Widely Linear State Space Filtering of Improper Complex Signals

by
Dahir H. Dini
MEng(Hons)

PhD Thesis

Communications and Signal Processing Research Group
Department of Electrical and Electronic Engineering
Imperial College London
2013

Copyright

Abstract

Complex signals are the backbone of many modern applications, such as power systems, communication systems, biomedical sciences and military technologies. However, standard complex valued signal processing approaches are suited to only a subset of complex signals known as proper, and are inadequate of the generality of complex signals, as they do not fully exploit the available information. This is mainly due to the inherent blindness of the algorithms to the complete second order statistics of the signals, or due to under-modelling of the underlying system. The aim of this thesis is to provide enhanced complex valued, state space based, signal processing solutions for the generality of complex signals and systems.

This is achieved based on the recent advances in the so called augmented complex statistics and widely linear modelling, which have brought to light the limitations of conventional statistical complex signal processing approaches. Exploiting these developments, we propose a class of widely linear adaptive state space estimation techniques, which provide a unified framework and enhanced performance for the generality of complex signals, compared with conventional approaches. These include the linear and nonlinear Kalman and particle filters, whereby it is shown that catering for the complete second order information and system models leads to significant performance gains. The proposed techniques are also extended to the case of cooperative distributed estimation, where nodes in a network collaborate locally to estimate signals, under a framework that caters for general complex signals, as well as the cross-correlations between observation noises, unlike earlier solutions. The analysis of the algorithms are supported by numerous case studies, including frequency estimation in three phase power systems, DIFAR sonobuoy underwater target tracking, and real-world wind modeling and prediction.

Acknowledgment

I would like to thank everyone who helped and supported me during my time at Imperial. In particular I would like to express my gratitude to Prof. Danilo Mandic for his supervision, technical support, and unflagging enthusiasm.

This work has also benefited greatly from the advice and discussions I have had over the years with many close friends. I would also like to thank my friends who have made my experiences inside and outside Imperial interesting and memorable.

I would like to reserve my biggest gratitude for my family, and thank them for their guidance and support.

This research was part of the University Defence Research Centre (UDRC) at Imperial College London, sponsored by the MoD and DSTL (Project Code: C3).

Contents

List of Figures

Statement of Originality

I declare that the content embodied in this thesis is the outcome of my research under the guidance of my thesis adviser Prof Danilo P. Mandic. Any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline. The material on this thesis has not been submitted for any degree at any other academic or professional institution.

Publications From This Thesis

Patent

A patent application regarding frequency estimation and fault identification in three phase power systems and smat-grids - UK Patent Application No. 1217737.4 (ref: 6279), October 2012.

Journal papers

[1] D.H. Dini, D.P. Mandic, S.J. Julier, "A Widely Linear Complex Unscented Kalman Filter," IEEE Signal Processing Letters, vol.18, no.11, pp.623-626, Nov. 2011

[2] D.H. Dini, C. Jahanchahi, D.P. Mandic, "Widely Linear Complex and Quaternion Valued Bearings Only Tracking," IET Signal Processing Special Issue on: Multi-Sensor Signal Processing for Defence, vol.6, no.5, pp.435-445, July 2012

[3] D.H. Dini, D.P. Mandic, "Class of Widely Linear Complex Kalman Filters," IEEE Transactions on Neural Networks and Learning Systems, vol.23, no.5, pp.775-786, May 2012

[4] D.H. Dini, D.P. Mandic, "Widely Linear Modeling for Frequency Estimation in Unbalanced Three-Phase Power Systems," IEEE Transactions on Instrumentation and Measurement, vol.62, no.2, pp.353-363, Feb. 2013

[5] D.H. Dini, P.M. Djuric, D.P. Mandic, "The Augmented Complex Particle Filter," IEEE Transactions on Signal Processing, vol.61, no.17, pp.4341-4346, Sept. 2013

Conference papers

[6] D.H. Dini, D.P. Mandic, "Analysis of the Widely Linear Complex Kalman Filter," Proceedings of Sensor Signal Processing for Defence (SSPD), Sep. 2010

[7] D.H. Dini, D.P. Mandic, "Widely Linear Complex Extended Kalman Filters," Proceedings of Sensor Signal Processing for Defence (SSPD), Sep. 2011

[8] D.H. Dini, D.P. Mandic, "Widely Linear State Space Models for Frequency Estimation in Unbalanced Three-Phase Systems," Proceedings of IEEE 7th Sensor Array and Multichannel Signal Processing Workshop (SAM), 2012, pp.9-12, 17-20 June 2012

[9] D.H. Dini, D.P. Mandic, "An Enhanced Bearing Estimation Technique for DIFAR Sonobuoy Underwater Target Tracking," Proceedings of Sensor Signal Processing for Defence (SSPD), Sep. 2012

[10] D.H. Dini, D.P. Mandic, "Cooperative Adaptive Estimation of Distributed Noncircular Complex Signals," Proceedings of ASILOMAR, pp.1518-1522, 4-7 Nov. 2012

[11] D.H. Dini, D.P. Mandic, "Exploiting Sparsity in Widely Linear Estimation," Proceedings of the 10th International Symposium on Wireless Communication Systems (ISWCS), vol., pp.1-5, Aug. 2013

Chapter 1

Introduction

WE live in an information age. The recent advances in sensor and computing technologies have not only vastly increased the availability of data, but also the capabilities to robustly deal with large data sets. Moreover, as technology continues to evolve, the need to make sense and inferences from ever more complex data takes ever greater precedence, and next generation techniques and solutions for achieving enhanced performance become paramount.

Another facet of modern technology is the reliance on multiple sensors, whereby measurements from a number of sensors, often with overlapping information, need to be simultaneously processed to enhance performance and increase system capabilities. However, in distributed systems consisting of sensor spread over a certain geographical region, such centralised methodologies requires large communication overheads, and distributed estimation frameworks, which relies on cooperation between neighbouring sensors, are often preferred to reduce the computational and communication overheads.

Typically, real-world data are corrupted by noise and interferences, exhibit latencies and coupling, and are nonstationary; and hence do not directly yield to analysis and information extraction. This necessitates mathematical data processing techniques that simultaneously facilitate the extraction of useful information, and the suppression of redundant interferences.

Signal processing is such a discipline, offering a mathematical framework for in-

formation retrieval from noisy corrupted data, and forms the backbone of many modern technologies, including communications and bio-medicine. Signal processing encompasses a large array of techniques ranging from image and audio processing to data compression and weather forecasting, however, as the need for more advanced data processing technologies increases, so does the requirement for the development of next-generation signal processing solutions capable of meeting the challenges of efficient, low-cost, fast and accurate data processing frameworks.

This thesis is on the subject of signal processing, and more specifically adaptive filters. Unlike other signal processing techniques, adaptive filters operate and optimise their performance in real-time with the arrival of new information, which has made them ubiquitous in many time-constrained technologies, such as satellite navigation, wireless communication, electrical smart-grids and brain-computer-interfaces. The aim of this research is focused on the development a novel theoretical framework and enhanced practical solutions for adaptive processing of complex valued signals, that is, signals with real and imaginary components.

Complex numbers are not artificially constructed concepts, but occur naturally when solving real valued problems. Although complex numbers came to prominence in the 16th century, they were not fully mainstream in the science community until the 19th century, when their geometrical interpretation was described and their usefulness in dealing with trigonometric identities identified.

Complex signals arise in a numerous real-world practical applications, as well as in transform domains such as Fourier and wavelet, where real valued data becomes complex after processing. The complex domain has distinct advantages including providing a convenient representation of bivariate data, and a natural way for preserving the characteristics of signals and the transformations they undergo, such as phase and magnitude distortions in communication systems.

Complex signal processing is the enabling technology behind applications, such as mobile communication and magnetic source imaging, however, standard, widely used, solutions inherently assume *proper* signal distributions. This is generally inadequate, given

that real-world signals almost invariably *improper*, and treating all signals as proper leads to algorithms that are unable to fully utilised the available information.

The work presented here is based on recent developments in the statistics of complex variables, called augmented complex statistics, and is used in conjunction with widely linear adaptive signal processing, to enable optimal processing for the generality of complex signals, both proper and improper. Augmented complex statistics allows for full utilisation of the available second order statistics of complex signals. Further, the effects of improper signals on the behavior of various algorithms, where second order propriety are normally assumed, is also examined. A number of widely linear (augmented) complex algorithms are proposed and analysed here, and their behavior are illustrated in a number of application employing real-world and synthetic data.

These include the gradient descent based least mean square together with the adaptive state space based Kalman and particle filtering techniques, which allow the modelling and estimation of nonstationary systems, and proposes solutions to enhance the performance of these techniques for improper data sources and widely linear system models.

This thesis is organised as follows. Each technical chapter and appendix starts with an introduction clearly detailing the original contribution of the author to the work contained within. Chapter 2 deals the background theory regarding complex signals, while Chapter 3 is the first technical chapter, and presents complex valued Kalman filters which are second order optimal for the generality of complex data. Chapter 4 concerns the application of the proposed Kalman filters for frequency estimation in unbalanced three phase power systems, whereas Chapter 5 extends the work in Chapter 3 to the case of distributed state space estimation in the presence of correlated measurement noises. In Chapter 6 some convergence issues of the gradient descent based augmented complex LMS (ACLMS) algorithm are addressed. Lastly, the chapter in Appendix A, explores the benefits of utilising density functions which cater for improper distributions within the framework of complex valued particle filters, while, the chapter in Appendix B proposes a new solution to the DIFAR sonobuoy bearing estimation problem for underwater acoustic sources based on the algorithms introduced in Chapter 3.

Novel Contributions

The main contributions of this thesis are presented in the following Chapters, however, concise summaries along with the relevant publications are presented below.

1. We introduce a class of widely linear complex Kalman filters, namely the augmented complex Kalman filter (ACKF), augmented complex extended Kalman filter (ACEKF) and augmented complex unscented Kalman filter (ACUKF), suited to the generality of complex signals, and analyse their performances under proper and improper signals. For rigour, a theoretical bound for the performance advantage of widely linear Kalman filters over their strictly linear conventional complex Kalman filters (CCKFs) is provided. The analysis also addresses the duality with bivariate real valued Kalman filters, together with several issues of implementation, and the Cramer-Rao lower bound (CRLB) for the widely linear Kalman filters is established. Our mean square analysis shows that the performance of CCKF is unaffected by the impropriety of the state and observation signals, however, the mean square characteristics of the complex extended Kalman filter (CEKF) and complex unscented Kalman filter (CUKF) are a functions of the impropriety of the state noise impropriety [1].

2. We revisit real-time frequency estimation in three phase power systems from a state space point of view, in order to provide a unified framework for frequency tracking in both balanced and unbalanced system conditions. We achieve this by using widely linear complex valued Kalman filters which are faster converging and more robust to noise and harmonic artifacts than the existing methods . It is shown that the Clarke's transformed three phase voltage is circular for balanced systems and noncircular for unbalanced ones, making the proposed widely linear estimation perfectly suited to both identify the fault and to provide accurate estimation in unbalanced conditions, critical issues where standard models typically fail. Our analysis and simulations show that the proposed approaches outperform the recently introduced widely linear stochastic gradient based frequency estimators, based on the augmented complex

least mean square (ACLMS) [2].

3. We introduce cooperative sequential state space estimation in the domain of augmented complex statistics, whereby nodes in a network collaborate locally to estimate improper complex signals. For rigour, a distributed augmented (widely linear) complex Kalman filter (D-ACKF) suited to the generality of complex signals is introduced, allowing for unified treatment of both proper (rotation invariant) and improper (rotation dependent) signal distributions. Our analysis and simulations show that unlike existing distributed Kalman filter solutions, the D-ACKF caters for both the improper data and the cross-correlations between the observation noises at neighbouring nodes, encountered when nodes are exposed to common noise (e.g. jamming noise), thus providing enhanced performance in real-world scenarios [3].

4. The distribution of complex random signals is typically improper, and conventional strictly linear models are only second order optimum for signals with proper distributions, while widely linear models are optimum for both proper and improper signals. Widely-linear models, however, are over-parameterised when the underlying system is strictly-linear, requiring twice the number of parameters to be estimated compared to strictly-linear models. This effects widely linear adaptive algorithms, such as the augmented complex least mean square (ACLMS) and augmented complex recursive least squares (ACRLS), and leads to slow convergence. We here address the problem of the over-parameterisation of the ACLMS through the use of regularised cost error functions. The conjugate weight regularised ACLMS (R-ACLMS) algorithm is presented and shown to converge faster than the ACLMS, while offering similar steady-state performance for strictly linear systems [4].

5. Current complex valued particle filters (PFs) have assumed (implicitly or explicitly) circular signal distributions, which for noncircular signals leads to suboptimal performance. We employ augmented complex statistics, and propose the augmented complex PF (ACPF) and the augmented complex Gaussian PF (ACGPF) for the sequential estimation of complex states in both circular and noncircular noise, and show through simulations the advantages for the proposed solutions [5].

6. We address the DIFAR sonobuoy bearing estimation problem for underwater acoustic sources. The standard arctangent based approach utilises the orthogonality between the observation noises for the different channels to form the bearing estimates, and ignores the correlation structure of the actual source signal. We propose a new state space technique, which exploits the correlations structure in the source signal to achieve enhanced performance, particularly in low signal-to-noise (SNR) conditions, compared to the standard arctangent estimator [6].

Chapter 2

Background

This chapter presents a background on complex numbers and a few topics on complex valued signal processing. This chapter contains a summary of the introductory chapters in [7], together with some works from [8] [9] [10] [11]. For more a complete overview of this subject, see the above references and the other works cited within this Chapter.

2.1 History of Complex Numbers

The concept of a "new number" has often arose because of a need to solve a practical problem. For instance to solve for the diagonal of a unit length square ($\sqrt{1^2 + 1^2} = \sqrt{2}$), irrational numbers needed to be introduced, whereas calculating the circumference of a circle required the use of the irrational π. Likewise complex numbers came from the necessity to solve equations involving the square root of negative numbers such as $x^2 = -4$.

Complex numbers arose to prominence in the 16th century when the Italian mathematicians Niccolo Fontana Tartaglia and Gerolamo Cardano sought to find closed form solutions to the roots of cubic and quartic polynomials. This led to expressions involving the square roots of negative numbers. They realized that even when only searching for real solutions, the manipulation of square roots of negative numbers was often required. For instance, Tartaglia's cubic formula

$$x^3 - x = 0 \tag{2.1}$$

has the following solution

$$\frac{1}{\sqrt{3}}\left((\sqrt{-1})^{\frac{1}{3}} + \frac{1}{(\sqrt{-1})^{\frac{1}{3}}}\right) \tag{2.2}$$

and when the three cube roots of -1 are substituted into this expression the three real roots, 0, 1 and -1 are found. Rafael Bombelli was the first to explicitly address these seemingly paradoxical solutions of cubic equations and developed the rules for manipulating complex numbers. In solving for the roots of

$$x^3 - 15x - 4 = 0 \tag{2.3}$$

he was able to show that

$$\left(2 + \sqrt{-1}\right) + \left(2 - \sqrt{-1}\right) = 4 \tag{2.4}$$

whereby, it was necessary to perform calculations in the field of complex numbers $\mathbb{C}$ in order to compute the real roots.

Complex numbers gained notoriety in the 18th century, as it was noted that computations involving trigonometric expressions could be simplified by utilising complex expressions. Abraham de Moivre, for example, used formal manipulation of complex expressions to show that identities relating trigonometric functions of an integer multiple of an angle could be re-expressed as powers of trigonometric functions of that same angle using the formula which bears his name, that is

$$(\cos\theta + j\sin\theta)^n = \cos n\theta + j\sin n\theta. \tag{2.5}$$

In 1748 Leonhard Euler went further and proposed the well-known Euler formula:

$$\cos\theta + j\sin\theta = e^{j\theta} \tag{2.6}$$

which reduces trigonometric identities to their simple exponential equivalents.

Complex numbers, however, did not become part of the mainstream until their

geometrical interpretation was described by Caspar Wessel in 1799. Carl Friedrich Gauss rediscovered these interpretations several years later and popularised it, and as a consequence the theory of complex numbers received a notable expansion. Although, the ideas behind the geometric representation of complex numbers had appeared as early as 1685, in Wallis's De Algebra Tractatus.

2.2 Motivations for Complex Valued Signal Processing

Complex valued signals and algorithms have proven to be useful in a wide range of theoretical and practical applications. The complex domain is the natural home for the representation and processing of numerous commonly encountered data, however, the usefulness of complex valued signals is generally application dependent. Next some applications, motivations and benefits behind complex valued systems and signals are discussed.

2.2.1 Examples of Complex Valued Signals

Fourier Analysis. The Fourier series decomposes periodic functions or signals into the sum of simple oscillating functions, namely complex exponentials. Fourier series were introduced by Joseph Fourier $(1768 - 1830)$ for the purpose of solving the heat equation in a metal plate. The real valued function $f[x]$ with a finite number of discontinuities and extrema has a Fourier series representation given by

$$f[t] = \sum_{n=-\infty}^{+\infty} c_n e^{j\omega_n t} \tag{2.7}$$

where the Fourier coefficients $\{c_n\}$ are computed as

$$c_n = \frac{1}{T} \int_{t_1}^{t_2} f[t] e^{-j\omega_n t} dt \tag{2.8}$$

and $T = t_2 - t_1$ is the period of the function $f[t]$.

The Fourier series along with the Fourier transform are perhaps the most widely used form of complex representation of real valued data. The original concept of Fourier

analysis has been extended over time to apply to more abstract and general situations, and the general field is often known as harmonic analysis. The applications of Fourier analysis are many and vary from filter bank design to modern cell phones or radio scanners.

Phasors. In mathematics and signal processing, a phase-vector ("phasor") is a very useful technique for conceptualising sinusoidally oscillating quantities. A phasor can be seen as a rotating vector. For instance, Euler's formula indicates that a sinusoidal signal $x[t] = |x|\, cos[\omega t + \phi]$ can be represented as

$$x[t] = \Re\{|x|\, e^{j\phi} \cdot e^{j\omega t}\} \tag{2.9}$$

where the operator $\Re\{\cdot\}$ denotes the real part of a complex number. The phasor can refer to either $|x|\, e^{j\phi} \cdot e^{j\omega t}$ or just the complex constant $|x|\, e^{j\phi}$. In the latter case, it is understood to be a shorthand notation, denoting the amplitude and phase of the underlying sinusoid function.

Phasors are used for the analysis of systems involving oscillating signals, such as three phase alternating current (AC) power systems, where three phasors, of equal magnitude and phases at 0, 120 and 240 degrees, are used to represent the three oscillating voltages. Phasor representations of the polyphase AC circuit variables allows for balanced systems to be simplified and unbalanced systems to be dealt with as algebraic combinations of symmetrical systems. This approach greatly simplifies the work required in calculating voltage drops, power flows, and short-circuit currents.

Analytic signals. Signals with no negative-frequency components are known as analytic. The analytic representation of a real valued function or signal facilitates mathematical manipulations, and offers a convenient way to obtain phase and instantaneous frequency information. The idea behind analytic signals is to remove the redundant frequency spectrum, that is due to the symmetry of the Fourier transform (spectrum) of real-valued function, the negative frequency components can be discarded without loss of information.

For a real valued signal $x[t]$ with a Fourier transform $X[f]$, the function $X_a[f]$

defined as

$$\begin{aligned} X_a[f] &= \begin{cases} 2X[f], & \text{if } f > 0 \\ X[f], & \text{if } f = 0 \\ 0, & \text{if } f < 0 \end{cases} \\ &= X[f] \cdot 2u[f] \end{aligned} \tag{2.10}$$

only contains the non-negative frequency components of $X[f]$, where $u[f]$ is the Heaviside step function. The inverse function exists due to symmetry of the spectrum of $X[f]$, that is

$$X[f] = \begin{cases} \frac{1}{2}X_a[f], & \text{if } f > 0 \\ X_a[f], & \text{if } f = 0 \\ \frac{1}{2}X_a^*[|f|], & \text{if } f < 0 \end{cases} \tag{2.11}$$

where $(\cdot)^*$ is the complex conjugate operator. The complex valued time domain representation of $X_a[f]$ is the analytic version of $x[t]$, and is given by

$$\begin{aligned} x_a[t] &= \mathcal{F}^{-1}\{X_a[f]\} \\ &= \mathcal{F}^{-1}\{X[f]\} * \mathcal{F}^{-1}\{2u[f]\} \\ &= x[t] + j\big(x[t] * \frac{1}{\pi t}\big) \end{aligned} \tag{2.12}$$

where the symbol $*$ denotes the convolution operator and $x[t] * \frac{1}{\pi t}$ is the Hilbert transform of $x[t]$.

Analytic representations of real signals are commonly utilised in signal processing and communication systems, whereby complex envelopes facilitates modulation and demodulation techniques together with the analysis of signals properties.

Native complex signals. Some signals can be seen as naturally complex, where an in-phase and a quadrature component is the natural representation which enables the full relationship between two components to be taken into account. Examples include radar

and directional processes as well as many communication signals such as binary phase shift keying (BPSK), quadrature phase shift keying (QPSK) and quadrature amplitude modulation (QAM). The MRI (magnetic resonance imaging) signal is also naturally complex because two orthogonal detectors are used to capture the images.

The complex domain can also be used to capture the magnitude and phase relationship between two real-valued signals. For example, wind signals have magnitude (wind intensity) and phase (wind direction), and have a natural complex representation.

2.2.2 Other Benefits of Complex Valued Processing

In addition to the examples above, signal processing in the complex domain has several distinct features and advantages; some of which are discussed below.

More powerful statistics. Recent developments in complex statistics have shown that statistics in $\mathbb{C}$ are not simple extensions of statistics in $\mathbb{R}$. The notions of proper and improper complex random variables, gives more degrees of freedom and hence greater potential for improved performance compared with standard modelling in $\mathbb{C}$. For example in blind source separation and extraction problems, complex signals with varying degrees of impropriety can be separated.

Simultaneous modelling and fusion of two variables. Complex domain modelling of directional processes, such as wind, not only provides a convenient representation, but also provides sequential data fusion. The magnitude and phase, which are of different natures, are fused to into a single scalar quantity.

Visualisation. Whereas real valued functions are represented by two dimensional graphs, complex functions are represented by four dimensional graphs (two axis for the real and imaginary parts of the function argument and two axis for the real and imaginary parts of the evaluated function). Hence to visualise a complex function, the two dimensional function argument is plotted against either the phase or magnitude of the evaluated function or the graph is colour coded to suggest the fourth dimension.

Compact and natural representation. The complex number $x = a+jb$ can be thought of as a single entity that satisfies all the standard rules of algebra. For algorithms such as the least mean square (LMS) or recursive least squares (RLS), where the desired signal (training signal) is a scalar, a complex version of these algorithms allows the desired signal to become bivariate because a scalar complex signal consists of real and imaginary parts. To account for a bivariate desired signal in real valued LMS and RLS algorithms, two filters need to be implemented.

An alternative domain. The complex domain offers an alternative to the real domain for formulating solutions. This is useful in numerous application, such as deriving recursive expressions to problems, which are necessary in adaptive filters. For example, consider the recursive expression for an exponential:

$$e^{j\omega k} = e^{j\omega} e^{j\omega(k-1)}$$

where k is the time index. The recursive nature of the real valued equivalent to this expression is not as intuitive, that is

$$\begin{aligned} e^{j\omega k} &= (\cos[\omega] + j\sin[\omega])(\cos[\omega(k-1)] + j\sin[\omega(k-1)]) \\ &= \cos[\omega]\cos[\omega(k-1)] - \sin[\omega]\sin[\omega(k-1)] + j\cos[\omega]\sin[\omega(k-1)] + j\sin[\omega]\cos[\omega(k-1)] \end{aligned}$$

Similarly, the recursive forms for sinusoids, for example $\cos[\omega k] = 1/2e^{j\omega k} + 1/2e^{-j\omega k}$, are more elegantly expressed as complex exponentials. Further, the complex domain can be considered a generalisation of the real domain, in that when the imaginary part vanishes, the two domains are equal.

Homomorphic filtering. Typically noise is additive, that is the observation consists of the summation of a signal and noise. However, there are often cases involving multiplicative noise, where the observation consists the product of a signal and noise. One approach for dealing with multiplicative signals is through homomorphic filtering, whereby the logarithm of the product is used to separate the signals, for example $\ln xy = \ln x + \ln y$. For real valued variables, the logarithm does not exist for $x \leq 0$ or $y \leq 0$, while the logarithm

of a nonzero complex signal $z = |z|e^{j\theta}$ is always defined as $\ln z = \ln |z| + j\theta$.

Derivative approximation. The first order Taylor series approximation of a real valued function $f[\cdot]$ with a real argument a is defined as

$$f'[a] = \frac{f[a+h] - f[a]}{h} + \mathcal{O}[h]$$

where h is the real valued argument increment. On the other hand, the Taylor series approximation for a complex valued argument is given by

$$f[a+jh] = f[a] + jhf'[a] - \frac{1}{2!}h^2 f''[a] - \frac{1}{3!}jh^3 f'''[a] + \cdots$$

where $f''[\cdot]$ and $f'''[\cdot]$ are the second and third order derivatives. Equating the imaginary parts yields

$$f'[a] = \frac{\Im\{f[a+jh]\}}{h} + \mathcal{O}[h^2]$$

The operator $\Im\{\cdot\}$ is the imaginary part of a complex number. Observe that there is no difference operation and the error is an $\mathcal{O}[h^2]$ operation when the argument is complex valued, hence the derivative is be better approximated by utilising a complex representation, given that $h << 1$ is typically chosen.

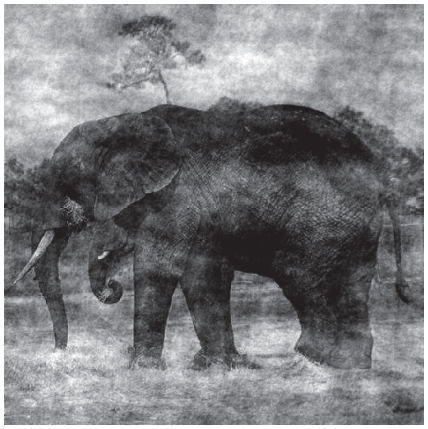
The importance of phase information. In a number of real world applications, the phase information is more important the magnitude information, e.g. image processing. Consider Figure 2.1 showing two original images along with the cases where their phase spectra are interchanged prior to taking the inverse Fourier transform. It is clear that most of the meaningful information is contained in the phase, as the appearance of the phase exchanged images are dominated by the phase information.

2.3 Complex Statistics and Widely Linear Estimation

The distribution of signals dictate the signal processing techniques and the nature of the estimators suited to dealing with them. For example, the optimal estimator, in the mean

(a) Image I1: Buffalos

(b) Image I2: Elephants



(c) I3 = $\mathcal{F}^{-1}\Big\{|\mathcal{F}(I1)|\exp(j\angle\mathcal{F}(I2))\Big\}$

(d) I4 = $\mathcal{F}^{-1}\Big\{|\mathcal{F}(I2)|\exp(j\angle\mathcal{F}(I1))\Big\}$

Figure 2.1: An illustration of the importance of phase in pictures. The Figures (a) and (b) are the original images, while figure (c) consists of the magnitude information from (a) and the phase information from (b), and vice versa for Figure (d). The visual perceptions of (c) and (d) are largely dominated by the phase information.

square error sense, for a linear Gaussian process is linear, while for nonlinear or non Gaussian processes the optimal estimator is generally untenable. Hence a thorough understanding of the distribution of complex signals, plays a fundamental role in developing the right algorithms for different problem sets. In this section, the statistical moments of complex signals are discussed with special emphasis on the second order moments.

2.3.1 Second Order Statistics of Complex Signals

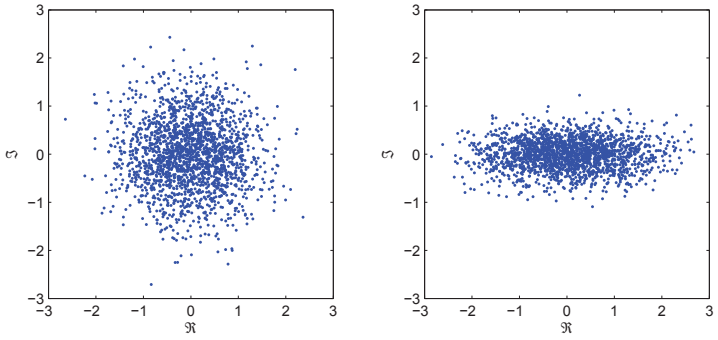
Second order statistics plays an important role in signal processing. Typically, the estimation error does not directly yield to minimisation due to non-convexity, and we seek to minimise convex functions of the estimation error. Among the possible choices, it is the mean square error (MSE) and its approximates which are often the default choice, given that for an unbiased estimator the MSE is equal to the error variance (power), while for biased estimator it is equal to sum of the error variance and the squared bias - thus MSE is essentially a second order statistical moment. From a practical view point, lower MSE corresponds to better estimation performance.

The prominence of complex signals necessitates the need for a deeper understanding of their statistics. The second order statistical properties of a zero mean[1] complex vector $\mathbf{z} = \mathbf{x} + j\mathbf{y}$ has conventionally been characterised by its covariance matrix $\mathbf{R}_\mathbf{z} = E\{\mathbf{z}\mathbf{z}^H\}$, where $(\cdot)^H$ indicates the complex-conjugate transpose operator. However, this is insufficient for a complete second-order description, and another moment function known as the pseudocovariance (also referred to as the relation function or complementary covariance) $\mathbf{P}_\mathbf{z} = E\{\mathbf{z}\mathbf{z}^T\}$, where $(\cdot)^T$ is the transpose operator, is also necessary. It is only for the special class of complex signals known as second order *proper* or*circular*, that is, those with rotation invariant probability distributions, characterised by a vanishing pseudocovariance, that their covariance function suffices to give the complete second order description. The covariance matrix captures the information regarding the total power of the signal, while the pseudocovariance captures the information about the power difference and cross-correlation between the real and imaginary parts of the signal.

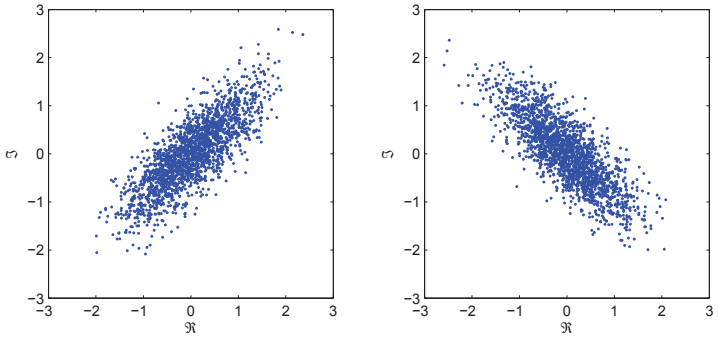
The term circularity comes from the following remarks. It is clear that the covariance of $\mathbf{z}$ and its rotated version $\bar{\mathbf{z}} = \mathbf{z}e^{j\theta}$ are equal for any real number θ. However, the pseudocovariances of $\mathbf{z}$ and $\bar{\mathbf{z}}$ are equal if and only if

$$\mathbf{P}_\mathbf{z} = E\{\mathbf{z}\mathbf{z}^T\} = \mathbf{P}_{\bar{\mathbf{z}}} = E\{\bar{\mathbf{z}}\bar{\mathbf{z}}^T\} = e^{j2\theta}E\{\mathbf{z}\mathbf{z}^T\} = e^{j2\theta}\mathbf{P}_\mathbf{z} \tag{2.13}$$

[1]The zero mean assumption simplifies the mathematical processing without loss of generality.



(a) Circular distribution (b) Uncorrelated noncircular



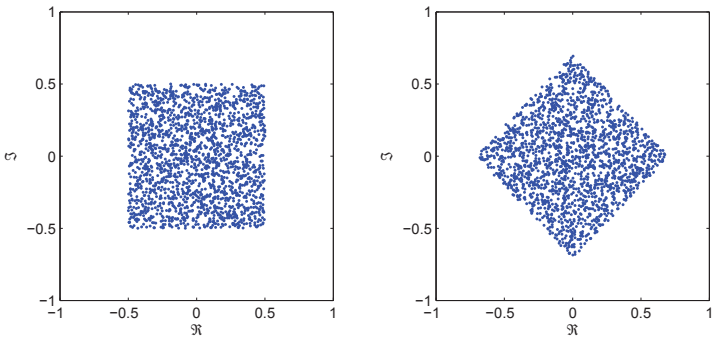
(c) Correlated noncircular (d) Correlated noncircular

Figure 2.2: A geometric view of circularity via a real-imaginary scatter plot of zero-mean complex white Gaussian distributions.

which leads to the solution $\mathbf{P_z} = \mathbf{0}$. A circular signal is a signal whose second order statistics are invariant for any phase rotation, thus the pseudocovariance vanishes for circular signals.

At this point it is worth noting that signals with zero pseudocovariances do not necessarily have circular distributions, however, signals with circular distributions always have zero pseudocovariances as shown in (2.13). Gaussian signals are a special case for whom a vanishing pseudocovariance implies circularity and vise-versa is also true.

To illustrate this point consider Figure 2.2 which shows a geometric view of the circularity of complex Gaussian distributions with identical variances with varying pseudocovariances. Figure 2.2a shows a circular signal with zero pseudocovariance, while the remaining figures have the same pseudocovariance magnitude. On the other hand, Figure 2.3 show a uniform distribution and its 45 degree rotated version, both have zero



(a) Uniform distribution (b) Rotated uniform distribution

Figure 2.3: A real-imaginary scatter plot of zero-mean complex white uniform distributions with zero pseudocovariances (both are proper).

pseudocovariances, but are noncircular.

To differentiate between these scenarios, for the remainder of this thesis, we will use the term *proper* to refer to all signals with vanishing pseudocovariances, and the term *circular* to refer to signals with rotation invariant distributions, which also implies zero pseudocovariances. Hence, the term proper is more general, and circular is as a special case of proper. Further, for real valued signals the covariance and pseudocovariance are equal, and as such real signals are always improper (noncircular).

The covariance matrix $\mathbf{R_z}$ by its definition is positive semi-definite, while the pseudocovariance matrix $\mathbf{P_z}$ is symmetric. To illuminate this, consider the decomposition of these two matrices in terms of the covariances and cross-correlations between the real and imaginary parts of the complex vector $\mathbf{z} = \mathbf{x} + j\mathbf{y}$, namely

$$\begin{aligned} \mathbf{R_z} &= E\{\mathbf{z}\mathbf{z}^H\} \\ &= \mathbf{R_x} + \mathbf{R_y} + j(\mathbf{R_{yx}} - \mathbf{R_{xy}}) \qquad (2.14) \\ \mathbf{P_z} &= E\{\mathbf{z}\mathbf{z}^T\} \\ &= \mathbf{R_x} - \mathbf{R_y} + j(\mathbf{R_{xy}} + \mathbf{R_{yx}}) \qquad (2.15) \end{aligned}$$

where $\mathbf{R_x} = E\{\mathbf{x}\mathbf{x}^T\}$, $\mathbf{R_y} = E\{\mathbf{y}\mathbf{y}^T\}$ and $\mathbf{R_{xy}} = \mathbf{R_{yx}}^T = E\{\mathbf{x}\mathbf{y}^T\}$. Based on (2.15), observe that for a complex signal to be circular, that is $\mathbf{P_z} = \mathbf{0}$, implies two strict conditions

on the real ($\mathbf{x}$) and imaginary ($\mathbf{y}$) parts of the signal, namely

1. $\mathbf{x}$ and $\mathbf{y}$ have identical covariances: $\mathbf{R_x} = \mathbf{R_y}$
2. $\mathbf{x}$ and $\mathbf{y}$ are orthogonal: $\mathbf{R_{xy}} = \mathbf{0}$

If either of these conditions is not met, the complex signal is then *improper*.

2.3.2 Complex White Noise

The concept of white noise is critical in signal processing, as it allows for modelling of uncertainties in systems. However, the framework for defining complex white noise is not a straight forward extension of real white noise. A wide sense stationary signal $z[k]$ is white if its covariance function $c_z[\tau] = E\{z[k]z^*[k-\tau]\}$ is a Dirac delta function or equivalently its power spectrum $\Gamma_z[f] = \mathcal{F}\{c_z[\tau]\}$ (the Fourier transform of c_z) is constant, that is

$$c_z[\tau] = a\delta[\tau] \quad \Leftrightarrow \quad \Gamma_z[f] = a = \text{constant} \tag{2.16}$$

The definition of whiteness also enforces constraints on the pseudocovariance function $p_z[\tau] = E\{z[k]z[k-\tau]\}$ or the spectral pseudocovariance $P_z[f] = \mathcal{F}\{p_z[\tau]\}$ (the Fourier transform of p_z), through the relationship between the covariance and pseudocovariance functions, see (2.14) and (2.15). It can be shown that given the power spectrum, the spectral pseudocovariance satisfy the following conditions [8]:

$$\begin{aligned} \Gamma_z[f] &\geq 0 \\ P_z[f] &= P_z[-f] \\ |P_z[f]|^2 &\leq \Gamma_z[f]\Gamma_z[-f] \end{aligned} \tag{2.17}$$

Hence, if a signal is second order stationary and white we have

$$\begin{aligned} c_z[k] &= c\delta[k] \\ |P_z[f]|^2 &\leq \Gamma_z[f]\Gamma_z[-f] = a^2 \end{aligned} \tag{2.18}$$

which implies that the absolute square of the spectral pseudocovariance is a constant with a value between zero and a^2. The spectral pseudocovariance function $P_z[f]$ is often assumed to be zero, which means that the white signal is proper. However, for a non-zero pseudocovariance function $P_z[f]$, we can have improper white noise. Therefore we can define two types of stationary complex white noises:

1. *Proper white noise* characterised by a constant power spectrum and a vanishing pseudocovariance function. The real and imaginary parts of the signal have equal variance and are uncorrelated, that is

$$c_z[\tau] = a\delta[\tau] \qquad \text{and} \qquad p_z[\tau] = 0 \tag{2.19}$$

   While the frequency domain equivalent is given by

$$\begin{aligned} \Gamma_z[f] &= a = \text{constant} \\ P_z[f] &= 0 \end{aligned} \tag{2.20}$$

2. *Doubly white noise* characterised by

$$c_z[\tau] = a\delta[\tau] \qquad \text{and} \qquad p_z[\tau] = b\delta[\tau]$$

   where the only condition on the pseudocovariance function is $|b| \leq a$. The power spectrum and the spectral pseudocovariance are then given by

$$\begin{aligned} \Gamma_z[f] &= a = \text{constant} \\ P_z[f] &= b = \text{constant} \end{aligned} \tag{2.21}$$

2.3.3 Widely Linear (Augmented) Complex Estimation

To introduce an optimal second order estimator for the generality of complex signals, consider the minimum mean square error (MSE) estimator of a real valued random vector $\mathbf{y}$ in terms of an observed real vector $\mathbf{x}$, that is, $\hat{\mathbf{y}} = E\{\mathbf{y}|\mathbf{x}\}$. For zero-mean, jointly

normal $\mathbf{y}$ and $\mathbf{x}$, the optimal estimator is linear, that is

$$\hat{\mathbf{y}} = \mathbf{A}\mathbf{x} \tag{2.22}$$

The aim is then to find the coefficient matrix $\mathbf{A}$ that minimises the MSE given by

$$\mathbf{\Sigma}^l = E\{[\mathbf{y} - \mathbf{A}\mathbf{x}][\mathbf{y} - \mathbf{A}\mathbf{x}]^H\} \tag{2.23}$$

Differentiating $\mathbf{\Sigma}^l$ with respect to $\mathbf{A}$, and setting the derivative to zero yields the solution

$$\mathbf{A} = \mathbf{R}_{\mathbf{yx}}\mathbf{R}_{\mathbf{x}}^{-1} \tag{2.24}$$

where $\mathbf{R}_{\mathbf{yx}} = E\{\mathbf{y}\mathbf{x}^H\}$. Standard, 'strictly linear' estimation in $\mathbb{C}$ assumes the same model but with complex valued $\mathbf{y} = \mathbf{y}_r + j\mathbf{y}_i$ and $\mathbf{x} = \mathbf{x}_r + j\mathbf{x}_i$, and the resulting solution has the same form: $\mathbf{A} = \mathbf{R}_{\mathbf{yx}}\mathbf{R}_{\mathbf{x}}^{-1}$ but is complex valued. Observe that this solution does not incorporate the pseudocovariance of the data, and is hence blind to the propriety of the signals.

Next consider the bi-variate estimation problem, whereby the aim is to estimate each of $\mathbf{y}_r$ and $\mathbf{y}_i$ based on $\mathbf{x}_r$ and $\mathbf{x}_i$, that is

$$\hat{\mathbf{y}}_r = E\{\mathbf{y}_r|\mathbf{x}_r, \mathbf{x}_i\} \tag{2.25}$$

$$\hat{\mathbf{y}}_i = E\{\mathbf{y}_i|\mathbf{x}_r, \mathbf{x}_i\} \tag{2.26}$$

Substituting in the complex representations for $\mathbf{x}_r = (\mathbf{x} + \mathbf{x}^*)/2$ and $\mathbf{x}_i = (\mathbf{x} - \mathbf{x}^*)/2\jmath$ yields

$$\hat{\mathbf{y}}_r = E\{\mathbf{y}_r|\mathbf{x}, \mathbf{x}^*\} \tag{2.27}$$

$$\hat{\mathbf{y}}_i = E\{\mathbf{y}_i|\mathbf{x}, \mathbf{x}^*\} \tag{2.28}$$

which highlights that $\mathbf{y}$ needs to be estimated in terms of both $\mathbf{x}$ and its conjugate $\mathbf{x}^*$. The optimal complex estimator can be derived formally by considering the problem of

estimating $\mathbf{y}$ as $E\{\mathbf{y}|\mathbf{x}, \mathbf{x}^*\}$. The MSE is then given by

$$\mathbf{\Sigma}^{wl} = E\{[\mathbf{y} - \mathbf{B}\mathbf{x} - \mathbf{C}\mathbf{x}^*][\mathbf{y} - \mathbf{B}\mathbf{x} - \mathbf{C}\mathbf{x}^*]^H\} \tag{2.29}$$

Next differentiating $\mathbf{\Sigma}^{wl}$ with respect to $\mathbf{B}$ and $\mathbf{C}$, and setting the derivatives to zero results in the *widely linear* complex estimator[2], that is

$$\hat{\mathbf{y}} = \mathbf{B}\mathbf{x} + \mathbf{C}\mathbf{x}^* \tag{2.30}$$

where the coefficient matrices are given as

$$\mathbf{B} = \mathbf{R}_{\mathbf{yx}}\mathbf{D} + \mathbf{P}_{\mathbf{yx}}\mathbf{E}^*$$

$$\mathbf{C} = \mathbf{R}_{\mathbf{yx}}\mathbf{E} + \mathbf{P}_{\mathbf{yx}}\mathbf{D}^*$$

with $\mathbf{D} = (\mathbf{R}_{\mathbf{x}} - \mathbf{P}_{\mathbf{x}}\mathbf{R}_{\mathbf{x}}^{*-1}\mathbf{P}_{\mathbf{x}}^*)^{-1}$ and $\mathbf{E} = -(\mathbf{R}_{\mathbf{x}} - \mathbf{P}_{\mathbf{x}}\mathbf{R}_{\mathbf{x}}^{*-1}\mathbf{P}_{\mathbf{x}}^*)^{-1}\mathbf{P}_{\mathbf{x}}\mathbf{R}_{\mathbf{x}}^{*-1}$.

The widely linear estimator (2.30) is optimal for the generality of complex signals, both proper and improper, as it caters for the covariance and pseudocovariance of the data. Observe that when $\mathbf{y}$ and $\mathbf{x}$ are jointly proper $\mathbf{P}_{\mathbf{yx}} = E\{\mathbf{y}\mathbf{x}^T\} = \mathbf{0}$, and $\mathbf{x}$ is proper $\mathbf{P}_{\mathbf{x}} = \mathbf{0}$, the widely linear linear solution degenerates to the standard strictly linear solution (2.22), that is $\mathbf{C} = \mathbf{0}$.

For convenience of representation, the widely linear model can be cast into an augmented representation[3]:

$$\hat{\mathbf{y}} = \mathbf{B}\mathbf{x} + \mathbf{C}\mathbf{x}^* = \mathbf{W}\mathbf{x}^a \tag{2.31}$$

where $\mathbf{x}^a = [\mathbf{x}^T, \mathbf{x}^H]^T$ is the augmented input vector, and $\mathbf{W} = [\mathbf{B}, \mathbf{C}]$ the optimal coefficient matrix. Further, the full second order information of the input $\mathbf{x}$ is contained in the

[2]The term "'widely linear"' indicates that the new estimator is a linear function of both $\mathbf{x}$ and $\mathbf{x}^*$, while the standard strictly linear estimator is a linear function of only $\mathbf{x}$.

[3]The term 'widely linear' model is associated with the signal generating system, whereas the term"augmented statistics" describe statistical properties of measured signals. Both the terms are used to name the resulting algorithms.

augmented covariance matrix

$$\mathbf{R}_{\mathbf{x}}^{a} = E\{\mathbf{x}^{a}\mathbf{x}^{aH}\} = \begin{bmatrix} \mathbf{R}_{\mathbf{x}} & \mathbf{P}_{\mathbf{x}} \\ \mathbf{P}_{\mathbf{x}}^{*} & \mathbf{R}_{\mathbf{x}}^{*} \end{bmatrix} \tag{2.32}$$

and as such, estimation based on $\mathbf{R}_{\mathbf{x}}^{a}$ incorporates both the covariance and pseudocovariance information. The augmented estimator then takes the form

$$\mathbf{A} = \mathbf{R}_{\mathbf{y}\mathbf{x}^{a}}[\mathbf{R}_{\mathbf{x}}^{a}]^{-1} \tag{2.33}$$

where $\mathbf{R}_{\mathbf{y}\mathbf{x}^{a}} = E\{\mathbf{y}\mathbf{x}^{a}\}$ is the cross correlation between $\mathbf{y}$ and the augmented input $\mathbf{x}^{a}$.

2.3.4 Benefit of Widely Linear Complex Estimation

The MSE performance difference between widely linear modelling over strictly linear modelling can be expressed as

$$\mathbf{\Delta} = \mathbf{\Sigma}^{wl} - \mathbf{\Sigma}^{l} \tag{2.34}$$

After some tedious algebraic manipulations and following the approach in [9], the MSE difference between the two estimators becomes [1]

$$\mathbf{\Delta} = (\mathbf{P}_{\mathbf{y}\mathbf{x}} - \mathbf{R}_{\mathbf{y}\mathbf{x}}\mathbf{R}_{\mathbf{x}}^{-1}\mathbf{P}_{\mathbf{x}})(\mathbf{R}_{\mathbf{x}}^{*} - \mathbf{P}_{\mathbf{x}}^{*}\mathbf{R}_{\mathbf{x}}^{-1}\mathbf{P}_{\mathbf{x}})^{-1}(\mathbf{P}_{\mathbf{y}\mathbf{x}} - \mathbf{R}_{\mathbf{y}\mathbf{x}}\mathbf{R}_{\mathbf{x}}^{-1}\mathbf{P}_{\mathbf{x}})^{H} \tag{2.35}$$

The matrix $\mathbf{\Delta}$ is positive semi-definite owing to the positive definiteness of the matrix $(\mathbf{R}_{\mathbf{x}}^{*} - \mathbf{P}_{\mathbf{x}}^{*}\mathbf{R}_{\mathbf{x}}^{-1}\mathbf{P}_{\mathbf{x}})$. The two estimators have the same MSE for $\mathbf{\Delta} = \mathbf{0}$, which is only the case when $(\mathbf{P}_{\mathbf{y}\mathbf{x}} - \mathbf{R}_{\mathbf{y}\mathbf{x}}\mathbf{R}_{\mathbf{x}}^{-1}\mathbf{P}_{\mathbf{x}}) = \mathbf{0}$, in other words, when $\mathbf{y}$ and $\mathbf{x}$ are jointly proper $\mathbf{P}_{\mathbf{y}\mathbf{x}} = E\{\mathbf{y}\mathbf{x}^{T}\} = \mathbf{0}$, and input $\mathbf{x}$ is proper $\mathbf{P}_{\mathbf{x}} = \mathbf{0}$. Hence, the widely linear estimator always performs the same or better than the strictly linear estimator, in a MSE sense.

2.4 Degree of Complex Impropriety

Complex random variables are classified as second order proper or improper. However, improper signals can take a wide range or degrees of impropriety. For example, when viewed geometrically, the circularity (propriety) of Gaussian signals can vary extensively, that is from a circular distribution to the extremely noncircular case where all the data are distributed on a line, for example when the real and imaginary parts are fully correlated the distribution is on a line. There is hence a need to quantify and measure the degrees of impropriety of complex variables. Further, the widely linear model has a larger computational overhead than the strictly linear model, and in some applications the degree of impropriety can determine whether the performance benefits of the widely linear model can offsets the extra computational overhead.

The circularity (or propriety) of a complex signal is preserved by linear transformations, which include scaling and rotation, but not by widely linear transformations. For instance, if the complex vector $\mathbf{z} = [z_1, ..., z_n]^T$ is proper then the linear transformation $\mathbf{G} \cdot \mathbf{z}$, where $\mathbf{G}$ is a nonsingular matrix, is also proper, while if $\mathbf{z}$ is improper, then its linear transform is also improper. However, under widely linear transformations, such as $\mathbf{G} \cdot \mathbf{z} + \mathbf{H} \cdot \mathbf{z}^*$ where both $\mathbf{G}$ and $\mathbf{H}$ are nonsingular, propriety is no longer preserved.

Hence, any measure of impropriety is also required to be invariant under linear transformations, but not widely linear transformations. This means that the measure must be a function of a complete set of invariants for the covariance $\mathbf{R_z}$ and pseudocovariance $\mathbf{P_z}$ under linear transformation. This has been shown to be given by the set of *canonical correlations* between $\mathbf{z}$ and its conjugate $\mathbf{z}^*$. The canonical correlations are also known as the *circularity coefficients* and play a key role in independent component analysis of complex signals [11].

The first step to computing the canonical correlations, involves whitening the signal by taking the square root decomposition of the covariance matrix, that is

$$\mathbf{R_z} = \mathbf{R_z}^{1/2}(\mathbf{R_z}^{1/2})^T = \mathbf{R_z}^{1/2}\mathbf{R_z}^{T/2} \tag{2.36}$$

where the invertible matrix $\mathbf{R_z}^{1/2}$ is defined as the square root of the covariance matrix $\mathbf{R_z}$. Then the vector $\bar{\mathbf{z}} = \mathbf{R_z}^{-1/2}\mathbf{z} = [\bar{z}_1, ..., \bar{z}_n]^T$ has covariance matrix

$$\mathbf{R}_{\bar{\mathbf{z}}} = E\{\bar{\mathbf{z}}\bar{\mathbf{z}}^H\} = \mathbf{R_z}^{-1/2}\mathbf{R_z}\mathbf{R_z}^{-T/2} = \mathbf{I} \tag{2.37}$$

and is therefore a unit variance white random vector. The canonical correlations are determined from the pseudocovariance of the whitened signal $\bar{\mathbf{z}}$, also known as the *coherence matrix*, that is

$$\mathbf{P}_{\bar{\mathbf{z}}} = E\{\bar{\mathbf{z}}\bar{\mathbf{z}}^T\} = \mathbf{R_z}^{-1/2}\mathbf{P_z}\mathbf{R_z}^{-T/2} = \mathbf{M} \tag{2.38}$$

The coherence matrix $\mathbf{M}$, being a pseudocovariance matrix, is complex symmetric, $\mathbf{M} = \mathbf{M}^T$, and can be decomposed using *Takagi factorisation* to yield

$$\mathbf{M} \;=\; \mathbf{F}\mathbf{K}\mathbf{F}^T$$

where $\mathbf{F}$ is a unitary matrix, that is $\mathbf{F}^H = \mathbf{F}^{-1}$, and the diagonal matrix $\mathbf{K} = \mathrm{diag}(k_1, k_2, \ldots, k_n)$ contains the canonical correlations $1 \geq k_1 \geq k_2 \geq \cdots \geq k_n \geq 0$ on its diagonal.

Further, the linear transformation $\acute{\mathbf{z}} = \mathbf{F}^H\bar{\mathbf{z}} = \mathbf{F}^H\mathbf{R_x}^{-1/2}\mathbf{z} = [\acute{z}_1, ..., \acute{z}_n]^T$, which simultaneously diagonalises both the covariance and pseudocovariance, is said to be given in *canonical coordinates*. The canonical coordinates have the special property of being white with unit variance, together with a diagonal pseudocovariance matrix of canonical correlations, that is

$$\mathbf{R}_{\acute{\mathbf{z}}} = E\{\acute{\mathbf{z}}\acute{\mathbf{z}}^H\} = \mathbf{I} \tag{2.39}$$

$$\mathbf{P}_{\acute{\mathbf{z}}} = E\{\acute{\mathbf{z}}\acute{\mathbf{z}}^T\} = \mathbf{K} \tag{2.40}$$

Vectors, such as $\acute{\mathbf{z}}$, with unit diagonal covariances, that are generally improper, are often referred to as *strongly uncorrelated*. The strongly uncorrelating transform is a useful framework for the analysis of complex signal processing algorithms.

There are a number of plausible functions for measuring impropriety based on the canonical correlations, however, one measure stands out because it relates the entropy of a noncircular Gaussian random variable with its circular counterpart. The entropy of an improper Gaussian random vector with augmented covariance matrix $\mathbf{R}_{\mathbf{z}^a}$ and the corresponding proper Gaussian random vector with covariance matrix $\mathbf{R}_{\mathbf{z}}$ can be expressed as [12]

$$\begin{aligned} H_{\text{improper}} &= \frac{1}{2}\ln[(\pi e)^{2n}\det\mathbf{R}_{\mathbf{z}^a}] \\ &= \underbrace{\ln[(\pi e)^n \det\mathbf{R}_{\mathbf{z}}]}_{H_{\text{proper}}} + \frac{1}{2}\ln\prod_{i=1}^{n}(1-k_i^2) \end{aligned} \tag{2.41}$$

where det is the matrix determinant operator. This illustrates the classical result that proper Gaussian random vectors maximise entropy $H_{\text{improper}} \leq H_{\text{proper}}$, while the entropy difference between the proper and improper signals is a function of the canonical correlations.

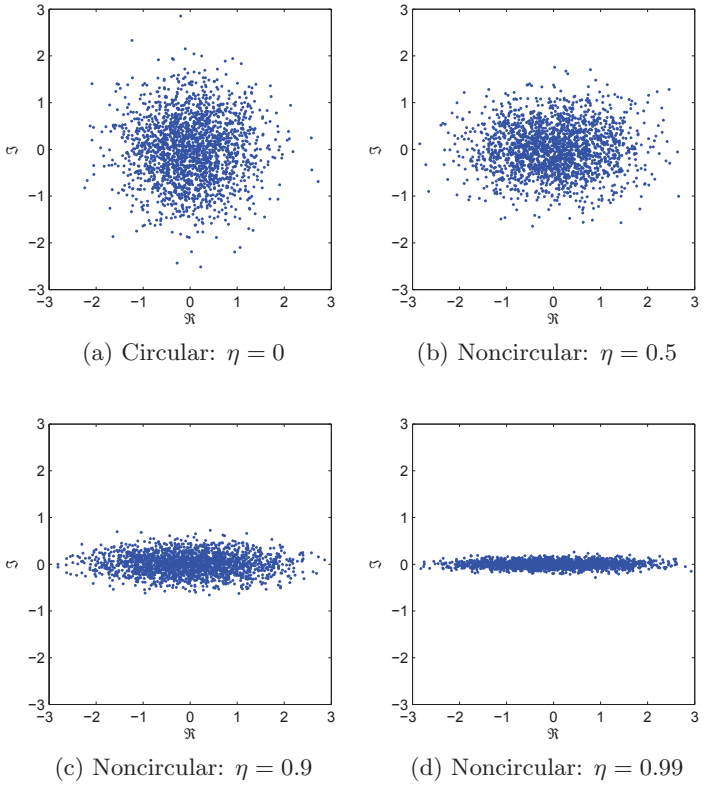
The circularity measure defined as [13]

$$d = 1 - \prod_{i=1}^{n}(1-k_i^2) = 1 - \det\mathbf{R}_{\mathbf{z}^a}[\det\mathbf{R}_{\mathbf{z}}]^{-2} \tag{2.42}$$

lends itself as the natural choice for measuring the degree of impropriety of complex random vectors. Moreover, this function is a compelling measure for several reasons:

- d is bounded as $0 \leq d \leq 1$, whereby for $d = 0$ the signal is circular and for $d = 1$ the signal is maximally improper.
- It connects the entropy of the proper and improper cases.
- It is a measure of the linear dependence between $\mathbf{z}$ and $\mathbf{z}^*$ and as such, can be used to design a generalized likelihood ratio test for impropriety.
- Tight bounds on the measure d can be obtained without the need to explicitly compute the canonical correlations, which can save on computational processing.

For a scalar random variable z with covariance $r_z = E\{|z|^2\}$ and pseudocovariance



(a) Circular: $\eta = 0$ (b) Noncircular: $\eta = 0.5$

(c) Noncircular: $\eta = 0.9$ (d) Noncircular: $\eta = 0.99$

Figure 2.4: A geometric view of circularity via a real-imaginary scatter plot of white complex Gaussian processes at different degrees of noncircularity (η), with orthogonal real and imaginary parts.

$p_z = E\{z^2\}$, the measure d simplifies to:

$$d = \frac{|p_z|^2}{r_z^2} \tag{2.43}$$

which is essentially the square of the ratio between the pseudocovariance and covariance. This in turn motivates the ratio between the pseudocovariance and covariance, known as the circularity quotient, to be taken as an impropriety measure, that is [10]

$$\varrho = \frac{p_z}{r_z} = \eta e^{j\theta} \tag{2.44}$$

Where $\eta = |\varrho_z| = \sqrt{d}$, $0 \leq \eta \leq 1$, is the *circularity coefficient*[4] and $\theta = arg(\varrho_z)$ is the *circularity angle*. The advantage of using the circularity quotient, over the circularity

[4]For Gaussian signals, we will refer to η as the 'degree of noncircularity'

measure d, is that it preserves the phase information contained within the pseudocovariance, and is simpler to compute. Figure 2.4 illustrates the distributions of white complex Gaussian noise with different degrees of noncircularity. In the following chapters, the circularity coefficient will be used to test the performance of algorithms at different degrees of impropriety.

Chapter 3

Complex Valued Kalman Filters

In this first technical Chapter, we propose second order optimal complex valued Kalman filters. The Kalman filter is an adaptive state space estimation technique with a wide range of applications, including space navigation and military technology development. Complex valued Kalman filters are commonly encountered in real world scenarios such as frequency estimation [14], training of neural networks [15][7], and wireless localization [16], however, conventionally they have explicitly or implicitly been designed for assumed proper (circular) signals, that is, signals that are uncorrelated with their complex conjugates, though, real world signals are typically improper. In an earlier work [15], the widely linear complex Kalman filter was proposed in the context of neural network training, but its performance characteristics and operation in general augmented state space models were not elaborated.

In this Chapter, we propose a class of widely linear complex Kalman filters and illuminate their performances under general improper state and observation signals. The effect of signal impropriety on the mean square behavior of the conventional complex Kalman filter (CCKF), complex extended Kalman filter (CEKF) and complex unscented Kalman filter (CUKF) are analysed, and the Cramer-Rao lower bound (CRLB) for the widely linear Kalman filters is established. While, computational complexity issues are addressed by exploiting the isomorphism between the bivariate-real and complex domains. Simulations on both benchmark and real world noncircular data support the analysis.

3.1 The Augmented Complex Kalman Filter (ACKF)

Consider the standard linear state space [17]

$$\mathbf{x}_n = \mathbf{F}_{n-1}\mathbf{x}_{n-1} + \mathbf{w}_n \tag{3.1a}$$

$$\mathbf{y}_n = \mathbf{H}_n\mathbf{x}_n + \mathbf{v}_n \tag{3.1b}$$

where $\mathbf{x}_n \in \mathbb{C}^L$ and $\mathbf{y}_n \in \mathbb{C}^K$ are the state to be estimated and the noisy observation (measurement) vectors at time instant n, respectively, while $\mathbf{F}_n$ and $\mathbf{H}_n$ are the state transition and observation matrices, whereas $\mathbf{w}_n \in \mathbb{C}^L$ and $\mathbf{v}_n \in \mathbb{C}^K$ denote the uncorrelated state and measurement noises, respectively, and are assumed to be doubly white[1] and zero-mean, hence, their covariance matrices are defined as

$$E\begin{bmatrix}\mathbf{w}_n\\ \mathbf{v}_n\end{bmatrix}\begin{bmatrix}\mathbf{w}_k\\ \mathbf{v}_k\end{bmatrix}^H = \begin{bmatrix}\mathbf{Q}_n & \mathbf{0}\\ \mathbf{0} & \mathbf{R}_n\end{bmatrix}\delta_{nk} \tag{3.2}$$

where δ_{nk} is the Kronecker delta function, and their pseudocovariance matrices as

$$E\begin{bmatrix}\mathbf{w}_n\\ \mathbf{v}_n\end{bmatrix}\begin{bmatrix}\mathbf{w}_k\\ \mathbf{v}_k\end{bmatrix}^T = \begin{bmatrix}\mathbf{P}_n & \mathbf{0}\\ \mathbf{0} & \mathbf{U}_n\end{bmatrix}\delta_{nk} \tag{3.3}$$

To cater for widely linear system models with improper state and observation noises, it is necessary to introduce a widely linear state space model. Based the widely linear model in (2.30), the widely linear version of the standard state space model in (3.1) is defined as[2]

$$\mathbf{x}_n = \mathbf{F}_{n-1}\mathbf{x}_{n-1} + \mathbf{A}_{n-1}\mathbf{x}_{n-1}^* + \mathbf{w}_n \tag{3.4a}$$

$$\mathbf{y}_n = \mathbf{H}_n\mathbf{x}_n + \mathbf{B}_n\mathbf{x}_n^* + \mathbf{v}_n \tag{3.4b}$$

[1] The term "doubly-white" refers to complex signals for which the covariance and pseudocovariance functions are Dirac delta functions.

[2] Observe that the noise models can also be widely linear, in which case:
$\mathbf{w}_n = \mathbf{C}_n\acute{\mathbf{w}}_n + \mathbf{D}_n\acute{\mathbf{w}}_n^*$ and $\mathbf{v}_n = \mathbf{E}_n\acute{\mathbf{v}}_n + \mathbf{G}_n\acute{\mathbf{v}}_n^*$, where $\mathbf{C}, \mathbf{D}, \mathbf{E}, \mathbf{G}$ are coefficient matrices and $\acute{\mathbf{w}}_n$ and $\acute{\mathbf{v}}_n$ are proper or improper noise models.

and can be expressed in a compact form using "augmented" complex vectors, such that

$$\mathbf{x}_n^a = \mathbf{F}_{n-1}^a \mathbf{x}_{n-1}^a + \mathbf{w}_n^a \tag{3.5a}$$

$$\mathbf{y}_n^a = \mathbf{H}_n^a \mathbf{x}_n^a + \mathbf{v}_n^a \tag{3.5b}$$

where $\mathbf{x}_n^a = [\mathbf{x}_n^T, \mathbf{x}_n^H]^T$ and $\mathbf{y}_n^a = [\mathbf{y}_n^T, \mathbf{y}_n^H]^T$, while,

$$\mathbf{F}_n^a = \begin{bmatrix} \mathbf{F}_n & \mathbf{A}_n \\ \mathbf{A}_n^* & \mathbf{F}_n^* \end{bmatrix} \quad \text{and} \quad \mathbf{H}_n^a = \begin{bmatrix} \mathbf{H}_n & \mathbf{B}_n \\ \mathbf{B}_n^* & \mathbf{H}_n^* \end{bmatrix}$$

The matrices $\mathbf{A}_n$ and $\mathbf{B}_n$ in (3.4) determine whether the state and observation equations are strictly or widely linear, whereby for $\mathbf{A} = \mathbf{0}$ and $\mathbf{B} = \mathbf{0}$, the state space equations assume strictly linear forms. However, even for strictly linear system models, the augmented state space representation offers the advantage of catering for improper state and observation noises, hence providing a complete second order statistical characterisation, unlike the standard strictly linear state space model. This point was not considered in earlier widely linear Kalman filters. The augmented covariance matrices of the noise vectors $\mathbf{w}_n^a = [\mathbf{x}_n^T, \mathbf{w}_n^H]^T$ and $\mathbf{v}_n^a = [\mathbf{v}_n^T, \mathbf{v}_n^H]^T$, defined as

$$\mathbf{Q}_n^a = E\{\mathbf{w}_n^a \mathbf{w}_n^{aH}\} = \begin{bmatrix} \mathbf{Q}_n & \mathbf{P}_n \\ \mathbf{P}_n^* & \mathbf{Q}_n^* \end{bmatrix} \tag{3.6}$$

$$\mathbf{R}_n^a = E\{\mathbf{v}_n^a \mathbf{v}_n^{aH}\} = \begin{bmatrix} \mathbf{R}_n & \mathbf{U}_n \\ \mathbf{U}_n^* & \mathbf{R}_n^* \end{bmatrix} \tag{3.7}$$

fully incorporate the noise covariance and pseudocovariance information. Once the augmented state space model and vectors are defined, the expressions for the augmented complex Kalman filter (ACKF) can be derived in same manner as the conventional complex Kalman filter (CCKF) [18] but employing the augmented vectors and augmented covariance matrices. Similar to the real valued Kalman filter, the ACKF is a minimum mean square error (MSE) estimator $\widehat{\mathbf{x}}_{n|n}^a = E[\mathbf{x}_n^a | \mathbf{y}_0^a, \mathbf{y}_1^a, ..., \mathbf{y}_n^a]$ of $\mathbf{x}_n^a$ based on the observations $\{\mathbf{y}_0^a, \mathbf{y}_1^a, ..., \mathbf{y}_n^a\}$, when the state and observation noises are Gaussian. The ACKF is summarised in Algorithm 1.

Algorithm 1: The augmented complex Kalman filter (ACKF)

Initialise with:

$$\begin{aligned}\widehat{\mathbf{x}}^a_{0|0} &= E\{\mathbf{x}^a_0\}\\ \mathbf{M}^a_{0|0} &= E\{(\mathbf{x}^a_0 - E\{\mathbf{x}^a_0\})(\mathbf{x}^a_0 - E\{\mathbf{x}^a_0\})^H\}\end{aligned}$$

State Prediction:

$$\widehat{\mathbf{x}}^a_{n|n-1} = \mathbf{F}^a_{n-1}\widehat{\mathbf{x}}^a_{n-1|n-1} \tag{3.8}$$

Prediction MSE:

$$\mathbf{M}^a_{n|n-1} = \mathbf{F}^a_{n-1}\mathbf{M}^a_{n-1|n-1}\mathbf{F}^{aH}_{n-1} + \mathbf{Q}^a_n \tag{3.9}$$

Kalman Gain:

$$\mathbf{G}^a_n = \mathbf{M}^a_{n|n-1}\mathbf{H}^{aH}_n\left(\mathbf{H}^a_n\mathbf{M}^a_{n|n-1}\mathbf{H}^{aH}_n + \mathbf{R}^a_n\right)^{-1} \tag{3.10}$$

State Update:

$$\widehat{\mathbf{x}}^a_{n|n} = \widehat{\mathbf{x}}^a_{n|n-1} + \mathbf{G}^a_n\left(\mathbf{y}^a_n - \mathbf{H}^a_n\widehat{\mathbf{x}}^a_{n|n-1}\right) \tag{3.11}$$

MSE Matrix Update:

$$\mathbf{M}^a_{n|n} = (\mathbf{I} - \mathbf{G}^a_n\mathbf{H}^a_n)\mathbf{M}^a_{n|n-1} \tag{3.12}$$

3.1.1 CCKF and ACKF Duality Analysis

For strictly linear state space models with proper state and observation noises, the CCKF and ACKF become equivalent and yield the identical state estimate at every time instant. These conditions can be summarised as follows:

$$\mathbf{Q}^a_n = \begin{bmatrix}\mathbf{Q}_n & \mathbf{0}\\ \mathbf{0} & \mathbf{Q}^*_n\end{bmatrix}, \mathbf{R}^a_n = \begin{bmatrix}\mathbf{R}_n & \mathbf{0}\\ \mathbf{0} & \mathbf{R}^*_n\end{bmatrix}, \mathbf{F}^a_n = \begin{bmatrix}\mathbf{F}_n & \mathbf{0}\\ \mathbf{0} & \mathbf{F}^*_n\end{bmatrix} \quad \text{and} \quad \mathbf{H}^a_n = \begin{bmatrix}\mathbf{H}_n & \mathbf{0}\\ \mathbf{0} & \mathbf{H}^*_n\end{bmatrix} \tag{3.13}$$

The duality between CCKF and ACKF for proper data and under the same initialisation can be illustrated as follows. Consider the prediction MSE matrix in the ACKF, which

can be expressed as a Riccati recursion, that is

$$\begin{aligned}\mathbf{M}^a_{n+1|n} =&\mathbf{F}^a_{n-1}\mathbf{M}^a_{n|n-1}(\mathbf{F}^a_{n-1})^H - \mathbf{F}^a_{n-1}\mathbf{M}^a_{n|n-1}(\mathbf{H}^a_n)^H[\mathbf{H}^a_n\mathbf{M}^a_{n|n-1}(\mathbf{H}^a_n)^H + \mathbf{R}^a_n]^{-1} \\ &\times (\mathbf{H}^a_n)\mathbf{M}^a_{n|n-1}(\mathbf{F}^a_{n-1})^H + \mathbf{Q}^a_n \end{aligned} \tag{3.14}$$

Note that the computations for $\mathbf{M}^a_{n|n-1}$ and $\mathbf{M}^a_{n|n}$ do not involve the observation vector, and as such can be calculated without taking any measurement into account. By substituting equation (3.10) into (3.12) and using the matrix inversion lemma, the augmented MSE matrix $\mathbf{M}^a_{n|n}$ can be expressed as

$$\begin{aligned}\mathbf{M}^a_{n|n} &= \mathbf{M}^a_{n|n-1} - \mathbf{M}^a_{n|n-1}(\mathbf{H}^a_n)^H\left[\mathbf{H}^a_n\mathbf{M}^a_{n|n-1}(\mathbf{H}^a_n)^H + \mathbf{R}^a_n\right]^{-1}\mathbf{H}^a_n\mathbf{M}^a_{n|n-1} \\ &= \left[(\mathbf{M}^a_{n|n-1})^{-1} + (\mathbf{H}^a_n)^H(\mathbf{R}^a_n)^{-1}\mathbf{H}^a_n\right]^{-1} \end{aligned} \tag{3.15}$$

Next, substituting (3.15) into (3.10), allows for the Kalman gain to be written as

$$\mathbf{G}^a_n = [(\mathbf{M}^a_{n|n-1})^{-1} + (\mathbf{H}^a_n)^H(\mathbf{R}^a_n)^{-1}\mathbf{H}^a_n]^{-1}(\mathbf{H}^a_n)^H(\mathbf{R}^a_n)^{-1} = \mathbf{M}^a_{n|n}(\mathbf{H}^a_n)^H(\mathbf{R}^a_n)^{-1} \tag{3.16}$$

Assuming that both CCKF and ACKF have the same initialisation, that is

$$\begin{aligned}\widehat{\mathbf{x}}^a_{0|0} &= \left[\widehat{\mathbf{x}}^T_{0|0}, \widehat{\mathbf{x}}^H_{0|0}\right]^T \\ \mathbf{M}^a_{0|0} &= \begin{bmatrix}\mathbf{M}_{0|0} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}^*_{0|0}\end{bmatrix}\end{aligned}$$

then $\widehat{\mathbf{x}}_{0|0}$ and $\mathbf{M}_{0|0}$ are, respectively, the the initial state and MSE for the strictly linear CCKF. Substituting the expressions in (3.13) into (3.16) yields a block diagonal Kalman gain, that is

$$\mathbf{G}^a_n = \begin{bmatrix}\mathbf{G}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^*_n\end{bmatrix} \tag{3.17}$$

where $\mathbf{G}_n = \mathbf{M}_{n|n}(\mathbf{H}_n)^H(\mathbf{R}_n)^{-1}$ is the Kalman gain for the CCKF at time instant n. Observe that CCKF and ACKF have the same Kalman gain (though ACKF has a

block-conjugate structure); and by substituting (3.17) into (3.11) it follows that the two filters yield identical state estimates.

Remark #1: For strictly linear state and observation models with circular state and observation noises (that is, the conditions illustrated in (3.13)), the ACKF and CCKF have identical performances at each time instant.

3.1.2 Mean square error performance analysis

We next illuminate the mean square error (MSE) performances of the CCKF and ACKF in to provide insight into the behavior of Kalman filters for proper and improper signals. The Kalman filter is a minimum MSE estimator for dynamic systems, and can alternatively be expressed in a nonrecursive form, similar to the standard Wiener (normal) solution, which is also a minimum MSE estimator but for stationary systems. From the state space model described by (3.1a) and (3.1b), the Kalman filter estimate $\widehat{\mathbf{x}}_{n|n}$ of the state $\mathbf{x}_n$ is based on the all observations up to time n, and can be written as a linear combination of the observation sequence, $\mathbf{z}_n = \left[\mathbf{y}_1^T, \mathbf{y}_2^T, ..., \mathbf{y}_n^T\right]^T$, that is

$$\widehat{\mathbf{x}}_{n|n} = E\{\mathbf{x}_0\} + \mathbf{W}_n \mathbf{z}_n \tag{3.18}$$

where $\mathbf{W}_n$ is the coefficient matrix, which is the solution to the normal equation, that is

$$\mathbf{W}_n = \mathbf{R}_{\mathbf{xz},n,n} \mathbf{R}_{\mathbf{z},n}^{-1} \tag{3.19}$$

with $\mathbf{R}_{\mathbf{xz},n,n} = E\big\{(\mathbf{x}_n - E\{\mathbf{x}_n\})(\mathbf{z}_n - E\{\mathbf{z}_n\})^H\big\}$ and $\mathbf{R}_{\mathbf{z},n} = E\big\{(\mathbf{z}_n - E\{\mathbf{z}_n\})(\mathbf{z}_n - E\{\mathbf{z}_n\})^H\big\}$. The MSE matrix is then given by

$$\begin{aligned} \mathbf{M}_{n|n} &= E\{(\mathbf{x}_n - \widehat{\mathbf{x}}_{n|n})(\mathbf{x}_n - \widehat{\mathbf{x}}_{n|n})^H\} \\ &= \mathbf{R}_{\mathbf{x},n} - \mathbf{R}_{\mathbf{xz},n,n} \mathbf{R}_{\mathbf{z},n}^{-1} \mathbf{R}_{\mathbf{xz},n,n}^H \end{aligned} \tag{3.20}$$

The Kalman filter output at each time instant is now summarised by state estimate (mean) in (3.18) and MSE (covariance) matrix in (3.20), although the computational complexity of

these expressions increase with time (as the dimensions of the observation vector increases), nonetheless, they are general and suffice for the analysis of the MSE performances of Kalman filters.

Next consider the state equation (3.1a) in its non-recursive form

$$\mathbf{x}_n = \mathbf{F}_{n:0}\mathbf{x}_0 + \sum_{i=1}^{n} \mathbf{F}_{n:i}\mathbf{w}_i \tag{3.21}$$

where $\mathbf{x}_0$ is the initial state[3], and the state transition matrix has the properties

$$\mathbf{F}_{n:i} = \mathbf{F}_n\mathbf{F}_{n-1}\cdots\mathbf{F}_i, \quad \mathbf{F}_{i:i} = \mathbf{I} \quad \text{and} \quad \mathbf{F}_0 = \mathbf{I}$$

This allows us to express the state covariance matrix as

$$\mathbf{R}_{\mathbf{x},n} = \mathbf{F}_{n:0}\mathbf{R}_{\mathbf{x},0}\mathbf{F}_{n:0}^H + \sum_{i=1}^{n} \mathbf{F}_{n:i}\mathbf{Q}_i\mathbf{F}_{n:i}^H \tag{3.22}$$

and the observation covariance as

$$\begin{aligned} \mathbf{R}_{\mathbf{y},n,m} &= E\{\mathbf{y}_n\mathbf{y}_m^H\} \\ &= \begin{cases} \mathbf{H}_n\mathbf{R}_{\mathbf{x},n}\mathbf{H}_n^H + \mathbf{R}_n & \text{if } n = m \\ \mathbf{H}_n\mathbf{R}_{\mathbf{x},n}\mathbf{H}_m^H & \text{if } n < m \\ \mathbf{H}_n\mathbf{R}_{\mathbf{x},m}\mathbf{H}_m^H & \text{if } n > m \end{cases} \end{aligned} \tag{3.23}$$

The cross-correlation between the state and observation can now be expressed as

$$\begin{aligned} \mathbf{R}_{\mathbf{xy},n,m} &= E\{\mathbf{x}_n\mathbf{y}_m^H\} \qquad n \geq m \\ &= E\{\mathbf{x}_n(\mathbf{H}_m\mathbf{x}_m + \mathbf{v}_m)^H\} \\ &= \mathbf{R}_{\mathbf{x},m}\mathbf{H}_m^H \end{aligned} \tag{3.24}$$

while the cross-correlation between the state $\mathbf{x}_n$ and the observation sequence $\mathbf{z}_n$ is given

[3]Without loss of generality, we assume $E\{\mathbf{x}_0\} = \mathbf{0}$.

by

$$\mathbf{R}_{\mathbf{xz},n,n} = \begin{bmatrix} \mathbf{R}_{\mathbf{xy},n,1} & \mathbf{R}_{\mathbf{xy},n,2} & \cdots & \mathbf{R}_{\mathbf{xy},n,n} \end{bmatrix} \tag{3.25}$$

and the covariance of the observation sequence by

$$\mathbf{R}_{\mathbf{z},n} = E\{\mathbf{z}_n \mathbf{z}_n^H\} = \begin{bmatrix} \mathbf{R}_{\mathbf{y},1} & \mathbf{R}_{\mathbf{y},1,2} & \cdots & \mathbf{R}_{\mathbf{y},1,n} \\ \mathbf{R}_{\mathbf{y},2,1} & \mathbf{R}_{\mathbf{y},2} & \cdots & \mathbf{R}_{\mathbf{y},2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{\mathbf{y},n,1} & \mathbf{R}_{\mathbf{y},n,2} & \cdots & \mathbf{R}_{\mathbf{y},n} \end{bmatrix} \tag{3.26}$$

Based on (3.18) and (3.21), observe that the estimate $\widehat{\mathbf{x}}_{n|n}$ is unbiased, that is

$$E\{\mathbf{e}_{n|n}\} \;=\; E\{(\mathbf{x}_n - \widehat{\mathbf{x}}_{n|n})\} = \mathbf{0} \tag{3.27}$$

and, as such, the mean characteristics of the CCKF is not effected by noncircular state and observation signals. While the expression in (3.20) shows that the mean square characteristics of the CCKF is dependent on the covariance matrices of the state and observation noises but not on their pseudocovariances.

Remark #2: The propriety of the state and observation data do not affect the performance of the linear conventional complex Kalman filter.

For the augmented complex Kalman filter (ACKF), the state estimate and MSE matrix are given by

$$\begin{aligned} \widehat{\mathbf{x}}^a_{n|n} &= E\{\mathbf{x}^a_0\} + \mathbf{W}^a_n \mathbf{z}^a_n = E\{\mathbf{x}^a_0\} + \mathbf{R}^a_{\mathbf{xz},n,n} (\mathbf{R}^a_{\mathbf{z},n})^{-1} \mathbf{z}^a_n \\ \mathbf{M}^a_{n|n} &= E\{(\mathbf{x}^a_n - \widehat{\mathbf{x}}^a_{n|n})(\mathbf{x}^a_n - \widehat{\mathbf{x}}^a_{n|n})^H\} \\ &= \mathbf{R}^a_{\mathbf{x},n} - \mathbf{R}^a_{\mathbf{xz},n,n} (\mathbf{R}^a_{\mathbf{z},n})^{-1} \mathbf{R}^{aH}_{\mathbf{xz},n,n} \end{aligned} \tag{3.28}$$

where $\mathbf{W}^a_n = \mathbf{R}^a_{\mathbf{xz},n,n} (\mathbf{R}^a_{\mathbf{z},n})^{-1}$, $\mathbf{z}^a_n = [\mathbf{z}^{aT}_n, \mathbf{z}^{aH}_n]^T$ and $\mathbf{R}^a_{\mathbf{xz},n,n} = E\{\mathbf{x}^a_n \mathbf{z}^{aH}_n\}$, while the matrix form of the augmented MSE matrix $\mathbf{M}^a_{n|n}$ has the block conjugate structure defined

as

$$\begin{bmatrix} \mathbf{M}_{wl,n|n} & \mathbf{P}_{wl,n|n} \\ \mathbf{P}^*_{wl,n|n} & \mathbf{M}^*_{wl,n|n} \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{\mathbf{x},n} & \mathbf{P}_{\mathbf{x},n} \\ \mathbf{P}^*_{\mathbf{x},n} & \mathbf{R}^*_{\mathbf{x},n} \end{bmatrix} - \begin{bmatrix} \mathbf{R}_{\mathbf{xz},n,n} & \mathbf{P}_{\mathbf{xz},n,n} \\ \mathbf{P}^*_{\mathbf{xz},n,n} & \mathbf{R}^*_{\mathbf{xz},n,n} \end{bmatrix} \times \begin{bmatrix} \mathbf{R}_{\mathbf{z},n} & \mathbf{P}_{\mathbf{z},n} \\ \mathbf{P}^*_{\mathbf{z},n} & \mathbf{R}^*_{\mathbf{z},n} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{R}_{\mathbf{xz},n,n} & \mathbf{P}_{\mathbf{xz},n,n} \\ \mathbf{P}^*_{\mathbf{xz},n,n} & \mathbf{R}^*_{\mathbf{xz},n,n} \end{bmatrix}^H$$

The terms $\mathbf{P}_{\mathbf{x},n}$ and $\mathbf{P}_{\mathbf{z},n}$ are the pseudocovariances of the state and observation sequence respectively, while $\mathbf{P}_{\mathbf{xz},n,n} = E\{\mathbf{x}_n\mathbf{z}_n^T\}$ is the pseudo-correlation between the state and observation sequence. It will be recognised that the matrix $\mathbf{M}_{wl,n|n}$ is just the widely linear error matrix for the state $\mathbf{x}_n$, that is

$$\mathbf{M}_{wl,n|n} = E\{(\mathbf{x}_n - \mathbf{x}_{wl,n|n})(\mathbf{x}_n - \mathbf{x}_{wl,n|n})^H\} \tag{3.29}$$

where $\mathbf{x}_{wl,n|n}$ is the widely linear estimate of $\mathbf{x}_n$. It is this that is to be compared with the strictly linear error matrix $\mathbf{M}_{n|n}$. Also note that the matrix $\mathbf{P}_{wl,n|n}$ is defined as

$$\mathbf{P}_{wl,n|n} = E\{(\mathbf{x}_n - \mathbf{x}_{wl,n|n})(\mathbf{x}_n - \mathbf{x}_{wl,n|n})^T\} \tag{3.30}$$

The inverse of the augmented covariance matrix $(\mathbf{R}^a_{\mathbf{z},n})^{-1}$ can be expressed as

$$\begin{bmatrix} \mathbf{R}_{\mathbf{z},n} & \mathbf{P}_{\mathbf{z},n} \\ \mathbf{P}^*_{\mathbf{z},n} & \mathbf{R}^*_{\mathbf{z},n} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{C} & \mathbf{D} \\ \mathbf{D}^* & \mathbf{C}^* \end{bmatrix}$$

where

$$\begin{aligned} \mathbf{C} &= (\mathbf{R}_{\mathbf{z},n} - \mathbf{P}_{\mathbf{z},n}\mathbf{R}^{*-1}_{\mathbf{z},n}\mathbf{P}^*_{\mathbf{z},n})^{-1} \\ \mathbf{D} &= -(\mathbf{R}_{\mathbf{z},n} - \mathbf{P}_{\mathbf{z},n}\mathbf{R}^{*-1}_{\mathbf{z},n}\mathbf{P}^*_{\mathbf{z},n})^{-1}\mathbf{P}_{\mathbf{z},n}\mathbf{R}^{*-1} \end{aligned}$$

and the ACKF MSE in (3.29) takes the form:

$$\mathbf{M}_{wl,n|n} = \mathbf{R}_{\mathbf{x},n} - \mathbf{R}_{\mathbf{xz},n,n}\mathbf{C}\mathbf{R}^H_{\mathbf{xz},n,n} - \mathbf{R}_{\mathbf{xz},n,n}\mathbf{D}\mathbf{P}^H_{\mathbf{xz},n,n} - \mathbf{P}_{\mathbf{xz},n,n}\mathbf{D}^*\mathbf{R}^H_{\mathbf{xz},n,n} - \mathbf{P}_{\mathbf{xz},n,n}\mathbf{C}^*\mathbf{P}^H_{\mathbf{xz},n,n}$$

After some tedious algebraic manipulations, the MSE difference between the CCKF and the ACKF is found to be [9]:

$$\begin{aligned}\Delta\mathbf{M}_n &= \mathbf{M}_{n|n} - \mathbf{M}_{wl,n|n} \\ &= (\mathbf{P}_{\mathbf{xz},n,n} - \mathbf{R}_{\mathbf{xz},n,n}\mathbf{R}_{\mathbf{z},n}^{-1}\mathbf{P}_{\mathbf{z},n})(\mathbf{R}_{\mathbf{z},n}^{*} - \mathbf{P}_{\mathbf{z},n}^{*}\mathbf{R}_{\mathbf{z},n}^{-1}\mathbf{P}_{\mathbf{z},n})^{-1}(\mathbf{P}_{\mathbf{xz},n,n} - \mathbf{R}_{\mathbf{xz},n,n}\mathbf{R}_{\mathbf{z},n}^{-1}\mathbf{P}_{\mathbf{z},n})^{H}\end{aligned} \tag{3.31}$$

Remark #3: The matrix $\Delta\mathbf{M}_n$ is always positive semidefinite owing to the positive definiteness of the matrix $(\mathbf{R}_{\mathbf{z},n}^{*} - \mathbf{P}_{\mathbf{z},n}^{*}\mathbf{R}_{\mathbf{z},n}^{-1}\mathbf{P}_{\mathbf{z},n})$, and consequently $\Delta\mathbf{M}_n = \mathbf{0}$ only when $(\mathbf{P}_{\mathbf{xz},n,n} - \mathbf{R}_{\mathbf{xz},n,n}^{-1}\mathbf{R}_{\mathbf{z},n}\mathbf{P}_{\mathbf{z},n}) = \mathbf{0}$. Therefore, the widely linear ACKF always has the same or better MSE performance than the strictly linear CCKF.

Remark #4: The CCKF and ACKF are equivalent when the observation sequence is proper ($\mathbf{P}_{\mathbf{z},n} = \mathbf{0}$) and the state and observation sequence are jointly proper ($\mathbf{P}_{\mathbf{xz},n,n} = \mathbf{0}$).

3.1.3 Duality Analysis of ACKF and real valued KF

Owing to the isomorphism between augmented complex vectors and bivariate real vectors, and the duality analysis for stochastic gradient filters [19], we next show that the ACKF algorithm has a dual bivariate real-valued Kalman filter (RKF). This duality can be exploited to reduce the computational complexity of ACKF in hardware implementations. A complex vector $\mathbf{z} = \mathbf{z}_r + j\mathbf{z}_i \in \mathbb{C}^q$ has a composite bivariate real representation in $\mathbb{R}^{2q}$ of the form

$$\mathbf{z}^a = \begin{bmatrix}\mathbf{z} \\ \mathbf{z}^*\end{bmatrix} = \underbrace{\begin{bmatrix}\mathbf{I} & j\mathbf{I} \\ \mathbf{I} & -j\mathbf{I}\end{bmatrix}}_{\equiv\mathbf{J}_\mathbf{z}} \underbrace{\begin{bmatrix}\mathbf{z_r} \\ \mathbf{z_i}\end{bmatrix}}_{=\mathbf{z}^r} \tag{3.32}$$

where $\mathbf{I}$ is the identity matrix (with appropriate dimensions), and the invertible orthogonal mapping[4] $\mathbf{J}_\mathbf{z} : \mathbb{C}^{2q} \to \mathbb{R}^{2q}$ is such that $\mathbf{J}_\mathbf{z}^{-1} = \frac{1}{2}\mathbf{J}_\mathbf{z}^{H}$ [20][21]. Based on this isomorphism,

[4]For a vector $\mathbf{z} \in \mathbb{C}^q$, the corresponding orthogonal matrix $\mathbf{J}_\mathbf{z}$ takes dimension $2q \times 2q$.

the real bivariate state space corresponding to the augmented complex state space in (3.5) is given by

$$\begin{aligned} \mathbf{x}_n^r &= \mathbf{F}_{n-1}^r \mathbf{x}_{n-1}^r + \mathbf{w}_n^r \\ \mathbf{y}_n^r &= \mathbf{H}_n^r \mathbf{x}_n^r + \mathbf{v}_n^r \end{aligned} \tag{3.33a}$$

where $\mathbf{x}_n^r = \mathbf{J}_{\mathbf{x}}^{-1}\mathbf{x}_n^a$, $\mathbf{y}_n^r = \mathbf{J}_{\mathbf{y}}^{-1}\mathbf{y}_n^a$, $\mathbf{F}_{n-1}^r = \mathbf{J}_{\mathbf{x}}^{-1}\mathbf{F}_{n-1}^a\mathbf{J}_{\mathbf{x}}$, $\mathbf{H}_n^r = \mathbf{J}_{\mathbf{y}}^{-1}\mathbf{H}_n^a\mathbf{J}_{\mathbf{x}}$, $\mathbf{w}_n^r = \mathbf{J}_{\mathbf{x}}^{-1}\mathbf{w}_n^a$ and $\mathbf{v}_n^r = \mathbf{J}_{\mathbf{y}}^{-1}\mathbf{v}_n^a$. In a similar manner, the real valued covariance matrices of $\mathbf{w}_n^r$ and $\mathbf{v}_n^r$ take the corresponding forms

$$\begin{aligned} \mathbf{Q}_n^r &= E\{\mathbf{w}_n^r \mathbf{w}_n^{rH}\} = \mathbf{J}_{\mathbf{x}}^{-1}\mathbf{Q}_n^a\mathbf{J}_{\mathbf{x}}^{-H} \\ \mathbf{R}_n^r &= E\{\mathbf{v}_n^r \mathbf{v}_n^{rH}\} = \mathbf{J}_{\mathbf{y}}^{-1}\mathbf{R}_n^a\mathbf{J}_{\mathbf{y}}^{-H} \end{aligned}$$

Next the ACKF and its dual RKF are shown to have the same performance. Assuming that ACKF is initiated at time $(n-1)$, with initial state $\widehat{\mathbf{x}}_{n-1|n-1}^a$ and MSE matrix $\mathbf{M}_{n-1|n-1}^a$, the corresponding dual RKF initialisation is given by

$$\begin{aligned} \widehat{\mathbf{x}}_{n-1|n-1}^r &= \mathbf{J}_{\mathbf{x}}^{-1}\widehat{\mathbf{x}}_{n-1|n-1}^a \\ \mathbf{M}_{n-1|n-1}^r &= \mathbf{J}_{\mathbf{x}}^{-1}\mathbf{M}_{n-1|n-1}^a\mathbf{J}_{\mathbf{x}}^{-H} \end{aligned} \tag{3.34}$$

It is now straightforward to show that the state and MSE matrix predictions of the Kalman filters are also related as

$$\begin{aligned} \widehat{\mathbf{x}}_{n|n-1}^r &= \mathbf{J}_{\mathbf{x}}^{-1}\widehat{\mathbf{x}}_{n|n-1}^a \\ \mathbf{M}_{n|n-1}^r &= \mathbf{J}_{\mathbf{x}}^{-1}\mathbf{M}_{n|n-1}^a\mathbf{J}_{\mathbf{x}}^{-H} \end{aligned} \tag{3.35}$$

and that the Kalman gains are related as

$$\begin{aligned}
\mathbf{G}_n^a &= \mathbf{M}_{n|n-1}^a \mathbf{H}_n^{aH} [\mathbf{H}_n^a \mathbf{M}_{n|n-1}^a \mathbf{H}_n^{aH} + \mathbf{R}_n^a]^{-1} \\
&= \mathbf{J}_{\mathbf{x}} \mathbf{M}_{n|n-1}^r \mathbf{J}_{\mathbf{x}}^H \mathbf{J}_{\mathbf{y}}^{-H} \mathbf{H}_n^{rH} \mathbf{J}_{\mathbf{x}}^H [\mathbf{J}_{\mathbf{y}} \mathbf{H}_n^r \mathbf{J}_{\mathbf{x}}^{-1} \mathbf{J}_{\mathbf{x}} \mathbf{M}_{n|n-1}^r \mathbf{J}_{\mathbf{x}}^H \mathbf{J}_{\mathbf{x}}^{-H} \mathbf{H}_n^{rH} \mathbf{J}_{\mathbf{y}}^H + \mathbf{J}_{\mathbf{y}} \mathbf{R}_n^r \mathbf{J}_{\mathbf{y}}^H]^{-1} \\
&= \mathbf{J}_{\mathbf{x}} \mathbf{M}_{n|n-1}^r \mathbf{H}_n^{rH} [\mathbf{H}_n^r \mathbf{M}_{n|n-1}^r \mathbf{H}_n^{rH} + \mathbf{R}_n^r]^{-1} \mathbf{J}_{\mathbf{y}}^{-1} \\
&= \mathbf{J}_{\mathbf{x}} \mathbf{G}_n^r \mathbf{J}_{\mathbf{y}}^{-1} \qquad (3.36)
\end{aligned}$$

Consequently, for the state estimates $\widehat{\mathbf{x}}_{n|n}^a$ and $\widehat{\mathbf{x}}_{n|n}^r$ we have

$$\begin{aligned}
\widehat{\mathbf{x}}_{n|n}^r &= \widehat{\mathbf{x}}_{n|n-1}^r + \mathbf{G}_n^r (\mathbf{y}_n^r - \mathbf{H}_n^r \widehat{\mathbf{x}}_{n|n-1}^r) \\
&= \mathbf{J}_{\mathbf{x}}^{-1} \widehat{\mathbf{x}}_{n|n-1}^a + \mathbf{J}_{\mathbf{x}}^{-1} \mathbf{G}_n^a \mathbf{J}_{\mathbf{y}} (\mathbf{y}_n^r - \mathbf{H}_n^r \mathbf{J}_{\mathbf{x}}^{-1} \widehat{\mathbf{x}}_{n|n-1}^a) \\
&= \mathbf{J}_{\mathbf{x}}^{-1} \widehat{\mathbf{x}}_{n|n}^a \qquad (3.37)
\end{aligned}$$

while, the MSE matrices are related as

$$\mathbf{M}_{n|n}^r = \mathbf{J}_{\mathbf{x}}^{-1} \mathbf{M}_{n|n}^a \mathbf{J}_{\mathbf{x}}^{-H} \qquad (3.38)$$

Observe that based on the expression in (3.37), the state estimates $\widehat{\mathbf{x}}_{n|n}^a$ and $\widehat{\mathbf{x}}_{n|n}^r$ are equivalent and are related by an invertible linear mapping. To show that ACKF and its dual real valued bivariate Kalman filter achieve the same mean square error (MSE), recall that the MSE for the real valued Kalman filter is given by

$$\epsilon_n^r = \mathrm{tr}\{\mathbf{M}_{n|n}^r\} \qquad (3.39)$$

where the symbol $\mathrm{tr}\{\cdot\}$ denotes the matrix trace operator. Similarly, the mean square error corresponding to the augmented MSE matrix $\mathbf{M}_{n|n}^a$ is given by the trace of (3.38), that is

$$\begin{aligned}
\mathrm{tr}\{\mathbf{M}_{n|n}^a\} &= \mathrm{tr}\{\mathbf{J}_{\mathbf{x}} \mathbf{M}_{n|n}^r \mathbf{J}_{\mathbf{x}}^H\} \\
&= \mathrm{tr}\{\mathbf{M}_{n|n}^r \mathbf{J}_{\mathbf{x}}^H \mathbf{J}_{\mathbf{x}}\} \\
&= 2 \cdot \mathrm{tr}\{\mathbf{M}_{n|n}^r\} \qquad (3.40)
\end{aligned}$$

where the expression $\mathbf{J}_{\mathbf{x}}^{H} = 2\mathbf{J}_{\mathbf{x}}^{-1}$ was utilised. At first, this result is misleading as it suggests that ACKF achieves twice the error of its dual real valued KF. However, this is because the error term is counted twice by the trace of $\mathbf{M}_{n|n}^{a}$, owing to the block diagonal structure of the augmented MSE covariance matrix, and hence needs to be halved to compute the true augmented MSE, that is

$$\epsilon_n^a = \frac{1}{2}\text{tr}\{\mathbf{M}_{n|n}^{a}\} = \epsilon_n^r$$

Remark #5: The ACKF and the its dual bivariate RKF are equivalent forms of the same state space model. They achieve the identical state estimates and MSEs at every time instant, regardless of the propriety of the processed signals.

By utilising the bivariate RKF, the computational complexity of ACKF is reduced, whereby the number of additions and multiplications required are approximately halved and quartered, respectively.

3.1.4 Posterior Cramer-Rao bound (PCRB)

For time invariant statistical models, the Cramer-Rao bound (CRB) provides a theoretical performance bound for all unbiased estimators, by establishing the lowest attainable mean square error (MSE). In time varying systems, such as the state space models where the state is driven by random noise, it is the Posterior Cramer-Rao bound (PCRB) that provides a lower bound on the MSE performance of a class of estimators [22].

For an unbiased estimator $\hat{\theta}[\mathbf{y}]$ of an r-dimensional random variable θ, obeying a condition of asymptotic unbiassedness, in the sense that

$$\lim_{||\theta||\to\infty} ||\theta - \hat{\theta}[\mathbf{y}]||\mathcal{P}_{\theta,\mathbf{y}}[\mathbf{\Theta}, \mathbf{Y}] = \mathbf{0} \tag{3.41}$$

for all values $\mathbf{y}$, where $\mathcal{P}_{\theta,\mathbf{y}}[\theta, \mathbf{Y}]$ is the joint probability density of θ and $\mathbf{y}$, the PCRB on

the estimation has the form [22]

$$\mathbf{\Gamma} = E\{(\theta - \hat{\theta}[\mathbf{y}])(\theta - \hat{\theta}[\mathbf{y}])^H\} \geq \mathbf{\Sigma}^{-1} \tag{3.42}$$

where $\mathbf{\Sigma}$ is the $r \times r$ dimensional Fisher information matrix with elements defined as

$$\Sigma_{lk} = -E\Big\{\frac{\partial^2 \log \mathcal{P}_{\theta,\mathbf{y}}[\theta, \mathbf{Y}]}{\partial\theta_l \partial\theta_k}\Big\} \qquad l, k = 1, \ldots, r \tag{3.43}$$

The inequality in (3.42) implies that the difference $\mathbf{\Gamma} - \mathbf{\Sigma}^{-1}$ is positive semidefinite. Consider a general state space model of the form

$$\mathbf{x}_n = \mathbf{f}[\mathbf{x}_{n-1}, \mathbf{w}_{n-1}] \tag{3.44a}$$

$$\mathbf{y}_n = \mathbf{h}[\mathbf{x}_n, \mathbf{v}_n] \tag{3.44b}$$

where $\mathbf{f}$ and $\mathbf{h}$ can be linear or nonlinear, possibly time varying vector valued functions, while $\mathbf{w}_n$ and $\mathbf{v}_n$ are independent white processes (not necessarily Gaussian). For this model, it was shown in [23] that the Fisher information matrix corresponding to the state $\mathbf{x}_{n+1}$ at time instant $(n+1)$ can be written in a computationally efficient recursive form, that is

$$\mathbf{\Sigma}_{n+1} = \mathbf{D}_n^{22} - \mathbf{D}_n^{21}(\mathbf{\Sigma}_n + \mathbf{D}_n^{11})^{-1}\mathbf{D}_n^{12} \tag{3.45}$$

where

$$\begin{aligned}
\mathbf{D}_n^{11} &= E\{-\Delta_{\mathbf{x}_n}^{\mathbf{x}_n} \log \mathcal{P}[\mathbf{x}_{n+1}|\mathbf{x}_n]\} \\
\mathbf{D}_n^{12} &= E\{-\Delta_{\mathbf{x}_n}^{\mathbf{x}_{n+1}} \log \mathcal{P}[\mathbf{x}_{n+1}|\mathbf{x}_n]\} \\
\mathbf{D}_n^{21} &= E\{-\Delta_{\mathbf{x}_{n+1}}^{\mathbf{x}_n} \log \mathcal{P}[\mathbf{x}_{n+1}|\mathbf{x}_n]\} = (\mathbf{D}_n^{12})^T \\
\mathbf{D}_n^{22} &= E\{-\Delta_{\mathbf{x}_{n+1}}^{\mathbf{x}_{n+1}} \log \mathcal{P}[\mathbf{x}_{n+1}|\mathbf{x}_n]\} \\
&\quad + E\{-\Delta_{\mathbf{x}_{n+1}}^{\mathbf{x}_{n+1}} \log \mathcal{P}[\mathbf{y}_{n+1}|\mathbf{x}_{n+1}]\}
\end{aligned}$$

and $\Delta_{\mathbf{b}}^{\mathbf{a}} = \frac{\partial^2}{\partial\mathbf{a}\partial\mathbf{b}}$, while the conditional probability densities $\mathcal{P}[\mathbf{x}_{n+1}|\mathbf{x}_n]$ and $\mathcal{P}[\mathbf{y}_{n+1}|\mathbf{x}_{n+1}]$ can be computed from (3.44).

Next, consider the application of the PCRB to the linear filtering problem characterised by the state space model in (3.5), where $\mathbf{w}_n$ and $\mathbf{v}_n$ are assumed zero-mean independent complex doubly white Gaussian noises with augmented covariance matrices $\mathbf{Q}_n^a$ and $\mathbf{R}_n^a$ respectively (see (3.6) and (3.7)), that is, $\mathbf{w}_n$, similar to $\mathbf{v}_n$, has a multivariate complex normal distribution defined as [24]

$$\mathcal{P}[\mathbf{w}_n] = \frac{1}{\pi^L (\det \mathbf{Q}_n^a)^{1/2}} \exp\Big(-\frac{1}{2}(\mathbf{w}_n - E\{\mathbf{w}_n\})^H \mathbf{Q}_n^{-a} (\mathbf{w}_n - E\{\mathbf{w}_n\})\Big)$$

It is now straightforward to show that

$$\begin{aligned}
\mathbf{D}_n^{11} &= \mathbf{F}_n^{aH} (\mathbf{Q}_{n+1}^a)^{-1} \mathbf{F}_n^a \\
\mathbf{D}_n^{12} &= -\mathbf{F}_n^{aH} (\mathbf{Q}_{n+1}^a)^{-1} \\
\mathbf{D}_n^{22} &= (\mathbf{Q}_{n+1}^a)^{-1} + \mathbf{H}_{n+1}^{aH} (\mathbf{R}_{n+1}^a)^{-1} \mathbf{H}_{n+1}^a
\end{aligned}$$

The distribution of the state estimate within the ACKF framework is Gaussian with a mean $\widehat{\mathbf{x}}_{n|n}^a$ and covariance $\mathbf{M}_{n|n}^a$, and by substituting these matrices in to the expression for the Fisher information matrix, it can be shown that the information matrix is the inverse of the state covariance matrix, that is

$$\mathbf{M}_{n|n}^a = \mathbf{\Sigma}_n^{-1} \tag{3.46}$$

Remark #6: The ACKF, like its real valued dual Kalman filter, achieves the Cramer-Rao lower bound [25][23] for linear systems with Gaussian signal distributions, since it essentially estimates the state $\mathbf{x}_n^a$ as $\widehat{\mathbf{x}}_{n|n}^a = E[\mathbf{x}_n^a | \mathbf{y}_0^a, \mathbf{y}_1^a, ..., \mathbf{y}_n^a]$. However, CCKF achieves the Cramer-Rao lower bound only when the state space models are strictly linear and all signals are proper at all time instants, see (3.13), which is generally not the case.

3.2 The Augmented Complex Extended Kalman Filter

The Kalman filter in its standard form is designed for linear systems, and is unsuitable for nonlinear state space models. A number of extensions have been introduced to address this

issue, including the extended Kalman filter (EKF) and unscented Kalman filter (UKF), which will be discussed in more details below.

Within the EKF framework, nonlinear systems are approximated by linear models, and as such, the state and observation functions need not be linear but differentiable. Consider the state space model defined as

$$\mathbf{x}_n = \mathbf{f}[\mathbf{x}_{n-1}] + \mathbf{w}_n \tag{3.47a}$$

$$\mathbf{y}_n = \mathbf{h}[\mathbf{x}_n] + \mathbf{v}_n \tag{3.47b}$$

where $\mathbf{f}[\cdot]$ and $\mathbf{h}[\cdot]$ are the vector valued nonlinear state and observation functions respectively, which may be time varying, and the remaining variables are as defined above. The extended Kalman filter approximates these nonlinear functions by their first order Taylor series expansions (TSE) about the state estimates. Calculating the complex derivative of a function requires the function to be analytic (differentiable) within the rigorous conditions set by the Cauchy-Riemann equations, though in practice, the functions $\mathbf{f}[\cdot]$ and $\mathbf{h}[\cdot]$ can be analytic or nonanalytic depending on the underlying physical model. For instance, a large class of functions, such as real functions of complex variables, do not satisfy the Cauchy-Riemann conditions thus severely restricting the set of allowable functions for nonlinear process and observations models.

The so called $\mathbb{CR}$ calculus [7] [21] exploits the isomorphism between the complex domain $\mathbb{C}$ and the real domain $\mathbb{R}^2$, and makes possible the TSE of both analytic and nonanalytic functions within the same framework. This way, the first order Taylor series approximation of a function $f[\cdot]$ about $\mathbf{z}$ is given by

$$f[\mathbf{z} + \Delta\mathbf{z}] = f[\mathbf{z}] + \frac{\partial f}{\partial \mathbf{z}}\Delta\mathbf{z} + \frac{\partial f}{\partial \mathbf{z}^*}\Delta\mathbf{z}^* \tag{3.48}$$

whereby for analytic functions (in the the Cauchy-Riemann sense), the term $\frac{\partial f}{\partial \mathbf{z}^*}\Delta\mathbf{z}^*$ vanishes. The partial derivatives in (3.48) are interpreted the following sense. For a real-valued function $f[\mathbf{z}] = g[\mathbf{z}_r, \mathbf{z}_i]$ of a complex variable with real and imaginary parts $\mathbf{z}_r$ and $\mathbf{z}_i$,

the '$\mathbb{R}$-derivative' and '$\mathbb{R}^*$-derivative' of f are defined as

$$\frac{\partial f}{\partial \mathbf{z}} = \frac{1}{2}\left(\frac{g}{\partial \mathbf{z_r}} - j\frac{g}{\partial \mathbf{z_i}}\right) \tag{3.49}$$

$$\frac{\partial f}{\partial \mathbf{z}^*} = \frac{1}{2}\left(\frac{g}{\partial \mathbf{z_r}} + j\frac{g}{\partial \mathbf{z_i}}\right) \tag{3.50}$$

The corresponding deriviatives for complex functions are obtained from the derivatives for the their real and imaginary parts.

Next, applying this approach, the first order approximations of the state and observation equations, (3.47a) and (3.47b), about the state estimates $\widehat{\mathbf{x}}_{n-1|n-1}$ and $\widehat{\mathbf{x}}_{n|n-1}$, yields

$$\mathbf{x}_n \approx \mathbf{F}_{n-1}\mathbf{x}_{n-1} + \mathbf{A}_{n-1}\mathbf{x}^*_{n-1} + \mathbf{w}_n + \mathbf{r}_{n-1} \tag{3.51}$$

$$\mathbf{y}_n \approx \mathbf{H}_n\mathbf{x}_n + \mathbf{B}_n\mathbf{x}^*_n + \mathbf{v}_n + \mathbf{z}_n \tag{3.52}$$

where the vectors $\mathbf{r}_n = \mathbf{f}[\widehat{\mathbf{x}}_{n-1|n-1}] - \mathbf{F}_{n-1}\widehat{\mathbf{x}}_{n-1|n-1} - \mathbf{A}_{n-1}\widehat{\mathbf{x}}^*_{n-1|n-1}$ and $\mathbf{z}_n = \mathbf{h}[\widehat{\mathbf{x}}_{n|n-1}] - \mathbf{H}_n\widehat{\mathbf{x}}_{n|n-1} - \mathbf{B}_n\widehat{\mathbf{x}}^*_{n|n-1}$, and the matrices $\mathbf{F}_{n-1}$, $\mathbf{A}_{n-1}$, $\mathbf{H}_n$ and $\mathbf{B}_n$ are the Jacobians defined as

$$\begin{aligned}
\mathbf{F}_{n-1} &= \frac{\partial \mathbf{f}}{\partial \mathbf{x}_{n-1}}\Big|_{\mathbf{x}_{n-1}=\widehat{\mathbf{x}}_{n-1|n-1}}, \ \mathbf{A}_{n-1} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}^*_{n-1}}\Big|_{\mathbf{x}^*_{n-1}=\widehat{\mathbf{x}}^*_{n-1|n-1}}, \\
\mathbf{H}_n &= \frac{\partial \mathbf{h}}{\partial \mathbf{x}_n}\Big|_{\mathbf{x}_n=\widehat{\mathbf{x}}_{n|n-1}} \text{ and } \mathbf{B}_n = \frac{\partial \mathbf{h}}{\partial \mathbf{x}^*_n}\Big|_{\mathbf{x}^*_n=\widehat{\mathbf{x}}^*_{n|n-1}}
\end{aligned}$$

From (3.51) and (3.52), observe that when the functions $\mathbf{f}[\cdot]$ and $\mathbf{h}[\cdot]$ are nonanalytic, we have $\mathbf{A}_{n-1} \neq \mathbf{0}$ and $\mathbf{B}_n \neq \mathbf{0}$, which means that the linearised state and observation models are widely linear, and thus cannot be implemented using the standard complex extended Kalman filter (CEKF). However, the state space equations become strictly linear for analytic functions, since the derivatives with respect to the complex conjugates vanish, that is, $\mathbf{A}_{n-1} = \mathbf{0}$ and $\mathbf{B}_n = \mathbf{0}$.

To simultaneously cater for the widely linear state and observation models, as well as the full second order statistics, an 'augmented' state space representation is required.

To this end, consider the nonlinear augmented state space model given by

$$\mathbf{x}_n^a = \mathbf{f}^a[\mathbf{x}_{n-1}^a] + \mathbf{w}_n^a \tag{3.53a}$$

$$\mathbf{y}_n^a = \mathbf{h}^a[\mathbf{x}_n^a] + \mathbf{v}_n^a \tag{3.53b}$$

with $\mathbf{f}^a[\mathbf{x}_{n-1}^a] = \left[\mathbf{f}^T[\mathbf{x}_{n-1}^a], \mathbf{f}^H[\mathbf{x}_{n-1}^a]\right]^T$ and $\mathbf{h}^a[\mathbf{x}_n^a] = \left[\mathbf{h}^T[\mathbf{x}_n^a], \mathbf{h}^H[\mathbf{x}_n^a]\right]^T$. The linearised augmented state space model then becomes

$$\mathbf{x}_n^a \approx \mathbf{F}_{n-1}^a \mathbf{x}_{n-1}^a + \mathbf{w}_n^a + \mathbf{r}_{n-1}^a \tag{3.54a}$$

$$\mathbf{y}_n^a \approx \mathbf{H}_n^a \mathbf{x}_n^a + \mathbf{v}_n^a + \mathbf{z}_n^a \tag{3.54b}$$

where $\mathbf{r}_n^a = \left[\mathbf{r}_n^T, \mathbf{r}_n^H\right]^T$, $\mathbf{z}_n^a = \left[\mathbf{z}_n^T, \mathbf{z}_n^H\right]^T$, $\mathbf{F}_n^a = \begin{bmatrix} \mathbf{F}_n & \mathbf{A}_n \\ \mathbf{A}_n^* & \mathbf{F}_n^* \end{bmatrix}$ and $\mathbf{H}^a = \begin{bmatrix} \mathbf{H}_n & \mathbf{B}_n \\ \mathbf{B}_n^* & \mathbf{H}_n^* \end{bmatrix}$.

Note that $\mathbf{F}_n^a = \frac{\partial \mathbf{f}^a}{\partial \mathbf{x}_n^a}$ and $\mathbf{H}_n^a = \frac{\partial \mathbf{h}^a}{\partial \mathbf{x}_n^a}$.

Therefore, in contrast to the conventional CEKF, the ACEKF allows for widely linear the state and observation models, and naturally caters for the pseudocovariances of the state and measurement noises. The derivation of the ACEKF follows from the derivation of the CEKF, but utilises the augmented state space model to derive the recursions. The ACEKF is summarised in Algorithm 2.

The novelty of the ACEKF algorithm presented in this work is that it does not assume a specific state or observation models, that is $\mathbf{f}[\cdot]$ and $\mathbf{h}[\cdot]$, which makes it a more general form of the ACEKF presented in [15]. Moreover, by utilising the $\mathbb{CR}$ calculus, we have shown how the ACEKF can be used for more general complex state space models. Appendix B presents a new solution to the DIFAR sonobuoy bearing estimation problem for underwater acoustic sources based on the ACEKF introduced in here.

3.3 The Augmented Complex Unscented Kalman Filter

The unscented Kalman filter (UKF) [26] addresses the problems arising from the first order approximation of nonlinearities withing the EKF. It aims to approximate the sta-

Algorithm 2: The augmented complex extended Kalman filter (ACEKF)

Initialise with:

$$\begin{aligned}\widehat{\mathbf{x}}^a_{0|0} &= E\{\mathbf{x}^a_0\}\\ \mathbf{M}^a_{0|0} &= E\{(\mathbf{x}^a_0 - E\{\mathbf{x}^a_0\})(\mathbf{x}^a_0 - E\{\mathbf{x}^a_0\})^H\}\end{aligned}$$

State Prediction:

$$\widehat{\mathbf{x}}^a_{n|n-1} = \mathbf{f}^a[\widehat{\mathbf{x}}^a_{n-1|n-1}] \tag{3.55}$$

Prediction Matrix:

$$\mathbf{M}^a_{n|n-1} = \mathbf{F}^a_{n-1}\mathbf{M}^a_{n-1|n-1}\mathbf{F}^{aH}_{n-1} + \mathbf{Q}^a_n \tag{3.56}$$

Kalman Gain:

$$\mathbf{G}^a_n = \mathbf{M}^a_{n|n-1}\mathbf{H}^{aH}_n\left(\mathbf{H}^a_n\mathbf{M}^a_{n|n-1}\mathbf{H}^{aH}_n + \mathbf{R}^a_n\right)^{-1} \tag{3.57}$$

State Update:

$$\widehat{\mathbf{x}}^a_{n|n} = \widehat{\mathbf{x}}^a_{n|n-1} + \mathbf{G}^a_n\left(\mathbf{y}^a_n - \mathbf{h}^a[\widehat{\mathbf{x}}^a_{n|n-1}]\right) \tag{3.58}$$

Matrix Update:

$$\mathbf{M}^a_{n|n} = (\mathbf{I} - \mathbf{G}^a_n\mathbf{H}^a_n)\mathbf{M}^a_{n|n-1} \tag{3.59}$$

tistical posterior distribution rather than approximating the nonlinearities [27]. The UKF utilises a deterministic sampling technique to select a set of sample points (known as sigma points) around the mean. These points are then propagated through the nonlinear state space models, from which the mean and covariance of the state estimate are recovered. This results in a filter which is able to more accurately capture the underlying statistical distribution of signals.

To illustrate the difference between the complex unscented transform (UT) and the augmented complex UT, consider the mapping

$$\mathbf{y} = \mathbf{f}[\mathbf{x}] = \mathbf{f}[\bar{\mathbf{x}} + \delta\mathbf{x}] \qquad \mathbf{x} \in \mathbb{C}^{L\times 1}, \quad \mathbf{y} \in \mathbb{C}^{K\times 1} \tag{3.60}$$

where $\mathbf{f}[\cdot]$ is a holomorphic nonlinear function, $\mathbf{y} = [y_1, \ldots, y_K]^T$ is the output, $\mathbf{x} = [x_1, \ldots, x_L]^T$ is the input with mean $\bar{\mathbf{x}} = E\{\mathbf{x}\}$, covariance $\mathbf{R}_\mathbf{x} = E\{(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^H\}$

and pseudocovariance $\mathbf{P}_{\mathbf{x}} = E\{(\mathbf{x}-\bar{\mathbf{x}})(\mathbf{x}-\bar{\mathbf{x}})^T\}$, while $\delta\mathbf{x} = \mathbf{x}-\bar{\mathbf{x}}$. The Taylor series expansion (TSE) of $\mathbf{y}$ about $\bar{\mathbf{x}}$ is then given by

$$\mathbf{y} = \mathbf{f}[\bar{\mathbf{x}}] + \nabla_{\delta\mathbf{x}}\mathbf{f} + \frac{1}{2!}\nabla^2_{\delta\mathbf{x}}\mathbf{f} + \frac{1}{3!}\nabla^3_{\delta\mathbf{x}}\mathbf{f} + \cdots \tag{3.61}$$

where the ith order term in the TSE for $\mathbf{f}[\cdot]$ about $\bar{\mathbf{x}}$ is [27]

$$\frac{1}{i!}\nabla^i_{\delta\mathbf{x}}\mathbf{f} = \frac{1}{i!}\Big(\sum_{k=1}^{L}\delta x_k\frac{\partial}{\partial x_k}\Big)^i\mathbf{f}[\mathbf{x}]\Big|_{\mathbf{x}=\bar{\mathbf{x}}} \tag{3.62}$$

with δx_k being the kth component of $\delta\mathbf{x}$. The expression (3.62) is an ith order polynomial in $\delta\mathbf{x}$ whose coefficients are given by the derivatives of $\mathbf{f}[\cdot]$. The mean of $\mathbf{y}$ can now be expressed as

$$\begin{aligned}\bar{\mathbf{y}} &= E\{\mathbf{f}[\bar{\mathbf{x}}+\delta\mathbf{x}]\}\\ &= \mathbf{f}[\bar{\mathbf{x}}] + E\Big\{\nabla_{\delta\mathbf{x}}\mathbf{f} + \frac{1}{2!}\nabla^2_{\delta\mathbf{x}}\mathbf{f} + \frac{1}{3!}\nabla^3_{\delta\mathbf{x}}\mathbf{f} + \cdots\Big\}\end{aligned}$$

where the ith term is given by

$$\begin{aligned}E\Big\{\frac{1}{i!}\nabla^i_{\delta\mathbf{x}}\mathbf{f}\Big\} &= \frac{1}{i!}E\Big\{\Big(\sum_{k=1}^{L}\delta x_k\frac{\partial}{\partial x_k}\Big)^i\Big\}\mathbf{f}[\mathbf{x}]\Big|_{\mathbf{x}=\bar{\mathbf{x}}}\\ &= \frac{1}{i!}\Big(m_{1,1,\cdots,1,1}\frac{\partial^i\mathbf{f}}{\partial x_1^i} + m_{1,1,\cdots,1,2}\frac{\partial^i\mathbf{f}}{\partial x_1^{i-1}\partial x_2} + \cdots\Big)\end{aligned}$$

The symbols $m_{a_1,a_2,\ldots,a_{i-1},a_i} = E\{\delta x_{a_1}\delta x_{a_2}\cdots\delta x_{a_{i-1}}\delta x_{a_i}\}$ denote the ith order central moments of the components $\mathbf{x}$ with $a_k \in [1,2,\ldots,L]$. Observe that the ith order term in the series for $\bar{\mathbf{y}}$ is a function of the ith order central moment of $\mathbf{x}$ multiplied by the ith derivative of $\mathbf{f}[\cdot]$. Hence, if the moments of $\mathbf{x}$ can be correctly evaluated up to the ith order, the mean $\bar{\mathbf{y}}$ can also be correctly evaluated up to the ith order. Similarly, the covariance matrix $\mathbf{R}_{\mathbf{y}} = E\{(\mathbf{y}-\bar{\mathbf{y}})(\mathbf{y}-\bar{\mathbf{y}})^H\}$ can be written as

$$\begin{aligned}\mathbf{R}_{\mathbf{y}} &= \frac{\partial\mathbf{f}}{\partial\mathbf{x}}\mathbf{R}_{\mathbf{x}}\Big(\frac{\partial\mathbf{f}}{\partial\mathbf{x}}\Big)^H + E\Big\{\frac{1}{3!}\nabla_{\delta\mathbf{x}}\mathbf{f}\big(\nabla^3_{\delta\mathbf{x}}\mathbf{f}\big)^H + \frac{1}{2!\times 2!}\nabla^2_{\delta\mathbf{x}}\mathbf{f}\big(\nabla^2_{\delta\mathbf{x}}\mathbf{f}\big)^H\\ &\quad + \frac{1}{3!}\nabla^3_{\delta\mathbf{x}}\mathbf{f}\big(\nabla_{\delta\mathbf{x}}\mathbf{f}\big)^H\Big\} - E\Big\{\frac{1}{2!}\nabla^2_{\delta\mathbf{x}}\mathbf{f}\Big\}E\Big\{\frac{1}{2!}\nabla^2_{\delta\mathbf{x}}\mathbf{f}\Big\}^H + \cdots\end{aligned}$$

and is accurate upto the ith term if the first i central moment of the input $\mathbf{x}$ are known. Within the complex unscented transform framework, the moments of the $p-$dimensional random variable $\mathbf{x}$ are approximated by a set $(2L+1)$ weighted (sigma) points $\{\mathcal{W}_i, \mathcal{X}_i\}_{i=0}^{2L+1}$, chosen so that their sample mean and covariance are equal to the true mean $\bar{\mathbf{x}}$ and covariance $\mathbf{R}_{\mathbf{x}}$. The nonlinear function $\mathbf{f}[\cdot]$ is then applied to each of these points to generate transformed points, $\mathcal{Y}_i = \mathbf{f}[\mathcal{X}_i]$, with a sample mean and covariance defined as

$$\begin{aligned}\hat{\bar{\mathbf{y}}} &= \sum_{i=0}^{2L} \mathcal{W}_i^{(m)} \mathcal{Y}_i \\ \hat{\mathbf{R}}_{\mathbf{y}} &= \sum_{i=0}^{2L} \mathcal{W}_i^{(c)} \big(\mathcal{Y}_i - \bar{\mathbf{y}}\big)\big(\mathcal{Y}_i - \bar{\mathbf{y}}\big)^H\end{aligned}$$

which are correct up to the second order statistical moments, given that statistics of the sigma points where correct upto the second order. The terms $\mathcal{W}_i^{(m)}$ and $\mathcal{W}_i^{(c)}$ are scalar weights used to compute the mean and covariance, respectively. For an improper $\mathbf{y}$, the output pseudocovariance $\mathbf{P}_{\mathbf{y}} = E\{(\mathbf{y}-\bar{\mathbf{y}})(\mathbf{y}-\bar{\mathbf{y}})^T\}$ is given by

$$\begin{aligned}\mathbf{P}_{\mathbf{y}} &= \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \mathbf{P}_{\mathbf{x}} \Big(\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\Big)^T + E\Big\{ \frac{1}{3!} \nabla_{\delta \mathbf{x}} \mathbf{f} \big(\nabla_{\delta \mathbf{x}}^3 \mathbf{f}\big)^T + \frac{1}{2! \times 2!} \nabla_{\delta \mathbf{x}}^2 \mathbf{f} \big(\nabla_{\delta \mathbf{x}}^2 \mathbf{f}\big)^T \\ &\quad + \frac{1}{3!} \nabla_{\delta \mathbf{x}}^3 \mathbf{f} \big(\nabla_{\delta \mathbf{x}} \mathbf{f}\big)^T \Big\} - E\Big\{ \frac{1}{2!} \nabla_{\delta \mathbf{x}}^2 \mathbf{f} \Big\} E\Big\{ \frac{1}{2!} \nabla_{\delta \mathbf{x}}^2 \mathbf{f} \Big\}^T + \cdots\end{aligned}$$

where again the accuracy depends on the knowledge of the moments of the input $\mathbf{x}$. The conventional complex unscented transform focuses on estimating the mean and covariance of the output, and does not cater for the input pseudocovariance and consequently the output pseudocovariance, due to the approach utilised for generating the sigma points, that is

$$\begin{aligned}\mathcal{X}_0 &= \bar{\mathbf{x}} \\ \mathcal{X}_i &= \bar{\mathbf{x}} + \Big(\sqrt{(p+\lambda)\mathbf{R}_{\mathbf{x}}}\Big)_i, i = 1, \ldots, L \\ \mathcal{X}_i &= \bar{\mathbf{x}} - \Big(\sqrt{(p+\lambda)\mathbf{R}_{\mathbf{x}}}\Big)_i, i = L+1, \ldots, 2L \end{aligned} \tag{3.63}$$

where $\left(\sqrt{(L+\lambda)\mathbf{R}_{\mathbf{x}}}\right)_i$ is the ith column of the matrix square root[5] and $\lambda = \alpha^2(2L+\kappa)-2L$ is a scaling parameter, while α determines the spread of the sigma points around the mean and is usually set to a small positive value (e.g., 10^{-3}), κ is a secondary scaling parameter which is usually set to 0, and β is used to incorporate prior knowledge of the distribution (for Gaussian distributions, $\beta = 2$ is optimal). From (3.63), observe that the sigma points do not incorporate the input pseudocovariance. To address this issue, consider the 'augmented' sigma points defined as

$$\begin{aligned}
\mathcal{X}_0^a &= \bar{\mathbf{x}}^a \\
\mathcal{X}_i^a &= \bar{\mathbf{x}}^a + \left(\sqrt{(L+\lambda)\mathbf{R}_{\mathbf{x}}^a}\right)_i, i = 1, \ldots, 2L \\
\mathcal{X}_i^a &= \bar{\mathbf{x}}^a - \left(\sqrt{(L+\lambda)\mathbf{R}_{\mathbf{x}}^a}\right)_i, i = 2L+1, \ldots, 4L
\end{aligned}$$

which are functions of the input mean, covariance and pseudocovariance, due to the use of the augmented covariance matrix, and can be used to propagate the second order statistics of improper inputs. The weights associated with the augmented sigma points are then given by

$$\mathcal{W}_0^{(m)} = \frac{\lambda}{2L+\lambda} \tag{3.64}$$

$$\mathcal{W}_0^{(c)} = \frac{\lambda}{2L+\lambda} + (1-\alpha^2+\beta)$$

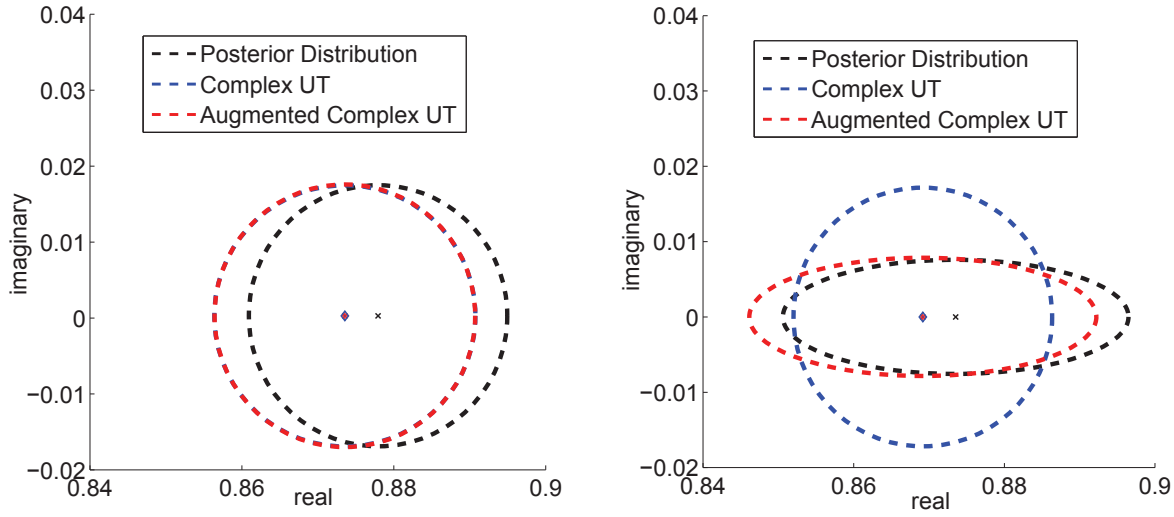
$$\mathcal{W}_i^{(m)} = \mathcal{W}_i^{(c)} = \frac{\lambda}{2(2L+\lambda)}, \qquad i = 1, \ldots, 4L \tag{3.65}$$

where the output mean and covariance are computed using the m and c super-scripted weights respectively.

To illustrate the benefits of the augmented complex UT over the standard complex UT, consider the system defined by

$$y = \cos[x] \tag{3.66}$$

[5]If $\mathbf{L}$ is the matrix square root of $\mathbf{R}_{\mathbf{x}} = \mathbf{L}\mathbf{L}^H$, then $\left(\sqrt{(L+\lambda)\mathbf{R}_{\mathbf{x}}}\right)_i$ is the ith column of the matrix $\sqrt{(L+\lambda)}\mathbf{L}$.



(a) circular input

(b) noncircular input

Figure 3.1: Performance of the complex UT and augmented complex UT

where the input x_n is a complex doubly white Gaussian white random variable with distribution $\mathcal{N}(0.5, 0.01)$. Figure 3.1a shows when input x is circular ($E\{x^2\} = 0$), the complex UT and the augmented complex UT had similar performance in capturing the distribution of the output y, while Figure 3.1b illustrates that when the input is noncircular ($E\{x^2\} = 0.008$), the augmented complex UT captures the orientation and power imbalance (pseudocovariance) of the distribution of the output, while the complex UT assumes a circular distribution.

The augmented complex unscented Kalman filter (ACUKF) corresponding to the nonlinear state space model defined in (3.47) is summarised in Algorithm 3. The novelty of the ACUKF algorithm presented in this work is that it does not assume a specific state or observation model, which makes it a more general form of the ACUKF presented in [7]. See Appendix A for a study of the advantages of utilising density functions which cater for improper distributions within the framework of complex valued particle filters with nonlinear state space models.

Algorithm 3: The augmented complex unscented Kalman filter (ACUKF)

Initialise with:

$$\begin{aligned}\widehat{\mathbf{x}}^a_{0|0} &= E\{\mathbf{x}^a_0\}\\ \mathbf{M}^a_{0|0} &= E\{(\mathbf{x}^a_0 - E\{\mathbf{x}^a_0\})(\mathbf{x}^a_0 - E\{\mathbf{x}^a_0\})^H\}\end{aligned}$$

Calculate sigma points for $i = 1, \ldots, 4L$

$$\begin{aligned}\mathcal{X}^a_{0,n-1} &= \widehat{\mathbf{x}}^a_{n-1|n-1}\\ \mathcal{X}^a_{i,n-1} &= \widehat{\mathbf{x}}^a_{n-1|n-1} \pm \left(\sqrt{(L+\lambda)\mathbf{M}^a_{n-1|n-1}}\right)_i\end{aligned} \tag{3.67}$$

Compute predictions:

$$\begin{aligned}\mathcal{X}^a_{i,n|n-1} &= \mathbf{f}^a[\mathcal{X}^a_{i,n-1}]\\ \widehat{\mathbf{x}}^a_{n|n-1} &= \sum_{i=0}^{4L} \mathcal{W}^{(m)}_i \mathcal{X}^a_{i,n|n-1}\\ \mathbf{M}^a_{n|n-1} &= \mathbf{Q}^a_n + \sum_{i=0}^{4L} \mathcal{W}^{(c)}_i \left(\mathcal{X}^a_{i,n|n-1} - \widehat{\mathbf{x}}^a_{n|n-1}\right)\left(\mathcal{X}^a_{i,n|n-1} - \widehat{\mathbf{x}}^a_{n|n-1}\right)^H\\ \mathcal{Y}^a_{i,n|n-1} &= \mathbf{h}^a[\mathcal{X}_{i,n|n-1}], \qquad i = 1, \ldots, 4L\\ \widehat{\mathbf{y}}^a_{n|n-1} &= \sum_{i=0}^{4L} \mathcal{W}^{(m)}_i \mathcal{Y}^a_{i,n|n-1}\end{aligned} \tag{3.68}$$

Measurement update:

$$\begin{aligned}\mathbf{R}^a_{\tilde{\mathbf{y}}^a,n|n-1} &= \mathbf{R}^a_n + \sum_{i=0}^{4L} \mathcal{W}^{(c)}_i \left(\mathcal{Y}^a_{i,n|n-1} - \widehat{\mathbf{y}}^a_{n|n-1}\right)\left(\mathcal{Y}^a_{i,n|n-1} - \widehat{\mathbf{y}}^a_{n|n-1}\right)^H\\ \mathbf{R}^a_{\mathbf{x}^a\mathbf{y}^a,n|n-1} &= \sum_{i=0}^{4L} \mathcal{W}^{(c)}_i \left(\mathcal{X}^a_{i,n|n-1} - \widehat{\mathbf{x}}^a_{n|n-1}\right)\left(\mathcal{Y}^a_{i,n|n-1} - \widehat{\mathbf{y}}^a_{n|n-1}\right)^H\\ \mathbf{G}^a_n &= \mathbf{R}^a_{\mathbf{x}^a\mathbf{y}^a,n|n-1}\left(\mathbf{R}^a_{\tilde{\mathbf{y}}^a,n|n-1}\right)^{-1}\\ \widehat{\mathbf{x}}^a_{n|n} &= \widehat{\mathbf{x}}^a_{n|n-1} + \mathbf{G}^a_n(\mathbf{y}^a_n - \widehat{\mathbf{y}}^a_{n|n-1})\\ \mathbf{M}^a_{n|n} &= \mathbf{M}^a_{n|n-1} - \mathbf{G}^a_n \mathbf{R}^a_{\tilde{\mathbf{y}}^a,n|n-1}\mathbf{G}^{aH}_n\end{aligned} \tag{3.69}$$

3.3.1 Performance analysis

Next we analyse the mean-square behavior of the the CEKF and CUKF for analytic state and observation functions. Consider the complex valued scalar state space given by

$$x_n = f[x_{n-1}] + w_n \tag{3.70}$$

$$y_n = h[x_n] + v_n \tag{3.71}$$

where $f[\cdot]$ and $h[\cdot]$ are holomorphic nonlinear state and observation models respectively, x_n and y_n are the state and observation, while w_n and v_n are uncorrelated zero-mean white complex-valued state (process) and observation (measurement) noises respectively. The process noise has variance $c_{w,n}$ and pseudocovariance $\rho_{w,n}$, while the measurement noise has a variance $c_{v,n}$ and pseudocovariance $\rho_{v,n}$.

The unscented and extended Kalman filters use the same general update expression, given by (3.69) and (3.58), to compute the estimate of the state, that is

$$\widehat{x}_{n|n} = \widehat{x}_{n|n-1} + g_n(y_n - \widehat{y}_{n|n-1}) \tag{3.72}$$

where g_n is the Kalman gain. This shows that the estimate comprises of a prediction term, $\widehat{x}_{n|n-1}$, and a weighted innovation term, $(y_n - \widehat{y}_{n|n-1})$. Substituting the state equation (3.70) in to the observation equation (3.71) yields

$$y_n = h\big[f[x_{n-1}] + w_n\big] + v_n \tag{3.73}$$

Now, let $z = f[x_{n-1}] + w_n$, then the TSE of the function $h[z] = h[f[x_{n-1}] + w_n]$ about $f[x_{n-1}]$ can be written as

$$h[f[x_{n-1}] + w_n] = h[f[x_{n-1}]] + \frac{\partial h}{\partial z} w_n + \frac{1}{2}\mathcal{H}_{zz} w_n^2 + \text{h.o.t.} \tag{3.74}$$

where h.o.t. stands for higher order terms, and the Jacobian $\frac{\partial h}{\partial z}$ and Hessian $\mathcal{H}_{zz} = \frac{\partial}{\partial z}\left(\frac{\partial h}{\partial z}\right)$ are evaluated at $z = f[x_{n-1}]$. Now subtract the true state, x_n, from the estimate given in

(3.72) to find the state estimation error

$$e_n = x_n - \widehat{x}_{n|n} = (f[x_{n-1}] + w_n) - \widehat{x}_{n|n-1} - g_n(y_n - \widehat{y}_{n|n-1}) \tag{3.75}$$

Substituting (3.73) and (3.74) into (3.75) yields

$$\begin{aligned} e_n = (f[x_{n-1}] + w_n) - \widehat{x}_{n|n-1} - g_n\Big(h[f[x_{n-1}]] \\ + \frac{\partial h}{\partial z} w_n + \frac{1}{2}\mathcal{H}_{zz} w_n^2 + \text{h.o.t.} + v_n - \widehat{y}_{n|n-1}\Big) \end{aligned} \tag{3.76}$$

Based on (3.76), the MSE, that is $E\{e_n e_n^*\}$, consists of a large number of terms, however, since we are interested in the effect of signal propriety, we shall restrict our analysis to terms related to the state and measurement noise pseudocovariances, that is

$$\begin{aligned} E\{e_n e_n^*\} = & -E\Big\{\frac{1}{2} g_n \mathcal{H}_{zz} w_n^2 \Big(f[x_{n-1}] - \widehat{x}_{n|n-1}\Big)^*\Big\} \\ & - E\Big\{\frac{1}{2}\Big(f[x_{n-1}] - \widehat{x}_{n|n-1}\Big) g_n^* \mathcal{H}_{zz}^* (w_n^*)^2\Big\} \\ & + E\Big\{\frac{1}{2} g_n \mathcal{H}_{zz} w_n^2 \Big(g_n\big(h[f[x_{n-1}] - \widehat{y}_{n|n-1}\big)\Big)^*\Big\} \\ & + E\Big\{\frac{1}{2}\Big(g_n\big(h[f[x_{n-1}] - \widehat{y}_{n|n-1}\big)\Big) g_n^* \mathcal{H}_{zz}^* (w_n^*)^2\Big\} \\ & + (\text{otherterms \& h.o.t.}) \\ = & -\Re\Big\{E\Big\{g_n \mathcal{H}_{zz}\Big(f[x_{n-1}] - \widehat{x}_{n|n-1}\Big)^*\Big\}\rho_{w,n}\Big\} \\ & + \Re\Big\{E\Big\{\,|g_n|^2\,\mathcal{H}_{zz}\big(h[f[x_{n-1}] - \widehat{y}_{n|n-1}\big)^*\Big\}\rho_{w,n}^*\Big\} \\ & + (\text{other terms \& h.o.t.}) \end{aligned} \tag{3.77}$$

where $\Re\{\cdot\}$ is the real part of a complex quantity.

Remark #7: From (3.77) observe that the MSE for the CUKF and CEKF depend on the pseudocovariance of the state noise ($\rho_{w,n}$), but not on the pseudocovariances of the observation noise. Hence, for nonlinear observation models, their mean square behaviors are affected by the impropriety of the state noise, regardless of whether the state equation is linear or nonlinear. However, for an arbitrary state space model it is

difficult to state how the performance is affected.

Remark #8: For linear state space models, the Hessian term $\mathcal{H}_{zz}$ in (3.77) vanishes, and consequently the terms shown in the MSE expression (3.77) dependent on the pseudocovariance. Therefore, the mean square characteristics of the conventional linear complex Kalman filter (CCKF) does not depend on the impropriety of the state or observation noises.

3.4 Application Examples

To illustrate the advantages of widely linear complex Kalman filters over their conventional counterparts, we considered the following case studies: 1) filtering for a noisy complex valued autoregressive process; 2) multistep ahead prediction for real-world noncircular and nonstationary wind data and the second order noncircular Lorenz attractor; 3) the nonlinear bearings only tracking. The results of these case studies are summarised in this section, and further results concerning the first and third studies are presented in Appendix A. For completeness the same explanatory texts regarding these studies are repeated in Appendix A.

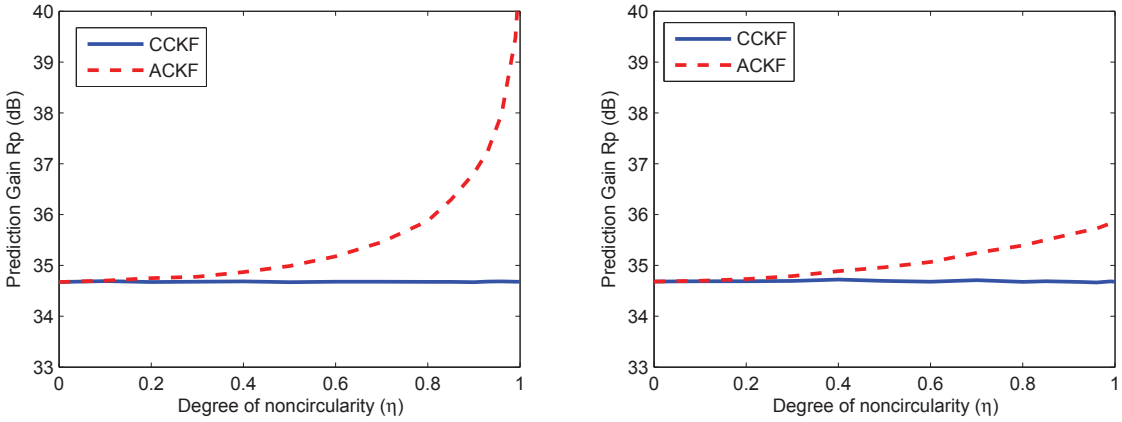
3.4.1 Complex autoregressive process

The performances of both the standard and widely linear Kalman filters were examined using the first order complex autoregressive process, AR(1), given by [7][28]

$$x_n = 0.9x_{n-1} + u_n$$

where the driving noise was u_n was doubly white Gaussian and zero-mean with variance 0.005 and a varying pseudocovariance. The observation equation for the linear filters, CCKF and ACKF, was given by

$$y_n = x_n + v_n$$



(a) Noncircular state noise

(b) Noncircular observation noise

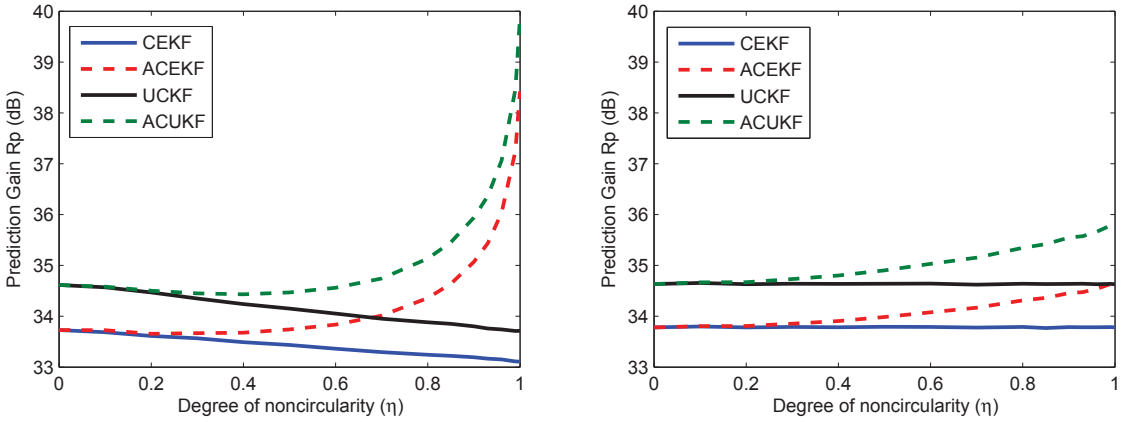
Figure 3.2: Steady-state performance comparison between CCKF and ACKF for the AR(1) filtering problem: (a) circular observation noise and a noncircular state noise with varying degrees of noncircularity; (b) circular state noise and noncircular observation noise with varying degrees of noncircularity.

while the observation equation for the nonlinear CEKF, CUKF and their augmented versions, was as follows:

$$y_n = \arctan[x_n] + v_n$$

where the function arctan is the inverse tangent function, and v_n a white Gaussian and zero-mean with variance 0.001 and a varying pseudocovariance. The ratio of pseudocovariance magnitude to covariance (circularity coefficient), that is $\eta = \frac{|\rho|}{c}$, was used as a measure for the degree of noncircularity of the complex state and measurement noises [10], where a complex random variable is circular for $\eta = 0$ and maximally noncircular for $\eta = 1$. 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 powers of the input (measurement) signal and the estimation error.

Figure 3.2 shows the performances of the CCKF and its corresponding widely linear (augmented) version, the ACKF. Figure 3.2a illustrates the results for a circular observation noise and a state noise with various degrees of noncircularity, while Figure 3.2b shows the results for a noncircular observation noise with a circular state noise. For both sets



(a) Noncircular state noise

(b) Noncircular observation noise

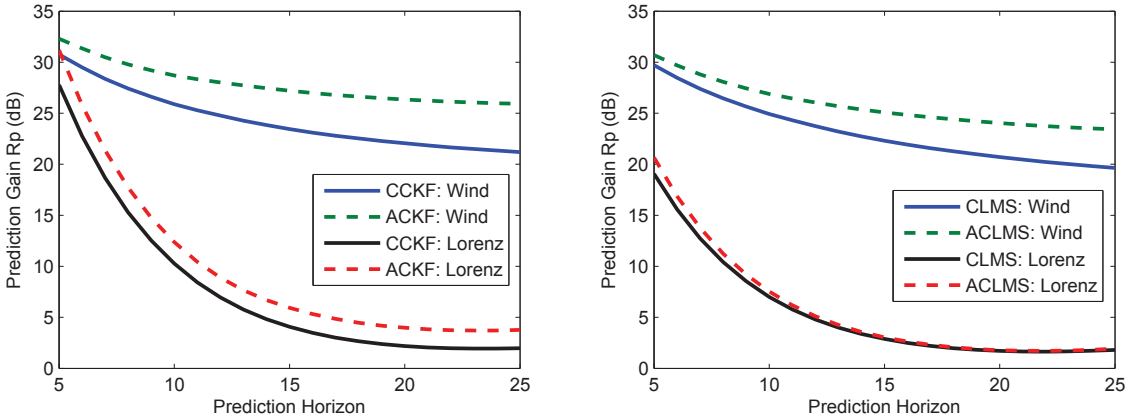
Figure 3.3: Steady-state performance comparison between CEKF, CUKF and their corresponding widely linear (augmented) versions for the AR(1) filtering problem: (a) circular observation noise and a noncircular state noise with varying degrees of noncircularity; (b) circular state noise and noncircular observation noise with varying degrees of noncircularity.

of simulations, when the noises were circular the ACKF had the same performance as the CCKF, while for noncircular noises the ACKF had superior performance as the degree of noise noncircularity (η) increased.

Figure 3.3 shows the corresponding results for the nonlinear CEKF, CUKF and their corresponding augmented versions, ACEKF and ACUKF. Similar to the ACKF case, the general pattern is that ACEKF and ACUKF outperform the CEKF and CUKF, respectively, if either of the state or observation noises are noncircular, while for circular noises they all had similar performances. However, when the state noise was noncircular, as illustrated in Figure 3.3a, the MSE behavior of CEKF and CUKF was dependent on the circularity of the state noise, while their performances were unaffected by the circularity of the observation noise, as shown in Figure 3.3b.

3.4.2 Multistep ahead prediction

The performances of the CCKF and ACKF were next assessed for the multistep ahead prediction of the improper *Lorenz* signal and real world improper and nonstationary *Wind* data. Simulations for the complex least mean square (CLMS) and its augmented (widely



(a) CCKF and ACKF (b) CLMS and ACLMS

Figure 3.4: Multistep ahead prediction of real-world Wind data and the Lorenz attractor using CCKF, CLMS and their corresponding widely linear versions

linear) version, the ACLMS, were also carried out to provide a performance comparison.

Figure 3.4 summarises the prediction performances for the *Lorenz* and the *Wind* data[6]. The ACKF was able to capture the underlying dynamics of the signals better than CCKF, which is indicated by its superior prediction performance. This can be attributed to the use of the widely linear 'augmented' model, which is better suited to capturing the full second order statistics of signals. Figure 3.4b shows the corresponding simulations for the CLMS and ACLMS, where the ACLMS is shown to have superior performance compared to the CLMS, but is worse than that of the ACKF.

3.4.3 Bearings only tracking

Bearings only tracking (BOT) is a problem encountered in many practical applications, including submarine tracking by passive sonar or aircraft surveillance. The objective is the online estimation of the kinematics (position and velocity) of a moving target using observer line of sight noise-corrupted bearing (phase) measurements[29]. As the range measurements are not available, the problem is inherently nonlinear. A single static sensor is unable to track targets using bearing measurements only, and in order to estimate

[6]The Wind signal (x_n), which has a magnitude (intensity) (ν_n) and direction (ϕ_n), is naturally represented as a complex signal ($x_n = \nu_n e^{\phi_n}$).

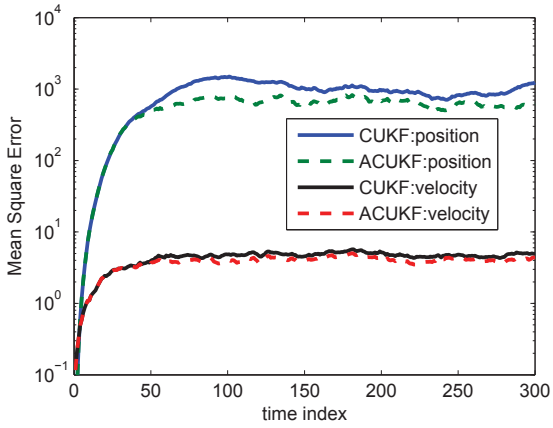


Figure 3.5: Performances of CUKF and ACUKF with second order noncircular state noise ($K = 0.9$)

the range, the sensor has to maneuver. However, for two or more stationary sensors, observability is not an issue, as the multiple bearing measurements can be used to form a range estimate.

To estimate the trajectory of a target at time instant n, that is, its position (x_n, y_n) and velocity $(\dot{x}_n, \dot{y}_n)$, for a system with L observers located at $(x^o_{i,n}, y^o_{i,n})$, $i = 1, 2, \ldots, L$, the complex BOT state space is defined as

$$\begin{aligned} \mathbf{x}_n &= \mathbf{F}\mathbf{x}_{n-1} + \mathbf{K}w_n \\ \mathbf{z}_n &= \mathbf{h}[\mathbf{x}_n] + \mathbf{v}_n \end{aligned}$$

with the variables defined as follows:

- $\mathbf{x}_n = \begin{bmatrix} x_n + jy_n & \dot{x}_n + j\dot{y}_n \end{bmatrix}^T$ is the complex target state vector,
- $\mathbf{F}$ and $\mathbf{K}$ are matrices defined as

$$\mathbf{F} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{K} = \begin{bmatrix} \frac{T^2}{2} \\ T \end{bmatrix}$$

  where T is the sampling interval,

- $\mathbf{z}_n$ is the observation vector and $\mathbf{h}[\mathbf{x}_n]$ is a vector function defined as

$$\mathbf{h}[\mathbf{x}_n] = \begin{bmatrix} \beta_{1,n} & \beta_{2,n} & \cdots & \beta_{L,n} \end{bmatrix}^T$$

  where $\beta_i = \tan^{-1} \frac{y_n - y^o_{i,n}}{x_n - x^o_{i,n}}$ is the target bearing at sensor i,

- $w_n = \ddot{x}_n + j\ddot{y}_n$ is the zero mean complex state noise (used to model unknown target accelerations), while $\mathbf{v}_n = \begin{bmatrix} v_{1,n} & v_{2,n} & \cdots & v_{L,n} \end{bmatrix}^T$ is the zero mean real valued observation noise with covariance $\mathbf{R}_{v,n}$.

The vector function $\mathbf{h}[\mathbf{x}_n]$ is real valued and it is straightforward to show that it does not satisfy the Cauchy Riemann-conditions, that is, $\frac{\partial \mathbf{h}[\mathbf{x}_n]}{\partial \mathbf{x}^*_n} \neq \mathbf{0}$, and is hence nonholomorphic.

To illustrate the benefits of ACUKF with over CUKF within the context of bearings only target motion analysis, consider a scenario with two static sensors located at $(-1200, 1300)$ and $(1000, 1500)$. The system described by (3.78) was simulated with a sampling interval of $T = 0.5$, and the mean square error (MSE) of the different algorithms were computed by averaging 100 independent trials.

The performances of the CUKF and the ACUKF were compared using a second order noncircular Gaussian state noise (with a degree of noncircularity of $\eta = 0.9$) with a distribution defined as

$$w_k \sim \mathcal{N}(0, 0.025) \qquad v_k \sim \mathcal{N}(0, 0.005)$$

The results, shown in Figure 3.5, illustrate that the ACUKF had a lower MSE in estimating both the position and velocity of the target compared to CUKF. This is due to the its ability to simultaneously cater for the covariance and pseudococariance of the signals.

3.5 Conclusions

We have readdressed the augmented complex Kalman filter (ACKF) and have examined its performance in relation to the conventional complex Kalman filter (CCKF). The anal-

ysis has shown that the ACKF offers significant performance gains over the CCKF for improper signals and the same performance as the CCKF for proper signals. A more general form of the augmented complex extended Kalman filter was also introduced, by utilising $\mathbb{CR}$ calculus which allows the Taylor series approximations of both holomorphic and nonholomorphic functions. The augmented complex unscented Kalman filter was also proposed under a framework employing augmented sigma points to cater for the complete second order statistical moments of signals. Analysis of the mean square characteristics of CCKF has shown that it is blind to the impropriety of the state and observation signals, however, the mean square characteristics of the complex extended Kalman filter (CEKF) and complex unscented Kalman filter (CUKF) are a function of the impropriety of the state noise noncircularity, when the observation equation is nonlinear.

Chapter 4

Widely Linear Frequency Estimation in Three-Phase Power Systems

In this chapter, we consider widely linear frequency estimation for three phase power systems utilising the Kalman filters developed in Chapter 3. The frequency of a power system is a crucial power quality parameter and is allowed to vary around its nominal value only within a prescribed tolerance level. Large frequency deviations are harmful to the system and arise in the presence of unbalanced system conditions, such as generation-consumption imbalances or unexpected conditions which require corrective actions. With the emergence of smart grids, system stability issues become even more pronounced, owing to more and more diversified generation and increasingly unpredictable power consumption. Frequency tracking and estimation in the context of smart grid is a key parameter for both the protection of power system and for improved power quality; for instance, frequent switching from the main grid to microgrids and electricity islands and dual natures of some loads require rapid frequency trackers to trigger corrective actions to maintain power quality.

A number of approaches for frequency tracking have been proposed, including least mean square adaptive filters [30], state space algorithms based on Kalman filters [31, 32], and Fourier transform based approaches [33, 34]. However, these techniques are either

only optimal for balanced systems (e.g. systems with line voltages of equal amplitudes) or are designed specifically for single-phase systems [35, 36, 37]; hence these techniques cannot fully characterise three-phase power systems where the line-to-line voltages also need to be taken into account.

In the future smart grid, the system frequency will undergo deviations due to: imbalance in generation (G) and load (L) (rise for $G > L$ and decay for $G < L$), single and dual phase faults or sags, dynamical loads and dual character of G an L, and a number of issues causing harmonics and transient stability issues (nonlinear loads, reactive power compensation). Accurate frequency estimators are a prerequisite for fault identification and troubleshooting, highlighting the need for a unified frequency estimation framework in three-phase power systems, which is:

- robust to measurement noise and harmonics in the system, including the slowly floating ones which are not integer multiplies of system frequency,
- real-time adaptive, fast converging, and asymptotically unbiased,
- minimum variance and statistically consistent, that is, approaching theoretical performance bounds,
- capable of catering for both balanced and unbalanced systems under the same umbrella, and at the same time tracking frequency and identifying system disturbance.

To deal with the three phase voltages simultaneously, standard frequency trackers employ Clarke's $\alpha\beta$ transformation which maps the three-phase voltages onto the variables v_0, v_α, and v_β, to produce the complex signal, $v = v_\alpha + jv_\beta$, with v_0 vanishing for a balanced system [32]. However, current strictly linear estimators are not capable of capturing full second order information for unbalanced voltage conditions, resulting in an oscillatory estimation error at twice the system frequency [38]. The recent work in [39] establishes that for unbalanced systems the $\alpha\beta$ voltage has a noncircular trajectory, and widely linear models are required for accurate system representation, achieved based on the augmented complex least mean square (ACLMS) [7].

The ACLMS framework is shown to cater for both balanced and unbalanced conditions, however, it is prone to slow convergence rates, and the assumption of noise free observations. This assumption is not practical, as real-world power systems are typically corrupted by harmonics and random noise sources. To this end, we here embark upon the stochastic gradient based widely linear frequency estimation framework in [39], and employ and analyse the performance of the widely linear (augmented) Kalman filters introduced in Chapter 3 for three phase frequency estimation. Owing to the underpinning state space representation, this approach offers enhanced accuracy and faster convergence, together with robustness to noise. Illustrative simulations on unbalanced and noisy real-world power systems support the analysis.

4.1 Background

4.1.1 Widely linear (augmented) Complex LMS (ACLMS)

The least mean square (LMS) algorithm is the most commonly used stochastic gradient adaptive filtering algorithm which adaptively estimates the filter coefficients that minimise the instantaneous squared error (the square of the difference between the desired signal and its estimate - the filter output). The standard complex least mean square (CLMS) algorithm is only suited to signals with circular (proper) distributions. With this mind, the widely linear (augmented) complex LMS (ACLMS) algorithm has been proposed to cater for both circular and noncircular signals [7], and is summarised in Algorithm 4: where d_k is the desired signal at time instant k, $\mathbf{x}_k^a = [\mathbf{x}_k^T, \mathbf{x}_k^H]^T$ the augmented input

Algorithm 4: Augmented complex LMS (ACLMS)

Filter output: $y_k = \mathbf{w}_k^{aT}\mathbf{x}_k^a$

Error: $e_k = d_k - y_k$

Weight updates:

$$\underbrace{\begin{bmatrix}\mathbf{h}_{k+1}\\ \mathbf{g}_{k+1}\end{bmatrix}}_{\equiv \mathbf{w}_{k+1}^a} = \underbrace{\begin{bmatrix}\mathbf{h}_k + \mu e_k \mathbf{x}_k^*\\ \mathbf{g}_k + \mu e_k \mathbf{x}_k\end{bmatrix}}_{\equiv \mathbf{w}_k^a + \mu e_k \mathbf{x}_k^{a*}} \tag{4.1}$$

regressor, $\mathbf{h}_k$ and $\mathbf{g}_k$ are the standard and conjugate filter weights, while $\mathbf{w}_k^a = \left[\mathbf{h}_k^T, \mathbf{g}_k^T\right]^T$ is the augmented weight vector.

4.2 Widely Linear Frequency Estimation

The instantaneous voltages in a three-phase power system are defined as

$$\begin{aligned} v_{a,k} &= V_{a,k}\cos(\omega kT + \phi) \\ v_{b,k} &= V_{b,k}\cos(\omega kT + \phi - 2\pi/3) \\ v_{c,k} &= V_{c,k}\cos(\omega kT + \phi + 2\pi/3) \end{aligned} \tag{4.2}$$

where $V_{a,k}$, $V_{b,k}$ and $V_{c,k}$ are the amplitudes of the of the three-phase voltages at time instant k, $\omega = 2\pi f$ the angular frequency with f being the system frequency, T the sampling interval, and ϕ the phase of the fundamental component. Clarke's transformation, given by

$$\begin{bmatrix} v_{0,k} \\ v_{\alpha,k} \\ v_{\beta,k} \end{bmatrix} = \sqrt{\frac{2}{3}} \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} v_{a,k} \\ v_{b,k} \\ v_{c,k} \end{bmatrix}, \tag{4.3}$$

maps the three-phase voltages onto a new domain where they can be conveniently represented by a scalar complex valued signal. In (4.3), the zero-sequence $v_{0,k}$ vanishes when the system is balanced, that is $V_{a,k} = V_{b,k} = V_{c,k}$, while $v_{\alpha,k} = A_k\cos(\omega kT + \phi)$ and $v_{\beta,k} = A_k\cos(\omega kT + \phi + \frac{\pi}{2})$ are orthogonal. In practice, the zero-sequence $v_{0,k}$ is not considered, and only v_α and v_β are used to form the complex valued recursive model for the system, that is[1]

$$v_k = v_{\alpha,k} + jv_{\beta,k} = A_k e^{j(\omega kT+\phi)} = v_{k-1}e^{j\omega T} \tag{4.4}$$

This model can be expressed as a state space model as shown in Algorithm 5, where the state x_k consists of the exponential $e^{j\omega T}$ whose argument contains the frequency f,

[1]The usual assumption in this type of estimation, that is $A_k \approx A_{k-1}$, is used.

Algorithm 5: State Space 1 - Linear (SS1-L)

$$\text{state equation:} \quad x_k = x_{k-1} + u_{k-1} \tag{4.6}$$
$$\text{observation equation:} \quad v_k = v_{k-1}x_k + n_k \tag{4.7}$$

v_k is the observation, while u_k and n_k are respectively the state and observations noise. This state space model is linear, and can be implemented using the conventional complex Kalman filter (CCKF). The selection of these noise variances is critical to the steady state error and convergence rates of Kalman filters, and will be discussed in more detail in the next section. The system frequency is derived from the state x as

$$\hat{f}_k = \frac{1}{2\pi T} \arcsin\left(\Im(x_k)\right) \tag{4.5}$$

where $\Im(\cdot)$ is the imaginary part of a complex quantity.

Example system operating conditions are illustrated in Figure 4.1, whereby for a balanced system Clarke's voltage v_k follows a circular trajectory, since the amplitude is time invariant and the angular frequency is proportional to the system frequency. However, this model is inaccurate when the system is operating under unbalanced conditions, the voltage amplitudes $V_{a,k}$, $V_{b,k}$ and $V_{c,k}$ are no longer equal, and the system trajectory becomes noncircular (ellipse in Figure 4.1). Therefore, for unbalanced systems the system takes on a widely linear model, that is [39]

$$v_k = v_{\alpha,k} + jv_{\beta,k} = A_k e^{j(\omega kT+\phi)} + B_k e^{-j(\omega kT+\phi)} \tag{4.8}$$

with

$$\begin{aligned} A_k &= \frac{\sqrt{6}(V_{a,k} + V_{b,k} + V_{c,k})}{6} \\ B_k &= \frac{\sqrt{6}(2V_{a,k} - V_{b,k} - V_{c,k})}{12} - \frac{\sqrt{2}(V_{b,k} - V_{c,k})}{4}j \end{aligned} \tag{4.9}$$

When the system is balanced and operating under nominal conditions, that is $V_{a,k} = V_{b,k} = V_{c,k}$, the coefficient B_k vanishes and system is accurately characterised by the

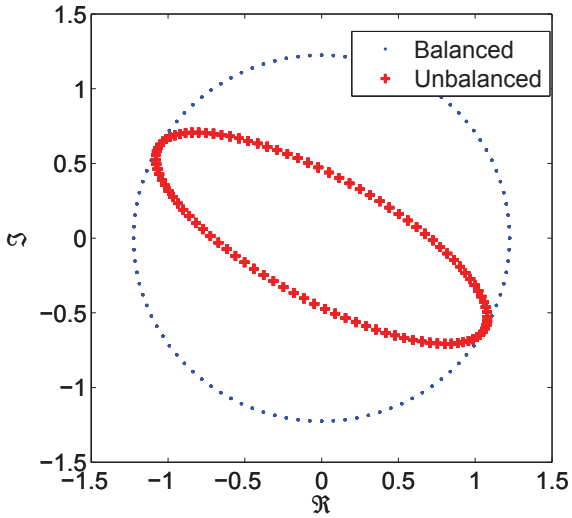


Figure 4.1: An illustration of the trajectory of Clarke voltage v_k for different operating conditions. For a balanced system, characterised by $V_{a,k} = V_{b,k} = V_{c,k}$, the trajectory of v_k is circular, while, for unbalanced systems, such as in the case of a 100% single-phase voltage sag illustrated by the ellipse in the figure (+), the trajectory of the output voltage becomes noncircular.

strictly linear model in (4.4). However, the expression in (4.8) is general and models the system under both balanced and unbalanced conditions, and can be written recursively as

$$v_k = v_{k-1}h_{k-1} + v_{k-1}^* g_{k-1} \tag{4.10}$$

which is a first-order widely linear autoregressive model, WLAR(1). The corresponding widely linear (augmented) state space model is defined in Algorithm 6, where the state vector consists of the strictly linear weight h_k and conjugate weight g_k, the observation v_k is a widely linear function of the previous observation, while, $\mathbf{u}_{h,k}$ and $\mathbf{u}_{g,k}$ are the state noise signals corresponding to h_k and g_k. The system frequency is now computed as

$$\hat{f}_k = \frac{1}{2\pi T} \arcsin\left(\Im(h_k + a_k g_k)\right) \tag{4.11}$$

$$a_k = \frac{-j\Im(h_k) + j\sqrt{\Im^2(h_k) - |g_k|^2}}{g_k}$$

Central to both the state space models SS1 and SS2 is the assumption of noise-free

Algorithm 6: State Space 2 - Widely Linear (SS2-WL)

state equation:

$$\begin{bmatrix} h_k \\ g_k \\ h_k^* \\ g_k^* \end{bmatrix} = \begin{bmatrix} h_{k-1} \\ g_{k-1} \\ h_{k-1}^* \\ g_{k-1}^* \end{bmatrix} + \begin{bmatrix} u_{h,k-1} \\ u_{g,k-1} \\ u_{h,k-1}^* \\ u_{g,k-1}^* \end{bmatrix} \tag{4.12}$$

observation equation:

$$\begin{bmatrix} v_k \\ v_k^* \end{bmatrix} = \begin{bmatrix} v_{k-1} & v_{k-1}^* & 0 & 0 \\ 0 & 0 & v_{k-1}^* & v_{k-1} \end{bmatrix} \begin{bmatrix} h_k \\ g_k \\ h_k^* \\ g_k^* \end{bmatrix} + \begin{bmatrix} n_k \\ n_k^* \end{bmatrix} \tag{4.13}$$

observations. If the three-phase voltages $v_{a,k}$, $v_{b,k}$ and $v_{c,k}$ are corrupted by zero-mean noises $n_{a,k}$, $n_{b,k}$ and $n_{c,k}$ respectively, then the output voltage is given by

$$v_k = v_{\alpha,k} + jv_{\beta,k} + n_k \tag{4.14}$$

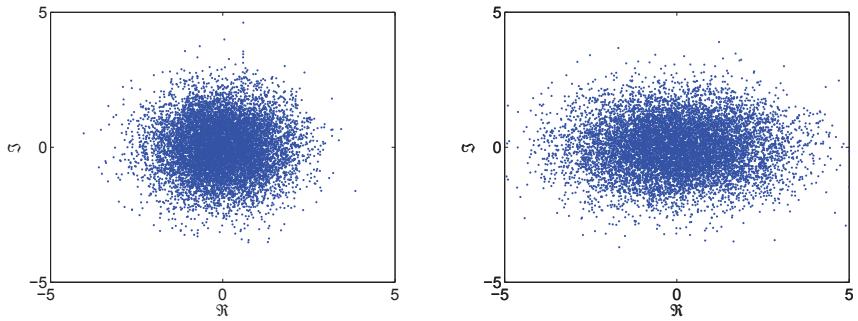
From Clarke's transform, the $\alpha\beta$ transformed noise is

$$\begin{aligned} n_k &= n_{\alpha,k} + jn_{\beta,k} \\ &= \sqrt{2/3}\Big(n_{a,k} - \frac{1}{2}n_{b,k} - \frac{1}{2}n_{c,k}\Big) + j\sqrt{2/3}\Big(\frac{\sqrt{3}}{2}n_{b,k} - \frac{\sqrt{3}}{2}n_{c,k}\Big) \end{aligned} \tag{4.15}$$

which is a zero-mean complex noise with pseudocovariance

$$\begin{aligned} p_{n,k} &= E\{n_k^2\} \\ &= E\Big\{\frac{2}{3}n_{a,k}^2 - \frac{1}{3}n_{b,k}^2 - \frac{1}{3}n_{c,k}^2 - \frac{2}{3}n_{a,k}n_{b,k} - \frac{2}{3}n_{a,k}n_{c,k} - -\frac{4}{3}n_{b,k}n_{c,k}\Big\} \\ &\quad jE\Big\{\frac{2\sqrt{3}}{3}n_{a,k}n_{b,k} - \frac{2\sqrt{3}}{3}n_{a,k}n_{c,k} - \frac{2\sqrt{3}}{3}n_{b,k}^2 + \frac{2\sqrt{3}}{3}n_{c,k}^2\Big\} \end{aligned} \tag{4.16}$$

The impropriety of the noise n_k is determined by the ratios of the variances and cross-correlations of the three phase observation noises $n_{a,k}$, $n_{b,k}$ and $n_{c,k}$. Figure 4.2 shows that the transformed noise n_k is proper, if and only if, the three phase noises are all uncorrelated and have identical variances, otherwise it is improper. Thus, equal line noise powers provide a circular (proper) Clarke's noise, whereas combinations of different noises



(a) $n_a, n_b, n_c \sim \mathcal{N}(0,1)$ (b) $n_a \sim \mathcal{N}(0,3), n_b, n_c \sim \mathcal{N}(0,1)$

(c) $n_a, n_c \sim \mathcal{N}(0,1), n_b \sim \mathcal{N}(0,3)$ (d) $n_a, n_b \sim \mathcal{N}(0,1), n_c \sim \mathcal{N}(0,3)$

Figure 4.2: Observation noise distributions after the three phase (independent, Gaussian and real valued) noises $n_{a,k}$, $n_{b,k}$ and $n_{c,k}$ undergo Clarke's $\alpha\beta$ transformation.

powers provide improper Clarke's noises with different degrees and natures of impropriety. For instance, the case in Figure 4.2b is improper with $n_{\alpha,k} \perp n_{\beta,k}$, whereas the cases in Figures 4.2c and 4.2d are improper with $n_{\alpha,k}$ and $n_{\beta,k}$ exhibiting different characters of correlations. Therefore, the impropriety of noise should be dealt with within the algorithm structure, such as in ACKF.

In the presence of noise, the recursion for Clarke's voltage can be found by substituting (4.14) into (4.10), that is

$$\begin{aligned} v_k &= \Big(v_{k-1} + n_{k-1}\Big)h_{k-1} + \Big(v^*_{k-1} + n^*_{k-1}\Big)g_{k-1} + n_k \\ &= v_{k-1}h_{k-1} + v^*_{k-1}g_{k-1} + n_{k-1}h_{k-1} + n^*_{k-1}g_{k-1} + n_k \end{aligned} \tag{4.17}$$

Remark #1: The observation noise n_k in (4.17) is additive, while the previous noise n_{k-1} has a multiplicative effect on the current observation. However, in practice, the

Algorithm 7: State Space 3 - Widely Linear (SS3-WL)

state equation:

$$\begin{bmatrix} h_k \\ g_k \\ \hat{v}_k \\ h_k^* \\ g_k^* \\ \hat{v}_k^* \end{bmatrix} = \begin{bmatrix} h_{k-1} \\ g_{k-1} \\ \hat{v}_{k-1}h_{k-1} + \hat{v}_{k-1}^* g_{k-1} \\ h_{k-1}^* \\ g_{k-1}^* \\ \hat{v}_{k-1}^* h_{k-1}^* + \hat{v}_{k-1} g_{k-1}^* \end{bmatrix} + \mathbf{u}_{k-1} \tag{4.18}$$

observation equation:

$$\begin{bmatrix} v_k \\ v_k^* \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} h_k \\ g_k \\ \hat{v}_k \\ h_k^* \\ g_k^* \\ \hat{v}_k^* \end{bmatrix} + \begin{bmatrix} n_k \\ n_k^* \end{bmatrix} \tag{4.19}$$

dynamics of real world three-phase systems do not follow this model, that is, the current observation is not a function of the previous observation noise, and using (4.17) to model noisy real world systems can lead to degraded or diverging estimates.

We next propose a more realistic sequential state space model where the current observation is independent of the previous observation noises. This is achieved by augmenting the state vector to include the output voltage. The resulting state space is summarised in Algorithm 7.

Remark #2: The state space model in Algorithm 7 does not use the previous observation to form the current observation, and hence does not propagate previous observation noises, thus providing a more a realistic and robust characterisation of real world systems. The state equation is nonlinear due to the coupling between v and x, and can be implemented using the augmented complex extended and unscented Kalman filters, ACEKF and ACUKF [1].

A strictly linear version of the state space model defined by (4.18) and (4.19) was proposed in [32], where the output voltage was estimated using the strictly linear model in (4.4). For convenience, this state space model is described is Algorithm 8.

Algorithm 8: State Space 4 - Linear (SS4-L)

state equation:

$$\begin{bmatrix} x_k \\ \hat{v}_k \end{bmatrix} = \begin{bmatrix} x_{k-1} \\ \hat{v}_{k-1} x_k \end{bmatrix} + \mathbf{u}_{k-1} \tag{4.20}$$

observation equation:

$$v_k = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} x_k \\ \hat{v}_k \end{bmatrix} + n_k \tag{4.21}$$

Remark #3: Owing to its strictly linear nature, Algorithm 8 suffers from the same limitations as the state space model described by Algorithm 5, namely it is not suited to systems operating in unbalanced conditions and in the presence of voltage sags or transients.

4.3 Robust Tracking Using the Innovation Process

The covariances of the state and observation noises govern the steady state error as well as the convergence speed of Kalman filters [40]. The noise statistics should ideally be matched to the system operating conditions: random interferences should be reflected in the statistics of the observation noise, while changes in the system dynamics, such as voltage sags and varying frequency, should be reflected in the state process.

Solutions for the estimation of the statistics of the state and observations noises mostly assume a degree of stationary. However, in a real world power system, the true noise statistics are generally unknown and almost invariably nonstationary, and the exact time instances at which changes occur in the system are generally unpredictable.

To cater for these uncertainties, we propose to employ the innovation process $\nu_k = \mathbf{y}_k^a - \mathbf{H}_k^a \widehat{\mathbf{x}}_{k|k-1}^a$ within the Kalman filter, that is, the difference between the actual and predicted observations, and use large changes in the innovation as an indication of changes in the system dynamics. To this end, we first show that if the state and observation noise covariance matrices are simultaneously scaled by the same factor, the Kalman gain and state estimate are unaltered [41]. In other words, it is the ratio between the state and

observation noise variances which determines the Kalman gain and not the exact values of these variances, hence the actual noise variances need not be known. This can be explained as follows. The time updates for the predicted state covariance matrix $\mathbf{M}^a_{k|k-1}$ can be expressed as

$$\begin{aligned}\mathbf{M}^a_{k+1|k} &= \mathbf{F}^a_{k-1}\mathbf{M}^a_{k|k-1}\mathbf{F}^{aH}_{k-1} - \mathbf{F}^a_{k-1}\mathbf{M}^a_{k|k-1}\mathbf{H}^{aH}_k[\mathbf{H}^a_k\mathbf{M}^a_{k|k-1}\mathbf{H}^{aH}_k + \mathbf{R}^a_{\mathbf{n},k}]^{-1} \\ &\quad \times\mathbf{H}^a_k\mathbf{M}^a_{k|k-1}\mathbf{F}^{aH}_{k-1} + \mathbf{R}^a_{\mathbf{u},k-1} \end{aligned} \tag{4.22}$$

that is, via a Riccati recursion, with initial condition $\mathbf{M}^a_{0|-1} = \mathbf{M}^a_0$. The computations for $\mathbf{M}^a_{k|k-1}$ and $\mathbf{M}^a_{k|k}$ are independent of the observations, and can be calculated without any knowledge of the observations. The state covariance matrix $\mathbf{M}^a_{k|k}$ can be computed from $\mathbf{M}^a_{k|k-1}$ by substituting expression (3.10) into (3.12) and using the matrix inversion lemma[2], that is

$$\begin{aligned}\mathbf{M}^a_{k|k} &= \mathbf{M}^a_{k|k-1} - \mathbf{M}^a_{k|k-1}\mathbf{H}^{aH}_k[\mathbf{H}^a_k\mathbf{M}^a_{k|k-1}\mathbf{H}^{aH}_k + \mathbf{R}^a_{\mathbf{n},k}]^{-1}\mathbf{H}^a_k\mathbf{M}^a_{k|k-1} \\ &= [(\mathbf{M}^a_{k|k-1})^{-1} + \mathbf{H}^{aH}_k(\mathbf{R}^a_{\mathbf{n},k})^{-1}\mathbf{H}^a_k]^{-1} \end{aligned} \tag{4.23}$$

Suppose $\mathbf{M}^a_0$, $\mathbf{R}^a_{\mathbf{n},k}$ and $\mathbf{R}^a_{\mathbf{u},k}$, $k \geq 0$, are replaced by scaled versions, that is

$$\begin{aligned}\bar{\mathbf{M}}^a_0 &= \kappa\mathbf{M}^a_0 \\ \bar{\mathbf{R}}^a_{\mathbf{n},k} &= \kappa\mathbf{R}^a_{\mathbf{n},k} \\ \bar{\mathbf{R}}^a_{\mathbf{u},k} &= \kappa\mathbf{R}^a_{\mathbf{u},k}\end{aligned}$$

where $\kappa > 0$ is a positive constant. Then a repeated application of (4.23) of (4.22) shows that the resulting covariance matrices satisfy

$$\begin{aligned}\bar{\mathbf{M}}^a_{k|k-1} &= \kappa\mathbf{M}^a_{k|k-1} \\ \bar{\mathbf{M}}^a_{k|k} &= \kappa\mathbf{M}^a_{k|k}\end{aligned} \tag{4.24}$$

[2]Woodbury matrix inversion identity is defined as: $(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B})^{-1}\mathbf{DA}^{-1}$

It then follows from (3.10) that the Kalman gain remains unchanged; that is $\bar{\mathbf{G}}_k^a = \mathbf{G}_k^a$.

Hence, the observation noise variance can be set to a positive constant, while the state noise variance can be chosen to balance between the convergence speed and the steady state error. With this in mind, for the remainder of this section the state and observation noise variances are assumed fixed, however, the system operating conditions (e.g. balanced to unbalanced) are allowed to change randomly.

A convenient way of detecting changes in system dynamics is to monitor the innovation ν_k. Large changes in the innovation indicate that the observed signal does not conform to the Kalman filter state estimate, and consequently the state estimate is inaccurate. In such scenarios, we propose to mitigate this issue by setting the state noise variance to a relatively large value, so that the state is re-estimated from the observations.

An estimate of the innovation power is an L sample moving average[3] given by

$$|\bar{\nu}_k|^2 = \frac{1}{L} \sum_{i=k-L-1}^{k} |\nu_i|^2, \tag{4.25}$$

At time k, if $|\nu_k|^2 > c|\bar{\nu}_{k-1}|^2$, where $c > 1$ is a threshold, then the state estimate is considered inaccurate and the state noise variance is increased for the next time instant. This allows for the detection of changes in system dynamics (e.g. the occurrence of voltage sags), and hence, the noise variances can be set accordingly.

4.4 Simulations

The proposed widely linear sequential state estimation algorithms were assessed for a simulated benchmark system using a 5kHz sampling rate, and were all initialised to 50.5Hz. The strictly linear state space models, SS1 and SS2, were implemented using the conventional complex Kalman filter (CCKF), SS4 was implemented using the conventional complex extended Kalman (CEKF), while the (widely linear) augmented CEKF (ACEKF) was used for SS3 [1]. Their performances were compared with those of with their stochas-

[3]In power systems operating at 50Hz, the cycle period is 20ms, and it is of interest to estimate changes in the system frequency in less than the duration of this cycle, hence a high sampling frequency is employed.

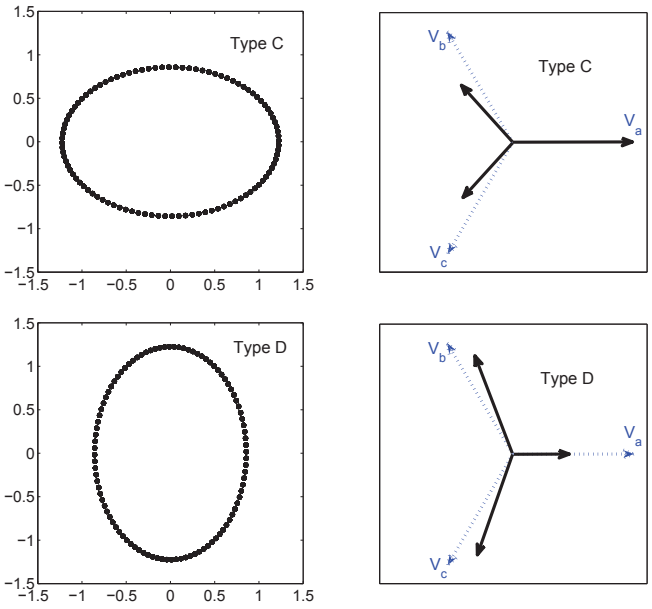


Figure 4.3: Geometric and phasor views of Type C and D voltage sags. The real-imaginary plots illustrate the noncircularity of Clarke's voltage in unbalanced conditions. The parameters of the circularity plot (ellipse) help identify the type of fault (voltage sag).

tic gradient based counterparts, the strictly linear complex least mean square (CLMS) and the augmented CLMS (ACLMS)[7].

In the first set of simulations, the performances of the algorithms were evaluated for an initially balanced system which became unbalanced after undergoing a Type C voltage sag starting at 0.1s, characterised by a 20% voltage drop and 10^o phase offset on both v_b and v_c, followed by a Type D sag starting at 0.25s, characterised by a 20% voltage drop at line v_a and a 10% voltage drop on both v_b and v_c with a 5^o phase angle offset, as illustrated in Figure 4.3. Observe from Figure 4.4 that for an unbalanced system, the widely linear algorithms, ACLMS, SS2 and SS3, were able to accurately estimate the system frequency, conforming with the analysis, while the strictly linear algorithms, CLMS, SS1 and SS4, yielded oscillating frequency estimates due to under-modelling of the system. The widely linear and strictly linear algorithms had similar performances under balanced conditions, as illustrated in the time interval upto 0.1s. Due to their stochastic gradient nature, CLMS and ACLMS had relatively slow convergence compared with the state space based Kalman

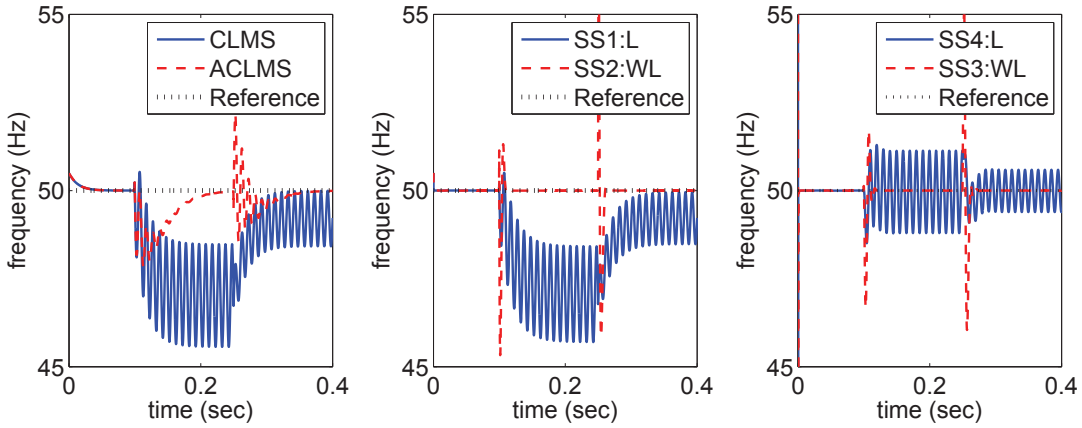


Figure 4.4: Frequency estimation for a system which is balanced up to 0.1s, after which the system becomes unbalanced due to the occurrence of voltage sags of differing natures.

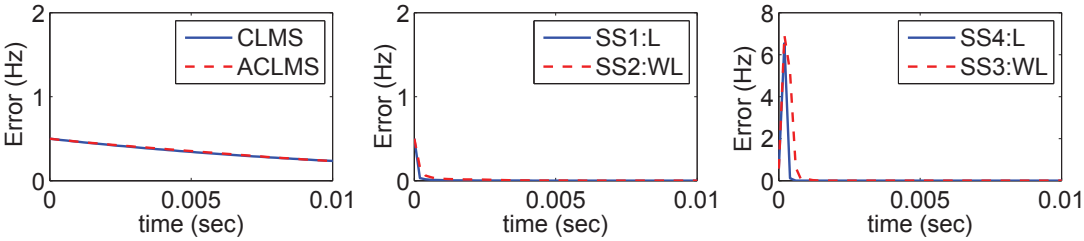


Figure 4.5: Initial transient behaviour for the simulations in Figure 5 (first 5ms), where all the Kalman filters where initialised as $\mathbf{M}^{a}_{k|k} = 10\mathbf{I}$.

filter algorithms, as illustrated in Figure 4.5.

Figure 4.6 illustrates frequency estimation in the presence of Gaussian noise. As expected, CLMS, ACLMS, SS1 and SS2, which assume noise free observations, gave inaccurate estimates, while the more general SS3 and SS4 provided accurate frequency estimates. Figure 4.7 illustrates frequency estimation in the presence of in-phase harmonic observation noise. Again, only SS3 and SS4, which do not assume noise free observations, converged to the true system frequency, and the remaining algorithms gave inaccurate estimates.

The performance of the algorithms for a power system which undergoes rise and decay in frequency, a typical case where generation does not match the load like in microgrids and islanding, is illustrated in Figure 4.8. The widely linear algorithms, ACLMS, SS2 and SS3, were able to accurately track the system frequency, as opposed to their corresponding strictly linear counterparts. Moreover, the stochastic gradient based ACLMS

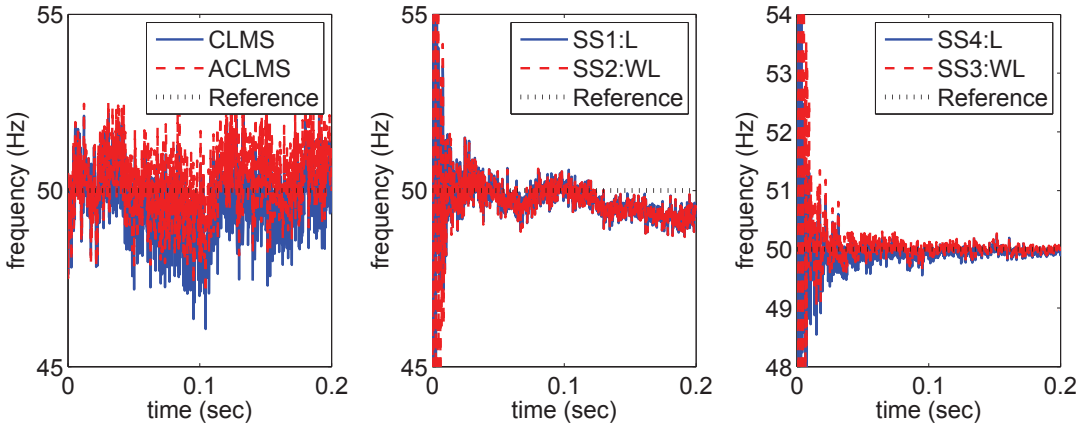


Figure 4.6: Frequency estimation for a balanced system in the presence of doubly white circular Gaussian noises at 20dB SNR.

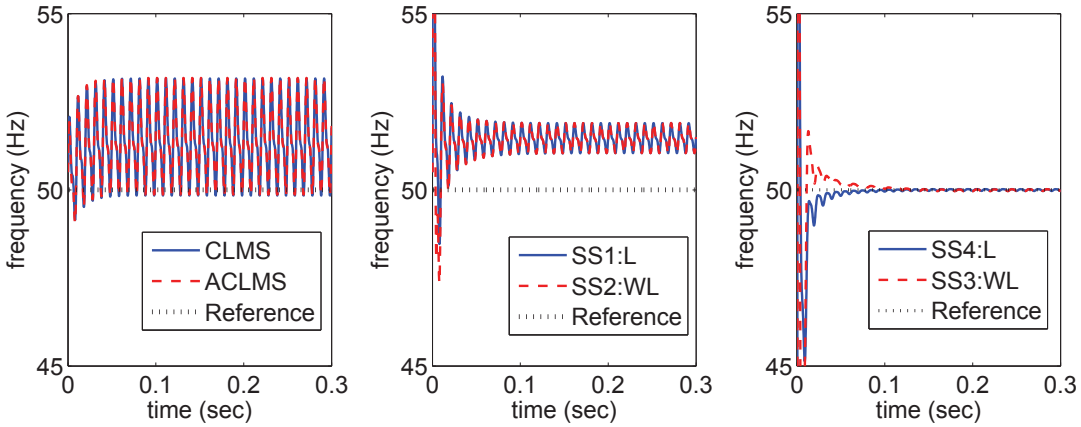


Figure 4.7: Frequency estimation when phase voltages are contaminated with in-phase harmonics at 10% p.u. for the 3rd and 5% p.u. for the 5th harmonics.

was outperformed by the widely linear Kalman filter algorithms, SS2 and SS3.

The statistical advantage of the widely linear estimators over their corresponding strictly linear estimators is illustrated by comparing the bias and mean square errors (MSEs) in the presence of Gaussian doubly white circular complex noise[4]. Figure 4.9 shows the performance of the algorithms for a system undergoing a Type D voltage sag (see Figure 4.3). The results in Figure 4.9a illustrate that the widely linear algorithms, ACLMS, SS2 and SS3, had decreasing MSEs as the signal to noise ratio (SNR) increased, while the strictly linear algorithms, CLMS, SS1 and SS4, yielded relatively large, almost

[4]For white Gaussian noise, $n_k = n_{r,k} + jn_{i,k}$, double whiteness implies $E\{n_k \cdot n_l^*\} = \sigma^2\delta_{k-l}$ and $E\{n_k \cdot n_l\} = \rho^2\delta_{k-l}$, where σ^2 and ρ^2 are the noise variance and pseudovariance.

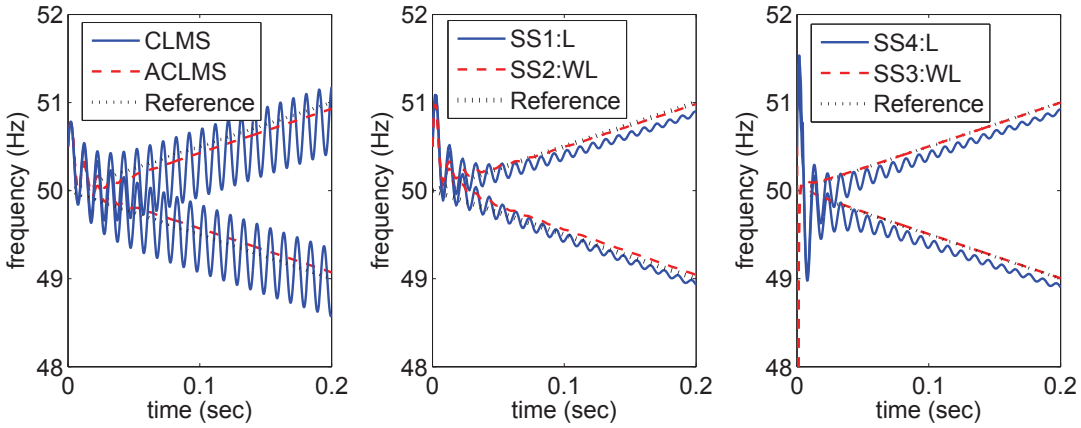
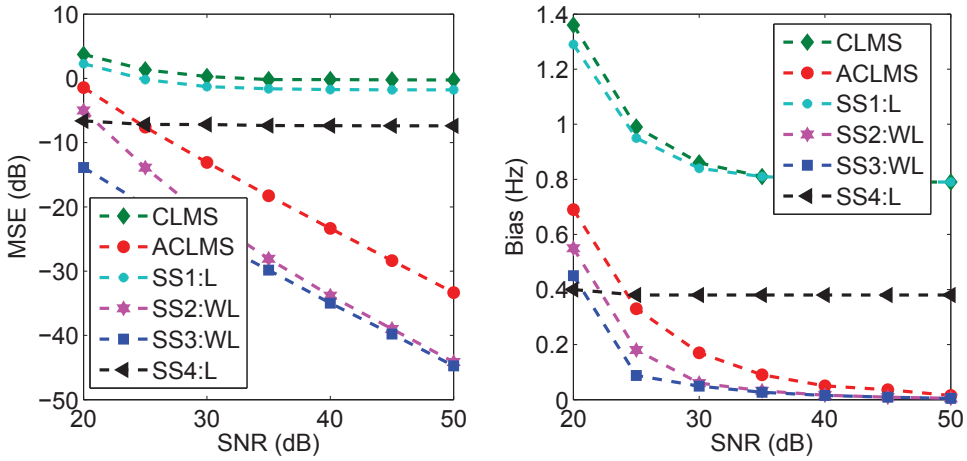


Figure 4.8: Frequency estimation for a power system which experiences a 5Hz/s rise and decay in system frequency.

constant, MSEs with increasing SNR. This can be attributed to the oscillating frequency estimates of these algorithms for unbalanced conditions that do not change with increasing SNR. Moreover, SS3 and SS4 had the best performances among the widely linear and strictly linear algorithms, respectively, because they did not assume noise free observations. Figure 4.9b shows the bias of the algorithms at different SNRs; observe that the algorithms based on the widely linear model offered the best performances, and that, again, the best results among the strictly linear and widely linear algorithms were achieved by SS4 and SS3 respectively. In other words, Figure 4.9 shows that the widely linear algorithms were asymptotically unbiased and statistically consistent.

Figure 4.10 illustrates the benefits of monitoring the innovation process to adjust the model to the changes in the system, evaluated for a system which undergoes a step change in frequency in the presence of additive white Gaussian observation noise using SS3 and SS4 (similar results can be shown for the other state space models). Observe the superior frequency estimation results, in terms of convergence speed and steady state error, when the state noise variance was set according to the changes in the innovation process, compared to using fixed noise variances.

The robustness of the proposed models to a combination of harmful events was examined using the setup in Figure 4.4, where an initially balanced system experienced consecutive voltage sags, together with the presence of doubly white Gaussian noise at

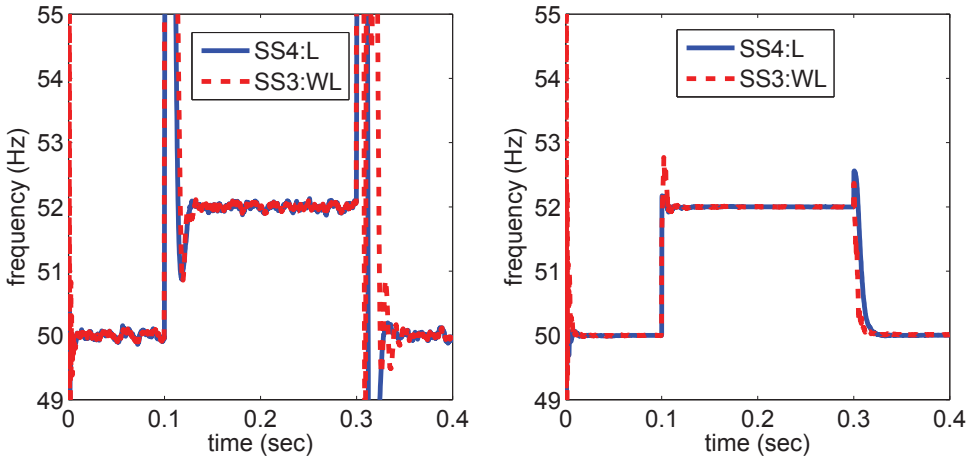


(a) MSE analysis (b) Bias analysis

Figure 4.9: Mean square error (MSE) and bias analysis for an unbalanced system undergoing a voltage sag (Type D).

40dB SNR, with the results shown in Figure 4.11. Observe that when the noise variances were set according to innovation power, the algorithm was more robust in the presence of the noise. Figure 4.11c shows the two peaks in the innovation process corresponding to the time instances when the system experienced the two different voltage sags.

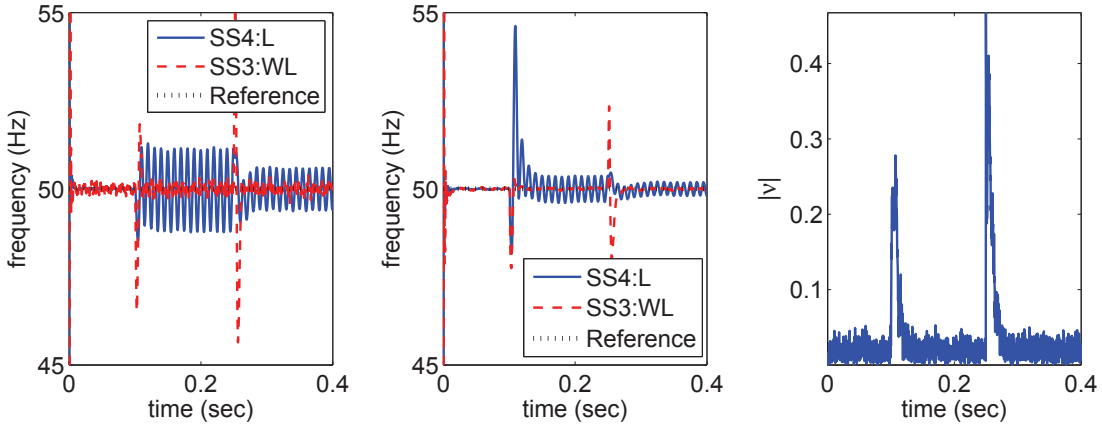
The last set of simulations explores frequency estimation for a real-world power system, where unbalanced three-phase voltages were recorded at a 110/20/10kV transformer station. The measured data come from a system with nominal frequency of 50Hz, sampled at a rate of 1kHz, and normalized with respect to their normal peak voltage value. The first set of results, for an unbalanced system (a single-phase short with earth), is shown in Figure 4.12, where the theoretical and practical superiority of the algorithms based on the widely linear model, ACLMS and SS3, compared with the strictly linear algorithms, CLMS and SS4, in unbalanced system conditions is highlighted. Figure 4.13 illustrates the performance of the algorithms for a real-world unbalanced system undergoing a two-phase short-cut with earth. Conforming with the analysis, the strictly linear algorithms, CLMS and SS4, yielded inaccurate, biased and oscillating frequency estimates, due to under-modelling of the system, while the algorithms based on the widely linear model, ACLMS and SS3, yielded accurate estimates, which were unbiased and with minimum variance, conforming with the ensemble analysis in Figure 4.9. In both simulations, the



(a) Fixed noise variances

(b) Noise variances set according to innovation power

Figure 4.10: Frequency estimation for a system which experiences a temporary step change in frequency from 50Hz to 52Hz in the presence of white circular Gaussian noises at 35dB SNR. In (a) the frequency is estimated using SS3 and SS4 with fixed state and observation noise variances, while in (b) the state noise variance was set according to the innovation power using the methodology described in Section 4.3.



(a) Fixed noise variances

(b) Noise variances set according to innovation power

(c) Innovation signal for SS3

Figure 4.11: An initially balanced system experiences a series of voltages sags, all in the presence of complex doubly white measurement noise.

state space based widely linear Kalman filter based algorithm, SS3, had a faster convergence rate and lower steady state error than the stochastic gradient based widely linear ACLMS algorithm.

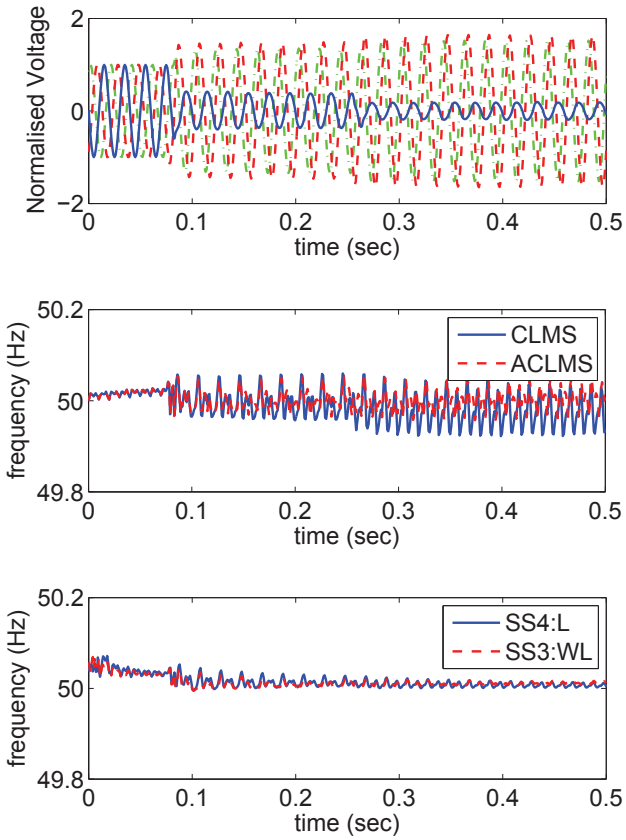


Figure 4.12: Frequency estimation for a real-world three-phase system, where an initially balanced system experienced a single-line short with earth.

4.5 Conclusions

We have introduced a novel, widely linear, framework for state space based frequency estimation in the context of three-phase power systems, under both balanced and unbalanced operating conditions. The signal, obtained from Clarke's $\alpha\beta$ transformation, is noncircular (improper) when the three-phase voltages are unbalanced, which makes the standard, strictly linear, estimation inadequate. It has been shown that accounting for noncircularity of amplitude distributions allows for both the development of second order optimal frequency estimation algorithms, and the identification of unbalanced conditions via circularity diagrams and degrees of impropriety. We have addressed frequency estimation from a state space perspective, and illustrated the superiority of the widely linear (augmented) complex Kalman filters over the stochastic gradient based augmented complex least mean square (ACLMS) algorithm. In order to increase convergence speed and reduce

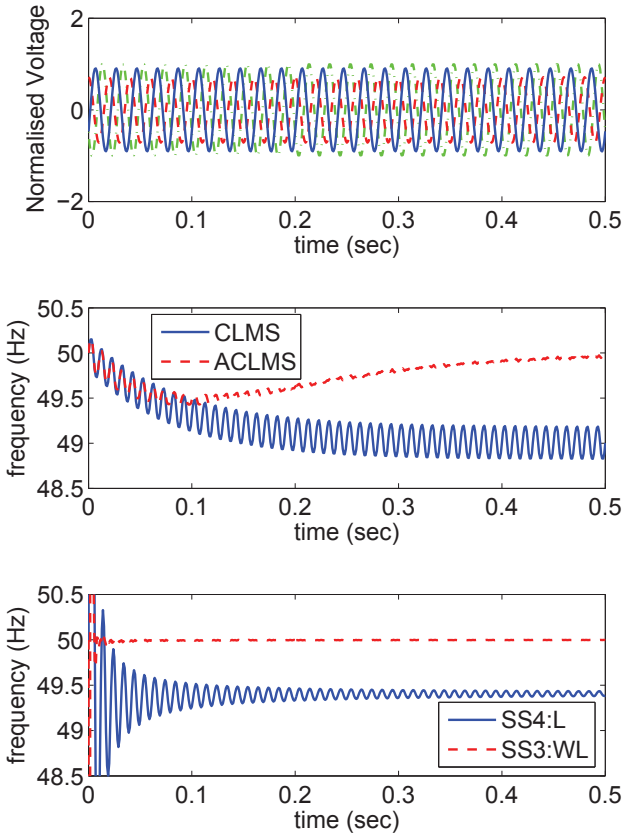


Figure 4.13: Frequency estimation for a real-world unbalanced three-phase system, where two lines experience a short with earth.

steady state error, a method based on the Kalman filter innovation process has also been proposed, and was shown to enhance the performance of the Kalman filters in terms of response and convergence rate. Comprehensive simulations over a range of power systems conditions evaluated for both balanced and unbalanced systems support the analysis.

Chapter 5

Distributed Widely Linear Complex Kalman Filters

In this Chapter, we extend the algorithms proposed in Chapter 3 to the case of cooperative sequential state space estimation, whereby nodes in a network collaborate locally to estimate improper complex signals. Distributed estimation and fusion has received significant attention in both military and civilian applications [42][43][44][45], and the recent advances in sensor technology and wireless communications have highlighted the usefulness of distributed networks in this context [46][42]. Such models rely on cooperation between the nodes (sensors) to provide more accurate and robust estimation compared to using independent uncooperative nodes, while approaching the performance of the more complex centralised systems.

This is achieved through nodes equipped with learning capabilities that take local measurements (observations) and share information with their neighbours, thus enhancing robustness to link and node failures and facilitating scalability [47][48][49]. A number of robust and scalable diffusion strategies for network cooperation have been developed for distributed least-mean-square estimation [50][51] and Kalman filtering [47][44], however, these are linked to a very restrictive class of proper signals, and are also inadequate for correlated measurement noises, a common case in practice. In this work, we focus on distributed solutions suited to general improper complex data arising from distributed

real-world applications, such as in wireless communication systems [52][53] and power systems [39].

Extending the recent work on widely linear estimation and distributed Kalman filters [7][1][47][44], we here propose a distributed augmented (widely linear) complex Kalman filter (D-ACKF) that caters for general complex signals, as well as the cross-correlations between the observation noises at neighbouring nodes. These are real-world scenarios encountered when node signals are exposed to common noise, such as:

- multi-sensor target tracking in the presence of observation jamming-noise;
- environmental noise in seismic arrays;
- signals from microphone array systems experiencing common interference;
- wireless sensor networks with overlapping user frequencies;
- distributed frequency estimation in smart grids experiencing common fault.

This work generalises earlier distributed Kalman filtering approaches [47][54][44], and illuminates the duality of D-ACKF with its corresponding bivariate real-valued distributed Kalman filter, highlighting several issues of implementation motivated by duality considerations. The performance of the D-ACKF is analysed, and supported by case studies on filtering autoregressive processes and projectile tracking, involving both proper and improper signals.

5.1 Diffusion Kalman Filtering

Consider the state space corresponding to a node i in a distributed system [18],

$$\mathbf{x}_n = \mathbf{F}_{n-1}\mathbf{x}_{n-1} + \mathbf{w}_n \tag{5.1a}$$

$$\mathbf{y}_{i,n} = \mathbf{H}_{i,n}\mathbf{x}_n + \mathbf{v}_{i,n} \tag{5.1b}$$

where $\mathbf{x}_n \in \mathbb{C}^L$ and $\mathbf{y}_{i,n} \in \mathbb{C}^K$ are respectively the state vector at time instant n and observation (measurement) vector at node i, while $\mathbf{F}_n$ and $\mathbf{H}_{i,n}$ are the state transition

and observation matrices, whereas $\mathbf{w}_n \in \mathbb{C}^L$ and $\mathbf{v}_{i,n} \in \mathbb{C}^K$ are respectively the white state and measurement noises at node i, and are assumed to be uncorrelated and zero-mean, with covariances and pseudocovariances defined as

$$E \begin{bmatrix} \mathbf{w}_n \\ \mathbf{v}_{i,n} \end{bmatrix} \begin{bmatrix} \mathbf{w}_k \\ \mathbf{v}_{i,k} \end{bmatrix}^H = \begin{bmatrix} \mathbf{Q}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{i,n} \end{bmatrix} \delta_{nk} \tag{5.2}$$

$$E \begin{bmatrix} \mathbf{w}_n \\ \mathbf{v}_{i,n} \end{bmatrix} \begin{bmatrix} \mathbf{w}_k \\ \mathbf{v}_{i,k} \end{bmatrix}^T = \begin{bmatrix} \mathbf{P}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_{i,n} \end{bmatrix} \delta_{nk} \tag{5.3}$$

where δ_{nk} is the Kronecker delta function.

5.1.1 Distributed Complex Kalman Filter

The distinguishing feature of the proposed class of distributed Kalman filters is that no assumption is made about the correlation of the observation noises at different nodes, thus extending earlier distributed Kalman filtering algorithms [44][47][54], and allowing us to deal more effectively with cases where the nodes experience common measurement noises.

Denote the neighbourhood of node i, that is, the set of nodes that can communicate directly with the node i (including itself) by $\mathcal{N}_i$, as illustrated in Figure 5.1. Within a distributed (diffusion) Kalman filtering framework, neighbouring nodes collaborate and share information to estimate the state vector $\mathbf{x}_n$. This can be achieved by using a diffusion technique to enable information sharing between neighbours.

Let $\underline{\widehat{\mathbf{x}}}_{i,n|n}$ denote the complex Kalman filter (CKF) state estimate at node i based on all the data from the neighbourhood $\mathcal{N}_i$ consisting of $M = |\mathcal{N}_i|$ nodes, where $|\mathcal{N}_i|$ denotes the number of nodes in the neighbourhood $\mathcal{N}_i$. The collective neighbourhood observation equation at node i is given by

$$\underline{\mathbf{y}}_{i,n} = \underline{\mathbf{H}}_{i,n}\mathbf{x}_n + \underline{\mathbf{v}}_{i,n} \tag{5.4}$$

with the collective (neighbourhood)) variables defined as

$$\underline{\mathbf{y}}_{i,n} = \left[\mathbf{y}_{i_1,n}^T, \mathbf{y}_{i_2,n}^T, \cdots, \mathbf{y}_{i_M,n}^T\right]^T$$
$$\underline{\mathbf{H}}_{i,n} = \left[\mathbf{H}_{i_1,n}^T, \mathbf{H}_{i_2,n}^T, \cdots, \mathbf{H}_{i_M,n}^T\right]^T$$
$$\underline{\mathbf{v}}_{i,n} = \left[\mathbf{v}_{i_1,n}^T, \mathbf{v}_{i_2,n}^T, \cdots, \mathbf{v}_{i_M,n}^T\right]^T$$

where $\{i_1, i_2, \ldots, i_M\}$ are all the nodes in the neighbourhood $\mathcal{N}_i$. The covariance and pseudocovariance of the collective observation noise vector are:

$$\underline{\mathbf{R}}_{i,n} = E\{\underline{\mathbf{v}}_{i,n}\underline{\mathbf{v}}_{i,n}^H\} = \begin{bmatrix} \mathbf{R}_{i_1,n} & \mathbf{R}_{i_1 i_2,n} & \cdots & \mathbf{R}_{i_1 i_M,n} \\ \mathbf{R}_{i_2 i_1,n} & \mathbf{R}_{i_2,n} & \cdots & \mathbf{R}_{i_2 i_M,n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{i_M i_1,n} & \mathbf{R}_{i_M i_2,n} & \cdots & \mathbf{R}_{i_M,n} \end{bmatrix}$$
$$\underline{\mathbf{U}}_{i,n} = E\{\underline{\mathbf{v}}_{i,n}\underline{\mathbf{v}}_{i,n}^T\} = \begin{bmatrix} \mathbf{U}_{i_1,n} & \mathbf{U}_{i_1 i_2,n} & \cdots & \mathbf{U}_{i_1 i_M,n} \\ \mathbf{U}_{i_2 i_1,n} & \mathbf{U}_{i_2,n} & \cdots & \mathbf{U}_{i_2 i_M,n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{U}_{i_M i_1,n} & \mathbf{U}_{i_M i_2,n} & \cdots & \mathbf{U}_{i_M,n} \end{bmatrix}$$

where $\mathbf{R}_{i_a,n} = E\{\mathbf{v}_{i_a,n}\mathbf{v}_{i_a,n}^H\}$, $\mathbf{R}_{i_a i_b,n} = E\{\mathbf{v}_{i_a,n}\mathbf{v}_{i_b,n}^H\}$, $\mathbf{U}_{i_a,n} = E\{\mathbf{v}_{i_a,n}\mathbf{v}_{i_a,n}^T\}$ and $\mathbf{U}_{i_a i_b,n} = E\{\mathbf{v}_{i_a,n}\mathbf{v}_{i_b,n}^T\}$, for $a, b \in \{1, 2, \ldots, M\}$.

Calculation of the neighbourhood state estimates is followed by the diffusion step, given by

$$\widehat{\mathbf{x}}_{i,n|n} = \sum_{k \in \mathcal{N}_i} c_{k,i}\underline{\widehat{\mathbf{x}}}_{k,n|n} \tag{5.5}$$

where the diffused state estimates $\widehat{\mathbf{x}}_{i,n|n}$ is the weighted estimates from the neighbourhood $\mathcal{N}_i$, and $c_{k,i} \geq 0$ are the weighting coefficients satisfying $\sum_{k \in \mathcal{N}_i} c_{k,i} = 1$. A number of fusion schemes have been proposed, including the Metropolis [50], Laplacian [55] and nearest neighbour method [44], however, the determination of the optimal weights for an arbitrary network of nodes is a difficult problem without accurate knowledge of the

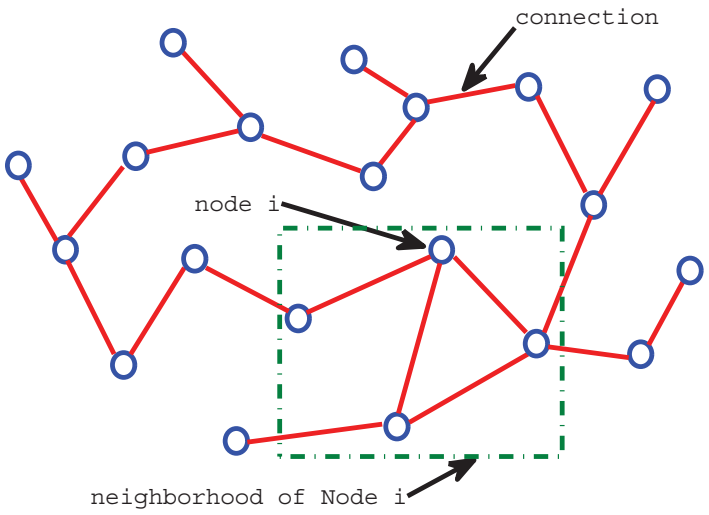


Figure 5.1: An illustrative example of a distributed network topology.

statistics of the local estimates [56].

The distributed complex Kalman filter (D-CKF) aims to approximate a centralised Kalman filter (with access to the observation data from all the nodes) via neighbourhood collaborations and diffusion, and is summarised in Algorithm 9. The D-CKF algorithm requires each node to form a collective observation equation as in (5.4) by gathering information from its neighbours, thereafter, each node computes a neighbourhood state estimate which are again transmitted to neighbours to be used for the diffusion step.

Remark #1: The D-CKF algorithm[1] is based on the standard (strictly linear) state space model (5.1), similar to existing algorithms [44] [57], and is thus inadequate for widely linear state space models or improper state and observation noises, where $\mathbf{P}_n \neq \mathbf{0}$ and $\mathbf{U}_{i,n} \neq \mathbf{0}$ for $i = 1, 2, \ldots, N$.

Remark #2: Observe that unlike existing distributed complex Kalman filters, the proposed D-CKF algorithm caters also for the cross-correlations between the

[1]The matrices $\mathbf{M}_{i,n|n}$ and $\mathbf{M}_{i,n|n-1}$ do not represent the covariances of $\widehat{\mathbf{x}}_{i,n|n}$ and $\widehat{\mathbf{x}}_{i,n|n-1}$, as is the case for the standard Kalman filter operating on linear Gaussian systems. This is due to the use of the suboptimal diffusion step, which updates the state estimate and not the covariance matrix $\mathbf{M}_{i,n|n}$.

Algorithm 9: The D-CKF

Initialisation: For each node $i = 1, 2, \ldots, N$

$$\begin{aligned} \widehat{\mathbf{x}}_{i,0|0} &= E\{\mathbf{x}_0\} \\ \mathbf{M}_{i,0|0} &= E\{(\mathbf{x}_0 - E\{\mathbf{x}_0\})(\mathbf{x}_0 - E\{\mathbf{x}_0\})^H\} \end{aligned}$$

For every time instant $n = 1, 2, \ldots$

– Evaluate at each node $i = 1, 2, \ldots, N$

$$\widehat{\mathbf{x}}_{i,n|n-1} = \mathbf{F}_{n-1}\widehat{\mathbf{x}}_{i,n-1|n-1} \tag{5.6}$$

$$\mathbf{M}_{i,n|n-1} = \mathbf{F}_{n-1}\mathbf{M}_{i,n-1|n-1}\mathbf{F}_{n-1}^H + \mathbf{Q}_n \tag{5.7}$$

$$\mathbf{G}_{i,n} = \mathbf{M}_{i,n|n-1}\underline{\mathbf{H}}_{i,n}^H\big(\underline{\mathbf{H}}_{i,n}\mathbf{M}_{i,n|n-1}\underline{\mathbf{H}}_{i,n}^H + \underline{\mathbf{R}}_{i,n}\big)^{-1} \tag{5.8}$$

$$\underline{\widehat{\mathbf{x}}}_{i,n|n} = \widehat{\mathbf{x}}_{i,n|n-1} + \mathbf{G}_{i,n}\big(\underline{\mathbf{y}}_{i,n} - \underline{\mathbf{H}}_{i,n}\widehat{\mathbf{x}}_{i,n|n-1}\big) \tag{5.9}$$

$$\mathbf{M}_{i,n|n} = (\mathbf{I} - \mathbf{G}_{i,n}\underline{\mathbf{H}}_{i,n})\mathbf{M}_{i,n|n-1} \tag{5.10}$$

– For every node i, compute the diffusion update as

$$\widehat{\mathbf{x}}_{i,n|n} = \sum_{k\in\mathcal{N}_i} c_{k,i}\underline{\widehat{\mathbf{x}}}_{k,n|n} \tag{5.11}$$

neighbourhood observation noises, while for uncorrelated nodal observation noises, it degenerates into Algorithm 9 in [44].

5.1.2 Distributed Augmented Complex Kalman Filter

To cater for widely linear state and observation models together with improper signals, the widely linear version of the distributed state space model (5.1) is defined as[1]

$$\mathbf{x}_n = \mathbf{F}_{n-1}\mathbf{x}_{n-1} + \mathbf{A}_{n-1}\mathbf{x}_{n-1}^* + \mathbf{w}_n \tag{5.12a}$$

$$\mathbf{y}_{i,n} = \mathbf{H}_{i,n}\mathbf{x}_n + \mathbf{B}_{i,n}\mathbf{x}_n^* + \mathbf{v}_{i,n} \tag{5.12b}$$

or in its augmented representation:

$$\mathbf{x}_n^a = \mathbf{F}_{n-1}^a\mathbf{x}_{n-1}^a + \mathbf{w}_n^a \tag{5.13a}$$

$$\mathbf{y}_{i,n}^a = \mathbf{H}_{i,n}^a\mathbf{x}_n^a + \mathbf{v}_{i,n}^a \tag{5.13b}$$

where $\mathbf{x}_n^a = [\mathbf{x}_n^T, \mathbf{x}_n^H]^T$ and $\mathbf{y}_n^a = [\mathbf{y}_n^T, \mathbf{y}_n^H]^T$, while

$$\mathbf{F}_n^a = \begin{bmatrix} \mathbf{F}_n & \mathbf{A}_n \\ \mathbf{A}_n^* & \mathbf{F}_n^* \end{bmatrix} \quad \text{and} \quad \mathbf{H}_{i,n}^a = \begin{bmatrix} \mathbf{H}_{i,n} & \mathbf{B}_{i,n} \\ \mathbf{B}_{i,n}^* & \mathbf{H}_{i,n}^* \end{bmatrix}$$

Similarly, the augmented covariance matrices of $\mathbf{w}_n^a = [\mathbf{x}_n^T, \mathbf{w}_n^H]^T$ and $\mathbf{v}_{i,n}^a = [\mathbf{v}_{i,n}^T, \mathbf{v}_{i,n}^H]^T$ are given by

$$\mathbf{Q}_n^a = E\{\mathbf{w}_n^a \mathbf{w}_n^{aH}\} = \begin{bmatrix} \mathbf{Q}_n & \mathbf{P}_n \\ \mathbf{P}_n^* & \mathbf{Q}_n^* \end{bmatrix} \tag{5.14}$$

$$\mathbf{R}_{i,n}^a = E\{\mathbf{v}_{i,n}^a \mathbf{v}_{i,n}^{aH}\} = \begin{bmatrix} \mathbf{R}_{i,n} & \mathbf{U}_{i,n} \\ \mathbf{U}_{i,n}^* & \mathbf{R}_{i,n}^* \end{bmatrix} \tag{5.15}$$

Remark #3: For strictly linear systems, $\mathbf{A}_n = \mathbf{0}$ and $\mathbf{B}_{i,n} = \mathbf{0}$, so that the widely linear (augmented) state space model degenerates into a strictly linear one, however, the augmented state space representation is still preferred in order to account for the pseudocovariances (impropriety) of the signals (*cf.* widely linear systems).

To enable collaborative estimation of the state within distributed networks, we employ neighbourhood observation equations comprising of all the neighbourhood observation data, that is

$$\underline{\mathbf{y}}_{i,n} = \underline{\mathbf{H}}_{i,n}\mathbf{x}_n + \underline{\mathbf{B}}_{i,n}\mathbf{x}_n^* + \underline{\mathbf{v}}_{i,n} \tag{5.16}$$

where the conjugate state matrix $\underline{\mathbf{B}}_{i,n} = \left[\mathbf{B}_{i_1,n}^T, \mathbf{B}_{i_2,n}^T, \ldots, \mathbf{B}_{i_M,n}^T\right]^T$, and $\{i_1, i_2, \ldots, i_M\} \in \mathcal{N}_i$. The augmented neighbourhood observation equations can now be written as

$$\underline{\mathbf{y}}_{i,n}^a = \underline{\mathbf{H}}_{i,n}^a \mathbf{x}_n^a + \underline{\mathbf{v}}_{i,n}^a \tag{5.17}$$

with augmented neighbourhood terms defined as

$$\underline{\mathbf{y}}_{i,n}^{a} = \begin{bmatrix} \underline{\mathbf{y}}_{i,n} \\ \underline{\mathbf{y}}_{i,n}^{*} \end{bmatrix}, \quad \underline{\mathbf{H}}_{i,n}^{a} = \begin{bmatrix} \underline{\mathbf{H}}_{i,n} & \underline{\mathbf{B}}_{i,n} \\ \underline{\mathbf{B}}_{i,n}^{*} & \underline{\mathbf{H}}_{i,n}^{*} \end{bmatrix}, \quad \underline{\mathbf{v}}_{i,n}^{a} = \begin{bmatrix} \underline{\mathbf{v}}_{i,n} \\ \underline{\mathbf{v}}_{i,n}^{*} \end{bmatrix} \tag{5.18}$$

The covariance of the augmented noise $\underline{\mathbf{v}}_{i,n}^{a}$ is then defined as

$$\underline{\mathbf{R}}_{i,n}^{a} = E\{\underline{\mathbf{v}}_{i,n}^{a}\underline{\mathbf{v}}_{i,n}^{aH}\} = \begin{bmatrix} \underline{\mathbf{R}}_{i,n} & \underline{\mathbf{U}}_{i,n} \\ \underline{\mathbf{U}}_{i,n}^{*} & \underline{\mathbf{R}}_{i,n}^{*} \end{bmatrix} \tag{5.19}$$

and caters for both the covariances $E\{\mathbf{v}_{i,n}\mathbf{v}_{i,n}^{H}\}$ and cross-correlations $E\{\mathbf{v}_{i,n}\mathbf{v}_{k,n}^{H}\}$, $i \neq k$ of the nodal observation noises through the covariance matrix $\underline{\mathbf{R}}_{i,n}$, and the pseudocovariances $E\{\mathbf{v}_{i,n}\mathbf{v}_{i,n}^{T}\}$ and cross-pseudocorrelations $E\{\mathbf{v}_{i,n}\mathbf{v}_{k,n}^{T}\}$ through the pseudocovariance matrix $\underline{\mathbf{U}}_{i,n}$. Finally, the augmented diffused state estimate becomes

$$\widehat{\mathbf{x}}_{i,n|n}^{a} = \sum_{k \in \mathcal{N}_i} c_{k,i}\widehat{\underline{\mathbf{x}}}_{k,n|n}^{a} \tag{5.20}$$

and represents a weighted average of the augmented (neighbourhood) state estimates. The proposed distributed augmented complex Kalman filter (D-ACKF), based on the widely linear state space model, is summarised in Algorithm 10.

Remark #4: For strictly linear systems ($\mathbf{A}_n = \mathbf{0}$ and $\mathbf{B}_{i,n} = \mathbf{0}$ for all n and i) with improper state and observation noises ($\mathbf{P}_n = \mathbf{0}$ and $\mathbf{U}_{i,n} = \mathbf{0}$ for all n and i), the D-ACKF and D-CKF algorithms are equivalent, in the sense that they yield identical state estimates for all time instants n.

However, if any of these conditions are not met, the D-ACKF assumes a more general form than the D-CKF. This can be illustrated as per the analysis in Chapter 3, which shows the advantages of the augmented complex Kalman filter (ACKF) over the conventional (strictly linear) complex Kalman filter (CKF) for the non-distributed case.

Remark #5: When the nodes are subject to uncorrelated observation noises, the

Algorithm 10: The D-ACKF

Initialisation: For each node $i = 1, 2, \ldots, N$

$$\begin{aligned} \widehat{\mathbf{x}}^a_{i,0|0} &= \left[E\{\mathbf{x}_0\}^T, E\{\mathbf{x}_0\}^H\right]^T \\ \mathbf{M}^a_{i,0|0} &= E\left\{(\mathbf{x}^a_0 - \widehat{\mathbf{x}}^a_{i,0|0})(\mathbf{x}^a_0 - \widehat{\mathbf{x}}^a_{i,0|0})^{aH}\right\} \end{aligned}$$

For every time instant $n = 1, 2, \ldots$

– Evaluate at each node $i = 1, 2, \ldots, N$

$$\widehat{\mathbf{x}}^a_{i,n|n-1} = \mathbf{F}^a_{n-1}\widehat{\mathbf{x}}^a_{i,n-1|n-1} \tag{5.21}$$

$$\mathbf{M}^a_{i,n|n-1} = \mathbf{F}^a_{n-1}\mathbf{M}^a_{i,n-1|n-1}\mathbf{F}^{aH}_{n-1} + \mathbf{Q}^a_n \tag{5.22}$$

$$\mathbf{G}^a_{i,n} = \mathbf{M}^a_{i,n|n-1}\underline{\mathbf{H}}^{aH}_{i,n}\left(\underline{\mathbf{H}}^a_{i,n}\mathbf{M}^a_{i,n|n-1}\underline{\mathbf{H}}^{aH}_{i,n} + \underline{\mathbf{R}}^a_{i,n}\right)^{-1} \tag{5.23}$$

$$\underline{\widehat{\mathbf{x}}}^a_{i,n|n} = \widehat{\mathbf{x}}^a_{i,n|n-1} + \mathbf{G}^a_{i,n}\left(\underline{\mathbf{y}}^a_{i,n} - \underline{\mathbf{H}}^a_{i,n}\widehat{\mathbf{x}}^a_{i,n|n-1}\right) \tag{5.24}$$

$$\mathbf{M}^a_{i,n|n} = (\mathbf{I} - \mathbf{G}^a_{i,n}\underline{\mathbf{H}}^a_{i,n})\mathbf{M}^a_{i,n|n-1} \tag{5.25}$$

– For every node i, compute the diffusion update as

$$\widehat{\mathbf{x}}^a_{i,n|n} = \sum\nolimits_{k\in\mathcal{N}_i} c_{k,i}\underline{\widehat{\mathbf{x}}}^a_{k,n|n} \tag{5.26}$$

information form of the D-ACKF, given in Algorithm 11, can be utilised to cater for the propriety of the signals without accounting for observation noise correlations at different nodes.

Further, depending on the correlation between the observation noises, different nodes in the distributed network can switch between the general D-ACKF in Algorithm 10 and the information form D-ACKF in Algorithm 11.

5.2 Analysis

5.2.1 Duality Analysis

Owing to the isomorphism between augmented complex vectors and bivariate real vectors, and the duality analysis for stochastic gradient filters [19], the D-ACKF algorithm has a dual bivariate distributed real valued Kalman filter (D-RKF) which can be used to reduce its computational complexity.

A complex vector $\mathbf{z} = \mathbf{z}_r + j\mathbf{z}_i \in \mathbb{C}^q$ has a composite bivariate real representation

Algorithm 11: The D-ACKF Information Form

Initialisation: For each node $i = 1, 2, \ldots, N$

$$\begin{aligned}\widehat{\mathbf{x}}^a_{i,0|0} &= \big[E\{\mathbf{x}_0\}^T, E\{\mathbf{x}_0\}^H\big]^T \\ \mathbf{M}^a_{i,0|0} &= E\big\{(\mathbf{x}^a_0 - \widehat{\mathbf{x}}^a_{i,0|0})(\mathbf{x}^a_0 - \widehat{\mathbf{x}}^a_{i,0|0})^{aH}\big\}\end{aligned}$$

For every time instant $n = 1, 2, \ldots$

– Evaluate at each node $i = 1, 2, \ldots, N$

$$\widehat{\mathbf{x}}^a_{i,n|n-1} = \mathbf{F}^a_{n-1}\widehat{\mathbf{x}}^a_{i,n-1|n-1} \tag{5.29}$$

$$\mathbf{M}^a_{i,n|n-1} = \mathbf{F}^a_{n-1}\mathbf{M}^a_{i,n-1|n-1}\mathbf{F}^{aH}_{n-1} + \mathbf{Q}^a_n \tag{5.30}$$

$$\mathbf{S}^a_{i,n} = \sum\nolimits_{k\in\mathcal{N}_i} \mathbf{H}^{aH}_{k,n}(\mathbf{R}^a_{k,n})^{-1}\mathbf{H}^a_{k,n} \tag{5.31}$$

$$\mathbf{r}^a_{i,n} = \sum\nolimits_{k\in\mathcal{N}_i} \mathbf{H}^{aH}_{k,n}(\mathbf{R}^a_{k,n})^{-1}\mathbf{y}^a_{k,n} \tag{5.32}$$

$$(\mathbf{M}^a_{i,n|n})^{-1} = (\mathbf{M}^a_{i,n|n-1})^{-1} + \mathbf{S}^a_{i,n} \tag{5.33}$$

$$\widehat{\chi}^a_{i,n|n} = \widehat{\mathbf{x}}^a_{i,n|n-1} + \mathbf{M}^a_{i,n|n}\big(\mathbf{r}^a_{i,n} - \mathbf{S}^a_{i,n}\widehat{\mathbf{x}}^a_{i,n|n-1}\big) \tag{5.34}$$

– For every node i, compute the diffusion update as

$$\widehat{\mathbf{x}}^a_{i,n|n} = \sum\nolimits_{k\in\mathcal{N}_i} c_{k,i}\widehat{\chi}^a_{i,n|n} \tag{5.35}$$

in $\mathbb{R}^{2q}$ of the form

$$\mathbf{z}^a = \begin{bmatrix}\mathbf{z}\\ \mathbf{z}^*\end{bmatrix} = \underbrace{\begin{bmatrix}\mathbf{I} & j\mathbf{I}\\ \mathbf{I} & -j\mathbf{I}\end{bmatrix}}_{\equiv\mathbf{J}_\mathbf{z}}\underbrace{\begin{bmatrix}\mathbf{z_r}\\ \mathbf{z_i}\end{bmatrix}}_{=\mathbf{z}^r} \tag{5.27}$$

where $\mathbf{I}$ is the identity matrix (with appropriate dimensions), and the invertible orthogonal mapping[2] $\mathbf{J}_\mathbf{z} : \mathbb{C}^{2q} \to \mathbb{R}^{2q}$ is such that $\mathbf{J}_\mathbf{z}^{-1} = \frac{1}{2}\mathbf{J}_\mathbf{z}^H$ [20][21]. Based on this isomorphism, the real bivariate state space corresponding to the augmented complex state space in (5.13) is given by

$$\begin{aligned}\mathbf{x}^r_n &= \mathbf{F}^r_{n-1}\mathbf{x}^r_{n-1} + \mathbf{w}^r_n \\ \mathbf{y}^r_n &= \mathbf{H}^r_n\mathbf{x}^r_n + \mathbf{v}^r_n\end{aligned} \tag{5.28a}$$

where $\mathbf{x}^r_n = \mathbf{J}_\mathbf{x}^{-1}\mathbf{x}^a_n$, $\mathbf{y}^r_n = \mathbf{J}_\mathbf{y}^{-1}\mathbf{y}^a_n$, $\mathbf{F}^r_{n-1} = \mathbf{J}_\mathbf{x}^{-1}\mathbf{F}^a_{n-1}\mathbf{J}_\mathbf{x}$, $\mathbf{H}^r_n = \mathbf{J}_\mathbf{y}^{-1}\mathbf{H}^a_n\mathbf{J}_\mathbf{x}$, $\mathbf{w}^r_n = \mathbf{J}_\mathbf{x}^{-1}\mathbf{w}^a_n$ and

[2]For a vector $\mathbf{z} \in \mathbb{C}^q$, the corresponding orthogonal matrix $\mathbf{J}_\mathbf{z}$ takes dimension $2q \times 2q$.

$\mathbf{v}_n^r = \mathbf{J}_{\mathbf{y}}^{-1}\mathbf{v}_n^a$. In the same vein, the real valued covariance matrices of $\mathbf{w}_n^r$ and $\mathbf{v}_n^r$ take the corresponding forms

$$\begin{aligned}\mathbf{Q}_n^r &= E\{\mathbf{w}_n^r\mathbf{w}_n^{rH}\} = \mathbf{J}_{\mathbf{x}}^{-1}\mathbf{Q}_n^a\mathbf{J}_{\mathbf{x}}^{-H}\\ \mathbf{R}_n^r &= E\{\mathbf{v}_n^r\mathbf{v}_n^{rH}\} = \mathbf{J}_{\mathbf{y}}^{-1}\mathbf{R}_n^a\mathbf{J}_{\mathbf{y}}^{-H}\end{aligned}$$

while the real valued counterpart of (5.17) is given by

$$\underline{\mathbf{y}}_{i,n}^r = \underline{\mathbf{H}}_{i,n}^r\mathbf{x}_n^r + \underline{\mathbf{v}}_{i,n}^r \tag{5.36}$$

with $\underline{\mathbf{y}}_n^r = \mathbf{J}_{\underline{\mathbf{y}}}^{-1}\underline{\mathbf{y}}_n^a$, $\underline{\mathbf{H}}_n^r = \mathbf{J}_{\underline{\mathbf{y}}}^{-1}\underline{\mathbf{H}}_n^a\mathbf{J}_{\mathbf{x}}$ and $\underline{\mathbf{v}}_n^r = \mathbf{J}_{\underline{\mathbf{y}}}^{-1}\underline{\mathbf{v}}_n^a$. Finally, the covariance matrix of $\underline{\mathbf{v}}_n^r$ is defined as

$$\underline{\mathbf{R}}_n^r = E\{\underline{\mathbf{v}}_n^r\underline{\mathbf{v}}_n^{rH}\} = \mathbf{J}_{\underline{\mathbf{y}}}^{-1}\underline{\mathbf{R}}_n^a\mathbf{J}_{\underline{\mathbf{y}}}^{-H}$$

The duality between the D-ACKF and the D-RKF is established through the following relationships:

$$\begin{aligned}\widehat{\mathbf{x}}_{i,n|n-1}^r &= \mathbf{J}_{\mathbf{x}}^{-1}\widehat{\mathbf{x}}_{i,n|n-1}^a\\ \mathbf{M}_{i,n|n-1}^r &= \mathbf{J}_{\mathbf{x}}^{-1}\mathbf{M}_{i,n|n-1}^a\mathbf{J}_{\mathbf{x}}^{-H}\\ \mathbf{G}_{i,n}^r &= \mathbf{J}_{\mathbf{x}}^{-1}\mathbf{G}_{i,n}^a\mathbf{J}_{\underline{\mathbf{y}}}\\ \underline{\widehat{\mathbf{x}}}_{i,n|n}^r &= \mathbf{J}_{\mathbf{x}}^{-1}\underline{\widehat{\mathbf{x}}}_{i,n|n}^a\\ \mathbf{M}_{i,n|n}^r &= \mathbf{J}_{\mathbf{x}}^{-1}\mathbf{M}_{i,n|n}^a\mathbf{J}_{\mathbf{x}}^{-H}\\ \widehat{\mathbf{x}}_{i,n|n}^r &= \mathbf{J}_{\mathbf{x}}^{-1}\widehat{\mathbf{x}}_{i,n|n}^a\end{aligned}$$

Therefore, the D-ACKF and D-RKF effectively implement the same state space model, but operate in the complex and real domains, respectively. Generally speaking, for systems naturally defined in the complex domain, it is desirable to keep the computations in the original complex domain in order to facilitate understanding of the signal transformations, together with benefiting from the well defined notions of phase and circularity.

5.2.2 Mean And Mean Square Analysis

For generality in analysis, we consider augmented complex variables: let $\underline{\mathbf{e}}^a_{i,n|n} = \mathbf{x}^a_n - \underline{\widehat{\mathbf{x}}}^a_{i,n|n}$ denote the local (non-diffused) error at node $i \in [1, N]$, $\mathbf{e}^a_{i,n|n-1} = \mathbf{x}^a_n - \widehat{\mathbf{x}}^a_{i,n|n-1}$ the prediction error, and $\mathbf{e}^a_{i,n|n} = \mathbf{x}^a_n - \widehat{\mathbf{x}}^a_{i,n|n}$ the diffused error. The difference between the true state in (5.13a) and the predicted state estimate in (5.24) now becomes

$$\begin{aligned}
\underline{\mathbf{e}}^a_{i,n|n} &= \mathbf{x}^a_n - \widehat{\mathbf{x}}^a_{i,n|n-1} + \mathbf{G}^a_{i,n}\big(\underline{\mathbf{y}}^a_{i,n} - \underline{\mathbf{H}}^a_{i,n}\widehat{\mathbf{x}}^a_{i,n|n-1}\big) \\
&= \mathbf{e}^a_{i,n|n-1} + \mathbf{G}^a_{i,n}\big(\underline{\mathbf{H}}^a_{i,n}\mathbf{x}^a_n + \underline{\mathbf{v}}^a_{i,n} - \underline{\mathbf{H}}^a_{i,n}\widehat{\mathbf{x}}^a_{i,n|n-1}\big) \\
&= \mathbf{e}^a_{i,n|n-1} + \mathbf{G}^a_{i,n}\big(\underline{\mathbf{H}}^a_{i,n}\mathbf{e}^a_{i,n|n-1} + \underline{\mathbf{v}}^a_{i,n}\big) \\
&= \big(\mathbf{I} + \mathbf{G}^a_{i,n}\underline{\mathbf{H}}^a_{i,n}\big)\mathbf{e}^a_{i,n|n-1} + \mathbf{G}^a_{i,n}\underline{\mathbf{v}}^a_{i,n}
\end{aligned} \tag{5.37}$$

Likewise, the difference between (5.13a) and (5.21) is given by

$$\mathbf{e}^a_{i,n|n-1} = \mathbf{F}^a_{n-1}\mathbf{e}^a_{i,n-1|n-1} + \mathbf{w}^a_n \tag{5.38}$$

while the diffused state estimation error can be expressed as

$$\mathbf{e}^a_{i,n|n} = \mathbf{x}^a_n - \sum_{k\in\mathcal{N}_i} c_{k,i}\underline{\widehat{\mathbf{x}}}^a_{k,n|n} = \sum_{k\in\mathcal{N}_i} c_{k,i}\underline{\mathbf{e}}^a_{k,n|n} \tag{5.39}$$

Substituting (5.37) and (5.38) into (5.39) and using $\mathbf{M}^a_{k,n|n}(\mathbf{M}^a_{k,n|n-1})^{-1} = \mathbf{I} - \mathbf{G}^a_{k,n}\underline{\mathbf{H}}^a_{k,n}$, we have

$$\begin{aligned}
\mathbf{e}^a_{i,n|n} &= \sum_{k\in\mathcal{N}_i} c_{k,i}\Big[\big(\mathbf{I} + \mathbf{G}^a_{k,n}\underline{\mathbf{H}}^a_{k,n}\big)\mathbf{F}^a_{n-1}\mathbf{e}^a_{k,n-1|n-1} + \big(\mathbf{I} + \mathbf{G}^a_{k,n}\underline{\mathbf{H}}^a_{k,n}\big)\mathbf{w}^a_n + \mathbf{G}^a_{k,n}\underline{\mathbf{v}}^a_{k,n}\Big] \\
&= \sum_{k\in\mathcal{N}_i} c_{k,i}\Big[\mathbf{M}^a_{k,n|n}(\mathbf{M}^a_{k,n|n-1})^{-1}\mathbf{F}^a_{n-1}\mathbf{e}^a_{k,n-1|n-1} + \mathbf{M}^a_{k,n|n}(\mathbf{M}^a_{k,n|n-1})^{-1}\mathbf{w}^a_n + \mathbf{G}^a_{k,n}\underline{\mathbf{v}}^a_{k,n}\Big]
\end{aligned} \tag{5.40}$$

Upon taking the statistical expectation, the recursion in (5.40) leads to a closed form expression for the mean error of the D-ACKF algorithm, given by

$$E\{\mathbf{e}^a_{i,n|n}\} = \sum_{k\in\mathcal{N}_i} c_{k,i}\mathbf{M}^a_{k,n|n}(\mathbf{M}^a_{k,n|n-1})^{-1}\mathbf{F}^a_{n-1}E\{\mathbf{e}^a_{k,n-1|n-1}\} = \mathbf{0} \tag{5.41}$$

Remark #6: Equation (5.41) demonstrates that the D-ACKF is an unbiased estimator of general complex processes, exhibiting both proper and improper statistics.

To derive the mean square error for the D-ACKF, we shall make the following assumptions, commonly used in the analysis of distributed state space estimators.

Assumption #1: *Convergence.* All the nodes (local Kalman filters) converge to the same state value by using their neighbourhood data, that is, there are no faulty nodes in the system. This can be restated as

$$\lim_{n\to\infty}\widehat{\mathbf{x}}_{i,n|n} = \lim_{n\to\infty}\widehat{\mathbf{x}}_{k,n|n} \quad \text{for all } i \text{ and } k \tag{5.42}$$

Assumption #2: *Time invariance.* The state space model (3.5) is time invariant, that is, $\mathbf{F}_n = \mathbf{F}$, $\mathbf{H}_{i,n} = \mathbf{H}$, $\mathbf{Q}_n = \mathbf{Q}$ and $\mathbf{R}_{i,n} = \mathbf{R}$, and the state transition matrix $\mathbf{F}$ is stable. This is a standard assumption for the steady-state analysis of Kalman filters.

It then follows that $\lim_{n\to\infty}\mathbf{M}^a_{i,n|n} = \mathbf{M}^a$ for $i \in \{1,\ldots,N\}$, that is, the matrix $\mathbf{M}^a_{i,n|n}$ is also time invariant at steady state. Next, we define the following terms for

convenience of notation:

$$\mathcal{E}_{i,n|n} = \left[\mathbf{e}_{i_1,n|n}^{aT}, \mathbf{e}_{i_2,n|n}^{aT}, \ldots, \mathbf{e}_{i_M,n|n}^{aT}\right]^T \in \mathbb{C}^{2ML}$$

$$\underline{\mathcal{W}}_{i,n} = \left[\mathbf{w}_n^{aT}, \mathbf{w}_n^{aT}, \ldots, \mathbf{w}_n^{aT}\right]^T \in \mathbb{C}^{2ML}$$

$$\underline{\mathcal{V}}_{i,n} = \left[\underline{\mathbf{v}}_{i_1,n}^{aT}, \underline{\mathbf{v}}_{i_2,n}^{aT}, \ldots, \underline{\mathbf{v}}_{i_M,n}^{aT}\right]^T$$

$$\underline{\mathcal{A}}_{i,n} = \begin{bmatrix} c_{i_1,i}\mathbf{F}_{n-1}^{aT}(\mathbf{M}_{i_1,n|n-1}^{a})^{-T}\mathbf{M}_{i_1,n|n}^{aT} \\ c_{i_2,i}\mathbf{F}_{n-1}^{aT}(\mathbf{M}_{i_2,n|n-1}^{a})^{-T}\mathbf{M}_{i_2,n|n}^{aT} \\ \vdots \\ c_{i_M,i}\mathbf{F}_{n-1}^{aT}(\mathbf{M}_{i_M,n|n-1}^{a})^{-T}\mathbf{M}_{i_M,n|n}^{aT} \end{bmatrix}^T$$

$$\underline{\mathcal{B}}_{i,n} = \begin{bmatrix} c_{i_1,i}(\mathbf{M}_{i_1,n|n-1}^{a})^{-T}\mathbf{M}_{i_1,n|n}^{aT} \\ c_{i_2,i}(\mathbf{M}_{i_2,n|n-1}^{a})^{-T}\mathbf{M}_{i_2,n|n}^{aT} \\ \vdots \\ c_{i_M,i}(\mathbf{M}_{i_M,n|n-1}^{a})^{-T}\mathbf{M}_{i_M,n|n}^{aT} \end{bmatrix}^T$$

$$\underline{\mathcal{G}}_{i,n} = \left[c_{i_1,i}\mathbf{G}_{i_1,n}^{a}, c_{i_2,i}\mathbf{G}_{i_2,n}^{a}, \ldots, c_{i_M,i}\mathbf{G}_{i_M,n}^{a}\right]$$

where $\{i_1, i_2, \ldots, i_M\} \in \mathcal{N}_i$, and $M = |\mathcal{N}_i|$ is the number of nodes in the neighbourhood $\mathcal{N}_i$. Based on (5.40), the mean square error $\mathbf{\Sigma}_{i,n}^{a} = E\{\mathbf{e}_{i,n|n}^{a}\mathbf{e}_{i,n|n}^{aH}\}$ at the node i then becomes

$$\mathbf{\Sigma}_{i,n}^{a} = \underline{\mathcal{A}}_{i,n}\mathcal{M}_{i,n-1}\underline{\mathcal{A}}_{i,n}^{H} + \underline{\mathcal{B}}_{i,n}\underline{\mathcal{Q}}_{i,n}\underline{\mathcal{B}}_{i,n}^{H} + \underline{\mathcal{G}}_{i,n}\underline{\mathcal{R}}_{i,n}\underline{\mathcal{G}}_{i,n} \tag{5.43}$$

where $\mathcal{M}_{i,n} = E\{\mathcal{E}_{i,n|n}\mathcal{E}_{i,n|n}^{H}\}$ is the neighbourhood error covariance matrix, $\underline{\mathcal{Q}}_{i,n} = E\{\underline{\mathcal{W}}_{i,n}\underline{\mathcal{W}}_{i,n}^{H}\}$ and $\underline{\mathcal{R}}_{i,n} = E\{\underline{\mathcal{V}}_{i,n}\underline{\mathcal{V}}_{i,n}^{H}\}$.

Remark #7: Under Assumption #2, the covariance matrices $\underline{\mathcal{Q}}_{i,n} = \underline{\mathcal{Q}}_{i}$ and $\underline{\mathcal{R}}_{i,n} = \underline{\mathcal{R}}_{i}$ are time invariant, while as $n \to \infty$ the terms $\underline{\mathcal{A}}_{i,n} = \underline{\mathcal{A}}_{i}$, $\underline{\mathcal{B}}_{i,n} = \underline{\mathcal{B}}_{i}$, $\underline{\mathcal{G}}_{i,n} = \underline{\mathcal{G}}_{i}$ also become time invariant. Then under Assumption #1, that is, provided all the nodes in the network converge to the same steady state value, the remaining error covariance term $\mathcal{M}_{i,n}$ also converges.

Further, observe that, based on (5.39), the MSE can alternatively be expressed as

$$\mathbf{\Sigma}_{i,n}^{a} = \sum_{j \in \mathcal{N}_i} \sum_{k \in \mathcal{N}_i} c_{j,i} c_{k,i} \underline{\mathbf{\Gamma}}_{jk,n}^{a} \tag{5.44}$$

where $\underline{\mathbf{\Gamma}}_{jk,n}^{a} = E\{\underline{\mathbf{e}}_{j,n|n}^{a} \underline{\mathbf{e}}_{k,n|n}^{aH}\}$ is the cross-correlation matrix between the neighbourhood errors.

Remark #8: Given that $\sum_{j \in \mathcal{N}_i} \sum_{k \in \mathcal{N}_i} c_{j,i} c_{k,i} = 1$, and the MSE at any node i, $\mathbf{\Sigma}_{i,n}^{a}$, is a weighted average of the MSEs of the nodes in its neighbourhood, then $\mathbf{\Sigma}_{i,n}^{a}$ is upper bounded by the MSE of the node in the neighbourhood of node i with the worst MSE, that is

$$\mathrm{tr}(\mathbf{\Sigma}_{i,n}^{a}) \leq \max_{k \in \mathcal{N}_i} \{\mathrm{tr}(\underline{\mathbf{\Gamma}}_{kk,n}^{a})\} \tag{5.45}$$

where $\mathrm{tr}(\cdot)$ is the matrix trace operator.

It then follows that at any time instant n, the upper bound for the average MSE of the whole distributed network is the MSE of the node with the highest MSE in the network.

From Remarks #6, #7 and #8, the D-ACKF converges both in the mean and mean square sense, hence it is a consistent estimator, while its MSE performance is upper bounded by the worst performing node in the network.

5.3 Application Examples

To illustrate the advantages of the widely linear D-ACKF over its strictly linear D-CKF counterpart, the following case studies were conducted: A) filtering of a noisy complex-valued autoregressive process; B) estimating and tracking the position of a projectile in two dimensions.

5.3.1 Filtering an Autoregressive Process

Consider a distributed network consisting of $N = 10$ nodes (see Figure 5.2), used for filtering the complex autoregressive (AR) process defined as

$$z_n = 1.2z_{n-1} - 0.8z_{n-2} + u_n$$

For rigour, u_n is a improper white complex Gaussian driving noise with variance $E\{|u_n|^2\} = 2$ and a varying pseudovariance $E\{u_n^2\}$. For each node i, the observation equation was a noisy measurement of the autoregressive output, that is improper white complex Gaussian driving noise with variance $E\{|u_n|^2\} = 2$ and a varying pseudovariance $E\{u_n^2\}$. For each node i, the observation equation was a noisy measurement of the autoregressive output, that is

$$y_{i,n} = z_n + v_{i,n}$$

where $v_{i,n}$ is the complex Gaussian white observation noise associated with node i, while the variances, pseudovariances and cross-correlations of the observation noises were $R_{i,n} = E\{|v_{i,n}|^2\} = 4 + 1/\sqrt{i}$, $U_{i,n} = E\{v_{i,n}^2\}$ and $R_{ik,n} = E\{v_{i,n}v_{k,n}^*\} = 4$ for $i, k \in \{1, 2, \ldots, N\}$ and $i \neq k$. Observe that the nodes in network experience correlated observation noises with different variances, modelled through the term $1/\sqrt{i}$ in the expression for $R_{i,n}$.

In the simulations, we used the ratio of the magnitude of pseudocovariance to covariance, that is $\eta_u = |E\{u^2\}|/E\{|u|^2\}$, as a measure for the degree of circularity of a (zero-mean) complex Gaussian signal $u = u_r + ju_i$, where a signal is circular for $\eta_u = 0$ and maximally noncircular for $\eta_u = 1$. The average mean square errors (MSEs) of all the nodes were used for a quantitative assessment of performance in a nearest neighbour diffusion scheme, which is as follows [44]. Let $|\mathcal{N}_k|$ denote the number of a neighbours (including itself) of node k; to compute the diffused state estimate for node i, the weight

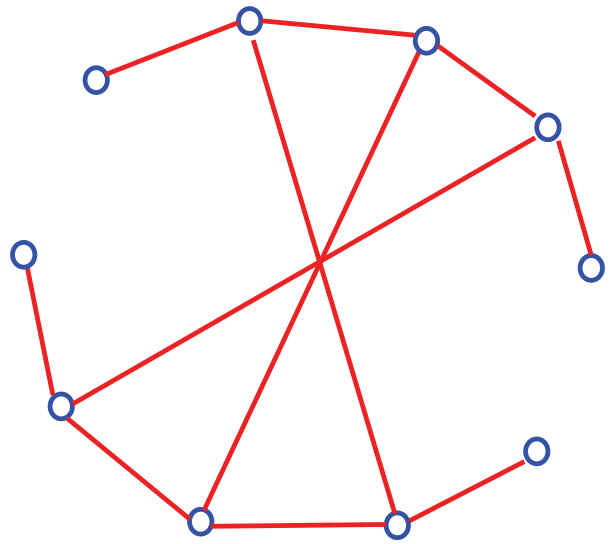


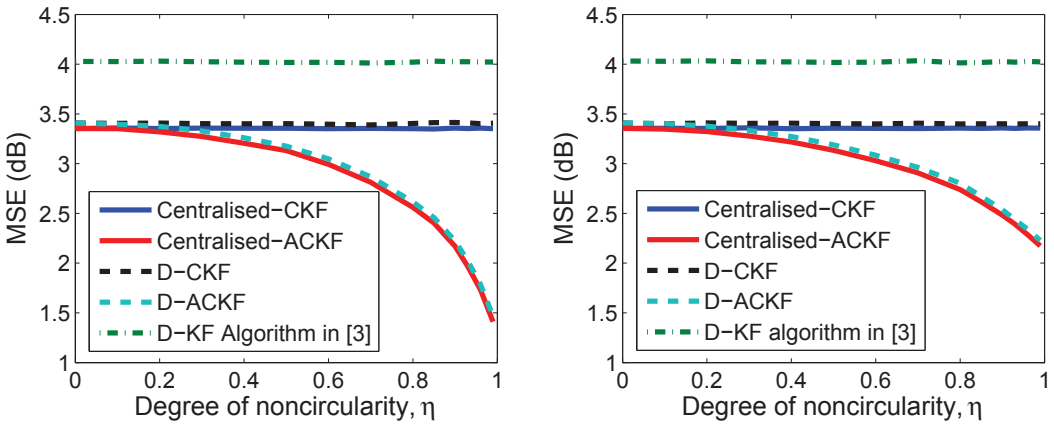
Figure 5.2: A distributed network with $N = 10$ nodes used in the simulations.

associated with a neighbour k is proportional to $|\mathcal{N}_k|$, that is

$$c_{k,i} = \begin{cases} |\mathcal{N}_k|/\alpha_i & \text{if } k \in \mathcal{N}_i \\ 0 & \text{otherwise} \end{cases}$$

where $\alpha_i = \sum_{k\in\mathcal{N}_i} |\mathcal{N}_k|$ is a normalisation parameter for node i which ensures that $\sum_{k\in\mathcal{N}_i} c_{k,i} = 1$.

Figure 5.3 compares the steady state performance of the diffusion Kalman filter in [44] (Algorithm 2), D-CKF and D-ACKF algorithms, along with the centralised versions of the D-CKF and D-ACKF (Centralised-CKF and Centralised-ACKF), with access to the observation data from all the nodes at each time instant. Figure 5.3a illustrates the results for circular observation noises ($U_{i,n} = E\{v_{i,n}^2\} = 0$ for $i = 1, 2, \ldots, N$) and a state (driving) noise with various degrees of noncircularity, whereas the results for a noncircular observation noise with a circular state noise ($P_n = 0$) are shown in Figure 5.3b. The variances of the state and observation noises were kept constant throughout, and only their pseudocovariances (degree of circularity) where changed. The results illustrate that for second order circular (proper) state and observation noises ($\eta_w = 0$ and $\eta_{v_i} = 0$), the



(a) Noncircular state noise (b) Noncircular observation noise

Figure 5.3: Steady state performance comparison for filtering the AR(2) process in the cases of: (a) circular observation noises and a noncircular driving noise with varying degrees of noncircularity; (b) circular state noise and noncircular observation noises with varying degrees of noncircularity, whereby all nodes have same degree of observation noise noncircularity.

strictly linear D-CKF and widely linear D-ACKF algorithms have identical performances, conforming with the analysis and Remark #3, while for proper noises ($\eta_w \neq 0$ and $\eta_{v_i} \neq 0$) the D-ACKF offered superior performance, as it catered for the pseudocovariances. Moreover, D-ACKF had decreasing MSE for an increasing degree of noise noncircularity, while D-CKF was unaffected by changes in the noncircularity of the noises, as it is not designed to recognise improper signals.

The performance comparison between the Centralised-CKF and centralised-ACKF algorithms also shows a similar trend, with centralised-ACKF offering better performance for noncircular signals. The D-CKF and D-ACKF algorithms outperformed the diffusion Kalman filter in [44] (Algorithm 2), because they cater for the cross-correlations between the the observation noises ($R_{ik,n} = E\{v_{i,n}v_{k,n}^*\}$), and only marginally underperformed compared with their centralised counterparts. Observe that for uncorrelated nodal observation noises with circular state and observation noises, the D-CKF, D-ACKF and the diffusion Kalman filter in [44] will have identical performances.

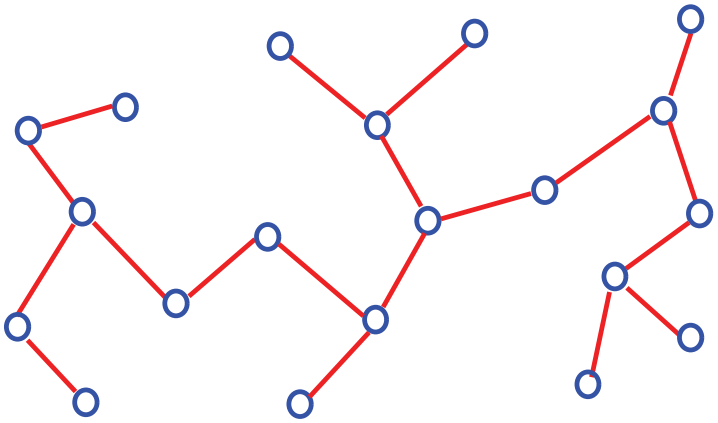


Figure 5.4: A distributed network with $N = 20$ nodes used in the simulations.

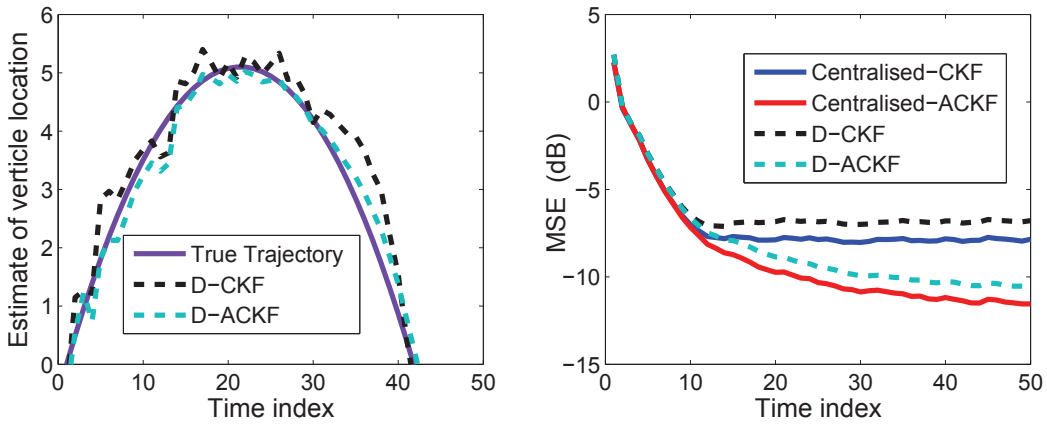
5.3.2 Projectile Tracking

We next considered the problem of estimating and tracking the position of a projectile in two dimensions, where only noisy measurements of its position are available. Let (x_n, y_n) and $(\dot{x}_n, \dot{y}_n)$ denote the position and velocity vectors of the projectile at time instant n, respectively, then the corresponding complex valued distributed state space model for the system is given by

$$\begin{aligned} \mathbf{x}_n &= \mathbf{F}\mathbf{x}_{n-1} - j\mathbf{K}g + \mathbf{K}w_n \\ z_{i,n} &= \mathbf{h}\mathbf{x}_n + v_{i,n} \end{aligned}$$

where:

- $\mathbf{x}_n = \begin{bmatrix} x_n + jy_n & \dot{x}_n + j\dot{y}_n \end{bmatrix}^T$ is the projectile state vector, and $g = 9.8m/s^2$ is the gravitational acceleration;



(a) Tracking Example (b) Performance Comparison

Figure 5.5: projectile tracking simulations: (a) Average performance (of all the nodes) for a trial run of the diffusion algorithms; (b) Transient performance of the centralised and diffusion algorithms.

- $\mathbf{F}$, $\mathbf{K}$ and $\mathbf{H}$ are time-invariant matrices and vectors defined as

$$\mathbf{F} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}, \quad \mathbf{K} = \begin{bmatrix} \frac{T^2}{2} \\ T \end{bmatrix} \quad \text{and} \quad \mathbf{h} = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

  where T is the sampling interval;

- $z_{i,n}$ is the observation at node i;

- w_n is the zero mean state noise (used to account for modelling inaccuracies), whereas, $\mathbf{v}_{i,n}$ is the zero mean observation noise at node i.

To illustrate the benefits of the proposed distributed algorithms, we considered a scenario with $N = 20$ nodes connected as in Figure 5.4, where a projectile was launched into the air with an initial velocity $(20, 10)$m/s, from location $(0, 0)$m. The sampling interval was set to $T = 0.05$s, and the mean square errors (MSEs) of the different algorithms were computed by averaging 1000 independent trials. The state and observation noises were noncircular Gaussian random processes, both with a degree of noncircularity of $\eta = 0.85$, and their respective distributions were defined as

$$w_n \sim \mathcal{N}(0, 5) \qquad v_{i,n} \sim \mathcal{N}(0, 1 + 2\sqrt{i})$$

where the observation noise cross-correlations were set to $E\{v_{i,n}v_{k,n}^*\} = 1$ for $i, k \in \{1, 2, \ldots, N\}$ and $i \neq k$.

A sample simulation run for the two diffusion algorithms D-CKF and D-ACKF is shown in Figure 5.5a, while Figure 5.5b illustrates the enhanced performance of the widely linear D-ACKF in estimating the projectile location compared with the strictly linear D-CKF. The (strictly linear) centralised-CKF and (widely linear) centralised-ACKF were able to outperform their distributed counterparts D-CKF and D-ACKF, respectively, due to their use of the full network observation data. However, this requires a high communication overhead, compared with that required for the diffusion algorithms, such as the scenario in the sparsely connected network shown in Figure 5.4.

5.4 Conclusions

Distributed complex state space estimation has been addressed in the context of collaborative networks for the general case of improper state and observation models and noises. The distributed (widely linear) augmented complex Kalman filter (D-ACKF) algorithm has been introduced for the sequential state estimation of both proper and improper signal distributions, within a framework which caters for correlated nodal observation noises. The analysis and simulations show that it provides unbiased and consistent estimates, and enhanced performance for improper signals, compared with the distributed complex Kalman filter (D-CKF). Simulations using both proper and improper signals illustrate the performance gains of the proposed solutions.

Chapter 6

Exploiting Sparsity in Widely Linear Estimation

The Kalman filter is powerful tool for dealing with dynamic systems, however, in a numerous applications it is the gradient descent based least mean square (LMS) algorithm that is preferred due to its simplicity and low computational cost. For adaptive algorithms such as the LMS, where the data pairs $\mathbf{x}$ and $\mathbf{y}$ are explicitly given, the aim is to estimate the coefficients of the underlying system [18], which for complex systems can take on strictly and widely linear forms. Widely linear algorithms are general and cater for both strictly or widely linear system models, that is, the conjugate coefficient converges to zero when the underlying transfer function is strictly linear. However, for the same steady-state performance, the convergence rate of the widely linear (augmented) complex LMS (ACLMS) algorithm is slower than its strictly linear counterpart the complex LMS (CLMS) [58].

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

6.1 Background

6.1.1 Complex Least Mean Square (CLMS)

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

$$d_k = \mathbf{x}_k^H \mathbf{w}^o + q_k \tag{6.1}$$

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

$$J = E\{e_k e_k^*\} = E\{|e_k|^2\} \tag{6.2}$$

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

$$y_k = \mathbf{x}_k^H \mathbf{w} \tag{6.3}$$

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

$$\hat{\mathbf{w}} = E\{\mathbf{x}_k \mathbf{x}_k^H\}^{-1} E\{d_k \mathbf{x}_k\} \tag{6.4}$$

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

$$J_k = e_k e_k^* = |e_k|^2 = |d_k - \mathbf{x}_k^H \mathbf{w}|^2 \tag{6.5}$$

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

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \mu \nabla_{\mathbf{w}} J_k \big|_{\mathbf{w}=\mathbf{w}_k} \tag{6.6}$$

where μ is the adaption gain, and $\nabla_{\mathbf{w}} J_k\big|_{\mathbf{w}=\mathbf{w}_k} = -e_k \mathbf{x}_k$ the derivative of J_k with respect to the weight vector [20]. The CLMS is a recursive algorithm, and can be summarised as

$$y_k = \mathbf{x}_k^H \mathbf{w}_k \tag{6.7}$$

$$e_k = d_k - y_k \tag{6.8}$$

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu e_k \mathbf{x}_k \tag{6.9}$$

6.1.2 Augmented CLMS (ACLMS)

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

$$d_k = \mathbf{x}_k^H \mathbf{g}^o + \mathbf{x}_k^T \mathbf{h}^o + q_k = \mathbf{x}_k^{aH} \mathbf{w}^{oa} + q_k \tag{6.10}$$

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

$$\begin{aligned} J_k^{wl} &= e_k e_k^* = |e_k|^2 \\ &= |d_k - \mathbf{x}_k^H \mathbf{g} - \mathbf{x}_k^T \mathbf{h}|^2 = |d_k - \mathbf{x}_k^{aH} \mathbf{w}^a|^2 \end{aligned} \tag{6.11}$$

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

$$\begin{aligned} y_k &= \mathbf{x}_k^H \mathbf{g}_k + \mathbf{x}_k^T \mathbf{h}_k = \mathbf{x}_k^{aH} \mathbf{w}_k^a \\ e_k &= d_k - y_k \\ \mathbf{w}_{k+1}^a &= \mathbf{w}_k^a + \mu e_k \mathbf{x}_k^a \end{aligned} \tag{6.12}$$

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

$$\mathbf{g}_{k+1} = \mathbf{g}_k + \mu e_k \mathbf{x}_k \tag{6.13}$$

$$\mathbf{h}_{k+1} = \mathbf{h}_k + \mu e_k \mathbf{x}_k^* \tag{6.14}$$

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

6.2 Regularised ACLMS (R-ACLMS)

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

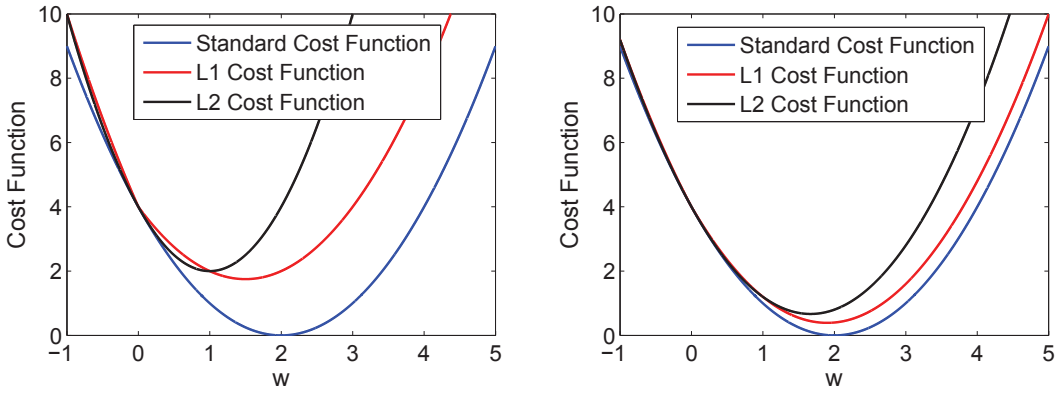
Regularisation can be used to balance between accuracy and model complexity by modifying the cost error function. A regularised version of the cost function (6.11) is given by

$$J_k^r = J_k^{wl} + \gamma ||\mathbf{w}^a||_p \tag{6.15}$$

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

The effect of regularisation in the cost function (6.15), is to enforce the coefficient estimates $\mathbf{g}$ and $\mathbf{h}$ to their minimum norm, which introduces an estimation bias when the true coefficients $\mathbf{g}^o$ and $\mathbf{h}^o$ are non-zero.

To illuminate this point, consider a real valued noiseless observation equation with



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

Figure 6.1: Comparison between the l_1- and l_2-norm regularised cost functions for different values of γ. The standard cost function is achieved by setting $\gamma = 0$.

a scalar weight, that is,

$$d_k = x_k^* w^o \tag{6.16}$$

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

$$J_k^r = |d_k - x_k^* w|^2 + \gamma ||w||_p \tag{6.17}$$

A plot of this cost function for different spans of w is shown in Figure 6.1, where for $\gamma \neq 0$ the minima of the cost functions do not correspond to the optimum weight $w^o = 2$, introducing bias into the estimation. However, the minima of the regularised cost functions approach w^o as γ is reduced. For $w^o = 0$, the cost function minimums are unbiased for any γ value.

6.2.1 Regularised Widely Linear Gradient Descent

The regularised cost function (6.15), regularises both coefficient vectors $\mathbf{g}$ and $\mathbf{h}$, which introduces an estimation bias when the true coefficients $\mathbf{g}^o$ and $\mathbf{h}^o$ are nonzero. In the case of the ACLMS, we are interested in preventing overfitting when the system to be estimated is strictly linear. Therefore, it is the conjugate weight ($\mathbf{h}$) that needs to be

regularised, and the regularised cost function takes the following form:

$$\begin{aligned} J_k^r &= J_k^{wl} + \gamma||\mathbf{h}||_p \\ &= |d_k - \mathbf{x}_k^H\mathbf{g} - \mathbf{x}_k^T\mathbf{h}|^2 + \gamma||\mathbf{h}||_p \end{aligned} \tag{6.18}$$

whereby the minima of (6.18) corresponds to the optimum weights when the systems to be estimated is strictly linear, that is $\mathbf{h}^o = \mathbf{0}$; otherwise for widely linear systems ($\mathbf{h}^o \neq \mathbf{0}$), the cost function minima is not aligned with the optimum weights. Based on (6.18), we have the following time updates for the filter coefficients

$$\begin{aligned} \mathbf{g}_{k+1} &= \mathbf{g}_k + \mu\nabla_{\mathbf{g}} J_k^r|_{\mathbf{g}=\mathbf{g}_k} \\ &= \mathbf{g}_k + \mu e_k\mathbf{x}_k \end{aligned} \tag{6.19}$$

$$\begin{aligned} \mathbf{h}_{k+1} &= \mathbf{h}_k + \mu\nabla_{\mathbf{h}} J_k^r|_{\mathbf{h}=\mathbf{h}_k} \\ &= \mathbf{h}_k + \mu e_k\mathbf{x}_k^* - \alpha\boldsymbol{\Sigma}_p(\mathbf{h}_k) \end{aligned} \tag{6.20}$$

where $\boldsymbol{\Sigma}_p(\mathbf{h}_k) = \big(\nabla_{\mathbf{h}}||\mathbf{h}||_p\big)\big|_{\mathbf{h}=\mathbf{h}_k} \in \mathbb{C}^{L\times 1}$ denotes the subgradient of the l_p-norm, and the term $\alpha = \mu\gamma$ governs the fraction of the conjugate weight updated due to regularisation. For the remainder of this paper, we will refer to the adaptive filters utilising the regularised update equations (6.19)–(6.20) as the regularised-ACLMS (R-ACLMS).

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

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

whereby for a vector input, we have

$$\boldsymbol{\Sigma}_1(\mathbf{h}_k) = [\mathrm{sgn}(h_k^{(1)}), \cdots, \mathrm{sgn}(h_k^{(L)})]^T \tag{6.21}$$

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

$$\mathbf{\Sigma}_2(\mathbf{h}_k) = \frac{1}{2}(\mathbf{h}_k^H \mathbf{h}_k)^{-\frac{1}{2}} \mathbf{h}_k = \frac{1}{2} \frac{\mathbf{h}_k}{||\mathbf{h}_k||_2^{1/2}} \tag{6.22}$$

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

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

$$y_k = \mathbf{x}_k^H \mathbf{g}_k + \mathbf{x}_k^T \mathbf{h}_k = \mathbf{x}_k^{aH} \mathbf{w}_k^a \tag{6.23}$$

$$e_k = d_k - y_k \tag{6.24}$$

$$\mathbf{w}_{k+1}^a = \mathbf{w}_k^a + \mu e_k \mathbf{x}_k^a - \alpha \mathbf{\Delta}_{p,k} \tag{6.25}$$

where $\mathbf{w}_k^a = [\mathbf{g}_k^T, \mathbf{h}_k^T]^T$, $\mathbf{\Delta}_{p,k} = [\mathbf{0}_L^T, \mathbf{\Sigma}_p(\mathbf{h}_k)^T]^T$, and $\mathbf{0}_L^T$ is a zero column vector of length L. The update for the two coefficients can be separated, as shown in (6.19)–(6.20).

6.2.2 Cost Function Bias Analysis

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

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

$$\nabla_{\mathbf{w}^a} J_k^r|_{\mathbf{w}^a} = -e_k \mathbf{x}_k^a + \gamma \mathbf{\Delta}_p \tag{6.26}$$

Setting this derivative to zero and rearranging, we have

$$\begin{aligned} \mathbf{w}_{\min}^a &= [\mathbf{x}_k^a \mathbf{x}_k^{aH}]^{-1} [d_k \mathbf{x}_k^a - \gamma \mathbf{\Delta}_p] \\ &= [\mathbf{x}_k^a \mathbf{x}_k^{aH}]^{-1} [\mathbf{x}_k^a \mathbf{x}_k^{aH} \mathbf{w}^{oa} + \mathbf{x}_k^a q_k - \gamma \mathbf{\Delta}_p] \\ &= \mathbf{w}^{oa} + [\mathbf{x}_k^a \mathbf{x}_k^{aH}]^{-1} [\mathbf{x}_k^a q_k - \gamma \mathbf{\Delta}_p] \end{aligned} \tag{6.27}$$

Utilising the independence assumptions and (6.27), the weight error becomes

$$\tilde{\mathbf{w}}^a = \mathbf{w}^a_{\min} - \mathbf{w}^{oa} = [\mathbf{x}^a_k \mathbf{x}^{aH}_k]^{-1}[\mathbf{x}^a_k q_k - \gamma \mathbf{\Delta}_p] \tag{6.28}$$

Finally, applying the expectation operator to both side yields

$$\begin{aligned} E\{\tilde{\mathbf{w}}^a\} &= E\{[\mathbf{x}^a_k \mathbf{x}^{aH}_k]^{-1}\mathbf{x}^a_k q_k\} - E\{[\mathbf{x}^a_k \mathbf{x}^{aH}_k]^{-1}\gamma \mathbf{\Delta}_p\} \\ &= -\gamma E\{[\mathbf{x}^a_k \mathbf{x}^{aH}_k]^{-1}\}\mathbf{\Delta}_p \end{aligned} \tag{6.29}$$

Remark #2: For the minima of the regularised cost function to align with the optimal weight, it is required that $\mathbf{\Delta}_p = [\mathbf{0}^T_L, \mathbf{\Sigma}_p(\mathbf{h})^T]^T = \mathbf{0}$. Based on (6.21) and (6.22), this is the case only when $\mathbf{w}^a = \mathbf{0}$, otherwise the weight estimate is biased.

6.2.3 Mean Convergence Analysis

We start by substituting the desired signal (6.10) and R-ACLMS output (6.23) into the error signal, that is

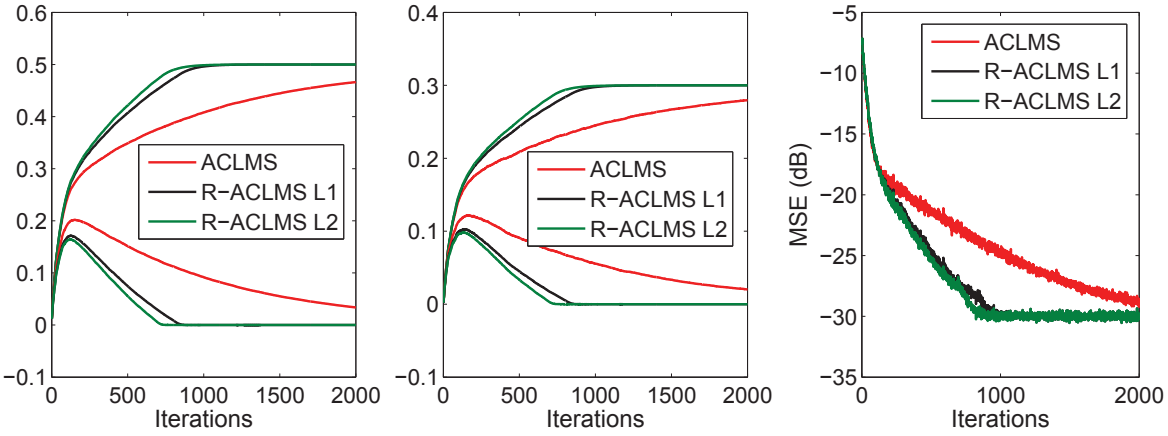
$$e_k = d_k - y_k = \mathbf{x}^{aH}_k \mathbf{w}^{oa} + q_k - \mathbf{x}^{aH}_k \mathbf{w}^a_k \tag{6.30}$$

then the weight update (6.25) can be written as

$$\mathbf{w}^a_{k+1} = \mathbf{w}^a_k + \mu\Big[\mathbf{x}^a_k \mathbf{x}^{aH}_k \mathbf{w}^{oa} + \mathbf{x}^a_k q_k - \mathbf{x}^a_k \mathbf{x}^{aH}_k \mathbf{w}^a_k\Big] - \alpha \mathbf{\Delta}_{p,k} \tag{6.31}$$

Subtracting the optimal weight vector $\mathbf{w}^o$ from both sides of (6.31) yields

$$\begin{aligned} \tilde{\mathbf{w}}^a_{k+1} &= \mathbf{w}^a_{k+1} - \mathbf{w}^o \\ &= \tilde{\mathbf{w}}^a_k - \mu \mathbf{x}^a_k \mathbf{x}^{aH}_k \tilde{\mathbf{w}}^a_k + \mu \mathbf{x}^a_k q_k - \alpha \mathbf{\Delta}_{p,k} \end{aligned} \tag{6.32}$$



(a) Real parts of the g_k and h_k (b) Imaginary parts of the g_k and h_k (c) MSE performance

Figure 6.2: Coefficient convergence of the ACLMS and R-ACLMS for a strictly linear system with a noncircular input signal.

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

$$\begin{aligned} E\{\tilde{\mathbf{w}}_{k+1}^a\} &= (\mathbf{I} - \mu\mathbf{R}_{\mathbf{x}}^a)E\{\tilde{\mathbf{w}}_k^a\} + \mu E\{\mathbf{x}_k^a q_k\} - \alpha E\{\boldsymbol{\Delta}_{p,k}\} \\ &= (\mathbf{I} - \mu\mathbf{R}_{\mathbf{x}}^a)E\{\tilde{\mathbf{w}}_k^a\} - \alpha E\{\boldsymbol{\Delta}_{p,k}\} \end{aligned} \tag{6.33}$$

Remark #3: Note that by setting $\alpha = 0$, the R-ACLMS become the ACLMS, and the convergence results for the ACLMS apply [58], whereby the ACLMS is stable for

$$0 < \mu < \frac{2}{\lambda_{\max}(\mathbf{R}_{\mathbf{x}}^a)} \tag{6.34}$$

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

6.3 Application Examples

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

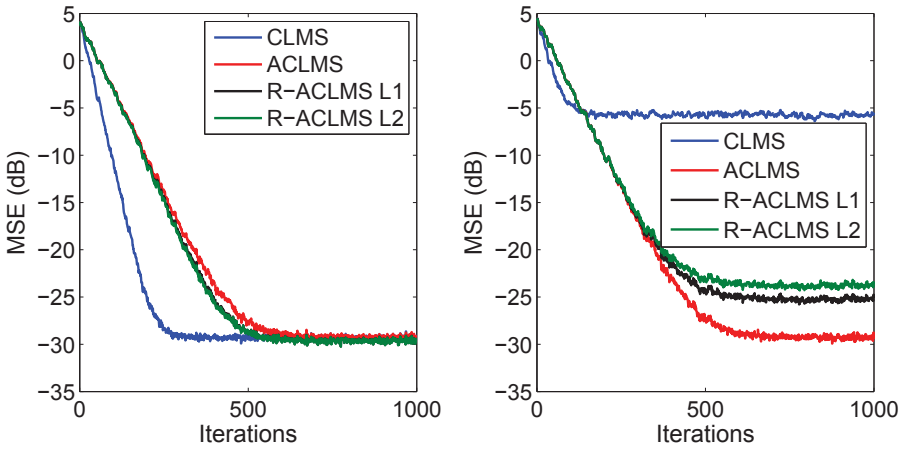
$$d_k = x_k^* w^o + q_k$$

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

Figure 6.2 shows the convergence of the real and imaginary parts of the standard weight g_k and conjugate weight h_k, together with the mean square error (MSE) performances. The results show that the R-ACLMS algorithms offer better weights and error convergence rates for noncircular inputs. However, the performance gains of the R-ACLMS algorithms are not inherited in widely linear systems as will be discussed in the following example.

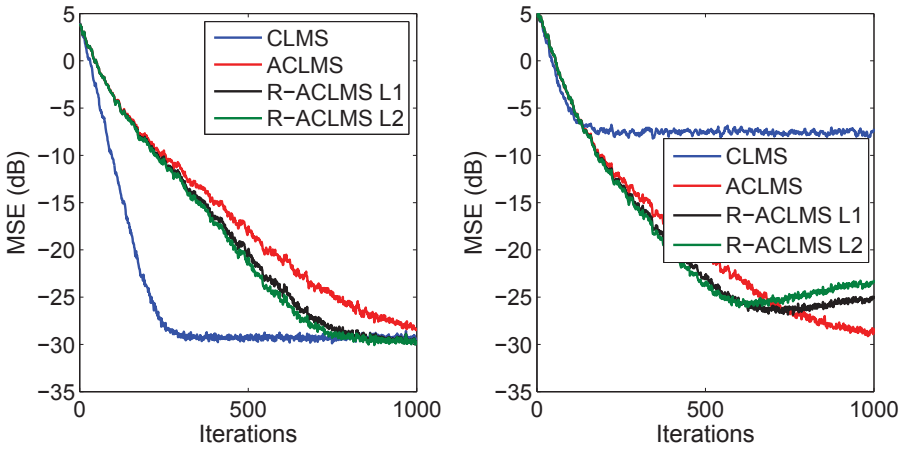
We next consider a system identification problem for (a) a strictly linear system with 15 complex weights (coefficients) and (b) a widely linear system with 15 standard complex weights and 15 conjugate weights. The input vector $\mathbf{x}_k$ was a complex white random process with identity covariance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

For the set of simulations shown in Figure 6.3, the input vector had a circular Gaussian distribution, and an adaptation gain of $\mu = 0.01$ was chosen for the CLMS, while for the ACLMS and R-ACLMS algorithms μ was set at half of this value to ensure that all the algorithms have the same steady-state performance as the CLMS. The results



(a) Strictly linear system (b) Widely linear system

Figure 6.3: Performance comparison between the CLMS, the widely linear ACLMS, the l_1- and l_2-norm regularised ACLMS (R-ACLMS) for strictly and widely linear systems with a circular input vector $E\{\mathbf{x}_k\mathbf{x}_k^T\} = \mathbf{0}$.



(a) Strictly linear system (b) Widely linear system

Figure 6.4: Performance comparison between the CLMS, the widely linear ACLMS, the l_1- and l_2-norm regularised ACLMS (R-ACLMS) for strictly and widely linear systems with a noncircular input vector $E\{\mathbf{x}_k\mathbf{x}_k^T\} = 0.6\mathbf{I}$.

for the strictly linear system, show that the CLMS had the best convergence rate [58], while the R-ACLMS algorithms converged slightly faster than the ACLMS. All the algorithms reached similar steady-states, and the R-ACLMS algorithms provided unbiased weight estimates. For the widely linear system in Figure 6.3b, the ACLMS offered the best steady-state performance due to its unbiased weight estimate, while the R-ACLMS weight

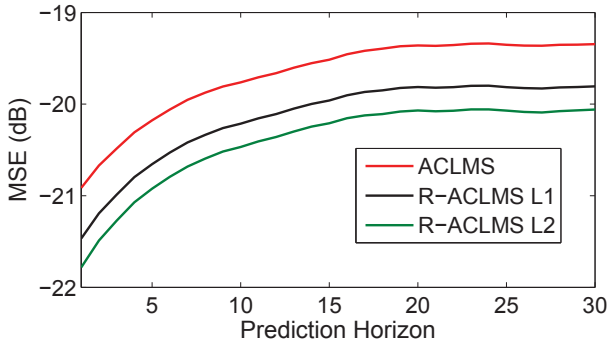


Figure 6.5: Performance comparison between the different algorithms for the prediction of real-world Wind data at different prediction horizons.

estimates were biased (see Remark 2), but outperform the CLMS which under-modelled the widely linear system.

Figure 6.4 illustrates the case when the input vector had a noncircular Gaussian distribution $E\{\mathbf{x}_k\mathbf{x}_k^T\} = 0.6\mathbf{I}$. Again, the comparative differences between the algorithms were the same: the CLMS had the best convergence rates, ACLMS had the best steady-state MSE for the widely linear system, while the performances of the R-ACLMS algorithms were somewhere in between the CLMS and ACLMS. The convergence rates of all the algorithms were effected by the noncircularity of the input [58], and this was especially pronounced for the ACLMS algorithm.

The final set of simulation in Figure 6.5, show the performance for multistep ahead prediction of real-world Wind data, whereby the filters employed widely linear 4th order autoregressive processes to make predictions. The results show that the R-ACLMS algorithms were better suited to tracking the improper and nonstationary Wind data compared with the ACLMS, due to their faster convergence rates.

6.4 Conclusions

The widely linear (augmented) complex least mean square (ACLMS) is suited to the generality of complex systems, both strictly and widely linear systems, but suffers from slow convergence speeds. In this work, the conjugate weight regularised ACLMS (R-ACLMS)

algorithm was presented to address the convergence issues of the ACLMS. The analysis shows that regularisation of the standard cost function introduces a weight estimation bias when the underlying system is widely linear, where the size of the bias is determined by the size of the regularisation factor. Simulation results show that the R-ACLMS converges faster than the ACLMS, and offers similar steady-state performance for strictly linear systems, which makes R-ACLMS algorithms better suited to improper nonstationary systems.

Chapter 7

Conclusions

This thesis has highlighted important theoretical and practical aspects regarding complex valued signal processing. Complex signals are natural and offer an alternative to real signals in deriving theoretical and practical solutions to problems. They also present a convenient and beneficial representation for various classes of data, including radar, wind, communication signals and MRI. Further, complex signals arise in many practical applications ranging from Fourier analysis to analytic signals.

The complex domain offers more powerful statistics, the notions of second-order proper and improper complex random variables, give more degrees of freedom. Additionally, a complex representation offers simultaneous modeling and fusion of two variables, as well as a compact way of treating two real valued variables as one number satisfying all the standard rules of algebra. This concluding chapter summarises the work presented in this thesis and suggests some future directions.

7.1 Summary of Work

The objective of this thesis has been to explore and design adaptive real-time complex valued signal processing algorithms and techniques for dealing with the generality of complex signal. An important component of this has been the recent developments which have brought to light problems concerning standard statistical complex signal processing

approaches. The second order statistical properties of a complex vector $\mathbf{z} = \mathbf{x} + j\mathbf{y}$ has conventionally been characterised by its covariance $\mathbf{C}_\mathbf{z} = E\{\mathbf{z}\mathbf{z}^H\}$, and the other statistical moment known as the pseudocovariance $\mathbf{P}_\mathbf{z} = E\{\mathbf{z}\mathbf{z}^T\}$ has typically been ignored implicitly or explicitly in the design of algorithms, which leads to suboptimal solutions from a second order statistical point of view. However, for a complete second order description, both these moments are necessary. It is only for the special class of complex signals known as *proper* and *circular* (rotation invariant probability distributions), that are characterised by a vanishing pseudocovariance (in other words, signals that are uncorrelated with their complex conjugates), that their covariance function suffices to capture the full second order information.

The so called augmented complex statistics has highlighted the benefits of simultaneously catering for both $\mathbf{C}_\mathbf{z}$ and $\mathbf{P}_\mathbf{z}$, through the estimation framework known as widely linear modeling, which seeks to estimate desired signals based on the observations and their complex conjugates. This thesis exploited these approaches to extend and develop adaptive complex signal processing algorithms suited to the generality of complex data. The main contributions of this thesis are as follows.

Chapter 3 presented a class of widely linear complex Kalman filters (KFs) for both linear and nonlinear systems. The augmented complex KF (ACKF) was introduced and shown to be second order optimal for the generality of complex signals (achieves the Cramer-Rao lower bound), and the analysis showed that it offers better mean square error (MSE) performance compared with the conventional complex KF (CCKF). For nonlinear systems, the augmented extended KF (ACEKF) and augmented unscented KF (ACEKF) were presented for systems with improper signal statistics to caters for the complete second order information. Analysis of the mean square characteristics of CCKF has shown that it is blind to the impropriety of the state and observation signals, however, the mean square characteristics of the complex extended Kalman filter (CEKF) and complex unscented Kalman filter (CUKF) are a function of the impropriety of the state noise noncircularity for the nonlinear observation models. In Appendix A, nonlinear state space estimation problems utilising complex particle filters in conjunction with augmented statistics was

explored. The results show that the proposed solutions offer enhanced performance for improper data. Appendix B proposes a widely linear solution to the DIFAR sonobuoy bearing estimation problem for underwater acoustic sources based on the ACEKF in Chapter 3.

In Chapter 4, we introduced a widely linear state space based frequency estimation technique in the context of three-phase power systems, under both balanced and unbalanced operating conditions. The analysis and simulations showed that owing to the underpinning state space representation, this approach offers enhanced accuracy and faster convergence, together with robustness to noises. Further, circularity diagrams of Clarke's $\alpha\beta$ transformed signals allowed for the identification of unbalanced faulty conditions.

We extended the Kalman filter algorithms proposed in Chapter 3 to the case of distributed cooperative state space estimation in Chapter 5, whereby nodes in a network collaborate locally with their neighbours to estimate signals. The propose distributed augmented (widely linear) complex KF (D-ACKF) caters for general improper complex signals, as well as the cross-correlations between the observation noises at neighbouring nodes, unlike earlier distributed Kalman filtering solutions. The analysis and simulations show that D-ACKF provides unbiased and consistent estimates, and enhanced performance for improper signals.

Chapter 6 addressed the convergence issues related to the widely linear (augmented) complex least mean square (ACLMS) algorithm through the use of widely linear regularised cost functions. The conjugate weight regularised ACLMS (R-ACLMS) algorithm was presented, and analysis showed that regularisation of the standard cost function introduces a weight estimation bias when the underlying system is widely linear, where the size of the bias is determined by the size regularisation factor. The simulation results showed that the R-ACLMS converges faster than the ACLMS, and offers similar steady-state performance for strictly linear systems, which makes R-ACLMS algorithms better suited to improper nonstationary systems.

7.2 Future Work

This thesis developed and illustrated the performance of widely linear complex state space algorithms, and illustrated their performances within a number of applications. However, these were by no means exhaustive. An interesting avenue of further research would be to consider applications of widely linear modelling algorithms and augmented complex statistics to a wider range of problems. A few topics are suggested below.

7.2.1 Complex Signals in Transform Domains

Complex signals are at the heart of transform domains such as the Fourier transforms, where real or complex time domain data are mapped onto the frequency domain. For example, we are often interested in estimating the sizes of the complex valued frequency bins and suppressing the noises associated with them, and due to the nonstationarity and impropriety of the analysed signals (e.g. speech), the second order optimum estimators of these bins will generally take widely linear forms [59]. There is also work to be done within spectral analysis techniques to address the possible benefits of augmented statistics and widely linear estimation.

7.2.2 Higher Order Propriety

The definition for complex propriety is not limited to second order statistics, but is also defined for higher order moments. There is one moment function corresponding to each statistical moment for real valued data, whereas for complex data the number of moment functions increases with the order of the moment. For example, a zero-mean complex signal z has four third order moments ($E\{zzz\}, E\{zzz^*\}, E\{zz^*z^*\} and E\{z^*z^*z^*\}$) - for a real random variable all these moments are equal [7]. In general, for a given order n, there are $(n+1)$ different moments[1]. Taking higher order information is particularly important in applications involving nonlinear estimation and/or non-Gaussian data sources. Widely linear kernels and neural networks have recently been introduced as way of incorporating

[1]Practically, only (m = floor$[n/2]+1$) are required for a complete nth order statistical description, since e.g. $E\{z^*z^*z^*\}$ is a deterministic transformation of $E\{zzz\}$.

higher order moments, however, this is an open avenue to further research.

7.2.3 Complex Signals in Communication Systems

Complex signals are ubiquitous within communication systems, where in-phase and quadrature components are combined to form complex exponentials, and estimation and detection are also typically carried out within the complex domain. An integral part of modern communication systems involves estimating the communication channel, and the arrival directions of signals and interferences. Due to the nonstationarity of the channel and interferences, the general form of minimum MSE receivers will be widely linear for improper data. Recently some initial research has been carried out to exploit widely linearity within this field, such as the widely linear channels equalization [60], and widely linear reception strategies [61], however, there is ample room for further research.

7.2.4 Complex Valued Imaging

Complex signals are also common in a number of imaging processing application [62]. For example, many coherent imaging systems such as synthetic aperture radar (SAR) have an inherent random phase, and are complex valued by nature [63]. Another example is ultrasound imaging which is typically carried out non-coherently on the extracted envelope of a signal, which is complex by an analytic representation or through a complex rotator demodulation of the signal.

7.2.5 Complex Biomedical Engineering

Complex signals play an increasing important role in biomedical engineering, where various signal processing techniques are being exploited to gain better insights [64]. These techniques include joint time-frequency analysis of non-stationary biomedical signals, analytic and Hilbert transform analysis of nonlinear systems and oscillators, together with biological clocks. Recent preliminary studies in this topic, such as the use of signal processing to classify between brain dead and coma patients, have shown promising results, and necessitate further investigations [65].

Appendix A

Particle Filtering and Augmented Complex Statistics

This Appendix explores nonlinear complex state space estimation problems utilising complex valued particle filters in conjunction with augmented statistics. The particle filter is a sequential Monte Carlo method offering a general numerical tool to approximate the Bayesian posterior distribution in nonlinear and non-Gaussian filtering problems. Compared to the Kalman filter and its extensions[40], which are generally designed for linear systems with Gaussian distributions (catering for only the first two statistical moments), the particle filter in both its real- and complex-valued forms has been shown to also effectively capture higher order moments, and has found a wide range of applications [66][67][68][69].

Conventional complex valued particle filters employing Gaussian distributions have been designed (implicitly or explicitly) for the special class of signals known as second order circular (proper) [68][70], that is, signals with rotation invariant probability distributions. However, complex Gaussian signals are typically second order noncircular (improper); this is due to the different signal powers in the real and imaginary parts, correlation of the real and imaginary parts, or finite sample size [7]. It has been shown recently for a number of complex valued algorithms that catering for the noncircularity of signals yields significantly improved estimation performance [1][71][72], and the ability to unify

the processing of proper and improper signals under one umbrella.

For instance, in the case of complex Gaussian distributions for general (noncircular) complex signals, the probability density function (PDF) is a function of the augmented covariance matrix [73] and consequently of both the covariance and pseudocovariance - it is only in the special case of circular signals, that the pseudocovariance vanishes, and the PDF is solely dependent on the covariance.

In this Appendix, we propose to employ augmented complex statistics in conjunction with particle filtering in order to provide sequential estimation for state space models with general Gaussian complex noises. We also illuminate the performance of such an augmented complex particle filter (ACPF), as well as that of the conventional complex particle filter (CCPF), for signals with various degrees of noncircularity. Further, the augmented complex Gaussian particle filter (ACGPF), which propagates the posterior distribution using the sample mean and both the sample covariance and pseudocovariance, is proposed and highlights the benefits of catering for noncircularity through the use of the pseudocovariance. Simulations on both proper and improper data support show the potential of the proposed filters.

A.1 Background

A.1.1 Generalised Multivariate Complex Gaussian Distribution

In standard statistics of complex signals, a zero mean complex variable $\mathbf{z} = \mathbf{z}_r + j\mathbf{z}_i \in \mathbb{C}^N$, the multivariate complex normal distribution (CND) $\mathbf{z} \sim \mathcal{N}(E\{\mathbf{z}\}, \mathbf{R}_\mathbf{z})$ is given by

$$\mathcal{P}[\mathbf{z}] = \frac{1}{\pi^N \det \mathbf{R}_\mathbf{z}} e^{-\mathbf{z}^H \mathbf{R}_\mathbf{z}^{-1} \mathbf{z}} \tag{A.1}$$

which has been a standard widely used in the literature. Recent results show that CND is only suitable for characterising Gaussian circular distributions [73], and that it is a special case of the generalised (multivariate) complex normal distribution (GCND) $\mathbf{z} \sim$

$\mathcal{N}(E\{\mathbf{z}\}, \mathbf{R_z}, \mathbf{P_z})$ with a PDF given by [7][73]

$$\mathcal{P}[\mathbf{z}^a] = \frac{1}{\pi^N (\det \mathbf{R}_{\mathbf{z}}^a)^{1/2}} e^{-\frac{1}{2}\mathbf{z}^{aH}(\mathbf{R}_{\mathbf{z}}^a)^{-1}\mathbf{z}^a} \tag{A.2}$$

which describes both circular and noncircular normal distributions, through the use of the augmented covariance matrix $\mathbf{R}_{\mathbf{z}}^a$. It is straightforward to show that GCND degenerates to CND for circular distributions ($\mathbf{P_z} = \mathbf{0}$).

A.2 Complex Particle Filtering

A.2.1 Conventional Complex PF (CCPF)

The particle filter is a sequential Monte Carlo technique that estimates the posterior state distribution using sequential importance sampling. The idea is to use a set of weighted samples (particles), drawn from the posterior distribution, to directly implement an optimal Bayesian estimate. Consider the complex valued state space model given by

$$\mathbf{x}_k = \mathbf{f}[\mathbf{x}_{k-1}, \mathbf{v}_k] \tag{A.3a}$$

$$\mathbf{y}_k = \mathbf{h}[\mathbf{x}_k, \mathbf{n}_k] \tag{A.3b}$$

where $\mathbf{x}_k$ is the state to be estimated at time instant k, $\mathbf{y}_k$ the noisy observation, $\mathbf{f}[\cdot]$ and $\mathbf{h}[\cdot]$ the nonlinear (possibly time-varying) state transition and measurement functions, while the vectors $\mathbf{v}_k$ and $\mathbf{n}_k$ comprise the uncorrelated state and measurement noises. In standard complex estimation, the posterior distribution of the full state trajectory $\mathbf{x}_{0:k} = \{\mathbf{x}_0, \ldots, \mathbf{x}_k\}$ based on the observation sequence $\mathbf{y}_{1:k} = \{\mathbf{y}_1, \ldots, \mathbf{y}_k\}$ is approximated as [74]

$$\hat{\mathcal{P}}[\mathbf{x}_{0:k}|\mathbf{y}_{1:k}] \approx \frac{1}{M}\sum_{i=1}^{M} \delta[\mathbf{x}_{0:k} - \mathbf{x}_{0:k}^{(i)}] \tag{A.4}$$

where the independent and identically distributed (i.i.d.) samples $\{\mathbf{x}_{0:k}^{(i)}\}_{i=1}^{M}$ are drawn from the posterior distribution $\mathcal{P}[\mathbf{x}_{0:k}|\mathbf{y}_{1:k}]$ and $\delta[\cdot]$ is the Dirac delta function. Expectations of

the form

$$E\{g[\mathbf{x}_{0:k}]\} = \int g[\mathbf{x}_{0:k}]\mathcal{P}[\mathbf{x}_{0:k}|\mathbf{y}_{1:k}]d\mathbf{x}_{0:k} \tag{A.5}$$

where $g[\cdot]$ is either a linear or nonlinear function, can then be approximated as

$$E\{g[\mathbf{x}_{0:k}]\} \approx \frac{1}{M}\sum_{i=1}^{M} g[\mathbf{x}_{0:k}^{(i)}] \tag{A.6}$$

whereby, as the number of particles M approaches infinity, the approximation (A.6) will converge to the true expectation almost surely. Note that although we are working with the more general full posterior distribution $\mathcal{P}[\mathbf{x}_{0:k}|\mathbf{y}_{1:k}]$, the filtering distribution $\mathcal{P}[\mathbf{x}_k|\mathbf{y}_{1:k}]$ is simply a marginal of the full distribution.

Sampling from the posterior density function is often mathematically intractable. However, by utilising the concept of importance sampling, the samples can be generated from a known proposal (or importance) density $\mathcal{Q}[\mathbf{x}_{0:k}|\mathbf{y}_{1:k}]$, instead of the true posterior density function. It follows that the expectation in (A.5) can be rewritten as

$$\begin{aligned} E\{g[\mathbf{x}_{0:k}]\} &= \int g[\mathbf{x}_{0:k}]\frac{\mathcal{P}[\mathbf{x}_{0:k}|\mathbf{y}_{1:k}]}{\mathcal{Q}[\mathbf{x}_{0:k}|\mathbf{y}_{1:k}]}\mathcal{Q}[\mathbf{x}_{0:k}|\mathbf{y}_{1:k}]d\mathbf{x}_{0:k} \\ &= \int g[\mathbf{x}_{0:k}]\frac{w_k}{\mathcal{P}[\mathbf{y}_{1:k}]}\mathcal{Q}[\mathbf{x}_{0:k}|\mathbf{y}_{1:k}]d\mathbf{x}_{0:k} \end{aligned} \tag{A.7}$$

where the variables w_k are the importance weights defined as

$$w_k = \frac{\mathcal{P}[\mathbf{x}_{0:k}]\mathcal{P}[\mathbf{y}_{1:k}|\mathbf{x}_{0:k}]}{\mathcal{Q}[\mathbf{x}_{0:k}|\mathbf{y}_{1:k}]} \tag{A.8}$$

Now, by using a proposal density which can be factorised as

$$\mathcal{Q}[\mathbf{x}_{0:k}|\mathbf{y}_{1:k}] = \mathcal{Q}[\mathbf{x}_{0:k-1}|\mathbf{y}_{1:k-1}]\mathcal{Q}[\mathbf{x}_k|\mathbf{x}_{0:k-1}, \mathbf{y}_{1:k}] \tag{A.9}$$

together with making the assumption that the current state is independent of future observations, allows for a sequential update of the importance weights, that is

$$w_k = \frac{\mathcal{P}[\mathbf{y}_k|\mathbf{x}_k]\mathcal{P}[\mathbf{x}_k|\mathbf{x}_{k-1}]}{\mathcal{Q}[\mathbf{x}_k|\mathbf{x}_{0:k-1}, \mathbf{y}_{1:k}]}w_{k-1} \tag{A.10}$$

In other words, the expression in (A.10) shows how to sequentially update the importance weights given the prior $\mathcal{P}[\mathbf{x}_k|\mathbf{x}_{k-1}]$, likelihood $\mathcal{P}[\mathbf{y}_k|\mathbf{x}_k]$, and proposal $\mathcal{Q}[\mathbf{x}_k|\mathbf{x}_{0:k-1}, \mathbf{y}_{1:k}]$ densities.

Based on the particles $\{\mathbf{x}_{0:k}^{(i)}\}_{i=1}^{M}$ generated from the proposal density, the expectation in (A.6) is approximated as

$$E\{g[\mathbf{x}_{0:k}]\} \approx \frac{\frac{1}{M}\sum_{i=1}^{M} w_k^{(i)} g[\mathbf{x}_{0:k}^{(i)}]}{\frac{1}{M}\sum_{i=1}^{M} w_k^{(i)}} = \sum_{i=1}^{M} \tilde{w}_k^{(i)} g[\mathbf{x}_{0:k}^{(i)}] \tag{A.11}$$

where $w_k^{(i)}$ is the weight corresponding to particle $\mathbf{x}_{0:k}^{(i)}$, and the variable $\tilde{w}_k^{(i)} = w_k^{(i)} / \sum_{n=1}^{M} w_k^{(n)}$ is the corresponding normalised weight. Provided the support of the proposal density includes the support of the true posterior distribution, the approximation in (A.11) will asymptotically converge, and the full posterior distribution can be estimated as

$$\hat{\mathcal{P}}[\mathbf{x}_{0:k}|\mathbf{y}_{1:k}] \quad = \quad \sum_{i=1}^{M} \tilde{w}_k^{(i)} \delta[\mathbf{x}_{0:k} - \mathbf{x}_{0:k}^{(i)}] \tag{A.12}$$

A major problem of particle filtering is the issue of degeneracy, where, after a few iterations, the majority of the particles are assigned negligible weights, which leads to deteriorating performance. Degeneracy, however, can be reduced by resampling the particles; several resampling schemes have been proposed in the literature, including sampling importance resampling (SIR), and residual resampling [70].

SIR involves remapping the particles and their corresponding weights to a set of particles with equal weights, whereby M samples are randomly selected from the set $\{\mathbf{x}_k^{(i)}\}_{i=1}^{M}$ with corresponding probabilities $\{\tilde{w}_k^{(i)}\}_{i=1}^{M}$. Resampling can be implemented as needed or at every time step. In this paper, we adopt the latter strategy.

A.2.2 Augmented Complex PF (ACPF)

To account for noncircular signal distributions, estimation of general Gaussian complex signals should be based on augmented statistics, catering for both circular or noncircular

distributions under the same umbrella. Hence, a more general approximation than that in (A.4), of the true posterior density becomes

$$\hat{\mathcal{P}}[\mathbf{x}^a_{0:k}|\mathbf{y}^a_{1:k}] \approx \frac{1}{M}\sum_{i=1}^{M}\delta[\mathbf{x}^a_{0:k} - \mathbf{x}^{a(i)}_{0:k}] \tag{A.13}$$

which approximates the state based on both the observation sequence and its conjugate. In (A.13), $\mathbf{x}^a_{0:k} = \{\mathbf{x}^a_0, \ldots, \mathbf{x}^a_k\}$ and $\mathbf{y}^a_{1:k} = \{\mathbf{y}^a_1, \ldots, \mathbf{y}^a_k\}$, while the samples $\{\mathbf{x}^{(i)}_{0:k}\}_{i=1}^M$ are drawn from the true posterior density $\mathcal{P}[\mathbf{x}^a_{0:k}|\mathbf{y}^a_{1:k}]$, as opposed to $\mathcal{P}[\mathbf{x}_{0:k}|\mathbf{y}_{1:k}]$ which only describes circular distributions. We use densities with augmented arguments, such as in (A.2) and $\mathcal{P}[\mathbf{x}^a_{0:k}|\mathbf{y}^a_{1:k}]$, to refer to densities which characterise both circular and noncircular distributions, while, densities with non-augmented arguments, such as in (A.1) and $\mathcal{P}[\mathbf{x}_{0:k}|\mathbf{y}_{1:k}]$, refer to densities which only describe circularity distributions.

The first aspect of the particle filtering framework where noncircularity can be incorporated is in the selection of proposal densities[1], which should be chosen to reflect the circularity of the true posterior density. A common choice of proposal density is the transition prior density, that is

$$\mathcal{Q}[\mathbf{x}_k|\mathbf{x}_{0:k-1}, \mathbf{y}_{1:k}] = \mathcal{P}[\mathbf{x}^a_k|\mathbf{x}^a_{k-1}] \tag{A.14}$$

The second aspect where noncircularity of signals can be accounted for is during the evaluation of the importance weights, and more specifically the likelihoods. The likelihoods of the particles can be examined using either density functions suited only to describing circular distribution [68][70] $\mathcal{P}[\mathbf{y}_k|\mathbf{x}_k]$, such as in (A.1), or using more general density functions that can characterise both circular and noncircular distributions $\mathcal{P}[\mathbf{y}^a_k|\mathbf{x}^a_k]$, such as that in (A.2).

For example, consider the case of a nonlinear state space model with additive

[1]The choice of proposal density is an important design issue for particle filters; in fact, the optimal choice of proposal density is the unknown true posterior density.

Gaussian noises,

$$\mathbf{x}_k = \mathbf{f}[\mathbf{x}_{k-1}] + \mathbf{v}_k \tag{A.15a}$$

$$\mathbf{y}_k = \mathbf{h}[\mathbf{x}_k] + \mathbf{n}_k \tag{A.15b}$$

Given the previous state $\mathbf{x}_{k-1}$, the transition prior is defined as in (A.1), that is

$$\mathcal{P}[\mathbf{x}_k|\mathbf{x}_{k-1}] \sim \mathcal{N}(\mathbf{f}[\mathbf{x}_{k-1}], \mathbf{R}_\mathbf{v}) \tag{A.16}$$

However, when the state noise is noncircular, the true prior distribution is described by (A.2), giving

$$\mathcal{P}[\mathbf{x}^a_k|\mathbf{x}^a_{k-1}] \sim \mathcal{N}(\mathbf{f}[\mathbf{x}_{k-1}], \mathbf{R}_\mathbf{v}, \mathbf{P}_\mathbf{v}) \tag{A.17}$$

This way, the distribution incorporates the circularity of the state noise through the use of the pseudocovariance $\mathbf{P}_\mathbf{v}$. Similarly, given the current state $\mathbf{x}_k$, the likelihood density is correctly evaluated using an augmented density function,

$$\mathcal{P}[\mathbf{y}^a_k|\mathbf{x}^a_k] \sim \mathcal{N}(\mathbf{h}[\mathbf{x}_k], \mathbf{R}_\mathbf{n}, \mathbf{P}_\mathbf{n}) \tag{A.18}$$

The augmented complex particle filter (ACPF) takes a similar form to the conventional complex particle filter, but utilises augmented complex densities to cater for noncircular signal distributions, and is summarised in Algorithm 12.

At this point it is worth pointing out that due to the topological isomorphism between augmented complex vectors and bivariate real vectors[7], real valued algorithms have dual corresponding augmented complex versions and vice-versa, with identical performances[1]. For any complex vector $\mathbf{z} = \mathbf{z}_r + j\mathbf{z}_i \in \mathbb{C}^N$, the duality mapping is given by

$$\mathbf{z}^a = \begin{bmatrix} \mathbf{z} \\ \mathbf{z}^* \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{I} & j\mathbf{I} \\ \mathbf{I} & -j\mathbf{I} \end{bmatrix}}_{\equiv \mathbf{J}_\mathbf{z}} \underbrace{\begin{bmatrix} \mathbf{z_r} \\ \mathbf{z_i} \end{bmatrix}}_{=\mathbf{z}^r} \tag{A.19}$$

Algorithm 12: The augmented complex PF (ACPF)

Initialisation:

- Draw state particles $\{\mathbf{x}_0^{(i)}\}_{i=1}^M$ from prior $\mathcal{P}[\mathbf{x}_0^a]$
- Set the weights as: $\{\tilde{w}_0^{(i)}\}_{i=1}^M = \frac{1}{M}$

For $k = 1, 2, ...$

- Sample from proposal density: $\mathbf{x}_k^{(i)} \sim \mathcal{Q}[\mathbf{x}_k^a | \mathbf{x}_{k-1}^{a,(i)}, \mathbf{y}_{1:k}^a]$
- Evaluate importance weights:

$$w_k^{(i)} = \frac{\mathcal{P}[\mathbf{y}_k^a | \mathbf{x}_k^{a,(i)}] \mathcal{P}[\mathbf{x}_k^{a,(i)} | \mathbf{x}_{k-1}^{a,(i)}]}{\mathcal{Q}[\mathbf{x}_k^{a,(i)} | \mathbf{x}_{0:k-1}^{a,(i)}, \mathbf{y}_{1:k}^a]}$$

- Normalise importance weights:

$$\tilde{w}_k^{(i)} = w_k^{(i)} / \Big(\sum_{n=1}^{M} w_k^{(n)} \Big)$$

- Compute state estimate:

$$\hat{\mathbf{x}}_k = \sum_{i=1}^{M} \tilde{w}_k^{(i)} \mathbf{x}_k^{(i)}$$

- Resample using an appropriate scheme, such as sampling importance resampling (SIR) or residual resampling

where $\mathbf{I}$ is the identity matrix (with appropriate dimensions), and the invertible orthogonal mapping[2] $\mathbf{J}_\mathbf{z} : \mathbb{C}^{2q} \to \mathbb{R}^{2q}$ is such that $\mathbf{J}_\mathbf{z}^{-1} = \frac{1}{2}\mathbf{J}_\mathbf{z}^H$ [20]. The derivation of the GCND above is also based on this isomorphism.

Generally, state space models are naturally defined in either the real or complex domain, and it is desirable to keep all of the computations in the original domain. This facilitates understanding of the transformations the signals undergo, and we also benefit from the notion of phase and circularity in the case of complex signals.

[2]For a vector $\mathbf{z} \in \mathbb{C}^q$, the corresponding orthogonal matrix $\mathbf{J}_\mathbf{z}$ takes dimension $2q \times 2q$.

A.2.3 Augmented Complex Gaussian PF (ACGPF)

The particle filter in its generic form has a number of disadvantages, including high computational complexity and problems when estimating constants, which has led to a number of variations and extensions which address these issues. One popular variation of the standard particle filter is the Gaussian particle filter (GPF), which essentially approximates the posterior density of the unknown state as a Gaussian distribution [66][75], that is, it only propagates the posterior mean and covariance as the extended Kalman filter (EKF). The benefits of GPF include reduced complexity and a unified framework for estimating both dynamic and constant states.

The main distinguishing point of the GPF algorithm is that after computation of the normalised particle weights, the sample mean and covariance of the state are evaluated as

$$\hat{\mu}_k = \sum_{i=1}^{M} \tilde{w}_k^{(i)} \mathbf{x}_k^{(i)} \tag{A.20}$$

$$\hat{\mathbf{R}}_{\mathbf{x},k} = \sum_{i=1}^{M} \tilde{w}_k^{(i)} \left(\mathbf{x}_k^{(i)} - \hat{\mu}_k\right)\left(\mathbf{x}_k^{(i)} - \hat{\mu}_k\right)^H \tag{A.21}$$

It is these two moments that are propagated to the next iteration, instead of the actual particles, which means that resampling of the particles can be avoided, thus, leading to reduced computational complexity.

However, as shown above, within a complex estimation framework the covariance alone is insufficient for a complete second order characterisation of the particle distributions, and an estimate of the pseudocovariance matrix, given by

$$\hat{\mathbf{P}}_{\mathbf{x},k} = \sum_{i=1}^{M} \tilde{w}_k^{(i)} \left(\mathbf{x}_k^{(i)} - \hat{\mu}_k\right)\left(\mathbf{x}_k^{(i)} - \hat{\mu}_k\right)^T \tag{A.22}$$

is also required for correct operation. A unified approach to the simultaneous estimation of covariance and pseudocovariance matrices can be achieved by utilising augmented particle

vectors $\mathbf{x}_k^{a,(i)} = [\mathbf{x}_k^{(i)T}, \mathbf{x}_k^{(i)H}]^T$ to estimate the augmented covariance matrix, that is

$$\begin{aligned} \hat{\mathbf{R}}_{\mathbf{x},k}^{a} &= \sum_{i=1}^{M} \tilde{w}_k^{(i)} \left(\mathbf{x}_k^{a,(i)} - \hat{\mu}_k^a\right)\left(\mathbf{x}_k^{a,(i)} - \hat{\mu}_k^a\right)^H \\ &= \begin{bmatrix} \hat{\mathbf{R}}_{\mathbf{x},k} & \hat{\mathbf{P}}_{\mathbf{x},k} \\ \hat{\mathbf{P}}_{\mathbf{x},k}^* & \hat{\mathbf{R}}_{\mathbf{x},k}^* \end{bmatrix} \end{aligned} \tag{A.23}$$

where the augmented mean $\hat{\mu}_k^a = [\hat{\mu}_k^T, \hat{\mu}_k^H]^T$. The individual estimates of the second order moments, (A.21) and (A.22), can be cast in to the augmented structure (A.23) to save in computations, thereafter, the distribution of the particles is approximated using the generalised complex multivariate normal distribution (A.2), that is, $\mathcal{N}(\hat{\mu}_k, \hat{\mathbf{R}}_{\mathbf{x},k}, \hat{\mathbf{P}}_{\mathbf{x},k})$. The augmented complex GPF (ACGPF) method is outlined in Algorithm 13.

A.3 Application Examples

To illustrate the advantages of the augmented particle filter algorithms over their conventional counterparts, we considered the following case studies: 1) filtering of a noisy complex-valued autoregressive process; 2) nonlinear bearings only tracking.

A.3.1 Complex autoregressive process

The performances of both the conventional complex particle filter (CCPF) and augmented complex particle filter (ACPF) were examined using 200 particles for a first order complex autoregressive process given by

$$\text{AR(1):} \quad x_k = 0.9x_{k-1} + u_k$$

where the symbol u_k denotes a noncircular white Gaussian driving noise with variance $E\{u_{k-i}u_{k-l}^*\} = E\{|u_{k-i}|^2\} = 0.001$ and a varying pseudovariance $E\{u_k^2\}$. The observation was a nonlinear function of the state, given by

$$y_k = \tanh[x_k] + n_k \tag{A.24}$$

Algorithm 13: The augmented complex Gaussian PF (ACGPF)

Initialisation:

- Draw state particles $\{\mathbf{x}_0^{(i)}\}_{i=1}^M$ from prior $\mathcal{P}[\mathbf{x}_0^a]$
- Set the weights as: $\{\tilde{w}_0^{(i)}\}_{i=1}^M = \frac{1}{M}$

For $k = 1, 2, ...$

- Sample from proposal density: $\mathbf{x}_k^{(i)} \sim \mathcal{Q}[\mathbf{x}_k^a | \mathbf{x}_{k-1}^{a,(i)}, \mathbf{y}_{1:k}^a]$
- Evaluate importance weights:

$$w_k^{(i)} = \frac{\mathcal{P}[\mathbf{y}_k^a | \mathbf{x}_k^{a,(i)}] \mathcal{P}[\mathbf{x}_k^{a,(i)} | \mathbf{x}_{k-1}^{a,(i)}]}{\mathcal{Q}[\mathbf{x}_k^{a,(i)} | \mathbf{x}_{0:k-1}^{a,(i)}, \mathbf{y}_{1:k}^a]}$$

- Normalise importance weights:

$$\tilde{w}_k^{(i)} = w_k^{(i)} / \Big(\sum_{n=1}^{M} w_k^{(n)} \Big)$$

- Estimate mean, covariance and pseudocovariance:

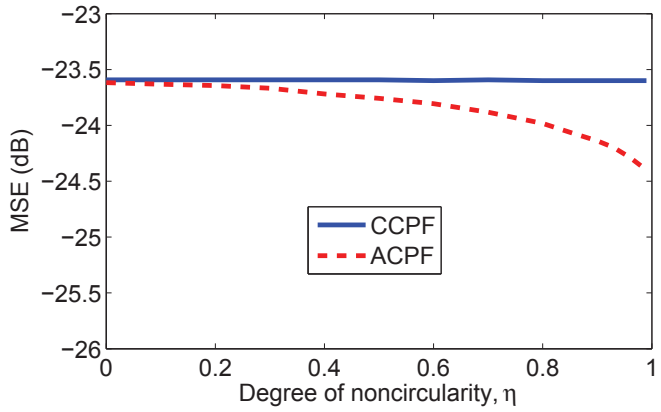
$$\begin{aligned} \hat{\mathbf{x}}_k &= \hat{\mu}_k = \sum_{i=1}^{M} \tilde{w}_k^{(i)} \mathbf{x}_k^{(i)} \\ \hat{\mathbf{R}}_{\mathbf{x},k} &= \sum_{i=1}^{M} \tilde{w}_k^{(i)} \Big(\mathbf{x}_k^{(i)} - \hat{\mu}_k \Big) \Big(\mathbf{x}_k^{(i)} - \hat{\mu}_k \Big)^H \\ \hat{\mathbf{P}}_{\mathbf{x},k} &= \sum_{i=1}^{M} \tilde{w}_k^{(i)} \Big(\mathbf{x}_k^{(i)} - \hat{\mu}_k \Big) \Big(\mathbf{x}_k^{(i)} - \hat{\mu}_k \Big)^T \end{aligned}$$

- Draw particles to be propagated from the GCND:

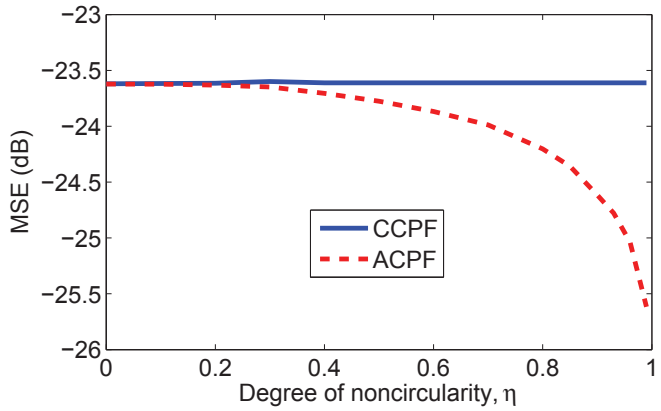
$$\mathbf{x}_k^{(i)} \sim \mathcal{N}(\hat{\mu}_k, \hat{\mathbf{R}}_{\mathbf{x},k}, \hat{\mathbf{P}}_{\mathbf{x},k}) =$$

where n_k is a additive Gaussian complex doubly-white observation noise with variance $E\{|n_{k-i}|^2\} = 0.1$ and a varying pseudovariance $E\{n_k^2\}$. The ratio of the magnitude of the pseudocovariance to covariance, $\eta = |p| / c$, was used as a measure of the degree of circularity of the complex noises, where a complex random variable is circular for $\eta = 0$ and maximally noncircular for $\eta = 1$.

Figure A.1a illustrates the filtering results for a circular observation noise and a



(a) Noncircular state noise



(b) Noncircular observation noise

Figure A.1: Steady-state performance comparison between the conventional complex particle filter (CCPF) and the augmented complex particle filter (ACPF) for the AR(1) filtering problem: (a) circular observation noise and a noncircular state noise with varying degrees of noncircularity; (b) circular state noise and noncircular observation noise with varying degrees of noncircularity.

state noise with various degrees of noncircularity, while Figure A.1b shows the results for a noncircular observation noise and a circular state noise. For both sets of simulations, the ACPF outperformed the CCPF - especially for high degrees of noncircularity. The ACPF had decreasing MSE as the degree of noise noncircularity η increased. As expected, the two algorithms had the same performance for circular state and observation noise distributions, that is, for $\eta = 0$. Observe that the performance of the CCPF was blind to the noncircularity of state and observation noises.

A.3.2 Bearings only tracking

Bearings only tracking (BOT) is a problem encountered in many practical applications, including submarine tracking by passive sonar or aircraft surveillance by a radar in a passive mode. The objective is the online estimation of the position and velocity of a moving target using observer line of sight noisy bearing (phase) measurements. As the range measurements are not available and the bearings are not linearly related to the target state, the problem is inherently nonlinear. A single static sensor is unable to track targets using bearing measurements only (due to the lack of range measurements, and in order to estimate the range, the sensor has to maneuver). For two or more stationary sensors, observability is not an issue, as the multiple bearing measurements can be used to form a range estimate.

To estimate the trajectory of a target at time instant k, that is, its position (x_k, y_k) and velocity $(\dot{x}_k, \dot{y}_k)$, for a system with two observers located at $(x^o_{1,k}, y^o_{1,k})$ and $(x^o_{2,k}, y^o_{2,k})$, the complex BOT state space is constructed as

$$\mathbf{x}_k = \mathbf{F}\mathbf{x}_{k-1} + \mathbf{B}v_k \qquad z_k = h[\mathbf{x}_k] + n_k$$

with the variables defined as follows:

- $\mathbf{x}_k = \begin{bmatrix} x_k + jy_k & \dot{x}_k + j\dot{y}_k \end{bmatrix}^T$ is the target state vector,
- $\mathbf{F}$ and $\mathbf{B}$ are matrices defined as

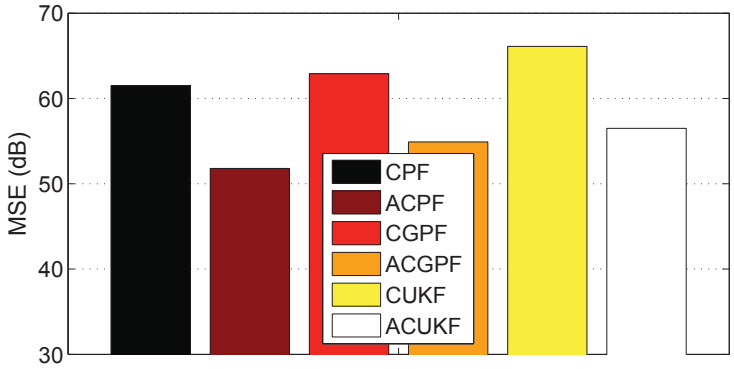
$$\mathbf{F} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} \frac{T^2}{2} \\ T \end{bmatrix}$$

  where T is the sampling interval,

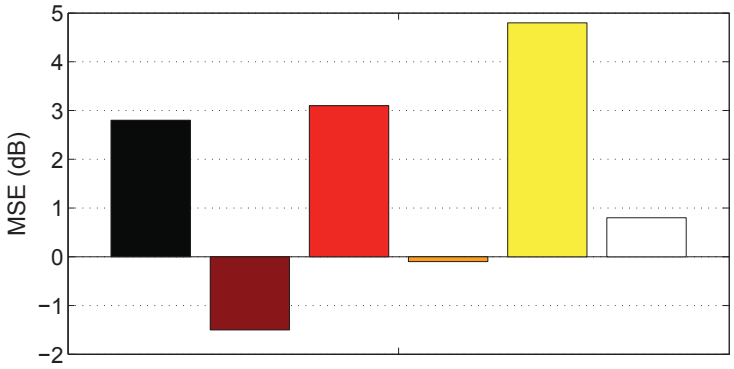
- z_k is the observation vector and $h[\mathbf{x}_k]$ is the scalar observation function defined as

$$h[\mathbf{x}_k] = \beta_{1,k} + j\beta_{2,k}$$

  where $\beta_{i,k} = \tan^{-1} \frac{y_k - y^o_{i,k}}{x_k - x^o_{i,k}}$ is the target bearing with respect to sensor i,



(a) Location estimation error



(b) Velocity estimation error

Figure A.2: Performance of standard and augmented complex filters for BOT problem with noncircular state and observation noises.

- $v_k = \ddot{x}_k + j\ddot{y}_k$ is the white state noise used to account for the unknown target accelerations, while n_k is the complex white observation noise.

The advantages of the augmented complex Gaussian particle filter (ACGPF) over the complex Gaussian particle filter (CGPF) within the context of bearings only target motion analysis are illustrated for a scenario with two static sensors located at $(-2200, -4300)$ and $(1500, 3500)$. The system was simulated using a sampling interval of $T = 1$, and the mean square errors (MSEs) of the different algorithms were computed by averaging 500 independent trials.

The particle filtering algorithms were compared using 1000 particles, noncircular state and observation noises (with 0.9 and 0.8 degrees of noncircularity respectively) with

distributions defined as

$$v_k \sim \mathcal{N}(0, 0.05) \qquad n_k \sim \mathcal{N}(0, 0.001)$$

Figures A.2a and A.2b show the average MSEs in estimating the position and velocity of the target over 100 seconds. The results for the complex unscented Kalman filter (CUKF) and its augmented version ACUKF are also shown for convenience [1]. All the augmented algorithms outperformed their corresponding strictly linear counterparts, as they were able to cater for the noncircular data distributions. The ACGPF underperformed compared with the ACPF, because it approximates the posterior distribution at every iteration as Gaussian, even when the true underlying distribution is non-Gaussian due to the nonlinear observation function.

A.4 Conclusions

Complex Gaussian random variables that are correlated with their conjugates are characterised by improper probability density functions. From a second order statistics point of view, while the covariance captures the information regarding the total power of complex signals, it is the second moment function known as the pseudocovariance, that conveys the nature of the propriety of the signal. Within the particle filtering framework, we have proposed the augmented complex PF (ACPF), based on augmented complex density functions that caters for both circular and circular distributions. The augmented complex Gaussian PF (ACGPF), which propagates the particle distribution based on the sample mean, covariance as well as the sample pseudocovariance, has also been proposed for improper signals. The performance of the filters have been tested using both circular and noncircular data.

Appendix B

An Enhanced Sonobuoy Bearing Estimation Technique

In this Appendix, we consider the DIFAR sonobuoy bearing estimation problem for underwater acoustic sources, and propose an new widely linear solution. Bearing or direction-of-arrival (DOA) estimation is a problem encountered in a wide range of applications, including navigation, surveillance and communication systems. In underwater environments, the DIFAR sonobuoy, consisting of two crossed dipoles and an omni-directional hydrophone, is a typical arrangement used to provide three observations of a source signal (target), which together allow for the bearing (angle) of a source (target) to be estimated. In the ocean, however, there are many sources of background noise, such as environmental noise from wind, rain and waves, and biological noise from whales and other marine mammals. These all contribute to the total power spectrum (both broadband and narrowband) of the observed signals. Moreover, the propagation of acoustic signals in the ocean is generally not uniform or isotropic, which also contributes to the difficulty of the bearing problem in underwater environments [76][77][78].

The standard solutions for sonobuoy target detection and bearing estimation are based on spectral analysis of the observed signals using the discrete Fourier transform (DFT) [76][79][80] or using spectral modelling approaches, such as autoregressive moving average (ARMA). However, these techniques usually suffer from limited frequency resolu-

tion, which becomes especially pronounced for low signal to noise ratios (SNRs), leading to poor performance. Moreover, due to their block-processing nature, these techniques are not suited to rapidly moving targets, where the target bearing is nonstationary during the collection of the data block used for the DFT. Among the popular solutions for underwater sonobuoy bearing estimation is the DFT based 'arctangent' estimator [79] which utilises time-averaged products of the observation data blocks to estimate the target bearing.

In this Appendix, embarking upon the recently introduced augmented complex statistics and widely linear modelling, we propose an online sonobuoy target bearing estimation solution, based on widely linear (augmented) complex state space model introduced in Chapter 3. The second order statistics of both the state and observation noises are estimated from the observation data, and their estimates are also updated online. It is shown that the state space model is inherently nonlinear, and the augmented complex extended Kalman filter is employed to address the problem. Simulations illustrate the robustness of the proposed technique, yielding enhanced performance compared to the standard arctangent estimator, especially in unfavourable signal-to-noise (SNR) conditions.

B.1 New State Space Formulation

Figure B.1 illustrates the arrangement of the sonobuoy sensors for a source at bearing β, the crossed-dipole sensor observes the following three waveforms [79]

$$y_{o,k} = s_k + v_{o,k} \tag{B.1a}$$

$$y_{c,k} = s_k \cos[\beta] + v_{c,k} \tag{B.1b}$$

$$y_{s,k} = s_k \sin[\beta] + v_{s,k} \tag{B.1c}$$

where the subscripts o, c and s denote the omni, cosine and sine channels respectively, while s_k is the signal emitted by the source (target) at time instant k, and $v_{o,k}$, $v_{c,k}$ and $v_{s,k}$ are the uncorrelated, zero-mean, observation noises. . In the standard arctangent bearing estimator, the discrete Fourier transform (DFT) of the observation signals are taken, and the frequency domain representation of the equations above assume the following forms

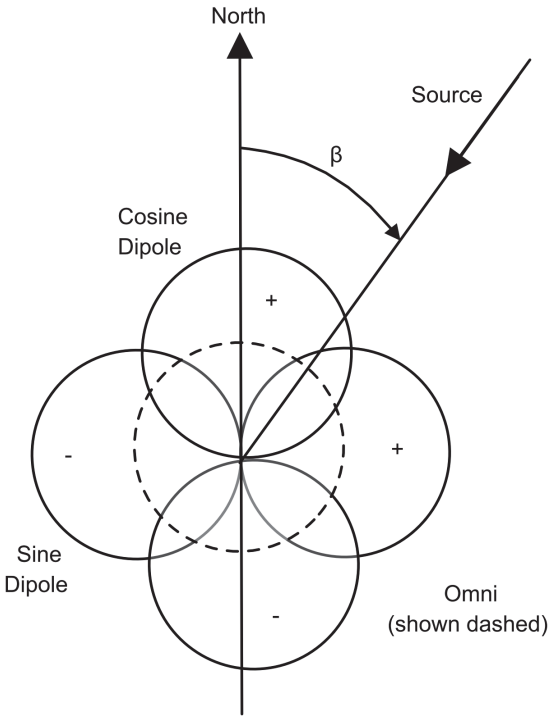


Figure B.1: A geometric view of the three sonobouy sensors (top view).

[79]

$$Y_{o,\omega} = S_\omega + V_{o,\omega} \tag{B.2a}$$

$$Y_{c,\omega} = S_\omega \cos[\beta] + V_{c,\omega} \tag{B.2b}$$

$$Y_{s,\omega} = S_\omega \sin[\beta] + V_{s,\omega} \tag{B.2c}$$

where ω is the frequency argument. A number of data snapshots or observations (M), are collected before taking the DFT, and the source bearing β is inherently assumed to be constant over this observation period. In the standard arctangent estimator, the target bearing is estimated as

$$\hat{\beta} = \arctan[\hat{s}/\hat{c}] \tag{B.3}$$

where the variables $\hat{c}$ and $\hat{s}$ are computed using the M observations, that is

$$\hat{c} = \Re\Big\{ \sum_{m=1}^{M} Y_{c,\omega}^{(m)} Y_{o,\omega}^{(m),*} \Big\} \tag{B.4}$$

$$\hat{s} = \Re\Big\{ \sum_{m=1}^{M} Y_{s,\omega}^{(m)} Y_{o,\omega}^{(m),*} \Big\} \tag{B.5}$$

The superscript m is in the range $1 \leq m \leq M$ and denotes the mth Fourier bin, while $\Re\{\cdot\}$ is the real part of a complex quantity. Observe that the variables $\hat{c}$ and $\hat{s}$ may alternatively be estimated in the time domain (without taking Fourier transforms) as shown in [76].

The arctangent estimator is essentially based on the time (or frequency) averaged products (correlations) of the omni directional sensor $y_{o,k}$ with the outputs from the sine and cosine sensors, $y_{s,k}$ and $y_{c,k}$. It does not attempt to cater for the dynamics of the source signal s_k, and deals with the individual observations (or frequency bins) independently of each other.

However, it is possible to model or exploit possible transitional (correlation) properties in the source signal s_k, which can be inferred from the M available observations, and updated online. For this purpose, we here propose utilising a random-walk (first order Markov) modelling of the signal s_k, that is

$$s_k = s_{k-1} + w_k \tag{B.6}$$

where w_k is the driving noise, together with an augmented complex state space formulation to address the bearing estimation problem, which takes on the following form

$$\mathbf{x}_k = \mathbf{x}_{k-1} + \mathbf{w}_k \tag{B.7}$$

$$\mathbf{y}_k = \mathbf{h}[\mathbf{x}_k] + \mathbf{v}_k \tag{B.8}$$

where $\mathbf{x}_k$ is the state vector to be estimated, $\mathbf{y}_k$ the noisy observation, $\mathbf{h}[\cdot]$ the nonlinear observation function, while $\mathbf{w}_k$ and $\mathbf{v}_k$ are respectively the state and observation noises with covariance matrices $\mathbf{Q}_k$ and $\mathbf{R}_k$ [17]. The state equation (B.7) can be explicitly

expressed as

$$\underbrace{\begin{bmatrix} s_k \\ z_k \\ z_k^* \end{bmatrix}}_{\mathbf{x}_k} = \underbrace{\begin{bmatrix} s_{k-1} \\ z_{k-1} \\ z_{k-1}^* \end{bmatrix}}_{\mathbf{x}_{k-1}} + \underbrace{\begin{bmatrix} w_k \\ e_k \\ e_k^* \end{bmatrix}}_{\mathbf{w}_k} \tag{B.9}$$

where $z_k = \cos[\beta] + j\sin[\beta] = e^{j\beta}$, and e_k is the state noise used to model nonstationary bearings β. Similarly, the observation equation in (B.8) takes the form

$$\begin{aligned} \underbrace{\begin{bmatrix} y_{o,k} \\ u_k \\ u_k^* \end{bmatrix}}_{\mathbf{y}_k} &= \underbrace{\begin{bmatrix} s_k \\ s_k z_k \\ s_k z_k^* \end{bmatrix}}_{\mathbf{h}[\mathbf{x}_k]} + \underbrace{\begin{bmatrix} v_{o,k} \\ n_k \\ n_k^* \end{bmatrix}}_{\mathbf{v}_k} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & s_k & 0 \\ 0 & 0 & s_k \end{bmatrix} \begin{bmatrix} s_k \\ z_k \\ z_k^* \end{bmatrix} + \begin{bmatrix} v_{o,k} \\ n_k \\ n_k^* \end{bmatrix} \end{aligned} \tag{B.10}$$

where $u_k = y_{c,k} + jy_{s,k}$ is the complex representation of the sine and cosine observations channels from (B.1a), and $n_k = v_{c,k} + jv_{s,k}$ is the corresponding noise.

The augmented (widely linear) state space model in (B.7) and (B.8) is nonlinear, and can be used in conjunction with a number of algorithms to estimate the source bearing, including the augmented complex extended and unscented Kalman filters as well as the augmented complex particle filter, [1].

B.1.1 Noise Statistics

In state space estimation we need to specify the second order statistics of the state and observation noises. To that end, given the observation noise variance of the omni channel, that is, $E\{v_{o,k}v_{o,k}^*\}$, the variances of the other two observation noises, $v_{c,k}$ and $v_{s,k}$, are

given by

$$E\{v_{c,k}v_{c,k}^*\} = E\{v_{s,k}v_{s,k}^*\} = \frac{1}{\gamma}E\{v_{o,k}v_{o,k}^*\} \tag{B.11}$$

where γ is the noise gain of either dipole, whereby $\gamma = 1/2$ or $\gamma = 1/3$ for 2D-isotropic or 3D-isotropic noise respectively [76], while the variance of the complex observation noise is $E\{n_k n_k^*\} = E\{v_{c,k}v_{c,k}^*\} + E\{v_{s,k}v_{s,k}^*\}$.

Therefore, the noise statistics to be computed are $E\{v_{o,k}v_{o,k}^*\}$ and $E\{w_k w_k^*\}$, and can be estimated online as follows. We start by forming a new variable defined as the difference between two consecutive omni channel observations, that is

$$r_k = y_{o,k} - y_{o,k-1} \tag{B.12}$$

then assuming that both noise processes, w_k and $v_{o,k}$, are white and stationary, it is can be shown that

$$E\{r_k r_k^*\} = 2E\{v_{o,k}v_{o,k}^*\} + E\{w_k w_k^*\} \tag{B.13}$$

and that the correlation between r_k and r_{k-1} becomes

$$E\{r_k r_{k-1}^*\} = -E\{v_{o,k-1}v_{o,k-1}^*\} = -E\{o, v_{o,k}v_{o,k}^*\} \tag{B.14}$$

therefore, from (B.13) and (B.14), we obtain

$$E\{v_{o,k}v_{o,k}^*\} = -E\{r_k r_{k-1}^*\} \tag{B.15}$$

$$E\{w_k w_k^*\} = E\{r_k r_k^*\} + 2E\{r_k r_{k-1}^*\} \tag{B.16}$$

Hence, the state and observation noise statistics of the state space model described by (B.9) and (B.10), can be estimated and tracked online based on the observation data.

Remark #1: The state space formulation of the problem enables tracking of the source (target) bearing in real-time, that is, the bearing estimate can be updated with

every new observation.

Remark #2: The random-walk model in conjunction with the preprocessing of the observation data (when computing the noise variances), allows for some of the correlation structure of the source signal to be incorporated into the state space model, even when the true source signal does not follow a random-walk model.

B.2 Simulations

To illustrate the potential of our augmented complex state space based solution for sonobuoy bearing estimation, we considered examples where the source signal s_k is modeled as a sinusoid (as in [79]) and as a first order autoregressive process. The augmented complex extended Kalman filter (ACEKF) is used to implement the approach described above and is compared with the standard arctangent (arctan) bearing estimator. In all the simulations, both the arctan and ACEKF algorithms utilise $M = 1024$ observations to estimate the bearing.

B.2.1 Signal Model: Sinusoid

Consider the case where the signal is a sinusoid, that is

$$s_k = \cos[2\pi f T k] + n_k \tag{B.17}$$

with a frequency of $f = 50Hz$, sampled at a rate of $f_s = \frac{1}{T} = 10kHz$.

Figure B.2 shows the enhanced performances of the proposed ACEKF based solution compared with the arctan estimator, for the case where the source signal is a pure sinusoid. The results show that the proposed technique was able to outperform the arctan algorithm for low signal to noise (SNR) levels, while the two algorithms had similar performances for SNRs greater than $0dB$.

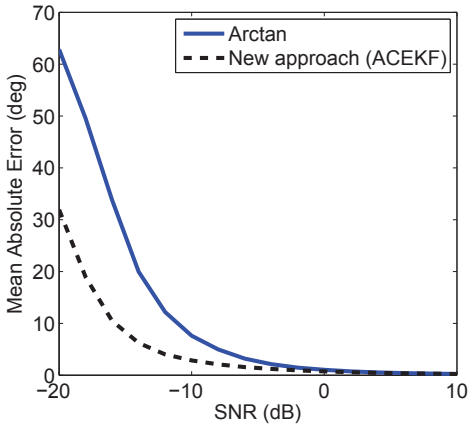


Figure B.2: Performance comparison between the proposed augmented complex state space approach and the arctan estimator for the case where the target source signal is a sinusoid.

B.2.2 Signal Model: Autoregressive

We next modelled the source signal as a first order autoregressive process, that is

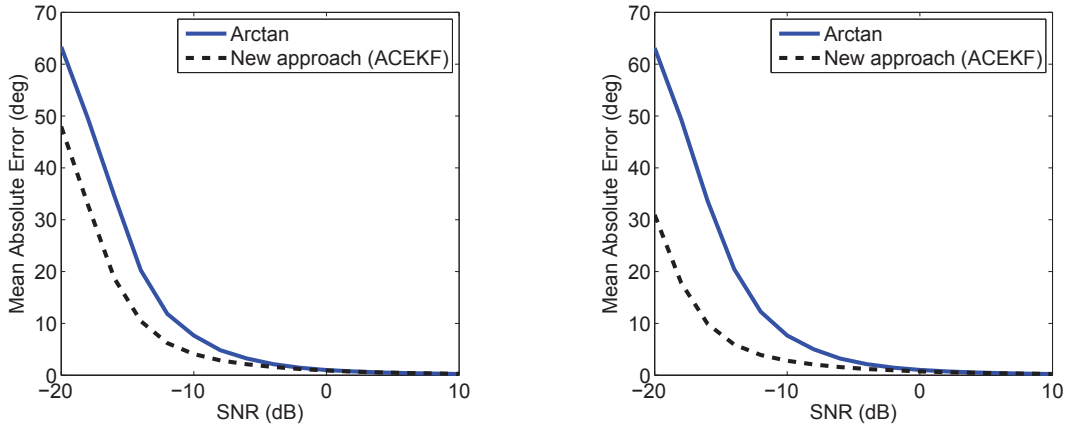
$$s_k = 0.9 s_{k-1} + n_k$$

where n_k is either a white Gaussian or uniform driving noise.

The results are shown in Figure B.3, where again the new ACEKF based algorithm achieved a lower bearing estimation error than the arctan estimator, for both Gaussian and uniform driving noises. Observe that the performance of the arctan estimator was similar in all the simulations, while the performance of the proposed approach was superior because fully exploits the correlation structure of the signals.

B.3 Conclusion

In this paper, we have proposed a new augmented (widely linear) complex state space solution for the DIFAR sonobuoy bearing estimation problem, with the aim of catering for the correlations in target source signals. This was archived through random-walk modelling of the source signal. It has been shown that the second order statistics of the state and observation noises can be estimated and updated online using the observation



(a) Gaussian driving noise

(b) Uniform driving noise

Figure B.3: Performance comparison between the proposed augmented complex state space approach and the arctan estimator for the case where source signal is an autoregressive process with (a) a Gaussian; (b) a uniform driving noise.

data; this together with the augmented state space model nature of our solution enables online tracking of target bearings. The enhanced performance of the proposed approach over the standard arctan bearing estimator has been illustrated for the cases where the source signals are sinusoidal and autoregressive processes.

Bibliography

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

[2] D. H. Dini and D. P. Mandic, "Widely linear modeling for frequency estimation in unbalanced three-phase power systems," *IEEE Transactions onInstrumentation and Measurement*, vol. 62, no. 2, pp. 353–363, 2013.

[3] D. H. Dini and D. P. Mandic, "Cooperative adaptive estimation of distributed noncircular complex signals," in *Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, pp. 1518–1522, 2012.

[4] D. H. Dini and D. P. Mandic, "Exploiting sparsity in widely linear estimation," in *Proceedings of the Tenth International Symposium on Wireless Communication Systems (ISWCS 2013)*, pp. 1–5, 2013.

[5] D. H. Dini, P. M. Djuric, and D. P. Mandic, "The augmented complex particle filter," *IEEE Transactions on Signal Processing*, vol. 61, no. 17, pp. 4341–4346, 2013.

[6] D. H. Dini and D. P. Mandic, "An enhanced bearing estimation technique for difar sonobuoy underwater target tracking," in *Sensor Signal Processing for Defence (SSPD 2012)*, pp. 1–4, 2012.

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

[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] B. Picinbono and P. Chevalier, "Widely Linear Estimation with Complex Data," *IEEE Transactions on Signal Processing*, vol. 43, no. 8, pp. 2030–2033, 1995.

[10] E. Ollila, "On the circularity of a complex random variable," *IEEE Signal Processing Letters*, vol. 15, pp. 841–844, 2008.

[11] P. Schreier, "Bounds on the degree of impropriety of complex random vectors," *IEEE Signal Processing Letters*, vol. 15, pp. 190–193, 2008.

[12] 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.

[13] P. Schreier, "The degree of impropriety (noncircularity) of complex random vectors," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3909–3912, 2008.

[14] P. Dash, A. Pradhan, and G. Panda, "Frequency estimation of distorted power system signals using extended complex Kalman filter," *IEEE Transactions on Power Delivery*, vol. 14, pp. 761–766, Jul. 1999.

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

[16] F. Cattivelli and A. Sayed, "Distributed nonlinear Kalman filtering with applications to wireless localization," in *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pp. 3522–3525, March 2010.

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

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

[19] D. P. Mandic, S. Still, and S. C. Douglas, "Duality between widely linear and dual channel adaptive filtering," in *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pp. 1745–1748, 2009.

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

[21] K. Kreutz-Delgado, "The Complex Gradient Operator and the CR-Calculus," Tech. Rep. ECE275CG-F05v1.3d, Electrical and Computer Engineering, Jacobs School of Engineering, University of California, San Diego, 2006.

[22] H. L. van Trees, *Detection, Estimation and Modulation Theory*. Wiley, 1968.

[23] P. Tichavsky, C.H. Muravchik, and A. Nehorai, "Posterior Cramer-Rao bounds for discrete-time nonlinear filtering," *IEEE Transactions on Signal Processing*, vol. 46, pp. 1386 –1396, May 1998.

[24] A. van den Bos, "The multivariate complex normal distribution-a generalization," *IEEE Transactions on Information Theory*, vol. 41, pp. 537–539, Mar. 1995.

[25] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Prentice-Hall, 1979.

[26] S. Julier and J. Uhlmann, "A new extension of the Kalman filter to nonlinear systems," *in Proceedings of AeroSense 11th International Symposium on Aerospace/Defense Sensing, Simulation and Controls*, pp. 182–193, 1997.

[27] S. Julier, J. Uhlmann, and H. Durrant-Whyte, "A New Method for the Nonlinear Transformation of Means and Covariances in Filters and Estimators," *IEEE Transactions on Automatic Control*, vol. 45, no. 3, pp. 477–482, 2000.

[28] J. Navarro-Moreno, "ARMA prediction of widely linear systems by using the innovations algorithm," *IEEE Transactions on Signal Processing*, vol. 56, pp. 3061–3068, July 2008.

[29] J.-P. Le Cadre and O. Tremois, "Bearings-only tracking for maneuvering sources," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 34, pp. 179–193, Jan. 1998.

[30] Y. Xiao, R. K. Ward, L. Ma, and A. Ikuta, "A new LMS-based fourier analyzer in the presence of frequency mismatch and applications," *IEEE Transactions on Circuits and Systems*, vol. 52, pp. 230–245, Jan. 2005.

[31] C. H. Huang, C. H. Lee, K. J. Shih, and Y. J. Wang, "Frequency estimation of distorted power system signals using a robust algorithm," *IEEE Transactions on Power Delivery,*, vol. 23, pp. 41–51, Jan. 2008.

[32] P. K. Dash, A. K. Pradhan, and G. Panda, "Frequency estimation of distorted power system signals using extended complex Kalman filter," *IEEE Transactions on Power Delivery*, vol. 14, pp. 761–766, July 1999.

[33] M. Wang and Y. Sun, "A practical, precise method for frequency tracking and phasor estimation," *IEEE Transactions on Power Delivery*, vol. 19, pp. 1547–1552, Oct. 2004.

[34] T. Lobos and J. Rezmer, "Real-time determination of power system frequency," *IEEE Transactions on Instrumentation and Measurement*, vol. 46, pp. 877 –881, Aug. 1997.

[35] P. Tichavsky and A. Nehorai, "Comparative study of four adaptive frequency trackers," *IEEE Transactions on Signal Processing*, vol. 45, pp. 1473–1484, Jun. 1997.

[36] S. Y. Park, Y. S. Song, H. J. Kim, and J. Park, "Improved method for frequency estimation of sampled sinusoidal signals without iteration," *IEEE Transactions on Instrumentation and Measurement,*, vol. 60, pp. 2828–2834, Aug. 2011.

[37] M. Karimi-Ghartemani, H. Karimi, and M. R. Iravani, "A magnitude/phase-locked loop system based on estimation of frequency and in-phase/quadrature-phase amplitudes," *IEEE Transactions on Industrial Electronics*, vol. 51, pp. 511–517, April 2004.

[38] D. Beeman, *Industrial Power System Handbook*. McGraw-Hill, 1955.

[39] Y. Xia and D. P. Mandic, "Widely linear adaptive frequency estimation of unbalanced three-phase power systems," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, pp. 74 –83, Jan. 2012.

[40] S. Haykin, *Kalman Filtering and Neural Networks*. John Wiley, 2001.

[41] S. Sangsuk-Iam and T. E. Bullock, "Behavior of the discrete-time Kalman filter under incorrect noise covariances," in *26th IEEE Conference on Decision and Control*, vol. 26, pp. 1594–1600, Dec. 1987.

[42] P. A. Stadter, A. A. Chacos, R. J. Heins, G. T. Moore, E. A. Olsen, M. S. Asher, and J. O. Bristow, "Confluence of navigation, communication, and control in distributed spacecraft systems," *IEEE Aerospace and Electronic Systems Magazine*, vol. 17, pp. 26–32, May 2002.

[43] D. P. Mandic, M. Golz, A. Kuh, D. Obradovic, and T. Tanaka, *Signal Processing Techniques for Knowledge Extraction and Information Fusion*. Springer, 2008.

[44] F. S. Cattivelli and A. H. Sayed, "Diffusion strategies for distributed Kalman filtering and smoothing," *IEEE Transactions on Automatic Control*, vol. 55, pp. 2069–2084, Sept. 2010.

[45] R. Olfati-Saber, "Flocking for multi-agent dynamic systems: algorithms and theory," *IEEE Transactions on Automatic Control*, vol. 51, pp. 401–420, Mar. 2006.

[46] S. Zhou, M. Zhao, X. Xu, J. Wang, and Y. Yao, "Distributed wireless communication system: A new architecture for future public wireless access," *IEEE Communications Magazine*, vol. 41, pp. 108–113, March 2003.

[47] R. Olfati-Saber, "Distributed Kalman filtering for sensor networks," in *46th IEEE Conference on Decision and Control*, pp. 5492–5498, Dec. 2007.

[48] R. Carli, A. Chiuso, L. Schenato, and S. Zampieri, "Distributed Kalman filtering based on consensus strategies," *IEEE Journal on Selected Areas in Communications*, vol. 26, pp. 622–633, May 2008.

[49] U. A. Khan and J. M. F. Moura, "Distributing the Kalman filter for large-scale systems," *IEEE Transactions on Signal Processing,*, vol. 56, pp. 4919–4935, Oct. 2008.

[50] C. G. Lopes and A. H. Sayed, "Diffusion least-mean squares over adaptive networks: Formulation and performance analysis," *IEEE Transactions on Signal Processing*, vol. 56, pp. 3122–3136, July 2008.

[51] Y. Xia, D. P. Mandic, and A. H. Sayed, "An adaptive diffusion augmented CLMS algorithm for distributed filtering of noncircular complex signals," *IEEE Signal Processing Letters*, vol. 18, pp. 659–662, Nov. 2011.

[52] Z. Gao, H.-Q. Lai, and K. J. R. Liu, "Differential space-time network coding for multi-source cooperative communications," *IEEE Transactions on Communications*, vol. 59, pp. 3146–3157, Nov. 2011.

[53] Y. Mao and M. Wu, "Tracing malicious relays in cooperative wireless communications," *IEEE Transactions on Information Forensics and Security*, vol. 2, pp. 198–212, June 2007.

[54] S. Kar and J. M. F. Moura, "Gossip and distributed Kalman filtering: Weak consensus under weak detectability," *IEEE Transactions on Signal Processing*, vol. 59, pp. 1766–1784, Apr. 2011.

[55] R. Bru, L. Elsner, and M. Neumann, "Convergence of infinite products of matrices and inner-outer iteration schemes," *Electronic Transactions on Numerical Analysis*, vol. 2, pp. 183–193, Dec. 1994.

[56] X.R. Li, Y. Zhu, J. Wang, and C. Han, "Optimal linear estimation fusion : Part I: Unified fusion rules," *IEEE Transactions on Information Theory*, vol. 49, pp. 2192–2208, Sep. 2003.

[57] F. S. Cattivelli and A. H. Sayed, "Distributed nonlinear Kalman filtering with applications to wireless localization," in *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pp. 3522–3525, Mar. 2010.

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

[59] J. Benesty, Jingdong Chen, and Y.A. Huang, "A widely linear distortionless filter for single-channel noise reduction," *IEEE Signal Processing Letters*, vol. 17, no. 5, pp. 469–472, 2010.

[60] W.H. Gerstacker, R. Schober, and A. Lampe, "Equalization with widely linear filtering," in *Proceedings of IEEE International Symposium on Information Theory*, pp. 265–269, 2001.

[61] S. Buzzi, M. Lops, and S. Sardellitti, "Widely linear reception strategies for layered space-time wireless communications," *IEEE Transactions on Signal Processing*, vol. 54, no. 6, pp. 2252–2262, 2006.

[62] G. Tanaka and K. Aihara, "Complex-valued multistate associative memory with nonlinear multilevel functions for gray-level image reconstruction," *IEEE Transactions on Neural Networks*, vol. 20, no. 9, pp. 1463–1473, 2009.

[63] S. Samadi, M. Cetin, and M.A. Masnadi-Shirazi, "Multiple feature-enhanced sar imaging using sparsity in combined dictionaries," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 4, pp. 821–825, 2013.

[64] S. Javidi and D. P. Mandic, "A fast algorithm for blind extraction of smooth complex sources with application in EEG conditioning," in *IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 397–402, 2010.

[65] Y. Xia, L. Li, J. Cao, M. Golz, and D. P. Mandic, "A collaborative filtering approach for quasi-brain-death EEG analysis," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),*, pp. 645–648, 2011.

[66] J.H. Kotecha and P.M. Djurić, "Gaussian particle filtering," *IEEE Transactions on Signal Processing*, vol. 51, pp. 2592–2601, Oct. 2003.

[67] B. Laska, M. Bolić, and R. Goubran, "Particle filter enhancement of speech spectral amplitudes," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, pp. 2155–2167, Nov. 2010.

[68] K. Huber and S. Haykin, "Application of particle filters to MIMO wireless communications," in *IEEE International Conference on Communications 2003*, vol. 4, pp. 2311–2315, May 2003.

[69] X. Zhong, A.B. Premkumar, and A.S. Madhukumar, "Particle filtering for acoustic source tracking in impulsive noise with alpha-stable process," *IEEE Sensors Journal*, vol. 13, no. 2, pp. 589–600, 2013.

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

[71] W. Dang and L.L. Scharf, "Extensions to the theory of widely linear complex Kalman filtering," *IEEE Transactions on Signal Processing*, vol. 60, pp. 6669–6674, Dec. 2012.

[72] 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.

[73] A. van-den-Bos, "The Multivariate Complex Normal Distribution a Generalization," *IEEE Transactions on Information Theory*, vol. 41, no. 2, pp. 537–539, 1995.

[74] M.S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, pp. 174–188, Feb. 2002.

[75] J.H. Kotecha and P.M. Djurić, "Gaussian sum particle filtering," *IEEE Transactions on Signal Processing*, vol. 51, pp. 2602–2612, Oct. 2003.

[76] S. Davies, "Bearing accuracies for arctan processing of crossed dipole arrays," in *Proc. Oceans Conference*, pp. 351–356, 1987.

[77] P. K. Tam and K. T. Wong, "Cramer-Rao bounds for direction finding by an acoustic vector sensor under nonideal gain-phase responses, noncollocation, or nonorthogonal orientation," *IEEE Sensors Journal*, vol. 9, pp. 969–982, Aug. 2009.

[78] J. Georgy, A. Noureldin, and G. R. Mellema, "Clustered mixture particle filter for underwater multitarget tracking in multistatic active sonobuoy systems," *IEEE Transactions on Systems, Man, and Cybernetics, Part C Applications and Reviews*, vol. 42, pp. 547–560, July 2012.

[79] B. H. Maranda, "The statistical accuracy of an arctangent bearing estimator," in *Proc. Oceans Conference*, vol. 4, pp. 2127–2132, 2003.

[80] A. R. Runnalls, "Likelihood function for a simple cardioid sonobuoy," *IEE Proceedings Radar, Sonar and Navigation*, vol. 153, pp. 417–426, Oct. 2006.