Maintaining the Integrity of Sources in Complex Learning Systems: Intraference and the Correlation Preserving Transform

Clive Cheong Took, *Senior Member, IEEE*, Scott C. Douglas, *Senior Member, IEEE*, and Danilo P. Mandic, *Fellow, IEEE*

***Abstract*— The correlation preserving transform (CPT) is introduced to perform bivariate component analysis via decorrelating matrix decompositions, while at the same time preserving the integrity of original bivariate sources. Specifically, unlike existing bivariate uncorrelating matrix decomposition techniques, CPT is designed to preserve both the order of the data channels within every bivariate source and their mutual correlation properties. We introduce the notion of intraference to quantify the effects of interchannel mixing artifacts within recovered bivariate sources, and show that the integrity of separated sources is compromised when not accounting for the intrinsic correlations within bivariate sources, as is the case with current bivariate matrix decompositions. The CPT is based on augmented complex statistics and involves finding the correct conjugate eigenvectors associated with the pseudocovariance matrix, making it possible to maintain the physical meaning of the separated sources. The benefits of CPT are illustrated in the source separation and clustering scenarios, for both synthetic and real-world data.**

***Index Terms*— Augmented complex statistics, bivariate data analysis, correlation preserving transform (CPT), noncircularity, widely linear modeling.**

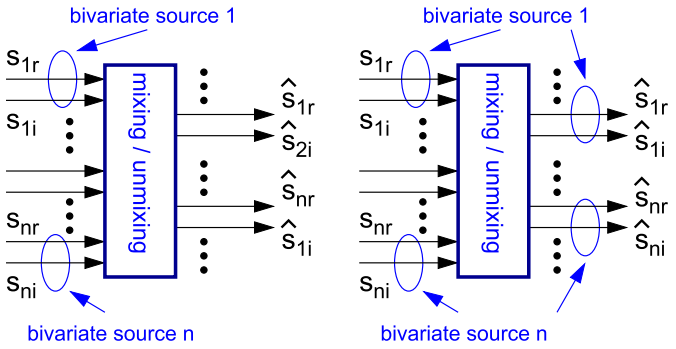

Fig. 1. Bivariate source separation. Left: current techniques both decorrelate the original bivariate sources and scatter the data channels within bivariate sources across the orthogonal component space. Right: we desire to maintain both the order and correlation properties of data channels within each bivariate source (integrity), while allowing for the permutation ambiguity of such coherent bivariate sources. The symbols S_{nr} and S_{ni} denote the real and imaginary parts of the nth bivariate source while the circumflex denotes the recovered sources.

I. Introduction

DECORRELATING (also known as prewhitening) transforms are commonly used in the design and analysis of complex learning systems, including neural networks, blind source separation [1], [2], nonnegative principal component analysis (PCA) [3], Hebbian-type learning techniques [4], and feature analysis [5]. These almost invariably operate on real vector-valued data (from univariate data through to CANDECOMP/PARAFAC tensor decompositions [6]–[8]). Current prewhitening transforms are not designed to preserve the inherent couplings that typically exist between the components in recovered vector-valued sources; for instance, the correlation between the real and imaginary parts within every complex (bivariate) source. The problem we consider is therefore generic and very important, especially when the physical meaning in complex learning systems needs to be maintained: we answer how to perform decorrelation between the sources while ensuring that the intrinsic correlation structure within the components of a given source is preserved. The shortcomings of overlooking the intrinsic correlation within the components of each bivariate source, while decorrelating the sources from one another, as is current practice, include the following:

1) *Source Integrity:* The well-known permutation ambiguity in source ordering after decorrelation means that the reconstructed sources appear in no particular order. This, in turn, causes the integrity problem, since the two univariate data channels belonging to a coherent bivariate source will not be aligned into a coherent bivariate-independent component (IC). Instead, the constituent data channels will be scattered anywhere across the space spanned by the separated data channels (see also Fig. 1).
2) *Indexing:* Upon correctly reconstructing every bivariate source from their mixtures, we need to index the channels pertaining to a particular source, and to keep track of the indexing. This is not trivial, as for nonstationary sources fixed indexing is not appropriate, since the intermittent new sources arising from the time-varying statistics are unlikely to be estimated in the original order.

Manuscript received January 31, 2013; revised March 29, 2014; accepted March 29, 2014. Date of publication June 17, 2014; date of current version February 16, 2015.

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

S. C. Douglas is with the Department of Electrical Engineering, Southern Methodist University at Dallas, Dallas, TX 75275-0338 USA (e-mail: d.mandic@imperial.ac.uk).

D. P. Mandic is with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K. (e-mail: douglas@engr.smu.edu).

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/TNNLS.2014.2316175

3) *Correlations:* A prerequisite for correct separation is for the original correlation structure to be preserved within each multidimensional recovered source, while at the same time decorrelating different multivariate signals (within the same vector space).

To overcome these obstacles, we propose the correlation preserving transform (CPT) that applies across complex learning systems. For convenience, we introduce it in the context of source separation, however, the concept straightforwardly generalizes to any type of complex neural networks. Our analysis is based on the strong uncorrelating transform (SUT) [9], [10], an elegant current technique, which is also prone to the above shortcomings.

Fig. 1 highlights the importance of designing matrix decompositions, which would account for the integrity of vector sources. To an extent, 2-D, 3-D, and 4-D data analyses in the context of complex- and quaternion-valued source separation address this issue [11], [12]. However, these do not exploit the full potential that the multivariate (augmented) statistics of such sources offer. Most importantly, by their design, current techniques destroyed the intrinsic cross-correlations inferring between the constituent source channels (components). To introduce the problem, we first consider the SUT for complex-valued data [9], [10], which has been designed based on the strong assumption of uncorrelatedness. As such, it is not suitable to estimate bivariate sources, which exhibit intrinsic correlation, a typical case in practice.

We first highlight that the problem of phase ambiguity is related to the treatment of the pseudocovariance matrix, which within SUT is kept real. Next, the CPT is proposed to provide a generic solution for bivariate source separation, which maintains the integrity of individual sources. The concept of **intraference** is introduced to quantify the extent to which the inherent statistical properties within original bivariate sources are violated. We show that not accounting for the **intraference** makes it impossible to associate physical meaning with recovered bivariate sources, as the intraference is manifested by the intrinsic mixing of the real and imaginary parts within every complex (bivariate) source. The analysis shows that the proposed CPT comprises the existing SUT, as a special case, highlighting the generality of CPT and the effects of the intraference present in SUT, which arises from the equivariant mixing–unmixing model. The performance of CPT is validated against PCA and SUT both analytically and through real-world simulation studies in brain computer interface (BCI) and wind modeling. The proposed statistical framework is shown to provide an enabling technology for the rigorous analysis of a wide range of complex-valued learning systems.

II. Bivariate Matrix Decompositions for Source Separation

For convenience, we introduce the concepts of intraference and correlation preserving bivariate transforms in the complex domain, as it is naturally suited to deal with phase information. The duality with bivariate vectors ensures that the concept also applies to real bivariate vectors [13]. To that end, basic notions from the augmented complex statistics and widely linear modeling are summarized in the following [14], [15].

For a complex-valued zero mean[1] random vector $\mathbf{x} = \mathbf{x}_r + \imath\mathbf{x}_i \in \mathbb{C}^N$, where $\imath = \sqrt{-1}$. The complete second-order information is provided only by accounting for both the covariance $\mathbf{C}$ and pseudocovariance $\mathbf{P}$ matrices, defined as [16]

$$\begin{aligned} \mathbf{C} &= E\{\mathbf{x}\mathbf{x}^H\} \\ \mathbf{P} &= E\{\mathbf{x}\mathbf{x}^T\} \end{aligned} \tag{1}$$

where $E\{\cdot\}$ is the statistical expectation operator. The covariance matrix $\mathbf{C}$ is symmetric and is commonly used to model second-order statistics of complex variables, whereas the pseudocovariance matrix $\mathbf{P}$ accounts for the correlation between the real and imaginary components and for a mismatch in the power between the real and imaginary data channel. A random vector with a vanishing pseudocovariance is termed as second-order circular or proper [17]. In general, the term circular refers to a rotation invariant probability distribution, while properness (second-order circularity) specifically refers to the second-order statistical properties (equal powers in data channels). Note that the majority of complex signals encountered in signal processing applications[2] are thus improper, highlighting the need to have a complete and unified treatment of such signals when designing complex learning systems.

A. Bivariate Matrix Analysis

Consider the vector of bivariate data $\mathbf{s}_B$, and a general case where a degree of correlation exists both within and between the data channels $\mathbf{s}_{BR}$ and $\mathbf{s}_{BI}$ of the source $\mathbf{s}_B = [\mathbf{s}_{BR}, \mathbf{s}_{BI}]$, so that the covariance matrix[3]

$$\mathbf{\Gamma}_B = E\{\mathbf{s}_B\mathbf{s}_B^T\} = \begin{bmatrix} \mathbf{\Lambda}_R & \mathbf{\Lambda}_{RI} \\ \mathbf{\Lambda}_{IR} & \mathbf{\Lambda}_I \end{bmatrix} \tag{2}$$

assumes a block diagonal structure. The covariance matrices of such bivariate real data channels (see real and imaginary) are denoted by $\mathbf{\Lambda}_R$, $\mathbf{\Lambda}_I$, and their cross-covariance by the matrices $\mathbf{\Lambda}_{IR}$ and $\mathbf{\Lambda}_{RI}$. Note that the symbol $\mathbf{\Lambda}$ reflects that the real-valued covariance $\mathbf{\Gamma}$ is diagonal. Therefore, to preserve the original correlation properties between the two data channels within the recovered individual bivariate components, it is necessary to design a procedure, which maintains the integrity of the block diagonal structure in (2). This is not straightforward to achieve, as generally the real-valued cross-covariance matrix $\mathbf{\Gamma}_{RI}$ between the two constitutive channels is not symmetric. Notice also the generality of (2), as when $\mathbf{\Gamma}_{RI} = \mathbf{0}$ we have the traditional case of a diagonal covariance matrix $\mathbf{\Gamma}_B = \mathbf{\Lambda}$, for which standard decorrelation and independent component analysis (ICA) models apply (circular case).

[1] In independent component analysis and blind source separation [11], it is common to assume centered sources.

[2] Either those complex by design, such as communications signals, or those made complex by convenience of representation, such as wind and EEG signals [14].

[3] For convenience, we denote the constituent data channels by $\mathbf{s}_{BR}$ and $\mathbf{s}_{BI}$ to facilitate the comparison with its dual complex representation.

An alternative way to address the block diagonal structure in (2) is to consider the corresponding complex-valued augmented covariance matrix $\mathbf{C}_A$, given by[4]

$$\mathbf{C}_A = E\{\mathbf{s}_A \mathbf{s}_A^H\} = \begin{bmatrix} \mathbf{C} & \mathbf{P} \\ \mathbf{P}^* & \mathbf{C}^* \end{bmatrix} \tag{3}$$

where

$$\begin{aligned} \mathbf{s}_A &= \mathbf{s}_{AR} + \imath \mathbf{s}_{AI} = \mathbf{s}_{BR} + \imath \mathbf{s}_{BI} \\ \mathbf{C} &= \mathbf{\Lambda}_R + \mathbf{\Lambda}_I \\ \mathbf{P} &= \mathbf{\Lambda}_R - \mathbf{\Lambda}_I + 2\imath \mathbf{\Lambda}_{RI}. \end{aligned} \tag{4}$$

The block diagonal structure (and thus the decorrelation between data channels) can be imposed on the augmented covariance matrix in (3) by the SUT [10], which jointly diagonalizes both the covariance $\mathbf{C}$ and the pseudocovariance $\mathbf{P}$ matrix; for more detail on SUT, see [9], [10], and [18]. Within the SUT, the pseudocovariance matrix is constrained to be real valued due to the use of a special form of singular-value decomposition, called the Takagi factorization. Thus, the real-valued bivariate covariance matrix $\mathbf{\Gamma}_B$ becomes strictly diagonal, causing the real-valued bivariate cross-covariance matrix $\mathbf{\Gamma}_{RI}$ to vanish, thus implying uncorrelatedness of the real and imaginary parts within every bivariate source and giving rise to the problem of source integrity addressed in this paper.

Remark 1: The SUT is blind to phase information as it forces the imaginary part of the pseudocovariance $\mathbf{P}$ in (4) to vanish. Therefore, when using Takagi factorization-based techniques, such as SUT and its variants [10], [19], [20], intrinsic mixing occurs within each individual bivariate recovered source, as the source pseudocovariance is rarely zero, yielding poor estimates of the sources stemming from the inherent inability of SUT to preserve source integrity (see also Footnote 5).

Our aim is therefore to introduce a novel diagonalization procedure, which preserves the inherent correlation, and thus physical meaning, within each recovered bivariate source, by also accounting for the cross-covariance matrix $\mathbf{\Gamma}_{RI}$.

B. Bivariate Matrices in the Context of Source Separation

To further illustrate the challenge in maintaining the original correlation structure (within bivariate sources) in (2), while decorrelating the sources from their mixtures, observe that the diagonal structures of the upper right and lower left submatrices in (2) require that in the dual, complex-valued space, the diagonal pseudocovariance matrix $\mathbf{P}$ in (3) is inherently complex valued.[5] However, the Takagi factorization within SUT which diagonalizes $\mathbf{P}$ enforces the estimated pseudocovariance to be real valued, leading to a complex-valued mixing–unmixing model $\mathbf{WH}$, which we formalize by considering the following ICA model:

$$\text{Mixing Model: } \mathbf{x}(n) = \mathbf{H}\mathbf{s}(n) \tag{5}$$

$$\text{Unmixing Model: } \mathbf{y}(n) = \mathbf{W}\mathbf{x}(n) = \mathbf{W}\mathbf{H}\mathbf{s}(n) \tag{6}$$

$$= \mathcal{P}\mathbf{D}\mathbf{s}(n) = \mathcal{P}\mathbf{\Delta}\mathbf{\Lambda}\mathbf{s}(n) \tag{7}$$

where $\mathbf{x}(n)$ is the vector of observed (mixed) bivariate signals, $\mathbf{y}(n)$ is the vector of recovered sources, $\mathbf{H}$ represents the linear mixing model or mixing coefficient matrix, and $\mathbf{W}$ is proportional to the inverse of $\mathbf{H}$ and is called the unmixing matrix. In the real domain, both the permutation matrix $\mathcal{P}$ and the scaling matrix $\mathbf{D}$ are trivial ambiguities, which can be overlooked in the analysis. In the complex domain, the situation is different, as the Euler representation of the scaling matrix, $\mathbf{D} = \mathbf{\Delta}\mathbf{\Lambda}$, whose diagonal elements $\delta_{ii} = \exp(\imath\theta_i)$ and $\lambda_{ii} = |d_{ii}|$, provides the desired phase and magnitude information (and hence the integrity of sources) for the diagonal elements of $\mathbf{D}$. In other words, our aim is to move beyond the usual complex-valued mixing matrix and real pseudocovariance model, and consider more physically meaningful real-valued mixing and complex pseudocovariance, and to show that this enables us to design a decorrelation technique, which preserves the integrity of bivariate sources, a key element in a number of practical applications.

Remark 2: To prevent intrinsic mixing between the real and imaginary parts of the extracted IC, we need to address the phase ambiguity, reflected in the matrix $\mathbf{\Delta}$.

The unmixing model in (6) leads to both the phase and real-valued scaling ambiguities, since

$$\begin{aligned} y(n) &\propto \delta s(n) \\ &\propto \underbrace{\left(\delta_R s_R(n) - \delta_I s_I(n)\right)}_{\text{real and imaginary parts of } s} + \imath\left(\delta_R s_I(n) + \delta_I s_R(n)\right) \end{aligned} \tag{8}$$

where δ denotes the complex-valued scaling ambiguity. Observe that $y(n)$ is a poor estimate of a source $s(n)$, since the real part $y_R(n)$ comprises of the real and imaginary parts of $s(n)$; a similar observation can also be made for the imaginary part $y_I(n)$. This problem was highlighted in [11, pp. 384–385] and [21] stating that it is not possible to solve this phase ambiguity.

To provide further evidence[6] that the phase ambiguity represented by the matrix $\mathbf{\Delta}$ arises due to the real-valued constraint imposed by the Takagi factorization, consider the pseudocovariance matrix of independent sources, which can be expressed as $\mathbf{P} = \mathbf{\Delta}_P \mathbf{\Lambda}_P$, to give

$$\mathbf{\Lambda}_P = \boldsymbol{W}\mathbf{H}\mathbf{P}\mathbf{H}^T\boldsymbol{W}^T \tag{9}$$

$$= \boldsymbol{W}\mathbf{H}\left(\mathbf{\Delta}_P^{1/2}\mathbf{\Lambda}_P\mathbf{\Delta}_P^{1/2}\right)\mathbf{H}^T\boldsymbol{W}^T \tag{10}$$

and consequently

$$\begin{aligned} \boldsymbol{W}\mathbf{H}\mathbf{\Delta}_P^{1/2} &= \mathbf{I} \\ \boldsymbol{W}\mathbf{H} &= \mathbf{\Delta}_P^{-1/2}. \end{aligned} \tag{11}$$

Remark 3: An appropriate mixing–unmixing model $\boldsymbol{W}\mathbf{H} \in \mathbb{C}$ would require a complex-valued $\mathbf{\Delta}_P^{-1/2}$, instead of a

[4]This duality is enabled by the isomorphism between $\mathbb{R}^2$ and the augmented complex field $\mathbb{C}$, see also [14].

[5] The pseudocovariance is always complex valued for correlated real and imaginary parts of a complex number. In addition, since the pseudocovariance $p = E\{xx^T\} = E\{|x|^2 e^{\imath 2\theta}\}$, the modulus $|x|^2$ is readily estimated by SUT, whereas SUT is blind to the phase θ, a subject of this paper.

[6]For clarity and without loss of generality, in the following we neglect the permutation ambiguity.

real-valued matrix used in current techniques, highlighting that the existing methods cannot reconstruct adequately general bivariate sources.

III. Correlation Preserving Transform

We shall now address the problem of integrity of bivariate sources stated in Section II in the complex domain, and provide a solution through the proposed CPT.

A. On Takagi Factorization and Its Computation

The Takagi factorization of a complex-valued symmetric matrix $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$ is a special case of the singular-value decomposition, whereby the diagonal matrix $\mathbf{\Lambda}$ is restricted to be real valued and positive. The key to the proposed CPT solution is to demonstrate that the unitary matrices $\mathbf{U}$ and $\mathbf{V}$ within the singular-value decomposition of a symmetric $\mathbf{A} = \mathbf{A}^T$ are related[7] by [18, p. 411]

$$\mathbf{V}^* = \mathbf{U}\mathbf{\Delta} \tag{12}$$

where $\mathbf{\Delta} = \text{diag}(e^{\imath\theta_1}, \ldots, e^{\imath\theta_n})$. For simplicity, it is generally assumed that $\mathbf{\Delta} = \mathbf{I}$, such that $\mathbf{U} = \mathbf{V}^*$, however, the matrix $\mathbf{\Delta}$ is not unique and may also depend on the numerical implementation[8] of the singular-value decomposition.

For rigor, we shall compute the Takagi factorization by exploiting the relationship in (12) as

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^H = \mathbf{U}\mathbf{S}\mathbf{\Delta}\mathbf{U}^T \tag{13}$$

$$= \mathbf{Q}\mathbf{S}\mathbf{Q}^T \tag{14}$$

where $\mathbf{Q} = \mathbf{U}\mathbf{\Delta}^{1/2}$ is a unitary matrix and the matrix $\mathbf{\Delta} = \mathbf{U}^H\mathbf{V}^*$ is diagonal.

Remark 4: There is a loss of phase information, when the matrix $\mathbf{\Delta}$ is absorbed into $\mathbf{U}$ in (13) so as to yield the matrix $\mathbf{Q}$ in Takagi factorization (14).

Remark 5: In addition to the phase ambiguity, since $\mathbf{Q}^H\mathbf{P}\mathbf{Q}^* = \mathbf{S} \in \mathbb{R}^{N\times N}$, we need to address the problem of finding conjugate eigenvectors (con-eigenvectors), which satisfy $\mathbf{A}\mathbf{U}^* = \mathbf{U}\mathbf{D}$. However, con-eigenanalysis is nonunique; for instance, if μ is a nonnegative con-eigenvalue of $\mathbf{A}$ then so too is $\mu\exp(\imath\theta)$, for any θ [18].

B. Strong Uncorrelating Transform

The SUT, formally introduced in [9] and then elaborated in the context of ICA in [10], performs decorrelation of multivariate complex data in two stages: 1) by performing decorrelation between the independent bivariate sources and 2) by subsequently decorrelating the real and imaginary parts of the bivariate ICs. In this way, the SUT diagonalizes both the covariance and pseudocovariance matrix as follows:

$$\begin{aligned}\mathbf{\Lambda}_C &= \boldsymbol{W}\mathbf{C}\boldsymbol{W}^H = \mathbf{I}\\ \mathbf{\Lambda}_P &= \boldsymbol{W}\mathbf{P}\boldsymbol{W}^T\end{aligned} \tag{15}$$

[7]The relation between the right and left eigenvector always holds for distinct singular values, however, when some of the singular values coincide this relationship may not hold [22].

[8]Real-valued singular values means that there is at least a phase ambiguity in the columns of $\mathbf{U}$ and $\mathbf{V}$.

where the symbols $\mathbf{\Lambda}_C$ and $\mathbf{\Lambda}_P$ denote, respectively, the real-valued diagonalized covariance and pseudocovariance matrices, and $\mathbf{I}$ and $\mathbf{W}$ are, respectively, the identity matrix and the SUT, given by

$$\boldsymbol{W} = \mathbf{Q}_P^H\mathbf{C}^{-\frac{1}{2}}. \tag{16}$$

The unitary matrix $\mathbf{Q}_P$ can be obtained from the Takagi factorization of the normalized pseudocovariance matrix $\mathbf{C}^{-(1/2)}\mathbf{P}\mathbf{C}^{-(1/2)^T} = \mathbf{Q}_P\mathbf{\Lambda}_P\mathbf{Q}_P^T$.

C. Proposed CPT

We have shown in Section II and in Remark 4 that the estimation of the phase matrix $\mathbf{\Delta}$ is a prerequisite for the correlation preservation within individual ICs, and also to prohibit the inherent mixing between the real and imaginary parts within every recovered bivariate source that occurs when using standard ICA, as stated in (8). A simple yet effective solution is based on the result in (11), which states that the SUT can be expressed as

$$\boldsymbol{W} = \mathbf{\Delta}_P^{-1/2}\mathbf{H}^{-1} \tag{17}$$

which implies that

$$\boldsymbol{W}^{-1} = \mathbf{H}\mathbf{\Delta}_P^{1/2}. \tag{18}$$

Upon dividing each element of $\boldsymbol{W}^{-1}$ by its magnitude, we obtain the phase matrix of the form

$$\mathbf{\Omega}_P = \begin{bmatrix} \pm\exp(\imath\theta_1) & \cdots & \pm\exp(\imath\theta_n) \\ \vdots & \ddots & \vdots \\ \pm\exp(\imath\theta_1) & \cdots & \pm\exp(\imath\theta_n) \end{bmatrix} \tag{19}$$

whose diagonal square root

$$\mathbf{\Delta}_P^{1/2} = \begin{bmatrix} \exp(\imath\theta_1) & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \exp(\imath\theta_n) \end{bmatrix}.$$

Observe that, if $\mathbf{\Omega}_P$ is not of the form in (19), that is, when not all the elements of the ith column are $\pm\exp(\imath\theta_i)$ but assume the value of unity, then $\mathbf{\Delta}_P^{1/2} = \mathbf{I}$. For this particular case, as shown later in (21) that CPT simplifies into SUT.

Additionally, there are many application scenarios, which only require the multiple bivariate sources to be mutually decorrelated, while at the same time, it is an imperative to preserve the cross-covariance structure between the constitutive components of each individual bivariate source. In this way, we preserve the integrity and physical meaning of the bivariate-independent components. We can also associate each diagonal element of $\mathbf{\Delta}_P^{1/2}$ with the square root of its corresponding normalized element p_{ii} of the pseudocovariance of the original data, that is

$$\mathbf{\Delta}_P^{1/2} = \begin{bmatrix} (p_{11}/|p_{11}|_2)^{1/2} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & (p_{nn}/|p_{nn}|_2)^{1/2} \end{bmatrix} \tag{20}$$

ensuring that the phase information regarding the individual bivariate components of the original multivariate data are preserved.

Remark 6: The form in (20) has the following desired properties: 1) the $\pm$ sign ambiguity problem is completely bypassed, as we can reconstruct the bivariate sources irrespective of this ambiguity and 2) the scaling ambiguity is also bypassed by the normalization of each element of $\boldsymbol{W}^{-1}$.

The proposed CPT therefore assumes the form

$$\boldsymbol{V} = \boldsymbol{\Delta}_P^{1/2}\mathbf{W} \tag{21}$$

where the SUT $\mathbf{W}$ can be calculated from (16). Observe that the computational complexity of the CPT in (21) is effectively the same as that of SUT, as $\boldsymbol{\Delta}_P^{1/2}$ is a diagonal matrix and the cost of premultiplication of the SUT matrix $\mathbf{W}$ with $\boldsymbol{\Delta}_P^{1/2}$ is negligible compared with the computational complexity of the two singular values involved in the SUT.

Remark 7: Observe that if the mixing matrix $\mathbf{H}$ in (5) is complex valued, it would not have been possible to perform the phase estimation accurately. Therefore, CPT generalizes SUT, as SUT does not cater for the phase ambiguity at all.

The advantages of accounting for phase information by CPT are supported by the simulation studies in Section V.

IV. Intraference Analysis of SUT and CPT

The performances of SUT and CPT are next analyzed, based on the standard small error assumption [18]. The error analysis is motivated by that in [23], however, our work focuses on the novel concept of intraference illustrated in (8), instead of the standard interference analyzed in [23].

A. Error Model

As illustrated above, the error in bivariate ICA arises from:

1) the imperfection of the mixing–unmixing model $\mathbf{WH}$;
2) the misspecification of the assumed statistics of the ICs, such as the assumption of real-valued nature of their pseudocovariance matrix.

To quantify those uncertainties, we employ the following error models:

$$\mathbf{WH} = \mathbf{I} + \boldsymbol{\Phi}, \quad \hat{\mathbf{C}} = \mathbf{I} + \boldsymbol{\xi}_c, \quad \hat{\mathbf{P}} = \mathbf{P} + \boldsymbol{\xi}_p \tag{22}$$

where $\boldsymbol{\Phi}$ is the intraference matrix, and $\boldsymbol{\xi}_c$ and $\boldsymbol{\xi}_p$ are the errors in estimating the covariance and the pseudocovariance matrices. This allows us to formulate the covariance and pseudocovariance of the estimated ICs in the form

$$(\mathbf{I}+\boldsymbol{\Phi})\hat{\mathbf{C}}\big(\mathbf{I}+\boldsymbol{\Phi}^H\big) = \mathbf{I} \quad (\mathbf{I}+\boldsymbol{\Phi})\hat{\mathbf{P}}\big(\mathbf{I}+\boldsymbol{\Phi}^T\big) = \tilde{\mathbf{P}} \tag{23}$$

where $\tilde{\mathbf{P}}$ denotes the pseudocovariance matrix estimated by either CPT or SUT. Expanding these expressions, we obtain the following approximations:

$$\begin{aligned} \mathbf{I} &\approx \mathbf{I} + \boldsymbol{\Phi} + \boldsymbol{\Phi}^H + \boldsymbol{\xi}_C \\ \boldsymbol{\xi}_C &\approx -\big(\boldsymbol{\Phi} + \boldsymbol{\Phi}^H\big) \end{aligned} \tag{24}$$

$$\begin{aligned} \tilde{\mathbf{P}} &\approx \mathbf{P} + \boldsymbol{\Phi}\mathbf{P} + \mathbf{P}\boldsymbol{\Phi}^T + \boldsymbol{\xi}_P \\ \tilde{\boldsymbol{\xi}}_P &\approx -\big(\boldsymbol{\Phi}\mathbf{P} + \mathbf{P}\boldsymbol{\Phi}^T\big) \end{aligned} \tag{25}$$

where $\tilde{\boldsymbol{\xi}}_P = \boldsymbol{\xi}_P - (\tilde{\mathbf{P}} - \mathbf{P})$.

B. Definition of the Intraference

Based on the error model in (22), the intraference for the kth bivariate source can be defined in terms of the deviation of diagonal term $[kk]$ of the mixing–unmixing model in (6) from unity

$$E\big\{\Phi^2[k,k]\big\} = E\{|(\mathbf{WH})[k,k]-1|^2\}. \tag{26}$$

This definition is based on the assumption that there is no permutation. In practice, to circumvent the permutation ambiguity, sources are sorted in a descending order of the absolute value of their pseudocovariance (noncircularity). This is logical because when Takagi factorization is employed, singular-value decomposition sorts the ICs in a descending order of their singular values.

C. Analysis of the Intraference

To derive the expression for the intraference $\boldsymbol{\Phi}[k,k]$ from (24) and (25), we consider all the ith diagonal elements of the matrices in (24) and (25) and express them as

$$\begin{aligned} \boldsymbol{\varepsilon}_k &\approx -\boldsymbol{\Psi}_k\boldsymbol{\theta}_k \\ \begin{bmatrix} \varepsilon_{CR}[k,k] \\ \tilde{\varepsilon}_{PR}[k,k] \\ \tilde{\varepsilon}_{PI}[k,k] \end{bmatrix} &\approx -\begin{bmatrix} 1 & 1 & 0 \\ P_R[k,k] & 0 & -P_I[k,k] \\ P_I[k,k] & 0 & P_R[k,k] \end{bmatrix}\begin{bmatrix} \Phi_R[k,k] \\ \Phi_R[k,k] \\ \Phi_I[k,k] \end{bmatrix}. \end{aligned} \tag{27}$$

The vector $\boldsymbol{\varepsilon}_k$ can also be expressed in terms of its complex-valued counterpart

$$\begin{aligned} \boldsymbol{\varepsilon}_k &= \frac{1}{2}\boldsymbol{\Omega}\boldsymbol{\epsilon}_k \\ \begin{bmatrix} \varepsilon_{CR}[k,k] \\ \tilde{\varepsilon}_{PR}[k,k] \\ \tilde{\varepsilon}_{PI}[k,k] \end{bmatrix} &= \frac{1}{2}\begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & -\imath & \imath \end{bmatrix}\begin{bmatrix} \xi_C[k,k] \\ \tilde{\xi}_P[k,k] \\ \tilde{\xi}_P^*[k,k] \end{bmatrix}. \end{aligned} \tag{28}$$

As $E\big\{\boldsymbol{\varepsilon}_k\boldsymbol{\varepsilon}_k^T\big\} = (1/4)\boldsymbol{\Omega}E\big\{\boldsymbol{\epsilon}_k\boldsymbol{\epsilon}_k^H\big\}\boldsymbol{\Omega}^H$, we have[9]

$$\begin{aligned} &E\big\{\boldsymbol{\varepsilon}_k\boldsymbol{\varepsilon}_k^T\big\} \\ &= \begin{bmatrix} 4\xi_C^2 & 2(\xi_C\tilde{\xi}_P^* + \xi_C\tilde{\xi}_P^*) & 2\imath(\xi_C\tilde{\xi}_P^* - \xi_C\tilde{\xi}_P) \\ 2(\tilde{\xi}_P\xi_C + \tilde{\xi}_P^*\xi_C) & 2\tilde{\xi}_P\tilde{\xi}_P^* + \tilde{\xi}_P^*\tilde{\xi}_P^* + \tilde{\xi}_P\tilde{\xi}_P & \imath(\tilde{\xi}_P^*\tilde{\xi}_P^* - \tilde{\xi}_P\tilde{\xi}_P) \\ 2\imath(\tilde{\xi}_P^*\xi_C - \tilde{\xi}_P\xi_C) & \imath(\tilde{\xi}_P^*\tilde{\xi}_P^* - \tilde{\xi}_P\tilde{\xi}_P) & -(\tilde{\xi}_P^*\tilde{\xi}_P^* + \tilde{\xi}_P\tilde{\xi}_P) \end{bmatrix} \\ &= \begin{bmatrix} \xi_C^2 & \xi_C\tilde{\xi}_{PR} & \xi_C\tilde{\xi}_{PI} \\ \xi_C\tilde{\xi}_{PR} & \tilde{\xi}_{PR}^2 & \tilde{\xi}_{PR}\tilde{\xi}_{PI} \\ \xi_C\tilde{\xi}_{PI} & \tilde{\xi}_{PR}\tilde{\xi}_{PI} & \tilde{\xi}_{PI}^2 \end{bmatrix} \end{aligned} \tag{29}$$

which can be used to compute the covariances of interest, that is, $E\big\{\boldsymbol{\theta}_k\boldsymbol{\theta}_k^T\big\} = \boldsymbol{\Psi}_k^{-1}E\big\{\boldsymbol{\varepsilon}_k\boldsymbol{\varepsilon}_k^T\big\}\boldsymbol{\Psi}_k^{-T}$, where

$$\begin{aligned} \boldsymbol{\Psi}_k^{-1} &= \frac{-1}{P_R^2[k,k] + P_I^2[k,k]} \\ &\quad\times\begin{bmatrix} 0 & P_R[kk] & P_I[k,k] \\ -(P_R^2[k,k] + P_I^2[k,k]) & -P_R[k,k] & -P_I[k,k] \\ 0 & -P_I[k,k] & P_R[k,k] \end{bmatrix}. \end{aligned} \tag{30}$$

Notice that to measure the intraference, only the knowledge of $E\big\{\Phi_R^2[k,k]\big\}$ and $E\big\{\Phi_I^2[k,k]\big\}$ is required, that is, it is

[9]For clarity, in the sequel we omit the index $[k,k]$.

sufficient to calculate just the diagonal elements [1, 1] and [3, 3] of the matrix $E\{\boldsymbol{\theta}_k\boldsymbol{\theta}_k^T\}$, which is achieved by

$$E\{\Phi_R^2\} = \frac{\left[P_R\tilde{\xi}_{PR} + P_I\tilde{\xi}_{PI}\right]^2}{\left[P_R^2 + P_I^2\right]^2} \text{ when } \xi_C = 0 \quad (31)$$

$$E\{\Phi_I^2\} = \frac{\left[P_R\tilde{\xi}_{PI} - P_I\tilde{\xi}_{PR}\right]^2}{\left[P_R^2 + P_I^2\right]^2}. \quad (32)$$

The value of intraference is thus given by $E\{\Phi^2[k,k]\} = E\{\Phi_R^2[k,k]\} + E\{\Phi_I^2[k,k]\}$, and assumes the value of

$$E\{\Phi^2[k,k]\} = E\{|\tilde{\xi}_P|^2\}/|P[k,k]|^2. \quad (33)$$

The so simplified expression for $E\{\Phi_R^2\}$ in (31) is perfectly valid, as the misspecification error of the covariance $\xi_C[k,k] = 1 - \left(\sum_N s_i s_i^*\right)/N$ can be overlooked due to the scaling ambiguity.

Remark 8: The phase ambiguity arises due to the error in the estimation of the pseudocovariance matrix and is not related to the errors in the estimation of the covariance matrix, highlighting the caveats of the existing bivariate matrix decorrelation algorithms and the advantages of the proposed CPT.

Remark 9: The intraference measure is invariant to the magnitude of the pseudocovariance when $|P[k,k]| = |\hat{P}[k,k]|$.

This can be explained as follows. The intraference (33) arises from both the misspecification error of the pseudocovariance of the kth IC and the estimation error during the demixing process. In a perfect scenario, where the estimated singular values of the ICs correspond to their exact values, i.e., $|P[k,k]| = |\hat{P}[k,k]|$, the error $\tilde{\xi}_P[k,k]$ can be expressed in an Euler form as

$$\tilde{\xi}_P[k,k] \approx \left|P[k,k]\right| E\{\left[\exp(\iota\theta_k) - \exp(\iota\hat{\theta}_k)\right]\} \quad (34)$$

where $\hat{\theta}$ is obtained from the pseudocovariance matrix of the estimated ICs. Thus, (33) can be simplified to

$$E\{\Phi^2[k,k]\} \approx E\{|\exp(\iota\theta_k) - \exp(\iota\hat{\theta}_k)|^2\}. \quad (35)$$

In practice, however, $|P[k,k]| \neq |\hat{P}[k,k]|$ and therefore the expression in (33) should be used to measure the intraference.

Remark 10: For the CPT, when the condition in (19) is satisfied then the estimate of CPT $\hat{\theta} \approx \theta$ leads to $E\{\Phi^2[k,k]\} \approx 0$. In the worst case Scenario, where $\theta_k \neq \hat{\theta}_k$ and the condition (19) is not satisfied, the CPT degenerates into the SUT, thus consistently performing no worse than the SUT.

Remark 11: For the standard SUT, $\hat{\theta}_k = 0$ always holds. Thus, if $\theta_i \neq 2\pi k$, for any integer k, the SUT is unconditionally biased, since in this particular case the intraference is given by $E\{\Phi^2[k,k]\} \approx 2[1 - E\{\cos(\theta_k)\}]$.

V. Simulations and Discussion

To verify the proposed CPT, comprehensive simulations were conducted for both synthetic and real-world data. Studies based on synthetic data were performed to assess: 1) the behavior of CPT for varying phase angles; 2) the performance of CPT against the varying number of sources; and 3) the robustness of CPT in the presence of noise. The experiments on real-world data were: 1) the removal of ocular artifacts in electroencephalography (EEG) and 2) clustering analysis of 3-D wind field to discover hidden patterns. Both classes of experiments were used to summaries the advantages of CPT over the existing SUT [9], [10], FastICA [11], BSE [24] algorithms, as shown in Table II.

A. Experiment 1: Performance Analysis for Synthetic Data

Each simulation was averaged over 100 independent trials, where the elements of the mixing matrix were generated from a Gaussian distribution. These experiments also demonstrated the validity of the theoretical approximation of the intraference derived in Section IV-C and highlighted the caveats of not addressing the intrinsic correlation within each bivariate source and phase ambiguity in complex-valued ICA. The degree of correlation between the real and imaginary parts of complex-Gaussian zero mean signals $x = x_R + \iota x_I$ was controlled by combining two zero mean uncorrelated variables x_1 and x_2 in the following way:

$$x_R(t) = x_1(t), \quad x_I(t) = \rho x_1(t) + x_2(t)\sqrt{1-\rho^2} \quad (36)$$

where ρ denotes the correlation coefficient used to generate a signal with nonvanishing phase angle ranging from uncorrelated ($\rho = 0$) to fully correlated ($\rho = 1$) phase angle. To assess the degree of improperness of the so generated complex-valued signals (with both the nonvanishing variance c and pseudocovariance $p = p_R + \iota p_I$, where $|p_R| + |p_I| = 1$ for consistency), the correlation coefficient ρ was related to the pseudocovariance through

$$\rho = \frac{p_I}{2E\{x_1^2\}} \quad (37)$$

since the pseudocovariance $p = E\{x_R^2 - x_I^2\} + 2\iota E\{x_R x_I\}$. Here, the variances of the real and the imaginary parts had to be chosen so as to satisfy the two constraints $p_R = E\{x_R^2 - x_I^2\}$ and $c = E\{x_R^2 + x_I^2\}$. For our experiments, the covariance was set to $c = 1.1$; for example, given $p = -0.1 + \iota 0.9$, representing a phase angle of 0.45 radians, the following parameters were used:

$$\begin{aligned} E\{x_R^2\} &= E\{x_1^2\} = 0.5 \\ E\{x_I^2\} &= 0.6 \\ E\{x_2^2\} &= \frac{1.1 - E\{x_1^2\}(1+\rho^2)}{1-\rho^2}. \end{aligned} \quad (38)$$

A unit variance was not selected, to avoid singular cases, such as when the pseudocovariance $p = -0.1 + \iota 0.9$. In this case, the conditions $E\{x_R^2\} = 0.45$ and $E\{x_I^2\} = 0.55$ need to be satisfied so that $p_R = -0.1$, however, this also means that $E\{x_2^2\} \to \infty$ in (38), as $\rho = 1$. Clearly, $E\{x_2^2\} \to \infty$ cannot be implemented in practice.[10] Finally, the sources were generated from the moving average model $y(n) =$

[10]Another potential caveat is when there are multiplicities of singular values of the pseudocovariance matrix. This can be avoided by selecting a pseudocovariance of different magnitude for each source.

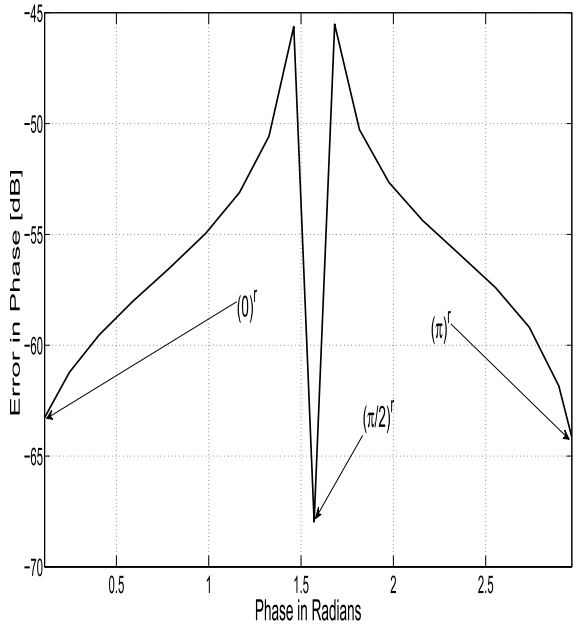

Fig. 2. Performance of CPT in terms of the phase angle of the pseudocovariance.

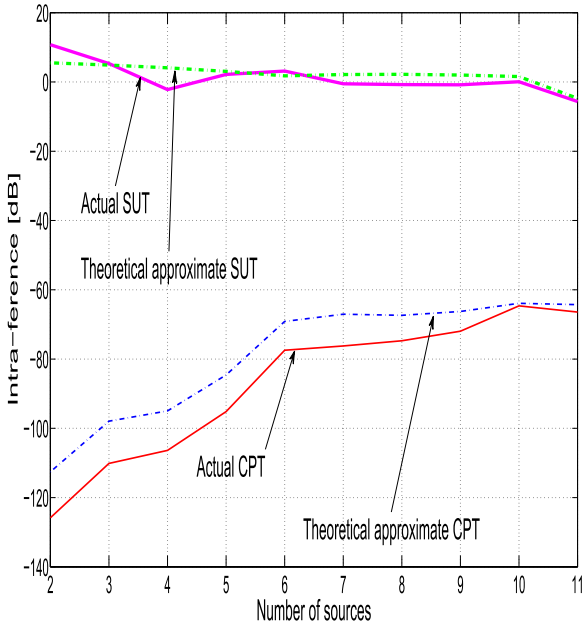

Fig. 3. Intraference in CPT and SUT as a function of the number of sources.

$x(n) + 0.9x(n-1) + 0.95x(n-2)$, as a complex valued filtering process would significantly alter the statistics of the signal x. We considered the mean value of intraference in the simulations, that is, $(1/N)\sum_{i=1}^{N} \Phi^2[i,i]$, with $\Phi^2[i,i]$ calculated from (26).

Remark 12: The intraference value can be expressed in terms of the correlation coefficient ρ as

$$E\{\Phi^2[k,k]\} = E\{|\tilde{\xi}_P|^2\}/\left[P_R^2 + 4\rho^2 E\{x_R^2\}\right]^2. \quad (39)$$

This is because using the relationship in (37) derived from the signal model (36), the imaginary part of the pseudocovariance can be obtained as $P_I^2 = 4\rho^2 E\{x_R^2\}$. This relationship can then be substituted into (33) to yield (39).

1) Performance as a Function of Phase Angle: In the first set of simulations, the performance of CPT was assessed against the phase angle associated with the pseudocovariance, $\angle\left(E\{x^2\}\right)$. Without loss of generality, a 2×2 source separation scenario was considered: one source had a zero phase angle $\angle\left(E\{x^2\}\right) = 0$ and the phase angle of the other source was nonzero.

Fig. 2 shows the excellent performance of CPT, which for a range of phase angles decorrelated bivariate sources with the intraference not exceeding -45 dB. Notice the three dips in the curve at the phase angles of $\{0, \pi/2, \pi\}$. The minimum points at $\angle\left(E\{x^2\}\right) = \{0, \pi\}$ correspond to the trivial scenarios corresponding to the sign ambiguity of ± 1, whereas for the phase angle $\angle\left(E\{x^2\}\right) = \pi/2$, CPT is required to estimate only the imaginary part of the pseudocovariance, instead of both the real and imaginary parts.

2) Performance Analysis for a Varying Number of Sources: In the second set of simulations, the source separation problem was considered for a varying number of bivariate sources. Fig. 3 shows the excellent performance of CPT, which maintained the intraference below -60 dB. On the other hand, the performance of SUT was unconditionally biased in the region of 0 dB, highlighting the caveat of not addressing the intraference shown in Remark 11. The theoretical approximation of the intraference for CPT followed quite closely the actual intraference; the difference between the theoretical approximation and the actual intraference decreased as the number of sources increased indicating that CPT is a consistent estimator. The estimation of the phase angle for the kth source was calculated by taking the average of the kth column of the matrix in (19).

3) Performance Analysis in the Presence of Noise: For rigor, the robustness of the proposed CPT was further investigated in the presence of additive doubly white Gaussian noise. Fig. 4 shows that the CPT outperformed the SUT, and that performance improved as the signal-to-noise ratio increased. The discrepancy between the theoretical approximation and the actual intraference at low signal-to-noise ratios stems from the small error assumption used in theoretical analysis. For high signal-to-noise ratios of 30 dB in Fig. 4, the small error assumption holds, the theoretical approximation thus followed closely the actual intraference.

B. Experiment 2: ICA for the Separation of Ocular Artifacts From EEG

The usefulness of CPT is next illustrated for the separation of eye muscle activity electrooculogram (EOG) from real-world EEG recordings. In real-time BCI it is desired to identify and remove such ocular artifacts from the contaminated EEG [25]. In our experiment, the EEG signals were from the electrodes Fp1, Fp2, while the EOG activity was recorded from the EOG1 and EOG2 channels with the electrodes placed above and on the side of the left eye socket [24]. The EEG data were sampled at 512 Hz and recorded for 12 s. Notice the pronounced presence of the eye blinks in the EEG mixtures corresponding to the frontal electrodes Fp1 and Fp2 (Fig. 5), and the clear presence of two signal components: the eye blink and the EEG activity.

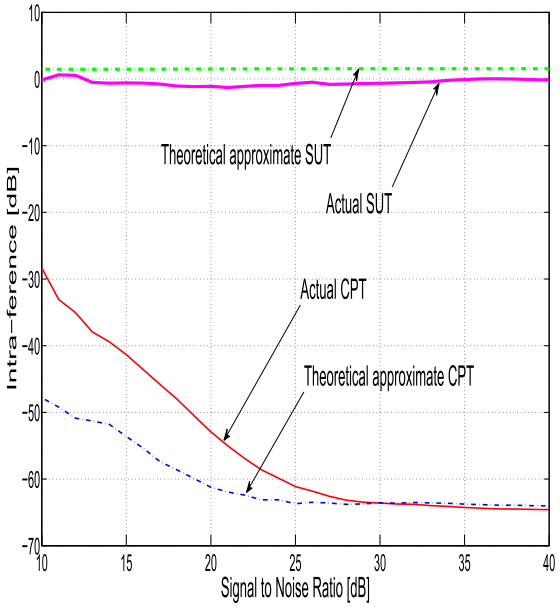

Fig. 4. Robustness of performance of CPT in presence of noise.

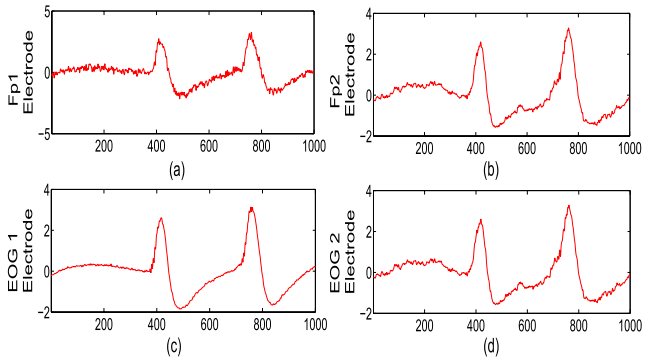

Fig. 5. Waveforms of EEG and EOG mixtures recorded at a sampling frequency of 512 Hz. (a) Mixed EEG and EOG. (b) Mixed EEG and EOG. (c) Pure EOG. (d) Pure EOG.

We adopted the same approach as in [24] and [26] to form the complex signals

$$\mathbf{x}_1 = \mathbf{Fp1} + \jmath\mathbf{Fp2}$$
$$\mathbf{x}_2 = \mathbf{EOG1} + \jmath\mathbf{EOG2}. \tag{40}$$

The left subplots of Fig. 6 show the separated ocular artifact (eye blink) is shown. Notice the noisier estimate produced by the SUT and the less pronounced EOG peaks in real-valued ICA estimates [11], compared with the reference in Fig. 5. Although the blind source extraction (BSE) technique in [24] produced a less noisy source estimate of the eye blink, it failed to extract an EEG source free from eye blink artifacts. Excellent and physically meaningful results were produced using the proposed CPT method (last row in Fig. 6).

Statistical validation of results: an important concept in BCI is the existence of synchrony between the two EEG sources [27]. Since the EOG signals can be regarded as the closest proxy to eye blinks (see Fig. 5), in our statistical analysis they were used to validate whether the extracted sources were either eye blinks or EEG activity.

Surrogate data analysis and hypothesis testing were next employed for statistical validation. Surrogate data are statistically similar to the original data, with identical mean, variance

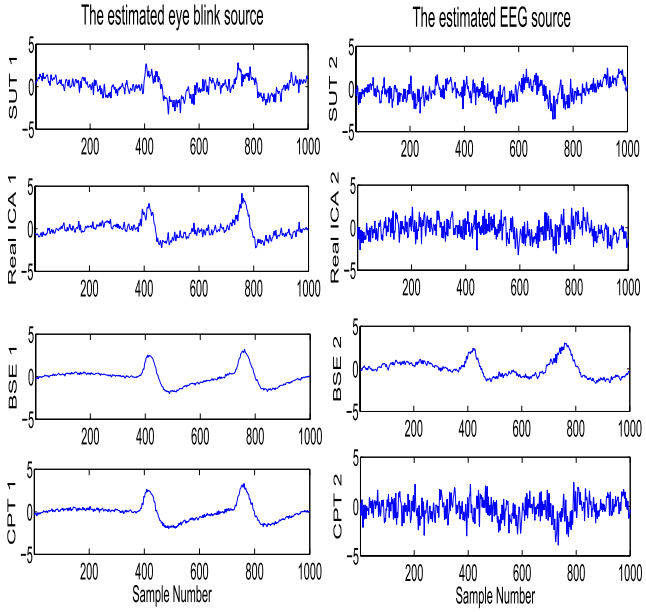

Fig. 6. Recovered eye blink (left column) and EEG (right column) sources estimated by SUT, real ICA, BSE, and CPT algorithms. Observe that the CPT retained perfect physical meaning of the EOG artifact and useful EEG with negligible intraference.

TABLE I

STATISTICAL VERIFICATION OF ESTIMATED SOURCES. THE THRESHOLD SYNCHRONY SCORE WAS 0.55 WITH A CONFIDENCE LEVEL OF 99%

Signal	Synchrony Score with EOG 1	Null Hypothesis	Correlation coef ρ with EOG 1
EOG 2 (test)	0.81	rejected	0.97
SUT 1	0.40	valid	0.71
SUT 2	0.39	valid	-0.10
Real ICA 1	0.57	rejected	0.90
Real ICA 2	0.39	valid	-0.02
BSE 1	0.83	rejected	0.99
BSE 2	0.86	rejected	0.87
CPT 1	0.85	rejected	0.99
CPT 2	0.42	valid	-0.13

and amplitude spectrum; however, the phase information of the original data are completely randomized using a uniform distribution between 0 and 2π. Any phase information is thus destroyed by the phase randomization step, and the null hypothesis considered was that there was no phase synchronization between the EOG source and other EEG activities. We generated 200 surrogates to calculate the synchrony score threshold (= 0.55), which governed the range corresponding to the validity of the null hypothesis with a confidence level of 99%. Table I shows that the synchrony score of test signal EOG 2 (Fig. 5) was above the threshold of 0.55, meaning that the null hypothesis was rejected and the recovered source must be an EOG signal. The rejected null hypothesis was also confirmed by the high degree of correlation between EOGs 1 and 2, shown in Table I.

C. Experiment 3: Clustering of 3-D Wind Measurements

3-D wind data were recorded in an urban environment[11] by two anemometers located 5 m apart in the North–South

[11]We thank Prof. Aihara and his team at Tokyo University for providing the wind dataset.

TABLE II
COMPARISONS BETWEEN SUT [9], [10], FASTICA [11], BSE [24], AND PROPOSED CPT

Properties	SUT	FastICA	BSE	CPT
1. Ability to cater for intrinsic correlation	No	N/A	Yes	Yes
2. Suitability to estimate multiple sources	Yes	Yes	No	Yes
3. Ability to prevent two different univariate sources from forming a bivariate source	No	N/A	Yes	Yes
4. Suitability for real-time processing	No	No	Yes	No
5. Simplicity in terms of ease of use	Yes	Yes	No	Yes
6. Use of Second Order Statistics	Yes	No	Yes	Yes

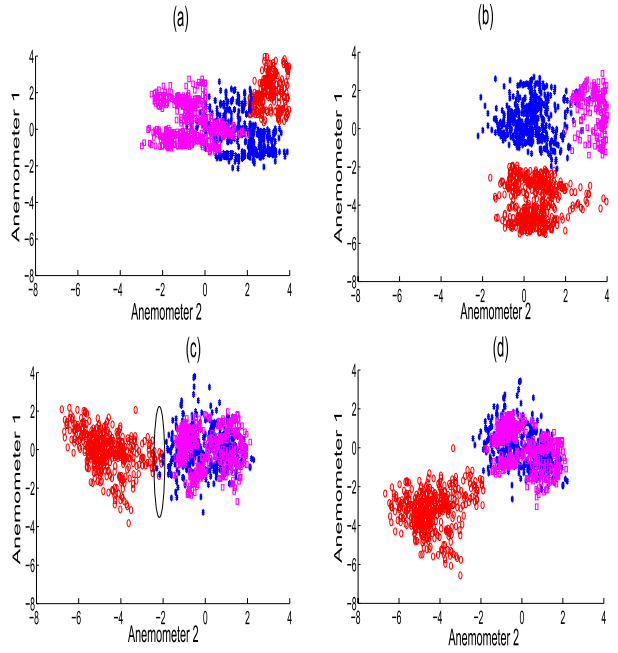

Fig. 7. Cluster analysis of 3-D wind data. (a) Original wind data. (b) Real-valued ICA. (c) SUT. (d) CPT. The red, blue, and magenta clusters correspond, respectively, to North–South, East–West, and vertical wind directions.

direction; the sampling frequency was 50 Hz. The first anemometer was located 5 m North from the second one. As wind speeds in the primary directions were correlated [28], [29], our aim was to discover the hidden orthogonal patterns along the three perpendicular wind directions: the North–South, East-West, and vertical direction. In doing so, the key was to preserve the correlation between the wind speeds recorded at different spatial locations, but pointing to the same direction.

Fig. 7 shows the scatter plots (real-imaginary) of the wind data, showing (a) the original data, and wind cluster estimates using (b) real-valued FastICA [11], (c) SUT and (d) CPT. For a fair comparison, all data were normalized. The original data in Fig. 7(a) suggests the existence of three distinctive wind components, partially overlapping. The real-valued ICA in Fig. 7(b) separated the three wind components, also suggesting three distinct data clusters. On the other hand, the SUT and CPT estimates correctly indicate only two different wind components, i.e., the North–South component and either East–West or vertical component. The overlap of the East–West and vertical wind components can be explained by the fact that the East–West wind component arose mainly due to the vertical wind component (reflection in urban space). These correct hidden patterns from SUT and CPT estimates were not obvious from the scatter plots of the original wind data, the reason being that the wind was blowing predominantly in the North direction. Indeed, the wind directions at the two anemometers had bearings of 11.39° and 8.07°, meaning the North–South wind component was dominant.

Observe two main differences between the PCA/SUT and CPT of the 3-D wind components. First, the correlation within any of the bivariate wind components was preserved in the CPT estimates, as shown in Fig. 7(d) and the bivariate components lying along the diagonal line $y = x$; a similar observation can be made for the original data in Fig. 7(a). Second, there was an overlap in the encircled region between the North–South and East–West components in the SUT estimates in Fig. 7(c). The much improved separation between the North–South and East–West/vertical components in the CPT estimates over the PCA/SUT estimates is evident in Fig. 7(d), facilitating general bivariate pattern classification.

Table II illustrates that our proposed CPT inherits all advantages of SUT and FastICA, and also exhibits two additional desired properties in rows 1 and 2 in the table. These two properties are also available in the BSE technique proposed in [24], however, the noncircular BSE technique is not straightforward to use due to its nonparametric nature.

VI. CONCLUSION

This paper has addressed critical open issues arising from the use of Takagi factorization in decorrelation of complex-valued matrices, a key step in the analysis of complex-valued learning systems. We have proposed a solution for preserving the integrity of bivariate sources; an issue largely overlooked in both current real-and complex-valued matrix decompositions. To that end, we have equipped matrix decompositions with the ability to maintain the correlation between the two data channels within each recovered bivariate source, thus making it possible to preserve the integrity of the separated sources. In doing so, we have solved the phase ambiguity problem, allowing for more degrees of freedom by accounting for the complex-valued nature of the pseudocovariance. In addition, a novel criterion for quantifying the degree to which the integrity of separated bivariate sources is preserved, referred to as the intraference, has been introduced highlighting that current bivariate matrix decompositions are subjected to an intrinsic mixing between the real and imaginary parts of each source. It has been further shown that the performance index of CPT in recovering bivariate sources has followed closely the theoretical performance, while the current techniques were inadequate for coupled source channels. Case studies in the

context of source separation for BCI and cluster analysis for 3-D wind data support the analysis.

REFERENCES

[1] S. Douglas and A. Cichocki, "Self-stabilized gradient algorithms for blind source separation with orthogonality constraints," *IEEE Trans. Neural Netw.*, vol. 11, no. 6, pp. 1490–1497, Nov. 2000.

[2] S. Douglas and A. Cichocki, "Neural networks for blind decorrelation of signals," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2829–2842, Nov. 1997.

[3] M. Plumbley, "A 'nonnegative PCA' algorithm for independent component analysis," *IEEE Trans. Neural Netw.*, vol. 15, no. 1, pp. 66–76, Jan. 2004.

[4] S. Choi, "Differential Hebbian-type learning algorithms for decorrelation and independent component analysis," *Electron. Lett.*, vol. 34, no. 9, pp. 900–901, 1998.

[5] C. Liu and H. Wechsler, "Independent component analysis of Gabor features for face recognition," *IEEE Trans. Neural Netw.*, vol. 14, no. 4, pp. 919–928, Jul. 2003.

[6] D. Nion, L. De Lathauwer, and I. Fijalkov, "Generalized PARAFAC methods for blind source extraction. Application in DS-CDMA," Ph.D. dissertation, Dept. Electr. Eng., Katholieke Univ. Leuven, Leuven, Belgium, 2007.

[7] H. Becker, P. Comon, L. Albera, M. Haardt, and I. Merlet, "Multi-way space-time-wave-vector analysis for EEG source separation," *EURASIP Signal Process. J.*, vol. 92, no. 4, pp. 1021–1031, Apr. 2012.

[8] D. Nion, K. Mokios, N. Sidiropoulos, and A. Potamianos, "Batch and adaptive PARAFAC-based blind separation of convolutive speech mixtures," *IEEE Trans. Audio, Speech Lang. Process.*, vol. 18, no. 6, pp. 1193–1207, Aug. 2010.

[9] L. De Lathauwer and B. De Moore, "On the blind separtion of non-circular source," in *Proc. 11th Eur. Signal Process. Conf.*, vol. 2. Sep. 2002, pp. 99–102.

[10] 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.

[11] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York, NY, USA: Wiley, 2001.

[12] S. Javidi, C. C. Took, and D. Mandic, "Fast independent component analysis algorithm for quaternion valued signals," *IEEE Trans. Neural Netw.*, vol. 22, no. 12, pp. 1967–1978, Dec. 2011.

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

[14] D. P. Mandic and V. S. L. Goh, *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models*. New York, NY, USA: Wiley, 2009.

[15] J. Navarro-Moreno, "ARMA prediction of widely linear systems by using the innovations algorithm," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3061–3068, Jul. 2008.

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

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

[18] R. A. Horn, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1990.

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

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

[21] P. Comon, "Independent component analysis, a new concept?" *Signal Process.*, vol. 36, no. 3, pp. 287–314, 1994.

[22] A. Bunse-Gerstner and W. B. Gragg, "Singular value decompositions of complex symmetric matrices," *J. Comput. Appl. Math.*, vol. 21, no. 1, pp. 41–54, 1988.

[23] A. Yeredor, "Performance analysis of the strong uncorrelating transformation in blind separation of complex-valued sources," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 478–483, Jan. 2012.

[24] 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.

[25] P. Georgiev, A. Cichocki, and H. Bakardjian, *Optimization Techniques for Independent Component Analysis With Applications to EEG Data, Quantitative Neuroscience: Models, Algorithms, Diagnostics, and Therapeutic Applications*. Norwell, MA, USA: Kluwer, 2004, ch. 3, pp. 53–68.

[26] S. Javidi, D. P. Mandic, C. C. Took, and A. Cichocki, "Kurtosis based blind source extraction of complex non-circular signals with application in EEG artifact removal in real time," *Frontiers Neurosci.*, vol. 5, pp. 1–18, Oct. 2011.

[27] T. Kreuz, A. Kraskov, R. Andrzejak, F. Mormann, K. Lehnertz, and P. Grassberger, "Measuring synchronization in coupled model systems: A comparison of different approaches," *Phys. D*, vol. 225, no. 1, pp. 29–42, 2007.

[28] S. L. Goh, M. Chen, D. H. Popovic, K. Aihara, D. Obradovic, and D. P. Mandic, "Complex valued forecasting of wind profile," *Renew. Energy*, vol. 31, no. 11, pp. 1733–1750, 2006.

[29] C. C. Took, G. Strbac, K. Aihara, and D. P. Mandic, "Quaternion-valued short term forecasting of three-dimensional wind and atmospheric parameters," *Renew. Energy*, vol. 36, no. 6, pp. 1754–1760, 2011.

Clive Cheong Took (SM'13) is a Lecturer with the Department of Computing, University of Surrey, Guildford, 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.

Dr. Took is an Associate Editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS. He has been involved in various conferences on machine learning, such as IEEE Workshop on Machine Learning for Signal Processing and the International Joint Conference on Neural Networks.

Scott C. Douglas (M'92–SM'98) received the B.S. (Hons.), M.S., and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, USA.

He is currently a Professor with the Department of Electrical Engineering, Southern Methodist University, Dallas, TX, USA. He has authored and co-authored two books, seven book chapters, and more than 190 articles in journals and conference proceedings. He is involved in the research and development activities, such as active noise control, blind deconvolution and source separation, wireless geolocation, microphone arrays, and geophysical signal processing and analysis.

Dr. Douglas has served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE SIGNAL PROCESSING LETTERS, and *Journal of Signal Processing Systems (Springer)*. He has served as a General Chair of the IEEE International Conference on Acoustics, Speech, and Signal Processing, the Chair of the Neural Networks for Signal Processing Technical Committee, and a Secretary of the Signal Processing Education Technical Committee, IEEE Signal Processing Society. He has played an integral role in developing and managing the Infinity Project, an effort among university faculty, high-tech industry, and civic educational leaders to bring an exciting and practical engineering curriculum to high school students both in the U.S. and internationally.

Danilo P. Mandic (F'13) is a Professor of Signal Processing with Imperial College London, London, U.K.

He has been involved in the nonlinear and multivariate adaptive signal processing and nonlinear dynamics. He has authored research monographs, such as *Recurrent Neural Networks for Prediction* and *Complex Valued Nonlinear Adaptive Filters*, and edited books, such as *Signal Processing Techniques for Knowledge Extraction and Information Fusion* and *Springer Hanbook of Bio- and Neuro-Informatics*. He has been a member of the IEEE Technical Committee on Signal Processing Theory and Methods, and an Associate Editor of the IEEE SIGNAL PROCESSING MAGAZINE, IEEE TRANSACTIONS ON SIGNAL PROCESSING, and IEEE TRANSACTIONS ON NEURAL NETWORKS. He was a recipient of Best Paper Awards in neurotechnology at the 2009 International Conference on Information and Computing Science and the 2010 International Symposium on Neural Networks, and the Eric Laithwaite Award from Imperial College on Ear-EEG.