

A CLASS OF WIDELY LINEAR COMPLEX-VALUED ADAPTIVE FILTERING ALGORITHMS

by
YILI XIA

A Thesis submitted in fulfilment of requirements for the degree of
Doctor of Philosophy of Imperial College London

Communications & Signal Processing Group
Department of Electrical & Electronic Engineering
Imperial College London
2011

Abstract

A large class of signals encountered in communications, biomedical engineering, renewable energy and power systems are conveniently processed in the complex domain $\mathbb{C}$, where traditional adaptive signal processing in $\mathbb{C}$ is regarded as a straightforward extension of the corresponding algorithms in the real domain $\mathbb{R}$. However, recent advances in widely linear modelling and augmented complex statistics show the suboptimality of such an assumption. In this work, based on the widely linear model, a class of linear and nonlinear adaptive filtering algorithms have been derived to process the generality of complex-valued signals (both second order circular and noncircular) in both noise-free and noisy environments, and their usefulness in real-world applications is demonstrated through case studies.

The focus of this thesis is on the use of augmented second order statistics and widely linear modelling. The so called Augmented Complex Least Mean Square (ACLMS) algorithm has already been extended from the standard CLMS algorithm to perform optimum mean square error(MSE) type of adaptive estimation for the generality of complex-valued signals and has been shown to outperform the CLMS algorithm, however, a theoretical understanding of its performance is still missing. To this end, this thesis first addresses this issue in terms of both convergence analysis and steady state analysis. Next, based on the generalised framework introduced by the derivation of the ACLMS algorithm, a class of widely linear adaptive algorithms have been introduced; these include the Regularised Normalised ACLMS (RNACLMS) algorithm, the Augmented Affine Projection algorithm (AAPA) for linear Finite Impulse Response (FIR) adaptive filters, and also in the context of reservoir computing, for the recently introduced random state space based Echo State Networks (ESNs). Furthermore, the widely linear model has been introduced in the

context of distributed networks, where the individual adaptive filters share information with their neighbours to achieve a cooperative estimation. The enhanced performances of the widely linear algorithms are illustrated in renewable energy and power system applications, in particular, for the prediction of wind profiles and frequency estimation of unbalanced three-phase power systems.

Acknowledgment

I would like to thank my supervisor Dr. Danilo P. Mandic for giving me the opportunity to work in such an exciting and promising research area, for which I feel most fortunate. Without his inspirational guidance and excellent supervision, this work would not have been possible. During the past four years, he has always patiently monitored the progress in my research, provided valuable feedback, and corrected hundreds of my technical drafts. Especially, at the very beginning of my studies, he taught me how to conduct and present research in an objective and efficient manner, for which I am truly grateful.

I would also like to give my gratitude to my colleagues and friends. Dr. Clive Cheong Took has always been there to provide his valuable comments for my questions. I also wish to address my appreciation to Dr. Sulee Goh, Dr. Beth Jelfs, Dr. Soroush Javidi, Dr. David Looney, and Dr. Ling Li for the invaluable discussions we have. Meanwhile, I have enjoyed the regular group meetings with Cheolsoo Park, Naveed ur Rehman, Che Ahmad Bukhari, Mosabber Ahmed, Cyrus Jahanchahi, and Dahir Dini, from whom inspirations have come. Thanks also go to Dr. Yizhou Wang, Jiang Han, Haishi Ning and Kai Luo for providing me a wonderful combination of research with a social life.

I wish to thank European Commission and EPSRC for financial support.

Last but not least, I wish to dedicate this work to my parents for their warm support, to my wife Yusi for her patience and kindness, and to my lovely daughter Tingxuan. Their constant love has always been there throughout my studies.

Publications

To the best of my knowledge, the following are original works in which I have participated.

Journal Publications:

1. 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, no. 11, pp. 659 - 662, 2011.
2. Y. Xia, S. C. Douglas, and D. P. Mandic, “Adaptive Frequency Estimation in Smart Grid Applications”, submitted to *IEEE Signal Processing Magazine, Special Issue on Signal Processing Techniques for Smart Grid*, 2011.
3. L. Li, Y. Xia, B. Jelfs, J. Cao and D. P. Mandic, “Modelling of Brain Consciousness based on Collaborative Adaptive Filters”, accepted in *Neurocomputing*, 2011.
4. Y. Xia, and D. P. Mandic, “Widely Linear Adaptive Frequency Estimation of Unbalanced Three-Phase Power Systems”, accepted in *IEEE Transactions on Instrumentation and Measurement*, 2011.
5. Y. Xia, B. Jelfs, M. M. Van Hulle, J. C. Príncipe, and D. P. Mandic, “An Augmented Echo State Network for Nonlinear Adaptive Filtering of Complex Noncircular Signals”, *IEEE Transactions on Neural Networks*, vol. 22, no. 1, pp. 74-83, 2011.
6. Y. Xia, C. Cheong Took, and D. P. Mandic, “An Augmented Affine Projection Algorithm for the Filtering of Complex Noncircular Signals”, *Signal Processing*, vol. 90, no. 6, pp. 1788-1799, 2010.

Conference Publications:

1. D. H. Dini, Y. Xia, D. P. Mandic and S. C. Douglas, "Widely Linear Kalman Filtering for Frequency Estimation in Unbalanced Three-Phase Power Systems", submitted to *the International Conference on Acoustics, Speech and Signal Processing, ICASSP*, 2012.
2. Y. Xia, S. C. Douglas, and D. P. Mandic, "Widely Linear Adaptive Frequency Estimation in Three-Phase Power Systems under Unbalanced Voltage Sag Conditions", in *proceedings of the International Joint Conference on Neural Networks, IJCNN*, pp. 1700 - 1705, 2011.
3. Y. Xia, L. Li, J. Cao, M. Golz, and D. P. Mandic, "A Collaborative Filtering Approach for Quasi-Brain-Death EEG Analysis", in *proceedings of International Conference on Acoustics, Speech and Signal Processing, ICASSP*, pp. 645-648, 2011.
4. D. P. Mandic, Y. Xia, and S. C. Douglas, "Steady State Analysis of the CLMS and Augmented CLMS Algorithms for Noncircular Complex Signals", in *proceedings of the 44th Asilomar Conference on Signals, Systems and Computers*, pp. 1635-1639, 2010.
5. Y. Xia, S. Javidi, and D. P. Mandic, "A Regularised Normalised Augmented Complex Least Mean Square Algorithm", in *proceedings of the International Symposium on Wireless Communication Systems, ISWCS*, pp. 355-359, 2010.
6. N. Rehman, Y. Xia, and D. P. Mandic, "Application of Multivariate Empirical Mode Decomposition for Seizure Detection in EEG Signals", in *proceedings of the International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC*, pp. 1650-1653, 2010.
7. L. Li, Y. Xia, B. Jelfs, J. Cao and D. P. Mandic, "Modelling of Brain Consciousness Based on Collaborative Adaptive Filters", accepted in *the International Symposium on Neural Networks, IJCNN*, 2010.

8. Y. Xia, C. Cheong Took, S. Javidi, and D. P. Mandic, “A Widely Linear Affine Projection Algorithm”, in *proceedings of the IEEE Workshop on Statistical Signal Processing, SSP*, pp. 373-376, 2009.
9. B. Jelfs, Y. Xia, D. P. Mandic, and S. C. Douglas, “Collaborative Adaptive Filtering in the Complex Domain”, in *proceedings of the IEEE Workshop on Machine Learning for Signal Processing, MLSP* pp. 421-425, 2008.
10. Y. Xia, D. P. Mandic, M. M. Van Hulle, and J. C. Príncipe, “A Complex Echo State Network for Nonlinear Adaptive Filtering”, in *proceedings of the IEEE Workshop on Machine Learning for Signal Processing, MLSP*, pp. 404-408, 2008.

Awards:

1. The paper “Modelling of Brain Consciousness Based on Collaborative Adaptive Filters” has been awarded for the Best Student Paper Award at the IEEE International Symposium on Neural Networks, 2010.

Contents

Abstract	3
Acknowledgment	5
Publications	7
Contents	11
List of Figures	17
List of Tables	23
Statement of Originality	25
Abbreviations	26
Mathematical Notations	29
Chapter 1. Introduction	33
1.1 Overview	33
1.2 Motivation and Aims	39
1.3 Organisation of the Thesis	39
Chapter 2. Background	41
2.1 Widely Linear Modelling	41

2.1.1	The duality between widely linear and dual channel adaptive filtering models	42
2.2	Augmented Complex Statistics	43
2.2.1	Seconder order noncircularity (impropriety)	44
2.2.2	Minimum mean square error (MMSE) analysis of strictly linear and widely linear estimators	45
2.2.3	Quantification of second order noncircularity (improperty)	46
2.3	Degree of noncircularity of the benchmarks and real-world wind profile . . .	47

Chapter 3. Mean Square Performance Analysis of the Augmented CLMS

Algorithm	51
3.1 Introduction	51
3.2 Derivation of ACLMS	53
3.3 Steady state mean square performance of CLMS and ACLMS on noncircular signals	54
3.3.1 Steady state analysis of CLMS algorithm for noncircular signals . .	55
3.3.2 Steady state analysis of ACLMS algorithm for noncircular signals . .	58
3.3.3 Mean and mean square stability analysis of ACLMS algorithm . . .	60
3.4 Simulations	62
3.5 Conclusions	65

Chapter 4. A Regularised Normalised Augmented Complex Least Mean Square Algorithm

4.1	Introduction	67
4.2	Derivation of RNACLMS algorithm	68
4.2.1	Gradient Adaptive Regularisation Factor	71
4.3	Simulations	72

4.4	Conclusions	75
Chapter 5. An Augmented Affine Projection Algorithm		77
5.1	Introduction	77
5.2	Derivation of Augmented APA	79
5.3	Steady state MSE performance of APA and AAPA	82
5.3.1	Steady state analysis of standard APA for noncircular signals	83
5.3.2	Steady state analysis of Augmented APA for noncircular signals . .	88
5.3.3	Comparison of steady state performance of APA and AAPA	91
5.4	Simulations	91
5.5	Conclusions	96
Chapter 6. An Augmented Echo State Network for Nonlinear Adaptive Filtering of Complex Noncircular Signals		97
6.1	Introduction	97
6.2	Derivation of the augmented ESN	100
6.2.1	Standard complex ESN with a linear output mapping	100
6.2.2	Augmented complex ESN with a linear readout neuron	101
6.2.3	Training of augmented ESNs	102
6.2.4	Complex ESN with a nonlinear readout neuron	104
6.2.5	ESNs with an adaptive amplitude of output nonlinearity	106
6.3	Merits of nonlinearity and widely linear model	107
6.4	Simulations	108
6.5	Conclusions	113
Chapter 7. An Adaptive Diffusion Augmented CLMS Algorithm for Distributed Filtering of Noncircular Complex Signals		115

7.1	Introduction	115
7.2	Derivation of D-ACLMS algorithm over cooperative networks	117
7.3	Stability Analysis of D-ACLMS algorithm	119
7.3.1	Performance dependence on the degree of noncircularity	121
7.4	Simulations	122
7.5	Conclusions	124
Chapter 8. Widely Linear Adaptive Frequency Estimation of Unbalanced Three-Phase Power Systems		125
8.1	Introduction	125
8.2	Frequency Estimation Based on Widely Linear Adaptive Filtering	127
8.2.1	The Suboptimality of CLMS for Frequency Estimation of Unbalanced Three-Phase Systems	133
8.3	Unbalanced System Conditions and Classification of Voltage Sags	134
8.4	Simulations	136
8.5	Conclusions	142
Chapter 9. Conclusions and Future Work		143
9.1	Conclusions	143
9.2	Future work	144
Bibliography		149
Appendix A. $\mathbb{CR}$ calculus		163
Appendix B. Derivation of RNCLMS Algorithm		165
Appendix C. Derivation of (8.20)		167
Appendix D. Cramer-Rao Lower Bound for Unbiased Frequency Estimator		

on Unbalanced Three-Phase Power Systems
--

169

List of Figures

2.1	Wind recordings as a complex vector representation using speed and direction information.	48
2.2	Complex wind signal magnitude. Three wind speed regions have been identified as <i>low</i> , <i>medium</i> and <i>high</i>	48
2.3	Geometric view of circularity via “real-imaginary” scatter plots. (a) Circular autoregressive AR(4) process (2.24); (b) Noncircular ARMA process (2.25); (c) Noncircular Ikeda map (2.27); (d) Wind (<i>low</i>) signal; (e) Wind (<i>medium</i>) signal; (f) Wind (<i>high</i>) signal.	49
3.1	Simulated MSE (dash) of ACLMS algorithm as a function of step-size μ for (a) Gaussian input (b) Uniform input.	63
3.2	Comparison of theoretical and simulated steady state MSE of CLMS and ACLMS on noncircular AR signals generated by (a) Gaussian input and (b) Uniform input.	63
3.3	Poles of the characteristic function of the AR(4) model.	64
3.4	Comparison of theoretical and simulated steady state MSE of CLMS and ACLMS on a system identification task. (a) noncircular noise ($\rho = 0.1$), and (b) noncircular noise ($\rho = 0.5$).	64

4.1	Learning curves for the NACLMS and RNACLMS algorithms. (a) noncircular ARMA signal (2.25) and (b) Ikeda map (2.27) with $\mu = 0.1, 1$ and 2.	73
4.2	Learning curves for standard RNCLMS and the proposed RNACLMS algorithms on one step ahead prediction of the noncircular ARMA signal (2.25) and Ikeda map (2.27) with $\mu = 0.1$	74
4.3	Prediction gains R_p of the NACLMS and RNACLMS algorithms against different μ and L , (a) noncircular ARMA signal (2.25), (b) Ikeda map (2.27), (c) wind (<i>low</i>) and wind (<i>medium</i>).	74
5.1	Learning curves of AAPA for the noncircular ARMA model (2.25) and noncircular Ikeda map (2.27) for $K = 1$ and various learning rate μ	92
5.2	Learning curves of APA and AAPA for the noncircular ARMA model (2.25) and noncircular Ikeda map (2.27), obtained by using $\mu = 0.1$ and $K = 1, 2, 4$, and 8.	92
5.3	MSE of AAPA as a function of step-size μ for the noncircular ARMA model (2.25) and Ikeda map (2.27).	93
5.4	Learning curves of APA and AAPA for the noncircular ARMA model (2.25) and noncircular Ikeda map (2.27) using different large μ and $K = 2$	93
5.5	Learning curves of APA and AAPA for the noncircular ARMA model (2.25) and noncircular Ikeda map (2.27) using $\mu = 1.5$ and $K = 2, 4$, and 8.	94
5.6	Learning curves of APA and AAPA for the proper AR(4) model (2.24) using different small and large μ and $K = 1$	94
5.7	Prediction gains of APA and AAPA for varying values of the control parameters of noncircular ARMA (2.25) and Ikeda map (2.27).	95
5.8	Comparison of theoretical and simulated steady state MSE of AAPA on noncircular ARMA (2.25) when $K = 1, 2$ and 3 using (a) (5.51) and (b) (5.52).	95

6.1	The architecture of an Echo State Network (ESN).	100
6.2	Probability density functions obtained from a doubly white circular noise after applying the widely linear model and tanh nonlinearity.	107
6.3	Mean Square Errors (MSE) of one-step ahead prediction for augmented and standard ESN trained by the AACNGD algorithm.	110
6.4	Comparison of performances of standard and augmented ESNs trained by different algorithms, over a range of degrees of connectivity on one-step ahead prediction of the wind (<i>medium</i>) and (<i>high</i>) signals.	111
6.5	Comparison of performances of standard and augmented ESNs trained by different algorithms, over a range of spectral radii on one-step ahead pre- diction of the noncircular wind (<i>medium</i>) and (<i>high</i>) signals.	111
6.6	Comparison of performances of standard and augmented ESNs trained by different algorithms, over a range of reservoir size on one-step ahead pre- diction of the noncircular ARMA and wind (<i>medium</i>) signals.	112
6.7	Comparison of performances of standard and augmented ESNs trained by different algorithms on multiple-step ahead prediction of the wind (<i>medium</i>) and (<i>high</i>) signals.	112
7.1	Two modes of cooperation strategy for a distributed network: (a) Incre- mental strategy, and (b) Diffusion strategy.	116
7.2	The topology of a distributed network with N nodes	118
7.3	Network topology used in the simulations.	122
7.4	Comparison of MSEs of all algorithms on noncircular signals (a) Noncircular ARMA and (b) Ikeda map.	123

- 8.1 Geometric view of circularity and noncircularity via a “real-imaginary” scatter plot in the complex plane. The circle denoted by ‘+’ represents a circular complex-valued signal $\hat{v}(k)$ obtained from a balanced situation where $V_a(k), V_b(k), V_c(k)$ are identical at 1-p.u. (per-unit) and time invariant. The ellipse denoted by ‘o’ represents a noncircular complex-valued $\hat{v}(k)$ obtained in an unbalanced condition with $V_a(k) = 1\text{-p.u.}$, $V_b(k) = 0.7\text{-p.u.}$ and $V_c(k) = 0.5\text{-p.u.}$ 129
- 8.2 Three-phase unbalanced sags due to two-phase-to-ground faults with characteristic complex voltage $V = 0.7$. (a) Phasor diagram (dotted line denotes normal operation). (b) Circularity via a “real-imaginary” plot. 134
- 8.3 Three-phase unbalanced sags due to two phase-to-ground faults with characteristic complex voltage $V = 0.7$. (a) Phasor diagram (dotted line denotes normal operation). (b) Circularity via a “real-imaginary” plot. 135
- 8.4 Time series of three-phase voltages for different sag types. 136
- 8.5 Frequency estimation using CLMS and ACLMS, for $\mu = 0.01$ when the three-phase power system undergoing three consecutive voltage sags of different natures. 137
- 8.6 Frequency estimation for the unbalanced three-phase voltage with SNR = 40dB. (a) The angle of $w(k)$ used in CLMS method, (b) Frequency estimation using CLMS with asin and atan functions, (c) The angle of $h(k) + a_1(k)g(k)$, (d) Frequency estimation using ACLMS with asin and atan functions. 137
- 8.7 Comparison of the proposed ACLMS and CLMS at different SNR, obtained by averaging 1000 independent trials. (a) bias error. (b) variance. 139
- 8.8 Frequency estimation for phase voltages contaminated with harmonics. The mean percentage error of CLMS and ACLMS algorithms over a range of amplitudes of both third harmonic and fifth harmonic at the steady-state. . 139

-
- 8.9 Frequency estimation using CLMS and ACLMS, for $\mu = 0.01$, when system
frequency undergoing the decay and rise at a rate of 5 Hz/sec. 140
- 8.10 Frequency estimation for real-world unbalanced three-phase voltage. . . . 141

List of Tables

2.1	Comparison of degrees of noncircularity s for the various classes of signals	49
3.1	Stability bounds of ACLMS algorithm (Gaussian input and uniform input)	62
4.1	Comparison of prediction gains R_p for the various classes of signals	73
5.1	Prediction gains R_p of standard APA and the proposed AAPA for various noncircular signals	91
6.1	Comparison of prediction gains R_p and their standard deviations (in bracket) for the various classes of signals averaged over 200 independent initialisations of ESN	108
6.2	Comparison of prediction gains R_p of previous discussed algorithms in Chapters 3 and 4 for the considered signals (200 independent trials on synthetic data and single trial on real world wind data).	108
6.3	The percentage of enhanced performance of augmented ESN algorithms for complex noncircular signals	110
7.1	Comparison of prediction gains R_p for the various classes of circular and noncircular signals	123
8.1	Four types of sags due to signal phase-to-ground fault in equation form . .	134
8.2	Three types of sags due to two phase-to-ground faults in equation form . .	134

8.3	Cauases of three-phase voltage sags and system unbalance	135
8.4	Mean square error (in Hz^2) at the steady state of frequency estimation using CLMS and ACLMS with asin and atan functions. The results were obtained by averaging 100 independent trails.	138

Statement of Originality

Substantial parts of this thesis are believed to be original contributions to the implementation of complex-valued linear and nonlinear adaptive filters and algorithms in the literature. As far as the author is aware, the following aspects of the thesis are believed to be original contributions:

- A class of augmented complex-valued algorithms for linear and nonlinear adaptive filters
- Mean square performance analysis on augmented complex-valued algorithms
- A widely linear adaptive frequency estimator for unbalanced three-phase power systems
- A widely linear diffusion ACLMS algorithm for distributed filtering of noncircular complex signals

Abbreviations

AACNGD:	Adaptive amplitude complex nonlinear gradient descent
AAPA:	Augmented affine projection algorithm
ACEKF:	Augmented complex Kalman filter
ACLMS:	Augmented complex least mean square
ACRTRL:	Augmented complex real time recurrent learning
AESN:	Augmented echo state network
AF:	Activation function
APA:	Affine projection algorithm
AR:	Autoregressive
ARMA:	Autoregressive moving average
ASK:	Amplitude shift keying
BPSK:	Binary phase shift keying
BPSK:	Binary phase shift keying
CLMS:	Complex least mean square
CNGD:	Complex nonlinear gradient descent
CRLB:	Cramer-Rao lower bound
CRTRL:	Complex real time recurrent learning
D-ACLMS:	Diffusion augmented complex least mean square
D-CLMS:	Diffusion complex least mean square
DS-CDMA:	Direct sequence code division multiple access
EEG:	Electroencephalogram
EKF:	Extended Kalman filter
EMSE:	Excess mean square error

ESN:	Echo state network
FIR:	Finite impulse response
GNGD:	Generalised normalised gradient descent
MMSE:	Minimum mean square error
MSE:	Mean square error
NACLMS:	Normalised augmented complex least mean square
NCLMS:	Normalised complex least mean square
NLMS:	Normalised least mean square
PDF:	Probability density function
PLL:	Phase locked loop
RHS:	Right hand side
RLS:	Recursive least squares
RNACLMS:	Regularised Normalised augmented complex least mean square
RNCLMS:	Regularised Normalised complex least mean square
WL-CLMS:	widely linear complex least mean square

Mathematical Notations

$\otimes$	Kronecker product
$ \cdot $	Modulus operator
$\ \cdot\ $	Vector or matrix norm
$\ \cdot\ ^2$	Euclidean norm
$(\cdot)^*$	Complex conjugate operator
$(\cdot)^{-1}$	Matrix inverse operator
$(\cdot)^T$	Vector or matrix transpose operator
$(\cdot)^H$	Complex conjugate transpose (Hermitian) operator
$(\cdot)_i$	Imaginary part of a complex-valued random variable
$(\cdot)_r$	Real part of a complex-valued random variable
$\lambda(\cdot)$	Eigenvalues of a matrix
Δ	Discriminant of a quadratic equation
λ	Lagrange multiplier
Λ	Lagrange multipliers
ΔT	Sampling interval
$\triangleq$	Defined as
∇	Gradient operator
∂	Partial derivative operator
$\mathbf{1}$	Vector or matrix with all unit elements

$\cos$	Cosine function
$\mathbb{C}$	Complex domain
$\mathbf{C}_{\mathbf{x}\mathbf{x}}$	Covariance matrix of a random vector $\mathbf{x}$
$\mathbf{C}_{\mathbf{x}^a\mathbf{x}^a}$	Augmented covariance matrix of random vector $\mathbf{x}$
d	Desired signal
$\det(\cdot)$	Matrix determinant operator
$\text{diag}(\cdot)$	Diagonal matrix of elements
e	Instantaneous output error
$E[\cdot]$	Expectation operator
$E[y x]$	Conditional expectation of y given x
f	System frequency
$\hat{f}$	Estimated system frequency
$\mathbf{g}$	Filter coefficient vector
$\mathbf{G}$	Transition matrix
$\mathbf{h}$	Filter coefficient vector
$\mathbf{I}$	Identity matrix
$\Im[\cdot]$	Imaginary part of a complex-valued number
j	$\sqrt{-1}$
J	Cost function
k	Discrete time instant
K	Number of observations
$\lim$	Limit
$n(k)$	Noise at time instant k
$\mathbf{n}(k)$	Noise vector at time instant k
$p[\mathbf{x}]$	Probability density function of a random vector $\mathbf{x}$
$\mathbf{P}_{\mathbf{x}\mathbf{x}}$	Pseudo-covariance matrix of a random vector $\mathbf{x}$

r	Degree of noncircularity
$\mathbb{R}$	Real domain
$\Re[\cdot]$	Real part of a complex-valued number
s	Degree of noncircularity
$\sin$	Sine function
$\text{Tr}(\cdot)$	Matrix trace operator
$\mathbf{w}$	Augmented filter coefficient vector
$\mathbf{w}^a$	Augmented filter coefficient vector
$\mathbf{x}(k)$	Input vector at time instant k
$\mathbf{X}(k)$	Input matrix at time instant k
$y(k)$	Output at time instant k
$y_i(k)$	i th output at time instant k
$\mathbf{y}(k)$	Output vector at time instant k
δ	Regularisation term
∞	Infinity
$\int$	Integration operator
μ	Step size
ω	Angular frequency
ρ	Step size
$\rho(x)$	Circularity quotient of a random variable x
σ_x^2	Variance of a random variable x
$\sum$	Summation operator
τ_x^2	Pseudo-variance of a random variable x

Chapter 1

Introduction

1.1 Overview

ADAPTIVE signal processing has found a number of applications, such as in system identification, noise cancellation, signal enhancement, and adaptive prediction [1–4]. Compared with digital filters with fixed coefficients which require static scenarios and predefined specifications, adaptive filters require no assumptions on the signal generating mechanism, and can be applied to situations where the specifications are not available or are time-variant [4].

The best known linear adaptive filtering algorithm is the least mean square (LMS) algorithm, derived by Widrow and Hoff, based on a recursive solution of least squares estimation [5]. Much research has been dedicated to the analysis of LMS with the aim to improve its performance based on the investigation of the step-size μ , which affects the convergence speed, steady state error, and stability of the adaptive filter, due to the fact that a fixed μ may not be optimal to respond to time-variant environments. Ideally, an adaptive algorithm is highly demanded with fast convergence speed and small steady state error when operating in a stationary environment, and adjusted step-size μ according to the dynamics of the input signal in nonstationary environment. To this end, a class of variable step-size LMS algorithms with the aim to adapt the LMS step-size μ in a linear stochastic fashion have been proposed [6–8]. Alternatively, the normalised LMS (NLMS) algorithm, which incorporates the step-size μ divided by the squared norm of the input

regressor [2], can be used for this purpose. However, in practice applications, to avoid the situation in which the norm of the input regressor gets close to zero, a small positive constant δ called the regularisation term is usually added [2]. To deal with signals with unknown and possibly very large dynamical range, the Generalised Normalised Gradient Descent (GNGD) algorithm has been proposed by employing a time-varying regularization parameter and updating it at every iteration. In this case, the GNGD algorithm adapts μ in a nonlinear manner, providing both improved stability and enhanced performance as compared with variable step-size LMS algorithms [9].

There are situations where it is possible to recycle the old data vector in order to improve convergence of the adaptive filtering algorithms. Data reusing algorithms, such as data reusing LMS [10], Affine Projection algorithm (APA) and its variants [11, 12], are considered an alternative to increase the speed of convergence in adaptive filtering algorithms in such situations. The penalty to be paid by data reusing is an increased estimation error in the steady state, and as usual the trade-off between steady state performance and convergence speed is achieved through the step-size.

The performance of an adaptive filter is measured in terms of its transient behaviour and its steady-state behaviour. The transient analysis is focused on the stability and the convergence rate of an adaptive filter, whereas the steady-state analysis provides information on the mean-square-error (MSE) of the filter when it reaches steady state. In the literature, there have been many works dedicated to the performance analysis of adaptive filters. Traditionally, the transient analysis was first investigated, and subsequently, steady-state analysis is deduced as the limiting case of a transient analysis [1, 4, 13, 14]. Although this procedure is versatile for simultaneously understanding both the steady-state and the transient behaviours of an adaptive algorithm, several difficulties are unavoidably encountered. Firstly, traditional transient analysis, involving recursion estimation for the weight-error variance, leads to a large number of computations, especially for adaptive filters with nonlinear update equations, such as NLMS and APA. Secondly, transient analysis requires some simplified assumptions as compared with steady-state analysis, hence, treating steady-state analysis as a fallout of a transient analysis results in suboptimality. To this end, a unified approach to the steady-state performance of a large class of adaptive

filters that bypasses several of the mentioned difficulties, called energy conservation principle, was proposed [15]. Based on the study of the energy flow through each iteration of an adaptive filter and a fundamental error variance relation that avoids the weight-error variance recursion, a steady-state analysis can be computationally efficiently obtained under weaker assumptions than those required in the previous approaches, and it has been successfully applied to several classes of adaptive filters with linear and nonlinear weight update equations [15–18].

The theory of linear adaptive filters is well established, whereas architectures and algorithms for nonlinear adaptive filtering are still emerging. Some recent results have shown that recurrent neural networks (RNNs) with feedback are powerful tools for nonlinear adaptive filtering of real-world data, due to their ability to uniformly approximate any continuous function on a compact domain [3, 19]. Given their ability to learn from examples, the application of neural networks in the area of nonlinear adaptive prediction and modelling offers potentially better performance as compared to standard statistical and linear filtering methods. In this context, the recurrent backpropagation (BP), the real time recurrent learning (RTRL), and extended Kalman filter (EKF) approaches have been developed for the training of RNNs. However, the main issue involved in the design and training of RNNs lies in the computation complexity associated with the updates of the weight matrix, and suffer from a variety of problems, such as the slow training resulting from the computational complexity and the possibility of instability. To address some of these problems, a class of discrete-time RNNs, called Echo State Networks (ESNs), has been introduced [20]. By employing a sparse and fixed hidden layer and by training only the weights connecting the internal layer to the readout neurons, the ESN approach significantly reduces the complexity of the representation associated with standard RNNs.

In modern disciplines, complex-valued signals arise frequently in applications as diverse as communications, biomedical and renewable energy. The complex domain $\mathbb{C}$ not only provides a convenient representation for these signals but also a natural way to represent the physical characteristics of the signals and the transformations they go through, such as the phase and magnitude distortion experienced by complex-valued communications signals, such as those using binary phase shift keying (BPSK), quadrature phase

shift keying (QPSK), and quadrature amplitude modulation (QAM). In all these cases, it is optimal to carry out signal processing directly in the complex domain to guarantee that the complete information, which is represented by the interrelationship of the real and imaginary parts or the magnitude and phase of the complex-valued signal, can be fully exploited. For instance, the white and gray matter within the magnitude of structural magnetic resonance imaging (MRI) data can be combined to form a complex number to make use of their interdependence [21]. In applications of adaptive systems, it is the signal magnitude which is considered as the main source of information [22]. Although this facilitates the established algorithms in $\mathbb{R}$, it is important to realise that this way the full information potential (phase information) within the signal is not exploited. Indeed real world processes with the “intensity” and “direction” component (radar, sonar, vector fields) require the consideration of such phase information which can be obtained by complex-valued representation. For instance recent results on the modelling of wind profile, indicate that the processing in (simultaneously speed and direction as a complex vector) has major benefits over the direct processing in $\mathbb{R}$ (speed only), or in $\mathbb{R}^2$ (speed and direction as independent processes) [23].

Properties of complex-valued signals are not only determined by their statistical nature but also in terms of their bivariate or complex nature, due to the fact that a complex signal can be represented either by its real and imaginary, or phase and amplitude components. To this end, adaptive processing of complex-valued signals can then be performed using three different approaches. Firstly, the real and imaginary components (or phase and amplitude) are considered as dual univariate signals and processed separately. Secondly, the two components can be considered as a real-valued bivariate signal and processed using a suitable real-valued two-dimensional algorithm. Alternatively, it would be natural to consider the signal in the complex domain $\mathbb{C}$, where complex-valued algorithms can be designed directly.

To directly operate on the complex-valued signals, in 1975 Widrow *et al.* introduced the complex Least Mean Square (CLMS) algorithm [24]. More recently, the design of complex-valued linear adaptive filtering algorithms and their performance analysis have been an active research area and a number of algorithms have been extended to the com-

plex domain $\mathbb{C}$ [18, 25], whereas in the field of RNNs, extensions of real-valued algorithms to the complex domain are not straightforward. One of the major issues to be solved is the design of a suitable complex nonlinear activation function (AF). According to Liouville's theorem, the only bounded and analytic function in $\mathbb{C}$ is constant [26], and to that cause, for convinience, previous studies have mostly focused on the so-called split-complex activation functions (AF). However, split-complex AFs are not analytic, where the Cauchy-Riemann equations do not apply, consequently, they are not universal approximators and algorithms based on them underperform in applications where complex signals exhibit strong component correlations. To this end, meromorphic functions have been introduced as activation functions, due to their property that they are analytic everywhere except for a discrete subset of $\mathbb{C}$. At their singular points, these functions tend to infinity, thus removing the possibility of encountering essential singularities. This enables to design fully-complex RNNs to deal with complex-valued signals directly and direct extensions of real-valued algorithms to the complex domain include the class of complex backpropagation algorithms, complex RTRL (CRTRL) [27] and complex EKF (CEKF).

Traditionally, in the signal processing literature, statistics in the complex domain $\mathbb{C}$ is regarded as an extension of that in $\mathbb{R}$, and consequently complex-valued algorithms were considered as simple extensions of the established corresponding algorithms in the real domain. In particular, statistical modelling of complex-valued random vectors was taken as straightforward extensions from $\mathbb{R}$. For instance, the covariance matrix of a zero mean complex vector $\mathbf{x}$, $E[\mathbf{x}\mathbf{x}^H]$, is transformed from the covariance matrix $E[\mathbf{x}\mathbf{x}^T]$ in $\mathbb{R}$, achieved by replacing the transpose operator $(\cdot)^T$ by the conjugate transpose (Hermitian) operator $(\cdot)^H$. In this manner, the distribution of a complex-valued random vector is implicitly assumed to be circular and rotation invariant in the complex domain, and indicates the independence between the real and imaginary components with the complex-valued signals. Thus, the aforementioned complex-valued algorithms based on this implicit assumption are only optimal for a subset of complex signals, those with a circularly symmetric distribution, and are not suitable to deal with the generality of complex-valued signals.

To this end, analysis on the properties of the generality of complex-valued signals

has become an active research area. Fundamental work by Picinbono, Neeser and Massey [28,29] proposed the concept of complex circularity based on the second order statistics of complex-valued random variables. In their analysis, a complex-valued random variable is defined to be second order circular (proper) if its distribution is rotation invariant, and is otherwise noncircular (improper). It was shown that the conventional covariance matrix cannot fully describe the statistics of noncircular signals and the pseudocovariance matrix is necessarily to be considered to fully capture the relation between the real and imaginary components of random vectors. This implies that in order to model the complete second order information available with in the generality of complex-valued signals, both the covariance and pseudocovariance matrices should be considered, and only in a special case where complex random variables are second order circular (proper), the pseudocovariance matrix vanishes.

The use of this augmented complex statistics has opened the possibility to design adaptive filtering algorithms suitable for processing both circular and noncircular signals. Such algorithms are based on a widely linear model, and are usually called “widely linear” or “augmented” algorithms. Examples include the work by Valkama *et al.* who have introduced widely linear algorithms in wireless communication applications to solve the complex-valued I/Q mismatch problem [30]. Other examples include the widely linear Complex Least Mean Square (WL-CLMS) algorithm for direct sequence code division multiple access (DS-CDMA) applications by Schober *et al.* [31], which used a real valued error. It was subsequently extended to the augmented CLMS (ACLMS) algorithm in [32] in order to cater for both complex signals and complex errors, and also by Kuh and Mandic in the context of complex augmented kernels [33]. The extensions of real-valued recurrent neural network structures to those in the complex domain based on widely linear models have been recently designed. Examples include the augmented complex real-time recurrent learning (ACRTRL) algorithm [34] and the augmented complex-valued extended Kalman filter (ACEKF) [35] for complex recurrent neural networks. These augmented algorithms were shown to outperform their standard complex counterparts for the generality of complex-valued signals. The performance of complex recurrent neural networks were compared for the task of wind profile prediction using a dual univariate model and

a complex model in [23, 27], where the complex representation of wind resulted in better performance when predicted using a trained CRTRL algorithm. In comparison, the ACRTRL algorithm achieved a better prediction performance, highlighting the associated benefits of considering augmented complex statistics.

1.2 Motivation and Aims

The focus of this work is on the generalisation of existing filtering algorithms in $\mathbb{C}$ using augmented complex statistics and widely linear modelling, to make them second order optimal for the processing of the circular and noncircular signals. Several contributions to adaptive signal processing of noncircular signals will be presented:

- Enlightened by previous research on the design of ACLMS algorithm, a class of linear and nonlinear adaptive filtering algorithms are derived based on the widely linear model and augmented complex statistics.
- Although the advantage of augmented algorithms over their strictly linear counterparts have been supported by both theory and simulations, a rigorous analysis on quantifying this advantage is still missing. To this end, mean square statistical performance analysis on CLMS and ACLMS, APA and AAPA algorithms to deal with noncircular signals is investigated.
- It is natural to design algorithms that can be efficiently applied in real-world, and this thesis provides solutions that demonstrate the suitability of the introduced algorithms to a variety of real world problems such as wind prediction and frequency estimation of three-phase power systems based on the convenient complex-valued representation of the real-world signals.

1.3 Organisation of the Thesis

The organisation of the rest of the thesis is as follows. Chapter 2 introduces the widely linear model and augmented complex statistics, which forms the statistical framework for the rest of work in this thesis. In Chapter 3, the ACLMS, a workhorse in $\mathbb{C}$ to deal with

the filtering of general complex-valued signals, is revisited, and its advantage over strictly CLMS is quantified based on the mean square performance analysis. In Chapter 4, the Normalised ACLMS (NACLMS) and Regularised NACLMS (RNACLMS) are introduced to deal with the situations where the value of the weight update term becomes extremely large due to close-to-zero input vectors. Chapter 5 derives the augmented affine projection (AAP) algorithm based on the data-reusing technique and the statistical performance analysis in Chapter 2 is extended. Chapter 6 introduces a computationally efficient augmented echo state networks (AESNs) for nonlinear adaptive filtering of noncircular signals. In Chapter 7, the widely linear model is used to derive an adaptive diffusion ACLMS (D-ACLMS) algorithm for collaborative processing of the generality of complex signals over distributed networks. In Chapter 8, the usefulness of the augmented adaptive filters is investigated in a real-world application, that is, frequency estimation of unbalanced three-phase power system. We first illustrate that the unbalanced three-phase voltage is second order noncircular and an adequate frequency estimator is derived based on widely linear model to eliminate frequency estimation oscillations which is always encountered by strictly linear estimator due to submodelling. Conclusions and future work are given in Chapter 9.

Several key concepts and algorithm derivations related to this work are highlighted in appendices. Appendix A provides an overview of the $\mathbb{C}\mathbb{R}$ calculus framework for the gradient calculation of real-valued and complex-valued functions in $\mathbb{C}$ with respect to complex variables. A detailed derivation of RNCLMS algorithm is presented in Appendix B. The second order noncircular (improper) nature of complex-valued voltage from unbalanced three-phase power systems is proved in Appendix C, while the Cramer-Rao Lower Bound (CRLB) on the estimation variance of unbiased frequency estimator under unbalanced conditions is derived in Appendix D and will be used as benchmark to assess the statistical performance of the proposed widely linear adaptive frequency estimator.

Chapter 2

Background

In this chapter, we introduce the motivation to use the widely linear model and augmented complex statistics for complex-valued linear estimation problems.

2.1 Widely Linear Modelling

Consider the real-valued mean square error (MSE) estimator

$$\hat{y} = E[y|x] \quad (2.1)$$

which estimates the values of signal y in terms of another observation x . For zero mean, jointly normal y and x , the linear model solution is

$$\hat{y} = \mathbf{x}^T \mathbf{h} \quad (2.2)$$

where $\mathbf{h} = [h_1, \dots, h_N]^T$ is a vector of fixed filter coefficients, and the past of the observed variable is contained in the regressor vector $\mathbf{x} = [x_1, \dots, x_N]^T$.

In the complex domain $\mathbb{C}$, it is assumed that we can use the same form of conditional mean estimator that for real-valued signals given in (2.1), leading to the standard complex linear MMSE estimator¹

$$\hat{y}_L = \mathbf{x}^H \mathbf{h} \quad (2.3)$$

¹Both $y = \mathbf{x}^T \mathbf{h}$ and $y = \mathbf{x}^H \mathbf{h}$ are correct yielding the same output and the mutually conjugate coefficient vectors. The latter form is more common and the former was used in the original CLMS paper [24]; in this work we will use the first form.

where symbol $(\cdot)^H$ denotes the Hermitian transform operator. However, the real-valued linear estimator in (2.2) applies to both the real and imaginary parts of complex variables, that is

$$\begin{aligned}\hat{y}_r &= E[y_r|x_r, x_i] \\ \hat{y}_i &= E[y_i|x_r, x_i]\end{aligned}\tag{2.4}$$

and a more general MSE estimator than that in (2.3) can be expressed as

$$\hat{y} = E[y_r|x_r, x_i] + jE[y_i|x_r, x_i]\tag{2.5}$$

Upon employing the identities $x_r = (x + x^*)/2$ and $x_i = (x - x^*)/2j$ we arrive at

$$\hat{y} = E[y_r|x, x^*] + jE[y_i|x, x^*]\tag{2.6}$$

leading to a *widely linear* (WL) estimator for complex valued data, given by

$$\hat{y}_{\text{WL}} = \mathbf{x}^T \mathbf{h} + \mathbf{x}^H \mathbf{g} = \mathbf{x}^{aT} \mathbf{w}^a\tag{2.7}$$

where $\mathbf{h}$ and $\mathbf{g}$ are complex-valued coefficient vectors. The widely linear model can also be expressed using augmented vectors $\mathbf{w}^a = [\mathbf{h}^T, \mathbf{g}^T]^T$ and $\mathbf{x}^{aT} = [\mathbf{x}^T, \mathbf{x}^H]^T$, which provides a more compact representation. This estimator is suitable for linear MMSE estimation of the generality of complex valued processes (both circular and noncircular) [36], as it accounts for complete second order information in $\mathbb{C}$, as will be shown in Section 2.2.

2.1.1 The duality between widely linear and dual channel adaptive filtering models

Consider the widely linear model in $\mathbb{C}$ as its counterpart in $\mathbb{R}^2$ via the one-to-one mapping between a point in the complex plane $\mathbb{C}$ and a point in $\mathbb{R}^2$, (2.7) can be expressed in the

form of a dual channel modelling as

$$\hat{y}_{\text{WL}_r} = \mathbf{x}_r^T \mathbf{a} + \mathbf{x}_i^T \mathbf{b} \quad (2.8)$$

$$\hat{y}_{\text{WL}_i} = \mathbf{x}_r^T \mathbf{c} + \mathbf{x}_i^T \mathbf{d} \quad (2.9)$$

where the real-valued weight vectors can be derived as

$$\mathbf{a} = \mathbf{h}_r + \mathbf{g}_r \quad \mathbf{b} = \mathbf{g}_i - \mathbf{h}_i \quad (2.10)$$

$$\mathbf{c} = \mathbf{h}_i + \mathbf{g}_i \quad \mathbf{d} = \mathbf{h}_r - \mathbf{g}_r \quad (2.11)$$

Therefore, the adaptive filters based on the widely linear model in $\mathbb{C}$ are isomorphic to their counterparts using the dual channel modelling in $\mathbb{R}^2$ [37]. However, by virtue of the complex-valued representation, the widely linear model based estimation can be written into a more compact form as $\mathbf{x}^{aT} \mathbf{w}^a$ rather than as a combination of four separate real-valued linear models. This gives mathematical convenience to deal with the theoretical performance analysis (both convergence and steady state) on adaptive filters where the derivation can be simplified by considering only a single weight vector $\mathbf{w}^a$. However, from a hardware implementation point of view, dual channel representation in $\mathbb{R}^2$ is usually preferable over widely linear representation in $\mathbb{C}$ because the former requires fewer real operations in terms of additions and multiplications than the latter.

2.2 Augmented Complex Statistics

From (2.7), it is clear that the covariance matrix, $\mathbf{C}_{\mathbf{xx}} = E[\mathbf{xx}^H]$, alone does not have sufficient degrees of freedom to describe full second order statistics in $\mathbb{C}$ [29], and in order to make use of all the available statistical information we also need to consider the pseudo-covariance matrix, $\mathbf{P}_{\mathbf{xx}} = E[\mathbf{xx}^T]$. In practice, to account for the improperness, the input vector $\mathbf{x}$ is concatenated with its conjugate $\mathbf{x}^*$, to produce an augmented $2N \times 1$ input vector $\mathbf{x}^{aT}$. This augmented input, together with the augmented weights $\mathbf{w}^a = [\mathbf{h}^T, \mathbf{g}^T]^T$, forms a widely linear estimate in (2.7) and its $2N \times 2N$ augmented covariance matrix is

given by [38]

$$\mathbf{C}_{\mathbf{x}^a \mathbf{x}^a} = E \begin{bmatrix} \mathbf{x} \\ \mathbf{x}^* \end{bmatrix} [\mathbf{x}^H \mathbf{x}^T] = \begin{bmatrix} \mathbf{C}_{\mathbf{x}\mathbf{x}} & \mathbf{P}_{\mathbf{x}\mathbf{x}} \\ \mathbf{P}_{\mathbf{x}\mathbf{x}}^* & \mathbf{C}_{\mathbf{x}\mathbf{x}}^* \end{bmatrix} \quad (2.12)$$

This matrix now contains the complete complex second-order statistical information available in the complex domain.

2.2.1 Second order noncircularity (impropriety)

Processes whose second order statistics can be completely described by the covariance matrix, that is, for which the pseudo-covariance matrix $\mathbf{P}_{\mathbf{x}\mathbf{x}} = \mathbf{0}$, are termed second order *circular* (or *proper*), or otherwise termed *noncircular* or *improper*. In general, the notion of circularity extends beyond second order statistics, to describe the class of signals with rotation-invariant distributions, $p[\cdot]$ for which $p[\mathbf{x}] = p[\mathbf{x}e^{j\theta}]$ for $\theta \in [0, 2\pi)$, whereas properness (also, propriety or second order circularity) specifically refers to the second order statistical properties. In most real world applications, complex signals are second order noncircular or improper, and their probability density functions are not rotation invariant.

For a complex-valued random scalar x with zero mean, the standard variance can be seen as the correlation of x and its complex conjugate x^* and is denoted by

$$\sigma_x^2 = E[xx^*] = E[|x|^2] \quad (2.13)$$

and the pseudo-variance is denoted by

$$\tau_x^2 = E[xx] = E[x^2] \quad (2.14)$$

Note that while the variance σ_x^2 is real-valued, the pseudo-variance τ_x^2 is normally complex-valued.

2.2.2 Minimum mean square error (MMSE) analysis of strictly linear and widely linear estimators

The MMSE of the widely linear approach using (2.7) can be derived as

$$E[|e_{\text{WL}}|^2] = E[|y - \hat{y}_{\text{WL}}|^2] \quad (2.15)$$

Using the compact form of $\hat{y}_{\text{WL}}$ in (2.7), the well-known Wiener-Hopf solution can be obtained as

$$\mathbf{w}^a = \mathbf{C}_{\mathbf{x}^a \mathbf{x}^a}^{-1} \mathbf{p}_{\hat{y}, \mathbf{x}^a} \quad (2.16)$$

where $\mathbf{p}_{\hat{y}, \mathbf{x}^a} = E[\hat{y}^* \mathbf{x}^a] \triangleq [\mathbf{c}_1^T, \mathbf{c}_2^T]^T$ is the cross-correlation between $\hat{y}$ and the augmented observation vector $\mathbf{x}^a$. The optimal coefficient vectors $\mathbf{h}_o$ and $\mathbf{g}_o$ can be subsequently obtained by virtue of the Cholesky block factorisation of $\mathbf{C}_{\mathbf{x}^a \mathbf{x}^a}^{-1}$, as given in [39], and (2.16) can be separated as

$$\begin{aligned} \mathbf{h}_o &= (\mathbf{C} - \mathbf{P} \mathbf{C}^{-*} \mathbf{P}^*)^{-1} (\mathbf{c}_1 - \mathbf{P} \mathbf{C}^{-*} \mathbf{c}_2^*) \\ \mathbf{g}_o &= (\mathbf{C}^* - \mathbf{P}^* \mathbf{C}^{-1} \mathbf{P})^{-1} (\mathbf{c}_2^* - \mathbf{P}^* \mathbf{C}^{-1} \mathbf{c}_2) \end{aligned} \quad (2.17)$$

where the subscripts have been omitted for clarity. The widely linear MMSE is then given by [36]

$$E[|e_{\text{WL}}|^2] = E[yy^*] - \mathbf{c}_1^T \mathbf{h}_o - \mathbf{c}_2^H \mathbf{g}_o \quad (2.18)$$

However, by considering the strictly linear model (2.3), the coefficient vector obtaining the MMSE is given by

$$\mathbf{h} = \mathbf{C}^{-1} \mathbf{c}_1 \quad (2.19)$$

and the corresponding MMSE is given by

$$E[|e_{\text{L}}|^2] = E[yy^*] - \mathbf{c}_1^H \mathbf{C}^{-1} \mathbf{c}_1 \quad (2.20)$$

Therefore, the advantage of the widely linear MMSE over the strictly linear MMSE can be characterised by the quantity $\Delta_{MMSE} = E[|e_L|^2] - E[|e_{WL}|^2]$, expressed as [36]

$$\Delta_{MMSE} = (\mathbf{c}_2^* - \mathbf{P}^* \mathbf{C}^{-1} \mathbf{c}_1)^H (\mathbf{C}^* - \mathbf{P}^* \mathbf{C}^{-1} \mathbf{P})^{-1} (\mathbf{c}_2^* - \mathbf{P}^* \mathbf{C}^{-1} \mathbf{c}_1) \quad (2.21)$$

Note that Δ_{MMSE} is always nonnegative due to the positive definite property of the matrix $(\mathbf{C}^* - \mathbf{P}^* \mathbf{C}^{-1} \mathbf{P})^{-1}$, indicating the widely linear solution yields a smaller MMSE compared to its strictly linear counterpart. It equals zero only when $\mathbf{c}_2^* - \mathbf{P}^* \mathbf{C}^{-1} \mathbf{c}_1 = \mathbf{0}$ and only for a second-order circular signal y and observation $\mathbf{x}$, such that $\mathbf{P} = \mathbf{0}$ and cross-correlation $\mathbf{c}_2 = \mathbf{0}$ [36]. Based on the above results, it is clear that the strictly linear model is suboptimal for the generality of complex-valued signals, and can be seen as a special case of the widely linear model to deal with second-order circular signals.

2.2.3 Quantification of second order noncircularity (improperity)

A natural choice to quantify the degree of noncircularity is to use the relationship between the standard complex variance and pseudo-variance, such as the circularity quotient $\rho(x) \triangleq \frac{\tau_x^2}{\sigma_x^2}$ [40, 41]. The magnitude of $\rho(x)$ can be used as a measure of degree of noncircularity in the complex-valued signals, denoted by

$$r = |\rho(x)| = \frac{|\tau_x^2|}{\sigma_x^2} \quad (2.22)$$

where r is normalised between $[0, 1]$, with 0 indicating perfect circularity and 1 maximum noncircularity. The circularity angle $\theta = \arg(\rho(x))$ is an indicator of orientation of distribution. Note that for a purely circular signal, $r = 0$, with θ not providing additional information about the distribution.

A more general numerical measure of degree of noncircularity is achieved by using the canonical correlations and strong uncorrelating transform on the augmented covariance matrix [42], where for a complex-valued random variable $\mathbf{x}$, its degree of noncircularity s is defined as

$$s = 1 - \det(\mathbf{C}_{\mathbf{x}^a \mathbf{x}^a}) \det^{-2}(\mathbf{C}_{\mathbf{xx}}), \quad s \in [0, 1] \quad (2.23)$$

where $\det(\cdot)$ denotes the matrix determinant operator. When $\mathbf{x}$ becomes a scalar, s is simplified into r^2 [42]. Other measures of noncircularity are also defined and may be used. A similar measure given by $(1 - r)$ is defined in [43], whereas an unbounded measure in $[1, \infty]$ based on the ratio of the standard deviations of the real and imaginary components of the complex random variable was introduced in [44]. Although the mentioned measures are quite similar, the generality of s in (2.23) and the embedded information on augmented covariance matrix, makes it a suitable noncircularity measure used in this work.

2.3 Degree of noncircularity of the benchmarks and real-world wind profile

This section describes the benchmark signals and the real-world wind profile data used in this work to test the advantage of widely linear augmented algorithms over their strictly linear counterparts.

The second order circular benchmark is a synthetised stable complex-valued AR(4) process, given by

$$x(k) = 1.79x(k-1) - 1.85x(k-2) + 1.27x(k-3) - 0.41x(k-4) + n(k) \quad (2.24)$$

driven by complex-valued doubly circular white Gaussian noise $n(k)$ with zero mean and unit variance. Double whiteness implies the uncorrelated real and imaginary channels, for circular white Gaussian noise, $n = n_r + jn_i$, $\sigma_{n_r}^2 = \sigma_{n_i}^2$, whereas for noncircular data, $\sigma_{n_r}^2 > \sigma_{n_i}^2$, where $\sigma_{n_r}^2$ and $\sigma_{n_i}^2$ are respectively the powers of the real and imaginary parts.

The benchmark noncircular signal was a complex auto-regressive moving average (ARMA) process, whose transfer function was a combination of the MA model in [45] and the stable AR model in (2.24), given by

$$\begin{aligned} r(k) = & 1.79r(k-1) - 1.85r(k-2) + 1.27r(k-3) - 0.41r(k-4) + 0.2r(k-5) + 2n(k) \\ & + 0.5n^*(k) + n(k-1) + 0.9n^*(k-1) \end{aligned} \quad (2.25)$$

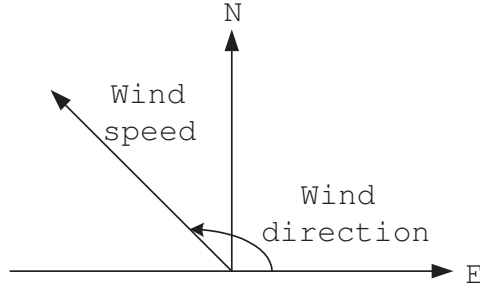


Figure 2.1: Wind recordings as a complex vector representation using speed and direction information.

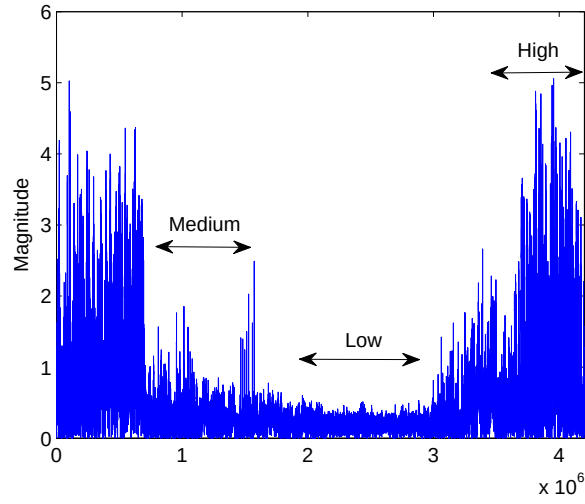


Figure 2.2: Complex wind signal magnitude. Three wind speed regions have been identified as *low*, *medium* and *high*.

with

$$E\{n(k-i)n^*(k-j)\} = \delta(i-j), \quad E\{n(k-i)n(k-j)\} = C\delta(i-j) \quad (2.26)$$

where $n(k)$ was complex-valued doubly circular white Gaussian noise and $C = 0.95$ [45].

A nonlinear and noncircular chaotic Ikeda map signal is given by [46]

$$\begin{aligned} x(k+1) &= 1 + u(x(k)\cos[t(k)] - y(k)\sin[t(k)]) \\ y(k+1) &= u(x(k)\sin[t(k)] + y(k)\cos[t(k)]) \end{aligned} \quad (2.27)$$

where $t(k) = 0.4 - \frac{6}{1+x^2(k)+y^2(k)}$ is the time instant and u is the map parameter to control

Table 2.1: Comparison of degrees of noncircularity s for the various classes of signals

	AR(4) (2.24)	Noncircular ARMA (2.25)	Ikeda map (2.27)	Wind (<i>low</i>)	Wind (<i>medium</i>)	Wind (<i>high</i>)
s	0.0016	0.9429	0.8936	0.1583	0.4305	0.8117

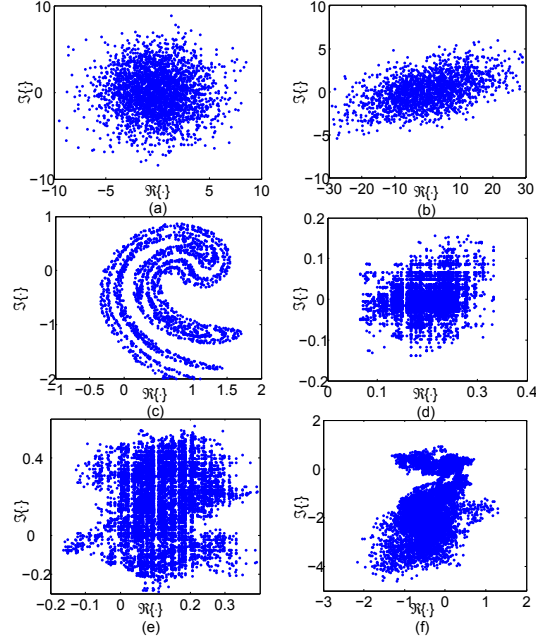


Figure 2.3: Geometric view of circularity via “real-imaginary” scatter plots. (a) Circular autoregressive AR(4) process (2.24); (b) Noncircular ARMA process (2.25); (c) Noncircular Ikeda map (2.27); (d) Wind (*low*) signal; (e) Wind (*medium*) signal; (f) Wind (*high*) signal.

bifurcation of attractor points. To make sure that the Ikeda map exhibits chaotic dynamics on a strange attractor, u is traditionally set to be 0.9 [46]. The complex-valued Ikeda attractor is built as $z(k) = x(k) + jy(k)$.

The real-world wind data was collected using an ultrasonic anemometer over a period of 24 hours sampled at 50Hz². As illustrated in Fig. 2.1, the wind profile was represented as a vector of speed and direction in the North-South (V_N) and East-West (V_E) coordinate system, which was used to create the complex wind signal as $V = ve^{j\varphi}$, and

$$v = \sqrt{V_E^2 + V_N^2}, \quad \varphi = \arctan\left(\frac{V_N}{V_E}\right) \quad (2.28)$$

where v is the wind speed, and φ is the wind direction. The magnitude of the wind profile

²The wind data was provided by Prof. Aihara’s team at the Institute of Industrial Science, University of Tokyo, Japan, which was recorded in an urban environment. The data was re-sampled at 5 Hz for simulation purposes.

is shown in Fig. 2.2. Based on the changes in the wind intensity, the noncircular wind data were identified as regions of *low*, *medium* and *high* dynamics. For more detail on the wind property see [41, 47].

Fig. 2.3 shows the scatter plots of the complex signals considered in simulations. Observe the circular symmetry (rotation invariance) for the AR(4) signal (2.24) and the noncircularity of the ARMA model (2.25), Ikeda map (2.27) and wind signals. Table 2.1 illustrates the degrees of noncircularity s for the various classes of signals. Observe the excellent match between the measure of noncircularity in Table 2.1 and scatter plot descriptions in Fig. 2.3; for instance, among the wind segments, the wind (*low*) region was least noncircular, whereas the wind (*high*) region exhibited strong noncircularity.

Chapter 3

Mean Square Performance

Analysis of the Augmented CLMS Algorithm

3.1 Introduction

The Least Mean Square (LMS) algorithm is a simple, yet most frequently used, algorithm for adaptive FIR filters in $\mathbb{R}$. As discussed before, to deal with complex-valued signals which are frequently encountered in modern disciplines, three different approaches can be used to design its extension to complex domain $\mathbb{C}$ according to the bivariate or complex nature of complex-valued data. Firstly, complex-valued signals can be separated into the real and imaginary components (magnitude and phase), and a dual univariate approach, which treats them independently, leads to a direct processing of a complex-valued signal using the real-valued LMS algorithm. However, due to this separate treatment, the cross-channel information would be missing, leading to inadequate performance. Alternatively, the dual channel LMS [48], a bivariate algorithm operating in $\mathbb{R}$, allows for the consideration of the available cross-information. A natural extension of the real-valued LMS algorithm for the adaptive filtering directly in the field of complex signals $\mathbb{C}$ was the Complex LMS (CLMS) [24]. Since CLMS treats the complex-valued signal as a whole, it enables simultaneous filtering of the real and imaginary components of complex-valued

data and accounts for second-order cross-information between the channels.

Although the CLMS algorithm has the advantage that it can directly operate on the complex-valued signals, it was derived under the implicit assumption that complex-valued signals and noises are second order circular (improper) with rotation invariant distributions, which accounts for only a subset of complex-valued signals. Based on the concept of augmented complex statistics and widely linear modelling, the widely linear CLMS was firstly introduced in the communications field for use in a direct-sequence code division multiple access (DS-CDMA) receiver [31]. However, due to the specific requirement in the DS-CDMA setting, where the desired input symbols are real-valued, i.e. amplitude-shift keying (ASK) or binary phase-shift keying (BPSK), this algorithm does not fully exploit the advantages of the widely linear model. To this end, it was subsequently extended to the augmented CLMS (ACLMS) algorithm in [32] in order to cater for general complex-valued signals and errors.

Although the ACLMS has already shown its potential to deal with noncircular complex signals over its conventional counterpart, CLMS, in real-world applications, such as wind profile prediction [47] and electroencephalogram (EEG) artifacts removal [49], a theoretical understanding of the performance comparisons between them is still missing.

To this end, in Section 3.2, we firstly revisit the derivation of ACLMS in an adaptive prediction context using an efficient tool, called the $\mathbb{CR}$ calculus, which provides a unified framework for the derivation of a real-valued function of complex variables and will be used throughout this work. Notice that in the original work of [32], the standard derivation method, which calculates the gradients by considering the derivative of the cost function with respect to the real and imaginary components of the weight vectors $\mathbf{h}$ and $\mathbf{g}$ separately, was used to derive ACLMS. Alternatively, the $\mathbb{CR}$ calculus framework facilitates a simpler derivation method by considering the cost function as a function of the conjugate coordinates of the weight vectors, that is $(\mathbf{h}, \mathbf{g}, \mathbf{h}^*, \mathbf{g}^*)$, allowing for the calculation of the derivatives directly in $\mathbb{C}$. A brief description of $\mathbb{CR}$ calculus is provided in Appendix A. In Section 3.3, an efficient approach to the steady state performance of adaptive filters, called the energy conservation principle [15, 50], which analyzes the energy flow through each iteration of adaptive filters, has been extended into widely linear modelling to quantify the

mean square errors of ACLMS and CLMS algorithms on noncircular signals. In Section 3.4, the improvement in the performance of ACLMS algorithm over standard CLMS algorithm is supported by both theoretical analysis and simulations performed on one step ahead prediction of noncircular signals.

3.2 Derivation of ACLMS

Consider the output of an adaptive filter,

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

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

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

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

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

where $d(k)$ is the desired signal. The update form of the standard part weight vector $\mathbf{h}(k)$ is given by

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

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

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

where

$$\begin{aligned}\frac{\partial J(k)}{\partial \mathbf{h}_r(k)} &= \frac{1}{2} \left[e(k) \frac{\partial e^*(k)}{\partial \mathbf{h}_r(k)} + e^*(k) \frac{\partial e(k)}{\partial \mathbf{h}_r(k)} \right] \\ &= -\frac{1}{2} [e(k) \mathbf{x}^*(k) + e^*(k) \mathbf{x}(k)]\end{aligned}$$

and

$$\begin{aligned}\frac{\partial J(k)}{\partial \mathbf{h}_i(k)} &= \frac{1}{2} \left[e(k) \frac{\partial e^*(k)}{\partial \mathbf{h}_i(k)} + e^*(k) \frac{\partial e(k)}{\partial \mathbf{h}_i(k)} \right] \\ &= \frac{j}{2} [e(k) \mathbf{x}^*(k) - e^*(k) \mathbf{x}(k)]\end{aligned}$$

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

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

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

$$\mathbf{h}(k+1) = \mathbf{h}(k) + \mu e(k) \mathbf{x}^*(k) \quad (3.1)$$

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

$$\mathbf{g}(k+1) = \mathbf{g}(k) + \mu e(k) \mathbf{x}(k) \quad (3.2)$$

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

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

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

3.3 Steady state mean square performance of CLMS and ACLMS on noncircular signals

The aim of this section is to prove the suitability of ACLMS algorithm for the filtering of noncircular complex data, for which the standard CLMS algorithm is suboptimal. Consider

a second order noncircular (improper) teaching signal $d(k)$ that stems from the widely linear model

$$d(k) = \underbrace{\mathbf{x}^T(k)\mathbf{h}_o}_{\text{standard part}} + \underbrace{\mathbf{x}^H(k)\mathbf{g}_o}_{\text{conjugate part}} + n(k) \quad (3.3)$$

where $\mathbf{h}_o$ and $\mathbf{g}_o$ are respectively the optimal standard and conjugate weight vectors, $n(k)$ represents the measurement noise with zero mean and variance σ_n^2 .

3.3.1 Steady state analysis of CLMS algorithm for noncircular signals

The conventional CLMS algorithm is given by [24]

$$\begin{aligned} y(k) &= \mathbf{x}^T(k)\mathbf{h}(k) \\ e(k) &= d(k) - y(k) \\ \mathbf{h}(k+1) &= \mathbf{h}(k) + \mu e(k)\mathbf{x}^*(k) \end{aligned} \quad (3.4)$$

The steady state performance is quantified by the mean square error (MSE) given by

$$\text{MSE} = \lim_{k \rightarrow \infty} E[|e(k)|^2] \quad (3.5)$$

where $e(k)$ is defined in (3.4). To find the closed form of MSE of CLMS for noncircular signals generated by the widely linear model (3.3), we shall first define the weight error vector $\tilde{\mathbf{h}}(k) = \mathbf{h}_o - \mathbf{h}(k)$; then upon subtracting the weight update in (3.4) from $\mathbf{h}_o$, we have

$$\tilde{\mathbf{h}}(k+1) = \tilde{\mathbf{h}}(k) - \mu \mathbf{x}^*(k)e(k) \quad (3.6)$$

Pre-multiplying both sides of (3.6) with $\mathbf{x}^T(k)$ gives

$$\mathbf{x}^T(k)\tilde{\mathbf{h}}(k+1) = \mathbf{x}^T(k)\tilde{\mathbf{h}}(k) - \mu \|\mathbf{x}(k)\|^2 e(k) \quad (3.7)$$

allowing us to introduce the *a posteriori* and *a priori* errors $e_p(k)$ and $e_a(k)$ as

$$e_p(k) = \mathbf{x}^T(k)\tilde{\mathbf{h}}(k+1) \quad \text{and} \quad e_a(k) = \mathbf{x}^T(k)\tilde{\mathbf{h}}(k) \quad (3.8)$$

and to rewrite the output error $e(k)$ as

$$e(k) = d(k) - y(k) = e_a(k) + n(k) + \mathbf{x}^H(k)\mathbf{g}_o \quad (3.9)$$

Compared with the expression in [4], an additional term $\mathbf{x}^H(k)\mathbf{g}_o$ arises due to the improper nature of the desired signal in (3.3). The relationship between the *a posteriori* and *a priori* errors in (3.8) is given by

$$e_p(k) = e_a(k) - \mu \|\mathbf{x}(k)\|^2 e(k) \quad (3.10)$$

and for nonzero input $\mathbf{x}(k)$, we have

$$\mu e(k) = \frac{e_a(k) - e_p(k)}{\|\mathbf{x}(k)\|^2} \quad (3.11)$$

The evolution of the weight error vector $\tilde{\mathbf{h}}(k)$ can now be expressed as

$$\tilde{\mathbf{h}}(k+1) = \tilde{\mathbf{h}}(k) - \frac{e_a(k) - e_p(k)}{\|\mathbf{x}(k)\|^2} \mathbf{x}^*(k) \quad (3.12)$$

or equivalently

$$\tilde{\mathbf{h}}(k+1) + \frac{\mathbf{x}^*(k)}{\|\mathbf{x}(k)\|^2} e_a(k) = \tilde{\mathbf{h}}(k) + \frac{\mathbf{x}^*(k)}{\|\mathbf{x}(k)\|^2} e_p(k) \quad (3.13)$$

With the standard assumption that at steady state $\tilde{\mathbf{h}}(k+1) \simeq \tilde{\mathbf{h}}(k)$, by estimating the energies of both sides of the above equation, we obtain the following energy relationship within the evolution of the weight error vector,

$$\|\tilde{\mathbf{h}}(k+1)\|^2 + \frac{|e_a(k)|^2}{\|\mathbf{x}(k)\|^2} = \|\tilde{\mathbf{h}}(k)\|^2 + \frac{|e_p(k)|^2}{\|\mathbf{x}(k)\|^2} \quad (3.14)$$

If we apply the statistical expectation of both sides of (3.14), and assume that at the steady state, $E[\|\tilde{\mathbf{h}}(k+1)\|^2] \simeq E[\|\tilde{\mathbf{h}}(k)\|^2]$, then

$$E\left[\frac{|e_a(k)|^2}{\|\mathbf{x}(k)\|^2}\right] = E\left[\frac{|e_p(k)|^2}{\|\mathbf{x}(k)\|^2}\right] \quad (3.15)$$

Substituting (3.10) into the right hand side (RHS) of (3.15) gives

$$E\left[\frac{|e_a(k)|^2}{\|\mathbf{x}(k)\|^2}\right] = E\left[\frac{|e_a(k) - \mu\|\mathbf{x}(k)\|^2 e(k)|^2}{\|\mathbf{x}(k)\|^2}\right] \quad (3.16)$$

and upon combining with (3.9), this yields

$$E\left[\frac{|e_a(k)|^2 - |e_a(k) - \mu\|\mathbf{x}(k)\|^2(e_a(k) + n(k) + \mathbf{x}^H(k)\mathbf{g}_o)|^2}{\|\mathbf{x}(k)\|^2}\right] = 0 \quad (3.17)$$

We can now use the standard assumptions that at the steady state (as $k \rightarrow \infty$), the noise sequence $n(k)$ is *i.i.d.* and statistically independent of the input sequence $\mathbf{x}(k)$, and that $e_a(k)$ is independent of $\mathbf{x}(k)$ to obtain

$$2\text{EMSE}_{\text{CLMS}} = \mu E[\|\mathbf{x}(k)\|^2 |e_a(k)|^2] + \mu E[\|\mathbf{x}(k)\|^2 |\mathbf{x}^H(k)\mathbf{g}_o|^2] + \mu \sigma_n^2 \text{Tr}(\mathbf{C}_{\mathbf{xx}}) \quad (3.18)$$

where $\text{EMSE} = \lim_{k \rightarrow \infty} E[|e_a(k)|^2]$. For sufficiently small step size μ , we can neglect the term $\mu E[\|\mathbf{x}(k)\|^2 |e_a(k)|^2]$, which is in the formulation of second order of $e_a(k)$, and obtain

$$\text{EMSE}_{\text{CLMS}} = \frac{\mu}{2} \sigma_n^2 \text{Tr}(\mathbf{C}_{\mathbf{xx}}) + \frac{\mu}{2} E[\|\mathbf{x}(k)\|^2 |\mathbf{x}^H(k)\mathbf{g}_o|^2] \quad (3.19)$$

For a large value of the step size μ , the second term on the RHS of (3.18) cannot be neglected, however, with a further assumption that at the steady state $\mu\|\mathbf{x}(k)\|^2$ is statistically independent of $e_a(k)$ ¹ [15]. This assumption becomes more realistic for filters with a longer length². It becomes exact for the constant modulus data which is encountered in some adaptive filtering applications. Within this assumption, we have

$$\text{EMSE}_{\text{CLMS}} = \frac{\mu \sigma_n^2 \text{Tr}(\mathbf{C}_{\mathbf{xx}}) + \mu E[\|\mathbf{x}(k)\|^2 |\mathbf{x}^H(k)\mathbf{g}_o|^2]}{2 - \mu \text{Tr}(\mathbf{C}_{\mathbf{xx}})} \quad (3.20)$$

¹By large values of μ , we do not mean a large μ , but rather step-size values that are not infinitesimally small and guarantee filter stability [51].

²To guarantee convergence, the algorithm step-size μ is normally chosen to be inversely proportional to the filter length L . Using the law of large numbers, $\|\mathbf{x}(k)\|^2$ can be considered as a random variable of variance proportional to L . Therefore, $\mu\|\mathbf{x}(k)\|^2$ has variance proportional to $1/L$, which decays with the filter length L . This means that for long-enough filters, the variations in $\mu\|\mathbf{x}(k)\|^2$ are very small, and can be considered to be independent of $e_a(k)$ [51].

Compared with the expressions of EMSE for standard CLMS operating on proper signals generated by a linear model (see [15]), both (3.19) and (3.20) contain an additional “conjugate” term $\mathbf{x}^H(k)\mathbf{g}_o$, arising from the widely linear nature of the teaching signal model in (3.3), thus illustrating that the standard CLMS is a suboptimal solution for the filtering of second order noncircular signals. Using (3.5) and (3.9), we finally obtain the steady state mean square error of standard CLMS for improper signals in the form

$$\text{MSE}_{\text{CLMS}} = \text{EMSE}_{\text{CLMS}} + E[|\mathbf{x}^H(k)\mathbf{g}_o|^2] + \sigma_n^2 \quad (3.21)$$

3.3.2 Steady state analysis of ACLMS algorithm for noncircular signals

Similar to the analysis in the previous subsection, in the context of ACLMS algorithm, by subtracting the $\mathbf{h}$ and $\mathbf{g}$ weight updates in (3.1) and (3.2) from their respective optimal weights $\mathbf{h}_o$ and $\mathbf{g}_o$ in (3.3), we obtain the corresponding weight error vectors $\tilde{\mathbf{h}}(k) = \mathbf{h}_o - \mathbf{h}(k)$ and $\tilde{\mathbf{g}}(k) = \mathbf{g}_o - \mathbf{g}(k)$, whose evolutions are described as

$$\tilde{\mathbf{h}}(k+1) = \tilde{\mathbf{h}}(k) - \mu \mathbf{x}^*(k)e(k) \quad (3.22)$$

$$\tilde{\mathbf{g}}(k+1) = \tilde{\mathbf{g}}(k) - \mu \mathbf{x}(k)e(k) \quad (3.23)$$

Multiplying both sides of (3.22) with $\mathbf{x}^T(k)$ and both sides of (3.23) with $\mathbf{x}^H(k)$ yields

$$\mathbf{x}^T(k)\tilde{\mathbf{h}}(k+1) = \mathbf{x}^T(k)\tilde{\mathbf{h}}(k) - \mu \|\mathbf{x}(k)\|^2 e(k) \quad (3.24)$$

$$\mathbf{x}^H(k)\tilde{\mathbf{g}}(k+1) = \mathbf{x}^H(k)\tilde{\mathbf{g}}(k) - \mu \|\mathbf{x}(k)\|^2 e(k) \quad (3.25)$$

The *a posteriori* and *a priori* errors $e_p(k)$ and $e_a(k)$ now become

$$e_p(k) = \mathbf{x}^T(k)\tilde{\mathbf{h}}(k+1) + \mathbf{x}^H(k)\tilde{\mathbf{g}}(k+1) \quad (3.26)$$

$$e_a(k) = \mathbf{x}^T(k)\tilde{\mathbf{h}}(k) + \mathbf{x}^H(k)\tilde{\mathbf{g}}(k) \quad (3.27)$$

Subtracting (3.27) from (3.26) yields

$$e_p(k) = e_a(k) - 2\mu \|\mathbf{x}(k)\|^2 e(k)$$

and

$$e(k) = \frac{1}{2\mu\|\mathbf{x}(k)\|^2} (e_a(k) - e_p(k)) \quad (3.28)$$

Substituting (3.28) into (3.22) and (3.23), with the further assumption that at steady state $\tilde{\mathbf{h}}(k+1) \simeq \tilde{\mathbf{h}}(k)$ and $\tilde{\mathbf{g}}(k+1) \simeq \tilde{\mathbf{g}}(k)$, we have

$$\|\tilde{\mathbf{h}}(k+1)\|^2 + \frac{\tilde{\mathbf{h}}^H(k)\mathbf{x}^*(k)e_a(k)}{2\|\mathbf{x}(k)\|^2} = \|\tilde{\mathbf{h}}(k)\|^2 + \frac{\tilde{\mathbf{h}}^H(k+1)\mathbf{x}^*(k)e_p(k)}{2\|\mathbf{x}(k)\|^2} \quad (3.29)$$

and

$$\|\tilde{\mathbf{g}}(k+1)\|^2 + \frac{\tilde{\mathbf{g}}^H(k)\mathbf{x}^*(k)e_a(k)}{2\|\mathbf{x}(k)\|^2} = \|\tilde{\mathbf{g}}(k)\|^2 + \frac{\tilde{\mathbf{g}}^H(k+1)\mathbf{x}^*(k)e_p(k)}{2\|\mathbf{x}(k)\|^2} \quad (3.30)$$

Adding up (3.29) and (3.30) gives the weight error energy conservation equation for the widely linear modelling

$$\|\tilde{\mathbf{h}}(k+1)\|^2 + \|\tilde{\mathbf{g}}(k+1)\|^2 + \frac{|e_a(k)|^2}{2\|\mathbf{x}(k)\|^2} = \|\tilde{\mathbf{h}}(k)\|^2 + \|\tilde{\mathbf{g}}(k)\|^2 + \frac{|e_p(k)|^2}{2\|\mathbf{x}(k)\|^2} \quad (3.31)$$

Taking the statistical expectation, and assuming that at the steady state $E[\|\tilde{\mathbf{h}}(k+1)\|^2] \simeq E[\|\tilde{\mathbf{h}}(k)\|^2]$, and $E[\|\tilde{\mathbf{g}}(k+1)\|^2] \simeq E[\|\tilde{\mathbf{g}}(k)\|^2]$, gives

$$E\left[\frac{|e_a(k)|^2}{\|\mathbf{x}(k)\|^2}\right] = E\left[\frac{|e_p(k)|^2}{\|\mathbf{x}(k)\|^2}\right] \quad (3.32)$$

Similar to the analysis in the previous subsection, it follows immediately that

$$2\text{EMSE}_{\text{ACLMS}} = \mu E[\|\mathbf{x}(k)\|^2 |e_a(k)|^2] + \mu \sigma_n^2 \text{Tr}(\mathbf{C}_{\mathbf{xx}}) \quad (3.33)$$

For a sufficiently small step size μ , we obtain

$$\text{EMSE}_{\text{ACLMS}} = \frac{\mu}{2} \sigma_n^2 \text{Tr}(\mathbf{C}_{\mathbf{xx}}) \quad (3.34)$$

where as for a large value of μ , we have

$$\text{EMSE}_{\text{ACLMS}} = \frac{\mu \sigma_n^2 \text{Tr}(\mathbf{C}_{\mathbf{xx}})}{2 - \text{Tr}(\mathbf{C}_{\mathbf{xx}})} \quad (3.35)$$

Note that from

$$\begin{aligned}
 e(k) &= d(k) - y(k) \\
 &= \mathbf{x}^T(k)(\mathbf{h}_o - \tilde{\mathbf{h}}(k)) + \mathbf{x}^H(k)(\mathbf{g}_o - \tilde{\mathbf{g}}(k)) + n(k) \\
 &= e_a(k) + n(k)
 \end{aligned} \tag{3.36}$$

the mean square error of ACLMS, denoted by $\text{MSE}_{\text{ACLMS}}$, has the form

$$\text{MSE}_{\text{ACLMS}} = \text{EMSE}_{\text{ACLMS}} + \sigma_n^2 \tag{3.37}$$

Consider, without loss of generality, a noncircular complex signal is generated from the widely linear model (3.3), $E[|\mathbf{x}^H(k)\mathbf{g}_o|^2] > 0$. This in turn indicates that at steady state, $\text{EMSE}_{\text{CLMS}} > \text{EMSE}_{\text{ACLMS}}$ and $\text{MSE}_{\text{CLMS}} > \text{MSE}_{\text{ACLMS}}$ for both small and large step sizes.

3.3.3 Mean and mean square stability analysis of ACLMS algorithm

We shall next investigate the bounds on the step size μ , giving insight into the convergence of the ACLMS algorithm. From Section 3.3.2, substituting (3.36) into (3.22) and (3.23), we have

$$\tilde{\mathbf{h}}(k+1) = (\mathbf{I} - \mu\mathbf{x}^*(k)\mathbf{x}^T(k))\tilde{\mathbf{h}}(k) - \mu\mathbf{x}^*(k)\mathbf{x}^H(k)\tilde{\mathbf{g}}(k) - \mu\mathbf{x}^*(k)n(k) \tag{3.38}$$

$$\tilde{\mathbf{g}}(k+1) = (\mathbf{I} - \mu\mathbf{x}(k)\mathbf{x}^H(k))\tilde{\mathbf{g}}(k) - \mu\mathbf{x}(k)\mathbf{x}^T(k)\tilde{\mathbf{h}}(k) - \mu\mathbf{x}(k)n(k) \tag{3.39}$$

where $\mathbf{I}$ is a $L \times L$ identity matrix. A compact vector-matrix form for the weight error vectors can be derived as

$$\begin{bmatrix} \tilde{\mathbf{h}}(k+1) \\ \tilde{\mathbf{g}}(k+1) \end{bmatrix} = \begin{bmatrix} \mathbf{I} - \mu\mathbf{x}^*(k)\mathbf{x}^T(k) & -\mu\mathbf{x}^*(k)\mathbf{x}^H(k) \\ -\mu\mathbf{x}(k)\mathbf{x}^T(k) & \mathbf{I} - \mu\mathbf{x}(k)\mathbf{x}^H(k) \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{h}}(k) \\ \tilde{\mathbf{g}}(k) \end{bmatrix} - \mu \begin{bmatrix} \mathbf{x}^*(k) \\ \mathbf{x}(k) \end{bmatrix} n(k) \tag{3.40}$$

Upon applying the statistical expectation operator, and using the augmented notation for the weight error vector $\tilde{\mathbf{w}}(k) = [\tilde{\mathbf{h}}^T(k), \tilde{\mathbf{g}}^T(k)]^T$, we obtain

$$E[\tilde{\mathbf{w}}(k+1)] = (\mathbf{I} - \mu \mathbf{C}_{\mathbf{x}^a \mathbf{x}^a}^*) E[\tilde{\mathbf{w}}(k)] \quad (3.41)$$

Performing the analysis of modes of convergence with respect to the conjugate augmented covariance matrix $\mathbf{C}_{\mathbf{x}^a \mathbf{x}^a}^*$ as in [52], and bearing in mind that the eigenvalues of $\mathbf{C}_{\mathbf{x}^a \mathbf{x}^a}^*$ are identical to those of $\mathbf{C}_{\mathbf{x}^a \mathbf{x}^a}$, it is easy to show that the step size is bounded by

$$0 < \mu < \frac{2}{\lambda_{\max}(\mathbf{C}_{\mathbf{x}^a \mathbf{x}^a})} \quad (3.42)$$

where $\lambda_{\max}(\mathbf{C}_{\mathbf{x}^a \mathbf{x}^a})$ is the largest eigenvalue of the augmented covariance matrix $\mathbf{C}_{\mathbf{x}^a \mathbf{x}^a}$.

Note that (3.41) satisfies the unified form of the evolution of the mean of the weight error vector, then, as shown in Chapter 9 of [2], the bounds on step size μ can be made tighter based on $\lambda(F(\mu)) < 1$, where

$$F(\mu) = \mathbf{I} - \mu \mathbf{A} + \mu^2 \mathbf{B} \quad (3.43)$$

where $\mathbf{A} = E[\mathbf{P}^T(k)] \otimes \mathbf{I} + \mathbf{I} \otimes E[\mathbf{P}^T(k)]$, $\mathbf{B} = E[\mathbf{P}^T(k)] \otimes E[\mathbf{P}(k)]$, and $\mathbf{P}(k) = \mathbf{x}^a(k) \mathbf{x}^{aH}(k)$. Following the argument used in [17], we can finally establish the bound on the step-size in the form

$$0 < \mu < \min\{1/\lambda_{\max}(\mathbf{A}^{-1}\mathbf{B}), 1/\max(\lambda(\mathbf{H}) \in \mathbb{R}^+)\} \quad (3.44)$$

where $\mathbf{H}$ is defined by

$$\mathbf{H} = \begin{bmatrix} (1/2)\mathbf{A} & -(1/2)\mathbf{B} \\ \mathbf{I} & \mathbf{0} \end{bmatrix} \quad (3.45)$$

Table 3.1: Stability bounds of ACLMS algorithm (Gaussian input and uniform input)

Types of input	$\frac{2}{\lambda_{\max}(\mathbf{C}_{\mathbf{x}^a \mathbf{x}^a})}$	$\frac{1}{\lambda_{\max}(\mathbf{A}^{-1} \mathbf{B})}$	$\frac{1}{\max(\lambda(\mathbf{H}) \in \mathbb{R}^+)}$	$\mu_{\max}$
Gaussian input	1.2795	1.2853	0.9109	0.9109
Uniform input	0.9769	0.9752	0.6614	0.6614

3.4 Simulations

The first set of simulations was performed in an one step prediction setting, for a noncircular complex signal generated by the widely linear autoregressive (AR) model

$$r(k+1) = 0.5r(k) + 0.2r^*(k) + n(k) \quad (3.46)$$

All the graphs were produced by averaging 200 independent trials. Our analysis does not require the input data to be Gaussian, and two cases of complex valued doubly white circular driving noise $n(k)$ were investigated: Gaussian and uniform, both with zero mean and unit variance. The filter length L is set to be 4 for both CLMS and ACLMS algorithms.

Table 3.1 presents the theoretical bounds on the step size μ of ACLMS algorithm based on the mean and mean square analysis. As expected, mean square analysis gives a better estimation. This is further illustrated in Fig. 3.1 which shows both the simulated MSE curves (dashed line) and theoretical bounds (solid line). Fig. 3.2 shows the simulated and theoretical steady state MSE curves of CLMS and ACLMS as a function of step size μ - they are in good agreement with the theoretical results, estimated using (3.21) and (3.37), when the step-size μ is relatively small between 10^{-4} and 10^{-2} , however, when the step-size is larger than 0.01, due to the not long enough filter length, the independence between $\mu \|\mathbf{x}(k)\|^2$ and $|e_a(k)|^2$ becomes unrealistic, and hence the mismatch between the theoretical results and their simulated counterparts becomes larger [15].

The next set of simulations was performed in a system identification setting, where the system to be identified was a strictly linear FIR channel of length $L = 4$ with coefficients $\mathbf{h}_o^T = [1.79, -1.85, 1.27, -0.41]$. The coefficients are selected to guarantee the corresponding autoregressive AR(4) model is stable with all the poles of its characteristic function inside the unit circle, as shown in Fig. 3.3; other coefficients to generate a stable

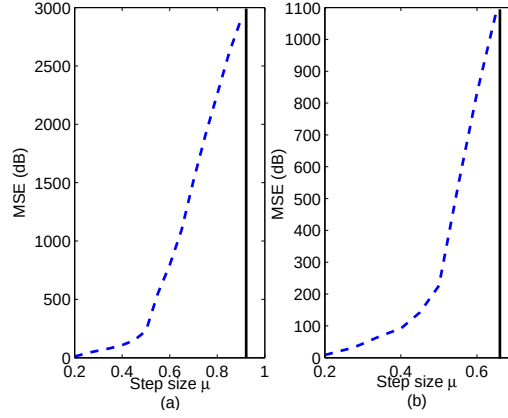


Figure 3.1: Simulated MSE (dash) of ACLMS algorithm as a function of step-size μ for (a) Gaussian input (b) Uniform input.

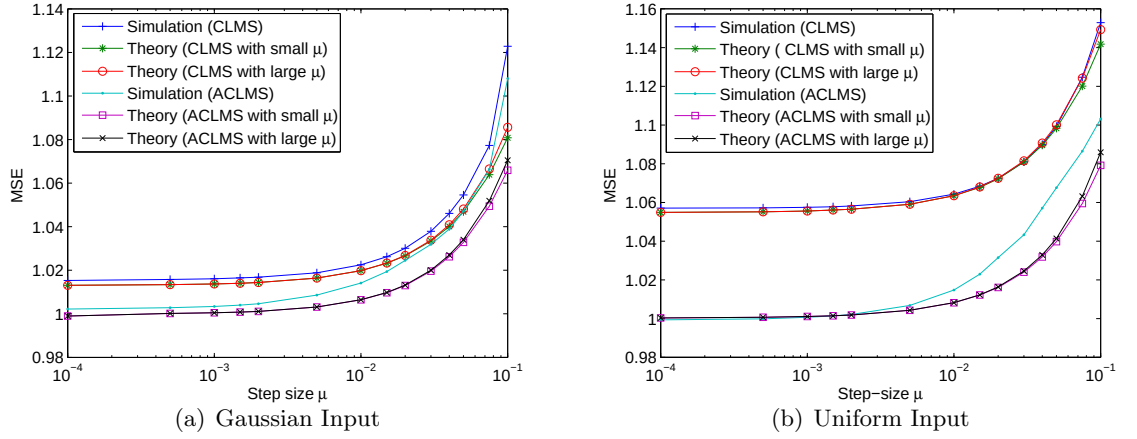


Figure 3.2: Comparison of theoretical and simulated steady state MSE of CLMS and ACLMS on noncircular AR signals generated by (a) Gaussian input and (b) Uniform input.

autoregressive process are also applicable. The input signal $x(k)$ was circular complex valued doubly white Gaussian noise with zero mean and unit variance. Following the discussion in [36], the desired signal was real valued $d(k) = \text{Re}\{\mathbf{x}^T(k)\mathbf{h}_o\}$, and the output of the filter was contaminated by noncircular Gaussian noise

$$n(k) = \sqrt{1 - \rho^2}n_r(k) + j\rho n_i(k) \quad (3.47)$$

at an SNR level of 30 dB; the components $n_r(k)$ and $n_i(k)$ are independent real valued white Gaussian processes with zero mean and unit variance (double whiteness). By chang-

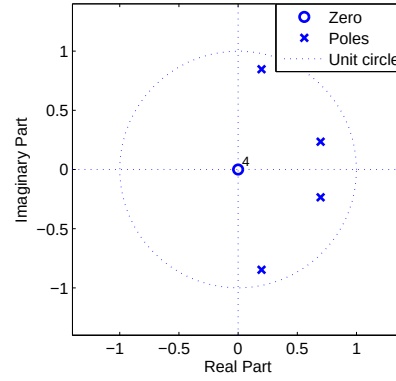


Figure 3.3: Poles of the characteristic function of the AR(4) model.

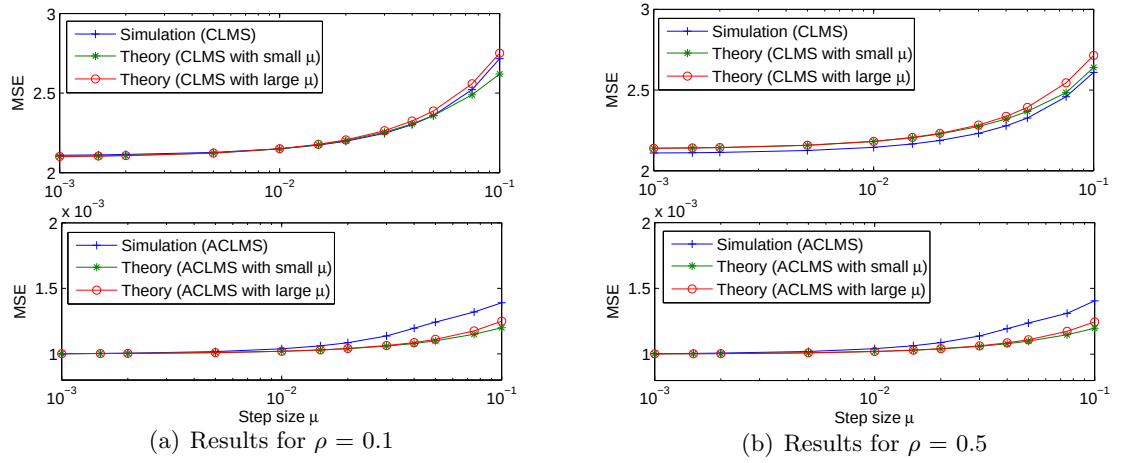


Figure 3.4: Comparison of theoretical and simulated steady state MSE of CLMS and ACLMS on a system identification task. (a) noncircular noise ($\rho = 0.1$), and (b) noncircular noise ($\rho = 0.5$).

ing the value of $\rho \in [0, 1]$, we can control the degree of noncircularity of $n(k)$. Fig. 3.4 illustrates the theoretical and simulated MSE of the CLMS and ACLMS algorithms for two cases of noncircular noise ($\rho = 0.1, 0.5$). In both the cases, theoretical results were in good agreement with the simulations with a small step-size μ . Due to the limited filter length, as step-size becomes large, the mismatch between theoretical results and their simulated counterparts becomes obvious, however, as desired, the ACLMS algorithm provided smaller MSE.

3.5 Conclusions

The standard complex LMS algorithm has been extended for the processing of the generality of complex-valued signals based on the augmented complex statistics and widely linear modelling. In comparison to the CLMS algorithm, the derived ACLMS, is capable of capturing the complete second-order information available within the signal. It has also been seen that the CLMS algorithm is a limiting case of the ACLMS algorithm, suitable only for processing proper signals, a subset of general complex-valued signals. Derivation of the proposed algorithm has been provided based on the $\mathbb{C}\mathbb{R}$ calculus framework, demonstrating its convenience for signal processing optimisation problems in the complex domain. In the second part of this chapter, we have provided the mean square steady state analysis of CLMS and ACLMS algorithms on the filtering of noncircular complex signals. The advantage of ACLMS over CLMS has been quantified via the energy conservation approach and without loss of generality, this measurement is not limited to Gaussian input data. Simulations on synthetic noncircular complex signals support the analysis.

Chapter 4

A Regularised Normalised Augmented Complex Least Mean Square Algorithm

4.1 Introduction

Stochastic gradient based adaptive filtering algorithms are designed to minimise an instantaneous error power and have been widely used in numerous signal processing applications, such as noise cancellation, system identification and adaptive prediction. The normalised least mean square (NLMS) algorithm is a standard method used in adaptive filtering due to its stability and low computational complexity [2]. The idea behind the NLMS algorithm is to employ the norm of the input vector in the denominator of the weight update term, and hence, provide decorrelation of the input and faster convergence than least mean square (LMS) algorithm. To deal with the situations where the value of the weight update term becomes extremely large due to close-to-zero input vectors, a compensation term δ , a small positive constant, is routinely employed. However, in practice, input signals with unknown and possibly very large dynamical range, and an ill-conditioned input autocorrelation matrix make it difficult to find an optimal value of δ , especially when required to meet the conflicting requirements of fast convergence and low misadjustment. To this end, the generalized normalized gradient descent (GNGD) algorithm, a variant of

the NLMS with a gradient adaptive regularisation term δ has been proposed [9]. It has been shown that the GNGD exhibits excellent stability and is robust to initialization of the compensation term δ and learning rate μ , and its sign variants have been introduced to reduce computational complexity [12].

In order to design an adequate adaptive filtering algorithm in $\mathbb{C}$ that caters for general complex-valued input signals with unknown and possibly very large dynamical range, we employ the GNGD in the ACLMS framework. We first derive the normalised ACLMS (NACLMS) algorithm and next make the regularisation term δ within NACLMS gradient adaptive. This way, the Regularised NACLMS (RNACLMS) algorithm adapts its learning rate according to the dynamics of the input signal, with the additional adaptive term also compensating for the simplifications in the derivation of NCLMS and ill-conditioned inputs. It is shown that the performance is bounded from below by the performance of the NACLMS, whereas the proposed algorithm exhibits enhanced convergence in various adverse environments where the NACLMS diverges. Simulations in the prediction setting for signals with varying degrees of circularity support the analysis.

4.2 Derivation of RNACLMS algorithm

The output $y(k)$ of an finite impulse response (FIR) adaptive filter based on the widely linear model can be written as

$$y(k) = \underbrace{\mathbf{x}^T(k)\mathbf{h}(k)}_{\text{conventional update}} + \underbrace{\mathbf{x}^H(k)\mathbf{g}(k)}_{\text{conjugate update}} \quad (4.1)$$

where $\mathbf{h}(k)$ and $\mathbf{g}(k)$ are the $L \times 1$ weight vectors, $\mathbf{x}(k)$ denotes the $L \times 1$ input vector $\mathbf{x}(k) = [x(k-1), \dots, x(k-L)]^T$ and the error $e(k)$ is obtained as $e(k) = d(k) - y(k)$, where $d(k)$ is the desired signal.

To update the conventional weight vector $\mathbf{h}(k)$ and the conjugate weight vector $\mathbf{g}(k)$, based on the minimum disturbance principle, the standard constrained minimisation problem for normalised CLMS (NCLMS) algorithm is modified to cater for the widely linear

model, that is

$$\begin{aligned}
J(k) &= \|\mathbf{h}(k+1) - \mathbf{h}(k)\|^2 + \|\mathbf{g}(k+1) - \mathbf{g}(k)\|^2 \\
&\quad + \lambda_r \Re[(d(k) - \mathbf{x}^T(k)\mathbf{h}(k+1) - \mathbf{x}^H(k)\mathbf{g}(k+1))] \\
&\quad + \lambda_i \Im[(d(k) - \mathbf{x}^T(k)\mathbf{h}(k+1) - \mathbf{x}^H(k)\mathbf{g}(k+1))]
\end{aligned} \tag{4.2}$$

where $J(k)$ is the real-valued cost function of complex variable, λ_r and λ_i are real-valued Lagrange multipliers, and $\Re[\cdot]$ and $\Im[\cdot]$ denote the real and imaginary parts of complex numbers respectively. In case of widely linear modelling, the constraint becomes $d(k) = \mathbf{x}^T(k)\mathbf{h}(k+1) + \mathbf{x}^H(k)\mathbf{g}(k+1)$. Now we can define a complex Lagrange multiplier $\lambda = \lambda_r + j\lambda_i$, and (4.2) can be rewritten in the form

$$\begin{aligned}
J(k) &= \|\mathbf{h}(k+1) - \mathbf{h}(k)\|^2 + \|\mathbf{g}(k+1) - \mathbf{g}(k)\|^2 + \Re[(d(k) \\
&\quad - \mathbf{x}^T(k)\mathbf{h}(k+1) - \mathbf{x}^H(k)\mathbf{g}(k+1))^* \cdot \lambda]
\end{aligned} \tag{4.3}$$

For the cost function $J(k)$, the maximum change in the gradient is in the direction of $\frac{\partial J(k)}{\partial \mathbf{g}^*(k+1)}$. Since $J(k)$ is real-valued and so is not directly differentiable in $\mathbb{C}$, we resort to the $\mathbb{C}\mathbb{R}$ calculus. Following the so-called $\mathbb{C}\mathbb{R}$ derivatives present in Appendix A, we obtain

$$\frac{\partial J(k)}{\partial \mathbf{g}^*(k+1)} = \frac{1}{2} \left(\frac{\partial J(k)}{\partial \mathbf{g}_r(k+1)} + j \frac{\partial J(k)}{\partial \mathbf{g}_i(k+1)} \right) \tag{4.4}$$

where $(\cdot)_r$ and $(\cdot)_i$ denote respectively the real and imaginary part of complex gradient. To calculate the update for the conjugate weight vector $\mathbf{g}$, the cost function $J(k)$ in (4.3) can be rewritten as

$$\begin{aligned}
J(k) &= [\mathbf{h}(k+1) - \mathbf{h}(k)]^H [\mathbf{h}(k+1) - \mathbf{h}(k)] + [\mathbf{g}(k+1) - \mathbf{g}(k)]^H [\mathbf{g}(k+1) - \mathbf{g}(k)] \\
&\quad + \frac{1}{2} [(d^*(k) - \mathbf{h}^H(k+1)\mathbf{x}^*(k) - \mathbf{g}^H(k+1)\mathbf{x}(k))\lambda] \\
&\quad + \frac{1}{2} [(d(k) - \mathbf{h}^T(k+1)\mathbf{x}(k) - \mathbf{g}^T(k+1)\mathbf{x}^*(k))\lambda^*]
\end{aligned}$$

and therefore

$$\frac{\partial J(k)}{\partial \mathbf{g}_r(k+1)} = \mathbf{g}(k+1) + \mathbf{g}^*(k+1) - \mathbf{g}(k) - \mathbf{g}^*(k) - \frac{1}{2}\mathbf{x}(k)\lambda - \frac{1}{2}\mathbf{x}^*(k)\lambda^* \quad (4.5)$$

and

$$\begin{aligned} \frac{\partial J(k)}{\partial \mathbf{g}_i(k+1)} &= -j\mathbf{g}(k+1) + j\mathbf{g}^*(k+1) + j\mathbf{g}(k) - j\mathbf{g}^*(k) + \frac{j}{2}\mathbf{x}(k)\lambda - \frac{j}{2}\mathbf{x}^*(k)\lambda^* \\ &= -j(\mathbf{g}(k+1) - \mathbf{g}^*(k+1) - \mathbf{g}(k) + \mathbf{g}^*(k) - \frac{1}{2}\mathbf{x}(k)\lambda + \frac{1}{2}\mathbf{x}^*(k)\lambda^*) \end{aligned} \quad (4.6)$$

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

$$\frac{\partial J(k)}{\partial \mathbf{g}^*(k+1)} = \mathbf{g}(k+1) - \mathbf{g}(k) - \frac{1}{2}\mathbf{x}(k)\lambda \quad (4.7)$$

Setting this derivative to zero, the update of conjugate weight vector $\mathbf{g}(k)$ is obtained as

$$\mathbf{g}(k+1) - \mathbf{g}(k) = \frac{1}{2}\mathbf{x}(k)\lambda \quad (4.8)$$

In the same way, the update of conventional weight vector $\mathbf{h}(k)$ is given by

$$\mathbf{h}(k+1) - \mathbf{h}(k) = \frac{1}{2}\mathbf{x}^*(k)\lambda \quad (4.9)$$

Premultiplying both sides of (4.8) by $\mathbf{x}^H(k)$ and both sides of (4.9) by $\mathbf{x}^T(k)$ yields

$$\begin{aligned} \mathbf{x}^H(k)(\mathbf{g}(k+1) - \mathbf{g}(k)) &= \|\mathbf{x}(k)\|^2 \frac{\lambda}{2} \\ \mathbf{x}^T(k)(\mathbf{h}(k+1) - \mathbf{h}(k)) &= \|\mathbf{x}(k)\|^2 \frac{\lambda}{2} \end{aligned}$$

Upon adding these two equations together and using the constraint (4.3) and the widely linear model (4.1), the Lagrange multiplier is obtained as

$$\lambda = \frac{e(k)}{\|\mathbf{x}(k)\|^2} \quad (4.10)$$

Thus, the NACLMS updates of the conventional weight vector $\mathbf{h}(k)$ and the conjugate

weight vector $\mathbf{g}(k)$ become

$$\mathbf{h}(k+1) = \mathbf{h}(k) + \frac{\mu e(k) \mathbf{x}^*(k)}{2\|\mathbf{x}(k)\|^2}$$

$$\mathbf{g}(k+1) = \mathbf{g}(k) + \frac{\mu e(k) \mathbf{x}(k)}{2\|\mathbf{x}(k)\|^2}$$

To preserve stability for close-to-zero input vectors, the compensation term δ is added to the above updates, giving the final weight updates of the NACLMS algorithm in the form

$$\mathbf{h}(k+1) = \mathbf{h}(k) + \frac{\mu e(k) \mathbf{x}^*(k)}{2\|\mathbf{x}(k)\|^2 + \delta} \quad (4.11)$$

$$\mathbf{g}(k+1) = \mathbf{g}(k) + \frac{\mu e(k) \mathbf{x}(k)}{2\|\mathbf{x}(k)\|^2 + \delta} \quad (4.12)$$

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

$$\mathbf{w}^a(k+1) = \mathbf{w}^a(k) + \frac{\mu e(k) \mathbf{x}^{a*}(k)}{\|\mathbf{x}^a(k)\|^2 + \delta} \quad (4.13)$$

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

4.2.1 Gradient Adaptive Regularisation Factor

When dealing with signals with unknown and possibly very large dynamical range, a fixed regularisation parameter δ may lead to divergence and poor performance of the NACLMS algorithm. To this end, following the standard gradient adaptive approach [9], we introduce an gradient adaptive δ , updated as

$$\delta(k) = \delta(k-1) - \rho \nabla_{\delta(k-1)} E(k) \quad (4.14)$$

where $E(k) = \frac{|e^2(k)|}{2} = \frac{e(k)e^*(k)}{2}$ is the standard cost function, ρ is step size parameter, and $\nabla_{\delta(k-1)} E(k)$ can be derived as

$$\begin{aligned} \nabla_{\delta(k-1)} E(k) &= \frac{1}{2} \left[e(k) \frac{\partial e^*(k)}{\partial \delta(k-1)} + e^*(k) \frac{\partial e(k)}{\partial \delta(k-1)} \right] \\ &= \Re \left[e^*(k) \frac{\partial e(k)}{\partial \delta(k-1)} \right] \end{aligned} \quad (4.15)$$

Since $e(k) = d(k) - y(k)$, using the chain rule, we have

$$\frac{\partial e(k)}{\partial \delta(k-1)} = -\frac{\partial y(k)}{\partial \delta(k-1)} = -\left[\mathbf{x}^{aT}(k) \frac{\partial \mathbf{w}^a(k)}{\partial \delta(k-1)}\right] \quad (4.16)$$

From (4.13), we then obtain

$$\frac{\partial \mathbf{w}^a(k)}{\partial \delta(k-1)} = -\frac{\mu e(k-1) \mathbf{x}^{a*}(k-1)}{[\|\mathbf{x}^a(k)\|^2 + \delta(k-1)]^2} \quad (4.17)$$

to give

$$\frac{\partial e(k)}{\partial \delta(k-1)} = \frac{\mu e(k-1) \mathbf{x}^{aT}(k) \mathbf{x}^{a*}(k-1)}{[\|\mathbf{x}^a(k)\|^2 + \delta(k-1)]^2}$$

and

$$\delta(k) = \delta(k-1) - \frac{\rho \mu \Re[e^*(k) e(k-1) \mathbf{x}^{aT}(k) \mathbf{x}^{a*}(k-1)]}{[\|\mathbf{x}^a(k)\|^2 + \delta(k-1)]^2}$$

The conventional NCLMS algorithm and the derivation of RNCLMS algorithm are given in the Appendix B.

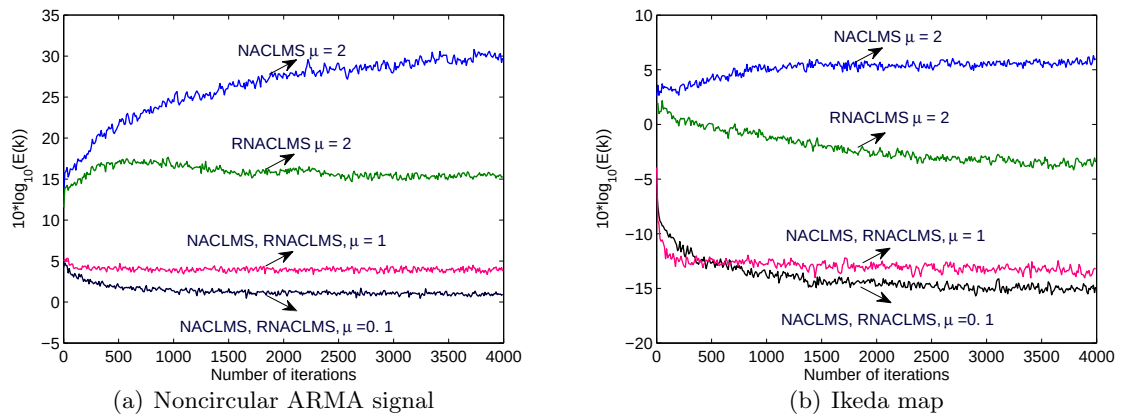
4.3 Simulations

In the simulations, the length of the FIR adaptive filters L was set to be 10 to guarantee a sufficient order for one-step ahead signal prediction task. The aim to add the regularisation factor $\delta(0)$ in the denominator of weight update of NLMS based algorithms was to avoid close-to-zero input vector, and in order not to affect the performance of adaptive filters, $\delta(0)$ should be kept small as compared with input signals. In all simulations, $\delta(0)$ was set to be 0.1, whereas its step-size parameter $\rho = 0.1$. For generality, both circular and noncircular benchmark and real-world nonstationary and noncircular wind signals were used in simulations. The performance of algorithms on synthetic signals was produced by averaging 100 independent trails, whereas single trial experiments were performed on real-world wind signals.

To illustrate the enhanced robustness and stability of employing a gradient adaptive regularisation factor, a performance comparison between the NACLMS and RNACLMS

Table 4.1: Comparison of prediction gains R_p for the various classes of signals

R_p [dB]	Circular AR(4) (2.24)	Noncircular ARMA (2.25)	Ikeda map (2.27)
CLMS	2.4372	5.2089	3.0157
ACLMS	2.3940	5.3920	5.1448
RNCLMS	2.5943	5.2618	3.5325
RNACLMS	2.5620	5.4152	5.2600
R_p [dB]	Wind (<i>low</i>)	Wind (<i>medium</i>)	Wind (<i>high</i>)
CLMS	3.5182	6.8953	10.0822
ACLMS	4.1213	7.2127	10.3517
RNCLMS	3.8854	7.4442	10.1501
RNACLMS	4.4291	7.7391	10.4543

Figure 4.1: Learning curves for the NACLMS and RNACLMS algorithms. (a) non-circular ARMA signal (2.25) and (b) Ikeda map (2.27) with $\mu = 0.1, 1$ and 2 .

algorithms was performed on prediction of the noncircular ARMA signal (2.25) and Ikeda map (2.27). Fig. 4.1 shows the corresponding learning curves for various values of μ . By design, the performance of RNACLMS was similar to that of NACLMS when NACLMS was stable, whereas RNACLMS did not diverge in cases when NACLMS did, as in the case of $\mu = 2$. Fig. 4.2 shows the learning curves of the conventional RNCLMS and RNACLMS algorithms on noncircular ARMA and Ikeda map with $\mu = 0.1$. Observe that both algorithms exhibited comparable rate of convergence, however, in the steady state, the RNACLMS algorithm achieved superior performance, indicating the suboptimality of standard RNCLMS algorithm when dealing with noncircular signals.

To further illustrate the advantage of using the widely linear model, in the next experiment, the prediction gain R_p was used as an overall performance measurement,

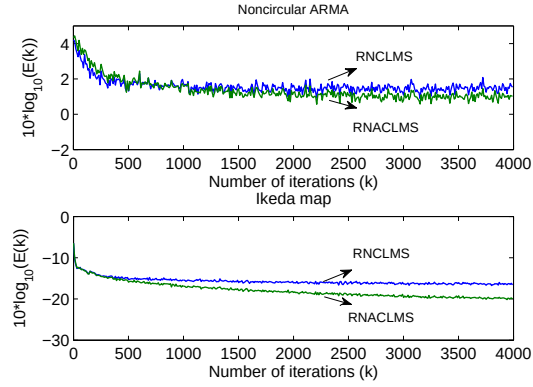


Figure 4.2: Learning curves for standard RNCLMS and the proposed RNACLMS algorithms on one step ahead prediction of the noncircular ARMA signal (2.25) and Ikeda map (2.27) with $\mu = 0.1$.

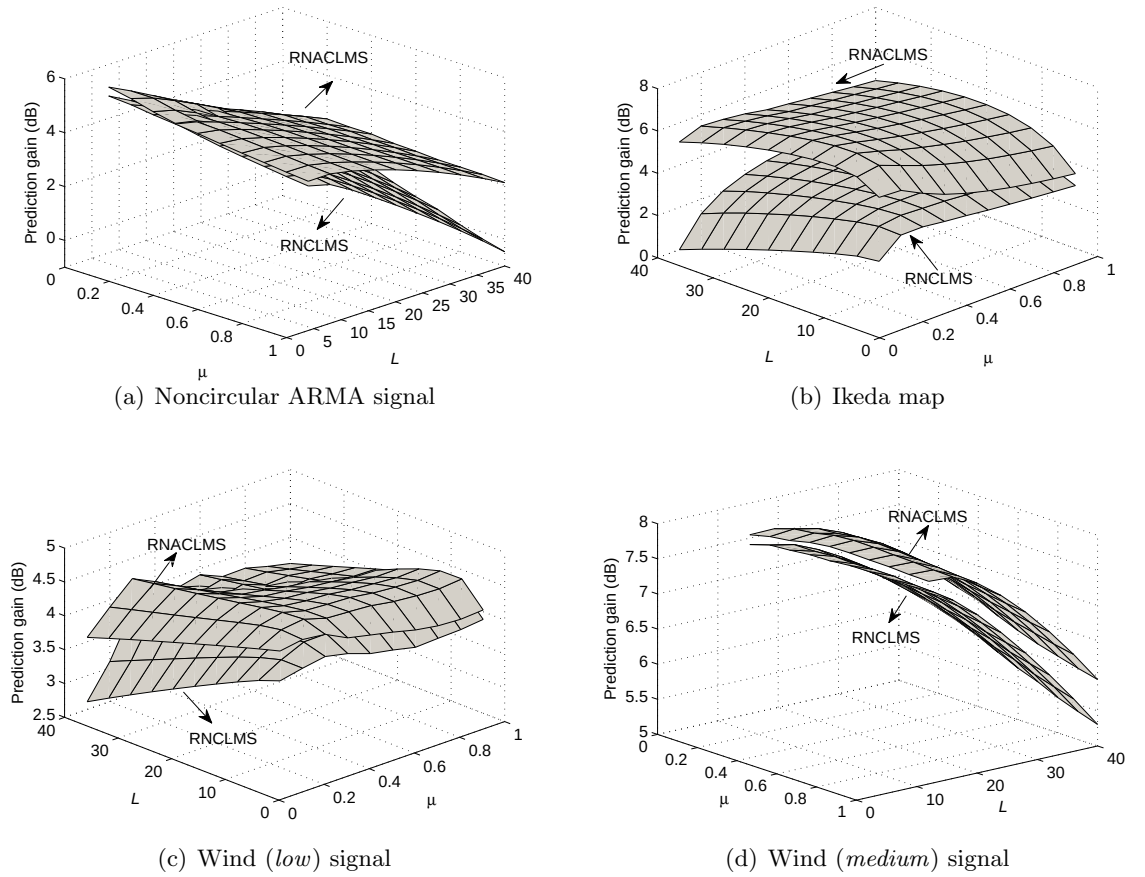


Figure 4.3: Prediction gains R_p of the NACLMS and RNACLMS algorithms against different μ and L , (a) noncircular ARMA signal (2.25), (b) Ikeda map (2.27), (c) wind (*low*) and wind (*medium*).

defined as

$$R_p = 10 \log_{10} (\sigma_x^2 / \hat{\sigma}_e^2) \text{ [dB]} \quad (4.18)$$

where σ_x^2 denotes the variance of the input signal $x(k)$, and $\hat{\sigma}_e^2$ denotes the variance of the prediction error $e(k)$ [53]. Table 4.1 shows the prediction gains R_p of the proposed RNACLMS, standard RNCLMS, ACLMS and standard CLMS for $\mu = 0.1$. As expected, for all the noncircular signals considered, there was significant improvement when augmented algorithms were employed over that of standard algorithms, and RNACLMS always achieved best performance. The prediction gains R_p of both algorithms against different step-size and filter length are given in Fig. 4.3, where the robustness of the advantage of widely linear RNACLMS over its strictly linear counterpart is observed.

4.4 Conclusions

A Regularised Normalised ACLMS (RNACLMS) algorithm has been introduced for adaptive filtering of general complex valued signals. It has been derived by utilising the widely linear model and a gradient adaptive regularisation factor within the NACLMS. Simulations on both synthetic and real world complex-valued signals have illustrated the advantage of RNACLMS algorithm over standard RNCLMS algorithm for adaptive filtering of noncircular signals. It also has been shown that by employing a gradient adaptive compensation term, the performance of RNACLMS algorithm is bounded from below by the performance of the NACLMS algorithm, whereas it achieves enhanced convergence and stability when NACLMS algorithm diverges.

Chapter 5

An Augmented Affine Projection Algorithm

5.1 Introduction

Stochastic gradient based adaptive filtering algorithms are designed to minimise an instantaneous error power and have been widely used in numerous signal processing applications, such as noise cancelation, system identification and adaptive prediction. The Normalized Least Mean Square (NLMS) algorithm is the most popular choice, due to its stability, fast convergence and low complexity. Learning algorithms based on the least squares estimation, such as the Recursive Least Squares (RLS) algorithm [2], converge faster, however, in practical applications, they suffer from high computational complexity. To combine the benefits of both approaches, that is excellent stability, low computational complexity and fast convergence, the affine projection algorithm (APA), based on affine subspace projections, has been introduced as a link between the NLMS and RLS algorithm [11, 54]. This way, the APA updates the weight vector on the basis of both past and current input vectors, resulting in faster convergence as compared to the NLMS algorithm [12, 55], and much fewer computations and enhanced stability as compared to the RLS algorithm [56].

These class of APAs have been originally introduced for real-valued signals. However, some real world processes, such as vector fields and directional signals with “intensity” and “direction” components, are best understood when considered complex-valued [41].

Unfortunately, standard adaptive filtering algorithms in $\mathbb{C}$ are straightforward extensions of the corresponding algorithms in $\mathbb{R}$, and do not make full use of the algebraic structure of the complex domain. For instance, it is commonly assumed that the covariance matrix of a zero mean complex vector $\mathbf{z}$ is $E\{\mathbf{z}\mathbf{z}^H\}$, and is seen as an extension of the real covariance $E\{\mathbf{z}\mathbf{z}^T\}$, achieved by replacing the vector transpose operator by the Hermitian transpose. However, recent advances in the so-called augmented complex statistics show that the complex covariance matrix $E\{\mathbf{z}\mathbf{z}^H\}$ is not sufficient to fully describe second order statistics in $\mathbb{C}$, and the pseudo-covariance matrix $E\{\mathbf{z}\mathbf{z}^T\}$ should also be considered. This proves particularly important when processing second order noncircular (improper) signals [38].

In the context of the APAs, a widely linear affine projection algorithm (WL-APA) for DS-CDMA applications has been proposed by Lima and Lamare [57]. This algorithm is based on the minimization of real-valued error vector, as required in the DS-CDMA setting, where the error vector is given by

$$\mathbf{e}_{APA}^{WL} = \mathbf{b}_{APA} - \Re[\mathbf{U}_{APA}\mathbf{w}_{APA}^{WL}] \quad (5.1)$$

and $\mathbf{U}_{APA}$ is the matrix of complex-valued received signals, deteriorated by complex noise, $\mathbf{b}_{APA}$ are the desired real-valued input symbols, i.e. amplitude-shift keying (ASK) or binary phase-shift keying (BPSK), $\mathbf{w}_{APA}^{WL}$ is the weight vector for estimating the coefficients of adaptive linear multiuser receivers, and $\Re[\cdot]$ denotes the real part of complex numbers. However, as this algorithm is developed for DS-CDMA and uses a real-valued output error, it does not fully exploit the advantages of the widely linear model [31]. Recent attempts to make full use of the information available in the widely linear model include the approach in [58], which optimises two separate cost functions for the “standard” and “conjugate” parts of the widely linear model.

A rigorously derived augmented APA algorithm for the processing of general complex signals (both circular and noncircular) is therefore still missing. To this end, we now provide a general augmented APA (AAPA), which uses a global cost function in the derivation and is a generic extension of standard APA for the filtering of second order noncircular signals. This algorithm simplifies into the normalized ACLMS when only

a scalar instantaneous error is considered. The analysis of steady state performance is provided based on the energy conservation principle [18] and illustrates the advantage of AAPA over APA for the filtering of general complex signals.

The chapter is organised as follows. We first use the unified framework, $\mathbb{CR}$ calculus introduced in Appendix A, to derive the augmented affine projection algorithm (AAPA) in an adaptive prediction context. The improvement in the performance of AAPA as compared to the standard APA is supported by both theoretical analysis and simulations performed on benchmark complex-valued both circular and noncircular signals, as well as on real world noncircular wind measurements.

5.2 Derivation of Augmented APA

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

$$\mathbf{y}(k) = \underbrace{\mathbf{X}^T(k)\mathbf{h}(k)}_{\text{standard part}} + \underbrace{\mathbf{X}^H(k)\mathbf{g}(k)}_{\text{conjugate part}} \quad (5.2)$$

where $\mathbf{h}(k)$ and $\mathbf{g}(k)$ are the $L \times 1$ weight vectors, and $\mathbf{X}(k)$ is the $L \times K$ data matrix in the filter memory at time instant k , defined as [11]

$$\mathbf{X}(k) = [\mathbf{x}(k - K + 1), \dots, \mathbf{x}(k)] \quad (5.3)$$

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

$$\mathbf{x}(k) = [x(k - 1), \dots, x(k - L)]^T \quad (5.4)$$

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

$$\mathbf{d}(k) = [d(k - K + 1), \dots, d(k)]^T \quad (5.5)$$

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

$$\mathbf{e}(k) = \mathbf{d}(k) - \mathbf{y}(k) \quad (5.6)$$

To update the weight vectors $\mathbf{g}(k)$ of the “conjugate” part and $\mathbf{h}(k)$ of the “standard” part of a widely linear affine projection filter, similar to the standard APA approach [4], based on the principle of minimum disturbance, we need to solve a constrained minimisation problem given by

$$J(k) = \|\mathbf{h}(k+1) - \mathbf{h}(k)\|^2 + \|\mathbf{g}(k+1) - \mathbf{g}(k)\|^2 + \Re[(\mathbf{d}(k) - \mathbf{X}^T(k)\mathbf{h}(k+1) - \mathbf{X}^H(k)\mathbf{g}(k+1))^H \cdot \mathbf{\Lambda}] \quad (5.7)$$

where the $K \times 1$ dimensional vector $\mathbf{\Lambda}$ comprises Lagrange multipliers and $J(k)$ is a real function of complex variables. Based on the widely linear model (5.2), the constraints in (5.7) become

$$\mathbf{d}(k) = \mathbf{X}^T(k)\mathbf{h}(k+1) + \mathbf{X}^H(k)\mathbf{g}(k+1) \quad (5.8)$$

To calculate the update for the “conjugate” part of the widely linear model, that is, for the weight vector $\mathbf{g}(k)$, the cost function $J(k)$ can be rewritten as

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

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

$$\frac{\partial J(k)}{\partial \mathbf{g}^*(k+1)} = \frac{1}{2} \left(\frac{\partial J(k)}{\partial \mathbf{g}_r(k+1)} + j \frac{\partial J(k)}{\partial \mathbf{g}_i(k+1)} \right) \quad (5.9)$$

where

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

and

$$\begin{aligned}\frac{\partial J(k)}{\partial \mathbf{g}_i(k+1)} &= -j\mathbf{g}(k+1) + j\mathbf{g}^*(k+1) + j\mathbf{g}(k) - j\mathbf{g}^*(k) + \frac{j}{2}\mathbf{X}(k)\mathbf{\Lambda} - \frac{j}{2}\mathbf{X}^*(k)\mathbf{\Lambda}^* \\ &= -j(\mathbf{g}(k+1) - \mathbf{g}^*(k+1) - \mathbf{g}(k) + \mathbf{g}^*(k) - \frac{1}{2}\mathbf{X}(k)\mathbf{\Lambda} + \frac{1}{2}\mathbf{X}^*(k)\mathbf{\Lambda}^*)(5.10)\end{aligned}$$

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

$$\frac{\partial J(k)}{\partial \mathbf{g}^*(k+1)} = \mathbf{g}(k+1) - \mathbf{g}(k) - \frac{1}{2}\mathbf{X}(k)\mathbf{\Lambda} \quad (5.11)$$

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

$$\mathbf{g}(k+1) - \mathbf{g}(k) = \frac{1}{2}\mathbf{X}(k)\mathbf{\Lambda} \quad (5.12)$$

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

$$\mathbf{h}(k+1) - \mathbf{h}(k) = \frac{1}{2}\mathbf{X}^*(k)\mathbf{\Lambda} \quad (5.13)$$

Following the standard APA approach [4], to eliminate the Lagrange multiplier vector $\mathbf{\Lambda}$ from (5.12) and (5.13), we shall premultiply both sides of (5.12) by $\mathbf{X}^H(k)$ and both sides of (5.13) by $\mathbf{X}^T(k)$ to yield

$$\begin{aligned}\mathbf{X}^H(k)(\mathbf{g}(k+1) - \mathbf{g}(k)) &= \frac{1}{2}\mathbf{X}^H(k)\mathbf{X}(k)\mathbf{\Lambda} \\ \mathbf{X}^T(k)(\mathbf{h}(k+1) - \mathbf{h}(k)) &= \frac{1}{2}\mathbf{X}^T(k)\mathbf{X}^*(k)\mathbf{\Lambda}\end{aligned}$$

Upon adding these two equations together and using the constraints (5.8) and the widely linear model (5.2), the output error becomes

$$\mathbf{e}(k) = \frac{1}{2}(\mathbf{X}^H(k)\mathbf{X}(k) + \mathbf{X}^T(k)\mathbf{X}^*(k))\mathbf{\Lambda} \quad (5.14)$$

giving the Lagrange multiplier vector

$$\mathbf{\Lambda} = 2[\mathbf{X}^H(k)\mathbf{X}(k) + \mathbf{X}^T(k)\mathbf{X}^*(k)]^{-1}\mathbf{e}(k) \quad (5.15)$$

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

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

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

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

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

$$\mathbf{h}(k+1) = \mathbf{h}(k) + \mu \mathbf{X}^*(k) [\mathbf{X}^H(k) \mathbf{X}(k) + \mathbf{X}^T(k) \mathbf{X}^*(k) + \delta \mathbf{I}]^{-1} \mathbf{e}(k) \quad (5.16)$$

$$\mathbf{g}(k+1) = \mathbf{g}(k) + \mu \mathbf{X}(k) [\mathbf{X}^H(k) \mathbf{X}(k) + \mathbf{X}^T(k) \mathbf{X}^*(k) + \delta \mathbf{I}]^{-1} \mathbf{e}(k) \quad (5.17)$$

In the derivation, the estimation is assumed to be of sufficient order. Compared with the weight update of standard APA [4], the AAPA includes both the estimate of the covariance matrix $\mathbf{X}^T(k) \mathbf{X}^*(k)$ and its complex conjugate $\mathbf{X}^H(k) \mathbf{X}(k)$ in the denominator term within the update for both the weight vectors $\mathbf{h}(k)$ and $\mathbf{g}(k)$. Observe that, as desired, when the observation length is $K = 1$, AAPA degenerates into the normalised ACLMS (NACLMS), given by

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

and when the denominator is omitted, we have the standard ACLMS algorithm, given in Chapter 3.

5.3 Steady state MSE performance of APA and AAPA

The aim of this section is to demonstrate analytically the suitability of the proposed AAPA to filter complex-valued noncircular data, for which the standard APA is suboptimal, and to provide insight into its convergence. Consider a second order noncircular (improper)

teaching signal $d(k)$ arising from the widely linear model

$$d(k) = \mathbf{x}^T(k)\mathbf{h}_o + \mathbf{x}^H(k)\mathbf{g}_o + n(k) \quad (5.18)$$

where $\mathbf{h}_o$ and $\mathbf{g}_o$ are unknown optimal weight vectors that we wish to estimate, $n(k)$ represents the measurement noise with zero mean and variance σ_n^2 , and $\mathbf{x}(k)$ is the input vector with a positive-definite covariance matrix $\mathbf{C}_{\mathbf{xx}} = E[\mathbf{xx}^H]$.

5.3.1 Steady state analysis of standard APA for noncircular signals

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

$$\text{MSE} = \lim_{k \rightarrow \infty} E[|e(k)|^2] \quad (5.19)$$

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

$$e(k) = d(k) - \mathbf{x}^T(k)\mathbf{h}(k) \quad (5.20)$$

and $\mathbf{h}(k)$ is the weight vector, for which, the update is given by [4]

$$\mathbf{h}(k+1) = \mathbf{h}(k) + \mu \mathbf{X}^*(k) [\mathbf{X}^T(k)\mathbf{X}^*(k) + \delta \mathbf{I}]^{-1} \mathbf{e}(k) \quad (5.21)$$

We next follow the approach from [16], based on the energy conservation principle, to formulate a closed form solution of the MSE for improper signals represented by the widely linear model (5.18). Using this approach, the update in (5.21) can be rewritten in terms of the weight error vector $\tilde{\mathbf{h}}(k) = \mathbf{h}_o - \mathbf{h}(k)$ as

$$\tilde{\mathbf{h}}(k+1) = \tilde{\mathbf{h}}(k) - \mu \mathbf{X}^*(k) [\mathbf{X}^T(k)\mathbf{X}^*(k) + \delta \mathbf{I}]^{-1} \mathbf{e}(k) \quad (5.22)$$

Multiplying both sides of (5.22) by $\mathbf{X}^T(k)$ gives

$$\mathbf{X}^T(k)\tilde{\mathbf{h}}(k+1) = \mathbf{X}^T(k)\tilde{\mathbf{h}}(k) - \mu \mathbf{X}^T(k)\mathbf{X}^*(k) [\mathbf{X}^T(k)\mathbf{X}^*(k) + \delta \mathbf{I}]^{-1} \mathbf{e}(k) \quad (5.23)$$

Then the *a posteriori* and *a priori* error vectors $\mathbf{e}_p(k)$ and $\mathbf{e}_a(k)$ can be introduced as [4]

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

Then from (5.23), the following equation stands

$$\mathbf{e}_p(k) = \mathbf{e}_a(k) - \mu \mathbf{X}^T(k) \mathbf{X}^*(k) [\mathbf{X}^T(k) \mathbf{X}^*(k) + \delta \mathbf{I}]^{-1} \mathbf{e}(k) \quad (5.24)$$

We can use (5.24) to solve for $\mathbf{e}(k)$, assuming $\mathbf{X}^T(k) \mathbf{X}^*(k)$ is invertible

$$[\mathbf{X}^T(k) \mathbf{X}^*(k) + \delta \mathbf{I}]^{-1} \mathbf{e}(k) = \frac{1}{\mu} [\mathbf{X}^T(k) \mathbf{X}^*(k)]^{-1} (\mathbf{e}_a(k) - \mathbf{e}_p(k)) \quad (5.25)$$

and substitute into (5.22) to obtain

$$\tilde{\mathbf{h}}(k+1) = \tilde{\mathbf{h}}(k) - \mathbf{X}^*(k) [\mathbf{X}^T(k) \mathbf{X}^*(k)]^{-1} (\mathbf{e}_a(k) - \mathbf{e}_p(k)) \quad (5.26)$$

which can be rearranged as

$$\tilde{\mathbf{h}}(k+1) + \mathbf{X}^*(k) [\mathbf{X}^T(k) \mathbf{X}^*(k)]^{-1} \mathbf{e}_a(k) = \tilde{\mathbf{h}}(k) + \mathbf{X}^*(k) [\mathbf{X}^T(k) \mathbf{X}^*(k)]^{-1} \mathbf{e}_p(k) \quad (5.27)$$

Making the standard assumption that at the steady state $\tilde{\mathbf{h}}(k+1) \approx \tilde{\mathbf{h}}(k)$, we find the following energy conservation principle [18]

$$\|\tilde{\mathbf{h}}(k+1)\|^2 + \mathbf{e}_a^H(k) [\mathbf{X}^T(k) \mathbf{X}^*(k)]^{-1} \mathbf{e}_a(k) = \|\tilde{\mathbf{h}}(k)\|^2 + \mathbf{e}_p^H(k) [\mathbf{X}^T(k) \mathbf{X}^*(k)]^{-1} \mathbf{e}_p(k) \quad (5.28)$$

Taking the statistical expectation operator $E[\cdot]$, and using the steady-state condition $E[\|\tilde{\mathbf{h}}(k+1)\|^2] = E[\|\tilde{\mathbf{h}}(k)\|^2]$ as $k \rightarrow \infty$, we can obtain

$$\mathbf{e}_a^H(k) [\mathbf{X}^T(k) \mathbf{X}^*(k)]^{-1} \mathbf{e}_a(k) = \mathbf{e}_p^H(k) [\mathbf{X}^T(k) \mathbf{X}^*(k)]^{-1} \mathbf{e}_p(k) \quad (5.29)$$

Substituting (5.24) into right hand side (RHS) of (5.29), we get

$$\begin{aligned} E[\mathbf{e}_p^H(k)[\mathbf{X}^T(k)\mathbf{X}^*(k)]^{-1}\mathbf{e}_p(k)] &= E[\mathbf{e}_a^H(k)[\mathbf{X}^T(k)\mathbf{X}^*(k)]^{-1}\mathbf{e}_a(k)] + \mu^2 E[\mathbf{e}^H(k)\mathbf{A}(k)\mathbf{e}(k)] \\ &\quad - \mu E[\mathbf{e}_a^H(k)\mathbf{B}(k)\mathbf{e}(k)] - \mu E[\mathbf{e}^H(k)\mathbf{B}(k)\mathbf{e}_a(k)] \end{aligned} \quad (5.30)$$

where

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

and

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

Comparing (5.29) with (5.30), we can obtain

$$\mu E[\mathbf{e}^H(k)\mathbf{A}(k)\mathbf{e}(k)] = E[\mathbf{e}_a^H(k)\mathbf{B}(k)\mathbf{e}(k)] + E[\mathbf{e}^H(k)\mathbf{B}(k)\mathbf{e}_a(k)] \quad (5.31)$$

Substituting (5.18) into (5.20), we can obtain

$$\mathbf{e}(k) = \mathbf{e}_a(k) + \mathbf{n}(k) + \mathbf{X}^H(k)\mathbf{g}_o \quad (5.32)$$

where $\mathbf{n}(k) = [n(k), n(k-1), \dots, n(k-K+1)]^T$. Note that since we assume a widely linear model, compared with the work in [18], an extra term “ $\mathbf{X}^H(k)\mathbf{g}_o$ ” is included. With the assumptions that the noise $n(k)$ is *i.i.d.* and statistically independent of the input matrix $\mathbf{X}(k)$, and at the state state $\mathbf{X}(k)$ is statistically independent of $\mathbf{e}_a(k)$, $E[\mathbf{e}_a(k)] = \mathbf{0}$ as $k \rightarrow \infty$, and the first term in (5.31) becomes

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

where the second term has the form

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

and the third term

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

Then (5.31) can be rewritten as

$$\begin{aligned} & \mu E[\mathbf{e}_a^H(k)\mathbf{A}(k)\mathbf{e}_a(k)] + \mu E[\mathbf{n}^H(k)\mathbf{A}(k)\mathbf{n}(k)] \\ & + \mu E[(\mathbf{X}^H(k)\mathbf{g}_o)^H \mathbf{A}(k)\mathbf{X}^H(k)\mathbf{g}_o] = 2E[\mathbf{e}_a^H(k)\mathbf{B}(k)\mathbf{e}_a(k)] \end{aligned} \quad (5.33)$$

To compute these statistical expectations, we shall further assume $E[\mathbf{e}_a(k)\mathbf{e}_a^H(k)] = E[|e_a(k)|^2] \cdot \mathbf{S}$, where $\mathbf{S} \approx \mathbf{I}$ for small μ , and $\mathbf{S} \approx \mathbf{1} \cdot \mathbf{1}^T$ for large μ , with $\mathbf{1} = [1, 0, \dots, 0]^T$.

As $k \rightarrow \infty$, the first term in (5.33) now becomes

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

where symbol $\text{Tr}(\cdot)$ denotes the matrix trace operator. Similarly, the second term in (5.33) becomes

$$\mu E[\mathbf{n}^H(k)\mathbf{A}(k)\mathbf{n}(k)] = \mu \sigma_n^2 \text{Tr}(E[\mathbf{A}(k)])$$

whereas the third and fourth terms become

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

and

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

where

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

With the introduction of following equations

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

When the regularisation parameter δ is small enough, $\mathbf{A}(k) = \mathbf{B}(k)$, and $\eta_\mu = \alpha_\mu$. The excess mean square error (EMSE), can be defined as

$$\text{EMSE} = \lim_{k \rightarrow \infty} E[|e_a(k)|^2] \quad (5.34)$$

In this case, the EMSE of the standard APA, denoted by EMSE_s , becomes

$$\text{EMSE}_s = \frac{\mu\sigma_n^2 \text{Tr}(E[\mathbf{A}(k)]) + \mu \mathbf{g}_o^H E[\mathbf{C}(k)] \mathbf{g}_o}{2\eta_\mu - \mu\alpha_\mu} \quad (5.35)$$

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

$$\text{EMSE}_s = \frac{\mu\sigma_n^2}{2 - \mu} + \frac{\mu \mathbf{g}_o^H E[\mathbf{C}(k)] \mathbf{g}_o}{\text{Tr}(E[\mathbf{A}(k)])} \quad (5.36)$$

On the other hand, for a large learning rate, $\mathbf{S} \approx \mathbf{1} \cdot \mathbf{1}^T$, (5.35) becomes

$$\text{EMSE}_s = \frac{\mu\sigma_n^2 \text{Tr}(E[\mathbf{A}(k)]) + \mu \mathbf{g}_o^H E[\mathbf{C}(k)] \mathbf{g}_o}{(2 - \mu)E[\mathbf{A}_k(1, 1)]}$$

where $\mathbf{A}_k(1, 1)$ is the first element of $\mathbf{A}(k)$. When δ is small enough and μ is large, the following assumptions hold [18]

$$\text{Tr}(E[\mathbf{A}(k)]) \simeq E\left[\frac{K}{\|\mathbf{x}(k)\|^2}\right] \quad \text{and} \quad \frac{1}{E[\mathbf{A}_k(1, 1)]} = \text{Tr}(\mathbf{C}_{xx}) \quad (5.37)$$

to get

$$\text{EMSE}_s = \frac{\mu\sigma_n^2}{2 - \mu} \text{Tr}(\mathbf{C}_{xx}) \left(E\left[\frac{K}{\|\mathbf{x}(k)\|^2}\right] + \mu \mathbf{g}_o^H E[\mathbf{C}(k)] \mathbf{g}_o \right) \quad (5.38)$$

Note that when δ is small enough, the following approximation exists

$$\mathbf{g}_o^H \mathbf{C}(k) \mathbf{g}_o \simeq \mathbf{g}_o^H \mathbf{X}(k) [\mathbf{X}^T(k) \mathbf{X}^*(k) + \delta \mathbf{I}]^{-1} (\mathbf{g}_o^H \mathbf{X}(k))^H \quad (5.39)$$

Since $\delta > 0$, the inversion term $[\mathbf{X}^T(k) \mathbf{X}^*(k) + \delta \mathbf{I}]^{-1}$ is positive definite [59]. For a complex-valued noncircular signal, generated from a widely linear model (5.18), $\mathbf{g}_o^H \mathbf{X}(k)$ is a $1 \times K$ vector with all non-zero elements. Following the properties of positive definite matrix [60], $\mathbf{g}_o^H \mathbf{C}(k) \mathbf{g}_o > 0$, then $\mu \mathbf{g}_o^H E[\mathbf{C}(k)] \mathbf{g}_o > 0$.

Compared with the expressions of EMSE_s for standard APA for circular signals in [18], both (5.36) and (5.38) contain the positive-valued additional term $\mathbf{g}_o^H \mathbf{C}(k) \mathbf{g}_o$ due to the “conjugate” term “ $\mathbf{X}^H(k) \mathbf{g}_o$ ” within the widely linear model (5.18), illustrating that the standard APA is a suboptimal solution for the filtering of second order improper (noncircular) signals. Using the first row in (5.32), we finally obtain the steady state mean square error of standard APA for improper signals in the form

$$\begin{aligned} \text{MSE}_s &= \lim_{k \rightarrow \infty} E[|e(k)|^2] \\ &= \text{EMSE}_s + E[|\mathbf{x}^H(k) \mathbf{g}_o|^2] + \sigma_n^2 \end{aligned} \quad (5.40)$$

5.3.2 Steady state analysis of Augmented APA for noncircular signals

Similarly to the analysis in Section 5.3.1, the update of “standard” weight vector of AAPA in (5.16) can be rewritten in terms of its weight error vector $\tilde{\mathbf{h}}(k) = \mathbf{h}_o - \mathbf{h}(k)$ as

$$\tilde{\mathbf{h}}(k+1) = \tilde{\mathbf{h}}(k) - \mu \mathbf{X}^*(k) [\mathbf{X}^H(k) \mathbf{X}(k) + \mathbf{X}^T(k) \mathbf{X}^*(k) + \delta \mathbf{I}]^{-1} \mathbf{e}(k) \quad (5.41)$$

Multiplying both sides of (5.16) by $\mathbf{X}^T(k)$ gives

$$\mathbf{X}^T(k) \tilde{\mathbf{h}}(k+1) = \mathbf{X}^T(k) \tilde{\mathbf{h}}(k) - \mu \mathbf{X}^T(k) \mathbf{X}^*(k) [\mathbf{X}^H(k) \mathbf{X}(k) + \mathbf{X}^T(k) \mathbf{X}^*(k) + \delta \mathbf{I}]^{-1} \mathbf{e}(k) \quad (5.42)$$

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

$$\tilde{\mathbf{g}}(k+1) = \tilde{\mathbf{g}}(k) - \mu \mathbf{X}(k) [\mathbf{X}^H(k) \mathbf{X}(k) + \mathbf{X}^T(k) \mathbf{X}^*(k) + \delta \mathbf{I}]^{-1} \mathbf{e}(k) \quad (5.43)$$

and

$$\mathbf{X}^H(k) \tilde{\mathbf{g}}(k+1) = \mathbf{X}^H(k) \tilde{\mathbf{g}}(k) - \mu \mathbf{X}^H(k) \mathbf{X}(k) [\mathbf{X}^H(k) \mathbf{X}(k) + \mathbf{X}^T(k) \mathbf{X}^*(k) + \delta \mathbf{I}]^{-1} \mathbf{e}(k) \quad (5.44)$$

Note that both the terms $\tilde{\mathbf{h}}(k)$ and $\tilde{\mathbf{g}}(k)$ have the same generic form, and that the denominator in the above updates is different from that of standard APA in (5.22). Then the a

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

$$\mathbf{e}_p(k) = \mathbf{X}^T(k)\tilde{\mathbf{h}}(k+1) + \mathbf{X}^H(k)\tilde{\mathbf{g}}(k+1) \quad (5.45)$$

$$\mathbf{e}_a(k) = \mathbf{X}^T(k)\tilde{\mathbf{h}}(k) + \mathbf{X}^H(k)\tilde{\mathbf{g}}(k) \quad (5.46)$$

and the sum of (5.42) and (5.44) gives the relationship between the *a posteriori* and *a priori* error vectors $\mathbf{e}_p(k)$ and $\mathbf{e}_a(k)$ in the form

$$\mathbf{e}_p(k) = \mathbf{e}_a(k) - \mu(\mathbf{X}^H(k)\mathbf{X}(k) + \mathbf{X}^T(k)\mathbf{X}^*(k))[\mathbf{X}^H(k)\mathbf{X}(k) + \mathbf{X}^T(k)\mathbf{X}^*(k) + \delta\mathbf{I}]^{-1}\mathbf{e}(k) \quad (5.47)$$

The next step is to use both the *a posteriori* and *a priori* error vectors in the update equations (5.41) and (5.43) to obtain the updates of weight error vectors as functions of $\mathbf{e}_a(k)$ and $\mathbf{e}_p(k)$

$$\tilde{\mathbf{h}}(k+1) = \tilde{\mathbf{h}}(k) - \mathbf{X}^*(k)[\mathbf{X}^H(k)\mathbf{X}(k) + \mathbf{X}^T(k)\mathbf{X}^*(k)]^{-1}(\mathbf{e}_a(k) - \mathbf{e}_p(k)) \quad (5.48)$$

and

$$\tilde{\mathbf{g}}(k+1) = \tilde{\mathbf{g}}(k) - \mathbf{X}(k)[\mathbf{X}^H(k)\mathbf{X}(k) + \mathbf{X}^T(k)\mathbf{X}^*(k)]^{-1}(\mathbf{e}_a(k) - \mathbf{e}_p(k)) \quad (5.49)$$

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

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

and

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

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

$$\begin{aligned} E[\mathbf{e}_a^H(k)[\mathbf{X}^H(k)\mathbf{X}(k) + \mathbf{X}^T(k)\mathbf{X}^*(k)]^{-1}\mathbf{e}_a(k)] = \\ E[\mathbf{e}_p^H(k)[\mathbf{X}^H(k)\mathbf{X}(k) + \mathbf{X}^T(k)\mathbf{X}^*(k)]^{-1}\mathbf{e}_p(k)] \end{aligned} \quad (5.50)$$

Similarly to the analysis in Section 5.3.1, it can be shown that, for a small learning rate μ , the EMSE of AAPA, denoted as EMSE_a , is given by

$$\text{EMSE}_a = \frac{\mu\sigma_n^2}{2 - \mu} \quad (5.51)$$

whereas for a large μ , we obtain

$$\text{EMSE}_a = \frac{\mu\sigma_n^2}{2 - \mu} \text{Tr}(\mathcal{C}_{xx}) E\left[\frac{K}{\|\mathbf{x}(k)\|^2}\right] \quad (5.52)$$

and

$$\text{MSE}_a = \text{EMSE}_a + \sigma_n^2 \quad (5.53)$$

Note that in [58], a widely linear APA (WL-APA) has been derived by solving two separate constrained minimisation problems, where

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

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

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

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

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

Table 5.1: Prediction gains R_p of standard APA and the proposed AAPA for various noncircular signals

R_p [dB]	Noncircular ARMA (2.25)	Ikeda (2.27)	Wind (<i>low</i>)	Wind (<i>medium</i>)	Wind (<i>high</i>)
APA	2.4641	1.8734	3.6774	5.8851	9.3610
AAPA	3.0072	4.6517	4.0500	6.2368	9.9181

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

5.3.3 Comparison of steady state performance of APA and AAPA

Consider, without loss in generality, a complex-valued noncircular signal is generated from a widely linear model (5.18), so that $\mu \mathbf{g}_o^H E[\mathbf{C}(k)] \mathbf{g}_o > 0$. This, in turn, makes $\text{EMSE}_s > \text{EMSE}_a$ ($\text{MSE}_s > \text{MSE}_a$) for both small and large learning rates at the steady state. On the other hand, when a complex-valued circular signal is considered, the weight vector $\mathbf{g}_o = \mathbf{0}$, leading to $\text{EMSE}_s = \text{EMSE}_a$ ($\text{MSE}_s = \text{MSE}_a$), resulting in both the APA and AAPA having the same performance at the steady state.

5.4 Simulations

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

Fig. 5.1 shows learning curves of AAPA for the noncircular ARMA process and Ikeda map, plotted for different small step-sizes, indicating that for noncircular signals, for a small step-size, increasing the step-size of AAPA will cause a larger MSE at the steady state, although with faster convergence. Fig. 5.2 illustrates the advantage of the AAPA

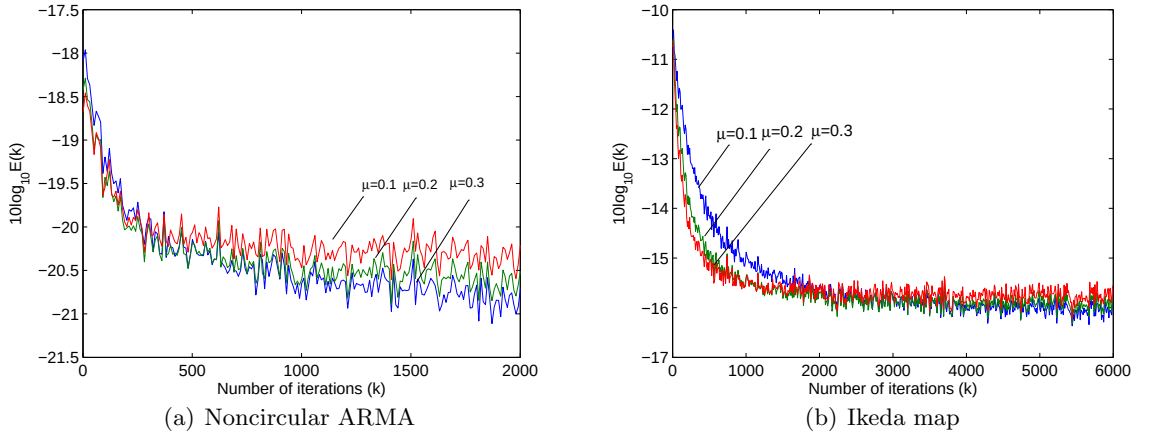


Figure 5.1: Learning curves of AAPA for the noncircular ARMA model (2.25) and noncircular Ikeda map (2.27) for $K = 1$ and various learning rate μ .

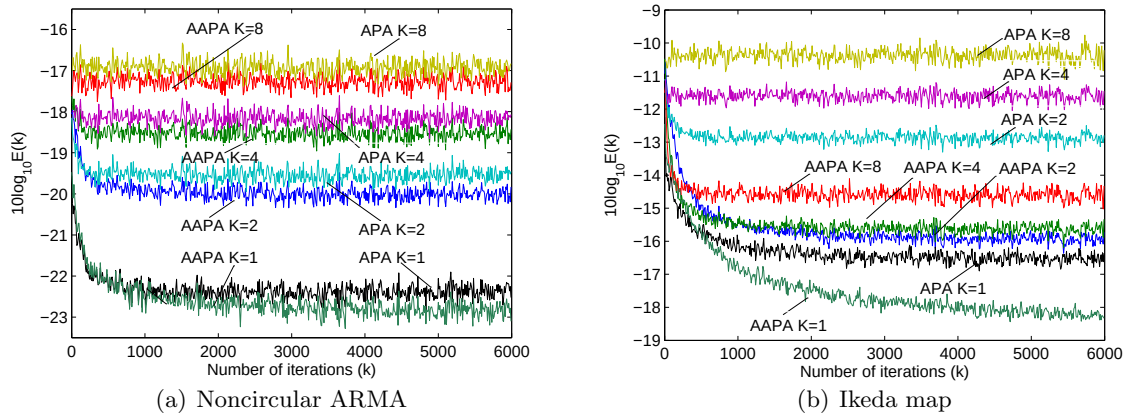


Figure 5.2: Learning curves of APA and AAPA for the noncircular ARMA model (2.25) and noncircular Ikeda map (2.27), obtained by using $\mu = 0.1$ and $K = 1, 2, 4$, and 8.

at steady state over the standard APA on both the noncircular ARMA model and Ikeda map for the step-size was $\mu = 0.1$, and observation lengths $K = 1, 2, 4$, and 8, wher in case of $K = 1$, the standard APA and AAPA degenerate into the normalised CLMS (NCLMS) and normalised ACLMS (NACLMS) derived in Chapter 4. Note that the main purpose to use K previous input vectors in weight update of affine projection based algorithms is to increase the speed of convergence of NCLMS-type filters at a computational cost that is still comparable to that of NCLMS, however, the penalty to be paid is the increased estimation error at the steady-state. In some special cases when the input signals are

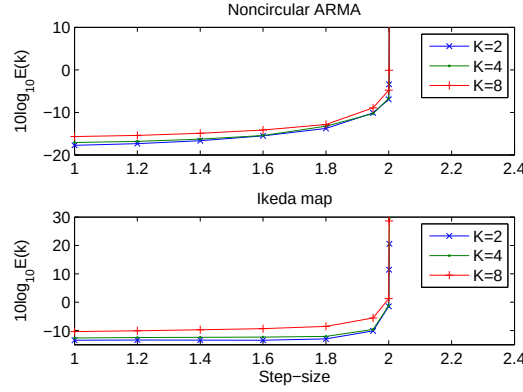


Figure 5.3: MSE of AAPA as a function of step-size μ for the noncircular ARMA model (2.25) and Ikeda map (2.27).

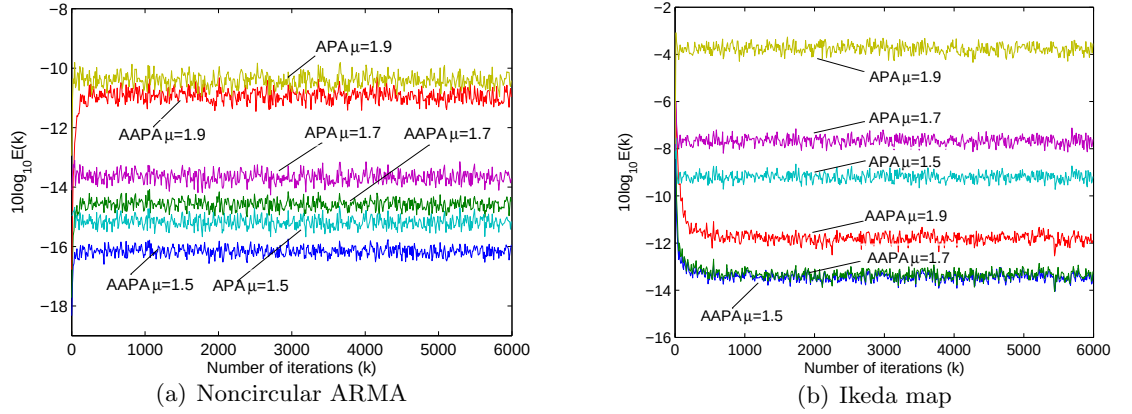


Figure 5.4: Learning curves of APA and AAPA for the noncircular ARMA model (2.25) and noncircular Ikeda map (2.27) using different large μ and $K = 2$.

highly correlated, which is commonly encountered in echo cancellation applications, both faster convergence and smaller steady-state error can be achieved by APA algorithm as compared with NCLMS [61, 62].

In the next experiment, the performances of APA and AAPA for large μ were considered. The results in [18] indicate that the stability bound on μ is approximately $0 < \mu < 2$ for APA, and Fig. 5.3 illustrates the stability bound of AAPA on the noncircular signals was $\mu < 2$. The simulation results in Fig. 5.4 illustrate that with a fixed observation length K , an increment in the step-size will cause MSE of AAPA to increase at the steady state of AAPA, whereas Fig. 5.5 shows that with a fixed large μ , the MSE at the steady state increases with an increment in K . Note that both Fig. 5.4 and Fig. 5.5 illustrate

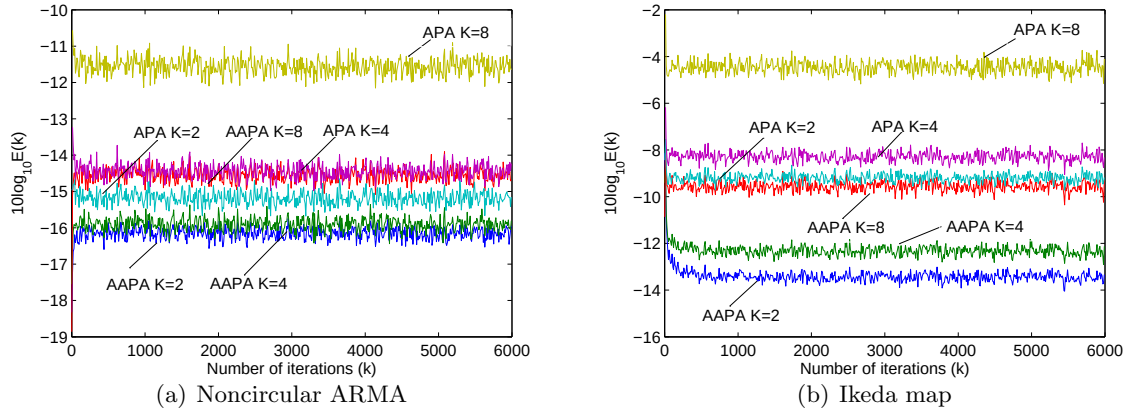


Figure 5.5: Learning curves of APA and AAPA for the noncircular ARMA model (2.25) and noncircular Ikeda map (2.27) using $\mu = 1.5$ and $K = 2, 4$, and 8 .

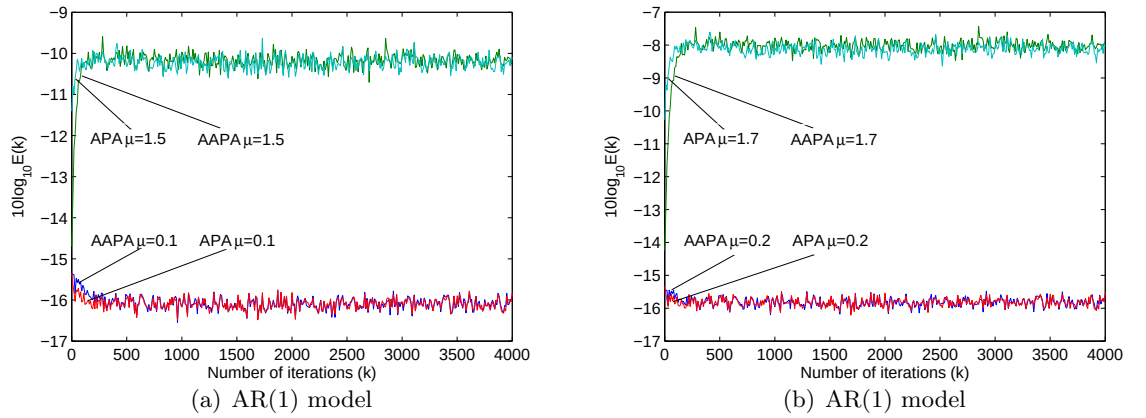


Figure 5.6: Learning curves of APA and AAPA for the proper AR(4) model (2.24) using different small and large μ and $K = 1$.

the advantage of AAPA over APA for noncircular signals considered with a large μ . The learning curves of both APA and AAPA on the circular AR(4) model for $K = 1$ and $\mu = 0.1, 0.2, 1.5$, and 1.7 are shown in Fig. 5.6, which indicates that APA converges faster than AAPA for the circular signal, however, at the steady state, there is no clear difference between the performances. Thus, although AAPA optimizes both the weight vectors $\mathbf{h}(k)$ and $\mathbf{g}(k)$ of the widely linear model, for circular signals, $\mathbf{g}_o = \mathbf{0}$, leading to the same model as that of standard APA.

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

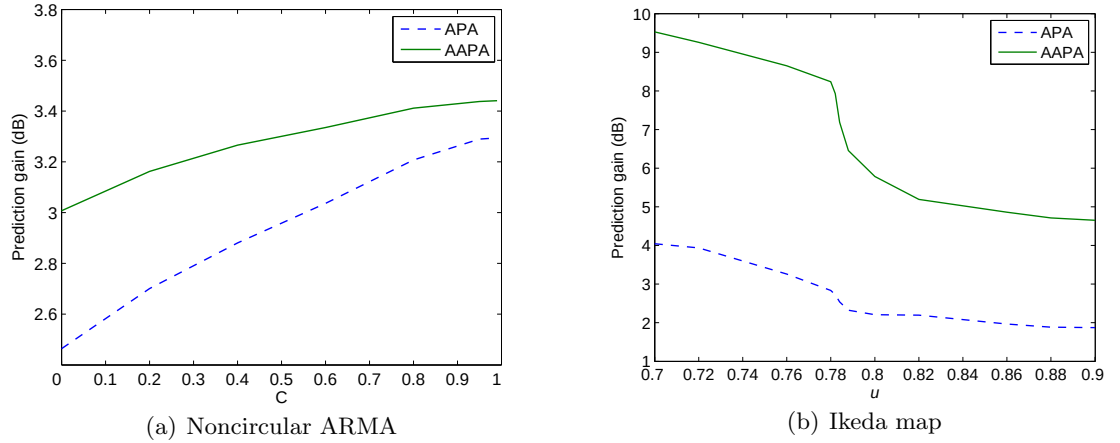


Figure 5.7: Prediction gains of APA and AAPA for varying values of the control parameters of noncircular ARMA (2.25) and Ikeda map (2.27).

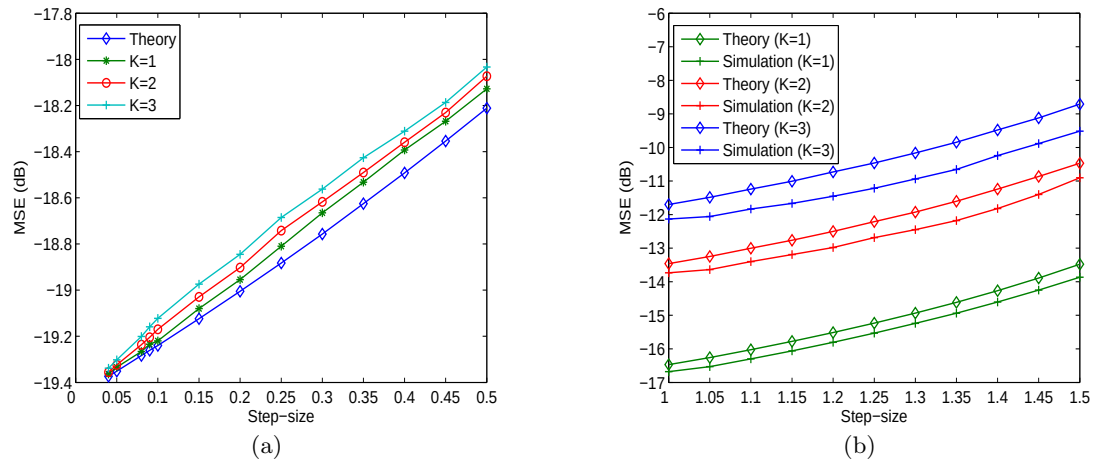


Figure 5.8: Comparison of theoretical and simulated steady state MSE of AAPA on noncircular ARMA (2.25) when $K = 1, 2$ and 3 using (a) (5.51) and (b) (5.52).

(4.18). Table 5.1 shows the prediction gains R_p of the proposed AAPA and standard APA for $\mu = 0.2$ and $K = 2$. As expected, for all noncircular signals considered, there was a significant improvement when the AAPA was employed over that of the standard APA. Fig. 5.7 shows the effects of the control parameters C (2.26) and u on the performances of standard APA and AAPA for noncircular ARMA (2.25) and Ikeda map (2.27), in all the cases, the performance of AAPA was superior than that of APA. Note that in Fig. 5.7(a), AAPA gradually loses advantage when C increases to 1, which conforms with the experimental results in [45].

We next performed statistical performance analysis on the noncircular ARMA signal in one step ahead prediction setting. Fig. 5.8 (a) and Fig. 5.8 (b) illustrate the theoretical steady state MSE of AAPA using (5.51) and (5.52) as a function of step-size with $K = 1, 2$, and 3 , and both showed good agreement with the simulation results.

5.5 Conclusions

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

Chapter 6

An Augmented Echo State Network for Nonlinear Adaptive Filtering of Complex Noncircular Signals

6.1 Introduction

Recurrent neural networks (RNNs) are a class of nonlinear adaptive filters with feedback, whose computational power stems from their ability to act as universal approximators for any continuous function on a compact domain [3, 63]. Owing to their rich inherent memory through feedback, RNNs have found applications in the modelling of highly nonlinear dynamic systems and the associated attractor dynamics. They are typically used in the system identification [64, 65], time series prediction [53], [66], and adaptive noise cancellation settings [67, 68], where for the nonstationary and nonlinear nature of the signals and typically long impulse responses, using the class of static feedforward networks or transversal filters would result in undermodelling [3, 69, 70].

Recently, a class of discrete time RNNs, called Echo State Networks (ESNs) have been introduced, with the aim to reduce the complexity of computation encountered by standard RNNs [20]. The principle behind ESNs is to separate the RNN architecture into

two constituent components: a recurrent architecture, called the ‘dynamical reservoir’ or ‘hidden layer’, and a memoryless output layer, called the ‘readout neuron’. The recurrent architecture consists of a randomly generated group of hidden neurons with a specified degree of recurrent connections, and should satisfy the so called ‘echo state property’ to maintain stability [71]. This way, the high computational complexity of RNNs is significantly reduced due to the sparse connections among the hidden neurons; in addition, the learning requirements are reduced to only the weights connecting the hidden layer and the readout neuron¹.

Many real world bivariate processes, such as vector fields and directional signals with “intensity” and “direction” components, are most conveniently represented when considered complex-valued [41]. However, the design of complex-valued RNNs is not straightforward, and one of the major issues is the suitability of complex nonlinear activation function. According to Liouville’s theorem, the only bounded and analytic function in $\mathbb{C}$ is constant [26]. In order to avoid the difficulties to find suitable activation functions (AF), for convenience, previous studies have mostly focused on the so-called split-complex activation functions, where the real and imaginary components of the complex-valued input signal are split and fed through two parallel real-valued activation functions [72, 73]. The split-complex approach has been shown to yield reasonable performance for some applications in channel equalisation [72, 74], where the coupling between the real and imaginary components within the complex-valued signals is not strong, however, for the common case where the two components are strongly correlated, RNNs based on split-complex activation function tend to yield poor performance [75]. To allow for a fully “complex” processing of complex-valued signals as a whole, meromorphic functions² have been employed as complex nonlinear activation functions due to their property that they are analytic everywhere except for a discrete subset of $\mathbb{C}$. Consequently, in the neural network literature, several important approaches have been extended to the complex domain; examples include a class of complex backpropagation [74, 76], the complex real-time recurrent learning (CRTRL)

¹A recent special issue of Neural Networks, vol. 20, no. 3, 2007, edited by H. Jaeger, W. Maass and J. C. Principe, was dedicated solely to echo state networks and liquid state machines.

²A meromorphic function is a single-valued function that is analytic in all but possibly a discrete subset of its domain, and at those singularities it must approach infinity “like a polynomial”.

algorithm [27, 72], and applications have been found in image enhancement [77], gray-scale image processing [78], and and geometric figure transformation [79].

The first extension of ESNs into the complex domain $\mathbb{C}$ was proposed in [80]; this network had a linear output mapping, and was trained by the complex-valued Wiener filter, thus making the network second order optimal for the processing of circular stationary data. Results in adaptive filtering dealing with real-world complex-valued data suggest that, due to the linearity of the output mapping, the degree of universal function approximation exhibited by standard ESNs may not be sufficient. To that end, a nonlinear output layer within ESNs, that is, the linear mapping followed by a nonlinear activation function, has been proposed in [81]. To deal with common problems experienced in neural network training, such as saturation and slow convergence, resulting from the unknown and large dynamics of inputs, the nonlinear output layer of ESNs has further been equipped with an adaptive amplitude of the nonlinearity [81].

Recently, augmented complex statistics have been introduced into several key learning algorithms in the context of RNNs to perform nonlinear adaptive filtering of general complex-valued signals (both second order circular and noncircular); examples include the augmented complex extended Kalman filter (ACEKF) [35], and augmented complex real-time recurrent learning (ACRTRL) [34]. Following on these results, we here introduce augmented statistics into the training of complex ESNs, allowing us to make use of all the available second order statistical information, and to produce optimal estimates for second order noncircular (improper) signals.

This chapter is organised as follows. In Section 6.2, the standard ESN is briefly introduced and the augmented complex ESN and its nonlinear variants are derived by using the $\mathbb{C}\mathbb{R}$ calculus. In Section 6.3, the merits of nonlinearity and widely linear modelling are discussed. Simulations on both synthetic circular and noncircular signals and real-world nonstationary and noncircular wind signals are given in Section 6.4, demonstrating the advantage of the augmented ESN over standard complex ESNs. Finally, Section 6.5 concludes the chapter.

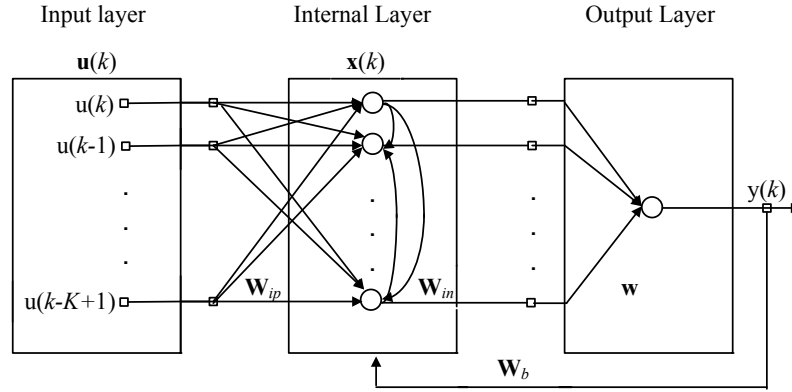


Figure 6.1: The architecture of an Echo State Network (ESN).

6.2 Derivation of the augmented ESN

6.2.1 Standard complex ESN with a linear output mapping

Fig. 6.1 shows the architecture of a standard ESN, which is composed of K external input neurons, L readout neurons, and N internal units. Without loss in generality, we shall address ESNs with one readout neuron ($L=1$), as this facilitates the nonlinear adaptive filtering setting within the ESN architecture. The input and internal weights are stored respectively in the $(N \times K)$ and $(N \times N)$ weight matrices $\mathbf{W}_{ip}$, $\mathbf{W}_{in}$, vector $\mathbf{w}_b$ comprises the the feedback weights connecting the readout neuron and the internal units, vector $\mathbf{x}(k)$ is the $(N \times 1)$ internal state vector, $\mathbf{u}(k)$ represents the $(K \times 1)$ input vector, and $y(k)$ is the overall output. The network state at time instant k , denoted by $\mathbf{q}(k)$, is a concatenation of the input $\mathbf{u}(k)$, internal state $\mathbf{x}(k)$ and the delayed output $y(k-1)$, that is

$$\mathbf{q}(k) = [u(k), \dots, u(k-K+1), x_1(k), \dots, x_N(k), y(k-1)]^T \quad (6.1)$$

whereas the internal unit dynamics are described by [71]

$$\mathbf{x}(k) = \Phi(\mathbf{W}_{ip}\mathbf{u}(k) + \mathbf{W}_{in}\mathbf{x}(k-1) + \mathbf{w}_b y(k-1)) \quad (6.2)$$

where $\Phi(\cdot)$ is a vector-valued nonlinear activation function of the neurons within the reservoir.

The echo state property is provided by randomly choosing an internal weight matrix

$\mathbf{W}_{in}$ and performing scaling to make the spectral radius $\rho(\mathbf{W}_{in}) < 1$, thus ensuring that the network is stable; the input and feedback weights can be initialized arbitrarily [71]. For an ESN with a linear output mapping, the output $y(k)$ is given by

$$y(k) = \mathbf{q}^T(k) \mathbf{w}(k) \quad (6.3)$$

where $\mathbf{w}(k)$ is the weight vector corresponding to the output layer. Its update can be performed, e.g. based on the complex least mean square (CLMS) algorithm, given by [24]

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu e(k) \mathbf{q}^*(k) \quad (6.4)$$

6.2.2 Augmented complex ESN with a linear readout neuron

Based on the widely linear model given in Section 2.1, we shall now derive the augmented widely linear stochastic gradient algorithm for the training of complex ESNs, thus making them suitable for processing general complex-valued signals (both circular and noncircular). To this end, we introduce the augmented network state $\mathbf{q}^a(k)$ as³

$$\mathbf{q}^a(k) = [u(k), \dots, u(k-K+1), x_1^a(k), \dots, x_N^a(k), y(k-1), u^*(k), \dots, u^*(k-K+1)]^T \quad (6.5)$$

Since the input weights of the ESN stored in matrix $\mathbf{W}_{ip}$ are randomly chosen prior to training, we can use another matrix $\mathbf{W}_{ip}^a$ to initialise the weights associated with the conjugate input vector $\mathbf{u}^*(k)$. The internal state transition within the augmented ESN is therefore described by

$$\mathbf{x}^a(k) = \Phi(\mathbf{W}_{ip} \mathbf{u}(k) + \mathbf{W}_{in} \mathbf{x}^a(k-1) + \mathbf{w}_b y(k-1) + \mathbf{W}_{ip}^a \mathbf{u}^*(k)) \quad (6.6)$$

Due to the specific properties of the ESN architecture, the output $y(k)$ of the augmented ESN with linear output mapping is governed by an asymmetric version of the widely linear

³This augmented state is not a straightforward application of the widely linear model in (2.7) and is specific to ESNs.

model in (2.7), to yield

$$y(k) = \mathbf{v}^T(k)\mathbf{h}(k) + \mathbf{u}^H(k)\mathbf{g}(k) \quad (6.7)$$

where $\mathbf{v}(k)$ is defined as $\mathbf{v}(k) = [u(k), \dots, u(k - K + 1), x_1^a(k), \dots, x_N^a(k), y(k - 1)]^T$, and $\mathbf{h}(k)$ and $\mathbf{g}(k)$ denote respectively the conventional and conjugate output weight vectors.

Note that the ESN has a local feedback (from the output to the internal state) and thus unlike standard feedback structures [41, 82], the output within the state vector (6.5) of the ESN does not require augmentation with its conjugate. Therefore due to local feedback, the conjugate weight vector $\mathbf{g}(k)$ is only associated with the conjugate input signal. The standard and conjugate weight vectors are thus of different dimensions, but as with all widely linear models, the conjugate weight vector $\mathbf{g}(k) = \mathbf{0}$ for a circular input signal.

6.2.3 Training of augmented ESNs

The ESN is trained based on the cost function

$$J(k) = \frac{1}{2}|e(k)|^2 = \frac{1}{2}e(k)e^*(k) \quad (6.8)$$

where $e(k)$ is the instantaneous output error $e(k) = d(k) - y(k)$, and $d(k)$ is the desired (teaching) signal. The update of the conjugate weight vector $\mathbf{g}(k)$ in (6.7) is given by

$$\mathbf{g}(k + 1) = \mathbf{g}(k) - \mu \nabla_{\mathbf{g}} J(k) \quad (6.9)$$

where μ is the learning rate. Note that $J(k)$ is a real-valued function dependent on both output errors $e(k)$ and $e^*(k)$. Following the $\mathbb{R}^*$ -derivative (A.3) in Appendix A, it can be shown that

$$\begin{aligned} \nabla_{\mathbf{g}} J(k) &= \frac{\partial J(k)}{\partial \mathbf{g}^*(k)} \\ &= \frac{1}{2} \left(\frac{\partial J(k)}{\partial \mathbf{g}_r(k)} + j \frac{\partial J(k)}{\partial \mathbf{g}_i(k)} \right) \end{aligned} \quad (6.10)$$

Since both the input and error $e(k)$ are complex, the gradient from (6.10) is evaluated as

$$\begin{aligned}\frac{\partial J(k)}{\partial \mathbf{g}_r(k)} &= \frac{1}{2} \left[e(k) \frac{\partial e^*(k)}{\partial \mathbf{g}_r(k)} + e^*(k) \frac{\partial e(k)}{\partial \mathbf{g}_r(k)} \right] \\ \frac{\partial J(k)}{\partial \mathbf{g}_i(k)} &= \frac{1}{2} \left[e(k) \frac{\partial e^*(k)}{\partial \mathbf{g}_i(k)} + e^*(k) \frac{\partial e(k)}{\partial \mathbf{g}_i(k)} \right]\end{aligned}\quad (6.11)$$

Therefore,

$$\begin{aligned}\frac{\partial J(k)}{\partial \mathbf{g}_r(k)} &= -\frac{1}{2} [e(k) \mathbf{u}(k) + e^*(k) \mathbf{u}^*(k)] \\ \frac{\partial J(k)}{\partial \mathbf{g}_i(k)} &= -\frac{j}{2} [-e(k) \mathbf{u}(k) + e^*(k) \mathbf{u}^*(k)]\end{aligned}\quad (6.12)$$

giving

$$\nabla_{\mathbf{g}} J(k) = \frac{1}{2} \left(\frac{\partial J(k)}{\partial \mathbf{g}_r(k)} + j \frac{\partial J(k)}{\partial \mathbf{g}_i(k)} \right) = -\frac{1}{2} e(k) \mathbf{u}(k) \quad (6.13)$$

The update of the conjugate weight vector $\mathbf{g}(k)$ then becomes⁴

$$\mathbf{g}(k+1) = \mathbf{g}(k) + \mu e(k) \mathbf{u}(k) \quad (6.14)$$

In a similar way, the update of the weight vector $\mathbf{h}(k)$ is given by

$$\mathbf{h}(k+1) = \mathbf{h}(k) + \mu e(k) \mathbf{q}^*(k) \quad (6.15)$$

In order to express (6.14) and (6.15) in a compact vector form, we can define the augmented weight vector $\mathbf{w}^a(k)$ as

$$\mathbf{w}^a(k) = [\mathbf{h}^T(k), \mathbf{g}^T(k)]^T \quad (6.16)$$

to give the update of augmented CLMS algorithm

$$\mathbf{w}^a(k+1) = \mathbf{w}^a(k) + \mu e(k) \mathbf{q}^{a*}(k) \quad (6.17)$$

where $\mathbf{q}^a(k)$ is the augmented network state in (6.5).

Owing to the use of the widely linear model and augmented complex statistics, the augmented ESN has clear theoretical advantages over the standard ESN for the processing

⁴The factor $\frac{1}{2}$ in (6.14) has been incorporated into the learning rate μ .

of noncircular complex signals. However, due to the sparse nature of the connectivity within the reservoir, the linear output mapping may not be powerful enough for efficient modelling of signals with large nonlinear dynamics. To this end, in the next section, we introduce a nonlinear readout neuron equipped with a trainable amplitude of nonlinearity.

6.2.4 Complex ESN with a nonlinear readout neuron

The output $y(k)$ of the standard ESN with a nonlinear output layer is given by [71]

$$y(k) = \Phi(\text{net}(k)) = \Phi(\mathbf{q}^T(k)\mathbf{w}(k)) \quad (6.18)$$

where $\Phi(\cdot)$ is the output activation function, and the weight vector $\mathbf{w}(k)$ is updated by minimising the cost function $\mathcal{J}(k)$, given in (6.8), to give

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \mu \nabla_{\mathbf{w}} J(k) \quad (6.19)$$

Following the stochastic gradient, that is, $\nabla J_{\mathbf{w}}(k) = \frac{\partial J(k)}{\partial \mathbf{w}^*(k)}$, we can obtain the weight update of the complex nonlinear gradient descent (CNGD) algorithm [76] as⁵

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu e(k) \Phi'^*(\text{net}(k)) \mathbf{q}^*(k) \quad (6.20)$$

To derive the corresponding augmented CNGD algorithm for the nonlinear output layer of the augmented ESN, recall that the output $y(k)$ is given by

$$y(k) = \Phi(\text{net}(k)) = \Phi(\mathbf{v}^T(k)\mathbf{h}(k) + \mathbf{u}^H(k)\mathbf{g}(k)) \quad (6.21)$$

The update of the conjugate weight vector $\mathbf{g}(k)$ in (6.21) is given by

$$\mathbf{g}(k+1) = \mathbf{g}(k) - \mu \nabla_{\mathbf{g}} J(k) \quad (6.22)$$

⁵For most standard nonlinear activation function in $\mathbb{C}$, we have $(\Phi^*)' = \Phi'^*$.

and following the $\mathbb{R}^*$ -derivative, (as Φ is holomorphic)

$$\nabla_{\mathbf{g}} J(k) = \frac{\partial J(k)}{\partial \mathbf{g}^*(k)} = \frac{1}{2} \left(\frac{\partial J(k)}{\partial \mathbf{g}_r(k)} + j \frac{\partial J(k)}{\partial \mathbf{g}_i(k)} \right) \quad (6.23)$$

Next we have,

$$\begin{aligned} \frac{\partial J(k)}{\partial \mathbf{g}_r(k)} &= \frac{1}{2} \left[e(k) \frac{\partial e^*(k)}{\partial \mathbf{g}_r(k)} + e^*(k) \frac{\partial e(k)}{\partial \mathbf{g}_r(k)} \right] \\ \frac{\partial J(k)}{\partial \mathbf{g}_i(k)} &= \frac{1}{2} \left[e(k) \frac{\partial e^*(k)}{\partial \mathbf{g}_i(k)} + e^*(k) \frac{\partial e(k)}{\partial \mathbf{g}_i(k)} \right] \end{aligned} \quad (6.24)$$

By using the chain rule, $\frac{\partial e(k)}{\partial \mathbf{g}_r(k)} = -\Phi'(net(k))\mathbf{u}(k)$, and since $\mathbf{g}_r(k)$ is real-valued, $\frac{\partial e^*(k)}{\partial \mathbf{g}_r(k)} = \left(\frac{\partial e(k)}{\partial \mathbf{g}_r(k)} \right)^*$, giving

$$\frac{\partial J(k)}{\partial \mathbf{g}_r(k)} = -\frac{1}{2} \left[e(k) \Phi'^*(net(k))\mathbf{u}(k) + e^*(k) \Phi'(net(k))\mathbf{u}^*(k) \right] \quad (6.25)$$

In a similar way, we have

$$\frac{\partial J(k)}{\partial \mathbf{g}_i(k)} = -\frac{j}{2} \left[e(k) \Phi'^*(net(k))\mathbf{u}(k) + e^*(k) \Phi'(net(k))\mathbf{u}^*(k) \right] \quad (6.26)$$

Then the weight update of conjugate weight vector $\mathbf{g}(k)$ becomes⁶

$$\mathbf{g}(k+1) = \mathbf{g}(k) + \mu e(k) \Phi'^*(net(k))\mathbf{u}(k) \quad (6.27)$$

In a similar way, the update of $\mathbf{h}(k)$ becomes,

$$\mathbf{h}(k+1) = \mathbf{h}(k) + \mu e(k) \Phi'^*(net(k))\mathbf{q}^*(k) \quad (6.28)$$

Then the update of the augmented weight vector $\mathbf{w}^a(k)$ can be written as

$$\mathbf{w}^a(k+1) = \mathbf{w}^a(k) + \mu e(k) \Phi'^*(net(k))\mathbf{q}^{a*}(k) \quad (6.29)$$

where $\mathbf{w}^a(k) = [\mathbf{h}^T(k), \mathbf{g}^T(k)]^T$.

⁶The factor $\frac{1}{2}$ has been incorporated into the learning rate μ .

6.2.5 ESNs with an adaptive amplitude of output nonlinearity

The nonlinear output layer has been introduced into ESNs to provide a sufficient degree of nonlinearity for enhanced modelling, however, this does not automatically guarantee optimal modelling, as some parameters, such as the amplitude of the nonlinear readout neuron, need to be chosen empirically [83]. To this end, we now introduce a gradient adaptive amplitude into the output nonlinearity of ESNs. In this case, the activation function can be expressed as

$$y(k) = \Phi(\text{net}(k)) = \lambda(k)\bar{\Phi}(\text{net}(k)) \quad (6.30)$$

where the real-valued $\lambda(k) \geq 0$ denotes the amplitude of the nonlinearity $\Phi(\text{net}(k))$ and $\bar{\Phi}(\text{net}(k))$ is the unit amplitude activation function, for which $\lambda(k) = 1$. In the stochastic gradient setting, the parameter λ can be made gradient adaptive as [81, 83]

$$\lambda(k+1) = \lambda(k) - \eta \nabla J_{\lambda}(k) \quad (6.31)$$

where η is the stepsize of the amplitude update. The gradient

$$\begin{aligned} \nabla_{\lambda} J(k) &= \frac{\partial J(k)}{\partial \lambda(k)} \\ &= \frac{1}{2} \frac{\partial [e(k)e^*(k)]}{\partial \lambda(k)} \\ &= \frac{1}{2} \left[e^*(k) \frac{\partial e(k)}{\partial \lambda(k)} + e(k) \frac{\partial e^*(k)}{\partial \lambda(k)} \right] \end{aligned} \quad (6.32)$$

and since $\lambda(k)$ is real-valued, $\frac{\partial e^*(k)}{\partial \lambda(k)} = \left(\frac{\partial e(k)}{\partial \lambda(k)} \right)^*$, giving

$$\frac{\partial e(k)}{\partial \lambda(k)} = -\bar{\Phi}(\text{net}(k)) \quad (6.33)$$

The update for the amplitude of nonlinearity within adaptive amplitude CNGD (AAC-NGD) algorithm is therefore given by

$$\lambda(k+1) = \lambda(k) + \frac{\eta}{2} \left[e^*(k)\bar{\Phi}(\text{net}(k)) + e(k)\bar{\Phi}^*(\text{net}(k)) \right] \quad (6.34)$$

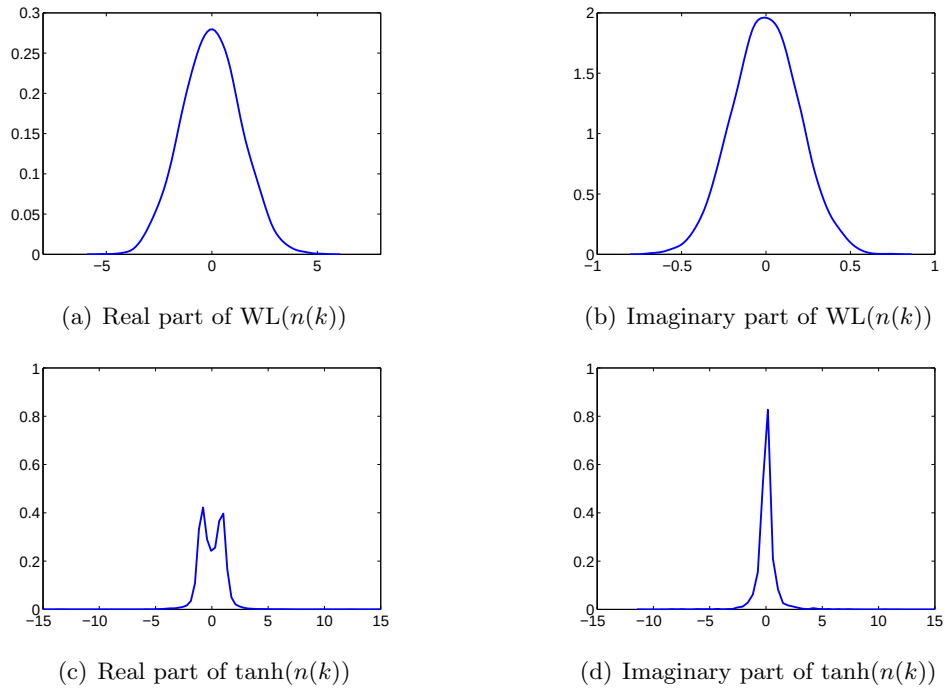


Figure 6.2: Probability density functions obtained from a doubly white circular noise after applying the widely linear model and tanh nonlinearity.

and applies to both standard (CGND) and augmented (ACNGD) learning algorithms; for more detail, see [41].

6.3 Merits of nonlinearity and widely linear model

To illustrate the effect of nonlinearity and the widely linear model, we generated a circular doubly white noise $n(k)$, with zero mean and unit variance, and passed it through a complex-valued tanh nonlinearity, defined in (6.35), and the widely linear model, given by $x(k) = WL(n(k)) = 0.6n(k) + 0.8n^*(k)$. Fig. 6.2(a) and Fig. 6.2(b) show that the application of the widely linear does not change the Gaussian natures of the real and imaginary parts - it only alters the power ratio. Fig. 6.2(c) and Fig. 6.2(d) show that the application of tanh nonlinearity alters the character of distribution, which cannot be achieved by using the widely linear model. It is therefore natural to combine the widely linear model and nonlinear processing in order to deal simultaneously with various aspects of the data nature. By introducing the adaptive amplitude of nonlinearity, we have an

Table 6.1: Comparison of prediction gains R_p and their standard deviations (in bracket) for the various classes of signals averaged over 200 independent initialisations of ESN

R_p [dB]	Circular AR(4) (2.24)	Noncircular ARMA (2.25)	Ikeda map (2.27)
Dual univariate ESN (LMS)	4.1928 (0.5849)	2.8064 (0.3211)	-0.0271 (0.1846)
Standard ESN (CLMS)	4.6701 (0.6112)	3.5341 (0.3541)	2.1134 (0.3619)
AESN (Augmented CLMS)	4.5489 (0.4792)	4.0396 (0.3624)	3.1967 (0.4258)
Standard ESN (CNGD)	4.6947 (0.5957)	3.6744 (0.3438)	2.1558 (0.3005)
AESN (Augmented CNGD)	4.5764 (0.4665)	4.1511 (0.3380)	3.2666 (0.5435)
Standard ESN (AACNGD)	6.6080 (0.5808)	5.0524 (0.5013)	2.4679 (0.5673)
AESN (Augmented AACNGD)	6.5357 (0.4970)	5.2484 (0.6149)	3.5912 (0.3053)
R_p [dB]	Wind (<i>low</i>)	Wind (<i>medium</i>)	Wind (<i>high</i>)
Dual univariate ESN (LMS)	2.3519 (0.5882)	4.0615 (0.9153)	8.1162 (0.9124)
Standard ESN (CLMS)	2.4635 (0.4516)	4.7938 (0.7519)	9.5917 (0.9746)
AESN (Augmented CLMS)	2.6417 (0.3620)	5.3029 (0.7781)	10.2063 (0.9537)
Standard ESN (CNGD)	2.6653 (0.4738)	5.0529 (0.7853)	9.9982 (1.0130)
AESN (Augmented CNGD)	2.8095 (0.3681)	5.7438 (0.8054)	10.7125 (0.9860)
Standard ESN (AACNGD)	4.0850 (0.3472)	6.3256 (0.7399)	11.7789 (0.9065)
AESN (Augmented AACNGD)	4.2571 (0.3499)	6.8286 (0.7968)	12.1900 (1.0272)

Table 6.2: Comparison of prediction gains R_p of previous discussed algorithms in Chapters 3 and 4 for the considered signals (200 independent trials on synthetic data and single trial on real world wind data).

R_p [dB]	Circular AR(4) (2.24)	Noncircular ARMA (2.25)	Ikeda map (2.27)
CLMS	3.8221	2.3428	1.5876
ACLMS	3.8181	3.1105	1.9852
NCLMS	4.1539	3.2107	2.0502
NACLMS	4.1326	3.4906	2.1667
R_p [dB]	Wind (<i>low</i>)	Wind (<i>medium</i>)	Wind (<i>high</i>)
CLMS	1.0169	2.9157	8.1790
ACLMS	1.4972	3.3788	8.7059
NCLMS	1.7728	4.9759	9.4764
NACLMS	2.3031	5.2714	9.7036

additional degree of freedom, allowing the nonlinear function to operate in a quasi-linear range if so required by the data nature.

6.4 Simulations

To verify the potential of the proposed augmented ESNs compared with standard complex ESNs, simulations were performed on both benchmark synthetic proper and improper signals, and for noncircular real-world wind data, given in Section .

For all the signals, experiments were undertaken by averaging 200 independent simulation trials in the adaptive prediction setting. The nonlinearity at the nonlinear output

layer of the ESNs was chosen to be the complex tanh function

$$\Phi(x) = \frac{e^{\beta x} - e^{-\beta x}}{e^{\beta x} + e^{-\beta x}} \quad (6.35)$$

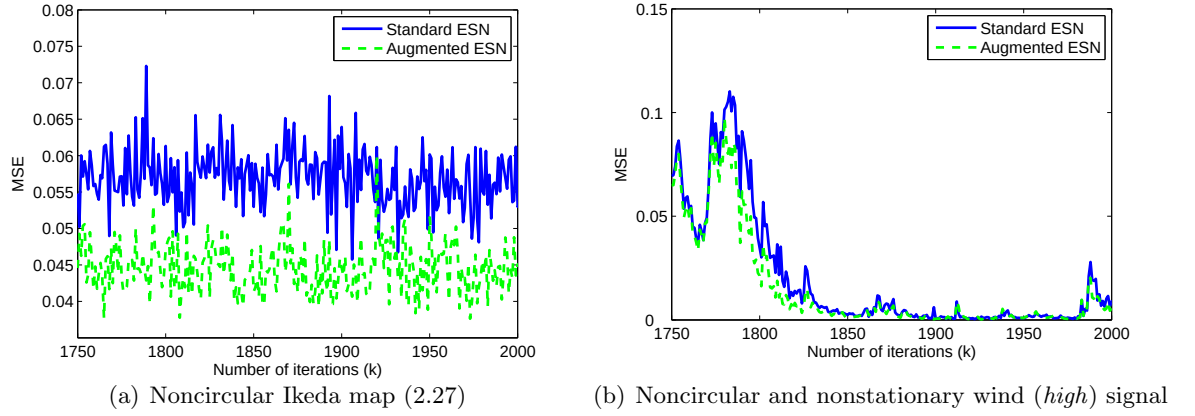
with slope $\beta = 1$. Ten neurons were used in the hidden layer, with the internal connection weights having 5% degree of connectivity. The input tap length was $K = 1$, with no bias input. The values of the randomly selected input, internal and feedback weights $\mathbf{W}_{ip}$, $\mathbf{W}_{in}$, and $\mathbf{w}_b$ were taken from a uniform distribution in the range $[-1, +1]$, and the spectral radius $\rho(\mathbf{W}_{in})$ was set to be 0.8. The learning rate was $\mu = 0.005$ for all the learning algorithms considered, with the initial amplitude for the AACNGD algorithm $\lambda(0) = 1$, and the step size of the adaptive amplitude update $\eta = 0.2$.

The standard prediction gain was employed to assess the performance, given in (4.18). Table 6.1 compares averaged prediction gains R_P [dB] and their standard deviations over 200 independent trials for the standard and augmented ESNs trained by CLMS, CNGD and AACNGD algorithms, as well as a dual univariate ESN trained by the LMS algorithm⁷, for the complex-valued signals considered. As expected, the dual univariate ESN, treating the real and imaginary parts of complex-valued data as two independent channels, had the worst performance. For the circular AR(4) signal (2.24), the performance of the augmented complex ESN was similar to that of standard ESN. For the noncircular signals, there was a significant improvement in the prediction gain when the augmented ESN was employed. As desired, the advantage of the nonlinear output layer over the linear output mapping was more pronounced in the prediction of the nonlinear synthetic signal and nonlinear and nonstationary real world wind signals. In practice, due to the randomly generated internal reservoir within an ESN, the augmented ESN cannot guarantee enhanced performance over its standard version in every trial, however, as illustrated in Table 6.3, on the average, in more than 90% of the trials the widely linear algorithms outperformed the corresponding standard ones. The advantage of using nonlinearity within reservoir computing over linear FIR filters was observed by comparing Table 6.1 and Table 6.2.

⁷The dual univariate approach deals with complex-valued data by splitting the input signals into its real and imaginary parts and treating them as independent real-valued quantities [41]. In a dual univariate ESN, two independent reservoirs are generated to model the two components of the input.

Table 6.3: The percentage of enhanced performance of augmented ESN algorithms for complex noncircular signals

	Noncircular ARMA	Ikeda map	Wind (<i>low</i>)	Wind (<i>medium</i>)	Wind (<i>high</i>)
Augmented CLMS	95.5%	98.5%	91.5%	94%	91%
Augmented CNGD	95%	98%	90%	95%	94.5%
Augmented AACNGD	93.5%	99%	89.5%	92.5%	92%

**Figure 6.3: Mean Square Errors (MSE) of one-step ahead prediction for augmented and standard ESN trained by the AACNGD algorithm.**

To further illustrate the advantage of using augmented complex statistics within complex valued ESNs, we compared the mean square errors (MSEs) of both the augmented and standard ESNs with adaptive amplitude of nonlinearity for the prediction of the complex-valued synthetic nonlinear and noncircular Ikeda map (2.27), and the noncircular wind (*high*) signal. Fig. 6.3 shows that in both cases, the augmented ESN with a nonlinear readout neuron trained by the augmented AACNGD outperformed its standard version.

We next investigated the influences of two parameters related to the generation of the internal layer, the degree of connectivity and the spectral radius $\rho(\mathbf{W}_{in})$, on the performance of standard and augmented ESNs. Fig. 6.4 and Fig. 6.5 show that in all the cases, for the prediction of real world wind (*medium*) and (*high*) signals, the augmented ESN trained by the augmented AACNGD algorithm achieved the best performance, and that for both learning strategies, it is desirable to keep a low degree of connectivity within the reservoir. This conforms with the ESN theory [71] that a small degree of connectivity can perform a relative decoupling of subnetworks with rich reservoir dynamics.

The size of the dynamical reservoir is another important parameter that influences

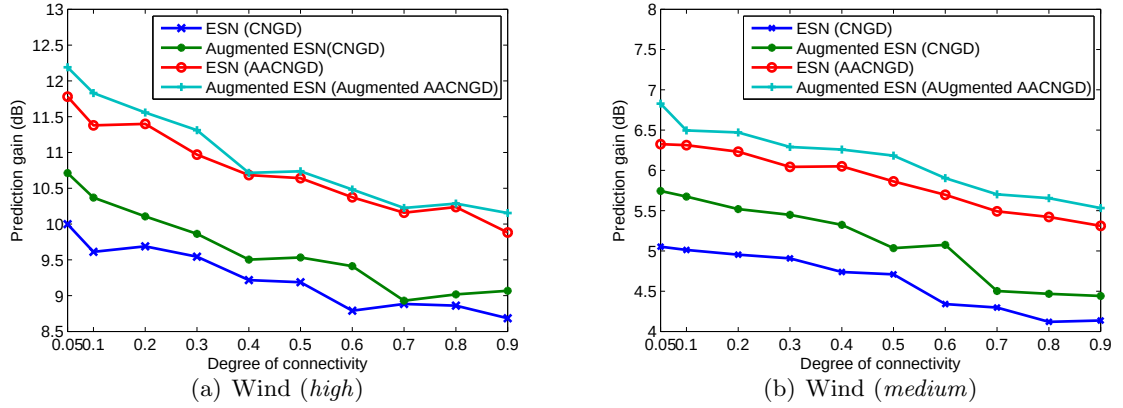


Figure 6.4: Comparison of performances of standard and augmented ESNs trained by different algorithms, over a range of degrees of connectivity on one-step ahead prediction of the wind (*medium*) and (*high*) signals.

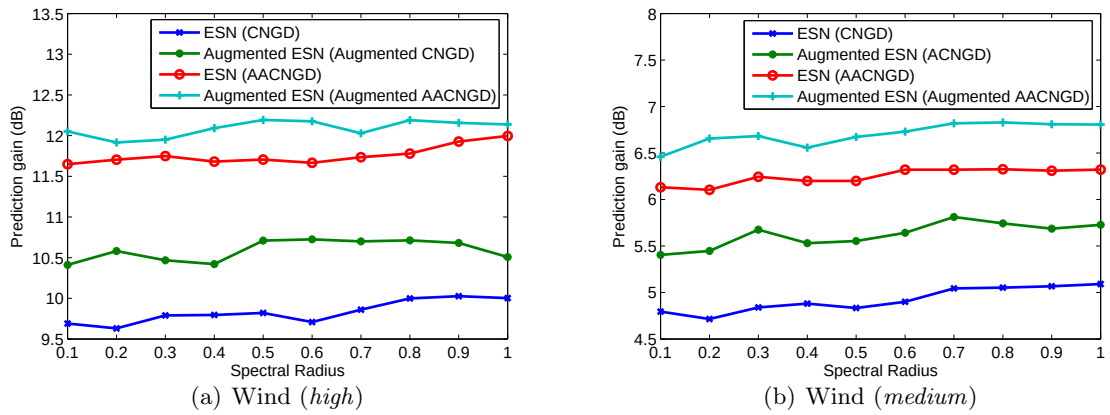


Figure 6.5: Comparison of performances of standard and augmented ESNs trained by different algorithms, over a range of spectral radii on one-step ahead prediction of the noncircular wind (*medium*) and (*high*) signals.

the performance of ESNs, as it reflects their universal approximation ability. An ESN with a larger reservoir size can learn the signal dynamics with a higher accuracy [84], as shown in Fig. 6.6(a) on one-step ahead prediction of the noncircular ARMA process in (2.25). This, however, applies to stationary signals, whereas for fast changing nonstationary processes, the larger reservoir caused saturation of internal neurons, resulting in performance degradation, as shown in Fig. 6.6(b) for the prediction of the nonstationary wind (*medium*) signal. Observe that in all the cases, the augmented ESNs outperformed their standard counterparts.

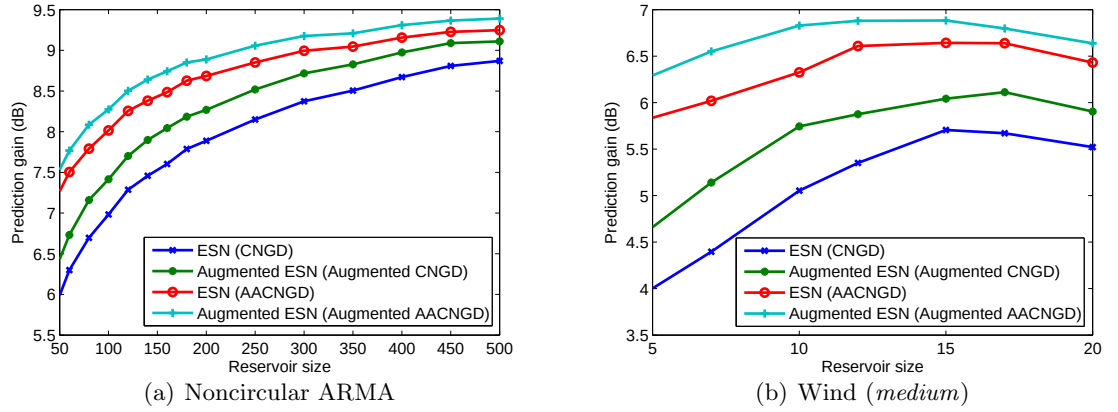


Figure 6.6: Comparison of performances of standard and augmented ESNs trained by different algorithms, over a range of reservoir size on one-step ahead prediction of the noncircular ARMA and wind (*medium*) signals.

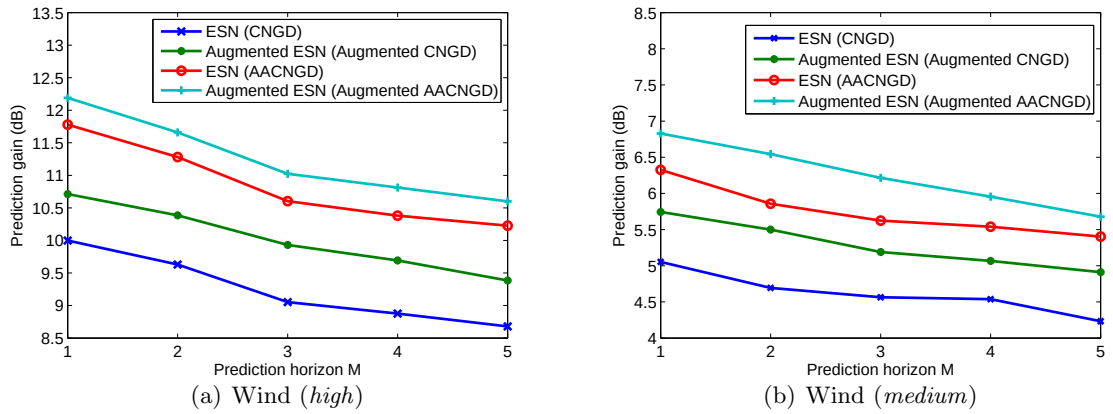


Figure 6.7: Comparison of performances of standard and augmented ESNs trained by different algorithms on multiple-step ahead prediction of the wind (*medium*) and (*high*) signals.

In the final set of simulations, we considered multi-step ahead prediction of the non-circular and nonstationary wind (*medium*) and (*high*) data. Fig. 6.7 shows the prediction gains of ESNs for a prediction horizon $M = 1, 2, 3, 4$ and 5 - in all the cases, the augmented ESN with adaptive amplitude of nonlinearity achieved the best performance.

6.5 Conclusions

An augmented complex echo state network (ESN) has been introduced for nonlinear adaptive filtering of the generality of complex-valued signals. The proposed ESN has been derived based on augmented complex statistics, thus making it suitable for both second order circular and noncircular signals. For generality, a nonlinear output layer has been introduced, and to deal with signals with large dynamics, an adaptive amplitude has been introduced into the output layer of the augmented ESN. The proposed augmented ESNs have been shown to exhibit theoretical and practical advantages over their conventional counterparts. This has been verified through comprehensive simulations on both synthetic noncircular data and real world wind measurements, and over a range of parameters.

Chapter 7

An Adaptive Diffusion Augmented CLMS Algorithm for Distributed Filtering of Noncircular Complex Signals

7.1 Introduction

In many typical applications of sensor networks, the main objective is to estimate a specific environmental parameter or phenomenon of interest. The traditional way to achieve this aim is to utilise a centralised network topology, where all the observations of a certain parameter of interest are firstly collected by surrounding sensors, and then transmitted to a central terminal, where the required estimation tasks are subsequently performed and the feedback is broadcast to the individual sensors surrounding it. Therefore, in this centralised mode of operation, a powerful central terminal for large data processing is highly demanded, together with extensive amount of communication between the nodes, which places a significant strain on communication and energy resources and limits the practical application of sensor networks [85].

In order to alleviate the computational demand in the traditional centralised solution, a distributed processing concept is proposed based on the localisation of data exchange and

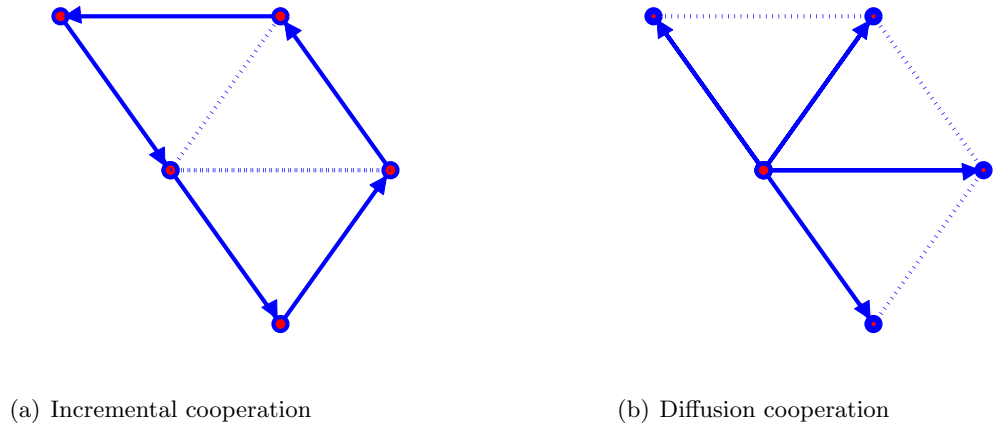


Figure 7.1: Two modes of cooperation strategy for a distributed network: (a) Incremental strategy, and (b) Diffusion strategy.

interactions within neighbouring sensor nodes. In this way, this distributed solution can still achieve a reliable estimation accuracy as compared with the centralized solution with tolerable computational complexity [85,86]. The idea behind the distributed processing is to deal with the extraction of information from data collected at nodes that are distributed over a geographic area. Within a distributed network, individual nodes collect observations related to a certain parameter of interest, and then interact with their neighbors in a certain manner, which is dictated by the network topology, in order to arrive at an estimate of the parameter of interest. One obvious advantage of distributed processing over its centralised counterpart lies in the fact that the individual nodes share the computational burden, and power and bandwidth usage are also consequently reduced [87]. In addition, the distributed processing is more robust against link failure, which is likely encountered by the data fusion nature of the centralised networks. Consequently, such distributed processing solution has found a wide range of real-world applications, such as in agriculture and environmental monitoring, transportation and factory instrumentation [85,86,88].

The distributed processing concept has been introduced into the context of adaptive filtering by Sayed *et al.* in 2007 [89], and distributed adaptive algorithms have been proposed for this purpose. According to the manner by which the individual nodes interact with neighbouring nodes for data processing, these distributed algorithms can be mainly classified into two groups, that is, either incremental or diffusion. As illustrated in Fig.

7.1, under the incremental strategy, information flows in a predefined sequential manner from one node to its adjacent one. One can imagine that due to the cyclic pattern of cooperation among the nodes, this mode of operation requires least amount of communication and power. In comparison, individual nodes under a diffusion strategy interact with all of their neighbours as dictated by the network topology. Although a higher amount of communication is demanded in this diffusion mode of operation as compared with its incremental counterpart, the nodes can get access to more information regarding the observations from their neighbors, and hence provide a better estimation performance and more robustness against the link failure. Several adaptive filtering algorithms have been extended into the operation of distributed estimation, including diffusion LMS [90], diffusion RLS [91], their incremental counterparts [89, 92], and diffusion Kalman filtering and smoothing algorithms [93].

Motivated by the widely linear modelling and augmented complex statistics introduced in Chapter 2, the Diffusion ACLMS (D-ACLMS) algorithm is introduced in this chapter in order to provide cooperative adaptive estimation of noncircular complex-valued signals, where the distributed nodes cooperate for the estimation of the same global task. The advantage of the widely linear D-ACLMS over the strictly linear D-CLMS is illustrated over simulations on both benchmark and real-world noncircular wind signals.

7.2 Derivation of D-ACLMS algorithm over cooperative networks

Consider a network with N distributed nodes, where each node has access to the local information $\{d_i(k), \mathbf{x}_i(k)\}$ at time instant k , where $d_i(k)$ is the desired signal, and $\mathbf{x}_i(k)$ is a regressor input vector of length L , defined as $\mathbf{x}_i(k) = [x_i(k-1), \dots, x_i(k-L)]^T$, both associated with node i . As illustrated in Fig. 7.2, each node in the network has access only to their peer neighbours, and the diffusion protocols allow every node i to combine information collected from its neighbourhood $\mathcal{N}_i$, which is the set of all the nodes linking to node i , including itself. By linearly combining the weight estimates in the neighbourhood of node i , and following the derivation in [90], we can motivate replacing the local weights

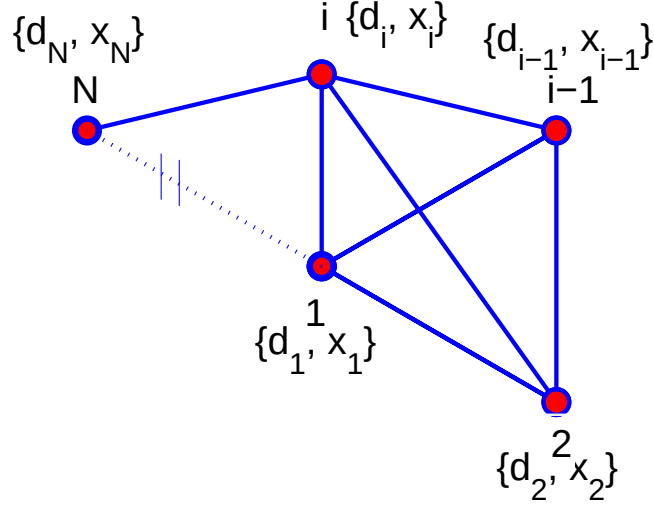


Figure 7.2: The topology of a distributed network with N nodes

for standard and conjugate parts, that is $\mathbf{h}_i(k)$ and $\mathbf{g}_i(k)$, with the weighted estimates

$$\underline{\mathbf{h}}_i(k) = \sum_{l \in \mathcal{N}_i} c(i, l) \mathbf{h}_l(k) \quad (7.1)$$

$$\underline{\mathbf{g}}_i(k) = \sum_{l \in \mathcal{N}_i} c(i, l) \mathbf{g}_l(k) \quad (7.2)$$

where the italic symbols with underscores denote the combined weights vectors $\underline{\mathbf{h}}_i(k)$ and $\underline{\mathbf{g}}_i(k)$ at node i for some combining coefficients $\{c_l \geq 0\}$ and satisfying $\sum_{l \in \mathcal{N}_i} c_l = 1$. This aggregate estimation at node i can be interpreted as a weighted linear combination which fuses information from nodes across the network into node i . The resulting aggregated weight vectors $\underline{\mathbf{h}}_i(k)$ and $\underline{\mathbf{g}}_i(k)$ at node i , can then be fed into its local adaptive filter in order to respond to local information and update the local weights $\mathbf{h}_i(k)$ and $\mathbf{g}_i(k)$. This way, the resulting cooperative adaptive network provides a peer-to-peer estimation framework that is robust to node and link failures and exploits network connectivity.

In the context of widely linear adaptive filtering, the output signal $y_i(k)$ at node i of a cooperative filter is given by

$$y_i(k) = \underbrace{\mathbf{x}_i^T(k) \underline{\mathbf{h}}_i(k)}_{\text{standard part}} + \underbrace{\mathbf{x}_i^H(k) \underline{\mathbf{g}}_i(k)}_{\text{conjugate part}} \quad (7.3)$$

whereas the corresponding output error at the node i is defined as $e_i(k) = d_i(k) - y_i(k)$. The local weight updates for the standard part and conjugate part at each node i can be derived as

$$\mathbf{h}_i(k+1) = \underline{\mathbf{h}}_i(k) + \mu_i e_i(k) \mathbf{x}_i^*(k) \quad (7.4)$$

$$\mathbf{g}_i(k+1) = \underline{\mathbf{g}}_i(k) + \mu_i e_i(k) \mathbf{x}_i(k) \quad (7.5)$$

7.3 Stability Analysis of D-ACLMS algorithm

To provide insight to the convergence of the D-ACLMS algorithm, we shall next investigate the role of the degree of cooperation and network topology on the system performance. To that end, rewrite the local and global quantities as

$$\begin{aligned} \mathbf{h}(k) &= \text{col}\{\mathbf{h}_1(k), \dots, \mathbf{h}_N(k)\}; \mathbf{g}(k) = \text{col}\{\mathbf{g}_1(k), \dots, \mathbf{g}_N(k)\} \\ \underline{\mathbf{h}}(k) &= \text{col}\{\underline{\mathbf{h}}_1(k), \dots, \underline{\mathbf{h}}_N(k)\}; \underline{\mathbf{g}}(k) = \text{col}\{\underline{\mathbf{g}}_1(k), \dots, \underline{\mathbf{g}}_N(k)\} \\ \mathbf{X}(k) &= \text{diag}\{\mathbf{x}_1(k), \dots, \mathbf{x}_N(k)\}; \mathbf{d}(k) = \text{col}\{d_1(k), \dots, d_N(k)\} \end{aligned}$$

in terms of the quantities that appear in (7.1), (7.2), (7.4) and (7.5), and denote by

$$\mathbf{D} = \text{diag}\{\mu_1 \mathbf{I}_L, \mu_2 \mathbf{I}_L, \dots, \mu_N \mathbf{I}_L\} \quad (7.6)$$

a diagonal matrix containing the local step-sizes.

To evaluate performance, without loss in generality consider an improper teaching signal $d_i(k)$, given by [41]

$$d_i(k) = \mathbf{x}_i^T(k) \mathbf{h}_o + \mathbf{x}_i^H(k) \mathbf{g}_o + n_i(k) \quad (7.7)$$

where $\mathbf{h}_o$ and $\mathbf{g}_o$ are respectively the unknown optimal local standard and conjugate weight vectors, and $n_i(k)$ denotes samples of doubly white noise with zero mean variance σ_n^2 , which are statistically independent of the input sequence $\mathbf{x}_i(k)$. The D-ACLMS algorithm can

now be written in a compact form

$$\begin{aligned}\underline{\mathbf{h}}(k) &= \mathbf{G}\mathbf{h}(k), \quad \underline{\mathbf{g}}(k) = \mathbf{G}\mathbf{g}(k) \\ \mathbf{h}(k+1) &= \underline{\mathbf{h}}(k) + \mathbf{D}(\mathbf{d}(k) - \mathbf{X}^T(k)\underline{\mathbf{h}}(k) - \mathbf{X}^H(k)\underline{\mathbf{g}}(k))\mathbf{X}^*(k) \\ \mathbf{g}(k+1) &= \underline{\mathbf{g}}(k) + \mathbf{D}(\mathbf{d}(k) - \mathbf{X}^T(k)\underline{\mathbf{h}}(k) - \mathbf{X}^H(k)\underline{\mathbf{g}}(k))\mathbf{X}(k)\end{aligned}\quad (7.8)$$

where $\mathbf{G} = \mathbf{C} \otimes \mathbf{I}_L$ is the transition matrix and $\mathbf{C} = [c(i, l)]$ is the $N \times N$ network connection matrix representing the network topology: a nonzero entry $c(i, l)$ means that node i is connected to node l . Moreover, $\mathbf{C}\mathbf{q}_N = \mathbf{q}_N$, where $\mathbf{q}_N = \text{col}\{1, \dots, 1\}_N$ is an $N \times 1$ column vector.

The standard and conjugate weight error vectors for all the nodes are given by

$$\tilde{\mathbf{h}}(k) = \mathbf{h}_{(o)} - \mathbf{h}(k) \quad \text{and} \quad \tilde{\mathbf{g}}(k) = \mathbf{g}_{(o)} - \mathbf{g}(k) \quad (7.9)$$

where

$$\mathbf{h}_{(o)} = \mathbf{q}_N \otimes \mathbf{h}_o \quad \text{and} \quad \mathbf{g}_{(o)} = \mathbf{q}_N \otimes \mathbf{g}_o \quad (7.10)$$

and their evolution is governed by

$$\tilde{\mathbf{h}}(k+1) = \mathbf{G}\tilde{\mathbf{h}}(k) - \mathbf{D}\mathbf{X}^*(k)(\mathbf{X}^T(k)\mathbf{G}\tilde{\mathbf{h}}(k) + \mathbf{X}^H(k)\mathbf{G}\tilde{\mathbf{g}}(k) + \mathbf{n}(k)) \quad (7.11)$$

$$\tilde{\mathbf{g}}(k+1) = \mathbf{G}\tilde{\mathbf{g}}(k) - \mathbf{D}\mathbf{X}(k)(\mathbf{X}^T(k)\mathbf{G}\tilde{\mathbf{h}}(k) + \mathbf{X}^H(k)\mathbf{G}\tilde{\mathbf{g}}(k) + \mathbf{n}(k)) \quad (7.12)$$

Therefore, the evolution of the augmented weight error vector $\tilde{\mathbf{w}}(k+1) = [\tilde{\mathbf{h}}^T(k+1), \tilde{\mathbf{g}}^T(k+1)]^T$ is described by

$$\begin{bmatrix} \tilde{\mathbf{h}}(k+1) \\ \tilde{\mathbf{g}}(k+1) \end{bmatrix} = \begin{bmatrix} \mathbf{I} - \mathbf{D}\mathbf{X}^*(k)\mathbf{X}^T(k) - \mathbf{X}^*(k)\mathbf{X}^H(k) \\ -\mathbf{X}(k)\mathbf{X}^T(k) \quad \mathbf{I} - \mathbf{D}\mathbf{X}(k)\mathbf{X}^H(k) \end{bmatrix} \mathbf{G} \begin{bmatrix} \tilde{\mathbf{h}}(k) \\ \tilde{\mathbf{g}}(k) \end{bmatrix} - \mathbf{D} \begin{bmatrix} \mathbf{X}^*(k) \\ \mathbf{X}(k) \end{bmatrix} \mathbf{n}(k)$$

and can be written more compactly as

$$\tilde{\mathbf{w}}(k+1) = (\mathbf{I} - \mathbf{D}\mathbf{X}^{a*}(k)\mathbf{X}^{aT}(k))\mathbf{G}\tilde{\mathbf{w}}(k) - \mathbf{D}\mathbf{X}^{a*}(k)\mathbf{n}(k) \quad (7.13)$$

where $\mathbf{X}^a(k) = [\mathbf{X}^T(k), \mathbf{X}^H(k)]^T$.

Under the assumption that the noise is statistically independent of the regressor vectors and the independence among the regressor vectors over space and time, upon applying the statistical expectation operator on both sides of the above equation, the mean error vector evolves according to

$$E[\tilde{\mathbf{w}}(k+1)] = (\mathbf{I} - \mathbf{D}\mathbf{C}_{\mathbf{X}^a\mathbf{X}^a}^*)\mathbf{G}E[\tilde{\mathbf{w}}(k)] \quad (7.14)$$

where $\mathbf{C}_{\mathbf{X}^a\mathbf{X}^a} = \text{diag}\{\mathbf{C}_{\mathbf{x}_1^a\mathbf{x}_1^a}, \dots, \mathbf{C}_{\mathbf{x}_N^a\mathbf{x}_N^a}\}$ is block diagonal and $\mathbf{C}_{\mathbf{x}_i^a\mathbf{x}_i^a} = E[\mathbf{x}_i^a\mathbf{x}_i^{aH}]$, where $i = 1, \dots, N$, is the augmented covariance matrix as defined in (2.12). Hence, for the iteration (7.14) to converge, that is, $|E[\tilde{\mathbf{w}}(k+1)]| < |E[\tilde{\mathbf{w}}(k)]|$ as $k \rightarrow \infty$, for stability in the mean

$$|\lambda_{\max}(\mathbf{B}\mathbf{G})| < 1 \quad i = 1, \dots, N \quad (7.15)$$

where $\mathbf{B} = \mathbf{I} - \mathbf{D}\mathbf{C}_{\mathbf{X}^a\mathbf{X}^a}^*$, and $\lambda_{\max}$ is the largest eigenvalue.

7.3.1 Performance dependence on the degree of noncircularity

The strictly linear D-CLMS can be summarised as [94]

$$\begin{aligned} y_i(k) &= \mathbf{x}_i^T(k)\mathbf{h}_i(k) \\ e_i(k) &= d_i(k) - y_i(k) \\ \mathbf{h}_i(k) &= \sum_{l \in \mathcal{N}_i} c(i, l)\mathbf{h}_l(k) \\ \mathbf{h}_i(k) &= \mathbf{h}_i(k) + \mu_i e_i(k)\mathbf{x}_i^*(k) \end{aligned} \quad (7.16)$$

and the evolution of its weight error vector follows the same generic form as that in (7.14), with the standard covariance matrix replacing the augmented covariance matrix. Since it does not account for the pseudocovariance, it cannot explain the noncircularity (rotation dependent distribution) of input signals, and performs suboptimally for improper data. Locally, for every node in the network, the analysis in [41, 95] shows that the advantage of the widely linear ACLMS over standard CLMS is more pronounced with higher degrees of noncircularity. Since the diffusion topology facilitates mutual communication between nodes, and each distributed node performs local adaptive estimation on the same global

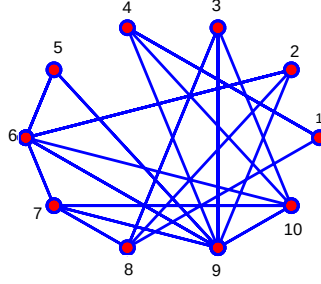


Figure 7.3: Network topology used in the simulations.

task, the performance advantage of D-ACLMS over D-CLMS is also more pronounced when processing improper signals, whereas for proper signals their performances are identical, as shown in the simulations.

7.4 Simulations

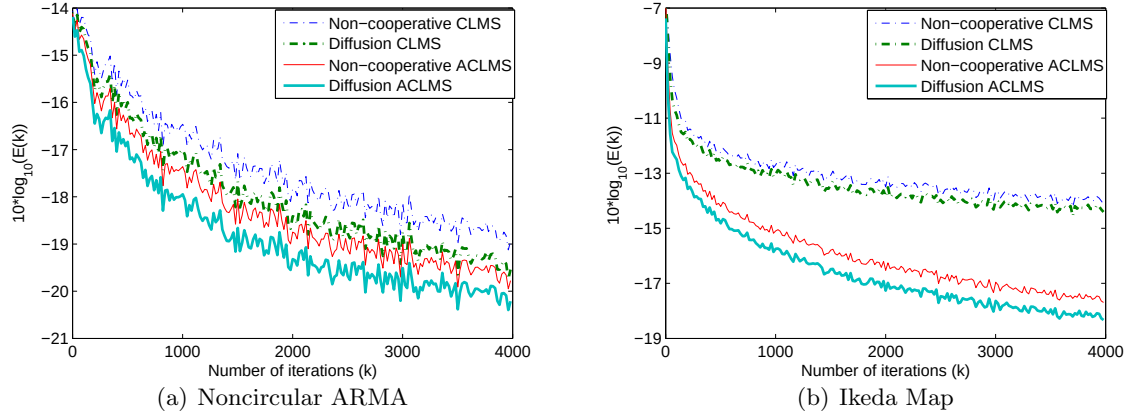
The proposed D-ACLMS was compared with the standard D-CLMS in a one step ahead prediction setting of both proper and improper signals. There are several methods to select the combination coefficients $c(i, l)$ and to maintain the stability of the cooperative networks. One such combiner the Metropolis rule, given by

$$c(i, l) = \begin{cases} 1/\max(n_i, n_l), & \text{if } l \neq i \text{ are connected} \\ 0, & \text{if } l \neq i \text{ are not connected} \\ 1 - \sum_{l \in \mathcal{N}_i/i} c(i, l), & \text{for } l = i \end{cases} \quad (7.17)$$

where n_i and n_l are respectively the numbers of nodes in the neighbourhood of nodes i and l . Other possible rules are the Laplacian and the nearest neighbour rules [94, 96]. In this work, a Metropolis network topology with $N = 10$ nodes was considered (see Fig. 7.3), with filter length $L = 4$ and the global step size $\mu = 0.01$. Independent doubly white Gaussian noise was added into each individual adaptive filter to give a 30dB signal-to-noise ratio (SNR). Learning curves comparing D-CLMS and D-ACLMS were produced by averaging 100 iterations of independent trials, performed on benchmark complex-valued

Table 7.1: Comparison of prediction gains R_p for the various classes of circular and noncircular signals

R_p [dB]	AR(4)	Noncircular ARMA	Ikeda map	Wind (<i>low</i>)	Wind (<i>medium</i>)	Wind (<i>high</i>)
CLMS	3.2152	3.0479	1.7229	4.9281	7.0285	13.1577
ACLMS	3.2068	3.8652	4.0542	5.1076	7.3248	13.3954
D-CLMS	3.7239	3.5672	2.1570	5.3175	7.6759	14.0499
D-ACLMS	3.7422	4.4098	4.8389	5.6024	8.1051	14.3064

**Figure 7.4: Comparison of MSEs of all algorithms on noncircular signals (a) Noncircular ARMA and (b) Ikeda map.**

stable circular and noncircular signals while single trial simulations were performed on real-world noncircular wind signals.

The standard prediction gain R_p [dB], given in (4.18), was employed to assess the performance, where the forward prediction error $e(k)$ which is the averaged error over all the nodes. Table 7.1 compares averaged prediction gains R_p over 100 independent trials for the D-CLMS and D-ACLMS trained cooperative networks. Observe that for the circular AR(4) signal, the performances of the standard algorithms (non-cooperative CLMS [24] and cooperative D-CLMS) were similar to those of their widely linear counterparts (ACLMS and D-ACLMS), whereas for the noncircular signals, there was an improvement in the prediction gain when the widely linear algorithms were employed. In all the cases, due to the cooperative mode of operation, the diffusion algorithms outperformed their single node counterparts.

Fig. 7.4 shows the evolution of mean-square errors (MSEs) for the noncircular ARMA process and Ikeda map for all the algorithms considered - observe that the diffusion al-

gorithms achieved both faster convergence and smaller MSEs than their non-cooperative counterparts, and that the widely linear algorithms outperformed the strictly linear ones. Note that in this work we assume that each individual adaptive filter receives the same source with different noise realisation within the cooperative network. However, the potential of cooperative networks has not been fully exploited due to the fact that real world wind observations was extracted from a single anemometer. In the future, a network of anemoeters will be established in a wind farm, and wind observations from multiple sensors will be transmitted into corresponding distributed adaptive filters within a cooperative network to perform adaptive wind prediction.

7.5 Conclusions

We have introduced a diffusion ACLMS (D-ACLMS) algorithm for distributed adaptive estimation of general complex-valued signals (both circular and noncircular) in a cooperative fashion. The advantage of the diffusion topology over non-cooperative structures in terms of convergence speed and smaller MSEs has been illustrated via analysis and simulations on both synthetic and real world noncircular complex signals. This work provides a stepping stone for a future class of widely linear diffusion estimation algorithms over cooperative networks.

Chapter 8

Widely Linear Adaptive Frequency Estimation of Unbalanced Three-Phase Power Systems

8.1 Introduction

In power systems, a precise detection of faulty or abnormal situations is necessary so that they can be eliminated and consequently return to the normal operation condition in time, and current power system protective measurements mainly monitor the three-phase voltage and current signals, as well as their frequency. Frequency is one of the main parameters to be detected in a power system, as during a faulty or undesired situation it will experience a significant alteration. In practice, frequency variations are considered as a consequence of the dynamic unbalance between the generation and the load. When the voltage frequency is lower than the nominal value, the generation-load system is indicated to be overloaded, while a frequency higher than the nominal value implies more generation than load within the system [97].

To this end, fast and accurate frequency estimation has attracted much attention. A variety of techniques and algorithms have been developed for this purpose, including the modified zero crossing technique [98, 99], phase-locked loop (PLL) [100–102], least-square error based adaptive filters [97, 103, 104], and recursive state estimation based nonlinear

observers and extended Kalman filters [105–107], among which, recursive state estimation based methods have been shown to achieve accurate frequency estimation in critical cases under unbalanced voltage conditions. However, most of these methods, being based on the measurement of a single phase of the system, are limited in terms of the characterisation of system frequency. In a three-phase system, if line-to-line voltages are also considered, no single-phase signal adequately characterizes system frequency, because six different single-phase voltage signals may exist [108]. Therefore, an optimal solution would be based on a framework which simultaneously considers all the three-phase voltages; this would enable a unified estimation of system frequency as a whole, and provide enhanced robustness.

To this end, Clarke’s $\alpha\beta$ transformation has been introduced with the aim to construct a complex-valued signal with the information provided by all the three-phase voltage in a simultaneous way [109], thus equipping the classical single phase methods with more robustness in characterising system frequency. Based on this transformation, a number of solutions have been developed in the complex domain $\mathbb{C}$ [110–115]. Among them, adaptive algorithms based on the minimisation of the mean square error are widely used owing to their simplicity, computational efficiency, and robust performance on frequency estimation in the presence of noise and harmonic distortions. However, in critical cases under unbalanced voltage conditions, such as when different amplitudes within the three-phase source voltage or a voltage sag in one or two phases are taking place, standard phase angle calculation techniques used in linear adaptive filter have proven suboptimal, resulting in an unavoidable oscillatory estimation error. This problem has been discussed in [116], where the complex-valued signal obtained from an unbalanced three-phase voltage source was represented as an orthogonal sum of the positive and negative sequences. Since the standard complex linear adaptive filter can only account for the positive sequences, the negative sequences introduce inevitable estimation error oscillating at twice the system frequency; recent attempts to estimate frequency under unbalanced conditions using the information extracted by all the three-phase voltage can be found in [117–119].

In this chapter, the issue of frequency estimation using adaptive filters under unbalanced conditions is addressed based on the widely linear modelling of the complex-valued signal, derived from three-phase voltages by the $\alpha\beta$ transformation. Using recent advances

in augmented complex-valued second order statistics, we illustrate that under unbalanced conditions, the complex-valued signal is second order noncircular (improper), for which the probability density function is not rotation invariant. For the modelling of noncircular signals, the standard linear estimation, which is based on the covariance matrix of a complex-valued random vector $\mathbf{x}$, that is $\mathbf{C}_{\mathbf{xx}} = E[\mathbf{xx}^H]$, is not adequate and the pseudocovariance matrix $\mathbf{P}_{\mathbf{xx}} = E[\mathbf{xx}^T]$ should also be taken into account to describe the complete second order behaviour [36], [38]. In practice, this is achieved by virtue of widely linear modelling [36, 41], where both $\mathbf{x}$ and its complex conjugate $\mathbf{x}^*$ are combined into the augmented input $\mathbf{x}_a = [\mathbf{x}^T, \mathbf{x}^H]^T$. Therefore, to deal with online frequency estimation of noncircular signals, we here propose to use a widely linear modelling based adaptive filtering method.

This chapter is organised as follows. In Section 8.2, the noncircular nature of the complex-valued signal generated from the $\alpha\beta$ transformation of unbalanced three-phase voltages is illustrated, and a robust frequency calculation method is derived based on widely linear modelling. In Section 8.3, the mathematical modelling of typical unbalanced power system conditions and voltage sags has been discussed. In Section 8.4, to illustrate the suitability of proposed method, the augmented CLMS (ACLMS) algorithm [47] is used and its superiority over the standard CLMS algorithm [112] is illustrated through analysis and via simulations on various unbalanced conditions, and also in presence of different levels of noise and higher order harmonics. Finally, Section 8.5 concludes this chapter.

8.2 Frequency Estimation Based on Widely Linear Adaptive Filtering

The voltages in power system in a noise-free environment can be represented in a discrete time form as

$$\begin{aligned} v_a(k) &= V_a(k)\cos(\omega k\Delta T + \phi) \\ v_b(k) &= V_b(k)\cos(\omega k\Delta T + \phi - \frac{2\pi}{3}) \\ v_c(k) &= V_c(k)\cos(\omega k\Delta T + \phi + \frac{2\pi}{3}) \end{aligned} \tag{8.1}$$

where $V_a(k), V_b(k), V_c(k)$ are the peak values of each fundamental voltage component at time instant k , ΔT the sampling interval, ϕ the phase of the fundamental component, and $\omega = 2\pi f$ the angular frequency of the voltage signal, with f being the system frequency. The time-dependent three-phase voltage is conveniently transformed by the orthogonal $\alpha\beta 0$ transformation matrix [109] into a zero-sequence v_0 and direct and quadrature-axis components, v_α and v_β , as Clarke's transform

$$\begin{bmatrix} v_0(k) \\ v_\alpha(k) \\ v_\beta(k) \end{bmatrix} = \sqrt{\frac{2}{3}} \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} v_a(k) \\ v_b(k) \\ v_c(k) \end{bmatrix} \quad (8.2)$$

The factor $\sqrt{2/3}$ is used to ensure that the system power is invariant under this transformation. When $V_a(k), V_b(k), V_c(k)$ are identical and time invariant, $v_0(k) = 0$, $v_\alpha(k) = A(k)\cos(\omega k\Delta T + \phi)$ and $v_\beta(k) = A(k)\cos(\omega k\Delta T + \phi + \frac{\pi}{2})$, with a constant amplitude $A(k) = A$, where $v_\alpha(k)$ and $v_\beta(k)$ are the orthogonal coordinates of a point whose position is time variant at a rate proportional to the system frequency. In practice, normally only the v_α and v_β parts are used in the modelling [110], (known as the $\alpha\beta$ transformation), whereas the zero sequence component v_0 is not necessary for analysis. The complex voltage signal $v(k)$ of the system, is therefore given by

$$\begin{aligned} v(k) &= v_\alpha(k) + jv_\beta(k) + n(k) \\ &= \hat{v}(k) + n(k) \end{aligned} \quad (8.3)$$

where $\hat{v}(k)$ is the desired voltage signal, $n(k)$ is the measurement noise and $v(k)$ is the noisy voltage observation. The desired voltage can be estimated iteratively as

$$\hat{v}(k+1) = Ae^{j(\omega(k+1)\Delta T + \phi)} = Ae^{j\omega\Delta T}e^{j(\omega k\Delta T + \phi)} = \hat{v}(k)e^{j\omega\Delta T} \quad (8.4)$$

where the instantaneous system frequency f is represented by the phasor $e^{j\omega\Delta T}$ ($f = \frac{\omega}{2\pi}$). Observe that in normal operating conditions (identical and time invariant $V_a(k), V_b(k), V_c(k)$), samples of $\hat{v}(k)$ are located on a circle in the complex plane with a

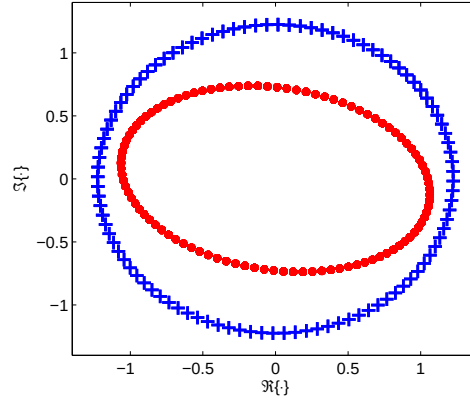


Figure 8.1: Geometric view of circularity and noncircularity via a “real-imaginary” scatter plot in the complex plane. The circle denoted by ‘+’ represents a circular complex-valued signal $\hat{v}(k)$ obtained from a balanced situation where $V_a(k), V_b(k), V_c(k)$ are identical at 1-p.u. (per-unit) and time invariant. The ellipse denoted by ‘o’ represents a noncircular complex-valued $\hat{v}(k)$ obtained in an unbalanced condition with $V_a(k) = 1\text{-p.u.}$, $V_b(k) = 0.7\text{-p.u.}$ and $V_c(k) = 0.5\text{-p.u.}$

constant radius A , depicted by ‘+’ in Fig. 8.1. For a constant sampling frequency, the probability density function of $v(k)$ is rotation invariant, since $\hat{v}$ and $\hat{v}e^{j\theta}$ have the same distribution for any real θ . This, in turn, means that $\hat{v}(k)$ is second order circular [28], and in this case the frequency estimation can be performed adequately by a standard linear adaptive filter, based on the strictly linear model in (2.3).

The complex least mean square (CLMS) algorithm [24], which is widely used in signal processing applications, has also been used for frequency estimation of three-phase voltages [112] and can be summarised as

$$\begin{aligned}\tilde{v}(k+1) &= v(k)w(k) \\ e(k) &= v(k+1) - \tilde{v}(k+1) \\ w(k+1) &= w(k) + \mu e(k)v^*(k)\end{aligned}\tag{8.5}$$

where $w(k)$ is the weight coefficient at time instant k , $\tilde{v}(k+1)$ is the estimate of $v(k+1)$, $e(k)$ the estimation error, and μ the step-size. Comparing (8.4) and the linear estimation model in (8.5), the system frequency can be estimated from

$$\hat{f}(k) = \frac{1}{2\pi\Delta T} \sin^{-1}(\Im(w(k)))\tag{8.6}$$

However, when the three-phase power system deviates from its normal condition, such as when the three channel voltages exhibit different levels of dips or transients, voltages $V_a(k), V_b(k), V_c(k)$ are not identical and can be time variant, and samples of $\hat{v}(k)$ are not allocated on a circle with a constant radius, as illustrated by the ellipse denoted by ‘o’ in Fig. 8.1, causing the iterative model of $\hat{v}(k)$ stated in (8.4) to break down. In this case, the distribution of $v(k)$ is rotation dependent (noncircular) and the signal is accurately expressed (see Appendix C for the derivation) only by using the widely linear model in (2.7), that is

$$\begin{aligned} v(k) &= Ae^{j(\omega k \Delta T + \phi)} + Be^{-j(\omega k \Delta T + \phi)} + n(k) \\ &= \hat{v}(k) + n(k) \end{aligned} \quad (8.7)$$

In other words, when $V_a(k), V_b(k), V_c(k)$ are not identical, $B \neq 0$, introducing a rotation variant probability density function, implying that in unbalanced conditions, $v(k)$ exhibits a certain degree of noncircularity. Since the widely linear model in (2.7) is the only second order optimal estimator for improper data, both $v(k)$ and its complex conjugate $v^*(k)$ should be considered in the frequency estimation in unbalanced cases, that is

$$\tilde{v}(k+1) = \underbrace{v(k)h(k)}_{\text{standard update}} + \underbrace{v^*(k)g(k)}_{\text{conjugate update}} \quad (8.8)$$

where $h(k)$ and $g(k)$ are respectively the filter weight coefficients corresponding to the standard and conjugate updates at time instant k , and the estimation error $e(k)$ and the cost function $J(k)$ can be defined as

$$e(k) = v(k+1) - \tilde{v}(k+1) \text{ and } J(k) = |e(k)|^2 = e(k)e^*(k) \quad (8.9)$$

The update of both the standard and the conjugate term weight coefficient can be obtained by using the steepest descent gradient as

$$h(k+1) = h(k) - \mu \nabla_h J(k) \quad (8.10)$$

$$g(k+1) = g(k) - \mu \nabla_g J(k) \quad (8.11)$$

Following the analysis in Chapter 3, the updates of the standard and conjugate weight coefficient $h(k)$ and $g(k)$ can be obtained as

$$h(k+1) = h(k) + \mu e(k) v^*(k) \quad (8.12)$$

$$g(k+1) = g(k) + \mu e(k) v(k) \quad (8.13)$$

Equations (8.12) and (8.13) describe the so called Augmented CLMS (ACLMS) algorithm [47], which is designed for training widely linear adaptive filters. To introduce the corresponding ACLMS based frequency estimation method for the three-phase unbalance system, by substituting (8.7) into (8.8) and neglecting the measurement noise, the estimate $\tilde{v}(k+1)$ can be expressed as

$$\begin{aligned} \tilde{v}(k+1) &= Ah(k)e^{j(\omega k \Delta T + \phi)} + Bh(k)e^{-j(\omega k \Delta T + \phi)} \\ &\quad + A^*g(k)e^{-j(\omega k \Delta T + \phi)} + B^*g(k)e^{j(\omega k \Delta T + \phi)} \\ &= (Ah(k) + B^*g(k))e^{j(\omega k \Delta T + \phi)} \\ &\quad + (A^*g(k) + Bh(k))e^{-j(\omega k \Delta T + \phi)} \end{aligned} \quad (8.14)$$

while from (8.7), the expression for $\hat{v}(k+1)$ can be re-written as

$$\hat{v}(k+1) = Ae^{j\omega \Delta T} e^{j(\omega k \Delta T + \phi)} + Be^{-j\omega \Delta T} e^{-j(\omega k \Delta T + \phi)} \quad (8.15)$$

Therefore, at the steady state, $\tilde{v}(k+1) \approx \hat{v}(k+1)$, the first term on the right hand side (RHS) of (8.15) can be estimated approximately by its counterpart in (8.14), hence, the term $e^{j\omega \Delta T}$ containing the frequency information can be estimated as

$$e^{j\omega \Delta T} = \frac{Ah(k) + B^*g(k)}{A} \quad (8.16)$$

Comparing the second term on the RHS of (8.14) and (8.15), the evolution of the term $e^{-j\omega \Delta T}$ can be expressed as

$$e^{-j\omega \Delta T} = \frac{A^*g(k) + Bh(k)}{B} \quad (8.17)$$

Upon taking the complex conjugate, we obtain

$$e^{j\hat{\omega}\Delta T} = \frac{Ag^*(k) + B^*h^*(k)}{B^*} \quad (8.18)$$

(8.16) and (8.18) can be simplified as

$$e^{j\hat{\omega}\Delta T} = h(k) + \frac{B^*}{A}g(k) \quad (8.19)$$

and

$$e^{j\hat{\omega}\Delta T} = h^*(k) + \frac{A}{B^*}g^*(k) \quad (8.20)$$

As shown in (C.2) in Appendix C, the coefficient A is real-valued whereas B is complex-valued, and thus $\frac{B^*}{A} = \left(\frac{B}{A}\right)^*$. Since (8.19) should be equal to (8.20), using $a(k) = \left(\frac{B}{A}\right)^*$, we can find the form of $a(k)$ by solving the following quadratic equation with complex-valued coefficients

$$g(k)a^2(k) + (h(k) - h^*(k))a(k) - g^*(k) = 0 \quad (8.21)$$

The discriminant of this quadratic equation is given by

$$\begin{aligned} \Delta &= \sqrt{(h(k) - h^*(k))^2 + 4|g(k)|^2} \\ &= 2\sqrt{-\Im^2(h(k)) + |g(k)|^2} \end{aligned} \quad (8.22)$$

where the operator $\Im(\cdot)$ represents the imaginary part of a complex-valued number. Since $a(k)$ is complex-valued, the discriminant is negative, and the two roots can be found as

$$\begin{aligned} a_1(k) &= \frac{-j\Im(h(k)) + j\sqrt{\Im^2(h(k)) - |g(k)|^2}}{g(k)} \\ a_2(k) &= \frac{-j\Im(h(k)) - j\sqrt{\Im^2(h(k)) - |g(k)|^2}}{g(k)} \end{aligned} \quad (8.23)$$

From (8.19), the phasor $e^{j\hat{\omega}\Delta T}$ is estimated either by using $h(k) + a_1(k)g(k)$ or $h(k) + a_2(k)g(k)$. Since the system frequency is far smaller than the sampling frequency, the imaginary part of $e^{j\hat{\omega}\Delta T}$ is positive, thus excluding the solution based on $a_2(k)$. The

system frequency $\hat{f}(k)$ is therefore estimated in the form

$$\hat{f}(k) = \frac{1}{2\pi\Delta T} \sin^{-1}(\Im(h(k) + a_1(k)g(k))) \quad (8.24)$$

The above equation is a generic widely linear extension of the standard linear frequency estimation method, and can be implemented by any type of widely linear adaptive filter [35, 120]. In addition, when the system is balanced, $g(k) = 0$, and the estimator in (8.24) simplifies into the standard linear CLMS based estimation.

8.2.1 The Suboptimality of CLMS for Frequency Estimation of Unbalanced Three-Phase Systems

This section gives a theoretical illustration of the suboptimality of strictly linear adaptive filters for frequency estimation of unbalanced three-phase voltage systems. In any unbalanced condition, expression (8.7) stands, and the estimator $\tilde{v}(k+1)$, obtained by using the CLMS algorithm, can be expressed as (neglecting the measurement noise)

$$\tilde{v}(k+1) = (Ae^{j(\omega k\Delta T + \phi)} + Be^{-j(\omega k\Delta T + \phi)})w(k)$$

In the steady state, $\tilde{v}(k+1) \approx \hat{v}(k+1)$, resulting in

$$w(k) = \frac{Ae^{j(\omega k\Delta T + \phi)}e^{j\omega\Delta T} + Be^{-j(\omega k\Delta T + \phi)}e^{-j\omega\Delta T}}{Ae^{j(\omega k\Delta T + \phi)} + Be^{-j(\omega k\Delta T + \phi)}}$$

We have

$$w(k) = e^{-j\omega\Delta T} + \frac{e^{j\omega\Delta T} - e^{-j\omega\Delta T}}{1 + \frac{B}{A}e^{-2j(\omega k\Delta T + \phi)}} \quad (8.25)$$

where $\frac{B}{A}$ is a unknown parameter, and $w(k)$ is periodic as $w(k) = w(k + \frac{1}{2f\Delta T})$. In (8.6), function $\sin^{-1}$ is a monotonic function, resulting in periodic oscillations in the estimated frequency $\hat{f}(k)$. The cycle frequency due to undermodelling is $2f$ when using standard linear adaptive filters for frequency estimation of unbalanced power systems, whereas for balanced power systems $B = 0$, and the standard linear estimate in (8.6) is adequate.

Table 8.1: Four types of sags due to signal phase-to-ground fault in equation form

Type A	Type B	Type C	Type D
$V_a = V$	$V_a = V$	$V_a = 1$	$V_a = V$
$V_b = (-\frac{1}{2} - \frac{j\sqrt{3}}{2})V$	$V_b = -\frac{1}{2} - \frac{j\sqrt{3}}{2}$	$V_b = -\frac{1}{2} - \frac{j\sqrt{3}}{2}V$	$V_b = -\frac{1}{2}V - \frac{j\sqrt{3}}{2}$
$V_c = (-\frac{1}{2} + \frac{j\sqrt{3}}{2})V$	$V_c = -\frac{1}{2} + \frac{j\sqrt{3}}{2}$	$V_c = -\frac{1}{2} + \frac{j\sqrt{3}}{2}V$	$V_c = -\frac{1}{2}V + \frac{j\sqrt{3}}{2}$

Table 8.2: Three types of sags due to two phase-to-ground faults in equation form

Type E	Type F	Type G
$V_a = 1$	$V_a = V$	$V_a = \frac{2}{3} + \frac{V}{3}$
$V_b = (-\frac{1}{2} - \frac{j\sqrt{3}}{2})V$	$V_b = -\frac{j\sqrt{3}}{3} - \frac{V}{2} - \frac{j\sqrt{3}V}{6}$	$V_b = -\frac{1}{3} - \frac{V}{6} - \frac{j\sqrt{3}V}{2}$
$V_c = (-\frac{1}{2} + \frac{j\sqrt{3}}{2})V$	$V_c = \frac{j\sqrt{3}}{3} - \frac{V}{2} + \frac{j\sqrt{3}V}{6}$	$V_c = -\frac{1}{3} - \frac{V}{6} + \frac{j\sqrt{3}V}{2}$

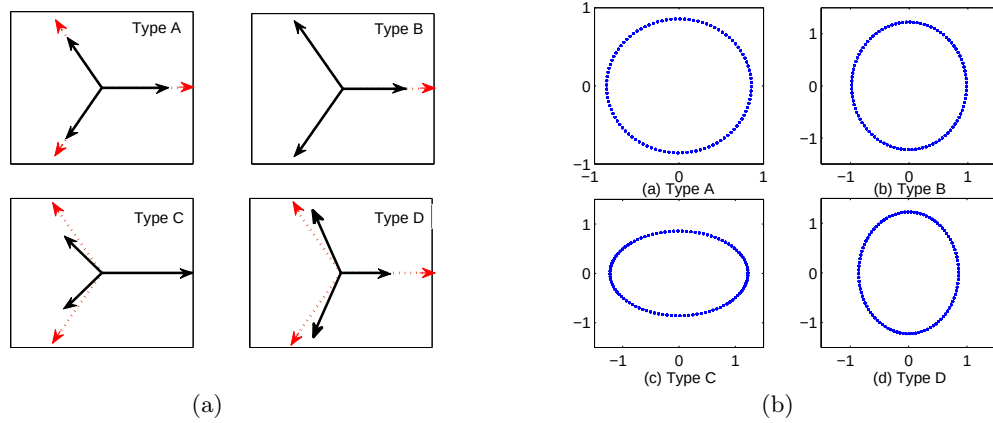


Figure 8.2: Three-phase unbalanced sags due to two-phase-to-ground faults with characteristic complex voltage $V = 0.7$. (a) Phasor diagram (dotted line denotes normal operation). (b) Circularity via a “real-imaginary” plot.

8.3 Unbalanced System Conditions and Classification of Voltage Sags

The severity of voltage sags is defined through its “characteristic complex voltage” V - the seven typical three-phase voltage sags and the cause of system unbalance are quantified in Table 8.1 and Table 8.2 according to the number of phase-to-ground faults, with the corresponding phasor diagrams given in Fig. 8.2(a) and Fig. 8.3(a).

To derive the closed form for sags due to single phase-to-ground fault, we first consider type B sag, which experiences a phase to-ground fault at phasor v_a when the power load is connected in a star structure. If the load is connected in a delta structure, where the equipment terminal voltages are the phase-to-phase voltages, the corresponding voltage

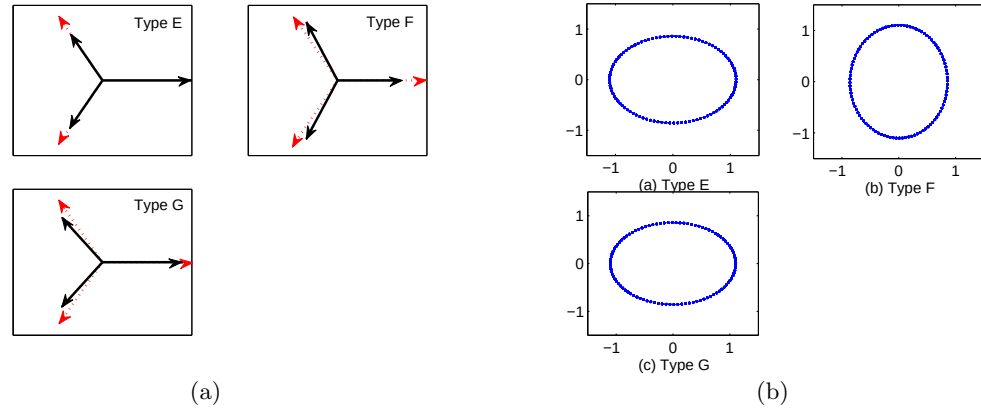


Figure 8.3: Three-phase unbalanced sags due to two phase-to-ground faults with characteristic complex voltage $V = 0.7$. (a) Phasor diagram (dotted line denotes normal operation). (b) Circularity via a “real-imaginary” plot.

Table 8.3: Causes of three-phase voltage sags and system unbalance

Fault type	Star-connected load	Delta-connected load
Three-phase	Type A	Type A
Two-phase-to-ground	Type E	Type F
Phase-to-phase	Type C	Type D
Single-phase	Type B	Type C

sag can be obtained by using the following star-delta transformation:

$$\begin{bmatrix} V'_a \\ V'_b \\ V'_c \end{bmatrix} = \frac{j}{\sqrt{3}} \begin{bmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} V_a \\ V_b \\ V_c \end{bmatrix} \quad (8.26)$$

where the factor $\sqrt{3}$ is used with to keep the system power invariant, while the 90° rotation by j aims to keep the axis of symmetry of the sag along the real axis. This way, a three-phase unbalanced voltage sag experienced by a delta-connected load due to a single-phase fault can be approximately classified as type C sag, usually considered to be caused by a phase-to-phase fault. In this case, the voltages in the two faulted phases move toward each other. After applying (8.26), the equivalent sag experienced in a phase-to-phase connected load can be classified as type D sag. The closed form of sags due to two phase-to-ground faults can be derived in the same way as single-phase and phase-to-phase faults. For example, we first assume that the non-faulted phasor is v_a , whereas both phasors v_b and v_c are affected by the fault, and this is type E sag as listed in Table 8.2. After

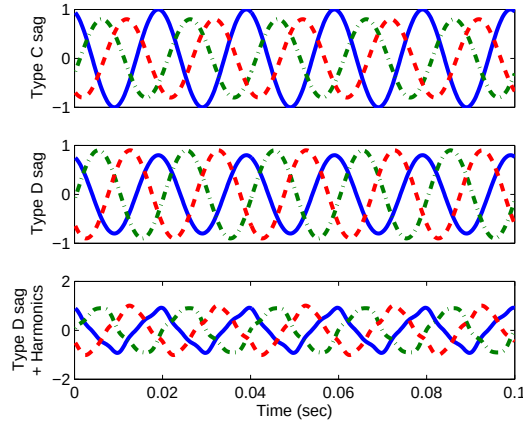


Figure 8.4: Time series of three-phase voltages for different sag types.

applying a star-delta transformation we obtain Type F sag, whereas applying twice star-delta transformation on type E sag gives type G sag. A full list of the relationship between fault, load connection and sag type is given in Table 8.3; for more detail see [121]. The geometric illustration of circularity via “real-imaginary” scatter plots in the complex plane is given in Fig. 8.2(b) and Fig. 8.3(b). Among these seven types of sags, type A sag is a symmetrical sag (second order circular with rotation independent distribution) whereas other types of sags are unsymmetrical and noncircular. Sags most commonly experienced by a three-phase system may be classified into types A, C, and D [122], however, since type A sag does not alter the circularity of the power system, the standard linear adaptive estimator is adequate, and in this work, we mainly focus on the latter two unsymmetrical voltage sags of type C and type D exhibiting noncircular amplitude distributions.

8.4 Simulations

The adaptive frequency estimator in (8.24) based on the widely linear ACLMS algorithm was applied to estimate the fundamental frequency variations from sampled values of voltage signals across several typical power system operating conditions, and was compared with the standard CLMS algorithm in (8.6). Simulations were performed in the Matlab programming environment with a sampling rate of 5kHz, and the step-size μ of both algorithms was set to be 0.01 in all simulations.

To quantify the degree of noncircularity in different unbalanced conditions the circu-

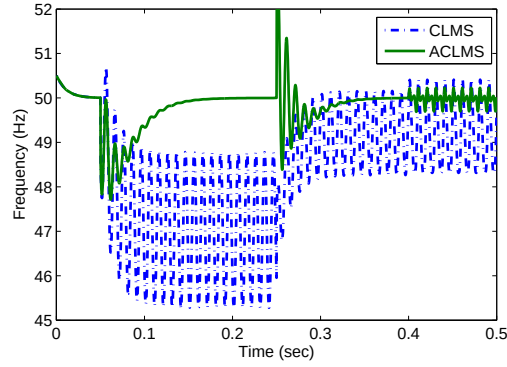


Figure 8.5: Frequency estimation using CLMS and ACLMS, for $\mu = 0.01$ when the three-phase power system undergoing three consecutive voltage sags of different natures.

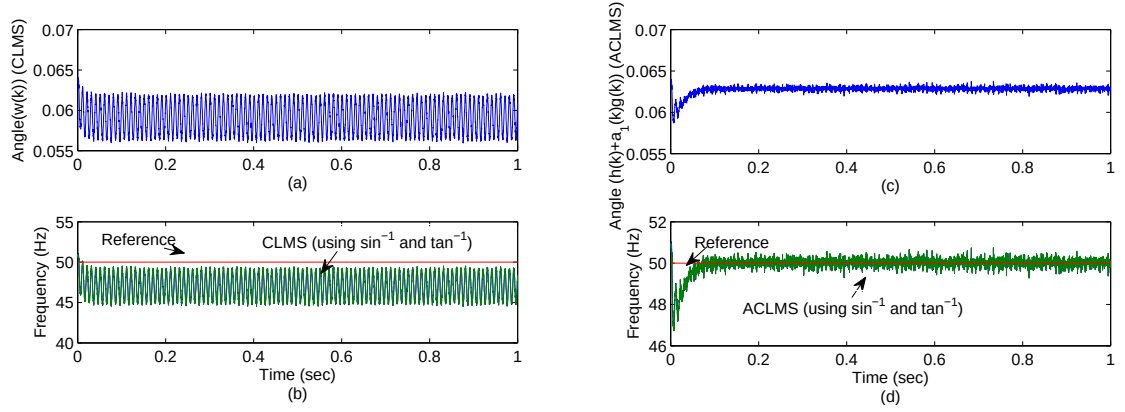


Figure 8.6: Frequency estimation for the unbalanced three-phase voltage with SNR = 40dB. (a) The angle of $w(k)$ used in CLMS method, (b) Frequency estimation using CLMS with asin and atan functions, (c) The angle of $h(k) + a_1(k)g(k)$, (d) Frequency estimation using ACLMS with asin and atan functions.

larity index r was used, given in (2.22). The values of the noncircularity index r lie in the interval $[0, 1]$, the value of 0 indicating that $v(k)$ is perfectly circular, otherwise indicating a second order noncircular (improper) $v(k)$.

The performance of frequency estimation of CLMS and ACLMS under both type C and D sag conditions (with characteristic complex voltage $V = 0.7$) is shown in Fig. 8.5, where the waveform of type C and D sags is shown in Fig. 8.4. Initially, the simulated power system was in its normal operation at 50 Hz with a balanced distortion-free three-phase voltages with unity magnitude. Both algorithms were initialised at 50.5Hz and converged to 50Hz in a very similar way. Then at $t = 0.05$ s, a type C sag occurred,

Table 8.4: Mean square error (in Hz^2) at the steady state of frequency estimation using CLMS and ACLMS with asin and atan functions. The results were obtained by averaging 100 independent trails.

SNR (dB)	50	45	40	35	30	25	20
CLMS ($\sin^{-1}$)	11.0174	11.2404	11.3346	11.3995	11.6185	13.3011	25.4697
CLMS ($\tan^{-1}$)	11.0089	11.0233	11.0788	11.0142	10.9041	11.6052	18.0988
ACLMS ($\sin^{-1}$)	0.0031	0.0062	0.0206	0.0835	0.1484	1.2030	2.5809
ACLMS ($\tan^{-1}$)	0.0031	0.0063	0.0255	0.0898	0.2936	1.6012	4.1548

with a 12% voltage drop and 9.5° phase angle offset on phases v_b and v_c , leading to an unbalanced three-phase power system with a degree of noncircularity $r = 0.3501$ (Fig. 8.4(b)). There was an inevitable biased oscillation error in CLMS based estimation due to the undermodelling of the system, whereas the advantage of the widely linear ACLMS based estimator in accurately estimating the frequency can be observed after convergence. At $t = 0.25$ s, a type D sag took place exhibiting a 30% voltage drop at phase v_a and 6.6% voltage drop at both phases v_b and v_c together with a 8° phase angle offset, whereby the degree of noncircularity r was 0.3433. Again ACLMS gave an unbiased performance, whereas the CLMS was not adequate. After $t = 0.4$ s, the unbalanced three-phase voltages were polluted with higher order harmonics; a 10% of the third harmonic and 5% of the fifth harmonic of the fundamental frequency were added into the unbalanced three-phase power system suffering from the same type D sag to give $r = 0.3910$. The estimated frequency was subject to an oscillatory steady state error; from $t = 0.4$ s, the ACLMS achieved better performance with a smaller oscillation error at the steady state as compared with CLMS.

In the literature, both the inverse trigonometric functions $\sin^{-1}$ and $\tan^{-1}$ have been used to extract the instantaneous frequency information from the weight of the adaptive filter [110,112]. When the system is noise free, the performances obtained by both functions are identical, however, when the noise is considered, it is expected that the two functions render different estimation results. Fig. 8.6(a) and (c) illustrate the angle estimated using in (8.6) and (8.24) for the unbalanced three phase system considered with $\text{SNR} = 40\text{dB}$, whereas Fig. 8.6(b) and (d) show the estimated frequency by CLMS and ACLMS using $\sin^{-1}$ and $\tan^{-1}$ functions respectively. In this high SNR region, the performances obtained by $\sin^{-1}$ and $\tan^{-1}$ functions show no much difference between them. However, from the results in Table 8.4, it is clear that for the unbiased ACLMS estimation method, the $\sin^{-1}$ function gives a better performance, which may result from the imaginary part is

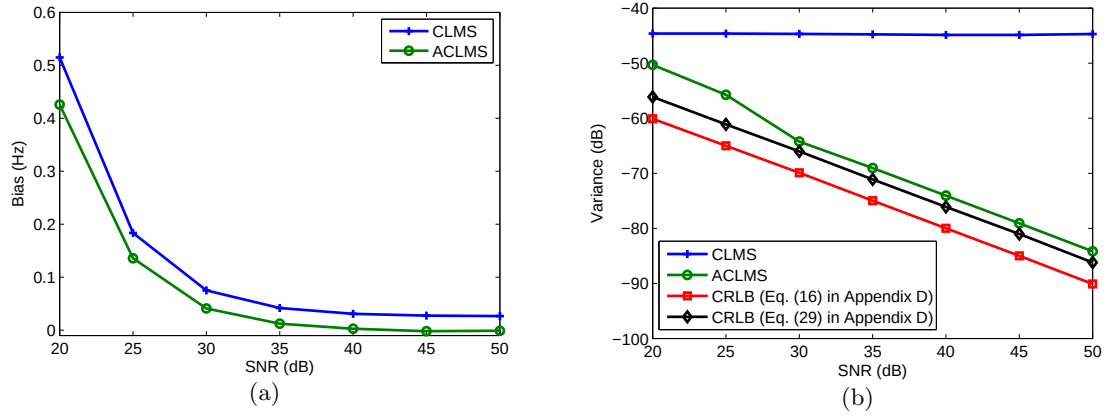


Figure 8.7: Comparison of the proposed ACLMS and CLMS at different SNR, obtained by averaging 1000 independent trials. (a) bias error. (b) variance.

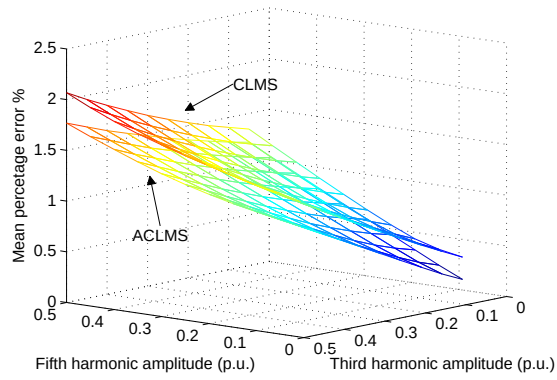


Figure 8.8: Frequency estimation for phase voltages contaminated with harmonics. The mean percentage error of CLMS and ACLMS algorithms over a range of amplitudes of both third harmonic and fifth harmonic at the steady-state.

only considered in $\sin^{-1}$ function, hence less contaminated by the noise as compared with the method using $\tan^{-1}$ function. However, the reason why $\tan^{-1}$ function gives a better performance in biased CLMS algorithm is still unclear, since the aim of this chapter is to proposed widely linear model based unbiased frequency estimation method for unbalanced power systems, the $\sin^{-1}$ function is still used for the following simulations.

To illustrate the statistical advantage of the ACLMS over CLMS based estimator, we next performed bias and variance analysis of both algorithms in a noisy environment when Type D sag happened. According to (8.7), and following the approach in [123], the Cramer-Rao Lower Bound (CRLB) which characterises the variance of a frequency

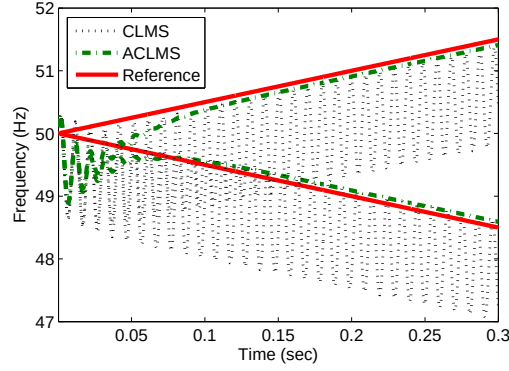


Figure 8.9: Frequency estimation using CLMS and ACLMS, for $\mu = 0.01$, when system frequency undergoing the decay and rise at a rate of 5 Hz/sec.

estimator on a single tone exponential contaminated by zero-mean complex-valued doubly white Gaussian noise¹, is derived in (D.16) in Appendix D under the assumption that the only unknown parameter is the system frequency f . The CRLB for all unknown parameters is further derived in (D.29) in Appendix D.

Fig. 8.7(a) and (b) illustrate the statistical bias and variance performance of CLMS and ACLMS on frequency estimation of the unbalanced system against different level of noise. The result was obtained by averaging 1000 independent trails. The unbiased property of ACLMS estimator can be observed in the high SNR region, whereas due to the submodelling, CLMS always performed a biased estimation. In Fig. 8.7(b), the ACLMS achieved a very small error variance approaching the CRLB (D.16) and (D.29) within 4 dB and 2dB respectively when SNR was between 30 to 50 dB. It is also interesting to note that the error variance of CLMS was nearly unaffected by the noise. This is due to the fact that compared with noise, the unavoidable estimation oscillation error experienced by CLMS (which can be observed from Fig. 8.5) constituted the major part of estimation variance.

The next set of simulations shows that when the input signal is contaminated with harmonics, the estimated frequency is subject to an oscillatory steady-state error. This advantage of ACLMS over CLMS was pronounced when the magnitudes of the third and fifth harmonics varied over a range of (0.0 - 0.5) p.u., as shown in Fig. 8.8.

In the following simulation, the performances of the proposed widely linear ACLMS

¹For white Gaussian noise, $n = n_r + jn_i$, $\sigma_{n_r}^2 = \sigma_{n_i}^2 = \frac{1}{2}\sigma_n^2$, and double whiteness implies the uncorrelated real and imaginary channels.

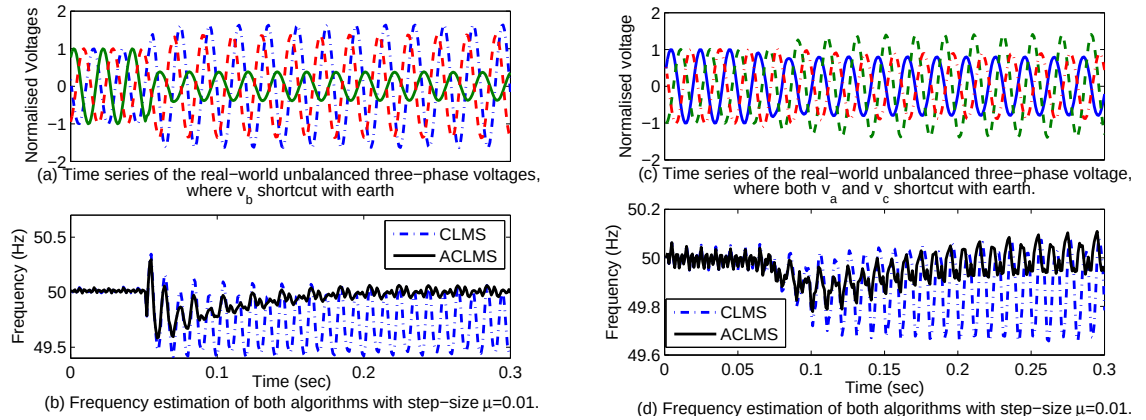


Figure 8.10: Frequency estimation for real-world unbalanced three-phase voltage.

and the strictly linear CLMS were compared for the case of frequency variation. In Fig. 8.9, the 50-Hz fundamental frequency of the unbalanced signal arose and decayed at a rate of 5 Hz/s; the estimated frequency obtained by the ACLMS algorithm followed the actual system frequency very closely with a delay of 0.05 s.

In the last set of simulations we considered a real-world problem, where unbalanced three phase voltages were recorded at a 110/20/10kV transformer station. The REL 531 numerical line distance protection terminal, produced by ABB Ltd, was installed in the station and was used to monitor changes in the three ‘phase-ground’ voltages. The device was set to record whenever the phase voltage value dropped below 90% of its nominal value for longer than 20 ms. The measured three ‘phase-ground’ voltages with system frequency 50 Hz were sampled at 1 kHz, and were normalised with respect to their normal peak voltage value². Two case studies are provided. In the first case, as shown in Fig. 8.10(a), at around $t = 0.05$ sec, a problem in phase v_b occurred (shortcut with earth), causing a 30% voltage drop, while the voltages in phase v_a and v_c simultaneously experienced respectively 60% and 37% voltage swells, to give a degree of noncircularity of $r = 0.1074$. In the second case study, as shown in Fig. 8.10(c), at around $t = 0.07$ sec, both phases v_a and v_c experienced shortcuts with earth, resulting in 20% and 11% voltage drops respectively and 41% voltage swell in phase v_b , to give a degree of noncircularity of $r = 0.0705$. The frequency tracking capabilities of the proposed ACLMS and standard CLMS methods are shown in Fig. 8.10(b) and (d). Both methods provided accurate

²The author would like to thank Z. Blazic of Elektroprenos, BiH, for providing real-world data, fruitful discussions, and expert advice.

responses under normal operating conditions, however, as expected, the CLMS failed to deal with unbalanced situations, whereas the fluctuations of the estimated frequency produced by ACLMS were much lower than those of the CLMS method.

8.5 Conclusions

We have shown that the complex-valued signal, obtained from the $\alpha\beta$ transformation of unbalanced three-phase voltages, is second order noncircular (improper), which renders strictly linear adaptive filtering based frequency estimation methods suboptimal. To deal with frequency estimation based on noncircular signals, a widely linear adaptive filter and a corresponding robust frequency estimation method have been introduced, achieving enhanced performance under various unbalanced conditions. The proposed method has also been shown to be less sensitive to higher order harmonics and noises as compared with standard linear adaptive filters. Simulations over a range of unbalanced system conditions support the approach.

Chapter 9

Conclusions and Future Work

9.1 Conclusions

In this thesis, a class of adaptive filtering algorithms suitable for the processing of the generality of complex-valued signals (both second order circular and noncircular) have been derived, and their performances have been analysed and tested in both synthetic cases and in real-world applications. This has been achieved based on a novel mathematical representation of general complex-valued signals, the so called widely linear modelling. A unified framework using the $\mathbb{CR}$ calculus has been introduced for the derivation of the proposed algorithms. The simplicity of the $\mathbb{CR}$ calculus has also been highlighted, due to the fact that $\mathbb{CR}$ calculus allows for the consideration of non-analytic functions in $\mathbb{C}$, such as the real-valued cost functions which are frequently encountered in complex domain signal processing.

This work has addressed both linear and nonlinear complex-valued adaptive algorithms and their usefulness has been shown through the analysis and simulations on benchmark complex-valued signals, as well as on real-world signals including complex wind vectors and voltage signals of three-phase power systems, which are made complex by convenience of representation.

The contributions in this thesis is summarised as following:

1. A class of augmented complex-valued linear and nonlinear adaptive algorithms have been derived based on the widely linear model, including regularised normalised ACLMS (RNACLMS) and augmented affine projection algorithm (APA) for linear

FIR filters, augmented echo state networks (AESNs) for reservoir computing, and diffusion ACLMS (D-ACLMS) algorithm for distributed adaptive filtering over cooperative networks. Unlike the standard complex algorithms based on the strictly linear model, the full second-order statistical model of the signal is used. It has been shown that the strictly linear algorithms can be regarded as a special case of their widely linear counterparts, which is only optimal for second order circular (proper) signals with rotation invariant distributions.

2. This thesis has addressed the theoretical mean square steady state performance analysis of widely linear adaptive algorithms on one step ahead prediction of complex-valued signals. Although the advantage of widely linear adaptive algorithms over their strictly linear counterparts have been illustrated by both theory and simulations in the literature, a quantitative analysis was still missing. To this end, a unified approach to analyse the steady-state performance of adaptive filters, called energy conservation principle, has been applied in the context of widely linear modelling to quantify the steady state performance of strictly linear and widely linear adaptive algorithms on complex-valued second order noncircular (improper) signals.
3. An unbiased frequency estimator for unbalanced three-phase power systems has been proposed based on widely linear modelling. It has been illustrated that when unbalanced conditions are encountered by three phase channels, the complex-valued voltage, obtained by the Clarke's $\alpha\beta$ transformation, is second order noncircular (improper), resulting in the suboptimality of traditional strictly linear modelled frequency estimators with unavoidable oscillatory estimation errors, while the proposed method is well matched to unbalanced system conditions and provides unbiased frequency estimation. It has also been shown to be less sensitive to higher order harmonics and noises as compared with strictly linear estimator.

9.2 Future work

The final consideration in this work is into future directions of research and open problems. The issues need to be addressed, including

1. Whilst mean square steady state analysis has been introduced for widely linear adap-

tive filters, a rigorous transient analysis to quantify their convergence is still an open issue. Recent research in [124] has shown that due to the double filter length required by the widely linear algorithms as compared with strictly linear counterparts, in the initialisation stage, the widely linear algorithms converge more slowly. A generalised energy conservation principle has been proposed in [10], which provides both quantitative transient and steady state analysis. However, this algorithm is derived under the implicit circularity assumption and will be derived in the context of widely linear modelling.

2. Although the widely linear modelling based frequency estimator has been introduced to deal with the frequency estimation on unbalanced three-phase voltage, the LMS type adaptive estimator suffers from slow convergence in real world applications and still renders large oscillatory estimation errors when the voltage is contaminated with higher order harmonics due to the simplicity of its implementation. To this end, computationally intelligent methods, such as the class of Kalman filters, and reservoir computing will be involved to design widely linear frequency estimator with more robustness and faster response.
3. Both the widely linear and strictly linear frequency estimators discussed in this thesis are using complex-valued adaptive filters to model the evolution of the complex-valued voltage $\hat{v}(k)$. Due to the fact that both the real and imaginary parts of $\hat{v}(k)$ contain the system frequency information, a potential solution with reduced complexity can be achieved by using a real-valued adaptive filter to deal with either part. In the future, performance comparison between this potential solution and proposed widely linear frequency estimator will be conducted.
4. In real world power generation-load networks, the centralised implementation of parameter monitors, although possibly optimal, is not robust to complex large-scale dynamical systems with their measurements distributed on a large geographical region. The large-scale power generation-load systems require extensive computations to implement the centralized procedure, and the span of the geographical region, over which the large-scale system is deployed or the physical phenomenon is observed, leads to a large communication burden, and subsequently introduces latency to the estimation mechanism. Therefore, to eliminate the difficulties encountered by

a centralised implementation, distributed estimation using cooperative networks is highly demanded. However, the potential of cooperative networks has not been fully exploited in this work, e.g. the individual node receives the same wind observations from a single anemometer without any geographic information. In the future, a network of anemometers will be established in a wind farm on a hilly terrain, and wind observations from multiple anemometers will be transmitted into corresponding distributed adaptive filters within a cooperative network to perform adaptive wind prediction. This concept of cooperative networks can also be introduced into large-scale transformer networks to estimation the fundamental frequency via interactions between distributed adaptive filters.

5. To meet the need of simultaneous processing of multichannel or multivariate data, extensions of signal processing techniques into hypercomplex domain, especially quaternion domain $\mathbb{H}$, has become an active research area and found wide applications in computer graphics [125], for the modelling of three-dimensional rotations, in biomedical applications [126], for artifacts extraction from multichannel EEG signals. In the area of adaptive filtering, the quaternion LMS (QLMS) has been introduced for four dimensional modelling and short-term forecasting of wind profile [127], where wind profile in the north-south, east-west, vertical direction, and air temperature information are simultaneously considered, and hence outperforms its counterparts in $\mathbb{R}^4$, where each dimension is processed separately without the consideration of cross-channel information. Subsequently, second order statistical analysis on improperness (noncircularity) has been addressed in $\mathbb{H}$ [128], and has led to a class of widely linear adaptive algorithms [129, 130], allowing the incorporation of the full second-order information. Therefore, it is natural to consider the extensions of the contributions of this work into $\mathbb{H}$, including the application of energy conservation principle into quaternion domain to facilitate statistical mean square analysis on both strictly linear and widely linear quaternion-valued adaptive filters, and the design of frequency estimator based on three dimensional modelling of the three-phase voltage by virtue of quaternion algebra. As illustrated in (8.2), the current frequency estimation methods established in $\mathbb{C}$ is based on the Clarke's transformation of three channel voltages, where a complex-valued voltage $v(k)$ is constructed as $v(k) = v_\alpha(k) + jv_\beta(k)$, and v_0 is excluded from analysis due to the limited dimensions

in the complex domain $\mathbb{C}$ and its zero-valued nature in balanced condition. However, under any unbalanced conditions, v_0 contains the system frequency information and cannot be neglected. It is therefore possible to construct a trivariate voltage signal using v_0 , v_α and v_β , and develop widely linear quaternion-valued adaptive frequency estimators for unbalanced three-phase power systems.

Bibliography

- [1] S. Haykin, *Adaptive Filter Theory*, New Jersey: Prentice Hall, 1996.
- [2] A. H. Sayed, *Fundamentals of Adaptive Filtering*, New York: Wiley, 2003.
- [3] D. P. Mandic and J. A. Chambers, *Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability*, John Wiley & Sons, 2001.
- [4] P. S. R. Diniz, *Adaptive Filtering: Algorithms and Practical Implementation*, Springer, 2008.
- [5] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, New Jersey: Prentice-Hall, 1985.
- [6] A. Benveniste, M. Metivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximation*, New York: Springer-Verlag, 1990.
- [7] V. J. Mathews and Z. Xie, “A stochastic gradient adaptive filter with gradient adaptive step size,” *IEEE Transactions on Signal Processing*, vol. 41, no. 6, pp. 2075–2087, 1993.
- [8] W. P. Ang and B. Farhang-Boroujeny, “A new class of gradient adaptive step-size LMS algorithms,” *IEEE Transactions on Signal Processing*, vol. 49, no. 4, pp. 805–810, 2001.
- [9] D. P. Mandic, “A generalized normalized gradient descent algorithm,” *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 115–118, 2004.
- [10] H. C. Shin, W. J. Song, and A. H. Sayed, “Mean-square performance of data-reusing adaptive algorithms,” *IEEE Signal Processing Letters*, vol. 12, no. 12, pp. 851–854, 2005.

- [11] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," *Transactions of the ICEIE*, vol. 67-A, no. 5, pp. 126–132, 1984.
- [12] Y. Choi, H. Shin, and W. Song, "Adaptive regularization matrix for affine projection algorithm," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 54, no. 12, pp. 1087–1091, 2007.
- [13] B. Widrow, J. McCool, M. G. Larimore, and C. R. Johnson, "Stationary and nonstationary learning characteristics of the LMS adaptive filter," *Proceedings of IEEE*, vol. 46, pp. 1151–1162, 1976.
- [14] N. J. Bershad, "Analysis of the normalized LMS algorithm with Gaussian inputs," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-34, pp. 793–806, 1986.
- [15] N. R. Yousef and A. H. Sayed, "A unified approach to the steady-state and tracking analyses of adaptive filters," *IEEE Transactions on Signal Processing*, vol. 49, no. 2, pp. 90–102, 2001.
- [16] M. Rupp and A. H. Sayed, "A time-domain feedback analysis of filtered error adaptive gradient algorithms," *IEEE Transactions on Signal Processing*, vol. 44, pp. 1428–1439, 1996.
- [17] T. Y. Al-Naffouri and A. H. Sayed, "Transient analysis of data normalized adaptive filters," *IEEE Transactions on Signal Processing*, vol. 51, no. 3, pp. 639–652, 2003.
- [18] H. C. Shin and A. H. Sayed, "Mean-square performance of a family of affine projection algorithms," *IEEE Transactions on Signal Processing*, vol. 52, no. 1, pp. 90–102, 2004.
- [19] J. T. Connor, R. D. Martin, and L. E. Atlas, "Recurrent neural networks and robust time series prediction," *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 240–254, 1994.
- [20] H. Jaeger and H. Haas, "Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication," *Science*, vol. 304, no. 5667, pp. 78–80, 2004.

- [21] L. Xu, G. D. Pearlson, and V. D. Calhoun, "Joint source based morphometry to identify sources of gray matter and white matter relative differences in schizophrenia versus healthy controls," *Proceedings of the International Society for Magnetic Resonance in Medicine, (ISMRM)*, pp. 2252–2252, 2008.
- [22] A. V. Oppenheim and J. S. Lim, "The importance of phase in signals," *Proceedings of the IEEE*, vol. 69, no. 5, pp. 529–550, 1981.
- [23] S. L. Goh, M. Chen, D. H. Popovic, K. Aihara, D. Obradovic, and D. P. Mandic, "Complex-valued forecasting of wind profile," *Renewable Energy*, vol. 31, no. 11, pp. 1733–1750, 2006.
- [24] B. Widrow, J. McCool, and M. Ball, "The complex LMS algorithm," *Proceedings of IEEE*, vol. 63, no. 4, pp. 719–720, 1975.
- [25] A. Tarighat and A. H. Sayed, "Least mean-phase adaptive filters with application to communications systems," *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 220–223, 2004.
- [26] T. Needham, *Visual Complex Analysis*, Oxford University Press, 1997.
- [27] S. L. Goh and D. P. Mandic, "A complex-valued RTRL algorithm for recurrent neural networks," *Neural Computation*, vol. 16, no. 12, pp. 2699–2713, 2004.
- [28] B. Picinbono, "On circularity," *IEEE Transactions on Signal Processing*, vol. 42, no. 12, pp. 3473–3482, 1994.
- [29] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Transactions on Information Theory*, vol. 39, no. 4, pp. 1293–1302, 1993.
- [30] M. Valkama, M. Renfors, and V. Koivunen, "Advanced methods for I/Q imbalance compensation in communication receivers," *IEEE Transactions on Signal Processing*, vol. 49, no. 10, pp. 2335–2344, 2001.
- [31] R. Schober, W. H. Gerstacker, and L. H. J. Lampe, "Data-aided and blind stochastic gradients for widely linear MMSE MAI suppression for DS-SSMA," *IEEE Transactions on Signal Processing*, vol. 52, no. 3, pp. 746–756, 2004.

- [32] S. Javidi, M. Pedzisz, S. L. Goh, and D. P. Mandic, "The augmented complex least mean square algorithm with application to adaptive prediction problems," *Proceedings of the 1st IARP Workshop on Cognitive Information Processing*, pp. 54–57, 2008.
- [33] A. Kuh and D. P. Mandic, "Applications of complex augmented kernels to wind profile prediction," *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, ICASSP*, pp. 3581–3584, 2009.
- [34] S. L. Goh and D. P. Mandic, "An augmented CRTRL for complex-valued recurrent neural networks," *Neural Networks*, vol. 20, no. 10, pp. 1061–1066, 2007.
- [35] S. L. Goh and D. P. Mandic, "An augmented extended Kalman filter algorithm for complex-valued recurrent neural networks," *Neural Computation*, vol. 19, no. 4, pp. 1039–1055, 2007.
- [36] B. Picinbono and P. Chevalier, "Widely linear estimation with complex data," *IEEE Transactions on Signal Processing*, vol. 43, no. 8, pp. 2030–2033, 1995.
- [37] D. P. Mandic, S. Still, and S. C. Douglas, "Duality between widely linear and dual channel adaptive filtering," *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, (ICASSP)*, pp. 1729–1732, 2009.
- [38] P. J. Schreier and L. L. Scharf, "Second-order analysis of improper complex random vectors and process," *IEEE Transactions on Signal Processing*, vol. 51, no. 3, pp. 714–725, 2003.
- [39] B. Picinbono, "Second-order complex random vectors and normal distributions," *IEEE Transactions on Signal Processing*, vol. 44, no. 10, pp. 2637–2640, 1996.
- [40] E. Ollila, "On the circularity of a complex random variable," *IEEE Signal Processing Letters*, vol. 15, pp. 841–844, 2008.
- [41] D. P. Mandic and S. L. Goh, *Complex-Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models*, John Wiley & Sons, 2009.
- [42] P. J. Schreier, "Bounds on the degree of impropriety of complex random vectors," *IEEE Signal Processing Letters*, vol. 15, pp. 190–193, 2008.

- [43] P. J. Schreier, L. L. Scharf, and A. Hanssen, "A generalized likelihood ratio test for impropriety of complex signals," *IEEE Signal Processing Letters*, vol. 13, no. 7, pp. 433–436, 2006.
- [44] M. Novey and T. Adali, "On extending the complex fast ICA algorithm to noncircular sources," *IEEE Transactions on Signal Processing*, vol. 56, no. 5, pp. 2148–2154, 2008.
- [45] J. Navarro-Moreno, "ARMA prediction of widely linear systems by using the innovations algorithm," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3061–3068, 2008.
- [46] K. Aihara, *Applied Chaos and Applicable Chaos*, Science-Sha, 1994.
- [47] D. P. Mandic, S. Javidi, S. L. Goh, A. Kuh, and K. Aihara, "Complex-valued prediction of wind profile using augmented complex statistics," *Renewable Energy*, vol. 34, no. 1, pp. 196–210, 2009.
- [48] Y. Huang and J. Benesty, *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, Springer, 2004.
- [49] S. Javidi, D. P. Mandic, and A. Cichocki, "Complex blind source extraction from noisy mixtures using second-order statistics," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 57, no. 7, pp. 1404–1416, 2010.
- [50] M. Rupp and A. H. Sayed, "A time-domain feedback analysis of filtered-error adaptive gradient algorithms," *IEEE Transactions on Signal Processing*, vol. 44, no. 6, pp. 1428–1439, 1996.
- [51] N. R. Yousef and A. H. Sayed, "Modeling head and hand orientation during motion using quaternions," *International Journal of Adaptive Control and Signal Processing*, vol. 17, pp. 237–258, 2003.
- [52] S. C. Douglas and D. P. Mandic, "Mean and mean-square analysis of the complex LMS algorithm for non-circular Gaussian signals," *Proceedings of IEEE International Workshop on Digital Signal Processing, DSP*, pp. 101–106, 2010.
- [53] S. Haykin and L. Li, "Nonlinear adaptive prediction of nonstationary signals," *IEEE Transactions on Signal Processing*, vol. 43, no. 2, pp. 526–535, 1995.

- [54] T. Hinamoto and S. Maekawa, “Extended theory of learning identification,” *Transactions of IEIJ*, vol. 95, no. 10, pp. 227–234, 1975.
- [55] K. Hwang and W. Song, “An affine projection adaptive filtering algorithm with selective regressors,” *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 54, no. 1, pp. 43–46, 2007.
- [56] M. Montazeri and P. Duhamel, “A set of algorithms linking NLMS and block RLS algorithms,” *IEEE Transactions on Signal Processing*, vol. 43, no. 2, pp. 444–453, 1995.
- [57] A. A. de Lima and R. C. de Lamare, “Adaptive detection using widely linear processing with data reusing for DS-CDMA systems,” *Proceedings of International Telecommunications Symposium (ITS)*, pp. 187–192, 2006.
- [58] Y. Xia, C. Cheong-Took, S. Javidi, and D. P. Mandic, “A widely linear affine projection algorithm,” *Proceedings of IEEE Workshop on Statistical Signal Processing, SSP*, pp. 373–376, 2009.
- [59] H. G. Ray, L. R. Vega, S. Tressens, and B. C. Frias, “Analysis of explicit regularization in affine projection algorithm: robustness and optimal choice,” *Proceedings of European Signal Processing Conference, EUSIPCO*, pp. 1809–1812, 2004.
- [60] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1985.
- [61] S. L. Gay and S. Tavathia, “The fast affine projection algorithm,” *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, (ICASSP)*, pp. 3023–3026, 1995.
- [62] S. Shimauchi and S. Makino, “Stereo projection echo canceller with true echo path estimation,” *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, (ICASSP)*, pp. 3059–3602, 1995.
- [63] G. Cybenko, “Approximation by superpositions of a sigmoidal function,” *Mathematics of Control, Signals and Systems*, vol. 2, no. 4, pp. 303–314, 1989.

- [64] K. S. Narendra and K. Parthasarathy, "Identification and control of dynamical systems using neural networks," *IEEE Transactions on Neural Networks*, vol. 1, no. 1, pp. 4–27, 1990.
- [65] C. H. Lee and C. C. Teng, "Identification and control of dynamic systems using recurrent fuzzy neural networks," *IEEE Transactions on Fuzzy Systems*, vol. 8, no. 4, pp. 349–366, 2000.
- [66] D. P. Mandic and J. A. Chambers, "Toward an optimal PRNN-based nonlinear predictor," *IEEE Transactions on Neural Networks*, vol. 10, no. 6, pp. 1435–1442, 1999.
- [67] S. A. Billings and C. F. Fung, "Recurrent radial basis function networks for adaptive noise cancellation," *Neural Networks*, vol. 8, no. 2, pp. 273–290, 1995.
- [68] T. Lin, M. Tsuji, and E. Yamada, "A wavelet approach to real time estimation of power system frequency," *Proceedings of the 40th Annual Conference of the Society of Instrument and Control Engineers Japan*, pp. 58–65, 2001.
- [69] J. L. Elman, "Finding structure in time," *Cognitive Science*, vol. 14, no. 2, pp. 179–211, 1990.
- [70] W. Liu, I. Park, J. C. Yeung, and Príncipe, "An information theoretic approach of designing sparse kernel adaptive filters," *IEEE Transactions on Neural Networks*, vol. 20, no. 12, pp. 1950–1961, 2009.
- [71] H. Jaeger, "The echo state approach to analyzing and training neural networks," GMD Report 148, German National Research Institute for Computer Science, 2002.
- [72] G. Kechriotis and E. S. Manolakos, "Frequency estimation of distorted power system signals using extended complex Kalman filter," *IEEE Transactions on Circuits and Systems II: Analog and Digital Processing*, vol. 41, no. 3, pp. 235–238, 1994.
- [73] T. Kim and T. Adali, "Approximation by fully complex multilayer perceptrons," *Neural Computation*, vol. 15, no. 7, pp. 1641–1666, 2003.
- [74] H. Leung and S. Haykin, "The complex backpropagation algorithm," *IEEE Transactions on Signal Processing*, vol. 3, no. 9, pp. 2101–2104, 1991.

- [75] T. Gautama, D. P. Mandic, and M. M. Van Hulle, "A non-parametric test for detecting the complex-valued nature of time series," *International Journal of Knowledge-based and Intelligent Engineering Systems*, vol. 8, pp. 99–106, 2004.
- [76] G. M. Georgiou and C. Koutsougeras, "Complex domain backpropagations," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 39, no. 5, pp. 330–334, 1992.
- [77] I. N. Aizenberg, N. N. Aizenberg, and J. P. L. Vandewalle, *Multi-Valued and Universal Binary Neurons: Theory, Learning and Applications*, Kluwer Academic Publishers, 2000.
- [78] S. Jankowski, A. Lozowski, and J. M. Zurada, "Complex-valued multistate neural associative memory," *IEEE Transactions on Neural Networks*, vol. 7, no. 6, pp. 1491–1496, 1996.
- [79] T. Nitta, "An extension of the back-propagation algorithm to complex numbers," *Neural Networks*, vol. 10, no. 8, pp. 1391–1415, 1997.
- [80] S. Seth, M. C. Ozturk, and J. C. Príncipe, "Signal processing with echo state networks in the complex domain," *Proceedings of IEEE International Workshop on Machine Learning for Signal Processing, MLSP*, pp. 408–412, 2007.
- [81] Y. Xia, D. P. Mandic, M. M. Van Hulle, and J. C. Príncipe, "A complex echo state network for nonlinear adaptive filtering," *Proceedings of IEEE International Workshop on Machine Learning for Signal Processing, MLSP*, pp. 404–408, 2008.
- [82] C. Cheong Took and D. P. Mandic, "Adaptive IIR filtering of noncircular complex signals," *IEEE Transactions on Signal Processing*, vol. 57, no. 10, pp. 4111–4118, 2009.
- [83] E. Trentin, "Networks with trainable amplitude of activation functions," *Neural Networks*, vol. 14, no. 4, pp. 471–493, 2001.
- [84] H. Jaeger, "Tutorial on training recurrent neural networks, covering BPTT, RTRL, EKF and the "echo state network" approach," GMD Report 159, German National Research Institute for Computer Science, 2002.

- [85] M. G. Rabbat and R. D. Nowak, “Quantized incremental algorithms for distributed optimization,” *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 4, pp. 798 – 808, 2005.
- [86] D. Estrin, L. Girod, G. Pottie, and M. Srivastava, “Instrumenting the world with wireless sensor networks,” *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, ICASSP*, pp. 2033–2036, 2001.
- [87] J. B. Predd, S. B. Kulkarni, and H. V. Poor, “Distributed learning in wireless sensor networks,” *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 56–69, 2006.
- [88] D. Culler, D. Estrin, and M. Srivastava, “Overview of sensor networks,” *Computer*, vol. 37, no. 8, pp. 41–49, 2004.
- [89] C. G. Lopes and A. H. Sayed, “Adaptive processing over distributed networks,” *IEEE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E90-A, no. 8, pp. 1504–1510, 2007.
- [90] F. S. Cattivelli and A. H. Sayed, “Diffusion LMS strategies for distributed estimation,” *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1035–1048, 2010.
- [91] F. S. Cattivelli, C. G. Lopes, and A. H. Sayed, “Diffusion recursive least-squares for distributed estimation over adaptive networks,” *IEEE Transactions on Signal Processing*, vol. 56, no. 5, pp. 1865–1877, 2008.
- [92] C. G. Lopes and A. H. Sayed, “Incremental adaptive strategies over distributed networks,” *IEEE Transactions on Signal Processing*, vol. 55, no. 8, pp. 4064–4077, 2007.
- [93] F. S. Cattivelli and A. H. Sayed, “Diffusion strategies for distributed Kalman filtering and smoothing,” *IEEE Transactions on Automatic Control*, vol. 56, no. 5, pp. 1865–1877, 2008.
- [94] 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, no. 7, pp. 3122–3136, 2008.

- [95] D. P. Mandic, Y. Xia, and S. C. Douglas, "Steady state analysis of the CLMS and augmented CLMS algorithms for noncircular complex signals," *Proceedings of the 44th Asilomar Conference on Signals, Systems, Computers*, pp. 1635–1639, 2010.
- [96] 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, 1994.
- [97] M. S. Sachdev and M. M. Giray, "A least error squares technique for determining power system frequency," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-104, no. 2, pp. 437–443, 1985.
- [98] G. P. Hancke, "The optimal frequency estimation of a noisy sinusoidal signal," *IEEE Transactions on Instrumentation and Measurement*, vol. 39, no. 6, pp. 843–846, 1990.
- [99] O. Vainio and S. Ovaska, "Digital filtering for robust 50/60 Hz zero-crossing detectors," *IEEE Transactions on Instrumentation and Measurement*, vol. 45, no. 2, pp. 426–430, 1996.
- [100] V. Kaura and V. Blasko, "Operation of a phase locked loop system under distorted utility conditions," *IEEE Transactions on Industry Applications*, vol. 33, no. 1, pp. 58 – 63, 1997.
- [101] S. K. Chung, "A phase tracking system for three phase utility interface inverters," *IEEE Transactions on Power Electronics*, vol. 15, no. 3, pp. 431–438, 2000.
- [102] H. Karimi, M. Karimi-Ghartemani, and M. R. Iravani, "Estimation of frequency and its rate of change for applications in power systems," *IEEE Transactions on Power Delivery*, vol. 19, no. 2, pp. 472–480, 2004.
- [103] I. Kamwa and R. Grondin, "Fast adaptive schemes for tracking voltage phasor and local frequency in power transmission and distribution systems," *IEEE Transactions on Power Delivery*, vol. 7, no. 2, pp. 789–795, 1992.
- [104] V. V. Terzija, "Improved recursive Newton-type algorithm for frequency and spectra estimation in power systems," *IEEE Transactions on Instrumentation and Measurement*, vol. 52, no. 5, pp. 1654–1659, 2003.

- [105] F. Nagy, "Measurement of signal parameters using nonlinear observers," *IEEE Transactions on Instrumentation and Measurement*, vol. 41, no. 1, pp. 152 – 155, 1992.
- [106] P. K. Dash, R. K. Jena, G. Panda, and Aurobinda Routray, "An extended complex Kalman filter for frequency measurement of distorted signals," *IEEE Transactions on Instrumentation and Measurement*, vol. 49, no. 4, pp. 746 – 753, 2000.
- [107] A. Routray, A. K. Pradhan, and K. P. Rao, "A novel Kalman filter for frequency estimation of distorted signals in power systems," *IEEE Transactions on Instrumentation and Measurement*, vol. 51, no. 3, pp. 469 – 479, 2002.
- [108] V. Eckhardt, P. Hippe, and G. Hosemann, "Dynamic measuring of frequency and frequency oscillations in multiphase power systems," *IEEE Transactions on Power Delivery*, vol. 4, no. 1, pp. 95–102, 1989.
- [109] E. Clarke, *Circuit Analysis of A.C. Power Systems*, New York: Wiley, 1943.
- [110] M. Akke, "Frequency estimation by demodulation of two complex signals," *IEEE Transactions on Power Delivery*, vol. 12, no. 1, pp. 157–163, 1997.
- [111] A. Cataliotti, V. Cosentino, and S. Nuccio, "A phase-locked loop for the synchronization of power quality instruments in the presence of stationary and transient disturbances," *IEEE Transactions on Instrumentation and Measurement*, vol. 56, no. 6, pp. 2232–2239, 2007.
- [112] A. K. Pradhan, A. Routray, and A. Basak, "Power system frequency estimation using least mean square technique," *IEEE Transactions on Power Delivery*, vol. 20, no. 3, pp. 761–766, 2005.
- [113] B. Subudhi, P. K. Ray, S. R. Mohanty, and A. M. Panda, "A comparative study on different power system frequency estimation techniques," *International Journal of Automation and Control*, vol. 3, no. 2-3, pp. 202–215, 2009.
- [114] 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, no. 3, pp. 761–766, 1999.

- [115] M. M. Canteli, A. O. Fernandez, L. I. Eguiluz, and C. R. Estebanez, "Three-phase adaptive frequency measurement based on Clarke's transformation," *IEEE Transactions on Power Delivery*, vol. 21, no. 3, pp. 1101–1105, 2006.
- [116] D. Beeman, *Industrial Power System Handbook*, McGraw-Hill, 1955.
- [117] H. S. Song and K. Nam, "Instantaneous phase-angle estimation algorithm under unbalanced voltage sag conditions," *Proceedings of IEE Genetations, Transmission and Distribution*, vol. 147, pp. 409–415, 2000.
- [118] P. Rodriguez, J. Pou, J. Bergas, J. I. Candela, R. P. Burgos, and D. Boroyevich, "Decoupled double synchronous reference frame PLL for power converter control," *IEEE Transactions on Power Electronics*, vol. 22, no. 2, pp. 584–592, 2007.
- [119] M. Mojiri, D. Yazdani, and A. Bakhshai, "Robust adaptive frequency estimation of three-phase power system," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 7, pp. 1793–1802, 2010.
- [120] Y. Xia, B. Jelfs, M. M. Van Hulle, J. C. Príncipe, and D. P. Mandic, "An augmented echo state network for nonlinear adaptive filtering of complex noncircular signals," *IEEE Transactions on Neural Networks*, vol. 22, no. 1, pp. 74–83, 2011.
- [121] M. H. J. Bollen, *Understanding Power Quality Problems: Voltages and Interruptions*, New York: IEEE Press, 1999.
- [122] M. H. J. Bollen, "Voltage sags in three-phase systems," *IEEE Power Engineering Review*, vol. 21, no. 9, pp. 8–15, 2001.
- [123] D. C. Rife and R. R. Boorstyn, "Single-tone parameter estimation from discrete-time observations," *IEEE Transactions on Information Theory*, vol. 20, no. 5, pp. 591–598, 1974.
- [124] S. C. Douglas and D. P. Mandic, "Performance analysis of the conventional complex LMS and augmented complex LMS algorithms," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP*, pp. 3794–3797, 2010.
- [125] S. B. Choe and J. J. Faraway, "Modeling head and hand orientation during motion using quaternions," *Journal of Aerospace*, vol. 113, no. 1, pp. 186–192, 2004.

- [126] S. Javid, C. Cheong Took, C. Jahanchahi, N. Le Bihan, and D. P. Mandic, “Blind extraction of improper quaternion sources,” *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, (ICASSP)*, pp. 3708–3711, 2011.
- [127] C. Cheong Took and D. P. Mandic, “The quaternion LMS algorithm for adaptive filtering of hypercomplex processes,” *IEEE Transactions on Signal Processing*, vol. 57, no. 4, pp. 1316–1327, 2009.
- [128] C. Cheong Took and D. P. Mandic, “Augmented second order statistics of quaternion random signals,” *Signal Processing*, vol. 91, pp. 214–224, 2011.
- [129] C. Cheong Took and D. P. Mandic, “A quaternion widely linear adaptive filter,” *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4427–4431, 2010.
- [130] B. C. Ujang, C. Cheong Took, and D. P. Mandic, “Augmented second order statistics of quaternion random signals,” *Accepted in IEEE Transactions on Neural Networks*, 2011.
- [131] M. J. Ablowitz and A. S. Fokas, *Complex Variables*, Cambridge University Press, 2003.
- [132] K. Kreutz-Delgado, “The complex gradient operator and the $\mathbb{CR}$ calculus,” *Lecture Supplement ECE275A*, pp. 1–74, 2006.
- [133] D. Franken, “A complex gradient operator and its application in adaptive array theory,” *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 44, no. 9, pp. 839–843, 1997.
- [134] D. H. Brandwood, “A complex gradient operator and its application in adaptive array theory,” *IEE Proceedings H (Microwaves, Optics and Antennas)*, vol. 130, no. 1, pp. 11–16, 1983.
- [135] R. J. McAulay and E. M. Hofstetter, “Barankin bounds on parameter estimation,” *IEEE Transactions on Information Theory*, vol. 17, pp. 669–676, 1971.
- [136] L. P. Seidman, “Performance limitations and error calculations for parameter estimation,” *Proceedings of IEEE*, vol. 58, pp. 644–652, 1970.

-
- [137] J. Ziv and M. Zakai, "Some lower bounds on signal parameter estimation," *IEEE Transactions on Information Theory*, vol. 15, pp. 386–391, 1969.
- [138] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, New Jersey: Prentice Hall, 1993.

Appendix A

$\mathbb{C}\mathbb{R}$ calculus

Consider a complex-valued function in $\mathbb{C}$, $f(\mathbf{x}) = u(\mathbf{x}_r, \mathbf{x}_i) + jv(\mathbf{x}_r, \mathbf{x}_i)$, where $\mathbf{x} = \mathbf{x}_r + j\mathbf{x}_i$.

The classical complex differentiability is defined as the limit

$$f'(\mathbf{x}_0) = \lim_{\Delta \mathbf{x} \rightarrow \mathbf{0}} \frac{f(\mathbf{x}_0 + \Delta \mathbf{x}) - f(\mathbf{x}_0)}{\Delta \mathbf{x}} \quad (\text{A.1})$$

which is independent of the direction in which $\Delta \mathbf{x}$ approaches $\mathbf{0}$ in the complex plane.

This however requires the Cauchy-Riemann equations [131]

$$\frac{\partial u}{\partial \mathbf{x}_r} = \frac{\partial v}{\partial \mathbf{x}_i} \quad \text{and} \quad \frac{\partial u}{\partial \mathbf{x}_i} = -\frac{\partial v}{\partial \mathbf{x}_r} \quad (\text{A.2})$$

be satisfied. These conditions are necessary for $f(\mathbf{x})$ to be differentiable. A function that is complex-differentiable on its entire domain is called holomorphic or analytic.

In statistical signal processing in the complex domain $\mathbb{C}$, real-valued cost functions, such as a likelihood function or a quadratic form, are frequently encountered. However, it is obvious that the Cauchy-Riemann conditions do not hold in such cases, since real-valued functions have $v(\mathbf{x}_r, \mathbf{x}_i) = 0$, and hence are not analytic. Indeed, the Cauchy-Riemann equations impose strict constraints on $u(\mathbf{x}_r, \mathbf{x}_i)$, $v(\mathbf{x}_r, \mathbf{x}_i)$ and thus $f(\mathbf{x})$.

A convenient framework on the calculation of the derivatives of function of complex variables (even when the Cauchy-Riemann equations do not hold), is the so called $\mathbb{C}\mathbb{R}$ calculus [132], which is also known as Wirtinger calculus¹ [133], and Brandwood's result [134]. It allows for the calculation of the gradients of any function of both $\mathbf{x}$ and $\mathbf{x}^*$, including the standard cost function $J(\mathbf{x}) : \mathbb{C}^N \mapsto \mathbb{R}$. Expanding the complex vectors $\mathbf{x}$ and $\mathbf{x}^*$ in terms of their real and imaginary components, $\mathbf{x}_r$ and $\mathbf{x}_i$, gives $J(\mathbf{x}_r, \mathbf{x}_i) : \mathbb{R}^N \times \mathbb{R}^N \mapsto \mathbb{R}$. The $\mathbb{C}\mathbb{R}$ derivatives of such functions are taken with respect to $\mathbf{x}$ and $\mathbf{x}^*$,

¹Wirtinger's result has been used largely within the German speaking engineering community.

while keeping the other variable constant, resulting in

$$\begin{aligned} \mathbb{R} - \text{derivative} : \left. \frac{\partial J}{\partial \mathbf{x}} \right|_{\mathbf{x}^* = \text{const}} &= \frac{1}{2} \left(\frac{\partial J}{\partial \mathbf{x}_r} - j \frac{\partial J}{\partial \mathbf{x}_i} \right) \\ \mathbb{R}^* - \text{derivative} : \left. \frac{\partial J}{\partial \mathbf{x}^*} \right|_{\mathbf{x} = \text{const}} &= \frac{1}{2} \left(\frac{\partial J}{\partial \mathbf{x}_r} + j \frac{\partial J}{\partial \mathbf{x}_i} \right) \end{aligned} \quad (\text{A.3})$$

The first extension (also called the $\mathbb{R}$ -derivative) is equivalent to the standard Cauchy-Riemann derivative for holomorphic complex functions, whereas the second expression (also called the $\mathbb{R}^*$ -derivative) is used for the calculation of gradients of functions which depend on both $\mathbf{x}$ and $\mathbf{x}^*$, as is the case with real-valued cost functions. For instance, consider a $\mathbb{C}$ -differentiable function, i.e. $f(\mathbf{x}) = \mathbf{x}$, the $\mathbb{R}$ -derivative in (A.3) which is equivalent to the standard Cauchy-Riemann derivative, is $f'(\mathbf{x}) = \mathbf{1}$, and the $\mathbb{R}^*$ -derivative vanishes, whereas for a non-analytic function, i.e. $f(\mathbf{x}) = \|\mathbf{x}\|^2 = \mathbf{x}^T \mathbf{x}^*$, the $\mathbb{R}$ -derivative gives $\mathbf{x}^*$, and the $\mathbb{R}^*$ -derivative gives $\mathbf{x}$. For more detail, see [41, 132].

In stochastic gradient based filtering in the complex domain, the cost function is dependent on both the weight vectors $\mathbf{w}$ and $\mathbf{w}^*$. It can be shown that the maximum change in the cost function on the error surface occurs in the direction of the $\mathbb{R}^*$ -derivative [24],

$$\nabla_{\mathbf{w}} J = \frac{\partial J}{\partial \mathbf{w}^*} \quad (\text{A.4})$$

which will be used in the derivation of the learning algorithms in this work.

Appendix B

Derivation of RNCLMS Algorithm

The conventional NCLMS algorithm is summarised as [2]

$$\begin{aligned} y(k) &= \mathbf{x}^T(k) \mathbf{w}(k) \\ e(k) &= d(k) - y(k) \\ \mathbf{w}(k+1) &= \mathbf{w}(k) + \frac{\mu e(k) \mathbf{x}^*(k)}{\|\mathbf{x}(k)\|^2 + \delta} \end{aligned} \quad (\text{B.1})$$

where $\mathbf{w}(k)$ is $L \times 1$ weight vector, and RNCLMS algorithm can be achieved by making $\delta(k)$ gradient adaptive as

$$\delta(k) = \delta(k-1) - \rho \nabla_{\delta(k-1)} E(k) \quad (\text{B.2})$$

where $\nabla_{\delta(k-1)} E(k) = \Re[e^*(k) \frac{\partial e(k)}{\partial \delta(k-1)}]$. Note that

$$\frac{\partial e(k)}{\partial \delta(k-1)} = - \left[\mathbf{x}^T(k) \frac{\partial \mathbf{w}(k)}{\partial \delta(k-1)} \right] \quad (\text{B.3})$$

and

$$\frac{\partial \mathbf{w}(k)}{\partial \delta(k-1)} = - \frac{\mu e(k-1) \mathbf{x}^*(k-1)}{[\|\mathbf{x}(k-1)\|^2 + \delta(k-1)]^2} \quad (\text{B.4})$$

Then we can obtain

$$\nabla_{\delta(k-1)} E(k) = \frac{\mu \Re[e^*(k) e(k-1) \mathbf{x}^T(k) \mathbf{x}^*(k-1)]}{[\|\mathbf{x}(k-1)\|^2 + \delta(k-1)]^2} \quad (\text{B.5})$$

and

$$\delta(k) = \delta(k-1) - \frac{\rho \mu \Re[e^*(k) e(k-1) \mathbf{x}^T(k) \mathbf{x}^*(k-1)]}{[\|\mathbf{x}(k)\|^2 + \delta(k-1)]^2}$$

Appendix C

Derivation of (8.20)

From the standard three-phase system in (8.1) and (8.2), the components $v_\alpha(k)$ and $v_\beta(k)$ of the complex voltage $v(k) = v_\alpha(k) + jv_\beta(k)$, obtained using the $\alpha\beta$ transformation, can be derived as

$$\begin{aligned}
 v_\alpha(k) &= \sqrt{\frac{2}{3}} \left(v_a(k) - \frac{v_b(k)}{2} - \frac{v_c(k)}{2} \right) \\
 &= \sqrt{\frac{2}{3}} \left(V_a(k) \cos(\omega k \Delta T + \phi) - \frac{V_b(k)}{2} \cos(\omega k \Delta T + \phi - \frac{2\pi}{3}) - \frac{V_c(k)}{2} \cos(\omega k \Delta T + \phi + \frac{2\pi}{3}) \right) \\
 &= \sqrt{\frac{2}{3}} \left(V_a(k) \cos(\omega k \Delta T + \phi) - \frac{V_b(k)}{2} \left(\cos(\omega k \Delta T + \phi) \cos \frac{2\pi}{3} + \sin(\omega k \Delta T + \phi) \sin \frac{2\pi}{3} \right) \right. \\
 &\quad \left. - \frac{V_c(k)}{2} \left(\cos(\omega k \Delta T + \phi) \cos \frac{2\pi}{3} - \sin(\omega k \Delta T + \phi) \sin \frac{2\pi}{3} \right) \right) \\
 &= \left(\frac{\sqrt{6}V_a(k)}{3} + \frac{\sqrt{6}(V_b(k) + V_c(k))}{12} \right) \cos(\omega k \Delta T + \phi) - \frac{\sqrt{2}(V_b(k) - V_c(k))}{4} \sin(\omega k \Delta T + \phi) \\
 v_\beta(k) &= \sqrt{\frac{2}{3}} \left(\frac{\sqrt{3}v_b(k)}{2} - \frac{\sqrt{3}v_c(k)}{2} \right) \\
 &= \frac{\sqrt{2}}{2} (v_b(k) - v_c(k)) \\
 &= \frac{\sqrt{2}}{2} \left(V_b(k) \left(\cos(\omega k \Delta T + \phi) \cos \frac{2\pi}{3} + \sin(\omega k \Delta T + \phi) \sin \frac{2\pi}{3} \right) \right. \\
 &\quad \left. - V_c(k) \left(\cos(\omega k \Delta T + \phi) \cos \frac{2\pi}{3} - \sin(\omega k \Delta T + \phi) \sin \frac{2\pi}{3} \right) \right) \\
 &= -\frac{\sqrt{2}(V_b(k) - V_c(k))}{4} \cos(\omega k \Delta T + \phi) + \frac{\sqrt{6}(V_b(k) + V_c(k))}{4} \sin(\omega k \Delta T + \phi)
 \end{aligned}$$

Keeping in mind that

$$\begin{aligned}\cos(\omega k \Delta T + \phi) &= \frac{e^{j(\omega k \Delta T + \phi)} + e^{-j(\omega k \Delta T + \phi)}}{2} \\ \sin(\omega k \Delta T + \phi) &= \frac{e^{j(\omega k \Delta T + \phi)} - e^{-j(\omega k \Delta T + \phi)}}{2j}\end{aligned}\quad (\text{C.1})$$

we have

$$\begin{aligned}\hat{v}(k) &= v_\alpha(k) + jv_\beta(k) \\ &= \left(\frac{\sqrt{6}V_a(k)}{3} + \frac{\sqrt{6}(V_b(k) + V_c(k))}{12} \right) \frac{e^{j(\omega k \Delta T + \phi)} + e^{-j(\omega k \Delta T + \phi)}}{2} \\ &\quad - \frac{\sqrt{2}(V_b(k) - V_c(k))}{4} \frac{e^{j(\omega k \Delta T + \phi)} - e^{-j(\omega k \Delta T + \phi)}}{2j} \\ &\quad - j \frac{\sqrt{2}(V_b(k) - V_c(k))}{4} \frac{e^{j(\omega k \Delta T + \phi)} + e^{-j(\omega k \Delta T + \phi)}}{2} \\ &\quad + j \frac{\sqrt{6}(V_b(k) + V_c(k))}{4} \frac{e^{j(\omega k \Delta T + \phi)} - e^{-j(\omega k \Delta T + \phi)}}{2j} \\ &= \underbrace{\frac{\sqrt{6}(V_a(k) + V_b(k) + V_c(k))}{6}}_A e^{j(\omega k \Delta T + \phi)} \\ &\quad + \underbrace{\frac{\sqrt{6}(2V_a(k) - V_b(k) - V_c(k))}{12} - \frac{\sqrt{2}(V_b(k) - V_c(k))}{4} j}_B e^{-j(\omega k \Delta T + \phi)}\end{aligned}\quad (\text{C.2})$$

Then the complex-valued $\hat{v}(k)$ in (C.2) can be written in the form of a standard part (left hand term) and a conjugate part (right hand term). Augmented complex statistics [28, 41] shows that $\hat{v}(k)$ is second order circular with rotation invariant probability density function in the complex plane if and only if B vanishes, which can only be achieved when $V_a(k), V_b(k), V_c(k)$ are identical at each time instant, and (C.2) simplifies into (8.4). In unbalanced conditions, A is real-valued, but $B \neq 0$, and can be complex-valued, which results in a rotation-variant distribution of $\hat{v}(k)$ in the complex plane, and a second order noncircular $\hat{v}(k)$.

Appendix D

Cramer-Rao Lower Bound for Unbiased Frequency Estimator on Unbalanced Three-Phase Power Systems

A lower bound on the variance of unbiased estimators is useful in practice. Firstly, it allows for the minimum variance estimation assessment in case the estimator can attain the bound for all values of the unknown parameter. Secondly, it provides a benchmark against which the performance of a specific unbiased estimator can be compared. Furthermore, it provides an insight into the infeasibility of the design of unbiased estimator whose variance appears less than the bound. Although many such variance bounds exist [135–137], the Cramer-Rao Lower Bound (CRLB) is the most straightforward and efficient, and allows us to immediately determine whether an estimator has attained its best theoretical performance [138]. In this appendix, the CRLB for unbiased frequency estimator in unbalanced three-phase power systems is derived.

For a scalar unknown parameter, consider a joint probability density function (PDF) that describes a random measurement process, $p(\mathbf{x}; \theta)$, which satisfies the ‘regularity’ condition [138]

$$E\left[\frac{\partial \ln(p(\mathbf{x}; \theta))}{\partial \theta}\right] = 0 \quad \text{for all } \theta \quad (\text{D.1})$$

where $E[\cdot]$ is the expectation operator, $\ln(\cdot)$ the natural logarithm operator, $\mathbf{x}$ a vector

of observations, and θ the unknown parameter of interest. Then, the variance of any unbiased estimator on $\hat{\theta}$ must satisfy [138]

$$\text{var}(\hat{\theta}) \geq \frac{1}{-E \left[\frac{\partial^2 \ln(p(\mathbf{x}; \theta))}{\partial \theta^2} \right]} \quad (\text{D.2})$$

where

$$E \left[\frac{\partial^2 \ln(p(\mathbf{x}; \theta))}{\partial \theta^2} \right] = \int \frac{\partial^2 \ln(p(\mathbf{x}; \theta))}{\partial \theta^2} p(\mathbf{x}; \theta) d\mathbf{x} \quad (\text{D.3})$$

In the vector case, when a set of parameters $\boldsymbol{\theta} = [\theta_1, \dots, \theta_i, \dots, \theta_L]^T$ are unknown, the CRLB for the i_{th} parameter is found as the $[i, i]$ element of the inverse of a matrix as

$$\text{var}(\hat{\theta}_i) \geq \mathbf{I}_{ii}^{-1}(\boldsymbol{\theta}) \quad (\text{D.4})$$

where $\mathbf{I}(\boldsymbol{\theta})$ is the $L \times L$ Fisher information matrix, defined as

$$\mathbf{I}_{ij}(\boldsymbol{\theta}) = -E \left[\frac{\partial^2 \ln(p(\mathbf{x}; \boldsymbol{\theta}))}{\partial \theta_i \partial \theta_j} \right] \quad (\text{D.5})$$

for $i = 1, 2, \dots, L; j = 1, 2, \dots, L$. To evaluate (D.5), the true value of $\boldsymbol{\theta}$ is used.

The widely linear modelling of the complex-valued voltage $v(k)$ obtained from noisy unbalanced three-phase power systems can be expressed as

$$v(k) = Ae^{j(\omega k \Delta T + \phi)} + Be^{-j(\omega k \Delta T + \phi)} + n(k) \quad (\text{D.6})$$

where $n(k)$ is complex-valued doubly white Gaussian noise with zero mean and variance σ_n^2 , that is, $n = n_r + jn_i$, $\sigma_{n_r}^2 = \sigma_{n_i}^2 = \frac{1}{2}\sigma_n^2$, and $n_r \perp n_i$. Alternatively, (D.6) can be rewritten in terms of its real and imaginary components $v_r(k)$ and $v_i(k)$ as

$$\begin{aligned} v(k) = & \underbrace{A \cos(2\pi f k \Delta T + \phi) + B_r \cos(2\pi f k \Delta T + \phi) + B_i \sin(2\pi f k \Delta T + \phi) + n_r(k)}_{v_r(k)} \\ & + j \underbrace{(A \sin(2\pi f k \Delta T + \phi) + B_i \cos(2\pi f k \Delta T + \phi) - B_r \sin(2\pi f k \Delta T + \phi) + n_i(k))}_{v_i(k)} \end{aligned} \quad (\text{D.7})$$

where B_r and B_i are respectively the real and imaginary parts of complex-valued B . The joint PDF (assuming doubly white Gaussian noise) when the parameter of interest is the

system frequency f is given by [123]

$$p(\mathbf{v}; f) = \left(\frac{1}{\pi\sigma_n^2}\right)^K \exp \left[-\frac{1}{\sigma_n^2} \sum_{k=1}^K ((v_r(k) - m_r(k))^2 + (v_i(k) - m_i(k))^2) \right] \quad (\text{D.8})$$

with its likelihood function

$$\ln(p(\mathbf{v}; f)) = -K \ln(\pi\sigma_n^2) - \frac{1}{\sigma_n^2} \sum_{k=1}^K ((v_r(k) - m_r(k))^2 + (v_i(k) - m_i(k))^2) \quad (\text{D.9})$$

where $\mathbf{v} = [v(1), v(2), \dots, v(K)]$ with K the number of observations, and

$$m_r(k) = A \cos(2\pi f k \Delta T + \phi) + B_r \cos(2\pi f k \Delta T + \phi) + B_i \sin(2\pi f k \Delta T + \phi) \quad (\text{D.10})$$

$$m_i(k) = A \sin(2\pi f k \Delta T + \phi) - B_r \sin(2\pi f k \Delta T + \phi) + B_i \cos(2\pi f k \Delta T + \phi) \quad (\text{D.11})$$

Therefore,

$$\frac{\partial \ln(p(\mathbf{v}; f))}{\partial f} = \frac{2}{\sigma_n^2} \sum_{k=1}^K \left((v_r(k) - m_r(k)) \frac{\partial m_r(k)}{\partial f} + (v_i(k) - m_i(k)) \frac{\partial m_i(k)}{\partial f} \right) \quad (\text{D.12})$$

and

$$E \left[\frac{\partial^2 \ln(p(\mathbf{v}; f))}{\partial f^2} \right] = -\frac{2}{\sigma_n^2} \sum_{k=1}^K \left(\left(\frac{\partial m_r(k)}{\partial f} \right)^2 + \left(\frac{\partial m_i(k)}{\partial f} \right)^2 \right) \quad (\text{D.13})$$

where

$$\frac{\partial m_r(k)}{\partial f} = 2\pi k \Delta T \left(-(A + B_r) \sin(2\pi f k \Delta T + \phi) + B_i \cos(2\pi f k \Delta T + \phi) \right) \quad (\text{D.14})$$

$$\frac{\partial m_i(k)}{\partial f} = 2\pi k \Delta T \left((A - B_r) \cos(2\pi f k \Delta T + \phi) - B_i \sin(2\pi f k \Delta T + \phi) \right) \quad (\text{D.15})$$

According to (D.2), the CRLB of frequency estimator on unbalanced three-phase power systems can be derived as

$$\text{var}(\hat{f}) \geq \frac{\sigma_n^2}{2 \sum_{k=1}^K \left(\left(\frac{\partial m_r(k)}{\partial f} \right)^2 + \left(\frac{\partial m_i(k)}{\partial f} \right)^2 \right)} \quad (\text{D.16})$$

When all the parameters within (D.7) are unknown, that is $\boldsymbol{\theta} = [f, A, B_r, B_i, \phi]^T$, the

likelihood function becomes

$$\ln(p(\mathbf{v}; \boldsymbol{\theta})) = -K \ln(\pi \sigma_n^2) - \frac{1}{\sigma_n^2} \sum_{k=1}^K ((v_r(k) - m_r(k))^2 + (v_i(k) - m_i(k))^2) \quad (\text{D.17})$$

and the elements within the Fisher information matrix can be derived as

$$\mathbf{I}_{ij}(\boldsymbol{\theta}) = \frac{2}{\sigma_n^2} \sum_{k=1}^K \left(\frac{\partial m_r}{\partial \theta_i} \frac{\partial m_r}{\partial \theta_j} + \frac{\partial m_i}{\partial \theta_i} \frac{\partial m_i}{\partial \theta_j} \right) \quad (\text{D.18})$$

where

$$\frac{\partial m_r(k)}{\partial f} = 2\pi k \Delta T \left(-(A + B_r) \sin(2\pi f k \Delta T + \phi) + B_i \cos(2\pi f k \Delta T + \phi) \right) \quad (\text{D.19})$$

$$\frac{\partial m_i(k)}{\partial f} = 2\pi k \Delta T \left((A - B_r) \cos(2\pi f k \Delta T + \phi) - B_i \sin(2\pi f k \Delta T + \phi) \right) \quad (\text{D.20})$$

$$\frac{\partial m_r(k)}{\partial A} = \cos(2\pi f k \Delta T + \phi) \quad (\text{D.21})$$

$$\frac{\partial m_i(k)}{\partial A} = \sin(2\pi f k \Delta T + \phi) \quad (\text{D.22})$$

$$\frac{\partial m_r(k)}{\partial B_r} = \cos(2\pi f k \Delta T + \phi) \quad (\text{D.23})$$

$$\frac{\partial m_i(k)}{\partial B_r} = -\sin(2\pi f k \Delta T + \phi) \quad (\text{D.24})$$

$$\frac{\partial m_r(k)}{\partial B_i} = \sin(2\pi f k \Delta T + \phi) \quad (\text{D.25})$$

$$\frac{\partial m_i(k)}{\partial B_i} = \cos(2\pi f k \Delta T + \phi) \quad (\text{D.26})$$

$$\frac{\partial m_r(k)}{\partial \phi} = -(A + B_r) \sin(2\pi f k \Delta T + \phi) + B_i \cos(2\pi f k \Delta T + \phi) \quad (\text{D.27})$$

$$\frac{\partial m_i(k)}{\partial \phi} = (A - B_r) \cos(2\pi f k \Delta T + \phi) - B_i \sin(2\pi f k \Delta T + \phi) \quad (\text{D.28})$$

hence

$$\begin{aligned}
 \mathbf{I}_{11}(\boldsymbol{\theta}) &= \frac{2}{\sigma_n^2} \sum_{k=1}^K (2\pi k \Delta T)^2 (A^2 + B_r^2 + B_i^2 - 2A(B_r \cos(4\pi f k \Delta T + 2\phi) + B_i \sin(4\pi f k \Delta T + 2\phi))) \\
 \mathbf{I}_{12}(\boldsymbol{\theta}) &= \mathbf{I}_{21}(\boldsymbol{\theta}) = \frac{2}{\sigma_n^2} \sum_{k=1}^K 2\pi k \Delta T (B_i \cos(4\pi f k \Delta T + 2\phi) - B_r \sin(4\pi f k \Delta T + 2\phi)) \\
 \mathbf{I}_{13}(\boldsymbol{\theta}) &= \mathbf{I}_{31}(\boldsymbol{\theta}) = \frac{2}{\sigma_n^2} \sum_{k=1}^K 2\pi k \Delta T (B_i - A \sin(4\pi f k \Delta T + 2\phi)) \\
 \mathbf{I}_{14}(\boldsymbol{\theta}) &= \mathbf{I}_{41}(\boldsymbol{\theta}) = \frac{2}{\sigma_n^2} \sum_{k=1}^K 2\pi k \Delta T (A \cos(4\pi f k \Delta T + 2\phi) - B_r) \\
 \mathbf{I}_{15}(\boldsymbol{\theta}) &= \mathbf{I}_{51}(\boldsymbol{\theta}) = \frac{2}{\sigma_n^2} \sum_{k=1}^K 2\pi k \Delta T (A^2 + B_r^2 + B_i^2 - 2A(B_r \cos(4\pi f k \Delta T + 2\phi) + B_i \sin(4\pi f k \Delta T + 2\phi))) \\
 \mathbf{I}_{22}(\boldsymbol{\theta}) &= \frac{2K}{\sigma_n^2} \\
 \mathbf{I}_{23}(\boldsymbol{\theta}) &= \mathbf{I}_{32}(\boldsymbol{\theta}) = \frac{2}{\sigma_n^2} \sum_{k=1}^K \cos(4\pi f k \Delta T + 2\phi) \\
 \mathbf{I}_{24}(\boldsymbol{\theta}) &= \mathbf{I}_{42}(\boldsymbol{\theta}) = \frac{2}{\sigma_n^2} \sum_{k=1}^K \sin(4\pi f k \Delta T + 2\phi) \\
 \mathbf{I}_{25}(\boldsymbol{\theta}) &= \mathbf{I}_{52}(\boldsymbol{\theta}) = \frac{2}{\sigma_n^2} \sum_{k=1}^K (B_i \cos(4\pi f k \Delta T + 2\phi) - B_r \sin(4\pi f k \Delta T + 2\phi)) \\
 \mathbf{I}_{33}(\boldsymbol{\theta}) &= \frac{2K}{\sigma_n^2} \\
 \mathbf{I}_{34}(\boldsymbol{\theta}) &= \mathbf{I}_{43}(\boldsymbol{\theta}) = 0 \\
 \mathbf{I}_{35}(\boldsymbol{\theta}) &= \mathbf{I}_{53}(\boldsymbol{\theta}) = \frac{2}{\sigma_n^2} \sum_{k=1}^K (-A \sin(4\pi f k \Delta T + 2\phi) + B_i \cos(4\pi f k \Delta T + 2\phi)) \\
 \mathbf{I}_{44}(\boldsymbol{\theta}) &= \frac{2K}{\sigma_n^2} \\
 \mathbf{I}_{45}(\boldsymbol{\theta}) &= \mathbf{I}_{54}(\boldsymbol{\theta}) = \frac{2}{\sigma_n^2} \sum_{k=1}^K (A \cos(4\pi f k \Delta T + 2\phi) - B_r) \\
 \mathbf{I}_{55}(\boldsymbol{\theta}) &= \frac{2}{\sigma_n^2} \sum_{k=1}^K (A^2 + B_r^2 + B_i^2 - 2A(B_r \cos(4\pi f k \Delta T + 2\phi) + B_i \sin(4\pi f k \Delta T + 2\phi)))
 \end{aligned}$$

A more general CRLB for unbiased frequency estimator on unbalanced three-phase power system can be obtained as

$$\text{var}(\hat{f}) \geq \mathbf{I}_{11}^{-1}(\boldsymbol{\theta}) \quad (\text{D.29})$$