Analysis of the Widely Linear Complex Kalman Filter

Dahir H. Dini [1], Danilo P. Mandic [2]

Department of Electrical and Electronic Engineering, Imperial College London
Exhibition Road, London, SW7 2BT, UK
[1] dahir.dini@imperial.ac.uk
[2] d.mandic@imperial.ac.uk

***Abstract*—The augmented complex Kalman filter (ACKF) has been recently proposed for the modeling of noncircular complex-valued signals for which widely linear modelling is more suitable than a strictly linear model. This has been achieved in the context of neural network training, however, the extent to which the ACKF outperforms the conventional complex Kalman filter (CCKF) in standard adaptive filtering applications remains unclear. In this paper, we show analytically that the ACKF algorithm achieves a lower mean squared error than the CCKF algorithm for noncircular signals. The analysis is supported by illustrative simulations.**

I. Introduction

Standard applications of adaptive systems normally use the signal magnitude as the main source of information. However, many real world applications rely on information from both the magnitude and direction or are multidimensional (e.g. radar, sonar and wind). This has lead to research towards extending the results from real-valued adaptive filters and standard complex domain filters to those employing augmented complex statistics and widely linear modelling. One popular adaptive system is the Kalman filter, which is optimal in the sense of second order statistics. Both the complex and real version of the Kalman filter exist, however the augmented complex version have only been deployed for the training of neutral networks and assume a random walk state foundations [1].

The implementation of the Kalman filter uses complete specifications of both dynamical and statistical model parameters of a system. In order to give the best results, the Kalman filter requires that the model parameters to be an accurate representation of the system. These statistical model parameters for complex valued signals are, however, not straightforward extensions of their real valued counterparts.

The second order statistical properties of a complex signal z is usually characterised by its covariance $r = E\{zz^*\}$. However, this is not sufficient for a complete second-order description, and it is necessary to consider another moment called the pseudocovariance $p = E\{zz\}$. It is only for the special class of complex signals known as second order *circular* or *proper*, that is, those with rotation invariant probability distributions and have vanishing pseudocovariance, that the their covariance function suffices to give the complete second-order description. However most real world processes are noncircular, either due to the different signal powers in the real and imaginary parts, or due to nonstationarity.

Several *augmented* or *widely linear* adaptive filtering algorithms have been developed by accounting for the information in both the covariance and pseudocovariance. These include the augmented complex least mean square [2], the augmented complex IIR filter [3], the augmented complex recursive least squares [4] and the augmented complex affine projection algorithm [5]. These are all based on the widely linear model

$$y = \mathbf{h}^T\mathbf{x} + \mathbf{g}^T\mathbf{x}^* \tag{1}$$

where y is the output, $\mathbf{h}$ and $\mathbf{g}$ are complex coefficient vectors, whereas $\mathbf{x}$ is the input vector and $\mathbf{x}^*$ is its complex conjugate.

The second order statistics of a complex random vector $\mathbf{z}$ are not fully described by its covariance, $\mathbf{R_z} = E\{\mathbf{z}\mathbf{z}^H\}$, as is the case for a real random vector. A second moment function called the pseudocovariance $\mathbf{P_z} = E\{\mathbf{z}\mathbf{z}^T\}$ (also known as the relation function or complementary covariance) is also needed in order to fully capture the second order statistics [6]. A complex-valued signal is said to be non-circular (or improper) if $\mathbf{z}$ and $e^{j\theta}\mathbf{z}$ have different probability density functions for any value of θ; otherwise it is circular (or proper) [7]. A circular signal, for which $\mathbf{P_z} = \mathbf{0}$, is fully described by its covariance alone. However for a non-circular signal, $\mathbf{P_z} \neq \mathbf{0}$, the pseudocovariance can not be ignored as it contains crucial information [8]. To cater for both $\mathbf{R_z}$ and $\mathbf{P_z}$ we can use the augmented signal vector $\mathbf{z}^a = \begin{bmatrix}\mathbf{z}^T & \mathbf{z}^H\end{bmatrix}^T$ for which the covariance matrix becomes

$$\mathbf{R_{z^a}} = \begin{bmatrix}\mathbf{R_z} & \mathbf{P_z} \\ \mathbf{P_z^*} & \mathbf{R_z^*}\end{bmatrix} \tag{2}$$

Hence, the use of widely linear (or augmented) signal models are expected to offer better second order performance and modelling capabilities for non-circular systems.

In this paper we consider the augmented complex Kalman Filter (ACKF), which uses both the pseudocovariance matrix and the covariance matrix in its statistical model parameters in order to achieve increased performance gains for noncircular signals. We show that the augmented complex Kalman filter always has the same or better performance than the conventional complex Kalman filter (CCKF), for the generality of complex signals, both circular and noncircular.

II. Analysis of the ACKF

A Kalman filter is an optimal sequential state estimator for linear dynamical systems, in the sense that it achieves the minimum mean squared error (MMSE). It is essentially a recursive filter that estimates the state of a linear dynamic system from a series of noisy observations. Its applications include state estimation for vehicular navigation systems, training of recurrent neural networks (RNNs) and time varying channel estimation. Consider a state space model given by [9]

$$\mathbf{x}_n = \mathbf{F}_{n-1}\mathbf{x}_{n-1} + \mathbf{w}_n \quad (3)$$
$$\mathbf{y}_n = \mathbf{H}_n\mathbf{x}_n + \mathbf{v}_n \quad (4)$$

where $\mathbf{x}_n$ is the state to be estimated (of dimension $p \times 1$) and $\mathbf{y}_n$ is the noisy observation (of dimension $q \times 1$). The vectors $\mathbf{w}_n$ and $\mathbf{v}_n$ are respectively the state noise and measurement noise. They are zero mean with covariance matrices $\mathbf{Q}_n$ and $\mathbf{R}_n$ respectively. The matrix $\mathbf{F}$ is the state transition matrix (of dimension $p \times p$) whereas $\mathbf{H}$ is the observation matrix (of dimension $q \times p$). Based on the widely linear model in (1) the augmented state space model can be written as [1]

$$\mathbf{x}^a_n = \mathbf{F}^a_{n-1}\mathbf{x}^a_{n-1} + \mathbf{w}^a_n \quad (5)$$
$$\mathbf{y}^a_n = \mathbf{H}^a_n\mathbf{x}^a_n + \mathbf{v}^a_n \quad (6)$$

where $\mathbf{x}^a_n = \begin{bmatrix}\mathbf{x}_n \\ \mathbf{x}^*_n\end{bmatrix}$, $\mathbf{y}^a_n = \begin{bmatrix}\mathbf{y}_n \\ \mathbf{y}^*_n\end{bmatrix}$, $\mathbf{w}^a_n = \begin{bmatrix}\mathbf{w}_n \\ \mathbf{w}^*_n\end{bmatrix}$, $\mathbf{v}^a_n = \begin{bmatrix}\mathbf{v}_n \\ \mathbf{v}^*_n\end{bmatrix}$, $\mathbf{F}^a_n = \begin{bmatrix}\mathbf{F}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{F}^*_n\end{bmatrix}$ and $\mathbf{H}^a = \begin{bmatrix}\mathbf{H}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{H}^*_n\end{bmatrix}$.

The covariance matrices of the augmented system and measurement noises, $\mathbf{w}_n$ and $\mathbf{v}_n$, then become

$$\mathbf{Q}^a_n = \begin{bmatrix}\mathbf{Q}_n & \mathbf{P}_n \\ \mathbf{P}^*_n & \mathbf{Q}^*_n\end{bmatrix} \quad (7)$$

$$\mathbf{R}^a_n = \begin{bmatrix}\mathbf{R}_n & \mathbf{U}_n \\ \mathbf{U}^*_n & \mathbf{R}^*_n\end{bmatrix} \quad (8)$$

where $\mathbf{P}_n$ and $\mathbf{U}_n$ are the pseudocovariance matrices of $\mathbf{w}_n$ and $\mathbf{v}_n$.

The MMSE estimator $\widehat{\mathbf{x}}^a_{n|n} = E[\mathbf{x}^a_n|\mathbf{y}^a_0, \mathbf{y}^a_1, ..., \mathbf{y}^a_n]$ of $\mathbf{x}^a_n$ based on $\{\mathbf{y}^a_0, \mathbf{y}^a_1, ..., \mathbf{y}^a_n\}$ can then be computed sequentially using the following recursion:

Prediction:

$$\widehat{\mathbf{x}}^a_{n|n-1} = \mathbf{F}^a_{n-1}\widehat{\mathbf{x}}^a_{n-1|n-1} \quad (9)$$

Minimum Prediction MSE Matrix:

$$\mathbf{M}^a_{n|n-1} = \mathbf{F}^a_{n-1}\mathbf{M}^a_{n-1|n-1}(\mathbf{F}^a_{n-1})^H + \mathbf{Q}^a_n \quad (10)$$

Kalman Gain Matrix:

$$\mathbf{G}_n = \mathbf{M}^a_{n|n-1}(\mathbf{H}^a_n)^H[\mathbf{H}^a_n\mathbf{M}^a_{n|n-1}(\mathbf{H}^a_n)^H + \mathbf{R}^a_n]^{-1} \quad (11)$$

Correction:

$$\widehat{\mathbf{x}}^a_{n|n} = \widehat{\mathbf{x}}^a_{n|n-1} + \mathbf{G}_n(\mathbf{y}^a_n - \mathbf{H}^a_n\widehat{\mathbf{x}}^a_{n|n-1}) \quad (12)$$

Minimum MSE Matrix:

$$\mathbf{M}^a_{n|n} = (\mathbf{I} - \mathbf{G}_n\mathbf{H}^a_n)\mathbf{M}^a_{n|n-1} \quad (13)$$

The minimum MSE estimate of $\mathbf{x}_n$ is $\widehat{\mathbf{x}}^a_{n|n}$ and can be expressed as

$$\widehat{\mathbf{x}}^a_{n|n} = (\mathbf{F}^a_{n-1} - \mathbf{G}_n\mathbf{H}^a_n\mathbf{F}^a_{n-1})\widehat{\mathbf{x}}^a_{n-1|n-1} + \mathbf{G}_n\mathbf{y}^a_n \quad (14)$$

For $\widehat{\mathbf{x}}^a_{0|0} = E\{\mathbf{x}_0\} = \mathbf{0}$ we have the following time evolution for $\widehat{\mathbf{x}}^a_{n|n}$

$$\begin{aligned}
\widehat{\mathbf{x}}^a_{0|0} &= \mathbf{0} \\
\widehat{\mathbf{x}}^a_{1|1} &= \mathbf{G}_1\mathbf{y}^a_1 \\
\widehat{\mathbf{x}}^a_{2|2} &= (\mathbf{F}^a_1 - \mathbf{G}_2\mathbf{H}^a_2\mathbf{F}^a_1)\mathbf{G}_1\mathbf{y}^a_1 + \mathbf{G}_2\mathbf{y}^a_2 \\
\widehat{\mathbf{x}}^a_{3|3} &= (\mathbf{F}^a_2 - \mathbf{G}_3\mathbf{H}^a_3\mathbf{F}^a_2)(\mathbf{F}^a_1 - \mathbf{G}_2\mathbf{H}^a_2\mathbf{F}^a_1)\mathbf{G}_1\mathbf{y}^a_1 \\
&\quad + (\mathbf{F}^a_2 - \mathbf{G}_3\mathbf{H}^a_3\mathbf{F}^a_2)\mathbf{G}_2\mathbf{y}^a_2 + \mathbf{G}_3\mathbf{y}^a_3 \\
\widehat{\mathbf{x}}^a_{4|4} &= (\mathbf{F}^a_3 - \mathbf{G}_4\mathbf{H}^a_4\mathbf{F}^a_3)(\mathbf{F}^a_2 - \mathbf{G}_3\mathbf{H}^a_3\mathbf{F}^a_2)(\mathbf{F}^a_1 - \mathbf{G}_2\mathbf{H}^a_2\mathbf{F}^a_1)\mathbf{G}_1\mathbf{y}^a_1 \\
&\quad + (\mathbf{F}^a_3 - \mathbf{G}_4\mathbf{H}^a_4\mathbf{F}^a_3)(\mathbf{F}^a_2 - \mathbf{G}_3\mathbf{H}^a_3\mathbf{F}^a_2)\mathbf{G}_2\mathbf{y}^a_2 \\
&\quad + (\mathbf{F}^a_3 - \mathbf{G}_4\mathbf{H}^a_4\mathbf{F}^a_3)\mathbf{G}_3\mathbf{y}^a_3 + \mathbf{G}_4\mathbf{y}^a_4 \\
&\vdots \\
\widehat{\mathbf{x}}^a_{n|n} &= \mathbf{W}_n\mathbf{Y}_n
\end{aligned} \quad (15)$$

where

$$\mathbf{Y}_n = [\mathbf{y}^{aT}_n, \mathbf{y}^a_{n-1}{}^T, \mathbf{y}^a_{n-2}{}^T, \ldots, \mathbf{y}^{aT}_1]^T$$

and

$$\mathbf{W}_n = \begin{bmatrix} [\mathbf{G}_n]^T \\ [(\mathbf{F}^a_{n-1} - \mathbf{G}_n\mathbf{H}^a_n\mathbf{F}^a_{n-1})\mathbf{G}_{n-1}]^T \\ [(\mathbf{F}^a_{n-1} - \mathbf{G}_n\mathbf{H}^a_n\mathbf{F}^a_{n-1})(\mathbf{F}^a_{n-2} - \mathbf{G}_{n-1}\mathbf{H}^a_{n-1}\mathbf{F}^a_{n-2})\mathbf{G}_{n-2}]^T \\ \vdots \\ [(\prod_2^{m=n}(\mathbf{F}^a_{m-1} - \mathbf{G}_m\mathbf{H}^a_m\mathbf{F}^a_{m-1}))\mathbf{G}_1]^T \end{bmatrix}^T$$

The length of the vector $\mathbf{Y}_n$ and columns of the matrix $\mathbf{W}_n$ increase with n. The weight matrix $\mathbf{W}_n$ can be seen as a function of the current and all the previous Kalman gains.

If the state and observations noises have a Gaussian distribution, then the Kalman filter is optimal in the MMSE sense and the weight matrix $\mathbf{W}_n$ is hence the minimum variance linear weighting matrix. The mean square error of any linear estimator,$\mathbf{W}'_n\mathbf{Y}_n$, may be compared with the mean square error for the optimal widely linear estimator $\widehat{\mathbf{x}}^a_{n|n} = \mathbf{W}_n\mathbf{Y}_n$ by writing [10, p. 327]

$$\begin{aligned}
&E\{(\mathbf{x}_n - \mathbf{W}'_n\mathbf{Y}_n)(\mathbf{x}_n - \mathbf{W}'_n\mathbf{Y}_n)^H\} \\
&\quad = E\{(\mathbf{x}_n - \mathbf{W}_n\mathbf{Y}_n)(\mathbf{x}_n - \mathbf{W}_n\mathbf{Y}_n)^H\} \\
&\qquad + E\{(\mathbf{W}_n\mathbf{Y}_n - \mathbf{W}'_n\mathbf{Y}_n)(\mathbf{W}_n\mathbf{Y}_n - \mathbf{W}'_n\mathbf{Y}_n)^H\} \\
&\quad \geq E\{(\mathbf{x}_n - \mathbf{W}_n\mathbf{Y}_n)(\mathbf{x}_n - \mathbf{W}_n\mathbf{Y}_n)^H\}
\end{aligned} \quad (16)$$

The cross-terms on the right-hand side vanish because the error $\mathbf{x}_n - \mathbf{W}_n\mathbf{Y}_n$ is orthogonal to every measurable function of $\mathbf{Y}_n$.

For an initialisation with $\mathbf{M}^a_{0|0} = E\{\mathbf{x}_0\mathbf{x}^H_0\} = \mathbf{0}$, the Kalman gain can be shown to be a recursive function of $\mathbf{Q}^a_n$

and $\mathbf{R}_n^a$. The optimal weight matrix $\mathbf{W}_n$, which is a function of the Kalman gain, is therefore also a function $\mathbf{Q}_n^a$ and $\mathbf{R}_n^a$. A particular choice for a potentially suboptimal weight matrix, $\mathbf{W}_n^{'}$, is a weight matrix that is a function of $\mathbf{Q}_n^L$ and $\mathbf{R}_n^L$, where

$$\mathbf{Q}_n^L = \begin{bmatrix} \mathbf{Q}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_n^* \end{bmatrix} \quad (17)$$

$$\mathbf{R}_n^L = \begin{bmatrix} \mathbf{R}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_n^* \end{bmatrix} \quad (18)$$

which corresponds to the conventional complex Kalman filter that does not take into account the pseudocovariances $\mathbf{P}_n$ and $\mathbf{U}_n$. Hence the augmented complex Kalman filter always has the same or better performance than the conventional complex Kalman filter, because of its utilisation of the pseudocovariances and consequently the full available augmented complex statistics.

III. SIMULATIONS

The performances of CCKF and ACKF were examined for their ability to track an autoregressive AR(4) process

$$\begin{aligned} x_n &= 1.79x_{n-1} - 1.85x_{n-2} + 1.27x_{n-3} \\ &\quad -0.41_{n-4} + u_n, n \geq 1 \end{aligned} \quad (19)$$

with the driving noise defined as

$$\begin{aligned} E\{u_{n-i}u_{n-l}^*\} &= c_u\delta_{i-l} \\ E\{u_{n-i}u_{n-l}\} &= p_u\delta_{i-l} \end{aligned} \quad (20)$$

Where δ is the discrete Dirac delta function. We used the ratio of the magnitude of the pseudocovariance (p) to the variance (c) as measure for the circularity of the complex random variables, that is

$$K = \frac{|p|}{c}$$

The driving noise $u(n)$ is circular if its pseudocovariance is zero (i.e. $K = 0$) and noncircular for all other values of K. Figure 1 shows a real-imaginary scatter plot for two different realisations of $u(n)$ with different levels of circularity. Note the circular symmetry for the circular signal and the non-circular shape for $K = 0.95$.

The performance of ACKF was assessed for the one-step ahead prediction of an AR(4) process, the noncircular *Lorenz* attractor, the chaotic *Ikeda* map and some real world *Wind* data using different AR models. The Wind data[1] was collected from measurements of the wind speeds in the north-south and east-west directions. These were then used to form the real and complex parts of a signal.

The Kalman filter was used to track the output of a noisy auto regressive process of order q, AR(q), that is generated according to the equation

$$x_n = \sum_{k=1}^{p} a_k x_{n-k} + u_n \quad (21)$$

[1]The wind data was provided by Prof. Aihara's team at the Institute of Industrial Science, University of Tokyo, Japan.

where u_n is the driving noise. It is assumed that x_n is observed in the presence of complex white noise v_n such that

$$y_n = x_n + v_n \quad (22)$$

The Kalman filter state space and observation equations then become [11]

$$\mathbf{x}_n = \begin{bmatrix} a_1 & a_2 & \dots & a_{p-1} & a_p \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix} \mathbf{x}_{n-1} + \begin{bmatrix} u_n \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (23)$$

$$y_n = [1, 0, ..., 0]\mathbf{x}_n + v_n \quad (24)$$

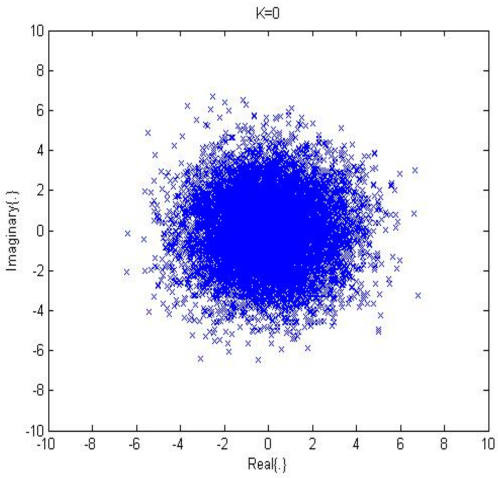
where $\mathbf{x}_n = [x_n, x_{n-1}, ..., x_{n-p+1}]^T$ is the state vector. We can then augment the Kalman filter parameters as in (3) and (4), to obtain the augmented complex Kalman filter.

For a quantitative assessment of the performance, the standard prediction gain $R_p = 10\log(\sigma_y^2/\sigma_e^2)$ was used, where σ_y^2 and σ_e^2 are the variances of the input signal and the output error. Figure 2 shows a comparison of the performance of the CCKF and the ACKF for the AR(4) process. Figure 2a illustrates the results for a circular observation noise of unit variance and a driving noise of unit variance but with various degrees of non-circularity, while Figure 2b shows the performance for a noncircular observation noise with a circular state noise. For both sets of results the two filters gave the same performance for the circular signals, i.e. $K = 0$. However, for noncircular noises, the ACKF outperformed the CCKF; the performance of the ACKF relative to CCKF, increased as the degree of non-circularity of the signals increased.

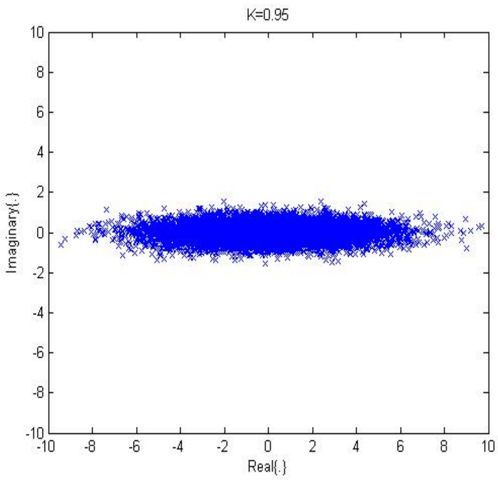
Table I summarises the prediction gains of CCKF and ACKF for the one-step ahead prediction of an AR(4) process, the *Lorenz* attractor, the *Ikeda* map and real world *Wind* data. In conformance with the analysis, in all the cases, the ACKF had the better prediction gain.

TABLE I
ONE-STEP AHEAD PREDICTION GAINS R_p FOR THE VARIOUS CLASSES SIGNALS

Signal	R_p(dB) (CCKF)	R_p(dB) (ACKF)
AR(4)(K=0)	11.97	11.97
AR(4)(K=0.9)	11.97	13.08
AR1(Lorenz)	67.46	70.07
AR4(Lorenz)	79.18	79.66
AR2(Ikeda)	1.22	2.27
AR4(Ikeda)	4.90	5.73
AR1(Wind)	22.13	22.98
AR2(Wind)	23.01	23.67
AR4(Wind)	23.33	23.64
AR6(Wind)	23.72	23.76

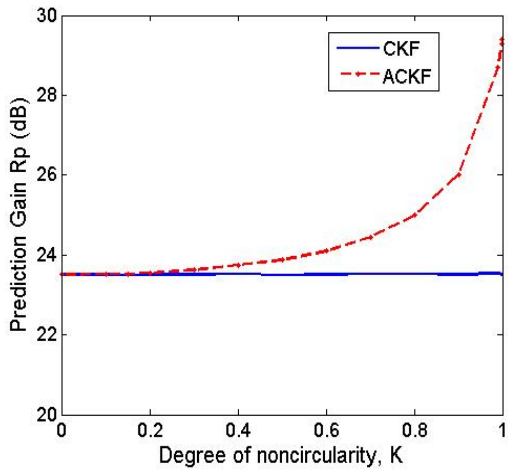


(a)

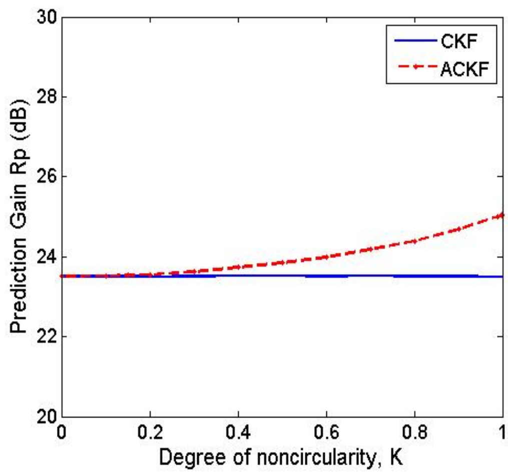


(b)

Fig. 1. A geometric view of circularity via a real-imaginary scatter plot. (a) a circular signal ($K = 0$); (b) a non-circular signal ($K = 0.95$).



(a)



(b)

Fig. 2. AR(4) Prediction gain comparison of the CCKF and ACKF for (a) a noncircular state noise and circular observation noise and (b) a circular state noise and noncircular observation noise

IV. Conclusion

We have introduced an augmented complex Kalman filter (ACKF) algorithm and have examined its performance in relation to the conventional complex Kalman filter (CCKF). The analysis has shown that it have the potential to offer significant performance gains over the CCKF for noncircular signals and similar performance to the CCKF for circular signals. Simulations for both synthetic and real world signals support the analysis.

References

[1] S. Goh and D. P. Mandic, "An Augmented Extended Kalman Filter Algorithm for Complex-Valued Recurrent Neural Networks," *Neural Computations*, vol. 19, pp. 1039–1055, 2007.

[2] S. Javidi, M. Pedzisz, S. Goh, and D.P. Mandic, "The augmented complex least mean square algorithm with application to adaptive prediction problems," *Proc. 1st IARP Workshop Cogn. Inf. Process*, pp. 54–57, 2008.

[3] C. Cheong-Took and D. Mandic, "Adaptive IIR filtering of noncircular complex signals," *IEEE Trans. Signal Process*, vol. 57, no. 10, pp. 4111–4118, 2009.

[4] S. Douglas, "Widely-linear recursive least-squares algorithm for adaptive beamforming," *Acoustics, Speech and Signal Processing*, pp. 2041–2044, 2009.

[5] Y. Xia, C. Cheong-Took, S. Javidi, and D.P. Mandic, "An augmented affine projection algorithm for the filtering of noncircular complex signals," *Signal Processing*, vol. 57, pp. 1788–1799, 2010.

[6] F. Neeser and J. Massey, "Proper Complex Random Processes with Applications to Information Theory," *IEEE Transactions on Information Theory*, vol. 39, no. 4, pp. 1293–1302, 1993.

[7] B. Picinbono, "On Circularity," *IEEE Transactions on Signal Processing*, vol. 42, no. 12, pp. 3473–3482, 1994.

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

[9] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall International, Inc, 1993.

[10] L. L. Scharf, *Statistical Signal Processing*. Addison Wesley, 1991.

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