

Statistical Signal Processing of Noncircular Quaternion Random Variables

Min Xiang

Supervisor: Prof. Danilo P. Mandic

A thesis submitted for the degree of
Doctor of Philosophy (Ph.D.)

Communications and Signal Processing Group
Department of Electrical and Electronic Engineering
Imperial College London
December 2017

Abstract

The focus of this thesis is on the development of theory and methods of statistical signal processing for noncircular quaternion random variables. Although it is straightforward to model three- and four-dimensional signals as real-valued trivariate and quadrivariate vectors, the division algebra of quaternions provides several crucial advantages. These include nonexistence of the gimbal lock associated with vector algebras and generic extensions of real and complex learning algorithms. However, traditional quaternion techniques are based on the assumption of circular (rotation-invariantly distributed) quaternion variables, a very limiting assumption for real-world applications. Only recently, advances in augmented quaternion statistics have revealed the suboptimality of such an assumption and have introduced the widely linear processing methodology, which is suitable for second-order noncircular signals.

First, a simultaneous diagonalisation of the covariance matrix and the three complementary covariance matrices of noncircular quaternion variables is developed. This forms a basis for the proposed quaternion approximate uncorrelating transform (QAUT), which can be used to decorrelate noncircular quaternion variables, and offers both computational and practical advantages in many applications.

Using the QAUT, the performance of existing minimum mean square error estimation techniques for noncircular quaternion signals is thoroughly explored. The performance advantage of widely linear estimation over strictly linear estimation is theoretically quantified, and the convergence performance of a class of quaternion adaptive filters is established in the mean and mean square sense.

Several novel adaptive estimation techniques for noncircular quaternion signals are proposed. A novel quaternion least mean magnitude-phase adaptive filter is shown to outperform the quaternion least mean square filter in several practical scenarios. Computationally efficient algorithms are also introduced for the quaternion estimation framework, through a special decomposition of the mean square

error. Finally, a class of quaternion multiple model adaptive estimation algorithms are proposed to effectively deal with model uncertainty in quaternion-valued state estimation.

Acknowledgements

First and foremost, I would like to express my gratitude to my supervisor Prof. Danilo Mandic. During the past four years, through his encouragement and guidance, he has given me opportunities to grow as a researcher in the field of signal processing. I am honoured to be part of his wonderful team and impressed by his professionalism, enthusiasm and determination. I am sure this will greatly benefit the rest of my life although there is still a lot left to learn from him.

Next, I am thankful to have been given the privilege to work with Dr. Yili Xia and Dr. Clive Cheong Took for their enlightening advice and comments which benefited this thesis and its related publications. I sincerely thank my former colleagues, Dr. Sithan Kanna and Dr. Sayed Pouria Talebi, for their guidance and assistance in all respects throughout my PhD. I am grateful to my colleagues and collaborators, Alex Stott and Bruno Scalzo Dees, from whom I have learnt a lot. It is important to thank my colleague, Theerasak Chanwimalueang, with whom I had multiple interesting experiences, including in writing this thesis. I also had the pleasure working with Ilia Kisil, Thayne Thanthawaritthisai, Giuseppe Giovanni Calvi, Apit Hemakom, Wilhelm von Rosenberg, Takashi Nakamura, Tricia Adjei and other colleagues in the CSP group. Furthermore, I thank the CSP group administrator Ms. Charlotte Grady, and graduate tutor, Dr. Imad Jaimoukha, for all their help during my PhD.

Last but not least, my heartfelt gratitude goes to my family for their love, understanding and support, not only during the past four years, but also throughout my education. I would not be at this level without them.

“Time is said to have only one dimension, and space to have three dimensions. ... The mathematical quaternion partakes of both these elements; in technical language it may be said to be “time plus space”, or “space plus time”: and in this sense it has, or at least involves a reference to, four dimensions. ... I have never been able to give a clearer statement of their nature and their aim than I have done in two lines of a sonnet addressed to Sir John Herschel:

*And how the One of Time, of Space the Three,
Might in the Chain of Symbols girdled be.”*

Sir William Rowan Hamilton

Contents

Abstract	2
Acknowledgements	4
Contents	6
List of Figures	11
List of Tables	14
Statement of Originality	15
Copyright Declaration	16
List of Principal Symbols	17
List of Abbreviations	19
Publications	22
1 Introduction	24
1.1 Statistical Signal Processing in the Real Domain	24
1.2 Statistical Signal Processing in the Complex Domain	25
1.3 Statistical Signal Processing in the Quaternion Domain	27
1.4 Contributions of the Thesis	29
1.5 Organisation of the Thesis	33

2	Statistical Signal Processing in the Quaternion Domain	38
2.1	Quaternion Algebra	38
2.1.1	Quaternion Matrices	41
2.2	Generalised $\mathbb{H}\mathbb{R}$ Calculus	42
2.3	Augmented Second-order Quaternion Statistics	45
2.3.1	Quaternion Properness	45
2.3.2	Quaternion Multivariate Gaussian Distribution	47
2.4	Widely Linear Quaternion Estimation	48
2.4.1	Quaternion Least Mean Square Filters	49
2.4.2	Quaternion Kalman Filters	50
I	Simultaneous Diagonalisation of the Covariance and Complementary Covariance Matrices	53
3	Quaternion Approximate Uncorrelating Transform	54
3.1	Introduction	54
3.2	Simultaneous Diagonalisation of the Covariance Matrix and One Complementary Covariance Matrix	56
3.3	Simultaneous Approximate Diagonalisation of the Covariance Matrix and Three Complementary Covariance Matrices	58
3.3.1	Principle of QAUT	60
3.4	Simulations	63
3.4.1	Performance of QAUT for Synthetic Improper Signals	63
3.4.2	Performance of QAUT for Real-World Signals	64
3.4.3	QAUT for Rank Reduction	67
3.4.4	QAUT for Imbalance Detection in Three-Phase Power Systems	70
3.5	Summary	72

II Performance Analysis of Minimum Mean Square Error Estimation Techniques 73

4 Performance Advantage of Widely Linear Quaternion Estimation 74

4.1	Introduction	74
4.2	Performance Advantage of Widely Linear Estimation	75
4.2.1	MMSE in Strictly Linear Estimation	76
4.2.2	MMSE in Widely Linear Estimation	76
4.2.3	MMSE Comparison Between Strictly and Widely Linear Estimation	77
4.3	Performance Analysis Using the QAUT	78
4.3.1	Impact of Improperity of the Input Signal	79
4.3.2	MMSE Analysis for Complex MMSE Estimation	81
4.4	Simulations	82
4.5	Summary	83

5 Convergence Performance of Quaternion Adaptive Filters in Non-stationary Environments 84

5.1	Introduction	85
5.2	A Class of Quaternion Adaptive Filters in Non-stationary Environments	87
5.3	Convergence Performance Analysis of Quaternion Adaptive Filters	89
5.3.1	Mean Analysis	89
5.3.2	Mean Square Analysis	90
5.3.3	Performance Analysis of QLMS	95
5.4	Mean Square Analysis of QLMS with Gaussian Regressors Using QAUT	99
5.4.1	SL-QLMS	103
5.4.2	SWL-QLMS	105

5.4.3	WL-QLMS	106
5.5	Simulations	108
5.5.1	SL-QLMS	108
5.5.2	SWL-QLMS	110
5.5.3	WL-QLMS	111
5.6	Summary	116

III Novel Adaptive Estimation Algorithms 117

6 The Quaternion Least Mean Magnitude-Phase Adaptive Filtering Algorithm 118

6.1	Introduction	118
6.2	The QLMMP Algorithm	120
6.3	Convergence Analysis	122
6.3.1	Convergence Analysis of Normalised QLMS with a Quaternion Step Size	123
6.3.2	Convergence Analysis of Normalised QLMMP	125
6.4	Special Cases of QLMMP	126
6.5	Simulations	127
6.6	Summary	131

7 Low-Complexity Quaternion Adaptive Filters 132

7.1	Introduction	132
7.2	Four-channel Quaternion Estimation Model	134
7.3	Four-channel Quaternion MMSE Estimation	135
7.3.1	FC-QLMS Algorithm	137
7.3.2	FC-QRLS Algorithm	139
7.3.3	FC-QNGD Algorithm	140

7.4	Performance of FC-QLMS	142
7.4.1	Equivalence of the FC-QLMS and WL-QLMS	142
7.4.2	Transient Performance	143
7.4.3	Computational Complexity	150
7.4.4	Adaptive Step Size	152
7.5	Simulations	152
7.5.1	QLMS	153
7.5.2	QRLS	154
7.5.3	QNGD	155
7.6	Summary	158
8	Widely Linear Quaternion Multiple Model Adaptive Estimation	159
8.1	Introduction	159
8.2	QMMAE Algorithms	161
8.2.1	WL-QIMM Algorithm	163
8.2.2	Performance Analysis of WL-QIMM	165
8.2.3	WL-QSMM Algorithm	170
8.2.4	Semi-Widely Linear and Strictly Linear QMMAE	172
8.3	Simulations	173
8.3.1	QIMM for Filtering Quaternion Signals	173
8.3.2	QSMM for Filtering Quaternion Signals	178
8.3.3	QIMM for 3D Target Tracking	181
8.4	Summary	184
9	Conclusions and Future Work	185
9.1	Conclusions	185
9.2	Future Work	187
	Bibliography	189

List of Figures

1.1	Thesis structure.	37
3.1	Squared diagonal error [in %] against the correlation degree and the data segment length for synthetic data.	65
3.2	Squared diagonal error [in %] against the correlation degree and the data segment length for EEG data.	66
3.3	Rank reduction using the QAUT.	68
3.4	Scatter plots showing high correlation between the original x_1 and x_2 variables.	68
3.5	Scatter plots showing no correlation between the original y_1 and y_2 variables.	68
3.6	The logarithm of the ratio of the variance explained by each variable to the total variance. <i>Insert diagram:</i> The cumulative ratio of variance against the number of variables for the original and decorrelated corn data.	69
3.7	Three-phase power signal and imbalance detection.	72
4.1	Effects of (a) widely linear nature of the system, (b) impropriety of the input signal, and (c) variance of the input signal on the MMSE advantage of the widely linear QLMS estimator.	83
5.1	Theoretical and simulated learning curves of SL-QLMS for an improper input with $r = 0.5$ and $\mu = 0.02$	109
5.2	Theoretical and simulated EMSEs of SL-QLMS as a function of the step size, for an $\mathbb{H}$ -proper input.	110
5.3	Theoretical and simulated EMSEs of SL-QLMS as a function of the step size, for an improper input with $r = 0.5$	110
5.4	Theoretical and simulated EMSEs of SL-QLMS as a function of the impropriety coefficient, for $\mu = 0.2$	111
5.5	Theoretical and simulated learning curves of SWL-QLMS for an improper input with $r = 0.5$ and $\mu = 0.02$	112

5.6	Theoretical and simulated EMSEs of SWL-QLMS as a function of the step size, for an $\mathbb{H}$ -proper input.	113
5.7	Theoretical and simulated EMSEs of SWL-QLMS as a function of the step size, for an improper input with $r = 0.5$	113
5.8	Theoretical and simulated EMSEs of SWL-QLMS as a function of the impropriety coefficient, for $\mu = 0.1$	113
5.9	Theoretical and simulated learning curves of WL-QLMS for an improper input with $r = 0.5$ and $\mu = 0.02$	114
5.10	Theoretical and simulated EMSEs of WL-QLMS as a function of the step size, for an $\mathbb{H}$ -proper input.	114
5.11	Theoretical and simulated EMSEs of WL-QLMS as a function of the step size, for an improper input with $r = 0.5$	115
5.12	Theoretical and simulated EMSEs of WL-QLMS as a function of the impropriety coefficient, for $\mu = 0.05$	115
6.1	Geometric interpretations of the QLMMP, QLMS, and QLMP. . . .	122
6.2	MSD curves for identification of the three-phase power system using the normalised QLMMP and normalised QLMS.	129
6.3	Phase error power curves for filtering three-phase voltages using the normalised QLMMP and normalised QLMP.	129
6.4	MSD curves for one-step ahead prediction of three-phase voltages using the normalised MQLMMP.	130
6.5	MSE curves for one-step ahead prediction of three-phase voltages using the normalised MQLMMP.	130
7.1	Architecture of the FC-QLMS algorithm.	137
7.2	Averaged weight trajectories for the estimation of a strictly linear MA(1) process driven by $\mathbb{H}$ -proper white Gaussian noise.	146
7.3	Averaged weight trajectories for the estimation of a strictly linear MA(1) process driven by $\mathbb{C}^i$ -improper white Gaussian noise.	147
7.4	Averaged weight trajectories for the estimation of a strictly linear MA(1) process driven by general improper white Gaussian noise. . .	149
7.5	The number of real-valued operations per iteration for the QLMS algorithms, for an adaptive filter of length L	151

7.6	NMSE curves of the QLMS algorithms for the estimation of MA systems.	154
7.7	NMSE curves of the QLMS algorithms for the wind forecasting. . .	155
7.8	NMSE curves of the QRLS algorithms for the estimation of MA systems.	156
7.9	NMSE curves of the QRLS algorithms for the wind forecasting. . .	156
7.10	Performance of the QNGD algorithms for the one-step ahead prediction of Saito's chaotic circuit data.	157
8.1	Learning curves of the three QIMM algorithms when the state and observation noise are improper.	175
8.2	Learning curves of the three QIMM algorithms when the state and observation noise are $\mathbb{C}^2$ -improper.	176
8.3	Learning curves of the three QIMM algorithms when the state and observation noise are $\mathbb{H}$ -proper.	177
8.4	The Monte Carlo (MC) simulated MSEs of the three QIMM algorithms and the MSE computed from the proposed performance analysis algorithm (Analysis) for WL-QIMM, as a function of the impropriety coefficient of noise.	178
8.5	Learning curves of the three QSMM algorithms when the state and observation noise are improper.	179
8.6	The Monte Carlo (MC) simulated MSEs of the three QSMM algorithms and the MSE computed from the proposed performance analysis algorithm (Analysis) for WL-QSMM, as a function of the impropriety coefficient of noise.	180
8.7	Learning curves of the three QIMM algorithms for target tracking in a 3D space.	183

List of Tables

5.1	Analysis results of quaternion adaptive filters.	103
7.1	The number of real-valued operations per iteration for the QLMS algorithms, for an adaptive filter of length L	150

Statement of Originality

I declare that the work presented in this thesis is the outcome of my research under the supervision of Prof. Danilo P. Mandic. Ideas and results from the work of others are appropriately acknowledged with the standard referencing practices employed in this discipline. The material in this thesis has not been submitted for any degree at any other academic or professional institution.

Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

List of Principal Symbols

$\mathbb{R}$	Field of real numbers
$\mathbb{R}^+$	Field of positive real numbers
$\mathbb{C}$	Field of complex numbers
$\mathbb{H}$	Field of quaternion numbers
x	Scalars are denoted by italic letters
$\mathbf{x}$	Column vectors are denoted by bold lowercase letters
$\mathbf{X}$	Matrices are denoted by bold uppercase letters
$\mathbf{I}_M$	$M \times M$ identity matrix
$\mathbf{1}_M$	M -dimensional all-ones column vector
$[\cdot]_{a,b}$	The element on the a -th row and b -th column of a matrix
$\text{diag}(\cdot)$	Create a column vector containing the diagonal entries of a matrix
$\text{Diag}(\cdot)$	Create a diagonal matrix with the augments on the diagonal
$\text{bdiag}(\cdot)$	Create a block-diagonal matrix of its arguments
$\text{Tr}(\cdot)$	Trace operator
$\text{vec}(\cdot)$	Constructs a column vector from a matrix by staking its columns
$(\cdot)^T$	Transpose operator for vectors and matrices
$(\cdot)^H$	Conjugate transpose operator for vectors and matrices
$(\cdot)^{-1}$	Inverse operator
$(\cdot)^*$	Conjugate operator

$ \cdot $	Modulus of a scalar or determinant of a matrix
$\ \cdot\ $	Euclidean norm of a vector
$\otimes$	Left Kronecker product
$\otimes_{\mathbf{R}}$	Right Kronecker product
$\imath, j, \kappa$	Imaginary units of quaternion field
$(\cdot)^{\imath}$	$\imath$ -involution operator
$(\cdot)^j$	j -involution operator
$(\cdot)^{\kappa}$	κ -involution operator
$\Re\{\cdot\}$	Real component of a quaternion/complex number
$\Im\{\cdot\}$	Imaginary component of a quaternion/complex number
$\Im_{\imath}\{\cdot\}$	$\imath$ -imaginary component of a quaternion number
$\Im_j\{\cdot\}$	j -imaginary component of a quaternion number
$\Im_{\kappa}\{\cdot\}$	κ -imaginary component of a quaternion number
e	Euler's number
$\exp\{\cdot\}$	Natural exponential function
$\ln\{\cdot\}$	Natural logarithm function
$E\{\cdot\}$	Statistical expectation operator
$E\{y x\}$	Conditional expectation of y given x
$\triangleq$	Equal by definition
∇	Gradient operator
∂	Partial differential operator
d	Total differential operator

List of Abbreviations

AR	Autoregressive
EMSE	Excess Mean Square Error
FC-QLMS	Four-Channel Quaternion Least Mean Square
FC-QNGD	Four-Channel Quaternion Nonlinear Gradient Decent
FC-QRLS	Four-Channel Quaternion Recursive Least Square
i.i.d.	independent and identically distributed
IMM	Interacting Multiple Model
LMS	Least Mean Square
MA	Moving Average
MMAE	Multiple Model Adaptive Estimation
MMSE	Minimum Mean Square Error
MC	Monte Carlo
MQLMMP	Modified Quaternion Least Mean Magnitude-Phase
MSD	Mean Square Deviation
MSE	Mean Square Error
QIMM	Quaternion Interacting Multiple Model
QKF	Quaternion Kalman Filter
QLMMP	Quaternion Least Mean Magnitude-Phase
QLMP	Quaternion Least Mean Phase

QLMS	Quaternion Least Mean Square
QMMAE	Quaternion Multiple Model Adaptive Estimation
QNGD	Quaternion Nonlinear Gradient Decent
QRLS	Quaternion Recursive Least Square
QMGD	Quaternion Multivariate Gaussian Distribution
QSMM	Quaternion Static Multiple Model
SL	Strictly Linear
SL-QLMS	Strictly Linear Quaternion Least Mean Square
SL-QMMAE	Strictly Linear Quaternion Multiple Model Adaptive Estimation
SL-QNGD	Strictly Linear Quaternion Nonlinear Gradient Decent
SL-QRLS	Strictly Linear Quaternion Recursive Least Square
SMM	Static Multiple Model
SWL	Semi-Widely Linear
SWL-QIMM	Semi-Widely Linear Quaternion Interacting Multiple Model
SWL-QLMS	Semi-Widely Linear Quaternion Least Mean Square
SWL-QMMAE	Semi-Widely Linear Quaternion Multiple Model Adaptive Estimation
SWL-QSMM	Semi-Widely Linear Quaternion Static Multiple Model
WL	Widely Linear
WL-QIMM	Widely Linear Quaternion Interacting Multiple Model

WL-QKF	Widely Linear Quaternion Kalman Filter
WL-QLMS	Widely Linear Quaternion Least Mean Square
WL-QMMAE	Widely Linear Quaternion Multiple Model Adaptive Estimation
WL-QNGD	Widely Linear Quaternion Nonlinear Gradient Decent
WL-QRLS	Widely Linear Quaternion Recursive Least Square
WL-QSMM	Widely Linear Quaternion Static Multiple Model

Publications

The following publications support the material given in this thesis:

Published papers

- Min Xiang, Clive Cheong Took, and Danilo P. Mandic, “Cost-effective quaternion minimum mean square error estimation: From widely linear to four-channel processing,” *Signal Processing*, vol. 136, pp. 81-91, 2017.
- Min Xiang, Sithan Kanna, Scott C. Douglas, and Danilo P. Mandic, “Performance advantage of quaternion widely linear estimation: An approximate uncorrelating transform approach,” in *Proceedings of 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, 2016, pp. 4348-4352.
- Min Xiang, Scott C. Douglas, and Danilo P. Mandic, “The quaternion least mean magnitude phase adaptive filtering algorithm,” in *the 22nd International Conference on Digital Signal Processing (DSP)*, London, 2017.

Accepted paper

- Min Xiang, Sithan Kanna, and Danilo P. Mandic, “Performance analysis of quaternion-valued adaptive filters in non-stationary environments,” accepted by *IEEE Transactions on Signal Processing*.

Submitted papers

- Min Xiang, Shirin Enshaeifar, Alexander E. Stott, Clive Cheong Took, Yili Xia, Sithan Kanna, and Danilo P. Mandic, “Simultaneous diagonalisation of covariance and complementary covariance matrices in quaternion widely linear signal processing,” submitted to *Signal Processing*.

- Min Xiang, Bruno Scalzo Dees, and Danilo P. Mandic, “Multiple model adaptive estimation for 3D and 4D signals: A widely linear quaternion approach,” submitted to *IEEE Transactions on Neural Networks and Learning Systems*.

The following work is not included in this thesis:

- Sithan Kanna, Min Xiang, and Danilo P. Mandic, “Real-time detection of rectilinear sources for wireless communication signals,” in *Proceedings of 2015 International Symposium on Wireless Communication Systems (ISWCS)*, Brussels, 2015, pp. 541-545.

Chapter 1

Introduction

Overview

In recent years advances in sensing technology have enabled more recording from three- and four-dimensional data sources, sparking a resurgence of study on statistical signal processing based on quaternions, as quaternions are a compact and natural representation of those multidimensional phenomena and avoid problems associated with vector algebras. Compared to real variables, quaternion variables have different statistics like the noncircularity, which have not been fully exploited in the literature. In this chapter, a survey of statistical signal processing in the real, complex, and quaternion domains that are relevant to the work in this thesis is presented. This helps clarify the motivations and contributions of the work in this thesis.

1.1 Statistical Signal Processing in the Real Domain

A stochastic process is a mathematical model for a phenomenon that evolves in time in an unpredictable manner, where the unpredictability arises from effects such as interference or noise in the information link. Statistical signal processing, a popular approach to signal processing, treats signals as stochastic processes and utilises their statistical properties to perform processing. A recurring theme in statistical signal processing is estimating some sets of signals based on information obtained from measurements of other signals. This has found a number of

applications, such as signal enhancement, system identification, and noise cancellation. The quantitative metrics for the assessment of estimation performance include mean square error (MSE), mean absolute error, mean mean signed error, etc., within which the MSE is the dominant metric. Aiming at minimising the MSE between the desired signal and the estimate, the minimum mean square error (MMSE) estimation has been extensively used since the method of least squares was proposed by Carl Friedrich Gauss in the late 18th century [1].

The Wiener filter is the optimal MMSE estimator when the statistics of the signal are known and time-invariant [2]. However, the design of the Wiener filter is impossible when the statistical information is unknown or time-varying. To mitigate this limitation, adaptive filters have been proposed and widely used in the past decades [3–5]. An adaptive filter, starting from some predetermined set of initial conditions, is self-designing in that it relies on a recursive algorithm making it adapt to variations in statistics and from their interactions with the environment. In a stationary environment, the filter converges to the optimum Wiener solution in some statistical sense after successive iterations (steady state), while in a non-stationary environment, the filter can track time variations in the statistics of the input data. The convergence performance of an adaptive filter is generally evaluated in terms of its transient and steady-state behaviours. The former characterises the stability and convergence rate of the adaptive filter, whereas the latter quantifies the filtering error at the steady state. It is a standard means of analysing filtering performance to evaluate the convergence performance in the sense of mean and mean square errors [3, 4].

1.2 Statistical Signal Processing in the Complex Domain

The need for processing multidimensional signals has been highlighted by development in sensing technology. The complex domain, $\mathbb{C}$, provides a natural and convenient manner to represent the physical characteristics of signals and the trans-

formations they go through. To fully exploit the complete information carried by the interrelationship of the real and imaginary parts or the magnitude and phase of the complex-valued signal, it is desirable to carry out signal processing directly in the complex domain.

Traditionally, statistics in $\mathbb{C}$ was regarded as an extension of that in the real domain $\mathbb{R}$, and consequently complex-valued algorithms were considered as simple extensions of the established corresponding algorithms in the real domain. For instance, the covariance matrix of a complex random vector $\mathbf{x}$, $E \left\{ (\mathbf{x} - E \{ \mathbf{x} \}) (\mathbf{x} - E \{ \mathbf{x} \})^H \right\}$ ($E \{ \cdot \}$ is the statistical expectation operator), was transformed from the covariance matrix $E \left\{ (\mathbf{x} - E \{ \mathbf{x} \}) (\mathbf{x} - E \{ \mathbf{x} \})^T \right\}$ in $\mathbb{R}$ by replacing the transpose operator $(\cdot)^T$ with the conjugate transpose (Hermitian) operator $(\cdot)^H$ [6]. In this manner, the second-order moments of the complex-valued random variables were implicitly assumed to be circular (rotation-invariant) in the complex domain. Thus, the complex-valued algorithms based on this implicit assumption are only optimal for circular complex signals, and suffer performance degradation for noncircular complex signals, as the covariance does not describe the inner-relation and power imbalance between the real and imaginary components of complex random variables. Further research found that both the covariance and the pseudo-covariance, $E \left\{ (\mathbf{x} - E \{ \mathbf{x} \}) (\mathbf{x} - E \{ \mathbf{x} \})^T \right\}$, should be considered to fully describe the second-order statistical information of general complex-valued variables [7]. This is the basis of widely linear complex signal processing, which uses the augmented vector $\mathbf{x}^a = [\mathbf{x}^T, \mathbf{x}^H]^T$, contrary to the strictly linear processing using only $\mathbf{x}$. The study of covariance and pseudo-covariance matrices in widely linear complex signal processing have received significant attention.

In gradient-based optimisation with respect to complex-valued variables, the $\mathbb{C}\mathbb{R}$ calculus (also known as Wirtinger calculus) [8,9] has been introduced to calculate the partial derivatives of real-valued functions with respect to complex-valued variables, which are not complex analytic, conditioned by the Cauchy-Riemann equations, and therefore are not differentiable in the standard complex derivative

sense. The resulted complex derivatives are an extension of the standard complex derivatives to non-analytic complex functions. The $\mathbb{CR}$ calculus allows for the consideration of general complex functions in a unified manner, and greatly simplifies the differentiation and analysis of the complex functions. The gradient decent method based on the $\mathbb{CR}$ calculus has been applied to a number of complex signal processing techniques, like the complex adaptive filtering algorithms [10].

1.3 Statistical Signal Processing in the Quaternion Domain

Quaternions, a four-dimensional skew field, were first described by William Rowan Hamilton in 1843 when he was posed with the problem to extend complex numbers into higher dimensions [11]. A quaternion has one-dimensional real component and three-dimensional imaginary component. Frobenius' finding that quaternion algebra is the highest associative division algebra made it more attractive to work with as compared to other hypercomplex spaces. Quaternions were also the first non-commutative division algebra to be discovered. Because of the non-commutativity of the quaternion algebra, quaternions have many different properties from real and complex numbers [12, 13].

Quaternions have been extensively used in mathematics and physics, in particular for calculations involving three-dimensional rotations, as their division algebra rectifies the gimbal lock problem associated with vector algebras, and pure quaternions provide a more concise and computationally efficient representation of three-dimensional rotations than rotation matrices [14]. With the four-dimensional nature, quaternions have been used to represent a number of three- and four-dimensional phenomena, such as image data, voltages in three-phase electric power systems, locations and orientations of objects in three-dimensional space, space-time-polarization block codes in communications, and Pauli matrices in quantum mechanics. Quaternions provide a compact model of mutual information between multiple dimensions and the inherent physically meaningful interpretation

for multidimensional data. Recently, quaternions have seen increasing applications in communications, motion tracking, and biomedical signal processing [15–18], since the state-of-the-art advances in sensing technology have enabled more recording from three- and four-dimensional data sources, such as measurements from seismometers [19], ultrasonic anemometers [20], and inertial body sensors [17].

Compared to real- and complex-valued statistical signal processing, quaternion-valued statistical signal processing is still in its infancy. Due to the lack of a suitable mathematical framework, most traditional quaternion signal processing techniques took the approach of essentially transforming quaternions into a vector of quadrivariate real-valued components, where most advantages of quaternions such as their division algebra and their underlying physical interpretation were lost. Within the quaternion domain, some algorithms were introduced as a straightforward extension of the counterpart algorithms in real and complex domains, where the freedom inherent in quaternions were lost and the performance degraded for general quaternion signals. The recently introduced mathematical tools for developing rigorous quaternion signal processing techniques, such as augmented second-order quaternion statistics [21, 22] and the $\mathbb{H}\mathbb{R}$ and generalised $\mathbb{H}\mathbb{R}$ calculi [23, 24], have triggered a resurgence of research on quaternion-valued signal processing including estimation [25], neural networks [26], independent component analysis [27], and Fourier transforms [28].

A mathematical framework termed augmented quaternion statistics has been established to exploit complete second-order statistical information of noncircular quaternion random variables [21, 22]. A quaternion variable is called circular if its probability distribution is rotation invariant, implying the independence and power balance between the real and imaginary components. Specifically, it is called second-order circular (proper) if its second-order moments are rotationally invariant; otherwise it is called second-order noncircular (improper)¹. Although the

¹In this thesis, only the lower-order quaternion statistics is considered, so the terms “circular/noncircular” and “proper/improper” are used interchangeably.

standard covariance fully explains the second-order statistics of circular quaternion variables, it does not characterise the unique second-order statistics of noncircular quaternion variables. Whereas, the unique statistical information is contained in the complementary covariances, the cross-covariances between the quaternion variable, $\mathbf{x}$, and its rotations around the three imaginary axes (involutions), $\mathbf{x}^i, \mathbf{x}^j, \mathbf{x}^k$. The augmented quaternion statistics form a basis for widely linear quaternion signal processing, which uses the augmented vector $\mathbf{x}^a = [\mathbf{x}^T, \mathbf{x}^{iT}, \mathbf{x}^{jT}, \mathbf{x}^{kT}]^T$, contrary to the strictly linear processing, which uses only $\mathbf{x}$ [21, 29]. The performance advantage of widely linear processing over the strictly linear counterpart has been proven in an indirect fashion via an intermediate, the semi-widely linear processing [30, 31].

Similar to the complex case, the non-differentiability of quaternion functions is a major block in gradient-based optimisation with respect to quaternion variables. It is known that a quaternion-valued function of quaternion-valued variables is quaternion differentiable if and only if its partial derivatives satisfy the Cauchy-Riemann-Fueter equation. But this condition does not hold for general polynomial functions or real-valued functions. Following the same path as the $\mathbb{C}\mathbb{R}$ calculus, a framework termed the $\mathbb{H}\mathbb{R}$ calculus [23], which allows for calculating the derivatives of quaternion functions directly in the quaternion domain, was developed by establishing a duality between quaternion functions in $\mathbb{H}$ and their corresponding quadrivariate real functions in $\mathbb{R}^4$. This theory has been further developed into the generalized $\mathbb{H}\mathbb{R}$ calculus [24, 32, 33] by employing quaternion rotations in a general orthogonal system, leading to novel product and chain rules of derivatives, which can be used to solve complicated quaternion optimisation problems.

1.4 Contributions of the Thesis

The focus of this work is on the development of theory and methods of statistical signal processing for noncircular quaternion random variables. The contributions of this thesis can be divided into three areas, roughly corresponding to how this

thesis is organised, as shown in Figure 1.1.

1. Simultaneous approximate diagonalisation of covariance and complementary covariance matrices

A novel technique is proposed for a simultaneous approximate decomposition of the covariance matrix and the three complementary covariance matrices of noncircular quaternion random variables. A singular value decomposition is proven to be sufficient to approximately diagonalise all of the four covariance matrices. This forms a basis for the proposed quaternion approximate uncorrelating transform (QAUT) which can be used to decorrelate improper quaternion signals. This contribution is elaborated in Chapter 3 and was introduced in [34].

Relevant Publication:

[34] Min Xiang, Shirin Enshaeifar, Alexander Stott, Clive Cheong Took, Yili Xia, Sithan Kanna, and Danilo P. Mandic, “Simultaneous diagonalisation of covariance and complementary covariance matrices in quaternion widely linear signal processing,” submitted to *Signal Processing*.

2. Performance analysis of existing MMSE estimation techniques

- Performance advantage of widely linear quaternion estimation

The performance advantage of widely linear processing over strictly linear processing in quaternion MMSE estimation is proved by comparing their optimal solutions. The relationship between this performance advantage and the properties of the quaternion input signal, especially the noncircularity, is established. This is achieved by examining the structure of the covariance and complementary covariance matrices of the input signal and by employing the QAUT. This contribution is elaborated in Chapter 4 and was introduced in [35].

Relevant Publication:

[35] Min Xiang, Sithan Kanna, Scott C. Douglas and Danilo P. Mandic, “Performance advantage of quaternion widely linear estimation: An approximate uncorrelating transform approach,” in *Proceedings of 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, 2016, pp. 4348-4352.

- Convergence performance of quaternion adaptive filters

An analysis framework for the mean and mean square convergence analyses of general quaternion adaptive filters in non-stationary environments is established. Transient and steady-state performance of a general class of quaternion adaptive filters for noncircular input signals are provided by exploiting the augmented quaternion statistics. The exploitation of QAUT produces intuitive closed-form expressions for the performance of quaternion least mean square (QLMS) filters with Gaussian inputs, providing new insights into the relationship between the filter behaviour and the complete second-order statistics of the input signal, that is, quaternion noncircularity. The closed-form expressions for the performance of strictly linear, semi-widely linear, and widely linear QLMS filters are investigated in detail. This contribution is elaborated in Chapter 5 and was introduced in [36].

Relevant Publication:

[36] Min Xiang, Sithan Kanna, and Danilo P. Mandic, “Performance analysis of quaternion-valued adaptive filters in non-stationary environments”, submitted to *IEEE Transactions on Signal Processing*.

3. Novel adaptive estimation algorithms

- Quaternion least mean magnitude-phase adaptive filtering algorithm

The MSE criterion widely used in existing quaternion-valued estimation techniques has only one degree of freedom. For the separate control of the magnitude and phase, a novel estimation criterion allowing for more

degrees of freedom is proposed. Based on the criterion, a quaternion least mean magnitude-phase (QLMMP) filtering algorithm is derived and is shown to provide a unified form for the QLMS algorithm, a magnitude filtering algorithm, and a phase filtering algorithm. The convergence analysis of QLMMP employs a connection to that of the QLMS with a quaternion-valued stepsize. This contribution is elaborated in Chapter 6 and was introduced in [37].

Relevant Publication:

[37] Min Xiang, Scott C. Douglas, and Danilo P. Mandic, “The quaternion least mean magnitude phase adaptive filtering algorithm,” in *the 22nd International Conference on Digital Signal Processing (DSP)*, London, 2017.

- Four-channel quaternion estimation

Quaternion-valued estimation algorithms are computationally intensive (e.g. one quaternion multiplication requires eight real multiplications) while standard methods to reduce this computational complexity cast the problem in the real domain, where the physical meaning inherent in the quaternion domain is lost. A four-channel quaternion estimation method is proposed by decomposing the standard MSE into the MSEs in estimating the real and imaginary parts. This allows for the development of computationally efficient quaternion-valued estimation algorithms while the processing still remains within the quaternion domain. This contribution is elaborated in Chapter 7 and was introduced in [38].

Relevant Publication:

[38] Min Xiang, Clive Cheong Took, and Danilo P. Mandic, “Cost-effective quaternion minimum mean square error estimation: From widely linear to four-channel processing,” *Signal Processing*, vol. 136, pp. 81-91, 2017.

- Widely linear quaternion multiple model adaptive estimation

Existing quaternion state estimation techniques are only suitable for dynamical systems represented by a single known model [39]. To deal with model uncertainty in real-world problems, a class of widely linear quaternion multiple model adaptive estimation (QMMAE) algorithms are introduced, based on widely linear quaternion Kalman filters and Bayesian inference. Within the widely linear QMMAE framework, a widely linear quaternion interacting multiple model algorithm is proposed to track time-variant model uncertainty, while a widely linear quaternion static multiple model algorithm is proposed for time-invariant model uncertainty. To efficiently evaluate the estimation error, a recursive algorithm for the performance analysis of the QMMAE algorithms is also proposed. This contribution is elaborated in Chapter 8 and was introduced in [40].

Relevant Publication:

[37] Min Xiang, Bruno Scalzo Dees, and Danilo P. Mandic, “Multiple model adaptive estimation for 3D and 4D signals: A widely linear quaternion approach”, submitted to *IEEE Transactions on Neural Networks and Learning Systems*.

1.5 Organisation of the Thesis

Figure 1.1 shows an overview of structure of the thesis which is organised as follows:

- **Chapter 2:** Statistical signal processing in the quaternion domain

The aim of this chapter is to provide the reader with the fundamental knowledge of quaternions and the necessary background on quaternion-valued statistical signal processing. This chapter includes an introduction to quaternion algebra, generalised $\mathbb{H}\mathbb{R}$ calculus, augmented second-order quaternion statistics, and widely linear quaternion estimation, within which some

estimation algorithms including QLMS and quaternion Kalman filtering algorithms are introduced.

The chapters outlining the contribution of the thesis are grouped into three parts.

Part I: Simultaneous diagonalisation of covariance and complementary covariance matrices

- **Chapter 3:** Quaternion approximate uncorrelating transform

This chapter introduces a simultaneous approximate decomposition of the covariance matrix and the three complementary covariance matrices of non-circular quaternion random variables. It has been shown that a singular value decomposition is sufficient to diagonalise approximately all of the four covariance matrices. This forms a basis for the proposed QAUT which can be used to decorrelate noncircular quaternion variables.

Part II: Performance analysis of existing MMSE estimation techniques

- **Chapter 4:** Performance advantage of widely linear quaternion estimation

This chapter provides a proof of the performance advantage of widely linear processing over strictly linear processing in quaternion MMSE estimation. By examining the structure of the covariance and complementary covariance matrices of the input signal and by employing the QAUT, the relationship between the performance advantage and the signal properties, including the noncircularity and power of the input signal and the degree of widely linear nature of the system, is established.

- **Chapter 5:** Convergence performance of quaternion adaptive filters in non-stationary environments

This chapter extends an analysis framework for real-valued adaptive filters to the mean and mean square convergence analyses of a general class of

quaternion adaptive filters in non-stationary environments. To model a non-stationary environment, the system weights are assumed to vary according to a first-order random-walk model. Transient and steady-state performance of the filters for noncircular input signals are derived. The exploitation of QAUT produces intuitive closed-form expressions for the performance of QLMS filters with Gaussian inputs, providing new insights into the relationship between the filter behaviour and the complete second-order statistics of the input signal, that is, quaternion noncircularity. The convergence performance of strictly linear, semi-widely linear, and widely linear QLMS filters are investigated in detail.

Part III: Novel adaptive estimation algorithms

- **Chapter 6:** Quaternion least mean magnitude-phase adaptive filter

This chapter decomposes the MSE cost function into the cost functions for estimating the magnitude and phase of a quaternion signal, whereby the magnitude and phase can be estimated separately. Accordingly, a QLMMP adaptive filtering algorithm is proposed. The QLMMP not only outperforms the existing QLMS algorithm by virtue of the enhanced degrees of freedom, but also provides a unified form for the QLMS, a magnitude filtering, and a phase filtering algorithm. Employing a connection to the normalised QLMS with a quaternion-valued step size, the convergence performance of the normalised QLMMP is analysed.

- **Chapter 7:** Low-complexity quaternion adaptive filters

This chapter elaborates on the four-channel estimation methodology which decomposes the MSE cost function into the MSEs in estimating the real and imaginary parts of a signal so that separate estimators can be used to minimise the our cost functions independently. The four-channel estimation scheme allows for the necessary degrees of freedom to model general

quaternion-valued signals, and enables computationally efficient quaternion adaptive filters while the processing still remains within the quaternion domain.

- **Chapter 8:** Widely linear quaternion multiple model adaptive estimation

In order to deal with model uncertainty in quaternion state estimation, this chapter proposes a class of widely linear quaternion multiple model adaptive estimation QMMAE algorithms based on Kalman filters and Bayesian inference. Within the proposed framework, a widely linear quaternion interacting multiple model algorithm is proposed to track time-variant model uncertainty, while a widely linear quaternion static multiple model algorithm is proposed for time-invariant model uncertainty. As a cost-effective alternative to the traditionally used Monte Carlo simulation method, a recursive algorithm for the performance analysis of the QMMAE algorithms is also proposed.

- **Chapter 9:** Conclusions and future work

Concluding remarks drawn from the thesis and ideas for future works are outlined in this chapter.

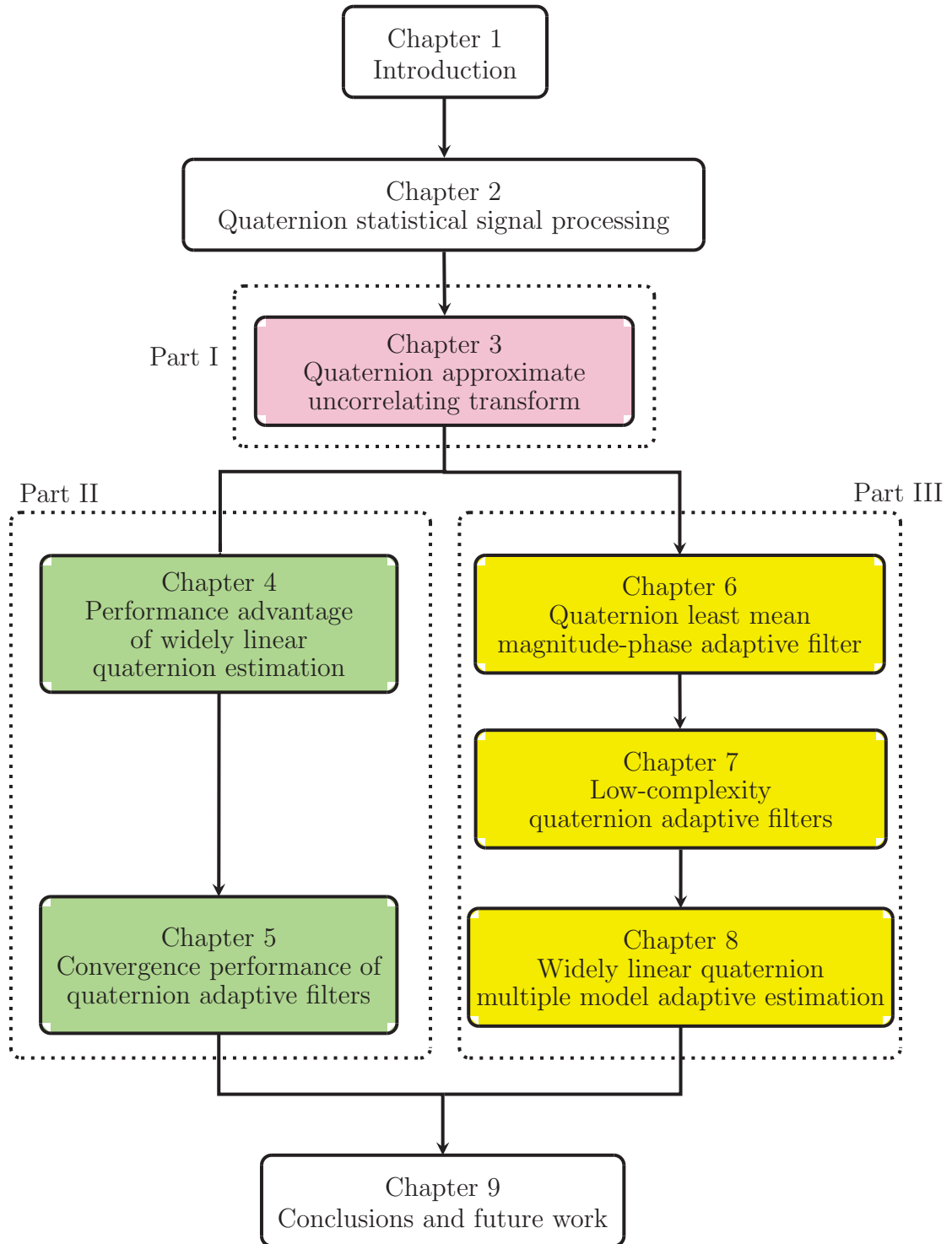


Figure 1.1: Thesis structure.

Chapter 2

Statistical Signal Processing in the Quaternion Domain

Overview

This chapter provides the reader with the fundamental knowledge of quaternions and necessary background on quaternion-valued statistical signal processing. It includes an introduction to quaternion algebra, generalised $\mathbb{H}\mathbb{R}$ calculus, augmented second-order quaternion statistics, and widely linear quaternion estimation. Some popular quaternion estimation algorithms including QLMS and quaternion Kalman filtering algorithms are reviewed.

2.1 Quaternion Algebra

The quaternion domain $\mathbb{H}$ is a four-dimensional vector space over the real field $\mathbb{R}$, spanned by the basis $\{1, \imath, \jmath, \kappa\}$. A random quaternion variable $x \in \mathbb{H}$ consists of a real part $\Re[\cdot]$ and an imaginary part $\Im[\cdot]$ which comprises three imaginary components, so that

$$x = \Re[x] + \Im[x] = \Re[x] + \Im_{\imath}[x]\imath + \Im_{\jmath}[x]\jmath + \Im_{\kappa}[x]\kappa \quad (2.1)$$

where $\Re[x], \Im_i[x], \Im_j[x], \Im_\kappa[x]$ are real variables and $\imath, j, \kappa$ are imaginary units with the properties

$$\begin{aligned} \imath^2 = j^2 = \kappa^2 &= -1, & \imath j &= -j \imath = \kappa \\ j \kappa &= -\kappa j = \imath, & \kappa \imath &= -\imath \kappa = j \end{aligned}$$

A quaternion can be represented by a pair of complex numbers through the Cayley-Dickson construction [41, 42], that is

$$x = z_1 + z_2 j \quad (2.2)$$

where

$$z_1 = \Re[x] + \Im_i[x] \imath, \quad z_2 = \Im_j[x] + \Im_\kappa[x] \imath \quad (2.3)$$

The conjugate of x is defined as

$$x^* = \Re[x] - \Im[x] = \Re[x] - \Im_i[x] \imath - \Im_j[x] j - \Im_\kappa[x] \kappa \quad (2.4)$$

The modulus of x is given by

$$|x| = \sqrt{\Re^2[x] + \Im_i^2[x] + \Im_j^2[x] + \Im_\kappa^2[x]}$$

and the product of two quaternions $x, y \in \mathbb{H}$ by

$$xy = \Re[x]\Re[y] - \Im[x] \cdot \Im[y] + \Re[x]\Im[y] + \Re[y]\Im[x] + \Im[x] \times \Im[y]$$

where “ $\cdot$ ” denotes the scalar product and “ $\times$ ” the vector product. The presence of the vector product causes non-commutativity of the quaternion product, that is, $xy \neq yx$. The quaternion product has the following properties:

$$|xy| = |x||y|, \quad (xy)^{-1} = y^{-1}x^{-1}, \quad (xy)^* = y^*x^*$$

A quaternion x is called a pure quaternion if $\Re[x] = 0$, and it is called a unit quaternion if $|x| = 1$. Similar to complex numbers, a quaternion x can be represented in the polar form as $x = |x| e^\zeta$, where ζ is a pure quaternion, and the exponential can be represented by the sum $e^\zeta = \sum_{n=0}^{\infty} \frac{\zeta^n}{n!}$ and satisfies $(e^\zeta)^* = e^{-\zeta}$ [43].

Another important notion is the quaternion involution [44], which defines a self-inverse mapping, analogous to the complex conjugate. The general involution of the quaternion variable x is defined as $x^\alpha = -\alpha x \alpha$, which represents the rotation of the vector part of x by π about a unit pure quaternion α . The quaternion involutions have the property: $(x^\alpha)^\alpha = x$. Accordingly, $(x^\alpha)^* = (x^*)^\alpha$, and they can be compactly denoted by $x^{\alpha*}$. Widely used in quaternion signal processing, three special cases of involutions about the $\imath$, j and κ imaginary axes are given by [44]

$$x^\imath = -\imath x \imath = \Re[x] + \Im_\imath[x] \imath - \Im_j[x] j - \Im_\kappa[x] \kappa \quad (2.5)$$

$$x^j = -j x j = \Re[x] - \Im_\imath[x] \imath + \Im_j[x] j - \Im_\kappa[x] \kappa \quad (2.6)$$

$$x^\kappa = -\kappa x \kappa = \Re[x] - \Im_\imath[x] \imath - \Im_j[x] j + \Im_\kappa[x] \kappa \quad (2.7)$$

Through the above defined quaternion involutions, the four real-valued components of $\mathbf{x}$ can be expressed as

$$\Re[x] = \frac{1}{4} (x + x^\imath + x^j + x^\kappa) \quad (2.8)$$

$$\Im_\imath[x] = -\frac{1}{4} \imath (x + x^\imath - x^j - x^\kappa) \quad (2.9)$$

$$\Im_j[x] = -\frac{1}{4} j (x - x^\imath + x^j - x^\kappa) \quad (2.10)$$

$$\Im_\kappa[x] = -\frac{1}{4} \kappa (x - x^\imath - x^j + x^\kappa) \quad (2.11)$$

2.1.1 Quaternion Matrices

For a general quaternion matrix $\mathbf{A}$, $(\mathbf{A}^*)^T = (\mathbf{A}^T)^*$, $(\mathbf{A}^*)^\alpha = (\mathbf{A}^\alpha)^*$, $(\mathbf{A}^\alpha)^T = (\mathbf{A}^T)^\alpha$, $(\mathbf{A}^\alpha)^H = (\mathbf{A}^H)^\alpha$, $(\mathbf{A}^\alpha)^\beta = (\mathbf{A}^\beta)^\alpha = \mathbf{A}^\gamma$, for distinct $\alpha, \beta, \gamma \in \{\iota, j, \kappa\}$. Two general multiplicative quaternion matrices, $\mathbf{A}$ and $\mathbf{B}$, have the properties: $(\mathbf{AB})^H = \mathbf{B}^H \mathbf{A}^H$, $(\mathbf{AB})^\alpha = \mathbf{A}^\alpha \mathbf{B}^\alpha$.

Due to the non-commutativity of quaternion products, quaternion matrices have many different properties from real and complex matrices. For example, for two general multiplicative quaternion matrices, $\mathbf{A}$ and $\mathbf{B}$, $(\mathbf{AB})^T \neq \mathbf{B}^T \mathbf{A}^T$, $(\mathbf{AB})^* \neq \mathbf{A}^* \mathbf{B}^*$. A quaternion square matrix $\mathbf{A} \in \mathbb{H}^{L \times L}$ has two types of eigenvalues [12]. If $\mathbf{Ax} = \mathbf{x}\lambda$ for $\lambda \in \mathbb{H}$ and some non-zero $\mathbf{x} \in \mathbb{H}^{L \times 1}$, then λ is called a right eigenvalue of $\mathbf{A}$, and $\mathbf{x}$ is a right eigenvector of $\mathbf{A}$. If $\mathbf{Ax} = \lambda\mathbf{x}$ for $\lambda \in \mathbb{H}$ and some non-zero $\mathbf{x} \in \mathbb{H}^{L \times 1}$, then λ is called a left eigenvalue of $\mathbf{A}$, and $\mathbf{x}$ is a left eigenvector of $\mathbf{A}$. Right eigenvalues have been well studied in literature, while left eigenvalues are less known and are not computationally well-posed [12, 45–47]. Relevant to our analysis, a Hermitian quaternion matrix $\mathbf{H} \in \mathbb{H}^{L \times L}$ has exactly L right eigenvalues, which are also left eigenvalues and are real-valued. Note that $\mathbf{H}$ can also have non-real left eigenvalues. There exists a unitary quaternion matrix $\mathbf{U} \in \mathbb{H}^{L \times L}$, which satisfies $\mathbf{U}^H \mathbf{U} = \mathbf{U} \mathbf{U}^H = \mathbf{I}_M$, such that $\mathbf{H} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^H$ is the eigendecomposition of $\mathbf{H}$ with the right eigenvalues of $\mathbf{H}$ on the diagonal of a diagonal real matrix, $\mathbf{\Lambda}$. A Hermitian quaternion matrix is called positive definite, positive semi-definite, negative definite, or negative semi-definite if $\mathbf{x}^H \mathbf{H} \mathbf{x}$ is respectively positive, non-negative, negative, or non-positive for any non-zero quaternion vector $\mathbf{x} \in \mathbb{H}^{L \times 1}$. This is the case if and only if all right eigenvalues of $\mathbf{H}$ are positive, non-negative, negative, or non-positive. In the rest of the thesis, without loss in generality, the eigendecomposition of Hermitian quaternion matrices, $\mathbf{H} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^H$, which only involves right eigenvalues, will be applied. For conciseness, use the terminology, eigenvalues and eigenvectors, to denote right eigenvalues and right eigenvectors of Hermitian quaternion matrices [13].

The Weyl's inequality about the eigenvalues of Hermitian complex matrices also applies to Hermitian quaternion matrices. If $\mathbf{A}$ and $\mathbf{B}$ are Hermitian quaternion matrices with eigenvalues $\lambda_1(\mathbf{A}) \geq \lambda_2(\mathbf{A}) \geq \dots \geq \lambda_L(\mathbf{A})$, $\lambda_1(\mathbf{B}) \geq \lambda_2(\mathbf{B}) \geq \dots \geq \lambda_L(\mathbf{B})$, then $\mathbf{C} = \mathbf{A} + \mathbf{B}$ is also Hermitian with eigenvalues $\lambda_1(\mathbf{C}) \geq \lambda_2(\mathbf{C}) \geq \dots \geq \lambda_L(\mathbf{C})$, satisfying

$$\begin{aligned} \lambda_i(\mathbf{A}) + \lambda_j(\mathbf{B}) &\leq \lambda_m(\mathbf{C}) \leq \lambda_k(\mathbf{A}) + \lambda_l(\mathbf{B}) \\ \text{if } k + l - 1 &\leq m \leq i + j - 1 \end{aligned} \quad (2.12)$$

$\forall i, j, k, l \in \{1, 2, \dots, L\}$ [48].

The singular value decomposition (SVD) has also been recently extended to the quaternion domain [12]. For any $\mathbf{A} \in \mathbb{H}^{L_1 \times L_2}$, there exist unitary quaternion matrices $\mathbf{U} \in \mathbb{H}^{L_1 \times L_1}$ and $\mathbf{V} \in \mathbb{H}^{L_2 \times L_2}$ such that $\mathbf{A} = \mathbf{U}\mathbf{M}\mathbf{V}$, where $\mathbf{M} \in \mathbb{R}^{L_1 \times L_2}$ is a diagonal matrix (may be rectangular) with non-negative diagonal elements. The positive diagonal elements of $\mathbf{M}$ are singular values of $\mathbf{A}$.

2.2 Generalised $\mathbb{H}\mathbb{R}$ Calculus

The differentiability of quaternion-valued functions of quaternion variables is challenging to deal with due to the non-commutative product. A function $f(x) : \mathbb{H} \rightarrow \mathbb{H}$ is defined to be quaternion-differentiable on the left at a point $x \in \mathbb{H}$ if the limit

$$\lim_{h \rightarrow 0} h^{-1} [f(x+h) - f(x)]$$

exists. It has been proven that $f(x)$ is quaternion-differentiable on the left if and only if the partial derivatives of f satisfy the left Cauchy-Riemann-Fueter equation given by

$$\frac{\partial f}{\partial \Re[x]} + i \frac{\partial f}{\partial \Im_i[x]} + j \frac{\partial f}{\partial \Im_j[x]} + \kappa \frac{\partial f}{\partial \Im_\kappa[x]} = 0$$

Likewise, $f(x)$ is defined to be quaternion-differentiable on the right at $x \in \mathbb{H}$ if the limit

$$\lim_{h \rightarrow 0} [f(x+h) - f(x)] h^{-1}$$

exists. $f(x)$ is quaternion-differentiable on the right if and only if the partial derivatives of f satisfy the right Cauchy-Riemann-Fueter equation given by

$$\frac{\partial f}{\partial \Re[x]} + \frac{\partial f}{\partial \Im_i[x]}i + \frac{\partial f}{\partial \Im_j[x]}j + \frac{\partial f}{\partial \Im_\kappa[x]}\kappa = 0$$

However, general polynomial functions (even the identity $f(x) = x$) and real-valued functions ($f(x) : \mathbb{H} \rightarrow \mathbb{R}$) do not satisfy the above Cauchy-Riemann-Fueter equations. This has been a major block in solving quaternion optimisation problems, most of which require the minimisation of a real-valued function of quaternion-valued variables. An elegant solution to this problem is the generalised $\mathbb{H}\mathbb{R}$ calculus [24], which includes the left and right forms because of the non-commutative quaternion multiplication. The left generalised $\mathbb{H}\mathbb{R}$ derivatives are defined as

$$\begin{aligned} \frac{\partial f}{\partial x^q} &\triangleq \frac{1}{4} \left(\frac{\partial f}{\partial \Re[x]} - \frac{\partial f}{\partial \Im_i[x]}i^q - \frac{\partial f}{\partial \Im_j[x]}j^q - \frac{\partial f}{\partial \Im_\kappa[x]}\kappa^q \right) \\ \frac{\partial f}{\partial x^{q*}} &\triangleq \frac{1}{4} \left(\frac{\partial f}{\partial \Re[x]} + \frac{\partial f}{\partial \Im_i[x]}i^q + \frac{\partial f}{\partial \Im_j[x]}j^q + \frac{\partial f}{\partial \Im_\kappa[x]}\kappa^q \right) \end{aligned}$$

while the right generalised $\mathbb{H}\mathbb{R}$ derivatives are defined as

$$\begin{aligned} \frac{\partial_r f}{\partial x^q} &\triangleq \frac{1}{4} \left(\frac{\partial f}{\partial \Re[x]} - i^q \frac{\partial f}{\partial \Im_i[x]} - j^q \frac{\partial f}{\partial \Im_j[x]} - \kappa^q \frac{\partial f}{\partial \Im_\kappa[x]} \right) \\ \frac{\partial_r f}{\partial x^{q*}} &\triangleq \frac{1}{4} \left(\frac{\partial f}{\partial \Re[x]} + i^q \frac{\partial f}{\partial \Im_i[x]} + j^q \frac{\partial f}{\partial \Im_j[x]} + \kappa^q \frac{\partial f}{\partial \Im_\kappa[x]} \right) \end{aligned}$$

for any $q \in \mathbb{H}$ and $q \neq 0$.

Based on the left and right generalised $\mathbb{H}\mathbb{R}$ derivatives, the differential of $f(x)$

can be represented by various forms given by

$$\begin{aligned}
df &= \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial x^i} dx^i + \frac{\partial f}{\partial x^j} dx^j + \frac{\partial f}{\partial x^\kappa} dx^\kappa \\
&= \frac{\partial f}{\partial x^*} dx^* + \frac{\partial f}{\partial x^{i*}} dx^{i*} + \frac{\partial f}{\partial x^{j*}} dx^{j*} + \frac{\partial f}{\partial x^{\kappa*}} dx^{\kappa*} \\
&= dx \frac{\partial f}{\partial x} + dx^i \frac{\partial f}{\partial x^i} + dx^j \frac{\partial f}{\partial x^j} + dx^\kappa \frac{\partial f}{\partial x^\kappa} \\
&= dx^* \frac{\partial f}{\partial x^*} + dx^{i*} \frac{\partial f}{\partial x^{i*}} + dx^{j*} \frac{\partial f}{\partial x^{j*}} + dx^{\kappa*} \frac{\partial f}{\partial x^{\kappa*}}
\end{aligned}$$

Both the left and right generalised $\mathbb{H}\mathbb{R}$ derivatives can be used in optimisation in the quaternion domain. For $f(x) : \mathbb{H} \rightarrow \mathbb{R}$, the left and right generalised $\mathbb{H}\mathbb{R}$ derivatives are identical. As the optimisation of real-valued functions of quaternion variables are studied in the thesis, the left generalised $\mathbb{H}\mathbb{R}$ derivatives will be used in subsequent analysis. They satisfy the product and chain rules given by

$$\frac{\partial (fg)}{\partial x^q} = f \frac{\partial g}{\partial x^q} + \frac{\partial f}{\partial x^{q*}} g$$

$$\frac{\partial (fg)}{\partial x^{q*}} = f \frac{\partial g}{\partial x^{q*}} + \frac{\partial f}{\partial x^{qq*}} g$$

$$\begin{aligned}
\frac{\partial f(g(x))}{\partial x^q} &= \frac{\partial f}{\partial g^\nu} \frac{\partial g^\nu}{\partial x^q} + \frac{\partial f}{\partial g^{\nu i}} \frac{\partial g^{\nu i}}{\partial x^q} + \frac{\partial f}{\partial g^{\nu j}} \frac{\partial g^{\nu j}}{\partial x^q} + \frac{\partial f}{\partial g^{\nu k}} \frac{\partial g^{\nu k}}{\partial x^q} \\
&= \frac{\partial f}{\partial g^{\nu*}} \frac{\partial g^{\nu*}}{\partial x^q} + \frac{\partial f}{\partial g^{\nu i*}} \frac{\partial g^{\nu i*}}{\partial x^q} + \frac{\partial f}{\partial g^{\nu j*}} \frac{\partial g^{\nu j*}}{\partial x^q} + \frac{\partial f}{\partial g^{\nu k*}} \frac{\partial g^{\nu k*}}{\partial x^q} \\
\frac{\partial f(g(x))}{\partial x^{q*}} &= \frac{\partial f}{\partial g^\nu} \frac{\partial g^\nu}{\partial x^{q*}} + \frac{\partial f}{\partial g^{\nu i}} \frac{\partial g^{\nu i}}{\partial x^{q*}} + \frac{\partial f}{\partial g^{\nu j}} \frac{\partial g^{\nu j}}{\partial x^{q*}} + \frac{\partial f}{\partial g^{\nu k}} \frac{\partial g^{\nu k}}{\partial x^{q*}} \\
&= \frac{\partial f}{\partial g^{\nu*}} \frac{\partial g^{\nu*}}{\partial x^{q*}} + \frac{\partial f}{\partial g^{\nu i*}} \frac{\partial g^{\nu i*}}{\partial x^{q*}} + \frac{\partial f}{\partial g^{\nu j*}} \frac{\partial g^{\nu j*}}{\partial x^{q*}} + \frac{\partial f}{\partial g^{\nu k*}} \frac{\partial g^{\nu k*}}{\partial x^{q*}}
\end{aligned}$$

It has been proven that the derivative $\frac{\partial f}{\partial x^*}$ defines the direction of the maximum rate of the change in $f(x)$.

It is straightforward to extend the above theory to the derivatives of quaternion-valued scalar functions with respect to quaternion vectors [32].

2.3 Augmented Second-order Quaternion Statistics

2.3.1 Quaternion Properness

The set of involutions in (2.5)-(2.7), together with the original quaternion x , forms the most frequently used basis for augmented quaternion statistics, the core of the recently proposed widely linear processing methodology [21, 22]. Benefiting from the involution basis, the augmented second-order statistics of a zero-mean random quaternion variable x is exploited by the $\imath$ -, j -, and κ - covariances, $C_{xx^\alpha} = E \{(x - E \{x\})(x - E \{x\})^{\alpha*}\}$, $\alpha = \imath, j, \kappa$, which are referred to as complementary covariances, together with the standard covariance $C_{xx} = E \{(x - E \{x\})(x - E \{x\})^*\}$. These covariances, taken together, enable the characterisation of quaternion impropriety (second-order noncircularity) arising from the degree of correlation and/or power imbalance between the imaginary components relative to the real component. The impropriety coefficients of a random quaternion variable x are defined as [49, 50]

$$r_\alpha \triangleq \left| \frac{C_{xx^\alpha}}{C_{xx}} \right|, \quad \alpha = \imath, j, \kappa \quad (2.13)$$

which measure the correlation between x and each of its involutions, normalised by the signal power. Note that $r_\alpha \in [0, 1]$. The four degrees of freedom in the quaternion domain allow for different notions of properness: $\mathbb{H}$ -properness, $\mathbb{R}^\alpha$ -properness and $\mathbb{C}^\alpha$ -properness.

A random quaternion variable x is $\mathbb{H}$ -proper if it is uncorrelated with its involutions $x^\imath$, x^j and x^κ , so that $C_{xx^\imath} = C_{xx^j} = C_{xx^\kappa} = \mathbf{0}$; x is $\mathbb{R}^\alpha$ -proper if it is only uncorrelated with the involution x^α , so that only C_{xx^α} among the three complementary covariances vanishes; x is $\mathbb{C}^\alpha$ -improper if it is only correlated with the involution x^α , so that all complementary covariances except C_{xx^α} vanish; x

is maximally improper (rectilinear) if its impropriety coefficients are maximal, so that $r_i = r_j = r_\kappa = 1$.

Similarly, the augmented second-order statistics of a zero-mean random quaternion column vector $\mathbf{x}$ are exploited by the i -, j - and κ - covariance matrices, $\mathbf{C}_{\mathbf{xx}^\alpha} \triangleq E\{(\mathbf{x} - E\{\mathbf{x}\})(\mathbf{x} - E\{\mathbf{x}\})^{\alpha H}\}$, $\alpha \in \{i, j, \kappa\}$, referred to as complementary covariance matrices, together with the standard Hermitian covariance matrix, $\mathbf{C}_{\mathbf{xx}} \triangleq E\{(\mathbf{x} - E\{\mathbf{x}\})(\mathbf{x} - E\{\mathbf{x}\})^H\}$. The α -complementary covariance matrix is α -Hermitian, which means $\mathbf{C}_{\mathbf{xx}^\alpha} = (\mathbf{C}_{\mathbf{xx}^\alpha})^{\alpha H}$. The knowledge of both the covariance and the three complementary covariance matrices is necessary to ensure the utilisation of complete second-order statistical information of $\mathbf{x}$ in $\mathbb{H}$.

A random quaternion vector $\mathbf{x}$ is $\mathbb{H}$ -proper if it is uncorrelated with its involutions $\mathbf{x}^i$, $\mathbf{x}^j$ and $\mathbf{x}^\kappa$, so that $\mathbf{C}_{\mathbf{xx}^i} = \mathbf{C}_{\mathbf{xx}^j} = \mathbf{C}_{\mathbf{xx}^\kappa} = \mathbf{0}$; $\mathbf{x}$ is $\mathbb{R}^\alpha$ -proper if it is only uncorrelated with the involution $\mathbf{x}^\alpha$, so that only $\mathbf{C}_{\mathbf{xx}^\alpha}$ among the three complementary covariances vanishes; $\mathbf{x}$ is $\mathbb{C}^\alpha$ -improper if it is only correlated with the involution $\mathbf{x}^\alpha$, so that all complementary covariances except $\mathbf{C}_{\mathbf{xx}^\alpha}$ vanish; $\mathbf{x}$ is maximally improper (rectilinear) if it is maximally correlated with its involutions, so that the impropriety coefficients of the elements of $\mathbf{x}$ are all unit.

In order to explore the statistical relationship between two quaternion random vectors, the concept of cross properness has been defined [21]. Two quaternion random vectors, $\mathbf{x}$ and $\mathbf{y}$, are cross $\mathbb{H}$ -proper if the three complementary cross-covariance matrices vanish, that is, $\mathbf{C}_{\mathbf{xy}^i} = \mathbf{C}_{\mathbf{xy}^j} = \mathbf{C}_{\mathbf{xy}^\kappa} = \mathbf{0}$; $\mathbf{x}$ and $\mathbf{y}$ are cross $\mathbb{R}^\alpha$ -proper if only $\mathbf{C}_{\mathbf{xy}^\alpha}$ among the three complementary cross-covariances vanishes; $\mathbf{x}$ and $\mathbf{y}$ are cross $\mathbb{C}^\alpha$ -proper if all complementary cross-covariances except $\mathbf{C}_{\mathbf{xy}^\alpha}$ vanishes.

Note that when $\mathbf{x}$ is zero-mean, the expressions of the covariance and complementary covariance matrices are simplified, e.g., $\mathbf{C}_{\mathbf{xx}} \triangleq E\{\mathbf{xx}^H\}$, $\mathbf{C}_{\mathbf{xx}^\alpha} \triangleq E\{\mathbf{xx}^{\alpha H}\}$. In Chapters 3-7, zero-mean quaternion random vectors are considered for simplicity, but this does not affect generality of the analysis results. In

Chapter 8, nonzero-mean quaternion random vectors are studied to analyse biased estimators.

2.3.2 Quaternion Multivariate Gaussian Distribution

The augmented second-order statistics pave the way for the theory of the Gaussian distribution of quaternion random variables. Using the augmented vector

$$\mathbf{x}^a \triangleq [\mathbf{x}^T, \mathbf{x}^{iT}, \mathbf{x}^{jT}, \mathbf{x}^{\kappa T}]^T \quad (2.14)$$

the Gaussian distribution of an N -dimensional improper quaternion random vector, $\mathbf{x}$, can be characterised by $\mathbf{x}^a \sim \mathcal{N}(\boldsymbol{\mu}^a, \mathbf{C}_{\mathbf{x}^a \mathbf{x}^a})$, where $\boldsymbol{\mu}^a \triangleq E\{\mathbf{x}^a\}$, with a probability density function given by [22]

$$f_G(\mathbf{x}^a | \boldsymbol{\mu}^a, \mathbf{C}^a) = \frac{4^N \exp\left\{-\frac{1}{2}(\mathbf{x}^a - \boldsymbol{\mu}^a)^H \mathbf{C}_{\mathbf{x}^a \mathbf{x}^a}^{-1}(\mathbf{x}^a - \boldsymbol{\mu}^a)\right\}}{\pi^{2N} \sqrt{|\mathbf{C}_{\mathbf{x}^a \mathbf{x}^a}|}} \quad (2.15)$$

For a $\mathbb{C}^\eta$ -improper quaternion random vector, the generic form of the quaternion multivariate Gaussian distribution (QMGD) in (2.15) simplifies to $\mathbf{x}^b \sim \mathcal{N}(\boldsymbol{\mu}^b, \mathbf{C}_{\mathbf{x}^b \mathbf{x}^b})$, where $\mathbf{x}^b \triangleq [\mathbf{x}^T, \mathbf{x}^{\eta T}]^T$, $\eta \in \{\imath, j, \kappa\}$, $\boldsymbol{\mu}^b \triangleq E\{\mathbf{x}^b\}$, with a probability density function given by

$$f_G(\mathbf{x}^b | \boldsymbol{\mu}^b, \mathbf{C}^b) = \frac{4^N \exp\left\{-(\mathbf{x}^b - \boldsymbol{\mu}^b)^H \mathbf{C}_{\mathbf{x}^b \mathbf{x}^b}^{-1}(\mathbf{x}^b - \boldsymbol{\mu}^b)\right\}}{\pi^{2N} |\mathbf{C}_{\mathbf{x}^b \mathbf{x}^b}|} \quad (2.16)$$

For an $\mathbb{H}$ -proper quaternion random vector, the QMGD in (2.15) simplifies to $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C}_{\mathbf{xx}})$, where $\boldsymbol{\mu} \triangleq E\{\mathbf{x}\}$, with a probability density function given by

$$f_G(\mathbf{x} | \boldsymbol{\mu}, \mathbf{C}) = \frac{4^N \exp\left\{-2(\mathbf{x} - \boldsymbol{\mu})^H \mathbf{C}_{\mathbf{xx}}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right\}}{\pi^{2N} |\mathbf{C}_{\mathbf{xx}}|^2} \quad (2.17)$$

2.4 Widely Linear Quaternion Estimation

Consider the estimation of a quaternion-valued variable, $y \in \mathbb{H}$, conditional to a set of quaternion-valued observations, $\mathbf{x} \in \mathbb{H}^{L \times 1}$. Quaternion estimation techniques aim to find the optimal estimate $\hat{y} = E\{y|\mathbf{x}\}$. The estimation model $\hat{y} = f(\mathbf{x})$, which represents the prior knowledge about the relation between y and $\mathbf{x}$, is crucial for estimation performance. Traditionally, the strictly linear (SL) model is employed as [25]

$$\hat{y} = \hat{\mathbf{h}}^H \mathbf{x} \quad (2.18)$$

where $\hat{\mathbf{h}} \in \mathbb{H}^{L \times 1}$ is the weight vector. This model achieves optimal estimation for quaternion signals driven by SL systems but is suboptimal for general quaternion signals because the lack of the knowledge about complementary covariances. To exploit the complete second-order statistical information, involutions of $\mathbf{x}$ can be incorporated to produce a widely linear (WL) model given by [29]

$$\hat{y} = \hat{\mathbf{h}}^H \mathbf{x} + \hat{\mathbf{g}}^H \mathbf{x}^i + \hat{\mathbf{u}}^H \mathbf{x}^j + \hat{\mathbf{v}}^H \mathbf{x}^\kappa = \hat{\mathbf{w}}^{aH} \mathbf{x}^a \quad (2.19)$$

where $\mathbf{x}^a \triangleq [\mathbf{x}^T, \mathbf{x}^{iT}, \mathbf{x}^{jT}, \mathbf{x}^{\kappa T}]^T$ is the augmented vector of the $\mathbf{x}$, and $\hat{\mathbf{h}}, \hat{\mathbf{g}}, \hat{\mathbf{u}}, \hat{\mathbf{v}} \in \mathbb{H}^{L \times 1}$ are the estimated weight vectors which can be compactly represented in the augmented form, $\hat{\mathbf{w}}^a \triangleq [\hat{\mathbf{h}}^T, \hat{\mathbf{g}}^T, \hat{\mathbf{u}}^T, \hat{\mathbf{v}}^T]^T$.

For semi-widely linear (SWL) systems, the WL model reduces to a SWL model given by [51, 52]

$$\hat{y} = \hat{\mathbf{w}}^{bH} \mathbf{x}^b \quad (2.20)$$

where $\mathbf{x}^b \triangleq [\mathbf{x}^T, \mathbf{x}^{\eta T}]^T$, $\eta \in \{i, j, \kappa\}$, is the semi-augmented vector, and $\hat{\mathbf{w}}^b \in \mathbb{H}^{2L \times 1}$ the weight vector.

Similar to the duality between complex filters and bivariate real filters [7, 53], quadrivariate real filters are isomorphic to widely linear quaternion filters, but are totally different from strictly and semi-widely linear quaternion filters.

2.4.1 Quaternion Least Mean Square Filters

A widely linear quaternion least mean square (WL-QLMS) filter for quaternion MMSE estimation has been introduced by implementing the WL model in an iterative fashion [24, 29]. In each iteration, the weights in the WL model are updated in the opposite direction of the gradient of the squared error $J_n = |y_n - \hat{y}_n|^2$, where $\hat{y}_n = \hat{\mathbf{w}}_n^{aH} \mathbf{x}_n^a$. The gradient of J_n with respect to $\hat{\mathbf{w}}_n^a$ is calculated through the generalised $\mathbb{H}\mathbb{R}$ calculus, and is given by

$$\nabla_{\mathbf{w}_n^{a*}} J_n = \frac{\partial |y_n - \hat{\mathbf{w}}_n^{aH} \mathbf{x}_n^a|^2}{\partial \hat{\mathbf{w}}_n^{a*}} = -0.5 \mathbf{x}_n^a (y_n - \hat{y}_n)^*$$

Incorporating a step size, μ , and absorbing the constant, 0.5, the weight is updated as

$$\hat{\mathbf{w}}_{n+1}^a = \hat{\mathbf{w}}_n^a - \mu \nabla_{\hat{\mathbf{w}}_n^{a*}} J_n = \hat{\mathbf{w}}_n^a + \mu \mathbf{x}_n^a (y_n - \hat{y}_n)^* \quad (2.21)$$

μ is fixed and real-valued in most applications and also in most part of this thesis, except in Section 7.4.4, where μ is time-varying, and in Section 6.3, where μ is quaternion-valued.

If $\mathbf{x}_n$ and y_n are cross $\mathbb{C}^\eta$ -improper, the WL-QLMS reduces to the semi-widely linear quaternion least mean square (SWL-QLMS) based on the SWL model in (2.20), that is [54]

$$\begin{aligned} \hat{y}_n &= \hat{\mathbf{w}}_n^{bH} \mathbf{x}_n^b \\ \hat{\mathbf{w}}_{n+1}^b &= \hat{\mathbf{w}}_n^b + \mu \mathbf{x}_n^b (y_n - \hat{y}_n)^* \end{aligned}$$

If $\mathbf{x}_n$ and y_n are cross $\mathbb{H}$ -proper, the WL-QLMS reduces to strictly linear quaternion least mean square (SL-QLMS) based on the SL model in (2.18), that is

$$\begin{aligned} \hat{y}_n &= \hat{\mathbf{w}}_n^H \mathbf{x}_n \\ \hat{\mathbf{w}}_{n+1} &= \hat{\mathbf{w}}_n + \mu \mathbf{x}_n (y_n - \hat{y}_n)^* \end{aligned}$$

2.4.2 Quaternion Kalman Filters

The widely linear quaternion state estimation based on the model

$$\begin{aligned}\mathbf{x}_n^a &= \mathbf{F}_{n-1}^a \mathbf{x}_{n-1}^a + \mathbf{u}_{n-1}^a \\ \mathbf{z}_n^a &= \mathbf{H}_n^a \mathbf{x}_n^a + \mathbf{v}_n^a\end{aligned}\tag{2.22}$$

has been proposed and has shown its optimality for processing widely linear quaternion dynamical systems with general (circular or noncircular) noise [39]. In (2.22), $\mathbf{x}_n^a$, $\mathbf{z}_n^a$, $\mathbf{u}_{n-1}^a$, and $\mathbf{v}_n^a$ are the augmented forms of the state variable vector $\mathbf{x}_n$, the observation vector $\mathbf{z}_n$, the state noise vector $\mathbf{u}_{n-1}$, and the observation noise vector $\mathbf{v}_n$, as in (2.14), while the state transition and observation matrices $\mathbf{F}_n^a$ and $\mathbf{H}_n^a$ have the structures

$$\mathbf{F}_{n-1}^a = \begin{bmatrix} \mathbf{F}_{n-1} & \mathbf{F}_{i,n-1} & \mathbf{F}_{j,n-1} & \mathbf{F}_{\kappa,n-1} \\ \mathbf{F}_{i,n-1}^i & \mathbf{F}_{n-1}^i & \mathbf{F}_{\kappa,n-1}^i & \mathbf{F}_{j,n-1}^i \\ \mathbf{F}_{j,n-1}^j & \mathbf{F}_{\kappa,n-1}^j & \mathbf{F}_{n-1}^j & \mathbf{F}_{i,n-1}^j \\ \mathbf{F}_{\kappa,n-1}^\kappa & \mathbf{F}_{j,n-1}^\kappa & \mathbf{F}_{i,n-1}^\kappa & \mathbf{F}_{n-1}^\kappa \end{bmatrix}$$

$$\mathbf{H}_n^a = \begin{bmatrix} \mathbf{H}_n & \mathbf{H}_{i,n} & \mathbf{H}_{j,n} & \mathbf{H}_{\kappa,n} \\ \mathbf{H}_{i,n}^i & \mathbf{H}_n^i & \mathbf{H}_{\kappa,n}^i & \mathbf{H}_{j,n}^i \\ \mathbf{H}_{j,n}^j & \mathbf{H}_{\kappa,n}^j & \mathbf{H}_n^j & \mathbf{H}_{i,n}^j \\ \mathbf{H}_{\kappa,n}^\kappa & \mathbf{H}_{j,n}^\kappa & \mathbf{H}_{i,n}^\kappa & \mathbf{H}_n^\kappa \end{bmatrix}$$

where $\mathbf{F}_{n-1}$, $\mathbf{F}_{i,n-1}$, $\mathbf{F}_{j,n-1}$, $\mathbf{F}_{\kappa,n-1}$, $\mathbf{H}_n$, $\mathbf{H}_{i,n}$, $\mathbf{H}_{j,n}$, $\mathbf{H}_{\kappa,n}$ are quaternion matrices.

The widely linear quaternion Kalman filtering (WL-QKF) algorithm is given by Algorithm 2.1, where $\mathbf{Q}_{n-1}^a$ and $\mathbf{R}_n^a$ are the augmented covariance matrices of $\mathbf{u}_{n-1}^a$ and $\mathbf{v}_n^a$, $\hat{\mathbf{x}}_{n|n-1}^a$ and $\hat{\mathbf{x}}_{n|n}^a$ the *a priori* and *a posteriori* estimates of $\mathbf{x}_n^a$, and $\mathbf{P}_{n|n-1}^a$ and $\mathbf{P}_{n|n}^a$ the augmented covariance matrices of the *a priori* and *a posteriori* state estimation error vectors.

Under the condition of $\mathbb{C}^\eta$ -improperness, the WL state space model in (2.22)

Algorithm 2.1: Widely linear quaternion Kalman filter.

Initialisation: $\hat{\mathbf{x}}_{0|0}^a = E\{\mathbf{x}_0^a\}$, $\mathbf{P}_{0|0} = E\left\{(\mathbf{x}_0^a - E\{\mathbf{x}_0^a\})(\mathbf{x}_0^a - E\{\mathbf{x}_0^a\})^H\right\}$.

To estimate $\mathbf{x}_n$ from the observation $\mathbf{z}_n$, the previous estimate $\hat{\mathbf{x}}_{n-1|n-1}$, and the previous augmented covariance matrix $\mathbf{P}_{n-1|n-1}^a$,

1. predict

$$\begin{aligned}\hat{\mathbf{x}}_{n|n-1}^a &= \mathbf{F}_{n-1}^a \hat{\mathbf{x}}_{n-1|n-1}^a \\ \mathbf{P}_{n|n-1}^a &= \mathbf{F}_{n-1}^a \mathbf{P}_{n-1|n-1}^a \mathbf{F}_{n-1}^{aH} + \mathbf{Q}_{n-1}^a\end{aligned}$$

2. update

$$\begin{aligned}\mathbf{S}_k^a &= \mathbf{H}_n^a \mathbf{P}_{n|n-1}^a \mathbf{H}_n^{aH} + \mathbf{R}_n^a \\ \mathbf{K}_n^a &= \mathbf{P}_{n|n-1}^a \mathbf{H}_n^{aH} \mathbf{S}_n^{a^{-1}} \\ \hat{\mathbf{x}}_{n|n} &= \hat{\mathbf{x}}_{n|n-1} + \mathbf{K}_n^a (\mathbf{z}_n^a - \mathbf{H}_n^a \hat{\mathbf{x}}_{n|n-1}^a) \\ \mathbf{P}_{n|n}^a &= (\mathbf{I} - \mathbf{K}_n^a \mathbf{H}_n^a) \mathbf{P}_{n|n-1}^a\end{aligned}$$

reduces to the SWL model given by [55]

$$\begin{aligned}\mathbf{x}_n^b &= \mathbf{F}_{n-1}^b \mathbf{x}_{n-1}^b + \mathbf{u}_{n-1}^b \\ \mathbf{z}_n^b &= \mathbf{H}_n^b \mathbf{x}_n^b + \mathbf{v}_n^b\end{aligned}\tag{2.23}$$

where the state transition and observation matrices $\mathbf{F}_n^b$ and $\mathbf{H}_n^b$ have the following structure

$$\mathbf{F}_{n-1}^b = \begin{bmatrix} \mathbf{F}_{n-1} & \mathbf{F}_{\eta,n-1} \\ \mathbf{F}_{\eta,n-1}^\eta & \mathbf{F}_{n-1}^\eta \end{bmatrix}, \quad \mathbf{H}_n^b = \begin{bmatrix} \mathbf{H}_n & \mathbf{H}_{\eta,n} \\ \mathbf{H}_{\eta,n}^\eta & \mathbf{H}_n^\eta \end{bmatrix}$$

The state estimation based on the SWL model in (2.23) is therefore optimal when the true system model is SWL with $\mathbb{C}^\eta$ -improper $\mathbf{u}_{n-1}$ and $\mathbf{v}_n$. Correspondingly, the SWL-QKF can be deduced from the WL-QKF by replacing the augmented forms of matrices and vectors with their semi-augmented forms, e.g., $\mathbf{x}_n^a$ with $\mathbf{x}_n^b$.

Under the condition of $\mathbb{H}$ -properness, the WL state space model in (2.22) reduces to the SL model given by

$$\begin{aligned}\mathbf{x}_n &= \mathbf{F}_{n-1} \mathbf{x}_{n-1} + \mathbf{u}_{n-1} \\ \mathbf{z}_n &= \mathbf{H}_n \mathbf{x}_n + \mathbf{v}_n\end{aligned}\tag{2.24}$$

The state estimation based on the SL model in (2.24) is optimal only when the true system model is SL while $\mathbf{u}_{n-1}$ and $\mathbf{v}_n$ are $\mathbb{H}$ -proper. Under this condition, the SL-QKF can be deduced from the WL-QKF by replacing the augmented forms of matrices and vectors with their original forms, e.g., $\mathbf{x}_n^a$ with $\mathbf{x}_n$.

Part I

Simultaneous Diagonalisation of the Covariance and Complementary Covariance Matrices

Chapter 3

Quaternion Approximate Uncorrelating Transform

Overview

Recent developments in quaternion-valued widely linear processing have established that the exploitation of complete second-order statistics requires consideration of both the standard covariance and three complementary covariance matrices. Although such matrices have a tremendous amount of structure and their decomposition is a powerful tool in a variety of applications, the noncommutative nature of the quaternion product has been prohibitive to the development of quaternion uncorrelating transforms. To this end, this chapter introduces a novel technique for simultaneous decomposition of the covariance and three complementary covariance matrices of noncircular quaternion variables, whereby the quaternion approximate uncorrelating transform is proposed.

3.1 Introduction

Diagonalisable matrices play a fundamental role in engineering applications, whereby the related uncorrelating transforms greatly simplify the analysis of complex problems, both in terms of enabling a solution and providing a methodological framework (e.g. performance bounds). Computational advantages associated with matrix diagonalisation include a reduction in the size of the parameter space from N^2 to N , while the statistical advantages include the possibility to match source properties (e.g., orthogonality) in blind source separation. In addition, joint dia-

gonalisation of matrices in some blind source separation applications has a close link with the canonical decomposition problem for tensors [56, 57]. The real linear algebra is a maturing area, however, covariance matrices in widely linear signal processing in the complex and quaternion division algebras and their structures have recently received significant attention.

In the context of complex-valued signal processing, for a complex random vector, both its covariance and pseudo-covariance matrices are necessary to capture complete second-order statistical information [7]. This is achieved by simultaneous diagonalisation of the Hermitian covariance matrix and the symmetric pseudo-covariance matrix, which is performed respectively via the eigenvalue decomposition (eigendecomposition) and the Takagi factorisation. While such decorrelations exist in their general forms in mathematics [58], De Lathauwer and De Moor [59] and Eriksson and Koivunen [60] have given them a practical value through an important engineering contribution called the strong uncorrelating transform (SUT), together with its extension, the generalised uncorrelating transform (GUT) [61]. The SUT and GUT not only decorrelate the complex variables in the random vector, but also eliminate the correlation between the real and imaginary components within each complex variable. Recent contributions in this case include a computationally efficient approximate uncorrelating transform (AUT) [62]. In some application scenarios, the multiple bivariate sources are required to be mutually decorrelated, while at the same time, the correlation between the constitutive components of each individual bivariate source must be preserved. To this end, the correlation preserving transform (CPT) has been proposed, preserving the correlation between the real and imaginary components within each complex variable in the random vector [63].

Whilst real-valued and complex-valued matrices are well understood [58], the quaternion-valued matrix algebra is still under development [12, 64]. The recently introduced theory of quaternion matrix derivatives provides a systematic framework for the calculation of derivatives of quaternion matrix functions with respect

to quaternion matrix variables [32]. Advances in structural quaternion matrix decompositions include tools to diagonalise quaternion matrices [65, 66], and in the context of quaternion widely linear processing, it was shown that for a quaternion random vector $\mathbf{x}$, the diagonalisation of the Hermitian covariance matrix can be performed straightforwardly using the quaternion eigendecomposition, whereas the diagonalisation of the complementary covariance matrices can be computed via the quaternion SVD [64]. To jointly diagonalise the covariance and complementary covariance matrices, the SUT was extended from the complex domain to the quaternion domain in [67]. However, the following issues remain unsolved.

- There is no closed-form solution¹ to perform a simultaneous diagonalisation of the covariance matrix and the three complementary covariance matrices.
- Given the dimensionality of the problem, a simple approximate diagonalising transform which would apply to both the covariance and pseudo-covariance matrices would be beneficial.

This chapter sets out to propose solutions to these open problems, and supports the analysis with illustrative examples.

3.2 Simultaneous Diagonalisation of the Covariance Matrix and One Complementary Covariance Matrix

The eigendecomposition of the covariance matrix $\mathbf{C}_{\mathbf{xx}}$ is given by $\mathbf{C}_{\mathbf{xx}} = \mathbf{Q}\mathbf{\Lambda}_{\mathbf{xx}}\mathbf{Q}^H$, where $\mathbf{Q}$ is a quaternion unitary matrix, and $\mathbf{\Lambda}_{\mathbf{xx}}$ is a real-valued non-negative diagonal matrix with the eigenvalues of $\mathbf{C}_{\mathbf{xx}}$ on the diagonal [12]. It has been proven that the α -complementary covariance matrix, $\mathbf{C}_{\mathbf{xx}^\alpha}$, $\alpha \in \{\iota, j, \kappa\}$, can be factorised as a special quaternion SVD, $\mathbf{C}_{\mathbf{xx}^\alpha} = \mathbf{Q}_\alpha\mathbf{\Lambda}_{\mathbf{xx}^\alpha}\mathbf{Q}_\alpha^{\alpha H}$, where $\mathbf{Q}_\alpha$ is a quaternion

¹Iterative solutions were proposed in the context of quaternion ICA [50].

Algorithm 3.1: Quaternion Uncorrelating Transform (QUT).

1. Compute the eigendecomposition of $\mathbf{C}_{\mathbf{xx}}$, as $\mathbf{C}_{\mathbf{xx}} = \mathbf{Q}\mathbf{\Lambda}_{\mathbf{xx}}\mathbf{Q}^H$.
 2. Compute the whitening matrix, $\mathbf{W} = \mathbf{Q}\mathbf{\Lambda}_{\mathbf{xx}}^{-\frac{1}{2}}\mathbf{Q}^H$.
 3. For an $\alpha \in \{\iota, j, \kappa\}$, calculate the α -complementary covariance matrix of the whitened data, $\mathbf{s} = \mathbf{W}\mathbf{x}$, as $\mathbf{C}_{\mathbf{ss}^\alpha} = E\{\mathbf{s}\mathbf{s}^{\alpha H}\}$.
 4. Compute the SVD of $\mathbf{C}_{\mathbf{ss}^\alpha}$, as $\mathbf{C}_{\mathbf{ss}^\alpha} = \mathbf{Q}_\alpha\mathbf{\Lambda}_{\mathbf{ss}^\alpha}\mathbf{Q}_\alpha^{\alpha H}$.
 5. Compute the QUT matrix, $\mathbf{V} = \mathbf{Q}_\alpha^H\mathbf{W}$. Transform $\mathbf{x}$ to $\mathbf{y} = \mathbf{V}\mathbf{x}$, then, $\mathbf{C}_{\mathbf{yy}} = \mathbf{I}$, $\mathbf{C}_{\mathbf{yy}^\alpha} = \mathbf{\Lambda}_{\mathbf{ss}^\alpha}$.
-

unitary matrix and $\mathbf{\Lambda}_{\mathbf{xx}^\alpha}$ is a real-valued non-negative diagonal matrix with the singular values of $\mathbf{C}_{\mathbf{xx}^\alpha}$ on the diagonal [64]. This enables the diagonalisation of quaternion complementary covariance matrices and degenerates into the Takagi factorisation of complex symmetric matrices if the quaternion vector $\mathbf{x}$ has two vanishing imaginary parts [68].

In order to simultaneously diagonalise the covariance matrix and a complementary covariance matrix of a zero-mean quaternion random vector $\mathbf{x}$, [67] has proposed the quaternion uncorrelating transform (QUT), which is summarised as Algorithm 3.1.

Recall that for a $\mathbb{C}^\alpha$ -improper vector $\mathbf{x}$, only the α -complementary covariance is non-vanishing while the other two complementary covariances vanish. Thus, the QUT of $\mathbb{C}^\alpha$ -improper signals, for example, for the case $\alpha = \kappa$, satisfies

$$\mathbf{C}_{\mathbf{yy}} = \mathbf{I} \quad \mathbf{C}_{\mathbf{yy}^\iota} = \mathbf{\Lambda}_{\mathbf{yy}^\iota} \quad \mathbf{C}_{\mathbf{yy}^j} = \mathbf{\Lambda}_{\mathbf{yy}^j} \quad \mathbf{C}_{\mathbf{yy}^\kappa} = \mathbf{\Lambda}_{\mathbf{yy}^\kappa}$$

where $\mathbf{\Lambda}_{\mathbf{yy}^\kappa}$ is a diagonal matrix and $\mathbf{\Lambda}_{\mathbf{yy}^\iota} = \mathbf{\Lambda}_{\mathbf{yy}^j} = \mathbf{0}$. In other words, for a $\mathbb{C}^\alpha$ -improper quaternion vector, the QUT can be regarded as the quaternion version of the SUT [59, 60]. For a general improper quaternion vector, however, the remaining two complementary covariance matrices are still not diagonalised by the

QUT. In fact, so far there is no closed-form solution to joint diagonalisation of four covariance matrices for general improper quaternion vectors. To circumvent this problem, the next section will introduce an approximate way to simultaneously diagonalise all four covariance matrices, a practically useful tool in most applications of quaternions.

3.3 Simultaneous Approximate Diagonalisation of the Covariance Matrix and Three Complementary Covariance Matrices

Following the approach in Section 3.2, the quaternion analogue of the AUT for complex-valued data [62] shall be derived. For a zero-mean quaternion random vector $\mathbf{x}$, consider finding a unitary transform $\mathbf{y} = \Phi \mathbf{x}$ which simultaneously diagonalises the covariance matrix, $\mathbf{C}_{\mathbf{y}\mathbf{y}} = E \{ \mathbf{y}\mathbf{y}^H \} = \Phi \mathbf{C}_{\mathbf{x}\mathbf{x}} \Phi^H$, and the three complementary covariance matrices, $\mathbf{C}_{\mathbf{y}\mathbf{y}^\alpha} = E \{ \mathbf{y}\mathbf{y}^{\alpha H} \} = \Phi \mathbf{C}_{\mathbf{x}\mathbf{x}^\alpha} \Phi^{\alpha H}$, $\alpha \in \{\iota, j, \kappa\}$. For N samples of $\mathbf{x}$, $N > L$, the covariance matrix, and the α -complementary covariance matrix can be estimated as

$$\mathbf{C}_{\mathbf{x}\mathbf{x}} = \frac{1}{N} \sum_{n_1=1}^N \mathbf{x}(n_1) \mathbf{x}^H(n_1) \quad (3.1)$$

$$\mathbf{C}_{\mathbf{x}\mathbf{x}^\alpha} = \frac{1}{N} \sum_{n_1=1}^N \mathbf{x}(n_1) \mathbf{x}^{\alpha H}(n_1) \quad (3.2)$$

Their quadratic forms, $\mathbf{C}_{\mathbf{x}\mathbf{x}} \mathbf{C}_{\mathbf{x}\mathbf{x}}^H$ and $\mathbf{C}_{\mathbf{x}\mathbf{x}^\alpha} \mathbf{C}_{\mathbf{x}\mathbf{x}^\alpha}^H$, are calculated as

$$\begin{aligned} \mathbf{C}_{\mathbf{x}\mathbf{x}} \mathbf{C}_{\mathbf{x}\mathbf{x}}^H &= \left[\frac{1}{N} \sum_{n_1=1}^N \mathbf{x}(n_1) \mathbf{x}^H(n_1) \right] \left[\frac{1}{N} \sum_{n_2=1}^N \mathbf{x}(n_2) \mathbf{x}^H(n_2) \right] \\ &= \frac{1}{N^2} \sum_{n_1=1}^N \sum_{n_2=1}^N \mathbf{x}(n_1) \left[\sum_{l=1}^L x_l^*(n_1) x_l(n_2) \right] \mathbf{x}^H(n_2) \end{aligned} \quad (3.3)$$

$$\begin{aligned}
\mathbf{C}_{\mathbf{xx}^\alpha} \mathbf{C}_{\mathbf{xx}^\alpha}^H &= \left[\frac{1}{N} \sum_{n_1=1}^N \mathbf{x}(n_1) \mathbf{x}^{\alpha H}(n_1) \right] \left[\frac{1}{N} \sum_{n_2=1}^N \mathbf{x}^\alpha(n_2) \mathbf{x}^H(n_2) \right] \\
&= \frac{1}{N^2} \sum_{n_1=1}^N \sum_{n_2=1}^N \mathbf{x}(n_1) \left[\sum_{l=1}^L x_l^*(n_1) x_l(n_2) \right]^\alpha \mathbf{x}^H(n_2) \quad (3.4)
\end{aligned}$$

If

$$\sum_{l=1}^L x_l^*(n_1) x_l(n_2) = \left[\sum_{l=1}^L x_l^*(n_1) x_l(n_2) \right]^\alpha \quad (3.5)$$

then, $\mathbf{C}_{\mathbf{xx}} \mathbf{C}_{\mathbf{xx}}^H = \mathbf{C}_{\mathbf{xx}^\alpha} \mathbf{C}_{\mathbf{xx}^\alpha}^H$, and hence $\mathbf{C}_{\mathbf{yy}} \mathbf{C}_{\mathbf{yy}}^H = \mathbf{C}_{\mathbf{yy}^\alpha} \mathbf{C}_{\mathbf{yy}^\alpha}^H$, which implies that if $\mathbf{C}_{\mathbf{yy}}$ is diagonal, so too is $\mathbf{C}_{\mathbf{yy}^\alpha}$, and *vice versa*. Furthermore, if

$$\sum_{l=1}^L x_l^*(n_1) x_l(n_2) \in \mathbb{R} \quad (3.6)$$

then, (3.5) holds for any $\alpha \in \{\iota, j, \kappa\}$, and thus the simultaneous diagonalisation applies to the covariance matrix and all of the three complementary covariance matrices. As shown in Section 3.3.1, (3.6) holds approximately if the four components of each quaternion variable in $\mathbf{x}$ are highly correlated. Therefore, the following *quaternion approximate uncorrelating transform* (QAUT) can be concluded.

QAUT: *Each of $\mathbf{C}_{\mathbf{xx}}$, $\mathbf{C}_{\mathbf{xx}^\iota}$, $\mathbf{C}_{\mathbf{xx}^j}$ and $\mathbf{C}_{\mathbf{xx}^\kappa}$ can be diagonalised by some unitary transform. A unitary transform, Φ , that diagonalises one of the four matrices can approximately diagonalise the other three, owing to their structural similarity. As $E\{\mathbf{xx}^T\} = \frac{1}{2}(\mathbf{C}_{\mathbf{xx}^\iota} + \mathbf{C}_{\mathbf{xx}^j} + \mathbf{C}_{\mathbf{xx}^\kappa} - \mathbf{C}_{\mathbf{xx}})$ [22], the pseudo-covariance matrix, $E\{\mathbf{xx}^T\}$, is also approximately diagonalised in this way.*

The QAUT agrees with the intuition that similar statistical information is contained in the covariance and complementary covariance matrices of a quaternion random vector, in which the variables consist of highly correlated real and imaginary components, so that the diagonalisation of one of the four matrices could also diagonalise the other three. For example, the transformation matrix Φ

can take the form of the eigenvector matrix of $\mathbf{C}_{\mathbf{xx}}$, that is

$$\begin{aligned}
\mathbf{C}_{\mathbf{xx}} &= \mathbf{Q}\mathbf{\Lambda}_{\mathbf{xx}}\mathbf{Q}^H \\
\mathbf{C}_{\mathbf{xx}^i} &\approx \mathbf{Q}\mathbf{\Lambda}_{\mathbf{xx}^i}\mathbf{Q}^{iH} \\
\mathbf{C}_{\mathbf{xx}^j} &\approx \mathbf{Q}\mathbf{\Lambda}_{\mathbf{xx}^j}\mathbf{Q}^{jH} \\
\mathbf{C}_{\mathbf{xx}^\kappa} &\approx \mathbf{Q}\mathbf{\Lambda}_{\mathbf{xx}^\kappa}\mathbf{Q}^{\kappa H}
\end{aligned} \tag{3.7}$$

3.3.1 Principle of QAUT

The QAUT is based on the assumption in (3.6). To examine the validity of this assumption for highly improper quaternion signals, consider the case where the four components of the quaternion variable, x_l , are highly correlated. x_l fits the Cayley-Dickson construction

$$x_l = x_a + x_b\imath + x_cj + x_d\kappa = c_1 + c_2j$$

where $c_1 = x_a + x_b\imath$ and $c_2 = x_c + x_d\imath$ are complex variables defined in the plane spanned by $\{1, \imath\}$. This gives

$$\begin{aligned}
x_l^*(n_1) x_l(n_2) &= c_1^*(n_1) c_1(n_2) + c_2(n_1) c_2^*(n_2) + \\
&\quad [c_1^*(n_1) c_2(n_2) - c_2(n_1) c_1^*(n_2)] j
\end{aligned} \tag{3.8}$$

Applying the statistical expectation operator to (3.8) yields

$$\begin{aligned}
E\{x_l^*(n_1) x_l(n_2)\} &= E\{c_1^*(n_1) c_1(n_2)\} + E\{c_2(n_1) c_2^*(n_2)\} \\
&\quad + E\{c_1^*(n_1) c_2(n_2) - c_2(n_1) c_1^*(n_2)\} j
\end{aligned} \tag{3.9}$$

where

$$\begin{aligned}
E\{c_1^*(n_1) c_1(n_2)\} &= E\{x_a(n_1) x_a(n_2) + x_b(n_1) x_b(n_2)\} + \\
&\quad E\{x_a(n_1) x_b(n_2) - x_a(n_2) x_b(n_1)\} \\
&\approx E\{x_a(n_1) x_a(n_2) + x_b(n_1) x_b(n_2)\} \\
&\in \mathbb{R}
\end{aligned}$$

in which the approximation is due to the high correlation between x_a and x_b . Likewise, $E\{c_2(n_1) c_2^*(n_2)\}$ is also approximately real-valued. Moreover, as c_1 and c_2 are highly correlated, they can be expressed as $c_2(t) \approx k c_1(t)$, where k is complex-valued, to give

$$\begin{aligned}
&E\{c_1^*(n_1) c_2(n_2) - c_2(n_1) c_1^*(n_2)\} \\
&\approx E\{c_1^*(n_1) k c_1(n_2) - k c_1(n_1) c_1^*(n_2)\} \\
&= 2k E\{\Im[c_1^*(n_1) c_1(n_2)]\} \\
&\approx 0
\end{aligned}$$

Therefore, it follows from (3.8) that

$$E\{x_l^*(n_1) x_l(n_2)\} \in \mathbb{R} \quad (3.10)$$

approximately holds. Given the assumption

$$E\{x_l^*(n_1) x_l(n_2)\} \approx \frac{1}{L} \sum_{l=1}^L x_l^*(n_1) x_l(n_2)$$

(3.6) holds approximately. The above analysis implies that the simultaneous diagonalisation holds exactly when four components of each quaternion variable are perfectly correlated (maximally improper). For general quaternion signals, however, these components are not perfectly correlated, causing the approximation

error influenced by the correlation coefficients defined by

$$\begin{aligned}\rho_{ab} &\triangleq \frac{\text{COV}[x_a, x_b]}{\sigma_{x_a} \sigma_{x_b}}, & \rho_{ac} &\triangleq \frac{\text{COV}[x_a, x_c]}{\sigma_{x_a} \sigma_{x_c}}, & \rho_{ad} &\triangleq \frac{\text{COV}[x_a, x_d]}{\sigma_{x_a} \sigma_{x_d}}, \\ \rho_{bc} &\triangleq \frac{\text{COV}[x_b, x_c]}{\sigma_{x_b} \sigma_{x_c}}, & \rho_{bd} &\triangleq \frac{\text{COV}[x_b, x_d]}{\sigma_{x_b} \sigma_{x_d}}, & \rho_{cd} &\triangleq \frac{\text{COV}[x_c, x_d]}{\sigma_{x_c} \sigma_{x_d}}\end{aligned}$$

within the range $[-1, 1]$, where “cov” denotes the covariance and σ the standard deviation. From the simultaneous diagonalisation condition in (3.5), the diagonalisation error of $\mathbf{C}_{\mathbf{xx}^\alpha}$ arises from and is positively correlated with

$$\begin{aligned}x_l^*(n_1) x_l(n_2) - [x_l^*(n_1) x_l(n_2)]^\alpha &= 2\mathfrak{I}_\beta [x_l^*(n_1) x_l(n_2)] \beta + 2\mathfrak{I}_\gamma [x_l^*(n_1) x_l(n_2)] \gamma \\ \text{for distinct, } \alpha, \beta, \gamma &\in \{\iota, j, \kappa\}\end{aligned}\tag{3.11}$$

Note that the imaginary part of $x_l^*(n_1) x_l(n_2)$ is given by

$$\begin{aligned}\mathfrak{I}[x_l^*(n_1) x_l(n_2)] &= [x_a(n_1) x_b(n_2) - x_a(n_2) x_b(n_1) + x_c(n_2) x_d(n_1) - x_c(n_1) x_d(n_2)] \iota \\ &+ [x_a(n_1) x_c(n_2) - x_a(n_2) x_c(n_1) + x_b(n_1) x_d(n_2) - x_b(n_2) x_d(n_1)] j \\ &+ [x_a(n_1) x_d(n_2) - x_a(n_2) x_d(n_1) + x_b(n_2) x_c(n_1) - x_b(n_1) x_c(n_2)] \kappa\end{aligned}\tag{3.12}$$

Similar to the complex AUT proposed in [62], (3.12) can be rewritten as

$$\begin{aligned}\mathfrak{I}[x_l^*(n_1) x_l(n_2)] &= \left(\xi_{ab} \sqrt{1 - \rho_{ab}^2} + \xi_{cd} \sqrt{1 - \rho_{cd}^2}\right) \iota + \left(\xi_{ac} \sqrt{1 - \rho_{ac}^2} + \xi_{bd} \sqrt{1 - \rho_{bd}^2}\right) j \\ &+ \left(\xi_{ad} \sqrt{1 - \rho_{ad}^2} + \xi_{bc} \sqrt{1 - \rho_{bc}^2}\right) \kappa\end{aligned}\tag{3.13}$$

where $\xi_{ab}, \xi_{cd}, \xi_{ac}, \xi_{bd}, \xi_{ad}, \xi_{bc}$ are real-valued coefficients dependent on the statistics of x_l .

It is now obvious that as the components of quaternion data become more correlated, (3.13) decreases and the diagonalisation error hence decreases. Owing to the cumulative nature, the error in assuming (3.6) increases with the data

segment length, L . Since this error leads to errors in all off-diagonal elements of the matrices in (3.3) and (3.4), the simultaneous diagonalisation error increases with L .

Remark 3.1. The simultaneous diagonalisation of the covariance and complementary covariance matrices becomes exact for quaternion variables with special second-order statistical properties, such as $\mathbb{H}$ -proper variables, the variables with fully correlated components, and the variables for which the covariance and complementary covariance matrices have the same singular vectors. This can be verified by exploring the structures of the matrices.

3.4 Simulations

The effectiveness of the proposed QAUT techniques is illustrated via simulations on both synthetic and real-world quaternion signals. In the experiments, the covariance and complementary covariance matrices were estimated based on (3.1) and (3.2).

3.4.1 Performance of QAUT for Synthetic Improper Signals

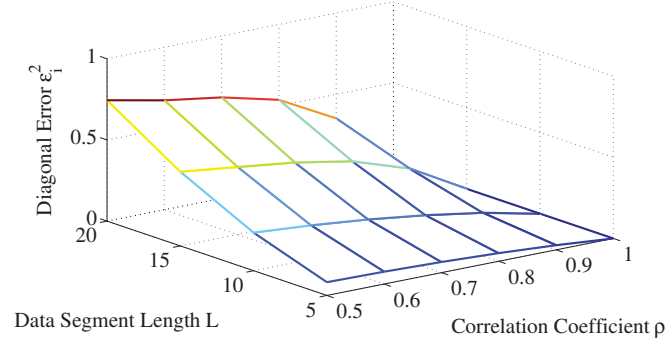
To assess the performance of QAUT, the squared diagonal error of the complementary covariance matrix $\mathbf{C}_{\mathbf{xx}^\alpha}$ was measured by an average power ratio of the off-diagonal, $\lambda_{\alpha,ij}$, versus diagonal, $\lambda_{\alpha,ii}$, elements of the approximately diagonal matrix $\mathbf{\Lambda}_{\mathbf{xx}^\alpha}$, given by

$$\varepsilon_\alpha^2 = \frac{\sum_{\substack{i,j=1 \\ i \neq j}}^L |\lambda_{\alpha,ij}|^2}{(N-1) \sum_{i=1}^L \lambda_{\alpha,ii}^2} \times 100\%$$

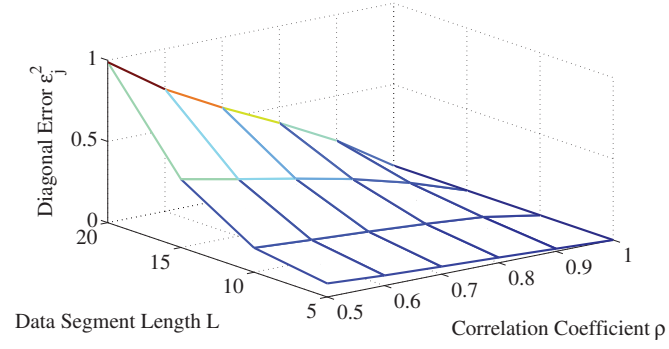
Synthetic quaternion signals with a varying degree of correlation between the real, i , j and k components were considered. For simplicity, the six correlation coefficients of the four components were assumed to be equally distributed in $(0, 1]$ and denoted by ρ , where $\rho = 1$ indicates full correlation and $\rho = 0$ no correlation. A synthetic quaternion data vector with a varying ρ was generated from a quaternion white Gaussian signal. Then, the performance of the QAUT was assessed comprehensively against the degree of correlation present in the data components, $\rho \in [0.5, 1]$, and the length of data segment, $L \in [5, 20]$. Conforming with the analysis, Figure 3.1 shows that the diagonalisation error increased as ρ or L increased. The increase with L reflects that the added data in the segment complicated the data correlation and hence hindered the simultaneous diagonalisation. The performance was excellent, for example, for a moderate correlation degree, $\rho = 0.5$, and a long data segment, $L = 20$, where the squared diagonalisation error was less than 1%, while for a high correlation degree, ρ is close to unit, where the error was negligible.

3.4.2 Performance of QAUT for Real-World Signals

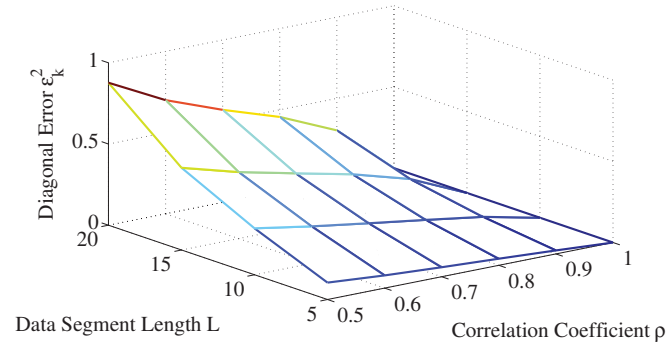
The QAUT was also applied to 4-D real-world electroencephalogram (EEG) signals by combining four adjacent channels of real-valued EEG signals into a quaternion-valued signal. The four channels measuring adjacent brain regions exhibited strong correlation with each other while the channels measuring distant brain regions were relatively weakly correlated [69, 70]. The QAUT was tested on such quaternion EEG signals with low, medium and high correlations between the four dimensions, with the data segment length L varying between 5 and 20. Figure 3.2 illustrates that the diagonalisation error increased with either an increase in L or a decrease in the degree of correlation present in the data. The squared diagonal error was less than 0.2%.



(a) Squared diagonal error of $\mathbf{C}_{\mathbf{x}\mathbf{x}^i}$

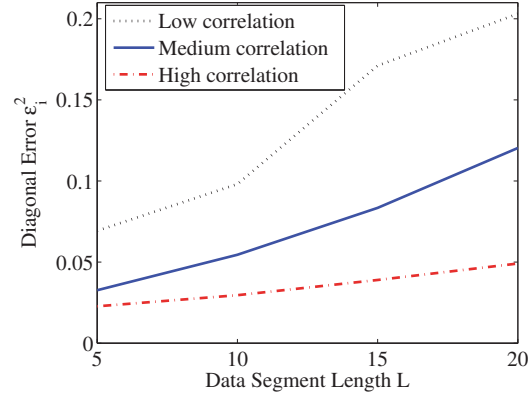


(b) Squared diagonal error of $\mathbf{C}_{\mathbf{x}\mathbf{x}^j}$

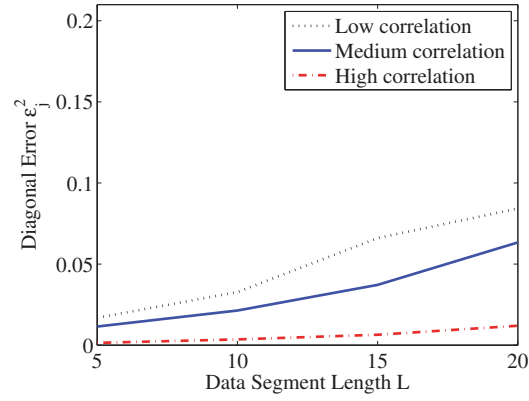


(c) Squared diagonal error of $\mathbf{C}_{\mathbf{x}\mathbf{x}^k}$

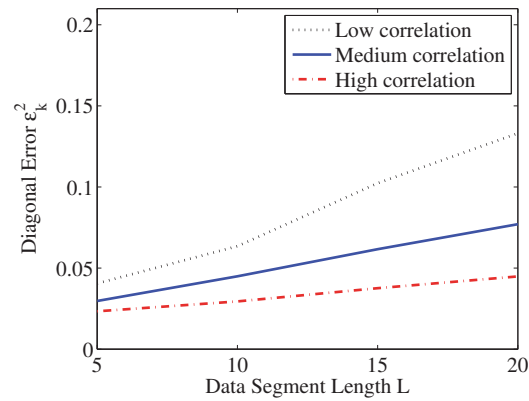
Figure 3.1: Squared diagonal error [in %] against the correlation degree and the data segment length for synthetic data.



(a) Squared diagonal error of $\mathbf{C}_{\mathbf{x}^i}$



(b) Squared diagonal error of $\mathbf{C}_{\mathbf{x}^j}$



(c) Squared diagonal error of $\mathbf{C}_{\mathbf{x}^k}$

Figure 3.2: Squared diagonal error [in %] against the correlation degree and the data segment length for EEG data.

3.4.3 QAUT for Rank Reduction

In many practical applications the acquired data are highly correlated or even collinear, with redundant elements that do not contribute to the accuracy of processing but require excessive computational resources and may affect stability of algorithms through the associated singular correlation matrices. It is therefore desirable to use the minimum number of variables to describe the information within a data set; this is typically achieved through a low-rank matrix approximation. For data requiring rank reduction, the assumption that the components of variables are highly correlated is sensible, that is, the approximation conditions for the QAUT hold. The QAUT can be therefore rearranged to decompose the data matrix $\mathbf{X}$ into a sum of uncorrelated rank one vector outer products

$$\mathbf{X} = \mathbf{\Phi}^H \mathbf{Y} = \sum_{m=1}^M \phi_m^H \mathbf{y}_m$$

where ϕ_m and $\mathbf{y}_m$ are rows in the matrices $\mathbf{\Phi}$ and $\mathbf{Y}$, respectively, and the transformed variables, $\mathbf{y}_m$, are monotonically non-increasing in the variance. The use of variables for which the variance is above a defined threshold and the disposal of the remaining variables, which account for noise, provides the reduced rank approximation of $\mathbf{X}$ in the form

$$\hat{\mathbf{X}} = \sum_{m=1}^P \phi_m^H \mathbf{y}_m, \quad \text{where } P < M$$

which is illustrated in Figure 3.3. This is analogous to the real domain where the SVD provides a change of coordinates to yield uncorrelated variables for which the variance is monotonically non-increasing.

To illustrate the potential of the QAUT in rank reduction of quaternion data, a 2D quaternion process, $\mathbf{x}_1$, $\mathbf{x}_2$, was generated where $\mathbf{x}_1$ and $\mathbf{x}_2$ were highly correlated as shown in Figure 3.4, which plots the component relationship in the

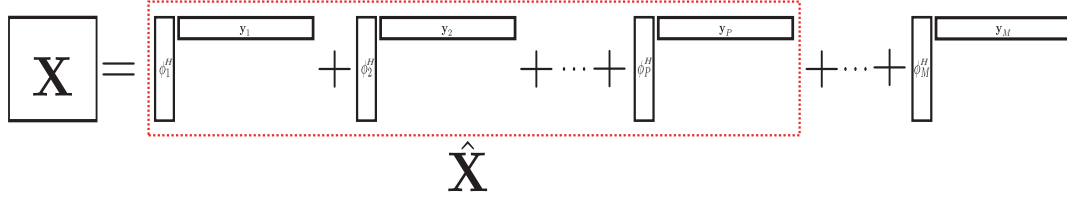


Figure 3.3: Rank reduction using the QAUT.

real, $\imath$, j and κ planes. The QAUT decorrelated $\mathbf{x}_1$, $\mathbf{x}_2$ into $\mathbf{y}_1$ and $\mathbf{y}_2$. Figure 3.5 shows significant variation only in the $\mathbf{y}_2$ axis while the $\mathbf{y}_1$ variation corresponds to noise, indicating that the true dimensionality of the process is one rather than two. Note that this method is not suitable when the noise power is comparable to signal power.

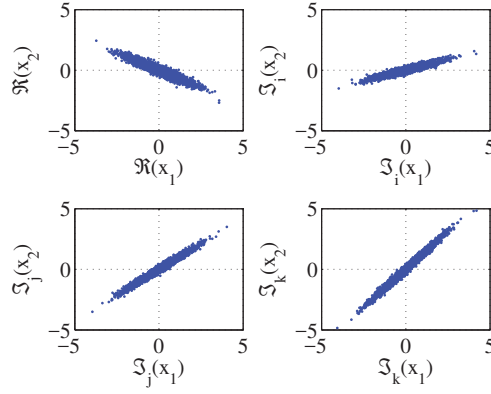


Figure 3.4: Scatter plots showing high correlation between the original x_1 and x_2 variables.

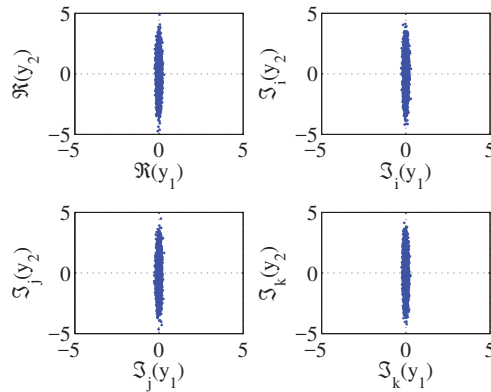


Figure 3.5: Scatter plots showing no correlation between the original y_1 and y_2 variables.

Rank reduction is commonly used in chemometrics where many sets of measurements are required to examine a process and the important information may not be directly observable. For example, Near Infrared Spectra (NIR) are used to examine the chemical content of a data sample and are often highly correlated or collinear. The QAUT was applied to the NIR obtained from 80 samples of corn using three different spectrometers [71]. A pure quaternion variable was formed for each corn sample by combining its spectra from the three spectrometers. The QAUT diagonalised the data covariance matrices successfully to show that one quaternion variable was sufficient to express most of the information, as the largest diagonal element was 39 compared to 0.01 for the next largest, as shown in Figure 3.6, which displays the logarithm of the ratio of variance explained by each variable, compared to the total variance in the data. The insert plot in Figure 3.6 shows the cumulative ratio of variance explained by including further variables for the original and transformed data. Observe that the transformed data with much fewer variables explain the same information as the original, thus verifying the QAUT.

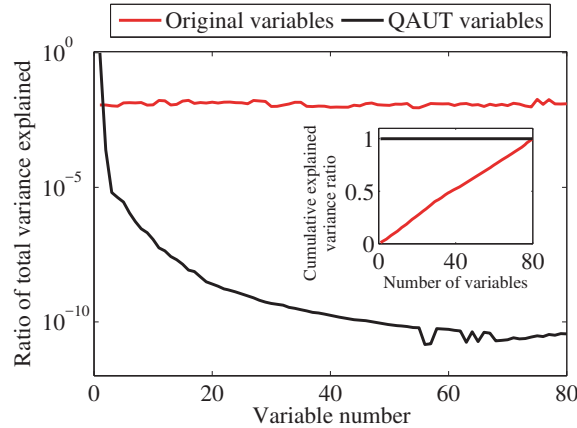


Figure 3.6: The logarithm of the ratio of the variance explained by each variable to the total variance. *Insert diagram:* The cumulative ratio of variance against the number of variables for the original and decorrelated corn data.

3.4.4 QAUT for Imbalance Detection in Three-Phase Power Systems

As shown in Section 3.3, when the components of data are highly correlated, the QAUT diagonalises all complementary covariance matrices at the expense of a small error, however, this error increases for lower degrees of correlation. This data dependence of the QAUT can be exploited to detect imbalance in a three-phase power system.

In general, the instantaneous voltages of a three-phase power system are given by [72]

$$\begin{aligned} v_1(n) &= V_1 \sin(2\pi f T n + \varphi_1) \\ v_2(n) &= V_2 \sin(2\pi f T n + \varphi_2) \\ v_3(n) &= V_3 \sin(2\pi f T n + \varphi_3) \end{aligned} \tag{3.14}$$

where V_1, V_2, V_3 are the amplitudes, $\varphi_1, \varphi_2, \varphi_3$ the phases, f the system frequency, and T the sampling interval. In a balanced three-phase system, $A_1 = A_2 = A_3$, $\varphi_2 - \varphi_1 = \frac{2\pi}{3}$, $\varphi_3 - \varphi_2 = \frac{2\pi}{3}$, and the correlation between each component is equivalent [73]. The power grid is usually designed to operate optimally under balanced conditions; however, faults in the power system can cause imbalanced operating conditions that propagate through the network, threatening its stability. Therefore, it is important in fault detection and mitigation applications to identify the incidences when the power grid is operating in an unbalanced fashion [74].

Pure quaternions have been used to deal with data recordings from all three phases simultaneously as $x(n) = v_1(n)\iota + v_2(n)j + v_3(n)\kappa$ [72]. Organise the recorded data into a pure quaternion vector, $\mathbf{x} = \mathbf{v}_1\iota + \mathbf{v}_2j + \mathbf{v}_3\kappa$, where $\mathbf{v}_m = [v_m(1), v_m(2), \dots, v_m(n)]^T$, $m = 1, 2, 3$. Then, the QAUT can be applied to $\mathbf{x}$ and the diagonalisation errors of the complementary covariance matrices can be employed as an indicator of power imbalance. This can be explained by the overall

diagonalisation error given by

$$\begin{aligned} \Im [x_l^* (n_1) x_l (n_2)] = & [v_2 (n_2) v_3 (n_1) - v_2 (n_1) v_3 (n_2)] \iota + [v_1 (n_1) v_3 (n_2) - v_1 (n_2) v_3 (n_1)] j \\ & + [v_1 (n_2) v_2 (n_1) - v_1 (n_1) v_2 (n_2)] \kappa \end{aligned} \quad (3.15)$$

which is a reduced form of (3.12) and is positively correlated with the diagonalisation error. When the system is balanced, by substituting (3.14) into (3.15), it can be derived that the three imaginary parts of $x_l^* (n_1) x_l (n_2)$ in (3.15) are equal, leading to identical diagonalisation errors of the three complementary covariance matrices, which is an implication of (3.11). On the other hand, when the system becomes unbalanced, the three imaginary parts of $x_l^* (n_1) x_l (n_2)$ differ, resulting in different diagonalisation errors of the three complementary covariance matrices, which is implied by (3.11). This allows the QAUT to be used in an imbalance detection application. To this end, the QAUT was applied to a sliding window of 0.04 seconds of a three-phase power system for which a fault occurred at 0.25s lasting for 0.25s before the system returned to a balanced state. During this fault, $\mathbf{v}_1$ decreased significantly in amplitude whereas $\mathbf{v}_2$ and $\mathbf{v}_3$ increased in amplitude ($\mathbf{v}_2 > \mathbf{v}_3$). Figure 3.7 shows that the diagonalisation error for $\mathbf{C}_{\mathbf{x}\mathbf{x}^j}$ and $\mathbf{C}_{\mathbf{x}\mathbf{x}^\kappa}$ increased while the error for $\mathbf{C}_{\mathbf{x}\mathbf{x}^\iota}$ decreased and therefore analysing the difference between the errors is an appropriate detection method for the power imbalance. This can be explained via (3.14), (3.15) and (3.11). The decrease in the amplitude of v_1 and the increase in the amplitudes of v_2 and v_3 induced an increased ι -imaginary part and decreased j -imaginary and κ -imaginary parts of $x_l^* (n_1) x_l (n_2)$, so that the diagonalisation error of $\mathbf{C}_{\mathbf{x}\mathbf{x}^\iota}$ declined and the diagonalisation errors of $\mathbf{C}_{\mathbf{x}\mathbf{x}^j}$ and $\mathbf{C}_{\mathbf{x}\mathbf{x}^\kappa}$ rose.

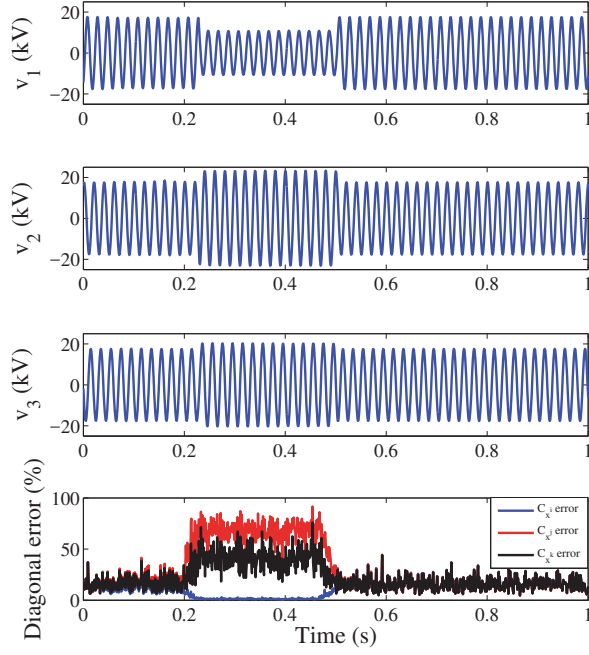


Figure 3.7: Three-phase power signal and imbalance detection.

3.5 Summary

A novel simultaneous matrix factorisation technique for the covariance and complementary covariance matrices, which arises in widely linear quaternion algebra, have been introduced. It has also been shown that for improper quaternion data, a typical case in quaternion signal processing applications, a single eigendecomposition of the covariance matrix is sufficient to diagonalise approximately the three complementary matrices. The analysis of the so introduced QAUT has demonstrated its usefulness for improper quaternion signals. Simulation studies on synthetic and real-world signals support the proposed QAUT technique.

Part II

Performance Analysis of Minimum Mean Square Error Estimation Techniques

Chapter 4

Performance Advantage of Widely Linear Quaternion Estimation

Overview

Widely linear processing has been shown to be superior to the traditional strictly linear processing in quaternion minimum mean square error (MMSE) estimation. However, a quantifiable performance difference between strictly and widely linear processing and the relationship between the performance and quaternion impropriety are still lacking. To this end, this chapter presents a proof for the performance advantage of widely linear estimation and relates the performance bounds to signal properties by exploiting the approximate joint diagonalisation of quaternion covariance matrices. This work can be seen as a generalisation of complex-valued MMSE estimation, and can thus also be applied to the complex-valued case.

4.1 Introduction

Conventional strictly linear quaternion MMSE estimators utilise the standard covariance and are optimal only for estimating second-order circular quaternion signals. Advances in quaternion statistics have established that in order to capture complete second-order statistics of quaternion signals, widely linear MMSE estimators that exploit three complementary covariances in addition to the standard covariance are needed [21, 22, 29]. Due to the non-trivial nature of quaternion al-

gebra, a closed-form performance comparison between strictly and widely linear MMSE estimators has not yet been presented. The only available result was achieved in an indirect fashion via the semi-widely linear estimator [30, 31, 75], which is a generalisation of the corresponding analysis for complex estimation [76] to the quaternion domain. In addition, the extent to which the performance bounds of these estimators are affected by the impropriety of signals remains unclear.

This chapter provides an analytical understanding and second-order performance comparison between strictly and widely linear estimators. This is achieved by examining the structure of the covariance and complementary covariance matrices of the input signal, and by employing the QAUT proposed in Chapter 3. In this way, the relationship between the MMSE and second-order statistics of the input signal is established, and the effect of signal properties on the performance is quantified. The analysis is supported by illustrative simulations.

4.2 Performance Advantage of Widely Linear Estimation

Consider the desired signal, y , which is generated by a widely linear system driven by noise, v , to give

$$y = \mathbf{h}^H \mathbf{x} + \mathbf{g}^H \mathbf{x}^i + \mathbf{u}^H \mathbf{x}^j + \mathbf{v}^H \mathbf{x}^\kappa + v = \mathbf{h}^H \mathbf{x} + \mathbf{w}^{cH} \mathbf{x}^c + v = \mathbf{w}^{aH} \mathbf{x}^a + v \quad (4.1)$$

where $\mathbf{w}^a = [\mathbf{h}^T, \mathbf{g}^T, \mathbf{u}^T, \mathbf{v}^T]^T$ denotes the true augmented weight vector of the widely linear system, while $\mathbf{w}^c = [\mathbf{g}^T, \mathbf{u}^T, \mathbf{v}^T]^T$ designates the true complementary weight vector of the widely linear system, $\mathbf{x}^c = [\mathbf{x}^{iT}, \mathbf{x}^{jT}, \mathbf{x}^{\kappa T}]^T$, and v is quaternion white Gaussian noise that is statistically independent of $\mathbf{x}$. The variance of v is denoted by σ_v^2 .

4.2.1 MMSE in Strictly Linear Estimation

It has been proved that the optimal weight of the strictly linear estimator in (2.18) is provided by the Wiener solution [75], given by

$$\hat{\mathbf{h}}_o = E \{ \mathbf{x} \mathbf{x}^H \}^{-1} E \{ \mathbf{x} y^* \} = \mathbf{h} + \mathbf{C}_{\mathbf{x}\mathbf{x}}^{-1} \mathbf{C}_{\mathbf{x}\mathbf{x}^c} \mathbf{w}^c \quad (4.2)$$

Note that $\hat{\mathbf{h}}_o \neq \mathbf{h}$ if $\mathbf{C}_{\mathbf{x}\mathbf{x}}^{-1} \mathbf{C}_{\mathbf{x}\mathbf{x}^c} \mathbf{w}^c \neq \mathbf{0}$, which occurs when the system is widely linear ($\mathbf{w}^c \neq \mathbf{0}$) and $\mathbf{x}$ is not $\mathbb{H}$ -proper ($\mathbf{C}_{\mathbf{x}\mathbf{x}^c} \neq \mathbf{0}$). The strictly linear estimator yields the steady-state error, e_{sl} , given by

$$e_{\text{sl}} \triangleq y - \hat{\mathbf{h}}_o^H \mathbf{x} = \mathbf{w}^{cH} \mathbf{x}^c - \mathbf{w}^{cH} \mathbf{C}_{\mathbf{x}\mathbf{x}^c}^H \mathbf{C}_{\mathbf{x}\mathbf{x}}^{-1} \mathbf{x} + v$$

and the corresponding MMSE is given by

$$E \{ |e_{\text{sl}}|^2 \} = \sigma_v^2 + \mathbf{w}^{cH} \mathbf{M} \mathbf{w}^c \quad (4.3)$$

where $\mathbf{M} = \mathbf{C}_{\mathbf{x}^c \mathbf{x}^c} - \mathbf{C}_{\mathbf{x}\mathbf{x}^c}^H \mathbf{C}_{\mathbf{x}\mathbf{x}}^{-1} \mathbf{C}_{\mathbf{x}\mathbf{x}^c}$ is the Schur complement of the augmented covariance matrix $\mathbf{C}_{\mathbf{x}^a \mathbf{x}^a}$, which has the following block structure:

$$\mathbf{C}_{\mathbf{x}^a \mathbf{x}^a} = \begin{bmatrix} \mathbf{C}_{\mathbf{x}\mathbf{x}} & \mathbf{C}_{\mathbf{x}\mathbf{x}^c} \\ \mathbf{C}_{\mathbf{x}^c \mathbf{x}} & \mathbf{C}_{\mathbf{x}^c \mathbf{x}^c} \end{bmatrix} \quad (4.4)$$

4.2.2 MMSE in Widely Linear Estimation

The optimal weight vector of the widely linear estimator in (2.19) is given by the Wiener solution

$$\hat{\mathbf{w}}_o^a = E \{ \mathbf{x}^a \mathbf{x}^{aH} \}^{-1} E \{ \mathbf{x}^a y^* \} = \mathbf{w}^a$$

resulting in the steady-state error, $e_{\text{wl}} = v$, and the corresponding MMSE

$$E \{ |e_{\text{wl}}|^2 \} = \sigma_v^2 \quad (4.5)$$

4.2.3 MMSE Comparison Between Strictly and Widely Linear Estimation

The difference between the MMSEs of the strictly and widely linear quaternion estimators, δe^2 , is obtained from (4.3) and (4.5), to yield

$$\delta e^2 \triangleq E \{ |e_{\text{sl}}|^2 \} - E \{ |e_{\text{wl}}|^2 \} = \mathbf{w}^{cH} \mathbf{M} \mathbf{w}^c \quad (4.6)$$

Since $\mathbf{C}_{\mathbf{x}^a \mathbf{x}^a}$ is positive semi-definite, its Schur complement, $\mathbf{M}$, is also positive semi-definite [77], and thus $\delta e^2 \geq 0$, indicating that the widely linear estimator has a smaller or equal MMSE compared to its strictly linear counterpart. However, their MMSEs are equal, that is, $\delta e^2 = 0$, if at least one of the following three conditions is satisfied:

1. The system being modelled is strictly linear, that is, $\mathbf{w}^c = \mathbf{0}$.
2. The input signal vector $\mathbf{x}$ is rectilinear, that is, $\mathbf{x} = \rho_i \mathbf{x}^i = \rho_j \mathbf{x}^j = \rho_\kappa \mathbf{x}^\kappa$ with probability 1, where ρ_α ($\alpha \in \{i, j, \kappa\}$) is a unit quaternion with a vanishing α -imaginary component. In this case, $\mathbf{C}_{\mathbf{x} \mathbf{x}^\alpha} = \mathbf{C}_{\mathbf{x} \mathbf{x}} \rho_\alpha$, and $\mathbf{M} = \mathbf{0}$.
3. The complementary weight vector $\mathbf{w}^c$ falls within the nullspace of $\mathbf{M}$, that is, $\mathbf{M} \mathbf{w}^c = \mathbf{0}$. This case is highly unlikely in practical engineering applications.

Remark 4.1. Widely linear processing has a lower MMSE than strictly linear processing when estimating signals generated by widely linear systems, unless the second or third condition holds. Otherwise, their MMSEs are equal.

Also notice that

$$\delta e^2 \leq \mathbf{w}^{cH} \mathbf{C}_{\mathbf{x}^c \mathbf{x}^c} \mathbf{w}^c$$

This equality holds if at least one of the following two conditions is satisfied:

1. $\mathbf{x}$ is $\mathbb{H}$ -proper, that is, $\mathbf{C}_{\mathbf{xx}^c} = \mathbf{0}$.
2. The complementary weight vector $\mathbf{w}^c$ falls within the nullspace of $\mathbf{C}_{\mathbf{xx}^c}^H \mathbf{C}_{\mathbf{xx}}^{-1} \mathbf{C}_{\mathbf{xx}^c}$, that is, $(\mathbf{C}_{\mathbf{xx}^c}^H \mathbf{C}_{\mathbf{xx}}^{-1} \mathbf{C}_{\mathbf{xx}^c}) \mathbf{w}^c = \mathbf{0}$. This case is highly unlikely in practice.

4.3 Performance Analysis Using the QAUT

The QAUT introduced in Chapter 3 allows for a simultaneous diagonalisation of the covariance matrix, $\mathbf{C}_{\mathbf{xx}}$, and the complementary covariance matrices, $\mathbf{C}_{\mathbf{xx}^i}, \mathbf{C}_{\mathbf{xx}^j}, \mathbf{C}_{\mathbf{xx}^\kappa}$, using a unitary transform. To understand the effect of matrix $\mathbf{M}$ on δe^2 in (4.6), the following assumption shall be used.

Assumption 4.1. There exists a quaternion unitary matrix $\mathbf{Q}$ such that

$$\mathbf{C}_{\mathbf{xx}} = \mathbf{Q} \mathbf{\Lambda}_{\mathbf{xx}} \mathbf{Q}^H, \quad \mathbf{C}_{\mathbf{xx}^\alpha} = \mathbf{Q} \mathbf{\Lambda}_{\mathbf{xx}^\alpha} \mathbf{Q}^{\alpha H}, \quad \alpha = i, j, \kappa \quad (4.7)$$

where $\mathbf{\Lambda}_{\mathbf{xx}}$ is a diagonal real-valued matrix containing the eigenvalues of $\mathbf{C}_{\mathbf{xx}}$ on the diagonal, and $\mathbf{\Lambda}_{\mathbf{ss}^\alpha}$ are diagonal quaternion matrices.

It can be verified that Assumption 4.1 holds exactly when one of the following conditions on $\mathbf{x}$ is met:

1. The covariance and complementary covariance matrices of $\mathbf{x}$ are related by scalar multiplication. This condition holds for quaternion random vectors with special second-order statistics, such as an $\mathbb{H}$ -proper vector ($\mathbf{C}_{\mathbf{xx}^\alpha} = 0\mathbf{C}_{\mathbf{xx}}$) and a rectilinear vector ($\mathbf{C}_{\mathbf{xx}^\alpha} = \mathbf{C}_{\mathbf{xx}}\rho_\alpha$).
2. $\mathbf{x}$ can be expressed by applying the unitary transform, $\mathbf{Q}$, to a white Gaussian random quaternion vector $\mathbf{v}$, that is, $\mathbf{x} = \mathbf{Q}\mathbf{v}$.

For quaternion inputs that do not meet the above two conditions, Assumption 3.1 holds approximately as the result of the QAUT. Under this assumption, the

block partitioned matrices in (4.4) can be expressed as

$$\mathbf{C}_{\mathbf{x}\mathbf{x}^c} = \mathbf{Q} \begin{bmatrix} \Lambda_{\mathbf{x}\mathbf{x}^i} & \Lambda_{\mathbf{x}\mathbf{x}^j} & \Lambda_{\mathbf{x}\mathbf{x}^\kappa} \end{bmatrix} \mathbf{U}^H \quad (4.8)$$

$$\mathbf{C}_{\mathbf{x}^c\mathbf{x}^c} = \mathbf{U} \begin{bmatrix} \Lambda_{\mathbf{x}\mathbf{x}} & \Lambda_{\mathbf{x}\mathbf{x}^\kappa}^i & \Lambda_{\mathbf{x}\mathbf{x}^j}^i \\ \Lambda_{\mathbf{x}\mathbf{x}^\kappa}^j & \Lambda_{\mathbf{x}\mathbf{x}} & \Lambda_{\mathbf{x}\mathbf{x}^i}^j \\ \Lambda_{\mathbf{x}\mathbf{x}^j}^\kappa & \Lambda_{\mathbf{x}\mathbf{x}^i}^\kappa & \Lambda_{\mathbf{x}\mathbf{x}} \end{bmatrix} \mathbf{U}^H \quad (4.9)$$

where $\mathbf{U} = \text{bdiag}(\mathbf{Q}^i, \mathbf{Q}^j, \mathbf{Q}^\kappa)$ is a unitary block diagonal matrix with the involutions of $\mathbf{Q}$ as the block diagonal entries. Therefore, the Schur complement matrix, $\mathbf{M}$, in (4.6), can be factorised as $\mathbf{M} = \mathbf{U}\mathbf{S}\mathbf{U}^H$ where

$$\mathbf{S} = \begin{bmatrix} \Lambda_{\mathbf{x}\mathbf{x}} - \Lambda_{\mathbf{x}\mathbf{x}^i}^* \Lambda_{\mathbf{x}\mathbf{x}}^{-1} \Lambda_{\mathbf{x}\mathbf{x}^i} & \Lambda_{\mathbf{x}\mathbf{x}^\kappa}^i - \Lambda_{\mathbf{x}\mathbf{x}^i}^* \Lambda_{\mathbf{x}\mathbf{x}}^{-1} \Lambda_{\mathbf{x}\mathbf{x}^j} & \Lambda_{\mathbf{x}\mathbf{x}^j}^i - \Lambda_{\mathbf{x}\mathbf{x}^i}^* \Lambda_{\mathbf{x}\mathbf{x}}^{-1} \Lambda_{\mathbf{x}\mathbf{x}^\kappa} \\ \Lambda_{\mathbf{x}\mathbf{x}^\kappa}^j - \Lambda_{\mathbf{x}\mathbf{x}^j}^* \Lambda_{\mathbf{x}\mathbf{x}}^{-1} \Lambda_{\mathbf{x}\mathbf{x}^i} & \Lambda_{\mathbf{x}\mathbf{x}} - \Lambda_{\mathbf{x}\mathbf{x}^j}^* \Lambda_{\mathbf{x}\mathbf{x}}^{-1} \Lambda_{\mathbf{x}\mathbf{x}^j} & \Lambda_{\mathbf{x}\mathbf{x}^i}^j - \Lambda_{\mathbf{x}\mathbf{x}^j}^* \Lambda_{\mathbf{x}\mathbf{x}}^{-1} \Lambda_{\mathbf{x}\mathbf{x}^\kappa} \\ \Lambda_{\mathbf{x}\mathbf{x}^j}^\kappa - \Lambda_{\mathbf{x}\mathbf{x}^\kappa}^* \Lambda_{\mathbf{x}\mathbf{x}}^{-1} \Lambda_{\mathbf{x}\mathbf{x}^i} & \Lambda_{\mathbf{x}\mathbf{x}^i}^\kappa - \Lambda_{\mathbf{x}\mathbf{x}^\kappa}^* \Lambda_{\mathbf{x}\mathbf{x}}^{-1} \Lambda_{\mathbf{x}\mathbf{x}^j} & \Lambda_{\mathbf{x}\mathbf{x}} - \Lambda_{\mathbf{x}\mathbf{x}^\kappa}^* \Lambda_{\mathbf{x}\mathbf{x}}^{-1} \Lambda_{\mathbf{x}\mathbf{x}^\kappa} \end{bmatrix}$$

and (4.6) becomes $\delta e^2 = \mathbf{w}^{cH} \mathbf{U} \mathbf{S} \mathbf{U}^H \mathbf{w}^c$.

4.3.1 Impact of Impropriety of the Input Signal

In order to clarify the relationship between the performance advantage of widely linear estimation and impropriety of $\mathbf{x}$, it is useful to discuss the following two special cases prior to the general case.

$\mathbb{H}$ -proper signal

For an $\mathbb{H}$ -proper $\mathbf{x}$,

$$\delta e^2 = \mathbf{w}^{cH} \mathbf{C}_{\mathbf{x}^c\mathbf{x}^c} \mathbf{w}^c = \mathbf{g}^H \mathbf{Q}^i \Lambda_{\mathbf{x}\mathbf{x}} \mathbf{Q}^{iH} \mathbf{g} + \mathbf{u}^H \mathbf{Q}^j \Lambda_{\mathbf{x}\mathbf{x}} \mathbf{Q}^{jH} \mathbf{u} + \mathbf{v}^H \mathbf{Q}^\kappa \Lambda_{\mathbf{x}\mathbf{x}} \mathbf{Q}^{\kappa H} \mathbf{v}$$

If $\mathbf{\Lambda}_{\mathbf{xx}} = \lambda \mathbf{I}$, where $\mathbf{I}$ is the identity matrix, then,

$$\delta e^2 = \lambda (\|\mathbf{g}\|^2 + \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2) = \lambda \|\mathbf{w}^c\|^2 \quad (4.10)$$

Remark 4.2. Without loss in generality, the performance difference between widely linear and strictly linear estimation increases as the Euclidean norm of the complementary weight vector, $\|\mathbf{w}^c\|$, increases. This statement is true regardless of the impropriety of the input signal since the complementary weight of the widely linear system cannot be modelled by the strictly linear estimator.

Uncorrelated improper signal

For an improper random quaternion vector $\mathbf{x}$ for which the elements are uncorrelated and have equal power and impropriety degree, the covariance and complementary covariance matrices are $\mathbf{C}_{\mathbf{xx}} = c_x \mathbf{I}$, $\mathbf{C}_{\mathbf{xx}^i} = c_i \mathbf{I}$, $\mathbf{C}_{\mathbf{xx}^j} = c_j \mathbf{I}$, $\mathbf{C}_{\mathbf{xx}^\kappa} = c_\kappa \mathbf{I}$, where c_x is a real-valued constant, and c_i, c_j, c_κ are quaternion-valued constants. The performance difference in (4.6) therefore becomes

$$\delta e^2 = \mathbf{w}^{cH} \left(\begin{bmatrix} c_x - c_x^{-1} |c_i|^2 & c_\kappa^i - c_x^{-1} c_i^* c_j & c_j^i - c_x^{-1} c_i^* c_\kappa \\ c_j^\kappa - c_x^{-1} c_j^* c_i & c_x - c_x^{-1} |c_j|^2 & c_i^j - c_x^{-1} c_j^* c_\kappa \\ c_j^\kappa - c_x^{-1} c_\kappa^* c_i & c_i^\kappa - c_x^{-1} c_\kappa^* c_j & c_x - c_x^{-1} |c_\kappa|^2 \end{bmatrix} \otimes \mathbf{I} \right) \mathbf{w}^c \quad (4.11)$$

where “ $\otimes$ ” denotes the Kronecker product [78].

General improper signal

To examine the effect of matrix $\mathbf{M}$ on the performance difference δe^2 , its trace shall be inspected. Since $\mathbf{U}$ is unitary, the trace of $\mathbf{M}$ is

$$\text{Tr}(\mathbf{M}) = \text{Tr} \left(3\mathbf{\Lambda}_{\mathbf{xx}} - \sum_{\alpha=i,j,\kappa} \mathbf{\Lambda}_{\mathbf{xx}^\alpha}^* \mathbf{\Lambda}_{\mathbf{xx}}^{-1} \mathbf{\Lambda}_{\mathbf{xx}^\alpha} \right) = \sum_{l=1}^L \lambda_l \left(3 - \sum_{\alpha=i,j,\kappa} r_{\alpha,l}^2 \right) \quad (4.12)$$

where λ_l is the l -th diagonal element of $\mathbf{\Lambda}_{\mathbf{xx}}$, and $r_{\alpha,l}$ the α -impropriety coefficient of the l -th element of the transformed random vector, $\mathbf{Q}^H \mathbf{x}$. As the degree of signal impropriety, which is reflected in $r_{\alpha,l}$, increases, $\text{Tr} [\mathbf{M}]$ decreases, so do eigenvalues of $\mathbf{M}$, and hence δe^2 decreases.

Remark 4.3. An $\mathbb{H}$ -proper quaternion random vector $\mathbf{x}$ and its involutions are uncorrelated, and the Wiener solution of strictly linear estimation is equal to $\mathbf{h}$, the system weight of $\mathbf{x}$ in (4.1). In contrast, an $\mathbb{H}$ -improper quaternion random vector $\mathbf{x}$ and its involutions are correlated, therefore some information about the involutions that is embedded in $\mathbf{x}$ can be exploited using the strictly linear model, in which case the low performance of strictly linear estimation is compensated for by the additional term $\mathbf{C}_{\mathbf{xx}}^{-1} \mathbf{C}_{\mathbf{xx}^c} \mathbf{w}^c$ in (4.2). This could explain why widely linear estimation shows only slight improvement in MSE performance over strictly linear estimation for highly improper signals, such as the 4D wind data in [79].

4.3.2 MMSE Analysis for Complex MMSE Estimation

The above analysis reduces to the complex case when the quaternion signal reduces to the complex one, that is, $\mathbf{x} = \mathbf{x}_1 + \imath \mathbf{x}_2$, where $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^{L \times 1}$. It follows that $\mathbf{C}_{\mathbf{xx}^i} = \mathbf{C}_{\mathbf{xx}}$, $\mathbf{C}_{\mathbf{xx}^j} = \mathbf{C}_{\mathbf{xx}^k} = E\{\mathbf{xx}^T\}$. Define the pseudo-covariance of $\mathbf{x}$ as $\mathbf{P}_{\mathbf{xx}} = E\{\mathbf{xx}^T\}$. The difference between the MMSEs of strictly linear and widely linear complex estimators [7] can now be obtained through a degeneration of (4.6), and is given by

$$\delta e^2 = E\{|e_{\text{sl}}|^2\} - E\{|e_{\text{wl}}|^2\} = \mathbf{g}^H (\mathbf{C}_{\mathbf{xx}}^* - \mathbf{P}_{\mathbf{xx}}^* \mathbf{C}_{\mathbf{xx}}^{-1} \mathbf{P}_{\mathbf{xx}}) \mathbf{g} \quad (4.13)$$

which conforms with the previous work [80], while (4.12) reduces to

$$\text{Tr} (\mathbf{M}) = \text{Tr} (\mathbf{\Lambda}_{\mathbf{xx}} - \mathbf{\Lambda}_p^* \mathbf{\Lambda}_{\mathbf{xx}}^{-1} \mathbf{\Lambda}_p) = \sum_{l=1}^L \lambda_l \left(1 - \left| \frac{\lambda_{p,l}}{\lambda_l} \right|^2 \right) \quad (4.14)$$

where $\mathbf{\Lambda}_p$ is a diagonal matrix with diagonal elements, $\lambda_{p,l}$, and $\left| \frac{\lambda_{p,l}}{\lambda_n} \right|$ reflects the degree of impropriety of the complex-valued $\mathbf{x}$. The performance advantage δe^2 of complex widely linear estimation increases with the widely linear nature of the system, propriety of $\mathbf{x}$, or power of $\mathbf{x}$. A higher degree of impropriety indicates higher correlation between $\mathbf{x}$ and its conjugate $\mathbf{x}^*$, meaning that some information about $\mathbf{x}^*$ that is embedded in $\mathbf{x}$ can be exploited via the strictly linear model, which compensates for the performance loss of strictly linear estimation.

4.4 Simulations

The simulation study was conducted by averaging the steady-state error power from 100 independent trials of quaternion least mean square (QLMS) algorithms. A quaternion random vector $\mathbf{x}$ with a varying degree of impropriety was generated and the widely linear system in (4.1) was constructed. To avoid the excess MSE of QLMS [4], the noise v was omitted from the system. The length of $\mathbf{x}$ was 4, and the complementary weights of the system were set as $\mathbf{g} = \mathbf{u} = \mathbf{v} = \beta \mathbf{h}$ with $\beta \in \mathbb{R}$, a varying coefficient reflecting the widely linear nature of the system. For simplicity, elements of $\mathbf{x}$ were generated to be uncorrelated and with equal power, so the theoretical performance advantage of widely linear estimation could be calculated from (4.11). The impropriety coefficients r_i, r_j, r_κ of $\mathbf{x}$ were equalised to a unique r . The QLMS algorithms based on the strictly linear and widely linear models were employed to estimate y . As shown in Figure 4.1, the theoretical and simulated performance advantage of the widely linear QLMS over the strictly linear QLMS were in agreement, and they both increased with the norm of the complementary weight vector of the widely linear system and the variance of the input signal, and decreased as the impropriety coefficient of $\mathbf{x}$ increased. The curves in Figure 4.1 (a) and (b) are parabolas, conforming with the quadratic forms of $\mathbf{w}^c$ and r_α in (4.6) and (4.12). The linear trend shown in Figure 4.1 (c) conforms with the linear form of λ in (4.12).

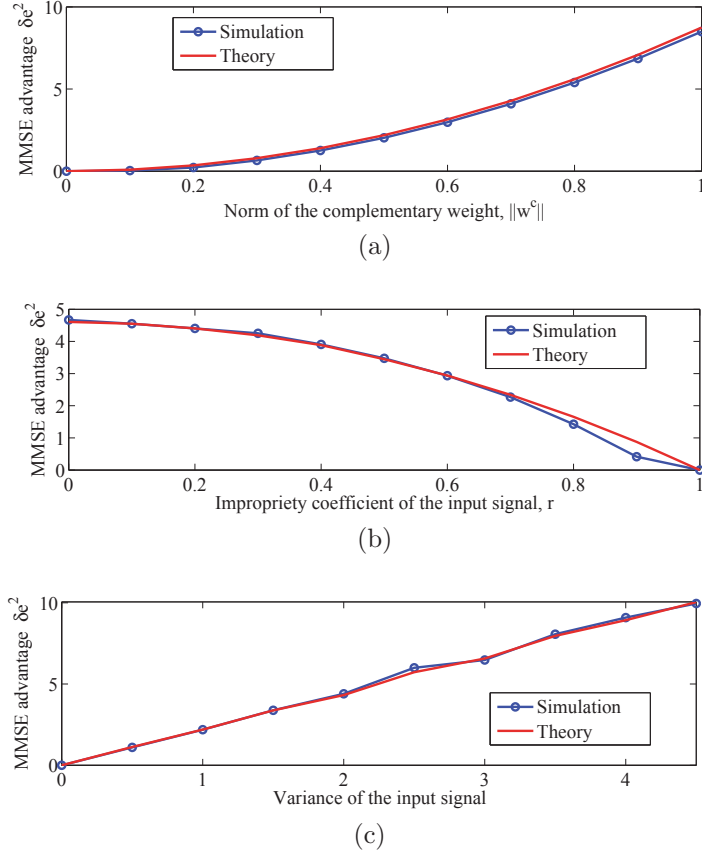


Figure 4.1: Effects of (a) widely linear nature of the system, (b) impropriety of the input signal, and (c) variance of the input signal on the MMSE advantage of the widely linear QLMS estimator.

4.5 Summary

This chapter has compared the performance of the strictly linear and widely linear models for MMSE estimation of quaternion signals. The widely linear model has been proved to achieve a lower or equal MSE than the strictly linear model, and this performance advantage has been found to be dependent on the degree of widely linear nature of the system, impropriety and power of the input signal. Simulations on synthetic signals support the analysis.

Chapter 5

Convergence Performance of Quaternion Adaptive Filters in Non-stationary Environments

Overview

Quaternion adaptive filters have been extensively used, but complete analysis of their performance is still lacking, partly due to the cumbersomeness of multivariate quaternion analysis. This causes difficulties in both understanding their behaviour and designing optimal filters. This chapter extends an analysis framework for real-valued adaptive filters to convergence analysis of a general class of quaternion adaptive filters in non-stationary environments. The extension is non-trivial, considering the non-commutative quaternion algebra, quaternion gradient, and multidimensional augmented quaternion statistics. Transient and steady-state performance of the considered filters is established. For the analysis of QLMS filters with Gaussian inputs, the QAUT proposed in Chapter 3 is utilised to yield intuitive closed-form solutions, which relates the filter behaviour to the complete second-order statistics of the input signal, that is, quaternion noncircularity. The closed-form expressions of the performance of widely, semi-widely, and strictly linear QLMS filters are investigated in detail.

5.1 Introduction

Contrary to the research on real-valued adaptive filters [4, 5, 81, 82], complete performance analysis of quaternion adaptive filters is still an open problem, causing difficulties in understanding their behaviour and in the design of optimal filtering strategies. Since the quadrivariate real filters are not isomorphic to the whole class of quaternion filters, it is impossible to use the performance of the former to straightforwardly analyse the latter, and thus the performance analysis of the latter must be undertaken in the quaternion domain. The mean convergence analysis for the QLMS filters was proposed in [83], but the non-commutativity of quaternion products poses a challenge to the application of classical mean square convergence analysis methods for real-valued and complex-valued adaptive filters to the quaternion domain. For the mean square convergence analysis, the work in [49] discussed a simple case, the univariate strictly linear QLMS algorithm, and derived the bounds of the step size. However, considering the difficulty in multivariate quaternion analysis due to the non-commutativity of quaternion products and the lack of quaternion matrix factorisation methods, it is challenging to extend the analysis in [49] to general multivariate QLMS algorithms.

Based on the analysis framework for real-valued adaptive filters developed in [81, 82], this chapter analyses the mean and mean square convergence of general quaternion adaptive filters with general (proper or improper) quaternion inputs in non-stationary environments, and thus quantifies their transient and steady-state performance. The proposed analysis not only treats different adaptation approaches uniformly, but also caters for different linear models for quaternion signal estimation. In order to model a non-stationary environment, the system weights are assumed to vary according to a first-order random-walk model. By decorrelating the quaternion regressor vector, the analysis of the QLMS filters with Gaussian inputs yields closed-form solutions relating to the second-order statistics of input data. The analytical results of the strictly linear QLMS (SL-QLMS),

widely linear QLMS (WL-QLMS), and semi-widely linear QLMS (SWL-QLMS) filters are investigated in detail. The key contributions of our work are as follows: (i) a unified performance analysis for a general class of quaternion adaptive filters, encompassing various adaptation approaches and quaternion linear estimation models; (ii) the mean square performance of QLMS filters with Gaussian inputs is linked to the second-order statistics of quaternion inputs by representing a fourth-order quaternion moment matrix in terms of covariance and complementary covariance matrices; (iii) the application of a new decorrelation technique for random quaternion vectors to the Gaussian regressors of QLMS filters leads to concise and physically meaningful closed-form results. This work is therefore a non-trivial extension of the real-valued analysis [81, 82] to the quaternion domain. Compared to [81, 82], careful attention is paid to the non-commutative quaternion algebra, and quaternion analysis techniques are employed to deal with augmented quaternion statistics. Compared to the mean square analysis of univariate SL-QLMS in [49], a general class of multivariate quaternion adaptive filters are considered by using a generic filter form and quaternion multivariate statistics. The so established analytical results bring new insights into the behaviour of the quaternion adaptive filters.

After an introduction to a class of quaternion adaptive filters to be considered, this chapter presents a convergence analysis framework to obtain the bounds of stability conditions and steady-state performances of such filters, from which the performance of QLMS filters with Gaussian regressors is deduced. Then, the QAUT proposed in Chapter 3 is used to derive concise and intuitive closed-form analytical results, which are exemplified with the SL-QLMS, SWL-QLMS and WL-QLMS.

5.2 A Class of Quaternion Adaptive Filters in Non-stationary Environments

Consider the problem of estimating $y_n \in \mathbb{H}$ from $\mathbf{x}_n \in \mathbb{H}^{L \times 1}$. Given a general form of linear estimation model

$$\hat{y}_n = \hat{\mathbf{w}}_n^H \mathbf{s}_n \quad (5.1)$$

where $\mathbf{s}_n \in \mathbb{H}^{M \times 1}$ is the regressor vector and $\hat{\mathbf{w}}_n \in \mathbb{H}^{M \times 1}$ the weight vector, the three linear estimation models introduced in (2.18)-(2.20) arise from (5.1) as follows

- Strictly linear model:

$$\mathbf{s}_n = \mathbf{x}_n, \quad M = L \quad (5.2)$$

- Semi-widely linear model:

$$\mathbf{s}_n = \mathbf{x}_n^b, \quad M = 2L \quad (5.3)$$

- Widely linear model:

$$\mathbf{s}_n = \mathbf{x}_n^a, \quad M = 4L \quad (5.4)$$

Based on (5.1) and the stochastic gradient decent approach to the minimisation of MSE, a number of quaternion adaptive filtering algorithms fit the following form:

$$\begin{aligned} \hat{y}_n &= \hat{\mathbf{w}}_n^H \mathbf{s}_n \\ e_n &= y_n - \hat{y}_n \\ \hat{\mathbf{w}}_{n+1} &= \hat{\mathbf{w}}_n + \mu f(\mathbf{s}_n) \mathbf{s}_n e_n^* \end{aligned} \quad (5.5)$$

where μ is the fixed real-valued step size, $f(\mathbf{s}_n)$ a real-valued scalar function of $\mathbf{s}_n$, $f(\mathbf{s}_n) : \mathbb{H}^{M \times 1} \rightarrow \mathbb{R}$. Specific values of $f(\mathbf{s}_n)$ correspond to specific filtering algorithms. For example, $f(\mathbf{s}_n) = 1$ results in the QLMS, $f(\mathbf{s}_n) = \|\mathbf{s}_n\|^{-2}$ results in the normalised QLMS, and $f(\mathbf{s}_n) = (\epsilon + \|\mathbf{s}_n\|^2)^{-1}$ results in the ϵ -normalised

QLMS [84]. The case of $f(\mathbf{s}_n)$ being a matrix-valued linear or nonlinear function worth further research. Equation (5.5) is similar to the quaternion nonlinear gradient decent algorithm given by [33]

$$\begin{aligned}\hat{y}_n &= \varphi(\hat{\mathbf{w}}_n^H \mathbf{s}_n) \\ e_n &= y_n - \hat{y}_n \\ \hat{\mathbf{w}}_{n+1} &= \hat{\mathbf{w}}_n + \mu \mathbf{s}_n \sum_{\nu \in \{1, i, j, \kappa\}} \frac{\partial \varphi(q)^\nu}{\partial q} \Big|_{q=\hat{\mathbf{w}}_n^H \mathbf{s}_n} e_n^{\nu*}\end{aligned}\tag{5.6}$$

which aims to minimise the cost function, $|e_n|^2$, in a stochastic gradient decent manner, where $\varphi(\cdot)$ is a quaternion-valued nonlinear function. Note that $\sum_{\nu \in \{1, i, j, \kappa\}} \frac{\partial \varphi(q)^\nu}{\partial q} \Big|_{q=\hat{\mathbf{w}}_n^H \mathbf{s}_n}$ is an counterpart to $f(\mathbf{s}_n)$. In the reduced case of linear function, e.g., $\varphi(q) = q$, (5.6) degenerates into the QLMS.

The estimation model in (5.1) is actually based on the assumption that the desired signal arises from the following linear system:

$$y_n = \mathbf{w}_n^H \mathbf{s}_n + v_n\tag{5.7}$$

where the system noise v_n is i.i.d., zero-mean, independent of $\mathbf{s}_n$, and with variance σ_v^2 . For the desired signal in (5.7), the optimal weight in the MMSE estimation based on (5.1) can be calculated as

$$E\{\mathbf{s}_n \mathbf{s}_n^H\}^{-1} E\{\mathbf{s}_n y_n^*\} = E\{\mathbf{w}_n\}\tag{5.8}$$

which is time-varying. To explore the filter behaviour in non-stationary environments, the system weight vector $\mathbf{w}_n \in \mathbb{H}^{M \times 1}$ is assumed to vary according to a widely used first-order random-walk model given by [4, 5, 84, 85]

$$\mathbf{w}_n = \mathbf{w}_{n-1} + \mathbf{q}_n\tag{5.9}$$

where $\mathbf{q}_n \in \mathbb{H}^{M \times 1}$ is a random quaternion vector that is i.i.d, zero-mean and

independent of $\mathbf{s}_n$. The expression in (5.9) is an approximation of a Markov model given by [86]

$$\mathbf{w}_n = \mathbf{A}\mathbf{w}_{n-1} + \mathbf{q}_n \quad (5.10)$$

where the coefficient matrix $\mathbf{A}$ is assumed to be close to $\mathbf{I}_M$ in a number of practical applications. For example, in modelling Rayleigh fading channels in a wireless communications environment, (5.10) is a first-order approximation of the variation of fading coefficients, and $\mathbf{A}$ is usually close to $\mathbf{I}_M$ [5]. In order to simplify the derivation during the convergence analysis, the real-valued adaptive filtering literature customarily assumes $\mathbf{A} = \mathbf{I}_M$ to employ the first-order random-walk model in (5.9), which can be straightforwardly extended to the quaternion domain. Despite the simplified non-stationary scenarios studied here, the essence of the problem that is common to more complicated situations is retained.

5.3 Convergence Performance Analysis of Quaternion Adaptive Filters

To analyse the convergence performance of the filters in the form of (5.5), define the weight-error vector as $\tilde{\mathbf{w}}_n \triangleq \mathbf{w}_n - \hat{\mathbf{w}}_n$, and express the filter error as $e_n = \tilde{\mathbf{w}}_n^H \mathbf{s}_n + v_n$. Then the weight-error vector recursion is expressed as

$$\tilde{\mathbf{w}}_{n+1} = \tilde{\mathbf{w}}_n - \mu f(\mathbf{s}_n) \mathbf{s}_n e_n^* + \mathbf{q}_n = [\mathbf{I}_M - \mu f(\mathbf{s}_n) \mathbf{s}_n \mathbf{s}_n^H] \tilde{\mathbf{w}}_n - \mu f(\mathbf{s}_n) \mathbf{s}_n v_n^* + \mathbf{q}_n \quad (5.11)$$

5.3.1 Mean Analysis

Since v_n and $\mathbf{q}_n$ are i.i.d., zero-mean and independent of $\mathbf{s}_n$, the expectation of (5.11) is given by

$$E\{\tilde{\mathbf{w}}_{n+1}\} = E\{\mathbf{I}_M - \mu f(\mathbf{s}_n) \mathbf{s}_n \mathbf{s}_n^H\} E\{\tilde{\mathbf{w}}_n\} \quad (5.12)$$

A convenient change of coordinates is enabled by the eigendecomposition given by $E \{f(\mathbf{s}_n) \mathbf{s}_n \mathbf{s}_n^H\} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^H$, where $\mathbf{U}$ is a unitary quaternion matrix and $\mathbf{\Lambda}$ is a real diagonal matrix containing eigenvalues of $E \{f(\mathbf{s}_n) \mathbf{s}_n \mathbf{s}_n^H\}$ on the diagonal. Define $\mathbf{w}'_n = \mathbf{U}^H \tilde{\mathbf{w}}_n$. Then, (5.12) becomes

$$E \{\mathbf{w}'_{n+1}\} = E \{\mathbf{I}_M - \mu \mathbf{\Lambda}\} E \{\mathbf{w}'_n\} \quad (5.13)$$

from which one can observe that $\lim_{n \rightarrow \infty} E \{\mathbf{w}'_n\} = \mathbf{0}$ if and only if the eigenvalues of the matrix $E \{\mathbf{I}_M - \mu \mathbf{\Lambda}\}$ are all within $(-1, 1)$, or equivalently [4]

$$0 < \mu < 2\lambda_{\max} \left(E \{f(\mathbf{s}_n) \mathbf{s}_n \mathbf{s}_n^H\} \right)^{-1} \quad (5.14)$$

5.3.2 Mean Square Analysis

Define the *a priori* and *a posteriori* estimation errors as

$$e_{a,n} \triangleq \mathbf{w}_n^H \mathbf{s}_n - \hat{\mathbf{w}}_n^H \mathbf{s}_n = \tilde{\mathbf{w}}_n^H \mathbf{s}_n$$

$$e_{p,n} \triangleq \mathbf{w}_n^H \mathbf{s}_n - \hat{\mathbf{w}}_{n+1}^H \mathbf{s}_n = (\tilde{\mathbf{w}}_{n+1} - \mathbf{q}_n)^H \mathbf{s}_n$$

and the weighted *a priori* and *a posteriori* errors as $e_{a,n}^\Sigma \triangleq \tilde{\mathbf{w}}_n^H \mathbf{\Sigma} \mathbf{s}_n$ and $e_{p,n}^\Sigma \triangleq (\tilde{\mathbf{w}}_{n+1} - \mathbf{q}_n)^H \mathbf{\Sigma} \mathbf{s}_n$, where $\mathbf{\Sigma}$ is a quaternion Hermitian positive definite weighting matrix. For notation conciseness, denote $\mathbf{s}_n^H \mathbf{\Sigma} \mathbf{s}_n$ by $\|\mathbf{s}_n\|_\Sigma^2$. From (5.11) and

$$e_{p,n}^\Sigma = e_{a,n}^\Sigma - \mu f(\mathbf{s}_n) e_n \mathbf{s}_n^H \mathbf{\Sigma} \mathbf{s}_n = e_{a,n}^\Sigma - \mu f(\mathbf{s}_n) e_n \|\mathbf{s}_n\|_\Sigma^2 \quad (5.15)$$

it can be derived that

$$\tilde{\mathbf{w}}_{n+1} + \mathbf{s}_n \frac{(e_{a,n}^\Sigma)^*}{\|\mathbf{s}_n\|_\Sigma^2} = \tilde{\mathbf{w}}_n + \mathbf{s}_n \frac{(e_{p,n}^\Sigma)^*}{\|\mathbf{s}_n\|_\Sigma^2} + \mathbf{q}_n \quad (5.16)$$

Evaluating energies of both sides of (5.16) yields

$$\|\tilde{\mathbf{w}}_{n+1}\|_{\Sigma}^2 + \frac{|e_{a,n}^{\Sigma}|^2}{\|\mathbf{s}_n\|_{\Sigma}^2} = \|\tilde{\mathbf{w}}_n\|_{\Sigma}^2 + \frac{|e_{p,n}^{\Sigma}|^2}{\|\mathbf{s}_n\|_{\Sigma}^2} + \|\mathbf{q}_n\|_{\Sigma}^2 + 2\Re \left[\mathbf{q}_n^H \Sigma \left(\tilde{\mathbf{w}}_n + \mathbf{s}_n \frac{(e_{p,n}^{\Sigma})^*}{\|\mathbf{s}_n\|_{\Sigma}^2} \right) \right]$$

which can be rewritten by combining (5.15) as

$$\|\tilde{\mathbf{w}}_{n+1}\|_{\Sigma}^2 = \|\tilde{\mathbf{w}}_n\|_{\Sigma_1}^2 + \mu^2 |v_n|^2 f(\mathbf{s}_n)^2 \|\mathbf{s}_n\|_{\Sigma}^2 + \|\mathbf{q}_n\|_{\Sigma}^2 + 2\Re \left[\mathbf{q}_n^H \Sigma \left(\tilde{\mathbf{w}}_n + \mathbf{s}_n \frac{(e_{p,n}^{\Sigma})^*}{\|\mathbf{s}_n\|_{\Sigma}^2} \right) \right]$$

—

$$2\mu\Re \left[f(\mathbf{s}_n) (e_{a,n}^{\Sigma} - \mu f(\mathbf{s}_n) \|\mathbf{s}_n\|_{\Sigma}^2 e_{a,n}) v_n \right]$$

$$\Sigma_1 \triangleq \Sigma - \mu f(\mathbf{s}_n) (\mathbf{s}_n \mathbf{s}_n^H \Sigma + \Sigma \mathbf{s}_n \mathbf{s}_n^H) + \mu^2 f(\mathbf{s}_n)^2 \|\mathbf{s}_n\|_{\Sigma}^2 \mathbf{s}_n \mathbf{s}_n^H$$

The application of the statistical expectation operator to the above equation yields

$$E \{ \|\tilde{\mathbf{w}}_{n+1}\|_{\Sigma}^2 \} = E \{ \|\tilde{\mathbf{w}}_n\|_{\Sigma_1}^2 \} + E \{ \|\mathbf{q}_n\|_{\Sigma}^2 \} + \mu^2 \sigma_v^2 E \{ f(\mathbf{s}_n)^2 \|\mathbf{s}_n\|_{\Sigma}^2 \} \quad (5.17)$$

For simplicity, assume that the sequence of $\mathbf{s}_n$ is i.i.d.. Thus (5.17) becomes

$$E \{ \|\tilde{\mathbf{w}}_{n+1}\|_{\Sigma}^2 \} = E \{ \|\tilde{\mathbf{w}}_n\|_{\Sigma'}^2 \} + E \{ \|\mathbf{q}_n\|_{\Sigma}^2 \} + \mu^2 \sigma_v^2 E \{ f(\mathbf{s}_n)^2 \|\mathbf{s}_n\|_{\Sigma}^2 \} \quad (5.18)$$

$$\Sigma' \triangleq E \{ \Sigma_1 \} = \Sigma + \mu^2 E \{ f(\mathbf{s}_n)^2 \|\mathbf{s}_n\|_{\Sigma}^2 \mathbf{s}_n \mathbf{s}_n^H \} - \mu (E \{ f(\mathbf{s}_n) \mathbf{s}_n \mathbf{s}_n^H \} \Sigma + \Sigma E \{ f(\mathbf{s}_n) \mathbf{s}_n \mathbf{s}_n^H \})$$

Stability condition

Consider now the transient behaviour of $E \{ \|\tilde{\mathbf{w}}_n\|^2 \}$ by letting $\Sigma = \mathbf{I}_M$ in (5.18). Similar to the analysis in [81], matrix Σ' can be vectorised into a vector referred to as $\text{vec}(\Sigma')$ by stacking the columns of Σ' on top of one another. Refer to $\|\cdot\|_{\Sigma'}^2$ as $\|\cdot\|_{\text{vec}(\Sigma')}^2$ and define

$$\mathbf{E} \triangleq \mathbf{I}_{M^2} - \mu \mathbf{A} + \mu^2 \mathbf{B}$$

where

$$\mathbf{A} \triangleq \mathbf{I}_M \otimes E \{ f(\mathbf{s}_n) \mathbf{s}_n \mathbf{s}_n^H \} + (E \{ f(\mathbf{s}_n) \mathbf{s}_n \mathbf{s}_n^H \})^T \otimes \mathbf{I}_M$$

$$\mathbf{B} \triangleq E \left\{ f(\mathbf{s}_n)^2 (\mathbf{s}_n \mathbf{s}_n^H)^T \otimes_{\mathbf{R}} (\mathbf{s}_n \mathbf{s}_n^H) \right\}$$

in which “ $\otimes$ ” denotes the left Kronecker product

$$\mathbf{A} \otimes \mathbf{B} \triangleq \begin{bmatrix} [\mathbf{A}]_{1,1} \mathbf{B} & \dots & [\mathbf{A}]_{1,M} \mathbf{B} \\ \vdots & \ddots & \vdots \\ [\mathbf{A}]_{M,1} \mathbf{B} & \dots & [\mathbf{A}]_{M,M} \mathbf{B} \end{bmatrix}$$

and “ $\otimes_{\mathbf{R}}$ ” the right Kronecker product [78]

$$\mathbf{A} \otimes_{\mathbf{R}} \mathbf{B} \triangleq \begin{bmatrix} \mathbf{B} [\mathbf{A}]_{1,1} & \dots & \mathbf{B} [\mathbf{A}]_{1,M} \\ \vdots & \ddots & \vdots \\ \mathbf{B} [\mathbf{A}]_{M,1} & \dots & \mathbf{B} [\mathbf{A}]_{M,M} \end{bmatrix}$$

Then, (5.18) can be rewritten as

$$E \left\{ \|\tilde{\mathbf{w}}_{n+1}\|^2 \right\} = E \left\{ \|\tilde{\mathbf{w}}_n\|_{\mathbf{E} \text{vec}(\mathbf{I}_M)}^2 \right\} + E \left\{ \|\mathbf{q}_n\|^2 \right\} + \mu^2 \sigma_v^2 E \left\{ f(\mathbf{s}_n)^2 \|\mathbf{s}_n\|^2 \right\} \quad (5.19)$$

Notice that the eigenvalues of $\mathbf{E}$ are all real-valued as $\mathbf{E}$ is a Hermitian quaternion matrix [13].

Based on the Cayley-Hamilton theorem, [81] has proved that (5.19) is stable for real variables if and only if all eigenvalues of $\mathbf{E}$ are within $(-1, 1)$, which is equivalent to

$$0 < \mu < \min \left\{ \lambda_{\max} (\mathbf{A}^{-1} \mathbf{B})^{-1}, \max \left\{ \lambda \left(\begin{bmatrix} \mathbf{A}/2 & -\mathbf{B}/2 \\ \mathbf{I}_{M^2} & \mathbf{0} \end{bmatrix} \right) \in \mathbb{R}^+ \right\}^{-1} \right\} \quad (5.20)$$

This conclusion also holds for quaternion variables since the Cayley-Hamilton theorem applies to the quaternion domain [12], implying that (5.20) is also the mean square stability condition of quaternion adaptive filters. In contrast to real and complex analyses [5, 7], the steady-state performance of quaternion adaptive filters is difficult to quantify in a closed form via the above analysis approach, since the re-

quired vectorisation of quaternion matrices is hindered by the non-commutativity of quaternion products.

Performance bounds

The sufficient and necessary stability condition shown in (5.20) is difficult to interpret owing to the complicated structure of matrices $\mathbf{A}$ and $\mathbf{B}$. Next, simple sufficient and necessary conditions for the mean square stability shall be derived. Set $\Sigma = \mathbf{I}_M$, then the evolution of $E \{ \|\tilde{\mathbf{w}}_{n+1}\|^2 \}$ can be deduced from (5.18) as

$$E \{ \|\tilde{\mathbf{w}}_{n+1}\|^2 \} = E \{ \|\tilde{\mathbf{w}}_n\|_{\Sigma'}^2 \} + E \{ \|\mathbf{q}_n\|^2 \} + \mu^2 \sigma_v^2 E \{ f(\mathbf{s}_n)^2 \|\mathbf{s}_n\|^2 \} \quad (5.21)$$

where

$$\Sigma' = \mathbf{I}_M - 2\mu E \{ f(\mathbf{s}_n) \mathbf{s}_n \mathbf{s}_n^H \} + \mu^2 E \{ f(\mathbf{s}_n)^2 \|\mathbf{s}_n\|^2 \mathbf{s}_n \mathbf{s}_n^H \}$$

Notice that the Cauchy–Schwarz inequality can be applied to the Hermitian quaternion matrix $E \{ f(\mathbf{s}_n)^2 \|\mathbf{s}_n\|^2 \mathbf{s}_n \mathbf{s}_n^H \}$ [87], yielding

$$\Sigma' \succeq \mathbf{I}_M - 2\mu E \{ f(\mathbf{s}_n) \mathbf{s}_n \mathbf{s}_n^H \} + \mu^2 E \{ f(\mathbf{s}_n) \mathbf{s}_n \mathbf{s}_n^H \}^2 = (\mathbf{I}_M - \mu E \{ f(\mathbf{s}_n) \mathbf{s}_n \mathbf{s}_n^H \})^2 \quad (5.22)$$

Observe that Σ' is positive semi-definite, and its eigendecomposition is given by $\Sigma' = \mathbf{T} \Lambda_{\Sigma'} \mathbf{T}^H$, where $\mathbf{T}$ is a quaternion unitary matrix, and $\Lambda_{\Sigma'}$ a real diagonal matrix with non-negative eigenvalues of Σ' on the diagonal. Thus,

$$E \{ \|\tilde{\mathbf{w}}_n\|_{\Sigma'}^2 \} = E \{ \tilde{\mathbf{w}}_n^H \mathbf{T} \Lambda_{\Sigma'} \mathbf{T}^H \tilde{\mathbf{w}}_n \} \quad (5.23)$$

From (5.23), observe

$$E \{ \|\tilde{\mathbf{w}}_n\|_{\Sigma'}^2 \} \leq \lambda_{\max}(\Sigma') E \{ \|\tilde{\mathbf{w}}_n\|^2 \} \quad (5.24)$$

where $\lambda_{\max}(\Sigma')$ is bounded by

$$\lambda_{\max}(\Sigma') \leq \xi_1 \triangleq 1 - 2\mu\lambda_{\min}(E\{f(\mathbf{s}_n)\mathbf{s}_n\mathbf{s}_n^H\}) + \mu^2\lambda_{\max}(E\{f(\mathbf{s}_n)^2\|\mathbf{s}_n\|^2\mathbf{s}_n\mathbf{s}_n^H\})$$

which is based on (2.12). This inequality, together with (5.21) and (5.24), yields

$$E\{\|\tilde{\mathbf{w}}_{n+1}\|^2\} \leq \xi_1 E\{\|\tilde{\mathbf{w}}_n\|^2\} + E\{\|\mathbf{q}_n\|^2\} + \mu^2\sigma_v^2 E\{f(\mathbf{s}_n)^2\|\mathbf{s}_n\|^2\} \quad (5.25)$$

which is stable if $\xi_1 < 1$, thus implying the following sufficient condition for the mean square stability

$$0 < \mu < \frac{2\lambda_{\min}(E\{f(\mathbf{s}_n)\mathbf{s}_n\mathbf{s}_n^H\})}{\lambda_{\max}(E\{f(\mathbf{s}_n)^2\|\mathbf{s}_n\|^2\mathbf{s}_n\mathbf{s}_n^H\})} \quad (5.26)$$

From (5.23), observe

$$E\{\|\tilde{\mathbf{w}}_n\|_{\Sigma'}^2\} \geq \lambda_{\min}(\Sigma') E\{\|\tilde{\mathbf{w}}_n\|^2\} \quad (5.27)$$

where $\lambda_{\min}(\Sigma')$ is bounded by

$$\lambda_{\min}(\Sigma') \geq \xi_2 \triangleq 1 - 2\mu\lambda_{\max}(E\{f(\mathbf{s}_n)\mathbf{s}_n\mathbf{s}_n^H\}) + \mu^2\lambda_{\min}(E\{f(\mathbf{s}_n)^2\|\mathbf{s}_n\|^2\mathbf{s}_n\mathbf{s}_n^H\})$$

This inequality, together with (5.21) and (5.27), yields

$$E\{\|\tilde{\mathbf{w}}_{n+1}\|^2\} \geq \xi_2 E\{\|\tilde{\mathbf{w}}_n\|^2\} + E\{\|\mathbf{q}_n\|^2\} + \mu^2\sigma_v^2 E\{f(\mathbf{s}_n)^2\|\mathbf{s}_n\|^2\} \quad (5.28)$$

Thus, a necessary condition for the convergence of $E\{\|\tilde{\mathbf{w}}_n\|^2\}$ is $\xi_2 < 1$, or equivalently

$$0 < \mu < \frac{2\lambda_{\max}(E\{f(\mathbf{s}_n)\mathbf{s}_n\mathbf{s}_n^H\})}{\lambda_{\min}(E\{f(\mathbf{s}_n)^2\|\mathbf{s}_n\|^2\mathbf{s}_n\mathbf{s}_n^H\})} \quad (5.29)$$

From (5.22), another necessary condition for the convergence of $E\{\|\tilde{\mathbf{w}}_n\|^2\}$ is that the eigenvalues of the matrix $(\mathbf{I}_M - \mu E\{f(\mathbf{s}_n)\mathbf{s}_n\mathbf{s}_n^H\})^2$ are all less than

unit, which yields the same step size bound as in (5.14).

Remark 5.1. Any value of μ that satisfies the mean square stability condition given in (5.20) also guarantees the mean stability.

Next, the steady-state performance of quaternion adaptive filters will be bounded through a similar approach. At the steady state, $\lim_{n \rightarrow \infty} E \{ \|\tilde{\mathbf{w}}_{n+1}\|^2 \} = \lim_{n \rightarrow \infty} E \{ \|\tilde{\mathbf{w}}_n\|^2 \}$, and thus the bounds of $\text{MSD} = \lim_{n \rightarrow \infty} E \{ \|\tilde{\mathbf{w}}_n\|^2 \}$ can be derived from (5.25) and (5.28) as

$$\begin{aligned} \text{MSD} &\leq \frac{\mu^2 \sigma_v^2 \text{Tr}(\mathbf{C}_{\text{ss}}) + \text{Tr}(\mathbf{C}_{\text{qq}})}{2\mu \lambda_{\min}(E\{f(\mathbf{s}_n) \mathbf{s}_n \mathbf{s}_n^H\}) - \mu^2 \lambda_{\max}(E\{f(\mathbf{s}_n)^2 \|\mathbf{s}_n\|^2 \mathbf{s}_n \mathbf{s}_n^H\})} \\ \text{MSD} &\geq \frac{\mu^2 \sigma_v^2 \text{Tr}(\mathbf{C}_{\text{ss}}) + \text{Tr}(\mathbf{C}_{\text{qq}})}{2\mu \lambda_{\max}(E\{f(\mathbf{s}_n) \mathbf{s}_n \mathbf{s}_n^H\}) - \mu^2 \lambda_{\min}(E\{f(\mathbf{s}_n)^2 \|\mathbf{s}_n\|^2 \mathbf{s}_n \mathbf{s}_n^H\})} \end{aligned} \quad (5.30)$$

Since

$$E \{ \|\tilde{\mathbf{w}}_n\|_{\mathbf{C}_{\text{ss}}}^2 \} = E \{ \tilde{\mathbf{w}}_n^H \mathbf{K} \mathbf{\Lambda}_{\text{ss}} \mathbf{K}^H \tilde{\mathbf{w}}_n \}$$

where $\mathbf{K} \mathbf{\Lambda}_{\text{ss}} \mathbf{K}^H$ is the eigendecomposition of $\mathbf{C}_{\text{ss}}$, it follows that

$$\lambda_{\min}(\mathbf{C}_{\text{ss}}) E \{ \|\tilde{\mathbf{w}}_n\|^2 \} \leq E \{ \|\tilde{\mathbf{w}}_n\|_{\mathbf{C}_{\text{ss}}}^2 \} \leq \lambda_{\max}(\mathbf{C}_{\text{ss}}) E \{ \|\tilde{\mathbf{w}}_n\|^2 \}$$

so that $\text{EMSE} = \lim_{n \rightarrow \infty} E \{ \|\tilde{\mathbf{w}}_n\|_{\mathbf{C}_{\text{ss}}}^2 \}$ is bounded by

$$\lambda_{\min}(\mathbf{C}_{\text{ss}}) \cdot \text{MSD} \leq \text{EMSE} \leq \lambda_{\max}(\mathbf{C}_{\text{ss}}) \cdot \text{MSD} \quad (5.31)$$

5.3.3 Performance Analysis of QLMS

As mentioned in Section 5.2, when $f(\mathbf{s}_n) = 1$, the general filter form in (5.5) reduces to the QLMS. Therefore, the performance of QLMS filters can be obtained by substituting $f(\mathbf{s}_n) = 1$ into the above analysis results for general quaternion adaptive filters. Specifically, the mean weight-error vector recursion in (5.12) becomes

$$E \{ \tilde{\mathbf{w}}_{n+1} \} = (\mathbf{I}_M - \mu \mathbf{C}_{\text{ss}}) E \{ \tilde{\mathbf{w}}_n \} \quad (5.32)$$

while the mean stability condition in (5.14) reduces to

$$0 < \mu < 2\lambda_{\max}^{-1}(\mathbf{C}_{\mathbf{ss}}) \quad (5.33)$$

which is the same as the mean stability condition obtained in [83].

The mean square stability condition in (5.20), which is also the overall stability condition, becomes

$$0 < \mu < \min \left\{ \lambda_{\max} \left(\bar{\mathbf{A}}^{-1} \bar{\mathbf{B}} \right)^{-1}, \max \left\{ \lambda \left(\begin{bmatrix} \bar{\mathbf{A}}/2 & -\bar{\mathbf{B}}/2 \\ \mathbf{I}_{M^2} & \mathbf{0} \end{bmatrix} \right) \in \mathbb{R}^+ \right\}^{-1} \right\}$$

$$\bar{\mathbf{A}} \triangleq \mathbf{I}_M \otimes \mathbf{C}_{\mathbf{ss}} + \mathbf{C}_{\mathbf{ss}}^T \otimes \mathbf{I}_M$$

$$\bar{\mathbf{B}} \triangleq E \left\{ (\mathbf{s}_n \mathbf{s}_n^H)^T \otimes_{\mathbf{R}} (\mathbf{s}_n \mathbf{s}_n^H) \right\} \quad (5.34)$$

Furthermore, if the regressor vector $\mathbf{s}_n$ arises from a Gaussian distribution, the analysis results shown in (5.26), (5.29), (5.30) and (5.31) can be simplified by expressing the fourth-order moment matrix $E \{ \|\mathbf{s}_n\|^2 \mathbf{s}_n \mathbf{s}_n^H \}$ in terms of covariance and complementary covariance matrices of $\mathbf{s}_n$ as

$$E \{ \|\mathbf{s}_n\|^2 \mathbf{s}_n \mathbf{s}_n^H \} = \text{Tr}(\mathbf{C}_{\mathbf{ss}}) \mathbf{C}_{\mathbf{ss}} + \frac{1}{2} \left(\mathbf{C}_{\mathbf{ss}}^2 + \sum_{\alpha=\iota, j, \kappa} \mathbf{C}_{\mathbf{ss}^\alpha} \mathbf{C}_{\mathbf{ss}^\alpha}^H \right) \quad (5.35)$$

which can be proved as follows.

Represent the quaternion variables using the Cayley-Dickson construction, that is, the m -th element of $\mathbf{s}_n$ is written as $s_m = c_m + d_m j$, where c_m and d_m are complex variables defined on the basis $\{1, \iota\}$ [42], and then apply Isserlis' theorem [88] to the constitutive complex variables. This leads to the element on the a -th row and b -th column of the fourth-order moment matrix $E \{ \|\mathbf{s}_n\|^2 \mathbf{s}_n \mathbf{s}_n^H \}$

as

$$\begin{aligned}
& \left[E \left\{ \|\mathbf{s}_n\|^2 \mathbf{s}_n \mathbf{s}_n^H \right\} \right]_{a,b} \\
&= E \left\{ \sum_{m=1}^M s_m s_m^* s_a s_b^* \right\} \\
&= E \left\{ \sum_{m=1}^M (c_m + d_m j) (c_m^* - d_m j) (c_a + d_a j) (c_b^* - d_b j) \right\} \\
&= \sum_{m=1}^M (E \{ c_m c_m^* c_a c_b^* + c_m c_m^* d_a d_b^* + d_m d_m^* c_a c_b^* + d_m d_m^* d_a d_b^* \} \\
&\quad + E \{ c_m c_m^* d_a c_b + d_m d_m^* d_a c_b - c_m c_m^* c_a d_b - d_m d_m^* c_a d_b \} j) \\
&= \sum_{m=1}^M E \{ c_m c_m^* + d_m d_m^* \} E \{ c_a c_b^* + d_a d_b^* + (d_a c_b - c_a d_b) j \} \\
&\quad + \sum_{m=1}^M (E \{ c_m c_a \} E \{ c_m^* c_b^* \} + E \{ c_m c_b^* \} E \{ c_m^* c_a \} + \\
&\quad E \{ c_m d_a \} E \{ c_m^* d_b^* \} + E \{ c_m d_b^* \} E \{ c_m^* d_a \} + E \{ d_m c_a \} E \{ d_m^* c_b^* \} + \\
&\quad E \{ d_m c_b^* \} E \{ d_m^* c_a \} + E \{ d_m d_a \} E \{ d_m^* d_b^* \} + E \{ d_m d_b^* \} E \{ d_m^* d_a \}) \\
&\quad + \sum_{m=1}^M (E \{ c_m d_a \} E \{ c_m^* c_b \} + E \{ c_m c_b \} E \{ c_m^* d_a \} + \\
&\quad E \{ d_m d_a \} E \{ d_m^* c_b \} + E \{ d_m c_b \} E \{ d_m^* d_a \} - E \{ c_m c_a \} E \{ c_m^* d_b \} - \\
&\quad E \{ c_m d_b \} E \{ c_m^* d_a \} - E \{ d_m c_a \} E \{ d_m^* d_b \} - E \{ d_m d_b \} E \{ d_m^* c_a \}) j
\end{aligned}$$

Observe that

$$\begin{aligned}
[\mathbf{C}_{\mathbf{ss}}]_{a,b} &= E \{ s_a s_b^* \} = E \{ (c_a + d_a j) (c_b^* - d_b j) \} \\
&= E \{ c_a c_b^* + d_a d_b^* \} + E \{ d_a c_b - c_a d_b \} j
\end{aligned}$$

$$\begin{aligned}
[\mathbf{C}_{\mathbf{ss}^i}]_{a,b} &= E \{ s_a s_b^{i*} \} = E \{ (c_a + d_a j) (c_b^* + d_b j) \} \\
&= E \{ c_a c_b^* - d_a d_b^* \} + E \{ d_a c_b + c_a d_b \} j
\end{aligned}$$

$$\begin{aligned}
[\mathbf{C}_{\mathbf{ss}^j}]_{a,b} &= E \{ s_a s_b^{j*} \} = E \{ (c_a + d_a j) (c_b - d_b^* j) \} \\
&= E \{ c_a c_b + d_a d_b \} + E \{ d_a c_b^* - c_a d_b^* \} j
\end{aligned}$$

$$\begin{aligned}
[\mathbf{C}_{\mathbf{ss}^\kappa}]_{a,b} &= E \{ s_a s_b^{\kappa*} \} = E \{ (c_a + d_a j) (c_b - d_b^* j) \} \\
&= E \{ c_a c_b - d_a d_b \} + E \{ d_a c_b^* + c_a d_b^* \} j
\end{aligned}$$

From the above equations, one can verify

$$[E \{ \|\mathbf{s}_n\|^2 \mathbf{s}_n \mathbf{s}_n^H \}]_{a,b} = \text{Tr}(\mathbf{C}_{\text{ss}}) [\mathbf{C}_{\text{ss}}]_{a,b} + \frac{1}{2} [\mathbf{C}_{\text{ss}}^2]_{a,b} + \frac{1}{2} \sum_{\alpha=\iota, j, \kappa} [\mathbf{C}_{\text{ss}^\alpha} \mathbf{C}_{\text{ss}^\alpha}^\alpha]_{a,b}$$

Therefore, (5.35) holds.

It follows (5.35) that

$$\lambda_{\max}(E \{ \|\mathbf{s}_n\|^2 \mathbf{s}_n \mathbf{s}_n^H \}) \leq \text{Tr}(\mathbf{C}_{\text{ss}}) \lambda_{\max}(\mathbf{C}_{\text{ss}}) + \frac{1}{2} \left[\lambda_{\max}^2(\mathbf{C}_{\text{ss}}) + \sum_{\alpha=\iota, j, \kappa} p_{\max}^2(\mathbf{C}_{\text{ss}^\alpha}) \right]$$

$$\lambda_{\min}(E \{ \|\mathbf{s}_n\|^2 \mathbf{s}_n \mathbf{s}_n^H \}) \geq \text{Tr}(\mathbf{C}_{\text{ss}}) \lambda_{\min}(\mathbf{C}_{\text{ss}}) + \frac{1}{2} \left[\lambda_{\min}^2(\mathbf{C}_{\text{ss}}) + \sum_{\alpha=\iota, j, \kappa} p_{\min}^2(\mathbf{C}_{\text{ss}^\alpha}) \right]$$

where $p_{\max}(\mathbf{C}_{\text{ss}^\alpha})$ and $p_{\min}(\mathbf{C}_{\text{ss}^\alpha})$ are maximum and minimum singular values of $\mathbf{C}_{\text{ss}^\alpha}$ respectively. A sufficient stability condition is deduced from (5.26) as

$$0 < \mu < \frac{4\lambda_{\min}(\mathbf{C}_{\text{ss}})}{2\text{Tr}(\mathbf{C}_{\text{ss}}) \lambda_{\max}(\mathbf{C}_{\text{ss}}) + \lambda_{\max}^2(\mathbf{C}_{\text{ss}}) + \sum_{\alpha=\iota, j, \kappa} p_{\max}^2(\mathbf{C}_{\text{ss}^\alpha})} \quad (5.36)$$

while a necessary stability condition is deduced from (5.29) as

$$0 < \mu < \frac{4\lambda_{\max}(\mathbf{C}_{\text{ss}})}{2\text{Tr}(\mathbf{C}_{\text{ss}}) \lambda_{\min}(\mathbf{C}_{\text{ss}}) + \lambda_{\min}^2(\mathbf{C}_{\text{ss}}) + \sum_{\alpha=\iota, j, \kappa} p_{\min}^2(\mathbf{C}_{\text{ss}^\alpha})} \quad (5.37)$$

The upper and lower bounds of MSD are then deduced from (5.30), as in

$$\begin{aligned} \text{MSD} &\leq \frac{2\mu^2 \sigma_v^2 \text{Tr}(\mathbf{C}_{\text{ss}}) + 2\text{Tr}(\mathbf{C}_{\text{qq}})}{4\mu \lambda_{\min}(\mathbf{C}_{\text{ss}}) - \mu^2 \left[2\text{Tr}(\mathbf{C}_{\text{ss}}) \lambda_{\max}(\mathbf{C}_{\text{ss}}) + \lambda_{\max}^2(\mathbf{C}_{\text{ss}}) + \sum_{\alpha=\iota, j, \kappa} p_{\max}^2(\mathbf{C}_{\text{ss}^\alpha}) \right]} \\ \text{MSD} &\geq \frac{2\mu^2 \sigma_v^2 \text{Tr}(\mathbf{C}_{\text{ss}}) + 2\text{Tr}(\mathbf{C}_{\text{qq}})}{4\mu \lambda_{\max}(\mathbf{C}_{\text{ss}}) - \mu^2 \left[2\text{Tr}(\mathbf{C}_{\text{ss}}) \lambda_{\min}(\mathbf{C}_{\text{ss}}) + \lambda_{\min}^2(\mathbf{C}_{\text{ss}}) + \sum_{\alpha=\iota, j, \kappa} p_{\min}^2(\mathbf{C}_{\text{ss}^\alpha}) \right]} \end{aligned} \quad (5.38)$$

The upper and lower bounds of EMSE can be calculated based on the MSD bounds in (5.38), as in (5.31).

5.4 Mean Square Analysis of QLMS with Gaussian Regressors Using QAUT

As shown in Section 5.3.2 and Section 5.3.3, unlike the analysis for real and complex adaptive filters [5, 7], the mean square performance of quaternion adaptive filters is difficult to evaluate exactly in a closed form, owing to the non-commutative quaternion algebra and the lack of structural quaternion matrix decomposition methods. To obtain a concise closed-form solution to the mean square convergence equation, (5.18), for QLMS with Gaussian regressors, the structures of the involved quaternion covariance matrices can be simplified by using the following assumption.

Assumption 5.1. There exists a unitary transform $\bar{\mathbf{s}}_n = \mathbf{G}^H \mathbf{s}_n$ such that

$$\mathbf{C}_{\bar{\mathbf{s}}\bar{\mathbf{s}}} = \mathbf{\Lambda}_{\mathbf{s}\mathbf{s}}, \quad \mathbf{C}_{\bar{\mathbf{s}}\bar{\mathbf{s}}^\alpha} = \mathbf{\Lambda}_{\mathbf{s}\mathbf{s}^\alpha}, \quad \alpha = \imath, \jmath, \kappa \quad (5.39)$$

where $\mathbf{\Lambda}_{\mathbf{s}\mathbf{s}}$ is a diagonal real matrix containing the eigenvalues of $\mathbf{C}_{\mathbf{s}\mathbf{s}}$ on the diagonal, and $\mathbf{\Lambda}_{\mathbf{s}\mathbf{s}^\alpha}$ are diagonal quaternion matrices.

It can be verified that Assumption 5.1 holds exactly when one of the following conditions on $\mathbf{s}_n$ is met:

1. The covariance and complementary covariance matrices of $\mathbf{s}_n$ are related by scalar multiplication. This condition holds for regressor vectors with specific second-order statistics, such as an $\mathbb{H}$ -proper regressor vector ($\mathbf{C}_{\mathbf{s}\mathbf{s}^\alpha} = 0\mathbf{C}_{\mathbf{s}\mathbf{s}}$) and a rectilinear regressor vector ($\mathbf{C}_{\mathbf{s}\mathbf{s}^\alpha} = \mathbf{C}_{\mathbf{s}\mathbf{s}}\rho_\alpha$, where ρ_α is a unit quaternion).
2. $\mathbf{s}_n$ can be expressed by applying a unitary transform to a white Gaussian random quaternion vector $\mathbf{v}_n$, that is, $\mathbf{s}_n = \mathbf{G}\mathbf{v}_n$, where $\mathbf{G}$ is a unitary quaternion matrix.

For quaternion inputs that do not meet the above two conditions, Assumption 5.1 holds approximately as the result of the QAUT. Under this assumption, define $\bar{\mathbf{w}}_n = \mathbf{G}^H \tilde{\mathbf{w}}_n$, $\bar{\mathbf{q}}_n = \mathbf{G}^H \mathbf{q}_n$, $\bar{\Sigma} = \mathbf{G}^H \Sigma \mathbf{G}$ and convert (5.18) into

$$\begin{aligned} E \{ \|\bar{\mathbf{w}}_{n+1}\|_{\bar{\Sigma}}^2 \} &= E \{ \|\bar{\mathbf{w}}_n\|_{\bar{\Sigma}'}^2 \} + \mu^2 \sigma_v^2 E \{ \|\bar{\mathbf{s}}_n\|_{\bar{\Sigma}}^2 \} + E \{ \|\mathbf{q}_n\|_{\Sigma}^2 \} \\ \bar{\Sigma}' &= \bar{\Sigma} - \mu (E \{ \bar{\mathbf{s}}_n \bar{\mathbf{s}}_n^H \} \bar{\Sigma} + \bar{\Sigma} E \{ \bar{\mathbf{s}}_n \bar{\mathbf{s}}_n^H \}) + \mu^2 E \{ \|\bar{\mathbf{s}}_n\|_{\bar{\Sigma}}^2 \bar{\mathbf{s}}_n \bar{\mathbf{s}}_n^H \} \end{aligned} \quad (5.40)$$

where

$$E \{ \|\bar{\mathbf{s}}_n\|_{\bar{\Sigma}}^2 \bar{\mathbf{s}}_n \bar{\mathbf{s}}_n^H \} = \text{Tr}(\bar{\Sigma} \Lambda_{\text{ss}}) \Lambda_{\text{ss}} + \frac{1}{2} \bar{\Sigma} \left(\Lambda_{\text{ss}}^2 + \sum_{\alpha=\iota, j, \kappa} \Lambda_{\text{ss}^\alpha} \Lambda_{\text{ss}^\alpha}^* \right) \quad (5.41)$$

is derived similarly to the derivation of (5.35). Let $\bar{\Sigma}$ be a diagonal real matrix, then $\bar{\Sigma}'$ is also diagonal and real, and hence these two matrices can be completely characterised by their diagonal entries. Define $\bar{\sigma} \triangleq \text{diag}(\bar{\Sigma})$, $\bar{\sigma}' \triangleq \text{diag}(\bar{\Sigma}')$, $\lambda_s \triangleq \text{diag}(\Lambda_{\text{ss}})$, $\lambda_q \triangleq \text{diag}(E \{ \bar{\mathbf{q}}_n \bar{\mathbf{q}}_n^H \})$, and refer to $\|\cdot\|_{\bar{\Sigma}}$ and $\|\cdot\|_{\bar{\Sigma}'}^2$ as $\|\cdot\|_{\bar{\sigma}}$ and $\|\cdot\|_{\bar{\sigma}'}^2$. The compact forms of (5.40) and (5.41) can be obtained as

$$E \{ \|\bar{\mathbf{w}}_{n+1}\|_{\bar{\sigma}}^2 \} = E \{ \|\bar{\mathbf{w}}_n\|_{\mathbf{E}\bar{\sigma}}^2 \} + (\mu^2 \sigma_v^2 \lambda_s + \lambda_q)^T \bar{\sigma} \quad (5.42)$$

$$\mathbf{E} = \mathbf{I}_M - \mu \mathbf{A} + \mu^2 \mathbf{B}$$

$$\mathbf{A} = 2\Lambda_{\text{ss}}$$

$$\mathbf{B} = \frac{1}{2} \left(\Lambda_{\text{ss}}^2 + \sum_{\alpha=\iota, j, \kappa} \Lambda_{\text{ss}^\alpha} \Lambda_{\text{ss}^\alpha}^* \right) + \lambda_s \lambda_s^T$$

Stability condition

Similar to the stability condition of (5.20), (5.42) is stable if and only if all eigenvalues of $\mathbf{E}$ are within $(-1, 1)$. These eigenvalues are easily seen to be always greater than -1. It can be proved that they are upper bounded by unit if and only

if $0 < \mu < \lambda_{\max}(\mathbf{A}^{-1}\mathbf{B})^{-1}$, which is verified to be equivalent to

$$0 < \sum_{m=1}^M \frac{2\mu\lambda_{\mathbf{s},m}}{4 - \mu\lambda_{\mathbf{s},m} \left(1 + \sum_{\alpha=i,j,\kappa} r_{\bar{\mathbf{s}}\alpha,m}^2 \right)} < 1 \quad (5.43)$$

where $\lambda_{\mathbf{s},m}$ is the m -th element of $\boldsymbol{\lambda}_{\mathbf{s}}$, and $r_{\bar{\mathbf{s}}\alpha,m}$ is the α -impropriety coefficient of the m -th element of $\bar{\mathbf{s}}_n$. According to Remark 5.1, the condition in (5.43) is sufficient and necessary for the algorithm stability both in the mean and mean square sense.

Remark 5.2. The stability condition in (5.43) indicates that the upper bound of μ decreases with the increase in the length of the regressor vector, M , the increase in the power of the regressor reflected by $\lambda_{\mathbf{s},m}$, and the increase in the impropriety degree of $\bar{\mathbf{s}}_n$ reflected by $r_{\bar{\mathbf{s}}i,m}, r_{\bar{\mathbf{s}}j,m}, r_{\bar{\mathbf{s}}\kappa,m}$.

MSD and EMSE

Iterating (5.42) gives

$$E \{ \|\bar{\mathbf{w}}_n\|_{\bar{\boldsymbol{\sigma}}}^2 \} = E \{ \|\bar{\mathbf{w}}_0\|_{\mathbf{E}^n \bar{\boldsymbol{\sigma}}}^2 \} + (\mu^2 \sigma_v^2 \boldsymbol{\lambda}_{\mathbf{s}} + \boldsymbol{\lambda}_{\mathbf{q}})^T \sum_{l=0}^{n-1} \mathbf{E}^l \bar{\boldsymbol{\sigma}} \quad (5.44)$$

The MSD and EMSE at iteration n can now be derived from (5.44) by letting $\bar{\boldsymbol{\sigma}} = \mathbf{1}_M$ and $\bar{\boldsymbol{\sigma}} = \boldsymbol{\lambda}_{\mathbf{s}}$, respectively. Equation (5.44) also yields the iteration

$$E \{ \|\bar{\mathbf{w}}_{n+1}\|_{\bar{\boldsymbol{\sigma}}}^2 \} \neq E \{ \|\bar{\mathbf{w}}_n\|_{\bar{\boldsymbol{\sigma}}}^2 \} + E \left\{ \|\bar{\mathbf{w}}_0\|_{\mathbf{E}^n(\mathbf{E} - \mathbf{I}_M) \bar{\boldsymbol{\sigma}}}^2 \right\} + (\mu^2 \sigma_v^2 \boldsymbol{\lambda}_{\mathbf{s}} + \boldsymbol{\lambda}_{\mathbf{q}})^T \mathbf{E}^n \bar{\boldsymbol{\sigma}} \quad (5.45)$$

which characterises the evolution of MSD when $\bar{\boldsymbol{\sigma}} = \mathbf{1}_M$, and characterises the evolution of EMSE when $\bar{\boldsymbol{\sigma}} = \boldsymbol{\lambda}_{\mathbf{s}}$.

At the steady state, $\lim_{n \rightarrow \infty} E \{ \|\bar{\mathbf{w}}_{n+1}\|_{\bar{\boldsymbol{\sigma}}}^2 \} = \lim_{n \rightarrow \infty} E \{ \|\bar{\mathbf{w}}_n\|_{\bar{\boldsymbol{\sigma}}}^2 \}$, so that (5.42) yields

$$\lim_{n \rightarrow \infty} E \left\{ \|\bar{\mathbf{w}}_n\|_{(\mu\mathbf{A} - \mu^2\mathbf{B})\bar{\boldsymbol{\sigma}}}^2 \right\} = (\mu^2 \sigma_v^2 \boldsymbol{\lambda}_{\mathbf{s}} + \boldsymbol{\lambda}_{\mathbf{q}})^T \bar{\boldsymbol{\sigma}}$$

Finally, the steady-state MSD and EMSE become

$$\begin{aligned}\text{MSD} &= \lim_{n \rightarrow \infty} E \left\{ \|\bar{\mathbf{w}}_n\|_{\mathbf{I}_M}^2 \right\} = (\mu^2 \sigma_v^2 \boldsymbol{\lambda}_s + \boldsymbol{\lambda}_q)^T (\mu \mathbf{A} - \mu^2 \mathbf{B})^{-1} \mathbf{1}_M \\ \text{EMSE} &= \lim_{n \rightarrow \infty} E \left\{ \|\bar{\mathbf{w}}_n\|_{\boldsymbol{\lambda}_s}^2 \right\} = (\mu^2 \sigma_v^2 \boldsymbol{\lambda}_s + \boldsymbol{\lambda}_q)^T (\mu \mathbf{A} - \mu^2 \mathbf{B})^{-1} \boldsymbol{\lambda}_s\end{aligned}$$

which can be rewritten as

$$\text{MSD} = \frac{2 \sum_{m=1}^M \frac{\mu^2 \sigma_v^2 + \lambda_{\mathbf{q},m} \lambda_{\mathbf{s},m}^{-1}}{4 - \mu \lambda_{\mathbf{s},m} \left(1 + \sum_{\alpha=i,j,\kappa} r_{\bar{\mathbf{s}}\alpha,m}^2 \right)}}{\mu - \mu^2 \sum_{m=1}^M \frac{2 \lambda_{\mathbf{s},m}}{4 - \mu \lambda_{\mathbf{s},m} \left(1 + \sum_{\alpha=i,j,\kappa} r_{\bar{\mathbf{s}}\alpha,m}^2 \right)}} \quad (5.46)$$

$$\text{EMSE} = \frac{2 \sum_{m=1}^M \frac{\mu^2 \sigma_v^2 \lambda_{\mathbf{s},m} + \lambda_{\mathbf{q},m}}{4 - \mu \lambda_{\mathbf{s},m} \left(1 + \sum_{\alpha=i,j,\kappa} r_{\bar{\mathbf{s}}\alpha,m}^2 \right)}}{\mu - \mu^2 \sum_{m=1}^M \frac{2 \lambda_{\mathbf{s},m}}{4 - \mu \lambda_{\mathbf{s},m} \left(1 + \sum_{\alpha=i,j,\kappa} r_{\bar{\mathbf{s}}\alpha,m}^2 \right)}} \quad (5.47)$$

Remark 5.3. Observe from (5.46) and (5.47) that the steady-state EMSE and MSD increase with the length of the regressor vector, M , the power of the regressor reflected by $\lambda_{\mathbf{s},m}$, the impropriety degree of $\bar{\mathbf{s}}_n$ reflected by $r_{\bar{\mathbf{s}}i,m}, r_{\bar{\mathbf{s}}j,m}, r_{\bar{\mathbf{s}}\kappa,m}$, and the variance of the system weight variation reflected by $\lambda_{\mathbf{q},m}$.

Table 5.1 summarises the sufficient stability conditions, necessary stability conditions, necessary and sufficient stability conditions, MSDs, and EMSEs for general quaternion adaptive filters and QLMS filters with Gaussian inputs. The closed-form representations of the performance of SL-QLMS, SWL-QLMS, and WL-QLMS with Gaussian regressors satisfying Assumption 5.1 can be obtained by replacing $\mathbf{s}$ with $\mathbf{x}_n$, $\dot{\mathbf{x}}_n$, and $\ddot{\mathbf{x}}_n$ respectively in (5.43), (5.46) and (5.47). Exploring the second-order statistics of input signal, the following three subsections further address the performance of the three classes of QLMS algorithms.

Table 5.1: Analysis results of quaternion adaptive filters. S: sufficient; N: necessary.

	Stability condition			Steady state	
	S	N	N&S	MSD	EMSE
General quaternion adaptive filter	Eq.(5.26)	Eq.(5.29)	Eq.(5.20)	Eq.(5.30) (bounds)	Eq.(5.31) (bounds)
QLMS with Gaussian inputs	Eq.(5.36)	Eq.(5.37)	Eq.(5.34)	Eq.(5.38) (bounds)	Eq.(5.31) (bounds)
QLMS with Gaussian inputs on Assumption 5.1	Eq.(5.36)	Eq.(5.37)	Eq.(5.43)	Eq.(5.46)	Eq.(5.47)

5.4.1 SL-QLMS

Scalar input

The stability condition of the SL-QLMS with a scalar regressor x_n is obtained from (5.43) as

$$0 < \mu < \frac{4}{\sigma_x^2 (3 + r_{x_i}^2 + r_{x_j}^2 + r_{x_\kappa}^2)} \quad (5.48)$$

where σ_x^2 is the variance of x_n . This condition is the same as the analysis result obtained in [49]. The steady-state performance can be derived from (5.46) and (5.47) as

$$\text{MSD} = \frac{2(\mu^2 \sigma_v^2 \sigma_x^2 + \sigma_q^2)}{\mu \sigma_x^2 [4 - \mu \sigma_x^2 (3 + r_{x_i}^2 + r_{x_j}^2 + r_{x_\kappa}^2)]} \quad (5.49)$$

$$\text{EMSE} = \frac{2(\mu^2 \sigma_v^2 \sigma_x^2 + \sigma_q^2)}{\mu [4 - \mu \sigma_x^2 (3 + r_{x_i}^2 + r_{x_j}^2 + r_{x_\kappa}^2)]} \quad (5.50)$$

where σ_q^2 is the variance of the weight variation q_n .

$\mathbb{H}$ -proper input vector

The diagonalisation of $\mathbf{C}_{\mathbf{x}\mathbf{x}}$ completely decorrelates an $\mathbb{H}$ -proper quaternion input vector $\mathbf{x}_n$. For such inputs, $r_{\bar{\mathbf{x}}\alpha, m} = 0, \forall m, \alpha$, so that the stability condition in

(5.43) and the steady-state performance in (5.46) and (5.47) become

$$0 < \sum_{m=1}^M \frac{2\mu\lambda_{\mathbf{x},m}}{4 - \mu\lambda_{\mathbf{x},m}} < 1 \quad (5.51)$$

$$\text{MSD} = \frac{2 \sum_{m=1}^M \frac{\mu^2 \sigma_v^2 + \lambda_{\mathbf{q},m} \lambda_{\mathbf{x},m}^{-1}}{4 - \mu\lambda_{\mathbf{x},m}}}{\mu \left(1 - \mu \sum_{m=1}^M \frac{2\lambda_{\mathbf{x},m}}{4 - \mu\lambda_{\mathbf{x},m}} \right)} \quad (5.52)$$

$$\text{EMSE} = \frac{2 \sum_{m=1}^M \frac{\mu^2 \sigma_v^2 \lambda_{\mathbf{x},m} + \lambda_{\mathbf{q},m}}{4 - \mu\lambda_{\mathbf{x},m}}}{\mu \left(1 - \mu \sum_{m=1}^M \frac{2\lambda_{\mathbf{x},m}}{4 - \mu\lambda_{\mathbf{x},m}} \right)} \quad (5.53)$$

Rectilinear input vector

If the Gaussian input vector $\mathbf{x}_n$ is rectilinear, it can be expressed as $\mathbf{x}_n = e^\zeta \mathbf{z}_n$, where $\mathbf{z}_n$ is a random Gaussian real vector, and ζ a constant pure quaternion scalar [89]. For such inputs, the covariance and complementary covariance matrices can be represented by $\mathbf{C}_{\mathbf{xx}} = \mathbf{C}_{\mathbf{zz}}$, $\mathbf{C}_{\mathbf{xx}^\alpha} = e^\zeta (e^{-\zeta})^\alpha \mathbf{C}_{\mathbf{zz}}$, $\alpha = i, j, \kappa$. Therefore, $\mathbf{C}_{\mathbf{xx}}$ and $\mathbf{C}_{\mathbf{xx}^\alpha}$ can be simultaneously diagonalised via the diagonalisation of the real symmetric matrix $\mathbf{C}_{\mathbf{zz}}$. Note $r_{\bar{\mathbf{x}}\alpha,m} = 1, \forall m, \alpha$, so that the stability condition in (5.43) and the steady-state performance in (5.46) and (5.47) become

$$0 < \sum_{m=1}^M \frac{\mu\lambda_{\mathbf{x},m}}{2 - 2\mu\lambda_{\mathbf{x},m}} < 1 \quad (5.54)$$

$$\text{MSD} = \frac{\sum_{m=1}^M \frac{\mu^2 \sigma_v^2 + \lambda_{\mathbf{q},m} \lambda_{\mathbf{x},m}^{-1}}{2 - 2\mu\lambda_{\mathbf{x},m}}}{\mu \left(1 - \mu \sum_{m=1}^M \frac{\lambda_{\mathbf{x},m}}{2 - 2\mu\lambda_{\mathbf{x},m}} \right)} \quad (5.55)$$

$$\text{EMSE} = \frac{\sum_{m=1}^M \frac{\mu^2 \sigma_v^2 \lambda_{\mathbf{x},m} + \lambda_{\mathbf{q},m}}{2 - 2\mu\lambda_{\mathbf{x},m}}}{\mu \left(1 - \mu \sum_{m=1}^M \frac{\lambda_{\mathbf{x},m}}{2 - 2\mu\lambda_{\mathbf{x},m}} \right)} \quad (5.56)$$

5.4.2 SWL-QLMS

$\mathbb{H}$ -proper input vector

If the random quaternion vector $\mathbf{x}_n$ is $\mathbb{H}$ -proper, $\mathbf{C}_{\dot{\mathbf{x}}\dot{\mathbf{x}}}$ reduces to a block diagonal matrix. Denote the eigendecomposition of $\mathbf{C}_{\mathbf{x}\mathbf{x}}$ by $\mathbf{C}_{\mathbf{x}\mathbf{x}} = \mathbf{Q}\mathbf{\Lambda}_{\mathbf{x}\mathbf{x}}\mathbf{Q}^H$, where $\mathbf{\Lambda}_{\mathbf{x}\mathbf{x}}$ is a diagonal real matrix containing all eigenvalues of $\mathbf{C}_{\mathbf{x}\mathbf{x}}$ on the diagonal, and $\mathbf{Q}$ is a unitary quaternion matrix containing eigenvectors of $\mathbf{C}_{\mathbf{x}\mathbf{x}}$. By using the unitary transform

$$\mathbf{G} = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{Q} & -\mathbf{Q} \\ \mathbf{Q}^\eta & \mathbf{Q}^\eta \end{bmatrix}$$

$\bar{\mathbf{s}}_n = \mathbf{G}^H \mathbf{x}_n^b$ is a $\mathbb{C}^\eta$ -improper vector and satisfies

$$\mathbf{C}_{\bar{\mathbf{s}}\bar{\mathbf{s}}} = \mathbf{I}_2 \otimes \mathbf{\Lambda}_{\mathbf{x}\mathbf{x}}, \quad \mathbf{C}_{\bar{\mathbf{s}}\bar{\mathbf{s}}^\eta} = \text{Diag}(1, -1) \otimes \mathbf{\Lambda}_{\mathbf{x}\mathbf{x}}$$

This reduces the stability condition in (5.43) and the steady-state performance in (5.46) and (5.47) to

$$0 < \sum_{m=1}^{M/2} \frac{2\mu\lambda_{\mathbf{x},m}}{2 - \mu\lambda_{\mathbf{x},m}} < 1 \quad (5.57)$$

$$\text{MSD} = \frac{\sum_{m=1}^{M/2} \frac{2\mu^2\sigma_v^2\lambda_{\mathbf{x},m} + \lambda_{\mathbf{q},m} + \lambda_{\mathbf{q},m+M/2}}{\lambda_{\mathbf{x},m}(2 - \mu\lambda_{\mathbf{x},m})}}{\mu \left(1 - 2\mu \sum_{m=1}^{M/2} \frac{\lambda_{\mathbf{x},m}}{2 - \mu\lambda_{\mathbf{x},m}} \right)} \quad (5.58)$$

$$\text{EMSE} = \frac{\sum_{m=1}^{M/2} \frac{2\mu^2\sigma_v^2\lambda_{\mathbf{x},m} + \lambda_{\mathbf{q},m} + \lambda_{\mathbf{q},m+M/2}}{2 - \mu\lambda_{\mathbf{x},m}}}{\mu \left(1 - 2\mu \sum_{m=1}^{M/2} \frac{\lambda_{\mathbf{x},m}}{2 - \mu\lambda_{\mathbf{x},m}} \right)} \quad (5.59)$$

Rectilinear input vector

As shown in Section 5.4.1, the covariance matrix of a rectilinear quaternion random vector $\mathbf{x}_n$ is real-valued, so its eigendecomposition can be represented by $\mathbf{C}_{\mathbf{x}\mathbf{x}} =$

$\mathbf{Q}\mathbf{\Lambda}_{\mathbf{xx}}\mathbf{Q}^T$, where $\mathbf{Q}$ is a real orthogonal matrix. By using the unitary transform

$$\mathbf{G} = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{Q} & -e^\zeta (e^{-\zeta})^\eta \mathbf{Q} \\ (e^\zeta)^\eta e^{-\zeta} \mathbf{Q} & \mathbf{Q} \end{bmatrix}$$

$\bar{\mathbf{s}}_n = \mathbf{G}^H \mathbf{x}_n^b$ satisfies

$$\begin{aligned} \mathbf{C}_{\bar{\mathbf{s}}\bar{\mathbf{s}}} &= \text{Diag}(2, 0) \otimes \mathbf{\Lambda}_{\mathbf{xx}} \\ \mathbf{C}_{\bar{\mathbf{s}}\bar{\mathbf{s}}^\eta} &= \text{Diag}(2e^\zeta (e^{-\zeta})^\alpha, 0) \otimes \mathbf{\Lambda}_{\mathbf{xx}}, \quad \alpha = \iota, j, \kappa \end{aligned}$$

which reduces the stability condition in (5.43) and the steady-state EMSE in (5.47) to

$$0 < \sum_{m=1}^{M/2} \frac{\mu \lambda_{\mathbf{x},m}}{1 - 2\mu \lambda_{\mathbf{x},m}} < 1 \quad (5.60)$$

$$\text{EMSE} = \frac{\sum_{m=1}^{M/2} \frac{2\mu^2 \sigma_v^2 \lambda_{\mathbf{x},m} + \lambda_{\mathbf{q},m}}{1 - 2\mu \lambda_{\mathbf{x},m}} + \sum_{m=1+M/2}^M \lambda_{\mathbf{q},m}}{2\mu \left(1 - \mu \sum_{m=1}^{M/2} \frac{\lambda_{\mathbf{x},m}}{1 - 2\mu \lambda_{\mathbf{x},m}} \right)} \quad (5.61)$$

In this case, the steady-state MSD is uncertain as the weight solution is not unique.

5.4.3 WL-QLMS

$\mathbb{H}$ -proper input vector

If $\mathbf{x}_n$ is $\mathbb{H}$ -proper, $\mathbf{C}_{\mathbf{x}^a \mathbf{x}^a}$ reduces to a block diagonal matrix. Consider the eigen-decomposition of $\mathbf{C}_{\mathbf{xx}}$, $\mathbf{C}_{\mathbf{xx}} = \mathbf{Q}\mathbf{\Lambda}_{\mathbf{xx}}\mathbf{Q}^H$. By using the unitary transform

$$\mathbf{G} = \frac{1}{2} \begin{bmatrix} \mathbf{Q} & -\mathbf{Q} & -\mathbf{Q} & -\mathbf{Q} \\ \mathbf{Q}^\iota & -\mathbf{Q}^\iota & \mathbf{Q}^\iota & \mathbf{Q}^\iota \\ \mathbf{Q}^j & \mathbf{Q}^j & -\mathbf{Q}^j & \mathbf{Q}^j \\ \mathbf{Q}^\kappa & \mathbf{Q}^\kappa & \mathbf{Q}^\kappa & -\mathbf{Q}^\kappa \end{bmatrix}$$

$\bar{\mathbf{s}}_n = \mathbf{G}^H \mathbf{x}_n^a$ satisfies

$$\begin{aligned}\mathbf{C}_{\bar{\mathbf{s}}\bar{\mathbf{s}}} &= \mathbf{I}_4 \otimes \mathbf{\Lambda}_{\mathbf{xx}} \\ \mathbf{C}_{\bar{\mathbf{s}}\bar{\mathbf{s}}^i} &= \text{Diag}(1, 1, -1, -1) \otimes \mathbf{\Lambda}_{\mathbf{xx}} \\ \mathbf{C}_{\bar{\mathbf{s}}\bar{\mathbf{s}}^j} &= \text{Diag}(1, -1, 1, -1) \otimes \mathbf{\Lambda}_{\mathbf{xx}} \\ \mathbf{C}_{\bar{\mathbf{s}}\bar{\mathbf{s}}^\kappa} &= \text{Diag}(1, -1, -1, 1) \otimes \mathbf{\Lambda}_{\mathbf{xx}}\end{aligned}$$

The stability condition in (5.43) and the steady-state performance in (5.46) and (5.47) become

$$0 < \sum_{m=1}^{M/4} \frac{2\mu\lambda_{\mathbf{x},m}}{1 - \mu\lambda_{\mathbf{x},m}} < 1 \quad (5.62)$$

$$\text{MSD} = \frac{\sum_{m=1}^{M/4} \frac{4\mu^2\sigma_v^2\lambda_{\mathbf{x},m} + \lambda_{\mathbf{q},m} + \lambda_{\mathbf{q},m+M/4} + \lambda_{\mathbf{q},m+M/2} + \lambda_{\mathbf{q},m+3M/4}}{2\lambda_{\mathbf{x},m}(1 - \mu\lambda_{\mathbf{x},m})}}{\mu \left(1 - 2\mu \sum_{m=1}^{M/4} \frac{\lambda_{\mathbf{x},m}}{1 - \mu\lambda_{\mathbf{x},m}} \right)} \quad (5.63)$$

$$\text{EMSE} = \frac{\sum_{m=1}^{M/4} \frac{4\mu^2\sigma_v^2\lambda_{\mathbf{x},m} + \lambda_{\mathbf{q},m} + \lambda_{\mathbf{q},m+M/4} + \lambda_{\mathbf{q},m+M/2} + \lambda_{\mathbf{q},m+3M/4}}{2(1 - \mu\lambda_{\mathbf{x},m})}}{\mu \left(1 - 2\mu \sum_{m=1}^{M/4} \frac{\lambda_{\mathbf{x},m}}{1 - \mu\lambda_{\mathbf{x},m}} \right)} \quad (5.64)$$

Rectilinear input vector

If $\mathbf{x}_n$ is rectilinear, the quaternion unitary matrix

$$\mathbf{G} = \frac{1}{2} \begin{bmatrix} \mathbf{Q} & -\mathbf{Q} & -\mathbf{Q} & -\mathbf{Q} \\ (e^\zeta)^i e^{-\zeta} \mathbf{Q} & -(e^\zeta)^i e^{-\zeta} \mathbf{Q} & (e^\zeta)^i e^{-\zeta} \mathbf{Q} & (e^\zeta)^i e^{-\zeta} \mathbf{Q} \\ (e^\zeta)^j e^{-\zeta} \mathbf{Q} & (e^\zeta)^j e^{-\zeta} \mathbf{Q} & -(e^\zeta)^j e^{-\zeta} \mathbf{Q} & (e^\zeta)^j e^{-\zeta} \mathbf{Q} \\ (e^\zeta)^\kappa e^{-\zeta} \mathbf{Q} & (e^\zeta)^\kappa e^{-\zeta} \mathbf{Q} & (e^\zeta)^\kappa e^{-\zeta} \mathbf{Q} & -(e^\zeta)^\kappa e^{-\zeta} \mathbf{Q} \end{bmatrix}$$

allows the transformed vector $\bar{\mathbf{s}}_n = \mathbf{G}^H \mathbf{x}_n^a$ to satisfy

$$\begin{aligned}\mathbf{C}_{\bar{\mathbf{s}}\bar{\mathbf{s}}} &= \text{Diag}(4, 0, 0, 0) \otimes \mathbf{\Lambda}_{\mathbf{xx}} \\ \mathbf{C}_{\bar{\mathbf{s}}\bar{\mathbf{s}}^\alpha} &= \text{Diag}(4e^\zeta (e^{-\zeta})^\alpha, 0, 0, 0) \otimes \mathbf{\Lambda}_{\mathbf{xx}}, \quad \alpha = i, j, \kappa\end{aligned}$$

which reduces the stability condition in (5.43) and the steady-state EMSE in (5.47) to

$$0 < \sum_{m=1}^{M/4} \frac{2\mu\lambda_{\mathbf{x},m}}{1 - 4\mu\lambda_{\mathbf{x},m}} < 1 \quad (5.65)$$

$$\text{EMSE} = \frac{\sum_{m=1}^{M/4} \frac{4\mu^2\sigma_v^2\lambda_{\mathbf{x},m} + \lambda_{\mathbf{q},m}}{1 - 4\mu\lambda_{\mathbf{x},m}} + \sum_{m=1+M/4}^M \lambda_{\mathbf{q},m}}{2\mu \left(1 - 2\mu \sum_{m=1}^{M/4} \frac{\lambda_{\mathbf{x},m}}{1 - 4\mu\lambda_{\mathbf{x},m}} \right)} \quad (5.66)$$

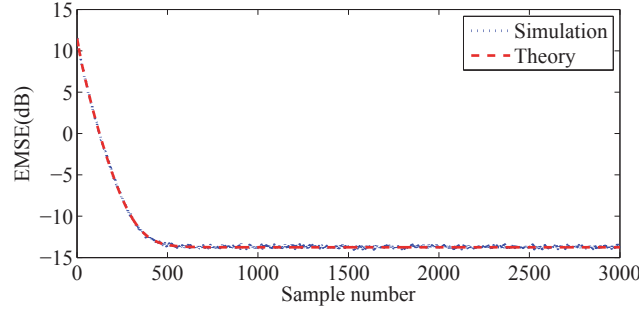
In this case, the steady-state MSD is uncertain as the weight solution is not unique.

5.5 Simulations

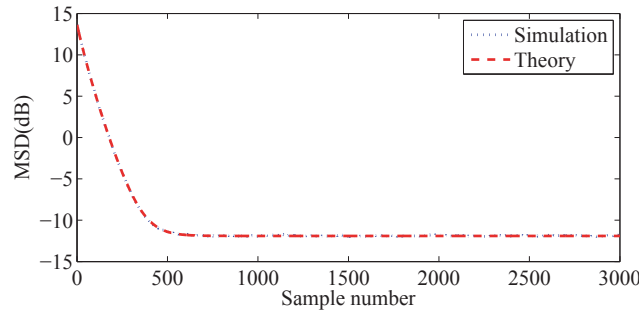
The performance of the three considered classes of QLMS algorithms with Gaussian inputs was evaluated by averaging results over 1000 independent numerical simulation trials. Throughout the simulations, the length of the synthetic correlated input vector $\mathbf{x}_n$ was $L = 4$, and the eigenvalues of $\mathbf{C}_{\mathbf{xx}}$ were kept the same. For simplicity, the elements of $\mathbf{x}_n$ were set to have an equal impropriety coefficient, that is, $r_{m,i} = r_{m,j} = r_{m,k} = r$.

5.5.1 SL-QLMS

In the simulations for the SL-QLMS, the desired signal was generated from the strictly linear model in (5.2) and (5.7). The variance of the additive Gaussian noise v_n was $\sigma_v^2 = 0.01$, and the Euclidean norm of the Gaussian weight variation $\mathbf{q}_n$ was 0.0016. Figure 5.1 depicts the theoretical and simulated EMSE and MSD learning curves of the SL-QLMS for an improper input with $r = 0.5$ and $\mu = 0.02$. These curves exhibit a good match between the theory and simulation results. Figures 5.2 and 5.3 illustrate the theoretical and simulated steady-state EMSEs for different values of the step size μ when $\mathbf{x}_n$ is $\mathbb{H}$ -proper and when $\mathbf{x}_n$ is improper with $r = 0.5$, respectively. The theoretical bounds of μ are also indicated. Both figures show that there was a good match between theory and simulations and their difference increased with μ , which is similar to the behaviour of real-valued LMS



(a) EMSE curves



(b) MSD curves

Figure 5.1: Theoretical and simulated learning curves of SL-QLMS for an improper input with $r = 0.5$ and $\mu = 0.02$.

filters [81]. Observe that the decrease in the stability bound for the improper input conforms with Remark 5.2. Figure 5.8 illustrates a match between the theoretical and simulated steady-state EMSEs for inputs with various impropriety coefficients $r \in [0, 1]$ and $\mu = 0.2$. The increase in the steady-state EMSEs with r conforms with Remark 5.3. As shown in Remark 5.2, Figure 5.2, and Figure 5.3, the upper bound of μ decreases with the increase in r . So, for a fixed μ , the theoretical EMSE for more improper inputs is less accurate, as μ is closer to the upper bound. This explains the difference between the theoretical and simulated steady-state EMSEs for a large r in Figure 5.4 as well as in Figures 5.8 and 5.12. The underlying cause of the performance degradation for highly improper input signals might be the complicated structure of those signals, which is worth further research.

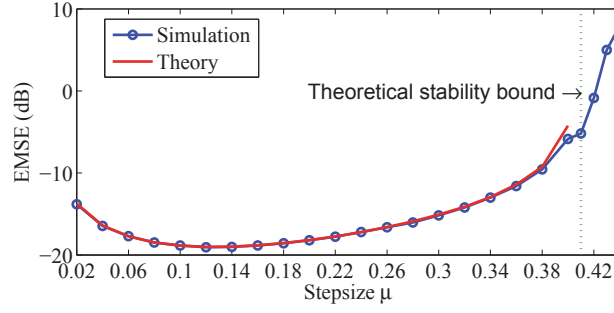


Figure 5.2: Theoretical and simulated EMSEs of SL-QLMS as a function of the step size, for an $\mathbb{H}$ -proper input.

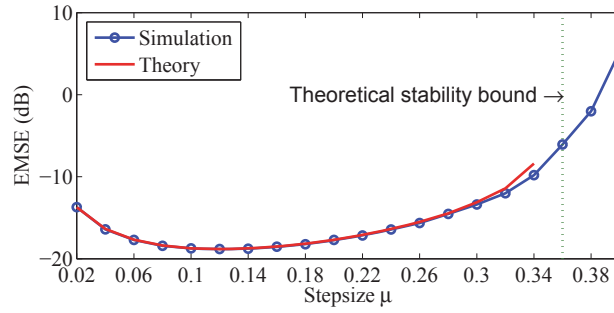


Figure 5.3: Theoretical and simulated EMSEs of SL-QLMS as a function of the step size, for an improper input with $r = 0.5$.

5.5.2 SWL-QLMS

In the simulations for the SWL-QLMS, the desired signal was generated from the semi-widely linear model in (5.3) and (5.7). The variance of the additive Gaussian noise v_n was $\sigma_v^2 = 0.01$, and the Euclidean norm of the Gaussian weight variation, $\mathbf{q}_n$, was 0.0008. Figure 5.5 depicts theoretical and simulated EMSE and MSD learning curves of the SL-QLMS for an improper input with $r = 0.5$ and $\mu = 0.02$. These curves exhibit a good match between the theory and simulation results. Figures 5.6 and 5.7 show theoretical and simulated steady-state EMSEs for different values of the step size μ when $\mathbf{x}_n$ is $\mathbb{H}$ -proper and when $\mathbf{x}_n$ is improper with $r = 0.5$. The theoretical bounds of μ are also indicated. Both figures illustrate a good match between theory and simulations and an increase in their difference with μ . The decrease in the stability bound for the improper input conforms with

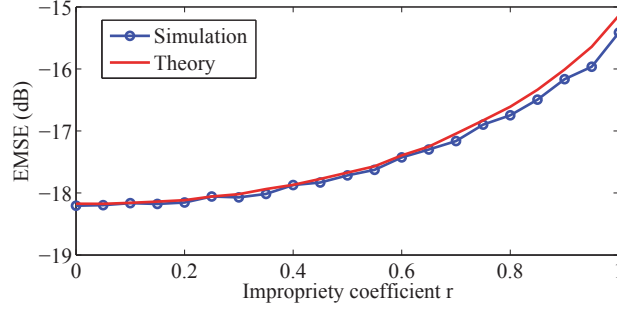
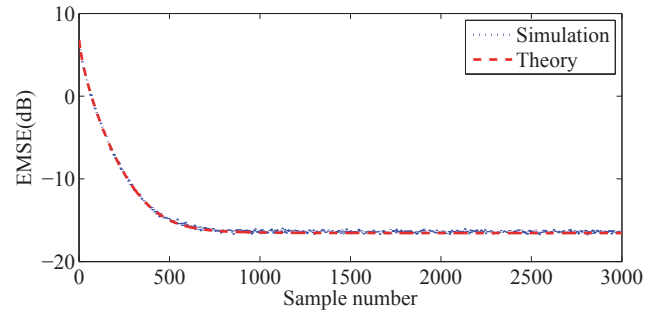


Figure 5.4: Theoretical and simulated EMSEs of SL-QLMS as a function of the impropriety coefficient, for $\mu = 0.2$.

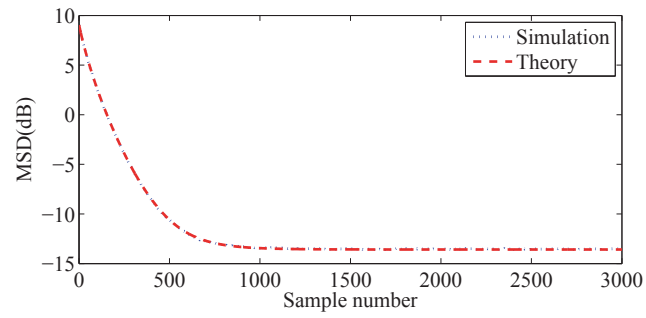
Remark 5.2. Figure 5.8 illustrates a match between theoretical and simulated steady-state EMSEs for inputs with various impropriety coefficients $r \in [0, 1]$ and $\mu = 0.1$. The increase in the steady-state EMSEs with r conforms with Remark 5.3.

5.5.3 WL-QLMS

In the simulations for the WL-QLMS, the desired signal was generated from the widely linear model in (5.4) and (5.7). The variance of the additive Gaussian noise v_n was $\sigma_v^2 = 0.01$, and the Euclidean norm of the Gaussian weight variation $\mathbf{q}_n$ was 0.0004. Figure 5.9 depicts theoretical and simulated EMSE and MSD learning curves of the SL-QLMS for an improper input with $r = 0.5$ and $\mu = 0.02$. These curves exhibit a good match between the theory and simulation results. Figures 5.10 and 5.11 show theoretical and simulated steady-state EMSEs for different values of the step size μ when $\mathbf{x}_n$ is $\mathbb{H}$ -proper and when $\mathbf{x}_n$ is improper with $r = 0.5$. The theoretical bounds of μ are also indicated. Both figures display a good match between theory and simulations and an increase in their difference with μ . The decrease in the stability bound for the improper input conforms with Remark 5.2. Figure 5.12 illustrates a match between theoretical and simulated steady-state EMSEs for inputs with various impropriety coefficients $r \in [0, 1]$ and $\mu = 0.05$. The increase in the steady-state EMSEs with r conforms with Remark



(a) EMSE curves



(b) MSD curves

Figure 5.5: Theoretical and simulated learning curves of SWL-QLMS for an improper input with $r = 0.5$ and $\mu = 0.02$.

5.3.

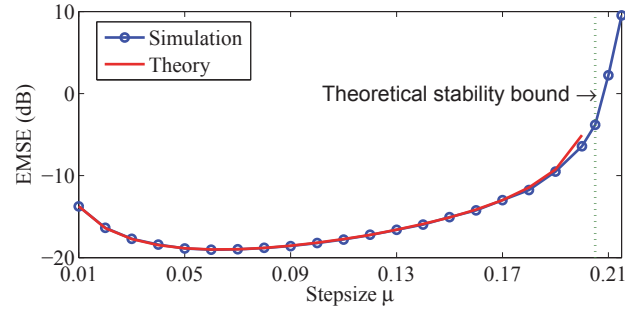


Figure 5.6: Theoretical and simulated EMSEs of SWL-QLMS as a function of the step size, for an $\mathbb{H}$ -proper input.

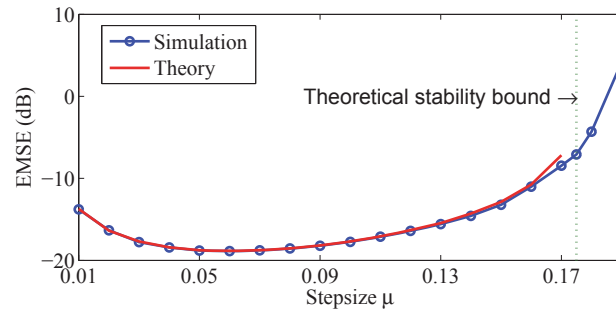


Figure 5.7: Theoretical and simulated EMSEs of SWL-QLMS as a function of the step size, for an improper input with $r = 0.5$.

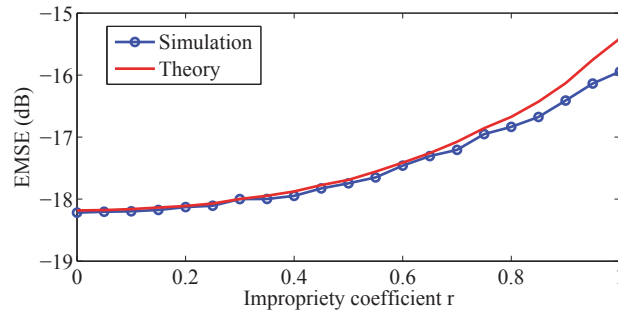
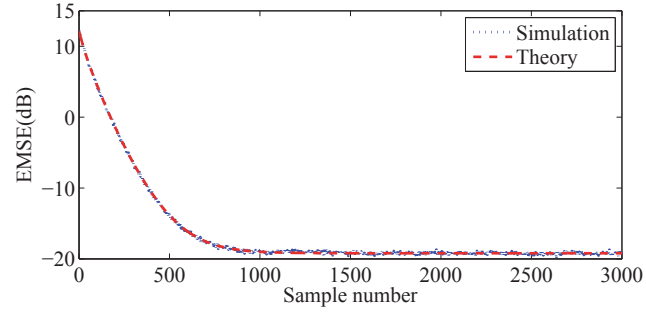
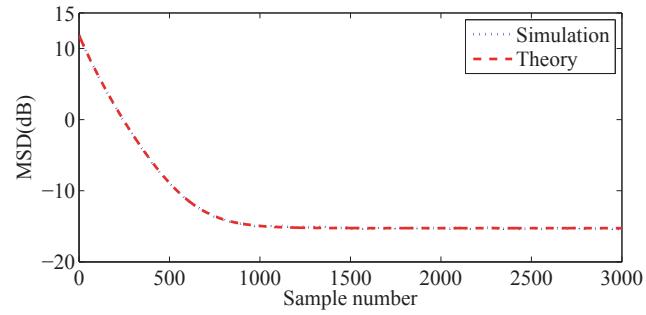


Figure 5.8: Theoretical and simulated EMSEs of SWL-QLMS as a function of the impropriety coefficient, for $\mu = 0.1$.



(a) EMSE curves



(b) MSD curves

Figure 5.9: Theoretical and simulated learning curves of WL-QLMS for an improper input with $r = 0.5$ and $\mu = 0.02$.

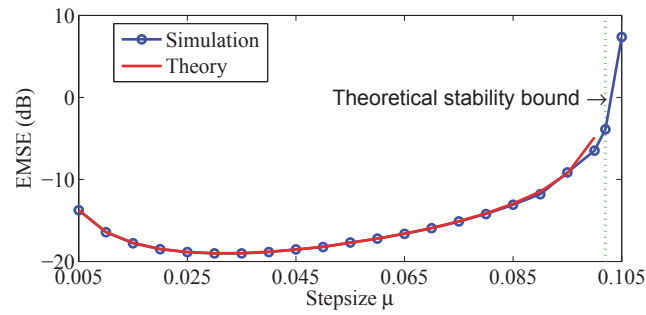


Figure 5.10: Theoretical and simulated EMSEs of WL-QLMS as a function of the step size, for an $\mathbb{H}$ -proper input.

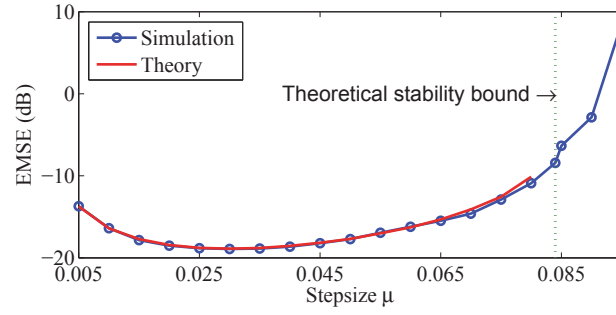


Figure 5.11: Theoretical and simulated EMSEs of WL-QLMS as a function of the step size, for an improper input with $r = 0.5$.

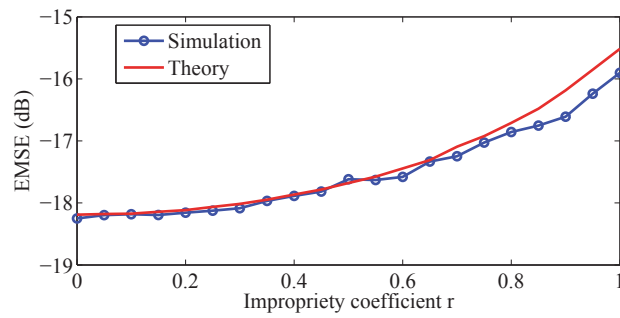


Figure 5.12: Theoretical and simulated EMSEs of WL-QLMS as a function of the improprity coefficient, for $\mu = 0.05$.

5.6 Summary

Convergence of quaternion adaptive filtering algorithms in non-stationary environments has been analysed by extending an analysis framework for real-valued adaptive filters to the quaternion domain. For rigour, the first-order random-walk model has been employed to model the variation of system weights in a non-stationary environment. The bounds on transient and steady-state performance of general quaternion adaptive filters have been derived. For QLMS filters with Gaussian inputs, an innovative quaternion decorrelation technique has simplified the analysis and enabled the step size bounds, MSDs and EMSEs to be expressed in closed forms. The analytical results, which are shown to be influenced by the second-order statistics of input signals, provide new insights into the statistical behaviour of the considered filtering algorithms. Effectiveness of the proposed analysis has been verified by numerical simulations for SL-QLMS, SWL-QLMS and WL-QLMS filters with correlated Gaussian inputs.

Part III

Novel Adaptive Estimation Algorithms

Chapter 6

The Quaternion Least Mean Magnitude-Phase Adaptive Filtering Algorithm

Overview

Quaternion-valued adaptive filters based on the MSE criterion have been extensively studied in recent years. However, the MSE cost function has only one degree of freedom. To circumvent this problem, this chapter introduces another criterion which enables separate control of the magnitude and phase. Accordingly, a quaternion least mean magnitude-phase (QLMMP) filtering algorithm is proposed and is shown to provide a unified form for the quaternion least mean square (QLMS) algorithm, a magnitude-only filtering algorithm, and a phase filtering algorithm. The convergence analysis of the normalised version of the QLMMP employs a connection to the normalised QLMS with a quaternion-valued step size. The proposed algorithms are validated over case studies of the identification in three-phase power systems.

6.1 Introduction

The LMS algorithm is known not to perform most efficiently for complex- and quaternion-valued signals with different dynamics in their magnitudes and phases, like those caused by Doppler effects [90]. To improve the filtering performance, several criteria have been proposed to separately extract the magnitude and/or phase

information from complex-valued signals, such as the least mean phase (LMP) algorithm for phase estimation [90], the constant modulus channel estimator for magnitude estimation [91], and the least mean magnitude-phase (LMMP) algorithm for advanced estimation of magnitude and phase variations [92, 93]. The LMMP has shown superior performance for the applications where different convergence performances in the magnitude and phase are needed. For example, when the signal is contaminated by larger magnitude noise and smaller phase noise, slower adaption in magnitude and faster adaption in phase is desired.

Magnitude/phase coupling is inherent to quaternion-valued signals. Based on a definition of phase in the quaternion domain, a quaternion LMP (QLMP) algorithm has been introduced [94]. However, it requires complicated computations and the phase change in each iteration is restricted to be less than $\pi/2$. To this end, this chapter proposes a quaternion LMMP (QLMMP) filtering algorithm which decomposes the instantaneous squared error cost into a composite of the squared magnitude error and a term representing the phase error between the desired signal and the estimate. In this way, a separate control of magnitude and phase is achieved, together with the corresponding weight updates. The convergence analysis for the normalised QLMMP algorithm is proposed by regarding the QLMMP as a special case of QLMS with an adaptive quaternion-valued step size. This provides a very general algorithm. By setting equal step sizes for the magnitude and phase update, the QLMMP is equivalent to the standard QLMS with a real-valued step size, while the vanishing phase (or magnitude) step size reduces the QLMMP to a magnitude-only (or phase) filtering algorithm. The phase filtering algorithm is shown to converge faster than the QLMP, and can be an alternative to the QLMS when combined with a simple magnitude adjustment. The performances of the proposed algorithms are validated by simulations in the context of frequency estimation of three-phase power systems.

6.2 The QLMMP Algorithm

Assume that the desired quaternion-valued signal at time n arises from a strictly linear model, $y_n = \mathbf{w}^H \mathbf{x}_n + v_n$, where $\mathbf{x}_n \in \mathbb{H}^{L \times 1}$ is the known input vector, $\mathbf{w} \in \mathbb{H}^{L \times 1}$ is the unknown optimal weight vector of the system, and $v_n \in \mathbb{H}$ is zero-mean white Gaussian noise with variance σ_v^2 . To get the estimate of y_n as $\hat{y}_n = \hat{\mathbf{w}}_n^H \mathbf{x}_n$, where $\hat{\mathbf{w}}_n \in \mathbb{H}^{L \times 1}$ is the weight vector, the QLMS algorithm, based on the minimisation of the instantaneous squared error cost $J_n = |y_n - \hat{y}_n|^2$, has been proposed [25]. Using a step size μ_1 and the gradient of J_n with respect to the conjugate of the weight vector, $\nabla_{\hat{\mathbf{w}}_n^*} J_n = -\mathbf{x}_n (y_n - \hat{y}_n)^*$, the weight update rule of the normalised QLMS is given by [24]

$$\hat{\mathbf{w}}_{n+1} = \hat{\mathbf{w}}_n - \frac{\mu_1}{\|\mathbf{x}_n\|^2} \nabla_{\hat{\mathbf{w}}_n^*} J_n = \hat{\mathbf{w}}_n + \frac{\mu_1}{\|\mathbf{x}_n\|^2} \mathbf{x}_n (y_n - \hat{y}_n)^*. \quad (6.1)$$

Following the LMMP approach for estimating complex-valued signals [92], the squared error cost can be decomposed into

$$J_n = \underbrace{(|y_n| - |\hat{y}_n|)^2}_{J_{m,n}} + 2 \underbrace{|\hat{y}_n| |y_n| (1 - \cos \theta_n)}_{J_{p,n}} \quad (6.2)$$

where $J_{m,n}$ is the magnitude cost, $J_{p,n}$ the phase cost, $\theta_n \in [0, \pi]$ the absolute value of the angle between $\hat{y}_n$ and y_n , and $\cos \theta_n = \Re[y_n^* \hat{y}_n] |y_n \hat{y}_n|^{-1}$. Similar to the complex LMMP algorithm [92], the minimisation of a linear weighted combination of $J_{m,n}$ and $J_{p,n}$, $J_{\text{lmmp},n} = \alpha_m J_{m,n} + \alpha_p J_{p,n}$, where $\alpha_m, \alpha_p \in \mathbb{R}^+$ denote the weights, leads to a new algorithm referred to as the QLMMP. Note that $J_{\text{lmmp},n} \geq 0$, where the equality holds only when $y_n = \hat{y}_n$. The gradients of $J_{\text{lmmp},n}$ with respect to the conjugate of the weight vector are then calculated as [23, 24]

$$\nabla_{\hat{\mathbf{w}}_n^*} J_{\text{lmmp},n} = 0.5 \mathbf{x}_n \left[\alpha_m \left(1 - \frac{|y_n|}{|\hat{y}_n|} \right) \hat{y}_n + \alpha_p \left(\frac{|y_n|}{|\hat{y}_n|} \hat{y}_n - y_n \right) \right]^*.$$

With the step size $\mu \in \mathbb{R}^+$, the weight update of the normalised QLMMP becomes

$$\begin{aligned}\hat{\mathbf{w}}_{n+1} &= \hat{\mathbf{w}}_n - \mu \|\mathbf{x}_n\|^{-2} \nabla_{\hat{\mathbf{w}}_n^*} J_{\text{lmp},n} \\ &= \hat{\mathbf{w}}_n + \frac{\mathbf{x}_n}{\|\mathbf{x}_n\|^2} \left[\mu_m \left(\frac{|y_n|}{|\hat{y}_n|} - 1 \right) \hat{y}_n + \mu_p \left(y_n - \frac{|y_n|}{|\hat{y}_n|} \hat{y}_n \right) \right]^*\end{aligned}\quad (6.3)$$

where $\mu_m = 0.5\mu\alpha_m$ and $\mu_p = 0.5\mu\alpha_p$ control the respective convergence rates in the magnitude and phase, so that they are termed magnitude and phase step sizes. By adjusting μ_m and μ_p , the QLMMP provides enhanced degrees of freedom in the minimisation of the cost function, and different convergence rates and accuracies in the magnitude and phase estimation. In contrast, the QLMS does not allow for such separate control of magnitude and phase.

From (6.3), given the new weight estimate $\hat{\mathbf{w}}_{n+1}$, the improvement in the estimate for d_n is

$$\Delta\hat{y} = \hat{\mathbf{w}}_{n+1}^H \mathbf{x}_n - \hat{y}_n = \underbrace{\mu_m \left(\frac{|y_n|}{|\hat{y}_n|} - 1 \right) \hat{y}_n}_{\Delta\hat{y}_m} + \underbrace{\mu_p \left(y_n - \frac{|y_n|}{|\hat{y}_n|} \hat{y}_n \right)}_{\Delta\hat{y}_p}$$

where $\Delta\hat{y}_m$ aligns with $\hat{y}_n$, and $\Delta\hat{y}_p$ rotates $\hat{y}_n$ towards y_n . Figure 6.1 visualises the estimation improvements in an iteration of QLMS [24], QLMP [94], and the proposed QLMMP. On the plane where y_n and $\hat{y}_n$ lie, the estimation improvement in the QLMS, $\Delta\hat{y}_{\text{qlms}}$, points to y_n , while the estimation improvement in the QLMP, $\Delta\hat{y}_{\text{qlmp}}$, is perpendicular to $\hat{y}_n$, indicating that the rotation angle of $\Delta\hat{y}_{\text{qlmp}}$ is always less than $\pi/2$. By weighting $\Delta\hat{y}_m$ and $\Delta\hat{y}_p$ via μ_m and μ_p , the estimation improvement in the QLMMP spans the area between $\Delta\hat{y}_{\text{qlms}}$ and $\Delta\hat{y}_{\text{qlmp}}$.

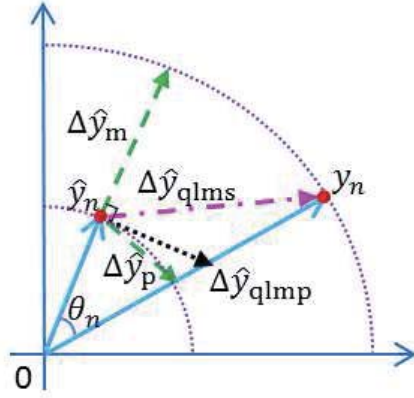


Figure 6.1: Geometric interpretations of the QLMMP, QLMS, and QLMP. Although $\mathbb{H}$ is a four-dimensional vector space over $\mathbb{R}$, y_n , $\hat{y}_n$, and $\Delta \hat{y}$ are on the same plane.

The above QLMMP algorithm can be extended to the widely linear QLMMP algorithm by incorporating the widely linear model. That is,

$$\hat{y}_n = \hat{\mathbf{w}}_n^{aH} \mathbf{x}_n^a$$

$$\hat{\mathbf{w}}_{n+1}^a = \hat{\mathbf{w}}_n^a + \frac{\mathbf{x}_n^a}{4 \|\mathbf{x}_n\|^2} \left[\mu_m \left(\frac{|y_n|}{|\hat{y}_n|} - 1 \right) \hat{y}_n + \mu_p \left(y_n - \frac{|y_n|}{|\hat{y}_n|} \hat{y}_n \right) \right]^*$$

6.3 Convergence Analysis

The step size of QLMS is confined to a real number in most applications but can be quaternion-valued alternatively. The quaternion-valued step size is an extension of the complex-valued step size in complex gradient learning algorithms [95–97]. From (6.1) and (6.3), at time n , the QLMMP and QLMS are equivalent if the step size in (6.1) is quaternion-valued, that is,

$$\mu_l = \mu_p + \frac{(\mu_m - \mu_p)(|y_n| - |\hat{y}_n|)}{|y_n - \hat{y}_n|^2 |\hat{y}_n|} \hat{y}_n (y_n - \hat{y}_n)^*. \quad (6.4)$$

This connection between the QLMMP and the QLMS with a quaternion-valued step size provides new insights into the convergence of the QLMMP, which is

difficult to analyse straightforwardly, owing to the high nonlinearity of the cost function, $\mu_m J_{m,n} + \mu_p J_{p,n}$. For a set of μ_m and μ_p , if all values of μ_l calculated from (6.4) for all y_n and $\hat{y}_n$ fall into the convergence region of the normalised QLMS, the corresponding normalised QLMMP achieves a monotonic decrease in MSE. Therefore, the convergence performance of the normalised QLMS with a quaternion-valued step size is a prerequisite for the convergence analysis of QLMMP.

6.3.1 Convergence Analysis of Normalised QLMS with a Quaternion Step Size

Denote the weight error vector by $\mathbf{v}_n = \mathbf{w} - \hat{\mathbf{w}}_n$, and (6.1) leads to

$$\mathbf{v}_{n+1} = (\mathbf{I} - \mu_l \|\mathbf{x}_n\|^{-2} \mathbf{x}_n \mathbf{x}_n^H) \mathbf{v}_n - \mu_l \|\mathbf{x}_n\|^{-2} \mathbf{x}_n v_n^* \quad (6.5)$$

Applying the statistical expectation operator to (6.5) yields

$$E\{\mathbf{v}_{n+1}\} = (\mathbf{I} - \mu_l \|\mathbf{x}_n\|^{-2} \mathbf{x}_n \mathbf{x}_n^H) E\{\mathbf{v}_n\}. \quad (6.6)$$

The eigendecomposition, $E\{\mathbf{x}_n \mathbf{x}_n^H\} = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^H$, enables a coordinate transform given by $\bar{\mathbf{x}}_n = \mathbf{Q}^H \mathbf{x}_n$ and $\bar{\mathbf{v}}_n = \mathbf{Q}^H \mathbf{v}_n$. Applying the statistical expectation operator to (6.5) yields

$$E\{\bar{\mathbf{v}}_{n+1}\} = (\mathbf{I} - \mu_l E\{\|\bar{\mathbf{x}}_n\|^{-2} \bar{\mathbf{x}}_n \bar{\mathbf{x}}_n^H\}) E\{\bar{\mathbf{v}}_n\}. \quad (6.7)$$

Thus, the sufficient condition for the mean convergence is

$$|1 - \lambda(E\{\|\bar{\mathbf{x}}_n\|^{-2} \bar{\mathbf{x}}_n \bar{\mathbf{x}}_n^H\}) \mu_l| < 1$$

which implies

$$2\Re[\mu_l] - |\mu_l|^2 \lambda(E\{\|\bar{\mathbf{x}}_n\|^{-2} \bar{\mathbf{x}}_n \bar{\mathbf{x}}_n^H\}) > 0 \quad (6.8)$$

where $\lambda(E\{\|\bar{\mathbf{x}}_n\|^{-2}\bar{\mathbf{x}}_n\bar{\mathbf{x}}_n^H\})$ represents any eigenvalue of $E\{\|\bar{\mathbf{x}}_n\|^{-2}\bar{\mathbf{x}}_n\bar{\mathbf{x}}_n^H\}$. It can be proved that $0 < \lambda(E\{\|\bar{\mathbf{x}}_n\|^{-2}\bar{\mathbf{x}}_n\bar{\mathbf{x}}_n^H\}) \leq 1$, and so the expression in (6.8) can be tightened to $2\Re[\mu_1] - |\mu_1|^2 > 0$.

To analyse the convergence in the mean square sense, the variance of the weight error vector can be calculated from (6.5) as

$$\begin{aligned}\|\mathbf{v}_{n+1}\|^2 = \mathbf{v}_{n+1}^H \mathbf{v}_{n+1} = & (|\mu_1|^2 - 2\Re[\mu_1]) \|\mathbf{x}_n\|^{-2} \mathbf{v}_n^H \mathbf{x}_n \mathbf{x}_n^H \mathbf{v}_n + \\ & \mathbf{v}_n^H \mathbf{v}_n + |\mu_1|^2 \sigma_\eta^2 \|\mathbf{x}_n\|^{-2} + \text{cross terms}.\end{aligned}$$

Upon taking statistical expectations of both sides of the above equation, the MSD is given by

$$\begin{aligned}E\{\|\mathbf{v}_{n+1}\|^2\} &= E\{\mathbf{v}_n^H \mathbf{E} \mathbf{v}_n\} + |\mu_1|^2 \sigma_\eta^2 E\{\|\mathbf{x}_n\|^{-2}\} \\ \mathbf{E} &= \mathbf{I} + (|\mu_1|^2 - 2\Re[\mu_1]) E\{\|\mathbf{x}_n\|^{-2} \mathbf{x}_n \mathbf{x}_n^H\}.\end{aligned}\tag{6.9}$$

The recursion in (6.9) converges if and only if all eigenvalues of $\mathbf{E}$ are within $(-1, 1)$ [81], which is equivalent to

$$-2 < (|\mu_1|^2 - 2\Re[\mu_1]) \lambda(E\{\|\mathbf{x}_n\|^{-2} \mathbf{x}_n \mathbf{x}_n^H\}) < 0.$$

It can be proved that $0 < \lambda(E\{\|\mathbf{x}_n\|_2^{-2} \mathbf{x}_n \mathbf{x}_n^H\}) \leq 1$, and so the above inequality can be tightened to $2\Re[\mu_1] - |\mu_1|^2 > 0$.

Therefore, the normalised QLMS converges in the mean and mean square sense for $2\Re[\mu_1] - |\mu_1|^2 > 0$.

Note that (6.9) yields

$$E\{\|\mathbf{v}_{n+1}\|^2\} = E\{\|\mathbf{v}_n\|^2\} + |\mu_1|^2 \sigma_\eta^2 E\{\|\mathbf{x}_n\|^{-2}\} + (|\mu_1|^2 - 2\Re[\mu_1]) E\{\|\mathbf{x}_n\|^{-2} |e_n^a|^2\}$$

where $e_n^a = \mathbf{v}_n^H \mathbf{x}_n$ is the *a priori* error. At the steady state, $\lim_{n \rightarrow \infty} E\{\|\mathbf{v}_{n+1}\|^2\} = \lim_{n \rightarrow \infty} E\{\|\mathbf{v}_n\|^2\}$. Through a further assumption that $\|\mathbf{x}_n\|^2$ is statistically indepen-

dent from $|e_n^a|^2$, the steady-state EMSE can be derived as

$$\lim_{n \rightarrow \infty} E \{ |e_n^a|^2 \} = |\mu_l|^2 \sigma_\eta^2 (2\Re[\mu_l] - |\mu_l|^2)^{-1}. \quad (6.10)$$

6.3.2 Convergence Analysis of Normalised QLMMP

(6.4) yields

$$\Re[\mu_l] = \mu_p + \frac{(\mu_m - \mu_p)(|y_n| - |\hat{y}_n|)}{|y_n - \hat{y}_n|^2} (|y_n| \cos \theta_n - |\hat{y}_n|)$$

$$|\mu_l|^2 = \frac{(\mu_m - \mu_p)(|y_n| - |\hat{y}_n|)}{|y_n - \hat{y}_n|^2} [(2\mu_p \cos \theta_n + \mu_m - \mu_p)|y_n| - (\mu_m + \mu_p)|\hat{y}_n|] + \mu_p^2$$

from where

$$2\Re[\mu_l] - |\mu_l|^2 = g(\cos \theta_n) |y_n|^2 |y_n - \hat{y}_n|^{-2} \quad (6.11)$$

with

$$g(\cos \theta_n) = 2 \cos \theta_n \left[(1 - \mu_p)(\mu_m - \mu_p) + (\mu_m + \mu_p - \mu_m \mu_p) \frac{|\hat{y}_n|}{|y_n|} \right] + 2\mu_p - \mu_p^2 + 2\mu_m \frac{|\hat{y}_n|^2}{|y_n|^2} + 2(\mu_p - \mu_m) \frac{|\hat{y}_n|}{|y_n|} - \left(\mu_m - \mu_p - \mu_m \frac{|\hat{y}_n|}{|y_n|} \right)^2$$

being a linear function of $\cos \theta_n \in [-1, 1]$ which attains the extrema when $\cos \theta_n$ attains the extrema. Therefore, if $g(1) > 0$ and $g(-1) > 0$, then $g(\cos \theta_n) > 0, \forall \theta_n \in [0, \pi]$, so that $2\Re[\mu_l] - |\mu_l|^2 > 0$ always holds.

Now the sufficient condition for $g(1) > 0$ and $g(-1) > 0$ can be deduced. It is obvious that $g(1) > 0$ if and only if $0 < \mu_m < 2$. Also note that

$$g(-1) = (2\mu_m - \mu_m^2) \frac{|\hat{y}_n|^2}{|y_n|^2} + 2(\mu_m^2 - 2\mu_m \mu_p + 2\mu_p) \frac{|\hat{y}_n|}{|y_n|} + (2\mu_p - \mu_m)(2 + \mu_m - 2\mu_p)$$

When $|\hat{y}_n| = 0$, it can be proved that $g(-1) > 0$ if and only if $\frac{\mu_m}{2} < \mu_p < 1 + \frac{\mu_m}{2}$. Under the condition $\frac{\mu_m}{2} < \mu_p < 1 + \frac{\mu_m}{2}$, $g(-1)$ increases with $\frac{|\hat{y}_n|}{|y_n|} \in [0, +\infty)$, so $g(-1) > 0$ for all y_n and $\hat{y}_n$. This yields a sufficient convergence condition for

QLMMP:

$$0 < \mu_m < 2, \quad \frac{1}{2}\mu_m < \mu_p < 1 + \frac{1}{2}\mu_m. \quad (6.12)$$

Practically, the step sizes can exceed the sufficient condition given by (6.12), as shown in Section 6.5.

6.4 Special Cases of QLMMP

1) For $\mu_m = \mu_p$, the QLMMP is equivalent to the QLMS with a real-valued step size μ_m . The sufficient convergence condition in (6.12) reduces to $0 < \mu_m < 2$, and the steady-state EMSE is obtained from (5.31) as $(2 - \mu_m)^{-1} \mu_m \sigma_v^2$.

2) For $\mu_m \neq 0$, $\mu_p = 0$, the QLMMP becomes a magnitude-only filtering algorithm which is equivalent to the QLMS with a real-valued step size μ_m under the constraint that the estimate $\hat{y}_n$ always aligns with the desired signal y_n , that is, $\theta_n \equiv 0$, $n = 1, 2, \dots$. The sufficient convergence condition in (6.12) then reduces to $0 < \mu_m < 2$, while the steady-state EMSE is obtained from (5.31) as $(2 - \mu_m)^{-1} \mu_m \sigma_v^2$.

3) For $\mu_m = 0$, $\mu_p \neq 0$, the QLMMP reduces to a phase filtering algorithm for which the following holds:

$$\hat{y}_{n+1} = \hat{y}_n + \mu_p (y_n - |\hat{y}_n|^{-1} \hat{y}_n) \quad (6.13)$$

- If $\mu_p = |\hat{y}_n| |y_n|^{-1}$, phase estimation is completed in one iteration, regardless of the value of phase difference. This is a fundamental advantage over the set of LMP algorithms introduced in [90, 94], for which the phase change in one iteration must be less than $\pi/2$.
- If $\mu_p < |\hat{y}_n| |y_n|^{-1}$, then $|\hat{y}_{n+1}| < |\hat{y}_n|$, and the phase difference θ_n monotonically converges to zero.
- If $\mu_p > |\hat{y}_n| |y_n|^{-1}$, then $\hat{y}_{n+1}$ and $\hat{y}_n$ lie on two sides of y_n on the plane shown in Figure 6.1, and $|\hat{y}_{n+1}| > |\hat{y}_n|$, so that the magnitude of the estimate incre-

ases until $\mu_p \leq |\hat{y}_n| |y_n|^{-1}$. This adaptive behaviour indicates the stability of the phase filtering algorithm for all positive μ_p .

When $\mu_m = 0$, $\mu_p \neq 0$, and for stationary signals, the steady-state condition is $\angle \hat{y}_n \approx \angle y_n$, $|y_n| \approx |y_{n+1}|$, $|\hat{y}_n| \approx |\hat{y}_{n+1}|$. Therefore, after each weight update, a magnitude adjustment to the output can be implemented by $\hat{y}_{n+1}^{\text{new}} = |y_n| |\hat{y}_n|^{-1} \hat{y}_{n+1}$, resulting in the complete estimation of both phase and magnitude of y_{n+1} . This algorithm, termed modified QLMMP (MQLMMP), can be used as an approximate alternative, but with superior stability, to the QLMS.

6.5 Simulations

The considered class of QLMS algorithms were used to estimate the frequency of a simulated balanced three-phase power system for which the voltages at time n are given by [73]

$$v_{m,n} = V \sin \left(2\pi f T n + \varphi + \frac{2\pi}{3} m - \frac{2\pi}{3} \right), \quad m = 1, 2, 3 \quad (6.14)$$

where V is the instantaneous magnitude, φ the initial phase, f the frequency, and T the sampling interval. The three voltages can be combined into a pure quaternion given by [72]

$$\begin{aligned} q_n &= v_1 v_{1,n} + j v_2 v_{2,n} + \kappa v_3 v_{3,n} \\ &= \sqrt{1.5} V [\zeta \sin(2\pi f T n + \varphi) + \zeta' \cos(2\pi f T n + \varphi)] \\ &= \sqrt{1.5} V e^{-\zeta''(2\pi f T n + \varphi)} \zeta' \end{aligned} \quad (6.15)$$

where $\zeta = \sqrt{\frac{2}{3}} (\iota - \frac{1}{2}j - \frac{1}{2}\kappa)$, $\zeta' = \sqrt{0.5} (j - \kappa)$, $\zeta'' = \sqrt{\frac{1}{3}} (\iota + j + \kappa)$ are the imaginary units. (6.15) leads to

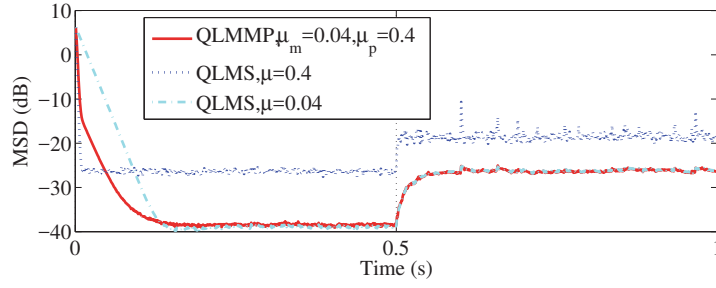
$$q_{n+1} = e^{-\zeta'' 2\pi f T} q_n. \quad (6.16)$$

Through the identification of the autoregressive (AR) system in (6.16), the frequency of the three-phase power system can be adaptively tracked, and one-step ahead prediction of the voltages can be achieved. A three-phase power system was simulated, for which $f = 50$ Hz, $V = 1$ V, $T = 1$ ms, and the following two cases of voltage measurements corrupted by different time-varying noise on magnitude and phase were considered.

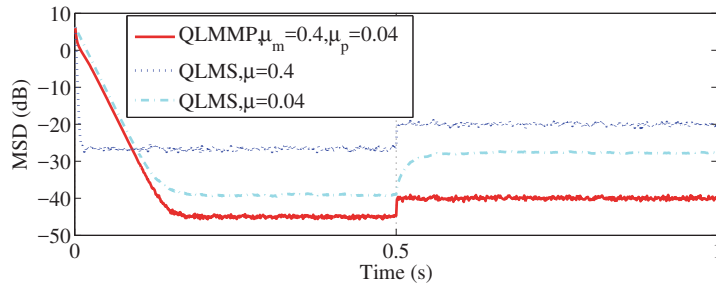
Signal 1: The magnitudes of the three voltages in (6.14) were corrupted by the same white Gaussian noise, a_n , with mean 1 and variance changing from 0.01 to 0.04 at 0.5 s. The quaternion-valued measurements of the voltages were additionally subject to additive noise, hence $\sqrt{1.5}a_n V e^{-\zeta''(2\pi f T n + \varphi)} \zeta' + v_n$, where v_n was quaternion-valued circular white Gaussian noise with variance 0.0003.

Signal 2: The phases of the three voltages in (6.14) were corrupted by the same zero-mean white Gaussian noise, β_n , with variance changing from 0.01 to 0.04 at 0.5 s. The quaternion-valued measurements of the voltages were additionally subject to additive noise, hence $\sqrt{1.5}V e^{-\zeta''(2\pi f T n + \varphi + \beta_n)} \zeta' + v_n$.

Figure 6.2 illustrates the evolution of the MSD averaged over 200 simulation runs for three algorithms: 1) normalised QLMMP with $\mu_m = 0.04$ and $\mu_p = 0.4$, 2) normalised QLMS with $\mu = 0.4$, and 3) normalised QLMS with $\mu = 0.04$. As shown in Figure 6.2 (a), for Signal 1, the QLMMP achieved a lower steady-state MSD than the QLMS with $\mu = 0.4$. Although the QLMS with $\mu = 0.04$ had a similar steady-state MSD to the QLMMP, its convergence was slower. Figure 6.2 (b) indicates that the QLMMP attained the least steady-state MSD among the three algorithms and also converged faster than the QLMS with $\mu = 0.04$. The performances of QLMMP and QLMP were compared in one-step ahead prediction of the phase of the quaternion variable, q_n . Figure 6.3 shows the evolution of the phase error power averaged over 200 simulation runs for two algorithms: 1) normalised QLMMP with $\mu_m = 0$ and $\mu_p = 0.4$, and 2) normalised QLMP with $\mu = 0.4$ [94]. Observe that the QLMMP converged faster than the QLMP in the phase estimation. The

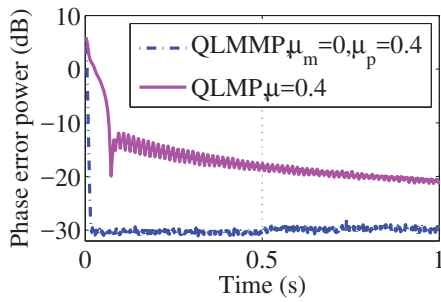


(a) Signal 1

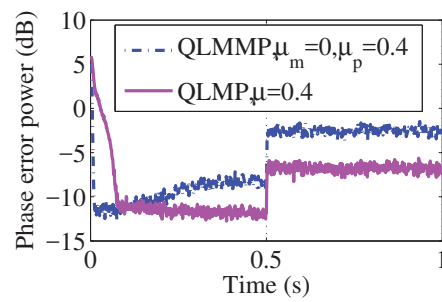


(b) Signal 2

Figure 6.2: MSD curves for identification of the three-phase power system using the normalised QLMMP and normalised QLMS.



(a) Signal 1



(b) Signal 2

Figure 6.3: Phase error power curves for filtering three-phase voltages using the normalised QLMMP and normalised QLMP.

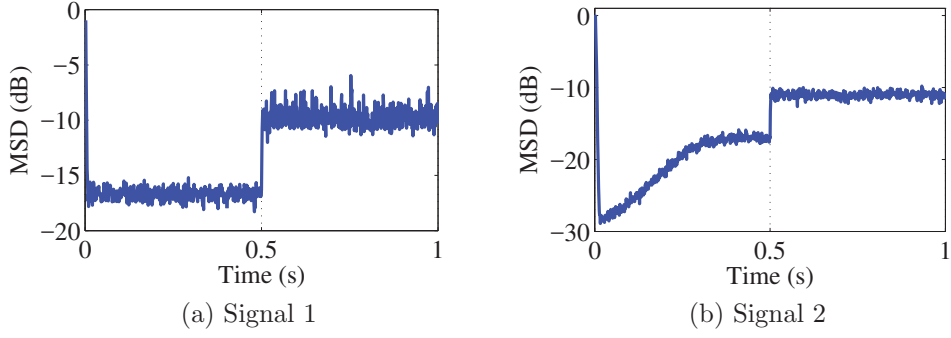


Figure 6.4: MSD curves for one-step ahead prediction of three-phase voltages using the normalised MQLMMP.

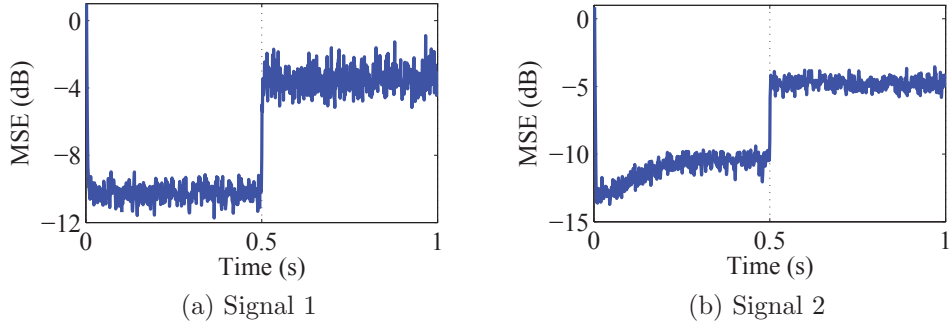


Figure 6.5: MSE curves for one-step ahead prediction of three-phase voltages using the normalised MQLMMP.

normalised MQLMMP was tested for one-step ahead prediction of the quaternion variable, q_n . Figures 6.4 and 6.5 show the evolution of the MSD and MSE averaged over 200 simulation runs for the normalised MQLMMP with $\mu_p = 0.4$ introduced in Section 6.4. Figures 6.2 and 6.4 indicate that, as expected, the MQLMMP can achieve one-step prediction of q_n , although with lower accuracy than the QLMMP and QLMS. Since the phase cost function, $J_{p,n}$ in (6.2), is not quadratic, the estimation errors of the QLMMP and MQLMMP decreased to a lower point before reaching their steady state when the signal is contaminated by high noise in phase, as shown in Figure 6.2 (b), 6.4 (b), and 6.5 (b).

6.6 Summary

This chapter has defined a novel cost function for quaternion adaptive filtering which represents a weighted combination of the magnitude error and phase error. The QLMMP algorithm has been derived based on such a cost function, and theoretical analysis has been conducted to find a range for the step sizes that guarantees the convergence. The QLMMP has been shown to outperform the QLMS and the QLMP by virtue of the enhanced degrees of freedom in the cost function.

Chapter 7

Low-Complexity Quaternion Adaptive Filters

Overview

Widely linear quaternion estimation caters for both proper and improper quaternion signals, but are computationally expensive owing to the use of augmented variables and statistics. To reduce the computation cost while maintaining the performance level, this chapter proposes a four-channel estimation framework as an efficient alternative to quaternion widely linear estimation. This is achieved by using four linear models to estimate the four components of quaternion signals. Furthermore, any of the four channels is able to replace a strictly linear quaternion estimator when estimating strictly linear systems. The proposed method is shown to reduce computational complexity and provide more flexible algorithms, while preserving the physical meaning inherent in the quaternion domain. The framework is applied to quaternion MMSE estimation to yield the reduced-complexity versions of the QLMS, quaternion recursive least squares (QRLS), and quaternion nonlinear gradient decent (QNGD) algorithms. The convergence performance of the proposed QLMS algorithm is investigated.

7.1 Introduction

Advances in quaternion statistics have established that in order to capture complete second-order statistics of quaternion signals, widely linear estimators which use the complementary covariances, in addition to the standard covariance, are

needed for estimation of general quaternion signals [29, 98]. However, the widely linear approach requires four times the number of parameters to update, inducing higher computational complexity. To reduce the computation cost, efficient algorithms casting the computations from $\mathbb{H}$ to $\mathbb{R}$ have been proposed, for example, the reduced-complexity widely linear quaternion least mean square (RC-WL-QLMS) algorithm employs a quaternion-valued weight vector and a real-valued input vector [99], while the multi-channel LMS (MLMS) algorithm exploits the duality between $\mathbb{R}^4$ and $\mathbb{H}$ [100]. Despite computational efficiency, these algorithms have no means of exploring the physical meaning inherent in $\mathbb{H}$, as the input and weight vectors no longer reside in $\mathbb{H}$.

In order to obtain physically meaningful estimates of quaternion signals at low computation cost, this chapter extends the recently proposed complex dual channel estimation [101] to a four-channel quaternion estimation framework, which comprises four sub-estimators for the four components of the quaternion. In this way, four degrees of freedom are provided to capture complete second-order statistical information; the so introduced estimation model and weight update rule reduce the computational cost compared with the conventional widely linear estimation, while achieving equivalent estimation performance. In this way, the physical meaning inherent to the quaternion domain is also preserved, together with enhanced flexibility as the convergence of the four sub-estimators can be controlled individually. When estimating strictly linear quaternion systems, the proposed four-channel estimation simplifies to a single-channel problem, which can also be used as an efficient alternative to conventional strictly linear estimation. For rigour, a class of four-channel quaternion MMSE estimation algorithms are derived, including the four-channel quaternion least mean square (FC-QLMS), the four-channel quaternion recursive least squares (FC-QRLS), and the four-channel quaternion nonlinear gradient decent (FC-QNGD). The convergence performance and computational complexity of the FC-QLMS algorithm are established, and an adaptive step size strategy based on the Barzilai-Borwein method is introduced.

ced. Simulations on synthetic and real-world multi-dimensional signals support the analysis.

7.2 Four-channel Quaternion Estimation Model

Consider a quaternion-valued desired signal, y , given by a widely linear system corrupted by independent noise, v , in the form

$$y = \mathbf{h}^H \mathbf{x} + \mathbf{g}^H \mathbf{x}^i + \mathbf{u}^H \mathbf{x}^j + \mathbf{v}^H \mathbf{x}^\kappa + v \quad (7.1)$$

where $\mathbf{h}, \mathbf{g}, \mathbf{u}, \mathbf{v} \in \mathbb{H}^{L \times 1}$ are the true weight vectors. According to the relationship between the components of a quaternion and its involutions, shown in (2.8)-(2.11), the four real-valued components of y can be represented by four linear models, as

$$\begin{aligned} \Re[y] &= \Re \left[(\mathbf{h} + \mathbf{g}^i + \mathbf{u}^j + \mathbf{v}^\kappa)^H \mathbf{x} + v \right] \\ \Im_i[y] &= \Im_i \left[(\mathbf{h} + \mathbf{g}^i - \mathbf{u}^j - \mathbf{v}^\kappa)^H \mathbf{x} + v \right] \\ \Im_j[y] &= \Im_j \left[(\mathbf{h} - \mathbf{g}^i + \mathbf{u}^j - \mathbf{v}^\kappa)^H \mathbf{x} + v \right] \\ \Im_\kappa[y] &= \Im_\kappa \left[(\mathbf{h} - \mathbf{g}^i - \mathbf{u}^j + \mathbf{v}^\kappa)^H \mathbf{x} + v \right] \end{aligned} \quad (7.2)$$

Therefore, the desired signal (7.1) can be expressed as

$$y = \Re \left[\mathbf{w}_1^H \mathbf{x} \right] + i \Im_i \left[\mathbf{w}_i^H \mathbf{x} \right] + j \Im_j \left[\mathbf{w}_j^H \mathbf{x} \right] + \kappa \Im_\kappa \left[\mathbf{w}_\kappa^H \mathbf{x} \right] + v \quad (7.3)$$

where $\mathbf{w}_1, \mathbf{w}_i, \mathbf{w}_j, \mathbf{w}_\kappa$ are quaternion-valued weight vectors. This motivates us to separately estimate the four real-valued components, $\Re[\hat{y}_1], \Im_i[\hat{y}_i], \Im_j[\hat{y}_j], \Im_\kappa[\hat{y}_\kappa]$, of y , in the form

$$\hat{y}_1 = \hat{\mathbf{w}}_1^H \mathbf{x}, \quad \hat{y}_i = \hat{\mathbf{w}}_i^H \mathbf{x}, \quad \hat{y}_j = \hat{\mathbf{w}}_j^H \mathbf{x}, \quad \hat{y}_\kappa = \hat{\mathbf{w}}_\kappa^H \mathbf{x} \quad (7.4)$$

which are respectively referred to as channel 1, $\imath$, j and κ . Then, these components can be combined into a quaternion-valued estimate

$$\hat{y} = \Re[\hat{y}_1] + \imath \Im[\hat{y}_\imath] + j \Im[\hat{y}_j] + \kappa \Im[\hat{y}_\kappa] \quad (7.5)$$

for which the estimation error is given by

$$e = y - \hat{y} = e_1 + \imath e_\imath + j e_j + \kappa e_\kappa$$

where

$$e_1 = \Re[y] - \Re[\hat{y}_1], \quad e_\imath = \Im_\imath[y] - \Im_\imath[\hat{y}_\imath], \quad e_j = \Im_j[y] - \Im_j[\hat{y}_j], \quad e_\kappa = \Im_\kappa[y] - \Im_\kappa[\hat{y}_\kappa]$$

are the respective estimation errors in the four channels (components). Refer to the equations in (7.4) and (7.5) as the four-channel linear estimation model, which avoids the summation of the four quaternion quantities in the widely linear estimation model (2.19) and consequently reduces the computation cost.

7.3 Four-channel Quaternion MMSE Estimation

On the basis of the four-channel linear model in (7.4) and (7.5), MMSE estimation of quaternion signals can be undertaken by analysing the four individual channels. Let $\hat{\mathbf{w}}_{1,o}$ denote the optimal estimate of the system weight $\mathbf{w}_1$. Then, the optimal estimate of $\Re[y]$ is given by

$$\Re[\hat{\mathbf{w}}_{1,o}^H \mathbf{x}] = \frac{1}{4} \left[\hat{\mathbf{w}}_{1,o}^H \mathbf{x} + \hat{\mathbf{w}}_{1,o}^{\imath H} \mathbf{x}^\imath + \hat{\mathbf{w}}_{1,o}^{jH} \mathbf{x}^j + \hat{\mathbf{w}}_{1,o}^{\kappa H} \mathbf{x}^\kappa \right]$$

which is the projection of $\Re[y]$ onto a Hilbert subspace spanned by the involution basis, and obeys the orthogonality principle

$$e_1 \perp \mathbf{x} \quad e_1 \perp \mathbf{x}^\imath \quad e_1 \perp \mathbf{x}^j \quad e_1 \perp \mathbf{x}^\kappa$$

where “ $\perp$ ” refers to orthogonality. This orthogonality condition is equivalent to

$$E \{ \mathbf{x} e_1^* \} = E \{ \mathbf{x}^i e_1^* \} = E \{ \mathbf{x}^j e_1^* \} = E \{ \mathbf{x}^\kappa e_1^* \} = \mathbf{0} \quad (7.6)$$

Substituting $e_1 = \Re \left[(\mathbf{w}_1 - \hat{\mathbf{w}}_{1,o})^H \mathbf{x} \right] + \Re[v]$ into (7.6) yields

$$E \{ \mathbf{x} \mathbf{x}^{aH} \} \begin{bmatrix} \hat{\mathbf{w}}_{1,o} \\ \hat{\mathbf{w}}_{1,o}^i \\ \hat{\mathbf{w}}_{1,o}^j \\ \hat{\mathbf{w}}_{1,o}^\kappa \end{bmatrix} = E \{ \mathbf{x} \mathbf{x}^{aH} \} \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_1^i \\ \mathbf{w}_1^j \\ \mathbf{w}_1^\kappa \end{bmatrix}$$

which implies $\hat{\mathbf{w}}_{1,o} = \mathbf{w}_1$. Similarly, the remaining three optimal estimated weights can be derived as $\hat{\mathbf{w}}_{i,o} = \mathbf{w}_i$, $\hat{\mathbf{w}}_{j,o} = \mathbf{w}_j$, $\hat{\mathbf{w}}_{\kappa,o} = \mathbf{w}_\kappa$. This indicates that the four-channel estimator is the optimal estimator for the desired signal shown in (7.1).

For $\mathbf{w}_1 = \mathbf{w}_i = \mathbf{w}_j = \mathbf{w}_\kappa$, Equation (7.3) reduces to a strictly linear model $y = \mathbf{w}_1^H \mathbf{x} + v$, and the optimal estimated weight vectors of the four-channel estimator are $\hat{\mathbf{w}}_{1,o} = \hat{\mathbf{w}}_{i,o} = \hat{\mathbf{w}}_{j,o} = \hat{\mathbf{w}}_{\kappa,o} = \mathbf{w}_1$. Therefore, a single sub-channel is sufficient for the estimation of the weight vector $\mathbf{w}_1$, and the signal estimate is consequently given by

$$\hat{y} = \Re \left[\hat{\mathbf{w}}_1^H \mathbf{x} \right] + i \Im_i \left[\hat{\mathbf{w}}_1^H \mathbf{x} \right] + j \Im_j \left[\hat{\mathbf{w}}_1^H \mathbf{x} \right] + \kappa \Im_\kappa \left[\hat{\mathbf{w}}_1^H \mathbf{x} \right] = \hat{\mathbf{w}}_1^H \mathbf{x}$$

This reduced form of the four-channel estimation is referred to as single-channel estimation.

Next, three adaptive algorithms for quaternion MMSE estimation within the proposed four-channel framework will be introduced.

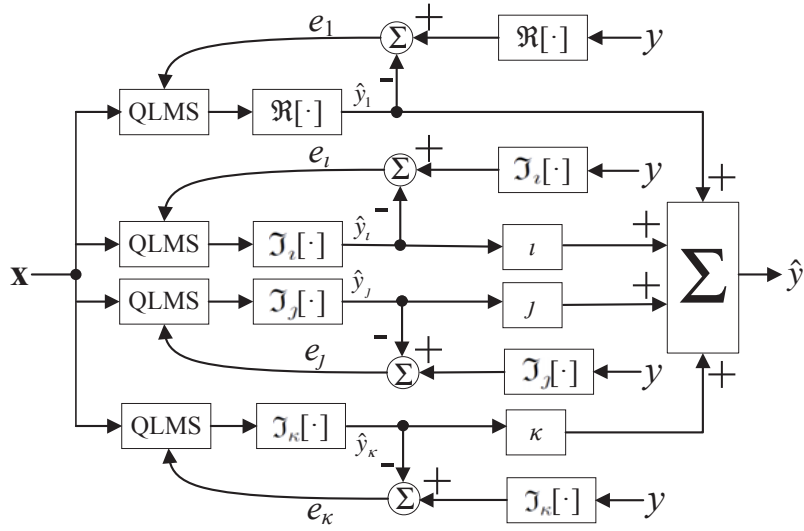


Figure 7.1: Architecture of the FC-QLMS algorithm.

7.3.1 FC-QLMS Algorithm

Based on the four-channel linear estimation model in (7.4) and (7.5), a FC-QLMS algorithm shall be derived. As shown in Figure 7.1, the desired signal at the n -th iteration, y_n , is estimated as

$$\hat{y}_n = \Re [\hat{\mathbf{w}}_{1,n}^H \mathbf{x}_n] + i \Im [\hat{\mathbf{w}}_{i,n}^H \mathbf{x}_n] + j \Im [\hat{\mathbf{w}}_{j,n}^H \mathbf{x}_n] + \kappa \Im [\hat{\mathbf{w}}_{\kappa,n}^H \mathbf{x}_n]$$

while the weights are updated to minimise the cost function

$$J_n^{\text{LMS}} = |y_n - \hat{y}_n|^2 = (e_{1,n}^2 + e_{i,n}^2 + e_{j,n}^2 + e_{\kappa,n}^2)$$

where

$$\begin{aligned} e_{1,n} &= \Re [y_n] - \Re [\hat{\mathbf{w}}_{1,n}^H \mathbf{x}_n] \\ e_{i,n} &= \Im [y_n] - \Im [\hat{\mathbf{w}}_{i,n}^H \mathbf{x}_n] \\ e_{j,n} &= \Im [y_n] - \Im [\hat{\mathbf{w}}_{j,n}^H \mathbf{x}_n] \\ e_{\kappa,n} &= \Im [y_n] - \Im [\hat{\mathbf{w}}_{\kappa,n}^H \mathbf{x}_n] \end{aligned}$$

As the maximum change of J_n^{LMS} occurs when the weight update is collinear with the conjugate derivative of J_n^{LMS} , which has been introduced in Chapter 2, the

weight vectors at the $(n + 1)$ -th iteration can be derived from the n -th iteration as

$$\hat{\mathbf{w}}_{\alpha,n+1} = \hat{\mathbf{w}}_{\alpha,n} - \mu_{\text{fc}} \nabla_{\hat{\mathbf{w}}_{\alpha}^*} J_n^{\text{LMS}}, \quad \alpha \in \{1, \iota, j, \kappa\} \quad (7.7)$$

where μ_{fc} is a positive step size. The derivative in (7.7) is calculated based on the generalised $\mathbb{H}\mathbb{R}$ calculus as

$$\nabla_{\hat{\mathbf{w}}_{\alpha}^*} J_n^{\text{LMS}} = -0.5 e_{\alpha,n} \mathbf{x}_n \alpha^{-1} \quad (7.8)$$

Therefore, the component-wise weight updates are given by

$$\hat{\mathbf{w}}_{\alpha,n+1} = \hat{\mathbf{w}}_{\alpha,n} + \mu_{\text{fc}} e_{\alpha,n} \mathbf{x}_n \alpha^{-1} \quad (7.9)$$

where the constant 0.5 in (7.8) is absorbed into the step size μ_{fc} .

If the desired signal obeys the strictly linear model, the four-channel estimation can be replaced with single-channel estimation, so that the FC-QLMS simplifies to a single-channel QLMS (SC-QLMS) algorithm. The SC-QLMS using channel 1 is therefore given by

$$\begin{aligned} \hat{y}_n &= \hat{\mathbf{w}}_{1,n}^H \mathbf{x}_n \\ \hat{\mathbf{w}}_{1,n+1} &= \hat{\mathbf{w}}_{1,n} + \mu_{\text{fc}} (\Re[y_n] - \Re[\hat{y}_n]) \mathbf{x}_n \end{aligned} \quad (7.10)$$

The proposed model is also scalable, for example, when the quaternion input signal reduces to a complex one, that is, $\mathbf{x} = \Re[\mathbf{x}] + \iota \Im[\mathbf{x}]$, the FC-QLMS algorithm reduces to a complex algorithm, as

$$\begin{aligned} \hat{y}_n &= \Re[\hat{\mathbf{w}}_{1,n}^H \mathbf{x}_n] + \iota \Im[\hat{\mathbf{w}}_{i,n}^H \mathbf{x}_n] \\ \hat{\mathbf{w}}_{1,n+1} &= \hat{\mathbf{w}}_{1,n} + \mu_{\text{fc}} \Re[y - \hat{\mathbf{w}}_{1,n}^H \mathbf{x}_n] \mathbf{x}_n \\ \hat{\mathbf{w}}_{i,n+1} &= \hat{\mathbf{w}}_{i,n} - \mu_{\text{fc}} \Im[y - \hat{\mathbf{w}}_{i,n}^H \mathbf{x}_n] \mathbf{x}_n \end{aligned}$$

which is equivalent to the dual channel complex LMS algorithm in [101].

In addition to the reduction of computational cost owing to the four-channel estimation model, by comparing (7.9) to (2.21), it is observed that the FC-QLMS has a more efficient weight update rule than existing QLMS algorithms [24, 33]. Another advantage of FC-QLMS is that the step sizes of the four channels can be different, and thus the convergence rate of each channel can be controlled individually, which provides the algorithm with more flexibility in practical applications.

7.3.2 FC-QRLS Algorithm

The class of QRLS algorithms include the strictly linear QRLS (SL-QRLS) algorithm given by [102]

$$\begin{aligned}
e_n^p &= y_n - \hat{\mathbf{h}}_{n-1}^H \mathbf{x}_n \\
\mathbf{p}_n &\triangleq \mathbf{P}_{n-1} \mathbf{x}_n (\rho + \mathbf{x}_n^H \mathbf{P}_{n-1} \mathbf{x}_n)^{-1} \\
\hat{\mathbf{h}}_n &= \hat{\mathbf{h}}_{n-1} + \mathbf{p}_n e_n^{p*} \\
\mathbf{P}_n &= \rho^{-1} (\mathbf{P}_{n-1} - \mathbf{p}_n \mathbf{x}_n^H \mathbf{P}_{n-1})
\end{aligned} \tag{7.11}$$

and the widely linear QRLS (WL-QRLS) algorithm given by [102]

$$\begin{aligned}
e_n^p &= y_n - \hat{\mathbf{w}}_{n-1}^{aH} \mathbf{x}_n^a \\
\mathbf{p}_n^a &\triangleq \mathbf{P}_{n-1}^a \mathbf{x}_n^a (\rho + \mathbf{x}_n^{aH} \mathbf{P}_{n-1}^a \mathbf{x}_n^a)^{-1} \\
\hat{\mathbf{w}}_n^a &= \hat{\mathbf{w}}_{n-1}^a + \mathbf{p}_n^a e_n^{p*} \\
\mathbf{P}_n^a &= \rho^{-1} (\mathbf{P}_{n-1}^a - \mathbf{p}_n^a \mathbf{x}_n^{aH} \mathbf{P}_{n-1}^a)
\end{aligned} \tag{7.12}$$

where $\mathbf{P}_n \in \mathbb{H}^{L \times L}$ and $\mathbf{P}_n^a \in \mathbb{H}^{4L \times 4L}$ are the estimated inverses of the covariance and augmented covariance matrices, and $\rho \in (0, 1]$ is the forgetting factor. A FC-QRLS version of the WL-QRLS algorithm can be derived to minimise the cost

function

$$J_n^{\text{RLS}} = \sum_{m=0}^n \rho^{n-m} |e_m|^2$$

$$e_m = y_m - \Re [\hat{\mathbf{w}}_{1,n}^H \mathbf{x}_m] - \imath \Im [\hat{\mathbf{w}}_{\imath,n}^H \mathbf{x}_m] - j \Im [\hat{\mathbf{w}}_{j,n}^H \mathbf{x}_m] - \kappa \Im [\hat{\mathbf{w}}_{\kappa,n}^H \mathbf{x}_m]$$

where the weight update is derived from the WL-QRLS weight in (7.12) according to the isomorphism in (7.2), to yield

$$\hat{\mathbf{w}}_{\alpha,n} = \hat{\mathbf{w}}_{\alpha,n-1} + 4\Psi [e_n^{p*}] \mathbf{r}_n \alpha \quad \alpha \in \{1, \imath, j, \kappa\} \quad (7.13)$$

In (7.13), $\mathbf{r}_n \in \mathbb{H}^{L \times 1}$ contains the first L entries of $\mathbf{p}_n^a$, $\Psi[\cdot] = \Re[\cdot]$ when $\alpha = 1$, $\Psi[\cdot] = \Im[\cdot]$ when $\alpha = \imath$, $\Psi[\cdot] = \Im[\cdot]$ when $\alpha = j$, $\Psi[\cdot] = \Im[\cdot]$ when $\alpha = \kappa$. The weight calculation in (7.13) is more efficient than the weight update of WL-QRLS in (7.12). If the desired signal obeys the strictly linear model, a single sub-channel is sufficient for optimal estimation, leading to a more efficient single-channel QRLS (SC-QRLS) algorithm.

7.3.3 FC-QNGD Algorithm

The class of QNGD algorithms include the strictly linear QNGD (SL-QNGD) algorithm, given by [33, 103]

$$\begin{aligned} q_n &= \hat{\mathbf{h}}_n^H \mathbf{x}_n \\ \hat{y}_n &= \varphi(q_n) \\ \hat{\mathbf{h}}_{n+1} &= \hat{\mathbf{h}}_n + \mu \mathbf{x}_n \sum_{\nu \in \{1, \imath, j, \kappa\}} \frac{\partial \varphi(q_n)^\nu}{\partial q_n} (y_n - \hat{y}_n)^{\nu*} \end{aligned}$$

and the widely linear QNGD (WL-QNGD) algorithm, given by [103]

$$\begin{aligned} q_n &= \hat{\mathbf{w}}_n^{aH} \mathbf{x}_n^a \\ \hat{y}_n &= \varphi(q_n) \\ \hat{\mathbf{w}}_{n+1}^a &= \hat{\mathbf{w}}_n^a + \mu \mathbf{x}_n^a \sum_{\nu \in \{1, \imath, j, \kappa\}} \frac{\partial \varphi(q_n)^\nu}{\partial q_n} (y_n - \hat{y}_n)^{\nu*} \end{aligned}$$

where $\varphi(q_n)$ is a nonlinear function of q_n , $\mu \in \mathbb{R}^+$ a step size.

The FC-QNGD algorithm can also be derived on the basis of the four-channel estimation model. Consider the minimisation of the cost function

$$J_n^{\text{NGD}} = |e_n|^2 = |y_n - \varphi(q_n)|^2$$

where

$$q_n = \Re[\hat{\mathbf{w}}_{1,n}^H \mathbf{x}_n] + \imath \Im[\hat{\mathbf{w}}_{\imath,n}^H \mathbf{x}_n] + \jmath \Im[\hat{\mathbf{w}}_{\jmath,n}^H \mathbf{x}_n] + \kappa \Im[\hat{\mathbf{w}}_{\kappa,n}^H \mathbf{x}_n]$$

The maximum change of J_n^{NGD} occurs when the weight update is collinear with the conjugate derivative of J_n^{NGD} , so the weight vectors can be updated as

$$\hat{\mathbf{w}}_{\alpha,n+1} = \hat{\mathbf{w}}_{\alpha,n} - \mu_{\text{fc}} \nabla_{\hat{\mathbf{w}}_{\alpha}^*} J_n^{\text{NGD}}, \quad \alpha \in \{1, \imath, \jmath, \kappa\}$$

where the derivative is calculated based on the generalised HIR calculus as

$$\nabla_{\hat{\mathbf{w}}_{\alpha}^*} J_n^{\text{NGD}} = \sum_{\nu \in \{1, \imath, \jmath, \kappa\}} \frac{\partial |e_n|^2}{\partial e_n^{\nu}} \frac{\partial e_n^{\nu}}{\partial \hat{\mathbf{w}}_{\alpha,n}^*} = -\frac{1}{8} \sum_{\nu \in \{1, \imath, \jmath, \kappa\}} e_n^{\nu*} \frac{\partial \varphi(q_n)^{\nu}}{\partial \Psi[q_n]} \mathbf{x}_n \alpha^{-1} \quad (7.14)$$

Therefore, the weight update is given by

$$\hat{\mathbf{w}}_{\alpha,n+1} = \hat{\mathbf{w}}_{\alpha,n} + \mu_{\text{fc}} \sum_{\nu \in \{1, \imath, \jmath, \kappa\}} e_n^{\nu*} \frac{\partial \varphi(q_n)^{\nu}}{\partial \Psi[q_n]} \mathbf{x}_n \alpha^{-1} \quad (7.15)$$

where the constant $1/8$ in (7.14) is absorbed into the step size μ_{fc} .

If $\varphi(q)$ is a split quaternion nonlinear function $\varphi(q) = \varphi_1(q) + \imath \varphi_{\imath}(q) + \jmath \varphi_{\jmath}(q) + \kappa \varphi_{\kappa}(q)$, where $\varphi_1(q)$ is a real-valued nonlinear function applied to $\Re[q]$, $\varphi_{\imath}(q)$ to $\Im_{\imath}[q]$, $\varphi_{\jmath}(q)$ to $\Im_{\jmath}[q]$, and $\varphi_{\kappa}(q)$ to $\Im_{\kappa}[q]$, then the cost function can be written as $J_n^{\text{NGD}} = \sum_{\alpha \in \{1, \imath, \jmath, \kappa\}} e_{\alpha,n}^2$, where $e_{\alpha,n} = \Psi[y_n] - \varphi_{\alpha}(q_n)$ are split errors. The

conjugate derivative of J_n^{NGD} is derived as

$$\nabla_{\hat{\mathbf{w}}_\alpha^*} J_n^{\text{NGD}} = \frac{\partial J_n^{\text{NGD}}}{\partial e_{\alpha,n}} \frac{\partial e_{\alpha,n}}{\partial \hat{\mathbf{w}}_{\alpha,n}^*} = -0.5e_{\alpha,n} \frac{\partial \varphi_\alpha(q_n)}{\partial \Psi[q_n]} \mathbf{x}_n \alpha^{-1} \quad (7.16)$$

and the weights are updated as

$$\hat{\mathbf{w}}_{\alpha,n+1} = \hat{\mathbf{w}}_{\alpha,n} - \mu_{\text{fc}} \nabla_{\hat{\mathbf{w}}_\alpha^*} J_n^{\text{NGD}} = \hat{\mathbf{w}}_{\alpha,n} + \mu_{\text{fc}} e_{\alpha,n} \frac{\partial \varphi_\alpha(q_n)}{\partial \Psi[q_n]} \mathbf{x}_n \alpha^{-1} \quad (7.17)$$

where the constant 0.5 in (7.16) is absorbed into the step size μ_{fc} .

The FC-QNGD weight update is obviously less complicated than the WL-QNGD weight update. When the desired signal obeys the strictly linear model, a single channel is sufficient for optimal estimation, yielding a more efficient single-channel QNGD (SC-QNGD) algorithm.

7.4 Performance of FC-QLMS

The performance of the FC-QLMS algorithm is next compared with the QLMS algorithms.

7.4.1 Equivalence of the FC-QLMS and WL-QLMS

Recently proposed QLMS algorithms include the strictly linear quaternion least mean square (SL-QLMS) algorithm [33]

$$\begin{aligned} \hat{y}_n &= \hat{\mathbf{h}}_n^H \mathbf{x}_n \\ \hat{\mathbf{h}}_{n+1} &= \hat{\mathbf{h}}_n + \mu \mathbf{x}_n (y_n - \hat{y}_n)^* \end{aligned} \quad (7.18)$$

and the widely linear quaternion least mean square (WL-QLMS) algorithm [24]

$$\begin{aligned} \hat{y}_n &= \hat{\mathbf{w}}_n^{aH} \mathbf{x}_n^a \\ \hat{\mathbf{w}}_{n+1}^a &= \hat{\mathbf{w}}_n^a + \mu \mathbf{x}_n^a (y_n - \hat{y}_n)^* \end{aligned} \quad (7.19)$$

where μ is a positive step size.

Let $\Psi[\cdot] \in \{\Re[\cdot], \Im_i[\cdot], \Im_j[\cdot], \Im_\kappa[\cdot]\}$, then, the four components of the FC-QLMS estimate have a unified form given by

$$\begin{aligned}\Psi[\hat{y}_n] &= \Psi[\hat{\mathbf{w}}_{\alpha,k}^H \mathbf{x}_n] \\ &= \Psi[(\hat{\mathbf{w}}_{\alpha,0}^H + \mu_{\text{fc}} e_{\alpha,0} \alpha \mathbf{x}_0^H + \cdots + \mu_{\text{fc}} e_{\alpha,n-1} \alpha \mathbf{x}_{n-1}^H) \mathbf{x}_n] \\ &= \Psi[\mathbf{w}_{\alpha,0}^H \mathbf{x}_n] + \mu_{\text{fc}} (\Re[\mathbf{x}_0^H \mathbf{x}_n] e_{\alpha,0} + \cdots + \Re[\mathbf{x}_{n-1}^H \mathbf{x}_n] e_{\alpha,n-1})\end{aligned}$$

where $\alpha = 1$ when $\Psi[\cdot] = \Re[\cdot]$, $\alpha = i$ when $\Psi[\cdot] = \Im_i[\cdot]$, $\alpha = j$ when $\Psi[\cdot] = \Im_j[\cdot]$, $\alpha = \kappa$ when $\Psi[\cdot] = \Im_\kappa[\cdot]$. Similarly, the components of the WL-QLMS estimate are represented as

$$\begin{aligned}\Psi[\hat{y}_n] &= \Psi[\hat{\mathbf{h}}_n^H \mathbf{x}_n] + \Psi[\hat{\mathbf{g}}_n^H \mathbf{x}_n^i] + \Psi[\hat{\mathbf{u}}_n^H \mathbf{x}_n^j] + \Psi[\hat{\mathbf{v}}_n^H \mathbf{x}_n^\kappa] \\ &= \Psi[(\hat{\mathbf{h}}_0^H + \mu e_0 \mathbf{x}_0^H + \cdots + \mu e_{n-1} \mathbf{x}_{n-1}^H) \mathbf{x}_n] + \Psi[(\hat{\mathbf{g}}_0^H + \mu e_0 \mathbf{x}_0^{iH} + \cdots + \mu e_{n-1} \mathbf{x}_{n-1}^{iH}) \mathbf{x}_n^i] \\ &\quad + \Psi[(\hat{\mathbf{u}}_0^H + \mu e_0 \mathbf{x}_0^{jH} + \cdots + \mu e_{n-1} \mathbf{x}_{n-1}^{jH}) \mathbf{x}_n^j] + \Psi[(\hat{\mathbf{v}}_0^H + \mu e_0 \mathbf{x}_0^{\kappa H} + \cdots + \mu e_{n-1} \mathbf{x}_{n-1}^{\kappa H}) \mathbf{x}_n^\kappa] \\ &= \Psi[\hat{\mathbf{h}}_0^H \mathbf{x}_n + \hat{\mathbf{g}}_0^H \mathbf{x}_n^i + \hat{\mathbf{u}}_0^H \mathbf{x}_n^j + \hat{\mathbf{v}}_0^H \mathbf{x}_n^\kappa] + 4\mu (\Re[\mathbf{x}_0^H \mathbf{x}_n] \Psi[e_0] + \cdots + \Re[\mathbf{x}_{n-1}^H \mathbf{x}_n] \Psi[e_{n-1}])\end{aligned}$$

Remark 7.1. From the above two formulas, the estimates of FC-QLMS and WL-QLMS are identical if $\mu_{\text{fc}} = 4\mu$, and the initial values of estimated weights satisfy $\hat{\mathbf{h}}_0 = \hat{\mathbf{g}}_0 = \hat{\mathbf{u}}_0 = \hat{\mathbf{v}}_0$, $\hat{\mathbf{w}}_{1,0} = 4\Re[\hat{\mathbf{h}}_0]$, $\hat{\mathbf{w}}_{i,0} = 4i\Im_i[\hat{\mathbf{h}}_0]$, $\hat{\mathbf{w}}_{j,0} = 4j\Im_j[\hat{\mathbf{h}}_0]$, $\hat{\mathbf{w}}_{\kappa,0} = 4\kappa\Im_\kappa[\hat{\mathbf{h}}_0]$. This also implies that the FC-QLMS and WL-QLMS algorithms have identical steady-state MSE if $\mu_{\text{fc}} = 4\mu$.

7.4.2 Transient Performance

For the FC-QLMS algorithm, the weight error vector in channel α at the n -th iteration is defined as $\tilde{\mathbf{w}}_{\alpha,n} \triangleq \mathbf{w}_\alpha - \hat{\mathbf{w}}_{\alpha,n}$, so that the estimation error in channel α becomes $e_{\alpha,n} = \Psi[\tilde{\mathbf{w}}_{\alpha,n}^H \mathbf{x}_n + v_n]$, and the weight error vector at the next iteration can be expressed as

$$\tilde{\mathbf{w}}_{\alpha,n+1} = \tilde{\mathbf{w}}_{\alpha,n} - \mu_{\text{fc}} e_{\alpha,n} \mathbf{x}_n \alpha^{-1} \quad (7.20)$$

which yields

$$\tilde{\mathbf{w}}_{1,n+1} = \tilde{\mathbf{w}}_{1,n} - \frac{1}{4}\mu_{\text{fc}}\mathbf{x}_n \left(\mathbf{x}_n^H \tilde{\mathbf{w}}_{1,n} + \mathbf{x}_n^{\iota H} \tilde{\mathbf{w}}_{1,n}^{\iota} + \mathbf{x}_n^{jH} \tilde{\mathbf{w}}_{1,n}^j + \mathbf{x}_n^{\kappa H} \tilde{\mathbf{w}}_{1,n}^{\kappa} + \Re[v_n] \right)$$

$$\tilde{\mathbf{w}}_{\iota,n+1} = \tilde{\mathbf{w}}_{\iota,n} - \frac{1}{4}\mu_{\text{fc}}\mathbf{x}_n \left(\mathbf{x}_n^H \tilde{\mathbf{w}}_{\iota,n} + \mathbf{x}_n^{\iota H} \tilde{\mathbf{w}}_{\iota,n}^{\iota} - \mathbf{x}_n^{jH} \tilde{\mathbf{w}}_{\iota,n}^j - \mathbf{x}_n^{\kappa H} \tilde{\mathbf{w}}_{\iota,n}^{\kappa} + \Im_{\iota}[v_n] \right)$$

$$\tilde{\mathbf{w}}_{j,n+1} = \tilde{\mathbf{w}}_{j,n} - \frac{1}{4}\mu_{\text{fc}}\mathbf{x}_n \left(\mathbf{x}_n^H \tilde{\mathbf{w}}_{j,n} - \mathbf{x}_n^{\iota H} \tilde{\mathbf{w}}_{j,n}^{\iota} + \mathbf{x}_n^{jH} \tilde{\mathbf{w}}_{j,n}^j - \mathbf{x}_n^{\kappa H} \tilde{\mathbf{w}}_{j,n}^{\kappa} + \Im_j[v_n] \right)$$

$$\tilde{\mathbf{w}}_{\kappa,n+1} = \tilde{\mathbf{w}}_{\kappa,n} - \frac{1}{4}\mu_{\text{fc}}\mathbf{x}_n \left(\mathbf{x}_n^H \tilde{\mathbf{w}}_{\kappa,n} - \mathbf{x}_n^{\iota H} \tilde{\mathbf{w}}_{\kappa,n}^{\iota} - \mathbf{x}_n^{jH} \tilde{\mathbf{w}}_{\kappa,n}^j + \mathbf{x}_n^{\kappa H} \tilde{\mathbf{w}}_{\kappa,n}^{\kappa} + \Im_{\kappa}[v_n] \right)$$

Applying the statistical expectation operator to the above four equations yields

$$\begin{aligned} E\{\tilde{\mathbf{w}}_{1,n+1}\} &= \mathbf{A}E\{\tilde{\mathbf{w}}_{1,n}\} - \frac{\mu_{\text{fc}}}{4} \left(\mathbf{C}_{\mathbf{xx}^{\iota}} E\{\tilde{\mathbf{w}}_{1,n}^{\iota}\} + \mathbf{C}_{\mathbf{xx}^j} E\{\tilde{\mathbf{w}}_{1,n}^j\} + \mathbf{C}_{\mathbf{xx}^{\kappa}} E\{\tilde{\mathbf{w}}_{1,n}^{\kappa}\} \right) \\ E\{\tilde{\mathbf{w}}_{\iota,n+1}\} &= \mathbf{A}E\{\tilde{\mathbf{w}}_{\iota,n}\} - \frac{\mu_{\text{fc}}}{4} \left(\mathbf{C}_{\mathbf{xx}^{\iota}} E\{\tilde{\mathbf{w}}_{\iota,n}^{\iota}\} - \mathbf{C}_{\mathbf{xx}^j} E\{\tilde{\mathbf{w}}_{\iota,n}^j\} - \mathbf{C}_{\mathbf{xx}^{\kappa}} E\{\tilde{\mathbf{w}}_{\iota,n}^{\kappa}\} \right) \\ E\{\tilde{\mathbf{w}}_{j,n+1}\} &= \mathbf{A}E\{\tilde{\mathbf{w}}_{j,n}\} - \frac{\mu_{\text{fc}}}{4} \left(-\mathbf{C}_{\mathbf{xx}^{\iota}} E\{\tilde{\mathbf{w}}_{j,n}^{\iota}\} + \mathbf{C}_{\mathbf{xx}^j} E\{\tilde{\mathbf{w}}_{j,n}^j\} - \mathbf{C}_{\mathbf{xx}^{\kappa}} E\{\tilde{\mathbf{w}}_{j,n}^{\kappa}\} \right) \\ E\{\tilde{\mathbf{w}}_{\kappa,n+1}\} &= \mathbf{A}E\{\tilde{\mathbf{w}}_{\kappa,n}\} - \frac{\mu_{\text{fc}}}{4} \left(-\mathbf{C}_{\mathbf{xx}^{\iota}} E\{\tilde{\mathbf{w}}_{\kappa,n}^{\iota}\} - \mathbf{C}_{\mathbf{xx}^j} E\{\tilde{\mathbf{w}}_{\kappa,n}^j\} + \mathbf{C}_{\mathbf{xx}^{\kappa}} E\{\tilde{\mathbf{w}}_{\kappa,n}^{\kappa}\} \right) \\ \mathbf{A} &\triangleq \mathbf{I} - \frac{1}{4}\mu_{\text{fc}}\mathbf{C}_{\mathbf{xx}} \end{aligned} \quad (7.21)$$

the solution to which is dependent on second-order statistics of $\mathbf{x}$. It is useful to discuss the following three special cases prior to considering the general case.

$\mathbb{H}$ -proper signal

If $\mathbf{x}$ is $\mathbb{H}$ -proper, the three complementary covariances vanish, and the four weight error vectors obey the same recursion form given by

$$E\{\tilde{\mathbf{w}}_{\alpha,n+1}\} = \left(\mathbf{I} - \frac{1}{4}\mu_{\text{fc}}\mathbf{C}_{\mathbf{xx}} \right) E\{\tilde{\mathbf{w}}_{\alpha,n}\}, \quad \alpha \in \{1, \iota, j, \kappa\} \quad (7.22)$$

The equivalence between the evolution of the four weight vectors is a consequence of the four components of $\mathbf{x}$ being uncorrelated and with the same variance. The weight error vectors converge if all the eigenvalues of the matrix $(\mathbf{I} - \frac{1}{4}\mu_{\text{fc}}\mathbf{C}_{\mathbf{xx}})$ are

within $(-1, 1)$, that is,

$$0 < \mu_{\text{fc}} < \frac{8}{\lambda_{\text{max}}}$$

where λ_{max} is the maximum eigenvalue of $\mathbf{C}_{\mathbf{xx}}$.

According to the expression for the SL-QLMS algorithm in (7.18), the weight error vector recursion of SL-QLMS is given by

$$E \left\{ \tilde{\mathbf{h}}_{n+1} \right\} = (\mathbf{I} - \mu \mathbf{C}_{\mathbf{xx}}) E \left\{ \tilde{\mathbf{h}}_n \right\} \quad (7.23)$$

Remark 7.2. For an $\mathbb{H}$ -proper quaternion signal $\mathbf{x}$, the ratio between the convergence rate of the FC-QLMS and the SL-QLMS is $\frac{\mu_{\text{fc}}}{4\mu}$. If $\mu_{\text{fc}} = 4\mu$, these two algorithms exhibit identical weight error evolution, which is illustrated in Figure 7.2 showing the averaged weight trajectories along the error surfaces while estimating a moving average (MA) process.

$\mathbb{C}^\eta$ -improper signal

For a $\mathbb{C}^\eta$ -improper quaternion signal $\mathbf{x}$, use η' and η'' to denote any other two imaginary units among $\imath, j, \kappa$. Then the weight error vector recursions in (7.21) reduce to

$$\begin{aligned} E \left\{ \tilde{\mathbf{w}}_{1,n+1} \right\} &= \left(\mathbf{I} - \frac{\mu_{\text{fc}}}{4} \mathbf{C}_{\mathbf{xx}} \right) E \left\{ \tilde{\mathbf{w}}_{1,n} \right\} - \frac{\mu_{\text{fc}}}{4} \mathbf{C}_{\mathbf{xx}\eta} E \left\{ \tilde{\mathbf{w}}_{1,n}^\eta \right\} \\ E \left\{ \tilde{\mathbf{w}}_{\eta,n+1} \right\} &= \left(\mathbf{I} - \frac{\mu_{\text{fc}}}{4} \mathbf{C}_{\mathbf{xx}} \right) E \left\{ \tilde{\mathbf{w}}_{\eta,n} \right\} - \frac{\mu_{\text{fc}}}{4} \mathbf{C}_{\mathbf{xx}\eta} E \left\{ \tilde{\mathbf{w}}_{\eta,n}^\eta \right\} \\ E \left\{ \tilde{\mathbf{w}}_{\eta',n+1} \right\} &= \left(\mathbf{I} - \frac{\mu_{\text{fc}}}{4} \mathbf{C}_{\mathbf{xx}} \right) E \left\{ \tilde{\mathbf{w}}_{\eta',n} \right\} + \frac{\mu_{\text{fc}}}{4} \mathbf{C}_{\mathbf{xx}\eta} E \left\{ \tilde{\mathbf{w}}_{\eta',n}^\eta \right\} \\ E \left\{ \tilde{\mathbf{w}}_{\eta'',n+1} \right\} &= \left(\mathbf{I} - \frac{\mu_{\text{fc}}}{4} \mathbf{C}_{\mathbf{xx}} \right) E \left\{ \tilde{\mathbf{w}}_{\eta'',n} \right\} + \frac{\mu_{\text{fc}}}{4} \mathbf{C}_{\mathbf{xx}\eta} E \left\{ \tilde{\mathbf{w}}_{\eta'',n}^\eta \right\} \end{aligned}$$

Observe that the evolutions of $\tilde{\mathbf{w}}_1$ and $\tilde{\mathbf{w}}_\eta$ are the same, while $\tilde{\mathbf{w}}_{\eta'}$ and $\tilde{\mathbf{w}}_{\eta''}$ also exhibit identical evolution, and the paths of the four weight errors are symmetric about the weight error path of SL-QLMS in (7.23). This is illustrated in Figure 7.3 showing the averaged weight trajectories in the estimation of an MA process. Figure 7.3 can also be interpreted by the Cayley-Dickson construction of $\mathbb{C}^\eta$ -improper

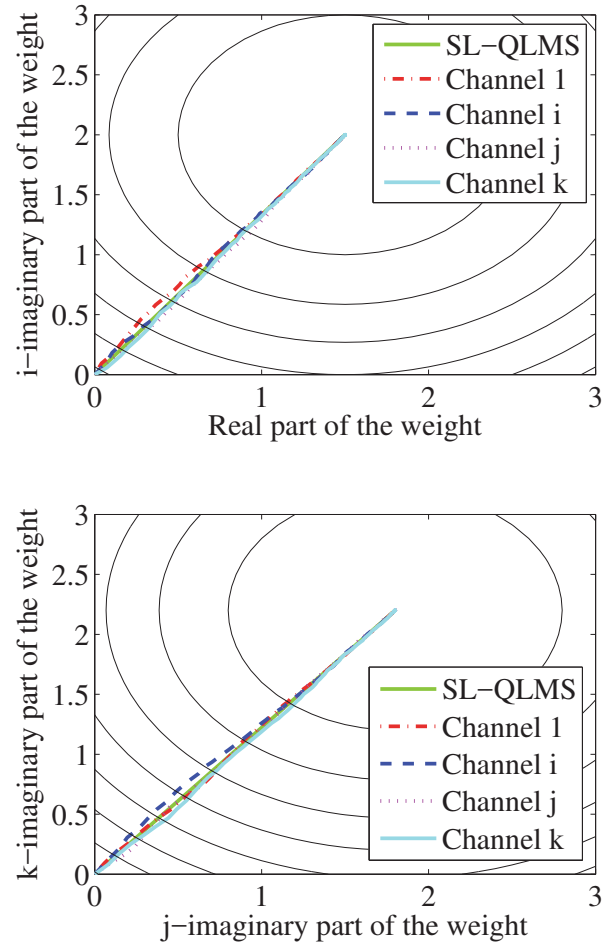


Figure 7.2: Averaged weight trajectories for the estimation of a strictly linear MA(1) process driven by $\mathbb{H}$ -proper white Gaussian noise.

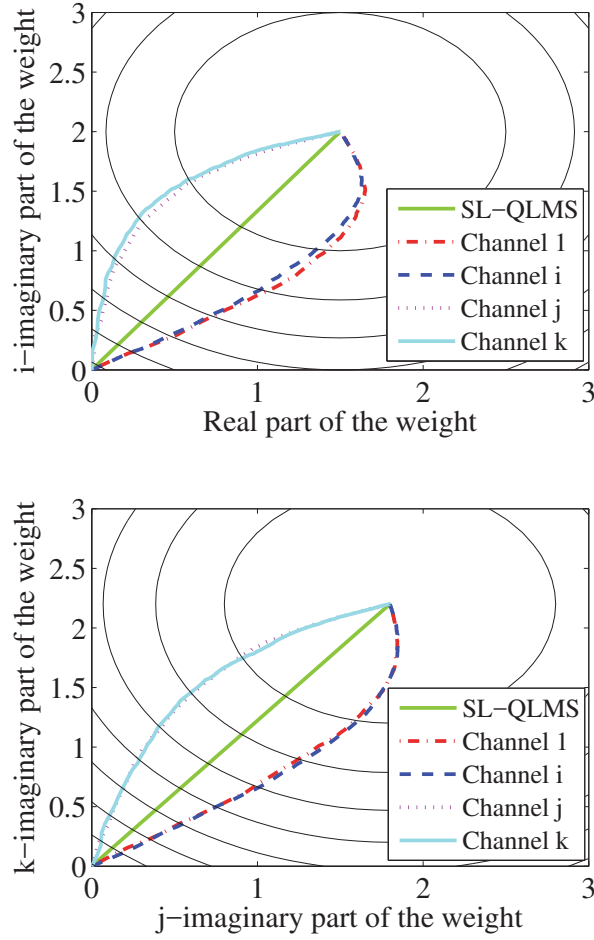


Figure 7.3: Averaged weight trajectories for the estimation of a strictly linear MA(1) process driven by $\mathbb{C}^2$ -improper white Gaussian noise.

quaternion vectors. A $\mathbb{C}^2$ -improper $\mathbf{x}$ can be expressed as $\mathbf{x} = \mathbf{z}_1 + \mathbf{z}_2 j$ where $\mathbf{z}_1$ and $\mathbf{z}_2$ are proper complex vectors defined in the plane spanned by $\{1, i\}$ [104], that is, the real and imaginary parts of each of $\mathbf{z}_1$ and $\mathbf{z}_2$ are uncorrelated and with the same variance, so that channel 1 and i or channel j and κ follow the same path, similarly to the $\mathbb{H}$ -proper case in (7.22). On the other hand, $\mathbf{z}_1$ and $\mathbf{z}_2$ are related or with different powers, so the two pairs of paths are in opposite directions.

General quaternion-valued signal with uncorrelated components

If the four components of $\mathbf{x}$ are uncorrelated, the covariance and complementary covariance matrices of $\mathbf{x}$ can be expressed as

$$\begin{aligned}\mathbf{C}_{\mathbf{xx}} &= \mathbf{C}_1 + \mathbf{C}_i + \mathbf{C}_j + \mathbf{C}_\kappa \\ \mathbf{C}_{\mathbf{xx}^i} &= \mathbf{C}_1 + \mathbf{C}_i - \mathbf{C}_j - \mathbf{C}_\kappa \\ \mathbf{C}_{\mathbf{xx}^j} &= \mathbf{C}_1 - \mathbf{C}_i + \mathbf{C}_j - \mathbf{C}_\kappa \\ \mathbf{C}_{\mathbf{xx}^\kappa} &= \mathbf{C}_1 - \mathbf{C}_i - \mathbf{C}_j + \mathbf{C}_\kappa\end{aligned}$$

where $\mathbf{C}_1 = E \left\{ \Re[\mathbf{x}] \Re[\mathbf{x}]^T \right\}$, $\mathbf{C}_i = E \left\{ \Im_i[\mathbf{x}] \Im_i[\mathbf{x}]^T \right\}$, $\mathbf{C}_j = E \left\{ \Im_j[\mathbf{x}] \Im_j[\mathbf{x}]^T \right\}$, and $\mathbf{C}_\kappa = E \left\{ \Im_\kappa[\mathbf{x}] \Im_\kappa[\mathbf{x}]^T \right\}$. The corresponding weight error vectors are given by

$$\begin{aligned}E \{ \tilde{\mathbf{w}}_{1,n+1} \} &= (\mathbf{I} - \mu_{\text{fc}} \mathbf{C}_1) E \{ \Re[\tilde{\mathbf{w}}_{1,n}] \} + \imath (\mathbf{I} - \mu_{\text{fc}} \mathbf{C}_i) E \{ \Im_i[\tilde{\mathbf{w}}_{1,n}] \} \\ &\quad + j (\mathbf{I} - \mu_{\text{fc}} \mathbf{C}_j) E \{ \Im_j[\tilde{\mathbf{w}}_{1,n}] \} + \kappa (\mathbf{I} - \mu_{\text{fc}} \mathbf{C}_\kappa) E \{ \Im_\kappa[\tilde{\mathbf{w}}_{1,n}] \} \\ E \{ \tilde{\mathbf{w}}_{i,n+1} \} &= (\mathbf{I} - \mu_{\text{fc}} \mathbf{C}_i) E \{ \Re[\tilde{\mathbf{w}}_{i,n}] \} + \imath (\mathbf{I} - \mu_{\text{fc}} \mathbf{C}_1) E \{ \Im_i[\tilde{\mathbf{w}}_{i,n}] \} \\ &\quad + j (\mathbf{I} - \mu_{\text{fc}} \mathbf{C}_\kappa) E \{ \Im_j[\tilde{\mathbf{w}}_{i,n}] \} + \kappa (\mathbf{I} - \mu_{\text{fc}} \mathbf{C}_j) E \{ \Im_\kappa[\tilde{\mathbf{w}}_{i,n}] \} \\ E \{ \tilde{\mathbf{w}}_{j,n+1} \} &= (\mathbf{I} - \mu_{\text{fc}} \mathbf{C}_j) E \{ \Re[\tilde{\mathbf{w}}_{j,n}] \} + \imath (\mathbf{I} - \mu_{\text{fc}} \mathbf{C}_\kappa) E \{ \Im_i[\tilde{\mathbf{w}}_{j,n}] \} \\ &\quad + j (\mathbf{I} - \mu_{\text{fc}} \mathbf{C}_1) E \{ \Im_j[\tilde{\mathbf{w}}_{j,n}] \} + \kappa (\mathbf{I} - \mu_{\text{fc}} \mathbf{C}_i) E \{ \Im_\kappa[\tilde{\mathbf{w}}_{j,n}] \} \\ E \{ \tilde{\mathbf{w}}_{\kappa,n+1} \} &= (\mathbf{I} - \mu_{\text{fc}} \mathbf{C}_\kappa) E \{ \Re[\tilde{\mathbf{w}}_{\kappa,n}] \} + \imath (\mathbf{I} - \mu_{\text{fc}} \mathbf{C}_j) E \{ \Im_i[\tilde{\mathbf{w}}_{\kappa,n}] \} \\ &\quad + j (\mathbf{I} - \mu_{\text{fc}} \mathbf{C}_i) E \{ \Im_j[\tilde{\mathbf{w}}_{\kappa,n}] \} + \kappa (\mathbf{I} - \mu_{\text{fc}} \mathbf{C}_1) E \{ \Im_\kappa[\tilde{\mathbf{w}}_{\kappa,n}] \}\end{aligned}\tag{7.24}$$

and converge if all the eigenvalues of the matrices $(\mathbf{I} - \mu_{\text{fc}} \mathbf{C}_1)$, $(\mathbf{I} - \mu_{\text{fc}} \mathbf{C}_i)$, $(\mathbf{I} - \mu_{\text{fc}} \mathbf{C}_j)$, and $(\mathbf{I} - \mu_{\text{fc}} \mathbf{C}_\kappa)$ are within $(-1, 1)$, that is,

$$0 < \mu_{\text{fc}} < \frac{2}{\sigma_{\max}}\tag{7.25}$$

where $\sigma_{\max}$ is the maximum eigenvalue of $\mathbf{C}_1$, $\mathbf{C}_i$, $\mathbf{C}_j$ and $\mathbf{C}_\kappa$. From (7.24), notice that when different step sizes are used in the four channels, their ranges are all identical to that shown in (7.25).

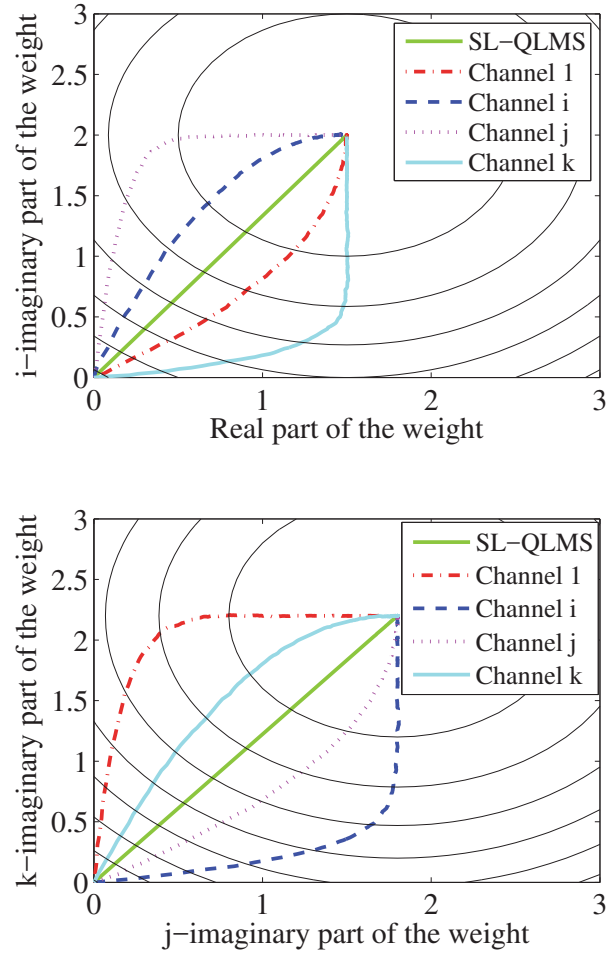


Figure 7.4: Averaged weight trajectories for the estimation of a strictly linear MA(1) process driven by general improper white Gaussian noise.

General improper signal

Equation (7.21) indicates that for a general improper quaternion signal, $\mathbf{x}$, the four weight error vectors obey different recursions which are symmetric about the weight error path of SL-QLMS, as shown in Figure 7.4.

7.4.3 Computational Complexity

Compared with the existing widely linear and strictly linear quaternion estimation approaches, the proposed four-channel estimation technique requires less computation cost. This is because the proposed estimation algorithm avoids the quaternion-quaternion addition in (7.4) and (7.5) while the proposed weight update rule in (7.9) replaces quaternion-quaternion multiplications, which cost 8 real multiplications and 28 real additions per multiplication [105], with real-quaternion multiplications costing only 4 real multiplications per multiplication. Table 7.1 and Figure 7.5 compare the FC-QLMS and SC-QLMS algorithms with the existing QLMS algorithms in terms of the number of real-valued operations required per iteration. Observe that the FC-QLMS and SC-QLMS require only about half the number of multiplications and an eighth of the number of additions, as compared to the WL-QLMS and the SL-QLMS. Therefore, the four-channel estimation framework allows for an efficient formulation of WL-QLMS and SL-QLMS in the form of FC-QLMS and SC-QLMS. Although the RC-WL-QLMS [99] and the MLMS [100] also have such low computational complexity, they bring about a loss of physical meaning inherent in the quaternion domain.

Table 7.1: The number of real-valued operations per iteration for the QLMS algorithms, for an adaptive filter of length L .

Algorithm	Real multiplications	Real additions
Estimation of widely linear systems		
<i>Proposed:</i> FC-QLMS	$32L + 4$	$32L$
WL-QLMS [24]	$64L + 4$	$256L$
RC-WL-QLMS [99]	$32L + 4$	$32L$
MLMS [100]	$32L + 4$	$32L$
Estimation of strictly linear systems		
<i>Proposed:</i> SC-QLMS	$8L + 1$	$8L$
SL-QLMS [33]	$16L + 4$	$64L$

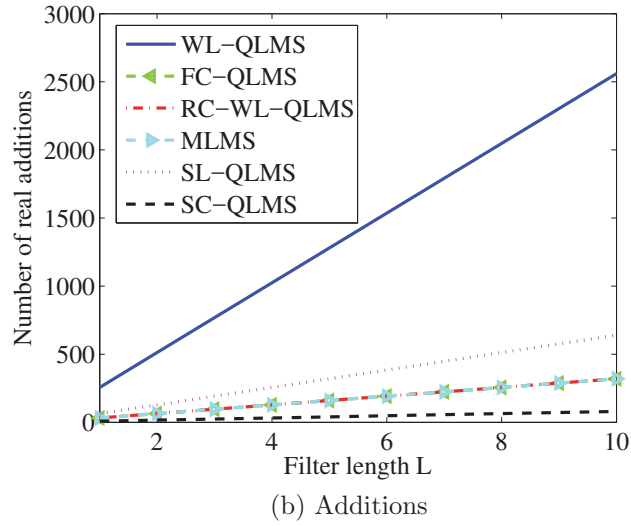
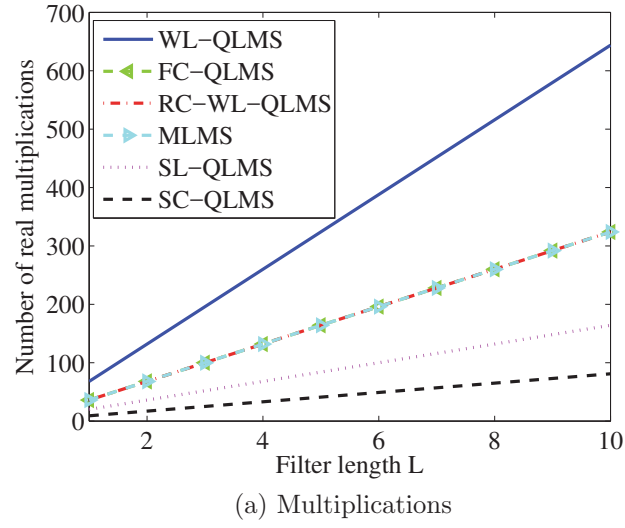


Figure 7.5: The number of real-valued operations per iteration for the QLMS algorithms, for an adaptive filter of length L .

7.4.4 Adaptive Step Size

Standard adaptive algorithms might experience degraded convergence when processing non-stationary signals with large dynamical ranges. This degradation can be circumvented by algorithms that adaptively optimise the step sizes. For the FC-QLMS algorithm, the four step sizes in the four channels can be optimised separately as $\mu_{1,n}$, $\mu_{\iota,n}$, $\mu_{j,n}$, and $\mu_{\kappa,n}$, before the weight vectors are updated as

$$\hat{\mathbf{w}}_{\alpha,n+1} = \hat{\mathbf{w}}_{\alpha,n} - \mu_{\alpha,n} \nabla_{\hat{\mathbf{w}}_{\alpha,n}^*} J_n^{\text{LMS}}, \quad \alpha \in \{1, \iota, j, \kappa\}$$

The Barzilai-Borwein method [106] can be used to obtain the optimal step sizes at the n -th iteration, as

$$\mu_{\alpha,n} = \arg \min_{\mu \in \mathbb{R}^+} \|\hat{\mathbf{w}}_{\alpha,n} - \hat{\mathbf{w}}_{\alpha,n-1} - \mu \mathbf{\Gamma}_{\alpha,n}\|^2, \quad \alpha \in \{1, \iota, j, \kappa\}$$

where

$$\mathbf{\Gamma}_{\alpha,n} = \frac{\partial \Psi[y_n - \hat{\mathbf{w}}_{\alpha,n}^H \mathbf{x}_n]^2}{\partial \hat{\mathbf{w}}_{\alpha,n}^*} - \frac{\partial \Psi[y_n - \hat{\mathbf{w}}_{\alpha,n-1}^H \mathbf{x}_n]^2}{\partial \hat{\mathbf{w}}_{\alpha,n-1}^*} = \Psi \left[(\hat{\mathbf{w}}_{\alpha,n} - \hat{\mathbf{w}}_{\alpha,n-1})^H \mathbf{x}_n \right] \mathbf{x}_n \alpha^{-1}$$

By setting

$$\nabla_{\mu_{\alpha,n}} \|\hat{\mathbf{w}}_{\alpha,n} - \hat{\mathbf{w}}_{\alpha,n-1} - \mu_{\alpha,n} \mathbf{\Gamma}_{\alpha,n}\|^2 = 0$$

the optimal step size can be obtained as

$$\mu_{\alpha,n} = \frac{\Re \left[(\hat{\mathbf{w}}_{\alpha,n} - \hat{\mathbf{w}}_{\alpha,n-1})^H \mathbf{\Gamma}_{\alpha,n} \right]}{\mathbf{\Gamma}_{\alpha,n}^H \mathbf{\Gamma}_{\alpha,n}} = \|\mathbf{x}_n\|^{-2} \quad (7.26)$$

which is equivalent to a unit step size after data normalisation [3].

7.5 Simulations

All estimation algorithms discussed were evaluated by simulations over synthetic and real-world signals. The performance index in the form of normalised MSE

(NMSE) was calculated at each iteration through averaging the error power normalised by the signal power from 100 independent trials, to yield

$$\text{NMSE}_n = \frac{1}{100} \sum_{l=1}^{100} \frac{|y_n^{(l)} - \hat{y}_n^{(l)}|^2}{|y_n^{(l)}|^2}$$

where $y_n^{(l)}$ is the desired signal and $\hat{y}_n^{(l)}$ the estimate at the l -th trial. For a fair comparison, according to the theoretical analysis result shown in Remark 7.1, the values of step sizes of the four-channel and single-channel algorithms were set to four times those of their widely linear and strictly linear counterparts, and the weight vectors in the algorithms were initialised to zero vectors.

7.5.1 QLMS

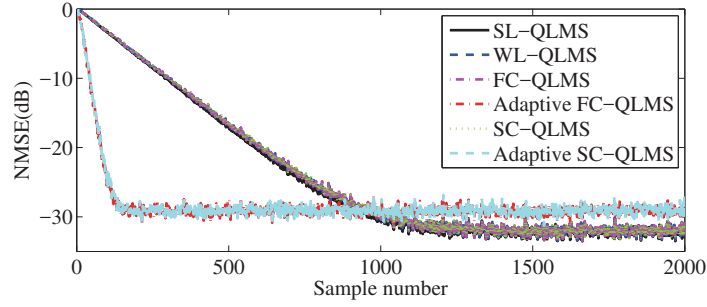
The QLMS algorithms were implemented to identify a strictly linear MA system, given by

$$y_n = b_0 x_n + b_1 x_{n-1} + b_2 x_{n-2} + b_3 x_{n-3} + v_n \quad (7.27)$$

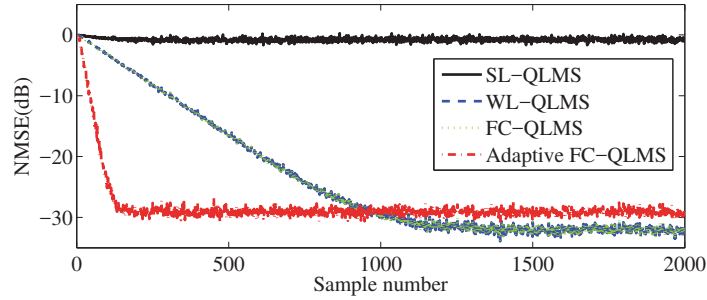
and the widely linear MA system

$$y_n = b_0 x_n + b_1 x_{n-1}^i + b_2 x_{n-2}^j + b_3 x_{n-3}^\kappa + v_n \quad (7.28)$$

where $b_0, b_1, b_2, b_3 \in \mathbb{H}$, the signal x and noise v were $\mathbb{H}$ -proper quaternion-valued white Gaussian data, and the signal-to-noise ratio was 10 dB. Figure 7.6a illustrates that when estimating the strictly linear MA system in (7.27), FC-QLMS, WL-QLMS, SC-QLMS, and SL-QLMS behaved identically, while the FC-QLMS and SC-QLMS with an adaptive step size (7.26), which are referred to as the adaptive FC-QLMS and adaptive SC-QLMS, exhibited faster convergence. Figure 7.6b shows that when estimating the widely linear MA system in (7.28), the FC-QLMS and the WL-QLMS behaved identically, and had a much smaller steady-state NMSE than the SL-QLMS, while the adaptive FC-QLMS converged



(a) Strictly linear MA system



(b) Widely linear MA system

Figure 7.6: NMSE curves of the QLMS algorithms for the estimation of MA systems.

fastest. The 2000 iterations of WL-QLMS took 11.6 s, while the 2000 iterations of FC-QLMS took only 6.1 s. The 2000 iterations of SL-QLMS took 2.4 s, whereas the 2000 iterations of SC-QLMS took only 1.9 s.

The considered QLMS algorithms were also used to forecast quaternion-valued wind data, the four components of which contain the wind speed in the north-south, east-west, and vertical directions, and the temperature. As shown in Figure 7.7, the FC-QLMS and the WL-QLMS behaved identically, while the adaptive FC-QLMS had the fastest convergence.

7.5.2 QRLS

Figure 7.8 and 7.9 illustrate the performance of QRLS algorithms in the identification of the MA systems in (7.27) and (7.28), and in the prediction of quaternion-

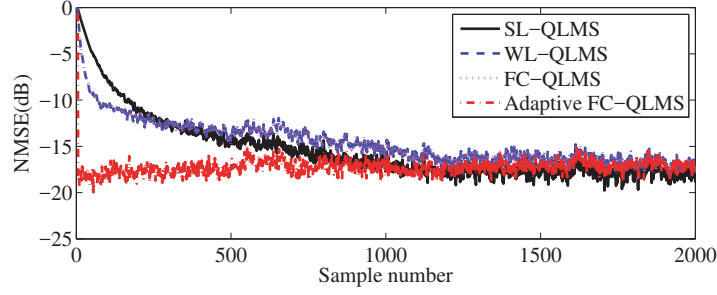


Figure 7.7: NMSE curves of the QLMS algorithms for the wind forecasting.

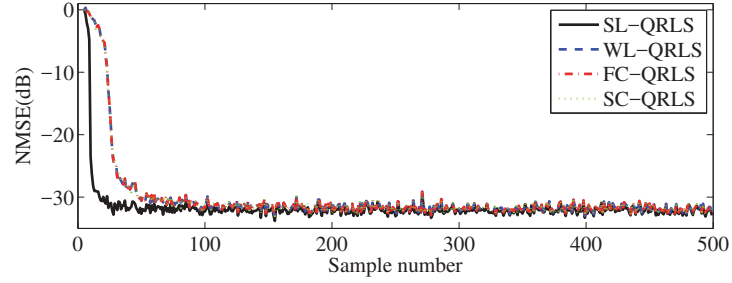
valued wind data. The FC-QRLS behaved identically to the WL-QRLS for both synthetic and wind data, while when identifying the strictly linear MA system, the SC-QRLS achieved equivalent performance. The 500 iterations of WL-QRLS took 3.3 s, while the 500 iterations of FC-QRLS took 3.0 s. The 500 iterations of SL-QRLS took 1.2 s, whereas the 500 iterations of SC-QRLS took 1.0 s. The advantage of FC-QRLS and SC-QRLS is not so significant as that of the FC-QLMS and SC-QLMS since the computational complexity of the intermediate variables in (7.11) and (7.12), $\mathbf{P}^a$, $\mathbf{p}^a$, $\mathbf{P}$, and $\mathbf{p}$, cannot be reduced by the proposed approach.

7.5.3 QNGD

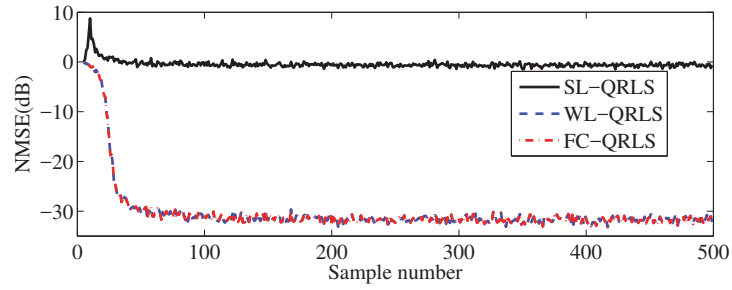
The QNGD algorithms were employed in a one-step ahead prediction of the 4D Saito's chaotic circuit data, which is governed by four state variables x_1, y_1, x_2, y_2 and five parameters, as [107]

$$\begin{bmatrix} \frac{\partial x_1}{\partial \tau} \\ \frac{\partial y_1}{\partial \tau} \end{bmatrix} = \begin{bmatrix} -1 & 1 \\ -\alpha_1 & -\alpha_1\beta_1 \end{bmatrix} \begin{bmatrix} x_1 - \gamma \frac{\beta_1}{1-\beta_1} h(z) \\ y_1 - \gamma \frac{1}{1-\beta_1} h(z) \end{bmatrix}$$

$$\begin{bmatrix} \frac{\partial x_2}{\partial \tau} \\ \frac{\partial y_2}{\partial \tau} \end{bmatrix} = \begin{bmatrix} -1 & 1 \\ -\alpha_2 & -\alpha_2\beta_2 \end{bmatrix} \begin{bmatrix} x_2 - \gamma \frac{\beta_2}{1-\beta_2} h(z) \\ y_2 - \gamma \frac{1}{1-\beta_2} h(z) \end{bmatrix}$$



(a) Strictly linear MA system



(b) Widely linear MA system

Figure 7.8: NMSE curves of the QRLS algorithms for the estimation of MA systems.

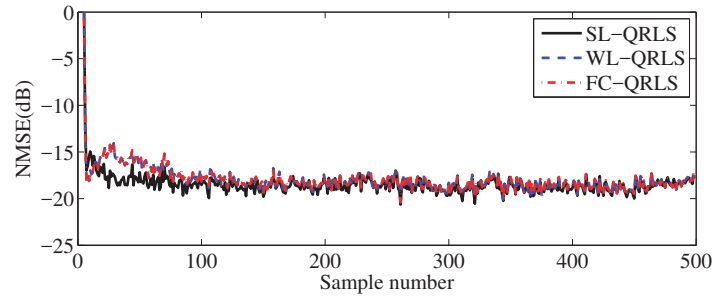
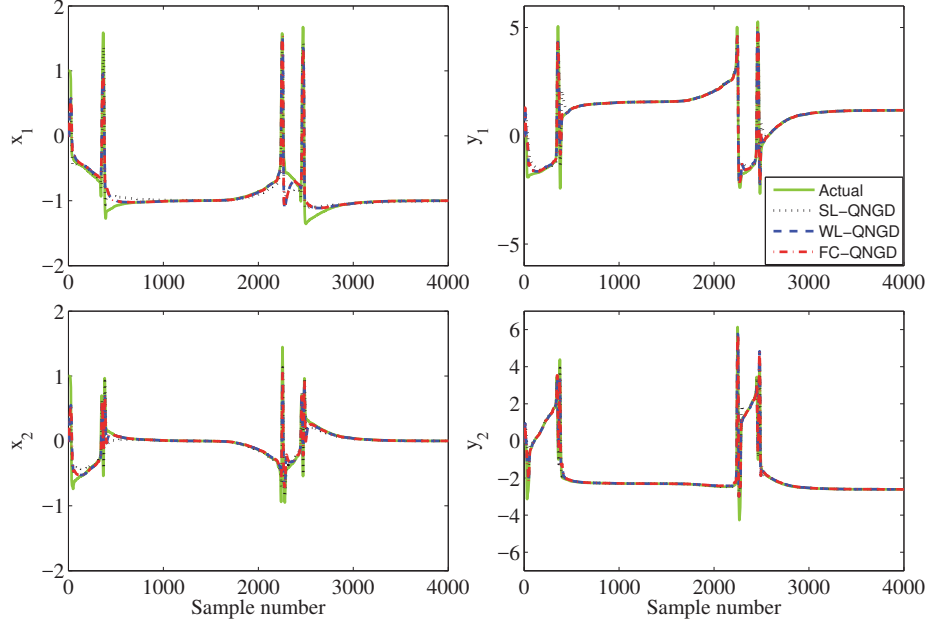
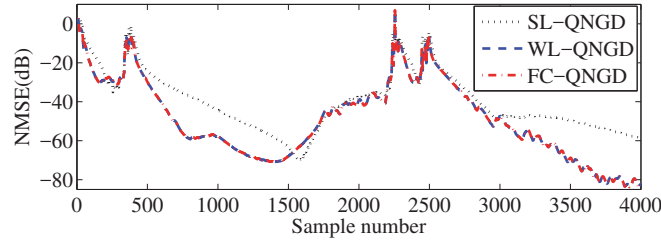


Figure 7.9: NMSE curves of the QRLS algorithms for the wind forecasting.



(a) Estimate curves



(b) NMSE curves

Figure 7.10: Performance of the QNGD algorithms for the one-step ahead prediction of Saito's chaotic circuit data.

$$h(z) = \begin{cases} 1, & z \geq -1 \\ -1, & z \leq 1 \end{cases}$$

$$z = x_1 + x_2$$

where τ is the time constant of the chaotic circuit. The parameter values were chosen as $\gamma = 1.3, \alpha_1 = 7.5, \alpha_2 = 15, \beta_1 = 0.16, \beta_2 = 0.097$. Figure 7.10 shows that the FC-QNGD and WL-QNGD behaved identically and outperformed the SL-QNGD. The 4000 iterations of WL-QNGD and SL-QNGD took 50 s and 31 s, respectively, while the 4000 iterations of FC-QNGD took only 29 s.

7.6 Summary

This chapter has introduced a four-channel linear model for quaternion signal estimation and has proposed the corresponding adaptive MMSE estimation algorithms, referred to as the FC-QLMS, FC-QRLS and FC-QNGD. These have been shown to be cost-effective alternatives to existing widely linear algorithms, while exhibiting identical performances. The proposed individual estimation of real and imaginary quaternion components has maintained the four degrees of freedom necessary for widely linear estimation, together with providing physical insight into quaternion estimation and enhanced flexibility through the estimation in each channel being controlled independently. In this way, the computational complexity has been dramatically reduced, while maintaining the performance of the original algorithms. Simulation studies on synthetic and real-world signals have validated the analysis.

Chapter 8

Widely Linear Quaternion Multiple Model Adaptive Estimation

Overview

Quaternion state estimation techniques have been used in various applications, yet they are only suitable for dynamical systems represented by a single known model. In order to deal with model uncertainty, this chapter proposes a class of widely linear quaternion multiple model adaptive estimation (QMMAE) algorithms based on widely linear quaternion Kalman filters (WL-QKFs) and Bayesian inference. The augmented second-order quaternion statistics is employed to capture complete second-order statistical information in improper quaternion signals. Within the QMMAE framework, a widely linear quaternion interacting multiple model algorithm is proposed to track time-variant model uncertainty, while a widely linear quaternion static multiple model algorithm is proposed for time-invariant model uncertainty. For improper quaternion signals, the proposed widely linear QMMAE algorithms exhibit enhanced performance over their strictly linear counterparts.

8.1 Introduction

Quaternion-valued state estimation methods have been proposed but are only suitable for dynamical systems represented by a single known model [39]. In reality, however, there is often considerable uncertainty about the system model. For

example, in target tracking, the target motion may follow the constant velocity model, the constant acceleration model, or other kinematic models [108]. For real-valued state estimation, an effective method for circumventing the system uncertainty is through multiple model adaptive estimation (MMAE), which can be seen as an application of the mixture of experts methodology [109, 110] to the state estimation problem. A real-valued MMAE algorithm typically consists of a bank of filters based on multiple models that represent possible system behaviour patterns, referred to as modes, and a fusion algorithm that fuses the state estimates of different filters to form the overall estimate. The fusion algorithm may be based on Bayesian inference [111] or neural network training [112]. Existing MMAE algorithms are divided into two categories: static multiple model (SMM) algorithms and interacting multiple model (IMM) algorithms. The former are valid for time-invariant system uncertainty and the latter are effective for time-variant system uncertainty. Both have been extensively used in practical applications such as target tracking, fault diagnosis, and intelligent control [108, 113–115], and have been recently extended to the complex domain [116]. The theoretical performance analysis of the MMAE algorithms, especially the IMM, however, is challenging. Recent research on this issue has been focused on recursive approaches, including the hybrid conditional averaging technique for predicting the mean square error [117–119], and the posterior Cramer-Rao lower bound technique for computing the lower performance bound [120]. These approaches have been shown to be computationally efficient alternatives to the traditionally used Monte Carlo simulation method.

However, quaternion MMAE (QMMAE) algorithms for hybrid systems, which have practical engineering applications, have not yet been investigated in the literature. The QMMAE is a non-trivial extension of the real-valued MMAE since the augmented second-order quaternion statistics must be incorporated to deal with improper quaternion signals. This chapter proposes a class of widely linear QMMAE (WL-QMMAE) algorithms based on the widely linear quaternion Kalman

filter (WL-QKF) introduced in Algorithm 2.1. This approach accounts for the complete second-order statistics of improper quaternion signals. For time-varying system uncertainty, a widely linear quaternion IMM (WL-QIMM) algorithm and a recursive algorithm for the analysis of its performance are proposed. For time-invariant system uncertainty, the WL-QIMM can simplify to a widely linear quaternion SMM (WL-QSMM) algorithm. In a similar spirit, for semi-widely linear systems with $\mathbb{C}^n$ -improper noise, the WL-QMMAE reduces to the semi-widely linear QMMAE (SWL-QMMAE), while for strictly linear systems with $\mathbb{H}$ -proper noise, the WL-QMMAE reduces to the strictly linear QMMAE (SL-QMMAE). For rigour, the convergence of WL-QSMM has also been proven.

8.2 QMMAE Algorithms

If a quaternion dynamical system is represented by the widely linear model in (2.22), but the state transition matrix, observation matrix, and properties of the noises are uncertain over time, a quaternion Kalman filter based on a single known model will suffer a great performance loss. In such cases, N_m potential models may be employed as candidates for the unknown system model, and each of them is represented by

$$\begin{aligned}\mathbf{x}_n^a &= \mathbf{F}_{n-1}^{a[l]} \mathbf{x}_{n-1}^a + \mathbf{u}_{n-1}^{a[l]} \\ \mathbf{z}_n^a &= \mathbf{H}_n^{a[l]} \mathbf{x}_n^a + \mathbf{v}_n^{a[l]}\end{aligned}\tag{8.1}$$

where $l \in \{1, 2, \dots, N_m\}$ is the model index, $\mathbf{u}_{n-1}^{a[l]}$ and $\mathbf{v}_n^{a[l]}$ are zero-mean improper state and observation noise, and $\mathbf{F}_{n-1}^{a[l]}$ and $\mathbf{H}_{n-1}^{a[l]}$ are state transition and observation matrices corresponding to model l . Refer to the system behaviour pattern as a mode. For example, if the system can be represented by model l at time n , the system is called in mode l at time n . By applying N_m WL-QKFs to the observations, N_m augmented state estimates and error covariances are given by

$$\hat{\mathbf{x}}_{n|n}^{a[l]} = E \{ \mathbf{x}_n^a \mid \mathbf{Z}_n^a, \tau_n^{[l]} \} \tag{8.2}$$

$$\mathbf{P}_{n|n}^{a[l]} = E \left\{ \left(\mathbf{x}_n^a - \hat{\mathbf{x}}_{n|n}^{a[l]} \right) \left(\mathbf{x}_n^a - \hat{\mathbf{x}}_{n|n}^{a[l]} \right)^H \mid \mathbf{Z}_n^a, \tau_n^{[l]} \right\} \quad (8.3)$$

By fusing the N_m estimates, the overall augmented estimate can be obtained as

$$E \{ \mathbf{x}_n^a \mid \mathbf{Z}_n^a \} = \sum_{l=1}^{N_m} p \left(\tau_n^{[l]} \mid \mathbf{Z}_n^a \right) \hat{\mathbf{x}}_{n|n}^{a[l]} \quad (8.4)$$

while the overall augmented error covariance can be derived as

$$\begin{aligned} & E \left\{ \left(\mathbf{x}_n^a - E \{ \mathbf{x}_n^a \mid \mathbf{Z}_n^a \} \right) \left(\mathbf{x}_n^a - E \{ \mathbf{x}_n^a \mid \mathbf{Z}_n^a \} \right)^H \mid \mathbf{Z}_n^a, \tau_n^{[l]} \right\} \\ &= \sum_{l=1}^{N_m} p \left(\tau_n^{[l]} \mid \mathbf{Z}_n^a \right) \left[\mathbf{P}_{n|n}^{a[l]} + \left(\hat{\mathbf{x}}_{n|n}^{a[l]} - \hat{\mathbf{x}}_{n|n}^a \right) \left(\hat{\mathbf{x}}_{n|n}^{a[l]} - \hat{\mathbf{x}}_{n|n}^a \right)^H \right] \\ &= \sum_{l=1}^{N_m} p \left(\tau_n^{[l]} \mid \mathbf{Z}_n^a \right) \left[\mathbf{P}_{n|n}^{a[l]} + \hat{\mathbf{x}}_{n|n}^{a[l]} \left(\hat{\mathbf{x}}_{n|n}^{a[l]} \right)^H \right] - \hat{\mathbf{x}}_{n|n}^a \left(\hat{\mathbf{x}}_{n|n}^a \right)^H \end{aligned} \quad (8.5)$$

where $\mathbf{Z}_n^a \triangleq [\mathbf{z}_1^{aT}, \dots, \mathbf{z}_n^{aT}]^T = [\mathbf{Z}_{n-1}^{aT}, \mathbf{z}_n^{aT}]^T$ is the set of measurements, and $\tau_n^{[l]}$ is the event that the system is in mode l at time instant n . For simplicity, assume that the system model is always included in the N_m candidate models. Then, the conditional probability in (8.4) and (8.5), $p \left(\tau_n^{[l]} \mid \mathbf{Z}_n^a \right)$, can be derived using Bayes' law, as

$$\begin{aligned} p \left(\tau_n^{[l]} \mid \mathbf{Z}_n^a \right) &= \frac{p \left(\mathbf{z}_n^a \mid \mathbf{Z}_{n-1}^a, \tau_n^{[l]} \right) p \left(\tau_n^{[l]} \mid \mathbf{Z}_{n-1}^a \right)}{\sum_{i=1}^{N_m} p \left(\mathbf{z}_n^a \mid \mathbf{Z}_{n-1}^a, \tau_n^{[i]} \right) p \left(\tau_n^{[i]} \mid \mathbf{Z}_{n-1}^a \right)} \\ &= \frac{p \left(\mathbf{z}_n^a \mid \mathbf{Z}_{n-1}^a, \tau_n^{[l]} \right) \sum_{i=1}^{N_m} p \left(\tau_n^{[i]} \mid \mathbf{Z}_{n-1}^a \right) p \left(\tau_n^{[l]} \mid \tau_{n-1}^{[i]} \right)}{\sum_{i=1}^{N_m} p \left(\mathbf{z}_n^a \mid \mathbf{Z}_{n-1}^a, \tau_n^{[i]} \right) \sum_{j=1}^{N_m} p \left(\tau_{n-1}^{[j]} \mid \mathbf{Z}_{n-1}^a \right) p \left(\tau_n^{[i]} \mid \tau_{n-1}^{[j]} \right)} \end{aligned} \quad (8.6)$$

where $p \left(\mathbf{z}_n^a \mid \mathbf{Z}_{n-1}^a, \tau_n^{[l]} \right)$ is the likelihood function of $\tau_n^{[l]}$. Note that $p \left(\mathbf{z}_n^a \mid \mathbf{Z}_{n-1}^a, \tau_n^{[l]} \right) = p \left(\mathbf{r}_n^{a[l]} \mid \tau_n^{[l]} \right)$, where

$$\mathbf{r}_n^{a[l]} \triangleq \mathbf{z}_n^a - \mathbf{H}_n^{[l]} \hat{\mathbf{x}}_{n|n-1}^{a[l]} \quad (8.7)$$

is referred to as the augmented residual vector. Following the principle commonly used for real-valued and complex-valued MMAE, the augmented residual vector can be assumed to be Gaussian distributed, with $\mathbf{r}_n^{a[l]} \sim \mathcal{N} \left(\mathbf{0}, \mathbf{S}_n^{a[l]} \right)$, where $\mathbf{S}_n^{a[l]}$ is the augmented covariance matrix of $\mathbf{r}_n^{a[l]}$. Thus, the estimate of the likelihood

function, denoted by $\mathcal{L}_n^{[l]}$, can be computed as

$$\mathcal{L}_n^{[l]} = f_G \left(\mathbf{r}_n^{a[l]} \mid \mathbf{0}, \mathbf{S}_n^{a[l]} \right) \quad (8.8)$$

Note that $\mathbf{S}_n^{a[l]}$ is an intermediate result in the WL-QKF algorithm corresponding to model l , shown in Algorithm 2.1. The estimated mode at time n is therefore given by $\arg \max_{l \in \{1, 2, \dots, N_m\}} p \left(\tau_n^{[l]} \mid \mathbf{Z}_n^a \right)$. The above constitutes the main framework of the proposed widely linear QMMAE (WL-QMMAE).

8.2.1 WL-QIMM Algorithm

To understand the WL-QMMAE, the probability $p \left(\tau_n^{[l]} \mid \mathbf{Z}_{n-1}^a \right)$ in (8.6) needs to be explained further. For systems with a time-varying mode, one can make the standard assumption that the mode transition is governed by a first-order Markov chain, that is

$$p \left(\tau_n^{[l]} \mid \tau_{n-1}^{[i]} \right) = \pi^{[il]}, \quad \forall i, l \in \{1, 2, \dots, N_m\} \quad (8.9)$$

where $\pi^{[il]}$ is the Markov transition probability from mode i to mode l . Upon incorporating this assumption and imposing a reinitialization step at the beginning of the estimation to couple the N_m WL-QKFs, the WL-QMMAE framework results in a practical WL-QIMM algorithm, an extension of the real-valued IMM [121]. An iteration of this WL-QIMM algorithm at time n can be described as follows.

Step 1 (Reinitialisation). For $l = 1, \dots, N_m$, $i = 1, \dots, N_m$, the probability of the system being in mode i at time $(n-1)$ given that the system is in mode l at time n , is computed as

$$\gamma_n^{[li]} \triangleq p \left(\tau_{n-1}^{[i]} \mid \tau_n^{[l]}, \mathbf{Z}_{n-1}^a \right) = \frac{\pi^{[il]} \mu_{n-1}^{[i]}}{\sum_{j=1}^{N_m} \pi^{[jl]} \mu_{n-1}^{[j]}} \quad (8.10)$$

where $\mu_{n-1}^{[i]}$, termed *mode probability*, is the estimate of $p \left(\tau_{n-1}^{[i]} \mid \mathbf{Z}_{n-1}^a \right)$ in the previous iteration. For $l = 1, \dots, N_m$, compute the augmented initial state of filter

l based on Bayes' law as

$$\mathbf{x}_{0n}^{a[l]} = E \left\{ \mathbf{x}_{n-1}^a \mid \tau_n^{[l]}, \mathbf{Z}_{n-1}^a \right\} = \sum_{i=1}^{N_m} \hat{\mathbf{x}}_{n-1|n-1}^{a[i]} \gamma_n^{[li]} \quad (8.11)$$

and the augmented initial error covariance of filter l as

$$\begin{aligned} \mathbf{P}_{0n}^{a[l]} &= \sum_{i=1}^{N_m} \gamma_n^{[li]} \left[\mathbf{P}_{n-1|n-1}^{a[i]} + \left(\mathbf{x}_{0n}^{[l]} - \hat{\mathbf{x}}_{n-1|n-1}^{a[i]} \right) \left(\mathbf{x}_{0n}^{[l]} - \hat{\mathbf{x}}_{n-1|n-1}^{a[i]} \right)^H \right] \\ &= \sum_{i=1}^{N_m} \gamma_n^{[li]} \left[\mathbf{P}_{n-1|n-1}^{a[i]} + \hat{\mathbf{x}}_{n-1|n-1}^{a[i]} \left(\hat{\mathbf{x}}_{n-1|n-1}^{a[i]} \right)^H \right] - \mathbf{x}_{0n}^{a[l]} \left(\mathbf{x}_{0n}^{a[l]} \right)^H \end{aligned}$$

Step 2 (Mode-conditioned Kalman filtering). Implement N_m WL-QKF's with the initial conditions $\mathbf{x}_{0n}^{a[l]}$ and $\mathbf{P}_{0n}^{a[l]}$. Specifically, for $l = 1, \dots, N_m$, compute the predicted augmented state:

$$\hat{\mathbf{x}}_{n|n-1}^{a[l]} = \mathbf{F}_{n-1}^{a[l]} \mathbf{x}_{0n}^{a[l]} \quad (8.12)$$

predicted augmented error covariance:

$$\mathbf{P}_{n|n-1}^{a[l]} = \mathbf{F}_{n-1}^{a[l]} \mathbf{P}_{0n}^{a[l]} \left(\mathbf{F}_{n-1}^{a[l]} \right)^H + \mathbf{Q}_{n-1}^{a[l]} \quad (8.13)$$

augmented residual:

$$\mathbf{r}_n^{a[l]} = \mathbf{z}_n^a - \mathbf{H}_n^{[l]} \hat{\mathbf{x}}_{n|n-1}^{a[l]} \quad (8.14)$$

augmented residual covariance:

$$\mathbf{S}_n^{a[l]} = \mathbf{H}_n^{a[l]} \mathbf{P}_{n|n-1}^{a[l]} \left(\mathbf{H}_n^{a[l]} \right)^H + \mathbf{R}_n^{a[l]} \quad (8.15)$$

filter gain:

$$\mathbf{K}_n^{a[l]} = \mathbf{P}_{n|n-1}^{a[l]} \left(\mathbf{H}_n^{a[l]} \right)^H \left(\mathbf{S}_n^{a[l]} \right)^{-1} \quad (8.16)$$

updated augmented state:

$$\hat{\mathbf{x}}_{n|n}^{a[l]} = \hat{\mathbf{x}}_{n|n-1}^{a[l]} + \mathbf{K}_n^{a[l]} \mathbf{r}_n^{a[l]} \quad (8.17)$$

updated augmented error covariance:

$$\mathbf{P}_{n|n}^{a[l]} = \mathbf{P}_{n|n-1}^{a[l]} - \mathbf{K}_n^{a[l]} \mathbf{H}_n^{a[l]} \mathbf{P}_{n|n-1}^{a[l]} \quad (8.18)$$

Step 3 (Combination). For $l = 1, \dots, N_m$, compute the likelihood for mode l as in (8.8), and the estimate of $p(\tau_n^{[l]} | \mathbf{Z}_n^a)$, denoted by $\mu_n^{[l]}$, as

$$\mu_n^{[l]} = \frac{\mathcal{L}_n^{[l]} \sum_{i=1}^{N_m} \pi^{[il]} \mu_{n-1}^{[i]}}{\sum_{i=1}^{N_m} \mathcal{L}_n^{[i]} \sum_{j=1}^{N_m} \pi^{[ji]} \mu_{n-1}^{[j]}} \quad (8.19)$$

according to (8.6) and (8.9). Based on (8.4), the overall augmented state estimate is a weighted sum of the augmented state estimates from N_m WL-QKF's, and is given by

$$\hat{\mathbf{x}}_{n|n}^a = \sum_{l=1}^{N_m} \mu_n^{[l]} \hat{\mathbf{x}}_{n|n}^{a[l]} \quad (8.20)$$

Based on (8.5), the overall augmented error covariance is given by

$$\begin{aligned} \mathbf{P}_{n|n}^a &= \sum_{l=1}^{N_m} \mu_n^{[l]} \left[\mathbf{P}_{n|n}^{a[l]} + \left(\hat{\mathbf{x}}_{n|n}^{a[l]} - \hat{\mathbf{x}}_{n|n}^a \right) \left(\hat{\mathbf{x}}_{n|n}^{a[l]} - \hat{\mathbf{x}}_{n|n}^a \right)^H \right] \\ &= \sum_{l=1}^{N_m} \mu_n^{[l]} \left[\mathbf{P}_{n|n}^{a[l]} + \hat{\mathbf{x}}_{n|n}^{a[l]} \left(\hat{\mathbf{x}}_{n|n}^{a[l]} \right)^H \right] - \hat{\mathbf{x}}_{n|n}^a \left(\hat{\mathbf{x}}_{n|n}^a \right)^H \end{aligned} \quad (8.21)$$

8.2.2 Performance Analysis of WL-QIMM

The stability condition of the real-valued IMM algorithm proved in [122] is readily extended to the QIMM, that is, if the system is uniformly controllable and uniformly observable, the covariances of the mode-conditioned Kalman filters are uniformly bounded from above and below. Furthermore, to efficiently evaluate the

estimation performance of WL-QIMM, this subsection extends the hybrid conditional averaging algorithms for the performance analysis of real-valued IMM [117,118] to the quaternion domain. The resulting recursive algorithm, without recourse to multiple trials, is more computationally efficient than the commonly used Monte Carlo simulation method [118]. Note that owing to the widely linear nature of the WL-QIMM algorithm, the performance analysis algorithm must also be in a widely linear form, since the augmented second-order quaternion statistics must be incorporated.

The performance of WL-QIMM depends on the operating scenario, so that the performance analysis must correspond to a definite operating scenario $\mathcal{S}$ in which the system can be represented by a widely linear quaternion state model in the form of (2.22). Compared with (8.1), the model in (2.22) is general in the sense that $\mathbf{F}_{n-1}^a$, $\mathbf{H}_n^a$, $\mathbf{u}_{n-1}^a$ and $\mathbf{v}_n^a$ may or may not belong to the candidate model sets used by the N_m WL-QKFs. Define $\bar{\mathbf{x}} \triangleq E\{\mathbf{x} \mid \mathcal{S}\}$ to denote the mean of the quaternion random vector $\mathbf{x}$ in a scenario $\mathcal{S}$, while $\text{Cov}(\mathbf{x}, \mathbf{y}) \triangleq E\{(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{y} - \bar{\mathbf{y}})^H \mid \mathcal{S}\}$ denotes the cross-covariance of two quaternion random vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathcal{S}$. Based on these definitions, a recursive algorithm can be derived to estimate the MSE of WL-QIMM without generating observation data in multiple trials. This is achieved by the conditional expectation operation mentioned above, through which the randomness of the error covariance due to the noise is averaged out. This recursive algorithm is similar to the performance analysis algorithm for real-valued IMM [118], which originated from [117], but is different in two respects: (i) the augmented forms of the quaternion variables are employed; (ii) the QMGD with the generic probability density function in (2.15) is incorporated into the evaluation of the average likelihood functions and the average mode probabilities. An iteration of the proposed performance analysis algorithm at time n is described as follows, while the derivation is omitted for its generic extension from the real-valued algorithm, for which the derivation can be found in [117,118].

Step 1. Compute the means of the initial conditions of the mode-conditioned

WL-QKFs. For $l = 1, \dots, N_m$, $i = 1, \dots, N_m$, compute the mean of the mixing probability as

$$\bar{\gamma}_n^{[li]} \triangleq E \{ \gamma_n^{[li]} \mid \mathcal{S} \} \approx \frac{\pi^{[il]} \bar{\mu}_{n-1}^{[i]}}{\sum_{j=1}^{N_m} \pi^{[jl]} \bar{\mu}_{n-1}^{[j]}} \quad (8.22)$$

For $l = 1, \dots, N_m$, compute the mean of the initial condition of filter l as

$$\bar{\mathbf{x}}_{0n}^{[l]} = \sum_{i=1}^{N_m} \bar{\gamma}_n^{[li]} \bar{\mathbf{x}}_{n-1|n-1}^{[i]} \quad (8.23)$$

$$\begin{aligned} \bar{\mathbf{P}}_{0n}^{a[l]} &= \sum_{i=1}^{N_m} \bar{\gamma}_n^{[li]} \left[\bar{\mathbf{P}}_{n-1|n-1}^{a[i]} + \left(\bar{\mathbf{x}}_{0n}^{a[l]} - \bar{\mathbf{x}}_{n-1|n-1}^{a[i]} \right) \left(\bar{\mathbf{x}}_{0n}^{a[l]} - \bar{\mathbf{x}}_{n-1|n-1}^{a[i]} \right)^H \right] \\ &= \sum_{i=1}^{N_m} \bar{\gamma}_n^{[li]} \left[\bar{\mathbf{P}}_{n-1|n-1}^{a[i]} + \bar{\mathbf{x}}_{n-1|n-1}^{a[i]} \left(\bar{\mathbf{x}}_{n-1|n-1}^{a[i]} \right)^H \right] - \bar{\mathbf{x}}_{0n}^{a[l]} \left(\bar{\mathbf{x}}_{0n}^{a[l]} \right)^H \end{aligned}$$

Step 2. Compute the means of residuals and estimates of the mode-conditioned WL-QKFs. For $l = 1, \dots, N_m$, define the estimation error of filter l as $\mathbf{e}_n^{[l]} \triangleq \mathbf{x}_n - \hat{\mathbf{x}}_{n|n}^{[l]}$, and the error of the initial state of filter l as $\mathbf{e}_{0n}^{[l]} \triangleq \mathbf{x}_{n-1} - \mathbf{x}_{0n}^{[l]}$. It can be proved that their means are related by

$$\bar{\mathbf{e}}_{0n}^{[l]} \approx \sum_{i=1}^{N_m} \bar{\gamma}_n^{[li]} \bar{\mathbf{e}}_{n-1}^{[i]} \quad (8.24)$$

For $l = 1, \dots, N_m$, define the model error as

$$\Delta^{a[l]} \triangleq \mathbf{H}_n^a \mathbf{F}_{n-1}^a - \mathbf{H}_n^{a[l]} \mathbf{F}_{n-1}^{a[l]} \quad (8.25)$$

compute the mean of the augmented residual as

$$\bar{\mathbf{r}}_n^{a[l]} = \mathbf{H}_n^{a[l]} \mathbf{F}_{n-1}^{a[l]} \bar{\mathbf{e}}_{0n}^{a[l]} + \Delta^{a[l]} \bar{\mathbf{x}}_{n-1}^a \quad (8.26)$$

and then compute

$$\mathbf{P}_{n|n-1}^{a[l]} = \mathbf{F}_{n-1}^{a[l]} \bar{\mathbf{P}}_{0n}^{a[l]} \left(\mathbf{F}_{n-1}^{a[l]} \right)^H + \mathbf{Q}_{n-1}^{a[l]} \quad (8.27)$$

$$\mathbf{S}_n^{a[l]} = \mathbf{H}_n^{a[l]} \mathbf{P}_{n|n-1}^{a[l]} (\mathbf{H}_n^{a[l]})^H + \mathbf{R}_n^{a[l]} \quad (8.28)$$

$$\mathbf{K}_n^{a[l]} = \mathbf{P}_{n|n-1}^{a[l]} (\mathbf{H}_n^{a[l]})^H (\mathbf{S}_n^{a[l]})^{-1} \quad (8.29)$$

$$\bar{\mathbf{x}}_{n|n}^{[l]} = \mathbf{F}_{n-1}^{a[l]} \mathbf{x}_{0n}^{a[l]} + \mathbf{K}_n^{a[l]} \bar{\mathbf{r}}_n^{a[l]} \quad (8.30)$$

$$\bar{\mathbf{P}}_{n|n}^{a[l]} = \bar{\mathbf{P}}_{n|n-1}^{a[l]} - \bar{\mathbf{K}}_n^{a[l]} \mathbf{H}_n^{a[l]} \bar{\mathbf{P}}_{n|n-1}^{a[l]} \quad (8.31)$$

$$\begin{aligned} \bar{\mathbf{e}}_n^{a[l]} = & \left(\mathbf{I} - \mathbf{K}_n^{a[l]} \mathbf{H}_n^{a[l]} \right) \mathbf{F}_{n-1}^{a[l]} \bar{\mathbf{e}}_{0n}^{a[l]} + \\ & \left(\mathbf{F}_{n-1}^a - \mathbf{F}_{n-1}^{a[l]} - \mathbf{K}_n^{a[l]} \Delta^{a[l]} \right) \bar{\mathbf{x}}_{n-1}^a \end{aligned} \quad (8.32)$$

Update

$$\bar{\mathbf{x}}_n^a = \mathbf{F}_{n-1}^a \bar{\mathbf{x}}_{n-1}^a \quad (8.33)$$

Step 3. Compute the cross-covariances of augmented residuals and augmented estimates of the mode-conditioned WL-QKF. For $l = 1, \dots, N_m$, $i = 1, \dots, N_m$, compute the cross-covariances of augmented residuals as

$$\begin{aligned} & \text{Cov} \left(\mathbf{r}_n^{a[l]}, \mathbf{r}_n^{a[i]} \right) \\ &= \mathbf{H}_n^a \mathbf{Q}_{n-1}^a \mathbf{H}_n^{aH} + \Delta^{a[l]} \text{Cov} \left(\mathbf{x}_{n-1}^a, \mathbf{x}_{n-1}^a \right) (\Delta^{a[l]})^H + \\ & \quad \mathbf{H}_n^{a[l]} \mathbf{F}_{n-1}^{a[l]} \text{Cov} \left(\mathbf{e}_{0n}^{a[l]}, \mathbf{e}_{0n}^{a[i]} \right) (\mathbf{H}_n^{a[i]} \mathbf{F}_{n-1}^{a[i]})^H + \\ & \quad \mathbf{H}_n^{a[l]} \mathbf{F}_{n-1}^{a[l]} \text{Cov} \left(\mathbf{e}_{0n}^{a[l]}, \mathbf{x}_{n-1}^a \right) (\Delta^{a[i]})^H + \\ & \quad \Delta^{a[l]} \text{Cov} \left(\mathbf{e}_{0n}^{a[i]}, \mathbf{x}_{n-1}^a \right)^H (\mathbf{H}_n^{a[i]} \mathbf{F}_{n-1}^{a[i]})^H + \mathbf{R}_n^a \end{aligned} \quad (8.34)$$

and the cross-covariances of augmented estimation errors as

$$\begin{aligned} & \text{Cov} \left(\mathbf{e}_n^{a[l]}, \mathbf{e}_n^{a[i]} \right) \\ &= \left(\mathbf{I} - \mathbf{K}_n^{a[l]} \mathbf{H}_n^{a[l]} \right) \mathbf{F}_{n-1}^{a[l]} \text{Cov} \left(\mathbf{e}_{0n}^{a[l]}, \mathbf{e}_{0n}^{a[i]} \right) \left(\mathbf{F}_{n-1}^{a[i]} - \mathbf{K}_n^{a[i]} \mathbf{H}_n^{a[i]} \mathbf{F}_{n-1}^{a[i]} \right)^H + \\ & \quad \left(\mathbf{F}_{n-1}^a - \mathbf{F}_{n-1}^{a[l]} - \mathbf{K}_n^{a[l]} \Delta^{a[l]} \right) \text{Cov} \left(\mathbf{x}_{n-1}^a, \mathbf{x}_{n-1}^a \right) \left(\mathbf{F}_{n-1}^a - \mathbf{F}_{n-1}^{a[i]} - \mathbf{K}_n^{a[i]} \Delta^{a[i]} \right)^H \\ & \quad + \left(\mathbf{F}_{n-1}^{a[l]} - \mathbf{K}_n^{a[l]} \mathbf{H}_n^{a[l]} \mathbf{F}_{n-1}^{a[l]} \right) \text{Cov} \left(\mathbf{e}_{0n}^{a[l]}, \mathbf{x}_{n-1}^a \right) \left(\mathbf{F}_{n-1}^a - \mathbf{F}_{n-1}^{a[i]} - \mathbf{K}_n^{a[i]} \Delta^{a[i]} \right)^H \\ & \quad + \left(\mathbf{F}_{n-1}^a - \mathbf{F}_{n-1}^{a[l]} - \mathbf{K}_n^{a[l]} \Delta^{a[l]} \right) \text{Cov} \left(\mathbf{e}_{0n}^{a[i]}, \mathbf{x}_{n-1}^a \right)^H \left(\mathbf{F}_{n-1}^{a[i]} - \mathbf{K}_n^{a[i]} \mathbf{H}_n^{a[i]} \mathbf{F}_{n-1}^{a[i]} \right)^H \\ & \quad + \left(\mathbf{I} - \mathbf{K}_n^{a[l]} \mathbf{H}_n^{a[l]} \right) \mathbf{Q}_{n-1}^a \left(\mathbf{I} - \mathbf{K}_n^{a[i]} \mathbf{H}_n^{a[i]} \right)^H + \mathbf{K}_n^{a[l]} \mathbf{R}_n^a \left(\mathbf{K}_n^{a[i]} \right)^H \end{aligned}$$

where

$$\text{Cov} \left(\mathbf{e}_{0n}^{a[l]}, \mathbf{x}_{n-1}^a \right) \approx \sum_{i=1}^{N_m} \bar{\gamma}_n^{[li]} \text{Cov} \left(\mathbf{e}_{n-1}^{a[i]}, \mathbf{x}_{n-1}^a \right) \quad (8.35)$$

$$\text{Cov} \left(\mathbf{e}_{0n}^{a[l]}, \mathbf{e}_{0n}^{a[i]} \right) \approx \sum_{k=1}^{N_m} \sum_{j=1}^{N_m} \bar{\gamma}_n^{[lk]} \bar{\gamma}_n^{[ij]} \text{Cov} \left(\mathbf{e}_{n-1}^{a[k]}, \mathbf{e}_{n-1}^{a[j]} \right) \quad (8.36)$$

For $l = 1, \dots, N_m$, update

$$\begin{aligned} \text{Cov} \left(\mathbf{e}_n^{a[l]}, \mathbf{x}_n^a \right) &= \left(\mathbf{I} - \mathbf{K}_n^{a[l]} \mathbf{H}_n^a \right) \mathbf{Q}_{n-1}^a + \left(\mathbf{I} - \mathbf{K}_n^{a[l]} \mathbf{H}_n^{a[l]} \right) \mathbf{F}_{n-1}^{a[l]} \text{Cov} \left(\mathbf{e}_{0n}^{a[l]}, \mathbf{x}_{n-1}^a \right) \mathbf{F}_{n-1}^{aH} \\ &\quad + \left(\mathbf{F}_{n-1}^a - \mathbf{F}_{n-1}^{a[l]} - \mathbf{K}_n^{a[l]} \Delta^a \right) \text{Cov} \left(\mathbf{x}_{n-1}^a, \mathbf{x}_{n-1}^a \right) \mathbf{F}_{n-1}^{aH} \end{aligned} \quad (8.37)$$

$$\text{Cov} \left(\mathbf{x}_n^a, \mathbf{x}_n^a \right) = \mathbf{F}_{n-1}^a \text{Cov} \left(\mathbf{x}_{n-1}^a, \mathbf{x}_{n-1}^a \right) \mathbf{F}_{n-1}^{aH} + \mathbf{Q}_{n-1}^a \quad (8.38)$$

Step 4. Compute the means of mode probabilities. For $l = 1, \dots, N_m$, employing the fact that $\mathbf{r}_n^{a[l]} \sim \mathcal{N} \left(\bar{\mathbf{r}}_n^{a[l]}, \mathbf{V}_n^{a[l]} \right)$, where $\mathbf{V}_n^{a[l]} = \text{Cov} \left(\mathbf{r}_n^{a[l]}, \mathbf{r}_n^{a[l]} \right)$, the mean of likelihood function can be computed as

$$\begin{aligned} \bar{\mathcal{L}}_n^{[l]} &= \int_{\mathbb{H}^N} \mathcal{L}_n^{[l]} f_G \left(\mathbf{r}_n^{a[l]} \mid \bar{\mathbf{r}}_n^{a[l]}, \mathbf{V}_n^{a[l]} \right) d\mathbf{r}_n^{[l]} \\ &= \int_{\mathbb{H}^N} f_G \left(\mathbf{r}_n^{a[l]} \mid \mathbf{0}, \mathbf{S}_n^{a[l]} \right) f_G \left(\mathbf{r}_n^{a[l]} \mid \bar{\mathbf{r}}_n^{a[l]}, \mathbf{V}_n^{a[l]} \right) d\mathbf{r}_n^{[l]} \\ &= \left(\frac{2}{\pi} \right)^{2N} \frac{\left[\left(\mathbf{S}_n^{a[l]} \right)^{-1} + \left(\mathbf{V}_n^{a[l]} \right)^{-1} \right]^{-1}}{\left| \mathbf{S}_n^{a[l]} \right|^{0.5} \left| \mathbf{V}_n^{a[l]} \right|^{0.5}} \exp \left\{ -0.5 \left(\bar{\mathbf{r}}_n^{a[l]} \right)^H \left(\mathbf{V}_n^{a[l]} + \mathbf{S}_n^{a[l]} \right)^{-1} \bar{\mathbf{r}}_n^{a[l]} \right\} \end{aligned} \quad (8.39)$$

and the mean of the mode probability as

$$\begin{aligned} \bar{\mu}_n^{[l]} &= \int_{\mathbb{H}^N} \mu_n^{[l]} f_G \left(\mathbf{r}_n^{a[l]} \mid \bar{\mathbf{r}}_n^{a[l]}, \mathbf{V}_n^{a[l]} \right) d\mathbf{r}_n^{[l]} \\ &= \int_{\mathbb{H}^N} \frac{\sum_{i=1}^{N_m} \pi^{[il]} \mu_{n-1}^{[i]} f_G \left(\mathbf{r}_n^a \mid \mathbf{0}, \mathbf{S}_n^{a[l]} \right) f_G \left(\mathbf{r}_n^a \mid \bar{\mathbf{r}}_n^{a[l]}, \mathbf{V}_n^{a[l]} \right)}{\sum_{i=1}^{N_m} \sum_{j=1}^{N_m} \pi^{[ji]} \mu_{n-1}^{[j]} f_G \left(\mathbf{r}_n^a \mid \mathbf{0}, \mathbf{S}_n^{a[i]} \right)} d\mathbf{r}_n \\ &= \bar{\mathcal{L}}_n^{[l]} \sum_{i=1}^{N_m} \pi^{[il]} \bar{\mu}_{n-1}^{[i]} \int_{\mathbb{H}^N} \frac{f_G \left(\mathbf{r}_n^a \mid \left[\mathbf{V}_n^{a[l]} \left(\mathbf{S}_n^{a[l]} \right)^{-1} + \mathbf{I} \right]^{-1} \bar{\mathbf{r}}_n^{a[l]}, \left[\left(\mathbf{S}_n^{a[l]} \right)^{-1} + \left(\mathbf{V}_n^{a[l]} \right)^{-1} \right]^{-1} \right)}{\sum_{i=1}^{N_m} \sum_{j=1}^{N_m} \pi^{[ji]} \bar{\mu}_{n-1}^{[j]} f_G \left(\mathbf{r}_n^a \mid \mathbf{0}, \mathbf{S}_n^{a[i]} \right)} d\mathbf{r}_n \end{aligned} \quad (8.40)$$

Step 5. Compute the MSE of the overall estimate. The covariance matrix of

the overall augmented estimation error is given by

$$\begin{aligned} & E \left\{ \left(\mathbf{x}_n^a - \hat{\mathbf{x}}_{n|n}^a \right) \left(\mathbf{x}_n^a - \hat{\mathbf{x}}_{n|n}^a \right)^H \mid \mathcal{S} \right\} \\ & \approx \sum_{l=1}^{N_m} \sum_{i=1}^{N_m} \bar{\mu}_n^{[l]} \bar{\mu}_n^{[i]} \left[\text{Cov} \left(\mathbf{e}_n^{a[l]}, \mathbf{e}_n^{a[i]} \right) + \bar{\mathbf{e}}_n^{a[l]} \left(\bar{\mathbf{e}}_n^{a[i]} \right)^H \right] \end{aligned} \quad (8.41)$$

and the MSE of the overall estimate is

$$E \left\{ \left(\mathbf{x}_n - \hat{\mathbf{x}}_{n|n} \right)^H \left(\mathbf{x}_n - \hat{\mathbf{x}}_{n|n} \right) \mid \mathcal{S} \right\} = \frac{1}{4} \text{Tr} \left(E \left\{ \left(\mathbf{x}_n^a - \hat{\mathbf{x}}_{n|n}^a \right) \left(\mathbf{x}_n^a - \hat{\mathbf{x}}_{n|n}^a \right)^H \mid \mathcal{S} \right\} \right) \quad (8.42)$$

Remark 8.1. In order to derive a closed-form solution, the approximations in (8.22), (8.35), (8.36) and (8.41) based on the zero correlation assumption between the involved variables have been applied.

8.2.3 WL-QSMM Algorithm

If the system mode is time-invariant, the WL-QIMM algorithm can reduce to a WL-QSMM algorithm, which does not require reinitialisation and exhibits simplified computation for the mode probabilities [108]. Specifically, the time-invariant system mode implies that

$$p \left(\tau_n^{[l]} \mid \tau_{n-1}^{[i]} \right) = \begin{cases} 1 & l = i \\ 0 & l \neq i \end{cases} \quad (8.43)$$

so that (8.19) becomes

$$\mu_n^{[l]} = \frac{\mathcal{L}_n^{[l]} \mu_{n-1}^{[l]}}{\sum_{i=1}^{N_m} \mathcal{L}_n^{[i]} \mu_{n-1}^{[i]}} \quad (8.44)$$

An iteration of the WL-QSMM algorithm includes two steps as follows.

Step 1 (Mode-conditioned Kalman filtering). Implement N_m parallel WL-QKFs with the initial states and covariances being the estimates obtained

in the previous iteration. This step is the same as Step 2 in WL-QIMM except that $\mathbf{x}_{0n}^{[l]}$ and $\mathbf{P}_{0n}^{a[l]}$ in (8.12) and (8.13) are replaced with $\hat{\mathbf{x}}_{n-1|n-1}^{a[l]}$ and $\mathbf{P}_{n-1|n-1}^{a[l]}$ for $l = 1, \dots, N_m$.

Step 2 (Combination). Compute the mode probabilities according to (8.8) and (8.44). The overall augmented estimate and overall augmented error covariance are given by (8.20) and (8.21).

The performance analysis algorithm presented in Section 8.2.2 applies to the WL-QSMM algorithm by replacing $\bar{\mathbf{e}}_{0n}^{a[l]}$, $\mathbf{x}_{0n}^{[l]}$ and $\mathbf{P}_{0n}^{a[l]}$ with $\bar{\mathbf{e}}_{n-1}^{a[l]}$, $\hat{\mathbf{x}}_{n-1|n-1}^{a[l]}$ and $\mathbf{P}_{n-1|n-1}^{a[l]}$, for $l = 1, \dots, N_m$.

Convergence analysis

Next, the WL-QSMM algorithm is proven to converge to the candidate model closest to the true system model in the sense of the Kullback–Leibler divergence. The Kullback–Leibler divergence between $\mathcal{N}(E\{\mathbf{r}_n^a\}, \mathbf{S}_n^a)$ and $\mathcal{N}(E\{\mathbf{r}_n^{a[l]}\}, \mathbf{S}_n^{a[l]})$, where $\mathbf{r}_n^a$ is the augmented residual corresponding to the filter that exactly matches the true system model and $\mathbf{S}_n^a$ is its covariance matrix, is given by

$$\begin{aligned} K_l &\triangleq \frac{1}{2} \ln \frac{|\mathbf{S}_n^{a[l]}|}{|\mathbf{S}_n^a|} + \frac{1}{2} \text{Tr} \left(\Re \left[\left(\mathbf{S}_n^{a[l]} \right)^{-1} \mathbf{S}_n^a \right] - \mathbf{I} \right) + \\ &\quad \frac{1}{2} \left(E \left\{ \mathbf{r}_n^{a[l]} \right\} - E \left\{ \mathbf{r}_n^a \right\} \right)^H \left(\mathbf{S}_n^{a[l]} \right)^{-1} \left(E \left\{ \mathbf{r}_n^{a[l]} \right\} - E \left\{ \mathbf{r}_n^a \right\} \right) \\ &= \frac{1}{2} \ln \frac{|\mathbf{S}_n^{a[l]}|}{|\mathbf{S}_n^a|} + \frac{1}{2} \text{Tr} \left(\Re \left[\left(\mathbf{S}_n^{a[l]} \right)^{-1} E \left\{ \mathbf{r}_n^{a[l]} \left(\mathbf{r}_n^{a[l]} \right)^H \right\} \right] - \mathbf{I} \right) \end{aligned} \quad (8.45)$$

The second equality in (8.45) is based on $E\{\mathbf{r}_n^a\} = \mathbf{0}$ and

$$E \left\{ \mathbf{r}_n^{a[l]} \left(\mathbf{r}_n^{a[l]} \right)^H \right\} - E \left\{ \mathbf{r}_n^{a[l]} \right\} \left(E \left\{ \mathbf{r}_n^{a[l]} \right\} \right)^H = \mathbf{S}_n^a \quad (8.46)$$

which is a straightforward extension of the analysis result for real-valued mismatched Kalman filters [123] to the quaternion domain. Denote the index of the candidate model closest to the true system model in the sense of Kullback–Leibler

divergence by $t = \arg \min_{l \in \{1, 2, \dots, N_m\}} K_l$. For $l \neq t$, define $\Lambda_n^{[l]} \triangleq \mu_n^{[l]} / \mu_n^{[t]}$ and obtain

$$\begin{aligned} \frac{\Lambda_n^{[l]}}{\Lambda_{n-1}^{[l]}} = \frac{\mathcal{L}_n^{[l]}}{\mathcal{L}_n^{[t]}} &= \sqrt{\frac{|\mathbf{S}_n^{a[t]}|}{|\mathbf{S}_n^{a[l]}|}} \exp \left\{ \frac{1}{2} \left[\left(\mathbf{r}_n^{a[t]} \right)^H \left(\mathbf{S}_n^{a[t]} \right)^{-1} \mathbf{r}_n^{a[t]} - \left(\mathbf{r}_n^{a[l]} \right)^H \left(\mathbf{S}_n^{a[l]} \right)^{-1} \mathbf{r}_n^{a[l]} \right] \right\} \\ &= \sqrt{\frac{|\mathbf{S}_n^{a[t]}|}{|\mathbf{S}_n^{a[l]}|}} \exp \left\{ \frac{1}{2} \text{Tr} \left(\Re \left[\left(\mathbf{S}_n^{a[t]} \right)^{-1} \mathbf{r}_n^{a[t]} \left(\mathbf{r}_n^{a[t]} \right)^H - \left(\mathbf{S}_n^{a[l]} \right)^{-1} \mathbf{r}_n^{a[l]} \left(\mathbf{r}_n^{a[l]} \right)^H \right] \right) \right\} \end{aligned} \quad (8.47)$$

Applying the the statistical expectation operator and the natural logarithm to 8.47 yields

$$\ln \left(E \left\{ \Lambda_n^{[l]} / \Lambda_{n-1}^{[l]} \right\} \right) = K_t - K_l < 0 \quad (8.48)$$

and $E \left\{ \Lambda_n^{[l]} / \Lambda_{n-1}^{[l]} \right\} < 1$, which implies that $\mu_n^{[l]}$ decreases to 0 and hence $\mu_n^{[t]}$ increases to 1 when $n \rightarrow \infty$. This proves the convergence of the WL-QSMM algorithm to model t .

Remark 8.2. Since the Kullback–Leibler divergence in (8.45) is nonnegative and vanishes only when the filter model exactly matches the true system, the WL-QSMM will converge to the true system model if the true model belongs to the candidate model set.

8.2.4 Semi-Widely Linear and Strictly Linear QMMAE

For semi-widely linear quaternion systems with $\mathbb{C}^n$ -improper quaternion Gaussian noise, the above WL-QMMAE algorithms reduce to corresponding SWL-QMMAE algorithms, by replacing the augmented forms of variables with the semi-augmented forms. Based on the probability density function in (2.16), the likelihood functions in the SWL-QIMM and SWL-QSMM algorithms are given by

$$\mathcal{L}_n^{[l]} = f_G \left(\mathbf{r}_n^{b[l]} \mid \mathbf{0}, \mathbf{S}_n^{b[l]} \right) \quad (8.49)$$

For strictly linear quaternion systems with $\mathbb{H}$ -proper quaternion Gaussian noise, the WL-QMMAE algorithms reduce to corresponding SL-QMMAE algo-

rithms, by replacing the augmented forms of variables with the original forms. Since the QMGD is associated with the probability density function in (2.17) under this condition, the likelihood functions in the SL-QIMM and SL-QSMM algorithms become

$$\mathcal{L}_n^{[l]} = f_G \left(\mathbf{r}_n^{[l]} \mid \mathbf{0}, \mathbf{S}_n^{[l]} \right) \quad (8.50)$$

The SL-QMMAE and SWL-QMMAE algorithms can be seen as reduced-order WL-QMMAE algorithms [124], and therefore are suboptimal for widely linear quaternion systems and general improper quaternion noise.

8.3 Simulations

8.3.1 QIMM for Filtering Quaternion Signals

This section presents simulation results for evaluating the SL-QIMM, SWL-QIMM and WL-QIMM algorithms and the corresponding performance analysis algorithms. Since the linear nature (strictly, semi-widely or widely linear) of the involved quaternion system essentially affects the estimation performance, a strictly linear quaternion hybrid system with noise at different levels of impropriety was employed, to study the impact of impropriety on the estimation performance. The considered system, which was used in [116], is represented by the following one-dimensional state-space model

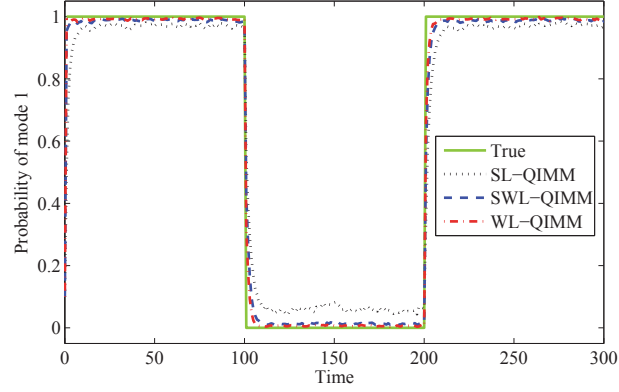
$$\begin{aligned} x_n &= b^{[l]}x_{n-1} + u_{n-1}^{[l]} \\ z_n &= c^{[l]}x_n + v_n^{[l]} \end{aligned} \quad (8.51)$$

where $l \in \{1, 2\}$, $b^{[1]} = 0.95$, $b^{[2]} = 0.85$, $c^{[1]} = 1$, $c^{[2]} = 0.4$, and $w_{n-1}^{[1]}$, $w_n^{[2]}$, $v_n^{[1]}$, $v_n^{[2]}$ are proper or improper quaternion Gaussian noise with power 2dB, 0dB, 0dB, 0dB. The system mode was 1 from $n = 1$ to 100 and from $n = 200$ to 300, and was 2 from $n = 100$ to 200. In each experiment, 500 Monte Carlo simulation runs of the three considered QIMM algorithms were implemented and the estimation performance was evaluated through the averaged MSE. The initial mode probabilities were set

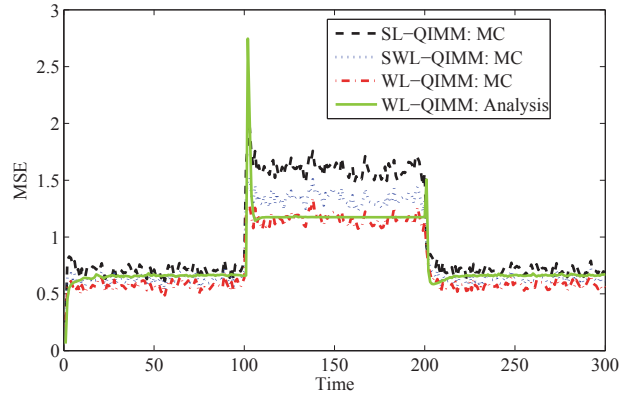
to $\mu_0^{[1]} = 0.1$ and $\mu_0^{[2]} = 0.9$. The proposed performance analysis algorithms were also implemented for comparison with Monte Carlo results.

Figure 8.1 illustrates the simulation results of the three QIMM algorithms for improper state noise with impropriety coefficients $r_i = r_j = r_\kappa = 0.2$ and improper observation noise with $r_i = r_j = r_\kappa = 0.9$. Figure 8.1 (a) shows the average probabilities of the hypothesising mode 1 over 500 Monte Carlo runs. Observe that the WL-QIMM converged faster to the true mode than the other two algorithms, while the SL-QIMM exhibited the slowest convergence. As shown in Figure 8.1 (b), the WL-QIMM achieved the smallest MSE while the SL-QSMM attained the largest MSE; the MSE computed from the proposed performance analysis algorithm for WL-QIMM matched the MSE averaged over the Monte Carlo runs, with 12% difference on average. The error arose from the approximations in (8.22), (8.35), (8.36), (8.41) and the integral calculation in (8.40). The 500 Monte Carlo runs of WL-QIMM took 79 minutes while the proposed performance analysis algorithm took only 2 minutes on a standard PC. Obviously, this computation efficiency advantage of the performance analysis algorithm will increase with the data size. Figures 8.1 (b), 8.2 (b) and 8.3 (b) display significant errors of the performance analysis algorithm in a short time after the mode transitions. The phenomenon is due to the slower response of the average probabilities calculated as (8.22) and (8.40) than the real probabilities in the WL-QIMM algorithm, because of the approximation in (8.22) and the integrity in (8.40). Further research is needed to improve the accuracy of those average probabilities.

Figure 8.2 illustrates the simulation results of the three QIMM algorithms for $\mathbb{C}^2$ -improper state and observation noise. Figure 8.2 (a) shows the average probabilities of the hypothesising mode 1 over 500 Monte Carlo runs, and Figure 8.2 (b) shows the MSEs of the three QIMM algorithms. As shown in Figure 8.2, the WL-QIMM and SWL-QIMM exhibited the same performance, achieving a lower MSE than the SL-QIMM, and the MSE computed from the performance analysis algorithm for WL-QIMM matched the MSE averaged over the Monte Carlo runs,



(a) Probability of the time-varying mode 1

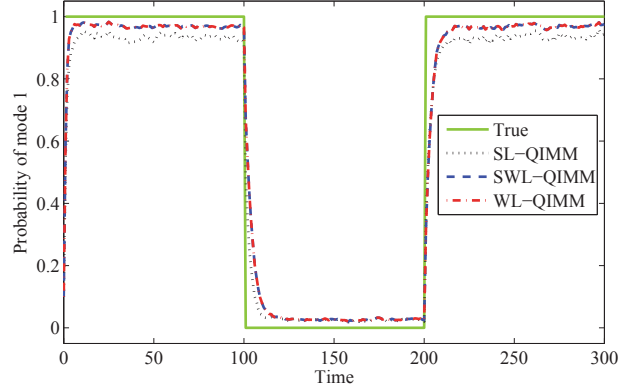


(b) MSE of the algorithms considered

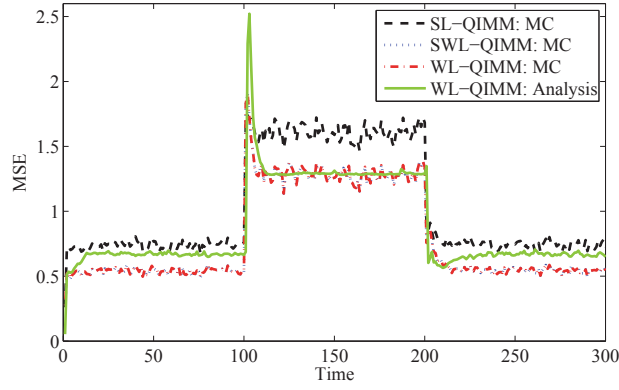
Figure 8.1: Learning curves of the three QIMM algorithms when the state and observation noise are improper.

with 16% difference on average.

Figure 8.3 illustrates the simulation results of the three QIMM algorithms for $\mathbb{H}$ -proper state and observation noise. Figure 8.3 (a) shows the average probabilities of the hypothesising mode 1 over the 500 Monte Carlo simulation runs, and Figure 8.3 (b) shows the MSEs of the three QIMM algorithms. As shown in Figure 8.3, the three QIMM algorithms had the same performance for $\mathbb{H}$ -proper noise, and the MSE computed from the performance analysis algorithm for WL-QIMM matched the MSE averaged over the Monte Carlo runs, with 14% difference on average.



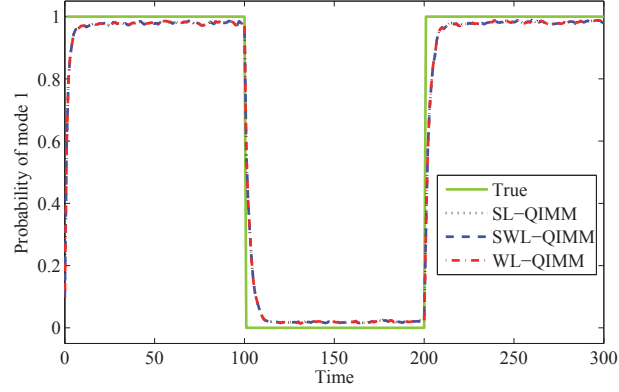
(a) Probability of the time-varying mode 1



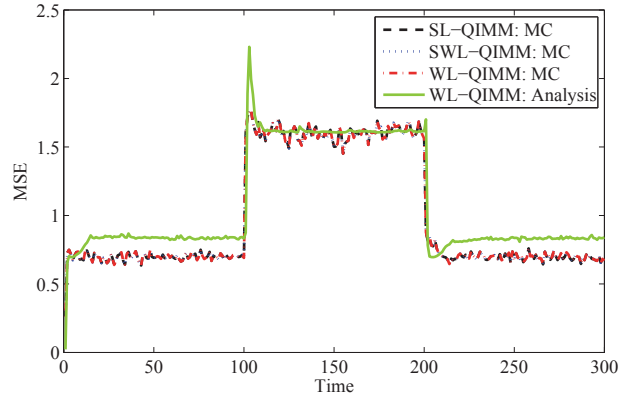
(b) MSE of the algorithms considered

Figure 8.2: Learning curves of the three QIMM algorithms when the state and observation noise are $\mathbb{C}^2$ -improper.

Figure 8.4 depicts the average MSEs of the three QIMM algorithms over the whole simulation period. Figure 8.4 (a) illustrates the results for $\mathbb{H}$ -proper observation noise and improper state noise with varying impropriety coefficients. For simplicity, the three impropriety coefficients of the state noise were set to be equal, that is, $r_i = r_j = r_\kappa = r$. Figure 8.4 (b) illustrates the results for $\mathbb{H}$ -proper state noise and improper observation noise with various impropriety coefficients. The three impropriety coefficients of the observation noise were also equal. Figure 8.4 shows that for the $\mathbb{H}$ -proper noise, the three QIMM algorithms exhibited the same MSE, while for the improper noise, the WL-QIMM achieved the smallest MSE and the SL-QIMM attained the largest MSE. As the impropriety degree of the



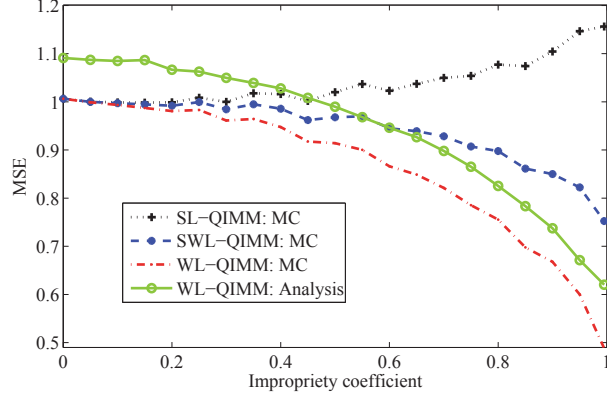
(a) Probability of the time-varying mode 1



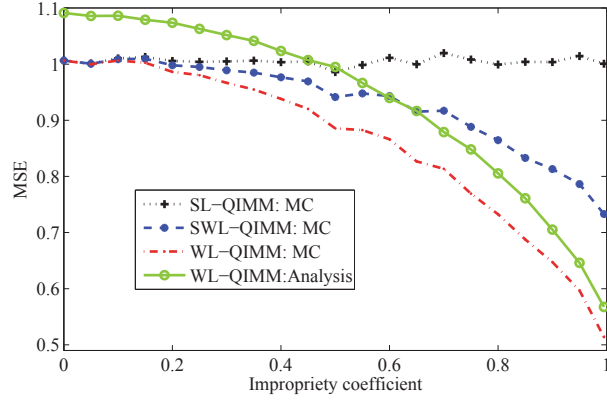
(b) MSE of the algorithms considered

Figure 8.3: Learning curves of the three QIMM algorithms when the state and observation noise are $\mathbb{H}$ -proper.

noise increased, the MSE difference between the three algorithms also increased. Interestingly, the MSE of SL-QIMM increased with the impropriety coefficient of state noise but was unaffected by the impropriety coefficient of observation noise, while the MSEs of SWL-QIMM and WL-QIMM decreased with the impropriety coefficients of state and observation noise. The MSE estimated from the performance analysis algorithm for WL-QIMM shows the same trend as the Monte Carlo simulation result, with about 10% difference in value.



(a) Improper state noise and $\mathbb{H}$ -proper observation noise

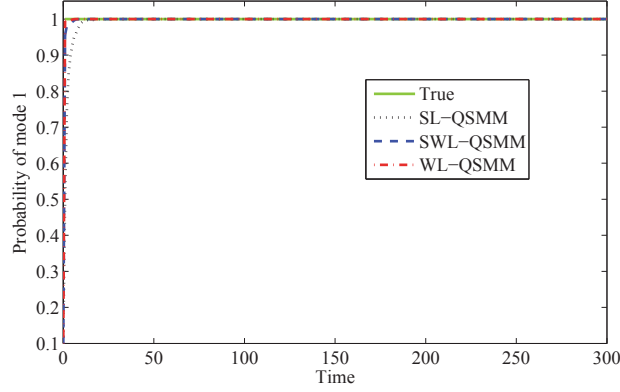


(b) $\mathbb{H}$ -proper state noise and improper observation noise

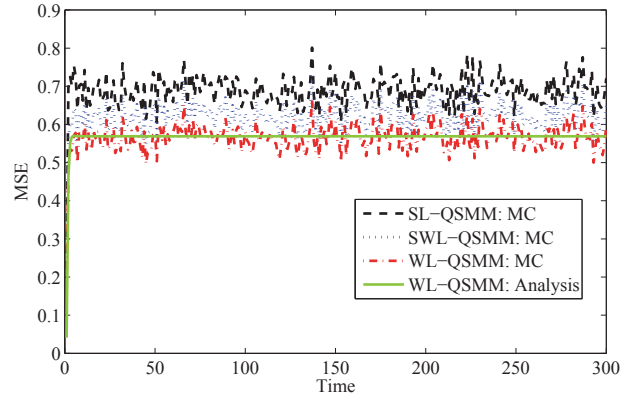
Figure 8.4: The Monte Carlo (MC) simulated MSEs of the three QIMM algorithms and the MSE computed from the proposed performance analysis algorithm (Analysis) for WL-QIMM, as a function of the improperity coefficient of noise.

8.3.2 QSMM for Filtering Quaternion Signals

This section presents simulation results for evaluating the SL-QSMM, SWL-QSMM and WL-QSMM algorithms and the performance analysis algorithm for WL-QSMM. The simulation settings were the same as that for the QIMM algorithms in Section 8.3.1, except that the system mode was constantly 1. Figure 8.5 illustrates the simulation results of the three QSMM algorithms for improper state noise with $r_i = r_j = r_\kappa = 0.2$ and improper observation noise with $r_i = r_j = r_\kappa = 0.9$. Figure 8.5 (a) shows the average probabilities of the hypothesising mode 1 over



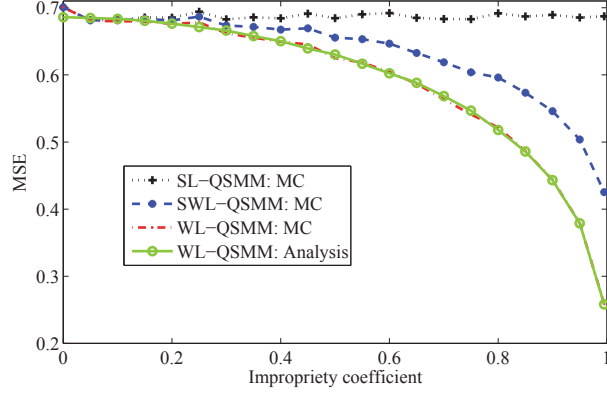
(a) Probability of the time-varying mode 1



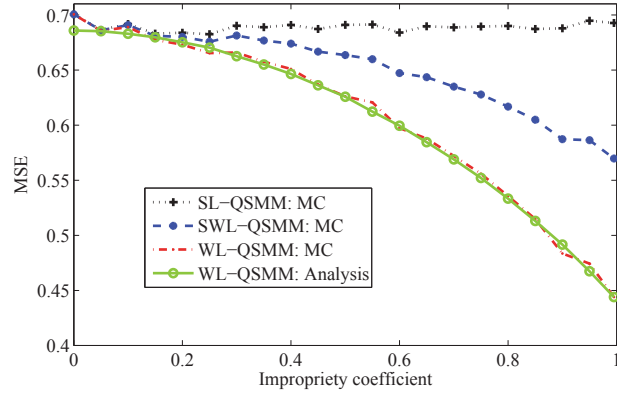
(b) MSE of the algorithms considered

Figure 8.5: Learning curves of the three QSMM algorithms when the state and observation noise are improper.

500 Monte Carlo runs. Observe that the WL-QSMM converged faster to the true mode than the other two algorithms, and the SL-QSMM exhibited the slowest convergence. As shown in Figure 8.5 (b), the WL-QSMM achieved the smallest MSE while the SL-QSMM attained the largest MSE; the MSE computed from the performance analysis algorithm for WL-QSMM matched the MSE averaged over the Monte Carlo runs, with 5% difference. The performance analysis algorithm for WL-QSMM was more accurate than the analysis algorithm for WL-QIMM simulated in Section 8.3.1 because of the less complicated mechanism of the QSMM algorithm compared to the QIMM algorithm.



(a) Improper state noise and $\mathbb{H}$ -proper observation noise



(b) $\mathbb{H}$ -proper state noise and improper observation noise

Figure 8.6: The Monte Carlo (MC) simulated MSEs of the three QSMM algorithms and the MSE computed from the proposed performance analysis algorithm (Analysis) for WL-QSMM, as a function of the impropriety coefficient of noise.

Figure 8.6 depicts the average MSEs of the three QSMM algorithms over the whole simulation period. Figure 8.6 (a) illustrates the results for $\mathbb{H}$ -proper observation noise and improper state noise, over a range of impropriety coefficients. For simplicity, the three impropriety coefficients of the state noise were set to be equal. Figure 8.6 (b) illustrates the results for $\mathbb{H}$ -proper state noise and improper observation noise with various impropriety coefficients. The three impropriety coefficients of the observation noise were also equal. Figure 8.6 shows that for $\mathbb{H}$ -proper noise, the three QSMM algorithms exhibited the same MSE, while for the improper noise, the WL-QSMM achieved the smallest MSE and the SL-QSMM

attained the largest MSE. As the impropriety degree of the noise increased, the MSE difference between the three algorithms increased. It is also noticeable that the MSEs of WL-QSMM and SWL-QSMM decreased with the impropriety coefficients of state and observation noise, and the MSE of SL-QSMM was unaffected by the impropriety coefficients. The MSE estimated from the performance analysis algorithm for WL-QSMM matched the Monte Carlo simulation result.

8.3.3 QIMM for 3D Target Tracking

The effectiveness of the proposed WL-QIMM algorithm is next illustrated in the context of target tracking in a 3D space. The task is to track the position and velocity of a moving target using the measurements of its position. Denote the position by (X_n, Y_n, Z_n) , velocity by $(\dot{X}_n, \dot{Y}_n, \dot{Z}_n)$, and acceleration by $(\ddot{X}_n, \ddot{Y}_n, \ddot{Z}_n)$. Consider two quaternion-valued kinematic models. The first one is the constant velocity model given by [108]

$$\mathbf{x}_n = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \mathbf{x}_{n-1} + \mathbf{u}_{n-1}$$

with the measurement given by

$$z_n = \begin{bmatrix} 1 & 0 \end{bmatrix} \mathbf{x}_n + v_n$$

where $\mathbf{x}_n = [\iota X_n + jY_n + \kappa Z_n; \iota \dot{X}_n + j\dot{Y}_n + \kappa \dot{Z}_n]$, Δt is the sampling interval, $\mathbf{u}_{n-1}$ a zero-mean pure quaternion Gaussian noise vector, and v_n zero-mean pure quaternion Gaussian noise. The covariance matrices of state noise are

$$\mathbf{C}_{\mathbf{u}\mathbf{u}} = \begin{bmatrix} \frac{1}{3}\Delta t^3 & \frac{1}{2}\Delta t^2 \\ \frac{1}{2}\Delta t^2 & \Delta t \end{bmatrix} \sigma_u^2$$

$$\mathbf{C}_{\mathbf{u}\mathbf{u}^\iota} = r_{u\iota} \mathbf{C}_{\mathbf{u}\mathbf{u}}, \quad \mathbf{C}_{\mathbf{u}\mathbf{u}^j} = r_{uj} \mathbf{C}_{\mathbf{u}\mathbf{u}}, \quad \mathbf{C}_{\mathbf{u}\mathbf{u}^\kappa} = r_{u\kappa} \mathbf{C}_{\mathbf{u}\mathbf{u}}$$

where σ_u^2 indicates the power of state noise, and $r_{ui}, r_{uj}, r_{uk} \in \mathbb{R}$ are the impropriety coefficients satisfying $r_{ui} + r_{uj} + r_{uk} = -1$. The covariances of observation noise are $C_{vv} = \sigma_v^2$, $C_{vv^i} = r_{vi}C_{vv}$, $C_{vv^j} = r_{vj}C_{vv}$, $C_{vv^\kappa} = r_{v\kappa}C_{vv}$, where σ_v^2 indicates the power of observation noise, and $r_{vi}, r_{vj}, r_{vk} \in \mathbb{R}$ are the impropriety coefficients satisfying $r_{vi} + r_{vj} + r_{vk} = -1$.

The second model is the constant acceleration model given by [108]

$$\mathbf{x}_n = \begin{bmatrix} 1 & \Delta t & \frac{1}{2}\Delta t^2 \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x}_{n-1} + \mathbf{u}_{n-1}$$

with the measurement given by

$$z_n = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \mathbf{x}_n + v_n$$

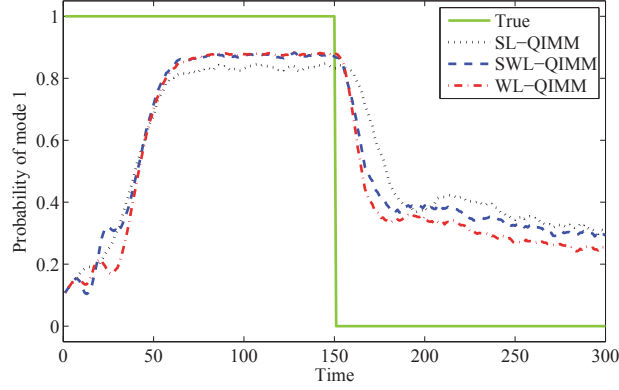
where

$$\mathbf{x}_n = \begin{bmatrix} iX_n + jY_n + \kappa Z_n \\ i\dot{X}_n + j\dot{Y}_n + \kappa\dot{Z}_n \\ i\ddot{X}_n + j\ddot{Y}_n + \kappa\ddot{Z}_n \end{bmatrix}$$

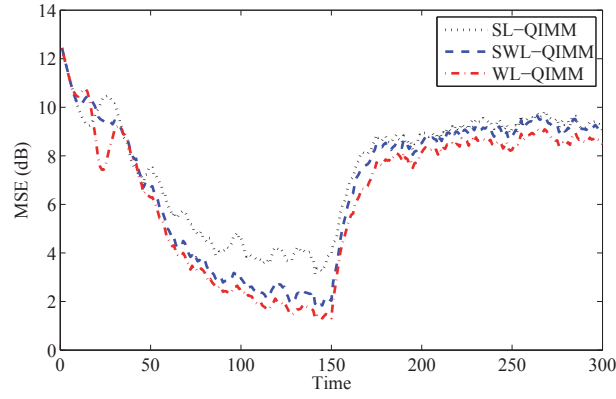
The pure quaternion Gaussian noise $\mathbf{u}_{n-1}$ and v_n in this model have the same properties as those in the constant velocity model, except for

$$\mathbf{C}_{\mathbf{u}\mathbf{u}} = \begin{bmatrix} \frac{1}{20}\Delta t^5 & \frac{1}{8}\Delta t^4 & \frac{1}{6}\Delta t^3 \\ \frac{1}{8}\Delta t^4 & \frac{1}{3}\Delta t^3 & \frac{1}{2}\Delta t^2 \\ \frac{1}{6}\Delta t^3 & \frac{1}{2}\Delta t^2 & dt \end{bmatrix} \sigma_u^2$$

In our simulation, $\Delta t = 0.1$ s, and the target initially had a position of $(0, 0, 0)$, velocity of $(1, 2, 0.5)$ and vanishing acceleration. The target followed the constant velocity model with $\sigma_w^2 = 0.01$ (mode 1) for 150 seconds before following the constant acceleration model with $\sigma_u^2 = 1$ (mode 2) for another 150 seconds. During the whole 300 seconds, $r_{ui} = r_{uj} = r_{uk} = -1/3$, $\rho_{vi} = \rho_{vj} = -25/27$, $\rho_{vk} = 23/27$,



(a) Probability of the time-varying mode 1



(b) MSE of the position and velocity estimation

Figure 8.7: Learning curves of the three QIMM algorithms for target tracking in a 3D space.

$\sigma_v^2 = 1$. In the estimation algorithms, the initial estimate of the state vector was set to $[2i + 2.1j + 2.2\kappa, 0.1i + 0.2j + 0.3\kappa, 0.05i + 0.05j + 0.05\kappa]$, and the initial mode probabilities were $\mu_0^{[1]} = 0.1$, $\mu_0^{[2]} = 0.9$. Figure 8.7 illustrates the Monte Carlo simulated performance of the three considered QIMM algorithms. As shown in Figure 8.7 (a), the probability of mode 1 increased to 0.9 in the first 150 seconds, and decreased towards 0.2 in the second 150 seconds. Figure 8.7 (b) shows that the WL-QIMM achieved a lower MSE of the position and velocity estimation, compared to the SL-QIMM and SWL-QIMM.

8.4 Summary

This chapter has proposed a class of WL-QMMAE algorithms, which employ the augmented second-order quaternion statistics to cater for both proper and improper quaternion Gaussian signals. The WL-QIMM algorithm for tracking time-variant system mode uncertainty and the WL-QSMM algorithm for tracking time-invariant system mode uncertainty have been proposed based on Bayesian inference. A recursive algorithm has been derived to efficiently assess the performance of WL-QIMM, and a convergence proof of WL-QSMM has been provided. It has shown that, as expected, the WL-QMMAE reduces to the semi-widely linear form for $\mathbb{C}^n$ -improper signals, and further reduces to the strictly linear form for $\mathbb{H}$ -proper signals. The effectiveness of the proposed algorithms has been verified by numerical simulations over three- and four-dimensional signal processing case studies.

Chapter 9

Conclusions and Future Work

9.1 Conclusions

Motivated by the natural ability of quaternions to model three- and four-dimensional phenomena, this thesis has developed the theory and methods of statistical signal processing for noncircular quaternion random variables. This has been achieved based on the recently proposed augmented second-order quaternion statistics, widely linear processing framework, and generalised $\mathbb{H}\mathbb{R}$ calculus.

The contributions of this thesis are summarised as following:

1. A technique for the simultaneous diagonalisation of the covariance matrix and the three complementary covariance matrices of noncircular quaternion random variables has been proposed. A singular value decomposition has been proven to be sufficient to approximately diagonalise all of the four quaternion covariance matrices, and this forms a basis for the QAUT, which can be used to decorrelate noncircular quaternion random variables. The QAUT, a fundamental tool for statistical analysis of noncircular quaternion variables, promises to pave the way for various quaternion-valued signal processing techniques and analyses.
2. This thesis has thoroughly explored the performance of existing MMSE estimation techniques for noncircular quaternion signals, including the MSE bound of quaternion MMSE estimation and the convergence performance of quaternion adaptive filters. The closed-form solutions of the quantitative

analysis derived by using the QAUT, have proven the lower MSE of widely linear processing than the strictly linear processing and have related their difference in MSE to the noncircularity and power of the quaternion input signal and the degree of widely linear nature of the system. The mean and mean square analyses of the widely linear, semi-widely linear, and strictly linear quaternion adaptive filters have been undertaken to provide the convergence condition, steady-state MSE and MSD, which have also been related to the noncircularity and power of the input signal. These essential findings provide new insights into the statistical behaviour of the quaternion MMSE estimation algorithms and is potentially useful for algorithm design.

3. Novel adaptive estimation algorithms for noncircular quaternion signals have been proposed and their performance have been illustrated by theory and simulations. First, a QLMMP filtering algorithm has been proposed by decomposing the MSE of quaternion estimation into the cost functions in the estimation of magnitude and phase, allowing for separate control of the magnitude and phase. This algorithm has been shown to outperform the existing QLMS algorithm by virtue of the enhanced degrees of freedom. Second, a four-channel quaternion estimation framework has been proposed by decomposing the MSE into the MSEs in the estimation of real and imaginary parts, allowing for the development of cost-effective alternatives to existing widely linear quaternion algorithms while the processing still remains within the quaternion domain. Finally, in order to deal with the uncertainty of model in quaternion-valued state estimation, a class of WL-QMMAE algorithms have been proposed by combining WL-QKFs and Bayesian inference. To efficiently assess the estimation performance, a recursive algorithm has been proposed in an off-line fashion, and this has much lower computational complexity than the traditionally used Monte Carlo simulation method.

9.2 Future Work

The research into quaternion-valued signal processing has brought to light new directions that are worth pursuing. These directions are as follows:

1. **Convergence performance analysis of undermodelled and overmodelled quaternion adaptive filters.** Whilst transient and steady state performance analysis of a class of quaternion adaptive filters have been introduced in Chapter 5, the analysis is based on the assumption that the length of the filter is equal to that of the system impulse response. In many applications, however, the length of the system impulse response is unknown, so that the filter used in practice might be underfitting or overfitting. A convergence performance analysis of such filters is still an open problem.
2. **Comparison between the convergence performance of widely, semi-widely, and strictly linear quaternion adaptive filters.** In Chapter 5, the convergence performance of the quaternion adaptive filters is based on the assumption that they are the optimal filters for the considered systems. However, in practical applications, the filter model might not match the system and is hence suboptimal, e.g., the strictly linear quaternion adaptive filter is used for a widely linear system. It is required to explore the convergence performance of such suboptimal filters in these scenarios.
3. **Extension of the proposed quaternion estimation algorithms.** Although the QLMMP algorithm, introduced in Chapter 6, is based on the stochastic gradient decent and thus is a variant of the QLMS algorithm, the framework based on separate estimation of the magnitude and phase can be generally applied to various quaternion estimation algorithms. The four-channel linear processing methodology, proposed in Chapter 7, can be applied to a wide range of applications in widely linear processing of quaternion signals, like widely linear series expansions [125]. Whereas the QMMAE

algorithms, presented in Chapter 8, use quaternion Kalman filters for linear state estimation of Gaussian quaternion signals, the estimation framework is general and therefore can incorporate other types of filters and neural networks for nonlinear estimation of non-Gaussian quaternion signals [126,127].

4. **Complete convergence performance analysis of QLMMP.** The performance analysis of the QLMMP filtering algorithm, proposed in Chapter 6, is far from sufficient, and the sufficient convergence condition derived is quite restricted. A complete analysis to quantify the convergence performance of QLMMP is challenging but worth further study. It is expected to obtain a necessary and sufficient condition for convergence and the steady-state performance.
5. **Extension to distributed quaternion-valued signal processing.** In recent years, the study of sensor networks for distributed learning tasks has gained popularity in both academia and industry. The work presented in this thesis can be extended to distributed statistical signal processing of noncircular quaternion random variables over networks, which has the potential to benefit the three- and four-dimensional signal processing in our increasingly interconnected world.

Bibliography

- [1] S. M. Stigler, “Gauss and the invention of least squares,” *The Annals of Statistics*, pp. 465–474, 1981.
- [2] N. Wiener, *Extrapolation, interpolation, and smoothing of stationary time series*, vol. 7. Cambridge, MA: MIT press, 1949.
- [3] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.
- [4] S. Haykin, *Adaptive Filter Theory*. 4 ed., 2002.
- [5] A. H. Sayed, *Adaptive Filters*. Hoboken, NJ: Wiley-IEEE Press, 2008.
- [6] B. Widrow, J. McCool, and M. Ball, “The complex LMS algorithm,” *Proceedings of the IEEE*, vol. 63, p. 719, 1975.
- [7] D. P. Mandic and V. S. L. Goh, *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models*. Hoboken, NJ: Wiley-Blackwell, 2009.
- [8] D. Brandwood, “A complex gradient operator and its application in adaptive array theory,” *IEE Proceedings F - Communications, Radar and Signal Processing*, vol. 130, no. 1, pp. 11–16, 1983.
- [9] K. Kreutz-Delgado, “The complex gradient operator and the CR-calculus,” *arXiv preprint arXiv:0906.4835*, 2009.
- [10] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-valued Data*. Cambridge, UK: Cambridge University Press, 2010.
- [11] W. R. Hamilton, “On quaternions; or on a new system of imaginaries in algebra,” *Philosophical Magazine*, vol. 25, no. 163, pp. 10–13, 1844.

- [12] F. Zhang, “Quaternions and matrices of quaternions,” *Linear Algebra and Its Applications*, vol. 251, pp. 21–57, 1997.
- [13] L. Rodman, *Topics in Quaternion Linear Algebra*. Princeton, NJ: Princeton University Press, 2014.
- [14] J. B. Kuipers, *Quaternions and Rotation Sequences: A Primer with Applications to Orbits, Aerospace and Virtual Reality*. Princeton, NJ: Princeton University Press, 1999.
- [15] S. Miron, N. L. Bihan, and J. I. Mars, “Quaternion-MUSIC for vector-sensor array processing,” *IEEE Transactions on Signal Processing*, vol. 54, no. 4, pp. 1218–1229, 2006.
- [16] J.-F. Gu and K. Wu, “Quaternion modulation for dual-polarized antennas,” *IEEE Communications Letters*, vol. 21, no. 2, pp. 286–289, 2017.
- [17] H. Fourati, N. Manamanni, L. Afilal, and Y. Handrich, “Complementary observer for body segments motion capturing by inertial and magnetic sensors,” *IEEE/ASME Transactions on Mechatronics*, vol. 19, no. 1, pp. 149–157, 2014.
- [18] K. Adhikari, S. Tatinati, W. T. Ang, K. C. Veluvolu, and K. Nazarpour, “A quaternion weighted Fourier linear combiner for modeling physiological tremor,” *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 11, pp. 2336–2346, 2016.
- [19] L. Krieger and F. Grigoli, “Optimal reorientation of geophysical sensors: A quaternion-based analytical solution,” *Geophysics*, vol. 80, no. 2, pp. F19–F30, 2015.
- [20] C. Cheong Took, G. Strbac, K. Aihara, and D. P. Mandic, “Quaternion-valued short-term joint forecasting of three-dimensional wind and atmospheric parameters,” *Renewable Energy*, vol. 36, no. 6, pp. 1754–1760, 2011.

- [21] J. Vía, D. Ramírez, and I. Santamaría, “Properness and widely linear processing of quaternion random vectors,” *IEEE Transactions on Information Theory*, vol. 56, no. 7, pp. 3502–3515, 2010.
- [22] C. Cheong Took and D. P. Mandic, “Augmented second-order statistics of quaternion random signals,” *Signal Processing*, vol. 91, no. 2, pp. 214–224, 2011.
- [23] D. P. Mandic, C. Jahanchahi, and C. Cheong Took, “A quaternion gradient operator and its applications,” *IEEE Signal Processing Letters*, vol. 18, no. 1, pp. 47–50, 2011.
- [24] D. Xu, C. Jahanchahi, C. Cheong Took, and D. P. Mandic, “Enabling quaternion derivatives: The generalized HR calculus,” *Royal Society Open Science*, vol. 2, no. 8, p. 150255, 2015.
- [25] C. Cheong Took and D. P. Mandic, “The quaternion LMS algorithm for adaptive filtering of hypercomplex processes,” *IEEE Transactions on Signal Processing*, vol. 57, no. 4, pp. 1316–1327, 2009.
- [26] Y. Xia, C. Jahanchahi, and D. P. Mandic, “Quaternion-valued echo state networks,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 4, pp. 663–673, 2015.
- [27] S. Javidi, C. Cheong Took, and D. P. Mandic, “Fast independent component analysis algorithm for quaternion valued signals,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 22, no. 12, pp. 1967–1978, 2011.
- [28] T. A. Ell, N. Le Bihan, and S. J. Sangwine, *Quaternion Fourier Transforms for Signal and Image Processing*. Hoboken, NJ: Wiley-ISTE, 2014.
- [29] C. Cheong Took and D. P. Mandic, “A quaternion widely linear adaptive filter,” *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4427–4431, 2010.

- [30] T. Nitta, “A theoretical foundation for the widely linear processing of quaternion-valued data,” *Applied Mathematics*, vol. 4, no. 12, pp. 1616–1620, 2013.
- [31] Y. Xia, C. Jahanchahi, T. Nitta, and D. P. Mandic, “Performance bounds of quaternion estimators,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 12, pp. 3287–3292, 2015.
- [32] D. Xu and D. P. Mandic, “The theory of quaternion matrix derivatives,” *IEEE Transactions on Signal Processing*, vol. 63, no. 3, pp. 1543–1556, 2015.
- [33] D. Xu, Y. Xia, and D. P. Mandic, “Optimization in quaternion dynamic systems: Gradient, Hessian, and learning algorithms,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 2, pp. 249–261, 2015.
- [34] M. Xiang, S. Enshaeifar, A. E. Stott, C. Cheong Took, Y. Xia, S. Kanna, and D. P. Mandic, “Simultaneous diagonalisation of covariance and complementary covariance matrices in quaternion widely linear signal processing.” Submitted to *Signal Processing*, 2017.
- [35] M. Xiang, S. Kanna, S. C. Douglas, and D. P. Mandic, “Performance advantage of quaternion widely linear estimation: An approximate uncorrelating transform approach,” in *Proceedings of 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (Shanghai), pp. 4348–4352, 2016.
- [36] M. Xiang, S. Kanna, and D. P. Mandic, “Performance analysis of quaternion-valued adaptive filters in non-stationary environments.” Submitted to *IEEE Transactions on Signal Processing*, 2017.
- [37] M. Xiang, S. C. Douglas, and D. P. Mandic, “The quaternion least mean magnitude phase adaptive filtering algorithm,” in *the 22nd International Conference on Digital Signal Processing (DSP)*, (London), 2017.

- [38] M. Xiang, C. Cheong Took, and D. P. Mandic, “Cost-effective quaternion minimum mean square error estimation: From widely linear to four-channel processing,” *Signal Processing*, vol. 136, pp. 81–91, 2017.
- [39] C. Jahanchahi and D. P. Mandic, “A class of quaternion Kalman filters,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 3, pp. 533–544, 2014.
- [40] M. Xiang, B. Scalzo Dees, and D. P. Mandic, “Multiple model adaptive estimation for 3D and 4D signals: A widely linear quaternion approach.” Submitted to *IEEE Transactions on Neural Networks and Learning Systems*, 2017.
- [41] J. P. Ward, *Quaternions and Cayley numbers: Algebra and Applications*. Springer, 1997.
- [42] T. Mizoguchi and I. Yamada, “An algebraic translation of Cayley-Dickson linear systems and its applications to online learning,” *IEEE Transactions on Signal Processing*, vol. 62, no. 6, pp. 1438–1453, 2014.
- [43] T. A. Ell and S. J. Sangwine, “Hypercomplex Fourier transforms of color images,” *IEEE Transactions on Image Processing*, vol. 16, no. 1, pp. 22–35, 2007.
- [44] T. A. Ell and S. J. Sangwine, “Quaternion involutions and anti-involutions,” *Computers & Mathematics with Applications*, vol. 53, no. 1, pp. 137–143, 2007.
- [45] L. Huang and W. So, “On left eigenvalues of a quaternionic matrix,” *Linear Algebra and Its Applications*, vol. 323, no. 1-3, pp. 105–116, 2001.
- [46] D. R. Farenick and B. A. Pidkowich, “The spectral theorem in quaternions,” *Linear Algebra and Its Applications*, vol. 371, pp. 75–102, 2003.

- [47] L. Zou, Y. Jiang, and J. Wu, “Location for the right eigenvalues of quaternion matrices,” *Journal of Applied Mathematics and Computing*, vol. 38, no. 1, pp. 71–83, 2012.
- [48] W. Fulton, “Eigenvalues, invariant factors, highest weights, and Schubert calculus,” *Bulletin of the American Mathematical Society*, vol. 37, no. 3, pp. 209–249, 2000.
- [49] S. P. Talebi, S. Kanna, and D. P. Mandic, “Real-time estimation of quaternion impropriety,” in *Proceedings of the 20th IEEE International Conference on Digital Signal Processing (DSP)*, (Singapore), pp. 557–561, 2015.
- [50] J. Vía, D. Palomar, L. Vielva, and I. Santamaría, “Quaternion ICA from second-order statistics,” *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1586–1600, 2011.
- [51] J. Navarro-Moreno, R. M. Fernández-Alcalá, and J. C. Ruiz-Molina, “Semi-widely simulation and estimation of continuous-time $\mathbb{C}^n$ -proper quaternion random signals,” *IEEE Transactions on Signal Processing*, vol. 63, no. 18, pp. 4999–5012, 2015.
- [52] J. Navarro-Moreno and J. C. Ruiz-Molina, “Semi-widely linear estimation of $\mathbb{C}^n$ -proper quaternion random signal vectors under Gaussian and stationary conditions,” *Signal Processing*, vol. 119, pp. 56–66, 2016.
- [53] D. P. Mandic, S. Still, and S. C. Douglas, “Duality between widely linear and dual channel adaptive filtering,” in *Proceedings of 2009 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (Taipei), pp. 1729–1732, 2009.
- [54] J. Vía, L. Vielva, I. Santamaría, and D. Palomar, “Independent component analysis of quaternion Gaussian vectors,” in *Proceedings of 2010 IEEE Workshop on Sensor Array and Multichannel Signal (SAM)*, pp. 145–148, 2010.

- [55] J. D. J. López, R. M. Fernández-Alcalá, J. Navarro-Moreno, and J. C. Ruiz-Molina, “A semi-widely linear filtering algorithm for $\mathbb{C}^n$ -proper quaternion signals based on randomly modeled observations,” in *Proceedings of 17th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, (Edinburgh), pp. 1–5, July 2016.
- [56] L. De Lathauwer, “A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization,” *SIAM Journal on Matrix Analysis and Applications*, vol. 28, no. 3, pp. 642–666, 2006.
- [57] A. Cichocki, D. P. Mandic, G. Zhou, Q. Zhao, C. Caiafa, H. Phan, and L. De Lathauwer, “Tensor decompositions for signal processing applications: From two-way to multiway component analysis,” *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 145–163, 2015.
- [58] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1990.
- [59] L. De Lathauwer and B. De Moor, “On the blind separation of non-circular sources,” in *Proceedings of the 11th European Signal Processing Conference (EUSIPCO)*, (Toulouse), pp. 1–4, 2002.
- [60] J. Eriksson and V. Koivunen, “Complex random vectors and ICA models: Identifiability, uniqueness, and separability,” *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 1017–1029, 2006.
- [61] E. Ollila and V. Koivunen, “Complex ICA using generalized uncorrelating transform,” *Signal Processing*, vol. 89, no. 4, pp. 365–377, 2009.
- [62] C. Cheong Took, S. Douglas, and D. P. Mandic, “On approximate diagonalization of correlation matrices in widely linear signal processing,” *IEEE Transactions on Signal Processing*, vol. 60, no. 3, pp. 1469–1473, 2012.
- [63] C. Cheong Took, S. Douglas, and D. P. Mandic, “Maintaining the integrity of sources in complex learning systems: Intraference and the correlation

- preserving transform,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 3, pp. 500–509, 2015.
- [64] C. Cheong Took, D. P. Mandic, and F. Zhang, “On the unitary diagonalisation of a special class of quaternion matrices,” *Applied Mathematics Letters*, vol. 24, no. 11, pp. 1806–1809, 2011.
- [65] S. Sangwine and N. L. Bihan, “Quaternion singular value decomposition based on bidiagonalization to a real matrix using quaternion Householder transformations,” *Applied Mathematics and Computation*, vol. 182, pp. 727–738, 2006.
- [66] N. L. Bihan and S. Sangwine, “Jacobi method for quaternion matrix singular value decomposition,” *Applied Mathematics and Computation*, vol. 187, pp. 1265–1271, 2007.
- [67] S. Enshaeifar, C. C. Took, S. Sanei, and D. P. Mandic, “Novel quaternion matrix factorisations,” in *Proceedings of 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (Shanghai), pp. 3946–3950, 2016.
- [68] F. Zhang, *The Schur Complement and Its Applications*. Springer, 2006.
- [69] C. Park, C. Cheong Took, and D. P. Mandic, “Augmented complex common spatial patterns for classification of noncircular EEG from motor imagery tasks,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 1, pp. 1–10, 2014.
- [70] S. Enshaeifar, C. Cheong Took, C. Park, and D. P. Mandic, “Quaternion common spatial patterns,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2016.
- [71] “NIR of corn samples for standardization benchmarking.” <http://www.eigenvector.com/data/Corn/>, June 2005.

- [72] S. P. Talebi and D. P. Mandic, “A quaternion frequency estimator for three-phase power systems,” in *Proceedings of 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (Brisbane), pp. 3956–3960, Apr. 2015.
- [73] Y. Xia, S. C. Douglas, and D. P. Mandic, “Adaptive frequency estimation in smart grid applications: Exploiting noncircularity and widely linear adaptive estimators,” *IEEE Signal Processing Magazine*, vol. 29, no. 5, pp. 44–54, 2012.
- [74] S. Kanna, D. H. Dini, Y. Xia, S. Y. Hui, and D. P. Mandic, “Distributed widely linear kalman filtering for frequency estimation in power networks,” *IEEE Transactions on Signal and Information Processing over Networks*, vol. 1, pp. 45–57, March 2015.
- [75] C. Jahanchahi, S. Kanna, and D. P. Mandic, “A theoretical bound for the performance advantage of quaternion widely linear estimation,” in *Proceedings of the 19th International Conference on Digital Signal Processing (DSP)*, (Hongkong), pp. 254–258, 2014.
- [76] B. Picinbono and P. Chevalier, “Widely linear estimation with complex data,” *IEEE Transactions on Signal Processing*, vol. 43, no. 8, pp. 2030–2033, 1995.
- [77] R. Sznajder, M. S. Gowda, and M. M. Moldovan, “More results on Schur complements in Euclidean Jordan algebras,” *Journal of Global Optimization*, vol. 53, no. 1, pp. 121–134, 2012.
- [78] D. Schulz and R. S. Thomä, “Using quaternion-valued linear algebra,” *arXiv preprint arXiv:1311.7488*, 2013.
- [79] C. Jahanchahi, C. Cheong Took, and D. P. Mandic, “On HR calculus, quaternion valued stochastic gradient, and adaptive three dimensional wind fo-

- recasting,” in *Proceedings of International Joint Conference on Neural Networks (IJCNN)*, (Barcelona), pp. 1–5, 2010.
- [80] S. C. Douglas and D. P. Mandic, “Performance analysis of the conventional complex LMS and augmented complex LMS algorithms,” in *Proceedings of 2010 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (Dallas), pp. 3794–3797, 2010.
- [81] T. Y. Al-Naffouri and A. H. Sayed, “Transient analysis of data-normalized adaptive filters,” *IEEE Transactions on Signal Processing*, vol. 51, no. 3, pp. 639–652, 2003.
- [82] N. R. Yousef and A. H. Sayed, “A unified approach to the steady-state and tracking analyses of adaptive filters,” *IEEE Transactions on Signal Processing*, vol. 49, no. 2, pp. 314–324, 2001.
- [83] C. Cheong Took, C. Jahanchahi, and D. P. Mandic, “A unifying framework for the analysis of quaternion valued adaptive filters,” in *Proceedings of the 45th Asilomar Conference on Signals, Systems and Computers*, (Pacific Grove), pp. 1771–1774, 2011.
- [84] A. H. Sayed, *Fundamentals of Adaptive Filtering*. Hoboken, NJ: Wiley-IEEE Press, 2003.
- [85] N. J. Bershad, E. Eweda, and J. C. M. Bermudez, “Stochastic analysis of the LMS and NLMS algorithms for cyclostationary white Gaussian inputs,” *IEEE Transactions on Signal Processing*, vol. 62, no. 9, pp. 2238–2249, 2014.
- [86] B. Widrow, J. M. McCool, M. G. Larimore, and C. R. Johnson, “Stationary and nonstationary learning characteristics of the LMS adaptive filter,” *Proceedings of the IEEE*, vol. 64, no. 8, pp. 1151–1162, 1976.
- [87] Y. Tian and G. P. Styan, “Some inequalities for sums of nonnegative definite matrices in quaternions,” *Journal of Inequalities and Applications*, vol. 2005, no. 5, pp. 449–458, 2005.

- [88] L. Isserlis, “On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables,” *Biometrika*, vol. 12, no. 1/2, pp. 134–139, 1918.
- [89] P. Chevalier and F. Pipon, “New insights into optimal widely linear array receivers for the demodulation of BPSK, MSK, and GMSK signals corrupted by noncircular interferences-application to SAIC,” *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 870–883, 2006.
- [90] A. Tarighat and A. H. Sayed, “Least mean-phase adaptive filters with application to communications systems,” *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 220–223, 2004.
- [91] M. Rupp, “On the separation of channel and frequency offset estimation,” in *Proceedings of the 32nd Asilomar Conference on Signals, Systems and Computers*, vol. 2, (Pacific Grove), pp. 1186–1190, 1998.
- [92] S. C. Douglas and D. P. Mandic, “The least-mean-magnitude-phase algorithm with applications to communications systems,” in *Proceedings of 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (Prague), pp. 4152–4155, May 2011.
- [93] S. C. Douglas and D. P. Mandic, “Stability analysis of the least-mean-magnitude-phase algorithm,” in *Proceedings of 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (Shanghai), pp. 4935–4939, Mar. 2016.
- [94] S. P. Talebi, D. Xu, A. Kuh, and D. P. Mandic, “A quaternion least mean phase adaptive estimator,” in *Proceedings of 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (Florence), pp. 6419–6423, May 2014.

- [95] H. Zhang and D. P. Mandic, “Is a complex-valued stepsize advantageous in complex-valued gradient learning algorithms?,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. PP, no. 99, pp. 1–6, 2015.
- [96] D. Nion and L. D. Lathauwer, “An enhanced line search scheme for complex-valued tensor decompositions. Application in DS-CDMA,” *Signal Processing*, vol. 88, no. 3, pp. 749 – 755, 2008.
- [97] Y. Chen, D. Han, and L. Qi, “New ALS methods with extrapolating search directions and optimal step size for complex-valued tensor decompositions,” *IEEE Transactions on Signal Processing*, vol. 59, no. 12, pp. 5888–5898, 2011.
- [98] J. Navarro-Moreno, R. M. Fernández-Alcalá, and J. C. Ruiz-Molina, “A quaternion widely linear model for nonlinear Gaussian estimation,” *IEEE Transactions on Signal Processing*, vol. 62, no. 24, pp. 6414–6424, 2014.
- [99] F. G. A. Neto and V. H. Nascimento, “A novel reduced-complexity widely linear QLMS algorithm,” in *Proceedings of 2011 IEEE Statistical Signal Processing Workshop (SSP)*, (Nice), pp. 81–84, 2011.
- [100] C. Cheong Took, D. P. Mandic, and J. Benesty, “Study of the quaternion LMS and four-channel LMS algorithms,” in *Proceedings of 2009 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (Taipei), pp. 3109–3112, 2009.
- [101] C. Jahanchahi, S. Kanna, and D. P. Mandic, “Complex dual channel estimation: Cost effective widely linear adaptive filtering,” *Signal Processing*, vol. 104, pp. 33–42, 2014.
- [102] C. Jahanchahi, C. Cheong Took, and D. P. Mandic, “The widely linear quaternion recursive least squares filter,” in *Proceedings of the 2nd International Workshop on Cognitive Information Processing (CIP)*, (Elba), pp. 87–92, 2010.

- [103] B. Che Ujang, C. Cheong Took, and D. P. Mandic, “Quaternion-valued non-linear adaptive filtering,” *IEEE Transactions on Neural Networks*, vol. 22, no. 8, pp. 1193–1206, 2011.
- [104] P. O. Amblard and N. Le Bihan, “On properness of quaternion valued random variables,” in *6th IMA Conference on Mathematics in Signal Processing*, pp. 23–26, 2004.
- [105] A. Cariow and G. Cariowa, “An unified approach for developing rationalized algorithms for hypercomplex number multiplication,” *Electric Review*, vol. 91, no. 2, pp. 36–39, 2015.
- [106] J. Barzilai and J. M. Borwein, “Two-point step size gradient methods,” *IMA Journal of Numerical Analysis*, vol. 8, no. 1, pp. 141–148, 1988.
- [107] K. Mitsubori and T. Saito, “Torus doubling and hyperchaos in a five dimensional hysteresis circuit,” in *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS)*, vol. 6, pp. 113–116, 1994.
- [108] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software*. Hoboken, NJ: Wiley-Blackwell, 2001.
- [109] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton, “Adaptive mixtures of local experts,” *Neural Computation*, vol. 3, no. 1, pp. 79–87, 1991.
- [110] S. E. Yuksel, J. N. Wilson, and P. D. Gader, “Twenty years of mixture of experts,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 8, pp. 1177–1193, 2012.
- [111] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice Hall, 1979.

- [112] M. Gwak, K. Jo, and M. Sunwoo, “Neural-network multiple models filter (NMM)-based position estimation system for autonomous vehicles,” *International Journal of Automotive Technology*, vol. 14, no. 2, pp. 265–274, 2013.
- [113] L. Chen and K. S. Narendra, “Nonlinear adaptive control using neural networks and multiple models,” *Automatica*, vol. 37, no. 8, pp. 1245–1255, 2001.
- [114] L. Chen and K. S. Narendra, “Intelligent control using multiple neural networks,” *International Journal of Adaptive Control and Signal Processing*, vol. 17, no. 6, pp. 417–430, 2003.
- [115] Y. Fu and T. Chai, “Nonlinear multivariable adaptive control using multiple models and neural networks,” *Automatica*, vol. 43, no. 6, pp. 1101–1110, 2007.
- [116] A. Mohammadi and K. N. Plataniotis, “Improper complex-valued multiple-model adaptive estimation,” *IEEE Transactions on Signal Processing*, vol. 63, no. 6, pp. 1528–1542, 2015.
- [117] X. R. Li and Y. Bar-Shalom, “Performance prediction of the interacting multiple model algorithm,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 29, no. 3, pp. 755–771, 1993.
- [118] C. E. Seah and I. Hwang, “Algorithm for performance analysis of the IMM algorithm,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, pp. 1114–1124, Apr. 2011.
- [119] R. W. Osborne and W. D. Blair, “Update to the hybrid conditional averaging performance prediction of the IMM algorithm,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, no. 4, pp. 2967–2974, 2011.
- [120] C. Hue, J.-P. Le Cadre, and P. Perez, “Posterior Cramer-Rao bounds for multi-target tracking,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 42, no. 1, pp. 37–49, 2006.

- [121] H. A. Blom and Y. Bar-Shalom, “The interacting multiple model algorithm for systems with Markovian switching coefficients,” *IEEE Transactions on Automatic Control*, vol. 33, no. 8, pp. 780–783, 1988.
- [122] I. Hwang, C. E. Seah, and S. Lee, “A study on stability of the interacting multiple model algorithm,” *IEEE Transactions on Automatic Control*, vol. 62, pp. 901–906, Feb. 2017.
- [123] P. D. Hanlon and P. S. Maybeck, “Characterization of Kalman filter residuals in the presence of mismodeling,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 36, no. 1, pp. 114–131, 2000.
- [124] B. Friedland, “On the properties of reduced order Kalman filters,” *IEEE Transactions on Automatic Control*, vol. 34, no. 3, pp. 321–324, 1989.
- [125] J. Navarro-Moreno, R. Fernandez-Alcala, and J. Ruiz-Molina, “A quaternion widely linear series expansion and its applications,” *IEEE Signal Processing Letters*, vol. 19, no. 12, pp. 868–871, 2012.
- [126] F. A. Tobar, S.-Y. Kung, and D. P. Mandic, “Multikernel least mean square algorithm,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 2, pp. 265–277, 2014.
- [127] A. Mohammadi and K. N. Plataniotis, “Complex-valued Gaussian sum filter for nonlinear filtering of non-Gaussian/non-circular noise,” *IEEE Signal Processing Letters*, vol. 22, no. 4, pp. 440–444, 2015.