

Quaternion Common Spatial Patterns

S. Enshaeifar, *Member, IEEE*, C. Cheong Took, *Senior Member, IEEE*,
C. Park, and D. P. Mandic, *Fellow, IEEE*

Abstract—A novel quaternion-valued common spatial patterns (QCSP) algorithm is introduced to model co-channel coupling of multi-dimensional processes. To cater for the generality of quaternion-valued non-circular data, we propose a generalized QCSP (G-QCSP) which incorporates the information on power difference between the real and imaginary parts of data channels. As an application, we demonstrate how G-QCSP can be used to provide high classification rates, even at a signal-to-noise ratio (SNR) as low as -10 dB. To illustrate the usefulness of our method in EEG analysis, we employ G-QCSP to extract features for discriminating between imagery left and right hand movements. The classification accuracy using these features is 70%. Furthermore, the proposed method is used to distinguish between Parkinson's disease (PD) patients and healthy control subjects, providing an accuracy of 87%.

Index Terms—Augmented statistics, common spatial pattern, quaternion domain.

I. INTRODUCTION

BRAIN-computer interface (BCI) refers to a computer-aided control system which uses the brain activity as a communication channel between human and computers [1]. One of the most common/convenient methods to measure the required brain activity is the scalp electroencephalogram (EEG) which measures the neurophysiological activity using a set of non-invasive electrodes. Despite this advantage, EEG has a poor spatial resolution and it provides rather blurred images of brain activity due to the presence of system and physiological noise [1], [2]. Thus, spatial filters are often used to improve the signal-to-noise ratio (SNR). A well-known spatial filtering algorithm in BCI research is the common spatial patterns (CSP) algorithm which enhances separability between the two classes of multi-channel EEG by providing their spatial patterns [2]. To this end, CSP simultaneously diagonalises the covariance matrices of the two classes to maximise the variance of signals belonging to one class, while minimizing the variance of signals of the other class. In other words, CSP aims to maximise the ratio

$$\frac{\mathbf{W}\mathbf{C}_a\mathbf{W}^T}{\mathbf{W}\mathbf{C}_b\mathbf{W}^T} \quad (1)$$

in which \mathbf{W} is the spatial filter and \mathbf{C}_a and \mathbf{C}_b are the covariance matrices for two classes a and b . In particular, the CSP algorithm is used in the analysis of brain activities

Manuscript received January 19, 2016; revised July 29, 2016 and September 23, 2016; accepted October 24, 2016. Date of publication November 3, 2016; date of current version August 7, 2017.

S. Enshaeifar and C. C. Took are with the Faculty of Engineering and Physical Sciences, University of Surrey, Guildford, GU2 7XH, U.K.

C. Park is with the Department of Computer Engineering, Kwangju University, Seoul 01897, South Korea.

D. P. Mandic is with the Department of Electrical and Electronic Engineering, Imperial College London, London, SW7 2AZ, U.K.

Digital Object Identifier 10.1109/TNSRE.2016.2625039

originated from spatially distributed regions, such as motor-imagery (MI) tasks¹ [3]. Generally, MI-based CSP applications exploit information on amplitude modulations of sensory motor rhythms (SMR), which represent the motor intention of the subject [4]. Hence, these applications have great potentials to be employed in motor rehabilitation and its usefulness for the motor function impaired patients [4].

Note that CSP is prone to overfitting and highly sensitive to noise. These issues were addressed in [5] by proposing a regularized CSP, and in [4] by exploiting local temporal correlation (LTC) information to enhance the estimation of covariance matrices. Furthermore, a complex-valued CSP (CCSP) algorithm was recently introduced to model the bi-variate correlation in EEG processing [6]. However, two-channel processing might not be adequate to model the correlation between *multiple* processes originating from different brain regions. Thus, we generalize the complex domain to quaternion (\mathbb{H}) to model four-dimensional signals considering their inherent multi-channel coupling. To this end, four real-valued channels $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_4$, form a single quaternion-valued process \mathbf{q} as

$$\begin{aligned} \mathbf{q} &= \mathbf{q}_1 + i\mathbf{q}_2 + j\mathbf{q}_3 + \kappa\mathbf{q}_4 \\ &= \Re[\mathbf{q}] + \Im_i[\mathbf{q}] + \Im_j[\mathbf{q}] + \Im_\kappa[\mathbf{q}] \end{aligned} \quad (2)$$

where $\Re[\cdot]$ is the scalar part and $\Im_\eta[\cdot]$ represents the η -imaginary component, $\eta \in \{i, j, \kappa\}$. Furthermore, to cater for quaternion-valued signals with non-circular probability distribution, we exploit the latest advances in quaternion statistics, so-called “augmented statistics” [7]. Modelling based on augmented quaternion has been previously deployed in biomedical applications such as ocular artefacts [8], sleep EEG analysis [9], and other applications [10]. In the same spirit, we introduce an augmented quaternion CSP (A-QCSP) and a generalized quaternion CSP (G-QCSP) to exploit the inherent coupling between four channels and cater for non-circular datasets.

To evaluate the performance of A-QCSP and G-QCSP methods, two real-world applications have been considered: 1) discrimination of two imagery tasks, and 2) classification of Parkinson's disease (PD) patients using multi-dimensional pressure sensors. The proposed methods have been compared with real-valued CSP and CCSP in both synthetic and real-world simulations.

II. QUATERNION COMMON SPATIAL PATTERNS

A complex extension of CSP was first introduced in [11] where the Hilbert transform was combined with the CCSP to exploit the phase information. Furthermore, [6] established

¹Positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) studies have demonstrated that motor-imagery (MI) tasks activate several brain regions which are spatially distributed.

rigorous extensions of CCSP, an augmented CCSP (ACCSP) and a CCSP equipped with strong-uncorrelating transform (SUTCCSP), to cater for the complex non-circularity distribution. In terms of MI task analysis, [6] illustrated that SUTCCSP provided the highest classification rate by considering the power difference between the real and imaginary parts of a complex-valued EEG channel. However, MI tasks involve several brain areas and two-channel pairing offered by complex domain may not be adequate to model the correlation between multiple processes. Therefore, for the first time, we propose the quaternion CSP (QCSP) which incorporates the inherent coupling between four channels for convenience of enhanced representation. Consider two classes a and b represented by quaternion-valued zero-mean processes² \mathbf{x}_a and $\mathbf{x}_b \in \mathbb{H}^{N \times T}$ where their covariance matrices are calculated as

$$\mathbf{C}_d = E[\mathbf{x}_d \mathbf{x}_d^H] \in \mathbb{H}^{N \times N}, d \in \{a, b\}. \quad (3)$$

The spatial covariance of each class can now be obtained by averaging the covariance matrices of all the processes in that class as $\bar{\mathbf{C}}_d$. Since the QCSP filter aims to simultaneously diagonalise these covariance matrices, the composite spatial covariance matrix is given by

$$\mathbf{C}_c = \bar{\mathbf{C}}_a + \bar{\mathbf{C}}_b. \quad (4)$$

Note that \mathbf{C}_c is a Hermitian matrix which can be factorized as $\mathbf{C}_c = \mathbf{U} \mathbf{\Sigma} \mathbf{U}^H$ using a quaternion unitary matrix \mathbf{U} and a diagonal matrix $\mathbf{\Sigma}$. Thus, \mathbf{C}_c can be whitened using the whitening matrix \mathbf{D} as

$$\begin{aligned} \mathbf{D} &= \mathbf{\Sigma}^{-\frac{1}{2}} \mathbf{U}^H \\ \mathbf{I} &= \mathbf{D} \mathbf{C}_c \mathbf{D}^H = \mathbf{D} \bar{\mathbf{C}}_a \mathbf{D}^H + \mathbf{D} \bar{\mathbf{C}}_b \mathbf{D}^H \end{aligned} \quad (5)$$

in which $\mathbf{G}_a = \mathbf{D} \bar{\mathbf{C}}_a \mathbf{D}^H$ and $\mathbf{G}_b = \mathbf{D} \bar{\mathbf{C}}_b \mathbf{D}^H$ share common eigenvectors \mathbf{F}

$$\mathbf{F}^H \mathbf{G}_a \mathbf{F} = \mathbf{\Sigma}_a \quad \text{and} \quad \mathbf{F}^H \mathbf{G}_b \mathbf{F} = \mathbf{\Sigma}_b \quad (\mathbf{\Sigma}_a + \mathbf{\Sigma}_b = \mathbf{I}). \quad (6)$$

Remark 1: The eigenvalues of $\mathbf{\Sigma}_a$ are sorted in a descending order and eigenvalues of $\mathbf{\Sigma}_b$ in an ascending order. Thus, the spatial filter $\mathbf{W} = \mathbf{F}^H \mathbf{D}$ discriminates between two classes by maximizing (1).

For a given quaternion-valued test data \mathbf{x} , the spatial filtered data is obtained as $\mathbf{y} = \mathbf{W} \mathbf{x}$ in which the first and last row vectors³ correspond to the largest eigenvalues in $\mathbf{\Sigma}_a$ and $\mathbf{\Sigma}_b$, respectively. Thus, the variances of the real and three-imaginary parts of these vectors are used as features

$$\begin{aligned} f_p^r &= \log \frac{\text{var}(\Re[\mathbf{y}_p])}{\text{var}(\Re[\mathbf{y}_1]) + \text{var}(\Re[\mathbf{y}_N])} \quad p \in \{1, N\} \\ f_p^i &= \log \frac{\text{var}(\Im_i[\mathbf{y}_p])}{\text{var}(\Im_i[\mathbf{y}_1]) + \text{var}(\Im_i[\mathbf{y}_N])} \\ f_p^j &= \log \frac{\text{var}(\Im_j[\mathbf{y}_p])}{\text{var}(\Im_j[\mathbf{y}_1]) + \text{var}(\Im_j[\mathbf{y}_N])} \end{aligned}$$

²Throughout this paper, N defines the channel number, T is the sample length, $E[\cdot]$ represents the expectation operator and $(\cdot)^H$ is Hermitian (i.e., conjugate transpose).

³In general, row vectors with indexes $p = 1, \dots, m$ and $p = N - m + 1, \dots, N$ are considered in CSP algorithms. However in this work, only the first and last row vectors were selected, i.e., $p \in \{1, N\}$. Note that as vectors are quaternion-valued, selecting a single pair of filter for QCSP is equivalent to selecting four pairs in the real domain—which follows the principle used in the literature [2,4,6].

$$f_p^\kappa = \log \frac{\text{var}(\Im_\kappa[\mathbf{y}_p])}{\text{var}(\Im_\kappa[\mathbf{y}_1]) + \text{var}(\Im_\kappa[\mathbf{y}_N])} \quad (7)$$

Observe that a conventional QCSP algorithm does not exploit the power difference or the correlation between the data channels. Therefore, it is necessary to make use of widely linear modelling of signals using augmented statistics—which is next discussed.

A. Quaternion Statistics

Recently, augmented statistics have been established to provide the complete second-order information and exploit the power difference or the correlation between the data channels [12]. In complex domain, the complete second-order information is obtained by considering both the standard covariance matrix $E[\mathbf{z}\mathbf{z}^H]$ and the pseudo-covariance matrix $E[\mathbf{z}\mathbf{z}^T]$. However, such convenient manipulation is not possible in quaternion domain.

The correspondence between the elements of a quaternion vector $\mathbf{q} = \mathbf{q}_1 + \iota \mathbf{q}_2 + j \mathbf{q}_3 + \kappa \mathbf{q}_4 \in \mathbb{H}$ and the elements of a quadrivariate vector in \mathbb{R}^4 can be obtained as

$$\Re[\mathbf{q}] = \frac{1}{2}(\mathbf{q} + \mathbf{q}^*), \quad \Im_\eta[\mathbf{q}] = \frac{1}{2\eta}(\mathbf{q} - (\mathbf{q}^\eta)^*) \quad \eta \in \{\iota, j, \kappa\} \quad (8)$$

where $(\cdot)^*$ is the conjugate operator, and $(\cdot)^\eta$ is the involution operator given by

$$\begin{aligned} \mathbf{q}^\eta &= -\eta \mathbf{q} \eta, \quad \eta \in \{\iota, j, \kappa\} \\ \text{e.g. } \mathbf{q}^\iota &= -\iota \mathbf{q} \iota = \mathbf{q}_1 + \iota \mathbf{q}_2 - j \mathbf{q}_3 - \kappa \mathbf{q}_4. \end{aligned} \quad (9)$$

According to (8) and (9), quaternion statistics should include all quaternion involutions \mathbf{q}^ι , \mathbf{q}^j , and \mathbf{q}^κ to access the complete second-order statistical information. In other words, second-order quaternion statistics should consider the standard covariance matrix $\mathbf{C}_\mathbf{q} = E[\mathbf{q}\mathbf{q}^H]$ and the complementary covariance matrices $\mathbf{C}_{\mathbf{q}^\iota}$, $\mathbf{C}_{\mathbf{q}^j}$, and $\mathbf{C}_{\mathbf{q}^\kappa}$ which are obtained as

$$\begin{aligned} \mathbf{C}_{\mathbf{q}^\iota} &= E[\mathbf{q}\mathbf{q}^{\iota H}] \quad \mathbf{C}_{\mathbf{q}^j} = E[\mathbf{q}\mathbf{q}^{j H}] \\ \mathbf{C}_{\mathbf{q}^\kappa} &= E[\mathbf{q}\mathbf{q}^{\kappa H}]. \end{aligned} \quad (10)$$

To this end, we propose an augmented quaternion CSP (A-QCSP) and a generalized quaternion CSP (G-QCSP) to incorporate the complementary covariance matrices.

III. AUGMENTED QUATERNION COMMON SPATIAL PATTERNS

The procedure of the A-QCSP algorithm is similar to the conventional QCSP. However, the A-QCSP algorithm is derived based on the augmented version of the data given by

$$\dot{\mathbf{x}} = [\mathbf{x}^T, \mathbf{x}^{\iota T}, \mathbf{x}^{j T}, \mathbf{x}^{\kappa T}]^T. \quad (11)$$

In other words, A-QCSP takes advantage of the augmented covariance matrix $\dot{\mathbf{C}}_d$ which is defined as

$$\begin{aligned} \dot{\mathbf{C}} &= E[\dot{\mathbf{x}}\dot{\mathbf{x}}^H] \\ &= \begin{bmatrix} \mathbf{C} & \mathbf{C}_{\mathbf{x}^\iota} & \mathbf{C}_{\mathbf{x}^j} & \mathbf{C}_{\mathbf{x}^\kappa} \\ \mathbf{C}_{\mathbf{x}^\iota}^i & \mathbf{C}^\iota & \mathbf{C}_{\mathbf{x}^\kappa}^i & \mathbf{C}_{\mathbf{x}^j}^i \\ \mathbf{C}_{\mathbf{x}^j}^j & \mathbf{C}_{\mathbf{x}^\kappa}^j & \mathbf{C}^j & \mathbf{C}_{\mathbf{x}^\iota}^j \\ \mathbf{C}_{\mathbf{x}^\kappa}^\kappa & \mathbf{C}_{\mathbf{x}^j}^\kappa & \mathbf{C}_{\mathbf{x}^\iota}^\kappa & \mathbf{C}^\kappa \end{bmatrix} \end{aligned} \quad (12)$$

Algorithm 1 A-QCSP Algorithm

- 1) Calculate the augmented version of data channels for both classes as $\dot{\mathbf{x}}_d = [\mathbf{x}_d^T, \mathbf{x}_d^{i^T}, \mathbf{x}_d^{j^T}, \mathbf{x}_d^{\kappa^T}]^T$, $d \in \{a, b\}$.
- 2) Generate the augmented covariance matrices as $\dot{\mathbf{C}}_d = E[\dot{\mathbf{x}}_d \dot{\mathbf{x}}_d^H] \in \mathbb{H}^{4N \times 4N}$, $d \in \{a, b\}$.
- 3) Define the composite augmented covariance matrix $\dot{\mathbf{C}}_c = \dot{\mathbf{C}}_a + \dot{\mathbf{C}}_b$.
- 4) Apply (5) and (6) on the composite augmented covariance matrix $\dot{\mathbf{C}}_c$ to obtain the corresponding $\dot{\mathbf{D}}$ and $\dot{\mathbf{F}}$.
- 5) Derive the augmented spatial filter $\dot{\mathbf{W}}_c = \dot{\mathbf{F}}^H \dot{\mathbf{D}}$.

where $\mathbf{C} = E[\mathbf{xx}^H]$ is the covariance matrix, and $\mathbf{C}_{x^\eta} = E[\mathbf{xx}^{\eta H}]$ is the η -covariance matrix $\eta \in \{i, j, \kappa\}$. For instance, $\mathbf{C}_{x^i} = E[\mathbf{xx}^{iH}]$ can be expanded as

$$\mathbf{C}_{x^i} = E[\mathbf{xx}^{iH}] = \begin{bmatrix} x_1 x_1^{i*} & x_1 x_2^{i*} & \cdots & x_1 x_N^{i*} \\ x_2 x_1^{i*} & x_2 x_2^{i*} & \cdots & x_2 x_N^{i*} \\ \vdots & \vdots & \ddots & \vdots \\ x_N x_1^{i*} & x_N x_2^{i*} & \cdots & x_N x_N^{i*} \end{bmatrix} \quad (13)$$

in which the diagonal elements are $(\cdot)^{i*}$ invariant, see Appendix A for further details.

Since $\dot{\mathbf{C}}$ includes the standard covariance and all the complementary covariance matrices, it provides the complete second-order information for the generality of quaternion-valued data.

The A-QCSP method is summarized in Algorithm 1. For a given test data \mathbf{x} , the augmented spatial filtered data is obtained as $\dot{\mathbf{y}} = \dot{\mathbf{W}}\dot{\mathbf{x}}$.

Remark 2: The performance of the A-QCSP algorithm is similar to that of real-valued CSP, subject to the scaling factor of $\frac{1}{4}$ [7]. This is due to diagonalisation of the real-valued covariance matrix in the whitening stage of A-QCSP.

Proof: For a zero-mean process $\mathbf{q} = \mathbf{q}_1 + i\mathbf{q}_2 + j\mathbf{q}_3 + \kappa\mathbf{q}_4$, the augmented version [7] is given by

$$\dot{\mathbf{q}} = [\mathbf{q}^T, \mathbf{q}^{i^T}, \mathbf{q}^{j^T}, \mathbf{q}^{\kappa^T}]^T = \Theta \mathbf{r} \quad (14)$$

where

$$\Theta = \begin{bmatrix} \mathbf{I} & i\mathbf{I} & j\mathbf{I} & \kappa\mathbf{I} \\ \mathbf{I} & i\mathbf{I} & -j\mathbf{I} & -\kappa\mathbf{I} \\ \mathbf{I} & -i\mathbf{I} & j\mathbf{I} & -\kappa\mathbf{I} \\ \mathbf{I} & -i\mathbf{I} & -j\mathbf{I} & \kappa\mathbf{I} \end{bmatrix} \quad \text{and} \quad \mathbf{r} = \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \\ \mathbf{q}_3 \\ \mathbf{q}_4 \end{bmatrix}. \quad (15)$$

Thus, the augmented covariance matrix is defined as

$$E[\dot{\mathbf{q}}\dot{\mathbf{q}}^H] = E[\Theta \mathbf{r} \mathbf{r}^H \Theta^H] = E[\Theta \mathbf{r} \mathbf{r}^T \Theta^H]. \quad (16)$$

Considering $\Theta^H \Theta = \Theta \Theta^H = 4\mathbf{I}$, the whitening transformation of augmented covariance matrix leads to

$$E[\dot{\mathbf{q}}\dot{\mathbf{q}}^H] = \mathbf{I} = \Theta E[\mathbf{r} \mathbf{r}^T] \Theta^H \quad (17)$$

$$\mathbf{I} = 4E[\mathbf{r} \mathbf{r}^T]$$

which justifies the duality between the real-valued CSP and A-QCSP with the scaling factor of $\frac{1}{4}$. ■

Note that A-QCSP algorithm does not provide information on the power difference of the data channels. To resolve this issue, we address a generalized quaternion CSP (G-QCSP) in the next section.

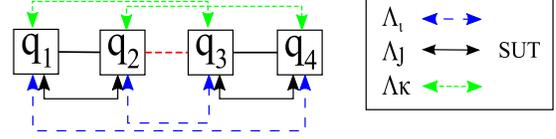


Fig. 1. Note that SUT models power difference between the real and imaginary parts of complex-valued channels, while the real-valued eigenvalues of i -, j -, and κ -covariance matrices hold power differences of all possible pairs of data channels.

IV. GENERALIZED QUATERNION COMMON SPATIAL PATTERNS

Consider four real-valued processes \mathbf{q}_1 , \mathbf{q}_2 , \mathbf{q}_3 and \mathbf{q}_4 which form two complex-valued channels \mathbf{z}_1 and \mathbf{z}_2 , and a quaternion-valued channel \mathbf{q} as

$$\begin{aligned} \mathbf{z}_1 &= \mathbf{q}_1 + i\mathbf{q}_2 & \mathbf{z}_2 &= \mathbf{q}_3 + i\mathbf{q}_4 \\ \mathbf{q} &= \mathbf{q}_1 + i\mathbf{q}_2 + j\mathbf{q}_3 + \kappa\mathbf{q}_4. \end{aligned} \quad (18)$$

According to [6], the SUTCCSP algorithm incorporates the power difference between the real and imaginary parts of the complex-valued channels. In other words, the SUTCCSP calculates the power difference for two specific pairs $\{\mathbf{q}_1, \mathbf{q}_2\}$ and $\{\mathbf{q}_3, \mathbf{q}_4\}$, as shown in Fig. 1. However, in this work, we propose a G-QCSP algorithm in which power differences of all possible pairs are considered using the i -, j -, and κ -complementary covariance matrices—as discussed below.

A. Analysis of Complementary Covariance Matrices

For the given quaternion-valued vector \mathbf{q} , the complementary covariance matrices \mathbf{C}_{q^η} are calculated⁴ via (10)

$$\begin{aligned} \mathbf{C}_{q^i} &= \Re[\mathbf{C}_{q^i}] + \Im[\mathbf{C}_{q^i}] \\ &= (\mathbf{C}_{q_1} + \mathbf{C}_{q_2} - \mathbf{C}_{q_3} - \mathbf{C}_{q_4}) + \Im[\mathbf{C}_{q^i}] \\ \mathbf{C}_{q^j} &= \Re[\mathbf{C}_{q^j}] + \Im[\mathbf{C}_{q^j}] \\ &= (\mathbf{C}_{q_1} - \mathbf{C}_{q_2} + \mathbf{C}_{q_3} - \mathbf{C}_{q_4}) + \Im[\mathbf{C}_{q^j}] \\ \mathbf{C}_{q^\kappa} &= \Re[\mathbf{C}_{q^\kappa}] + \Im[\mathbf{C}_{q^\kappa}] \\ &= (\mathbf{C}_{q_1} - \mathbf{C}_{q_2} - \mathbf{C}_{q_3} + \mathbf{C}_{q_4}) + \Im[\mathbf{C}_{q^\kappa}]. \end{aligned} \quad (19)$$

Observe that \mathbf{C}_{q^η} is η -Hermitian ($\mathbf{C}_{q^\eta} = \mathbf{C}_{q^\eta}^{\eta H}$, $\eta \in \{i, j, \kappa\}$) and it can be factorized using the quaternion Takagi (QTakagi) factorization established in [13]. Hence, each η -covariance matrix is rewritten as $\mathbf{C}_{q^\eta} = \mathbf{Q}_\eta \Lambda_\eta \mathbf{Q}_\eta^{\eta H}$, where \mathbf{Q}_η is a quaternion unitary matrix and Λ_η denotes a real-valued diagonal matrix of eigenvalues. Since the obtained eigenvalues are real-valued, they hold information on the power differences of various pairs, in other words

- $\Lambda_i = \mathbf{C}_{q_1} + \mathbf{C}_{q_2} - \mathbf{C}_{q_3} - \mathbf{C}_{q_4}$ is the power difference for pairs $\{\mathbf{q}_1, \mathbf{q}_4\}$ and $\{\mathbf{q}_2, \mathbf{q}_3\}$,
- $\Lambda_j = \mathbf{C}_{q_1} - \mathbf{C}_{q_2} + \mathbf{C}_{q_3} - \mathbf{C}_{q_4}$ determines the power difference for pairs $\{\mathbf{q}_1, \mathbf{q}_2\}$ and $\{\mathbf{q}_3, \mathbf{q}_4\}$ - same information provided by SUTCCSP,
- $\Lambda_\kappa = \mathbf{C}_{q_1} - \mathbf{C}_{q_2} - \mathbf{C}_{q_3} + \mathbf{C}_{q_4}$ is the power difference for pairs $\{\mathbf{q}_1, \mathbf{q}_3\}$ and $\{\mathbf{q}_4, \mathbf{q}_2\}$, see Fig. 1.

⁴In this work, we use the “real-valued” eigenvalues of the complementary covariance matrices. This is why we focus on real-valued components of these matrices. Further details can be found in [7].

Algorithm 2 G-QCSP Algorithm⁵

- 1) Generate the covariance matrices for classes a and b
 $\mathbf{C}_d = E[\mathbf{x}_d \mathbf{x}_d^H]$, $d \in \{a, b\}$
- 2) Define the composite spatial covariance \mathbf{C}_c using the averaged covariance matrices $\mathbf{C}_c = \bar{\mathbf{C}}_a + \bar{\mathbf{C}}_b$
- 3) Decompose the composite spatial covariance matrix and derive the whitening matrix \mathbf{D}
 $\mathbf{C}_c = \mathbf{U} \mathbf{\Sigma} \mathbf{U}^H$ $\mathbf{D} = \mathbf{\Sigma}^{-\frac{1}{2}} \mathbf{U}^H$
- 4) Factorize $\mathbf{G}_a = \mathbf{D} \bar{\mathbf{C}}_a \mathbf{D}^H$ and $\mathbf{G}_b = \mathbf{D} \bar{\mathbf{C}}_b \mathbf{D}^H$ to obtain the common eigenvector \mathbf{F}
- 5) Define the whitened data \mathbf{v}_d and generate the η -covariances of the whitened data, $\mathbf{v}_d = \mathbf{D} \mathbf{x}_d$
 $\mathbf{C}_{d\eta} = E[\mathbf{v}_d \mathbf{v}_d^{\eta H}]$, $d \in \{a, b\}$ and $\eta \in \{t, j, \kappa\}$
- 6) Calculate the composite η -covariance matrices as $\mathbf{C}_{c\eta} = \bar{\mathbf{C}}_{a\eta} + \bar{\mathbf{C}}_{b\eta}$
- 7) Apply QTakagi factorization [13] on the η -covariance matrices such that $\mathbf{C}_{c\eta} = \mathbf{Q}_\eta \mathbf{\Lambda}_\eta \mathbf{Q}_\eta^H$
- 8) Derive the standard covariance spatial filter \mathbf{W}_c and η -covariance spatial filters \mathbf{W}_η . $\mathbf{W}_c = \mathbf{F}^H \mathbf{D}$ and $\mathbf{W}_\eta = \mathbf{\Lambda}_\eta^{-\frac{1}{2}} \mathbf{Q}_\eta^H$, $\eta \in \{t, j, \kappa\}$.

B. The G-QCSP Algorithm

Given two zero-mean quaternion-valued processes \mathbf{x}_a and \mathbf{x}_b , the G-QCSP procedure is fully described in Algorithm 2.

Remark 3: The spatial filters generated via the G-QCSP method satisfy the following conditions:

$$\mathbf{W}_c \mathbf{C}_c \mathbf{W}_c^H = \mathbf{I} \quad \text{and} \quad \mathbf{W}_\eta \mathbf{C}_{c\eta} \mathbf{W}_\eta^H = \mathbf{I}, \quad \eta \in \{t, j, \kappa\} \quad (20)$$

Remark 4: For a given test data \mathbf{x} , the filtered data sets via G-QCSP are calculated as $\mathbf{y} = \mathbf{W}_c \mathbf{x}$ and $\mathbf{y}_\eta = \mathbf{W}_\eta \mathbf{x}$, where $\eta \in \{t, j, \kappa\}$. Thus, there are total number of 16 features calculated using (7).

Observe that compared to the SUTCCSP algorithm, G-QCSP holds more rigorous information on the power difference and it is expected to provide higher reliability.

V. SIMULATIONS

Performance of the A-QCSP and G-QCSP methods were evaluated using both synthetic and real-world EEG data. To highlight the advantages of proposed methods, they were compared with real-valued CSP and SUTCCSP algorithms.

In all experiments, classification was performed using a support vector machine (SVM) with a Gaussian kernel. Performance of the automated classification was assessed using accuracy (α), sensitivity (δ), and Cohen's kappa (τ) measured from the confusion matrix of manual labels versus automatic classification, see Table I. Accuracy is the percentage of all correctly labelled tasks by automated algorithms, while sensitivity is the ratio of correctly identified tasks for an individual class. For instance using Table I, accuracy and sensitivities are calculated as

$$\alpha = \frac{p_a + p_b}{p_a + f_a + p_b + f_b}, \quad \delta_a = \frac{p_a}{p_a + f_a} \quad \text{and} \quad \delta_b = \frac{p_b}{p_b + f_b} \quad (21)$$

⁵The commented MATLAB code for G-QCSP method is available at http://www.surrey.ac.uk/cs/research/nice/people/clive_cheong_tool.

TABLE I
CONFUSION MATRIX FOR TWO CLASSES a AND b

	Automated	a	b
Manual		p_a	f_a
		f_b	p_b

TABLE II
QUALITATIVE INTERPRETATION OF τ

Agreement	Kappa value	Agreement	Kappa value
Almost Perfect	> 0.8	Fair	0.21-0.40
Substantial	0.61-0.8	Slight	0.00-0.20
Moderate	0.41-0.60	Poor	< 0

Remark 5: For same number of trials in classes a and b , accuracy equals to the mean sensitivity⁶ providing $\alpha = \frac{\delta_a + \delta_b}{2}$. This is why in Sections V-A and V-B, we only analyze the averaged sensitivity.

Unlike α and δ , kappa coefficient considers the probability of agreement due to chance alone and it is more robust than simple percentage agreement. Cohen's kappa quantifies the inter-rater agreement and its interpretation is summarized in Table II.

Remark 6: In both synthetic and real-world simulations, the final α , δ and τ values were obtained by averaging the outcomes of five-fold cross-validation⁷ procedure.

A. Synthetic Data

The first experiment evaluated the classification accuracy of the A-QCSP and G-QCSP methods under different levels of additive white Gaussian noise. Following the example of Koles and Soong [14], the sum of sinusoids were used to generate the synthetic EEG data to validate their source localization. For this purpose, two sources \mathbf{s}_a and \mathbf{s}_b were initiated by summing 10 sinusoids with the frequency ranges⁸ of 8–30 Hz and sample length of $N = 1000$. Then, sources were used to compose two classes \mathbf{x}_a and \mathbf{x}_b as

$$\mathbf{x}_d = \mathbf{M}_d \mathbf{s}_d + \mathbf{e}_d, \quad d \in \{a, b\} \quad (22)$$

in which $\mathbf{x}_d \in \mathbb{H}^{2 \times N}$ defines the two-dimensional quaternion-valued input of class d and \mathbf{e} is the white Gaussian noise. To accumulate the averaged results, 50 sets of data were generated and each set involved 40 trials. In order to assess the effect of noise and correlation on the classification accuracy white Gaussian noise was added with different levels of SNR varying from -10 dB to 10 dB, and the correlation between real and imaginary parts of the noise was adjusted from 0.1 to 0.9. Furthermore to set the specific spatial pattern for each

⁶Positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) studies have demonstrated that motor-imagery (MI) tasks activate several brain regions which are spatially distributed.

⁷Each dataset was divided into 80% training and 20% testing segments, and the classification was repeated five times in various sample orders.

⁸Since the basis for MI tasks are the mu (8–12 Hz), beta (14–20 Hz), and higher beta (above 20 Hz) rhythms [15], [16], our synthetic simulations were performed in the frequency range of 8–30 Hz.

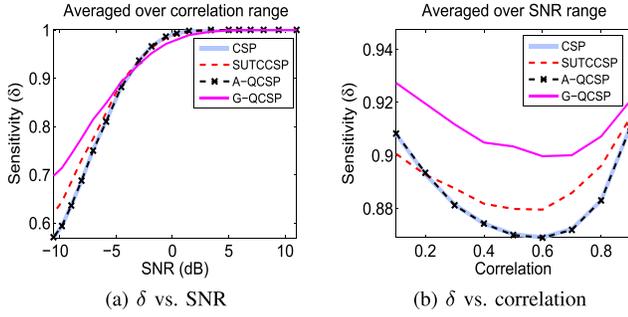


Fig. 2. The sensitivity averaged over 50 sets of data. To clarify the effect of additive noise and correlation, (a) δ was averaged over the correlation range, and (b) δ was averaged over the SNR range.

class, mixing matrices \mathbf{M}_a and \mathbf{M}_b were defined as⁹

$$\mathbf{M}_a = \begin{bmatrix} 1 & 1.05\iota & 1.10j & 1.15\kappa \\ 1.11 & 1.15\iota & 1.19j & 1.23\kappa \end{bmatrix}$$

$$\mathbf{M}_b = \begin{bmatrix} 1.18 & 1.17\iota & 1.16j & 1.15\kappa \\ 1.02 & 1.04\iota & 1.06j & 1.08\kappa \end{bmatrix} \quad (23)$$

so that 1) second channel of class a had larger amplitude than the first channel (vice versa for class b), and 2) class a had higher power differences than class b (0.05 and 0.04 for class a and 0.01 and 0.02 for class b).

Note that inputs for the real-valued CSP (\mathbf{r}_d) and SUTCCSP (\mathbf{z}_d) algorithms were given by

$$\mathbf{r}_d = [\Re\{\mathbf{x}_d\}, \Im\{\mathbf{x}_d\}, \Im\{j\mathbf{x}_d\}, \Im\{\kappa\mathbf{x}_d\}]^T \in \mathbb{R}^{8 \times N}$$

$$\mathbf{z}_d = [\Re\{\mathbf{x}_d\} + \iota\Im\{\mathbf{x}_d\}, \Im\{j\mathbf{x}_d\} + \iota\Im\{\kappa\mathbf{x}_d\}]^T \in \mathbb{C}^{4 \times N}. \quad (24)$$

To perform the classification, all algorithms were applied on the generated datasets to extract the features using¹⁰ (7). These features were then fed to the SVM classifier and δ and τ values were calculated for each set. Fig. 2 illustrates the δ value averaged over 50 sets of data. To distinguish the effect of additive noise and correlation, the δ was first averaged over the correlation range, and then it was averaged over the SNR range, see Fig. 2(a) and (b).

Remark 7: Note that the CSP and A-QCSP algorithms showed similar performances—as expected from Remark 2.

Remark 8: The G-QCSP and SUTCCSP algorithms hold additional information on the power difference and they mostly provided higher sensitivities compared to CSP and A-QCSP.

Remark 9: Since the G-QCSP structure involved all the possible power differences of a quaternion-valued channel, this method mainly outperformed the SUTCCSP and provided the highest sensitivity.

Remark 10: As shown, for higher correlations results of the G-QCSP and SUTCCSP algorithms were closer to the ones of CSP and A-QCSP. This is due to the decrease of power difference for higher correlation values as shown in Fig. 3.

⁹To confirm the robustness of the proposed method, further experiments were performed to consider negative values in the mixing matrices, see Appendix B.

¹⁰Note that (7) was proposed for quaternion-valued CSP algorithms. However, same concept was used to obtain features for real-valued and complex-valued CSP.

TABLE III
THE AVERAGED τ AND δ VALUES

Methods	CSP	SUTCCSP	A-QCSP	G-QCSP
Cohen's kappa (τ)	0.76	0.78	0.76	0.82
Sensitivity (δ)	0.87	0.89	0.87	0.91

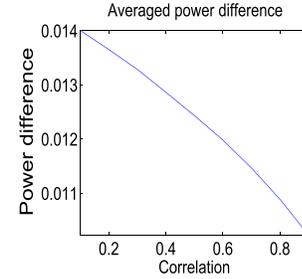


Fig. 3. The averaged power difference versus correlation. To calculate the mean power difference, it was first averaged over the SNR range for each dataset and then it was averaged for all 50 sets of data.

Overall, the outputs of 50 sets were averaged over both SNR and correlation to measure the final δ and τ values obtained from each method, see Table III.

Remark 11: Although CSP, SUTCCSP, and A-QCSP methods provided substantial agreements ($\tau > 0.61$), our proposed algorithm G-QCSP outperformed the others by offering almost perfect classification ($\tau = 0.82$) and the highest sensitivity of $\delta = 0.91$. Observe the similar values for the CSP and A-QCSP algorithms.

B. Motor Imagery EEG Data

The second experiment involved G-QCSP with application to MI EEG data. To this end, the Physiobank motor/mental imagery (MMI) database¹¹ was used to evaluate our proposed G-QCSP algorithm [18]. This database involved 109 subjects performing both motor and imagery tasks while EEG was recorded using 64-channels according to the standard 10–10 system.

In this work 48 channels were selected, and the EEG data was band-pass filtered in the range of 8–30 Hz using a FIR filter. Note that positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) studies have demonstrated that MI tasks activate several brain regions which are spatially distributed. According to [3], some of these regions include: supplementary motor area (SMA), pre-motor area (PMA), superior and inferior parietal lobule, primary motor cortex (M1), primary and secondary somatosensory cortices (S1), and cerebellum (CRB). Thus, specific electrodes were chosen to cover most scalp regions in these areas.

The spatially adjacent electrodes were paired to form complex-valued channels, and two complex-valued channels were then merged to generate a quaternion-valued signal, as shown in Fig. 4. Pairing spatially adjacent electrodes naturally facilitated the use of cross-information to model the inherent correlation [6]. Furthermore, the power difference of the

¹¹The Physiobank MMI database is recorded using the BCI2000 system [17] and is available through <http://www.physionet.org/pn4/eegmmidb>.

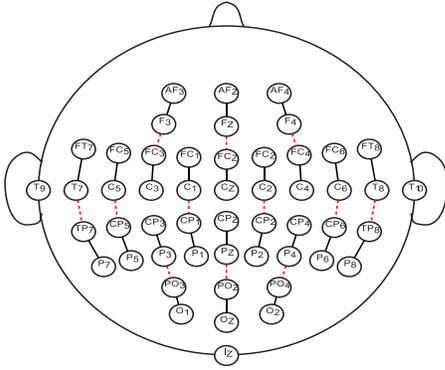


Fig. 4. EEG Channel structure for CCSP and QCSP: 1) pairs of electrodes connected via a solid line were used as complex-valued EEG data, and 2) pairs of complex channels merged via a dash line were used as quaternion-valued EEG channels.

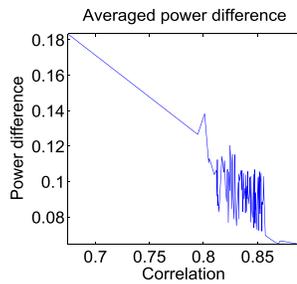


Fig. 5. Real-world EEG power difference versus correlation for 109 subjects.

TABLE IV
THE FINAL δ VALUE AVERAGED OOVER 72 SUBJECTS

Methods	CSP	SUTCCSP	A-QCS	G-QCSP
Sensitivity (δ)	0.63	0.67	0.64	0.70

combined electrodes was used as an additional information to classify the left and right hand movements using only imagery trials. Fig. 5 shows the averaged power difference versus correlation for 109 subjects. Observe that power difference was smaller for higher correlation values, similar to the trend of synthetic data shown in Fig. 3. Hence, synthetic data is valid and realistic as in real-world EEG.

The δ and τ values were calculated for each subject by averaging the results of five-fold cross-validation procedure. Among the whole database, subjects for whom the classification was slight or poor ($\tau < 0.21$) using all four methods were discarded. Fig. 6 illustrates the sensitivity of the remaining 72 subjects using whisker diagram in which the red lines define median. Note that on average G-QCSP and SUTCCSP which included the power difference provided higher performance, and as expected G-QCSP outperformed the SUTCCSP. Same observation was perceived by comparing the final δ values averaged over 72 subjects, see Table IV. In addition, Fig. 7 displays number of subjects whose best classification was obtained using one of the CSP, SUTCCSP, A-QCSP, and G-QCSP methods. As shown, G-QCSP leads to 35 subjects which is more than twice the number of subjects obtained via SUTCCSP method.

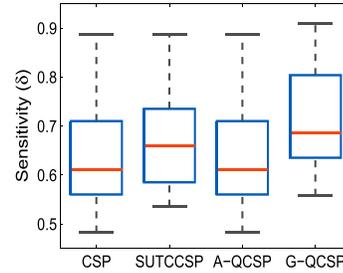


Fig. 6. Whisker diagram of sensitivity values obtained in the classification of right and left hand movement using imagery trials. Note that subjects who had slight or poor agreement ($\tau < 0.21$) for all methods CSP, SUTCCAP, A-QCSP, and G-QCSP, were discarded and this diagram includes the remaining 72 subjects.

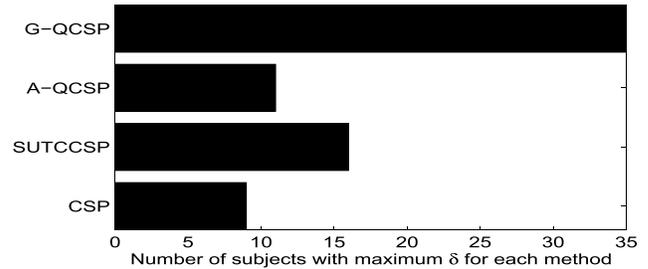


Fig. 7. Number of subjects whose best classification rates were obtained using CSP, SUTCCSP, A-QCSP, and G-QCSP.

TABLE V
THE AVERAGED POWER DIFFERENCE FOR SUBJECTS WHO HAD THE HIGHEST SENSITIVITY USING SUTCCSP AND G-QCSP

Methods	Power difference \pm std	
	SUTCCSP	G-QCSP
Λ_i	0.09 \pm 0.05	0.12 \pm 0.05
Λ_j	0.06 \pm 0.04	0.07 \pm 0.03
Λ_κ	0.06 \pm 0.02	0.08 \pm 0.03

To highlight the impact of power differences, i -, j -, and κ -differences were averaged for subjects who had the best performance using SUTCCSP and G-QCSP, see Table V. As shown, the averaged power differences were higher for the G-QCSP algorithm.

Furthermore, as an illustrative example, the proposed G-QCSP was used to compare the topology distributions of the left and right imagery hand movements. To this end, the standard covariance spatial filter W_c was considered to exploit the spatial patterns of a *single trial* EEG for each class, see Fig. 8. Observe that these topologies confirm the contra-lateral distribution during the MI tasks of the right and left hand movements—which is in agreement with the literature [2], [3]. Given that each topology was obtained using a *single trial EEG* for the frequency range of 8–30 Hz, the proposed G-QCSP provided good discrimination between the left and right hand movements.

C. Gait Analysis for Parkinson’s Disease

Besides the BCI application, the G-QCSP algorithm was applied on gait data to classify idiopathic Parkinson’s disease (IPD) patients and healthy controls (HC). In this work, we

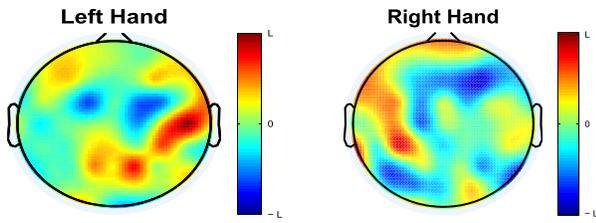


Fig. 8. Topology distributions of the left and right imagery hand movements. Note that topologies were obtained for a single trial EEG within the frequency range of 8–30 Hz. To provide a statistical analysis, the percentages of highly active regions were determined (details of the calculations are included in Appendix C). These percentages were found to be 20% and 17% for the left and right subplots. The trivial difference of 3% between these two values shows the consistency of the highly active regions in the human brain for both left and right hand movements.

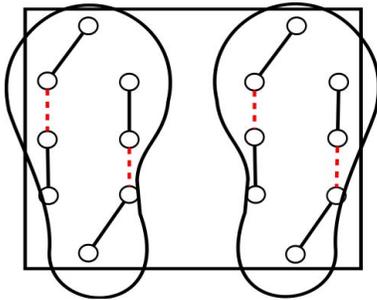


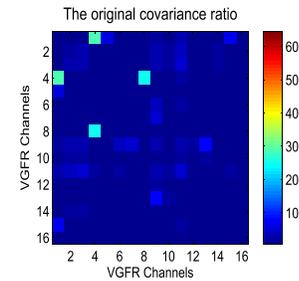
Fig. 9. Position of the pressure sensors. Note that pairs of sensors connected via a solid line were used as complex-valued channels, and two complex channels merged via a dash line were used as quaternion-valued channels.

used the Physiobank gate dataset¹² including 93 IPD patients (mean age: 66.3 years; 63% men), and 73 healthy controls (mean age: 66.3 years; 55% men) [18]. This database provided the vertical ground reaction force (VGRF) recorded at a rate of 100 Hz using eight pressure-sensitive sensors underneath each foot, as shown in Fig. 9. All subjects were asked to walk for 2 min on level ground at their self-selected comfortable pace.

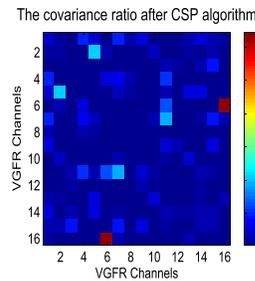
To perform the classification of gait data, healthy controls and IPD patients were considered as classes \mathbf{x}_a and \mathbf{x}_b , and a window of 3000 samples (half a minute) was considered as a single trial, i.e., 2 min data generated four trials. In order to use all CSP-based algorithms, the VGRF recording of each class was represented by four quaternion-valued, eight complex-valued or sixteen real-valued channels. The corresponding CSP-based algorithm was then applied to increase the covariance ratio of two classes by maximizing (1). Fig. 10 represents the covariance ratio before and after applying the CSP and G-QCSP algorithms. As shown, the covariance ratio of the raw data had uniform small values, while CSP and G-QCSP filtered data showed larger covariance ratio by the presence of numerous non-blue pixels. Note that compared to the conventional CSP, our proposed method provided higher covariance ratio and better discrimination, compare Fig. 10(b) and (c).

Furthermore, Fig. 11 illustrates how the CSP filtering works in 2-D, as expected from literature [2]. Note that CSP maps the raw samples of Fig. 11(a) to those in Fig. 11(b) where

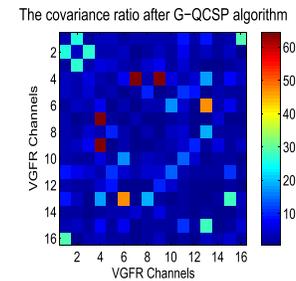
¹²The Physiobank gate database is available through <http://physionet.org/physiobank/database/gaitpdb>.



(a) Covariance ratio of the original data ($\bar{C}_{\mathbf{x}_a} / \bar{C}_{\mathbf{x}_b}$)

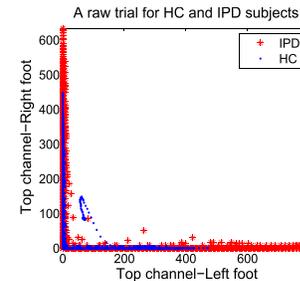


(b) Covariance ratio of CSP filtered data ($\bar{C}_{\mathbf{y}_a} / \bar{C}_{\mathbf{y}_b}$)

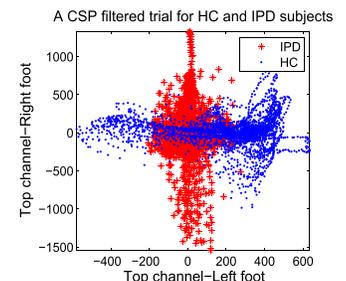


(c) Covariance ratio of G-QCSP filtered data ($\bar{C}_{\mathbf{y}_a} / \bar{C}_{\mathbf{y}_b}$)

Fig. 10. Effect of the CSP-based algorithms on the covariance ratio of two classes: (a) the raw covariance ratio, and (b) the CSP filtered covariance ratio, and (c) the G-QCSP filtered covariance ratio. Observe the largest covariance ratio obtained using the G-QCSP filter.



(a) Scatter plot of a raw data



(b) Scatter plot of a CSP filtered data

Fig. 11. Illustration of CSP algorithm in 2-D space using top channels of left and right feet. In (a), scatter plot of a raw trial for a HC and an IPD is shown, whereas (b) shows scatter plot of the same subjects after CSP filtering.

the distributions are uncorrelated and horizontal axis shows the largest variance for HC subject—and vice versa for IPD patients.

To compare the real-, complex-, and quaternion- CSP algorithms, the corresponding methods extracted features using (7). These features were then imported into our SVM classifier for a five-fold cross-validation procedure. Table VI shows the confusion matrix of classification results obtained via G-QCSP. The accuracy, sensitivity, and kappa values for real-valued CSP, SUTCCSP, A-QCSP, and G-QCSP are given in Table VII. As expected SUTCCSP and G-QCSP provided higher reliable results, and G-QCSP offered the best performance among all CSP-based algorithms.

Remark 12: Note that same dataset was previously used in [19] to classify IPD patients using 40 Wavelet-based features and neural network function. That study offered $\alpha = 0.77$

TABLE VI
CONFUSION MATRIX FOR G-QCSP METHOD

Manual \ Automated	HC-trials	IPD-trials	Sensitivity
HC-trials	175	35	0.83
IPD-trials	30	255	0.89

TABLE VII
THE FINAL α , δ , AND τ VALUES FOR FIVE-FOLD CROSS VALIDATION

Methods	CSP	SUTCCSP	A-QCS	G-QCSP
Accuracy (α)	0.74	0.78	0.73	0.87
Sensitivity of HC (δ_a)	0.74	0.77	0.75	0.83
Sensitivity of IPD (δ_b)	0.73	0.79	0.70	0.89
Cohen's kappa (τ)	0.45	0.55	0.45	0.73

TABLE VIII
THE AVERAGED τ AND δ VALUES

Methods	CSP	SUTCCSP	A-QCSP	G-QCSP
Cohen's kappa (τ)	0.69	0.72	0.69	0.76
Sensitivity (δ)	0.84	0.86	0.84	0.88

and $\delta = 0.81$. However, our proposed G-QCSP algorithm provided $\alpha = 0.87$ and $\delta = 0.86$ using only 16 features for each class. The simulation studies therefore validate our approach.

VI. CONCLUSION

A quaternion version of CSP has been introduced for the first time to address four-dimensional processing of both circular and non-circular data. For rigour, it has been shown that our proposed method G-QCSP leveraged the additional information on the power difference in the analysis provided in Section IV-A. As such, our approach has provided higher classification rate compared to the real-valued CSP and SUTCCSP. The robustness of proposed method has been demonstrated over a wide range of SNR and correlation in synthetic EEG datasets. Furthermore, in terms of real-world applications, our proposed G-QCSP algorithm provided 1) $\alpha = 70\%$ for discrimination of two mental tasks, and 2) $\alpha = 87\%$ for classification of Parkinson's patients.

APPENDIX A

Based on (9), the *conjugate* involution operator conjugates only one imaginary component, i.e., for $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2\iota + \mathbf{x}_3J + \mathbf{x}_4\kappa$ we have

$$\begin{aligned} \mathbf{x}^{\iota*} &= \mathbf{x}_1 - \mathbf{x}_2\iota + \mathbf{x}_3J + \mathbf{x}_4\kappa \\ \mathbf{x}^{J*} &= \mathbf{x}_1 + \mathbf{x}_2\iota - \mathbf{x}_3J + \mathbf{x}_4\kappa \\ \mathbf{x}^{\kappa*} &= \mathbf{x}_1 + \mathbf{x}_2\iota + \mathbf{x}_3J - \mathbf{x}_4\kappa. \end{aligned} \quad (25)$$

Note that the only difference between \mathbf{x} and $\mathbf{x}^{\eta*}$, for $\eta \in \{\iota, J, \kappa\}$, is the sign of the η -imaginary component. Thus, a quaternion vector \mathbf{x} is called $(\cdot)^{\eta*}$ invariant if its η -imaginary components are zero. For example, if \mathbf{x} has a vanishing ι -imaginary part, it is $(\cdot)^{\iota*}$ invariant, i.e., $\mathbf{x}^{\iota*} = \mathbf{x}$. This property is particularly important for the structure of the ι -, J -, and

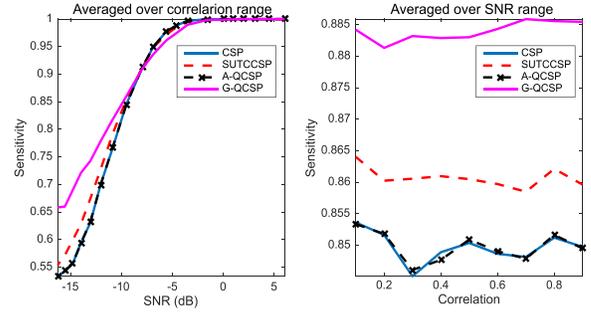


Fig. 12. The sensitivity averaged over 50 sets of data. To clarify the effect of additive noise and correlation, δ was first averaged over the correlation range (left), and then δ was averaged over the SNR range (right).

κ -complementary covariance matrices, which are given by [7]

$$\mathbf{C}_{\mathbf{x}^\eta} = E[\mathbf{x}\mathbf{x}^{\eta H}] = \begin{bmatrix} x_1x_1^{\eta*} & x_1x_2^{\eta*} & \cdots & x_1x_N^{\eta*} \\ x_2x_1^{\eta*} & x_2x_2^{\eta*} & \cdots & x_2x_N^{\eta*} \\ \vdots & \vdots & \ddots & \vdots \\ x_Nx_1^{\eta*} & x_Nx_2^{\eta*} & \cdots & x_Nx_N^{\eta*} \end{bmatrix} \quad (26)$$

where $\eta \in \{\iota, J, \kappa\}$. Observe that the diagonal entries are $(\cdot)^{\eta*}$ invariant, meaning that $(z)^{\eta*} = z$. This is due to the fact that η -imaginary components are zero, and $(\cdot)^{\eta*}$ corresponds to the conjugation of the η -imaginary component only.

APPENDIX B

The proposed G-QCSP algorithm is robust to other coefficient values considered in the mixing matrix, even for negative values. To confirm this advantage, we have performed further simulations in which mixing matrices \mathbf{M}_a and \mathbf{M}_b were defined as

$$\begin{aligned} \mathbf{M}_a &= \begin{bmatrix} -1 & -1.05\iota & -1.10J & -1.15\kappa \\ 1.11 & 1.15\iota & 1.19J & 1.23\kappa \end{bmatrix} \\ \mathbf{M}_b &= \begin{bmatrix} 1.18 & 1.17\iota & 1.16J & 1.15\kappa \\ -1.02 & -1.04\iota & -1.06J & -1.08\kappa \end{bmatrix} \end{aligned} \quad (27)$$

Note that our simulations considered negative values in the mixing matrices, while maintaining the initial spatial pattern for each class. To evaluate the robustness, the averaged sensitivity (δ) and kappa coefficient (τ) were measured following the same steps in Section V -A. The results are shown in Fig. 12. Observe that the CSP and A-QCSP algorithms showed similar performances, whereas the G-QCSP and SUTCCSP algorithms exploited additional information on the power difference and they provided higher sensitivities on average. To provide numerical assessments, the outputs of 50 sets were averaged over both SNR and correlation to measure the final δ and τ values obtained from each method, see the following table.

Note that all four algorithms provided substantial agreements ($\tau > 0.61$), and the proposed G-QCSP outperformed the others, since $\tau = 0.76$ and had the highest sensitivity $\delta = 0.88$. These results confirm that the proposed algorithm is robust regardless of the values of mixing matrices.

APPENDIX C

To provide a statistical analysis of the highly active regions in the context of the topology shown in Fig. 8, we have measured the percentage of highly active regions by calculating the following equation:

$$P = \frac{\sum_j x_j}{\sum_i x_i} \quad \text{where} \quad \begin{cases} 0 < x_i \leq L \\ 60\% \times L \leq x_j \leq L \end{cases} \quad (28)$$

in which L is the maximum value of each topoplot shown in Fig. 8, $\sum x_i$ represents the summation of regions with positive activity, and $\sum x_j$ corresponds to the accumulation of highly active regions with activity levels higher than 60%. Thus, the statistical measure P quantifies and complements the qualitative activity patterns shown in Fig. 8.

REFERENCES

- [1] S. Sanei, *Adaptive Processing of Brain Signals*. Hoboken, NJ, USA: Wiley, 2013.
- [2] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K. R. Müller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 41–56, Jan. 2008.
- [3] 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.
- [4] R. Zhang *et al.*, "Local temporal correlation common spatial patterns for single trial EEG classification during motor imagery," *Comput. Math. Methods Med.*, vol. 2013, Oct. 2013, Art. no. 591216.
- [5] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 2, pp. 355–362, Feb. 2011.
- [6] C. Park, C. C. Took, and D. P. Mandic, "Augmented complex common spatial patterns for classification of noncircular EEG from motor imagery tasks," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 1, pp. 1–10, Jan. 2014.
- [7] C. C. Took and D. P. Mandic, "Augmented second-order statistics of quaternion random signals," *Signal Process.*, vol. 91, no. 2, pp. 214–224, Feb. 2011.
- [8] S. Javidi, C. C. Took, and D. P. Mandic, "Fast independent component analysis algorithm for quaternion valued signals," *IEEE Trans. Neural Netw.*, vol. 22, no. 12, pp. 1967–1978, Dec. 2011.
- [9] S. Enshaeifar, S. Kouchaki, C. C. Took, and S. Sanei, "Quaternion singular spectrum analysis of electroencephalogram with application in sleep analysis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 1, pp. 57–67, Jan. 2016.
- [10] J. Navarro-Moreno, R. M. Fernandez-Alcala, and J. C. Ruiz-Molina, "A quaternion widely linear series expansion and its applications," *IEEE Signal Process. Lett.*, vol. 19, no. 12, pp. 868–871, Dec. 2012.
- [11] O. Falzon, K. 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.
- [12] D. P. Mandic and V. S. L. Goh, *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models*, vol. 59. Hoboken, NJ, USA: Wiley, 2009.
- [13] C. C. Took, D. P. Mandic, and F. Zhang, "On the unitary diagonalisation of a special class of quaternion matrices," *Appl. Math. Lett.*, vol. 24, no. 11, pp. 1806–1809, 2011.
- [14] Z. J. Koles and A. C. Soong, "EEG source localization: Implementing the spatio-temporal decomposition approach," *Electroencephalogr. Clin. Neurophysiol.*, vol. 107, no. 5, pp. 343–352, 1998.
- [15] G. Pfurtscheller, C. Brunner, A. Schlögl, and F. L. D. Silva, "Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks," *NeuroImage*, vol. 31, no. 1, pp. 153–159, 2006.
- [16] 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, Mar. 2000.
- [17] 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.
- [18] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
- [19] S.-H. Lee and J. S. Lim, "Parkinson's disease classification using gait characteristics and wavelet-based feature extraction," *Expert Syst. Appl.*, vol. 39, no. 8, pp. 7338–7344, 2012.

Shirin Enshaeifar (M'14) received the M.Sc. degree in medical imaging, and the Ph.D. degree in biomedical signal processing from University of Surrey, Surrey, U.K., in 2013 and 2016, respectively.

She is currently working as a Research Fellow within the Department of Electronic Engineering at University of Surrey, Surrey, U.K. Her research interests include biomedical signal processing, machine learning, and neural networks, with applications in brain-computer interface, smart healthcare, wearable technology, and internet of things (IoT).



C. Cheong Took (SM'13) received the B.S. degree in telecommunication engineering from Kings College London, London, U.K., in 2004, and read the 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 in Computational Intelligence at the University of Surrey U.K. His current research interests include machine learning, signal processing, and neural networks, with applications in telecommunication, finance, renewable energy, and biomedicine.

Dr. Took is an Associate Editor for IEEE Transactions on Neural Networks and Learning Systems. He is involved in various conference committees such as the International Joint Conference on Neural Networks 2014 (Beijing) and the International Conference on Acoustics, Speech, and Signal Processing 2019 (Brighton).



C. Park received the B.Eng. degree in electrical engineering from Sogang University, Seoul, South Korea, and the M.Sc. degree in biomedical engineering department from Seoul National University, Seoul, South Korea. In 2012, he received the Ph.D. degree in adaptive nonlinear signal processing from Imperial College London, London, U.K., and worked as a postdoctoral researcher in bioengineering department at University of California, San Diego, CA, USA.

He is an Assistant Professor of Computer Engineering Department at Kwangwoon University, Seoul, South Korea. His research interests are mainly in the area of machine learning, adaptive and statistical signal processing, with applications in brain-computer interface, computational neuroscience and wearable technology.



D. 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.

Dr. Mandic 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 ICIC09 and ISNN10, and his work on Ear-EEG has been nominated for the Annual BCI Award in 2012.