Signal Processing 152 (2018) 350–362

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Widely linear complex partial least squares for latent subspace regression

Alexander E. Stott*, Sithan Kanna, Danilo P. Mandic

Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, UK

ARTICLE INFO

Article history:
Received 24 November 2017
Revised 24 May 2018
Accepted 18 June 2018
Available online 19 June 2018

Keywords:
Regression
Complex-valued signal processing
Multi-dimensional signal processing
Partial least squares
Regularisation
Component analysis

ABSTRACT

The method of partial least squares (PLS) has become a preferred tool for ill-posed linear estimation problems in the real domain, both in the regression and correlation analysis context. However, many modern applications involve complex-valued data (e.g. smart grid, sensor networks) and would benefit from corresponding well-posed latent variable regression analyses. To this end, we propose a PLS algorithm for physically meaningful latent subspace regression with complex-valued data. For rigour, this is achieved by taking into account full complex second-order augmented statistics to produce a robust widely linear estimator for general improper complex-valued data which may be highly correlated or colinear. The so-derived widely linear complex PLS (WL-CPLS) is shown to allow for effective joint latent variable decomposition of complex-valued data, while accounting for computational intractabilities in the calculation of a generalised inverse. This makes it possible to also determine the joint-subspace identified within the proposed algorithm, when applied to univariate outputs. The analysis is supported through both simulations on synthetic data and a real-world application of frequency estimation in unbalanced power grids. Finally, the ability of WL-CPLS to identify physically meaningful components is demonstrated through simultaneous complex covariance matrix diagonalisation.

1. Introduction

Developments in sensor technology and the increasing availability of computational power and computer memory have made it possible to obtain and process very large and often high-dimensional datasets. Such real-world datasets, typically have a rich structure which creates an opportunity for physically meaningful analysis, at the expense of computational tractability. For example, data from high-density sensor networks are frequently highly-correlated (colinear), which renders traditional regression methods ill-posed. It is therefore of particular interest to develop signal processing techniques that both account for these numerical issues and at the same time take advantage of any structure present in the data.

For many applications a widely accepted method to exploit structure in bivariate data is through complex-valued signal processing. The complex representation transforms complicated expressions in $\mathbb{R}^2$, such as rotations, into compact and easy to interpret forms in $\mathbb{C}$. This has led to advances in analysis of wind profiles [1], power systems [2,3], acoustics [4], and communications [5,6]. More recently, advances in so-called "augmented" statistics [7] have shown that a full second-order description of a complex-valued random variable, $\mathbf{z}$, includes both the pseudocovariance matrix, $\mathbf{P} = E[\mathbf{z}\mathbf{z}^{\mathrm{T}}]$, and the standard covariance matrix, $\mathbf{C} = E[\mathbf{z}\mathbf{z}^{\mathrm{H}}]$. Therefore, only the consideration of such "augmented" complex statistics can yield signal analysis tools which make use of features intrinsic to the complex domain, such as complex second-order noncircularity [8–10].

When it comes to determining the relationship between two sets of variables, linear regression is probably the most common data analysis method, whereby the variable $y \in \mathbb{R}$ is estimated through a linear combination, $\hat{y} = \mathbf{a}^{\mathrm{T}}\mathbf{x}$, of the independent variables, $\mathbf{x} \in \mathbb{R}^{m\times 1}$, by the vector of coefficients, $\mathbf{a} \in \mathbb{R}^{m\times 1}$. The vector $\mathbf{a}$ is calculated so as to minimise the mean square error (MSE) between the observation, y, and its prediction, $\hat{y}$. An extension to the complex domain has been developed by Picinbono and Chevalier [11], whereby the optimal estimate, $\hat{y}$, for complex-valued data, $y \in \mathbb{C}$, is given by $\hat{y} = \mathbf{h}^{\mathrm{H}}\mathbf{x} + \mathbf{g}^{\mathrm{H}}\mathbf{x}^*$, where the coefficient vectors, $\mathbf{h} \in \mathbb{C}^{m\times 1}$ and $\mathbf{g} \in \mathbb{C}^{m\times 1}$, describe the relation with the independent variables $\mathbf{x} \in \mathbb{C}^{m\times 1}$ and their conjugate $\mathbf{x}^*$. This so-called widely linear estimator is linear in both $\mathbf{x}$ and $\mathbf{x}^*$, and has found use in numerous applications including adaptive estimation of system frequency in distributed power systems [12].

A direct application of linear regression to dense sensor arrays has a very limited scope, as such solutions become ill-posed when

* Corresponding author.
E-mail address: alexander.stott10@imperial.ac.uk (A.E. Stott).

https://doi.org/10.1016/j.sigpro.2018.06.018
0165-1684/

the data are highly-correlated or colinear [13]. This can cause the covariance matrix, the inverse of which is inherent to regression methods, to have a large condition number or to become sub-rank which makes it difficult to compute its inverse. As a remedy, regularisation methods, such as Ridge-Regression [14], add a constant to the matrix diagonal to enforce well-posedness, however, this includes spurious information in the calculation. An alternative approach is to use the class of component analysis methods to factorise the original variables, which in addition to extracting the relevant information also provides a representation that is straightforwardly invertible. One such technique is principal component regression (PCR), which uses principal component analysis (PCA) to describe the original data matrix of regressors, **X**, through orthogonal latent components [15]. This allows for the separation of the desired information from noise related latent variables and admits a straightforward calculation of the generalised inverse of **X**, thus stabilising linear regression [16].

It is important to note that the so-obtained PCR solution creates a latent variable decomposition based only on the information in the independent variables, **X**, which means that it may contain erroneous information for use in the prediction of the dependent variables, **Y**. To this end, the partial least squares (PLS) regression algorithm integrates component analysis into the regression calculation. This is achieved by finding latent variables that explain only the joint input-output relation between the variables, **X** and **Y**, thus rendering the problem well-posed [13]. Real-world applications of the PLS are found in chemometrics and are emerging in signal processing [17–19].

The original real-valued PLS has been established as a robust data-analysis methodology [20]. The several types of PLS can be broadly split into two groups: *i)* those used for regression calculations (PLS1/2 in [20]) and *ii)* those used for dataset cross-covariance analysis (PLS Mode-A, PLS-SB in [20]). The PLS algorithms that aim to calculate a regression (NIPALS[1] and SIMPLS [21]) produce an orthogonal decomposition of the independent variable data block **X**. This leads to the most parsimonious model of the data for a regression calculation, because dimensionality reduction is at the heart of this approach. On the other hand, for dataset cross-covariance analysis it is often desirable that the latent variable decomposition is symmetric between the **X** and **Y** blocks, in which case the scores are not generally orthogonal. In the latter format, there are strong similarities to canonical correlation analysis (CCA), however, these type of methods are not usually used for prediction. The PLS framework therefore offers an in-depth data analysis tool through a combination of a linear regression and its latent variable decomposition.

It is crucial that the derived latent variables provide a useful and physically meaningful interpretation of the data, which can be further enhanced through constraints on the components such as non-negativity or sparseness [22]. Component analysis tools based on augmented complex statistics have recently been developed for complex-valued data and include the Strong Uncorrelating Transform (SUT) [23,24] and the Approximate Uncorrelating Transform (AUT) [25], while an extension of the PLS to complex-valued data has been proposed [26]. However, this version of PLS is structurally equivalent to the real-valued PLS-SB method in [20] and is presented from the viewpoint of dataset cross-covariance analysis. Such a decomposition therefore inherits the properties of the data-covariance analysis class of methods: the latent variables are not in general orthogonal and the relation between the **X** and **Y** block is symmetric. On the contrary, the proposed WL-CPLS algorithm is designed as a generic extension of the NIPALS algorithm for PLS-regression [13,27] to complex-valued data, taking into account full second-order augmented statistics. This generates the desirable property of the orthogonality of the obtained latent variables, unlike that proposed in [26], and naturally incorporates the calculation of a widely-linear regression. This important feature is shown to be useful beyond the field of regression for complex data and, in Section 4.2, its use is demonstrated to yield an uncorrelating transform. The analysis shows that the WL-CPLS algorithm caters for non-circular data without any restriction and in a generic way, unlike existing algorithms.

Our main technical contributions are threefold. We provide a method to calculate the widely linear regression coefficients akin to the real-domain PLS algorithm. Next, the properties of the WL-CPLS model residuals are determined and the algorithm convergence is proved for a univariate output. Finally, the WL-CPLS is verified on practical applications of complex-valued covariance matrix diagonalisation and for smart grid frequency estimation.

The paper is structured as follows. The background on PLS and widely linear regression is given in Section 2. We then derive the WL-CPLS algorithm in Section 3 based on a critical review of the PLS algorithm. The WL-CPLS algorithm is analysed in Section 4 and its application for simultaneous complex covariance matrix diagonalisation is introduced. The utility of WL-CPLS for complex-valued regression is illustrated through simulations on synthetic data in Section 5. The WL-CPLS is then applied to the real-world application of estimating the frequency of an unbalanced multi-nodal power grid in Section 6, confirming its capabilities over existing techniques.

Boldfaced capital letters denote matrices, **A**, lower case boldfaced letters vectors, **a**, and lightfaced italic letters scalars, *a*. The superscripts $(\cdot)^+$, $(\cdot)^{\mathsf{T}}$, $(\cdot)^{\mathsf{H}}$ and $(\cdot)^*$ denote respectively the generalised inverse, transpose, Hermitian transpose and conjugate operators respectively. The operator $\mathrm{Eig}_{\max}\{\cdot\}$ returns the eigenvector corresponding to the largest eigenvalue of the matrix in the argument.

2. Background and review

2.1. Partial least squares regression

Consider the linear regression problem of predicting a matrix of p dependent variables, $\mathbf{Y} \in \mathbb{R}^{N\times p}$, from a matrix of m independent variables, $\mathbf{X} \in \mathbb{R}^{N\times m}$, through a matrix of coefficients, $\mathbf{B} \in \mathbb{R}^{m\times p}$, described by

$$\hat{\mathbf{Y}} = \mathbf{X}\mathbf{B}, \tag{1}$$

where $\hat{\mathbf{Y}}$ denotes the estimate of **Y** and N denotes the number of observations. The general solution for the regression coefficients, **B**, has the form

$$\mathbf{B} = \mathbf{X}^{+}\mathbf{Y}, \tag{2}$$

which requires the calculation of the generalised matrix inverse $\mathbf{X}^+$ [28]. The ordinary least squares solution is then given by

$$\mathbf{X}^{+} = (\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathsf{T}}. \tag{3}$$

If the variables in **X** (its columns) are highly-correlated or colinear, then **X** is sub-rank, which is prohibitive to the calculation of the inverse of the matrix $\mathbf{X}^{\mathsf{T}}\mathbf{X}$. To counteract this issue, the method of Partial Least Squares (PLS) produces a latent variable decomposition of the matrix **X** from which a generalised inverse is straightforwardly calculated [13,17,29]. The advantage of PLS compared to other component analysis regression methods (*e.g.* PCR) is that the latent components are selected so as to explain the joint dynamics (shared latent variables) between **X** and **Y**, while the PCR solution produces a decomposition of **X** without consideration of **Y**, thus

[1] Throughout the paper we refer to the NIPALS algorithm for the PLS-regression method known as PLS1/2 in [20]

yielding a less parsimonious, and typically less physically meaningful, model than PLS.

The PLS decomposition is performed through the factorisations [17]

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}}, \tag{4}$$

$$\mathbf{Y} = \mathbf{T}\mathbf{C}^{\mathrm{T}}, \tag{5}$$

$$\mathbf{Y} = \mathbf{U}\mathbf{Q}^{\mathrm{T}}, \tag{6}$$

where $\mathbf{T} \in \mathbb{R}^{N\times r}$ is the matrix which comprises r latent variables in $\mathbf{X}$ with the loadings given by the matrix $\mathbf{P} \in \mathbb{R}^{m\times r}$ while $\mathbf{U} \in \mathbb{R}^{N\times r}$ is the corresponding matrix of r latent variables in $\mathbf{Y}$ with loadings given by the matrix $\mathbf{Q} \in \mathbb{R}^{p\times r}$. The matrix $\mathbf{C} \in \mathbb{R}^{p\times r}$ describes the "inner-relation" of PLS, that is, the regression between $\mathbf{Y}$ and $\mathbf{T}$, and indicates the extent to which the latent variables, $\mathbf{T}$, are good descriptors of both $\mathbf{X}$ and $\mathbf{Y}$. These relations demonstrate the utility of the PLS solution as both a regression and a component analysis tool, which offers new data analysis opportunities. For example, Abdi [30] shows that plotting the latent variables against one another (through biplots) reveals information about different groups within data.

Algorithm 1 The NIPALS algorithm for PLS.

Initialise: $\mathbf{X}_1 = \mathbf{X}$, $\mathbf{Y}_1 = \mathbf{Y}$
for $i = 1, \ldots, r$ **do**
  $\mathbf{w}_i = \mathrm{Eig}_{\max}\{\mathbf{X}_i^{\mathrm{T}}\mathbf{Y}_i\mathbf{Y}_i^{\mathrm{T}}\mathbf{X}_i\}$
  $\mathbf{t}_i = \mathbf{X}_i\mathbf{w}_i$
  $\mathbf{c}_i = \mathbf{Y}_i^{\mathrm{T}}\mathbf{t}_i / \mathbf{t}_i^{\mathrm{T}}\mathbf{t}_i$
  $\mathbf{u}_i = \mathbf{Y}_i\mathbf{c}_i$
  $\mathbf{p}_i = \mathbf{X}_i^{\mathrm{T}}\mathbf{t}_i / \mathbf{t}_u^{\mathrm{T}}\mathbf{t}_i$
  $\mathbf{q}_i = \mathbf{Y}_i^{\mathrm{T}}\mathbf{u}_i / \mathbf{u}_i^{\mathrm{T}}\mathbf{u}_i$
  $\mathbf{X}_{i+1} = \mathbf{X}_i - \mathbf{t}_i\mathbf{p}_i^{\mathrm{T}}$, $\mathbf{Y}_{i+1} = \mathbf{Y}_i - \mathbf{t}_i\mathbf{c}_i^{\mathrm{T}}$
  Store $\mathbf{t}_i$, $\mathbf{u}_i$, $\mathbf{p}_i$, $\mathbf{q}_i$, $\mathbf{c}_i$ and $\mathbf{w}_i$
end for

The multivariate PLS solution is typically calculated by the iterative NIPALS algorithm[2] [17], outlined in Algorithm 1. Each iteration calculates and stores a rank-1 approximation of $\mathbf{X}$ and $\mathbf{Y}$, and on completion of the algorithm, this results in the relations (4)–(6). In order to calculate these rank-1 approximations, a score vector $\mathbf{t}$ is found as the component in $\mathbf{X}$ which has the maximum cross-covariance with $\mathbf{Y}$. This is achieved by finding a weight vector, $\mathbf{w}$, that yields the maximal projection of the cross-covariance matrix $\mathbf{X}^{\mathrm{T}}\mathbf{Y}$. In other words, the solution, $\mathbf{w}$, arises from the optimisation problem

$$\mathbf{w} = \underset{||\mathbf{w}||=1}{\arg\max}\, ||\mathbf{w}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{Y}||_2^2, \tag{7}$$

and is given as the eigenvector which corresponds to the largest eigenvalue of the matrix $\mathbf{X}^{\mathrm{T}}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{X}$. This matrix is symmetric and positive-semidefinite, hence its eigenvectors are orthogonal and represent a valid basis for the cross-covariance between $\mathbf{X}$ and $\mathbf{Y}$ [20]. The "score" of this vector is given by

$$\mathbf{t} = \mathbf{X}\mathbf{w}, \tag{8}$$

and represents a latent variable in $\mathbf{X}$, which is optionally normalised to $\mathbf{t}^{\mathrm{T}}\mathbf{t} = 1$ so that $\mathbf{T}^{\mathrm{T}}\mathbf{T} = \mathbf{I}$. The NIPALS algorithm next calculates the regressions of $\mathbf{X}$ and $\mathbf{Y}$ to the computed $\mathbf{t}$, resulting in the rank-1 approximations $\tilde{\mathbf{X}} = \mathbf{t}\mathbf{p}^{\mathrm{T}}$ and $\tilde{\mathbf{Y}} = \mathbf{t}\mathbf{c}^{\mathrm{T}}$. The score vector in $\mathbf{Y}$ corresponding to $\mathbf{t}$ is $\mathbf{u} = \mathbf{Y}\mathbf{c}$ and the rank-1 approximation $\tilde{\mathbf{Y}} = \mathbf{u}\mathbf{q}^{\mathrm{T}}$ is then calculated. Before proceeding to the next iteration, the data matrices $\mathbf{X}$ and $\mathbf{Y}$ are "deflated", that is, the rank-1 contribution of the current score $\mathbf{t}$ is removed from $\mathbf{X}$ and $\mathbf{Y}$, which makes the components calculated in the following iteration orthogonal to the already extracted components. The vectors $\mathbf{w}$, $\mathbf{t}$, $\mathbf{p}$, $\mathbf{c}$, $\mathbf{u}$ and $\mathbf{q}$ are stored in the columns of their respective matrices and form the PLS decompositions in (4)–(6). The algorithm is iterated until all significant components are found, with the stopping criterion typically based on a negligible contribution of a new component to the variance in the prediction of $\mathbf{Y}$.

In summary, the NIPALS PLS-regression algorithm provides an approximation of $\mathbf{X}$, denoted by $\tilde{\mathbf{X}}$, in the form

$$\tilde{\mathbf{X}} = \mathbf{T}(\mathbf{P}^{\mathrm{T}}\mathbf{W})\mathbf{W}^{\mathrm{T}}, \tag{9}$$

which is based on a linear combination of columns (given by each vector $\mathbf{w}$) from a repeatedly deflated $\mathbf{X}$, accounted for by the matrix $(\mathbf{P}^{\mathrm{T}}\mathbf{W})$ which has an upper-triangular form [29]. The generalised inverse of the matrix $\tilde{\mathbf{X}}$ is now straightforward to calculate due to its structure, as it is an orthogonal matrix multiplied by an upper-triangular matrix multiplied by orthogonal matrices, which are all straightforwardly invertible. Hence, the generalised inverse is calculated as $\tilde{\mathbf{X}}^{+} = \mathbf{W}(\mathbf{P}^{\mathrm{T}}\mathbf{W})^{-1}\mathbf{T}^{\mathrm{T}}$ to give the PLS solution $\hat{\mathbf{B}}_{\mathrm{PLS}} = \tilde{\mathbf{X}}^{+}\mathbf{Y}$.

Remark 1. The PLS-regression calculates a joint decomposition (in a second-order statistical sense) of variables $\mathbf{X}$ and $\mathbf{Y}$ in order to produce a solution of a regularised generalised inverse problem.

2.2. Widely linear regression

Prior to introducing a complex domain extension of the PLS algorithm, we shall provide a brief review of complex-valued regression. The aim is to find an estimate, $\hat{\mathbf{Y}}$, of dependent variables, $\mathbf{Y} \in \mathbb{C}^{N\times p}$, from independent variables $\mathbf{X} \in \mathbb{C}^{N\times m}$ through the minimisation of the mean square error (MSE)

$$MSE = Tr\{E[(\mathbf{Y} - \hat{\mathbf{Y}})^{\mathrm{H}}(\mathbf{Y} - \hat{\mathbf{Y}})]\}. \tag{10}$$

When $\mathbf{X}$ and $\mathbf{Y}$ are complex-valued, the general solution should account for second-order noncircularity (improperness) of the data, this is achieved through a widely linear form [11]

$$\hat{\mathbf{Y}} = \mathbf{X}\mathbf{H} + \mathbf{X}^{*}\mathbf{G}, \tag{11}$$

with coefficient matrices $\mathbf{H} \in \mathbb{C}^{m\times p}$ and $\mathbf{G} \in \mathbb{C}^{m\times p}$. Such an optimal estimator of $\mathbf{Y}$ is, in general, linear in terms of both $\mathbf{X}$ and $\mathbf{X}^{*}$, in contrast with the strictly linear result in (1). Observe that the widely linear regression coefficient matrices, $\mathbf{H}$ and $\mathbf{G}$, cannot be calculated through a generalised inverse of only $\mathbf{X}$. In order to derive the solution, we observe that the residual $(\mathbf{Y} - \hat{\mathbf{Y}})$ is orthogonal to both $\mathbf{X}$ and $\mathbf{X}^{*}$, which yields the expectations

$$E[\hat{\mathbf{Y}}^{\mathrm{H}}\mathbf{X}] = E[\mathbf{Y}^{\mathrm{H}}\mathbf{X}], \qquad E[\hat{\mathbf{Y}}^{\mathrm{H}}\mathbf{X}^{*}] = E[\mathbf{Y}^{\mathrm{H}}\mathbf{X}^{*}], \tag{12}$$

leading to

$$\mathbf{H}^{\mathrm{H}}\mathbf{C} + \mathbf{G}^{\mathrm{H}}\mathbf{P} = \mathbf{R}, \qquad \mathbf{H}^{\mathrm{H}}\mathbf{P}^{*} + \mathbf{G}^{\mathrm{H}}\mathbf{C}^{*} = \mathbf{S}^{*}, \tag{13}$$

where $\mathbf{R} = \mathbf{Y}^{\mathrm{H}}\mathbf{X}$, $\mathbf{S} = \mathbf{Y}^{\mathrm{T}}\mathbf{X}$, $\mathbf{C} = \mathbf{X}^{\mathrm{H}}\mathbf{X}$ and $\mathbf{P} = \mathbf{X}^{\mathrm{T}}\mathbf{X}$ are respectively the empirical cross-covariance, cross-pseudocovariance, covariance and pseudocovariance matrices. A simultaneous solution of (13) yields the widely linear regression coefficients given by

$$\begin{aligned}\mathbf{H} &= [\mathbf{C} - \mathbf{P}^{*}\mathbf{C}^{*-1}\mathbf{P}]^{-1}[\mathbf{R}^{\mathrm{H}} - \mathbf{P}^{*}\mathbf{C}^{*-1}\mathbf{S}^{\mathrm{T}}]\\ \mathbf{G} &= [\mathbf{C}^{*} - \mathbf{P}\mathbf{C}^{-1}\mathbf{P}^{*}]^{-1}[\mathbf{S}^{\mathrm{T}} - \mathbf{P}\mathbf{C}^{-1}\mathbf{R}^{\mathrm{H}}].\end{aligned} \tag{14}$$

The augmented forms, $\underline{\mathbf{X}} = [\mathbf{X}, \mathbf{X}^{*}]$ and $\underline{\mathbf{Y}} = [\mathbf{Y}, \mathbf{Y}^{*}]$, provide a compact representation for the widely linear regression in (11) as

$$\underline{\mathbf{Y}} = \underline{\mathbf{X}}\mathbf{B}, \tag{15}$$

[2] Other multivariate PLS algorithms do exist, such as the SIMPLS [21].

where

$$\mathbf{B} = \begin{bmatrix} \mathbf{H} & \mathbf{G}^* \\ \mathbf{G} & \mathbf{H}^* \end{bmatrix}.$$

The widely linear regression coefficients in $\mathbf{B}$ are consequently obtained as $\mathbf{B} = \underline{\mathbf{X}}^+\underline{\mathbf{Y}}$. The generalised inverse in this context is given $\underline{\mathbf{X}}^+ = (\underline{\mathbf{X}}^{\mathrm{H}}\underline{\mathbf{X}})^{-1}\underline{\mathbf{X}}^{\mathrm{H}}$ (akin to the linear least squares solution in (3)) which leads to

$$\underline{\mathbf{X}}^+ = \begin{bmatrix} \mathbf{C} & \mathbf{P}^* \\ \mathbf{P} & \mathbf{C}^* \end{bmatrix}\underline{\mathbf{X}}^{\mathrm{H}}.$$

A direct application of block matrix inversion [31] yields a derivation of the widely-linear estimation in augmented form which requires the same calculations for the regression coefficient matrices $\mathbf{G}$ and $\mathbf{H}$ as those derived in (14). This highlights that the augmented form in (15) allows a representation of complex-valued regression in the same form as a standard real-valued linear regression in (1).

Remark 2. Widely linear regression is solved through a generalised inverse of the augmented complex matrix $\underline{\mathbf{X}}$, which is a generic extension of standard regression based on a generalised inverse of the matrix $\mathbf{X}$. This is equivalent to the solution in (14).

2.3. Duality between complex-valued processing and processing in $\mathbb{R}^2$

Complex valued data can be equivalently represented and processed in $\mathbb{R}^2$ [7,32]. To illustrate this duality, consider a matrix, $\mathbf{X} \in \mathbb{C}^{N\times m}$, which can be written in terms of its real and imaginary parts as $\mathbf{X} = (\mathbf{X}_R + j\mathbf{X}_I)$. The individual parts, $\mathbf{X}_R \in \mathbb{R}^{N\times m}$ and $\mathbf{X}_I \in \mathbb{R}^{N\times m}$, can be represented as

$$\mathbf{X}_{Re} = [2\mathbf{X}_R, 2\mathbf{X}_I], \tag{16}$$

where $\mathbf{X}_{Re} \in \mathbb{R}^{N\times 2m}$. The isomorphism between this representation in $\mathbb{R}^{N\times 2m}$ and the augmented form of complex-valued data, given by

$$\underline{\mathbf{X}} = [\mathbf{X}, \mathbf{X}^*],$$

is described by the transform matrix, $\mathbf{\Gamma}_m$, given by

$$\mathbf{\Gamma}_m = \begin{bmatrix} \mathbf{I}_m & -j\mathbf{I}_m \\ \mathbf{I}_m & j\mathbf{I}_m \end{bmatrix}, \tag{17}$$

where $\mathbf{I}_m \in \mathbb{R}^{m\times m}$ is the identity matrix of appropriate dimensions. The transform matrix, $\mathbf{\Gamma}_m$, is unitary with a scale factor of 2 so that $\mathbf{\Gamma}_m^{\mathrm{H}}\mathbf{\Gamma}_m = 2\mathbf{I}_{2m}$ and the mapping in (16) then becomes

$$[2\mathbf{X}_R, 2\mathbf{X}_I] = [\mathbf{X}, \mathbf{X}^*]\mathbf{\Gamma}_m.$$

2.4. The notion of circularity

Widely linear regression is optimal for second-order non-circular[3] (improper) data, which is a more general case than circular rotation invariant (proper) data [7,33]. The concept of circularity admits an insightful geometric interpretation of a Complex Random Variable (CRV) [8]. The degree of circularity can be described by the circularity coefficient, the absolute value of the ratio of pseudocovariance, $p = E[z^2]$, to covariance, $c = E[z^*z]$, given by [34]

$$\rho = \frac{|p|}{c}.$$

For a proper variable $\rho = 0$, while for a maximally non-circular variable (*i.e.* a real number represented as complex-valued) $\rho = 1$. Moreover, the circularity quotient (the fraction $\frac{p}{c}$) reveals whether the non-circularity arose due to the power imbalance or correlation between the data channels. A CRV $z = z_r + jz_i$ has pseudocovariance $p = E[z_r^2 - z_i^2 + jz_rz_i]$ and covariance $c = E[z_r^2 + z_i^2]$, and therefore, the real part of their ratio quantifies power imbalance between z_r and z_i whereas the imaginary part quantifies their correlation [7].

[3] Note that propriety is a second-order property whereas circularity accounts for all statistical moments.

3. Widely linear complex partial least squares

Since the fundamental aim of the PLS algorithm is to provide a linear regression estimator, an extension for general complex-valued data should be based on a widely linear model of the form $\mathbf{Y} = \mathbf{XH} + \mathbf{X}^*\mathbf{G}$. The solution is then created through a latent variable decomposition of $\mathbf{X}$ and based on full joint second-order complex statistics. The so-obtained latent variable decomposition, owing to its structure, will admit a straightforward calculation of the generalised inverse of the augmented data matrix $\underline{\mathbf{X}}$, and in doing so, yield a regularised widely linear regression solution. In general, the requirements for the widely linear complex PLS (WL-CPLS) are to:

- **R1** Create a joint latent variable decomposition of the data matrices, $\mathbf{X}$ and $\mathbf{Y}$, that describes the complete complex second-order statistics;
- **R2** Account for a widely linear relationship between the $\mathbf{X}$ and $\mathbf{Y}$ data blocks;
- **R3** Obtain approximations $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$ that admit a straightforward and tractable computation of a widely linear regression $\mathbf{Y} = \tilde{\mathbf{X}}\mathbf{H} + \tilde{\mathbf{X}}^*\mathbf{G}$.

3.1. A complex-valued PLS algorithm

The proposed complex PLS algorithm for widely linear regression (WL-CPLS) is next derived as an extension of Wold's real-valued NIPALS PLS-regression algorithm [17], outlined in Algorithm 1. To extend NIPALS for a general complex-valued input $\mathbf{X} \in \mathbb{C}^{N\times m}$ and output $\mathbf{Y} \in \mathbb{C}^{N\times p}$, we must find the components in $\mathbf{X}$ that represent the maximal joint second-order information with $\mathbf{Y}$, in the form of vectors, $\mathbf{w}$, which give "score" vectors $\mathbf{t} = \mathbf{Xw}$. The augmented representation is then given by

$$[\mathbf{t}, \mathbf{t}^*] = [\mathbf{X}, \mathbf{X}^*]\begin{bmatrix} \mathbf{w} & \mathbf{0} \\ \mathbf{0} & \mathbf{w}^* \end{bmatrix}.$$

This structure is required as the vectors $\mathbf{w}$ must form an orthogonal basis for the inputs $\mathbf{X}$. To find the vector $\mathbf{w}$ that meets this criterion and produces the required structure we shall employ the isomorphism between $\mathbb{R}^2$ and $\mathbb{C}$, given in (16), to transform $\underline{\mathbf{X}}$ and $\underline{\mathbf{Y}}$ as

$$\mathbf{X}_{Re} = \underline{\mathbf{X}}\mathbf{\Gamma}_m, \qquad \mathbf{Y}_{Re} = \underline{\mathbf{Y}}\mathbf{\Gamma}_n,$$

where $\mathbf{\Gamma}_m$, and $\mathbf{\Gamma}_n$ are defined in (17). The matrices $\mathbf{X}_R \in \mathbb{R}^{N\times m}$, $\mathbf{Y}_R \in \mathbb{R}^{N\times p}$, $\mathbf{X}_I \in \mathbb{R}^{N\times m}$ and $\mathbf{Y}_I \in \mathbb{R}^{N\times p}$ denote the real and imaginary parts of $\mathbf{X}$ and $\mathbf{Y}$ respectively. The cross-covariance criterion for the real-valued PLS in (7) is then extended to the complex domain as

$$\mathbf{w}_{Re} = \underset{||\mathbf{w}_{Re}||=1}{\arg\max} ||\mathbf{w}_{Re}^{\mathrm{T}}\mathbf{X}_{Re}^{\mathrm{T}}\mathbf{Y}_{Re}||_2^2. \tag{18}$$

The solution, $\mathbf{w}_{Re} \in \mathbb{R}^{2m}$, is found as the eigenvector of $\mathbf{X}_{Re}^{\mathrm{T}}\mathbf{Y}_{Re}\mathbf{Y}_{Re}^{\mathrm{T}}\mathbf{X}_{Re}$ which corresponds to its largest eigenvalue. The resulting vector, $\mathbf{w}_{Re}$, is of the form $[\mathbf{w}_R^{\mathrm{T}}, -\mathbf{w}_I^{\mathrm{T}}]^{\mathrm{T}}$ and is transformed back to the complex domain through

$$[\mathbf{w}^{\mathrm{T}}, \mathbf{w}^{\mathrm{H}}]^{\mathrm{T}} = \mathbf{\Gamma}_m\mathbf{w}_{Re}. \tag{19}$$

Notice that the computation of the vector, $\mathbf{w}$, in the real domain before converting it back to the complex domain does not affect any benefit gained by the complex representation.

We have seen that the augmented form of a complex-valued variable makes it possible to capture the full second-order statistics for complex-valued data. However, in this case, an SVD of the augmented cross-covariance matrix does not provide the orthogonal basis for the $\mathbf{X}$ block in the structure required for the WL-CPLS algorithm, and the result would not be usable for the NIPALS extension. On the other hand, the proposed calculation of $\mathbf{w}$ does fulfil this structural requirement. The problem of an appropriate structure for a complex-valued PLS is also addressed in [26], where the computation is performed in the real-domain before transforming back to the complex-domain. This process, therefore, is implicitly complex-valued and is not equivalent to an application of the NIPALS algorithm on complex data transformed to the real domain.

The so-obtained complex-valued vector $\mathbf{w} \in \mathbb{C}^{m\times 1}$ caters for full joint second-order statistics between the input $\mathbf{X}$ and the output $\mathbf{Y}$. The corresponding component vector, $\mathbf{t}$, is then calculated as

$$\mathbf{t} = \mathbf{X}\mathbf{w}.$$

Remark 3. The latent variable, $\mathbf{t}$, is obtained through a consideration of both cross-covariance and cross-pseudocovariance between the input $\mathbf{X}$ and output $\mathbf{Y}$, this fulfils Requirement **R1** for a WL-CPLS.

The joint approximations of the input $\mathbf{X}$ and the output $\mathbf{Y}$ are produced by regressing the so-obtained $\mathbf{t}$ onto $\mathbf{X}$ and $\mathbf{Y}$. For complex-valued data, these regressions now become widely linear in terms of $\mathbf{t}$ and $\mathbf{t}^*$, which results in the approximations[4] $\mathbf{X} = \mathbf{t}\mathbf{p}_1^{\mathrm{H}} + \mathbf{t}^*\mathbf{p}_2^{\mathrm{H}}$ and $\mathbf{Y} = \mathbf{t}\mathbf{c}_1^{\mathrm{H}} + \mathbf{t}^*\mathbf{c}_2^{\mathrm{H}}$, calculated in an augmented form

$$\underline{\mathbf{p}} = (\underline{\mathbf{t}}^{+}\underline{\mathbf{X}})^{\mathrm{H}}, \qquad \underline{\mathbf{c}} = (\underline{\mathbf{t}}^{+}\underline{\mathbf{Y}})^{\mathrm{H}}, \tag{20}$$

where $\underline{\mathbf{p}} = [\mathbf{p}_1, \mathbf{p}_2]$, $\underline{\mathbf{t}} = [\mathbf{t}, \mathbf{t}^*]$ and $\underline{\mathbf{c}} = [\mathbf{c}_1, \mathbf{c}_2]$.

The latent variables in $\mathbf{Y}$, denoted by the vectors $\mathbf{u}$ that correspond to the latent variables $\mathbf{t}$ in $\mathbf{X}$, are obtained as

$$\mathbf{u} = \mathbf{Y}\mathbf{c}_1 + \mathbf{Y}^*\mathbf{c}_2. \tag{21}$$

Similarly, the regression of $\mathbf{u}$ to $\mathbf{Y}$ is an approximation $\mathbf{Y} = \mathbf{u}\mathbf{q}_1^{\mathrm{H}} + \mathbf{u}^*\mathbf{q}_2^{\mathrm{H}}$ which can be represented in an augmented form as

$$\underline{\mathbf{q}} = (\underline{\mathbf{u}}^{+}\underline{\mathbf{Y}})^{\mathrm{H}}$$

where $\underline{\mathbf{q}} = [\mathbf{q}_1, \mathbf{q}_2]$ and $\underline{\mathbf{u}} = [\mathbf{u}, \mathbf{u}^*]$. The above steps describe the WL-CPLS decomposition for the latent variable $\mathbf{t}$ in each PLS iteration. Before the next latent variable can be calculated, the impact of the currently extracted component, $\mathbf{t}_n$, must be removed from the data matrices $\mathbf{X}$ and $\mathbf{Y}$. This is achieved by deflating (subtracting) the respective approximations (23) and (24) from $\mathbf{X}$ and $\mathbf{Y}$ to give

$$\mathbf{X}_{i+1} = \mathbf{X}_i - \mathbf{t}_i\mathbf{p}_{1,i}^{\mathrm{H}} - \mathbf{t}_i^*\mathbf{p}_{2,i}^{\mathrm{H}}, \qquad \mathbf{Y}_{i+1} = \mathbf{Y}_i - \mathbf{t}_i\mathbf{c}_{1,i}^{\mathrm{H}} - \mathbf{t}_i^*\mathbf{c}_{2,i}^{\mathrm{H}}, \tag{22}$$

or in an augmented form

$$\underline{\mathbf{X}}_{i+1} = \underline{\mathbf{X}}_i - \underline{\mathbf{t}}_i\underline{\mathbf{p}}_i^{\mathrm{H}}, \qquad \underline{\mathbf{Y}}_{i+1} = \underline{\mathbf{Y}}_i - \underline{\mathbf{t}}_i\underline{\mathbf{c}}_i^{\mathrm{H}},$$

where the subscript i indicates the iteration number with $\mathbf{X}_1 = \mathbf{X}$ and $\mathbf{Y}_1 = \mathbf{Y}$. In the next iteration, the matrices $\mathbf{X}_{i+1}$ and $\mathbf{Y}_{i+1}$ are used in place of $\mathbf{X}_i$ and $\mathbf{Y}_i$, which ensures that the new extracted score (latent variable) will be orthogonal to the component extracted in the previous iterations. As such, the information expressed by each extracted component is accounted for separately, that is

$$\mathbf{X} = \mathbf{T}\mathbf{P}_1^{\mathrm{H}} + \mathbf{T}^*\mathbf{P}_2^{\mathrm{H}}, \tag{23}$$

$$\mathbf{Y} = \mathbf{T}\mathbf{C}_1^{\mathrm{H}} + \mathbf{T}^*\mathbf{C}_2^{\mathrm{H}}, \tag{24}$$

[4] This is a rank-2 approximation unless a strictly linear regression of the component $\mathbf{t}$ to $\mathbf{X}$ and $\mathbf{Y}$ is sufficient.

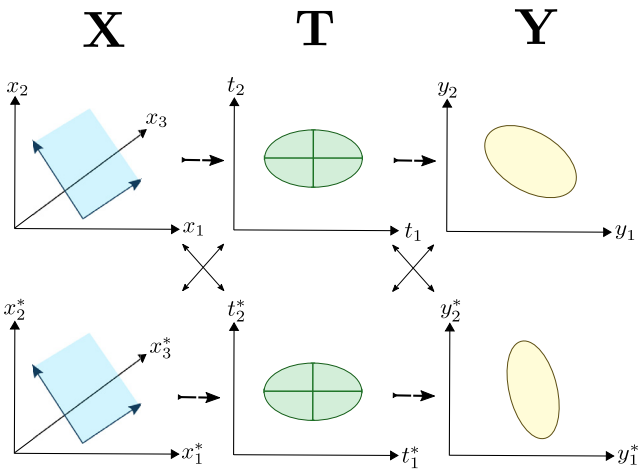

Fig. 1. Geometric view of WL-CPLS regression.

which gives the WL-CPLS decomposition corresponding to (6), in the form

$$\mathbf{Y} = \mathbf{U}\mathbf{Q}_1^{\mathrm{H}} + \mathbf{U}^*\mathbf{Q}_2^{\mathrm{H}} \tag{25}$$

The complete WL-CPLS algorithm is outlined in Algorithm 2 and is iterated until r latent components have been found. The required number of components is determined based on a stopping criterion, described in Section 5.2.1.

Algorithm 2 The NIPALS algorithm for widely linear complex PLS (WL-CPLS).

Initialise: $\underline{\mathbf{X}}_1 = [\mathbf{X}, \mathbf{X}^*]$, $\underline{\mathbf{Y}}_1 = [\mathbf{Y}, \mathbf{Y}^*]$
for $i = 1, \ldots, r$ **do**
  $\mathbf{X}_{i,Re} = \underline{\mathbf{X}}_i\mathbf{\Gamma}_m$, and $\mathbf{Y}_{i,Re} = \underline{\mathbf{Y}}_i\mathbf{\Gamma}_n$
  $\mathbf{w}_{i,Re} = \mathrm{Eig}_{\max}\{\mathbf{X}_{i,Re}^{\mathrm{T}}\mathbf{Y}_{i,Re}\mathbf{Y}_{i,Re}^{\mathrm{T}}\mathbf{X}_{i,Re}\}$
  $[\mathbf{w}_i^{\mathrm{T}}, \mathbf{w}_i^{\mathrm{H}}]^{\mathrm{T}} = \mathbf{\Gamma}_m\mathbf{w}_{i,Re}$
  $\mathbf{t}_i = \mathbf{X}_i\mathbf{w}_i$, $\underline{\mathbf{t}}_i = [\mathbf{t}_i, \mathbf{t}_i^*]$
  $\underline{\mathbf{c}}_i = (\underline{\mathbf{t}}_i^{+}\ \underline{\mathbf{Y}}_i)^{\mathrm{H}}$
  $\underline{\mathbf{p}}_i = (\underline{\mathbf{t}}_i^{+}\ \underline{\mathbf{X}}_i)^{\mathrm{H}}$
  $\mathbf{u}_i = \mathbf{Y}_i\mathbf{c}_{1,i} + \mathbf{Y}_i^*\mathbf{c}_{2,i}$, $\underline{\mathbf{u}}_i = [\mathbf{u}_i, \mathbf{u}_i^*]$
  $\underline{\mathbf{q}}_i = (\underline{\mathbf{u}}_i^{+}\ \underline{\mathbf{Y}}_i)^{\mathrm{H}}$
  $\underline{\mathbf{X}}_{i+1} = \underline{\mathbf{X}}_i - \underline{\mathbf{t}}_i\underline{\mathbf{p}}_i^{\mathrm{H}}$, $\underline{\mathbf{Y}}_{i+1} = \underline{\mathbf{Y}}_i - \underline{\mathbf{t}}_i\underline{\mathbf{c}}_i^{\mathrm{H}}$
  Store $\mathbf{t}_i$, $\mathbf{u}_i$, $\mathbf{p}_{1,i}$, $\mathbf{p}_{2,i}$, $\mathbf{q}_{1,i}$, $\mathbf{q}_{2,i}$, $\mathbf{c}_{1,i}$, $\mathbf{c}_{2,i}$ and $\mathbf{w}_i$
end for

Remark 4. For a regularised regression application it is desirable to identify a latent subspace within the regressors, $\mathbf{X}$, that is used to predict the dependent variables, $\mathbf{Y}$. The matrix $\mathbf{W}$ within the proposed WL-CPLS in Algorithm 2 provides such a basis, which is then used to sequentially calculate the score vectors in the matrix $\mathbf{T}$. Owing to the optimisation process in (18), this subspace is chosen so as to contain the full joint complex second-order information. The widely linear regression of the scores, $\mathbf{T}$, to the data matrices $\mathbf{X}$ and $\mathbf{Y}$ described in (23) and (24) then give the WL-CPLS approximations of the identified subspace and account for the required degrees of freedom needed in complex-valued estimation. As such, the WL-CPLS algorithm is a generic extension of the real-valued NIPALS algorithm described in Section 2.1 for PLS-regression.

Remark 5. The complex-valued WL-CPLS takes into account the widely linear relationship between the $\mathbf{X}$ and $\mathbf{Y}$ components and thus satisfies Requirement **R2** for a WL-CPLS.

Fig. 1 shows a geometric interpretation of WL-CPLS. In contrast to real-valued PLS, the scores $\mathbf{T}$ in WL-CPLS define a regression

subspace in both $\mathbf{X}$ and $\mathbf{X}^*$. This subspace is then used to provide a widely linear regression to predict dependent variables $\mathbf{Y}$.

Remark 6. An important aspect of the WL-CPLS algorithm is that the scores $\mathbf{T}$ are fully uncorrelated in terms of augmented complex-valued second order statistics, that is, the score covariance and pseudocovariance matrices, $\mathbf{T}^{\mathrm{H}}\mathbf{T}$ and $\mathbf{T}^{\mathrm{T}}\mathbf{T}$, are diagonal. For normalised score vectors, $\mathbf{t}$, their covariance matrix becomes $\mathbf{T}^{\mathrm{H}}\mathbf{T} = \mathbf{I}$. This, in turn, means that the diagonal elements of the score pseudocovariance matrix, $\mathbf{T}^{\mathrm{T}}\mathbf{T}$, readily give the circularity quotient (as defined in Section 2.4), thus further highlighting the physical insight of WL-CPLS. This would not be available using the NIPALS algorithm on complex data cast into the real domain.

3.2. Alternative WL-CPLS algorithms

The WL-CPLS described in Algorithm 2 can serve as a basis for a class of WL-CPLS results, as the vector, $\mathbf{w}$, chosen to reflect the maximum cross-covariance and cross-pseudocovariance between the input $\mathbf{X}$ and the output $\mathbf{Y}$, can be replaced by a different criterion. For example, we can choose $\mathbf{w}$ to reflect only the maximum cross-covariance between the input $\mathbf{X}$ and the output $\mathbf{Y}$. This is achieved by the following optimisation

$$\mathbf{w} = \underset{||\mathbf{w}||=1}{arg\,max}\, ||\mathbf{w}^{\mathrm{H}}\mathbf{X}^{\mathrm{H}}\mathbf{Y}||_2^2.$$

The solution, $\mathbf{w}$, is the largest eigenvector of the matrix, $\mathbf{X}^{\mathrm{H}}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}\mathbf{X}$. This alternative WL-CPLS is summarised in Algorithm 3. It is important to note, however, that this alternative vector, $\mathbf{w}$, is not capable of detecting a latent variable, $\mathbf{t}$, that represents only cross-psuedocovariance between $\mathbf{X}$ and $\mathbf{Y}$.

Algorithm 3 The WL-CPLS NIPALS algorithm with an alternative cross-covariance criterion.

Initialise: $\underline{\mathbf{X}}_1 = [\mathbf{X}, \mathbf{X}^*]$, $\underline{\mathbf{Y}}_1 = [\mathbf{Y}, \mathbf{Y}^*]$
for $i = 1, \ldots, r$ **do**
  $\mathbf{w}_i = \mathrm{Eig}_{\max}\{\mathbf{X}_i^{\mathrm{H}}\mathbf{Y}_i\mathbf{Y}_i^{\mathrm{H}}\mathbf{X}_i\}$
  $\mathbf{t}_i = \mathbf{X}_i\mathbf{w}_i$, $\underline{\mathbf{t}}_i = [\mathbf{t}_i, \mathbf{t}_i^*]$
  $\underline{\mathbf{c}}_i = (\underline{\mathbf{t}}_i^+\ \underline{\mathbf{Y}}_i)^{\mathrm{H}}$
  $\underline{\mathbf{p}}_i = (\underline{\mathbf{t}}_i^+\ \underline{\mathbf{X}}_i)^{\mathrm{H}}$
  $\mathbf{u}_i = \mathbf{Y}_i\mathbf{c}_{1,i} + \mathbf{Y}_i^*\mathbf{c}_{2,i}$, $\underline{\mathbf{u}}_i = [\mathbf{u}_i, \mathbf{u}_i^*]$
  $\underline{\mathbf{q}}_i = (\underline{\mathbf{u}}_i^+\ \underline{\mathbf{Y}}_i)^{\mathrm{H}}$
  $\underline{\mathbf{X}}_{i+1} = \underline{\mathbf{X}}_i - \underline{\mathbf{t}}_i\underline{\mathbf{p}}_i^{\mathrm{H}}$, $\underline{\mathbf{Y}}_{i+1} = \underline{\mathbf{Y}}_i - \underline{\mathbf{t}}_i\underline{\mathbf{c}}_i^{\mathrm{H}}$
  Store $\mathbf{t}_i$, $\mathbf{u}_i$, $\mathbf{p}_{1,i}$, $\mathbf{p}_{2,i}$, $\mathbf{q}_{1,i}$, $\mathbf{q}_{2,i}$, $\mathbf{c}_{1,i}$, $\mathbf{c}_{2,i}$ and $\mathbf{w}_i$
end for

4. Analysis of the WL-CPLS

The focus of our analysis of the WL-CPLS algorithm is on the calculation of the widely linear regression coefficients, in (11), based on the structure of the WL-CPLS decomposition of $\mathbf{X}$ and $\mathbf{Y}$, so as to satisfy Requirement **R3**. This also highlights the ability of the WL-CPLS algorithms to produce a regularised widely linear regression solution. Next, the application of the WL-CPLS algorithm as a complex covariance matrix diagonalisation transform is demonstrated, including a special case where the result is equivalent to an existing technique, the strong uncorrelating transform (SUT) [23]. The orthogonality between the latent variables, $\mathbf{t}$, and the model residuals is then examined. Finally, the convergence of the WL-CPLS algorithm is proven for a univariate output, $\mathbf{y}$, through a recurrence relation for the latent variables, $\mathbf{t}$.

4.1. PLS For regularised widely linear estimation

We have already shown that the PLS method calculates a joint latent variable decomposition of both the input $\mathbf{X}$ and output $\mathbf{Y}$. In real-valued PLS, this property is used to calculate a regularised regression solution in scenarios where the input matrix $\mathbf{X}$ is sub-rank. We next show that the WL-CPLS algorithm allows a straightforward calculation of the coefficient matrices $\mathbf{H}$ and $\mathbf{G}$ in the widely linear regression model

$$\hat{\mathbf{Y}} = \tilde{\mathbf{X}}\mathbf{H} + \tilde{\mathbf{X}}^*\mathbf{G},$$

where $\tilde{\mathbf{X}}$ is the WL-CPLS approximation of the input $\mathbf{X}$ given

$$\tilde{\mathbf{X}} = \mathbf{T}(\mathbf{P}_1^{\mathrm{H}}\mathbf{W})\mathbf{W}^{\mathrm{H}} + \mathbf{T}^*(\mathbf{P}_2^{\mathrm{H}}\mathbf{W})\mathbf{W}^{\mathrm{H}}. \tag{26}$$

This is a rigorous generalisation of the real-valued NIPALS algorithm, where the matrix $\mathbf{W}$ is unitary and the matrices $\mathbf{P}_1^{\mathrm{H}}\mathbf{W}$ and $\mathbf{P}_2^{\mathrm{H}}\mathbf{W}$ are upper triangular [29]. Upon substituting (26) into (14), the WL-CPLS solutions for $\mathbf{H}$ and $\mathbf{G}$ are calculated as

$$\begin{aligned}\mathbf{H}_{\text{WL-CPLS}} &= [\tilde{\mathbf{C}} - \tilde{\mathbf{P}}^*\tilde{\mathbf{C}}^{*+}\tilde{\mathbf{P}}]^+[\tilde{\mathbf{R}}^{\mathrm{H}} - \tilde{\mathbf{P}}^*\tilde{\mathbf{C}}^{*+}\tilde{\mathbf{S}}^{\mathrm{T}}]\\ \mathbf{G}_{\text{WL-CPLS}} &= [\tilde{\mathbf{C}}^* - \tilde{\mathbf{P}}\tilde{\mathbf{C}}^{+}\tilde{\mathbf{P}}^*]^+[\tilde{\mathbf{S}}^{\mathrm{T}} - \tilde{\mathbf{P}}\tilde{\mathbf{C}}^{+}\tilde{\mathbf{R}}^{\mathrm{H}}],\end{aligned} \tag{27}$$

where $\tilde{\mathbf{R}} = \tilde{\mathbf{Y}}^{\mathrm{H}}\tilde{\mathbf{X}}$, $\tilde{\mathbf{S}} = \tilde{\mathbf{Y}}^{\mathrm{T}}\tilde{\mathbf{X}}$, $\tilde{\mathbf{C}} = \tilde{\mathbf{X}}^{\mathrm{H}}\tilde{\mathbf{X}}$, $\tilde{\mathbf{P}} = \tilde{\mathbf{X}}^{\mathrm{T}}\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$ is given by the decomposition in (24). Therefore, the computation of $\mathbf{H}$ and $\mathbf{G}$ requires calculation of the generalised inverses[5] of $\tilde{\mathbf{C}}$ and $\tilde{\mathbf{C}} - \tilde{\mathbf{P}}^*\tilde{\mathbf{C}}^{*+}\tilde{\mathbf{P}}$, which requires special attention.

Lemma 1 (Miller [35])**.** *Consider a matrix $\mathbf{M} = \mathbf{N} + \mathbf{O}$, where $\mathbf{N} \in \mathbb{C}^{n\times n}$ is invertible and $\mathbf{O} \in \mathbb{C}^{n\times n}$ can be split into rank-1 matrices, $\mathbf{O} = \mathbf{E}_1 + \mathbf{E}_2 + \cdots + \mathbf{E}_r$, with r denoting the rank of $\mathbf{O}$. Then $\mathbf{M}$ can be inverted iteratively as*

$$\mathbf{M}_{k+1}^{-1} = \mathbf{M}_k^{-1} + \frac{\mathbf{M}_k^{-1}\mathbf{E}_k\mathbf{M}_k^{-1}}{1 + Tr(\mathbf{M}_k^{-1}\mathbf{E}_k)}, \tag{28}$$

where $\mathbf{M}_1 = \mathbf{N}$. The iteration is terminated at $k = r$.

Proposition 1. *The inverses of the matrices $\tilde{\mathbf{C}}$ and $\tilde{\mathbf{C}} - \tilde{\mathbf{P}}^*\tilde{\mathbf{C}}^{*+}\tilde{\mathbf{P}}$ can be calculated through the exploitation of their special structure and with the aid of Lemma 1.*

Proof. The proof consists of two parts, the calculation of $\tilde{\mathbf{C}}^+$ and the calculation of $(\tilde{\mathbf{C}} - \tilde{\mathbf{P}}^*\tilde{\mathbf{C}}^{*+}\tilde{\mathbf{P}})^+$.

In order to calculate the generalised inverse of $\tilde{\mathbf{C}}$, the WL-CPLS approximation (26) can be used to give

$$\tilde{\mathbf{C}} = \mathbf{W}(\mathbf{F}_1 + \mathbf{F}_2 + \mathbf{F}_3 + \mathbf{F}_4)\mathbf{W}^{\mathrm{H}}, \tag{29}$$

with $\mathbf{F}_1$, $\mathbf{F}_2$, $\mathbf{F}_3$ and $\mathbf{F}_4$ defined as the **LDU** decompositions

$$\begin{aligned}\mathbf{F}_1 &= (\mathbf{P}_1^{\mathrm{H}}\mathbf{W})^{\mathrm{H}}(\mathbf{P}_1^{\mathrm{H}}\mathbf{W}) = \mathbf{L}_1\mathbf{L}_1^{\mathrm{H}}\\ \mathbf{F}_2 &= (\mathbf{P}_1^{\mathrm{H}}\mathbf{W})^{\mathrm{H}}\mathbf{D}^*(\mathbf{P}_2^{\mathrm{H}}\mathbf{W}) = \mathbf{L}_1\mathbf{D}^*\mathbf{L}_2^{\mathrm{H}}\\ \mathbf{F}_3 &= (\mathbf{P}_2^{\mathrm{H}}\mathbf{W})^{\mathrm{H}}\mathbf{D}(\mathbf{P}_1^{\mathrm{H}}\mathbf{W}) = \mathbf{L}_2\mathbf{D}\mathbf{L}_1^{\mathrm{H}}\\ \mathbf{F}_4 &= (\mathbf{P}_2^{\mathrm{H}}\mathbf{W})^{\mathrm{H}}(\mathbf{P}_2^{\mathrm{H}}\mathbf{W}) = \mathbf{L}_2\mathbf{L}_2^{\mathrm{H}},\end{aligned} \tag{30}$$

where $\mathbf{D} = \mathbf{T}^{\mathrm{T}}\mathbf{T}$ is diagonal, and the implicit assumption of normalised scores, $\mathbf{T}^{\mathrm{H}}\mathbf{T} = \mathbf{I}$ was employed. Since $\mathbf{W}$ is unitary, a generalised inverse for $\mathbf{C}$ is then given by

$$\tilde{\mathbf{C}}^+ = \mathbf{W}(\mathbf{F}_1 + \mathbf{F}_2 + \mathbf{F}_3 + \mathbf{F}_4)^{-1}\mathbf{W}^{\mathrm{H}}, \tag{31}$$

which boils down to the problem of inverting the matrix $(\mathbf{F}_1 + \mathbf{F}_2 + \mathbf{F}_3 + \mathbf{F}_4)$. Although this matrix is non-singular, we show that it is possible to calculate its inverse using only its structure. This is equivalent to the real-domain NIPALS algorithm, where the necessary generalised inverses for a regularised regression are calculated solely due to the decompositions' structure [29].

From the **LDU** decompositions of $\mathbf{F}_n$, $n = 1, 2, 3, 4$ given in (30), we further factorise $(\mathbf{F}_1 + \mathbf{F}_2 + \mathbf{F}_3 + \mathbf{F}_4)$ as

$$\begin{aligned}\mathbf{F}_1 + \mathbf{F}_2 + \mathbf{F}_3 + \mathbf{F}_4 &= \mathbf{L}_1(\mathbf{L}_1^{\mathrm{H}} + \mathbf{D}^*\mathbf{L}_2^{\mathrm{H}}) + \mathbf{L}_2(\mathbf{L}_2^{\mathrm{H}} + \mathbf{D}\mathbf{L}_1^{\mathrm{H}})\\ &= \mathbf{L}_1\mathbf{U}_A + \mathbf{L}_2\mathbf{U}_B\end{aligned}$$

[5] Note that if the matrices in (14) are generally invertible then $\tilde{\mathbf{C}}^*$ and $\tilde{\mathbf{C}}^* - \tilde{\mathbf{P}}\tilde{\mathbf{C}}^{+}\tilde{\mathbf{P}}^*$ are also generally invertible.

$$= \mathbf{A} + \mathbf{B}.$$

Upon substituting $\mathbf{N} = \mathbf{A}$ and $\mathbf{O} = \mathbf{B}$ into Lemma 1, this form permits the calculation of the inverse whereby the **LU** structure of $\mathbf{A}$ guarantees that the generalised inverse of $\tilde{\mathbf{C}}$ is readily obtained.

We next consider the inversion of $\tilde{\mathbf{C}} - \tilde{\mathbf{P}}^{*}\tilde{\mathbf{C}}^{*+}\tilde{\mathbf{P}}$. Similarly to $\tilde{\mathbf{C}}$, the matrix $\tilde{\mathbf{P}}$ can be factorised using the WL-CPLS approximation in (26), to yield

$$\tilde{\mathbf{P}} = \mathbf{W}^{*}(\mathbf{K}_1 + \mathbf{K}_2 + \mathbf{K}_3 + \mathbf{K}_4)\mathbf{W}^{\mathrm{H}}, \tag{32}$$

where

$$\begin{aligned}
\mathbf{K}_1 &= (\mathbf{P}_1^{\mathrm{H}}\mathbf{W})^{\mathrm{T}}\mathbf{D}(\mathbf{P}_1^{\mathrm{H}}\mathbf{W}) = \mathbf{L}_1^{*}\mathbf{D}\mathbf{L}_1^{\mathrm{H}} \\
\mathbf{K}_2 &= (\mathbf{P}_1^{\mathrm{H}}\mathbf{W})^{\mathrm{T}}(\mathbf{P}_2^{\mathrm{H}}\mathbf{W}) = \mathbf{L}_1^{*}\mathbf{L}_2^{\mathrm{H}} \\
\mathbf{K}_3 &= (\mathbf{P}_2^{\mathrm{H}}\mathbf{W})^{\mathrm{T}}(\mathbf{P}_1^{\mathrm{H}}\mathbf{W}) = \mathbf{L}_2^{*}\mathbf{L}_1^{\mathrm{H}} \\
\mathbf{K}_4 &= (\mathbf{P}_2^{\mathrm{H}}\mathbf{W})^{\mathrm{T}}\mathbf{D}^{*}(\mathbf{P}_2^{\mathrm{H}}\mathbf{W}) = \mathbf{L}_2^{*}\mathbf{D}^{*}\mathbf{L}_2^{\mathrm{H}}.
\end{aligned} \tag{33}$$

Denote $\mathbf{K}_1 + \mathbf{K}_2 + \mathbf{K}_3 + \mathbf{K}_4 = \mathbf{K}$ and $\mathbf{F}_1 + \mathbf{F}_2 + \mathbf{F}_3 + \mathbf{F}_4 = \mathbf{F}$, we can now obtain $\mathbf{C} = \mathbf{W}\mathbf{F}\mathbf{W}^{\mathrm{H}}$ and $\mathbf{P} = \mathbf{W}^{*}\mathbf{K}\mathbf{W}^{\mathrm{H}}$. The generalised inverse of $\tilde{\mathbf{C}} - \tilde{\mathbf{P}}^{*}\tilde{\mathbf{C}}^{*+}\tilde{\mathbf{P}}$ can therefore be obtained as

$$(\tilde{\mathbf{C}} - \tilde{\mathbf{P}}^{*}\tilde{\mathbf{C}}^{*+}\tilde{\mathbf{P}})^{+} = \mathbf{W}(\mathbf{F} - \mathbf{K}^{*}\mathbf{F}^{-1*}\mathbf{K})^{-1}\mathbf{W}^{\mathrm{H}}.$$

To calculate the inverse of $(\mathbf{F} - \mathbf{K}^{*}\mathbf{F}^{-1*}\mathbf{K})^{-1}$ we can employ again Lemma 1 with $\mathbf{N} = \mathbf{F}$ and $\mathbf{O} = -\mathbf{K}^{*}\mathbf{F}^{-1*}\mathbf{K}$. The inverse of $\mathbf{F}$ is retained from the calculation of the generalised inverse of $\tilde{\mathbf{C}}$ which satisfies the requirements of Lemma 1. The so obtained generalised inverses of $\tilde{\mathbf{C}}$ and $\tilde{\mathbf{C}} - \tilde{\mathbf{P}}^{*}\tilde{\mathbf{C}}^{*+}\tilde{\mathbf{P}}$ conclude the proof of Proposition 1. □

The steps in the calculation of the WL-CPLS regression coefficient matrices, $\mathbf{H}_{\text{WL-CPLS}}$ and $\mathbf{G}_{\text{WL-CPLS}}$, are summarised in Algorithm 4.

Algorithm 4 Calculation of $\mathbf{H}_{\text{WL-CPLS}}$ and $\mathbf{G}_{\text{WL-CPLS}}$.

1: Use the WL-CPLS in Algorithm 2 to obtain: $\mathbf{T}$, $\mathbf{P}_1$, $\mathbf{P}_2$ and $\mathbf{W}$
2: Calculate WL-CPLS approximation $\tilde{\mathbf{X}} = \mathbf{T}(\mathbf{P}_1^{\mathrm{H}}\mathbf{W})\mathbf{W}^{\mathrm{H}} + \mathbf{T}^{*}(\mathbf{P}_2^{\mathrm{H}}\mathbf{W})\mathbf{W}^{\mathrm{H}}$
3: Obtain $\tilde{\mathbf{C}} = \tilde{\mathbf{X}}^{\mathrm{H}}\tilde{\mathbf{X}}$, $\tilde{\mathbf{P}} = \tilde{\mathbf{X}}^{\mathrm{T}}\tilde{\mathbf{X}}$, $\tilde{\mathbf{R}} = \tilde{\mathbf{Y}}^{\mathrm{H}}\tilde{\mathbf{X}}$ and $\tilde{\mathbf{S}}^{\mathrm{T}} = \tilde{\mathbf{Y}}^{\mathrm{T}}\tilde{\mathbf{X}}$
4: Obtain $\mathbf{L}_1 = (\mathbf{P}_1^{\mathrm{H}}\mathbf{W})^{\mathrm{H}}$, $\mathbf{L}_2 = (\mathbf{P}_2^{\mathrm{H}}\mathbf{W})^{\mathrm{H}}$ and $\mathbf{D} = \mathbf{T}^{\mathrm{T}}\mathbf{T}$
5: Obtain $\mathbf{U}_A = (\mathbf{L}_1^{\mathrm{H}} + \mathbf{D}^{*}\mathbf{L}_2^{\mathrm{H}})$ and $\mathbf{U}_B = (\mathbf{L}_2^{\mathrm{H}} + \mathbf{D}\mathbf{L}_1^{\mathrm{H}})$
6: Define $\mathbf{F} = \mathbf{L}_1\mathbf{U}_A + \mathbf{L}_2\mathbf{U}_B$
7: Calculate $\mathbf{F}^{-1}$ using Lemma 1 with $\mathbf{N} = \mathbf{L}_1\mathbf{U}_A$ and $\mathbf{O} = \mathbf{L}_2\mathbf{U}_B$
8: Calculate $\tilde{\mathbf{C}}^{+} = \mathbf{W}\mathbf{F}^{-1}\mathbf{W}^{\mathrm{H}}$
9: Define $\mathbf{K} = \mathbf{L}_1^{*}\mathbf{D}\mathbf{L}_1^{\mathrm{H}} + \mathbf{L}_1^{*}\mathbf{L}_2^{\mathrm{H}} + \mathbf{L}_2^{*}\mathbf{L}_1^{\mathrm{H}} + \mathbf{L}_2^{*}\mathbf{D}^{*}\mathbf{L}_2^{\mathrm{H}}$
10: Obtain $(\mathbf{F} - \mathbf{K}^{*}\mathbf{F}^{-1*}\mathbf{K})^{-1}$ using Lemma 1 with $\mathbf{N} = \mathbf{F}$ and $\mathbf{O} = -\mathbf{K}^{*}\mathbf{F}^{-1*}\mathbf{K}$
11: Calculate $[\tilde{\mathbf{C}} - \tilde{\mathbf{P}}^{*}\tilde{\mathbf{C}}^{*+}\tilde{\mathbf{P}}]^{+} = \mathbf{W}^{\mathrm{H}}(\mathbf{F} - \mathbf{K}^{*}\mathbf{F}^{-1*}\mathbf{K})^{-1}\mathbf{W}$
12: Calculate $\mathbf{H}_{\text{WL-CPLS}} = [\tilde{\mathbf{C}} - \tilde{\mathbf{P}}^{*}\tilde{\mathbf{C}}^{*+}\tilde{\mathbf{P}}]^{+}[\tilde{\mathbf{R}}^{\mathrm{H}} - \tilde{\mathbf{P}}^{*}\tilde{\mathbf{C}}^{*+}\tilde{\mathbf{S}}^{\mathrm{T}}]$
13: Calculate $\mathbf{G}_{\text{WL-CPLS}} = ([\tilde{\mathbf{C}} - \tilde{\mathbf{P}}^{*}\tilde{\mathbf{C}}^{*+}\tilde{\mathbf{P}}]^{+})^{*}[\tilde{\mathbf{S}}^{\mathrm{T}} - \tilde{\mathbf{P}}\tilde{\mathbf{C}}^{+}\tilde{\mathbf{R}}^{\mathrm{H}}]$

Remark 7. The widely linear complex PLS algorithm in Algorithm 2 provides an approximation of the matrix $\mathbf{X}$, denoted by $\tilde{\mathbf{X}}$, which admits a regularised widely linear regression $\mathbf{Y} = \tilde{\mathbf{X}}\mathbf{H}_{\text{WL-CPLS}} + \tilde{\mathbf{X}}^{*}\mathbf{G}_{\text{WL-CPLS}}$. The regression coefficients are obtained from Algorithm 4, owing to the inherent structure of the WL-CPLS decomposition. As a result, the final Requirement **R3** for the WL-CPLS is met. We note that the required generalised inverses can also be calculated through other methods, not utilising the matrix structures.

4.2. WL-CPLS as a covariance matrix diagonalisation transform

Diagonalisation of covariance matrices is an essential data analysis tool, and for real-valued data this is accomplished through PCA/SVD. However, for the applications in the complex domain [24], both the covariance and pseudocovariance matrices must be simultaneously diagonalised. This can be achieved through the strong uncorrelating transform (SUT) [23], given by

$$\mathbf{X}_{SUT} = \mathbf{X}\mathbf{\Phi}, \tag{34}$$

where $\mathbf{\Phi} \in \mathbb{C}^{m\times m}$ is a transform, obtained through Algorithm 5, which maps the measured variables, $\mathbf{X} \in \mathbb{C}^{N\times m}$, on to the uncorrelated SUT variables, $\mathbf{X}_{SUT} \in \mathbb{C}^{N\times m}$. The transformed covariance matrix $\mathbf{C}_{\mathbf{X}_{SUT}} = \mathbf{X}_{SUT}^{\mathrm{H}}\mathbf{X}_{SUT}$ is an identity matrix, and the pseudocovariance matrix $\mathbf{P}_{\mathbf{X}_{SUT}} = \mathbf{X}_{SUT}^{\mathrm{T}}\mathbf{X}_{SUT}$ is diagonal and with real-valued entries which represent the circularity quotient (defined in Section 2.4) of each component (column vector) in $\mathbf{X}_{SUT}$.

Algorithm 5 The SUT algorithm.

1: **Initialise:** The data matrix $\mathbf{X}$ is provided
2: Calculate empirical covariance matrix $\mathbf{C} = \mathbf{X}^{\mathrm{H}}\mathbf{X}$
3: Apply the SVD to give $\mathbf{C} = \mathbf{U}\mathbf{\Delta}\mathbf{U}^{\mathrm{H}}$
4: Obtain new variables $\hat{\mathbf{X}} = \mathbf{X}\mathbf{U}\mathbf{\Delta}^{-1/2}$
5: Calculate empirical psuedocovariance matrix $\mathbf{P} = \hat{\mathbf{X}}^{\mathrm{T}}\hat{\mathbf{X}}$
6: Apply the Takagi factorisation to give $\mathbf{P} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\mathrm{T}}$
7: Obtain SUT variables $\tilde{\mathbf{X}}_{SUT} = \mathbf{X}\mathbf{U}\mathbf{\Delta}^{-1/2}\mathbf{V}^{*} = \mathbf{X}\mathbf{\Phi}$

We now show that the WL-CPLS result can be viewed, similarly to the SUT, as an uncorrelating transform by considering the augmented form of the decomposition of $\mathbf{X}$, given by

$$\underline{\mathbf{T}} = \underline{\mathbf{X}}\begin{pmatrix} (\mathbf{P}_1^{\mathrm{H}}\mathbf{W})\mathbf{W}^{\mathrm{H}} & ((\mathbf{P}_2^{\mathrm{H}}\mathbf{W})\mathbf{W}^{\mathrm{H}})^{*} \\ (\mathbf{P}_2^{\mathrm{H}}\mathbf{W})\mathbf{W}^{\mathrm{H}} & ((\mathbf{P}_1^{\mathrm{H}}\mathbf{W})\mathbf{W}^{\mathrm{H}})^{*} \end{pmatrix}^{+}. \tag{35}$$

The matrix $\underline{\mathbf{T}}^{\mathrm{H}}\underline{\mathbf{T}}$ is block diagonal, and hence, the transformed covariance matrix, $\mathbf{T}^{\mathrm{H}}\mathbf{T}$, and the pseudocovariance matrix, $\mathbf{T}^{\mathrm{T}}\mathbf{T}$, are both diagonal. Therefore, the matrix

$$\begin{pmatrix} (\mathbf{P}_1^{\mathrm{H}}\mathbf{W})\mathbf{W}^{\mathrm{H}} & ((\mathbf{P}_2^{\mathrm{H}}\mathbf{W})\mathbf{W}^{\mathrm{H}})^{*} \\ (\mathbf{P}_2^{\mathrm{H}}\mathbf{W})\mathbf{W}^{\mathrm{H}} & ((\mathbf{P}_1^{\mathrm{H}}\mathbf{W})\mathbf{W}^{\mathrm{H}})^{*} \end{pmatrix}^{+},$$

obtained by the WL-CPLS solution from regressing $\mathbf{X}$ to itself can be considered as an uncorrelating transform.

To validate the performance of the so-obtained WL-CPLS uncorrelating transform for dimensionality reduction, we consider a case where a data matrix

$$\mathbf{X} = \mathbf{Z} + \mathbf{N}, \tag{36}$$

contains a low-rank "signal" subspace component, denoted by the matrix $\mathbf{Z}$, and a full-rank "noise" subspace component, denoted by the matrix $\mathbf{N}$. This is a common real-world scenario and if the signal accounted for the majority of the total variance in $\mathbf{X}$ then a practical uncorrelating transform would clearly identify the low-rank subspace. To this end, we generated $N = 1000$ samples of 20 independent, identically distributed (i.i.d.), non-circular, Gaussian sources which were mixed to give a matrix $\mathbf{Z} \in \mathbb{C}^{1000\times 50}$. A noise source, $\mathbf{N} \in \mathbb{C}^{1000\times 50}$, was added to the matrix, $\mathbf{Z}$, to give the data matrix $\mathbf{X} = \mathbf{Z} + \mathbf{N}$, in (36), drawn from a circular, Gaussian i.i.d. distribution with an SNR=26dB, where the SNR is defined as

$$SNR = 10\log_{10}\frac{Tr\{E[\mathbf{Z}^{\mathrm{H}}\mathbf{Z}]\}}{Tr\{E[\mathbf{N}^{\mathrm{H}}\mathbf{N}]\}}, \tag{37}$$

The SUT and WL-CPLS transforms were then performed. Fig. 2 shows the percentage of the total variance in $\mathbf{X}$ that is explained by the approximation obtained from each score described by the metric

$$VE = 100\frac{Tr\{\tilde{\mathbf{A}}^{\mathrm{H}}\tilde{\mathbf{A}}\}}{Tr\{\mathbf{A}^{\mathrm{H}}\mathbf{A}\}}. \tag{38}$$

For the WL-CPLS result $\tilde{\mathbf{A}} = \underline{\mathbf{t}}_i\underline{\mathbf{p}}_i^{\mathrm{H}}$, where i is the iteration number for the respective WL-CPLS score and loading and for the SUT result $\tilde{\mathbf{A}} = \mathbf{x}_{SUT,i}\mathbf{\Phi}_i^{+}$, where $\mathbf{x}_{SUT,i}$ is the *ith* component (column of

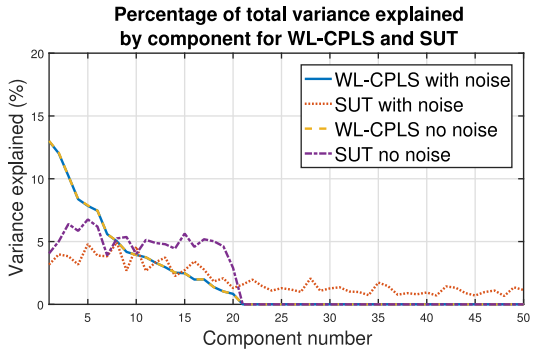

Fig. 2. Proportion of the variance (in %) explained by each component in the SUT and WL-CPLS transform, *VE* in (38).

$\mathbf{X}_{SUT}$) and $\mathbf{\Phi}_i^+$ is the *ith* row of the pseudoinverse of the SUT transform $\mathbf{\Phi}$. The metric *VE* was calculated for both the WL-CPLS/SUT results obtained from both the data matrix $\mathbf{A} = \mathbf{X}$ and for the noiseless case $\mathbf{A} = \mathbf{Z}$ from above in (36).

Fig. 2 indicates that the uncorrelating transform performed by the WL-CPLS is more powerful for subspace identification than the SUT. Observe that the WL-CPLS concentrates the energy in the signal into as few components as possible, owing to the fact it selects a component in each iteration which explains the maximum variance in $\mathbf{X}$ (due to the cross-covariance optimisation problem in (18)). On the other hand for the SUT, owing to the inherent whitening, each element accounts for a similar amount of variance. This is more clearly seen for the results of the data matrix, $\mathbf{X}$, compared to those for just the low-rank signal matrix, $\mathbf{Z}$. Clearly, in the absence of noise (which increases the rank), the SUT can identify well a low-rank subspace. For even a small amount of noise, however, the signal power is spread amongst all components, which makes it diffcult for the SUT to identify a subspace. On the other hand, the WL-CPLS performs similarly in both cases. Note that the computational requirements of the SUT are only two SVDs whereas the WL-CPLS transform requires one SVD per latent component extracted.

4.2.1. Duality with SUT

A special case where the application of the WL-CPLS algorithm as a transform and the SUT yield identical results exists for a unitary input, $\mathbf{X}$, that has an empirical covariance matrix $\mathbf{C} = \mathbf{X}^{\mathrm{H}}\mathbf{X} = \mathbf{I}$. For such data, both methods result in the transform matrix $\mathbf{V}^*$ obtained from the Takagi factorisation of the matrix $\mathbf{P} = \mathbf{X}^{\mathrm{T}}\mathbf{X} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\mathrm{T}}$. In other words, for "white" data, the SUT transformed variables, $\mathbf{X}_{SUT}$ in (34), are equal to the scores matrix, $\mathbf{T}$, produced by the WL-CPLS algorithm.

To prove this equivalence, observe first that the whitening transform in the SUT calculation (outlined in Steps 3 and 4 of Algorithm 5) is no longer required, and so the required Takagi factorisation can be calculated directly from the empirical pseudo-covariance matrix $\mathbf{P}$. On the other hand, for the WL-CPLS result, the input and output deflations (Step 11 of Algorithm 2) are symmetric and are achieved by a strictly linear regression. This means that $\mathbf{p}_1 = \mathbf{c}_1 = \mathbf{w}$ and $\mathbf{p}_2 = \mathbf{c}_2 = \mathbf{0}$. The WL-CPLS solution can now be calculated in a closed form through the SVD of $\mathbf{S}_{Re}\mathbf{S}_{Re}^{\mathrm{T}}$ (where $\mathbf{S}_{Re} = \mathbf{X}_{Re}^{\mathrm{T}}\mathbf{Y}_{Re}$) as the first m (where m is the number of columns of $\mathbf{X}$) singular vectors are cast into the complex domain according to (19). To demonstrate the equivalence to the Takagi factorisation of the empirical pseudocovariance matrix, $\mathbf{P}$, consider the link between the SVDs of $\mathbf{X}_{Re}^{\mathrm{T}}\mathbf{X}_{Re} = \mathbf{U}_{Re}\mathbf{\Sigma}_{Re}\mathbf{U}_{Re}^{\mathrm{T}}$ and $\underline{\mathbf{X}}^{\mathrm{H}}\underline{\mathbf{X}} = \mathbf{U}_C\mathbf{\Sigma}_C\mathbf{U}_C^{\mathrm{H}}$. It is well known that $\mathbf{\Sigma}_{Re} = 2\mathbf{\Sigma}_C$ [36] and hence, the transformed singular vectors $\mathbf{\Gamma}_m\mathbf{U}_{Re}$ produce the same diagonalisation of the matrix $\underline{\mathbf{X}}^{\mathrm{H}}\underline{\mathbf{X}}$ as the SVD, up to a factor of 2.

Lemma 2 (Horn [37]). *Let $\mathbf{A} \in \mathbb{C}^{n\times n}$ be a square matrix. There exists a unitary matrix, $\mathbf{X}$, a diagonal matrix, $\mathbf{\Lambda}$, with non-negative entries, and a matrix $\mathbf{Y}$ with orthonormal rows such that $\mathbf{A} = \mathbf{X}\mathbf{\Lambda}\mathbf{Y}$. The columns of matrix $\mathbf{X}$ are the eigenvectors of $\mathbf{A}\mathbf{A}^{\mathrm{H}}$ and the diagonal entries of the matrix $\mathbf{\Lambda}$ are the square root of the corresponding eigenvalues of $\mathbf{A}\mathbf{A}^{\mathrm{H}}$. If $\mathbf{A}\mathbf{A}^{\mathrm{H}}$ has distinct eigenvalues, then $\mathbf{X}$ is determined up to a right diagonal factor $\mathbf{D} = \mathrm{diag}(e^{i\theta_1}, \ldots, e^{i\theta_n})$ with all $\theta \in \mathbb{R}$ and $|d_{ij}| = |e^{i\theta}| = 1$; that is, if $\mathbf{A} = \mathbf{X}_1\mathbf{\Lambda}\mathbf{Y}_1 = \mathbf{X}_2\mathbf{\Lambda}\mathbf{Y}_2$ then $\mathbf{X}_2 = \mathbf{X}_1\mathbf{D}$.*

Lemma 2 (which is a property of the SVD and is proved in [37]) states that if a square matrix $\mathbf{A}$ can be represented as $\mathbf{A} = \mathbf{X}_1\mathbf{\Lambda}\mathbf{Y}_1 = \mathbf{X}_2\mathbf{\Lambda}\mathbf{Y}_2$ (where $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are unitary matrices) then $\mathbf{X}_2 = \mathbf{X}_1\mathbf{D}$ where $\mathbf{D} = \mathrm{diag}(e^{i\theta_1}, \ldots, e^{i\theta_n})$, $\theta \in \mathbb{R}$ and $|d_{ij}| = |e^{i\theta}| = 1$. Using Lemma 2 we can now derive the relationship

$$\mathbf{\Gamma}_m\mathbf{U}_{Re} = \sqrt{2}\mathbf{U}_C\mathbf{D}. \tag{39}$$

Therefore, the latent vectors obtained by the WL-CPLS are identical to those obtained from the SVD of $\underline{\mathbf{X}}^{\mathrm{H}}\underline{\mathbf{X}}$, each rotated to be of the form

$$\sqrt{2}\mathbf{U}_C\mathbf{D} = \begin{pmatrix} \mathbf{U}_{S1} & \mathbf{U}_{S2} \\ \mathbf{U}_{S1}^* & \mathbf{U}_{S2}^* \end{pmatrix}, \tag{40}$$

where the matrix $[\mathbf{U}_{S1}^{\mathrm{T}}, \mathbf{U}_{S1}^{\mathrm{H}}]^{\mathrm{T}}$ represents the matrix of the first m (of $2m$) eigenvectors of $\mathbf{S}_{Re}\mathbf{S}_{Re}^{\mathrm{T}}$ transformed to the complex domain (as in (19)) and the matrix $[\mathbf{U}_{S2}^{\mathrm{T}}, \mathbf{U}_{S2}^{\mathrm{H}}]^{\mathrm{T}}$ represents the remaining m eigenvectors of $\mathbf{S}_{Re}\mathbf{S}_{Re}^{\mathrm{T}}$. The matrix $\mathbf{U}_{S1}$ is then identical to the matrix $\mathbf{W}$ for the WL-CPLS result. In order for the matrix $\sqrt{2}\mathbf{U}_C\mathbf{D}$ to diagonalise $\underline{\mathbf{X}}^{\mathrm{H}}\underline{\mathbf{X}}$ in this special case, the sum $\mathbf{U}_{S1}^{\mathrm{H}}\mathbf{P}^*\mathbf{U}_{S1}^* + \mathbf{U}_{S1}^{\mathrm{T}}\mathbf{P}\mathbf{U}_{S1} + \mathbf{U}_{S1}^{\mathrm{H}}\mathbf{U}_{S1} + \mathbf{U}_{S1}^{\mathrm{T}}\mathbf{U}_{S1}^*$ must be diagonal. This condition (the derivation is given in Appendix A) is satisfied if $\mathbf{U}_{S1} = \mathbf{V}^*$ from the Takagi factorisation of the matrix $\mathbf{P} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\mathrm{T}}$. Therefore, the WL-CPLS result for $\mathbf{X}$ serving as both the input and output is equivalent to that of the SUT for data for which the empirical covariance matrix is given by $\mathbf{C} = \mathbf{I}$.

4.3. The convergence of the WL-CPLS algorithm

The convergence analysis of real-valued PLS algorithms has proved a difficult task [38]. For a univariate output, $\mathbf{y}$, it has been shown by Helland [39] that the vectors $\mathbf{w}$ span a space defined by the vectors in a Krylov sequence $\mathbf{K}_i = (\mathbf{s}, \mathbf{S}\mathbf{s}, \ldots, \mathbf{S}^{i-1}\mathbf{s})$ (defined as the Krylov space $\mathcal{K}_i(\mathbf{S}, \mathbf{s}) = \mathrm{span}(\mathbf{K}_i)$) where $\mathbf{s} = \mathbf{X}^{\mathrm{T}}\mathbf{y}$, $\mathbf{S} = \mathbf{X}^{\mathrm{T}}\mathbf{X}$ and i is the number of components, while the $\mathbf{t}$ vectors span a Krylov sequence, $\mathbf{K}_i$, where $\mathbf{s} = \mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{y}$ and $\mathbf{S} = \mathbf{X}\mathbf{X}^{\mathrm{T}}$ [38,40]. Furthermore, the work of Bro [38] highlights that the residuals for $\mathbf{X}$ and $\mathbf{y}$ are orthogonal to the space defined by the obtained $\mathbf{T}$, an important property for a regression algorithm.

4.3.1. The orthogonality of the model residuals

Consider the WL-CPLS approximation with i components of $\mathbf{X}$ in (23). We can write the residual of $\mathbf{X}$ as

$$\mathbf{X}_{i+1} = \mathbf{X} - \underline{\mathbf{T}}_i\overline{\mathbf{P}}_i^{\mathrm{H}}, \tag{41}$$

where $\overline{\mathbf{P}}_i = [\mathbf{P}_{1,i}, \mathbf{P}_{2,i}]$. Note that $\mathbf{X}_{i+1}$ is equivalent to the deflation step in (22) at the ith iteration. The matrix $\overline{\mathbf{P}}_i$ can be obtained (from (23)) as

$$\overline{\mathbf{P}}_i = \mathbf{X}^{\mathrm{H}}\underline{\mathbf{T}}_i(\underline{\mathbf{T}}_i^{\mathrm{H}}\underline{\mathbf{T}}_i)^{-1},$$

and the residuals from (41) can be expressed as

$$\mathbf{X}_{i+1} = (\mathbf{I} - \mathcal{P}_{t,i})\mathbf{X}, \tag{42}$$

where $\mathcal{P}_{t,i} = \underline{\mathbf{T}}_i(\underline{\mathbf{T}}_i^{\mathrm{H}}\underline{\mathbf{T}}_i)^{-1}\underline{\mathbf{T}}_i^{\mathrm{H}}$ is a projection onto the space spanned by $\underline{\mathbf{T}}_i$. A similar relationship can be derived for the output $\mathbf{Y}$ using the WL-CPLS model (24), where the model residuals are given as

$\mathbf{Y}_{i+1} = \mathbf{Y} - \underline{\mathbf{T}}_i \bar{\mathbf{C}}_i^{\mathrm{H}}$ with $\bar{\mathbf{C}}_i = \mathbf{Y}^{\mathrm{H}}\underline{\mathbf{T}}_i(\underline{\mathbf{T}}_i^{\mathrm{H}}\underline{\mathbf{T}}_i)^{-1}$. As above, this leads to a relationship of the form

$$\mathbf{Y}_{i+1} = (\mathbf{I} - \mathcal{P}_{t,i})\mathbf{Y}. \tag{43}$$

From (43) and (42), the residuals for both multivariate $\mathbf{X}_i$ and $\mathbf{Y}_i$ must be orthogonal to the subspace defined by the augmented latent variables $\underline{\mathbf{T}}_i$ as $\mathcal{P}_{t,i}$ is a projection matrix, verified by the computation of $\mathcal{P}_{t,i}^{\mathrm{H}}\mathcal{P}_{t,i} = \mathcal{P}_{t,i}$.

4.3.2. Proof of WL-CPLS convergence to a Krylov space

Consider the univariate output $\mathbf{y}$. In this case we can derive a recurrence relation for the augmented scores $\underline{\mathbf{t}} = [\mathbf{t}_i, \mathbf{t}_i^*]$ obtained in the alternative form of WL-CPLS described in Algorithm 3, where the vector $\mathbf{w}_i$ in each iteration is obtained as the largest eigenvector of the matrix $\mathbf{X}_i^{\mathrm{H}}\mathbf{y}_i\mathbf{y}_i^{\mathrm{H}}\mathbf{X}_i$. This eigenvector is given by [39]

$$\mathbf{w}_i = \frac{\mathbf{X}_i^{\mathrm{H}}\mathbf{y}}{||\mathbf{X}_i^{\mathrm{H}}\mathbf{y}||},$$

and is proportional to the vector $\mathbf{X}_i^{\mathrm{H}}\mathbf{y}$. The score vector is then given by $\mathbf{t}_i = \mathbf{X}_i\mathbf{w}_i$ which is proportional to $\mathbf{X}_i\mathbf{X}_i^{\mathrm{H}}\mathbf{y}$. Utilising the relation obtained in (42) we can now write

$$\mathbf{t}_{i+1} \propto (\mathbf{I} - \mathcal{P}_{t,i})\mathbf{X}\mathbf{X}^{\mathrm{H}}(\mathbf{I} - \mathcal{P}_{t,i})\mathbf{y}.$$

This can be written in the form of a recurrence relationship of the augmented scores vector

$$\underline{\mathbf{t}}_{i+1} \propto [(\mathbf{I} - \mathcal{P}_{t,i})\mathbf{X}\mathbf{X}^{\mathrm{H}}(\mathbf{I} - \mathcal{P}_{t,i})\mathbf{y},\ (\mathbf{I} - \mathcal{P}_{t,i})^*\mathbf{X}^*\mathbf{X}^{\mathrm{T}}(\mathbf{I} - \mathcal{P}_{t,i})^*\mathbf{y}^*]. \tag{44}$$

Proposition 2. *The augmented scores matrix,* $\underline{\mathbf{T}}_i = [\mathbf{T}_i, \mathbf{T}_i^*]$, *obtained from the WL-CPLS in Algorithm 3, form a basis for the space defined by the vectors* $\mathbf{K}_i = [\mathbf{s}, \mathbf{S}\mathbf{K}_{i-1}, \mathbf{S}\mathbf{K}_{i-1}^*]$ *and its conjugate* $\mathbf{K}_i^*$, *where* $\mathbf{s} = \mathbf{X}\mathbf{X}^{\mathrm{H}}\mathbf{y}$, $\mathbf{S} = \mathbf{X}\mathbf{X}^{\mathrm{H}}$ *and* $\mathbf{K}_1 = [\mathbf{s}]$.

Proof. In order to show that the columns of $\underline{\mathbf{T}}_i$ form a basis of the space spanned by the columns of $[\mathbf{K}_i, \mathbf{K}_i^*]$, it is sufficient to show that they can be created through a linear combination of the columns. The proof is obtained by induction, in the same way as the proof of Proposition 3.1 in [39]. It has been shown that $\underline{\mathbf{t}}_1 \propto [\mathbf{s}, \mathbf{s}^*] = [\mathbf{K}_1, \mathbf{K}_1^*]$ and therefore our hypothesis is true for the base case $i = 1$. We then assume that it is true that the columns of the matrix $\underline{\mathbf{T}}_i$ are a linear combination of the column vectors in $\mathbf{K}_i$ and $\mathbf{K}_i^*$. The matrix $\mathcal{P}_{t,i}$ is therefore also a combination of the vectors in $\mathbf{K}_i$ and $\mathbf{K}_i^*$. The recursion in (44) represents the composite of the vector

$$\mathbf{t}_{i+1} \propto \mathbf{s} - \mathbf{S}\mathcal{P}_{t,i}\mathbf{y} - \mathcal{P}_{t,i}\mathbf{s} + \mathcal{P}_{t,i}\mathbf{S}\mathcal{P}_{t,i}\mathbf{y},$$

and the vector

$$\mathbf{t}_{i+1}^* \propto \mathbf{s}^* - \mathbf{S}^*\mathcal{P}_{t,i}^*\mathbf{y}^* - \mathcal{P}_{t,i}^*\mathbf{s}^* + \mathcal{P}_{t,i}^*\mathbf{S}^*\mathcal{P}_{t,i}^*\mathbf{y}^*.$$

Since the matrix $\mathcal{P}_{t,i}$ is a linear combination of the columns of $[\mathbf{K}_i, \mathbf{K}_i^*]$, the vectors $\mathbf{t}_{i+1}$ and $\mathbf{t}_{i+1}^*$ are then a linear combination of the columns of $[\mathbf{K}_{i+1}, \mathbf{K}_{i+1}^*]$ which is $[\mathbf{s}, \mathbf{s}^*, \mathbf{S}\mathbf{K}_i, \mathbf{S}^*\mathbf{K}_i, \mathbf{S}\mathbf{K}_i^*, \mathbf{S}^*\mathbf{K}_i^*]$ and so if $\underline{\mathbf{T}}_i$ is a linear combination of the columns of $[\mathbf{K}_i, \mathbf{K}_i^*]$ then the columns of $\underline{\mathbf{T}}_{i+1}$ are a linear combination of the columns of $[\mathbf{K}_{i+1}, \mathbf{K}_{i+1}^*]$ which proves Proposition 2. □

5. Simulation results

The performance of the proposed WL-CPLS algorithm is verifed in a variety of scenarios. The performance metric used was the prediction mean square error (MSE), defined as

$$MSE = E[\|\mathbf{Y} - \hat{\mathbf{Y}}\|_F^2], \tag{45}$$

where $\mathbf{Y}$ are the original "correct" output variables and $\hat{\mathbf{Y}}$ are their predictions from the WL-CPLS model, $\hat{\mathbf{Y}} = \tilde{\mathbf{X}}\mathbf{H} + \tilde{\mathbf{X}}^*\mathbf{G}$, calculated from the approximation $\tilde{\mathbf{X}} = \mathbf{T}(\mathbf{P}_1^{\mathrm{H}}\mathbf{W})\mathbf{W}^{\mathrm{H}} + \mathbf{T}^*(\mathbf{P}_2^{\mathrm{H}}\mathbf{W})\mathbf{W}^{\mathrm{H}}$. For rigour, we examined the key factors that affect the performance of the WL-CPLS algorithm: the number of WL-CPLS components selected and the noise level in $\mathbf{X}$ and $\mathbf{Y}$.

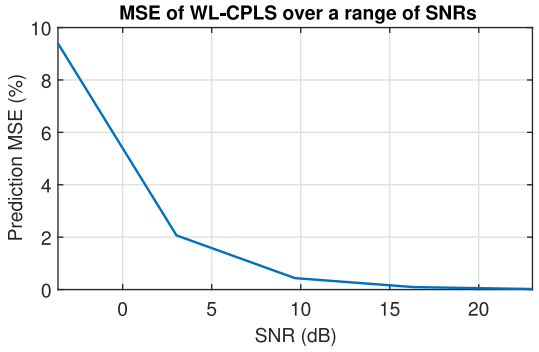

Fig. 3. Prediction MSE for $\hat{\mathbf{Y}}$ (as a percentage of total variance) against the SNR in $\mathbf{Y}$.

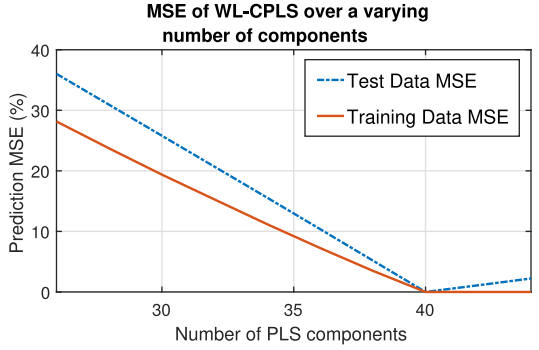

Fig. 4. Performance, in terms of MSE, of the prediction of $\mathbf{Y}$ (as a percentage of total variance) for a varying number of WL-CPLS components, in both training and test data.

5.1. Prediction MSE for a varying SNR in $\mathbf{Y}$

The performance of the WL-CPLS estimator of the output, $\mathbf{Y}$, was assessed for a varying SNR in $\mathbf{Y}$. This is a classical linear regression scenario which assumes that the input variables, $\mathbf{X}$, are accurate, whereas the output variables, $\mathbf{Y}$, are corrupted by noise. The data were generated as described in Table 1 with $N = 1000$, $r = 20$, $m = p = 100$, $\sigma_{\mathbf{X}} = 0$ while $\sigma_{\mathbf{Y}}$ was varied to give a range of SNRs defined as in (37) where $\mathbf{Z} = \mathbf{Y}$ and $\mathbf{N} = \mathbf{N}_{\mathbf{Y}}$.

The WL-CPLS estimator was obtained for 20 components from training data and the average prediction MSE was then calculated for an ensemble of 100 realisations. The results are shown in Fig. 3 and the prediction MSE was calculated as a percentage of the total variance in $\mathbf{Y}$. Observe that even for negative SNRs the WL-CPLS produced a regularised regression estimator with less than a 10% error.

5.2. Prediction MSE for a varying number of WL-CPLS components

The WL-CPLS solution was next examined over a varying number of components, in order to reflect the rank of the approximation for $\tilde{\mathbf{X}}$. In this set of simulations, the average prediction MSE was calculated over an ensemble of 100 trials, for a varying SNR and a number of components in WL-CPLS. The data was generated as described in Table 1, with $N = 1000$, $r = 40$, $m = p = 100$ and no noise added ($\sigma_{\mathbf{X}} = \sigma_{\mathbf{Y}} = 0$). The MSE (as a percentage of the total variance) between the training data and the corresponding WL-CPLS estimate is shown by the solid red line in Fig. 4. Observe a characteristic "elbow" when more than 40 components are used

Table 1
Generation of synthetic test data matrices $\mathbf{X}$ and $\mathbf{Y}$.

1 :	Initialise inputs : N, r, m, p, $\sigma_{\mathbf{X}}$ and $\sigma_{\mathbf{Y}}$
2 :	Define $\mathbf{s}_n \in \mathbb{C}^{N\times 1}$ as vector of N samples generated such that $\Re\{\mathbf{s}_n\} = \mathcal{N}(0,1)$ and $\Im\{\mathbf{s}_n\} = \mathcal{N}(0,1)$
3 :	Define $\mathbf{0}_n$ as a vector of N zeros
4 :	Generate matrix $\mathbf{M} \in \mathbb{R}^{m\times m}$ such that $\mathbf{M}^{\mathsf{T}}\mathbf{M} = \mathbf{I}$
5 :	Create matrix $\mathbf{X} = [\mathbf{s}_1, \mathbf{s}_2, \ldots \mathbf{s}_r, \mathbf{0}_1, \mathbf{0}_2, \ldots, \mathbf{0}_{m-r}]\mathbf{M}$
6 :	Generate matrices $\mathbf{H} \in \mathbb{C}^{m\times p}$ and $\mathbf{G} \in \mathbb{C}^{m\times p}$ such that $\mathbf{H}^{\mathsf{H}}\mathbf{H} = \mathbf{I}$ and $\mathbf{G}^{\mathsf{H}}\mathbf{G} = \mathbf{I}$
7 :	Define $\mathbf{n}_n \in \mathbb{C}^{N\times 1}$ as vector of N samples generated as $\Re\{\mathbf{n}_n\} = \mathcal{N}(0, \sigma_{\mathbf{Y}}^2)$ and $\Im\{\mathbf{n}_n\} = \mathcal{N}(0, \sigma_{\mathbf{Y}}^2)$
8 :	Create matrix $\mathbf{N}_{\mathbf{Y}} = [\mathbf{n}_1, \ldots, \mathbf{n}_p]$
9 :	Create matrix $\mathbf{Y} = \mathbf{X}\mathbf{H} + \mathbf{X}^*\mathbf{G} + \mathbf{N}_{\mathbf{Y}}$
10 :	Define $\mathbf{n}_n \in \mathbb{C}^{N\times 1}$ as vector of N samples generated as $\Re\{\mathbf{n}_n\} = \mathcal{N}(0, \sigma_{\mathbf{X}}^2)$ and $\Im\{\mathbf{n}_n\} = \mathcal{N}(0, \sigma_{\mathbf{X}}^2)$
11:	Create matrix $\mathbf{N}_{\mathbf{X}} = [\mathbf{n}_1, \ldots, \mathbf{n}_m]$
12 :	Create matrix $\mathbf{X} = \mathbf{X} + \mathbf{N}_{\mathbf{X}}$

to calculate the WL-CPLS solution, beyond this point there was no significant information added by further components and the WL-CPLS model "over-fits". This is demonstrated by the ensemble average prediction MSE for test data using the WL-CPLS estimator and shown by the dashed blue line in Fig. 4. The MSE was lowest for 40 WL-CPLS components, which confirms that the number of WL-CPLS latent variables selected should be the number of independent components used to create the joint process.

5.2.1. Stopping criteria for WL-CPLS algorithm

Notice that the WL-CPLS formulations in Algorithms 2 and 3 are iterated until r components are found. The value of r is often not known *a priori* but can be determined based on a stopping criterion. For example, the WL-CPLS solution is computed for a range of component numbers, r, then the prediction MSE for both training and test data is determined, and the variable r is selected as the index of the last component after which adding a new component to the data no longer significantly improves the MSE. This is indicated by the "elbow" in Fig. 4.

5.3. Prediction MSE for a varying SNR in $\boldsymbol{X}$ and $\boldsymbol{Y}$

We next assessed the performance of the WL-CPLS in the presence of noise in both $\mathbf{X}$ and $\mathbf{Y}$. The data were generated as shown in Table 1 with $N = 1000$, $r = 40$, $m = p = 100$, while $\sigma_{\mathbf{X}}$ and $\sigma_{\mathbf{Y}}$ were varied so as to give a range of SNRs, defined as in (37), where $\mathbf{Z}$ is either $\mathbf{X}$ or $\mathbf{Y}$ and $\mathbf{N}$ is $\mathbf{N}_{\mathbf{X}}$ or $\mathbf{N}_{\mathbf{Y}}$, respectively.

Remark 8. The ordinary least squares (OLS) regression model assumes that the input variables, $\mathbf{X}$, are accurate and only the output variables, $\mathbf{Y}$, may contain error, however, this is not generally the case in real-world scenarios. The PLS aims to improve the estimate by using only the relevant subspace shared between the variables $\mathbf{X}$ and $\mathbf{Y}$, so as to eliminate spurious correlations from erroneous variables from the regression calculation.

Fig. 5 shows the average prediction MSE (as a percentage of the total variance) of an ensemble of 100 realisations for the WL-CPLS solution obtained from training data (generated as before) with varying noise levels for $\mathbf{X}$ and $\mathbf{Y}$ (the noise level for $\mathbf{X}$ was defined similarly to the noise in $\mathbf{Y}$ above) where the training data $\mathbf{X}$ had 40 independent components mixed over $m = 100$ variables and transformed to 100 variables in $\mathbf{Y}$. For comparison, we used the Moore-Penrose pseudoinverse [28] to provide the inversions required in Section 4.1, and Fig. 6 shows the average prediction MSE for the same ensemble as in Fig. 5. Observe that the MSE of WL-CPLS was lower, especially for higher noise levels in input data $\mathbf{X}$.

6. Distributed frequency estimation in power grids

The problem complex-valued of frequency estimation in multi-node systems is important in modern smart grids [12,41,42], where

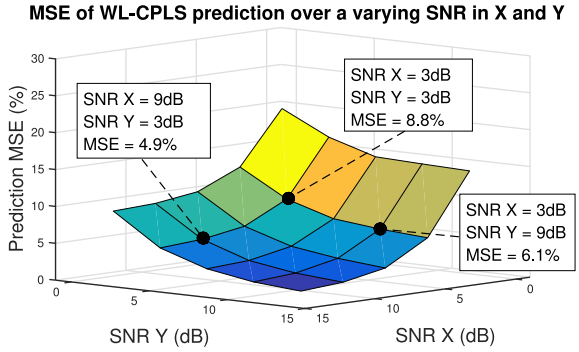

Fig. 5. Prediction MSE against the noise level in $\mathbf{Y}$ (as a percentage of total variance), with noise also present in $\mathbf{X}$.

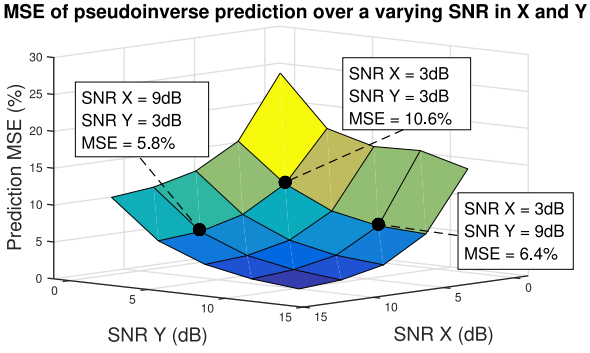

Fig. 6. Prediction obtained from the Moore–Penrose pseudoinverse against the noise level in $\mathbf{Y}$ (as a percentage of total variance), with noise also present in $\mathbf{X}$.

any imbalance is indicated by a noncircular behaviour of a voltage phasor.

6.1. Frequency estimation: Problem specification

Consider a network of M voltage sensors where each node has access to sampled three-phase voltage measurements, at the discrete time instant k, given by [12]

$$\mathbf{s}_k = \begin{bmatrix} v_{a,k} \\ v_{b,k} \\ v_{c,k} \end{bmatrix} = \begin{bmatrix} V_a \cos(\omega k + \phi_a) \\ V_b \cos\left(\omega k + \phi_b - \dfrac{2\pi}{3}\right) \\ V_c \cos\left(\omega k + \phi_c + \dfrac{2\pi}{3}\right) \end{bmatrix}. \tag{46}$$

The amplitudes of the phase voltages $v_{a,k}$, $v_{b,k}$, $v_{c,k}$, are V_a, V_b, V_c, while the phase values are denoted by ϕ_a, ϕ_b, ϕ_c and the angular frequency is $\omega = 2\pi fT$, with f the fundamental power system frequency. Observe that both the frequency ω and phasors (amplitude and phases) are assumed to be identical over a local area.

The three-phase representation of the $\mathbf{s}_k$ in (46) is over-parametrised and is routinely represented as a compact "two-phase" Clarke voltage, $v_{\alpha,k}$ and $v_{\beta,k}$, via the Clarke transform, given by [3]

$$\begin{bmatrix} v_{\alpha,k} \\ v_{\beta,k} \end{bmatrix} \stackrel{\text{def}}{=} \underbrace{\sqrt{\frac{2}{3}} \begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix}}_{\text{Clarke matrix}} \begin{bmatrix} v_{a,k} \\ v_{b,k} \\ v_{c,k} \end{bmatrix}. \tag{47}$$

Moreover, the Clarke transform enables, $v_{\alpha,k}$ and $v_{\beta,k}$, to be conveniently represented jointly as a complex-valued scalar,

$$s_k \stackrel{\text{def}}{=} v_{\alpha,k} + j v_{\beta,k}. \tag{48}$$

The complex $\alpha\beta$ voltage in (48) therefore admits a widely linear auto-regressive (WLAR) representation given by Xia et al. [2,3]

$$s_k = h^* s_{k-1} + g^* s^*_{k-1}, \tag{49}$$

where the WLAR coefficients h and g contain the information of the system frequency, ω_i, and level of imbalance in the system, that is, the degree of improperness. The system frequency is then calculated as [3]

$$e^{j\omega} = \text{Re}\{h\} + j\sqrt{\text{Im}^2\{h\} - |g|^2} \tag{50}$$

$$\Rightarrow \omega = \text{angle}\left\{\text{Re}\{h\} + j\sqrt{\text{Im}^2\{h\} - |g|^2}\right\}. \tag{51}$$

Note that if the system is in a balanced condition, only a single parameter, h, is required to estimate the system frequency (a strictly linear system).

An important task in electricity grids is to estimate the system frequency, ω, given noisy observations of the Clarke voltage s_k in (48), which can be expressed

$$z_{i,k} = s_k + \eta_{i,k}, \tag{52}$$

where $\eta_{i,k}$, is a zero-mean complex-valued white Gaussian noise process, with variance $\sigma^2_{\eta_i} = E|\eta_{i,k}|^2$.

6.2. Balanced multiple node case

The noisy voltage measurements in (52) at each node are given by

$$\mathbf{z}_{i,k} \stackrel{\text{def}}{=} \left[z_{i,k}, \quad z_{i,k+1}, \quad \ldots, \quad z_{i,k+N-1}\right]^{\mathbf{T}}.$$

To construct a classical strictly linear least squares problem, while exploiting all the measurements in the network, we fold a collection of voltage measurements at each node into a single column vector

$$\begin{aligned} \mathbf{z}_{-1} &= \left[\mathbf{z}_{1,k-1}, \quad \cdots, \quad \mathbf{z}_{M,k-1}\right]^{\mathsf{T}}, \\ \mathbf{z} &= \left[\mathbf{z}_{1,k}, \quad \cdots, \quad \mathbf{z}_{M,k}\right]^{\mathsf{T}}, \end{aligned} \tag{53}$$

where the subscript "−1" indicates that the sample at a given time index is delayed by one time instant compared to the vector $\mathbf{z}$. This gives the formulation of the strictly linear least squares solution in the form

$$\hat{h} = (\mathbf{z}^{\mathsf{H}}_{-1}\mathbf{z}_{-1})^{-1}\mathbf{z}^{\mathsf{H}}_{-1}\mathbf{z}, \tag{54}$$

from which the system frequency is calculated as

$$\hat{\omega}_{\text{SL}} = \text{angle}\{\hat{h}\}.$$

6.3. Unbalanced multiple node case

Three-phase systems under unbalanced conditions require a widely linear solution [3], given by

$$\mathbf{z}_{i,k} = h_i \mathbf{z}_{i,k-1} + g_i \mathbf{z}^*_{i,k-1}, \tag{55}$$

where the regression can be represented through the augmented matrix of the vector of delayed system voltages, $\mathbf{z}_{-1}$, in (53), defined as

$$\underline{\mathbf{z}}_{-1} = \left[\mathbf{z}_{-1}, \quad \mathbf{z}^*_{-1}\right],$$

such that the widely linear model of the voltage in (55) assumes an augmented form

$$\mathbf{z} = \underline{\mathbf{z}}_{-1}\mathbf{h}^a, \tag{56}$$

where $\mathbf{h}^a = [h, g]^{\mathsf{H}}$. The WL-LS solution is then given by

$$\hat{\mathbf{h}}^a = \left[h, \quad g\right]^{\mathsf{H}} = \mathbf{Z}^{+}\mathbf{y},$$

where the frequency can be obtained from (51).

6.4. Exploiting redundancy with a WL-CPLS solution

The formulations proposed so far rearrange the data from each node into a single composite vector and, hence, destroy any spatial information. Alternatively, consider the following matrix of regression variables

$$\begin{aligned} \mathbf{Z}_{-1} &= \left[\mathbf{z}_{1,k-1}, \quad \mathbf{z}_{2,k-1}, \quad \ldots, \quad \mathbf{z}_{M,k-1}\right], \\ \mathbf{Z} &= \left[\mathbf{z}_{1,k}, \quad \mathbf{z}_{2,k}, \quad \ldots, \quad \mathbf{z}_{M,k}\right]. \end{aligned}$$

The problem now assumes a multivariate WL-LS form

$$\mathbf{Z} = \left[\begin{array}{cc} \mathbf{Z}_{-1}, & \mathbf{Z}^*_{-1} \end{array}\right]\mathbf{B}^a.$$

Note that $\mathbf{Z}_{-1}$ is a rank-1 matrix (as the system voltage signal at each node should be the same) corrupted by noise, a natural scenario for the WL-CPLS algorithm as the widely linear OLS solution is computationally intractable. The augmented regression coefficient matrix can now be expressed as

$$\mathbf{B}^a = \begin{bmatrix} h_1 & \cdot & \cdots & \cdot \\ \cdot & h_2 & \cdots & \cdot \\ \vdots & \vdots & \ddots & \vdots \\ \cdot & \cdot & \cdots & h_M \\ g_1 & \cdot & \cdots & \cdot \\ \cdot & g_2 & \cdots & \cdot \\ \vdots & \vdots & \ddots & \vdots \\ \cdot & \cdot & \cdots & g_M \end{bmatrix}.$$

From (49), the elements $h_1, h_2, \ldots, h_M$ and $g_1, g_2, \ldots, g_M$ are the coefficients of the WLAR model, for each node, while the matrix $\mathbf{B}^a$ can be estimated using the WL-CPLS algorithm with a single component. This admits a new estimator of the grid frequency in the form

$$\omega = \frac{1}{M}\sum_{i=1}^{M} \text{angle}\left\{\text{Re}\{h_i\} + j\sqrt{\text{Im}^2\{h_i\} - |g_i|^2}\right\}.$$

6.5. Simulation comparison of estimators

The performance of the three system frequency estimators was evaluated for a power grid in an unbalanced condition. Synthetic data were generated for a system with 100 nodes over a range of SNRs. The three phase voltages, v_a, v_b and v_c, were generated as $f = 50Hz$ sinusoids sampled at $f_s = 5kHz$ for $t = 0.3s$ and with a $\frac{2\pi}{3}$ phase difference. The system imbalance was Type B sag which was caused by a resistive fault on v_a causing it to drop to half

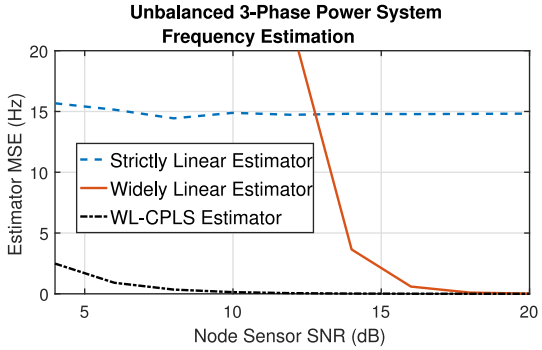

Fig. 7. Comparison of MSE for an unbalanced 3-phase power grid frequency estimator, for a range of SNR.

the amplitude of the other phases, and the fault was propagated over $M = 100$ nodes corrupted with i.i.d. Gaussian noise to give the corresponding SNR. An ensemble MSE of the three system frequency estimators was calculated over 100 realisations over a range of SNR, and the results are given in Fig. 7. Observe that the strictly linear estimator did not have enough degrees of freedom to provide an accurate result. The performance of the widely linear least squares estimator rapidly degraded for low SNRs, exhibiting a "blow up" at SNR=12dB. This can be attributed to the fact that autoregressive modelling of sinusoids (strictly or widely linear) produces biased estimates in the presence of measurement noise [43]. The estimator derived from the WL-CPLS algorithm with only one component, however, provided an accurate estimate even for low SNRs. This demonstrates the ability of the WL-CPLS to find a common subspace between noisy data blocks, a common application in dimensionality reduction [44]. Moreover, the WL-CPLS allows for spatial information to be maintained and, since all the quantities (e.g. phasors, widely linear auto-regressive model) involved are complex-valued, the computation is performed in a physically meaningful manner[6].

7. Conclusion

A widely linear complex partial least squares regression algorithm (WL-CPLS) has been derived as a generalisation of the NIPALS algorithm in [17,27] to noncircular complex data. It has been shown that the WL-CPLS provides a latent variable decomposition of a data matrix, $\mathbf{X}$, which, in turn, admits a tractable computation of the generalised inverses of the required matrices in order to calculate a widely linear regression. Moreover, the proposed algorithm has been shown to extend the cross-covariance criterion of real-valued PLS to suitably select components based on the maximum joint second-order information between the complex-valued input and output blocks. In this way, both significant data components are prioritised and the subspace containing relevant information is identified. The so derived WL-CPLS has been shown to fulfil the requirements for the PLS class of algorithms and its performance has been analysed in terms of the properties of the residuals and through convergence analysis for the case of a univariate output. The utility of the WL-CPLS latent variable decomposition has been demonstrated through its application as a covariance matrix diagonalisation transform, which exhibits useful properties for dimensionality reduction and physically meaningful data analysis. Finally, the benefits from the inherent structure of the complex-valued representation within the WL-CPLS framework have been exemplified through real-world multi-node frequency estimation in unbalanced power grids.

[6] For example, the variable g_i in the WL complex solution indicates whether the system is balanced or not.

Acknowledgements

A. E. Stott was supported by an EPSRC Doctoral Training Award.

Appendix A. SUT and WL-CPLS duality proof condition derivation

To show that both the WL-CPLS transform and the SUT are determined by the conjugate of the matrix $\mathbf{V}$ which arises from the Takagi factorisation of the matrix $\mathbf{P} = \mathbf{X}^{\mathrm{T}}\mathbf{X} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{\mathrm{T}}$, we now derive the condition that the sum $\mathbf{U}_{S1}^{\mathrm{H}}\mathbf{P}^{*}\mathbf{U}_{S1}^{*} + \mathbf{U}_{S1}^{\mathrm{T}}\mathbf{P}\mathbf{U}_{S1} + \mathbf{U}_{S1}^{\mathrm{H}}\mathbf{U}_{S1} + \mathbf{U}_{S1}^{\mathrm{T}}\mathbf{U}_{S1}^{*}$ must be a diagonal matrix. Consider the transform of the original variables $\underline{\mathbf{X}}$ (in augmented form) to the new variables $\tilde{\underline{\mathbf{X}}}$ by means of the matrix $\boldsymbol{\Gamma}_m\mathbf{U}_{Re}$ which, through (40), gives

$$\tilde{\underline{\mathbf{X}}} = \underline{\mathbf{X}}\begin{pmatrix}\mathbf{U}_{S1} & \mathbf{U}_{S2}\\ \mathbf{U}_{S1}^{*} & \mathbf{U}_{S2}^{*}\end{pmatrix}.$$

This is precisely the result produced by the WL-CPLS transform. The covariance matrix of the transformed variables is then given by

$$\begin{aligned}\tilde{\underline{\mathbf{X}}}^{\mathrm{H}}\tilde{\underline{\mathbf{X}}} &= \begin{pmatrix}\mathbf{U}_{S1}^{\mathrm{H}}\mathbf{U}_{S1}^{\mathrm{T}} & \\ \mathbf{U}_{S2}^{\mathrm{H}} & \mathbf{U}_{S2}^{\mathrm{T}}\end{pmatrix}\begin{pmatrix}\mathbf{I} & \mathbf{P}^{*}\\ \mathbf{P} & \mathbf{I}\end{pmatrix}\begin{pmatrix}\mathbf{U}_{S1} & \mathbf{U}_{S2}\\ \mathbf{U}_{S1}^{*} & \mathbf{U}_{S2}^{*}\end{pmatrix},\\ &= \begin{pmatrix}\mathbf{U}_{S1}^{\mathrm{H}}\mathbf{I}\mathbf{U}_{S1} + \mathbf{U}_{S1}^{\mathrm{T}}\mathbf{P}\mathbf{U}_{S1} + \mathbf{U}_{S1}^{\mathrm{H}}\mathbf{P}^{*}\mathbf{U}_{S1}^{*} + \mathbf{U}_{S1}^{\mathrm{T}}\mathbf{I}\mathbf{U}_{S1}^{*} & \\ \mathbf{U}_{S2}^{\mathrm{H}}\mathbf{I}\mathbf{U}_{S1} + \mathbf{U}_{S2}^{\mathrm{T}}\mathbf{P}\mathbf{U}_{S1} + \mathbf{U}_{S2}^{\mathrm{H}}\mathbf{P}^{*}\mathbf{U}_{S1}^{*} + \mathbf{U}_{S2}^{\mathrm{T}}\mathbf{I}\mathbf{U}_{S1}^{*} & \end{pmatrix}\\ &\qquad \begin{matrix} & \mathbf{U}_{S1}^{\mathrm{H}}\mathbf{I}\mathbf{U}_{S2} + \mathbf{U}_{S1}^{\mathrm{T}}\mathbf{P}\mathbf{U}_{S2} + \mathbf{U}_{S1}^{\mathrm{H}}\mathbf{P}^{*}\mathbf{U}_{S2}^{*} + \mathbf{U}_{S1}^{\mathrm{T}}\mathbf{I}\mathbf{U}_{S2}^{*}\\ & \mathbf{U}_{S2}^{\mathrm{H}}\mathbf{I}\mathbf{U}_{S2} + \mathbf{U}_{S2}^{\mathrm{T}}\mathbf{P}\mathbf{U}_{S2} + \mathbf{U}_{S2}^{\mathrm{H}}\mathbf{P}^{*}\mathbf{U}_{S2}^{*} + \mathbf{U}_{S2}^{\mathrm{T}}\mathbf{I}\mathbf{U}_{S2}^{*}\end{matrix}\end{pmatrix}. \end{aligned} \tag{A.1}$$

This matrix is known to be diagonal, which gives rise to the condition that $\mathbf{U}_{S1}^{\mathrm{H}}\mathbf{P}^{*}\mathbf{U}_{S1}^{*} + \mathbf{U}_{S1}^{\mathrm{T}}\mathbf{P}\mathbf{U}_{S1} + \mathbf{U}_{S1}^{\mathrm{H}}\mathbf{U}_{S1} + \mathbf{U}_{S1}^{\mathrm{T}}\mathbf{U}_{S1}^{*}$ must be diagonal. Through inspection, this is satisfied when $\mathbf{U}_{S1} = \mathbf{V}^{*}$.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.sigpro.2018.06.018

References

[1] D.P. Mandic, S. Javidi, S. Goh, A. Kuh, K. Aihara, Complex-valued prediction of wind profile using augmented complex statistics, Renew. Energy 34 (1) (2009) 196–201.
[2] Y. Xia, D.P. Mandic, Widely linear adaptive frequency estimation of unbalanced three-phase power systems, IEEE Trans. Instrum. Meas. 61 (1) (2012) 74–83.
[3] Y. Xia, S.C. Douglas, D.P. Mandic, Adaptive frequency estimation in smart grid applications: exploiting noncircularity and widely linear adaptive estimators, IEEE Signal Process. Mag. 29 (5) (2012) 44–54.
[4] C. Stanciu, J. Benesty, C. Paleologu, T. Gänsler, S. Ciochină, A widely linear model for stereophonic acoustic echo cancellation, Signal Process. 93 (2) (2013) 511–516.
[5] P. Rabiei, W. Namgoong, N. Al-Dhahir, Reduced-complexity joint baseband compensation of phase noise and I/Q imbalance for MIMO-OFDM systems, IEEE Trans. Wirel. Commun. 9 (11) (2010) 3450–3460.
[6] D. Mattera, L. Paura, F. Sterle, MMSE WL equalizer in presence of receiver IQ imbalance, IEEE Trans. Signal Process. 56 (4) (2008) 1735–1740.
[7] D.P. Mandic, V.S.L. Goh, Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models, John Wiley & Sons, 2009.
[8] S. Kanna, S.C. Douglas, D.P. Mandic, A real time tracker of complex circularity, in: Proceedings of the Eighth IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM), 2014, pp. 129–132.
[9] D. Looney, D.P. Mandic, Augmented complex matrix factorisation, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2011, pp. 4072–4075.
[10] D.H. Dini, D.P. Mandic, Class of widely linear complex Kalman filters, IEEE Trans. Neural Netw. Learn. Syst. 23 (5) (2012) 775–786.
[11] B. Picinbono, P. Chevalier, Widely linear estimation with complex data, IEEE Trans. Signal Process. 43 (8) (1995) 2030–2033.
[12] S. Kanna, S.P. Talebi, D.P. Mandic, Diffusion widely linear adaptive estimation of system frequency in distributed power grids, in: Proceeding of the IEEE International Energy Conference (ENERGYCON), 2014, pp. 772–778.

[13] S. Wold, A. Ruhe, H. Wold, W. Dunn III, The collinearity problem in linear regression. the partial least squares (PLS) approach to generalized inverses, SIAM J. Sci. Stat. Comput. 5 (3) (1984) 735–743.
[14] A.E. Hoerl, R.W. Kennard, Ridge regression: biased estimation for nonorthogonal problems, Technometrics 12 (1) (1970) 55–67.
[15] J. Shlens, A tutorial on principal component analysis, arXiv preprint arXiv:1404.1100.
[16] P. Geladi, B. Kowalski, Partial least-squares regression: a tutorial, Anal. Chim. Acta 185 (1986) 1–17.
[17] S. Wold, M. Sjöström, L. Eriksson, PLS-Regression: a basic tool of chemometrics, Chemom. Intell. Lab. Syst. 58 (2) (2001) 109–130.
[18] A.E. Stott, S. Kanna, D.P. Mandic, W.T. Pike, An online NIPALS algorithm for partial least squares, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2017, pp. 4177–4181.
[19] T.J. Cleophas, A.H. Zwinderman, Factor analysis and partial least squares for complex-data reduction (250 Patients), in: Machine Learning in Medicine–Cookbook, Springer, 2014, pp. 43–49.
[20] R. Rosipal, N. Krämer, Overview and recent advances in partial least squares, in: Subspace, Latent Structure and Feature Selection, Springer, 2006, pp. 34–51.
[21] S. De Jong, SIMPLS: an alternative approach to partial least squares regression, Chemom. Intell. Lab. Syst. 18 (3) (1993) 251–263.
[22] A. Cichocki, R. Zdunek, A.H. Phan, S. Amari, Nonnegative Matrix and Tensor Factorisations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation, John Wiley & Sons, 2009.
[23] J. Eriksson, V. Koivunen, Complex-valued ICA using second order statistics, in: Proceedings of the Fourteenth IEEE Signal Processing Society Workshop on Machine Learning for Signal Processing, 2004, pp. 183–192.
[24] L. De Lathauwer, B. De Moor, On the blind separation of non-circular sources, in: Proceedings of the Eleventh European Signal Processing Conference, 2002, pp. 1–4.
[25] C.C. Took, S.C. Douglas, D.P. Mandic, On approximate diagonalisation of correlation matrices in widely linear signal processing, IEEE Trans. Signal Process. 60 (3) (2012) 1469–1473.
[26] P.J. Schreier, A unifying discussion of correlation analysis for complex random vectors, IEEE Trans. Signal Process. 56 (4) (2008) 1327–1336.
[27] H. Wold, Estimation of principal components and related models by iterative least squares, in: Multivariate Analysis, 6, Academic Press, New York, 1966, pp. 391–420.
[28] R. Penrose, A generalized inverse for matrices, in: Mathematical Proceedings of the Cambridge Philosophical Society, 51, Cambridge University Press, 1955, pp. 406–413.
[29] R. Manne, Analysis of two partial-least-squares algorithms for multivariate calibration, Chemom. Intell. Lab. Syst. 2 (1–3) (1987) 187–197.
[30] H. Abdi, Partial least squares regression and projection on latent structure regression (PLS regression), Wiley Interdiscip. Rev. Comput. Stat. 2 (1) (2010) 97–106.
[31] K.B. Petersen, M.S. Pedersen, et al., The Matrix Cookbook 7 (15) (2008) 510.
[32] D.P. Mandic, S. Still, S.C. Douglas, Duality between widely linear and dual channel adaptive filtering, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2009, pp. 1729–1732.
[33] C. Jahanchahi, S. Kanna, D. Mandic, Complex dual channel estimation: cost effective widely linear adaptive filtering, Signal Process. 104 (2014) 33–42.
[34] E. Ollila, On the circularity of a complex random variable, IEEE Signal Process. Lett. 15 (2008) 841–844.
[35] K.S. Miller, On the inverse of the sum of matrices, Math. Mag. 54 (2) (1981) 67–72.
[36] P.J. Schreier, L.L. Scharf, Second-order analysis of improper complex random vectors and processes, IEEE Trans. Signal Process. 51 (3) (2003) 714–725.
[37] R.A. Horn, R.J. Johnson, Matrix Analysis, Cambridge University Press, 1990.
[38] R. Bro, L. Elden, PLS works, J. Chemom. 23 (2) (2009) 69–71.
[39] I.S. Helland, On the structure of partial least squares regression, Commun. Stat. Simul. Comput. 17 (2) (1988) 581–607.
[40] A. Phatak, F. de Hoog, Exploiting the connection between PLS, Lanczos methods and conjugate gradients: alternative proofs of some properties of PLS, J. Chemom. 16 (7) (2002) 361–367.
[41] S. Kanna, D.H. Dini, Y. Xia, S.Y. Hui, D.P. Mandic, Distributed widely linear Kalman filtering for frequency estimation in power networks, IEEE Trans. Signal Inf. Process. Netw. 1 (1) (2015) 45–57, doi:10.1109/TSIPN.2015.2442834.
[42] G.C. Paap, Symmetrical components in the time domain and their application to power network calculations, IEEE Trans. Power Syst. 15 (2) (2000) 522–528, doi:10.1109/59.867135.
[43] H. So, P. Ching, Adaptive algorithm for direct frequency estimation, IEEE Proc. Radar Sonar Navig. 151 (6) (2004) 359–364, doi:10.1049/ip-rsn:20041001.
[44] G. Zhou, A. Cichocki, Y. Zhang, D.P. Mandic, Group component analysis for multiblock data: common and individual feature extraction, IEEE Trans. Neural Netw. Learn. Syst. 27 (11) (2016) 2426–2439.