

COLLABORATIVE ADAPTIVE FILTERING FOR MACHINE LEARNING

by
BETH JELFS

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
2009

Abstract

Quantitative performance criteria for the analysis of machine learning architectures and algorithms have long been established. However, qualitative performance criteria, which identify fundamental signal properties and ensure any processing preserves the desired properties, are still emerging. In many cases, whilst offline statistical tests exist such as assessment of nonlinearity or stochasticity, online tests which not only characterise but also track changes in the nature of the signal are lacking. To that end, by employing recent developments in signal characterisation, criteria are derived for the assessment of the changes in the nature of the processed signal.

Through the fusion of the outputs of adaptive filters a single collaborative hybrid filter is produced. By tracking the dynamics of the mixing parameter of this filter, rather than the actual filter performance, a clear indication as to the current nature of the signal is given. Implementations of the proposed method show that it is possible to quantify the degree of nonlinearity within both real- and complex-valued data. This is then extended (in the real domain) from dealing with nonlinearity in general, to a more specific example, namely sparsity. Extensions of adaptive filters from the real to the complex domain are non-trivial and the differences between the statistics in the real and complex domains need to be taken into account. In terms of signal characteristics, nonlinearity can be both split- and fully-complex and complex-valued data can be considered circular or noncircular. Furthermore, by combining the information obtained from hybrid filters of different natures it is possible to use this method to gain a more complete understanding of the nature of the nonlinearity within a signal. This also paves the way for building multidimensional feature spaces and their application in

data/information fusion.

To produce online tests for sparsity, adaptive filters for sparse environments are investigated and a unifying framework for the derivation of proportionate normalised least mean square (PNLMS) algorithms is presented. This is then extended to derive variants with an adaptive step-size. In order to create an online test for noncircularity, a study of widely linear autoregressive modelling is presented, from which a proof of the convergence of the test for noncircularity can be given. Applications of this method are illustrated on examples such as biomedical signals, speech and wind data.

Acknowledgement

There are many people I would like to thank for helping in their own way throughout this PhD.

First and foremost I would like to thank my supervisor Danilo Mandic for his support and guidance. His constant enthusiasm is an inspiration to us all. Most importantly I am grateful for his constant attempts to improve my use of punctuation. I have tried my best to make sure his efforts have not gone to waste.

I would also like to thank my colleagues from the Communications and Signal Processing group especially Ling Li, David Looney, Soroush Javidi and Phebe Vayanos. I would like to acknowledge the British Council and the DAAD for their support of the work on EEG and BSE under the project ARC1303. Thanks also go to Loic Baboulaz, Jon Gudnason, Nick Gaubitch, Jesse Berent and Mark Thomas for proving you can combine a PhD with a social life.

Further, the Imperial College Fencing Club, in particular Chris Namih and my teammates Joan Chang, Hannah Bryars and Emily Bottle, deserve thanks for keeping me sane. They gave me an escape in the moments that I needed a break from the PhD.

For all their love and support my parents deserve my never ending gratitude. My brother Sam deserves thanks, if for nothing else, not finishing his PhD before me.

Finally, to Chris, a fencer, colleague, source of data, and so much more, for his constant patience and understanding.

Contents

Abstract	3
Acknowledgement	5
Contents	7
List of Figures	11
List of Tables	17
Abbreviations	19
List of Symbols	21
Chapter 1. Introduction	25
1.1 Background	25
1.2 Motivation	27
1.3 Original Contributions	28
1.4 Thesis Outline	32
Chapter 2. Signal Characterisation	35
2.1 Introduction	35
2.1.1 Characterisation in the Complex Domain	36
2.2 Existing Methods for Signal Characterisation	38
2.2.1 Offline Statistical Tests	39
2.2.2 Measures of Sparsity	40

2.2.3	Noncircularity Measures	42
2.2.4	Online Model Based Tests	43
2.3	Conclusions	44
Chapter 3.	Hybrid Filters for Signal Modality Characterisation	45
3.1	Introduction	45
3.2	Hybrid Filter Structure	47
3.2.1	Convergence and Computational Complexity	48
3.3	Performance of Hybrid Filters	49
3.4	Hybrid Filters for Identifying & Tracking Nonlinearity	53
3.5	Conclusions	56
Chapter 4.	Sparse Adaptive Filters	59
4.1	Introduction	59
4.2	Derivation of the Class of PNLMS Algorithms	61
4.2.1	Convergence in the Mean	62
4.2.2	Convergence in the Mean Square	63
4.2.3	Introducing PNLMS via Normalisation of PLMS	65
4.3	Unifying Approach to GASS Algorithms for PNLMS	68
4.3.1	Linear Update of the Adaptive Step-Size	70
4.3.2	Nonlinear Update of the Adaptive Step-Size	72
4.3.3	Robust Regularisation	72
4.4	Notes on Convergence and Computational Complexity	74
4.5	Performance of the Class of PNLMS Algorithms	75
4.6	Conclusions	79
Chapter 5.	Hybrid Filters for Sparsity and Combining Multiple Signal Modality Trackers	81
5.1	Introduction	81
5.2	Comparison of Algorithms for Sparse Hybrid Filters	82
5.2.1	Effect of Additive Noise	84

5.3 Multiple Signal Modality Trackers	84
5.4 Conclusions	86

Chapter 6. Hybrid Filtering & Signal Modality Characterisation in the

Complex Domain	89
6.1 Introduction	89
6.2 Identification and Tracking of Nonlinearity in Complex Signals	90
6.2.1 Split-Complex vs. Fully-Complex	92
6.3 Widely Linear Autoregressive Modelling	94
6.3.1 Widely Linear Extension to the Standard AR Model	97
6.4 Collaborative Adaptive Filter for Circularity	99
6.4.1 Convergence of the Mixing Parameter	100
6.4.2 Performance of the Hybrid Filter	104
6.5 Tracking Circularity of a Signal	106
6.6 Conclusions	109

Chapter 7. Real World Applications

7.1 Introduction	111
7.2 EEG Data	112
7.2.1 Epileptic Seizure Data	112
7.2.2 Identification of Consciousness States	114
7.3 EMG Data	117
7.4 Speech	119
7.5 Radar	121
7.6 Wind	123
7.6.1 Nonlinearity	125
7.6.2 Split-/Fully-Complex	126
7.6.3 Circularity	127
7.7 Conclusions	129

Chapter 8. Conclusions	131
8.1 Thesis Summary	131
8.2 Future Research	133
Appendix A. Derivation of GASS Algorithms for PNLMS	135
A.1 Derivation of Adaptive Step-Size Algorithm	135
A.2 Derivation of Adaptive Regularisation Algorithm	136
Appendix B. Widely Linear Autoregressive Modelling	139
B.1 Advantage of Widely Linear Model Over Standard Linear Model	139
B.2 Least Squares Solution for the Widely Linear Model	141
Appendix C. Expectation of the Mixing Parameter of Hybrid Filter for Complex Circularity	143
Appendix D. Blind Source Extraction Using a Widely Linear Predictor	147
Bibliography	153

List of Figures

1.1	Deterministic vs. stochastic nature and linear vs. nonlinear nature of real world signals (Modified from [82]).	26
2.1	Block diagram of DVV method [16].	40
2.2	Sparsity in time and frequency domains	41
2.3	Sparsity in time-frequency domain.	42
2.4	Estimated degree of signal nonlinearity for an input signal alternating from linear to nonlinear.	44
3.1	Convexity	46
3.2	Hybrid filter structure	47
3.3	DVV plots and scatter diagrams for linear and nonlinear signals [17]. . .	50
3.4	Performance of the NLMS, NNGD and corresponding hybrid filter for a linear AR(4) signal.	52
3.5	Performance of the NLMS, NNGD and corresponding hybrid filter for a benchmark nonlinear signal.	52
3.6	Evolution of the mixing parameter λ for a hybrid filter combining the NLMS and NNGD algorithms for linear AR(4) signal and benchmark nonlinear signal.	54
3.7	Evolution of the mixing parameter λ for a hybrid filter combining the NLMS and NNGD algorithms for an input signal which alternates from a linear AR(4) signal to a nonlinear signal every 200 samples.	55

3.8	Evolution of the mixing parameter λ for a hybrid filter combining the NLMS and NNGD algorithms for an input signal which alternates from a linear AR(4) signal to a nonlinear signal every 100 samples.	55
3.9	Evolution of the mixing parameter λ for a hybrid filter combining NLMS and NNGD subfilters for an input signal alternating from linear AR(4) to AR(1) to nonlinear every 200 samples.	57
4.1	Performance comparison of the GASS PNLMS algorithms with the standard PNLMS for $\mu = 0.1$	76
4.2	Performance comparison of the GASS PNLMS algorithms with the standard PNLMS for $\mu = 1.95$	77
4.3	Performance comparison of the PNLMS with the adaptive ε PNLMS for variations in μ and ρ	78
4.4	Performance comparison of the adaptive ε PNLMS with the sign regularised PNLMS for variations in μ and ρ	78
5.1	Comparison of the evolution of mixing parameter λ for two different parameter settings of the hybrid filter combining NLMS and PNLMS for an input signal alternating between a linear AR(4) signal to a sparse signal every 200 samples.	82
5.2	Evolution of the mixing parameter λ for a hybrid filter combining the NLMS and SSLMS algorithms for an input signal which alternates from a linear AR(4) signal to a sparse signal every 200 samples.	83
5.3	Evolution of the mixing parameter λ for a hybrid filter combining the NLMS and SSLMS algorithms for an input signal which alternates from a linear AR(4) signal to a sparse signal every 200 samples, with varying degrees of noise.	84
5.4	Evolution of the mixing parameter λ for hybrid filters combining NLMS with either NNGD (<i>broken line</i>) or SSLMS (<i>solid line</i>) for an input signal alternating from linear to nonlinear and linear to sparse every 200 samples.	85

5.5	Comparison of the evolution of the mixing parameters for linear/nonlinear and sparse/nonsparse for an input signal alternating every 200 samples. <i>Solid line:</i> linear samples, <i>broken line:</i> nonlinear samples, <i>dotted line:</i> sparse samples.	87
6.1	Evolution of the mixing parameter λ for a hybrid filter combining the CNNGD and CNLMS, for input nature alternating between linear and nonlinear every 200 samples.	92
6.2	Evolution of the mixing parameter λ for prediction of Ikeda map	93
6.3	Learning curves for fully-complex vs. split-complex prediction of Ikeda map	94
6.4	Distributions of AR(4) with different driving noise	96
6.5	Covariance and pseudocovariance of a standard AR(4) model driven by different driving noises	97
6.6	Signal distributions of the Ikeda map	99
6.7	Covariance and pseudocovariance of the Ikeda map	100
6.8	Convergence curves of the hybrid filter and constituent subfilters for a circular AR(4) process (3.8)	105
6.9	Convergence curves of the hybrid filter and constituent subfilters for a noncircular Ikeda map (6.11)	106
6.10	Evolution of the mixing parameter λ for the linear circular AR(4) process, the widely linear AR(4) model of the Ikeda map (noncircular) and the noncircular Ikeda signals	107
6.11	Evolution of the mixing parameter λ for a linear circular AR(4) signal driven by different noises	108
6.12	Evolution of the mixing parameter λ for signals alternating from linear AR(4) to either widely linear AR(4) model of Ikeda map or Ikeda map .	108
6.13	Evolution of the mixing parameter λ for a signal alternating from linear AR(4) to widely linear AR(4) with a varying alternation period	109

7.1	Top panels: EEG epileptic seizure data. Middle panels: corresponding evolution of λ solid line: nonlinear hybrid filter, broken line: sparse hybrid filter. Bottom panels comparison of nonlinearity and sparsity, evolution over time starting from coordinates (1,1)	113
7.2	The Electrode placement.	114
7.3	One minute of each data set for one EEG channel.	115
7.4	Evolution of mixing parameter λ for different states of coma for one EEG channel ($\lambda \rightarrow 0$ linear, $\lambda \rightarrow 1$ nonlinear).	115
7.5	One minute of each data set averaged across all EEG channels.	116
7.6	Evolution of mixing parameter λ for different states of coma averaged across all EEG channels ($\lambda \rightarrow 0$ linear, $\lambda \rightarrow 1$ nonlinear).	116
7.7	Evolution of the mixing parameter λ for EMG data with the subject tensing and relaxing arm muscles and corresponding response when λ takes discrete values.	118
7.8	Speech signal S1 and corresponding evolution of mixing parameter λ for nonlinearity and sparsity tracking.	120
7.9	Speech signal S3 and corresponding evolution of mixing parameter λ for nonlinearity and sparsity tracking.	122
7.10	Evolution of the mixing parameter λ for alternating blocks of 400 data samples of radar data from the “low” to “high” sea state	123
7.11	Wind recordings as a complex [speed,direction] vector representation . .	124
7.12	Wind speed readings.	125
7.13	Evolution of the mixing parameter λ for a hybrid combination of CN-NGD and CNLMS for a set of wind data.	126
7.14	Wind magnitude over a 24 hour period in an urban environment	127
7.15	Evolution of the mixing parameter λ for the prediction of wind	128
8.1	BSE with a linear predictor	134

D.1 Blind source extraction example showing sources and corresponding ex- traction performance index for both the widely linear predictor and a standard linear adaptive filter.	151
--	-----

List of Tables

4.1	Computational requirements for the NLMS, PNLMS and its GASS variants	75
7.1	Prediction gain for the CLMS, ACLMS and hybrid filters for wind data and circular and non-circular synthetic data	128

Abbreviations

ACLMS:	Augmented Complex Least Mean Square
AF:	Activation Function
AR:	Autoregressive
ARMA:	Autoregressive Moving Average
BSE:	Blind Source Extraction
C3:	Third Order Autocovariance
CLMS:	Complex Least Mean Square
CNLMS:	Complex Normalised Least Mean Square
CNGD:	Complex Nonlinear Gradient Descent
CNNGD:	Complex Normalised Nonlinear Gradient Descent
CRTRL:	Complex Real Time Recurrent Learning
DV:	Delay Vector
DVS:	Deterministic Versus Stochastic
DVV:	Delay Vector Variance
EEG:	Electroencephalography
EMD:	Empirical Mode Decomposition
EMG:	Electromyography
EMSE:	Excess Mean Square Error
GASS:	Gradient Adaptive Step-Size
GLRT:	Generalised Likelihood Ratio Test
iAAFT:	Iterative Amplitude Adjusted Fourier Transform
i.i.d.:	Independent Identically Distributed

IPNLMS:	Improved Proportionate Normalised Least Mean Square
LMS:	Least Mean Square
ML:	Maximum Likelihood
MMSE:	Minimum Mean Square Error
MSPE:	Mean Square Prediction Error
MPNLMS:	μ -Law Proportionate Normalised Least Mean Square
NARMA:	Nonlinear Autoregressive Moving Average
NLMS:	Normalised Least Mean Square
NNGD:	Normalised Nonlinear Gradient Descent
PI:	Performance Index
PLMS:	Proportionate Least Mean Square
PNLMS:	Proportionate Normalised Least Mean Square
REV:	Time Reversibility
SSLMS:	Signed Sparse Least Mean Square
TSE:	Taylor Series Expansion
WLAR:	Widely Linear Autoregressive

List of Symbols

$\mathbf{a}$	Autoregressive Coefficients
$\mathbf{A}$	Mixing Matrix
$\mathbf{c}$	Correlation Vector
$\mathbf{C}_{\mathbf{z}\mathbf{z}}$	Covariance Matrix
$d(k)$	Desired Signal
$e(k)$	Instantaneous Error
$\mathbf{F}$	Unitary Matrix
$g(k)$	Individual Coefficient Gain
$\mathbf{g}(k)$	Augmented Filter Coefficient Vector
$\mathbf{g}_o$	Optimal Augmented Filter Coefficient Vector
$\mathbf{G}(k)$	Proportional Tap Selection (Gain) Matrix
$\mathbf{h}(k)$	Filter Coefficient Vector
$\mathbf{h}_o$	Optimal Filter Coefficient Vector
$\mathbf{I}$	Identity Matrix
$\mathcal{J}(k)$	Cost Function
k	Time Index
k_i	Canonical Correlation
$\mathbf{K}$	Diagonal Matrix of Canonical Correlations
$\mathbf{K}(k)$	Weight Error Correlation Matrix
M	Misadjustment
$\mathbf{M}$	Coherence Matrix
N	Filter Length

$\mathbf{p}$	Pseudocorrelation Vector
$\mathcal{P}_{\mathbf{zz}}$	Pseudocovariance Matrix
$q(k)$	Noise Source
$\mathbf{Q}$	Matrix of Eigenvectors
$\mathbf{R}$	Correlation Matrix
$\mathcal{R}_{\mathbf{zz}}$	Augmented Covariance Matrix
$\mathbf{s}(k)$	Source Signal Vector
$\mathbf{v}(k)$	Weight Error Vector
$\mathbf{w}(k)$	Filter Coefficient Vector
$\mathbf{w}_o$	Optimal Weight Vector
$\mathbf{x}(k)$	Input Signal Vector
$\mathbf{z}$	Complex Signal Vector
$\mathbf{z}_a$	Augmented Complex Vector
α	Multiplicative Constant for GASS algorithms
β	Multiplicative Constant for GASS algorithms
$\gamma(k)$	Derivative of Filter Coefficients by Adaptive Learning Rates
δ	Initialisation Parameter
ε	Regularisation Parameter
$\eta(k)$	Adaptive Filter Learning Rate
$\theta(k)$	Random Variable Uniformly Distributed Over $[0, 2\pi]$
θ	Angle between Two Vectors
ϑ	Higher Order Terms for GASS algorithms
κ_{ii}	Diagonal Elements of $\mathbf{K}$
λ_n	Eigenvalue
$\lambda(k)$	Hybrid Filter Mixing Parameter
$\mathbf{\Lambda}$	Vector of Eigenvalues
μ	Adaptive Filter Step-Size
$\xi(k)$	Expected Value of Squared Error
ρ	Proportionality (Activation) Parameter

$\rho(k)$	Uniformly Distributed Random Variable
σ_x^2	Variance of x
ϕ	Proportionality Function
$\Phi(\cdot)$	Nonlinear Activation Function

Chapter 1

Introduction

1.1 Background

Research on signal modality characterisation is becoming an increasingly important area of multidisciplinary research. Whilst research in this area started in physics in the mid 1990s [42,84,96], its applications in machine learning and signal processing are only just beginning to emerge [18,30]. Signal modality characterisation reveals the changes in the nature of data, such as the degree of nonlinearity, stochasticity, noncircularity, and large effort has been put into devising efficient algorithms for this purpose, as the accurate characterisation of the nature of signals is a key prerequisite for choosing a signal processing framework. The use of inappropriate models can lead to problems in their application or training and in some situations (such as the use of nonlinear model in absence of nonlinearity) can add unnecessary computational complexity.

With this in mind, the area of signal characterisation which has received the most attention is methods for identifying nonlinearity. It is well known that the use of linear models in the presence of nonlinearity can lead to inaccurate representations, but so too can the use of incorrect nonlinear models. Much of the initial work in nonlinear analysis has focussed on chaotic signals, and more recently it has been discussed that the assumption of chaos is too readily applied to all areas where nonlinearity is suspected [82]. While the use of chaotic modelling can allow signals which appear erratic to be predicted, by modelling them as deterministic equations, even small errors in

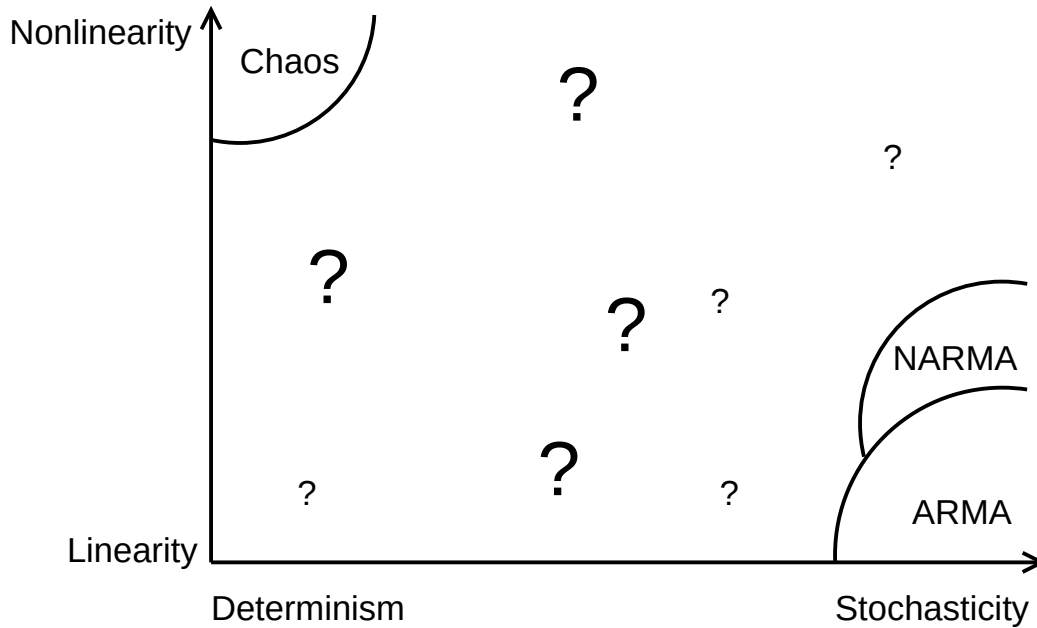


Figure 1.1: Deterministic vs. stochastic nature and linear vs. nonlinear nature of real world signals (Modified from [82]).

initial conditions can grow with time leading to inaccuracy.

Real-world processes are typically mixtures of linear and nonlinear signal components (which can be either deterministic or stochastic) and noise, yet it is common practice to process them using linear, mathematically tractable, models. Consider the range of signals spanned by the fundamental signal characteristics of nonlinearity and stochasticity, shown in Fig. 1.1. In terms of computational models, there are some small areas which are well understood, however, these are very specialised cases and tend to be extremes in nature, such as purely nonlinear deterministic signals (chaos), or linear and stochastic signals represented by autoregressive moving average (ARMA) models. These extremes do not cover the majority of real world signals which do not fit into such classifications. For instance, due to the presence of factors such as noise or uncertainty this leads to most real world signals being represented in the areas '??'; these are significant areas about which we know little or nothing.

Thus, current aspects of signal modality characterisation consider not only methods of characterisation but the types of classification. This involves expanding beyond chaos to different types of nonlinearity and also from nonlinearity to other characteristics.

Knowing more about the nature of the signal being processed and being able to track changes in modality in real-time can provide valuable information in many areas such as health, weather or finance conditions.

Signal modality characterisation has particular application in prediction and signal separation and can be used to provide prior knowledge to a blind algorithm. In addition, by combining the information from several signal modality trackers it becomes possible to build a signal “fingerprint”. As such, signal modality characterisation can provide a powerful tool for machine learning, detailing changes within the signal dynamics along time.

1.2 Motivation

The requirement for adaptive methods of signal characterisation is a multifaceted problem. Many existing approaches to signal characterisation are based upon hypothesis testing, describing the signal in a statistical manner [29, 31, 84]. There is, however, a need for online approaches to signal characterisation. Online methods of signal characterisation allow for real time analysis of data to not only identify the nature of a signal but also track any changes in signal modality. Many existing tests are block based in nature, meaning they test only one section of the data at a time and require accurate selection of the size of these blocks. Incorrect selection of the block size can lead to problems such as confusion between nonlinearity of a signal and nonstationarity, whereas an adaptive method can overcome this. Most existing tests also consider synthetic data only or when they have been applied to real world data it has been under strong assumptions e.g. chaos, stationarity, or circularity.

Existing online approaches also have disadvantages, mainly due to their tendency to rely on underlying models [67], making their application somewhat limited. To overcome these limitations a much more flexible method based on collaborative adaptive filtering [6, 92] is presented. Using a hybrid filter provides a fast, simple to implement test, which is supported by a maturing background theory.

Adaptive filters provide a flexible means of predicting signals through the use of

recursive algorithms, used to estimate the optimal settings for a particular input. This mode of adaptation allows operation in nonstationary environments and provides a greater degree of freedom than standard filtering. However, an adaptive filter still requires some knowledge of the desired output of the system or the use of a training signal to derive the filter parameters. As such, there are many different algorithms for training adaptive filters each based upon a certain class of desired signal properties. It is therefore possible to present a signal as the input to two different adaptive filters to provide a useful simple test for the nature of a signal. However, this method does not provide a truly online adaptive solution, and there are additional ambiguities due to the need to choose many parameters of the corresponding filters. In addition, this approach does not rely on the “synergy” between the filters considered.

Previous implementations of hybrid filters based on convex combinations of adaptive subfilters have focused mainly on the quantitative improvement in performance. The approach presented here instead relies on observing the evolution of the mixing parameter within such a structure. It has been shown that for the convex combination of two filters with different characteristics, the resultant hybrid filter has better tracking capabilities for the generality of real world signals [5, 92, 111]. This can be observed when the system being processed changes with time. The constituent subfilters each having different characteristics will be better suited to different sections of the data. Thus, the hybrid filter gives an overall improved response over either constituent filter, making the hybrid filter better suited to processing a wider range of signals. As such, the hybrid filter is designed to have two constituent subfilters with sufficiently different characteristics so that when the mixing parameter is observed it provides an insight into the nature of the signals. In addition, through the evolution of the mixing parameter, any changes in the modality of the signals can also be tracked.

1.3 Original Contributions

The main contribution of this thesis is the use of convex combinations of adaptive filters for the purpose of signal modality characterisation. The following aspects are believed

to be original contributions:

- **Tracking the Convex Mixing Parameter:** Previous applications of hybrid filters have focussed on the quantitative performance advantage of the hybrid filter, compared to the constituent subfilters. The method presented here relies on having subfilters with different characteristics and observing the behaviour of the convex mixing parameter at the output of the subfilters. This is a novel approach to the application of an already existing structure. The validity and application of this method has been illustrated in publications [II] and [IV]. The tracking of the mixing parameter in one hybrid filter, has been extended by comparing the mixing parameters from two such structures to give a more complete picture of the nature of the signal being analysed, this approach has been presented in [II] and [VI].
- **Derivation and Convergence Analysis of PNLMS:** The proportionate normalised least mean square (PNLMS) algorithm [26] has been presented based on empirical evidence and as such the stability and convergence have received little attention. The matter is also confused as the original version of the algorithm has been superseded by a more commonly used variant [11]. To introduce rigour into the analysis of the PNLMS, a derivation for the PNLMS starting from a proportionate least mean square (PLMS) algorithm through to the optimal step-size for the normalised version has been proposed. The stability analysis shows that the optimal version of the algorithm is actually a combination of aspects from both variants. From the basis of this derivation, adaptive step-size variants of the PNLMS are also presented. The algorithms presented are considered to be novel algorithms which improve on the stability and performance of the PNLMS in critical operating conditions. This work has lead to the publications [III], [VII], [VIII].
- **Widely Linear Autoregressive Modelling:** Current methods of complex autoregressive modelling rely on the circular nature of the noise driving the signal [75, 77]. However, this is not realistic, and the study of widely linear au-

toregressive modelling presents an alternative method of generating more general noncircular benchmark data. Whilst widely linear modelling is not a new premise, the extension of the Yule-Walker equations to include augmented statistics presents a novel approach to autoregressive modelling. This study gives a more complete overview of autoregressive modelling in the complex domain than presented elsewhere and is included in [I].

- **Convergence of the Mixing Parameter λ in Hybrid Filter for Identifying Complex Noncircularity:** A hybrid filter for operation in the complex domain has been presented in [V]. It was then desirable to use the mixing parameter within this hybrid filter to track the degree of noncircularity of signals. In order to assess the capability of such a structure to identify the presence of noncircularity within a signal the behaviour of the mixing parameter was analysed. Previous analysis of the hybrid filter has relied on the convergence and performance of the overall hybrid filter [6,13]. In this case the convergence of each subfilter is assumed based on previous analysis and the response of the mixing parameter for extremes in signal nature is assessed. This results in the novel proof of convergence of the mixing parameter λ presented in [I].

Together with the implementations of the hybrid filter structure for identifying and tracking different signal characteristics such as those in [IX] and [X], this thesis also presents several corresponding real world applications, these include EEG, speech and wind data analysis. The original work outlined has lead to the following publications:

- [I] B. Jelfs, D. Mandic and S. Douglas, “Collaborative Adaptive Filters for Noncircularity Detection”, submitted to *IEEE Signal Processing Letters*, 2009.
- [II] B. Jelfs, S. Javidi, P. Vayanos and D. Mandic, “Characterisation of Signal Modality: Exploiting Signal Nonlinearity in Machine Learning and Signal Processing” accepted to appear in *Journal of Signal Processing Systems, special issue on Machine Learning for Signal Processing*, 2009.
- [III] B. Jelfs, D. Mandic and A. Kuh, “Unifying Framework for Proportionate NLMS

Algorithms and Their Adaptive Step-Size Variants”, invited to *Journal of The Franklin Institute, special issue on Modelling and Simulation in Advanced Communications*, 2009.

- [IV] B. Jelfs, P. Vayanos, S.L. Goh and D. Mandic, “Collaborative Adaptive Filters for Online Knowledge Extraction and Information Fusion”, in *Signal Processing Techniques for Knowledge Extraction and Information Fusion*, D. Mandic, M. Golz, A. Kuh, D. Obradovic and T. Tanaka Eds., Springer, pp. 3–21, 2008.
- [V] B. Jelfs, Y. Xia, D. Mandic and S. Douglas, “Collaborative Adaptive Filtering in the Complex Domain”, in *Proc. IEEE Workshop on Machine Learning for Signal Processing*, pp. 421–425, 2008.
- [VI] B. Jelfs and D. Mandic, “Signal Modality Characterisation Using Collaborative Adaptive Filters”, in *Proc. IAPR Workshop on Cognitive Information Processing*, pp. 58–62, 2008.
- [VII] B. Jelfs, D. Mandic and A. Cichocki, “A Unifying Approach to the Derivation of the Class of PNLMS Algorithms”, in *Proc. International Conference on Digital Signal Processing*, pp. 35–38, 07/2007.
- [VIII] B. Jelfs, D. Mandic and J. Benesty, “A Class of Adaptively Regularised PNLMS Algorithms”, in *Proc. International Conference on Digital Signal Processing*, pp. 19–22, 2007.
- [IX] B. Jelfs and D. Mandic, “Towards Online Monitoring of the Changes in Signal Modality: The Degree of Sparsity”, in *Proc. IMA International Conference on Mathematics in Signal Processing*, pp. 29–32, 2006.
- [X] B. Jelfs, P. Vayanos, M. Chen, S.L. Goh, C. Boukis, T. Gautama, T. Rutkowski, A. Kuh and D. Mandic, “An Online Method For Detecting Nonlinearity Within a Signal”, in *Proc. International Conference on Knowledge-Based & Intelligent Information & Engineering Systems (Lecture Notes in A.I.)*, vol. 4253, pp. 1216–1223, Springer, 2006.

1.4 Thesis Outline

The structure of the thesis is as follows:

Chapter 2 outlines key signal characteristics and existing methods for signal modality characterisation. Signal characteristics (beyond the commonly used second order statistics) in the real domain are discussed along with the differences between modality characterisation in the complex and real domains. After considering the signal characteristics to be identified existing methods of signal characterisation are discussed. These include offline statistical hypothesis based tests and online model based tests. Finally the need for online nonparametric adaptive filtering based tests is highlighted.

In Chapter 3 the general structure of the hybrid filter is outlined. The performance of the hybrid filter is discussed and a hybrid filter for the real time characterisation of signal nonlinearity is presented. The results are then compared against an existing offline based test to illustrate the validity of such an approach.

After considering the general characterisation of nonlinearity, a subset of nonlinearity in the form of sparsity is considered. For this purpose Chapter 4 outlines one of the key adaptive methods for use in sparse environments, the PNLMS algorithm. For the first time the derivation and convergence proofs are presented based on a unifying framework of generalised normalised gradient descent. This framework also allows the development of adaptive step size methods to improve the stability and performance of the PNLMS algorithm in critical operating conditions.

The discussion of adaptive filters for sparse environments gives a basis for the creation of hybrid filters for identifying the sparsity of signals presented in Chapter 5. Once a stable hybrid filter for tracking changes in the sparsity of signals has been presented, the effect of additive noise on such a structure is studied. Finally, the characteristics of nonlinearity and sparsity are treated simultaneously by combining the responses of the mixing parameters of the two different hybrid filters.

In Chapter 6 signal modality characterisation is extended into the complex domain. The nature of nonlinearity in $\mathbb{C}$ is investigated and the differences between split-complex and fully-complex nonlinearity are identified. To support the classification of noncir-

cularity in complex signals, autoregressive modelling in the complex domain is first studied, and a widely linear autoregressive model is presented. Based on a combination of standard and widely linear filters, a test for identification and tracking of the noncircularity of a signal is proposed. The analysis of the mixing parameter λ in this structure is presented, showing that the mixing parameter tracks the changes in circularity of the input signal.

Real world applications of the collaborative adaptive filtering method for signal modality characterisation are presented in Chapter 7. In the real domain the nature of electroencephalography (EEG), electromyography (EMG), and speech data are investigated. In the complex domain the nature of radar and wind data is studied.

Finally Chapter 8 concludes the thesis, giving a summary of the thesis and some remarks on possible directions of future research.

Chapter 2

Signal Characterisation

2.1 Introduction

Before discussing characterisation of signal modalities, certain key properties for defining the nature of a signal should be outlined [58, 85]:

- 1 Linear (strict definition):** A linear signal is generated by a linear time invariant system, driven by white Gaussian noise;
- 2 Linear (commonly adopted):** Definition 1. is relaxed somewhat by allowing the distribution of the signal to deviate from the Gaussian one, which can be interpreted as a linear signal from 1 measured by a static (possibly nonlinear) observation function;
- 3 Nonlinear:** A signal that cannot be generated in the above way is considered nonlinear;
- 4 Deterministic (predictable):** A signal is considered deterministic if it can be precisely described by a set of equations;
- 5 Stochastic:** A signal that is not deterministic.

These are key properties which are adhered to in the following chapters, it is clear from these definitions that the notion of nonlinearity covers a wide range of signals with a large variation in characteristics. Therefore, when considering the nature of a signal

it is not only of interest to, for example, be able to identify the degree of nonlinearity of a signal, but also the nature of the nonlinearity. For the purpose of identifying nonlinearity in general, the algorithms used employ saturation type nonlinearities such as $\tanh$ or $\exp$, however, this only helps to identify nonlinear ARMA (NARMA) models ($\tanh$) and exponential models ($\exp$).

One form of nonlinearity which occurs in many different applications is sparsity; sparse environments are those where the impulse response of an unknown system has a large number of zero elements and only relatively few active ones. Sparse systems occur naturally within many real-world applications such as acoustics, seismics and chemical processes. One field where algorithms for sparse systems have received much attention is in network echo cancellers, where the typical impulse response of the echo path is extremely sparse. As sparsity occurs commonly in nature and much effort has been put into developing algorithms suited to sparse environments, it is natural that this is one form of nonlinearity that would be of particular interest in terms of signal modality characterisation.

2.1.1 Characterisation in the Complex Domain

Signal representations in the complex domain provide a natural processing platform for spectrum estimation allowing the inclusion of phase components [36], whereas, for example, in digital communications the symbols are made complex by design. Directional processes such as radar or sonar are regularly represented by intensity (amplitude) and direction (phase) components, resulting in multidimensional solutions with benefits over real domain solutions [60]. As analysis in the complex domain is a standard in statistical signal processing, signal modality characterisation also needs to be extended to the complex domain.

Characterisation in the complex domain $\mathbb{C}$ is not a straightforward extension of characterisation in the real domain $\mathbb{R}$. For instance, the nature of nonlinearity in $\mathbb{C}$ is fundamentally different from that in $\mathbb{R}$. Whilst being able to identify the nonlinearity of a signal is important and can give key knowledge about the signal under observation, within nonlinear adaptive filtering in $\mathbb{C}$ one of the biggest problems is the choice of

nonlinear complex activation function (AF). There are three main methods to deal with this:- processing the real and imaginary components separately using a real nonlinearity; processing in the complex domain using a so called “split-complex” nonlinearity; or using a so called “fully-complex” nonlinearity.

A fully-complex nonlinearity is a function $f : \mathbb{C} \rightarrow \mathbb{C}$ and these type of functions are the most efficient in using higher order statistics within a signal [45]. For a split-complex function the real and imaginary components of the input are separated and fed through the dual real valued activation function $f_{real}(x) = f_{imag}(x)$, $x \in \mathbb{R}$. A split complex AF can be represented as

$$\Phi_{split}(z) = f_R(z^r) + j f_I(z^i) = u(z^r) + j v(z^i) \quad (2.1)$$

Whilst algorithms using split-complex AFs have been shown to give good results, their reliance on the real and imaginary weight updates being mutually exclusive can lead to problems when calculating their gradients. This results in split-complex activation functions being unsuitable for processing signals where the real and imaginary components of a signal are strongly correlated. Therefore, in the complex domain, on identification of the presence nonlinearity it can also be important to identify the type of nonlinearity, whether split- or fully-complex.

Another key characteristic in $\mathbb{C}$ is circularity (or propriety). Circularity assumes a rotation-invariant distribution, that is $P[\mathbf{z}] = P[\mathbf{z}e^{j\theta}]$, and a second-order circular signal is termed *proper* [70, 73]. A circular complex signal can be sufficiently represented using the standard complex covariance matrix $E[\mathbf{z}\mathbf{z}^H]$. However, circular data occurs only in a limited number of applications and to accurately model the generality of complex signals we need to take into account not only the usually defined covariance matrix but also the pseudocovariance matrix. For any complex random vector $\mathbf{z} \in \mathbb{C}^m$ with $E\{\mathbf{z}\} = 0$, the two covariance matrices are defined as [101]

$$\mathcal{C}_{\mathbf{z}\mathbf{z}} = E\{\mathbf{z}\mathbf{z}^H\}, \quad \mathcal{P}_{\mathbf{z}\mathbf{z}} = E\{\mathbf{z}\mathbf{z}^T\}. \quad (2.2)$$

where $(\cdot)^H$ is the Hermitian (complex conjugate) transpose and $\mathcal{C}_{\mathbf{z}\mathbf{z}}$ and $\mathcal{P}_{\mathbf{z}\mathbf{z}}$ are the *covariance matrix* and *pseudocovariance matrix* respectively. For a proper signal the pseudocovariance matrix $\mathcal{P}_{\mathbf{z}\mathbf{z}} = 0$,¹. Propriety, on the other hand, is a property of second order statistics and only indicates a zero pseudocovariance matrix [86].

For accurate modelling of complex signals so called augmented complex statistics take into account not only the usually defined covariance matrix $\mathcal{C}_{\mathbf{z}\mathbf{z}}$ but also the pseudocovariance matrix $\mathcal{P}_{\mathbf{z}\mathbf{z}}$ (2.2). By defining an augmented complex vector $\mathbf{z}_a$ as

$$\mathbf{z}_a = \begin{bmatrix} \mathbf{z} \\ \mathbf{z}^* \end{bmatrix}, \quad (2.3)$$

then the augmented covariance matrix is given by [89]

$$\mathcal{R}_{\mathbf{z}\mathbf{z}} = \begin{bmatrix} \mathcal{C}_{\mathbf{z}\mathbf{z}} & \mathcal{P}_{\mathbf{z}\mathbf{z}} \\ \mathcal{P}_{\mathbf{z}\mathbf{z}}^* & \mathcal{C}_{\mathbf{z}\mathbf{z}}^* \end{bmatrix}, \quad (2.4)$$

and contains information from both the covariance and pseudocovariance matrices of $\mathbf{z}$. This enables statistical signal processing of both second order circular and noncircular signals. The modelling of noncircular data based on only the covariance matrix generally² leads to undermodelling [69, 70] that is the system model has been generated based on too few model parameters. On the other hand, the modelling of proper complex processes based on augmented statistics is accurate but involves additional computational complexity.

2.2 Existing Methods for Signal Characterisation

The existing methods for signal modality characterisation can largely be split into two main categories. The majority are offline methods based upon hypothesis testing, these describe the signal in a statistical manner. These methods provide excellent results,

¹Note, circularity is a wider notion than propriety, as it is a property of the probability density function (rotation-invariance).

²Unless we have a standard autoregressive process driven by a doubly white noise with different powers of the real and imaginary part, as shown in Section 6.3.

but there is also a need for online approaches to signal characterisation which can not only identify the nature of a signal but also track any changes in signal modality. The existing online approaches use parametric models to effectively model the system in order to provide a true indication of the nature of a signal.

2.2.1 Offline Statistical Tests

Many offline methods for detecting the nonlinear structure within a signal have been proposed, such as the surrogate data method [96]. The surrogate data method has been extensively used in the context of statistical nonlinearity testing. A surrogate time series, is a realisation of a ‘composite’ null hypothesis. For nonlinearity testing the null hypothesis is that the original signal is linear. Then, a discriminating statistic is calculated for both the original time series and a set of surrogate data. If the statistics for the original time series do not lie in the range of those for the surrogate data, the null hypothesis is rejected, and the original data is judged to be nonlinear, otherwise, it is judged to be linear [83, 96]. In order to decrease the spurious rejections of the null hypothesis, several modified methods for the generation of surrogate data have been proposed. Commonly used is the iterative amplitude adjusted Fourier Transform (iAAFT) surrogate method [83]. The iAAFT surrogate data have their amplitude spectra similar and their amplitude distribution identical to those of the original time series. There exist many discriminating statistics, the commonly used ones include the third order autocovariance (C3) and the asymmetry due to time reversal (REV) [84], deterministic versus stochastic (DVS) plot [104], or δ - ε method [43].

One method which is straightforward to visualise, and which makes use of notions from nonlinear dynamics and chaos theory, e.g., embedding dimension and phase space, is the recently proposed delay vector variance (DVV) method [31]. The DVV method examines simultaneously the determinism and nonlinearity within a signal. The DVV algorithm is summarised in Fig. 2.1. For a predictable signal, the principle of the DVV method is:- if two delay vectors (DVs) lie close to one another in terms of their Euclidean distance, they should also have similar targets. The smaller the Euclidean distance between them, the more similar targets they have, for further details refer

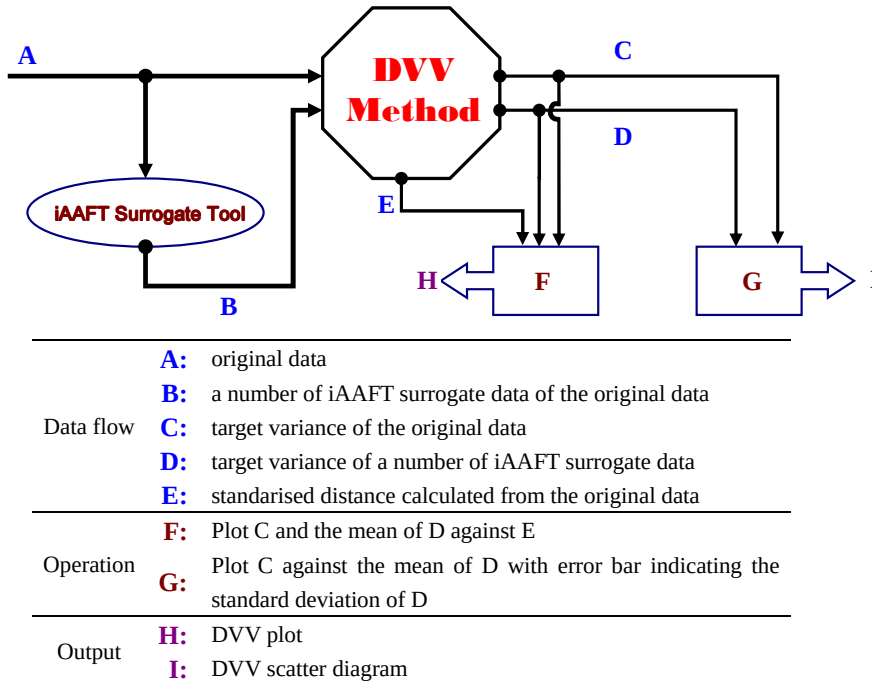


Figure 2.1: Block diagram of DVV method [16].

to [31, 58]. The DVV method has been shown to provide excellent results for the detection of the presence of nonlinearity within a signal, the deterministic/stochastic nature of a signal and estimating the amount of noise within a signal. As such the DVV provides an excellent basis for comparison with the online methods.

2.2.2 Measures of Sparsity

As nonlinearity covers a range of signals and characteristics and as such is a large topic, for accurate characterisation, measures of specific types of nonlinearity are also required. As discussed previously, focussing specifically on sparsity as an example of nonlinearity, standard measures of sparseness are now considered.

The idea of sparseness is becoming more commonly used, however, ways of defining and measuring the sparsity of a signal are not as clear. When considering measures of sparsity, different types of sparseness need to be accounted for as signals may be sparse in some domains whilst not in others. Figures 2.2 and 2.3 illustrate this problem, Figs. 2.2(a) and 2.2(c) show a signal sparse in time but not in frequency. Figures 2.2(b) and 2.2(d) show a signal sparse in frequency but not in time. Figures 2.3(a) and 2.3(b)

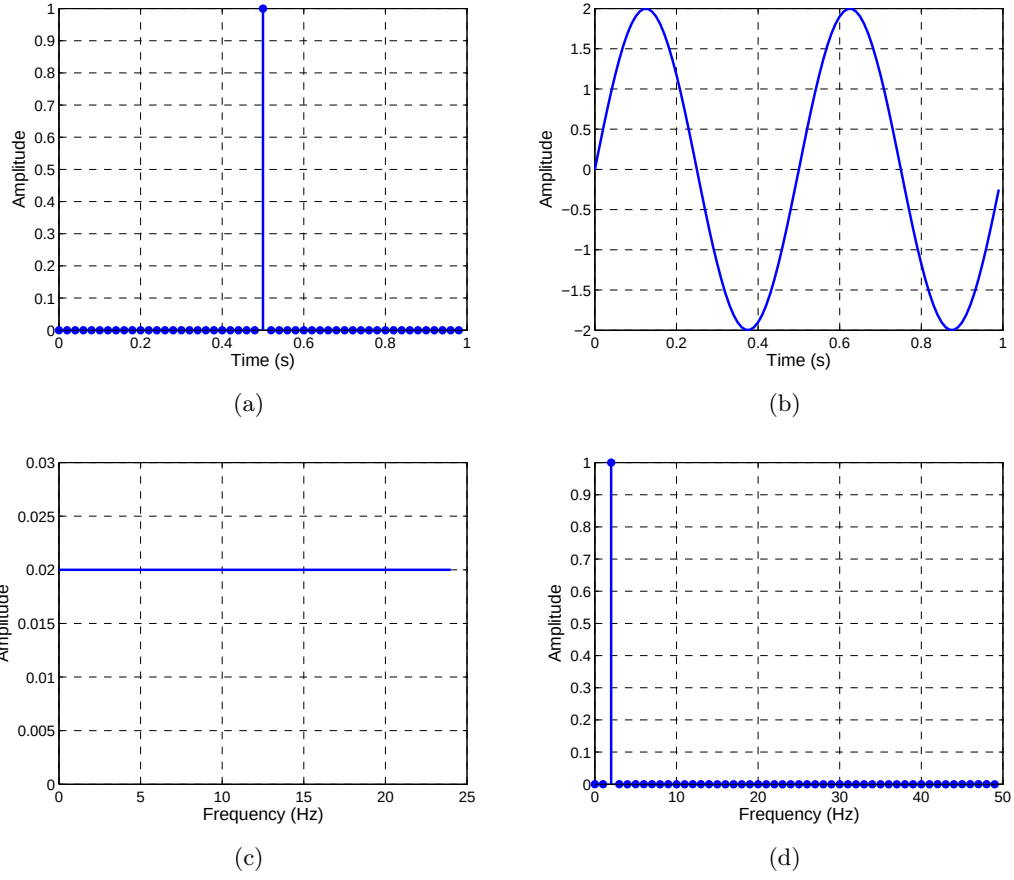


Figure 2.2: Sparsity in time and frequency domains

show a signal which is neither sparse in time or frequency but is sparse in the time-frequency domain, Fig. 2.3(c).

There are a number of possible measures of sparsity including those based on the kurtosis, tanh functions and $\ell^p, p = 1, \dots, N$ norms [44] with the ℓ^1 norm

$$\|x\|_1 = \sum_{i=1}^N |x_i|, \quad (2.5)$$

often being used as unlike the ℓ^0 norm which provides an accurate count of the non zero elements, the ℓ^1 norm is less affected by small changes in the values of x making it robust to the presence of noise.

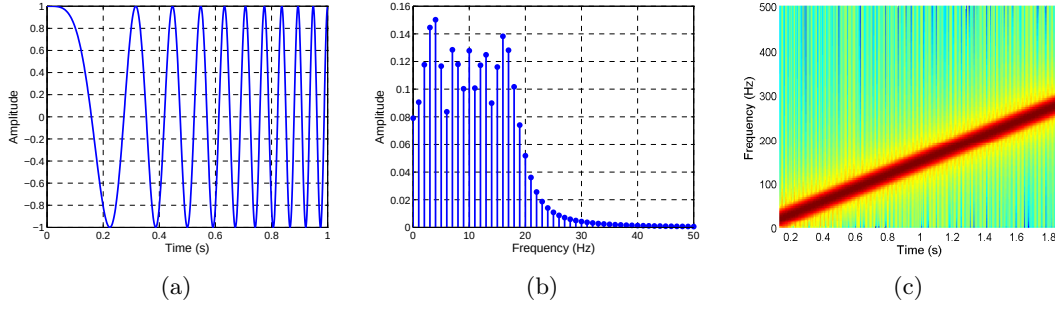


Figure 2.3: Sparsity in time-frequency domain.

2.2.3 Noncircularity Measures

Measures for the degree of noncircularity/impropriety are generally given as functions of the circularity coefficients k_i . To find the circularity coefficients consider the singular value decomposition (SVD) of the coherence matrix

$$\mathbf{M} = \mathbf{C}_{\mathbf{zz}}^{-1/2} \mathbf{P}_{\mathbf{zz}} \mathbf{C}_{\mathbf{zz}}^{-T/2} \quad (2.6)$$

$$= \mathbf{F} \mathbf{K} \mathbf{F}^T, \quad (2.7)$$

where $\mathbf{F}$ is a unitary matrix and $\mathbf{K} = \text{diag}[k_1, k_2, \dots, k_N]$ contains the canonical correlations (circularity coefficients). Following the approaches for measures of multivariate association in real valued vectors, measures of noncircularity given as functions of the canonical correlations have been proposed in [87]

$$d_1 = 1 - \prod_{i=1}^n (1 - k_i^2) = 1 - \det \mathcal{R}_{\mathbf{zz}} \det^{-2} \mathbf{C}_{\mathbf{zz}} \quad (2.8)$$

$$d_2 = \prod_{i=1}^n k_i^2 = \det(\mathbf{P}_{\mathbf{zz}} \mathbf{C}_{\mathbf{zz}}^{-*} \mathbf{P}_{\mathbf{zz}}^*) \det^{-1} \mathbf{C}_{\mathbf{zz}} \quad (2.9)$$

$$d_3 = \frac{1}{n} \sum_{i=1}^n k_i^2 = \frac{1}{n} \text{tr}(\mathbf{C}_{\mathbf{zz}}^{-1} \mathbf{P}_{\mathbf{zz}} \mathbf{C}_{\mathbf{zz}}^{-*} \mathbf{P}_{\mathbf{zz}}^*), \quad (2.10)$$

where $\mathcal{R}_{\mathbf{zz}}$ is the augmented correlation matrix $\mathcal{R}_{\mathbf{zz}} = \begin{bmatrix} \mathbf{C}_{\mathbf{zz}} & \mathbf{P}_{\mathbf{zz}} \\ \mathbf{P}_{\mathbf{zz}}^* & \mathbf{C}_{\mathbf{zz}}^* \end{bmatrix}$. The measure d_1 is a measure of the linear dependence between $\mathbf{z}$ and $\mathbf{z}^*$ and can therefore be extended to give an adaptive test based upon likelihood measures. One such example is the

generalised likelihood ratio test (GLRT) [88] given by

$$\ell = \frac{\det \hat{\mathcal{R}}_{\mathbf{zz}}}{(\det \hat{\mathcal{C}}_{\mathbf{zz}})^2}, \quad (2.11)$$

where $\hat{\mathcal{R}}_{\mathbf{zz}}$ and $\hat{\mathcal{C}}_{\mathbf{zz}}$ are the maximum likelihood (ML) estimates of the augmented covariance matrix $\mathcal{R}_{\mathbf{zz}}$ and covariance matrix $\mathcal{C}_{\mathbf{zz}}$.

2.2.4 Online Model Based Tests

Online methods alternatively provide less statistically rigorous results, this is due to difficulties in accurately estimating statistical measures online. However, online methods do allow for tracking in the changes in signal modality. One intuitive method to determine the nature of a signal has been to present the signal as input to two adaptive filters with different characteristics. By comparing the responses of each filter, this can be used to identify the nature of the signal. The approach in [67] has been shown to successfully track the degree of nonlinearity of a signal using adaptive algorithms. Figure 2.4 shows an implementation of this method which uses a third order Volterra filter (nonlinear subfilter) and a linear subfilter trained by the normalised least mean square (NLMS) algorithm with a step size $\mu = 0.008$ to update the system parameters. The system input and output can be described by

$$u[n] = \sum_{i=0}^I a_i x[n-i] \text{ where } I = 2 \text{ and } a_0 = 0.5, a_1 = 0.25, a_2 = 0.125 \quad (2.12)$$

$$y[n] = F(u[n]; n) + \eta[n] \quad (2.13)$$

where $x[n]$ are independent identically distributed (i.i.d.) and uniformly distributed over the range $[-0.5, 0.5]$ and $\eta[n] \sim \mathcal{N}(0, 0.0026)$. The function $F(u[n]; n)$ varies with the range of n , as follows

$$F(u[n]; n) = \begin{cases} u^3[n], & \text{for } 10000 < n \leq 20000 \\ u^2[n], & \text{for } 30000 < n \leq 40000 \\ u[n], & \text{at all other times} \end{cases} \quad (2.14)$$

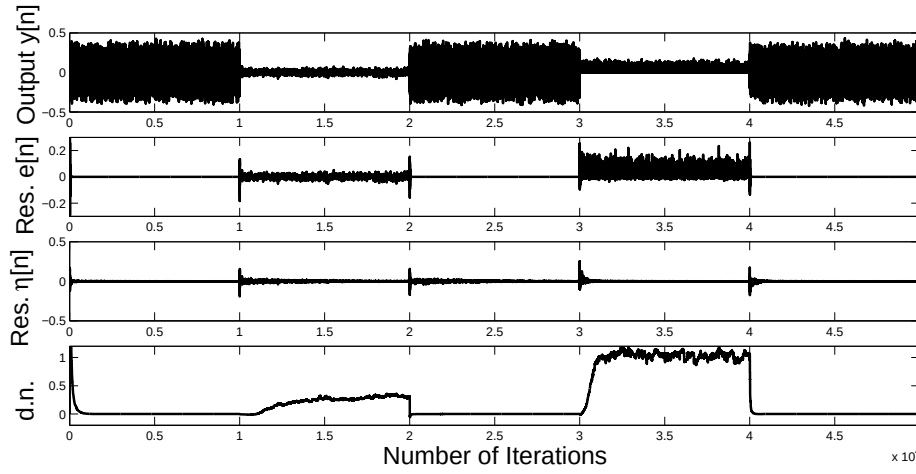


Figure 2.4: Estimated degree of signal nonlinearity for an input signal alternating from linear to nonlinear.

The signal $y[k]$ can be seen in the first trace of Fig. 2.4. The second and third traces show the residual estimation errors of the optimal linear system and Volterra system respectively; the final trace is the estimated degree of system nonlinearity. These results show that this approach can detect changes in nonlinearity and is not affected by the presence of noise.

2.3 Conclusions

Existing methods of signal modality characterisation have been presented. These include offline measures for different signal properties and an online method of tracking nonlinearity. Whilst the results of the online method presented here show this method successfully tracking nonlinearity, this may be due to the type of nonlinearity used. This type of method is dependent on the nature of the nonlinearity under observation being suited to the signal model being used. In real world situations it is not always possible to know the nonlinearity in advance, making their application somewhat limited. Hence to overcome these limitations in the following chapters a much more flexible method based on collaborative adaptive filtering; by means of hybrid filters is proposed.

Chapter 3

Hybrid Filters for Signal Modality Characterisation

3.1 Introduction

Hybrid filters develop on the well established tracking capabilities of adaptive filters by using combinations of adaptive subfilters in a natural way to produce a single hybrid filter without the need for any knowledge of underlying signal generation models. Hybrid filters consist of multiple individual adaptive subfilters operating in parallel and all feeding into a mixing algorithm which produces the single output of the filter [7, 46]. This mixing algorithm is also adaptive and combines the output of each of the subfilters based on their instantaneous output error i.e. the estimate of their current performance on the input signal. For signal modality characterisation a combination of two subfilters is used, this gives a basis for comparing the performance of the filters in an adaptive manner. Combinations of two adaptive filters is a commonly adopted structure for hybrid filters, with the two main forms of mixing algorithms being affine combinations [13] and convex combinations [6]. As the purpose of this method is to observe the mixing parameter within such a structure, the convex combination is used as it provides a bounded value of mixing parameter.

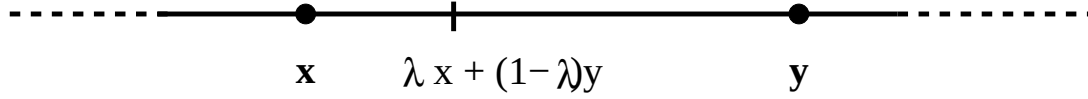


Figure 3.1: Convexity

Convexity can be described as [20]

$$\lambda x + (1 - \lambda)y \text{ where } \lambda \in [0, 1]. \quad (3.1)$$

For x and y being two points on a line, as shown in Fig. 3.1, the resultant convex mixture (3.1) will lie on the same line between x and y . By using a convex combination of subfilters with different properties, it is possible to produce hybrid filters which can outperform either of the constituent subfilters. For convex mixing of the outputs of adaptive filters, it is intuitively clear that initially λ will adapt to favour the filter with the faster learning rate and following convergence it will favour the filter with better steady state properties¹. Thus, hybrid filters using convex combinations of two subfilters both being trained by the same algorithm, with different learning rates, have been shown to perform well in stationary environments and always perform at least as well as the better of the two subfilters in the steady state [4]. Alternatively by using subfilters trained by different algorithms, to take advantage of the desired properties of each subfilter, a hybrid filter can be produced which performs well in a wide range of environments for a variety of different input signals.

Taking this into account, first the hybrid filtering structure is derived and the performance of the hybrid filter in terms of the one step ahead prediction error is outlined. Following this the use of the hybrid filter to identify and track nonlinearity is illustrated, with simulations on benchmark synthetic data supporting the approach.

¹Unlike traditional search then converge approaches this method allows for potentially nonstationary data.

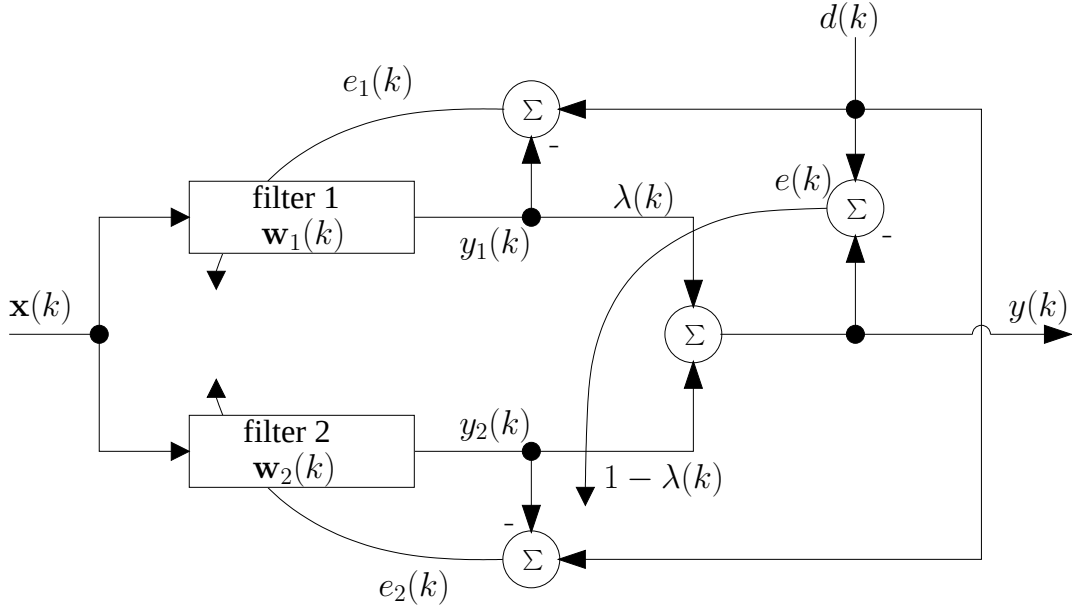


Figure 3.2: Hybrid filter structure

3.2 Hybrid Filter Structure

The hybrid filter, shown in Figure 3.2, consists of two subfilters, each being adapted independently, with a convex combination of the two filters then taken as the output of the hybrid filter. The two subfilters within the hybrid filtering architecture operate in the prediction setting, sharing the common input vector $\mathbf{x}(k) = [x(k-1), \dots, x(k-N)]^T$ for filters of length N . The outputs of the two subfilters are dependent on the algorithms used to train the subfilters, and are given by $y_1(k) = \mathbf{x}^T(k)\mathbf{w}_1(k)$ and $y_2(k) = \mathbf{x}^T(k)\mathbf{w}_2(k)$. The corresponding weight vectors are $\mathbf{w}_1(k) = [w_{1,1}(k), \dots, w_{1,N}(k)]^T$ and $\mathbf{w}_2(k) = [w_{2,1}(k), \dots, w_{2,N}(k)]^T$; to preserve the inherent characteristics, each subfilter is updated by its own error $e_1(k)$ and $e_2(k)$, using a common desired signal $d(k)$. The convex combination of the subfilter outputs $y_1(k)$ and $y_2(k)$ forms the overall system output $y(k)$, given by

$$y(k) = \lambda(k)y_1(k) + (1 - \lambda(k))y_2(k), \quad (3.2)$$

where $\lambda(k)$ is the mixing parameter, which is made adaptive, and is updated by minimising the cost function

$$\mathcal{J}(k) = \frac{1}{2}|e(k)|^2 = \frac{1}{2}|d(k) - y(k)|^2. \quad (3.3)$$

To obtain the update for $\lambda(k)$, approaches such as using a stochastic gradient based adaptation have been proposed [9]. In this case, a least mean square (LMS) based adaptation is used, whereby

$$\lambda(k+1) = \lambda(k) - \mu_\lambda \nabla_\lambda \mathcal{J}(k)|_{\lambda=\lambda(k)}, \quad (3.4)$$

and μ_λ is the step size. From (3.2) and (3.4), the λ update can be shown to be

$$\begin{aligned} \lambda(k+1) &= \lambda(k) - \frac{\mu_\lambda}{2} \frac{\partial e^2(k)}{\partial \lambda(k)} \\ &= \lambda(k) + \mu_\lambda e(k) (y_1(k) - y_2(k)). \end{aligned} \quad (3.5)$$

Unlike, for example, affine combinations [13], using a convex combination provides a distinct range of values for the mixing parameter. However, for stability and as the basis of using this structure for a test for signal characterisation, it is key that the mixing parameter λ remain in the range $[0, 1]$. Existing algorithms have used a nonlinear sigmoid function to not only ensure the mixing parameter remains in the required range but also to reduce gradient noise as the update of the mixing parameter approaches these limits [6, 111]. However, as this method relies on observing the behaviour of the mixing parameter λ , this is not a valid option as it will artificially change the response of the mixing parameter. Instead a hard limit is placed on the values of λ can take.

3.2.1 Convergence and Computational Complexity

Hybrid filters have also been shown to improve the overall stability of learning, as should one of the subfilters fail to converge the values of λ adapt such that the hybrid filter follows the stable subfilter [6]. The computational complexity of the training of

the hybrid filter is naturally a combination of the computational complexities of the algorithms used to update the constituent subfilters. The additional complexity required for the update of the mixing parameter λ is minimal, requiring only an additional 4 multiplications and 5 additions, and only becomes relevant if the overall weight update $\mathbf{w}(k) = \lambda(k)\mathbf{w}_1(k) + (1 - \lambda(k))\mathbf{w}_2(k)$ is also of interest.

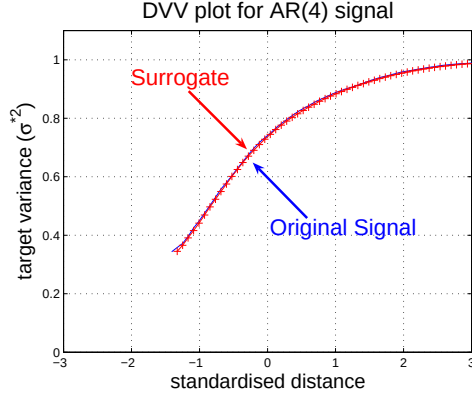
3.3 Performance of Hybrid Filters

It has been shown that convex combinations of subfilters with different characteristics in terms of their responses to input signals can outperform the constituent adaptive filters for the generality of signals [53]. To illustrate this, a convex combination of subfilters trained by the NLMS and normalised nonlinear gradient descent (NNGD) algorithms has been implemented. The aim of using this combination is to take advantage of the stability and excellent steady state properties of the NLMS and the faster convergence and better tracking capabilities of the NNGD. The output of the subfilter trained by the NLMS y_{NLMS} is given by [106]

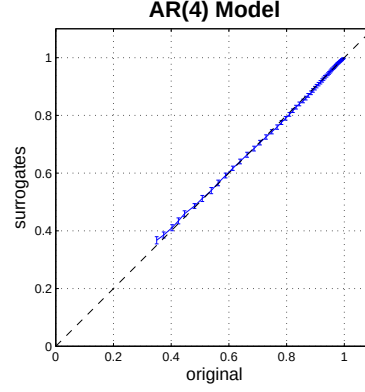
$$\begin{aligned} y_{NLMS}(k) &= \mathbf{x}^T(k) \mathbf{w}_{NLMS}(k) \\ e_{NLMS}(k) &= d(k) - y_{NLMS}(k) \\ \mathbf{w}_{NLMS}(k+1) &= \mathbf{w}_{NLMS}(k) + \frac{\mu_{NLMS}}{\|\mathbf{x}(k)\|_2^2 + \varepsilon} e_{NLMS}(k) \mathbf{x}(k) \end{aligned} \quad (3.6)$$

and the output of the NNGD trained subfilter y_{NNGD} is given by [55]

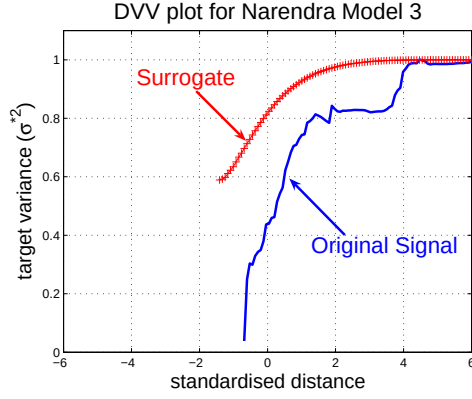
$$\begin{aligned} y_{NNGD}(k) &= \Phi(net(k)), \\ net(k) &= \mathbf{x}^T(k) \mathbf{w}_{NNGD}(k), \\ e_{NNGD}(k) &= d(k) - y_{NNGD}(k), \\ \mathbf{w}_{NNGD}(k+1) &= \mathbf{w}_{NNGD}(k) + \eta(k) \Phi'(net(k)) e_{NNGD}(k) \mathbf{x}(k), \\ \eta(k) &= \frac{\mu}{C + [\Phi'(net(k))] \|x(k)\|_2^2}, \end{aligned} \quad (3.7)$$



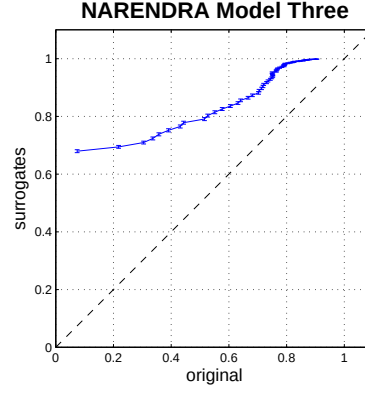
(a) AR(4) DVV plot.



(b) AR(4) DVV scatter diagram.



(c) Nonlinear signal (3.9) DVV plot.



(d) Nonlinear signal (3.9) DVV scatter diagram.

Figure 3.3: DVV plots and scatter diagrams for linear and nonlinear signals [17].

where the step-size parameter of the NLMS filter is μ_{NLMS} and ε is the regularisation term. In the case of the NNGD, μ_{NNGD} is the step-size, $\Phi(\cdot)$ represents a nonlinear activation function and C a constant representing the ignored higher terms, for simulation purposes these were $\tanh(\cdot)$ and unity respectively, and in all cases $\mu_{NLMS} = \mu_{NNGD}$.

The performance of the hybrid filter was evaluated over a set of 100 independent simulation runs and the average of the squared prediction error taken. To illustrate the performance in different settings the hybrid filter was presented with a linear autoregressive AR(4) process given by

$$x(k) = 1.79x(k-1) - 1.85x(k-2) + 1.27x(k-3) - 0.41x(k-4) + n(k), \quad (3.8)$$

and a benchmark nonlinear signal [68] given by

$$x(k) = \frac{x(k-1)}{1 + x^2(k-1)} + n^3(k), \quad (3.9)$$

where $n(k)$ is a zero mean, unit variance white Gaussian process. To illustrate the input signals have the desired linear and nonlinear natures the DVV method described in the previous chapter was used to assess the signals (3.8) and (3.9), the results of which are shown in Fig. 3.3. Figures 3.3(a) and 3.3(c) show the DVV plots of the target variance as a function of standardised distance and Fig. 3.3(b) and 3.3(d) show the DVV scatter plots of the target variance of the original data against the mean of the target variances of the surrogate data. The results shown here are in line with those obtained from the hybrid filter. For the AR(4) signal the DVV scatter diagram lies on the bisector line, indicating its linear nature. Whereas the corresponding plot for the nonlinear signal deviates from the bisector line, indicating its nonlinear nature.

Figure 3.4 shows the convergence curves for the NLMS, NNGD and hybrid filter for the linear AR(4) signal (3.8); in this example both subfilters were reasonably closely matched, and the performance of the hybrid filter was in line with that of the subfilter trained by the NLMS algorithm. In the second example, Fig. 3.5 shows the convergence curves for the nonlinear signal given in (3.9). In this case, there is a clear difference in the performance of the two different subfilters, with the NLMS converging faster but to a larger steady state error. This highlights the advantage of using such a combination, as the hybrid filter initially follows the faster convergence of the NLMS before going on to benefit from the smaller steady state error of the NNGD. As the hybrid filter outperforms both constituent subfilters over the two different signals, it illustrates the possibility of using such a scheme in situations where the nature of the input signal is unknown or where the dynamics of the input signal vary over time.

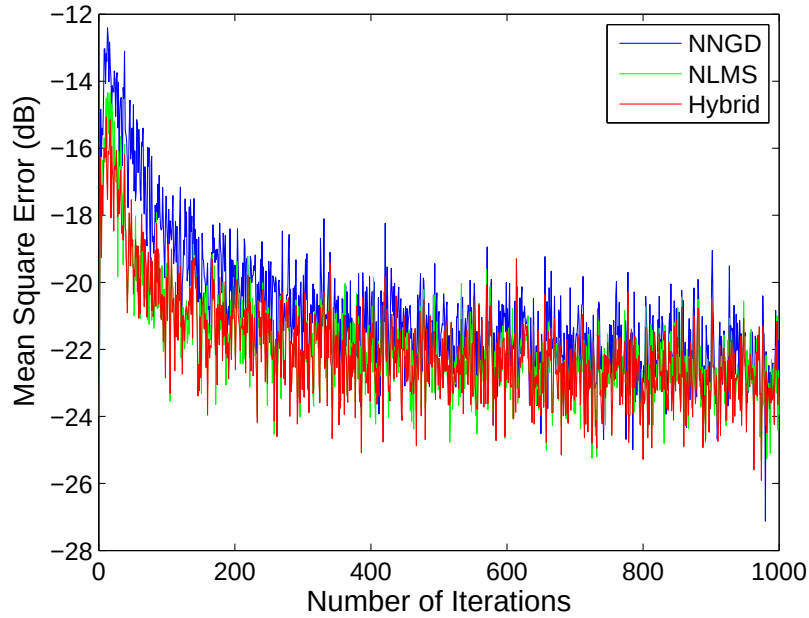


Figure 3.4: Performance of the NLMS, NNGD and corresponding hybrid filter for a linear AR(4) signal.

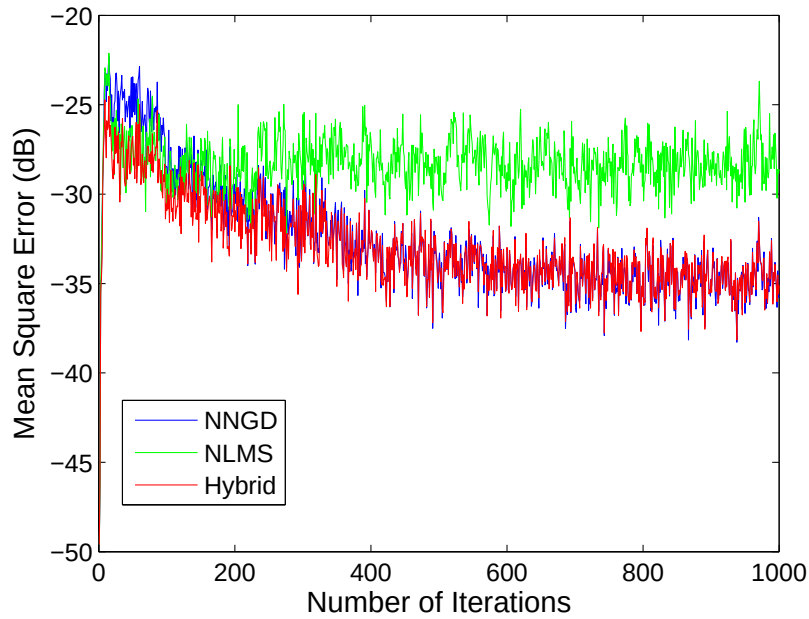


Figure 3.5: Performance of the NLMS, NNGD and corresponding hybrid filter for a benchmark nonlinear signal.

3.4 Hybrid Filters for Identifying & Tracking Nonlinearity

Previous applications of hybrid filters have focused mainly on the improved performance they can offer over the individual constituent filters. However, by understanding that the output of the hybrid filter will always be dominated by the better performing of the two subfilters, it is natural to assume that this information can be obtained by observing the adaptive mixing parameter λ . Therefore, with appropriately chosen subfilters, using algorithms with known different properties, it is intuitive that the behaviour of λ within such a combination will reveal which filter is currently giving the best response to the input signal. In which case, the response of λ can also be used to reveal knowledge about the nature of the input signal.

Figure 3.6 shows the behaviour of the convex mixing parameter λ for the previously described hybrid filter for both the linear AR(4) signal (3.8) and the benchmark nonlinear signal (3.9), where a value of $\lambda = 1$ corresponds to the output of the hybrid filter being dominated by the output of the NNGD trained subfilter and $\lambda = 0$ by the NLMS trained subfilter. As can be seen, for the nonlinear signal the output of the hybrid filter quickly moves to be dominated by the output of the NNGD subfilter and for the linear signal the converse is true. These results are in line with the results from the previous section where the NNGD has a smaller steady error for the nonlinear signal and the NLMS for the linear signal. As expected these results also correspond to the results obtained from the DVV method. By using the convex combination of the NNGD and NLMS, it is clear that the behaviour of λ is significantly different depending on whether the input signal is linear or nonlinear. It is also clear that by observing the behaviour of λ , it is possible to draw conclusions about the nature of the input signal.

Knowing that the behaviour of λ changes dependent on which subfilter is dominating the output of the hybrid filter, it is natural to extend this to consider tracking changes in the nature of the signal. In this case the hybrid filter was presented with an input signal which alternated between linear and nonlinear input signals every 200 samples. Figure 3.7 shows that the value of λ clearly adapts to be dominated by the subfilter

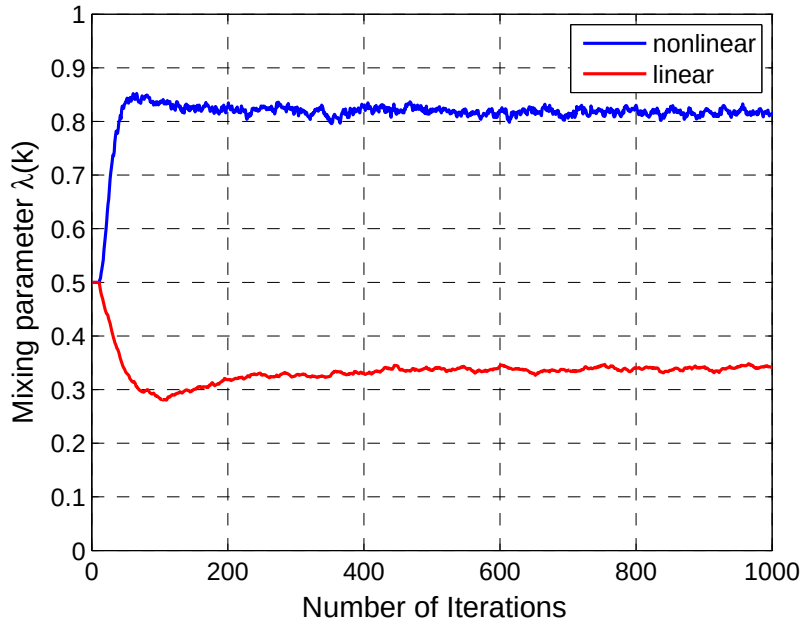


Figure 3.6: Evolution of the mixing parameter λ for a hybrid filter combining the NLMS and NNGD algorithms for linear AR(4) signal and benchmark nonlinear signal.

most suited to the current dynamics of the input signal. To further illustrate this, the period over which the input signal was alternated was decreased to 100 samples, the result of which is shown in Fig. 3.8. At this point the hybrid filter clearly still tracks the changes in the nature of the input signal, however, reducing the period further limits the accuracy. Due to the nature of the filter, the period over which the signal nature can alternate and the filter still be capable of tracking the changes is restricted by the length of the filter, in this case $N = 10$.

The previous examples show changes in the nature of the signal from one extreme to another. To ascertain whether these changes are related to the nature of the input signal or are brought about by a change in the response of the filter because the input signal has been changed, a second linear and nonlinear signal are introduced. The linear signal is given by an AR(1) process

$$x(k) = 0.9x(k-1) + n(k), \quad (3.10)$$

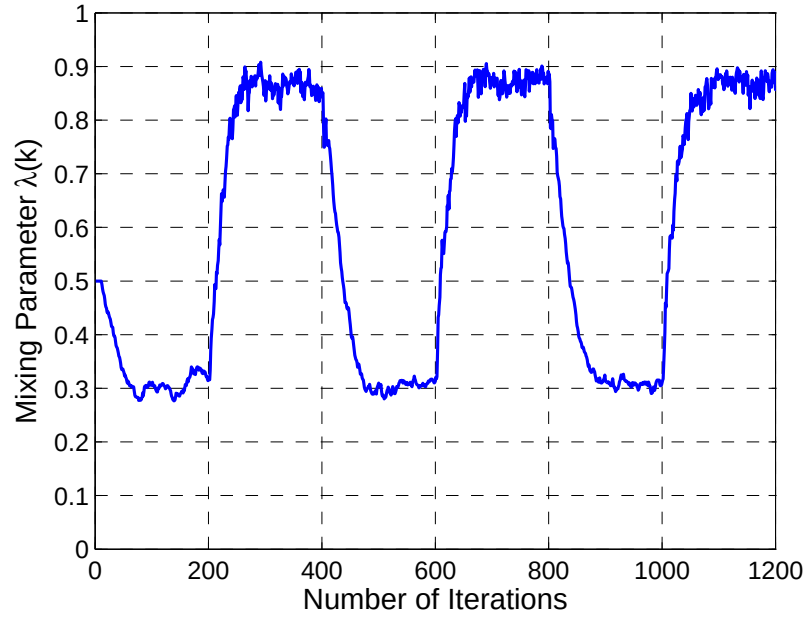


Figure 3.7: Evolution of the mixing parameter λ for a hybrid filter combining the NLMS and NNGD algorithms for an input signal which alternates from a linear AR(4) signal to a nonlinear signal every 200 samples.

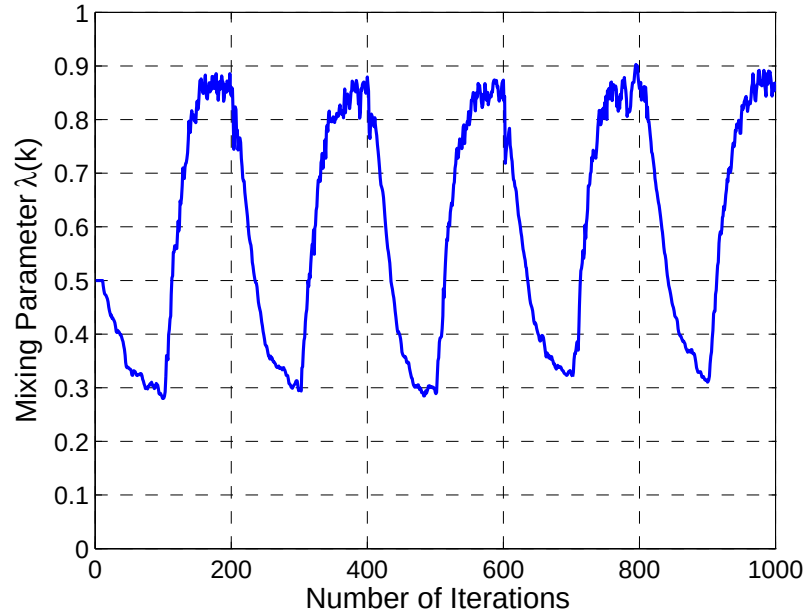


Figure 3.8: Evolution of the mixing parameter λ for a hybrid filter combining the NLMS and NNGD algorithms for an input signal which alternates from a linear AR(4) signal to a nonlinear signal every 100 samples.

and the nonlinear signal is given by [68]

$$x(k) = \frac{x^2(k-1)(x(k-1) + 2.5)}{1 + x^2(k-1) + x^2(k-2)} + n(k-1). \quad (3.11)$$

By presenting the hybrid filter with an input signal which alternates every 200 samples from the AR(4) signal (3.8) to AR(1) (3.10), followed by each of the benchmark nonlinear signals (3.9) and (3.11) in turn, it is possible observe changes in λ when the input signal changes but the nature of the signal does not. Figure 3.9 shows that when the input signal changes between the two AR signals there is a slight change in the response of λ but overall the hybrid filter is still dominated by the subfilter trained by the NLMS. Whereas for the nonlinear signals there is little change in the response of λ when the input signal changes. In this case it is clear that the only significant changes occur when the nature of the input signal changes from linear to nonlinear. In the situation where after a change neither constituent filter is well suited for the current input signal, the hybrid filter will still perform in line with the subfilter with the better performance and hence the mixing parameter will still indicate this. In certain situations the hybrid filter will use a combination of the outputs of each subfilter to give a smaller error than either constituent filter. In these situations the mixing parameter will remain in the mid range, giving no conclusive result about which subfilter is dominating.

3.5 Conclusions

A method for the tracking of signal modality in an online manner has been introduced. This is achieved by a convex combination of two adaptive filters for which the transient responses are significantly different. The improvement in performance obtained from using a hybrid filtering architecture has previously been described and an implementation of such a structure has been illustrated. However, by training the two filters with different algorithms, it is possible to exploit the difference in the performance capabilities of each. The evolution of the adaptive convex mixing parameter λ , helps determine which filter is more suited to the current input signal dynamics, and thereby

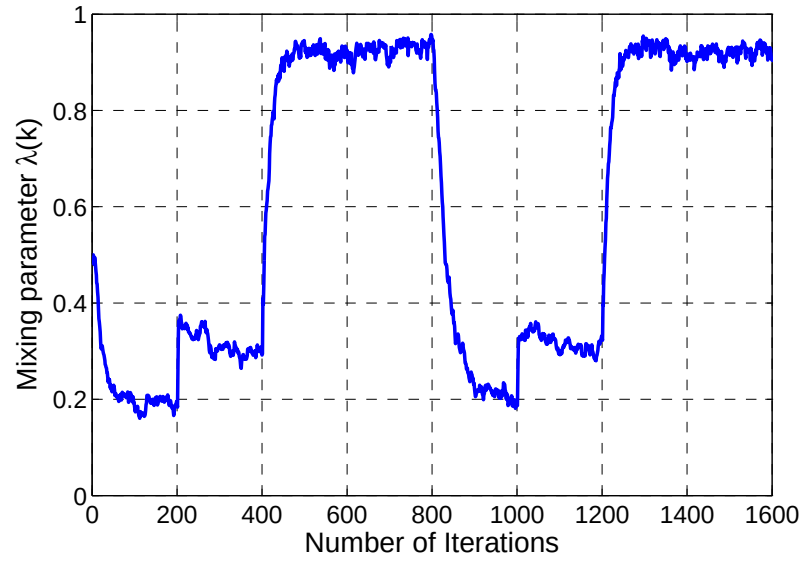


Figure 3.9: Evolution of the mixing parameter λ for a hybrid filter combining NLMS and NNGD subfilters for an input signal alternating from linear AR(4) to AR(1) to nonlinear every 200 samples.

gain information about the nature of the signal. The algorithms used to train the subfilters of the hybrid filter can be selected dependant on the characteristics of the input signals which of interest. The analysis and simulations illustrate that there is significant potential for the use of this method for online tracking of some fundamental properties of the input signal, and that the proposed approach is capable of tracking the nonlinearity within synthesized data.

Chapter 4

Sparse Adaptive Filters

4.1 Introduction

The previous chapter addressed identification and tracking of nonlinearity in general using a saturation type nonlinearity. However, as previously stated, in order to build a more complete picture of the nature of the signal it is also of interest to be able to identify not only whether the signal is linear or nonlinear but also to identify different types of nonlinearity, with one such example of a subset of nonlinearity being sparsity. In sparse settings standard algorithms, such as the LMS and NLMS, are known to perform in a suboptimal manner [26, 62]. In order to facilitate identification of signals which are sparse in nature, this chapter looks at adaptive filters designed specifically for sparse environments.

One algorithm in particular, the proportionate NLMS (PNLMS) algorithm [26], is commonly used in network echo cancellation, where the impulse response is particularly sparse. The PNLMS takes advantage of the fact that the impulse response is sparse to give an update which is proportional relative to the size of the filter coefficients. The PNLMS in its original form has been introduced based on empirical evidence and subsequently most of its variants have also been designed in a similar manner. Since its introduction, a number of modifications have been proposed such as the the PNLMS++ [32] which uses both the PNLMS and NLMS updates. The MPNLMS [22, 23] which employs the μ -law (commonly used in compression for telecommunications), within the

filter coefficient update. The improved PNLMS (IPNLMS) [11] which focuses on resolving problems with the PNLMS when the impulse response of the system is dispersive. However, the stability of the algorithm has received little attention, leading to the PNLMS still being constrained to the stability limits of the NLMS [26, 71]. The existing approaches such as Doroslovački and Deng’s analysis [24], provide convergence proofs based on existing algorithms, such as steepest descent, and are not explicitly derived for PNLMS. It is therefore beneficial before moving on to consider the performance of the algorithms to provide the derivation and analysis of the class of PNLMS algorithms based on a unified theoretical platform.

From this platform, the results are then extended to provide adaptive step-size variants of the PNLMS algorithms. The selection of learning rate is critical to the performance of all LMS-type algorithms. Ideally, we desire an algorithm for which the speed of convergence is fast and the steady state error is small when operating in a stationary environment. In a nonstationary environment the algorithm should change the learning rate according to the dynamics of the input signal, to achieve as good a tracking performance as possible. A convenient way to improve the convergence of linear adaptive filters in nonstationary environments is to introduce a gradient adaptive step-size (GASS). Approaches with a “linear” gradient adaptive learning rate based on $\partial\mathcal{J}/\partial\mu$ include the algorithms by Benveniste et al. [12], Mathews and Xie [64], and Ang and Farhang [3]. Alternatively, a “nonlinear” gradient adaptive step-size based on $\partial\mathcal{J}/\partial\varepsilon$, where ε is the NLMS regularisation parameter in (3.6), such as the Generalised Normalised Gradient Descent (GNGD) algorithm [56], can be employed. Variants of the PNLMS with a fixed regularisation parameter ε have been proposed [11, 32] but this does not guarantee optimal performance in statistically nonstationary environments. Both approaches are considered in relation to improving the performance of the PNLMS algorithm. To complete the analysis the performance of the PNLMS and its adaptive step-size variants are illustrated on benchmark sparse systems.

4.2 Derivation of the Class of PNLMS Algorithms

Originally Duttweiler [26] introduced the PNLMS algorithm as a version of NLMS based on the NLMS update, the proportional update of the PNLMS is achieved by introducing a diagonal “tap selection matrix” $\mathbf{G}(k)$ within the coefficient update (3.6), giving

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu \frac{\mathbf{G}(k)e(k)\mathbf{x}(k)}{\|\mathbf{x}(k)\|_2^2}, \quad (4.1)$$

where

$$\mathbf{G}(k) = \text{diag}[g_1(k), \dots, g_N(k)]. \quad (4.2)$$

The diagonal elements of $\mathbf{G}(k)$ define the proportionate amounts that each coefficient is updated by, where the elements $g_n(k)$ are given by

$$\begin{aligned} \bar{\phi}(k) &= 1/N \sum_{n=1}^N \phi_n(k), \\ \phi_n(k) &= \max \{ \rho \max [\delta, \|\mathbf{w}(k)\|_\infty], |w_n(k)| \}, \\ g_n(k) &= \frac{\phi_n(k)}{\bar{\phi}(k)}, \quad n = 1, \dots, N \end{aligned} \quad (4.3)$$

and $\|\cdot\|_\infty$ is the infinity norm.

However, to unify the analysis it is advantageous to start from the LMS. In the same vein as the Duttweiler result, the LMS is modified to suit sparse environments, thus providing a basis to derive the class of PNLMS algorithms in a generic way. The “proportionate LMS” (PLMS) algorithm is therefore given by

$$\begin{aligned} e(k) &= d(k) - \mathbf{x}^T(k)\mathbf{w}(k), \\ \mathbf{w}(k+1) &= \mathbf{w}(k) + \mu \mathbf{G}(k)e(k)\mathbf{x}(k), \end{aligned} \quad (4.4)$$

that is, the PLMS is equipped with the “tap selective” term $\mathbf{G}(k)$. The use of tap selection also has a geometric justification, as for an N-tap LMS the weight update

lives in $\mathbb{R}^N$ and the direction of that update is dominated by the largest element of the input vector $\mathbf{x}$ [93]. The following sections shall now provide a rigorous derivation of the standard PNLMS, based on the convergence analysis of PLMS in (4.4); this approach is supported by the original analysis by Duttweiler [26] which analyses a form of the PLMS algorithm with fixed gain distributors $\mathbf{G}$.

4.2.1 Convergence in the Mean

Assume without a loss in generality the desired response of the filter can be expressed as [25]

$$d(k) = \mathbf{x}^T(k) \mathbf{w}_o + q(k), \quad (4.5)$$

where $\mathbf{w}_o$ is the optimal weight vector and $q(k)$ is zero mean Gaussian noise with variance σ_q^2 , uncorrelated with $\mathbf{x}(k)$. This allows us to describe the PLMS by

$$\begin{aligned} e(k) &= \mathbf{x}^T(k) \mathbf{w}_o + q(k) - \mathbf{x}^T(k) \mathbf{w}(k), \\ \mathbf{w}(k+1) &= \mathbf{w}(k) + \mu \mathbf{G}(k) \mathbf{x}(k) \mathbf{x}^T(k) \mathbf{w}_o - \mu \mathbf{G}(k) \mathbf{x}(k) \mathbf{x}^T(k) \mathbf{w}(k) + \mu q(k) \mathbf{G}(k) \mathbf{x}(k). \end{aligned} \quad (4.6)$$

Subtract the optimal weight vector $\mathbf{w}_o$ from both sides, to give the weight error vector $\mathbf{v}(k) = \mathbf{w}(k) - \mathbf{w}_o$ expressed as

$$\mathbf{v}(k+1) = \mathbf{v}(k) - \mu \mathbf{G}(k) \mathbf{x}(k) \mathbf{x}^T(k) \mathbf{v}(k) + \mu q(k) \mathbf{G}(k) \mathbf{x}(k). \quad (4.7)$$

Applying the statistical expectation operator and employing the independence assumptions¹ yields

$$\begin{aligned} E[\mathbf{v}(k+1)] &= (\mathbf{I} - \mu \mathbf{G}(k) E[\mathbf{x}(k) \mathbf{x}^T(k)]) E[\mathbf{v}(k)] + \mu \mathbf{G}(k) E[q(k) \mathbf{x}(k)], \\ &= (\mathbf{I} - \mu \mathbf{G}(k) \mathbf{R}) E[\mathbf{v}(k)]. \end{aligned} \quad (4.8)$$

¹Namely that the input signal and filter coefficient vectors are zero mean, stationary, jointly normal and with finite moments; the successive increments of tap weights are independent of one another and the error and input vector sequences are statistically independent of one another [37, 80, 106].

Since the correlation matrix $\mathbf{R} = E[\mathbf{x}\mathbf{x}^T]$ is symmetric and positive semidefinite, it has the following decomposition

$$\mathbf{R} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T, \quad (4.9)$$

where $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N)$ are the eigenvalues and $\mathbf{Q}$ is the orthogonal matrix of the corresponding eigenvectors. A rotation of the weight error vector $\mathbf{v}(k)$ by the eigenmatrix $\mathbf{Q}$, that is, $\mathbf{v}'(k) = \mathbf{Q}^T \mathbf{v}(k)$, gives

$$\mathbf{v}'(k+1) = (\mathbf{I} - \mu \mathbf{G}(k) \mathbf{\Lambda}) \mathbf{v}'(k). \quad (4.10)$$

As $(\mathbf{I} - \mu \mathbf{G}(k) \mathbf{\Lambda})$ is diagonal, every element of $\mathbf{v}'(k)$ evolves independently and converges to zero if for all the eigenvalues $|1 - \mu g_n \lambda_n| < 1$. Therefore, the PLMS converges only if it converges for the maximum mode of convergence, that is, for λ_{max} . Since $\lambda_{max} \leq \sum(\text{diagonal elements of } \mathbf{R}) = \text{tr}[\mathbf{R}]$, the condition for the convergence in the mean of the PLMS becomes

$$0 < \mu < \frac{2}{\text{tr}[\mathbf{G}(k) \mathbf{R}]}. \quad (4.11)$$

Note that the term $\text{tr}[\mathbf{G}(k) \mathbf{R}]$ is equivalent to $\mathbf{x}^T(k) \mathbf{G}(k) \mathbf{x}(k)$, however, for simplicity, for a white i.i.d. input for which $\mathbf{R} = \sigma_x^2 \mathbf{I}$, and using the identity $\text{tr}[AB] \leq \text{tr}[A] \text{tr}[B]$ and $\text{tr}[\mathbf{G}(k)] = N$, results in

$$0 < \mu < \frac{2}{N \sigma_x^2}. \quad (4.12)$$

4.2.2 Convergence in the Mean Square

To converge in the mean square the algorithm must firstly converge in the mean, that is, (4.11) must be satisfied. Based on the output error (4.6), we can arrive at the mean square error by squaring and taking the expectation of both sides, to give

$$E[e^2(k)] = \sigma_q^2 + E[(\mathbf{x}^T(k) \mathbf{v}(k))^2] - 2E[q(k) \mathbf{x}^T(k) \mathbf{v}(k)]. \quad (4.13)$$

Using the independence assumptions [37, 80, 106] and $\mathbf{x}^T(k)\mathbf{v}(k) = \mathbf{v}^T(k)\mathbf{x}(k)$ and defining the weight error vector correlation matrix as $\mathbf{K}(k) = E[\mathbf{v}(k)\mathbf{v}^T(k)]$, results in

$$\begin{aligned}
 E \left[\left(\mathbf{x}^T(k)\mathbf{v}(k) \right)^2 \right] &= \text{tr} \left[E \left[\mathbf{v}^T(k)\mathbf{x}(k)\mathbf{x}^T(k)\mathbf{v}(k) \right] \right] \\
 &= \text{tr} \left[E \left[\mathbf{v}^T(k) E \left[\mathbf{x}(k)\mathbf{x}^T(k) \right] \mathbf{v}(k) \right] \right] \\
 &= E \left[\text{tr} \left[\mathbf{v}^T(k)\mathbf{R}\mathbf{v}(k) \right] \right] \\
 &= E \left[\text{tr} \left[\mathbf{v}(k)\mathbf{v}^T(k)\mathbf{R} \right] \right] \\
 &= \text{tr} \left[E \left[\mathbf{v}(k)\mathbf{v}^T(k) \right] \mathbf{R} \right] \\
 &= \text{tr} \left[\mathbf{R}\mathbf{K}(k) \right].
 \end{aligned} \tag{4.14}$$

Thus, the expected value of the squared error becomes

$$\begin{aligned}
 \xi(k) = E[e^2(k)] &= \xi_{min} + \xi_{EMSE}(k) \\
 &= \sigma_q^2 + \text{tr}[\mathbf{R}\mathbf{K}(k)]
 \end{aligned} \tag{4.15}$$

where ξ_{min} is the minimum mean square error defined by the power of the noise, σ_q^2 , $\xi_{EMSE}(k)$ is the excess mean square error and due to the independence assumptions the final term of (4.13) disappears. The excess mean square error is a result of the filter coefficients fluctuating around their optimum values as they begin to converge. Using again the rotation (4.9) and $\mathbf{K}'(k) = \mathbf{Q}^T\mathbf{K}(k)\mathbf{Q}$ gives

$$\xi(k) = \sigma_q^2 + \text{tr}[\mathbf{K}'(k)\mathbf{\Lambda}]. \tag{4.16}$$

Since matrix $\mathbf{\Lambda}$ is diagonal, this results in

$$\xi(k) = \sigma_q^2 + \sum_{i=1}^N \lambda_i \kappa'_{ii}(k), \tag{4.17}$$

where κ'_{ii} are the diagonal elements of $\mathbf{K}'$.

It is convenient to assess the mean square performance of an algorithm in terms of

the misadjustment

$$M = \frac{\xi_{EMSE}}{\xi_{min}} = \frac{\xi_{EMSE}}{\sigma_q^2}, \quad (4.18)$$

which following the approach from [27, 28] can be shown to be

$$M = \frac{\sum_{i=1}^N \frac{\mu g_i \lambda_i}{2(1-\mu g_i(k) \lambda_i)}}{1 - \sum_{i=1}^N \frac{\mu g_i(k) \lambda_i}{2(1-\mu g_i(k) \lambda_i)}} \quad (4.19)$$

$$= \mu \frac{\frac{1}{2} \text{tr}[\mathbf{G}(k) \mathbf{R}]}{1 - \frac{1}{2} \mu \text{tr}[\mathbf{G}(k) \mathbf{R}]} \quad (4.20)$$

Thus, for a white i.i.d. input, the mean square error $\xi(k)$ converges asymptotically to $\xi(\infty) = \mathcal{J}_{min} = \sigma_q^2$ for

$$0 < \mu < \frac{2}{\text{tr}[\mathbf{G}(k) \mathbf{R}]} = \frac{2}{\sigma_x^2 N}. \quad (4.21)$$

which gives the bound on the step-size of PLMS.

4.2.3 Introducing PNLMS via Normalisation of PLMS

It has been shown that for convergence of PLMS in both the mean and the mean square, the bound on the step-size is given by

$$0 < \mu < \frac{2}{\mathbf{x}^T(k) \mathbf{G}(k) \mathbf{x}(k)}. \quad (4.22)$$

Incorporating this step-size into the PLMS (4.4) gives the normalised PLMS (PNLMS)

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \frac{\eta}{\mathbf{x}^T(k) \mathbf{G}(k) \mathbf{x}(k)} \mathbf{G}(k) e(k) \mathbf{x}(k), \quad (4.23)$$

where $0 < \eta < 2$.

As the expansion of $\mathbf{K}'(k)$ to obtain (4.19) is not straightforward, an alternative method is to recast the PNLMS into the optimisation task performed by NLMS type algorithms, that is, to minimise the *a posteriori* error $d(k) - \mathbf{x}^T(k) \mathbf{w}(k+1)$, as opposed to LMS which minimises the *a priori* error $e(k)$. However, minimisation of the *a posteriori*

error may be mathematically intractable, so in order to arrive at the PNLMS (following the approach from [94]) the Taylor series expansion (TSE) of the instantaneous output error $e(k+1)$ around $e(k)$ is performed, that is

$$e(k+1) = e(k) + \sum_{i=1}^N \frac{\partial e(k)}{\partial w_i(k)} \Delta w_i(k) + \sum_{i=1}^N \frac{\partial e(k)}{\partial x_i(k)} \Delta x_i(k) + \frac{\partial e(k)}{\partial d(k)} \Delta d(k) + h.o.t. \quad (4.24)$$

where the term h.o.t. denotes the usually neglected (due to the linearity of the filter) higher order terms. Unlike in the approach from [94], generally the partial derivatives with respect to $\mathbf{x}$ and d as well as higher order terms in TSE (4.24), cannot be neglected due to the nonlinearity of the filter update. Firstly, from (4.4), we obtain the partial derivatives in (4.24) as

$$\begin{aligned} \frac{\partial e(k)}{\partial w_i(k)} &= -x(k-i+1) = -x_i(k), & i &= 1, 2, \dots, N \\ \frac{\partial e(k)}{\partial x_j(k)} &= -w_j(k), & j &= 1, 2, \dots, N \\ \frac{\partial e(k)}{\partial d(k)} &= 1, \end{aligned} \quad (4.25)$$

and the update $\Delta w_i(k)$ in (4.4), is given by

$$\Delta w_i(k) = \mu e(k) g_i(k) x_i(k), \quad i = 1, 2, \dots, N \quad (4.26)$$

Finally, after substituting (4.25)–(4.26) into (4.24) we obtain

$$e(k+1) = e(k) - \mu e(k) \left(\sum_{i=1}^N x_i^2(k) g_i(k) \right) - \left(\sum_{i=1}^N w_i(k) \Delta x_i(k) \right) + \Delta d(k) + \dots \quad (4.27)$$

To find the optimal learning rate η in the minimum mean squared error (MMSE) sense, differentiate $\frac{1}{2}e^2(k+1)$ with respect to μ

$$\frac{\partial \frac{1}{2}e^2(k+1)}{\partial \mu} = 0 \Rightarrow e(k+1) = 0, \quad (4.28)$$

and solve for μ , to yield

$$\mu_{opt} = \frac{e(k) - \left(\sum_{i=1}^N w_i(k) \Delta x_i(k) \right) + \Delta d(k)}{e(k) \sum_{i=1}^N x_i^2(k) g_i(k)}. \quad (4.29)$$

This learning rate is optimal, providing the first order TSE gives a good approximation of the instantaneous output error $e(k+1)$. Whenever the values $\Delta x_i(k)$ and $\Delta d(k)$ can be calculated or are known (as in the case of off-line training schemes or ensemble learning), they should be included. In the case of on-line linear adaptive filters, $\Delta d(k)$ cannot be computed and in addition, the variation of first order terms associated with $w_i(k)$, $x_i(k)$ and $d(k)$ between two successive discrete time instants may be insignificant. For simplicity, and using the independence assumptions, the PNLMS can be derived from (4.4) and (4.24), by neglecting the first order partial derivatives with respect to $\mathbf{x}$ and d , and also the second and higher order terms. The original PNLMS was introduced empirically and uses the standard NLMS update to obtain its optimal step-size. By minimising the error $e(k+1)$, from (4.29), gives²

$$\mu_{opt} = \frac{1}{\mathbf{x}^T(k) \mathbf{G}(k) \mathbf{x}(k)}, \quad (4.30)$$

that is, the Proportionate LMS is now equipped with an optimal learning rate, to yield

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu \frac{\mathbf{G}(k) e(k) \mathbf{x}(k)}{\mathbf{x}^T(k) \mathbf{G}(k) \mathbf{x}(k)}. \quad (4.31)$$

The introduction of the constant term μ is due to the assumptions in the TSE of $e(k+1)$, which results in an overestimate of the optimal learning rate, hence μ should be less than unity. This update has precisely the form of what has become the standard version of PNLMS [11].

Comparison of Different Formulations of PNLMS

It is worth noting that the above formulation of the PNLMS, whilst having an update of $\mathbf{w}(k)$ in line with that of the standard version of the PNLMS, does differ in one

²We shall ignore the trivial solution $e(k) = 0$.

notable point. In the original version of the PNLMS the update of the proportionate terms $\mathbf{G}(k)$ are given by (4.3) where

$$\bar{\phi}(k) = 1/N \sum_{n=1}^N \phi_n(k), \quad (4.32)$$

resulting in $\text{tr}[\mathbf{G}(k)] = N$. However, the more commonly used version of the PNLMS has the term $\bar{\phi}$ given by

$$\bar{\phi}(k) = \sum_{n=1}^N \phi_n(k), \quad (4.33)$$

which results in $\text{tr}[\mathbf{G}(k)] = 1$. In this case the bound on the step-size (4.21) becomes

$$0 < \mu < \frac{2}{\text{tr}[\mathbf{R}]} = \frac{2}{\mathbf{x}^T(k)\mathbf{x}(k)}, \quad (4.34)$$

and results in an update of $\mathbf{w}(k)$ in the form of the original PNLMS (4.1). As such, the version of the update of $\mathbf{w}(k)$ depends on the update of $\mathbf{G}(k)$ used; and the version of the PNLMS presented here is an optimal version in terms of the MSE, based upon the standard independence assumptions. In this sense neither the original Duttweiler version nor the standard version can be considered an ideal representation.

4.3 Unifying Approach to GASS Algorithms for PNLMS

Upon obtaining the PNLMS as a result of an optimisation procedure, it is natural to extend this approach in terms of fast and robust learning, by considering adaptive learning rates. The class of gradient adaptive step-size (GASS) algorithms, such as the algorithm proposed in [64] is based upon a gradient adaptation of the learning rate of LMS, based on $\frac{\partial \mathcal{J}(k)}{\partial \mu}$ where $\mathcal{J}(k) = 1/2e^2(k)$ is the cost function. This algorithm is essentially a simplified version of the Benveniste algorithm [12], where the time varying filtered versions of the instantaneous gradients of the cost function with respect to the learning rate are replaced by their instantaneous values. In the algorithm of Ang and Farhang [3], this time variant filtering of instantaneous gradients is replaced by a low

pass filter with a fixed coefficient.

To arrive at this class of algorithms for PNLMS, it is necessary to account for the exclusion of the neglected higher order terms of TSE (4.24) and to provide an insight into the extent the higher order terms influence the algorithm. The optimal learning rate obtained from the TSE of the output error $e(k+1)$ uses only the first term of the expansion, that is, the partial derivatives with respect to $w_i, i = 1, \dots, N$, and thus the TSE (4.24) can be rewritten as

$$e(k+1) = e(k) - \mu e(k) \sum_{i=1}^N x_i^2(k) g_i(k) + h.o.t.(k). \quad (4.35)$$

The higher order terms from (4.35) for the class of GASS algorithms are expressed as

$$h.o.t.(k) = \vartheta(k)e(k). \quad (4.36)$$

On the other hand, consider the adaptive step-size in the update of the NLMS (3.6), to facilitate the operation in a nonstationary environment, that is, to allow the filter to adapt independently of the signal power in the tap input, the NLMS [37, 106] uses an adaptive step-size

$$\frac{\mu}{\|\mathbf{x}(k)\|_2^2 + \varepsilon}, \quad (4.37)$$

where $\|\cdot\|_2$ is the Euclidean norm. For practical reasons, the regularisation parameter ε is included to prevent the weight update becoming unstable for input vectors comprising of near to zero elements, typical values of $\varepsilon = 0.01$. Without this it would otherwise be interpreted as effectively a large learning rate $\eta(k)$, with values outside the bound for convergence and hence lead to divergence. The GNGD class of algorithms [56], make the parameter ε in the step-size of NLMS gradient adaptive based on $\frac{\partial \mathcal{J}(k)}{\partial \varepsilon(k-1)}$, and have been introduced to provide compensation for the assumptions in the derivation of the NLMS. By introducing this form of adaptive regularisation into the update of the PNLMS gives an alternative method of allowing for the neglected higher order terms within the adaptive step-size. Following this approach the higher order terms being

expressed as

$$h.o.t.(k) = -\mu e(k)\varepsilon(k). \quad (4.38)$$

By setting (4.35) to zero and solving for μ , we can now express the optimal step-sizes for the Mathews' GASS algorithm for PNLMS as

$$\begin{aligned} e(k+1) &= e(k) - \mu e(k)\mathbf{x}^T(k)\mathbf{G}(k)\mathbf{x}(k) + \vartheta(k)e(k) \\ \mu e(k)\mathbf{x}^T(k)\mathbf{G}(k)\mathbf{x}(k) &= e(k) + \vartheta(k)e(k) \\ \mu &= \frac{e(k) + \vartheta(k)e(k)}{e(k)\mathbf{x}^T(k)\mathbf{G}(k)\mathbf{x}(k)} \\ \mu &= \frac{1 + \vartheta(k)}{\mathbf{x}^T(k)\mathbf{G}(k)\mathbf{x}(k)} \\ \mu_{opt} &= \frac{\eta(k)}{\mathbf{x}^T(k)\mathbf{G}(k)\mathbf{x}(k)} \end{aligned} \quad (4.39)$$

and the GNGD step-size update for PNLMS as

$$\begin{aligned} e(k+1) &= e(k) - \mu e(k)\mathbf{x}^T(k)\mathbf{G}(k)\mathbf{x}(k) - \mu e(k)\varepsilon(k) \\ \mu e(k)\mathbf{x}^T(k)\mathbf{G}(k)\mathbf{x}(k) + \mu e(k)\varepsilon(k) &= e(k) \\ \mu &= \frac{e(k)}{e(k)\mathbf{x}^T(k)\mathbf{G}(k)\mathbf{x}(k) + e(k)\varepsilon(k)} \\ \mu_{opt} &= \frac{1}{\mathbf{x}^T(k)\mathbf{G}(k)\mathbf{x}(k) + \varepsilon(k)}. \end{aligned} \quad (4.40)$$

Notice that there is a fundamental difference between the variable step-size algorithms with a “linear” multiplicative adaptive factor (GASS framework) and GNGD, which employs a “nonlinear” update of the adaptive learning rate. The exact derivation of the update in either case depends crucially on the weight sensitivities with respect to the adaptive term within the step-size.

4.3.1 Linear Update of the Adaptive Step-Size

In this case, using the term $h.o.t.(k)$ written in the form $\vartheta(k)e(k)$, to derive a form of a gradient adaptive step-size algorithm similar to that of the Mathews' algorithm for

the LMS, gives

$$\begin{aligned}
\mathbf{w}(k+1) &= \mathbf{w}(k) + \frac{\eta(k)}{\mathbf{x}^T(k)\mathbf{G}(k)\mathbf{x}(k)}\mathbf{G}(k)e(k)\mathbf{x}(k), \\
\eta(k+1) &= \eta(k) - \beta \frac{\partial \mathcal{J}(k)}{\partial \eta(k-1)}, \\
\frac{\partial \mathcal{J}(k)}{\partial \eta(k-1)} &= \frac{\partial \mathcal{J}(k)}{\partial e(k)} \frac{\partial e(k)}{\partial \mathbf{w}(k)} \frac{\partial \mathbf{w}(k)}{\partial \eta(k-1)}, \\
\eta(k+1) &= \eta(k) + \beta \frac{e(k)e(k-1)\mathbf{x}^T(k)\mathbf{G}(k-1)\mathbf{x}(k-1)}{\mathbf{x}^T(k-1)\mathbf{G}(k-1)\mathbf{x}(k-1)}, \tag{4.41}
\end{aligned}$$

where β is a small positive constant. In practice, it is frequently the case that learning rates need to be smaller than the optimal ones³ and the term $\eta(k)$ from above is often replaced by a small constant μ which multiplies an adaptive parameter, for instance $\eta(k) = \mu\vartheta(k)$.

The result in (4.41) is only an approximation of the rigorous analysis provided by Benveniste et al. [12] where,

$$\eta(k+1) = \eta(k) + \beta e(k)\mathbf{x}^T(k)\gamma(k) \tag{4.42}$$

and

$$\gamma(k) = [1 - \eta(k-1)]\gamma(k-1) + \frac{\mathbf{G}(k-1)e(k-1)\mathbf{x}(k-1)}{\mathbf{x}^T(k-1)\mathbf{G}(k-1)\mathbf{x}(k-1)}, \tag{4.43}$$

or alternatively using the approximation of Ang and Farhang [3]

$$\gamma(k) = \alpha\gamma(k-1) + \frac{\mathbf{G}(k-1)e(k-1)\mathbf{x}(k-1)}{\mathbf{x}^T(k-1)\mathbf{G}(k-1)\mathbf{x}(k-1)}, \tag{4.44}$$

where α is a constant such that $\alpha < 1$. For further detail see Appendix A.1.

³The optimal learning rates are usually derived for white inputs and based on the independence assumptions, however, in practice a smaller learning rate can account for the violation of the independence assumptions.

4.3.2 Nonlinear Update of the Adaptive Step-Size

In the second case, the term *h.o.t.* from (4.35) is written as *h.o.t.* = $-\mu e(k)\varepsilon(k)$, to arrive at the GNGD type PNLMS algorithm, where the variation of the time varying μ_{opt} is governed in a nonlinear manner. The weight update of this algorithm is derived starting from

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \frac{\mu}{\mathbf{x}(k)^T \mathbf{G}(k) \mathbf{x}(k) + \varepsilon(k)} \mathbf{G}(k) e(k) \mathbf{x}(k). \quad (4.45)$$

The parameter ε in (4.45) is made gradient adaptive as

$$\begin{aligned} \varepsilon(k+1) &= \varepsilon(k) - \beta \nabla_{\varepsilon(k-1)} \mathcal{J}(k), \\ \frac{\partial \mathcal{J}(k)}{\partial \varepsilon(k-1)} &= \frac{\partial \mathcal{J}(k)}{\partial e(k)} \frac{\partial e(k)}{\partial \mathbf{w}(k)} \frac{\partial \mathbf{w}(k)}{\partial \varepsilon(k-1)}, \\ &= \mu \frac{e(k) e(k-1) \mathbf{x}^T(k) \mathbf{G}(k) \mathbf{x}(k-1)}{[\mathbf{x}^T(k-1) \mathbf{G}(k-1) \mathbf{x}(k-1) + \varepsilon(k-1)]^2}, \\ \varepsilon(k+1) &= \varepsilon(k) - \beta \mu \frac{e(k) e(k-1) \mathbf{x}^T(k) \mathbf{G}(k-1) \mathbf{x}(k-1)}{[\mathbf{x}^T(k-1) \mathbf{G}(k-1) \mathbf{x}(k-1) + \varepsilon(k-1)]^2}, \end{aligned} \quad (4.46)$$

where β is a small positive constant. The adaptation of the time-varying term in the denominator of the above gradient adaptive learning rate compensates for the deficiencies in the derivation of the step-size of the NLMS [94] and hence the PNLMS. The complete derivation can be found in Appendix A.2.

4.3.3 Robust Regularisation

One method of improving problems with the steady state performance of adaptive algorithms is to introduce noise into the updates [103], thus helping to avoid convergence to local and spurious minima especially for small values of learning rates. One algorithm which aims to do just that has appeared in both the machine learning community [91] and the adaptive filtering community [19]. Therefore, following such an approach, gradient noise is introduced into the update of the regularisation parameter ε , with the aim of improving the performance of the adaptive ε PNLMS algorithm. This is achieved based on normalising the gradient of the cost function $\mathcal{J}(k)$ with respect to

ε , this extension is not trivial however, due to the tap-selective nature of the PNLMS update (4.45). Normalising the gradient (4.46) gives

$$\frac{\nabla_{\varepsilon(k-1)}\mathcal{J}(k)}{\|\nabla_{\varepsilon(k-1)}\mathcal{J}(k)\|_2} = \text{sgn}\left[\nabla_{\varepsilon(k-1)}\mathcal{J}(k)\right]. \quad (4.47)$$

This results in the sign regularised update given by

$$\varepsilon(k+1) = \varepsilon(k) - \beta \text{sgn} \left\{ \frac{e(k)e(k-1)\mathbf{x}^T(k)\mathbf{G}(k-1)\mathbf{x}(k-1)}{[\mathbf{x}^T(k-1)\mathbf{G}(k-1)\mathbf{x}(k-1) + \varepsilon(k-1)]^2} \right\}. \quad (4.48)$$

Notice in the above equations that the denominator from (4.46), which is the source of much of the computational complexity, can be omitted since it is always positive, giving

$$\varepsilon(k+1) = \varepsilon(k) - \beta \text{sgn} [e(k)e(k-1)\mathbf{x}^T(k)\mathbf{G}(k-1)\mathbf{x}(k-1)]. \quad (4.49)$$

From (4.49), due to the sparse representation, there is also the $\mathbf{G}(k-1)$ term to consider; since $\mathbf{G}(k-1)$ is always positive it can be omitted from the update (4.49) to give

$$\varepsilon(k+1) = \varepsilon(k) - \beta \text{sgn} [e(k)e(k-1)\mathbf{x}^T(k)\mathbf{x}(k-1)]. \quad (4.50)$$

By following the approach of [19], where by it has been noted that the sign GNGD algorithm makes use of the gradient vectors $\nabla_{\mathbf{w}}\mathcal{J}(k) = -e(k)\mathbf{x}(k)$ and $\nabla_{\mathbf{w}}\mathcal{J}(k-1) = -e(k-1)\mathbf{x}(k-1)$, and the inner product of these two vectors is given by

$$\nabla_{\mathbf{w}}^T\mathcal{J}(k)\nabla_{\mathbf{w}}\mathcal{J}(k-1) = \|\nabla_{\mathbf{w}}\mathcal{J}(k)\| \cdot \|\nabla_{\mathbf{w}}\mathcal{J}(k-1)\| \cdot \cos\theta, \quad (4.51)$$

where θ is the angle between the vectors. Since $\|\nabla_{\mathbf{w}}\mathcal{J}(k)\|$ and $\|\nabla_{\mathbf{w}}\mathcal{J}(k-1)\|$ are always positive, (4.50) reduces to

$$\begin{aligned} \varepsilon(k+1) &= \varepsilon(k) - \beta \text{sgn} [\nabla_{\mathbf{w}}^T\mathcal{J}(k)\nabla_{\mathbf{w}}\mathcal{J}(k-1)] \\ &= \varepsilon(k) - \beta \text{sgn} [\cos\theta]. \end{aligned} \quad (4.52)$$

Attention should be paid to the role of the matrix $\mathbf{G}(k-1)$, as this has an effect on the direction of $\mathbf{x}(k-1)$ and hence the angle between the two gradients. To cause a change in the sign of the ε update, we need $|\theta| > \pi/2$, and as the difference between $\mathbf{x}(k)$ and $\mathbf{x}(k-1)$ is generally small, slight changes to the direction of $\mathbf{x}(k-1)$ caused by $\mathbf{G}(k-1)$ are not significant and the change in sign, and consequent deviation from the learning curve is not likely.

4.4 Notes on Convergence and Computational Complexity

Due to the adaptive regularisation parameter in the GNGD and sign regularised versions, we have to introduce one more stability bound, given by

$$\varepsilon(k) > -\frac{\|x(k)\|_2^2}{2}. \quad (4.53)$$

The convergence and sensitivity analysis of this bound is in correspondence with the analysis of the GNGD [56].

The adaptive regularisation parameter update has two competing unconstrained optimisation processes, one based on $\nabla_{\mathbf{w}}\mathcal{J}$ and the other on $\nabla_{\varepsilon}\mathcal{J}$, for which the convergence speed is generally different, and would benefit from constraints on the allowable values of their parameters. Indeed, in practice for the Adaptive ε PNLMS there was no need to include a minimum bound on ε , but with the sign regularised version this is not expected, due to the simplified (and hence noisy) updates of ε . In this case and due to the unconstrained nature of (4.50), a lower bound of $\varepsilon_{\min} = 0$ has to be introduced to ensure the stable operation of the algorithm.

A comparison of computational complexity of the various algorithms considered is listed in Table 4.1. As well as computations listed in Table 4.1 all the PNLMS algorithms also contain $2N + 1$ maximum functions, which can be extremely computationally intensive. Notice that the sign regularised version does not result in a significant increase in computational complexity over PNLMS and offers a considerable reduction

Algorithm	Multiplications	Divisions	Additions
NLMS	$2N + 3$	1	$2N + 3$
Duttweiler PNLMS	$3N + 4$	$N + 2$	$3N + 1$
PNLMS	$5N + 2$	$N + 2$	$4N - 1$
Benveniste Type PNLMS	$9N + 4$	$2N + 2$	$5N + 1$
Ang Type PNLMS	$9N + 4$	$2N + 2$	$5N$
Mathews Type PNLMS	$7N + 5$	$N + 3$	$5N - 1$
Adaptive ε PNLMS	$7N + 7$	$N + 3$	$5N$
Sign Regularised PNLMS	$5N + 8$	$N + 2$	$4N + 2$

Table 4.1: Computational requirements for the NLMS, PNLMS and its GASS variants

over the Adaptive ε PNLMS.

4.5 Performance of the Class of PNLMS Algorithms

To illustrate the performance of the algorithms derived within the proposed GASS framework, simulations in a system identification setting were ran for a range of the algorithm parameters, including 3D graphs of the sensitivity of some of the algorithms to their parameters. Learning curves were produced using the normalised misalignment in dB, given by $10 \log_{10} \|\mathbf{w}_{\text{opt}} - \mathbf{w}\|_2^2 / \|\mathbf{w}_{\text{opt}}\|_2^2$, averaged over 100 independent trials, where $\mathbf{w}_{\text{opt}} = [w_{1 \text{ opt}}, \dots, w_{N \text{ opt}}]^T$ is the optimal filter coefficient vector. For all the algorithms, the parameters ρ and δ were set to the recommended values of $\rho = 5/N$ and $\delta = 0.01$ [26]. The regularisation parameter ε was set to $\varepsilon = 0.01$ or for the adaptive ε algorithms $\varepsilon(0) = 0.01$, whereas for the Ang type learning rate $\alpha = 0.9$. The sparse system under consideration was the benchmark system analysed in [62], a 100 tap channel with four non-zero taps

$$\begin{bmatrix} 0.1, & 1, & -0.5, & 0.1 \end{bmatrix} \text{ in positions } \begin{bmatrix} 1, & 30, & 35, & 85 \end{bmatrix}. \quad (4.54)$$

The performance of the standard PNLMS compared with that of the proposed algorithms for a step-size of μ and $\eta(0) = 0.1$ is shown in Fig. 4.1. Initially for all the linear adaptive learning rate updates the step-size was $\beta = 0.01$ (Fig. 4.1(a)). In this case, the Benveniste and Ang algorithms offered faster rates of convergence,

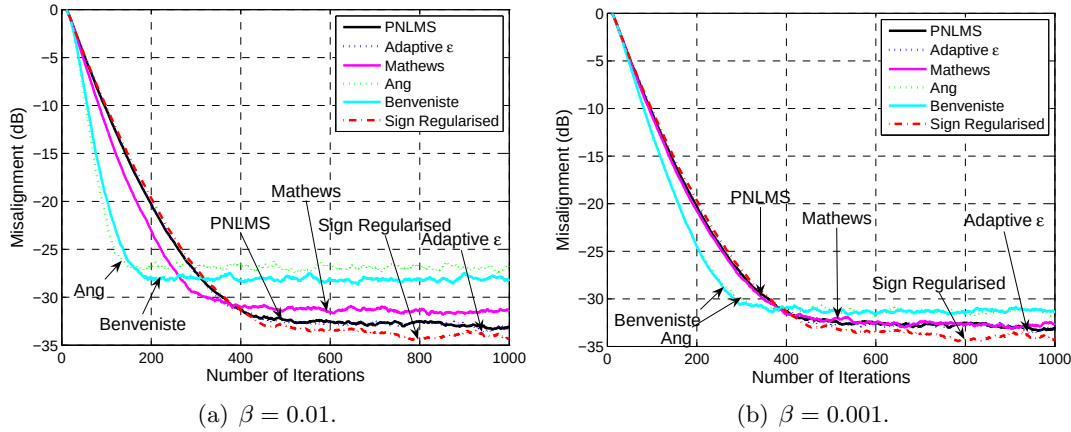


Figure 4.1: Performance comparison of the GASS PNLMS algorithms with the standard PNLMS for $\mu = 0.1$.

but the PNLMS and adaptive ε algorithms had a smaller steady state error, with the Mathews algorithm in between. Figure 4.1(b) shows convergence curves for the same parameter settings, except for the linear learning rates, where the parameter β was set to $\beta = 0.001$. A comparison of the performances of algorithms with “linear” learning rates in Fig. 4.1(b), with those in Fig. 4.1(a) highlights the sensitivity issues associated with the linear learning rates and also the selection of their parameters.

To illustrate the behaviour of the algorithms in typical critical operating conditions (simulating the effect of close to zero inputs⁴), the value of μ and $\eta(0)$ was increased to 1.95, at which point PNLMS is on the limit of stability. As shown in Fig. 4.2(a), with the value of $\beta = 0.01$ the Benveniste algorithm did not converge, and the Ang algorithm offered a smaller steady state error than the Mathews algorithm. Observe that when the algorithms were operating in critical conditions, the performance of the algorithm with an adaptive regularisation factor offers slight improvement compared to the PNLMS and the sign regularised version offers much improved performance compared to all the algorithms. Figure 4.2(b) illustrates that by reducing the value of β to $\beta = 0.001$ causes the Benveniste algorithm to converge, however, the performance was in line with that of the adaptive ε algorithm.

The sensitivity of the PNLMS and the adaptive ε PNLMS for a range of parameter

⁴Notice from (4.39)–(4.40) as $\mathbf{x}(k)\mathbf{G}(k)\mathbf{x}(k) \rightarrow 0$ then effectively $\mu(k) \rightarrow \infty$. By making μ and $\eta(0)$ large we achieve the same effect, allowing a convenient comparison of the fixed, linear, and variable ε methods.

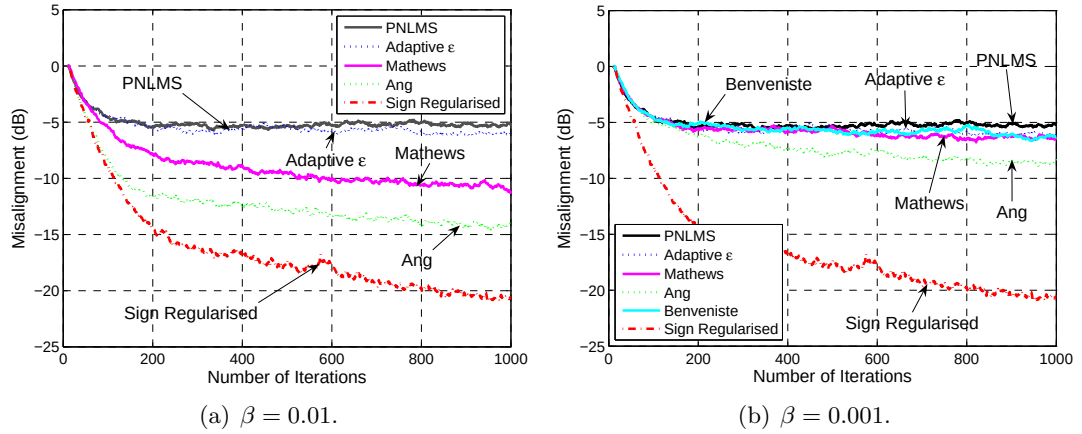


Figure 4.2: Performance comparison of the GASS PNLMS algorithms with the standard PNLMS for $\mu = 1.95$.

values within the algorithm are shown in Fig. 4.3. The parameter ρ was varied between zero and unity, for which the behaviour of the PNLMS approaches that of the NLMS. Changes in δ were not investigated, as it is known to only have an effect on the algorithm initially when all filter coefficients are zero. To show how the PNLMS behaves in extreme circumstances, the value of μ was varied between zero and three, with $\mu \in [2, 3)$ being outside the normal range of stability for the PNLMS. The PNLMS and adaptive ε PNLMS algorithms exhibited similar performance for $0 < \mu \leq 1.8$, but beyond this the difference between the two algorithms is clearly apparent with the PNLMS becoming extremely unstable. Figure 4.4 shows the above experiment repeated for adaptive ε PNLMS and sign regularised PNLMS. Whilst the behaviour of both algorithms was clearly still under control it can be seen that for all settings with $\mu \leq 2.5$, sign regularised PNLMS had similar or better performance than adaptive ε PNLMS. For situations with $\mu > 2.5$, ε PNLMS showed better stability than sign regularised PNLMS, due to its more rigorous derivation, however the sign regularised PNLMS offers a viable alternative as a trade-off between performance and computational complexity (Table 4.1).

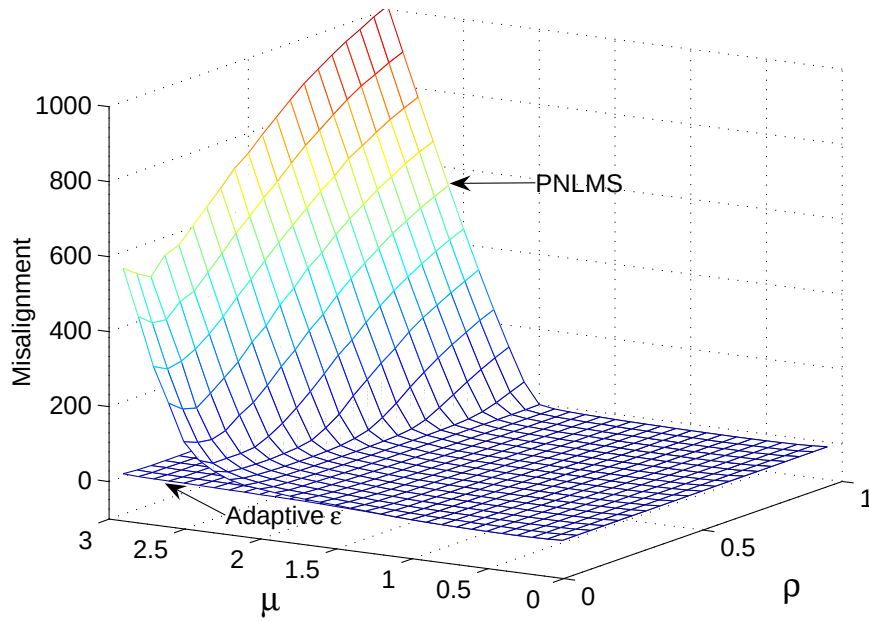


Figure 4.3: Performance comparison of the PNLMS with the adaptive ε PNLMS for variations in μ and ρ .

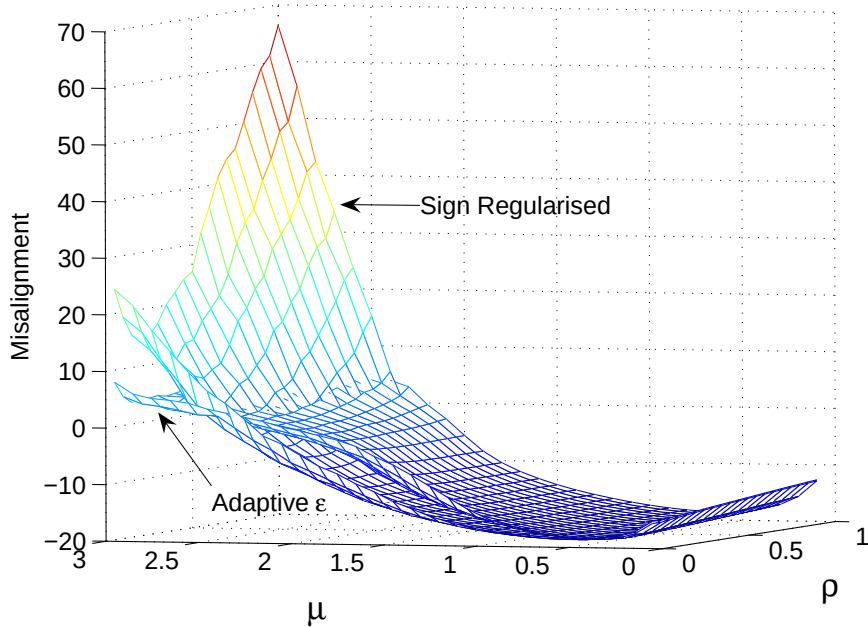


Figure 4.4: Performance comparison of the adaptive ε PNLMS with the sign regularised PNLMS for variations in μ and ρ .

4.6 Conclusions

With the aim of introducing a hybrid filter to identify the sparseness of a signal, the behaviour of the PNLMS algorithm is investigated. To avoid problems experienced with PNLMS when processing inputs with large dynamics and ill-conditioned tap input correlation matrices, a class of algorithms with GASS has been implemented. A unified approach to the derivation of the class of PNLMS algorithms, starting from a standard LMS through to the PNLMS and its adaptive step-size variants has been achieved in a generic way, and has allowed the introduction of two classes of adaptive step-size algorithms within the same framework. This framework can be used as a unifying framework for the derivation and analysis of the entire class of PNLMS algorithms.

For the example given, the Benveniste algorithm, although the most mathematically rigorous, when combined with the empirically derived PNLMS, has been shown to lack robustness. In comparison, the other “linear” adaptive learning rate PNLMS algorithms, Ang’s and Mathews’, have exhibited enhanced robustness. The “nonlinear” adaptive ε algorithm has proved to be significantly easier to tune and less sensitive to initial conditions and when used in an ill-conditioned environment remains stable. To overcome problems incurred in steady state and to reduce computational complexity, an algorithm based upon a robust adaptive regularisation parameter has also been derived.

Chapter 5

Hybrid Filters for Sparsity and Combining Multiple Signal Modality Trackers

5.1 Introduction

Having investigated the performance of one of the key algorithms for sparse environments and having illustrated the use of the collaborative adaptive filtering architecture to identify and track the nonlinearity of a signal, it follows naturally to combine the two. For the purpose of identifying the sparsity of a signal the hybrid structure from Chapter 3 is employed, replacing the nonlinear filter with one suited for sparse environments. The linear filter in the hybrid structure remained trained by the NLMS algorithm. Initially the performance of the PNLMS algorithm is assessed and then compared with that of another filter specifically designed for sparse environments, the signed sparse LMS (SSLMS). Following this the effect of additive noise on the performance of the hybrid filter is investigated.

Once a stable collaborative adaptive filtering approach to identifying and tracking changes in the sparsity of a signal has been identified, the hybrid filters for sparsity and nonlinearity are compared. By combining the trackers from both hybrid filters allows

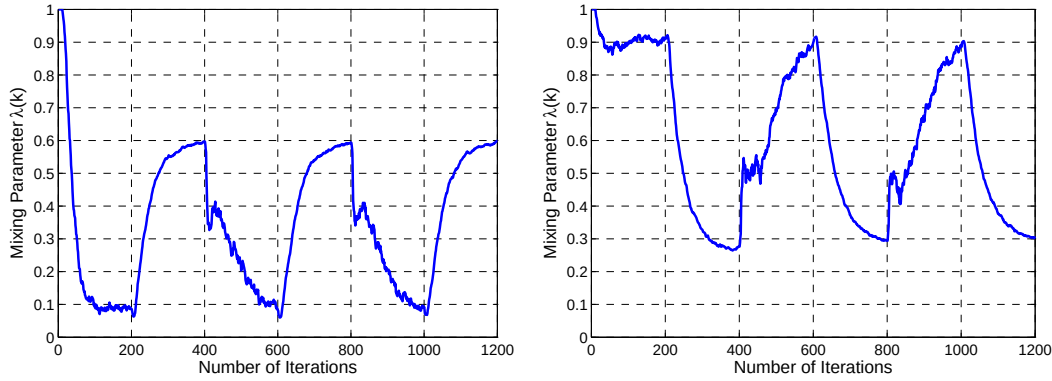


Figure 5.1: Comparison of the evolution of mixing parameter λ for two different parameter settings of the hybrid filter combining NLMS and PNLMS for an input signal alternating between a linear AR(4) signal to a sparse signal every 200 samples.

for building a more complete picture when identifying the nature of a signal.

5.2 Comparison of Algorithms for Sparse Hybrid Filters

Whilst it was found that it was fairly straightforward to detect the changes in modality, it was not always possible to detect the direction of the changes, whether from sparse to nonsparse or vice versa, or the degree of the change. Figure 5.1 shows the hybrid filter consisting of the NLMS and PNLMS for the same input signal, alternating from linear to nonlinear (sparse) every 200 samples. The linear signal was again the AR(4) signal (3.8) and the sparse signal was generated from the benchmark system [62] used in the previous chapter (4.54). It can clearly be seen that whilst both plots accurately show the occurrence of the changes in the input signal, they do not agree on which filter is performing better for each section of the data. These inconclusive results occur due to the different settings of the parameters of the constituent subfilters. This can be attributed to similarities in the nature of the PNLMS and NLMS algorithms, leading to the PNLMS having the ability to process the nonsparse signal equally as well as the NLMS. Similar problems were also encountered with the proposed algorithms from Chapter 4. In an attempt to overcome these limitations, the simulations were repeated

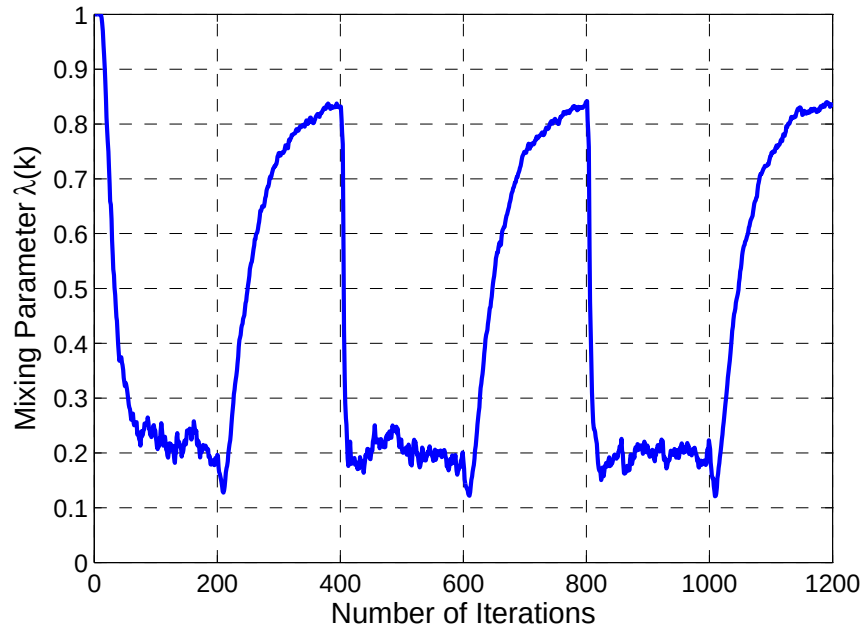


Figure 5.2: Evolution of the mixing parameter λ for a hybrid filter combining the NLMS and SSLMS algorithms for an input signal which alternates from a linear AR(4) signal to a sparse signal every 200 samples.

replacing the PNLMS with the SSLMS [62]. The SSLMS is given by

$$\begin{aligned}
 y(k) &= \mathbf{x}^T(k) \mathbf{w}(k) \\
 e(k) &= d(k) - y(k) \\
 \mathbf{w}(k+1) &= \mathbf{w}(k) + \mu (|\mathbf{w}(k)| + \varepsilon) e(k) \mathbf{x}(k)
 \end{aligned} \tag{5.1}$$

Unlike in the previous example the SSLMS filter did not perform well compared to the NLMS for nonsparse signals, making it better suited to the task than the PNLMS. Figure 5.2 shows the output of the hybrid filter consisting of the NLMS and the SSLMS for the alternating signal. In this case no matter what the initial parameter settings of the subfilters, the hybrid filter always showed the changes in the signal nature to be in the same direction.

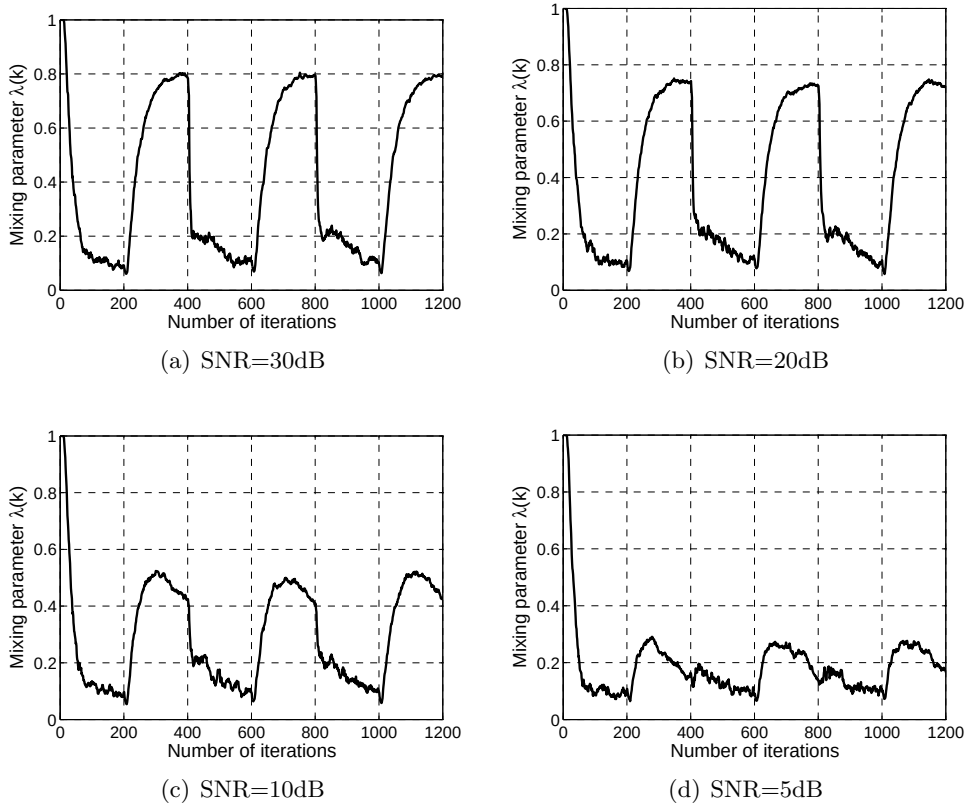


Figure 5.3: Evolution of the mixing parameter λ for a hybrid filter combining the NLMS and SSLMS algorithms for an input signal which alternates from a linear AR(4) signal to a sparse signal every 200 samples, with varying degrees of noise.

5.2.1 Effect of Additive Noise

To illustrate the performance of the filter in unfavourable circumstances, the simulation was repeated with the same filter settings and alternating input, but with increasing amounts of additive noise. As can be seen in Fig. 5.3 providing the signal to noise ratio is greater than 10dB it is possible to identify the changes in the signal nature.

5.3 Multiple Signal Modality Trackers

To consider the difference between nonlinearity in general and sparsity both hybrid filters were presented with an input signal which alternated from linear (3.8) to non-linear (3.9), then back to linear (3.8) and finally to sparse (4.54). The input signal was alternated every 100 samples and the corresponding dynamics of the mixing parameters

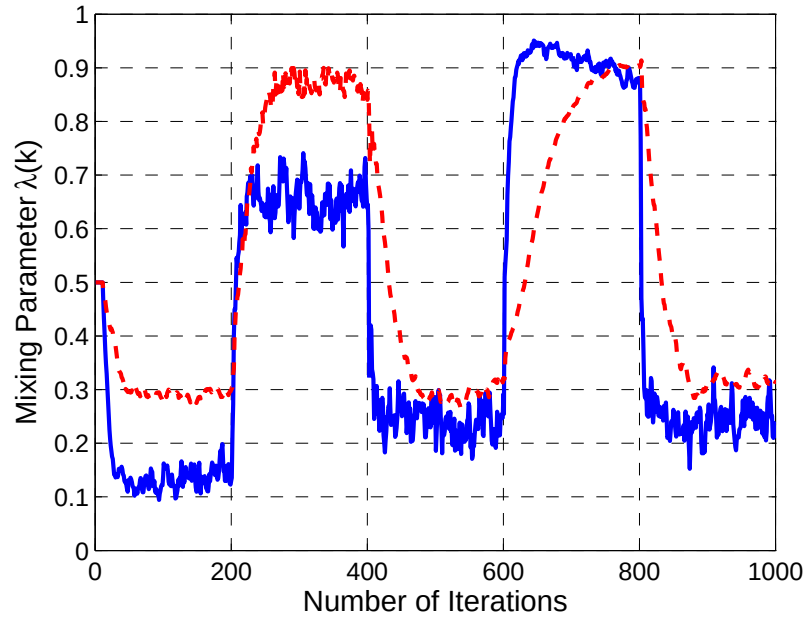


Figure 5.4: Evolution of the mixing parameter λ for hybrid filters combining NLMS with either NNGD (*broken line*) or SSLMS (*solid line*) for an input signal alternating from linear to nonlinear and linear to sparse every 200 samples.

$\lambda(k)$ are shown in Fig. 5.4, where a value of $\lambda = 1$ corresponds to the output of the NNGD/SSLMS trained subfilters and a value of $\lambda = 0$ corresponds to the output of the NLMS trained subfilters. It is clear from Fig. 5.4 that the value of λ adapts to be dominated by the filter most suited to the current dynamics of the input signal. As expected (as sparsity can be considered a subset of nonlinearity) the nonlinear hybrid filter obtains similar results for both the nonlinear and sparse inputs, whereas the sparse hybrid filter shows a marked difference in levels of sparsity for the same inputs.

It is natural to consider whether these results can therefore be combined to allow us to track not only changes in nonlinearity but also at the same time changes in sparsity. Intuitively there should be a certain degree of correlation between the changes in the evolution of the mixing parameters λ of the sparse hybrid filter and that of the nonlinear hybrid filter. This allows us to distinguish not only between linear and nonlinear signals but also to distinguish particular types of nonlinearity of interest allowing us to build up a more complete interpretation of the nature of the signal in question. It is worth noting at this point that in all cases the outputs of the hybrid filters and their subfilters

converged for all the input signals. The MSEs of each gave good results with only small differences in their performances.

Figure 5.5 shows the response of λ for both the nonlinear hybrid filter and the sparse hybrid filter for the alternating input signal previously described, with the solid lines representing the linear sections, the broken lines the nonlinear sections and the dotted lines the sparse sections. For the linear sections although the evolution of the two λ s do not follow the same path, there is an obvious correlation between them and the difference in responses can be attributed to the different learning rates of the subfilters of each hybrid filter. For the nonlinear and the sparse signals, however, the sparsity and saturation type nonlinearity are different phenomena and the sparse and nonlinear hybrid filters behaved differently. The results obtained from this method give a clearer picture of the nature of the nonlinearity of the signal, than any results obtained using a hybrid filter with nonlinear and sparse subfilters. This is due to the fact any signals which are sparse are also nonlinear so there is little advantage gained from the sparse subfilter over the nonlinear subfilter and therefore a hybrid filter using this combination would result in less clearly defined values of λ .

This representation is similar to the phase space representation in chaos theory, and allows for the signal modality characterisation to be considered within the framework of nonlinear dynamics. These results highlight the use of this technique in building a complete understanding of the nature of signals and has natural extensions both by using third dimensions and also by using alternative filters to explore different signal characteristics.

5.4 Conclusions

The hybrid filter structure for detecting changes in signal modality has been changed, from looking at changes in nonlinearity generally, to focussing more specifically on changes in sparsity. Comparison of different sparse and nonsparse filters has been carried out, indicating that the NLMS and SSLMS algorithms are particularly well suited to perform this task. The analysis and simulations again illustrate that there is

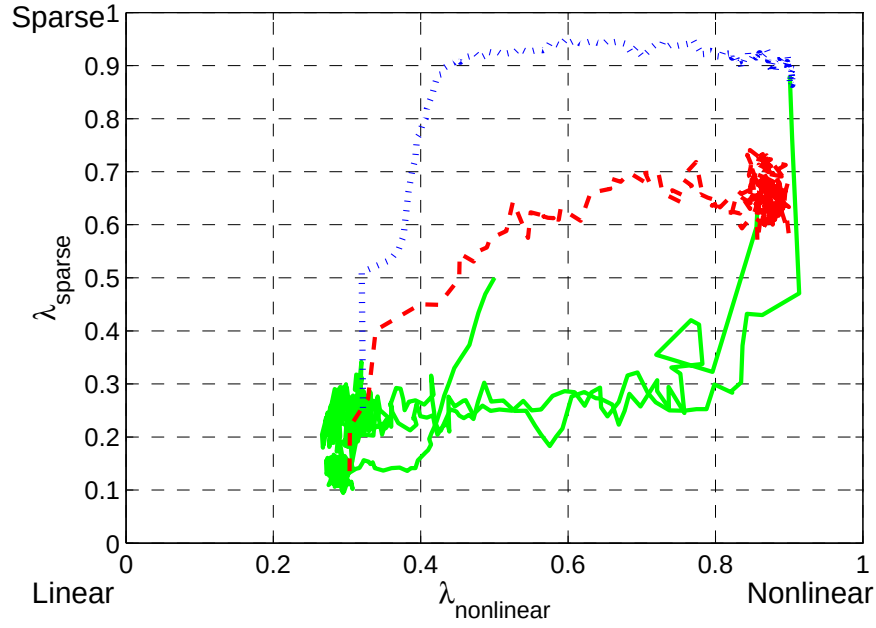


Figure 5.5: Comparison of the evolution of the mixing parameters for linear/nonlinear and sparse/nonsparse for an input signal alternating every 200 samples. *Solid line:* linear samples, *broken line:* nonlinear samples, *dotted line:* sparse samples.

significant potential for the use of this method and that in this case it is also reasonably robust to additive noise. A method has also been presented by which it is possible to build on this information by combining the responses of several mixing parameters to obtain a more complete understanding of the nature of signals.

Chapter 6

Hybrid Filtering & Signal Modality Characterisation in the Complex Domain

6.1 Introduction

For generality, building upon the identification and tracking of signal modality in the real domain $\mathbb{R}$, it is natural to extend this analysis into the complex domain $\mathbb{C}$. In order to facilitate this, the update of λ (3.5) has been extended into the complex domain. Note that since the input to the filters is complex, the error $e(k)$ is also complex, and therefore the mixing parameter $\lambda(k)$ is obtained from

$$\lambda(k+1) = \lambda(k) - \mu_\lambda \nabla_\lambda \mathcal{J}(k)|_{\lambda=\lambda(k)}, \quad (6.1)$$

$$\nabla_\lambda \mathcal{J}(k)|_{\lambda=\lambda(k)} = \left\{ e(k) \frac{\partial e^*(k)}{\partial \lambda(k)} + e^*(k) \frac{\partial e(k)}{\partial \lambda(k)} \right\}. \quad (6.2)$$

The two gradient terms from (6.2) can be evaluated as

$$\frac{\partial e(k)}{\partial \lambda(k)} = y_1(k) - y_2(k), \quad (6.3)$$

$$\frac{\partial e^*(k)}{\partial \lambda(k)} = (y_1(k) - y_2(k))^*, \quad (6.4)$$

yielding the update of the mixing parameter in the form

$$\lambda(k+1) = \lambda(k) + \mu_\lambda [e(k) (y_1(k) - y_2(k))^* + e^*(k) (y_1(k) - y_2(k))], \quad (6.5)$$

where $(\cdot)^*$ denotes the complex conjugation operator.

To introduce signal modality characterisation in the complex domain, a hybrid filter with a combination of the complex NLMS (CNLMS) and the complex NNGD (CNNGD) algorithms is employed to track the nonlinearity within a complex signal. Following this to identify the type of nonlinearity of the signal, the split-complex or fully-complex nature of signals is assessed. This is achieved using two filters both trained by the complex nonlinear gradient descent (CNGD) one with a split complex activation function the other with a fully-complex activation function. Moving on from nonlinearity the circularity of complex signals is investigated, through a collaborative adaptive filter consisting of the complex LMS (CLMS) and augmented CLMS (ACLMS). It can be shown that for a noncircular signal the convex combination of these two filters offers the faster convergence speed of the CLMS with the better steady state error of the ACLMS. Following this result, it is shown that by tracking the value of the mixing parameter within the update of the hybrid filter it is possible to identify and track the circularity of a signal. To this end the problem of autoregressive modelling in $\mathbb{C}$ in light of augmented statistics is addressed to perform widely linear autoregressive modelling of the generality of complex signals. The convergence analysis of the mixing parameter in the hybrid filter shows that the hybrid filter converges uniquely to the subfilter which is best suited to the data nature.

6.2 Identification and Tracking of Nonlinearity in Complex Signals

For the purposes of tracking nonlinearity of complex signals, the subfilters of the hybrid filter were trained by the CNLMS and the CNNGD. The CLMS is a natural extension of the LMS algorithm and benefits from the stability and robustness of the LMS. The

CLMS is described by [105]

$$\begin{aligned} e(k) &= d(k) - y(k), \\ y(k) &= \mathbf{h}^T(k) \mathbf{z}(k), \\ \mathbf{h}(k+1) &= \mathbf{h}(k) + \mu e(k) \mathbf{z}^*(k). \end{aligned} \quad (6.6)$$

The CNGD for a fully complex nonlinear activation function is described by

$$\begin{aligned} e(k) &= d(k) - net(k), \\ net(k) &= \Phi(\mathbf{x}^T(k) \mathbf{w}(k)), \\ \mathbf{h}(k+1) &= \mathbf{h}(k) + \mu e(k) (\Phi'[net(k)])^* \mathbf{x}^*(k), \end{aligned} \quad (6.7)$$

where $\mathbf{h}(k) = [h_1(k), h_2(k), \dots, h_N(k)]^T$ is the filter weight coefficient vector for a filter of length N . For the normalised variants, CNLMS and CNNGD the step-size μ is replaced with

$$\eta(k) = \frac{\mu}{\|\mathbf{x}(k)\|_2^2 + \varepsilon}, \quad (6.8)$$

$$\eta(k) = \frac{\mu}{[\Phi'(net(k))]^2 \|\mathbf{x}(k)\|_2^2 + \varepsilon}. \quad (6.9)$$

The synthetic benchmark data used was as in Chapter 3 with the process $n(k)$ replaced with a complex white Gaussian process, $n(k) = n_r(k) + jn_i(k)$, with zero mean and unit variance. Where the real and imaginary components of n are mutually independent sequences having equal variances so that $\sigma_n^2 = \sigma_{n_r}^2 + \sigma_{n_i}^2$. For the NNGD the nonlinearity used was the sigmoid function

$$\Phi(z) = \frac{1}{1 + e^{-z}}, \text{ where } z \in \mathbb{C}. \quad (6.10)$$

The results shown in Figure 6.1 demonstrate the robustness of this combination, indeed, the hybrid filter performed well on a range of synthetically generated signals and accurately detected both the direction of the change from linear to nonlinear and vice

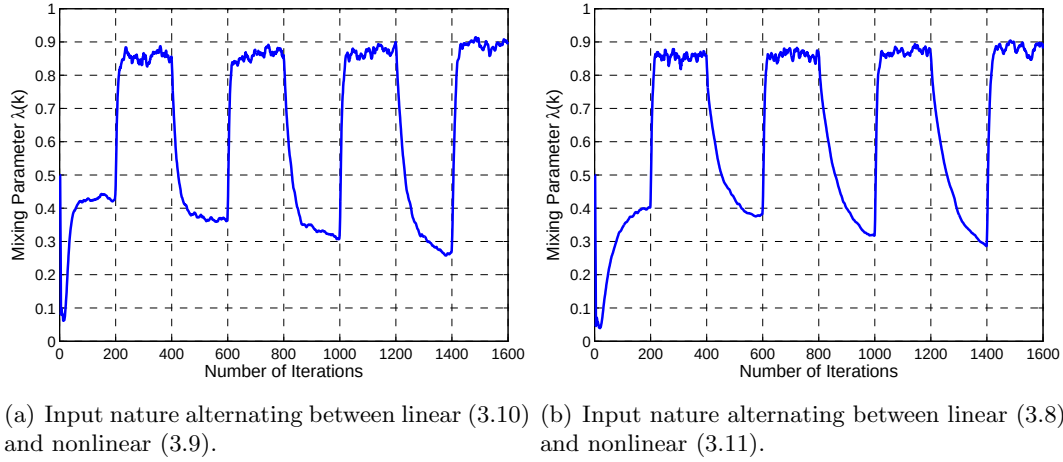


Figure 6.1: Evolution of the mixing parameter λ for a hybrid filter combining the CNGD and CNLMS, for input nature alternating between linear and nonlinear every 200 samples.

versa, and the degree of the change. Furthermore, compared to the problems incurred with signal conditioning when using the CLMS and CNGD [54], the tracking of the degree of nonlinearity in the input has been shown to improve in terms of the range swept by λ . The values of λ obtained for the standard versions correctly indicated the changes in the nature of the signal. The difference, however, between the steady state values was not as large as those obtained from the normalised versions of the algorithms.

6.2.1 Split-Complex vs. Fully-Complex

Following having shown that the proposed collaborative adaptive filtering approach is capable of tracking changes in the nonlinearity of a complex valued signal, next the aim is to address the issue of identifying the nature of the nonlinearity in terms of the so called “split-complex” nonlinearity or “fully-complex” nonlinearity. For the purpose of this the hybrid filter is equipped with subfilters both trained by the CNGD algorithm, one with a split-complex activation the other with a fully-complex activation function.

Consider the Ikeda map, a well known benchmark signal in chaos theory [1], given

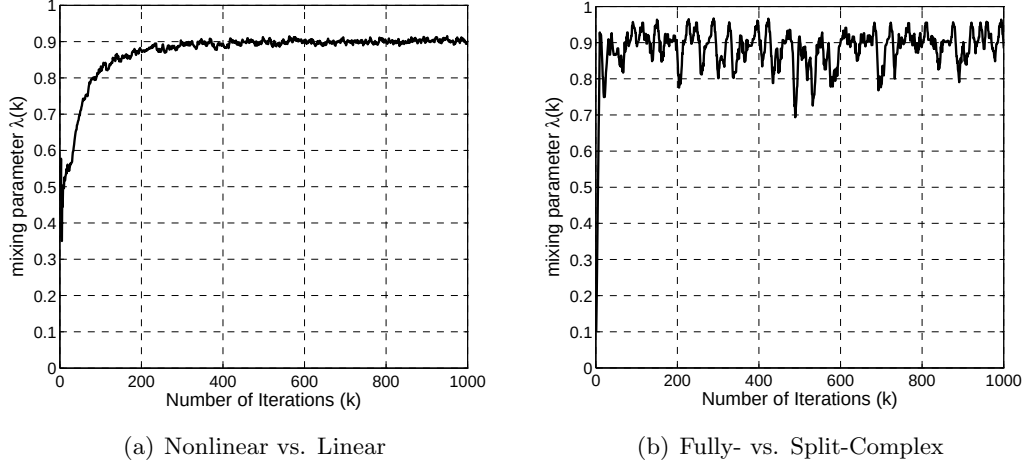


Figure 6.2: Evolution of the mixing parameter λ for prediction of Ikeda map

by

$$\begin{aligned} x(k+1) &= 1 + u[x(k) \cos t(k) - y(k) \sin t(k)] \\ y(k+1) &= u[x(k) \sin t(k) + y(k) \cos t(k)] \end{aligned} \quad (6.11)$$

where u is a parameter and

$$t(k) = 0.4 - \frac{6}{1 + x^2(k) + y^2(k)}. \quad (6.12)$$

Figure 6.2(a) illustrates that the hybrid filter can clearly identify the Ikeda map as nonlinear when presented with it as an input. It is natural, however, to expect that as the signal generation mechanism is in the form of coupled difference equations, by representing the pair $[x(n), y(n)]$ as a vector in $\mathbb{C}$, the Ikeda map (6.11) will represent a fully complex signal.

This is confirmed as expected (by design), from Fig. 6.2, the Ikeda map is a nonlinear signal (Fig. 6.2(a)) which exhibits fully-complex nonlinear properties (Fig. 6.2(b)). To illustrate this further, Fig. 6.3 shows the performance of the complex real time recurrent learning (CRTRL) algorithm [33] for both split- and fully-complex learning on the prediction of the Ikeda map; observe that the split-complex CRTRL did not respond as well as the fully complex version to prediction of the Ikeda map.

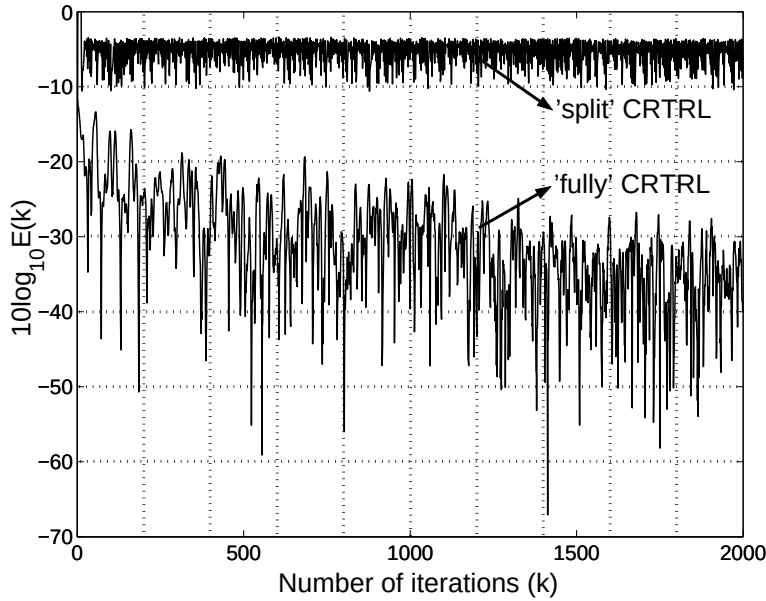


Figure 6.3: Learning curves for fully-complex vs. split-complex prediction of Ikeda map

6.3 Widely Linear Autoregressive Modelling

Before moving on to consider characterisation of circularity of signals, it is beneficial to consider autoregressive modelling in the complex domain. A better understanding of complex autoregressive modelling not only gives a better understanding of the nature of signals but also allows the production of benchmark synthetic data. Similarly to the augmented covariance matrix, when performing autoregressive (AR) modelling based on augmented statistics, second order noncircular (improper) data needs to be accounted for. The standard AR(n) model (in $\mathbb{R}$ or $\mathbb{C}$) is described by

$$z(k) = a_1 z(k-1) + \dots + a_n z(k-n) + q(k), \quad (6.13)$$

where $a_1, \dots, a_n$ are the AR coefficients, $z(k-1), \dots, z(k-n)$ are the samples of the regressor vector and $q(k)$ are samples of the driving white Gaussian noise. Using the

Yule-Walker equations, the AR coefficients can be found from [80]

$$\mathbf{a}^* = \mathcal{C}_{\mathbf{z}\mathbf{z}}^{-1} \mathbf{c}$$

$$\begin{bmatrix} a_1^* \\ a_2^* \\ \vdots \\ a_n^* \end{bmatrix} = \begin{bmatrix} \mathcal{C}_{\mathbf{z}\mathbf{z}}(0) & \mathcal{C}_{\mathbf{z}\mathbf{z}}^*(1) & \dots & \mathcal{C}_{\mathbf{z}\mathbf{z}}^*(n-1) \\ \mathcal{C}_{\mathbf{z}\mathbf{z}}(1) & \mathcal{C}_{\mathbf{z}\mathbf{z}}(0) & \dots & \mathcal{C}_{\mathbf{z}\mathbf{z}}^*(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{C}_{\mathbf{z}\mathbf{z}}(n-1) & \mathcal{C}_{\mathbf{z}\mathbf{z}}(n-2) & \dots & \mathcal{C}_{\mathbf{z}\mathbf{z}}(0) \end{bmatrix}^{-1} \begin{bmatrix} \mathcal{C}_{\mathbf{z}\mathbf{z}}(1) \\ \mathcal{C}_{\mathbf{z}\mathbf{z}}(2) \\ \vdots \\ \mathcal{C}_{\mathbf{z}\mathbf{z}}(n) \end{bmatrix}, \quad (6.14)$$

where $\mathbf{c}$ is the time shifted correlation vector. Using this model, it is possible to generate both proper and improper linear data, depending on the driving noise. Consider the signal

$$z(k) = \mathbf{a}^T \mathbf{z}(k) + q(k), \quad (6.15)$$

where $\mathbf{z}(k) = [z(k-1), z(k-2), \dots, z(k-N)]^T$. The linear $\hat{z}_l(k)$ and widely linear $\hat{z}_{wl}(k)$ estimates of $z(k)$ are given respectively by [76]

$$\hat{z}_l(k) = \mathbf{a}^T \mathbf{z}(k), \quad (6.16)$$

$$\hat{z}_{wl}(k) = \mathbf{h}^T \mathbf{z}(k) + \mathbf{g}^T \mathbf{z}^*(k). \quad (6.17)$$

The corresponding optimal AR estimation errors are given by

$$\begin{aligned} e_l^2 &= E[|z(k)|^2] - E[|\hat{z}_l(k)|^2] \\ &= \mathbf{c}^T \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathbf{c}^* + \sigma_q^2 - \mathbf{c}^T \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathbf{c}^*, \end{aligned} \quad (6.18)$$

$$\begin{aligned} e_{wl}^2 &= E[|z(k)|^2] - E[|\hat{z}_{wl}(k)|^2] \\ &= \mathbf{c}^T \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathbf{c}^* + \sigma_q^2 - \mathbf{r}^T \mathcal{R}_{\mathbf{z}\mathbf{z}}^{*-1} \mathbf{r}^*. \end{aligned} \quad (6.19)$$

Therefore depending on the nature of the driving noise $q(k)$ there are three possible scenarios (distributions for the example of an AR(4) model are shown in Fig. 6.4 and the corresponding covariance and pseudocovariance in Fig. 6.5).

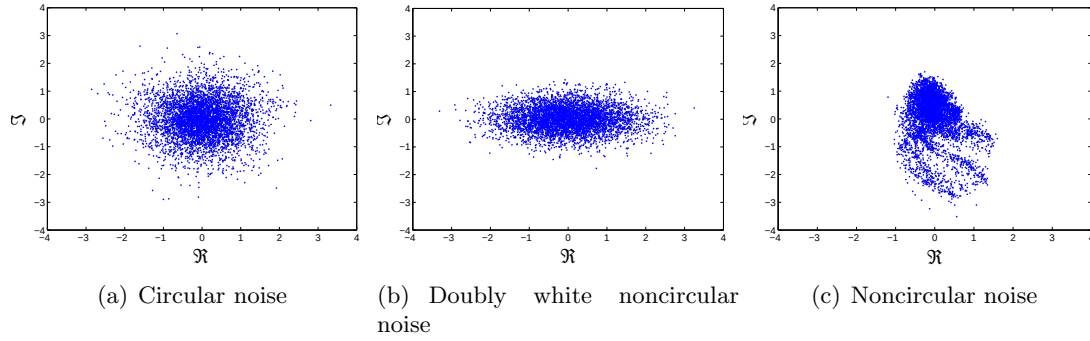


Figure 6.4: Distributions of AR(4) with different driving noise

- **Doubly White Circular Noise:** In this case $\mathcal{P}_{\mathbf{q}\mathbf{q}} = \mathbf{0}$, $q_r(k) \perp q_i(k)$, and $\sigma_{q_r}^2 = \sigma_{q_i}^2$, and the resultant signal is also circular (Fig. 6.5(a)) hence

$$e_l^2 = e_{wl}^2, \quad (6.20)$$

and the signal can be sufficiently modelled by standard AR model.

- **Doubly White Improper Noise:** In this case $\mathcal{P}_{\mathbf{q}\mathbf{q}} \neq \mathbf{0}$, $q_r(k) \perp q_i(k)$ and $\sigma_{q_r}^2 > \sigma_{q_i}^2$, and the resultant signal is noncircular (Fig. 6.5(b)). However, in this case there is no correlation between the current value of $z(k)$ and the previous samples $\mathbf{z}(k)$, meaning $\mathbf{c} = \mathbf{0}$ and therefore (as the pseudocovariance by its nature must be smaller than the covariance) $\mathbf{p} = \mathbf{0}$. Thus, in this case, the errors of the standard and widely linear estimator are the same,

$$e_l^2 = e_{wl}^2 = \sigma_q^2. \quad (6.21)$$

Therefore when noncircular linear processes are generated by a doubly white noncircular noise driving standard AR models, there is no advantage in using a widely linear model [74].

- **Noncircular Noise:** In this case $\mathcal{P}_{\mathbf{q}\mathbf{q}} \neq \mathbf{0}$ and the resultant AR signal is non-circular (Fig. 6.5(c)). The advantage of the widely linear model over the linear

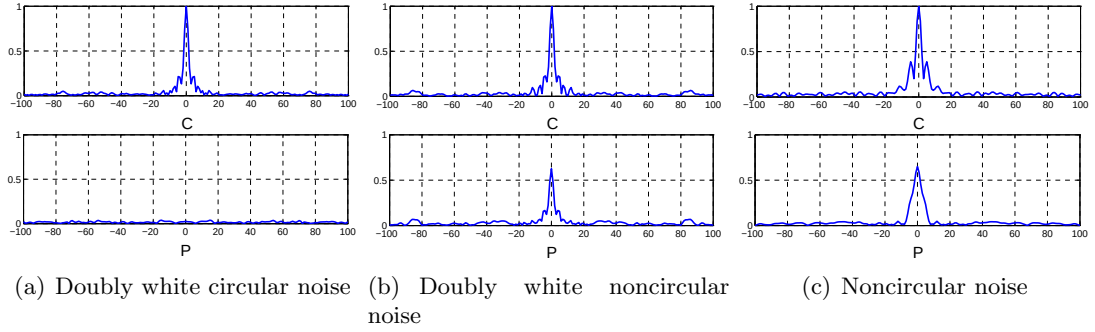


Figure 6.5: Covariance and pseudocovariance of a standard AR(4) model driven by different driving noises

model is quantified from

$$\begin{aligned}
 \delta e^2 &= e_l^2 - e_{wl}^2 \\
 &= E[|z(k)|^2] - E[|\hat{z}_l(k)|^2] - (E[|z(k)|^2] - E[|\hat{z}_{wl}(k)|^2]) \\
 &= E[|\hat{z}_{wl}(k)|^2] - E[|\hat{z}_l(k)|^2] \\
 &= \mathbf{r}^T \mathcal{R}_{\mathbf{z}\mathbf{z}}^{*-1} \mathbf{r}^* - \mathbf{c}^T \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathbf{c}^*, \tag{6.22}
 \end{aligned}$$

which following the approach from [77] can also be expressed as

$$\delta e^2 = \left[\mathbf{p} - \mathcal{P}_{\mathbf{z}\mathbf{z}}^* \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathbf{c}^* \right]^H \left[\mathcal{C}_{\mathbf{z}\mathbf{z}} - \mathcal{P}_{\mathbf{z}\mathbf{z}} \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathcal{P}_{\mathbf{z}\mathbf{z}}^* \right]^{-1} \left[\mathbf{p} - \mathcal{P}_{\mathbf{z}\mathbf{z}} \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathbf{c}^* \right]. \tag{6.23}$$

For further details see Appendix B.1.

6.3.1 Widely Linear Extension to the Standard AR Model

For linear AR modelling of noncircular data the standard linear AR model needs to be extended to incorporate augmented statistics, leading to the widely linear AR (WLAR) model given by

$$z(k) = h_1 z(k-1) + g_1 z^*(k-1) + \dots + h_n z(k-n) + g_n z^*(k-n) + q(k). \tag{6.24}$$

The equivalent of the Yule-Walker equations for the widely linear model are given by (for more detail see Appendix B.2)

$$\begin{bmatrix} \mathbf{h}^* \\ \mathbf{g}^* \end{bmatrix} = \begin{bmatrix} \mathcal{C}_{\mathbf{z}\mathbf{z}} & \mathcal{P}_{\mathbf{z}\mathbf{z}} \\ \mathcal{P}_{\mathbf{z}\mathbf{z}}^* & \mathcal{C}_{\mathbf{z}\mathbf{z}}^* \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{c} \\ \mathbf{p}^* \end{bmatrix}, \quad (6.25)$$

where $\mathbf{h}$ denotes the coefficient vector of the standard complex AR model and $\mathbf{g}$ the coefficient vector of the widely linear part of the WLAR.

To illustrate the validity of this method consider a real valued AR(4) model with coefficient vector $\mathbf{h}_1$ and a complex valued AR(2) example with coefficient vector $\mathbf{h}_2$ [110], given below

$$\mathbf{h}_1 = \begin{bmatrix} 1.79 \\ -1.85 \\ 1.27 \\ -0.41 \end{bmatrix} \quad \mathbf{h}_2 = \begin{bmatrix} 0.72 + 0.50i \\ -0.30 - 0.25i \end{bmatrix}, \quad (6.26)$$

where in all cases $\mathbf{g} = \mathbf{0}$. The widely linear AR solution (6.25) gives the results

$$\mathbf{h}_1 = \begin{bmatrix} 1.7937 - 0.0062i \\ -1.8471 + 0.0057i \\ 1.2631 + 0.0007i \\ -0.4103 + 0.0037i \end{bmatrix} \quad \mathbf{g}_1 = \begin{bmatrix} -0.0044 - 0.0050i \\ 0.0105 - 0.0035i \\ -0.0167 - 0.0161i \\ 0.0098 + 0.0240i \end{bmatrix}, \quad (6.27)$$

$$\mathbf{h}_2 = \begin{bmatrix} 0.7106 + 0.5025i \\ -0.2820 - 0.2548i \end{bmatrix} \quad \mathbf{g}_2 = \begin{bmatrix} -0.0105 + 0.0119i \\ -0.0042 - 0.0052i \end{bmatrix}, \quad (6.28)$$

which are in line with the expected results. The widely linear normal equations (6.25) are now applied to estimate parameters of an AR(4) widely linear model generated from the nonlinear and noncircular Ikeda map (6.11), shown in Fig. 6.6(a). The resulting

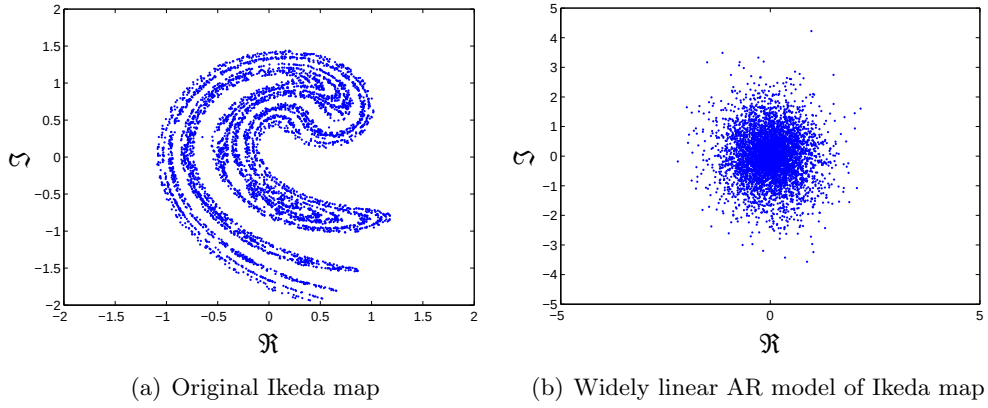


Figure 6.6: Signal distributions of the Ikeda map

widely linear AR(4) coefficient vectors are given by

$$\mathbf{h} = \begin{bmatrix} -0.2529 - 1.0954i \\ 0.4607 - 0.5744i \\ 0.2540 - 0.1165i \\ 0.0596 + 0.0414i \end{bmatrix} \quad \mathbf{g} = \begin{bmatrix} 0.2666 - 0.4688i \\ 0.1178 - 0.5964i \\ -0.0845 - 0.2897i \\ -0.0731 - 0.0848i \end{bmatrix}, \quad (6.29)$$

and the corresponding distribution is shown in Fig 6.6(b).

Comparing the resulting covariance and pseudocovariance functions (Fig. 6.7(b)) with those originally obtained from the Ikeda map (Fig. 6.7(a)) and those obtained from the standard linear AR(4) model where the pseudocovariance is close to zero (Fig. 6.7(c)) it is clear that the normal equations for the widely linear model produce linear noncircular signal distributions, as desired.

6.4 Collaborative Adaptive Filter for Circularity

The proposed approach for the assessment of signal noncircularity is based on a hybrid filter for complex data with a combination of the CLMS (6.6) and augmented CLMS (ACLMS) algorithms [41, 81]. The recently introduced ACLMS algorithm extends the CLMS algorithm to use augmented statistics and thereby utilise the full second order

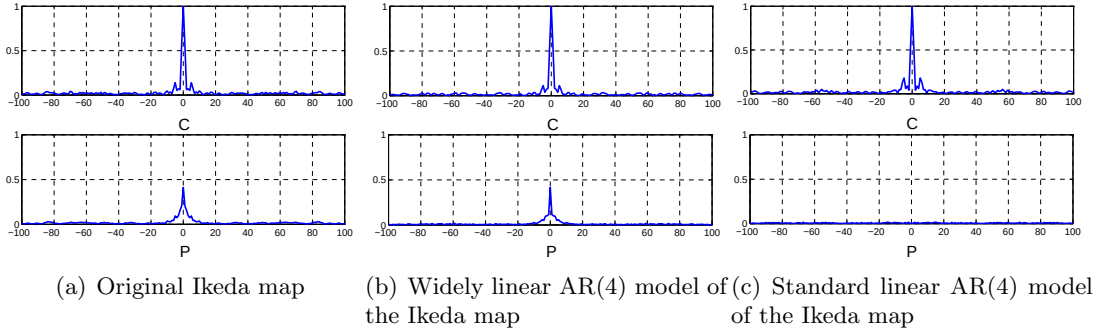


Figure 6.7: Covariance and pseudocovariance of the Ikeda map

statistical information available within the signal. The ACLMS is given by

$$\begin{aligned}
 e_a(k) &= d(k) - y_a(k), \\
 y_a(k) &= \mathbf{h}_a^T(k) \mathbf{z}(k) + \mathbf{g}_a^T(k) \mathbf{z}^*(k), \\
 \mathbf{h}_a(k+1) &= \mathbf{h}_a(k) + \mu e_a(k) \mathbf{z}^*(k), \\
 \mathbf{g}_a(k+1) &= \mathbf{g}_a(k) + \mu e_a(k) \mathbf{z}(k).
 \end{aligned} \tag{6.30}$$

However, this results in a filter which is effectively twice the length of that of the CLMS, resulting in the CLMS having a faster convergence rate than that of the ACLMS. Using a convex combination of these algorithms allows the hybrid filter to take advantage of both the improved performance of the ACLMS for filtering of noncircular signals, along with the faster convergence of the CLMS.

6.4.1 Convergence of the Mixing Parameter

To illustrate the desired response of the hybrid filter consider the behaviour of the mixing parameter $\lambda(k)$ for extremes in signal nature. Without loss in generality assume that the desired response can be expressed as [25]

$$d(k) = \mathbf{h}_o^T \mathbf{z}(k) + \mathbf{g}_o^T \mathbf{z}^*(k) + q(k). \tag{6.31}$$

Whilst, in the steady state both subfilters converge towards the optimal values of their coefficient vectors, $\mathbf{h}_{co}$, $\mathbf{h}_{ao}$ and $\mathbf{g}_{ao}$, and the subfilter outputs become

$$y_c(k) = \mathbf{h}_{co}^T \mathbf{z}(k), \quad (6.32)$$

$$y_a(k) = \mathbf{h}_{ao}^T \mathbf{z}(k) + \mathbf{g}_{ao}^T \mathbf{z}^*(k). \quad (6.33)$$

This then gives two cases:

- **Circular Input Signal:** In this case $\mathbf{h}_{co} = \mathbf{h}_{ao} = \mathbf{h}_o$ and $\mathbf{g}_{ao} = \mathbf{g}_o = 0$, and the steady state instantaneous output error of the hybrid filter becomes

$$\begin{aligned} e(k) &= \mathbf{h}_o^T \mathbf{z}(k) + \mathbf{g}_o^T \mathbf{z}^*(k) + q(k) - \lambda(k)y_c(k) - (1 - \lambda(k))y_a(k) \\ &= \mathbf{h}_o^T \mathbf{z}(k) + \mathbf{g}_o^T \mathbf{z}^*(k) + q(k) - \lambda(k)\mathbf{h}_o^T \mathbf{z}(k) - (1 - \lambda(k))\mathbf{h}_o^T \mathbf{z}(k) \\ &= q(k). \end{aligned} \quad (6.34)$$

Substituting this result into the update of the mixing parameter (6.5) results in

$$\begin{aligned} \lambda(k+1) &= \lambda(k) + \mu_\lambda [e(k)(y_c(k) - y_a(k))^* + e^*(k)(y_c(k) - y_a(k))] \\ &= \lambda(k) + \mu_\lambda q(k) (\mathbf{h}_o^H \mathbf{z}^*(k) - \mathbf{h}_o^H \mathbf{z}^*(k)) + \mu_\lambda q^*(k) (\mathbf{h}_o^T \mathbf{z}(k) - \mathbf{h}_o^T \mathbf{z}(k)) \\ &= \lambda(k) \end{aligned} \quad (6.35)$$

In practice, the CLMS algorithm converges towards $\mathbf{h}_o$ faster than the ACLMS, and therefore the value of λ initially favours that of the CLMS before reaching the steady state.

- **Noncircular Input Signal:** For noncircular inputs $\mathbf{h}_{ao} = \mathbf{h}_o \neq \mathbf{h}_{co}$ and $\mathbf{g}_{ao} = \mathbf{g}_o$, and the output error becomes

$$\begin{aligned} e(k) &= \mathbf{h}_o^T \mathbf{z}(k) + \mathbf{g}_o^T \mathbf{z}^*(k) + q(k) - \lambda(k)y_c(k) - (1 - \lambda(k))y_a(k) \\ &= \mathbf{h}_o^T \mathbf{z}(k) + \mathbf{g}_o^T \mathbf{z}^*(k) + q(k) - \lambda(k)\mathbf{h}_{co}^T \mathbf{z}(k) - (1 - \lambda(k))(\mathbf{h}_o^T \mathbf{z}(k) + \mathbf{g}_o^T \mathbf{z}^*(k)) \\ &= q(k) + \lambda(k) (\mathbf{h}_o^T \mathbf{z}(k) + \mathbf{g}_o^T \mathbf{z}^*(k) - \mathbf{h}_{co}^T \mathbf{z}(k)) \end{aligned} \quad (6.36)$$

Substituting this result into the update of the mixing parameter (6.5) results in

$$\begin{aligned}
\lambda(k+1) &= \lambda(k) + \mu_\lambda \left[e(k)(y_c(k) - y_a(k))^* + e^*(k)(y_c(k) - y_a(k)) \right] \\
&= \lambda(k) + \mu_\lambda e(k)y_c^*(k) - \mu_\lambda e(k)y_a^*(k) + \mu_\lambda e^*(k)y_c(k) - \mu_\lambda e^*(k)y_a(k) \\
&= \lambda(k) + \mu_\lambda \left[q(k) + \lambda(k)(\mathbf{h}_o^T \mathbf{z}(k) + \mathbf{g}_o^T \mathbf{z}^*(k) - \mathbf{h}_{co}^T \mathbf{z}(k)) \right] \mathbf{h}_{co}^H \mathbf{z}^*(k) \\
&\quad - \mu_\lambda \left[q(k) + \lambda(k)(\mathbf{h}_o^T \mathbf{z}(k) + \mathbf{g}_o^T \mathbf{z}^*(k) - \mathbf{h}_{co}^T \mathbf{z}(k)) \right] (\mathbf{h}_o^H \mathbf{z}^*(k) + \mathbf{g}_o^H \mathbf{z}(k)) \\
&\quad + \mu_\lambda \left[q^*(k) + \lambda(k)(\mathbf{h}_o^H \mathbf{z}^*(k) + \mathbf{g}_o^H \mathbf{z}(k) - \mathbf{h}_{co}^H \mathbf{z}^*(k)) \right] \mathbf{h}_{co}^T \mathbf{z}(k) \\
&\quad - \mu_\lambda \left[q^*(k) + \lambda(k)(\mathbf{h}_o^H \mathbf{z}^*(k) + \mathbf{g}_o^H \mathbf{z}(k) - \mathbf{h}_{co}^H \mathbf{z}^*(k)) \right] (\mathbf{h}_o^T \mathbf{z}(k) \\
&\quad + \mathbf{g}_o^T \mathbf{z}^*(k)) \\
&= \lambda(k) + \mu_\lambda \left[q(k)\mathbf{h}_{co}^H \mathbf{z}^*(k) + q^*(k)\mathbf{h}_{co}^T \mathbf{z}(k) - q(k)\mathbf{h}_o^H \mathbf{z}^*(k) \right. \\
&\quad - q^*(k)\mathbf{h}_o^T \mathbf{z}(k) - q(k)\mathbf{g}_o^H \mathbf{z}(k) - q^*(k)\mathbf{g}_o^T \mathbf{z}^*(k) \\
&\quad + 2\lambda(k) \left(\mathbf{h}_{co}^T \mathbf{z}(k)\mathbf{h}_o^H \mathbf{z}^*(k) + \mathbf{h}_{co}^T \mathbf{z}(k)\mathbf{g}_o^H \mathbf{z}(k) - \mathbf{h}_{co}^T \mathbf{z}(k)\mathbf{h}_{co}^H \mathbf{z}^*(k) \right. \\
&\quad + \mathbf{h}_{co}^H \mathbf{z}^*(k)\mathbf{h}_o^T \mathbf{z}(k) + \mathbf{h}_{co}^H \mathbf{z}^*(k)\mathbf{g}_o^T \mathbf{z}^*(k) - \mathbf{h}_o^T \mathbf{z}(k)\mathbf{h}_o^H \mathbf{z}^*(k) \\
&\quad \left. \left. - \mathbf{h}_o^T \mathbf{z}(k)\mathbf{g}_o^H \mathbf{z}(k) - \mathbf{g}_o^T \mathbf{z}^*(k)\mathbf{h}_o^H \mathbf{z}^*(k) - \mathbf{g}_o^T \mathbf{z}^*(k)\mathbf{g}_o^H \mathbf{z}(k) \right) \right]. \quad (6.37)
\end{aligned}$$

Taking the expectation and employing the independence assumptions yields

$$\begin{aligned}
E[\lambda(k+1)] &= E[\lambda(k)] \left[1 + 2\mu \left(E[\mathbf{h}_{co}^T \mathbf{z}(k)\mathbf{z}^H \mathbf{h}_o^*] + E[\mathbf{h}_{co}^T \mathbf{z}(k)\mathbf{z}^T(k)\mathbf{g}_o^*] \right. \right. \\
&\quad - E[\mathbf{h}_{co}^T \mathbf{z}(k)\mathbf{z}^H(k)\mathbf{h}_{co}] + E[\mathbf{h}_{co}^H \mathbf{z}^*(k)\mathbf{z}^T(k)\mathbf{h}_o] \\
&\quad + E[\mathbf{h}_{co}^H \mathbf{z}^*(k)\mathbf{z}^H(k)\mathbf{g}_o] - E[\mathbf{h}_o^T \mathbf{z}(k)\mathbf{z}^H(k)\mathbf{h}_o^*] - E[\mathbf{h}_o^T \mathbf{z}(k)\mathbf{z}^T(k)\mathbf{g}_o^*] \\
&\quad \left. \left. - E[\mathbf{g}_o^T \mathbf{z}^*(k)\mathbf{z}^H(k)\mathbf{h}_o^*] - E[\mathbf{g}_o^T \mathbf{z}^*(k)\mathbf{z}^T(k)\mathbf{g}_o^*] \right) \right]. \quad (6.38)
\end{aligned}$$

If we define $\mathbf{A} = \left(\mathcal{C}_{\mathbf{z}\mathbf{z}} - \mathcal{P}_{\mathbf{z}\mathbf{z}} \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathcal{P}_{\mathbf{z}\mathbf{z}}^* \right)^{-1}$ and $\mathbf{A}^* = \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} + \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathcal{P}_{\mathbf{z}\mathbf{z}}^* \mathbf{A} \mathcal{P}_{\mathbf{z}\mathbf{z}} \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1}$,

then from (B.4) and (B.5)

$$\begin{aligned}
\mathbf{h}_o &= [\mathcal{C}_{zz}^{*-1} + \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* \mathbf{A} \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1}] [\mathbf{c}^* - \mathcal{P}_{zz}^* \mathcal{C}_{zz}^{-1} \mathbf{p}] \\
&= \mathcal{C}_{zz}^{*-1} \mathbf{c}^* - \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* \mathcal{C}_{zz}^{-1} \mathbf{p} + \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* \mathbf{A} \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathbf{c}^* - \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* \mathbf{A} \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* \mathcal{C}_{zz}^{-1} \mathbf{p}, \\
&= \mathcal{C}_{zz}^{*-1} \mathbf{c}^* - \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* [\mathcal{C}_{zz}^{-1} + \mathbf{A} \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* \mathcal{C}_{zz}^{-1}] \mathbf{p} + \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* \mathbf{A} \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathbf{c}^* \\
&= \mathcal{C}_{zz}^{*-1} \mathbf{c}^* - \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* \mathbf{A} \mathbf{p} + \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* \mathbf{A} \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathbf{c}^* \tag{6.39}
\end{aligned}$$

$$\mathbf{h}_o^* = \mathbf{A} [\mathbf{c} - \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathbf{p}^*] \tag{6.40}$$

$$\mathbf{g}_o = \mathbf{A} [\mathbf{p} - \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathbf{c}^*] \tag{6.41}$$

$$\begin{aligned}
\mathbf{g}_o^* &= [\mathcal{C}_{zz}^{*-1} + \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* \mathbf{A} \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1}] [\mathbf{p}^* - \mathcal{P}_{zz}^* \mathcal{C}_{zz}^{-1} \mathbf{c}] \\
&= \mathcal{C}_{zz}^{*-1} \mathbf{p}^* - \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* \mathcal{C}_{zz}^{-1} \mathbf{c} + \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* \mathbf{A} \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathbf{p}^* - \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* \mathbf{A} \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* \mathcal{C}_{zz}^{-1} \mathbf{c}, \\
&= \mathcal{C}_{zz}^{*-1} \mathbf{p}^* - \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* [\mathcal{C}_{zz}^{-1} + \mathbf{A} \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* \mathcal{C}_{zz}^{-1}] \mathbf{c} + \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* \mathbf{A} \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathbf{p}^* \\
&= \mathcal{C}_{zz}^{*-1} \mathbf{p}^* - \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* \mathbf{A} \mathbf{c} + \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* \mathbf{A} \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathbf{p}^* \tag{6.42}
\end{aligned}$$

Substituting these into the expectations from (6.38) (for more detail see Appendix C) results in the expression for the expectation of λ given by

$$\begin{aligned}
E[\lambda(k+1)] &= E[\lambda(k)] \left(1 + 2\mu \left[-\mathbf{c}^H \mathcal{C}_{zz}^{-1} \mathcal{P}_{zz}^* \mathbf{A}^T \mathcal{P}_{zz} \mathbf{A} \mathbf{c} + \mathbf{p}^T \mathbf{A}^T \mathcal{P}_{zz} \mathbf{A} \mathbf{c} - \mathbf{p}^T \mathbf{A}^T \mathbf{p}^* \right. \right. \\
&\quad \left. \left. + \mathbf{c}^H \mathcal{C}_{zz}^{-1} \mathcal{P}_{zz}^* \mathbf{A}^T \mathbf{p}^* + \mathbf{c}^H \mathcal{C}_{zz}^{-1} \mathcal{P}_{zz}^* \mathbf{A}^T \mathcal{P}_{zz} \mathcal{C}_{zz}^{-1} \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* \mathbf{A} \mathbf{c} \right. \right. \\
&\quad \left. \left. - \mathbf{p}^T \mathbf{A}^T \mathcal{P}_{zz} \mathcal{C}_{zz}^{-1} \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* \mathbf{A} \mathbf{c} \right] \right) \\
&= E[\lambda(k)] \left(1 + 2\mu \left[\mathbf{p}^T \mathbf{A}^T (\mathcal{P}_{zz} \mathbf{A} \mathbf{c} - \mathbf{p}^* - \mathcal{P}_{zz} \mathcal{C}_{zz}^{-1} \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* \mathbf{A} \mathbf{c}) \right. \right. \\
&\quad \left. \left. + \mathbf{c}^H \mathcal{C}_{zz}^{-1} \mathcal{P}_{zz}^* \mathbf{A}^T (\mathbf{p}^* + \mathcal{P}_{zz} \mathcal{C}_{zz}^{-1} \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* \mathbf{A} \mathbf{c} - \mathcal{P}_{zz} \mathbf{A} \mathbf{c}) \right] \right) \\
&= E[\lambda(k)] \left(1 + 2\mu \left[(\mathbf{p}^T - \mathbf{c}^H \mathcal{C}_{zz}^{-1} \mathcal{P}_{zz}^*) \mathbf{A}^T (\mathcal{P}_{zz} \mathbf{A} \mathbf{c} - \mathbf{p}^* \right. \right. \\
&\quad \left. \left. - \mathcal{P}_{zz} \mathcal{C}_{zz}^{-1} \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* \mathbf{A} \mathbf{c}) \right] \right)
\end{aligned}$$

$$\begin{aligned}
&= E[\lambda(k)] \left(1 + 2\mu \left[(\mathbf{p}^T - \mathbf{c}^H \mathcal{C}_{\mathbf{zz}}^{-1} \mathcal{P}_{\mathbf{zz}}^*) \mathbf{A}^T (\mathcal{P}_{\mathbf{zz}} [\mathbf{I} \right. \right. \\
&\quad \left. \left. - \mathcal{C}_{\mathbf{zz}}^{-1} \mathcal{P}_{\mathbf{zz}} \mathcal{C}_{\mathbf{zz}}^{*-1} \mathcal{P}_{\mathbf{zz}}^*] \mathbf{A} \mathbf{c} - \mathbf{p}^*) \right] \right) \\
&= E[\lambda(k)] \left(1 + 2\mu \left[(\mathbf{p}^T - \mathbf{c}^H \mathcal{C}_{\mathbf{zz}}^{-1} \mathcal{P}_{\mathbf{zz}}^*) \mathbf{A}^T (\mathcal{P}_{\mathbf{zz}} \mathcal{C}_{\mathbf{zz}}^{-1} [\mathcal{C}_{\mathbf{zz}} \right. \right. \\
&\quad \left. \left. - \mathcal{P}_{\mathbf{zz}} \mathcal{C}_{\mathbf{zz}}^{*-1} \mathcal{P}_{\mathbf{zz}}^*] \mathbf{A} \mathbf{c} - \mathbf{p}^*) \right] \right) \\
&= E[\lambda(k)] \left(1 + 2\mu \left[(\mathbf{p}^T - \mathbf{c}^H \mathcal{C}_{\mathbf{zz}}^{-1} \mathcal{P}_{\mathbf{zz}}^*) \mathbf{A}^T (\mathcal{P}_{\mathbf{zz}} \mathcal{C}_{\mathbf{zz}}^{-1} \mathbf{c} - \mathbf{p}^*) \right] \right) \\
&= E[\lambda(k)] \left(1 + 2\mu \left[(\mathcal{P}_{\mathbf{zz}} \mathcal{C}_{\mathbf{zz}}^{-1} \mathbf{c} - \mathbf{p}^*)^T \mathbf{A} (\mathbf{p}^T - \mathbf{c}^H \mathcal{C}_{\mathbf{zz}}^{-1} \mathcal{P}_{\mathbf{zz}}^*)^T \right] \right) \\
&= E[\lambda(k)] \left(1 + 2\mu \left[(\mathcal{P}_{\mathbf{zz}} \mathcal{C}_{\mathbf{zz}}^{-1} \mathbf{c} - \mathbf{p}^*)^T \mathbf{A} (\mathbf{p} - \mathcal{P}_{\mathbf{zz}} \mathcal{C}_{\mathbf{zz}}^{*-1} \mathbf{c}^*) \right] \right) \\
&= E[\lambda(k)] \left(1 + 2\mu \left[(\mathcal{P}_{\mathbf{zz}} \mathcal{C}_{\mathbf{zz}}^{-1} \mathbf{c} - \mathbf{p}^*)^T (\mathcal{C}_{\mathbf{zz}} - \mathcal{P}_{\mathbf{zz}} \mathcal{C}_{\mathbf{zz}}^{*-1} \mathcal{P}_{\mathbf{zz}}^*)^{-1} (\mathbf{p} \right. \right. \\
&\quad \left. \left. - \mathcal{P}_{\mathbf{zz}} \mathcal{C}_{\mathbf{zz}}^{*-1} \mathbf{c}^*) \right] \right) \\
&= E[\lambda(k)] \left(1 - 2\mu \delta e^2 \right), \tag{6.43}
\end{aligned}$$

where δe^2 is the advantage of the widely linear model over the standard linear model (6.23). Therefore, $\lambda(k)$ converges towards zero (the ACLMS subfilter) for all situations where the widely linear model has an advantage over the standard linear model i.e. when the input signal is noncircular. In the situation where there is no advantage obtained from the widely linear model (6.43) reduces to (6.35).

6.4.2 Performance of the Hybrid Filter

Initial simulations were performed to illustrate the behaviour of the subfilters of the hybrid filter for both circular and noncircular data, and highlight the advantages of using the hybrid filter over either of the constituent subfilters. The circular signal used was the stable AR(4) process (3.8) where $n(k)$ is a random uniform circular variable given by

$$n(k) = \rho(k) \cos(\theta(k)) + j\rho(k) \sin(\theta(k)) \tag{6.44}$$

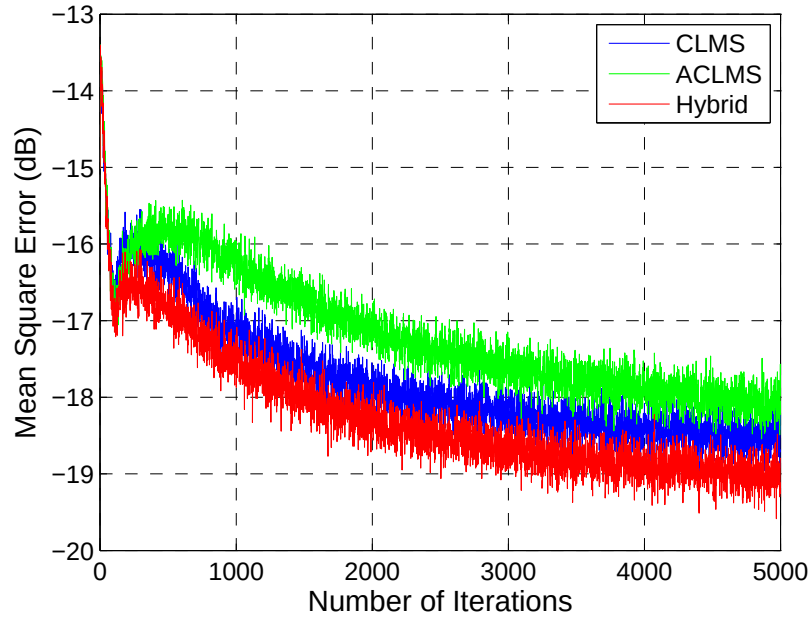


Figure 6.8: Convergence curves of the hybrid filter and constituent subfilters for a circular AR(4) process (3.8)

and ρ and θ are random variables uniformly distributed on the unit interval and $[0, 2\pi]$ respectively. The noncircular data used was the Ikeda map (6.11). For the simulations filters of length $N = 100$ were used, the performance of the filters was assessed based on $10 \log_{10} e^2(k)$ averaged over a set of 1000 independent trials with an average variance between trials of between 0.5 and 1.5.

From Fig. 6.8 it can be seen that both the CLMS and ACLMS perform well on the circular data, with the CLMS converging slightly faster than the ACLMS. This is to be expected as the ACLMS is in effect twice the length of the CLMS and therefore takes longer to converge. The key point to note here is that in this instance the hybrid filter outperforms both of its constituent subfilters converging slightly faster and with an improved steady state error. For the noncircular data, shown in Fig. 6.9, the CLMS again converges faster than the ACLMS but in this instance does not perform as well in the steady state as the ACLMS. Again, the hybrid filter outperforms both subfilters by following first the CLMS and then the slower ACLMS when it surpasses the CLMS to give a better combined output.

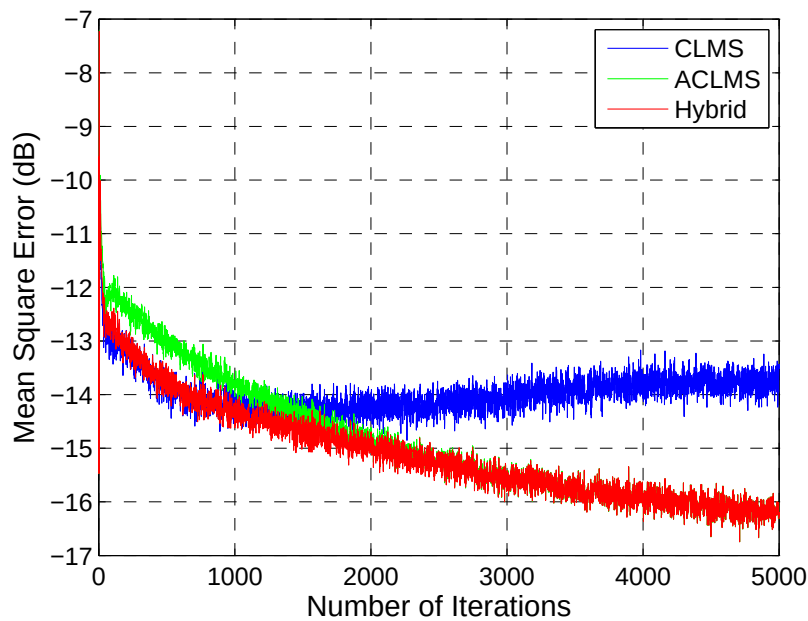


Figure 6.9: Convergence curves of the hybrid filter and constituent subfilters for a noncircular Ikeda map (6.11)

6.5 Tracking Circularity of a Signal

Moving on to consider the tracking of the circularity of signals recall that a value of $\lambda = 1$ corresponds to the hybrid filter being dominated by the CLMS subfilter which relates to a circular signal, whereas a value of $\lambda = 0$ means the hybrid filter is dominated by the ACLMS subfilter which indicates a noncircular signal.

Figure 6.10 shows the evolution of the mixing parameter λ for a linear AR(4) signal and the widely linear AR(4) model of the Ikeda map (6.29), both driven by doubly white noise, and the noncircular Ikeda map. It can be seen that for the linear AR(4) signal the value of the mixing parameter is between 0.8 and 0.9 throughout, corresponding to the hybrid filter being dominated by the CLMS. Alternatively, for the Ikeda Map the value of the mixing parameter moves towards zero showing the filter is dominated almost entirely by the widely linear ACLMS. This is as expected and supports the observed outputs of the hybrid filter from the previous section. For the widely linear AR(4) signal, the mixing parameter moves initially towards $\lambda = 1$ indicating the faster convergence of the CLMS, but after this initial spike it settles to approximately $\lambda = 0.1$

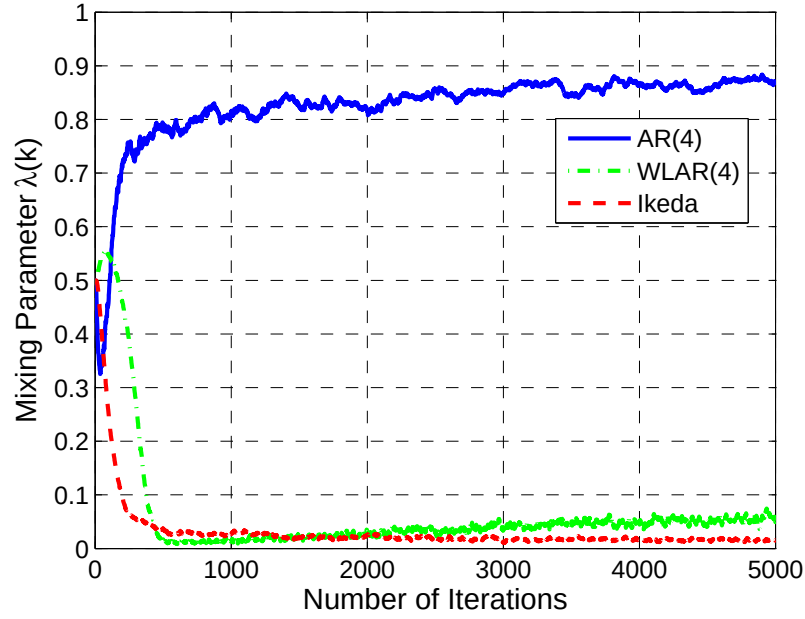


Figure 6.10: Evolution of the mixing parameter λ for the linear circular AR(4) process, the widely linear AR(4) model of the Ikeda map (noncircular) and the noncircular Ikeda signals

showing that the signal is better modelled by the ACLMS. To investigate the effect of the driving noise, in the next set of experiments the standard AR(4) model was driven by doubly white circular noise, doubly white noncircular noise and noncircular noise. The results are shown in Fig. 6.11. Conforming with the analysis in (6.20)–(6.23), both signals driven by the doubly white noise tend towards 1 and are therefore better represented by the CLMS. This is due to the fact that while neither subfilter has a steady state advantage over the other the CLMS has a faster convergence. The signal driven by noncircular noise, however, results in a noncircular signal and is better represented by the ACLMS algorithm, hence the value of λ approaches 1. To illustrate the ability to track changes in the circularity of a signal, the input signal was alternated between the circular AR(4) signal and the Ikeda map in the first setting, and between the AR(4) signal and the widely linear AR(4) model of the Ikeda map in the second setting. The results shown in Fig. 6.12 show that the hybrid filter effectively tracks these changes as illustrated by the values of the convex mixing parameter λ following the circularity of the nature of the signal. Figure 6.13 shows the behaviour of λ when the period

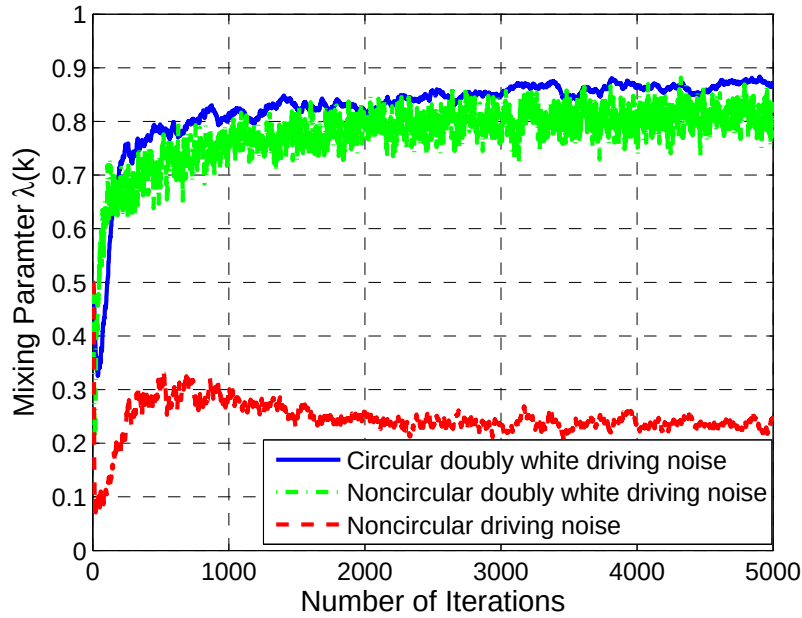


Figure 6.11: Evolution of the mixing parameter λ for a linear circular AR(4) signal driven by different noises

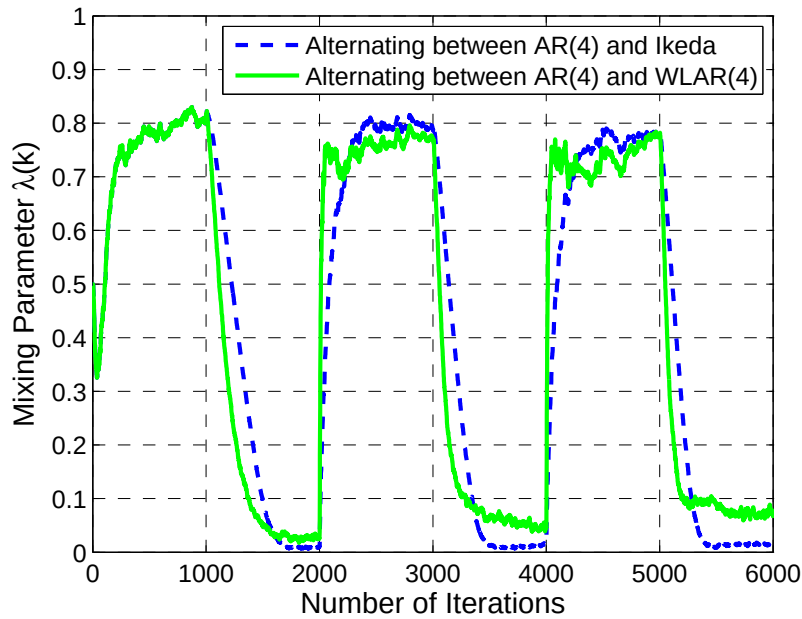


Figure 6.12: Evolution of the mixing parameter λ for signals alternating from linear AR(4) to either widely linear AR(4) model of Ikeda map or Ikeda map

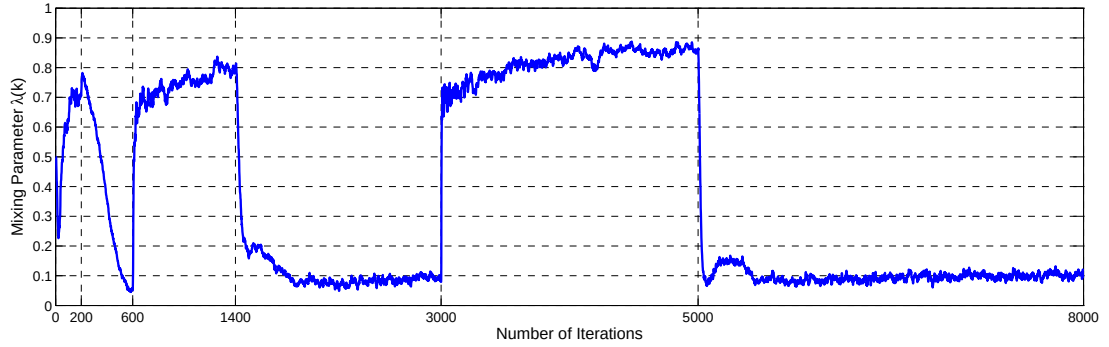


Figure 6.13: Evolution of the mixing parameter λ for a signal alternating from linear AR(4) to widely linear AR(4) with a varying alternation period

over which the signal is alternating changes over time, illustrating the flexibility of the proposed test.

6.6 Conclusions

Signal modality characterisation through hybrid filtering has been extended to the complex domain. Tests for nonlinearity and the split-complex or fully-complex nature of nonlinearity of complex valued signals have been proposed, with simulations on benchmark synthetic data supporting the approach. A hybrid filter consisting of a convex combination of the CLMS and ACLMS algorithms has been introduced. The hybrid filter takes advantage of the faster convergence speeds of the CLMS and the improved performance in the steady state for noncircular data of the ACLMS. It has been shown that the hybrid filter can outperform both of the subfilters for synthetic circular and noncircular data. Using this collaborative adaptive filtering approach an online adaptive test for the circularity of a signal has been produced. It has been shown that for circular data the mixing parameter λ within this structure will favour the CLMS algorithm and for noncircular data it favours the ACLMS algorithm. The use of AR processes for the generation of circular and noncircular data has been illustrated, through the use of both the standard and widely linear AR models. With simulations on the thus generated AR and widely linear AR processes it has been shown that through this method it is possible to both identify and track the circularity of a signal.

Chapter 7

Real World Applications

7.1 Introduction

The previous chapters have illustrated a collaborative adaptive filtering approach to signal modality characterisation on synthetic benchmark data. Having shown the efficacy of the hybrid filter structure at identifying and tracking various different signal characteristics this chapter considers the characterisation of real world signals. In the real domain $\mathbb{R}$ various biomedical signals are considered. Firstly, two different sets of electroencephalography (EEG) data are analysed, one set for a group of coma patients, the second for epileptic seizure patients. For the epileptic patients the aim is to identify the onset of epileptic seizures. For this both the nonlinearity and the sparsity of the data is considered and a comparison of the two given. For the coma patients the nonlinearity of the signal is considered with the aim of identifying whether the patients are merely in a coma or whether they may also be considered brain dead. Following this the sparsity of electromyography (EMG) data is considered, before moving on to consider both the nonlinearity and sparsity of speech.

In the complex domain $\mathbb{C}$ the first example used is a set of maritime radar data for different sea states, in which the fully-complex/split-complex nature is considered. Finally a set of wind data is analysed, both in terms of the nonlinearity of the data and the nature of the nonlinearity. Finally, a comparison of the performance of the different algorithms for the detection of circularity is given.

7.2 EEG Data

The accurate assessment of EEG data can reveal more about the health conditions in question, with measures of nonlinearity proving useful for the classification of EEG data [18, 30, 95] and as such online assessment can provide a valuable tool. For the purpose of identifying the nature of EEG signals, two sets of data showing the onset of epileptic seizures are analysed, both in terms of the nonlinearity and sparsity of the data, along with the comparison of the two responses. For epileptic data it has been shown that the optimal embedding dimension of epileptic data is larger than normal data and that the embedding dimension varies largely during seizures and suggests that epileptic data is more deterministic [109].

The second group of data analysed for its nonlinearity considers the brain consciousness state of several different patients. The accurate assessment of the coma data is particularly key as there are obvious legal implications of declaring a patient brain dead. Testing for brain death can require the patients to be removed from life support equipment, need repeating several times and put undue stress on already compromised organs. To address some of these issues a preliminary EEG test has been proposed [15] to determine whether further medical tests are required. If further tests are deemed to be required, a diagnosis of quasi-brain death is given until a final diagnosis is made.

7.2.1 Epileptic Seizure Data

The epileptic seizure data contains 16 channels which for the purpose of the analysis were averaged together to give a single channel. This averaged electrode amplitude was presented as the input to both the hybrid filter consisting of the convex combination of the NNGD and NLMS and also to the convex combination of the SSLMS and NLMS. Figure 7.1 shows the averaged electrode amplitude, along with the corresponding evolution of the mixing parameters λ for both hybrid filters and the resultant changes in nonlinearity against sparsity. These results show the proposed approach can effectively detect changes in the nature of the EEG signals which can be very difficult to achieve otherwise, and can also differentiate between changes in nonlinearity and sparsity.

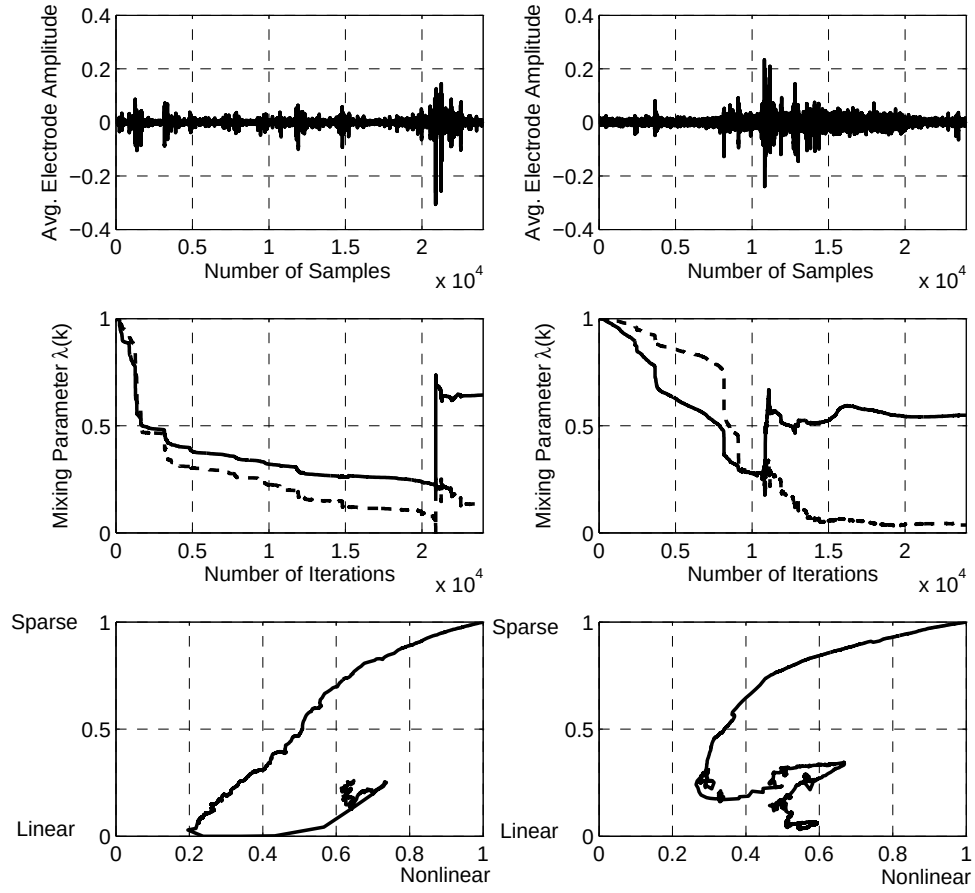


Figure 7.1: Top panels: EEG epileptic seizure data. Middle panels: corresponding evolution of λ solid line: nonlinear hybrid filter, broken line: sparse hybrid filter. Bottom panels comparison of nonlinearity and sparsity, evolution over time starting from coordinates (1,1)

7.2.2 Identification of Consciousness States

The coma data¹ consisted of seven channels sampled at 1000Hz, recorded from electrodes placed as in Fig. 7.2, where the two electrodes placed on the ears (A1 and A2 respectively) act as a reference for the measurements. In this case both the averaged electrode data and the first channel of the EEG data were considered. To assess the nonlinearity of the EEG data the hybrid filter consisting of the convex combination of the NNGD and NLMS was used. Eight sets of data were used, one awake patient, three coma patients and four patients which have been assessed by existing medical tests performed at the hospital to be quasi-brain dead. Figure 7.3 and Fig. 7.5 show the data sets for each consciousness state for one channel and the averaged channel respectively. For all the sets of data, 10 seconds, 1 minute and 5 minutes of data were considered with Fig. 7.4 showing the results from the the first channel and Fig. 7.6 the averaged electrode channel. Comparing the results from Fig. 7.4 and Fig. 7.6 shows similar results are obtained for

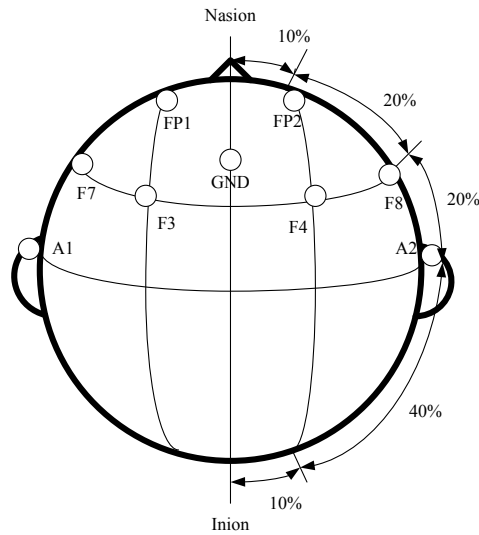


Figure 7.2: The Electrode placement.

the assessment of brain consciousness whether using only the first channel of the EEG data or the average of all the channels. It is clear from these results that there is a difference in the nature of the EEG signals for patients that are awake compared to those that are considered to be brain dead. For the coma patients, however, there can be more variation in the nature of the signal over time, this is particularly clear in Fig. 7.6(c) where one of the coma signals goes from being predominantly linear to predominantly nonlinear then back again. These results illustrate the possibility of using a collaborative adaptive approach for the assessment of brain consciousness states and the advantage over offline tests. Due to their nature offline tests analyse only short

¹The data was provided by Prof. Jianting Cao at Brain Science Institute, RIKEN, Japan and was recorded at HuaShan Hospital, Shanghai, China.

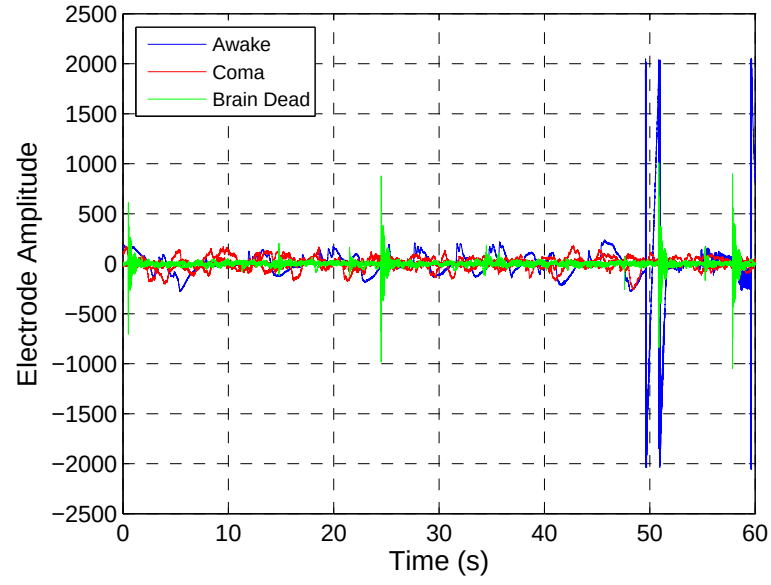
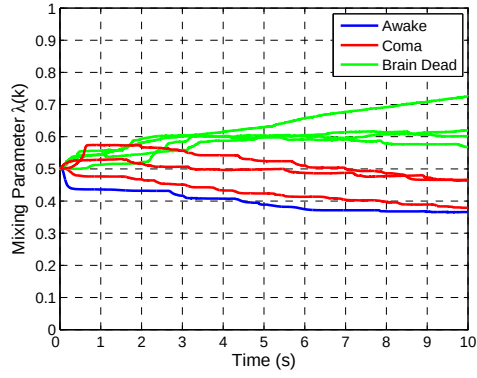
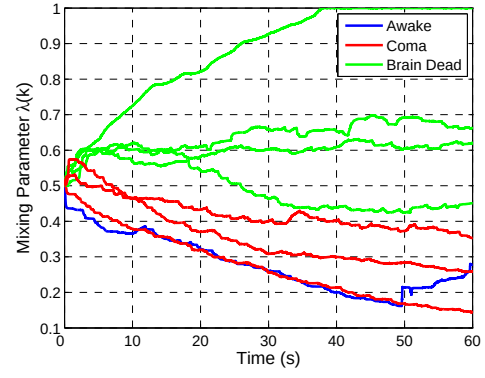


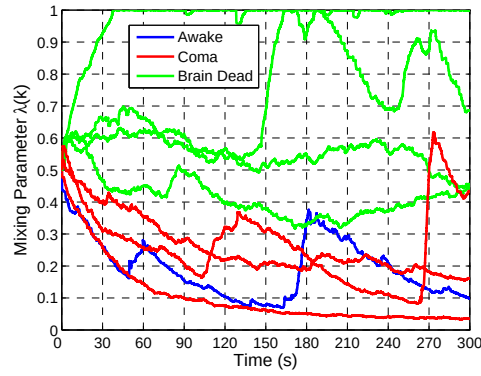
Figure 7.3: One minute of each data set for one EEG channel.



(a) 10 secs.



(b) 1 min.



(c) 5 min.

Figure 7.4: Evolution of mixing parameter λ for different states of coma for one EEG channel ($\lambda \rightarrow 0$ linear, $\lambda \rightarrow 1$ nonlinear).

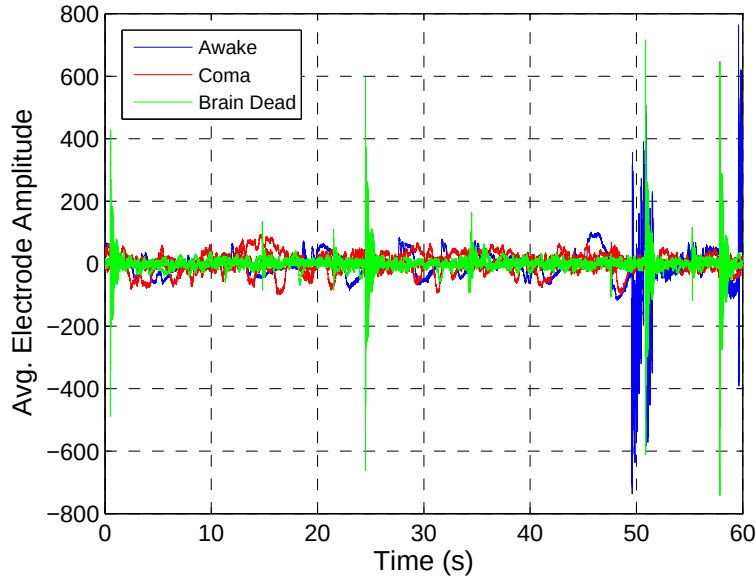
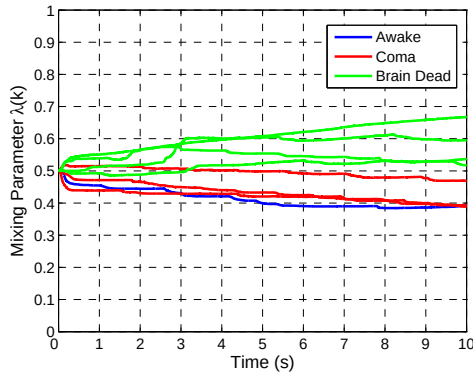
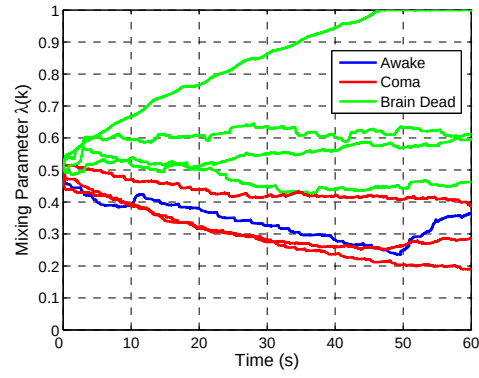


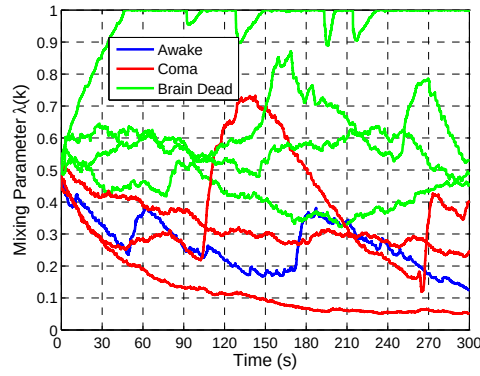
Figure 7.5: One minute of each data set averaged across all EEG channels.



(a) 10 secs.



(b) 1 min.



(c) 5 min.

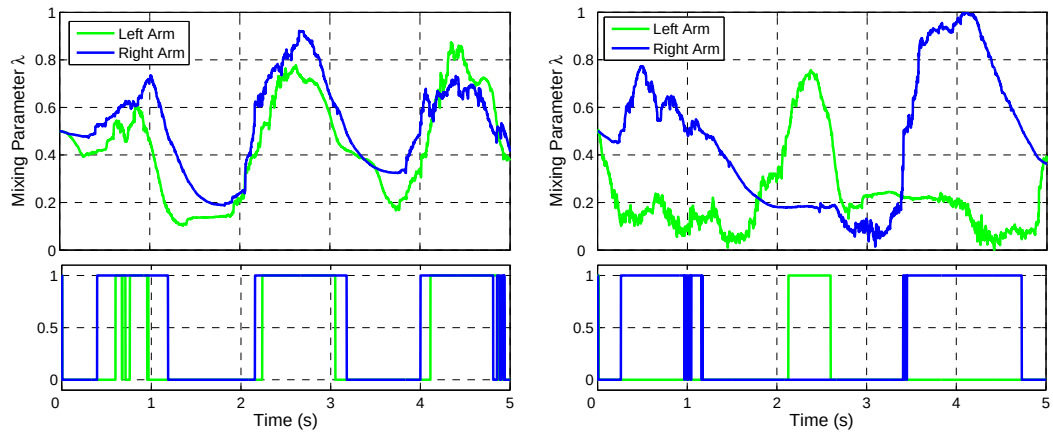
Figure 7.6: Evolution of mixing parameter λ for different states of coma averaged across all EEG channels ($\lambda \rightarrow 0$ linear, $\lambda \rightarrow 1$ nonlinear).

term sections of the data making it possible to incorrectly assess the long term nature of the signal. Even testing several sections of data can miss underlying long term trends in the data. Further results inline with those given here are presented in [49] including error bounds and support vector machine classification (SVM) results.

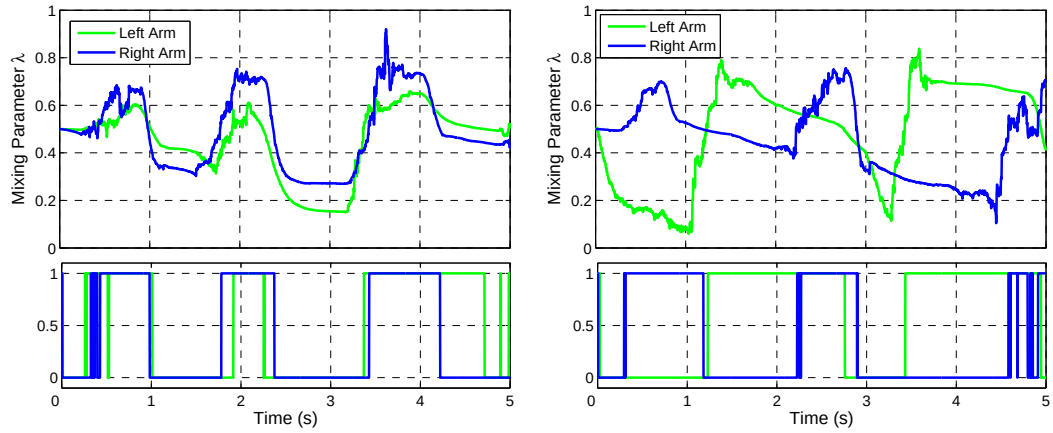
7.3 EMG Data

EMG data records the potentials created by muscles when they contract, and has attracted much interest in relation to prosthetics and medical diagnosis, and therefore the classification of EMG data [34,52]. One other area which has received attention is in the control of robotics. EMG data has been used to provide an interface for controlling an electronic car based on the frequency of the signal [8] and to control a robot for use in an electronic wheelchair using empirical mode decomposition (EMD) [72]. Whilst these methods have proved effective, the filtering method requires tuning for each individual operator and the EMD based approach is a block based method and as such for effective online operation relies on the length of the block used. By using a collaborative adaptive filtering approach, provides a simple online test for whether a muscle is tensed or not.

The EMG data was recorded from the arms of two different subjects for 5 second periods sampled at 512Hz. The electrodes were placed on the forearms of each subject, for the first experiment the subject tensed then relaxed both arms at the same time, in the second experiment the subjects alternated tensing each arm in turn. For the tracking of the nature of the EMG data the hybrid filter from Chapter 5 using the combination of the SSLMS and NLMS was used to test for sparsity. The choice of hybrid filter can either be made based on existing knowledge about the data or alternatively testing initially for a more general characteristic such as nonlinearity before more specific tests. In this case it is common practice to analyse surface EMGs using nonlinear methods and studies have shown good results can be obtained using sparse methods [97]. Figure 7.7 shows the evolution of the mixing parameter λ in the hybrid filter structure along with the corresponding results when the value of λ is given a discrete value zero or one. From these initial results the hybrid filter provides an accu-



(a) Subject tensing both arms at the same time. (b) Subject alternating tensing each arm in turn.



(c) Subject tensing both arms at the same time. (d) Subject alternating tensing each arm in turn.

Figure 7.7: Evolution of the mixing parameter λ for EMG data with the subject tensing and relaxing arm muscles and corresponding response when λ takes discrete values.

rate assessment of the nature of the EMG data being processed and as such whether the muscle was tensed or not. However as can be seen for the second subject when alternating tensing each arm in turn (Fig. 7.7(d)) there is some overlap between the left arm relaxing and the right arm tensing. This can be attributed to the contraction in the muscle in the left arm being less pronounced than that in the right arm, a problem found to some extent in all the experiments and to be expected as the subjects were right handed.

7.4 Speech

The area of speech modality characterisation is only just emerging despite extensive body of work on speech processing. Characterisation of speech is not a straightforward task due to the nonstationary nature of speech, to overcome this recently much effort has been put in developing accurate models for the speech production system. Accurate knowledge of speech characteristics will prove beneficial in several areas of speech processing, including speech coding and speech synthesis. Consider the vocal tract, the vocal tract is the cavity where sound is filtered, it starts at the vocal folds (vocal cords) and consists of the laryngeal cavity, the pharynx, the oral and nasal cavities. The vocal tract is normally modelled as an all-pole filter, i.e. using a linear difference equation, which provides a straightforward well understood model for the vocal tract. Several studies have suggested that linear models do not sufficiently model the human vocal tract [90,98] and considering the physical nature of the vocal tract itself it seems natural to consider the nature of speech in a nonlinear framework.

Studies into the nonlinearity of speech have typically been based on classification between vowels and consonants or between voiced and unvoiced sounds². It is known that all vowels and certain consonants are voiced, voiced sounds are highly periodic in nature with a periodic excitation source. In the case of unvoiced consonants, the folds may be completely open, for example the /s/, /sh/ and /f/ sounds, or partially open, for /h/ sound, resulting in a noise like waveform [21,78,79]. While the majority

²A sound is considered “voiced” when the vocal folds are vibrating, otherwise it is considered “unvoiced”.

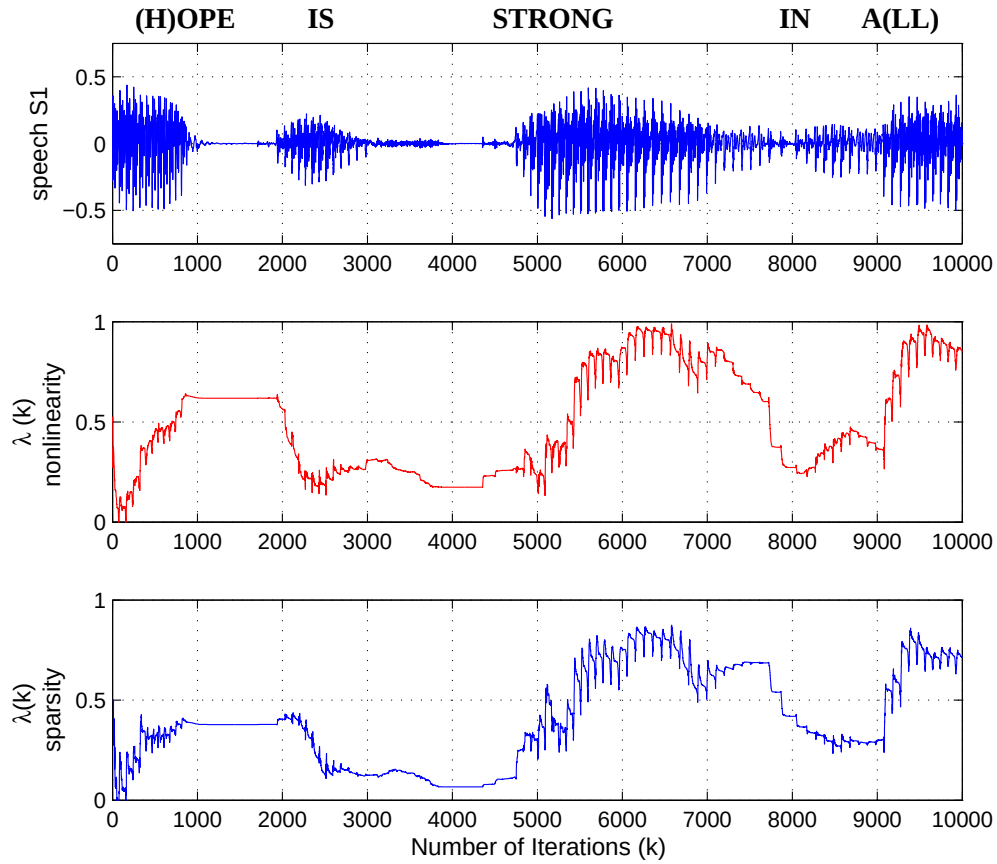


Figure 7.8: Speech signal S1 and corresponding evolution of mixing parameter λ for nonlinearity and sparsity tracking.

of the studies so far have suggested a nonlinear nature of voiced speech, the form of fundamental nonlinearity is still unknown, Kubin [48] suggests there are several nonlinearities in the human vocal tract, whereas linear autoregressive models are fully adequate for unvoiced speech. Turenen et al. [100] agree speech may contain different types of nonlinear characteristics and that for example, vowels may be modelled by either chaotic features or types of higher order nonlinear features, while consonants may be modelled by random processes. Several other studies [10, 63, 66] also suggest chaotic behaviour is found in voiced sounds such as vowels and nasals like /n/ and /m/, and in [99] the speech signal is modelled as a chaotic process. Finally, hybrid methods combining linear and nonlinear structures have been applied to speech processing [35, 61, 108].

In order to investigate the use of hybrid filters in the analysis of speech signals both the combination of NLMS and NNGD from Chapter 3 and NLMS and SSLMS from Chapter 5 were presented with the speech waveforms S1 and S3 from [57]. By presenting these hybrid filters with the speech signals as an input gives an indication of not only which filter is currently performing best but also of the degree of nonlinearity. An insight into the nature of the signal can be drawn even if neither subfilter is an ideal match for the data. Providing at least one of the subfilters converges, this will still indicate which response better represents the input signal. All amplitude signals were standardised so that the amplitude range was between $[-0.5, 0.5]$ and a value of λ close to 1 indicated the predominantly nonlinear or sparse nature of speech and vice versa for $\lambda \rightarrow 0$. The first trace from Fig. 7.8 and Fig. 7.9 show the speech waveform while the second and third traces respectively show the corresponding variations of λ for tracking the degree of nonlinearity and sparsity in the waveform. It can be seen from the figures, that certain parts of a speech signal are better modelled using nonlinear structures ($\lambda \rightarrow 1$), while for others, linear structures are sufficient ($\lambda \rightarrow 0$) and as expected the correlation between nonlinearity and sparsity is confirmed. Furthermore, voiced speech appears to be indicated by regions where λ exhibits a “spiky” behaviour. From Fig. 7.8, it can be noticed that the noise like sounds /z/ (around samples 2800-3200) and /s/ (around samples 4100-4200) are linear which agree with previous findings in the field and from Fig. 7.9, it can be inferred that highly voiced sounds such as /a/ in “trying” are more nonlinear [102].

7.5 Radar

The radar data comes from a maritime radar (IPIX, publicly available from [65]), for different sea states, “low” (calm sea) and “high” (turbulent) states. Whilst there are offline statistical tests for radar data [39] and it has been shown that radar data is predominantly fully complex in nature when the target is in the beam [58], online estimation algorithms are lacking and it is clear that it is important to track the onset of changes in the nature whilst recording.

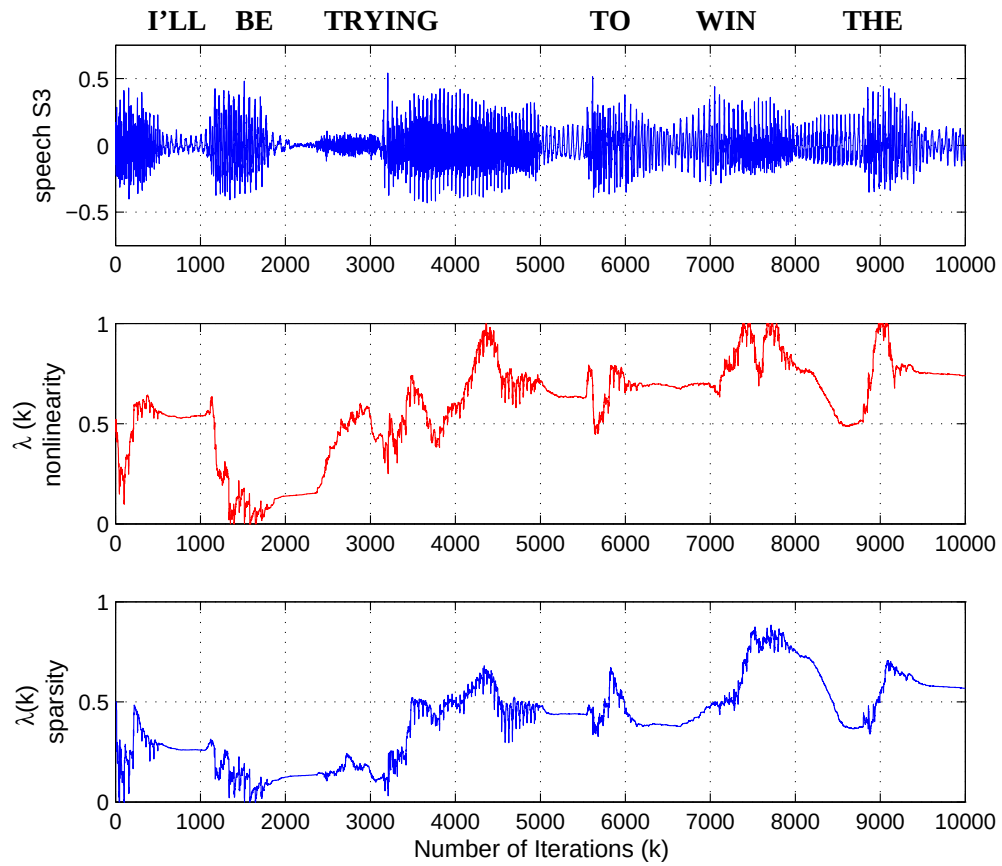


Figure 7.9: Speech signal S3 and corresponding evolution of mixing parameter λ for nonlinearity and sparsity tracking.

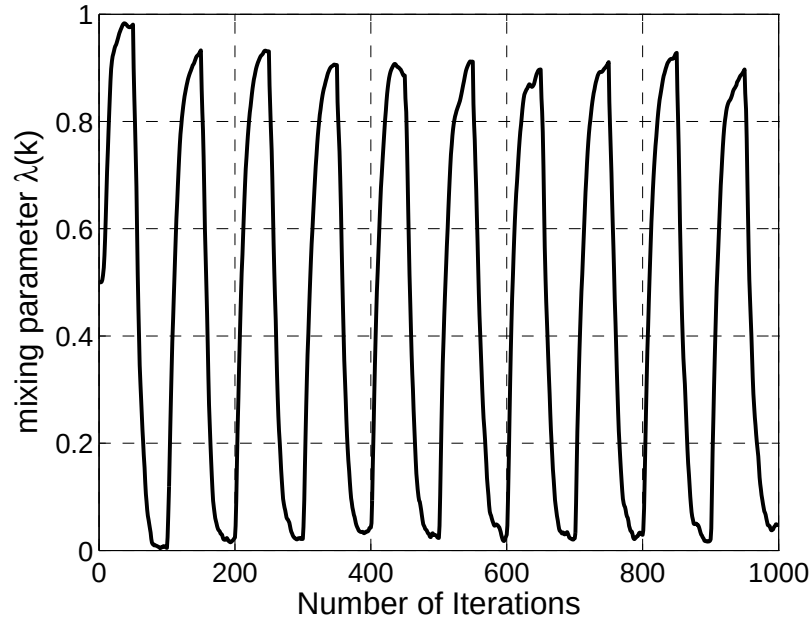


Figure 7.10: Evolution of the mixing parameter λ for alternating blocks of 400 data samples of radar data from the “low” to “high” sea state

For the purpose of tracking the nonlinearity of the radar data the complex hybrid filter for identifying the split-/fully-complex nature of a signal was employed. In this case though the data used was real world data, it was presented as input to the hybrid filter in an artificial way in order to ensure changes in the nature of the signal thus allowing the tracking capabilities of the hybrid filter to be shown. Figure 7.10 shows the evolution of λ when predicting radar data that was alternated every 50 samples from the low sea state to the high sea state. All the data sets were standardised so the magnitudes were in the range $[-1, 1]$. The initial weight vectors were set to zero and the filter order $N = 10$. Figure 7.10 shows that the modality of the high sea state was predominantly fully complex and similarly the low sea state was predominantly split complex.

7.6 Wind

Wind modelling is an illustration of the need for complex valued techniques; wind is normally measured either as a bivariate process of direction and speed [2] or, despite

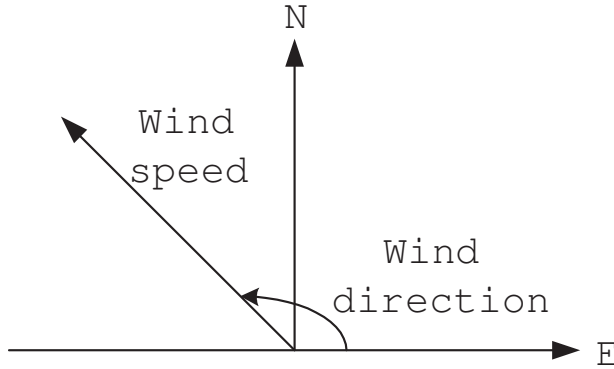


Figure 7.11: Wind recordings as a complex [speed,direction] vector representation

the clear interdependence between the components, only the speed component is taken into account. From Fig. 7.11 it is clear that wind can be represented as a vector of speed and direction components in the North – East coordinate system. The wind vector $\mathbf{v}$ can be made complex through the heterogeneous fusion of the wind speed $\mathbf{v}$ and direction θ to form a complex signal

$$\mathbf{V} = \mathbf{v} \cdot e^{j\theta}, \quad (7.1)$$

where $\mathbf{v}$ is the speed and θ the direction modelled as a single complex value.

The sets wind data used were each measured over a 24 hour period sampled at 50Hz in an urban environment³. The wind speed readings were taken in the north–south ($\mathbf{V}_N$) and east–west ($\mathbf{V}_E$) directions where

$$\mathbf{v} = \sqrt{\mathbf{V}_E^2 + \mathbf{V}_N^2} \quad \text{and} \quad \theta = \arctan\left(\frac{\mathbf{V}_N}{\mathbf{V}_E}\right), \quad (7.2)$$

were used to give the complex signal (7.1). To reduce the effects of high frequency noise a moving average filter with window length 10s was applied to the data before resampling at 1Hz. For the purposes of building a complete picture of the nature of wind all three hybrid filters discussed in Chapter 6 are considered. Firstly, the nonlinearity of the wind is addressed followed by the split-/fully-complex nature of the signal, followed

³The wind data was provided by Prof. Kazuyuki Aihara and Dr Yoshito Hirata from the Institute of Industrial Science, University of Tokyo

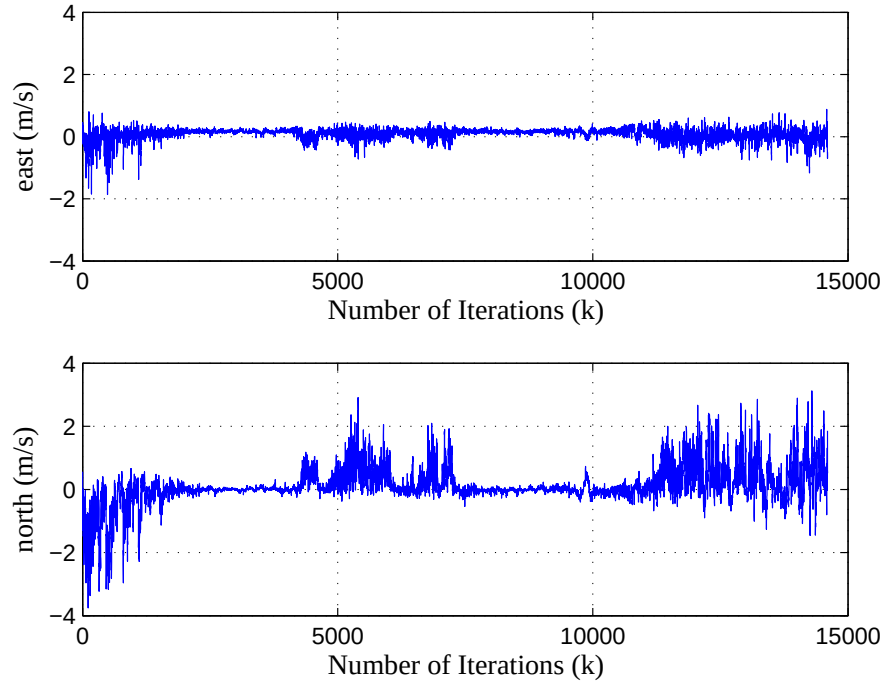


Figure 7.12: Wind speed readings.

finally by the degree of circularity.

7.6.1 Nonlinearity

The wind speed readings for the first set of wind data are shown in Fig. 7.12 and the nonlinearity of the complex signal obtained from these readings was considered. In this case the hybrid filter consisting of the convex combination of the CNNGD and CNLMS was clearly capable of tracking the changes in the nonlinearity of the intermittent and nonstationary wind data as shown in Fig. 7.13. From Fig. 7.12 the wind was clearly changing in the regions between (1-2500), (5000-7000) and (10000-15000) samples, which is also reflected in the values of λ in Fig. 7.13. In comparison for the steady wind, the nature of wind exhibited medium degrees of nonlinearity. Meaning the ideal representation of the data is neither purely linear nor the nonlinear model, as the signal does not display high degrees of nonlinearity. For a more in depth insight into the nature of the signal further tests are required.

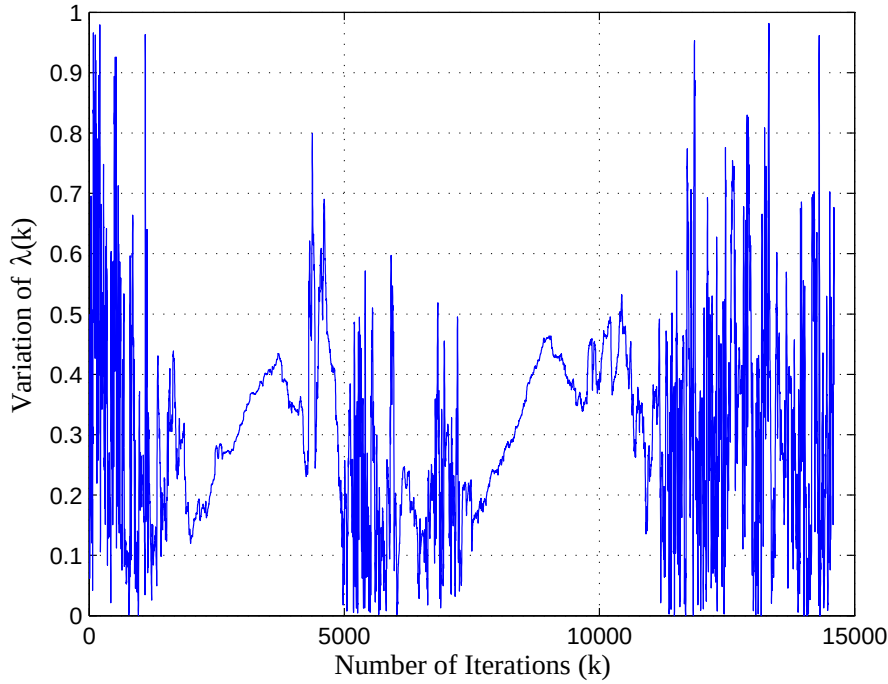


Figure 7.13: Evolution of the mixing parameter λ for a hybrid combination of CNNGD and CNLMS for a set of wind data.

7.6.2 Split-/Fully-Complex

For the second set of wind data, the magnitude of the complex signal obtained from combining the two speed readings for the course of the 24 hours is shown in Fig. 7.14. As it has already been shown that for periods of fluctuating wind the nature of the signal in terms of nonlinearity also fluctuates, for this set of data the split-/fully-complex nature of the signal is investigated. This is achieved using the hybrid filter combining two complex NGD algorithms one with split-complex activation function the other with fully-complex activation function and the corresponding values of λ are shown in Fig. 7.15. For the majority of the time the value of λ is around 0.9 indicating the nature of the wind signal can be considered to mainly be fully-complex. Again for the sections where the measurements fluctuate in magnitude it is also clear that the measurements are more unstable in nature. In these sections λ oscillated in the range $[0.5, 0.9]$, indicating that the signal was mainly fully-complex but at times becomes more split-complex in nature. In fact as these measurements were taken from

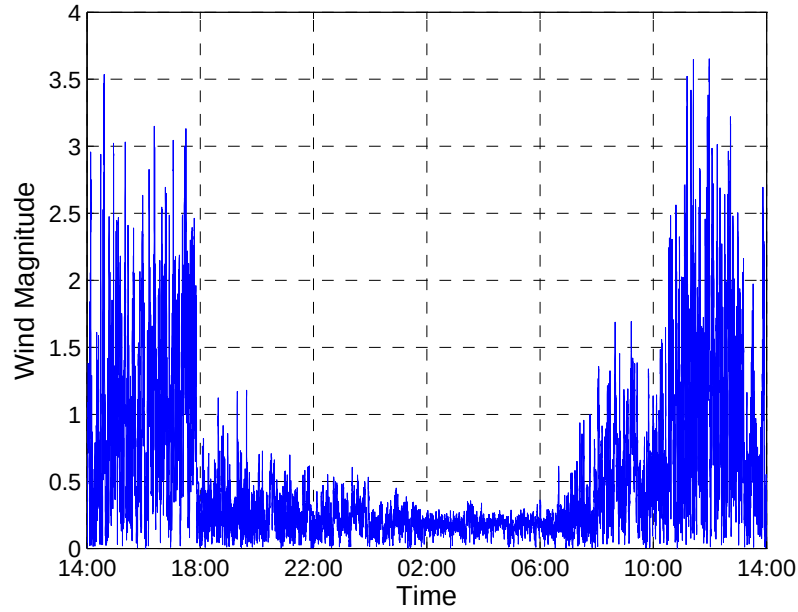


Figure 7.14: Wind magnitude over a 24 hour period in an urban environment

a 24 hour period starting from 14:00, these sections correspond to the recordings taken between 14:00-18:00 and 08:00-14:00 the next day. Fluctuations are to be expected during these times as the wind is changing rapidly compared to the “calm” period in the late evening and the early morning.

7.6.3 Circularity

To best assess the performance of the hybrid filter for detecting the level of noncircularity of the wind data, two periods of approximately 90 mins were assessed from the second set of data, Fig. 7.14, one from the ‘calm’ period between 04:00-05:30 and one from the ‘high’ wind state between 16:00-17:30. The prediction gain $R_p = 10 \log_{10} \frac{\sigma_y^2}{\sigma_e^2}$ is a standard measure of the quantitative performance of learning algorithms [38]. As such, Table 7.1 lists the prediction gains of the subfilters and hybrid filter for the calm and high wind periods. For comparison the corresponding prediction gains for the circular AR(4) signal and noncircular Ikeda map are also shown. For the wind data it can be seen from Table 7.1 that in the calm wind state the CLMS performs better than the ACLMS in the steady state indicating a circular nature. For the high wind state where

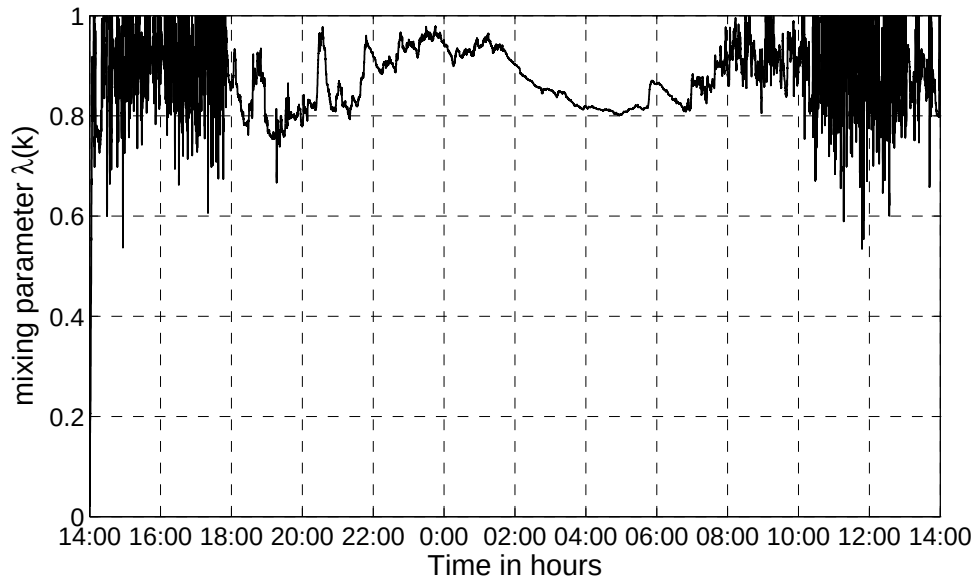


Figure 7.15: Evolution of the mixing parameter λ for the prediction of wind

	AR(4)	Ikeda	'Calm' wind	'High' wind
CLMS	5.25	0.65	7.03	3.26
ACLMS	4.73	3.77	6.87	4.35
Hybrid	5.66	3.73	7.33	4.48

Table 7.1: Prediction gain for the CLMS, ACLMS and hybrid filters for wind data and circular and non-circular synthetic data

there are large variations in the signal dynamics, the ACLMS performs better than the CLMS indicating a noncircular nature. It should also be noted that more significantly in all but one case the hybrid filter outperforms both of the subfilters regardless of which subfilter is responding best to the current input. In the case of the noncircular Ikeda map the hybrid filter only fractionally under performs the ACLMS in terms of steady state prediction gain and as illustrated in Fig. 6.6 has a faster convergence than the ACLMS.

7.7 Conclusions

This chapter has illustrated the use of the hybrid filters discussed in the previous chapters for the characterisation of real world signals. Through the tracking of the mixing parameter in the hybrid filter structure, the results laid out in this chapter have shown that the collaborative adaptive filtering approach can effectively track the changes in the nature of the signals presented. Thus providing a method for online tracking of some fundamental properties of signals.

Chapter 8

Conclusions

8.1 Thesis Summary

An online method for signal modality characterisation through a collaborative adaptive filtering approach has been presented. This is achieved through a hybrid filtering structure, whereby two adaptive filters are combined in a convex manner. By tracking the mixing parameter within such a structure it has been shown that this method can successfully track key signal characteristics in both the real $\mathbb{R}$ and complex $\mathbb{C}$ domains. In the real domain this method has been employed to investigate nonlinearity and sparsity. In the complex domain nonlinearity has again been studied along with the split-complex or fully-complex nature of the nonlinearity, and was then extended to include the circularity (propriety) of the signal. By comparing the responses of the mixing parameters of two different hybrid filters, it was shown that it was possible to build up a more complete understanding of the nature of the signal being processed. Along with the key characteristics of the nature of the signal, the effects of changing between two signals with the same characteristics and the effect of additive noise on the response of the mixing parameter λ were also studied.

When considering algorithms for sparse environments, the derivation of the PNLMS was implemented in a unifying framework starting from the convergence in the mean and mean square of the proportionate LMS algorithm. From this framework adaptive step-size variants have then been developed to improve the performance of the PNLMS

in critical operating conditions.

In the complex domain when identifying and tracking the circularity of a signal a study of widely linear autoregressive modelling was presented. It has been shown that for standard AR models the driving noise is a key factor in the nature of the given signal. For producing nonlinear signals a widely linear extension to AR modelling using augmented statistics was presented. The convergence of the mixing parameter λ within the hybrid filter combining the CLMS and ACLMS has also been given, showing that in the steady state the response of λ was determined by the circular or noncircular nature of the input signal.

These methods were then implemented for various real world signals, showing the possible applications of such an approach. Within this framework, it has been shown that it is possible to track and identify the nonlinearity of different consciousness states in EEG coma patients. It has also been illustrated that the nonlinearity and sparsity of EEG epileptic seizure data and the sparsity of EMG data can be used to assess changes in the underlying nature of the data. For the epileptic seizure data a comparison of the changes of the two different characteristics has also been presented and the nonlinearity and sparsity of speech data has been considered for identifying different sounds in the speech. In the complex domain, the split-/fully-complex nature of radar data has been considered along with the nonlinearity, type of nonlinearity and circularity of wind data.

There are however limitations to this method, when the data is available the use of offline tests will be preferable as by their nature they provide a more accurate measure of the signal characteristics. It is clear from the results in Section 5.2.1 that any noise in the system introduces a certain degree of uncertainty. As noise can be considered to be linear and stochastic, this would in theory lead to the expectation that introducing noise would move all signal natures towards the linear measure, thus reducing the range over which the mixing parameter λ adapts. This does in fact occur for the example given in Chapter 5, though not until SNR is less than 10dB. Another consideration is the test for circularity, which will not identify AR processes driven by doubly white noise with different variances of the real and imaginary parts as being noncircular.

Though in many cases this is not an issue as in terms of the processing such signals can be sufficiently modelled using a standard AR model.

8.2 Future Research

The final consideration is into future directions of research and open problems. Whilst the results provided here have illustrated the validity of this method for signal modality characterisation, there are still issues to be addressed;

- A generic proof of convergence for the mixing parameter λ is required;
- The nature of nonlinearity needs to be more thoroughly examined. The examples given here illustrate nonlinearity in a general sense using saturation type nonlinearities (NARMA) and sparsity as a more specific example. However, as the range of signals spanned by nonlinearity is so wide more precise tests are required;
- Consider multiple modality trackers to give a more complete understanding of the nature of signals. One alternative would be to expand the hybrid filter beyond the convex structure to include multiple subfilters [7] to assess whether this method can successfully identify the predominant signal characteristic.

For the real world examples used more in depth studies need to be performed:

- One topic which may be of interest is to perform simulations using long-term prediction, for example in speech using a prediction delay of one pitch period [14], with the aim of predicting changes in the nature of the signal before the onset;
- For the EMG data rather than identify whether a muscle is tensed or relaxed, identify the degree of tension. This can be used in applications such as speed controls for robotics as in [8];
- The EEG data naturally leads to the consideration of the characteristics of the different electrodes rather than looking at all the data as a whole.

One natural progression is to extend this work from supervised adaptive filtering to unsupervised techniques. The first stage in this process has been to incorporate

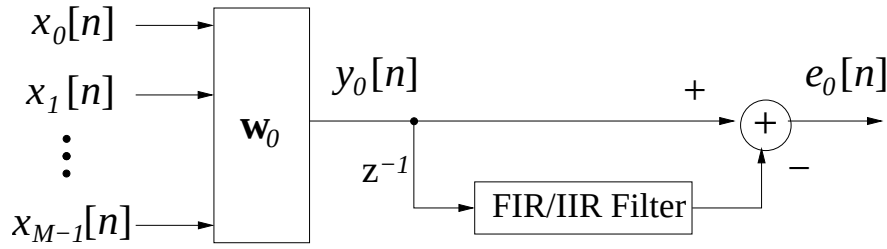


Figure 8.1: BSE with a linear predictor

the supervised filters into a semi-blind structure as shown in Fig. 8.1 for blind source extraction. By using an adaptive filter suited to particular signal characteristics as a predictor in this configuration it's possible to extract signals from a mixture based on the signal nature, one example of which for circularity has been shown in [40], for further details see Appendix D. However, problems with this method occur when the nature of the signal changes during the extraction process, so by combining this method with the hybrid filtering technique to track the nature of the signal throughout with the aim of reducing problems with signal swapping during extraction. This semi-blind structure paves the way to using unsupervised methods and is just one possible extension of the work presented here.

Appendix A

Derivation of GASS Algorithms for PNLMS

A.1 Derivation of Adaptive Step-Size Algorithm

Following the approach from [12], we briefly sketch a rigorous derivation of adaptive step-size PNLMS algorithms. In this approach, the term $\frac{\partial \mathbf{w}(k)}{\partial \eta(k-1)} = \gamma(k)$ is derived based on the PNLMS update (4.31). The step-size update in gradient adaptive step-size algorithms now becomes

$$\eta(k+1) = \eta(k) + \beta e(k) \mathbf{x}(k) \gamma(k) \quad (\text{A.1})$$

and $\gamma(k)$ is obtained from a matrix equation [3, 12]

$$\begin{aligned} \gamma(k) = & \frac{\partial \mathbf{w}(k-1)}{\partial \eta(k-1)} + \frac{\partial \eta(k-1)}{\partial \eta(k-1)} \cdot \frac{\mathbf{G}(k-1)e(k-1)\mathbf{x}(k-1)}{\mathbf{x}^T(k-1)\mathbf{G}(k-1)\mathbf{x}(k-1)} \\ & + \frac{\partial \mathbf{G}(k-1)}{\partial \eta(k-1)} \cdot \frac{\eta(k-1)e(k-1)\mathbf{x}(k-1)}{\mathbf{x}^T(k-1)\mathbf{G}(k-1)\mathbf{x}(k-1)} \\ & + \frac{\partial e(k-1)}{\partial \eta(k-1)} \cdot \frac{\eta(k-1)\mathbf{G}(k-1)\mathbf{x}(k-1)}{\mathbf{x}^T(k-1)\mathbf{G}(k-1)\mathbf{x}(k-1)} \\ & + \frac{\partial [\mathbf{x}^T(k-1)\mathbf{G}(k-1)\mathbf{x}(k-1)]^{-1}}{\partial \eta(k-1)} \cdot \eta(k-1)\mathbf{G}(k-1)e(k-1)\mathbf{x}(k-1). \quad (\text{A.2}) \end{aligned}$$

Assuming that for small η , $\eta(k-1) \approx \eta(k) \therefore \frac{\partial \mathbf{w}(k)}{\partial \eta(k-1)} = \frac{\partial \mathbf{w}(k)}{\partial \eta(k)} = \gamma(k)$ and $\gamma(k-1) = \frac{\partial \mathbf{w}(k-1)}{\partial \eta(k-1)}$ and substituting $\lambda(k)$ with $\frac{\partial \mathbf{G}(k)}{\partial \eta(k)}$, we have

$$\begin{aligned}
\gamma(k) &= \gamma(k-1) + \frac{\mathbf{G}(k-1)e(k-1)\mathbf{x}(k-1)}{\mathbf{x}^T(k-1)\mathbf{G}(k-1)\mathbf{x}(k-1)} + \lambda(k-1) \frac{\eta(k-1)e(k-1)\mathbf{x}(k-1)}{\mathbf{x}^T(k-1)\mathbf{G}(k-1)\mathbf{x}(k-1)} \\
&\quad - \mathbf{x}^T(k-1)\gamma(k-1) \frac{\eta(k-1)\mathbf{G}(k-1)\mathbf{x}(k-1)}{\mathbf{x}^T(k-1)\mathbf{G}(k-1)\mathbf{x}(k-1)} \\
&\quad - \frac{\mathbf{x}^T(k-1)\lambda(k-1)\mathbf{x}(k-1)}{[\mathbf{x}^T(k-1)\mathbf{G}(k-1)\mathbf{x}(k-1)]^2} \eta(k-1)\mathbf{G}(k-1)e(k-1)\mathbf{x}(k-1), \\
\gamma(k) &= \left[1 - \eta(k-1) \frac{\mathbf{x}^T(k-1)\mathbf{G}(k-1)\mathbf{x}(k-1)}{\mathbf{x}^T(k-1)\mathbf{G}(k-1)\mathbf{x}(k-1)} \right] \gamma(k-1) \\
&\quad + \frac{\mathbf{G}(k-1)e(k-1)\mathbf{x}(k-1)}{\mathbf{x}^T(k-1)\mathbf{G}(k-1)\mathbf{x}(k-1)} \\
&\quad + \left[1 - \frac{\mathbf{x}^T(k-1)\mathbf{G}(k-1)\mathbf{x}(k-1)}{\mathbf{x}^T(k-1)\mathbf{G}(k-1)\mathbf{x}(k-1)} \right] \frac{\eta(k-1)e(k-1)\mathbf{x}(k-1)}{\mathbf{x}^T(k-1)\mathbf{G}(k-1)\mathbf{x}(k-1)} \lambda(k-1), \\
\gamma(k) &= [1 - \eta(k-1)] \gamma(k-1) + \frac{\mathbf{G}(k-1)e(k-1)\mathbf{x}(k-1)}{\mathbf{x}^T(k-1)\mathbf{G}(k-1)\mathbf{x}(k-1)}. \tag{A.3}
\end{aligned}$$

In the algorithm [3], the time varying term in the square brackets from the above is replaced by a constant $\alpha < 1$, thus, performing fixed parameter low pass filtering. The algorithm from [64], which we have addressed in section 4.3.1, is therefore a special case of [12] and [3] when the parameter α is set to $\alpha = 0$.

A.2 Derivation of Adaptive Regularisation Algorithm

In this case, $\frac{\partial \mathbf{w}(k)}{\partial \varepsilon(k-1)} = \gamma(k)$ and the update of the regularisation factor now becomes

$$\varepsilon(k+1) = \varepsilon(k) + \beta e(k)\mathbf{x}(k)\gamma(k) \tag{A.4}$$

whereas $\gamma(k)$ is obtained from

$$\begin{aligned}
\gamma_i(k) &= \frac{\partial \mathbf{w}(k-1)}{\partial \varepsilon(k-1)} + \frac{\partial \mathbf{G}(k-1)}{\partial \varepsilon(k-1)} \cdot \mu \frac{e(k-1)\mathbf{x}(k-1)}{\mathbf{x}^T(k-1)\mathbf{G}(k-1)\mathbf{x}(k-1) + \varepsilon(k-1)} \\
&\quad + \frac{\partial e(k-1)}{\partial \varepsilon(k-1)} \cdot \mu \frac{\mathbf{G}(k-1)\mathbf{x}(k-1)}{\mathbf{x}^T(k-1)\mathbf{G}(k-1)\mathbf{x}(k-1) + \varepsilon(k-1)} \\
&\quad + \frac{\partial [\mathbf{x}^T(k-1)\mathbf{G}(k-1)\mathbf{x}(k-1) + \varepsilon(k-1)]^{-1}}{\partial \varepsilon(k-1)} \cdot \mu \mathbf{G}(k-1)e(k-1)\mathbf{x}(k-1), \\
\gamma(k) &= \gamma(k-1) + \lambda(k-1)\mu \frac{e(k-1)\mathbf{x}(k-1)}{\mathbf{x}^T(k-1)\mathbf{G}(k-1)\mathbf{x}(k-1) + \varepsilon(k-1)} \\
&\quad - \mathbf{x}^T(k-1)\gamma(k-1)\mu \frac{\mathbf{G}(k-1)\mathbf{x}(k-1)}{\mathbf{x}^T(k-1)\mathbf{G}(k-1)\mathbf{x}(k-1) + \varepsilon(k-1)} \\
&\quad - \frac{\mathbf{x}^T(k-1)\lambda(k-1)\mathbf{x}(k-1) + 1}{[\mathbf{x}^T(k-1)\mathbf{G}(k-1)\mathbf{x}(k-1) + \varepsilon(k-1)]^2} \mu \mathbf{G}(k-1)e(k-1)\mathbf{x}(k-1), \\
\gamma(k) &= \left[1 - \mu \frac{\mathbf{x}^T(k-1)\mathbf{G}(k-1)\mathbf{x}(k-1)}{\mathbf{x}^T(k-1)\mathbf{G}(k-1)\mathbf{x}(k-1) + \varepsilon(k-1)} \right] \gamma(k-1) \\
&\quad - \mu \frac{\mathbf{G}(k-1)e(k-1)\mathbf{x}(k-1)}{[\mathbf{x}^T(k-1)\mathbf{G}(k-1)\mathbf{x}(k-1) + \varepsilon(k-1)]^2} + \\
&\quad \left[1 - \frac{\mathbf{x}^T(k-1)\mathbf{G}(k-1)\mathbf{x}(k-1)}{\mathbf{x}^T(k-1)\mathbf{G}(k-1)\mathbf{x}(k-1) + \varepsilon(k-1)} \right] \cdot \\
&\quad \cdot \mu \frac{e(k-1)\mathbf{x}(k-1)}{\mathbf{x}^T(k-1)\mathbf{G}(k-1)\mathbf{x}(k-1) + \varepsilon(k-1)} \lambda(k-1). \tag{A.5}
\end{aligned}$$

Allowing $\gamma(k-1) = \frac{\partial \mathbf{w}(k-1)}{\partial \varepsilon(k-1)}$ and $\lambda(k) = \frac{\partial \mathbf{G}(k)}{\partial \varepsilon(k)}$, for small ε gives

$$\gamma(k) = [1 - \mu] \gamma(k-1) - \mu \frac{\mathbf{G}(k-1)e(k-1)\mathbf{x}(k-1)}{[\mathbf{x}^T(k-1)\mathbf{G}(k-1)\mathbf{x}(k-1) + \varepsilon(k-1)]^2}. \tag{A.6}$$

The equation derived in (4.46) is a simplified version of this strict update.

Appendix B

Widely Linear Autoregressive Modelling

B.1 Advantage of Widely Linear Model Over Standard Linear Model

To obtain the advantage of the widely linear AR model over the standard linear AR model we can use the normal equation for the widely linear AR model to find $\mathbf{R}$ from

$$\mathbf{R}^* \begin{bmatrix} \mathbf{h} \\ \mathbf{g} \end{bmatrix} = \mathbf{r}^*$$

$$\begin{bmatrix} \mathcal{C}_{zz}^* & \mathcal{P}_{zz}^* \\ \mathcal{P}_{zz} & \mathcal{C}_{zz} \end{bmatrix} \begin{bmatrix} \mathbf{h} \\ \mathbf{g} \end{bmatrix} = \begin{bmatrix} \mathbf{c}^* \\ \mathbf{p} \end{bmatrix}, \quad (\text{B.1})$$

$$\mathcal{C}_{zz}^* \mathbf{h} + \mathcal{P}_{zz}^* \mathbf{g} = \mathbf{c}^*$$

$$\mathbf{h} = \mathcal{C}_{zz}^{*-1} [\mathbf{c}^* - \mathcal{P}_{zz}^* \mathbf{g}], \quad (\text{B.2})$$

$$\mathcal{P}_{zz} \mathbf{h} + \mathcal{C}_{zz} \mathbf{g} = \mathbf{p}$$

$$\mathbf{g} = \mathcal{C}_{zz}^{-1} [\mathbf{p} - \mathcal{P}_{zz} \mathbf{h}]. \quad (\text{B.3})$$

Solving this system of equations gives

$$\begin{aligned}
\mathcal{C}_{zz}^* \mathbf{h} + \mathcal{P}_{zz}^* [\mathcal{C}_{zz}^{-1} [\mathbf{p} - \mathcal{P}_{zz} \mathbf{h}]] &= \mathbf{c}^* \\
[\mathcal{C}_{zz}^* - \mathcal{P}_{zz}^* \mathcal{C}_{zz}^{-1} \mathcal{P}_{zz}] \mathbf{h} &= \mathbf{c}^* - \mathcal{P}_{zz}^* \mathcal{C}_{zz}^{-1} \mathbf{p} \\
\mathbf{h} &= [\mathcal{C}_{zz}^* - \mathcal{P}_{zz}^* \mathcal{C}_{zz}^{-1} \mathcal{P}_{zz}]^{-1} [\mathbf{c}^* - \mathcal{P}_{zz}^* \mathcal{C}_{zz}^{-1} \mathbf{p}], \quad (\text{B.4})
\end{aligned}$$

$$\begin{aligned}
\mathcal{P}_{zz} [\mathcal{C}_{zz}^{*-1} [\mathbf{c}^* - \mathcal{P}_{zz}^* \mathbf{g}]] + \mathcal{C}_{zz} \mathbf{g} &= \mathbf{p} \\
[\mathcal{C}_{zz} - \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^*] \mathbf{g} &= \mathbf{p} - \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathbf{c}^* \\
\mathbf{g} &= [\mathcal{C}_{zz} - \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^*]^{-1} [\mathbf{p} - \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathbf{c}^*]. \quad (\text{B.5})
\end{aligned}$$

Using these values of $\mathbf{h}$ and $\mathbf{g}$ in the difference between the errors 6.22 gives

$$\begin{aligned}
\delta e^2 &= \mathbf{c}^T [\mathcal{C}_{zz}^* - \mathcal{P}_{zz}^* \mathcal{C}_{zz}^{-1} \mathcal{P}_{zz}]^{-1} [\mathbf{c}^* - \mathcal{P}_{zz}^* \mathcal{C}_{zz}^{-1} \mathbf{p}] \\
&\quad + \mathbf{p}^H [\mathcal{C}_{zz} - \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^*]^{-1} [\mathbf{p} - \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathbf{c}^*] - \mathbf{c}^T \mathcal{C}_{zz}^{*-1} \mathbf{c}^* \\
&= \mathbf{c}^T \left[\mathcal{C}_{zz}^{*-1} + \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* [\mathcal{C}_{zz} - \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^*]^{-1} \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \right] [\mathbf{c}^* - \mathcal{P}_{zz}^* \mathcal{C}_{zz}^{-1} \mathbf{p}] \\
&\quad + \mathbf{p}^H [\mathcal{C}_{zz} - \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^*]^{-1} [\mathbf{p} - \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathbf{c}^*] - \mathbf{c}^T \mathcal{C}_{zz}^{*-1} \mathbf{c}^* \\
&= \mathbf{c}^T \mathcal{C}_{zz}^{*-1} \mathbf{c}^* - \mathbf{c}^T \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* \mathcal{C}_{zz}^{-1} \mathbf{p} + \mathbf{c}^T \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* [\mathcal{C}_{zz} - \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^*]^{-1} \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathbf{c}^* \\
&\quad - \mathbf{c}^T \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* [\mathcal{C}_{zz} - \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^*]^{-1} \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* \mathcal{C}_{zz}^{-1} \mathbf{p} \\
&\quad + \mathbf{p}^H [\mathcal{C}_{zz} - \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^*]^{-1} [\mathbf{p} - \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathbf{c}^*] - \mathbf{c}^T \mathcal{C}_{zz}^{*-1} \mathbf{c}^* \\
&= \mathbf{p}^H [\mathcal{C}_{zz} - \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^*]^{-1} [\mathbf{p} - \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathbf{c}^*] \\
&\quad + \mathbf{c}^T \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* [\mathcal{C}_{zz} - \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^*]^{-1} \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathbf{c}^* \\
&\quad - \mathbf{c}^T \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* [\mathcal{C}_{zz}^{-1} + [\mathcal{C}_{zz} - \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^*]^{-1} \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* \mathcal{C}_{zz}^{-1}] \mathbf{p} \\
&= \mathbf{p}^H [\mathcal{C}_{zz} - \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^*]^{-1} [\mathbf{p} - \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathbf{c}^*] \\
&\quad + \mathbf{c}^T \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* [\mathcal{C}_{zz} - \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^*]^{-1} \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathbf{c}^* \\
&\quad - \mathbf{c}^T \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* [\mathcal{C}_{zz} - \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^*] \mathbf{p} \\
&= [\mathbf{p}^H - \mathbf{c}^T \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^*] [\mathcal{C}_{zz} - \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^*]^{-1} [\mathbf{p} - \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathbf{c}^*], \quad (\text{B.6})
\end{aligned}$$

which can be rewritten as

$$\delta e^2 = \left[\mathbf{p} - \mathcal{P}_{\mathbf{z}\mathbf{z}}^* \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathbf{c}^* \right]^H \left[\mathcal{C}_{\mathbf{z}\mathbf{z}} - \mathcal{P}_{\mathbf{z}\mathbf{z}} \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathcal{P}_{\mathbf{z}\mathbf{z}}^* \right]^{-1} \left[\mathbf{p} - \mathcal{P}_{\mathbf{z}\mathbf{z}} \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathbf{c}^* \right]. \quad (\text{B.7})$$

This means, that as the matrix $\left[\mathcal{C}_{\mathbf{z}\mathbf{z}} - \mathcal{P}_{\mathbf{z}\mathbf{z}} \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathcal{P}_{\mathbf{z}\mathbf{z}}^* \right]$ is positive definite, $\delta e^2 = 0$ only when either $\left[\mathbf{p} - \mathcal{P}_{\mathbf{z}\mathbf{z}}^* \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathbf{c}^* \right] = 0$ or $\left[\mathbf{p} - \mathcal{P}_{\mathbf{z}\mathbf{z}} \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathbf{c}^* \right] = 0$.

B.2 Least Squares Solution for the Widely Linear Model

To find the WLAR coefficients, consider a least square estimator

$$\mathbf{A}\mathbf{x} = \mathbf{b} \quad (\text{B.8})$$

for which the solution is

$$\mathbf{x} = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \mathbf{b}. \quad (\text{B.9})$$

To acquire a solution for the AR coefficients it is possible to solve the normal equations for the AR(1) model

$$\begin{bmatrix} z(2) \\ z(3) \\ \vdots \\ z(N) \end{bmatrix} = \begin{bmatrix} z(1) & z^*(1) \\ z(2) & z^*(2) \\ \vdots & \vdots \\ z(N-1) & z^*(N-1) \end{bmatrix} \begin{bmatrix} h_1 \\ g_1 \end{bmatrix}, \quad (\text{B.10})$$

thus, in our case

$$\mathbf{A} = \begin{bmatrix} z(1) & z^*(1) \\ z(2) & z^*(2) \\ \vdots & \vdots \\ z(N-1) & z^*(N-1) \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} z(2) \\ z(3) \\ \vdots \\ z(N) \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} h_1 \\ g_1 \end{bmatrix}. \quad (\text{B.11})$$

Expand the Least Square estimator in (B.9) to give

$$\begin{aligned}
 \mathbf{A}^H \mathbf{A} &= \begin{bmatrix} z^*(1) & z^*(2) & \dots & z^*(N-1) \\ z(1) & z(2) & \dots & z(N-1) \end{bmatrix} \begin{bmatrix} z(1) & z^*(1) \\ z(2) & z^*(2) \\ \vdots & \vdots \\ z(N-1) & z^*(N-1) \end{bmatrix} \\
 &= \begin{bmatrix} \sum_{i=1}^{N-1} z^*(i)z(i) & \sum_{i=1}^{N-1} z^*(i)z^*(i) \\ \sum_{i=1}^{N-1} z(i)z(i) & \sum_{i=1}^{N-1} z(i)z^*(i) \end{bmatrix} \\
 &= \begin{bmatrix} C_{zz}^*(0) & P_{zz}^*(0) \\ P_{zz}(0) & C_{zz}(0) \end{bmatrix}, \tag{B.12}
 \end{aligned}$$

$$\begin{aligned}
 \mathbf{A}^H \mathbf{b} &= \begin{bmatrix} z^*(1) & z^*(2) & \dots & z^*(N-1) \\ z(1) & z(2) & \dots & z(N-1) \end{bmatrix} \begin{bmatrix} z(2) \\ z(3) \\ \vdots \\ z(N) \end{bmatrix} \\
 &= \begin{bmatrix} \sum_{i=1}^{N-1} z^*(i)z(i+1) \\ \sum_{i=1}^{N-1} z(i)z(i+1) \end{bmatrix} \\
 &= \begin{bmatrix} C_{zz}^*(1) \\ P_{zz}(1) \end{bmatrix}, \tag{B.13}
 \end{aligned}$$

$$\begin{bmatrix} h_1 \\ g_1 \end{bmatrix} = \begin{bmatrix} C_{zz}^*(0) & P_{zz}^*(0) \\ P_{zz}(0) & C_{zz}(0) \end{bmatrix}^{-1} \begin{bmatrix} C_{zz}^*(1) \\ P_{zz}(1) \end{bmatrix}. \tag{B.14}$$

This is the solution for the normal equations for the widely linear AR model. We can now rewrite this to give the solution in the form of the Yule-Walker equations (6.14)

$$\begin{bmatrix} h_1^* \\ g_1^* \end{bmatrix} = \begin{bmatrix} C_{zz}(0) & P_{zz}(0) \\ P_{zz}^*(0) & C_{zz}^*(0) \end{bmatrix}^{-1} \begin{bmatrix} C_{zz}(1) \\ P_{zz}^*(1) \end{bmatrix}. \tag{B.15}$$

Appendix C

Expectation of the Mixing Parameter of Hybrid Filter for Complex Circularity

Using the equations (6.39)-(6.42), and $\mathbf{A} = \left(\mathcal{C}_{\mathbf{z}\mathbf{z}} - \mathcal{P}_{\mathbf{z}\mathbf{z}} \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathcal{P}_{\mathbf{z}\mathbf{z}}^* \right)^{-1}$ and $\mathbf{A}^* = \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} + \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathcal{P}_{\mathbf{z}\mathbf{z}}^* \mathbf{A} \mathcal{P}_{\mathbf{z}\mathbf{z}} \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1}$, the expectations from (6.38) can be expressed as

$$\begin{aligned}
 E[\mathbf{h}_{co}^T \mathbf{z}(k) \mathbf{z}^H \mathbf{h}_o^*] &= [\mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathbf{c}^*]^T \mathcal{C}_{\mathbf{z}\mathbf{z}} \mathbf{A} [\mathbf{c} - \mathcal{P}_{\mathbf{z}\mathbf{z}} \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathbf{p}^*] \\
 &= \mathbf{c}^H \mathcal{C}_{\mathbf{z}\mathbf{z}}^{-1} \mathcal{C}_{\mathbf{z}\mathbf{z}} \mathbf{A} [\mathbf{c} - \mathcal{P}_{\mathbf{z}\mathbf{z}} \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathbf{p}^*] \\
 &= \mathbf{c}^H \mathbf{A} \mathbf{c} - \mathbf{c}^H \mathbf{A} \mathcal{P}_{\mathbf{z}\mathbf{z}} \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathbf{p}^*
 \end{aligned} \tag{C.1}$$

$$\begin{aligned}
 E[\mathbf{h}_{co}^T \mathbf{z}(k) \mathbf{z}^T(k) \mathbf{g}_o^*] &= [\mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathbf{c}^*]^T \mathcal{P}_{\mathbf{z}\mathbf{z}} [\mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathbf{p}^* - \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathcal{P}_{\mathbf{z}\mathbf{z}}^* \mathbf{A} \mathbf{c} + \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathcal{P}_{\mathbf{z}\mathbf{z}}^* \mathbf{A} \mathcal{P}_{\mathbf{z}\mathbf{z}} \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathbf{p}^*] \\
 &= \mathbf{c}^H \mathcal{C}_{\mathbf{z}\mathbf{z}}^{-1} \mathcal{P}_{\mathbf{z}\mathbf{z}} [\mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathbf{p}^* - \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathcal{P}_{\mathbf{z}\mathbf{z}}^* \mathbf{A} \mathbf{c} + \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathcal{P}_{\mathbf{z}\mathbf{z}}^* \mathbf{A} \mathcal{P}_{\mathbf{z}\mathbf{z}} \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathbf{p}^*] \\
 &= \mathbf{c}^H \mathcal{C}_{\mathbf{z}\mathbf{z}}^{-1} \mathcal{P}_{\mathbf{z}\mathbf{z}} \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathbf{p}^* - \mathbf{c}^H \mathcal{C}_{\mathbf{z}\mathbf{z}}^{-1} \mathcal{P}_{\mathbf{z}\mathbf{z}} \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathcal{P}_{\mathbf{z}\mathbf{z}}^* \mathcal{C}_{\mathbf{z}\mathbf{z}}^{-1} \mathbf{c} \\
 &\quad + \mathbf{c}^H \mathcal{C}_{\mathbf{z}\mathbf{z}}^{-1} \mathcal{P}_{\mathbf{z}\mathbf{z}} \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathcal{P}_{\mathbf{z}\mathbf{z}}^* \mathbf{A} \mathcal{P}_{\mathbf{z}\mathbf{z}} \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathbf{p}^* \\
 &= \mathbf{c}^H [\mathcal{C}_{\mathbf{z}\mathbf{z}}^{-1} + \mathcal{C}_{\mathbf{z}\mathbf{z}}^{-1} \mathcal{P}_{\mathbf{z}\mathbf{z}} \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathcal{P}_{\mathbf{z}\mathbf{z}}^* \mathbf{A}] \mathcal{P}_{\mathbf{z}\mathbf{z}} \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathbf{p}^* - \mathbf{c}^H \mathcal{C}_{\mathbf{z}\mathbf{z}}^{-1} \mathcal{P}_{\mathbf{z}\mathbf{z}} \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathcal{P}_{\mathbf{z}\mathbf{z}}^* \mathcal{C}_{\mathbf{z}\mathbf{z}}^{-1} \mathbf{c} \\
 &= \mathbf{c}^H \mathbf{A} \mathcal{P}_{\mathbf{z}\mathbf{z}} \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathbf{p}^* - \mathbf{c}^H \mathcal{C}_{\mathbf{z}\mathbf{z}}^{-1} \mathcal{P}_{\mathbf{z}\mathbf{z}} \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathcal{P}_{\mathbf{z}\mathbf{z}}^* \mathcal{C}_{\mathbf{z}\mathbf{z}}^{-1} \mathbf{c}
 \end{aligned} \tag{C.2}$$

$$\begin{aligned}
-E[\mathbf{h}_{co}^T \mathbf{z}(k) \mathbf{z}^H(k) \mathbf{h}_{co}^*] &= -[\mathcal{C}_{zz}^{*-1} \mathbf{c}^*]^T \mathcal{C}_{zz} [\mathcal{C}_{zz}^{*-1} \mathbf{c}^*]^* \\
&= -\mathbf{c}^H \mathcal{C}_{zz}^{-1} \mathcal{C}_{zz} \mathcal{C}_{zz}^{-1} \mathbf{c} \\
&= -\mathbf{c}^H \mathcal{C}_{zz}^{-1} \mathbf{c} \tag{C.3}
\end{aligned}$$

$$\begin{aligned}
E[\mathbf{h}_{co}^H \mathbf{z}^*(k) \mathbf{z}^T(k) \mathbf{h}_o] &= [\mathcal{C}_{zz}^{*-1} \mathbf{c}^*]^H \mathcal{C}_{zz}^* [\mathcal{C}_{zz}^{*-1} \mathbf{c}^* - \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* \mathbf{A} \mathbf{p} + \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* \mathbf{A} \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathbf{c}^*] \\
&= \mathbf{c}^T \mathcal{C}_{zz}^{*-1} \mathcal{C}_{zz}^* [\mathcal{C}_{zz}^{*-1} \mathbf{c}^* - \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* \mathbf{A} \mathbf{p} + \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* \mathbf{A} \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathbf{c}^*] \\
&= \mathbf{c}^T \mathcal{C}_{zz}^{*-1} \mathbf{c}^* - \mathbf{c}^T \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* \mathbf{A} \mathbf{p} + \mathbf{c}^T \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* \mathbf{A} \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathbf{c}^* \tag{C.4}
\end{aligned}$$

$$\begin{aligned}
E[\mathbf{h}_{co}^H \mathbf{z}^*(k) \mathbf{z}^H(k) \mathbf{g}_o] &= [\mathcal{C}_{zz}^{*-1} \mathbf{c}^*]^H \mathcal{P}_{zz}^* \mathbf{A} [\mathbf{p} - \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathbf{c}^*] \\
&= \mathbf{c}^T \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* \mathbf{A} [\mathbf{p} - \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathbf{c}^*] \\
&= \mathbf{c}^T \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* \mathbf{A} \mathbf{p} - \mathbf{c}^T \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* \mathbf{A} \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathbf{c}^* \tag{C.5}
\end{aligned}$$

$$\begin{aligned}
-E[\mathbf{h}_o^T \mathbf{z}(k) \mathbf{z}^H(k) \mathbf{h}_o^*] &= -[\mathcal{C}_{zz}^{*-1} \mathbf{c}^* - \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* \mathbf{A} \mathbf{p} + \mathcal{C}_{zz}^{*-1} \mathcal{P}_{zz}^* \mathbf{A} \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathbf{c}^*]^T \mathcal{C}_{zz}^* \mathbf{A} [\mathbf{c} \\
&\quad - \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathbf{p}^*] \\
&= -[\mathbf{c}^H \mathcal{C}_{zz}^{-1} - \mathbf{p}^T \mathbf{A}^T \mathcal{P}_{zz} \mathcal{C}_{zz}^{-1} + \mathbf{c}^H \mathcal{C}_{zz}^{-1} \mathcal{P}_{zz}^* \mathbf{A}^T \mathcal{P}_{zz} \mathcal{C}_{zz}^{-1}] \mathcal{C}_{zz}^* \mathbf{A} [\mathbf{c} \\
&\quad - \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathbf{p}^*] \\
&= -\mathbf{c}^H \mathbf{A} \mathbf{c} + \mathbf{c}^H \mathbf{A} \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathbf{p}^* + \mathbf{p}^T \mathbf{A}^T \mathcal{P}_{zz} \mathbf{A} \mathbf{c} \\
&\quad - \mathbf{p}^T \mathbf{A}^T \mathcal{P}_{zz} \mathbf{A} \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathbf{p}^* - \mathbf{c}^H \mathcal{C}_{zz}^{-1} \mathcal{P}_{zz}^* \mathbf{A}^T \mathcal{P}_{zz} \mathbf{A} \mathbf{c} \\
&\quad + \mathbf{c}^H \mathcal{C}_{zz}^{-1} \mathcal{P}_{zz}^* \mathbf{A}^T \mathcal{P}_{zz} \mathbf{A} \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathbf{p}^* \tag{C.6}
\end{aligned}$$

$$\begin{aligned}
-E[\mathbf{g}_o^T \mathbf{z}^*(k) \mathbf{z}^H(k) \mathbf{h}_o^*] &= -\left[\mathbf{A} [\mathbf{p} - \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathbf{c}^*] \right]^T \mathcal{P}_{zz}^* \mathbf{A} [\mathbf{c} - \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathbf{p}^*] \\
&= -[\mathbf{p}^T - \mathbf{c}^H \mathcal{C}_{zz}^{-1} \mathcal{P}_{zz}^*] \mathbf{A}^T \mathcal{P}_{zz}^* \mathbf{A} [\mathbf{c} - \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathbf{p}^*] \\
&= -\mathbf{p}^T \mathbf{A}^T \mathcal{P}_{zz}^* \mathbf{A} \mathbf{c} + \mathbf{p}^T \mathbf{A}^T \mathcal{P}_{zz}^* \mathbf{A} \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathbf{p}^* \\
&\quad + \mathbf{c}^H \mathcal{C}_{zz}^{-1} \mathcal{P}_{zz}^* \mathbf{A}^T \mathcal{P}_{zz}^* \mathbf{A} \mathbf{c} - \mathbf{c}^H \mathcal{C}_{zz}^{-1} \mathcal{P}_{zz}^* \mathbf{A}^T \mathcal{P}_{zz}^* \mathbf{A} \mathcal{P}_{zz} \mathcal{C}_{zz}^{*-1} \mathbf{p}^* \tag{C.7}
\end{aligned}$$

$$\begin{aligned}
-E[\mathbf{g}_o^T \mathbf{z}^*(k) \mathbf{z}^T(k) \mathbf{g}_o^*] &= - \left[\mathbf{A}[\mathbf{p} - \mathcal{P}_{\mathbf{z}\mathbf{z}} \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathbf{c}^*] \right]^T \mathcal{C}_{\mathbf{z}\mathbf{z}}^* [\mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathbf{p}^* - \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathcal{P}_{\mathbf{z}\mathbf{z}}^* \mathbf{A} \mathbf{c} \\
&\quad + \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathcal{P}_{\mathbf{z}\mathbf{z}}^* \mathbf{A} \mathcal{P}_{\mathbf{z}\mathbf{z}} \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathbf{p}^*] \\
&= - [\mathbf{p}^T - \mathbf{c}^H \mathcal{C}_{\mathbf{z}\mathbf{z}}^{-1} \mathcal{P}_{\mathbf{z}\mathbf{z}}^*] \mathbf{A}^T \mathcal{C}_{\mathbf{z}\mathbf{z}}^* [\mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathbf{p}^* - \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathcal{P}_{\mathbf{z}\mathbf{z}}^* \mathbf{A} \mathbf{c} \\
&\quad + \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathcal{P}_{\mathbf{z}\mathbf{z}}^* \mathbf{A} \mathcal{P}_{\mathbf{z}\mathbf{z}} \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathbf{p}^*] \\
&= - \mathbf{p}^T \mathbf{A}^T \mathbf{p}^* + \mathbf{p}^T \mathbf{A}^T \mathcal{P}_{\mathbf{z}\mathbf{z}}^* \mathbf{A} \mathbf{c} - \mathbf{p}^T \mathbf{A}^T \mathcal{P}_{\mathbf{z}\mathbf{z}}^* \mathbf{A} \mathcal{P}_{\mathbf{z}\mathbf{z}} \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathbf{p}^* \\
&\quad + \mathbf{c}^H \mathcal{C}_{\mathbf{z}\mathbf{z}}^{-1} \mathcal{P}_{\mathbf{z}\mathbf{z}}^* \mathbf{A}^T \mathbf{p}^* - \mathbf{c}^H \mathcal{C}_{\mathbf{z}\mathbf{z}}^{-1} \mathcal{P}_{\mathbf{z}\mathbf{z}}^* \mathbf{A}^T \mathcal{P}_{\mathbf{z}\mathbf{z}}^* \mathbf{A} \mathbf{c} \\
&\quad + \mathbf{c}^H \mathcal{C}_{\mathbf{z}\mathbf{z}}^{-1} \mathcal{P}_{\mathbf{z}\mathbf{z}}^* \mathbf{A}^T \mathcal{P}_{\mathbf{z}\mathbf{z}}^* \mathbf{A} \mathcal{P}_{\mathbf{z}\mathbf{z}} \mathcal{C}_{\mathbf{z}\mathbf{z}}^{*-1} \mathbf{p}^* \tag{C.9}
\end{aligned}$$

Substituting these expectations back into (6.38) results in (6.43).

Appendix D

Blind Source Extraction Using a Widely Linear Predictor

To highlight the use of signal characterisation in blind source extraction (BSE) consider the extraction of complex signals using a widely linear predictor. The BSE problem can be expressed as, an observation $\mathbf{x}(k)$ at time instant k , modelled as

$$\mathbf{x}(k) = \mathbf{A}\mathbf{s}(k), \quad (\text{D.1})$$

where $\mathbf{A}$ is a mixing matrix and $\mathbf{s}(k)$ the source vector. The extracted signal $y(k)$ is given by

$$y(k) = \mathbf{w}^H \mathbf{x}(k), \quad (\text{D.2})$$

where the aim is to find a $\mathbf{w}$ such that $\mathbf{w}^H \mathbf{A} = [0, \dots, e^{j\varphi}, \dots, 0]$ extracting a single source, based on a desired criterion. One method of extraction is based on a linear predictor [50] as shown in Fig 8.1. However for complex signals the linear predictor is replaced with a widely linear predictor

$$y_{WL}(k) = \mathbf{h}^T(k) \mathbf{y}(k) + \mathbf{g}^T(k) \mathbf{y}^*(k), \quad (\text{D.3})$$

$$e(k) = y(k) - y_{WL}(k), \quad (\text{D.4})$$

where $\mathbf{y}(k) = [y(k-1), \dots, y(k-M)]^T$ for a predictor of length M . Thus the mean square prediction error (MSPE) can be expressed as

$$\begin{aligned} E[e^2(k)] &= E[e(k)e^*(k)] \\ &= \Re \left\{ \mathbf{w}^H \mathbf{A} \hat{\mathcal{C}}_{\text{ss}} \mathbf{A}^H \mathbf{w} - \mathbf{w}^H \mathbf{A} \hat{\mathcal{P}}_{\text{ss}} \mathbf{A}^T \mathbf{w}^* \right\}, \end{aligned} \quad (\text{D.5})$$

where

$$\hat{\mathcal{C}}_{\text{ss}} = \mathcal{C}_{\text{ss}}(0) - 2 \sum_{m=1}^M h_m^*(k) \mathcal{C}_{\text{ss}}(m) + 2 \sum_{m,l=1}^M [h_m(k) h_l^*(k) + g_m^*(k) g_l(k)] \mathcal{C}_{\text{ss}}(l-m), \quad (\text{D.6})$$

$$\hat{\mathcal{P}}_{\text{ss}} = -2 \sum_{m=1}^M g_m^*(k) \mathcal{P}_{\text{ss}}(m) + 2 \sum_{m,l=1}^M h_m(k) g_l(k) \mathcal{P}_{\text{ss}}(l-m), \quad (\text{D.7})$$

and $\mathcal{C}_{\text{ss}}$ and $\mathcal{P}_{\text{ss}}$ are the covariance and pseudocovariance matrices of the source signals respectively. As the sources are assumed to be uncorrelated the covariance and pseudocovariance are diagonal matrices. It should be noted that the diagonal elements of the matrices $\hat{\mathcal{C}}_{\text{ss}}$ and $\hat{\mathcal{P}}_{\text{ss}}$ represent the mean square prediction errors introduced by the corresponding source signals and for circular signals $\hat{\mathcal{P}}_{\text{ss}} = 0$.

To overcome ambiguity associated with power levels of the source signals using the normalised MSPE which does not change with variation in the source signal amplitude, the cost function can be defined as

$$\mathcal{J}(\mathbf{w}, \mathbf{w}^*) = \frac{E[e^2(k)]}{E[y^2(k)]}. \quad (\text{D.8})$$

The weight coefficients $\mathbf{w}$ are then updated by applying a gradient descent update

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \mu \nabla_{\mathbf{w}^*} \mathcal{J}, \quad (\text{D.9})$$

where μ is the step-size and the direction of the steepest descent is in the direction of the conjugate vector $\mathbf{w}^*$. The gradient $\nabla_{\mathbf{w}^*} \mathcal{J}$ is derived using $\mathbb{C}\mathbb{R}$ calculus, $\mathbb{C}\mathbb{R}$ calculus [47] (Wirtinger calculus [107]) allows the derivatives of complex functions for

cases where the Cauchy-Riemann equations do not hold to be calculated. Typically in signal processing cost functions are real-valued and the conditions imposed by the Cauchy Riemann equations do now allow for their evaluation [59]. Using $\mathbb{CR}$ calculus, the cost function is defined as a function of the vector $\mathbf{z}$ and its conjugate $\mathbf{z}^*$ and the derivative of such functions are taken with respect to $\mathbf{z}$ and $\mathbf{z}^*$, while keeping the other variable constant. As such

$$\begin{aligned} \nabla_{\mathbf{w}^*} \mathcal{J} = & \left[\frac{\partial e^2(k)}{\partial e(k)} \frac{\partial e(k)}{\partial \mathbf{w}^*} + \frac{\partial e^2(k)}{\partial e^*(k)} \frac{\partial e^*(k)}{\partial \mathbf{w}^*} \right] \frac{1}{E[y^2(k)]} \\ & - \left[\frac{\partial y^2(k)}{\partial y(k)} \frac{\partial y(k)}{\partial \mathbf{w}^*} + \frac{\partial y^2(k)}{\partial y^*(k)} \frac{\partial y^*(k)}{\partial \mathbf{w}^*} \right] \frac{E[e^2(k)]}{E[(y^2(k))^2]}, \end{aligned} \quad (\text{D.10})$$

where

$$\frac{\partial e(k)}{\partial \mathbf{w}^*} = \mathbf{x}(k) - \sum_{m=1}^M h_m(k) \mathbf{x}(k-m) \triangleq \hat{\mathbf{x}}_h(k), \quad (\text{D.11})$$

$$\frac{\partial e^*(k)}{\partial \mathbf{w}^*} = - \sum_{m=1}^M g_m^*(k) \mathbf{x}(k-m) \triangleq \hat{\mathbf{x}}_g(k), \quad (\text{D.12})$$

$$\frac{\partial y(k)}{\partial \mathbf{w}^*} = \mathbf{x}(k), \quad (\text{D.13})$$

$$\frac{\partial y^*(k)}{\partial \mathbf{w}^*} = 0. \quad (\text{D.14})$$

Substituting these terms into the gradient D.10 gives

$$\nabla_{\mathbf{w}^*} \mathcal{J} = \left[e^*(k) \hat{\mathbf{x}}_h(k) - e(k) \hat{\mathbf{x}}_g(k) - \frac{\sigma_e^2(k)}{\sigma_y^2(k)} y^*(k) \mathbf{x}(k) \right] \frac{1}{\sigma_y^2(k)}. \quad (\text{D.15})$$

The MSPE, $\sigma_e^2(k)$ and the power of the extracted signal $\sigma_y^2(k)$ can be estimated recursively by

$$\sigma_e^2(k) = \beta_e \sigma_e^2(k-1) + (1 - \beta_e) e^2(k), \quad (\text{D.16})$$

$$\sigma_y^2(k) = \beta_y \sigma_y^2(k-1) + (1 - \beta_y) y^2(k), \quad (\text{D.17})$$

where β_e and β_y are the corresponding forgetting factors. The resulting online update

for $\mathbf{w}$ can now be given by

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \frac{\mu}{\sigma_y^2(k)} \left[e^*(k) \hat{\mathbf{x}}_h(k) - e(k) \hat{\mathbf{x}}_g(k) - \frac{\sigma_e^2(k)}{\sigma_y^2(k)} y^*(k) \mathbf{x}(k) \right]. \quad (\text{D.18})$$

After each update the value of $\mathbf{w}$ is normalised to avoid problems when the value of the norm of $\mathbf{w}$ becomes too small,

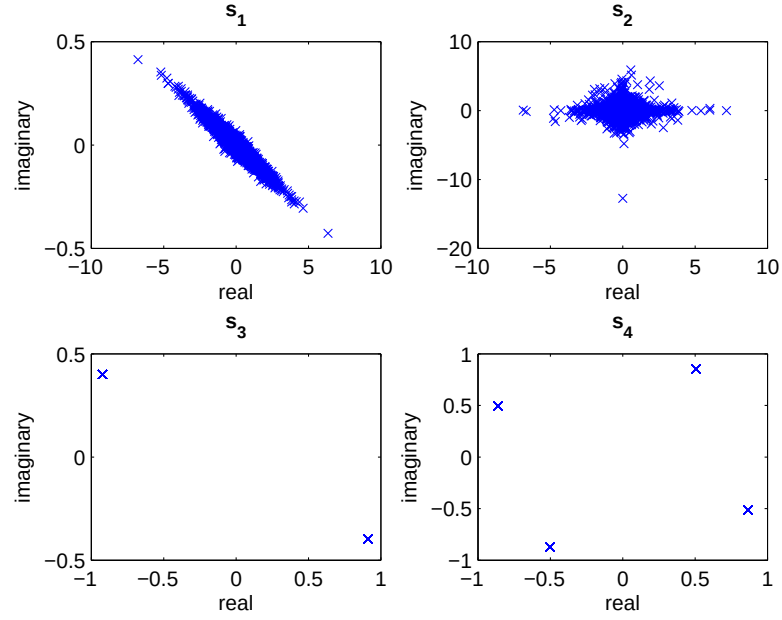
$$\mathbf{w}(k+1) \leftarrow \frac{\mathbf{w}(k+1)}{\|\mathbf{w}(k+1)\|_2^2}. \quad (\text{D.19})$$

It follows that using the update D.18 and adapting the widely linear predictor with the ACLMS algorithm complex signals can be extracted based on their predictability. Using the example from [40], 4 sources of 5000 samples each are combined using a complex mixing matrix $\mathbf{A}$ to give the observations $\mathbf{x}$. Figure D.1(a) shows the complex sources used, s_1 is a noncircular Gaussian signal, s_2 a circular super-Gaussian, s_3 a noncircular sub-Gaussian BPSK signal and s_4 a circular 4-QAM signal, with s_2 having the smallest MSPE. To assess the quality of the extraction process the performance index (PI) was used, PI is defined as

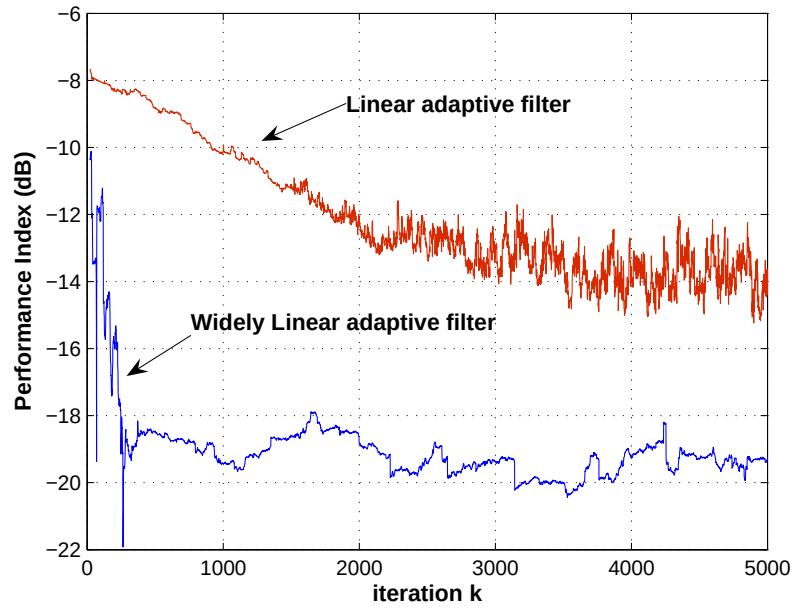
$$PI = 10 \log_{10} \left(\frac{1}{M} \left(\sum_{i=1}^M \frac{c_i^2}{\max\{c_1^2, \dots, c_M^2\}} \right) \right) \quad (\text{D.20})$$

where $\mathbf{c} = \mathbf{w}^T \mathbf{A} = [c_1, \dots, c_M]$. The PI for extraction of s_2 is shown in Fig. D.1(b), for comparison the PI when using a standard linear predictor is also shown. It is clear that for extraction of complex signals the widely linear predictor provides more accurate results than using a standard linear predictor.

This paves the way for extracting signals based on fundamental signal characteristics. Extensions may include the use of different cost functions, or by extending the results presented in [51], where an extraction method using two linear predictors in parallel is presented, to consider such a method with two different filters to allow the extraction of signals based on different characteristics.



(a) The complex source signals.



(b) Comparison of performance index using widely linear and linear adaptive filters.

Figure D.1: Blind source extraction example showing sources and corresponding extraction performance index for both the widely linear predictor and a standard linear adaptive filter.

Bibliography

- [1] K. Aihara, Ed., *Applied Chaos and Applicable Chaos*. Tokyo: Science-Sha, 1994.
- [2] M. C. Alexiadis, P. S. Dokopoulos, H. S. Sahsamanoglou, and I. M. Manousadiris, “Short-term forecasting of wind speed and related electrical power,” *Solar Energy*, vol. 63, no. 1, pp. 61–68, 1998.
- [3] 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.
- [4] J. Arenas-García, A. R. Figueiras-Vidal, and A. H. Sayed, “Steady state performance of convex combinations of adaptive filters,” in *Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 4, 2005, pp. 33–36.
- [5] ———, “Tracking properties of a convex combination of two adaptive filters,” in *Proceedings IEEE Workshop on Statistical Signal Processing*, 2005, pp. 109–114.
- [6] ———, “Mean-square performance of a convex combination of two adaptive filters,” *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 1078–1090, 2006.
- [7] J. Arenas-García, V. Gómez-Verdejo, and A. R. Figueiras-Vidal, “New algorithms for improved adaptive convex combination of LMS transversal filters,” *IEEE Transactions on Instrumentation and Measurement*, vol. 54, no. 6, pp. 2239–2249, 2005.

- [8] S. Aso, A. Sasaki, H. Hashimoto, and C. Ishii, "Driving electric car by using EMG interface," in *Proceedings of IEEE Conference on Cybernetics and Intelligent Systems*, 2006, pp. 1–5.
- [9] L. A. Azpicueta-Ruiz, A. R. Figueiras-Vidal, and J. Arenas-Garcia, "A normalized adaptation scheme for the convex combination of two adaptive filters," in *Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing*, 2008, pp. 3301–3304.
- [10] M. Banbrook, S. McLaughlin, and I. Mann, "Speech characterisation and synthesis by nonlinear methods," in *Proceedings of the International Conference on Speech and Audio Processing*, vol. 7, 1999, pp. 1–17.
- [11] J. Benesty and S. L. Gay, "An improved PNLMS algorithm," in *Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 2, 2002, pp. 1881–1884.
- [12] A. Benveniste, M. Metivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations*. Springer-Verlag, 1990.
- [13] N. J. Bershad, J. C. M. Bermudez, and J.-Y. Tourneret, "An affine combination of two LMS adaptive filterstransient mean-square analysis," *IEEE Transactions on Signal Processing*, vol. 56, no. 5, pp. 1853–1864, 2008.
- [14] M. Birgmeier, H.-P. Bernhard, and G. Kubin, "Nonlinear long-term prediction of speech signals," in *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, 1997, pp. 1283–1286.
- [15] J. Cao, "Analysis of the quasi-brain-death EEG data based on a robust ICA approach," in *Proceedings International Conference on Knowledge-Based & Intelligent Information & Engineering Systems (Lecture Notes in A.I.)*, vol. 4253, 2006, pp. 1240–1247.
- [16] M. Chen, T. Gautama, M. Van Hulle, D. Obradovic, and D. Mandic, "Towards qualitative assessment of machine learning algorithms: Utilising signal modality

- characterisation,” in *Proceedings IEEE Workshop on Machine Learning for Signal Processing*, 2006, pp. 365–370.
- [17] M. Chen, T. Rutkowski, B. Jelfs, G. Souretis, J. Cao, and D. Mandic, “Assessment of nonlinearity in brain electrical activity: A DVV approach,” in *Proceedings RISP International Workshop on Nonlinear Circuits and Signal Processing*, 2007, pp. 461–464.
- [18] M. Chen, D. Sommer, S. L. Goh, T. Gautama, D. Obradovic, M. Golz, M. Morrell, H. Wang, and D. Mandic, “A novel tool for sequential fusion of nonlinear features: A sleep psychology application,” in *Proceedings IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, 2006, pp. 474–478.
- [19] Y.-S. Choi, H.-C. Shin, and W.-J. Song, “Robust regularization for normalized LMS algorithms,” *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 53, no. 8, pp. 627–631, 2006.
- [20] A. Cichocki and R. Unbehauen, *Neural Networks for Optimisation and Signal Processing*. Wiley, 1993.
- [21] J. R. Deller, J. G. Proakis, and H. L. Hansen, *Discrete Time Processing of Speech Signals*. Prentice Hall, 1987.
- [22] H. Deng and M. Doroslovački, “Improving convergence of the PNLMS algorithm for sparse impulse response identification,” *IEEE Signal Processing Letters*, vol. 12, no. 3, pp. 181–184, 2005.
- [23] —, “Proportionate adaptive algorithms for network echo cancellation,” *IEEE Transactions on Signal Processing*, vol. 54, no. 5, pp. 1794–1803, 2006.
- [24] M. Doroslovački and H. Deng, “On convergence of proportionate-type NLMS adaptive algorithms,” in *Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 3, 2006, pp. 105–108.

- [25] S. C. Douglas and W. Pan, "Exact expectation analysis of the LMS adaptive filter," *IEEE Transactions on Signal Processing*, vol. 43, no. 12, pp. 2863–2871, 1995.
- [26] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 5, pp. 508–518, 2000.
- [27] B. Farhang-Boroujeny, *Adaptive Filters: Theory and Applications*. Wiley, 1998.
- [28] A. Feuer and E. Weinstein, "Convergence analysis of LMS filters with uncorrelated Gaussian data," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 1, pp. 222–230, 1985.
- [29] T. Gautama, D. Mandic, and M. Van Hulle, "Signal nonlinearity in fMRI: a comparison between BOLD and MION," *IEEE Transactions on Medical Imaging*, vol. 22, no. 5, pp. 636–644, 2003.
- [30] T. Gautama, D. P. Mandic, and M. M. Van Hulle, "Indications of nonlinear structures in brain electrical activity," *Physical Review E*, vol. 67, no. 4, p. 046204, 2003.
- [31] —, "The delay vector variance method for detecting determinism and nonlinearity in time series," *Physica D*, vol. 190, no. 3–4, pp. 167–176, 2004.
- [32] S. L. Gay, "An efficient, fast converging adaptive filter for network echo cancellation," in *Record of Asilomar Conference on Signals, Systems and Computers*, vol. 1, 1998, pp. 394–398.
- [33] S. L. Goh and D. P. Mandic, "A general complex valued RTRL algorithm for nonlinear adaptive filters," *Neural Computation*, vol. 16, no. 12, pp. 2699–2731, 2004.
- [34] N. F. Güler and S. Koçer, "Classification of EMG signals using PCA and FFT," *Journal of Medical Systems*, vol. 29, no. 3, pp. 241–250, 2005.

- [35] J. Hansen, L. Gavidia-Ceballos, and J. Kaiser, "A nonlinear operator-based speech feature analysis method with applications to vocal fold pathology assessment," *IEEE Transactions on Biomedical Engineering*, vol. 45, no. 3, pp. 300–313, 1998.
- [36] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*. Wiley, 1996.
- [37] S. Haykin, *Adaptive Filter Theory*, 4th ed. Prentice Hall, 2002.
- [38] S. Haykin and L. Li, "Nonlinear adaptive prediction of nonstationary signals," *IEEE Transactions on Signal Processing*, vol. 43, no. 2, pp. 526–535, 1995.
- [39] S. Haykin and J. Principe, "Making sense of a complex world [chaotic events modeling]," *IEEE Signal Processing Magazine*, vol. 15, no. 3, pp. 66–81, 1998.
- [40] S. Javidi, B. Jelfs, and D. P. Mandic, "Blind extraction of noncircular complex signals using a widely linear predictor," in *Proceedings of the IEEE Workshop on Statistical Signal Processing*, 2009, pp. 501–504.
- [41] S. Javidi, M. Pedzisz, S. L. Goh, and D. P. Mandic, "The augmented complex least mean square algorithm with application to adaptive prediction problems," in *Proceedings of the IAPR Workshop on Cognitive Information Processing*, 2008, pp. 54–57.
- [42] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*. Cambridge University Press, 1997.
- [43] D. Kaplan, "Exceptional events as evidence for determinism," *Physica D*, vol. 73, no. 1, pp. 38–48, 1994.
- [44] J. Karvanen and A. Cichocki, "Measuring sparseness of noisy signals," in *Proceedings of the International Symposium on Independent Component Analysis and Blind Signal Separation*, 2003, pp. 125–130.
- [45] T. Kim and T. Adalı, "Approximation by fully complex multilayer perceptrons," *Neural Computation*, vol. 15, no. 7, pp. 1641–1666, 2003.

- [46] S. S. Kozat and A. C. Singer, "Multi-stage adaptive signal processing algorithms," in *Proceedings of the IEEE Sensor Array and Multichannel Signal Processing Workshop*, 2000, pp. 380–384.
- [47] K. Kreutz-Delgado, "The complex gradient operator and the $\mathbb{CR}$ -calculus," Dept. of Electrical and Computer Engineering, UC San Diego, Course Lecture Supplement ECE275A, 2006.
- [48] G. Kubin, "Nonlinear processing of speech," in *Speech coding and synthesis*, W. Kleijn and K. Paliwal, Eds. Elsevier Science B.V., 1995, pp. 557–610.
- [49] L. Li, Y. Xia, B. Jelfs, J. Cao, and D. P. Mandic, "Modelling of brain consciousness based on collaborative adaptive filters," in *International Symposium on Neural Networks*, 2010.
- [50] W. Liu, D. P. Mandic, and A. Cichocki, "Blind source extraction based on a linear predictor," *IET Signal Processing*, vol. 1, no. 1, pp. 29–34, 2007.
- [51] —, "A dual-linear predictor approach to blind source extraction for noisy mixtures," in *IEEE Workshop on Sensor Array and Multichannel Signal Processing*, 2008, pp. 515–519.
- [52] Y.-H. Liu, H.-P. Huang, and C.-H. Weng, "Recognition of electromyographic signals using cascaded kernel learning machine," *IEEE/ASME Transactions on Mechatronics*, vol. 12, no. 3, pp. 253–264, 2007.
- [53] D. Mandic, P. Vayanos, C. Boukis, B. Jelfs, S. L. Goh, T. Gautama, and T. Rutkowski, "Collaborative adaptive learning using hybrid filters," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. III, 2007, pp. 921–924.
- [54] D. Mandic, P. Vayanos, S. Javidi, B. Jelfs, and K. Aihara, "Online tracking of the degree of nonlinearity within complex signals," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 2008, pp. 2061–2064.

- [55] D. P. Mandic, "NNGD algorithm for neural adaptive filters," *Electronics Letters*, vol. 36, no. 9, pp. 845–846, 2000.
- [56] —, "A generalized normalized gradient descent algorithm," *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 115–118, 2004.
- [57] D. P. Mandic, J. Baltersee, and J. A. Chambers, "Nonlinear prediction of speech with a pipelined recurrent neural network and advanced learning algorithms," in *Signal Analysis and Prediction*, A. Prochazka, J. Uhler, P. J. W. Rayner, and N. G. Kingsbury, Eds. Birkhauser, 1998, pp. 291–309.
- [58] D. P. Mandic, M. Chen, T. Gautama, M. M. Van Hulle, and A. Constantinides, "On the characterisation of the deterministic/stochastic and linear/nonlinear nature of time series," *Proceedings of the Royal Society A*, vol. 464, pp. 1141–1160, 2008.
- [59] D. P. Mandic and S. L. Goh, *Complex Valued Nonlinear Adaptive Filters: Non-circularity, Widely Linear and Neural Models*. Wiley, 2009.
- [60] D. P. Mandic, S. Javidi, G. Souretis, and S. L. Goh, "Why a complex valued solution for a real domain problem," in *Proceedings IEEE International Workshop on Machine Learning for Signal Processing*, 2007, pp. 384–389.
- [61] P. Maragos, T. Quatieri, and J. Kaiser, "Speech nonlinearities, modulations, and energy operators," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, vol. 1, 1991, pp. 421–424.
- [62] R. K. Martin, W. A. Sethares, R. C. Williamson, and C. R. Johnson, Jr., "Exploiting sparsity in adaptive filters," *IEEE Transactions on Signal Processing*, vol. 50, no. 8, pp. 1883–1894, 2002.
- [63] F. Martinez, A. Guillamon, J. Alcaraz, and M. Alcaraz, "Detection of chaotic behaviour in speech signals using the largest Lyapunov exponent," in *Proceedings IEEE International Conference on Digital Signal Processing*, vol. 1, 2002, pp. 317–320.

- [64] 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.
- [65] The McMaster IPIX radar sea clutter database. McMaster University. [Online]. Available: <http://soma.crl.mcmaster.ca/ipix/>
- [66] T. Miyano, A. Nagami, I. Tokuda, and K. Aihara, "Detecting nonlinear determinism in voiced sounds of Japanese vowel /a/," *International Journal of Bifurcation and Chaos*, vol. 10, no. 8, pp. 1973–1979, 2000.
- [67] H. Mizuta, M. Jibu, and K. Yana, "Adaptive estimation of the degree of system nonlinearity," in *Proceedings IEEE Adaptive Systems for Signal Processing and Control Symposium*, 2000, pp. 352–356.
- [68] 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.
- [69] 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.
- [70] F. Neeser and J. Massey, "Proper complex random processes with application to information theory," *IEEE Transactions on Information Theory*, vol. 39, no. 4, pp. 1293–1302, 1993.
- [71] M. Nekuui and M. Atarodi, "A fast converging algorithm for network echo cancellation," *IEEE Signal Processing Letters*, vol. 11, no. 4, pp. 427–430, 2004.
- [72] Y. Park, "Control of the robotic wheelchair using electromyogram (EMG) signals," Master's thesis, Imperial College London, 2008.
- [73] B. Picinbono, "On circularity," *IEEE Transactions on Signal Processing*, vol. 42, no. 12, pp. 3473–3482, 1994.

- [74] ———, “Wide-sense linear mean square estimation and prediction,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 3, 1995, pp. 2032–2035.
- [75] B. Picinbono and P. Bondon, “Second-order statistics of complex signals,” *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 411–420, 1997.
- [76] B. Picinbono and M. Bouvet, “Complex white noises and autoregressive signals,” in *Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 9, 1984, pp. 596–599.
- [77] B. Picinbono and P. Chevalier, “Widely linear estimation with complex data,” *IEEE Transactions On Signal Processing*, vol. 43, no. 8, pp. 2030–2033, 1995.
- [78] L. Rabiner and B. H. Juang, *Fundamentals of Speech Recognition*. Prentice Hall, 1993.
- [79] L. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*. Prentice Hall, 1978.
- [80] A. H. Sayed, *Fundamentals of Adaptive Filtering*. Wiley, 2003.
- [81] R. Schober, W. H. Gerstacker, and L. H.-J. Lampe, “A widely linear LMS algorithm for MAI suppression for DS-CDMA,” in *IEEE International Conference on Communications*, vol. 4, 2003, pp. 2520–2525.
- [82] T. Schreiber, “Interdisciplinary application of nonlinear time series methods,” *Physics Reports*, vol. 308, no. 1, pp. 1–64, 1999.
- [83] T. Schreiber and A. Schmitz, “Improved surrogate data for nonlinearity tests,” *Physical Review Letters*, vol. 77, no. 4, pp. 635–638, 1996.
- [84] ———, “On the discrimination power of measures for nonlinearity in a time series,” *Physical Review E*, vol. 55, no. 5, pp. 5443–5447, 1997.
- [85] ———, “Surrogate time series,” *Physica D*, vol. 142, pp. 346–382, 2000.

- [86] P. Schreier and L. Scharf, "Second-order analysis of improper complex random vectors and processes," *IEEE Transactions on Signal Processing*, vol. 51, no. 3, pp. 714–725, 2003.
- [87] P. J. Schreier, "The degree of impropriety (noncircularity) of complex random vectors," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 2008, pp. 3909–3912.
- [88] 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, p. 433, 2006.
- [89] P. J. Schreier, L. L. Scharf, and C. T. Mullis, "Detection and estimation of improper complex random signals," *IEEE Transactions on Information Theory*, vol. 51, no. 1, pp. 306–312, 2005.
- [90] J. Schroeter and M. Sondhi, "Speech coding based on physiological models of speech production," in *Advances in Speech Signal Processing*, S. Furui and M. Sondhi, Eds. Marcel Dekker, 1992, pp. 231–268.
- [91] W. C. Siaw, S. L. Goh, A. I. Hanna, C. Boukis, and D. P. Mandic, "Fully adaptive neural nonlinear FIR filters," in *Proceedings IEEE Workshop on Neural Networks for Signal Processing*, 2002, pp. 279–288.
- [92] M. T. M. Silva and V. H. Nascimento, "Improving the tracking capability of adaptive filters via convex combination," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3137–3149, 2008.
- [93] E. Soria, J. Calpe, J. Chambers, M. Martinez, G. Camps, and J. D. M. Guerrero, "A novel approach to introducing adaptive filters based on the LMS algorithms and its variants," *IEEE Transactions on Education*, vol. 47, no. 1, pp. 127–133, 2004.
- [94] E. Soria-Olivas, J. Calpe-Maravilla, J. F. Guerrero-Martinez, M. Martinez-Sober, and J. Espi-Lopez, "An easy demonstration of the optimum value of the adapta-

- tion constant in the LMS algorithm,” *IEEE Transactions on Education*, vol. 41, no. 1, p. 81, 1998.
- [95] C. J. Stam, T. Woerkom, and W. S. Pritchard, “Use of nonlinear EEG measures to characterize EEG changes during mental activity,” *Electroencephalography and Clinical Neurophysiology*, vol. 99, pp. 214–224, 1996.
- [96] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. Farmer, “Testing for nonlinearity in time series: The method of surrogate data,” *Physica D*, vol. 58, pp. 77–94, 1992.
- [97] F. J. Theis and G. A. Garcia, “On the use of sparse signal decomposition in the analysis of multi-channel surface electromyograms,” *Signal Processing*, vol. 86, no. 3, pp. 603–623, 2006.
- [98] J. Thyssen, H. Nielsen, and S. Hansen, “Non-linear short-term prediction in speech coding,” in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, vol. 1, 1994, pp. 185–188.
- [99] B. Townshend, “Nonlinear prediction of speech,” in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, vol. 1, 1991, pp. 425–428.
- [100] J. Turunen, J. T. Tiantu, and P. Loula, “Hammerstein model for speech coding,” *EURASIP Journal on Applied Signal Processing*, vol. 2003, no. 12, pp. 1238–1249, 2003.
- [101] A. van den Bos, “The multivariate complex normal distribution,” *IEEE Transactions on Information Theory*, vol. 41, no. 41, pp. 537–539, 1995.
- [102] P. Vayanos, M. Chen, B. Jelfs, and D. P. Mandic, “Exploiting nonlinearity in signal processing: Qualitative assessment of adaptive filtering algorithms and signal modality characterisation,” in *Advances in Nonlinear Speech Processing (Lecture Notes in Computer Science)*. Springer, 2007, vol. 4885, pp. 57–77.

- [103] C. Wang and J. C. Principe, "Training neural networks with additive noise in the desired signal," *IEEE Transactions on Neural Networks*, vol. 10, no. 6, pp. 1511–1517, 1999.
- [104] A. S. Weigend and N. A. Gershenfeld, Eds., *Time Series Prediction: Forecasting the Future and Understanding the Past*, ser. Santa Fe Institute Studies in the Sciences of Complexity. Addison–Wesley Publishing Company, 1994.
- [105] B. Widrow, J. McCool, and M. Ball, "The complex LMS algorithm," *Proceedings of the IEEE*, vol. 63, no. 4, pp. 719–720, 1975.
- [106] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Prentice–Hall, 1985.
- [107] W. Wirtinger, "Zur formalen theorie der funktionen von mehr komplexen veränderlichen," *Mathematische Annalen*, vol. 97, no. 1, pp. 357–375, 1927.
- [108] W. Wokurek, "Time-frequency analysis of the glottal opening," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, vol. 2, 1997, pp. 1435–1438.
- [109] Y. Yuan, Y. Li, and D. P. Mandic, "Comparison analysis of embedding dimension between normal and epileptic EEG time series," *Journal of Physiological Sciences*, vol. 58, no. 4, pp. 239–247, 2008.
- [110] Q. T. Zhang, S. Haykin, and K. M. Wong, "Target detection in a nonstationary noise field: A comparison-based approach," in *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, 1991, pp. 3549–3552.
- [111] Y. Zhang and J. A. Chambers, "Convex combination of adaptive filters for a variable tap-length LMS algorithm," *IEEE Signal Processing Letters*, vol. 13, no. 10, p. 628, 2006.