Signal Processing 90 (2010) 1788–1799

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

# An augmented affine projection algorithm for the filtering of noncircular complex signals

Yili Xia *, Clive Cheong Took, Danilo P. Mandic

*Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2BT, UK*

ARTICLE INFO

*Article history:*
Received 15 June 2009
Received in revised form
5 October 2009
Accepted 21 November 2009
Available online 1 December 2009

*Keywords:*
Widely linear modelling
Affine projection algorithm
Complex-valued noncircular signals
Augmented complex statistics
Wind modelling

ABSTRACT

An augmented affine projection adaptive filtering algorithm (AAPA), utilising the full second order statistical information in the complex domain is proposed. This is achieved based on the widely linear model and the joint optimisation of the direct and conjugate data channel parameters. The analysis illustrates that the use of augmented complex statistics and widely linear modelling makes the AAPA suitable for the processing of both second order complex circular (proper) and noncircular (improper) signals. The derivation is supported by the analysis of convergence in the energy conservation setting. Simulations on both benchmark and real-world noncircular wind signals support the analysis.



## 1. Introduction

Stochastic gradient based adaptive filtering algorithms are designed to minimise an instantaneous error power and have been widely used in numerous signal processing applications, such as noise cancellation, system identification and adaptive prediction. The normalised least mean square (NLMS) algorithm is the most popular choice, due to its stability, fast convergence and low complexity. Learning algorithms based on the least squares estimation, such as the recursive least squares (RLS) algorithm [1], converge faster, however, in practical applications, they suffer from high computational complexity. To combine the benefits of both approaches, that is excellent stability, low computational complexity and fast convergence, the affine projection algorithm (APA), based on affine subspace projections, has been introduced as a link between the NLMS and RLS algorithm [2,3]. This way, the APA updates the weight vector on the basis of both past and current input vectors, resulting in faster convergence as compared to the NLMS algorithm [4,5], and much fewer computations and enhanced stability as compared to the RLS algorithm [6].

These class of APAs have been originally introduced for real-valued signals. However, some real world processes, such as vector fields and directional signals with "intensity" and "direction" components, are best understood when considered complex-valued [7]. Unfortunately, standard adaptive filtering algorithms in $\mathbb{C}$ are straightforward extensions of the corresponding algorithms in $\mathbb{R}$, and do not make full use of the algebraic structure of the complex domain. For instance, it is commonly assumed that the covariance matrix of a zero mean complex vector $\mathbf{z}$ is $E\{\mathbf{z}\mathbf{z}^H\}$, and is seen as an extension of the real covariance $E\{\mathbf{z}\mathbf{z}^T\}$, achieved by replacing the vector transpose operator by the Hermitian transpose. However, recent advances in the so-called augmented complex statistics show that the complex covariance matrix $E\{\mathbf{z}\mathbf{z}^H\}$ is not sufficient to fully describe second order statistics in $\mathbb{C}$, and the pseudo-covariance matrix $E\{\mathbf{z}\mathbf{z}^T\}$ should also be considered. This proves particularly important when processing second order noncircular (improper) signals [8].

* Corresponding author.
*E-mail addresses:* yili.xia06@ic.ac.uk (Y. Xia), c.cheong-took@ic.ac.uk (C.C. Took), d.mandic@ic.ac.uk (D.P. Mandic).

0165-1684/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2009.11.026

The use of augmented complex statistics has opened the possibility to design adaptive filtering algorithms suitable for processing both circular and noncircular signals. Such algorithms are based on a widely linear model [8,9], and are usually called "widely linear" or "augmented" algorithms. Examples include the work by Valkama et al. who have introduced widely linear algorithms in wireless communication applications to solve the complex-valued I/Q mismatch problem [10,11]. Other examples include the widely linear complex least mean square (WL-CLMS) algorithm for direct sequence code division multiple access (DS-CDMA) applications by Schober et al. [12], which used a real valued error. It was subsequently extended to the augmented CLMS (ACLMS) algorithm in [13] in order to cater for both complex signals and complex errors. A widely linear recursive least squares (WL-RLS) algorithm was proposed by Douglas [14], and also by Kuh and Mandic [15] in the context of complex augmented kernels. An extension of widely linear filtering for feedback systems was addressed in [16].

In the context of the APAs, a widely linear affine projection algorithm (WL-APA) for DS-CDMA applications has been proposed by Lima and Lamare [17]. This algorithm is based on the minimisation of real-valued error vector, as required in the DS-CDMA setting, where the error vector is given by

$$\mathbf{e}_{APA}^{WL} = \mathbf{b}_{APA} - \Re[\mathbf{U}_{APA}\mathbf{w}_{APA}^{WL}] \tag{1}$$

and $\mathbf{U}_{APA}$ is the matrix of complex-valued received signals, deteriorated by complex noise, $\mathbf{b}_{APA}$ are the desired real-valued input symbols, i.e. amplitude-shift keying (ASK) or binary phase-shift keying (BPSK), $\mathbf{w}_{APA}^{WL}$ is the weight vector for estimating the coefficients of adaptive linear multiuser receivers, and $\Re[\cdot]$ denotes the real part of complex numbers. However, as this algorithm is developed for DS-CDMA and uses a real-valued output error, it does not fully exploit the advantages of the widely linear model [12]. Recent attempts to make full use of the information available in the widely linear model include the approach in [18], which optimises two separate cost functions for the "standard" and "conjugate" parts of the widely linear model.

A rigorously derived augmented APA algorithm for the processing of general complex signals (both circular and noncircular) is therefore still missing. To this end, we now provide a general augmented APA (AAPA), which uses a global cost function in the derivation and is a generic extension of standard APA for the filtering of second order noncircular signals. This algorithm simplifies into the normalised ACLMS when only a scalar instantaneous error is considered. The analysis of steady state performance is provided based on the energy conservation principle [19] and illustrates the advantage of AAPA over APA for the filtering of general complex signals.

The paper is organised as follows. We first provide mathematical foundations for complex-valued second order statistics and the $\mathbb{CR}$ (Wirtinger) calculus [7,20,21], which is required for the derivation of a real-valued function of complex variables. Next, the derivation of the augmented affine projection algorithm (AAPA) is provided in an adaptive prediction context. The improvement in the performance of AAPA as compared to the standard APA is supported by both theoretical analysis and simulations performed on benchmark complex-valued both circular and noncircular signals, as well as on real world noncircular wind measurements. In this paper, the following notations are adopted: $(\cdot)_r$ and $(\cdot)_i$ denote respectively the real and imaginary parts of complex variables in partial derivative respectively, $E[\cdot]$ the statistical expectation operator, $|\cdot|$ the absolute value of a variable, $\|\cdot\|$ the Euclidean norm of a vector, $(\cdot)^*$ the complex conjugate, $(\cdot)^T$ the transpose of a vector or a matrix, $(\cdot)^H$ the Hermitian conjugate of a vector or a matrix, $\mathrm{Tr}(\cdot)$ the trace of a matrix, and $\Re(\cdot)$ the real part of complex numbers.

## 2. Augmented complex statistics and $\mathbb{CR}$ calculus

### 2.1. Augmented complex statistics

For a complex random vector (RV) $\mathbf{z} \in \mathbb{C}^L$, with $E\{\mathbf{z}\} = \mathbf{0}$, we can define two correlation matrices, $\mathcal{C}_{\mathbf{zz}} = E\{\mathbf{z}\mathbf{z}^H\}$ and $\mathcal{P}_{\mathbf{zz}} = E\{\mathbf{z}\mathbf{z}^T\}$, which are called respectively the covariance matrix and pseudo-covariance matrix [9]. Based on augmented complex statistics, in order to allow for completing the second order statistical information available within a complex RV to be utilised, the signal model should be based on the augmented $2L \times 1$ complex vector $\mathbf{z}^a = [\mathbf{z}^T, \mathbf{z}^H]^T$ [22]. Then, the augmented covariance matrix $\mathcal{C}_{\mathbf{z}^a\mathbf{z}^a}$ contains information from both the covariance and pseudo-covariance matrices of $\mathbf{z}$, and is given by [8]

$$\mathcal{C}_{\mathbf{z}^a\mathbf{z}^a} = E\begin{bmatrix}\mathbf{z}\\ \mathbf{z}^*\end{bmatrix}[\mathbf{z}^H\mathbf{z}^T] = \begin{bmatrix}\mathcal{C}_{\mathbf{zz}} & \mathcal{P}_{\mathbf{zz}}\\ \mathcal{P}_{\mathbf{zz}}^* & \mathcal{C}_{\mathbf{zz}}^*\end{bmatrix} \tag{2}$$

When $\mathcal{P}_{\mathbf{zz}} = \mathbf{0}$, the complex random vector is called second order circular (proper), indicating that the probability density functions for such processes are rotation invariant. However, in most applications, complex signals are noncircular, and the probability density functions are not rotation invariant. Second order noncircular signals are termed improper.

### 2.2. $\mathbb{CR}$ calculus

The Cauchy–Riemann conditions for a holomorphic complex function $f(z) = u(x,y) + \mathrm{j}v(x,y)$ impose very strict constraints, i.e. $\partial u/\partial x = \partial v/\partial y$ and $\partial u/\partial y = -\partial v/\partial x$. However, for real-valued functions with complex arguments, these conditions are not satisfied. For instance, our usual cost function, $E = \frac{1}{2}ee^* = \frac{1}{2}|e|^2$, a real-valued function of complex variables, is not analytic, and so it does not satisfy the Cauchy–Riemann conditions. Since a holomorphic function $f(z): \mathbb{C} \to \mathbb{C}$ can be viewed as a real bivariate function of its real and imaginary components $g(x,y): \mathbb{R}^2 \to \mathbb{R}^2$, the stringent conditions of standard complex derivative ($\mathbb{C}$−derivative) of $f(z)$ can be circumvented by using $\mathbb{CR}$ calculus. The main idea behind $\mathbb{CR}$ calculus is to introduce the so-called *conjugate coordinates*, which apply to any real or complex-valued function

dependent on both $z$ and $z^*$. This way, the $\mathbb{R}-$derivative of a real function of a complex variable $f=f(z,z^*)$ is given by

$$\frac{\partial f}{\partial z_{|z^*=const}}=\frac{1}{2}\left(\frac{\partial f}{\partial x}-\mathrm{j}\frac{\partial f}{\partial y}\right) \tag{3}$$

where $\partial f/\partial x$ and $\partial f/\partial y$ are the partial derivatives of the function $f(z)=f(z,z^*)=g(x,y)$, whereas the conjugate $\mathbb{R}-$derivative ($\mathbb{R}^*-$derivative) of a function $f(z)=f(z,z^*)$ is given by

$$\frac{\partial f}{\partial z^*_{|z=const}}=\frac{1}{2}\left(\frac{\partial f}{\partial x}+\mathrm{j}\frac{\partial f}{\partial y}\right) \tag{4}$$

For a complex analytic function, the $\mathbb{R}-$derivative is equivalent to the standard $\mathbb{C}-$derivative and the $\mathbb{R}^*-$derivative vanishes. For real functions of complex variables (cost function), the $\mathbb{R}^*-$derivative gives the standard pseudo-gradients. Thus, the generalised Cauchy–Riemann conditions can be expressed as $\partial f/\partial z^*=0$, that is, in stochastic gradient filtering in the complex domain, the gradient should be calculated with respect to the conjugate weight vector. This was used in the derivation of the proposed AAPA. For more detail, see an excellent overview by Kreutz-Delgado [20] and a recent book by Mandic and Goh [7].

### 2.3. Augmented CLMS

To illustrate the convenience of the use of $\mathbb{CR}$ calculus and to connect standard augmented CLMS (ACLMS) and AAPA, we shall now use $\mathbb{CR}$ calculus to derive ACLMS. Consider the output of an adaptive filter,

$$\mathbf{y}(k)=\underbrace{\mathbf{x}^T(k)\mathbf{h}(k)}_{\text{standard part}}+\underbrace{\mathbf{x}^H(k)\mathbf{g}(k)}_{\text{conjugate part}}$$

where $\mathbf{h}(k)$ and $\mathbf{g}(k)$ are the $L\times 1$ weight vectors of filter coefficients, and $\mathbf{x}(k)$ denotes the $L\times 1$ input vector at time instant $k$, defined as

$$\mathbf{x}(k)=[x(k),\ldots,x(k-L+1)]^T$$

The error $e(k)$ and cost function $J(k)$ can be defined as

$$e(k)=d(k)-y(k)\quad\text{and}\quad J(k)=\tfrac{1}{2}|e(k)|^2=\tfrac{1}{2}e(k)e^*(k)$$

and the update form of the standard part weight vector $\mathbf{h}(k)$ is given by

$$\mathbf{h}(k+1)=\mathbf{h}(k)-\mu\nabla J_{\mathbf{h}}(k)$$

where $\mu$ is the learning rate. Note that since $J(k)$ is a real-valued function of complex variables, the steepest decent points to the direction of $\partial J(k)/\mathbf{h}^*(k)$ (see Section 2.2). Following the $\mathbb{R}^*-$derivative (4)

$$\nabla J_{\mathbf{h}}(k)=\frac{\partial J(k)}{\partial \mathbf{h}^*(k)}=\frac{1}{2}\left(\frac{\partial J(k)}{\partial \mathbf{h}_r(k)}-\mathrm{j}\frac{\partial J(k)}{\partial \mathbf{h}_i(k)}\right)$$

and

$$\frac{\partial J(k)}{\partial \mathbf{h}_r(k)}=-\frac{1}{2}e(k)\mathbf{x}^*(k)\quad\text{and}\quad\frac{\partial J(k)}{\partial \mathbf{h}_i(k)}=-\frac{\mathrm{j}}{2}e(k)\mathbf{x}^*(k)$$

Finally, the derivative of cost function $J(k)$ with respect to $\mathbf{h}^*(k)$ becomes

$$\frac{\partial J(k)}{\partial \mathbf{h}^*(k)}=\frac{1}{2}e(k)\mathbf{x}^*(k)$$

and the update of $\mathbf{h}(k)$ becomes

$$\mathbf{h}(k+1)=\mathbf{h}(k)+\mu e(k)\mathbf{x}^*(k) \tag{5}$$

Note that a factor $\frac{1}{2}$ has been incorporated into $\mu$. In a similar way, we can obtain

$$\mathbf{g}(k+1)=\mathbf{g}(k)+\mu e(k)\mathbf{x}(k) \tag{6}$$

Combining $\mathbf{h}(k)$ and $\mathbf{g}(k)$ into an augmented weight vector $\mathbf{w}_a(k)$, we have

$$\mathbf{w}_a(k+1)=\mathbf{w}_a(k+1)+\mu e(k)\mathbf{x}_a^*(k)$$

where $\mathbf{x}_a(k)$ is the augmented input vector $\mathbf{x}_a(k)=[\mathbf{x}^T(k),\mathbf{x}^H(k)]^T$.

## 3. Augmented affine projection algorithm

The output vector $\mathbf{y}(k)$ of a FIR adaptive filter trained by AAPA can be written as

$$\mathbf{y}(k)=\underbrace{\mathbf{X}^T(k)\mathbf{h}(k)}_{\text{standard part}}+\underbrace{\mathbf{X}^H(k)\mathbf{g}(k)}_{\text{conjugate part}} \tag{7}$$

where $\mathbf{h}(k)$ and $\mathbf{g}(k)$ are the $L\times 1$ weight vectors, and $\mathbf{X}(k)$ is the $L\times K$ data matrix in the filter memory at time instant $k$, defined as [3]

$$\mathbf{X}(k)=[\mathbf{x}(k-K+1),\ldots,\mathbf{x}(k)] \tag{8}$$

where $K$ is the observation length, and $\mathbf{x}(k)$ denotes the $L\times 1$ input vector,

$$\mathbf{x}(k)=[x(k-1),\ldots,x(k-L)]^T \tag{9}$$

The vector of desired signals $\mathbf{d}(k)$ from the $K$ most recent observations is given by

$$\mathbf{d}(k)=[d(k-K+1),\ldots,d(k)]^T \tag{10}$$

whereas the error vector $\mathbf{e}(k)$ for the augmented affine projection filter is

$$\mathbf{e}(k)=\mathbf{d}(k)-\mathbf{y}(k) \tag{11}$$

To update the weight vectors $\mathbf{g}(k)$ of the "conjugate" part and $\mathbf{h}(k)$ of the "standard" part of a widely linear affine projection filter, similar to the standard APA approach [23], based on the principle of minimum disturbance, we need to solve a constrained minimisation problem given by

$$\begin{aligned}J(k)=&\|\mathbf{h}(k+1)-\mathbf{h}(k)\|^2+\|\mathbf{g}(k+1)-\mathbf{g}(k)\|^2\\&+\Re[(\mathbf{d}(k)-\mathbf{X}^T(k)\mathbf{h}(k+1)-\mathbf{X}^H(k)\mathbf{g}(k+1))^H\cdot\mathbf{\Lambda}]\end{aligned} \tag{12}$$

where the $K\times 1$ dimensional vector $\mathbf{\Lambda}$ comprises Lagrange multipliers and $J(k)$ is a real function of complex variables. Based on the widely linear model (7), the constraints in (12) become

$$\mathbf{d}(k)=\mathbf{X}^T(k)\mathbf{h}(k+1)+\mathbf{X}^H(k)\mathbf{g}(k+1) \tag{13}$$

To calculate the update for the "conjugate" part of the widely linear model, that is, for the weight vector $\mathbf{g}(k)$, the

cost function $J(k)$ can be rewritten as

$$J(k)=\|\mathbf{h}(k+1)-\mathbf{h}(k)\|^2+[\mathbf{g}(k+1)-\mathbf{g}(k)]^H[\mathbf{g}(k+1)-\mathbf{g}(k)]+\tfrac{1}{2}[(\mathbf{d}^H(k)-\mathbf{h}^H(k+1)\mathbf{X}^*(k)-\mathbf{g}^H(k+1)\mathbf{X}(k))\boldsymbol{\Lambda}]+\tfrac{1}{2}[(\mathbf{d}^T(k)-\mathbf{h}^T(k+1)\mathbf{X}(k)-\mathbf{g}^T(k+1)\mathbf{X}^*(k))\boldsymbol{\Lambda}^*]$$

As $J(k)$ is a real-valued function of complex variables (as shown in Section 2.2) the steepest descent points to the direction of $\partial J(k)/\partial\mathbf{g}^*(k+1)$. Following the $\mathbb{R}^*-$derivative in (4), we have

$$\frac{\partial J(k)}{\partial\mathbf{g}^*(k+1)}=\frac{1}{2}\left(\frac{\partial J(k)}{\partial\mathbf{g}_r(k+1)}+\mathrm{j}\frac{\partial J(k)}{\partial\mathbf{g}_i(k+1)}\right) \tag{14}$$

and

$$\frac{\partial J(k)}{\partial\mathbf{g}_r(k+1)}=\mathbf{g}(k+1)-\mathbf{g}(k)-\frac{1}{2}\mathbf{X}(k)\boldsymbol{\Lambda}$$

$$\frac{\partial J(k)}{\partial\mathbf{g}_i(k+1)}=-\mathrm{j}(\mathbf{g}(k+1)-\mathbf{g}(k))+\frac{\mathrm{j}}{2}\mathbf{X}(k)\boldsymbol{\Lambda}$$

Finally, the derivative of the cost function $J(k)$ with respect to the coefficients vector $\mathbf{g}^*(k+1)$ becomes

$$\frac{\partial J(k)}{\partial\mathbf{g}^*(k+1)}=\mathbf{g}(k+1)-\mathbf{g}(k)-\frac{1}{2}\mathbf{X}(k)\boldsymbol{\Lambda} \tag{15}$$

Setting this derivative to zero, the update of the "conjugate" part of the AAPA model is obtained as

$$\mathbf{g}(k+1)-\mathbf{g}(k)=\tfrac{1}{2}\mathbf{X}(k)\boldsymbol{\Lambda} \tag{16}$$

In the same way, the update of weight vector $\mathbf{h}(k)$ of the "standard" part of the AAPA model is given by

$$\mathbf{h}(k+1)-\mathbf{h}(k)=\tfrac{1}{2}\mathbf{X}^*(k)\boldsymbol{\Lambda} \tag{17}$$

Following the standard APA approach [23], to eliminate the Lagrange multiplier vector $\boldsymbol{\Lambda}$ from (16) and (17), we shall premultiply both sides of (16) by $\mathbf{X}^H(k)$ and both sides of (17) by $\mathbf{X}^T(k)$ to yield

$$\mathbf{X}^H(k)(\mathbf{g}(k+1)-\mathbf{g}(k))=\tfrac{1}{2}\mathbf{X}^H(k)\mathbf{X}(k)\boldsymbol{\Lambda}$$

$$\mathbf{X}^T(k)(\mathbf{h}(k+1)-\mathbf{h}(k))=\tfrac{1}{2}\mathbf{X}^T(k)\mathbf{X}^*(k)\boldsymbol{\Lambda}$$

Upon adding these two equations together and using the constraints (13) and the widely linear model (7), the output error becomes

$$\mathbf{e}(k)=\tfrac{1}{2}(\mathbf{X}^H(k)\mathbf{X}(k)+\mathbf{X}^T(k)\mathbf{X}^*(k))\boldsymbol{\Lambda} \tag{18}$$

giving the Lagrange multiplier vector

$$\boldsymbol{\Lambda}=2[\mathbf{X}^H(k)\mathbf{X}(k)+\mathbf{X}^T(k)\mathbf{X}^*(k)]^{-1}\mathbf{e}(k) \tag{19}$$

Thus, the update of the "conjugate" weight vector within the widely linear AAPA becomes

$$\mathbf{g}(k+1)=\mathbf{g}(k)+\mu\mathbf{X}(k)[\mathbf{X}^H(k)\mathbf{X}(k)+\mathbf{X}^T(k)\mathbf{X}^*(k)]^{-1}\mathbf{e}(k)$$

In a similar way, for the update of the "standard" weight vector within AAPA, we have

$$\mathbf{h}(k+1)=\mathbf{h}(k)+\mu\mathbf{X}^*(k)[\mathbf{X}^H(k)\mathbf{X}(k)+\mathbf{X}^T(k)\mathbf{X}^*(k)]^{-1}\mathbf{e}(k)$$

To avoid singularities due to the inversion of a rank deficient matrix, a positive constant $\delta$, called the regularisation parameter, is added to the above updates, giving

$$\mathbf{h}(k+1)=\mathbf{h}(k)+\mu\mathbf{X}^*(k)[\mathbf{X}^H(k)\mathbf{X}(k)+\mathbf{X}^T(k)\mathbf{X}^*(k)+\delta\mathbf{I}]^{-1}\mathbf{e}(k) \tag{20}$$

$$\mathbf{g}(k+1)=\mathbf{g}(k)+\mu\mathbf{X}(k)[\mathbf{X}^H(k)\mathbf{X}(k)+\mathbf{X}^T(k)\mathbf{X}^*(k)+\delta\mathbf{I}]^{-1}\mathbf{e}(k) \tag{21}$$

In the derivation, the estimation is assumed to be of sufficient order. Compared with the weight update of standard APA [23], the AAPA includes both the estimate of the covariance matrix $\mathbf{X}^T(k)\mathbf{X}^*(k)$ and its complex conjugate $\mathbf{X}^H(k)\mathbf{X}(k)$ in the denominator term within the update for both the weight vectors $\mathbf{h}(k)$ and $\mathbf{g}(k)$. Observe that, as desired, when the observation length is $K=1$, AAPA degenerates into the normalised ACLMS [7], given by

$$\mathbf{h}(k+1)=\mathbf{h}(k)+\frac{\mu\mathbf{x}^*(k)e(k)}{\mathbf{x}^H(k)\mathbf{x}(k)+\mathbf{x}^T(k)\mathbf{x}^*(k)+\delta}$$

$$\mathbf{g}(k+1)=\mathbf{g}(k)+\frac{\mu\mathbf{x}(k)e(k)}{\mathbf{x}^H(k)\mathbf{x}(k)+\mathbf{x}^T(k)\mathbf{x}^*(k)+\delta}$$

and when the denominator is omitted, we have the standard ACLMS algorithm [7]. Table 1 presents the computational complexity of the proposed AAPA compared with the standard APA based on [1]. Note that the differences in the computational complexity are marginal.

**Table 1**
Comparison of the computational complexity between standard APA and proposed AAPA.

| | Standard APA | Proposed AAPA |
|---|---|---|
| Multiplications | $(K^2+2K)L+K^3+K^2$ | $(2K^2+4K)L+K^3+K^2$ |
| Additions | $(K^2+2K)L+K^3+K^2$-$K$ | $(K^2+4K)L+K^3+2K^2$ |

## 4. Steady state mean square error (MSE) performance of APA and AAPA

The aim of this section is to demonstrate analytically the suitability of the proposed AAPA to filter complex-valued noncircular data, for which the standard APA is suboptimal, and to provide insight into its convergence. Consider a second order noncircular (improper) teaching signal $d(k)$ arising from the widely linear model

$$d(k)=\mathbf{x}^T(k)\mathbf{h}_o+\mathbf{x}^H(k)\mathbf{g}_o+\nu(k) \tag{22}$$

where $\mathbf{h}_o$ and $\mathbf{g}_o$ are unknown optimal weight vectors that we wish to estimate, $\nu(k)$ represents the measurement noise and $\mathbf{x}(k)$ is the input vector with a positive-definite covariance matrix $C_{\mathbf{xx}}=E[\mathbf{x}\mathbf{x}^H]$.

### 4.1. Steady state analysis of standard APA for noncircular signals

The steady state mean square error performance of the standard APA is evaluated based on the mean square error

$$\mathrm{MSE}=\lim_{k\to\infty}E[|e(k)|^2] \tag{23}$$

where $e(k)$, the first element of $\mathbf{e}(k)$, is the error at the time instant $k$, given by

$$e(k) = d(k) - \mathbf{x}^T(k)\mathbf{h}(k) \tag{24}$$

and $\mathbf{h}(k)$ is the weight vector, for which, the update is given by [23]

$$\mathbf{h}(k+1) = \mathbf{h}(k) + \mu\mathbf{X}^*(k)[\mathbf{X}^T(k)\mathbf{X}^*(k) + \delta\mathbf{I}]^{-1}\mathbf{e}(k) \tag{25}$$

We next follow the approach from [19], based on the energy conservation principle, to formulate a closed form solution of the MSE for improper signals represented by the widely linear model (22). Using this approach, the update in (25) can be rewritten in terms of the weight error vector $\tilde{\mathbf{h}}(k) = \mathbf{h}_o - \mathbf{h}(k)$ as

$$\tilde{\mathbf{h}}(k+1) = \tilde{\mathbf{h}}(k) - \mu\mathbf{X}^*(k)[\mathbf{X}^T(k)\mathbf{X}^*(k) + \delta\mathbf{I}]^{-1}\mathbf{e}(k) \tag{26}$$

Multiplying both sides of (26) by $\mathbf{X}^T(k)$ gives

$$\mathbf{X}^T(k)\tilde{\mathbf{h}}(k+1) = \mathbf{X}^T(k)\tilde{\mathbf{h}}(k) - \mu\mathbf{X}^T(k)\mathbf{X}^*(k)\cdot[\mathbf{X}^T(k)\mathbf{X}^*(k) + \delta\mathbf{I}]^{-1}\mathbf{e}(k) \tag{27}$$

Then the *a posteriori* and *a priori* error vectors $\mathbf{e}_p(k)$ and $\mathbf{e}_a(k)$ can be introduced as [23]

$$\mathbf{e}_p(k) = \mathbf{X}^T(k)\tilde{\mathbf{h}}(k+1) \quad \text{and} \quad \mathbf{e}_a(k) = \mathbf{X}^T(k)\tilde{\mathbf{h}}(k)$$

The MSE analysis of standard APA is given in [24], however, we here need to account for the differences coming from the widely linear model, as the standard APA is optimal only for second order circular (proper) data. This results in the following expression for $\mathbf{e}(k)$, obtained by substituting (22) into (24),

$$\mathbf{e}(k) = \mathbf{e}_a(k) + \mathbf{v}(k) + \mathbf{X}^H(k)\mathbf{g}_o \tag{28}$$

where $\mathbf{v}(k) = [v(k), v(k-1), \ldots, v(k-K+1)]^T$. Note that since we assume a widely linear model, compared with the work in [24], an extra term "$\mathbf{X}^H(k)\mathbf{g}_o$" is included. The excess mean square error (EMSE), can be defined as

$$\mathrm{EMSE} = \lim_{k\to\infty} E[|e_a(k)|^2] \tag{29}$$

We can now employ the energy conservation principle to derive the excess mean square error (EMSE) of standard APA, denoted by $\mathrm{EMSE}_s$. A sketch of the derivation is given in Appendix A; for a full derivation for standard APA applied to circular signals we refer to [24]. When the regularisation parameter $\delta$ is small enough, $\mathrm{EMSE}_s$ can be conveniently analysed for two cases: for a small learning rate $\mu$, we obtain

$$\mathrm{EMSE}_s = \frac{\mu\sigma_v^2}{2-\mu} + \frac{\mu\mathbf{g}_o^H E[\mathbf{C}_s(k)]\mathbf{g}_o}{\mathrm{Tr}(E[\mathbf{A}_s(k)])} \tag{30}$$

whereas, for a large learning rate, we obtain

$$\mathrm{EMSE}_s = \frac{\mu\sigma_v^2}{2-\mu}\mathrm{Tr}(\mathcal{C}_{xx})\left(E\left[\frac{K}{\|\mathbf{x}(k)\|^2}\right] + \mu\mathbf{g}_o^H E[\mathbf{C}_s(k)]\mathbf{g}_o\right) \tag{31}$$

where

$$\mathbf{A}_s(k) = [\mathbf{X}^T(k)\mathbf{X}^*(k) + \delta\mathbf{I}]^{-1}\mathbf{X}^T(k)\mathbf{X}^*(k)[\mathbf{X}^T(k)\mathbf{X}^*(k) + \delta\mathbf{I}]^{-1}$$

and

$$\mathbf{C}_s(k) = \mathbf{X}(k)\mathbf{A}_s(k)\mathbf{X}^H(k) \tag{32}$$

Compared with the expressions of $\mathrm{EMSE}_s$ for standard APA for circular signals in [24], both (30) and (31) contain an additional term due to the "conjugate" term "$\mathbf{X}^H(k)\mathbf{g}_o$" within the widely linear model (22), illustrating that the standard APA is a suboptimal solution for the filtering of second order improper (noncircular) signals.

### 4.2. Steady state analysis of augmented APA for noncircular signals

Similarly to the analysis in Section 4.1, the update of "standard" weight vector of AAPA in (20) can be rewritten in terms of its weight error vector $\tilde{\mathbf{h}}(k) = \mathbf{h}_o - \mathbf{h}(k)$ as

$$\tilde{\mathbf{h}}(k+1) = \tilde{\mathbf{h}}(k) - \mu\mathbf{X}^*(k)[\mathbf{X}^H(k)\mathbf{X}(k) + \mathbf{X}^T(k)\mathbf{X}^*(k) + \delta\mathbf{I}]^{-1}\mathbf{e}(k) \tag{33}$$

Multiplying both sides of (20) by $\mathbf{X}^T(k)$ gives

$$\mathbf{X}^T(k)\tilde{\mathbf{h}}(k+1) = \mathbf{X}^T(k)\tilde{\mathbf{h}}(k) - \mu\mathbf{X}^T(k)\mathbf{X}^*(k)[\mathbf{X}^H(k)\mathbf{X}(k) + \mathbf{X}^T(k)\mathbf{X}^*(k) + \delta\mathbf{I}]^{-1}\mathbf{e}(k) \tag{34}$$

For the "conjugate" weight error vector $\tilde{\mathbf{g}}(k)$, we have

$$\tilde{\mathbf{g}}(k+1) = \tilde{\mathbf{g}}(k) - \mu\mathbf{X}(k)[\mathbf{X}^H(k)\mathbf{X}(k) + \mathbf{X}^T(k)\mathbf{X}^*(k) + \delta\mathbf{I}]^{-1}\mathbf{e}(k) \tag{35}$$

and

$$\mathbf{X}^H(k)\tilde{\mathbf{g}}(k+1) = \mathbf{X}^H(k)\tilde{\mathbf{g}}(k) - \mu\mathbf{X}^H(k)\mathbf{X}(k)[\mathbf{X}^H(k)\mathbf{X}(k) + \mathbf{X}^T(k)\mathbf{X}^*(k) + \delta\mathbf{I}]^{-1}\mathbf{e}(k) \tag{36}$$

Note that both the terms $\tilde{\mathbf{h}}(k)$ and $\tilde{\mathbf{g}}(k)$ have the same generic form, and that the denominator in the above updates is different from that of standard APA in (26). Then the *a posteriori* and *a priori* error vectors $\mathbf{e}_p(k)$ and $\mathbf{e}_a(k)$ can be introduced as

$$\mathbf{e}_p(k) = \mathbf{X}^T(k)\tilde{\mathbf{h}}(k+1) + \mathbf{X}^H(k)\tilde{\mathbf{g}}(k+1) \tag{37}$$

$$\mathbf{e}_a(k) = \mathbf{X}^T(k)\tilde{\mathbf{h}}(k) + \mathbf{X}^H(k)\tilde{\mathbf{g}}(k) \tag{38}$$

and the sum of (34) and (36) gives the relationship between the *a posteriori* and *a priori* error vectors $\mathbf{e}_p(k)$ and $\mathbf{e}_a(k)$ in the form

$$\mathbf{e}_p(k) = \mathbf{e}_a(k) - \mu(\mathbf{X}^H(k)\mathbf{X}(k) + \mathbf{X}^T(k)\mathbf{X}^*(k)) \cdot [\mathbf{X}^H(k)\mathbf{X}(k) + \mathbf{X}^T(k)\mathbf{X}^*(k) + \delta\mathbf{I}]^{-1}\mathbf{e}(k) \tag{39}$$

The next step is to use both the *a posteriori* and *a priori* error vectors in the update equations (33) and (35) to obtain the updates of weight error vectors as functions of $\mathbf{e}_a(k)$ and $\mathbf{e}_p(k)$

$$\tilde{\mathbf{h}}(k+1) = \tilde{\mathbf{h}}(k) - \mathbf{X}^*(k)[\mathbf{X}^H(k)\mathbf{X}(k) + \mathbf{X}^T(k)\mathbf{X}^*(k)]^{-1}\cdot(\mathbf{e}_a(k) - \mathbf{e}_p(k)) \tag{40}$$

and

$$\tilde{\mathbf{g}}(k+1) = \tilde{\mathbf{g}}(k) - \mathbf{X}(k)[\mathbf{X}^H(k)\mathbf{X}(k) + \mathbf{X}^T(k)\mathbf{X}^*(k)]^{-1}\cdot(\mathbf{e}_a(k) - \mathbf{e}_p(k)) \tag{41}$$

Making the standard assumption that at the steady state $\tilde{\mathbf{h}}(k+1) \approx \tilde{\mathbf{h}}(k)$, and $\tilde{\mathbf{g}}(k+1) \approx \tilde{\mathbf{g}}(k)$, the application of the energy conservation principle gives

$$\|\tilde{\mathbf{h}}(k+1)\|^2 + \tilde{\mathbf{h}}^H(k)\mathbf{X}^*(k)[\mathbf{X}^H(k)\mathbf{X}(k) + \mathbf{X}^T(k)\mathbf{X}^*(k)]^{-1}\mathbf{e}_a(k) = \|\tilde{\mathbf{h}}(k)\|^2 + \tilde{\mathbf{h}}^H(k+1)\mathbf{X}^*(k)[\mathbf{X}^H(k)\mathbf{X}(k) + \mathbf{X}^T(k)\mathbf{X}^*(k)]^{-1}\mathbf{e}_p(k)$$

and

$$\|\tilde{\mathbf{g}}(k+1)\|^2+\tilde{\mathbf{g}}^H(k)\mathbf{X}(k)[\mathbf{X}^H(k)\mathbf{X}(k)+\mathbf{X}^T(k)\mathbf{X}^*(k)]^{-1}\mathbf{e}_a(k)$$
$$=\|\tilde{\mathbf{g}}(k)\|^2+\tilde{\mathbf{g}}^H(k+1)\mathbf{X}(k)[\mathbf{X}^H(k)\mathbf{X}(k)+\mathbf{X}^T(k)\mathbf{X}^*(k)]^{-1}\mathbf{e}_p(k)$$

Upon adding together the above two equations and applying the statistical expectation operator, as $k\to\infty$, $E[\|\tilde{\mathbf{h}}(k+1)\|^2]=E[\|\tilde{\mathbf{h}}(k)\|^2]$ and $E[\|\tilde{\mathbf{g}}(k+1)\|^2]=E[\|\tilde{\mathbf{g}}(k)\|^2]$, and we obtain

$$E[\mathbf{e}_a^H(k)[\mathbf{X}^H(k)\mathbf{X}(k)+\mathbf{X}^T(k)\mathbf{X}^*(k)]^{-1}\mathbf{e}_a(k)]$$
$$=E[\mathbf{e}_p^H(k)[\mathbf{X}^H(k)\mathbf{X}(k)+\mathbf{X}^T(k)\mathbf{X}^*(k)]^{-1}\mathbf{e}_p(k)] \quad (42)$$

Similarly to the analysis in Appendix A, it can be shown that, for a small learning rate $\mu$, the EMSE of AAPA, denoted as $\mathrm{EMSE}_a$, is given by

$$\mathrm{EMSE}_a=\frac{\mu\sigma_v^2}{2-\mu} \quad (43)$$

whereas for a large $\mu$, we obtain

$$\mathrm{EMSE}_a=\frac{\mu\sigma_v^2}{2-\mu}\mathrm{Tr}(\mathcal{C}_{xx})E\left[\frac{K}{\|\mathbf{x}(k)\|^2}\right] \quad (44)$$

Observe that the EMSE is proportional to the observation length $K$.

Note that in [18], a widely linear APA (WL-APA) has been derived by solving two separate constrained minimisation problems, where

$$J_1(k)=\|\mathbf{h}(k+1)-\mathbf{h}(k)\|^2+\Re[(\mathbf{d}(k)-\mathbf{X}^T(k)\mathbf{h}(k+1)-\mathbf{X}^H(k)\mathbf{g}(k+1))^H\cdot\mathbf{\Lambda}_1] \quad (45)$$

$$J_2(k)=\|\mathbf{g}(k+1)-\mathbf{g}(k)\|^2+\Re[(\mathbf{d}(k)-\mathbf{X}^T(k)\mathbf{h}(k+1)-\mathbf{X}^H(k)\mathbf{g}(k+1))^H\cdot\mathbf{\Lambda}_2] \quad (46)$$

Compared with the global cost function of the proposed AAPA, this WL-APA employs two different Lagrange multiplier vectors, with additional assumptions that for the update of $\mathbf{h}(k)$, $\mathbf{g}(k+1)\approx\mathbf{g}(k)$, and vice versa, leading to the following weight updates:

$$\mathbf{h}(k+1)=\mathbf{h}(k)+\mu\mathbf{X}^*(k)[\mathbf{X}^T(k)\mathbf{X}^*(k)+\delta\mathbf{I}]^{-1}\mathbf{e}(k)$$

$$\mathbf{g}(k+1)=\mathbf{g}(k)+\mu\mathbf{X}(k)[\mathbf{X}^H(k)\mathbf{X}(k)+\delta\mathbf{I}]^{-1}\mathbf{e}(k)$$

In this case, the weight update forms of $\mathbf{h}(k)$ and $\mathbf{g}(k)$ do not share the same denominator, which causes additional difficulties to find the closed form relationship between the *a posteriori* and *a priori* error vectors $\mathbf{e}_p(k)$ and $\mathbf{e}_a(k)$ in widely linear model as compared with AAPA, rendering the algorithm suboptimal.

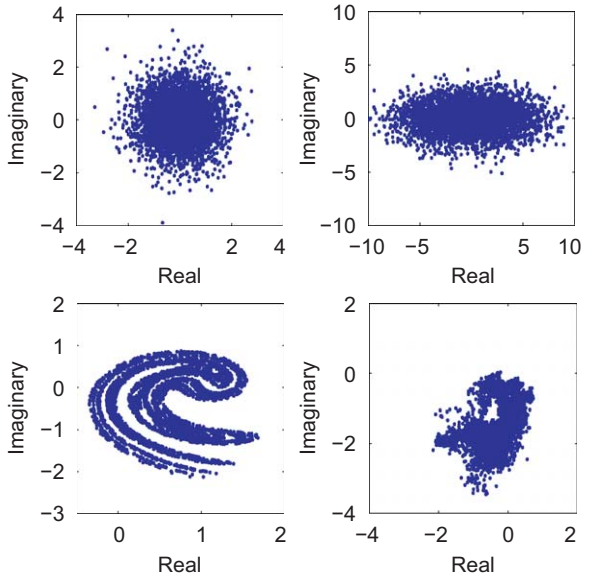


**Fig. 1.** Geometric view of circularity. (a) Circular AR(1) model (47); (b) noncircular ARMA model (48); (c) Ikeda map (50); and (d) wind (*high*) signal.

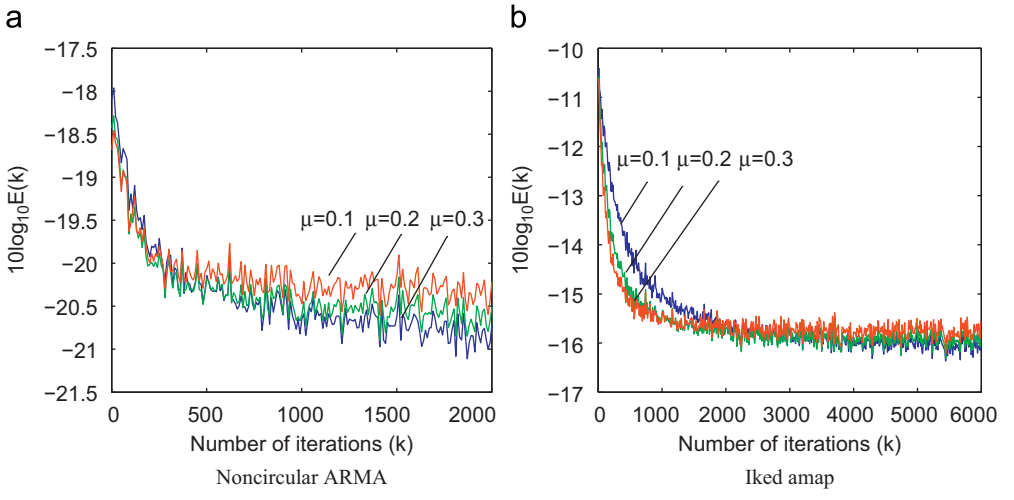


**Fig. 2.** Learning curves of AAPA for the noncircular ARMA model (48) and noncircular Ikeda map (50) for $K=1$ and various learning rate $\mu$. (a) Noncircular ARMA and (b) Ikeda map.

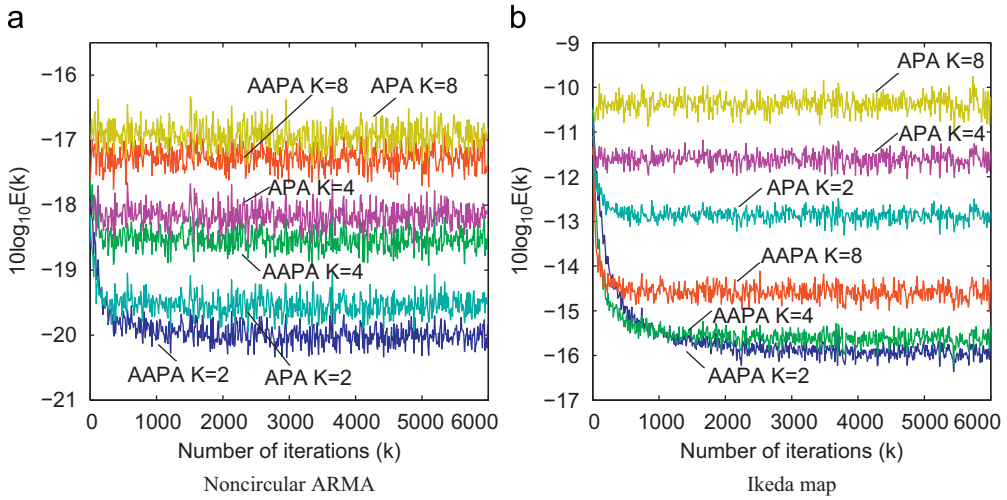


**Fig. 3.** Learning curves of APA and AAPA for the noncircular ARMA model (48) and noncircular Ikeda map (50), obtained by using $\mu = 0.1$ and $K = 2$, 4, and 8. (a) Noncircular ARMA and (b) Ikeda map.

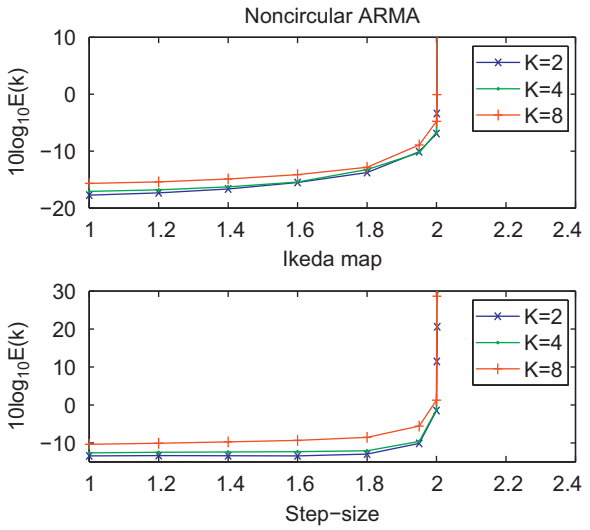


**Fig. 4.** MSE of AAPA as a function of step-size $\mu$ for the noncircular ARMA model (48) and Ikeda map (50).

### 4.3. Comparison of steady state performance of APA and AAPA

Consider, without loss in generality, a complex-valued noncircular signal is generated from a widely linear model (22), so that $\mu \mathbf{g}_o^H E[\mathbf{C}_s(k)]\mathbf{g}_o > 0$ (proof is given in Appendix B). This, in turn, makes $\mathrm{EMSE}_s > \mathrm{EMSE}_a$ ($\mathrm{MSE}_s > \mathrm{MSE}_a$, since $\mathrm{MSE} = \mathrm{EMSE} + \sigma_v^2$) for both small and large learning rates at the steady state. On the other hand, when a complex-valued circular signal is considered, the weight vector $\mathbf{g}_o = \mathbf{0}$, leading to $\mathrm{EMSE}_s = \mathrm{EMSE}_a$ ($\mathrm{MSE}_s = \mathrm{MSE}_a$), resulting in both the APA and AAPA having the same performance at the steady state.

## 5. Simulations

To verify the benefits of AAPA as compared with standard APA, simulations were conducted for both small and large learning rates, in a one step ahead prediction setting. In all the simulations, the filter length was $L = 10$, and the regularisation parameter $\delta = 0.01$. Comprehensive statistical tests comparing APA and AAPA were performed on benchmark complex-valued circular and noncircular signals, and the graphs were produced by averaging 200 iterations of independent trials. In addition, single trial simulations were performed on real-world nonstationary and noncircular wind signals.[1] The test signals employed in the simulations were:

- The linear circular complex signal was a stable AR (1) model, given by

$$r(k) = 0.5r(k-1) + n(k) \tag{47}$$

  where $n(k)$ is complex valued doubly white Gaussian noise with unit variance ($n_r \perp n_i$ and $\sigma_{n_r}^2 = \sigma_{n_i}^2$).
- The benchmark widely linear noncircular complex signal was an autogressive moving average ARMA complex process, made by combining the MA model in [25] and the stable AR(1) model, given by

$$r(k) = 0.5r(k-1) + 2n(k) + 0.5n^*(k) + n(k-1) + 0.9n^*(k-1) \tag{48}$$

  with

$$E\{n(k-i)n^*(k-j)\} = \delta(i-j)$$

$$E\{n(k-i)n(k-j)\} = \varepsilon\delta(i-j) \tag{49}$$

[1] Wind can be represented as a bivariate process of its speed and direction, that is, $w(k) = |v(k)|e^{j\Phi(k)}$, where $v(k)$ denotes the speed and $\Phi(k)$ the direction.

where $n(k)$ was complex valued doubly white Gaussian noise with unit variance and $\varepsilon = 0$.

- The nonlinear and noncircular chaotic Ikeda map signal, given by [26]

$$x(k+1) = 1 + u(x(k)\cos[t(k)] - y(k)\sin[t(k)])$$

$$y(k+1) = u(x(k)\sin[t(k)] + y(k)\cos[t(k)]) \quad (50)$$

where typically $u = 0.9$ and $t(k) = 0.4 - \{6/(1 + x^2(k) + y^2(k))\}$.

- The noncircular wind signals employed were with different dynamical characteristics, identified as regions of *high*, *medium* and *low* dynamics, based on the changes in the wind intensity.

Fig. 1 shows the scatter plots of the complex-valued signals considered. Observe the circular symmetry of the distribution for the AR(1) model (47) and the noncircularity of the ARMA model (48), Ikeda map (50) and Wind (*high*) signal; for a detailed account of the noncircularity of complex signals we refer to [7]. Fig. 2 shows learning curves of AAPA for the noncircular ARMA process and Ikeda map, plotted for different small step-sizes, indicating that for noncircular signals, for a small step-size, increasing the step-size of AAPA will cause a larger MSE at the steady state, although with faster convergence. Fig. 3 illustrates the advantage of the AAPA at steady state over the standard APA on both the noncircular ARMA model and Ikeda map for the step-size was $\mu = 0.1$, and observation lengths $K = 2$, 4, and 8.

In the next experiment, the performances of APA and AAPA for large $\mu$ were considered. The results in [24] indicate that the stability bound on $\mu$ is approximately $0 < \mu < 2$ for APA, and Fig. 4 illustrates the stability bound of AAPA on the noncircular signals was $\mu < 2$. The simulation results in Fig. 5 illustrate that with a fixed

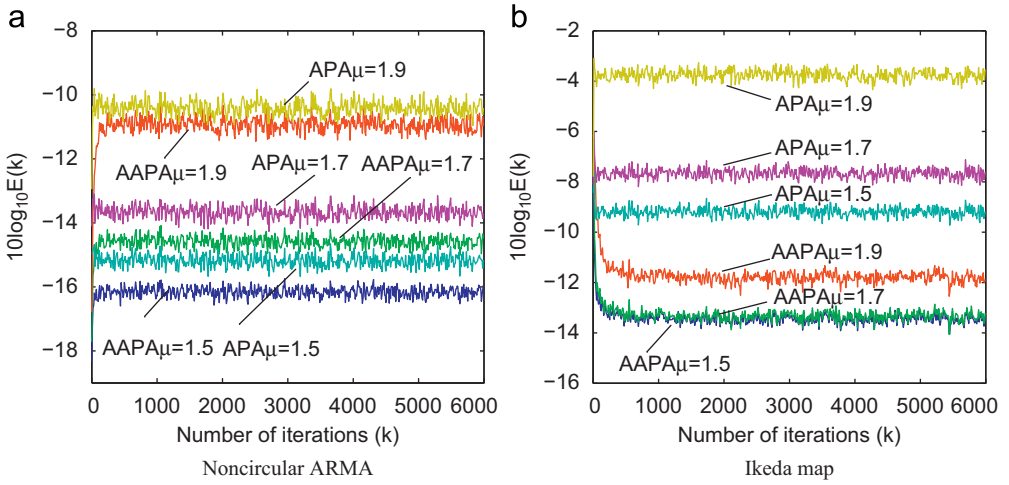


**Fig. 5.** Learning curves of APA and AAPA for the noncircular ARMA model (48) and noncircular Ikeda map (50) using different large $\mu$ and $K = 2$. (a) Noncircular ARMA and (b) Ikeda map.

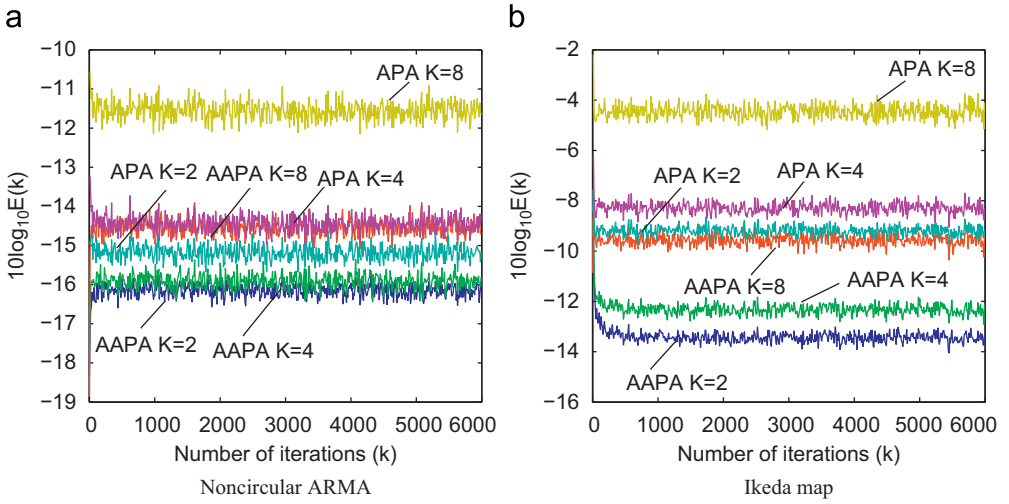


**Fig. 6.** *Learning* curves of APA and AAPA for the noncircular ARMA model (48) and noncircular Ikeda map (50) using $\mu = 1.5$ and $K = 2$, 4, and 8. (a) Noncircular ARMA and (b) Ikeda map.

observation length $K$, an increment in the step-size will cause MSE of AAPA to increase at the steady state of AAPA, whereas Fig. 6 shows that with a fixed large $\mu$, the MSE at the steady state increases with an increment in $K$. Note that both Figs. 5 and 6 illustrate the advantage of AAPA over APA for noncircular signals considered with a large $\mu$. The learning curves of both APA and AAPA on the circular AR(1) model for $K=1$ and $\mu=0.1$, 0.2, 1.5, and 1.7 are shown in Fig. 7, which indicates that APA converges faster than AAPA for the circular signal, however, at the steady state, there is no clear difference between the performances. Thus, although AAPA optimises both the weight vectors $\mathbf{h}(k)$ and $\mathbf{g}(k)$ of the widely linear model, for circular signals, $\mathbf{g}_o=\mathbf{0}$, leading to the same model as that of standard APA.

To further illustrate the advantage of using the AAPA over APA, in the next experiment, the prediction gain $R_p$ was used as an overall performance measurement, defined as

$$R_p(k)=10\log_{10}\left(\frac{\sigma_x^2}{\hat{\sigma}_e^2}\right)[\mathrm{dB}] \tag{51}$$

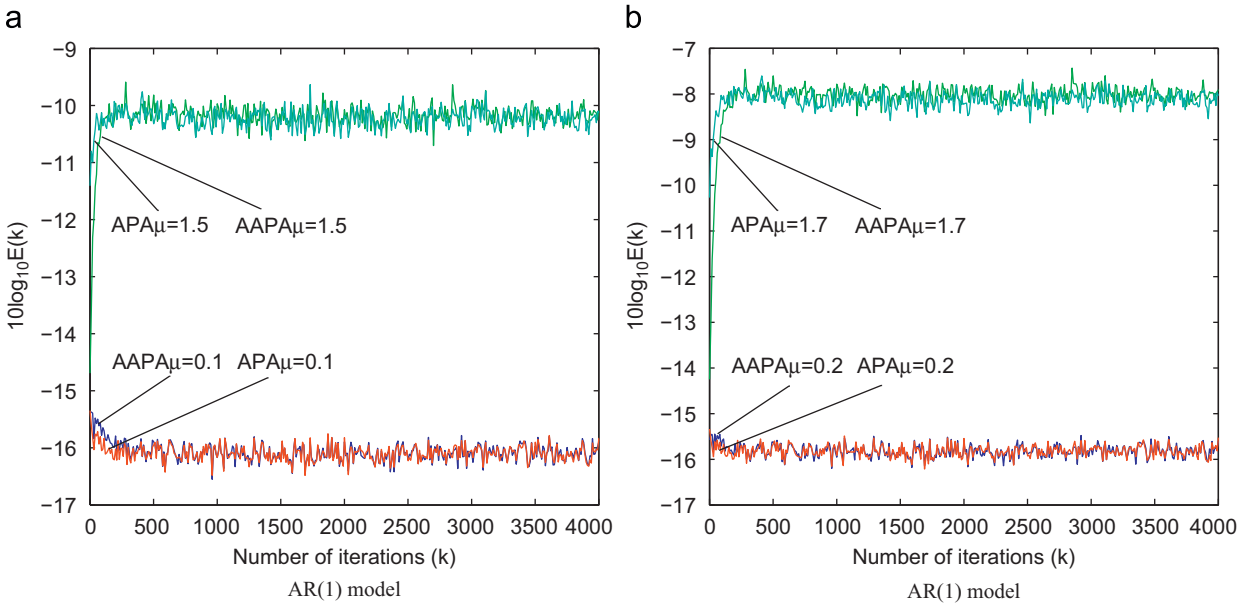


**Fig. 7.** Learning curves of APA and AAPA for the proper AR(1) model (47) using different small and large $\mu$ and $K=1$. (a) AR(1) model and (b) AR(1) model.

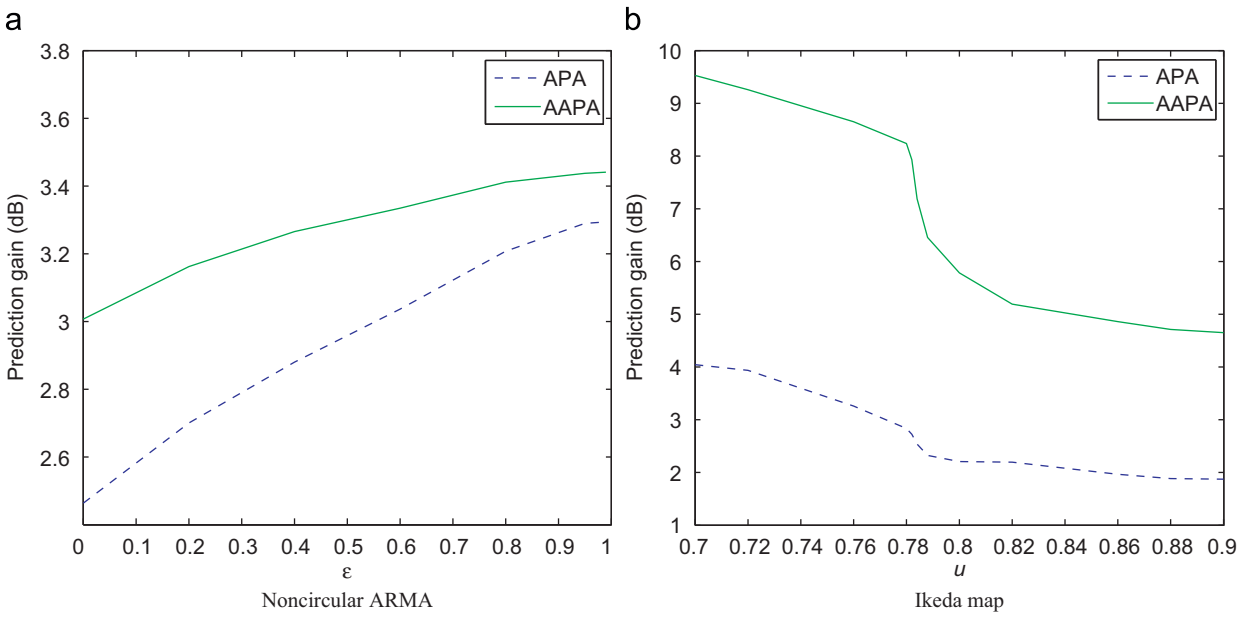


**Fig. 8.** Prediction gains of APA and AAPA for varying values of the control parameters of noncircular ARMA (48) and Ikeda map (50). (a) Noncircular ARMA and (b) Ikeda map.

**Table 2**
Prediction gains $R_p$ of standard APA and the proposed AAPA for various noncircular signals.

| $R_p$ (dB) | Noncircular ARMA | Ikeda | Wind (*low*) | Wind (*medium*) | Wind (*high*) |
|---|---|---|---|---|---|
| APA | 2.4641 | 1.8734 | 3.6774 | 5.8851 | 9.3610 |
| AAPA | 3.0072 | 4.6517 | 4.0500 | 6.2368 | 9.9181 |

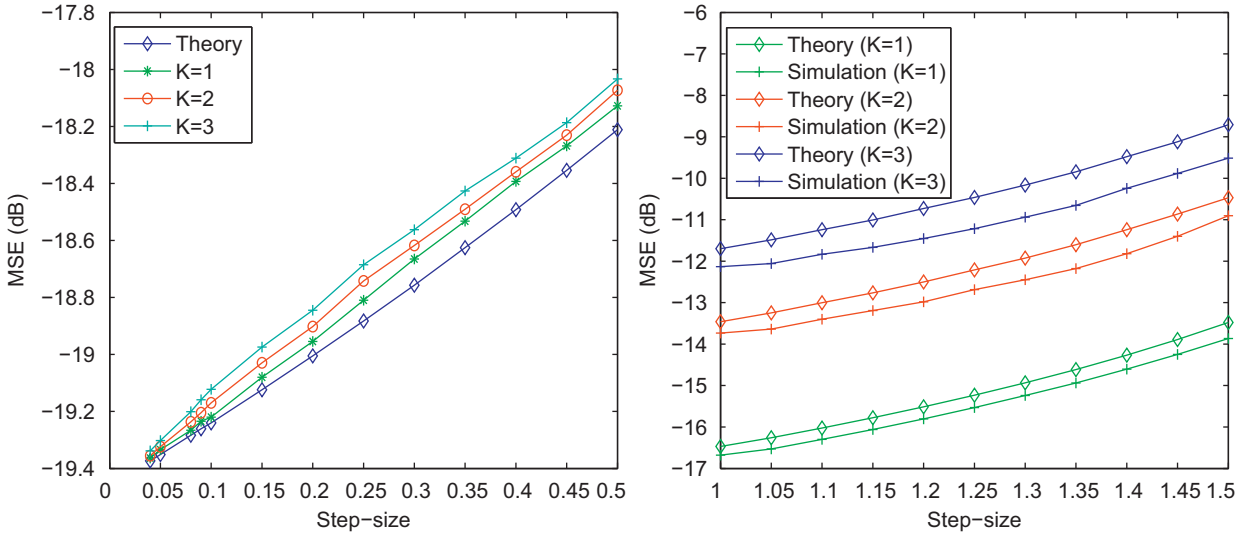


**Fig. 9.** Comparison of theoretical and simulated steady state MSE of AAPA on noncircular ARMA (48) when $K=1$, 2 and 3 using (a) (43) and (b) (44).

where $\sigma_x^2$ denotes the variance of the input signal $x(k)$, and $\hat{\sigma}_e^2$ denotes the variance of the prediction error $e(k)$ [27]. Table 2 shows the prediction gains $R_p$ of the proposed AAPA and standard APA for $\mu=0.2$ and $K=2$. As expected, for all noncircular signals considered, there was a significant improvement when the AAPA was employed over that of the standard APA. Fig. 8 shows the effects of the control parameters $\varepsilon$ (49) and $u$ (50) on the performances of standard APA and AAPA for noncircular ARMA and Ikeda map, in all the cases, the performance of AAPA was superior than that of APA. Note that in Fig. 8(a), AAPA gradually loses advantage when $\varepsilon$ increases to 1, which conforms with the experimental results in [25].

We next performed one step ahead prediction on noisy noncircular ARMA signal, where the measurement noise was added to the noncircular ARMA signal in (48) such that the signal-to-noise ratio (SNR) was 30 dB. Fig. 9(a) and (b) illustrate the theoretical steady state MSE of AAPA using (43) and (44) as a function of step-size with $K=1$, 2, and 3, and both showed good agreement with the simulation results.

## 6. Conclusions

We have introduced an augmented affine projection algorithm (AAPA) for adaptive filtering of complex valued noncircular signals. The AAPA has been derived based on recent advances in the so-called augmented complex statistics, the widely linear model and $\mathbb{CR}$ calculus. Steady state analysis of MSE performance of AAPA and APA has proved the advantage of AAPA over standard APA for adaptive filtering of second order noncircular signals, and comparable performance for adaptive filtering of circular signals. Comprehensive simulations on both synthetic and real world complex-valued, circular and noncircular signals support the analysis.

## Acknowledgements

We would like to thank Gill Instruments Ltd, who have provided ultrasonic anemometers used for our wind measurements.

## Appendix A. Derivation of EMSE of APA for noncircular signals

Following the approach in [1,24], the energy conservation relation is given by

$$\|\tilde{\mathbf{h}}(k+1)\|^2+\mathbf{e}_a^H(k)[\mathbf{X}^T(k)\mathbf{X}^*(k)]^{-1}\mathbf{e}_a(k) = \|\tilde{\mathbf{h}}(k)\|^2+\mathbf{e}_p^H(k)[\mathbf{X}^T(k)\mathbf{X}^*(k)]^{-1}\mathbf{e}_p(k) \quad (52)$$

Taking statistical expectation of both sides, and as $k\to\infty$, $E[\|\tilde{\mathbf{h}}(k+1)\|^2]=E[\|\tilde{\mathbf{h}}(k)\|^2]$, the following equations hold:

$$\mu E[\mathbf{e}^H(k)\mathbf{A}_s(k)\mathbf{e}(k)]=E[\mathbf{e}_a^H(k)\mathbf{B}_s(k)\mathbf{e}(k)]+E[\mathbf{e}^H(k)\mathbf{B}_s(k)\mathbf{e}(k)] \quad (53)$$

where

$$\mathbf{A}_s(k)=[\mathbf{X}^T(k)\mathbf{X}^*(k)+\delta\mathbf{I}]^{-1}\mathbf{X}^T(k)\mathbf{X}^*(k)[\mathbf{X}^T(k)\mathbf{X}^*(k)+\delta\mathbf{I}]^{-1}$$

and

$$\mathbf{B}_s(k)=[\mathbf{X}^T(k)\mathbf{X}^*(k)+\delta\mathbf{I}]^{-1}$$

Note that

$$\mathbf{e}(k)=\mathbf{e}_a(k)+\mathbf{v}(k)+\mathbf{X}^H(k)\mathbf{g}_o \tag{54}$$

With the assumptions that the noise $v(k)$ is *i.i.d.* and statistically independent of the input matrix $\mathbf{X}(k)$, and at the steady state $\mathbf{X}(k)$ is statistically independent of $\mathbf{e}_a(k)$, $E[\mathbf{e}_a(k)]=\mathbf{0}$ as $k\to\infty$, and the first term in (53) becomes

$$\begin{aligned}\mu E[\mathbf{e}^H(k)\mathbf{A}_s(k)\mathbf{e}(k)]&=\mu E[\mathbf{e}_a^H(k)\mathbf{A}_s(k)\mathbf{e}_a(k)]\\&\quad+\mu E[\mathbf{v}^H(k)\mathbf{A}_s(k)\mathbf{v}(k)]\\&\quad+\mu E[(\mathbf{X}^H(k)\mathbf{g}_o)^H\mathbf{A}_s(k)\mathbf{X}^H(k)\mathbf{g}_o]\end{aligned}$$

where the second term has the form

$$E[\mathbf{e}_a^H(k)\mathbf{B}_s(k)\mathbf{e}(k)]=E[\mathbf{e}_a^H(k)\mathbf{B}_s(k)\mathbf{e}_a(k)]$$

and the third term

$$E[\mathbf{e}^H(k)\mathbf{B}_s(k)\mathbf{e}_a(k)]=E[\mathbf{e}_a^H(k)\mathbf{B}_s(k)\mathbf{e}_a(k)]$$

Then (53) can be rewritten as

$$\mu E[\mathbf{e}_a^H(k)\mathbf{A}_s(k)\mathbf{e}_a(k)]+\mu E[\mathbf{v}^H(k)\mathbf{A}_s(k)\mathbf{v}(k)]+\mu E[(\mathbf{X}^H(k)\mathbf{g}_o)^H\mathbf{A}_s(k)\mathbf{X}^H(k)\mathbf{g}_o]$$
$$=2E[\mathbf{e}_a^H(k)\mathbf{B}(k)\mathbf{e}_a(k)] \tag{55}$$

To compute these statistical expectations, we shall further assume $E[\mathbf{e}_a\mathbf{e}_a^H]=E[|e_a(k)|]^2\cdot S$, where $S\approx\mathbf{I}$ for small $\mu$, and $S\approx\mathbf{1}\cdot\mathbf{1}^T$ for large $\mu$, with $\mathbf{1}=[1,0,\ldots,0]^T$. As $k\to\infty$, the first term in (55) now becomes

$$\begin{aligned}\mu E[\mathbf{e}_a^H(k)\mathbf{A}_s(k)\mathbf{e}_a(k)]&=\mu\,\mathrm{Tr}(E[\mathbf{e}_a^H(k)\mathbf{e}_a(k)\mathbf{A}_s(k)])\\&=\mu E[|e_a(k)|]^2\,\mathrm{Tr}(S\cdot E[\mathbf{A}_s(k)])\end{aligned}$$

where symbol $\mathrm{Tr}(\cdot)$ denotes the matrix trace operator. Similarly, the second term in (55) becomes

$$\mu E[\mathbf{v}^H(k)\mathbf{A}_s(k)\mathbf{v}(k)]=\mu\sigma_v^2\,\mathrm{Tr}(E[\mathbf{A}_s(k)])$$

whereas the third and fourth terms become

$$2E[\mathbf{e}_a^H(k)\mathbf{B}_s(k)\mathbf{e}_a(k)]=2E|e_a(k)|^2\,\mathrm{Tr}(S\cdot E[\mathbf{B}_s(k)])$$

and

$$\mu E[(\mathbf{X}^H(k)\mathbf{g}_o)^H\mathbf{A}_s(k)\mathbf{X}^H(k)\mathbf{g}_o]=\mu\mathbf{g}_o^HE[\mathbf{C}_s(k)]\mathbf{g}_o$$

where

$$\mathbf{C}_s(k)=\mathbf{X}(k)\mathbf{A}_s(k)\mathbf{X}^H(k)$$

With the introduction of following equations:

$$\alpha_\mu\triangleq\mathrm{Tr}(S\cdot E[\mathbf{A}_s(k)])\quad\text{and}\quad\eta_\mu\triangleq\mathrm{Tr}(S\cdot E[\mathbf{B}_s(k)])$$

When the regularisation parameter $\delta$ is small enough, $\mathbf{A}_s(k)=\mathbf{B}_s(k)$, and $\eta_\mu=\alpha_\mu$. In this case, the EMSE of the standard APA, denoted by $\mathrm{EMSE}_s$, becomes

$$\mathrm{EMSE}_s=\frac{\mu\sigma_v^2\,\mathrm{Tr}(E[\mathbf{A}_s(k)])+\mu\mathbf{g}_o^HE[\mathbf{C}_s(k)]\mathbf{g}_o}{2\eta_\mu-\mu\alpha_\mu}$$

For a small enough $\delta$ and for a small learning rate $\mu$, $\mathbf{S}\approx\mathbf{I}$, we obtain

$$\mathrm{EMSE}_s=\frac{\mu\sigma_v^2}{2-\mu}+\frac{\mu\mathbf{g}_o^HE[\mathbf{C}_s(k)]\mathbf{g}_o}{\mathrm{Tr}(E[\mathbf{A}_s(k)])} \tag{56}$$

On the other hand, for a large learning rate, $\mathbf{S}\approx\mathbf{1}\cdot\mathbf{1}^T$, and we obtain

$$\mathrm{EMSE}_s=\frac{\mu\sigma_v^2}{2-\mu}\,\mathrm{Tr}(\mathcal{C}_{xx})\left(E\left[\frac{K}{\|\mathbf{x}(k)\|^2}\right]+\mu\mathbf{g}_o^HE[\mathbf{C}_s(k)]\mathbf{g}_o\right) \tag{57}$$

## Appendix B

To prove that $\mu\mathbf{g}_o^HE[\mathbf{C}_s(k)]\mathbf{g}_o>0$, observe that when $\delta$ is small enough, $\mathbf{g}_o^H\mathbf{C}_s(k)\mathbf{g}_o$ can be written as $\mathbf{g}_o^H\mathbf{X}(k)[\mathbf{X}^T(k)\mathbf{X}^*(k)+\delta\mathbf{I}]^{-1}(\mathbf{g}_o^H\mathbf{X}(k))^H$. Since $\delta>0$, the inversion term $[\mathbf{X}^T(k)\mathbf{X}^*(k)+\delta\mathbf{I}]^{-1}$ is positive definite [28]. Note that $\mathbf{g}_o^H\mathbf{X}(k)$ is a $1\times K$ vector with all nonzero elements, since for complex-valued noncircular signal $\mathbf{g}_o$ is with all nonzero elements. Following the properties of positive definite matrix [29], $\mathbf{g}_o^H\mathbf{C}_s(k)\mathbf{g}_o>0$, then $\mu\mathbf{g}_o^HE[\mathbf{C}_s(k)]\mathbf{g}_o>0$.

## References


[1] A.H. Sayed, Fundamentals of Adaptive Filtering, Wiley, New York, 2003.
[2] T. Hinamoto, S. Maekawa, Extended theory of learning identification, Transactions of IEEJ 95 (10) (1975) 227–234.
[3] K. Ozeki, T. Umeda, An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties, Transactions of the ICEIE 67-A (5) (1984) 126–132.
[4] Y. Choi, H. Shin, W. Song, Adaptive regularization matrix for affine projection algorithm, IEEE Transactions on Circuits and Systems II: Express Briefs 54 (12) (2007) 1087–1091.
[5] K. Hwang, W. Song, An affine projection adaptive filtering algorithm with selective regressors, IEEE Transactions on Circuits and Systems II: Express Briefs 54 (1) (2007) 43–46.
[6] M. Montazeri, P. Duhamel, A set of algorithms linking NLMS and block RLS algorithms, IEEE Transactions on Signal Processing 43 (2) (1995) 444–453.
[7] D.P. Mandic, S.L. Goh, Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models, Wiley, New York, 2009.
[8] P.J. Schreier, L.L. Scharf, Second-order analysis of improper complex random vectors and process, IEEE Transactions on Signal Processing 51 (3) (2003) 714–725.
[9] B. Picinbono, P. Chevalier, Widely linear estimation with complex data, IEEE Transactions on Signal Processing 43 (8) (1995) 2030–2033.
[10] M. Valkama, M. Renfors, V. Koivunen, Advanced methods for I/Q imbalance compensation in communication receivers, IEEE Transactions on Signal Processing 49 (10) (2001) 2335–2344.
[11] M. Valkama, V. Koivunen, Circular and non-circular complex random signals with applications in wireless communications, Tutorial in Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, ICASSP, 2007.
[12] R. Schober, W.H. Gerstacker, L.H.J. Lampe, Data-aided and blind stochastic gradients for widely linear MMSE MAI suppression for DS-CDMA, IEEE Transactions on Signal Processing 52 (3) (2004) 746–756.
[13] S. Javidi, M. Pedzisz, S.L. Goh, D.P. Mandic, The augmented complex least mean square algorithm with application to adaptive prediction problems, in: 1st IARP Workshop on Cognitive Information Processing, 2008, pp. 54–57.
[14] S.C. Douglas, Widely-linear recursive least-squares algorithm for adaptive beamforming, in: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, ICASSP, 2009, pp. 2041–2044.
[15] A. Kuh, D.P. Mandic, Applications of complex augmented kernels to wind profile prediction, in: Proceedings of the International Conference on Acoustics Speech and Signal Processing, ICASSP, 2009, pp. 3581–3584.
[16] C. Cheong Took, D.P. Mandic, Adaptive IIR filtering of noncircular complex signals, IEEE Transactions on Signal Processing 57 (10) (2009) 4111–4118.
[17] A.A. de Lima, R.C. de Lamare, Adaptive detection using widely linear processing with data reusing for DS-CDMA systems, in: International Telecommunications Symposium, 2006, pp. 187–192.

[18] Y. Xia, C. Cheong Took, S. Javidi and D.P. Mandic, A widely linear affine projection algorithm, in: IEEE Workshop on Statistical Signal Processing, SSP, 2009, pp. 373–376.
[19] M. Rupp, A.H. Sayed, A time-domain feedback analysis of filtered error adaptive gradient algorithms, IEEE Transactions on Signal Processing 44 (1996) 1428–1439.
[20] K. Kreutz-Delgado, The complex gradient operator and the $\mathbb{CR}$ calculus, in: Lecture Supplement ECE275A, 2006, pp. 1–74.
[21] D.H. Brandwood, A complex gradient operator and its application in adaptive array theory, IEE Proceedings H 130 (1) (1983) 11–16.
[22] A. van den Bos, The multivariate complex normal distribution: a generalization, IEEE Transactions on Information Theory 41 (2) (1995) 537–539.
[23] P.S.R. Diniz, Adaptive Filtering: Algorithms and Practical Implementation, Springer, Berlin, 2008.
[24] H. Shin, A.H. Sayed, Mean-square performance of a family of affine projection algorithms, IEEE Transactions on Signal Processing 52 (2004) 90–102.
[25] J. Navarro-Moreno, ARMA prediction of widely linear systems by using the innovations algorithm, IEEE Transactions on Signal Processing 56 (7) (2008) 3061–3068.
[26] K. Aihara, Applied Chaos and Applicable Chaos, Science-Sha, 1994.
[27] D.P. Mandic, J.A. Chambers, Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability, Wiley, New York, 2001.
[28] H.G. Ray, L.R. Vega, S. Tressens, B.C. Frias, Analysis of explicit regularization in affine projection algorithm: robustness and optimal choice, in: Proceedings of EUSIPCO, 2004, pp. 1809–1812.
[29] R.A. Horn, C.R. Johnson, Matrix Analysis, Cambridge University Press, Cambridge, UK, 1985.