Quaternion Matrices: Statistical Properties and Applications to Signal Processing and Wavelets

A thesis presented for the degree of Doctor of Philosophy of Imperial College and the Diploma of Imperial College by

Paul Ginzberg

Department of Mathematics Imperial College 180 Queen's Gate, London SW7 2BZ

NOVEMBER 28, 2013

I certify that this thesis, and the research to which it refers, are the product of my own work, and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline.

Signed:

Copyright

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



Acknowledgements

First and foremost, I would like to thank my advisor Andrew Walden, whose serene optimism and regular guidance have led me to this point.

This work would not have been possible without the financial generosity of the EPSRC, and by extension of the UK taxpayer.

The transfer committee members Alastair Young and Ajay Jasra and viva examiners Alastair Young (again) and Stephen Sangwine deserve thanks for their time, comments and encouragement.

I credit John Gibbons, Chris Sisson and Rusudan Svanidze for the lack of administrative delays or hurdles; and Nigel Lawrence et al. for the ICMathsThesis LATEX document class.

For bringing pleasant lightheartedness to what would have otherwise been an unbearably solitary pursuit, I would like to thank the Huxley 526 crew: Orlando Doehrig, Christopher Minas, James Martin, Swati Chandna, Anna Fowler, Dean Bodenham, Georg Hahn, Zhana Kuncheva, Diletta Martinelli and Ricardo Monti; along with fellow PhD students Ed Cohen, Elly Ehrlich, Lewis Evans, Din Lau, Aidan O'Sullivan and Adam Persing; and also for liberally dispensing their wisdom, the staff members of the Statistics section, especially Niall Adams, Axel Gandy, Nick Heard, Emma McCoy and Giovanni Montana.

I thank my girlfriend Claire Hood for her support in everything off-campus.

This thesis is dedicated to my family.

Abstract

Similarly to how complex numbers provide a possible framework for extending scalar signal processing techniques to 2-channel signals, the 4-dimensional hypercomplex algebra of quaternions can be used to represent signals with 3 or 4 components.

For a quaternion random vector to be suited for quaternion linear processing, it must be (second-order) proper. We consider the likelihood ratio test (LRT) for propriety, and compute the exact distribution for statistics of Box type, which include this LRT. Various approximate distributions are compared. The Wishart distribution of a quaternion sample covariance matrix is derived from first principles.

Quaternions are isomorphic to an algebra of structured 4×4 real matrices. This mapping is our main tool, and suggests considering more general real matrix problems as a way of investigating quaternion linear algorithms.

A quaternion vector autoregressive (VAR) time-series model is equivalent to a structured real VAR model. We show that generalised least squares (and Gaussian maximum likelihood) estimation of the parameters reduces to ordinary least squares, but only if the innovations are proper. A LRT is suggested to simultaneously test for quaternion structure in the regression coefficients and innovation covariance.

Matrix-valued wavelets (MVWs) are generalised (multi)wavelets for vector-valued signals. Quaternion wavelets are equivalent to structured MVWs. Taking into account orthogonal similarity, all MVWs can be constructed from non-trivial MVWs. We show that there are no non-scalar non-trivial MVWs with short support [0,3]. Through symbolic computation we construct the families of shortest non-trivial 2×2 Daubechies MVWs and quaternion Daubechies wavelets.

Table of contents

A	bstra	\mathbf{ct}		5
\mathbf{Li}	st of	Figur	es	9
\mathbf{Li}	st of	Table	s	10
\mathbf{Li}	st of	Publi	cations	11
In	trod	uction		13
1	Qua	aternio	on Linear Algebra	18
	1.1	Introd	luction	18
	1.2	Quate	rnions	21
		1.2.1	An algebraic introduction	21
		1.2.2	Algebraic significance	24
		1.2.3	Matrix representation	26
	1.3	Quate	rnion matrices	27
		1.3.1	Representation as real matrices	27
		1.3.2	Left-linear quaternion matrix multiplication	31
		1.3.3	The matrix product as a projection and ensemble	33
		1.3.4	Determinant, trace and norm	36
		1.3.5	Special matrices and decompositions	37
2	Qua	aternio	n Probability Distributions	41
	2.1	Introd	luction	41
	2.2	Chara	cteristic functions	43
	2.3	(Proper) normal distribution		
	2.4	Wisha	rt distribution	48
		2.4.1	A review of literature related to the quaternion Wishart distri-	
			bution	51
	2.5	Impro	per normal distribution	53
	2.6	Chara	cterisations of propriety and second-order propriety	58

3	The	Quate	ernion Vector Autoregressive Model	61			
	3.1	Introdu	iction	61			
	3.2	Quater	nion multivariate linear regression	64			
	3.3	Quater	nion VAR as a structured real VAR	70			
		3.3.1	Quaternion VAR parameter estimation	72			
	0.4	3.3.2	Numerical evaluation	75			
	3.4	Widely	<i>r</i> -linear quaternion VAR as a real VAR	80			
	3.5	Testing	g for VAR propriety	81			
4	Likelihood Ratio Testing for Quaternion-Structured Covariance Ma-						
		es		85			
	4.1	Introdu		85			
	4.2	The LI	1 for quaternion propriety	87			
		4.2.1	Maximum likelihood estimators of covariance	88			
	4.0	4.2.2	The LRT statistic and its moments	90			
	4.3	The di	stribution of statistics of Box type	92			
		4.3.1	Exact distribution	95			
		4.3.2	Approximations	101			
		4.3.3	Numerical comparison of approximations	112			
5	Qua	ternio	n Wavelets and Matrix-Valued Wavelets	120			
	5.1	Introdu	action	120			
	5.2	A revie	ew of literature on quaternion wavelet transforms	124			
		5.2.1	Different types of quaternion wavelet transform	124			
		5.2.2	Problems with existing quaternion wavelet constructions	127			
	5.3	Matrix	and vector multiresolution analyses	132			
	5.4	Matrix	-valued scaling filters	137			
		5.4.1	Orthogonality	139			
		5.4.2	Vanishing moments	141			
		5.4.3	The fast matrix-valued wavelet transform	142			
		5.4.4	Computing matrix-valued wavelet filters	143			
	5.5	Trivial	matrix-valued scaling filters	145			
		5.5.1	Orthogonal similarity	145			
		5.5.2	Decomposition of filters	146			
		5.5.3	Computational complexity	148			
		5.5.4	Triviality of MVSFs of length $L \leq 3$	148			
	5.6	Daube	chies matrix-valued scaling filters	150			
		5.6.1	Triviality of Daubechies MVSFs of length $L \leq 4$	153			
	5.7	Matrix	representation of quaternion and algebra-valued wavelets	155			
		5.7.1	Quaternion propriety	158			
		5.7.2	Orthogonal similarity for quaternions	159			
		5.7.3	The biquaternion Fourier transform	160			

7

	5.8 5.9	Examples of non-trivial Daubechies MVSFs \ldots \ldots 5.8.1The 2 × 2 Daubechies MVSFs of length $L = 6$ \ldots 5.8.2The quaternion Daubechies MVSFs of length $L = 10$ \ldots On the use of MVWs in practice \ldots \ldots	161 162 165 170
Co	onclu	sion	173
Re	eferei	nces	176
A	Add A.1 A.2 A.3	itional ResultsA note on rotation invarianceAdditional results on random variables of Box typeA note on the matrix Karhunen-Loève transform	193 193 194 196
в	Pro B.1 B.2 B.3 B.4 B.5 B.6	ofs Proof of Theorem 2.15 Proof of Remark 3.2 Proof of Proposition 4.8 Proof of Theorem 4.33 Proof of Proposition 5.3 Proof of Proposition 5.18	198 198 204 205 206 207 207
С	Con C.1 C.2	Apputer Code Matlab code for wavelet filter computation Maple code for the design of scaling filters	210 210 212
D	Peri	nission to use IEEE Copyrighted material	221
\mathbf{E}	Ada Vect	ptive Orthogonal Matrix-Valued Wavelets and Compression o cor-Valued signals	f 224

8

List of Figures

57	Commutative diagram containing the structured real and augmented quaternion approaches to real/complex/quaternion linear algebra	2.1
77 79	Error in the estimation of \widetilde{a}_1 for varying sample length $N \ldots \ldots$ Error in the estimation of \widetilde{a} for $N = 100$ and varying degrees of impropriety	$3.1 \\ 3.2$
	Relative errors of approximate CDFs at the 95 th percentile for varying	4.1
115	$p \text{ and } N = 4p \dots$ Relative errors of approximate CDFs at the 95 th percentile for varying	4.2
116	p and $N = 5p$	4.3
116	$p \text{ and } N = 8p. \dots$ Relative errors of approximate CDFs at the 95 th percentile for $p = 6$	4.4
117	and varying N	4.5
119	x , or equivalently varying percentiles $\dots \dots \dots$	4.0
$\overline{5}.166$	Absolute entries of the frequency response $\hat{\boldsymbol{G}}(f)$ (and $\hat{\boldsymbol{H}}(f)$) for the 2×2 Daubechies MVSF of length $L = 6$ with parameter choice $x = \sqrt{2}$	5.1
166	Scaling function $\Phi(t)$ for the 2 × 2 Daubechies MVSF of length $L = 6$ with parameter choice $x = \sqrt{5}$.	5.2
167	Wavelet $\Psi(t)$ for the 2 × 2 Daubechies MVSF of length $L = 6$ with parameter choice $x = \sqrt{5}$.	5.3
	Absolute entries of the frequency response $\hat{G}(f)$ (and $\hat{H}(f)$) for the quaternion Daubechies scaling (and wavelet) filter of length $L = 10$	5.4
170	with parameter choice $x = \pi/2$	5.5
171	with parameter $x = \pi/2$	0.0

List of Tables

4.1	Legend	115
4.2	Approximate rejection probabilities (in %) for the the 1% level critical region.	118
5.1	Number of operations required for a $n \times n$ matrix wavelet transform when the scaling and wavelet filters are diagonal, diagonal up to or- thogonal similarity (highly trivial), block-diagonal up to orthogonal similarity (trivial) or non-trivial	149

List of Publications

P. Ginzberg and A. T. Walden. Testing for quaternion propriety. *IEEE Transactions on Signal Processing*, 59(7):3025–3034, 2011.

In this paper we consider the problem of testing whether a multivariate quaternion normal distribution is proper, from a set of independent samples. The likelihood ratio test is given, and the exact distribution of the test statistic under the null hypothesis of propriety is derived. As this is in terms of Meijer's G-function, various approximation methods are compared. Based on the Pearson system of curves, we suggest an improved high-accuracy F approximation. Chapter 4 of this thesis includes these results, and extends them by considering general test statistics of Box type and additional approximations. ©IEEE.

P. Ginzberg and A. T. Walden. Quaternion VAR modelling and estimation. IEEE Transactions on Signal Processing, 61(1):154–158, 2013b.

This short paper demonstrates how a quaternion vector autoregression (VAR) can be treated as a special case of structured real VAR. We show that generalised least squares and (Gaussian) maximum likelihood estimation of the model regression parameters reduce to simple least squares estimation if the innovations are (secondorder) quaternion proper. Chapter 3 of this thesis shows that this simplification applies more generally to quaternion multivariate linear regression and discusses some consequences of the real VAR interpretation. ©IEEE. P. Ginzberg and A. T. Walden. Matrix-valued and quaternion wavelets. *IEEE Trans*actions on Signal Processing, 61(6):1357–1367, 2013a.

We compare the matrix-valued wavelet (MVW) based multiresolution analysis of matrix-valued and (more appropriate) vector-valued signals. We construct a novel family of non-trivial orthogonal 2×2 (resp. 4×4) MVWs having 3 (resp. 5) vanishing moments. These can be considered generalisations of the real and complex Daubechies wavelets, and the latter construction represents a symmetric quaternion wavelet. Some useful uniqueness and non-existence results for scaling filters with certain lengths and numbers of vanishing moments are proved. This material is presented in Chapter 5 of this thesis. ©IEEE.

Material from the three papers listed above is included in this thesis with permission of IEEE (see Appendix D).

P. Ginzberg and A. T. Walden. Adaptive orthogonal matrix-valued wavelets and compression of vector-valued signals. In *Proceedings of the 9th IMA International Conference on Mathematics in Signal Processing*, Birmingham, UK, 2012.

We parameterise the set of all 3×3 Daubechies matrix-valued scaling filters (MVSFs) of length L = 6. We note that for each MVSF, the corresponding wavelet filter (or the computed wavelet coefficients) can be rotated arbitrarily. We show how a modified SIMPLIMAX algorithm can be used to adaptively optimise this choice of rotation. All parameters are jointly optimised numerically for compressing the standard colour test image *Lena*. We show that this approach fails to outperform an optimally rotated naive Daubechies wavelet. However, optimisation of the free rotation for the naive Daubechies wavelet — which can be interpreted as adaptive colourspace selection — can on its own decrease the root mean squared error of compressed images by over 20%. Because these conference proceedings are hard to obtain, we attach a copy of this paper in Appendix E.

Introduction

The quaternions $\mathbb{H} = \{a + bi + cj + dk : a, b, c, d \in \mathbb{R}\}$ are a four dimensional generalisation of the two dimensional complex algebra $\mathbb{C} = \{a + bi : a, b \in \mathbb{R}\}$. Similarly to how complex numbers can describe both points and linear operations in the plane, quaternions can describe both points and linear operations in three or four dimensions.¹ Historically, the development of quaternions runs parallel to the development of real linear algebra and matrix theory. Thus they provided a framework for dealing with vector quantities before the widespread popularisation of matrices and vector calculus in mathematics and physics.

Since then, quaternions have continued to be studied in detail, and have inspired the development of more general 'hypercomplex' geometric algebras, such as Clifford algebras. In practical applications quaternions are most commonly used to represent 3D rotations or orientations.

More recently, the use of quaternions as a way of expressing and manipulating 3and 4-dimensional quantities has seen a resurgence. Examples of intrinsically vectorvalued signals — such as those collected by vector sensors — which have been treated as quaternion-valued include those from 3D anemometers (Cheong Took and Mandic, 2009), 3D geophones (Grandi et al., 2007; Sajeva, 2009), EEG (Javidi et al., 2011), gyroscopes (Jahanchahi et al., 2013), colour images (Sangwine and Ell, 2000) and multispectral images (Xu et al., 2012).

Various common signal processing and image processing algorithms have been

¹The set of all possible linear operations in the plane can be generated from addition, complex multiplication and complex conjugation. In the case of quaternions, addition and quaternion multiplication from both the right and the left are sufficient.

generalised to work in the quaternion domain. These include the singular value decomposition (SVD) (Sangwine and Le Bihan, 2006) which can be used for blind source separation of polarised waves (Le Bihan and Mars, 2004), e.g. Rayleigh wave extraction (Sajeva, 2009), and for video quality assessment (Zhang et al., 2009). This has been extended to the quaternion polynomial SVD for convolutive mixtures (Menanno and Le Bihan, 2010), and to quaternion MUSIC (Miron et al., 2006), which estimates the direction of propagation and polarisation of the sources. The quaternion eigenvalue decomposition (the quaternion SVD of a Hermitian matrix) gives us quaternion principal component analysis (Sangwine and Ell, 2000; Xu et al., 2012), which provides low-rank approximations to quaternion covariance matrices. This is also treated by Vía et al. (2010a), along with quaternion versions of multivariate linear regression, canonical correlation analysis and partial correlation analysis in a unified approach.

For quaternion-linear and real-linear modelling (and prediction and filtering) of quaternion-valued time-series, various algorithms have been adapted, such as Yule-Walker vector autoregressive modelling (a.k.a. Wiener filtering) (Navarro-Moreno et al., 2013), recursive least squares (Jahanchahi et al., 2010), least means squared (stochastic gradient descent) (Cheong Took and Mandic, 2009, 2010a,b), and affine projection (Jahanchahi et al., 2013). We will take a step back from these approaches — which are mostly adaptive and online — to consider the underlying basics of quaternion VAR modelling and least squares parameter estimation in Chapter 3.

Other recent applications of quaternion signal processing include seismic velocity analysis (Grandi et al., 2007), seismic waveform deconvolution (Menanno, 2010) and block coding for wireless communications (Seberry et al., 2008; Wysocki et al., 2009).

This thesis hopes to provide a rigorous foundation for quaternion-based statistical signal processing by clarifying its relationship to standard real statistical signal processing and collecting useful results on quaternion linear algebra and related probability distributions. We then consider in detail two important signal processing tools: vector autoregressive (VAR) time-series modelling and wavelet transforms. VAR is the fundamental model for interacting short-memory stationary time-series. We will focus particularly on orthogonal discrete wavelet transforms, which generate a sparse representation for piecewise smooth signals, and can be computed in linear time through filtering and down-sampling.

As a vector space, \mathbb{H} is isomorphic to \mathbb{R}^4 . If we wish to also preserve the multiplicative structure, \mathbb{H} can be represented by real quaternion-structured matrices of the form

$$\left(\begin{array}{rrrrr} a & -b & -c & -d \\ b & a & -d & c \\ c & d & a & -b \\ d & -c & b & a \end{array}\right).$$

This representation is a *-algebra homomorphism of the quaternion algebra into the matrix algebra $\mathbb{R}^{4\times4}$. It allows us to map problems from the quaternion domain (or quaternion matrix domain) to the more familiar real matrix domain, where we can exploit the extensive machinery of real linear algebra and real multichannel signal processing. This allows for simple or even trivial proofs for some of the questions that arise when working in the quaternion linear setting.

We wish to develop a statistical theory of quaternion-valued random variables. One way to do this is to simply note that \mathbb{H}^n is a vector space isomorphic to \mathbb{R}^{4n} and use the usual real theory on \mathbb{R}^{4n} . This is the 'improper' approach which ignores the multiplicative structure of quaternions. The natural 'proper' extension of second-order statistical theory to quaternions requires that we restrict ourselves to real covariance matrices which have quaternion structure. This restriction then allows for algorithms based on quaternion linear transformations rather than real linear transformations (Vía et al., 2010a). Real linear transformations can still be expressed in the quaternion domain, where they are called widely-linear transformations. However, we argue that doing so often complicates matters unnecessarily.

Although quaternion signal processing has mostly been developed as a generalisation of scalar real signal processing, with the ability to process vector-valued signals whilst making only minor adjustments to the underlying algorithms; we believe that wherever possible a more informative approach is achieved by viewing quaternion signal processing as a special case of real vector signal processing with structural assumptions. This wider context clarifies the implicit assumptions and restrictions of the quaternion domain, and the possible benefits. For example, this allows us to show

INTRODUCTION

that for quaternion VAR parameter estimation (and more general multivariate linear regression) optimality of the least squares solution requires a propriety assumption on the errors, in the absence of which the generalised least squares (maximum like-lihood) solution offers better results. This philosophy is obviously harder to apply in areas where the corresponding vector signal processing approach is not sufficiently well understood (which may be the original motivator behind the use of quaternions). Even then, it may lead to interesting questions about the general vector case. For example, in our work on quaternion wavelets, we were drawn to prove various results concerning the more general matrix-valued wavelets.

It is worth noting that the methods we have developed here can be generalised to algebras other than the quaternions, e.g. Clifford algebras. This can either be done by adapting them directly to the relevant structured real matrix representation, or by decomposing the algebra into a direct sum of unstructured real, complex and quaternion matrix algebras.²

This thesis is organised as follows.

In **Chapter 1** we collect standard results on quaternions and quaternion linear algebra which will be needed for later chapters, allowing this thesis to be mostly selfcontained. The vector space isomorphism and algebra isomorphism are introduced and their properties examined. The relationship between quaternion left-linearity and right-linearity is explained. We also note that every semi-simple finite-dimensional real algebra can be constructed as a product of real, complex and quaternion matrix algebras. The only original result in this chapter is that general quaternion multiplication can be interpreted in the real domain in terms of ensemble averaging and in terms of an orthogonal projection imposing quaternion structure.

In **Chapter 2** we define the proper quaternion normal distribution (resolving some inconsistencies in the literature) and point out its fundamental relationship with quaternion linearity. The improper quaternion normal approach is also examined. Interpreting the quaternion sample covariance matrix in terms of an orthogonal projection allows for a simple derivation of the quaternion Wishart characteristic

 $^{^{2}}$ The latter approach can be applied to any semi-simple real algebra (see Section 1.2.2).

function. We also give a novel derivation of the quaternion Wishart density (See Appendix B.1).

In **Chapter 3** we define the (proper) quaternion VAR time-series model. We show that quaternion VAR modelling is a type of structured real VAR modelling. We prove that, for a quaternion general (left-)linear model with uncorrelated (right-)proper vector errors, least squares and generalised least squares estimation are equivalent. As a particular case of this new result, generalised least squares (and maximum likelihood) estimation of the parameters of a quaternion VAR model reduces to least squares estimation. The likelihood ratio test (LRT) for propriety of a VAR time-series is given. This chapter is an extension of the author's paper Ginzberg and Walden (2013b).

Many likelihood ratio test statistics are of Box type, including the LRT for quaternion propriety of a multivariate normal sample. In **Chapter 4** we find the exact density (PDF) and distribution function (CDF) for an arbitrary random variable of Box type. Using the LRT for quaternion propriety (which we re-derive using the orthogonal projection interpretation) as an example, we compare a wide range of approximations which have been suggested for this distribution. A new F approximation is also considered. This chapter is largely based on the author's paper Ginzberg and Walden (2011).

We show in **Chapter 5** that previous examples of discrete quaternion wavelets in the literature are either incorrect or trivial. Using the real matrix representation, we note that quaternion wavelets are simply matrix-valued wavelets (MVWs) with quaternion structure. The MVW transform treats a vector-valued signal holistically, as opposed to independent scalar wavelet transforms of the components. We prove some non-existence results for short non-trivial orthogonal MVWs, and by solving a set of quadratic design equations symbolically through Gröbner bases, give the first example of (a family of) non-trivial Daubechies MVWs. We also construct the (family of) shortest non-trivial quaternion Daubechies wavelets. This chapter is largely based on the author's paper Ginzberg and Walden (2013a).

Chapter 1

Quaternion Linear Algebra

1.1 Introduction

Quaternions were invented by William R. Hamilton in 1843 as a four dimensional generalisation of complex numbers (Hamilton, 1866). They have since seen a variety of uses, most notably to represent and manipulate 3D rotations and orientations in engineering (Crassidis et al., 2007) and computer graphics (Shoemake, 1985), where they avoid the gimbal lock problem of Euler angles, and the high redundancy of 3×3 special orthogonal matrices.

In signal processing, algorithms based on quaternions can be used to deal with 2, 3 or 4-channel data arising from vector-sensors. Although quaternions are non-commutative, extending mathematical methods based on complex (or real) numbers to quaternions can often be done with few (or no) adjustments.

Non-linear quaternion methods, such as quaternion neural networks (Buchholz and Le Bihan, 2006), have been considered. However, signal processing algorithms often boil down to a particular application of a linear algebra algorithm, such as the SVD, linear equation solving or change of basis. Replacing real matrices with quaternion matrices in these methods allows for conceptually simple joint processing of 3 or 4 component signals, but also introduces a restriction to quaternion-linearity. This in turn introduces the implicit symmetry assumption of quaternion propriety for the data, which will be discussed in Chapter 2. As a preliminary to the study of quaternion signal processing, this chapter collects important known results on quaternion linear algebra, which will be used in later chapters. For the most part, the properties of complex and quaternion matrices are the same. We point out the occasional differences, which require special attention. For example, conjugation is an automorphism for \mathbb{C} , but not for \mathbb{H} .

The quaternion linear structure appears to be rare in practical applications,¹ however we note that it is a crucial building block that allows for generalisation to a wide range of algebraic structures. As we note in Section 1.2.2, the importance of studying the algebra of quaternions comes in part from Frobenius's theorem, which states that \mathbb{R} , \mathbb{C} and \mathbb{H} are the only (finite-dimensional) real algebras in which every non-zero element has an inverse. When combined with the Artin-Wedderburn theorem, this implies that every finite-dimensional semi-simple algebra can be written as the direct sum of matrix algebras with real, complex or quaternion entries. This is true in particular for hypercomplex Clifford algebras. Thus, results which can be proved for real complex and quaternion matrices can be immediately generalised to matrices with entries in a Clifford algebra.

Although not as ubiquitous as complex numbers, Clifford algebras appear in physics. Examples include the algebra of physical space $\mathcal{C}\ell_{3,0}(\mathbb{R})$ used in classical and relativistic physics and — in the form of Pauli spin matrices — in quantum methanics,² and the Minkowski space-time algebra $\mathcal{C}\ell_{1,3}(\mathbb{R})$ used in special relativity³ (Baylis, 2004). For a recent review of Clifford algebra applications see Hitzer et al. (2013), which includes uses of the conformal geometric algebra $\mathcal{C}\ell_{4,1}(\mathbb{R})^4$ and applications in image analysis.

The key results of this chapter, which we use extensively throughout the rest of

¹With the exception of polarised waves, which are intrinsically two dimensional and thus satisfy a stronger structure; and problems involving unit orientation/rotation quaternions, which are not closed under addition.

 $^{{}^{2}\}mathcal{C}\ell_{3,0}(\mathbb{R})$ is isomorphic to the algebra of biquaternions (quaternions with complex coefficients) and to the matrix algebra $\mathbb{C}^{2\times 2}$.

 $^{{}^{3}\}mathcal{C}\ell_{1,3}(\mathbb{R})$ is isomorphic to the quaternion matrix algebra $\mathbb{H}^{2\times 2}$

 $^{{}^{4}\}mathcal{C}\ell_{4,1}(\mathbb{R})$ is isomorphic to the complex matrix algebra $\mathbb{C}^{4\times 4}$

this thesis, are Theorem 1.26, which shows that the structured real representation of quaternion matrices preserves the vector space, multiplicative and involutive structure; and Remark 1.33 which shows that when viewed as a (real-)linear operator on \mathbb{R}^{4n} , the real matrix representation is equivalent to the quaternion matrix viewed as a linear operator on \mathbb{H}^n .

A similar structured real representation exists for complex numbers. Unlike quaternions, complex numbers are commutative. Thus, all algebraic manipulations of equalities remain valid when we change the domain of the variables from \mathbb{R} to \mathbb{C} . This will often make extending real methods to the complex domain seamless. Because matrix multiplication is not commutative, treating complex numbers as structured real matrices obscures this critical property and can be counterproductive. When dealing with quaternions, this downside of the matrix representation is not present, whilst the non-commutativity simultaneously makes it harder to intuitively and seamlessly replace the real domain with the quaternion domain directly. This makes the general-purpose use of real representation techniques particularly attractive in quaternion signal processing. Obviously, quaternion-domain thinking can still be simpler at times (e.g. when interpreting quaternions as rotations), and the isomorphism between quaternions and quaternion-structured real matrices allows for changing between approaches.

In Section 1.3.3 we show that in general quaternion matrix multiplication can be viewed as an ensemble average or an orthogonal projection of (unstructured) real matrix products. This suggests two possible general methods for both interpreting and implementing quaternion linear algorithms based on their real equivalent. Although special cases of the result are implicitly key to existing proofs (e.g. Andersson et al. (1983, Theorem 3)), we have not seen this general insight expressed in the literature.

1.2 Quaternions

1.2.1 An algebraic introduction

In this section we give essential definitions and properties of quaternions. All proofs are straightforward and will be omitted.

Definition 1.1. A real algebra \mathbb{A} is a vector space over \mathbb{R} with a multiplication satisfying $\forall x, y, z \in \mathbb{A}, \ \forall a, b \in \mathbb{R}$

$$\begin{aligned} x(y+z) &= xy + xz, \\ (y+z)x &= yx + zx, \\ (ax)(by) &= (ab)(xy). \end{aligned}$$

Definition 1.2. The quaternions are the four-dimensional real algebra

$$\mathbb{H} = \{a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k} : a, b, c, d \in \mathbb{R}\}.$$

Let $q = a + bi + cj + dk \in \mathbb{H}$, $q_0 = a_0 + b_0i + c_0j + d_0k \in \mathbb{H}$; their product is defined by

$$qq_0 = (aa_0 - bb_0 - cc_0 - dd_0) + (ab_0 + ba_0 + cd_0 - dc_0)i + (ac_0 - bd_0 + ca_0 + db_0)j + (ad_0 + bc_0 - cb_0 + da_0)k.$$
(1.1)

Remark 1.3. The multiplication table for the four basis elements 1, i, j, k is:⁵

	1	i	j	k
1	1	i	j	k
i	i	-1	k	-j
j	j	-k	-1	i
k	k	j	-i	-1

Definition 1.4. Let $q = a + bi + cj + dk \in \mathbb{H}$.

⁵Entry=(basis element of row)·(basis element of column). For example, $i \cdot j = k$

The real and imaginary parts of q are given respectively by

$$\begin{aligned} \Re(q) &= a \\ \Im(q) &= \begin{pmatrix} \Im_{i}(q) \\ \Im_{j}(q) \\ \Im_{k}(q) \end{pmatrix} = \begin{pmatrix} b \\ c \\ d \end{pmatrix}. \end{aligned}$$

q is said to be real iff (if and only if) $\mathfrak{S}(q) = \mathbf{0}$, and (pure) imaginary iff $\mathfrak{R}(q) = 0$. We identify the subalgebra of quaternions which are real with real scalars, so $\mathbb{R} \subset \mathbb{H}$.

Definition 1.5. The conjugate of q is

$$\bar{q} = a - bi - ci - dk.$$

Definition 1.6. The amplitude or norm of q is the euclidean norm on \mathbb{R}^4 , i.e.

$$|q| = \sqrt{q\bar{q}} = \sqrt{a^2 + b^2 + c^2 + d^2}.$$

Definition 1.7. q is said to be a unit quaternion iff |q| = 1.

Remark 1.8. q is a pure imaginary unit quaternion iff $q^2 = -1$

Proposition 1.9. Conjugation $\overline{\bullet} : \mathbb{H} \to \mathbb{H}$ is a ring involution, i.e. for $q, q_0 \in \mathbb{H}$

$$\begin{split} \bar{\bar{q}} &= q, \\ \overline{q+q_0} &= \bar{q} + \bar{q}_0, \\ \overline{qq_0} &= \bar{q}_0 \bar{q}. \end{split}$$

Equipped with this involution, \mathbb{H} is a *-algebra.⁶

⁶A *-algebra is an algebra with an algebra involution. All quaternion ring involutions are also algebra involutions, i.e. for any $\lambda \in \mathbb{R}$, they satisfy the additional condition $\overline{\lambda q} = \lambda \overline{q}$. The corresponding involution on \mathbb{R} is the identity function. (In fact, the identity function is the only involution on \mathbb{R} .)

Definition 1.10. Let i_0 be a pure imaginary unit quaternion. Then

$$q^{(i_0)} = i_0 q i_0^{-1} = i_0 q \bar{i_0} = -i_0 q i_0.$$

Proposition 1.11. Let i_0 be a pure imaginary unit quaternion. Then $\bullet^{(i_0)} : \mathbb{H} \to \mathbb{H}$ is a ring anti-involution, *i.e.*

$$(q^{(i_0)})^{(i_0)} = q,$$

$$(q + q_0)^{(i_0)} = q^{(i_0)} + q_0^{(i_0)},$$

$$(qq_0)^{(i_0)} = q^{(i_0)}q_0^{(i_0)}.$$

The terms 'involution' and 'anti-involution' are often used interchangeably, since the distinction is only relevant for non-commutative rings. Since anti-involutions are ring automorphisms, they are also known as involutive automorphisms. For an extensive treatment of quaternion involutions and anti-involutions, see Ell and Sangwine (2007).

Proposition 1.12. \mathbb{H} is a unital⁷ associative normed division algebra, i.e. for $q, q_0, q_1 \in \mathbb{H}$

$$egin{aligned} q &= 1q = q1, \ q(q_0q_1) &= (qq_0)q_1, \ |qq_0| &= |q||q_0|, \end{aligned}$$

and if $q \neq 0$ then it has a unique inverse

$$q^{-1} = \frac{\bar{q}}{|q|^2}$$

Corollary 1.13. \mathbb{H} is a division ring (a.k.a. a skew-field). It satisfies all the axioms of a field except for commutativity of multiplication.

⁷All associative division algebras are unital.

Multiplication between quaternions and real numbers however commutes. Indeed the subalgebra $\mathbb{R} \subset \mathbb{H}$ is the center of \mathbb{H} .

Each of the subalgebras $\{x + yi : x, y \in \mathbb{R}\}, \{x + yj : x, y \in \mathbb{R}\}, \{x + yk : x, y \in \mathbb{R}\}$ is isomorphic to \mathbb{C} . In fact, if i_0 is an arbitrary pure imaginary unit quaternion, then since $i_0^2 = -1$, the subalgebra $\{x + yi_0 : x, y \in \mathbb{R}\}$ is isomorphic to \mathbb{C} .

Remark 1.14. For any $q = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$ we can write $q = a + y\mathbf{i}_0$ where

$$y = \sqrt{b^2 + c^2 + d^2}$$

i₀ =
$$\begin{cases} i & \text{if } y = 0\\ y^{-1}(bi + cj + dk) & \text{if } y > 0 \end{cases}$$

This gives us a way to extend the definitions of standard complex functions $f : \mathbb{C} \to \mathbb{C}$ to quaternion functions $f : \mathbb{H} \to \mathbb{H}$.

Example 1.15. Define

$$\exp(q) = \exp(a + y\mathbf{i}_0) = \exp(a)(\cos(y) + \sin(y)\mathbf{i}_0).$$

Note that generalising functions of more than one variable is not as straightforward, since i_0 may then be different for each variable.

1.2.2 Algebraic significance

The following uniqueness theorems make \mathbb{H} a particularly interesting algebraic structure to consider.

Theorem 1.16 (Frobenius' Theorem, (Palais, 1968)). \mathbb{R} , \mathbb{C} and \mathbb{H} are the only finitedimensional real associative division algebras up to isomorphism.

Definition 1.17. The direct sum of two matrices $A \in \mathbb{A}_1^{m \times m}$, $B \in \mathbb{A}_2^{n \times n}$ is the block-diagonal matrix

$$oldsymbol{A} \oplus oldsymbol{B} = \left(egin{array}{cc} oldsymbol{A} & oldsymbol{0}_{m imes m} \ oldsymbol{0}_{n imes m} & oldsymbol{B} \end{array}
ight).$$

The direct sum of two (matrix) algebras is correspondingly

$$\mathbb{A}_1^{m \times m} \oplus \mathbb{A}_2^{n \times n} = \left\{ \boldsymbol{A} \oplus \boldsymbol{B} : \boldsymbol{A} \in \mathbb{A}_1^{m \times m}, \boldsymbol{B} \in \mathbb{A}_2^{n \times n} \right\},\$$

equipped with block matrix multiplication.

Consider also the following result.

Theorem 1.18 (Artin–Wedderburn theorem (Grillet, 2007)). A (Artinian) ring is semisimple if and only if it is isomorphic to a direct sum⁸

$$\mathbb{A}_1^{n_1 \times n_1} \oplus \cdots \oplus \mathbb{A}_s^{n_s \times n_s}$$

of finitely many matrix rings over division rings $\mathbb{A}_1, \ldots, \mathbb{A}_s$.

Combining Theorems 1.16 and 1.18 gives us the following:

Corollary 1.19. Every (finite-dimensional) real semi-simple algebra is isomorphic to a direct sum

$$\mathbb{A}_1^{n_1 \times n_1} \oplus \cdots \oplus \mathbb{A}_s^{n_s \times n_s}$$

of finitely many matrix algebras where $\mathbb{A}_1, \ldots, \mathbb{A}_s \in \{\mathbb{R}, \mathbb{C}, \mathbb{H}\}$.

Remark 1.20. A sufficient condition for a finite-dimensional algebra to be semi-simple is that it has no non-trivial nilpotent right-ideals.

In particular, as shown by Garling (2011, pp. 97–98), Clifford algebras (with nondegenerate inner product) ⁹ are isomorphic to either $\mathbb{A}^{n \times n}$ or $\mathbb{A}^{n \times n} \oplus \mathbb{A}^{n \times n}$, where $\mathbb{A} = \mathbb{R}, \mathbb{C}$ or \mathbb{H} . Tian (1998) shows how to construct such isomorphisms explicitly.

⁸Direct sum and direct product are equivalent in this finite context.

⁹A typical Clifford algebra Cl(p,q) is generated by p grade 1 basis elements squaring to 1 and q grade 1 basis elements squaring to -1, and is fully determined by the (non-positive-definite) inner product $\langle x, y \rangle = \Re(xy)$ defined for linear combinations of grade 1 basis elements x, y. More general Clifford algebras Cl(p,q,r) can be defined, with an additional r grade 1 basis elements which square to 0. In such cases the inner product is degenerate in the sense that $\exists x \neq 0 : \langle x, x \rangle = 0$.

1.2.3 Matrix representation

 \mathbb{C} is a two-dimensional real vector space with basis 1, *i*. Consider complex multiplication by a + bi as a linear operator on \mathbb{R}^2 with basis 1, *i*, then its matrix is given by:

$$\left(\begin{array}{cc} a & -b \\ b & a \end{array}\right). \tag{1.2}$$

Such structured 2 × 2 real matrices form a real algebra which is isomorphic to \mathbb{C} . Similarly, if we consider multiplication on the left by a quaternion q = a + bi + cj + dk as a real linear operator on \mathbb{R}^4 with basis 1, i, j, k, then we can see from (1.1) that its matrix is

$$\begin{pmatrix}
a & -b & -c & -d \\
b & a & -d & c \\
c & d & a & -b \\
d & -c & b & a
\end{pmatrix}.$$
(1.3)

Such structured 4×4 real matrices form a real algebra isomorphic to \mathbb{H} .

These representations provide the crucial connection between operations performed in the real, complex and quaternion domains. They allow us to view complex and quaternion statistical theory as specialisations of real statistical theory, with structured linear transformations and structured covariance matrices.¹⁰ We note in particular that the quaternion matrix representation of a complex number is the direct sum of two copies of its complex representation.

These isomorphisms can be generalised to matrices with complex or quaternion entries. This will be covered in Section 1.3.1, and provides the main tool of quaternion linear algebra.

¹⁰It is worth noting that the isomorphism used is not unique since $\mathbb{R}^{4\times4}$ has many automorphisms, namely similarity transformations $M \to PMP^{-1}$, or — if we wish to preserve the *–algebra structure — orthogonal similarity transformations. In addition, there are representations of quaternions as structured 2 × 2 complex matrices.

1.3 Quaternion matrices

In this section we will present certain standard results on quaternion matrices which will be of use in later sections. We show that to a large extent, we can manipulate quaternion matrices as we would complex matrices. For a more comprehensive treatment of quaternion linear algebra, see Davis (2009); Farenick and Pidkowich (2003); Zhang (1997).

Consider a quaternion matrix $\boldsymbol{Q} = \boldsymbol{A} + \boldsymbol{B}\mathbf{i} + \boldsymbol{C}\mathbf{j} + \boldsymbol{D}\mathbf{k} \in \mathbb{H}^{m \times n}$, where $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D} \in \mathbb{R}^{m \times n}$. Let $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{H}^n$ be quaternion (column) vectors and let $q, q_0 \in \mathbb{H}$. Then $\boldsymbol{Q}(\boldsymbol{u}q + \boldsymbol{v}q_0) = (\boldsymbol{Q}\boldsymbol{u})q + (\boldsymbol{Q}\boldsymbol{v})q_0$, but generally $\boldsymbol{Q}(q\boldsymbol{u} + q_0\boldsymbol{v}) \neq q(\boldsymbol{Q}\boldsymbol{u}) + q_0(\boldsymbol{Q}\boldsymbol{v})$, i.e. \boldsymbol{Q} is right quaternion linear, but in general not left quaternion linear. This motivates us to view \mathbb{H}^n as a right module over \mathbb{H} , i.e. a "vector space" where quaternion scalars multiply on the right. Quaternion matrices are then the (right-) linear operators on \mathbb{H}^n .

An alternate but equivalent theory can be developed by treating quaternion matrices as left-linear operators. We will touch on this topic in Section 1.3.2.

1.3.1 Representation as real matrices

Definition 1.21. Define the real vector space isomorphism $\mathcal{V} : \mathbb{H}^{m \times n} \to \mathbb{R}^{4m \times n}$

$$\mathcal{V}(\boldsymbol{A} + \boldsymbol{B}\mathbf{i} + \boldsymbol{C}\mathbf{j} + \boldsymbol{D}\mathbf{k}) = egin{pmatrix} \boldsymbol{A} \ \boldsymbol{B} \ \boldsymbol{C} \ \boldsymbol{D} \end{pmatrix}$$

 $\boldsymbol{Q} \in \mathbb{H}^{n \times n}$ can be thought of as a (real-) linear operator on $\mathbb{R}^{4n} = \mathcal{V}(\mathbb{H}^n)$.

Definition 1.22. Define $\widetilde{\bullet}$: $\mathbb{H}^{m \times n} \to \mathbb{R}^{4m \times 4n}$ by

$$\widetilde{\bullet} : \mathbf{A} + \mathbf{B}\mathbf{i} + \mathbf{C}\mathbf{j} + \mathbf{D}\mathbf{k} \mapsto \begin{pmatrix} \mathbf{A} & -\mathbf{B} & -\mathbf{C} & -\mathbf{D} \\ \mathbf{B} & \mathbf{A} & -\mathbf{D} & \mathbf{C} \\ \mathbf{C} & \mathbf{D} & \mathbf{A} & -\mathbf{B} \\ \mathbf{D} & -\mathbf{C} & \mathbf{B} & \mathbf{A} \end{pmatrix}.$$
(1.4)

Remark 1.23. For each choice of $m, n \in \mathbb{N}$ we define a different function named $\tilde{\bullet}$. This abuse of notation is unambiguous since the dimensions of the quaternion matrix implicitly determine which definition is used. This remark also applies to Definition 1.21 and similar operators defined later in this thesis.

Definition 1.24. Let $\widetilde{\mathbb{H}^{m \times n}}$ be the image of $\mathbb{H}^{m \times n}$ under $\widetilde{\bullet}$, i.e.

$$\widetilde{\mathbb{H}^{m \times n}} = \left\{ \begin{pmatrix} A & -B & -C & -D \\ B & A & -D & C \\ C & D & A & -B \\ D & -C & B & A \end{pmatrix} : A, B, C, D \in \mathbb{R}^{m \times n} \right\}.$$
 (1.5)

Matrices in $\widetilde{\mathbb{H}^{m \times n}}$ are said to have quaternion structure. Similarly, matrices of the form

$$\begin{pmatrix} \boldsymbol{A} & -\boldsymbol{B} \\ \boldsymbol{B} & \boldsymbol{A} \end{pmatrix}, \ \boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{m \times n}$$
(1.6)

are said to have complex structure.

Remark 1.25. In terms of tensor products, we have $\mathbb{H}^{m \times n} = \mathbb{R}^{m \times n} \otimes \mathbb{H}$ and $\widetilde{\mathbb{H}^{m \times n}} = \mathbb{R}^{m \times n} \otimes \widetilde{\mathbb{H}}$. Obviously, we could use instead $\widetilde{\mathbb{H}} \otimes \mathbb{R}^{m \times n}$ as a representation, i.e. $m \times n$ block matrices with 4×4 quaternion-structured blocks.¹¹

As in Andersson et al. (1983); Kabe (1984), we can define a (proper) quaternion normal distribution by giving the real and imaginary parts a joint real normal

¹¹We identify the abstract tensor product of two vector spaces with the vector space generated by the Kronecker products of their elements (see Lemma 3.4).

distribution with a quaternion-structured covariance matrix. We will consider the quaternion normal distribution in detail in Section 2.3.

Theorem 1.26. Let $Q \in \mathbb{H}^{m \times n}$, $R \in \mathbb{H}^{n \times p}$, $x \in \mathbb{R}$. Then

$$\widetilde{\boldsymbol{Q}\boldsymbol{R}} = \widetilde{\boldsymbol{Q}}\widetilde{\boldsymbol{R}}$$
(1.7)

$$\widetilde{\boldsymbol{Q}+\boldsymbol{R}} = \widetilde{\boldsymbol{Q}} + \widetilde{\boldsymbol{R}}$$
(1.7)

$$\widetilde{(\boldsymbol{x}\boldsymbol{Q})} = \boldsymbol{x}\widetilde{\boldsymbol{Q}}$$
(1.8)

Proof. To prove (1.7), note that (1.1) holds when $a, b, c, d, a_0, b_0, c_0, d_0$ are replaced by real matrices, and use block matrix multiplication on the right hand side. The remaining equalities are straightforward to check.

Remark 1.27. The matrix transpose operator \bullet^T is an involution for $\mathbb{R}^{n \times n}$, and the conjugate (Hermitian) transpose operator \bullet^H is an involution for $\mathbb{H}^{n \times n}$.

For square matrices, Theorem 1.26 can be summarised as:

Corollary 1.28. $\widetilde{\bullet}$: $\mathbb{H}^{n \times n} \to \widetilde{\mathbb{H}^{n \times n}} \subset \mathbb{R}^{4n \times 4n}$ is an isomorphism of real *-algebras.

In particular, note that for the $n \times n$ identity matrix I_n we have $\widetilde{I}_n = I_{4n}$.

Corollary 1.29. Q is invertible iff \widetilde{Q} is invertible, also

$$\widetilde{(\boldsymbol{Q}^{-1})} = \left(\widetilde{\boldsymbol{Q}}\right)^{-1}.$$

Definition 1.30. We denote by $GL_n(\mathbb{H})$ the set of invertible quaternion $n \times n$ matrices.

Proposition 1.31. Let $\boldsymbol{Q} \in GL_n(\mathbb{H})$, then $(\boldsymbol{Q}^H)^{-1} = (\boldsymbol{Q}^{-1})^H$

Proof.
$$(\widetilde{\boldsymbol{Q}^{H}})^{-1} = \left(\widetilde{\boldsymbol{Q}}^{T}\right)^{-1} = \left(\widetilde{\boldsymbol{Q}}^{-1}\right)^{T} = (\widetilde{\boldsymbol{Q}}^{-1})^{H}$$

Remark 1.32. For complex matrices, $\overline{\bullet}$ is an anti-involution and \bullet^T is an involution. This is no longer the case with quaternion matrices. Also, in general $(\overline{Q})^{-1} \neq \overline{(Q^{-1})}$ and $(Q^T)^{-1} \neq (Q^{-1})^T$. One way of thinking about this is that, $\overline{\bullet}$ maps between the right-module and left-module of quaternions, and these happen to be the same in the complex case due to commutativity.

Remark 1.33. $\tilde{\bullet}$ and \mathcal{V} preserve the linear operator structure associated with a matrix, i.e. for $\mathbf{Q} \in \mathbb{H}^{m \times n}$, $\mathbf{v} \in \mathbb{H}^n$

$$\mathcal{V}(\boldsymbol{Q}\boldsymbol{v}) = \widetilde{\boldsymbol{Q}}\,\mathcal{V}(\boldsymbol{v}).$$

Proof. This is the first column of the matrix equality $\widetilde{Qv} = \widetilde{Q}\widetilde{v}$.

More generally Remark 1.33 holds when replacing v by $V \in \mathbb{H}^{n \times k}$. This suggests a simple way of coding quaternion matrix multiplication using real matrix multiplication. In terms of computational complexity, the real-domain product $\tilde{Q} \mathcal{V}(V)$ requires exactly the same operations as the quaternion-domain product QV. The product $\tilde{Q}\tilde{V}$ however requires four times as many operations. Storing a 4×4 real matrix also uses four times as much memory as storing a quaternion. Thus, although using real matrix algorithms with quaternion-structured inputs will typically yield the desired results, the use of specialised quaternion algorithms (or equivalently computing and storing only the first block column of quaternion-structured matrices) can potentially quadruple efficiency.

It is noted in Andersson et al. (1983) that $\tilde{\bullet}$ and \mathcal{V} however do *not* preserve the bilinear (sesquilinear) form structure associated with a matrix. Instead, for $\mathbf{Q} \in \mathbb{H}^{m \times n}$, $\mathbf{v}, \mathbf{w} \in \mathbb{H}^{n}$

$$\Re \left(\boldsymbol{v}^{H} \boldsymbol{Q} \boldsymbol{w} \right) = \mathcal{V}(\boldsymbol{v})^{T} \widetilde{\boldsymbol{Q}} \mathcal{V}(\boldsymbol{w}).$$
(1.9)

This is the top left corner of the matrix equality $\widetilde{v}^H \widetilde{Q} w = \widetilde{v}^T \widetilde{Q} \widetilde{w}$. In particular, for $v, w \in \mathbb{H}^n$ we can write the real euclidean inner product as

$$\Re \left(\boldsymbol{v}^{H} \boldsymbol{w} \right) = \mathcal{V}(\boldsymbol{v})^{T} \, \mathcal{V}(\boldsymbol{w}). \tag{1.10}$$

Remark 1.34. By considering the subalgebra $\mathbb{C}^{n \times n} = \{ \mathbf{A} + \mathbf{B} i : \mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n} \} \subseteq \mathbf{H}^{n \times n}$ and noting that the first $2n \times 2n$ block of the quaternion representation $\tilde{\bullet}$ is

the complex representation (1.6), we can obtain the equivalent complex versions of all results in this section.

1.3.2 Left-linear quaternion matrix multiplication

In the above, we implicitly defined the product of two matrices $\boldsymbol{Q} \in \mathbb{H}^{m \times n}$ and $\boldsymbol{S} \in \mathbb{H}^{n \times p}$ as $\boldsymbol{P} = \boldsymbol{Q}\boldsymbol{S}$ with $(i, j)^{\text{th}}$ element $p_{i,j} = \sum_{\ell=1}^{n} q_{i,\ell} s_{\ell,j}$.

Because quaternion multiplication is non-commutative, we can define an alternate right multiplication $\boldsymbol{P} = \boldsymbol{Q} *^{R} \boldsymbol{S}$, where $(i, j)^{\text{th}}$ element of \boldsymbol{P} is now given by $p_{i,j} = \sum_{\ell=1}^{n} s_{\ell,j} q_{i,\ell}$. With this new multiplication, \mathbb{H}^{n} is a left module and quaternion matrices are left-linear operators.

Proposition 1.35. Let $Q \in \mathbb{H}^{m \times n}$ and $S \in \mathbb{H}^{n \times p}$. Then $\overline{QS} = \overline{Q} *^R \overline{S}$ and $\overline{Q} *^R \overline{S} = \overline{QS}$.

Proof. By Proposition 1.9
$$\sum_{\ell=1}^{n} \overline{q_{\ell,j} s_{i,\ell}} = \sum_{\ell=1}^{n} \overline{s_{i,\ell} q_{\ell,j}}$$
 and $\sum_{\ell=1}^{n} \overline{s_{i,\ell} q_{\ell,j}} = \sum_{\ell=1}^{n} q_{\ell,j} \overline{s_{i,\ell}}$

The two types of multiplication are related by the fact that conjugation $\overline{\bullet}$ defines an isomorphism between the algebra \mathbb{H} and the alternate quaternion algebra which we would obtain by taking $*^R$ instead of the usual quaternion multiplication.¹² This implies in particular that our choice for the definition of matrix multiplication is made without loss of generality.

Thinking of Q^{*R} as a real linear operator leads to the following alternate structured real matrix representation of quaternion matrices.

Definition 1.36. Define $\widetilde{\bullet}^R : \mathbb{H}^{m \times n} \to \mathbb{R}^{4m \times 4n}$ by

$$\widetilde{\bullet}^{R}: \mathbf{A} + \mathbf{B}\mathbf{i} + \mathbf{C}\mathbf{j} + \mathbf{D}\mathbf{k} \mapsto \begin{pmatrix} \mathbf{A} & -\mathbf{B} & -\mathbf{C} & -\mathbf{D} \\ \mathbf{B} & \mathbf{A} & \mathbf{D} & -\mathbf{C} \\ \mathbf{C} & -\mathbf{D} & \mathbf{A} & \mathbf{B} \\ \mathbf{D} & \mathbf{C} & -\mathbf{B} & \mathbf{A} \end{pmatrix}.$$
 (1.11)

¹²Another way of interpreting the alternate left-linear matrix multiplication $*^R$ is to note that $\boldsymbol{Q} *^R \boldsymbol{S} = (\boldsymbol{S}^T \boldsymbol{Q}^T)^T$, and work with row-vectors being multiplied by matrices on the right instead of column vectors being multiplied by matrices on the left.

Remark 1.37. \tilde{q}^R is the real matrix corresponding to multiplication by q on the right: for $\lambda, q \in \mathbb{H}, \tilde{q}^R \mathcal{V}(\lambda) = \mathcal{V}(\lambda q)$.¹³

More generally, let $\boldsymbol{Q} \in \mathbb{H}^{m \times n}$ and $\lambda \in \mathbb{H}$. Then

$$\mathcal{V}(\boldsymbol{Q}\lambda) = \widetilde{\lambda \boldsymbol{I}_m}^R \mathcal{V}(\boldsymbol{Q}) \tag{1.12}$$

Lemma 1.38. Let $M \in \mathbb{R}^{4m \times 4n}$. Then $M \in \widetilde{\mathbb{H}^{m \times n}}$ if and only if $M \lambda \widetilde{I_n}^R = \lambda \widetilde{I_m}^R M$ $\forall \lambda \in \mathbb{H}$.

Proof. For clarity, we omit the qualifiers $\forall q \in \mathbb{H}^n, \forall \lambda \in \mathbb{H}$ which apply to all equalities in this proof.

Let $M : \mathbb{H}^n \to \mathbb{H}^m$ be given by $M(q) = \mathcal{V}^{-1}(\mathcal{M}\mathcal{V}(q))$. *M* is real-linear, hence it is quaternion linear iff it satisfies $M(q\lambda) = M(q)\lambda$. Now by (1.12)

$$\mathcal{V}(M(\boldsymbol{q}\lambda)) = \boldsymbol{M} \, \mathcal{V}(\boldsymbol{q}\lambda) = \boldsymbol{M} \widetilde{\lambda \boldsymbol{I}_n}^R \mathcal{V}(\boldsymbol{q}),$$

and again by (1.12)

$$\mathcal{V}(M(\boldsymbol{q})\lambda) = \widetilde{\lambda \boldsymbol{I}_m}^R \mathcal{V}(M(\boldsymbol{q})) = \widetilde{\lambda \boldsymbol{I}_n}^R \boldsymbol{M} \mathcal{V}(\boldsymbol{q})$$

Since equality must hold $\forall \mathcal{V}(\boldsymbol{q}) \in \mathbb{R}^{4n}$, M is quaternion linear iff $\boldsymbol{M} \widetilde{\lambda \boldsymbol{I}_n}^R = \widetilde{\lambda \boldsymbol{I}_n}^R \boldsymbol{M}$.

Quaternion linearity of M is equivalent to the existence of $\mathbf{Q} \in \mathbb{H}^{m \times n}$ such that $M(\mathbf{q}) = \mathbf{Q}\mathbf{q}$,¹⁴ so that by Remark 1.33 $\mathbf{M}\mathcal{V}(\mathbf{q}) = \widetilde{\mathbf{Q}}\mathcal{V}(\mathbf{q})$.

Remark 1.39. In Lemma 1.38 it is actually sufficient to consider $\lambda \in \{i, j, k\}$ or even just $\lambda \in \{i, j\}$ instead of $\lambda \in \mathbb{H}$.¹⁵

Taken individually, commuting with right multiplication by $\lambda = i$ is equivalent to \mathbb{C}^{i} -linearity, as defined in Vía et al. (2010b) and similarly for other pure unit

 $^{^{13}\}text{The}$ matrix representation \widetilde{q}^R was first introduced by Ickes (1970) and referred to as the 'quaternion transmuted matrix'.

¹⁴To prove this, note that linear transformations are uniquely determined by their action on basis elements, which is uniquely encoded by the columns of Q.

¹⁵This is due to the real linearity assumption and k = ij.

quaternions.

Definition 1.40. We define the sum of two sets A, B as $A+B = \{a+b : a \in A, b \in B\}$, and the product with a set as $Ab = \{ab : a \in A\}$.

Corollary 1.41. Let $\lambda \in \mathbb{H}$. If $M \in \widetilde{\mathbb{H}^{m \times n}} + \widetilde{\mathbb{H}^{m \times n}} \lambda \widetilde{I_n}^R$ then $M \lambda \widetilde{I_n}^R = \lambda \widetilde{I_m}^R M$ $\forall \lambda \in \mathbb{H}$.

Proof. This follows immediately from Lemma 1.38.

1.3.3 The matrix product as a projection and ensemble

In the interest of brevity, within this section we introduce the notation $\lambda_0 = 1$, $\lambda_1 = i$, $\lambda_2 = j$, $\lambda_3 = \lambda_1 \lambda_2 = k$.

Definition 1.42. Let $\hat{h} : \mathbb{R}^{4m \times 4n} \to \mathbb{R}^{4m \times 4n}$ be given by

$$\hat{h}(\boldsymbol{M}) = \frac{1}{4} \sum_{i=0}^{3} \widetilde{\lambda_{i} \boldsymbol{I}_{m}}^{R} \boldsymbol{M} \widetilde{\lambda_{i} \boldsymbol{I}_{n}}^{RT}$$
$$= \frac{1}{4} \boldsymbol{M} - \frac{1}{4} \sum_{i=1}^{3} \widetilde{\lambda_{i} \boldsymbol{I}_{m}}^{R} \boldsymbol{M} \widetilde{\lambda_{i} \boldsymbol{I}_{n}}^{R}$$

Definition 1.43. For a pure unit quaternion η , let $\hat{c}^{\eta} : \mathbb{R}^{4m \times 4n} \to \mathbb{R}^{4m \times 4n}$ be given by

$$\hat{c}^{\eta}(\boldsymbol{M}) = rac{1}{2}\boldsymbol{M} + rac{1}{2}\widetilde{\eta \boldsymbol{I}_m}^R \boldsymbol{M} \widetilde{\eta \boldsymbol{I}_n}^{RD}$$

Remark 1.44. $\hat{h}=\hat{c}^{\mathbf{k}}\circ\hat{c}^{\mathbf{j}}$

Proof.

$$\begin{aligned} 4\hat{c}^{k}\left(\hat{c}^{j}(\boldsymbol{M})\right) &= 2\hat{c}^{k}\left(\boldsymbol{M}+\widetilde{\boldsymbol{j}\boldsymbol{I}_{m}}^{R}\boldsymbol{M}\widetilde{\boldsymbol{j}\boldsymbol{I}_{n}}^{RT}\right) \\ &= \boldsymbol{M}+\widetilde{\boldsymbol{j}\boldsymbol{I}_{m}}^{R}\boldsymbol{M}\widetilde{\boldsymbol{j}\boldsymbol{I}_{n}}^{RT}+\widetilde{\boldsymbol{k}\boldsymbol{I}_{m}}^{R}\boldsymbol{M}\widetilde{\boldsymbol{k}\boldsymbol{I}_{n}}^{RT}+\widetilde{\boldsymbol{k}\boldsymbol{j}\boldsymbol{I}_{m}}^{R}\boldsymbol{M}\widetilde{\boldsymbol{k}\boldsymbol{j}\boldsymbol{I}_{n}}^{RT} \\ &= 4\hat{h}(\boldsymbol{M}) \end{aligned}$$

We proceed to prove that \hat{h} is the orthogonal projection onto $\mathbb{H}^{m \times n}$. \hat{c}^{η} on the other hand is the orthogonal projection onto $\widetilde{\mathbb{H}^{m \times n}} + \widetilde{\mathbb{H}^{m \times n}} \eta \widetilde{I_n}^R$, and this could be proved in the same fashion.

Lemma 1.45. Let $M \in \mathbb{R}^{4m \times 4n}$ and $N \in \widetilde{\mathbb{H}^{m \times n}} + \widetilde{\mathbb{H}^{m \times n}} \eta \widetilde{I_n}^R$. Then

$$\operatorname{tr}\left(\boldsymbol{N}^{T}\hat{c}^{\eta}(\boldsymbol{M})\right) = \operatorname{tr}\left(\boldsymbol{N}^{T}\boldsymbol{M}\right)$$

Proof. Note that $\mathbf{N}^T \in \widetilde{\mathbb{H}^{n \times m}} + \widetilde{\mathbb{H}^{m \times n}} \widetilde{\eta \mathbf{I}_m}^R$. ¹⁶ Also note that $\overline{\eta} = -\eta$ so $\widetilde{\lambda \mathbf{I}_n}^{RT} = -\widetilde{\lambda \mathbf{I}_n}^R$. Using Corollary 1.41,

$$2\operatorname{tr} \left(\boldsymbol{N}^{T} \hat{c}^{\eta}(\boldsymbol{M}) \right) = \operatorname{tr} \left(\boldsymbol{N}^{T} \boldsymbol{M} \right) - \operatorname{tr} \left(\boldsymbol{N}^{T} \widetilde{\boldsymbol{\eta} \boldsymbol{I}_{m}}^{R} \boldsymbol{M} \widetilde{\boldsymbol{\eta} \boldsymbol{I}_{n}}^{R} \right)$$
$$= \operatorname{tr} \left(\boldsymbol{N}^{T} \boldsymbol{M} \right) - \operatorname{tr} \left(\widetilde{\boldsymbol{\eta} \boldsymbol{I}_{n}}^{R} \widetilde{\boldsymbol{\eta} \boldsymbol{I}_{n}}^{R} \boldsymbol{N}^{T} \boldsymbol{M} \right)$$
$$= 2 \operatorname{tr} \left(\boldsymbol{N}^{T} \boldsymbol{M} \right)$$

Lemma 1.46. Let $M \in \mathbb{R}^{4m \times 4n}$ and $N \in \widetilde{\mathbb{H}^{m \times n}}$. Then

$$\operatorname{tr}\left(\boldsymbol{N}^{T}\hat{h}(\boldsymbol{M})\right) = \operatorname{tr}\left(\boldsymbol{N}^{T}\boldsymbol{M}\right).$$

Proof. $\widetilde{\mathbb{H}^{m \times n}} \subseteq \widetilde{\mathbb{H}^{n \times m}} + \widetilde{\mathbb{H}^{m \times n}} \widetilde{\eta I_m}^R$ for $\eta = j, k$. Hence using Remark 1.44 and applying Lemma 1.45 twice,

$$\operatorname{tr}\left(\boldsymbol{N}^{T}\hat{h}(\boldsymbol{M})\right) = \operatorname{tr}\left(\boldsymbol{N}^{T}\hat{c}^{k}(\hat{c}^{j}(\boldsymbol{M}))\right) = \operatorname{tr}\left(\boldsymbol{N}^{T}\hat{c}^{j}(\boldsymbol{M})\right) = \operatorname{tr}\left(\boldsymbol{N}^{T}\boldsymbol{M}\right).$$

Proposition 1.47. \hat{h} is the orthogonal projection of $\mathbb{R}^{4m \times 4n}$ onto $\widetilde{\mathbb{H}^{m \times n}}$ *Proof.* By Lemma 1.38, for any $\boldsymbol{M} \in \widetilde{\mathbb{H}^{m \times n}}$ and i = 0, 1, 2, 3,

$$\widetilde{\lambda_i I_m}^R M \widetilde{\lambda_i I_n}^{RT} = \widetilde{\lambda_i \overline{\lambda_i I_m}}^R M = M$$

¹⁶ Let $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{H}^{m \times n}$. Then by Lemma 1.38 $\left(\widetilde{\boldsymbol{A}} + \widetilde{\boldsymbol{B}} \widetilde{\eta \boldsymbol{I}_n}^R\right)^T = \widetilde{\boldsymbol{A}}^T - \widetilde{\eta \boldsymbol{I}_n}^R \widetilde{\boldsymbol{B}}^T = \widetilde{\boldsymbol{A}}^H - \widetilde{\boldsymbol{B}}^H \widetilde{\eta \boldsymbol{I}_m}^R$.

Hence $\hat{h}(\boldsymbol{M}) = \boldsymbol{M}$.

Now consider $\boldsymbol{M} \in \mathbb{R}^{m \times n}$.

$$4\widetilde{\lambda_{1}I_{m}}^{R}\widehat{h}(\boldsymbol{M})\widetilde{\lambda_{1}I_{n}}^{RT} = \sum_{i=0}^{n}\widetilde{\lambda_{i}\lambda_{1}I_{m}}^{R}\boldsymbol{M}\widetilde{\lambda_{i}\lambda_{1}I_{n}}^{RT}$$
$$= \widetilde{\lambda_{1}I_{m}}^{R}\boldsymbol{M}\widetilde{\lambda_{1}I_{n}}^{RT} + \widetilde{\lambda_{0}I_{m}}^{R}\boldsymbol{M}\widetilde{-\lambda_{0}I_{n}}^{RT}$$
$$\widetilde{-\lambda_{3}I_{m}}^{R}\boldsymbol{M}\widetilde{-\lambda_{3}I_{n}}^{RT} + \widetilde{\lambda_{2}I_{m}}^{R}\boldsymbol{M}\widetilde{\lambda_{2}I_{n}}^{RT}$$
$$= 4\widehat{h}(\boldsymbol{M}).$$

Multiplying the first and last expression by $\frac{1}{4}\widetilde{\lambda_1 I_n}^R$ on the right we get $\widetilde{\lambda_1 I_n}^R \hat{h}(\boldsymbol{M}) = \hat{h}(\boldsymbol{M})\widetilde{\lambda_1 I_n}^R$. Similarly we can obtain $\widetilde{\lambda_2 I_m}^R \hat{h}(\boldsymbol{M}) = \hat{h}(\boldsymbol{M})\widetilde{\lambda_2 I_n}^R$. Hence by Remark 1.39, $\hat{h}(\boldsymbol{M}) \in \widetilde{\mathbb{H}^{m \times n}}$. Thus, $\hat{h} \circ \hat{h} = \hat{h}$ and \hat{h} is a projection onto $\widetilde{\mathbb{H}^{m \times n}}$.

Now to prove orthogonality, since $\langle \mathbf{N}, \mathbf{M} \rangle = \operatorname{tr}(\mathbf{N}^T \mathbf{M})$ is the scalar product on $\mathbb{R}^{4m \times 4n}$, it is sufficient to show that for any $\mathbf{N} \in \widetilde{\mathbb{H}^{m \times n}}$, $\langle \mathbf{N}, \mathbf{M} - \hat{h}(\mathbf{M}) \rangle = 0$, or equivalently $\langle \mathbf{N}, \hat{h}(\mathbf{M}) \rangle = \langle \mathbf{N}, \mathbf{M} \rangle$, which we know by Lemma 1.46.

Note from (1.4) that

$$\widetilde{\boldsymbol{Q}} = \begin{bmatrix} \mathcal{V}(\boldsymbol{Q}) & \mathcal{V}(\boldsymbol{Q}i) & \mathcal{V}(\boldsymbol{Q}j) & \mathcal{V}(\boldsymbol{Q}k) \end{bmatrix}$$
 (1.13)

Consider $X \in \mathbb{H}^{m \times k}$, $Y \in \mathbb{H}^{n \times k}$. We wish to gain some insight on the quaternion product XY^{H} and its real representation. From (1.13) and (1.12) we have

$$\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{Y}}^{T} = \mathcal{V}(\boldsymbol{X})\mathcal{V}(\boldsymbol{Y})^{T} + \mathcal{V}(\boldsymbol{X}i)\mathcal{V}(\boldsymbol{Y}i)^{T} + \mathcal{V}(\boldsymbol{X}j)\mathcal{V}(\boldsymbol{Y}j)^{T} + \mathcal{V}(\boldsymbol{X}k)\mathcal{V}(\boldsymbol{Y}k)^{T}(1.14)$$

$$= \sum_{i=0}^{3} \widetilde{\lambda_{i}}\widetilde{\boldsymbol{I}_{m}}^{R}\mathcal{V}(\boldsymbol{X})\mathcal{V}(\boldsymbol{Y})^{T}\widetilde{\lambda_{i}}\widetilde{\boldsymbol{I}_{n}}^{RT}$$

$$= 4\hat{h}(\mathcal{V}(\boldsymbol{X})\mathcal{V}(\boldsymbol{Y})^{T}). \qquad (1.15)$$

 $\mathcal{V}(\mathbf{X}) \mathcal{V}(\mathbf{Y})^T$ can be thought of as a "block matrix outer product". Up to a multiplicative factor of 4, through (1.14) we can think of $\widetilde{\mathbf{X}}\widetilde{\mathbf{Y}}^T$ as an ensemble average of these products, taken over the ensemble of pairs $\{(\mathbf{X}\lambda_i, \mathbf{Y}\lambda_i) : i = 0, 1, 2, 3\}$. On the other hand (1.15) allows us to interpret quaternion multiplication in terms of the

quaternion-structured projection of this "outer product".

The results of this section suggest that it will often be possible to interpret algorithms using quaternion matrices as real-valued algorithms which work on an artificial ensemble of observations, and/or as real-valued algorithms imposing an assumed quaternion structure through projection(s).

1.3.4 Determinant, trace and norm

There are multiple possible definitions for quaternionic determinants, which have been reviewed by Aslaksen (1996). We will use the Dieudonné determinant (Dieudonné, 1943).

Definition 1.48. Denote by $|\bullet|_{\mathbb{C}}$ the usual determinant for real and complex matrices. The (quaternionic) determinant $|\bullet| : \mathbb{H}^{n \times n} \to \mathbb{R}$ of Q is¹⁷

$$|\boldsymbol{Q}| = \left| \boldsymbol{\widetilde{Q}} \right|_{\mathbb{C}}^{\frac{1}{4}}.$$
 (1.16)

Remark 1.49. By Corollary 1.29, \boldsymbol{Q} is invertible iff $|\boldsymbol{Q}| \neq 0$.

Remark 1.50. The determinant of a quaternion scalar $(1 \times 1 \text{ matrix})$ is its norm.

Remark 1.51. The quaternionic determinant of a real matrix is the absolute value of its real determinant, and hence does not generalise the real (or complex) determinant. However, $|\bullet|$ and $|\bullet|_{\mathbb{C}}$ are equal for real symmetric positive semidefinite matrices.

Proposition 1.52 (Dieudonné (1943)). Let $Q, R \in \mathbb{H}^{n \times n}$. Then

$$\begin{aligned} \boldsymbol{Q}\boldsymbol{R} &= & |\boldsymbol{Q}| \, |\boldsymbol{R}| \,, \\ |\boldsymbol{Q}^{T}| &= & |\boldsymbol{Q}|. \end{aligned} \tag{1.17}$$

Corollary 1.53. Let $Q \in \mathbb{H}^{n \times n}$. Then

$$\left| \bar{\boldsymbol{Q}} \right| = \left| \boldsymbol{Q}^H \right| = \left| \boldsymbol{Q} \right|.$$

¹⁷It can be shown that the determinant of a quaternion-structured matrix is always non-negative by considering its singular value decomposition and the eigenvalue decomposition of the orthogonal factors.
Proof. Apply (1.17) followed by (1.16) and (1.8).

Proposition 1.54. Let $Q \in \mathbb{H}^{n \times n}$. Then

$$\operatorname{tr}\left(\widetilde{\boldsymbol{Q}}\right) = 4\Re\operatorname{tr}(\boldsymbol{Q}).$$
 (1.18)

Proof. This is immediate from (1.4) and the linearity of the trace.

The quaternion trace generalises the real and complex traces.

Corollary 1.55. Let $Q, R \in \mathbb{H}^{m \times n}$. Then

$$\Re \operatorname{tr}(\boldsymbol{R}^{T}\boldsymbol{Q}) = \Re \operatorname{tr}(\boldsymbol{Q}\boldsymbol{R}^{T}).$$
(1.19)

Remark 1.56. In general $\operatorname{tr}(\boldsymbol{R}^T\boldsymbol{Q}) \neq \operatorname{tr}(\boldsymbol{Q}\boldsymbol{R}^T)$. Instead we have $\operatorname{tr}(\boldsymbol{R}^T\boldsymbol{Q}) = \operatorname{tr}(\boldsymbol{Q}*^R\boldsymbol{R}^T)$. This differs from the complex case. However, it still holds that

$$\operatorname{tr}(\boldsymbol{R}^{T}\boldsymbol{Q}) = \sum_{i=1}^{m} \sum_{j=1}^{n} r_{i,j} q_{i,j} = \operatorname{tr}(\boldsymbol{R}\boldsymbol{Q}^{T}).$$
(1.20)

Definition 1.57. The Frobenius or L^2 norm of $Q \in \mathbb{H}^{m \times n}$ is given by

$$||\boldsymbol{Q}|| = \left[\sum_{i=1}^{m} \sum_{j=1}^{n} |q_{i,j}|^2\right]^{\frac{1}{2}}$$

This generalises the usual real and complex Frobenius norm. Also note that by Proposition 1.54

$$||\widetilde{\boldsymbol{Q}}||^{2} = \operatorname{tr}\left(\widetilde{\boldsymbol{Q}^{H}\boldsymbol{Q}}\right) = 4\operatorname{tr}\left(\boldsymbol{Q}^{H}\boldsymbol{Q}\right) = 4||\boldsymbol{Q}||^{2}, \qquad (1.21)$$

1.3.5 Special matrices and decompositions

Definition 1.58. $Q \in \mathbb{H}^{n \times n}$ is said to be

Normal iff $Q^H Q = Q Q^H$

Unitary iff $Q^H Q = Q Q^H = I_n$

Hermitian iff $Q^H = Q$

Hermitian positive definite (QHPD) iff Q is Hermitian and $v^H Q v > 0 \ \forall v \in \mathbb{H}^n \setminus \{0\}$

Upper (resp. lower) triangular iff $q_{i,j} = 0 \ \forall j < (resp. >)i$.

Remark 1.59. All unitary matrices and all Hermitian matrices are normal.

Remark 1.60. \boldsymbol{Q} is Hermitian (resp. QHPD) iff \boldsymbol{A} is symmetric (resp. positive definite) and $\boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}$ are skew-symmetric¹⁸.

Lemma 1.61. Let $M \in GL_n(\mathbb{H})$. Consider the map $g_M : \mathbb{H}^{n \times n} \to \mathbb{H}^{n \times n}$ given by

$$oldsymbol{Q}\mapsto M^{\scriptscriptstyle H}oldsymbol{Q}M.$$

When restricted to the appropriate subset it is:

- 1. A bijection of $\mathbb{H}^{n \times n}$ onto itself
- 2. A bijection of $\operatorname{GL}_n(\mathbb{H})$ onto itself
- 3. A bijection of the space of $n \times n$ Hermitian matrices onto itself
- 4. A bijection of the space of $n \times n$ QHPD matrices onto itself

Proof.

- 1. $g_{\boldsymbol{M}}^{-1}(\boldsymbol{Q}) = g_{\boldsymbol{M}^{-1}}(\boldsymbol{Q})$. Hence $g_{\boldsymbol{M}}$ is invertible. Let $X \subseteq \mathbb{H}^{n \times n}$. If $\forall \boldsymbol{M} \in \operatorname{GL}_n(\mathbb{H}), g_{\boldsymbol{M}}(X) \subseteq X$, then $\forall \boldsymbol{M} \in \operatorname{GL}_n(\mathbb{H}), g_{\boldsymbol{M}^{-1}}(X) \subseteq X$. Hence to prove that $g_{\boldsymbol{M}}$ is a bijection for the set of matrice with a certain property, it is sufficient to show that for arbitrary $\boldsymbol{M}, g_{\boldsymbol{M}}$ preserves that property.
- 2. $(g_{\boldsymbol{M}}(\boldsymbol{Q}))^{-1} = g_{\boldsymbol{M}^{H^{-1}}}(\boldsymbol{Q}^{-1}).$
- 3. $g_{\boldsymbol{M}}(\boldsymbol{Q})^{H} = g_{\boldsymbol{M}}(\boldsymbol{Q}^{H}) = g_{\boldsymbol{M}}(\boldsymbol{Q}).$

¹⁸ \boldsymbol{B} is skew-symmetric iff $\boldsymbol{B}^T = -\boldsymbol{B}$.

4. Taking $\boldsymbol{v} = \boldsymbol{M}\boldsymbol{w}$ gives $\boldsymbol{w}^H g_{\boldsymbol{M}}(\boldsymbol{Q})\boldsymbol{w} = \boldsymbol{v}^H \boldsymbol{Q}\boldsymbol{v} > 0$.

Definition 1.62. For $Q \in \mathbb{H}^{n \times n}$, a (right) eigenvalue-eigenvector pair is a pair $\lambda \in \mathbb{H}, v \in \mathbb{H}^n \setminus \{0\}$ satisfying

$$Qv = v\lambda$$
.

Remark 1.63. We can define left eigenvalues in a similar fashion. The theory behind left eigenvalues however is not immediately comparable to the complex case, and a topic of current research (Davis, 2009).¹⁹

Theorem 1.64 (Quaternion Spectral Theorem). Let $Q \in \mathbb{H}^{n \times n}$. Then Q is normal if and only if there exist $U, D \in \mathbb{H}^{n \times n}$ such that:

- 1. U is unitary, D is diagonal and $Q = U^H D U$,
- 2. the diagonal entries of D are in \mathbb{C} and have non-negative imaginary part,
- 3. $\lambda \in \mathbb{H}$ is a (right) eigenvalue of \mathbf{Q} iff $\exists r \in \mathbb{H} \setminus \{0\} : r^{-1}\lambda r$ is a diagonal entry of \mathbf{D} .

Proof. See Farenick and Pidkowich (2003) for the 'only if' statement. For the 'if' statement,

$$Q^{H}Q = U^{H}D^{H}UU^{H}DU = U^{H}D^{H}DU = U^{H}DD^{H}U = U^{H}DUU^{H}D^{H}U = QQ^{H}$$

Remark 1.65. More generally, we can perform a quaternion singular value decomposition on any $Q \in \mathbb{H}^{m \times n}$ (Le Bihan and Mars, 2004; Sangwine and Le Bihan, 2006).

Corollary 1.66. In Theorem 1.64 we have furthermore,

¹⁹Based on Section 1.3.2, the theory of right eigenvalues for right-linear transformations is equivalent to the theory of left eigenvalues for left-linear transformations, but the theory of left eigenvalues for right-linear transformations is not.

- Q is Hermitian iff all entries of D are real.
- Q is QHPD iff all diagonal entries of D are real and positive.

Proof. By Lemma 1.61, \boldsymbol{Q} is Hermitian (resp. QHPD) iff \boldsymbol{D} is Hermitian (resp. QHPD).

Lemma 1.67. Let $\Sigma \in \mathbb{H}^{n \times n}$ be QHPD and $\Theta \in \mathbb{H}^{n \times n}$ be Hermitian. Then there exist $M \in GL_n(\mathbb{H})$ and a diagonal matrix D with real entries such that

$$egin{array}{rcl} M^H \Sigma M &=& I \ M^H \Theta M &=& D. \end{array}$$

Furthermore, if Θ is QHPD, then **D** has positive entries.

Proof. By Theorem 1.64 \exists : V unitary, G diagonal with real positive entries, such that $\Sigma = V^H G V$. Let $Q = G^{\frac{1}{2}} V^{20} Q^{H^{-1}} \Theta Q^{-1}$ is Hermitian (or QHPD) by Lemma 1.61. By Theorem 1.64 \exists : U unitary s.t. $Q^{H^{-1}} \Theta Q^{-1} = U^H D U$. Set $M = Q^{-1} U^{-1}$.

Theorem 1.68 (Quaternion Cholesky Decomposition). Let Σ be QHPD. Then there exists a unique upper triangular matrix T with positive real diagonal elements such that $\Sigma = T^H T$.

Proof. The proof of Stewart (1998, Theorem 2.7) for the complex Cholesky decomposition can be applied to the quaternion case without adjustments. \Box

 $^{{}^{20}}G^{rac{1}{2}}$ is obtained by taking the square root of each real positive entry in the diagonal.

Chapter 2

Quaternion Probability Distributions

2.1 Introduction

In this chapter we provide the necessary definitions for a rigorous treatment of quaternion-valued random variables. We will give the densities and characteristic functions for the (proper) quaternion normal and Wishart distributions.

There are two main approaches to defining a quaternion normal distribution. The *improper* approach defines a vector \boldsymbol{q} to be quaternion normal iff the real vector containing its components $\mathcal{V}(\boldsymbol{q})$ is real normal. Similarly to the complex case, the quaternion covariance matrix fails to capture the full second order properties of an improper quaternion random vector (i.e the real covariance matrix of $\mathcal{V}(\boldsymbol{q})$ cannot be computed from the quaternion covariance matrix of \boldsymbol{q}). Further information is contained in three complementary quaternion covariance matrices. Thus a treatment of improper distributions in the quaternion domain typically relies on augmented quaternions and their covariance matrix, as we describe in Section 2.5. We are more interested in the *proper* approach. A proper (a.k.a. \mathbb{H} -proper) quaternion normal distribution is a special case of the improper distribution where the complementary covariance matrices are assumed to be $\mathbf{0}$ so that all second-order information is contained in the quaternion covariance matrix. Vía et al. (2010a,b) show that using

quaternion linear processing for partial least squares, principal component analysis, multivariate linear regression or canonical correlation analysis is optimal for proper quaternion random vectors, as opposed to the improper case where widely-linear transformations are required.¹

Working in \mathbb{H}^n with widely-linear transformations is equivalent to working in \mathbb{R}^{4n} with real linear transformations. Working in the quaternion domain may still be useful if the quaternion-linear part and/or the complementary parts of a real linear transformation have meaningful interpretations for the problem at hand, since it helps visualise and separate the corresponding four orthogonal subspaces of $\mathbb{R}^{m \times n}$. Widelylinear complex modelling for example is popular for rotational processes because the complex-linear and complementary parts correspond to counter-clockwise and clockwise components (Rubin-Delanchy, 2008; Schreier, 2010). However, we have yet to find a practical application of widely-linear quaternion signal processing where this is the case.

Various sometimes inconsistent definitions and parameterisations of the proper quaternion normal distribution have been suggested. In particular, we show that right- and left-proper quaternion random vectors are conjugates. We choose to work with the definition which best generalises the usual proper (a.k.a. circular) complex normal distribution.² This provides the foundation for later statistical work, especially Chapter 4.

When quaternions are used to represent orientations they are restricted to have unit norm, so distributions on the hypersphere S^3 and multiplicative errors are a more appropriate model than the quaternion normal distribution and additive errors. One such distribution is the Bingham distribution (a.k.a. von Mises-Fisher distribution) which was used by Glover and Kaelbling (2013) in a Kalman-like filter for orientation tracking. We will not explore this avenue of research.

¹Weaker assumptions than joint propriety are sufficient for partial least squares and multivariate linear regression (Vía et al., 2010a, Table II). Also, even for improper data, when the sample size is small proper models may be more efficient (Vía et al., 2010b, Figure 1).

²Arguably our choice of right-propriety instead of left-propriety is arbitrary. However, it is the more common choice, possibly because of a preference for interpreting vectors as column vectors.

We will present a novel first-principles derivation of the quaternion Wishart distribution (the distribution of a quaternion sample covariance matrix), which we obtain by adapting the method used by Goodman (1963) to derive the complex Wishart distribution (see also Appendix B.1). We also discuss other derivations and generalisations in Section 2.4.1.

2.2 Characteristic functions

Definition 2.1. Let q be a quaternion random variable. Its characteristic function is given by $\phi_q : \mathbb{H} \to \mathbb{C}$,

$$\phi_q(\theta) = \mathbb{E}\left[\exp\left(\mathrm{i}\Re\left(\bar{\theta}q\right)\right)\right].$$

Let \mathbf{q} be a quaternion (column) vector random variable. Its characteristic function is given by $\phi_q : \mathbb{H}^n \to \mathbb{C}$,

$$\phi_{\boldsymbol{q}}(\boldsymbol{\theta}) = \mathbb{E}\left[\exp\left(\mathrm{i}\Re\left(\boldsymbol{\theta}^{H}\boldsymbol{q}\right)\right)\right].$$

Let Q be a $m \times n$ quaternion matrix random variable. Its characteristic function is given by $\phi_Q : \mathbb{H}^{m \times n} \to \mathbb{C}$,

$$\phi_{\boldsymbol{Q}}(\boldsymbol{\Theta}) = \mathbb{E}\left[\exp\left(i\Re\operatorname{tr}\left(\boldsymbol{\Theta}^{H}\boldsymbol{Q}\right)\right)\right].$$
(2.1)

If Q is Hermitian, then it is enough to specify $\phi_Q(\Theta)$ for Θ Hermitian.³

Proposition 2.2. Let q, q and Q be a quaternion random scalar, vector and matrix respectively. Then

$$\phi_q(\theta) = \phi_{\mathcal{V}(q)}(\mathcal{V}(\theta)), \qquad (2.2)$$

$$\phi_{\boldsymbol{q}}(\boldsymbol{\theta}) = \phi_{\mathcal{V}(\boldsymbol{q})}\left(\mathcal{V}(\boldsymbol{\theta})\right), \qquad (2.3)$$

$$\phi_{\boldsymbol{Q}}(\boldsymbol{\Theta}) = \phi_{\widetilde{\boldsymbol{Q}}}\left(\frac{1}{4}\widetilde{\boldsymbol{\Theta}}\right) \tag{2.4}$$

$$= \phi_{\Re(\boldsymbol{Q}),\Im_{i}(\boldsymbol{Q}),\Im_{j}(\boldsymbol{Q}),\Im_{k}(\boldsymbol{Q})}\left(\Re(\boldsymbol{\Theta}),\Im_{i}(\boldsymbol{\Theta}),\Im_{j}(\boldsymbol{\Theta}),\Im_{k}(\boldsymbol{\Theta})\right), \qquad (2.5)$$

³This effectively gives us the joint characteristic function of $q_{1,1}, q_{2,2}, \ldots, q_{n,n}, 2q_{1,2}, 2q_{1,3}, \ldots, 2q_{1,n}, 2q_{2,3}, \ldots, 2q_{n-1,n}$.

where the usual characteristic functions of real statistical theory are used on the right hand side, and (2.5) is a joint characteristic function.

Proof. (2.2)-(2.3) follow from (1.10). (2.4) follows from Proposition 1.54. (2.5) can be shown by expanding $\Re \operatorname{tr}(\Theta^H Q)$ in (2.1)

Since the quaternion characteristic function is directly related to the real characteristic function, the same existence and uniqueness results apply.

2.3 (Proper) normal distribution

We will now consider (proper) quaternion random vectors which can be described and manipulated by quaternion-linear transformations. This construction takes into account the multiplicative structure of quaternions in addition to the real vector space structure $\mathcal{V}(\mathbb{H}^p) = \mathbb{R}^{4p}$.

A common way of constructing or generating real normal (a.k.a. Gaussian) random vectors $\boldsymbol{x}_0 \sim \mathcal{N}^{\mathbb{R}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is to first generate a vector of independent and identically distributed (i.i.d.) standard normal random variables \boldsymbol{x} and then take the linear (or affine) combination $\boldsymbol{x}_0 = \boldsymbol{T}\boldsymbol{x} + \boldsymbol{\mu}$, where we factor $\boldsymbol{\Sigma} = \boldsymbol{T}\boldsymbol{T}^T$. The most straightforward way of constructing a standard complex normal random variable $z \sim \mathcal{N}^{\mathbb{C}}(0, 1)$ is to set $z = \frac{1}{\sqrt{2}} (x + yi)$ where x and y are i.i.d. real standard normal ($\mathcal{N}^{\mathbb{R}}(0, 1)$) random variables.⁴ We can then construct a general (proper) complex normal random vector $\boldsymbol{z}_0 \sim \mathcal{N}^{\mathbb{C}}(\boldsymbol{\mu}, \boldsymbol{C}\boldsymbol{C}^H)$ by taking

$$oldsymbol{z}_0 = oldsymbol{C}oldsymbol{z} + oldsymbol{\mu},$$

where \boldsymbol{z} is a vector of independent $\mathcal{N}^{\mathbb{C}}(0,1)$ random variables, \boldsymbol{C} is a complex matrix and $\boldsymbol{\mu}$ is a constant complex vector. This leads to the usual definition of a (proper) complex normal random vector (Goodman, 1963; Wooding, 1956).

The same process can be applied to quaternions. Define a standard quaternion normal $q \sim \mathcal{N}^{\mathbb{H}}(0,1)$ as $q = \frac{1}{2} (a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k})$ with a, b, c, d i.i.d. $\mathcal{N}^{\mathbb{R}}(0,1)$. Then we can

⁴The normalisation ensures that $\mathbb{E}\left[|z|^2\right] = 1$.

construct a general (proper) quaternion normal random vector $q_0 \sim \mathcal{N}^{\mathbb{H}}(\mu, QQ^H)$ by taking

$$\boldsymbol{q}_0 = \boldsymbol{Q} \boldsymbol{q} + \boldsymbol{\mu},$$

where \boldsymbol{q} is a vector of i.i.d. $\mathcal{N}^{\mathbb{H}}(0,1)$ random variables, \boldsymbol{Q} is a quaternion matrix and $\boldsymbol{\mu}$ is a constant quaternion vector. Any target covariance matrix $\boldsymbol{\Sigma}$ can be written as $\boldsymbol{\Sigma} = \boldsymbol{Q}\boldsymbol{Q}^{H}$ by Theorem 1.68. This desired link between quaternion (right-)linear transformations and quaternion propriety is obtained by using the following definition.

Definition 2.3.

Let μ ∈ ℝ^p, Σ ∈ ℝ^{p×p} symmetric positive definite. The real p-dimensional normal distribution N^ℝ(μ, Σ) has density

$$f_{\mathcal{N}^{\mathbb{R}}(\boldsymbol{\mu},\boldsymbol{\Sigma})}(\boldsymbol{x}) = (2\pi)^{-\frac{p}{2}} |\boldsymbol{\Sigma}|_{\mathbb{C}}^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{T}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)$$

 Let μ ∈ C^p, Σ ∈ C^{p×p} Hermitian positive definite. The (proper) complex pdimensional normal distribution N^C(μ, Σ) has density (Goodman, 1963, eqn. (1.5))

$$f_{\mathcal{N}^{\mathbb{C}}(\boldsymbol{\mu},\boldsymbol{\Sigma})}(\boldsymbol{x}) = \pi^{-p} \left| \boldsymbol{\Sigma} \right|_{\mathbb{C}}^{-1} \exp \left(-(\boldsymbol{x}-\boldsymbol{\mu})^{H} \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})
ight).$$

Let μ ∈ ℍ^p, Σ ∈ ℍ^{p×p} QHPD. The (right proper) quaternion p-dimensional normal distribution N^ℍ(μ, Σ) has density

$$f_{\mathcal{N}^{\mathbb{H}}(\boldsymbol{\mu},\boldsymbol{\Sigma})}(\boldsymbol{x}) = \left(\frac{2}{\pi}\right)^{2p} |\boldsymbol{\Sigma}|^{-2} \exp\left(-2(\boldsymbol{x}-\boldsymbol{\mu})^{H}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right).$$
(2.6)

Proposition 2.4. \boldsymbol{q} is distributed as $\mathcal{N}^{\mathbb{H}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ iff $\mathcal{V}(\boldsymbol{q})$ is distributed as $\mathcal{N}^{\mathbb{R}}\left(\mathcal{V}(\boldsymbol{\mu}), \frac{1}{4}\widetilde{\boldsymbol{\Sigma}}\right)$.

Proof. In (2.6), the term in the exponential is real, so we can apply (1.9). Finally, $|\Sigma|^2 = 4^{2p} \left| \frac{1}{4} \widetilde{\Sigma} \right|_{\mathbb{C}}^{\frac{1}{2}}$ and the real dimensionality is $p_0 = 4p$.

Corollary 2.5. Let $q \in \mathbb{H}^n$, $q \sim \mathcal{N}^{\mathbb{H}}(\mu, \Sigma)$ and $M \in \mathbb{H}^{m \times n}$. Then

$$Mq \sim \mathcal{N}^{\mathbb{H}}(M\mu, M\Sigma M^{H}).$$

Proof. From Proposition 2.4,
$$\mathcal{V}(\boldsymbol{q}) \sim \mathcal{N}^{\mathbb{R}}\left(\mathcal{V}(\boldsymbol{\mu}), \frac{1}{4}\widetilde{\Sigma}\right)$$
. From the real case $\widetilde{\boldsymbol{M}} \mathcal{V}(\boldsymbol{q}) \sim \mathcal{N}^{\mathbb{R}}\left(\widetilde{\boldsymbol{M}} \mathcal{V}(\boldsymbol{\mu}), \frac{1}{4}\widetilde{\boldsymbol{M}}\widetilde{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{M}}^{T}\right) = \mathcal{N}^{\mathbb{R}}\left(\mathcal{V}(\boldsymbol{M}\boldsymbol{\mu}), \frac{1}{4}(\widetilde{\boldsymbol{M}}\widetilde{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{M}}^{H})\right)$. Apply Proposition 2.4.

Remark 2.6. For any unit quaternion $u, q \sim \mathcal{N}^{\mathbb{H}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ iff $qu \sim \mathcal{N}^{\mathbb{H}}(\boldsymbol{\mu}u, \boldsymbol{\Sigma})$.

Proof. This can be shown by applying a change of variables $\boldsymbol{y} = \boldsymbol{x}u$ to (2.6),⁵ or through Proposition 2.4 by applying Lemma 1.38 to the covariance matrix in the real domain.

This invariance fully characterises quaternion propriety (Vía et al., 2010a, Lemma 9). Since multiplication on the left by a quaternion is a special case of Corollary 2.5, a general 4D rotation $\boldsymbol{q} \mapsto v\boldsymbol{q}u$, with u,v unit quaternions gives us $v\boldsymbol{q}u \sim \mathcal{N}^{\mathbb{H}}(v\boldsymbol{\mu}u, v\boldsymbol{\Sigma}\bar{v})$. Hence propriety is not a basis-dependent notion and in particular the basis element 1 plays no special role. This generalises the invariance under rotations $\boldsymbol{q} \mapsto \boldsymbol{u}\boldsymbol{q}\bar{\boldsymbol{u}}$ of the 3D space of pure imaginary quaternions shown by Vía et al. (2010a).

Proposition 2.7.

 Let μ ∈ ℝ^p, Σ ∈ ℝ^{p×p} symmetric positive definite. The real p-dimensional normal distribution N^ℝ(μ, Σ) has characteristic function

$$\phi_{\mathcal{N}^{\mathbb{R}}(\boldsymbol{\mu},\boldsymbol{\Sigma})}(\boldsymbol{\theta}) = \exp\left(\boldsymbol{\theta}^{T}\boldsymbol{\mu}\mathbf{i} - \frac{1}{2}\boldsymbol{\theta}^{T}\boldsymbol{\Sigma}\boldsymbol{\theta}\right).$$
(2.7)

 Let μ ∈ C^p, Σ ∈ C^{p×p} Hermitian positive definite. The (proper) complex pdimensional normal distribution N^C(μ, Σ) has characteristic function (Wooding, 1956, eqn. (20))

$$\phi_{\mathcal{N}^{\mathbb{C}}(\boldsymbol{\mu},\boldsymbol{\Sigma})}(\boldsymbol{\theta}) = \exp\left(\Re(\boldsymbol{\theta}^{H}\boldsymbol{\mu})\mathrm{i} - \frac{1}{4}\boldsymbol{\theta}^{H}\boldsymbol{\Sigma}\boldsymbol{\theta}
ight)$$

⁵And noting that $(\boldsymbol{y}\bar{u}-\boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}\bar{u}-\boldsymbol{\mu}) = u(\boldsymbol{y}-\boldsymbol{\mu}u)^H \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}-\boldsymbol{\mu}u)\bar{u} = u\bar{u}(\boldsymbol{y}-\boldsymbol{\mu}u)^H \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}-\boldsymbol{\mu}u).$

 Let μ ∈ ℍ^p, Σ ∈ ℍ^{p×p} QHPD. The (right proper) quaternion p-dimensional normal distribution N^ℍ(μ, Σ) has characteristic function

$$\phi_{\mathcal{N}^{\mathbb{H}}(\boldsymbol{\mu},\boldsymbol{\Sigma})}(\boldsymbol{\theta}) = \exp\left(\Re(\boldsymbol{\theta}^{H}\boldsymbol{\mu})\mathbf{i} - \frac{1}{8}\boldsymbol{\theta}^{H}\boldsymbol{\Sigma}\boldsymbol{\theta}\right).$$
(2.8)

Proof. We will only prove (2.8). By Proposition 2.2 and Proposition 2.4

$$\begin{split} \phi_{\mathcal{N}^{\mathbb{H}}(\boldsymbol{\mu},\boldsymbol{\Sigma})}(\boldsymbol{\theta}) &= \phi_{\mathcal{N}^{\mathbb{R}}\left(\mathcal{V}(\boldsymbol{\mu}),\frac{1}{4}\widetilde{\boldsymbol{\Sigma}}\right)}(\mathcal{V}(\boldsymbol{\theta})) \\ &= \exp\left(\mathcal{V}(\boldsymbol{\theta})^{T} \, \mathcal{V}(\boldsymbol{q}) \mathbf{i} - \frac{1}{8} \, \mathcal{V}(\boldsymbol{\theta})^{T} \widetilde{\boldsymbol{\Sigma}} \, \mathcal{V}(\boldsymbol{\theta})\right) \\ &= \exp\left(\Re(\boldsymbol{\theta}^{H} \boldsymbol{\mu}) \mathbf{i} - \frac{1}{8} \Re\left(\boldsymbol{\theta}^{H} \boldsymbol{\Sigma} \boldsymbol{\theta}\right)\right) \end{split}$$

where we used (2.7), (1.10) and (1.9). Finally, note that since Σ is Hermitian, $\theta^H \Sigma \theta$ is real.

The definition of quaternion normal distribution used by Andersson (1975) and implicitly by Andersson et al. (1983); Møller (1986) is given in terms of $\mathcal{V}(\boldsymbol{q})$. Hence the covariance parameter they choose is $\frac{1}{4}\widetilde{\Sigma}$. Kabe (1984) however chooses $\frac{1}{8}\widetilde{\Sigma}$ as a parameter instead.⁶ Both these definitions are equivalent to ours by Proposition 2.4. Vakhania (1999) uses a *left* quaternion normal distribution. The difference flows from their choice to treat \boldsymbol{H}^n (and more general quaternion Hilbert spaces) as a quaternion left module.⁷ The characteristic function of this left quaternion normal distribution is given as

$$\phi_{\mathcal{N}^{\mathbb{H}\mathrm{Left}}(\boldsymbol{\mu},\boldsymbol{\Sigma})}(\boldsymbol{\theta}) = \exp\left(\Re(\boldsymbol{\theta}^{H}\boldsymbol{\mu})\mathrm{i} - \frac{1}{8}\boldsymbol{\theta}^{T}\boldsymbol{\Sigma}\bar{\boldsymbol{\theta}}\right).$$
(2.9)

The right and left normal theories are equivalent, since the conjugate operator $\overline{\bullet}$ is an isomorphism between the right and left modules, as we explained in Section 1.3.2. This should allow a careful reader to use the theory of quaternion distributions on

⁶We find this choice rather unusual, as is their choice to parameterise the multivariate real normal distribution in terms of 2Σ instead of Σ .

⁷The "scalar product" for the left module of quaternion column vectors is then given by $\langle \boldsymbol{q}, \boldsymbol{q}_0 \rangle = \boldsymbol{q}^T \bar{\boldsymbol{q}}_0$.

Hilbert spaces developed by Vakhania (1999) under either convention.

Proposition 2.8. q follows a (right) quaternion normal distribution $\mathcal{N}^{\mathbb{H}}(\mu, \Sigma)$ iff \bar{q} follows a left quaternion normal distribution $\mathcal{N}^{\mathbb{HLeft}}(\bar{\mu}, \Sigma)$.

Proof. From (1.20) and (1.19), $\Re(\boldsymbol{\theta}^T \boldsymbol{q}) = \Re(\boldsymbol{q}^H \bar{\boldsymbol{\theta}}) = \Re \operatorname{tr}(\bar{\boldsymbol{q}} \boldsymbol{\theta}^H) = \Re(\boldsymbol{\theta}^H \bar{\boldsymbol{q}})$. Hence $\phi_{\boldsymbol{q}}(\bar{\boldsymbol{\theta}}) = \phi_{\bar{\boldsymbol{q}}}(\boldsymbol{\theta})$. Comparing (2.8) and (2.9), $\phi_{\mathcal{N}^{\mathbb{H}}(\boldsymbol{\mu},\boldsymbol{\Sigma})}(\boldsymbol{\theta}) = \phi_{\mathcal{N}^{\mathbb{H}}(\boldsymbol{\mu},\boldsymbol{\Sigma})}(\bar{\boldsymbol{\theta}}) = \phi_{\mathcal{N}^{\mathbb{HLeft}}(\bar{\boldsymbol{\mu}},\boldsymbol{\Sigma})}(\boldsymbol{\theta})$.

Cheong Took and Mandic (2011) also consider the quaternion proper normal distribution. They find that a p dimensional quaternion normal random vector is equivalent to a 4p dimensional spherical real normal random vector, i.e. all components are independent with equal variance. However, this result is incorrect for p > 1. This is however the correct characterisation for quaternion normal random vectors which are simultaneously right and left proper, implying 4D rotation invariance (see Appendix A.1).

We believe that our definition and parameterisation for the quaternion normal distribution are the most consistent with the usual complex normal distribution. It also allows us to write $\Sigma = \mathbb{E}\left[(\boldsymbol{q} - \boldsymbol{\mu})(\boldsymbol{q} - \boldsymbol{\mu})^H\right]$ for $\boldsymbol{q} \sim \mathcal{N}^{\mathbb{H}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, so the covariance parameter has its usual interpretation.

2.4 Wishart distribution

We are interested in the real (resp. complex/quaternion) distribution, denoted $\mathcal{W}_p^{\mathbb{R}/\mathbb{C}/\mathbb{H}}$, of

$$\boldsymbol{W} = \sum_{i=1}^{N} \boldsymbol{v}_i \boldsymbol{v}_i^H, \qquad (2.10)$$

where the v_i are N i.i.d. samples from a $\mathcal{N}^{\mathbb{R}/\mathbb{C}/\mathbb{H}}(\mathbf{0}, \Sigma)$ distribution.

Remark 2.9. We will be assuming in this section that the samples \boldsymbol{v}_i have mean zero. If the mean is known, we may subtract it without loss of generality. If the mean is unknown, let $\hat{\boldsymbol{\mu}} = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{v}_i$. Then we have instead $\boldsymbol{W} = \sum_{i=1}^{N} (\boldsymbol{v}_i - \hat{\boldsymbol{\mu}}) (\boldsymbol{v}_i - \hat{\boldsymbol{\mu}})^H \sim \mathcal{W}_p^{\mathbb{R}/\mathbb{C}/\mathbb{H}}(\boldsymbol{\Sigma}, N-1)$. Again there is no loss of generality. Remark 2.10. The real/complex/quaternion sample covariance matrix is also Wishart distributed. If $\boldsymbol{W} \sim \mathcal{W}_p^{\mathbb{R}/\mathbb{C}/\mathbb{H}}(\boldsymbol{\Sigma}, N)$, then $\hat{\boldsymbol{\Sigma}} = \frac{1}{N} \boldsymbol{W} \sim \mathcal{W}_p^{\mathbb{R}/\mathbb{C}/\mathbb{H}}\left(\frac{1}{N}\boldsymbol{\Sigma}, N\right)$.⁸

We will prove that the sample covariance matrices considered here are the maximum likelihood estimators in Section 4.2.1.

Let $\boldsymbol{v}_i \sim \mathcal{N}^{\mathbb{H}}(\boldsymbol{0}, \boldsymbol{\Sigma})$ and let $\boldsymbol{S} = \sum_{i=1}^N \mathcal{V}(\boldsymbol{v}_i) \mathcal{V}(\boldsymbol{v}_i)^T \sim \mathcal{W}_{4p}^{\mathbb{R}}\left(\frac{1}{4}\widetilde{\boldsymbol{\Sigma}}, N\right)$. Then, as we have shown in Section 1.3.3, we can interpret the quaternion product(s) in (2.10) as a projection, so that $\widetilde{\boldsymbol{W}} = 4\hat{h}(\boldsymbol{S})$. We can also interpret $\widetilde{\boldsymbol{W}}$ as an ensemble average of 4 real wishart matrices, obtained from the 4 ensembles of N samples $\mathcal{V}(2\boldsymbol{v}_i), \mathcal{V}(2\boldsymbol{v}_i\mathbf{i}), \mathcal{V}(2\boldsymbol{v}_i\mathbf{j}), \mathcal{V}(2\boldsymbol{v}_i\mathbf{k}).$

Definition 2.11. \boldsymbol{W} follows a Wishart distribution $\boldsymbol{W} \sim \mathcal{W}_p^{\mathbb{H}}(\boldsymbol{\Sigma}, N)$ iff $\widetilde{\boldsymbol{W}} = 4\hat{h}(\boldsymbol{S})$ for some $\boldsymbol{S} \sim \mathcal{W}_{4p}^{\mathbb{R}}\left(\frac{1}{4}\widetilde{\boldsymbol{\Sigma}}, N\right)$.

Remark 2.12. As with the real Wishart distribution, the quaternion Wishart distribution can be defined for non-integer N.

Proposition 2.13.

 Let Σ ∈ ℝ^{p×p} symmetric positive definite. The characteristic function of the (real) Wishart distribution is (Muirhead, 1982, Theorem 3.2.3)⁹

$$\phi_{\mathcal{W}_p^{\mathbb{R}}(\boldsymbol{\Sigma},N)}(\boldsymbol{\Theta}) = |\boldsymbol{I}_p - 2\mathrm{i}\boldsymbol{\Sigma}\boldsymbol{\Theta}|_{\mathbb{C}}^{-\frac{N}{2}}.$$
(2.11)

Let Σ ∈ C^{p×p} Hermitian positive definite. The characteristic function of the complex Wishart distribution is (Goodman, 1963)

$$\phi_{\mathcal{W}_p^{\mathbb{C}}(\boldsymbol{\Sigma},N)}(\boldsymbol{\Theta}) = |\boldsymbol{I}_p - \mathrm{i}\boldsymbol{\Sigma}\boldsymbol{\Theta}|_{\mathbb{C}}^{-N}$$

• Let $\Sigma \in \mathbb{H}^{p \times p}$ QHPD. The characteristic function of the quaternion Wishart distribution is

$$\phi_{\mathcal{W}_{p}^{\mathbb{H}}(\boldsymbol{\Sigma},N)}(\boldsymbol{\Theta}) = \left| \boldsymbol{I}_{4p} - \frac{\mathrm{i}}{2} \widetilde{\boldsymbol{\Sigma}} \widetilde{\boldsymbol{\Theta}} \right|_{\mathbb{C}}^{-\frac{N}{2}}.$$
 (2.12)

⁸In particular, if $\boldsymbol{W} \sim \mathcal{W}_p^{\mathbb{R}/\mathbb{C}/\mathbb{H}}(\boldsymbol{\Sigma}, N)$ then $\mathbb{E}[\boldsymbol{W}] = N\boldsymbol{\Sigma}$ so that $\mathbb{E}\left[\frac{1}{N}\boldsymbol{W}\right] = \boldsymbol{\Sigma}$

⁹Where Γ is replaced by 2 Θ , due to our different way of defining the characteristic function for a symmetric matrix (see footnote 3).

Note that since the random matrices being considered are symmetric/Hermitian, we restrict ourselves to Θ symmetric/Hermitian.

Proof. We will only prove (2.12). Let $\boldsymbol{W} \sim \mathcal{W}_p^{\mathbb{H}}(\boldsymbol{\Sigma}, N)$ then $\widetilde{\boldsymbol{W}} = 4\hat{h}(\boldsymbol{S})$ where $\boldsymbol{S} \sim \mathcal{W}_{4p}^{\mathbb{R}}\left(\frac{1}{4}\widetilde{\boldsymbol{\Sigma}}, N\right)$. So using (2.1), followed by Lemma 1.46 and (2.11).

$$\begin{split} \phi_{\mathcal{W}_{p}^{\mathbb{H}}(\boldsymbol{\Sigma},N)}(\boldsymbol{\Theta}) &= \phi_{\widetilde{W}}\left(\frac{1}{4}\widetilde{\boldsymbol{\Theta}}\right) \\ &= \mathbb{E}\left[\mathrm{e}^{\mathrm{i}\operatorname{tr}\left(\widetilde{\boldsymbol{\Theta}}^{T}\frac{1}{4}\widetilde{\boldsymbol{W}}\right)}\right] \\ &= \mathbb{E}\left[\mathrm{e}^{\mathrm{i}\operatorname{tr}\left(\widetilde{\boldsymbol{\Theta}}^{T}\hat{h}(\boldsymbol{S})\right)}\right] \\ &= \mathbb{E}\left[\mathrm{e}^{\mathrm{i}\operatorname{tr}\left(\widetilde{\boldsymbol{\Theta}}^{T}\boldsymbol{S}\right)}\right] \\ &= \phi_{\mathcal{W}_{4p}^{\mathbb{R}}\left(\frac{1}{4}\widetilde{\boldsymbol{\Sigma}},N\right)}(\widetilde{\boldsymbol{\Theta}}) \\ &= \left|\boldsymbol{I}_{4p} - \frac{\mathrm{i}}{2}\widetilde{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{\Theta}}\right|_{\mathbb{C}}^{-\frac{N}{2}} \end{split}$$

Note that the critical result used in the above proof is that \hat{h} is an orthogonal projection (through Lemma 1.46). The proof can be thus generalised to any structured random matrix that can be constructed as the orthogonal projection of a real random matrix with known characteristic function. This includes for example the complex Wishart distribution, allowing a simpler proof than that of Goodman (1963). More generally, as can be seen from Jensen (1988, p. 304), the maximum likelihood covariance estimators for multivariate normal distributions with structured covariances will fall into this category when the structure is linear for both the covariance and inverse covariance.

Remark 2.14. The characteristic function of the improper quaternion Wishart distribution¹⁰ can be obtained by simply replacing $\frac{1}{4}\widetilde{\Sigma}$ in (2.12) with the covariance matrix of $\mathcal{V}(\boldsymbol{q})$.

Theorem 2.15. Assume N > p - 1.

¹⁰The distribution of the quaternion sample covariance matrix of an improper quaternion normal sample.

Let Σ ∈ ℝ^{p×p} symmetric positive definite. The (real) Wishart density (Wishart, 1928) is

$$f_{\mathcal{W}_p^{\mathbb{R}}(\boldsymbol{\Sigma},N)}(\boldsymbol{W}) = C_1 |\boldsymbol{\Sigma}|^{-\frac{N}{2}} |\boldsymbol{W}|^{\frac{N-p-1}{2}} \exp\left(-\frac{1}{2} \operatorname{tr}\left(\boldsymbol{\Sigma}^{-1} \boldsymbol{W}\right)\right)$$

Let Σ ∈ C^{p×p} Hermitian positive definite. The complex Wishart density (Goodman, 1963) is

$$f_{\mathcal{W}_p^{\mathbb{C}}(\boldsymbol{\Sigma},N)}(\boldsymbol{W}) = C_2 |\boldsymbol{\Sigma}|^{-N} |\boldsymbol{W}|^{N-p} \exp\left(-\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1} \boldsymbol{W}\right)\right).$$

• Let $\Sigma \in \mathbb{H}^{p \times p}$ QHPD. The quaternion Wishart density is

$$f_{\mathcal{W}_p^{\mathbb{H}}(\boldsymbol{\Sigma},N)}(\boldsymbol{W}) = C_3 |\boldsymbol{\Sigma}|^{-2N} |\boldsymbol{W}|^{2N-2p+1} \exp\left(-2\Re \operatorname{tr}\left(\boldsymbol{\Sigma}^{-1} \boldsymbol{W}\right)\right).$$
(2.13)

The normalisation constants are

$$C_{1} = \left(2^{\frac{Np}{2}}\pi^{\frac{p(p-1)}{4}}\prod_{m=1}^{p}\Gamma\left(\frac{N+1-m}{2}\right)\right)^{-1},$$

$$C_{2} = \left(\pi^{\frac{p(p-1)}{2}}\prod_{m=1}^{p}\Gamma\left(N+1-m\right)\right)^{-1},$$

$$C_{3} = \left(2^{-2Np}\pi^{p(p-1)}\prod_{m=1}^{p}\Gamma\left(2(N+1-m)\right)\right)^{-1}.$$
(2.14)

Proof. We will only prove (2.13), (2.14). See Appendix B.1.

Note that for $p \leq N < 4p$, the quaternion sample covariance matrix has a density even though the corresponding unstructured real sample covariance matrix is singular and does not.

2.4.1 A review of literature related to the quaternion Wishart distribution

Kabe (1984, eq. (10)) gives the quaternion Wishart density (2.13), with $\Sigma/2$ as a

covariance parameter, and sketches a proof¹¹ using hypercomplex matrix calculus. Andersson et al. (1983) describe it in terms of $\widetilde{\Sigma}$, \widetilde{W} with respect to the group invariant measure $|W|^{-2p+1} dW$, and give a proof using abstract results on group invariance. Møller (1986) gives the density in terms of $\widetilde{\Sigma}$ and the sample covariance matrix $\frac{1}{N}\widetilde{W}$, but provides no proof.

The work of Loots et al. (2012) is very similar to ours.¹² They choose the same definition for the quaternion normal distribution and also make use of the real representation to derive the quaternion Wishart characteristic function and density.¹³ Their Fourier inversion of the characteristic function relies on a series expansion in zonal polynomials and on the hypergeometric function of a quaternion matrix argument. We believe that their derivation of the characteristic function relies on a confusion between $\frac{1}{4}\widetilde{W}$ and S due to notation.

Li and Xue (2010) derive the density of the quaternion Wishart distribution in the very general singular non-central¹⁴ case.¹⁵ Again, the derivation is based on the hypergeometric function of a quaternion matrix argument. It is worth noting that their definition of a quaternion normal (column) vector corresponds to our (right) quaternion normal distribution, but in their definition of a quaternion normal matrix, the rows are *left* quaternion normal.¹⁶ Nonetheless, their quaternion Wishart distribution still reduces to ours in the central non-singular case because the position of the conjugate transpose is swapped in their version of (2.10).

¹¹More specifically, it simply refers the reader to the proof of the complex case in Khatri (1965).

 $^{^{12}}$ We would like to point out that the author's related work in this chapter dates from his 2011 transfer report and thus predates Loots et al. (2012).

¹³In both cases their findings agree with ours, with their Σ_0^* corresponding to our $\frac{1}{4}\widetilde{\Sigma}$.

¹⁴The underlying multivariate normal samples may have a singular covariance matrix, they may be correlated (with a known possibly singular real covariance which is corrected for when constructing the Wishart matrix), and may have (possibly different) non-zero means (which are not corrected for).

¹⁵They also consider the related quaternion matrix-valued F and beta distributions.

¹⁶The columns are restricted to having a real-valued quaternion covariance matrix. In other words, their real vector version is real block spherical. Thus it is the rows, not the columns, which are to be interpreted as individual samples.

Andersson (1975) shows that covariance models defined by group invariance (under an arbitrary subgroup of the group of orthogonal matrices) can be constructed using just models with real, complex or quaternion structure, as in Corollary 1.19. Thus the corresponding Wishart distributions can be mapped under an appropriate isomorphism to collections of independent real, complex and quaternion Wishart distributions. Jensen (1988) extends the classification result of Andersson (1975) to include any linear constraints which translate to linear constraints on the inverse covariance matrix. In that case the decomposition may include simple Jordan algebras of degree 2⁻¹⁷ in addition to real complex and quaternion matrix algebras. Another generalisation, considered by Käuff (2012) (and references therein), combines group invariance with graphical models and gives the corresponding Wishart distribution as a generalised Riesz distribution. Andersson and Wojnar (2004); Wojnar (1999) consider an even more general Wishart distribution, which can be applied to any set of "covariance matrices" parametrised by an open proper convex homogeneous cone.

2.5 Improper normal distribution

For ease of exposition, we will again assume in this section that all normal random vectors are zero-mean. All results can however be easily generalised.

Definition 2.16. We introduce the following notation for covariances

$$egin{aligned} oldsymbol{\Sigma}_{oldsymbol{x},oldsymbol{y}} &= \mathbb{E}\left[oldsymbol{x}oldsymbol{y}^H
ight] \ oldsymbol{\Sigma}_{oldsymbol{x}} &= oldsymbol{\Sigma}_{oldsymbol{x},oldsymbol{x}} &= \mathbb{E}\left[oldsymbol{x}oldsymbol{x}^H
ight]. \end{aligned}$$

In particular, if \boldsymbol{x} and \boldsymbol{y} are real

$$egin{aligned} & \mathbf{\Sigma}_{m{x},m{y}} = \mathbb{E}\left[m{x}m{y}^T
ight] \ & \mathbf{\Sigma}_{m{x}} = \mathbf{\Sigma}_{m{x},m{x}} = \mathbb{E}\left[m{x}m{x}^T
ight] \end{aligned}$$

¹⁷Up to isomorphism a simple Jordan algebra of degree 2 and dimension n has the basis $1, e_1, \ldots, e_{n-1}$ with multiplication $e_i e_j = \delta_{i,j}$

A general 2*p*-dimensional real normal distribution can be expressed as an improper *p*-dimensional complex normal distribution (Schreier, 2010). The distribution of an improper complex normal random vector \boldsymbol{z} depends not only on the complex covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{z},\boldsymbol{\bar{z}}}$, but also on the complementary covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{z},\boldsymbol{\bar{z}}} = \mathbb{E} [\boldsymbol{z}\boldsymbol{z}^T]$. The distribution is proper iff $\boldsymbol{\Sigma}_{\boldsymbol{z},\boldsymbol{\bar{z}}} = \boldsymbol{0}$, i.e. when \boldsymbol{z} and $\boldsymbol{\bar{z}}$ are uncorrelated. A similar approach is possible for quaternions.

Let $a, b, c, d \in \mathbb{R}^p$ be jointly normal real random vectors, and q = a + bi + cj + dk. We can express the arbitrary 4*p*-dimensional real normal distribution of $\mathcal{V}(q)$ by an improper quaternion normal distribution. Define the augmented vector

$$\underline{\boldsymbol{q}} = \begin{pmatrix} \boldsymbol{q} \\ \boldsymbol{q}^{(i)} \\ \boldsymbol{q}^{(j)} \\ \boldsymbol{q}^{(k)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{q} \\ -\mathrm{i}\boldsymbol{q}\mathrm{i} \\ -\mathrm{j}\boldsymbol{q}\mathrm{j} \\ -\mathrm{k}\boldsymbol{q}\mathrm{k} \end{pmatrix} = \mathfrak{A}_{p} \mathcal{V}(\boldsymbol{q}), \qquad (2.15)$$

where

$$\mathfrak{A}_p = \left(egin{array}{cccc} oldsymbol{I}_p & \mathrm{i}oldsymbol{I}_p & \mathrm{j}oldsymbol{I}_p & \mathrm{k}oldsymbol{I}_p \ oldsymbol{I}_p & \mathrm{i}oldsymbol{I}_p & -\mathrm{j}oldsymbol{I}_p & -\mathrm{k}oldsymbol{I}_p \ oldsymbol{I}_p & -\mathrm{i}oldsymbol{I}_p & \mathrm{j}oldsymbol{I}_p & -\mathrm{k}oldsymbol{I}_p \ oldsymbol{I}_p & -\mathrm{i}oldsymbol{I}_p & -\mathrm{j}oldsymbol{I}_p & \mathrm{k}oldsymbol{I}_p \end{array}
ight)$$

This matrix satisfies

$$\mathfrak{A}_p\mathfrak{A}_p^H=4\boldsymbol{I}_{4p}$$

Hence we have

$$\Sigma_{\boldsymbol{q}} = \mathfrak{A}_{p} \Sigma_{\mathcal{V}(\boldsymbol{q})} \mathfrak{A}_{p}^{H}$$
(2.16)

$$\Sigma_{\mathcal{V}(\boldsymbol{q})} = \frac{1}{16} \mathfrak{A}_p^H \Sigma_{\boldsymbol{q}} \mathfrak{A}_p.$$
 (2.17)

Remark 2.17. Other definitions are possible for the augmented quaternion vector \underline{q} .

For example, Cheong Took and Mandic (2010b) use

$$\left(egin{array}{c} m{q} \ ar{m{q}} \ ar{m{q}} \ ar{m{q}}^{(\mathrm{i})} \ ar{m{q}}^{(\mathrm{j})} \end{array}
ight).$$

However, we believe that (2.15), which agrees with Vía et al. (2010a), is the most elegant choice.

Note that for a pure imaginary unit quaternion η ,

$$\boldsymbol{\Sigma}_{\boldsymbol{x},\boldsymbol{y}}^{(\eta)} = -\eta \mathbb{E}\left[\boldsymbol{x}\boldsymbol{y}^{H}\right] \eta = \eta \mathbb{E}\left[\boldsymbol{x}\eta\eta\boldsymbol{y}^{H}\right] \eta = \mathbb{E}\left[(-\eta\boldsymbol{x}\eta)(-\eta\boldsymbol{y}\eta)^{H}\right] = \boldsymbol{\Sigma}_{\boldsymbol{x}^{(\eta)},\boldsymbol{y}^{(\eta)}}.$$

Hence

$$\Sigma_{\underline{q}} = \begin{pmatrix} \Sigma_{q} & \Sigma_{q,q^{(i)}} & \Sigma_{q,q^{(j)}} & \Sigma_{q,q^{(k)}} \\ \Sigma_{q^{(i)},q} & \Sigma_{q^{(i)},q^{(i)}} & \Sigma_{q^{(i)},q^{(j)}} & \Sigma_{q^{(i)},q^{(k)}} \\ \Sigma_{q^{(j)},q} & \Sigma_{q^{(j)},q^{(i)}} & \Sigma_{q^{(j)},q^{(j)}} & \Sigma_{q^{(j)},q^{(k)}} \\ \Sigma_{q^{(k)},q} & \Sigma_{q^{(k)},q^{(i)}} & \Sigma_{q^{(k)},q^{(j)}} & \Sigma_{q^{(k)},q^{(k)}} \end{pmatrix} \\ = \begin{pmatrix} \Sigma_{q} & \Sigma_{q,q^{(i)}} & \Sigma_{q,q^{(j)}} & \Sigma_{q,q^{(k)}} \\ \Sigma_{q,q^{(i)}} & \Sigma_{q}^{(i)} & \Sigma_{q,q^{(k)}} & \Sigma_{q,q^{(i)}} \\ \Sigma_{q,q^{(i)}}^{(j)} & \Sigma_{q,q^{(k)}}^{(j)} & \Sigma_{q,q^{(i)}} \\ \Sigma_{q,q^{(k)}}^{(k)} & \Sigma_{q,q^{(k)}}^{(j)} & \Sigma_{q,q^{(i)}}^{(j)} \\ \Sigma_{q,q^{(k)}}^{(k)} & \Sigma_{q,q^{(k)}}^{(k)} & \Sigma_{q,q^{(i)}}^{(k)} \end{pmatrix}. \quad (2.18)$$

Since all the blocks in (2.18) can be derived from the first row of blocks through involutions, the second-order properties of q (or equivalently $\mathcal{V}(q)$) can be described by specifying the covariance Σ_q along with three complementary covariance matrices

$$\begin{split} \boldsymbol{\Sigma}_{\boldsymbol{q},\boldsymbol{q}^{(\mathrm{i})}} &= -E\{\boldsymbol{q}\mathrm{i}\boldsymbol{q}^{H}\mathrm{i}\}\\ \boldsymbol{\Sigma}_{\boldsymbol{q},\boldsymbol{q}^{(\mathrm{j})}} &= -E\{\boldsymbol{q}\mathrm{j}\boldsymbol{q}^{H}\mathrm{j}\}\\ \boldsymbol{\Sigma}_{\boldsymbol{q},\boldsymbol{q}^{(\mathrm{k})}} &= -E\{\boldsymbol{q}\mathrm{k}\boldsymbol{q}^{H}\mathrm{k}\}. \end{split}$$

Proposition 2.18. The pdf for an improper quaternion normal distribution with mean **0** and augmented covariance matrix Σ_q is

$$f_{\mathcal{N}^{\mathbb{H}\mathrm{Improper}}(\mathbf{0},\boldsymbol{\Sigma}_{\underline{q}})}(\boldsymbol{q}) = \left(\frac{2}{\pi}\right)^{2p} \left|\boldsymbol{\Sigma}_{\underline{q}}\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\underline{\boldsymbol{q}}^{H}\boldsymbol{\Sigma}_{\underline{q}}^{-1}\underline{\boldsymbol{q}}\right).$$

Proof. Let $\boldsymbol{x} = \mathcal{V}(\boldsymbol{q})$ be a 4*p*-dimensional real normal random vector. From (2.15) and (2.16), $\underline{\boldsymbol{q}}^H \boldsymbol{\Sigma}_{\underline{\boldsymbol{q}}}^{-1} \underline{\boldsymbol{q}} = \boldsymbol{x}^T \mathfrak{A}_p^H \mathfrak{A}_p \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \mathfrak{A}_p^H \mathfrak{A}_p \boldsymbol{x} = \boldsymbol{x}^T \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \boldsymbol{x}$. Also, $\left| \boldsymbol{\Sigma}_{\underline{\boldsymbol{q}}} \right| = \left| \mathfrak{A}_p \boldsymbol{\Sigma}_{\boldsymbol{x}} \mathfrak{A}_p^H \right| = \left| \mathfrak{A}_p^H \mathfrak{A}_p \right| |\boldsymbol{\Sigma}_{\boldsymbol{x}}| = 4^{4p} \left| \boldsymbol{\Sigma}_{\boldsymbol{x}} \right|_{\mathbb{C}}$.

We would like to stress that the improper quaternion distributions are equivalent to conventional real normal distributions, and that quaternion widely-linear processing in \mathbb{H}^p is equivalent to conventional real-linear processing in \mathbb{R}^{4p} . The use of augmented quaternions and of the algebra isomorphism $\mathbf{M} \mapsto \frac{1}{4}\mathfrak{A}_p \mathbf{M}\mathfrak{A}_p^H$ between $\mathbb{R}^{4p\times 4p}$ and quaternion widely-linear transformations provides a notation which may be convenient and insightful when comparing propriety to impropriety or quaternion-linearity to real-linearity. This is because quaternion widely linear notation effectively separates real linear transformations into four orthogonal components $\mathbb{R}^{4m\times 4n} = \widetilde{\mathbb{H}^{m\times n}} + \widetilde{\mathbb{H}^{m\times n}} i \widetilde{\mathbf{I}_n}^R + \widetilde{\mathbb{H}^{m\times n}} j \widetilde{\mathbf{I}_n}^R + \widetilde{\mathbb{H}^{m\times n}} k \widetilde{\mathbf{I}_n}^{R_{18}}$ with meaningful interpretations in the quaternion domain. Although augmented quaternion approaches may for example aid in interpreting results obtained through real-linear processing when the quaternion-linear component is physically meaningful, we believe that in general it will be simpler to develop and use multichannel real-linear techniques, algorithms and results in the familiar real matrix (or tensor) domain.

For a graphical representation of the relationship between the representations behind the real structured and widely linear approaches, see the commutative diagram Figure 2.1.

¹⁸Here we are treating $\mathbb{R}^{4m \times 4n}$ as a vector space. Note that $\widetilde{\mathbb{H}^{m \times n}} \eta \widetilde{I_n} = \widetilde{\mathbb{H}^{m \times n}}$, so the use of involutions or quaternion scalar multiplication on the right is equivalent in this context.



Figure 2.1: Commutative diagram containing the structured real (top three rows) and augmented quaternion (fourth row) approaches to real/complex/quaternion linear algebra. Hooked arrows represent injective real (*-)algebra homomorphisms, whilst arrows with ~ represent real (*-)algebra isomorphisms. Unlabeled arrows correspond to the identity function. $\tilde{\bullet}^{\mathbb{C}}$ denotes the complex representation (1.6).

2.6 Characterisations of propriety and second-order propriety

If we drop Gaussianity assumptions, we may still impose quaternion structure via second-order propriety.

Definition 2.19. A random vector $\boldsymbol{q} \in \mathbb{H}^n$ is second-order proper iff $\boldsymbol{\Sigma}_{\mathcal{V}(\boldsymbol{q})} \in \widetilde{\mathbb{H}^{n \times n}}$.

We can generalise Proposition 2.4 to the second order properties of non-normal random vectors.

Lemma 2.20. Let $q \in \mathbb{H}^n$ be a quaternion random vector. Then

$$\mathbb{E}\left[\mathcal{V}(\boldsymbol{q})\right] = \mathcal{V}\left(\mathbb{E}\left[\boldsymbol{q}
ight]
ight)$$

and

$$\hat{h}\left(\boldsymbol{\Sigma}_{\mathcal{V}(\boldsymbol{q})}\right) = \frac{1}{4}\widetilde{\boldsymbol{\Sigma}_{\boldsymbol{q}}}.$$

In particular,

$$\boldsymbol{\Sigma}_{\mathcal{V}(\boldsymbol{q})} = \frac{1}{4} \widetilde{\boldsymbol{\Sigma}_{\boldsymbol{q}}}$$

iff q is second-order proper.

Proof. The first equation holds because \mathcal{V} is a linear operator. For the second note that $\tilde{\bullet}$ and \hat{h} are also linear operators, so that using (1.15)

$$\widetilde{\boldsymbol{\Sigma}_{\boldsymbol{q}}} = \widetilde{\mathbb{E}[\boldsymbol{q}\boldsymbol{q}^{H}]} \\ = \mathbb{E}\left[\widetilde{\boldsymbol{q}\boldsymbol{q}^{H}}\right] \\ = \mathbb{E}\left[\widetilde{\boldsymbol{q}}\widetilde{\boldsymbol{q}}^{T}\right] \\ = \mathbb{E}\left[4\hat{h}(\mathcal{V}(\boldsymbol{q}) \ \mathcal{V}(\boldsymbol{q})^{T})\right] \\ = 4\hat{h}\left(\mathbb{E}\left[\mathcal{V}(\boldsymbol{q}) \ \mathcal{V}(\boldsymbol{q})^{T}\right]\right) \\ = 4\hat{h}\left(\boldsymbol{\Sigma}_{\mathcal{V}(\boldsymbol{q})}\right)$$

The last statement follows by Proposition 1.47.

The "meaning" of propriety can be summarised by the following proposition.

Proposition 2.21. Given a zero-mean improper p-dimensional quaternion normal random vector $\boldsymbol{q} \sim \mathcal{N}^{\mathbb{H}Improper}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_{\boldsymbol{q}})$, the following statements are equivalent

- 1. q is proper.
- 2. $\Sigma_{\mathcal{V}(q)} \in \widetilde{\mathbb{H}^{p \times p}}$. (i.e. q is second-order proper.)
- 3. $\Sigma_{q,q^{(\mathrm{i})}} = \Sigma_{q,q^{(\mathrm{j})}} = \Sigma_{q,q^{(\mathrm{k})}} = \mathbf{0}_{p \times p}$
- 4. There exist $M \in \mathbb{H}^{p \times p}$ and $s \sim \mathcal{N}^{\mathbb{H}}(\mathbf{0}_{1 \times p}, I_{p \times p})$ such that $q = \mu + Ms$.
- 5. For any unit quaternion $u, q \mu$ and $(q \mu)u$ are identically distributed.

Proof. 1. \Leftrightarrow 2. by Proposition 2.4.

 $3. \Rightarrow 2$. Is given in Vía et al. (2010a, Lemma 8). It can be shown directly by expanding (2.17).

 $2. \Rightarrow 3$. Can be shown by by expanding (2.16).

1. \Leftrightarrow 4. As we discussed in Section 2.3, this follows from Corollary 2.5 and Theorem 1.68 by taking the Cholesky decomposition $\Sigma_q = MM^H$.

3. \Leftrightarrow 5. Is equivalent to Vía et al. (2010a, Lemma 9), since u is a unit quaternion iff $u = e^{i_0 \theta}$ for some pure imaginary unit i_0 and some $\theta \in \mathbb{R}$ (see Example 1.15). Alternatively, 2. \Leftrightarrow 5. follows from Lemma 1.38.

Part 5. of Proposition 2.21 gives us a geometric characterisation of propriety to complement the constructive approach 4. and the structural approach 2.

Multiplying a real vector by a real scalar corresponds to scaling (and/or reflecting) the underlying space \mathbb{R} , so real-linear operators are invariant to such scaling. Multiplying a complex vector by a complex scalar rotates the underlying space \mathbb{C} in addition to scaling it. Complex-linear transformations and (zero-mean) proper complex distributions are precisely those that are invariant to such rotations. With quaternions the geometric interpretation is less intuitive. Quaternion right-linear transformations and (zero-mean) proper distributions are not invariant under all rotations $q \mapsto vqu$, but only the subgroup of 'right isoclinic' rotations $q \mapsto qu$. One way of thinking about isoclinic rotations is to treat quaternions as a pair of complex numbers through the Cayley-Dickson decomposition. Under an appropriate choice of the basis i, j, k of the set of pure imaginary quaternions, ¹⁹ the isoclinic rotation will rotate the two complex numbers by the same angle. The direction of rotation may be inverted depending on whether the rotation is right or left isoclinic, and whether we choose to decompose the quaternion as (a + bi) + (c + di)j or (a + bi) + j(c - di).

One might think that it would be more interesting to consider full rotation invariance instead of invariance under right isoclinic rotations. However, no 4D signal (or nD signal with $n \ge 3$) having correlated components can be invariant under general 4D (or nD) rotations. Indeed, the structure corresponding to full rotation invariance (for $n \ge 3$) is block sphericity. We show this in Appendix A.1.

¹⁹A basis rotation $q \mapsto uqu^{-1}$ is an algebra automorphism of \mathbb{H} and thus can be performed without loss of generality.

Chapter 3

The Quaternion Vector Autoregressive Model

3.1 Introduction

The vector autoregressive (VAR) model is the fundamental model of linear multiple time-series analysis. As noted by Lütkepohl (2006, p. 25), "Under quite general conditions, every stationary, purely nondeterministic process [a process minus its deterministic component] can be approximated well by a finite order VAR process."

Another approach to modelling vector time series has been to use scalar complex or quaternion AR processes. Complex-valued AR processes (Picinbono and Bondon, 1997) have been applied to temperature forecasting (Gu and Jiang, 2005), character recognition (Nakatani et al., 1999) and shape recognition and extraction (Sekita et al., 1992; Umeyama, 1997). A synthetic quaternion AR process was considered by Cheong Took and Mandic (2010b),¹ and adaptive (i.e. time-varying) quaternion AR filters have been applied to short-term wind forecasting (Cheong Took and Mandic, 2009) and hand orientation modelling (Jahanchahi et al., 2013).

Navarro-Moreno et al. (2013) study the problem of linear prediction for stationary quaternion-valued time-series. The method proposed corresponds to fitting an AR

¹The process chosen had real coefficients however, so the components can be interpreted as an ensemble of four independent realisations of a real AR process.

model by solving the Yule-Walker equations. These are the models which we will cover in this chapter. We will however choose to fit them through (forward) least squares, and generalise them by allowing for multiple quaternion time-series. Through the Yule-Walker equations, we will prove that VAR propriety defined through quaternion linearity and innovation propriety is equivalent to VAR propriety defined through the autocovariance matrix as in Chapter 2. This intuitive result underpins the validity of their approach in the proper case.

Because of the structure imposed by quaternion (or complex) linearity, quaternion (or complex) AR models are not appropriate for general vector signals. Widely-linear approaches have been suggested to overcome this limitation. For example, Navarro-Moreno et al. (2013) apply the AR model to 3D wind speed time-series and 4D wind speed and air temperature time-series prediction, and find that the widely-linear model outperforms the quaternion linear model. However, widely-linear VARs are just reformulations of equivalent unstructured real VAR approaches.²

The value of quaternion linear modelling in this context is not the capacity to write vector models as scalar models. It is a fourfold reduction in the number of real parameters to estimate, which improves efficiency when the assumption of quaternion linearity is (at least approximately) satisfied, and/or low sample size causes overfitting. This advantage persists when moving from scalar AR to VAR.

Baddour and Beaulieu (2002) simulate the fading of telecommunication signals using complex-valued VAR processes.³ Complex VARs also appear in the eigensystem VAR model (Krippner, 2010) whenever there are complex eigenvalues.⁴ We have not found any application of quaternion VARs in the literature, but we believe that it is a

²Some differences in interpretation may nevertheless appear when using an improper quaternion formulation over a real VAR formulation. For example, the size of updates in the usual real formulation of multivariate stochastic gradient descent is proportional to the estimation error in the relevant component. In the widely-linear quaternion formulation of Cheong Took and Mandic (2010b) however, the average estimation error across all components is used instead.

³The alternative real VAR approach is also proposed to allow for the improper case, avoiding widely-linear formulations.

⁴However, these complex processes are improper and singular. The imaginary parts can be computed deterministically from the real parts.

model worth treating, since Corollary 1.19 implies that real complex and quaternion VARs are the building blocks for AR models in arbitrary finite-dimensional real semisimple algebras, including Clifford algebras.

In this chapter, we define the proper quaternion vector autoregressive model, and show that using the real matrix representation it can be treated as a real VAR model, with quaternion structure assumptions imposed on the regression coefficients *and* the innovation covariance. Thus proper quaternion VAR modelling is a special case of real VAR modelling with linear structure constraints. A treatment of this general theory can be found in Lütkepohl (2006, Sections 4, 5 and 9), covering parameter estimation, asymptotic estimator distributions and hypothesis testing. We use this to develop a likelihood ratio test for quaternion propriety which combines the regression coefficient and innovation covariance assumptions.

For an unrestricted real VAR model generalised least squares (GLS) estimation reduces to least squares (LS) estimation. In other words, the efficiency of LS estimation is not degraded by anisotropic innovations. However, this is no longer true in general when constraints (like quaternion structure) are imposed on the coefficients. We prove that for proper quaternion VAR processes the result does hold. This implies that LS estimation gives the Gaussian maximum likelihood and best linear unbiased solution.

We will prove the equivalence between LS and GLS for a quaternion multivariate linear regression (MLR) — the linear regression for a proper quaternion AR process being a special case — and show that this requires an assumption of second-order (right-) propriety of the errors for a left-linear model, and vice-versa. This new optimality result is much stronger than the one given by Vía et al. (2010a).

Some of the material in this chapter was published in Ginzberg and Walden (2013b), see p. 11.

3.2 Quaternion multivariate linear regression

Proposition 3.1. Consider the standard real-valued (multiple) linear regression model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e} \tag{3.1}$$

with error covariance matrix Σ_e . If there exists a matrix S such that $\Sigma_e X = XS$, then the LS estimator

$$\hat{\boldsymbol{\beta}}^{\text{LS}} = \left(\boldsymbol{X}^T \boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \boldsymbol{y}$$
(3.2)

and the GLS estimator

$$\hat{\boldsymbol{\beta}}^{\text{GLS}} = \left(\boldsymbol{X}^T \boldsymbol{\Sigma}_{\boldsymbol{e}}^{-1} \boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \boldsymbol{\Sigma}_{\boldsymbol{e}}^{-1} \boldsymbol{y}$$
(3.3)

are equal. We assume for simplicity that $X^T X$, Σ_e and S are invertible.⁵

Proof. First note that

$$\boldsymbol{S}^T \boldsymbol{X}^T \boldsymbol{\Sigma}_{\boldsymbol{e}}^{-1} = (\boldsymbol{X} \boldsymbol{S})^T \boldsymbol{\Sigma}_{\boldsymbol{e}}^{-1} = (\boldsymbol{\Sigma}_{\boldsymbol{e}} \boldsymbol{X})^T \boldsymbol{\Sigma}_{\boldsymbol{e}}^{-1} = \boldsymbol{X}^T \boldsymbol{\Sigma}_{\boldsymbol{e}}^T \boldsymbol{\Sigma}_{\boldsymbol{e}}^{-1} = \boldsymbol{X}^T.$$

So we have

$$\hat{\boldsymbol{\beta}}^{\text{GLS}} = \left(\boldsymbol{X}^T \boldsymbol{\Sigma}_{\boldsymbol{e}}^{-1} \boldsymbol{X}\right)^{-1} \boldsymbol{S}^{T-1} \boldsymbol{S}^T \boldsymbol{X}^T \boldsymbol{\Sigma}_{\boldsymbol{e}}^{-1} \boldsymbol{y}$$

$$= \left(\boldsymbol{S}^T \boldsymbol{X}^T \boldsymbol{\Sigma}_{\boldsymbol{e}}^{-1} \boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \boldsymbol{y}$$

$$= \left(\boldsymbol{X}^T \boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \boldsymbol{y}$$

$$= \hat{\boldsymbol{\beta}}^{\text{LS}}.$$

Remark 3.2. Zyskind (1967) shows that the existence of S is both necessary and sufficient, and we can drop all invertibility assumptions in Proposition 3.1.

⁵Invertibility of $X^T X$ is equivalent to X having full rank (and dimension(y) \geq dimension(β)). $X^T \Sigma_e^{-1} X$ is then also invertible since rank(X) = rank($\Sigma_e^{-\frac{1}{2}} X$).

Proof. See Appendix B.2

Definition 3.3. Denote by vec : $\mathbb{R}^{m \times n} \to \mathbb{R}^{mn}$ the operator which stacks the columns of a matrix.

Lemma 3.4. Let \otimes denote the Kronecker product. Let $U \in \mathbb{R}^{n \times m}$, $V \in \mathbb{R}^{m \times \ell}$, and $P \in \mathbb{R}^{\ell \times k}$. Then

$$\operatorname{vec}(\boldsymbol{U}\boldsymbol{V}\boldsymbol{P}) = [\boldsymbol{P}^T \otimes \boldsymbol{U}]\operatorname{vec}(\boldsymbol{V}).$$
 (3.4)

Let $U \in \mathbb{R}^{n \times m}$, $V \in \mathbb{R}^{\ell \times k}$, $P \in \mathbb{R}^{m \times q}$, and $M \in \mathbb{R}^{k \times p}$, instead. Then

$$[\boldsymbol{U} \otimes \boldsymbol{V}][\boldsymbol{P} \otimes \boldsymbol{M}] = \boldsymbol{U}\boldsymbol{P} \otimes \boldsymbol{V}\boldsymbol{M}.$$
(3.5)

Proof. We can see (3.4) as a definition of \otimes , in which case (3.5) follows by the associativity of matrix multiplication. Alternatively, see e.g. Bernstein (2009, Propositions 7.1.6 & 7.1.9).

Remark 3.5. For $\lambda \in \mathbb{H}$, $\widetilde{\lambda I_n} = \widetilde{\lambda} \otimes I_n$ and $\widetilde{\lambda I_n}^R = \widetilde{\lambda}^R \otimes I_n$.

Lemma 3.6. Since $\tilde{\bullet} \circ \mathcal{V}^{-1} = \mathcal{V}^{-1}(\bullet) : \mathbb{R}^{4m \times n} \to \mathbb{R}^{4m \times 4n}$ is a real linear operator, we can write it in matrix form as

$$\operatorname{vec}(\widetilde{\boldsymbol{Q}}) = \boldsymbol{\Upsilon}^{(m \times n)} \operatorname{vec}\left(\mathcal{V}(\boldsymbol{Q})\right),$$

where

$$\boldsymbol{\Upsilon}^{(m\times n)} = \begin{pmatrix} \boldsymbol{I}_n \otimes \boldsymbol{I}_4 \otimes \boldsymbol{I}_m \\ \boldsymbol{I}_n \otimes \tilde{\mathbf{i}}^R \otimes \boldsymbol{I}_m \\ \boldsymbol{I}_n \otimes \tilde{\mathbf{j}}^R \otimes \boldsymbol{I}_m \\ \boldsymbol{I}_n \otimes \tilde{\mathbf{k}}^R \otimes \boldsymbol{I}_m \end{pmatrix} = \begin{pmatrix} \widetilde{\boldsymbol{I}}_n \\ \widetilde{\boldsymbol{I}}_n \tilde{\mathbf{i}}^R \\ \widetilde{\boldsymbol{I}}_n \tilde{\mathbf{j}}^R \\ \widetilde{\boldsymbol{I}}_n \tilde{\mathbf{k}}^R \end{pmatrix} \otimes \boldsymbol{I}_m \in \mathbb{R}^{16mn \times 4mn}.$$
(3.6)

Proof. This follows immediately from (1.13), (1.12) and Remark 3.5.

65

Lemma 3.7. Let $M \in \mathbb{H}^{k \times m}$. Then

$$(oldsymbol{I}_{4n}\otimes \widetilde{oldsymbol{M}})oldsymbol{\Upsilon}^{(m imes n)}=oldsymbol{\Upsilon}^{(k imes n)}(oldsymbol{I}_n\otimes \widetilde{oldsymbol{M}})$$

Proof. Consider an arbitrary $\boldsymbol{Q} \in \mathbb{H}^{m \times n}$. Then

$$\begin{split} (\boldsymbol{I}_{4n}\otimes\widetilde{\boldsymbol{M}})\boldsymbol{\Upsilon}^{(m\times n)}\operatorname{vec}(\boldsymbol{\mathcal{V}}(\boldsymbol{Q})) &= (\boldsymbol{I}_{4n}\otimes\widetilde{\boldsymbol{M}})\operatorname{vec}(\widetilde{\boldsymbol{Q}}) \\ &= \operatorname{vec}(\widetilde{\boldsymbol{M}}\widetilde{\boldsymbol{Q}}) \\ &= \boldsymbol{\Upsilon}^{(k\times n)}\operatorname{vec}(\boldsymbol{\mathcal{V}}(\boldsymbol{M}\boldsymbol{Q})) \\ &= \boldsymbol{\Upsilon}^{(k\times n)}\operatorname{vec}(\widetilde{\boldsymbol{M}}\,\boldsymbol{\mathcal{V}}(\boldsymbol{Q})) \\ &= \boldsymbol{\Upsilon}^{(k\times n)}\operatorname{vec}(\widetilde{\boldsymbol{M}}\,\boldsymbol{\mathcal{V}}(\boldsymbol{Q})) \\ &= \boldsymbol{\Upsilon}^{(k\times n)}(\boldsymbol{I}_n\otimes\widetilde{\boldsymbol{M}})\operatorname{vec}(\boldsymbol{\mathcal{V}}(\boldsymbol{Q})). \end{split}$$

Alternatively, Lemma 3.7 can be proven using (3.6) and Lemma 1.38. It is then clear that the quaternion structure of \widetilde{M} is a necessary condition.

Theorem 3.8. Consider the quaternion multivariate linear regression (a.k.a. general linear model)

$$\boldsymbol{Q} = \boldsymbol{B}\boldsymbol{W} + \boldsymbol{E},\tag{3.7}$$

were $\mathbf{Q} \in \mathbb{H}^{m \times N}$, $\mathbf{W} \in \mathbb{H}^{k \times N}$ are the observed quaternion response and regressor matrices respectively, and $\mathbf{E} \in \mathbb{H}^{m \times N}$ is an error matrix whose columns are uncorrelated second-order proper quaternion random vectors with common covariance matrix. Then the least squares estimator and generalised least squares estimator of the regression coefficients $\mathbf{B} \in \mathbb{H}^{m \times k}$ are equal and (assuming $\mathbf{W}\mathbf{W}^{H}$ invertible) given by

$$\hat{\boldsymbol{B}} = \boldsymbol{Q}\boldsymbol{W}^{H}(\boldsymbol{W}\boldsymbol{W}^{H})^{-1}.$$
(3.8)

Proof. Let $\boldsymbol{y} = \operatorname{vec}(\mathcal{V}(\boldsymbol{Q})), \boldsymbol{\beta} = \operatorname{vec}(\mathcal{V}(\boldsymbol{B}))$ and $\boldsymbol{e} = \operatorname{vec}(\mathcal{V}(\boldsymbol{E}))$. Then

$$\begin{split} \boldsymbol{y} &= \operatorname{vec}(\mathcal{V}(\boldsymbol{B}\boldsymbol{W})) + \boldsymbol{e} \\ &= \operatorname{vec}\left(\widetilde{\boldsymbol{B}} \, \mathcal{V}(\boldsymbol{W})\right) + \boldsymbol{e} \\ &= (\mathcal{V}(\boldsymbol{W})^T \otimes \boldsymbol{I}_{4m}) \operatorname{vec}(\widetilde{\boldsymbol{B}}) + \boldsymbol{e} \\ &= (\mathcal{V}(\boldsymbol{W})^T \otimes \boldsymbol{I}_{4m}) \Upsilon^{(m \times k)} \operatorname{vec}(\mathcal{V}(\boldsymbol{B})) + \boldsymbol{e} \\ &= \boldsymbol{X} \boldsymbol{\beta} + \boldsymbol{e}, \end{split}$$

where we define $\boldsymbol{X} = (\mathcal{V}(\boldsymbol{W})^T \otimes \boldsymbol{I}_{4m}) \Upsilon^{(m \times k)}$.

Let $\Sigma_{E_{\bullet,1}} \in \mathbb{H}^{m \times m}$ be the common covariance matrix of the columns of E, and let $\widetilde{M} = \Sigma_{\mathcal{V}(E_{\bullet,1})} = \frac{1}{4} \widetilde{\Sigma_{E_{\bullet,1}}}$. Then $\Sigma_e = I_N \otimes \widetilde{M}$

Using the above, (3.5) and Lemma 3.7

$$\begin{split} \boldsymbol{\Sigma_{e}X} &= (\boldsymbol{I}_{N} \otimes \widetilde{\boldsymbol{M}}) (\boldsymbol{\mathcal{V}}(\boldsymbol{W})^{T} \otimes \boldsymbol{I}_{4m}) \boldsymbol{\Upsilon}^{(m \times k)} \\ &= (\boldsymbol{\mathcal{V}}(\boldsymbol{W})^{T} \otimes \widetilde{\boldsymbol{M}}) \boldsymbol{\Upsilon}^{(m \times k)} \\ &= (\boldsymbol{\mathcal{V}}(\boldsymbol{W})^{T} \otimes \boldsymbol{I}_{4m}) (\boldsymbol{I}_{4k} \otimes \widetilde{\boldsymbol{M}}) \boldsymbol{\Upsilon}^{(m \times k)} \\ &= (\boldsymbol{\mathcal{V}}(\boldsymbol{W})^{T} \otimes \boldsymbol{I}_{4m}) \boldsymbol{\Upsilon}^{(m \times k)} (\boldsymbol{I}_{k} \otimes \widetilde{\boldsymbol{M}}) \\ &= \boldsymbol{XS}, \end{split}$$

where $S = I_k \otimes \widetilde{M}$. Hence by Proposition 3.1, the LS and GLS estimators are equal.

Now to prove (3.8), let $\hat{E} = Q - \hat{B}W$. The LS estimator is the value of \hat{B} which minimises the sum of squared (absolute) errors, which is given by the squared Frobenius norm $||\hat{E}||^2 = ||Q - \hat{B}W||^2$. By (1.21) $4||\hat{E}||^2 = ||\tilde{E}||^2$, so that we may equivalently minimise $||\tilde{Q} - \tilde{B}\widetilde{W}||^2$. This is a structured real least squares problem since we are restricted to $\tilde{B} \in \widetilde{\mathbb{H}}^{m \times k}$. However, if we drop this restriction, it becomes a standard real LS problem with solution $\tilde{B} = \widetilde{Q}\widetilde{W}^T \left(\widetilde{W}\widetilde{W}^T\right)^{-1}$ (see e.g. Lütkepohl (2006, pp. 71–72)). By Theorem 1.26 and Corollary 1.29 this is equal to $\tilde{\bullet} \left(QW^H (WW^H)^{-1} \right) \in \widetilde{\mathbb{H}}^{m \times k}$ and hence solves the original structured least squares problem.

Remark 3.9. In practice, computing the LS estimator of \boldsymbol{B} using a standard real LS

algorithm on the real representation (such as $\hat{B} = \tilde{Q}/\tilde{W}$ in Matlab) will be faster and more numerically stable than using (3.8) explicitly. Even better, one may extrapolate the rest of \tilde{B} from its first row (which can be computed as $\mathcal{V}\left(\hat{B}^{H}\right)^{T} = \mathcal{V}\left(Q^{H}\right)^{T}/\tilde{W}$).

Note that in our proof of (3.8) we reduce the quaternion LS problem to a real LS problem using the real representation $\tilde{\bullet}$. Similarly, Jiang and Chen (2007) reduce the quaternion LS problem to a complex LS problem by using the complex representation.

The LS estimate is the choice of \hat{B} minimising the sum of (estimated) squared errors $||\hat{E}||^2 = \operatorname{tr}\left(\hat{E}^H\hat{E}\right)$, whereas the GLS estimate minimises

$$\operatorname{tr}\left(\mathcal{V}(\hat{\boldsymbol{E}})^T \boldsymbol{\Sigma}_{\mathcal{V}(\boldsymbol{E}_{\bullet,1})}^{-1} \, \mathcal{V}(\hat{\boldsymbol{E}})\right). \tag{3.9}$$

In the improper case, minimising tr $\left(\hat{E}^{H}\Sigma_{E_{\bullet,1}}^{-1}\hat{E}\right)$ instead of (3.9) — or equivalently (by Theorem 3.8) using LS instead of GLS — effectively amounts to using the closest quaternion-structured approximation to the true error covariance matrix. This misspecification of the error covariance matrix will then lead to a loss of efficiency.

An interesting aspect of Theorem 3.8 is that (3.7) is *left*-linear in the parameter⁶ \boldsymbol{B} whereas \boldsymbol{E} is assumed to be *right* second-order proper.⁷ As we discussed in Section 1.3.2, applying the quaternion conjugate $\bar{\bullet}$ to \mathbb{H} maps left-linear operators to right-linear operators and vice-versa, so in a quaternion right-linear version of Theorem 3.8 we would need to assume that \boldsymbol{E} is left second-order proper. The need for propriety assumptions is in itself somewhat counterintuitive, since one might expect that — as in the real case — each row of (3.7) could be treated independently as a linear regression.

Vía et al. (2010a) interpret the parameter \boldsymbol{B} in MLR as a real- or quaternion- linear transformation maximising the correlation between transformed regressors and response. This allows them to fit MLR within a unified approach also covering canonical correlation analysis and partial least squares. From this point of view,

⁶The 'linear' in 'linear regression' refers to linearity with respect to the parameter.

⁷This means that if we were to write the quaternion MLR (3.7) as a quaternion version of the linear regression model (3.1), the product $X\beta$ would actually have to be replaced by a left-linear product $X *^R \beta$, and the error vector e would be (right) second-order proper.

right-propriety of the errors goes along with right-linearity of the transformation in Theorem 3.8. They show that quaternion LS estimation is optimal in the following sense:

Given the MLR (3.7) Let $\mathbf{Y} = \mathcal{V}(\mathbf{Q})$ and $\mathbf{X} = \mathcal{V}(\mathbf{W})$. Then (3.7) becomes $\mathbf{Y} = \widetilde{\mathbf{B}}\mathbf{X} + \mathcal{V}(\mathbf{E})$. If we ignore the assumption of quaternion structure on $\widetilde{\mathbf{B}}$ then this is a real MLR. Now if we treat the columns of \mathbf{W} as random and assume they are second-order proper and furthermore assume that the cross-covariance between \mathbf{X} and \mathbf{Y} has quaternion structure, then the exact real MLR solution for $\widetilde{\mathbf{B}}$ has quaternion structure. The real (or quaternion widely-linear) MLR then reduces to a quaternion MLR (Vía et al., 2010b, Figure 1).

The projection interpretation of quaternion matrix multiplication from Section 1.3.3 sheds some light on the role played by the above assumptions. Assume for simplicity that all variables are mean-adjusted. We can then interpret $\hat{\Sigma}_{X_{\bullet,i}} = n^{-1}XX^{T}$ as an estimator of the regressor covariance matrix (where each column $X_{\bullet,i}$ of X is treated as a sample) and $\hat{\Sigma}_{Y_{\bullet,i},X_{\bullet,i}} = n^{-1}YX^{T}$ as an estimator of the cross covariance between response and regressor (where each column $Y_{\bullet,i}$ of Y is treated as a corresponding sample). Now, ignoring structural assumptions, the real LS solution is

$$\hat{\widetilde{B}}^{\mathbb{R}} = YX^T (XX^T)^{-1}$$

$$= \hat{\Sigma}_{Y_{\bullet,i},X_{\bullet,i}}\hat{\Sigma}_{X_{\bullet,i}}^{-1}.$$

Whilst from (3.8) using (1.15) the quaternion-linear LS solution is

$$\widetilde{\widehat{B}}^{\mathbb{H}} = \widetilde{Q}\widetilde{W}^{T} \left(\widetilde{W}\widetilde{W}^{T}\right)^{-1} = 4\hat{h} \left(\mathcal{V}(Q) \mathcal{V}(W)^{T}\right) (4n)^{-1} (4n) \left(4\hat{h} \left(\mathcal{V}(W) \mathcal{V}(W)^{T}\right)\right)^{-1} = \hat{h} \left(\hat{\Sigma}_{\mathbf{Y}_{\bullet,i}, \mathbf{X}_{\bullet,i}}\right) \hat{h} \left(\hat{\Sigma}_{\mathbf{X}_{\bullet,i}}\right)^{-1}.$$

This makes it clear that the difference between the real MLR and quaternion MLR approaches is *precisely* that the latter forces quaternion structure on $\hat{\Sigma}_{Y_{\bullet,i},X_{\bullet,i}}$ and $\hat{\Sigma}_{X_{\bullet,i}}$ through orthogonal projection.

Next, we will consider the quaternion VAR time-series model, and apply Theorem 3.8 to VAR parameter estimation.

3.3 Quaternion VAR as a structured real VAR

Definition 3.10. Let $A_1, \ldots, A_p \in \mathbb{R}^{n \times n}$, $\mu \in \mathbb{R}^n$, and let $\epsilon_t \in \mathbb{R}^n$ be a sequence of uncorrelated zero-mean random vectors with a common covariance matrix $\Sigma_{\epsilon} = E\{\epsilon_t \epsilon_t^T\}$. The process

$$\boldsymbol{y}_t = \boldsymbol{\mu} + \boldsymbol{A}_1 \boldsymbol{y}_{t-1} + \ldots + \boldsymbol{A}_p \boldsymbol{y}_{t-p} + \boldsymbol{\epsilon}_t$$
(3.10)

is a real VAR process $AR_n^{\mathbb{R}}(p)$.

Definition 3.11. Let $A_1, \ldots, A_p \in \mathbb{H}^{n \times n}$, $\mu \in \mathbb{H}^n$, and let $\epsilon_t \in \mathbb{H}^n$ be a sequence of uncorrelated zero-mean second-order proper random vectors with a common covariance matrix $\Sigma_{\epsilon} = E\{\epsilon_t \epsilon_t^H\} \in \mathbb{H}^{n \times n}$. Then the process

$$oldsymbol{q}_t = oldsymbol{\mu} + oldsymbol{A}_1 oldsymbol{q}_{t-1} + \ldots + oldsymbol{A}_p oldsymbol{q}_{t-p} + oldsymbol{\epsilon}_t$$

is a proper quaternion VAR process, i.e., proper $AR_n^{\mathbb{H}}(p)$.

Proposition 3.12. Let \boldsymbol{q}_t be the proper $\operatorname{AR}_n^{\mathbb{H}}(p)$ process of Definition 3.11. Then $\boldsymbol{y}_t = \mathcal{V}(\boldsymbol{q}_t)$ is the $\operatorname{AR}_{4n}^{\mathbb{R}}(p)$ process

$$\boldsymbol{y}_{t} = \mathcal{V}(\boldsymbol{\mu}) + \widetilde{\boldsymbol{A}}_{1}\boldsymbol{y}_{t-1} + \ldots + \widetilde{\boldsymbol{A}}_{p}\boldsymbol{y}_{t-p} + \mathcal{V}(\boldsymbol{\epsilon}_{t}).$$
(3.11)

Furthermore, the innovations covariance is

$$\Sigma_{\mathcal{V}(\boldsymbol{\epsilon})} = \frac{1}{4} \widetilde{\Sigma_{\boldsymbol{\epsilon}}} \in \widetilde{\mathbb{H}^{n \times n}}.$$
(3.12)

Conversely, consider an arbitrary $\operatorname{AR}_{4n}^{\mathbb{R}}(p)$ process \boldsymbol{y}_t as in Definition 3.10. Then $\boldsymbol{q}_t = \mathcal{V}^{-1}(\boldsymbol{y}_t)$ is a proper $\operatorname{AR}_n^{\mathbb{H}}(p)$ process if both the regression coefficients \boldsymbol{A}_i and the innovations covariance $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ belong to $\widetilde{\mathbb{H}^{n \times n}}$ *Proof.* (3.11) is immediate from Remark 1.33 and the (real) linearity of \mathcal{V} . (3.12) is given by Lemma 2.20. The converse is proved the same way.

By considering the real form $\mathcal{V}(\boldsymbol{q}_t)$ of a proper quaternion VAR process, we can immediately translate standard definitions and theoretical results from real VARs to quaternion VARs. For example, we can define \boldsymbol{q}_t to be stable (resp. stationary or Gaussian) if and only if $\mathcal{V}(\boldsymbol{q}_t)$ is stable (resp. stationary or Gaussian). Then, as in the real case (Lütkepohl, 2006, Proposition 2.1), every stable quaternion VAR is stationary.

The particularity of the quaternion VAR is that we impose quaternion structure on the parameters. Thus, to translate *some* results from real VARs to quaternion VARs, we need to check (usually effortlessly) that the structure is preserved. For example, any $AR_n^{\mathbb{H}}(p)$ process \boldsymbol{q}_t can be rewritten as a $AR_{np}^{\mathbb{H}}(1)$ process $(\boldsymbol{q}_t^T, \ldots, \boldsymbol{q}_{t-p+1}^T)^T$, even though the standard $AR_{4np}^{\mathbb{R}}(1)$ version of its $AR_{4n}^{\mathbb{R}}(p)$ representation will not have quaternion structure due to the ordering of its components.

Theorem 3.13. The joint distribution of values from a stable proper $AR_n^{\mathbb{H}}(p)$ process q_t is second-order proper.

Conversely, for any stable $\operatorname{AR}_{4n}^{\mathbb{R}}(p)$ process \boldsymbol{y}_t , $\boldsymbol{q}_t = \mathcal{V}^{-1}(\boldsymbol{y}_t)$ is a proper $\operatorname{AR}_n^{\mathbb{H}}(p)$ process if its values are jointly second-order proper.

Proof. By considering the $\operatorname{AR}_{np}^{\mathbb{H}}(1)$ process $(\boldsymbol{q}_t^T, \ldots, \boldsymbol{q}_{t-p+1}^T)^T$, we may assume without loss of generality that p=1. Let $\boldsymbol{y}_t = \mathcal{V}(\boldsymbol{q}_t)$. From Lütkepohl (2006, (2.1.18)&(2.1.22)) we can write

$$\Sigma_{\boldsymbol{y}_t, \boldsymbol{y}_{t- au}} = \sum_{i=0}^{\infty} \Phi_{ au+i} \Sigma_{\mathcal{V}(\boldsymbol{\epsilon})} \Phi_i^T,$$

where $\Phi_i \in \mathbb{R}^{4n \times 4n}$ are the coefficients of the moving average representation of y_t , and are given by

$$oldsymbol{\Phi}_i = \widetilde{oldsymbol{A}}_1^i$$
 .

Hence $\Phi_i \in \widetilde{\mathbb{H}^{n \times n}}$ and $\Sigma_{y_t, y_{t-\tau}} \in \widetilde{\mathbb{H}^{n \times n}}$. Finally note that pairwise second-order propriety implies full second-order propriety.

For the converse, we may again assume without loss of generality that p = 1by considering $\mathcal{V}^{-1}\left(\left(\boldsymbol{q}_{t}^{T},\ldots,\boldsymbol{q}_{t-p+1}^{T}\right)^{T}\right)$. Second-order propriety implies $\boldsymbol{\Sigma}_{\boldsymbol{y}_{t},\boldsymbol{y}_{t-\tau}} =$ $\frac{1}{4}\widetilde{\Sigma_{q_t,q_{t-\tau}}} \in \widetilde{\mathbb{H}^{n \times n}}$, and the Yule-Walker equations give us (Lütkepohl, 2006, p. 86)

$$oldsymbol{A}_1 = oldsymbol{\Sigma}_{oldsymbol{y}_t, oldsymbol{y}_{t-1}} oldsymbol{\Sigma}_{oldsymbol{y}_t}^{-1} \in \widetilde{\mathbb{H}^{n imes n}}.$$

We may assume without loss of generality that Σ_{y_t} is non-singular.⁸

Corollary 3.14. A Gaussian stable $AR_n^{\mathbb{H}}(p)$ process is proper in the sense of Definition 3.11 if and only if it⁹ is proper in the sense of Definition 2.3

Although Theorem 3.13 has been assumed implicitly by e.g. Navarro-Moreno et al. (2013), we have not found a proof of it in the literature.

3.3.1 Quaternion VAR parameter estimation

Let \boldsymbol{q}_t be proper $AR_n^{\mathbb{H}}(p)$ as in Definition 3.11. Define

$$Q = \begin{bmatrix} q_1 & \dots & q_N \end{bmatrix} \in \mathbb{H}^{n \times N}$$

$$B = \begin{bmatrix} \mu & A_1 & \dots & A_p \end{bmatrix} \in \mathbb{H}^{n \times (np+1)}$$

$$W = \begin{bmatrix} 1 & 1 & \dots & 1 \\ q_0 & q_1 & \dots & q_{N-1} \\ q_{-1} & q_0 & \dots & q_{N-2} \\ \vdots & & \vdots \\ q_{-p+1} & q_{-p+2} & \dots & q_{N-p} \end{bmatrix} \in \mathbb{H}^{(np+1) \times N}$$

$$E = \begin{bmatrix} \epsilon_1 & \dots & \epsilon_N \end{bmatrix} \in \mathbb{H}^{n \times N}$$

so that

$$\boldsymbol{Q} = \boldsymbol{B}\boldsymbol{W} + \boldsymbol{E},\tag{3.13}$$

⁸If Σ_{y_t} is singular we may remove constant deterministic components from q_t by applying a quaternion eigenvalue decomposition to Σ_{q_t} . If we do not remove these deterministic components there may be superficially improper solutions in addition to the proper solution(s) $A_1 \in \widetilde{\mathbb{H}^{n \times n}}$.

⁹or technically any finite subseries, since we have only defined the finite-dimensional quaternion normal distribution.
Let $y_t = \mathcal{V}(q_t)$. We can do the same with the real representation y_t and define

$$\begin{split} \boldsymbol{B}^{\star} &= \left[\begin{array}{ccc} \mathcal{V}(\boldsymbol{\mu}) & \widetilde{\boldsymbol{A}}_{1} & \dots & \widetilde{\boldsymbol{A}}_{p} \end{array} \right] \in \mathbb{R}^{4n \times (4np+1)} \\ \boldsymbol{Y} &= \left[\begin{array}{cccc} \boldsymbol{y}_{1} & \dots & \boldsymbol{y}_{N} \end{array} \right] \in \mathbb{R}^{4n \times N} \\ & & & \\ \boldsymbol{W}^{\star} &= \left[\begin{array}{ccccc} 1 & 1 & \dots & 1 \\ \boldsymbol{y}_{0} & \boldsymbol{y}_{1} & \dots & \boldsymbol{y}_{N-1} \\ \boldsymbol{y}_{-1} & \boldsymbol{y}_{0} & \dots & \boldsymbol{y}_{N-2} \\ \vdots & & \vdots \\ \boldsymbol{y}_{-p+1} & \boldsymbol{y}_{-p+2} & \dots & \boldsymbol{y}_{N-p+1} \end{array} \right] \in \mathbb{R}^{(4np+1) \times N} \\ \boldsymbol{E}^{\star} &= \left[\begin{array}{c} \mathcal{V}(\boldsymbol{E}) = \left[\begin{array}{c} \mathcal{V}(\boldsymbol{\epsilon}_{1}) & \dots & \mathcal{V}(\boldsymbol{\epsilon}_{N}) \end{array} \right] \in \mathbb{R}^{4n \times N}, \end{split}$$

so that $\boldsymbol{Y} = \boldsymbol{B}^{\star} \boldsymbol{W}^{\star} + \boldsymbol{E}^{\star}$.

There are multiple ways of viewing the parameter estimation problem. Treating it as a structured real VAR model we have $\mathbf{Y} = \mathbf{B}^* \mathbf{W}^* + \mathbf{E}^*$. This would be the approach of Lütkepohl (2006). Mapping (3.13) to the real domain directly we get instead $\mathbf{Y} = \mathcal{V}(\mathbf{Q}) = \mathcal{V}(\mathbf{B}\mathbf{W}) + \mathcal{V}(\mathbf{E}) = \widetilde{\mathbf{B}}\mathcal{V}(\mathbf{W}) + \mathbf{E}^*$. This is the approach used in the proof of Theorem 3.8. Yet another interpretation comes from considering that $\mathcal{V}(\mathbf{Q}^H) = \widetilde{\mathbf{W}}^T \mathcal{V}(\mathbf{B}^H) + \mathcal{V}(\mathbf{E}^H)$. This is a real regression problem with an *unstructured* parameter, but the correlation structure of $\mathcal{V}(\mathbf{E}^H)$ is harder to describe (due to our choices of representation and notation).

Corollary 3.15. The LS and GLS estimators for the parameter **B** of a proper $AR_n^{\mathbb{H}}(p)$ process are equal. If the process is Gaussian, then the maximum likelihood estimator (MLE) is also equal to the LS estimator.

Proof. The LS and GLS estimators are equal by Theorem 3.8.

Under a Gaussianity assumption, the MLE of \boldsymbol{B} is equal to the GLS estimator, except that the covariance matrix $\Sigma_{\mathcal{V}(\boldsymbol{E}_{\bullet,1})} = \frac{1}{4}\widetilde{\Sigma_{\epsilon}}$ appearing in (3.9) is replaced by its MLE (Lütkepohl, 2006, eqn. 5.2.17). Because we assume that the innovations are proper, the covariance MLE is restricted to have quaternion structure (see also Proposition 4.3). Since the LS estimator does not depend on $\Sigma_{\mathcal{V}(\boldsymbol{E}_{\bullet,1})}$, replacing $\Sigma_{\mathcal{V}(\boldsymbol{E}_{\bullet,1})}$ with another quaternion-structured covariance matrix has no effect. *Remark* 3.16. When computing the MLE of a Gaussian $AR_n^{\mathbb{H}}(p)$ process, the first p samples are treated as constants describing the initial state, rather than random variables from a stationary process. This differs from the interpretation behind the Yule-Walker approach.

Remark 3.17. For a scalar $(AR_1^{\mathbb{H}}(p))$ quaternion autoregressive process, propriety of the innovations covariance is equivalent to circularity $\Sigma_{\mathcal{V}(\epsilon)} = \sigma^2 I_4$. In this case the equivalence between LS and GLS estimation is obvious since $\Sigma_e = \sigma^2 I_{4nN}$.

Let \boldsymbol{q}_t be an $\operatorname{AR}_n^{\mathbb{H}}(p)$ process and consider the matrix-valued time-series $\widetilde{\boldsymbol{q}}_t$. Similarly to Section 1.3.3, using (1.13), the four columns of this matrix form can be interpreted as an ensemble of four $\operatorname{AR}_{4n}^{\mathbb{R}}(p)$ time series $\mathcal{V}(\boldsymbol{q}_t), \mathcal{V}(\boldsymbol{q}_t\mathrm{i}), \mathcal{V}(\boldsymbol{q}_t\mathrm{j}), \mathcal{V}(\boldsymbol{q}_t\mathrm{k}),$ having shared regression parameters. (To be more precise, the constant term is different for each of the four time series, and is given by $\mathcal{V}(\boldsymbol{\mu}), \mathcal{V}(\boldsymbol{\mu}\mathrm{i}), \mathcal{V}(\boldsymbol{\mu}\mathrm{j}), \mathcal{V}(\boldsymbol{\mu}\mathrm{k}),$ respectively.) The first N columns of $\widetilde{\boldsymbol{E}} = \widetilde{\boldsymbol{Q}} - \widetilde{\boldsymbol{B}}\widetilde{\boldsymbol{W}}$ are given by

$$\mathcal{V}(\boldsymbol{Q}) - \widetilde{\boldsymbol{B}} \, \mathcal{V}(\boldsymbol{W}) = \boldsymbol{Y} - \boldsymbol{B}^{\star} \boldsymbol{W}^{\star},$$

and the following three blocks of N columns are given by

$$egin{aligned} \mathcal{V}(oldsymbol{Q}\mathrm{i}) &- \widetilde{oldsymbol{B}}\,\mathcal{V}(oldsymbol{W}\mathrm{i}), \ \mathcal{V}(oldsymbol{Q}\mathrm{j}) &- \widetilde{oldsymbol{B}}\,\mathcal{V}(oldsymbol{W}\mathrm{j}), \ \mathcal{V}(oldsymbol{Q}\mathrm{k}) &- \widetilde{oldsymbol{B}}\,\mathcal{V}(oldsymbol{W}\mathrm{k}), \end{aligned}$$

respectively, which are the corresponding matrices for $\mathcal{V}(\boldsymbol{q}_t \mathbf{i}), \mathcal{V}(\boldsymbol{q}_t \mathbf{j})$ and $\mathcal{V}(\boldsymbol{q}_t \mathbf{k})$.

We can see from the above that treating the columns of \tilde{q}_t as an ensemble of real $\operatorname{AR}_{4n}^{\mathbb{R}}(p)$ observations and computing the ensemble least squares parameter estimator (without imposing structural assumptions directly) gives us the desired structured least squares solution. This ensemble-based approach can be generalised to any process whose regression parameters are invariant under the action of a finite group.¹⁰

¹⁰Although quaternion propriety implies that the coefficients are invariant under the infinite group of transformations of the form $q_t \mapsto q_t u$ with u an arbitrary unit quaternion, we see that it is not actually necessary to integrate over all unit quaternions.

Other methods of parameter estimation can be considered. Navarro-Moreno et al. (2013) show that a quaternion Durbin-Levinson algorithm can be used to efficiently solve the Yule-Walker equations and compute parameter estimates from estimates of the lagged covariances $\hat{\Sigma}_{q_t,q_{t-\tau}}$. This method assumes stability, so by Theorem 3.13 assuming $4\Sigma_{y_t,y_{t-\tau}} = \widetilde{\Sigma}_{q_t,q_{t-\tau}}$ will lead to a proper solution. For the widely-linear case, this is simply a reformulation of the block Durbin-Levinson algorithm of Akaike (1973). In the proper case, the quaternion-domain algorithm is still equivalent to the block Durbin-Levinson algorithm, where the blocks are set to have quaternion structure.¹¹ In practice the proper Yule-Walker approach differs only through taking $\hat{h}(\hat{\Sigma}_{y_t,y_{t-\tau}})$ as the autocorrelation estimator instead of $\Sigma_{y_t,y_{t-\tau}}$.

We will also consider a 'naive method', in which an unrestricted real LS estimator for the parameters is computed, and then projected onto the nearest structured solution using \hat{h} . Note that orthogonal projection onto a space containing the true value always improves estimates. Indeed $\forall \boldsymbol{M} \in \widetilde{\mathbb{H}^{m \times n}}, \hat{\boldsymbol{M}} \in \mathbb{R}^{4m \times 4n}$,

$$||\hat{M} - M||^2 = ||\hat{h}(\hat{M}) - M||^2 + ||\hat{M} - \hat{h}(\hat{M}) - M||^2.$$

3.3.2 Numerical evaluation

As an example, we will consider the quaternion linear but improper $AR_1^{\mathbb{H}}(1)$ process¹²

$$q_t = a_1 q_{t-1} + \epsilon_t,$$

¹¹Taking account of this structure however allows for improvements to the algorithm which reduce the number of required operations.

¹²As a consequence of Remark 1.14, every proper $AR_1^{\mathbb{H}}(1)$ process can be expressed as a pair of uncorrelated realisations from a $AR_1^{\mathbb{C}}(1)$ process.

where $a_1 = 0.7 + 0.5i$ and ϵ_t is an *improper* quaternion normal random variable with covariance matrix

$$\boldsymbol{\Sigma}_{\mathcal{V}(\epsilon)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \gamma & 0 & 0 \\ 0 & 0 & \gamma & 0 \\ 0 & 0 & 0 & \gamma \end{bmatrix}.$$
 (3.14)

We first set $\gamma = 0.1$ in (3.14). For various N, we generate a sample time-series of length N + p = N + 1 (allowing for a burn-in period of 10 000 samples to avoid initialisation effects). We then estimate $\tilde{a_1}$ by $\hat{a_1}$ using quaternion LS estimation, quaternion GLS estimation, unstructured real LS estimation (which is equal to GLS), and the naive LS method (the quaternion-structured projection of the real LS solution). The estimation error is given by the L^2 (or Frobenius) distance $||\tilde{a_1} - \hat{a_1}||$, which is equal to $2|a_1 - \hat{a_1}|$ (except in the unstructured real case where $\hat{a_1}$ will not have quaternion structure). We average the estimation error over 100 independent simulations to obtain an approximate average error. The results are given in Figure 3.1. We see that quaternion GLS marginally outperforms LS which in turn marginally outperforms the naive method.

An $\operatorname{AR}_{4}^{\mathbb{R}}(1)$ model has 16 real regression parameters in addition to 4 real mean parameters and 10 degrees of freedom in the error covariance matrix. A proper $\operatorname{AR}_{1}^{\mathbb{H}}(1)$ model on the other hand has only 4 real regression parameters in addition to 4 real mean parameters and 1 degree of freedom in the error covariance matrix. We see that here reducing the number of parameters being estimated from 20 to 8 improves estimation accuracy by an amount comparable to an order of magnitude increase in sample size.

In Ginzberg and Walden (2013b) we concluded that the loss of efficiency when using LS instead of GLS for a quaternion VAR with improper error covariance was minor in practice, especially when compared with the improvement obtained by reducing the number of parameters through the assumption of quaternion structure. This is confirmed by Figure 3.1 and will usually be true, however it is worth noting that this is no longer true for cases of extreme impropriety. If the error covariance is



Figure 3.1: Error in the estimation of \tilde{a}_1 for varying sample length N. From top to bottom the methods used are real LS (squares), the naive projection of the real LS (circles), quaternion LS (plain line) and quaternion GLS (dashed)

singular, say

$$\boldsymbol{\Sigma}_{\mathcal{V}(\boldsymbol{\epsilon})} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

1

then we can estimate the last row of \tilde{a}_1 exactly, which then gives us an exact estimate for the whole of \tilde{a}_1 . So in cases of extreme impropriety where $\Sigma_{\mathcal{V}(\epsilon)}$ is near-singular, one may get an arbitrarily large improvement from using the GLS estimator over the LS estimator.

In Figure 3.2 we examine this effect by varying γ in (3.14) to change the degree of impropriety. $\gamma = 1$ corresponds to propriety while very large or small values of γ indicate high impropriety. We see that the GLS does indeed offer a large improvement for highly improper errors, and no improvement when propriety holds ($\gamma = 1$). For $\gamma \ll 1$, the resulting anisotropy in the steady state distribution of \boldsymbol{y}_t makes estimating the first row of \tilde{a}_1 difficult for the real least squares algorithm, since the regressors are small relative to the noise. This leads to large errors for the real LS and also by extension for the naive method. We see that if not for this effect, the naive method can provide a reasonably good approximation to quaternion LS, as we can see for $\gamma \geq 1$.

It is worth noting that, unlike typical linear regression, in the VAR context parameter estimation is not affected by the overall level of noise. Indeed, scaling of the error covariance matrix leads to equal scaling of the process y_t and thus of the regressors (after mean-adjustment).

The approach in Navarro-Moreno et al. (2013) is based on Yule-Walker rather than LS estimation of the regression coefficients, and they consider an example widelylinear quaternion AR process. They also find that including improper errors decreases the efficiency of the proper quaternion parameter estimates. However, we believe that the mechanism for this effect is different in their example. Namely, we believe that it is simply due to an increase in the overall impropriety of the process and thus an increase in bias when the projection \hat{h} is mistakenly applied to the autocovariance matrices.



Figure 3.2: Error in the estimation of \tilde{a} for N = 100 and varying degrees of impropriety controlled by γ . $\gamma = 1$ corresponds to propriety. From top to bottom the methods used are real LS (squares), the naive projection of the real LS (circles), quaternion LS (plain line) and quaternion GLS (dashed)

3.4 Widely-linear quaternion VAR as a real VAR

Widely linear quaternion AR modelling has been used for example by Jahanchahi et al. (2010) for wind forecasting.

Using the approach in Section 2.5 (see also Vía et al. (2010a)), and similarly to improper complex signal processing (Schreier, 2010), a unified treatment of proper and improper quaternion signals in the quaternion domain can be obtained by allowing additional operations on the three involutions $q^{(\eta)} = -\eta q \eta$, $\eta = i, j, k$. However, as was noted by Rubin-Delanchy (2008) for the complex case, improper quaternion AR (or VAR) modelling is simply a more complicated reformulation of standard real VAR modelling (see also Figure 2.1).

Definition 3.18. Let $A_1, B_1, C_1, D_1, \ldots, A_p, B_p, C_p, D_p \in \mathbb{H}^{n \times n}$, $\mu \in \mathbb{H}^n$, and let $\epsilon_t \in \mathbb{H}^n$ be a sequence of (possibly improper) uncorrelated zero-mean innovations with common covariance $\Sigma_{\mathcal{V}(\epsilon)}$. The process

$$\boldsymbol{q}_{t} = \boldsymbol{\mu} + \boldsymbol{A}_{1}\boldsymbol{q}_{t-1} + \boldsymbol{B}_{1}\boldsymbol{q}_{t-1}^{(\mathrm{i})} + \boldsymbol{C}_{1}\boldsymbol{q}_{t-1}^{(\mathrm{j})} + \boldsymbol{D}_{1}\boldsymbol{q}_{t-1}^{(\mathrm{k})} + \dots + \boldsymbol{D}_{p}\boldsymbol{q}_{t-p}^{(\mathrm{k})} + \boldsymbol{\epsilon}_{t}$$
(3.15)

is a widely-linear quaternion VAR process, i.e., widely-linear $AR_{n}^{\mathbb{H}}(p)$.

Using the augmented quaternion formalism from Section 2.5, we may write

$$\underline{\boldsymbol{q}}_{t} = \mathfrak{A}_{n} \, \mathcal{V}(\boldsymbol{q}_{t}) = \underline{\boldsymbol{\mu}} + \underline{\boldsymbol{A}}_{1} \underline{\boldsymbol{q}}_{t-1} + \ldots + \underline{\boldsymbol{A}}_{p} \underline{\boldsymbol{q}}_{t-p} + \underline{\boldsymbol{\epsilon}}_{t},$$

where

$$\underline{\boldsymbol{A}}_{\ell} = \begin{pmatrix} \boldsymbol{A}_{\ell} & \boldsymbol{B}_{\ell} & \boldsymbol{C}_{\ell} & \boldsymbol{D}_{\ell} \\ \boldsymbol{A}_{\ell}^{(\mathrm{i})} & \boldsymbol{B}_{\ell}^{(\mathrm{i})} & \boldsymbol{C}_{\ell}^{(\mathrm{i})} & \boldsymbol{D}_{\ell}^{(\mathrm{i})} \\ \boldsymbol{A}_{\ell}^{(\mathrm{j})} & \boldsymbol{B}_{\ell}^{(\mathrm{j})} & \boldsymbol{C}_{\ell}^{(\mathrm{j})} & \boldsymbol{D}_{\ell}^{(\mathrm{j})} \\ \boldsymbol{A}_{\ell}^{(\mathrm{k})} & \boldsymbol{B}_{\ell}^{(\mathrm{k})} & \boldsymbol{C}_{\ell}^{(\mathrm{k})} & \boldsymbol{D}_{\ell}^{(\mathrm{k})} \end{pmatrix} = \frac{1}{4} \mathfrak{A}_{n} \boldsymbol{A}_{\ell}^{\star} \mathfrak{A}_{n}^{H}$$

It will usually be simpler to consider the $AR_{4n}^{\mathbb{H}}(p)$ representation

$$\boldsymbol{y}_t = \mathcal{V}(\boldsymbol{q}_t) = \frac{1}{4} \mathfrak{A}_n^H \underline{\boldsymbol{q}}_t = \mathcal{V}(\boldsymbol{\mu}) + \boldsymbol{A}_1^* \boldsymbol{y}_{t-1} + \ldots + \boldsymbol{A}_p^* \boldsymbol{y}_{t-p} + \mathcal{V}(\boldsymbol{\epsilon}_t),$$

with coefficients given by

$$\boldsymbol{A}_{\ell}^{\star} = \widetilde{\boldsymbol{A}}_{\ell} + \widetilde{\boldsymbol{B}}_{\ell} \mathrm{i} \widetilde{\boldsymbol{I}}_{n}^{R} + \widetilde{\boldsymbol{C}}_{\ell} \mathrm{j} \widetilde{\boldsymbol{I}}_{n}^{R} + \widetilde{\boldsymbol{D}}_{\ell} \mathrm{k} \widetilde{\boldsymbol{I}}_{n}^{R}.$$

This equation shows that the widely-linear quaternion VAR (3.15) is simply a reformulation of the standard real VAR which draws attention to the orthogonal decomposition $\mathbb{R}^{4m \times 4n} = \widetilde{\mathbb{H}^{m \times n}} + \widetilde{\mathbb{H}^{m \times n}} \widetilde{\mathbf{I}_n}^R + \widetilde{\mathbb{H}^{m \times n}} \widetilde{\mathbf{J}_n}^R + \widetilde{\mathbb{H}^{m \times n}} \widetilde{\mathbf{K}_n}^R$.

Because there is no assumed structure in the $AR_{4n}^{\mathbb{R}}(p)$ representation of a widelylinear $AR_n^{\mathbb{H}}(p)$ process, all results about real VARs can be applied directly, including parameter estimation in the real domain.

Based on the ideas of Vía et al. (2010a), we can define \mathbb{C}^k -proper $\operatorname{AR}_n^{\mathbb{H}}(p)$ processes as a structured subclass of widely-linear $\operatorname{AR}_n^{\mathbb{H}}(p)$ processes. See Ginzberg and Walden (2013b) for more details.

3.5 Testing for VAR propriety

When given a time-series, one may want to check whether a quaternion proper VAR model is appropriate before imposing any parameter restrictions. In this section we give the LRT for quaternion propriety of an $AR_n^{\mathbb{H}}(p)$ time-series. Since this is a special case of the more general structured VAR testing problem, we simply apply the results of Lütkepohl (2006, Sections 4 and 5 and Appendix C.7) and note that propriety involves simultaneous parameter constraints on the regression parameters and on the residual covariance matrix.

A different approach to the model selection problem was considered by Ujang et al. (2013). They adaptively estimate both quaternion-linear and widely-linear AR models (through stochastic gradient descent), and combine them by taking a weighted average. The weights are adaptively tuned to favor the model with the lower prediction error. This allows them to combine fast convergence and efficiency for proper processes with the ability to model improper processes. The following heuristic can then be used: When the algorithm puts a weight close to 1 on the widely-linear model, then the process is believed to be improper. When it is the quaternion-linear model which has a weight close to 1, the process is believed to be proper. Their approach has the advantages of being applicable to both stationary and non-stationary signals, being online, and being based directly on the relative performance of proper and improper models (for prediction).¹³ A disadvantage of their approach compared to hypothesis testing is that the degree of certainty about the presence of quaternion structure is hard to quantify.¹⁴

Proposition 3.19. Let us have an observation of length N + p from a stable $AR_n^{\mathbb{H}}(p)$ process \mathbf{q}_t such that the errors $\mathcal{V}(\boldsymbol{\epsilon}_t)$ are i.i.d. with bounded fourth moment. Let $\hat{\Sigma}_{\mathbb{H}}$ and $\hat{\Sigma}_{\mathbb{R}}$ be the (Gaussian) maximum likelihood estimators of the error covariance $\Sigma_{\mathcal{V}(\boldsymbol{\epsilon})}$ with and without propriety assumptions respectively. The (Gaussian) LR test statistic T for testing H_0 : \mathbf{q}_t is proper against H_1 : \mathbf{q}_t is improper is given by

$$-2\ln(T) = N(\ln(|\hat{\boldsymbol{\Sigma}}_{\mathbb{R}}|) - \ln(|\hat{\boldsymbol{\Sigma}}_{\mathbb{H}}|))$$

and $-2\ln(T)$ is asymptotically distributed as χ^2_d with $d = 12n^2p + 6n^2 + 3n$ degrees of freedom under the null hypothesis H_0 .

Proof. In terms of the real representation $y_t = \mathcal{V}(q_t)$, propriety corresponds to linear restrictions which reduce the number of free parameters in the regression coefficients (excluding the mean) from $(4n)^2 p$ to $4n^2 p$, and the number of free parameters in the error covariance matrix from $\frac{1}{2}(4n)(4n+1)$ to n + 2n(n-1).¹⁵ The total reduction in degrees of freedom is

$$d = (16n^2p + 8n^2 + 2n) - (4n^2p + 2n^2 - n) = 12n^2p + 6n^2 + 3n.$$

By Lütkepohl (2006, Appendix C.7), including covariance matrix restrictions does not fundamentally alter the standard results on LR testing which are applied to re-

 $^{^{13}}$ The latter is an important distinction since proper modelling may outperform improper modelling when the true degree of impropriety and/or the sample size are small (Vía et al., 2010b, Figure 1).

¹⁴It may however still be possible to assign p-values based on Monte Carlo methods.

¹⁵Note that the diagonal elements of a quaternion covariance matrix are real.

gression parameter restrictions. Thus, the proposition follows from Lütkepohl (2006, Proposition 4.1). $\hfill \Box$

By Corollary 3.15, the Gaussian MLE of the regression parameters B^{\star} under H_0 is given by the structured LS estimator corresponding to the quaternion LS estimator (3.7), which we denote by $\hat{B}_{\mathbb{H}}^{\star}$, and which does not depend on $\Sigma_{\mathcal{V}(\epsilon)}$. The MLE of B^{\star} under H_1 is given by the standard (unstructured) real LS estimator, which we denote by $\hat{B}_{\mathbb{R}}^{\star}$ and also does not depend on $\Sigma_{\mathcal{V}(\epsilon)}$. As in Lütkepohl (2006, (4.2.11)-(4.2.12)), the Gaussian MLE of $\Sigma_{\mathcal{V}(\epsilon)}$ under H_1 is

$$\hat{\boldsymbol{\Sigma}}_{\mathbb{R}} = \frac{1}{N} \hat{\boldsymbol{E}}_{\mathbb{R}}^{\star} \hat{\boldsymbol{E}}_{\mathbb{R}}^{\star^{T}},$$

where $\hat{E}_{\mathbb{R}}^{\star} = Y - \hat{B}_{\mathbb{R}}^{\star} W^{\star}$.

If we did not impose quaternion structure on $\hat{\Sigma}_{\mathbb{H}}$, it would be similarly be given by $\frac{1}{N}\hat{E}_{\mathbb{H}}^{\star}\hat{E}_{\mathbb{H}}^{\star}^{T}$, where $\hat{E}_{\mathbb{H}}^{\star} = Y - \hat{B}_{\mathbb{H}}^{\star}W^{\star}$. We can see from Lütkepohl (2006, (3.4.5)) that, for fixed B^{\star} , the Gaussian likelihood function of E^{\star} is the likelihood function for a sample of N i.i.d $\mathcal{N}^{\mathbb{R}}\left(\mathbf{0}_{4n\times 1}\Sigma_{\mathcal{V}(\epsilon)}\right)$ random variables. As we will show in Proposition 4.3, under the assumption of quaternion structure, the MLE of a covariance matrix is obtained by orthogonal projection of the unstructured MLE. Thus the MLE under H_0 is

$$\hat{\boldsymbol{\Sigma}}_{\mathbb{H}} = \frac{1}{N} \hat{h} \left(\hat{\boldsymbol{E}}_{\mathbb{H}}^{\star} \hat{\boldsymbol{E}}_{\mathbb{H}}^{\star^{T}} \right).$$

We now have all the necessary elements to compute the LRT.

We will look at likelihood ratio testing for quaternion propriety and maximum likelihood estimation in much more detail in the following chapter, where we consider the problem of testing from an i.i.d. sample whether a multivariate normal distribution is quaternion proper. Note however that the results in the next chapter are not directly applicable to VAR modelling, since the (block-) Toeplitz structure of the signal covariance matrix must be taken into account (in addition to any a-priori specification of p).¹⁶

¹⁶Note that although general Toeplitz covariance structure cannot be described by group invariance, circulant structure corresponds to circular shift invariance and can be. Thus it may be possible

to use the results of the next section to obtain the exact distributions for propriety and other structure tests applied to VAR time-series, through circulant embedding. Alternatively, the results of Wojnar (1999) could be used. However, covariance estimates will be rank deficient if n > 1 and no assumption is made on p. This problem is left for future research.

Chapter 4

Likelihood Ratio Testing for Quaternion-Structured Covariance Matrices

4.1 Introduction

This chapter looks at the basic problem of determining whether an i.i.d quaternionvector-valued sample is proper. As we noted previously, because the covariance structure of quaternion propriety arises naturally from the way quaternion multiplication is defined, its presence begs for a treatment in the quaternion domain. Quaternion linear methods will then outperform their real linear counterparts due to having fewer free parameters. Conversely, the real domain is more appropriate when handling improper data, since the full flexibility of real linearity is then required (or equivalently, one may work with quaternion wide-linearity in the augmented quaternion domain). Mistakenly assuming quaternion propriety will introduce bias, whilst failing to acknowledge quaternion propriety will harm efficiency.

We will first describe the likelihood ratio test (LRT) for propriety of a multivariate quaternion normal distribution. Following the spirit of Andersson et al. (1983), and expanding on certain points for clarity, we show that because quaternion-structured real covariance matrices are a subset of complex-structured real covariance matrices (see the third row of Figure 2.1), the LRT for quaternion propriety is the product of the LRT for complex propriety and the LRT for quaternion propriety given complex propriety.

Vía et al. (2011) give an augmented quaternion domain derivation, and link the LRT to the information-theoretic Kullback-Leibler divergence. Vía and Vielva (2011) introduce the locally most powerful invariant test as a better alternative to the LRT when the amount of impropriety is small.

By computing the moments of the LRT, we show that it belongs to the general class of random variables of Box type. Many common tests are of Box type, including e.g. Wilks' statistic for multivariate analysis of variance and its quaternionic counterpart (Loots et al., 2012). Wojnar (1999) shows that under very general conditions, a LRT between two nested covariance models is of Box type. As was pointed out by Jensen (1991), the ten LRT statistics of Andersson et al. (1983) are of Box type.¹ Any LRT between nested group-invariance-based covariance structures can be obtained as a product of these ten tests, similarly to our derivation for the quaternion propriety LRT (Andersson, 1975). This also allows one to easily obtain the moments. Käufl (2012) shows that the LRT between two nested invariant (Gaussian) graphical models is of Box type. This includes as special cases testing (non-invariant) nested graphical models, as well as the group invariance structures considered by Andersson (1975).

We derive the exact density (PDF) and distribution function (CDF) for general random variables of Box type. These are given in terms of Meijer's G-function, or more generally Fox's H-function. Because routine computation of the exact CDF is impractical, we review many approximations which have been suggested in the literature for random variables of Box type. Knowledge of the exact distribution allows us to compare their accuracy for the LRT for quaternion propriety. A novel approximation based on the Pearson system of curves (Craig, 1936), which consists in fitting the moments of an F distribution exactly, is also suggested and found to

¹Except technically the statistic for testing equality of covariances, which we must first multiply by the constant $\left(\frac{(N_1+N_2)^{N_1+N_2}}{N_1^{N_1}N_2^{N_2}}\right)^{\delta p}$, where $\delta = 1, 2, 4$ for the real, complex and quaternion cases respectively.

be extremely accurate. Simpler methods fitting a gamma or chi-squared distribution are however more appropriate for large samples.

Some of the material in this chapter was published in Ginzberg and Walden (2011), see p. 11.

4.2 The LRT for quaternion propriety

Let $\boldsymbol{q} = \boldsymbol{a} + \boldsymbol{b}\mathbf{i} + \boldsymbol{c}\mathbf{j} + \boldsymbol{d}\mathbf{k}$ be a (possibly improper) *p*-dimensional quaternion normal random vector with mean **0**. Let $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_N$ be an i.i.d. sample from the distribution of \boldsymbol{q} and let $\boldsymbol{r}_1 = \mathcal{V}(\boldsymbol{q}_1), \ldots, \boldsymbol{r}_N = \mathcal{V}(\boldsymbol{q}_N)$ be the corresponding i.i.d. 4*p*-dimensional real-valued random vectors, $\boldsymbol{r}_i \sim \mathcal{N}^{\mathbb{R}}(\boldsymbol{0}_{p\times 1}, \boldsymbol{\Sigma}_r)$.

We shall assume that $N \ge 4p$ and denote the (unrestricted) maximum likelihood estimator of Σ_r by $\hat{\Sigma}_{\mathbb{R}}$, where

$$\hat{\boldsymbol{\Sigma}}_{\mathbb{R}} = \frac{1}{N} \sum_{\ell=1}^{N} \boldsymbol{r}_{\ell} \boldsymbol{r}_{\ell}^{T} \sim \mathcal{W}_{4p}^{\mathbb{R}} \left(\frac{1}{N} \boldsymbol{\Sigma}_{\boldsymbol{r}}, N \right).$$
(4.1)

Definition 4.1. Let $P_{\mathbb{R}}$ denote the set of $4p \times 4p$ symmetric positive definite matrices. Let $P_{\mathbb{C}} \subset P_{\mathbb{R}}$ and $P_{\mathbb{H}} \subset P_{\mathbb{R}}$ denote the set of symmetric positive definite matrices with complex structure (1.6) and quaternion structure (1.5) respectively.

We are interested in testing whether q is proper, which by Proposition 2.4 is equivalent to testing whether Σ_r has quaternion structure. In other words, we consider the hypothesis test

 H_0 : The $4p \times 4p$ real covariance matrix Σ_r has quaternion structure. $[H_0: \Sigma_r \in P_{\mathbb{H}}$ versus $H_1: \Sigma_r \in P_{\mathbb{R}} \setminus P_{\mathbb{H}}.]$

Since $P_{\mathbb{H}} \subset P_{\mathbb{C}}$, we can break this down into two tests with nested hypotheses.

- 1. Test 1. $H_0^{\mathbb{C}}$: The $4p \times 4p$ real covariance matrix Σ_r has complex structure. $[H_0^{\mathbb{C}}: \Sigma_r \in P_{\mathbb{C}}$ versus $H_1^{\mathbb{C}}: \Sigma_r \in P_{\mathbb{R}} \setminus P_{\mathbb{C}}.]$
- 2. Test 2. $H_0^{\mathbb{H}}$: The $4p \times 4p$ covariance matrix Σ_r with complex structure has quaternion structure. $[H_0^{\mathbb{H}}: \Sigma_r \in P_{\mathbb{H}} \text{ versus } H_1^{\mathbb{H}}: \Sigma_r \in P_{\mathbb{C}} \setminus P_{\mathbb{H}}.]$

4.2.1 Maximum likelihood estimators of covariance

Let

$$oldsymbol{z} = \left(egin{array}{c} oldsymbol{a} + oldsymbol{c} \mathrm{i} \ oldsymbol{b} + oldsymbol{d} \mathrm{i} \end{array}
ight)$$

Then $\Sigma_r \in P_{\mathbb{R}}$ can be written as

$$\mathbf{\Sigma}_{m{r}} = egin{pmatrix} \mathbf{\Sigma}_{\Re(m{z})} & \mathbf{\Sigma}_{\Re(m{z}),\Im(m{z})} \ \mathbf{\Sigma}_{\Im(m{z}),\Re(m{z})} & \mathbf{\Sigma}_{\Im(m{z})} \end{pmatrix}.$$

 $\Sigma_r \in P_{\mathbb{C}}$ is equivalent to complex propriety of z which is equivalent to \mathbb{C}^{j} -propriety of q as defined by Vía et al. (2010a). Also, $P_{\mathbb{C}} = P_{\mathbb{R}} \cap \left(\widetilde{\mathbb{H}^{p \times p}} + \widetilde{\mathbb{H}^{p \times p}} \widetilde{jI_p}^R\right)$ and $P_{\mathbb{H}} = P_{\mathbb{R}} \cap \widetilde{\mathbb{H}^{p \times p}}$. This section relies on the nestedness implied by the third row of Figure 2.1, where $\widetilde{\mathbb{H}^{p \times p}} + \widetilde{\mathbb{H}^{p \times p}} \widetilde{jI_p}^R$ was denoted $\widetilde{\mathbb{C}^{2n \times 2n}}^{\mathbb{C}}$.

Proposition 4.2. Let us consider the restriction of \hat{c}^{j} (Definition 1.43) to $P_{\mathbb{R}}$.

$$\hat{c}^{j}(\boldsymbol{\Sigma}_{\boldsymbol{r}}) = \frac{1}{2}\boldsymbol{\Sigma}_{\boldsymbol{r}} + \frac{1}{2}\tilde{\boldsymbol{j}}\widetilde{\boldsymbol{I}}_{p}^{R}\boldsymbol{\Sigma}_{\boldsymbol{r}}\tilde{\boldsymbol{j}}\widetilde{\boldsymbol{I}}_{p}^{RT}$$

$$= \frac{1}{2} \begin{pmatrix} \boldsymbol{\Sigma}_{\Re(\boldsymbol{z})} + \boldsymbol{\Sigma}_{\Im(\boldsymbol{z})} & \boldsymbol{\Sigma}_{\Re(\boldsymbol{z}),\Im(\boldsymbol{z})} - \boldsymbol{\Sigma}_{\Im(\boldsymbol{z}),\Re(\boldsymbol{z})} \\ \boldsymbol{\Sigma}_{\Re(\boldsymbol{z}),\Re(\boldsymbol{z})} - \boldsymbol{\Sigma}_{\Re(\boldsymbol{z}),\Im(\boldsymbol{z})} & \boldsymbol{\Sigma}_{\Re(\boldsymbol{z})} + \boldsymbol{\Sigma}_{\Im(\boldsymbol{z})} \end{pmatrix}, \quad (4.3)$$

where

$$\widetilde{\mathbf{jI}_p}^R = \left(egin{array}{cc} \mathbf{0}_{2p} & -\mathbf{I}_{2p} \ \mathbf{I}_{2p} & \mathbf{0}_{2p} \end{array}
ight)$$

Then $\hat{\Sigma}_{\mathbb{C}} = \hat{c}^{j}(\hat{\Sigma}_{\mathbb{R}})$ is the maximum likelihood estimator of Σ_{r} under $H_{0}^{\mathbb{C}}$.

Proof. It is clear from (4.3) that $\hat{\Sigma}_{\mathbb{C}}$ has complex structure, and from (4.2) that it is a convex combination of positive definite matrices and hence positive definite. Thus $\hat{\Sigma}_{\mathbb{C}} \in P_{\mathbb{C}}$.

We wish to maximise the normal likelihood function

$$(2\pi)^{-2pN} \left| \boldsymbol{\Sigma}_{\mathbb{C}} \right|^{-N/2} \exp\left(-\operatorname{tr}\left(\boldsymbol{\Sigma}_{\mathbb{C}}^{-1} \hat{\boldsymbol{\Sigma}}_{\mathbb{R}}\right) / 2\right)$$

$$(4.4)$$

over $\Sigma_{\mathbb{C}} \in P_{\mathbb{C}}$. Since $\hat{\Sigma}_{\mathbb{C}} = \hat{c}^{j}(\hat{\Sigma}_{\mathbb{R}})$ and $\Sigma_{\mathbb{C}}^{-1}$ has complex structure,² by Lemma 1.45 (4.4) is equal to

$$(2\pi)^{-2pN} |\mathbf{\Sigma}_{\mathbb{C}}|^{-N/2} \exp\left(-\operatorname{tr}\left(\mathbf{\Sigma}_{\mathbb{C}}^{-1} \hat{\mathbf{\Sigma}}_{\mathbb{C}}\right)/2\right).$$

Since the sample covariance is the maximum likelihood estimator in the unstructured real case, this is maximised over $\Sigma_{\mathbb{C}} \in P_{\mathbb{R}} \supset P_{\mathbb{C}}$ by setting $\Sigma_{\mathbb{C}} = \hat{\Sigma}_{\mathbb{C}}$. Since $\hat{\Sigma}_{\mathbb{C}} \in P_{\mathbb{C}}$ this is also the restricted solution.

(See also Andersson et al. (1983, eqn. 12, Theorem 1).)

Proposition 4.3. Let us consider the restriction of \hat{h} (Definition 1.42) to $P_{\mathbb{R}}$. Then $\hat{\Sigma}_{\mathbb{H}} = \hat{h}(\hat{\Sigma}_{\mathbb{R}}) = \hat{c}^{k}(\hat{\Sigma}_{\mathbb{C}})$ is the maximum likelihood estimator of Σ_{r} under $H_{0}^{\mathbb{H}}$.

Proof. By Proposition 1.47, $\hat{\Sigma}_{\mathbb{H}}$ has quaternion structure, and from (1.13) it is a convex combination of positive definite matrices. Hence $\hat{\Sigma}_{\mathbb{H}} \in P_{\mathbb{H}}$.

The rest of this proof is identical to the proof of Proposition 4.2, with \mathbb{C} replaced by \mathbb{H} , \hat{c}^{j} replaced by \hat{h} and Lemma 1.45 replaced by Lemma 1.46.

(See also Andersson et al. (1983, eqn. 56, Theorem 3).)

Let us now consider how the above fits with the ideas of Section 1.3.3. As we already discussed in Section 2.4, if we let $\hat{\Sigma}_{\boldsymbol{q}} = \sum_{i=1}^{N} \boldsymbol{q}_{i} \boldsymbol{q}_{i}^{H}$. Then since $\widetilde{\boldsymbol{q}_{i} \boldsymbol{q}_{i}^{H}} = 4\hat{h} \left(\mathcal{V}(\boldsymbol{q}_{1}) \mathcal{V}(\boldsymbol{q}_{i})^{T} \right)$

$$\hat{\boldsymbol{\Sigma}}_{\mathbb{H}} = \hat{h}(\hat{\boldsymbol{\Sigma}}_{\mathbb{R}}) = \frac{1}{4} \widetilde{\hat{\boldsymbol{\Sigma}}_{\boldsymbol{q}}}.$$

Similarly, if we let

$$\hat{\boldsymbol{\Sigma}}_{oldsymbol{z}} = \sum_{i=1}^N oldsymbol{z}_i oldsymbol{z}_i^H,$$

then

$$\hat{\boldsymbol{\Sigma}}_{\mathbb{C}} = \frac{1}{2} \begin{pmatrix} \Re(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{z}}) & -\Im(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{z}}) \\ \Im(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{z}}) & \Re(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{z}}) \end{pmatrix}$$

This is because a similar relationship exists between complex multiplication and the projection onto complex-structured matrices.

²Corollary 1.29 also applies to complex structure by Remark 1.34

Remark 4.4. If the mean $\boldsymbol{\mu}$ of a sample $\boldsymbol{r}_i \sim \mathcal{N}^{\mathbb{R}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_r)$ is unknown, then the MLE of $\boldsymbol{\mu}$ is $\hat{\boldsymbol{\mu}} = \frac{1}{N} \sum_{\ell=1}^{N} \boldsymbol{r}_{\ell}$, and the MLE of $\boldsymbol{\Sigma}_r$ is

$$\hat{\boldsymbol{\Sigma}}_{\mathbb{R}} = \frac{1}{N} \sum_{\ell=1}^{N} (\boldsymbol{r}_{\ell} - \hat{\boldsymbol{\mu}}) (\boldsymbol{r}_{\ell} - \hat{\boldsymbol{\mu}})^{T} \sim \mathcal{W}_{4p}^{\mathbb{R}} \left(\frac{1}{N} \boldsymbol{\Sigma}_{\boldsymbol{r}}, N - 1\right).$$

4.2.2 The LRT statistic and its moments

Proposition 4.5. The likelihood ratio (LR) for testing $H_0^{\mathbb{H}}$ versus $H_1^{\mathbb{H}}$ is

$$T_{\mathbb{H}} = \left(\frac{|\hat{\boldsymbol{\Sigma}}_{\mathbb{C}}|}{|\hat{\boldsymbol{\Sigma}}_{\mathbb{H}}|}\right)^{N/2}$$

Proof. The likelihood for a real normal sample is

$$(2\pi)^{-2pN} |\mathbf{\Sigma}_{\mathbf{r}}|^{-N/2} \exp\left(-\operatorname{tr}\left(\mathbf{\Sigma}_{\mathbf{r}}^{-1} \hat{\mathbf{\Sigma}}_{\mathbb{R}}\right)/2\right).$$

The LR is the ratio of maximum likelihoods, so by Propositions 4.2 and 4.3 it is given by

$$\left[\frac{|\hat{\boldsymbol{\Sigma}}_{\mathbb{C}}|}{|\hat{\boldsymbol{\Sigma}}_{\mathbb{H}}|}\right]^{N/2} \cdot \frac{\exp\left(-\operatorname{tr}\left(\hat{\boldsymbol{\Sigma}}_{\mathbb{H}}^{-1}\hat{\boldsymbol{\Sigma}}_{\mathbb{R}}\right)/2\right)}{\exp\left(-\operatorname{tr}\left(\hat{\boldsymbol{\Sigma}}_{\mathbb{C}}^{-1}\hat{\boldsymbol{\Sigma}}_{\mathbb{R}}\right)/2\right)}$$

By Corollary 1.29, $\hat{\Sigma}_{\mathbb{H}}^{-1} \in \widetilde{\mathbb{H}^{p \times p}}$, and by Remark 1.34, $\hat{\Sigma}_{\mathbb{C}}^{-1}$ similarly has complex structure. By Lemmas 1.45 and 1.46,

$$\operatorname{tr}\left(\hat{\Sigma}_{\mathbb{H}}^{-1}\hat{\Sigma}_{\mathbb{R}}\right) = \operatorname{tr}\left(\hat{\Sigma}_{\mathbb{H}}^{-1}\hat{\Sigma}_{\mathbb{H}}\right) = \operatorname{tr}\left(\boldsymbol{I}_{4p}\right) = \operatorname{tr}\left(\hat{\Sigma}_{\mathbb{C}}^{-1}\hat{\Sigma}_{\mathbb{C}}\right) = \operatorname{tr}\left(\hat{\Sigma}_{\mathbb{C}}^{-1}\hat{\Sigma}_{\mathbb{R}}\right),$$

so the exponential terms cancel.

Proposition 4.6. The LR for testing $H_0^{\mathbb{C}}$ versus $H_1^{\mathbb{C}}$ is

$$T_{\mathbb{C}} = \left(\frac{|\hat{\boldsymbol{\Sigma}}_{\mathbb{R}}|}{|\hat{\boldsymbol{\Sigma}}_{\mathbb{C}}|}\right)^{N/2}$$

Proof. This proof is almost identical to the proof of Proposition 4.5, with \mathbb{C} replaced by \mathbb{R} and \mathbb{H} replaced by \mathbb{C} . (Note that $\operatorname{tr}\left(\hat{\Sigma}_{\mathbb{R}}^{-1}\hat{\Sigma}_{\mathbb{R}}\right) = \operatorname{tr}\left(I_{4p}\right)$.)

(See also Walden and Rubin-Delanchy (2009, p. 828).)

Corollary 4.7. The LR for testing H_0 versus H_1 is

$$T = T_{\mathbb{C}} \cdot T_{\mathbb{H}} = \left(\frac{|\hat{\boldsymbol{\Sigma}}_{\mathbb{R}}|}{|\hat{\boldsymbol{\Sigma}}_{\mathbb{H}}|}\right)^{N/2}.$$
(4.5)

Proof. H_0 and $H_0^{\mathbb{H}}$ are equal. Since $\mathbb{P}(\hat{\Sigma}_{\mathbb{R}} \in P_{\mathbb{C}} \setminus P_{\mathbb{H}}) \leq \mathbb{P}(\hat{\Sigma}_{\mathbb{R}} \in P_{\mathbb{C}}) = 0$, the maximum likelihood under $H_1^{\mathbb{C}}$ and under H_1 are almost surely equal.³

Since T is a LR and $P_{\mathbb{H}} \subset P_{\mathbb{R}}$, we have $0 \leq T \leq 1$. We reject the null hypothesis for small values of T, or equivalently for large values of $M = -2\log(T)$.

From Andersson et al. (1983, Theorem 1), $T_{\mathbb{C}}$ and $\hat{\Sigma}_{\mathbb{C}}$ are independent under $H_0^{\mathbb{C}}$. $T_{\mathbb{H}}$ is a function of $\hat{\Sigma}_{\mathbb{C}}$ and hence $T_{\mathbb{H}}$ and $T_{\mathbb{C}}$ are independent. Now let us consider the moments of T. By independence $E\{T^h\} = E\{T^h_{\mathbb{C}}\}E\{T^h_{\mathbb{H}}\}$.

Proposition 4.8. Under H_0 the LRT statistic T for H_0 versus H_1 has moments

$$E\{T^{h}\} = K \prod_{j=1}^{\left\lceil \frac{3p}{2} \right\rceil} \frac{\Gamma[N(h+1) - 4p + 2j - 1]}{\Gamma\left[N(h+1) + \frac{2-j - \left\lfloor \frac{j-1}{3} \right\rfloor}{2}\right]},$$
(4.6)

where |x| is the integer part of x, [x] is the smallest integer greater or equal to x, and K does not depend on h.

Proof. See Appendix B.3.

Remark 4.9. By Remark 4.4, if the distribution of r had an unknown mean μ , then Proposition 4.8 would still hold, with N replaced by N-1.

As an immediate consequence of Proposition 4.8, we see that the LRT for quaternion propriety is of Box type (see Section 4.3.3 for details).

91

³Alternatively, we could argue that H_1 and $H_1^{\mathbb{C}}$ are equivalent since $P_{\mathbb{R}} \setminus P_{\mathbb{C}}$ and $P_{\mathbb{R}} \setminus P_{\mathbb{H}}$ have the same topological closure $P_{\mathbb{R}}$.

4.3 The distribution of statistics of Box type

Box (1949) describes various approximations to the distribution of a random variable

$$M = -2\log(W),$$

when the moments of W are products and ratios of gamma functions. W is typically a LRT statistic.

Definition 4.10. A random variable $0 \le W \le 1$ is said to be of Box type if

$$\mathbb{E}\left[W^{h}\right] = K\left(\frac{\prod_{j=1}^{k} y_{j}^{y_{j}}}{\prod_{i=1}^{m} x_{i}^{x_{i}}}\right)^{h} \frac{\prod_{i=1}^{m} \Gamma\left(x_{i}\left(1+h\right)+\xi_{i}\right)}{\prod_{j=1}^{k} \Gamma\left(y_{j}\left(1+h\right)+\eta_{j}\right)} \ \forall h \in \mathbb{N},$$
(4.7)

where K is such that $\mathbb{E}[W^0] = 1$ i.e.

$$K = \frac{\prod_{j=1}^{k} \Gamma\left(y_j + \eta_j\right)}{\prod_{i=1}^{m} \Gamma\left(x_i + \xi_i\right)},\tag{4.8}$$

and

$$\sum_{i=1}^{m} x_i = \sum_{j=1}^{k} y_j$$

$$x_i > 0 \quad \forall i$$

$$y_j > 0 \quad \forall j.$$

$$(4.9)$$

Remark 4.11. The assumption $W \leq 1$ is redundant. (See Appendix A.2)

Box's χ^2 expansion, detailed in Section 4.3.2.2, and our derivation of the exact distribution in Section 4.3.1 require the stronger assumption that (4.7) holds for all $h \in \mathbb{C}$ where the gamma functions are defined. In other words, we will make the additional assumption that the moment generating function of $M = -2\log(W)$ is

given by^4

$$\phi_M(s) = \mathbb{E}\left[e^{sM}\right] = K\left(\frac{\prod_{j=1}^k (2y_j)^{2y_j}}{\prod_{i=1}^m (2x_i)^{2x_i}}\right)^{-s} \frac{\prod_{i=1}^m \Gamma\left(x_i(1-2s) + \xi_i\right)}{\prod_{j=1}^k \Gamma\left(y_j(1-2s) + \eta_j\right)}$$
(4.10)

and is valid for $s \in \mathbb{C}$ except at a countable number of poles. In particular it is valid on the half plane $\Re(s) < s_0$ where $s_0 > 0$ is the smallest pole.⁵

Remark 4.12. By analytic extension, the additional assumption will hold whenever (4.7) or (4.10) holds for h or s on some interval.

Since W is bounded, the moments of W completely determine its distribution and hence also completely determine the distribution of M. Hence by uniqueness of characteristic functions, the additional assumption will hold whenever (4.10) is a valid characteristic function on the imaginary axis.

To the author's best knowledge, in all cases where random variables of Box type are considered in practice, (4.10) holds. The following proposition gives yet another way of checking (4.10).

Proposition 4.13. Assume W has moments given by (4.7) where m = k, $x_i = y_i \forall i$ and $\eta_i > \xi_i > -x_i \forall i$.⁶ Then (4.10) holds and W is distributed as a product of powers of independent beta random variables

$$\prod_{i=1}^{m} X_i^{x_i},\tag{4.11}$$

where $X_i \sim \beta(\xi_i + x_i, \eta_i - \xi_i)$.

⁴we just use $\sum x_i - \sum y_i = 0$ and rearrange in order to match (4.10) and (4.7)

⁶Note that one may need to reorder the parameters to satisfy this inequality.

 $^{{}^{5}0 &}lt; \mathbb{E}[W^{h}] \leq 1 \ \forall h > 0$, so if $s_{0} \leq 0$ we obtain a contradiction by taking $h \to -2s_{0} \geq 0$. However, note that we do not necessarily have $x_{i} + \xi_{i} > 0 \ \forall i$. Indeed there may hypothetically be positive removable singularities if a pole in the numerator and the denominator match. Such removable singularities would pose no theoretical problem.

Proof.

$$\mathbb{E}\left[X_i^{x_ih}\right] = \frac{\Gamma(x_i + \eta_i)\Gamma\left(\xi_i + x_i(1+h)\right)}{\Gamma(x_i + \xi_i)\Gamma\left(\eta_i + x_i(1+h)\right)} \ \forall h \in \mathbb{C} : \Re(h) > -\xi_i - x_i$$

Hence the moments of W and of (4.11) match and so does their distribution which is uniquely determined by the moments.

See also Mathai et al. (2009, p.122) or Anderson (1958, p.203). \Box

Definition 4.14. The degrees of freedom associated with W are

$$f = -2\left(\sum_{i=1}^{m} \xi_i - \sum_{j=1}^{k} \eta_j - \frac{m-k}{2}\right).$$
 (4.12)

Remark 4.15. $f \ge 0$. Also, f = 0 iff M has a mass at 0. (See Appendix A.2)

Proposition 4.16. The cumulants of M are

$$\kappa_{1} = 2\left(\sum_{i=1}^{m} x_{i} \log(x_{i}) - \sum_{i=1}^{k} y_{r} \log(y_{r}) - \sum_{i=1}^{m} x_{i} \psi\left(x_{i} + \xi_{i}\right) + \sum_{i=1}^{k} y_{r} \psi\left(y_{r} + \eta_{r}\right)\right)$$
$$\kappa_{j} = \sum_{i=1}^{m} (-2x_{i})^{j} \psi^{(j-1)}(x_{i} + \xi_{i}) - \sum_{i=1}^{k} (-2y_{i})^{j} \psi^{(j-1)}(y_{i} + \eta_{i}), \ j \ge 2,$$

where $\psi(x) = \frac{d \log \Gamma(x)}{dx}$ is the digamma function, and its derivatives $\psi^{(j)}(x)$ are polygamma functions.

Proof. As in Jensen (1991), these are obtained directly by differentiating the cumulant generating function $\log \phi_M(s)$, since $\kappa_j = \frac{d^j \log \phi_M(s)}{ds^j}\Big|_{s=0}$.

Remark 4.17. $\kappa_1 = \mathbb{E}[M]$ is the mean, $\kappa_2 = \operatorname{Var}(M)$ is the variance, $\kappa_3/\kappa_2^{3/2}$ is the skewness and κ_4/κ_2^2 is the excess kurtosis.

Lemma 4.18. The following are well known properties of the gamma function

$$\Gamma(n) = (n-1)! \ \forall n \in \mathbb{N}, n > 0$$

$$\Gamma(z+1) = z\Gamma(z)$$

$$\Gamma(z)\Gamma\left(z+\frac{1}{2}\right) = \sqrt{\pi}2^{1-2z}\Gamma(2z)$$
(4.13)

$$\prod_{k=0}^{n-1} \Gamma\left(z+\frac{k}{n}\right) = (2\pi)^{(n-1)/2} m^{1/2-nz} \Gamma(nz).$$
(4.14)

Proposition 4.19. If in (4.7) the x_i and y_j are rational $\forall i, j$, there is an alternate parameterisation of (4.7) which satisfies m = k and $x_i = y_j \forall i, j$.

Proof. Write all x_i, y_j as fractions with a common positive denominator $d \in \mathbb{N}$. Each $\Gamma(x_i(1+h)+\xi_i) = \Gamma\left(dx_i \frac{x_i(1+h)+\xi_i}{dx_i}\right)$ can be expanded into a product of dx_i terms of the form $\Gamma\left(\frac{1}{d}(1+h)+\frac{\xi_i+k}{dx_i}\right)$ using (4.14), and similarly for $\Gamma(y_i(1+h)+\eta_i)$. \Box

4.3.1 Exact distribution

We now proceed to show that both the PDF and CDF of random variables of Box type can be given in terms of Fox's *H*-function (or in simpler cases Meijer's *G*-function).

Pham-Gia (2008) expresses the density of the generalised Wilks' statistic in terms of H-functions and G-functions. The density of an arbitrary product of independent beta random variables is also given. Special cases had already been treated for small dimensions, where the H function reduces to simpler functions. For example with Votaw's criterion⁷ (Consul, 1969) and with the likelihood ratio test for sphericity (Consul, 1967). The exact densities of products of powers of independent gamma and beta random variables are given as H functions by Mathai et al. (2009). More generally, we will show that this can be done with random variables of Box type.

Note that the class of densities which can be expressed as H-functions is even more general than Box type, since products and ratios of some random variables which are

⁷Votaw's criterion is for testing whether the distribution of a p+q dimensional normal is invariant under permutations of the first p or last q indices. It is thus a LRT for group invariance structure.

not necessarily of Box type⁸ can be written as H-functions with equal ease, see Carter and Springer (1977); Mathai et al. (2009); Springer and Thompson (1970).

Definition 4.20. Fox's H-function is defined by the following Mellin-Barnes integral (Carter and Springer, 1977)

$$H_{p,q}^{m,n}\left(z \left| \begin{array}{c} (a_{1},\alpha_{1}),\ldots,(a_{p},\alpha_{p}) \\ (b_{1},\beta_{1}),\ldots,(b_{q},\beta_{q}) \end{array} \right) \\ = \frac{1}{2\pi i} \int_{L} \frac{\prod_{j=1}^{m} \Gamma(b_{j}-\beta_{j}s)}{\prod_{j=m+1}^{q} \Gamma(1-b_{j}+\beta_{j}s)} \frac{\prod_{j=1}^{n} \Gamma(1-a_{j}+\alpha_{j}s)}{\prod_{j=n+1}^{p} \Gamma(a_{j}-\alpha_{j}s)} z^{s} ds, \qquad (4.15)$$

where the path of integration L is chosen such that the poles $\frac{b_j+k}{\beta_j}$, $j = 1, \ldots, m$, $k \in \mathbb{N}$ lie on the right and the poles $\frac{a_j-1-k}{\alpha_j}$, $j = 1, \ldots, n$, $k \in \mathbb{N}$ lie on the left. The parameters are $0 \le n \le p$, $0 \le m \le q$, $\alpha_i, \beta_j \ge 0$, $a_i, b_j \in \mathbb{C}$.

Remark 4.21. (4.15) is an inverse Mellin transform.

Remark 4.22. The choice of branch cut of the logarithm in $z^s = e^{s \log(z)}$ $(z \neq 0)$ determines the choice of branch cut for the *H* function. We will however only need to work with $z \in [0, \infty[$ and the principal value of the logarithm.

Remark 4.23. In all our uses, we will have n = 0, m = q, and the path of integration L will be a vertical line from $\gamma - i\infty$ to $\gamma + i\infty$.

Definition 4.24. Meijer's G-function is a special case of the H-function

$$G_{p,q}^{m,n}\left(z \middle| \begin{array}{c} a_1, \dots, a_p \\ b_1, \dots, b_q \end{array}\right) = H_{p,q}^{m,n}\left(z \middle| \begin{array}{c} (a_1, 1), \dots, (a_p, 1) \\ (b_1, 1), \dots, (b_q, 1) \end{array}\right).$$

Proposition 4.25. The following 3 expressions are equal for arbitrary $c \in \mathbb{C}, \lambda > 0$

⁸With $x_i < 0$ for some i

Chapter 4. Likelihood Ratio Testing for Quaternion-Structured Covariance Matrices

(Carter and Springer, 1977), (Mathai et al., 2009, p.12)

$$H_{p,q}^{m,n}\left(z \begin{vmatrix} (a_{1},\alpha_{1}),\ldots,(a_{p},\alpha_{p})\\ (b_{1},\beta_{1}),\ldots,(b_{q},\beta_{q}) \end{vmatrix}\right)$$
$$\lambda H_{p,q}^{m,n}\left(z^{\lambda} \begin{vmatrix} (a_{1},\lambda\alpha_{1}),\ldots,(a_{p},\lambda\alpha_{p})\\ (b_{1},\lambda\beta_{1}),\ldots,(b_{q},\lambda\beta_{q}) \end{vmatrix}\right)$$
(4.16)

$$z^{-c}H_{p,q}^{m,n}\left(z \left|\begin{array}{c} (a_1+c\alpha_1,\alpha_1),\ldots,(a_p+c\alpha_p,\alpha_p)\\ (b_1+c\beta_1,\beta_1),\ldots,(b_q+c\beta_q,\beta_q) \end{array}\right).$$
(4.17)

Remark 4.26. All *H* functions with rational α_i, β_j can be written as *G* functions of the form

$$dG_{p,q}^{m,n}\left(z^d \middle| \begin{array}{c} a_1,\ldots,a_p \\ b_1,\ldots,b_q \end{array}\right)$$

Proof. This follows from the proof of Proposition 4.19 and (4.16).

The theory surrounding G and H functions is expressed most conveniently in terms of Mellin transforms. We have however chosen to use equivalent Fourier transforms for familiarity.

Theorem 4.27. Let M satisfy (4.10). Then the pdf of M is given by

$$f_M(x) = KH_{k,m}^{m,0} \left(\frac{\prod_{i=1}^m (2x_i)^{2x_i}}{\prod_{j=1}^k (2y_j)^{2y_j}} e^{-x} \middle| \begin{array}{c} (y_1 + \eta_1, 2y_1), \dots, (y_k + \eta_k, 2y_k) \\ (x_1 + \xi_1, 2x_1), \dots, (x_m + \xi_m, 2x_m) \end{array} \right)$$
(4.18)

on x > 0. In particular when $\alpha = x_i = y_j \ \forall i, j \ and \ m = k$ This simplifies to

$$f_M(x) = \frac{K e^{-\frac{x}{2}}}{2\alpha} G_{m,m}^{m,0} \left(e^{\frac{-x}{2\alpha}} \middle| \begin{array}{c} \eta_1, \dots, \eta_m \\ \xi_1, \dots, \xi_m \end{array} \right).$$
(4.19)

The constant K is defined in (4.8).

Proof. To obtain (4.18), simply notice that the integral for inverting the characteristic function of M

$$f_M(x) = \frac{1}{2\pi i} \int_{-i\infty}^{+i\infty} e^{-sx} \phi_M(s) ds$$
(4.20)

is of the form of (4.15). (4.19) is then obtained by applying (4.16) with $\lambda = \frac{1}{2\alpha}$ followed by (4.17) with c = -N.

Remark 4.28. Applying Mathai et al. (2009, Theorem 1.1 p.4), the integral (4.20) converges for $x \neq 0$. Proposition A.2 gives the tail of the characteristic function as $\phi_M(s) = O\left(|s|^{-\frac{f}{2}}\right)$. Hence if f > 2 the characteristic function will be absolutely integrable and the density will be uniformly continuous on \mathbb{R} . The condition f > 2 is consistent with the fact that the χ_1^2 and χ_2^2 densities are discontinuous at 0.

Since $M = -2\log(W)$, by change of variables we have

$$f_W(x) = \frac{2}{x} f_M(-2\log(x)).$$

In particular, when (4.19) holds

$$f_W(x) = \frac{K}{\alpha} G_{m,m}^{m,0} \left(x^{\frac{1}{\alpha}} \middle| \begin{array}{c} \eta_1, \dots, \eta_m \\ \xi_1, \dots, \xi_m \end{array} \right)$$

Theorem 4.29. Let M satisfy (4.10), and assume f > 0. Then the CDF of M is given by

$$F_M(x) = KH_{k+1,m+1}^{m+1,0} \left(\frac{\prod_{i=1}^m (2x_i)^{2x_i}}{\prod_{j=1}^k (2y_j)^{2y_j}} e^{-x} \middle| \begin{array}{c} (y_1 + \eta_1, 2y_1), \dots, (y_k + \eta_k, 2y_k), (1,1) \\ (x_1 + \xi_1, 2x_1), \dots, (x_m + \xi_m, 2x_m), (0,1) \end{array} \right).$$

$$(4.21)$$

In particular, when $\alpha = x_i = y_j \ \forall i, j \ and \ m = k \ this \ simplifies \ to$

$$F_M(x) = K e^{-\frac{x}{2}} G_{m+1,m+1}^{m+1,0} \left(e^{-\frac{x}{2\alpha}} \middle| \begin{array}{c} \eta_1, \dots, \eta_m, 1-\alpha \\ \xi_1, \dots, \xi_m, -\alpha \end{array} \right).$$
(4.22)

Proof. Choose some arbitrary $\gamma < 0$ and some arbitrary $\alpha_0 > 0$. Then integrating

Chapter 4. Likelihood Ratio Testing for Quaternion-Structured Covariance Matrices

the PDF from 0 to x or using Lévy's inversion formula (Loeve, 1977, p. 199) gives

$$F_M(x) = \frac{1}{2\pi i} \int_{-i\infty}^{+i\infty} \left[\frac{1}{s} - \frac{e^{-sx}}{s} \right] \phi_M(s) ds$$

$$= \frac{1}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} \left[\frac{2\alpha_0 e^{-sx}}{-2\alpha_0 s} - \frac{1}{-s} \right] \phi_M(s) ds$$

$$= \frac{2\alpha_0}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} e^{-sx} \frac{\Gamma(-2\alpha_0 s)}{\Gamma(1-2\alpha_0 s)} \phi_M(s) ds - \frac{1}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} \frac{1}{-s} \phi_M(s) ds.$$

The path can be shifted by γ since no poles are crossed. Consider the first integral. It is of the form of (4.15) and hence it is equal to

$$2\alpha_0 K H_{k+1,m+1}^{m+1,0} \left[\frac{\prod_{i=1}^m (2x_i)^{2x_i}}{\prod_{j=1}^k (2y_j)^{2y_j}} e^{-x} \middle| \begin{array}{c} (y_1 + \eta_1, 2y_1), \dots, (y_k + \eta_k, 2y_k), (1, 2\alpha_0) \\ (x_1 + \xi_1, 2x_1), \dots, (x_m + \xi_m, 2x_m), (0, 2\alpha_0) \end{array} \right].$$

By choosing $\alpha_0 = \frac{1}{2}$, this gives us (4.21).

By choosing instead $\alpha_0 = \alpha$, (4.22) is obtained by applying (4.16) with $\lambda = \frac{1}{2\alpha}$ followed by (4.17) with $c = -\alpha$.

Using (A.1), the tail of the second integrand is $O\left(|s|^{-1-\frac{f}{2}}\right)$. Since f > 0, the integral converges. Taking $\gamma \to -\infty$, the second integral goes to 0, hence it is equal to 0. As with Remark 4.28, f > 0 ensures that the CDF will be uniformly continuous.⁹

Since $M = -2\log(W)$, by change of variables we have

$$F_W(x) = F_M(-2\log(x)).$$

In particular, when (4.22) holds

$$F_W(x) = K x G_{m+1,m+1}^{m+1,0} \left(x^{\frac{1}{\alpha}} \middle| \begin{array}{c} \eta_1, \dots, \eta_m, 1-\alpha \\ \xi_1, \dots, \xi_m, -\alpha \end{array} \right).$$

⁹The condition f > 0 is consistent with Remark 4.15

4.3.1.1 Numerical evaluation of the exact distribution

As noted by Mathai (1973a), (4.21) and (4.18) are simply statements about the moment generating function of M unless values of the G-function can be computed. One possible method is to write the G-function as a sum of generalised hypergeometric functions using Slater's theorem, however this cannot be used in our case because the Mellin-Barnes integral will have non-simple poles (Marichev, 1983, pp. 56-58 & 66-67).

A general algorithm for the numerical evaluation of Meijer's G function by summing over the residues is described in Liakhovetski (2001).¹⁰ See also Cook (1981); Springer (1987). Mathai (1973b) reviews various methods for computing the exact distributions of products of independent beta or gamma random variables. Notable amongst these is the "method of calculus of residues" which expresses the PDF (or CDF) of W as a (possibly infinite) sum of terms of the form $a_i x^{b_i} \log(x)^{c_i}$. This is equivalent to expressing the distribution of M as a series of gamma distributions. Dennis (1994) describes an algorithm to apply this method in general to products of independent beta random variables. In essence, the method is equivalent to the algorithm of Liakhovetski (2001).

Specific likelihood ratio tests for which this method has been applied include Wilks' criterion¹¹ (Schatzoff, 1966), the complex Wilks' criterion (Gupta, 1971), circular symmetry (Nagar et al., 2004), diagonality (uncorrelatedness) (Mathai and Katiyar, 1979), sphericity given diagonality (Mathai, 1979) and more (Mathai, 1972).

As noted in Schatzoff (1966), numerical evaluation of the series must be performed to many extra significant digits, since it suffers from large cancellation errors.

Modern symbolic computation engines such as Maple, Mathematica and MuPad (Matlab) have arbitrary-precision implementations of Meijer's G-function. However, this does not make numerical evaluation of the G-function trivial, since we have encountered problems with each of these implementations. Maple(v13.0) improperly applies Mathai et al. (2009, property 1.6 p. 12) (see Mathai et al. (2009, note 1.6))

¹⁰It assumes that the path of integration is a loop from $-\infty$ to $-\infty$. This is a valid choice of path for 0 < x.

¹¹When either p or q is even.

for some values of the parameters. This relies on making the substitution

$$\frac{\Gamma(s-\alpha)}{\Gamma(s+1-\alpha)} = \frac{1}{s-\alpha} = \frac{-1}{\alpha-s} = -\frac{\Gamma(\alpha-s)}{\Gamma(1+\alpha-s)}$$

which moves the pole at α from one group to the other and hence changes the implied path of integration. This leads to (4.22) being incorrectly evaluated to $F_M(x) - 1$ instead of $F_M(x)$.

Matlab(v7.9)'s symbolic engine MuPad on the other hand, provides inaccurate results for $m \gtrsim 10$, even when computation is performed to many extra significant digits. Finally, Mathematica(v6.0.1) was found to be significantly slower than the other options.¹² To compute exact quantiles for Section 4.3.3, we have used Maple, correcting for the misallocated residue at α when appropriate. Still, numerical inversion of the *G*-function to obtain quantiles is impractically slow for large p.¹³ These hurdles motivate us to consider approximations to the CDF in the next section.

4.3.2 Approximations

In this section we will describe various approximations to the distribution of $M = -2 \log W$. In Section 4.3.3 we will compare their accuracy by applying them to the LRT for quaternion propriety.

4.3.2.1 Asymptotic distribution

Proposition 4.30. If for all i, j we let $x_i, y_j \to \infty$, then M is asymptotically distributed as χ^2_f , where f is given by (4.12).

Proof. By applying Stirling's approximation (A.2), we can show

$$\frac{\Gamma(x_i(1+h)+\xi_i)}{x_i^{x_ih}\Gamma(x_i+\xi_i)} = (1+h)^{x_i(1+h)+\xi_i}(1+o(1)),$$

 $^{12}(\sim 50\times)$ slower

 $^{^{13}}$ It is the author's opinion that due to the importance of G-functions in symbolic computation, software for its evaluation will improve promptly.

and similarly for terms in y_j , η_j . By taking a product of such terms, simplifying with (4.9), and substituting h = -2s,

$$\phi_M(s) = (1 - 2s)^{-\frac{1}{2}}(1 + o(1))$$

Hence the characteristic function of M converges to the characteristic function of a χ_f^2 distribution. The proposition follows by Levy's continuity theorem.

4.3.2.2 Box's chi-squared series

Box (1949) obtains an asymptotic expansion of the distribution of ρM as a series of χ^2 distributions, for some arbitrary $\rho \geq 0$. Gupta and Tang (1988) shows that the asymptotic series also converges to the true distribution when the number of terms taken tends to infinity, except in the right tail. Thus, the Box series can in principle be used to compute (most of) the distribution to arbitrary precision.

Definition 4.31 (Bernoulli Polynomial). The Bernoulli polynomials $B_n(x)$ are given by

$$B_n(x) = \sum_{k=0}^n \binom{n}{k} B_{n-k} x^k ,$$

where $B_n = B_n(0)$ are the Bernoulli numbers.

$$\mathbf{B}_{n} = \begin{cases} 1 & \text{if } n = 0\\ \frac{1}{2} & \text{if } n = 1\\ 0 & \text{if } n > 1, n \text{ odd}\\ (-1)^{\frac{n}{2} + 1} \frac{2 \cdot n!}{(2\pi)^{n}} \zeta(n) & \text{if } n > 1, n \text{ even} \end{cases}$$

and $\zeta(z)$ is the Riemann zeta function. The first polynomials are $B_0(x) = 1$, $B_1(x) = x - \frac{1}{2}$, $B_2(x) = x^2 - x + \frac{1}{6}$, $B_3(x) = x^3 - \frac{3x^2}{2} + \frac{x}{2}$.

Lemma 4.32 (Barnes (1899, p.121)).

$$\log \Gamma(z+h) = \left(z+h-\frac{1}{2}\right)\log(z) - z + \frac{1}{2}\log(2\pi) + \sum_{j=1}^{n} (-1)^{j+1} \frac{B_{j+1}(h)}{j(j+1)z^j} + R_{n+1}^{**}$$
(4.23)

where

$$\left|R_{n+1}^{**}\right| = \left|\int_{-n-\frac{1}{2}-i\infty}^{-n-\frac{1}{2}+i\infty} \frac{\zeta(s,h)z^{s}i}{2s\sin(\pi s)} ds\right| = O(|z|^{-n-1}).$$
¹⁴

Theorem 4.33. Let

$$\omega_j = \frac{(-1)^{j+1}}{j(j+1)} \left[\sum_{i=1}^m \frac{\mathbf{B}_{j+1} \left((1-\rho)x_i + \xi_i \right)}{(\rho x_i)^j} - \sum_{i=1}^k \frac{\mathbf{B}_{j+1} \left((1-\rho)y_i + \eta_i \right)}{(\rho y_i)^j} \right], \quad (4.24)$$

and let a_j be the coefficient of t^j in the series expansion of $\exp\left(\sum_{j=1}^n \omega_j t^j\right)$. Then

$$f_M(x) = \rho K_B \sum_{j=0}^n a_j f_{\chi^2_{f+2j}}(\rho x) + \mathcal{O}(x_0^{-n-1}), \qquad (4.25)$$

$$F_M(x) = K_B \sum_{j=0}^n a_j F_{\chi^2_{j+2j}}(\rho x) + \mathcal{O}(x_0^{-n-1}), \qquad (4.26)$$

where

$$\log(K_B) = -\sum_{j=1}^n \omega_j + \mathcal{O}(x_0^{-n-1}).$$

Proof. This is a result of Box (1949). We present his derivation in Appendix B.4. \Box *Remark* 4.34. Typically, x_0 is proportional to the sample size. *Remark* 4.35. If m = k and $\alpha = x_i = y_j \forall i, j$, then

$$K_B = K(\rho\alpha)^{-\frac{J}{2}}.$$

An asymptotic expansion will usually be a divergent series. If so, for fixed x_0 , there is a finite number $(O(x_0))$ of terms after which adding more terms decreases

¹⁴Whittaker and Watson (1927, pp.277-278) only give $|R_{n+1}^{**}| = O(|z|^{-n-\frac{1}{2}})$. However the stronger result of Barnes (1899) is correct.

the accuracy of the approximation. Indeed, (4.23) diverges when $n \to \infty$, as pointed out by Bayes (1763).

Gupta and Tang (1988) provide an efficient iterative scheme to compute the a_j from the ω_j :¹⁵

$$a_0 = 1$$

$$a_j = \frac{1}{j} \sum_{i=1}^j i\omega_i a_{j-i} \quad \forall j > 0.$$

This is essentially a reformulation of the classic iterative scheme for computing the moments of a distribution from its cumulants. Gupta and Tang (1988) also show that the series (4.25) (and hence also (4.26)¹⁶) is convergent as $n \to \infty$ for fixed $x < 4\pi x_0$. The proof relies implicitly on the fact that $\phi_M(s)$ is an analytic function, and as a result its asymptotic expansion can be differentiated term by term (Estrada and Kanwal, 1994, Theorem 10 p. 25).

By choosing a suitable value for ρ , the terms in the series can be made to decrease faster. As suggested by Box (1949), we will choose ρ such that $\omega_1 = a_1 = 0$. Gleser and Olkin (1975) points out that this corresponds to

$$\rho = 1 - \frac{1}{f} \left(\sum_{i=1}^{m} \frac{B_2(\xi_i)}{x_i} - \sum_{j=1}^{k} \frac{B_2(\eta_j)}{y_j} \right).$$
(4.27)

Numerical tests show that, roughly speaking, choosing a ρ larger than (4.27) will lead to a slower decay in the coefficients, whereas choosing a smaller ρ will introduce oscillations between positive and negative a_j for small j, and thus potential loss of numerical accuracy.

Another version of the χ^2 expansion, used by Anderson (1958), is obtained by further expanding K_B and collecting terms of equal order. However, this greatly

¹⁵The formula is given twice, the first time there is a typo.

¹⁶We can write (4.25) as $e^{-\frac{x}{2}}$ times a power series in x, and convergent power series are uniformly convergent on compact subsets of their disc of convergence. Thus (4.25) is uniformly convergent on the range of integration [0, x].

complicates the expansion when n is large. Box (1949) suggests either computing K_B exactly or approximating it by truncating the infinite sum in (B.8).

Another option would be to set K_B so that

$$K_B \sum_{j=0}^{n} a_j = 1. (4.28)$$

This was the choice made by e.g. Conradsen et al. (2003), who used a Box series with n = 3 to approximate the distribution of the LRT for equality of two complex covariance matrices.¹⁷ This simpler choice ensures that the approximate CDF (4.26) goes to 1 as $x \to \infty$, and possibly defines a valid CDF.¹⁸ This is particularly important if we wish to invert it numerically to find quantiles. Alternatively, Davis (1971) develops an analytically inverted version of Box's series for computing quantiles directly.

4.3.2.3 Bartlett adjustments

We wish to approximate the distribution of M by that of a random variable of the form $C\chi_f^2$, where C is chosen so that the cumulants of $C\chi_f^2$ match those of M up to an error of order $O(N^{-2})$.

Box's constant is obtained by computing from $(4.24)^{19}$

$$A_1 = \frac{2\omega_1}{f}\Big|_{\rho=1}$$
$$A_2 = \frac{4\omega_2}{f}\Big|_{\rho=1}$$

¹⁷This corresponds to the complex case of test (f) of Andersson et al. (1983)

¹⁸If there are some negative a_j , we are taking a non-convex discrete mixture of χ^2 distributions and would need to prove that the pdf is non-negative. This will not be the case for example when the last coefficient is negative.

¹⁹The criterion $0 < A_1^2 - 2A_2$ isn't given explicitly in Box (1949), but rather a heuristic distinction is made between when $A_2 \approx 0$ and when $A_2 - A_1^2 \gtrsim 0$

$$C_{\text{box}} = \begin{cases} 1 + A_1 & \text{if } 0 < A_1^2 - 2A_2 \\ (1 - A_1)^{-1} & \text{otherwise} \end{cases}.$$
 (4.29)

Remark 4.36. Choosing $C_{\text{box}} = (1 - A_1)^{-1}$ is equivalent to using Box's χ^2 series of order $O(N^{-2})$ with ρ given by (4.27)= $1 - A_1$ and $K_B = 1$.

A more accurate approximation is obtained by fitting the mean exactly (Jensen, 1991; Møller, 1986), i.e. choosing

$$C_{\text{exact}} = \frac{\kappa_1}{f}.$$

 $C_{\text{box}} = 1 + A_1$ and C_{exact} correspond to the Bartlett adjustments b_1 and b_3 of Møller (1986) respectively. Their accuracy is compared numerically, along with a more complicated Bartlett adjustment b_2 , the performance of which is in between that of b_1 and b_3 .

Møller (1986) claims that C_{exact} gives a O $\left(N^{-\frac{3}{2}}\right)$ approximation to the density. It is clear from Remark 4.36 that the Bartlett adjustments considered actually yield O(N^{-2}) approximations to the exact distribution. Barndorff-Nielsen and Hall (1988) show that this is indeed the case under general conditions for Bartlett adjustments of likelihood ratio criteria.

4.3.2.4 Box's F approximation

Box (1949) improves on the χ^2 approximation of section 4.3.2.3 by using the Pearson system of curves along with asymptotic approximations. For $A_2 - A_1^2 > 0$, Box finds that M is approximately distributed as $b \cdot F(f, f_2)$ where

$$f_2 = \frac{f+2}{A_2 - A_1^2} \tag{4.30}$$

$$b = \frac{f}{1 - A_1 - \frac{f_1}{f_2}}.$$
(4.31)

Note that f_2 need not be an integer, hence we are technically generalising the F

distribution to non-integer parameters.

For $A_2 - A_1^2 < 0$, Box finds that M is approximately distributed as $b \cdot \beta(\frac{f}{2}, \frac{f_2}{2})$ where

$$f_2 = \frac{f+2}{A_1^2 - A_2}$$

$$b = \frac{f_2}{1 - A_1 + \frac{2}{f_2}}$$

With these approximations, the first four cumulants are fitted up to an error of order $O(N^{-3})$.

A numerical study by Foerster and Stemmler (1990) establishes how large N must be to obtain accurate F approximations in the test for equality of covariance matrices.

4.3.2.5 A new F approximation

As with the $C_{\text{exact}}\chi^2$ approximation, we can improve on Box's F approximation by fitting the first three cumulants exactly, i.e. by following the steps in Box (1949) without taking asymptotic approximations. We first compute the exact value of Box's discriminant $\tau = \frac{\kappa_1 \kappa_3}{2\kappa_2^2}$. Then if $\tau > 1$ use a F distribution and if $\tau < 1$ use a beta distribution. For the $bF(f_1, f_2)$ distribution, the fitted parameters are²⁰

$$f_1 = \frac{4\kappa_1 \left(\kappa_1^2 \kappa_2 - \kappa_2^2 + \kappa_1 \kappa_3\right)}{4\kappa_1 \kappa_2^2 - \kappa_1^2 \kappa_3 + \kappa_2 \kappa_3}$$
(4.32)

$$f_2 = \frac{4\kappa_1^2\kappa_2 - 8\kappa_2^2 + 6\kappa_1\kappa_3}{\kappa_1\kappa_3 - 2\kappa_2^2}$$
(4.33)

$$b = \frac{2\kappa_1 \left(\kappa_1^2 \kappa_2 - \kappa_2^2 + \kappa_1 \kappa_3\right)}{2\kappa_1^2 \kappa_2 - 4\kappa_2^2 + 3\kappa_1 \kappa_3}.$$
(4.34)

²⁰Assuming $f_2 > 6$, so that the cumulants exist.

For the $bBeta\left(\frac{f_1}{2}, \frac{f_2}{2}\right)$ distribution, we have

$$f_{1} = \frac{4\kappa_{1} (\kappa_{1}^{2}\kappa_{2} - \kappa_{2}^{2} + \kappa_{1}\kappa_{3})}{4\kappa_{1}\kappa_{2}^{2} - \kappa_{1}^{2}\kappa_{3} + \kappa_{2}\kappa_{3}}$$

$$f_{2} = \frac{4\kappa_{2} (2\kappa_{1}\kappa_{2} + \kappa_{3}) (\kappa_{1}^{2}\kappa_{2} - \kappa_{2}^{2} + \kappa_{1}\kappa_{3})}{(\kappa_{1}\kappa_{3} - 2\kappa_{2}^{2}) (\kappa_{1}^{2}\kappa_{3} - 4\kappa_{1}\kappa_{2}^{2} - \kappa_{2}\kappa_{3})}$$

$$b = \frac{\kappa_{1}^{2}\kappa_{3} - 4\kappa_{1}\kappa_{2}^{2} - \kappa_{2}\kappa_{3}}{\kappa_{1}\kappa_{3} - 2\kappa_{2}^{2}}.$$

Note that some values of the cumulants will yield invalid negative parameters. This happens when the Pearson curve to be fitted is neither F nor β . For example, to approximate the LRT for quaternion propriety of Section 4.2 with p = 4, N = 16, a Pearson type IV distribution should be used. This problem will not arise if N is sufficiently large. For the LRT for quaternion propriety for example, $N \ge 4.13p + 0.5$ is a sufficient condition.²¹ See Craig (1936) for more details on the Pearson system of curves.

Remark 4.37. $\kappa_1 = \mathbb{E}[M], \kappa_2 = \operatorname{Var}(M)$ and κ_3 can be calculated with Proposition 4.16.

Remark 4.38. Since M is asymptotically χ_f^2 , when an F distribution is fitted, $f_2 \to \infty$ for large samples. We have found that for large f_2 , the implementation of the $F \text{ CDF}^{22}$ in GSL (Galassi et al., 2009) is more numerically accurate than the implementation in Matlab. Hence we will use it in our numerical evaluation.

4.3.2.6 Gamma approximation

Jensen (1991) suggests approximating the distribution of M with a $\Gamma(\lambda, \theta)$ distribution

$$f_{\Gamma(\lambda,\theta)}(z) = \theta^{-\lambda} \Gamma(\lambda)^{-1} z^{\lambda-1} e^{-\frac{z}{\theta}}$$

 $^{^{21}{\}rm This}$ conclusion was reached by a numerical study of all cases $p\leq 10\,000,\,N\leq 4p+2\,000.$ Note that $N\geq 4p$ is assumed.

 $^{^{22}}$ More specifically, the implementation of the incomplete beta function, from which the F CDF is computed.
by fitting the first two cumulants exactly, i.e. choosing

$$\lambda = \frac{\kappa_1^2}{\kappa_2} \tag{4.35}$$

$$\theta = \frac{\kappa_2}{\kappa_1}.\tag{4.36}$$

This approach amounts to treating the degrees of freedom in the χ^2 approximation as a free parameter.

4.3.2.7 Large deviation saddlepoint estimate

Consider the exponentially tilted²³ random variable M_s with density

$$f_{M_s}(x) = \frac{\mathrm{e}^{sx} f_M(x)}{\phi_M(s)}$$

Then

$$\phi_{M_s}(t) = \frac{\phi_M(t+s)}{\phi_M(s)}$$

Thus M_s is also of Box type, with parameters

$$\begin{aligned} \xi_{si} &= -2x_i s + \xi_i, \\ \eta_{sj} &= -2y_j s + \eta_j. \end{aligned}$$

The valid range of s is $s < s_0$ where $s_0 > 0$ is the leftmost pole of $\phi_M(s)$. Since $f_M(x) = \phi_M(s) e^{-sx} f_{M_s}(x)$, we can obtain an approximation to $f_M(x)$ by choosing a suitable s and approximating the tilted density $f_{M_s}(x)$ instead.

Jensen (1991) chooses s such that $x = \mathbb{E}[M_s]$. The tilted density is then approximated using the gamma approximation of Section 4.3.2.6. The approximate $f_M(x)$ is then integrated (keeping s fixed) to get an approximation for $F_M(x)$. The corrected

109

 $^{^{23}}$ The exponentially tilted family is also called conjugate family of distributions. Note however that there is no relationship with the use of the term in Bayesian statistics.

formula, given in Jensen (1995) is

$$F_M(x) \approx \phi_M(s) \left(1 + \frac{s\sigma_s^2}{\mu_s}\right)^{-\lambda} F_{\Gamma(\lambda,1)}\left(\mu_s s + \lambda\right), \qquad (4.37)$$

where

$$\mu_s = \mathbb{E}[M_s] = x$$

$$\sigma_s^2 = \operatorname{Var}[M_s]$$

$$\lambda = \frac{\mu_s^2}{\sigma_s^2}.$$

 μ_s and σ_s^2 can be calculated with Proposition 4.16. Solving $\mu_s = x$ for s must in general be done numerically. Since $\sigma_s^2 = \frac{d\mu_s}{ds} > 0$ the solution is unique and the gradient is easily computable. Jensen (1991) suggests using the Newton-Raphson method.

4.3.2.8 Lugannani & Rice saddlepoint approximation

The Lugannani & Rice saddlepoint approximation (truncated to two terms) is (Lugannani and Rice, 1980)

$$F_M(x) \approx F_{\mathcal{N}(0,1)}(x^*) + f_{\mathcal{N}(0,1)}(x^*) \left(\frac{1}{x^*} - \frac{1}{s\sigma_s}\right),$$
 (4.38)

where

$$x^* = \operatorname{sgn}(s)\sqrt{2\left(sx - \log(\phi_M(s))\right)}.$$

and, as in the previous section, s solves $\mu_s = x$.

The distribution of M is asymptotically χ_f^2 , not asymptotically normal. Hence the Lugannani & Rice approximation will not converge to the true distribution for large sample sizes. In the context of approximating the distribution to the Bartlett-Nanda-Pillai trace statistic, Butler et al. (1992) discusses how this problem affects various saddlepoint approximations.

Wood et al. (1993) generalise the Lugannani & Rice approximation so that we

Chapter 4. Likelihood Ratio Testing for Quaternion-Structured Covariance Matrices

may use a non-normal first term. If we choose a $\Gamma(k, 1)$ basis distribution,²⁴ then Wood et al. (1993) suggests taking $\lambda = \frac{4\sigma_s^6}{\kappa_3_s^2}$. Let $s^* = 1 - \frac{\lambda}{x^*}$ play the role of s in tilting the gamma basis distribution. We must solve $\phi_{\Gamma(\lambda,1)}(s^*)e^{-s^*x^*} = \phi_M(s)e^{-sx}$. The two real solutions can be expressed in terms of the multi-valued Lambert W function $(x^* = -\lambda \text{LambertW}(-e^{\frac{x}{\lambda}-1}))$. The larger real solution is chosen if $x > \mu_0$, and the smaller real solution is chosen if $x < \mu_0$. Let $\mu_{s^*}^* = x^*$ and $\sigma_{s^*}^*^2 = \frac{\lambda}{(1-s^*)^2}$ denote the mean and variance of the tilted gamma distribution respectively. Then

$$F_M(x) \approx F_{\Gamma(\lambda,1)}(x^*) + f_{\Gamma(\lambda,1)}(x^*) \left(\frac{1}{s^*} - \frac{\sigma_{s^*}^*}{s\sigma_s}\right).$$

$$(4.39)$$

For applications of various types of saddlepoint approximations to some particular test statistics of Box type, see e.g. Butler et al. (1992, 1993); Srivastava and Yau (1989).

(4.37), (4.38) and (4.39) can all be inverted numerically to obtain approximate quantiles. Alternatively, Maesono and Penev (1998) gives an asymptotic inversion of the Lugannani & Rice approximation; however we have found its performance to be poor.²⁵

4.3.2.9 Monte Carlo method

The distribution of M can be approximated by an empirical CDF, which we obtain by simulating M repeatedly. For fixed x, the empirical estimate of $F_M(x)$ obtained from n simulations will be random and distributed as $\frac{1}{n}$ Binomial $(n, F_M(x))$.

When applying this method to the LRT for quaternion propriety, the moments (4.6), which fully determine the null distribution of W, and hence of M, do not depend on the true covariance matrix Σ_r . Hence we can assume without loss of generality that $\Sigma_r = I_{4p}$ when simulating M. For each normal sample of size N the maximum likelihood estimator $\hat{\Sigma}_{\mathbb{R}}$ is computed from (4.1). $\hat{\Sigma}_{\mathbb{H}}$ is then given by Proposition 4.3

²⁴The choice of location and scale parameters does not influence the final outcome.

 $^{^{25}}$ We believe that this is because it relies on having an accurate initial normal approximation. Also, the initial normal estimate of the tilting parameter s may be outside the valid range.

and Definition 1.42, and W by (4.5). Each simulated value of M requires Np standard normal samples. Note that because in this case the assumptions of Proposition 4.13 hold, we could instead simulate W as a product of only $\lceil 3p/2 \rceil$ independent beta random variables.

Walden and Rubin-Delanchy (2009) applies the Monte Carlo method to the LRT for complex propriety. When comparing with the exact distribution — which we obtain from Theorem 4.29 — we have found however that by using $n = 30\,000$ simulations they achieved only a couple of digits of accuracy in computing the quantiles of W.²⁶

For large *n* the sample quantile \hat{x}_{MC} obtained by this Monte Carlo method will be approximately distributed as $\mathcal{N}^{\mathbb{R}}\left(x, \frac{F_W(x)(1-F_W(x))}{nf_W(x)^2}\right)$ (Walker, 1968).²⁷ Thus every additional digit of desired accuracy requires that we increase the number of simulations by a factor of 100.

More sophisticated Monte Carlo methods such as importance sampling may also be used (Glynn, 1996).

4.3.3 Numerical comparison of approximations

Comparing (4.6) and (4.7) we see that the LRT for quaternion propriety T is a statistic of Box type with

$$\begin{split} m &= k = \lceil 3p/2 \rceil, \\ x_i &= y_j = N, \\ \xi_i &= -4p + 2i - 1, \\ \eta_j &= \frac{1}{2} \left(2 - j - \lfloor (j - 1)/3 \rfloor \right), \end{split}$$

²⁶Sometimes only one digit, 0 digits for p = 6, N = 20, 4 digits when N = 1000

²⁷Consider for example $W \sim \chi_9^2$. Then the 99th percentile is $x \approx 21.666$ and the distribution of \hat{x}_{MC} is approximately $\mathcal{N}^{\mathbb{R}}\left(x, \frac{785.53}{n}\right)$. So in order to ensure that an an estimation error of more than ± 0.5 has a probability of occurring of less than 1% one would need over 20 000 simulations.

and K in (4.6) is

$$K = \prod_{i=1}^{\left\lceil \frac{3p}{2} \right\rceil} \frac{\Gamma(N+\eta_i)}{\Gamma(N+\xi_i)}.$$

Since W = T is of Box type and satisfies the assumptions of Proposition 4.13, all of the results of Section 4.3 can be applied to compute the distribution of $M = -2 \log T$. In particular, the asymptotic distribution of M is χ_f^2 with

$$f = \sum_{i=1}^{\lceil 3p/2 \rceil} (8p + 4 - 5i - \lfloor (i-1)/3 \rfloor)$$

= $3p(2p+1).$

As expected from Wilks' theorem (Young and Smith, 2005, p. 132), this is equal to the difference between the number of free parameters in the covariance matrix under H_1 , namely 2p(4p + 1), and under H_0 , namely p(2p - 1).

Also, the Box Bartlett adjustment in this case is

$$C_{\rm Box} = \frac{12N}{12N + 1 - 20p}.$$

In this section we will compare the accuracy of the various approximations described in Section 4.3.2, by applying them to the LRT statistic for quaternion propriety.

For a chosen combination of p and N we define the relative error of the approximation $F_{\approx}(x)$ to $F_M(x)$ as

$$\left|\frac{F_{\approx}(x) - F_M(x)}{\min\left\{F_M(x), 1 - F_M(x)\right\}}\right|,$$

where $F_M(x)$ is the exact CDF and $F_{\approx}(x)$ is any of the approximate CDFs considered in this section. The effect of the divisor is to make the error relate to the corresponding tail probability, depending on whether x corresponds to a value in the left or right tail. Since we will reject the hypothesis of propriety when M is larger than some critical value, we will be most interested in the region around $F_M(x) = 0.95$ and 0.99. The relative error at the 95th percentile for example is $\left|\frac{F_{\approx}(z_{0.95})-0.95}{0.05}\right|$ where F_{\approx} is the approximate CDF and $F_M(z_{0.95}) = 0.95$.

We have chosen not to include the Monte Carlo method in the figures, since it would hinder readability and provide little insight. Indeed, for figures 4.1, 4.2, 4.3 and 4.4, the relative errors for the Monte Carlo method would simply be an i.i.d. sequence of |Binomial(100 000, 0.05) - 5000|/100 000 random variables. These are approximately half-normal distributed as $|\mathcal{N}^{\mathbb{R}}(0, 4.75 \cdot 10^{-7})|$ and have a mean of approximately $5.5 \cdot 10^{-4}$.

vc								
	Label	Description	Line Style					
	$C_{\rm box}$	χ^2 approximation with Bartlett adjustment (4.29)						
	$F_{\rm box}$	Box's F approximation (4.30)-(4.31)	+					
	F_{exact}	New F approximation (4.32)-(4.34)	——————————————————————————————————————					
	Γ	Γ approximation (4.35)-(4.36)						
	JLDE	Jensen's large deviation estimate (4.37)						
	$L\&R\mathcal{N}$	Lugannani & Rice approximation (4.38)	·					
	$L\&R\Gamma$	Generalised Lugannani & Rice (4.39)						
	$\chi^2 5$	Box χ^2 series of order O(N^{-5}) (4.26),						
		with renormalisation (4.28)						
	MC	100 000 Monte Carlo simulations (Section 4.3.2.9)	N/A					

Chapter 4. Likelihood Ratio Testing for Quaternion-Structured Covariance Matrices

Table 4.1: Legend



Figure 4.1: Relative errors of approximate CDFs at the 95th percentile for varying p and N = 4p. For $p \ge 4$, the new F approximation yields invalid parameter values.

Fitting lines to the curves in Figure 4.4, between N = 500 and N = 1000, we obtain the following relative errors for the various approximations:



Figure 4.2: Relative errors of approximate CDFs at the 95th percentile for varying p and N = 5p.



Figure 4.3: Relative errors of approximate CDFs at the 95th percentile for varying p and N = 8p.

Chapter 4. Likelihood Ratio Testing for Quaternion-Structured Covariance Matrices



Figure 4.4: Relative errors of approximate CDFs at the 95th percentile for p = 6 and varying N.

C_{box} :	$8.91 \cdot 10^2 \cdot N^{-2.03},$
Γ:	$1.26 \cdot N^{-2.04},$
F_{box} :	$1.21 \cdot 10^4 \cdot N^{-3.04},$
F_{exact} :	$3.85 \cdot 10^{-1} \cdot N^{-2.99},$
JLDE:	$3.30 \cdot N^{-2.04},$
L&R \mathcal{N} :	$3.02 \cdot 10^{-6} \cdot N^{0.036},$
$L\&R\Gamma$:	$4.46 \cdot 10^{-2} \cdot N^{-2.05},$
χ^2 5:	$7.37 \cdot 10^7 \cdot N^{-5.34}.$

These agree closely with the theoretical order of the errors.

Surprisingly, using a gamma basis instead of the normal basis in the Lugannani & Rice approximation does not improve the precision noticeably, except for large N as seen in Figure 4.4. The Box χ^2 series performs poorly for moderate N or large p. This is also counterintuitive given its high order of approximation and non-negligible complexity.

Box's Bartlett adjustment provides an approximation so simple that it does not require the use of a computer, but should only be used when the sample size is very large. Our new F approximation on the other hand is extremely accurate, even for small N. The gamma approximation is a simpler alternative with intermediate perfor-

р	Ν	$C_{\rm box}$	Г	$F_{\rm box}$	F_{exact}	JLDE	$L\&R\mathcal{N}$	$L\&R\Gamma$	$\chi^2 5$	MC
1	4	0.039774	0.797285	0.147933	0.980560	1.033382	0.996140	0.985760	0.305001	1.009
1	5	0.312620	0.902056	0.555284	0.997154	1.007755	0.979037	0.971145	0.822674	1.004
1	8	0.768796	0.975642	0.915373	1.000009	0.998012	0.989318	0.986453	0.992721	0.934
1	16	0.953481	0.995624	0.991747	1.000011	0.999662	0.999082	0.997682	0.999896	1.005
2	8	0.000247	0.631786	0.004430	0.981219	1.064562	0.991405	0.980947	0.003468	1.062
2	10	0.142918	0.887571	0.363418	0.998806	1.008592	0.978844	0.976115	0.482437	1.020
2	16	0.678962	0.979091	0.881933	0.999897	1.002074	0.997900	0.997594	0.971254	1.007
2	32	0.933936	0.996533	0.988992	0.999993	1.000379	0.999908	0.999728	0.999605	0.978
3	12	0.000002	0.554584	0.000146	0.979849	1.081944	0.988963	0.978544	0.000025	0.978
3	15	0.083270	0.905234	0.284950	0.998159	1.009231	0.989446	0.988778	0.286228	0.993
3	24	0.616741	0.984185	0.860928	0.999874	1.001752	0.999448	0.999374	0.940046	1.015
3	48	0.918455	0.997425	0.987029	0.999993	1.000282	0.999981	0.999929	0.999123	0.963
4	16	0.000000	0.511526	0.000005	NaN	1.092964	0.987269	0.976912	0.000000	0.973
4	20	0.049552	0.921868	0.229535	0.998152	1.008401	0.995224	0.995035	0.163298	0.968
4	32	0.560705	0.987543	0.840782	0.999898	1.001321	0.999793	0.999766	0.894254	0.974
4	64	0.903368	0.997983	0.985043	0.999994	1.000208	0.999994	0.999973	0.998249	1.051
6	24	0.000000	0.467119	0.000000	NaN	1.105843	0.984892	0.974666	0.000000	0.953
6	- 30	0.016723	0.944156	0.148510	0.998668	1.005919	0.998775	0.998744	0.048562	0.995
6	48	0.461552	0.991373	0.800892	0.999940	1.000832	0.999948	0.999941	0.767637	1.042
6	96	0.873663	0.998606	0.980953	0.999997	1.000131	0.999999	0.999993	0.994481	0.987
12	48	0.000000	0.428737	0.000000	NaN	1.119405	0.980828	0.970926	0.000000	0.996
12	60	0.000377	0.971059	0.035333	0.999528	1.002628	0.999895	0.999893	0.000832	0.991
12	96	0.249391	0.995573	0.688035	0.999981	1.000372	0.999995	0.999994	0.390801	1.014
12	192	0.788828	0.999284	0.968377	0.999999	1.000059	1.000000	0.999999	0.954162	1.031

Table 4.2: Approximate rejection probabilities (in %) for the the 1% level critical region. Entries are $100(1 - F_{\approx}(z_0))$ %, where $F_M(z_0) = 0.99$. NaN values indicate that the parameters computed for F_{exact} were invalid.



Chapter 4. Likelihood Ratio Testing for Quaternion-Structured Covariance Matrices

Figure 4.5: Relative errors of approximate CDFs for p = 4 and N = 32, for varying x, or equivalently varying percentiles. The sharp dips correspond to points where F_{\approx} and F_M cross and the error changes sign.

mance. These three approximations are true distributions, and allow for immediate computation of the CDF, PDF and quantiles, something which is not true of the three saddlepoint approximations considered. Thus, we would recommend using one of these three methods, depending on the sample size and the user's preference for simplicity.

Chapter 5

Quaternion Wavelets and Matrix-Valued Wavelets

5.1 Introduction

Wavelet transforms (Daubechies, 1992) are a tool for signal decomposition and analysis and have been succesfully applied to many signal processing problems in the past two decades. We will consider in particular orthogonal wavelets. These are functions $\psi(t)$ whose translations and dilations $2^{-\frac{j}{2}}\psi(2^{j}t-k)$, $j,k \in \mathbb{Z}$ generate an orthonormal basis for the signal space (commonly $L^{2}(\mathbb{R},\mathbb{R})$). The wavelet transform of a signal is given by its coefficients in this basis. By putting upper limits on the size of dilations j, linear subspaces of the signal space can be generated with varying granularity. This produces a multiresolution analysis (MRA) and allows for scale-based signal decomposition. Unless otherwise specified, the term 'wavelet' will refer to 'orthogonal wavelet' throughout this chapter (and similarly for wavelet filter, scaling filter, scaling function and MRA).

Scalar (real) wavelet techniques can be applied to vector-valued signals in $L^2(\mathbb{R}, \mathbb{R}^n)$ by simply treating each component independently as a scalar signal. However, this common 'naive' approach ignores potentially useful dependencies between components. As with AR time-series modelling, one holistic approach to vector-valued signals is to treat them as algebra-valued signals so that one may apply corresponding algebra-valued wavelets. For example, the complex Daubechies wavelets of Lina and Mayrand (1995) can be used to analyse signals in $L^2(\mathbb{R}, \mathbb{R}^2)$.

It is important not to confuse the wavelets required for the analysis of algebravalued signals with other types of wavelet also labeled as complex-, quaternion- or Clifford-algebra-valued. The latter are typically designed for analysing real scalar signals, and we will describe these briefly in Section 5.2.1,

Another approach to processing vector-valued signals holistically is given by matrixvalued wavelets (MVWs). Through the vector space and algebra isomorphisms, quaternion wavelets can be seen as special cases of MVWs having quaternion structure. We show in Section 5.7 that such an approach can be used for any real algebra.¹ Thus most of the material in this chapter will be presented within the more general framework of orthogonal MVWs.

More general types of MVW transform have also been considered in the literature, such as biorthogonal MVWs (Agreste and Vocaturo, 2009b; Bacchelli et al., 2002; Chen et al., 2006; Cui et al., 2009), *m*-band MVWs (Chen and Shi, 2008; Cui and Zhang, 2008) and MVW packets (Chen and Shi, 2008). MVWs can also be considered as a special case of generalised multiwavelets. Multiwavelet transforms can be applied directly to vector-valued signals (since they require vectorisation of scalar signals), and the matrix-valued and multiwavelet versions of the fast DWT algorithm differ only in their choice of matrix-valued filters. Fowler and Hua (2002a) show however that in practice this leads to very poor results, highlighting the need for wavelets specifically designed for vector-valued signals.

Note that within the literature on MVWs, a plurality of alternate names for them are used. These include the original name vector-valued wavelet (Xia and Suter, 1996), multiple vector-valued wavelet (Chen et al., 2006), multichannel wavelet (Agreste and Vocaturo, 2009a; Bacchelli, 2002), omnidirectionally balanced multiwavelet (Fowler and Hua, 2002a), and wavelet with a full rank (multi-)filter (Agreste and Vocaturo, 2009b; Bacchelli et al., 2002). The name we have chosen to use (matrixvalued wavelet) seems to be the most common (Walden and Serroukh, 2002; Xia, 1997;

¹Assuming that orthogonality in the algebra is defined based on an involution which maps to matrix transposition in the algebra's matrix representation.

Yu, 2011).

We will prove in Section 5.3 that two competing interpretations of MVWs — as $\mathbb{R}^{n \times n}$ -algebra-valued wavelets generating a matrix MRA of $L^2(\mathbb{R}, \mathbb{R}^{n \times n})$ and as a collection of n vector-valued wavelets which jointly generate a vector MRA of $L^2(\mathbb{R}, \mathbb{R}^n)$ — are fundamentally equivalent. In general, algebra-valued MRAs can be treated as special cases of vector MRAs.

We define a $n \times n$ MVW to be trivial if it can be decomposed into independent lower-dimensional MVWs (in some appropriate orthogonal basis of \mathbb{R}^n). Every MVW is then composed of one or more non-trivial MVWs. In particular, real and complex² wavelets are the trivial examples of quaternion wavelets. Indeed, within the algebravalued framework, the naive approach corresponds to the special case where the wavelet used is real-valued. He and Yu (2005); Peng and Zhao (2004) have constructed quaternion wavelets. However, we show in Section 5.2.2 that all examples given are either incorrect or trivial. In Section 5.8.2 we give the first example of a non-trivial orthogonal quaternion wavelet.

We prove various results showing a lack of non-trivial MVWs: There are no nontrivial matrix-valued scaling filters (MVSFs) of length $L \leq 3$ and no non-trivial Daubechies MVSFs of length L = 4 (i.e. with 2 vanishing moments) except for the real scalar Haar and Daubechies filters respectively. We also show computationally that there are no non-trivial quaternion Daubechies scaling filters of length L < 10and there are no non-trivial 3×3 Daubechies MVSFs of length L = 6. For any filter length, matrix Daubechies filters differ from their naive counterpart only by an all-pass filter.

To construct a MVW, it is sufficient to specify an appropriate MVSF $\{G_k\}$. A matrix-valued wavelet filter, matrix-valued scaling function and MVW can then be computed from the MVSF. Constructing trivial MVSFs from non-trivial MVSFs is simple, however constructing new non-trivial MVSFs is harder. Agreste and Vocaturo (2009b) develop a method for constructing biorthogonal MVSFs through a multichannel lifting scheme, which leads to the explicit designs by Bacchelli et al.

 $^{^{2}}$ Note that as in Remark 1.14, quaternion wavelets can still be considered complex if the imaginary unit is a pure unit quaternion other than i.

(2002, Table 4.1) (2 × 2) and Agreste and Vocaturo (2009b, p. 4) (3 × 3). Fowler and Hua (2002b) design biorthogonal wavelets by symbolically solving a set of design equations, with explicit 2 × 2 designs given by Hua and Fowler (2002, pp. 3–7).³ We will use a similar approach for the orthogonal case. Another method worth mentioning, which allows for orthogonal constructions, is the spectral factorisation of interpolatory vector subdivision schemes suggested by Conti et al. (2008).⁴ We will construct examples of non-trivial MVSFs by symbolically solving a set of quadratic design equations imposing orthogonality and vanishing moments. More specifically, we obtain the family of non-trivial 2 × 2 Daubechies MVSFs of length L = 6 and the family of non-trivial quaternion Daubechies scaling filters of length L = 10.

Except for the cases mentioned above, the explicit constructions of compactlysupported MVWs we have found in the literature are limited to toy examples. For many of these, the author's desire to obtain closed-form solutions for the matrixvalued wavelet filter (as a function of the MVSF) narrows design possibilities. For example Chen and Shi (2008); Chen et al. (2006) only consider filters of length L = 3(which are trivial);⁵ the constructions of Cui et al. (2009); He and Huang (2012); Walden and Serroukh (2002) focus on controlling the eigenvalues of the Fourier tranform of the scaling filter; and Cui and Zhang (2008)⁶ impose that certain products of coefficients be symmetric matrices. As we noted in Ginzberg and Walden (2013a, Section VII), a general algorithm for obtaining multiwavelet filters from multiscaling filters can be applied to MVSFs, rendering these restrictions unnecessary.

Walden and Serroukh (2002) use a 2×2 MVW to compress four financial timeseries by interpreting them as a $\mathbb{R}^{2\times 2}$ -valued time-series. Such use of the matrix MRA interpretation of MVWs in practice is inappropriate as it amounts to independent

³The OBSA5-3 filter of Fowler and Hua (2002b) is trivial, but the OBSA7-5 filter is non-trivial.

⁴One difficulty with this approach is that to design MVSFs with additional properties, one must find and impose corresponding constraints on the interpolatory filter to be factorised.

⁵Chen and Shi (2008) generalise the approach to *m*-band filters of length m + 1.

⁶Note that the filter given by Cui and Zhang (2008, pp.180-181) only satisfies the necessary condition (5.10) for parameter values $\alpha = \tan\left(\frac{\pi}{8}\right)$ and $\alpha = -2 - \tan\left(\frac{\pi}{8}\right)$, both of which lead to trivial filters.

analysis of the matrix rows. Other applications of MVWs — which use the better vector MRA interpretation — include the compression and denoising of colour images with 3×3 MVWs (Agreste and Vocaturo, 2009a,b) and of 2D wind fields with a 2×2 MVW (Hua and Fowler, 2004; Westenberg and Ertl, 2005); and digital watermarking of colour images with a 3×3 MVW (Agreste and Vocaturo, 2009c).⁷

Some of the material in this chapter was published in Ginzberg and Walden (2013a), see p. 11.

5.2 A review of literature on quaternion wavelet transforms

5.2.1 Different types of quaternion wavelet transform

With the exception of this section, within this chapter we are interested in quaternion wavelets which are quaternion-valued functions $\psi \in L^2(\mathbb{R}, \mathbb{H})$, and in associated quaternion wavelet transforms suited to analysing signals $f \in L^2(\mathbb{R}, \mathbb{H})$ (or more generally $L^2(\mathbb{R}^m, \mathbb{H})$). The continuous wavelet transform $\mathcal{W}_{\psi}(f)$ is then given by convolving the signal with dilated versions of the (mother) wavelet

$$\mathcal{W}_{\psi}(f)(a,b) = \int_{-\infty}^{\infty} f(t)a^{-\frac{1}{2}}\overline{\psi\left(\frac{t-b}{a}\right)} \mathrm{d}t.$$

We are particularly interested in orthogonal wavelets, whose dyadic dilations $a = 2^j$, $j \in \mathbb{Z}$ and integer translations $b = 2^j k$, $k \in \mathbb{Z}$ form an orthonormal basis for the quaternion (left-)module $L^2(\mathbb{R}, \mathbb{H})$. The DWT limits itself to these discrete values of a and b and can be computed from a digital signal through convolutions with a scaling filter and wavelet filter. We will also assume for simplicity that the wavelet has compact support, or equivalently that the filters are of finite length.

The terms 'quaternion wavelet' and 'quaternion wavelet transform' are however used in the literature to refer to a number of different things. We will now discuss these different approaches. One way in which they all differ from ours is that they

⁷These are signals which are 2D in 'time' (in this case (x, y) spatial position) in addition to being 2D or 3D in the number of components (in this case wind speed in x and y directions or red green and blue colour intensity).

are designed specifically to analyse signals in $L^2(\mathbb{R}^2, \mathbb{R})$ (or for some also $L^2(\mathbb{R}^2, \mathbb{H})$).

Zhao and Peng (2007) define a continuous quaternion wavelet transform which decomposes a signal $L^2(\mathbb{R}^2, \mathbb{H})$ according to orientation in addition to scale (or frequency) and space (or time), i.e. the signal is analysed by convolving it with scaled and rotated versions of a mother wavelet. Bahri et al. (2012) derive some theoretical results for this type of QWT, but because of their use of a quaternion Fourier transform (with kernel $e^{\frac{i+j+k}{\sqrt{3}}\omega t}$), these require additional ad-hoc commutativity assumptions.⁸ We will discuss other cases where the use of quaternion Fourier transforms is problematic in Section 5.2.2.

The most widely used category of quaternion wavelet transforms are those generalising the dual-tree complex wavelet transform, similarly to how bivariate quaternion Fourier transforms generalise the complex Fourier transform. For a review of the dual-tree (and related) complex wavelet transform see Selesnick et al. (2005); Shukla (2003). As noted by Selesnick et al. (2005, p. 131), the dual-tree complex wavelet transform is $2 \times$ redundant for both real and complex signals (it is based on a wavelet tight frame rather than a wavelet basis). It generates two real MRAs (of $L^2(\mathbb{R},\mathbb{R})$) which are (approximate) Hilbert transforms of one another rather than a complex MRA, and it is computed via two independent real wavelet transforms producing the real and imaginary parts. The corresponding complex wavelet is (approximately) analytic.

Chan et al. (2008) refer to the quaternion generalisation — which applies to signals in $L^2(\mathbb{R}^2, \mathbb{R})$ — as the dual-tree quaternion wavelet transform, and show that the coefficients of the dual-tree quaternion wavelet transform and a corresponding 2D dual-tree complex wavelet transform are related by a simple linear transformation. The advantage of the quaternion formulation over the complex pair formulation is that, in their polar form, quaternion coefficients can be interpreted in terms of a single shift-invariant amplitude and three phases, two of which vary with horizontal and vertical shifts respectively. Another relationship with complex wavelets is that

⁸We believe that these additional assumptions will not hold in practice unless the quaternion wavelet is trivial.

the quaternion wavelet $\Psi \in L^2(\mathbb{R}^2, \mathbb{H})$ can be decomposed as

$$\Psi(x,y) = \psi(x) e^{\pi k/4} \psi(y) e^{-\pi k/4}, \qquad (5.1)$$

where $\psi \in L^2(\mathbb{R}, \mathbb{C})$ is a complex wavelet.⁹ The most popular wavelets of this type are the quaternion Gabor wavelets developed by Bayro-Corrochano (2006).

Dual-tree quaternion wavelets have been succesfully applied to greyscale images for texture classification (Li et al., 2013), compression (Soulard and Carré, 2010) (despite the $4 \times$ redundancy), speckle denoising (Liu et al., 2012) and optical flow estimation (Bayro-Corrochano, 2006).

Closely related to the use of dual-tree quaternion wavelets is the hyperanalytic wavelet transform (Olhede, 2007), which computes the real wavelet transform of each component of the (hyper)analytic version of a signal $f \in L^2(\mathbb{R}^2, \mathbb{R})$. For the continuous wavelet transform and the maximum overlap DWT,¹⁰ the use of the analytic version of a real wavelet, of the analytic version of a signal or of the analytic version of the real wavelet transform coefficients are equivalent. For the standard DWT the three approaches are subtly different.

By using the Riesz transform to generalise the Hilbert transform to higher dimensions (instead of using Hilbert transforms along the vertical and horizontal directions), a different quaternion generalisation of the complex analytic signal — called monogenic signal — is obtained. The monogenic versions of real wavelets yield monogenic quaternion wavelets for analysing $L^2(\mathbb{R}^2, \mathbb{R})$ or more generally Clifford wavelets for $L^2(\mathbb{R}^m, \mathbb{R})$ (Held et al., 2010). These monogenic wavelets exhibit rotation-invariance of the wavelet coefficient amplitude when the real part is isotropic, in addition to the shift-invariance offered by analytic wavelets.

A generalisation of monogenic wavelet analysis to vector signals such as color images has been suggested by Soulard et al. (2013), but relies on a reinterpretation of

⁹Note that (5.1) would yield a quaternion wavelet suitable for analysing quaternion signals if the complex wavelet ψ were suitable for complex signals (and hence also quaternion signals). It could still in a sense be considered trivial as a tensor product of trivial wavelets.

¹⁰The maximum overlap DWT (a.k.a. shift-invariant or cycle-spinning DWT) is a redundant version of the DWT which does not downsample coarse-scale coefficients.

marginal transforms. The use of redundant quaternion analytic or monogenic wavelets in greyscale image analysis to obtain wavelet coefficients with amplitude-phase interpretations is of a fundamentally different nature from the use of quaternions to encode pixel colour in colour image analysis. The latter allows for simple formulations of geometric colourspace transformations (Ell, 2007), and can account for them in e.g. image registration (Moxey et al., 2003).¹¹ This latter use of quaternions is more closely related to the type of quaternion wavelet considered in the rest of this chapter.

For an extensive bibliography on quaternion and Clifford-algebra-valued wavelet (and Fourier) transforms, see the recent review paper by Brackx et al. (2013).

5.2.2 Problems with existing quaternion wavelet constructions

Quaternion wavelets are investigated by Bahri (2010) and He and Yu (2005) using two different (but fundamentally equivalent) representations of quaternions as structured matrices in $\mathbb{C}^{2\times 2}$. They work predominantly in the frequency domain by making use of quaternion Fourier transforms, however we will show that these are not a suitable choices of Fourier transform for this task. The fundamental problem is that although every pure imaginary unit quaternion generates a complex subalgebra of \mathbb{H} (as we noted in Section 1.2.1) — and can thus be used as an imaginary unit in a Fourier kernel — \mathbb{H} is not a complex algebra and hence there is no quaternion-valued Fourier kernel which commutes with all quaternions. This problem can be solved by extending \mathbb{H} to the complex algebra of biquaternions (see Section 5.7.3).

Sangwine and Ell (2012) show that real matrix representations allow for a unified understanding of quaternion Fourier transforms and other hypercomplex Fourier transforms, since in the real matrix domain they all use kernels of the form $\cos(\omega t)\mathbf{I}_n$ + $\sin(\omega t)\mathbf{M}$, differing only in their choice of 'imaginary unit' matrix \mathbf{M} (which must

¹¹Moxey et al. (2003) show that the quaternion correlation between two colour images can be used for image registration when — in addition to a spatial shift — the image has suffered from a colour space distortion (modeled as a rotation, scaling and translation of the colour basis) or has been converted to greyscale.

satisfy $M^2 = -I_n$.¹²

To better compare and understand the complex-matrix-domain approaches of He and Yu (2005) and Bahri (2010) we map them to the quaternion domain. In the following $\phi \in L^2(\mathbb{R}, \mathbb{H})$ denotes a quaternion scaling function and $\{g_\ell\}$ denotes a quaternion scaling filter.

He and Yu (2005, eqns. (1.7), (2.1), (2.2) and (3.1)) use the continuous quaternion Fourier transform

$$\hat{\phi}^{\mathrm{H}Y}(f) = \int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}2\pi ft} \phi(t) \mathrm{d}t,$$

and the discrete quaternion Fourier transform

$$\hat{G}^{\mathrm{HY}}(f) = \sum_{\ell \in \mathbb{Z}} g_{\ell} \,\mathrm{e}^{-\mathrm{i}2\pi f\ell}.$$

They claim that the two-scale dilation equation

$$\phi(t) = \sqrt{2} \sum_{\ell \in \mathbb{Z}} g_\ell \phi(2t - \ell)$$
(5.2)

is given in the Fourier domain by

$$\hat{\phi}^{\rm HY}(f) = \frac{1}{\sqrt{2}} \hat{G}^{\rm HY}\left(\frac{f}{2}\right) \hat{\phi}^{\rm HY}\left(\frac{f}{2}\right).$$
(5.3)

This is a standard result in the real and complex case. However in the quaternion case, if the scaling function and scaling filter are non-trivial, then they will not commute with the Fourier kernel and (5.3) will not hold. This creates problems for their frequency domain design method. Three constructions are given:

Design 1 is simply the real Haar scaling filter $g_0 = g_1 = 2^{-\frac{1}{2}}$. The two other scaling filter designs however do not produce orthogonal MVWs because they are not

¹²In the quaternion case, n = 4. The extension to vector time is obtained by taking products of such kernels. The case of two-sided quaternion Fourier transforms can be accommodated by using imaginary units $M_1 \in \widetilde{\mathbb{H}}$ for the left kernel and $M_2 \in \widetilde{\mathbb{H}}^R$ for the right kernel (with both kernels appearing on the left in the real matrix domain and the signal appearing as a vector).

orthogonal to their even shifts, i.e. they do not satisfy

$$\sum_{\ell \in \mathbb{Z}} g_{\ell} \bar{g}_{\ell+2k} = \delta_{k,0}, \tag{5.4}$$

(see also (5.11)).

Design 2 is given in the quaternion domain by

$$g_{0} = \frac{3}{4\sqrt{2}}(1+j)$$

$$g_{1} = \frac{1}{\sqrt{2}}$$

$$g_{2} = \frac{3}{4\sqrt{2}}(1-j).$$

In addition to the fact that this filter is trivial, we have

$$g_0\bar{g}_2 = \frac{9}{16}\mathbf{j} \neq 0$$

which contradicts (5.4) and thus precludes orthogonality.

Design 3 is given by

$$g_{0} = 0$$

$$g_{1} = \frac{1}{8\sqrt{2}} \left(2 - \sqrt{3}j - 3k\right)$$

$$g_{2} = \frac{1}{\sqrt{2}}$$

$$g_{3} = \frac{1}{8\sqrt{2}} \left(6 + \sqrt{3}j + 3k\right).$$

In addition to the fact that this filter is also trivial, 13 we have

$$g_0\bar{g}_2 + g_1\bar{g}_3 = g_1\bar{g}_3 = -\frac{1}{16}\left(\sqrt{3}j + 3k\right) \neq 0,$$

which again contradicts (5.4).

 $^{13}\mathrm{It}$ is complex with imaginary unit $i_0=\frac{j+\sqrt{3}k}{2}.$

Bahri (2010, eqns. (20) and (33)) uses the continuous quaternion Fourier transform

$$\hat{\phi}^{\text{Bahri}}(f) = \int_{-\infty}^{\infty} \phi(t) \,\mathrm{e}^{-\mathrm{k}2\pi ft} \mathrm{d}t$$

and the discrete quaternion Fourier transform

$$\hat{G}^{\mathrm{Bahri}}(f) = \frac{1}{\sqrt{2}} \sum_{\ell \in \mathbb{Z}} g_{\ell} \mathrm{e}^{-\mathrm{k}2\pi f\ell}.$$

Bahri (2010, eqn. (34)) claims that

$$\hat{\phi}^{\text{Bahri}}(f) = \hat{G}^{\text{Bahri}}\left(\frac{f}{2}\right)\hat{\phi}^{\text{Bahri}}\left(\frac{f}{2}\right).$$

However, the proof given for this equality incorrectly assumes commutativity between $e^{-k\pi f\ell}$ and $\phi(2t-\ell)$, so it does not hold for non-trivial scaling functions.

Peng and Zhao (2004) use the biquaternion Fourier transform, which leads to correct frequency-domain results (see Section 5.7.3). They obtain three symmetric quaternion scaling filters by a method similar to the one we will use, i.e. by solving the quadratic equations corresponding to the various design constraints. However, all three constructions are trivial.

The first construction is given by

$$g_0 = g_3 = x + yi_0$$

 $g_1 = g_2 = (1/\sqrt{2}) - x - yi_0$

where $y = [(x/\sqrt{2}) - x^2]^{1/2}$, $x \in [0, (1/\sqrt{2})]$ is a free parameter and i_0 is an arbitrary pure imaginary unit quaternion.

The second construction is the symmetric complex Daubechies scaling filter of

length L = 6 of Lina and Mayrand (1995, Eqn. (2.21)) (with complex unit i_0).¹⁴

$$g_{0} = g_{5} = -\frac{1}{32\sqrt{2}} \left(3 + \sqrt{15}i_{0}\right)$$

$$g_{1} = g_{4} = \frac{1}{32\sqrt{2}} \left(5 - \sqrt{15}i_{0}\right)$$

$$g_{2} = g_{3} = \frac{1}{16\sqrt{2}} \left(15 + \sqrt{15}i_{0}\right).$$

The third construction is given by

$$g_{0} = g_{7} = \frac{-155 + \sqrt{1583470i_{0}}}{8448\sqrt{2}}$$

$$g_{1} = g_{6} = 3g_{0} + \frac{1}{16\sqrt{2}}$$

$$g_{2} = g_{5} = g_{0} + \frac{5}{16\sqrt{2}}$$

$$g_{3} = g_{4} = -5g_{0} + \frac{10}{16\sqrt{2}}.$$

In addition to being trivial,

$$g_0\bar{g}_6 + g_1\bar{g}_7 = \frac{35}{1056} \neq 0,$$

which contradicts (5.4). This attempt by Peng and Zhao (2004) to produce a symmetric Daubechies quaternion scaling filter of length L = 8 failed because no such filter exists. The authors did not notice the problem because only a subset of the design equations was used in the derivation.

In the more general field of Clifford-valued wavelets, Askari Hemmat and Rahbani (2010) give two constructions of $C\ell_{4,0}(\mathbb{R})$ -valued MVSFs. Case I is the real Haar filter. Case II is both trivial and fails to be orthogonal.¹⁵

¹⁴Hence the filter has 3 vanishing moments even though the authors only imposed 2 in the design.

¹⁵Or rather, it fails to be orthonormal. After an appropriate rotation, it is equal to the direct sum of the real Haar filter, a shifted version thereof and two scaled delay filters.

5.3 Matrix and vector multiresolution analyses

 $L^2(\mathbb{R}, \mathbb{R}^{m \times n})$ denotes the space of $m \times n$ matrix-valued functions defined on \mathbb{R} with values in $\mathbb{R}^{m \times n}$ having finite Frobenius (a.k.a. L^2) norm.

Within this chapter, it will often be helpful to think of matrices as column vectors of row vectors, i.e. $\mathbb{R}^{m \times n} = (\mathbb{R}^{1 \times n})^m$ and $L^2(\mathbb{R}, \mathbb{R}^{m \times n}) = (L^2(\mathbb{R}, \mathbb{R}^{1 \times n}))^m$. We will also tend to think of matrices as linear operators multiplying row vectors from the right.

Similarly to Section 1.3.2, we could have alternatively proceeded by treating matrices as row vectors of column vectors and as linear operators multiplying column vectors from the left. The latter approach would have allowed for a treatment of quaternions more consistent with the rest of the thesis, but would have required a reversion of the order of operations in the various equations, making comparisons with scalar wavelets and most MVW literature (e.g. Walden and Serroukh (2002)) less obvious. The two approaches are however equivalent, and the transpose operator \bullet^T maps between them.

Definition 5.1. The symbol "inner product" of $F_1, F_2 \in L^2(\mathbb{R}, \mathbb{R}^{m \times n})$ is given by

$$\langle \mathbf{F}_1, \mathbf{F}_2 \rangle_{m \times m} = \int_{-\infty}^{\infty} \mathbf{F}_1(t) \mathbf{F}_2^T(t) \mathrm{d}t.$$

The (usual) inner product is instead given by

$$\langle \boldsymbol{F}_{1}, \boldsymbol{F}_{2}
angle = \mathrm{tr}\left(\langle \boldsymbol{F}_{1}, \boldsymbol{F}_{2}
angle_{m imes m}
ight)$$

We may similarly define these two types of inner product for matrices in $\mathbb{R}^{m \times n}$ and for (square-summable) sequences in $\ell^2(\mathbb{Z}, \mathbb{R}^{m \times n})$.

The inner product $\langle \bullet, \bullet \rangle$ is consistent with interpreting $\mathbb{R}^{m \times n}$ as a *mn*-dimensional vector space, and gives $L^2(\mathbb{R}, \mathbb{R}^{m \times n})$ a Hilbert space structure. The Frobenius norm follows from this inner product, i.e.

$$||\boldsymbol{F}|| = \sqrt{\langle \boldsymbol{F}, \boldsymbol{F} \rangle}.$$

(see also Definition 1.57)

Definition 5.2. A set V is a (left-) \mathbb{A} -module for a ring \mathbb{A} if it is closed under (left-) \mathbb{A} -linear combinations, i.e. for any $x, y \in V$, $a, b \in \mathbb{A}$

$$ax + by \in V.$$

The symbol inner product $\langle \bullet, \bullet \rangle_{m \times m}$ is consistent with interpreting $\mathbb{R}^{m \times n}$ and $L^2(\mathbb{R}, \mathbb{R}^{m \times n})$ as $\mathbb{R}^{m \times m}$ -modules. With the exception of the case m = 1 — for which the two inner products are equal — $\langle \bullet, \bullet \rangle_{m \times m}$ does not define a true inner product, since it takes values in the algebra $\mathbb{R}^{m \times m}$ which is not a field. However, it is bilinear (sesquilinear) and symmetric in the sense that for $A, B \in \mathbb{R}^{m \times m}$ and $F_1, F_2, F_3 \in L^2(\mathbb{R}, \mathbb{R}^{m \times n})$

$$egin{array}{rcl} \langle oldsymbol{A}oldsymbol{F}_1,oldsymbol{B}oldsymbol{F}_2
angle_{m imes m} &=& oldsymbol{A}oldsymbol{F}_1,oldsymbol{F}_2
angle_{m imes m} &=& oldsymbol{A}oldsymbol{F}_2,oldsymbol{F}_1
angle_{m imes m} \ oldsymbol{A}oldsymbol{F}_2,oldsymbol{F}_1
angle_{m imes m} &=& oldsymbol{A}oldsymbol{F}_2,oldsymbol{F}_1
angle_{m imes m} \ oldsymbol{A}oldsymbol{F}_2,oldsymbol{F}_1
angle_{m imes m} &=& oldsymbol{A}oldsymbol{F}_2,oldsymbol{F}_1
angle_{m imes m} \ oldsymbol{A}oldsymbol{F}_2,oldsymbol{F}_1
angle_{m imes m} \ oldsymbol{A}oldsymbol{F}_2,oldsymbol{F}_1
angle_{m imes m} \ oldsymbol{A}oldsymbol{A}oldsymbol{F}_2,oldsymbol{F}_1
angle_{m imes m} \ oldsymbol{A}oldsymbol{F}_2,oldsymbol{F}_2,oldsymbol{F}_1
angle_{m imes m} \ oldsymbol{A}oldsymbol{B}oldsymbol{A}oldsymbol{A} \ oldsymbol{A}oldsymbol{F}_2,oldsymbol{F}_2,oldsymbol{F}_2
angle_{m imes m} \ oldsymbol{A} \ oldsymbol{A} \ oldsymbol{A} \ oldsymbol{A} \ oldsymbol{F}_2,oldsymbol{A} \ oldsymbol{A} \ oldsymbol{F}_2,oldsymbol{A} \ oldsymbol{A} \$$

By considering matrices of the form λI_m , $\lambda \in \mathbb{R}$ it is clear that every $\mathbb{R}^{m \times m}$ module is also a real vector space. The following stronger result shows that the two notions are to a large extent interchangeable.

Proposition 5.3. Every (left-) $\mathbb{R}^{m \times m}$ -module V is of the form $V = S^m$, where S is a real vector space. Conversely, if S is a real vector space, then $V = S^m$ is a (left-) $\mathbb{R}^{m \times m}$ -module.

Proof. See Appendix B.5 or Ginzberg and Walden (2013a, Proposition 1). \Box

We will use the following notation

Definition 5.4. Given a matrix (or matrix-valued function) \mathbf{F} , let $\mathbf{F}^{(i,\bullet)}$ denote its *i*th row. Given a set V, let $V^{(i,\bullet)} = \{\mathbf{F}^{(i,\bullet)} : \mathbf{F} \in V\}$. **Definition 5.5.** $\delta_{i,j}$ denotes the Kronecker delta

$$\delta_{i,j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Definition 5.6. A (finite or countable) sequence $\mathbf{F}_k \in V$ forms a $\mathbb{R}^{m \times m}$ -orthonormal basis for the (left-) $\mathbb{R}^{m \times m}$ -module V iff

$$\langle \boldsymbol{F}_i, \boldsymbol{F}_j \rangle_{m \times m} = \delta_{i,j} \boldsymbol{I}_m \ \forall k, l,$$

and for every $F \in V$ there exists a sequence $A_k \in \mathbb{R}^{m \times m}$ such that

$$oldsymbol{F} = \sum_k oldsymbol{A}_k oldsymbol{F}_k$$

Remark 5.7. $\mathbb{R}^{m \times m}$ -orthogonality is stronger than vector-space orthogonality $(\langle \bullet, \bullet \rangle)$ and corresponds to orthogonality of the rows, i.e. $\langle F_1, F_2 \rangle_{m \times m} = \mathbf{0}_{m \times m}$ iff $\forall i, j \langle F_1^{(i,\bullet)}, F_2^{(j,\bullet)} \rangle = 0$.

Proof.
$$\left\langle \mathbf{F}_{1}^{(i,\bullet)}, \mathbf{F}_{2}^{(j,\bullet)} \right\rangle$$
 is the (i, j) -entry of $\left\langle \mathbf{F}_{1}, \mathbf{F}_{2} \right\rangle_{m \times m}$.

A MRA defines nested spaces of finer and coarser-scale signal approximations. When the signal space considered is $L^2(\mathbb{R}, \mathbb{R}^{n \times n})$, as in e.g. Walden and Serroukh (2002); Xia and Suter (1996), the following definition arises:¹⁶

Definition 5.8. A (orthogonal $n \times n$) matrix MRA (MMRA) is a sequence of closed sub- $\mathbb{R}^{n \times n}$ -modules $V_j \subset L^2(\mathbb{R}, \mathbb{R}^{n \times n}), j \in \mathbb{Z}$ satisfying

- 1. $V_j \subset V_{j-1} \ \forall j \in \mathbb{Z}$.
- 2. $\bigcup_{j\in\mathbb{Z}} V_j$ is dense in $L^2(\mathbb{R}, \mathbb{R}^{n\times n})$ and $\bigcap_{j\in\mathbb{Z}} V_j = \{\mathbf{0}_{n\times n}\}.$
- 3. $\mathbf{F}(t) \in V_0 \Leftrightarrow \mathbf{F}(t-k) \in V_0 \ \forall k \in \mathbb{Z}.$
- 4. $\mathbf{F}(t) \in V_j \Leftrightarrow \mathbf{F}(2^j t) \in V_0 \ \forall j \in \mathbb{Z}.$

¹⁶Note that Xia and Suter (1996) refers to our notion of matrix MRA as vector MRA.

- 5. There exists $\mathbf{\Phi} \in V_0$ such that its integer translates $\mathbf{\Phi}(t-k)$, $k \in \mathbb{Z}$ form an orthonormal basis for V_0 .
- Φ is a $(n \times n)$ scaling function, and we say that Φ generates the matrix MRA.

Remark 5.9. Given an $n \times n$ scaling function Φ , the $n \times n$ MMRA it generates is unique. However, $\{V_j\}$ is also generated by $O\Phi(t-k)$ for any orthogonal matrix Oand $k \in \mathbb{Z}$. This ambiguity can be resolved by assuming that $\int_{-\infty}^{\infty} \Phi(t) dt = I_n$ and that Φ has compact support [0, L-1] for some L > 1.

If the signal space considered is $L^2(\mathbb{R}, \mathbb{R}^n)$ — or equivalently $L^2(\mathbb{R}, \mathbb{R}^{1 \times n})$ for ease of comparison — then the following definition arises (Chen and Cheng, 2007):

Definition 5.10. A (orthogonal n-dimensional) vector MRA is a sequence of closed linear spaces $V_j \subset L^2(\mathbb{R}, \mathbb{R}^{1 \times n}), j \in \mathbb{Z}$ satisfying

- 1. $V_j \subset V_{j-1} \ \forall j \in \mathbb{Z}$.
- 2. $\bigcup_{j\in\mathbb{Z}} V_j$ is dense in $L^2(\mathbb{R}, \mathbb{R}^{1\times n})$ and $\bigcap_{j\in\mathbb{Z}} V_j = \{\mathbf{0}_{1\times n}\}.$

3.
$$\boldsymbol{f}(t) \in V_0 \Leftrightarrow \boldsymbol{f}(t-k) \in V_0 \ \forall k \in \mathbb{Z}$$

- 4. $\boldsymbol{f}(t) \in V_j \Leftrightarrow \boldsymbol{f}(2^j t) \in V_0 \ \forall j \in \mathbb{Z}.$
- 5. There exist $\phi_1, \ldots, \phi_n \in V_0$ such that their integer translates $\phi_i(t-k), k \in \mathbb{Z}$, $i = 1, \ldots, n$ form an orthonormal basis for V_0 .

The ϕ_i are vector scaling functions, and we say that they generate the vector MRA.

Proposition 5.11. An $n \times n$ matrix-valued function Φ generates a MMRA $\{V_j\}$ if and only if its rows $\Phi^{(i,\bullet)}$ generate a VMRA $\{S_j\}$. Furthermore, we then have $V_j = S_j^n$.

Proof. For the "if" case, by Proposition 5.3 we can write $V_j = S_j^n$. For the "only if" case, set $V_j = S_j^n$ (the uniqueness of this construction then follows from the "if" case). We need to show that closedness and conditions 1 to 5 are satisfied by S_j^n iff they are satisfied by S_j . For conditions 1, 3 and 4 this is trivial.

For closedness, note that the norms in $L^2(\mathbb{R}, \mathbb{R}^{n \times n})$ and $L^2(\mathbb{R}, \mathbb{R}^{1 \times n})$ are related by $||\mathbf{F}||^2 = \sum_{i=0}^n ||\mathbf{F}^{(i,\bullet)}||^2$. For condition 2, this implies that a sequence $\mathbf{F}_k \in \bigcup_{j \in \mathbb{Z}} S_j^n = \left(\bigcup_{j \in \mathbb{Z}} S_j\right)^n$ converges to \mathbf{F} iff for each i the sequence $\mathbf{F}_k^{(i,\bullet)}$ converges to $\mathbf{F}^{(i,\bullet)}$. For condition 5, this follows from Remark 5.7 and

 $\boldsymbol{F}(t) = \sum_{k \in \mathbb{Z}} \boldsymbol{A}_k \boldsymbol{\Phi}(t-k) \Leftrightarrow \forall i, \ \boldsymbol{F}^{(i,\bullet)}(t) = \sum_{k \in \mathbb{Z}} \sum_{j=1}^n a_{i,j,k} \boldsymbol{\Phi}^{(j,\bullet)}(t-k),$

where $a_{i,j,k}$ is the (i, j)-entry of A_k .

The (a) MVW associated with a MMRA V_j is a function $\Psi \in L^2(\mathbb{R}, \mathbb{R}^{n \times n})$ such that its integer translates $\Psi(t-k), k \in \mathbb{Z}$ form a $\mathbb{R}^{n \times n}$ -orthonormal basis of

$$V_{-1} \ominus V_0 = \left\{ \boldsymbol{F} \in V_{-1} : \left\langle \boldsymbol{F}, \boldsymbol{\Phi}(t-k) \right\rangle_{n \times n} = \boldsymbol{0}_{n \times n} \; \forall k \in \mathbb{Z} \right\},\$$

the orthogonal complement of V_0 in V_{-1} . Then $2^{\frac{j}{2}}\Psi(2^{-j}t-k)$, $j,k \in \mathbb{Z}$ form a $\mathbb{R}^{n \times n}$ -orthonormal basis of $L^2(\mathbb{R}, \mathbb{R}^{n \times n})$ since $\bigcup_{j \in \mathbb{Z}} (V_{j-1} \ominus V_j) = \bigcup_{j \in \mathbb{Z}} V_j$ is dense in $L^2(\mathbb{R}, \mathbb{R}^{n \times n})$. Similarly to the proof of Proposition 5.11, the rows of Ψ form an orthonormal basis of $L^2(\mathbb{R}, \mathbb{R}^{1 \times n})$.

In a theoretical setting, whether to use a vector or matrix MRA formulation is largely a matter of taste. MMRAs allow us to think of MVWs as $\mathbb{R}^{n \times n}$ -algebra-valued wavelets. This conveniently leads to formulas and notation which are very similar to the familiar real and complex cases. However, it is the vector MRA which describes the correct *practical* application of MVWs, and the use by e.g. Walden and Serroukh (2002) of MMRA in a practical setting should be considered inappropriate.

Consider a matrix-valued signal $\mathbf{F} \in L^2(\mathbb{R}, \mathbb{R}^{n \times n})$. An arbitrary DWT coefficient (a coefficient in the wavelet basis) is given by

$$\mathcal{W}_{\boldsymbol{\Psi}}(\boldsymbol{F})(2^{j},2^{j}k) = \left\langle \boldsymbol{F}, 2^{\frac{j}{2}}\boldsymbol{\Psi}(2^{-j}t-k) \right\rangle = 2^{\frac{j}{2}} \int_{-\infty}^{\infty} \boldsymbol{F}(t)\boldsymbol{\Psi}^{T}(2^{-j}t-k) \mathrm{d}t$$

Hence each row $\mathcal{W}_{\Psi}(\mathbf{F})(2^{j}, 2^{j}k)^{(i,\bullet)}$ depends only on the corresponding row $\mathbf{F}^{(i,\bullet)}$, and a DWT (or continuous wavelet transform) of a matrix-valued signal is equivalent

to n independent vector-valued transforms of its rows (see also Appendix B.6).

If we instead consider the vector-valued signal $\operatorname{vec}(\mathbf{F})^T \in L^2(\mathbb{R}, \mathbb{R}^{1 \times n^2})$, a truly holistic analysis can be obtained by using an appropriate $n^2 \times n^2$ MVW. The matrixvalued approach corresponds in this latter context to using the $n^2 \times n^2$ MVW $\Psi(t) \otimes \mathbf{I}_n$.

Although treating a n^2 -dimensional signal as $n \times n$ -matrix-valued is inappropriate, one may choose to treat an *n*-dimensional signal $\mathbf{f} \in L^2(\mathbb{R}, \mathbb{R}^{1 \times n})$ as $n \times n$ -matrixvalued. This can be done by setting $\mathbf{F}^{(i,\bullet)} = \delta_{i,1}\mathbf{f}$, $i = 1, \ldots, n^{17}$ or alternatively by setting $\mathbf{F}^{(i,\bullet)} = \mathbf{f}$, $i = 1, \ldots, n$ as done by Fowler and Hua (2002a). Such approaches handle vector-valued signals without requiring an explicit theory of vector MRA.

Remark 5.12. Generalising the Karhunen-Loève transform to the matrix algebra case leads to similar issues. See Appendix A.3.

5.4 Matrix-valued scaling filters

Let Φ be a $n \times n$ scaling function associated with a MMRA $\{V_j\}$. $\Phi \in V_0 \subset V_{-1}$, hence it satisfies the two-scale dilation equation

$$\mathbf{\Phi}(t) = \sqrt{2} \sum_{k \in \mathbb{Z}} \mathbf{G}_k \mathbf{\Phi}(2t - k).$$
(5.5)

We call the sequence of $n \times n$ matrices $\{G_k\}$ the matrix-valued scaling filter (MVSF).

We will assume that $\{G_k\}$ is of the form $\ldots, \mathbf{0}_{n \times n}, G_0, \ldots, G_{L-1}, \mathbf{0}_{n \times n}, \ldots$, where L is the finite length of the filter.¹⁸ This is equivalent to assuming that Φ has compact support [0, L-1] as per Remark 5.9 (Strang and Nguyen, 1996, pp. 185-186).

For a MVW Ψ , since $\Psi \in V_{-1} \ominus V_0 \subset V_{-1}$ we have

$$\Psi(t) = \sqrt{2} \sum_{k \in \mathbb{Z}} \boldsymbol{H}_k \Phi(2t - k).$$
(5.6)

¹⁷Compare this row-embedding with the column-embedding of Xia (1997, p. 9), where the MVW transform is effectively performed independently on each component as a type of redundant multi-wavelet transform.

¹⁸This assumption is made without loss of generality for finite-length filters. Most results will remain valid for filters of infinite-length in $\ell^2(\mathbb{R}^{n \times n})$ (i.e. such that $\sum_{k \in \mathbb{Z}} ||\mathbf{G}_k||^2 < \infty$).

We call $\{H_k\}$ the matrix-valued wavelet filter.

The matrix Fourier transform which we will use is simply a scalar Fourier transform applied to each entry, i.e.

$$\hat{\Phi}(f) = \int_{-\infty}^{\infty} \Phi(t) e^{-i2\pi f t} dt$$

$$\hat{G}(f) = \sum_{k \in \mathbb{Z}} G_k e^{-i2\pi f k}.$$
(5.7)

In the frequency domain (5.5) becomes

$$\hat{\Phi}(f) = \frac{1}{\sqrt{2}} \hat{G}\left(\frac{f}{2}\right) \hat{\Phi}\left(\frac{f}{2}\right).$$
(5.8)

The assumption $\int_{-\infty}^{\infty} \Phi(t) dt = I_n$ from Remark 5.9 is given in the frequency domain by $\hat{\Phi}(0) = I_n$. Since $\Phi(t)$ has compact support, $\hat{\Phi}(f)$ is continuous. By iterating (5.8) we obtain in the limit

$$\hat{\Phi}(f) = \prod_{m=1}^{\infty} \frac{\hat{G}(f/2^m)}{\sqrt{2}}.$$
 (5.9)

Note that by convention the product expands from left to right. (5.9) allows us to compute the scaling function from the filter coefficients. In practice, values are computed on a dyadic grid by truncating the infinite product after finitely many terms, as explained by Walden and Serroukh (2002, Appendix A). Thus we may concentrate on designing the filter $\{G_k\}$. This amounts to choosing n^2L real scalar values for the coefficient entries.

In the remainder of this section we will give necessary conditions for $\{G_k\}$ to be a valid MVSF, and also express further design conditions in terms of G_k . These lead to a system of quadratic (and linear) equations in n^2L real variables which we solve in Section 5.8 to produce novel MVWs.

Unlike the scalar case, filter lengths may a-priori be odd or even. In order to easily cover both cases by a single equation, we define L' to be the even length of a filter of

length L, i.e.

$$L' = \begin{cases} L & \text{if } L \text{ is even} \\ L+1 & \text{if } L \text{ is odd} \end{cases}$$

We may think of a filter $\{G_k\}$ with odd length L as a filter with even length L' = L+1satisfying $G_{L'-1} = \mathbf{0}_{n \times n}$.

Setting f = 0 in (5.8) gives us the scaling equation

$$\hat{\boldsymbol{G}}(0) = \sum_{k=0}^{L'-1} \boldsymbol{G}_k = \sqrt{2} \boldsymbol{I}_n.$$
(5.10)

Fowler and Hua (2002a) refer to the property (5.10) as "omnidirectional balancing". It is the condition which sets MVWs apart from standard multiwavelets.¹⁹ Walden and Serroukh (2002) noted that (5.10) implies that the filter $\{G_k\}$ preserves constant signals.

We intend to work with $\{G_k\}$ with no a-priori knowledge of Φ or $\{V_j\}$. Thus, we must check that Φ (and by extension $\{V_j\}$) is well-defined through (5.9).

Corollary 5.13. Let $\{G_k\}$ be a finite-length filter satisfying (5.10). Then the infinite product (5.9) converges uniformly on compact sets.

Proof. This is an immediate corollary of Heil and Colella (1996, Proposition 5.2), since $\left(\frac{\hat{G}(0)}{\sqrt{2}}\right)^{\infty} = I_n^{\infty} = I_n$.

5.4.1 Orthogonality

Proposition 5.14. Orthonormality of $\{\Phi(t-k)\}$ implies

$$\sum_{k=0}^{L'-1-2m} \boldsymbol{G}_k \boldsymbol{G}_{k+2m}^T = \delta_{m,0} \boldsymbol{I}_n, \ m = 0, \dots, (L'/2) - 1.$$
 (5.11)

¹⁹For standard multiwavelets, one eigenvalue of $\hat{G}(0)/\sqrt{2}$ is equal to 1, and all other eigenvalues are strictly less than 1 in absolute value.

Proof. Using (5.5)

$$\begin{split} \delta_{m,0} \boldsymbol{I}_n &= \langle \boldsymbol{\Phi}(t), \boldsymbol{\Phi}(t+m) \rangle_{n \times n} \\ &= \sum_{k \in \mathbb{Z}} \sum_{\ell \in \mathbb{Z}} \boldsymbol{G}_k \left\langle \sqrt{2} \boldsymbol{\Phi}(2t-k), \sqrt{2} \boldsymbol{\Phi}(2t+2m-\ell) \right\rangle_{n \times n} \boldsymbol{G}_\ell^T \\ &= \sum_{k \in \mathbb{Z}} \sum_{\ell \in \mathbb{Z}} \delta_{k,\ell-2m} \boldsymbol{G}_k \boldsymbol{G}_\ell^T \\ &= \sum_{k \in \mathbb{Z}} \boldsymbol{G}_k \boldsymbol{G}_{k+2m}^T \end{split}$$

(5.11) is a necessary (but not sufficient) condition for orthogonality of $\{\Phi(t-k)\}$. A sufficient (but not necessary) condition is given by

Proposition 5.15. Let $\{G_k\}$ be a finite length filter satisfying (5.10) and (5.11). If $\det(\hat{G}(f)) \neq 0$ for $|f| \leq \frac{1}{4}$, then $\Phi(t)$ defined by (5.9) is a matrix-valued scaling function for a MMRA.

Proof. This is a reformulation of (Xia, 1997, Theorem 3.4) (See also He and Yu (2005, Theorem 2.2)) which requires that $\inf_{|f|<\frac{1}{4}} |\lambda(f)| > 0$ for all eigenvalue functions $\lambda(f)$ of $\hat{\boldsymbol{G}}(f)$. All eigenvalues are non-zero iff their product, the determinant, is non-zero. This remains true in the infimum limit since

$$|\lambda(f)| \le ||\hat{\boldsymbol{G}}(f)|| \le \sum_{k=1}^{L} ||\boldsymbol{G}_L|| < \infty$$

 $\det\left(\hat{G}(f)\right)$ is a finite trigonometric polynomial. Hence it is continuous and

$$\inf_{|f| < \frac{1}{4}} \left| \det \left(\hat{\boldsymbol{G}}(f) \right) \right| = 0 \Leftrightarrow \exists f \in \left[-\frac{1}{4}, \frac{1}{4} \right] : \det \left(\hat{\boldsymbol{G}}(f) \right) = 0.$$

Unlike Cui et al. (2009); He and Huang (2012); Walden and Serroukh (2002), we will not focus on satisfying this technical sufficient condition in our MVSF designs.

We will instead check the sufficient condition *after* obtaining explicit formulas for our MVSF constructions. In all cases it was satisfied. Because we will construct only Daubechies MVSFs, we will see in Remark 5.32 that this is actually unnecessary.

5.4.2 Vanishing moments

Definition 5.16. The MVSF $\{G_k\}$ has A vanishing moments iff

$$\sum_{k=0}^{L-1} (-1)^k k^d \boldsymbol{G}_k = \boldsymbol{0}_{n \times n}, \ d = 0, \dots, A-1.$$
(5.12)

As in the scalar case, an alternate formulation of the vanishing moment condition (5.12) is that

$$\hat{\boldsymbol{G}}(f) = (1 + e^{-i2\pi f})^A \hat{\boldsymbol{J}}(f)$$
 (5.13)

for some filter $\{J_k\}$ of length L - A. Having A vanishing moments for the MVSF is equivalent to the ability of (linear combinations of shifted versions of) the scaling filter to reproduce (matrices of) polynomials of order A (Bacchelli et al., 2002, Theorem 3.1). Vanishing moments are also desirable because they are related to the smoothness of Φ : Having A vanishing moments is a necessary (but not sufficient) condition for the existence of an A-fold derivative $\frac{d^A \Phi}{dt^A} \in L^2(\mathbb{R}, \mathbb{R}^{n \times n})$ (Micchelli and Sauer, 1997, Theorem 5.1).

Proposition 5.17. Every MVSF has at least one vanishing moment. Proof. Let $\mathbf{X} = \sum_{k \in \mathbb{Z}} (-1)^k \mathbf{G}_k$. Then using (5.11)

$$2\boldsymbol{I}_{n}\boldsymbol{I}_{n}^{T} + \boldsymbol{X}\boldsymbol{X}^{T} = \left(\sum_{k\in\mathbb{Z}}\boldsymbol{G}_{k}\right)\left(\sum_{k\in\mathbb{Z}}\boldsymbol{G}_{k}\right) + \left(\sum_{k\in\mathbb{Z}}(-1)^{k}\boldsymbol{G}_{k}\right)\left(\sum_{k\in\mathbb{Z}}(-1)^{k}\boldsymbol{G}_{k}\right)^{T}$$
$$= 2\sum_{k\in\mathbb{Z}}\sum_{\ell\in\mathbb{Z}}\left(1 + (-1)^{k+\ell}\right)\boldsymbol{G}_{k}\boldsymbol{G}_{\ell}^{T}$$
$$= 2\sum_{m\in\mathbb{Z}}\sum_{k\in\mathbb{Z}}\boldsymbol{G}_{k}\boldsymbol{G}_{k+2m}^{T}$$
$$= 2\sum_{m\in\mathbb{Z}}\delta_{m,0}\boldsymbol{I}_{n}$$
$$= 2\boldsymbol{I}_{n}.$$

Hence $\boldsymbol{X}\boldsymbol{X}^T = \boldsymbol{0}_{n \times n}$, the trace of which implies $\boldsymbol{X} = \boldsymbol{0}_{n \times n}$. (see also Walden and Serroukh (2002, eqn. (2.5)).)

5.4.3 The fast matrix-valued wavelet transform

Most of the results in this section can also be found in Xia and Suter (1996, Section V). However, our conventions differ, so we reformulate them here for clarity. The J^{th} level DWT of a vector-valued signal $\mathbf{f} \in L^2(\mathbb{R}, \mathbb{R}^{1 \times n})$ decomposes it into a linear combination

$$\mathbf{f}(t) = \sum_{k \in \mathbb{Z}} \mathbf{s}_{J,k} 2^{-J/2} \Phi(2^{-J}t - k) + \sum_{j \le J} \sum_{k \in \mathbb{Z}} \mathbf{w}_{j,k} 2^{-j/2} \Psi(2^{-j}t - k),$$
(5.14)

where $s_{J,k}, w_{j,k} \in \mathbb{R}^{1 \times n}$ are called respectively the scaling and wavelet coefficients. This decomposition is directly related to the notion of vector MRA, since it follows the decomposition of $L^2(R, R^{1 \times n})$ into the orthogonal subspaces V_J and $V_{j-1} \ominus V_j$, $j = J, J - 1, \ldots$

We will assume that $\mathbf{f}(t) \in V_0$, and that we are given the 0th level scaling coefficients $\mathbf{s}_{0,k}$. Often, for a discrete or discretely sampled signal, one will simply set $\mathbf{s}_{0,k} = \mathbf{f}(k)$ instead. Also, for signals of finite-length T, we will assume that periodic boundary conditions are imposed, i.e. $\mathbf{s}_{0,k} = \mathbf{s}_{0,k \mod T}$. This induces periodicity in the wavelet and scaling coefficients, allowing for a non-redundant transform.

Proposition 5.18. The coefficients $s_{J,k}$ and $w_{j,k}$ in (5.14) can be obtained through the fast wavelet transform (a.k.a. Mallat's pyramid algorithm) by iteratively computing

$$\boldsymbol{s}_{j+1,k} = \sum_{\ell=2k}^{2k+L-1} \boldsymbol{s}_{j,\ell} \boldsymbol{G}_{\ell-2k}^{T}$$
$$\boldsymbol{w}_{j+1,k} = \sum_{\ell\in\mathbb{Z}} \boldsymbol{s}_{j,\ell} \boldsymbol{H}_{\ell-2k}^{T}.$$
(5.15)

The original signal $s_{0,k}$ can then be recovered through the reconstruction algorithm

which iteratively computes

$$\boldsymbol{s}_{j-1,k} = \sum_{\ell \in \mathbb{Z}} \left(\boldsymbol{s}_{j,\ell} \boldsymbol{G}_{k-2\ell} + \boldsymbol{w}_{j,\ell} \boldsymbol{H}_{k-2\ell} \right).$$
(5.16)

Proof. See Appendix B.6

Note that the fast MVW transform differs from the fast multiwavelet transform (of a vectorised scalar signal) only in the choice of filters.

For matrix-valued signals $\mathbf{F} \in L^2(\mathbb{R}, \mathbb{R}^{n \times n})$, the same algorithm applies, with $\mathbf{s}_{j,k}, \mathbf{w}_{j,k}$ replaced by matrix coefficients in $\mathbb{R}^{n \times n}$. This is however equivalent to independent transforms of the rows (see Appendix B.6).

Although the scaling and wavelet functions are important for interpreting MRA and the transform coefficients, only the filters $\{G_k\}$ and $\{H_k\}$ are required to compute a DWT.

5.4.4 Computing matrix-valued wavelet filters

In the scalar case (n = 1) it is well known that a wavelet filter $\{h_k\}$ can be computed from a scaling filter $\{g_k\}$ by the simple quadrature mirror relationship

$$h_k = (-1)^{k+1} g_{L-1-k}.$$
(5.17)

Xia and Suter (1996) note that in order for the construction (5.17) to be valid for matrix-valued filters, $\hat{G}(f)$ should commute with $\hat{G}\left(f+\frac{1}{2}\right)$ for all values of f. This condition will hold in the case of 2 × 2 MVSFs with complex structure – and more generally for trivial filters which are orthogonally similar to a direct sum of filters for which it holds – but is very restrictive in the general matrix case.

Chen et al. (2006, Corollary 1) give a procedure for the computation of matrixvalued wavelet filters from MVSFs of length $L \leq 3$. We will show however, that these are all cases where (5.17) is applicable in Corollary 5.28.

Yu (2011) suggests the general construction

$$\hat{\boldsymbol{H}}(f) = e^{-2\pi f \mathbf{i}} \boldsymbol{P}\left(f + \frac{1}{2}\right) \boldsymbol{U}^{H}(f),$$

based on a frequency-by-frequency polar decomposition $G^{H}(f) = U(f)P(f)$, where U(f) is unitary and P(f) is Hermitian positive semi-definite. However, this construction will in general lead to a matrix-valued wavelet filter of infinite length.

The method we will use for matrix-valued wavelet filter computation is paraunitary completion of the polyphase matrix, as suggested by Xia and Suter (1996). This method is applicable to generalised multiwavelets, of which both multiwavelets and MVWs are special cases (Keinert, 2003, Corollary 10.2).

In Ginzberg and Walden (2013a, Section VII) we describe the paraunitary completion method in detail, and by using the formulation of Keinert (2003, Theorem 9.2) we note that the resulting matrix-valued wavelet filter will have length at most L'.

In practice we will perform paraunitary completion using the projection_factorization function from the mw Matlab toolbox by Keinert (2004). This function supports both numeric and symbolic computation. We present our Matmatrix-valued lab code for wavelet filter computation using projection_factorization in Appendix C.1.

Remark 5.19. For a given MVSF there are infinitely many possible choices of matrixvalued wavelet filter (and hence of MVW). Any two such filters $\{H_k\}$ and $\{J_k\}$ are related by

$$\hat{\boldsymbol{H}}(f) = \hat{\boldsymbol{O}}(2f)\hat{\boldsymbol{J}}(f) \tag{5.18}$$

for some paraunitary $\hat{O}(f)$. Conversely, if $\{J_k\}$ is a valid matrix-valued wavelet filter for a given MVSF, then so is $\{H_k\}$ defined by (5.18).²⁰

Proof. The existence of $\hat{O}(f)$ is a reformulation of (Keinert, 2003, Theorem 10.1), and the converse follows from Xia and Suter (1996, Proposition 1).

²⁰Note in particular that O(f) may be taken to be a constant orthogonal matrix.
5.5 Trivial matrix-valued scaling filters

We will consider in this section methods for constructing new MVSFs, which we label 'trivial', from existing MVSFs. Excluding trivial and orthogonally similar MVSFs from our later constructions of Daubechies MVSFs will allow us to significantly reduce the number of free parameters.

5.5.1 Orthogonal similarity

Definition 5.20. Two filters $\{G_k\}$ and $\{J_k\}$ are orthogonally similar iff there exists an orthogonal matrix O such that

$$\boldsymbol{G}_k = \boldsymbol{O} \boldsymbol{J}_k \boldsymbol{O}^T, \forall k \in \mathbb{Z}.$$
(5.19)

We will refer to maps of the form $\mathbf{M} \mapsto \mathbf{O}\mathbf{M}\mathbf{O}^T$ and of the form $\{\mathbf{J}_k\} \mapsto \{\mathbf{O}\mathbf{J}_k\mathbf{O}^T\}$, where \mathbf{O} is an orthogonal matrix, as orthogonal similarity transformations (OSTs).

OSTs account for $\frac{n(n-1)}{2}$ degrees of freedom in the design of MVSFs.

Proposition 5.21. If $\{G_k\}$ is an MVSF of length L with A vanishing moments, then any orthogonally similar filter $\{OG_kO^T\}$ is also an MVSF of length L with A vanishing moments. If furthermore the matrix-valued scaling function Φ associated with $\{G_k\}$ generates a MMRA $\{V_j\}$, then the matrix-valued scaling function associated with $\{OG_kO^T\}$ generates the MMRA $\{V_jO^T\}$, where $V_jO^T = \{FO^T : F \in V_j\}$.

Proof. This mostly follows from the fact that an OST is a *-algebra automorphism of $\mathbb{R}^{n \times n}$. For example, it is clear that $\{OG_kO^T\}$ has A vanishing moments by applying the OST to both sides of (5.12). By (5.9), the scaling function corresponding to $\{OG_kO^T\}$ is $O\Phi(t)O^T$, which is orthogonal to its integer shifts since

$$\left\langle \boldsymbol{O}\boldsymbol{\Phi}(t-k)\boldsymbol{O}^{T},\boldsymbol{O}\boldsymbol{\Phi}(t-l)\boldsymbol{O}^{T}\right\rangle_{n\times n}=\boldsymbol{O}\left\langle \boldsymbol{\Phi}(t-k),\boldsymbol{\Phi}(t-l)\right\rangle_{n\times n}\boldsymbol{O}^{T}=\delta_{k,l}\boldsymbol{I}_{n}$$

Finally, consider an arbitrary $F(t) = \sum_{k \in \mathbb{Z}} A_k \Phi(t-k) \in V_0$. Then

$$\boldsymbol{F}(t)\boldsymbol{O}^{T} = \sum_{k\in\mathbb{Z}} \left(\boldsymbol{A}_{k}\boldsymbol{O}^{T}\right) \left(\boldsymbol{O}\boldsymbol{\Phi}(t-k)\boldsymbol{O}^{T}\right).$$

Hence $\{\boldsymbol{O}\boldsymbol{\Phi}(t-k)\boldsymbol{O}^T\}$ is an orthogonal basis for $V_0\boldsymbol{O}^T$.

Orthogonal similarity is an equivalence relation, and hence it makes sense to work on filters 'up to (or modulo) orthogonal similarity'. The following helps us with choosing a representative filter.

Lemma 5.22. Every $M \in \mathbb{R}^{n \times n}$ is orthogonally similar to a matrix of the form D + A, where D is diagonal and A is anti-symmetric, (i.e., $A = -A^T$).

Proof. M = S + B where $S = \frac{1}{2}(M + M^T)$ is symmetric and $B = \frac{1}{2}(M - M^T)$ is anti-symmetric. By the real spectral theorem $S = ODO^T$ for some orthogonal matrix O and diagonal matrix D. M is orthogonally similar to $O^TMO = D + A$ where $A = O^TBO = -O^TB^TO = -A^T$ is anti-symmetric.

Corollary 5.23. Given a filter $\{G_k\}$, we may assume up to orthogonal similarity that $G_0 = D + A$ for some diagonal matrix D and anti-symmetric matrix A.

Note that Corollary 5.23 will usually not be sufficient to select a *unique* representative element from an equivalence class of orthogonally similar matrices, since for example OSTs where O is a permutation matrix preserve diagonal and anti-symmetric matrices. However, when all diagonal entries of D are different, assuming that they appear in decreasing order does fix a unique representative element.

5.5.2 Decomposition of filters

Consider a diagonal $n \times n$ scaling function

$$\mathbf{\Phi}(t) = \bigoplus_{i=1}^{n} \phi_n(t) = \begin{pmatrix} \phi_1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \phi_n \end{pmatrix}.$$

Then each ϕ_i is a scalar scaling function. The vector MRA $\{V_j\}$ generated by $\boldsymbol{\Phi}$ simply contains vectors of functions, the *i*th entry of which belongs to the scalar MRA generated by ϕ_i . The MVSF $\{\boldsymbol{G}_k\}$ will also be a diagonal direct sum of scalar scaling filters, and the matrix wavelet transform of a vector signal will be given by independent scalar wavelet transforms of its components. Such an approach clearly does *not* offer a holistic alternative to the naive use of scalar wavelets (which corresponds to the special case where all the ϕ_i are equal, i.e. $\boldsymbol{\Phi}(t) = \phi(t) \boldsymbol{I}_n$).

This is true more generally of any MVSF which can be split into a direct sum of lower-dimensional components, since the signal space can then also be split into corresponding subspaces being analysed independently.

Definition 5.24. A filter (resp. scaling function or wavelet) is trivial²¹ iff it is orthogonally similar to the direct sum of two (or more) filters (resp. scaling functions or wavelets), i.e. to a block-diagonal filter (resp. scaling function or wavelet).

Theorem 5.25. Every filter $\{G_k\}$ is orthogonally similar to a direct sum of nontrivial filters, i.e.

$$\boldsymbol{G}_{k} = \boldsymbol{O}\left(\bigoplus_{i=1}^{m} \boldsymbol{J}_{k}^{(i)}\right) \boldsymbol{O}^{T},$$
(5.20)

where **O** is an orthogonal matrix and each $\{\mathbf{J}_{k}^{(i)}\}, i = 1, ..., m$ is non-trivial $(m \ge 1)$.

Proof. The theorem holds for non-trivial filters by taking m = 1. All scalar (1×1) filters are non-trivial, hence the theorem holds for n = 1. We proceed by strong induction on n. Every trivial $(n+1) \times (n+1)$ filter is orthogonally similar to a direct sum of filters, each of which is of size at most $n \times n$. We may assume that each of those filters in turn is orthogonally similar to a direct sum of non-trivial filters. Since direct sums of orthogonal matrices are orthogonal, and the product of two orthogonal matrices is orthogonal, this completes the proof.

Remark 5.26. Let each $\{J_k^{(i)}\}$ in (5.20) be a $n_i \times n_i$ MVSF of length L_i^{22} with A_i

 $^{^{21} \}rm One$ may wish based on convention to refer to trivial (resp. non-trivial) filters as 'non-simple' (resp. simple) instead.

 $^{^{22}\}mathrm{In}$ the sense that $\boldsymbol{J}_k^{(i)} = \boldsymbol{0}_{n_i \times n_i}$ for k < 0 and for $k \geq L_i$

vanishing moments. Then G_k is a $n \times n$ MVSF of length L with A vanishing moments where

$$n = \sum_{i=1}^{m} n_i$$
$$L = \max_i L_i$$
$$A = \min_i A_i.$$

Proof. This follows from Proposition 5.21.

Despite Remark 5.26, trivial filters may satisfy some desirable properties which are absent in the filters from which they are assembled. One such example is the 'symmetric-antisymmetric' condition, which implies that all the matrix entries are linear-phase (see Fowler and Hua (2002b); Ginzberg and Walden (2013a, Section IV.A)).

5.5.3 Computational complexity

An advantage of trivial filters is that their wavelet transforms can be computed through lower-dimensional transforms (in the appropriate basis of \mathbb{R}^n), and this require less computation than a general implementation of non-trivial filters.

Multiplication of a vector by a general $n \times n$ matrix requires n^2 multiplications and n(n-1) additions. For a block-diagonal matrix this can be broken down into lower-dimensional products, and in the extreme case where the matrix is diagonal, only n multiplications and 0 additions are required.

Compared with (block-)diagonal filters, trivial filters will however generally require changing the signal to and from the basis of \mathbb{R}^n in which the filter is block-diagonal. The number of operations in each case is given in Table 5.1. Asymptotically for large n the computational complexity when using a non-trivial filter is $O(n^2)$, wherease for a trivial filter it is $O(\max_i n_i^2)$.

5.5.4 Triviality of MVSFs of length $L \leq 3$

Proposition 5.27. Every non-scalar MVSF of length L = 3 is trivial.

Chapter 5. Quaternion Wavelets and Matrix-Valued Wavelets

filter type	multiplications	additions
diagonal	$nLN\gamma$	$n(L-1)N\gamma$
highly trivial	$2nN + nLN\gamma$	$2(n-1)N + n(L-1)N\gamma$
trivial	$2nN + \sum_{i=1}^{m} n_i^2 LN\gamma$	$2(n-1)N + \sum_{i=1}^{m} n_i(n_iL - 1)N\gamma$
non-trivial	$n^2 LN\gamma$	$n(nL-1)N\gamma$

Table 5.1: Number of operations required for a $n \times n$ matrix wavelet transform when the scaling and wavelet filters are diagonal, diagonal up to orthogonal similarity (highly trivial), block-diagonal up to orthogonal similarity (trivial) or non-trivial. Here L is the length of the filters, N is the length of the signal, and $1 \leq \gamma = (2 - 2^{1-J}) < 2$ where $J \leq \log_2(N)$ is the number of transform levels computed.

Proof. (5.10) is

$$oldsymbol{G}_0+oldsymbol{G}_1+oldsymbol{G}_2=\sqrt{2}oldsymbol{I}_n.$$

Proposition 5.17 and (5.12) imply

$$\boldsymbol{G}_0 - \boldsymbol{G}_1 + \boldsymbol{G}_2 = \boldsymbol{0}_{n \times n}.$$

Subtracting the latter from the former we have

$$G_1 = 2^{-\frac{1}{2}} I_n. \tag{5.21}$$

Hence

$$G_2 = 2^{-\frac{1}{2}} I_n - G_0. \tag{5.22}$$

These allow us to express (5.11) in terms of G_0 only. Adding twice the equation for m = 1 to the equation for m = 0 we have

$$\boldsymbol{G}_{0}\boldsymbol{G}_{0}^{T} + 2^{-1}\boldsymbol{I}_{n} + (2^{-\frac{1}{2}}\boldsymbol{I}_{n} - \boldsymbol{G}_{0})(2^{-\frac{1}{2}}\boldsymbol{I}_{n} - \boldsymbol{G}_{0})^{T} + 2\boldsymbol{G}_{0}(2^{-\frac{1}{2}}\boldsymbol{I}_{n} - \boldsymbol{G}_{0})^{T} = \boldsymbol{I}_{n},$$

which simplifies to

$$oldsymbol{G}_0 - oldsymbol{G}_0^T = oldsymbol{0}_{n imes n}.$$

 G_0 has no antisymmetric part, and hence by Corollary 5.23, up to an OST, G_0 is diagonal. By (5.21) and (5.22), G_1 and G_2 are then also diagonal.

Corollary 5.28. Every $n \times n$ MVSF of length $L \leq 3$ is of the form

$$\boldsymbol{G}_{k} = \boldsymbol{O}\left(d_{k}\boldsymbol{I}_{m} \oplus d_{2-k}\boldsymbol{I}_{n-m}\right)\boldsymbol{O}^{T},$$

where O is an orthogonal matrix, $0 \le m \le n$, and $\{d_k\}$ are the coefficients of the scalar Haar filter, i.e. $d_0 = d_1 = 2^{-\frac{1}{2}}$ and $d_k = 0$ otherwise.

Proof. This follows from Proposition 5.27. There are no truly odd-length scalar scaling filters, and the only scalar scaling filter with length L = 2 is the Haar filter. By Remark 5.26 this implies that the only scalar scaling filters which can be components of a diagonal MVSF of length $L \leq 3$ are $\{d_k\}$ and $\{d_{2-k}\} = \{d_{k-1}\}$. The order of these diagonal elements can be fixed without loss of generality, since permutation matrices are orthogonal.

Corollary 5.29. The only $n \times n$ MVSF of length L = 2 is the matrix Haar filter

$$\boldsymbol{G}_0 = \boldsymbol{G}_1 = \frac{1}{\sqrt{2}} \boldsymbol{I}_n.$$

Proof. This follows immediately from Corollary 5.28, since $\{G_k\}$ is invariant under OSTs. It can also be shown directly from Proposition 5.17, (5.12) and (5.10).

Corollary 5.28 implies that the (orthogonal) MVSFs of length $L \leq 3$ found in the literature are either trivial (e.g. Walden and Serroukh, 2002, Design 1 and Design 2(i)); or incorrect (e.g. Chen et al., 2006, Example 2).²³

5.6 Daubechies matrix-valued scaling filters

Definition 5.30. A MVSF of length L = 2A with A vanishing moments is a Daubechies MVSF. Corresponding matrix-valued wavelets (resp. scaling functions or wavelet filters) are Daubechies matrix-valued wavelets (resp. scaling functions or wavelet filters).

 $^{^{23}}$ The filter given by Chen et al. (2006, Example 2) does not satisfy (5.10) and generates a multiwavelet suitable for the analysis of scalar signals rather than a matrix-value wavelet.

Definition 5.30 reduces to the usual Daubechies wavelets of Daubechies (1988) in the scalar case, and also generalises the complex Daubechies wavelets of Lina and Mayrand (1995) (see Section 5.7).

Proposition 5.31. In the frequency domain, every $n \times n$ Daubechies MVSF $\{G_k\}$ of length L is of the form

$$\hat{\boldsymbol{G}}(f) = \hat{\boldsymbol{U}}(f)\hat{g}(f),$$

where $\{g_k\}$ is the²⁴ scalar Daubechies scaling filter of length L, $\hat{U}(f)$ is a (normalised) paraunitary matrix (i.e. $\hat{U}(f)\hat{U}(f)^H = I_n$), and $\hat{U}(0) = I_n$.

Proof. Let G(z) be the z-transform of the scaling filter $\{G_k\}$, i.e.

$$\boldsymbol{G}(z) = \sum_{k \in \mathbb{Z}} \boldsymbol{G}_k z^{-k}$$

Note that $\hat{\boldsymbol{G}}(f) = \boldsymbol{G}(e^{2\pi f i})$ and that filter convolution is equivalent to polynomial multiplication in the z-transform domain. In particular, setting $m = k - \ell$,

$$\sum_{\ell \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \boldsymbol{G}_{k} \boldsymbol{G}_{k+\ell}^{T} \boldsymbol{z}^{-\ell} = \left(\sum_{k \in \mathbb{Z}} \boldsymbol{G}_{k} \boldsymbol{z}^{-k} \right) \left(\sum_{m \in \mathbb{Z}} \boldsymbol{G}_{m} \boldsymbol{z}^{m} \right)$$
$$= \boldsymbol{G}(\boldsymbol{z}) \boldsymbol{G}(\boldsymbol{z}^{-1})^{T}.$$

Orthonormality of $\{G_k\}$ with respect to its integer shifts would be written in the ztransform domain as $G(z)G(z^{-1})^T = I_n$. Since for any Laurent polynomial J(z) the even coefficients are given by $\frac{1}{2}(J(z) + J(-z))$, orthonormality of $\{G_k\}$ with respect to its even shifts (5.11) can be written in the z-transform domain as²⁵

$$\frac{1}{2} \left(\boldsymbol{G}(z) \boldsymbol{G}^{T}(z^{-1}) + \boldsymbol{G}(-z) \boldsymbol{G}^{T}(-z^{-1}) \right) = \boldsymbol{I}_{n}.$$

Let $Q(z) = z^{L-1}G(z)G^{T}(z^{-1})$. Then the above can be written as a polynomial equa-

²⁴Except for L = 2, there are multiple Daubechies scaling filters of a given length. We may however choose $\{g_k\}$ to be any particular one, e.g. the minimum phase filter.

²⁵Note that setting $z = e^{2\pi f i}$ this gives us in particular the Fourier-domain characterisation of MVSF orthonormality $\hat{\boldsymbol{G}}(f)\hat{\boldsymbol{G}}(f)^{H} + \hat{\boldsymbol{G}}\left(f + \frac{1}{2}\right)\hat{\boldsymbol{G}}\left(f + \frac{1}{2}\right)^{H}$.

tion

$$\boldsymbol{Q}(z) - \boldsymbol{Q}(-z) = 2z^{L-1}\boldsymbol{I}_n.$$

The left hand side is twice the odd coefficients of Q(z), and hence this equation implies that the polynomials in the off-diagonal entries of Q(z) contain only even powers of z. A polynomial containing only even powers is a symmetric function, and hence its roots come in pairs r, -r. By (5.13), a MVSF $\{G_k\}$ of length L is a Daubechies MVSF iff each entry of $G(z^{-1})$ has $\frac{L}{2}$ roots at -1. Hence each entry of Q(z) must have L roots at -1. The off-diagonal entries must then also have L roots at 1 for a total of 2L roots. The entries of $G(z^{-1})$ have degree at most L - 1 and hence the entries of Q(z) have degree at most 2(L - 1) < 2L. Since 0 is the only polynomial having more roots than its degree, this implies that the off-diagonal entries of Q(z)must be 0, i.e. Q(z) is diagonal.

The diagonal entries of Q(z) satisfy the design equations found in the original derivation of the scalar Daubechies wavelets (Daubechies, 1988, Section 4.B), where it is shown that there exists a unique minimum-degree solution. Hence $Q(z) = q(z)I_n$, where $q(z) = z^{L-1}g(z)g(z^{-1})$ and g(z) is the z-transform of the (a) Daubechies scaling filter $\{g_k\}$ of length L.²⁶

Let $U(z) = \frac{1}{q(z)}G(z)$. Then

$$\begin{aligned} \boldsymbol{U}(z) \boldsymbol{U}(z^{-1})^T &=& \frac{z^{L-1}}{z^{L-1} g(z) g(z^{-1})} \boldsymbol{G}(z) \boldsymbol{G}(z^{-1})^T \\ &=& \frac{1}{q(z)} \boldsymbol{Q}(z) \\ &=& \boldsymbol{I}_n. \end{aligned}$$

Finally set $z = e^{2\pi f i}$.

A paraunitary filter is a filter which preserves for each frequency the total signal power across all channels, generalising the concept of a scalar all-pass filter. In particular, a paraunitary filter applied to white noise will not affect its statistical properties.

²⁶Since the minimum-degree solution for Q(z) has degree 2(L-1), this implies that $G(z^{-1})$ must have degree L-1. Hence there are no MVSFs with A vanishing moments of length L < 2A.

Proposition 5.31 is a generalisation of the fact that, for a given filter length, different scalar Daubechies scaling filters differ only by an all-pass filter.

We may obtain a corresponding Daubechies matrix-valued wavelet filter by setting $\hat{H}(f) = \hat{U}(f)\hat{h}(f)$, where $\{h_k\} = \{(-1)^k g_{L-k-1}\}$ is the corresponding scalar Daubechies wavelet filter. Note however that this wavelet filter may have infinite length, unlike the length L wavelet filter constructed in Section 5.4.4. With this choice of wavelet filter, the DWT obtained using a Daubechies MVW differs from the DWT obtained using the corresponding scalar Daubechies wavelet only through a pre-filtering of the input by the paraunitary filter $\hat{U}(f)$ at each step.

Remark 5.32. If the sufficient condition of Proposition 5.15 holds for a scalar Daubechies scaling filter, then Proposition 5.31 implies that it holds for all Daubechies MVSFs of same length.

Proof. For each
$$f$$
, $\hat{U}(f)$ is unitary and hence $\left|\det\left(\hat{G}(f)\right)\right| = |\hat{g}(f)|^n$.

5.6.1 Triviality of Daubechies MVSFs of length $L \leq 4$

Proposition 5.33. Every non-scalar Daubechies MVSF of length $L \leq 4$ is trivial.

Proof. By Proposition 5.27 we need only prove the case L = 4.

(5.10) and (5.12) give us

$$G_0 + G_1 + G_2 + G_3 = \sqrt{2}I_n$$
 (5.23)

$$G_0 - G_1 + G_2 - G_3 = 0_n$$
 (5.24)

$$-G_1 + 2G_2 - 3G_3 = 0_n.$$
 (5.25)

This system simplifies to

$$G_1 = 2^{-3/2} I_n + G_0 \tag{5.26}$$

$$G_2 = 2^{-1/2} I_n - G_0 \tag{5.27}$$

$$G_3 = 2^{-3/2} I_n - G_0. (5.28)$$

(5.26) is obtained by adding $\frac{1}{2}$ times (5.25) and subtracting $\frac{5}{4}$ times (5.24) from $\frac{1}{4}$ times (5.23). (5.27) is obtained by adding $\frac{1}{2}$ times (5.24) to $\frac{1}{2}$ times (5.24). (5.28) is obtained by adding $\frac{3}{4}$ times (5.24) and subtracting $\frac{1}{2}$ times (5.25) from $\frac{1}{4}$ times (5.23).

(5.26)-(5.28) allow us to write (5.11) in terms of G_0 only. Adding $2^{-1/2}$ times the equation for m = 0 to $2^{1/2}$ times the equation for m = 1 gives us

$$2^{-\frac{1}{2}}\boldsymbol{I}_{n} = \left(4 \cdot 2^{-\frac{1}{2}} - 2 \cdot 2^{\frac{1}{2}}\right)\boldsymbol{G}_{0}\boldsymbol{G}_{0}^{T} + \left(2 \cdot 2^{-\frac{1}{2}-3} + 2^{-\frac{1}{2}-1} + 2^{\frac{1}{2}-3}\right)\boldsymbol{I}_{n} \\ + \left(2^{-\frac{1}{2}-\frac{3}{2}} - 2^{-\frac{1}{2}-\frac{1}{2}} - 2^{-\frac{1}{2}-\frac{3}{2}} + 2^{\frac{1}{2}-\frac{1}{2}} + 2^{\frac{1}{2}-\frac{3}{2}}\right)\boldsymbol{G}_{0} \\ + \left(2^{-\frac{1}{2}-\frac{3}{2}} - 2^{-\frac{1}{2}-\frac{1}{2}} - 2^{-\frac{1}{2}-\frac{3}{2}} - 2^{\frac{1}{2}-\frac{3}{2}}\right)\boldsymbol{G}_{0}^{T},$$

which simplifies to

$$oldsymbol{G}_0 = oldsymbol{G}_0^T$$

 G_0 has no antisymmetric part, and hence by Corollary 5.23, up to an OST, G_0 is diagonal. By (5.26)-(5.28), G_1 , G_2 and G_3 are then also diagonal.

Corollary 5.34. Every $n \times n$ Daubechies MVSF of length L = 4 is of the form

$$\boldsymbol{G}_{k} = \boldsymbol{O}\left(d_{k}\boldsymbol{I}_{m} \oplus d_{3-k}\boldsymbol{I}_{n-m}\right)\boldsymbol{O}^{T},$$

where O is an orthogonal matrix, $0 \le m \le n$ and $\{d_k\}$ is the scalar Daubechies minimum phase (a.k.a. extremal phase or minimum delay) scaling filter of length 4:

$$d_0 = \frac{1+\sqrt{3}}{4\sqrt{2}}; d_1 = \frac{3+\sqrt{3}}{4\sqrt{2}}; d_2 = \frac{3-\sqrt{3}}{4\sqrt{2}}; d_3 = \frac{1-\sqrt{3}}{4\sqrt{2}},$$

and $d_k = 0$ otherwise.

Proof. Similarly to the proof of Corollary 5.28, this follows from Proposition 5.33 because this Daubechies filter and its time-reversed (maximum phase) version are the only scalar Daubechies filters of length L = 4.

5.7 Matrix representation of quaternion and algebra-valued wavelets

Let \mathbb{A} denote an arbitrary *n*-dimensional real *-algebra with involution $\overline{\bullet}$. We may treat $L^2(\mathbb{R}, \mathbb{A})$ as a (left-) \mathbb{A} -module, with symbol 'inner product'

$$\langle f_1, f_2 \rangle_{\mathbb{A}} = \int_{-\infty}^{\infty} f_1(t) \overline{f_2(t)} \mathrm{d}t.$$

Definition 5.35. An (orthogonal) A-valued-MRA is a sequence of closed sub-Amodules $V_j \subset L^2(\mathbb{R}, \mathbb{A}), j \in \mathbb{Z}$ satisfying

- 1. $V_j \subset V_{j-1} \ \forall j \in \mathbb{Z}$.
- 2. $\bigcup_{i \in \mathbb{Z}} V_j$ is dense in $L^2(\mathbb{R}, \mathbb{A})$ and $\bigcap_{i \in \mathbb{Z}} V_j = \{0\}$.
- 3. $f(t) \in V_0 \Leftrightarrow f(t-k) \in V_0 \ \forall k \in \mathbb{Z}.$
- 4. $f(t) \in V_j \Leftrightarrow f(2^j t) \in V_0 \; \forall j \in \mathbb{Z}.$
- 5. There exists $\phi \in V_0$ such that its integer translates $\phi(t-k)$, $k \in \mathbb{Z}$ form an \mathbb{A} -orthonormal basis for V_0 .

 ϕ is an A-valued scaling function, and we say that ϕ generates the A-valued-MRA.

Every *n*-dimensional real algebra \mathbb{A} is a vector space isomorphic to \mathbb{R}^n . We may thus define a vector space isomorphism $\mathcal{V}^* : \mathbb{A} \to \mathbb{R}^{1 \times n}$. In the case of quaternions $(\mathbb{A} = \mathbb{H})$, for consistency we take $\mathcal{V}^*(\bullet) = \mathcal{V}(\bar{\bullet})^T$.

Every such vector isomorphism defines a unique algebra isomorphism $\tilde{\bullet}^* : \mathbb{A} \to \tilde{\mathbb{A}}^* \subseteq \mathbb{R}^{n \times n}$ by letting \tilde{a}^* be the linear transformation $\mathcal{V}^*(b) \mapsto \mathcal{V}^*(ba)$. Note that here we choose to think of matrices as multiplying row vectors on the right. In the case of quaternions we have $\tilde{\bullet}^* = \tilde{\bullet}$.

Also for consistency with the quaternion approach, we may assume without loss of generality that $\mathcal{V}^*(1) = (1, 0, 0, \dots, 0)$, so that $\mathcal{V}^*(x) = \mathcal{V}^*(1)\tilde{x}^*$ is the first row of \tilde{x}^* .

 $\tilde{\bullet}^*$ will be a *-algebra isomorphism iff the involution $\bar{\bullet}$ satisfies

$$\widetilde{\overline{x}}^* = \widetilde{x}^{*T}.$$
(5.29)

This can however be assumed: For an algebra \mathbb{A} without an a-priori *-algebra structure, we will *define* $\overline{\bullet}$ to be the unique involution satisfying (5.29).

Lemma 5.36. Any two algebra representations $\widetilde{\bullet}^*, \widetilde{\bullet}^* : \mathbb{A} \to \mathbb{R}^{n \times n}$, of an n-dimensional real algebra \mathbb{A} satisfy $\widetilde{\bullet}^* = M \widetilde{\bullet}^* M^{-1}$ for some $M \in GL_n(\mathbb{R})$. If furthermore $\widetilde{\bullet}^*$ and $\widetilde{\bullet}^*$ are *-algebra representations, then M is orthogonal.

Proof. Let $\mathbf{e}_1 = (1, 0, 0, \dots, 0)$ and define $\mathcal{V}^* : a \mapsto \mathbf{e}_1 \widetilde{a}^*$ and $\mathcal{V}^* : a \mapsto \mathbf{e}_1 \widetilde{a}^*$. These are both vector space isomorphisms, and hence $\mathcal{V}^* \circ \mathcal{V}^{*-1}$ is an automorphism of $\mathbb{R}^{1 \times n}$, i.e. $\mathcal{V}^*(\bullet) = \mathbf{M} \mathcal{V}^*(\bullet)$ for some $\mathbf{M} \in GL_n(\mathbb{R})$.

 $\tilde{\bullet}^*$ (resp. $\tilde{\bullet}^*$) can in turn be obtained from \mathcal{V}^* (resp. \mathcal{V}^*) as described above. Hence for $a \in \mathbb{A}$, \tilde{a}^* is the linear transformation

$$x \mapsto \mathcal{V}^{\star}\left(\mathcal{V}^{\star-1}(x)a\right) = x \mapsto \mathcal{M}\mathcal{V}^{\star}\left(\mathcal{V}^{\star-1}(\mathcal{M}^{-1}x)a\right),$$

and $\widetilde{a}^{\star} = M \widetilde{a}^{\star} M^{-1}$.

If $\tilde{\bullet}^*$ is a *-algebra representation, then

$$\mathcal{V}^*(a) \, \mathcal{V}^*(a)^T = \mathbf{e}_1 \widetilde{a}^* \widetilde{a}^{*T} \mathbf{e}_1^T$$
$$= \mathbf{e}_1 \widetilde{a} \overline{\overline{a}}^* \mathbf{e}_1^T$$
$$= \mathbf{e}_1 (a \overline{a}) \mathbf{I}_n \mathbf{e}_1^T$$
$$= a \overline{a}.$$

Hence \mathcal{V}^* is an isometry. Similarly, \mathcal{V}^* is also an isometry, and hence $\mathcal{V}^* \circ \mathcal{V}^{*-1}$ is an isometry and M is orthogonal.

Proposition 5.37. Let \mathbb{A} be an n-dimensional real algebra and let $\phi \in L^2(\mathbb{R}, \mathbb{A})$ be a scaling function generating an \mathbb{A} -valued MRA $\{S_j\}$. Then the rows of its matrix representation $\widetilde{\phi(t)}^{*(i,\bullet)}$, i = 1, ..., n generate the n-dimensional vector MRA $\{\mathcal{V}^*(S_j)\}$.

Proof. The topologies on $L^2(\mathbb{R}, \mathbb{A})$ and $L^2(\mathbb{R}, \mathbb{R}^{1 \times n}) = L^2(\mathbb{R}, \mathcal{V}^*(\mathbb{A}))$ are equivalent, since by choosing $\overline{\bullet}$ appropriately, \mathcal{V}^* is an isometry (see the proof of Lemma 5.36). Also, $\langle \bullet, \bullet \rangle_{\mathbb{A}}$ - and $\langle \widetilde{\bullet}^*, \widetilde{\bullet}^* \rangle_{n \times n}$ -orthogonality (and hence orthogonality of the rows) are equivalent. It remains to show that (right-)A-linear combinations of $\phi(t-k)$ correspond to real linear combinations of the $\widetilde{\phi(t)}^{*(i,\bullet)}$. This follows from

$$\mathcal{V}^*\left(\sum_{k\in\mathbb{Z}}a_k\phi(t-k)\right) = \sum_{k\in\mathbb{Z}}\mathcal{V}^*\left(a_k\right)\widetilde{\phi(t-k)}^* = \sum_{k\in\mathbb{Z}}\sum_{i=1}^n\mathcal{V}^*\left(a_k\right)_{1,i}\widetilde{\phi(t-k)}^{*(i,\bullet)}.$$

Corollary 5.38. Let \mathbb{A} by a n-dimensional real algebra and let $\phi \in L^2(\mathbb{R}, \mathbb{A})$ be a scaling function generating an \mathbb{A} -valued-MRA $\{S_j\}$. Then its matrix representation $\widetilde{\phi(t)}^*$ generates the $n \times n$ MMRA $\{\mathcal{V}^*(S_j)^n\}$.

Proof. This follows from Proposition 5.37 and Proposition 5.11. \Box

Definition 5.8 is a special case of Definition 5.35, with $\mathbb{A} = \mathbb{R}^{n \times n}$, i.e. a MMRA is an $\mathbb{R}^{n \times n}$ -algebra-valued-MRA. Corollary 5.38 however shows that algebra-valued scaling functions (resp. wavelets or filters) can be seen as a special case of matrixvalued scaling functions (resp. wavelets or filters). The proof of Proposition 5.37 also shows that conversely, if the rows of $\phi(t)^*$ generates a vector MRA, then ϕ generates an \mathbb{A} -valued-MRA. In other words, the matrix-valued scaling functions (resp. wavelets) corresponding to \mathbb{A} -valued scaling functions (resp. wavelets) are precisely those with the corresponding matrix structure, i.e. those in $L^2(\mathbb{R}, \widetilde{\mathbb{A}}^*)$.

Note that $\{\tilde{S}_j^*\}$ will not be a MMRA (except for $\mathbb{A} = \mathbb{R}$), since it is a (left-) \mathbb{A}^* -module and not a (left-) $\mathbb{R}^{n \times n}$ -module. Because \mathbb{A}^* and $\mathcal{V}^*(\mathbb{A})$ are isomorphic as vector spaces, the *structured* matrix MRA of $L^2(\mathbb{R}, \mathbb{A}^*)$ generated by the matrix representation of an \mathbb{A} -valued scaling function and the vector MRA of $L^2(\mathbb{R}, \mathcal{V}^*(\mathbb{A}))$ are however equivalent. For example, the quaternion fast wavelet transform can be written as both the vector and matrix versions of (5.15)-(5.16), with the former parsimoniously computing only the first row of \mathbb{H} -valued coefficients appearing in the latter. This is enough to infer the remaining rows and requires the same computations as a quaternion-domain algorithm. The same applies to general real algebras.

Definition 5.39. An A-algebra-valued scaling filter $\{g_k\}$ is trivial iff, under some *-algebra representation $\tilde{\bullet}^*$, its matrix-valued image $\{\tilde{g}_k^*\}$ is trivial.

This definition does not depend on choice of $\tilde{\bullet}^*$ by Lemma 5.36.

Corollary 5.40. Let \mathbb{A} be an n-dimensional real semi-simple algebra and furthermore assume that \mathbb{A} is not simple, i.e. not isomorphic to $\mathbb{R}^{\sqrt{n} \times \sqrt{n}}$, $\mathbb{C}^{\sqrt{\frac{n}{2}} \times \sqrt{\frac{n}{2}}}$, or $\mathbb{H}^{\sqrt{\frac{n}{4}} \times \sqrt{\frac{n}{4}}}$. Then (under an appropriate choice of involution on \mathbb{A}) every \mathbb{A} -valued filter is trivial.

Proof. By Corollary 1.19 there exists a block-diagonal algebra representation of \mathbb{A} in $\mathbb{R}^{n \times n}$. Define the involution on \mathbb{A} to be the one induced by the involution \bullet^T on $\mathbb{R}^{n \times n}$, so that this is a *-algebra representation.

5.7.1 Quaternion propriety

Corollary 5.41. The wavelet transform coefficients $s_{J,k}$, $w_{j,k}$ of a quaternion DWT are jointly left-proper (resp. second-order left-proper) if and only if the signal $s_{0,k}$ is left-proper (resp. second-order left-proper).

Proof. The wavelet transform (5.15) consists entirely of quaternion left-linear operations, and the same is true of the inverse wavelet transform (5.16). Hence this follows from Corollary 2.5 (resp. Corollary 1.28).

Corollary 5.41 holds in particular for the noise component in a 'signal + noise' model.

Note that we refer to left-propriety in Corollary 5.41 whilst the rest of this thesis concentrates on right-propriety and right- \mathbb{H} -modules. One way of inverting the handedness of results would be to take the matrix transpose of all MVW-related definitions, as mentioned previously. More simply however, if the isomorphism used to interpret the quaternion DWT as a vector DWT is taken to be $\mathcal{V}(\bullet)^T$ instead of $\mathcal{V}^*(\bullet) = \mathcal{V}(\bar{\bullet})^T$, then by Proposition 2.8 the resulting transform will preserve right-propriety instead of left-propriety.

Note that quaternion proper i.i.d. Gaussian noise is both right-proper and leftproper. In this case, by considering the equivalent real vector formulation, it is clear

5.7.2 Orthogonal similarity for quaternions

For the design of quaternion scaling filters, Lemma 5.22 is unhelpful, since every matrix in $\widetilde{\mathbb{H}}$ is the sum of a diagonal matrix (corresponding to the real part) and an antisymmetric matrix. This section describes an alternative strategy for selecting a representative element amongst orthogonally similar quaternion scaling filters.

By Remark 1.14, we can write any quaternion in the form $q = a + (b^2 + c^2 + d^2)^{1/2}i_0$, where i_0 is a pure unit quaternion. Since we can rotate i_0 onto i, another way of interpreting Remark 1.14 is the following.

Remark 5.42. Let q = a + bi + cj + dk. Then there exists a unit quaternion u such that

$$uq\bar{u} = a + (b^2 + c^2 + d^2)^{1/2}i.$$
(5.30)

Proof. For example, we can take

$$u = \exp\left(\frac{b(-d\mathbf{j} + c\mathbf{k})}{2(b^2 + c^2 + d^2)^{\frac{1}{2}}(c^2 + d^2)^{\frac{1}{2}}}\right).$$

 $u \in \mathbb{H}$ is a unit quaternion iff \tilde{u} is an orthogonal matrix,²⁷ and hence for any unit quaternion u the 3D rotation $q \mapsto uq\bar{u}$ is an OST.

Lemma 5.43. Let $q_0, q_1 \in \mathbb{H}$. Then there exists a unit quaternion u such that

$$\Im_{\mathbf{j}}(uq_0\bar{u}) = \Im_{\mathbf{k}}(uq_0\bar{u}) = \Im_{\mathbf{j}}(uq_1\bar{u}) = 0$$

²⁷This follows from Definition 1.6 and Theorem 1.26 since $|u|^2 = u\bar{u} = 1 \Leftrightarrow \tilde{u}\tilde{u}^T = I_4$.

5.7 Matrix representation of quaternion and algebra-valued wavelets 160

Proof. By Remark 5.42 there exists a unit quaternion v such that $\Im_j(vq_0\bar{v}) = \Im_k(vq_0\bar{v}) = 0$. Let

$$w = \exp\left(-\frac{1}{2}\tan\left(\frac{\Im_{j}(vq_{1}\bar{v})}{\Im_{k}(vq_{1}\bar{v})}\right)i\right).$$

Then $wvq_0\bar{v}\bar{w} = vq_0\bar{v}$ and $\Im_j(wvq_1\bar{v}\bar{w}) = 0$. Setting u = wv completes the proof. \Box

Note that a quaternion filter (or function) is orthogonally similar to its conjugate, since $\tilde{\bar{q}} = \mathbf{O}\tilde{q}\mathbf{O}^T$ with

$$\boldsymbol{O} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

even though $O \notin \widetilde{\mathbb{H}}^{.28}$ This implies that the set of right- and left- quaternion scaling filters are equal.

As we already mentioned, a quaternion filter is trivial iff it is real or complex (with respect to some imaginary unit i_0), since $\widetilde{\mathbb{R}}$ and $\widetilde{\mathbb{C}}$ are respectively the diagonal and the block-diagonal matrices in $\widetilde{\mathbb{H}}$.

5.7.3 The biquaternion Fourier transform

Biquaternions are an 8-dimensional real algebra isomorphic to $\mathcal{C}l_{3,0}(\mathbb{R})$ (and to $\mathcal{C}l_{0,2}(\mathbb{C})$) obtained by allowing the coefficients a, b, c, d of a quaternion to be complex-valued, thus introducing a new imaginary unit which commutes with i, j, k.

Frequency-domain interpretation of matrix-valued filters and functions relies on the matrix-valued Fourier transform (5.7). This can also be applied to the special case of quaternion-structured MVWs. The matrix Fourier transform can then be interpreted as a biquaternion Fourier transform, by extending $\tilde{\bullet}$ to a representation of biquaternions in $\mathbb{C}^{4\times 4}$.

For complex wavelets the usual complex Fourier transform is not directly equivalent to the Fourier transform for complex structured 2×2 real matrices, which

 $^{^{28}}$ It is not clear to us whether more generally given a MVSF $\{\boldsymbol{G}_k\},\;\{\boldsymbol{G}_k^T\}$ will also be a valid MVSF.

transforms the real and imaginary parts independently. However both approaches are valid.

5.8 Examples of non-trivial Daubechies MVSFs

Like Fowler and Hua (2002b); Hua and Fowler (2002); Peng and Zhao (2004), we will design the scaling filters $\{G_k\}$ by directly solving a set of polynomial design equations. Our method is implemented in Appendix C.2 as a Maple worksheet. We will consider in particular Daubechies MVSFs, but the approach can be used for any design constraints which can be expressed as polynomial equations.

For an $n \times n$ Daubechies scaling filter of length L, the polynomial system is composed of L + 1 matrix equations, and hence $n^2(L + 1)$ (scalar) equations. These are respectively n^2 linear equations from the single matrix scaling equation (5.10), $n^2A = n^2\frac{L}{2}$ linear equations from the vanishing moment conditions (5.12), and $n^2\frac{L'}{2}$ quadratic equations from the necessary orthogonality conditions (5.11).

The n^2L unknowns in this system of equations can be reduced to $n^2L - \frac{n(n-1)}{2}$ by Corollary 5.23, or to nL unknowns when working with an *n*-dimensional real algebra. If the MVSF is assumed to have quaternion structure, then the number of unknowns can be further reduced to 4L - 3 by Lemma 5.43.

We first solve the linear equations in the system. This leaves us with a system of quadratic equations with fewer unknowns. This set of polynomials is pre-processed by computing a lexicographic Gröbner basis. This is a particular set of polynomials which has the same (complex) roots as our original system (because it generates the same ideal), but can be more readily solved. Lebrun and Selesnick (2004) give a detailed introduction to Gröbner bases, and use them in a similar approach to design multiwavelets.

To obtain a Gröbner basis, first an ordering of the unknowns is chosen, and this induces a lexicographic ordering of monomials.²⁹ The lexicographic Gröbner basis

²⁹If the unknowns are ordered as $x_1 > x_2 > \ldots$, then the lexicographic monomial ordering is given by $x_1^{i_1} x_2^{i_2} x_3^{i_3} \cdots > x_1^{i_1} x_2^{i_2} x_3^{i_3} \cdots$ if $i_1 > j_1$; or if $i_1 = j_1$ and $i_2 > j_2$; or if $i_1 = j_1$, $i_2 = j_2$ and $i_3 > j_3$; etc.

can then be computed by a procedure similar to Gaussian elimination, eliminating at each step the largest monomial (in the lexicographic ordering) from all but one of the remaining polynomials. The polynomial obtained at the last step will contain only the smallest monomials and hence only the smallest unknowns. Once this polynomial is solved, the remaining unknowns can be obtained by a kind of back-substitution. We simply use the Groebner:-Basis command in Maple for Gröbner basis computation.

Note that in order to avoid the presence of the irrational constant $\sqrt{2}$ from (5.10) in the system of equations, and thus limit polynomial coefficients to the field of rational numbers and accelerate computation, we will use the entries of $\{\sqrt{2}\mathbf{G}_k\}$ as our unknowns instead of the entries of $\{\mathbf{G}_k\}$.³⁰

5.8.1 The 2 \times 2 Daubechies MVSFs of length L = 6

Solving the design equations corresponding to the 2×2 Daubechies MVSF of length L = 6,³¹ we obtain — in addition to the trivial diagonal solutions — the following non-trivial family of solutions.

³⁰Under some authors' conventions, $\{\sqrt{2}G_k\}$ is defined as the scaling filter, rather than $\{G_k\}$.

³¹This corresponds to setting the parameters to n = 2, L = 6, Nvm = 0, Nvmplus = 3, cstr = false, qstr = false and sym = false in Appendix C.2.

$$G_{0} = \frac{1}{32\sqrt{2}} \begin{pmatrix} x^{2} - 2x - 3 & y \\ -y & x^{2} + 2x - 3 \end{pmatrix},$$

$$G_{1} = \frac{1}{32\sqrt{2}} \begin{pmatrix} x^{2} - 6x + 5 & y \\ -y & x^{2} + 6x + 5 \end{pmatrix},$$

$$G_{2} = \frac{1}{16\sqrt{2}} \begin{pmatrix} -x^{2} - 2x + 15 & -y \\ y & -x^{2} + 2x + 15 \end{pmatrix},$$

$$G_{3} = \frac{1}{16\sqrt{2}} \begin{pmatrix} -x^{2} + 2x + 15 & -y \\ y & -x^{2} - 2x + 15 \end{pmatrix},$$

$$G_{4} = \frac{1}{32\sqrt{2}} \begin{pmatrix} x^{2} + 6x + 5 & y \\ -y & x^{2} - 6x + 5 \end{pmatrix},$$

$$G_{5} = \frac{1}{32\sqrt{2}} \begin{pmatrix} x^{2} + 2x - 3 & y \\ -y & x^{2} - 2x - 3 \end{pmatrix},$$
(5.31)

where 32

$$y = \sqrt{-x^4 + 10x^2 + 15},$$

and x is a free parameter. Since y must be real, the free parameter is limited to $|x| \leq \sqrt{5 + 2\sqrt{10}} \approx 3.3652.$

The filter obtained by replacing x with -x is the time-reversal $\{G_{5-k}\}$, which is orthogonally similar to $\{G_k\}$, with

$$oldsymbol{O} = \left(egin{array}{cc} 0 & 1 \ -1 & 0 \end{array}
ight).$$

Thus we may restrict ourselves without loss of generality to $x \ge 0$.

y reaches its minimum (y = 0) for $x = \sqrt{5 + 2\sqrt{10}}$, giving the trivial diagonal scaling filter $\{d_{5-k} \oplus d_k\}$, where $\{d_k\}$ is the scalar minimum-phase Daubechies scaling

³²We may choose the positive square root without loss of generality since replacing y with -y gives an orthogonally similar filter with $O = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

filter of length L = 6 (Daubechies, 1992, Table 6.1). Setting x = 0 in (5.31) we obtain (the real matrix representation of) the symmetric complex-valued Daubechies filter of length L = 6 of Lina and Mayrand (1995, p. 222).

y reaches its maximum $(y = 2\sqrt{10} \approx 6.3246)$ for $x = \sqrt{5}$, giving the filter

$$G_{0} = \frac{1}{16\sqrt{2}} \begin{pmatrix} 1 - \sqrt{5} & \sqrt{10} \\ -\sqrt{10} & 1 + \sqrt{5} \end{pmatrix}$$

$$G_{1} = \frac{1}{16\sqrt{2}} \begin{pmatrix} 5 - 3\sqrt{5} & \sqrt{10} \\ -\sqrt{10} & 5 + 3\sqrt{5} \end{pmatrix}$$

$$G_{2} = \frac{1}{8\sqrt{2}} \begin{pmatrix} 5 - \sqrt{5} & -\sqrt{10} \\ \sqrt{10} & 5 + \sqrt{5} \end{pmatrix}$$

$$G_{3} = \frac{1}{8\sqrt{2}} \begin{pmatrix} 5 + \sqrt{5} & -\sqrt{10} \\ \sqrt{10} & 5 - \sqrt{5} \end{pmatrix}$$

$$G_{4} = \frac{1}{16\sqrt{2}} \begin{pmatrix} 5 + 3\sqrt{5} & \sqrt{10} \\ -\sqrt{10} & 5 - 3\sqrt{5} \end{pmatrix}$$

$$G_{5} = \frac{1}{16\sqrt{2}} \begin{pmatrix} 1 + \sqrt{5} & \sqrt{10} \\ -\sqrt{10} & 1 - \sqrt{5} \end{pmatrix}.$$
(5.32)

The matrix-valued wavelet filter corresponding to (5.31), was then obtained through the Matlab implementation in Appendix C.1 of the method described in Ginzberg and Walden (2013a, Section VII) (see also Section 5.4.4). It is given by

$$H_{0} = \frac{1}{176\sqrt{2}} \begin{pmatrix} -11+9\sqrt{5} & 10\sqrt{2}+11\sqrt{10} \\ 10\sqrt{2}-11\sqrt{10} & -11-9\sqrt{5} \end{pmatrix},$$

$$H_{1} = \frac{1}{176\sqrt{2}} \begin{pmatrix} 55-27\sqrt{5} & -30\sqrt{2}-11\sqrt{10} \\ -30\sqrt{2}+11\sqrt{10} & 55+27\sqrt{5} \end{pmatrix},$$

$$H_{2} = \frac{1}{88\sqrt{2}} \begin{pmatrix} -55+9\sqrt{5} & 10\sqrt{2}-11\sqrt{10} \\ 10\sqrt{2}+11\sqrt{10} & -55-9\sqrt{5} \end{pmatrix},$$

$$H_{3} = \frac{1}{88\sqrt{2}} \begin{pmatrix} 55+9\sqrt{5} & 10\sqrt{2}+11\sqrt{10} \\ 10\sqrt{2}-11\sqrt{10} & 55-9\sqrt{5} \end{pmatrix},$$

$$H_{4} = \frac{1}{176\sqrt{2}} \begin{pmatrix} -55-27\sqrt{5} & -30\sqrt{2}+11\sqrt{10} \\ -30\sqrt{2}-11\sqrt{10} & -55+27\sqrt{5} \end{pmatrix},$$

$$H_{5} = \frac{1}{176\sqrt{2}} \begin{pmatrix} 11+9\sqrt{5} & 10\sqrt{2}-11\sqrt{10} \\ 10\sqrt{2}+11\sqrt{10} & 11-9\sqrt{5} \end{pmatrix}.$$
(5.33)

For the filters (5.32) and (5.33) we computed the corresponding frequency responses (Fourier transforms) $\hat{G}(f)$ and $\hat{H}(f)$, the absolute values of which are shown in Figure 5.1. The scaling function $\Phi(t)$ and wavelet $\Psi(t)$ were computed according to (5.9) and (5.6), using the method described by Walden and Serroukh (2002, Appendix A), and are shown in Figures 5.2 and 5.3.

Our attempt to design a 3×3 Daubechies MVSF of length L = 6 produced no non-trivial solutions.

5.8.2 The quaternion Daubechies MVSFs of length L = 10

We show in Ginzberg and Walden (2013a, Proposition 7) that there are no odd-length MVSFs with symmetry. Similarly,

Remark 5.44. There are no complex or quaternion scaling filters of odd length.

Proof. Let $\{g_k\}$ be a scaling filter of length L. By the last equality of 5.11, we have $g_0\overline{g_{L-1}} = 0$, which implies that $g_0 = 0$ or $g_{L-1} = 0$ since \mathbb{H} (resp. \mathbb{C}) is a division algebra.



Figure 5.1: Absolute entries of the frequency responses $\hat{\boldsymbol{G}}(f)$ (full line) and $\hat{\boldsymbol{H}}(f)$ (dashed line) for the 2 × 2 Daubechies MVSF (resp. wavelet filter) of length L = 6 with parameter choice $x = \sqrt{5}$.



Figure 5.2: Entries of the scaling function $\Phi(t)$ for the 2 × 2 Daubechies MVSF of length L = 6 with parameter choice $x = \sqrt{5}$. ©IEEE. Reprinted with permission.



Figure 5.3: Entries of the wavelet $\Psi(t)$ for the 2 × 2 Daubechies MVSF of length L = 6 with parameter choice $x = \sqrt{5}$. ©IEEE. Reprinted with permission.

By Proposition 5.27 there are no non-trivial quaternion scaling filters of length $L \leq 3$ and by Proposition 5.33 there are no non-trivial quaternion Daubechies scaling filters of length L = 4.

A 4 × 4 MVSF $\{G_k\}$ is block-diagonal with 2 × 2 blocks iff its entries are roots of

$$\sum_{k=0}^{L-1} \left(g_{3,1,k}^2 + g_{4,1,k}^2 + g_{3,2,k}^2 + g_{4,2,k}^2 + g_{1,3,k}^2 + g_{2,3,k}^2 + g_{1,4,k}^2 + g_{2,4,k}^2 \right).$$

If this polynomial belongs to the ideal generated by the design equations, then all solutions must be roots and thus all solutions are block diagonal. This sufficient condition for the non-existence of non-trivial solutions can be checked as follows: Once a (not necessarily lexicographic) Gröbner basis is found for the design equations, any polynomial can be reduced to normal form by taking the remainder of (multivariate) polynomial division with respect to the elements of the basis. A polynomial belongs to the ideal generated by the design equations iff its normal form is 0. See Appendix C.2 (9).

Through the above computational procedure, we have shown that there are no non-trivial quaternion Daubechies scaling filters of lengths L = 6 and L = 8,³³ i.e.

³³The parameters in Appendix C.2 were set to n = 4, L = 6, Nvm = 0, Nvmplus = 3, cstr = false, qstr = true and sym = false (resp. n = 4, L = 8, Nvm = 0, Nvmplus = 4, cstr = false,

the only quaternion Daubechies filters of length $L \leq 8$ are the corresponding real and (for L = 6, 8) complex Daubechies filters.

The shortest non-trivial quaternion Daubechies scaling filters are obtained for L = 10, and these are discussed next. All non-trivial solutions are symmetric and can be parameterised (up to orthogonal similarity) as

$$g_{0} = g_{9} = \frac{1}{256\sqrt{2}} (y_{1} + y_{2}i)$$

$$g_{1} = g_{8} = \frac{1}{256\sqrt{2}} ((y_{1} - 10) + y_{2}^{-1} (y_{2}^{2} + 10y_{1} - 70)i + y_{3}k)$$

$$g_{2} = g_{7} = \frac{1}{256\sqrt{2}} ((-4y_{1} - 14) - 2y_{2}^{-1} (2y_{2}^{2} - 15y_{1} + 105)i + 3y_{3}k)$$

$$g_{3} = g_{6} = \frac{1}{256\sqrt{2}} ((-4y_{1} + 70) - 2y_{2}^{-1} (2y_{2}^{2} - 5y_{1} + 35)i + y_{3}k)$$

$$g_{4} = g_{5} = \frac{1}{256\sqrt{2}} ((6y_{1} + 210) + 2y_{2}^{-1} (3y_{2}^{2} - 25y_{1} + 175)i - 5y_{3}k),$$

where x is a free parameter and

$$y_1 = \sqrt{70} \cos(x)$$

$$y_2 = \sqrt{70} \sin(x)$$

$$y_3 = 2y_2^{-1} \sqrt{60y_2^2 - 8y_2^2y_1 + 350y_1 - 2975}.$$

The range of x is $1.0995 \leq x \leq 2.1764$, so that $60y_2^2 - 8y_2^2y_1 + 350y_1 - 2975 \geq 0$ and y_3 is real. The two extreme values of x lead to $y_3 = 0$, and the resulting filters are the two different symmetric complex Daubechies filters of length 10.

qstr = true and sym = false).

If we choose $x = \pi/2$, then $y_1 = 0$, $y_3 = \sqrt{70}$ and

$$g_{0} = g_{9} = \frac{\sqrt{35}}{256}i$$

$$g_{1} = g_{8} = \frac{1}{256} \left(-5\sqrt{2} + \sqrt{35}k\right)$$

$$g_{2} = g_{7} = \frac{1}{256} \left(-7\sqrt{2} - 7\sqrt{35}i + 3\sqrt{35}k\right)$$

$$g_{3} = g_{6} = \frac{1}{256} \left(35\sqrt{2} - 5\sqrt{35}i + \sqrt{35}k\right)$$

$$g_{4} = g_{5} = \frac{1}{256} \left(105\sqrt{2} + 11\sqrt{35}i - 5\sqrt{35}k\right), \qquad (5.34)$$

The quaternion wavelet filter corresponding to (5.34), was obtained by applying the Matlab implementation in Appendix C.1 of the method described in Ginzberg and Walden (2013a, Section VII) to the matrix representation $\{\tilde{g}_k\}$. It is anti-symmetric and given by

$$h_{0} = -h_{9} = \frac{1}{24576} \left(89\sqrt{35}i + 35\sqrt{2}j - 35\sqrt{35}k \right)$$

$$h_{1} = -h_{8} = \frac{1}{24576} \left(-480\sqrt{2} + 35\sqrt{35}i - 175\sqrt{2}j + 79\sqrt{35}k \right)$$

$$h_{2} = -h_{7} = \frac{1}{3072} \left(84\sqrt{2} - 91\sqrt{35}i + 35\sqrt{2}j + \sqrt{35}k \right)$$

$$h_{3} = -h_{6} = \frac{1}{256} \left(35\sqrt{2} + 5\sqrt{35}i - \sqrt{35}k \right)$$

$$h_{4} = -h_{5} = \frac{1}{12288} \left(-5040\sqrt{2} + 577\sqrt{35}i - 245\sqrt{2}j + 5\sqrt{35}k \right). \quad (5.35)$$

The (absolute) frequency response of the scaling and wavelet filter entries is shown in Figure 5.4. Quaternion scaling and wavelet functions were computed from (5.34)and (5.35) using the method of Walden and Serroukh (2002, Appendix A(b)), and are shown in Figure 5.5.



Figure 5.4: Absolute entries of the frequency responses $\hat{\boldsymbol{G}}(f)$ (full line) and $\hat{\boldsymbol{H}}(f)$ (dashed line) for the quaternion Daubechies scaling filter (resp. wavelet filter) of length L = 10 with parameter choice $x = \frac{\pi}{2}$. Note that the axes have different scales. Subscripts refer to the quaternion-structured matrix representation based on $\boldsymbol{G}_k = \tilde{g}_k$, $\boldsymbol{H}_k = \tilde{h}_k$.

5.9 On the use of MVWs in practice

As we mentioned in Section 5.4.3, the fast MVW transform is identical to the fast multiwavelet transform of a vectorised scalar signal, but for the choice of matrixvalued filters. Although multiwavelet filters behave poorly when used on vector signals (Fowler and Hua, 2002a), there is no such problem with using MVW filters on vectorised scalar signals. MVWs are *balanced* (generalised) multiwavelets, i.e. unlike unbalanced multiwavelets they do not require the use of pre- or post-processing filters. Also, through matrix MRA, the theory of MVWs is more similar to that of scalar wavelets.

Despite the above advantages, it is the author's opinion that the framework of MVWs is not well suited to the design of multiwavelets. The design condition (5.10), which applies to MVWs but not multiwavelets, greatly restricts possible constructions. For example, Strang and Strela (1994) construct a multiwavelet of length L = 3 with A = 2 vanishing moments. By Corollary 5.28 (or more generally footnote 26 p. 152) this cannot be achieved with MVWs. Heuristically, whilst MVWs of



Figure 5.5: Quaternion Daubechies scaling and wavelet functions of length L = 10, with parameter $x = \frac{\pi}{2}$. Note that the axes have different scales. Subscripts refer to the quaternion-structured matrix representation $\Phi(t) = \widetilde{\phi(t)}, \Psi(t) = \widetilde{\psi(t)}$. ©IEEE. Reprinted with permission.

length L are comparable to real or complex wavelets of length L, it may be fairer to compare multiwavelets of length L with real wavelets of length nL.

As we mentioned in the Introduction to this chapter, MVW transforms have been applied to compression and denoising of colour images (Agreste and Vocaturo, 2009a,b), and of wind field data (Hua and Fowler, 2004; Westenberg and Ertl, 2005). Westenberg and Ertl (2005) show superior denoising performance for MVWs compared to the naive use of scalar wavelets. However, as they note, this may be due to the use of vector-thresholding in the first case and scalar-thresholding in the latter, rather than the choice of wavelets. We show in Ginzberg and Walden (2013a) that our quaternion Daubechies wavelet (Figure 5.5) could outperform the corresponding real Daubechies wavelet for compressing a synthetic quaternion orientation time-series. However, further analysis has revealed that although the quaternion Daubechies wavelet outperformed the *minimum-phase* real Daubechies wavelet of length L = 10, it was in turn outperformed by the *least-asymmetric* real Daubechies wavelet of length L = 10. It is visually clear from Agreste and Vocaturo (2009a, Figure 2) that the compressed versions of the standard colour test image *Lena* obtained by MVW transform are of significantly poorer quality than those which would be obtained using a naive approach.

We know from Appendix A.1 that, with the exception of complex wavelets, the naive component-wise approach is the only one which is invariant under OSTs (and hence under rotation of the signal space and wavelets). We conjecture that in typical applications³⁴ MVWs will not outperform real wavelets unless they are tailored to take advantage of *specific* (and anisotropic) properties of interchannel correlation in the type of signal being processed. In Ginzberg and Walden (2012) (available in Appendix E) we adaptively optimise all free parameters of the family of 3×3 Daubechies MVWs of length L = 6 (all of which are trivial) to compress the colour image *Lena*. However, no significant improvement is obtained over the naive approach if we allow both methods to take advantage of *instantaneous* interchannel correlation through a simple rotation of the wavelet coefficient basis.³⁵ Although vector MRA and MVWs arguably provide the correct theoretical framework for wavelet-based analysis of vector-valued signals, further research is required to determine which combinations of signal and MVW (including algebra-valued wavelets) — if any — will lead to significant practical benefits compared to the naive use of real wavelets.

 $^{^{34}}$ In certain less typical applications, such as watermarking or — as noted by Walden and Serroukh (2002) — encryption, it is plausible that the mixing of channels obtained through the use of MVWs is in and of itself valuable. Indeed Agreste and Vocaturo (2009c) show superior performance of MVW based watermarking compared to real wavelet based watermarking for certain kinds of attack.

³⁵We optimise the rotation of the wavelet coefficient basis using a modified SIMPLIMAX algorithm. One could instead use principal component analysis on the RGB colourspace for a similar result. Robinson (2001) applies the latter approach to machine vision.

Conclusion

The set of $n \times n$ covariance matrices (or equivalently multivariate normal distributions) which are invariant under the action of some group³⁶ can be conveniently interpreted as belonging to some semi-simple real algebra. This is a major motivation for the study of statistics in algebras other than \mathbb{R} . Since all semi-simple real algebras can be constructed from the simple algebras of real, complex and quaternion matrices, these are important special cases.

As real and complex linear algebra are well studied, we turned our attention to the use of quaternion linear algebra in statistics. Despite quaternions' lack of commutativity, we note that quaternion matrices can in most respects be manipulated similarly to complex matrices. One particularly useful tool for handling quaternion matrices is the (*-)algebra isomorphism between $n \times n$ quaternion matrices and $4n \times 4n$ quaternion-structured real matrices.

Two of the most basic statistical problems involving quaternions are 'how to test whether the interpretation of a sample as being quaternion-valued is proper' and 'how to fit a quaternion multivariate linear regression (a.k.a. general linear model)'.

The former problem can be answered by the likelihood ratio test for quaternion structure in a sample covariance matrix. We have shown that the distribution of this LRT is given by a product of independent beta random variables and is of Box type. Multiple suggested approximations to this distribution were shown to be acceptably accurate. In addition, the exact distribution (CDF and PDF) of a general random variable of Box type was derived in closed form in terms of Meijer's G-function (and Fox's H-function). This exact distribution can be applied to many commonly (and

³⁶More specifically, a subgroup of the orthogonal group.

less commonly) used likelihood ratio tests, especially tests for covariance structure, and by extension group invariance.

Analysing quaternion multivariate linear regression we have shown that in addition to the usual assumption of i.i.d. vector errors, one must assume propriety of the errors to ensure that the ordinary least squares estimator is equal to the generalised least squares estimator (which is also the best linear unbiased estimator and the Gaussian maximum likelihood estimator). This result is applicable in particular to least squares estimation of the coefficients of a quaternion VAR process. More generally, group invariance for a real VAR process can be modeled by interpreting it as an A-valued process with a semi-simple algebra A. The A-linear least squares parameter estimator will then be the best linear unbiased estimator if the common covariance of the innovations is also group invariant.

In our last chapter, we considered algebraic extensions to yet another linear signal processing tool: wavelet transforms. This was done through the theory of matrixvalued wavelets, which generalise wavelet transforms and multiresolution analysis to vector signals in $L^2(\mathbb{R}, \mathbb{R}^n)$. We elucidated the fundamental equivalence of three multiresolution analysis frameworks, based on vector-valued, matrix-valued and algebravalued signals respectively. Since every finite-dimensional real algebra has a matrix representation, we may reduce the study of algebra-valued wavelets to special cases of matrix-valued wavelets. In particular, quaternion wavelets are equivalent to quaternion-structured 4×4 MVWs.

In the design of MVWs, the degrees of freedom offered by orthogonal similarity transformations can be isolated by working 'up to' or 'modulo' orthogonal similarity. We have made an important distinction between trivial wavelets — which operate independently on orthogonal subspaces of \mathbb{R}^n — and non-trivial wavelets, from which all matrix-valued wavelets can be constructed. Many examples of MVWs in the literature are orthogonally similar to a direct sum of scalar wavelets, and hence trivial. By symbolically solving a system of quadratic equations, we obtained the scaling filters corresponding to the shortest non-trivial 2×2 and quaternion-valued Daubechies wavelets.

MVWs are a promising approach to holistic processing of vector-valued signals.

However, more research is needed to understand how and for which type of signal the additional degrees of freedom available in the design of MVWs can be effectively used to improve performance over the naive use of scalar wavelets component-wise. This may require MVWs to be chosen adaptively for each signal.

Whilst some generalisations of univariate statistical tools to vector signals such as multivariate linear regression, multivariate analysis of variance, multipleinput multiple-output filters and vector autoregression — are well established and in frequent use, others — such as hypercomplex Fourier transforms and MVW transforms — are somewhat niche and not fully understood.³⁷ Replacing the real numbers used in univariate algorithms with another algebra (especially the division algebras \mathbb{C} and \mathbb{H} or commutative algebras) often requires only minor modifications. Interpreting vector-valued signals as algebra-valued can thus be an attractive approach to vector signal processing.

Methods based on real algebras should, in this author's opinion, be studied whenever possible within a wider context of vector methods. This can be achieved with matrix representations, and in many cases reduces problems to familiar real linear algebra. In particular, widely-linear methods can be simpler in their real-linear form. The ad-hoc use of algebra-based methods for vector signal processing may not be appropriate, and the wider context clarifies the implicit constraints imposed by such methods. Where there is additional signal structure imposed by known group-invariance, the use of algebras is however clearly justified.

A majority of methods in statistical signal processing are linear and based on the second-order properties of a signal. They can hence be generalised to algebra-valued signals and account for group-invariance. A general and comprehensive approach to algebra-valued signal processing would be an interesting objective for future research.

³⁷Note that vector autoregression, discrete MVW transforms and 1D discrete hypercomplex Fourier transforms are special cases of multiple-input multiple-output filtering.

References

- S. Agreste and A. Vocaturo. Multichannel wavelet scheme for color image processing. In Applied and Industrial Mathematics in Italy III: Selected Contributions from the 9th SIMAI Conference, Rome, Italy 15-19 September 2008, volume 82, page 1, 2009a.
- S. Agreste and A. Vocaturo. A new class of full rank filters in the context of digital color image processing. In *Proceedings of the 10th European Congress of ISS*, MIRIAM Project, pages 1–6, Bologna, Italy, 2009b. ESCULAPIO Pub. Co.
- S. Agreste and A. Vocaturo. Wavelet and multichannel wavelet based watermarking algorithms for digital color images. In *Communications to SIMAI Congress*, volume 3, pages 242–252, 2009c.
- H. Akaike. Block toeplitz matrix inversion. SIAM Journal on Applied Mathematics, 24(2):234–241, 1973.
- T. W. Anderson. An introduction to multivariate statistical analysis. Wiley, 1958.
- S. A. Andersson. Invariant normal models. *The Annals of Statistics*, 3(1):132–154, 1975.
- S. A. Andersson and G. G. Wojnar. Wishart distributions on homogeneous cones. Journal of Theoretical Probability, 17(4):781–818, 2004.
- S. A. Andersson, H. K. Brons, and S. T. Jensen. Distribution of eigenvalues in multivariate statistical analysis. *The Annals of Statistics*, 11(2):392–415, 1983.

- A. Askari Hemmat and Z. Rahbani. Clifford wavelets and clifford-valued MRAs. Iranian Journal of Mathematical Sciences and Informatics, 5(1):7–18, 2010.
- H. Aslaksen. Quaternionic determinants. The Mathematical Intelligencer, 18(3):57– 65, 1996.
- S. Bacchelli. Wavelets for multichannel signals. *Advances in Applied Mathematics*, 29(4):581–598, 2002.
- S. Bacchelli, M. Cotronei, and T. Sauer. Multifilters with and without prefilters. BIT Numerical Mathematics, 42(2):231–261, 2002.
- K. Baddour and N. Beaulieu. Accurate simulation of multiple cross-correlated fading channels. In *IEEE International Conference on Communications*, 2002, volume 1, pages 267–271, New York, NY, 2002.
- M. Bahri. Construction of quaternion-valued wavelets. *MATEMATIKA*, 26(1): 107–114, 2010.
- M. Bahri, R. Ashino, and R. Vaillancourt. Two-dimensional quaternion Fourier transform of type II and quaternion wavelet transform. In 2012 International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR), pages 359–364, 2012.
- O. E. Barndorff-Nielsen and P. Hall. On the level-error after Bartlett adjustment of the likelihood ratio statistic. *Biometrika*, 75(2):374–378, 1988.
- E. W. Barnes. The theory of the gamma function. *Messenger of Mathematics*, 29: 64–128, 1899.
- T. Bayes. A letter from the late reverend Mr. Thomas Bayes, F.R.S. to John Canton, M.A. and F.R.S., 1763.
- W. E. Baylis. Applications of Clifford algebras in physics. In R. Abamowicz and G. Sobczyk, editors, *Lectures on Clifford (Geometric) Algebras and Applications*. Springer, 2004.

- E. Bayro-Corrochano. The theory and use of the quaternion wavelet transform. *Journal of Mathematical Imaging and Vision*, 24(1):19–35, 2006.
- D. S. Bernstein. *Matrix Mathematics: Theory, Facts, and Formulas*. Princeton University Press, 2nd edition, 2009.
- G. E. P. Box. A general distribution theory for a class of likelihood criteria. *Biometrika*, 36(3/4):317–346, 1949.
- F. Brackx, E. Hitzer, and S. J. Sangwine. History of quaternion and Clifford Fourier transforms and wavelets. In E. Hitzer and S. J. Sangwine, editors, *Quaternion and Clifford Fourier Transforms and Wavelets*, pages xi–xxvii. Birkhäuser, 2013.
- S. Buchholz and N. Le Bihan. Optimal separation of polarized signals by quaternionic neural networks. In *Eusipco 2006. 14th European Signal Processing Conference*, pages 4–8, 2006.
- R. W. Butler, S. Huzurbazar, and J. G. Booth. Saddlepoint approximations for the Bartlett-Nanda-Pillai trace statistic in multivariate analysis. *Biometrika*, 79(4): 705, 1992.
- R. W. Butler, S. Huzurbazar, and J. G. Booth. Saddlepoint approximations for tests of block independence, sphericity and equal variances and covariances. *Journal of* the Royal Statistical Society. Series B (Methodological), 55(1):171–183, 1993.
- B. D. Carter and M. D. Springer. The distribution of products, quotients and powers of independent H-function variates. SIAM Journal on Applied Mathematics, 33(4): 542–558, 1977.
- W. L. Chan, H. Choi, and R. Baraniuk. Coherent multiscale image processing using dual-tree quaternion wavelets. *IEEE Transactions on Image Processing*, 17(7): 1069–1082, 2008.
- Q. Chen and Z. Cheng. A study on compactly supported orthogonal vector-valued wavelets and wavelet packets. *Chaos, Solitons & Fractals*, 31(4):1024–1034, 2007.

- Q. Chen and Z. Shi. Construction and properties of orthogonal matrix-valued wavelets and wavelet packets. *Chaos, Solitons & Fractals*, 37(1):75–86, 2008.
- Q. J. Chen, Z. X. Cheng, and C. L. Wang. Existence and construction of compactly supported biorthogonal multiple vector-valued wavelets. *Journal of Applied Mathematics and Computing*, 22(3):101–115, 2006.
- C. Cheong Took and D. P. Mandic. The quaternion LMS algorithm for adaptive filtering of hypercomplex processes. *IEEE Transactions on Signal Processing*, 57 (4):1316–1327, 2009.
- C. Cheong Took and D. P. Mandic. Quaternion-valued stochastic gradient-based adaptive IIR filtering. *IEEE Transactions on Signal Processing*, 58(7):3895–3901, 2010a.
- C. Cheong Took and D. P. Mandic. A quaternion widely linear adaptive filter. *IEEE Transactions on Signal Processing*, 58(8):4427–4431, 2010b.
- C. Cheong Took and D. P. Mandic. Augmented second-order statistics of quaternion random signals. *Signal Processing*, 91(2):214–224, 2011.
- K. Conradsen, A. A. Nielsen, J. Schou, and H. Skriver. A test statistic in the complex Wishart distribution and its application to change detection in polarimetric SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 41:4–19, 2003.
- P. C. Consul. On the exact distributions of the criterion W for testing sphericity in a p-variate normal distribution. *The Annals of Mathematical Statistics*, 38(4): 1170–1174, 1967.
- P. C. Consul. On the exact distributions of Votaw's criteria for testing compound symmetry of a covariance matrix. *The Annals of Mathematical Statistics*, 40(3): 836–843, 1969.
- C. Conti, M. Cotronei, and T. Sauer. Full rank positive matrix symbols: interpolation and orthogonality. *BIT Numerical Mathematics*, 48(1):5–27, 2008.

REFERENCES

- J. Cook. The H-Function and Probability Density Functions of Certain Algebraic Combinations of Independent Random Variables with H-Function Probability Distribution. PhD thesis, Air Force Institute of Technology, 1981.
- C. C. Craig. A new exposition and chart for the Pearson system of frequency curves. The Annals of Mathematical Statistics, 7(1):16–28, 1936.
- J. L. Crassidis, F. L. Markley, and Y. Cheng. Survey of nonlinear attitude estimation methods. *Journal of Guidance Control and Dynamics*, 30(1):12, 2007.
- L. Cui and T. Zhang. m-band orthogonal vector-valued multiwavelets for vectorvalued signals. Journal of Applied Mathematics and Computing, 28(1-2):165–184, 2008.
- L. Cui, B. Zhai, and T. Zhang. Existence and design of biorthogonal matrix-valued wavelets. *Nonlinear Analysis: Real World Applications*, 10(5):2679–2687, 2009.
- I. Daubechies. Orthonormal bases of compactly supported wavelets. *Communications* on Pure and Applied Mathematics, 41(7):909–996, 1988.
- I. Daubechies. Ten lectures on wavelets. SIAM, 1992.
- A. W. Davis. Percentile approximations for a class of likelihood ratio criteria. Biometrika, 58(2):349–356, 1971.
- M. Davis. Quaternionic linear algebra. Technical report, 2009.
- S. Y. Dennis. On the distribution of products of independent beta variates. Communications in Statistics - Theory and Methods, 23(7):1895, 1994.
- J. Dieudonné. Les déterminants sur un corps non commutatif. Bulletin de la Société Mathématique de France, 71(171-180):95, 1943.
- T. Ell. Hypercomplex color affine filters. In *IEEE International Conference on Image Processing*. ICIP 2007, volume 5, pages 249–252, 2007.
- T. A. Ell and S. J. Sangwine. Quaternion involutions and anti-involutions. Computers & Mathematics with Applications, 53(1):137–143, 2007.
- R. Estrada and R. P. Kanwal. Asymptotic analysis: a distributional approach. Birkhäuser, 1994.
- D. R. Farenick and B. A. F. Pidkowich. The spectral theorem in quaternions. *Linear Algebra and its Applications*, 371:75–102, 2003.
- F. Foerster and G. Stemmler. When can we trust the F-approximation of the Boxtest? *Psychometrika*, 55(4):727–728, 1990.
- J. E. Fowler and L. Hua. Omnidirectionally balanced multiwavelets for vector wavelet transforms. 2002a.
- J. E. Fowler and L. Hua. Wavelet transforms for vector fields using omnidirectionally balanced multiwavelets. *IEEE Transactions on Signal Processing*, 50(12):3018– 3027, 2002b.
- M. Galassi, J. Theiler, and J. Davies. *GNU Scientific Library Reference Manual*. Network Theory Limited, 3rd edition, 2009.
- D. J. H. Garling. *Clifford Algebras: An Introduction*. Cambridge University Press, 2011.
- P. Ginzberg and A. T. Walden. Testing for quaternion propriety. *IEEE Transactions* on Signal Processing, 59(7):3025–3034, 2011.
- P. Ginzberg and A. T. Walden. Adaptive orthogonal matrix-valued wavelets and compression of vector-valued signals. In *Proceedings of the 9th IMA International Conference on Mathematics in Signal Processing*, Birmingham, UK, 2012.
- P. Ginzberg and A. T. Walden. Matrix-valued and quaternion wavelets. *IEEE Transactions on Signal Processing*, 61(6):1357–1367, 2013a.
- P. Ginzberg and A. T. Walden. Quaternion VAR modelling and estimation. *IEEE Transactions on Signal Processing*, 61(1):154–158, 2013b.

- L. Gleser and I. Olkin. A note on Box's general method of approximation for the null distributions of likelihood criteria. Annals of the Institute of Statistical Mathematics, 27(1):319–326, 1975.
- J. Glover and L. P. Kaelbling. Tracking 3-D rotations with the quaternion Bingham filter. Technical Report MIT-CSAIL-TR-2013-005, Massachusetts Institute of Technology, 2013.
- P. W. Glynn. Importance sampling for Monte Carlo estimation of quantiles. In Mathematical Methods in Stochastic Simulation and Experimental Design: Proceedings of the 2nd St. Petersburg Workshop on Simulation, page 180–185, 1996.
- N. R. Goodman. Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction). The Annals of Mathematical Statistics, 34(1):152– 177, 1963.
- A. Grandi, A. Mazzotti, and E. Stucchi. Multicomponent velocity analysis with quaternions. *Geophysical Prospecting*, 55(6):761–777, 2007.
- P. A. Grillet. Abstract Algebra. Graduate Texts in Mathematics. Springer, New York, NY, 2007.
- X. Gu and J. Jiang. A complex autoregressive model and application to monthly temperature forecasts. *Annales Geophysicae*, 23:3229–3235, 2005.
- A. Gupta. Distribution of Wilks' likelihood-ratio criterion in the complex case. Annals of the Institute of Statistical Mathematics, 23(1):77–87, 1971.
- A. Gupta and J. Tang. On a general distribution theory for a class of likelihood criteria. Australian & New Zealand Journal of Statistics, 30(3):359–366, 1988.
- S. W. R. Hamilton. *Elements of quaternions*. Longmans, Green, & co., 1866.
- J. He and S. Huang. Constructions of vector-valued filters and vector-valued wavelets. Journal of Applied Mathematics, 2012:1–18, 2012.

- J. X. He and B. Yu. Wavelet analysis of quaternion-valued time-series. International Journal of Wavelets, Multiresolution and Information Processing, 3:233–46, 2005.
- C. Heil and D. Colella. Matrix refinement equations: Existence and uniqueness. Journal of Fourier Analysis and Applications, 2:363—377, 1996.
- S. Held, M. Storath, P. Massopust, and B. Forster. Steerable wavelet frames based on the Riesz transform. *IEEE Transactions on Image Processing*, 19(3):653–667, 2010.
- E. Hitzer, T. Nitta, and Y. Kuroe. Applications of Clifford's geometric algebra. Advances in Applied Clifford Algebras, 23(2):377–404, 2013.
- L. Hua and J. E. Fowler. Technical details on a family of omnidirectionally balanced symmetric-antisymmetric multiwavelets. Technical Report MSSU-COE-ERC-02-08, Engineering Research Center, Mississippi State University, 2002.
- L. Hua and J. E. Fowler. Wavelet-based coding of time-varying vector fields of oceansurface winds. *IEEE Transactions on Geoscience and Remote Sensing*, 42(6):1283– 1290, 2004.
- B. P. Ickes. A new method for performing digital control system attitude computations using quaternions. *AIAA Journal*, 8(1):13–17, 1970.
- C. Jahanchahi, C. Cheong Took, and D. P. Mandic. The widely linear quaternion recursive least squares filter. In 2010 2nd International Workshop on Cognitive Information Processing (CIP), pages 87–92, 2010.
- C. Jahanchahi, C. Cheong Took, and D. P. Mandic. A class of quaternion valued affine projection algorithms. *Signal Processing*, 93(7):1712–1723, 2013.
- S. Javidi, C. Cheong Took, C. Jahanchahi, N. Le Bihan, and D. P. Mandic. Blind extraction of improper quaternion sources. In 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 3708–3711, 2011.

- J. L. Jensen. A large deviation-type approximation for the 'Box class' of likelihood ratio criteria. Journal of the American Statistical Association, 86(414):437–440, 1991.
- J. L. Jensen. Correction: A large deviation-type approximation for the 'Box class' of likelihood ratio criteria. Journal of the American Statistical Association, 90(430): 812, 1995.
- S. T. Jensen. Covariance hypotheses which are linear in both the covariance and the inverse covariance. *The Annals of Statistics*, 16(1):302–322, 1988.
- T. Jiang and L. Chen. Algebraic algorithms for least squares problem in quaternionic quantum theory. *Computer Physics Communications*, 176(7):481–485, 2007.
- D. G. Kabe. Classical statistical analysis based on a certain hypercomplex multivariate normal distribution. *Metrika*, 31(1):63–76, 1984.
- A. Käufl. The distribution of the maximum likelihood estimator in invariant Gaussian graphical models and its application to likelihood ratio tests. 2012.
- F. Keinert. Wavelets and Multiwavelets. Chapman and Hall/CRC, 1st edition, 2003.
- F. Keinert. Wavelets and multiwavelets, 2004. URL http://orion.math.iastate.edu/keinert/book.html.
- C. G. Khatri. Classical statistical analysis based on a certain multivariate complex Gaussian distribution. *The Annals of Mathematical Statistics*, 36(1):98–114, 1965.
- L. Krippner. The eigensystem vector autogression model (AMW discussion draft). 2010.
- N. Le Bihan and J. Mars. Singular value decomposition of quaternion matrices: a new tool for vector-sensor signal processing. *Signal Processing*, 84(7):1177–1199, 2004.
- J. Lebrun and I. Selesnick. Gröbner bases and wavelet design. Journal of Symbolic Computation, 37(2):227–259, 2004.

- C. Li, J. Li, and B. Fu. Magnitude-phase of quaternion wavelet transform for texture representation using multilevel copula. *IEEE Signal Processing Letters*, 20(8):799– 802, 2013.
- F. Li and Y. Xue. The density functions of the singular quaternion normal matrix and the singular quaternion Wishart matrix. *Communications in Statistics - Theory* and Methods, 39(18):3316–3331, 2010.
- G. V. Liakhovetski. An algorithm for a series expansion of the Meijer G-function. Integral Transforms and Special Functions, 12(1):53–64, 2001.
- J.-M. Lina and M. Mayrand. Complex daubechies wavelets. Applied and Computational Harmonic Analysis, 2(3):219–229, 1995.
- Y. Liu, J. Jin, Q. Wang, and Y. Shen. Phase-preserving speckle reduction based on soft thresholding in quaternion wavelet domain. *Journal of Electronic Imaging*, 21 (4):043009–1–043009–11, 2012.
- M. Loeve. Probability Theory I. Springer, 4th edition, 1977.
- M. T. Loots, A. Bekker, M. Arashi, and J. J. Roux. On the real representation of quaternion random variables. *Statistics*, to be published, 2012.
- R. Lugannani and S. Rice. Saddle point approximation for the distribution of the sum of independent random variables. *Advances in Applied Probability*, 12(2):475–490, 1980.
- H. Lütkepohl. New introduction to multiple time series analysis. Birkhäuser, 2006.
- Y. Maesono and S. I. Penev. Higher order relations between Cornish-Fisher expansions and inversions of saddlepoint approximations. *Journal of the Japan Statistical Society*, 28(1):21–38, 1998.
- O. I. Marichev. Handbook of integral transforms of higher transcendental functions: theory and algorithmic tables. Ellis Horwood Ltd., 1983.

- A. Mathai. A few remarks about some recent articles on the exact distributions of multivariate test criteria: I. Annals of the Institute of Statistical Mathematics, 25 (1):557–566, 1973a.
- A. Mathai, R. K. Saxena, and H. J. Haubold. The H-Function: Theory and Applications. Springer, 1st edition, 2009.
- A. M. Mathai. The exact distributions of three multivariate statistics associated with Wilks' concept of generalized variance. Sankhyā: The Indian Journal of Statistics, Series A, 34(2):161–170, 1972.
- A. M. Mathai. A review of the different techniques used for deriving the exact distributions of multivariate test criteria. Sankhyā: The Indian Journal of Statistics, Series A, 35(1):39–60, 1973b.
- A. M. Mathai. The exact distributions and the exact percentage points for testing equality of variances in independent normal populations. *Journal of Statistical Computation and Simulation*, 9(3):169, 1979.
- A. M. Mathai and R. S. Katiyar. Exact percentage points for testing independence. *Biometrika*, 66(2):353–356, 1979.
- G. M. Menanno. Seismic multicomponent deconvolution and wavelet estimation by means of quaternions. PhD thesis, Università di Pisa, Pisa Italy, 2010.
- G. M. Menanno and N. Le Bihan. Quaternion polynomial matrix diagonalization for the separation of polarized convolutive mixture. *Signal Processing*, 90(7):2219– 2231, 2010.
- C. A. Micchelli and T. Sauer. Regularity of multiwavelets. Advances in Computational Mathematics, 7:455–545, 1997.
- S. Miron, N. Le Bihan, and J. Mars. Quaternion-MUSIC for vector-sensor array processing. *IEEE Transactions on Signal Processing*, 54(4):1218–1229, 2006.

- J. Møller. Bartlett adjustments for structured covariances. Scandinavian Journal of Statistics, 13(1):1–15, 1986.
- C. Moxey, S. Sangwine, and T. Ell. Hypercomplex correlation techniques for vector images. *IEEE Transactions on Signal Processing*, 51(7):1941 – 1953, 2003.
- R. J. Muirhead. Aspects of multivariate statistical theory. Wiley, New York; Chichester, 1982.
- D. K. Nagar, J. Chen, and A. K. Gupta. Distribution and percentage points of the likelihood ratio statistic for testing circular symmetry. *Computational Statistics & Data Analysis*, 47(1):79–89, 2004.
- Y. Nakatani, D. Sasaki, Y. Iiguni, and H. Maeda. Online recognition of handwritten hiragana characters based upon a complex autoregressive model. *IEEE Transac*tions on Pattern Analysis and Machine Intelligence, 21(1):73–76, 1999.
- J. Navarro-Moreno, R. M. Fernandez-Alcala, and J. C. Ruiz-Molina. A quaternion widely linear series expansion and its applications. *IEEE Signal Processing Letters*, 19(12):868-871, 2012.
- J. Navarro-Moreno, R. M. Fernández-Alcalá, C. Cheong Took, and D. P. Mandic. Prediction of wide-sense stationary quaternion random signals. *Signal Processing*, 93(9):2573–2580, 2013.
- S. Olhede. Hyperanalytic denoising. *IEEE Transaction on Image Processing*, 16(6): 1522–1537, 2007.
- R. S. Palais. The classification of real division algebras. The American Mathematical Monthly, 75(4):366–368, 1968.
- L. Peng and J. Zhao. Quaternion-valued smooth orthogonal wavelets with short support and symmetry. In Advances in analysis and geometry: new developments using Clifford algebras, Trends in Mathematics, pages 365–376. Birkhäuser, 2004.

- T. Pham-Gia. Exact distribution of the generalized Wilks's statistic and applications. Journal of Multivariate Analysis, 99(8):1698–1716, 2008.
- B. Picinbono and P. Bondon. Second-order statistics of complex signals. *IEEE Transactions on Signal Processing*, 45(2):411–420, 1997.
- D. A. Robinson. Method and system for identifying an image feature and method and system for determining an optimal color space for use therein, 2001.
- P. T. G. Rubin-Delanchy. Some New Results in the Analysis of Complex-Valued Time Series. PhD thesis, Imperial College, London, 2008.
- A. Sajeva. Quaternion SVD Methods for the Extraction of Rayleigh Waves. PhD thesis, Università di Pisa, Pisa Italy, 2009.
- S. Sangwine and T. Ell. Colour image filters based on hypercomplex convolution. *IEE Proceedings - Vision, Image and Signal Processing*, 147(2):89–93, 2000.
- S. J. Sangwine and T. A. Ell. Complex and hypercomplex discrete Fourier transforms based on matrix exponential form of Euler's formula. *Applied Mathematics and Computation*, 219(2):644–655, 2012.
- S. J. Sangwine and N. Le Bihan. Quaternion singular value decomposition based on bidiagonalization to a real or complex matrix using quaternion Householder transformations. *Applied Mathematics and Computation*, 182(1):727–738, 2006.
- M. Schatzoff. Exact distributions of Wilks's likelihood ratio criterion. *Biometrika*, 53 (3/4):347–358, 1966.
- L. S. P. Schreier. Statistical Signal Processing of Complex-Valued Data. Cambridge University Press, 1st edition, 2010.
- J. Seberry, K. Finlayson, S. Adams, T. Wysocki, T. Xia, and B. Wysocki. The theory of quaternion orthogonal designs. *IEEE Transactions on Signal Processing*, 56(1): 256–265, 2008.

- I. Sekita, T. Kurita, and N. Otsu. Complex autoregressive model for shape recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(4): 489–496, 1992.
- I. W. Selesnick, R. G. Baraniuk, and N. C. Kingsbury. The dual-tree complex wavelet transform. Signal Processing Magazine, IEEE, 22(6):123–151, 2005.
- K. Shoemake. Animating rotation with quaternion curves. In Proceedings of the 12th annual conference on Computer graphics and interactive techniques - SIGGRAPH '85, pages 245–254, San Francisco, CA, 1985.
- P. D. Shukla. *Complex wavelet transforms and their applications*. MPhil, University of Strathclyde, Strathclyde, Scotland, 2003.
- R. Soulard and P. Carré. Quaternionic wavelets for image coding. In EUSIPCO-2010 Proceedings, page 125–129, Aalborg, Danemark, 2010.
- R. Soulard, P. Carre, and C. Fernandez-Maloigne. Vector extension of monogenic wavelets for geometric representation of color images. *IEEE Transactions on Image Processing*, 22(3):1070–1083, 2013.
- M. D. Springer. Evaluation of the H-function inversion integral for real variables using Jordan's lemma and residues. SIAM Journal on Applied Mathematics, 47(2): 416–424, 1987.
- M. D. Springer and W. E. Thompson. The distribution of products of beta, gamma and Gaussian random variables. SIAM Journal on Applied Mathematics, 18(4): 721–737, 1970.
- M. S. Srivastava and W. K. Yau. Saddlepoint method for obtaining tail probability of Wilks' likelihood ratio test. *Journal of Multivariate Analysis*, 31(1):117–126, 1989.
- G. W. Stewart. Matrix Algorithms: Basic Decompositions. Society for Industrial Mathematics, 1998.
- G. Strang and T. Nguyen. Wavelets and Filter Banks. SIAM, 1996.

- G. Strang and V. Strela. Orthogonal multiwavelets with vanishing moments. Optical Engineering, 33(7):2104–2107, 1994.
- Y. Tian. Universal similarity factorization equalities over real Clifford algebras. Advances in Applied Clifford Algebras, 8(2):365–402, 1998.
- B. C. Ujang, C. Jahanchahi, C. Cheong Took, and D. P. Mandic. Adaptive convex combination approach for the identification of improper quaternion processes. *IEEE Transactions on Neural Networks and Learning Systems*, to be published, 2013.
- S. Umeyama. Contour extraction using a complex autoregressive model. Systems and Computers in Japan, 28(1):66–73, 1997.
- N. N. Vakhania. Random vectors with values in quaternion hilbert spaces. *Theory of Probability and its Applications*, 43(1):99–115, 1999.
- J. Vía and L. Vielva. Testing quaternion properness: generalized likelihood ratios and locally most powerful invariants. In 19th European Signal Processing Conference (EUSIPCO), Barcelona, Spain, 2011.
- J. Vía, D. Ramírez, and I. Santamaría. Properness and widely linear processing of quaternion random vectors. *IEEE Transactions on Signal Processing*, 56(7):3502– 3515, 2010a.
- J. Vía, D. Ramírez, I. Santamaría, and L. Vielva. Widely and semi-widely linear processing of quaternion vectors. In *IEEE International Conference on Acoustics* Speech and Signal Processing (ICASSP), pages 3946–3949, Dallas, Texas, USA, 2010b.
- J. Vía, D. Palomar, and L. Vielva. Generalized likelihood ratios for testing the properness of quaternion Gaussian vectors. *IEEE Transactions on Signal Processing*, 59 (4):1356–1370, 2011.
- A. T. Walden and P. T. G. Rubin-Delanchy. On testing for impropriety of complexvalued Gaussian vectors. *IEEE Transactions on Signal Processing*, 57(3):825–834, 2009.

- A. T. Walden and A. Serroukh. Wavelet analysis of matrix-valued time-series. Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 458(2017):157 –179, 2002.
- A. M. Walker. A note on the asymptotic distribution of sample quantiles. Journal of the Royal Statistical Society. Series B (Methodological), 30(3):570–575, 1968.
- M. A. Westenberg and T. Ertl. Denoising 2-D vector fields by vector wavelet thresholding. *Journal of WSCG*, 13, 2005.
- E. T. Whittaker and G. N. Watson. A Course of Modern Analysis. Cambridge University Press, 1927.
- J. Wishart. The generalised product moment distribution in samples from a normal multivariate population. *Biometrika*, 20A(1/2):32–52, 1928.
- G. G. Wojnar. Generalized Wishart models on convex homogeneous cones. PhD, Indiana University, 1999.
- A. T. A. Wood, J. G. Booth, and R. W. Butler. Saddlepoint approximations to the CDF of some statistics with nonnormal limit distributions. *Journal of the American Statistical Association*, 88(422):680–686, 1993.
- R. A. Wooding. The multivariate distribution of complex normal variables. Biometrika, 43(1/2):212–215, 1956.
- T. A. Wysocki, B. J. Wysocki, and S. Spence Adams. Correction to "The theory of quaternion orthogonal designs" [jan 08 256-265]. *IEEE Transactions on Signal Processing*, 57(8):3298, 2009.
- X.-G. Xia. Orthonormal matrix valued wavelets and matrix Karhunen-Loève expansion. In Wavelets, multiwavelets, and their applications, number 216 in Contemporary Mathematics, pages 159–175. American Mathematical Society, 1997.
- X.-G. Xia and B. Suter. Vector-valued wavelets and vector filter banks. *IEEE Trans*actions on Signal Processing, 44(3):508–518, 1996.

- X. Xu, Z. Guo, C. Song, and Y. Li. Multispectral palmprint recognition using a quaternion matrix. *Sensors*, 12(12):4633–4647, 2012.
- G. A. Young and R. L. Smith. Essentials of statistical inference: G.A. Young, R.L. Smith. Cambridge University Press, 2005.
- B. M. Yu. On existence of matrix-valued wavelets. Advanced Materials Research, 282-283:153–156, 2011.
- F. Zhang. Quaternions and matrices of quaternions. Linear Algebra and its Applications, 251:21–57, 1997.
- F. Zhang, J. Li, G. Chen, and J. Man. Assessment of color video quality based on quaternion singular value decomposition. In Sixth International Conference on Fuzzy Systems and Knowledge Discovery. FSKD '09, volume 4, pages 7–10, 2009.
- J. Zhao and L. Peng. Quaternion-valued admissible wavelets and orthogonal decomposition of L2(IG(2), H). Frontiers of Mathematics in China, 2(3):491–499, 2007.
- G. Zyskind. On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models. *The Annals of Mathematical Statistics*, 38(4):1092–1109, 1967.

Appendix A

Additional Results

A.1 A note on rotation invariance

Lemma A.1. Let \mathbb{A} be a unital real algebra, $n \geq 3$ and $\mathbf{M} \in \mathbb{A}^{n \times n}$. Then $\mathbf{M} = \mathbf{R}^T \mathbf{M} \mathbf{R}$ for all rotations $\mathbf{R} \in SO(n)$ of \mathbb{R}^n if and only if $\mathbf{M} = \lambda \mathbf{I}_n$ for some $\lambda \in \mathbb{A}$.

Proof. If $M = \lambda I_n$ then because $R \in SO(n) \subseteq \mathbb{R}^{n \times n} \subseteq \mathbb{A}^{n \times n}$, and A is a real algebra, λ commutes with R.

For the converse, fist assume n = 3 and consider the rotations

$$\boldsymbol{R} = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ and } \boldsymbol{R} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}.$$

These give us

$$\boldsymbol{M} = \begin{pmatrix} m_{2,2} & -m_{2,1} & -m_{2,3} \\ -m_{1,2} & m_{1,1} & m_{1,3} \\ -m_{3,2} & m_{3,1} & m_{3,3} \end{pmatrix} = \begin{pmatrix} m_{1,1} & -m_{1,3} & m_{1,2} \\ -m_{3,1} & m_{3,3} & -m_{3,2} \\ m_{2,1} & -m_{2,3} & m_{2,2} \end{pmatrix}.$$

In particular $m_{1,1} = m_{2,2} = m_{3,3}$ and $m_{1,2} = -m_{2,1} = -m_{1,3} = m_{2,3} = m_{1,3} = -m_{3,2}$ (and hence $m_{1,3} = 0$). This is our required result.

For n > 3 proceed by induction. The $(n-1) \times (n-1)$ submatrix obtained by

A.2 Additional results on random variables of Box type

deleting the first row and column is invariant under $SO(n-1) \cong \mathbf{I}_1 \oplus SO(n-1) \subset$ SO(n), and similarly for the submatrix obtained by deleting the second row and column or the last row and column. Hence these submatrices are of the form $\mathbf{I}_{n-1}\alpha$, $\mathbf{I}_{n-1}\beta$ and $\mathbf{I}_{n-1}\gamma$ respectively and $\alpha = \beta = \gamma = \lambda$.

Lemma A.1 applies in particular to block partitioned matrices, by taking $\mathbb{A} = \mathbb{R}^{m \times m}$. Consider two random vectors $\boldsymbol{u}, \boldsymbol{v}$ in \mathbb{R}^n . Their joint second-order properties are given by the covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{2n \times 2n}$ of $(u_1, v_1, \ldots, u_n, v_n)$. $\boldsymbol{\Sigma}$ can be partitioned into $n^2 \ 2 \times 2$ blocks. If $n \ge 3$ then the joint second-order properties will be invariant under rotations of \mathbb{R}^n if and only if $\boldsymbol{\Sigma}$ is block-diagonal, i.e. u_i and v_i are both uncorrelated with u_j and v_j for all $i \ne j$. In other words, second-order rotation invariance in dimensions n > 2 is equivalent to block sphericity, as opposed to the case n = 2 where it is equivalent to complex structure. Thus, rotation invariance in dimensions $n \ge 3$ implies lack of correlation.¹ In particular, a Gaussian signal taking values in \mathbb{R}^n , $n \ge 3$ cannot be rotation-invariant unless its components are independent.

A.2 Additional results on random variables of Box type

Proposition A.2. Let $\mathbb{E}[W^h]$ be given by (4.7). Then for $h \to \infty$

$$\mathbb{E}\left[W^{h}\right] = C_{7}h^{\frac{-f}{2}}(1+o(1)) \tag{A.1}$$

where C_7 is some positive constant.

Proof. Substitute Stirling's approximation

$$\Gamma(t+1) = \sqrt{2\pi h} \left(\frac{t}{e}\right)^t (1+o(1)) \tag{A.2}$$

¹Consider for example a time-series $\boldsymbol{x}(t) \in \mathbb{R}^n$. Then taking $\boldsymbol{u} = \boldsymbol{x}(t_1)$ and $\boldsymbol{v} = \boldsymbol{x}(t_2)$ shows that $x_i(t_1)$ is uncorrelated with $x_j(t_2)$ for $i \neq j$ for every t_1, t_2 .

A.2 Additional results on random variables of Box type

into equation (4.7), and simplify with (4.9). This yields (A.1) with

$$C_{7} = (2\pi)^{\frac{m-k}{2}} \prod_{i=1}^{m} \left[x_{i}^{\xi_{i}-\frac{1}{2}} \left(\Gamma \left(x_{i}+\xi_{i} \right) \right)^{-1} \right] \prod_{j=1}^{k} \left[y_{j}^{-\eta_{j}+\frac{1}{2}} \Gamma \left(y_{j}+\eta_{j} \right) \right].$$

Remark A.3. Proposition A.2 holds for complex h when $|h| \to \infty$, as long as $|\arg(h)| < \pi - \epsilon$ for some $\epsilon > 0$.

Corollary A.4. If the moments of W are given by (4.7) then $||W||_{\infty} = 1$. Proof.

$$||W||_{\infty} = \lim_{h \to \infty} \mathbb{E} \left[|W|^{h} \right]^{\frac{1}{h}}$$
$$= \lim_{h \to \infty} C_{7}^{\frac{1}{h}} \exp \left(\frac{-f}{2h} \log(h) \right)$$
$$= 1.$$

Remark A.5. $||W||_{\infty} = 1$ implies $W \le 1$ almost surely.

Lemma A.6. Let W be a random variable such that $0 \le W \le 1$. Then

$$\lim_{h\to\infty}\mathbb{E}\left[W^h\right]=\mathbb{P}(W=1)$$

Proof. Let $0 < \epsilon < 1$ be arbitrary, then

$$\begin{split} \mathbb{E}\left[W^{h}\right] &= \mathbb{E}\left[W^{h}\mathbbm{1}_{W \leq 1-\epsilon}\right] + \mathbb{E}\left[W^{h}\mathbbm{1}_{1-\epsilon < W}\right] \\ &\leq (1-\epsilon)^{h}\mathbb{P}\left(W \leq 1-\epsilon\right) + \mathbb{P}\left(1-\epsilon < W\right). \end{split}$$

Hence taking $h \to \infty$ and then $\epsilon \to 0^+$

$$\lim_{h \to \infty} \mathbb{E}\left[W^{h}\right] \leq \lim_{\epsilon \to 0^{+}} \mathbb{P}\left(1 - \epsilon < W\right) = \mathbb{P}\left(W = 1\right).$$

Also

$$\mathbb{E}\left[W^{h}\right] = \mathbb{E}\left[W^{h}\mathbb{1}_{W\neq 1}\right] + \mathbb{E}\left[W^{h}\mathbb{1}_{W=1}\right] \geq \mathbb{P}\left(W=1\right).$$

Hence

$$\lim_{h \to \infty} \mathbb{E}\left[W^h\right] = \mathbb{P}\left(W = 1\right).$$

Corollary A.7. f > 0, except for the degenerate case where f = 0 and W has a mass at 1.

Proof. When f < 0, (A.1) implies $\mathbb{P}(W = 1) = \lim_{h \to \infty} \mathbb{E}[W^h] = \infty > 1$, and when f = 0 it implies $\mathbb{P}(W = 1) = \lim_{h \to \infty} \mathbb{E}[W^h] = C_7 > 0$.

A.3 A note on the matrix Karhunen-Loève transform

Navarro-Moreno et al. (2012, 2013) consider the problems of estimating and testing for the presence of a (possibly random) continuous-time quaternion-valued signal measured with additive noise. No propriety assumptions are made and the approach is widely-linear. For the purpose of obtaining a Karhunen-Loève (KL) expansion (the continuous-time equivalent to principal component analysis), the quaternion problem is reduced to the scalar real case by concatenating the four quaternion components in time. Note that the approach can be viewed as a real vector generalisation of the scalar KL transform.

This is the same trick used previously by Xia (1997) for the more general matrix KL expansion. Note that as with the competing notions of matrix-valued and vector-valued multiresolution analysis discussed in section 5.3, the matrix KL expansion defined in Xia (1997) and its vector counterpart are largely equivalent. The equivalence is not immediately obvious because of three subtleties. Firstly, one needs to assume that the eigen-matrix-values found are diagonal. This can be done without loss of generality from Xia (1997, Theorem 5.1). Secondly, note that in an eigen-matrix-function with diagonal eigen-matrix-values, each row is an eigen-vector-function with corresponding scalar eigenvalue. Thirdly note that the covariance matrix used is the

A.3 A note on the matrix Karhunen-Loève transform

sum of the covariance matrices which would be obtained for each row of the signal taken independently. Because of this the matrix KL transform, as it is formulated, cannot be expressed as a parallel implementation of multiple vector KL transforms. It provides a single basis of vector-valued functions that can decorrelate each of the vector-valued signals given by the rows of the matrix-valued signal, wherease applying seperate transforms would in general produce a different basis for each row.

Another point worth noting is that although it is claimed that the matrix KL transform fully decorrelates a signal, this is based on a weak notion of orthogonality, so that the coefficients corresponding to one row of the matrix-valued signal may still be correlated to the coefficients corresponding to a different row, i.e. each of the vector-valued signals is decorrelated, but they are not *jointly* decorrelated.

Appendix B

Proofs

B.1 Proof of Theorem 2.15

Proof. We will adapt to the quaternion case the proof given in Goodman (1963) for the complex Wishart distribution. This relies on computing the characteristic function corresponding to (2.13) and comparing it with the characteristic function given in Proposition 2.13.

Consider the integral

$$c_k(\boldsymbol{\Sigma}^{-1}, \boldsymbol{\Theta}) = \int_{\text{QHPD}} |\boldsymbol{W}|^k \exp\left(-2\Re \operatorname{tr}\left(\boldsymbol{\Sigma}^{-1} \boldsymbol{W}\right) + \mathrm{i}\Re \operatorname{tr}\left(\boldsymbol{\Theta} \boldsymbol{W}\right)\right) \mathrm{d}\boldsymbol{W},$$

where we will be assuming that Σ is QHPD, Θ is Hermitian and k > -1. we integrate over the space of quaternion Hermitian positive definite matrices using the Lebesgue measure

$$\mathrm{d}\boldsymbol{W} = \prod_{i=1}^{p} \left[\mathrm{d}w_{i,i} \prod_{j=i+1}^{p} \mathrm{d}w_{i,j,1} \mathrm{d}w_{i,j,2} \mathrm{d}$$

Let D_1 be diagonal with positive real diagonal elements and D_i be real diagonal. Write $D = D_1 + iD_i$. Note that we will use the notation \bullet_1 , \bullet_i , \bullet_j , \bullet_k as an abbreviation to $\Re(\bullet)$, $\Im_i(\bullet)$, $\Im_j(\bullet)$ and $\Im_k(\bullet)$ respectively. i, j, k and i, j, k are not to be confused.

We will first consider $c_k \left(\frac{1}{2}\boldsymbol{D}_1, -\boldsymbol{D}_i\right)$.

Remark B.1. Note that Goodman (1963) only looks at real diagonal elements here where he should consider complex ones too.

Let $\mathbf{T} = (t_{i,j})_{i,j}$ be the upper triangular matrix with real positive elements on the diagonal that arises from the Cholesky decomposition of \mathbf{W} , i.e. $\mathbf{W} = \mathbf{T}^H \mathbf{T}$.

$$w_{i,j} = \sum_{k=1}^{i} \bar{t}_{k,i} t_{k,j} \ \forall i \le j$$
(B.2)

In particular

$$w_{i,i} = \sum_{k=1}^{i} |t_{k,i}|^2 \quad \forall i$$

The jacobian matrix between dW given in (B.1), and

$$d\boldsymbol{T} = \prod_{i=1}^{p} \left[dt_{i,i} \prod_{j=i+1}^{p} dt_{i,j,1} dt_{i,j,i} dt_{i,j,j} dt_{i,j,k} \right]$$
(B.3)

is lower triangular if we take that ordering (i.e. the ordering obtained by expanding the products in (B.1) and (B.3) without commuting.). Indeed, if $i \leq j$ then (B.2) shows that $w_{i,j}$ only depends on $t_{k,i}, t_{k,j}, k = 1, ..., i$ and the last term is

$$\bar{t}_{i,i}t_{i,j} = \begin{cases} t_{i,i}^2 & \text{if } i = j\\ t_{i,i}t_{i,j,1} + t_{i,i}t_{i,j,1}i + t_{i,i}t_{i,j,1}j + t_{i,i}t_{i,j,k}k & \text{if } i < j \end{cases}$$

(so $\Re(w_{i,j})$ doesn't depend on $\Im(t_{i,j})$ etc.)

$$\frac{\partial w_{i,i}}{\partial t_{i,i}} = 2t_{i,i} \;\forall i$$

$$\frac{\partial w_{i,i}}{\partial t_{i,i}} = 2t_{i,i} \;\forall i$$

$$\frac{\partial w_{i,j}}{\partial t_{i,j}} = t_{i,i} \mathbf{I}_4 \ \forall i < j$$

Hence the jacobian (determinant) is

$$\left|\frac{\partial \boldsymbol{W}}{\partial \boldsymbol{T}}\right| = 2^p \prod_{i=1}^p t_{i,i}^{1+4p-4i}.$$

(1 contributed from i = j and 4(p - i) from $i \neq j$.)

$$\begin{split} |\boldsymbol{W}|^{k} &= |\boldsymbol{T}|^{2k} = \prod_{i=1}^{p} t_{i,i}^{2k} \\ \operatorname{tr}\left(\boldsymbol{D}\boldsymbol{W}\right) &= \sum_{j=1}^{p} \sum_{i=1}^{j} d_{j,j} |t_{i,j}|^{2} \\ c_{k}\left(\frac{1}{2}\boldsymbol{D}_{1}, -\boldsymbol{D}_{i}\right) &= \int_{\mathrm{QHPD}} |\boldsymbol{W}|^{k} \exp\left(-\Re\operatorname{tr}\left(\boldsymbol{D}_{1}\boldsymbol{W}\right) - \mathrm{i}\Re\operatorname{tr}\left(\boldsymbol{D}_{i}\boldsymbol{W}\right)\right) \mathrm{d}\boldsymbol{W} \\ &= \int_{\mathrm{QHPD}} |\boldsymbol{W}|^{k} \exp\left(-\operatorname{tr}\left(\boldsymbol{D}\boldsymbol{W}\right)\right) \mathrm{d}\boldsymbol{W} \\ &= \int_{\mathrm{QHPD}} |\boldsymbol{W}|^{k} \exp\left(-\operatorname{tr}\left(\boldsymbol{D}\boldsymbol{W}\right)\right) \mathrm{d}\boldsymbol{W} \\ &= \int_{\mathrm{Triang}^{+}} 2^{p} \prod_{j=1}^{p} t_{j,j}^{1+2k+4p-4j} \exp\left(-\sum_{j=1}^{p} \sum_{i=1}^{j} d_{j,j} |t_{i,j}|^{2}\right) \mathrm{d}\boldsymbol{T} \\ &= \prod_{j=1}^{p} \left[L_{j,j} \prod_{i=1}^{j-1} L_{i,j} \right], \end{split}$$

where, using the Gamma pdf and characteristic function

$$L_{j,j} = \int_0^\infty t_{j,j}^{2k+4p-4j} e^{-d_{j,j}t_{j,j}^2} 2t_{j,j} dt_{j,j}$$

$$= \int_0^\infty u^{k+2p-2j} e^{-d_{j,j,1}u} e^{-id_{j,j,1}u} du$$

$$= \Gamma(1+k+2p-2j) d_{j,j,1}^{-1-k-2p+2j} (1+i\frac{d_{j,j,1}}{d_{j,j,1}})^{-1-k-2p+2j}$$

$$= \Gamma(1+k+2p-2j) d_{j,j}^{-1-k-2p+2j},$$

$$L_{i,j} = \int_{\mathbb{H}} e^{-d_{j,j}|t_{i,j}|^2} dt_{i,j}$$

= $\left(\int_{-\infty}^{\infty} e^{-d_{j,j}t_{i,j,1}^2} dt_{i,j,1}\right)^4$
= $\left(2\int_0^{\infty} e^{-d_{j,j,1}u} e^{-id_{j,j,1}u} \frac{1}{2}u^{-\frac{1}{2}} du\right)^4$
= $\left(\Gamma\left(\frac{1}{2}\right) d_{j,j,1}^{-\frac{1}{2}} \left(1 + i\frac{d_{j,j,1}}{d_{j,j,1}}\right)^{-\frac{1}{2}}\right)^4$
= $\pi^2 d_{j,j}^{-2}.$

The product gives

$$c_k \left(\frac{1}{2}\boldsymbol{D}_1, -\boldsymbol{D}_i\right) = \prod_{j=1}^p \left[\Gamma(1+k+2p-2j)d_{j,j}^{-1-k-2p+2j} \prod_{i=1}^{j-1} \pi^2 d_{j,j}^{-2} \right]$$
$$= \pi^{p(p-1)} \prod_{j=1}^p d_{j,j}^{1-k-2p} \Gamma(1+k+2p-2j)$$
$$= \pi^{p(p-1)} |\boldsymbol{D}|_{\mathbb{C}}^{1-k-2p} \prod_{i=1}^p \Gamma(-1+k+2i)$$
$$= c_k(\boldsymbol{I}, \boldsymbol{0}) \left| \frac{1}{2}\boldsymbol{D}_1 - \frac{\mathrm{i}}{2}\boldsymbol{D}_i \right|_{\mathbb{C}},$$

and in particular

$$c_k(\boldsymbol{D}_1, \boldsymbol{0}) = c_k(\boldsymbol{I}, \boldsymbol{0}) |\boldsymbol{D}_1|^{1-k-2p},$$

where

$$c_k(\mathbf{I}, \mathbf{0}) = \pi^{p(p-1)} 2^{p(1-k-2p)} \prod_{i=1}^p \Gamma(-1+k+2i).$$

We wish for some invertible \boldsymbol{M} to calculate the jacobian (determinant) $|J(\boldsymbol{M})| = \left|\frac{\partial g_{\boldsymbol{M}}(\boldsymbol{W})}{\partial \boldsymbol{W}}\right|$. By Lemma 1.61, the map $g_{\boldsymbol{M}}: \boldsymbol{W} \to \boldsymbol{M}^{H} \boldsymbol{W} \boldsymbol{M}$ is a bijection on the space of QHPD matrices. Because the map is linear, its jacobian is a function of \boldsymbol{M} only.

$$c_{k}\left(g_{M}(\boldsymbol{\Sigma})^{-1}, g_{M^{H^{-1}}}(\boldsymbol{\Theta})\right)$$

$$= c_{k}\left(g_{M^{H^{-1}}}(\boldsymbol{\Sigma}^{-1}), g_{M^{H^{-1}}}(\boldsymbol{\Theta})\right)$$

$$= \int_{\boldsymbol{W}\in\text{QHPD}} |\boldsymbol{W}|^{k} \exp\left(-2\Re \operatorname{tr}\left(\boldsymbol{M}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{M}^{H^{-1}}\boldsymbol{W}\right) + i\Re \operatorname{tr}\left(\boldsymbol{M}^{-1}\boldsymbol{\Theta}\boldsymbol{M}^{H^{-1}}\boldsymbol{W}\right)\right) d\boldsymbol{W}$$

$$= \int_{g_{M^{-1}}(\boldsymbol{W})\in\text{QHPD}} |\boldsymbol{M}^{H}g_{M^{-1}}(\boldsymbol{W})\boldsymbol{M}|^{k}$$

$$\cdot \exp\left(-2\Re \operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}g_{M^{-1}}(\boldsymbol{W})\right) i\Re \operatorname{tr}\left(\boldsymbol{\Theta}g_{M^{-1}}(\boldsymbol{W})\right)\right) J\left(\boldsymbol{M}^{-1}\right)^{-1} dg_{M^{-1}}(\boldsymbol{W})$$

$$= \int_{g_{M^{-1}}(\boldsymbol{W})\in\text{QHPD}} |\boldsymbol{M}^{H}\boldsymbol{M}|^{k} |g_{M^{-1}}(\boldsymbol{W})|^{k}$$

$$\cdot \exp\left(-2\Re \operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}g_{M^{-1}}(\boldsymbol{W})\right) i\Re \operatorname{tr}\left(\boldsymbol{\Theta}g_{M^{-1}}(\boldsymbol{W})\right)\right) J(\boldsymbol{M}) dg_{M^{-1}}(\boldsymbol{W})$$

$$= |J(\boldsymbol{M})| |\boldsymbol{M}^{H}\boldsymbol{M}|^{k} c_{k}\left(\boldsymbol{\Sigma}^{-1}, \boldsymbol{\Theta}\right)$$
(B.4)

By Theorem 1.64, there exist U unitary and D_1 diagonal with real positive elements s.t. $\Sigma = U^H D_1 U$. (B.4) gives

$$c_k\left(g_{\boldsymbol{U}}(\boldsymbol{I})^{-1},\boldsymbol{0}\right) = |J(\boldsymbol{U})| \left|\boldsymbol{U}^H \boldsymbol{U}\right|^k c_k\left(\boldsymbol{I},\boldsymbol{0}\right) = |J(\boldsymbol{U})| c_k\left(g_{\boldsymbol{U}}(\boldsymbol{I})^{-1},\boldsymbol{0}\right)$$

Hence J(U) = 1. Hence

$$c_{k} \left(\boldsymbol{\Sigma}^{-1}, \boldsymbol{0} \right) = c_{k} \left(g_{\boldsymbol{U}}(\boldsymbol{D}_{1})^{-1}, \boldsymbol{0} \right)$$

$$= c_{k} \left(\boldsymbol{D}_{1}^{-1}, \boldsymbol{0} \right)$$

$$= \pi^{p(p-1)} 2^{p(1-k-2p)} |\boldsymbol{D}_{1}|^{-1+k+2p} \prod_{i=1}^{p} \Gamma(-1+k+2i)$$

$$= c_{k} \left(\boldsymbol{I}, \boldsymbol{0} \right) |\boldsymbol{\Sigma}|^{-1+k+2p}$$
(B.5)

Using Lemma 1.67 (slightly modified), there is an invertible M and a diagonal matrix D_{i} with real diagonal entries s.t. $\Sigma^{-1} = M^{H}M$ and $\frac{1}{2}\Theta = M^{H}D_{i}M$. (By

using these equations to define Σ and Θ , M can be made arbitrary.) This gives

$$c_k \left(\boldsymbol{\Sigma}^{-1}, \boldsymbol{0} \right) = c_k \left(g_{\boldsymbol{M}^{H^{-1}}}(\boldsymbol{I})^{-1}, \boldsymbol{0} \right)$$

= $\left| J \left(\boldsymbol{M}^{H^{-1}} \right) \right| \left| \boldsymbol{\Sigma} \right|^k c_k(\boldsymbol{I}, \boldsymbol{0}).$

Comparing with (B.5) gives

$$ig|J\left(oldsymbol{M}^{H^{-1}}
ight)ig| = |oldsymbol{\Sigma}|^{2p-1} = |oldsymbol{M}|^{2-4p}$$
 $|J(oldsymbol{M})| = |oldsymbol{M}|^{4p-2}$

(This jacobian calculation is general for \boldsymbol{M} invertible)

$$c_{k} \left(\Sigma^{-1}, \Theta \right) = c_{k} \left(g_{M}(I), g_{M}(D_{i}) \right)$$

$$= \left| J(M^{H^{-1}}) \right| \left| M^{-1}M^{H^{-1}} \right|^{k} c_{k}(I, D_{i})$$

$$= \left| M^{H}M \right|^{1-k-2p} \left| I - \frac{i}{2}D_{i} \right|_{\mathbb{C}}^{1-k-2p} c_{k}(I, 0)$$

$$= \left| \widetilde{M}^{T}\widetilde{M} \right|_{\mathbb{C}}^{\frac{1-k-2p}{4}} \left| \widetilde{I} - \frac{i}{2}\widetilde{D}_{i} \right|_{\mathbb{C}}^{\frac{1-k-2p}{4}} c_{k}(I, 0)$$

$$= \left| \widetilde{M}^{T}\widetilde{M} - \frac{i}{2}\widetilde{M}^{T}\widetilde{D}_{i}\widetilde{M} \right|_{\mathbb{C}}^{\frac{1-k-2p}{4}} c_{k}(I, 0)$$

$$= \left| \widetilde{\Sigma}^{-1} - \frac{i}{2}\widetilde{\Theta} \right|_{\mathbb{C}}^{\frac{1-k-2p}{4}} c_{k}(I, 0)$$

Consider a random QHPD matrix \boldsymbol{W} with density

$$(c_k(\boldsymbol{\Sigma}^{-1}, \mathbf{0}))^{-1} |\boldsymbol{W}|^k \exp(-2\Re \operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{W})).$$
 (B.6)

B.2 Proof of Remark 3.2

Its characteristic function is then given by

$$\frac{c_k \left(\boldsymbol{\Sigma}^{-1}, \boldsymbol{\Theta}\right)}{c_k \left(\boldsymbol{\Sigma}^{-1}, \boldsymbol{0}\right)} = \left|\boldsymbol{\Sigma}\right|^{1-k-2p} \left| \widetilde{\boldsymbol{\Sigma}}^{-1} - \frac{\mathrm{i}}{2} \widetilde{\boldsymbol{\Theta}} \right|_{\mathbb{C}}^{\frac{1-k-2p}{4}} = \left| \widetilde{\boldsymbol{I}} - \frac{\mathrm{i}}{2} \widetilde{\boldsymbol{\Sigma}} \widetilde{\boldsymbol{\Theta}} \right|_{\mathbb{C}}^{\frac{1-k-2p}{4}}.$$
(B.7)

If we choose k = 1 - 2p + 2N then N > p - 1 implies k > -1, (B.7)=(2.12) and (B.6)=(2.13) completing the proof.

B.2 Proof of Remark 3.2

Proof. It is well known that (for constant X), $\hat{\beta}^{\text{GLS}}$ is the best linear unbiased estimator (BLUE). Now from Zyskind (1967, Theorem 2) either of the following two conditions (quoted verbatim) are necessary and sufficient for the simple linear LS estimator $\hat{\beta}^{\text{LS}}$ to be the BLUE estimator $\hat{\beta}^{\text{GLS}}$.

- 1. A matrix S exists satisfying the relation $\Sigma_e X = XS$, and further, for Σ_e non-singular, a matrix R exists satisfying $\Sigma_e^{-1}X = XR$.
- 2. A matrix R exists such that $\Sigma_e^+ X = XR$. (Note that when Σ_e is non-singular $\Sigma_e^+ = \Sigma_e^{-1}$).

Here \bullet^+ indicating the Moore-Penrose generalized inverse. However from Zyskind (1967, p. 1099), $\Sigma_e X = XS$ for some S if and only if $\Sigma_e^+ X = XR$ for some R. Putting this together with 2. we see that 1. can be reduced to the simple requirement that a matrix S exists satisfying the relation $\Sigma_e X = XS$.

B.3 Proof of Proposition 4.8

B.3 Proof of Proposition 4.8

Proof. Test 1 corresponds to test (a) in Andersson et al. (1983). Changing p to 2p in Andersson et al. (1983, eqn. 101) gives

$$E\{T_{\mathbb{C}}^{h}\} = K_{0} \prod_{j=1}^{2p} \frac{\Gamma\left(\left[N(h+1) - 2p - j + 1\right]/2\right)}{\Gamma\left(\left[N(h+1) - j + 2\right]/2\right)},$$

where K_0 does not depend on h. Applying (4.13) we can write

$$\Gamma\left[\frac{[N(h+1)-2p-2k+1]}{2}\right] \Gamma\left[\frac{[N(h+1)-2p-2k+2]}{2}\right]$$
$$= \sqrt{\pi} 2^{2p+2k-N(h+1)} \Gamma\left[N(h+1)-2p-2k+1\right]$$

and

$$\Gamma\left[\frac{[N(h+1)-2k+2]}{2}\right] \Gamma\left[\frac{[N(h+1)-2k+3]}{2}\right] = \sqrt{\pi} \ 2^{2k-1-N(h+1)} \Gamma\left[N(h+1)-2k+2\right].$$

Hence

$$E\{T_{\mathbb{C}}^{h}\} = K_{1} \prod_{j=1}^{p} \frac{\Gamma[N(h+1) - 2p - 2j + 1]}{\Gamma[N(h+1) - 2j + 2]},$$

while, from Andersson et al. (1983, eqn. 103)

$$E\{T_{\mathbb{H}}^{h}\} = K_{2} \prod_{j=1}^{p} \frac{\Gamma[N(h+1) - p - j + 1]}{\Gamma[N(h+1) + \frac{3}{2} - j]},$$

where K_1 and K_2 do not depend on h. So $E\{T^h\} = E\{T^h_{\mathbb{C}}\}E\{T^h_{\mathbb{H}}\}$ takes the form

$$K\prod_{j=1}^{p} \frac{\Gamma\left[N(h+1) - 2p - 2j + 1\right]\Gamma\left[N(h+1) - p - j + 1\right]}{\Gamma\left[N(h+1) - 2j + 2\right]\Gamma\left[N(h+1) - j + \frac{3}{2}\right]}.$$

This expression can be simplified by canceling terms and reordering. Even and odd p can be considered separately. $\lfloor p/2 \rfloor$ of the $\Gamma[N(h+1) - 2j + 2]$ terms cancel with

 $\Gamma[N(h+1) - p - j + 1]$ terms. The remaining numerator terms of the product can be juxtaposed and the remaining denominator terms interspersed to form a monotone pattern. Finally by inverting the order of the product in the numerator we obtain (4.6).

B.4 Proof of Theorem 4.33

Proof. What follows is a slight rewording of the derivation by Box (1949). First the cumulant generating function of ρM is expanded along the imaginary axis using Lemma 4.32.

$$\log \phi_{M}(\rho ti) = \sum_{i=1}^{k} \log \Gamma (y_{r} + \eta_{r}) - \sum_{i=1}^{m} \log \Gamma (x_{i} + \xi_{i}) \\ + 2it\rho \left(\sum_{i=1}^{m} x_{i} \log(x_{i}) - \sum_{j=1}^{k} y_{j} \log(y_{j}) \right) \\ + \sum_{i=1}^{m} \log \Gamma (\rho x_{i}(1 - 2ti) + (1 - \rho)x_{i} + \xi_{i}) \\ - \sum_{i=1}^{k} \log \Gamma (\rho y_{r}(1 - 2ti) + (1 - \rho)y_{r} + \eta_{r}) \\ = \log(K_{B}) - \frac{f}{2} \log(1 - 2ti) \\ + \sum_{j=1}^{n} \omega_{j}(1 - 2ti)^{-j} + R_{n+1}^{***}(t)$$

where

$$\log(K_B) = \log(K) + \frac{1}{2}(m-k)\log(2\pi) - \frac{f}{2}\log(\rho) + \sum_{i=1}^{m} \left(x_i + \xi_i - \frac{1}{2}\right)\log(x_i) - \sum_{i=1}^{k} \left(y_i + \eta_i - \frac{1}{2}\right)\log(y_i) = -\sum_{j=1}^{n} \omega_j + R_{n+1}^{***}(0) , \qquad (B.8)$$

B.5 Proof of Proposition 5.3

$$\omega_j = \frac{(-1)^{j+1}}{j(j+1)} \left[\sum_{i=1}^m \frac{\mathbf{B}_{j+1} \left((1-\rho)x_i + \xi_i \right)}{\left(\rho x_i\right)^j} - \sum_{i=1}^k \frac{\mathbf{B}_{j+1} \left((1-\rho)y_i + \eta_i \right)}{\left(\rho y_i\right)^j} \right]$$
(B.9)

and $R_{n+1}^{***}(t) = O\left(\left(x_0\sqrt{1+4t^2}\right)^{-n-1}\right)$ with $x_0 = \min_{i,j}(x_i, y_j)$. This leads to an asymptotic expansion of the characteristic function

$$\phi_M(\rho ti) = K_B \sum_{j=0}^n a_j (1 - 2ti)^{-\frac{f+2j}{2}} + R_{n+1}^{****}(t)$$
$$= K_B \sum_{j=0}^n a_j \phi_{\chi_{f+2j}^2}(t) + R_{n+1}^{****}(t),$$

where a_j is the coefficient of t^j in the series expansion of $\exp\left(\sum_{j=1}^n \omega_j t^j\right)$, and $R_{n+1}^{****}(t) = O\left(\left(x_0\sqrt{1+4t^2}\right)^{-n-1}\right)$.

The asymptotic series for the pdf and CDF can now be obtained through termby-term integration of the characteristic function. $\hfill \square$

B.5 Proof of Proposition 5.3

Proof. Let $\mathbf{F}^{(i,\bullet)} \in V^{(i,\bullet)}$. Let $\mathbf{M} \in \mathbb{R}^{m \times m}$ have (j,i)-entry equal to 1 and all other entries 0. Then $\mathbf{F}^{(i,\bullet)} = (\mathbf{M}\mathbf{F})^{(j,\bullet)} \in V^{(j,\bullet)}$. Hence $V^{(i,\bullet)} = V^{(j,\bullet)} = S \forall i, j$ and $V \subseteq S^m$. Let $\mathbf{F} \in S^m$. For $i = 1, \ldots, m$, choose $\mathbf{F}_i \in V$ such that $\mathbf{F}_i^{(i,\bullet)} = \mathbf{F}^{(i,\bullet)} \in$ $S = V^{(i,\bullet)}$ and let $\mathbf{M}_i \in \mathbb{R}^{m \times m}$ have (i,i)-entry equal to 1 and all other entries 0. Then $\mathbf{F} = \sum_{i=1}^m \mathbf{M}_i \mathbf{F}_i \in V$, and so $S^m \subseteq V$. Hence $V = S^m$. Linearity of $S = V^{(1,\bullet)}$ follows directly from that of V. For the converse, note that for any $\mathbf{M} \in \mathbb{R}^{m \times m}$ and $\mathbf{F} \in S^m$, $(\mathbf{M}\mathbf{F})^{(i,\bullet)}$ is a linear combination of the rows of \mathbf{F} .

B.6 Proof of Proposition 5.18

We will prove the matrix-valued version of Proposition 5.18, given below. Proposition 5.18 follows by setting $S_{j,k}^{(i,\bullet)} = \delta_{i,1} s_{j,k}$ and $W_{j,k}^{(i,\bullet)} = \delta_{i,1} w_{j,k}$. Note that in this proof the matrix subscripts $\bullet_{j,k}$ do not indicate matrix (j, k) entries.

B.6 Proof of Proposition 5.18

Proposition B.2. Given a signal

$$\boldsymbol{F}(t) = \sum_{k \in \mathbb{Z}} \boldsymbol{S}_{0,k} \boldsymbol{\Phi}(t-k) \in V_0 \subset L^2(\mathbb{R}, \mathbb{R}^{n \times n}),$$

the coefficients $\boldsymbol{S}_{J,k}$ and $\boldsymbol{W}_{j,k}$ in the decomposition

$$\boldsymbol{F}(t) = \sum_{k \in \mathbb{Z}} \boldsymbol{S}_{J,k} 2^{-J/2} \boldsymbol{\Phi}(2^{-J}t - k) + \sum_{k \in \mathbb{Z}} \sum_{j=0}^{J} \boldsymbol{W}_{j,k} 2^{-\frac{j}{2}} \boldsymbol{\Psi}(2^{-j}t - k)$$
(B.10)

can be obtained through the fast wavelet transform (a.k.a. Mallat's pyramid algorithm) by iteratively computing

$$egin{array}{rcl} m{S}_{j+1,k} &=& \sum_{\ell=2k}^{2k+L-1} m{S}_{j,\ell} m{G}_{\ell-2k}^T \ m{W}_{j+1,k} &=& \sum_{\ell=2k}^{2k+L-1} m{W}_{j,\ell} m{H}_{\ell-2k}^T. \end{array}$$

The original signal $S_{0,k}$ can then be recovered through the reconstruction algorithm which iteratively computes

$$oldsymbol{S}_{j-1,k} = \sum_{\ell \in \mathbb{Z}} \left(oldsymbol{S}_{j,\ell} oldsymbol{G}_{k-2\ell} + oldsymbol{W}_{j,\ell} oldsymbol{H}_{k-2\ell}
ight).$$

Proof. By (5.5)

$$S_{j+1,k} = \left\langle F(t), 2^{-\frac{j+1}{2}} \Phi(2^{-j-1}t-k) \right\rangle_{n \times n}$$
$$= \left\langle F(t), \sum_{m \in \mathbb{Z}} G_m 2^{-\frac{j}{2}} \Phi(2^{-j}t-2k-m) \right\rangle_{n \times n}$$
$$= \sum_{m \in \mathbb{Z}} \left\langle F(t), 2^{-\frac{j}{2}} \Phi(2^{-j}t-2k-m) \right\rangle_{n \times n} G_m^T$$
$$= \sum_{m \in \mathbb{Z}} S_{j,2k+m} G_m^T.$$

Finally, note that $G_m = \mathbf{0}_{n \times n}$ except for $0 \le m \le L-1$ and set $\ell = m+2k$. Similarly,

B.6 Proof of Proposition 5.18

by (5.6) one obtains $W_{j+1,k}$.

For the reconstruction, first note that $\Psi(2^{-j}t - k) \in V_{j-1} \ominus V_j$ and hence is orthogonal to all functions in V_i , $i \ge j$. Thus $\Psi(2^{-i}t - k)$ is orthogonal to all functions in V_{j-1} for all i < j. From (B.10)

$$\begin{split} S_{j-1,k} &= \left\langle F(t), 2^{-\frac{j-1}{2}} \Phi(2^{-j+1}t-k) \right\rangle_{n \times n} \\ &= \sum_{\ell \in \mathbb{Z}} S_{j,\ell} \left\langle 2^{-\frac{j}{2}} \Phi(2^{-j}t-\ell), 2^{-\frac{j-1}{2}} \Phi(2^{-j+1}t-k) \right\rangle_{n \times n} \\ &+ \sum_{\ell \in \mathbb{Z}} \sum_{i=0}^{j} W_{i,\ell} \left\langle 2^{-\frac{i}{2}} \Psi(2^{-i}t-\ell), 2^{-\frac{j-1}{2}} \Phi(2^{-j+1}t-k) \right\rangle_{n \times n} \\ &= \sum_{\ell \in \mathbb{Z}} \left(S_{j,\ell} \left\langle 2^{-\frac{j}{2}} \Phi(2^{-j}t-\ell), 2^{-\frac{j-1}{2}} \Phi(2^{-j+1}t-k) \right\rangle_{n \times n} \right. \\ &+ W_{j,\ell} \left\langle 2^{-\frac{j}{2}} \Psi(2^{-j}t-\ell), 2^{-\frac{j-1}{2}} \Phi(2^{-j+1}t-k) \right\rangle_{n \times n} \right). \end{split}$$

Now, using (5.5)

$$\begin{split} &\left\langle 2^{-\frac{j}{2}} \mathbf{\Phi}(2^{-j}t-\ell), 2^{-\frac{j-1}{2}} \mathbf{\Phi}(2^{-j+1}t-k) \right\rangle_{n \times n} \\ &= \sum_{m \in \mathbb{Z}} \mathbf{G}_m \left\langle 2^{-\frac{j-1}{2}} \mathbf{\Phi}(2^{-j+1}t-2\ell-m), 2^{-\frac{j-1}{2}} \mathbf{\Phi}(2^{-j+1}t-k) \right\rangle_{n \times n} \\ &= \sum_{m \in \mathbb{Z}} \mathbf{G}_\ell \delta_{2\ell+m,k} \\ &= \mathbf{G}_{k-2\ell}. \end{split}$$

Similarly, from (5.6) we deduce

$$\left\langle 2^{-\frac{j}{2}} \Psi(2^{-j}t-\ell), 2^{-\frac{j-1}{2}} \Phi(2^{-j+1}t-k) \right\rangle_{n \times n} = H_{k-2\ell}.$$

Hence substituting the above we obtain

$$oldsymbol{S}_{j-1,k} = \sum_{\ell \in \mathbb{Z}} \left(oldsymbol{S}_{j,\ell} oldsymbol{G}_{k-2\ell} + oldsymbol{W}_{j,\ell} oldsymbol{H}_{k-2\ell}
ight).$$

Appendix C

Computer Code

C.1 Matlab code for wavelet filter computation

function [G,H] = G2GH(G)

```
\ INPUT: A matrix-valued scaling filter G (as a n*n*L array)
```

```
% OUTPUT: A matrix-valued scaling filter G
```

```
% and a corresponding matrix-valued wavelet filter H
```

```
% (as n*n*L(+1) arrays)
```

[G,H]=polyphase2GH(G2polyphase(G)); end

function P = G2polyphase(G) % OUTPUT: A 2n*2n polyphase matrix P of class mpoly % INPUT: A n*n*L array G representing a matrix-valued scaling filter

```
[n,m,L]=size(G);
if mod(L,2)==1
    G(:,:,L+1)=zeros(n);
    L=L+1;
```

end

* Write the n*2n top half of the polyphase matrix as a mpoly object P1

C.1 Matlab code for wavelet filter computation

```
Pg=zeros(n, 2*n, L/2);
if isa(G, 'sym')
    Pg=sym(Pg);
end
for k=1:L/2
   Pg(:,:,k) = [G(:,:,2*k-1),G(:,:,2*k)];
end
P1=mpoly(Pg,0,'polyphase',2,2);
% Compute the projection factorisation of the polyphase matrix
% using the mw package by Fritz Keinert.
% http://orion.math.iastate.edu/keinert/book.html
F=projection_factorization(P1);
* Compute the full 2n*2n polyphase matrix from the factorisation by
% unitary completion of the constant coefficient.
lf=length(F);
P=[eye(n), eye(n); -eye(n), eye(n)]/sqrt(2);
for k=2:1f
   P=P*F\{k\};
end
end
```

function [G,H] = polyphase2GH(P)

```
% INPUT: A 2n*2n polyphase matrix P of class mpoly
% OUTPUT: A n*n matrix-valued scaling filter G and wavelet filter H
% as n*n*L(+1) arrays
```

```
P=P.coef;
```

```
L=2*size(P,3);
n=size(P,1)/2;
G=zeros(n,n,L);
if isa(P,'sym')
    G=sym(G);
end
H=G;
for k=1:L/2
    G(:,:,2*k-1)=P(1:n,1:n,k);
```

```
G(:,:,2*k)=P(1:n,n+1:2*n,k);
H(:,:,2*k-1)=P(n+1:2*n,1:n,k);
H(:,:,2*k)=P(n+1:2*n,n+1:2*n,k);
end
end
```

C.2 Maple code for the design of scaling filters

```
with(LinearAlgebra) :
         #Parameters
        n := 4 : #Dimensions
        L := 10 : #Filter Length
       Nvm := 0: #Number of vanishing moments (incorrect weak definition)
        Numplus := 5 : #Number of vanishing moments (correct strong definition)
       cstr := false : #Impose complex structure?
       qstr := true : #Impose quaternion structure?
        sym := false : #Impose symmetry?
 > #Preliminaries
  if type(L, odd) then isodd := 1 : L := L + 1 : else isodd := 0 : end if:
   > Inn := IdentityMatrix(n, n):
   > Onn := Matrix(1..n, 1..n, 0):
   > ONES := Matrix(1..n, 1..n, 1):
> g := \operatorname{Array}(1..n, 1..n, 1..L, (i, j, k) \rightarrow gg_{i, j, k}):
       if isodd = 1 then for kmi from 1 to n do for kmj from 1 to n do g := subs(g[kmi, kmj, L] = 0, g)
                      end do: end do: end if: #Set last entry to 0 if length is odd
         #Impose the diagonal+antisymmetric condition on the first coefficient, or complex structure, or
                     quaternion structure,
        if not cstr and not qstr then for kmi from 1 to n - 1 do for kmj from kmi + 1 to n do g
                        := subs( g[kmi, kmj, 1] = -g[kmj, kmi, 1], g) : end do: end do: end if:
        if cstr then for kk from 1 to L do g := subs(\{g[1, 2, kk] = -g[2, 1, kk], g[2, 2, kk\} = g[1, 1, kk], g[2, 2, kk] = g[1, 1, kk], g[2, 2, kk], g[2, k
                     kk]}, g) : end do: end if:
> if qstr then for kk from 1 to L do g :=
         subs(\{g[1, 2, kk] = -g[2, 1, kk], g[1, 3, kk] = -g[3, 1, kk], g[1, 4, kk] = -g[4, 1, kk], g[1, 2, kk], g[1, 2, kk], g[1, 2, kk], g[1, 2, kk]
          g[2, 2, kk] = g[1, 1, kk], g[2, 3, kk] = -g[4, 1, kk], g[2, 4, kk] = g[3, 1, kk],
          g[3, 2, kk] = g[4, 1, kk], g[3, 3, kk] = g[1, 1, kk], g[3, 4, kk] = -g[2, 1, kk],
          g[4, 2, kk] = -g[3, 1, kk], g[4, 3, kk] = g[2, 1, kk], g[4, 4, kk] = g[1, 1, kk], g[4, 2, kk] = g[1, 1, kk]
             end do:
          g := subs(\{g[3, 1, 1] = 0, g[4, 1, 1] = 0, g[3, 1, 2] = 0\}, g):
         end if:
> if sym then for kk from 1 to \frac{(L-isodd)}{2} do for kmi from 1 to n do for kmj from 1 to n do g
                        := subs(g[kmi, kmj, L - kk + 1 - isodd] = g[kmi, kmj, kk], g) : end do: end do: end do:
                      end if:
         #orthogonality equations
        ort := []: for kk from 0 by 2 to L - 2 do ort := [op(ort), add(Matrix(g[1...n, 1...n, k]))]
                      .Matrix(g[1..n, 1..n, k+kk])^+, k=1..L-kk)]: end do:
         #vanishing moments equations
         vm := []: for kk from Nvmplus + 1 to Nvm do vm := [op(vm), add(Matrix(g[1..n, 1..n, k]) .ONES \cdot (1-k)^{kk-1}, k=1..L)]: end do:
```

for kk from 1 to Nymplus do $vm := [op(vm), add(Matrix(g[1..n, 1..n, k]) \cdot (k-1)^{kk-1} \cdot (k-1)^{kk-1})$ $(-1)^k, k = 1...L$: end do: #scaling equation sca := add(Matrix(g[1..n, 1..n, k]), k=1..L): $constr := \{ \}$: # Additional ad-hoc constraints, given as a set of polynomial equations in the coefficients which are set to 0 #generate list of polynomial equations $equify := (s, r) \rightarrow \{seq(s(i) - r(i), i = 1..n^2)\}:$ twoorone := 2 :#Set to 1 for standard scaling, set to 2 so that the scaling is rational $eqset := \left(equify(ort[1], twoorone \cdot Inn) \text{ union `} (seq\left(equify(ort[k], Onn), k = 2 \dots \frac{L}{2} \right) \right)$ **union** union (seq(equify(vm[k], Onn), k=1..nops(vm))) **union** equify(sca, sqrt(twoorone)) $\cdot 2$) $\cdot Inn$) **union** constr) $\setminus \{0\}$: varlist := [op(indets(g) union indets(constr))]:eqlist := [op(eqset)]:nops(varlist); #Number of unknowns 37 (1) nops(eqset); #Number of equations 71 (2) # Pre-solve the linear equations thus reducing the number of variables in the polynomial equations islinear := map(evalb, simplify(map(degree, eqlist) = 1)): *lineareqset* $:= \{ \}$: for k from 1 to nops(eqlist) do if islinear[k] then lineareqset $:= \{op(lineareqset), eqlist[k]\} : end if: end do:$ $polyeqset := eqset \setminus linear eqset$: nops(lineareqset); # Number of linear equations (3) 42 nops(polyeqset); # Number of polynomial equations (4) 29 *linsol* := *solve*(*lineareqset*) : # *Solve the linear equations* polyeqlist := [op(simplify(eval(polyeqset, linsol)))]:# Substitute the solutions into the polynomial equations. # Use the simplified polynomial equations as our list of equations. *fulleqlist* := *eqlist* : eqlist := polyeqlist: varset := indets(eqlist): varlist := [op(varset)]:nops(varlist); # Number of remaining unknowns 13 (5) #Manipulate the system of polynomial equations Groebner:-IsProper(eqlist); #Are there any solutions? (6) true

C.2 Maple code for the design of scaling filters

Groebner:-HilbertDimension(eqlist); #How many free parameters will there be in the solutions? (7) varord := Groebner:-SuggestVariableOrder(eqlist, varlist) : #Choose an ordering of variables #Compute the lexicographic Groebner basis for the given set of equations and variable ordering GB := Groebner:-Basis(eqlist, plex(varord)):#GB2:=Groebner:-Basis(eqlist, tdeg(varord)):#Non-lexicographic Groebner Basis. Will usually be much faster to compute. *#Tests for special properties* $sumoffdiag := add(add(add(g[i, j, k]^2 + g[j, i, k]^2, i = j + 1 ...n), j = 1 ...n), k = 1 ...L):$ #=0 iff g is diagonal > sumoffblockdiag := $add\left(add\left(g[i, j, k]^2 + g[j, i, k]^2, i = j + \operatorname{ceil}\left(\frac{n}{2}\right)..n\right), j = 1..n\right), k = 1$..L : > symmetrycond := $add \left(add \left(add \left((g[i,j,k] - g[i,j,L+1 - isodd - k])^2, k = 1 ... \frac{L}{2} - isodd \right), i = 1 ... \frac{L}{2} - isodd \right)$ = 1 ... n, j = 1 ... n; # = 0 iff g is symmetric > # We can the Groebner Basis to check whether certain polynomials belong to the ideal (which would imply that all solutions are roots). Note that this is a sufficient but not necessary condition for all solutions to be roots. evalb(Groebner:-NormalForm(eval(sumoffdiag, linsol), GB, plex(varord)) = 0); #Are all solutions diagonal? (8) false evalb(Groebner:-NormalForm(eval(sumoffblockdiag, linsol), GB, plex(varord)) = 0); #Are all solutions block diagonal with two evenly sized blocks (and hence trivial)? false (9) evalb(Groebner:-NormalForm(eval(symmetrycond, linsol), GB, plex(varord)) = 0);#Are all solutions symmetric? false (10) #Solve the pre-processed system of equations #sollist:=SolveTools:-PolynomialSystem(eqset) : #groebner2:=Groebner:-Solve(eqlist, varlist) : # Either of these lines should in theory solve the system of equations with our pre-processing performed implicitly, but fail. $freevars := \{ \} : #Set free variables a-priori, e.g. freevars := \{g[1, 2, 1]\}$ # Solve the system of polynomial equations given by the Groebner basis. $sollist := RealDomain :- solve(\{op(GB)\}, varset \land freevars, DropMultiplicity = true, explicit \}$ = true, all solutions = true) : **use** *RealDomain* **in** *sollist* := *op*(*simplify*(*expand*([*sollist*]))) : **end use**: #*Simplify solutions* $Nsols := nops(\{sollist\});$ #Number of solutions. Note that many of these may be orthogonally similar and or reparameterisations of one-another, and that each of these may contain free parameters.

/11\

Nsols := 20(11) > renorm := 1
sqrt(twoorone) :
#Undo the effect of setting twoorone to 2, so that coefficients have energy 1 > solve $(16384 \cdot gg_{1, 1, 8}^2 + 1792 \cdot gg_{1, 1, 8} - 231)$ $-\frac{7}{128} + \frac{1}{64}\sqrt{70}, -\frac{7}{128} - \frac{1}{64}\sqrt{70}$ (12) > sqrt(4096) (13) *Nsolsprint* := *Nsols* :#*Number of solutions to print* # Print the solutions > for ns from 1 to Nsolsprint do # Note that this for loop is a single execution group sol := sollist[ns]:> $gs := simplify(eval(eval(g, linsol), sol) \cdot renorm) :$ > isdiag := simplify(eval(eval(sumoffdiag, linsol), sol)) : #Check whether the solution is diagonal (and hence trivial) isblockdiag := simplify(eval(eval(sumoffblockdiag, linsol), sol)):*issymmetric* := *simplify*(*eval*(*eval*(*symmetrycond*, *linsol*), *sol*)) : # Make ad - hoc simplifications or substitutions to improve legibility of output assume(Y2 > 0): $gs := simplify \left(subs \left(\sqrt{-16 \ YI \ Y2^2 + 120 \ Y2^2 + 700 \ YI - 5950} = \frac{1}{\sqrt{2}} \cdot Y2 \cdot Y3, simplify \left(algsubs \left(70 - YI^2 = Y2^2, simplify \left(algsubs \left(-64 \cdot \left(\frac{7}{128} + g[1, 1, 8] \right) = YI, 1 + 128 \right) \right) \right) \right)$ gs)))))) : #For Maple v17 $#gs := simplify \left(subs \left(\sqrt{1050 \ Yl - 24 \ Yl \ Y2^2 - 8925 + 180 \ Y2^2} = \frac{\sqrt{3}}{2} \cdot |Y2| \cdot Y3 \right) \right)$ $simplify(algsubs(70 - Yl^2 = Y2^2, algsubs(-105 + 128 gg_{1, 1, 6} = 3 \cdot Yl, gs))))))$: #For Maple v15 use this substitution instead #gs := evalf(gs) : # To print all solutions in approximate decimal form if nops(indets(gs)) = 0 then gs := evalf(gs) : end if: # To do so only in the absence of free variables print("#", ns) : if isdiag = 0 then print("DIAGONAL"); # Do not print diagonal solutions elif isblockdiag = 0 then print("BLOCK DIAGONAL"); # Or block diagonal solutions (this line can be commented out) else if issymmetric = 0 then print("SYMMETRIC"); Lprint := $\frac{L}{2}$: else Lprint := L : end if: # Only print first half for symmetric solutions for k from 1 to Lprint do if qstr or cstr then print((Matrix(gs[1..n, 1, k]))); else print(Matrix(gs[1..n, 1..n, k])); end if; end do; # Only print first column for complex\quaternion structured solutions end if: end do: "#", 1
```
"SYMMETRIC"
           0. 0.02310968665 0. 0.
     -0.02762135863 0. 0. 0.02310968665
-0.03866990209 -0.1617678066 0. 0.06932905996
  0.1933495104 -0.1155484333 0. 0.02310968665
  0.5800485313 0.2542065532 0. -0.1155484333
                    "#", 2
                "SYMMETRIC"
          0. -0.02310968665 0. 0.
     -0.02762135863 0. 0. 0.02310968665
-0.03866990209 0.1617678066 0. 0.06932905996
 0.1933495104 0.1155484333 0. 0.02310968665
 0.5800485313 -0.2542065532 0. -0.1155484333
                    "#", 3
                "SYMMETRIC"
           0. 0.02310968665 0. 0.
    -0.02762135863 0. 0. -0.02310968665
-0.03866990209 -0.1617678066 0. -0.06932905996
0.1933495104 -0.1155484333 0. -0.02310968665
  0.5800485313 0.2542065532 0. 0.1155484333
                    "#", 4
                "SYMMETRIC"
          0. -0.02310968665 0. 0.
    -0.02762135863 0. 0. -0.02310968665
-0.03866990209 0.1617678066 0. -0.06932905996
  0.1933495104 0.1155484333 0. -0.02310968665
  0.5800485313 -0.2542065532 0. 0.1155484333
                    "#", 5
                "DIAGONAL"
                    "#", 6
                "DIAGONAL"
                    "#", 7
```

"DIAGONAL"
"#", 8
"DIAGONAL"
"#", 9
"BLOCK DIAGONAL"
"#", 10
"BLOCK DIAGONAL"
"#", 10
"BLOCK DIAGONAL"
"#", 12
"BLOCK DIAGONAL"
"#", 12
"BLOCK DIAGONAL"
"#", 14
"BLOCK DIAGONAL"
"#", 16
"BLOCK DIAGONAL"
"#", 17
"SYMMETRIC"

$$\left[\frac{1}{512}\sqrt{2}(-10+YI) - \frac{1}{512}\frac{\sqrt{2}YI(-10+YI)}{Y2^{-2}} - 0 - \frac{1}{512}\sqrt{2}Y3\right]$$

 $\left[-\frac{1}{256}\sqrt{2}(-35+2YI) - \frac{1}{256}\frac{\sqrt{2}(-2Y2^{-2}+5YI-105)}{Y2^{-2}} - 0 - \frac{1}{512}\sqrt{2}Y3\right]$
 $\left[\frac{3}{256}\sqrt{2}(35+YI) - \frac{1}{256}\frac{\sqrt{2}(-3Y2^{-2}+25YI-175)}{Y2^{-2}} - 0 - \frac{5}{512}\sqrt{2}Y3\right]$
 $\left[\frac{3}{256}\sqrt{2}(35+YI) - \frac{1}{256}\frac{\sqrt{2}(-3Y2^{-2}+25YI-175)}{Y2^{-2}} - 0 - \frac{5}{512}\sqrt{2}Y3\right]$
 $\left[\frac{1}{512}\sqrt{2}YI - \frac{1}{512}Y2^{-2}\sqrt{2} - 0 - 0\right]$
 $\left[\frac{1}{512}\sqrt{2}(-10+YI) - \frac{1}{512}\frac{\sqrt{2}YI(-10+YI)}{Y2^{-2}} - 0 - \frac{1}{512}\sqrt{2}Y3\right]$

$$\begin{bmatrix} -\frac{1}{256}\sqrt{2}(7+2YI) - \frac{1}{256}\frac{\sqrt{2}(-2Y2-^{2}+15YI-105)}{Y2-} & 0 - \frac{3}{512}\sqrt{2}Y3 \end{bmatrix}$$

$$\begin{bmatrix} -\frac{1}{256}\sqrt{2}(-35+2YI) - \frac{1}{256}\frac{\sqrt{2}(-2Y2-^{2}+5YI-35)}{Y2-} & 0 - \frac{1}{512}\sqrt{2}Y3 \end{bmatrix}$$

$$\begin{bmatrix} \frac{3}{256}\sqrt{2}(35+YI) \frac{1}{256}\frac{\sqrt{2}(-3Y2-^{2}+25YI-175)}{Y2-} & 0 - \frac{5}{512}\sqrt{2}Y3 \end{bmatrix}$$

$$\begin{bmatrix} \frac{1}{512}\sqrt{2}(35+YI) \frac{1}{256}\frac{\sqrt{2}(-3Y2-^{2}+25YI-175)}{Y2-} & 0 - \frac{5}{512}\sqrt{2}Y3 \end{bmatrix}$$

$$\begin{bmatrix} \frac{1}{512}\sqrt{2}(-10+YI) - \frac{1}{512}\frac{\sqrt{2}YI}{Y2-} & 0 - 0 \end{bmatrix}$$

$$\begin{bmatrix} \frac{1}{512}\sqrt{2}(-10+YI) - \frac{1}{512}\frac{\sqrt{2}(-2Y2-^{2}+15YI-105)}{Y2-} & 0 - \frac{3}{512}\sqrt{2}Y3 \end{bmatrix}$$

$$\begin{bmatrix} -\frac{1}{256}\sqrt{2}(-35+2YI) - \frac{1}{256}\frac{\sqrt{2}(-2Y2-^{2}+5YI-35)}{Y2-} & 0 - \frac{1}{512}\sqrt{2}Y3 \end{bmatrix}$$

$$\begin{bmatrix} \frac{3}{256}\sqrt{2}(35+YI) - \frac{1}{256}\frac{\sqrt{2}(-3Y2-^{2}+25YI-175)}{Y2-} & 0 - \frac{5}{512}\sqrt{2}Y3 \end{bmatrix}$$

$$\begin{bmatrix} \frac{1}{512}\sqrt{2}(-10+YI) - \frac{1}{512}\frac{\sqrt{2}YI}{Y2-} & 0 - \frac{5}{512}\sqrt{2}Y3 \end{bmatrix}$$

$$\begin{bmatrix} \frac{1}{512}\sqrt{2}(-10+YI) - \frac{1}{512}\frac{\sqrt{2}(-3Y2-^{2}+25YI-175)}{Y2-} & 0 - \frac{5}{512}\sqrt{2}Y3 \end{bmatrix}$$

$$\begin{bmatrix} \frac{1}{512}\sqrt{2}(-10+YI) - \frac{1}{512}\frac{\sqrt{2}(-2Y2-^{2}+15YI-105)}{Y2-} & 0 - \frac{5}{512}\sqrt{2}Y3 \end{bmatrix}$$

$$\begin{bmatrix} -\frac{1}{256}\sqrt{2}(-10+YI) - \frac{1}{512}\frac{\sqrt{2}(-2Y2-^{2}+25YI-175)}{Y2-} & 0 - \frac{5}{512}\sqrt{2}Y3 \end{bmatrix}$$

$$\begin{bmatrix} -\frac{1}{256}\sqrt{2}(-10+YI) - \frac{1}{512}\frac{\sqrt{2}(-2Y2-^{2}+15YI-105)}{Y2-} & 0 - \frac{1}{512}\sqrt{2}Y3 \end{bmatrix}$$

$$\begin{bmatrix} -\frac{1}{256}\sqrt{2}(-35+2YI) - \frac{1}{256}\frac{\sqrt{2}(-2Y2-^{2}+5YI-35)}{Y2-} & 0 - \frac{1}{512}\sqrt{2}Y3 \end{bmatrix}$$

$$\begin{bmatrix} -\frac{1}{256}\sqrt{2}(-35+2YI) - \frac{1}{256}\frac{\sqrt{2}(-2Y2-^{2}+5YI-35)}{Y2-} & 0 - \frac{1}{512}\sqrt{2}Y3 \end{bmatrix}$$

$$\begin{bmatrix} -\frac{1}{256}\sqrt{2}(-35+2YI) - \frac{1}{256}\frac{\sqrt{2}(-2Y2-^{2}+5YI-35)}{Y2-} & 0 - \frac{5}{512}\sqrt{2}Y3 \end{bmatrix}$$

$$\begin{bmatrix} -\frac{1}{256}\sqrt{2}(-35+2YI) - \frac{1}{256}\frac{\sqrt{2}(-2Y2-^{2}+5YI-35)}{Y2-} & 0 - \frac{5}{512}\sqrt{2}Y3 \end{bmatrix}$$

$$\begin{bmatrix} -\frac{1}{256}\sqrt{2}(-35+2YI) - \frac{1}{256}\frac{\sqrt{2}(-3Y2-^{2}+25YI-175)}{Y2-} & 0 - \frac{5}{512}\sqrt{2}Y3 \end{bmatrix}$$

$$\begin{bmatrix} -\frac{1}{256}\sqrt{2}(-35+2YI) - \frac{1}{256}\frac{\sqrt{2}(-3Y2-^{2}+25YI-175)}{Y2-} & 0 - \frac{5}{512}\sqrt{2}Y3 \end{bmatrix}$$

gs1 := eval(eval(g, linsol), sollist[19]) · renorm : #Select one of the solutions
 Ghat := add(Matrix(gs1[1..n, 1..n, k]) · z^{k-1}, k=1..L) : #z-transform of the scaling filter
 solve({Determinant(Ghat) = 0, abs(z) = 1}) #Check the sufficient condition for orthogonality

(1*E*)



220

Appendix D

Permission to use IEEE Copyrighted material

Comments/Response to Case ID: 003BD3BB

ReplyTo: Copyrights@ieee.org
From: Jacqueline Hansson
Date: 02/15/2013
Subject: Re: Copyright query
Send To: "Walden, Andrew T" <a.walden@imperial.ac.uk>
cc: "Ginzberg, Paul" <paul.ginzberg05@imperial.ac.uk>

Dear Andrew Walden,

It's my understanding that your PhD student or any PhD student will only use portions of their elsewhere published works in their t heses. If that is true and he is referenicing his sources properly and indicating IEEE copyright, there is no probem with Statement being used in his thesis.

Sincerely,

APPENDIX D. PERMISSION TO USE IEEE COPYRIGHTED MATERIAL 222

Jacqueline Hansson, Coordinator IEEE Intellectual Property Rights Office 445 Hoes Lane Piscataway, NJ 08855-1331 USA +1 732 562 3828 (phone) +1 732 562 1746(fax) e-mail: j.hansson@ieee.org IEEE Fostering technological innovation and excellence for the benefit of humanity. cccccccccccccccccccccccc

Hi,

I am the supervisor of a PhD student who has published (with me) 3 papers in IEEE Transactions on Signal Processing resulting from his thesis work. Imperial College London are asking students to include the following in their thesis which he is currently completing:

[1] `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'

In the IEEE Copyright form we signed for each of the 3 papers it says
[2] `Authors/employers may reproduce or authorize others to reproduce the
Work, material extracted verbatim from the Work, or derivative works for

APPENDIX D. PERMISSION TO USE IEEE COPYRIGHTED MATERIAL 223

the author's personal use or for company use, provided that the source and the IEEE copyright notice are indicated, the copies are not used in any way that implies IEEE endorsement of a product or service of any employer, and the copies themselves are not offered for sale.'

My question is this. If the student includes the statement [1] can he still include verbatim bits of the final draft of the 3 papers post refereeing provided that the source and the IEEE copyright notice are indicated as in [2] ?

I have seen US theses which seem to do this, but I wanted to check with you.

Thank you very much,

Andrew Walden

Andrew Walden Professor of Statistics, Dept. Mathematics, Imperial College London, London SW7 2AZ UK http://stats.ma.ic.ac.uk/~atw http://www3.imperial.ac.uk/people/a.walden

Appendix E

Adaptive Orthogonal Matrix-Valued Wavelets and Compression of Vector-Valued signals

ADAPTIVE ORTHOGONAL MATRIX-VALUED WAVELETS AND COMPRESSION OF VECTOR-VALUED SIGNALS

P. Ginzberg and A. T. Walden

Department of Mathematics, Imperial College London, 180 Queen's Gate, London SW7 2BZ, UK. (e-mail: paul.ginzberg05@imperial.ac.uk and a.walden@imperial.ac.uk)

ABSTRACT

Wavelet transforms using matrix-valued wavelets (MVWs) can process the components of vector-valued signals jointly, and thus offer potential advantages over scalar wavelets. For every matrix-valued scaling filter, there are infinitely many matrix-valued wavelet filters corresponding to rotated bases. We show how the arbitrary orthogonal factor in the choice of wavelet filter can be selected adaptively with a modified SIMPLIMAX algorithm. The 3 × 3 orthogonal matrix-valued scaling filters of length 6 with 3 vanishing moments have one intrinsic free scalar parameter in addition to three scalar rotation parameters. Tests suggest that even when optimising over these parameters, no significant improvement is obtained when compared to the naive scalar-based filter. We have found however in an image compression test that, for the naive scaling filter, adaptive basis rotation can decrease the RMSE by over 20%.

Index Terms— multichannel wavelet, vector-valued wavelet, matrix-valued wavelet, basis rotation, SIMPLIMAX, compression, scalar thresholding

1. INTRODUCTION

The naive approach for applying wavelet-based methods to vector-valued data is to transform each component independently with a scalar wavelet transform. An $n \times n$ matrix-valued wavelet (MVW) is a type of wavelet which is specifically designed to jointly transform the components of *n*-vector-valued signals [6, 14]. The coefficients of a matrixvalued scaling filter (MVSF) or matrix-valued wavelet filter (MVWF) are $n \times n$ matrices. The increased number of degrees of freedom offered by such filters allows one, for example, to build finite impulse-response (FIR) MVSFs which are orthogonal, symmetric, and have high vanishing moments, such as the quaternion (4 × 4) construction in [7].

In a search of the MVW literature, we have come across only four explicit MVW designs of practical interest. [4] devised a procedure based on multichannel lifting to construct biorthogonal MVWs, and gives coefficients for the 2×2 case. A 3×3 example based on the same method is given in [2]. [8] construct two examples of biorthogonal MVWs by solving a set of design equations symbolically. The construction from [2] has been applied to the compression, denoising [1] and watermarking [3] of colour images. The construction from [8] has been applied to the compression and denoising of 2-D vector wind fields [9, 13]. In addition to these, the authors have constructed 2×2 , and quaternion (4 × 4) orthogonal MVWs [7].

One characteristic which all these constructions share is that they contain free parameters which must be specified. In [8], the free parameters are chosen such that the scaling and wavelet filters resemble ideal lowpass and highpass filters as closely as possible. In [2, Fig. 7] the performance for a few parameter choices are compared. In this paper, we develop a method which allows us to systematically select the free parameter in the orthogonal 3×3 construction based on [7], in order to optimise its performance for signal compression. Since the optimisation can be performed for a specific signal, this can be considered as a method for implementing an adaptive wavelet transform. However, whilst the adaptive optimisation of the wavelet filter for a given scaling filter can be done in a computationally efficient manner, we will use brute force (trying a large number of parameter values) to optimise the scaling filter.

In Section 2 we introduce MVWs. In Sections 3, 4 and 5 we classify the three types of free parameters. These are, respectively, an arbitrary orthogonal similarity transformation of the scaling filter, an intrinsic parameter in the scaling filter design, and an arbitrary rotation of the wavelet filter which controls the wavelet coefficient basis. We suggest an algorithm for the adaptive optimisation of the latter in Section 6. Section 4 describes the set of all 3×3 orthogonal MVSFs of length 6 with 3 vanishing moments. In Section 7 we systematically test the effects of parameter choices for these filters on a test image, which leads to some insights on MVW design.

2. MATRIX-VALUED WAVELETS

Paul Ginzberg thanks the EPSRC (UK) for financial support.

A (discrete) MVW transform decomposes a vector-valued signal $\pmb{f}(t)\in L^2(R,R^{1\times n})$ into a linear combination

$$\boldsymbol{f}(t) = \sum_{k \in \mathbb{Z}} \boldsymbol{s}_k 2^{-J/2} \boldsymbol{\Phi}(2^{-J}t \cdot k) + \sum_{k \in \mathbb{Z}, j < J} \boldsymbol{w}_{j,k} 2^{-j/2} \boldsymbol{\Psi}(2^{-j}t \cdot k)$$
(1)

of the translations and dilations of a matrix-valued (MV) scaling function $\Phi(t) \in L^2(R, R^{n \times n})$, and a MV wavelet function $\Psi(t) \in L^2(R, R^{n \times n})$, with coefficients $s_k, w_{j,k} \in \mathbb{R}^{1 \times n}$.

 Φ and Ψ satisfy the dilation equations

$$\boldsymbol{\Phi}(t) = \sqrt{2} \sum_{k \in \mathbb{Z}} \boldsymbol{G}_k \boldsymbol{\Phi}(2t-k), \ \boldsymbol{\Psi}(t) = \sqrt{2} \sum_{k \in \mathbb{Z}} \boldsymbol{H}_k \boldsymbol{\Phi}(2t-k),$$

where $\{G_k\}$ and $\{H_k\}$ are $n \times n$ matrix-valued sequences, called the matrix-valued scaling filter (MVSF) and matrix-valued wavelet filter (MVWF) respectively.

MVWs are a type of generalized multiwavelet. Indeed, the MVW transform can be implemented as a fast multiwavelet transform. There is however no need for vectorization, pre-filtering or post-filtering steps since the signal is already in vector form. MVSF coefficients satisfy $2^{-\frac{1}{2}} \sum_{k \in \mathbb{Z}} G_k = I_n$. This sets them apart from standard multiwavelets, for which the sum has one eigenvalue equal to 1, and all other eigenvalues strictly less than 1 in absolute value.

In this paper, we will deal only with orthogonal MVWs, i.e. MVWs for which the basis of $L^2(\mathbb{R}, \mathbb{R}^{1 \times n})$ used in the decomposition (1) is orthonormal. Also, we will only deal with MVSFs $\{G_k\}$ having finite length L, i.e. $G_k \neq \mathbf{0}_{n \times n}$ only for $0 \le k < L$. Particular attention will be given to the case n = 3 and L = 6.

3. ORTHOGONAL SIMILARITY TRANSFORMATIONS

Definition 1 Two filters $\{G_k\}$ and $\{J_k\}$ are orthogonally similar iff

$$\boldsymbol{J}_k = \boldsymbol{O}\boldsymbol{G}_k \boldsymbol{O}^T, \forall k \in \mathbb{Z}$$
(2)

for some orthogonal matrix O, (i.e., $OO^T = I_n$).

The map $\{G_k\} \mapsto \{OG_kO^T\}$ is called an orthogonal similarity transformation (OST).

OSTs preserve orthogonality, filter length and vanishing moments [7].

For a given MVSF $\{G_k\}$, we can generate a whole family of MVSFs $\{OG_kO^T\}$ by taking OSTs. It is convenient to group MVSFs into orthogonally similar families, which can be described by an arbitrarily chosen representative element.

Given a scaling and wavelet filter pair $\{G_k\}$, $\{H_k\}$, we will apply any OST to both filters, to obtain a valid scaling and wavelet filter pair $\{OG_kO^T\}, \{OH_kO^T\}$.

Let O(3) denote the set of 3×3 orthogonal matrices and $SO(3) = \{ \boldsymbol{O} \in O(3) : \det(\boldsymbol{O}) = 1 \}$ denote the set of 3×3 rotation matrices. Then $O(3) = SO(3) \cup (-SO(3))$. However, for any $\boldsymbol{G}_k, \boldsymbol{O} \in \mathbb{R}^{3 \times 3}, (-\boldsymbol{O})\boldsymbol{G}_k(-\boldsymbol{O})^T = \boldsymbol{O}\boldsymbol{G}_k\boldsymbol{O}^T$. Hence, we only need to consider OSTs with rotations $\boldsymbol{O} \in SO(3)$.

We can parameterise the rotations $O \in SO(3)$ using 3 Euler angles $(\theta_1, \theta_2, \theta_3) \in]-\pi, \pi] \times [0, \pi] \times]-\pi, \pi]$. Given a MVSF $\{G_k\}$, we will want to select an optimal MVSF within the family of orthogonally similar filters that it generates, by choosing appropriate values for the parameters $\theta_1, \theta_2, \theta_3$.

If we wish to select these parameters *before* observing the signal to be compressed (i.e. non-adaptively), and the properties of the unknown signal are a-priori invariant under rotations (e.g. the coordinate system used for the signal is unknown and arbitrary), then the choice of OST is irrelevant and we may arbitrarily set $\theta_1 = \theta_2 = \theta_3 = 0$, ($O = I_3$).

4. INTRINSIC PARAMETERS

After considering OSTs, there may still be additional free parameters in the design of the MVSF. For the set of 3×3 MVSFs of length 6 with 3 vanishing moments there is one such free parameter, denoted x^* .

We can describe the set of all orthogonal 3×3 MVSFs of length 6 with 3 vanishing moments as follows:

There are two naive filters. One is given by $\{g_k I_3\}$, where $\{g_k\}$ is the scalar minimum phase Daubechies scaling filter, and the other by its time-reversal $\{g_{5-k}I_3\}$.

The non-naive filters are either orthogonally similar to

$$\left\{ \left(\begin{array}{cc} g_k & 0 & 0 \\ 0 & J_k(x) \\ 0 & J_k(x) \end{array} \right) \right\},\tag{3}$$

or to its time-reversal, where g_k is as above, and J_k is the non-trivial 2×2 MVSF construction of length 6 with 3 vanishing moments, given in [7] with free parameter $0 \le x \le C = [5 + 2\sqrt{10}]^{1/2}$;

We treat the non-naive filters as a single family, parameterised by $-1 \le x^* \le 1$ as follows: If $0 \le x^* \le 1$ then select (3)., with $x = Cx^*$. If $-1 \le x^* < 0$ then take the time-reversal of (3) with $x = -Cx^*$.

5. BASIS SELECTION AND ROTATION OF THE WAVELET FILTER

For a given MVSF $\{G_k\}$, a corresponding MVWF $\{H_k\}$ can be computed using the method described in [10, Thm. 10.2, Coroll. 9.2] (see also [7]). However, the choice of MVWF for a given MVSF is not unique. Indeed, any filter of the form $\{RH_k\}$ where R is an orthogonal matrix is also valid. Hence, we may wish to optimise the choice of R. Consider the matrix W whose rows are given by the various wavelet coefficients $w_{j,k}$ obtained from (1). (We assume that in practice the signal being transformed is finite and discrete, so that there are only finitely many wavelet coefficients.) Then the matrix of wavelet coefficients obtained by using $\{RH_k\}$ as our wavelet filter instead of $\{H_k\}$ (and hence $R\Psi$ instead of Ψ) will simply be WR^T . In other words, choosing R is equivalent to selecting the orthonormal basis under which we will encode the wavelet coefficients.

When applying the MVW transform to images (or more generally using transforms with more than one time dimension or wavelet packet transforms) the effects of rotating the wavelet filter or rotating the wavelet coefficient basis are subtly different due to further filtering being applied after the wavelet filter. Thus, treating this situation in its full generality requires that we consider two separate rotation parameters \mathbf{R} . We will avoid this complication resulting from the non-commutative interaction between vertical and horizontal transform components by considering only the problem of finding an optimal rotation of the wavelet coefficient basis. This is the more tractable rotation to optimise, since rotated wavelet coefficients can be obtained without recomputing the MVW transform.

For certain applications, such as those based on vector thresholding, the choice of basis is irrelevant. We can then arbitrarily choose $\mathbf{R} = \mathbf{I}_n$. In the context of compression by scalar thresholding however, selecting an appropriate basis can significantly improve performance. When n = 3, since choosing $\mathbf{R} = -\mathbf{I}_3$ will not affect results, we again need only consider $\mathbf{R} \in SO(3)$, parameterised by three Euler angles $\tilde{\theta}_1, \tilde{\theta}_2, \tilde{\theta}_3$. (Since inversions and permutations of the axes will not affect results, we could decide to restrict the 3D range of $(\tilde{\theta}_1, \tilde{\theta}_2, \tilde{\theta}_3)$ by a factor of 24. This is done by "quotienting out" the rotation group of the cube from SO(3).)

6. MODIFIED SIMPLIMAX ALGORITHM

Let $\tau_p : \mathbb{R}^{m \times n} \to \mathbb{R}^{m \times n}$ denote the hard scalar thresholding operator which sets the 100p% smallest entries of a wavelet coefficient matrix \boldsymbol{W} to 0. We wish to minimise the L^2 distance between the original signal and the signal reconstructed from the thresholded coefficients. We call this quantity the root mean squared error (RMSE). Since the orthogonal wavelet transform is an isometry, this is given by $\mathrm{RMSE} = ||\tau_p(\boldsymbol{WR}^T) - \boldsymbol{WR}^T||_2$, where $||\bullet||_2$ denotes the Frobenius norm. The problem of minimizing this quantity over $\boldsymbol{R} \in O(3)$ can be solved by a simpler orthogonal variant of the SIMPLIMAX algorithm used in factor analysis, as hinted at in [11, p. 578]. The algorithm is based on [5, Case II] and proceeds as follows:

Start from an initial guess R_0 and recursively set $R_{k+1} = U_k V_k^T R_k$, where U_k and V_k are obtained from the singular value decomposition $M_k = U_k D_k V_k^T$ of $M_k = R_k W^T \tau_p (W R_k^T)$. The RMSE decreases at each iteration,

until convergence.

Like many non-convex optimisation routines, this procedure suffers from the fact that it may converge to a local minimum. To mitigate this problem, random initial guesses are used in addition to the default choice $\mathbf{R}_0 = \mathbf{I}_n$. In our applications, we computed the RMSE for 2000 random \mathbf{R} , and selected the best 4 rotations as additional random starting values \mathbf{R}_0 .

Uniformly distributed random rotations are generated using the rotation-invariant (Haar) measure [12].

Remark 1 The same algorithm can be applied with quantization operators other than τ_p .

7. NUMERICAL RESULTS AND INTERPRETATION

We will take as our signal f the well known 512×512 test colour image *Lena* in 24-bit RGB format.

We considered 56 values of x^* and (due to computational time constraints) 100 OSTs ($O = I_3$ and a further 99 uniformly distributed random OSTs). For each combination of x^* and O, we computed the full MVW transform of the image and optimised the choice of wavelet coefficient basis rotation R through the modified SIMPLIMAX algorithm. The relative RMSE, rRMSE = RMSE $||W||_2^{-1}$, was computed after thresholding p = 90% of wavelet coefficients.

The naive filter built from the minimum phase scalar Daubechies scaling filter of length 6 gives an rRMSE of 8.75%. When $O = R = I_3$, the lowest rRMSE is obtained for the diagonal MVSF corresponding to $x^* = -1$ and equals 8.74%. To remove the influence of our choice of representative element amongst orthogonally similar wavelets, we average results over the 100 OSTs. Then the lowest average rRMSE obtained from non-naive filters is 8.82%, for $x^* = 0$. Hence the unoptimised MVSFs are generally underperforming relative to the naive filter. We see from Fig. 1 that even after optimising the choice of both O and x^* , the decrease in rRMSE relative to the naive filter is less than 2%. Again $x^* = 0$ is optimal.

Optimisation over R on the other hand can provide a significant improvement in performance at a much lower computational cost. This optimisation is however particularly effective for the naive filter, leading to a 12.9% decrease in rRMSE to 7.62%. Again, MVSFs underperform.

Experiments on the 512×512 images *mandrill*, *peppers* and *airplane* give qualitatively similar results to Fig. 1, except for different ranges of rRMSE. The values for the naive filter before and after optimisation are given in Table 1.

We believe that optimisation over \mathbf{R} is particularly effective for the naive filters because the phases of the filters applied to each component match, leading to better alignment of the large wavelet coefficients across the 3 columns of \mathbf{W} . This explanation is consistent with the fact that optimisation over \mathbf{R} is more effective for $x^* = 0$ and $x^* = \pm 1$, values at



Fig. 1. Relative RMSE after setting p = 90% of coefficients in the MVW transform of *Lena* to 0, for varying x^* and various degrees of optimisation. From top to bottom, the dashdotted curve is for no optimisation (averaged over OSTs), the dotted curve is after optimising O, the dashed curve is after optimising R (averaged over OSTs), the full curve is after jointly optimising both O and R. The horizontal lines correspond to the naive minimum-phase filter, before (square markers) and after (round markers) optimisation of R.

which two out of the three filter dimensions will have matching phases, in some sense. Lack of proper alignment of the wavelet coefficients is also problematic for applications based on vector thresholding, and may be at the root of the overall disappointing performance of the non-naive 3×3 wavelets. Although symmetric (zero phase) MVWs exist for n = 2, 4, currently no example exists for odd n.

Optimisation of \mathbf{R} is useful because the distribution of wavelet coefficients in \mathbb{R}^3 is anisotropic. Indeed, for naive wavelet filters, the wavelet coefficients which encode a sharp edge between two uniformly coloured regions will lie along a line through the origin. One of the reasons for the lesser effectiveness of basis selection for MVWs may be that they do not exhibit this behavior. If we treat the anisotropy as ellipsoidal, then the major and minor axes provide a heuristic choice of basis. In other words we may choose \mathbf{R} such that $\mathbf{W}^T \mathbf{W} = \mathbf{R}^T \mathbf{D} \mathbf{R}$, with \mathbf{D} diagonal. This heuristic can also be used as a starting guess for the SIMPLIMAX algorithm.

8. REFERENCES

 S. Agreste and A. Vocaturo, "Multichannel wavelet scheme for color image processing." In Applied and Industrial Mathematics in Italy III: Selected Contribu-

image	rRMSE	optimised	improvement
	(%)	rRMSE (%)	(%)
Lena	8.75	7.65	12.9
mandrill	25.3	19.7	22.1
peppers	8.25	8.02	2.86
airplane	7.88	5.71	27.5

Table 1. rRMSE obtained for the naive Daubechies filter before and after optimisation of \mathbf{R} . p = 90% of the wavelet coefficients are set to 0.

tions from the 9th SIMAI Conference, Rome, Italy 15-19 September 2008, pp. 1–12, 2009.

- [2] S. Agreste and A. Vocaturo, "A new class of full rank filters in the context of digital color image processing." In *Proceedings of the 10th European Congress of ISS*, Bologna, Italy, 2009, pp. 1–6.
- [3] S. Agreste and A. Vocaturo, "Wavelet and multichannel wavelet based watermarking algorithms for digital color images." In *Communications to SIMAI Congress*, vol. 3, 2009, pp. 242.1–242.11.
- [4] S. Bacchelli, M. Cotronei & T. Sauer, "Multifilters with and without prefilters," *BIT Numerical Mathematics*, vol. 42, pp. 231–261, 2001.
- [5] N. Cliff, "Orthogonal rotation to congruence," *Psychometrika*, vol. 31, pp. 33–42, 1966.
- [6] J. E. Fowler and L. Hua, "Wavelet transforms for vector fields using omnidirectionally balanced multiwavelets," *IEEE Trans. Signal Process.*, vol. 50, pp. 3018–3027, 2002.
- [7] P. Ginzberg and A. T. Walden, "Matrix-valued and quaternion wavelets," submitted to *IEEE Trans. Signal Process*, 2012.
- [8] L. Hua and J. E. Fowler, "Technical details on a family of omnidirectionally balanced symmetric-antisymmetric multiwavelets." Technical report, Mississippi State University. http://citeseerx.ist.psu.edu/viewdoc/summary ?doi=10.1.1.11.3802, 2002.
- [9] L. Hua and J. E. Fowler, "Wavelet-based coding of time-varying vector fields of ocean-surface winds," *IEEE Trans. Geosci. Remote Sensing*, vol. 42, pp. 1283–1290.
- [10] F. Keinert, *Wavelets and Multiwavelets*. Chapman and Hall/CRC, 2003.
- [11] H. A. L. Kiers, "Simplimax: Oblique rotation to an optimal target with simple structure," *Psychometrika*, vol. 59, pp. 567–579, 1994.
- [12] R. E. Miles, "On random rotations in R³," *Biometrika*, vol. 52, pp. 636–639, 1965.
- [13] M. A. Westenberg and E. Thomas, "Denoising 2-D vector fields by vector wavelet thresholding," J. WSCG vol. 13, pp. 33–40, 2005.
- [14] X.-G. Xia and B. W. Suter, "Vector-valued wavelets and vector filter banks," *IEEE Trans. Signal Process.*, vol. 44, pp. 508–18, 1996.