Complex Blind Source Extraction From Noisy Mixtures Using Second-Order Statistics

Soroush Javidi, *Student Member, IEEE*, Danilo P. Mandic, *Senior Member, IEEE*, and
Andrzej Cichocki, *Senior Member, IEEE*

Abstract—A class of second-order complex domain blind source extraction algorithms is introduced to cater for signals with noncircular probability distributions, which is a typical case in real-world scenarios. This is achieved by employing the so-called augmented complex statistics and based on the temporal structures of the sources, thus permitting widely linear (WL) predictability to be the extraction criterion. For rigor, the analysis of the existence and uniqueness of the solution is provided based on both the covariance and the pseudocovariance and for both noise-free and noisy cases, and serves as a platform for the derivation of the algorithms. Both direct solutions and those requiring prewhitening are provided based on a WL predictor, thus making the methodology suitable for the generality of complex signals (both circular and noncircular). Simulations on synthetic noncircular sources support the uniqueness and convergence study, followed by a real-world example of electrooculogram artifact removal from electroencephalogram recordings in real time.

Index Terms—Augmented complex least mean square (ACLMS), blind source extraction (BSE), complex noncircularity, complex pseudocovariance, electroencephalogram (EEG) artifact removal, noisy mixtures, widely linear (WL) model.

I. Introduction

BLIND SOURCE separation (BSS) is an active topic of research in the signal processing community and has found application in a wide range of areas, including biomedical engineering, communications, radar, and sonar [1]. Various methodologies have been discussed in the past two decades to separate latent sources from their linear (and nonlinear) mixtures in both noise-free and noisy environments [2]–[7]. The BSS paradigm aims to find an inverse of the mixing system in order to recover the original sources from the observed mixtures without explicit knowledge of the mixing process or the sources. This is achieved by a variety of optimization methods, including the minimization of mutual information and maximization of likelihood and non-Gaussianity [8], [9].

S. Javidi and D. P. Mandic are with the Communications and Signal Processing Research Group, Department of Electrical and Electronic Engineering, Imperial College London, SW7 2AZ London, U.K. (e-mail: soroush.javidi@imperial.ac.uk; d.mandic@imperial.ac.uk).

A. Cichocki is with the Laboratory for Advanced Brain Signal Processing, Brain Science Institute, RIKEN, Saitama 351-0198, Japan, and also with the IBS, Polish Academy of Science (PAN), 01-477 Warsaw, Poland (e-mail: cia@brain.riken.jp).

Algorithms based on these methods use either deflationary or symmetric orthogonalization procedures to separate the source signals, that is, one by one or simultaneously. However, in some situations, it may be more appropriate to extract only a single source of interest based on a certain fundamental signal property; this procedure is called blind source extraction (BSE) [2]. BSE offers advantages compared with standard BSS methods as it can be used in large-scale problems where only a small subset of sources with specific properties may be of interest. This leads to lower computational complexity and the possibility to relax the need for preprocessing or postprocessing. Algorithms for the BSE of real-valued sources utilize both second- and higher-order statistical properties of signals to discriminate between the sources. Algorithms based on higher-order statistics achieve this by minimizing cost functions typically based on skewness [10] and kurtosis (and generalized kurtosis) [2], [11]–[13]. Alternatively, the predictability of the sources (arising from their temporal structure) leads to another class of algorithms that minimize cost functions based on the mean square prediction error (MSPE) [2], [14], [15].

The algorithms described in [2] and [14] minimize the MSPE at the output of a linear predictor and in an online adaptive manner. While the different sources have different prediction errors, due to the changes in the signal magnitude through mixing, their power levels and thus the prediction errors can vary. The normalized MSPE was thus proposed as an alternative extraction criterion to remove the ambiguity associated with the error power levels [16]. A modified version of this cost function was subsequently used to extract source signals from noisy mixtures based on their temporal features [17], [18].

The recent resurgence of complex domain signal processing [19], [20] has been made possible due to advances in complex statistics [21]–[23]. The enhanced modeling of complex signals and the utilization of the full statistical information available has been achieved by considering both the pseudocovariance $E\{\mathbf{x}\mathbf{x}^T\}$ and the traditional covariance $E\{\mathbf{x}\mathbf{x}^H\}$. In addition, the so-called $\mathbb{CR}$ calculus (also known as Wirtinger calculus) has allowed for another perspective in the analysis of complex functions, specifically those that do not satisfy the stringent Cauchy–Riemann conditions for analytic (holomorphic) functions [24], [25].

The advantages offered by augmented complex statistics have already been exploited in supervised learning, where algorithms such as augmented complex least mean square (ACLMS), widely linear (WL) affine projection algorithm, and WL infinite-impulse-response (IIR) filters have been demonstrated to be suitable for processing the generality of real-world

data [26]–[28]. These algorithms are based on the concept of WL modeling, whereby complex data are modeled as a linear function of both the complex signal and its complex conjugate. In this manner, the available "augmented" second-order information is completely utilized, resulting in improved performance.

In the field of unsupervised learning, complex BSS algorithms [29]–[31] utilizing the augmented complex statistics are WL extensions of the traditional ones [1], [32]. As BSE based on the temporal structure of sources utilizes linear prediction, it is therefore natural to extend the BSE paradigm to the complex domain. This will offer a new perspective and generalization of real-valued BSE methods, as by design these algorithms cater both for noncircular (improper) and second-order circular (proper) sources; they simplify into their real-valued counterparts when operating on real signals.

In our previous work [33], it was shown that by utilizing a WL predictor, it was possible to extract both complex circular and non-circular signals. Generalizing this idea, we here propose a class of algorithms for the blind extraction of the generality of complex-valued sources from both noise-free and noisy mixtures. Algorithms based on prewhitened mixtures are also derived and shown to provide simpler solutions. By considering a general complex doubly white noise model, these algorithms are designed so as to successfully extract sources from noisy mixtures with both circular and noncircular additive noise.

This paper is organized as follows: In Section II, we provide an overview of complex statistics $\mathbb{CR}$ calculus and discuss complex-valued noise and WL modeling. In Section III, the complex BSE problem is introduced, the noise-free cost function and the online algorithm are briefly reviewed, and the cost function and the respective online algorithm for the noisy case are given. Simulations using synthesized complex signals and real-world electroencephalogram (EEG) signals in Section IV demonstrate the performance of the mentioned algorithms followed by analysis of the results. Finally, Section V concludes this paper.

II. PRELIMINARIES

A. Second-Order Statistics of Complex Random Vectors

For a complex random vector $\mathbf{z} = \mathbf{z}_r + \jmath \mathbf{z}_i \in \mathbb{C}^N$, the complete second-order information is provided by the covariance $\mathcal{C}$ and pseudocovariance $\mathcal{P}$ matrices defined as [22]

$$\begin{aligned} \mathcal{C} &= E\{\mathbf{z}\mathbf{z}^H\} \\ \mathcal{P} &= E\{\mathbf{z}\mathbf{z}^T\}. \end{aligned} \quad (1)$$

The covariance $\mathcal{C}$ is a standard complex covariance, whereas the pseudocovariance $\mathcal{P}$ accounts for the correlation between the real and imaginary components. A random vector with a vanishing pseudocovariance is termed second-order circular or proper [21], [23]. In general, the term *circular* refers to a signal with rotation invariant probability distribution, whereas properness (second-order circularity) specifically refers to the second-order statistical properties. Note that the majority of complex signals encountered in signal processing applications[1] are non-

circular. It is therefore necessary to have a complete and unified treatment of such signals when designing complex BSS algorithms.

Complex random vectors are considered uncorrelated if the covariance and pseudocovariance matrices are diagonal [34]. In this case, the diagonal elements of the covariance and pseudovariance matrices are denoted by the variance σ_n^2 and the pseudovariance τ_n^2, which is normally complex valued [35].

B. Brief Overview of $\mathbb{CR}$ Calculus

When dealing with functions in the complex domain, it is required that the Cauchy–Riemann equations are satisfied when calculating the gradients. Most cost functions encountered in signal processing are nonanalytic real-valued functions of complex variables (error power), and thus, standard calculus in $\mathbb{C}$ does not allow for a straightforward evaluation of their derivatives. However, by using $\mathbb{CR}$ calculus, it is possible to perform derivatives of real-valued cost functions directly in $\mathbb{C}$ and in a straightforward manner (for more detail,[2] see [20] and [25]).

In the context of $\mathbb{CR}$ calculus, a cost function $\mathcal{J}(\mathbf{z}) : \mathbb{C}^N \mapsto \mathbb{R}$ can be considered as a function of a complex vector $\mathbf{z}$ and its complex conjugate $\mathbf{z}^*$ such that $\mathcal{J}(\mathbf{z}, \mathbf{z}^*) : \mathbb{C}^N \times \mathbb{C}^N \mapsto \mathbb{R}$, where $\mathbf{z}$ and $\mathbf{z}^*$ are the conjugate coordinates. Alternatively, it can be written in terms of the real and imaginary components of the complex variable $\mathbf{z}$ and given as $\mathcal{J}(\mathbf{x}, \mathbf{y}) : \mathbb{R}^N \times \mathbb{R}^N \mapsto \mathbb{R}$. The derivatives are taken with respect to both $\mathbf{z}$ and $\mathbf{z}^*$ while keeping the other variable constant, that is,

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

It can be shown that the direction of steepest descent is given by the derivative with respect to $\mathbf{z}^*$. The use of $\mathbb{CR}$ calculus is not limited to nonanalytic functions and can be applied for any general complex function. The elegance of this framework lies in the fact that when applied to holomorphic functions, the derivative $\partial/\partial\mathbf{z}^*$ vanishes and so is equal to the standard complex derivative defined based on the Cauchy–Riemann equations ($\mathbb{R}$-derivative), whereas when applied to nonholomorphic functions such as real-valued cost functions, it is equal to the standard pseudogradient ($\mathbb{R}^*$-derivative).

C. Complex-Valued Noise

In real-valued BSS, the additive noise $\mathbf{v}(k)$ is commonly assumed to be white Gaussian and independent of the source signals. In the complex domain, the situation is different, and we shall consider noise in the following two forms [22].

1) *Circular white noise*: While the assumption of whiteness holds for the covariance matrix, the pseudocovariance $\mathcal{P}_{\mathbf{vv}}$ vanishes, that is,

$$\begin{aligned} \mathcal{C}_{\mathbf{vv}}(\delta) &= E\left\{\mathbf{v}(k)\mathbf{v}^H(k-\delta)\right\} \stackrel{\delta\equiv0}{=} \sigma_v^2\mathbf{I} \\ \mathcal{P}_{\mathbf{vv}}(\delta) &= E\left\{\mathbf{v}(k)\mathbf{v}^T(k-\delta)\right\} \stackrel{\delta\equiv0}{=} \mathbf{0} \end{aligned}$$

---

[1]Either those complex by design such as communications signals or those made complex by convenience of representation such as wind and EEG signals [20].

[2]The online material [25] by K. Kreutz-Delgado provides an excellent and comprehensive account of $\mathbb{CR}$ calculus, whereas [20] uses $\mathbb{CR}$ calculus concepts to introduce widely linear adaptive filtering algorithms.

where $\mathbf{I}$ denotes the identity matrix. In other words, the powers $\sigma^2_{\mathbf{v}_r}$ and $\sigma^2_{\mathbf{v}_i}$ of the real and imaginary parts of the complex noise $\mathbf{v}(k) = \mathbf{v}_r(k) + J\mathbf{v}_i(k)$ are equal, and as $E\{\mathbf{v}(k)\mathbf{v}^T(k)\} = E\{\mathbf{v}_r(k)\mathbf{v}_r^T(k)\} - E\{\mathbf{v}_i(k)\mathbf{v}_i^T(k)\} = \mathbf{0}$, the pseudocovariance matrix of proper signals vanishes.

2) *Doubly white noncircular noise*: The condition of whiteness is assumed for both covariance and pseudocovariance matrices so that

$$\mathcal{C}_{\mathbf{vv}}(\delta) \stackrel{\delta=0}{\equiv} \sigma^2_v \mathbf{I}$$

$$\mathcal{P}_{\mathbf{vv}}(\delta) \stackrel{\delta=0}{\equiv} \tau^2_v \mathbf{I}.$$

The power levels and distributions can be different for the real and imaginary parts. Thus, the probability density functions of doubly white noise are in general non-circular (not rotation invariant) and are named after the distribution of the amplitude, as shown in Fig. 1.

The circularity measure ξ is given as the following ratio of the standard deviation of the real to imaginary component of the signal [29], [36], [37]:

$$\xi = \sqrt{\sigma^2_{z_r}/\sigma^2_{z_i}} \tag{3}$$

where the value of $\xi = 1$ indicates equal powers in the real and imaginary components and thus a proper signal, whereas $\xi > 1$ indicates improperness.

D. Widely Linear Modeling

To allow for the design of algorithms that cater for the generality of complex signals, we utilize the WL model [38] for the prediction filter. The WL model is linear in both the complex input $\mathbf{x}$ and its complex conjugate $\mathbf{x}^*$ and can be written as

$$y = \mathbf{h}^T\mathbf{x} + \mathbf{g}^T\mathbf{x}^* \tag{4}$$

where $\mathbf{h}$ and $\mathbf{g}$ are complex-valued weight vectors. The advantage of this approach can be seen by the form of the correlation matrices introduced by the augmented input vector $\mathbf{x}^a = [\mathbf{x}, \mathbf{x}^*]^T$, that is,

$$\mathcal{C}_{\mathbf{x}^a\mathbf{x}^a} = E\begin{bmatrix}\mathbf{x}\\\mathbf{x}^*\end{bmatrix}\begin{bmatrix}\mathbf{x}^H, & \mathbf{x}^T\end{bmatrix} = \begin{bmatrix}\mathcal{C}_{\mathbf{xx}} & \mathcal{P}_{\mathbf{xx}}\\\mathcal{P}^*_{\mathbf{xx}} & \mathcal{C}^*_{\mathbf{xx}}\end{bmatrix}. \tag{5}$$

This demonstrates that the second-order information available in the signal is fully modeled by the WL approach, fully describing noncircular signals that have a nonzero pseudocovariance, whereas for circular signals, the value of $\mathbf{g}$ in (4) is zero and the pseudocovariance vanishes. Utilizing this approach, in [33], a complex BSE algorithm using a WL predictor was shown to outperform the corresponding one based on a linear predictor. We will therefore consider the WL model within BSE schemes for both noise-free and noisy signals.

1) Brief Derivation of the ACLMS Algorithm: Based on the discussion on WL modeling of complex signals and by combining elements of $\mathbb{CR}$ calculus, we briefly provide the derivation of the ACLMS algorithm [26], [39], which is the basis for our proposed prediction filter.

Consider the complex-valued input $\mathbf{x}(k)$ and a finite-impulse-response WL filter in the prediction setting, with filter coeffi-
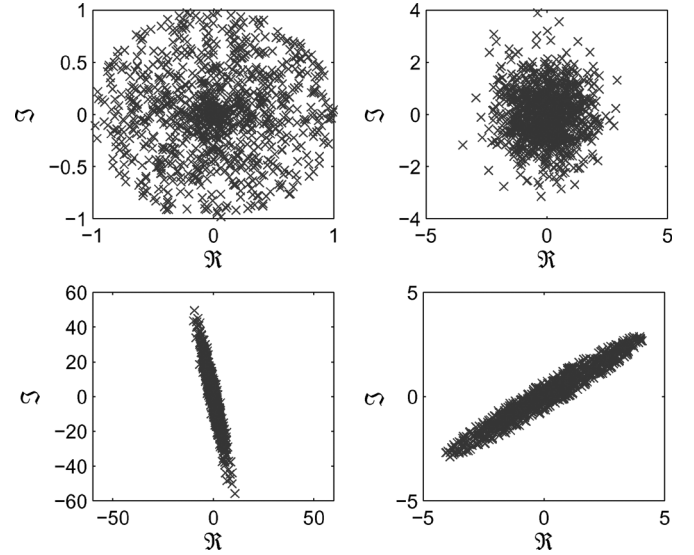

Fig. 1. Scatter plots of complex white noise realizations. (Top row, left) Circular uniform noise and (right) circular Gaussian noise. (Bottom row, left) Non-circular uniform noise ($\xi = 5$) and (right) noncircular Gaussian noise ($\xi = 16$). The circularity measure ξ is defined in (3).

cients $\mathbf{h}$ and $\mathbf{g}$, output $y(k)$, and the desired signal $d(k)$. By using an adaptive version of the WL model (4), the error signal $e(k)$ is given by

$$e(k) = d(k) - y(k) = d(k) - \mathbf{h}^T(k)\mathbf{x}(k) - \mathbf{g}^T(k)\mathbf{x}^*(k). \tag{6}$$

The corresponding cost function $\mathcal{J}_{\text{ACLMS}} = |e(k)|^2 = e(k)e^*(k)$ is minimized with respect to the two coefficient vectors of the WL filter using a steepest-descent adaptation given by

$$\mathbf{h}(k+1) = \mathbf{h}(k) - \mu_h\nabla_{\mathbf{h}^*}\mathcal{J}_{\text{ACLMS}} \tag{7}$$

$$\mathbf{g}(k+1) = \mathbf{g}(k) - \mu_g\nabla_{\mathbf{g}^*}\mathcal{J}_{\text{ACLMS}}. \tag{8}$$

Recall that the direction of steepest descent is given by the $\mathbb{R}^*$-derivative for both update equations. By using $\mathbb{CR}$ calculus and the chain rule,[3] we then simply calculate

$$\nabla_{\mathbf{h}^*} = -e(k)\mathbf{x}^*(k)$$

$$\nabla_{\mathbf{g}^*} = -e(k)\mathbf{x}(k)$$

and substitute in (7) and (8) to form the complete update equations for the ACLMS algorithm [20]

$$\mathbf{h}(k+1) = \mathbf{h}(k) + \mu_h e(k)\mathbf{x}^*(k) \tag{9}$$

$$\mathbf{g}(k+1) = \mathbf{g}(k) + \mu_g e(k)\mathbf{x}(k). \tag{10}$$

It is also possible to consider the complex vectors as "augmented" vectors given by the pair of complex vector and its complex conjugate, to obtain

$$\mathbf{w}^a(k+1) = \mathbf{w}^a(k) + \eta e^a(k)\mathbf{x}^{a^*}(k) \tag{11}$$

[3]For a complex vector-valued composite function $\mathbf{f} \circ \mathbf{g}$, the chain rule states that $(\partial\mathbf{f}(\mathbf{g})/\partial\mathbf{z}) = (\partial\mathbf{f}/\partial\mathbf{g})(\partial\mathbf{g}/\partial\mathbf{z}) + (\partial\mathbf{f}/\partial\mathbf{g}^*)(\partial\mathbf{g}^*/\partial\mathbf{z})$ and $(\partial\mathbf{f}(\mathbf{g})/\partial\mathbf{z}^*) = (\partial\mathbf{f}/\partial\mathbf{g})(\partial\mathbf{g}/\partial\mathbf{z}^*) + (\partial\mathbf{f}/\partial\mathbf{g}^*)(\partial\mathbf{g}^*/\partial\mathbf{z}^*)$ [25].

where $\eta = \mu_h = \mu_g$, $\mathbf{w}^a(k) = [\mathbf{h}^T(k), \mathbf{g}^T(k)]^T$ is the augmented coefficient vector, $\mathbf{x}^a(k) = [\mathbf{x}^T(k), \mathbf{x}^H(k)]^T$ is the augmented input vector, and $e^a(k) = d(k) - \mathbf{x}^{a^T}(k)\mathbf{w}^a(k)$ is a complex scalar value measuring the distance of the output of the predictor to the desired signal.

The performance of this algorithm for the prediction of complex-valued wind signals was analyzed in [39] and demonstrated superiority over the standard complex least mean square (CLMS) algorithm in the prediction of noncircular signals. As derived, in the prediction of circular signals, the WL filter behaves in a similar fashion to a standard filter as the "conjugate" part of the update (10), that is, $\mathbf{g}$, vanishes to form the standard CLMS algorithm. This makes the ACLMS an ideal candidate for BSE-based linear prediction of both proper and improper signals in the complex domain.

III. COMPLEX BSE OF NOISE-FREE AND NOISY MIXTURES

A. Normalized MSPE

The mixing vector $\mathbf{x}(k) \in \mathbb{C}^N$ at time index k is observed from the linear mixture of the complex sources $\mathbf{s}(k) \in \mathbb{C}^{N_s}$ as

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

where $\mathbf{A} \in \mathbb{C}^{N \times N_s}$ is the mixing matrix, and $\mathbf{v}(k) \in \mathbb{C}^N$ denotes the additive noise. Here, it is assumed that the number of observations is equal to that of the sources. The next section shows how the overdetermined case can be used for the estimation of the second-order statistics of the noise $\mathbf{v}(k)$.

The sources $\mathbf{s}(k)$ are assumed to be stationary and spatially uncorrelated with unit variance and zero mean, with no assumptions regarding their second-order circularity. For a lag δ, we can therefore formulate the covariance $\mathcal{C}_{\mathbf{ss}}$ and pseudocovariance $\mathcal{P}_{\mathbf{ss}}$ as

$$
\begin{aligned}
\mathcal{C}_{\mathbf{ss}}(\delta) &= E\left\{\mathbf{s}(k)\mathbf{s}^H(k-\delta)\right\} \\
&= \operatorname{diag}\left(\sigma_1^2(\delta), \ldots, \sigma_{N_s}^2(\delta)\right) \\
\mathcal{P}_{\mathbf{ss}}(\delta) &= E\left\{\mathbf{s}(k)\mathbf{s}^T(k-\delta)\right\} \\
&= \operatorname{diag}\left(\tau_1^2(\delta), \ldots, \tau_{N_s}^2(\delta)\right).
\end{aligned} \tag{13}
$$

Fig. 2 shows the blind extraction architecture for complex signals based on the minimization of the MSPE. For the observation vector $\mathbf{x}(k)$, the extracted signal $y(k)$ is formed as

$$y(k) = \mathbf{w}^H\mathbf{x}(k). \tag{14}$$

The aim of the demixing process is to find a demixing vector $\mathbf{w}$ such that $\mathbf{u}^H = \mathbf{w}^H\mathbf{A} = [0, \ldots, u_n, \ldots, 0]$ and thus extract only a single source with the smallest MSPE. The prediction error is given by

$$e(k) = y(k) - y_{WL}(k) \tag{15}$$

where $y_{WL}(k)$ denotes the output of the prediction filter and given by

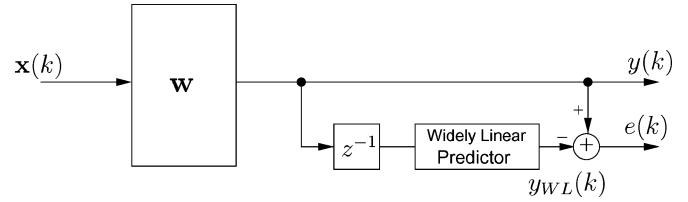$$y_{WL}(k) = \mathbf{h}^T(k)\mathbf{y}(k) + \mathbf{g}^T(k)\mathbf{y}^*(k) \tag{16}$$

Fig. 2. Complex BSE algorithm using a WL predictor.

where $\mathbf{h}(k)$ and $\mathbf{g}(k)$ are the coefficient vectors of length M, and $\mathbf{y}(k)$ is a delayed version of the extracted signal given by $\mathbf{y}(k) = [y(k-1), \ldots, y(k-M)]^T$. The length M of the filter affects the performance of the predictor such that sources with rapid variations can be extracted using a short tap length while smoother sources require a much larger tap length [14]. By updating the coefficient vectors adaptively, it is possible to introduce the largest relative difference in the MSPE as a criterion[4] for extraction [16].

The MSPE $E\{|e(k)|^2\}$ can then be calculated as

$$
\begin{aligned}
E\left\{|e(k)|^2\right\} &= E\left\{e(k)e^*(k)\right\} \\
&= \mathbf{w}^H\mathbf{A}\hat{\mathcal{C}}_{\mathbf{ss}}\mathbf{A}^H\mathbf{w} + \Re\left\{\mathbf{w}^H\mathbf{A}\hat{\mathcal{P}}_{\mathbf{ss}}\mathbf{A}^T\mathbf{w}^*\right\}
\end{aligned} \tag{17}
$$

where

$$
\begin{aligned}
\hat{\mathcal{C}}_{\mathbf{ss}} &= \mathcal{C}_{\mathbf{ss}}(0) - 2\Re\left\{\sum_{m=1}^{M} h_m^*(k)\mathcal{C}_{\mathbf{ss}}(m)\right\} \\
&\quad + \sum_{m,\ell=1}^{M} [h_m(k)h_\ell^*(k) + g_m^*(k)g_\ell(k)]\mathcal{C}_{\mathbf{ss}}(\ell - m) \\
\hat{\mathcal{P}}_{\mathbf{ss}} &= -2\sum_{m=1}^{M} g_m^*(k)\mathcal{P}_{\mathbf{ss}}(m) \\
&\quad + 2\sum_{m,\ell=1}^{M} h_m(k)g_\ell^*(k)\mathcal{P}_{\mathbf{ss}}(\ell - m).
\end{aligned}
$$

The operator $\Re\{\cdot\}$ denotes the real part of a complex quantity. We can see that the prediction error is a function of both covariance and pseudocovariance of the sources, and as the sources are assumed uncorrelated, $\hat{\mathcal{C}}_{\mathbf{ss}}$ and $\hat{\mathcal{P}}_{\mathbf{ss}}$ are diagonal matrices with the value of the nth element corresponding to the error of the nth source $s_n(k)$. Denoting this value by $e_n(k)$, the MSPE relating to $s_n(k)$ is given as

$$
\begin{aligned}
E\left\{|e_n(k)|^2\right\} &= E\left\{\left|s_n(k) - \mathbf{h}^T(k)\mathbf{s}_n(k)\right|^2 \right. \\
&\quad - 2\Re\left\{\left(s_n(k) - \mathbf{h}^T(k)\mathbf{s}_n(k)\right)\left(\mathbf{g}^H(k)\mathbf{s}_n(k)\right)\right\} \\
&\quad \left. + \left|\mathbf{g}^H(k)\mathbf{s}_n(k)\right|^2\right\}
\end{aligned} \tag{18}
$$

where $\mathbf{s}_n(k) = [s_n(k-1), \ldots, s_n(k-M)]^T$. Due to the vanishing pseudocovariance ($\mathcal{P}_{\mathbf{ss}} = \mathbf{0}$) of complex circular sources, the MSPE in (17) and that given in (18) for the nth source simplify and are only functions of the covariance matrix.

---

[4]It is also possible to assign fixed values to the coefficient vectors $\mathbf{h}$ and $\mathbf{g}$; however, this results in poorer performance.

A complete derivation of the extraction based on MSPE as the extraction criterion is given in the Appendix.

B. Noise-Free Complex BSE

1) Cost Function: The algorithms derived for the complex BSE of noise-free mixtures are based on a cost function that minimizes the normalized MSPE. As described in [16], the variation in the magnitude of source signals results in an ambiguity of the power levels, and so algorithms based on the minimization of the MSPE cannot effectively extract a source of interest. This can be seen by considering (17) and noticing that changes in the values of $\hat{\mathcal{C}}_{\mathbf{ss}}$ and $\hat{\mathcal{P}}_{\mathbf{ss}}$ can effectively be absorbed into the mixing matrix, thus enabling the minimization independent of the source power levels. This way, by using the MSPE, this ambiguity is removed as different signals exhibit different degrees of normalized predictability despite the varying power levels.

The normalized MSPE cost function is therefore given as

$$\mathcal{J}_1(\mathbf{w}, \mathbf{h}, \mathbf{g}) = \frac{E\left\{|e(k)|^2\right\}}{E\left\{|y(k)|^2\right\}} \tag{19}$$

where $\mathcal{J}_1 \in \mathbb{R}$ and is a function of the demixing vector and the coefficient vectors. In the noise-free case, the optimization problem for the demixing vector can be expressed as

$$\mathbf{w}_{\mathrm{opt}} = \arg \max_{\|\mathbf{w}\|_2=1} \mathcal{J}_1(\mathbf{w}, \mathbf{h}, \mathbf{g}) \tag{20}$$

where the norm of $\mathbf{w}$ is constrained to unity, and $\mathbf{u}_{\mathrm{opt}}^H = \mathbf{w}_{\mathrm{opt}}^H \mathbf{A}$ only has a single nonzero value with unit magnitude that corresponds to the source with the smallest normalized MSPE. This can be illustrated by observing the cost function in (19) and its components in (17), that is,

$$E\left\{|y(k)|^2\right\} = \mathbf{w}^H \mathcal{C}_{\mathbf{xx}}(0)\mathbf{w}$$
$$= \mathbf{w}^H \mathbf{A} \mathcal{C}_{\mathbf{ss}}(0)\mathbf{A}^H \mathbf{w} + \mathbf{w}^H \mathcal{C}_{\mathbf{vv}}(0)\mathbf{w} \tag{21}$$

and noting that the sources have unit variance and the noise variance is zero. The cost function [see (19)] then becomes

$$\mathcal{J}_1(\mathbf{w}, \mathbf{h}, \mathbf{g}) = \frac{\mathbf{u}^H \hat{\mathcal{C}}_{\mathbf{ss}} \mathbf{u} + \Re\{\mathbf{u}^H \hat{\mathcal{P}}_{\mathbf{ss}} \mathbf{u}^*\}}{\mathbf{u}^H \mathbf{u}}. \tag{22}$$

Consider a new variable $\breve{\mathbf{u}} = \mathbf{u}/\|\mathbf{u}\|$ and the associated cost function

$$\mathcal{J}_1(\mathbf{w}, \mathbf{h}, \mathbf{g}) = \breve{\mathbf{u}}^H \hat{\mathcal{C}}_{\mathbf{ss}} \breve{\mathbf{u}} + \Re\{\breve{\mathbf{u}}^H \hat{\mathcal{P}}_{\mathbf{ss}} \breve{\mathbf{u}}^*\} \tag{23}$$

where $\breve{\mathbf{u}}^H \breve{\mathbf{u}} = 1$. With this constraint, the minimum of (23) is a vector $\breve{\mathbf{u}}_{\mathrm{opt}}$ with a single nonzero element with arbitrary phase and unit magnitude at a position corresponding to the smallest combination of the diagonal elements of $\hat{\mathcal{C}}_{\mathbf{ss}}$ and $\Re\{\hat{\mathcal{P}}_{\mathbf{ss}}\}$. In the case of circular sources, this argument simplifies so that only the smallest diagonal element of $\hat{\mathcal{C}}_{\mathbf{ss}}$ is considered. This solution is similar for $\mathbf{u}_{\mathrm{opt}}$ with only a single nonzero value. Likewise, the optimal value of the demixing vector can be recovered as $\mathbf{w}_{\mathrm{opt}} = \mathbf{A}^{H\#} \mathbf{u}_{\mathrm{opt}}$, where the symbol $(\cdot)^{\#}$ denotes the matrix pseudoinverse. As described in [17], if a value $\mathbf{w}_{\mathrm{opt}}$ exists such

that $\mathbf{u}$ and hence $\breve{\mathbf{u}}$ respectively assume their optimal value to be $\mathbf{u}_{\mathrm{opt}}$ and $\breve{\mathbf{u}}_{\mathrm{opt}}$, then the cost function of (19) can successfully be minimized with respect to $\mathbf{w}$.

2) Algorithms for the Noise-Free Case: We will use a gradient-descent approach to update the values of the demixing vector $\mathbf{w}$ and the coefficient vectors $\mathbf{h}$ and $\mathbf{g}$. As mentioned earlier, the value of the demixing vector is constrained to unit norm and is normalized after each update. The complex gradients are thus calculated as

$$\nabla_{\mathbf{w}^*} \mathcal{J}_1 = \left[e^*(k)\hat{\mathbf{x}}_h(k) - e(k)\hat{\mathbf{x}}_g(k) \right.$$
$$\left. - \frac{\sigma_e^2(k)}{\sigma_y^2(k)} y^*(k)\mathbf{x}(k) \right] \frac{1}{\sigma_y^2(k)}$$
$$\nabla_{\mathbf{h}^*} \mathcal{J}_1 = \frac{-1}{\sigma_y^2(k)} e(k)\mathbf{y}^*(k)$$
$$\nabla_{\mathbf{g}^*} \mathcal{J}_1 = \frac{-1}{\sigma_y^2(k)} e(k)\mathbf{y}(k) \tag{24}$$

where

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

and the MSPE $\sigma_e^2(k)$ and variance of the extracted signal $\sigma_y^2(k)$ are estimated by an online moving average relation [2]

$$\sigma_e^2(k) = \beta_e \sigma_e^2(k-1) + (1 - \beta_e)|e(k)|^2$$
$$\sigma_y^2(k) = \beta_y \sigma_y^2(k-1) + (1 - \beta_y)|y(k)|^2 \tag{25}$$

where β_e and β_y are the corresponding forgetting factors for the MSPE and the signal power.

The update algorithm [that is, second-order complex domain blind source extraction algorithm (P-cBSE)] for the demixing vector $\mathbf{w}$ is given as

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \mu \left[e^*(k)\hat{\mathbf{x}}_h(k) - e(k)\hat{\mathbf{x}}_g(k) \right.$$
$$\left. - \frac{\sigma_e^2(k)}{\sigma_y^2(k)} y^*(k)\mathbf{x}(k) \right] \frac{1}{\sigma_y^2(k)},$$
$$\mathbf{w}(k+1) \leftarrow \frac{\mathbf{w}(k+1)}{\|\mathbf{w}(k+1)\|_2} \tag{26}$$

for the noise-free case, and the filter coefficient updates are given by

$$\mathbf{h}(k+1) = \mathbf{h}(k) + \mu_h \frac{1}{\sigma_y^2(k)} e(k)\mathbf{y}^*(k) \tag{27}$$

$$\mathbf{g}(k+1) = \mathbf{g}(k) + \mu_g \frac{1}{\sigma_y^2(k)} e(k)\mathbf{y}(k). \tag{28}$$

From the expressions for gradients $\nabla_{\mathbf{h}^*} \mathcal{J}_1$ and $\nabla_{\mathbf{g}^*} \mathcal{J}_1$ in (24), we notice that the update equations (27) and (28) can be combined to form a normalized ACLMS-type adaptation [26]. Recall that for circular sources, the pseudocovariance matrix vanishes; thus, a standard complex linear predictor (say based on

CLMS) can be used. However, this case is already incorporated within the WL predictor as, e.g., the conjugate part of the ACLMS weight vector vanishes for circular data ($\mathbf{g} = \mathbf{0}$), demonstrating the flexibility of the proposed approach.

One way to remove the effects of source power ambiguity is to prewhiten the observation vector $\mathbf{x}(k)$ to make the power levels of the output (extracted) signals constant. This also helps to orthogonalize an ill-conditioned mixing matrix; however, performing prewhitening for an online algorithm is not convenient. Denoting the prewhitening matrix $\mathbf{V} = \mathbf{D}^{-1/2}\mathbf{E}$, where $\mathbf{D}$ is a diagonal matrix containing the eigenvalues of $\mathcal{C}_{\mathbf{xx}}(0)$, and $\mathbf{E}$ is an orthogonal matrix whose columns are the eigenvectors of $\mathcal{C}_{\mathbf{xx}}(0)$, the covariance matrix $\mathcal{C}_{\breve{\mathbf{x}}\breve{\mathbf{x}}}(0) = \mathbf{V}\mathcal{C}_{\mathbf{xx}}\mathbf{V}^H = \mathbf{I}$; the symbol $\breve{x}(k)$ denotes a prewhitened observation vector. From (21) and the constraint on the norm of $\mathbf{w}$, that is, $E\{|y(k)|^2\} = \mathbf{w}^H\mathbf{w} = 1$, the cost function in (19) can be simplified to

$$\nabla_{\mathbf{w}^*}\mathcal{J}_1 = [e^*(k)\hat{\mathbf{x}}_h(k) - e(k)\hat{\mathbf{x}}_g(k)]$$
$$\nabla_{\mathbf{h}^*}\mathcal{J}_1 = -e(k)\mathbf{y}^*(k)$$
$$\nabla_{\mathbf{g}^*}\mathcal{J}_1 = -e(k)\mathbf{y}(k). \tag{29}$$

The resulting coefficient updates

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \mu\left[e^*(k)\hat{\mathbf{x}}_h(k) - e(k)\hat{\mathbf{x}}_g(k)\right] \tag{30}$$
$$\mathbf{h}(k+1) = \mathbf{h}(k) + \mu_h e(k)\mathbf{y}^*(k) \tag{31}$$
$$\mathbf{g}(k+1) = \mathbf{g}(k) + \mu_h e(k)\mathbf{y}(k) \tag{32}$$

are simpler than those in (26)–(28), and the coefficients of the WL predictor in (31) and (32) are updated using the ACLMS algorithm.

C. Noisy Complex BSE

1) Cost Function: The algorithms described above do not account for the effect of the additive noise $\mathbf{v}(k)$ and thus underperform for the extraction of sources from noisy mixtures. By modifying the cost function, it is possible to derive a new class of algorithms for the extraction of complex sources from noisy mixtures. We can make use of the modified cost function described in [17], which employs a normalized MSPE-type cost function, to remove the effect of noise from the MSPE and output variance.

Taking a closer look at the covariance and pseudocovariance of the observation vector with additive noise

$$\mathcal{C}_{\mathbf{xx}}(\delta) = \mathbf{A}\mathcal{C}_{\mathbf{ss}}(\delta)\mathbf{A}^H + \mathcal{C}_{\mathbf{vv}}(\delta)$$
$$\mathcal{P}_{\mathbf{xx}}(\delta) = \mathbf{A}\mathcal{P}_{\mathbf{ss}}(\delta)\mathbf{A}^T + \mathcal{P}_{\mathbf{vv}}(\delta) \tag{33}$$

we note that the MSPE can be divided into two parts e_s^2 and e_v^2, where the first term is related to the MSPE relevant to the sources in (17), and the second term pertains to that of the noise, and so $E\{|e(k)|^2\} = e_s^2 + e_v^2$. The expression for e_v^2 is derived in the Appendix. The cost function for the noisy BSE thus becomes

$$\mathcal{J}_2(\mathbf{w}, \mathbf{h}, \mathbf{g}) = \frac{E\left\{|e(k)|^2\right\} - e_v^2}{E\left\{|y(k)|^2\right\} - \mathbf{w}^H\mathcal{C}_{\mathbf{vv}}(0)\mathbf{w}}$$
$$= \frac{E\left\{|e(k)|^2\right\} - c_1\sigma_v^2 - \Re\left\{c_2\tau_v^2\mathbf{w}^H\mathbf{w}^*\right\}}{E\left\{|y(k)|^2\right\} - \sigma_v^2} \tag{34}$$

where the signal variance is given in (21). The existence of a solution to the minimization of the cost function can be considered similarly to that of the noiseless case. By removing the effect of noise from $\mathcal{J}_2$, the resultant cost function is expanded exactly as in (22), and a similar argument can be used for the analysis.

2) Algorithms for the Noisy Case: The cost function in (34) is minimized using steepest descent, and the coefficient vectors $\mathbf{w}$, $\mathbf{h}$, and $\mathbf{g}$ are updated via an online algorithm, similar to the noise-free case. The corresponding equations are calculated as

$$\nabla_{\mathbf{w}^*}\mathcal{J}_2 = \frac{1}{\sigma_y^2(k) - \sigma_v^2}$$
$$\times \left[e^*(k)\hat{\mathbf{x}}_h(k) - e(k)\hat{\mathbf{x}}_g(k) - \Re\left\{c_2\tau_v^2\right\}\mathbf{w}^*(k)\right.$$
$$- \frac{\sigma_e^2(k) - c_1\sigma_v^2 - \Re\left\{c_2\tau_v^2\mathbf{w}^H(k)\mathbf{w}^*(k)\right\}}{\sigma_y^2(k) - \sigma_v^2}$$
$$\left.\times y^*(k)\mathbf{x}(k)\right]$$
$$\nabla_{\mathbf{h}^*}\mathcal{J}_2 = \frac{-1}{\sigma_y^2(k) - \sigma_v^2}\left(e(k)\mathbf{y}^*(k) + \sigma_v^2\mathbf{h}(k)\right)$$
$$\nabla_{\mathbf{g}^*}\mathcal{J}_2 = \frac{-1}{\sigma_y^2(k) - \sigma_v^2}$$
$$\times \left(e(k)\mathbf{y}(k) + \sigma_v^2\mathbf{g}(k) + \Re\left\{\mathbf{w}^H\mathbf{w}^*\tau_v^2\right\}\mathbf{h}(k)\right) \tag{35}$$

where

$$c_1 = 1 + \mathbf{h}^H(k)\mathbf{h}(k) + \mathbf{g}^H(k)\mathbf{g}(k)$$
$$c_2 = 2\mathbf{g}^H(k)\mathbf{h}(k)$$

and the demixing vector $\mathbf{w}(k+1)$ is normalized after each update so that

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

For circular white additive noise, the pseudovariance τ_v^2 is zero and thus the terms related to the pseudocovariance. It is apparent that the estimation of noise variance and pseudovariance is necessary for the operation of this BSE method, as discussed in the next section. The algorithms for the BSE of noisy mixtures are given as

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \frac{\mu}{\sigma_y^2(k) - \sigma_v^2}$$
$$\times \left[e^*(k)\hat{\mathbf{x}}_h(k) - e(k)\hat{\mathbf{x}}_g(k) - \Re\left\{c_2\tau_v^2\right\}\mathbf{w}^*(k)\right.$$
$$- \frac{\sigma_e^2(k) - c_1\sigma_v^2 - \Re\left\{c_2\tau_v^2\mathbf{w}^H(k)\mathbf{w}^*(k)\right\}}{\sigma_y^2(k) - \sigma_v^2}$$
$$\left.\times y^*(k)\mathbf{x}(k)\right] \tag{36}$$
$$\mathbf{h}(k+1) = \mathbf{h}(k) + \mu_h\frac{1}{\sigma_y^2(k) - \sigma_v^2}\left(e(k)\mathbf{y}^*(k) + \sigma_v^2\mathbf{h}(k)\right) \tag{37}$$
$$\mathbf{g}(k+1) = \mathbf{g}(k) + \mu_g\frac{1}{\sigma_y^2(k) - \sigma_v^2}$$
$$\times \left(e(k)\mathbf{y}(k) + \sigma_v^2\mathbf{g}(k) + \Re\left\{\mathbf{w}^H\mathbf{w}^*\tau_v^2\right\}\mathbf{h}(k)\right) \tag{38}$$

The case of the prewhitened observation vector $\breve{\mathbf{x}}$ is next considered, where the variance of the extracted signal is constant, and the resulting algorithms are somewhat simpler. The prewhitened covariance and pseudocovariance are now given as

$$
\begin{aligned}
\mathcal{C}_{\breve{\mathbf{x}}\breve{\mathbf{x}}}(0) &= \breve{\mathbf{A}}\mathcal{C}_{\mathbf{ss}}(0)\breve{\mathbf{A}}^H + \breve{\mathcal{C}}_{\mathbf{vv}}(0) = \mathbf{I} \\
\mathcal{P}_{\breve{\mathbf{x}}\breve{\mathbf{x}}}(0) &= \breve{\mathbf{A}}\mathcal{P}_{\mathbf{ss}}(0)\breve{\mathbf{A}}^T + \breve{\mathcal{P}}_{\mathbf{vv}}(0)
\end{aligned}
\tag{39}
$$

with $\breve{\mathbf{A}} = \mathbf{VA}$, $\breve{\mathcal{C}}_{\mathbf{vv}}(0) = \mathbf{V}\mathcal{C}_{\mathbf{vv}}(0)\mathbf{V}^H$, and $\breve{\mathcal{P}}_{\mathbf{vv}}(0) = \mathbf{V}\mathcal{P}_{\mathbf{vv}}(0)\mathbf{V}^T$. It is possible to use a strong uncorrelating transform (SUT) [34] to whiten the covariance matrix and diagonalize the pseudocovariance such that $\mathcal{P}_{\breve{\mathbf{x}}\breve{\mathbf{x}}}(0) = \mathbf{\Lambda}$ contains nonnegative real values. In the case of circular signals, the SUT simplifies to a standard whitening.

This way, the term e_v^2 can be expanded as

$$
e_v^2 = \mathbf{w}^H \hat{\breve{\mathcal{C}}}_{\mathbf{vv}}\mathbf{w} + \Re\{\mathbf{w}^H \hat{\breve{\mathcal{P}}}_{\mathbf{vv}}\mathbf{w}^*\}
\tag{40}
$$

and the variance of the extracted signal

$$
\begin{aligned}
E\left\{|y(k)|^2\right\} &= \mathbf{w}^H \mathcal{C}_{\breve{\mathbf{x}}\breve{\mathbf{x}}}(0)\mathbf{w} \\
&= \mathbf{w}^H \breve{\mathbf{A}}\mathcal{C}_{\mathbf{ss}}(0)\breve{\mathbf{A}}^H\mathbf{w} + \mathbf{w}^H \breve{\mathcal{C}}_{\mathbf{vv}}(0)\mathbf{w} \\
&= \mathbf{w}^H\mathbf{w} = 1.
\end{aligned}
\tag{41}
$$

The cost function in (34) can thus be rewritten as

$$
\begin{aligned}
\mathcal{J}_2 &= \frac{E\left\{|e(k)|^2\right\} - \mathbf{w}^H \hat{\breve{\mathcal{C}}}_{\mathbf{vv}}\mathbf{w} + \Re\{\mathbf{w}^H \hat{\breve{\mathcal{P}}}_{\mathbf{vv}}\mathbf{w}^*\}}{E\left\{|y(k)|^2\right\} - \mathbf{w}^H \breve{\mathcal{C}}_{\mathbf{vv}}(0)\mathbf{w}} \\
&= \frac{E\left\{|e(k)|^2\right\} - c_1\mathbf{w}^H \hat{\breve{\mathcal{C}}}_{\mathbf{vv}}(0)\mathbf{w} + \Re\{c_2\mathbf{w}^H \hat{\breve{\mathcal{P}}}_{\mathbf{vv}}(0)\mathbf{w}^*\}}{\mathbf{w}^H\mathbf{w} - \mathbf{w}^H \breve{\mathcal{C}}_{\mathbf{vv}}(0)\mathbf{w}} \\
&= E\left\{|e(k)|^2\right\} - c_1 E\left\{|y(k)|^2\right\} \\
&\quad - \Re\left\{c_2\tau_v^2\mathbf{w}^H\mathbf{V}\mathbf{V}^T\mathbf{w}^*\right\} + 1
\end{aligned}
\tag{42}
$$

where the demixing vector is normalized as

$$
\mathbf{w}(k+1) \leftarrow \frac{\mathbf{w}(k+1)}{\sqrt{\mathbf{w}(k+1)^H\left[\mathbf{I} - \breve{\mathcal{C}}_{\mathbf{vv}}(0)\right]\mathbf{w}(k+1)}}.
\tag{43}
$$

The gradients within the updates of the online algorithms for noisy BSE can be calculated as

$$
\begin{aligned}
\nabla_{\mathbf{w}^*}\mathcal{J}_2 &= e^*(k)\hat{\mathbf{x}}_h(k) - e(k)\hat{\mathbf{x}}_g(k) \\
&\quad - \Re\left\{c_2\tau_v^2\right\}\mathbf{V}^T\mathbf{V}\mathbf{w}^*(k) - c_1 y^*(k)\mathbf{x}(k) \\
\nabla_{\mathbf{h}^*}\mathcal{J}_2 &= -e(k)\mathbf{y}^*(k) \\
\nabla_{\mathbf{g}^*}\mathcal{J}_2 &= -\left(e(k)\mathbf{y}(k) + \Re\{\tau_v^2\mathbf{w}^H\mathbf{V}\mathbf{V}^T\mathbf{w}^*\}\mathbf{h}(k)\right)
\end{aligned}
\tag{44}
$$

to form the final online update for the BSE of prewhitened noisy mixtures, with the update algorithm for the demixing vector given as

$$
\begin{aligned}
\mathbf{w}(k+1) &= \mathbf{w}(k) - \mu e^*(k)\hat{\mathbf{x}}_h(k) - e(k)\hat{\mathbf{x}}_g(k) \\
&\quad - \Re\left\{c_2\tau_v^2\right\}\mathbf{V}^T\mathbf{V}\mathbf{w}^*(k) - c_1 y^*(k)\mathbf{x}(k)
\end{aligned}
\tag{45}
$$

with the update equations for the filter coefficient vectors given by

$$
\mathbf{h}(k+1) = \mathbf{h}(k) + \mu_h e(k)\mathbf{y}^*(k)
\tag{46}
$$

$$
\begin{aligned}
\mathbf{g}(k+1) = \mathbf{g}(k) + \mu_g &\left(e(k)\mathbf{y}(k) \right. \\
&\left. + \Re\left\{\tau_v^2\mathbf{w}^H\mathbf{V}\mathbf{V}^T\mathbf{w}^*\right\}\mathbf{h}(k)\right).
\end{aligned}
\tag{47}
$$

D. Remark on the Estimation of Noise Variance and Pseudovariance

The adaptive algorithms derived in the previous section require estimation of the noise variance and pseudovariance for their operation. As mentioned in Section II, the noise is considered to have a constant and equal variance σ_v^2 and pseudovariance τ_v^2 so that $\mathcal{C}_{\mathbf{vv}} = \sigma_v^2\mathbf{I}$ and $\mathcal{P}_{\mathbf{vv}} = \tau_v^2\mathbf{I}$. Furthermore, two variants of complex noise were discussed: circular white noise and doubly white noise. One possible method for the estimation of the variance of circular white noise is by means of a subspace method [40] and can intuitively be extended for the estimation of the pseudovariance of doubly white noise, as detailed below.

Consider the number of observations larger than that of the sources ($N > N_s$). It is then possible to estimate the noise variance and pseudovariance based on

$$
\begin{aligned}
\mathcal{C}_{\mathbf{xx}} &= \mathbf{A}\mathcal{C}_{\mathbf{ss}}\mathbf{A}^H + \mathcal{C}_{\mathbf{vv}} = \mathbf{\Theta} + \sigma_v^2\mathbf{I} \\
\mathcal{P}_{\mathbf{xx}} &= \mathbf{A}\mathcal{P}_{\mathbf{ss}}\mathbf{A}^T + \mathcal{P}_{\mathbf{vv}} = \mathbf{\Xi} + \tau_v^2\mathbf{I}.
\end{aligned}
\tag{48}
$$

For both cases, by assuming that the matrix $\mathbf{A}$ is of full column rank $\text{Rank}(\mathbf{A}) = N_s$ and that $\mathbf{s}$ is nonsingular, then $\text{Rank}(\mathbf{\Theta}) = \text{Rank}(\mathbf{\Xi}) = N_s$, and so the $(N - N_s)$ smallest eigenvalues of $\mathbf{\Theta}$ and $\mathbf{\Xi}$ are zero. Hence, the $(N - N_s)$ smallest eigenvalues of $\mathcal{C}_{\mathbf{xx}}$ and $\mathcal{P}_{\mathbf{xx}}$ are respectively equal to σ_v^2 and τ_v^2.

IV. SIMULATIONS AND DISCUSSION

A. Performance Analysis for Synthetic Data

The performances of the proposed algorithms were analyzed using sources with different degrees of noncircularity and distributions and in various simulation settings comprising of noise-free and noisy mixtures. The performances of the algorithms were measured using the performance index (PI) [2], which for $\mathbf{u} = \mathbf{A}^H\mathbf{w} = [u_1, \ldots, u_M]$ is given as

$$
\text{PI} = 10\log_{10}\left(\frac{1}{M}\left(\sum_{i=1}^{M} \frac{|u_i|^2}{\max\{|u_1|^2, \ldots, |u_M|^2\}}\right)\right)
\tag{49}
$$

and indicates the closeness of $\mathbf{u}$ to having only a single nonzero element.

The values of the step sizes μ, μ_h and μ_g were set empirically, the mixing matrix $\mathbf{A}$ was generated randomly, and in all the experiments the forgetting factors $\beta_e = \beta_y = 0.975$. The additive noise $\mathbf{v}(k)$ had a Gaussian distribution in two variants of proper white ($\xi = 1$) and doubly white improper ($\xi = 5$). Its variance and pseudovariance were estimated using the subspace method [see (48)].
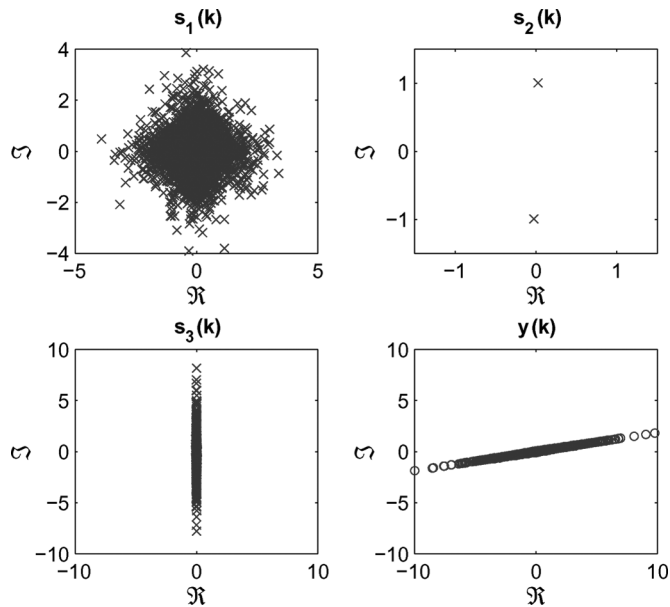
Fig. 3. Scatter plots of the complex sources $s_1(k)$, $s_2(k)$, and $s_3(k)$, whose properties are described in Table I. Scatter plot of the extracted signal $y(k)$, corresponding to the source $s_3(k)$, is given in the bottom right plot.
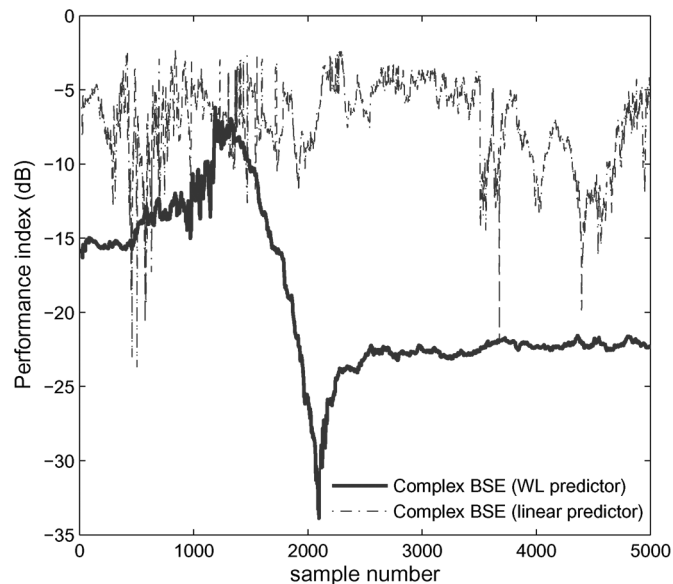

Fig. 4. Learning curves for the extraction of complex sources from noise-free mixtures using the algorithm in (26)–(28) based on (solid line) WL predictor and (broken line) linear predictor.

TABLE I
SOURCE PROPERTIES FOR NOISE-FREE EXTRACTION EXPERIMENTS

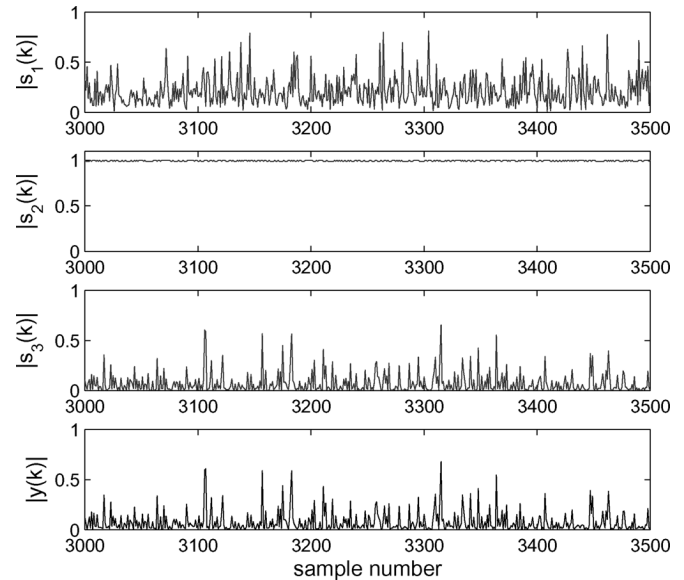| Source	distribution	circ. measure (ξ)	norm. MSPE
$s_1(k)$	Super-Gaussian	1.0000	1.4504
$s_2(k)$	Sub-Gaussian	39.3550	1.6869
$s_3(k)$	Super-Gaussian	106.7113	**1.3438**

Fig. 5. Normalized absolute values of the sources $s_1(k)$, $s_2(k)$, and $s_3(k)$, whose properties are described in Table I. The extracted source $y(k)$, shown in the bottom plot, is obtained from a noise-free mixture using the algorithm in (26)–(28).

The algorithm was able to extract the source with the smallest normalized MSPE, with the PI reaching a value of -22 dB at steady state after 2000 samples (Fig. 4). The normalized absolute values of the sources $s_i(k)$, $i = 1, 2, 3$, and $y(k)$ are shown in Fig. 5, illustrating that the desired source $s_3(k)$, with the smallest MSPE, was extracted successfully. Fig. 3 shows the scatter plots of the three sources and the extracted signal. The scatter plot of the extracted signal $y(k)$ is a scaled and rotated version of $s_3(k)$ due to the ambiguity problem of BSS. Next, for the same setting, the resulting mixture was prewhitened, and extraction was performed using the algorithm (30)–(32). The resulting learning curve shown in Fig. 6 exhibits slow convergence with an average steady-state value of -19 dB after 4000 samples. The step-size parameters were set to $\mu = 5 \times 10^{-3}$ and $\mu_h = \mu_g = 10^{-4}$.

For comparison, we next demonstrate the performance of the algorithm (26)–(28), which uses a standard linear predictor for the extraction of the complex sources. We performed the extraction of the noncircular sources (whose properties are given in Table I) using the same mixing matrix as in the previous experiments. This is straightforward by assuming the conjugate part of the coefficient vector of the WL predictor in (27) and (28) $\mathbf{g} = \mathbf{0}$ and updating only the coefficient vector $\mathbf{h}$, as shown in Section III. As shown in the analysis, the linear predictor is not suited for modeling the full second-order information and did not provide satisfactory extraction (as seen from Fig. 4), reaching an average PI of only -6.5 dB as opposed to -22 dB

In the first set of experiments, $N_s = 3$ sources with 5000 samples were generated (Fig. 3) and subsequently mixed to form a noise-free mixture. The sources were mixed using a 3×3 mixing matrix, and the resultant observation vector was input to the adaptive algorithm of (26) with a step size of $\mu = 5 \times 10^{-3}$. The coefficients of the WL predictor were updated using (27) and (28) with filter length $M = 20$ and $\mu_h = \mu_g = 10^{-5}$. The resultant learning curve shown in Fig. 4 was averaged over 100 independent trials. The source properties are shown in Table I, which also includes the circularity measure and the value of the normalized MSPE corresponding to the source [see (18)].
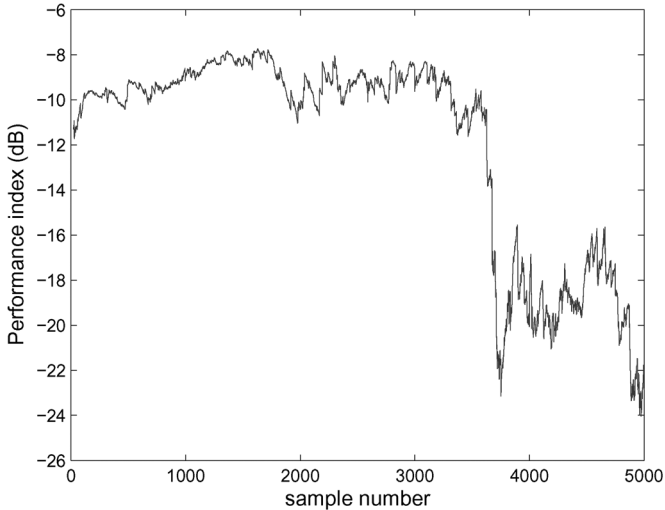
Fig. 6. Extraction of complex sources from a noise-free prewhitened mixture using the algorithm in (30)–(32) based on a WL predictor.

TABLE II
SOURCE PROPERTIES FOR NOISY EXTRACTION EXPERIMENTS

Source	distribution	circ. measure (ξ)	norm. MSPE
$s_1(k)$	Super-Gaussian	1.0140	2.8101
$s_2(k)$	Sub-Gaussian	190.9176	2.8268
$s_3(k)$	Super-Gaussian	20.9222	**2.7991**
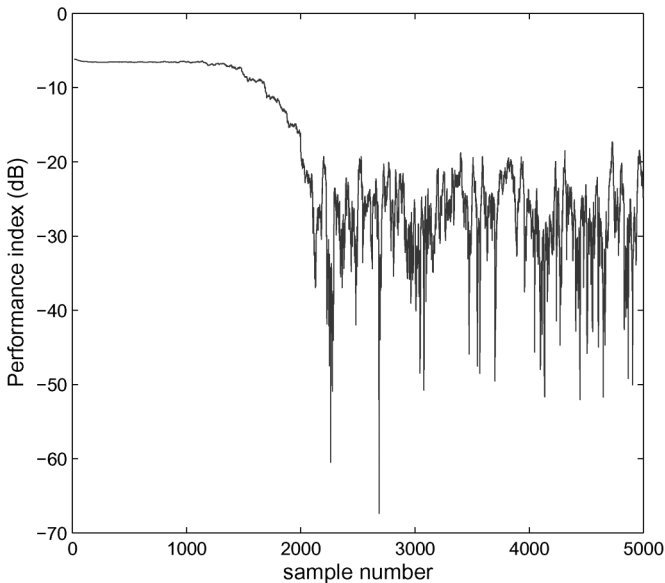
Fig. 7. Extraction of complex sources from a noisy mixture with additive circular white Gaussian noise using the algorithm in (36)–(38) with a WL predictor.

for the WL case using the ACLMS. In the next set of experiments, the performance of the proposed algorithms for the noisy case were investigated. A new set of three complex source signals was generated with 5000 samples, whose properties are described in Table II, and the 4×3 mixing matrix was generated randomly. Circular white Gaussian noise with variance $\sigma_v^2 = 0.1$ was added to the mixture to create the observed noisy mixture. The algorithm given in (36) was used to minimize the cost function and extract the source with the smallest normal-

ized MSPE. The values of the WL predictor coefficient vectors were updated via (37) and (38) with filter length $M = 20$ and step-size values $\mu = 5 \times 10^{-3}$ and $\mu_h = \mu_g = 10^{-3}$. The learning curve in Fig. 7 demonstrates the performance of the algorithm, reaching steady state after 2000 samples and with an average PI of -30 dB, indicating a successful extraction of the source $s_3(k)$.

We next investigated the effect of doubly white noncircular Gaussian noise with circularity measure $\xi = 5$ while keeping the source and mixing matrix values unchanged. The noise variance was $\sigma_v^2 = 0.1$, and the estimated pseudovariance of the noise was $\tau_v^2 = -0.0894 - j0.0002$ [using the subspace method in (48)]. The learning curve in Fig. 8 indicates the algorithm in (36)–(38) converging to a solution in around 1500 samples and with an average steady-state value of -21 dB for the step sizes $\mu = 5 \times 10^{-3}$ and $\mu_h = \mu_g = 10^{-5}$. For comparison, the learning curve using the algorithm in (26)–(28) is also included, illustrating the inability to extract the desired source from the noisy noncircular mixture. Finally, the input was prewhitened and sources extracted based on (45) for the update of the de-mixing vector, and using (46) and (47) for the update of the coefficient vectors, to produce the learning curve in Fig. 9. In this scenario, the step-size parameters were chosen as $\mu = 10^{-4}$ and $\mu_h = \mu_g = 10^{-6}$, leading to slow convergence.

B. EEG Artifact Extraction

We next demonstrate the usefulness of the proposed complex BSE scheme on the task of the extraction of eye muscle activity [electrooculogram (EOG)] from real-world EEG recordings. In real-time brain computer interfaces, it is desirable to identify and remove such artifacts from the contaminated EEG [41].

In our experiment, the EEG signals used were from electrodes Fp1, Fp2, C5, C6, O1, and O2 with the ground electrode placed at Cz, as shown in Fig. 10(a). In addition, EOG activity was also recorded from vEOG and hEOG channels to provide a reference for the performance assessment of the extraction.[5] Data were sampled at 512 Hz and recorded for 30 s. Notice that the effects of the artifacts diminish with the distance from the eyes, being most pronounced for the frontal electrodes Fp1 and Fp2 [Fig. 10(b)].

Pairing spatially symmetric electrodes to form complex signals facilitates the use of cross information and simultaneous modeling of the amplitude–phase relationships. Thus, pairs of symmetric electrodes were combined to form three temporal complex EEG signals given by

$$\mathbf{x}_1 = \mathbf{Fp1} + j\mathbf{Fp2}$$
$$\mathbf{x}_2 = \mathbf{C5} + j\mathbf{C6}$$
$$\mathbf{x}_3 = \mathbf{O1} + j\mathbf{O2}. \qquad (50)$$

First, the algorithm in (26)–(28) was used to remove EOG using the step size $\mu = 5 \times 10^{-3}$ with filter length $M = 70$ and step-sizes $\mu_h = \mu_g = 10^{-4}$ for the standard and conjugate coefficients of the ACLMS. The estimated EOG artifact was represented by the real component of the extracted signal $\Re\{y(k)\}$,

[5]As we do not have knowledge of the mixing matrix, we used the comparison of power spectra of the original and extracted EOG to validate the performance of the proposed complex BSE algorithms.
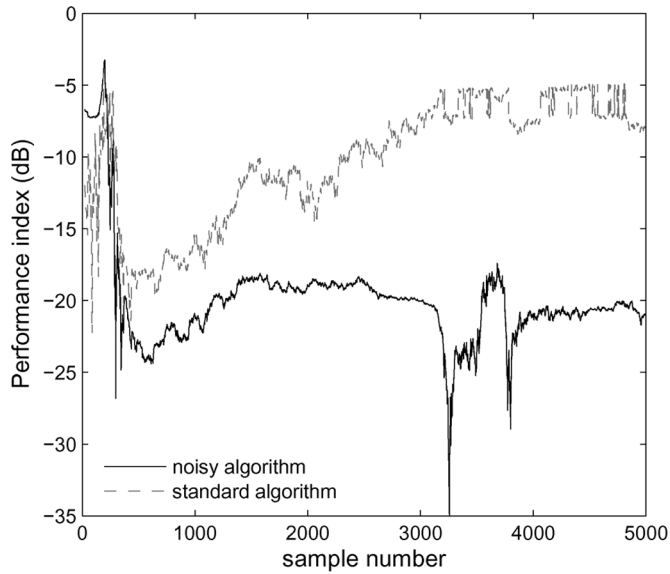
Fig. 8. Extraction of complex sources from a noisy mixture with additive doubly white noncircular Gaussian noise using (solid line) the algorithm in (36)–(38) and (broken line) the algorithm in (26)–(28) with a WL predictor.
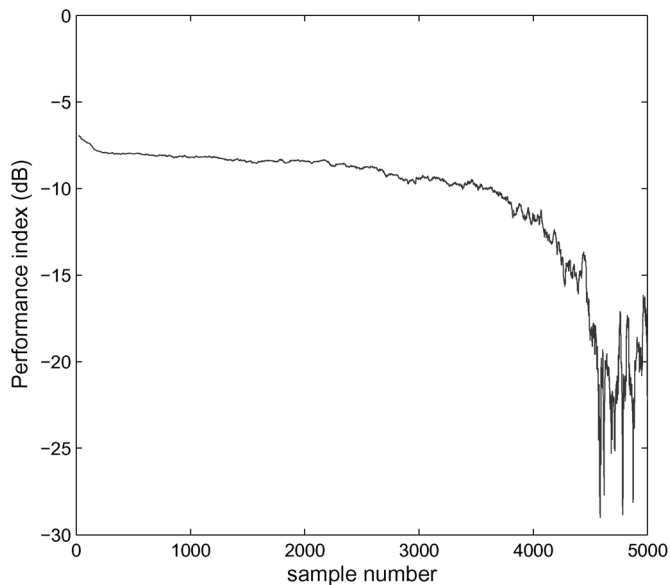

Fig. 9. Extraction of complex sources from a prewhitened noisy mixture with additive doubly white noncircular Gaussian noise using the algorithm in (45)–(47) with a WL predictor.

as illustrated in Fig. 10(c), in both time and frequency domains (the normalized power spectrum). The original vEOG signal is included for reference, confirming the successful extraction of the EOG artifact from EEG.

V. CONCLUSION

The BSE of complex signals from both noise-free and noisy mixtures has been addressed. The normalized MSPE, measured at the output of a WL predictor, is utilized as a criterion to extract sources based on their degree of predictability. The effectiveness of the WL model in this context has been demonstrated, verifying that the proposed approach is suitable for both second-order circular (proper) and noncircular (improper) signals and for general doubly white additive complex noises (improper).
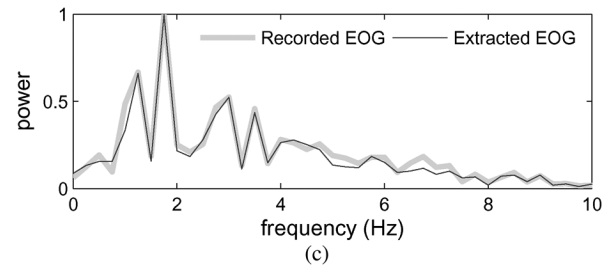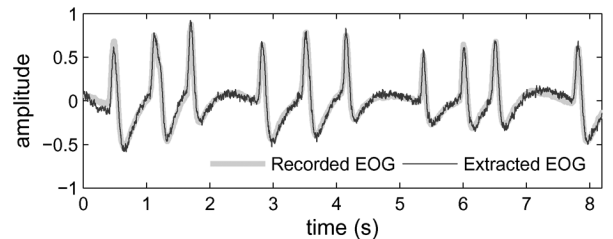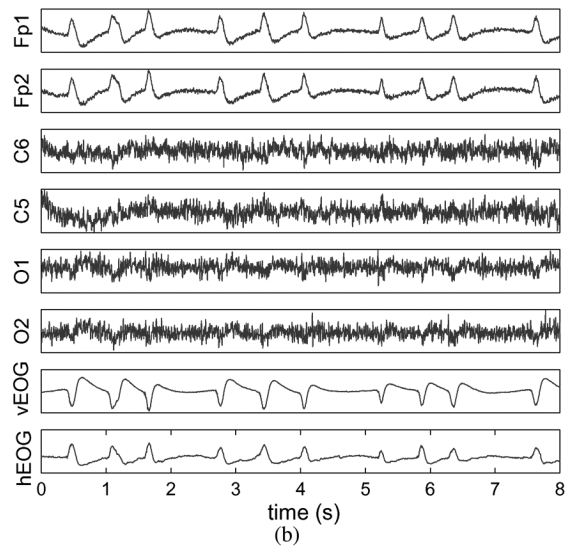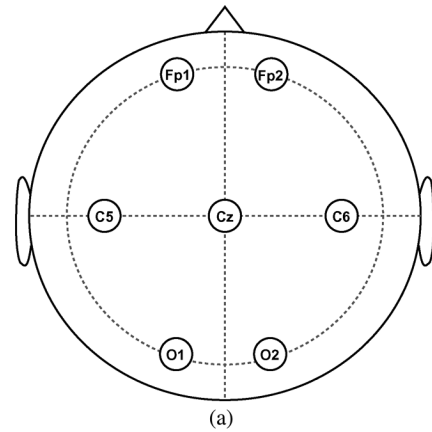

Fig. 10. Extraction of the EOG artifact due to eye movement from EEG data using the algorithm in (26)–(28). (a) EEG channels used in the experiment (according to the 10–20 system). (b) First 8 s of the EEG and EOG recordings. (c) (Top) First 8 s of the extracted EOG signal (thick grey line) and recorded vEOG signal (thin line), after normalizing amplitudes. (Bottom) Normalized power spectra of the extracted EOG signal (thin line) and the original vEOG signal (thick grey line).

For circular sources, the proposed BSE approach (P-cBSE) has been shown to perform as good as standard approaches, whereas for noncircular sources, it exhibits theoretical and practical advantages over the existing methods. The performance of the pro-

posed algorithm has been illustrated by simulations in noise-free and noisy conditions. In addition, the application of the proposed method has been demonstrated in the extraction of artifacts from corrupted EEG signals directly in the time domain.

APPENDIX
DERIVATION OF THE MSPE

The error at the output of the WL predictor $e(k)$ can be written as

$$
\begin{aligned}
e(k) &= y(k) - y_{WL}(k) \\
&= y(k) - \mathbf{h}^T(k)\mathbf{y}(k) - \mathbf{g}^T(k)\mathbf{y}^*(k) \\
&= \mathbf{w}^H \underbrace{\left(\mathbf{x}(k) - \sum_{m=1}^{M} h_m(k)\mathbf{x}(k-m) \right)}_{\triangleq \bar{\mathbf{x}}_h(k)} \\
&\quad - \mathbf{w}^T \underbrace{\sum_{m=1}^{M} g_m(k)\mathbf{x}^*(k-m)}_{\triangleq \bar{\mathbf{x}}_g(k)} \\
&= \mathbf{w}^H \bar{\mathbf{x}}_h(k) - \mathbf{w}^T \bar{\mathbf{x}}_g(k)
\end{aligned}
\tag{51}
$$

and the MSPE can be expanded as

$$
\begin{aligned}
E\left\{ |e(k)|^2 \right\} &= E\left\{ \left(\mathbf{w}^H \bar{\mathbf{x}}_h(k) - \mathbf{w}^T \bar{\mathbf{x}}_g(k) \right) \right. \\
&\quad \left. \times \left(\bar{\mathbf{x}}_h^H(k)\mathbf{w} - \bar{\mathbf{x}}_g^H(k)\mathbf{w}^* \right) \right\} \\
&= \mathbf{w}^H E_1 \mathbf{w} - \mathbf{w}^H E_2 \mathbf{w}^* - \mathbf{w}^T E_3 \mathbf{w} + \mathbf{w}^T E_4 \mathbf{w}^*
\end{aligned}
\tag{52}
$$

where

$$
E_1 = E\left\{ \bar{\mathbf{x}}_h(k)\bar{\mathbf{x}}_h^H(k) \right\}, \ E_2 = E\left\{ \bar{\mathbf{x}}_h(k)\bar{\mathbf{x}}_g^H(k) \right\}
$$
$$
E_3 = E\left\{ \bar{\mathbf{x}}_g(k)\bar{\mathbf{x}}_h^H(k) \right\}, \ E_4 = E\left\{ \bar{\mathbf{x}}_g(k)\bar{\mathbf{x}}_g^H(k) \right\}.
$$

Recall that the observation $\mathbf{x}(k) = \mathbf{A}\mathbf{s}(k) + \mathbf{v}(k)$, so the MSPE can be divided into terms relating to the source (denoted by e_s^2) and those relating to the noise (denoted by e_v^2), giving $E\{|e(k)|^2\} = e_s^2 + e_v^2$. Assuming a noise-free case, that is, $e_v^2 = 0$, the values of E_i, $i = \{1,2,3,4\}$, can be expressed as

$$
\begin{aligned}
E_1 &= \mathcal{C}_{\mathbf{ss}}(0) - \sum_{m=1}^{M} h_m(k)\mathcal{C}_{\mathbf{ss}}(-m) - \sum_{m=1}^{M} h_m^*(k)\mathcal{C}_{\mathbf{ss}}(m) \\
&\quad + \sum_{m,\ell=1}^{M} h_m(k)h_\ell^*(k)\mathcal{C}_{\mathbf{ss}}(\ell - m)
\end{aligned}
\tag{53}
$$

$$
\begin{aligned}
E_2 &= \sum_{m=1}^{M} g_m^*(k)\mathcal{P}_{\mathbf{ss}}(m) \\
&\quad - \sum_{m,\ell=1}^{M} h_m(k)g_\ell^*(k)\mathcal{P}_{\mathbf{ss}}(\ell - m)
\end{aligned}
\tag{54}
$$

$$
\begin{aligned}
E_3 &= \sum_{m=1}^{M} g_m(k)\mathcal{P}_{\mathbf{ss}}^*(m) \\
&\quad - \sum_{m,\ell=1}^{M} h_m^*(k)g_\ell(k)\mathcal{P}_{\mathbf{ss}}^*(\ell - m)
\end{aligned}
\tag{55}
$$

$$
E_4 = \sum_{m,\ell=1}^{M} g_m^*(k)g_\ell(k)\mathcal{C}_{\mathbf{ss}}(\ell - m).
\tag{56}
$$

Since $E_3 = E_2^H$ and $z + z^* = 2\Re\{z\}$, (53)–(56) can be simplified and substituted in (52) to produce the final result $E\{|e(k)|^2\} = e_s^2$, as given in (17).

To derive the MSPE relating to the nth source, notice that the sources are assumed uncorrelated, and so the covariance and pseudocovariance matrices are diagonal. It is then straightforward to express the nth diagonal element of (53)–(56) to produce (18).

In the noisy case, the values of E_i pertaining to e_v^2 (denoted by $E_{i,v}$) can be evaluated in a similar fashion to that in (53)–(56), noticing that $\mathcal{C}_{\mathbf{vv}}(\delta) = \mathcal{P}_{\mathbf{vv}}(\delta) = \mathbf{0}$ for $\delta \neq 0$. Thus

$$
E_{1,v} = \mathcal{C}_{\mathbf{vv}}(0) + \sum_{m=1}^{M} h_m(k)h_m^*(k)\mathcal{C}_{\mathbf{vv}}(0)
\tag{57}
$$

$$
E_{2,v} = -\sum_{m=1}^{M} h_m(k)g_m^*(k)\mathcal{P}_{\mathbf{vv}}(0)
\tag{58}
$$

$$
E_{3,v} = -\sum_{m=1}^{M} h_m^*(k)g_m(k)\mathcal{P}_{\mathbf{vv}}(0)
\tag{59}
$$

$$
E_{4,v} = \sum_{m=1}^{M} g_m(k)g_m^*(k)\mathcal{C}_{\mathbf{vv}}(0)
\tag{60}
$$

which when substituted in (52) and simplified results in

$$
e_v^2 = \mathbf{w}^H \hat{\mathcal{C}}_{\mathbf{vv}} \mathbf{w} + \Re\{\mathbf{w}^H \hat{\mathcal{P}}_{\mathbf{vv}} \mathbf{w}^*\}
\tag{61}
$$

$$
\hat{\mathcal{C}}_{\mathbf{vv}} = \left[1 + \mathbf{h}^H(k)\mathbf{h}(k) + \mathbf{g}^H(k)\mathbf{g}(k) \right] \sigma_v^2 \mathbf{I}
\tag{62}
$$

$$
\hat{\mathcal{P}}_{\mathbf{vv}} = \begin{cases} 2\mathbf{g}^H(k)\mathbf{h}(k)\tau_v^2 \mathbf{I}, & \mathbf{v}(k) \text{ for doubly white} \\ \mathbf{0}, & \mathbf{v}(k) \text{ for circular white} \end{cases}
\tag{63}
$$

where $\hat{\mathcal{C}}_{\mathbf{vv}}$ and $\hat{\mathcal{P}}_{\mathbf{vv}}$ are written in their vector form.

ACKNOWLEDGMENT

REFERENCES

[1] J. Anemüller, T. J. Sejnowski, and S. Makeig, "Complex independent component analysis of frequency-domain electroencephalographic data," *Neural Netw.*, vol. 16, no. 9, pp. 1311–1323, Nov. 2003.

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

[3] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. Hoboken, NJ: Wiley, 2001.

[4] J. Särelä and H. Valpola, "Denoising source separation," *J. Mach. Learn. Res.*, vol. 6, pp. 233–272, 2005.

[5] J. Cardoso, "Blind signal separation: Statistical principles," *Proc. IEEE*, vol. 86, no. 10, pp. 2009–2025, Oct. 1998.

[6] S. Douglas, "Blind signal separation and blind deconvolution," in *Handbook of Neural Network Signal Processing*. Boca Raton, FL: CRC Press, 2002, ch. 7.

[7] W. Leong and D. Mandic, "Noisy component extraction (NoiCE)," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 57, no. 3, pp. 664–671, Mar. 2010.

[8] A. J. Bell and T. J. Sejnowski, "An information-maximisation approach to blind separation and blind deconvolution," *Neural Comput.*, vol. 7, no. 6, pp. 1129–1159, Nov. 1995.

[9] S. Amari, A. Cichocki, and H. Yang, "A new learning algorithm for blind signal separation," in *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press, 1996, pp. 757–763.

[10] Q. Shi, R. Wu, and S. Wang, "A novel approach to blind source extraction based on skewness," in *Proc. ICSP*, 2006 , vol. 4, pp. 3187–3190.

[11] P. Georgiev and A. Cichocki, "Robust blind source separation utilizing second and fourth order statistics," in *Proc. ICANN*, New York, 2002, vol. 2415, pp. 1162–1167.

[12] A. Cichocki, R. Thawonmas, and S. Amari, "Sequential blind signal extraction in order specified by stochastic properties," *Electron. Lett.*, vol. 33, no. 1, pp. 64–65, Jan. 1997.

[13] W. Liu and D. P. Mandic, "A normalised kurtosis-based algorithm for blind source extraction from noisy measurements," *Signal Process.*, vol. 86, no. 7, pp. 1580–1585, Jul. 2006.

[14] D. P. Mandic and A. Cichocki, "An online algorithm for blind extraction of sources with different dynamical structures," in *Proc. 4th Int. Symp. Ind. Compon. Anal. Blind Signal Separation (ICA)*, 2003, pp. 645–650.

[15] B.-Y. Wang and W. Zheng, "Blind extraction of chaotic signal from an instantaneous linear mixture," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 53, no. 2, pp. 143–147, Feb. 2006.

[16] W. Liu, D. P. Mandic, and A. Cichocki, "Blind source extraction of instantaneous noisy mixtures using a linear predictor," in *Proc. IEEE Int. Symp. Circuits Syst.*, 2006, pp. 4199–4202.

[17] W. Liu, D. P. Mandic, and A. Cichocki, "Blind second-order source extraction of instantaneous noisy mixtures," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 53, no. 9, pp. 931–935, Sep. 2006.

[18] 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, no. 10–12, pp. 2344–2355, Jun. 2008.

[19] D. P. Mandic, S. Javidi, G. Souretis, and S. L. Goh, "Why a complex valued solution for a real domain problem," in *Proc. IEEE Workshop Mach. Learn. Signal Process.*, Aug. 2007, pp. 384–389.

[20] D. P. Mandic and S. L. Goh, *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models*. Hoboken, NJ: Wiley, 2009.

[21] B. Picinbono, "On circularity," *IEEE Trans. Signal Process.*, vol. 42, no. 12, pp. 3473–3482, Dec. 1994.

[22] B. Picinbono and P. Bondon, "Second-order statistics of complex signals," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 411–420, Feb. 1997.

[23] F. Neeser and J. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1293–1302, Jul. 1993.

[24] W. Wirtinger, "Zur formalen theorie der funktionen von mehr komplexen veränderlichen," *Mathematische Annalen*, vol. 97, no. 1, pp. 357–375, Dec. 1927.

[25] K. Kreutz-Delgado, The complex gradient operator and the $\mathbb{CR}$-calculus Dept. Elect. Comput. Eng., Univ. California, San Diego, CA, Course Lecture Supplement No. ECE275A, 2006.

[26] S. Javidi, M. Pedzisz, S. Goh, and D. P. Mandic, "The augmented complex least mean square algorithm with application to adaptive prediction problems," in *Proc. 1st IARP Workshop Cogn. Inf. Process.*, 2008, pp. 54–57.

[27] Y. Xia, C. Cheong-Took, S. Javidi, and D. Mandic, "A widely linear affine projection algorithm," in *Proc. IEEE Workshop Stat. Signal Process.*, 2009, pp. 373–376.

[28] C. Cheong-Took and D. Mandic, "Adaptive IIR filtering of noncircular complex signals," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 4111–4118, Oct. 2009.

[29] M. Novey and T. Adalı, "On extending the complex FastICA algorithm to noncircular sources," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 2148–2154, May 2008.

[30] S. Douglas, J. Eriksson, and V. Koivunen, "Fixed-point complex ICA algorithms for the blind separation of sources using their real or imaginary components," in *Independent Component Analysis and Blind Signal Separation*. New York: Springer-Verlag, 2006, pp. 343–351.

[31] A. Erdogan, "On the convergence of ICA algorithms with symmetric orthogonalization," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2209–2221, Jun. 2009.

[32] E. Bingham and A. Hyvärinen, "A fast fixed point algorithm for independent component analysis of complex valued signals," *Int. J. Neural Syst.*, vol. 10, no. 1, pp. 1–8, Feb. 2000.

[33] S. Javidi, B. Jelfs, and D. P. Mandic, "Blind extraction of noncircular signals using a widely linear predictor," in *Proc. IEEE Workshop Stat. Signal Process.*, 2009, pp. 501–504.

[34] J. Eriksson and V. Koivunen, "Complex random vectors and ICA models: Identifiability, uniqueness, and separability," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1017–1029, Mar. 2006.

[35] E. Ollila and V. Koivunen, "Complex ICA using generalized uncorrelating transform," *Signal Process.*, vol. 89, no. 4, pp. 365–377, Apr. 2009.

[36] P. Schreier, "Bounds on the degree of impropriety of complex random vectors," *IEEE Signal Process. Lett.*, vol. 15, pp. 190–193, 2008 .

[37] A. Walden and P. Rubin-Delanchy, "On testing for impropriety of complex-valued Gaussian vectors," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 825–834, Mar. 2009.

[38] B. Picinbono and P. Chevalier, "Widely linear estimation with complex data," *IEEE Trans. Signal Process.*, vol. 43, no. 8, pp. 2030–2033, Aug. 1995.

[39] D. P. Mandic, S. Javidi, S. L. Goh, A. Kuh, and K. Aihara, "Complex-valued prediction of wind profile using augmented complex-valued statistics," *Renew. Energy*, vol. 34, pp. 196–201, Nov. 2007.

[40] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1996.

[41] P. Georgiev, A. Cichocki, and H. Bakardjian, "Optimization techniques for independent component analysis with applications to EEG data," in *Quantitative Neuroscience: Models, Algorithms, Diagnostics, and Therapeutic Applications*. Norwell, MA: Kluwer, 2004, ch. 3, pp. 53–68.

Soroush Javidi (S'09) received the M.Eng. degree in electrical and electronic engineering in 2006 from Imperial College London, London, U.K., where he is currently working toward the Ph.D. degree with the Department of Electrical and Electronic Engineering, under the supervision of Dr. D. Mandic.

His research interests include adaptive and statistical signal processing and complex-valued blind source separation. His research has focused on the use of augmented complex statistics and widely linear models in practical applications, such as EEG, wind modeling, and energy-efficient adaptive signal processing algorithms (green circuits).

Danilo Mandic (M'99–SM'03) is a Reader in signal processing with Imperial College London. London, U.K. He has been working in the area of nonlinear adaptive signal processing and nonlinear dynamics. His publication record includes two research monographs *Recurrent Neural Networks for Prediction* (1st ed., Aug. 2001) and *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models* (1st ed., Wiley, Apr. 2009), an edited book *Signal Processing for Information Fusion* (Springer, 2008), and more than 200 publications on signal and image processing. He has been a Guest Professor with K.U. Leuven Belgium, TUAT Tokyo, Japan, and Westminster University, U.K., and a Frontier Researcher with RIKEN Japan.

Dr. Mandic is a member of the London Mathematical Society. He is a member of the IEEE Technical Committee on Machine Learning for Signal Processing and an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—II: EXPRESS BRIEFS, IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE TRANSACTIONS ON NEURAL NETWORKS, and the *International Journal of Mathematical Modelling and Algorithms*. He has produced award-winning papers and products resulting from his collaboration with industry.

Andrzej Cichocki (M'96–SM'07) received the M.Sc. (with honors), Ph.D., and Habilitate Doctorate (Dr.Sc.) degrees from Warsaw University of Technology, Warsaw, Poland, in 1972, 1975, and 1982, respectively, all in electrical engineering.

Since 1972, he has been with the Institute of Theory of Electrical Engineering and Electrical Measurements, Warsaw University of Technology, where he became a full Professor in 1991. Since 1999, he has been Head of the Laboratory for Advanced Brain Signal Processing, Brain Science Institute RIKEN, Saitama, Japan. His current research interests include biomedical signal and image processing (particularly blind signal/image processing), neural networks and their applications, learning theory and robust algorithms, generalization and extensions of independent and principal component analysis, optimization problems, and nonlinear circuits and systems theory and their applications. He is the coauthor of four books *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis* (Wiley, 2009, with R. Zdunek, A.-H Phan and S.-. Amari), *Adaptive Blind Signal and Image Processing* (New York: Wiley, 2003 with Prof. S.-. Amari), *MOS Switched-Capacitor and Continuous-Time Integrated Circuits and Systems* (New York: Springer-Verlag, 1989, with Prof. R. Unbehauen), and *Neural Networks for Optimization and Signal Processing* (New York: Wiley/Teubner-Verlag, 1993/1994, with Prof. R. Unbehauen) and more than 150 papers.

Dr. Cichocki is a member of several international scientific committees and the Associated Editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS (since January 1998).