Augmented Complex Statistics for Signal Prediction

M. Pedzisz and D. P. Mandic

Technical Report: TR-ICU-EPSRC-08-07-TR002

This material is a part of the second year report
for the EPSRC grant EP/D061709/1

August 24, 2007

Contents

1 Introduction

Standard statistical signal processing approaches have been based essentially on second–order statistical measures, that is, the variance and covariance for random variables, correlation and crosscorrelation for signals in the time domain and power spectral density and cross-spectral power density for signals in the frequency domain.

After the pioneering work in the domain of Higher-Order Statistics (HOS) (Shirayev, 1960; Rosenblatt, 1962; Brillinger, 1981), the potential of these tools has gained considerable attention, giving a raise to a number of research publications and applications (Giannakis, 1987; Mendel, 1991; Giannakis & Tsatsanis, 1992; Nikias & Petropolu, 1993; Nikias & Raghuveer, 1987; Lacoume, Gaeta, & Amblard, 1998). This field of research is particularly well suited to the higher-order properties of multidimensional variables and signals: multicorrelations in the time domain (Brillinger, 1981), and multispectra in the frequency domain (Shirayev, 1960), where a tensorial approach is a basis for the development of higher-order moments and cumulants (McCullagh, 1987).

Very few such results are available for complex random variables and signals in general. The Gaussian complex model is commonly used in the classical theory of complex variables and in communication applications, and is well documented (Wooding, 1956; Goodman, 1963). In particular, the distinct properties of the statistics of complex variables include Gaussian complex circularity, which is the basis of the theory of linear systems.

More recently, the lack of a general tool for complex-valued modelling was brought to light in (?, ?), where the notion of "proper complex random process" (their term for circular process) was introduced. This particular aspect of complex random variables (especially in terms of the bispectrum) has been presented in (Jouny & Moses, 1992).

With the increasing use of complex models in practical applications, it is necessary to develop a general statistical framework for dealing with complex random variables and signals.

This chapter provides an overview of the recent developments in this area, and suggests applications for "augmented statistics" in signal processing. It is divided into three main parts. In the first part, we introduce with complex random variables. Next, we advance to complex signals and their characterisation by means of augmented statistics. Finally, we deal with widely linear estimation of complex signals.

2 Complex Random Variables

The definition of a complex random variable (CRV) can be derived from definitions of two real random variables (RV). Given two real random vectors (RVs) X and Y, a complex random variable (CRV) $Z$is defined as

$$Z = X + jY, \qquad j^2 = -1 \tag{1}$$

In order to provide full statistical description (first, second, and higher order moments) of a CRV, consider the characteristic function[1]

$$\Phi_{Z,Z^*}(w, w^*) = E\left[\exp\left(j\frac{Z^*w + Zw^*}{2}\right)\right] \tag{2}$$

[1]This is a 2D FFT of a 2D probability density function.

where $E[\cdot]$ is the ensemble mean and $(\cdot)^*$ the complex conjugate operator. After applying Taylor series expansion (TSE) to the characterisation function (2), we have

$$E\left[\exp\left(j\frac{Z^*w+Zw^*}{2}\right)\right]=\sum_{n=1}^{\infty}\frac{1}{n!}\frac{j^n}{2^n}\sum_{p=0}^{n}b(n,p)w^{n-p}w^{*p}E\left[Z^{*n-p}Z^p\right] \tag{3}$$

where $b(n,p)$ are binomial coefficients. From (3), it is important to note that for a given order n, there are $(n+1)$ different moments $E\left[Z^{*n-p}Z^p\right]$ for complex random variables[2].

For a given characteristic function, moments of an arbitrary order n can be derived from

$$E\left[Z^{*n-p}Z^p\right]=\frac{2^n}{j^n}\frac{\partial^n\Phi_{Z,Z^*}(0,0)}{\partial w^{n-p}\partial w^{*p}}. \tag{4}$$

and similarly, by introducing the so called "second" characteristic function

$$\Psi_{Z,Z^*}(w,w^*)=\log\left[\Phi_{Z,Z^*}(w,w^*)\right] \tag{5}$$

we can calculate the cumulants as

$$\mathrm{Cum}\left[Z^{*n-p},Z^p\right]=\frac{2^n}{j^n}\frac{\partial^n\Psi_{Z,Z^*}(0,0)}{\partial w^{n-p}\partial w^{*p}} \tag{6}$$

Similarly to (3), there are, in general, $(n+1)$ different cumulants of order n. The theory of random variables states that

i) In the case of Gaussian random variable, cumulants of order greater than two $(n \geq 2)$ vanish;

ii) Joint cumulants of two statistically independent variables are zero.

Property *ii)* implies that for two statistically independent random variables X and Y

$$\mathrm{Cum}\left[X+Y\right]=\mathrm{Cum}\left[X\right]+\mathrm{Cum}\left[Y\right]$$

that is, cumulants of the sum of two independent random variables are equal to the sum of the cumulants of those variables.

An extension of the definitions of moments and cumulants to the multidimensional case using tensorial notation can be found in (Amblard, Gaeta, & Lacoume, 1996b).

2.1 Complex Circular Random Variables

The notion of "circularity" has been used to characterise Gaussian complex random variables (CCRV) (Goodman, 1963). However, in order to introduce strict statistical description of circularity, we need to involve general signal distributions and higher order statistics (HOS).

Circularity is intimately related to rotation in the geometric sense, in the case of a complex random variable Z, rotation by angle ϕ is achieved by multiplication of a RV by $\mathrm{e}^{j\phi}$, giving $Z_\phi = Z\mathrm{e}^{j\phi}$. Intuitively, a random variable Z is circular if its statistical properties are "invariant under a rotation".

A circular complex random variable (CCRV) can be now defined as (Comon, 1993)

[2]For example, for $n=2$, we have three different moments: $E\left[Z^2\right]$, $E\left[ZZ^*\right]$, and $E\left[Z^{*2}\right]$, for real random variables, all these moments are equal.

A single-dimensional complex random variable Z is called circular if for any angle ϕ both Z and $Ze^{j\phi}$, that is its rotation by angle ϕ, have the same probability distribution.

Alternatively, the condition of circularity in terms of the probability density function (pdf) can be expressed as

A single–dimensional CRV Z is circular if and only if its pdf is a function of only the product zz^*, that is[3]

$$p_{Z,Z^*}(z, z^*) = p_{Z_\phi, Z_\phi^*}(z_\phi, z_\phi^*) \tag{7}$$

In the special case of Gaussian CCRVs, this yields (Goodman, 1963)

$$p_{Z,Z^*}(z, z^*) = \frac{1}{\pi\sigma^2} e^{-zz^*/\sigma^2} \tag{8}$$

where σ^2 denotes the variance of the CRV.

The condition of circularity for characteristic functions:

A complex random variable Z is circular if and only if its (first or second) characteristic function depends only on the product ww^*, that is

$$\Phi_{Z_\phi, Z_\phi^*}(w, w^*) = \Phi_{Z,Z^*}(we^{-j\phi}, w^* e^{j\phi}). \tag{9}$$

When the statistical moments and cumulants do exist, then the condition of circularity can be expressed as (Amblard et al., 1996b)

A complex random variable Z is circular if and only if the only non–zero moments and cumulants are the moments and cumulants that have the same power in Z and Z^*.

This property follows directly from the fact that the partial derivatives in (6) taken at $(0, 0)$ are non–zero only when they are of the same order both in w and w^*. Conversely, when all the statistical moments or cumulants that are not symmetric are equal to zero, the Taylor series expansion of the characteristic function is a function of ww^*, for example, $\mathrm{Cum}[Z^2] = 0$, $\mathrm{Cum}[Z^4] = 0$, ..., $\mathrm{Cum}[Z^{2p}] = 0$. An extension of these results to the multidimensional case can be found in (Amblard et al., 1996b).

2.2 Design of Complex Circular Random Variables

For a deeper insight into the notion of CCRV, it is very helpful to be able to create CCRVs "by design". For simplicity, consider the single-dimensional case and zero-mean variables (circularity requires that all the even moments and cumulants vanish).

The Gaussian case is rather unique and is well understood. The Gaussian complex random variable $Z = X + jY$ is circular if

$$E[Z^2] = E[X^2] - E[Y^2] + jE[XY] = 0 \tag{10}$$

[3]This also implies that the pdf of a CCRV is function of only the modulus of Z.

In order to construct a CCRV, we therefore have to use two Gaussian RVs X and Y, and to verify that

$$E[X^2] = E[Y^2], \quad \text{and} \quad E[XY] = 0 \tag{11}$$

which boils down to having two independent RVs with equal powers.

For non-Gaussian RVs, this construction of a CCRV can be based upon the properties of pdf, as follows

1. Take a RV ρ with a pdf $p(\rho)$;
2. Take another RV θ uniformly distributed on $[0, 2\pi]$ and independent of ρ;
3. Construct $Z = X + jY$ as

$$X = \rho \cos(\theta), \qquad Y = \rho \sin(\theta). \tag{12}$$

The above procedure intuitively follows the general idea of circularity; in the polar coordinate system, this yields the following condition

$$p(\rho, \theta) = p(\rho, \theta - \phi) \tag{13}$$

where ϕ is the rotation angle.

3 Complex Signals

A natural extension of the concept of CRVs is to the analysis of higher-order statistical properties of complex–valued signals. The mathematical tools used are multicorrelations and multispectra for stationary signals, whereas for non-stationary signals they include higher–order time–frequency distributions. Applications of such techniques include blind deconvolution and equalisation, blind separation of sources, modulation recognition, detection of quadratic phase coupling, and Volterra filtering. A comprehensive introduction to these higher-order tools with numerous examples is provided in (Amblard, Gaeta, & Lacoume, 1996a).

3.1 Multicorrelations of Complex Signals

For a complex–valued random signal $z(t)$, its multicorrelation of order $(p+q)$ is defined as

$$C_{z,p+q,p}(\mathbf{t}) = \text{Cum}[z(t_0), \ldots, z(t_{p-1}), z^*(t_p), \ldots, z^*(t_{p+q-1})] \tag{14}$$

where $\text{Cum}[\cdot]$ is the cumulant and $\mathbf{t} = (t_0, \ldots, t_{p+q-1})$. Note that in this definition, the order of multicorrelation is $(p+q)$ with q referring to the conjugated and p referring to the nonconjugated components. Statistical properties of multicorrelations are direct consequences of the corresponding properties of cumulants:

- multicorrelations are multilinear
- they are equal to zero for $(p+q) > 2$ for Gaussian signals

Furthermore, if the signal is white in the strict sense (the random variables $z(t_0), \ldots, z(t_{p+q-1})$ are statistically independent for all $(p+q)$), the multicorrelations vanish except for the same index t_i in both z and z^*.

3.2 Multicorrelations and Multispectra for Stationary Complex Signals

A signal is called stationary if the statistics of all random vectors it induces are invariant under a time shift operation. More precisely, let $z(t)$ be a complex random signal and $\mathbf{z}_n = (z(t_1), \ldots, z(t_n))$ an n-dimensional induced vector. Then the signal is said to be stationary of order k if the pdf of $\mathbf{z}_n$ is invariant under a time shift for all $n \leq k$.

For the multicorrelation of order $(p+q)$, the property of stationarity leads to

$$C_{z,p+q,p}(\mathbf{t} + \boldsymbol{\tau}) = C_{z,p+q,p}(\boldsymbol{\tau}) \tag{15}$$

for all p and q such that $(p+q) \leq n$ and for all time lags $\boldsymbol{\tau}$. This shows that for stationary complex signals, the multicorrelation is no longer $(p+q)$ dimensional but $(p+q-1)$-dimensional, and is a function of only $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_{p+q-1})$, as illustrated by

$$C_{z,p+q,p}(\boldsymbol{\tau}) = \mathrm{Cum}[z(t), z(t+\tau_1), \ldots, z(t+\tau_{p-1}), z^*(t-\tau_p), \ldots, z^*(t-\tau_{p+q-1})] \tag{16}$$

In such a case, the multispectrum of order $(p+q)$ is defined as the Fourier transform of the corresponding multicorrelation, that is

$$S_{z,p+q,p}(\boldsymbol{\nu}) = \int C_{z,p+q,p}(\boldsymbol{\tau}) \exp(-j2\pi\boldsymbol{\tau}^{\mathrm{T}}\boldsymbol{\nu}) d\boldsymbol{\tau} \tag{17}$$

where $(\cdot)^{\mathrm{T}}$ denotes the vector transpose operator.

The extension to cross-multicorrelations and cross-multispectra and to various classes of multidimensional signals is straightforward and can be found in (Amblard et al., 1996a).

3.3 Higher Order Statistics of Complex Circular Stationary Signals

Finally, the notion of circularity can be extended to general complex valued signals in the following way. We say that a signal is circular of order n if the induced vectors of order lower or equal to n are circular. Further, if a random vector is circular of order n, then the statistics of the order lower or equal to n, containing a number of conjugated terms different from that of nonconjugated terms are zero. Therefore, this characterization of circularity suffices to prove that, for circular signals of order n

$$\forall_{p,q} \quad C_{z,p+q,p}(\boldsymbol{\tau}) = 0, \quad \text{such that} \quad p+q \leq n, \quad \text{and} \quad p \neq q. \tag{18}$$

Similar results can be obtained in the frequency domain, and in particular, a complex Gaussian signal is strictly circular (i.e. circular for all orders) if the only nonzero multispectra are $S_{z,2,0}(\boldsymbol{\nu})$, $S_{z,2,1}(\boldsymbol{\nu})$ and $S_{z,2,2}(\boldsymbol{\nu})$. If this signal is also analytic, only $S_{z,2,1}(\boldsymbol{\nu})$ is nonzero.

The properties of multispectra for other signal types can be found in (Amblard et al., 1996a). In particular: for any analytic signal, multispectra of the type $S_{z,p,p}(\boldsymbol{\nu})$ are identically zero; for any band-limited signal, the only nonzero multispectra are of the type $S_{z,2p,p}(\boldsymbol{\nu})$. To visualize the property of circularity of complex signals, Figure 1 presents samples of two signals: complex white Gaussian noise (complex circular) and the first two coordinates of the Lorenz attractor (non-circular signal). Observe the shapes of the signal distributions, the Gaussian signal clearly exhibits circularity, which is not the case for the Lorenz signal.

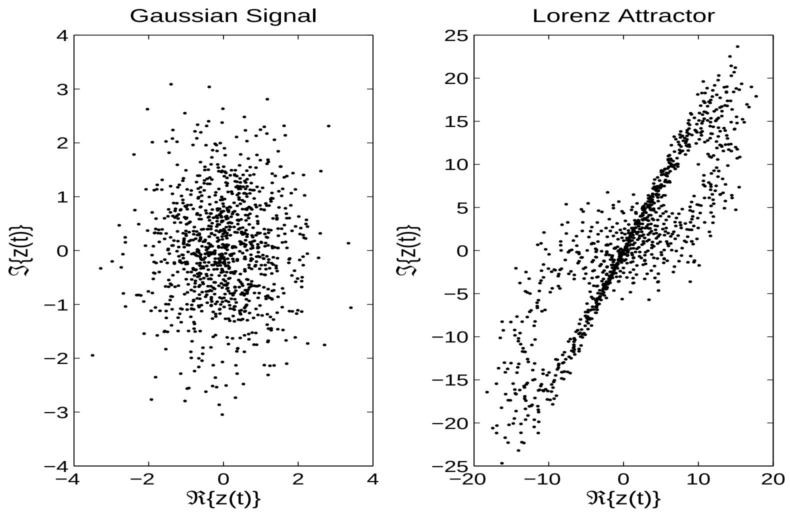

Figure 1: The real–imaginary part scatter plot of a circular (left) and noncircular signal (right).

4 Second Order Characterization of Complex Signals

In the preceding section, we have presented a general mathematical framework for higher-order statistical characterization of complex random variables and signals. In many practical situations (like MMSE based adaptive filtering) such a description can be restricted to only Second Order Statistics (SOS), giving a considerable reduction in the complexity of applied algorithms. In this section, we will consider such statistics for complex-valued signals (referred also as augmented statistics (Schreier, Scharf, & Hanssen, 2006)), and in particular, we will focus on their circular properties (Picinbono, 1994) and modelling capabilities (Picinbono & Chevalier, 1995).

4.1 Augmented Statistics of Complex Signals

For a discrete, zero-mean complex signal z, its second order statistics are described by the covariance function (CF) defined by

$$\gamma_z(k_1, k_2) = E[z(k_1)z^*(k_2)] \tag{19}$$

As has already been shown, this function is not sufficient to entirely describe the second-order statistics of $z(k)$. For this purpose, we have to introduce another relation function (RF) (Picinbono, 1996) (also called pseudo-covariance (Neeser & Massey, 1993) or complementary covariance (Schreier & Scharf, 2003)), defined as

$$r_z(k_1, k_2) = E[z(k_1)z(k_2)]. \tag{20}$$

In various instances, this RF is equal zero and therefore can be omitted[4]. Indeed, a second–order circular signal can be alternatively defined by the property $r_z(k_1, k_2) = 0$. As an obvious consequence, real-valued signals cannot be circular.

However, in general, there is no reason for the RF to be equal to zero, and therefore, it is necessary to describe the second–order statistics of complex signals completely, both in

[4]This is the case for the analytic signal of any stationary process, and more generally, for any circular signal (Picinbono, 1994).

terms of their "proper" and "improper" members. In particular, it has been demonstrated (Picinbono & Bondon, 1997), that the CF cannot be an arbitrary function, because it must be non–negative definite. In the same vein, the RF cannot be an arbitrary function, and the necessary conditions for its existence can be derived as:

- A complex signal $z(k)$ is said to be wide sense stationary (WSS) if its mean is constant and its CF (19) is only a function of the lag $\tau = k_1 - k_2$. This definition does not imply any condition on the RF (20);
- We say that a complex signal is second–order (SO) stationary if it is WSS and if its RF is only depending on τ.

It is clear that for real signals, the concepts of WSS and SO stationarity are equivalent. On the other hand, for complex signals, WSS stationary does not imply SO stationarity.

Let $\Gamma_z(\nu)$ be a Fourier Transform (FT) of CF, or in other words, the power spectrum of $z(k)$. Since stationarity implies that RF is symmetric, then the same property holds for its FT, i.e. $R_z(\nu) = R_z(-\nu)$, where $R_z(\nu)$ is a general a complex function.

Let us introduce a 2×1 complex random vector

$$\mathbf{z}(k) = [z(k), z^*(k)]^{\mathrm{T}} \tag{21}$$

Its CF is the 2×2 matrix $E[\mathbf{z}(k)\mathbf{z}^{\mathrm{H}}(k-\tau)]$, where the symbol $(\cdot)^{\mathrm{H}}$ denotes a simultaneous transposition and complex conjugation (Hermitian transpose). The FT of this matrix is called the spectral matrix of $\mathbf{z}(k)$, denoted $\mathbf{\Gamma_z}(\nu)$, and is known to be nonnegative definite (NND) (Picinbono, 1993). Simple calculation gives

$$\mathbf{\Gamma_z}(\nu) = \begin{bmatrix} \Gamma_z(\nu) & R_z(\nu) \\ R_z^*(\nu) & \Gamma_z(-\nu) \end{bmatrix} \tag{22}$$

This matrix is non–negative definite if and only if one of its diagonal elements and its determinant are nonnegative. This property is obvious for the diagonal elements. By making use of the symmetry of $R_z(\nu)$, the non–negativity condition on the determinant of (22) yields the bound

$$|R_z(\nu)|^2 \leq \Gamma_z(\nu)\Gamma_z(-\nu) \tag{23}$$

This is a necessary condition that the function $R_z(\nu)$ must satisfy in order to be FT of an RF of a signal with power spectrum $\Gamma_z(\nu)$. This condition is also sufficient (Picinbono & Bondon, 1997). A direct transposition of this condition to the case of nonstationary random signals yields

$$|R_z(\nu_1, \nu_2)|^2 \leq \Gamma_z(\nu_1, -\nu_1)\Gamma_z(\nu_2, -\nu_2). \tag{24}$$

These conditions can be equally derived in the time domain, using the properties of CF and RF. For a zero-mean, complex vector $\mathbf{z} = \mathbf{x} + j\mathbf{y}$ of $\mathbb{C}^n$, the corresponding covariance $\boldsymbol{\gamma}$ and relation $\mathbf{r}$ matrices are given by

$$\boldsymbol{\gamma} = E[\mathbf{z}\mathbf{z}^{\mathrm{H}}], \quad \text{and} \quad \mathbf{r} = E[\mathbf{z}\mathbf{z}^{\mathrm{T}}]. \tag{25}$$

If we define an augmented covariance matrix $\mathbf{A}$ (Schreier & Scharf, 2003) as

$$\mathbf{A} = \begin{bmatrix} \boldsymbol{\gamma} & \mathbf{r} \\ \mathbf{r}^* & \boldsymbol{\gamma}^* \end{bmatrix} \tag{26}$$

this allows us to redefine the condition (23) in the time domain; this is equivalent to the condition that the expression (Schur complement)

$$\boldsymbol{\gamma}^* - \mathbf{r}^{\mathrm{H}}\boldsymbol{\gamma}^{-1}\mathbf{r} \tag{27}$$

is non–negative definite (Picinbono, 1996). If $\mathbf{r} = \mathbf{0}$, then the complex vector $\mathbf{z}$ is called circular (or proper (Neeser & Massey, 1993)), otherwise it is called noncircular (or improper).

A detailed account of these results for a generalized single-sideband modulator is given in (Schreier & Scharf, 2003), whereas applications in the areas of telecommunications and information theory are outlined in (Neeser & Massey, 1993). A low-rank approximation of improper complex random vectors is provided in (Schreier & Scharf, 2001), and its applications in Wiener filtering are given in (Schreier & L.L., 2002). Finally, in (Schreier et al., 2006) a generalized likelihood ratio test for impropriety of complex signals is proposed.

4.2 Widely Linear Estimation of Complex Signals

Mean square estimation is one of the most fundamental principles of statistical signal processing. The basic problem can be stated as follows: let y be a scalar random variable to be estimated in terms of an observation random vector $\mathbf{z}$. The estimate that minimizes the mean square error (MSE) is then regression or the conditional expectation $E[y|\mathbf{z}]$. This result is usually given when both y and $\mathbf{z}$ are real, however, it also remains valid when these quantities are complex-valued.

If y and $\mathbf{z}$ are real jointly normal and with zero mean, then the regression is linear. This is no longer true for normal complex-valued data, where the regression is linear in both $\mathbf{z}$ and $\mathbf{z}^*$ and is called widely linear (WL). Without the normality assumption, this problem can be solved in terms of least mean square estimation (LMSE), given by (Picinbono & Chevalier, 1995)

$$\hat{y} = \mathbf{h}^{\mathrm{H}}\mathbf{z} + \mathbf{g}^{\mathrm{H}}\mathbf{z}^* \tag{28}$$

where both $\mathbf{h}$ and $\mathbf{g}$ are complex vectors of filter coefficients.

Since the moments of order k of y are completely defined from the corresponding moments of order k of $\mathbf{z}$ and $\mathbf{z}^*$, expression (28) characterizes a form of linearity, and (28) is very often called a wide sense linear filter or system. In practical applications, estimators based on (28) are not widely used due to the fact that in almost all calculations using complex Gaussian distributions, circularity is explicitly (or implicitly) assumed. Indeed, this assumption holds in many practical applications, causing the second term in (28) to vanish. However, there is no reason for this assumption to be generally accepted.

In general, the MMSE estimation problem in $\mathbb{C}$ can be formulated as:

Given (28), find the vectors $\mathbf{h}$ and $\mathbf{g}$ that minimize MSE $E[|y - \hat{y}|^2]$.

Using the principle of orthogonality (Picinbono & Chevalier, 1995), the solution is given in terms of expectations

$$E[\hat{y}^*\mathbf{z}] = E[y^*\mathbf{z}], \quad \text{and} \quad E[\hat{y}^*\mathbf{z}^*] = E[y^*\mathbf{z}^*]. \tag{29}$$

Replacing $\hat{y}$ with (28), gives

$$\boldsymbol{\gamma}\mathbf{h} + \mathbf{r}\mathbf{g} = \mathbf{u} \tag{30}$$

$$\mathbf{r}^*\mathbf{h} + \boldsymbol{\gamma}^*\mathbf{g} = \mathbf{v}^* \tag{31}$$

where $\boldsymbol{\gamma}$ and $\mathbf{r}$ are defined by (25), $\mathbf{u} = E[y^*\mathbf{z}]$, and $\mathbf{v} = E[y\mathbf{z}]$.

From (30) and (31), the solution becomes

$$\mathbf{h} = [\boldsymbol{\gamma} - \mathbf{r}(\boldsymbol{\gamma}^{-1})^*\mathbf{r}^*]^{-1}[\mathbf{u} - \mathbf{r}(\boldsymbol{\gamma}^{-1})^*\mathbf{v}^*] \tag{32}$$
$$\mathbf{g} = [\boldsymbol{\gamma}^* - \mathbf{r}^*\boldsymbol{\gamma}^{-1}\mathbf{r}]^{-1}[\mathbf{v}^* - \mathbf{r}^*\boldsymbol{\gamma}^{-1}\mathbf{u}] \tag{33}$$

and the corresponding MSE is then given by

$$e^2 = E[|y|^2] - (\mathbf{h}^{\mathrm{H}}\mathbf{u} + \mathbf{g}^{\mathrm{H}}\mathbf{v}^*) \tag{34}$$

and is smaller than e_L^2, which is the error that is obtained with a strictly LMSE-like procedure

$$e_L^2 = E[|y|^2] - \mathbf{u}^{\mathrm{H}}\boldsymbol{\gamma}^{-1}\mathbf{u}. \tag{35}$$

The advantage of the widely linear MSE (WLMSE) over LMSE is characterized by the quantity $\delta e^2 = e_L^2 - e^2$, and can be expressed as

$$\delta e^2 = [\mathbf{v}^* - \mathbf{r}^*\boldsymbol{\gamma}^{-1}\mathbf{u}]^{\mathrm{H}}[\boldsymbol{\gamma}^* - \mathbf{r}^*\boldsymbol{\gamma}^{-1}\mathbf{r}]^{-1}[\mathbf{v}^* - \mathbf{r}^*\boldsymbol{\gamma}^{-1}\mathbf{u}]. \tag{36}$$

This quantity is always nonnegative because the matrix $[\boldsymbol{\gamma}^* - \mathbf{r}^*\boldsymbol{\gamma}^{-1}\mathbf{r}]$ is positive definite, and consequently, $\delta e^2 = 0$ only when $[\mathbf{v}^* - \mathbf{r}^*\boldsymbol{\gamma}^{-1}\mathbf{u}] = \mathbf{0}$.

The consequences of applying the WLMSE procedure instead of LMSE, can be straightforwardly determined for some common scenarios (Picinbono & Chevalier, 1995). In the jointly circular case (well known in the normal case), characterised by

$$\mathbf{r} = \mathbf{0}, \quad \text{and} \quad \mathbf{v} = \mathbf{0} \tag{37}$$

it immediately results from (33) that (37) implies $\mathbf{g} = \mathbf{0}$. Similarly, (32) gives $\mathbf{h} = \boldsymbol{\gamma}^{-1}\mathbf{u}$, thus WLMSE becomes strictly linear, $\delta e^2 = 0$, and there is no advantage over standard LMSE procedure.

If the second assumption in (37) is neglected (circular observation), then (32), (33), and (36) can be respectively greatly simplified into

$$\mathbf{h} = \boldsymbol{\gamma}^{-1}\mathbf{u}, \qquad \mathbf{g}^* = \boldsymbol{\gamma}^{-1}\mathbf{v}, \qquad \text{and} \qquad \delta e^2 = \mathbf{v}^{\mathrm{H}}\boldsymbol{\gamma}^{-1}\mathbf{v} \tag{38}$$

which leads to the fact that a non–zero vector $\mathbf{v}$ necessarily implies an increase in the performance of WLMSE.

The general conclusion is that when the complex data are not jointly circular, the LMSE is not the best estimation technique in the domain of second-order statistics of the signals.

5 Prediction of Complex Signals

Recent progress in biomedicine, wireless and mobile communications, seismic, sonar and radar signal processing has brought to the light new problems where data models are often complex-valued or have a higher-dimensional compact representation (eg. wind speed and direction). To process such signals, much effort has been applied toward extending the results from real-valued adaptive filters to their complex counterparts.

One such algorithm is the complex least mean square (CLMS) for finite impulse response (FIR) adaptive filters (Widrow, McCool, & Ball, 1975). Other authors have considered complex-valued learning algorithms for training neural networks (Kim & Adali, 2001; Leung & Haykin, 1991; Goh & Mandic, 2004a, 2004b).

Recently, an extended Kalman filter (EKF) training of neural networks has been extended to the complex domain (Huang & Chen, 2000). This work has given rise to the first (to our knowledge) application of augmented statistics used for learning EKF in the framework of complex-valued recurrent neural networks (Goh & Mandic, 2007).

5.1 Augmented LMS Algorithm

Given the widespread use of LMS algorithms in practice, it is natural to consider the extent to which WLMSE has advantages over LMSE in adaptive filtering applications. To answer this question, consider a widely linear prediction model with the input $\mathbf{z}(k)$ at the time instant k composed of N successive samples, represented by a delay vector given by

$$\mathbf{z}(k) = [z(k-1),\, z(k-2),\, \ldots,\, z(k-N)]^{\mathrm{T}}. \tag{39}$$

This vector is widely linearly combined with the adaptable weights $h_i(k)$ and $g_i(k)$ to form the output $y(k)$

$$y(k) = \sum_{i=1}^{N} [h_i(k) z(k-i) + g_i(k) z^*(k-i)], \quad \Longleftrightarrow \quad y(k) = \mathbf{h}^{\mathrm{T}}(k)\mathbf{z}(k) + \mathbf{g}^{\mathrm{T}}(k)\mathbf{z}^*(k) \tag{40}$$

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

To find an optimal estimate of $d(k)$, it is necessary to minimize a performance criterion $E(k)$, typically the square of the norm of the instantaneous error $e(k)$, given by

$$E(k) = \tfrac{1}{2}|e(k)|^2 = \tfrac{1}{2}\left[e_R^2(k) + e_I^2(k)\right], \quad \text{with} \quad e(k) = d(k) - y(k) \tag{41}$$

where $e_R(k)$ and $e_I(k)$ are the real and imaginary part of the complex instantaneous error $e(k)$.

Using gradient-based learning, the aim is to update iteratively the filter weights so that $E(k)$ is minimized. In such a case, a "general weight update" $\Delta w_i(k)$ can be derived from

$$\Delta w_i(k) = -\mu \frac{\partial E(k)}{\partial w_i(k)} = -\mu \left(\frac{\partial E(k)}{\partial w_i^R(k)} + j \frac{\partial E(k)}{\partial w_i^I(k)} \right) \tag{42}$$

where $w_i(k) = w_i^R(k) + j w_i^I(k)$, μ is the learning rate, a small positive constant; and the partial updates over the real and imaginary part are expressed by

$$\frac{\partial E(k)}{\partial w_i^R(k)} = e_R(k)\frac{\partial e_R(k)}{\partial w_i^R(k)} + e_I(k)\frac{\partial e_I(k)}{\partial w_i^R(k)} = -e_R(k)\frac{\partial y_R(k)}{\partial w_i^R(k)} - e_I(k)\frac{\partial y_I(k)}{\partial w_i^R(k)} \tag{43}$$

$$\frac{\partial E(k)}{\partial w_i^I(k)} = e_R(k)\frac{\partial e_R(k)}{\partial w_i^I(k)} + e_I(k)\frac{\partial e_I(k)}{\partial w_i^I(k)} = -e_R(k)\frac{\partial y_R(k)}{\partial w_i^I(k)} - e_I(k)\frac{\partial y_I(k)}{\partial w_i^I(k)}. \tag{44}$$

For the considered widely linear scenario (40), the corresponding partial derivatives can be calculated as

$$\frac{\partial E(k)}{\partial h_i^R(k)} = -e_R(k)\frac{\partial y_R(k)}{\partial h_i^R(k)} - e_I(k)\frac{\partial y_I(k)}{\partial h_i^R(k)} = -e_R(k)z_R(k-i) - e_I(k)z_I(k-i) \quad (45)$$

$$\frac{\partial E(k)}{\partial h_i^I(k)} = -e_R(k)\frac{\partial y_R(k)}{\partial h_i^I(k)} - e_I(k)\frac{\partial y_I(k)}{\partial h_i^I(k)} = e_R(k)z_I(k-i) - e_I(k)z_R(k-i) \quad (46)$$

$$\frac{\partial E(k)}{\partial g_i^R(k)} = -e_R(k)\frac{\partial y_R(k)}{\partial g_i^R(k)} - e_I(k)\frac{\partial y_I(k)}{\partial g_i^R(k)} = -e_R(k)z_R(k-i) + e_I(k)z_I(k-i) \quad (47)$$

$$\frac{\partial E(k)}{\partial g_i^I(k)} = -e_R(k)\frac{\partial y_R(k)}{\partial g_i^I(k)} - e_I(k)\frac{\partial y_I(k)}{\partial g_i^I(k)} = -e_R(k)z_I(k-i) - e_I(k)z_R(k-i) \quad (48)$$

and the corresponding weight updates can be expressed as

$$\begin{aligned}\Delta h_i(k) &= -\mu\frac{\partial E(k)}{\partial h_i(k)} = -\mu\left(\frac{\partial E(k)}{\partial h_i^R(k)} + j\frac{\partial E(k)}{\partial h_i^I(k)}\right)\\ &= \mu\left[(e_R(k)z_R(k-i) + e_I(k)z_I(k-i)) + j(e_I(k)z_R(k-i) - e_R(k)z_I(k-i))\right]\\ &= \mu e(k)z^*(k)\end{aligned} \quad (49)$$

$$\begin{aligned}\Delta g_i(k) &= -\mu\frac{\partial E(k)}{\partial g_i(k)} = -\mu\left(\frac{\partial E(k)}{\partial g_i^R(k)} + j\frac{\partial E(k)}{\partial g_i^I(k)}\right)\\ &= \mu\left[(e_R(k)z_R(k-i) - e_I(k)z_I(k-i)) + j(e_R(k)z_I(k-i) + e_I(k)z_R(k-i))\right]\\ &= \mu e(k)z(k)\end{aligned} \quad (50)$$

In the matrix formulation, the filter weights are updated as follows

$$\mathbf{h}(k+1) = \mathbf{h}(k) + \mu e(k)\mathbf{z}^*(k) \quad (51)$$

$$\mathbf{g}(k+1) = \mathbf{g}(k) + \mu e(k)\mathbf{z}(k) \quad (52)$$

which completes the derivation of the augmented CLMS (ACLMS) algorithm, which is a widely linear extension of the CLMS algorithm proposed in (Widrow et al., 1975).

5.2 Experiments

To assess the performances of the ACLMS algorithm, simulations were performed for a 4-tap (4 taps of **h** and 4 taps of **g**) widely linear model (ACLMS) and for a standard LMS-like FIR adaptive filter. The main objective of these experiments was to provide an insight into the performance gains of the WL estimation compared to the strictly linear one for both circular and noncircular complex-valued signals.

As the learning sets, we used:

Linear AR(4) model ('ar4'). This linear model is proposed in (Mandic, 2004) (first experiment, page 116) and described by

$$y(k) = 1.79y(k-1) - 1.85y(k-2) + 1.27y(k-3) - 0.41y(k-4) + z(k) \quad (53)$$

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

Wind ('wind'). The samples used in the experiments were obtained from the Iowa Department of Transportation[5] and contain data acquired every minute from AWOS (Automated

[5] Publicly available from `http://mesonet.agron.iastate.edu/request/awos/1min.php`.

Weather Observing System) sensors. We have chosen the Washington (AWG) station, and the gathered data corresponds to the wind speed and direction observed in January 2004.

Lorenz Attractor ('lorenz'). Lorenz attractor is a chaotic map which shows how the state of a dynamical system evolves over time in a complex, non-repeating pattern. This system is nonlinear, three-dimensional, deterministic and is described by coupled equations (Lorenz, 1963)

$$\frac{dx}{dt} = \sigma(y - x), \qquad \frac{dy}{dt} = x(\rho - z) - y, \qquad \frac{dz}{dt} = xy - \beta z \tag{54}$$

with (usually) $\sigma = 10$, $\beta = 8/3$, $\rho = 28$. For simulations we have used the first two coordinates of the Lorenz attractor.

Ikeda Map ('ikeda'). This is another complex chaotic map, described by the following nonlinear system (Hammel, Jones, & Moloney, 1985)

$$x(n+1) = 1 + u\,(x(n)\cos[t(n)] - y(n)\sin[t(n)]) \tag{55}$$

$$y(n+1) = u\,(x(n)\sin[t(n)] + y(n)\cos[t(n)]) \tag{56}$$

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

$$t(n) = 0.4 - \frac{6}{1 + x^2(n) + y^2(n)}. \tag{57}$$

For simulations we have used the first two coordinates of this chaotic map.

For all experiments, the following two learning scenarios were considered:

- Batch training for 1000 epochs ($\mu = 0.001$) where each signal was composed of 1000 samples. This scenario applies to the case where the number of available samples is relatively large and the objective is to find the best (in terms of the prediction gain) data model;
- On-line adaptation ($\mu = 0.01$) for each signal composed of 1000 samples. This on-line scenario corresponds to situations where the number of samples is relatively large, the required processing time is small, and at the same time, prediction gain should be close to that obtained by batch training.

As a quantitative measure of performance we used a prediction gain R_p given by (Haykin & Li, 1995)

$$R_p \triangleq \frac{\sigma_{\hat{y}}^2}{\sigma_e^2} \tag{58}$$

where $\sigma_{\hat{y}}^2$ denotes the variance of the predicted signal $\hat{y}(k)$, and σ_e^2 the variance of the instantaneous prediction error $e(k)$.

In the first experiment, the performances of the considered algorithms are compared in the left hand part of Figure 2. The dotted lines correspond to the CLMS algorithm, whereas the solid lines correspond to the ACLMS algorithm. Observe that for 'ar4' signal (which is strictly circular) and 'wind' signal (which is almost circular for the given sample rate and data length), there is almost no difference in performances between CLMS and ACLMS algorithms. A totally different situation was observed for the 'lorenz' and 'ikeda' signals (purely noncircular, see also Figure 1) – at the end of training, the prediction gain of the ALMS algorithm was about 3.36 (for 'lorenz') and 2.24 (for 'ikeda') times bigger than that corresponding to the CLMS algorithm. These results are perfectly in line with the background theory introduced in Section 4.2: when the complex data are

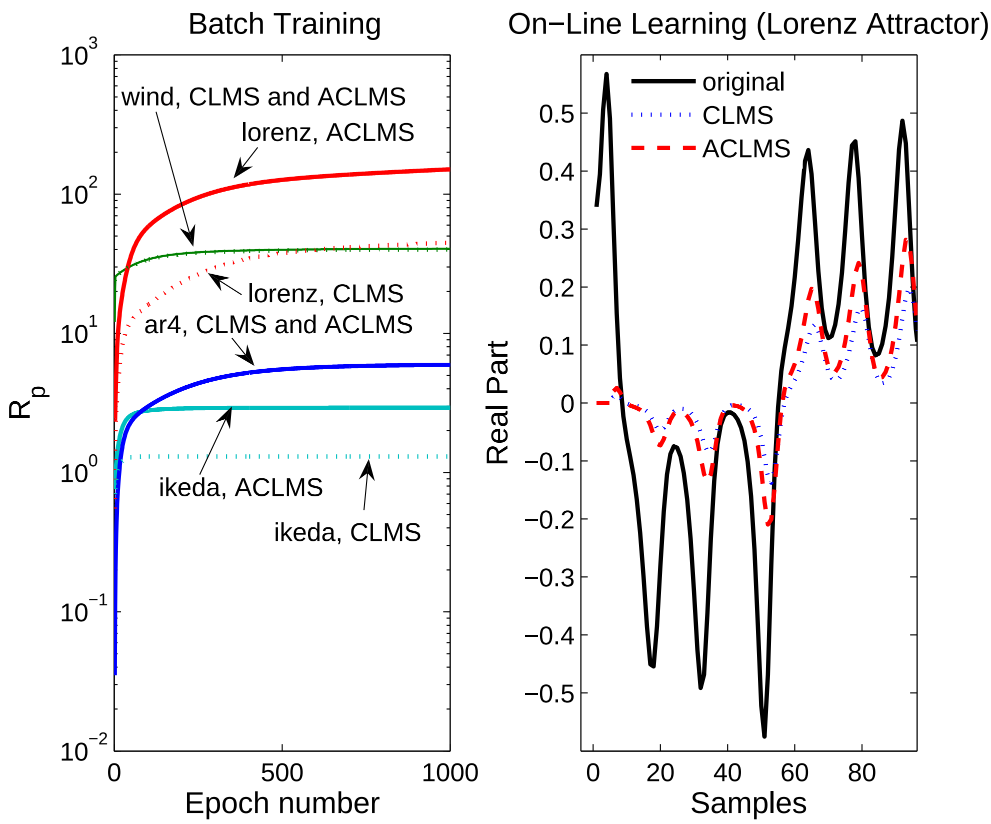

Figure 2: Advantages of the ACLMS algorithm over standard CLMS

not circular, the optimal solution, in the mean square sense, for signal estimation is based on the augmented statistics (full second-order statistical description). Similar results were obtained in the second experiment, where on-line learning was applied for signal prediction. The solid curve in Figure 2 (right) corresponds to the real part of the 'lorenz' signal, the dotted curve to the prediction based on the CLMS, and the dashed curve to the prediction based on the ACLMS algorithm. It is clearly visible that, for the same learning rate, the ACLMS algorithm converges faster to the desired signal than the CLMS algorithm.

References

Amblard, P., Gaeta, M., & Lacoume, J. (1996a). Statistics for Complex Variables and Signals – Part II: Signals. Signal Processing, 53, 15-25.

Amblard, P., Gaeta, M., & Lacoume, J. (1996b). Statistics for Complex Variables and Signals – Part I: Variables. Signal Processing, 53, 1-13.

Brillinger, D. (1981). Time Series Analysis: Data Analysis and Theory. New York: McGraw-Hill.

Comon, P. (1993, September). Course on Higher-Order Statistics (Tech. Rep.). School of Theoretical Physics at Les Houches, France.

Giannakis, G. (1987, September). Cumulants: A Powerfull Tool in Signal Processing. Proceedings of the IEEE, 75(9), 1333-1334.

Giannakis, G., & Tsatsanis, M. (1992, March). A Unifying Maximum-Likelihood View of Cumulants and Polyspectra Measures for Non-Gaussian Classification and Estimation. IEEE Transactions on Information Theory, 38(2), 386-406.

Goh, S., & Mandic, D. (2004a). Complex-Valued Estimation of Wind Profile and Wind Power. Proceedings of the 12th IEEE Mediterranen Electrotechnical Conference, MELECON '04, 3, 1037-1040.

Goh, S., & Mandic, D. (2004b). A Complex-Valued RTRL Algorithm for Recurrent Neural Networks. Neural Computation, 16(12), 2699-2713.

Goh, S., & Mandic, D. (2007). An Augmented Extended Kalman Filter Algorithm for Complex-Valued Recurrent Neural Networks. Neural Computation, 19, 1039-1055.

Goodman, N. (1963). Statistical Analysis Based on Certain Multivariate Complex Gaussian Distribution. Annals of Mathematics and Statistics, 34, 152-176.

Hammel, S., Jones, C., & Moloney, J. (1985). Global Dynamical Behavior of the Optical Field in a Ring Cavity. Journal of the Optical Society of America, Optical Physics, 2(4), 552-564.

Haykin, S., & Li, L. (1995). Nonlinear Adaptive Prediction of Nonstationary Signals. IEEE Transactions on Signal Processing, 43(2), 526-535.

Huang, R., & Chen, M. (2000). Adaptive Equalization Using Complex-Valued Multilayered Neural Networ Based on the Extended Kalman Filter. Proceeding of the International Conference on Signal Processing, WCCC-ICSP '00, 1, 519-524.

Jouny, I., & Moses, R. (1992, November). The Bispectrum of Complex Signals: Definitions and Properties. IEEE Transactions on Signal Processing, 40, 2833-2836.

Kim, T., & Adali, T. (2001). Complex Backpropagation Neural Network Using Elementary Transcendental Activation Functions. Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, ICASSP' 01, 2, 1281-1284.

Lacoume, J., Gaeta, M., & Amblard, P. (1998). From Order Two to Higher Order Statistics: New Tools and Applications. Signal Processing VI, Theory and Applications, Elsevier, 1, 91-98.

Leung, H., & Haykin, S. (1991, September). The Complex Backpropagation Algorithm. IEEE Transactions on Signal Processing, 39(9), 2101-2104.

Lorenz, E. (1963, March). Deterministic Nonperiodic Flow. Journal of the Atmospheric Sciences, 20(2), 130-141.

Mandic, D. (2004, February). A Generalized Normalized Gradient Descent Algorithm. IEEE Signal Processing Letters, 11(2), 115-118.

McCullagh, P. (1987). Tensor Methods in Statistics. London, UK: Chapman and Hall.

Mendel, J. (1991, March). Tutorial on Higher-Order Statistics (spectra) in Signal Processing and System Theory: Theoretical Results and Some Applications. Proceedings of the IEEE, 79, 278-305.

Neeser, F., & Massey, J. (1993, July). Proper Complex Random Processes with Application to Information Theory. IEEE Transactions on Information Theory, 39(4), 1293-1302.

Nikias, C., & Petropolu, A. (1993). Higher-Order Statistics: A Non-Linear Siganal Processing Framework. New York: Prentice-Hall.

Nikias, C., & Raghuveer, M. (1987). Bispectrum Estimation: A Digital Signal Processing Framework. Proceedings of the IEEE, 75, 869-891.

Picinbono, B. (1993). Random Signals and Systems. NJ, USA: Englewood Cliffs, Prentice Hall.

Picinbono, B. (1994, December). On Circularity. IEEE Transactions on Signal Processing, 42(12), 3473-3482.

Picinbono, B. (1996, October). Second-Order Complex Random Vectors and Normal Distributions. IEEE Transactions on Signal Processing, 44, 2637-2640.

Picinbono, B., & Bondon, P. (1997, February). Second-Order Statistics of Complex Signals. IEEE Transactions on Signal Processing, 45(2), 411-420.

Picinbono, B., & Chevalier, P. (1995, August). Widely Linear Estimation with Complex Data. IEEE Transactions on Signal Processing, 43(8), 2030-2033.

Rosenblatt, M. (1962). Random Process. New York: Oxford Press.

Schreier, P., & L.L., S. (2002). Canonical coordinates for reduced-rank estimation of improper complex random vectors. In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, ICASSP '02 (Vol. 2, p. 1153-1156).

Schreier, P., & Scharf, L. (2001, November). Low-Rank Approximation of Improper Complex Random Vectors. In Conference Record of the Thirty-Fifth Asilomar Conference on Signals, Systems and Computers (Vol. 1, p. 597-601).

Schreier, P., & Scharf, L. (2003, March). Second-Order Analysis of Improper Complex Random Vectors and Processes. IEEE Transactions on Signal Processing, 51(3), 714-725.

Schreier, P., Scharf, L., & Hanssen, A. (2006, July). A generalized likelihood ratio test for impropriety of complex signals. IEEE Signal Processing Letters, 13(7), 433-436.

Shirayev, A. (1960, March). Some Problems in the Spectral Theory of Higher-Order Moments. Theory of Probability and Applications, 5.

Widrow, B., McCool, J., & Ball, M. (1975, April). The Complex LMS Algorithm. Proceedings of the IEEE, 63(4), 719-720.

Wooding, R. (1956). The Multivariate Distribution of Complex Normal Variables. Biometrica, 43, 212-215.