

On Approximate Diagonalization of Correlation Matrices in Widely Linear Signal Processing

C. Cheong Took, S. C. Douglas, and D. P. Mandic

Abstract—The so called “augmented” statistics of complex random variables has established that both the covariance and pseudocovariance are necessary to fully describe second order properties of noncircular complex signals. To jointly decorrelate the covariance and pseudocovariance matrix, the recently proposed strong uncorrelating transform (SUT) requires two singular value decompositions (SVDs). In this correspondence, we further illuminate the structure of these matrices and demonstrate that for univariate noncircular data it is sufficient to diagonalize the pseudocovariance matrix—this ensures that the covariance matrix is also approximately diagonal. The proposed approach is shown to result in lower computational complexity and enhanced numerical stability, and to enable elegant new formulations of performance bounds in widely linear signal processing. The analysis is supported by illustrative case studies and simulation examples.

Index Terms—Approximate uncorrelating transform, improperness, noncircular complex signals, strong uncorrelating transform, widely linear modeling.

I. INTRODUCTION

Diagonalization of covariance matrices underpins several key areas in statistical signal processing, including blind source separation (BSS) and principal component analysis (PCA). In their most frequently used (and most general) form, diagonalization procedures are applied to covariance matrices of multivariate data (represented by a data matrix), in order to e.g. decorrelate the data channels in BSS type of applications. Diagonalization of correlation matrices that arise from single data channels (data vectors) is usually considered as a special case within this general framework and is used, for instance, in the analysis of convergence of adaptive filters.

The recently introduced augmented complex statistics have played a pivotal role in the treatment of second order noncircular (improper) processes in statistical signal processing [1], [2], allowing us to utilise both the covariance \mathbf{C}_x and the pseudocovariance \mathbf{P}_x matrices, given by

$$\mathbf{R}_x = \begin{pmatrix} E\{x(t)\chi(t)\} & \cdots & E\{x(t)\chi(t+L)\} \\ \vdots & \ddots & \vdots \\ E\{x(t+L)\chi(t)\} & \cdots & E\{x(t+L)\chi(t+L)\} \end{pmatrix}$$

where $\mathbf{C}_x = \mathbf{R}_x$ when $\chi = x^*$ and $\mathbf{P}_x = \mathbf{R}_x$ when $\chi = x$, and L is the length of the data segment considered (e.g. filter length). An important contribution to augmented statistics is the strong uncorrelating transform (SUT) [3], which allows for the covariance and pseudocovariance matrices to be diagonalized simultaneously [4]. It employs

Manuscript received June 27, 2011; revised October 10, 2011; accepted November 22, 2011. Date of publication December 07, 2011; date of current version February 10, 2012. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Suleyman S. Kozat.

C. Cheong Took and D. P. Mandic are with the Department of Electrical and Electronic Engineering, Imperial College London, SW7 2AZ, U.K. (e-mail: c.cheong-took@imperial.ac.uk; d.mandic@imperial.ac.uk).

S. C. Douglas is with the Department of Electrical Engineering, Southern Methodist University, Dallas, TX 75275-0338 USA (e-mail: douglas@enr.smu.edu).

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

Digital Object Identifier 10.1109/TSP.2011.2178603

two singular value decompositions (SVDs) and imposes strict requirements on the structure of diagonalized covariance and pseudocovariance matrices—unit data channel powers for the covariance matrix and ordered circularity coefficients for the pseudocovariance matrix—and has been used in a variety of applications [3], [5].

The SUT method was designed specifically for multivariate random vectors described by the data matrix¹ $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]^T$; however, correlation matrices of univariate data (be they proper or improper) have simpler structures than those of multivariate data and are of particular interest in many applications, including single channel noise reduction [6], beamforming [7], and communications [8]. Our aim is, therefore, to exploit their univariate natures in order to introduce a new approximate method for a joint diagonalization of the covariance and pseudocovariance matrix, which does not alter the fundamental properties of the original correlation matrices. The proposed method makes use of the eigenstructures of the quadratic forms of correlation matrices, and requires only one SVD, thus offering reduced computational complexity as compared with SUT. It is also shown that the new factorization allows us to express several performance bounds in widely linear signal processing in a simpler and more intuitive form. The analysis is supported by simulations over a range of synthetic and real-world signals with varying degrees of circularity.

II. FACTORIZATION OF COMPLEX COVARIANCE MATRICES

The SUT was introduced in [3] to decorrelate² multivariate complex-valued data, by simultaneously diagonalizing the covariance, \mathbf{C}_x , and pseudocovariance, \mathbf{P}_x , matrices as

$$\mathbf{C}_x^d = \Phi \mathbf{C}_x \Phi^H = \mathbf{I} \quad \mathbf{P}_x^d = \Phi \mathbf{P}_x \Phi^T = \Lambda_p \quad (1)$$

where the symbols \mathbf{C}_x^d and \mathbf{P}_x^d denote the decorrelated (and, hence, diagonalized) covariance and pseudocovariance matrices, and \mathbf{I} , Λ_p , and Φ are, respectively, the identity matrix, a real-valued positive diagonal matrix, and the strong uncorrelating transform. The SUT can be computed as

$$\Phi = \mathbf{U}_p^H \mathbf{C}_x^{-\frac{1}{2}} \quad (2)$$

where $\mathbf{C}_x^{-\frac{1}{2}} = \mathbf{U}_c \Lambda_c^{-\frac{1}{2}} \mathbf{U}_c^H$, and the unitary matrices \mathbf{U}_p and \mathbf{U}_c and the diagonal matrix Λ_c are obtained from the following singular value decompositions:

$$\begin{aligned} \mathbf{C}_x &= \mathbf{U}_c \Lambda_c \mathbf{U}_c^H \\ \mathbf{C}_x^{-\frac{1}{2}} \mathbf{P}_x \mathbf{C}_x^{-\frac{1}{2}T} &= \mathbf{U}_p \Lambda_p \mathbf{U}_p^T \end{aligned} \quad (3)$$

In doing so, SUT modifies the structure of the original correlation matrices: via $\mathbf{C}_x^{-\frac{1}{2}}$, it first decorrelates the data channels from one another, it then decorrelates the real and imaginary parts of the individual data channels, and finally it rotates the distributions in order to align them along the real and imaginary axes (via \mathbf{U}_p^H). This way, the entries in the diagonalized correlation matrices are ordered from most to least improper and the powers in the data channels are scaled to unity.

Notice that SUT, being SVD based, is not unique and that, for all its usefulness in blind signal processing, it may not be the most intuitive

¹Vectors \mathbf{x}_i , $i = 1, \dots, N$ are column vectors comprising the data channels, $\mathbf{x}_i = [x_i(t), \dots, x_i(t+T)]^T$, $i = 1, \dots, N$.

²Observe that zero correlation between the real and imaginary parts of a complex variable implies their orthogonality and any decorrelating technique thus affects the phase information of the complex variable.

diagonalization tool for adaptive signal processing for the following reasons.

- 1) Singular values of the ‘‘diagonalized’’ covariance matrix \mathbf{C}_x^d in (1) are restricted to unity, and, hence, do not contain information about the powers in the actual data channels.
- 2) Singular values of the ‘‘diagonalized’’ pseudocovariance $\hat{\mathbf{P}}_x^d$ in (1) pertain to those of $\mathbf{C}_x^{-\frac{1}{2}} \mathbf{P}_x \mathbf{C}_x^{-\frac{1}{2}T}$ and not to those of \mathbf{P}_x (see also (3)).
- 3) From (3), SUT requires two SVDs for \mathbf{C}_x and \mathbf{P}_x .

For univariate complex processes, we propose to address these issues by exploiting the eigenstructure of their correlation matrices, in order to introduce a new diagonalization procedure that i) requires only a single SVD on the pseudocovariance matrix; ii) preserves the original eigenvalues of $E\{\mathbf{x}\mathbf{x}^H\}$ and $E\{\mathbf{x}\mathbf{x}^T\}$, which is beneficial when working on real world problems; and iii) helps to simplify, and to make more intuitive, various statistical analyses and performance bounds in widely linear signal processing.

It should be mentioned that an efficient technique to compute the SUT was proposed in [9] and is referred to as the generalized uncorrelating transform (GUT). The method finds directly the subspace of the matrix pseudoproduct of the covariance and pseudocovariance matrices and achieves computational complexity reduction equivalent to the computation of one SVD, thus satisfying the issue 3) above. Notice that GUT keeps the same eigenstructure as SUT and does not address the issues 1) and 2) above, crucial in the statistical analysis of adaptive filtering algorithms.

A. The Approximate Uncorrelating Transform

Recall that for single channel data (univariate), the covariance and pseudocovariance matrices have the form $\mathbf{C}_x = E\{\mathbf{x}\mathbf{x}^H\}$ and $\mathbf{P}_x = E\{\mathbf{x}\mathbf{x}^T\}$. To show that the diagonalization of both the sample covariance and pseudocovariance matrices³ (of size $L \times L$) requires only one SVD, consider the diagonalized pseudocovariance matrix $\hat{\mathbf{P}}_x^d = \mathbf{U}_p^H \hat{\mathbf{P}}_x \mathbf{U}_p$, where \mathbf{U}_p can be obtained from the Takagi factorization $\hat{\mathbf{P}}_x = \mathbf{U}_p \mathbf{\Lambda}_p \mathbf{U}_p^T$. The quadratic form $\hat{\mathbf{P}}_x^d \hat{\mathbf{P}}_x^{dH}$ is guaranteed to be diagonal, and can be estimated from the data as

$$\begin{aligned}
\hat{\mathbf{P}}_x^d \hat{\mathbf{P}}_x^{dH} &= \left[\frac{1}{T} \sum_{k=1}^T \mathbf{x}(t+k) \mathbf{x}^T(t+k) \right] \\
&\quad \times \left[\frac{1}{T} \sum_{l=1}^T \mathbf{x}^*(t+l) \mathbf{x}^H(t+l) \right] \\
&= \frac{1}{T^2} \sum_{i=1}^T \sum_{j=1}^T \mathbf{x}(t+k) \\
&\quad \times \underbrace{\left(\mathbf{x}^T(t+k) \mathbf{x}^*(t+l) \right)}_{\epsilon_{kl}} \mathbf{x}^H(t+l) \\
&\approx \frac{1}{T^2} \sum_{i=1}^T \sum_{j=1}^T \mathbf{x}(t+k) \\
&\quad \times \underbrace{\left(\mathbf{x}^H(t+k) \mathbf{x}(t+l) \right)}_{\approx \epsilon_{kl}} \mathbf{x}^H(t+l) \\
&\approx \left[\frac{1}{T} \sum_{k=1}^T \mathbf{x}(t+k) \mathbf{x}^H(t+k) \right] \\
&\quad \times \left[\frac{1}{T} \sum_{l=1}^T \mathbf{x}(t+l) \mathbf{x}^H(t+l) \right] \\
&\approx \hat{\mathbf{C}}_x^d \hat{\mathbf{C}}_x^{dH}. \tag{4}
\end{aligned}$$

³An implicit assumption is second order stationarity over T samples.

Therefore, the diagonalization of the pseudocovariance matrix at the same time implies an approximately diagonalized covariance matrix, requiring only one SVD for their joint diagonalization. This is possible⁴ because for $k = l$ the scalar ϵ_{kl} in (4) is real valued, allowing us to substitute $\epsilon_{kl} = \mathbf{x}^T(t+k) \mathbf{x}^*(t+l) = \mathbf{x}^H(t+k) \mathbf{x}(t+l)$, whereas for $k \neq l$ and $\forall k \leq l$, $\epsilon_{kl} = \mathbf{x}^T(t+k) \mathbf{x}^*(t+l) \approx 0$, since $\hat{\mathbf{P}}_x^d$ is diagonal.

To validate this approximation, recall that for $x = x_r + jx_i$, the ‘covariance’ term $xx^* = x_r^2 + x_i^2$ is independent of the improperness of the signal, the ‘pseudocovariance’ term⁵ is $xx^T = x_r^2 - x_i^2 + 2jx_r x_i$, and the correlation coefficient between the real and imaginary part is given by $\rho = \frac{E\{x_r x_i\}}{(\sigma_{x_r} \sigma_{x_i})}$ where σ_{x_r} and σ_{x_i} denote the respective standard deviations. Consider now the following two special cases for improper signals:

- 1) For $\rho = 1$, that is, a perfect correlation between the real x_r and imaginary x_i parts of x (a strongly improper signal), the term $\epsilon_{kl} = \mathbf{x}^T(t+k) \mathbf{x}^*(t+l) = \sum_{\ell=1}^L x_\ell(t+k) x_\ell^*(t+l)$ in (4) is a real scalar and the approximation in (4) holds, because

$$\begin{aligned}
E\{x(k)x^*(l)\} &= E\{x_r(k)x_r(l)\} + E\{x_i(k)x_i(l)\} \\
&\quad + j(E\{x_r(l)x_i(k)\} - E\{x_r(k)x_i(l)\}) \\
&\approx E\{x_r(k)x_r(l)\} + E\{x_i(k)x_i(l)\} \\
&\quad + j(E\{x_i(l)x_i(k)\} - E\{x_i(k)x_i(l)\}) \\
&\approx E\{x_r(k)x_r(l)\} + E\{x_i(k)x_i(l)\} \in \mathbb{R}. \tag{5}
\end{aligned}$$

- 2) For $\rho = 0$, that is, for a doubly white signal, from (5) we have $E\{x_r(l)x_i(k)\} = E\{x_r(k)x_i(l)\} \approx 0$ and the approximation in (4) holds, and similarly for either $x_i = 0$ or $x_r = 0$, where ϵ_{kl} is real.

Without loss in generality, the approximation error in a general case can be expressed as $\varepsilon = \xi_{kl} \sqrt{1 - \rho^2}$ (see the Appendix for more details). Observe that ξ_{kl} can be assumed to be negligible when compared with the diagonal elements of the correlation matrices; the approximation in (4) thus holds, which is confirmed in Section IV. Notice that for $\rho = 1$, the approximation becomes exact.

Remark 1: We have shown that both the pseudocovariance, \mathbf{P}_x and covariance, \mathbf{C}_x , matrices can be diagonalized by the same unitary matrix \mathbf{Q} , obtained from the singular value decomposition⁶ of $\mathbf{P}_x = \mathbf{Q} \mathbf{\Lambda}_p \mathbf{Q}^T$. Thus, the proposed approximate uncorrelating transform (AUT) has the form

$$\begin{aligned}
\mathbf{C}_x &\approx \mathbf{Q} \mathbf{\Lambda}_c \mathbf{Q}^H \\
\mathbf{P}_x &= \mathbf{Q} \mathbf{\Lambda}_p \mathbf{Q}^T. \tag{6}
\end{aligned}$$

Remark 2: The proposed diagonalization procedure applies to correlation matrices of univariate processes and not to those of multivariate data. To show this, consider a p -variate signal represented by the $p \times T$ matrix $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_p]$, whose $p \times p$ pseudocovariance matrix is $\mathbf{P}_x = E\{\mathbf{X}\mathbf{X}^T\}$. Diagonalizing such a matrix implies that the off-diagonal terms $\hat{\mathbf{P}}_{\mathbf{x}_k \mathbf{x}_l} = \frac{1}{T} \sum_t x_k(t) x_l(t) = 0 \forall k \neq l$, however, this does not mean that all the individual products $x_k(t+i) x_l(t+j) = 0$. These constitute the off-diagonal elements of the $L \times L$ matrix $\mathbf{X}^T \mathbf{X}$, and, hence, the condition $\mathbf{X}^T \mathbf{X}^* \approx \mathbf{X}^H \mathbf{X}$ does not hold for multivariate data, and we require SUT for such diagonalizations.

⁴These conditions also apply for negative indices, as the autocorrelation function is symmetric.

⁵For improper (second-order noncircular) signals, either their real and imaginary parts are correlated, or they are uncorrelated but have different powers.

⁶The SVD of a symmetric matrix, such as \mathbf{P}_x , is known as the Takagi/Autonne factorization [10], and it ensures that the singular values have been sorted in a descending order.

III. SOME APPLICATIONS OF AUT DIAGONALIZATION

We next briefly illustrate the usefulness of the proposed diagonalization procedure in some standard optimization problems in statistical signal processing of noncircular complex random variables, that is, those with rotation dependent distributions.

A. Convergence of Widely Linear Adaptive Filters

Consider the augmented complex least mean square (ACLMS) algorithm, whose output is governed by the widely linear model $y(t) = \mathbf{x}^H(t)\mathbf{h}(t) + \mathbf{x}^T(t)\mathbf{g}(t)$, and its update is given by [1], [5]

$$\mathbf{w}_a(t+1) = \mathbf{w}_a(t) + \mathbf{x}_a(t)e_{w1}(t) \quad (7)$$

where the filter coefficient vector $\mathbf{w}_a(t) = [\mathbf{h}^T(t)\mathbf{g}^T(t)]^T$, the augmented input $\mathbf{x}_a(t) = [\mathbf{x}^T(t)\mathbf{x}^H(t)]^T$, output error $e_{w1}(t) = d(t) - \mathbf{x}_a^H(t)\mathbf{w}_a(t)$, and the desired signal $d(t) = \mathbf{x}_a^H(t)\mathbf{w}_a^o(t) + q(t)$, with $q(t)$ being doubly white Gaussian noise uncorrelated with $\mathbf{x}(t)$, and $\mathbf{w}_a^o(t)$ the optimum adaptive coefficient vector that minimizes the cost function $J(t) = |e_{w1}^2(t)|$. The evolution of the weight error vector $\mathbf{v}(t) = \mathbf{w}_a(t) - \mathbf{w}_a^o(t)$ is given by

$$E\{\mathbf{v}(t+1)\} = (\mathbf{I} - \mathbf{C}_x^a)E\{\mathbf{v}(t)\} \quad (8)$$

where the augmented covariance matrix

$$\mathbf{C}_x^a = \begin{bmatrix} \mathbf{C}_x & \mathbf{P}_x \\ \mathbf{P}_x^* & \mathbf{C}_x^* \end{bmatrix}$$

comprises both the covariance matrix \mathbf{C}_x and the pseudocovariance matrix \mathbf{P}_x . When analyzing ‘‘convergence in the mean,’’ we need to diagonalize $\mathbf{C}_x^a = \mathbf{U}^a \mathbf{\Lambda}^a \mathbf{U}^{aH}$ in order to determine the stepsize range as a function of eigenvalues of \mathbf{C}_x^a ; however, doing this directly does not allow us to examine how the individual parameters of the covariance matrices \mathbf{C}_x and \mathbf{P}_x affect the convergence. On the other hand, the proposed diagonalization in (6) allows us to express \mathbf{C}_x^a in terms of the eigenvalues of \mathbf{C}_x and \mathbf{P}_x , that is

$$\mathbf{C}_x^a = \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}^* \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda}_c & \mathbf{\Lambda}_p \\ \mathbf{\Lambda}_p & \mathbf{\Lambda}_c \end{bmatrix} \begin{bmatrix} \mathbf{Q}^H & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}^T \end{bmatrix}. \quad (9)$$

The middle term above can be diagonalized as [5]

$$\begin{bmatrix} \mathbf{\Lambda}_c & \mathbf{\Lambda}_p \\ \mathbf{\Lambda}_p & \mathbf{\Lambda}_c \end{bmatrix} = \frac{1}{2} \begin{bmatrix} \mathbf{I} & -\mathbf{I} \\ \mathbf{I} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda}_c + \mathbf{\Lambda}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_c - \mathbf{\Lambda}_p \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{I} \\ -\mathbf{I} & \mathbf{I} \end{bmatrix}$$

giving a simple and intuitive demonstration that the convergence of ACLMS is governed by the condition number of \mathbf{C}_x^a , with the mismatch between covariance and pseudocovariance given by

$$\kappa(\mathbf{C}_x^a) = \frac{\max(\lambda_c) + \max(\lambda_p)}{\min(\lambda_c) - \min(\lambda_p)}. \quad (10)$$

Remark 3: The proposed diagonalization procedure allows us to express the condition number in terms of the true eigenvalues of \mathbf{C}_x and \mathbf{P}_x , instead of those of \mathbf{C}_x^a which are less straightforward to physically interpret in terms of, e.g., the degree of circularity or signal power.

Remark 4: When using SUT for the same analysis, the imposed constraint $\max(\lambda_c) = \min(\lambda_c) = 1$ might not reflect the true dynamics of ACLMS convergence [5]. This is because the SUT-based condition number $\kappa(\mathbf{C}_x^a)$ is not computed based on the eigenvalues of the covariance matrices of the ‘‘actual’’ data, but on the eigenvalues of the covariance matrices of the ‘‘whitened’’ data.

B. Performance Advantage of Widely Linear Models

The performance advantage of the widely linear model $y = \mathbf{h}^H \mathbf{x} + \mathbf{g}^H \mathbf{x}^*$ over the strictly linear model $y = \mathbf{h}^H \mathbf{x}$ when modeling improper processes is given by [11, eq. (2.10)]

$$\delta\epsilon^2 = \epsilon_1^2 - \epsilon_{w1}^2 = [\mathbf{P}_{yx}^* - \mathbf{P}_x^* \mathbf{C}_x^{-1} \mathbf{P}_{yxx}]^H \times [\mathbf{C}_x^* - \mathbf{P}_x^* \mathbf{C}_x^{-1} \mathbf{P}_x]^{-1} [\mathbf{P}_{yx}^* - \mathbf{P}_x^* \mathbf{C}_x^{-1} \mathbf{P}_{yxx}] \quad (11)$$

and evaluates the difference between the corresponding mean squared errors of the strictly linear model (denoted by ϵ_1^2) and the widely linear model (denoted by ϵ_{w1}^2).

Expression (11) is difficult to interpret, and at best, it can be simplified to $\mathbf{w}^H \mathbf{\Pi}^{-1} \mathbf{w} + \epsilon_{w1}^2$, where $\mathbf{\Pi} = \mathbf{C}_x^* - \mathbf{P}_x^* \mathbf{C}_x^{-1} \mathbf{P}_x$ is the Schur complement of \mathbf{C}_x^a . Since $\mathbf{\Pi}$ is a positive-definite matrix, this guarantees that the widely linear model always performs no worse than the strictly linear model. We can now apply the proposed diagonalization in (6) to obtain

$$\begin{aligned} \delta\epsilon^2 &= \mathbf{w}^H \mathbf{Q}^* [\mathbf{\Lambda}_c - \mathbf{\Lambda}_p \mathbf{\Lambda}_c^{-1} \mathbf{\Lambda}_p]^{-1} \mathbf{Q}^T \mathbf{w} \\ &= \tilde{\mathbf{w}}^H \underbrace{[\mathbf{\Lambda}_c - \mathbf{\Lambda}_p \mathbf{\Lambda}_c^{-1} \mathbf{\Lambda}_p]}_{\boldsymbol{\beta}} \tilde{\mathbf{w}} \\ \boldsymbol{\beta} &= \text{diag}(\beta_1, \dots, \beta_L), \quad \beta_i = \left[\lambda_{c,i} - \frac{\lambda_{p,i}^2}{\lambda_{c,i}} \right]^{-1} \\ &= \frac{\lambda_{c,i}}{\lambda_{c,i}^2 - \lambda_{p,i}^2}. \end{aligned} \quad (12)$$

The result in (11) has now been simplified to (12), providing physical meaning to the analysis. For instance, it is now clear that as the degree of noncircularity increases (i.e., an increase in $\lambda_{p,i}$), β_i increases as well, indicating a larger performance advantage of the widely linear over the strictly linear model.

C. Numerical Stability and Complexity Reduction

In the context of single-channel modelling, inversion of correlation matrices is performed routinely, but can be very challenging due to a possibly high condition number or rank deficiency. To provide insight into the sensitivity of such inversions and to have control over numerical stability, it would be advantageous to factorize such matrices into their diagonal forms. We shall now illustrate the usefulness of the proposed AUT relationships in (6) when performing an inverse of the augmented covariance matrix.

The standard inverse of the augmented covariance is given by [7]

$$\begin{bmatrix} \mathbf{C}_x & \mathbf{P}_x \\ \mathbf{P}_x^* & \mathbf{C}_x^* \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A} & \mathbf{D} \\ \mathbf{D}^* & \mathbf{A}^* \end{bmatrix} \quad (13)$$

where $\mathbf{A} = [\mathbf{C}_x - \mathbf{P}_x \mathbf{C}_x^{-1} \mathbf{P}_x^*]^{-1}$ and $\mathbf{D} = \mathbf{A} \mathbf{P}_x \mathbf{C}_x^{*-1}$. To deal with numerical instability, we can perform a further factorization (e.g., as in square-root Kalman filters), as

$$\mathbf{C}_x^{-1} = \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}^* \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} & \boldsymbol{\gamma} \\ \boldsymbol{\gamma} & \boldsymbol{\beta} \end{bmatrix} \begin{bmatrix} \mathbf{Q}^H & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}^T \end{bmatrix} \quad (14)$$

where $\boldsymbol{\beta}$ is defined in (12) and $\boldsymbol{\gamma} = -\boldsymbol{\beta} \mathbf{\Lambda}_p \mathbf{\Lambda}_c^{-1}$. Therefore, to guarantee numerical stability of the inverse, we need to ensure all diagonal terms of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are nonzero, leading to a full rank matrix. Since the submatrices $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are not only diagonal but also real-valued, this condition is met. The additional benefit of this result is its generic nature, as it can be applied in many applications requiring an inverse of the correlation matrices such as in Wiener filters or square root Kalman filters.

The proposed approximate diagonalization also reduces the computational complexity of the inversion of \mathbf{C}_x^a . The usual tradeoff between

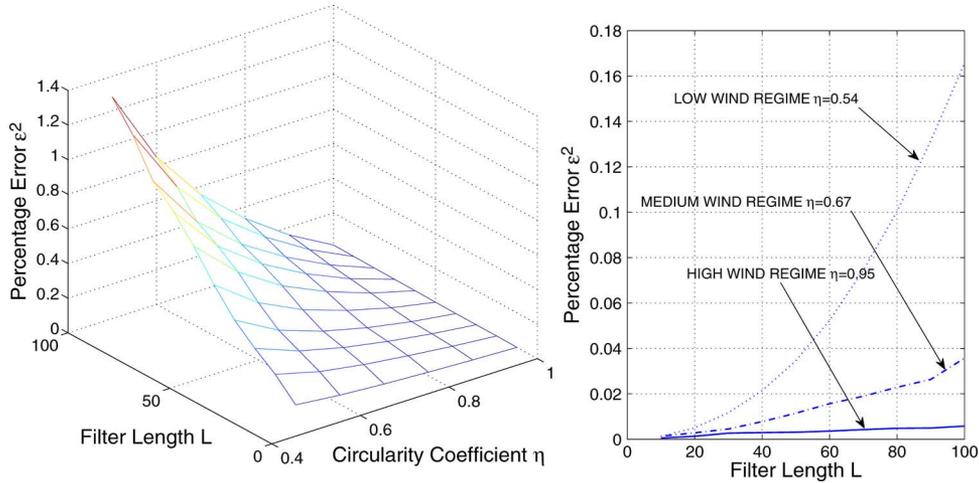


Fig. 1. Performance of the approximate diagonalization in (6) in terms of the percentage error ϵ^2 . Left: for synthetic data as a function of the filter length L and the circularity coefficient η . Right: for the “low”, “medium”, and “high” dynamics of a real world 2D wind signal.

reducing the computational complexity of an algorithm and the accuracy of the algorithm applies, i.e. a decrease in the computational complexity would lead to a less accurate Wiener solution, yet a reasonable approximation due to the negligible error (see Section IV). Also, notice from (17) that when we have perfect correlation ($\rho = 1$) between the real and imaginary part of a complex number, the approximation of widely linear Wiener solution by the proposed AUT technique becomes exact.

D. Extension to Widely Linear Modeling in the Quaternion Domain

Augmented quaternion statistics employ the covariance matrix \mathbf{C}_x and complementary covariance matrices⁷ $\mathbf{C}_x^i = E\{\mathbf{x}\mathbf{x}^{iH}\}$, $\mathbf{C}_x^j = E\{\mathbf{x}\mathbf{x}^{jH}\}$, and $\mathbf{C}_x^k = E\{\mathbf{x}\mathbf{x}^{kH}\}$ [12]. Using (4), it can be shown that $\varepsilon = \mathbf{x}^H \mathbf{x} \approx (\mathbf{x}^H \mathbf{x})^i \approx (\mathbf{x}^H \mathbf{x})^j \approx (\mathbf{x}^H \mathbf{x})^k$ which either vanishes or is real-valued. Following on the proposed result for complex data, one SVD is adequate to diagonalize all the quaternion-valued correlation matrices,⁸ which helps in the analysis of quaternion-valued adaptive filters [13].

IV. SIMULATIONS

The performance index used was a ratio of the powers of the off-diagonal, c_{ij} , versus diagonal, c_{ii} , elements of the correlation matrix, defined as $\varepsilon^2 = \frac{E\{|c_{ij}|^2\}}{E\{c_{ii}^2\}} \times 100 \forall i \neq j$ and the performance was assessed against: i) size of the correlation matrices L and ii) degree of circularity

$$\eta = |E\{x^2\}| / E\{|x|^2\} \quad (15)$$

where $\eta = 1$ corresponds to a highly second-order noncircular (improper) signal, whereas $\eta = 0$ corresponds to a circular, proper, signal.

In the first set of simulations, 100 independent trials were averaged for an MA synthetic signal with a varying degree of noncircularity. It was generated by filtering a complex doubly white circular or noncircular Gaussian signal with a moving average filter whose coefficients were taken from a uniform distribution within $[-0.5, 0.5]$. The left-hand

⁷Quaternion involution operation $(\cdot)^\eta \forall \eta \in \{i, j, k\}$ corresponds to keeping the real and the η -imaginary part intact and conjugating the two other imaginary parts, see [12] for more detail.

⁸Notice that real and complex numbers are special instances of quaternions, and as such this generic quaternion result decomposes into the complex solution, provided the same two imaginary parts of the quaternion variables vanish.

plot in Fig. 1 illustrates an excellent performance of the proposed approximation; the diagonalization error had a very small value of 1.2% even for a large filter length of $L = 100$ and only a moderately non-circular signal. The diagonalization error ε stems from the imaginary part of the approximation of the term ε_{kl} in (4), i.e., $\Im\{\sum_{\ell=1}^L x_\ell(t+k)x_\ell^*(t+l)\}$, which explains the behavior in Fig. 1 for large L . The Appendix shows that ε is a decreasing function of the circularity coefficient η , that is, $\varepsilon \sim \sqrt{1 - \eta^2}$.

The right-hand plot in Fig. 1 illustrates the performance of the proposed diagonalization method for a two-dimensional wind signal w , where the speed v and direction θ were measured at 50 Hz using a 2D ultrasonic anemometer, and the signal was made complex valued as $w = v e^{j\theta}$. The degree of noncircularity of the “high” wind regime was the highest ($\eta = 0.95$), followed by the “medium” ($\eta = 0.67$) and “low” ($\eta = 0.54$) regimes. Observe a near perfect approximate diagonalization for small sizes of correlation matrices ($L < 20$) and an excellent approximation for large sizes of correlation matrices ($L \rightarrow 100$). The approximation was better for signals with higher degree of noncircularity, conforming with the analysis.

V. CONCLUSION

We have demonstrated that a single singular value decomposition of the pseudocovariance matrix is sufficient to simultaneously diagonalize both the covariance and pseudocovariance matrices of single channel noncircular complex data. This has been achieved by examining the structure of the quadratic form of the sample pseudocorrelation matrix. The proposed approximate diagonalization has been shown to preserve the original eigenvalues of the covariance and pseudocovariance matrix, and to make several standard analyses of performance bounds in widely linear modelling simpler and more intuitive. Case studies in the context of widely linear adaptive filtering, together with illustrative simulations, support the approach.

APPENDIX

Without loss in generality assume that the real and imaginary part of a complex random variable $x = x_r + jx_i$ are generated from zero mean, unit variance, uncorrelated variables x_1 and x_2 as

$$x_r(t) = x_1(t) \quad x_i(t) = \rho x_1(t) + x_2(t) \sqrt{1 - \rho^2} \quad (16)$$

where the correlation coefficient $\rho = \frac{E\{x_1 x_2\}}{(\sigma_{x_1} \sigma_{x_2})}$ allows us to control the correlation between x_r and x_i between uncorrelated ($\rho = 0$) to fully

correlated ($\rho = 1$). From (5), the diagonalization error arises from the imaginary part of ϵ_{kl} , that is, $\alpha = \Im\{\epsilon_{kl}\} = x_r(l)x_i(k) - x_r(k)x_i(l)$. We can now write

$$\begin{aligned} \alpha &= x_1(l) \left(\rho x_1(k) + \sqrt{1 - \rho^2} x_2(k) \right) \\ &\quad - x_1(k) \left(\rho x_1(l) + \sqrt{1 - \rho^2} x_2(l) \right) \\ &= \sqrt{1 - \rho^2} (x_1(l)x_2(k) - x_2(l)x_1(k)) = \xi_{kl} \sqrt{1 - \rho^2}. \end{aligned} \quad (17)$$

For the case $E\{x_r^2\} = E\{x_i^2\} = 1$, the circularity coefficient $\eta = \rho$, and the diagonalization error increases with a decrease in η . It is possible to show that the same diagonalization principle also applies to the correlation matrices in augmented quaternion statistics [12].

REFERENCES

- [1] D. P. Mandic and V. S. L. Goh, *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models*. New York: Wiley, 2009.
- [2] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*. Cambridge, U.K.: Cambridge Univ. Press, 2010.
- [3] J. Eriksson and V. Koivunen, "Complex random vectors and ICA models: Identifiability, uniqueness, and separability," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1017–1029, 2006.
- [4] L. DeLathauwer and B. DeMoore, "On the blind separation of noncircular source," in *Proc. 11th Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2002, vol. 2, pp. 99–102.
- [5] S. C. Douglas and D. P. Mandic, "Performance analysis of the conventional complex LMS and augmented complex LMS algorithms," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2010, pp. 3794–3797.
- [6] J. Benesty, J. Chen, and Y. Huang, "A widely linear distortionless filter for single-channel noise reduction," *IEEE Signal Process. Lett.*, vol. 5, no. 17, pp. 469–472, 2010.
- [7] P. Chevalier and A. Blin, "Widely linear MVDR beamformers for the reception of an unknown signal corrupted by noncircular interferences," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5323–5336, 2007.
- [8] H. Gerstacker, R. Schober, and A. Lampe, "Receivers with widely linear processing for frequency-selective channels," *IEEE Trans. Commun.*, vol. 51, no. 9, pp. 1512–1523, 2003.
- [9] E. Ollila and V. Koivunen, "Complex ICA using generalized uncorrelating transform," *Signal Process.*, vol. 89, no. 4, pp. 365–377, 2009.
- [10] R. A. Horn, *Topics in Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1991.
- [11] B. Picinbono and P. Chevalier, "Widely linear estimation with complex data," *IEEE Trans. Signal Process.*, vol. 43, no. 8, pp. 2030–2033, 1995.
- [12] C. Cheong Took, D. P. Mandic, and F. Zhang, "On the unitary diagonalisation of a special class of quaternion matrices," *Appl. Math. Lett.*, vol. 24, no. 11, pp. 1806–1809, 2011.
- [13] C. Cheong Took and D. P. Mandic, "A quaternion widely linear adaptive filter," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4427–4431, 2010.

Global Stabilization of the Least Mean Fourth Algorithm

Eweda Eweda

Abstract—The least mean fourth algorithm has several stability problems. Its stability depends on the variance and distribution type of the adaptive filter input, the noise variance, and the initialization of the filter weights. The present correspondence provides a global solution to all these stability problems. This is achieved by normalizing the weight vector update term by a term that is fourth order in the regressor and second order in the estimation error. The former property stabilizes the algorithm against the variance and distribution type of the filter input, while the latter stabilizes the algorithm against the noise variance and the weight initialization. The obtained algorithm is stable for all values of the step-size between 0 and 2. The stability of the algorithm is supported by simulations.

Index Terms—Adaptive filtering, least mean fourth algorithm, normalized least mean fourth algorithm.

I. INTRODUCTION

The least mean fourth (LMF) algorithm [1]–[20] outperforms the well-known least mean square (LMS) algorithm [1], [2] in achieving a better tradeoff between the transient and steady-state performances of the adaptive filter. However, the LMF algorithm has several stability problems that may put a limitation to its use in applications. In [3], it is shown that the stability of the algorithm around the Wiener solution depends on both the input power of the adaptive filter and the noise power. In [7], [10], and [12], it is shown that the stability of the algorithm depends also on the initial setting of the adaptive filter weights. More recently, it is shown in [13], [15], and [16] that the LMF algorithm with unbounded regressors is not mean square stable whatever small is the algorithm step size. The above arguments show that the stabilization of the LMF algorithm is more complicated than that of the LMS algorithm, whose stability, for a given step-size, depends solely on the input power of the adaptive filter.

To improve the stability of the LMF algorithm, normalized versions of the algorithm have been introduced in the literature [11], [14], [17]–[19]. In spite of the improved performance, none of the available normalized LMF (NLMF) algorithms provides a global remedy to the above mentioned stability problems. This is explained as follows. In [11] and [17], the weight vector update term of the LMF algorithm is normalized by the squared norm of the regressor. The stability of this algorithm is shown to be dependent on the input signal power [19], [20]. In [14] and [18], the weight vector update term of the LMF algorithm is normalized by a weighted sum of the squared norm of the regressor and the squared norm of the error vector. It is shown that this algorithm diverges when the input power exceeds a threshold value that depends on the algorithm step-size [19]. This problem takes place in the above NLMF algorithms because the numerator of the weight vector update term is fourth order in the regressor, while the denominator is second order in the regressor. To overcome this problem, an NLMF algorithm in which the weight vector update term is normalized by the fourth power of the norm of the regressor is proposed in [19]. This normalization results in an improvement of the

Manuscript received October 10, 2011; revised November 18, 2011; accepted November 19, 2011. Date of publication December 02, 2011; date of current version February 10, 2012. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jonathon A. Chambers.

The author is with Department of Electrical Engineering, Ajman University of Science and Technology, Ajman, United Arab Emirates (e-mail: eweda@ieee.org).

Digital Object Identifier 10.1109/TSP.2011.2177976