Augmented Complex Common Spatial Patterns for Classification of Noncircular EEG From Motor Imagery Tasks

Cheolsoo Park, Clive Cheong Took, *Senior Member, IEEE*, and Danilo P. Mandic, *Fellow, IEEE*

***Abstract*—A novel augmented complex-valued common spatial pattern (CSP) algorithm is introduced in order to cater for general complex signals with *noncircular* probability distributions. This is a typical case in multichannel electroencephalogram (EEG), due to the power difference or correlation between the data channels, yet current methods only cater for a very restrictive class of circular data. The proposed complex-valued CSP algorithms account for the generality of complex *noncircular* data, by virtue of the use of augmented complex statistics and the strong-uncorrelating transform (SUT). Depending on the degree of power difference of complex signals, the analysis and simulations show that the SUT based algorithm maximizes the inter-class difference between two motor imagery tasks. Simulations on both synthetic *noncircular* sources and motor imagery experiments using real-world EEG support the approach.**

***Index Terms*—Augmented complex common spatial pattern (ACCSP), brain–computer interface (BCI), common spatial pattern (CSP), complex noncircularity, complex pseudocovariance, motor imagery paradigm, strong-uncorrelating transform (SUT).**

I. Introduction

BRAIN–COMPUTER interface (BCI) research aims to provide computer-aided control using exclusively brain activity, and has been instrumental to advances across bioengineering fields, such as in neuroprosthetics. For real-time BCI systems, the electroencephalogram (EEG) is the most convenient means to measure neurophysiological activity due to its noninvasive nature and affordable recording equipment [1], [2]. In particular, the EEG in response to the motor imagery task, defined as the "mental rehearsal of simple or complex motor acts that is not accompanied by overt body movements" [3], [4] has been a popular BCI paradigm [5].

The basis for such "motor imagery" BCI tasks are the mu (8–12 Hz) and $beta$ rhythms (14–20 Hz), together with higher beta/gamma (above 20 Hz) in EEG, that have been observed when subjects plan and execute their hand or finger movement [6]–[9]. Nikouline *et al.* [10] demonstrated that somatosensory stimuli suppressed mu rhythms at somatosensory cortex (SI), and Yuan *et al.* [9] suggested that the mu and $beta$ rhythms are changed by the reflection of phase coherence in thalamocortical circuits. In addition, positron emission tomography (PET) or functional magnetic resonance imaging (fMRI) imaging studies have established that several active brain regions are observed during motor imagery tasks, including supplementary motor area, superior and inferior parietal lobule, dorsal and ventral lateral pre-motor cortices, pre-frontal areas, inferior frontal gyrus, superior temporal gyrus, primary motor cortex (M1), primary sensory cortex, secondary sensory area, insular cortex, anterior cingulate cortex, superior temporal gyrus, basal ganglia, and cerebellum [11]. In the analysis of such brain activities coming from spatially distributed regions, the common spatial patterns (CSP) is a common algorithm that extracts features relevant to motor imagery tasks [12]–[14].

The CSP decomposes multichannel EEG coming from two classes into spatial patterns, defined through a simultaneous diagonalization of the corresponding data covariance matrices. In doing so, it maximizes the variance of signals belonging to one class while simultaneously minimizing the variance of signals of the other class, thus enhancing separability between the two classes. The CSP algorithm was first applied to EEG for abnormality detection [15] and has been recently employed to discriminate between groups of movement-related tasks [12].

When modeling EEG channel interaction pair-wise, it is convenient and physically meaningful to pair spatially adjacent real-valued EEG signals x and y to form complex-valued data, $z = x + jy$. This allows for the coupling (e.g., phase information) between the two channels to be naturally exploited; for instance, power difference and correlation between two channels affect the level of *noncircularity* (rotation invariant) of the distribution of the complex signal. To account for *noncircularity* that always exists in real-world complex-valued data, recent advances in complex statistics [16]–[19] consider both the pseudocovariance $E[\mathbf{x}\mathbf{x}^T]$ and the traditional covariance $E[\mathbf{x}\mathbf{x}^H]$ matrix, in order to utilize full statistical information available [20], [21]. On the other hand, traditional studies on modeling pair-wise EEG channel interaction such as connectivity and synchrony mostly rely on correlation analysis [11], [22]. The complex extension of CSP was introduced by Falzon *et al.* using the analytic signal in order to enhance the modeling

Manuscript received August 08, 2013; revised October 04, 2013 and November 18, 2013; accepted November 24, 2013. Date of publication December 12, 2013; date of current version January 06, 2014.

C. Park is with the Department of Bioengineering, University of California-San Diego, La Jolla, CA 92093 USA (e-mail: charles586@gmail.com).

C. C. Took is with the Department of Computing, University of Surrey, GU2 7XH Surrey, U.K. (e-mail: c.cheongtook@surrey.ac.uk).

D. P. Mandic is with the Department of Electrical and Electronic Engineering, Imperial College London, SW7 2AZ London, U.K. (e-mail: d.mandic@imperial.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNSRE.2013.2294903

by considering the phase information from a single channel EEG signal [23], [24]. However, as shown later the analytic signal based CSP method is not optimal for multichannel data which exhibit different powers or degrees of correlation, as in this way the pseudocovariance information is omitted in the modeling.

To this end, we propose to use augmented complex statistics in order to also access the information contained in the pseudocovariance. Modelling based on augmented statistics has been successfully exploited in supervised learning, including the augmented complex least mean square (ACLMS) [25] and augmented complex matrix factorization [26]—demonstrating enhanced and rigorous modeling of real-world data.

In this paper, augmented complex statistics are applied to design optimal complex extensions of the common spatial pattern method. To extract signatures of motor imagery tasks, we employ the strong-uncorrelating transform (SUT) to diagonalise the covariance and pseudocovariance matrix simultaneously [27]–[29]. In particular, we show that SUT produces the information of power difference between the real and imaginary parts of complex data as well as the power information for each data channel. In this way, SUT naturally helps to maximize the difference in the variance between two groups of data, provided a significant power difference between two parts exists. Since the network change of the brain activity corresponding to different motor imagery tasks is considered to cause a variation in the power difference between the EEG channels [11], [22], considered our aim is to show that it is convenient and physically meaningful to pair, augmented complex statistics using SUT provides more information and hence better features to classify the brain responses to different tasks, compared to the real-valued and the existing complex valued CSP algorithm.

We here introduce the augmented complex CSP using SUT (SUTCCSP) and show that SUTCCSP is the optimal CSP method for the analysis of pair-wise motor imagery data, which is almost invariably *noncircular*. Unlike the real-valued CSP and the existing analytic signal-based complex CSP, the analysis and simulations show that SUTCCSP exploits the full available statistics between two-channel EEG data yielding improved classification performance. The proposed method is verified on a bench mark motor imagery dataset of 109 subjects, whereby subjects were imaging their left and right hand movement. The usefulness of SUTCCSP is supported by simulations.

II. Complex Common Spatial Pattern Methods

A complex extension of common spatial filters was recently introduced by O. Falzon *et al.* for the discrimination of mental tasks [23], [24], [30]. By combining the Hilbert transform with complex CSP, they exploited phase information from EEG data, leading to the algorithm termed the analytic signal-based CSP (ACSP). However, statistics in $\mathbb{C}$ are not an analytical continuation of the corresponding statistics in $\mathbb{R}$ [17] since general complex random variables are *noncircular*.[1] In the domain of second order statistics, a *circular* signal is called proper and is manifested by equal powers in the real and imaginary part, while second order *noncircular* signal is called improper [31]. *Circular* signals have rotation invariant distributions, whereas improper signals exhibit different power levels in their real and imaginary parts. We aim to introduce rigorous complex extensions of the CSP algorithm catering for complex *noncircularity*, including augmented complex CSP (ACCSP) and complex CSP equipped with the strong-uncorrelating transform, termed SUTCCSP.

A. Covariance Matrix of Complex Common Spatial Pattern

Consider complex-valued zero-mean data matrices, $\mathbf{Z}_a$ and $\mathbf{Z}_b \in \mathbb{C}^{N\times T}$, corresponding to the classes a and b, where N is the number of data channels and T the number of samples per channel. The covariance matrix of $\mathbf{Z}_a$ can be calculated as

$$\mathbf{C}_a = cov(\mathbf{Z}_a) = E\left[\mathbf{Z}_a\mathbf{Z}_a^H\right] \tag{1}$$

where $E[\cdot]$ denotes the statistical expectation operator, and $(\cdot)^H$ the conjugate transpose. The spatial covariance $\bar{\mathbf{C}}_{d\in[a,b]}$ for the class a or b is obtained by averaging the covariance matrices of all the complex-valued data channels, thus, $\bar{\mathbf{C}}_a$ and $\bar{\mathbf{C}}_b$ represent a pair of $N \times N$ Hermitian positive-semidefinite matrices,[2] whose elements are real-valued along the main diagonal and the off-diagonal elements are complex-valued. The composite spatial covariance matrix is given as

$$\mathbf{C}_c = \bar{\mathbf{C}}_a + \bar{\mathbf{C}}_b \tag{2}$$

and can be factored as

$$\mathbf{C}_c = \mathbf{U}_c\mathbf{\Lambda}_c\mathbf{U}_c^H. \tag{3}$$

The columns of $\mathbf{U}_c$ are the eigenvectors corresponding to the real-valued eigenvalues in the diagonal eigenvalue matrix $\mathbf{\Lambda}_c$. This allows for $\mathbf{C}_c$ to be whitened by multiplying with $\mathbf{G} = \mathbf{\Lambda}_c^{-(1/2)}\mathbf{U}_c^H$ such that

$$\mathbf{I} = \mathbf{G}\mathbf{C}_c\mathbf{G}^H = \mathbf{G}\bar{\mathbf{C}}_a\mathbf{G}^H + \mathbf{G}\bar{\mathbf{C}}_b\mathbf{G}^H \tag{4}$$

where the symbol $\mathbf{I}$ denotes the identity matrix. Next, let $\mathbf{S}_a = \mathbf{G}\bar{\mathbf{C}}_a\mathbf{G}^H$ and $\mathbf{S}_b = \mathbf{G}\bar{\mathbf{C}}_b\mathbf{G}^H$, so that $\mathbf{S}_a$ and $\mathbf{S}_b$ share a common eigenvector matrix, that is

$$\mathbf{B}^{-1}\mathbf{S}_a\mathbf{B} = \mathbf{\Lambda}_a \text{ and } \mathbf{B}^{-1}\mathbf{S}_b\mathbf{B} = \mathbf{\Lambda}_b \quad (\mathbf{\Lambda}_a + \mathbf{\Lambda}_b = \mathbf{I}). \tag{5}$$

By design, the eigenvalues of $\mathbf{\Lambda}_a$ are sorted in a descending order, this at the same time implies that the eigenvalues of $\mathbf{\Lambda}_b$ are sorted in an ascending order. Then, the spatial filter that maximizes the variance for one class of data and minimises the variance for the other is obtained as

$$\mathbf{W} = \mathbf{B}^{-1}\mathbf{G}. \tag{6}$$

[1] A Hilbert transformed analytic signal is always *circular*.

[2] For any non-zero vector $\mathbf{x}$ ($\in \mathbb{C}^n$), a $n \times n$ Hermitian matrix $\mathbf{M}$ is called positive-semidefinite if $\mathbf{x}^H\mathbf{M}\mathbf{x} \geq 0$ [32].

In this way, for a given data matrix $\mathbf{Z}$, a new set of data $\mathbf{V}$ is calculated as

$$\mathbf{V} = \mathbf{W}\mathbf{Z} \tag{7}$$

where each row vector $\mathbf{w}_j$ $(j = 1, \ldots, N)$ of $\mathbf{W}$ is called a spatial filter or simply a filter.

In order to discriminate between two classes, the variances of the real and imaginary parts of the spatial filtered complex data $\mathbf{V}$ in (7) are used as features. The row vectors corresponding to the real and imaginary parts of $\mathbf{V}$, that is, $\Re[\mathbf{v}_p]$ and $\Im[\mathbf{v}_p]$ ($p = 1, \ldots, m$ and $N - m + 1, \ldots, N$) are associated with the largest eigenvalues in $\mathbf{\Lambda}_a$ and $\mathbf{\Lambda}_b$. These signals are the m first and last rows of $\mathbf{V}$ due to the filtering process of $\mathbf{W}$, so that the features from real and imaginary parts are obtained as

$$f_p^R = \log\left(\frac{\mathtt{var}\left(\Re[\mathbf{v}_p]\right)}{\sum_{i=1,\ldots,m \text{ and } N-m+1,\ldots,N} \mathtt{var}\left(\Re[\mathbf{v}_i]\right)}\right) \tag{8}$$

$$f_p^I = \log\left(\frac{\mathtt{var}\left(\Im[\mathbf{v}_p]\right)}{\sum_{i=1,\ldots,m \text{ and } N-m+1,\ldots,N} \mathtt{var}\left(\Im[\mathbf{v}_i]\right)}\right) \tag{9}$$

where the symbol $\mathtt{var}(\cdot)$ denotes the variance of $(\cdot)$.

B. Analytic Signal-Based Common Spatial Pattern

In 2010, Falzon *et al.* proposed the ACSP by transforming input signals into their analytic representations and performing joint diagonalization of the obtained complex-valued covariance matrices [23]. After the signals are represented in their analytic forms, the phase information in the data is used to distinguish between two classes of mental processes.

For a signal $x(t)$, its analytic representation $z(t)$ is obtained through its Hilbert transform as

$$\hat{x}(t) = H\left(x(t)\right) = \frac{1}{\pi} K \int_{-\infty}^{\infty} \frac{x(t')}{t - t'} dt' \tag{10}$$

where K is the Cauchy principal value, to give the complex signal

$$z(t) = x(t) + j\hat{x}(t) = a(t)e^{j\phi(t)}. \tag{11}$$

Such a complex signal $z(t)$ is described by its amplitude and phase functions, $a(t)$ and $\phi(t)$, which are calculated as

$$a(t) = \sqrt{x^2(t) + \hat{x}^2(t)} \text{ and } \phi(t) = arctan\left(\frac{\hat{x}(t)}{x(t)}\right). \tag{12}$$

Based upon the so derived complex-valued analytic signal $z(t)$, the same procedure as in Section II-A is used to obtain the ACSP features.

However, it is important to notice that the Hilbert transform is only used for a narrowband signal and it cannot provide a full description of the frequency content for general data [33], [34], also leading to negative instantaneous frequencies in the time-frequency domain, when the amplitude variations are faster than the phase variations. To this end, to bypass the bandwidth limitation of the Hilbert transform, the Fourier transform or empirical mode decomposition [34], [35] should be considered in conjunction with ACSP [36].

C. Augmented Complex Common Spatial Pattern (ACCSP)

For a zero-mean complex random vector variable $\mathbf{z}$ the covariance matrix is given by $\underline{\mathbf{C}} = E[\mathbf{z}\mathbf{z}^H]$ and is used in standard second-order statistical signal processing. However, complex statistics are not a straightforward extension of real-valued statistics [17], [37], [38], since $\underline{\mathbf{C}}$ does not completely describe the second order statistics of $\mathbf{z}$, and a statistical descriptor called the pseudocovariance, $\underline{\mathbf{P}} = E[\mathbf{z}\mathbf{z}^T]$, also needs to be considered.

The covariance of a random variable $z = z_r + jz_i$ is given by $E[zz^*] = E[z_r^2 + z_i^2]$ (>0, unless $z = 0 + j0$), where $(\cdot)^*$ denotes the complex conjugate operator, while the pseudocovariance $E[zz] = E[z_r^2] - E[z_i^2] + 2jE[z_r z_i]$ vanishes only if z_r and z_i are uncorrelated and with the same variance. Signals are called second-order *circular* or proper if their pseudo-covariance is zero. However, due to short window observations, anisotropic noises, unequal powers of data channels and reflections, the pseudocovariance of real world data is often non-zero even if the data are *circular* [26]. Therefore, second order statistical modeling in $\mathbb{C}$ should examine joint statistical properties of $\mathbf{z}$ and $\mathbf{z}^*$, that is, it should be based on the augmented form of the complex variable $\hat{\mathbf{z}} = [\mathbf{z}^T, \mathbf{z}^H]^T$ [17], [26], [39], and corresponding the augmented covariance matrix

$$\underline{\mathbf{C}}_{a^2} = E[\hat{\mathbf{z}}\hat{\mathbf{z}}^H] = \begin{bmatrix} \mathcal{C} & \mathcal{P} \\ \mathcal{P}^* & \mathcal{C}^* \end{bmatrix} \tag{13}$$

which contains complete second order statistics of both covariance and pseudocovariance.

As mentioned earlier, the existing ACSP does not account for the *noncircularity* in the data. To that end, we propose the augmented complex common spatial pattern (ACCSP) algorithm, which produces second order practical spatial filters using the augmented covariance matrix. Given complex-valued data matrices with zero-mean, $\mathbf{Z}_a$ and $\mathbf{Z}_b \in \mathbb{C}^{N \times T}$ (for classes a and b), their augmented covariance matrices are calculated as

$$\hat{\mathbf{C}}_a = E\left[\hat{\mathbf{Z}}_a \hat{\mathbf{Z}}_a^H\right] = \begin{bmatrix} \mathbf{C}_a & \mathbf{P}_a \\ \mathbf{P}_a^* & \mathbf{C}_a^* \end{bmatrix} \tag{14}$$

$$\hat{\mathbf{C}}_b = E\left[\hat{\mathbf{Z}}_b \hat{\mathbf{Z}}_b^H\right] = \begin{bmatrix} \mathbf{C}_b & \mathbf{P}_b \\ \mathbf{P}_b^* & \mathbf{C}_b^* \end{bmatrix} \tag{15}$$

where $\hat{\mathbf{Z}}_a = [\mathbf{Z}_a^T, \mathbf{Z}_a^H]^T$ and $\hat{\mathbf{Z}}_b = [\mathbf{Z}_b^T, \mathbf{Z}_b^H]^T$. The augmented spatial filter, $\hat{\mathbf{W}}$, can then be obtained from the common eigenvector matrix between the whitened augmented covariance matrices of the classes, similar to (4) and (5). For a given augmented form of data $\hat{\mathbf{Z}}$, new set of signals $\hat{\mathbf{V}}$ can be obtained as

$$\hat{\mathbf{V}} = \hat{\mathbf{W}}\hat{\mathbf{Z}}. \tag{16}$$

The row vectors $\Re[\hat{\mathbf{v}}_p]$ and $\Im[\hat{\mathbf{v}}_p]$ ($p = 1, \ldots, m$ and $2N - m + 1, \ldots, 2N$) from $\hat{\mathbf{V}}$ can be now used to extract features to distinguish between two classes, using (8) and (9).

D. Augmented Complex Common Spatial Pattern With the Strong-Uncorrelating Transform

We next employ the strong-uncorrelating transform (SUT), an extension of the conventional whitening transform for improper complex random variables [28], [29] in the general context of CSP. This way, both the covariance and the pseudocovariance matrices are diagonalized, to guarantee that the resulting complex data are uncorrelated. The SUT transform $\mathbf{Q}$ has the following properties [29]

$$\mathbf{QCQ}^H = \mathbf{I} \text{ and } \mathbf{QPQ}^T = \mathbf{\Lambda} \tag{17}$$

where $\mathbf{C}$ is the covariance matrix, $\mathbf{P}$ the pseudocovariance matrix and $\mathbf{\Lambda}$ diagonal matrix of eigenvalues $\{\lambda_i\}, i \in \{1, \ldots, N\}$. In other words, by SUT both the covariance and the pseudocovariance are diagonalized simultaneously.

Given zero-mean complex-valued data matrices $\mathbf{Z}_a$ and $\mathbf{Z}_b$, the associated covariance and pseudocovariance matrices can be calculated from $\mathbf{C}_c = E[\mathbf{Z}_a\mathbf{Z}_a^H] + E[\mathbf{Z}_b\mathbf{Z}_b^H]$ and $\mathbf{P}_c = E[\mathbf{Z}_a\mathbf{Z}_a^T] + E[\mathbf{Z}_b\mathbf{Z}_b^T]$. After applying the whitening transformation matrix $\mathbf{G} = \mathbf{\Lambda}_c^{-(1/2)}\mathbf{U}_c^H$ to the covariance $\mathbf{C}_c$ ($\mathbf{\Lambda}_c$ and $\mathbf{U}_c$ from $\mathbf{C}_c = \mathbf{U}_c\mathbf{\Lambda}_c\mathbf{U}_c^H$), the symmetric pseudocovariance matrix can be decomposed as

$$\bar{\mathbf{P}} = \mathbf{GP}_c\mathbf{G}^T = \mathbf{Y\Lambda Y}^T. \tag{18}$$

Note that it is Takagi's factorization that enables such decomposition, which for symmetric matrices yields $\mathbf{Y}$ and $\mathbf{\Lambda}$ [40], based on which the SUT matrix $\mathbf{Q}$ can be defined as

$$\mathbf{Q} = \mathbf{Y}^H\mathbf{G} \tag{19}$$

and can diagonalise both the covariance and pseudocovariance matrices such that

$$\mathbf{QC}_c\mathbf{Q}^H = \mathbf{QC}_a\mathbf{Q}^H + \mathbf{QC}_b\mathbf{Q}^H = \mathbf{I} \tag{20}$$

$$\mathbf{QP}_c\mathbf{Q}^T = \mathbf{QP}_a\mathbf{Q}^T + \mathbf{QP}_b\mathbf{Q}^T = \mathbf{\Lambda}. \tag{21}$$

Let $\mathbf{S}_a = \mathbf{QC}_a\mathbf{Q}^H$ and $\mathbf{S}_b = \mathbf{QC}_b\mathbf{Q}^H$, then the estimates of common eigenvectors from the covariance matrices are obtained as

$$\mathbf{B}^{-1}\mathbf{S}_a\mathbf{B} = \mathbf{\Lambda}_a \text{ and } \mathbf{B}^{-1}\mathbf{S}_b\mathbf{B} = \mathbf{\Lambda}_b \quad (\mathbf{\Lambda}_a + \mathbf{\Lambda}_b = \mathbf{I}). \tag{22}$$

Similarly, if we wish to estimate common eigenvectors from the pseudocovariance matrices, then

$$\hat{\mathbf{Q}} = \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{Y}^H\mathbf{G} \tag{23}$$

$$\hat{\mathbf{S}}_a = \hat{\mathbf{Q}}\mathbf{P}_a\hat{\mathbf{Q}}^T \text{ and } \hat{\mathbf{S}}_b = \hat{\mathbf{Q}}\mathbf{P}_b\hat{\mathbf{Q}}^T \tag{24}$$

$$\hat{\mathbf{Q}}\mathbf{P}_c\hat{\mathbf{Q}}^T = \hat{\mathbf{S}}_a + \hat{\mathbf{S}}_b = \mathbf{I} \tag{25}$$

$$\hat{\mathbf{B}}^{-1}\hat{\mathbf{S}}_a\hat{\mathbf{B}} = \hat{\mathbf{\Lambda}}_a \text{ and } \hat{\mathbf{B}}^{-1}\hat{\mathbf{S}}_b\hat{\mathbf{B}} = \hat{\mathbf{\Lambda}}_b \ (\hat{\mathbf{\Lambda}}_a + \hat{\mathbf{\Lambda}}_b = \mathbf{I}). \tag{26}$$

In order to maximize the difference in variance between the two groups, we need to ensure that the SUT of the pseudocovariance matrix is an identity matrix, like that of the covariance matrix.[3] Then, two spatial filters for the covariance and pseudocovariance matrices can be obtained from

$$\mathbf{W} = \mathbf{B}^{-1}\mathbf{Q} \text{ and } \hat{\mathbf{W}} = \hat{\mathbf{B}}^{-1}\hat{\mathbf{Q}} \tag{27}$$

so that the transformed sets of data become

$$\mathbf{V} = \mathbf{WZ} \text{ and } \hat{\mathbf{V}} = \hat{\mathbf{W}}\mathbf{Z} \tag{28}$$

allowing us to calculate the two sets of features using (8) and (9).

Since the eigenvalues from the Takagi factorization of the pseudocovariance are guaranteed to be real-valued (from $\lambda = E[zz] = E[z_r^2] - E[z_i^2] + j2E[z_r z_i]$, when $z = z_r + jz_i$), this demonstrates that the eigenvalue $\lambda = E[z_r^2] - E[z_i^2]$ provides an additional information of the power difference between real and imaginary parts of complex sources. In this way, the power difference between real and imaginary parts of complex sources (two channel data) is accounted for by SUTCCSP, unlike the conventional CSP.

III. Analysis of Augmented Complex Common Spatial Pattern Methods

We shall next demonstrate the duality between the augmented complex CSP and the real-valued CSP, while the ACCSP is more compact and physically intuitive. Given a zero-mean complex random signal, $z(t) = z_r(t) + jz_i(t)$, where $z_r(t)$ and $z_i(t)$ are real-valued random signals, an augmented form of the complex signal is given by $\mathbf{Z}_a = [z(t); z^*(t)]$. The augmented complex matrix $\mathbf{Z}_a$ can be presented using $z_r(t)$ and $z_i(t)$ such that

$$\mathbf{Z}_a = \begin{bmatrix} z(t) \\ z^*(t) \end{bmatrix} = \begin{bmatrix} 1 & j \\ 1 & -j \end{bmatrix} \times \begin{bmatrix} z_r(t) \\ z_i(t) \end{bmatrix}. \tag{29}$$

Denote the transfer matrix and the real-valued matrix by

$$\mathbf{\Phi} = \begin{bmatrix} 1 & j \\ 1 & -j \end{bmatrix} \tag{30}$$

$$\mathbf{R} = \begin{bmatrix} z_r(t) \\ z_i(t) \end{bmatrix} \tag{31}$$

then the covariance matrix of $\mathbf{Z}_a$ is

$$\mathbf{Z}_a\mathbf{Z}_a^H = \mathbf{\Phi RR}^H\mathbf{\Phi}^H = \mathbf{\Phi RR}^T\mathbf{\Phi}^H. \tag{32}$$

[3]When the diagonalized composite covariance and pseudocovariance are identity matrices, their eigenvalue matrices also become identity matrices, shown in (22) and (26). These properties make the variances of the first rows of $\mathbf{V}$ and $\hat{\mathbf{V}}$ in (28) maximum for the trials of group "a" and at the same time minimum for the trials of group "b". On the other hand, the variances of the last rows of $\mathbf{V}$ and $\hat{\mathbf{V}}$ are minimal for the trials of group "a," and at the same time maximal for the trials of group "b".

After the whitening transformation of the covariance matrix, such that $E\{\mathbf{Z}_a\mathbf{Z}_a^H\} = \mathbf{I}$, we have

$$\mathbf{I} = \mathbf{\Phi} E\{\hat{\mathbf{R}}\hat{\mathbf{R}}^T\}\mathbf{\Phi}^H \tag{33}$$

$$\mathbf{\Phi}^{-1}(\mathbf{\Phi}^H)^{-1} = E\{\hat{\mathbf{R}}\hat{\mathbf{R}}^T\} \tag{34}$$

$$\frac{1}{2}\mathbf{I} = E\{\hat{\mathbf{R}}\hat{\mathbf{R}}^T\}. \tag{35}$$

Therefore, the real-valued covariance matrix $E\{\hat{\mathbf{R}}\hat{\mathbf{R}}^T\}$ is also diagonalized when the augmented complex covariance matrix $\mathbf{Z}_a\mathbf{Z}_a^H$ is diagonalized. This means the performance of ACCSP is similar to that of CSP, subject to the scaling factor 1/2.

IV. Experiments

We now illustrate the performance of the proposed algorithms via simulations using both synthetic data and real-world EEG data. A support vector machine (SVM) [41] with a Gaussian kernel[4] was used to obtain the classification results for the features from the algorithms. All the data sets were divided into 80% training and 20% testing sets. The classification procedure was repeated five times while changing the sample order (five-fold cross-validation), and the average of these outcomes was the final classification rate. The constant m in (8) and (9) was set to unity and three for the synthetic and EEG data, defining the number of features for all CSP algorithms.

A. Synthetic Data

Two synthetic datasets were designed to model different EEG-like scenarios. The first dataset used sinusoids, to follow the example of Koles and Soong [43] who employed sum of sinusoids as a synthetic EEG to validate their source localization algorithm, and ACSP was also verified using sinusoids [23]. The second dataset used bandpass filtered EEG signal containing the frequency components in the range of alpha and beta bands (8–30 Hz), which was employed by Park *et al.* [35]. The synthetic datasets were composed by summing two sinusoids, whose frequencies were in the range of alpha (8–13 Hz) and beta (13–30 Hz) bands, and corrupted by complex-valued random noise, to generate

$$\text{Class A}$$

$$(50 \text{ data sets with different complex} - \text{valued noises})$$

$$\begin{bmatrix} s_1(t) + 1.05 \times js_1(t) + v_1(t) \\ 1.11 \times s_1(t) + 1.15 \times js_1(t) + v_2(t) \end{bmatrix}$$

$$\text{Class B}$$

$$(50 \text{ data sets with different complex} - \text{valued noises})$$

$$\begin{bmatrix} 1.18 \times s_2(t) + 1.17 \times js_2(t) + v_3(t) \\ 1.02 \times s_2(t) + 1.04 \times js_2(t) + v_4(t) \end{bmatrix}$$

$$s_1(t) = sin(2\pi f_1 t) + sin(2\pi f_2 t)$$

$$s_2(t) = sin(2\pi f_3 t) + sin(2\pi f_4 t)$$

$$f_1 = 10 \text{ Hz}, f_2 = 19 \text{ Hz}, f_3 = 9 \text{ Hz and } f_4 = 17 \text{ Hz}$$

where $v_1(t), \ldots, v_4(t)$ are different realizations of correlated random noises with signal-to-noise ratios (SNR) varying from -9.5 dB to -16.3 dB.

[4]The MATLAB code can be downloaded from [42].

The second synthetic dataset using bandpass filtered EEG were designed in the same way as the sinusoid synthetic dataset except $s_1(t)$ and $s_2(t)$. Two EEG signals were obtained from electrode Cz according to the 10-10 system, sampled at 160 Hz, and filtered using a Butterworth filter, occupying the alpha and beta bands (8–30 Hz), to produce $s_1(t)$ and $s_2(t)$. The complex-valued dataset of Class A had larger amplitudes in the second channel than in the first one, while Class B had the larger amplitudes in the first channel. The power differences of the complex-valued data were 0.05 and 0.04 for class A and 0.01 and 0.02 for class B, which will be used as additional features for SUTCCSP. The information about these amplitude and power difference were identified by the SUTCCSP spatial patterns[5] owing to the use of both the covariance and pseudocovariance, as shown in Fig. 1 for the first and the last spatial patterns of the sinusoid synthetic dataset. The correlation between real and imaginary parts of the complex-valued random noise varied from 0.1 to 0.9. Fig. 2 illustrates the distribution of the complex-valued data with correlation degrees of 0 and 0.9 in -9.5 dB SNR. Their degrees of circularity[6] are governed by the correlation between the real and imaginary parts; lower correlation produces more *circular* data and higher correlation highly *noncircular* data. In addition, Fig. 3 shows the power difference[7] between real and imaginary parts decreases as the correlation increases in SNR -9.5 and -14.8 dB. Owing to this additional feature of power difference provided by augmented complex statistics, the classification rates of SUTCCSP were higher than those of CSP, ACSP, and ACCSP algorithms for most correlated data of sinusoid and bandpass filtered EEG synthetic datasets. The classification rates become close to the results of CSP and ACCSP for the data with higher degrees of correlation, due to their small power differences. As seen in Fig. 4, the improvement offered by SUTCCSP is reconfirmed over various SNRs from -9.5 to -16.3 dB for the correlation degree of 0.5. In particular, the better performance of SUTCCSP is visible in the critical case of low SNRs. Additionally, the CSP and ACCSP had similar performances as discussed in Section III.

B. Motor Imagery EEG Datasets

The Physiobank Motor/Mental Imagery (MMI) database recorded using the BCI2000 system [44], which is available through Physionet[8] [45], was used to test our CSP algorithms. The datasets consists of 109 subjects performing different motor imagery tasks while 64-channel EEG were recorded according to 10-10 system, sampled at 160 Hz. We chose the blocks where the subjects imagined movement of *left hand* and *right hand*. Subjects performed a total of 45 trials and imagined one of the two tasks for a duration of 4 s. Out of the 64 EEG

[5]Spatial patterns using covariance matrix, SP_{cov}, and pseudocovariance matrix, SP_{pcov}, are obtained by the inverse matrices of spatial filters, $\mathbf{W}$ and $\hat{\mathbf{W}}$ in (28)

[6]*Circularity* refers to shape of a scatter plot of complex-valued data, *circular* meaning rotation-invariant. A complex signal is circular if its real and imaginary parts are uncorrelated and have the same power. The *circularity* degree of a zero-mean complex signal $z(t)$ $(t = 1 \ldots T)$, ρ, is defined as $\rho = |E[\sum_{t=1}^{T} z(t)z(t)]|/E[\sum_{t=1}^{T} z(t)z^*(t)]$ $(0 < \rho < 1)$.

[7]The normalized power difference of a complex signal $z(t) = z_r(t) + jz_i(t)$, δ, is defined as $\delta = |E[z_r^2(t) - z_i^2(t)]|/E[z_r^2(t) + z_i^2(t)]$ $(0 < \delta < 1)$

[8]http://www.physionet.org/pn4/eegmmidb/#experimental-protocol.

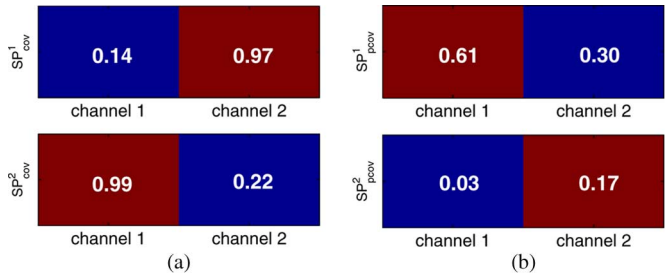


Fig. 1. Spatial patterns of the synthetic sinusoid dataset ($\mathrm{correlation} = 0.1$ and $\mathrm{SNR} = -9.5$ dB) using both the covariance and pseudocovariance matrices. The numbers in the diagrams denote the coefficient of spatial patterns. Note that the spatial patterns using pseudocovariance contains the information about the power difference between the real and imaginary parts, while spatial patterns using covariance contain the information about the magnitude of the complex data. (a) Spatial patterns using covariance. (b) Spatial patterns using pseudocovariance.

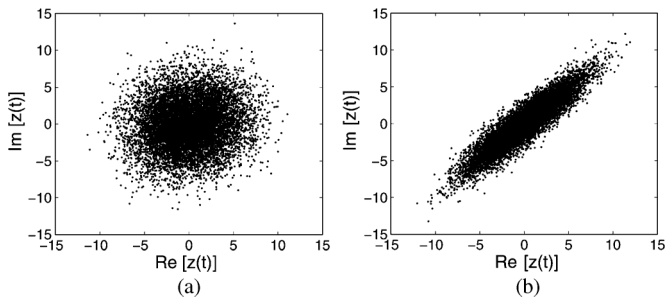


Fig. 2. Geometric view of *circularity* via "real-imaginary" scatter plots: (a) *circular* data ($\mathrm{correlation} = 0$); (b) *noncircular* data ($\mathrm{correlation} = 0.9$).

channels, 58 shown in Fig. 5 were selected for the analysis, and the motor imagery EEG data was band-pass filtered to occupy the frequency band 8–30 Hz [12], [13]. Pairing spatially adjacent electrodes, as shown in Fig. 5, to form complex-valued data facilitated the use of cross-information and a simultaneous estimation of the spatial amplitude relationships. Solodkin *et al.* [11] investigated brain connectivity within functional networks during motor imagery, and demonstrated networks underlying the mental behavior on M1 (primary motor cortex), S1 (primary and secondary somatosensory cortices), LPMC (lateral premotor cortex, dorsal), SMA (supplementary and pre-supplementary motor areas), CRB (cerebellum), PAR (superior parietal lobule and intra-parietal sulcal area), IF (inferior frontal cortex, LPMC ventral; inferior frontal gyrus and anterior insular cortex), and OCC (occipital lobe) using correlation analysis. The chosen 58 electrodes in Fig. 5 cover most scalp regions in these areas, in particular, they observed strong connection between the adjacent brain regions. Owing to their high correlation, smaller power differences can be expected between the adjacent areas, as shown in Fig. 3, that is, the higher correlation between two data sources the lower the power difference. In addition, the changes in brain connectivity depending on the different motor imagery tasks have been observed by Chung *et al.* [22]. Therefore, the power difference between the paired electrodes is dynamically changing depending on the mental tasks, which can be useful information to distinguish between the different tasks.

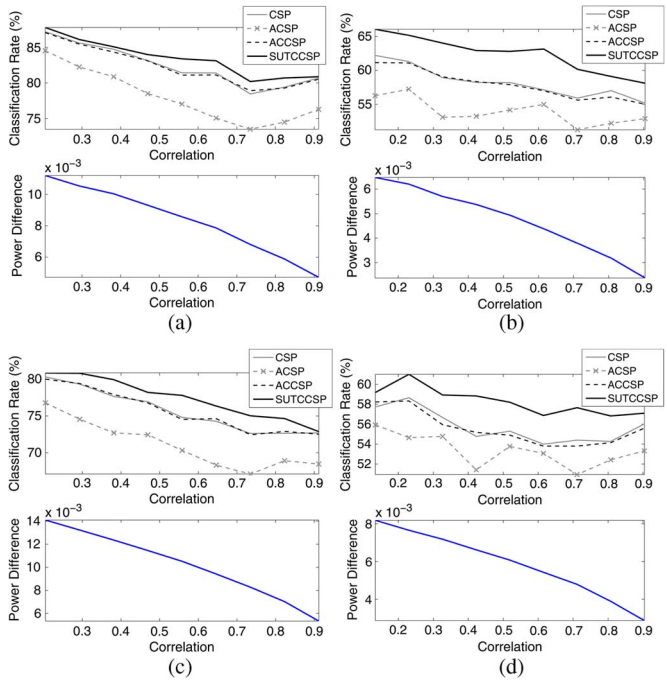


Fig. 3. Classification results for synthetic complex data using sinusoids and bandpass filtered EEG corresponding to the channel correlation and power difference in SNR -9.5 dB and -14.8 dB. Note the better performances of SUTCCSP compared to CSP, ACSP, and ACCSP for both data, and the performance of SUTCCSP becomes close to those of CSP and ACCSP when the power difference is small. The more correlated the data, the less power difference observed. (a) Sinusoid ($\mathrm{SNR} = -9.5$ dB). (b) Sinusoid ($\mathrm{SNR} = -14.8$ dB). (c) Bandpass filtered EEG ($\mathrm{SNR} = -9.5$ dB). (d) Bandpass filtered EEG ($\mathrm{SNR} = -14.8$ dB).

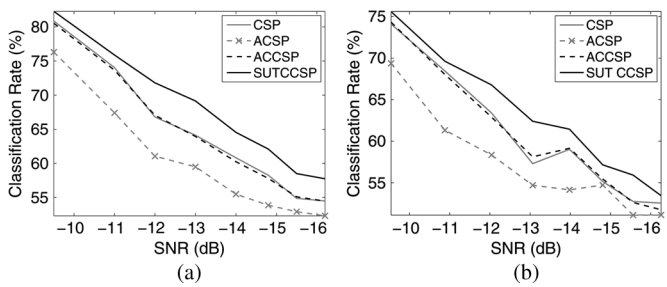


Fig. 4. Classification results for synthetic complex data using sinusoid and bandpass filtered EEG (for the 0.5 correlation coefficient between real and imaginary part) corresponding to various SNRs. Note the better performance of SUTCCSP compared to CSP, ACSP, and ACCSP in all conditions for both sinusoid and bandpass filtered EEG synthetic data. (a) Sinusoid. (b) Bandpass filtered EEG.

C. Classification Results

The classification accuracies estimated using real and complex CSP methods were compared and illustrated in Fig. 6. The upper limit of confidence intervals between two classes was 64.0% for 45 trials (cf. [46]). Among the 109 subjects in Physiobank MMI database, the subjects for whom the classification results were below 64% using all four CSP, ACSP, ACCSP, and SUTCCSP methods, were discarded so that Fig. 6 displays the results of 56 significant subjects using whisker diagram. The red lines and cross in the boxplot denote median and outlier. On the average, the SUTCCSP gave the best classification performance of 72.37%, a 4.29% improvement over CSP, a 5.09% im-

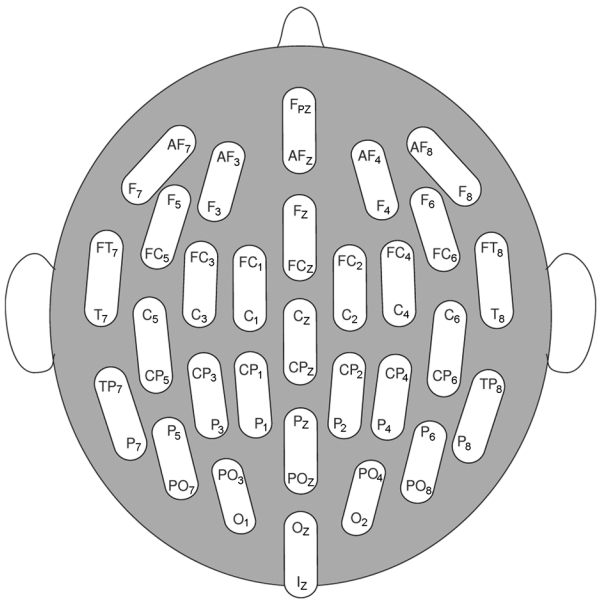


Fig. 5. EEG montage of paired channels for complex-valued EEG data.

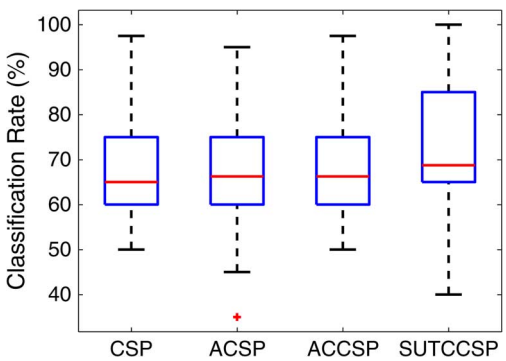


Fig. 6. Whisker diagrams of the classification rates of motor imagery EEG for 56 subjects out of 109, who had significant separation results. The red lines and cross in the boxplot denote median and outlier. The enhanced performance of SUTCCSP compared to CSP, ACSP, and ACCSP are confirmed by a one-way analysis of variance yielding a p-value less than 0.05.

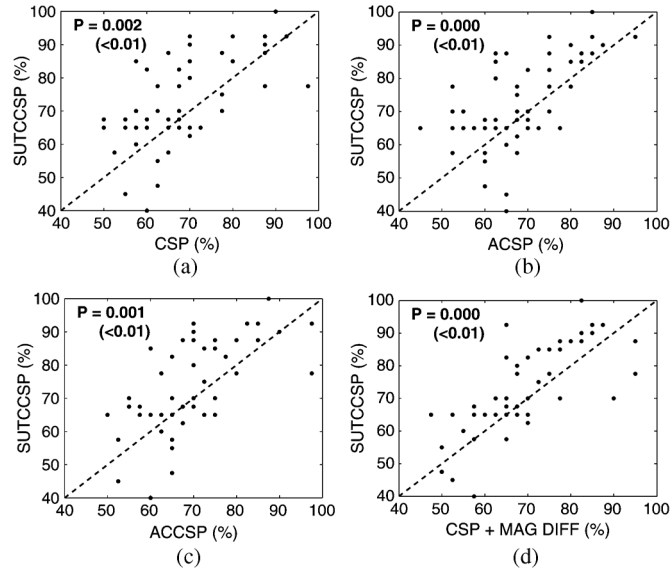


Fig. 7. Scatter-plot of classification results obtained using CSP, ACSP, ACCSP, SUTCCSP, and CSP which accounts for magnitude difference. Dots above the diagonal denote the cases when SUTCCSP outperforms the others. Significance of these improvements are confirmed by the one-tailed p-values of the t-test (less than 0.01). (a) SUTCCSP versus CSP. (b) SUTCCSP versus ACSP. (c) SUTCCSP versus ACCSP. (d) SUTCCPS versus CSP+MAG DIFF.

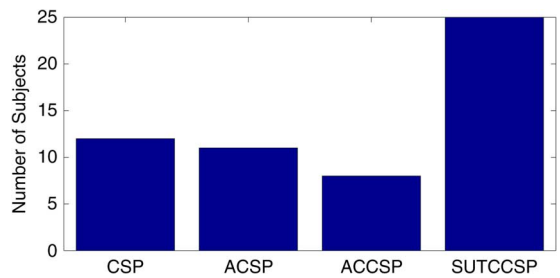
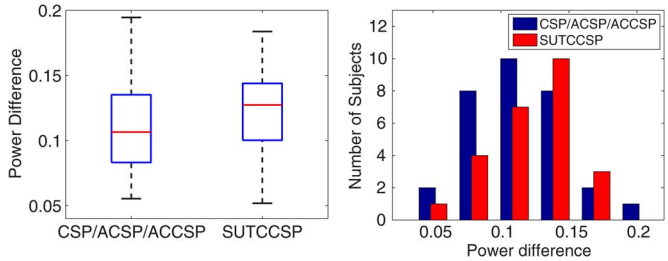


Fig. 8. Number of subjects whose best classification rates were obtained using one of CSP, ACSP, ACCSP, and SUTCCSP. Note the high performance when using SUTCCSP than the others.

provement over ACSP, and a 4.38% improvement over ACCSP. The enhanced performance of SUTCCSP were confirmed by a one-way analysis of variance (ANOVA) yielding a p-value 0.049 (<0.05). As expected, CSP and ACCSP had similar performances as exemplified in the simulations in Section IV-A.

Fig. 7(a)–(c) displays a series of scatter-plots of classification results between SUTCCSP and the other methods considered using the results in Fig. 6. The values should lie on the diagonal if the two algorithms have the same performance, however, it is notable that many of values lie above the diagonal, indicating an improved performance of SUTCCSP compared to the other methods. For rigor, the difference in classification rates between SUTCCSP and the other algorithms was analyzed using a one-tailed t-test. The corresponding p-values in Fig. 7(a)–(c) prove that the proposed SUTCCSP performs significantly better than the other algorithms (p-values less than 0.01). In order to demonstrate the effective feature extraction of SUTCCSP compared to CSP, the classification performance of the CSP feature with magnitude differences of the paired channels was calculated and compared with those of SUTCCSP in Fig. 7(d). The scatter-plot displays more dots above the diagonal, which indicates CSP cannot outperform SUTCCSP even though it also uses the information about magnitude difference.

We next calculated a histogram of subjects, whose best classification rates were obtained using one of CSP, ACSP, ACCSP, and SUTCCSP in Fig. 8. Observe the highest number of subjects for SUTCCSP, 25, which is twice the second highest one, obtained for CSP. In order to investigate the contribution of power difference feature to the performance of SUTCCSP, the average power differences between the paired EEG data for all significant 56 subjects were calculated, and their distributions based on the histogram in Fig. 8 were investigated, that is, one distribution for the subjects who made the best performance using CSP, ACSP or ACCSP versus the other distribution for the subjects who made the best using SUTCCSP. Fig. 9 illustrates the two distributions using the whisker diagram and histogram, where the red line in the boxplot denotes the median. Note that the power differences of SUTCCSP are distributed above those of



(a) Whisker diagram (b) Histogram

Fig. 9. Whisker diagrams (a) and histograms (b) of power differences between the paired EEG data of the subjects, who made the best classification rates using "CSP, ACSP, or ACCSP" and "SUTCCSP" based on Fig. 8. Red lines in the boxplot denote median. Note the higher power difference distribution of SUTCCSP compared to the other methods. (a) Whisker diagram. (b) Histogram.

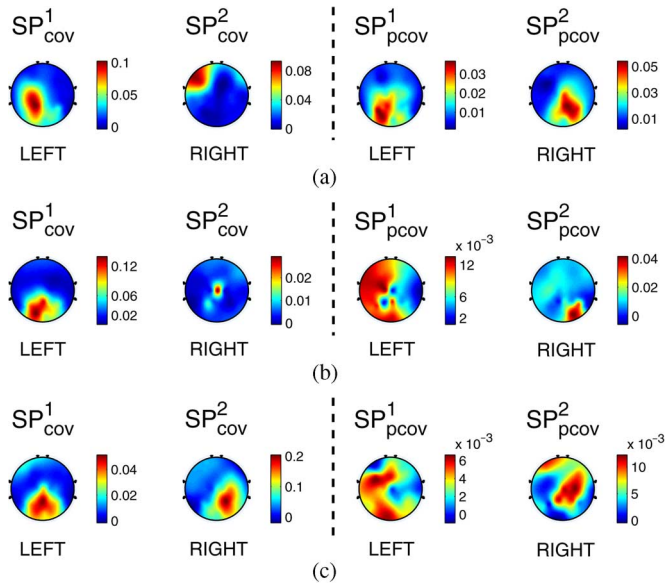


Fig. 10. Spatial patterns [obtained by the inverse matrices of $\mathbf{W}$ and $\hat{\mathbf{W}}$ in (28)] of the subjects who had classification rates over 90% using SUTCCSP. Symbols SP^1_{cov} and SP^2_{cov} denote two most significant spatial patterns obtained using covariance matrices, and SP^1_{pcov} and SP^2_{pcov} those using pseudocovariance matrices. "LEFT" and "RIGHT" denote left- and right-hand motor imagery, respectively. Note the different spatial patterns between the results of covariance and pseudocovariance matrices. (a) Subject 29. (b) Subject 48. (c) Subject 94.

CSP/ACSP/ACCSP, which is consistent with the results of simulation in Section IV-A. This suggests that the power difference feature is accounted for using SUTCCSP by its very design, and that this improves the classification performance for motor imagery data. Also, since the paired real-world EEG data cannot be identical, their complex-valued versions always contain certain amount of power difference between the real and imaginary parts (*noncircularity* of complex-valued EEG data).

Fig. 10 illustrates the spatial patterns of three subjects using SUTCCSP, whose classification rates were over 90%. This includes the spatial patterns using covariance, SP_{cov}, and pseudocovariance, SP_{pcov}, which illustrate some amount of brain lateralization for the two different mental tasks ("LEFT" left-hand motor imagery and "RIGHT" right-hand motor imagery). The brain lateralization of event-related desynchronization during the motor imagery of left and right hands is a well known phenomenon [9], and Chung *et al.* [22] observed the contralateral connectivity during the motor imagery of left and right hands. In particular, subject 29 and 94 show the salient decrease of SP_{pcov} coefficients on the contralateral motor cortex area, while the contralateral differences of SP_{cov} corresponding to the motor tasks are less clear. This is a good example to show that the spatial filters using the pseudocovariance can produce the complementary information, and how such information can be used to estimate true spatial patterns of multichannel EEG.

Finally, the performance of ACSP was most of the time worse than those of CSP, ACCSP, and SUTCCSP for both synthetic data and motor imagery EEG data. As mentioned earlier, the inadequacy of the Hilbert transform for nonbandlimited data is the most likely cause of the poor performance of ACSP.

V. Conclusion

A complex extension of CSP has been introduced to deal with general data exhibiting complex *noncircularity*, a unique feature of real world complex-valued signals. It has been shown that the complex CSP equipped with the strong uncorrelating transform (SUTCCSP) can improve the classification performance between two groups, when there exists a degree of power difference between the real and imaginary parts of complex data, a typical case in practice. The robustness of the proposed approaches has also been demonstrated by the low SNR (from -9.5 to -16.3 dB) in synthetic EEG datasets. In the context of real-world EEG, we have considered a motor imagery dataset of 109 subjects, where the optimal performance in distinguishing between two mental tasks has been achieved based on augmented complex statistics, using the SUTCCSP method applied to the complex-valued EEG composed by combining two closely located real-valued EEG signals.

References

[1] S. Sanei and J. Chambers, *EEG Signal Processing*. New York: Wiley, 2007.

[2] A. Akrami, S. Solhjoo, A. Motie-Nasrabadi, and M. Hashemi-Golpayegani, "EEG-based mental task classification: Linear and nonlinear classification of movement imagery," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2005, pp. 4626–4629.

[3] M. Jeannerod, "Mental imagery in the motor context," *Neuropsychologia*, vol. 33, no. 11, pp. 1419–1432, 1995.

[4] C. A. Porro, M. P. Francescato, V. Cettolo, M. E. Diamond, P. Baraldi, C. Zuiani, M. Bazzocchi, and P. E. di Prampero, "Primary motor and sensory cortex activation during motor performance and motor imagery: A functional magnetic resonance imaging study," *J. Neurosci.*, vol. 16, no. 23, pp. 7688–7698, 1996.

[5] G. Pfurtscheller, C. Neuper, D. Flotzinger, and M. Pregenzer, "EEG-based discrimination between imagination of right and left hand movement," *Electroencephalogr. Clin. Neurophysiol.*, vol. 103, no. 6, pp. 642–651, 1997.

[6] C. Park, D. Looney, N. Rehman, A. Ahrabian, and D. P. Mandic, "Classification of motor imagery BCI using multivariate empirical mode decomposition," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 1, pp. 10–22, Jan. 2013.

[7] D. J. McFarland, L. A. Miner, T. M. Vaughan, and J. R. Wolpaw, "Mu and beta rhythm topographies during motor imagery and actual movements," *Brain Topogr.*, vol. 12, no. 3, pp. 177–186, 2000.

[8] G. Pfurtscheller and F. H. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: Basic principles," *Clin. Neurophysiol.*, vol. 110, no. 11, pp. 1842–1857, 1999.

[9] H. Yuan, A. Doud, A. Gururajan, and B. He, "Cortical imaging of event-related (de)synchronization during online control of brain-computer interface using minimum-norm estimates in frequency domain," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 16, no. 5, pp. 425–431, Oct. 2008.

[10] V. V. Nikouline, K. Linkenkaer-Hansen, H. Wikstrom, M. Kesaniemi, E. V. Antonova, R. J. Ilmoniemi, and J. Huttunen, "Dynamics of mu-rhythm suppression caused by median nerve stimulation: A magnetoencephalographic study in human subjects," *Neurosci. Lett.*, vol. 294, no. 3, pp. 163–166, 2000.

[11] A. Solodkin, P. Hlustik, E. E. Chen, and S. L. Small, "Fine modulation in network activation during motor execution and motor imagery," *Cerebral Cortex*, vol. 14, no. 11, pp. 1246–1255, 2004.

[12] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, pp. 441–446, Dec. 2000.

[13] J. Muller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," *Clini. Neurophysiol.*, vol. 110, no. 5, pp. 787–798, 1999.

[14] H. Lu, H. Eng, C. Guan, K. Plataniotis, and A. Venetsanopoulos, "Regularized common spatial pattern with aggregation for EEG classification in small-sample setting," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 12, pp. 2936–2946, Dec. 2010.

[15] Z. J. Koles, "The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG," *Electroencephalogr. Clin. Neurophysiol.*, vol. 79, pp. 440–447, 1991.

[16] 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.*, 2007, pp. 384–389.

[17] D. P. Mandic and S. L. Goh, *Complex Valued Nonlinear Adaptive Filters*. Hoboken, NJ: Wiley, 2009.

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

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

[20] S. Javidi, D. P. Mandic, and A. Cichocki, "Complex blind source extraction from noisy mixtures using second-order statistics," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 57, no. 7, pp. 1404–1416, Jul. 2010.

[21] S. Javidi, D. Mandic, C. C. Took, and A. Cichocki, "Kurtosis based blind source extraction of complex noncircular signals with application in EEG artifact removal in real-time," *Front. Neurosci.*, vol. 5, no. 105, pp. 1–18, 2011.

[22] Y. Chung, J. Kang, and S. Kim, "Analysis of correlated EEG activity during motor imagery for brain-computer interfaces," in *Proc. Control, Automat. Syst.*, 2011, pp. 337–341.

[23] O. Falzon, K. P. Camilleri, and J. Muscat, "Complex-valued spatial filters for task discrimination," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2010, pp. 4707–4710.

[24] O. Falzon, K. P.Camilleri, and J. Muscat, "The analytic common spatial patterns method for EEG-based BCI data," *J. Neural Eng.*, vol. 9, no. 4, p. 045009, 2012.

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

[26] D. Looney and D. P. Mandic, "Augmented complex matrix factorisation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2011, pp. 4072–4075.

[27] C. C. Took, D. P. Mandic, and S. C. Douglas, "On approximate diagonalisation of correlation matrices in widely linear signal processing," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1469–1473, Mar. 2012.

[28] L. D. Lathauwer and B. D. Moore, "On the blind separation of noncircular source," in *Proc. 11th Eur. Signal Process. Conf.*, Sep. 2002, vol. 2, pp. 99–102.

[29] 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.

[30] O. Falzon, K. P. Camilleri, and J. Muscat, "Complex-valued spatial filters for SSVEP-based BCIs with phase coding," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 9, pp. 2486–2495, Sep. 2012.

[31] J. Navarro-Moreno, M. D. Estudillo-Martinez, R. M. Fernandex-Alcala, and J. C. Ruiz-Molina, "Estimation of improper complex-valued random signals in colored noise by using the Hilbert space theory," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2859–2867, Jun. 2009.

[32] R. Bhatia, *Positive Definite Matrices*, ser. Princeton series in applied mathematics. Princeton, NJ: Princeton Univ. Press, 2007.

[33] S. R. Long, N. E. Huang, C. C. Tung, M. L. Wu, R. Q. Lin, E. Mollo-Christensen, and Y. Yuan, "The Hilbert techniques: An alternate approach for non-steady time series analysis," *IEEE Geosci. Remote Sens. Soc. Lett.*, vol. 3, pp. 6–11, 1995.

[34] N. E. Huang, Z. Shen, S. R. Long, M. L. Wu, H. H. Shih, Z. Quanan, 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," in *Proc. R. Soc. A*, 1998, vol. 454, no. 1971, pp. 903–995.

[35] C. Park, D. Looney, P. Kidmose, M. Ungstrup, and D. P. Mandic, "Time-frequency analysis of EEG asymmetry using bivariate empirical mode decomposition," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 19, no. 4, pp. 366–373, Aug. 2011.

[36] D. P. Mandic, N. Rehman, Z. Wu, and N. Huang, "Empirical mode decomposition based time-frequency analysis of multivariate signals," *IEEE Signal Process. Mag.*, vol. 30, no. 6, pp. 74–86, Nov. 2013.

[37] P. J. Schreier and L. L. Scharf, "Second-order analysis of improper complex random vectors and processes," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 714–725, Mar. 2003.

[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] Y. Xia, B. Jelfs, M. M. V. Hulle, J. C. Principe, and D. P. Mandic, "An augmented echo state network for nonlinear adaptive filtering of complex noncircular signals," *IEEE Trans. Neural Netw.*, vol. 22, no. 1, pp. 74–83, Jan. 2011.

[40] S. C. Douglas, J. Eriksson, and V. Koivunen, "Adaptive estimation of the strong uncorrelating transform with applications to subspace tracking," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2006, pp. 941–944.

[41] H. Drucker, C. J. Burges, L. Kaufman, A. Smola, and V. Vapnik, *Support Vector Regression Machines*. Cambridge, MA: MIT Press, 1996.

[42] S. Canu, Y. Grandvalet, V. Guigue, and A. Rakotomamonjy, "SVM and kernel methods MATLAB toolbox," in *Percept. Syst. Inf.*, Rouen, France, 2005.

[43] Z. J. Koles and A. C. K. Soong, "EEG source localization: Implementing the spatio-temporal decomposition approach," *Electroencephalogr. Clin. Neurophysiol.*, vol. 107, no. 5, pp. 343–352, 1998.

[44] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "BCI2000: A general purpose brain-computer interface (BCI) system," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1034–1043, Jun. 2004.

[45] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. Ivanov, R. Mark, J. Mietus, G. Moody, C.-K. Peng, and H. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, pp. e215–e220, 2000.

[46] G. R. Müller-Putz, R. Scherer, C. Brunner, R. Leeb, and G. Pfurtscheller, "Better than random? A closer look on BCI results," *Int. J. Bioelectromagn.*, vol. 10, no. 1, pp. 52–55, 2008.

Cheolsoo Park received the B.Eng. degree in electrical engineering from Sogang University, Seoul, South Korea, the M.Sc. degree in biomedical engineering department from Seoul National University, Seoul, South Korea, and the Ph.D. degree in adaptive nonlinear signal processing from Imperial College London, London, U.K., in 2012.

Currently, he is a Postdoctoral Researcher at University California-San Diego, La Jolla, CA, USA. His research interests are mainly in the area of machine learning, adaptive and statistical signal processing, with applications in brain–computer interface and computational neuroscience.

Clive Cheong Took (SM'13) received the B.S. degree in telecommunication engineering from King's College London, London, U.K., in 2004, and read his doctoral degree in signal processing at Cardiff University, Cardiff, U.K., in 2007.

From 2007 to 2012, he worked as Research Associate within the Department of Electronic Engineering at Imperial College London. He is currently a Lecturer with the Department of Computing, University of Surrey, Surrey, U.K. His current research interests include machine learning, adaptive and blind signal processing, and neural networks, with applications in telecommunication, finance, renewable energy, and biomedicine.

Danilo P. Mandic (F'13) is a Professor of signal processing at Imperial College London, London, U.K. He has been working in the area of nonlinear and multivariate adaptive signal processing and nonlinear dynamics. He has published two research monographs titled "Recurrent Neural Networks for Prediction" and "Complex Valued Nonlinear Adaptive Filters." He has been a guest Professor at KU Leuven, Belgium and frontier researcher at RIKEN Japan.

He has been a member of the IEEE Technical Committee on Signal Processing Theory and Methods, and an Associate Editor of IEEE TRANSACTIONS ON SIGNAL PROCESSING and IEEE TRANSACTIONS ON NEURAL NETWORKS. He has won best paper awards in the area of neurotechnology in ICIC'09 and ISNN'10, and his work on Ear-EEG has been nominated for the Annual BCI Award in 2012.