Exploiting Sparsity in Widely Linear Estimation

Dahir H. Dini and Danilo P. Mandic

Department of Electrical and Electronic Engineering
Imperial College London, Exhibition Road, London, SW7 2BT
Email:{dahir.dini, d.mandic}@imperial.ac.uk

Abstract—**The distribution of complex random signals is typically improper. It has recently been established that conventional strictly linear models are only second order optimum for signals with proper distributions, while so called "widely-linear models" are optimum for the generality of complex signals, both proper and improper. Widely-linear models, however, are over-parameterised when the underlying system is strictly-linear, requiring twice the number of parameters to be estimated compared to strictly-linear models. This effects widely-linear adaptive algorithms, such as the augmented complex least mean square (ACLMS) and augmented complex recursive least squares (ACRLS), and leads to slow convergence. We here address the problem of the over-parameterisation of the ACLMS through the use of regularised cost error functions, and illustrate its effects through analysis and simulations.**

Index Terms—**Widely linear, ACLMS, regularisation**

I. Introduction

Complex signals are used in a variety of applications ranging from power systems to communications systems and military technology [1] [2] [3], owing to their concise natural representation for bivariate data [4]. The inherent transformation they offer between the Cartesian and polar (magnitude and phase) domains, also naturally leads to the generic extensions of real valued signal-processing techniques to their complex valued counterparts.

Complex valued filters have typically been derived as straightforeword extensions of their real valued equivalents, making them suited to a special class of signals known as circular (proper), that is, signals with rotation invariant probability distributions [5]. However, general complex signals assume noncircular (improper) distributions due to imbalance in the powers of the real and imaginary parts or their correlation, finite sample sizes, or nonlinear transformations [4]; which makes conventional complex valued filters inadequate.

Recent advances in augmented complex statistics have highlighted that for a general (improper) zero-mean complex vector $\mathbf{x}$, estimation based on the covariance $\mathbf{R}_{\mathbf{x}} = E\{\mathbf{x}\mathbf{x}^H\}$ is suboptimal and the pseudocovariance $\mathbf{P}_{\mathbf{x}} = E\{\mathbf{x}\mathbf{x}^T\}$ is also required to fully capture the second order statistics. Note that for real signals the covariance and pseudocovariance are equal. Augmented complex valued algorithms incorporate this information, are suited to the generality of complex signals, both proper and improper, and have been shown to outperform their conventional complex counterparts [2] [6].

To illuminate this point, consider the minimum mean square error (MSE) estimator of a zero-mean real valued random vector $\mathbf{y}$ in terms of an observed zero-mean real vector $\mathbf{x}$, that is, $\hat{\mathbf{y}} = E\{\mathbf{y}|\mathbf{x}\}$. For jointly normal $\mathbf{y}$ and $\mathbf{x}$, the optimal linear estimator is

$$\hat{\mathbf{y}} = \mathbf{A}\mathbf{x} \quad (1)$$

where $\mathbf{A} = \mathbf{R}_{\mathbf{yx}}\mathbf{R}_{\mathbf{x}}^{-1}$ is a coefficient matrix, and $\mathbf{R}_{\mathbf{yx}} = E\{\mathbf{y}\mathbf{x}^H\}$. Standard, 'strictly linear' estimation in $\mathbb{C}$ assumes the same model but with complex valued $\mathbf{y}, \mathbf{x}$, and $\mathbf{A}$. However, when $\mathbf{y}$ and $\mathbf{x}$ are jointly improper $\mathbf{P}_{\mathbf{yx}} = E\{\mathbf{y}\mathbf{x}^T\} \neq \mathbf{0}$, and $\mathbf{x}$ is improper $\mathbf{P}_{\mathbf{x}} \neq \mathbf{0}$, then the optimal estimator becomes [5]

$$\hat{\mathbf{y}} = \mathbf{B}\mathbf{x} + \mathbf{C}\mathbf{x}^* = \mathbf{W}\mathbf{x}^a \quad (2)$$

where $\mathbf{B} = \mathbf{R}_{\mathbf{yx}}\mathbf{D} + \mathbf{P}_{\mathbf{yx}}\mathbf{E}^*$ and $\mathbf{C} = \mathbf{R}_{\mathbf{yx}}\mathbf{E} + \mathbf{P}_{\mathbf{yx}}\mathbf{D}^*$ are coefficient matrices, with $\mathbf{D} = (\mathbf{R}_{\mathbf{x}} - \mathbf{P}_{\mathbf{x}}\mathbf{R}_{\mathbf{x}}^{*-1}\mathbf{P}_{\mathbf{x}}^*)^{-1}$ and $\mathbf{E} = -(\mathbf{R}_{\mathbf{x}} - \mathbf{P}_{\mathbf{x}}\mathbf{R}_{\mathbf{x}}^{*-1}\mathbf{P}_{\mathbf{x}}^*)^{-1}\mathbf{P}_{\mathbf{x}}\mathbf{R}_{\mathbf{x}}^{*-1}$, while $\mathbf{x}^a = [\mathbf{x}^T, \mathbf{x}^H]^T$ is the augmented input vector, and $\mathbf{W} = [\mathbf{B}, \mathbf{C}]$ the optimal coefficient matrix. The estimator in (2) is optimal for the generality of complex signals, both proper and improper. Further, the full second order information is contained in the augmented covariance matrix

$$\mathbf{R}_{\mathbf{x}}^a = E\{\mathbf{x}^a\mathbf{x}^{aH}\} = \begin{bmatrix} \mathbf{R}_{\mathbf{x}} & \mathbf{P}_{\mathbf{x}} \\ \mathbf{P}_{\mathbf{x}}^* & \mathbf{R}_{\mathbf{x}}^* \end{bmatrix} \quad (3)$$

and as such, estimation based on $\mathbf{R}_{\mathbf{x}}^a$ incorporates both the covariance and pseudocovariance.

For adaptive algorithms such as the least mean square (LMS), where the data pairs $\mathbf{x}$ and $\mathbf{y}$ are explicitly given, the aim is to estimate the coefficients of the underlying system [7]. Widely linear algorithms are general and cater for both strictly or widely linear system models, that is, the conjugate coefficient $\mathbf{C}$ in (2) converges to zero when the underlying transfer function is strictly linear. However, for the same steady-state performance, the convergence rate of the widely linear (augmented) complex LMS (ACLMS) algorithm is slower than its strictly linear counterpart, the complex LMS (CLMS) [8].

In this paper, we address some convergence issues of the augmented complex LMS (ACLMS) algorithm through the use of widely linear regularised cost functions, analyse the effects of regularisation on the performance of the filter, and provide illustrative simulations to illuminate the analysis.

ISBN 978-3-8007-3529-7

II. Background

A. Complex Least Mean Square (CLMS)

Without loss in generality, consider a measurement equation which relates a desired (observed) signal $d_k \in \mathbb{C}$ at time instant n to a regressor vector $\mathbf{x}_k \in \mathbb{C}^{L\times 1}$ such that

$$d_k = \mathbf{x}_k^H \mathbf{w}^o + q_k \tag{4}$$

where $q_k \in \mathbb{C}$ is a zero-mean white noise process, and $\mathbf{w}^o \in \mathbb{C}^{L\times 1}$ is the weight vector to be estimated. The minimum MSE solution is found by minimising the standard cost function

$$J = E\{e_k e_k^*\} = E\{|e_k|^2\} \tag{5}$$

where the error $e_k = d_k - y_k$ is the difference between the desired signal d_k and the filter output

$$y_k = \mathbf{x}_k^H \mathbf{w} \tag{6}$$

whereby $\mathbf{w}$ is the filter coefficient estimate. The cost function is convex, and the minimum of its derivative with respect to $\mathbf{w}^*$ yields the Wiener solution

$$\hat{\mathbf{w}} = E\{\mathbf{x}_k \mathbf{x}_k^H\}^{-1} E\{d_k \mathbf{x}_k\} \tag{7}$$

In practice, the true statistical moments in the Wiener solution are often unknown and non-stationary. The CLMS is a gradient descent based algorithm, and approximates these moments by their instantaneous estimates. The cost function is redefined to minimise the instantaneous error, that is

$$J_k = e_k e_k^* = |e_k|^2 = |d_k - \mathbf{x}_k^H \mathbf{w}|^2 \tag{8}$$

and is now time varying. The weight update vector can be expressed as

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \mu \nabla_{\mathbf{w}} J_k \big|_{\mathbf{w}=\mathbf{w}_k} \tag{9}$$

where μ is the adaption gain, and $\nabla_{\mathbf{w}} J_k \big|_{\mathbf{w}=\mathbf{w}_k} = -e_k \mathbf{x}_k$ the derivative of J_k with respect to the weight vector [9]. The CLMS is a recursive algorithm, and can be summarised as

$$y_k = \mathbf{x}_k^H \mathbf{w}_k \tag{10}$$

$$e_k = d_k - y_k \tag{11}$$

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu e_k \mathbf{x}_k \tag{12}$$

B. Augmented CLMS (ACLMS)

The ACLMS is the widely linear (augmented) extension of the CLMS, and is suited to estimating the coefficients associated with the more general observation equation

$$d_k = \mathbf{x}_k^H \mathbf{g}^o + \mathbf{x}_k^T \mathbf{h}^o + q_k = \mathbf{x}_k^{aH} \mathbf{w}^{oa} + q_k \tag{13}$$

where $\mathbf{g}^o \in \mathbb{C}^{L\times 1}$ and $\mathbf{h}^o \in \mathbb{C}^{L\times 1}$ are weight vectors to be estimated, and $\mathbf{w}^{oa} = [\mathbf{g}^{oT}, \mathbf{h}^{oT}]^T$ and $\mathbf{x}_k^a = [\mathbf{x}_k^T, \mathbf{x}_k^H]^T$ are the augmented coefficient and input vectors respectively. The ACLMS cost function is of the form

$$\begin{aligned} J_k^{wl} &= e_k e_k^* = |e_k|^2 \\ &= |d_k - \mathbf{x}_k^H \mathbf{g} - \mathbf{x}_k^T \mathbf{h}|^2 = |d_k - \mathbf{x}_k^{aH} \mathbf{w}^a|^2 \end{aligned} \tag{14}$$

where the aim is to find the two weights $\mathbf{w}^a = [\mathbf{g}^T, \mathbf{h}^T]^T$ which minimise the cost function. Following the same derivation as the CLMS, the ACLMS can be summarised as

$$\begin{aligned} y_k &= \mathbf{x}_k^H \mathbf{g}_k + \mathbf{x}_k^T \mathbf{h}_k = \mathbf{x}_k^{aH} \mathbf{w}_k^a \\ e_k &= d_k - y_k \\ \mathbf{w}_{k+1}^a &= \mathbf{w}_k^a + \mu e_k \mathbf{x}_k^a \end{aligned} \tag{15}$$

Equivalently, the update for the two coefficients can be separated, that is

$$\mathbf{g}_{k+1} = \mathbf{g}_k + \mu e_k \mathbf{x}_k \tag{16}$$

$$\mathbf{h}_{k+1} = \mathbf{h}_k + \mu e_k \mathbf{x}_k^* \tag{17}$$

Note that the ACLMS is more general then the CLMS, however, it has a slower convergence rate than the CLMS due to the excess number weights to be estimated.

III. Regularised ACLMS (R-ACLMS)

Regularisation is used to avoid overfitting to a particular dataset by introducing additional information. It is usually implemented as a penalty for complexity, e.g. through bounds on the vector space norm. Examples of regularisation include model order selection techniques, such as the Akaike information criterion (AIC), minimum description length (MDL), and the Bayesian information criterion (BIC); in these cases regularisation is used to find a balance between performance, model order, and coefficient size.

Regularisation can be used to balance between accuracy and model complexity by modifying the cost error function. A regularised version of the cost function (14) is given by

$$J_k^r = J_k^{wl} + \gamma ||\mathbf{w}^a||_p \tag{18}$$

where $||\mathbf{w}^a||_p$ denotes the l_p-norm of $\mathbf{w}^a$, and the term $\gamma \geq 0$ controls the degree of regularisation, for example when $\gamma = 0$ the cost function J_k^r becomes the widely linear cost function J_k^{wl}.

The effect of regularisation in the cost function (18), is to enforce the coefficient estimates $\mathbf{g}$ and $\mathbf{h}$ to their minimum norm, which introduces an estimation bias when the true coefficients $\mathbf{g}^o$ and $\mathbf{h}^o$ are non-zero.

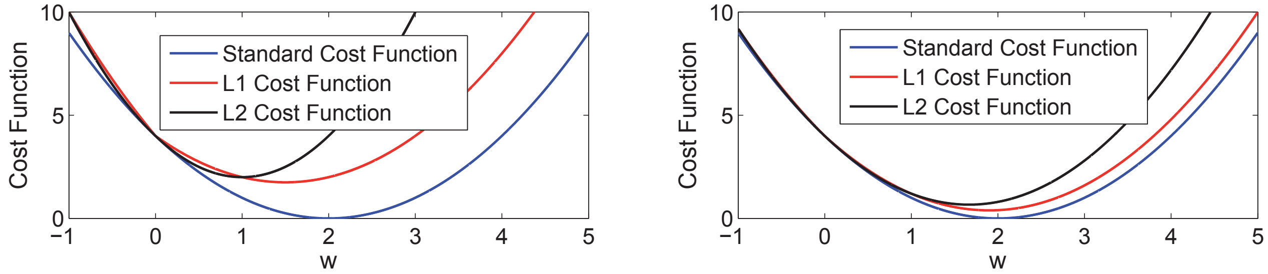
To illuminate this point, consider a real valued noiseless observation equation with a scalar weight, that is,

$$d_k = x_k^* w^o \tag{19}$$

where the optimum weight $w^o = 2$, the input $x_k = 1$ is a constant. The regularised cost function is then given by

$$J_k^r = |d_k - x_k^* w|^2 + \gamma ||w||_p \tag{20}$$

A plot of this cost function for different spans of w is shown in Figure 1, where for $\gamma \neq 0$ the minima of the cost functions do not correspond to the optimum weight $w^o = 2$, introducing bias into the estimation. However, the minima of the regularised cost functions approach w^o as γ is reduced. For $w^o = 0$, the cost function minimums are unbiased for any γ value.



(a) $\gamma = 1$ (b) $\gamma = 0.2$

Fig. 1. Comparison between the l_1- and l_2-norm regularised cost functions for different values of γ. The standard cost function is achieved by setting $\gamma = 0$.

A. Regularised Widely Linear Gradient Descent

The regularised cost function (18), regularises both coefficient vectors $\mathbf{g}$ and $\mathbf{h}$, which introduces an estimation bias when the true coefficients $\mathbf{g}^o$ and $\mathbf{h}^o$ are nonzero. In the case of the ACLMS, we are interested in preventing overfitting when the system to be estimated is strictly linear. Therefore, it is the conjugate weight $(\mathbf{h})$ that needs to be regularised, and the regularised cost function takes the following form:

$$\begin{aligned} J_k^r &= J_k^{wl} + \gamma||\mathbf{h}||_p \\ &= |d_k - \mathbf{x}_k^H\mathbf{g} - \mathbf{x}_k^T\mathbf{h}|^2 + \gamma||\mathbf{h}||_p \end{aligned} \tag{21}$$

whereby the minima of (21) corresponds to the optimum weights when the systems to be estimated is strictly linear, that is $\mathbf{h}^o = \mathbf{0}$; otherwise for widely linear systems $(\mathbf{h}^o \neq \mathbf{0})$, the cost function minima is not aligned with the optimum weights. Based on (21), we have the following time updates for the filter coefficients

$$\begin{aligned} \mathbf{g}_{k+1} &= \mathbf{g}_k + \mu\nabla_{\mathbf{g}} J_k^r|_{\mathbf{g}=\mathbf{g}_k} \\ &= \mathbf{g}_k + \mu e_k \mathbf{x}_k \end{aligned} \tag{22}$$

$$\begin{aligned} \mathbf{h}_{k+1} &= \mathbf{h}_k + \mu\nabla_{\mathbf{h}} J_k^r|_{\mathbf{h}=\mathbf{h}_k} \\ &= \mathbf{h}_k + \mu e_k \mathbf{x}_k^* - \alpha\mathbf{\Sigma}_p(\mathbf{h}_k) \end{aligned} \tag{23}$$

where $\mathbf{\Sigma}_p(\mathbf{h}_k) = \left(\nabla_{\mathbf{h}}||\mathbf{h}||_p\right)\big|_{\mathbf{h}=\mathbf{h}_k} \in \mathbb{C}^{L\times 1}$ denotes the subgradient of the l_p-norm, and the term $\alpha = \mu\gamma$ governs the fraction of the conjugate weight updated due to regularisation. For the remainder of this paper, we will refer to the adaptive filters utilising the regularised update equations (22)–(23) as the regularised-ACLMS (R-ACLMS).

In this work, we restrict our analysis to regularisation involving l_1- and l_2-norms. The l_1-norm subgradient is the component-wise sign function given by

$$\mathbf{\Sigma}_1(u) = \text{sgn}(u) = \begin{cases} u/|u| & \text{if } u \neq 0 \\ 0 & \text{if } u = 0 \end{cases}$$

whereby for a vector input, we have

$$\mathbf{\Sigma}_1(\mathbf{h}_k) = [\text{sgn}(h_k^{(1)}), \cdots, \text{sgn}(h_k^{(L)})]^T \tag{24}$$

with $h_k^{(i)}$ being the ith component of $\mathbf{h}_k$; while, the l_2-norm subgradient is defined as

$$\mathbf{\Sigma}_2(\mathbf{h}_k) = \frac{1}{2}(\mathbf{h}_k^H\mathbf{h}_k)^{-\frac{1}{2}}\mathbf{h}_k = \frac{1}{2}\frac{\mathbf{h}_k}{||\mathbf{h}_k||_2^{1/2}} \tag{25}$$

Remark 1: Note that each component in the l_1-norm gradient, $\mathbf{\Sigma}_1(\mathbf{h}_k)$, consists of the normalisation of $\mathbf{h}_k$ such that each component has unit magnitude; while the l_2-norm gradient, $\mathbf{\Sigma}_2(\mathbf{h}_k)$, consists of the normalisation of the conjugate vector by its l_2-norm.

The augmented form for the R-ACLMS is as follows:

$$y_k = \mathbf{x}_k^H\mathbf{g}_k + \mathbf{x}_k^T\mathbf{h}_k = \mathbf{x}_k^{aH}\mathbf{w}_k^a \tag{26}$$

$$e_k = d_k - y_k \tag{27}$$

$$\mathbf{w}_{k+1}^a = \mathbf{w}_k^a + \mu e_k \mathbf{x}_k^a - \alpha\mathbf{\Delta}_{p,k} \tag{28}$$

where $\mathbf{w}_k^a = [\mathbf{g}_k^T, \mathbf{h}_k^T]^T$, $\mathbf{\Delta}_{p,k} = [\mathbf{0}_L^T, \mathbf{\Sigma}_p(\mathbf{h}_k)^T]^T$, and $\mathbf{0}_L^T$ is a zero column vector of length L. The update for the two coefficients can be separated, as shown in (22)–(23).

B. Cost Function Bias Analysis

For the analysis of the R-ACLMS, we utilise the standard independence assumptions, that is, $\mathbf{x}_k$ is independent and identically distributed in time with augmented covariance $\mathbf{R}_{\mathbf{x}}^a$, and uncorrelated with the white observation noise process q_k.

The derivative of the cost function J_k^r with respect to $\mathbf{w}_k^a$ is given by

$$\nabla_{\mathbf{w}^a} J_k^r|_{\mathbf{w}^a} = -e_k\mathbf{x}_k^a + \gamma\mathbf{\Delta}_p \tag{29}$$

Setting this derivative to zero and rearranging, we have

$$\begin{aligned} \mathbf{w}_{\text{min}}^a &= [\mathbf{x}_k^a\mathbf{x}_k^{aH}]^{-1}[d_k\mathbf{x}_k^a - \gamma\mathbf{\Delta}_p] \\ &= [\mathbf{x}_k^a\mathbf{x}_k^{aH}]^{-1}[\mathbf{x}_k^a\mathbf{x}_k^{aH}\mathbf{w}^{oa} + \mathbf{x}_k^a q_k - \gamma\mathbf{\Delta}_p] \\ &= \mathbf{w}^{oa} + [\mathbf{x}_k^a\mathbf{x}_k^{aH}]^{-1}[\mathbf{x}_k^a q_k - \gamma\mathbf{\Delta}_p] \end{aligned} \tag{30}$$

Utilising the independence assumptions and (30), the weight error becomes

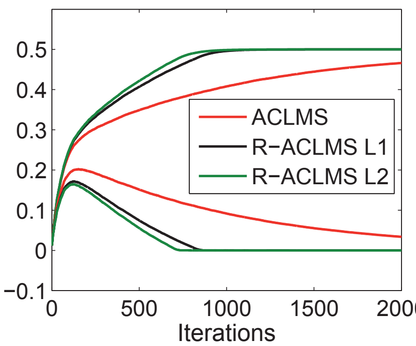
$$\tilde{\mathbf{w}}^a = \mathbf{w}_{\text{min}}^a - \mathbf{w}^{oa} = [\mathbf{x}_k^a\mathbf{x}_k^{aH}]^{-1}[\mathbf{x}_k^a q_k - \gamma\mathbf{\Delta}_p] \tag{31}$$

Finally, applying the expectation operator to both side yields

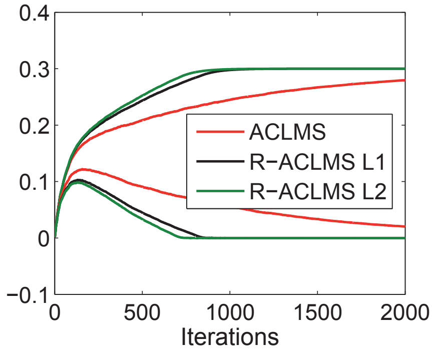
$$\begin{aligned} E\{\tilde{\mathbf{w}}^a\} &= E\{[\mathbf{x}_k^a\mathbf{x}_k^{aH}]^{-1}\mathbf{x}_k^a q_k\} - E\{[\mathbf{x}_k^a\mathbf{x}_k^{aH}]^{-1}\gamma\mathbf{\Delta}_p\} \\ &= -\gamma E\{[\mathbf{x}_k^a\mathbf{x}_k^{aH}]^{-1}\}\mathbf{\Delta}_p \end{aligned} \tag{32}$$

Remark 2: For the minima of the regularised cost function to align with the optimal weight, it is required that $\mathbf{\Delta}_p = [\mathbf{0}_L^T, \mathbf{\Sigma}_p(\mathbf{h})^T]^T = \mathbf{0}$. Based on (24) and (25), this is the case only when $\mathbf{w}^a = \mathbf{0}$, otherwise the weight estimate is biased.

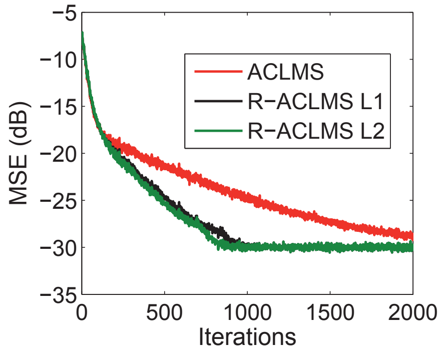
ISBN 978-3-8007-3529-7



(a) Real parts of the g_k and h_k coefficients



(b) Imaginary parts of the g_k and h_k coefficients



(c) MSE performance

Fig. 2. Coefficient convergence of the ACLMS and R-ACLMS for a strictly linear system with a noncircular input signal.

C. Mean Convergence Analysis

We start by substituting the desired signal (13) and R-ACLMS output (26) into the error signal, that is

$$e_k = d_k - y_k = \mathbf{x}_k^{aH}\mathbf{w}^{oa} + q_k - \mathbf{x}_k^{aH}\mathbf{w}_k^a \tag{33}$$

then the weight update (28) can be written as

$$\mathbf{w}_{k+1}^a = \mathbf{w}_k^a + \mu\Big[\mathbf{x}_k^a\mathbf{x}_k^{aH}\mathbf{w}^{oa} + \mathbf{x}_k^a q_k - \mathbf{x}_k^a\mathbf{x}_k^{aH}\mathbf{w}_k^a\Big] - \alpha\boldsymbol{\Delta}_{p,k} \tag{34}$$

Subtracting the optimal weight vector $\mathbf{w}^o$ from both sides of (34) yields

$$\begin{aligned}\tilde{\mathbf{w}}_{k+1}^a &= \mathbf{w}_{k+1}^a - \mathbf{w}^o \\ &= \tilde{\mathbf{w}}_k^a - \mu\mathbf{x}_k^a\mathbf{x}_k^{aH}\tilde{\mathbf{w}}_k^a + \mu\mathbf{x}_k^a q_k - \alpha\boldsymbol{\Delta}_{p,k}\end{aligned} \tag{35}$$

Applying the statistical expectation operator to both sides and employing the independence assumption, we have

$$\begin{aligned}E\{\tilde{\mathbf{w}}_{k+1}^a\} &= (\mathbf{I} - \mu\mathbf{R}_{\mathbf{x}}^a)E\{\tilde{\mathbf{w}}_k^a\} + \mu E\{\mathbf{x}_k^a q_k\} - \alpha E\{\boldsymbol{\Delta}_{p,k}\} \\ &= (\mathbf{I} - \mu\mathbf{R}_{\mathbf{x}}^a)E\{\tilde{\mathbf{w}}_k^a\} - \alpha E\{\boldsymbol{\Delta}_{p,k}\}\end{aligned} \tag{36}$$

Remark 3: Note that by setting $\alpha = 0$, the R-ACLMS become the ACLMS, and the convergence results for the ACLMS apply [8], whereby the ACLMS is stable for

$$0 < \mu < \frac{2}{\lambda_{\max}(\mathbf{R}_{\mathbf{x}}^a)} \tag{37}$$

with $\lambda_{\max}(\mathbf{R}_{\mathbf{x}}^a)$ is the largest eigenvalue of $\mathbf{R}_{\mathbf{x}}^a$. Otherwise, for $\alpha \neq 0$, the R-ACLMS has an extra degree of freedom, and α can be chosen to set performance characteristics.

IV. Application Examples

We firstly illustrate the coefficient convergence of the adaptive filters for the strictly linear system given by

$$d_k = x_k^* w^o + q_k$$

where $w^o = 0.5 + j0.3$, the input x_k was a zero-mean noncircular ($E\{x_k^2\} = 0.9$), unit variance, complex white process, and q_k was a complex circular observation noise with variance 0.001. An adaption gain of $\mu = 0.01$ was chosen for the ACLMS and R-ACLMS algorithms, the regularisation parameter α for the l_1 and l_2 R-ACLMS algorithms were set to 0.0004 and 0.001 respectively, and filter coefficient were initialised to zero.

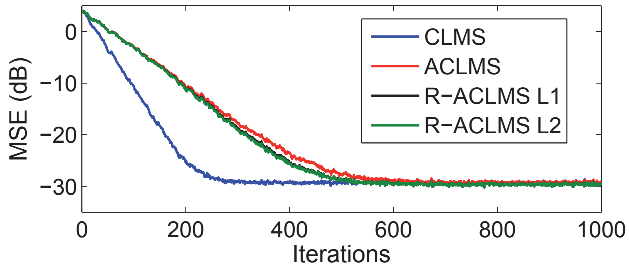
Figure 2 shows the convergence of the real and imaginary parts of the standard weight g_k and conjugate weight h_k, together with the mean square error (MSE) performances. The results show that the R-ACLMS algorithms offer better weights and error convergence rates for noncircular inputs. However, the performance gains of the R-ACLMS algorithms are not inherited in widely linear systems as will be discussed in the following example.

We next consider a system identification problem for (a) a strictly linear system with 15 complex weights (coefficients) and (b) a widely linear system with 15 standard complex weights and 15 conjugate weights. The input vector $\mathbf{x}_k$ was a complex white random process with identity covariance matrix $\mathbf{R}_{\mathbf{x}} = \mathbf{I}$, while the variance of the complex white observation noise q_k and the regularisation parameters for the l_1-norm and l_2-norm R-ACLMS algorithms were as above. In the figures that follow, the mean square error (MSE) of the algorithms were computed by averaging 500 trails.

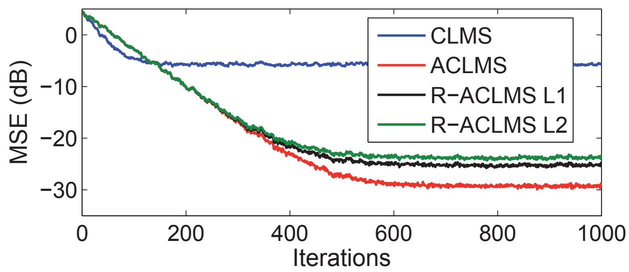
For the set of simulations shown in Figure 3, the input vector had a circular Gaussian distribution, and an adaptation gain of $\mu = 0.01$ was chosen for the CLMS, while for the ACLMS and R-ACLMS algorithms μ was set at half of this value to ensure that all the algorithms have the same steady-state performance as the CLMS. The results for the strictly linear system, show that the CLMS had the best convergence rate [8], while the R-ACLMS algorithms converged slightly faster than the ACLMS. All the algorithms reached similar steady-states, and the R-ACLMS algorithms provided unbiased weight estimates. For the widely linear system in Figure 3b, the ACLMS offered the best steady-state performance due to its unbiased weight estimate, while the R-ACLMS weight estimates were biased (see Remark 2), but outperform the CLMS which under-modelled the widely linear system.

Figure 4 illustrates the case when the input vector had a noncircular Gaussian distribution $E\{\mathbf{x}_k\mathbf{x}_k^T\} = 0.6\mathbf{I}$. Again, the comparative differences between the algorithms were the same: the CLMS had the best convergence rates, ACLMS had the best steady-state MSE for the widely linear system, while the performances of the R-ACLMS algorithms were some-

ISBN 978-3-8007-3529-7

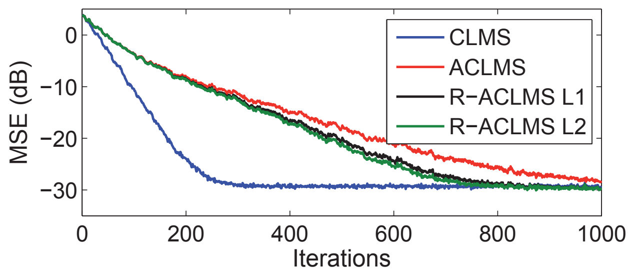


(a) Strictly linear system

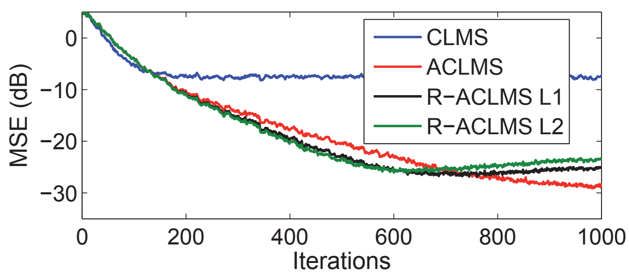


(b) Widely linear system

Fig. 3. Performance comparison between the CLMS, the widely linear ACLMS, the l_1- and l_2-norm regularised ACLMS (R-ACLMS) for strictly and widely linear systems with a circular input vector $E\{\mathbf{x}_k\mathbf{x}_k^T\} = \mathbf{0}$.



(a) Strictly linear system



(b) Widely linear system

Fig. 4. Performance comparison between the CLMS, the widely linear ACLMS, the l_1- and l_2-norm regularised ACLMS (R-ACLMS) for strictly and widely linear systems with a noncircular input vector $E\{\mathbf{x}_k\mathbf{x}_k^T\} = 0.6\mathbf{I}$.

where in between the CLMS and ACLMS. The convergence rates of all the algorithms were effected by the noncircularity of the input [8], and this was especially pronounced for the ACLMS algorithm.

The final set of simulation in Figure 5, show the performance for multistep ahead prediction of real-world Wind data, whereby the filters employed widely linear 4th order autoregressive processes to make predictions. The results show that the R-ACLMS algorithms were better suited to tracking the noncircular and nonstationary Wind data compared with the ACLMS, due to their faster convergence rates.

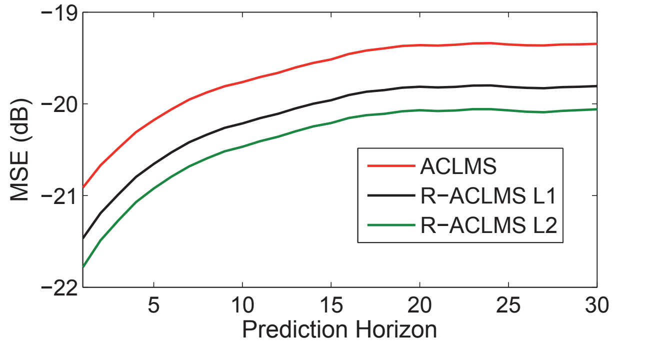


Fig. 5. Performance comparison between the different algorithms for the prediction of real-world Wind data at different prediction horizons.

V. Conclusion

The widely linear (augmented) complex least mean square (ACLMS) is suited to the generality of complex systems, both strictly and widely linear systems, but suffers from slow convergence speeds. In this work, the conjugate weight regularised ACLMS (R-ACLMS) algorithm was presented to address the convergence issues of the ACLMS. The analysis shows that regularisation of the standard cost function introduces a weight estimation bias when the underlying system is widely linear, where the size of the bias is determined by the regularisation factor. Simulation results show that the R-ACLMS converges faster than the ACLMS, and offers similar steady-state performance for strictly linear systems, which makes R-ACLMS algorithms better suited to noncircular nonstationary systems.

References

[1] Y. Xia and S. C. Douglas and D. P. Mandic, "Adaptive frequency estimation in smart grid applications: Exploiting noncircularity and widely linear adaptive estimators," *IEEE Signal Processing Magazine*, In print, 2012.

[2] S. Douglas, "Widely-linear recursive least-squares algorithm for adaptive beamforming," *in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2009*, pp. 2041–2044, 2009.

[3] P.M. Djurić, J.H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M.F. Bugallo, and J. Miguez, "Particle filtering," *IEEE Signal Processing Magazine*, vol. 20, pp. 19 – 38, Sep. 2003.

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

[5] B. Picinbono and P. Bondon, "Second-order Statistics of Complex Signals," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 411–420, 1997.

[6] D. H. Dini and D. P. Mandic, "Class of widely linear complex Kalman filters," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, pp. 775 –786, May 2012.

[7] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*. John Wiley & Sons, 1996.

[8] S.C. Douglas and D. P. Mandic, "Performance analysis of the conventional complex LMS and augmented complex LMS algorithms," in *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pp. 3794–3797, 2010.

[9] D.H. Brandwood, "A complex gradient operator and its application in adaptive array theory," *IEE Proceedings on Communications, Radar and Signal Processing,*, vol. 130, pp. 11 –16, Feb 1983.

ISBN 978-3-8007-3529-7