13

Widely Linear Estimation and Augmented CLMS (ACLMS)

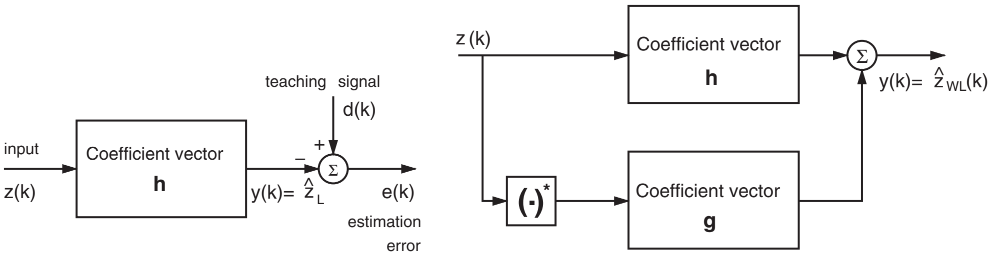
It has been shown in Chapter 12 that the full-second order statistical description of a general complex valued process can be obtained only by using the augmented complex statistics, that is, by considering both the covariance and pseudocovariance functions. It is therefore natural to ask how much we can gain in terms of the performance of statistical signal processing algorithms by doing so. To that end, this chapter addresses linear estimation for both circular and noncircular (proper and improper) complex signals; this is achieved based on a finite impulse response (FIR) system model and for both the second-order regression modelling with fixed coefficients (autoregressive modelling) and for linear adaptive filters for which the filter coefficients are adaptive. Based mainly on the work by Picinbono [239, 240] and Schreier and Scharf [268], Sections 13.1 – 13.3 show that for general complex signals (noncircular), the optimal linear model is the 'widely linear' (WL) model, which is linear both in z and z^*. Next, based on the widely linear model, for adaptive filtering of general complex signals, the augmented complex least mean square (ACLMS) algorithm is derived, and by comparing the performances of ACLMS and CLMS, we highlight how much is lost by treating improper signals in the conventional way.

13.1 Minimum Mean Square Error (MMSE) Estimation in $\mathbb{C}$

The estimation of one signal from another is at the very core of statistical signal processing, and is illustrated in Figure 13.1, where $\{z(k)\}$ is the input signal, $\mathbf{z}(k) = [z(k-1), \ldots, z(k-N)]^T$ is the regressor vector in the filter memory, $d(k)$ is the teaching signal, $e(k)$ is the instantaneous output error, $y(k)$ is the filter output[1], and $\mathbf{h} = [h_1, \ldots, h_N]^T \in \mathbb{C}^{N\times 1}$ is the vector of filter coefficients.

[1]For prediction applications $d(k) = z(k)$ and $y(k) = \hat{z}_L(k)$, where the subscript 'L' refers to the standard linear estimator.

Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models
Danilo Mandic and Vanessa Su Lee Goh



(a) Standard linear filtering model (b) Widely linear filtering model

Figure 13.1 Complex linear estimators

A solution that minimises the mean squared (MS) error is linear regression [110, 113]

$$\hat{z}_{\mathrm{L}}(k) = E[y(k)|\mathbf{z}(k)] \quad \Leftrightarrow \quad y(k) = \mathbf{h}^{\mathrm{T}}\mathbf{z}(k) \tag{13.1}$$

which estimates a scalar random variable y (real or complex) from an observation vector $\mathbf{z}(k)$, and is linear in $\mathbf{z}$ if both y and $\mathbf{z}$ are zero mean, and jointly normal. To perform this linear regression, we need to decide on the order N of the system model (typically a finite impulse response (FIR) system), and also on how to measure best fit of the data (error criterion). The estimator chooses those values of $\mathbf{h}$ which make $z(k)$ closest to $d(k)$, where closeness is measured by an error criterion, which should be reasonably realistic for the task in hand and it should be analytically tractable. Depending on the character of estimation, the commonly used error criteria are:

- *Deterministic error criterion*, given by

$$J = \min_{\mathbf{h}} \sum_{k} |e(k)|^p = \sum_{k=0}^{N-1} \big(d(k) - y(k)\big)^p \tag{13.2}$$

  which for $p = 2$ is known as the *Least Squares (LS) problem*, and its solution is known as the Yule–Walker solution (the basis for autoregressive (AR) modelling in $\mathbb{C}$), given by [33]

$$\mathbf{h} = \mathbf{R}^{-1}\mathbf{r} \tag{13.3}$$

  where $\mathbf{R}$ is the input correlation matrix and $\mathbf{r} = E[z(k)\mathbf{z}^*(k)]$.
- *Stochastic error criterion*, given by

$$J = \min_{\mathbf{h}} E\{|e(k)|^p\} \tag{13.4}$$

For $p = 2$ this optimisation problem is known as the *Wiener filtering problem*, for which the solution is given by Wiener–Hopf equations[2]

$$\mathbf{h} = \mathbf{R}_{\mathbf{z},\mathbf{z}}^{-1}\mathbf{r}_{d,\mathbf{z}} \tag{13.5}$$

where $\mathbf{R}_{\mathbf{z},\mathbf{z}}$ is the tap input correlation matrix and $\mathbf{r}_{d,\mathbf{z}}$ is the cross-correlation vector between the teaching signal and the tap input.

From (13.1) and (13.5), the output of the linear estimator is given by

$$y(k) = \hat{z}_L(k) = \left(\mathbf{R}_{\mathbf{z},\mathbf{z}}^{-1}\mathbf{r}_{d,\mathbf{z}}\right)^{\mathrm{T}} \mathbf{z}(k) \tag{13.6}$$

A stochastic gradient based iterative solution to this problem, which bypasses the requirement of piece-wise stationarity of the signal is the complex least mean square (CLMS) algorithm [307].

13.1.1 Widely Linear Modelling in $\mathbb{C}$

As shown in Chapter 12, for complete second-order statistical description of general complex signals we need to consider the statistics of the augmented input vector (in the prediction setting)

$$\mathbf{z}^a(k) = \left[z(k-1), \ldots, z(k-N), z^*(k-1), \ldots, z^*(k-N)\right]^{\mathrm{T}} = \left[\mathbf{z}^{\mathrm{T}}(k), \mathbf{z}^{\mathrm{H}}(k)\right]^{\mathrm{T}} \tag{13.7}$$

Thus a linear estimator in $\mathbb{C}$ should be linear in both $\mathbf{z}$ and $\mathbf{z}^*$, that is

$$\hat{z}_{WL}(k) = y(k) = \mathbf{h}^{\mathrm{T}}\mathbf{z}(k) + \mathbf{g}^{\mathrm{T}}\mathbf{z}^*(k) = \mathbf{q}^{\mathrm{T}}\mathbf{z}^a(k) \tag{13.8}$$

where $\mathbf{h}$ and $\mathbf{g}$ are complex vectors of filter coefficients and $\mathbf{q} = \left[\mathbf{h}^{\mathrm{T}}, \mathbf{g}^{\mathrm{T}}\right]^{\mathrm{T}}$. Statistical moments of random variable $y(k)$ in Equation (13.8) are defined by the corresponding moments of the augmented input $\mathbf{z}^a(k)$, and the signal model (Equation 13.8) is referred to as a *wide sense linear* or *widely linear* (WL) estimator [239, 240, 271], depicted in Figure 13.1(b).

From Equations (13.6) and (13.8), the optimum widely linear coefficient vector is given by

$$\mathbf{q} = \left[\mathbf{h}^{\mathrm{T}}, \mathbf{g}^{\mathrm{T}}\right]^{\mathrm{T}} = \mathcal{C}_a^{-1}\mathbf{r}_{d,\mathbf{z}^a} \tag{13.9}$$

where $\mathcal{C}_a$ is the augmented covariance matrix given in Equation (12.50) and $\mathbf{r}_{d,\mathbf{z}^a}$ is the cross-correlation vector between the augmented input $\mathbf{z}^a(k)$ and the teaching signal $d(k)$. The widely linear MMSE solution can now be expressed as

$$y(k) = \hat{z}_{WL}(k) = \mathbf{q}^{\mathrm{T}}\mathbf{z}^a(k) = \left(\mathcal{C}_a^{-1}\mathbf{r}_{d,\mathbf{z}^a}\right)^{\mathrm{T}}\mathbf{z}^a(k) = \left(\begin{bmatrix} \mathcal{C} & \mathcal{P} \\ \mathcal{P}^* & \mathcal{C}^* \end{bmatrix}^{-1} \begin{bmatrix} \mathcal{C}_{d,\mathbf{z}} \\ \mathcal{P}_{d,\mathbf{z}} \end{bmatrix}\right)^{\mathrm{T}} \begin{bmatrix} \mathbf{z}(k) \\ \mathbf{z}^*(k) \end{bmatrix} \tag{13.10}$$

[2]To indicate the block nature of the solution, a piece-wise stationary segment of the data is considered.

The widely linear signal model therefore utilises information from both the covariance $\mathcal{C}$ and pseudocovariance $\mathcal{P}$ matrices (given in Equation 12.49) and from ordinary crosscorrelations $\mathcal{C}_{d,\mathbf{z}}$ and $\mathcal{P}_{d,\mathbf{z}}$, and as such it is suitable for the estimation of general complex signals. In the special case when the complementary statistics vanish, that is $\mathcal{P} = 0$ and $\mathcal{P}_{d,\mathbf{z}} = 0$, the widely linear estimator (Equation 13.8) degenerates into the standard linear estimator (Equation 13.1).

Practical applications of wide sense linear filters are only emerging, this is due to the fact that almost all existing applications assume circularity (explicitly or implicitly), however, this assumption cannot be generally accepted. A stochastic gradient based iterative solution to the widely linear estimation problem, which caters for nonstationary signals is called the augmented CLMS (ACLMS) and is introduced in Section 13.4.

13.2 Complex White Noise

Central to the autoregressive modelling and prediction is the concept of white noise;[3] a wide sense stationary signal $z(k)$ is said to be white if its power spectrum $\Gamma(\nu)$ is constant, or equivalently, if its covariance function c_z is a Dirac delta function, that is (see also Equation 12.44)

$$c_z(m) = c_z\delta(m) \quad \Leftrightarrow \quad \Gamma_z(\nu) = \text{const}$$

The concept of white noise is inherited from the analysis of real random variables, however, the fact that the power spectrum of a wide sense stationary white noise is constant does not imply any constraint on the pseudocovariance[4] p_z or spectral pseudocovariance $R_z(\nu) = \mathcal{F}\{p_z\}$ (see also Equation 12.43).

It has been shown in Section 12.4.1 that spectral covariance $\Gamma_z(\nu)$ and spectral pseudocovariance $R_z(\nu)$ of a second-order stationary complex signal z need to satisfy

$$\begin{aligned} \Gamma_z(\nu) &\geq 0 \\ R_z(\nu) &= R_z(-\nu) \\ |R_z(\nu)|^2 &\leq \Gamma_z(\nu)\Gamma_z(-\nu) \end{aligned} \tag{13.11}$$

In other words, for a *second-order stationary white noise* signal we have [239]

$$\begin{aligned} c_z(m) &= c_z\delta(m) \\ |R_z(\nu)|^2 &\leq \Gamma_z(\nu)\Gamma_z(-\nu) = c_z^2 \end{aligned} \tag{13.12}$$

It is often implicitly assumed that the spectral pseudocovariance function $R_z(\nu)$ of a second-order white signal vanishes, however, this would only mean that such white signal is *circular* [239].

- A *second-order circular white noise* signal is characterised by a constant power spectrum and vanishing spectral pseudocovariance function, that is, $\Gamma_z(\nu) = \text{const}$ and $R_z(\nu) = 0$. The real and imaginary parts of circular white noise are white and uncorrelated.

[3] Whiteness, in terms of multicorrelations, has already been introduced in Section 12.3.1.
[4] It can even be nonstationary.

Since the concept of whiteness is intimately related with the correlation structure of a signal (only instantaneous relationships are allowed, that is, there is no memory in the system), whiteness can also be defined in the time domain. One special case of a second-order white signal, which is a direct extension of the real valued white noise, is called *doubly white* noise.

- A second-order white signal is called *doubly white*, if

$$c_z(m) = c_z\delta(m)$$
$$p_z(m) = p_z\delta(m) \tag{13.13}$$

where the only condition on the pseudocovariance function is $|p_z| \le c_z$. The spectral covariance and pseudocovariance functions of doubly white noise are then given by [239]

$$\Gamma_w(\nu) = c_w \qquad \text{and} \qquad R_w(\nu) = p_w \tag{13.14}$$

13.3 Autoregressive Modelling in $\mathbb{C}$

The task of autoregressive (AR) modelling is, given a set of data, to find a regression of order p which approximates the given dataset. The standard autoregressive model in $\mathbb{C}$ takes the same form as the AR model for real valued signals, that is

$$z(k) = h_1 z(k-1) + \cdots + h_p z(k-p) + w(k) = \mathbf{h}^{\mathrm{T}}\mathbf{z}(k), \quad \mathbf{h} \in \mathbb{C}^{N\times 1} \tag{13.15}$$

where $\{z(k)\}$ is the random process to be modelled and $\{w(k)\}$ is white Gaussian noise (also called the driving noise), $\mathbf{h} = \left[h_1, \ldots, h_p\right]^{\mathrm{T}}$ and $\mathbf{z} = \left[z(k-1), \ldots, z(k-p)\right]^{\mathrm{T}}$. If the driving noise is assumed to be doubly white, we need to find the coefficient vector $\mathbf{h}$ and the covariance and pseudocovariance of the noise which provide best fit to the data in the minimum mean square error sense. Equivalently, in terms of the transfer function $H(\nu)$, we have (for more detail see [239])

$$|H(\nu)|^2 = \Gamma(\nu)$$
$$p_w H(\nu)H(-\nu) = R(\nu) \quad \Leftrightarrow \quad |R(\nu)|^2 = |p_w|^2\Gamma(\nu)\Gamma(-\nu) \tag{13.16}$$

In autoregressive modelling, it is usually assumed that the driving noise has zero mean and unit variance, and we can assume $c_w = 1$, which then implies[5] $|p_w| \le 1$. Then, linear autoregressive modelling (based on the deterministic error criterion, Equation 13.2) has a solution only if both Equations (13.11) and (13.16) are satisfied. This happens, for instance, when $R(\nu) = 0$ and $p_w = 0$, that is, when the driving noise is white and second-order circular, which is the usual assumption in standard statistical signal processing literature [110, 113]. In this case, the solution has the same form as in the real case [33], that is

$$\mathbf{h} = \mathbf{R}^{-1}\mathbf{r} \tag{13.17}$$

where $\mathbf{R} = E\left[\mathbf{z}(k)\mathbf{z}^{\mathrm{H}}(k)\right]$ and $\mathbf{r} = E\left[z(k)\mathbf{z}^{*}(k)\right]$.

[5]The pseudocovariance p_w can even be a complex quantity, hence the modulus operator.

Thus, a general complex signal cannot be modelled by a linear filter driven by doubly white noise.

13.3.1 Widely Linear Autoregressive Modelling in $\mathbb{C}$

The widely linear autoregressive (WLAR) model in $\mathbb{C}$ is linear in both $\mathbf{z}$ and $\mathbf{z}^*$, that is

$$\begin{aligned} z(k) &= \sum_{i=1}^{p} h_i z(k-i) + \sum_{i=1}^{p} g_i z^*(k-i) + h_0 w(k) + g_0 w^*(k) \\ &= \mathbf{h}^{\mathrm{T}}\mathbf{z}(k) + \mathbf{g}^{\mathrm{T}}\mathbf{z}^*(k) + [h_0, g_0] w^a(k) \end{aligned} \tag{13.18}$$

and has more degrees of freedom and hence potentially improved performance over the standard linear model. The gain in performance, however, depends on the degree of circularity of the signal at hand.
Autoregressive modelling is intimately related to prediction, that is (since $E[w(k)] = 0$)

$$\hat{z}_{\mathrm{WL}}(k) = E\left[\mathbf{h}^{\mathrm{T}}\mathbf{z}(k) + \mathbf{g}^{\mathrm{T}}\mathbf{z}^*(k) + [h_0, g_0] w^a(k)\right] = \mathbf{h}^{\mathrm{T}}\mathbf{z}(k) + \mathbf{g}^{\mathrm{T}}\mathbf{z}^*(k) = \mathbf{q}^{\mathrm{T}}\mathbf{z}^a(k) \tag{13.19}$$

When the driving noise w is circular, the widely linear model has no advantage over the standard linear model, whereas for noncircular signals we expect improvement in the performance proportional to the degree of noncircularity within the signal.

13.3.2 Quantifying Benefits of Widely Linear Estimation

The goal of widely linear estimation is to find coefficient vectors $\mathbf{h}$ and $\mathbf{g}$ that minimise the mean squared error $E[|d(k) - y(k)|^2]$ of the regression

$$\hat{y} = \mathbf{h}^{\mathrm{T}}\mathbf{z} + \mathbf{g}^{\mathrm{T}}\mathbf{z}^* \tag{13.20}$$

Following the approach from [240], to find the solution, apply the principle of orthogonality to obtain[6]

$$E[\hat{y}^*\mathbf{z}] = E[y^*\mathbf{z}], \quad \text{and} \quad E[\hat{y}^*\mathbf{z}^*] = E[y^*\mathbf{z}^*] \tag{13.21}$$

and replace $\hat{y}$ in (13.21) with its widely linear estimate (13.20), to yield[7]

$$\mathcal{C}\mathbf{h} + \mathcal{P}\mathbf{g} = \mathbf{u} \tag{13.22}$$

$$\mathcal{P}^*\mathbf{h} + \mathcal{C}^*\mathbf{g} = \mathbf{v}^* \tag{13.23}$$

where $\mathcal{C}$ and $\mathcal{P}$ are defined in (12.49), $\mathbf{u} = E[y^*\mathbf{z}]$, and $\mathbf{v} = E[y\mathbf{z}]$.

[6] We have $(y - \hat{y}) \perp \mathbf{z}$ and $(y - \hat{y}) \perp \mathbf{z}^*$, and as a consequence the orthogonality can be expressed in terms of expectations, as given in Equation (13.21).
[7] For convenience of the derivation, in Equations (13.22) and (13.28) the expression $z(k) = \mathbf{h}^{\mathrm{T}}\mathbf{z}(k) + \mathbf{g}^{\mathrm{T}}\mathbf{z}^*(k)$ is replaced by $z(k) = \mathbf{h}^{\mathrm{H}}\mathbf{z}(k) + \mathbf{g}^{\mathrm{H}}\mathbf{z}^*(k)$. This is a deterministic transformation and does not affect the generality of the results.

From Equations (13.22) and (13.23), the coefficient vectors that minimise the MSE of the widely linear model (Equation 13.8) are given by

$$\mathbf{h} = [\mathcal{C} - \mathcal{P}\mathcal{C}^{-1*}\mathcal{P}^*]^{-1}[\mathbf{u} - \mathcal{P}\mathcal{C}^{-1*}\mathbf{v}^*] \tag{13.24}$$

$$\mathbf{g} = [\mathcal{C}^* - \mathcal{P}^*\mathcal{C}^{-1}\mathcal{P}]^{-1}[\mathbf{v}^* - \mathcal{P}^*\mathcal{C}^{-1}\mathbf{u}] \tag{13.25}$$

and the corresponding widely linear mean square error (WLMSE) e^2_{WL} is given by

$$e^2_{\mathrm{WL}} = E[|y|^2] - (\mathbf{h}^{\mathrm{H}}\mathbf{u} + \mathbf{g}^{\mathrm{H}}\mathbf{v}^*) \tag{13.26}$$

whereas the mean square error (LMSE) e^2_{L} obtained with standard linear estimation is given by

$$e^2_{\mathrm{L}} = E[|y|^2] - \mathbf{u}^{\mathrm{H}}\mathcal{C}^{-1}\mathbf{u} \tag{13.27}$$

The advantage of widely linear estimation over standard linear estimation can be illustrated by comparing the corresponding mean square estimation errors $\delta e^2 = e^2_{\mathrm{L}} - e^2_{\mathrm{WL}}$, that is

$$\delta e^2 = [\mathbf{v}^* - \mathcal{P}^*\mathcal{C}^{-1}\mathbf{u}]^{\mathrm{H}}[\mathcal{C}^* - \mathcal{P}^*\mathcal{C}^{-1}\mathcal{P}]^{-1}[\mathbf{v}^* - \mathcal{P}^*\mathcal{C}^{-1}\mathbf{u}] \tag{13.28}$$

Due to the positive definiteness of the term $[\mathcal{C}^* - \mathcal{P}^*\mathcal{C}^{-1}\mathcal{P}]$ from Equation (13.28)

δe^2 is always non-negative;

$\delta e^2 = 0$ only when $[\mathbf{v}^* - \mathcal{P}^*\mathcal{C}^{-1}\mathbf{u}] = \mathbf{0}$.

that is, *widely linear estimation outperforms standard linear estimation for general complex signals; the two models produce identical results for circular signals.*

Exploitation of widely linear modelling promises several benefits, including:

- identical performance for circular signals and improved performance for noncircular signals;
- in blind source separation we may be able to deal with more sources than observations;
- improved signal recovery in communications modulation schemes (BPSK, GMSK);
- different and more realistic bounds on minimum variance unbiased (MVU) estimation;
- improved 'direction of arrival' estimation in augmented array signal processing;
- the analysis of augmented signal processing algorithms benefits from special matrix structures which do not exist in standard complex valued signal processing.

13.4 The Augmented Complex LMS (ACLMS) Algorithm

We now consider the extent to which widely linear mean square estimation has advantages over standard linear mean square estimation in the context of linear adaptive prediction. To answer this question, consider a widely linear adaptive prediction model for which the tap input $\mathbf{z}(k)$ to a finite impulse response filter of length N at the time instant k is given by

$$\mathbf{z}(k) = [z(k-1),\ z(k-2),\ \ldots,\ z(k-N)]^{\mathrm{T}} \tag{13.29}$$

Within widely linear regression, the augmented tap input delay vector $\mathbf{z}^a(k) = [\mathbf{z}^T(k), \mathbf{z}^H(k)]^T$ is 'widely linearly' combined with the adjustable filter weights $\mathbf{h}(k)$ and $\mathbf{g}(k)$ to form the output[8]

$$y(k) = \sum_{n=1}^{N} \left[h_n(k)z(k-n) + g_n(k)z^*(k-n)\right] \quad \Longleftrightarrow \quad y(k) = \mathbf{h}^T(k)\mathbf{z}(k) + \mathbf{g}^T(k)\mathbf{z}^*(k) \tag{13.30}$$

where $\mathbf{h}(k)$ and $\mathbf{g}(k)$ are the $N \times 1$ column vectors comprising the filter weights at time instant k, and $y(k)$ is the estimate of the desired signal $d(k)$.

For adaptive filtering applications,[9] similarly to the derivation of the standard complex least mean square (CLMS) algorithms, we need to minimise the cost function [113, 307]

$$J(k) = \frac{1}{2}|e(k)|^2 = \frac{1}{2}\left[e_r^2(k) + e_i^2(k)\right], \quad \text{with} \quad e(k) = d(k) - y(k) \tag{13.31}$$

where $e_r(k)$ and $e_i(k)$ are the respectively the real and imaginary part of the instantaneous output error $e(k)$. For simplicity, consider a generic weight update in the form[10]

$$\Delta w_n(k) = -\mu \nabla_{w_n} J(k) = -\mu \frac{\partial J(k)}{\partial w_n(k)} = -\mu \left(\frac{\partial J(k)}{\partial w_n^r(k)} + \jmath \frac{\partial J(k)}{\partial w_n^i(k)}\right) \tag{13.32}$$

where $w_n(k) = w_n^r(k) + \jmath w_n^i(k)$ is a complex weight and μ is the learning rate, a small positive constant. The real and imaginary parts of the gradient $\nabla_{w_n} J(k)$ can be expressed respectively as

$$\frac{\partial J(k)}{\partial w_n^r(k)} = e_r(k)\frac{\partial e_r(k)}{\partial w_n^r(k)} + e_i(k)\frac{\partial e_i(k)}{\partial w_n^r(k)} = -e_r(k)\frac{\partial y_r(k)}{\partial w_n^r(k)} - e_i(k)\frac{\partial y_i(k)}{\partial w_n^r(k)} \tag{13.33}$$

$$\frac{\partial J(k)}{\partial w_n^i(k)} = e_r(k)\frac{\partial e_r(k)}{\partial w_n^i(k)} + e_i(k)\frac{\partial e_i(k)}{\partial w_n^i(k)} = -e_r(k)\frac{\partial y_r(k)}{\partial w_n^i(k)} - e_i(k)\frac{\partial y_i(k)}{\partial w_n^i(k)}. \tag{13.34}$$

Similarly to Equations (13.33) and (13.34), the error gradients with respect to the elements of the weight vectors $\mathbf{h}(k)$ and $\mathbf{g}(k)$ of the widely linear variant of CLMS can be calculated as

[8] For consistent notation, we follow the original derivation of the complex LMS from [307].

[9] The widely linear LMS (WLLMS) and widely linear blind LMS (WLBLMS) algorithms for multiple access interference supression in DS-CDMA communications were derived in [262].

[10] We here provide a step by step derivation of ACLMS. The $\mathbb{CR}$ calculus (see Chapter 5) will be used in Chapter 15 to simplify the derivations for feedback and nonlinear architectures.

$$\frac{\partial J(k)}{\partial h_n^r(k)} = -e_r(k)\frac{\partial y_r(k)}{\partial h_n^r(k)} - e_i(k)\frac{\partial y_i(k)}{\partial h_n^r(k)} = -e_r(k)z_r(k-n) - e_i(k)z_i(k-n) \tag{13.35}$$

$$\frac{\partial J(k)}{\partial h_n^i(k)} = -e_r(k)\frac{\partial y_r(k)}{\partial h_n^i(k)} - e_i(k)\frac{\partial y_i(k)}{\partial h_n^i(k)} = e_r(k)z_i(k-n) - e_i(k)z_r(k-n) \tag{13.36}$$

$$\frac{\partial J(k)}{\partial g_n^r(k)} = -e_r(k)\frac{\partial y_r(k)}{\partial g_n^r(k)} - e_i(k)\frac{\partial y_i(k)}{\partial g_n^r(k)} = -e_r(k)z_r(k-n) + e_i(k)z_i(k-n) \tag{13.37}$$

$$\frac{\partial J(k)}{\partial g_n^i(k)} = -e_r(k)\frac{\partial y_r(k)}{\partial g_n^i(k)} - e_i(k)\frac{\partial y_i(k)}{\partial g_n^i(k)} = -e_r(k)z_i(k-n) - e_i(k)z_r(k-n) \tag{13.38}$$

giving the updates

$$\begin{aligned}\Delta h_n(k) &= -\mu\frac{\partial J(k)}{\partial h_n(k)} = -\mu\left(\frac{\partial J(k)}{\partial h_n^r(k)} + \jmath\frac{\partial J(k)}{\partial h_n^i(k)}\right)\\ &= \mu\Big[\big(e_r(k)z_r(k-n) + e_i(k)z_i(k-n)\big) + \jmath\big(e_i(k)z_r(k-n) - e_r(k)z_i(k-n)\big)\Big]\\ &= \mu e(k)z^*(k)\end{aligned} \tag{13.39}$$

$$\begin{aligned}\Delta g_n(k) &= -\mu\frac{\partial J(k)}{\partial g_n(k)} = -\mu\left(\frac{\partial J(k)}{\partial g_n^r(k)} + \jmath\frac{\partial J(k)}{\partial g_n^i(k)}\right)\\ &= \mu\Big[\big(e_r(k)z_r(k-n) - e_i(k)z_i(k-n)\big) + \jmath\big(e_r(k)z_i(k-n) + e_i(k)z_r(k-n)\big)\Big]\\ &= \mu e(k)z(k)\end{aligned} \tag{13.40}$$

These weight updates can we written in vector form as

$$\mathbf{h}(k+1) = \mathbf{h}(k) + \mu e(k)\mathbf{z}^*(k) \tag{13.41}$$

$$\mathbf{g}(k+1) = \mathbf{g}(k) + \mu e(k)\mathbf{z}(k) \tag{13.42}$$

To further simplify the notation, we can introduce an augmented weight vector $\mathbf{w}^a(k) = \big[\mathbf{h}^{\mathrm{T}}(k), \mathbf{g}^{\mathrm{T}}(k)\big]^{\mathrm{T}}$, and rewrite the ACLMS in its compact form as [132, 194]

$$\mathbf{w}^a(k+1) = \mathbf{w}^a(k) + \mu e(k)\mathbf{z}^{a*}(k) \tag{13.43}$$

where the ‘augmented’ instantaneous error[11] is $e(k) = d(k) - \mathbf{z}^{a\mathrm{T}}(k)\mathbf{w}^a(k)$. This completes the derivation of the augmented CLMS (ACLMS) algorithm, a widely linear extension of standard CLMS.

The ACLMS algorithm has the same generic form as the standard CLMS, it is simple to implement, yet it takes into account the full available second-order statistics of complex valued inputs (noncircularity).

[11]The output error itself is not augmented, but it is calculated based on the linear combination of the augmented input and weight vectors.

13.5 Adaptive Prediction Based on ACLMS

Simulations were performed for a 4-tap (4 taps of $\mathbf{h}$ and 4 taps of $\mathbf{g}$) FIR filter trained with ACLMS and the performances were compared to those of standard CLMS for a range of both synthetic and real world data, denoted by **DS1** – **DS4**. The synthetic benchmark signals were a linear circular complex $AR(4)$ process and two noncircular chaotic series,[12] whereas wind was used as a real world dataset.

DS1 *Linear AR(4) process* ('AR4'), given by [180]

$$y(k) = 1.79y(k-1) - 1.85y(k-2) + 1.27y(k-3) - 0.41y(k-4) + n(k) \qquad (13.44)$$

where $n(k)$ is a complex, white Gaussian noise with variance $\sigma^2 = 1$.

DS2 *Wind* ('wind'), containing wind speed and direction data averaged over one minute.[13]

DS3 *Lorenz Attractor* ('lorenz'), is a nonlinear, three-dimensional, deterministic system given by the coupled differential equations [173]

$$\frac{dx}{dt} = \sigma(y - x), \quad \frac{dy}{dt} = x(\rho - z) - y, \quad \frac{dz}{dt} = xy - \beta z \qquad (13.45)$$

where (typically) $\sigma = 10$, $\beta = 8/3$, $\rho = 28$.

DS4 *Ikeda Map* ('ikeda'), described by [104]

$$\begin{aligned} x(k+1) &= 1 + u\,(x(k)\cos[t(k)] - y(k)\sin[t(k)]) \\ y(k+1) &= u\,(x(k)\sin[t(k)] + y(k)\cos[t(k)]) \end{aligned} \qquad (13.46)$$

where u is a parameter (typically $u = 0.8$) and

$$t(k) = 0.4 - \frac{6}{1 + x^2(k) + y^2(k)}. \qquad (13.47)$$

Both batch[14] and online[15] learning scenarios were considered, and the standard prediction gain $R_p = 10 \log \frac{\sigma_y^2}{\sigma_e^2}$ was used as a quantitative measure of performance.

Batch learning scenario. Learning curves for the batch learning scenario are shown in the left-hand part of Figure 13.2. The dotted lines correspond to the learning curves of the CLMS algorithm, whereas the solid lines correspond to those of the ACLMS algorithm. For 'AR4'

---

[12]The two chaotic time series are generated by coupled difference equations, and were made complex by 'convenience of representation', that is, by taking the x and y components from Equations (13.45) and (13.46) and building a complex signal $z(k) = x(k) + \jmath y(k)$.

[13]The data used are from AWOS (Automated Weather Observing System) sensors obtained from the Iowa Department of Transportation. The Washington (AWG) station was chosen, and the dataset analysed corresponds to the wind speed and direction observed in January 2004. This dataset is publicly available from http://mesonet.agron.iastate.edu/request/awos/1min.php.

[14]For 1000 epochs with $\mu = 0.001$ and for 1000 data samples; for more detail on batch learning see Appendix G.

[15]With $\mu = 0.01$ and for 1000 samples of **DS1** – **DS4**.

signal (strictly circular) and 'wind' signal (almost circular for the given averaging interval and data length), there was almost no difference in performances of CLMS and ACLMS. A completely different situation occurred for the strongly noncircular 'lorenz' (see also Figure 12.2) and 'ikeda' signals. After the training, the prediction gains for the ACLMS algorithm were respectively about 3.36 (for 'lorenz') and 2.24 (for 'ikeda') times bigger than those of the corresponding CLMS algorithm. These results are perfectly in line with the background theory, that is, for noncircular signals, the minimum mean square error solution is based on augmented complex statistics.

Online learning scenario. Learning curves for adaptive one step ahead prediction are shown in Figure 13.2 (right), where the solid line corresponds to the real part of the original 'lorenz' signal, the dotted line represents the prediction based on CLMS, and the dashed line corresponds to the ACLMS based prediction. For the same filter setting, ACLMS was able to track the desired signal more accurately than CLMS.

Figure 13.2 also shows that the more noncircular the signal in question, the greater the performance advantage of ACLMS over CLMS, which conforms with the analysis in Section 13.3.2.

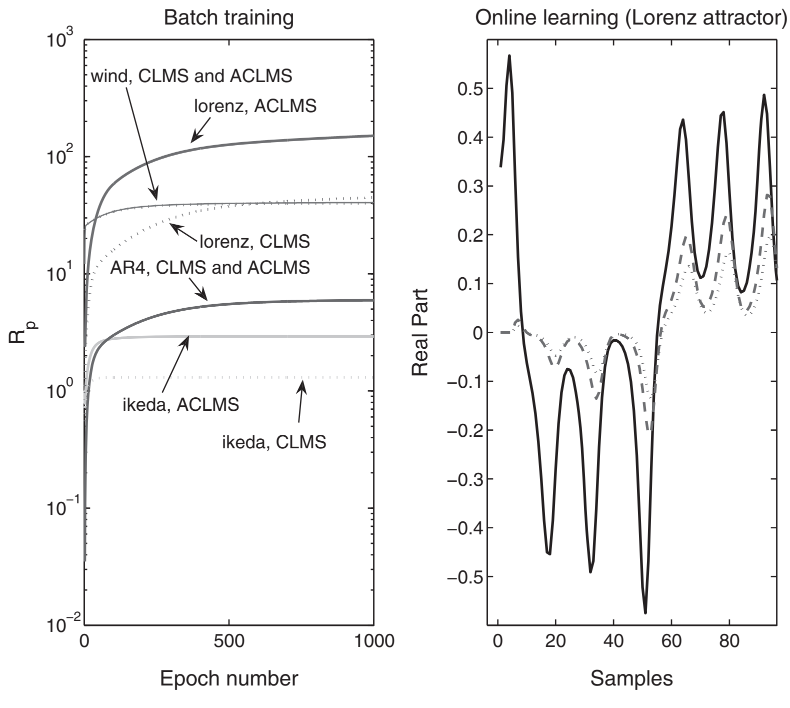


Figure 13.2 Comparison of the ACLMS and standard CLMS. Left: Prediction gains R_p for signals **DS1** – **DS4**. Right: Tracking performance for the Lorenz signal; solid line represents the original signal, dotted line the CLMS based prediction, and dashed line the ACLMS based prediction

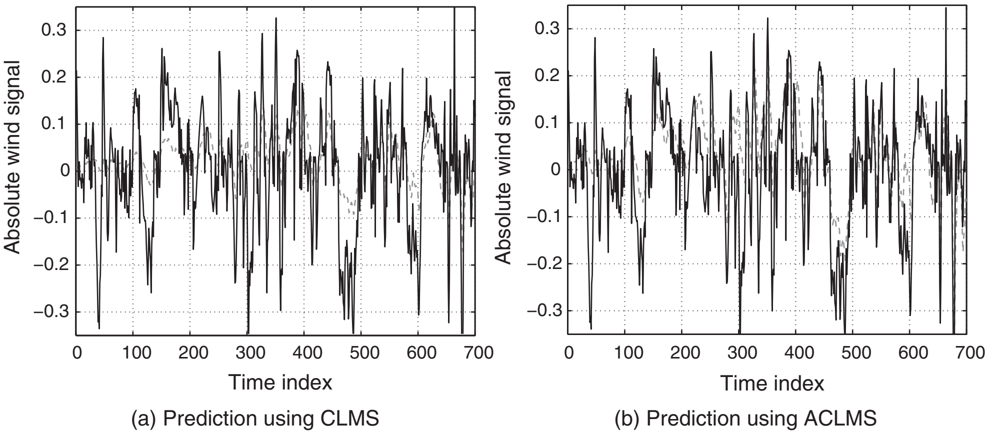


Figure 13.3 The original signal (solid line) and one step ahead prediction (dashed line)

13.5.1 Wind Forecasting Using Augmented Statistics

In the first experiment, adaptive one step ahead prediction of the original wind signal[16] was performed for a $N = 10$ tap FIR filter trained with CLMS and ACLMS. The time waveforms of the original and predicted signal are shown in Figure 13.3, indicating that the ACLMS was better suited to the statistics of the wind signal considered. A segment from Figure 13.3 is enlarged in Figure 13.4, showing the ACLMS being able to track the changes in wind dynamics more accurately than CLMS.

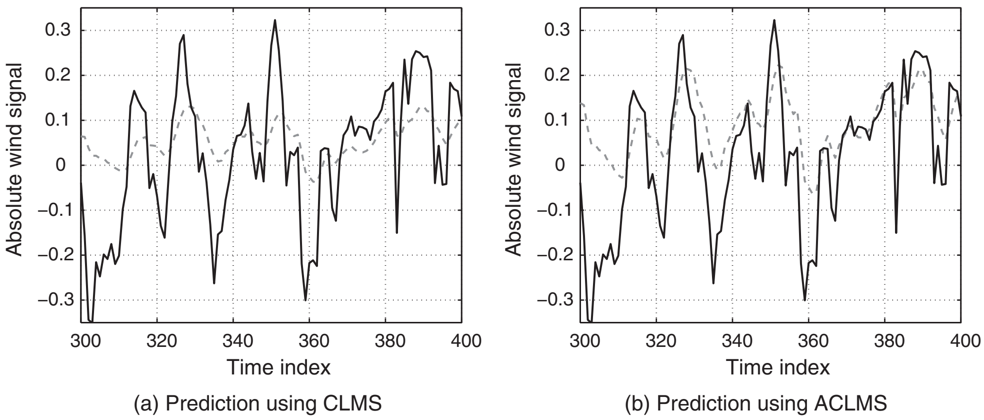


Figure 13.4 The original signal (solid line) and one step ahead prediction (dashed line)

[16]IOWA wind data averaged over 3 hours (see Footnote 13), which facilitates Gaussianity and widely linear modelling.

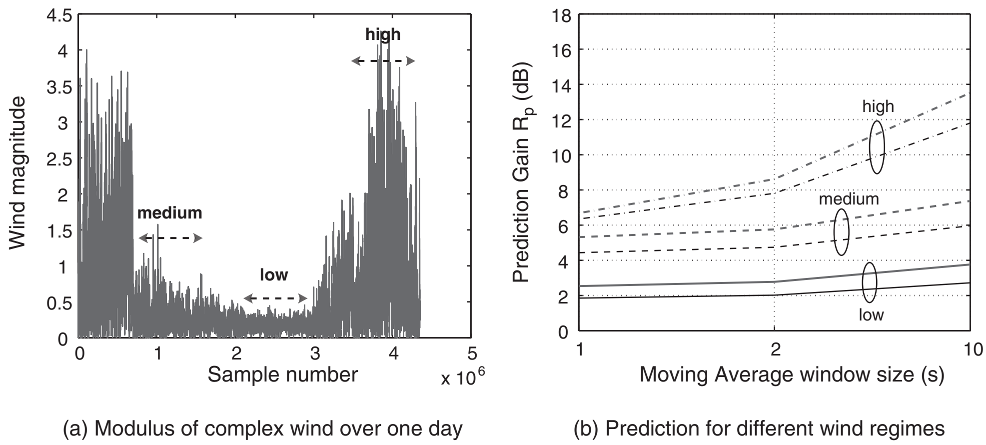


Figure 13.5 Performance of CLMS and ACLMS for different wind regimes. Thick lines correspond to ACLMS and thin lines to CLMS

In the second experiment, tracking performances of CLMS and ACLMS were investigated over a long period of time. Figure 13.5 shows the modulus of complex wind measurements recorded over one day at a sampling frequency of 50 Hz, and the performances of CLMS and ACLMS for the wind regimes denoted (according to the wind dynamics) by ‘low’, ‘medium’ and ‘high’. The prediction was performed on the raw data, and also on the data averaged over 2 and 10 seconds, and in all the cases, due to the noncircular nature of the considered wind data, the widely linear ACLMS outperformed standard CLMS.

To summarise:

- For nonlinear and noncircular signals (chaotic Lorenz and Ikeda maps), and signals with a large variation in the dynamics (‘high’ wind from Figure 13.5), the augmented statistics based modelling exhibited significant advantages over standard modelling for both the batch and online learning paradigms;
- For signals with relatively mild dynamics (linear ‘AR4’, heavily averaged wind from Figure 13.3, and the ‘medium’ and ‘low’ regions from Figure 13.5), the widely linear model outperformed the standard complex model; the improvement in the performance, however, varied depending on the degree of circularity within the signal;
- In practical applications, the pseudocovariance matrix is estimated from short segments of data in the filter memory and in the presence of noise; such estimate will be nonzero even for circular sources, and widely linear models are a natural choice.

It is therefore natural to ask whether it is possible to design a rigorous statistical testing framework which would reveal the second-order circularity properties of the ‘complex by convenience’ class of signals, a subject of Chapter 18.