Kurtosis Based Blind Source Extraction of Complex Noncircular Signals with Application in EEG Artifact Removal

Soroush Javidi and **Danilo P. Mandic**
Communications and Signal Processing Research Group,
Department of Electrical and Electronic Engineering,
Imperial College London, London, SW7 2AZ, U.K.
email: {soroush.javidi, d.mandic}@imperial.ac.uk
URL: http://www.commsp.ee.ic.ac.uk/~mandic

Abstract

An online blind extraction algorithm, suitable for the generality of complex-valued sources, both complex circular and noncircular, is introduced. This is achieved based on higher order statistics of latent sources, and using the deflation approach. The novelty of the proposed approach is that the cost function is designed so as to be robust to both circular and noncircular second order additive noise. The analysis is supported by simulations in both noise-free and noisy scenarios, and by practical example of conditioning of EEG signals for the removal of artifacts.

I. INTRODUCTION

Blind source separation (BSS) is the paradigm whereby latent sources are recovered from an observed mixture. It aims to reconstruct the original sources by identifying the inverse mixing process without explicit knowledge of the mixing process or the sources [1], and has found application in diverse areas including biomedical engineering, communications, sonar and radar [2], [3]. Standard BSS methods use cost functions based on second- and higher-order statistics and maximisation of likelihood and entropy [4]–[6]. To better facilitate the modelling of real-world systems, noisy environments and post-nonlinear mixtures have been recently studied in real domain algorithms [7], [8].

Within the BSS methodology, the 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 latent sources based on a fundamental signal property (nonlinearity, sparsity), effectively inducing an order in the separation process. The benefit of BSE becomes apparent in large-scale problems where only a small subset of the sources are required from a large-scale mixture of latent sources, making it possible to extract only the sources of interest. This also reduces the computational complexity and the requirement for pre- or post-processing of the mixture or separated sources that may be necessary if BSS techniques are employed.

Real domain algorithms performing extraction based on the temporal structure (predictability) of signals were studied in [1], [9], [10] and modifications to the cost function were proposed to cater for noisy mixtures [11], [12]. Similarly, the algorithm in [13] demonstrated the feasibility of extraction of real-valued sources based on the degree of kurtosis, while [14] proposed a modified cost function for the extraction based in noisy environments. An overview and discussion on these class of algorithms is also provided in [15].

Recent developments in complex statistics [16], [17] have resulted in the introduction of a new hierarchy of complex domain signal processing algorithms, capable of catering for the generality of complex signals [3]. This is achieved through the consideration of the circular symmetry of the distribution, where rotational invariance of the distribution signifies a complex *circular* random variable. However, most complex-valued signals encountered in signal processing application are *noncircular*. Using the paradigm of augmented statistics, it is possible to consider the complete second-order information available in a complex-valued random variable. In this manner, the second-order statistics are not only based the standard covariance matrix $E\{\mathbf{x}\mathbf{x}^H\}$, but the pseudo-covariance $E\{\mathbf{x}\mathbf{x}^T\}$ as well. A complex-valued random vector with a vanishing pseudo-covariance is terms *proper* which signifies second-order circularity, and otherwise is considered *improper* [18], [19]. In this light, it is possible to design complex-valued signals through convenience of representation, while providing a full complex second-order statistical model of it [2]. Likewise, widely linear models [20] in supervised adaptive signal processing allow for the design of algorithms capable of processing both complex circular and noncircular signals.

In complex-valued blind source separation research, recent algorithms have been designed based on augmented statistics principles so as to cater for the generality of complex signals [21], [22], with applications in fMRI and communications [23]. In comparison to standard complex BSS methodology [24], [25], which assume complex circular sources, these algorithms have been shown to exhibit enhanced performance. Similarly, the feasibility of blind source extraction of complex sources based on the temporal structure of the latent sources was studied in [26] where it was shown that by using a widely linear predictor, it is possible to extract both complex circular and noncircular sources. The study concluded that a WL predictor has

a much better and consistent performance to a linear predictor in blind extraction of complex signals. A class of algorithms for blind extraction from noisy complex-valued mixtures, based on predictability has also been recently proposed, demonstrating successful extraction of signals from both complex circular and noncircular white noise [27].

In this report, we propose a new class of complex BSE algorithms designed for the extraction based on the degree of kurtosis, and in the presence of complex-valued additive noise. This provides an extension of the methodology presented in [14] to the generality of complex signals, both complex circular and noncircular. A modified cost function, based on the concept of kurtosis in the complex domain, is proposed so as to cater for blind extraction from general complex-valued noisy mixtures. Extensive studies of the extraction of various artifacts from electroencephalograph (EEG) signals demonstrate the application of the algorithm, with the study of the artifact extraction performance aided by several qualitative and quantitative metrics.

This report is organised as follows. Section II provides an overview of complex statistics, complex-valued noise and $\mathbb{CR}$ calculus. In Section III, the proposed cost function is given, its convergence considered and the resulting online algorithm introduced. Section IV illustrates the performance of the proposed algorithm in blind extraction of both synthetic and real-world sources, including communication signals and the EEG for brain computer interfacing. Conclusions are presented in Section VI.

II. Circularity, Complex Noise and $\mathbb{CR}$ Calculus

A. Complex statistics: second-order circularity

Second-order circularity is based on the general concept of complex circularity, where the distribution of a complex random variable $\mathbf{z}$ and its rotation $e^{j\varphi}\mathbf{z}$ are equal for any φ and circularly symmetric [16]. In defining the second-order statistics, the circular symmetry of a complex random variable is considered by using both the covariance $\mathcal{C}_{\mathbf{zz}}$ and pseudo-covariance $\mathcal{P}_{\mathbf{zz}}$ [18]

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

where the covariance matrix is similar to that defined in the real domain and the pseudo-covariance provides more depth in describing the relationship between the real and imaginary components. For second-order circular (also termed proper) random variables, the pseudo-covariance matrix vanishes, $\mathcal{P}_{\mathbf{zz}} = \mathbf{0}$, whereas for second-order noncircular (improper) random variables the pseudo-covariance is non-zero, $\mathcal{P}_{\mathbf{zz}} \neq \mathbf{0}$ and is usually complex-valued. The pseudo-covariance $\mathcal{P}_{\mathbf{zz}}$ can be written in terms of the covariance of its real and imaginary components, given as

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

where the vanishing pseudo-covariance is owing to an equal covariance in the real and imaginary components, while the cross-covariance is skew-symmetric, as shown in [17]. Thus

$$E\{\mathbf{z}_r\mathbf{z}_r^T\} = E\{\mathbf{z}_i\mathbf{z}_i^T\}, \quad E\{\mathbf{z}_i\mathbf{z}_r^T\} = -\big(E\{\mathbf{z}_i\mathbf{z}_r^T\}\big)^T \tag{2}$$

and while only the diagonal elements of the cross-covariance are required to be zero (that is, uncorrelated real and imaginary components $z_{r,k}$ and $z_{i,k}$), for a white random variable, the cross-covariance matrix is completely zero.

A complex random vector is considered uncorrelated if both the covariance and pseudo-covariance matrices are diagonal [28], where the covariance and pseudo-covariance values are respectively denoted by σ_k^2 and τ_k^2 [29]. Examples of complex circular signals encountered in signal processing research are QPSK and BPSK signals in communications, while most real-world complex signals are noncircular such as EEG signals, and nonstationary signals such as wind [3].

Consider the second-order stationary augmented complex random signal $\mathbf{z}^a(k) = [\mathbf{z}(k), \mathbf{z}^*(k)]^T$ which can be used to define the augmented covariance matrix, given by

$$\mathcal{C}_{\mathbf{z}^a\mathbf{z}^a}(\delta) = E\begin{bmatrix} \mathbf{z}(k) \\ \mathbf{z}^*(k) \end{bmatrix}\begin{bmatrix} \mathbf{z}^H(k-\delta), \mathbf{z}^T(k-\delta) \end{bmatrix}$$
$$= \begin{bmatrix} \mathcal{C}_{\mathbf{zz}}(\delta) & \mathcal{P}_{\mathbf{zz}}(\delta) \\ \mathcal{P}_{\mathbf{zz}}^*(\delta) & \mathcal{C}_{\mathbf{zz}}^*(\delta) \end{bmatrix}, \tag{3}$$

providing a complete description of the second-order statistics of $\mathbf{z}(k)$. The transformation of this matrix to the frequency domain is used to define the augmented spectral matrix [18], [19]

$$\mathcal{S}_{\mathbf{z}^a}(\omega) = \begin{bmatrix} \mathcal{S}_{\mathbf{z}}(\omega) & \check{\mathcal{S}}_{\mathbf{z}}(\omega) \\ \check{\mathcal{S}}_{\mathbf{z}}^*(-\omega) & \mathcal{S}_{\mathbf{z}}(-\omega) \end{bmatrix}, \tag{4}$$

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

$$\mathcal{S}_{\mathbf{z}}(\omega) = \mathcal{F}\big(\mathcal{C}_{\mathbf{zz}}(\delta)\big) = \mathcal{F}\big(E\{\mathbf{z}(k)\mathbf{z}^H(k-\delta)\}\big)$$

$$\breve{S}_{\mathbf{z}}(\omega) = \mathcal{F}\big(\mathcal{P}_{\mathbf{zz}}(\delta)\big) = \mathcal{F}\big(E\{\mathbf{z}(k)\mathbf{z}^T(k-\delta)\}\big) \tag{5}$$

and δ is a discrete time lag.

The matrix $\mathcal{S}_{\mathbf{z}}(\omega)$ denotes the power spectral density (also power spectrum, or PSD) of the random variable $\mathbf{z}$, whereas $\breve{S}_{\mathbf{z}}(\omega)$ can be interpreted as its pseudo-power spectral density (also pseudo-spectrum, or pPSD). While the power spectrum provides information on the distribution of power over a frequency range, the magnitude of the pseudo-spectrum characterises the second-order circularity of the random variable in the frequency domain. Thus, as with the consideration of both the covariance and pseudo-covariance matrices in second-order time-domain analysis, it is acceptable to consider both the spectrum and pseudo-spectrum in frequency-domain analysis. Finally, the augmented spectral matrix in (4) is nonnegative definite which results in the condition [18]

$$|\breve{S}_{\mathbf{z}}(\omega)|^2 \leq S_{\mathbf{z}}(\omega) \cdot S_{\mathbf{z}}(-\omega). \tag{6}$$

B. Complex statistics: Kurtosis

Kurtosis, is a well understood concept in statistics of real-valued random variables, and has been used to design contrast functions in BSS, such as in the FastICA [30], and BSE algorithms [13]. It is common to use the normalised kurtosis $K_R(\cdot)$ instead of the standard kurtosis $\text{kurt}_R(\cdot)$ as it allows for the comparison of the Gaussianity of random variables, irrespective of the range of amplitudes. In [31], the extension and relevance of this concept to the complex domain, as well as the relation between the kurtosis of the real and imaginary components of a complex random variable, $\text{kurt}_R(z_r)$ and $\text{kurt}_R(z_i)$, and the kurtosis of the complex random variable $\text{kurt}_c(z)$ has been discussed.

The real-valued normalised kurtosis of a complex random variable can be defined in several forms, where

$$
\begin{aligned}
K_c(z) &= \frac{\text{kurt}_c(z)}{(E\{|z|^2\})^2} \\
&= \frac{E\{|z|^4\}}{(E\{|z|^2\})^2} - \frac{|E\{z^2\}|^2}{(E\{|z|^2\})^2} - 2
\end{aligned} \tag{7}
$$

is the most common definition, with

$$\text{kurt}_c(z) = E\{|z|^4\} - |E\{z^2\}|^2 - 2(E\{|z|^2\})^2. \tag{8}$$

The first term in (7) is the normalised fourth order moment whereas the second term is the square of the circularity coefficient [32], and $\text{kurt}_c(z)$ in (8) 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, in this measure, kurtosis values of a sub-Gaussian complex random variable is negative and that of a super-Gaussian complex random variable is positive, irrespective of the circularity/noncircularity of the random variable.

C. Complex-valued Noise

For the purpose of analysis, we will make use of the circularity measure r, defined as in [32] as the magnitude of the circularity quotient $\rho(z) = re^{j\theta} \triangleq \tau_z^2/\sigma_z^2$

$$r = |\rho(z)| = \frac{|\tau_z^2|}{\sigma_z^2}, \quad r \in [0, 1] \tag{9}$$

where r measures the degree of noncircularity in the complex signal, and for completeness, the circularity angle is defined as $\theta = \arg(\rho(z))$ which can be utilised to indicate the orientation of the distribution. Note that for a purely circular signal, the measure $r = 0$, while the value of θ does not provide additional information about its 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 $\theta = 2\alpha$ [32, 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 $r = 1$, the ellipse becomes elongated with a maximal major axis and minor axis of length zero. Also note that the pseudo-variance of a general complex Gaussian distribution is then related to the elliptic shape by $\tau^2 = \epsilon^2 e^{j2\theta}$ [29].

The treatment of a noise signal $\mathbf{v}(k)$ in $\mathbb{C}$ is different to that in the real domain [18]. While in $\mathbb{R}$ only the variance σ_v^2 of the noise signal is of concern, in $\mathbb{C}$ it is necessary to consider the pseudo-variance τ_v^2 as well in order to completely model the noise. We therefore have the following cases.

- *Circular white noise*, where the noise is considered white in terms of covariance matrix, while the pseudo-covariance matrix vanishes. This is demonstrated by

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

where $\mathbf{I}$ denotes the identity matrix.

In the frequency domain, the power power spectrum $S_{\mathbf{v}}(\omega)$ of the circular white noise contains equal power across all
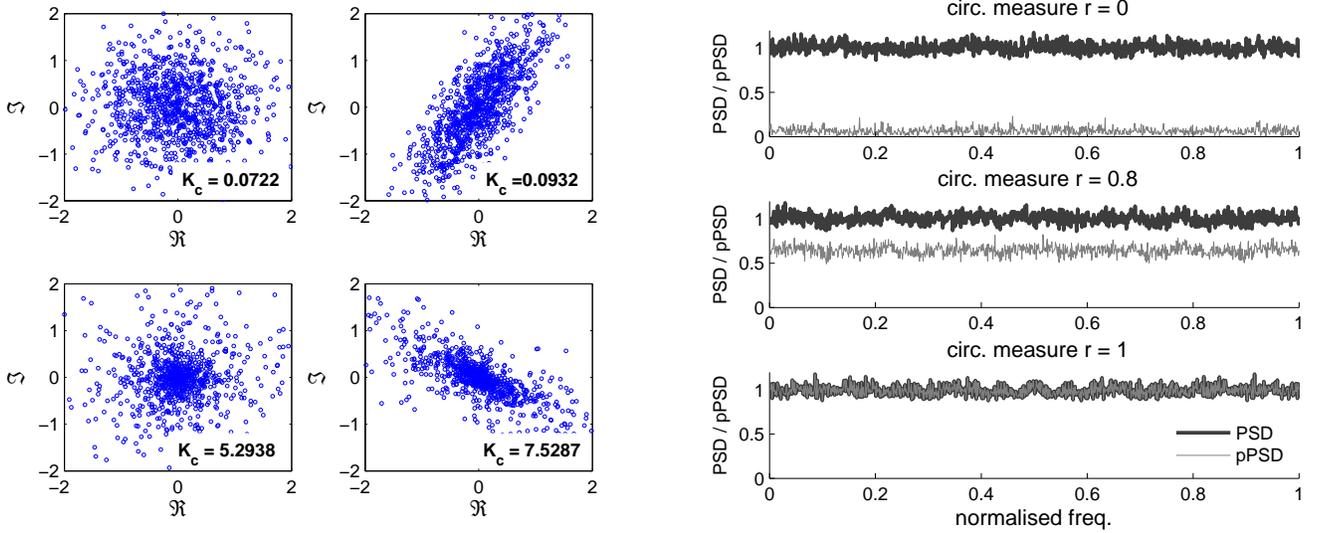
(a) Scatter plots of complex white noise realisations. *Top row:* circular Gaussian (b) Power spectrum (thick gray line) and pseudo-power spectrum (thin gray line) noise (left) and noncircular Gaussian noise ($r = 0.81$) (right). *Bottom row:* of complex Gaussian noise with varying degree of noncircularity $r = \{0, 0.8, 1\}$ circular Laplacian noise (left) and noncircular Laplacian noise ($r = 0.81$) (right). The circularity measure r is defined in (9). The kurtosis values K_c are given for each case.

Fig. 1.   Scatter plots and spectra of complex-valued noise

frequencies, while the pseudo-spectrum $\breve{\mathcal{S}}_\mathbf{v}(\omega)$ is zero.

- *Noncircular doubly white noise*, where whiteness is assumed for both the covariance and pseudo-covariance matrices, however, the distribution and power levels may be different in each, 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 of the white noise is uniform across all frequencies while the pseudo-spectrum is non-zero with its magnitude depending on the pseudo-variance. As the noise becomes more noncircular ($r \rightarrow 1$), the pseudo-spectrum approaches its upper-bound defined in (6), where for highly noncircular noise ($r \approx 1$), the magnitude of the pPSD and PSD are similar. For a scalar complex white noise signal $v(k)$, the relation between the correlation and pseudo-correlation and the repsective spectra are given as

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

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

Examples of circular white Gaussian and Laplacian noise with unit variance are illustrated in the left column of Fig. 1(a) while the right 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 $\tau_v^2 = -0.38 + \jmath0.71$, 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 $0.45 - \jmath0.66$. Also note that in Fig. 1(a) the value of kurtosis is approximately zero for both the circular and noncircular Gaussian noise signals, while the kurtosis values are different for the circular and noncircular super-Gaussian noise signals.

Fig. 1(b) depicts the PSD and pPSD of circular circular ($r = 0$) white and noncircular doubly white Gaussian noise for circularity measures $r = \{0.8, 1\}$. The pseudo-spectrum is zero for the circular noise, while it has a magnitude of 0.64 for the second noise realisation, and reaches it upper-bound of 1 in the third realisation where the noise is highly noncircular. For the Gaussian noise, the spectrum $\mathcal{S}(\omega) = 1$ and the pseudo-spectrum $\breve{\mathcal{S}}(\omega) = |\tau_v^2| = |\epsilon^2 e^{\jmath2\theta}| = |\epsilon^2| = r$, thus by increasing the eccentricity of the ellipse (degree of noncircularity), the magnitude of the pPSD approaches the its maximum value of 1.

D. $\mathbb{CR}$ calculus: Brief overview

The so called $\mathbb{CR}$ calculus (also known as Wirtinger calculus) [33] 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 in signal processing problems. Consider a typical cost function $F(\mathbf{z}) : \mathbb{C}^N \mapsto \mathbb{R}$, a real function of complex variables, which does not
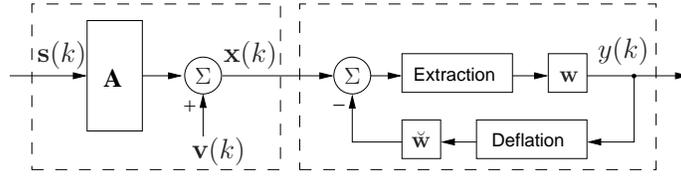
Fig. 2. The noisy mixture model, and BSE architecture.

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, as

$$F(\mathbf{z}, \mathbf{z}^*) : \mathbb{C}^N \times \mathbb{C}^N \mapsto \mathbb{R}. \tag{10}$$

Note that although $\mathbf{z}$ and $\mathbf{z}^*$ are not statistically independent, this does not affect the calculation of derivatives, defined respectively as [33]

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

$$\mathbb{R}^*\text{–derivative}: \quad \left.\frac{\partial F}{\partial \mathbf{z}^*}\right|_{\mathbf{z}=const} = \frac{1}{2}\left(\frac{\partial F}{\partial \mathbf{z}_r} + \jmath\frac{\partial F}{\partial \mathbf{z}_i}\right). \tag{11}$$

Also note that the direction of steepest descent is given by the derivative with respect to $\mathbf{z}^*$, the $\mathbb{R}^*$–derivative. Using the Taylor Series Expansion of F [34], the magnitude of a small change in the function F is given by

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

where using the Cauchy-Schwarz Inequality it is shown 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 $\delta\mathbf{z}$ is in the direction of $\frac{\partial F}{\partial \mathbf{z}^*}$ [33], [35].

Furthermore, in calculating derivatives of analytic functions, as expected, the $\mathbb{R}^*$–derivative vanishes and only the standard derivative is present, demonstrating the flexibility of the framework. This can be illustrated through a simple example. Consider the widely used and non-analytic error power 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 and comprehensive analysis and insight into $\mathbb{CR}$ calculus, we refer to the material in [3], [33].

III. BSE of Complex Noisy Mixtures

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

$$\mathbf{x}(k) = \mathbf{A}\mathbf{s}(k) + \mathbf{v}(k) \tag{12}$$

and $\mathbf{s}(k) \in \mathbb{C}^{N_s}$ is the vector of latent sources, $\mathbf{A} \in C^{N \times N_s}$ is the mixing matrix and $\mathbf{v}(k) \in \mathbb{C}^N$ is the vector of additive doubly white Gaussian noise. 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 equal to that of the sources, however, in the case of noisy mixtures, an overdetermined mixture is necessary, to estimate the second-order statistics of noise parameters shown in the next section.

The adaptive gradient descent algorithm at the extraction stage adapts the parameters of the demixing vector $\mathbf{w}$ such that the source signal with the largest (smallest) kurtosis, denoted by $y(k)$, is first extracted where

$$y(k) = \mathbf{w}^H \mathbf{x}(k)$$
$$= \underbrace{\mathbf{w}^H \mathbf{A}}_{\triangleq \mathbf{u}^H} \mathbf{s}(k) + \mathbf{w}^H \mathbf{v}(k) \tag{13}$$

The variance of $y(k)$ can be written in an expanded form using (13), as

$$E\{|y(k)|^2\} = \mathbf{u}^H \mathcal{C}_{\mathbf{ss}}(0)\mathbf{u} + \mathbf{w}^H \mathcal{C}_{\mathbf{vv}}(0)\mathbf{w}$$

$$= \mathbf{u}^H\mathbf{u} + \sigma_v^2\mathbf{w}^H\mathbf{w} \tag{14}$$

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).

Similarly, the normalised kurtosis of $y(k)$ can be written as

$$K_c(y) = \sum_{n=1}^{N_s} K_c(u_n^* s_n) + \sum_{n=1}^{N} \underbrace{K_c(w_n^* v_n)}_{=0}$$

$$= \sum_{n=1}^{N_s} |u_n|^4 E\{|s_n|^4\} - 2|u_n|^4 (E\{|s_n|^2\})^2 - |u_n|^4 |E\{s_n^2\}|^2$$

$$= \sum_{n=1}^{N_s} |u_n|^4 K_c(s_n) \tag{15}$$

where the normalised kurtosis of the noise is zero. In a vectorised form, this is equivalent to

$$K_c(y) = \hat{\mathbf{u}}^H \mathbf{K}_c(\mathbf{s})\hat{\mathbf{u}} \tag{16}$$

where

$$\hat{\mathbf{u}} = [u_1^2, \ldots, u_{N_s}^2]$$
$$\mathbf{K}_c(\mathbf{s}) = diag\big(K_c(s_1), \ldots, K_c(s_{N_s})\big). \tag{17}$$

The next stage within BSE 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 $\breve{\mathbf{w}}$ is the deflation weight coefficient vector, updated using an adaptive gradient descent algorithm detailed later in this section. In essence, for $y(k)$ being an estimate of one of the original sources, say $s_n(k)$, the ideal deflation weight vector is equal to the nth column of the mixing matrix $\mathbf{A}$, such that the effect of the particular source is removed from the mixture. Finally, a threshold can be set on the deflation process, such that extraction is continued until some or all required source have been successfully extracted [36].

A. Cost function

The cost function proposed for the extraction of the generality of complex sources from a noisy mixture is given by

$$\mathcal{J}(\mathbf{w}) = -\beta \frac{\text{kurt}_c\big(y(k)\big)}{\big(E\{|y(k)|^2\} - \mathbf{w}^H\mathcal{C}_{\mathbf{vv}}(0)\mathbf{w}\big)^2} \tag{18}$$

where $\mathcal{J} \in \mathbb{R}$, and is based on a modified version of the normalised kurtosis defined in (7) and is a generalisation of the methodology presented in [14]. As illustrated in (14), the variance of $y(k)$ contains the noise variance σ_v^2 so as to remove the effect of noise from (18) 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, 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, as demonstrated in Section IV.

In the cost function (18), the parameter β dictates the order of extraction where for

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

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

$$\mathbf{w}_{opt} = \arg \max_{\|\mathbf{w}\|_2^2=1} \mathcal{J}(\mathbf{w}) \tag{19}$$

where the norm of the demixing vector is constrained to unity to avoid very small coefficient values.

Rewriting and simplifying (18) in terms of (14) and (17) results in

$$\mathcal{J}(\mathbf{w}) = -\frac{\hat{\mathbf{u}}^H |\mathbf{K}_c(\mathbf{s})|\hat{\mathbf{u}}}{(\mathbf{u}^H\mathbf{u})^2} = -\acute{\mathbf{u}}^H |\mathbf{K}_c(\mathbf{s})|\acute{\mathbf{u}} \tag{20}$$

where

$$\acute{\mathbf{u}}^H \triangleq \frac{\hat{\mathbf{u}}^H}{\mathbf{u}^H\mathbf{u}} = \frac{\hat{\mathbf{u}}^H}{\|\mathbf{u}\|_2^2}. \tag{21}$$

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 (20) 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 $\mathbf{w}_{opt} = \mathbf{A}^{H\#}\mathbf{u}_{opt}$, where the symbol $(\cdot)^\#$ denotes the matrix pseudo-inverse operation [14].

B. Adaptive algorithm for extraction

Optimisation of (18) 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 II-D, the gradient[1] is thus expressed as

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

where $\phi\big(y(k)\big)$ is used for simplification and $m_\ell(y)$ and $p_\ell(y)$ are respectively the ℓth moments and pseudo-moment at time instance k (the time index dropped), and they are estimated using the moving average estimates

$$m_\ell\big(y(k)\big) = (1-\alpha)m_\ell\big(y(k-1)\big) + \alpha|y(k)|^\ell, \ \ell = \{2, 4\}$$
$$p_\ell\big(y(k)\big) = (1-\alpha)p_\ell\big(y(k-1)\big) + \alpha\big(y(k)\big)^\ell, \ \ell = 2 \tag{22}$$

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 in expanded form as

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \frac{\mu\beta\Big[y^*(k)\big(m_4(y) - 2m_2^2(y) - |p_2(y)|^2\big) + \big(m_2(y) - \sigma_v^2\big)\big(-y(k)y^{*2}(k) + 2m_2(y)y^*(k) + p_2^*(y)y(k)\big)\Big]}{(m_2(y) - \sigma_v^2)^3}\mathbf{x}(k), \tag{23}$$

where μ is the small positive step-size. To preserve the vector magnitude to unit norm and avoid small small magnitudes, the demixing weights are then normalised at each iteration, that is

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

so as to preserve the vector magnitude.

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, while the noise variance σ_v^2 is utilised in the online algorithm. The estimation of the noise variance can be performed using a subspace method, as described in [37]. 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 consideration of the eigenvalues of the observation covariance matrix $\mathcal{C}_{\mathbf{xx}}$, where

$$\mathcal{C}_{\mathbf{xx}} = \mathbf{A}\mathcal{C}_{\mathbf{ss}}\mathbf{A}^H + \mathcal{C}_{\mathbf{vv}} = \boldsymbol{\Upsilon} + \sigma_v^2\mathbf{I}. \tag{24}$$

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

1) Modifications to the update algorithm: In order to enhance the performance of the online gradient descent algorithm, implementation of adaptive step-size update algorithms are considered, where the performance of these modifications are compared to the standard proposed algorithm and are detailed in the next section. The justification for the use of these modifications is that the empirical nature of selecting the step-size μ may result in inappropriate step-sizes as the algorithm converges to a local minima of the cost function and thus the algorithm in (23) should be modified by also updating the step-size μ at each iteration. We consider the complex-valued implementation of the Farhang-Ang type variable step-size (VSS)

---

[1] Since the normalised kurtosis $\mathcal{J}$ is real valued, the $\mathbb{R}^*$–derivative $\frac{\partial \mathcal{J}}{\partial \mathbf{w}^*}$ is utilised for the calculation of the gradient.

algorithm [38] and the complex-valued generalised normalised gradient descent (GNGD) type algorithm [39].

At each iteration k, the VSS algorithm minimises the cost function $\mathcal{J}$ in (18) with respect to $\mu(k-1)$ to provide the most optimum value of the step-size, given as

$$\mu(k) = \mu(k-1) - \eta \nabla_\mu \mathcal{J}\big|_{\mu=\mu(k-1)}$$

$$\nabla_\mu \mathcal{J} = \nabla_{\mathbf{w}^*} \mathcal{J} \cdot \frac{\partial \mathbf{w}^*(k)}{\partial \mu(k-1)}$$

$$\boldsymbol{\psi}(k) = \gamma \boldsymbol{\psi}(k-1) - \nabla_{\mathbf{w}^*} \mathcal{J}\big|_{\mathbf{w}^*=\mathbf{w}^*(k-1)} \tag{25}$$

where $\boldsymbol{\psi}(k) \triangleq \frac{\partial \mathbf{w}^*(k)}{\partial \mu(k-1)} \approx \frac{\partial \mathbf{w}^*(k)}{\partial \mu(k)}$ and η and γ are step-sizes.

The GNGD-type algorithm is based on a normalised variation of (23), which can be shown to be given by

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \frac{\mu}{|\phi(y(k))|^2 \cdot \|\mathbf{x}(k)\|_2^2 + \epsilon(k)} \phi(y(k)) \mathbf{x}(k) \tag{26}$$

where $\epsilon(k)$ is an adaptive regularisation parameter. In a standard normalised algorithm, the parameter $\epsilon(k)$ is a small positive constant to avoid instability due to a zero-normed input vector, where in comparison the GNGD methodology allows for faster convergence and greater stability by allowing a gradient adaptive regularisation parameter, given by

$$\epsilon(k+1) = \epsilon(k) - \rho\mu \frac{\Re\{\phi(y(k))\mathbf{x}^T(k)\phi^*(y(k-1))\mathbf{x}^*(k-1)\}}{\left[|\phi(y(k-1)) \cdot \|\mathbf{x}(k-1)\|_2^2 + \epsilon(k-1)\right]^2} \tag{27}$$

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

C. 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 using an adaptive gradient descent method similar to [36], where the cost function

$$\mathcal{J}_d(\breve{\mathbf{w}}) = \|\mathbf{x}_{n+1}(k)\|^2 = \mathbf{x}_{n+1}^H(k)\mathbf{x}_{n+1}(k) \tag{28}$$

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

$$\mathbf{x}_{n+1}(k) = \mathbf{x}_n(k) - \breve{\mathbf{w}}(k)y_n(k). \tag{29}$$

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

$$\nabla_{\breve{\mathbf{w}}^*} \mathcal{J}_d = \frac{\partial \mathcal{J}_d}{\partial \mathbf{x}_{n+1}^*} \cdot \frac{\partial \mathbf{x}_{n+1}^*}{\partial \breve{\mathbf{w}}^*} = -y_n^*(k)\mathbf{x}_{n+1}(k) \tag{30}$$

and the online algorithm for BSE then becomes

$$\breve{\mathbf{w}}(k+1) = \breve{\mathbf{w}}(k) + \mu_d y_n^*(k)\mathbf{x}_{n+1}(k) \tag{31}$$

and will be used to estimate $\mathbf{x}_{n+1}(k+1)$ using (29), with μ_d a step-size. The drawback to 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 $y_n(k)$.

As mentioned earlier, in the design of complex adaptive algorithms, it is common to utilise a WL model to ensure that the algorithm is capable of processing the generality of complex signals [3]. In the case of the update for the deflation weight coefficient (31), however, a linear model is considered as the original BSS mixing model (12) is linear and thus a WL deflation model is not required.

IV. SIMULATIONS AND DISCUSSIONS

A. Performance analysis

We will consider extraction of both synthetic and real-world sources from noise-free and noisy mixtures with various degrees of noise levels. The performance of the synthesised simulations are measured using the Performance Index (PI) [1] given as

$$PI = 10\log_{10}\left(\frac{1}{M}\left(\sum_{i=1}^{M} \frac{|u_i|^2}{\max\{|u_1|^2, \ldots, |u_M|^2\}}\right)\right). \tag{32}$$
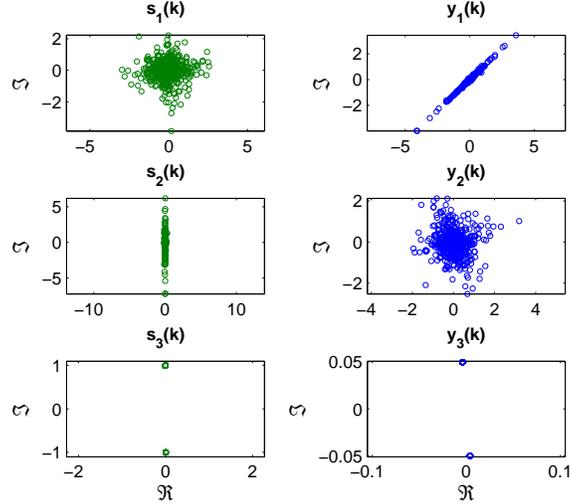
Fig. 3. Scatter plot of the complex-valued sources $s_1(k)$, $s_2(k)$ and $s_3(k)$, with the signal properties described in Table I (*left column*). Scatter plot of estimated sources $y_1(k)$, $y_2(k)$ and $y_3(k)$, extracted according to a decreasing order of kurtosis ($\beta = 1$) (*right column*).

where $\mathbf{u} = \mathbf{A}^H \mathbf{w} = [u_1, \ldots, u_M]$ and indicates the proximity of $\mathbf{u}$ to having only a single non-zero element and thus a successful convergence, as shown in the convergence analysis of Section III.

For each synthetic simulation, the results were produced through 100 independent trials and averaged. The mixing matrix $\mathbf{A}$ was generated randomly as a full rank complex matrix. The values of the extraction and deflation step-size μ was set empirically and adjusted for each consecutive extraction stage and the forgetting factor α in (22) was set as 0.975 in all the experiments. The complex Gaussian noise was generated in the forms of circular white with circularity measure $r = 0$ and noncircular doubly white with $r = 0.93$, although the degree of circularity of the noise does not affect the performance, as shown in the analysis and simulations.

In the first simulation, 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. The signals are illustrated in Fig. 3 and their properties listed in Table I. Extraction was performed in order from high to low kurtosis, with the value of $\beta = 1$ in (18).

In the first experiment, the performance of the algorithm (23) using a standard step-size was contrasted to that produced by utilising the VSS, normalised and GNGD methods in the extraction of the first source, $s_1(k)$, with the value of μ set to 0.01 for all methods. It can be seen from the performance curves in Fig. 4 that the best performance was achieved using the GNGD method, demonstrating the best performance with a PI of around -45dB at the steady-state. The performance curve resulting from the normalised method provided good 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. It can therefore be seen that the modification to the step-size enhances the performance of the algorithm.

In the next set of simulations, we considered the full extraction of all three sources (Fig. 3). The value of μ was set respectively as $0.01, 0.008$ and 10^{-5} for the consecutive extraction stages. As shown in Fig. 5, the algorithm successfully extracted all the three sources, as shown by a PI of less than -20dB at the steady-state for the extraction iteration $i = \{1, 2, 3\}$, with the algorithm 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 in 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 I; the scale and rotation ambiguities of the source estimates are also visible.

B. Extraction of communication sources

Communication sources are example of sources made complex by design and we next demonstrate the extraction of BPSK, QPSK and 16-QAM sources, illustrated in Fig. 6, from a noise-free mixture; the source properties are given in Table II. The BSPK source is noncircular, while the QPSK and 16-QAM sources are second-order circular and each has a distinct degree of Gaussianity which is exploited in the extraction process. The value of $\beta = -1$, such that source with the smallest kurtosis is extracted first (BSPK), followed on to the least sub-Gaussian (16-QAM). The number of samples generated was $N = 5000$ and the value of μ was chosen empirically and set respectively to 0.95, 2 and 0.1 for each iteration $i = \{1, 2, 3\}$ of the extraction stage. The algorithm has a very fast convergence in extracting the source signals (see Fig. 7) in the desired
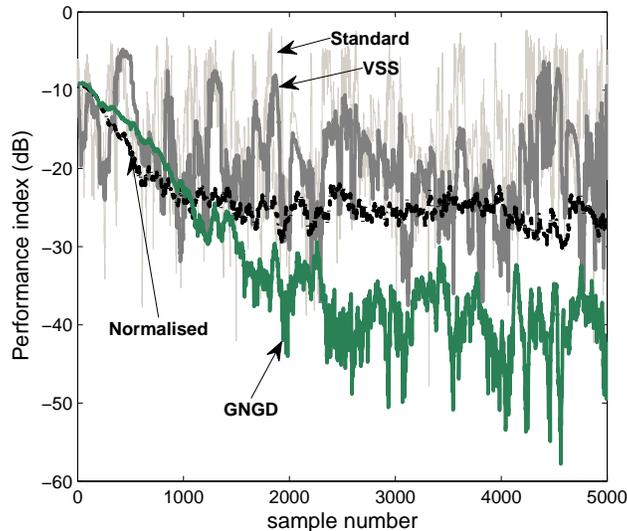
Fig. 4. Comparison of the effect of step-size adaptation on the performance of algorithm (23) for the extraction of a single source.
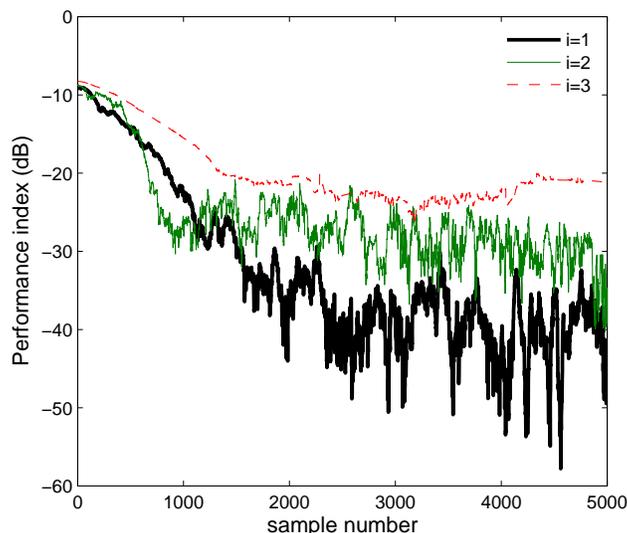

Fig. 5. Extraction of complex circular and noncircular sources from a noise-free mixture based on kurtosis.

order. The scatter plot of the extracted sources are given in Fig. 6 with the respective normalised kurtosis values calculated as $K_c(y_1) = -1.9996, K_c(y_2) = 1.0000$ and $K_c(y_3) = -0.6650$ which are close proximity to the true kurtosis values in Table II.

In the next experiment, we consider the extraction of a complex-valued source from a noisy mixture. Three sources of $N = 5000$ samples were considered (see Table III, Fig. 8) 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 III. 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. 8 with a calculated normalised kurtosis of $K_c(y_1) = -1.8002$, which is within a 10% range of the true value, given in Table III. The Performance Index, shown in Fig. 9, 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 III that the performance of the algorithm (23) 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 (23). The circularity measure r was varied from a value of $r = 0$ (circular) to a value of $r = 0.9998$ (highly noncircular), 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
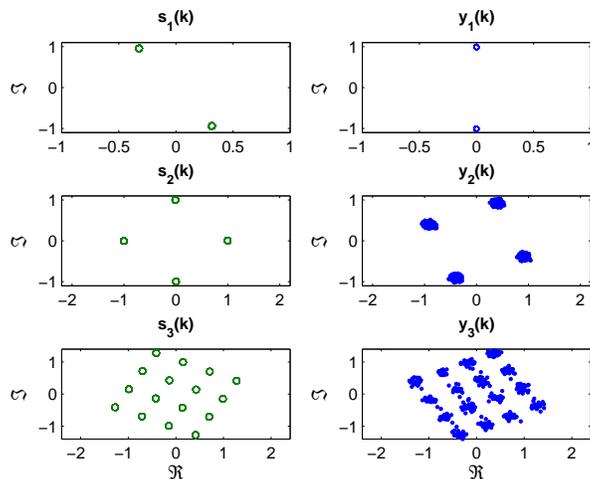
Fig. 6. Scatter plot of the BSPK, QPSK and 16-QAM sources $s_1(k), s_2(k)$ and $s_3(k)$, with properties given in Table II (*left column*), and the estimated sources $y_1(k), y_2(k)$ and $y_3(k)$ (*right column*).
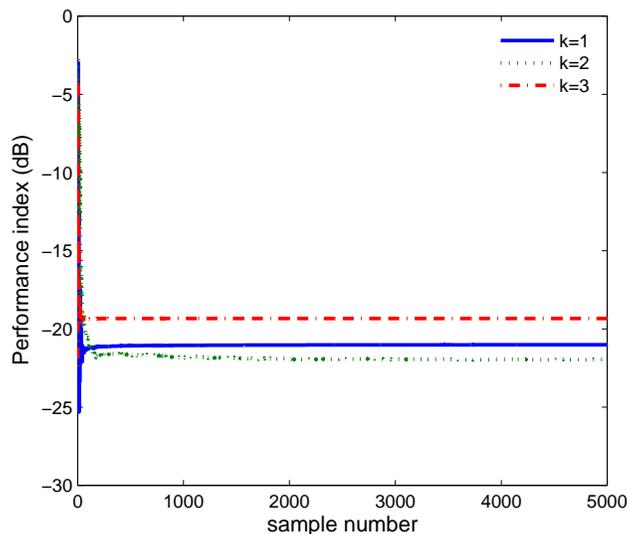

Fig. 7. Extraction of communication sources (properties given in Table II) in a noise-free environment.

of -10dB. The initial values were generated randomly and PI was averaged over 100 trials. Fig. 10 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 1dB.

C. EEG artifact extraction

In order to obtain useful information from recorded EEG data, it is often necessary to perform post-processing operations to remove artifacts such as line noise and biological artifacts including those pertaining to eye muscle activity, also known as electrooculogram (EOG) and facial muscle activity known 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 [40]–[45]. While offline algorithms are suitable for processing the recorded EEG data for further analysis, it is necessary to utilise online algorithms for real-time applications such as those encountered in brain computer interface (BCI) scenarios.

In [44] 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
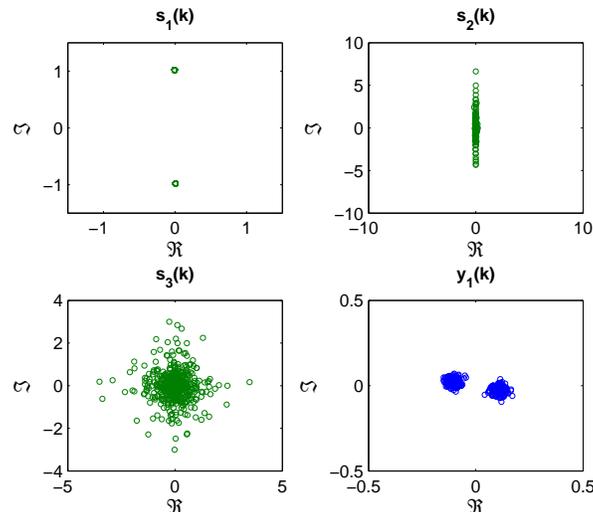
Fig. 8.   Scatter plots of the original sources $s_1(k), s_2(k)$ and $s_3(k)$. The first estimated source $y_1(k)$ is shown in the bottom-right plot.
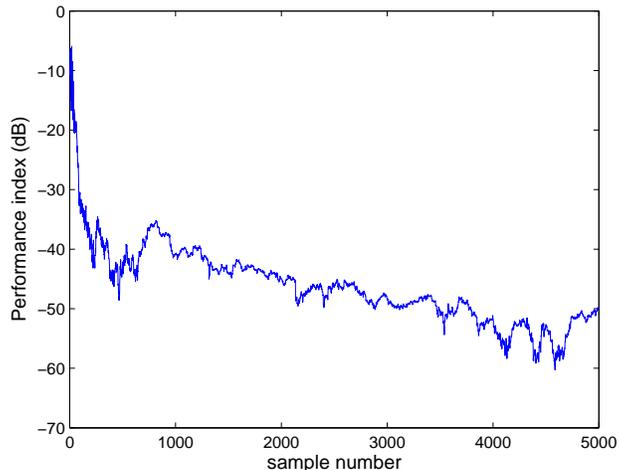

Fig. 9.   Extraction of a complex-valued source from a noisy mixture, with the source properties given in Table III.

perform all the necessary processing in the time domain, as 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 inspection. As detailed in [40], the ICA algorithm separates the recorded EEG mixture into its original sources as independent components (ICs), and artifact sources are then identified and removed. In semi-automatic [41] and automatic [43] 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, the IC kurtosis and entropy have both been utilised to identify and isolate the artifacts. While the recorded EEG mixture typically has a near-zero kurtosis value, indicating a possible Gaussian distribution, artifacts such as EOG illustrate a peaked distribution with a highly positive kurtosis value [43], while periodic power line noise has a highly negative kurtosis value. This has been used as the main concept in defining classifications based on the the fourth order moment.

Using the concept discussed, our aim is to extract artifacts as independent sources from the recorded EEG mixture directly in the time domain. Therefore through the convenience of representation [3], the corresponding contaminated EEG signals are paired as the real and imaginary components of a complex signal and processed using the architecture described in Section III. In this manner, the full mutual information between the corresponding electrodes and the resultant recorded EEG is maintained, while allowing 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 instead, pipelined to a further post processing stage, which would then extract

TABLE I
SOURCE PROPERTIES FOR NOISE-FREE EXTRACTION SIMULATIONS

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 II
PROPERTIES OF THE BPSK, QPSK AND 16-QAM SOURCES USED IN SIMULATIONS

Source	Type	Distribution	Kurtosis	circ. measure (r)
$s_1(k)$	BSPK	Sub-Gaussian	-1.9996	1.0000
$s_2(k)$	QPSK	Sub-Gaussian	-1.0000	0.0016
$s_3(k)$	16-QAM	Sub-Gaussian	-0.6804	0.0004

the EEG signals based on a desired fundamental property, such as predictability.

We now evaluate the performance of the proposed K-cBSE algorithm (23) in the extraction of several artifacts from recorded EEG signals. The electrodes were placed according to the 10-20 system (Fig. 11), 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 at the side of the left eye socket. As we have no knowledge of the mixing process, the Performance Index (32) is not applicable for this case and thus the frequency spectra of the extracted and original EOG signals are used instead for the evaluation of the algorithm performance.

Three separate studies were performed with the aim of assessing the performance of the algorithm in removing a unique artifact simultaneously with the line noise, which in Europe, is 50Hz. While the rejection of the power line noise artifact is feasible by passing the recorded EEG signals through a bandpass filter, this hard-limited solution also leads to the removal of information within the 50Hz range pertaining to the EEG signals, in particular those within the gamma band (25Hz-100Hz). It would therefore be desired 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), and in the second study eye muscle artifacts from rolling the eyes ('EYEROLL' set) and in the third study muscle activity from raising the eyebrow ('EYEBROW' set) are extracted.

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

$$x_1(k) = \text{Fp1}(k) + \jmath\text{Fp2}(k)$$
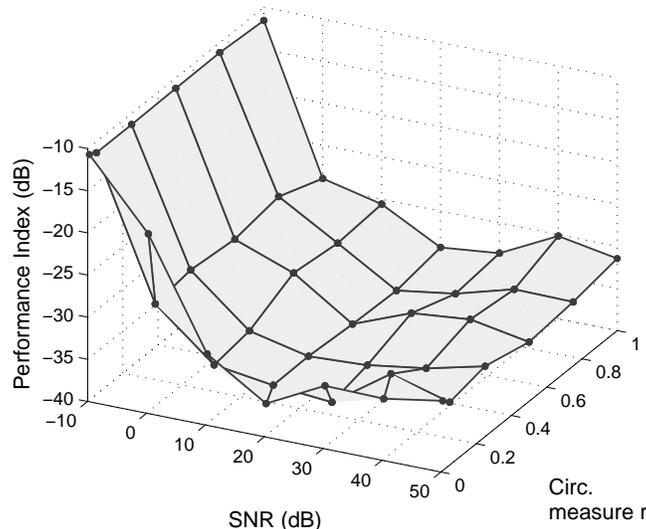$$x_2(k) = \text{C3}(k) + \jmath\text{C4}(k)$$

Fig. 10. Comparison of the performance of algorithm (23) with respect to changes in the SNR and noise circularity.

TABLE III
SOURCE PROPERTIES FOR EXTRACTION SIMULATIONS FROM A NOISY MIXTURE

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

$$x_3(k) = \text{O1}(k) + \jmath\text{O2}(k). \tag{33}$$

The construction of the complex EEG signals through convenience of representation using the specific channel pairs is owing to the cross-information available in corresponding pairs from each sidelobe, allowing for the simultaneous processing of the amplitude and phase information using the K-cBSE algorithm (23).

As we have no knowledge of the mixing process, the Performance Index (32) is not applicable for this case and thus several quantitative and qualitative measures are used instead for the evaluation of the algorithm performance. These are briefly discussed below.

Quantitative metrics

- *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.
- *Power spectra Correlation*: In a similar manner to [45], 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.

Qualitative measures

- *Hilbert-Huang Time-Frequency Analysis*: By employing time-frequency (T-F) analysis using the Hilbert-Huang (H-H) transform [46], [47], 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 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 necessary intrinsic mode functions (IMFs) required by the H-H transform were obtained using a multivariate empirical mode decomposition (MEMD) algorithm [48], 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.
- *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 this manner, the distribution of power in the extracted sources is compared to the original artifact. In addition, the pseudo-spectrum demonstrates the quality of the proposed method in extracting general complex source. For implementation purposes, the PSD and pPSD were calculated directly from the autocorrelation and pseudo-autocorrelation of the complex-valued signals.

1) Case Study 1 – EOG extraction: The 'EYEBLINK' set was used for this study, where the EEG recordings were contaminated with eye blink activity as well as line noise. The recorded EEG and EOG signals are plotted in Fig. 12(a), where the effect of the EOG activity is seen to be very strong in the frontal lobe Fp1 and Fp2 channels, with the effect diminishing as the distance of the electrodes to the eyes increases. The effect of the line noise is also visible on the occipital O1 and O2 channels, as well as the hEOG recording. The H-H T-F spectrogram (Fig 12(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 activity are visible with a duration of around 1 seconds, while EEG activity is not clearly visible. In addition, 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 (33) 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 Table IV and V, where the normalised kurtosis of both the extracted complex signals and their respective components are given.

TABLE IV
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

The order of the extracted complex signals are as expected, with the first extracted source $y_1(k)$ being sub-Gaussian and $y_2(k)$ a super-Gaussian source. This is also reflected in the kurtosis of the individual real and imaginary components of the signals. The imaginary component of $y_1(k)$ has the smallest kurtosis, and is 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, are taken as the extracted EOG sources. Fig. 12(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. 12(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 (33), and the extracted sources $y_1(k)$ and $y_2(k)$, depicted in Fig. 12(e). As described in Section II, the high degree of noncircularity of the constructed EOG signals in reflected in the similar magnitudes of its spectrum and pseudo-spectrum. Notice that the distribution of power $\mathcal{S}_{EOG}$ and pseudo-power $\breve{\mathcal{S}}_{EOG}$ is concentrated respectively in the the frequency range 0Hz-5Hz 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 0Hz-5Hz frequency range. These results can also be seen by comparing the frequency components of the recorded EEG mixture and extracted source around the 50Hz range, in Fig. 12(f). While the presence of the power line artifact is evident in all recorded channels, the 50Hz frequency component is only available in $\Im\{y_1(k)\}$ after the extraction procedure. Likewise, the spectra of $y_2(k)$ illustrates 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 PSD and pPSD and those of the extracted sources were calculated [45] and presented in Table V. For the extracted source $y_1(k)$ these values are respectively 0.2313 and 0.2847, while measuring 0.9698 and 0.9822 for $y_2(k)$. 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.

2) Case Study 2 – Eye muscle artifact extraction: In this study, the 'EYEROLL' set contained artifacts from circular movement of the eye during the recording session with inadvertent EOG activity from eye blinks, shown in Fig. 13(a) and kurtosis values
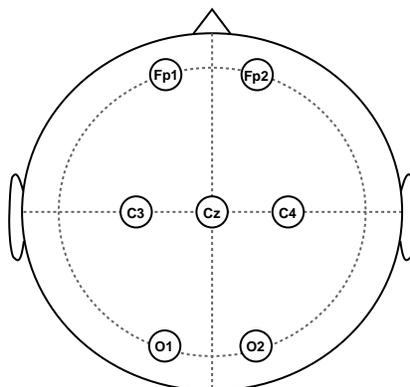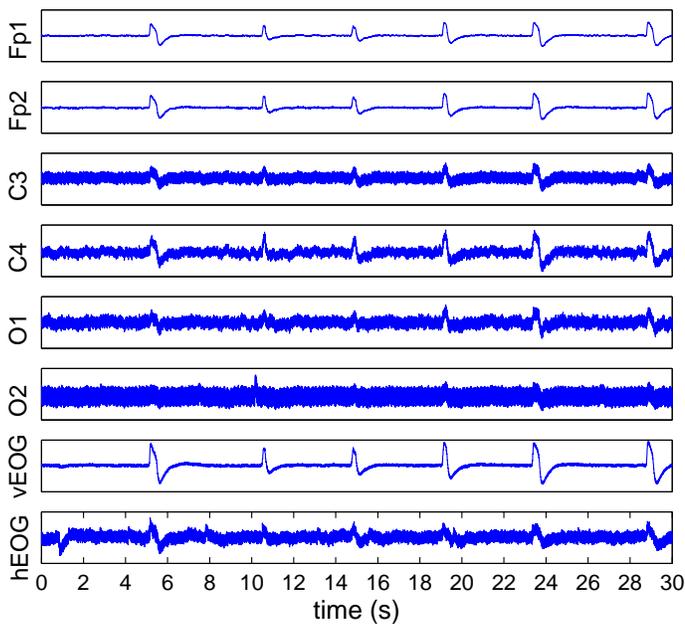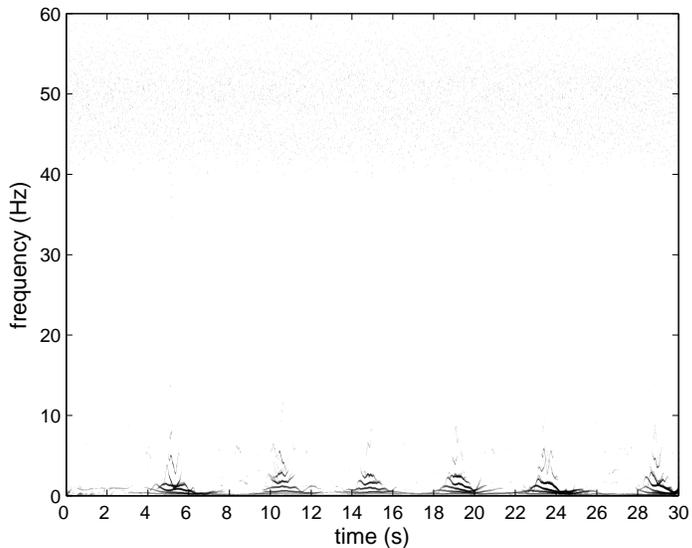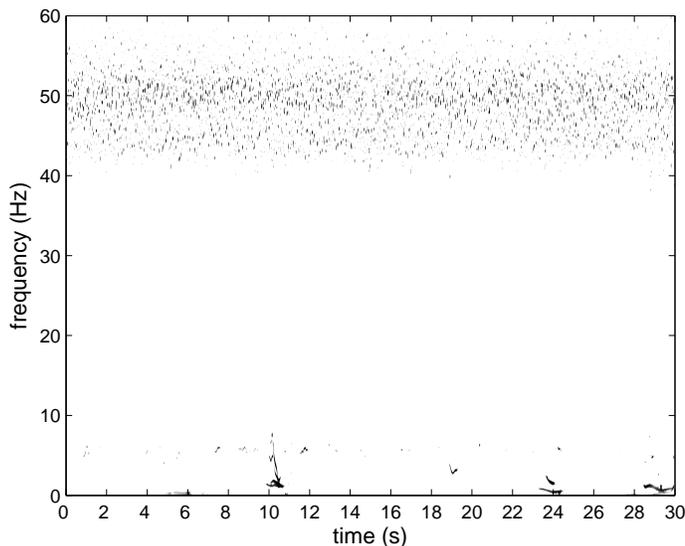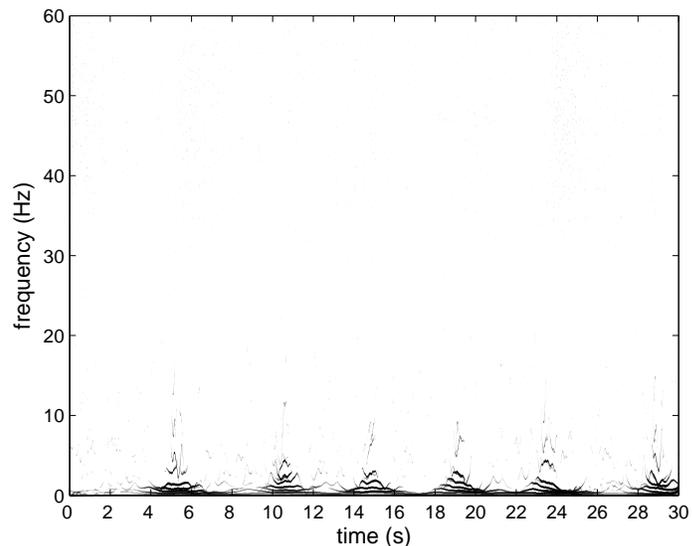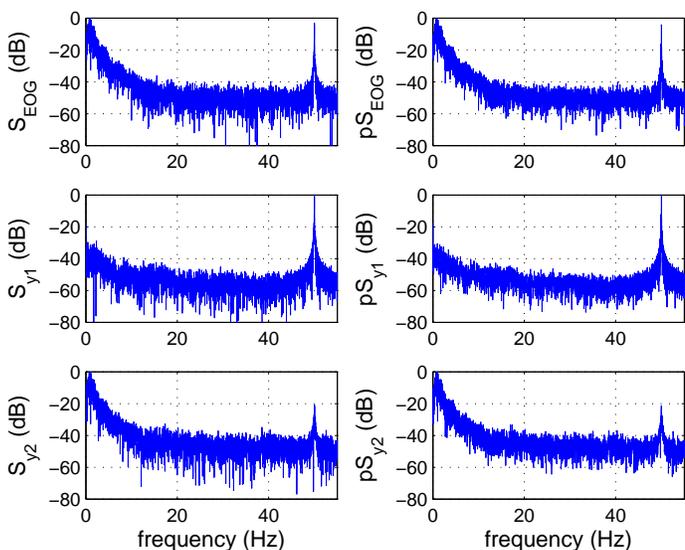

Fig. 11. Placement of electrodes on the scalp according to the 10-20 system for EEG recording.

(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 $\Im\{y_1(k)\}$

(d) The Hilbert-Huang time-frequency plot of the extracted EOG $\Re\{y_2(k)\}, \Im\{y_2(k)\}$

(e) The power spectrum (S on graph) and pseudo-spectrum (pS on graph) of the recorded EOG and extracted signals $y_1(k)$ and $y_2(k)$
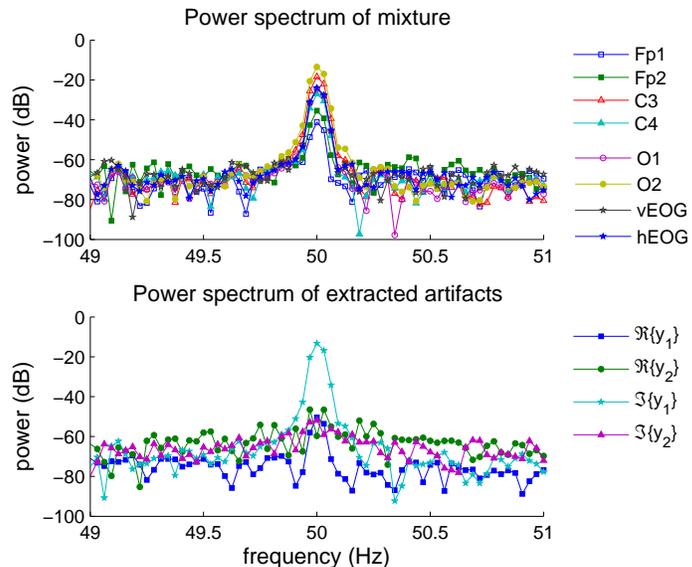
(f) Frequency components of the recorded EEG signals and extracted signals around the 50Hz range

Fig. 12. Recorded and extracted artifacts from the 'EYEBLINK' set

TABLE V
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

| Set | Signal | K_c | $K_R(\Re, \Im)$ | Spectra corr. | |
				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

given in Table IV. 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 is 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. 13(b) illustrates the presence of frequency components up to 10Hz, as well as scattered frequencies belonging to the 50Hz power line noise.

For the extraction procedure, the step-size of the K-cBSE algorithm $\mu = \{5, 0.2\}$ and $\beta = \{-1, 1\}$, while $\alpha = 0.975$. The T-F analysis of the extraction are illustrated in Fig. 13(c)–(d), and the kurtosis values of the complex-valued extracted signals and their real and imaginary components given in Table V. 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. 13(d), while the lack of power line noise frequency components in the 50Hz range is visible.

These results are also confirmed from the power spectra of the recorded artifacts and the extracted sources, given in Fig. 13(e). While the PSD and pPSD of the complex-valued $y_1(k)$ contains the 50Hz components, these are diminished to -40dB in the spectra of $y_2(k)$. The frequency components of the mixture channels and extracted sources in the 50Hz range also show that the line noise artifact was successfully removed (see Fig. 13(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 0-10Hz range of the PSD and pPSD of the complex-valued EOG. Quantitatively, 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 is 0.8222, in comparison these values are respectively 0.0792 and 0.1844 for $y_1(k)$.

3) Case Study 3 – EMG extraction: In the 'EYEBROW' set, the EEG mixture was contaminated with EMG artifacts resulting from the raising of the eyebrows shown in Fig. 14(a) with kurtosis values given in Table IV. 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. 14(b)) the EMG frequency range has 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 scatterings belonging the power line noise can also be seen.

The extraction of the artifacts was performed using the K-cBSE algorithm with step-size $\mu = \{2, 0.2\}$, $\beta = \{-1, 1\}$ and $\alpha = 0.975$. Owing to the large amplitude of the EMG signals and wide frequency range, the extraction of the artifacts was not straightforward. However, as shown in Fig. 14(c) and Fig. 14(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 the real and imaginary components of $y_2(k)$ in Fig. 14(d) it is seen that the complete EMG frequency component range has been successfully extracted with power line noise frequency components not present.

Considering the power spectra $\mathcal{S}_{EMG}$ and $\check{\mathcal{S}}_{EMG}$ in Fig. 14(e), the spectral distribution of the power and pseudo-power are strong in the 0Hz-10Hz range with an amplitude of around -10dB and in the 20Hz-40Hz range, though having a much lower value. In addition a single spike at 50Hz of amplitude -10dB denotes the presence of power line noise. After the extraction, the power line noise is contained in the spectra of the $y_1(k)$ while the 0Hz-10Hz and 20Hz-40Hz frequency components are present in the PSD and pPSD of $y_2(k)$. For the 'EYEBROW' set, the spectra correlation coefficients between $\mathcal{S}_{EMG}$ and $\check{\mathcal{S}}_{EMG}$ and those of $y_1(k)$ and $y_2(k)$ were respectively $\{0.1287, 0.1078\}$ and $\{0.7593, 0.7906\}$. Also consider the 50Hz frequency range for the contaminated mixture and the extracted source shown in Fig. 14(f). It is 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. 12(f) and Fig. 13(f)), $\Re\{y_2(k)\}$ and $\Im\{y_2(k)\}$ have a higher power level in this range, reflecting the wider frequency range of the EMG artifact.
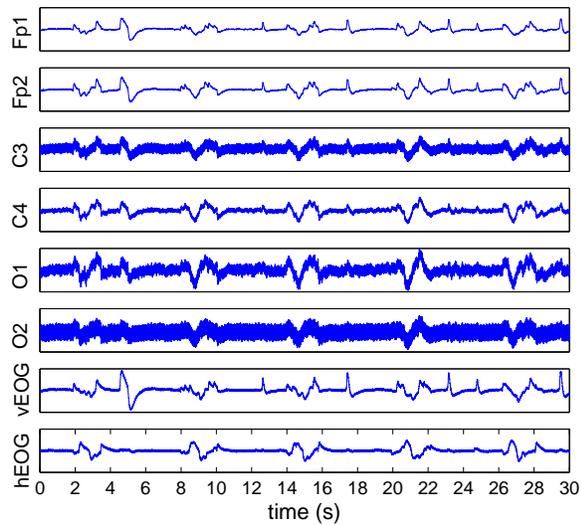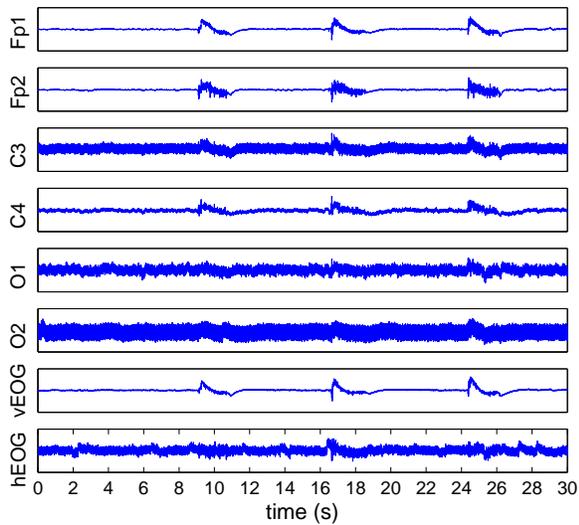
(a) Recorded EEG signals from the 'EYEROLL' set
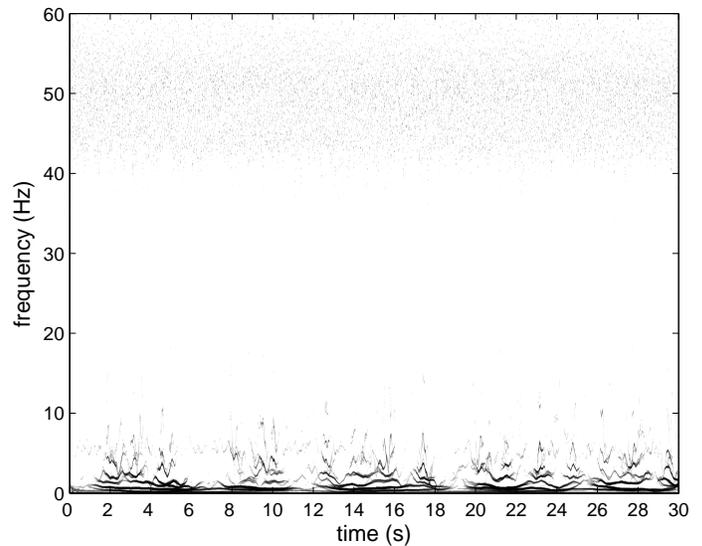
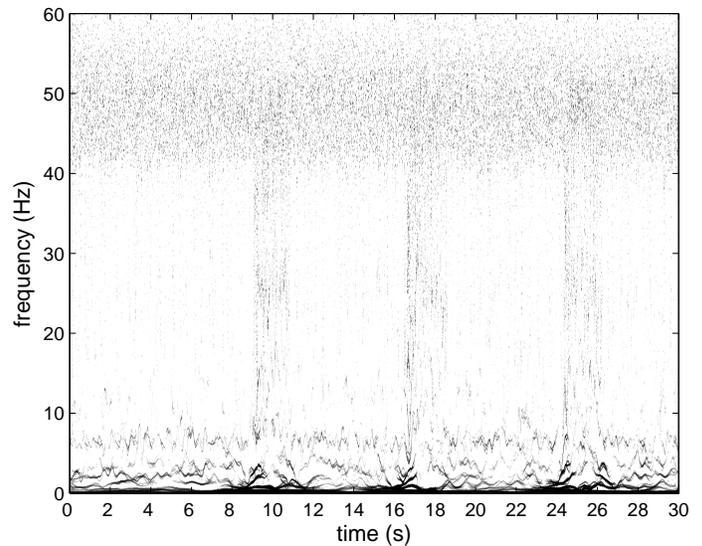(b) The Hilbert-Huang time-frequency plot of the recorded EEG signals
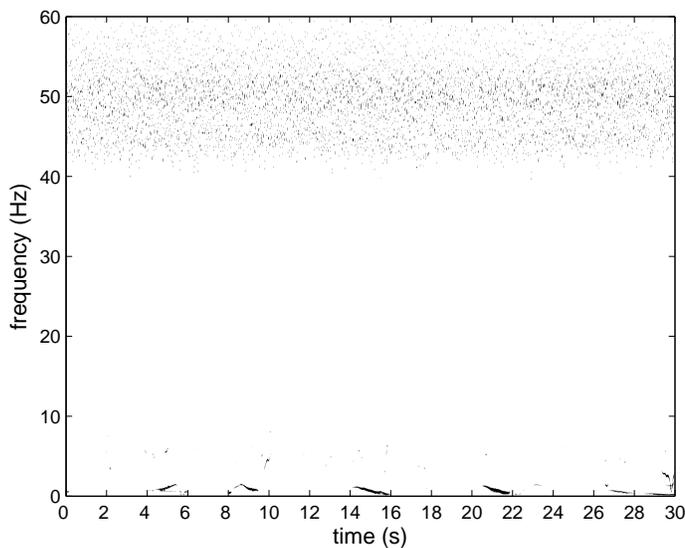
(c) The Hilbert-Huang time-frequency plot of the extracted line noise $\Re\{y_1(k)\}$
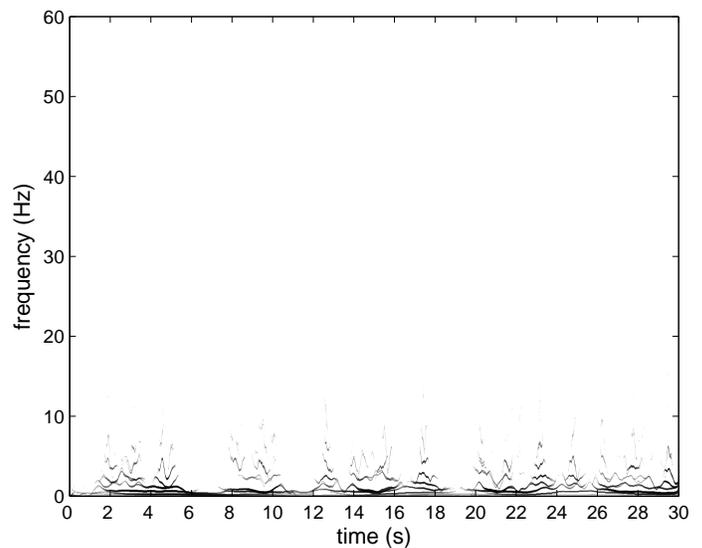
(d) The Hilbert-Huang time-frequency plot of the extracted EOG $\Re\{y_2(k)\}, \Im\{y_2(k)\}$
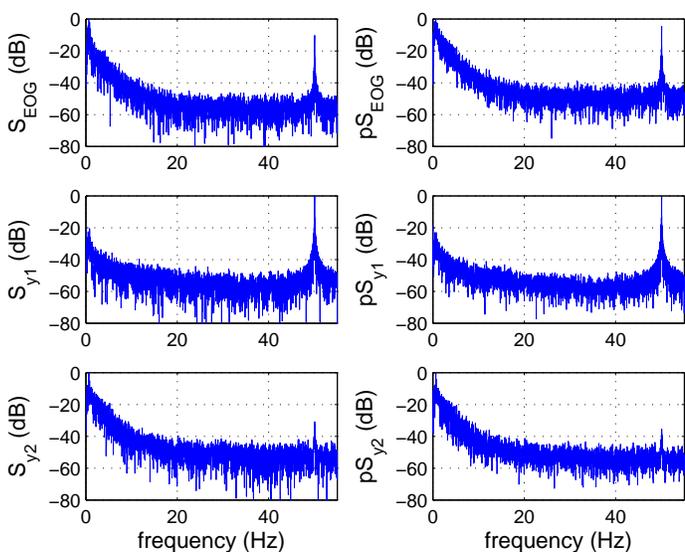
(e) The power spectrum (S on graph) and pseudo-spectrum (pS on graph) of the recorded EOG and extracted signals $y_1(k)$ and $y_2(k)$
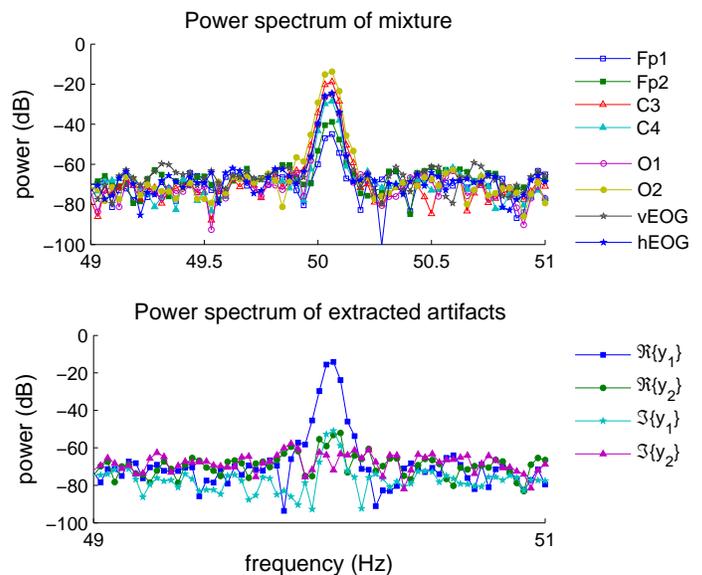
(f) Frequency components of the recorded EEG signals and extracted signals around the 50Hz range
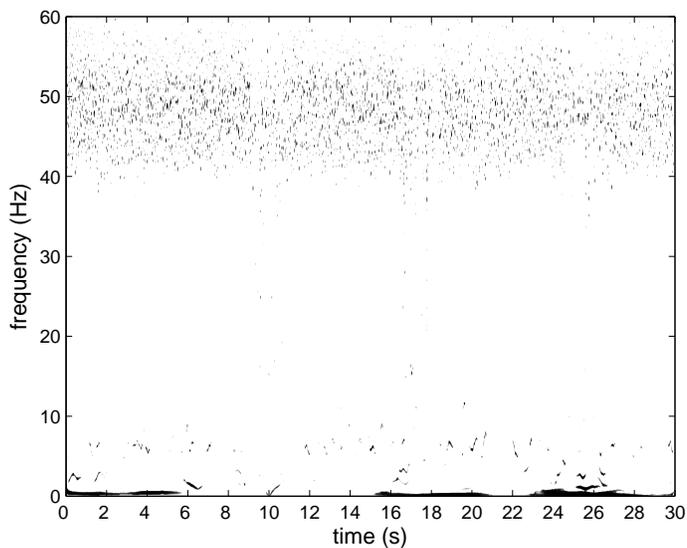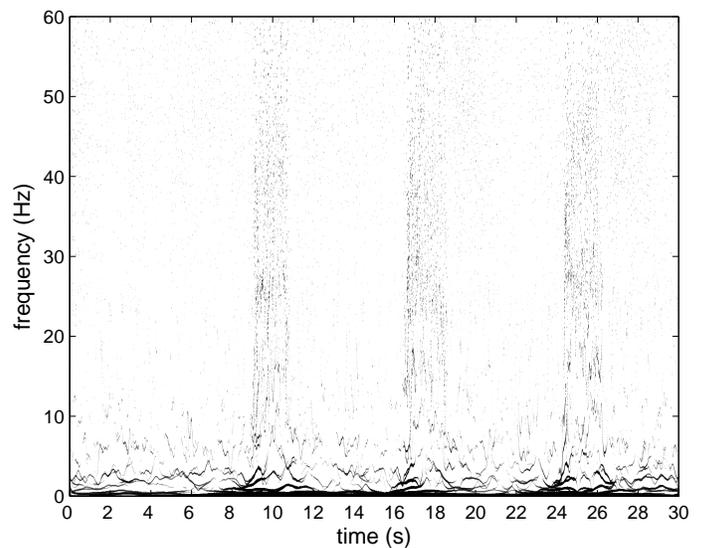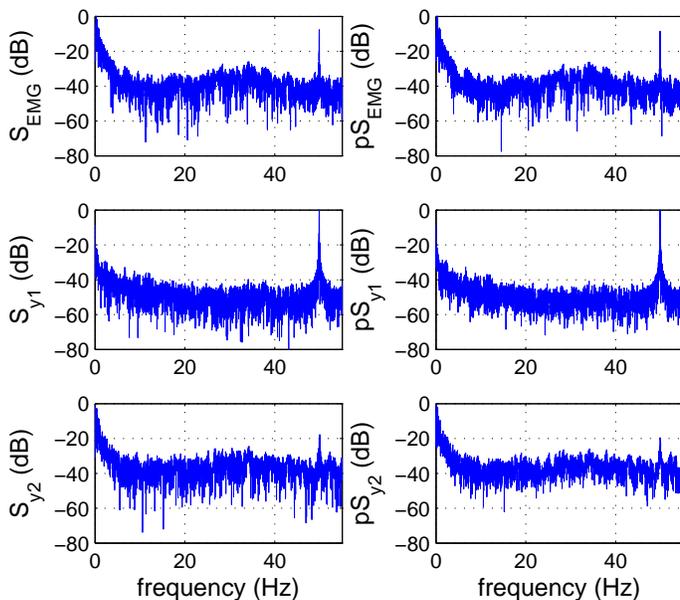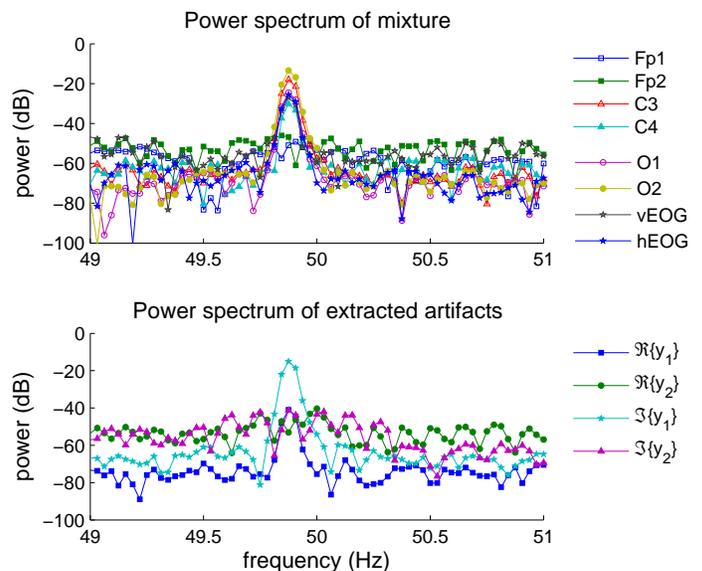
Fig. 13. Recorded and extracted artifacts from the 'EYEROLL' set

(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 spectrum (S on graph) and pseudo-spectrum (pS on graph) of the recorded EMG and extracted signals $y_1(k)$ and $y_2(k)$

(f) Frequency components of the recorded EEG signals and extracted signals around the 50Hz range

Fig. 14. Recorded and extracted artifacts from the 'EYEBROW' set

V. EEG ARTIFACT EXTRACTION – HIGHER SAMPLING RATE

Simulations for EEG artifact extraction using the K-cBSE algorithm from the 'EYEBLINK', 'EYEROLL' and 'EYEBROW' EEG recording dataset, but using a sampling rate of 4800Hz are presented. A total of 16 electrodes were used, with 12 electrodes placed on the scalp (Fig. 15), and 4 electrodes placed above and to the side of both eye sockets, where the difference between respective electrodes measured the vEOG and hEOG recordings. The ground electrode was placed on the forehead and the reference electrode on the right earlobe. Note that the A1 and A2 electrodes were placed respectively on the left and right mastoid (A1 is referred to as ML and A2 as MR in figures, to fix.).

A total of 6 complex EEG signals were formed using symmetric pairs of electrodes, given as

$$x_1(k) = \text{AF7}(k) + \jmath\text{AF8}(k)$$
$$x_2(k) = \text{AF3}(k) + \jmath\text{AF4}(k)$$
$$x_3(k) = \text{A1}(k) + \jmath\text{A2}(k)$$
$$x_4(k) = \text{C3}(k) + \jmath\text{C4}(k)$$
$$x_5(k) = \text{PO7}(k) + \jmath\text{PO8}(k)$$
$$x_6(k) = \text{PO3}(k) + \jmath\text{PO4}(k)$$

$$(34)$$

Artifact extraction was performed on a 10 second segment of each dataset, and the extracted artifacts were subsequently removed from the recorded mixture. The results are shown in Fig. V–18
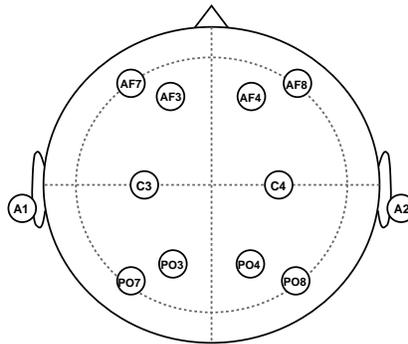

Fig. 15.   Electrode locations

VI. CONCLUSIONS

Blind source extraction of the generality of complex-valued signals based on the degree of non-Gaussianity and from noisy mixtures has been been addressed. A cost function based on the normalised kurtosis is utilised to perform blind extraction, and the corresponding online algorithm (K-cBSE) has been derived. The algorithm is shown to be convergent and variable step-size variants of to the algorithm have been discussed. It has been shown that the algorithm is robust to the 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 performance of the algorithm for the successful extraction of both circular and non-circular signals, while the extraction of EOG and EMG artifacts from recorded EEG signals 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 (26) as

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

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

$$\nabla_\epsilon\mathcal{J} = \left(\nabla_{\mathbf{w}^*}\mathcal{J}\right)^T \cdot \frac{\partial\mathbf{w}^*(k)}{\partial\upsilon(k-1)} \cdot \frac{\partial\upsilon(k-1)}{\partial\epsilon(k-1)} \tag{35}$$
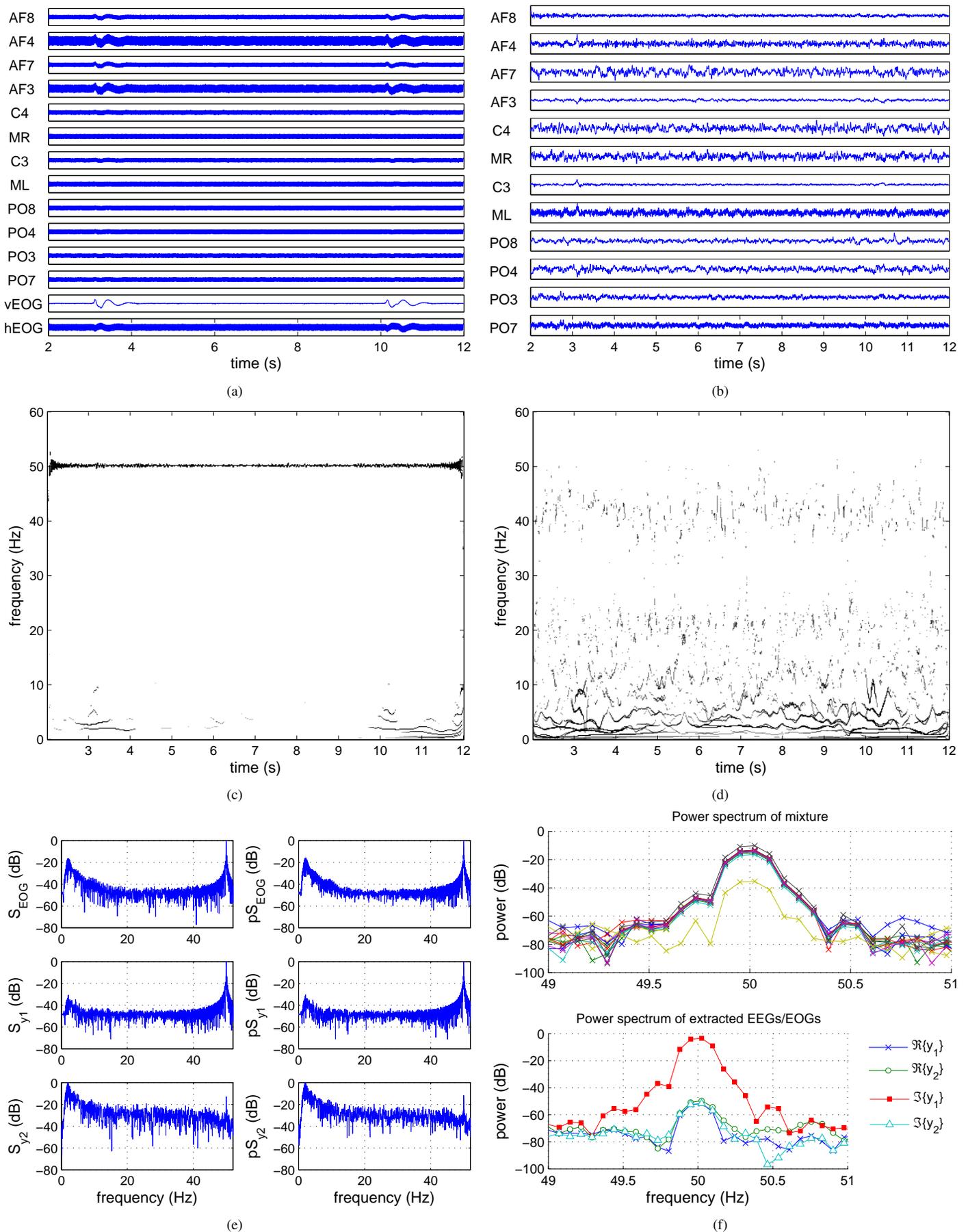
Fig. 16.   Recorded and extracted artifacts from the 'EYEBLINK' set. (a) Recorded EEG signals from the 'EYEBLINK' set. (b) The EEG mixture after artifact removal. (c) The Hilbert-Huang time-frequency plot of the extracted line noise $\Im\{y_1(k)\}$. (d) The Hilbert-Huang time-frequency plot of the extracted 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 signals around the 50Hz frequency range. After extraction, the power line noise is contained in $\Im\{y_1\}$.

Fig. 17.   Recorded and extracted artifacts from the 'EYEROLL' set. (a) Recorded EEG signals from the 'EYEROLL' set. (b) The EEG mixture after artifact removal. (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 signals around the 50Hz frequency range. After extraction, the power line noise is contained in $\Re\{y_1\}$.
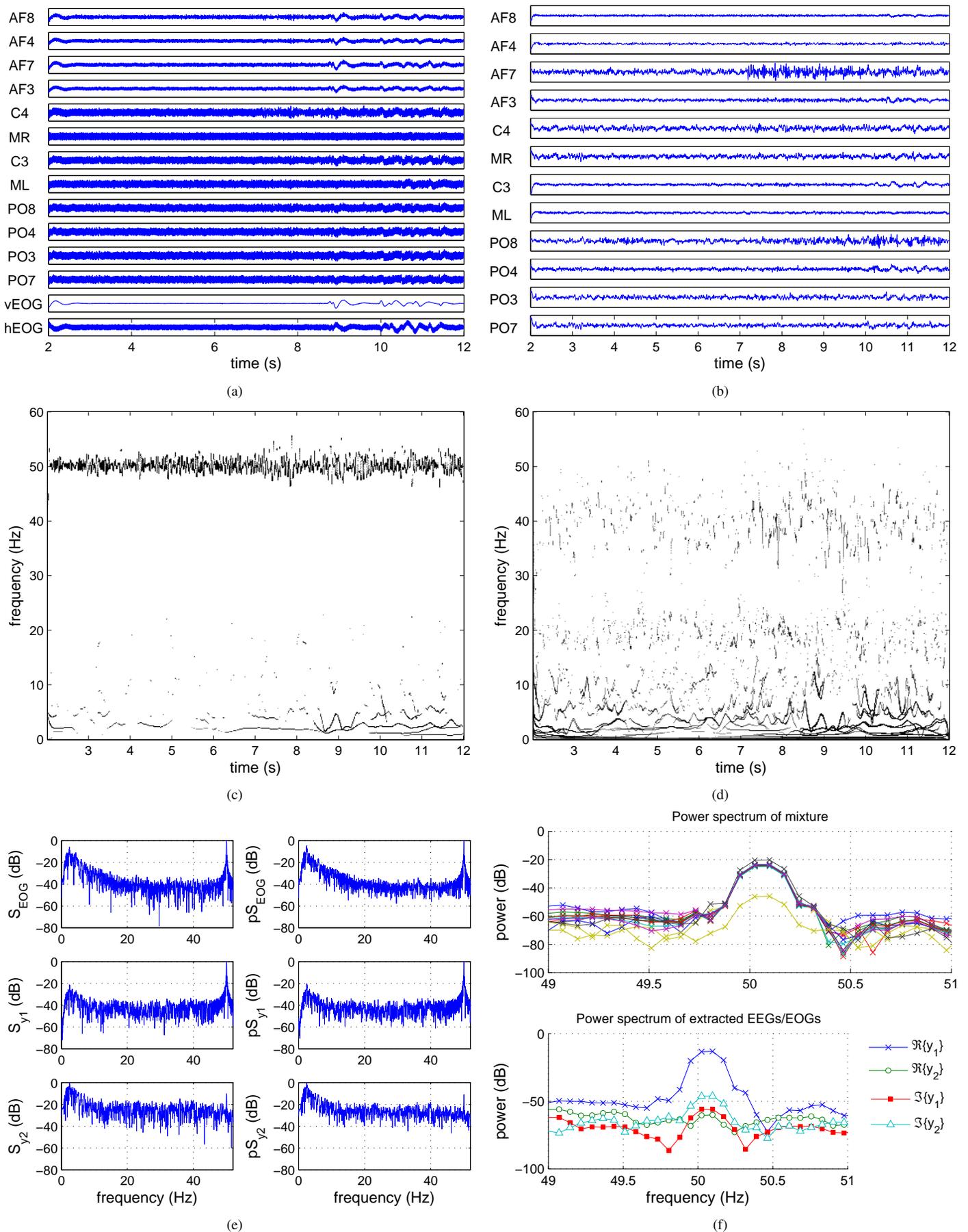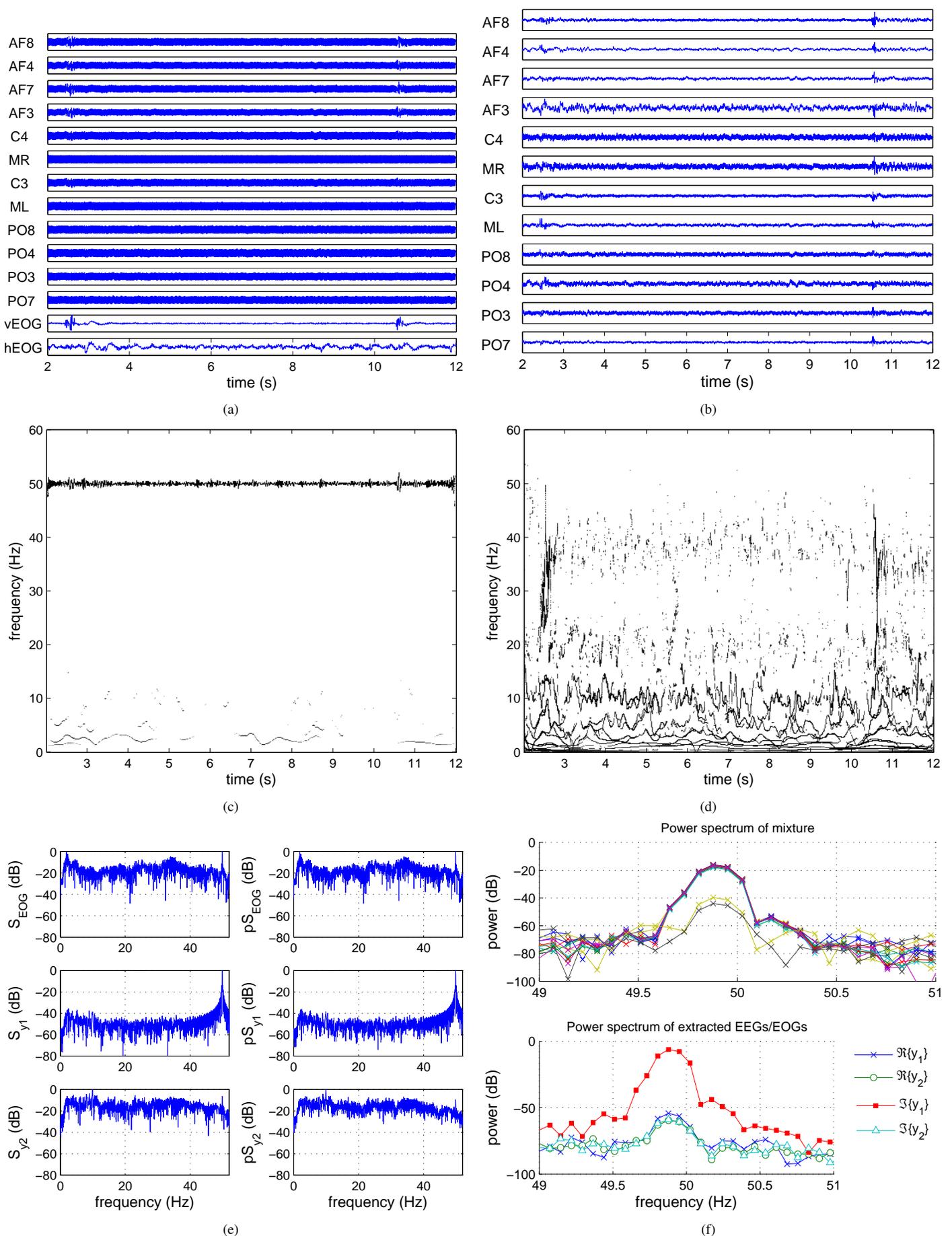
Fig. 18. Recorded and extracted artifacts from the 'EYEBROW' set. (a) Recorded EEG signals from the 'EYEBROW' set. (b) The EEG mixture after artifact removal. (c) The Hilbert-Huang time-frequency plot of the extracted line noise $\Im\{y_1(k)\}$. (d) The Hilbert-Huang time-frequency plot of the extracted 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 signals around the 50Hz frequency range. After extraction, the power line noise is contained in $\Im\{y_1\}$.

where

$$\frac{\partial \mathbf{w}^*(k)}{\partial \upsilon(k-1)} = \frac{\partial \mathbf{w}^*(k)}{\partial \upsilon(k-1)} - \phi^*\big(y(k-1)\big)\mathbf{x}^*(k-1)-$$

$$\frac{\partial \phi^*\big(y(k-1)\big)}{\partial \upsilon(k-1)}\upsilon(k-1)\mathbf{x}^*(k-1)$$

$$\approx -\phi^*\big(y(k-1)\big)\mathbf{x}^*(k-1)$$

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

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

While the derivative in (35) is calculated with respect to a complex value in the $\mathbb{CR}$ calculus framework, $\epsilon(k)$ is real-valued and so only the real component of the $\mathbb{R}^*$–derivative in (35) is required. This leads to the update equation given in (27).

References

[1] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing, Learning Algorithms and Applications*. Wiley, 2002.
[2] D. P. Mandic, S. Javidi, G. Souretis, and S. L. Goh, "Why a complex valued solution for a real domain problem," in *IEEE Workshop on Machine Learning for Signal Processing*, Aug. 2007, pp. 384–389.
[3] D. P. Mandic and S. L. Goh, *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models*. Wiley, 2009.
[4] S. Amari, S. C. Douglas, A. Cichocki, and H. H. Yang, "Novel on-line adaptive learning algorithms for blind deconvolution using the natural gradient approach," in *Proc. 11th IFAC Symp. System Identification*, vol. 3, July 1997, pp. 1057–1062.
[5] A. J. Bell and T. J. Sejnowski, "An information-maximisation approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, pp. 1129–1159, 1995.
[6] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. Wiley, 2001.
[7] W. Y. Leong and D. P. Mandic, "Post-nonlinear blind extraction in the presence of ill-conditioned mixing," *IEEE Transactions on Circuits and Systems I*, vol. 55, pp. 2631–2638, Oct. 2008.
[8] J. Särelä and H. Valpola, "Denoising source separation," *The Journal of Machine Learning Research*, vol. 6, pp. 233–272, 2005.
[9] A. K. Barros and A. Cichocki, "Extraction of specific signals with temporal structure," *Neural Computation*, vol. 13, no. 9, pp. 1995–2003, 2001.
[10] D. P. Mandic and A. Cichocki, "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)*, 2003, pp. 645–650.
[11] W. Liu, D. P. Mandic, and A. Cichocki, "Blind source extraction of instantaneous noisy mixtures using a linear predictor," in *Proc. IEEE International Symposium on Circuits and Systems*, 2006, pp. 4199–4202.
[12] ——, "Blind second-order source extraction of instantaneous noisy mixtures," *IEEE Transactions of Circuits and Systems II*, vol. 53, no. 9, pp. 931–935, 2006.
[13] A. Cichocki, R. Thawonmas, and S. Amari, "Sequential blind signal extraction in order specified by stochastic properties," *Electronics Letters*, vol. 33, pp. 64–65, 1997.
[14] W. Liu and D. P. Mandic, "A normalised kurtosis-based algorithm for blind source extraction from noisy measurements," *Signal Processing*, vol. 86, no. 7, pp. 1580–1585, 2006.
[15] 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.
[16] B. Picinbono, "On circularity," *IEEE Transactions on Signal Processing*, vol. 42, no. 12, pp. 3473–3482, 1994.
[17] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Transactions on Information Theory*, vol. 39, no. 4, pp. 1293–1302, 1993.
[18] B. Picinbono and P. Bondon, "Second-order statistics of complex signals," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 411–420, 1997.
[19] P. J. Schreier and L. L. Scharf, "Second-order analysis of improper complex random vectors and processes," *IEEE Transactions on Signal Processing*, vol. 51, no. 3, pp. 714–725, 2003.
[20] B. Picinbono and P. Chevalier, "Widely linear estimation with complex data," *IEEE Transactions on Signal Processing*, vol. 43, no. 8, pp. 2030–2033, 1995.
[21] A. T. Erdogan, "On the convergence of ICA algorithms with symmetric orthogonalization," *IEEE Transactions on Signal Processing*, vol. 57, no. 6, pp. 2209–2221, June 2009.
[22] S. C. Douglas, "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*, 2005, pp. 1320–1325.
[23] M. Novey and T. Adalı, "On extending the complex FastICA algorithm to noncircular sources," *IEEE Transactions on Signal Processing*, vol. 56, no. 5, pp. 2148–2154, 2008.
[24] J. Anemüller, T. J. Sejnowski, and S. Makeig, "Complex independent component analysis of frequency-domain electroencephalographic data," *Neural Networks*, vol. 16, no. 9, pp. 1311–1323, Nov. 2003.
[25] E. Bingham and A. Hyvärinen, "A fast fixed point algorithm for independent component analysis of complex valued signals," *Journal of Neural Systems*, vol. 10, pp. 1–8, 2000.
[26] S. Javidi, B. Jelfs, and D. P. Mandic, "Blind extraction of noncircular signals using a widely linear predictor," in *IEEE Workshop on Statistical Signal Processing*, 2009, pp. 501–504.
[27] S. Javidi, D. P. Mandic, and A. Cichocki, "Complex blind source extraction from noisy mixtures using second order statistics," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 57, no. 7, pp. 1404–1416, Jul. 2010.
[28] 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.
[29] E. Ollila and V. Koivunen, "Complex ICA using generalized uncorrelating transform," *Signal Processing*, vol. 89, no. 4, pp. 365 – 377, 2009.
[30] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Computation*, vol. 9, no. 7, pp. 1483–1492, 1997.
[31] E. Ollila and V. Koivunen, "Adjusting the generalized likelihood ratio test of circularity robust to non-normality," in *IEEE 10th Workshop on Signal Processing Advances in Wireless Communications*, 2009, pp. 558–562.
[32] E. Ollila, "On the circularity of a complex random variable," *IEEE Signal Processing Letters*, vol. 15, pp. 841–844, 2008.
[33] 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. [Online]. Available: http://arxiv.org/abs/0906.4835v1

[34] A. van den Bos, "Complex gradient and Hessian," *IEE Proceedings of Vision, Image and Signal Processing*, vol. 141, no. 6, pp. 380–383, 1994.

[35] D. H. Brandwood, "A complex gradient operator and its application in adaptive array theory," *IEE Proceedings F: Communications, Radar and Signal Processing*, vol. 130, no. 1, pp. 11–16, Feb. 1983.

[36] R. Thawonmas, A. Cichocki, and S. Amari, "A cascade neural network for blind signal extraction without spurious equilibria," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 81, no. 9, pp. 1833–1846, 1998.

[37] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*. Wiley, 1996.

[38] W.-P. Ang and B. Farhang-Boroujeny, "A new class of gradient adaptive step-size LMS algorithms," *IEEE Transactions on Signal Processing*, vol. 49, no. 4, pp. 805–810, 2001.

[39] D. P. Mandic, "A generalized normalized gradient descent algorithm," *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 115–118, 2004.

[40] R. N. Vigário, "Extraction of ocular artefacts from EEG using independent component analysis," *Electroencephalography and Clinical Neurophysiology*, vol. 103, no. 3, pp. 395–404, 1997.

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

[42] A. Delorme, T. Sejnowski, and S. Makeig, "Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis," *NeuroImage*, vol. 34, no. 4, pp. 1443–1449, 2007.

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

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

[45] G. Barbati, C. Porcaro, F. Zappasodi, P. M. Rossini, and F. Tecchio, "Optimization of an independent component analysis approach for artifact identification and removal in magnetoencephalographic signals," *Clinical Neurophysiology*, vol. 115, no. 5, pp. 1220–1232, 2004.

[46] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "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*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.

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

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