel at time instants 5s, 13s, 17s, 23s, 25s and 29s, and eye muscle activity present more clearly on the hEOG channel with a duration of around 2s. The eye muscle artifact was present on all six EEG channels, while the EOG artifact is strong on the Frontal lobe electrodes and the effect of the power line noise is seen more strongly on the central and occipital lobe electrodes. The H-H T-F analysis of Fig. 11(b) illustrates the presence of frequency components up to 10Hz, as well as scattered frequencies belonging to the 50Hz power line noise.

In the extraction procedure, the step-size of the K-cBSE algorithm was $\mu = \{5, 0.2\}$ and $\beta = \{-1, 1\}$, while $\alpha = 0.975$. The T-F analysis of the extraction are illustrated in Fig. 11(c)–(d), and the kurtosis values of the complex-valued extracted signals and their real and imaginary components given in Table 4.

The real component of the first extracted source, $\Re\{y_1(k)\}$, having the smallest kurtosis of $K_c(\Re\{y_1) = -1.1958$ contained the power line noise artifact. The eye muscle activity and EOG artifacts were collectively extracted using the real and imaginary components of the second extracted source $y_2(k)$. The five instances of the eye muscle activity and the EOG can be detected in Fig. 11(d), while the lack of power line noise frequency components in the 50Hz range is visible.

These results were also confirmed based on the power spectra of the recorded artifacts and the extracted sources, given in Fig. 11(e). While the PSD and pPSD of the complex-valued $y_1(k)$ contained the 50Hz components, these were suppressed to -40dB in the spectra of $y_2(k)$. The frequency components of the mixture channels and extracted artifacts in the 50Hz range also showed that the line noise artifact was successfully removed (see Fig. 11(f)). Conversely, the spectral components pertaining to the eye muscle and EOG artifacts are present in the PSD and pPSD of $y_2(k)$ corresponding to the (0-10)Hz range of the PSD and pPSD of the complex-valued EOG. The correlation coefficient between the PSD spectra of the complex-valued recorded EOG channel and extracted source $y_2(k)$ is 0.8244, while the correlation between the pPSD spectra was 0.8222; these values were respectively 0.0792 and 0.1844 for $y_1(k)$.

3.3.5 Case Study 3 – EMG extraction In the 'EYEBROW' set, the EEG mixture was heavily contaminated with EMG artifacts from raising the eyebrows, and are shown in Fig. 12(a) with kurtosis values given in Table 3. The EMG signals were recorded using the vEOG and hEOG electrodes, with the effect more prominent on the vEOG recording. All EEG channels were affected by the artifact, though this is not clearly visible in the occipital lobe channels due to the strong presence of power line noise. In the T-F domain (Fig. 12(b)) the EMG frequency range had a large span containing both low and high frequency components, present in the duration of the raising of the eyebrows and lasting for around 2s. In addition, the 50Hz frequency component cloud reflecting the power line noise can also be seen.

The extraction of the artifacts was performed using the K-cBSE algorithm (22) with step-size $\mu = \{2, 0.2\}$, $\beta = \{-1, 1\}$ and $\alpha = 0.975$.

Table 4. Normalised kurtosis values of the extracted artifacts, and the correlation coefficient of the power and pseudo-power spectra respectively with the spectra of the recorded EOG

				Spectra corr.	
Set	Signal	K_c	$K_R(\Re, \Im)$	PSD	pPSD
'EYEBLINK'	$y_1(k)$	-1.2223	-0.0893, -1.2392	0.2313	0.1847
	$y_2(k)$	7.3914	7.5051, 5.1583	0.9698	0.9822
'EYEROLL'	$y_1(k)$	-1.1744	-1.1958, -0.0341	0.0792	0.1844
	$y_2(k)$	3.0644	3.5217, 2.7289	0.8244	0.8222
'EYEBROW'	$y_1(k)$	-1.0100	-0.7254, -1.1319	0.1287	0.1078
	$y_2(k)$	4.5144	5.4278, 6.3792	0.7593	0.7906

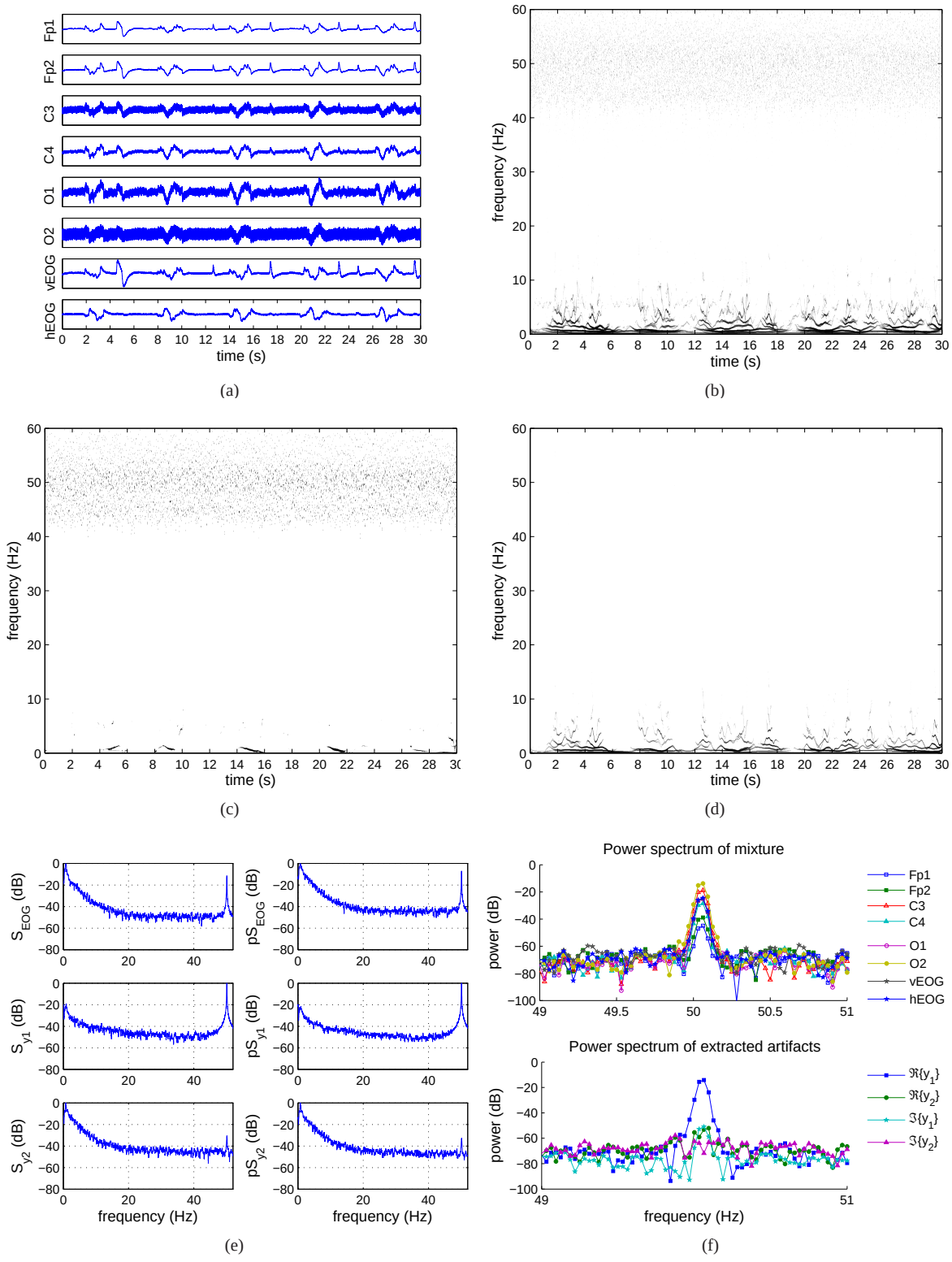


Fig. 11. Recorded and extracted artifacts from the 'EYEROLL' set. (a) Recorded EEG signals from the 'EYEROLL' set. (b) The Hilbert-Huang time-frequency plot of the recorded EEG signals. (c) The Hilbert-Huang time-frequency plot of the extracted line noise $\Re\{y_1(k)\}$. (d) The Hilbert-Huang time-frequency plot of the extracted EOG $\Re\{y_2(k)\}, \Im\{y_2(k)\}$. (e) The power spectra (S) and pseudo-spectra(pS) of the recorded EOG, and the extracted signals $y_1(k)$ and $y_2(k)$. (f) Frequency components of the recorded EEG signals and the extracted artifacts around the 50Hz frequency range. After extraction, the power line noise is contained in $\Re\{y_1\}$.

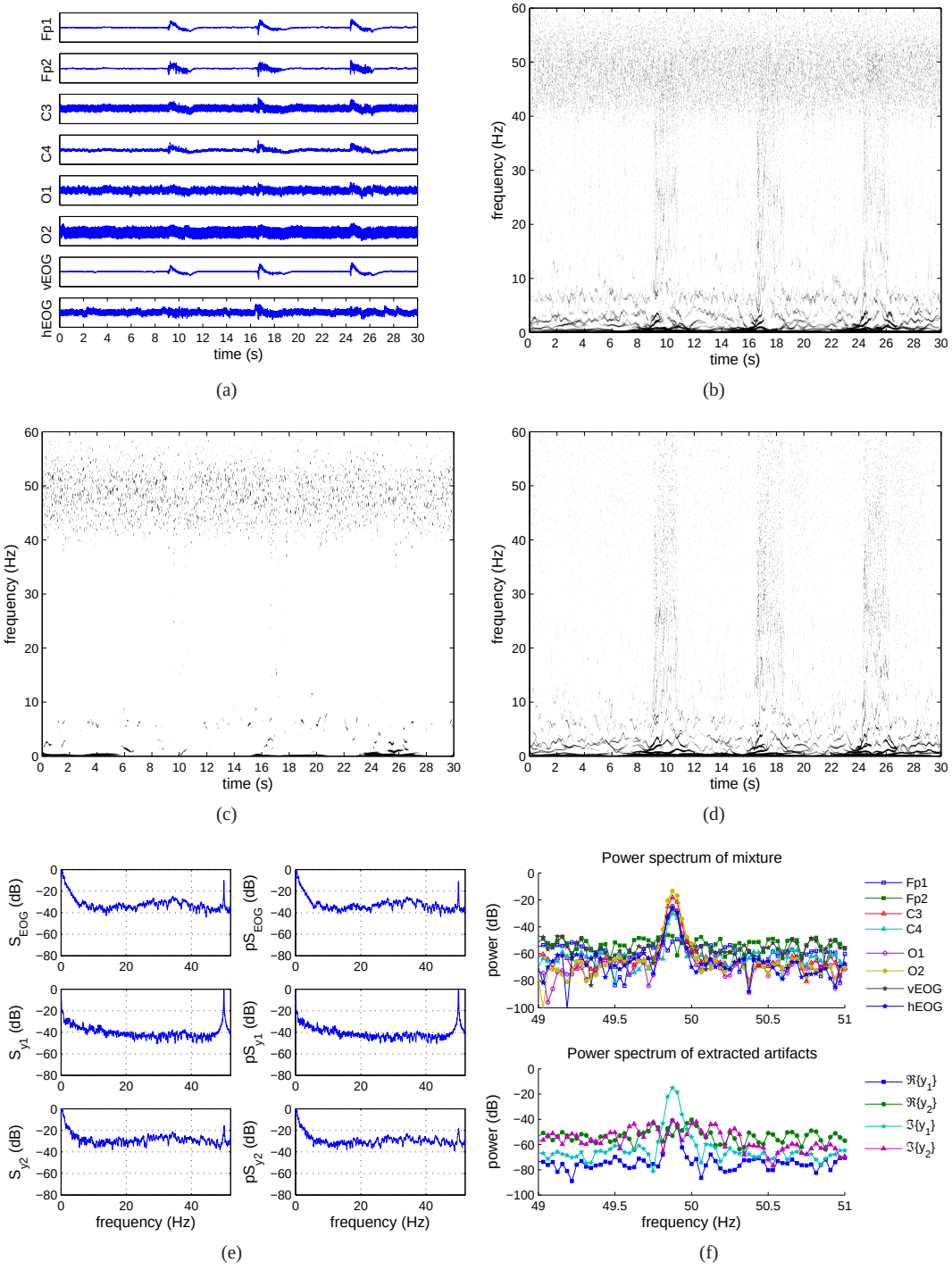


Fig. 12. Recorded and extracted artifacts from the 'EYEBROW' set. (a) Recorded EEG signals from the 'EYEBROW' set. (b) The Hilbert-Huang time-frequency plot of the recorded EEG signals. (c) The Hilbert-Huang time-frequency plot of the extracted line noise $\Im\{y_1(k)\}$. (d) The Hilbert-Huang time-frequency plot of the extracted EMG $\Re\{y_2(k)\}, \Im\{y_2(k)\}$. (e) The power spectra (S) and pseudo-spectra (pS) of the recorded EMG, and the extracted signals $y_1(k)$ and $y_2(k)$. (f) Frequency components of the recorded EEG signals and the extracted artifacts around the 50Hz frequency range. After extraction, the power line noise is contained in $\Im\{y_1\}$.

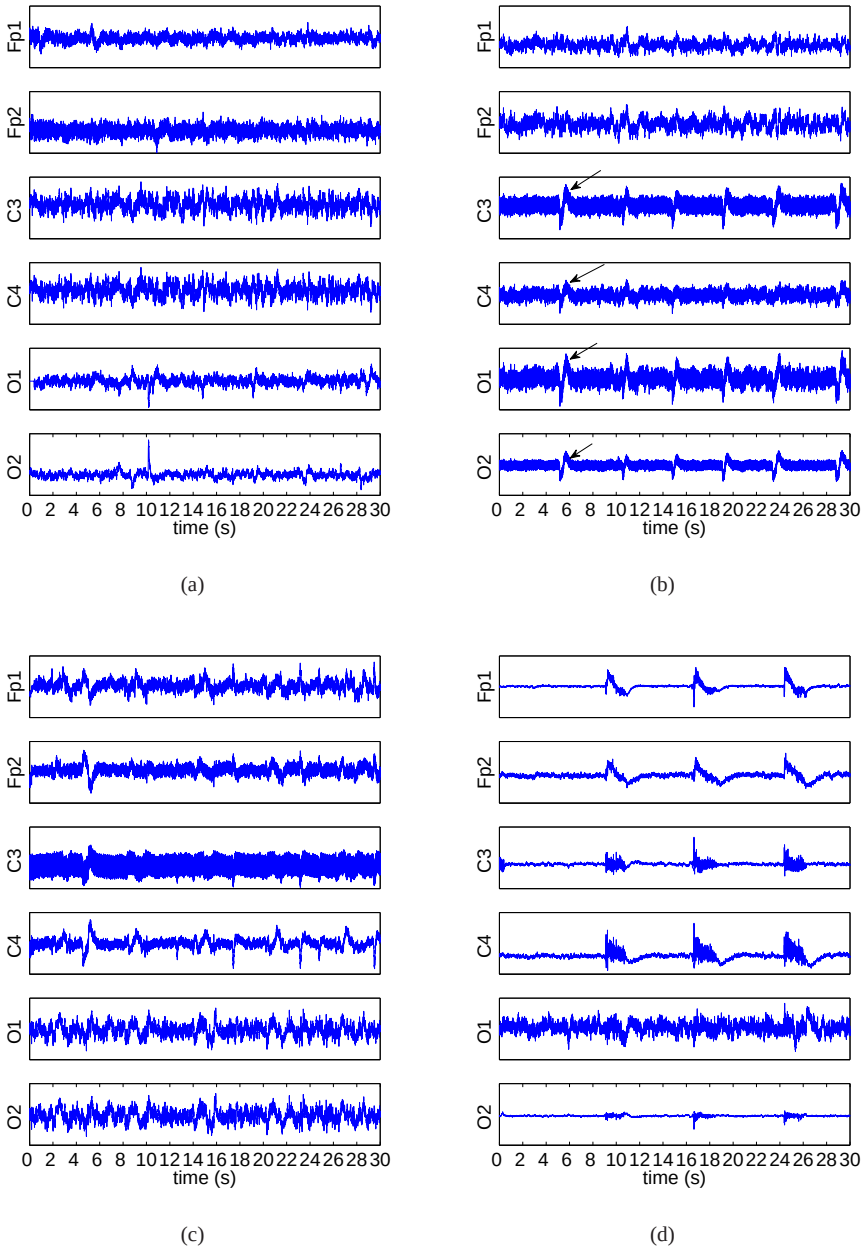


Fig. 13. EEG after extracting artifacts from the 'EYEBLINK', 'EYEROLL' and 'EYEBROW' set. (a) EYEBLINK: Kurtosis-based method. (b) EYEBLINK: Predictor-based method in (Javidi et al., 2010). (c) EYEROLL: Kurtosis-based method. (d) EYEBROW: Kurtosis-based method.

As shown in Fig. 12(c) and Fig. 12(d), the algorithm successfully extracted the power line noise as the imaginary component of the first extracted signal $y_1(k)$ and the EMG signal as the real and imaginary components of the second extracted signal $y_2(k)$. From the T-F plot of $y_2(k)$ in Fig. 12(d), the complete EMG frequency component range was successfully extracted, with power line noise frequency components not present.

Considering the power spectra $\mathcal{S}_{EMG}$ and pseudo-power spectra $\breve{\mathcal{S}}_{EMG}$ in Fig. 12(e), the spectral distribution of the power and pseudo-power spectral density were strong in the (0-10)Hz range with an amplitude of around -10dB and in the (20-40)Hz range, though having a much lower value. In addition, a single spike at 50Hz of amplitude -10dB indicates the presence of power line noise. After the extraction, the power line noise was contained in the spectra of the $y_1(k)$ while the (0-10)Hz and (20-40)Hz frequency components were present in the PSD and pPSD of $y_2(k)$. For the 'EYEBROW' set, the spectra correlation coefficients between $\mathcal{S}_{EMG}$ and $\breve{\mathcal{S}}_{EMG}$ and those of $y_1(k)$ and $y_2(k)$ were respectively

$\{0.1287, 0.1078\}$ and $\{0.7593, 0.7906\}$. Also, the 50Hz frequency range for the contaminated mixture and the extracted artifacts are shown in Fig. 12(f). It can be seen that after the extraction procedure, the 50Hz component is contained in $\Im\{y_1(k)\}$, while in comparison to the EOG and eye muscle extracted components from the 'EYEBLINK' and 'EYEROLL' studies (see Fig. 10(f) and Fig. 11(f)), components $\Re\{y_2(k)\}$ and $\Im\{y_2(k)\}$ had a higher power level in this range, reflecting the wider frequency range of the EMG artifact.

3.3.6 EEGs after EOG and 50 Hz power line artifacts removal. Fig. 13 shows the EEG waveforms after the extraction of 'EYEBLINK', 'EYEROLL', 'EYEBROW' and 50 Hz noise artifacts. The top two plots compare the proposed method with the widely linear prediction based one in (Javidi et al., 2010). Notice for the first two EEG electrodes Fp1 and Fp2, the predictor-based technique in (Javidi et al., 2010) performed well, with the successful removal of the EYEBLINK artifact. However, it performed poorly in terms of the 50Hz noise removal, which caused the 'EYEBLINK' artifact to be present (but attenuated) in the remaining EEG electrodes. Comparing Fig. 13(c) with Fig. 11(a), it is clear that the 'EYEROLL' artifact was either heavily attenuated or removed; whereas comparing Fig. 13(d) with 12(a) demonstrates that it is quite challenging to remove completely the 'EYEBROW' artifact; however, the 50 Hz noise has been removed almost perfectly, as illustrated by comparing the bottom plots of Figs. 13(d) and 12(a).

4 DISCUSSION

Both qualitative and quantitative metrics have showed that the kurtosis-based extraction method yields enhanced results for real-time extraction of artifacts. Excellent results were obtained for the removal of eye blink, eye roll and power line artifacts. Although artifacts arising from eye rolling and raising the eye brow might seem similar to that of an eye blink, it is much more challenging to perform their complete removal in the context of real-time EEG processing, as they involve longer firing of larger groups of muscles. These are critical cases, as the EMG source goes into saturation; and to our knowledge, these artifacts have not been considered before in the literature. These results are promising, as our technique operates real-time, in contrast to methods such as in (Vigário, 1997; Jung et al., 2000; Delorme et al., 2001; Barbati et al., 2004; Greco et al., 2005; Delorme et al., 2007; Kumar et al., 2009). The advantage of the proposed kurtosis based method as compared to our previous method (Javidi et al., 2010) is also in that the proposed method allows us to select a particular artifact to be extracted. For instance, if we wish only the EOG artifact such as eye-blink to be removed, the parameter β in (22) can be set to unity; whereas in (Javidi et al., 2010), we do not have full control over which artifact is going to be extracted.

5 CONCLUSIONS

Blind source extraction of the generality of complex-valued signals based on the degree of non-Gaussianity and from noisy mixtures has been addressed. A cost function based on the normalised kurtosis has been utilised to perform blind extraction, and the corresponding online algorithm (K-cBSE) has been derived. The existence and uniqueness of the solutions have been discussed and variable step-size variants of to the algorithm have been addressed. It has been shown that the algorithm is robust to the degree of noncircularity of the additive noise and the success of the algorithm over increasing noise levels has been demonstrated. Simulations in noise-free and noisy environments illustrate the successful performance of the algorithm in the extraction of both circular and non-circular signals, while the extraction of EOG and EMG artifacts from recorded EEG signals in real-time demonstrate a practical application for the proposed methodology.

APPENDIX: UPDATE OF $\epsilon(k)$ FOR THE GNGD-TYPE COMPLEX BSE

The gradient descent update for the regularisation parameter $\epsilon(k)$ is written as

$$\epsilon(k+1) = \epsilon(k) - \rho \nabla_\epsilon \mathcal{J}\big|_{\epsilon=\epsilon(k-1)}$$

and the gradient derived as follows. Defining the adaptive step-size in (25) as

$$\upsilon(k) \triangleq \frac{\mu}{|\phi\big(y(k)\big)|^2 \cdot \|\mathbf{x}(k)\|_2^2 + \epsilon(k)}$$

the gradient $\nabla_\epsilon \mathcal{J}$ is given by

$$\nabla_\epsilon \mathcal{J} = \big(\nabla_{\mathbf{w}^*} \mathcal{J}\big)^T \cdot \frac{\partial \mathbf{w}^*(k)}{\partial \upsilon(k-1)} \cdot \frac{\partial \upsilon(k-1)}{\partial \epsilon(k-1)} \tag{33}$$

where

$$\begin{aligned}\frac{\partial \mathbf{w}^*(k)}{\partial \upsilon(k-1)} &= \frac{\partial \mathbf{w}^*(k)}{\partial \upsilon(k-1)} - \phi^*\big(y(k-1)\big)\mathbf{x}^*(k-1) - \\ &\quad \frac{\partial \phi^*\big(y(k-1)\big)}{\partial \upsilon(k-1)} \upsilon(k-1)\mathbf{x}^*(k-1) \\ &\approx -\phi^*\big(y(k-1)\big)\mathbf{x}^*(k-1)\end{aligned}$$

and only the driving term of the recursion is considered, and

$$\frac{\partial \upsilon(k-1)}{\partial \epsilon(k-1)} = \frac{-\mu}{\big[|\phi\big(y(k-1)\big)|^2 \cdot \|\mathbf{x}(k-1)\|_2^2 + \epsilon(k-1)\big]^2}.$$

While the derivative in (33) is calculated according to the $\mathbb{CR}$ calculus, $\epsilon(k)$ is real-valued and so only the real component of the $\mathbb{R}^*$–derivative in (33) is required. This leads to the update equation given in (26).

REFERENCES

Amari, S., Douglas, S. C., Cichocki, A., Yang, H. H., July 1997. Novel on-line adaptive learning algorithms for blind deconvolution using the natural gradient approach. In: Proc. 11th IFAC Symp. System Identification. Vol. 3. pp. 1057–1062.

Anemüller, J., Sejnowski, T. J., Makeig, S., Nov. 2003. Complex independent component analysis of frequency-domain electroencephalographic data. Neural Networks 16 (9), 1311–1323.

Ang, W.-P., Farhang-Boroujeny, B., 2001. A new class of gradient adaptive step-size LMS algorithms. IEEE Transactions on Signal Processing 49 (4), 805–810.

Barbati, G., Porcaro, C., Zappasodi, F., Rossini, P. M., Tecchio, F., 2004. Optimization of an independent component analysis approach for artifact identification and removal in magnetoencephalographic signals. Clinical Neurophysiology 115 (5), 1220–1232.

Barros, A. K., Cichocki, A., 2001. Extraction of specific signals with temporal structure. Neural Computation 13 (9), 1995–2003.

Bell, A. J., Sejnowski, T. J., 1995. An information-maximisation approach to blind separation and blind deconvolution. Neural Computation 7, 1129–1159.

Bingham, E., Hyvärinen, A., 2000. A fast fixed point algorithm for independent component analysis of complex valued signals. Journal of Neural Systems 10, 1–8.

Brandwood, D. H., Feb. 1983. A complex gradient operator and its application in adaptive array theory. IEE Proceedings F: Communications, Radar and Signal Processing 130 (1), 11–16.

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

Cichocki, A., Thawonmas, R., Amari, S., 1997. Sequential blind signal extraction in order specified by stochastic properties. Electronics Letters 33, 64–65.

Delorme, A., Makeig, S., Sejnowski, T., 2001. Automatic artifact rejection for EEG data using high-order statistics and independent component analysis. In: International Workshop on ICA. pp. 457–462.

Delorme, A., Sejnowski, T., Makeig, S., 2007. Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. NeuroImage 34 (4), 1443–1449.

Douglas, S. C., 2005. Fixed-point FastICA algorithms for the blind separation of complex-valued signal mixtures. In: Conference Record of the Thirty-Ninth Asilomar Conference on Signals, Systems and Computers. pp. 1320–1325.

Erdogan, A., June 2009. On the Convergence of ICA Algorithms With Symmetric Orthogonalization. IEEE Transactions on Signal Processing 57 (6), 2209–2221.

Eriksson, J., Koivunen, V., 2006. Complex random vectors and ICA models: Identifiability, uniqueness, and separability. IEEE Transactions on Information Theory 52 (3), 1017–1029.

Greco, A., Mammone, N., Morabito, F. C., Versaci, M., 2005. Semi-automatic artifact rejection procedure based on kurtosis, Renyi's entropy and independent component scalp maps. In: International Enformatika Conference. pp. 22–26.

Hayes, M. H., 1996. Statistical Digital Signal Processing and Modeling. Wiley.

Huang, N. E., Shen, S. S., 2005. Hilbert-Huang transform and its applications. World Scientific.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., Liu, H. H., Mar. 1998. The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis. Proceedings of the Royal Society of London. Series A 454 (1971), 903–995.

Hyvärinen, A., Karhunen, J., Oja, E., 2001. Independent Component Analysis. Wiley.

Hyvärinen, A., Oja, E., 1997. A fast fixed-point algorithm for independent component analysis. Neural Computation 9 (7), 1483–1492.

Javidi, S., Jelfs, B., Mandic, D. P., 2009. Blind extraction of noncircular signals using a widely linear predictor. In: IEEE Workshop on Statistical Signal Processing. pp. 501–504.

Javidi, S., Mandic, D. P., Cichocki, A., 2010. Complex blind source extraction from noisy mixtures using second order statistics. IEEE Transactions on Circuits and System I 57 (7), 1404 – 1416.

Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., Mckeown, M. J., Iragui, V., Sejnowski, T. J., 2000. Removing electroencephalographic artifacts by blind source separation. Psychophysiology 37 (02), 163–178.

Kreutz-Delgado, K., 2006. The complex gradient operator and the $\mathbb{CR}$-calculus. Dept. of Electrical and Computer Engineering, UC San Diego, Course Lecture Supplement No. ECE275A, 1–74. URL `http://arxiv.org/abs/0906.4835v1`

Kumar, P. S., Arumuganathan, R., Sivakumar, K., Vimal, C., 2009. An adaptive method to remove ocular artifacts from EEG signals using wavelet transform. Journal of Applied Sciences Research 5, 711–745.

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

Leong, W. Y., Mandic, D. P., Oct. 2008. Post-nonlinear blind extraction in the presence of ill-conditioned mixing. IEEE Transactions on Circuits and Systems I 55, 2631–2638.

Liu, W., Mandic, D. P., 2006. A normalised kurtosis-based algorithm for blind source extraction from noisy measurements. Signal Processing 86 (7), 1580–1585.

Liu, W., Mandic, D. P., Cichocki, A., Sept 2006a. Blind second-order source extraction of instantaneous noisy mixtures. IEEE Transactions of Circuits and Systems II 53 (9), 931–935.

Liu, W., Mandic, D. P., Cichocki, A., 2006b. Blind source extraction of instantaneous noisy mixtures using a linear predictor. In: Proc. IEEE International Symposium on Circuits and Systems. pp. 4199–4202.

Mandic, D. P., 2004. A generalized normalized gradient descent algorithm. IEEE Signal Processing Letters 11 (2), 115–118.

Mandic, D. P., Cichocki, A., April 2003. An online algorithm for blind extraction of sources with different dynamical structures. In: 4th Internation Symposium of Independent Component Analysis and Blind Signal Separation (ICA 2003). pp. 645–650.

Mandic, D. P., Goh, S. L., 2009. Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models. Wiley.

Neeser, F., Massey, J., 1993. Proper complex random processes with applications to information theory. IEEE Transactions on Information Theory 39 (4), 1293–1302.

Novey, M., Adalı, T., 2008. On extending the complex FastICA algorithm to noncircular sources. IEEE Transactions on Signal Processing 56 (5), 2148–2154.

Ollila, E., 2008. On the circularity of a complex random variable. IEEE Signal Processing Letters 15, 841–844.

Ollila, E., Koivunen, V., 2009a. Adjusting the generalized likelihood ratio test of circularity robust to non-normality. In: IEEE 10th Workshop on Signal Processing Advances in Wireless Communications. pp. 558–562.

Ollila, E., Koivunen, V., 2009b. Complex ICA using generalized uncorrelating transform. Signal Processing 89 (4), 365 – 377.

Palmer, J. A., Makeig, S., Kreutz-Delgado, K., 2009. A complex cross-spectral distribution model using normal variance mean mixtures. In: IEEE International Conference on Acoustics, Speech and Signal Processing. pp. 3569–3572.

Picinbono, B., 1994. On circularity. IEEE Transactions on Signal Processing 42 (12), 3473–3482.

Picinbono, B., Bondon, P., 1997. Second-order statistics of complex signals. IEEE Transactions on Signal Processing 45 (2), 411–420.

Picinbono, B., Chevalier, P., 1995. Widely linear estimation with complex data. IEEE Transactions on Signal Processing 43 (8), 2030–2033.

Rehman, N., Mandic, D. P., 2010. Multivariate empirical mode decomposition. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science 466, 1291–1302.

Sanei, S., Chambers, J. A., 2007. EEG signal processing. John Wiley & Sons Ltd.

Särelä, J., Valpola, H., 2005. Denoising source separation. The Journal of Machine Learning Research 6, 233–272.

Schreier, P., Scharf, L., 2003. Second-order analysis of improper complex random vectors and processes. IEEE Transactions on Signal Processing 51 (3), 714–725.

Thawonmas, R., Cichocki, A., Amari, S., 1998. A cascade neural network for blind signal extraction without spurious equilibria. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences 81 (9), 1833–1846.

van den Bos, A., 1994. Complex gradient and Hessian. IEE Proceedings of Vision, Image and Signal Processing 141 (6), 380–383.

Vigário, R. N., 1997. Extraction of ocular artefacts from EEG using independent component analysis. Electroencephalography and Clinical Neurophysiology 103 (3), 395–404.