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, $\mathbf{P} = E[\mathbf{z}\mathbf{z}^{\mathsf{T}}]$, and the standard covariance matrix, $\mathbf{C} = E[\mathbf{z}\mathbf{z}^{\mathsf{H}}]$. 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, $\hat{y} = \mathbf{a}^{\mathsf{T}}\mathbf{x}$, 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 $\hat{y} = \mathbf{h}^{\mathsf{H}}\mathbf{x} + \mathbf{g}^{\mathsf{H}}\mathbf{x}^*$, 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

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]