A wavelet approach to cardiac signal processing for low-power hardware applications

Citation for published version (APA):

Karel, J. M. H. (2009). *A wavelet approach to cardiac signal processing for low-power hardware applications*. [Doctoral Thesis, Maastricht University]. Datawyse / Universitaire Pers Maastricht. <https://doi.org/10.26481/dis.20091215jk>

Document status and date:

Published: 01/01/2009

DOI:

[10.26481/dis.20091215jk](https://doi.org/10.26481/dis.20091215jk)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Download date: 26 Apr. 2023

**A WAVELET APPROACH TO CARDIAC SIGNAL PROCESSING FOR
LOW-POWER HARDWARE APPLICATIONS**

Joël M.H. Karel

©Copyright Joël M.H. Karel, Maastricht 2009
Universitaire Pers Maastricht
ISBN 978-90-5278-887-6

All rights reserved. No part of this thesis may be reproduced, stored in a retrieval system of any nature, or transmitted in any form by any means, electronic, mechanical, photocopying, recording or otherwise, included a complete or partial transcription, without the permission of the author.

A WAVELET APPROACH TO CARDIAC SIGNAL PROCESSING FOR LOW-POWER HARDWARE APPLICATIONS

PROEFSCHRIFT

**ter verkrijging van de graad van doctor aan de Universiteit Maastricht, op
gezag van de Rector Magnificus, Prof. mr. G.P.M.F. Mols volgens het
besluit van het College van Decanen, in het openbaar te verdedigen op
dinsdag 15 december 2009 om 14.00 uur**

door

Joël Matheus Hendrikus Karel

Promotor:

Prof. dr. ir. R.L.M. Peeters

Copromotor:

Dr. R.L. Westra

Beoordelingscommissie:

Prof. dr. H. Kingma (voorzitter)

Prof. dr. M.P.F. Berger

Prof. dr. Y. Rudy (Washington University in St. Louis)

Dr. ir. W.A. Serdijn (Technische Universiteit Delft)

Dr. P.G.A. Volders

Contents

Contents	i
1 Introduction	3
1.1 Research setting	3
1.2 Research questions and thesis outline	4
2 Cardiac signal processing	7
2.1 Introduction	7
2.2 The human heart	8
2.3 Electrocardiogram data	10
2.3.1 The Electrocardiogram	10
2.3.2 Signal archives	14
2.4 Cardiac rhythms and pathologies	14
2.5 Signal processing	15
2.6 Fourier transforms	15
2.7 Laplace and z-transforms	18
2.8 Linear systems	19
2.9 Filtering signals	22
3 Wavelet transformations	25
3.1 Continuous wavelets	26
3.2 Wavelets from filter banks	27
3.2.1 Perfect reconstruction	29
3.2.2 Orthogonal filter banks	29
3.2.3 Multi-resolution analysis	31
3.2.4 Wavelet and scaling functions	31
3.2.5 Vanishing moments	33
3.2.6 Linear phase	34
3.2.7 Polyphase filtering	34
3.3 The stationary wavelet transform	36

3.4	Multiwavelets	40
4	Analog implementation of wavelets	43
4.1	Dynamic translinear systems	43
4.2	Wavelet transformations as linear systems	45
4.3	Padé approximation of wavelet functions	48
4.4	L2 approximation of wavelet functions	51
4.4.1	Wavelets and the L2 space	51
4.4.2	Parameterization	53
4.4.3	Vanishing moments	55
4.4.4	Obtaining a good starting point	57
4.5	Empirical results	64
5	Orthogonal wavelet design	71
5.1	Measures for the quality of a given representation	72
5.2	Wavelet parameterization and design	74
5.2.1	Lattice structure	75
5.2.2	Enforcing additional vanishing moments	79
5.2.3	Design and optimization	82
5.2.4	Experimentation	87
5.3	Multiwavelet parameterization and design	89
5.3.1	Parameterization of lossless systems	90
5.3.2	Parameterization of scalar wavelets	95
5.3.3	Parameterization of multiwavelets	98
5.3.4	Balanced vanishing moments	101
5.3.5	Multiwavelet design	106
6	Biomedical applications of wavelet design	109
6.1	Applications of wavelet design in cardiology	109
6.1.1	Detecting the QRS complex using orthogonal wavelet design . . .	109
6.1.2	QT time measurement using designed multiwavelets	113
6.2	Bias field removal from MR images using wavelet design	118
6.2.1	Radio Frequency inhomogeneities in magnetic resonance images .	118
6.2.2	Bias field removal in magnetic resonance images	118
6.2.3	Wavelet design for RF inhomogeneity detection	119
6.2.4	Filtering MR images with designed wavelets	121
7	Conclusions and directions for further research	125
	Bibliography	129
	Summary	139
	Samenvatting	141

Curriculum Vitae	143
Lists of Symbols and Abbreviations	145
Index	149

Acknowledgements

The BioSens project is a cooperation with the Maastricht University and the Delft University of Technology. It is financially supported by the Technology Foundation STW (project number DTC 6418), applied science division of NWO and the technology programme of the Dutch Ministry of Economic Affairs. In addition it is supported by the following research partners: Medtronic Bakken Research Center Maastricht, Medtronic Subcutaneous Diagnostics & Monitoring in Arnhem, Maastricht Instruments B.V., Twente Medical Systems International B.V. in Enschede, SystematIC design B.V. in Delft and Weijand R&D Consultancy, B.V. in Hellevoetsluis.

A Ph.D. thesis is a long-term project and it is impossible to complete it without the support of many people around you. I would like to express my gratitude to all of them.

My gratitude goes to my promotor and friend Ralf Peeters who was an incredible source of new ideas and made the link between the design of orthogonal wavelets and lossless systems in which he is an expert. My colleague Jordi Heijman was a great help regarding the background on cardiology.

Furthermore I would like to express my gratitude to Ronald Westra, my direct colleague and the project leader of the BioSens project in Maastricht, who, together with Wouter Serdijn and Richard Houben made the BioSens project possible. As a copromotor Ronald enthusiastically proofread several sections of my thesis and helped to make them more enjoyable to read.

A special word of thanks goes to my colleagues from the Delft University of Technology: Sandro Haddad and Wouter Serdijn. Our kind cooperation lead to a considerable number of publications. The trips to Delft were always enjoyable.

To the industry partners. In particular Richard Houben, Jan Peuscher, Frans Smeets and Kiwi Smit who's active participation in the users committee contributed to the success of this project.

To my colleagues of the Department of Knowledge Engineering Who supported me throughout my Ph.D. time and ensured that I had enough time beside the educational load to complete my thesis.

My gratitude also goes to Orazio Gambino who provided the data and supported me with the application on bias field correction.

And last but not least to my friends, family, parents and wife. They endured a long period in which I had little time for them, but also ensured that I was able to take my distance from my thesis when necessary. Due to unforeseen circumstances, handing in the full draft of my Ph.D. thesis to my promotor and sending the manuscript to the reading committee almost coincided with the date of our marriage and the birth of our daughters Romy and Sofia, respectively. However my wife Erica courageously withstood the extra stress involved.

Joël Karel

Chapter 1

Introduction

Paraphrasing Jane Austen [9], one can state that: *It is a truth universally acknowledged, that people die.* Moreover, it is an established fact that among the causes of death, the cardiovascular related diseases worldwide rank as number one [123]. Technical advances, such as pacemakers, help to reduce the share of these diseases in the total number of deaths. Paramount in the future advancement of pacemakers and similar implantable devices, is the progress that can be achieved in their ability to sift, process and interpret sensor data. In this light, this work addresses two relevant aspects of new generation implantable devices: low-power implementation combined with high computing power, and efficient signal processing, using wavelets and various aspects of systems theory.

1.1 Research setting

For readers unfamiliar with cardiology a brief introduction is provided in Sections 2.2–2.4. One possibility of interest for decreasing the mortality among patients at risk of a cardiovascular disorder, is the implantation of a therapeutic device or a monitoring device. A classical example of such a therapeutic device is a pacemaker [122, 51], which is further discussed in Section 2.1. The fact that permanent artificial pacemakers are *implanted* in the human body makes it currently still impossible to recharge them, limiting the lifespan of the device to the battery lifetime.

Nowadays a sensing circuit is present that controls the frequency at which a therapy is applied, contributing to the patients' comfort. The signal processing techniques employed in production pacemakers are relatively simple [51] in general. More advanced algorithms have the clear potential to give a considerable increase in patient comfort by tuning the mode of operation to the electrophysiological state of the heart. However, since the required sensing circuit is always active [49], this introduces a new source of power consumption. The power consumption tends to increase with the complexity of the sense amplifier, and power saving techniques are therefore a relevant topic.

In the last few decades a novel signal processing technique called wavelet transforms has emerged. As is well known, the classical Fourier transform (see Section 2.6) represents a signal in terms of its frequency components. Wavelets, as discussed in Chapter 3, on the other hand represent the signal in terms of both pseudo-frequency (scale) and place (time). In addition, wavelet analysis is capable of flexibly adapting its focus, such that it narrows the time window and widens the frequency window for high frequencies, and widens the time window and narrows the frequency window for low frequencies. This is referred to as the “zoom-in” property of wavelets. Wavelets are nowadays widely used in the field of cardiac signal processing [61, 78, 83, 102, 98, 112, 87] and signal processing in general. However, for a realistic integration in implantable devices, a power-efficient implementation is essential. As argued in Chapter 4, it is opportunistic, from a power consumption perspective, to perform as many computations as possible in the analog domain. This is due to the fact that analog-to-digital converters account for a considerable proportion of the power consumption.

1.2 Research questions and thesis outline

Three main research questions are addressed in this thesis:

1. *How to approximate wavelets for the implementation in continuous-time analog systems.*

In earlier work [53, 50, 49, 54, 48] analog dynamic translinear systems (see Section 4.1) were used as a platform to implement continuous wavelet transforms (see Section 3.1). This involved the approximation of the wavelet transform with the impulse response of a linear system as discussed in Section 4.2. A brief background on linear systems is provided in Section 2.8. Using the technique of Padé approximation the authors of [53] obtained a rational approximation of the Laplace transform of a given wavelet function. In Section 4.3 this approach, that is only suited for the approximation of a limited number of wavelet functions, is discussed in more detail. A new approach, based on L_2 approximation, that can be used for a wider range of wavelet functions is discussed in Sections 4.4 and 4.5. This approach can even be employed for the approximation of the wavelet function of discrete wavelets such as the Daubechies 3 wavelet as demonstrated in Section 4.5. The approximation of discrete wavelets is possible due to the fact that the dilation and wavelet equation, that are discussed in Section 3.2, provide a connection between discrete-time and continuous-time. This is illustrated in Figure 1.1

2. *How to design (multi)wavelets for an application at hand.*

Another important issue of interest is the choice of a wavelet basis [103]. When employing the Fourier transform the basis functions are fixed, i.e., sines and cosines. However for wavelets this is not the case and the basis needs to satisfy the conditions associated with the wavelet framework. One possibility for wavelet selection is to use a parameterized wavelet that can be tuned to the application at hand

Figure 1.1: Overview of the design and analog implementation of wavelets. Continuous-time wavelets can be approximated by matching the associated wavelet function with the impulse response of a linear system. This can then be implemented using the dynamic translinear circuit approach. Orthogonal discrete wavelets have an associated filter bank. Under conditions, using the dilation and wavelet equations, an associated wavelet function can be found that can be approximated. For orthogonal wavelets from filter banks a parameterization and a design criterion are available that allow for the design of wavelets.

[101]. Another possibility is to design a custom wavelet as is the case in this study [69]. Earlier work on the design of wavelets matched the amplitude spectrum and the phase spectrum of a wavelet to a reference signal in the Fourier domain separately [24]. The authors of [46] describe an algorithm for the design of biorthogonal and semi-orthogonal wavelets. The design criterion comes down to the maximization of the energy in the approximation coefficients, however no clear motivation is provided by the authors. In [92] a parameterization of orthogonal wavelets, involving polyphase filters and the lattice structure as for example in [107] (see also Section 3.2), was discussed, however no clear design criterion was provided. In Section 5.1 two design criteria are introduced for discrete wavelets, that can be used to measure the quality of a discrete wavelet for representing a given signal. In this context, a wavelet of “good quality” ensures that the signal’s energy is clustered in the wavelet (time-frequency) domain, i.e., it maximizes sparsity. In Section 5.2 a parameterization of discrete-time orthogonal wavelets involving polyphase filterbanks and the lattice structure as in [92, 107] is discussed. The enforcement of additional vanishing moments and the use of the design criterion is discussed as well. An integral approach for orthogonal wavelet filter design is developed and presented in that section. As discussed in [68] and illustrated in Figure 1.1 it is possible to compute a wavelet function associated with the designed wavelet filter bank. If this wavelet function is sufficiently smooth, it is possible to approximate the designed wavelet in analog circuits.

When multiple well-defined features in a signal need to be distinguished simultaneously, multiwavelets [76, 77, 107] can be employed. In the multiwavelet multiple waveforms are used to jointly generate a basis for the L_2 space. In Section 3.4 it is discussed how these multiwavelets can be parameterized in polyphase terms and how the condition of orthogonality comes down to requiring that a system in polyphase is “lossless” [117]. In order to design multiwavelets, a parameterization of lossless systems is required, as is discussed in Section 5.3. For this parameterization in Section 5.3.1 the “tangential Schur algorithm” [55] is used. This parameterization is first provided for scalar wavelets in Section 5.3.2 and for multiwavelets in Section 5.3.3. One of the issues involved is the enforcement of a first vanishing moment which is a requirement for a valid wavelet multiresolution structure.

3. *To investigate the practical potential of this approach for cardiac applications.*

In Section 6.1 two practical applications of (multi)wavelet design with respect to cardiac signal processing are discussed. In Section 6.1.1 it is demonstrated how scalar wavelet design is useful for the detection of QRS complexes in ECG signals. In Section 6.1.2 the length of the QT interval is estimated using designed multiwavelets. In Section 6.2 the use of wavelet design for an application in 2D image analysis is discussed. This describes the removal of an artifact in magnetic resonance imaging, called the “bias field”, using specially designed wavelets [64].

Chapter 2

Cardiac signal processing

In the first part of this chapter a brief introduction is provided for readers unfamiliar with cardiology in Sections 2.2–2.4. One possibility of interest to decrease the mortality of patients at risk of a cardiovascular disorder, is the implantation of a therapeutic device or a monitoring device. A classical example of such a therapeutic device is a pacemaker [122, 51], as is further discussed in Section 2.1. For this research various datasets with ECG data were used as discussed in Section 2.3. In Section 2.4 it is discussed that due to pathologies various morphologies and rhythms may be exhibited in ECG signals.

In the second part a short introduction to signal processing is provided in Section 2.5. Three classical transforms of interest in this work are the Fourier transform (Section 2.6), the Laplace transform (Section 2.7) and the z -transform (Section 2.7). A brief background in linear systems is given in Section 2.8, followed by an introduction to filters in Section 2.9.

2.1 Introduction

Cardiovascular related diseases are still the major cause of death in The Netherlands, as well as worldwide, according to WHO data [123], accounting for 33% of the total mortalities in 2004 [60]. On average this means over 45,000 death per year or over a hundred deaths per day in The Netherlands alone. However the proportion of deaths due to cardiovascular disorders is decreasing. Prevention, better health care and technological advances account for this development.

Among these technological advances are new drugs, therapies, monitoring devices and implantable therapeutic and diagnostic devices. These devices measure signals that may give an indication of the state of operation of the heart. However, since these signals are often corrupted with noise, a denoising step may be required. Denoising is one of the aspects addressed by the field of *Signal processing* and, in general, monitoring and diagnostic devices rely heavily on signal processing techniques. For diagnostic purposes

it may be necessary to store the signals that have been measured. Some devices, such as implantable diagnostic and monitoring devices, may detect events in order to decide whether or not to store the signals that are being measured.

The *pacemaker* is the most prominent implantable therapeutic device. Annually over 600,000 pacemakers are implanted worldwide [122]. Early artificial pacemakers did not possess a sensing circuit and only consisted of a power source, a pulse generator and an electrode [51]. They simply delivered a fixed-rate pulse to the heart, regardless of the state of the heart. Since this type of pacemaker can compete with the spontaneous activity of the heart, it could cause arrhythmias and other undesired effects. In addition the therapy was also applied during normal operation of the heart, i.e., when it was unnecessary. A consequence of this, is that the built-in battery was depleted relatively fast. To overcome these shortcomings, demand pacemakers were developed that do take intrinsic heart activity into account (see e.g. [51]). As a result of this development, nowadays a sensing circuit is present in pacemakers. In this sensing circuit signal processing is performed as part of the decision stage to determine whether or not stimuli are needed. Therapeutic devices may be critical in order to sustain the life of the patient, therefore their decision algorithms have to be very robust. Failure of these devices may result in lawsuits and costly claims. As a result medical companies are conservative if it comes to employing new technologies [51] in implantable devices.

Because a persons life depends on the correct interpretation of the signals by the pacemaker and other medical devices, it can be concluded that signal processing is of great importance to the field of cardiology.

2.2 The human heart

The human heart is a special organ. It has a size that is slightly bigger than a fist and it mainly consists of muscle tissue, the myocardium, that contracts up to several billions of times during a persons lifespan. The muscle tissue of which the heart consists, is unique in the human body. It is a special kind of involuntary muscle. The heart basically acts as a pump that consists of four chambers. Oxygen-poor blood enters the heart in the *right atrium* (RA). This blood is pumped to the *right ventricle* (RV). From there it is pumped through the lungs for *oxygenation* and this oxygenated blood reenters the heart in the *left atrium* (LA). From there it is pumped to the *left ventricle* (LV), which pumps the oxygenated blood through the rest of the body. Since the atria only pump the blood to the adjacent ventricles their tissue mass is relatively small. In addition the tissue mass of the RV is small compared to that of the LV since the RV only has to pump the blood through the lungs whereas the LV has to pump the blood through the whole body. The cells of the myocardium conduct current and if stimulated by an electrical pulse these muscle cells contract. This is called excitation-contraction coupling [14]. In their resting state these cells are said to be *polarized* and their electrical potential is at *resting potential*. If electrical activation of a cell is initiated, Na^+ and Ca^+ ions travel through ion channels in a membrane in and out the cell. This process is called *depolarization*.

Figure 2.1: Schematic representation of the heart

In the case of ventricular depolarization the Na^+ ions accounts for 99% of the effect. During the 150-300 ms after the heart cell is *depolarized*, the dynamic balance of the electrical currents maintains what is known as an *action potential*. This action potential ends with the *repolarization* of the cell, which involves moving ions into and out of the cell in the opposite direction, allowing the cell to conduct electrical impulses again. For a more detailed background on electrophysiology the reader is referred to [124].

In order for the heart to effectively pump the blood through the body, the contractions of the muscle cells have to be coordinated. Special *pacemaker cells* control the heart rate. This process goes through a number of steps during normal activity:

- i) The *sino-atrial* (SA) node, located at the top of the RA generates an electrical impulse that propagates relatively slowly through the muscles of the atria, such that these are depolarized. The conduction of this pulse from the AV node to the LA tracts mainly over Bachmann's bundle.
- ii) When this electrical pulse reaches the *atrio-ventricular* (AV) node after traversing from the SA node over the anterior, middle, and posterior internodal tracts, the

propagation is delayed. During this delay the transfer of blood from the atria to the ventricles is completed.

- iii) The pulse travels through the His-bundle, the bundle branch and the Purkinje fibers to the endocardium of the ventricles.
- iv) The ventricles contract from bottom to top stimulated by a fast moving pulse. The *anisotropy* of the tissue is effective here; the tissue of the heart has a certain orientation and the electrical pulses travel faster along this direction.

If the regular coordination in the cardiac activity is disturbed this is called an *arrythmia*. The excitation and conduction system consists of the SA node, the internodal tracks, Bachmann's bundle, the AV node, the bundle of His, bundle branches and Purkinje fibers. The AV node is innervated by the autonomic nervous system, however the intrinsic rate of the SA node during normal sinus rhythm is the highest and as a result under non-pathological circumstances the SA node controls the heart rate. Certain chemicals, e.g. epinephrine, may affect the heart rate.

2.3 Electrocardiogram data

A number of different measurements can be conducted on the cardiovascular system. The electrical activity of the heart can be measured with the *electrocardiogram* (ECG or EKG). The heart sounds or vibrations can be measured as the phonocardiogram, and the pressure waves propagating through the arteries as the arterial pulse. The ECG will be discussed in this section, where the reader is referred to [100] for a more thorough discussion on this subject.

2.3.1 The Electrocardiogram

A signal of particular interest in this research is the ECG. The ECG is the measurement of the electrical potential between various points *on* or *in* the body. The measurement electrodes are called *leads*. There are various possible set-ups for lead placement. Typically a 12-lead ECG is acquired by placing the leads on the body (surface ECG). Since the heart is a 3-dimensional (3D) object it is important that the ECG provides information about the hearts activity in three (approximately) orthogonal directions [36]. The standard lead placement for the 12-lead ECG ensures that this spatial information is provided. For a more thorough discussion on lead placement the reader is referred to for example [100, 35]. An example of the 12-lead ECG plus three extra leads for the PTB database (see Section 2.3.2) is displayed in Figure 2.2.

It is also possible to measure the ECG by placing an electrode directly on the tissue of the heart. One then speaks of the *intracardiac ECG* (IECG) [51]. The measurement of an IECG is obviously more invasive than the measurement of the surface ECG. Yet another possibility for placing leads is to implant a monitoring device under the skin that records the subcutaneous ECG.

Figure 2.2: Excerpt of the ECG of patient 1 from the PTB database over 4500 ms. The conventional 12 lead ECG is displayed (leads i, ii, iii, avr, avl, avf, v1, v2, v3, v4, v5, v6) along the three Frank leads (vx, vy, vz). The ECG was digitized at 1 kHz. In this ECG, right-bundle branch block is visible.

Figure 2.3: Segments of the ECG

From the ECG recordings various waves can be distinguished that correspond to the propagation of electrical activity across the heart as described in Section 2.2. The morphology and variability of these waves can vary from lead to lead due to the fact that the heart is a 3D organ. Also the muscle mass is not uniform throughout the heart, and as a consequence it is easier to see the activity of the ventricles than of the atria. The leads may also pick up signals from other sources, such as respiration, muscle movement and electrode movement. Respiration can cause low-frequency noise. But muscle movement, and to a lesser extent electrode movement, may cause in-band noise which can be much harder to filter out. The heart may also suffer from an arrhythmia that gives rise to a different morphology. A number of examples of morphologically different ECG waves is displayed in Figure 2.4.

Due to the step-wise operation of the heart during normal sinus rhythm, a number of physiological meaningful segments can be identified in the ECG (see for example [59]) as displayed in Figure 2.3:

P wave Depolarization of the atria. Corresponds to step i on page 10.

PQ segment Propagation delay in the AV node (step ii) on page 10) corresponds to iso-electric segment.

QRS complex Ventricular depolarization (steps iii and iv on page 10). Due to the muscle mass of the ventricles this generally is the dominant complex. Meanwhile atrial repolarization occurs.

ST segment Ventricular muscle cells maintain their action potential resulting in an iso-electric segment.

T wave Ventricular repolarization.

Sometimes other small amplitude waves such as the “U wave” are visible [59].

A number of intervals of interest can be derived from these segments:

Figure 2.4: Examples of smoothed ECG beats from the MIT-BIH arrhythmia database.

QT interval Duration of ventricular depolarization and repolarization. Also indicated as the duration of the ventricular electrical systole, which is the electrical activity that stimulates the ventricles to contract. During the diastole the myocardium relaxes.

RR interval Duration of a complete ventricular cardiac cycle.

PP interval Duration of a complete atrial cardiac cycle.

Due to the physical meaning of these time segments their proper identification and duration is essential for the assessment of the patients condition. In addition the various segments may have a certain morphology. The duration of the segments and their morphology may be related to a certain rhythm; that may in turn correspond to a certain arrhythmia. In clinical situations a cardiologist judges the ECG, and using his/her expert knowledge, determines the rhythm. However this decision is a challenging task for automated systems. In particular if the requirement is that the system has to work in real-time, and especially if it concerns an implantable device.

2.3.2 Signal archives

The main source of data used in this study are the Physionet signal archives [43]. This is a collection of databases of various types of signals, of which ECG recordings form the majority. The most prominent recordings are the MIT-BIH databases that were recorded by the Beth Israel Deaconess Medical Center and Massachusetts Institute of Technology. In particular the MIT-BIH arrhythmia database is widely used for both the evaluation of arrhythmia detectors as well as for basic research into cardiac dynamics. Additionally the Physikalisch-Technische Bundesanstalt (PTB) database [16] which is also available through Physionet has been used in this research. This same database was used for the Computers in Cardiology 2006 Challenge [88] on QT interval estimation. One of the participants of this challenge developed a set of manual annotations that has been published in [25].

In addition datasets from industry partners were available. Medtronic supplied a number of datasets that were recorded with a Medtronic Reveal® subcutaneous leadless ECG recorder.

2.4 Cardiac rhythms and pathologies

From the morphology and frequency of the various complexes in the ECG, cardiologists can determine the involved cardiac rhythm. Furthermore the patient may suffer from a certain pathology that is associated with the rhythm at hand, or with a specific morphology of one of the complexes in the ECG. In order to arrive at more effective therapies the diagnostic capabilities of therapeutic devices have to be improved. Detecting morphologies in ECGs may be of great importance for expanding these capabilities.

The human heart is a syncytium, which means that electrical pulses can propagate freely from cell to cell in any direction on the myocardium. In normal sinus rhythm this propagation is done in an orderly fashion. However if this process is disturbed, turbulent and disorganized propagation of the waves causes the myocardium to contract in a chaotic manner; it fibrillates. This can occur in the atria, hence atrial fibrillation (AF). Often AF is asymptomatic, i.e., the patient does not experience any grief from this arrhythmia. However, there is an elevated risk of stroke since due to the poor circulation of blood during AF the blood may pool and clot. Ventricular fibrillation (VF), puts the patient in acute risk of sudden cardiac death. In order to restore normal sinus rhythm in case of VF, a defibrillator is used which generates a large electrical discharge to “reset” the heart. For patients at chronic risk of VF, a cardioverter-defibrillator can be implanted that monitors the heart, and in case an arrhythmia is detected, therapy is applied.

There are a large number of arrhythmias and pathologies that can occur [13]. For effective diagnosis and treatment, signal processing is a very important tool, certainly in the absence of a physician when automated diagnosis is required such as in implantable medical devices.

2.5 Signal processing

Signal processing is the field that aims to analyze, manipulate and interpret signals. This includes the removal of noise, separation of signals from various sources, feature extraction, storage (optionally with compression) and signal representation.

Signals can be either *continuous-time* or *discrete-time*, with either *analog* or *digital* values. It is popular convention not to make the distinction between whether the signal is sampled (continuous/discrete-time) and how these samples are quantified (analog/digital). Instead “analog” is a popular term for continuous-time analog signals and “digital” is a popular term for discrete-time digital signals. In this work the terms “analog” and “digital” will refer to the representation of the signal values, regardless whether they are sampled or not.

Most sensor information concerns continuous-time analog signals with the signal values represented in voltage or in ampères. In order to obtain a digital signal, the sensor data has to be quantified. This is done by an analog-to-digital (A/D) converter. This converter has a certain number of bits that together with the range of both the converter and the signal determines the precision it can obtain.

The A/D converter usually also samples the input signal such that the output signal becomes a discrete-time digital signal. The number of samples per time unit is the *sampling rate* or the *sampling frequency*. The *Nyquist frequency* is a property of a discrete-time system and defined as half the sampling frequency. The *Nyquist rate* [94, 104] is twice the highest frequency that can be perfectly reconstructed in a continuous-time signal, using interpolation, from a discrete-time signal which has been sampled at a frequency equal to this Nyquist rate. So if the Nyquist frequency (half of the sampling frequency) of the sampled system is larger than the bandwidth (half of the Nyquist rate) of the continuous-time source signal, then perfect reconstruction should, at least in theory, be possible.

Signals can be represented in various domains. Sensor information such as ECGs is usually represented in the time domain. If one wants to examine certain characteristics of the signal it can be opportunistic to examine the signal in a different domain. For example in order to view the frequency content of a signal one can use the frequency domain, or if one wants to analyze time information and frequency information simultaneously, one can use the wavelet domain. To convert a signal to the frequency domain or to the wavelet domain, the Fourier (or Laplace) transform or the wavelet transform can be used, respectively.

2.6 Fourier transforms

The *Fourier transform* is a linear operator that maps a finite energy time domain function $f(t)$ to a frequency domain function $F(\omega)$ [15, 17]. The function $f(t)$ is expressed in terms of $F(\omega)$ as:

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} F(\omega) e^{i\omega t} d\omega, \quad (2.1)$$

where ω is the angular frequency variable (in radians per second) and i is the complex number. This formula is used in the “synthesis”.

The Fourier transform (used in the “analysis”) is defined as:

$$F(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-i\omega t} f(t) dt, \quad (2.2)$$

Note that the product of the normalization factors of (2.2) and (2.1) has to be equal to $\frac{1}{2\pi}$. In this case the normalization factors are chosen to be symmetrical.

A time domain signal $f(t)$ is said to have *finite energy* if its L_2 -norm is finite:

$$\|f\|_2 = \sqrt{\int_{-\infty}^{\infty} |f(t)|^2 dt} < \infty \quad (2.3)$$

The *angular frequency* ω is related to the frequency h in Hertz (Hz) by the identities

$$\omega = 2\pi h \quad \text{and} \quad h = \frac{\omega}{2\pi}. \quad (2.4)$$

When $F(\omega)$ is expressed in polar coordinates as $F(\omega) = r(\omega)e^{i\phi(\omega)}$, further meaning can be given to the complex function $F(\omega)$ (2.1). The function $r(\omega)$ represents the magnitude, i.e. the amplitude, of the complex harmonics $e^{i\omega t}$ per frequency, and $\phi(\omega)$ their initial phase angles. This can be expressed in a pair of *Bode plots*. The Bode magnitude plot shows the logarithm of the magnitude in relation to the frequency, plotted on a logarithmic frequency axis. The frequency response of a linear, time-invariant system (see Section 2.8) can for example be visualized in such manner. The Bode phase plot shows the initial phase angle in relation to the frequency, again plotted on a logarithmic frequency axis.

Using Euler’s formula (see e.g. [39])

$$e^{i\phi} = \cos \phi + i \sin \phi, \quad (2.5)$$

it becomes clear that the Fourier transform decomposes the signal in terms of (orthogonal) sine and cosine basis functions.

If the Fourier transformed signal has compact support in the sense that there exists an angular frequency ω_0 such that:

$$F(\omega) = 0, \quad \forall |\omega| > \omega_0, \quad (2.6)$$

then the signal is said to be bandlimited with bandwidth ω_0 .

If the function f is periodic then it will obviously have infinite energy. However the *power*, i.e., the energy per time unit, can still be required to be finite. The Fourier transform over a single period will then be calculated and one then sometimes speaks of a *Fourier series*. However it is popular convention to reserve this term for the discrete Fourier transform of a periodic function.

Of the many interesting properties of the Fourier transform, one of particular interest is that *convolution* in the time domain of two functions f and g

$$h(t) = (f * g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t - \tau)d\tau, \quad (2.7)$$

becomes multiplication in the Fourier domain:

$$H(\omega) = \sqrt{2\pi}F(\omega)G(\omega), \quad (2.8)$$

where $F(\omega)$, $G(\omega)$ and $H(\omega)$ are the Fourier transforms of $f(t)$, $g(t)$ and $h(t)$ respectively.

This property is particularly useful in the field of signal processing where convolution is extensively used. The Fourier transform has many more properties. For more details the reader is referred to for example [15, 84].

When determining the discrete Fourier transform (analysis) X of a (possibly complex) discrete-time signal $x = \{x[0], \dots, x[N - 1]\}$, one treats this finite length signal x as if it were periodic. If the complex harmonics $e^{-i\Omega n}$ that are used as a basis for the decomposition are chosen such that an integer number of periods fit in the length of the signal x , this indeed gives the desired result. One then speaks of the *discrete Fourier series*. The discrete Fourier series is defined as:

$$X[k] = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x[n]e^{-i\Omega n}, \quad (2.9)$$

where $\Omega = 2\pi k/N$, $k = 0, 1, \dots, N - 1$. Its inverse transform (synthesis) is given by:

$$x[n] = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X[k]e^{i\Omega n}. \quad (2.10)$$

For a pair of vectors (e.g. finite length discrete-time signals) a and b the convolution product $c = a * b$ is defined as the vector:

$$c[n] = (a * b)[n] = \sum_{\ell=0}^n a[\ell]b[n - \ell], \quad n \in \{0, 1, \dots, N - 1\}. \quad (2.11)$$

In the Fourier domain this again becomes multiplication:

$$C[k] = \sqrt{N}A[k]B[k], \quad (2.12)$$

where A , B and C are the Fourier series of a , b and c respectively.

Since the basis functions of the Fourier transform are sinusoids that extend over the entire real line $\mathbb{R}$, the Fourier transform is not well equipped for studying transient phenomena. A classical way around this issue is to use a windowing function $w(t)$ to arrive at:

$$F_w(\tau, \omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} w(t - \tau)e^{-i\omega t}f(t)dt, \quad (2.13)$$

This transform is called the *windowed Fourier transform* alias the *short time Fourier transform* [40]. The time window $w(t - \tau)$ localizes the transform around time τ and the exponential $e^{-i\omega t}$ in frequency around frequency ω . Due to the Heisenberg uncertainty principle [58] from the field of quantum dynamics, one cannot obtain an arbitrary accurate time-frequency measurement. Instead the product of the “standard deviations” of the time- and frequency windows, i.e., the area of the *uncertainty rectangle*, obeys a certain lower bound: $\sigma_\tau \sigma_\omega \geq \frac{1}{2}$. This area is for instance minimal if $w(t)$ is a Gaussian function, i.e., in the case of the *Gabor transform*. For the windowed Fourier transform the uncertainty rectangle has constant dimensions and a variable placement in the time-frequency plane (see for example [80]).

2.7 Laplace and z-transforms

The (one-sided) *Laplace transform* (see e.g. [62]) is conveniently introduced for functions $y(t)$ defined only for $t \geq 0$. It is a function of a complex variable s and given by:

$$Y(s) = \mathcal{L}\{y(t)\} = \int_{t=0}^{\infty} y(t)e^{-st}dt. \quad (2.14)$$

Formally, it is defined for values of s for which the integral in (2.14) converges, which constitutes the *region of convergence*. This usually is a half-plane. An important class of functions used in signal processing and filtering involves sines, cosines, exponentials, polynomials and from the non-standard functions the impulse functions, all of which have rational Laplace transforms [62].

The continuous-time impulse function is the *Dirac delta-function* $\delta(t)$ is given by:

$$\begin{cases} \delta(t) = 0, & \forall t \neq 0 \\ \int_{-\infty}^{\infty} \delta(t)dt = 1 \end{cases} \quad (2.15)$$

yielding an infinitesimally narrow, infinitely tall pulse which integrates to unity. On the other hand the discrete-time impulse can be associated (via *zero-order hold* (ZOH)) with a continuous time block function, called the *Kronecker delta function*:

$$\begin{cases} \delta[n] = 0, & \forall n \neq 0 \\ \delta[n] = 1, & \text{for } n = 0. \end{cases} \quad (2.16)$$

Writing $s = \sigma + i\omega$, it is clear that the Fourier transform is obtained for $\sigma = 0$:

$$\begin{array}{ccc} Y(i\omega) & = & Y(\omega). \\ \text{(Laplace)} & & \text{(Fourier)} \end{array} \quad (2.17)$$

Note that with abuse of notation we denote the Laplace transform of $y(t)$ as $Y(s)$, the Laplace transform for $s = i\omega$ as $Y(i\omega)$, and the Fourier transform of $y(t)$ as $Y(\omega)$.

As is well-known, for the Laplace transform $Y(s)$ of a function $y(t)$ the *initial value theorem*:

$$y(0) = \lim_{|s| \rightarrow \infty} sY(s), \quad (2.18)$$

and the *final value theorem*:

$$\lim_{t \rightarrow \infty} y(t) = \lim_{s \rightarrow 0} sY(s), \quad (2.19)$$

hold.

In discrete time, a similar transform which is called the z -transform is used. Like s for the Laplace transform, z is a complex variable. The z -transform of an (infinite) sequence $y[k], k \in \mathbb{Z}$ is denoted by $\mathcal{Z}\{y_k\} = Y(z)$ and is given by:

$$Y(z) = \mathcal{Z}\{y_k\} = \sum_k y_k z^{-k}. \quad (2.20)$$

Observe the similarity with the definition of the Laplace transform in (2.14). The z -transform is the Laplace transform of an impulse train of a continuous function, with sampling period $T \in \mathbb{R}^+$. Given this instantaneous sampled function:

$$\sum_{n=-\infty}^{\infty} y(nT) \delta(t - nT), \quad (2.21)$$

where $\delta(k)$ is the *Dirac delta* as in (2.7). One can take the Laplace transform of this sampled function and substitute $z = e^{sT}$ to obtain the z -transform of $y(t)$. When taking $z = e^{i\Omega}$ in (2.20) the Fourier series as in (2.9) is obtained.

2.8 Linear systems

Let S be a *linear system* with input u and output y . Then the system S has the following defining properties:

Homogeneity If the input u to a linear system S is scaled with a factor α , the output of that system is scaled with a factor α as well: $S(\alpha u) = \alpha S(u)$.

Superposition If two inputs are added and passed through a linear system, then the output equals the sum of the outputs as if the two input signals were passed through the system individually: $S(u_1 + u_2) = S(u_1) + S(u_2)$.

Linear systems can be discrete-time or continuous-time, but in this section the focus will be on continuous-time systems. Linear systems can also be *time-invariant*: for any input u producing a corresponding output $y = S(u)$, it holds that $y_\tau = S(u_\tau)$ for any time shift τ , where u_τ, y_τ denote the signals u, y time shifted by τ . If a system is both linear and time-invariant, one calls it a *linear time-invariant system* or *LTI system* for short. The input/output relation of LTI systems can be represented by differential equations. As an example of such an n^{th} order differential equation takes the following form:

$$\begin{aligned} y^{(n)}(t) + a_1 y^{(n-1)}(t) + \dots + a_{n-1} y^{(1)}(t) + a_n y(t) \\ = b_0 u^{(n)}(t) + b_1 u^{(n-1)}(t) + \dots + b_{n-1} u^{(1)}(t) + b_n u(t), \end{aligned} \quad (2.22)$$

where $y^{(k)}(t)$ denotes the k^{th} derivative of $y(t)$ with respect to t .

A system is *causal* if any input u with $u(t) = 0, \forall t < 0$ produces an output y for which it holds that $y(t) = 0, \forall t < 0$.

LTI systems have an associated impulse response, step response and transfer function, each characterizing the system. For a detailed discussion on these topics the reader is referred to for example [62]. The *impulse response function* $h(t)$ is defined as the output of a system, corresponding to the *Dirac delta* $\delta(t)$ (2.15) as an input for continuous-time systems, and the *Kronecker delta* $\delta[n]$ (2.16) for discrete-time systems.

The transfer function the Laplace transform of the impulse response. It is well known that in the case of zero initial conditions the following relation holds for the Laplace transform of the input $U(s)$, the Laplace transform of the output $Y(s)$ and the *transfer function* $H(s)$ of the LTI system:

$$Y(s) = H(s)U(s). \quad (2.23)$$

The Laplace transform of $\delta(t)$ is: $\mathcal{L}\{\delta(t)\} = D(s) = 1$ so that for the Laplace transformed output it holds that:

$$Y(s) = H(s)D(s) = H(s). \quad (2.24)$$

If a causal LTI system is of finite order, it will possess a *proper rational transfer function* [62] $H(s)$:

$$H(s) = \frac{Y(s)}{U(s)} = \frac{b_0s^n + b_1s^{n-1} + \dots + b_n}{s^n + a_1s^{n-1} + \dots + a_n}. \quad (2.25)$$

The roots of the numerator polynomial in (2.25) are called the *zeros* of the system and the roots of the denominator polynomial the *poles*. The order n of the system is defined as the McMillan degree of $H(s)$. For single-input single-output (SISO) systems, the order of the system equals the degree of the denominator of the transfer function $H(s)$, after canceling all common factors. In this work linear systems will be assumed to be finite order, time-invariant and causal, unless stated otherwise.

Given a system with impulse response function $h(t)$ and input $u(t)$, the output $y(t)$ becomes the convolution of the input $u(t)$ and the impulse response $h(t)$:

$$y(t) = (h * u)(t) = \int_{-\infty}^{\infty} h(\tau)u(t - \tau)d\tau. \quad (2.26)$$

The impulse response function is the derivative of the *step response function*, and therefore this latter one can be calculated from the transfer function as:

$$\mathcal{L}^{-1}\left\{\frac{1}{s}H(s)\right\}. \quad (2.27)$$

An LTI system is said to be *asymptotically stable* if the output corresponding to an impulse input damps out to zero. Formally this comes down to the requirement that the poles of its transfer function are all in the open complex left halfplane.

LTI systems can be represented in *state-space*:

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + Du(t). \end{aligned} \quad (2.28)$$

The vector $x(t)$ is the *state vector*. The $n \times n$ matrix A is called the *system matrix*, the *dynamical matrix* or the *state matrix*, the $n \times 1$ vector B is called the *input vector*, the $1 \times n$ vector C the *output vector* and the scalar D is called the *direct feedthrough* term, alias direct current or DC-term. A system may be denoted by this quadruple of matrices $S = (A, B, C, D)$.

We usually consider systems as input/output systems, hence focusing on its external interactions. The system is identified with its transfer function. Its state-space representations is merely used as a convenient “tool”, and having meaningful states in this representation is not a requirement. To arrive at computationally convenient canonical forms and parameterizations, similarity transforms are used. Given any non-singular matrix T , a *similarity transformation* from a basis $x(t)$ to a basis $\hat{x}(t) = Tx(t)$ may be performed yielding an equivalent system:

$$\begin{aligned}\dot{\hat{x}}(t) &= \hat{A}x(t) + \hat{B}u(t) \\ y(t) &= \hat{C}x(t) + Du(t),\end{aligned}\tag{2.29}$$

where

$$\hat{A} = TAT^{-1}\tag{2.30}$$

$$\hat{B} = TB\tag{2.31}$$

$$\hat{C} = CT^{-1}\tag{2.32}$$

A *canonical form* is a “standard” form. We deal with a canonical form if for all systems that are mutually equivalent to some given system, we transform it to the *same* member of that equivalence class. As a result, in a canonical form different choices of the parameters produce systems that are never equivalent. Using similarity transformations, systems can be brought into a canonical form and from one canonical form to another.

For state-space systems the order n has the interpretation that it is the *state-space dimension* that is minimally required to represent the system.

From the state-space representation $S = (A, B, C, D)$ the transfer function can be determined:

$$H(s) = D + C(sI - A)^{-1}B.\tag{2.33}$$

The poles of the system correspond to the eigenvalues of the system matrix A . As one would expect from the above it is also possible to compute the impulse response associated with the system:

$$h(t) = Ce^{At}B + D\delta(t), \quad t \geq 0.\tag{2.34}$$

Insted of a SISO system a multi-input multi-output system (MIMO) is considered with p inputs and q outputs. The vectors B and C and the scalar D change accordingly. B then becomes an $n \times p$ *input matrix*, C a $q \times n$ *output matrix* and D a $q \times p$ direct feedthrough. All other formulas previously discussed in this section still apply.

For discrete-time systems similar relations and properties as for continuous-time systems hold. The state-space description in (2.28) takes the form:

$$\begin{aligned} x[k+1] &= Ax[k] + Bu[k] \\ y[k] &= Cx[k] + Du[k]. \end{aligned} \quad (2.35)$$

The transfer function can be obtained from the state-space representation similarly as in (2.33) as

$$H(z) = D + C(zI - A)^{-1}B, \quad (2.36)$$

and (2.34) changes to:

$$h[k] = \begin{cases} D & k = 0 \\ CA^{k-1}B & k > 0. \end{cases} \quad (2.37)$$

For a more in-depth discussion on linear systems and their properties the reader is referred to the relevant literature, for example [62, 115].

2.9 Filtering signals

It is well-known that a sinusoid $u_s(t) = \sin(\omega t)$ used as input for a stable linear system with transfer function $H(s)$ yields a steady state output $y_s(t)$ as:

$$y_s(t) = A \sin(\omega t + \phi), \quad (2.38)$$

where $A = |H(i\omega)|$ is that absolute value and $\phi = \angle H(i\omega)$ is the angle of the complex number $H(i\omega)$ in the complex plane. The function $H(i\omega)$ is called the *frequency response function* of the system and is the Fourier transform of the impulse response $h(t)$. Equation (2.38) is called the property of *sinusoidal fidelity*. This property implies that if a sinusoid is used as an input to such a system, then a sinusoid of the same frequency will constitute the output. However the output sinusoid exhibits a phase shift ϕ relative to the input sinusoid and an amplitude gain of $|H(i\omega)|$. If the phase response of the LTI system is linear with the frequency, it is said that the system has *linear phase*. In a linear phase system all frequencies have equal delay times, so that there is no phase distortion.

As discussed in Section 2.6, each signal can be decomposed in terms of sines and cosines. Consequently, if a Bode plot of the frequency response of the system is considered, one can examine how each frequency in an input $u(t)$ signal is affected by the system. How each frequency is affected by the system depends on the placement of the poles and zeros in the complex s -plane: frequencies near zeros will be suppressed and frequencies near poles will be amplified. However, if poles and zeros are sufficiently close there will be interaction between the poles and zeros. By appropriate placement of the poles different type of filters can be created such as low-pass, high-pass, band-pass and notch filters. These can have properties such as fast roll-off, no pass/stop-band ripple, etc. Classically, these filters are designed to have certain properties in the Fourier domain.

It is useful to consider filtering in the time-domain. From (2.26) it is clear that if a signal $u(t)$ is used as an input, it is convoluted with the impulse response $h(t)$ of the system, which is the inverse Laplace transform of the transfer function.

A similar property holds in discrete time. For a discrete-time signal that is a sampled version of a continuous-time signal, as discussed on page 15, meaning can be given to the sampling frequency f in Hertz. Since sensor data is mostly continuous-time data, for most applications this f is important to relate to the frequencies in the source signal. For applications where the data is inherently discrete-time one can choose $f = 1$. Given an input $u_s[k] = \sin\left(\frac{\omega k}{f}\right)$, the corresponding output $y_s[k]$ of a system with transfer function $H(z)$ is:

$$y_s[k] = A \sin\left[\frac{\omega k}{f} + \phi\right], \quad (2.39)$$

where where $A = |H(e^{i\omega/f})|$ and $\phi = \angle H(e^{i\omega/f})$. The function $H(e^{i\omega/f})$ is the discrete-time frequency response. Frequencies in input signals are thus similarly affected as in the continuous-time case.

As in (2.20) let $u(z) = u_0 + u_1 z^{-1} + u_2 z^{-2} + \dots$ be the z -transform of a discrete-time signal $u = \{u_0, u_1, u_2, \dots\}$. For discrete-time filters with a proper rational transfer function of the following form:

$$H(z) = \frac{b_0 z^n + b_1 z^{n-1} + \dots + b_n}{z^n} = b_0 + b_1 z^{-1} + \dots + b_n z^{-n}, \quad (2.40)$$

the z -transformed output corresponding to a transformed input $u(z)$ can be computed as:

$$\begin{aligned} H(z)u(z) &= b_0 u_0 + (b_0 u_1 + b_1 u_0) z^{-1} + (b_0 u_2 + b_1 u_1 + b_2 u_0) z^{-2} \\ &\quad + (b_0 u_3 + b_1 u_2 + b_2 u_1 + b_3 u_0) z^{-3} \dots \end{aligned} \quad (2.41)$$

The impulse response $u_0 = 1$, $u_k = 0$ for $k > 0$ can be easily read off from this expression yielding $\{b_0, b_1, \dots, b_n, 0, 0, \dots\}$ and will have a finite length, i.e., $h[k] = 0, \forall k > n$. This system is said to be of *finite impulse response* (FIR). Since in linear systems the input is convoluted with the impulse response, such systems exhibit a convenient one-to-one correspondence of the transformation from input to output in the time and z -domain. The output at each time instance is the weighted mean of the input around that instance, where the weighing factors are defined by the numerator polynomial of the transfer function, hence the name *moving average (MA) filters*. The filter banks associated with the discrete wavelet transform in Section 3.2 take such a form (2.40).

Historically filters are designed based on frequency domain properties. However in the case of in-band noise, such as muscle artifact noise in ECGs, the signal is not effectively separated from the noise in the frequency representation and consequently the noise cannot be removed using this approach. However the noise may have a different morphology from that of the signal and filtering techniques that take morphology into account, such as wavelets, may be beneficial.

Chapter 3

Wavelet transformations

Nowadays, a quarter of a century after Morlet and Grossman first coined the word “wavelet” and formalized the corresponding transform [44], wavelets are recognized as a fundamental and powerful tool in signal analysis, and widely applied in practice, including in the field of biomedical engineering. The remarkable rise and success of wavelets, starting from the 70s, may largely be attributed to its added value to Fourier analysis. The wavelet transform offers *both* time and frequency localization, whereas the Fourier transform only offers frequency information, making the wavelet transform a powerful tool for signal analysis [79, 107]. The windowed Fourier transform (2.13) offers the ability to obtain both time and frequency information too, but this method does not have the flexibility of the wavelet transform since the basis is restricted to sines and cosines (harmonics) and it does not possess the “zoom-in” property of wavelets as discussed on page 26. In the late 1980s Yves Meyer published work on orthogonal wavelets (e.g.[85]). This inspired Stéphane Mallat to construct a multi-resolution theory on wavelets [79], whereas Ingrid Daubechies constructed a well-known set of orthonormal wavelets [30]. Despite its recent invention, wavelet theory is already well established and employed in many standards such as the jpeg standard for image compression.

Many biomedical applications have been developed around wavelets. For example the detection of characteristic points in ECGs [105, 61, 78, 83, 102, 2], filtering [98, 112], separation of fetal cardiac activity [87] and many others. In [101] it has been shown that wavelets can be finetuned for a given application. Due to the work in [107, 92] on orthogonal filter banks a parameterization is available for scalar wavelets that can be used as a framework for wavelet design. However no optimization criterion was discussed. In Chapter 5 an optimization criterion is introduced for the design of optimal wavelets.

In this chapter we will briefly introduce some basic concepts of wavelets, and then focus in some depth on the relation between wavelets and filter banks. This chapter will close with a discussion on multiwavelets and their relation with lossless systems.

3.1 Continuous wavelets

The wavelet transformation is aimed at the decomposition of a signal, localized in both time and frequency simultaneously by inducing a change of basis for the signal in question. In order to accomplish this a so called “wavelet” function is used as a basis, and this wavelet function has the property that it is localized in both time and frequency. More formally, the wavelet transformation of a real signal $x(t)$ using a real wavelet $\psi(t)$ is defined as the cross-covariance at lag τ of that signal with the normalized dilated wavelet $\frac{1}{\sqrt{\sigma}}\psi(\frac{t}{\sigma})$ (see for example [44, 80]):

$$W(\tau, \sigma) = \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{\sigma}} \psi\left(\frac{t - \tau}{\sigma}\right) dt, \quad \tau \in \mathbb{R}, \sigma \in \mathbb{R}^+. \quad (3.1)$$

The parameter σ is a scale parameter that determines the (pseudo)frequency localization of the wavelet transform and the parameter τ determines the time localization of the wavelet transform. For a specific scale σ and a specific time τ the L_2 -inner product between the signal $x(t)$ and the normalized, time-shifted and scaled wavelet $\frac{1}{\sqrt{\sigma}}\psi(\frac{t - \tau}{\sigma})$ is thus computed. The form in (3.1) is known as the *continuous wavelet transform* (CWT) or as the *integral wavelet transform*. The continuous wavelet transform is an invertible transform as shown below:

$$x(t) = \frac{1}{C} \int_0^\infty \int_{-\infty}^\infty W(\tau, \sigma) \frac{1}{\sqrt{\sigma}} \psi\left(\frac{t - \tau}{\sigma}\right) d\tau \frac{d\sigma}{\sigma^2}, \quad t \in \mathbb{R}, C \in \mathbb{R}^+, \quad (3.2)$$

where C is a constant that depends on the wavelet $\psi(t)$ at hand. Not any arbitrary function can be used as a wavelet basis and therefore the function $\psi(t)$ must obey a so-called admissibility condition [44]. The admissibility condition comes down to requiring that $\psi(t)$ has zero average and that its Fourier transform $\Psi(\omega)$ is continuously differentiable [80]. For a more detailed discussion on this subject the reader is referred to for example [80, 26].

We denote dilated and time-shifted wavelet functions as

$$\psi_{\tau, \sigma}(t) = \frac{1}{\sqrt{\sigma}} \psi\left(\frac{t - \tau}{\sigma}\right). \quad (3.3)$$

Their corresponding Fourier transform is:

$$\Psi_{\tau, \sigma}(\omega) = e^{-i\tau\omega} \Psi(\sigma\omega). \quad (3.4)$$

From (3.4) it can be seen that unlike in (2.13) where the Heisenberg uncertainty rectangle has constant dimensions in the time-scale plane. In this case when moving to the coarser scales (i.e. lower frequencies) the window tightens in the scale dimension and widens in the time dimension. Conversely the window widens in the scale dimension and tightens in the time dimension when moving to finer scales (higher frequencies). The dimensions of the uncertainty rectangle remain constant when moving it along the time dimension. This behavior of the uncertainty rectangle facilitates the “zoom-in” property of wavelets.

Figure 3.1: The Gaussian and the Mexican Hat wavelet

In cardiac signal processing the Gaussian wavelet $\psi(t) = -2(\frac{2}{\pi})^{\frac{1}{4}}te^{-t^2}$ is very popular (see for example [5, 102]), partly due to the fact that the wave-like shape as shown in Figure 3.1 resembles the QRS complex, partly due to some of its other nice properties such as the fact that it is a derivative of a smoothing function [80]. The Gaussian wavelet is constructed by taking the normalized derivative of the Gaussian probability density function (pdf) with mean zero and variance 0.5. Higher order derivatives of the Gaussian pdf can be taken to obtain Gaussian wavelets of various orders such as the seconde derivative of the Gaussian pdf: the Mexican Hat wavelet $\psi(t) = \frac{2}{\pi^{0.25}\sqrt{3}}(t^2 - 1)e^{-0.5t^2}$ which is a very popular wavelet for ECG processing [2, 20] too. When the question is posed what wavelet to use, the answer depends heavily on the application and signals at hand. Considering the large variety of functions that obey the admissability conditions there is a good chance that one can do better than by choosing a popular and well-known wavelet.

3.2 Wavelets from filter banks

The wavelet transform can also be performed on discrete time signals, thereby still using a continuous wavelet function $\psi(t)$. One way to do this is with discrete filters from filter banks as in the *discrete wavelet transform* (DWT). In this thesis the framework of DWT wavelets from filter banks as described in for example [107] is used as a setting for wavelet design. This framework has a number of advantages and restrictions that make it a convenient setting for wavelet design. Filter banks may have, when imposed, five important properties of interest to us, which are further addressed below:

- Perfect reconstruction
- Orthogonality of the filter bank and the underlying wavelet based multi-resolution structure

Figure 3.2: Wavelet analysis and synthesis

- Flatness of the filters and vanishing moments in the wavelets
- Smoothness of the wavelets
- Linear phase

In discrete-time the input sequence is processed by two filters in parallel that split the input sequence in terms of frequency. These filters relate to the wavelet function in such a way that the downsampled output corresponds to the filtering of the signal with the wavelet function at a particular scale. The discrete-time input signal $\{x_n\}$ is essentially fed through a low-pass filter $H_0(z)$ and a high-pass filter $H_1(z)$ in parallel [38]. The filters $H_0(z)$ and $H_1(z)$ form a filter bank. The pair of filters in the filter bank are designed in such a way that after *downsampling* by a factor 2 no information is lost and there is no redundancy; they are *critically sampled*. In formulas, downsampling with a factor two will be written as $\downarrow 2$. The output of the low-pass channel is a sequence of *approximation coefficients* (alias scaling coefficients) and the output from the high-pass channel a sequence of *detail coefficients* (alias wavelet coefficients). The original input can be reconstructed by upsampling and filtering each channel with the corresponding synthesis filter $F_0(z)$ or $F_1(z)$ and adding the two channels as illustrated in Figure 3.2. The downsampling can be implemented with frequency modulation [107]:

$$a(z) = \downarrow 2 H_0(z) X(z) = \frac{1}{2} \left(H_0(z^{\frac{1}{2}}) X(z^{\frac{1}{2}}) + H_0(-z^{\frac{1}{2}}) X(-z^{\frac{1}{2}}) \right). \quad (3.5)$$

The *aliasing* term (with $-z$) cancels the odd part. This works similarly for the detail coefficients $b(z)$. To upsample $a(z)$ again take $a(z^2)$, which comes down to inserting a 0 after each element in the sequence.

Example 3.2.1. As an example consider the Haar wavelet with low-pass filter $H_0(z) = \frac{1}{2} + \frac{1}{2}z$ and high-pass filter $H_1(z) = -\frac{1}{2} + \frac{1}{2}z$. For input $x = \{4, 6, 8, 7\}$ the low-pass output will be $a = \{5, 7.5\}$ and the high-pass output $b = \{-1, 0.5\}$.

3.2.1 Perfect reconstruction

From this point on $H_0(z)$ and $H_1(z)$ are assumed to be FIR (finite impulse response) filters:

$$H_0(z) = c_0 + c_1 z^{-1} + \dots + c_N z^{-N}, \quad (3.6)$$

$$H_1(z) = d_0 + d_1 z^{-1} + \dots + d_N z^{-N}. \quad (3.7)$$

This will result in wavelets that are compactly supported [107] and makes the conditions in this paragraph easy to satisfy.

To have *perfect reconstruction* it means that if a signal is decomposed with an analysis filter bank and then reconstructed with a corresponding synthesis filter bank, the output signal is equal to the input signal except for a possible delay. To ensure perfect reconstruction there may be no aliasing and no distortion.

The aliasing effect [38] is present in both channels and must be canceled by the synthesis filters F_0 and F_1 :

$$F_0(z)H_0(-z) + F_1(z)H_1(-z) = 0. \quad (3.8)$$

To ensure that there is *no distortion* in the output of Figure 3.2, the following condition must hold:

$$F_0(z)H_0(z) + F_1(z)H_1(z) = 2z^{-Q}, \quad (3.9)$$

where Q is the overall delay of the filter bank [107].

The conditions (3.8) and (3.9) can be written in terms of a single modulation matrix:

$$\begin{pmatrix} F_0(z) & F_1(z) \\ F_0(-z) & F_1(-z) \end{pmatrix} \begin{pmatrix} H_0(z) & H_0(-z) \\ H_1(z) & H_1(-z) \end{pmatrix} = \begin{pmatrix} 2z^{-Q} & 0 \\ 0 & 2z^{-Q} \end{pmatrix}. \quad (3.10)$$

3.2.2 Orthogonal filter banks

For orthogonal filter banks it holds that the impulse responses of the synthesis filters are the time reverses of the impulse responses of the analysis filters [106] as expressed by (3.13). To see this, we start by observing that the two filters in the filter bank $H_0(z)$ and $H_1(z)$ form a power complementary set [117, page 183]. Since on the unit circle the complex conjugate of z is z^{-1} the orthogonality condition from [117], taking the differences in notation into account, comes down to:

$$H_0(z^{-1})H_0(z) + H_1(z^{-1})H_1(z) = c, \quad (3.11)$$

where $c = 2$ for consistency with (3.9), that is, to ensure conservation of energy.

In this orthogonal framework the condition (3.8) can be enforced by constructing the synthesis filters from the analysis filters by *alternating signs* [31]:

$$F_0(z) = H_1(-z) \text{ and } F_1(z) = -H_0(-z). \quad (3.12)$$

To satisfy the condition in (3.9) the power complementary property can be used [117]:

$$F_0(z) = z^{-N} H_0(z^{-1}) \text{ and } F_1(z) = z^{-N} H_1(z^{-1}), \quad (3.13)$$

with the result that for orthogonal filters it holds that in (3.9) $Q = N$.

To satisfy both (3.12) and (3.13) for an N^{th} order filter one can construct the high-pass filter from the low-pass filter by the *alternating flip construction* [31]. From the left side of (3.12) and (3.13) it is obtained that:

$$H_1(-z) = z^{-N} H_0(z^{-1}). \quad (3.14)$$

After substituting $z \rightarrow -z$ it is obtained that:

$$H_1(z) = (-z)^{-N} H_0(-z^{-1}). \quad (3.15)$$

And from the right side of (3.12) and (3.13) it is obtained that:

$$H_1(z^{-1}) = z^N F_1(z) = -z^N H_0(-z). \quad (3.16)$$

After substituting $z \rightarrow z^{-1}$ it is obtained that:

$$H_1(z) = -z^{-N} H_0(-z^{-1}). \quad (3.17)$$

From (3.15) and (3.17) it follows that N must be odd yielding the alternating flip construction:

$$H_1(z) = (-z)^{-N} H_0(-z^{-1}), \quad \text{for } N = 2n - 1. \quad (3.18)$$

In terms of (3.6) and (3.7) with $N = 2n - 1$ it follows that:

$$d_k = (-1)^k c_{N-k} \quad (k = 0, 1, \dots, N). \quad (3.19)$$

Filters that are constructed in this manner are called *quadrature mirror filters* [38, 86, 30].

For orthogonality the following constraints hold for the scaling filter coefficients c_k and the wavelet filter coefficients d_k of $H_0(z)$ and $H_1(z)$ respectively:

Normalization $\sum_{k=0}^N c_k^2 = 1$ and $\sum_{k=0}^N d_k^2 = 1$

Double-shift orthogonality $\sum_{k=0}^N c_k c_{k-2\ell} = 0$ and $\sum_{k=0}^N d_k d_{k-2\ell} = 0 \quad \forall \ell \in \mathbb{Z} \setminus \{0\}$

Double-shift orthogonality between the filters $\sum_{k=0}^N c_k d_{k-2\ell} = 0 \quad \forall \ell \in \mathbb{Z}$

In the last two conditions negatively indexed coefficients are all zero by convention (c_0, d_k for $k < 0$ or $k > N$).

Figure 3.3: Wavelet analysis tree structure

3.2.3 Multi-resolution analysis

The frequency localization for the discrete wavelet transform comes from using dilated basis functions, known as the *scaling* and *wavelet function*, as will be discussed in Section 3.2.4. However this multi-resolution analysis can be implemented by a cascade of filter banks, consisting of an orthogonal pair of low- and high-pass filters. The output of the low-pass filter is used as the input of the next level as in Mallat's algorithm [79]. This procedure is illustrated in Figure 3.3 and gives rise to *multi-resolution analysis* (MRA). Note that we now have approximation and detail coefficients at various levels j that correspond to dyadic scales 2^j .

The basic principle is as follows: One starts with a signal x and feeds it through an analysis filter bank. The output of the low-pass filter is then $a^{(1)} = \downarrow 2 H_0 x$ and that of the high-pass filter is $b^{(1)} = \downarrow 2 H_1 x$. Next $a^{(1)}$ is used as the input of the filter bank to yield $a^{(2)} = \downarrow 2 H_0 a^{(1)}$ and $b^{(2)} = \downarrow 2 H_1 a^{(1)}$. Then again repeating this procedure a number of times, the output of the low-pass filter is used so that one eventually ends up with $\{b^{(1)}, b^{(2)}, \dots, b^{(L)}, a^{(L)}\}$. The signal x can be reconstructed by a cascade of synthesis filter banks. The length of the signal x and the size of the filter N determines how many levels this cascade can have. If the maximum number of levels L has been reached, the signal is said to be fully *decomposed*.

3.2.4 Wavelet and scaling functions

In this section it will be discussed how filter banks can be interpreted using the wavelet paradigm. For an excellent exposition on this topic the reader is referred to [107].

Consider the following multi-resolution structure for the L_2 -space, consisting of a nested sequence of linear subspaces V^ℓ , called *approximation spaces*:

$$\dots \subset V^1 \subset V^0 \subset V^{-1} \subset \dots, \quad (3.20)$$

with $\bigcap_{\ell \in \mathbb{Z}} V_\ell = \{0\}$ and $\overline{\bigcup_{\ell \in \mathbb{Z}} V_\ell} = L^2(\mathbb{R})$, where the bar indicates closure. Assume that there exists a function $\varphi(t)$, called the *scaling function*, that spans the subspace V^0 .

This scaling function generates a shift invariant orthonormal basis $\{\varphi(t - k) | k \in \mathbb{Z}\}$ of V^0 . From V^0 the orthonormal bases are induced for all spaces V^ℓ [80]:

$$\begin{aligned}\varphi_k^{(\ell)}(t) &= \frac{1}{\sqrt{2^\ell}} \varphi\left(\frac{t - 2^\ell k}{2^\ell}\right) = 2^{-\frac{\ell}{2}} \varphi(2^{-\ell}t - k) \\ V^\ell &= \text{span}\{\varphi_k^{(\ell)} | k \in \mathbb{Z}\}.\end{aligned}\tag{3.21}$$

Furthermore, consider the subspaces W^ℓ called *detail spaces*, that are the orthogonal complements of the approximation spaces V^ℓ :

$$V^\ell \oplus W^\ell = V^{\ell-1}.\tag{3.22}$$

Assume that the approximation space W^0 is spanned by a the *wavelet function* $\psi(t)$, and its dilated versions are shift-invariant orthonormal bases for the spaces W^ℓ :

$$\begin{aligned}\psi_k^{(\ell)}(t) &= \frac{1}{\sqrt{2^\ell}} \psi\left(\frac{t - 2^\ell k}{2^\ell}\right) = 2^{-\frac{\ell}{2}} \psi(2^{-\ell}t - k) \\ W^\ell &= \text{span}\{\psi_k^{(\ell)} | k \in \mathbb{Z}\}.\end{aligned}\tag{3.23}$$

From (3.20) and (3.21), it follows that there exists a sequence of coefficients $\{c_k\}$ such that the scaling function can be written as a linear combination of dilated versions of itself in the *dilation equation*:

$$\varphi(t) = \sqrt{2} \sum_{k=0}^N c_k \varphi(2t - k).\tag{3.24}$$

Likewise from (3.20), (3.22) and (3.23), it follows that there exists a sequence of coefficients $\{d_k\}$ such that the wavelet function can be written as a linear combination of dilated versions the scaling function in the *wavelet equation*:

$$\psi(t) = \sqrt{2} \sum_{k=0}^N d_k \varphi(2t - k).\tag{3.25}$$

The coefficients c_k and d_k in (3.24) and (3.25) are the same coefficients as in Section 3.2.2, with all properties involved, linking the filter coefficients to the wavelet and scaling function. This allows us to reverse the approach and given coefficients c_k and d_k try to find a scaling function $\varphi(t)$ and a wavelet function $\psi(t)$. An iteration scheme which allows for the exact computation of $\varphi(t)$ at all the dyadic points up to an arbitrary resolution can be found, for instance, in [107, section 6.1]. We have adopted this method in this work. It is important to note that it may well happen that the resulting sequence exhibits a discontinuous and fractal structure and may not converge to an actual function, i.e., $\varphi(t)$ and $\psi(t)$ may not exist for a given set of filters. The continuity of the wavelet and scaling functions is further investigated in terms of the “joint spectral radius” in [110, Section 2.2].

Now let $f(t)$ be a function, corresponding to a sequence $\{x_k\}$ of wavelet coefficients, that can be expressed in terms of the wavelet basis for V^0 as:

$$f(t) = \sum_k x_k \varphi(t - k), \quad (3.26)$$

then the coefficients resulting from the DWT give an expansion for $f(t)$ in terms of the orthonormal functions $\varphi_k^{(\ell)}(t)$ and $\psi_k^{(\ell)}(t)$:

$$f(t) = \sum_k a_k^{(L)} \varphi_k^{(L)}(t) + \sum_{\ell=1}^L \sum_k b_k^{(\ell)} \psi_k^{(\ell)}(t), \quad (3.27)$$

where L is no more than the maximum level as in Section 3.2.3. For an orthogonal wavelet structure it then also holds [91] that:

$$a_k^{(\ell)} = \int_{-\infty}^{\infty} \varphi_k^{(\ell)}(t) f(t) dt, \quad (3.28)$$

$$b_k^{(\ell)} = \int_{-\infty}^{\infty} \psi_k^{(\ell)}(t) f(t) dt, \quad (3.29)$$

which connects the continuous wavelet transform in (3.1) and discrete wavelet transform (3.29). Note that with the discrete wavelet transform these expansion coefficients $\{a_k^{(\ell)}\}$ and $\{b_k^{(\ell)}\}$ can be calculated without the explicit use of the scaling and wavelet function, but only by making use of a filter bank, using the approach in Section 3.2.3.

3.2.5 Vanishing moments

For some applications it is beneficial that the wavelets have vanishing moments. If a wavelet has p vanishing moments then the wavelet function $\psi(t)$ is orthogonal to polynomials of up to degree $p - 1$ [80]. This leads to the following equation, which is familiar for moments from physics and statistics:

$$\int_{-\infty}^{\infty} t^k \psi(t) dt = 0, \quad \text{for } 0 \leq k < p. \quad (3.30)$$

Due to the orthogonality between the wavelet and scaling function it follows that if a wavelet has p vanishing moments then the space $V^0 = \text{span}\{\varphi_k | k \in \mathbb{Z}\}$ spanned by the scaling function contains all the polynomials of degree less than p [107, Section 7.1]. As familiar from Taylor series, smooth functions can locally be approximated by polynomials. If a wavelet has p vanishing moments, then the detail signal will contain no energy of polynomials up to degree $p - 1$. If a signal $f(t)$ can locally in neighborhood v be represented as a polynomial part $f_{v,p}(t)$ with degree less than p and some error term $\varepsilon_v(t)$ then the wavelet transform of f will be [80, Section 6.1]:

$$W_{f_v}(\tau, \sigma) = \int_{-\infty}^{\infty} f_v(t) \frac{1}{\sqrt{\sigma}} \psi\left(\frac{t - \tau}{\sigma}\right) dt = \int_{-\infty}^{\infty} \varepsilon_v(t) \frac{1}{\sqrt{\sigma}} \psi\left(\frac{t - \tau}{\sigma}\right) dt = W_{\varepsilon_v}(\tau, \sigma). \quad (3.31)$$

When attempting to locally approximate the signal f with polynomials of degree less than p , ϵ represents the approximation error. The approximation error will have a non-zero contribution to the detail coefficients. This property can be used to measure the Lipschitz regularity (also known as the Hölder exponent) of a signal [80, Section 6.1]. The approximation error decays with the level j (scale 2^j) and the number of vanishing moments as $O(2^{jp})$ [107, Section 7.1].

For discrete time signals and wavelets, vanishing moments have to be imposed on the impulse response of $H_1(z)$, that is, moments up to order $p-1$ have to be imposed on the wavelet filter. To have one vanishing moment amounts to the condition $d_0 + d_1 + \dots + d_N = 0$ which is equivalent to the commonly imposed condition that the integral of the mother wavelet $\psi(t)$ is equal to zero: $\int_{-\infty}^{\infty} \psi(t) dt = 0$. In Section 3.2.4 it is discussed how the wavelet function is obtained from the filter coefficients. For a wavelet filter to have p vanishing moments the constraints on the high-pass coefficients are:

$$\sum_{\ell=0}^N \ell^k d_\ell = 0, \quad \text{for } 0 \leq k < p. \quad (3.32)$$

The conditions on the wavelet filter (3.32) to have p vanishing moments can be translated to the scaling and wavelet function using the dilation equation (3.24) and the wavelet equation (3.25). From this it follows that the conditions on the filter are a linear combination of the conditions on the wavelet function (3.30) to have p vanishing moments, vice versa.

For a wavelet filter to have vanishing moments, it is required that $H_0(z)$ and $H_1(z)$ have zeros at $z = -1$ and $z = 1$ respectively [80, Theorem 7.4][107, Theorem 7.1]. For filters this corresponds to requiring zeros at respectively π radians (Nyquist frequency) for the scaling filter and 0 radians for the wavelet filter. The zeros for the scaling filter at π radians make it a low-pass filter with a corresponding degree of flatness and the wavelet filter a high-pass filter with a certain degree of flatness. The Daubechies orthogonal wavelet family is constructed by using all freedom to impose vanishing moments.

3.2.6 Linear phase

As discussed on page 22, linear phase helps to avoid phase distortion. There only exists a single orthogonal wavelet with linear phase: the Haar wavelet (see e.g. [120]). In order to obtain both orthogonality and linear phase simultaneously, the theory of multiwavelets (see Section 3.4) can be employed.

3.2.7 Polyphase filtering

In the traditional discrete wavelet transform approach, the high- and low-pass filters operate at full rate, after which the outputs are downsampled. In other words: half of the processed information is discarded. The use of *polyphase filters* avoids this, by first splitting the input into two or more (M) phases, then applying a polyphase filter to each phase separately and finally combining the polyphase outputs to generate the intended output.

Figure 3.4: Polyphase low-pass filter

The polyphase filters operate at reduced rate $\frac{1}{M}$. Suppose an input signal $X(z)$ and an analysis low-pass filter $H_0(z)$. The goal is to obtain $\downarrow 2H_0X$; the downsampled result of filtering $X(z)$ with H_0 . To implement this with polyphase filters, with e.g., $M = 2$, take $H_{0,e}(z) = c_0 + c_2z^{-1} + \dots + c_{2n-2}z^{-(n-1)}$ and $H_{0,o}(z) = c_1 + c_3z^{-1} + \dots + c_{2n-1}z^{-(n-1)}$ as respectively the even and odd phase of the filter $H_0(z)$, and $X_e(z)$ and $X_o(z)$ as respectively the even and odd phase of the input signal $X(z)$. (Since $M = 2$ it is convenient to call one of the phases “even” and the other one “odd”.) Then the desired output $\downarrow 2H_0x$ can be calculated [107, Section 4.2] with polyphase filters as:

$$\begin{aligned}
 \mathcal{Z} \{ \downarrow 2H_0x \} &= (H_0(z)X(z))_e \\
 &= X_e(z)H_{0,e}(z) + z^{-1}H_{0,o}(z)X_o(z) \\
 &= \begin{bmatrix} H_{0,e}(z) & H_{0,o}(z) \end{bmatrix} \begin{bmatrix} X_e(z) \\ z^{-1}X_o(z) \end{bmatrix},
 \end{aligned} \tag{3.33}$$

This process is illustrated in Figure 3.4. The basic idea behind this approach to find the even part, magister dixit, “So it is really odd *plus* odd, and even *plus* even” [107, p. 115].

Now consider a filter bank with low-pass filter H_0 and high-pass filter H_1 . These filters can be split into two phases and a *polyphase matrix* $H_p(z)$ can be constructed as:

$$H_p(z) = \begin{bmatrix} H_{0,e}(z) & H_{0,o}(z) \\ H_{1,e}(z) & H_{1,o}(z) \end{bmatrix}, \tag{3.34}$$

Then this polyphase matrix $H_p(z)$ can be used to implement the low- and high-pass filter in parallel by acting on the same polyphase input:

$$\begin{bmatrix} V_0(z) \\ V_1(z) \end{bmatrix} = \begin{bmatrix} (H_0(z)X(z))_e \\ (H_1(z)X(z))_e \end{bmatrix} = \begin{bmatrix} H_{0,e}(z) & H_{0,o}(z) \\ H_{1,e}(z) & H_{1,o}(z) \end{bmatrix} \begin{bmatrix} X_e(z) \\ z^{-1}X_o(z) \end{bmatrix}. \tag{3.35}$$

For reconstruction the synthesis filter can be implemented as a polyphase filter $F_p(z)$

Figure 3.5: Polyphase analysis and synthesis filters

too:

$$F_p(z) = \begin{bmatrix} F_{0,o}(z) & F_{1,o}(z) \\ F_{0,e}(z) & F_{1,e}(z) \end{bmatrix} \quad (3.36)$$

$$\hat{X}(z) = \begin{bmatrix} z^{-1} & 1 \end{bmatrix} F_p(z^2) \begin{bmatrix} V_0(z) \\ V_1(z) \end{bmatrix}. \quad (3.37)$$

In the polyphase approach the analysis and synthesis filters are adjacent, in contrast to the filterbank approach where they are separated by down- and upsampling. As a result the condition for perfect reconstruction becomes elegant and simple for polyphase filters:

$$F_p(z)H_p(z) = z^{-q}I. \quad (3.38)$$

For practical purposes, the delay q should be as small as possible. The overall delay induced by the filter system due to the fact that the filters are causal is $Q = 2q + 1$, so that $\hat{X}(z) = z^{-Q}X(z)$. In the case of orthogonality as in Section 3.2.2 it holds that $Q = N = 2n - 1$. The elegant condition in (3.38) facilitates the parameterization of wavelet filters. Due to the fact that the polyphase filters operate at half rate this approach is also beneficial from a computational viewpoint.

3.3 The stationary wavelet transform

The multi-resolution approach discussed in Section 3.2.3 is well suited for compression purposes since it is *critically sampled*, i.e., all resulting wavelet coefficients are required to ensure perfect reconstruction without redundancy. If a sequence x of length 2^m is filtered with wavelet filters H_0 and H_1 , then x is mapped to $\downarrow 2H_0x$ and $\downarrow 2H_1x$ by means of a projection. Both $\downarrow 2H_0x$ and $\downarrow 2H_1x$ will then have a length of approximately 2^{m-1} , effectively reducing the resolution by a factor two. After rewriting the upper right-hand side of (3.23) as

$$\begin{aligned} \psi_k^{(\ell)}(t) &= 2^{-\frac{\ell}{2}} \psi(2^{-\ell}(t - 2^\ell k)) \\ &= 2^{-\frac{\ell}{2}} \psi(2^{-\ell}(t - \tau)), \quad \tau = 2^\ell k, k \in \mathbb{Z}, \end{aligned} \quad (3.39)$$

one can see that the timeshifts of the dilated wavelet functions $2^{-\frac{\ell}{2}} \psi(2^{-\ell}t)$ are integer multiples of 2^ℓ with an initial offset of zero.

When employing the wavelet transform for detection purposes, this approach has two important drawbacks, in casu the lack of shift invariance and the loss of resolution at coarse scales. These problems can be avoided by using an *overcomplete wavelet transform*. The so-called *stationary wavelet transform* [91] achieves this by taking timeshifts equal to all integer values instead of 2^ℓ as in (3.39).

From (3.29) the connection between the continuous and discrete wavelet transform becomes clear. If (3.29) is applied at an appropriate resolution at each scale, one obtains a transform with the same resolution as the original signal x of length $n = 2^m$. The regular DWT uses downsampling at each successive scale, which causes loss of resolution at those scales. One way to avoid this is to perform the DWT 2^L times (L being the number of scales in the MRA) and each time shifting the sequence x by one.

The results can then be combined to have full resolution at each scale; all the shifted wavelet transforms combined give a shift invariant, redundant transform. A major drawback of this approach is that it is very inefficient, since such a large number of shifts are not required at fine scales.

Another approach is to split the wavelet decomposition at each scale. Let $a^{(\ell-1)}$ be the input to scale ℓ . Then one can perform a wavelet transform on the approximation signal at the preceding level $a^{(\ell-1)}$ and on its time-shifted version $z^{-1}a^{(\ell-1)}$. This results in a tree that branches out exponentially fast compared to the the decomposition tree in the regular discrete wavelet transform

$$\begin{aligned} a^{(\ell),0} &= \downarrow 2 H_0 a^{(\ell-1)} \\ b^{(\ell),0} &= \downarrow 2 H_1 a^{(\ell-1)} \\ a^{(\ell),1} &= \downarrow 2 H_0 z^{-1} a^{(\ell-1)} \\ b^{(\ell),1} &= \downarrow 2 H_1 z^{-1} a^{(\ell-1)}. \end{aligned} \tag{3.40}$$

This process is also illustrated in Figure 3.6. Due to recursion we get

$$a^{(\ell),[0,S]} = \downarrow 2 H_0 a^{(\ell-1),S}, \tag{3.41}$$

where S is a bit string that encodes through what downsampling types (even or odd) the signal has passed, with the most recent type up front. A 0 in S indicates that the input passed through a downsampling $\downarrow 2$, i.e., the regular detail and approximation coefficients are involved, and an 1 indicates that it has passed through a delay followed by downsampling $\downarrow 2z^{-1}$, i.e., the detail and approximation coefficients of the shifted signal are involved.

At scale k (scale $k = 0$ corresponds to the signal) there are thus 2^k sets of approximation and detail coefficients, with each a associated bit string. The sets of detail coefficients at each scale have to be interlaced such that a single set of coefficients is

Figure 3.6: Stationary wavelet transform

acquired at full resolution. For example at level $l + 1$ we have:

$$b_{\text{even}}^{(l+1),[0,S]} = b^{(l+1),[0,0,S]}, \quad (3.42)$$

$$b_{\text{odd}}^{(l+1),[0,S]} = b^{(l+1),[1,0,S]}, \quad (3.43)$$

$$b_{\text{even}}^{(l+1),[1,S]} = b^{(l+1),[0,1,S]}, \quad (3.44)$$

$$b_{\text{odd}}^{(l+1),[1,S]} = b^{(l+1),[1,1,S]}, \quad (3.45)$$

$$b_{\text{even}}^{(l+1),S} = b^{(l+1),[0,S]}, \quad (3.46)$$

$$b_{\text{odd}}^{(l+1),S} = b^{(l+1),[1,S]}. \quad (3.47)$$

This process is displayed in Figure 3.7. For the maximum level L the sets of approximation coefficients also have to be interlaced, likewise to obtain a wavelet decomposition with full resolution at each dyadic scale.

The stationary wavelet transform can also be calculated in polyphase. The stationary wavelet transform in polyphase representation is partially displayed in Figure 3.8, where H_p is a polyphase filter with two inputs (the phases) and two outputs.

Another approach to calculate the stationary wavelet transform is to modify the filters. Such an approach, which gives the same results, is discussed in [91].

Figure 3.7: Interlacing for the stationary wavelet transform

Figure 3.8: Stationary wavelet transform in polyphase representation

3.4 Multiwavelets

Regular scalar wavelets have a number of limitations. For example both linear phase and orthogonality is only possible for the Haar wavelet as discussed in Section 3.2.6. Another problem is that for a given choice of basis function this basis function may correlate with the features in a signal in such manner, that it is impossible to distinguish these features in the wavelet domain. *Multiwavelets* are a generalization of wavelets in the sense that instead of that V^0 is spanned in the L_2 space by a basis generated (through integer shifts) from a single scalar function $\varphi(t)$, it is spanned by a vector (multiscaling) function $\underline{\varphi}(t)$ [119, 76, 77, 107], which is a vector of r scaling functions $\underline{\varphi}(t) = [\varphi_{[0]}(t), \dots, \varphi_{[r-1]}(t)]^T$. The dilation equation is a vector generalization of (3.24):

$$\underline{\varphi}(t) = \sqrt{2} \sum_{k=0}^N C_k \underline{\varphi}(2t - k), \quad (3.48)$$

with $H_0(z) = C_0 + C_1 z^{-1} + \dots + C_N z^{-N}$ where each C_k is an $r \times r$ coefficient matrix. Likewise a multiwavelet function, which is a vector function $\underline{\psi}(t)$ is introduced as:

$$\underline{\psi}(t) = \sqrt{2} \sum_{k=0}^N D_k \underline{\varphi}(2t - k). \quad (3.49)$$

The entries of $\underline{\psi}(t)$ (with their integer translates) constitute a basis for W^0 with the multiwavelet filter $H_1(z) = D_0 + D_1 z^{-1} + \dots + D_N z^{-N}$. The Smith-Barnwell orthogonality conditions can be imposed as:

$$H_0(z)H_0(z^{-1})^T + H_0(-z)H_0(-z^{-1})^T = 2I_r, \quad (3.50)$$

$$H_1(z)H_1(z^{-1})^T + H_1(-z)H_1(-z^{-1})^T = 2I_r, \quad (3.51)$$

$$H_0(z)H_1(z^{-1})^T + H_0(-z)H_1(-z^{-1})^T = 0, \quad (3.52)$$

generalizing (3.10). The vectorfunctions $\underline{\varphi}(t)$ and $\underline{\psi}(t)$ both have compact support in the interval $[0, N]$ due to the assumed FIR property of the related filters. They generate a multiresolution structure for the inner-product space $L^2(\mathbb{R})$: $\dots, V_{-1}, V_0, V_1, \dots$, with $\bigcap_{\ell \in \mathbb{Z}} V_\ell = \{0\}$ and $\overline{\bigcup_{k \in \mathbb{Z}} V_k} = L^2(\mathbb{R})$, analogous to the regular DWT case. Note however that all inputs and outputs are vector sequences with r components. The input is now recomposed of the vector sequence containing the r phases.

Multiwavelets can also be implemented with polyphase filters. The first step is to split the input signal $X(z)$, the low-pass filters $H_0(z)$ and the high-pass filters $H_1(z)$ into $2r$ phases.

$$X(z) = X_e(z^2) + z^{-1}X_o(z^2) \quad (3.53)$$

$$H_0(z) = H_{0,e}(z^2) + z^{-1}H_{0,o}(z^2) \quad (3.54)$$

$$H_1(z) = H_{1,e}(z^2) + z^{-1}H_{1,o}(z^2). \quad (3.55)$$

Figure 3.9: Multiwavelet with $r = 3$ in polyphase representation

Next one constructs the $2r \times 2r$ polyphase FIR filter $H_p(z)$:

$$H_p(z) = \begin{pmatrix} H_{0,e}(z) & H_{0,o}(z) \\ H_{1,e}(z) & H_{1,o}(z) \end{pmatrix}. \quad (3.56)$$

Then the polyphase filter $H_p(z)$ is applied, resulting in r vectors with detail coefficients $b_k^{(\ell,m)}$ (ℓ being the scale and $m = 0, \dots, r-1$) and r vectors with approximation coefficients $a_k^{(\ell,m)}$. The approximation coefficients are then split into two phases and iterated through the polyphase filter $H_p(z)$ as illustrated in Figure 3.9. We have:

$$Y_k(z) = \sum_{m=0}^{2r-1} H_{km}(z) U_m(z), \quad (3.57)$$

where $H_{km}(z)$ is the transfer function from the m^{th} input to the k^{th} output. And this can be rewritten in vector form as $Y(z) = H_p(z)U(z)$. The orthogonality conditions translate in terms of the polyphase representation into:

$$H_p(z)H_p(z^{-1})^T = I_{2r}, \quad (3.58)$$

where I_{2r} is the $2r \times 2r$ identity matrix.

Lossless systems [116, 117] or *stable all-pass systems* are systems that retain the energy from the (Fourier transformable) input U to the output Y (with a possible scale

factor):

$$\frac{1}{2\pi} \int_0^{2\pi} |Y(e^{i\omega})|^2 d\omega = \frac{c}{2\pi} \int_0^{2\pi} |U(e^{i\omega})|^2 d\omega, \quad c \in \mathbb{R}^+. \quad (3.59)$$

It holds [117] for the transfer matrix of a stable all-pass system that:

$$H(z)^\dagger H(z) = cI, \quad \forall |z| = 1, c \in \mathbb{R}^+, \quad (3.60)$$

where † indicates the *Hermitian transpose* or *conjugate transpose* which is defined as transposition followed by complex conjugation: $(A^\dagger)_{a,b} = \overline{A_{b,a}}$. In order to build in the condition (3.58) the constant c is chosen to be $c = 1$. Note that wavelets filters are clearly stable since they are FIR and that they are all-pass. As a result it additionally holds that:

$$H(z)^\dagger H(z) \leq I, \quad |z| > 1, \quad (3.61)$$

$$H(z)^\dagger H(z) \geq I, \quad |z| < 1. \quad (3.62)$$

Similarly from the properties of lossless systems in [117]:

$$\tilde{H}(z)H(z) = I_s, \quad z = e^{i\omega}, \quad (3.63)$$

$$\tilde{H}(z) = H(z^{-1})_*^T, \quad (3.64)$$

where the subscript $H(z)_*$ stands for conjugation of the coefficients of $H(z)$, i.e., $H(z)$ has to be unitary (orthogonal for real matrices) on $|z| = 1$. Since we have real coefficients in $H_p(z)$ in (3.58) and because (3.64) extends to the whole complex plane due to analytic continuation, (3.58) is equivalent to the property of lossless systems in (3.64). Hence the condition of orthonormality for multiwavelets comes down to requiring that $H_p(z)$ is lossless. A real lossless polyphase matrix $H_p(z)$ of order $n - 1$ thus corresponds with a pair of polynomial matrices $H_0(z)$ and $H_1(z)$ that form an orthogonal FIR (multi)wavelet filterbank of order $2n - 1$. An interpolation condition on the unit circle for the lossless system can be used to impose a vanishing moment for the wavelets as will be further discussed in Section 5.3. This is useful for parameterization purposes.

Chapter 4

Continuous-time analog implementation of wavelets

In implantable medical devices such as pacemakers *power consumption* is a critical issue because battery lifetime is limited. This especially holds for sensing circuits since they are permanently active. To perform discrete-time digital signal processing, an analog to digital (A/D) converter and sampling (continuous time to discrete time) is required in order to transfer continuous-time analog sensor information to the digital domain. Depending on the number of bits used, the A/D conversion is a heavy power consuming operation. For power consumption considerations it therefore is preferable to perform as many computations as possible in the *continuous-time analog domain*. In earlier work [53, 50, 49, 54, 48] analog dynamic translinear systems (see Section 4.1) were used as a platform to implement continuous wavelet transforms (see Section 3.1). This involved the approximation of the wavelet transform with the impulse response of a linear system as discussed in Section 4.2. Using the technique of Padé approximation the authors of [53] obtained a rational approximation of the Laplace transform of the wavelet function under approximation. In Section 4.3 this approach, that only can approximate a limited number of wavelet functions, is discussed in more detail. A key problem of this Padé approach is that it does not behave equally well for each time instance. An new alternative approach, based on L_2 approximation, that can be used for a wider range of wavelet functions is discussed in Sections 4.4 and 4.5. This approach can even be employed to approximate the wavelet function of discrete wavelets such as the Daubechies 3 wavelet as demonstrated in Section 4.5.

4.1 Dynamic translinear systems

The value of an input or output signal is commonly represented in circuits in the voltage domain (volts). However *Dynamic Translinear (DTL) circuits* [90, 37] are current-based

Figure 4.1: Linear log-domain lossy integrator. Note that Q_3 is not fundamental for the operation.

circuits (amps). As a result capacitors and transistors are used in the DTL circuit, but no resistors. The voltage-to-current transformation corresponds to an exponential transformation [90] if, for example, bipolar transistors are used or CMOS transistors in the region where the Voltage density is small, i.e., in the weak inversion region.

In Figure 4.1 it is illustrated how an integrator can be implemented in a DTL circuit. It is commonly said that the integrator is implemented in the “log-domain”, referring to the logarithmic relation between the internal state variables and capacitance voltages and input/output voltages, since the current-to-voltage transformation corresponds to a logarithmic transform and the voltage-to-current transformation to an exponential transform. The nonlinear integrator is the log-domain filter in this figure [37]. The outputs are related linearly to currents. The integrator block is nonlinear, however the overall scheme is linearized by transforming the input/output voltages to overall input/output currents respectively. It is well known that for the product of two exponentials it holds that $\exp a \exp b = \exp a + b$ and therefore this logarithmic relationship in DTL

circuits makes it possible to implement multipliers. Furthermore the derivative of an exponential function is equal to the exponential function times the derivative of the exponent. Not only linear systems can be implemented with the DTL approach but also some non-linear operations, such as nonlinear differential equations. Four advantages of DTL circuits of interest are:

1. The *current based* scheme is potentially less power-consuming than a voltage based scheme.
2. The absence of resistors in the physical implementation allows a smaller circuit.
3. Both linear systems as various non-linear operations (such as an RMS D/C converter, an oscillator with limit cycle or multipliers) can be implemented with DTL circuits.
4. An attractive potential of the DTL technique are the *dynamic possibilities*: The behavior of the circuit can be changed almost instantaneously by changing the magnitude of the currents that determine the implemented state-space system. This can for example be used to scale the system such that a wavelet can be dilated in order to adapt to the specific frequency that is relevant for a sense amplifier in a given state.

In [54] a DTL approach is described which aims to implement the Gaussian wavelet transform. This Gaussian wavelet transform is of interest for ECG processing (see e.g. [5, 102]). The implementation is done in the analog domain for the purpose of cardiac signal analysis. In Figure 4.2 a DTL implementation of a state-space system is displayed. The performance of such an implementation depends largely on the accuracy of the approximations involved in this approach. From a technological point of view, the quality of the hardware components used in the manufacturing process may have a considerable impact on the performance of the IC, but such issues will not be discussed here. From a conceptual point of view, one of the critical steps concerns the *approximation of the Laplace transform of the (time-reversed and shifted) Gaussian wavelet function by means of a strictly proper rational function of low order*. For this purpose the classical technique of Padé approximation [10, 19] was previously proposed. We will be discussing the drawbacks of this approach and propose an alternative approach: L_2 approximation of wavelets, that overcomes the problems discussed for the Padé approximation.

4.2 Wavelet transformations as linear systems

The available IC design methods only allow for a limited class (e.g. finite order and causal) of linear filters to be implemented. The IC design of linear filters (Section 2.8) of finite order is quite well understood. If a time signal $f(t)$ is passed through a linear system, then $f(t)$ is *convoluted with the impulse response* $h(t)$ of that linear system, producing the output signal as in (2.26). If the continuous wavelet transform $W(\tau, \sigma)$ of $f(t)$ associated with a given mother wavelet $\psi(t)$ on a scale σ is considered, this

Figure 4.2: Implementation of a state-space system with a DTL circuit. Picture at the courtesy of Sandro Haddad. Note that in this specific case I_{C10} is negligible small [47, p. 181].

transform is obtained as the integral defined in (3.1). Note that the wavelet transform at a fixed scale σ involves a *linear* filter operation. Therefore, the analog computation of $W(\tau, \sigma)$ can be achieved through the implementation of a linear filter of which the impulse response satisfies

$$h(t) = \frac{1}{\sqrt{\sigma}} \psi\left(\frac{-t}{\sigma}\right). \quad (4.1)$$

For linear systems of a finite order this equation can in general not be satisfied exactly, however a reasonably good approximation may be sufficient for the intended application. Equation (4.1) can also be reformulated in the Laplace domain:

$$H(s) = \int_0^\infty \frac{1}{\sqrt{\sigma}} \psi\left(\frac{-t}{\sigma}\right) e^{-st} dt. \quad (4.2)$$

For obvious physical reasons only the hardware implementation of *strictly causal stable filters* of sufficiently low order is feasible. In other words, an implementable linear filter will have a (strictly) proper rational transfer function $H(s)$ that has all its poles in the left half of the complex plane to ensure stability. In a causal system the degree of the numerator is less than or equal to the degree of the denominator, yielding a direct feed-

order	norm	shift 2.0	shift 2.5	shift 3.0	shift 3.5
3	l_1 -norm	0.997	1.466	1.871	1.980
5	l_1 -norm	0.117	0.265	0.496	0.789
7	l_1 -norm	0.017	0.027	0.071	0.153
9	l_1 -norm	0.016	0.002	0.007	0.021
3	l_2 -norm	0.397	0.551	0.678	0.696
5	l_2 -norm	0.048	0.101	0.178	0.269
7	l_2 -norm	0.007	0.010	0.025	0.053
9	l_2 -norm	0.010	0.001	0.003	0.007
3	l_∞ -norm	0.490	0.539	0.485	0.430
5	l_∞ -norm	0.077	0.151	0.232	0.303
7	l_∞ -norm	0.005	0.017	0.041	0.076
9	l_∞ -norm	0.002	0.001	0.004	0.011
3	CPU time	0.630	0.820	1.380	1.700
5	CPU time	0.860	0.990	1.240	1.430
7	CPU time	17.010	2.340	1.980	1.820
9	CPU time	933.170	62.390	10.390	5.520
energy loss		$5.5 \cdot 10^{-4}$	$7.4 \cdot 10^{-6}$	$3.5 \cdot 10^{-8}$	$6.1 \cdot 10^{-11}$

Table 4.1: Effect of time-shift on required order, norm of misfit and energy loss illustrated on the Gaussian wavelet. The approximations were determined by using a deterministic starting point as in Section 4.4.4. Note that the l_2 norm of the 9th order approximation with a time-shift of two is less well than for the 7th order approximation. This is due to the fact that in this case the deterministic starting point leads to a local optimum. If the 7th order approximation was used as a starting point for the 9th order approximation a better result would have been obtained.

through in the case of an equality. Strict causality ensures that the direct feed-through is absent.

Due to this strict causality $h(t)$ will be zero for negative t , so that any time-reversed mother wavelet $\psi(-t)$ which does not have this property must be *time-shifted*, by some value t_0 , yielding, to facilitate an accurate approximation of its (correspondingly time-shifted) wavelet transform $\widetilde{W}(\tau, \sigma)$:

$$\widetilde{W}(\tau, \sigma) = \int_{-\infty}^{\infty} x(t) \widetilde{\psi}_{\text{trunc}}(\tau - t) dt, \quad (4.3)$$

where

$$\widetilde{\psi}(t) = \psi(t_0 - t), \quad (4.4)$$

$$\widetilde{\psi}_{\text{trunc}}(t) = \begin{cases} \widetilde{\psi} & \text{for } t \geq 0 \\ 0 & \text{for } t < 0 \end{cases} \quad (4.5)$$

In case $\psi(t_0 - t)$ is nonzero for $t > t_0$, a truncation error results, which should be kept small. This is illustrated in Table 4.1, where the time-reversed and time-shifted Gaussian wavelet is approximated, using various delays. Note that an approximation error will also

occur due to the fact that a wavelet does not usually possess a rational Laplace transform which is a requirement that follows from the restrictions to linear filters of finite order.

A possible approach for this *filter synthesis* problem is to use N building blocks with each an impulse response function $h_k, k = 1, \dots, N$ and to combine them with a weight vector w_k such that $h(t) = \sum_{k=1}^N w_k h_k(t) \approx \tilde{\psi}_{\text{trunc}}(t)$. This approach is for example discussed in [12]. Due to the high requirements on power consumption this approach however is not well suited for the application at hand. The system should be tailored for the desired impulse response to keep the order and thus the power consumption as low as possible.

4.3 Padé approximation of wavelet functions

Padé approximation provides a method to obtain a rational approximation to a given function. In the current application a rational function in the Laplace domain is needed. Therefore the intended application requires one to work in the *frequency domain*. Any *Padé approximation* $H(s)$ of $\tilde{\Psi}(s) = \mathcal{L}\{\tilde{\psi}(t)\}$ is characterized by the property that the coefficients of the *Taylor series expansion*:

$$\tilde{\Psi}(s) = \sum_{k=0}^{\infty} \frac{1}{k!} \tilde{\Psi}^{(k)}(s_0) (s - s_0)^k \quad (4.6)$$

of $\tilde{\Psi}(s)$ around a selected point $s = s_0$ coincide with the corresponding Taylor series coefficients of $H(s)$ up to the highest possible order, given the pre-specified degrees of the numerator and denominator polynomials of $H(s)$. If we denote the Padé approximation $H(s)$ at $s = s_0$ and of order (n, m) with $n \leq m$ by

$$H(s) = \frac{p_0(s - s_0)^n + p_1(s - s_0)^{n-1} + \dots + p_n}{(s - s_0)^m + q_1(s - s_0)^{m-1} + \dots + q_m} \quad (4.7)$$

then there are $m + n + 1$ degrees of freedom, which generically makes it possible to match exactly the first $m + n + 1$ coefficients of the Taylor series expansion of $\tilde{\Psi}(s)$ around $s = s_0$. In fact, this matching problem can easily be rewritten as a system of $m + n + 1$ linear equations in the $m + n + 1$ variables $p_0, p_1, \dots, p_n, q_1, \dots, q_m$. See, e.g., [10].

This brings us to one of the main advantages of Padé approximation: the linear system of equations will generically yield a *unique solution* which is easy to compute. Moreover, a good match is guaranteed between the given function $\tilde{\Psi}(s)$ and its approximation $H(s)$ in a neighborhood of the selected point s_0 . However, there are also some disadvantages which limit the practical applicability of this technique in the setting of this research.

One important issue concerns the selection of the point s_0 . Note that a good approximation of $\tilde{\Psi}(s)$ over the entire (complex) Laplace domain is not a requirement *per se*. Instead, an approximation is needed which performs well when used for convolution in the time domain. Since the function $\tilde{\psi}(t)$ is a wavelet, it effectively will have compact support and in particular it should be approximated well in the region of the time domain where it ‘lives’: which is somewhere near $t = 0$. Now, the initial value theorem for the

Figure 4.3: Top figure: Padé approximations of the Gaussian wavelet. Lower figure: The value of all function has been increased by 1 such they are all positive and they can be plotted on a logarithmic scale. It is clear that the Padé approximation with $s_0 \rightarrow \infty$ is unstable and the Padé approximation with $s_0 = 0$ has a poor fit at the beginning of the impulse response.

Laplace transformation (2.18) motivates the choice $s_0 = \infty$. This choice will lead to a good approximation of $\tilde{\psi}(t)$ near $t = 0$, as is demonstrated in Figure 4.3 for a 7th order approximation of the Gaussian wavelet.

A second important issue concerns stability. The approximation $h(t)$ of $\tilde{\psi}(t)$ is required to tend to zero for large values of t , since $\tilde{\psi}(t)$ has this property too. However, stability does not automatically result from the Padé approximation technique. Indeed, if emphasis is put on obtaining a good fit near $t = 0$ by choosing $s_0 = \infty$, it may easily happen that the resulting approximation becomes unstable: see again the 7th order approximation in Figure 4.3. The selection of a suitable point s_0 involves a choice between a good fit near $t = 0$ and stability, yielding a non-trivial problem which may be difficult to handle, depending on the wavelet at hand. In this respect it may be of interest to note that the 5th order approximation described in [54] and illustrated in Figure 4.3 was obtained with the choice $s_0 = 0$ which corresponds to a good fit in the time domain for large values of t as a result of the final value theorem (2.19). This however usually results in a poor fit at the beginning of the signal in the time domain, which is clearly visible in Figure 4.3 too.

A third issue concerns the choice of the degrees m and n of the numerator and denominator polynomials of the rational approximation $H(s)$. An unfortunate choice may yield an inconsistent system of equations or an unstable approximation. Changing m or n may solve this problem, but the converse may also happen: one may run into stability problems even if s_0 is left unchanged.

In [19] an overview is given of various generalizations and extensions of Padé approximation which aim to deal with some of the problems just mentioned. For instance, it is possible to choose some or all of the poles of the rational approximation $H(s)$ in advance. This offers a possibility to deal with the stability issue, but no clear theory exists on the optimal choices for these poles. Another generalization involves the possibility to use more than one interpolation point, which for instance offers a method to deal with the trade-off between $s_0 = \infty$ and $s_0 = 0$ in a more systematic way. Yet another possibility is to deal with many interpolation points $s_0, s_1, \dots, s_k$ and to require only a match between the values of $H(s_i)$ and $\tilde{\Psi}(s_i)$ for $i = 0, 1, \dots, k$, no longer taking any derivatives at these points into account. One may even specify more interpolation points than the number of unknowns and use a linear least squares estimation technique to arrive at a unique solution. The advantage of such an approach is that one can optimize the function over a better distributed and controlled set of points than with classical Padé approximation. One may choose complex values here. A 9th order approximation of the Gaussian wavelet function obtained with this method is illustrated in Figure 4.3. This shows the feasibility of the approach, but for low order approximations the choice of interpolation points becomes more critical and the results are markedly less well.

However, all of these techniques remain to have one important drawback in common: the quality of the *approximation of the wavelet is not measured directly in the time domain but in the Laplace domain*, and the *criterion used does not allow for a direct interpretation in system theoretic terms*.

4.4 L_2 approximation of wavelet functions

The theory of L_2 -approximation, can be formulated equally well in the time-domain and in the frequency domain, providing an alternative framework for studying the problem of wavelet approximation which offers a number of advantages over the Padé-approach. The advantages of this technique will be discussed, as well as an approach to find a suitable approximation.

4.4.1 Wavelets and the L_2 space

On the conceptual level it is quite appropriate to use the L_2 -norm to measure the quality of an approximation $h(t)$ of the function $\tilde{\psi}(t)$. Indeed, the very definition of the wavelet transform itself involves the L_2 -inner product between the signal $f(t)$ and the mother wavelet $\psi(t)$. It is also desirable that the approximation $h(t)$ of $\tilde{\psi}(t)$ behaves equally well for all time instances t since $h(t)$ is used as a *convolution kernel* with any arbitrary shift. This property holds naturally for L_2 -approximation, but it is not supported by the Padé approximation approach.

Another advantage of L_2 -approximation is that it allows for a description in the time domain as well as in the Laplace domain, so that both frameworks can be exploited to develop further insight. According to *Parseval's identity* [80] the squared L_2 -norm of the difference between $\tilde{\psi}(t)$ and $h(t)$ can be expressed as:

$$\|\tilde{\psi} - h\|^2 = \int_{-\infty}^{\infty} \left(\tilde{\psi}(t) - h(t) \right)^2 dt, \quad (4.8)$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left| \tilde{\Psi}(i\omega) - H(i\omega) \right|^2 d\omega. \quad (4.9)$$

Minimization of $\|\tilde{\psi} - h\|$ is therefore equivalent to minimization of the L_2 -norm of the difference between the Laplace transforms $\tilde{\Psi}_{\text{trunc}}(s)$ and $H(s)$ over the imaginary axis $s = i\omega$. Note that this observation provides a rationale for the choice of interpolation points in a generalized Padé-approximation approach. In addition note that due to the causality of $h(t)$ we can also write:

$$\|\tilde{\psi} - h\|^2 = \int_{-\infty}^0 \left(\tilde{\psi}(t) \right)^2 dt + \int_0^{\infty} \left(\tilde{\psi}(t) - h(t) \right)^2 dt, \quad (4.10)$$

$$= \zeta_{\psi, t_0} + \int_0^{\infty} \left(\tilde{\psi}(t) - h(t) \right)^2 dt, \quad (4.11)$$

so since ζ_{ψ, t_0} does not depend on $h(t)$ we may restrict the criterion to the positive real axis and the use of $\tilde{\psi}(t)$ in the optimization criterion is equivalent to the use of $\tilde{\psi}(t)$.

One of the disadvantages of an L_2 -approximation approach is that there is a risk that the numerical optimization of $\|\tilde{\psi} - h\|$ ends in a local, non-global optimum. Several global optimization techniques and software packages exist and if the problem can be rewritten in a specific form, then a global optimum can be found. However in general there is no guarantee whether a global optimum will be found or even whether it has

been found. Different starting points can give different local optima and thus can be used to find better solutions. Also, the outcomes of other approximation techniques can be used as starting points for L_2 -approximation.

A question that arises is whether we can predict how good our approximation of the *wavelet transformation* is; i.e. not how well the basis functions are approximated, but how well the resulting detail coefficients match the detail coefficients that come from the actual transformation.

Theorem 4.4.1. *The squared error of the approximated wavelet transformation at any given point at any given scale $|W(\tau, \sigma) - \hat{W}(\tau, \sigma)|$ is no greater than the energy $E_{s(t)}$ of the signal $s(t)$ multiplied by the squared error of the wavelet function approximation $\|\tilde{\psi}(t) - h(t)\|$.*

Proof. Denote the error of the wavelet approximation as:

$$\|\tilde{\psi}(t) - h(t)\| = \varepsilon \quad (4.12)$$

Given a scaled version of the wavelet:

$$\psi_{\tau, \sigma} = \frac{1}{\sqrt{\sigma}} \psi \left(\frac{t - \tau}{\sigma} \right), \quad (4.13)$$

and the resulting wavelet transform of signal $s(t)$ at a given scale σ and place τ :

$$W(\tau, \sigma) = \langle s(t), \psi_{\tau, \sigma}(t) \rangle = \int_{-\infty}^{\infty} s(t) \psi_{\tau, \sigma}(t) dt. \quad (4.14)$$

We can take their respective approximations:

$$\begin{aligned} \frac{1}{\sqrt{\sigma}} \psi \left(\frac{t - \tau}{\sigma} \right) &\sim \frac{1}{\sqrt{\sigma}} h \left(t_0 - \frac{t - \tau}{\sigma} \right), \\ &= \frac{1}{\sqrt{\sigma}} h \left(\frac{-t + (\tau + t_0 \sigma)}{\sigma} \right), \\ &= \hat{\psi}_{\tau, \sigma}(t) \end{aligned} \quad (4.15)$$

$$\begin{aligned} \hat{W}(\tau, \sigma) &= \left\langle s(t), \hat{\psi}_{\tau, \sigma}(t) \right\rangle, \\ &= \int_{-\infty}^{\infty} s(t) \hat{\psi}_{\tau, \sigma}(t) dt. \end{aligned} \quad (4.16)$$

We can calculate the error of the approximation as:

$$|W(\tau, \sigma) - \hat{W}(\tau, \sigma)| = |\langle s(t), \psi_{\tau, \sigma}(t) \rangle - \langle s(t), \hat{\psi}_{\tau, \sigma}(t) \rangle|. \quad (4.17)$$

Due to the linearity of the inner product in each of the arguments:

$$|W(\tau, \sigma) - \hat{W}(\tau, \sigma)| = |\langle s(t), \psi_{\tau, \sigma}(t) - \hat{\psi}_{\tau, \sigma}(t) \rangle|. \quad (4.18)$$

Using the Cauchy-Schwarz inequality we obtain an upper bound:

$$|W(\tau, \sigma) - \hat{W}(\tau, \sigma)| \leq \|s(t)\| \|\psi_{\tau, \sigma}(t) - \hat{\psi}_{\tau, \sigma}(t)\| = \sqrt{E_{s(t)}} \varepsilon. \quad (4.19)$$

□

A number of remarks are in place here. First the upper bound depends on the energy of the total signal, whereas the wavelet functions have (effective) compact support. As a result the upper bound can be reduced by taking this into account. The upper bound then becomes dependent on the scale and properties of the wavelet. A second point is that the wavelets and their approximations can possess beneficial properties that further reduce the error, among which the possession of a vanishing moment as described in Section 4.4.3. This property avoids a bias in the approximated wavelet transform.

4.4.2 Parameterization

Particularly in the case of low-order approximation, the L_2 -approximation problem can be approached in a simple and straightforward way in the time domain. As is well known from linear systems theory (see, e.g., [62]) any strictly causal linear filter of finite order n can be represented in the time domain as a state-space system (A, B, C) as in (2.28). The impulse response function $h(t)$ and its Laplace transform $H(s)$ (i.e., the transfer function of the system) are then given by:

$$h(t) = Ce^{At}B, \quad (4.20)$$

$$H(s) = C(sI - A)^{-1}B. \quad (4.21)$$

For the generic situation of stable systems with distinct poles, the impulse response function $h(t)$ is a *linear combination of damped exponentials and exponentially damped harmonics*. For low-order systems, this makes it possible to propose an explicitly parameterized class of impulse response functions among which to search for a good approximation of $\psi(t)$ as described in [67].

The type of functions that come from (4.20) are the class of *Bohl functions* [97, Remark 3.5.3]. Bohl functions are sums of the products of polynomials and exponentials, which, in the real case, comes down to the sum of the product of polynomials, real exponentials, sines and cosines. The matrix exponential is defined as:

$$e^{At} \stackrel{\text{def}}{=} I + At + \frac{A^2t^2}{2} + \frac{A^3t^3}{3!} + \dots \quad (4.22)$$

The matrix A in (4.22) can be transformed, using a similarity transform, to Jordan form (a block-diagonal matrix).

$$A = S^{-1}BS \quad (4.23)$$

$$e^{At} = S^{-1}e^{Bt}S, \quad B \text{ in Jordan form} \quad (4.24)$$

These blocks do not show interaction and each gives rise to a specific Bohl function. Scalar blocks obviously form pure exponentials. Products of polynomials and exponentials correspond to special Jordan structures in the matrix A and since they are not generic, they are not used, unless one has other reasons to consider them, in the parameterization. Typically, the functions that are used in the parameterization are *damped*

exponentials and *exponentially damped sines and cosines*. The latter can model the oscillatory behavior of wavelets.

Sufficiently smooth continuous-time wavelets can be reasonably well approximated by strictly causal stable linear systems with distinct poles. The selected parameterization consists of damped exponentials and exponentially damped harmonics. The parameterized class of impulse response functions has the following form:

$$h(t) = [\alpha_1 e^{p_1 t} + \dots + \alpha_n e^{p_n t}] + [\beta_1 e^{q_1 t} \sin(r_1 t) + \gamma_1 e^{q_1 t} \cos(r_1 t) + \dots + \beta_m e^{q_m t} \sin(r_m t) + \gamma_m e^{q_m t} \cos(r_m t)], \quad (4.25)$$

where the parameters p_k and q_k must be strictly negative for reasons of stability.

For instance, if a 5th order approximation is attempted, this parameterized class of functions $h(t)$ may typically have the following form:

$$h(t) = \alpha_1 e^{p_1 t} + \beta_1 e^{q_1 t} \sin(r_1 t) + \gamma_1 e^{q_1 t} \cos(r_1 t) + \beta_2 e^{q_2 t} \sin(r_2 t) + \gamma_2 e^{q_2 t} \cos(r_2 t), \quad (4.26)$$

Note that (4.26) can be rewritten as a sum of complex exponentials:

$$h(t) = \alpha_1 e^{p_1 t} + \eta_1 e^{(q_1 + ir_1)t} + \eta_1^* e^{(q_1 - ir_1)t} + \eta_2 e^{(q_2 + ir_2)t} + \eta_2^* e^{(q_2 - ir_2)t}, \quad (4.27)$$

with $p_k, q_k \in \mathbb{R}^-$, $\alpha_k, \beta_k, \gamma_k, r_k \in \mathbb{R}$, $\eta_k = \frac{\sqrt{\beta_k^2 + \gamma_k^2}}{2} e^{i(\arctan(\frac{\beta_k}{\gamma_k}) + \mathbf{1}_{R^-(\beta_k)}\pi - \frac{\pi}{2})}$ and $\mathbf{1}_A(x)$ is the indicator function: $\mathbf{1}_A(x) = 1$ iff $x \in A$.

The parameterization has in this form some similarity to the classical problem of *separation of exponentials* [74, Chapter IV-23]. The problem deals with a function given in the following form:

$$f(x) = \rho_1 e^{-\lambda_1 x} + \rho_2 e^{-\lambda_2 x} + \dots + \rho_m e^{-\lambda_m x}, \quad (4.28)$$

and aims to find the amplitudes ρ_i and damping coefficients λ_i . A difference with the problem at hand is that complex exponentials need to be found, and as a result the method would need modification.

For the purpose of IC design it is useful to have a state-space representation (A, B, C) associated with (4.26) available. Such a representation is for instance provided by:

$$A = \begin{pmatrix} p_1 & 0 & 0 & 0 & 0 \\ 0 & q_1 & r_1 & 0 & 0 \\ 0 & -r_1 & q_1 & 0 & 0 \\ 0 & 0 & 0 & q_2 & r_2 \\ 0 & 0 & 0 & -r_2 & q_2 \end{pmatrix}, \quad B = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}, \quad C = (\alpha_1 \quad \beta_1 \quad \gamma_1 \quad \beta_2 \quad \gamma_2). \quad (4.29)$$

Given the explicit form of the wavelet $\tilde{\psi}_{\text{trunc}}(t)$ and the parameterized class of functions $h(t)$, the L_2 -norm of the difference $\tilde{\psi}_{\text{trunc}}(t) - h(t)$ can now be minimized in a straightforward way using standard numerical optimization techniques and software. In

[67, 66] a numerical approach was used that involves discretizing both the wavelet and the parameterized impulse response with a very fine mesh, and locally searching for the least-squares minimum of the difference between the two. The problem of avoiding local optima is discussed in Section 4.4.4. The negativity constraints on $\{p_k\}$ and $\{q_k\}$ which enforce stability are not difficult to handle for most optimization packages.

4.4.3 Vanishing moments

One common property of a wavelet function $\tilde{\psi}(t)$ that was undiscussed so far is that it must have at least one *vanishing moment*. Enforcing the first vanishing moment comes down to the requirement that the integral of the wavelet function is equal to zero:

$$\int_0^\infty \tilde{\psi}(t) dt = 0. \quad (4.30)$$

If this property is not shared by the approximation $h(t)$, this will cause an unwanted bias in the approximation of the wavelet transform as can be seen in the simulation in Figure 4.4, where the $h(t)$ obtained in this manner is denoted “ L_2 normal” in the figure. This is likely to happen in a situation where a truncation error occurs.

Proposition 4.4.2. *The condition $\int_0^\infty h(t) dt = 0$ is equivalent to $H(0) = 0$.*

Proof.

$$0 = \int_0^\infty h(t) dt = \int_0^\infty e^{-st} h(t) dt, \quad \text{with } s = 0 \quad (4.31)$$

$$= H(s) \quad \text{with } s = 0 \quad (4.32)$$

$$= H(0). \quad (4.33)$$

□

In terms of linear filters, the property that the integral of the impulse response function $h(t)$ is zero is equivalent to the property that the step response of the filter tends to zero for large t as follows from (2.27) and the final value theorem (2.19). Indeed, if the wavelet transform is computed for a step input signal, then a bias will be manifest if such a property is not satisfied.

In terms of a state-space representation (A, B, C) we have that

$$H(0) = -CA^{-1}B. \quad (4.34)$$

As an example, for the representation (4.29) it is not difficult to compute A^{-1} since it is block diagonal.

$$A^{-1} = \begin{pmatrix} p_1 & 0 & 0 & 0 & 0 \\ 0 & q_1 & r_1 & 0 & 0 \\ 0 & -r_1 & q_1 & 0 & 0 \\ 0 & 0 & 0 & q_2 & r_2 \\ 0 & 0 & 0 & -r_2 & q_2 \end{pmatrix}^{-1} \quad (4.35)$$

Figure 4.4: A scaled dataset is displayed along with its transformation with a wavelet. The black, thin, dashed line displays the output of a system that has the dataset as an input and attempts to approximate the wavelet transform without enforcing that the step response tends to zero for large t . The reader can observe that a bias manifests that correlates with the input data. The diamonds show the output of another system that does enforce the condition on the step response. For this system the bias is absent in the output data.

$$A^{-1} = \begin{pmatrix} \frac{1}{p_1} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{q_1^2 + r_1^2} \begin{pmatrix} q_1 & -r_1 \\ r_1 & q_1 \end{pmatrix} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{q_2^2 + r_2^2} \begin{pmatrix} q_2 & -r_2 \\ r_2 & q_2 \end{pmatrix} & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad (4.36)$$

This yields the explicit condition:

$$H(0) = \frac{\alpha_1}{p_1} + \frac{-\beta_1 r_1 + \gamma_1 q_1}{q_1^2 + r_1^2} + \frac{-\beta_2 r_2 + \gamma_2 q_2}{q_2^2 + r_2^2} = 0 \quad (4.37)$$

If such an extra nonlinear condition is not conveniently handled by the optimization software, then it can easily be used to eliminate one of the variables from the problem:

$$\alpha_1 = -p_1 \left(\frac{\gamma_1 q_1 - \beta_1 r_1}{q_1^2 + r_1^2} + \frac{\gamma_2 q_2 - \beta_2 r_2}{q_2^2 + r_2^2} \right) \quad (4.38)$$

Figure 4.5: Impulse and step responses of the “normal” and “step corrected” L_2 -approximation

4.4.4 Obtaining a good starting point

When optimizing the parameterized class of impulse response functions to match a certain wavelet function in an L_2 -sense, it is a non-trivial task to find the global optimum. There generally exist a large number of local optima and it is very hard to find the global optimum. In fact the problem of global optimization in this setting is unsolved and no guarantees in general exist that a global optimum can or has been found. For specific problems these guarantees do exist and one can attempt to rewrite problems into such a form, however no suitable form has been found for the problem at hand. The optimization technique used in this study is to locally minimize a discretized least-squares criterion, since an l_2 -criterion is to be minimized. In order to prevent the optimization technique from terminating in an unsatisfactory local optimum, one can attempt to find a good starting point. The choice of the *starting point* can have a considerable impact on the solution found by the L_2 -approximation approach.

A methodology is presented in [66] to obtain a good starting point for the L_2 -approximation approach in an automated fashion. One starts by constructing a high-order model and applying model reduction techniques such that an initial model with the appropriate order n_{req} is obtained. A number of intermediate steps are required as illustrated in Figure 4.6 and listed below:

1. **Sampling the wavelet $\tilde{\psi}$**

Figure 4.6: Automated approximation of functions

The wavelet function $\tilde{\psi}$ is sampled with a sufficiently high resolution over a large enough time interval.

2. Construction of a high order discrete-time FIR-model

The sampled wavelet is used to construct a *high order discrete-time FIR (or moving-average) model*, which has an impulse response that exactly matches the sampled wavelet.

3. Conversion to an intermediate order discrete-time IIR-model

The state-space model is balanced and truncated to yield an accurate *reduced* order discrete-time model, referred to as an intermediate order model.

4. Conversion of the discrete-time IIR-model to a continuous-time IIR-model

The discrete-time model is then *converted back to continuous-time*, using the ZOH-principle. Until here, all steps in the procedure have to be performed just once.

5. Reduction of the continuous-time IIR-model to the desired lower order

The intermediate order continuous-time model is *reduced* to a specified lower order n_{req} , to be used as a starting point for the optimization technique in the next step. Various reduced orders can be attempted until a satisfactory result is obtained, steps 1-4 do not have to be repeated.

6. L₂-approximation of the wavelet function

The low order model obtained in the previous step is used as a *starting point* for

solving the minimization problem described above under the constraint that for (4.34) it must hold that $H(0) = 0$, using an iterative local search optimization technique.

Sampling the wavelet $\tilde{\psi}$

To obtain a good starting point one can start with a sampled version of the time-reversed and shifted wavelet $\tilde{\psi}(t)$ using sample intervals of size Δt :

$$f_k = \tilde{\psi}(k\Delta t), k = 0 \dots n. \quad (4.39)$$

The horizon $n\Delta t$ has to be selected in such a way that stability is obtained. However the computational complexity grows with n . In the case of non-compactly supported wavelets the horizon can be selected by choosing a threshold α and setting $n\Delta t$ close to the smallest $t_{\max}$ for which $|\tilde{\psi}(t)| < \alpha, \forall t > t_{\max}$. One should select α in such a way that not only stability is obtained, but also the truncation error is sufficiently small. Stability is not guaranteed, but implicitly obtained in practice. For compactly supported wavelets n and Δt should be selected in such a way that $\tilde{\psi}(n\Delta t)$ is well outside the support of the wavelet to ensure stability, but n is small enough to keep the calculations feasible. A typical value for Δt is 0.01 and for $n\Delta t$ a typical value is 20, depending on the decay of the function at hand.

Construction of a high order discrete-time FIR-model

The sequence $\{f_k\}$ will be used as the impulse response of a discrete-time system. Caution needs to be taken here since the impulse function in continuous-time is inherently different from the impulse response in discrete-time. The continuous-time impulse function is the *Dirac delta-function* (2.15), yielding an infinitesimally narrow, infinitely tall pulse which integrates to unity. On the other hand the discrete-time impulse can be associated (via *zero-order hold*) with a continuous time block function, called the *Kronecker delta function* (2.16). As a result a correction is required by filtering the sequence $\{f_k\}$ by a FIR block filter $\frac{1}{2} + \frac{1}{2}z^{-1}$ to obtain $\{\hat{f}_k\}$, which corresponds to a discrete-time impulse response. This sequence can be then be used to define a discrete-time system.

The impulse response of the system under construction is chosen to be as follows: $h[0] = 0, h[1] = \hat{f}_0, h[2] = \hat{f}_1, \dots, h[n+1] = \hat{f}_n$. The impulse response of a discrete-time state-space system (A, B, C, d) is given by:

$$h[k] = \begin{cases} d & k = 0 \\ CA^{k-1}B & k > 0. \end{cases} \quad (4.40)$$

Therefore the required impulse response can be obtained by using a system $M_{\text{dh}} = (A_{\text{dh}}, B_{\text{dh}}, C_{\text{dh}}, d_{\text{dh}})$ in controllable companion form and a zero term d_{dh} . The matrix

A_{dh} then has the following form that shifts the state variables in time:

$$A_{\text{dh}} = \left(\begin{array}{ccc|c} 0 & \dots & 0 & 0 \\ \hline & & & 0 \\ & I & & \vdots \\ & & & 0 \end{array} \right) \quad (4.41)$$

The input is only fed to the first state and the vector C determines the output given the function that has to be fitted:

$$B_{\text{dh}} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (4.42)$$

$$C_{\text{dh}} = \begin{pmatrix} \hat{f}_0 & \hat{f}_1 & \dots & \hat{f}_n \end{pmatrix}$$

$$d_{\text{dh}} = 0$$

Eventually the model in (4.42) has to be converted to a low-order continuous time model. It may however not be easily converted directly to a continuous time model because A_{dh} has all its poles at the origin, which makes it impossible to take a matrix logarithm as can be seen from (4.57). Therefore the model M_{dh} is first reduced, then converted to continuous time and finally is reduced again to obtain a continuous-time model of the required low order.

Conversion to an intermediate order discrete-time IIR-model

To reduce the model M_{dh} the *balance and truncate* procedure can be used for which the reader is referred to [89, 96, 121].

M_{dh} has to be balanced first. A stable, discrete-time system is balanced if the following system of equations hold where the first two equations are the well-known discrete-time *Lyapunov-Stein* equations:

$$P - APA^T = BB^T \quad (4.43)$$

$$Q - A^TQA = C^TC \quad (4.44)$$

$$P = Q = \text{diag}\{\sigma_1, \dots, \sigma_n\} \quad (4.45)$$

for positive numbers σ_k , known as the *Hankel singular values* of the system. The system does always admit a balanced realization and the Hankel singular values can be ordered in a decreasing fashion. Observe that (4.43) does only depend on A and B , and as a result not on the wavelet $\tilde{\psi}(t)$ that was fitted to M_{dh} . It is not difficult to see that the matrix P , i.e., the *controllability Grammian*, corresponding to the model M_{dh} is the identity matrix I . As a consequence P will be unaffected by orthogonal state-space transformations.

For (4.45) to hold, the matrix Q or the *observability Grammian* has to be diagonalized first by an orthogonal state-space transformation and then the system will be transformed by an additional diagonal state-space transformation to ensure $P = Q$.

The diagonalization of Q is indeed possible with an orthogonal state-space transform of the form UQU^T since the matrix Q is always positive definite in case of stability and minimality. The transform will have an influence on A, B and C , but not on P ; an identity matrix, as mentioned earlier. This makes it possible to transform Q with singular value decomposition.

$$\bar{A}_{\text{dh}} = U^T A_{\text{dh}} U \quad (4.46)$$

$$\bar{B}_{\text{dh}} = U^T B_{\text{dh}} \quad (4.47)$$

$$\bar{C}_{\text{dh}} = C_{\text{dh}} U \quad (4.48)$$

$$\bar{P} = P = I \quad (4.49)$$

$$\bar{Q} = U^T Q U \quad (4.50)$$

The orthogonal transformation matrix U will be such that $\bar{Q}$ is a diagonal matrix. Condition (4.45) can now be ensured by an additional diagonal state-space transformation.

The procedure discussed in [11] arrives at the same balanced realization in a slightly different way, by observing that for the controllable companion form realization of a moving-average system, the associated Hankel matrix $\mathcal{H}$ built from the finite impulse response $0, h[1], h[2], \dots, h[n]$ as

$$\mathcal{H} = \begin{pmatrix} h[1] & h[2] & h[3] & \dots & h[n] \\ h[2] & h[3] & & h[n] & 0 \\ h[3] & & 0 & & \vdots \\ \vdots & h[n] & 0 & & \vdots \\ h[n] & 0 & \dots & \dots & 0 \end{pmatrix} \quad (4.51)$$

is diagonalized by the same matrix U . This avoids the computation of Q , of which the condition number is the square of the condition number of $\mathcal{H}$.

The controllability grammian P is defined as:

$$P = \sum_{k=0}^{\infty} A^k B B^T (A^T)^k = \mathfrak{C} \mathfrak{C}^T, \quad (4.52)$$

where $\mathfrak{C} = (B \ AB \ A^2 B \ \dots)$, i.e., the infinite controllability matrix. Similarly $Q = \mathfrak{D} \mathfrak{D}^T$, where $\mathfrak{D}^T = (C^T \ A^T C^T \ (A^T)^2 C^T \ \dots)$, i.e. the observability matrix. Since $h[k] = C A^{k-1} B$ and because $A^{k-1} B = 0$ for $k > n$, it is not necessary to consider an infinite matrix $\mathcal{H}$. The Hankel matrix can be written in terms of the observability and controllability matrices as: $\mathcal{H} \mathcal{H}^T = \mathfrak{D} \mathfrak{C} \mathfrak{C}^T \mathfrak{D}^T = \mathfrak{D} \mathfrak{D}^T$, since $P = I$.

After balancing $\bar{Q}$ is a diagonal matrix with the singular values $\{\sigma_1, \dots, \sigma_n\}$ of Q in decreasing order. Upon choosing a threshold ϵ_σ , the state vector may be truncated, retaining as many n_m state components as there are singular values above the threshold, yielding an n_m th order discrete-time system $M_{dm} = (A_{dm}, B_{dm}, C_{dm}, 0)$.

Conversion of the discrete-time IIR-model to a continuous-time IIR-model

The matrix A_{dm} will usually no longer contain zero eigenvalues and therefore the system M_{dm} can be converted to continuous time as described in for example [115]:

$$A_{\text{dm}} = e^{A_{\text{cm}}\Delta t} \quad (4.53)$$

$$B_{\text{dm}} = \int_0^{\Delta t} e^{A_{\text{cm}}\tau} B_{\text{cm}} d\tau \quad (4.54)$$

$$C_{\text{dm}} = C_{\text{cm}} \quad (4.55)$$

with A_{dm} , A_{cm} , B_{dm} and B_{cm} further conveniently related by (see [118]):

$$N = \begin{pmatrix} A_{\text{cm}} & B_{\text{cm}} \\ 0 & 0 \end{pmatrix} \quad (4.56)$$

$$e^{N\Delta t} = \begin{pmatrix} A_{\text{dm}} & B_{\text{dm}} \\ 0 & 1 \end{pmatrix} \quad (4.57)$$

To compute the matrix N first diagonalize the matrix $F = \begin{pmatrix} A_{\text{dm}} & B_{\text{dm}} \\ 0 & 1 \end{pmatrix}$, using that

$$F' = V^{-1} F V = \text{diag}(k), \quad (4.58)$$

where V is the matrix with eigenvectors of F , k is the vector of eigenvalues of F and the function diag gives a diagonal matrix the given vector as its diagonal elements, produces a diagonal matrix F' if F . Next the matrix logarithm of F can be calculated as:

$$N\Delta t = \ln F = V \ln F' V^{-1} \quad (4.59)$$

From (4.59) it can be seen that the logarithm of a matrix, of which the diagonal elements contain the diagonal entries of the matrix A_{dm} has to be taken. Since the logarithm of zero is undefined one cannot transform matrices A_{dm} that have zeros on the diagonal to continuous-time directly. Note that the controllable companion form with a matrix A_{cm} as in (4.41) does have zeros on the diagonal and hence cannot be directly transformed to continuous-time.

Reduction of the continuous-time IIR-model to the desired lower order

The model M_{cm} will not yet have the required order n_{req} , therefore the continuous time model that has now been obtained has to be reduced even further with, for instance, the balance and truncate procedure. This procedure now involves the continuous-time Lyapunov equations, so that eventually a continuous-time model M_{cr} of reduced order n_{req} is obtained.

Figure 4.7: Automatic approximation of the Morlet wavelet

L2-approximation of the wavelet function

M_{cr} is a single input - single output, strictly proper, asymptotically stable, continuous-time state-space system which will not in general be of the form as in (4.25), with a corresponding state-space realization of a form similarly as the example in (4.29) and will not obey the constraint as described in Section 4.4.3. To ensure that the step response of M_{cr} tends to zero, the constant term in the numerator of the transfer function associated with M_{cr} is simply set to zero to enforce a zero at zero. Next the system has to be brought in the model form as in (4.25). The complex eigenvalues associated with A_{cr} will be ordered in complex pairs where the imaginary part is positive. The states corresponding to real eigenvalues follow last. The A matrix will have a block structure, from which the required parameters can be easily read off. Note that A_{cr} fixes the number of complex pairs that the approximating model will have.

In Figure 4.7 the approximation of the Morlet wavelet with systems of various reduced orders is illustrated. In the lower half of this figure the deviation from the ideal wavelet is shown. Since the “ideal” wavelet is truncated, it does not have a zero integral. The approximations however, do have this property, as described in Section 4.4.3, and therefore a perfect fit is not possible.

Abbreviation	Matlab Release	Function	Algorithm
lsqc2008bTrr	R2008b	"lsqcurvefit"	"Algorithm"="trust-region-reflective"
fmin2008bTrr	R2008b	"fminsearch"	"Algorithm"="trust-region-reflective"
lsqc14-3GN	R14 SP3	"lsqcurvefit"	"LargeScale"="off", "LevenbergMarquardt"="off"
lsqc2007bGN	R2007b	"lsqcurvefit"	"LargeScale"="off", "LevenbergMarquardt"="off"
lsqc2008bGN	R2008b	"lsqcurvefit"	"LargeScale"="off", "LevenbergMarquardt"="off"
fmin14-3NM	R14 SP3	"lsqcurvefit"	"LargeScale"="off", "LevenbergMarquardt"="off"
fmin2007bNM	R2007b	"fminsearch"	"LargeScale"="off", "LevenbergMarquardt"="off"
fmin2008bNM	R2008b	"fminsearch"	"LargeScale"="off", "LevenbergMarquardt"="off"

Table 4.2: Used Matlab settings and function for wavelet approximation. The settings "LargeScale"="off", "LevenbergMarquardt"="off" attempts to use Gauss-Newton optimization for lsqcurvefit and Nelder-Mead simplex direct search for fminsearch. Under certain conditions the optimization may switch to Levenberg-Marquardt or to a large-scale method. The lsqcurvefit function offers the possibility to specify upper and lower bounds for each of the parameters.

Settings	12 th order	10 th order	8 th order	7 th order	6 th order	5 th order	3 rd order
lsqc2008bTrr	0.0022	0.0009	0.0056	0.0070	0.0132	0.0475	0.3975
fmin2008bTrr	0.0073	0.0068	0.0066	0.0073	0.0132	0.0475	0.3975
lsqc14-3GN	0.0163	0.0198	0.0165	0.0173	0.0329	0.1168	0.9973
lsqc2007bGN	0.0044	0.0103	0.0056	0.0070	0.0132	0.0475	0.3975
lsqc2008bGN	0.0044	0.0103	0.0056	0.0070	0.0132	0.0475	0.3975
fmin14-3NM	0.0164	0.0164	0.0173	0.0199	0.0329	0.1168	0.9973
fmin2007bNM	0.0076	0.0068	0.0065	0.0077	0.0132	0.0475	0.3975
fmin2008bNM	0.0073	0.0068	0.0066	0.0073	0.0132	0.0475	0.3975

Table 4.3: Approximation of Gaussian wavelet with $t_0 = 2.0$ and Matlab settings as in Table 4.2.

4.5 Empirical results

The L_2 approximation approach for wavelets allows for the implementation of a wide variety of wavelets. In [52] it was discussed how complex wavelets can be implemented and in [65] a large number of wavelets are approximated with this approach. Besides the Gaussian wavelet, also the Morlet wavelet and the Mexican Hat wavelet, which are of interest for ECG processing [63, 3, 2, 20], have been approximated. Not only continuous wavelets have been approximated, but also discrete wavelets such as some of the Daubechies wavelets, provided that they are sufficiently smooth. The required order to obtain a satisfying approximation of these wavelets is quite high though.

In [65] it is shown that a 4th order L_2 approximation of a Gaussian wavelet outperforms a 5th order Padé approximation, with respect to a multitude of evaluation methods, thus showing a significant improvement in performance. The approximated wavelets are shown in Figure 4.8.

In order to illustrate how well various wavelets are approximated the l_2 norms of the approximation errors of various wavelets will be shown.

As can be seen from Tables 4.3–4.8, accuracy does not always increase with the order. The use of a lower order solution as a starting point should give a performance increase in such cases. If a local optimum is nearby, the optimization may terminate in this local optimum. A way to overcome the latter problem is to introduce random disturbances on the starting point and use random starting point in the area around the deterministic starting points. Another observation is that with the same settings, the

Figure 4.8: Approximations of various wavelet functions. The thick gray line represents the wavelet function and the thin dashed black line represents the impulse response of the approximating system. The following wavelets were approximated with the specified orders n_{req} and time shifts t_0 : (a) Gaussian $n_{\text{req}} = 5, t_0 = 2$, (b) Morlet $n_{\text{req}} = 5, t_0 = 2.5$, (c) Mexican Hat $n_{\text{req}} = 6, t_0 = 3.2$, (d) Daubechies 3 $n_{\text{req}} = 6, t_0 = -1$, (e) Daubechies 7 $n_{\text{req}} = 8, t_0 = -4$, (f) Coiflet 5 $n_{\text{req}} = 8, t_0 = -11$.

Settings	12 th order	10 th order	8 th order	7 th order	6 th order	5 th order	3 rd order
lsqc2008bTrr	0.0026	1.5043	0.0277	0.0636	0.0926	0.2372	0.6318
fmin2008bTrr	0.0030	0.0043	0.0277	0.0636	0.0926	0.2372	0.6398
lsqc14-3GN	0.0026	0.0037	0.0277	0.0636	0.0926	0.2372	0.6318
lsqc2007bGN	0.0026	0.0037	0.0277	0.0636	0.0926	0.2372	0.6318
lsqc2008bGN	0.0026	0.0037	0.0277	0.0636	0.0926	0.2372	0.6318
fmin14-3NM	0.0029	0.0043	0.0277	0.0636	0.0926	0.2372	0.6398
fmin2007bNM	0.0029	0.0043	0.0277	0.0636	0.0926	0.2372	0.6318
fmin2008bNM	0.0030	0.0043	0.0277	0.0636	0.0926	0.2372	0.6318

Table 4.4: Approximation of Morlet wavelet with $t_0 = 2.5$ and Matlab settings as in Table 4.2.

Settings	12 th _{order}	10 th _{order}	8 th _{order}	7 th _{order}	6 th _{order}	5 th _{order}	3 rd _{order}
lsqc2008bTrr	0.0302	0.0051	0.0074	0.1447	0.1469	0.1782	0.9939
fmin2008bTrr	0.1077	0.1320	0.1163	0.5844	0.1720	0.1721	0.5019
lsqc14-3GN	0.0048	0.0082	0.0074	0.0189	0.0613	0.1782	0.5019
lsqc2007bGN	0.3790	1.0549	1.1359	3.0077	0.1108	1.1009	0.5019
lsqc2008bGN	0.6411	1.2933	1.0450	0.0189	0.0613	0.9974	0.5019
fmin14-3NM	0.0060	0.0052	0.0083	0.0189	0.0613	0.1782	0.5019
fmin2007bNM	0.1237	0.1871	0.2179	0.3885	0.4874	0.4850	0.5019
fmin2008bNM	0.1077	0.1320	0.1163	0.5844	0.1720	0.1721	0.5019

Table 4.5: Approximation of Mexican hat wavelet with $t_0 = 3.2$ and Matlab settings as in Table 4.2.

Settings	12 th _{order}	10 th _{order}	8 th _{order}	7 th _{order}	6 th _{order}	5 th _{order}	3 rd _{order}
lsqc2008bTrr	0.8419	0.5685	0.9998	1.0001	1.0000	1.0001	1.0001
lsqc14-3GN	0.1348	0.2785	0.6065	0.6065	0.6065	0.6086	1.0001
lsqc2007bGN	0.8512	0.5835	0.8015	0.8477	1.0001	1.0001	1.0001
lsqc2008bGN	0.8419	0.5685	0.9998	1.0001	1.0000	1.0001	1.0001
fmin2007bNM*	0.1440	1.0001	1.0000	0.9998	1.0001	1.0001	1.0001
fmin2008bNM*	0.5749	0.5728	0.9999	1.0000	1.0000	1.0000	1.0000

Table 4.6: Approximation of Daubechies 3 wavelet with $t_0 = 0.0$ and Matlab settings as in Table 4.2. The * indicates that a different starting point was used to obtain the approximation. For results close to one, the impulse response of the obtained systems looks very much like an impulse; a very undesirable approximation.

Settings	12 th _{order}	10 th _{order}	8 th _{order}	7 th _{order}	6 th _{order}	5 th _{order}	3 rd _{order}
lsqc2008bTrr	0.0369	0.0429	0.0993	0.2430	0.2510	0.5241	0.8400
fmin2008bTrr	0.0440	0.0449	0.0993	0.2430	0.2510	0.5241	0.8400
lsqc14-3GN	0.0369	0.0429	0.0993	0.2430	0.2510	0.5241	0.8400
lsqc2007bGN	0.0369	0.0429	0.0993	0.2430	0.2510	0.5241	0.8400
lsqc2008bGN	0.0440	0.0449	0.0993	0.2430	0.2510	0.5241	0.8400
fmin14-3GN	0.0440	0.0453	0.0993	0.2430	0.2510	0.5241	0.8400
fmin2007bNM	0.0439	0.0443	0.0993	0.2430	0.2510	0.5241	0.8400
fmin2008bNM	0.0440	0.0449	0.0993	0.2430	0.2510	0.5241	0.8400

Table 4.7: Approximation of Daubechies 7 wavelet with $t_0 = -4.0$ and Matlab settings as in Table 4.2.

Settings	12 th _{order}	10 th _{order}	8 th _{order}	7 th _{order}	6 th _{order}	5 th _{order}	3 rd _{order}
lsqc2008bTrr	0.0282	0.0706	0.1832	0.3510	0.4340	0.6335	0.8840
fmin2008bTrr	0.0329	0.0730	0.1857	0.3510	0.4955	0.6335	0.8783
lsqc14-3GN	0.0284	0.0706	0.1832	0.3510	0.4340	0.6335	0.8840
lsqc2007bGN	0.0284	0.0706	0.1832	0.3510	0.4340	0.6335	0.8840
lsqc2008bGN	0.0284	0.0706	0.1832	0.3510	0.4340	0.6335	0.8840
fmin14-3GN	0.0329	0.0730	0.1857	0.3510	0.4955	0.6335	0.8783
fmin2007bGN	0.0329	0.0730	0.1894	0.3510	0.4955	0.6335	0.8783
fmin2008bGN	0.0329	0.0730	0.1857	0.3510	0.4955	0.6335	0.8783

Table 4.8: Approximation of Coiflet 5 wavelet with $t_0 = -11.0$ and Matlab settings as in Table 4.2.

Figure 4.9: Approximations of Mexican Hat wavelet. The time-shifted, time-reversed Mexican Hat wavelet is displayed along its 18th order approximation that results from step 4 of the procedure to find a suitable starting point. The 7th order approximation from step 5 (which does not necessarily have an integral of 1) and the 7th order approximation from step 6 are displayed as well and show a poor fit with respect to the Mexican Hat wavelet.

acquired approximation using Matlab Release 14 Service Pack 3 is markedly better than with the newer Matlab versions.

As can be seen from Table 4.5, the 7th order approximation of the Mexican Hat wavelet, with the `lsqcurvefit` function and the Gauss-Newton algorithm, is dramatic. This approximation is illustrated in Figure 4.9. It is clear that the starting point for the L_2 wavelet design is very inconvenient. If one considers the Hankel singular values corresponding to the 18th order approximation in Figure 4.10, one can observe that all but one states seem to contain relevant information about the system. This is further illustrated if the corresponding impulse responses are taken into consideration in Figure 4.11. The 17th order system is still a good starting point and the quality quickly decreases. How dramatic the decrease is depends on a range of factors such as the wavelet at hand, the involved time shift, etc. And the effect of the quality of the starting point depends on the specific optimization surface at hand, the optimization routine and even the specific version of the software as becomes apparent from Tables 4.2–4.8.

Figure 4.10: Hankel singular values of the 18th order approximation of the Mexican Hat wavelet that results from step 4 of the procedure to find a suitable starting point.

Figure 4.11: Impulse responses of systems that are used as a starting point for the L_2 approximation of the Mexican Hat wavelet.

Chapter 5

Orthogonal wavelet design

An important practical issue of interest is the choice of a wavelet basis [103]. When employing the Fourier transform the basis functions are fixed, i.e., sines and cosines. However for wavelets more freedom exists and the basis needs to satisfy the conditions associated with the wavelet framework. One possibility for wavelet selection is to use a parameterized wavelet that can be tuned for the application at hand [101]. Another possibility is to design a custom wavelet as is the case in this study [69]. Note that the current motivation for wavelet design is not to achieve a good quality lossy compression as is commonly seen, but feature detection, better representations and other signal processing applications. Earlier work on the design of wavelets matched the amplitude spectrum and the phase spectrum of a wavelet to a reference signal in the Fourier domain separately [24]. The authors of [46] describe an algorithm for the design of biorthogonal and semi-orthogonal wavelets. The design criterion comes down to the maximization of the energy in the approximation coefficients, however no clear motivation is provided by the authors. In [92] a parameterization of orthogonal wavelets, involving polyphase filters and the lattice structure as for example in [107] (see also Section 3.2), was discussed, however no clear design criterion was provided.

In Section 5.1 two design criteria are introduced for discrete wavelets, that can be used to measure the quality of a discrete wavelet. This quality measure is with respect to a given signal and the key idea behind it is the maximization of the sparsity. In Section 5.2 a parameterization of discrete-time orthogonal wavelets through polyphase filter banks with a lattice structure as in [92, 107] is discussed. Next the enforcement of additional vanishing moments and the use of the design criterion is discussed. Together with the results of Chapters 3 and 4 a complete procedure for orthogonal wavelet filter design is provided here. Using the wavelet equation (3.25) it is possible to compute a wavelet function associated with the designed wavelet filter bank. If this wavelet function is sufficiently smooth, it is possible to approximate the designed wavelet in analog circuits. The steps needed to come to an approximation for the implementation in analog circuits

is illustrated in Figure 1.1.

When multiple well-defined features in a signal need to be distinguished simultaneously, multiwavelets [76, 77, 107] can be employed. In the multiwavelet case multiple bases are used to span the l_2 space. In Section 3.4 it has been discussed how these multiwavelets can be parameterized in polyphase form and how the condition of orthogonality comes down to requiring that system in polyphase form is “lossless” as previously discussed in [116, 117, 107]. In order to design multiwavelets, a concrete parameterization of lossless systems is required, which is discussed in Section 5.3. For this parameterization in Section 5.3.1 the “tangential Schur algorithm” [55] is used. This parameterization is first provided for scalar wavelets, followed by the parameterization for multiwavelets. In this section the enforcement of a first vanishing moment, which is a requirement for a valid wavelet multiresolution structure, is discussed.

5.1 Measures for the quality of a given representation

When faced with an application, the choice of a wavelet is not a trivial task. A given wavelet can have desirable properties such as a large number of vanishing moments, optimal ratio in the Heisenberg uncertainty rectangle, linear phase, etcetera, but this may not give a direct answer to the question “what wavelet should be used?”. In this work a quantitative measure is described to assess the quality of a given wavelet for a number of application types.

For compression purposes a wavelet is good for a signal if the signal can be represented in the wavelet domain with only a few nonzero coefficients. This means that it has a *sparse representation in the wavelet domain* and is well located in both time and frequency. This is also a good property for detection purposes: if the wavelet is optimized for a certain feature, one then can identify it at a certain scale and relate it to a specific time. If one wants to detect anomalies, a deviation from this sparsity is a good indicator for this.

The recursive application of the wavelet filter bank to the low-pass filter outputs (see Figure 3.3) gives a decomposition in terms of detail coefficients at all levels and in terms of approximation coefficient(s) at the coarsest level. In case *orthonormal* wavelets are employed, the energy of the signal is preserved in the detail and approximation coefficients (using the notation in Section 3.2.3) [107, page 27]:

$$\sum_k x_k^2 = \sum_{j=1}^{j_{max}} \sum_k \left(b_k^{(j)}\right)^2 + \sum_k \left(a_k^{(j_{max})}\right)^2 \quad (5.1)$$

Note that this is the wavelet analogue of Parseval’s identity [80] as discussed on page 4.4.1. We introduce a vector w to consist of all these detail and approximation coefficients, i.e.

$$w = (b_1^{(1)}, \dots, b_k^{(1)}, b_1^{(2)}, \dots, b_k^{(2)}, \dots, b_1^{(j_{max})}, \dots, b_k^{(j_{max})}, a_1^{(j_{max})}, \dots, a_k^{(j_{max})}). \quad (5.2)$$

From (5.1) it follows that the l_2 -norm of this vector w (from now on one is implicitly referred to this vector when measuring the wavelet decomposition) is not an appropriate measure for the quality of the wavelet.

The guiding principle to obtain sparsity that was proposed in [69] is to aim for the *maximization of the variance* [81]. This is worked out in the following two ways which are relevant for orthogonal wavelet transforms:

1. maximization of the variance of the absolute values of the wavelet coefficients
2. maximization of the variance of the squared wavelet coefficients

The latter one maximizes the variance of the energy distribution over the detail and approximation coefficients at the various scales, and is shown below to correspond to maximization of the l_4 -norm. The former one corresponds to minimization of the l_1 -norm, which is a well-known criterion to achieve sparsity in various other contexts, see for example [33, 34, 22, 21]

Theorem 5.1.1. *Let w be the vector of the wavelet and the approximation coefficients as in (5.2), resulting from the processing of a signal $x = (x_0, x_1, x_2, \dots, x_m)$ by means of an orthogonal filter bank. Then:*

1. *Maximization of the variance of the sequence of absolute values $|w_k|$ is equivalent to minimization of the l_1 -norm $V_1 = \sum_{k=0}^m |w_k|$.*
2. *Maximization of the variance of the sequence of energies $|w_k|^2$ is equivalent to maximization of the l_4 -norm $V_4 = (\sum_{k=0}^m |w_k|^4)^{1/4}$.*

Proof. We will prove 1. and 2. separately:

1. As expressed in (5.1), the energy E in a signal is given by $E = \sum_k |x_k|^2 = \sum_k |w_k|^2$, irrespective of the choice of orthogonal wavelet basis. The variance of the vector of absolute values $\{|w_k|\}$ is given by

$$\frac{\sum_k |w_k|^2}{m+1} - \left(\frac{\sum_k |w_k|}{m+1} \right)^2 = \frac{E}{m+1} - \left(\frac{\sum_k |w_k|}{m+1} \right)^2,$$

in which E and m are constant. Hence maximization of this quantity is equivalent to minimization of V_1 .

2. The variance of the vector of energies $\{|w_k|^2\}$ is given by

$$\frac{\sum_k |w_k|^4}{m+1} - \left(\frac{\sum_k |w_k|^2}{m+1} \right)^2 = \frac{\sum_k |w_k|^4}{m+1} - \left(\frac{E}{m+1} \right)^2,$$

in which E and m are constant. Hence maximization of this quantity is equivalent to maximization of the V_4 .

□

Figure 5.1: (a) Smoothed ECG beat of 256 samples. (b) Designed wavelets for ECG beat with filter length 8. (c) Wavelet decomposition of ECG beat with l_1 -optimized wavelet. In the top row the coarsest detail coefficients are displayed. The intensity of the blocks corresponds to the coefficient values. The lower rows show finer detail coefficients. The approximation coefficients have been omitted since their large values may hamper the readability of the detail coefficients. (d) Wavelet decomposition of ECG beat with l_4 -optimized wavelet.

Example 5.1.2. As an example, take the smoothed ECG beat that is displayed in Figure 5.1 along with the wavelets with filter length 8 that have been designed for this signal by optimizing the criteria V_1 and V_4 , respectively. In the lower part of the figure the wavelet decomposition of the signal using the respective designed wavelets is displayed. From Figure 5.1c and 5.1d it can be clearly seen that just a few wavelet coefficients are significantly larger than the rest as expected. $\square$

5.2 Wavelet parameterization and design

For wavelet design, it is inconvenient to take the set of filter coefficients directly as a search space for numerical optimization, on which additional constraints are imposed

Figure 5.2: Two examples of lattices

to meet all of the necessary and desired conditions. Instead we aim to construct a parameterization which builds the desired properties in. In this research the choice is made to restrict to orthogonal wavelets. One of the advantages of this choice is that the inverse transform will be the adjoint of the forward transform $H^{-1} = H^\dagger$. Using the polyphase representation (Section 3.2.7) offers the advantage that the analysis and synthesis filters are adjacent, without down- and upsampling in between, making is easier to build in the required conditions for orthogonal wavelets from Section 3.2: normalization, double shift orthogonality and a first vanishing moment. The lattice structure [107] or all-pass systems are used to ensure that the filter is orthogonal as discussed below.

5.2.1 Lattice structure

If one has a orthonormal polyphase matrix $H_p(z)$, it can be expressed in *lattice form*. A lattice structure implements a filter $H_p(z)$ as a cascade of simple building blocks that each implement a single multiplication. In lattice form, wavelets can be implemented as a cascade of constant matrices and delays (see for example [107, section 4.5] and [116]). In Figure 5.2 two examples of lattices are displayed. The lattice in Figure 5.2a corresponds to the following polyphase filter matrix:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}. \quad (5.3)$$

And the lattice in Figure 5.2b with the matrix as in (5.6).

As noted in for example [107] in order to have symmetry and consequently linear phase filters one can choose $a = d$ and $b = c$. For orthonormal filters one chooses $a = d$ and $b = -c$ and, if desired, one appropriately normalizes the coefficients. The lattice structure is convenient since a cascade of linear phase filters is again linear phase and a cascade of orthogonal filters is again orthogonal. Depending on the design, each lattice will be constructed to be linear phase or orthogonal; quantization errors with respect to

Figure 5.3: Lattice structure for a polyphase filter bank

the parameters will not change this. Any 2-channel orthonormal filter can be expressed in lattice form [107].

When choosing $a = d$ and $b = -c$ and using a suitable normalization, the lattice structure can also be implemented with a rotation matrix:

$$R = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}. \quad (5.4)$$

With this structure the $2n$ low pass filter coefficients (as in (3.6)) $c_0, c_1, \dots, c_{2n-1}$ can be reparameterized in terms of n new parameters $\theta_1, \theta_2, \dots, \theta_n$. Note that, unlike in [107], conventional rotation matrices are used.

For $k = 1, \dots, n$, let

$$R(\theta_k) = \begin{bmatrix} \cos \theta_k & -\sin \theta_k \\ \sin \theta_k & \cos \theta_k \end{bmatrix}, \quad (5.5)$$

and let

$$\Lambda(z) = \begin{bmatrix} 1 & 0 \\ 0 & z^{-1} \end{bmatrix}. \quad (5.6)$$

Then consider the 2×2 matrix product

$$H(z) = R(\theta_n)\Lambda(z)R(\theta_{n-1})\Lambda(z) \cdots R(\theta_2)\Lambda(z)R(\theta_1)\Lambda(-1). \quad (5.7)$$

The polyphase filter structure that follows from (5.7) is illustrated in Figure 5.3. With this lattice structure the orthogonality constraints from Section 3.2.2 are automatically satisfied [107].

In Section 3 the convention is used that the energy of the wavelet low- and high-pass filter coefficients is normalized to one: $\sum_k c_k^2 = \sum_k d_k^2 = 1$. A filter $H(z) = \sum_{k=0}^{n-1} \begin{pmatrix} c_{2k} & c_{2k+1} \\ d_{2k} & d_{2k+1} \end{pmatrix} z^{-k}$ that forms a wavelet filter bank will now be constructed. The coefficients c_k and d_k should therefore obey the conditions discussed in Section 3 (Normalization and double shift orthogonality and a vanishing moment). The coefficients d_k will be related to the coefficients c_k according to the alternating flip construction $d_k = (-1)^k c_{2n-1-k}$. For the sum of each phase (even and odd) of the filters the following holds [107]:

Proposition 5.2.1. *For relation between 1) the sum of the even phase of the low-pass coefficients $c_{2(\ell-1)}$, 2) the sum of the odd phase of the low-pass coefficients $c_{2\ell-1}$, 3) the sum of the even phase of the high-pass coefficients $d_{2(\ell-1)}$ and 4) the sum of the odd phase of the high-pass coefficients $d_{2\ell-1}$ respectively, and the parameters θ_k of the polyphase representation of wavelets in lattice form the following holds:*

$$\sum_{\ell=1}^n c_{2(\ell-1)} = \cos \left(\sum_{k=1}^n \theta_k \right) \quad (5.8)$$

$$\sum_{\ell=1}^n c_{2\ell-1} = \sin \left(\sum_{k=1}^n \theta_k \right) \quad (5.9)$$

$$\sum_{\ell=1}^n d_{2(\ell-1)} = -\sin \left(\sum_{k=1}^n \theta_k \right) \quad (5.10)$$

$$\sum_{\ell=1}^n d_{2\ell-1} = \cos \left(\sum_{k=1}^n \theta_k \right). \quad (5.11)$$

Proof. We shall prove this proposition by induction. For $n = 1$ the wavelet filter coefficients become: $c_0 = \cos \theta_1$, $c_1 = \sin \theta_1$, $d_0 = -\sin \theta_1$ and $d_1 = \cos \theta_1$ for which the theorem clearly holds. Now assume that for $n = p$ the proposition holds, making the following equations true: $\sum_{\ell=1}^p c_{2(\ell-1)} = \cos(\sum_{k=1}^p \theta_k)$, $\sum_{\ell=1}^p c_{2\ell-1} = \sin(\sum_{k=1}^p \theta_k)$, $\sum_{\ell=1}^p d_{2(\ell-1)} = -\sin(\sum_{k=1}^p \theta_k)$ and $\sum_{\ell=1}^p d_{2\ell-1} = \cos(\sum_{k=1}^p \theta_k)$. Now consider Figure 5.3. From the lattice structure it can be seen that for $n = p + 1$ it holds for the even phase that:

$$\begin{aligned} \sum_{\ell=1}^n c_{2(\ell-1)} &= \cos \theta_{p+1} \cos \left(\sum_{k=1}^p \theta_k \right) - \sin \theta_{p+1} \sin \left(\sum_{k=1}^p \theta_k \right) \\ &= \cos \left(\sum_{k=1}^{p+1} \theta_k \right). \end{aligned}$$

For the odd phase it holds that:

$$\begin{aligned} \sum_{\ell=1}^n c_{2\ell-1} &= \sin \theta_{p+1} \cos \left(\sum_{k=1}^p \theta_k \right) + \cos \theta_{p+1} \sin \left(\sum_{k=1}^p \theta_k \right) \\ &= \sin \left(\sum_{k=1}^{p+1} \theta_k \right). \end{aligned}$$

For the even phase of the high-pass filter it holds that

$$\begin{aligned} \sum_{\ell=1}^n d_{2(\ell-1)} &= -\sin \theta_{p+1} \cos \left(\sum_{k=1}^p \theta_k \right) - \cos \theta_{p+1} \sin \left(\sum_{k=1}^p \theta_k \right) \\ &= -\sin \left(\sum_{k=1}^{p+1} \theta_k \right). \end{aligned}$$

For the odd phase it holds that:

$$\begin{aligned} \sum_{\ell=1}^n d_{2\ell-1} &= -\sin \theta_{p+1} \sin \left(\sum_{k=1}^p \theta_k \right) + \cos \theta_{p+1} \cos \left(\sum_{k=1}^p \theta_k \right) \\ &= \cos \left(\sum_{k=1}^{p+1} \theta_k \right). \end{aligned}$$

which completes the proof. $\square$

Due to (5.8) and (5.9) the identity

$$\left(\sum_{\ell} c_{2(\ell-1)} \right)^2 + \left(\sum_{\ell} c_{2\ell-1} \right)^2 = 1 \quad (5.12)$$

holds.

From Figure 5.3, the relation between the parameters θ_k and the filter coefficients can be observed. If the filter coefficients for a lattice structure with $n = p$ are known, then their relation with the filter coefficients for the case $n = p + 1$ and the additionally involved parameter θ_{p+1} can be easily found.

Observation 5.2.2. Suppose that the $p \times 2$ array with filter coefficients for the case $n = p$ is known to be

$$\begin{pmatrix} c_0 & c_1 \\ c_2 & c_3 \\ \vdots & \vdots \\ c_{2(p-1)} & c_{2p-1} \end{pmatrix}.$$

Then the $(p + 1) \times 2$ array with filter coefficients for the case $n = p + 1$ is:

$$\begin{pmatrix} c_0 \cos \theta_{p+1} & c_0 \sin \theta_{p+1} \\ c_2 \cos \theta_{p+1} - c_1 \sin \theta_{p+1} & c_2 \sin \theta_{p+1} + c_1 \cos \theta_{p+1} \\ c_4 \cos \theta_{p+1} - c_3 \sin \theta_{p+1} & c_4 \sin \theta_{p+1} + c_3 \cos \theta_{p+1} \\ \vdots & \vdots \\ c_{2(p-1)} \cos \theta_{p+1} - c_{2p-3} \sin \theta_{p+1} & c_{2(p-1)} \sin \theta_{p+1} + c_{2p-3} \cos \theta_{p+1} \\ -c_{2p-1} \sin \theta_{p+1} & c_{2p-1} \cos \theta_{p+1} \end{pmatrix} = \begin{pmatrix} c_0 & 0 \\ c_2 & c_1 \\ c_4 & c_3 \\ \vdots & \vdots \\ c_{2(p-1)} & c_{2p-3} \\ 0 & c_{2p-1} \end{pmatrix} R(\theta_{p+1})^T \quad (5.13)$$

as follows from the lattice structure in Figure 5.2. $\square$

Example 5.2.3. As an example the low-pass coefficients for the filters of orders $n = 2$ and $n = 3$ are given:

$n = 2$	c_0	$= \cos \theta_1 \cos \theta_2$
	c_1	$= \cos \theta_1 \sin \theta_2$
	c_2	$= -\sin \theta_1 \sin \theta_2$
	c_3	$= \sin \theta_1 \cos \theta_2$
$n = 3$	c_0	$= \cos \theta_1 \cos \theta_2 \cos \theta_3$
	c_1	$= \cos \theta_1 \cos \theta_2 \sin \theta_3$
	c_2	$= -\sin \theta_1 \sin \theta_2 \cos \theta_3 - \cos \theta_1 \sin \theta_2 \sin \theta_3$
	c_3	$= -\sin \theta_1 \sin \theta_2 \sin \theta_3 + \cos \theta_1 \sin \theta_2 \cos \theta_3$
	c_4	$= -\sin \theta_1 \cos \theta_2 \sin \theta_3$
	c_5	$= \sin \theta_1 \cos \theta_2 \cos \theta_3$

In order to avoid a bias in the wavelet transform and in order to make it possible to obtain a well-defined scaling and wavelet function (see Section 3.2.4 and [107, section 6.1]), the wavelet filter must possess at least one vanishing moment. As explained in Section 3.2.5, when a wavelet has vanishing moments, this offers a number of advantages such as giving the wavelet regularity and smoothness. In terms of the filter coefficients this comes down to the constraint:

$$\sum_{\ell} c_{2(\ell-1)} - \sum_{\ell} c_{2\ell-1} = 0. \quad (5.14)$$

In terms of the parameters $\theta_1, \dots, \theta_n$ this comes down to the following:

Theorem 5.2.4. *Consider a polyphase filter in lattice structure with the parameters $\theta_1, \dots, \theta_n$. For this filter to have at least a single vanishing moment the condition is [107, Theorem 4.6]:*

$$\sum_{k=1}^n \theta_k = \frac{\pi}{4} + \ell 2\pi, \quad \ell \in \mathbb{Z} \quad (5.15)$$

Proof. In terms of the filter coefficients the condition comes down to $\sum_{\ell=1}^n c_{2(\ell-1)} = \sum_{\ell=1}^n c_{2\ell-1}$. From Proposition 5.2.1 it follows that this is equivalent with the condition: $\cos(\sum_{k=1}^n \theta_k) = \sin(\sum_{k=1}^n \theta_k)$. This condition is obviously satisfied if $\sum_{k=1}^n \theta_k = \frac{\pi}{4} + \ell\pi$, $\ell \in \mathbb{Z}$. However, when using the sign convention that $\sum_{k=1}^{2n-1} c_k = \sqrt{2}$, it is required that $\cos(\sum_{k=1}^n \theta_k) = \sin(\sum_{k=1}^n \theta_k) = \frac{1}{2}\sqrt{2}$, which excludes half of the previous solutions. Therefore the constraint becomes: $\sum_{k=1}^n \theta_k = \frac{\pi}{4} + \ell 2\pi$, $\ell \in \mathbb{Z}$, which completes the proof. $\square$

5.2.2 Enforcing additional vanishing moments

To impose additional vanishing moments is possible, but substantially more complicated since the parameters $\theta_1, \dots, \theta_n$ enter the equations in a nonlinear way. As an example the condition to have a second vanishing moment is now worked out.

Theorem 5.2.5. *In order for the wavelet low-pass filter resulting from the lattice structure to have two vanishing moments, for the parameters $\theta_1, \dots, \theta_n$ it must hold in addition to (5.15) that*

$$\sum_{k=1}^{n-1} \left(\cos \left(2 \sum_{\ell=1}^k \theta_\ell \right) \right) + \frac{1}{2} = 0 \quad (5.16)$$

Proof. The condition of the first vanishing moment $H_1(1) = 0$ with the corresponding condition on the parameters $\{\theta_k\}$ in (5.15) is assume to hold. For the second vanishing moment the additional condition $H'_1(1) = 0$ must hold. Since $H(z^2) \begin{pmatrix} 1 \\ z^{-1} \end{pmatrix} = \begin{pmatrix} H_0(z) \\ H_1(z) \end{pmatrix}$, we have that:

$$\frac{d}{dz} \begin{pmatrix} H_0(z) \\ H_1(z) \end{pmatrix} = 2zH'(z^2) \begin{pmatrix} 1 \\ z^{-1} \end{pmatrix} + H(z^2) \begin{pmatrix} 0 \\ -z^{-2} \end{pmatrix}. \quad (5.17)$$

From (5.7) we find:

$$H(1) = \begin{pmatrix} \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} \\ \frac{1}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2} \end{pmatrix}, \quad (5.18)$$

so that the last term of (5.17) at $z = 1$ yields:

$$H(1) \begin{pmatrix} 0 \\ -1 \end{pmatrix} = \begin{pmatrix} -\frac{1}{2}\sqrt{2} \\ \frac{1}{2}\sqrt{2} \end{pmatrix}. \quad (5.19)$$

Now we determine $H'(z)$:

$$\begin{aligned} H'(z) &= R(\theta_n)\Lambda'(z)R(\theta_{n-1})\Lambda(z)R(\theta_{n-2})\Lambda(z) \dots \Lambda(z)R(\theta_1)\Lambda(-1) \\ &\quad + R(\theta_n)\Lambda(z)R(\theta_{n-1})\Lambda'(z)R(\theta_{n-2})\Lambda(z) \dots \Lambda(z)R(\theta_1)\Lambda(-1) \\ &\quad + R(\theta_n)\Lambda(z)R(\theta_{n-1})\Lambda(z)R(\theta_{n-2})\Lambda'(z) \dots \Lambda(z)R(\theta_1)\Lambda(-1) \\ &\quad + \dots \\ &\quad + R(\theta_n)\Lambda(z)R(\theta_{n-1})\Lambda(z)R(\theta_{n-2})\Lambda(z) \dots \Lambda'(z)R(\theta_1)\Lambda(-1) \end{aligned} \quad (5.20)$$

Since we are interested in $H'(z^2)$ at $z = 1$ we note that $\Lambda(1) = I$, $\Lambda(z)$ becomes identity and $\Lambda'(1) = \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}$. The following is obtained:

$$\begin{aligned} H'(1) &= \sum_{k=1}^{n-1} \left(\begin{pmatrix} \cos \left(\sum_{\ell=k+1}^n \theta_\ell \right) & -\sin \left(\sum_{\ell=k+1}^n \theta_\ell \right) \\ \sin \left(\sum_{\ell=k+1}^n \theta_\ell \right) & \cos \left(\sum_{\ell=k+1}^n \theta_\ell \right) \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix} \right. \\ &\quad \left. \begin{pmatrix} \cos \left(\sum_{\ell=1}^k \theta_\ell \right) & -\sin \left(\sum_{\ell=1}^k \theta_\ell \right) \\ \sin \left(\sum_{\ell=1}^k \theta_\ell \right) & \cos \left(\sum_{\ell=1}^k \theta_\ell \right) \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right) \\ &= \sum_{k=1}^{n-1} \begin{pmatrix} \sin \left(\sum_{\ell=k+1}^n \theta_\ell \right) \sin \left(\sum_{\ell=1}^k \theta_\ell \right) & -\sin \left(\sum_{\ell=k+1}^n \theta_\ell \right) \cos \left(\sum_{\ell=1}^k \theta_\ell \right) \\ -\cos \left(\sum_{\ell=k+1}^n \theta_\ell \right) \sin \left(\sum_{\ell=1}^k \theta_\ell \right) & \cos \left(\sum_{\ell=k+1}^n \theta_\ell \right) \cos \left(\sum_{\ell=1}^k \theta_\ell \right) \end{pmatrix}. \end{aligned} \quad (5.21)$$

Now $H'(1) \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is determined:

$$\begin{aligned}
H'(1) \begin{pmatrix} 1 \\ 1 \end{pmatrix} &= \sum_{k=1}^{n-1} \begin{pmatrix} \sin \left(\sum_{\ell=k+1}^n \theta_\ell \right) \left(\sin \left(\sum_{l=1}^k \theta_l \right) - \cos \left(\sum_{l=1}^k \theta_l \right) \right) \\ \cos \left(\sum_{\ell=k+1}^n \theta_\ell \right) \left(\cos \left(\sum_{l=1}^k \theta_l \right) - \sin \left(\sum_{l=1}^k \theta_l \right) \right) \end{pmatrix}, \\
&= \sum_{k=1}^{n-1} \begin{pmatrix} \sqrt{2} \sin \left(\sum_{\ell=k+1}^n \theta_\ell \right) \sin \left(\sum_{l=1}^k \theta_l - \frac{\pi}{4} \right) \\ \sqrt{2} \cos \left(\sum_{\ell=k+1}^n \theta_\ell \right) \sin \left(\frac{\pi}{4} - \sum_{l=1}^k \theta_l \right) \end{pmatrix}, \\
&= \sum_{k=1}^{n-1} \begin{pmatrix} -\sqrt{2} \sin^2 \left(\frac{\pi}{4} - \sum_{l=1}^k \theta_l \right) \\ \sqrt{2} \cos \left(\frac{\pi}{4} - \sum_{l=1}^k \theta_l \right) \sin \left(\frac{\pi}{4} - \sum_{l=1}^k \theta_l \right) \end{pmatrix}, \\
&= \sum_{k=1}^{n-1} \begin{pmatrix} -\sqrt{2} \sin^2 \left(\frac{\pi}{4} - \sum_{l=1}^k \theta_l \right) \\ \frac{1}{2} \sqrt{2} \sin \left(\frac{\pi}{2} - 2 \sum_{l=1}^k \theta_l \right) \end{pmatrix}, \\
&= \sum_{k=1}^{n-1} \begin{pmatrix} -\sqrt{2} \sin^2 \left(\frac{\pi}{4} - \sum_{l=1}^k \theta_l \right) \\ \frac{1}{2} \sqrt{2} \cos \left(2 \sum_{l=1}^k \theta_l \right) \end{pmatrix}. \tag{5.22}
\end{aligned}$$

Due to Theorem 5.2.4 it also holds that:

$$H'(1) \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \sum_{k=1}^{n-1} \begin{pmatrix} \frac{1}{2} \sqrt{2} \cos \left(2 \sum_{\ell=k+1}^n \theta_\ell \right) - \frac{1}{2} \sqrt{2} \\ \frac{1}{2} \sqrt{2} \sin \left(2 \sum_{\ell=k+1}^n \theta_\ell \right) \end{pmatrix}. \tag{5.23}$$

Since the condition for the second vanishing moment is $D'(1) = 0$ the bottom row of (5.17) needs to be equal to zero. From (5.17), (5.19) and (5.22) we find that the condition for the second vanishing moment comes down to:

$$\sum_{k=1}^{n-1} \left(\cos \left(2 \sum_{\ell=1}^k \theta_\ell \right) \right) + \frac{1}{2} = 0, \tag{5.24}$$

which completes the proof. $\square$

The reader should note that this condition cannot be conveniently enforced by eliminating a parameter. As a concrete example, consider the case with three parameters ($n = 3$) of which θ_1 is fixed in order to enforce a first vanishing moment. In this case the condition (5.16) becomes:

$$\cos(2\theta_2 + 2\theta_3) + \cos(2\theta_3) + \frac{1}{2} = 0. \tag{5.25}$$

Given that the condition for the first vanishing moment is satisfied it may happen that it is not possible to satisfy the condition for the second vanishing moment by fixing another parameter. For $n = 3$, the curve implicitly parameterized by (5.24) is displayed in Figure 5.4.

Figure 5.4: Curve where (5.25) is satisfied. The dot marks the location of the Daubechies 3 wavelet in the parameter space.

5.2.3 Design and optimization

The parameterized lattice structure discussed previously allows for the design of wavelets, where orthogonality is built into the model class. The parameters $\theta_1, \dots, \theta_n$ can be chosen to optimize a certain design goal as discussed in Section 5.1. The issues involved with optimizing over the previously discussed parameter space are discussed in this section.

Local optima

With the criteria described in Section 5.1, combined with the theory of polyphase filters and the lattice structure, it is possible to construct an orthogonal wavelet that is optimal in the sense of a chosen criterion function. To compute this orthogonal wavelet, a local search method can be employed to search the parameter space $\theta_1, \dots, \theta_n$ for an optimal wavelet. As often happens with local search techniques, there is a risk that the optimization may terminate in a local optimum, as will be illustrated in the following example.

Example 5.2.6. The smoothed ECG signal from Figure 5.1a is used as a prototype signal. The number of parameters is chosen to be $n = 3$, yielding two degrees of freedom and a resulting filter of length 6. As design criterion the l_1 -norm of the wavelet coefficients of the wavelet transform is minimized. A single vanishing moment was enforced. During the experimentation we discovered a number of local minima, as illustrated in Figures 5.5, 5.6 and 5.7.

Figure 5.5: The existence of local minima

In Figure 5.6 and Figure 5.7 the local optima are further investigated in detail. In the upper left corner of the figure the location of the local optima is displayed in a contour plot of the l_1 -criterion over the free parameter space. In the other subfigures the wavelet and scaling function corresponding to the optima in the contour plot are displayed. From this figure it can be seen that some of these local optima can be quite similar, where one is a flipped or shifted version of another one, yielding almost the same result. Note that local optimum 3 is so close to number 8 that their difference is negligible. It is contained in the same gully due to the periodicity of the parameter space. The fact that this local optimum is found from both sides of the parameter space substantiates that this is indeed a local optimum and not a border artefact. If one considers the low-pass filter $[-0.0016, -0.1225z^{-1}, 0.2217z^{-2}, 0.8359z^{-3}, 0.4870z^{-4}, -0.0062z^{-5}]$ and the high-pass filter $[-0.0062, -0.4870z^{-1}, 0.8359z^{-2}, -0.2217z^{-3}, -0.1225z^{-4}, 0.0016z^{-5}]$, associated with this local optimum, it is easy to see that these effectively have a filter of length four, since leading or trailing coefficients are not significantly large.

Observation 5.2.7. *An interesting consequence of the preceding example is that if one compares the significant filter coefficients of local optima 3 and 8 with the low-pass filter of the Daubechies 2 wavelet: $[-0.1294, 0.2241z^{-1}, 0.8365z^{-2}, 0.4830z^{-3}]$ and high-pass*

Figure 5.6: Local optima in detail. The wavelet and scaling functions corresponding to the local optima found in the design of a wavelet for the signal in Figure 5.1a using three free parameters.

Figure 5.7: The location of the local optima with respect to the free parameters θ_2 and θ_3 as in Figure 5.6 is displayed.

Figure 5.8: Algorithm to find an optimal wavelet

filter of the Daubechies 2 wavelet: $[-0.4830, 0.8365z^{-1}, -0.2241z^{-2}, -0.1294z^{-3}]$, they are nearly identical. This observation provides a rationale for the use of the Daubechies 2 wavelet for ECG processing.

It can be concluded that local optima are present in the search space associated with the parameterization described in this section and the criteria described in Section 5.1. The problem of finding a global optimum in the presence of local optima is very hard. There are a number of ways in which these local optima can be avoided. There are a number of techniques that attempt to find a global optimum in the presence of local optima such as for example simulated annealing [71], genetic algorithms [42] and branch and bound methods [6]. However in general these methods cannot guarantee that a global optimum can be found and even that an optimum that has been found is indeed a global or local optimum. In light of these drawbacks and the applications at hand it was decided to run the local search algorithm multiple times with a number of different, randomly selected starting points as a heuristic of finding a global optimum.

Optimization algorithm

As discussed in the previous section a local search technique was employed and this search algorithm was used a number of times with random starting points. This approach is illustrated in Figure 5.8, where the blocks with solid borders visualize the local search method and the blocks with the dashed borders represent the loop with random starting points. This approach is implemented in `Matlab`, using the implementation of the Nelder-Mead direct search (simplex) algorithm [73] that is available in the function `fminsearch`

of the Matlab Optimization Toolbox.

5.2.4 Experimentation

In order to verify the effectiveness of the proposed approach a number of tests were conducted.

Test for reproducibility

In order to validate the approach, first consider that if a signal has a sparse representation in the wavelet domain (when transformed with a certain wavelet), this wavelet is likely to give a good performance for the criteria discussed earlier with respect to the wavelet decomposition of the signal in question. Note that although both the criteria V_1 and V_4 aim at maximizing the variance, they still may give different results. However when starting with a sparse representation in the wavelet domain and reconstructing an artificial signal, the chance that differently shaped wavelets perform equally well becomes smaller. As a validation of the approach, we start with a sparse set of wavelet coefficients and wavelet filter. The associated time series is used as an input for the wavelet design procedure, and it is verified whether the wavelet filter that is discovered as an optimum is equal to the wavelet filter used to reconstruct the signal from the wavelet coefficients.

As example of the test, first random parameters θ_2 and θ_3 are selected and θ_1 is fixed such that (5.15) holds and the corresponding sequences of scaling and wavelet filter coefficients (resp. H_0 and H_1) are determined. Next the full wavelet decomposition of a signal of length 256 is considered and only a few of these wavelet coefficients are assigned non-zero values. Carrying out the synthesis algorithm, we are constructing a signal in the wavelet domain. These wavelet coefficients are considered to be a sparse representation of an unknown signal x in terms of H_0 and H_1 . Now the signal x is determined by taking the sparse wavelet decomposition and performing an inverse wavelet transform on it with orthogonal wavelet filters H_0 and H_1 , i.e. the signal x is reconstructed from the sparse wavelet representation using the wavelet filters H_0 and H_1 .

We now have a signal x for which we know that H_0 and H_1 will give a sparse representation, thus this wavelet filter bank should perform well with respect to the optimization criterion and there should be a reasonably large chance that this wavelet filter bank is found as an optimum. To verify this, the signal x was taken and the optimal wavelet was determined with the approach described in this report. Optionally, before determining the optimal wavelet, additive white Gaussian noise was added to the signal x in order to test the effect of noise on the approach. The testing procedure is visualized in Figure 5.9 and the results for 250 trials are displayed in Table 5.1. If the l_1 -norm of the difference of the six actual and calculated filter coefficients is larger than 0.01, then the test is considered to be a failure.

Without noise the original wavelet was recovered in virtually all cases as can be seen in Table 5.1. If the SNR decreases a clear sparse representation of the signal in the wavelet domain may no longer exist but, in particular for the l_4 criterion, the original

Figure 5.9: Test for reproducibility

SNR	l_1 recovery rate	l_4 recovery rate
∞	99.2%	100%
40 dB	99.6%	99.6%
20 dB	32.4%	44.4%
10 dB	3.60%	9.60%
0 dB	0.00%	1.60%

Table 5.1: Percentage of times that a wavelet that provided a sparse representation of the prototype signal was recovered. Noise was added to the prototype signal yielding a Signal to Noise Ratio as indicated in the table.

wavelet was still found as an optimum in many cases. Taking the presence of local optima in consideration the results are fair.

Filter length versus criterion value

The filter length of the optimized wavelet is twice the number of parameters n . A large n gives more freedom to optimize the wavelet with respect to the criterion since in order to enforce the required first vanishing moment, a single parameter has to be fixed. However, as n increases, the complexity of the optimization increases as well, and it becomes more difficult to avoid local optima. The “curse of dimensionality” applies here: choosing random initial points to cover the domain becomes more and more cumbersome; there is an exponential increase of effort. To determine the effect of the choice of n on the criterion value that is achieved during the optimization a test was conducted. For the data in Figure 5.1 the optimal wavelet was determined for $n = 1, 2, \dots, 25$. For each choice of the number of parameters, i.e. for each choice of the filter length, a thousand random starting points were generated and the best criterion value in terms of both the

Figure 5.10: criterion value vs n for the l_1 criterion in (a) and for the l_4 criterion in (b). Note that the values of the l_4 criterion have been multiplied with -1 such that the optimization problem becomes a minimization problem.

l_1 and l_4 criterion were stored. The results are illustrated in Figure 5.10. From this figure it can be concluded that it is not beneficial for the given signal and criterion to use more than eight parameters. The increase in criterion value for a large number parameters is a consequence of the existence of local optima.

5.3 Multiwavelet parameterization and design

If one wants to discriminate between two features in a signal with a different morphology then multiwavelets (Section 3.4) [119, 116, 76, 77, 107] are a powerful tool. As discussed in Section 3.4 multiwavelets are a number of mutually orthogonal wavelet functions. In that section it was discussed that for orthogonal multiwavelets with compact support in a polyphase representation the polyphase matrices become lossless [116, 117] and FIR. These orthogonal multiwavelets can be designed using a parameterization as lossless systems, by jointly optimizing each multiwavelet function with respect to different segments of a prototype signal as first discussed in [95]. The class of orthogonal multiwavelets with compact support is parameterized with the results from [55]. The same design criteria as discussed in Section 5.1 are used for multiwavelet design.

First, in the following subsection, the parameterization for (multi)wavelets using lossless systems will be discussed. This employs interpolation theory and the theory of balanced realizations of lossless systems by means of the tangential Schur algorithm [55]. The approach consists of a number of steps. A key observation to start from is that a balanced realization of a lossless system has a unitary realization matrix. One way of parameterizing unitary realization matrices is by building them as a product of elementary unitary matrices. For this purpose, mappings $\mathcal{F}_{U,V}$ are introduced. These

mappings are a special simple kind of so called “linear fractional transformations”, which are extensively used in interpolation theory. The tangential Schur algorithm allows one to parameterize lossless transfer functions through a recursive procedure which involves linear fractional transformations; this procedure was cast in state-space terms in [55]. To parameterize real FIR lossless polyphase matrices, a number of special choices for the parameters in the tangential Schur algorithm can be made, which substantially simplify the expressions.

Then, in the next subsection, the use of the tangential Schur algorithm [55] to parameterize *scalar* wavelets is explained. It is shown how to recover the product formula (5.7) encountered in the lattice filter implementation by making suitable choices in the algorithm. The current set-up allows us to generalize this to the case of orthogonal *multiwavelets* with compact support. We also manage to build in a first “balanced vanishing moment”, which is much needed for practical purposes. This balanced vanishing moment was previously enforced in the literature as a constraint, but in this work a parameterization for multiwavelets with a built-in balanced vanishing moment is presented. In the last subsection a design approach for multiwavelets is discussed in which the multiwavelet parameterization is used to arrive at concrete results.

5.3.1 Parameterization of lossless systems

As discussed in Section 3.4 the condition of orthogonality for polyphase multiwavelet filters comes down to requiring that $H_p(z)$ is lossless, i.e., stable all-pass. We recall that for an arbitrary all-pass system $G(z)$ of size $p \times p$ it holds that:

$$G(e^{i\omega})^\dagger G(e^{i\omega}) = I_p, \quad \forall \omega \in \mathbb{R}. \quad (5.26)$$

Conversely, if (5.26) holds then $G(z)$ is all-pass. If $G(z)$ satisfies the following properties:

$$G(z)^\dagger G(z) = I_p \quad \text{for } |z| = 1, \quad (5.27)$$

$$G(z)^\dagger G(z) \geq I_p \quad \text{for } |z| < 1, \quad (5.28)$$

$$G(z)^\dagger G(z) \leq I_p \quad \text{for } |z| > 1, \quad (5.29)$$

where each two equations imply the remaining third one, then $G(z)$ is also stable, thus lossless.

For any proper rational matrix function $R(z)$ we introduce the following notation:

$$R^\dagger(z) = R(\bar{z})^\dagger = R_0^\dagger + R_1^\dagger z^{-1} + R_2^\dagger z^{-2} + \dots \quad (5.30)$$

For $z = e^{i\omega}$ it holds that $R^\dagger(z^{-1}) = R(z)^\dagger$. Consequently an all-pass system $G(z)$ has the property that an inverse exists and is given by:

$$G(z)^{-1} = G^\dagger(z^{-1}), \quad (5.31)$$

for $z = e^{i\omega}$, hence for *all* complex z by analytical continuation (see for example [74]).

When considering a $2r \times 2r$ lossless polyphase multiwavelet filter $H_p(z)$ with multiplicity r , as described in Section 3.4, it can be partitioned as:

$$H_p(z) = \left(\begin{array}{c|c} H_{0,e}(z) & H_{0,o}(z) \\ \hline H_{1,e}(z) & H_{1,o}(z) \end{array} \right), \quad (5.32)$$

with $H_{0,e}(z)$, $H_{0,o}(z)$, $H_{1,e}(z)$ and $H_{1,o}(z)$ the even and odd parts of the low- and high-pass multiwavelet filters, respectively. Due to (3.58) and (5.31) we have that:

$$\left(\begin{array}{c|c} H_{0,e}(z) & H_{0,o}(z) \\ \hline H_{1,e}(z) & H_{1,o}(z) \end{array} \right) \left(\begin{array}{c|c} H_{0,e}^\dagger(z^{-1}) & H_{1,e}^\dagger(z^{-1}) \\ \hline H_{0,o}^\dagger(z^{-1}) & H_{1,o}^\dagger(z^{-1}) \end{array} \right) = \left(\begin{array}{c|c} I_r & 0 \\ \hline 0 & I_r \end{array} \right). \quad (5.33)$$

These conditions must be exactly met in order to generate a valid orthogonal multiwavelet structure. Although they can be enforced by adding them as constraints to an optimization routine during the multiwavelet design, this is not convenient as they are nonlinear and they will often be satisfied merely approximately. A better and more elegant way is to find a parameterization in which these conditions are built-in. Procedures for the recursive construction of lossless systems exist in the literature; the procedures for building all-pass systems from [55] will be used in the remainder of this chapter to parameterize orthogonal multiwavelets.

A construction method for $p \times p$ lossless systems of order n will now be described. When compact support is required, corresponding to the FIR property of the filters $H_0(z)$, $H_1(z)$ as well as $H_p(z)$, special choices can be made to parameterize this subclass too. Additional properties such as a (balanced) vanishing moment can also be incorporated. However, the problem of building in vanishing moments of higher order, and in fact of finding the appropriate balancing conditions, is currently not completely solved [95, 27, 77, 76, 109, 23, 108].

A key result which underlies the parameterization procedure for lossless systems is the following. From [55, Proposition 3.2] it follows that if the realization matrix associated with a realization (A, B, C, D) of some rational function $G(z)$ is unitary, then $G(z)$ will be lossless. If A is additionally asymptotically stable then (A, B, C, D) will actually be a minimal and balanced representation. Recall that a balanced realization satisfies the Lyapunov-Stein equations (4.43)-(4.45), while in the balanced lossless case it holds that $P = Q = I_n$.

Theorem 5.3.1. *Let (A, B, C, D) (with dimensions $n \times n$, $n \times p$, $p \times n$ and $p \times p$ respectively) be a balanced state-space realization of an n^{th} order lossless p -input, p -output transfer function $G(z) = D + C(zI_n - A)^{-1}B$. Then the $(p+n) \times (p+n)$ realization matrix $R = \begin{pmatrix} D & C \\ B & A \end{pmatrix}$ is unitary.*

Conversely, let R be such a unitary block-partitioned matrix, then $G(z) = D + C(zI_n - A)^{-1}B$ is $p \times p$ lossless of degree less than or equal to n . It is of degree n if and only if A is asymptotically stable, in which case the realization is balanced.

For a proof see [56, 55]. This theorem allows the problem of parameterizing lossless functions to be studied in terms of unitary realization matrices, associated with balanced

realizations of lossless functions. Such approach was first carried out in [56] for scalar (1×1) lossless transfer functions, where a balanced canonical form was constructed with a corresponding upper triangular reachability matrix. The associated realization matrix allows for a factorization into a product of n more simple unitary building blocks. This approach was generalized in [55] to multivariable $(p \times p)$ lossless functions transfer functions of degree n . There, it is described how to parameterize such lossless systems using the tangential Schur algorithm. The associated canonical forms (A, B, C, D) are parameterized with the associated Schur parameter vectors, while other quantities serve to index local coordinate charts. In the multivariable case instead of unitary matrix multiplications, so-called linear fractional transformations are employed to set up a recursive procedure to construct the realization matrix. It was also found for which choices in the tangential Schur algorithm this procedure in fact reduces to unitary matrix multiplications for realization matrices, which is useful from a practical viewpoint. These procedures will now be described, where the following definitions are helpful.

On the level of transfer functions, for a proper rational matrix $G(z)$, we introduce a mapping $\mathcal{F}_{U,V}$:

$$\mathcal{F}_{U,V} : G(z) \rightarrow \tilde{G}(z) \quad (5.34)$$

defined by:

$$\tilde{G}(z) = F_1(z) + \frac{F_2(z)F_3(z)}{z - F_4(z)}, \quad (5.35)$$

where

$$\left(\begin{array}{c|c} F_1(z) & F_2(z) \\ \hline F_3(z) & F_4(z) \end{array} \right) = V \left(\begin{array}{c|ccc} 1 & 0 & \dots & 0 \\ \hline 0 & & & \\ \vdots & & G(z) & \\ 0 & & & \end{array} \right) U^\dagger, \quad (5.36)$$

and U and V are $(p+1) \times (p+1)$ matrices, $F_1(z)$ is a $p \times p$ matrix, $F_2(z)$ a $p \times 1$ vector, $F_3(z)$ a $1 \times p$ vector and $F_4(z)$ a scalar.

Additionally, it is proven in [55] that if $G(z)$ is proper then $\tilde{G}(z)$ will be well-defined. Note that lossless transfer functions are always proper.

When using unitary matrices U and V , the mapping $\mathcal{F}_{U,V}(G(z))$ takes lossless functions to lossless functions. In fact, it constructs a lossless transfer function $\tilde{G}(z)$ with an order at most one larger than the order of the lossless system $G(z)$.

Theorem 5.3.2. *Let U, V be two unitary $(p+1) \times (p+1)$ matrices and let $G(z)$ be $p \times p$ lossless of degree n . Then $\tilde{G}(z) = \mathcal{F}_{U,V}(G(z))$ is also lossless of degree less than or equal to $n+1$.*

For a proof see [55].

On the level of balanced state-space realizations of lossless transfer functions, the mapping $\mathcal{F}_{U,V}$ can be implemented as described by the following result (see again [55]):

Theorem 5.3.3. *Let U, V be two unitary $(p+1) \times (p+1)$ matrices and let $G(z)$ be $p \times p$ lossless of degree n , with a balanced realization (A, B, C, D) . Then $\tilde{G}(z) = \mathcal{F}_{U,V}(G(z))$*

has the realization $(\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D})$ given by:

$$\left(\begin{array}{c|c} V & 0 \\ \hline 0 & I_n \end{array} \right) \left(\begin{array}{c|cc} 1 & 0 & 0 \\ \hline 0 & D & C \\ 0 & B & A \end{array} \right) \left(\begin{array}{c|c} U^\dagger & 0 \\ \hline 0 & I_n \end{array} \right) = \left(\begin{array}{c|c} \tilde{D} & \tilde{C} \\ \hline \tilde{B} & \tilde{A} \end{array} \right), \quad (5.37)$$

where $\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D}$ respectively have dimensions $(n+1) \times (n+1)$, $(n+1) \times p$, $p \times (n+1)$ and $p \times p$; hence the state-space dimension is $n+1$. It is minimal and balanced if and only if $\tilde{A}$ happens to be asymptotically stable.

This theory allows for a parameterization of lossless systems in state-space by iteratively applying unitary matrix multiplications. It is also possible to parameterize lossless systems in terms of transfer functions by using so-called “linear fractional transformations” as in the more general framework of interpolation theory, more specifically using the tangential Schur algorithm. A lossless function $g(z)$ increases in McMillan degree by exactly one into a lossless function $\tilde{G}(z)$, in each step of this tangential Schur algorithm.

The tangential Schur algorithm employs linear fractional transformations (LFT) associated with J -inner matrices. In [55] LFTs are considered of the form:

$$\mathcal{T}_\Theta : G \rightarrow (\Theta_4 G + \Theta_3)(\Theta_2 G + \Theta_1)^{-1}, \quad (5.38)$$

where $\Theta = \begin{pmatrix} \Theta_1 & \Theta_2 \\ \Theta_3 & \Theta_4 \end{pmatrix}$ is a $2p \times 2p$ block-partitioned rational matrix function of McMillan degree m , with blocks of size $p \times p$ and G is a $p \times p$ rational matrix of McMillan degree n .

Theorem 5.3.4. *If $G(z)$ is $p \times p$ lossless of degree n and $\Theta(z)$ is $2p \times 2p$ J -inner of degree m , then $\tilde{G}(z) = \mathcal{T}_{\Theta(z)}(G(z))$ is $p \times p$ lossless of degree less than or equal to $n+m$.*

Here the function $\Theta(z)$ is called J -inner if it holds that:

$$\Theta(z)^\dagger J \Theta(z) = J \quad \text{for } |z| = 1, \quad (5.39)$$

$$\Theta(z)^\dagger J \Theta(z) \leq J \quad \text{for } |z| < 1, \quad (5.40)$$

$$\Theta(z)^\dagger J \Theta(z) \geq J \quad \text{for } |z| > 1, \quad (5.41)$$

with $J = \begin{pmatrix} I_p & 0 \\ 0 & -I_p \end{pmatrix}$. For a proof see [55, Proposition 4.4]. Note that the inverse of a J -inner function $\Theta(z)$ is given by:

$$\Theta(z)^{-1} = J \Theta^\dagger(z^{-1}) J. \quad (5.42)$$

In the tangential Schur algorithm, a rational $p \times p$ lossless function of McMillan degree n is reduced in n recursion steps to such a function of degree 0; that is, to a constant unitary matrix. The degree is reduced by $m = 1$ in each step and the J -inner functions associated with these steps are also of McMillan degree one. Such function are called “elementary J -inner factors”.

The elementary J -inner factors that are employed in [55] are those which have their pole outside the closed unit disk at $z = \bar{w}^{-1}$. They are represented as:

$$\Theta(u, v, w, \xi, M)(z) = \left(I_{2p} + \left(\frac{\left(\frac{z-w}{1-\bar{w}z} \right)}{\left(\frac{\xi-w}{1-\bar{w}\xi} \right)} - 1 \right) \frac{\begin{bmatrix} u \\ v \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}^\dagger J}{(\|u\|^2 - \|v\|^2)} \right) M. \quad (5.43)$$

In this expression, the matrix $M \in \mathbb{C}^{2p \times 2p}$ is a J -unitary constant matrix, $u \in \mathbb{C}^{p \times 1}$ with $\|u\| = 1$ is a normalized direction vector, $v \in \mathbb{C}^{p \times 1}$ with $\|v\| < 1$ is a Schur vector, $w \in \mathbb{C}$ with $|w| < 1$ is an interpolation point and $\xi \in \mathbb{C}$ with $|\xi| = 1$ is a normalization point on the unit circle. Note that $\Theta(u, v, w, \xi, M)(\xi) = M$.

From [55, Proposition 5.3] we have that:

Theorem 5.3.5. *Let $G(z)$ be $p \times p$ lossless of McMillan degree n , then $\tilde{G}(z) = \mathcal{T}_{\Theta(u, v, w, \xi, H)(z)}(G(z))$ is $p \times p$ lossless of McMillan degree $n + 1$, satisfying the interpolation condition $\tilde{G}(\bar{w}^{-1})u = v$.*

Conversely from [55, Proposition 5.4] we have:

Theorem 5.3.6. *Let $\tilde{G}(z)$ be $p \times p$ lossless of McMillan degree $n + 1$, let $w \in \mathbb{C}$ be such that $|w| < 1$, let $u \in \mathbb{C}^{p \times 1}$ with $\|u\| = 1$ be such that $v := \tilde{G}(\bar{w}^{-1})u$ satisfies $\|v\| < 1$. Let $\xi \in \mathbb{C}$ satisfy $|\xi| = 1$ and $M \in \mathbb{C}^{2p \times 2p}$ be constant J -unitary. Then there exists a unique lossless $p \times p$ function $G(z)$ of McMillan degree n such that $\tilde{G}(z)$ can be obtained as $\tilde{G}(z) = \mathcal{T}_{\Theta(u, v, w, \xi, H)(z)}(G(z))$.*

In each step of the parameterization procedure, the order is increased by one, by choosing values for w, u, ξ and M and by letting v vary over the space of vectors in $\mathbb{C}^p$ with $\|v\| < 1$. The interpolation conditions make clear that this gives us coordinate charts for the space, i.e. the manifold, of $p \times p$ lossless functions of degree n . Multiple charts are needed to cover the manifold; it can be shown that no single Euclidean coordinate chart is capable of covering the manifold entirely.

In our implementation, a lossless system is preferably built in state-space using a recursion, i.e., the space of $p \times p$ lossless function is parameterized by application of the reversed tangential Schur algorithm, carried over to the state-space level, starting from a constant $p \times p$ unitary matrix. To achieve this, we describe which choices in the tangential Schur algorithm allow an LFT $\tilde{G}(z) = \mathcal{T}_{\Theta(u, v, w, \xi, H)(z)}(G(z))$ to be represented as a mapping $\tilde{G}(z) = \mathcal{F}_{u, v}(G(z))$ for suitable unitary matrices U and V . This is discussed in [55, Theorem 6.4].

Theorem 5.3.7. *If $\mathcal{T}_{\Theta(u, v, w, \xi, M)(z)}$ coincides with a mapping $\mathcal{F}_{U, V}$ with U and V unitary, then*

$$\Theta(u, v, w, \xi, M)(z) = \mathcal{H}(uv^\dagger)S_{u, w}(z)\mathcal{H}(\bar{w}uv^\dagger) \begin{pmatrix} P & 0 \\ 0 & Q \end{pmatrix}, \quad (5.44)$$

for some $u, v \in \mathbb{C}^p$ with $\|u\| = 1$, $\|v\| < 1$, some $w \in \mathbb{C}$ with $|w| < 1$, and some $p \times p$ unitary matrices P and Q . Equivalently it holds that

$$M = \mathcal{H}(uv^\dagger)S_{u,w}(\xi)\mathcal{H}(\bar{w}uv^\dagger) \begin{pmatrix} P & 0 \\ 0 & Q \end{pmatrix}. \quad (5.45)$$

In that case one can take $U = \hat{U} \begin{pmatrix} 1 & 0 \\ 0 & P \end{pmatrix}$ and $V = \hat{V} \begin{pmatrix} 1 & 0 \\ 0 & Q \end{pmatrix}$, where

$$\hat{U} = \begin{pmatrix} \frac{\sqrt{1-|w|^2}}{\sqrt{1-|w|^2}\|v\|^2}u & I_p - (1 + \frac{w\sqrt{1-\|v\|^2}}{\sqrt{1-|w|^2}\|v\|^2})uu^\dagger \\ \frac{\bar{w}\sqrt{1-\|v\|^2}}{\sqrt{1-|w|^2}\|v\|^2} & \frac{\sqrt{1-|w|^2}}{\sqrt{1-|w|^2}\|v\|^2}u^\dagger \end{pmatrix}, \quad (5.46)$$

$$\hat{V} = \begin{pmatrix} \frac{\sqrt{1-|w|^2}}{\sqrt{1-|w|^2}\|v\|^2}v & I_p - (1 - \frac{\sqrt{1-\|v\|^2}}{\sqrt{1-|w|^2}\|v\|^2})\frac{vv^\dagger}{\|v\|^2} \\ \frac{\sqrt{1-\|v\|^2}}{\sqrt{1-|w|^2}\|v\|^2} & -\frac{\sqrt{1-|w|^2}}{\sqrt{1-|w|^2}\|v\|^2}v^\dagger \end{pmatrix}. \quad (5.47)$$

In this theorem, the notation $\mathcal{H}(E)$ is used to denote the Halmos extension of a strictly contractive $p \times p$ matrix E :

$$\mathcal{H}(E) = \begin{pmatrix} (I_p - EE^\dagger)^{-1/2} & E(I_p - E^\dagger E)^{-1/2} \\ E^\dagger(I_p - EE^\dagger)^{-1/2} & (I_p - E^\dagger E)^{-1/2} \end{pmatrix}. \quad (5.48)$$

It holds that $\mathcal{H}(E)$ is Hermitian, J -unitary and invertible with inverse $\mathcal{H}(E)^{-1} = \mathcal{H}(-E)$.

Also, the matrix function $S_{u,w}(z)$ is defined by:

$$S_{u,w}(z) := \begin{pmatrix} I_p + \left(\frac{z-w}{1-\bar{w}z} - 1\right)uu^\dagger & 0 \\ 0 & I_p \end{pmatrix}. \quad (5.49)$$

For lossless FIR systems, the special case $w = 0$ is important. In that case:

$$\hat{U} = \begin{pmatrix} u & I_p - uu^\dagger \\ 0 & u^\dagger \end{pmatrix}, \quad (5.50)$$

$$\hat{V} = \begin{pmatrix} v & I_p - (1 - \sqrt{1-\|v\|^2})\frac{vv^\dagger}{\|v\|^2} \\ \sqrt{1-\|v\|^2} & -v^\dagger \end{pmatrix}. \quad (5.51)$$

In the following subsection we will discuss the use of the tangential Schur algorithm to parameterize scalar wavelets. It is shown how to recover the product formula (5.7) by making suitable choices in the algorithm.

5.3.2 Parameterization of scalar wavelets

In Section 5.2 a lattice filter implementation of 2×2 lossless systems was presented. The associated parameterization procedure employs the product formula (5.7), which can be generated through a recursion which involves premultiplication in each iteration step:

$$G^{(k)}(z) = R_k \Lambda(z) G^{(k-1)}(z) \quad (5.52)$$

$$G^{(1)}(z) = R_1 \Lambda(-1). \quad (5.53)$$

Here, the function $G^{(k)}(z)$ denotes a lossless function of degree $k-1$, parameterized by the k parameters $\theta_1, \theta_2, \dots, \theta_k$. When attempting to rewrite this in terms of the LFTs and the tangential Schur algorithm of the previous subsection, it is noted that postmultiplication is more conveniently represented in that formalism than premultiplication. For that reason, we address the transposed expressions, which are given by:

$$\left(G^{(k)}(z)\right)^T = \left(G^{(k-1)}(z)\right)^T \Lambda(z) R_k^T \quad (5.54)$$

$$\left(G^{(1)}(z)\right)^T = \Lambda(-1) R_1^T. \quad (5.55)$$

From these equations it is not difficult to read off a valid interpolation condition and to construct an associated J -inner matrix $\Theta^{(k)}(z)$ which allows one to rewrite an iteration step in terms of an LFT:

$$\left(G^{(k)}(z)\right)^T = \mathcal{T}_{\Theta^{(k)}(z)} \left(\left(G^{(k-1)}(z)\right)^T \right), \quad (5.56)$$

for

$$\Theta^{(k)}(z) = \begin{pmatrix} (\Lambda(z) R_k^T)^{-1} & 0 \\ 0 & I_2 \end{pmatrix} = \begin{pmatrix} R_k \Lambda(z^{-1}) & 0 \\ 0 & I_2 \end{pmatrix} \quad (5.57)$$

The following choices can then be made for the parameters in (5.43), with the recursion index k running from $k = 2$ to $k = n$, to recover (5.7) for the scalar wavelet case:

$$u_k = \begin{pmatrix} \sin(\theta_k) \\ -\cos(\theta_k) \end{pmatrix} \quad (5.58)$$

$$v_k = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (5.59)$$

$$w_k = 0 \quad (5.60)$$

$$\xi_k = 1 \quad (5.61)$$

$$M_k = \begin{pmatrix} \cos(\theta_k) & -\sin(\theta_k) & 0 & 0 \\ \sin(\theta_k) & \cos(\theta_k) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (5.62)$$

M_k is a block diagonal matrix with the unitary matrices $P_k = R_k$ and $Q_k = I_2$ on its diagonals. Choosing $w_k = 0$ and $v_k = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ in each recursion step, implies that $\left(G^{(k)}(z)\right)^T$ has all its poles at $z = 0$, thus ensuring the FIR property for this polyphase filter. (This corresponds to a Potapov decomposition of $\left(G^{(k)}(z)\right)^T$.) The parameter θ_k shows up in the normalized direction vector u_k , which gives a non-standard choice for parameterizing lossless functions. The standard choice to parameterize lossless functions requires u_k , w_k , ξ_k and M_k to be fixed and v_k to contain the free parameters subject to

the constraint $\|v_k\| < 1$. The current non-standard choice, where v_k is kept fixed and u_k is varied, does produce a parameterization however which works properly, although some lossless functions can be obtained for different choices of parameters.

It is now straightforward to verify that the LFT in each recursion step admits a representation as a mapping $\mathcal{F}_{U,V}$. Referring to Theorem 5.3.7 we have that:

$$\begin{aligned}\widehat{U} &= \begin{pmatrix} \frac{\sqrt{1-|w_k|^2}}{\sqrt{1-|w_k|^2}\|v_k\|^2} u_k & I_p - \left(1 + \frac{w_k \sqrt{1-\|v_k\|^2}}{\sqrt{1-|w_k|^2}\|v_k\|^2}\right) u_k u_k^\dagger \\ \frac{\bar{w}_k \sqrt{1-\|v_k\|^2}}{\sqrt{1-|w_k|^2}\|v_k\|^2} & \frac{\sqrt{1-|w_k|^2}}{\sqrt{1-|w_k|^2}\|v_k\|^2} u_k^\dagger \end{pmatrix} \\ &= \begin{pmatrix} \sin(\theta_k) & \cos^2(\theta_k) & \sin(\theta_k) \cos(\theta_k) \\ -\cos(\theta_k) & \sin(\theta_k) \cos(\theta_k) & \sin^2(\theta_k) \\ 0 & \sin(\theta_k) & -\cos(\theta_k) \end{pmatrix}. \end{aligned} \quad (5.63)$$

$$\begin{aligned}\widehat{V} &= \begin{pmatrix} \frac{\sqrt{1-|w_k|^2}}{\sqrt{1-|w_k|^2}\|v_k\|^2} v_k & I_p - \left(1 + \frac{\sqrt{1-\|v_k\|^2}}{\sqrt{1-|w_k|^2}\|v_k\|^2}\right) \frac{v_k v_k^\dagger}{\|v_k\|^2} \\ \frac{\sqrt{1-\|v_k\|^2}}{\sqrt{1-|w_k|^2}\|v_k\|^2} & \frac{-\sqrt{1-|w_k|^2}}{\sqrt{1-|w_k|^2}\|v_k\|^2} v_k^\dagger \end{pmatrix} \\ &= \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}. \end{aligned} \quad (5.64)$$

The matrices U and V that are used in the mappings $\mathcal{F}_{U,V}$ in the steps of the corresponding state-space recursion (5.37) are given by:

$$U = \widehat{U} \left(\begin{array}{c|cc} 1 & 0 & 0 \\ \hline 0 & & \\ 0 & P_k & \end{array} \right) = \left(\begin{array}{cc|c} \sin(\theta_k) & \cos(\theta_k) & 0 \\ -\cos(\theta_k) & \sin(\theta_k) & 0 \\ \hline 0 & 0 & -1 \end{array} \right) \quad (5.65)$$

$$V = \widehat{V} \left(\begin{array}{c|cc} 1 & 0 & 0 \\ \hline 0 & & \\ 0 & Q_k & \end{array} \right) = \left(\begin{array}{c|cc} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \hline 1 & 0 & 0 \end{array} \right). \quad (5.66)$$

The resulting state-space recursion (5.37) then attains the following form:

$$\begin{aligned} & \left(\begin{array}{c|c} D^{(k)} & C^{(k)} \\ \hline B^{(k)} & A^{(k)} \end{array} \right) \\ &= \left(\begin{array}{cc|cccc} \cos(\theta_k) D_{1,1}^{(k-1)} & \sin(\theta_k) D_{1,1}^{(k-1)} & -D_{1,2}^{(k-1)} & C_{1,1}^{(k-1)} & \cdots & C_{1,k-1}^{(k-1)} \\ \cos(\theta_k) D_{2,1}^{(k-1)} & \sin(\theta_k) D_{2,1}^{(k-1)} & -D_{2,2}^{(k-1)} & C_{2,1}^{(k-1)} & \cdots & C_{2,k-1}^{(k-1)} \\ \hline \sin(\theta_k) & -\cos(\theta_k) & 0 & 0 & \cdots & 0 \\ \cos(\theta_k) B_{1,1}^{(k-1)} & \sin(\theta_k) B_{1,1}^{(k-1)} & -B_{1,2}^{(k-1)} & & & \\ \vdots & \vdots & \vdots & & A^{(k-1)} & \\ \cos(\theta_k) B_{k-1,1}^{(k-1)} & \sin(\theta_k) B_{k-1,1}^{(k-1)} & -B_{k-1,2}^{(k-1)} & & & \end{array} \right) \end{aligned} \quad (5.67)$$

From the previous equation it can be seen that the dynamical matrix $A^{(k-1)}$ is extended by means of a zero row and a first column that is equal to the second column of $B^{(k-1)}$. Hence $A^{(k)}$ is strictly lower triangular, having a zero diagonal. Consequently its eigenvalues are zero and the system is FIR.

Note that (5.67) yields a balanced realization of $G^{(k)}(z)^T$ and it may be transposed in order to obtain a balanced realization of $G^{(k)}(z)$. It follows that $A^{(k)T}$ is a strictly upper triangular matrix: it is in real Schur form. (Balancing leaves the freedom of the orthogonal group which allowed us to bring $A^{(k)T}$ into that real Schur form.) Since the real Schur form is not necessarily unique it is not a canonical form; this is a consequence of the non-standard parameterization mentioned earlier.

5.3.3 Parameterization of multiwavelets

The parameterization in the previous section can be extended to the multiwavelet case. For the parameters of the elementary J -inner factors from (5.43) the following choices are made: Since we are building FIR filters we take $w = 0$ and $v = (0 \dots 0)^T$. The parameter ξ can be chosen freely on the unit circle; the choice $\xi = 1$ is made so that $S_{u,w}(\xi) = I_{4r}$. In order to again ensure that $\mathcal{T}_{\Theta(u,v,w,\xi,M)}(z)$ coincides with a mapping $\mathcal{F}_{U,V}$ as in Theorem 5.3.7, due to the convenient choice of ξ we need to have that $M = \begin{pmatrix} P & 0 \\ 0 & Q \end{pmatrix}$.

Theorem 5.3.8. *Let $G^{(k+1)}(z)$ be a FIR all-pass filter of order $k+1$. Then with the choice of $w_k = 0$, $v_k = (0 \dots 0)^T$, $\|u_k\| = 1$, $\xi_k = 1$, P and Q unitary, it follows that $G^{(k+1)}(z)$ can be factored as:*

$$G^{(k+1)}(z) = (I_{2r} + (z^{-1} - 1)u_k u_k^\dagger) P_k G^{(k)}(z) Q_k^\dagger. \quad (5.68)$$

where $G^{(k)}(z)$ is a FIR all-pass filter of order k .

Proof. Obtaining a state-space recursion for a block diagonal matrix G^k involves a factorization of $\Theta^{(k)}(z)$ as in [55, Proposition 5.2]:

$$\Theta^{(k)}(u_k, v_k, w_k, \xi_k, M_k)(z) = \mathcal{H}(u_k v_k^\dagger) S_{u_k, w_k}(z) S_{u_k, w_k}(\xi_k)^{-1} \mathcal{H}(u_k v_k^\dagger)^{-1} M_k. \quad (5.69)$$

Also observe that for the choices made [55, Theorem 6.4] holds, such that:

$$\Theta^{(k)}(u_k, v_k, w_k, \xi_k, M_k)(z) = \widehat{\Theta}^{(k)}(u_k, v_k, w_k)(z) \begin{pmatrix} P_k & 0 \\ 0 & Q_k \end{pmatrix}, \quad (5.70)$$

with

$$\widehat{\Theta}^{(k)}(u_k, v_k, w_k)(z) = S_{u_k, 0}(z) = \begin{pmatrix} I_{2r} + (z - 1)u_k u_k^\dagger & 0 \\ 0 & I_{2r} \end{pmatrix}. \quad (5.71)$$

Using the LFT formula in (5.38) we see that the inverse of the upper right quadrant of (5.71) is used as a post multiplying factor. However we are working on the transposed

system as noted in (5.56), such that after transposition the inverse factor appears as a pre-multiplying factor. This leads to the recursion formula:

$$G^{(k+1)}(z)^T = \mathcal{T}_{\Theta(u_k, v_k, w_k, \xi_k, M_k)(z)} \left(G^{(k)}(z)^T \right) = Q_k G^{(k)}(z)^T P_k^\dagger (I_{2r} + (z^{-1} - 1)u_k u_k^\dagger), \quad (5.72)$$

such that:

$$G^{(k+1)}(z) = (I_{2r} + (z^{-1} - 1)u_k u_k^\dagger) P_k G^{(k)}(z) Q_k^\dagger \quad (5.73)$$

□

Observe that this is a more general parameterization than previously available in the literature. One can obtain a lattice decomposition in terms of Givens rotation matrices from Theorem 5.3.8 with the following choices:

We relate the unitary matrix P to the vector u , which will again contain the parameters. Analogous to the scalar case we require that $P^\dagger u = -e_{2r}$ and any such P will do. Here e_n denotes the n^{th} standard basis vector. As in the scalar case we choose $Q = I_{2r}$.

For the parameter vector u it must again hold that $\|u\| = 1$ and as a choice for a fully parameterized u one can for example make:

$$u = \begin{pmatrix} \sin(\alpha_1) \sin(\alpha_2) \sin(\alpha_3) \\ -\cos(\alpha_3) \sin(\alpha_1) \sin(\alpha_2) \\ \cos(\alpha_2) \sin(\alpha_1) \\ -\cos(\alpha_1) \end{pmatrix}. \quad (5.74)$$

Corollary 5.3.9. *Let $G^{(k+1)}(z)$ be a FIR all-pass filter of order $k+1$. Then with the choice of $w_k = 0$, $v_k = (0 \dots 0)^T$, $\|u_k\| = 1$, $\xi_k = 1$, $Q = I_{2r}$ and $P_k^\dagger u_k = -e_{2r}$ it follows that $G^{(k+1)}(z)$ can be factored in a lattice decomposition as:*

$$G^{(k+1)}(z) = P_k \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & z^{-1} \end{pmatrix} G^{(k)}(z), \quad (5.75)$$

where $G^{(k)}(z)$ is a FIR all-pass filter of order k .

Proof. Observe that due to the fact that P_k is a unitary $2r \times 2r$ matrix and that $P_k u_k = -e_{2r}$, the following equalities hold:

$$u_k = -P_k e_{2r}, \quad (5.76)$$

$$u_k^\dagger P_k = -e_{2r}^\dagger P_k^\dagger P_k = -e_{2r}^\dagger, \quad (5.77)$$

Such that (5.68) can be rewritten as:

$$G^{(k+1)}(z) = P_k (I_{2r} + (z^{-1} - 1)e_{2r} e_{2r}^\dagger) G^{(k)}(z) Q_k^\dagger \quad (5.78)$$

$$= P_k \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & z^{-1} \end{pmatrix} G^{(k)}(z) Q_k^\dagger. \quad (5.79)$$

Since $Q_k = I_{2r}$ the following product formula is obtained:

$$G^{(k+1)}(z) = P_k \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & z^{-1} \end{pmatrix} G^{(k)}(z), \quad (5.80)$$

which completes the proof. $\square$

A practical choice for P relating to the choice of u as in (5.74) is to build it as a product of Givens rotation matrices as in for example [116]. Such a parameterization for $r = 2$ is given by:

$$P(\alpha_1, \alpha_2, \alpha_3) = \begin{pmatrix} \cos(\alpha_3) & -\sin(\alpha_3) & 0 & 0 \\ \sin(\alpha_3) & \cos(\alpha_3) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\alpha_2) & -\sin(\alpha_2) & 0 \\ 0 & \sin(\alpha_2) & \cos(\alpha_2) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos(\alpha_1) & -\sin(\alpha_1) \\ 0 & 0 & \sin(\alpha_1) & \cos(\alpha_1) \end{pmatrix}$$

This results in the following parameterized matrix P :

$$P = \begin{pmatrix} \cos(\alpha_3) - \cos(\alpha_2) \sin(\alpha_3) & \cos(\alpha_1) \sin(\alpha_2) \sin(\alpha_3) & -\sin(\alpha_1) \sin(\alpha_2) \sin(\alpha_3) \\ \sin(\alpha_3) \cos(\alpha_2) \cos(\alpha_3) & -\cos(\alpha_1) \cos(\alpha_3) \sin(\alpha_2) & \cos(\alpha_3) \sin(\alpha_1) \sin(\alpha_2) \\ 0 & \sin(\alpha_2) & \cos(\alpha_1) \cos(\alpha_2) & -\cos(\alpha_2) \sin(\alpha_1) \\ 0 & 0 & \sin(\alpha_1) & \cos(\alpha_1) \end{pmatrix}, \quad (5.81)$$

which indeed satisfies the condition $P_k^\dagger u_k = -e_{2r}$.

The parameterization as in the so-called “great factorization theorem”, cf. [107, 117], can be obtained with the choice of $P_k = Q_k = I_{2r}$.

Corollary 5.3.10. *Let $G^{(k+1)}(z)$ be a FIR all-pass filter of order $k+1$. Then with the choice of $w_k = 0$, $v_k = (0 \dots 0)^T$, $\|u_k\| = 1$, $\xi_k = 1$, $Q = I_{2r}$ and $P = I_{2r}$ it follows that $G^{(k+1)}(z)$ can be factored in a Householder decomposition as:*

$$G^{(k+1)}(z) = \left(I_{2r} - u_k u_k + u_k u_k^\dagger z^{-1} \right) G^{(k)}(z), \quad (5.82)$$

where $G^{(k)}(z)$ is a FIR all-pass filter of order k .

Proof. Consider (5.68) from Theorem 5.3.8. With the choice of $P_k = Q_k = I_{2r}$, (5.68) can be rewritten as:

$$G^{(k+1)}(z) = \left(I_{2r} - u_k u_k + u_k u_k^\dagger z^{-1} \right) G^{(k)}(z). \quad (5.83)$$

$\square$

Using the parameterization in Corollary 5.3.9 of multiwavelets as polyphase all-pass systems all requirements but the first (balanced) vanishing moment are enforced. In the next section it is discussed how a first balanced vanishing moment can be built in as first discussed in [95]. It will turn out that the first balanced vanishing moment can be enforced by initializing the recursion in a specific way. In Section 5.3.5 a step-by-step design procedure for orthogonal multiwavelets will be provided.

5.3.4 Balanced vanishing moments

When considering a constant signal, we may investigate how it is processed when employing Mallat's cascade algorithm [79]. For scalar wavelets, the approximation coefficients will all be identical, i.e., a constant coefficient vector. Therefore the sampled signal is commonly used to initialize the multi resolution analysis, because the sampled signal and the approximation coefficients constitute similar sequences. This avoids an actual computation of the scaling coefficients that express the input signal in terms of the basis spanned by shifted versions of the scaling function to initialize the multi resolution analysis and it also avoids the actual computation of the multiscaling and multiwavelet function.

When processing a constant signal in the multiwavelet case, one finds that each subsequence of scaling coefficients, for each of the multiwavelets, is also constant. However, when these constants differ per multiwavelet, then one cannot simply use the sampled signal to initialize the multi resolution analysis. The necessary condition to use a sampled signal to initialize the multiresolution analysis is that the low-pass synthesis operator should exactly preserve polynomials of order p for a p^{th} order balanced vanishing moment [77]. If this condition were not satisfied a preprocessing step would be required, which is unattractive. The alternative is to impose a so-called *balancing* condition on the multiwavelet structure and to allow again the direct use of the sampled signal.

A 0^{th} order balanced scaling function preserves constant signals [27]:

$$\int_{\mathbb{R}} 1\phi_{-[0]}(t)dt = \dots = \int_{\mathbb{R}} 1\phi_{-[r-1]}(t)dt. \quad (5.84)$$

Due to power complementarity this is equivalent to the condition that the detail coefficients are not affected by constant signals: $0 = \int_{\mathbb{R}} 1\psi_{[k]}(t)dt$. When taking the dilation equation (3.24) and wavelet equation (3.25) and integrating both sides over the real axis we obtain:

$$\zeta_0 = \int_{\mathbb{R}} \phi(t)dt = \sqrt{2} \sum_{k=0}^{2n-1} C_k \int_{\mathbb{R}} \phi(2t+k)dt, \quad (5.85)$$

$$\eta_0 = \int_{\mathbb{R}} \psi(t)dt = \sqrt{2} \sum_{k=0}^{2n-1} D_k \int_{\mathbb{R}} \phi(2t+k)dt. \quad (5.86)$$

The desired *vanishing moment* comes down to the condition $\eta_0 = 0$. The *balancing*

condition comes down to $\zeta_0 = c \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$, for some constant scalar c . Using the change of

basis $\tau = 2t + k$, $d\tau = 2dt$ and $dt = \frac{1}{2}d\tau$ we obtain:

$$\int_{\mathbb{R}} \underline{\phi}(2t + k) dt = \int_{\mathbb{R}} \underline{\phi}(\tau) \frac{1}{2} d\tau = \frac{1}{2} \zeta_0, \quad (5.87)$$

from which it follows that:

$$\zeta_0 = \frac{1}{2} \sqrt{2} H_0(1) \zeta_0, \quad (5.88)$$

$$\eta_0 = \frac{1}{2} \sqrt{2} H_1(1) \zeta_0. \quad (5.89)$$

This can be rewritten in terms of the polyphase matrix $H_p(z)$ as:

$$\begin{pmatrix} \zeta_0 \\ \eta_0 \end{pmatrix} = \frac{1}{2} \sqrt{2} H_p(1) \begin{pmatrix} \zeta_0 \\ \zeta_0 \end{pmatrix}. \quad (5.90)$$

Since $H_p(z)$ is all-pass and thus orthogonal at $z = 1$ we have that:

$$\frac{1}{2} \|H_p(1) \begin{pmatrix} \zeta_0 \\ \zeta_0 \end{pmatrix}\|^2 = \frac{1}{2} \left\| \begin{pmatrix} \zeta_0 \\ \zeta_0 \end{pmatrix} \right\|^2 = \|\zeta_0\|^2, \quad (5.91)$$

$$\left\| \begin{pmatrix} \zeta_0 \\ \eta_0 \end{pmatrix} \right\|^2 = \|\zeta_0\|^2 + \|\eta_0\|^2, \quad (5.92)$$

and since these expressions are equal due to (5.90) it follows that $\|\eta_0\|^2 = 0$ and $\eta_0 = 0$. It can then be concluded that the first vanishing moment is built-in, meaning that every feasible (multi)wavelet structure which fits the orthogonal set-up as discussed previously in this chapter, already obeys the conditions for a first vanishing moment. The conclusion here is that the first vanishing moment is implicitly required and restricts the class of lossless polyphase matrices $H_0(z), H_1(z)$. The conditions for a first *balanced* vanishing moment can be enforced in the tangential Schur algorithm by means of an interpolation condition on the unit circle. The condition on $H_p^{(k)}(z)$ for the first balanced vanishing moment is as in [95]:

Theorem 5.3.11. *If $H_p^{(k)}(z)$ is a $2r \times 2r$ real FIR all-pass filter of order k associated with the corresponding FIR filters $H_0(z), H_1(z)$ and vector functions $\underline{\phi}(t)$ and $\underline{\psi}(t)$, satisfying the dilation and wavelet equation, and $\sqrt{2}$ is a simple eigenvalue of $H_0(1)$, then $\underline{\psi}(t)$ has a balanced vanishing moment of order 0 if and only if*

$$(1, 1, \dots, 1 | 1, 1, \dots, 1) H_p^{(k)}(1)^T = \sqrt{2} (1, 1, \dots, 1 | 0, 0, \dots, 0) \quad (5.93)$$

Proof. This theorem directly follows from (5.90). Note that the condition that $\sqrt{2}$ is a simple eigenvalue of $H_0(1)$ ensures that (5.88) has a solution ζ_0 which is determined up to a nonzero scaling factor. Orthogonality of the multiresolution structure and sign conventions for $\underline{\phi}(t)$ and $\underline{\psi}(t)$ fix this scaling factor. $\square$

The vanishing moment condition is now rewritten in terms of $H_0(z)$, $H_1(z)$. The following relation holds:

$$\frac{1}{2}\sqrt{2}H_p(z^2) \begin{pmatrix} I_r & 0 \\ 0 & z^{-1}I_r \end{pmatrix} \begin{pmatrix} I_r & I_r \\ I_r & -I_r \end{pmatrix} = \frac{1}{2}\sqrt{2} \begin{pmatrix} H_0(z) & H_0(-z) \\ H_1(z) & H_1(-z) \end{pmatrix}. \quad (5.94)$$

This equation allows one to relate the derivatives of $H_p(z)$ to the derivatives of $H_0(z)$, $H_1(z)$, $H_0(-z)$ and $H_1(-z)$, in particular at $z = 1$. Additionally, due to orthogonality, we have that $H_1(1) = 0, H_1'(1) = 0, H_1''(1) = 0$ is equivalent to $H_0(-1) = 0, H_0'(-1) = 0, H_0''(-1) = 0$. This specifies the vanishing moment condition entirely in terms of the low-pass filter. Note that all matrices on the left hand side of (5.94) are real lossless matrices, hence the matrix on the right hand side is real lossless and thus orthogonal at $|z| = 1$ and more specifically at $z = 1$:

$$\frac{1}{2}\sqrt{2} \begin{pmatrix} H_0(1) & H_0(-1) \\ H_1(1) & H_1(-1) \end{pmatrix} = H_p(1)I_{2r}\frac{1}{2}\sqrt{2} \begin{pmatrix} I_r & I_r \\ I_r & -I_r \end{pmatrix}. \quad (5.95)$$

Using the right hand side of (5.95) and that $\eta_0 = 0$, (5.90) can be rewritten as:

$$\begin{pmatrix} \zeta_0 \\ 0 \end{pmatrix} = H_p(1)\frac{1}{2}\sqrt{2} \begin{pmatrix} I_r & I_r \\ I_r & -I_r \end{pmatrix} \begin{pmatrix} \zeta_0 \\ 0 \end{pmatrix} \quad (5.96)$$

$$= \frac{1}{2}\sqrt{2} \begin{pmatrix} H_0(1) & H_0(-1) \\ H_1(1) & H_1(-1) \end{pmatrix} \begin{pmatrix} \zeta_0 \\ 0 \end{pmatrix}. \quad (5.97)$$

One can investigate how this condition translates to $H_p(1)$ in the scalar case, i.e., $r = 1$. The conditions become: $\sqrt{2}\zeta_0 = H_0(1)\zeta_0$ and $0 = H_1(1)\zeta_0$. Since $\zeta_0 \neq 0$, it must hold that $H_0(1) = \sqrt{2}$ and $H_1(1) = 0$ and due to power complementarity that $H_0(-1) = 0$ and thus that $H_1(-1) = \pm\sqrt{2}$. The sign choice for $H_1(-1)$ corresponds to a sign choice for the wavelet $\underline{\psi}(t)$. We shall use the convention $H_1(-1) = \sqrt{2}$. Hence $H_p(1) = \frac{1}{2}\sqrt{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$. In the multiwavelet case balancing is required and we have to work with the scalar multiples.

Using (5.97), the condition for the first (0th order) vanishing moment becomes [77]:

$$\sqrt{2}(1, 1, \dots, 1)^T = H_0(1)(1, 1, \dots, 1)^T \quad (5.98)$$

$$(0, 0, \dots, 0)^T = H_1(1)(1, 1, \dots, 1)^T \quad (5.99)$$

As a side remark it can be verified that the results in this section are consistent with the condition for a first balanced vanishing moment in [77, Theorem 2]. Taking into account the differences in notation and terminology the balanced vanishing moment is characterized in [77] by the conditions:

$$(1, 1, \dots, 1)H_0(1) = \sqrt{2}(1, 1, \dots, 1) \quad (5.100)$$

$$(1, 1, \dots, 1)H_0(-1) = \sqrt{2}(0, 0, \dots, 0) \quad (5.101)$$

The conditions in (5.100) and (5.101) relate to the conditions in (5.98) and (5.99) due to the following lemma, which is an application to (5.93):

Lemma 5.3.12. *Consider an orthogonal real matrix N with a real eigenvalue λ and a corresponding real eigenvector v . Then the equality $Nv = \lambda v$ is equivalent to $v^T N = \lambda v^T$.*

Proof. Consider the following equality:

$$Nv = \lambda v \quad (5.102)$$

The norm on both sides needs to be equal as well:

$$v^T N^T N v = |\lambda|^2 v^T v, \quad (5.103)$$

$$v^T v = |\lambda|^2 v^T v, \quad (5.104)$$

hence $|\lambda|^2 = 1$ and thus for real eigenvalues we have $\lambda = 1$ or $\lambda = -1$. Now construct an orthogonal matrix N_2 which has v as its first column: $v = N_2 e_1$. We can then rewrite (5.102) as:

$$N N_2 e_1 = \lambda N_2 e_1 \quad (5.105)$$

$$N_2^T N N_2 e_1 = \lambda e_1 \quad (5.106)$$

Due to the fact that $|\lambda| = 1$, the orthogonal matrix $N_2^T N N_2$ will have the following structure:

$$N_2^T N N_2 = \left(\begin{array}{c|ccc} \lambda & 0 & \dots & 0 \\ \hline 0 & & & \\ \vdots & & N_3 & \\ 0 & & & \end{array} \right),$$

where N_3 is another orthogonal matrix. Hence we can write:

$$e_1^T N_2^T N N_2 = \lambda e_1^T, \quad (5.107)$$

$$e_1^T N_2^T N = \lambda e_1^T N_2^T, \quad (5.108)$$

$$v^T N = \lambda v^T, \quad (5.109)$$

which completes the proof. $\square$

A consequence of this lemma is that the transpose in (5.93) can be omitted.

The fact that the parameterization in Corollary 5.3.9 has the convenient property that $H_p^{(k+1)}(1) = H_p^{(k)}(1)$, makes it possible to build in a vanishing moment with respect to the multiwavelet in a straightforward manner.

Theorem 5.3.13. *An orthogonal multiwavelet of multiplicity r having a balanced vanishing moment and a FIR filter of order k can be built as a polyphase all-pass $H_p^{(k)}(z)$ by initializing the factorization in Corollary 5.3.9 with a zeroth order polyphase all-pass $H_p^{(0)}$ that is factored as:*

$$H_p^{(0)T} = \mathcal{H}_\alpha \begin{pmatrix} 1 & 0 \\ 0 & L \end{pmatrix} \mathcal{H}_\beta, \quad (5.110)$$

where $\mathcal{H}_\alpha$ and $\mathcal{H}_\beta$ are fixed orthogonal Householder matrices of the form:

$$\mathcal{H}_\alpha = I_{2r} - 2 \frac{\alpha \alpha^T}{\alpha^T \alpha}, \quad (5.111)$$

with:

$$\alpha^T = (1, 1, \dots, 1 | 0, 0, \dots, 0) / \sqrt{r} - (1, 0, \dots, 0 | 0, 0, \dots, 0) \quad (5.112)$$

$$\beta^T = (1, 1, \dots, 1 | 1, 1, \dots, 1) / \sqrt{2r} - (1, 0, \dots, 0 | 0, 0, \dots, 0), \quad (5.113)$$

and L is an arbitrary $(2r - 1) \times (2r - 1)$ orthogonal matrix.

Proof. First observe that α and β are of the form $\alpha = \frac{a-b}{\|a-b\|}$. The Householder reflection $\mathcal{H}_\alpha x$ reflects the component of x in the direction a to b . Components that are orthogonal to a are left unchanged.

Since (5.71) shows that $\hat{\Theta}^{(k)}(u_k, v_k, w_k)(1) = I_{2r}, \forall k$ interpolation conditions on $H_p^{(k)}(1)$ that are built-in in H_0 are preserved throughout the recursion. It is now left to correctly initialize this recursion.

The interpolation condition of interest is the balancing condition from Theorem 5.3.11:

$$H_p^{(0)}(1, 1, \dots, 1 | 1, 1, \dots, 1)^T = (1, 1, \dots, 1 | 0, 0, \dots, 0)^T. \quad (5.114)$$

Now suppose that we factor $H_p^{(0)}$ as the product of three matrices $H_p^{(0)} = WRU$, with W and U symmetrical. Then we get the condition:

$$WRU(1, 1, \dots, 1 | 1, 1, \dots, 1)^T = (1, 1, \dots, 1 | 0, 0, \dots, 0)^T, \quad (5.115)$$

which can be rewritten as (using the symmetry property):

$$RU(1, 1, \dots, 1 | 1, 1, \dots, 1)^T = W(1, 1, \dots, 1 | 0, 0, \dots, 0)^T. \quad (5.116)$$

If the matrix W is chosen to be $\mathcal{H}_\alpha$ then the right hand side of (5.116) simplifies to e_1 and similarly if U is chosen to be $\mathcal{H}_\beta$ then the left hand side of (5.116) simplifies to Re_1 as a direct consequence of the theory on Householder transformations. Thus with this choice of W and U any matrix of the following form satisfies the condition in (5.116):

$$\begin{pmatrix} 1 & 0 \\ 0 & L \end{pmatrix}. \quad (5.117)$$

Since we want to obtain an orthogonal system the initial matrix $H_p^{(0)}$ and thus L has to be orthogonal. Since it is multiplied with orthogonal matrices during the recursion orthogonality will be preserved. $\square$

Note that such an orthogonal matrix L can be parameterized by $(2 - 1)(2r - 1)$ free angular parameters in a straightforward way. For example for $r = 2$ a 3×3 matrix Q can be parameterized as:

$$\begin{pmatrix} 1 & 0 \\ 0 & L \end{pmatrix} = \begin{pmatrix} I_2 & 0 \\ 0 & R(\theta_3) \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & R(\theta_2) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} I_2 & 0 \\ 0 & R(\theta_1) \end{pmatrix}. \quad (5.118)$$

$$R(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \quad (5.119)$$

5.3.5 Multiwavelet design

An explicit algorithm for constructing a multiwavelet filter of order n with $r = 2$ is given by:

1. Initialize the recursion:

- a) Select initial angles $\theta_1^0, \theta_2^0, \theta_3^0$.
- b) Construct the matrix $\begin{pmatrix} 1 & 0 \\ 0 & L \end{pmatrix}$ from (5.118) using parameters $\theta_1^0, \theta_2^0, \theta_3^0$.
- c) Construct the matrices U and W from (5.116) as:

$$U = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \end{pmatrix} \quad (5.120)$$

$$W = \begin{pmatrix} \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} & 0 & 0 \\ \frac{1}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (5.121)$$

- d) Construct the zeroth order all-pass $H_p^{(0)}$ as:

$$H_p^{(0)} = WRU. \quad (5.122)$$

2. Set $k = 1$ and enter the recursion

3. Recursion step

- a) Select angles $\theta_1^k, \theta_2^k, \theta_3^k$.
- b) Use these angles to construct the matrix P_k as in (5.81).
- c) Recursively construct a k^{th} order multiwavelet filter $H_p^{(k)}(z)$ from a $(k-1)^{\text{th}}$ order filter in accordance to Corollary 5.3.9:

$$H_p^{(k)} = P_k \begin{pmatrix} I_3 & 0 \\ 0 & z^{-1} \end{pmatrix} H_p^{(k-1)} \quad (5.123)$$

4. If $k < n$ then set $k = k + 1$ and goto 3.

The parameterization of multiwavelets as stable all-pass systems in the previous section allows for the design of multiwavelets for a specific purpose. The discussion on a suitable criterion for wavelet design as in Section 5.1 applies here as well since energy preservation holds when considering the overall multiwavelet filter.

Figure 5.11: Windowing for multiwavelet design

The orthogonality between the different filters in multiwavelets make them conceptually an appealing tool to distinguish and detect features simultaneously. In that respect it is opportunistic to employ r time-scale masks (as in Figure 5.11) to the wavelet coefficients. So for a multiwavelet with $r = 2$ two time-scale masks are used such that two features can be discriminated. When using these masks for optimization the conservation of energy assumption no longer holds. Also it is not convenient to use a criterion (such as l_1 minimization) that involves minimization since it is possible that the optimization routine converges to an optimum that is an effective delay such that features are pushed outside the mask. It is still possible to use the l_4 maximization, however since we want to put the energy at a specific place it is also possible to use l_2 maximization. Another point of interest is that for detection purposes it is most convenient to have full resolution at each scale, i.e. to use an undecimated multiwavelet transform.

As was the case for scalar wavelet design the optimization is performed by means of a local search technique with random starting points. The choice of parameterization ensures that all required conditions in order to obtain an orthogonal stable all-pass with at least one vanishing moment is built-in. However for applications such as filtering

additional smoothness might be needed. For the smoothness of multiwavelet functions a vanishing moment is not sufficient. Additionally the multiwavelet needs to be balanced [76, 77, 27]. Balancing requires an additional interpolation condition that is difficult to handle in the current parameterization and is still an open problem. It is still not guaranteed that the resulting system is indeed a multiwavelet since it is not enforced that each set of wavelet coefficients forms a high-pass filter and the associated scaling coefficients form a low-pass filter.

Chapter 6

Biomedical applications of orthogonal (multi)wavelet design

In this chapter the potential of the approaches, as introduced in Chapter 5, will be demonstrated by means of a number of practical applications. In Section 6.1 two practical applications of (multi)wavelet design with respect to cardiac signal processing are discussed. In Section 6.1.1 it is demonstrated how scalar wavelet design is useful for the detection of QRS complexes in ECG signals. In Section 6.1.2 the length of the QT interval is estimated using designed multiwavelets. In Section 6.2 the use of wavelet design for an application in 2D image analysis is discussed. This describes the removal an artifact in magnetic resonance imaging, called the “bias field”, using specially designed wavelets [64].

6.1 Applications of orthogonal (multi)wavelet design in cardiology

6.1.1 Detecting the QRS complex using orthogonal wavelet design

Using the wavelet design approach in Chapter 5 it is possible to design a wavelet for a given signal of interest. Given an ECG signal (see Section 2.3) one can average the beats in the signal and use them as a prototype for the wavelet design algorithm. If an ECG signal consisting of multiple beats was used as a prototype, the wavelet design would have searched for a wavelet that performs well for a superposition of the beats in the ECG signal. Hence it is justified to construct a prototype by aligning the ECG beats. The annotated R peak is used as a reference for the alignment. Since the QRS complex is the dominant part of an ECG beat, the location of this complex is generally not hard to detect. However as discussed in Section 2.2, the heart is a 3D organ. As such, the signals that are measured extracardiacally may exhibit different morphologies, dependent on the position of the lead, relative to the heart, complicating the detection.

Figure 6.1: Wavelets used for the detection of the QRS complex in episode 103.

Furthermore patient-to-patient variation and pathologies may affect the morphology of ECG signals.

In order to assess whether the wavelet design procedure from Section 5.2 indeed yields a beneficial distribution of the energy in the wavelet domain for the detection of the QRS complex, the ECG beat from Figure 5.1(a) was used as a prototype signal for wavelet design for two to five parameters and for both criteria. This prototype is a superposition of the normal beats from episode 103 of the MIT-BIH arrhythmia database [43]. Next the stationary wavelet transform (see Section 3.3, [91]) was calculated for 2^{19} samples of episode 103 of the MIT-BIH arrhythmia database. The choice of this length is a result of the used implementation of the SWT that accepts only signal with a length that is a power of two. For the detection step only a single scale was taken into account. Note that a sparse representation in the wavelet domain may yield a certain pattern in space-scale that can be used as a detector for a given morphology. A QRS complex was detected if the absolute value of the wavelet coefficients at that particular scale exceeded a fixed threshold value. This threshold value was chosen for each wavelet and each scale such that

	c_0	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9
db2	0.4830	0.8365	0.2241	-0.1294						
$l_1, 2p$	0.5010	0.8313	0.2061	-0.1242						
$l_1, 3p$	0.1860	0.7394	0.6336	-0.0607	-0.1125	0.0283				
$l_1, 4p$	-0.1077	0.1915	0.7849	0.5750	0.0167	-0.0668	0.0132	0.0074		
$l_1, 5p$	0.0159	0.0279	-0.0615	0.1402	0.7140	0.6652	0.0593	-0.1380	-0.0206	0.0117
$l_4, 2p$	0.1039	0.7868	0.6032	-0.0797						
$l_4, 3p$	-0.1111	0.1612	0.8158	0.5443	0.0024	0.0017				
$l_4, 4p$	-0.1143	0.2268	0.8451	0.4699	-0.0190	0.0128	-0.0047	-0.0024		
$l_4, 5p$	-0.1017	0.2915	0.8666	0.3894	-0.0311	0.0023	-0.0170	0.0273	-0.0097	-0.0034

Table 6.1: Low-pass filters used for the detection of the QRS complex in episode 103. The number before “p” indicates the number of parameters used to design the filter.

the sum of the false negatives and false positives was minimized. Since the effectiveness of the approach is evaluated here and not the robustness of a detector, this is a valid choice for the goal at hand. If the threshold value was exceeded within 20 samples (55ms) of an annotated QRS complex this was accepted as a valid detection. If it was outside this interval a false positive was reported. This is a very basic detection scheme since we do not aim to find a detector for the QRS complex which is nearly flawless, or even tries to find the exact location of the complex, but we are assessing the potential of the wavelet design approach. Note that the false positives may be clotted together; in this case all of them are still considered individual false positives. As reference wavelet the Daubechies 2 wavelet was chosen due to the fact that it is quite popular (see e.g. [105, 114, 93, 1]) and that the steep slopes in the signal are captured well by the Daubechies 2 wavelet function.

All designed wavelets and the Daubechies 2 wavelet were able to correctly find all or all but one of the QRS complexes at level 3 and 4 of the wavelet transform in the 1690 beats of episode 103 that were considered. Upon examining the wavelet and scaling functions in Figure 6.1, we see that indeed the (possibly mirrored) Daubechies 2 wavelet was found as an optimum for some of the designed wavelets, even though in some of these cases a longer filter size was used. From the corresponding low-pass filter coefficients in Table 6.1, we see that for a large number of higher order filters a number of coefficients are close to zero and that the remaining coefficients are in the same order of magnitude as for the Daubechies 2 wavelet. This corresponds to Observation 5.2.7.

Next, as a second test, a dataset was used which exhibits more variation in terms of morphology. The selected dataset: episode 215 of the MIT-BIH arrhythmia database [43] does not only contain normal beats, but also atrial premature contractions and premature ventricular contractions. The normal beats as illustrated in Figure 6.2 have a different morphology than the normal beats of episode 103 as displayed in Figure 5.1(a). Of episode 215 an excerpt consisting of 2718 beats was considered. The results are displayed in Table 6.2.

From Table 6.2 it can be seen that the l_1 designed wavelets perform better than the Daubechies 2 wavelet for this dataset. The results for the l_4 designed wavelets are not as good. The Daubechies 3 wavelet performs quite well. It can however be seen from the table that an increase in order for the Daubechies wavelets does not automatically yield an improvement in terms of detection performance. Among the higher order wavelets,

Figure 6.2: Smoothed normal beat from episode 215 of the MIT-BIH database.

wavelet	false neg.	false pos.	total errors	level	thres.
Daubechies 2	9	3	12	4	295
Daubechies 3	4	3	7	4	243
Daubechies 4	20	0	20	4	311
Daubechies 5	8	4	12	4	256
Daubechies 6	23	5	28	4	304
Daubechies 7	78	27	105	4	367
Daubechies 8	71	17	88	4	347
l_1 , 2 parameters	8	1	9	4	255
l_1 , 4 parameters	8	3	11	4	271
l_1 , 5 parameters	6	0	6	4	252
l_1 , 6 parameters	4	0	4	4	240
l_1 , 8 parameters	3	0	3	4	230
l_4 , 2 parameters	18	0	18	4	380
l_4 , 3 parameters	9	4	13	4	303
l_4 , 4 parameters	8	4	12	4	294

Table 6.2: Detection of the QRS complex in 2718 beats of episode 215 of the MIT-BIH arrhythmia database. The used wavelet filter, the number of missed beats, the number of false detections, the total number of errors, the used wavelet scale and the used threshold value are displayed in the columns respectively.

Figure 6.3: Wavelets used for the detection of the QRS complex in episode 215 of the MIT-BIH database.

the designed wavelets tend to perform better in this table. Summarizing, for low-order wavelet filters, the property to have vanishing moments may thus be beneficial in a morphologically mixed set, however once a required number of vanishing moments has been reached one can gain more by maximizing sparsity instead of imposing additional vanishing moments. The required vanishing moments can be incorporated in the parameterization as discussed in Chapter 5. The designed wavelets are displayed in Figure 6.3. Once again wavelets strongly resembling the Daubechies 2 wavelets have been found as an optimum relatively often.

6.1.2 QT time measurement using designed multiwavelets

As explained in Section 2.3 ECG signals can be separated into various segments. The QRS segment physically corresponds with the depolarization of the ventricles, which causes the ventricles to contract, and meanwhile atrial repolarization occurs, but due to the muscle mass of the ventricles the ventricular activity will be dominant. During the

Figure 6.4: Example of an ECG beat from the PTB database. The beat is extended to have a length that is a power of two. In order to cope with the end points of the signal when calculating the wavelet transform, periodic extension is employed.

following ST segment the ventricular muscle cells hold their action potential resulting in an iso-electric segment. The T wave corresponds to the ventricular repolarization. The start of the Q wave until the end of the T wave thus corresponds to the total ventricular activation (depolarization and repolarization) and gives the duration of the electrical systole. As previously discussed in [70] the QT interval is relevant for a large number of medical applications. QT prolongation is considered an indicator for sudden cardiac death, see [99]. The QT interval is used to calculate the beat-to-beat variability of repolarization (BVR) which currently is an important indicator for some cardiac pathologies, see [111]. The QT interval is also linked to non-cardiac pathologies such as diabetic autonomic neuropathy, see e.g. [41].

The accurate measurement of the QT time is a challenging task and was the subject of the PhysioNet/Computers in Cardiology Challenge 2006 [88]. Many recent methods [75, 57, 4] attempt to exploit the information that is contained in multiple leads since it shows a different projection of the same phenomenon (see Section 2.3).

Using the orthogonal multiwavelet design approach from Section 5.3 it is possible to construct a multiwavelet such that wavelet function in the multiwavelet structure is designed for the QRS complex and the other wavelet function in the multiwavelet structure is designed for the T wave, aiming at distinguishing them in the wavelet domain.

In Figure 6.5 the design of a multiwavelet for the simultaneous detection of both QRS complexes and T waves is displayed. In the top left of the figure the stationary wavelet decomposition with the first wavelet in the multiwavelet structure is shown. The white dashed rectangle is a time-scale mask, used during the optimization, where the time component corresponds to the QRS complex. In the lower left subfigure the absolute values of wavelet coefficients of this decomposition are displayed. In the top right of Figure 6.5 the stationary wavelet decomposition using the second wavelet in the multiwavelet structure is displayed. The rectangle is the time-scale mask used during the design, where the time component corresponds to the location of the T wave.

Because a large number of different morphologies can manifest in an ECG signal (see Section 2.3) it is interesting to see whether a multiwavelet can be designed that is

Figure 6.5: Design of a multiwavelet for the simultaneous detection of the QRS complex and the T wave for the ECG beat in Figure 6.4. In the upper two figures the detail coefficients are given at full resolution (SWT) from coarse to fine, and in the lower two figures the corresponding absolute values. The solid rectangle is a time-scale mask that indicated the region of interest with respect to the QRS complex. The dashed rectangle is the mask corresponding to the T wave.

applicable in a number of different situations (morphologies). To this end the criterion for wavelet design has to be modified a bit. Instead of calculating the multiwavelet decomposition of a single prototype signal and calculating the l_1 or the l_4 norm of this decomposition, the l_1 or l_4 norms of the decompositions of a vector of prototype signals is calculated yielding a vector of criterion values $(v_1, v_2, \dots, v_k)$, where k is the number of prototypes. Next these norms have to be merged into a single design criterion. This can be accomplished by e.g. taking the sums of the entries in this vector $\sum_{\ell=1}^k v_k$, taking the maximum over the entries of this vector $\max(v)$ or taking the product of the sum and the maximum of the entries of this vector $\max(v) \cdot \sum_{\ell=1}^k v_k$. In Figure 6.6 a multiwavelet that has been designed in this way for a vector of prototypes is displayed. A test signal was constructed for this figure that stitches a number of different morphologies together. As a reference the decomposition using the Daubechies 2 wavelet was used. Although a large number of T waves exhibit improved visibility some of them are still not clearly visible. This applies in particular to the third T wave in Figure 6.6 which has a low amplitude.

Since the aim is to detect the Q onset and T end points one could argue that the wavelet is to be designed to detect these points. For this purpose a new multiwavelet

Figure 6.6: Design of a multiwavelet for QRS and T wave detection in a range of morphologies. (a) decomposition with Daubechies 2 wavelet, (b) decomposition with first multiwavelet, (c) decomposition with second multiwavelet, (d) test signal

was designed, for adapted location of the two masks. This is illustrated in Figure 6.7. As can be observed in this figure, the detection of the location of the T end point from this decomposition is not convenient.

For this reason a testing algorithm was developed that performs a multiwavelet transform on a band-pass filtered ECG signal, considering only a single lead, and then detects the presence of the QRS complex and the T wave from multiple scales of the stationary multiwavelet decomposition. The exact location of the QRS onset and T end points is then determined by a rule-based system that involves the derivative of the single lead ECG signal. Preliminary results of the current algorithm show that it can detect the location of the Q onset of 548 beats from the PTB database [16] with a standard deviation of the error of $\sigma = 15.22\text{ms}$, for the Q onset point, $\sigma = 30.29\text{ms}$ for the T end point and $\sigma = 29.87\text{ms}$ for the QT time, with respect to the used gold standard [28] which consists of manual annotations for beats in the PTB database as published in [25]. It is also found that the means of the errors are quite high (-2.44ms , -22.82ms and -20.37ms) respectively and indicate a systematic error that can be corrected at the end of the algorithm and as a result is not a good measure of the performance of the algorithm. The

Figure 6.7: Design of a multiwavelet for Q onset and T end detection in a range of morphologies. (a) decomposition with Daubechies 2 wavelet, (b) decomposition with first multiwavelet, (c) decomposition with second multiwavelet, (d) test signal

obtained result is not spectacular but most reasonable (see e.g. [75, 57, 4]). Note that when considering a single lead, the choice of the lead may affect the results [82].

6.2 Bias field removal from magnetic resonance images using wavelet design

The wavelet design procedures from Chapter 5 can also be used for 2D applications, such as image processing, by taking the Cartesian product of 1D wavelet transforms as in [79]. As a concrete example, bias field removal from magnetic resonance images is considered as published in [64].

Magnetic Resonance Imaging (MRI), or Nuclear Magnetic Resonance Imaging (NMRI) as it was called in the early days, is widely used in the medical practice. It provides a much better contrast for soft tissue than for example x-ray / Computed Tomography (CT). The technique uses a magnetic field that is generated by coils to align the magnetization of hydrogen atoms. The alignment of this magnetization is altered with radio waves such that a rotating magnetic field is created that is picked up by the scanner.

The technology is still undergoing major improvements, both when it comes to hardware performance (such as field strength increase to improve the visibility of tissue anomalies) and software development (e.g. for sophisticated MR image processing). An example of such a new development is 4D acquisition: a time-lapsed 3D image of a subject.

6.2.1 Radio Frequency inhomogeneities in magnetic resonance images

An important type of artifact of interest is the *Radio Frequency (RF) inhomogeneity* or *bias field artifact*; see e.g. [7]. This type of artifact affects MR images and may seriously hamper reliable diagnosis in practice. The artifact is caused by the fact that the intensity of the machine magnetic field varies, depending on the scan sequence, tissues being imaged and the type of coil being used. Regarding the coil type, mainly surface coil images suffer from this type of artifact (whereas body coil images suffer less from this type of artifact, but contain more noise). The effect of RF homogeneity artifacts is that the image suffers from a non-uniform illumination. This expresses itself as, for example, extended luminescence or an image shade variation, but also as white stripes as illustrated in for example Figure 6.8.

6.2.2 Bias field removal in magnetic resonance images

One possible approach to remove this artifact, is to first image a phantom (a container with a known substance) and then to use this to determine the specific magnetic field intensity variations, see [32, 113]. From the phantom image a degradation model is computed, which is then used to account for a bias field in other MRI images obtained under comparable circumstances. In [64], which is used as a basis for this section, an approach was described that is intended to facilitate bias field removal from a corrupted image *without* prior knowledge of the specific field variations. A closely related approach with similar goals has previously been described in [7, 8] using classical Butterworth filters.

Figure 6.8: Knee MRI corrupted with bias field

6.2.3 Wavelet design for RF inhomogeneity detection

In the literature the RF inhomogeneity is commonly viewed and treated as *multiplicative noise*. As discussed in [7, 45] one can apply *homomorphic filtering* by taking the logarithm of each pixel ($\ln(\cdot)$) of the corrupted image ($C' = \ln(C)$) first, to make the bias field additive such that it can more easily be removed. In our case the image values are in the range 0–255, and since the logarithm of zero is undefined the image values are increased by 1.

In order to suppress the bias field from MR images in a wavelet-based approach, in [64] wavelets are designed by extending the approach discussed in [69] from 1D to 2D. This extension is carried out by taking the Cartesian product of 1D wavelet transforms as in [79]. The design is performed by computing the 2D wavelet decomposition of a prototype image and optimizing it according to an application dependent place-scale mask. Since the bias field in the present application covers the whole image, and it is a low frequency artifact, a scale mask is used which uniformly takes only the coarsest scales into account.

As a prototype, the natural logarithm of an artificial bias field F_B from BrainWeb [18, 72, 29] was used: $F_C = \ln(F_B)$. To optimize the wavelet design criterion, the 2D wavelet transform of F_C was calculated over 5 scales in the following way. First, the 1D wavelet transform was calculated in the horizontal direction of each row of F_C giving arrays $a1$ and $d1$. Next the 1D wavelet transforms were calculated in the vertical directions of $a1$ and $d1$, giving arrays $aa1$ and $ad1$ for $a1$, and $da1$ and $dd1$ for $d1$. This splits the image into an approximation part $aa1$ and a horizontal $ad1$, a vertical $da1$ and a diagonal $dd1$ detail part. The procedure is then reapplied iteratively on the

Figure 6.9: 2D wavelet transform

approximation coefficient array $aa1$ only, yielding the arrays $aa2$, $ad2$, $da2$ and $dd2$. This procedure is then applied recursively to $aa2$. This process is illustrated in Figure 6.9. The depth of this recursion is limited by the size of the image and the length of the filters. Once the array $aa5$ is obtained, the l_4 norm of its entries is calculated, providing the design criterion V from [64] that is to be maximized by tuning the free filter coefficients:

$$V(F_C) = \max \left(\sum_i \sum_j |aa5_{i,j}|^4 \right) \quad (6.1)$$

This procedure assists in creating a wavelet which pushes the energy of the prototype bias field into the approximation scale. The key idea is to exploit the fact that the bias field is a slowly varying artifact. This process is illustrated in Figure 6.10. The reader should note that the scaling of the prototype bias field is not necessarily the same as the bias field in the actual image. The actual number of scales in the transformation should be selected in accordance with the actual resolution of the image. In this manner the low-pass FIR filter

$$\begin{aligned} & -0.0230 + 0.0173z^{-1} + 0.1533z^{-2} + 0.1349z^{-3} + 0.5148z^{-4} + 0.7898z^{-5} + 0.0455z^{-6} \\ & -0.2567z^{-7} + 0.0165z^{-8} + 0.0219z^{-9} \end{aligned}$$

Figure 6.10: Wavelet designed to maximize the l_4 -norm of the approximation coefficients on scale 5. As a prototype the logarithm of an artificial bias field from BrainWeb was taken. For the optimization $n - 1 = 4$ free parameters were used, which provided a FIR filter of length $2n = 10$.

and the associated power complementary high-pass FIR filter

$$0.0219 - 0.0165z^{-1} - 0.2567z^{-2} - 0.0455z^{-3} + 0.7898z^{-4} - 0.5148z^{-5} + 0.1349z^{-6} \\ - 0.1533z^{-7} + 0.0173z^{-8} + 0.0230z^{-9}$$

were obtained.

6.2.4 Filtering MR images with designed wavelets

In order to filter the actual image, the SWT is applied. This technique that is discussed in Section 3.3 is extended to two dimensions by means of a Cartesian product. The advantage of this approach is that artifacts that can appear with the regular wavelet transform (see e.g. Figure 6.11) are avoided when reconstructing the bias field from the approximation coefficients only. A discussion on artifacts related to such smoothness issues can be found in, e.g., [107, Chapter 10].

Figure 6.11: Original image and reconstructed images with each time the n finest detail coefficients omitted where n is the title of the image. Note that artifacts are introduced, due to a lack of smoothness.

Figure 6.12: In clockwise order: the MRI image of a hip, the logarithm of this image, a 5-scale wavelet decomposition of image, and the approximation coefficients of the SWT of the image.

Another technicality which requires attention is that when applying homomorphic filtering, streak artifacts are created at the boundary between tissue and background. Assume that there is a mask M that indicates the tissue region. In order to avoid these artifacts *normalized convolution* [45] is used. The natural logarithm $\ln(F)$ of the bias field F in the original image O in Figure 6.12 can be reconstructed from the approximation coefficients of the stationary wavelet decomposition of $\ln(O) \bullet M$, where the $\bullet$ denotes the *Hadamard product*, i.e., the entry-wise product. The mask M is filtered in exactly the same manner giving $w(M)$, where $w(\cdot)$ indicates that it has been filtered using a wavelet approach. The actual estimated bias field is then obtained by first applying normalized convolution followed by taking the exponential of each entry $\exp(\cdot)$:

$$F = \underline{\exp} \left(\frac{w(\ln(O) \bullet M)}{w(M)} \right). \quad (6.2)$$

The obtained bias field is shown in the center of Figure 6.13. The original image at the left side of Figure 6.13 is divided by this bias field to obtain the restored image at the right side of Figure 6.13. For the criteria discussed in [64] the image needs to be segmented. For each segment i the mean of the pixel values in the segment μ_i are calculated as well as the standard deviation σ_i . Next as previously discussed in [7] the coefficient of variance $cv_i = \frac{\sigma_i}{\mu_i}$ of each segment and the coefficient of contrast $cc_{i,j} = \frac{\mu_i}{\mu_j}$ between each adjacent pair of segments is calculated. The measures of performance used here are, as discussed in [64], the minimum of the absolute values of the coefficients of contrast of the adjacent segments $\min(bcc|bcc > 0) = \min_{\text{segments } i,j \text{ adjacent}} \{\text{abs}(\ln(cc_{i,j}))\}$, and the mean of the negative logarithm of the coefficients of variance $\text{mean}(lcv) = \text{mean}_i \{-\ln(cv_i)\}$. An overall criterion can be computed as $opc = \min(bcc|bcc > 0)\text{mean}(lcv)$. In terms of these performance measures a score of $\min(bcc|bcc > 0) = 0.392$ was obtained for the original image and $\min(bcc|bcc > 0) = 0.588$ for the reconstructed image. For the reduction in variance the original image had a score of $\text{mean}(lcv) = 0.0695$ and the reconstructed image a score of $\text{mean}(lcv) = 1.024$. The respective products are $opc = 0.0272$ and $opc = 0.602$, which constitutes a major improvement for the reconstructed image.

The current approach appears to work well on images with strong small details. The results for e.g. brain MR images are currently not as good. If the wavelet transform for these images is calculated on more scales one can see an improvement in the results. However the maximum scale that can be calculated is limited by both the image size and the filter size. The relevant components of the image obviously live at the same coarse scale as the bias field. A possible way to overcome this problem is to design multiwavelets (Section 5.3) and to simultaneously optimize one wavelet for the bias field and another orthogonal wavelet for the relevant image such that they become separable. This approach is more flexible than the current one.

Figure 6.13: From left to right: the original image, the bias field as detected by the method and the filtered image after removal of the multiplicative bias field.

Chapter 7

Conclusions and directions for further research

Technological advancements in the field of biomedical engineering help to lengthen the lifespan of patients that suffer from cardiovascular related illnesses. One of such advancements are implantable devices such as pacemakers. Battery life is a critical issue for bio-implantables. To reduce the power demand the frequency of therapies need to be reduced, i.e., good sense amplifiers are crucial and the power consumption of these sense amplifiers needs to be as low as possible. In recent years the technique of wavelet transforms that offer simultaneous time-frequency resolution with a zoom-in property, that offer a choice of basis and that can be used to measure the regularity of signals, has gained popularity in the field of biomedical engineering. Wavelets are a signal processing technique that can make sense amplifiers in pacemakers more powerful. However, A/D conversion is power consuming and for low-power applications a low resolution A/D converter is required. This implies that one should perform as many computations as possible in the analog domain. It is possible to approximate the wavelet transform by matching the impulse response of a linear system with the time-reversed, time-shifted wavelet function. Previously this matching was established by means of Padé approximation. It is discussed however that this technique has a number of drawbacks:

- The matching is performed in the Laplace domain in the neighborhood of a point s_0 which has to be chosen in advance.
- Stability of the filter is not automatically guaranteed. The choice $s_0 = 0$ will help to obtain stability but unfortunately it is likely to give a poor fit at the beginning of the impulse response in the region where the wavelet “lives”.
- The choice of the degrees of the numerator and denominator polynomials which constitute the Padé approximation is not a trivial problem and may strongly influence the quality of the results.

A number of variations to Padé approximation are discussed that attempt to cope with these problems, but the drawback remains that the quality of the approximation is not measured directly in the time domain, but in the Laplace domain, with the consequence that it does not allow for a direct interpretation in system theoretic terms.

In this thesis a novel approach is introduced based on L_2 approximation. This approach offers a number of conceptual and practical advantages:

- The wavelet transform involves the L_2 -inner product between the wavelet function and the signal with an arbitrary time shift. The L_2 approximation approach treats all time instances equal and is appropriate for reliably computing L_2 -inner products. It allows one to determine an error bound for the approximations obtained in this way. Optionally it allows for the use of weighing.
- Due to Parseval's equality the L_2 -norm allows for a description in both the time domain and in the Laplace domain, which we use to our advantage.

In order to approximate wavelet functions with the impulse response of linear systems a parameterization of these impulse responses was proposed that is appropriate for the approximation of wavelet functions. Additionally it is shown that it is required to enforce a first vanishing moment in order to avoid a bias in the approximate wavelet transform. It is shown how the condition for this vanishing moment translates in terms of the parameters that are used to build linear systems which effectively implement a single scale of the wavelet transform.

Since the optimization surface corresponding to the wavelet approximation error may contain multiple local optima, a procedure is proposed to deterministically find a good starting point for the iterative local search optimization routine. The L_2 approximation methodology was demonstrated on the Gaussian wavelet, the Morlet wavelet, the Mexican hat wavelet, the Daubechies 3 wavelet, the Daubechies 7 wavelet and the Coiflet 5 wavelet, of which only one has been successfully implemented directly using the Padé approach. The results of these approximations are quite satisfactory as a specified approximation quality is now obtained with approximations of lower order than previously found with Padé approximation. However it may still happen that the approximation corresponds to an unsatisfactory local optimum. One may approach this problem in several ways:

- To develop a different model order reduction technique that specifically attempts to preserve the shape of the impulse response, rather than to optimize one of the currently available criteria for model order reduction.
- To employ or develop a global optimization technique for the problem at hand. General global optimization techniques, such as simulated annealing, may have the drawback of being computationally intensive. For L_2 approximation, however, theoretical results are available in discrete-time which guarantee the existence of only a finite number of local optima.

- To develop a practical approach which employs a large number of different starting points which are well distributed over the class of candidate approximations.

The choice of a specific wavelet may have substantial impact on the performance of signal processing algorithms such as detection and compression algorithms. This raises the research question of how to select an appropriate wavelet for the signals and problem at hand. In order to support this decision it was argued that for applications such as detection and compression, a sparse representation in the wavelet domain is beneficial. To obtain sparsity we have pursued the principle of maximization of the variance of the wavelet coefficients. For orthogonal wavelets from filter banks using finite-length signals, the energy of the detail and approximation coefficients, computed with periodic extension, will be a constant equalling the energy of the signal. As a result it is possible to either maximize the variance of the absolute values of the wavelet coefficients which is shown to boil down to minimizing their l_1 -norm, or to maximize the variance of the squared wavelet coefficients which comes down to maximization of their l_4 -norm. These criteria can be used as optimization criteria for the design of wavelets.

As a parameterization of orthogonal wavelets, the existing lattice structure of wavelet filters in polyphase form was used. Additional constraints to enforce vanishing moments are worked out and made explicit in terms of the parameterization used. The first vanishing moment can be built-in by eliminating a parameter. A second vanishing moment is given in terms of a constraint on the parameters. This novel combination of the parameterization, the newly introduced design goals and the enforcement of vanishing moments allows for the design of orthogonal wavelets. It was shown that there exist local optima on the optimization surface and appropriate measures need to be taken to arrive at satisfactory results. The applicability of the approach was demonstrated two examples in Chapter 6: on the detection of the QRS complex in ECG signals and on a 2D application: bias field removal in MR images. For low-order wavelet filters, the property to have many vanishing moments can be beneficial in a morphologically mixed set of signals. However this does not hold in general and once a limited number of vanishing moments has been reached one can gain more by maximizing sparsity instead of imposing additional vanishing moments.

One of the major achievements of this thesis is that we have developed a complete procedure to go from a prototype signal, related to a specific application, to a linear filter which performs approximate wavelet analysis in the analog domain, ready for implementation in low-power hardware:

- Wavelet filter design by optimization with a sparsity criterion over a parameterized class
- Computation of the wavelet and scaling function
- L_2 approximation to construct a linear filter

When one wants to distinguish between morphologies in a given signal, multiwavelets can be employed. Orthogonal multiwavelets can, as previously discussed in the literature

and further explained in this work, be formulated in polyphase as lossless systems. In this work a method is introduced that allows for the recursive construction of orthogonal multiwavelets by employing balanced realizations of lossless systems and the reversed tangential Schur algorithm. At the heart of the tangential Schur algorithm are linear fractional transforms that build a lossless polyphase matrix in state-space form by means of a recursive procedure. This provides us with a general framework to deal with orthogonal multiwavelets which encompasses all the existing parameterizations in the literature, and which has a number of advantages from a computational point of view. Numerical experience shows that a well behaved implementation is feasible, as orthogonal matrices are employed.

In this work it is also presented how balanced vanishing moments can be built in to obtain a valid multiwavelet structure that can be used on arbitrary signals without the need of prefiltering. The first balanced vanishing moment corresponds to an interpolation condition on the unit circle. Starting from a low-order system with a built-in balanced vanishing moment, it is shown how the tangential Schur algorithm can be used to recursively build a lossless system of the desired order, which then gives rise to a multiwavelet structure. Effectively these ingredients provide a parameterization of orthogonal multiwavelets with compact support and built-in balanced vanishing moments for which no parameterization was previously available in the literature. This facilitates the design of orthogonal multiwavelets. Additionally, a design procedure is developed which involves masks to highlight different morphologies in a signal to which multiwavelets are adapted. The potential of this approach is demonstrated on the detection of the Q onset and T end points in ECG signals for the purpose of QT interval estimation. For image processing the multiwavelets have been found to be currently too irregular. Additional balanced vanishing moments are known to enhance the regularity of multiwavelets. In the future additional interpolation conditions on the unit circle might be imposed in order to enforce such additional balanced vanishing moments.

Bibliography

- [1] Nurettin Acir. Classification of ECG beats by using a fast least square support vector machines with a dynamic programming feature selection algorithm. *Neural Computing & Applications*, 14(4):299–309, December 2005. [cited at p. 111]
- [2] P.S. Addison. Wavelet transforms and the ECG: A review. *Physiological Measurement*, 26:R155–R199, 2005. [cited at p. 25, 27, 64]
- [3] P.S. Addison, J.N. Watson, G.R. Clegg, M. Holzer, and C.E. Robertson F. Sterz. Evaluating arrhythmias in ECG signals using wavelet transforms. *IEEE Eng. Med. Biol. Mag.*, 19(5):104–109, September–October 2000. [cited at p. 64]
- [4] R. Almeida, J.P. Martínez, A.P. Rocha, S. Olmos, and P. Laguna. Automatic multilead VCG based approach for QT interval measurement. In *Computers in Cardiology*, 2006. [cited at p. 114, 117]
- [5] Rodrigo Varejão Andreão and Jérôme Boudy. Combining wavelet transform and hidden markov models for ECG segmentation. *EURASIP Journal on Advances in Signal Processing*, 2007. [cited at p. 27, 45]
- [6] I. Androulakis, C. Maranas, and C. Floudas. fBB: a global optimization method for general constrained nonconvex problems. *Journal of Global Optimization*, 7:337–363, 1995. [cited at p. 86]
- [7] E. Ardizzone, R. Pirrone, and O. Gambino. Exponential entropy driven HUM on knee MR images. In *IEEE Eng. Med Biol. Conf.*, pages 1769 – 1772, September 2005. [cited at p. 118, 119, 123]
- [8] E. Ardizzone, R. Pirrone, and O. Gambino. Morphological exponential entropy driven-HUM. In *IEEE Eng. Med. Biol. Conf.*, pages 3771 – 3774, August 2006. [cited at p. 118]
- [9] Jane Austen. *Pride and Prejudice*. T. Egerton, Whitehall, January 1813. [cited at p. 3]
- [10] G.A. Baker Jr. *Essentials of Padé Approximants*. Academic Press, 1975. [cited at p. 45, 48]
- [11] B. Beliczynski, I. Kale, and G.D. Cain. Approximation of FIR by IIR digital filters: an algorithm based on balanced model reduction. *IEEE Trans. Signal Processing*, 40(3):532–542, March 1992. [cited at p. 61]

- [12] Aharon Ben-Tal and Arkadi Mermirovski. *Lectures on Modern Convex Optimization*. MPS/SIAM Series on Optimization. MPS-SIAM, 2001. [cited at p. 48]
- [13] David H. Bennett. *Cardiac Arrhythmias: Practical Notes on Interpretation and Treatment*. Hodder Arnold, 7th edition edition, 2006. [cited at p. 14]
- [14] Donald M. Bers. Cardiac excitation-contraction coupling. *Nature*, 415:198–205, January 2002. [cited at p. 8]
- [15] S. Bochner and K. Chandrasekharan. *Fourier transforms*. Princeton University press, 1949. [cited at p. 15, 17]
- [16] R. Bousseljot, D. Kreiseler, and A. Schnabel. Nutzung der EKG-signaldatenbank CARDIODAT der PTB über das internet. *Biomedizinische Technik*, 40:317, 1995. [cited at p. 14, 116]
- [17] R. Bracewell. *The Fourier Transform and Its Applications*. McGraw-Hill, New York, 3rd edition edition, 1999. [cited at p. 15]
- [18] BrainWeb. Simulated brain database. <http://www.bic.mni.mcgill.ca/brainweb/>. [cited at p. 119]
- [19] A. Bultheel and M. van Barel. Padé techniques for model reduction in linear system theory: a survey. *Journal of Computational and Applied Mathematics*, 14(3):401–438, March 1986. [cited at p. 45, 50]
- [20] M.J. Burke and M. Nasor. Wavelet based analysis and characterization of the ECG signal. *Journal of Medical Engineering and Technology*, 28(2):47–55, March-April 2004. [cited at p. 27, 64]
- [21] E. J. Candès and M. Wakin. An introduction to compressive sampling. *IEEE Signal Processing Magazine*, 25(2):21–30, March 2008. [cited at p. 73]
- [22] E. J. Candès, M. Wakin, and S. Boyd. Enhancing sparsity by reweighted l1 minimization. *J. Fourier Anal. Appl.*, 14:877–905, 2007. [cited at p. 73]
- [23] Kuei-Fang Chang, Sue-Jen Shih, and Chiou-Mei Chang. Regularity and vanishing moments of multiwavelets. *Taiwanese Journal of Mathematics*, 1(3):303–314, September 1997. [cited at p. 91]
- [24] J.O. Chapa and R.M. Rao. Algorithms for designing wavelets to match a specified signal. *IEEE Trans. Signal Processing*, 48(12):3395– 3406, December 2000. [cited at p. 6, 71]
- [25] Ivaylo Christov, Ivan Dotsinsky, Iana Simova, Rada Prokopova, Elina Trendafilova, and Stefan Naydenov. Dataset of manually measured QT intervals in the electrocardiogram. *Biomedical Engineering Online*, 5(31), May 2006. [cited at p. 14, 116]
- [26] Charles K. Chui. *Wavelets: A Mathematical Tool for Signal Analysis*. SIAM monographs on mathematical modeling and computation. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 1997. [cited at p. 26]
- [27] C.K. Chui and Q.T. Jiang. Balanced multiwavelets in R^s . *Mathematics of Computation*, 74:1323–1344, 2005. [cited at p. 91, 101, 108]

- [28] Jurgen A. H. R. Claassen. The gold standard: not a golden standard. *BJM*, 330:1121, May 2005. [cited at p. 116]
- [29] D.L. Collins, A.P. Zijdenbos, V. Kollokian, J.G. Sled, N.J. Kabani, C.J. Holmes, and A.C. Evans. Design and construction of a realistic digital brain phantom. *IEEE Trans. Med. Imaging*, 17(3):463–468, 1998. [cited at p. 119]
- [30] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Comm. on Pure and Appl. Math.*, 41(7):909–996, 1988. [cited at p. 25, 30]
- [31] Ingrid Daubechies. *Ten Lectures on Wavelets*. SIAM, 1992. [cited at p. 29, 30]
- [32] B.M. Dawant, A.P. Zijdenbos, and R.A. Margolin. Correction of intensity variations in MR images for computer-aided tissue classification. *IEEE Trans. Med. Imaging*, 12:770–781, 1993. [cited at p. 118]
- [33] David L. Donoho. For most underdetermined systems of linear equations, the minimal l1-norm near-solution approximates the sparsest near-solution. *Comm. Pure Appl. Math.*, 59(6):797–829, June 2006. [cited at p. 73]
- [34] David L. Donoho and Michael Elad. Optimally sparse representation from overcomplete dictionaries via l1 norm minimization. *Proc. Natl. Acad. Sci.*, 100(5):2197–2002, March 2003. [cited at p. 73]
- [35] Barbara J. Drew. Pitfalls and artifacts in electrocardiography. *Cardiology Clinics*, 24(3):309–315, August 2006. [cited at p. 10]
- [36] W. Einthoven. The different forms of the human electrocardiogram and their signification. *Lancet*, 1:853–861, 1912. [cited at p. 10]
- [37] Mourad N. El-Gamal and Gordon W. Roberts. A 1.2-V N-P-N-only integrator for log-domain filtering. *IEEE Trans. Circuits Syst. II*, 49(4):257–265, April 2002. [cited at p. 43, 44]
- [38] D. Esteban and C. Galand. Application of quadrature mirror filters to split-band voice coding. In *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, volume 2, pages 191–195, April 1977. [cited at p. 28, 29, 30]
- [39] Richard Feynman. *The Feynman Lectures on Physics*. Addison Wesley Longman, 1970. [cited at p. 16]
- [40] D. Gabor. Theory of communication. *Journal I.E.E.*, 93(26):429–457, November 1946. [cited at p. 18]
- [41] W.H. Gispen, B. Bravenboer, P.-H. Hendriksen, P.L. Oey, A.C. van Huffelen, and D.W. Erkelens. Is the corrected QT interval a reliable indicator of the severity of diabetic autonomic neuropathy? *American Diabetes Care*, 16(9):1249–1253, 1993. [cited at p. 114]
- [42] David E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Kluwer Academic Publishers, 1989. [cited at p. 86]
- [43] A.L. Goldberger, L.A.N. Amaral, L. Glass, J.M. Hausdorff, P.Ch. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.-K. Peng, and H.E. Stanley. Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, June 2000. [cited at p. 14, 110, 111]

- [44] A. Grossman and J. Morlet. Decomposition of hardy functions into square integrable wavelets of constant shape. *Siam J. on Math. Analysis*, 15(4):723–736, 1984. [cited at p. 25, 26]
- [45] R. Guillemaud. Uniformity correction with homomorphic filtering on region of interest. In *IEEE Int. conf. Image Processing*, volume 2, pages 872–875, 1998. [cited at p. 119, 123]
- [46] Anubha Gupta, Shiv Dutt Joshi, and Surendra Prasad. A new method of estimating wavelet with desired features from a given signal. *Signal Processing*, 85(1):147–161, January 2005. [cited at p. 6, 71]
- [47] Sandor A.P. Haddad. *Ultra Low-Power Biomedical Signal Processing: An Analog Wavelet Filter Approach for Pacemakers*. PhD thesis, Delt University of Technology, December 2006. [cited at p. 46]
- [48] S.A.P. Haddad, S. Bagga, and W.A. Serdijn. Log-domain wavelet bases. In *Proceedings of the 2004 International Symposium on Circuits and Systems, ISCAS '04*, volume 1, pages 1100–1103, May 2004. [cited at p. 4, 43]
- [49] S.A.P. Haddad, S. Gieltsjes, R.P.M. Houben, and W.A. Serdijn. An ultra low-power dynamic translinear sense amplifier for pacemakers. In *proc. IEEE International Symposium on Circuits and Systems*, 2003. [cited at p. 3, 4, 43]
- [50] S.A.P. Haddad, R. Houben, and W.A. Serdijn. Analog wavelet transform employing dynamic tranlinear circuits for cardiac signal characterization. In *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS)*, volume 1, pages 121–124, May 2003. [cited at p. 4, 43]
- [51] S.A.P. Haddad, R.P.M. Houben, and W.A. Serdijn. The evolution of pacemakers. *IEEE Eng. Med. Biol. Mag.*, 25(3):38–48, May-June 2006. [cited at p. 3, 7, 8, 10]
- [52] S.A.P. Haddad, J.M.H. Karel, R.L.M. Peeters, R.L. Westra, and W.A. Serdijn. Complex wavelet transform for analog signal processing. In *proc.ProRISC'2004, Veldhoven*, November 2004. [cited at p. 64]
- [53] S.A.P. Haddad and W.A. Serdijn. Mapping the wavelet transform onto silicon: the dynamic translinear approach. In *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS)*, volume V, pages 621–624, May 2002. [cited at p. 4, 43]
- [54] S.A.P. Haddad, N. Verwaal, R. Houben, and W.A. Serdijn. Optimized dynamic translinear implementation of the gaussian wavelet transform. In *Proceedings of the IEEE International Symposium on Circuits and Systems*, volume I, pages 145–148, May 2004. [cited at p. 4, 43, 45, 50]
- [55] B. Hanzon, M. Olivi, and R.L.M. Peeters. Balanced realizations of discrete-time stable all-pass systems and the tangential Schur algorithm. *Linear Algebra and its Applications*, 418(2-3):793–820, October 2006. [cited at p. 6, 72, 89, 90, 91, 92, 93, 94, 98]
- [56] Bernard Hanzon and Ralf L.M. Peeters. Balanced parametrizations of stable siso all-pass systems in discrete time. *Mathematics of Control, Signals, and Systems (MCSS)*, 13(3):240–276, September 2000. [cited at p. 91, 92]
- [57] D. Hayn, A. Kollmann, and G. Schreier. Automated QT interval measurement from multilead ECG signals. In *Computers in Cardiology*, 2006. [cited at p. 114, 117]

- [58] W. Heisenberg. Über den anschaulichen inhalt der quantentheoretischen kinematik und mechanik. *Zeitschrift für Physik*, 43:172–198, 1927. [cited at p. 18]
- [59] J. Willis Hurst. Naming of the waves in the ECG, with a brief account of their genesis. *Circulation*, 98:1937–1942, 1998. [cited at p. 12]
- [60] M.H. Jager-Geurts, R.J.G. Peters, S.J. van Dis, and M.L. Bots. Hart- en vaatziekten in Nederland 2006. Technical report, Nederlandse Hartstichting, 2006. [cited at p. 7]
- [61] S. Kadambe, R. Murray, and G.F. Boudreaux-Bartels. Wavelet transform-based QRS complex detector. *IEEE Trans. Biomed. Eng.*, 46(7):838–848, July 1999. [cited at p. 4, 25]
- [62] Thomas Kailath. *Linear Systems*. Prentice Hall Inc., 1980. [cited at p. 18, 20, 22, 53]
- [63] A. Kapela, R.D. Berger, A. Achim, and A. Bezerianos. Wavelet variance analysis of high-resolution ECG in patients prone to VT/VF during cardiac electrophysiology studies. In *Proceedings of the 14th International Conference on Digital Signal Processing (DSP 2002)*, volume 2, pages 1133–1136, July 2002. [cited at p. 64]
- [64] J.M.H. Karel, K. Fischer, R.L. Westra, and R.L.M. Peeters. Measures for the evaluation of bias field removal procedures in MRI and the design of wavelets for suppression of RF inhomogeneities. In *Proc. Eng.Med.Biol. Conf*, 2008. (submitted). [cited at p. 6, 109, 118, 119, 120, 123]
- [65] J.M.H. Karel, R.L.M. Peeters, R.L. Westra, S.A.P. Haddad, and W.A. Serdijn. L2-norm based wavelet approximation for analog implementation. Technical Report M 04-05, Maastricht University, P.O. Box 616 6200 MD Maastricht The Netherlands, December 2004. [cited at p. 64]
- [66] J.M.H. Karel, R.L.M. Peeters, R.L. Westra, S.A.P. Haddad, and W.A. Serdijn. An L2-based approach for wavelet approximation. In *Proceedings of the CDC-ECC 2005*, 2005. [cited at p. 55, 57]
- [67] J.M.H. Karel, R.L.M. Peeters, R.L. Westra, S.A.P. Haddad, and W.A. Serdijn. Wavelet approximation for implementation in dynamic translinear circuits. In *Proceedings of the 16th IFAC World Congress*. International Federation of Automatic Control, 2005. [cited at p. 53, 55]
- [68] J.M.H. Karel, R.L.M. Peeters, R.L. Westra, S.A.P. Haddad, and W.A. Serdijn. Wavelet design for analog implementation. In *25th Benelux Meeting on Systems and Control*, Heeze, The Netherlands, March 2006. [cited at p. 6]
- [69] J.M.H. Karel, R.L.M. Peeters, R.L. Westra, K.M.S. Moermans, S.A.P. Haddad, and W.A. Serdijn. Optimal discrete wavelet design for cardiac signal processing. In *Proc. 27th Ann. Int. Conf. of the IEEE Enigneering in Medicine and Biology Society (EMBC 2005)*, Shanghai, China, September 1-4. IEEE, September 2005. [cited at p. 6, 71, 73, 119]
- [70] Joël Karel, Ralf Peeters, Ronald Westra, Sandro Haddad, and Wouter Serdijn. QT interval measurement in cardiac signal processing with multiwavelets. In *First Dutch Conference on Bio-Medical Engineering*, page 184, January 2007. [cited at p. 114]
- [71] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983. [cited at p. 86]

- [72] R.K.-S. Kwan, A.C. Evans, and G.B. Pike. An extensible MRI simulator for post-processing evaluation. In *Visualization in Biomedical Computing (VBC'96)*, volume 1131 of *Lecture Notes in Computer Science*, pages 135–140, 1996. [cited at p. 119]
- [73] J.C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright. Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM Journal of Optimization*, 9(1):112–147, 1998. [cited at p. 86]
- [74] Cornelius Lanczos. *Applied analysis*. Prentice-Hall Mathematics Series. Prentice Hall, Inc., 1956. [cited at p. 54, 90]
- [75] P. Langley, F.E. Smith, S.T. King, D. Zheng, A.J. Haigh, and A. Murray. Fully automated computer measurement of QT interval from the 12-lead electrocardiogram. In *Computers in Cardiology*, 2006. [cited at p. 114, 117]
- [76] J. Lebrun and M. Vetterli. Balanced multiwavelets theory and design. *IEEE Trans. Signal Processing*, 46(4):1119–1125, April 1998. [cited at p. 6, 40, 72, 89, 91, 108]
- [77] J. Lebrun and M. Vetterli. High-order balanced multiwavelets: theory, factorization, and design. *IEEE Trans. Signal Processing*, 49(9):1918–1930, 2001. [cited at p. 6, 40, 72, 89, 91, 101, 103, 108]
- [78] Cuiwei Li, Chongxun Zheng, and Changfeng Tai. Detection of ECG characteristic points using wavelet transforms. *IEEE Trans. Biomed. Eng.*, 42(1):21–28, January 1995. [cited at p. 4, 25]
- [79] Stéphane Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Machine Intell.*, 11(7):674 – 693, July 1989. [cited at p. 25, 31, 101, 118, 119]
- [80] Stéphane Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 1999. [cited at p. 18, 26, 27, 32, 33, 34, 51, 72]
- [81] Stéphane Mallat. *A Wavelet Tour of Signal Processing: The Sparse Way*. Academic Press, 2008. [cited at p. 73]
- [82] J.P. Martínez, R. Almeida, S. Olmos, A.P. Rocha, and P. Laguna. Stability of QT measurements in the PTB database depending on the selected lead. In *Computers in Cardiology*, 2006. [cited at p. 117]
- [83] Juan Pablo Martínez, Rute Almeida, Salvador Olmos, Ana Paula Rocha, and Pablo Laguna. A wavelet-based ECG delineator: Evaluation on standard databases. *IEEE Trans. Biomed. Eng.*, 51(4):570–581, April 2004. [cited at p. 4, 25]
- [84] Hamish D. Meikle. *A New Twist to Fourier Transforms*. Technology & Industrial Arts. Wiley-VCH, 2004. [cited at p. 17]
- [85] Y. Meyer. Constructions de bases orthonormées d’ondelettes. *Rev. Mat. Iberoamericana*, 4:31–39, 1988. [cited at p. 25]
- [86] Fred Mintzer. Filters for distortion-free two-band multirate filter banks. *IEEE Trans. Acoustics, Speech and Signal Proc.*, 33(3):626–630, June 1985. [cited at p. 30]

- [87] F. Mochimaru, Y. Fujimoto, and Y. Ishikawa. Detecting the fetal electrocardiogram by wavelet theory-based methods. *Progress in Biomedical Research*, 7(3):185–193, September 2002. [cited at p. 4, 25]
- [88] G.B. Moody, H. Koch, and U. Steinhoff. The physionet/computers in cardiology challenge 2006: QT interval measurement. In *Computers in Cardiology*, 2006. [cited at p. 14, 114]
- [89] B.C. Moore. Principal component analysis in linear systems: Controllability, observability, and model reduction. *IEEE Trans. Automat. Contr.*, 26(1):17–32, 1981. [cited at p. 60]
- [90] Jan Mulder, Wouter Serdijn, Albert C. van der Woerd, and Arthur H.M. van Roermund. Dynamic translinear circuits - an overview. *Analog Integrated Circuits and Signal Processing*, 22(2-3):111–126, 2000. [cited at p. 43, 44]
- [91] Guy P. Nason and Bernard W. Silverman. The stationary wavelet transform and some statistical applications. In Anestis Antoniadis and Georges Oppenheim, editors, *Lecture Notes in Statistics: Wavelets and Statistics*, volume 103 of *Lecture Notes in Statistics*, pages 281–299. Springer-Verlag, 1995. [cited at p. 33, 37, 38, 110]
- [92] N. Neretti and N. Intrator. An adaptive approach to wavelet filter design. In *Proceedings of the IEEE international workshop on neural networks for signal processing*, September 2002. [cited at p. 6, 25, 71]
- [93] Nikolay Nikolaev and Atanas Gotchev. ECG signal denoising using wavelet domain Wiener filtering. In *European Singal Processing Conference (Eusipco)*, September 2000. [cited at p. 111]
- [94] H. Nyquist. Certain topics in telegraph transmission theory. *Proceedings of the IEEE*, 90(2):280–305, February 2002. [cited at p. 15]
- [95] R.L.M. Peeters, J.M.H. Karel, R.L. Westra, S.A.P. Haddad, and W.A. Serdijn. Multi-wavelet design for cardiac signal processing. In *Proc. 28th Ann. Int. Conf. of the IEEE Enigneering in Medicine and Biology Society (EMBC 2006)*, New York City, August 30 - September 3, pages 1682–1685, 2006. [cited at p. 89, 91, 101, 102]
- [96] L. Pernebo and L.M. Silverman. Model reduction via balanced state space representations. *IEEE Trans. Automat. Contr.*, 27(2):382–387, 1982. [cited at p. 60]
- [97] Jan Willem Polderman and Jan C. Willems. *Introduction to Mathematical Systems Theory, A Behavioral Approach*. Texts in applied mathematics: 26. Springer, 1997. [cited at p. 53]
- [98] Ivo Provazník, Jiří Kozumplík, Jana Bardoňová, Marie Nováková, and Zuzana Nováková. Wavelet transform in ECG signal processing. In *Proceedings of the 15th International Conference BIOSIGNAL 2000*, pages 21–25. Vutium Press Brno, 2000. [cited at p. 4, 25]
- [99] P.E. Puddu and M.G. Bourassa. Prediction of sudden death from QTc interval prolongation in patients with chronic ischemic disease. *Journal of Electrocardiology*, 19:203–212, 1986. [cited at p. 114]
- [100] R.M. Rangayyan. *Biomedical Signal Analysis: A case-study approach*, volume IEEE Press series on biomedical engineering. IEEE Press, 2002. [cited at p. 10]
- [101] Philippe Ravier and Olivier Buttelli. Robust detection of QRS complex using Klauder wavelets. In *EUSIPCO 2004*, 2004. [cited at p. 6, 25, 71]

- [102] J. Sahambi, S. Tandon, and R. Bhatt. Using wavelet transforms for ECG characterization. *IEEE Eng. Med. Biol. Mag.*, 16(1):77–83, 1997. [cited at p. 4, 25, 27, 45]
- [103] S.C. Saxena, V. Kumar, and S.T. Hamde. QRS detection using new wavelets. *Journal of Medical Engineering and Technology*, 26(1):7–15, January–February 2002. [cited at p. 4, 71]
- [104] C.E. Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, January 1949. [cited at p. 15]
- [105] D. A. Sherman. An introduction to wavelets with electrocardiology applications. *Herzschrittmachertherapie und Elektrophysiologie*, 9:42–52, 1998. [cited at p. 25, 111]
- [106] M.J.T. Smith and T. P. Barnwell III. A procedure for designing exact reconstruction filter banks for tree-structured subband coders. In *IEEE Int. Conf on Acoustics, Speech and Signal Processing*, volume 9, pages 421–424, March 1984. [cited at p. 29]
- [107] Gilbert Strang and Truong Nguyen. *Wavelets and Filter Banks*. Wellesley-Cambridge Press, 1996. [cited at p. 6, 25, 27, 28, 29, 31, 32, 33, 34, 35, 40, 71, 72, 75, 76, 79, 89, 100, 121]
- [108] Gilbert Strang and Vasily Strela. Orthogonal multiwavelets with vanishing moments. *Journal of Optical Engineering*, 33:2104–2107, 1994. [cited at p. 91]
- [109] Vasily Strela. Multiwavelets: Regularity, orthogonality and symmetry via two-scale similarity transform. *Studies in Applied Mathematics*, 98:335–354, 1997. [cited at p. 91]
- [110] Jacques Theys. *Joint Spectral Radius: theory and approximations*. PhD thesis, Université Catholique de Louvain, May 2005. [cited at p. 32]
- [111] M.B. Thomsen, S.C. Verduyn, M. Stengl, J.D.M. Beekman, G. de Pater, J. van Opstal, P.G.A. Volders, and M.A. Vos. Increased short-term variability of repolarization predicts d-sotalol-induced torsades de pointes in dogs. *Circulation*, 110:2453–2459, 2004. [cited at p. 114]
- [112] P. E. Tikkänen. Nonlinear wavelet and wavelet packet denoising of electrocardiogram signal. *Biological Cybernetics*, 80(4):259 – 267, April 1999. [cited at p. 4, 25]
- [113] M. Tincher, C.R. Meyer, R. Gupta, and D.M. Williams. Polynomial modelling and reduction of RF body coil spatial inhomogeneity in MRI. *IEEE Trans. on Med. Imaging*, 12:361–365, 1993. [cited at p. 118]
- [114] Hüseyin Tirtoma, Mehmet Engin, and Erkan Zeki Engin. Enhancement of time-frequency properties of ecg for detecting micropotentials by wavelet transform based method. *Expert Systems with Applications*, 34(1):746–753, January 2008. [cited at p. 111]
- [115] Richard J. Vaccaro. *Digital Control: A State-Space Approach*. McGraw-Hill Series in Electrical and Computer Engineering. McGraw-Hill, 1995. [cited at p. 22, 62]
- [116] P.P. Vaidyanathan. Theory and design of m-channel maximally decimated quadrature mirror filters with arbitrary m, having the perfect-reconstruction property. *IEEE Trans. Acoustics, Speech and Signal Proc.*, 35(4):476–492, April 1987. [cited at p. 41, 72, 75, 89, 100]
- [117] P.P. Vaidyanathan and Zinnur Doğanata. The role of lossless systems in modern digital signal processing: analog tutorial. *IEEE Trans. Education*, 32(3):181–197, August 1989. [cited at p. 6, 29, 30, 41, 42, 72, 89, 100]

- [118] C.F. van Loan. Computing integrals involving the matrix exponential. *IEEE Trans. Automat. Contr.*, 23:395–404, 1978. [cited at p. 62]
- [119] Martin Vetterli. Filter banks allowing perfect reconstruction. *Signal Processing*, 10:219–244, 1986. [cited at p. 40, 89]
- [120] Martin Vetterli and Cormac Herley. Wavelets and filter banks: relationships and new results. In *IEEE Conference on Acoustics, Speech and Signal Processing*, volume 3, pages 1723–1726, 1990. [cited at p. 34]
- [121] Jan C. Willems and Paolo Rapisarda. Balanced state representation from higher-order models. In *Proceedings Equadiff 2003*, Hasselt, Belgium, 2003. [cited at p. 60]
- [122] Mark A. Wood and Kenneth A. Ellenbogen. Cardiac pacemakers from the patient’s perspective. *Circulation*, 105(18):2136–2138, 2002. [cited at p. 3, 7, 8]
- [123] World Health Organization. Cardiovascular diseases - fact sheet 317. <http://www.who.int/mediacentre/factsheets/fs317/en/index.html>, February 2007. [cited at p. 3, 7]
- [124] Douglas P. Zipes and José Jalife. *Cardiac electrophysiology: from cell to bedside*. Saunders, 2000. [cited at p. 9]

Summary

Each year a large number of people die due to cardiovascular disorders. Due to medical advancements the contribution of these disorders to the total mortality is reduced. One of these advancements is a joint medical and engineering development: the implantable devices, among which the pacemaker. Early pacemakers electro-stimulated the heart at a fixed rate, ensuring that the heart contracts according to that rate and pumps blood throughout the body. This approach however has a number of drawbacks. First of all, the continuous stimulation of the heart is power consuming, which is a major issue for a device that is surgically inserted into the body and is not easily rechargeable. Secondly, unnecessary stimulation of the heart can be harmful. Thirdly, the fixed rate of the pacemaker can interfere with the intrinsic rate of the heart. Nowadays pacemakers are equipped with a sensing circuit that monitors the electrical currents on the heart. To accomplish this a monitoring mechanism is required that performs signal processing, leading to a decision step, in which the decision is made whether or not to stimulate the heart. Considering that this monitoring circuit is always active, it must be ensured that this circuit is not power-consuming. Obviously this circuit must also be robust since a patient's life depends on it.

A relatively young signal processing technique that is interesting for the use in pacemakers is the so-called “wavelet transform”. With this technique a signal can simultaneously be represented in both the time- and frequency-domain. It is easy to implement in a computer. However, a computer generally operates in the digital domain, whereas the sensor information is in the analog domain. An analog signal thus has to be converted to the digital domain, which is a power-consuming operation. An energy saving solution is to implement the wavelets in the analog domain, and in this manner to reduce the amount of analog to digital conversion. To achieve this, wavelets have to be approximated by means of linear systems that can be applied in microelectronics. This is not a trivial task. In the current work it is discussed why L_2 approximation is a relevant technique. A complete approach is discussed to approximate wavelet functions (associated with both continuous and discrete-time wavelets) with this technique. Since the optimization surface may contain various local optima, it is discussed how a good starting point for a local search algorithm can be found. Various examples illustrate the

flexibility of this approach.

The simultaneous time-frequency representation characteristic is not the only distinction between wavelets and Fourier transforms. Unlike Fourier transforms, wavelet transforms offer a choice of bases. Two criteria, that operate in the wavelet domain, are introduced to determine how good a certain wavelet is for compressing or detecting a given signal. These criteria are then used, along with a parameterization of orthogonal wavelets based on “polyphase filters” and the ‘lattice structure”, to design custom wavelets for an application at hand. Not only scalar wavelets are of interest, but also multiwavelets. These involve multiple orthogonal wavelet and scaling functions that enable them to separate orthogonal components in a signal. A parameterization in terms of “lossless systems” is introduced for these multiwavelets. This parameterization is more general than the parameterizations that are known from the literature. For a number of these parameterizations it is discussed how these follow from the introduced parameterization as special cases. For the new parameterization it is additionally discussed how balanced vanishing moments can be built-in, which is required in order to use these designed multiwavelets directly on measured signals. These balanced vanishing moments were not explicitly build into earlier parameterizations.

To demonstrate the potential of the discussed techniques, three examples are worked out. Firstly the designed scalar wavelets are used to detect the QRS complex in an ECG. Experiments show that the designed wavelets indeed offer advantages. An interesting observation is that the Daubechies 2 wavelets is often found as an optimum. As a second application designed multiwavelets are used to simultaneously distinguish the Q and the T peak in ECGs. A rule-based decision algorithm that has been designed as a proof of principle readily shows promising results. As a third application wavelet design is used to facilitate the processing of MR images. Using wavelet filtering, low-frequency multiplicative noise is successfully removed from images of the pelvis and the knee. The current technique is not as successful on MR images of the brain. Recommendations to increase the performance are made.

Samenvatting

Jaarlijks sterft een groot aantal mensen ten gevolge van cardiovasculaire aandoeningen. Door vooruitgang op het gebied van de medische wetenschap, waaronder meer specifiek implanteerbare apparaten, wordt het aandeel in de totale sterfte verkleind. Een van de eerste van dergelijke implanteerbare apparaten is de pacemaker. De vroege uitvoeringen geven met een vast ritme elektrische pulsen aan het hart, waardoor dit met het vast-gestelde ritme samenknijpt en bloed door het lichaam pompt. Hier kleeft echter een aantal nadelen aan. Zo kost het voortdurend stimuleren van het hart veel energie, wat een groot probleem vormt omdat een dergelijk apparaat chirurgisch in het lichaam van de patiënt is geplaatst en daardoor moeilijk opgeladen kan worden. Ook kan het onnodig stimuleren van het hart schadelijk zijn. Daarnaast kan het vaste ritme van de pacemaker interfereren met de frequentie waarop het hart zelf wil samentrekken. Tegenwoordig zijn pacemakers daarom uitgerust met sensoren die de elektrische stromen op het hart registreren die vervolgens worden gebruikt in een beslissingsstap waarin wordt besloten of het hart gestimuleerd moet worden of niet. Dit vereist een monitoringsmechanisme dat signaalverwerkingstaken uitvoert. Aangezien het monitoring circuit altijd actief is, is het belangrijk dat dit zuinig met de energie omgaat. Daarnaast moet het de juiste beslissingen nemen, omdat deze van levensbelang kunnen zijn.

Een relatief recente signaalverwerkingstechniek die interessant is voor het gebruik in pacemakers is de zogenaamde "wavelet transformatie". Deze techniek maakt het mogelijk om een signaal tegelijkertijd in tijd en frequentie af te beelden. Het is gemakkelijk om deze techniek in een computer te implementeren. Echter een computer werkt digitaal en de sensor informatie in pacemakers is analoog. Er moet dus een analoog signaal omgezet worden in een digitaal signaal wat veel energie vergt. Een energiezuinige oplossing is om de wavelets op een analoge manier te implementeren en zo de analoog/digitaal omzetting te beperken. Om dit te doen moeten de wavelet functies benaderd worden met behulp van lineaire systemen die in de microelektronica toepasbaar zijn. Dit is echter geen triviale opgave. In het huidige werk wordt beargumenteerd waarom L_2 approximatie een goede aanpak is hiervoor. Een complete aanpak om wavelet functies (zowel van continue als discrete wavelets) met deze techniek te benaderen is uitgewerkt. Aangezien het optimalisatieoppervlak diverse locale optima kan bevatten, wordt er besproken hoe

een goed startpunt voor een locale zoektechniek kan worden gevonden. Met diverse voorbeelden wordt de flexibiliteit van de aanpak aangetoond.

Buiten het feit dat wavelets een signaal tegelijkertijd in termen van tijd als frequentie laten zien is er nog een ander duidelijk verschil met een klassieke techniek als Fourier transformaties: er zijn geen vaste basissen en er is keuzevrijheid. Dit geeft meteen een keuzeprobleem. Er worden twee criteria behandeld die in het wavelet domein bepalen hoe goed een gegeven wavelet is om een bepaald signaal te comprimeren of te detecteren. Daarnaast worden deze criteria met een parameterizatie op basis van “polyphase filters” en de “lattice” structuur van discrete, orthogonale wavelets, gebruikt om wavelets te ontwerpen. Naast reguliere wavelets zijn ook zogenaamde “multiwavelets” interessant. Deze beschikken over meerdere orthogonale wavelet en schalingsfuncties en kunnen diverse orthogonale componenten in een signaal onderscheiden. Voor deze multiwavelets is een parameterisatie opgezet in termen van zogenaamde “lossless” systemen. Deze parameterisatie is algemener dan eerdere parameterisaties die bekend zijn uit de literatuur. Van een aantal van deze bestaande parameterisaties wordt besproken hoe ze volgen uit de geïntroduceerde, algemenere parameterisatie. Daarnaast wordt besproken hoe met de geïntroduceerde parameterisatie gebalanceerde momenten kunnen worden ingebouwd, hetgeen noodzakelijk is om de, met deze parameterisatie, ontworpen multiwavelets direct op gemeten signalen toe te passen. Deze gebalanceerde momenten werden in eerdere parameterisaties niet ingebouwd.

Om de kracht van de besproken technieken te demonstreren wordt een drietal voorbeelden uitgewerkt. Als eerste worden ontworpen wavelets gebruikt om het QRS complex in een hartsignaal te detecteren. Uit numerieke gegevens blijkt dat deze ontworpen wavelets inderdaad voordelen bieden. Een opmerkelijk resultaat is dat een bepaalde wavelet, nl. de Daubechies 2 wavelet dikwijls als optimum gevonden wordt. Als tweede applicatie worden ontworpen multiwavelets gebruikt om tegelijkertijd de Q piek en de T piek in een hartsignaal te onderscheiden. Een regelgebaseerd detectiealgoritme dat als voorbeeld is ontworpen laat reeds bemoedigende resultaten zien. Als derde toepassing worden wavelets ontworpen en gebruikt voor het verwerken van afbeeldingen afkomstig van een MRI scanner. Met behulp van deze technieken wordt succesvol laagfrequente multiplicatieve ruis verwijderd bij afbeeldingen van het bekken en de knie. Op MRI afbeeldingen van hersenen is de huidige techniek minder succesvol, echter aanbevelingen om dit verder te verbeteren worden gedaan.

Curriculum Vitae

1978	Born on 10 April 1978 in Maastricht, The Netherlands
1990–1997	Secondary School, Jeanne d’Arc College Maastricht, HAVO/VWO
1997–2001	Master of Science in Knowledge Engineering, business mathematics major. Joint programme of the Maastricht University (The Netherlands) and Limburgs Universitair Centrum (Nowadays Hasselt University, Belgium). Obtained the Belgian diploma “Kandidaat Informatica” (Computer Science) from the Limburgs Universitair Centrum in 1999. Attended a special four-week programme on knowledge engineering and computer science in 1999 at Baylor University, Waco, Texas. Master thesis at Medtronic Bakken Research Center b.v., Maastricht, The Netherlands, on the mathematical modeling of atrial fibrillation, completing a five years dual track at the Maastricht University in four years in 2001.
2001–2004	Joint position at MaTeUM b.v., Maastricht, The Netherlands as a developer and at the Maastricht University, Faculty of General Sciences for research on simulation.
2004–2008	Ph.D. Student on the STW funded BioSens project at the Maastricht University, Faculty of Humanities and Sciences, MICC, Department of Mathematics.
2007–2008	Half-time position as a lecturer at the Maastricht University, Faculty of Humanities and Sciences, MICC, Department of Mathematics.
2008–current	Assistant Professor at the Maastricht University, Faculty of Humanities and Sciences, Department of Knowledge Engineering.

List of Symbols and Abbreviations

Symbol	Description	Definition
$\bullet$	Hadamard product	page 123
$*$	convolution operator	page 17
$\dagger$	Hermitian transpose	page 42
$\downarrow 2$	downsampling by 2; $(\downarrow 2x)_k = x_{2k}$	page 28
A	system matrix for a linear system in state-space representation	page 21
a_k	approximation or scaling coefficients	page 31
B	input matrix for a linear system in state-space representation	page 21
b_k	detail or wavelet coefficients	page 31
C	output matrix for a linear system in state-space representation	page 21
c_k	scaling filter coefficients	page 30
$\delta[n]$	Kronecker delta	page 18
$\delta(t)$	Dirac delta	page 18
D	direct feed-through matrix for a linear system in state-space representation	page 21
d_k	wavelet filter coefficients	page 30
e_n	n^{th} standard basis vector	page 99
$\exp$	element-wise exponential	page 123
$\overline{\mathcal{F}}_{U,V}$	mapping for a proper rational matrix	page 92
$\mathcal{H}(E)$	Halmos extension	page 95
$\mathcal{H}$	Hankel matrix	page 61
$h(t)$	impulse response	page 20
$H(s)$	transfer function	page 20
$H_0(z)$	wavelet low-pass filter	page 28
$H_1(z)$	wavelet high-pass filter	page 28
$H_p(z)$	polyphase matrix in z	page 35

Symbol	Description	Definition
i	complex number	page 16
$\mathbf{1}_A(x)$	indicator function	page 54
$\mathcal{L}$	Laplace transform	page 18
$\underline{\ln}$	Element-wise natural logarithm	page 119
ω	frequency variable	page 16
ϕ	phase angle	page 16
$u(t)$	input for a system	page 19
$\varphi(t)$	scaling function	page 31
$\underline{\varphi}(t)$	multi scaling function	page 40
$\psi(t)$	wavelet function	page 26
$\tilde{\psi}(t)$	time-reversed and time-shifted wavelet function	page 47
$\hat{\psi}(t)$	causal time-reversed and time-shifted wavelet function	page 47
$\underline{\psi}(t)$	multi wavelet function	page 40
σ	scale	page 26
$\mathcal{T}$	linear fractional transformation	page 93
$W(\tau, \sigma)$	wavelet transform	page 26
$x(t)$	state vector	page 21
$y(t)$	output for a system	page 19
$Y(i\omega)$	Laplace transform of $y(t)$ for $s = i\omega$	page 18
$Y(\omega)$	Fourier transform of $y(t)$	page 18
$Y(s)$	Laplace transform of $y(t)$	page 18
$\mathcal{Z}$	Z-transform	page 19

Abbreviation	Description	Definition
3D	three-dimensional	page 10
A/D	analog-to-digital	page 15
AF	atrial fibrillation	page 14
AV	atrio-ventricular	page 9
BVR	beat-to-beat variability of repolarization	page 114
CT	computed tomography	page 118
CWT	continuous wavelet transform	page 26
DC	direct current	page 21
DTL	dynamic translinear	page 43
DWT	discrete wavelet transform	page 27
ECG	electrocardiogram	page 10
EKG	electrocardiogram	page 10
FIR	finite impulse response	page 23
Hz	Hertz	page 16
IECG	intracardiac electrocardiogram	page 10
LA	left atrium	page 8
LFT	linear fractional transformations	page 93
LTI	linear time-invariant	page 19
LV	left ventricle	page 8

Abbreviation	Description	Definition
MA	moving average	page 23
MIMO	multi-input multi-output	page 21
MIT-BIH	Massachusetts Institute of Technology - Beth Israel Deaconess Medical Center	page 14
MR	magnetic resonance	page 118
MRA	multi-resolution analysis	page 31
MRI	magnetic resonance imaging	page 118
NMRI	nuclear magnetic resonance imaging	page 118
pdf	probability density function	page 27
RA	right atrium	page 8
RF	radio frequency	page 118
RV	right ventricle	page 8
SA	sino-atrial	page 9
SISO	single-input single-output	page 20
VF	ventricular fibrillation	page 14
ZOH	zero-order-hold	page 18

Index

- l_1 -norm, 73
- l_4 -norm, 73
- A/D converter, 15, 43
- action potential, 9
- admissability condition, 26
- aliasing, 28
- all-pass systems, 41, 90
- alternating flip, 30
- alternating signs, 29
- analog, 15
- analysis, 16
- angular frequency, 16
- anisotropy, 10
- approximation
 - L_2 , 51
 - coefficients, 28
 - error, 34
 - Padé, 48
- approximation space, 31
- arrhythmia, 10
- asymptotic stability, 20, 91
- atrial fibrillation, 14
- atrio-ventricular node, 9
- atrium, 8
- balance and truncate, 62
- balancing, 60, 98, 101, 108
- bandlimited, 16
- bandwidth, 16
- bias field artifact, 118
- Bode plot, 16
- Bohl functions, 53
- canonical form, 21
- cardiovascular, 7
- cascade of filter banks, 31, 75
- Cauchy-Schwarz inequality, 52
- causality, 20
- closure, 31
- coefficient of contrast, 123
- coefficient of variance, 123
- coefficients
 - approximation, 28
 - scaling, 28
 - wavelet, 28
- compact support, 29, 40
- Computed Tomography, 118
- conjugate transpose, 42
- conservation of energy, 29, 72
- continuous
 - Fourier transform, 16
 - time, 15
 - wavelet transform, 26
- controllability Grammian, 60
- controllability matrix, 61
- controllable companion form, 61
- convolution, 17, 45
- convolution kernel, 51
- critical sampling, 28, 36
- Daubechies wavelets, 34
- decision stage, 8
- decomposition, 31
- defibrillator, 14
- depolarization, 8
- depolarized, 9
- design, wavelet, 72
- detail
 - coefficients, 28
 - signal, 33
- detail space, 32
- digital, 15
- dilation, 26
 - equation, 32
- Dirac delta, 18–20, 59
- direct feedthrough, 21
- discrete

- Fourier series, 17
 - Fourier transform, 17
 - time, 15
 - wavelet transform, 27
- domain, 15
- double-shift orthogonality, 30
- downsampling, 28
- dyadic points, 32
- dyadic scales, 31
- dynamic translinear circuits, 43
- dynamical matrix, 21
- ECG, 10
 - intracardiac, 10
 - subcutaneous, 10
 - surface, 10
- electrical systole, 13
- electrocardiogram, 10
- elementary J-inner factors, 93
- energy distribution, 73
- Euler's formula, 16
- even phase, 35
- exponent
 - Hölder, 34
 - Lipschitz, 34
- fibrillation
 - atrial, 14
 - ventricular, 14
- filter bank, 27, 72
- filter synthesis, 48
- final value theorem, 19
- finite energy, 16
- finite impulse response, 23, 29, 89
- Fourier
 - series, 16
 - discrete, 17
 - transform, 15
 - continuous, 16
 - discrete, 17
 - windowed, 18
- frequency domain, 15
- frequency response function, 22
- full rate, 34
- function, vector, 40
- Gabor transform, 18
- Gaussian wavelet, 27
- global optimization, 51, 57, 86
- Hölder exponent, 34
- Haar wavelet, 28
- Hadamard product, 123
- half rate, 35
- Halmos extension, 95
- Hankel matrix, 61
- Hankel singular values, 60
- heart, 8
- Heisenberg uncertainty rectangle, 26
- Hermitian transpose, 42
- high-pass filter, 28
- homomorphic filtering, 119
- implantable devices, 7, 43
- impulse
 - continuous-time, 18, 20
 - discrete-time, 18
 - response function, 20
- impulse response, 45
- indicator function, 54
- initial phase angles, 16
- initial value theorem, 18
- inner-product, 51
- input matrix, 21
- input vector, 21
- integral wavelet transform, 26
- intracardiac ECG, 10
- iteration scheme, 32
- Kronecker delta, 18, 20, 59
- L_2 -approximation, 51
- Laplace transform, 18
- lattice form, 75
- lead, 10
- linear fractional transform, 93
- linear phase, 22, 34
- linear system, 19
 - time-invariant, 19
- Lipschitz exponent, 34
- local optima, 82
- lossless systems, 41, 89, 90
- low-pass filter, 28
- Lyapunov-Stein equations, 60
- magnetic resonance imaging, 118
- Mallat's algorithm, 31
- matrix exponential, 53
- McMillan degree, 20
- Mexican Hat wavelet, 27
- modulation matrix, 29
- mother wavelet, 34
- moving average filters, 23
- multi-resolution, 25, 31
- multiplicative noise, 119
- multiwavelet, 34, 40, 89
 - design, 89
 - parameterization, 42

- myocardium, 8
- normalized convolution, 123
- nuclear magnetic resonance imaging, 118
- Nyquist frequency, 15
- Nyquist rate, 15
- observability grammian, 61
- observability matrix, 61
- odd phase, 35
- orthogonal complement, 32
- orthogonal filter banks, 29
- orthogonal wavelets, 75
- orthogonality condition, 29
- orthonormal wavelets, 25
- output matrix, 21
- output vector, 21
- overcomplete wavelet transform, 37
- P wave, 12
- pacemaker, 8
- pacemaker cells, 9
- Padé approximation, 45
- Padé approximation, 48
- Parseval's identity, 51
- perfect reconstruction, 29
- phantom, 118
- phase, 34
- polarized, 8
- poles, 20
- polyphase filters, 34, 75, 76
- polyphase matrix, 35
- power, 16
- power complementary, 29
- PP interval, 13
- PQ segment, 12
- proper rational transfer function, 20
- pseudofrequency, 26
- QRS complex, 12
- QT interval, 13, 114
- quadrature mirror filters, 30
- radio frequency inhomogeneity, 118
- reduced rate, 35
- region of convergence, 18
- regularity, 34
- repolarization, 9, 12
- resting potential, 8
- rhythm, 14
- rotation matrix, 76
- RR interval, 13
- sampling
 - frequency, 15
 - rate, 15
- scaling
 - coefficients, 28
 - filter coefficients, 30
 - function, 31
- Schur
 - form, 98
 - tangential algorithm, 93
 - vector, 94
- segments (ECG), 12
- separation of exponentials, 54
- short time Fourier transform, 18
- signal processing, 7
- similarity transformation, 21
- sino-atrial node, 9
- sinusoidal fidelity, 22
- Smith-Barnwell orthogonality conditions, 40
- space, approximation, 31
- space,detail, 32
- sparsity, 72
- ST segment, 12
- stability, 50, 54, 90
- starting point, 57
- state matrix, 21
- state vector, 21
- state-space, 20
 - dimension, 21
 - representation, 21, 54
- stationary wavelet transform, 37
- stationary wavelet transform, polyphase, 38
- step response function, 20
- strictly proper rational function, 45
- subcutaneous ECG, 10
- sudden cardiac death, 14
- surface ECG, 10
- syncytium, 14
- synthesis, 16
- system matrix, 21
- T wave, 12, 114
- tangential Schur algorithm, 93
- Taylor series expansion, 48
- time domain, 15
- time-frequency localization, 25
- time-invariant systems, 19
- time-shifting, 26, 47
- transfer function, 20
 - proper rational, 20
- transform
 - Fourier, 15
 - continuous, 16
 - discrete, 17

- windowed, 18
- Gabor, 18
- Laplace, 18
- wavelet, 25
 - continuous, 26
 - discrete, 27
- truncation error, 47

- uncertainty rectangle, 18, 26
- upsampling, 28

- vanishing moments, 33, 55, 79, 80, 101
- variance maximization, 73
- vector function, 40
- vectorfunctions, 40
- ventricle, 8
- ventricular fibrillation, 14

- wavelet, 71
 - approximation, 45
 - basis, 33, 71
 - choosing a, 72
 - coefficients, 28
 - design, 72
 - equation, 32
 - filter coefficients, 30
 - function, 31, 32
 - Gaussian, 27
 - Mexican Hat, 27
 - multi, 40, 89
 - orthonormal, 25, 75
 - transform, 25
 - continuous, 26
 - discrete, 27
 - overcomplete, 37
 - stationary, 37
- wavelets
 - Daubechies, 34
- waves (ECG), 12
- windowed Fourier transform, 18

- zero-order hold, 18, 59
- zeros, 20
- zoom-in property, 26