



AALBORG UNIVERSITY

DENMARK

P2 PROJECT
MATHEMATICAL ENGINEERING

Fault Localisation in Wind Turbine Machinery

Technological Application of Linear Algebra

Authors:

Alexander Topholm Bertelsen
Benjamin Bitsch Knudsen
Lasse Ryge Andersen
Mads Arnløv Jørgensen
Magnus Berg Ladefoged
Sofus Gøgsig Broberg

Supervisor:

Jan Østergaard
Assistant supervisor:
Henrik Worm Routhe

27th May 2019

Copyright © Aalborg University 2018

The project is written in Overleaf v2 and compiled with pdfL^AT_EX. Calculations are conducted in Python 3.6.5 and Maple.



Dept. of Mathematical Sciences

Skjernvej 4A
DK-9220 Aalborg Ø
<http://math.aau.dk>

AALBORG UNIVERSITY

STUDENT REPORT

Title:

Fault Localisation in Wind Turbine
Machinery

Theme:

Technological Application of Linear Algebra

Project Period:

Spring Semester 2019

Project Group:

MATTEK A-411

Participant(s):

Alexander Topholm Bertelsen
Benjamin Bitsch Knudsen
Lasse Ryge Andersen
Mads Arnløv Jørgensen
Magnus Berg Ladefoged
Sofus Gøgsig Broberg

Supervisor(s):

Jan Østergaard
Henrik Worm Routhe

Copies: 1

Numbered Pages: 109

Date of Completion:

27th May 2019

Abstract:

The purpose of the project is to locate a faulty component in wind turbine machinery. This localisation is facilitated using the discrete Wavelet transform and the discrete Fourier transform separately. The project concludes that the discrete Wavelet transform can be used to facilitate an accurate localisation of an impulse sound with fair consistency. Furthermore, the project concludes, that the discrete Fourier transform is suitable for detecting whether a fault, which results in a change of frequency, has occurred. However, the discrete Fourier transform did not prove suitable for facilitating the localisation of a faulty component.

Preface

We have written this project to gain insight of a technological application of linear algebra, by applying the discrete Wavelet transform and the discrete Fourier transform to a technological problem. Furthermore, the purpose of the project is to design and implement numerical software to analyse data gathered from an experiment. The project explores the transforms and some of their applications in the fields of mathematics and engineering. The distribution of the content of the project is deemed to be weighted evenly among mathematics and engineering.

We would like to thank our supervisors Jan Østergaard and Henrik Worm Routhe for advising us during the project. We would also like to thank our lecturer Jan Dimon Bendtsen for recommending some reliable sources related to signals and sampling.

SYMBOL TABLE

Symbol	Description
A	Matrix
A^{-1}	Inverse of a matrix
A^T	The transpose of a matrix
\mathbf{u}	Vector
$[A \mid \mathbf{u}]$	Augmented matrix
$card(\mathbf{u})$	Cardinality of a vector
$\ \mathbf{u}\ $	The norm of a vector
•	Dot product
\oplus	Direct sum
*	Linear convolution
\circledast	Circular convolution
★	Cross-correlation
$(\downarrow 2)$	Downsampling by 2
$(\uparrow 2)$	Upsampling by 2
$\mathcal{F}()$	discrete Fourier transform
\mathbb{N}	Set of natural numbers
\mathbb{Z}	Set of integers
\mathbb{R}	Set of real numbers
\mathbb{C}	Set of complex numbers

Contents

1	Introduction	1
1.1	Problem Analysis	1
1.1.1	Reliability	1
1.1.2	Availability	2
1.1.3	Operation and Maintenance Costs	3
1.1.4	Monitoring	4
1.2	Problem Statement	4
1.3	Project Scope	5
2	System Description	7
3	Linear Algebra	9
3.1	Introduction	9
3.2	Matrices	9
3.2.1	Matrix Transposes	11
3.3	Vectors	12
3.4	Matrix-Vector Product	17
3.5	Matrix Multiplication	20
3.6	Linear and Matrix Transformations	21
3.7	Systems of Linear Equations	22
3.7.1	Reduced Row Echelon Form	23
3.8	Invertibility	24
3.9	Orthogonality and Subspaces	26
4	Signals and Sampling	33
4.1	Introduction	33
4.2	Signals	33
4.2.1	Sampling	34
4.3	Sound Sensors	36
5	Convolution	39
5.1	Introduction	39
5.2	Discrete Convolution	39
5.2.1	Cross-Correlation	41
6	Filters	43
6.1	Introduction	43
6.2	Low-pass and High-pass Filters	43
6.3	Frequency Response	44

6.4	Filter Bank	45
6.5	Half-band Filters	48
7	The Discrete Wavelet Transform	51
7.1	Introduction	51
7.2	Matrix Formulation	51
7.3	Properties of the Transform	55
7.4	The Transform and Subspaces	55
8	Selected Filters	57
8.1	Introduction	57
8.2	Daubechies 16	57
8.2.1	Frequency Response	58
9	Filtering	59
9.1	Introduction	59
9.2	Filtering with the DWT	59
10	The Discrete Fourier Transform	63
10.1	Introduction	63
10.2	Prerequisite	63
10.3	Fourier Series	63
10.4	Discrete Fourier Transform	64
11	Identifying a New Frequency	67
11.1	Introduction	67
11.2	Identifying a New Frequency through the DFT	67
12	Localisation through Time Delay Estimation	69
12.1	Introduction	69
12.2	Time Delay Estimation through Cross-correlation	69
12.3	Triangulation of Position in Two Dimensions	70
12.3.1	Practical Procedure	71
13	Practical Experiment	73
13.1	Introduction	73
13.2	Setup	73
13.3	Procedure	74
14	Processing of Data from Experiment	77
14.1	Introduction	77
14.2	DWT Analysis	78
14.2.1	Description of the DWT Analysis Procedure	78
14.2.2	Analysis of Impulses	79
14.3	DFT Analysis	84
14.3.1	Description of the DFT Analysis Procedure	84
14.3.2	Analysis of Frequency Altering	84

15 Discussion	89
15.1 Introduction	89
15.2 Experimental Uncertainties and Errors	89
15.3 Results from the DWT Analyses	90
15.3.1 Successful Localisation	90
15.3.2 Unsuccessful Localisation	90
15.4 Results from the DFT Analysis	91
15.4.1 Errors of the Localisation	91
16 Conclusion	93
Appendices	97
A Experiment	99
A.1 Equipment	99
A.2 Table of Conducted Recordings	99
A.3 Practical Setup	100
B Hyperbolas from Data Processing	103
B.1 Recording 1	103
B.2 Recording 3	104
B.3 Recording 4	105
B.4 Recording 5	106
B.5 Recording 7	107
B.6 Recording 8	108

1 | Introduction

1.1 Problem Analysis

A crisis is occurring in present time; fossil fuels are causing drastic climate changes, and human society is looking to expand on sustainable energy. [NASA, 2009] Because of this, renewable energy sources are on a rise in recent decades. Among these sustainable energy sources is wind power, which has been expanding drastically since the early 2000s according to the Global Wind Energy Council (GWEC). In Figure 1.1, the expansion in global wind capacity from 2001 to 2017 is shown. According to the GWEC, the amount of installed wind capacity will continue to grow even more rapidly in the upcoming years. [GWEC, 2017]

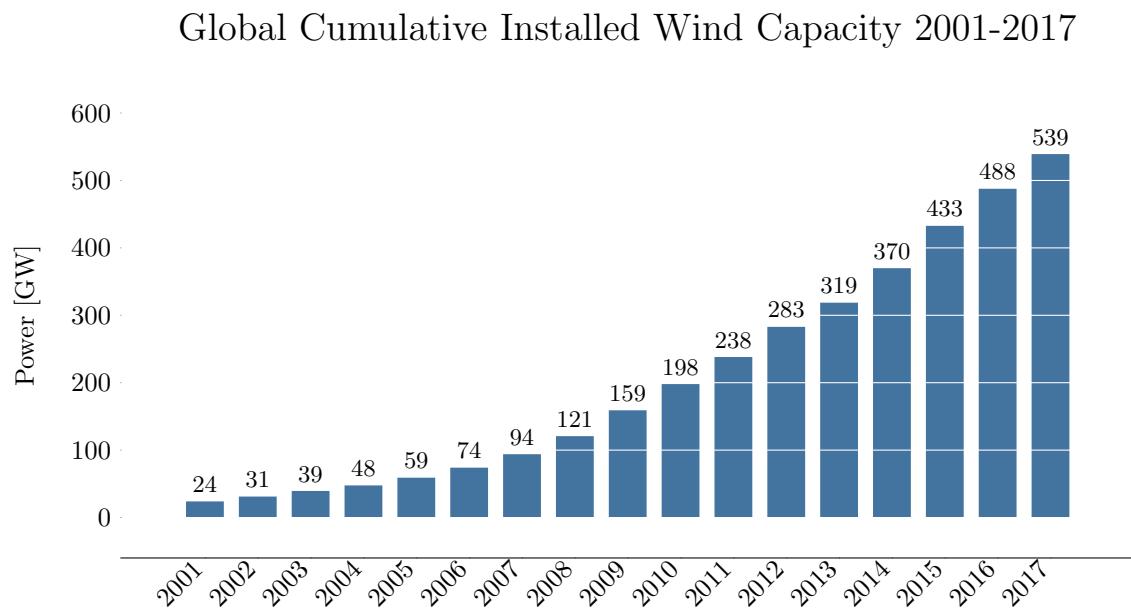


Figure 1.1: The chart is based on statistics by GWEC. [GWEC, 2017]

1.1.1 Reliability

Although there has been an increased interest in wind turbines as a renewable and sustainable energy source in the later years, they still present some challenging aspects for the modern world, one of which is their reliability. The reliability of a wind turbine is determined by its ability to operate as required without failure under the given conditions. Wind turbines are very complex structures that consist of many different components each of which has a possibility of developing faults. This can result in dire consequences, and

in a worst-case scenario it can lead to death of people nearby. As the number of active wind turbines increase so does the number of accidents happening every year. The data in Figure 1.2 contains reported accidents involving blade failures, fire outbreaks, structural failures, component failures, and mechanical failures. Many accidents are not publicly reported, and the actual number of yearly accidents is presumably higher.

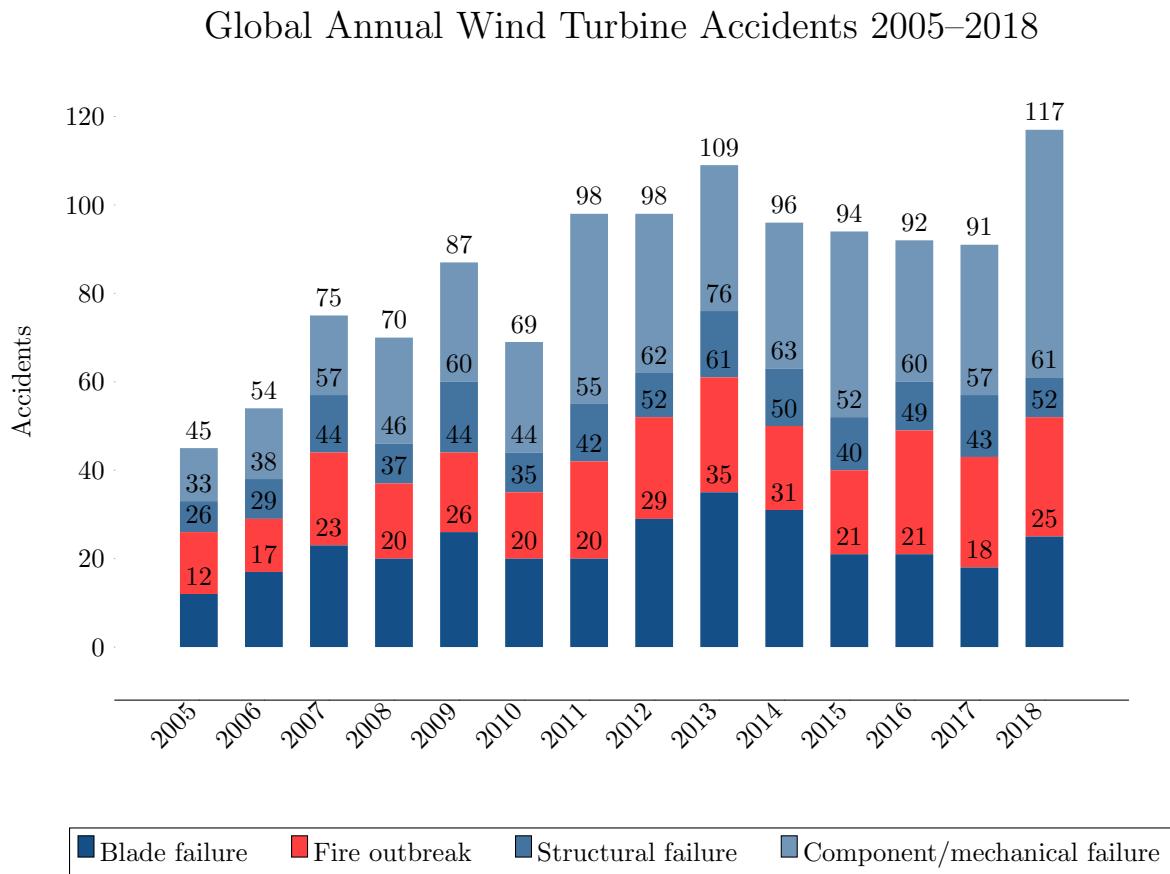


Figure 1.2: Each bar accumulates the amount of accidents of the four types. The chart is based on statistics by CWIF. [CWIF, 2018]

It is desirable to improve the reliability since it increases the operational uptime of the wind turbine and reduces maintenance costs.

1.1.2 Availability

To maximise the potential of wind turbines, it is important to guarantee high performance uptime. If the wind and temperature are within the limits of which the wind turbine is able to operate, then the availability is defined as [GL, 2017]

$$\text{Availability} = \frac{\text{Active hours during a period}}{\text{Total hours during a period}} \cdot 100\%.$$

The factors that can decrease the active hours of the wind turbines during the period are: faults, planned outages, and maintenance. The availability is a key factor for evaluating the efficiency of wind turbines and ought to be as high as possible.

Except for being a measure of energy estimates and performance, the availability is also used for warranties and can have influence on legal, financial, and technical matters.

The availability can be increased by reducing the downtime caused by component failures. A survey conducted by WMEP (Wissenschaftlichen Mess- und Evaluierungsprogramms) gathered 64,000 maintenance reports from 1,500 on-shore wind turbines in Europe during the period 1989–2006. The extent of the survey is 15,357 wind turbine years and highlights the failure rates and downtime of the wind turbines.

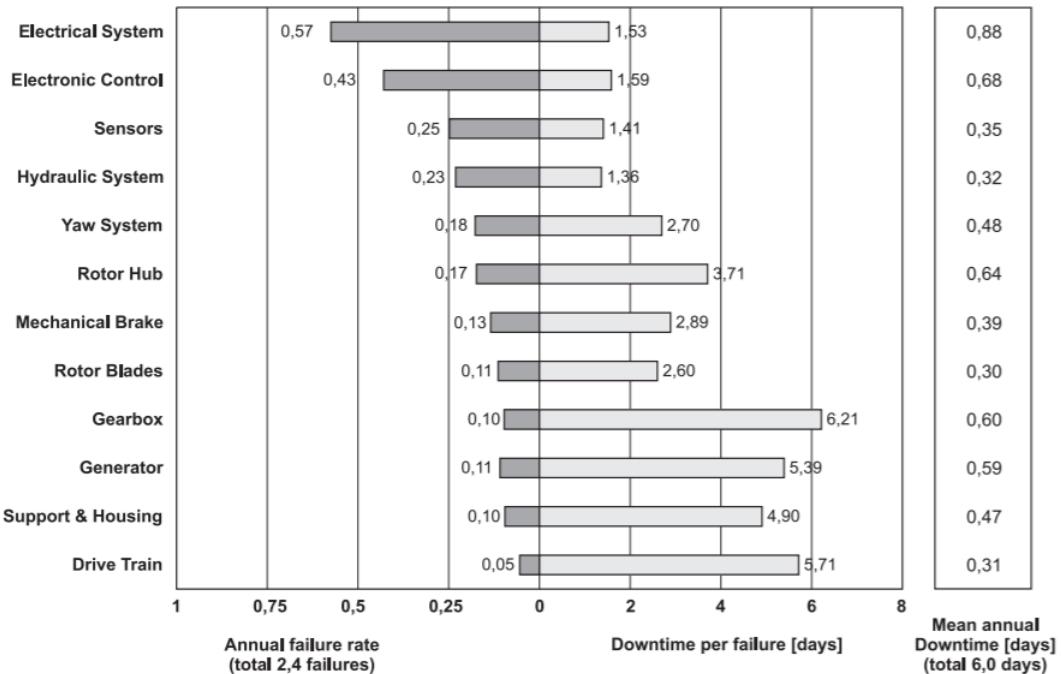


Figure 1.3: Failure rate of 1,500 wind turbines in the span of 17 years. [S. Faulstich, 2011]

According to Figure 1.3, the most frequently occurring failures in wind turbines are related to the electrical system(s). However, the downtime of the mechanical components is much higher per failure than the downtime of electrical components. Despite the low frequency of mechanical failures, the average annual downtime is still high for those components due to the long downtime per failure and excessive maintenance required. [S. Faulstich, 2011]

1.1.3 Operation and Maintenance Costs

The profitability is another challenge that arises with the increased demand of wind turbines. The profitability of wind turbines is affected both by their reliability and availability. Malfunctions and faults require maintenance while decreasing the efficiency of wind turbines.

Naturally it is desirable to lower the cost of such maintenance and increase the uptime of wind turbines. The operation and maintenance costs make up a significant amount of the total lifetime cost distribution of a wind turbine project. It typically accounts for 20%–25%. [IRENA, 2018]

The cost of operation and maintenance of wind turbines is especially significant due to the strategy concerning the maintenance. Periodical maintenance is a strategy which involves inspection of the wind turbines at certain times. This is usually conducted twice a year. This maintenance strategy has some flaws since it requires a stoppage of the wind turbines for assessment and maintenance. Furthermore, the wind turbines are assessed at

fixed intervals, and therefore, the condition of the wind turbines is unknown between the intervals. This can lead to an accumulation of faults and thus a lack of maintenance or unnecessary shutdowns and thus excessive maintenance.

A preferable maintenance strategy is remotely monitoring the condition of the wind turbines. This enables a frequent assessment of the components of the wind turbines without interrupting the operation and thereby planning of maintenance practice when needed.

1.1.4 Monitoring

The health monitoring of a wind turbines is conducted online by collecting data on the conditions of the components and subsystems. This data can then be analysed and used to detect and locate faults or even predict impending damage. Among the most common health monitoring approaches are vibration, electrical, and thermal monitoring to gain features of the faults. Signal processing based analysis can then be utilised to diagnose and predict faults.

The mechanical machinery in a wind turbine emits sound due to the moving components. For constant rotational velocities the sound emitted tends to be tonal and thus periodic. In Figure 1.4 a typical amplitude spectrum of machinery induced noise from a wind turbine can be seen.

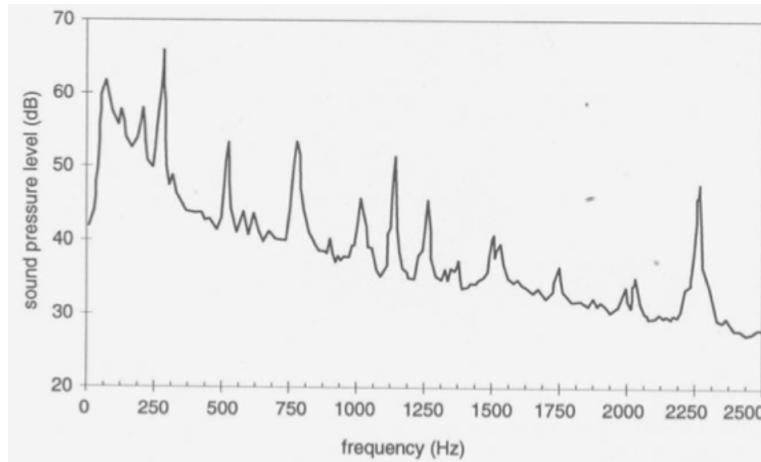


Figure 1.4: Typical frequency spectrum of machinery induced noise. [Lehto, 2014, p.6]

A mechanical fault can cause a component to emit a detectable sound. The detectable sounds can consist of impulses caused by a crack in a component and a change of frequency when the fault changes the rotational velocity.

1.2 Problem Statement

The wind industry is continuously growing, and the total installed wind capacity reached 539 GW at the end of 2017 and is predicted to further increase in the upcoming years. In the wind industry a key factor for profit is to ensure system safety and uninterrupted operation of the wind turbines. If the wind turbines should malfunction it can result in increasingly large profit losses. To avoid failures which may cause malfunctions and thereby unplanned interruptions of the wind turbines, monitoring the conditions of the components of the wind turbines is crucial. If slight disturbances that can cause failures can be detected

and localised during operation, then the planning and performance of maintenance of the wind turbines can be done more efficiently.

The goal of this project is to detect when and where faults occur within a wind turbine, so that they can be repaired and maintained quickly, before causing further problems.

The faults can be measured as a change in the sound signals produced by the components, and this change can indicate that a fault has occurred. Naturally, a lot of noise will occur when measuring a signal in a nacelle due to the many components as well as the reverb. Locating the faulty component can be done using cross-correlation to approximate a time delay, but since this method does not perform well on noisy signals, the recorded sound signals will have to be denoised.

Two frequently used methods for signal processing are the Fourier transform and the Wavelet transform. They can among other things be used to denoise signals. The two transforms have different properties making them preferable for different scenarios, however, both can be used to facilitate the localisation of a change in a signal.

The discrete Fourier transform is well suited for analysing periodic signals as it can accurately find the specific frequency components of the signal. The discrete Wavelet transform has different properties depending on the applied filters and is capable of filtering the signal for either low or high frequencies, which makes it suitable for detecting specific signal structures such as impulses. The discrete Wavelet transform is an application of linear algebra, and it will be studied using matrix formulations.

This leads to the problem statement:

How can the Wavelet and Fourier transforms separately facilitate the localisation of potential mechanical failures in wind turbine machinery?

1.3 Project Scope

When monitoring wind turbine machinery, different types of signals can be measured. This project is limited to sound signals recorded by microphones. In a wind turbine the sound spectrum is composed of many sounds produced by all the different components. Four components with a high average annual downtime are the gearbox, generator, brake, and pitch system. Therefore, the project is limited to a sound signal consisting of four sounds as a proof of concept.

The experiment will be conducted in a room with reverb to simulate the reverb in a nacelle. Noise is added to the sounds in the experiment to simulate the ambient sounds surrounding wind turbines such as wind.

The localisation of the faulty component will be performed in two dimensions. An angle is deemed sufficient as a result of the localisation of the faulty component.

In this project the Wavelet transform is limited to its discrete form. The project mainly focuses on the theory of the discrete Wavelet transform and less on the Fourier transform. Filters are an essential part of the Wavelet transform, and this project is limited to causal FIR-filters.

2 | System Description

The concept of detecting and locating a fault using the discrete Fourier transform and the discrete Wavelet transform is outlined in Figures 2.1a and 2.1b respectively. The purpose is to compute how accurately the faulty component can be located by the discrete Fourier transform and the discrete Wavelet transform respectively. The applicability of the two transforms can thus be discussed and evaluated.

The procedure presented in Figure 2.1a involves sampling a signal and analysing it using the discrete Fourier transform. The purpose is to identify a frequency that was not present in the signal at first, indicating a fault. That specific frequency is then used to compute a time delay between the microphones using cross-correlation. The time delay estimation (TDE) can then be used to locate the faulty component using triangulation.

The procedure in Figure 2.1b describes a similar process, except the sampled signal is analysed using the discrete Wavelet transform instead of the discrete Fourier transform. The purpose of the discrete Wavelet transform is to filter out unwanted frequencies in order to compute a TDE with cross-correlation and thereby allowing triangulation.

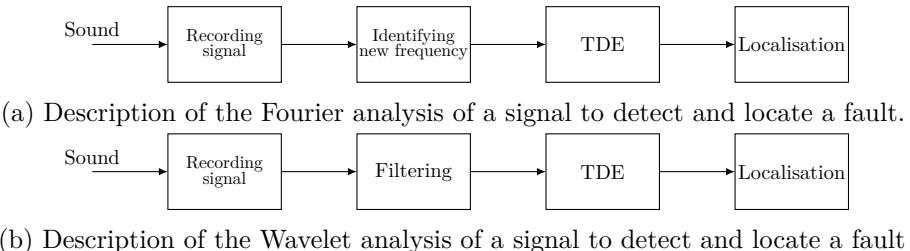


Figure 2.1: System description of signal analysis.

3 | Linear Algebra

3.1 Introduction

In this chapter the essential concepts of linear algebra crucial to the understanding of filters and the discrete Wavelet transform are presented. This chapter is based on the book “Elementary Linear Algebra”. [Spence E., 2008]

3.2 Matrices

A matrix is a data structure that can be used to store information. A matrix is a combination of rows and columns, and a matrix with m rows and n columns is called an m by n matrix (denoted $m \times n$). The matrix is square if and only if $m = n$, and it is otherwise rectangular. The entry at the i th row and j th column is referred to as the (i, j) -entry of the matrix. Matrices are denoted by uppercase letters and have their entries denoted by the corresponding lowercase letter. An example matrix A can be written as

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix},$$

where the real numbers m and n are referred to as scalars in this and subsequent chapters.

Addition of matrices is defined in Definition 3.1, and the scalar-matrix product is defined in Definition 3.2.

Definition 3.1 (Matrix Addition)

Let A , B , and C be $m \times n$ matrices. Addition is then defined as

$$A + B = C,$$

such that

$$a_{ij} + b_{ij} = c_{ij}.$$

[Spence E., 2008, p. 5]

Definition 3.2 (Scalar-Matrix Product)

Let A be an $m \times n$ matrix and let $c \in \mathbb{R}$ be a constant, then the scalar-matrix product is given by

$$cA = \begin{bmatrix} ca_{11} & ca_{12} & \dots & ca_{1n} \\ ca_{21} & ca_{22} & \dots & ca_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ ca_{m1} & ca_{m2} & \dots & ca_{mn} \end{bmatrix}.$$

[Spence E., 2008, p. 5]

Matrices have certain properties for addition and scalar multiplication as described in Theorem 3.1.

Theorem 3.1 (Properties for Matrix Addition and Scalar Multiplication)

Let A , B , and C be $m \times n$ matrices and let s and t be real-valued arbitrary scalars. The matrices have the following properties for addition and multiplication.

- a) $A + B = B + A$. (Commutative law of matrix addition)
- b) $(A + B) + C = A + (B + C)$. (Associative law of matrix addition)
- c) $A + O = A$. (O is the $m \times n$ zero matrix)
- d) $A + (-A) = O$.
- e) $(st)A = s(tA)$.
- f) $s(A + B) = sA + sB$.
- g) $(s + t)A = sA + tA$.

[Spence E., 2008, p. 6]

Proof 3.1

Let A , B , and C be $m \times n$ matrices and let s and t be real-valued arbitrary scalars.

- a) By Definition 3.1 the (i,j) -entry of $A + B$ is given by the sum of the (i,j) -entry of A and B respectively. Thus

$$\begin{aligned} a_{ij} + b_{ij} &= b_{ij} + a_{ij} \\ A + B &= B + A, \end{aligned}$$

since the law of scalar addition holds. Thereby a) is proven.

- b) By Definition 3.1, the (i,j) -entry of $(A + B) + C$ is given by the sum of the (i,j) -entries of A , B , and C as

$$\begin{aligned} (a_{ij} + b_{ij}) + c_{ij} &= a_{ij} + (b_{ij} + c_{ij}) \\ (A + B) + C &= A + (B + C), \end{aligned}$$

since the law of associativity holds for scalars. Thereby b) is proven.

- c) Let O be the $m \times n$ zero matrix for which all entries are zero. Since matrix addition is entrywise as $a_{ij} + 0 = a_{ij}$, it leaves the matrix A unchanged:

$$a_{ij} + 0 = a_{ij} \Rightarrow A + 0 = A,$$

proving c).

- d) Matrix addition is entrywise, thus the (i,j) -entry sum of $A + (-A)$ is given by

$$a_{ij} + (-a_{ij}) = 0,$$

resulting in all entries of A being 0, thereby proving d).

- e) The (i,j) -entry of the scalar-matrix product of $(st)A$ is given by

$$(st)a_{ij} = sta_{ij}.$$

This is true since the associative law applies to multiplication of scalars. Multiplying the matrix with the scalar t does not change the result

$$s(ta_{ij}) = sta_{ij}.$$

As such e) is proven.

- f) The (i,j) -entry of the scalar-matrix product $s(A + B)$ is given by

$$s(a_{ij} + b_{ij}) = sa_{ij} + sb_{ij},$$

which corresponds to the entries in $sA + sB$. The proof of f) is thereby complete.

- g) The (i,j) -entry of the scalar-matrix product $(s + t)A$ is given by

$$(s + t)a_{ij} = sa_{ij} + ta_{ij}.$$

This corresponds to the (i,j) -entries of $sA + tA$. Thus ending the proof of g). ■

3.2.1 Matrix Transposes

The rows and columns of a matrix A can be switched resulting in a matrix called the transpose of A .

Definition 3.3 (The Matrix Transpose)

Let A be an $m \times n$ matrix. A matrix B of size $n \times m$ whose (i,j) -entries are the (j,i) -entries of A is called the transpose of A . The transpose of a matrix A is denoted as $A^T = B$, and $B^T = A$.

[Spence E., 2008, p. 7]

Theorem 3.2 (Properties of the Transpose of a Matrix)

Let A and B be $m \times n$ matrices, and let s be an arbitrary real-valued scalar. For the matrices A and B the following properties apply:

- a) $(A + B)^\top = A^\top + B^\top$.
- b) $(sA)^\top = sA^\top$.
- c) $(A^\top)^\top = A$.

[Spence E., 2008, p. 7]

Proof 3.2

Let A and B be $m \times n$ matrices.

- a) By the definition of the transpose of a matrix it is known that the (i, j) -entry of A^\top equals the (j, i) -entry of A . The same applies for B , thus the entrywise addition of $A^\top + B^\top$ is by Definition 3.1: $a_{ji} + b_{ji} = c_{ji}$. The entrywise addition of $A + B$ is $a_{ij} + b_{ij} = c_{ij}$ with the transpose $c_{ij}^\top = c_{ji}$. Thus ending the proof a).
- b) Entrywise $(sA)^\top$ can be written as $s a_{ji}$ since the entry of (sA) is $s a_{ij}$. Whereas sA^\top can be written as $s a_{ji}$ since the entry of A^\top is a_{ji} . Thus ending the proof of b).
- c) By definition the transpose of A is another matrix denoted A^\top , whose (i, j) -entry is the (j, i) -entry of A . By the same definition the transpose of A^\top is a matrix $(A^\top)^\top$ whose (j, i) -entry is the (i, j) -entry of A^\top . Since the (j, i) -entry of both A and $(A^\top)^\top$ equals the (i, j) -entry of A^\top , the proof of c) is completed.

■

3.3 Vectors

A vector is either a matrix with only one row ($1 \times n$) or one column ($m \times 1$), either one referred to as either a row or a column vector respectively. Most commonly column vectors will be used. A matrix A can also be expressed as either a row vector of column vectors or a column vector of row vectors:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} = [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \dots \quad \mathbf{a}_n] = \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \vdots \\ \mathbf{r}_m \end{bmatrix},$$

where $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ are column vectors in \mathbb{R}^m and $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_m$ are row vectors in \mathbb{R}^n .

The fundamental operations of arithmetic for matrices apply to vectors. The cardinality and norm of a vector is respectively defined in Definition 3.4 and 3.5.

Definition 3.4 (Cardinality)

Let $\mathbf{u} = \{u_1, u_2, \dots, u_n\}$ be a vector in \mathbb{R}^n , then the cardinality of \mathbf{u} is

$$\text{card}(\mathbf{u}) = n.$$

Definition 3.5 (Norm of a Vector)

Let $\mathbf{u} = \{u_1, u_2, \dots, u_n\}$ be a vector in \mathbb{R}^n , the norm of the vector denoted $\|\mathbf{u}\|$ is then

$$\|\mathbf{u}\| = \sqrt{u_1^2 + u_2^2 + \dots + u_n^2}.$$

[Spence E., 2008, p. 361]

A vector whose norm is 1 is called a unit vector. For vectors the operator called dot product is the summation of the components of two vectors multiplied entrywise.

Definition 3.6 (Dot Product)

Let $\mathbf{u} = \{u_1, u_2, \dots, u_n\}$ and $\mathbf{v} = \{v_1, v_2, \dots, v_n\}$ be vectors with $\text{card}(\mathbf{u}) = \text{card}(\mathbf{v}) = n$, then the dot product is given by

$$\mathbf{u} \cdot \mathbf{v} = \sum_{k=1}^n u_k v_k.$$

[Spence E., 2008, p. 363]

The dot product has certain properties that are utilised in later proofs. These properties are described in Theorem 3.3.

Theorem 3.3 (Geometrical Properties of Vectors)

Let \mathbf{u} , \mathbf{v} , and \mathbf{w} be vectors in \mathbb{R}^n and let c be a real-valued arbitrary scalar, then the following applies.

- a) $\mathbf{u} \cdot \mathbf{u} = \|\mathbf{u}\|^2$
- b) $\mathbf{u} \cdot \mathbf{u} = 0$ if $\mathbf{u} = \mathbf{0}$. ($\mathbf{0}$ is the zero vector)
- c) $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$
- d) $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$.
- e) $(c\mathbf{u}) \cdot \mathbf{v} = c(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \cdot (c\mathbf{v})$.
- f) $\|c\mathbf{u}\| = |c|\|\mathbf{u}\|$.

[Spence E., 2008, p. 364]

Proof 3.3

Let \mathbf{u}, \mathbf{v} , and \mathbf{w} be vectors in \mathbb{R}^n and let $c \in \mathbb{R}$ be an arbitrary scalar.

- a) From Definition 3.6 it follows that

$$\mathbf{u} \cdot \mathbf{u} = (u_1^2 + u_2^2 + \cdots + u_n^2) = \left(\sqrt{u_1^2 + u_2^2 + \cdots + u_n^2} \right)^2 = \|\mathbf{u}\|^2,$$

as the dot product is a sum of exponentiated scalars the operations are valid, completing the proof of a).

- b) Assuming $\mathbf{u} = \mathbf{0}$ the dot product follows from Definition 3.6

$$\mathbf{u} \cdot \mathbf{u} = (0^2 + 0^2 + \cdots + 0^2) = 0,$$

which is the sought result, ending the proof of b).

- c) It does not matter in which order the dot product of two vectors is calculated. Following Definition 3.6

$$\begin{aligned} \mathbf{u} \cdot \mathbf{v} &= u_1 v_1 + u_2 v_2 + \cdots + u_n v_n \\ \mathbf{v} \cdot \mathbf{u} &= v_1 u_1 + v_2 u_2 + \cdots + v_n u_n, \end{aligned}$$

which is a sum of multiplied scalars for which the associative law applies, completing the proof of c).

- d) By Definition 3.6

$$\begin{aligned} \mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) &= u_1(v_1 + w_1) + u_2(v_2 + w_2) + \cdots + u_n(v_n + w_n) \\ &= (u_1 v_1 + u_2 v_2 + \cdots + u_n v_n) + (u_1 w_1 + u_2 w_2 + \cdots + u_n w_n), \end{aligned}$$

which is equal to the sum of the two dot products

$$\mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w} = (u_1 v_1 + u_2 v_2 + \cdots + u_n v_n) + (u_1 w_1 + u_2 w_2 + \cdots + u_n w_n),$$

thus proving d).

- e) By Definition 3.6, the dot product of two vectors is defined for their components as

$$\begin{aligned} (\mathbf{c}\mathbf{u}) \cdot \mathbf{v} &= (cu_1)v_1 + (cu_2)v_2 + \cdots + (cu_n)v_n \\ c(\mathbf{u} \cdot \mathbf{v}) &= c(u_1 v_1 + u_2 v_2 + \cdots + u_n v_n) \\ \mathbf{u} \cdot (c\mathbf{v}) &= u_1(cv_1) + u_2(cv_2) + \cdots + u_n(cv_n), \end{aligned}$$

for which the properties of scalars apply, proving e).

- f) The norm of a vector is given by Definition 3.5, and by Definition 3.2 vector scaling is defined entrywise, yielding

$$\begin{aligned} \|\mathbf{c}\mathbf{u}\| &= \sqrt{c^2 u_1^2 + c^2 u_2^2 + \cdots + c^2 u_n^2} \\ |c| \|\mathbf{u}\| &= |c| \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2}. \end{aligned} \tag{3.1}$$

Equation (3.1) can be rewritten by the rules of square roots for scalars:

$$\begin{aligned} \sqrt{c^2 u_1^2 + c^2 u_2^2 + \cdots + c^2 u_n^2} &= \sqrt{c^2} \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2} \\ &= |c| \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2}. \end{aligned}$$

Thereby ending the proof of f). ■

From the dot product orthogonality of vectors can be defined.

Definition 3.7 (Orthogonal Vectors)

Let \mathbf{u} and \mathbf{v} be vectors in \mathbb{R}^n . The two vectors are orthogonal if and only if

$$\mathbf{u} \cdot \mathbf{v} = 0.$$

[Spence E., 2008, p. 363]

Note that two vectors are said to be orthonormal if they are orthogonal unit vectors.

Theorem 3.4 (Orthogonality)

Let \mathbf{u} and \mathbf{v} be vectors in \mathbb{R}^n . If the two vectors are orthogonal if

$$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2.$$

[Spence E., 2008, p. 365]

Proof 3.4

Let \mathbf{u} and \mathbf{v} be vectors in \mathbb{R}^n . By Definitions 3.5 and 3.6

$$\begin{aligned}\|\mathbf{u} + \mathbf{v}\|^2 &= \left(\sqrt{(u_1 + v_1)^2 + (u_2 + v_2)^2 + \cdots + (u_n + v_n)^2} \right)^2 \\ &= (\mathbf{u} + \mathbf{v}) \cdot (\mathbf{u} + \mathbf{v}) = \mathbf{u} \cdot \mathbf{u} + \mathbf{v} \cdot \mathbf{v} + 2\mathbf{u} \cdot \mathbf{v},\end{aligned}$$

it is noted that $\|\mathbf{u}\|^2 = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2}^2 = \mathbf{u} \cdot \mathbf{u}$ which also applies for \mathbf{v} . By Definition 3.7 the two vectors \mathbf{u} and \mathbf{v} are only orthogonal if $\mathbf{u} \cdot \mathbf{v} = 0$, thus the result is reduced to

$$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2,$$

thereby completing the proof. ■

All vectors can be expressed as a linear combination of scaled standard vectors.

Definition 3.8 (Linear Combination)

A linear combination of the vectors $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k$ is a vector \mathbf{v} on the form

$$\mathbf{v} = c_1 \mathbf{u}_1 + c_2 \mathbf{u}_2 + \cdots + c_k \mathbf{u}_k,$$

where c_1, c_2, \dots, c_k are real-valued scalars and the vectors $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k$ and \mathbf{v} are vectors in \mathbb{R}^n .

[Spence E., 2008, p. 14]

Definition 3.9 (Standard Vectors)

The standard vectors in \mathbb{R}^n are defined as

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} \quad \dots \quad \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}.$$

[Spence E., 2008, p. 17]

By Definitions 3.8 and 3.9, a vector \mathbf{v} in \mathbb{R}^n can be expressed as the linear combination

$$\mathbf{v} = c_1\mathbf{e}_1 + c_2\mathbf{e}_2 + \dots + c_n\mathbf{e}_n = [v_1 \ v_2 \ \dots \ v_n],$$

where c_1, c_2, \dots, c_n are scalars.

From standard vectors the identity matrix can be defined.

Definition 3.10 (Identity Matrix)

The identity matrix I_n is defined as the matrix of the standard vectors $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$:

$$I_n = [\mathbf{e}_1 \ \mathbf{e}_2 \ \dots \ \mathbf{e}_n] = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}.$$

[Spence E., 2008, p. 22]

The term for all the linear combinations of a set of vectors in \mathbb{R}^n is defined in Definition 3.11

Definition 3.11 (Span of Vectors)

For a nonempty set $S = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ of vectors in \mathbb{R}^n , the span of S , denoted by $\text{Span}\{S\}$, is defined as the set of all linear combinations S .

[Spence E., 2008, p. 66]

If a generating set of vectors for a span has the smallest possible size the vectors are linearly independent.

Definition 3.12 (Linear Independence)

A set of vectors $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ is linearly independent if the only scalars c_1, c_2, \dots, c_k such that

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \dots + c_k\mathbf{u}_k = \mathbf{0}$$

are $c_1 = c_2 = \dots = c_k = 0$.

[Spence E., 2008, p. 81]

Orthogonality is a sufficient requirement for linear independence of a set of vectors.

Theorem 3.5 (Linear Independence of Orthogonal Set)

Any orthogonal set of nonzero vectors is linearly independent.

[Spence E., 2008, p. 375]

Proof 3.5

Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ be an orthogonal subset of \mathbb{R}^n , and let c_1, c_2, \dots, c_k be scalars such that

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k = \mathbf{0}.$$

Then for any vector $\mathbf{v}_i \in S$

$$\begin{aligned} 0 &= \mathbf{0} \cdot \mathbf{v}_i \\ &= (c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_i\mathbf{v}_i + \cdots + c_k\mathbf{v}_k) \cdot \mathbf{v}_i \\ &= c_1\mathbf{v}_1 \cdot \mathbf{v}_i + c_2\mathbf{v}_2 \cdot \mathbf{v}_i + \cdots + c_i\mathbf{v}_i \cdot \mathbf{v}_i + \cdots + c_k\mathbf{v}_k \cdot \mathbf{v}_i \\ &= c_i(\mathbf{v}_i \cdot \mathbf{v}_i) \\ &= c_i\|\mathbf{v}_i\|^2. \end{aligned}$$

Since $\|\mathbf{v}_i\|^2 \neq 0$ as $\mathbf{v}_i \neq \mathbf{0}$, hence $c_i = 0$ thus $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ is linearly independent, completing the proof. ■

3.4 Matrix-Vector Product

A matrix-vector product is a linear combination of the components of a vector and the column vectors of a matrix.

Definition 3.13 (Matrix-Vector Product)

Let A be an $m \times n$ matrix with the column vectors $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ and let \mathbf{u} be a column vector in \mathbb{R}^n , the matrix-vector product is then

$$A\mathbf{u} = u_1\mathbf{a}_1 + u_2\mathbf{a}_2 + \cdots + u_n\mathbf{a}_n.$$

[Spence E., 2008, p. 19]

Note that the matrix-vector product is a vector in \mathbb{R}^m , meaning that the matrix is an operator that maps a vector in \mathbb{R}^n to a vector in \mathbb{R}^m .

Theorem 3.6 (Properties of the Matrix-Vector Product)

Let A and B be $m \times n$ matrices and let \mathbf{u} and \mathbf{v} be vectors in \mathbb{R}^n . Then

- a) $A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v}$.
- b) $A(c\mathbf{u}) = c(A\mathbf{u}) = (cA)\mathbf{u} \quad \forall c \in \mathbb{R}$.
- c) $(A + B)\mathbf{u} = A\mathbf{u} + B\mathbf{u}$.
- d) $A\mathbf{e}_j = \mathbf{a}_j$, for $j = 1, 2, \dots, n$ where \mathbf{e}_j is the j th standard vector in \mathbb{R}^n .
- e) If B is an $m \times n$ matrix such that $B\mathbf{w} = A\mathbf{w} \quad \forall \mathbf{w} \in \mathbb{R}^n$, then $B = A$.
- f) $A\mathbf{0}$ is the $m \times 1$ zero-vector.
- g) If O is the $m \times n$ zero-matrix, then $O\mathbf{v}$ is the $m \times 1$ zero-vector.
- h) $I_n\mathbf{v} = \mathbf{v}$.

[Spence E., 2008, p. 24]

Proof 3.6

Let A and B be $m \times n$ matrices and let \mathbf{u} and \mathbf{v} be vectors in \mathbb{R}^n .

- a) By Definition 3.13 the matrix-vector product is given as

$$\begin{aligned} A(\mathbf{u} + \mathbf{v}) &= (u_1 + v_1)\mathbf{a}_1 + (u_2 + v_2)\mathbf{a}_2 + \cdots + (u_n + v_n)\mathbf{a}_n \\ &= (u_1 + v_1)a_{1i} + (u_2 + v_2)a_{2i} + \cdots + (u_n + v_n)a_{ni} \quad i = 1, 2, \dots, m, \end{aligned}$$

which is simply addition and multiplication of scalars. This is equal to the sum of the matrix-vector product for each vector:

$$\begin{aligned} A\mathbf{u} + A\mathbf{v} &= (u_1\mathbf{a}_1 + u_2\mathbf{a}_2 + \cdots + u_n\mathbf{a}_n) + (v_1\mathbf{a}_1 + v_2\mathbf{a}_2 + \cdots + v_n\mathbf{a}_n) \\ &= (u_1a_{1i} + u_2a_{2i} + \cdots + u_na_{ni}) + (v_1a_{1i} + v_2a_{2i} + \cdots + v_na_{ni}) \quad i = 1, 2, \dots, m. \end{aligned}$$

This applies for scalars, thus completing the proof.

- b) First $A(c\mathbf{u})$ is reduced to a sum of scalar products:

$$\begin{aligned} A(c\mathbf{u}) &= (cu_1)\mathbf{a}_1 + (cu_2)\mathbf{a}_2 + \cdots + (cu_n)\mathbf{a}_n \\ &= (cu_1)a_{1i} + (cu_2)a_{2i} + \cdots + (cu_n)a_{ni} \quad i = 1, 2, \dots, m \\ &= cu_1a_{1i} + cu_2a_{2i} + \cdots + cu_na_{ni} \quad i = 1, 2, \dots, m. \end{aligned}$$

The other two equations are similarly reduced to scalars:

$$\begin{aligned} c(A\mathbf{u}) &= c(u_1\mathbf{a}_1 + u_2\mathbf{a}_2 + \cdots + u_n\mathbf{a}_n) \\ &= c(u_1a_{1i} + u_2a_{2i} + \cdots + u_na_{ni}) \quad i = 1, 2, \dots, m \\ &= cu_1a_{1i} + cu_2a_{2i} + \cdots + cu_na_{ni} \quad i = 1, 2, \dots, m, \end{aligned}$$

and

$$\begin{aligned}
 (cA)\mathbf{u} &= u_1(c\mathbf{a}_1) + u_2(c\mathbf{a}_2) + \cdots + u_n(c\mathbf{a}_n) \\
 &= u_1(ca_{1i}) + u_2(ca_{2i}) + \cdots + u_n(ca_{ni}) \quad i = 1, 2, \dots, m \\
 &= u_1ca_{1i} + u_2ca_{2i} + \cdots + u_nca_{ni} \quad i = 1, 2, \dots, m.
 \end{aligned}$$

The three expressions can be reduced to the same sum of scalar multiplications, consisting of the same scalars thus completing the proof.

- c) $(A + B)\mathbf{u}$ can be expressed as an entrywise sum multiplied by the components of \mathbf{u} :

$$(a_{ij} + b_{ij})\mathbf{u} = (a_{1i} + b_{1i})u_1 + (a_{2i} + b_{2i})u_2 + \cdots + (a_{ni} + b_{ni})u_n, \quad i = 1, 2, \dots, m.$$

Similarly the two matrix-vector products are reduced to scalar calculation:

$$\begin{aligned}
 A\mathbf{u} + B\mathbf{u} &= (u_1a_{1i} + u_2a_{2i} + \cdots + u_na_{ni}) + (u_1b_{1i} + u_2b_{2i} + \cdots + u_nb_{ni}) \quad i = 1, 2, \dots, m \\
 &= (a_{1i} + b_{1i})u_1 + (a_{2i} + b_{2i})u_2 + \cdots + (a_{ni} + b_{ni})u_n, \quad i = 1, 2, \dots, m,
 \end{aligned}$$

completing the proof as the equations are equal and consists of scalars.

- d) Let \mathbf{e}_j be the j th standard vector, the matrix vector product $A\mathbf{e}_j$ is then

$$\begin{aligned}
 A\mathbf{e}_j &= \mathbf{a}_10 + \mathbf{a}_20 + \cdots + \mathbf{a}_j1 + \cdots + \mathbf{a}_n0 \\
 &= \mathbf{a}_j,
 \end{aligned}$$

thus completing the proof.

- e) First the matrix-vector products are expressed:

$$\begin{aligned}
 B\mathbf{w} &= A\mathbf{w} \\
 \mathbf{b}_1w_1 + \mathbf{b}_2w_2 + \cdots + \mathbf{b}_nw_n &= \mathbf{a}_1w_1 + \mathbf{a}_2w_2 + \cdots + \mathbf{a}_nw_n.
 \end{aligned}$$

To satisfy the equation $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n$ must be identical to $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$. Thus $A = B$, thereby completing the proof.

- f) The matrix-vector product $A\mathbf{0}$, where $\mathbf{0} \in \mathbb{R}^n$ is given by

$$A\mathbf{0} = \mathbf{a}_10 + \mathbf{a}_20 + \cdots + \mathbf{a}_n0 = \mathbf{0} \in \mathbb{R}^m,$$

completing the proof.

- g) Let O denote the $m \times n$ matrix consisting only of zeroes, the matrix-vector product $O\mathbf{v}$ is then

$$O\mathbf{v} = \mathbf{0}v_1 + \mathbf{0}v_2 + \cdots + \mathbf{0}v_n = \mathbf{0} \in \mathbb{R}^m.$$

The vector $\mathbf{0} \in \mathbb{R}^m$ is the $m \times 1$ zero vector, thereby completing the proof.

- h) By Definition 3.10 I_n is defined as the matrix of standard vectors, and the matrix-vector product $I_n\mathbf{v}$ is then

$$I_n\mathbf{v} = \mathbf{e}_1v_1 + \mathbf{e}_2v_2 + \cdots + \mathbf{e}_nv_n = \mathbf{v},$$

thus completing the proof. ■

3.5 Matrix Multiplication

Let A be an $m \times n$ matrix and B be an $n \times p$ matrix. The matrix-vector product $B\mathbf{v}$ for any $p \times 1$ vector \mathbf{v} is an $n \times 1$ vector. Thus $A(B\mathbf{v})$ is an $m \times 1$ vector. By Definition 3.13 and Theorem 3.6

$$\begin{aligned} A(B\mathbf{v}) &= A(v_1\mathbf{b}_1 + v_2\mathbf{b}_2 + \cdots + v_p\mathbf{b}_p) \\ &= A(v_1\mathbf{b}_1) + A(v_2\mathbf{b}_2) + \cdots + A(v_p\mathbf{b}_p) \\ &= v_1A\mathbf{b}_1 + v_2A\mathbf{b}_2 + \cdots + v_pA\mathbf{b}_p \\ &= [A\mathbf{b}_1, A\mathbf{b}_2, \dots, A\mathbf{b}_p] \mathbf{v}. \end{aligned}$$

Let C be the $m \times p$ matrix $[A\mathbf{b}_1, A\mathbf{b}_2, \dots, A\mathbf{b}_p]$. Then $A(B\mathbf{v}) = C\mathbf{v}$. Thus, by Theorem 3.6(e), C is the only matrix with this property. This leads to the following definition.

Definition 3.14 (Matrix Product)

Let A be an $m \times n$ matrix and B be an $n \times p$ matrix. Then the matrix product AB is the $m \times p$ matrix C whose j th column is $A\mathbf{b}_j$.

$$C = [A\mathbf{b}_1, A\mathbf{b}_2, \dots, A\mathbf{b}_p].$$

[Spence E., 2008, p. 97]

The associative law is a property which assures, that the order of operations is indifferent.

Theorem 3.7 (Associativity)

For any $m \times n$ matrix A , any $n \times p$ matrix B , and any $p \times 1$ vector \mathbf{v} ,

$$(AB)\mathbf{v} = A(B\mathbf{v}).$$

[Spence E., 2008, p. 97]

Proof 3.7

By Definition 3.13 and Theorem 3.6

$$\begin{aligned} A(B\mathbf{v}) &= A(v_1\mathbf{b}_1 + v_2\mathbf{b}_2 + \cdots + v_p\mathbf{b}_p) \\ &= A(v_1\mathbf{b}_1) + A(v_2\mathbf{b}_2) + \cdots + A(v_p\mathbf{b}_p) \\ &= v_1A\mathbf{b}_1 + v_2A\mathbf{b}_2 + \cdots + v_pA\mathbf{b}_p \\ &= [A\mathbf{b}_1, A\mathbf{b}_2, \dots, A\mathbf{b}_p] \mathbf{v}. \end{aligned}$$

Let C be the $m \times p$ matrix $[A\mathbf{b}_1, A\mathbf{b}_2, \dots, A\mathbf{b}_p]$. Then $A(B\mathbf{v}) = C\mathbf{v}$. Thus, by Theorem 3.6(e), C is the only matrix with this property.

By Definition 3.14

$$(AB)\mathbf{v} = C\mathbf{v},$$

thus completing the proof. ■

It can be proven that the associative property also applies for matrix multiplication.

Theorem 3.8 (Associativity for Matrix Multiplication)

Let A be a $k \times m$ matrix, C be an $m \times n$, and P be an $n \times p$ matrix. Then the following statement is true

$$A(CP) = (AC)P.$$

[Spence E., 2008, p. 100-101]

Proof 3.8

$A(CP)$ and $(AC)P$ are $k \times p$ matrices. Let \mathbf{u}_j denote column j of CP . Since $\mathbf{u}_j = C\mathbf{p}_j$, column j of $A(CP)$ is $A\mathbf{u}_j = A(C\mathbf{p}_j)$. Furthermore, column j of $(AC)P$ is $(AC)\mathbf{p}_j = A(C\mathbf{p}_j)$ by Theorem 3.7. It follows that the corresponding columns of $A(CP)$ and $(AC)P$ are equal. Thereby completing the proof. ■

3.6 Linear and Matrix Transformations

Before defining linear and matrix transformations, it is necessary to define a function f that maps from \mathbb{R}^n to \mathbb{R}^m denoted by $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$.

Definition 3.15 (Mapping Function)

Let S_1 and S_2 be subsets of \mathbb{R}^n and \mathbb{R}^m , respectively. A function f from S_1 to S_2 , written $f : S_1 \rightarrow S_2$, is a rule that assigns to each vector \mathbf{v} in S_1 a unique vector $f(\mathbf{v})$ in S_2 . The vector $f(\mathbf{v})$ is called the image of \mathbf{v} (under f). The set S_1 is called the domain of a function f , and the set S_2 is called the codomain of f . The range of f is defined to be the set of images $f(\mathbf{v})$ for all \mathbf{v} in S_1 .

[Spence E., 2008, p. 167]

A matrix transformation such as the matrix-vector product defined in Definition 3.13 satisfies the conditions of Definition 3.15.

Definition 3.16 (Matrix Transformation)

Let A be an $m \times n$ matrix. The mapping function $T_A : \mathbb{R}^n \rightarrow \mathbb{R}^m$ defined by $T_A(\mathbf{x}) = A\mathbf{x}$ for all \mathbf{x} in \mathbb{R}^n is called the matrix transformation induced by A .

[Spence E., 2008, p. 168]

For a function to be linear two conditions must be satisfied. The linearity properties: additivity and homogeneity both apply to matrix transformations, and the mapping from one subset to another is regarded as a transformation.

Theorem 3.9 (Linearity of Transformation)

For any $m \times n$ matrix A and any vectors \mathbf{u} and \mathbf{v} in \mathbb{R}^n , the following statements are true:

a) $T_A(\mathbf{u} + \mathbf{v}) = T_A(\mathbf{u}) + T_A(\mathbf{v}).$ (Additivity)

b) $T_A(c\mathbf{u}) = cT_A(\mathbf{u}), \quad \forall c \in \mathbb{R}.$ (Homogeneity)

[Spence E., 2008, p. 170]

Proof 3.9

By Definition 3.16 the transformation $T_A : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is induced by the matrix A , and the mapping is a matrix-vector product.

- a) The additivity property is equivalent to Theorem 3.6(a) and is proven in Proof 3.6.
- b) The homogeneity property is equivalent to Theorem 3.6(b) and is proven in Proof 3.6.

■

Definition 3.17 (Linear Transformation)

A function $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is called a linear transformation if, for all vectors \mathbf{u} and \mathbf{v} in \mathbb{R}^n and all scalars $c \in \mathbb{R}$, both of the following conditions hold:

1. $T(\mathbf{u} + \mathbf{v}) = T\mathbf{u} + T\mathbf{v}.$ (T preserves vector addition.)

2. $T(c\mathbf{u}) = cT(\mathbf{u}).$ (T preserves scalar multiplication.)

[Spence E., 2008, p. 171]

3.7 Systems of Linear Equations

A linear equation is an equation which can be written on the form

$$a_1x_1 + a_2x_2 + \dots + a_nx_n = b,$$

where a_1, a_2, \dots, a_n and $b \in \mathbb{R}$, and x_1, x_2, \dots, x_n are the variables. The scalars a_1, a_2, \dots, a_n , are called the coefficients of the equation and the scalar b is called the constant term.

A system of linear equations is a series consisting of m linear equations with n variables. A system of linear equations can be written on the form

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1$$

$$a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2$$

⋮

$$a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m.$$

The solution to the system is a vector \mathbf{s} in \mathbb{R}^n , which satisfies every equation in the system when each variable x_j is respectively substituted by the corresponding vector component s_j . Every system of linear equations has one of three solutions: no solution, exactly one solution, or infinitely many solutions. If the system has one or more solutions it is called consistent. Contrarily, if the system has no solution it is considered inconsistent.

To solve a system of linear equations or determine whether it is inconsistent, a simpler version of the system is advantageous. Such a system has the exact same solutions and it is therefore called equivalent.

When creating a simpler equivalent version of a system of linear equations it is easier to express it as a matrix. This is done by expressing the system of linear equations as a matrix equation $A\mathbf{x} = \mathbf{b}$:

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ a_{21} & \dots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

The augmented matrix $[A | \mathbf{b}]$ is the favourable representation of the system of linear equations when simplifying the system to find a solution:

$$[A | \mathbf{b}] = \left[\begin{array}{ccc|c} a_{11} & \dots & a_{1n} & b_1 \\ a_{21} & \dots & a_{2n} & b_2 \\ \vdots & \ddots & \vdots & \vdots \\ a_{m1} & \dots & a_{mn} & b_n \end{array} \right]$$

3.7.1 Reduced Row Echelon Form

Certain operations are applicable to the augmented matrix when simplifying. These operations are referred to as the elementary row operations.

Definition 3.18 (Elementary Row Operations)

Any of the following three operations are called an elementary row operation when performed on a matrix:

1. **Interchange operation:** Interchange any two rows of the matrix.
2. **Scaling operation:** Multiply every entry of a row by the same nonzero scalar.
3. **Row addition operation:** Add a multiple of one row to another row in the matrix.

[Spence E., 2008, p. 32]

An equivalent form of the augmented matrix is the reduced row echelon form obtained by applying elementary row operations to the augmented matrix. Note that the leftmost nonzero entry of a nonzero row is called a leading entry.

Definition 3.19 (Reduced Row Echelon Form)

For a matrix to be on row echelon form the following conditions has to be satisfied:

1. Any nonzero row has to lie above every zero row
2. The leading entry of a nonzero row must lie in a column to the right of a column containing a leading entry in any preceding row
3. If a column contains the leading entry of some row, then all entries below the leading entry in the same column has to be 0

If the following conditions are also satisfied the matrix is on reduced row echelon form:

1. If a column contains a leading entry of some row, then all other entries in the column must be 0
2. The leading entry of each nonzero row has to be 1

[Spence E., 2008, p. 33]

3.8 Invertibility

A desired property for linear transformations is invertibility. For a linear transformation T_A to be invertible the matrix A inducing the transformation must be invertible. The matrix is invertible if it satisfies certain conditions.

Definition 3.20 (Invertible Matrix)

An $n \times n$ matrix A is called invertible if there exist an $n \times n$ matrix B such that

$$AB = BA = I_n.$$

Under these circumstances B is the inverse of A and contrariwise.

[Spence E., 2008, p. 122]

The inverse of a matrix A is denoted by A^{-1} .

Theorem 3.10 (Properties of Inverse Matrices)

Let A and B be $n \times n$ matrices.

- a) If A is invertible, then A^{-1} is invertible and $(A^{-1})^{-1} = A$.
- b) If A and B are invertible, then AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$.
- c) If A is invertible, then A^\top is invertible and $(A^\top)^{-1} = (A^{-1})^\top$.

[Spence E., 2008, p. 125]

Proof 3.10

a) By Definition 3.20

$$\begin{aligned} AA^{-1} &= I_n \\ A^{-1}(A^{-1})^{-1} &= I_n, \end{aligned}$$

thus $(A^{-1})^{-1} = A$, thereby completing the proof.

b) Let A and B be invertible matrices. Then by Theorem 3.8

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AI_nA^{-1} = AA^{-1} = I_n.$$

Likewise $(B^{-1}A^{-1})(AB) = I_n$, proving that AB satisfies that $(AB)^{-1} = (B^{-1}A^{-1})$.

c) Let A be invertible so that $A^{-1}A = I_n$. By Theorem 3.2(a) $(AB)^\top = A^\top B^\top$ it is obtained that:

$$(A^{-1})^\top A^\top = (A^{-1}A)^\top = I_n^\top = I_n$$

proving that A^\top is invertible, and $(A^{-1})^\top = (A^\top)^{-1}$.

■

The elementary row operations that reduce a matrix to reduced row echelon (RREF) form can be expressed by a transformation matrix.

Theorem 3.11 (RREF by an Invertible Matrix)

Let A be an $m \times n$ matrix with reduced row echelon form R . Then there exists an invertible $m \times m$ matrix P such that

$$PA = R.$$

[Spence E., 2008, p. 127]

Proof 3.11

The matrix equation $A\mathbf{x} = \mathbf{b}$ has the same solutions as $R\mathbf{x} = \mathbf{c}$ if $[R \mid \mathbf{c}]$ is the reduced row echelon form of the augmented matrix $[A \mid \mathbf{b}]$. In that case, there exists an invertible matrix P such that

$$P[A \mid \mathbf{b}] = [R \mid \mathbf{c}].$$

Hence $PA = R$, $P\mathbf{b} = \mathbf{c}$, and P is invertible it follows that $A = P^{-1}R$ and $\mathbf{b} = P^{-1}\mathbf{c}$. Suppose \mathbf{v} is a solution of $A\mathbf{x} = \mathbf{b}$. Then $A\mathbf{v} = \mathbf{b}$, so

$$R\mathbf{v} = (PA)\mathbf{v} = P(A\mathbf{v}) = P\mathbf{b} = \mathbf{c}.$$

Therefore \mathbf{v} is a solution of $R\mathbf{x} = \mathbf{c}$. Conversely, suppose that \mathbf{v} is a solution of $R\mathbf{x} = \mathbf{c}$. Then $R\mathbf{v} = \mathbf{c}$, such that

$$A\mathbf{v} = (P^{-1}R)\mathbf{v} = P^{-1}(R\mathbf{v}) = P^{-1}\mathbf{c} = \mathbf{b}.$$

Therefore \mathbf{v} is a solution of $R\mathbf{x} = \mathbf{c}$. Thus the equations $A\mathbf{x} = \mathbf{b}$ and $R\mathbf{x} = \mathbf{c}$ have the same solutions completing the proof. ■

The reduced row echelon form of a matrix can be used to guarantee that the inverse of the matrix exists.

Theorem 3.12 (Sufficient Requirement for Invertibility)

Let A an $n \times n$ matrix. Then A is invertible if the reduced row echelon form of A is I_n .
[Spence E., 2008, p. 135]

Proof 3.12

Suppose that the reduced row echelon form of A equals I_n . Then, as a result of Theorem 3.11, there exists an invertible $n \times n$ matrix P such that $PA = I_n$, so

$$A = I_n A = (P^{-1}P)A = P^{-1}(PA) = P^{-1}I_n = P^{-1}.$$

By Theorem 3.10(a), P^{-1} is an invertible matrix, and therefore A is invertible. ■

As a result of Theorem 3.12 it is possible to provide a method for determining whether a matrix is invertible or not. If so, the method can compute the inverse matrix.

By Theorem 3.11, it is known that any $n \times n$ matrix A can be converted into a matrix R in reduced row echelon form. Applying the same operations to the augmented $n \times 2n$ matrix $[A \mid I_n]$ gives an $n \times 2n$ matrix $[R \mid B]$ for some $n \times n$ matrix B . Thus, there is an invertible matrix P such that

$$[R \mid B] = P [A \mid I_n] = [PA \mid PI_n] = [PA \mid P].$$

By Theorem 3.12, if $R \neq I_n$, then A is not invertible. If $R = I_n$, then A is invertible. Furthermore, by Definition 3.20 $PA = I_n$ and $P = B$, hence $B = A^{-1}$.

3.9 Orthogonality and Subspaces

A method for determining the inverse of a matrix is presented in the previous section, however this method is not necessary when determining the inverse of orthogonal matrices.

Definition 3.21 (Orthogonal Matrix)

Let A be an $m \times n$ matrix, and S be the set of column vectors of A . Matrix A is said to be orthogonal if and only if the following conditions hold true:

- 1. $\mathbf{u} \cdot \mathbf{v} = 0 \quad \forall \mathbf{u}, \mathbf{v} \in S$.
- 2. $\|\mathbf{u}\| = \|\mathbf{v}\| = 1 \quad \forall \mathbf{u}, \mathbf{v} \in S$.

Some properties of orthogonal matrices is listed in Theorem 3.13.

Theorem 3.13 (Properties of Orthogonal Matrices)

The following conditions for an $n \times n$ matrix A are equivalent

- a) A is orthogonal
- b) $A^\top A = I_n$
- c) A is invertible with $A^\top = A^{-1}$
- d) $A\mathbf{u} \cdot A\mathbf{v} = \mathbf{u} \cdot \mathbf{v}$ for any \mathbf{u} and \mathbf{v} in \mathbb{R}^n
- e) $\|A\mathbf{u}\| = \|\mathbf{u}\|$ for any $\mathbf{u} \in \mathbb{R}^n$

[Spence E., 2008, p. 413]

Proof 3.13

Let A be an $n \times n$ orthogonal matrix. Notice that the entry (i,j) in $A^\top A$ is the dotproduct of \mathbf{a}_i and \mathbf{a}_j . As all columns in A are orthonormal it follows that if $j = i$ then $\mathbf{a}_i \cdot \mathbf{a}_j = 1$ and if $j \neq i$ then $\mathbf{a}_i \cdot \mathbf{a}_j = 0$. Notice that $A^\top A = I_n$ this proves that (a) implies (b). Suppose that $A^\top A = I_n$, then A is invertible and $A^\top = A^{-1}$ this proves that (b) implies (c). To prove that (c) implies (d) suppose that (c) holds. For any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$,

$$A\mathbf{u} \cdot A\mathbf{v} = \mathbf{u} \cdot A^\top A\mathbf{v} = \mathbf{u} \cdot \mathbf{v}.$$

To show that (d) implies (e), assume that (d) holds. Then for any $\mathbf{u} \in \mathbb{R}^n$,

$$\|A\mathbf{u}\| = \sqrt{A\mathbf{u} \cdot A\mathbf{u}} = \sqrt{\mathbf{u} \cdot \mathbf{u}} = \|\mathbf{u}\|.$$

Since

$$\|\mathbf{u}_j\| = \|A\mathbf{e}_j\| = \|\mathbf{e}_j\| = 1,$$

and

$$\|\mathbf{u}_i + \mathbf{u}_j\|^2 = \|A(\mathbf{e}_i + \mathbf{e}_j)\|^2 = \|\mathbf{u}_i\|^2 + \|\mathbf{u}_j\|^2,$$

(e) implies (a). Thereby completing the proof. ■

A set of vectors in \mathbb{R}^n such as the column vectors of a matrix, span a space in \mathbb{R}^n and certain subsets of \mathbb{R}^n have desirable properties. These subsets will be defined in Definition 3.22.

Definition 3.22 (Subspaces)

A set of vectors S in \mathbb{R}^n is called a subspace of \mathbb{R}^n if and only if the following conditions holds true:

1. $\mathbf{0} \in S$.
2. $(\mathbf{u} + \mathbf{v}) \in S$ for any two vectors $\mathbf{u}, \mathbf{v} \in S$.
3. $c\mathbf{u} \in S$ for any vector $\mathbf{u} \in S$ and $c \in \mathbb{R}$.

[Spence E., 2008, p. 227]

Subspaces have relation to the span of vectors.

Theorem 3.14 (Span and Subspace)

The span of a finite nonempty subset of \mathbb{R}^n is a subspace in \mathbb{R}^n .

[Spence E., 2008, p. 231]

Proof 3.14

Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ where $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n \in \mathbb{R}^n$. It is noticed that

$$0\mathbf{v}_1 + 0\mathbf{v}_2 + \dots + 0\mathbf{v}_n = \mathbf{0},$$

in accordance with Definition 3.11, and therefore $\mathbf{0}$ is in the span of S .

Let a_1, a_2, \dots, a_n and b_1, b_2, \dots, b_n be scalars, and $\mathbf{u}, \mathbf{w} \in \text{Span}\{S\}$ meaning that:

$$\mathbf{u} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \dots + a_n\mathbf{v}_n, \quad \mathbf{w} = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \dots + b_n\mathbf{v}_n.$$

The sum of these two vectors can be written as

$$\mathbf{u} + \mathbf{w} = (a_1 + b_1)\mathbf{v}_1 + (a_2 + b_2)\mathbf{v}_2 + \dots + (a_n + b_n)\mathbf{v}_n.$$

Notice that this is another linear combination (see Definition 3.8) of the vectors in S thereby placing $\mathbf{u} + \mathbf{w}$ in the span of S .

For any scalar c , the following applies:

$$\begin{aligned} c\mathbf{u} &= c(a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \dots + a_n\mathbf{v}_n) \\ &= (ca_1)\mathbf{v}_1 + (ca_2)\mathbf{v}_2 + \dots + (ca_n)\mathbf{v}_n. \end{aligned}$$

Again this is a linear combination of the vectors in S , placing $c\mathbf{u}$ in the span of S . Notice that the three conditions from Definition 3.22 are satisfied, thereby completing the proof. ■

By the definition of orthogonality of vectors, orthogonality of subspaces which is a desirable property, can be defined.

Definition 3.23 (Orthogonality of Subspaces)

Let S_1 and S_2 be subspaces of \mathbb{R}^n . The subspaces S_1 and S_2 are said to be orthogonal if and only if

$$\mathbf{v} \cdot \mathbf{u} = 0,$$

for all $\mathbf{v} \in S_1$ and $\mathbf{u} \in S_2$.

The vectors of a subspace can be expressed by a set of linearly independent vectors spanning that subspace.

Definition 3.24 (Basis)

Let S be a nonzero subspace of \mathbb{R}^n . A basis for S is a linearly independent generating set for S .

[Spence E., 2008, p. 241]

Notice that \mathbb{R}^n has an infinite number of bases. The term standard basis for \mathbb{R}^n refers to the set of vectors $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$.

Definition 3.25 (Vector Representation in Terms of Basis)

Let $\{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$ be a basis \mathcal{B} for the subspace \mathbb{R}^n . Any vector \mathbf{u} in \mathbb{R}^n can be written uniquely as

$$\mathbf{u} = c_1 \mathbf{b}_1 + c_2 \mathbf{b}_2 + \cdots + c_n \mathbf{b}_n.$$

Where c_1, c_2, \dots, c_n are called the coordinates of \mathbf{u} in terms of the basis \mathcal{B} . This is denoted

$$[\mathbf{u}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}.$$

[Spence E., 2008, p. 265]

Normally a vector \mathbf{u} is represented in terms of the standard basis E which is denoted $[\mathbf{u}]_E$. As such a conversion can be done that translates a vector \mathbf{u} from the standard basis to any basis \mathcal{B} and contrariwise.

Theorem 3.15 (Basis Translation)

Let \mathcal{B} be a basis for \mathbb{R}^n , $P_{\mathcal{B}} = [\mathbf{b}_1 \ \mathbf{b}_2 \ \dots \ \mathbf{b}_n]$ be the matrix containing the basis vectors of \mathcal{B} , $\mathbf{u} \in \mathbb{R}^n$ be a vector in standard basis, and $[\mathbf{u}]_{\mathcal{B}}$ be \mathbf{u} in terms of \mathcal{B} , then

- a) $\mathbf{u} = P_{\mathcal{B}}[\mathbf{u}]_{\mathcal{B}}$
- b) $[\mathbf{u}]_{\mathcal{B}} = P_{\mathcal{B}}^{-1}\mathbf{u}$.

[Spence E., 2008, p. 267]

Proof 3.15

Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$ be a basis for \mathbb{R}^n , and \mathbf{u} be a vector in \mathbb{R}^n . If

$$[\mathbf{u}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix},$$

then following Definition 3.25

$$\begin{aligned} \mathbf{u} &= c_1 \mathbf{b}_1 + c_2 \mathbf{b}_2 + \cdots + c_n \mathbf{b}_n \\ &= [\mathbf{b}_1 \quad \mathbf{b}_2 \quad \dots \quad \mathbf{b}_n] \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} \\ &= P_{\mathcal{B}}[\mathbf{u}]_{\mathcal{B}}. \end{aligned}$$

$P_{\mathcal{B}}$ is the matrix containing the vectors in \mathcal{B} . Since \mathcal{B} is a basis, the vectors in $P_{\mathcal{B}}$ are linearly independent in accordance with Definition 3.24. Hence $P_{\mathcal{B}}$ is invertible following Theorem 3.12, thus

$$\mathbf{u} = P_{\mathcal{B}}[\mathbf{u}]_{\mathcal{B}} \Leftrightarrow [\mathbf{u}]_{\mathcal{B}} = P_{\mathcal{B}}^{-1}\mathbf{u},$$

thereby completing the proof. ■

Vectorspaces are beyond the scope of this project, but to present a definition for direct sums, which is used in Chapter 7, the sum of vectorspaces must be defined.

Definition 3.26 (Sum of Vectorspaces)

Let $X, Y \subseteq V$. Then the sum of X and Y is:

$$X + Y = \{\mathbf{x} + \mathbf{y} \mid \mathbf{x} \in X, \mathbf{y} \in Y\}.$$

[Vujicic, 2008, p. 139]

Definition 3.27 (Direct Sum)

Let W_1 and W_2 be subspaces of V . With $W_1 \cap W_2 = \{0\}$. If

$$V = W_1 + W_2$$

V is called the direct sum of W_1 and W_2 denoted

$$V = W_1 \oplus W_2.$$

[Vujicic, 2008, p. 139]

Notice that V is the direct sum of W_1 and W_2 if and only if all elements in V can be expressed uniquely as $\mathbf{w}_1 + \mathbf{w}_2 | \mathbf{w}_1 \in W_1, \mathbf{w}_2 \in W_2$. [Vujicic, 2008, p. 139]

4 | Signals and Sampling

4.1 Introduction

Any physical system emits information that can be represented as signals which convey the information of said systems. This information could be the behaviour of a physical system or the characteristics of a certain phenomenon. Information can be recorded using different sensors depending on the type of information. Among the most common signals are vibrations, electrical, and electromagnetic signals. This project uses microphones to measure continuous sound signals. These measurements are stored as discrete values thus becoming a discrete signal. Digital signals can be converted to analogue signals, which for sound is accomplished by speakers that convert electrical signals to acoustic signals.

4.2 Signals

A signal may be continuous or discrete. If a signal is discrete it can be represented as a sequence. A sequence of numbers \mathbf{x} , where the n th value in the sequence is denoted $\mathbf{x}[n]$ is formally written as

$$\mathbf{x} = \{\mathbf{x}[n]\} \quad -\infty < n < \infty,$$

where $n \in \mathbb{Z}$. A sequence is typically a result of periodic measurements of a continuous signal $x_c(t)$.

$$\mathbf{x}[n] = x_c(nT) \quad -\infty < n < \infty.$$

Notice that T is the period of the measurements and the reciprocal value is its frequency. [Oppenheim, 1989, p. 10]

Definition 4.1 (Continuous Signal)

A continuous-time signal $x_c(t)$ is any piecewise continuous function

$$x_c(t) : \mathbb{R}_+ \rightarrow \mathbb{R}^n,$$

defined on a closed interval, that carries information of a behaviour in time of a measurable physical size.
[Rabbath, 2013, p. 13]

Definition 4.2 (Periodic Continuous Signal)

A continuous-time signal $x_c(t)$ is said to be periodic with period $T > 0$ if

$$x_c(t) = x_c(t + T), \quad t \in \mathbb{R}.$$

[Beerends, 2003, p.12]

Definition 4.3 (Periodic Discrete Signal)

A discrete-time signal $\mathbf{x}[n]$ is said to be periodic with period $N \in \mathbb{N}$ if

$$\mathbf{x}[n] = \mathbf{x}[n + N], \quad n \in \mathbb{Z}.$$

[Beerends, 2003, p.12]

4.2.1 Sampling

The process of discretising continuous signals is called sampling.

Definition 4.4 (Sampling)

A continuous signal $x_c(t)$ is discretised using the sampling interval T . The discretised signal is then:

$$\mathbf{x}[n] = x_c(nT), \quad T > 0, \quad n \in \mathbb{Z}.$$

[Oppenheim, 1989, p.140]

The sampling frequency is given by $f_s = \frac{1}{T}$. In practice the process of sampling consists of two electrical circuits: a sample/hold circuit and an analogue to digital converter circuit. A block diagram of the sampling process is seen in Figure 4.1.

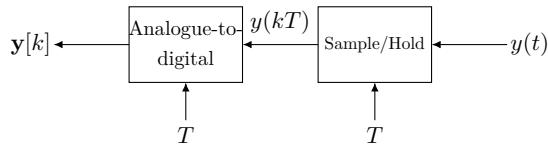


Figure 4.1: Block diagram of the sampling process. The sample/hold block measures the analogue input at every kT , $k \in \mathbb{Z}$. The measurements are then discretised in the analogue-to-digital block and results in the digital representation $\mathbf{y}[k]$.

The external voltage $y(t)$ is an analogue signal. The analogue signal is a result of sensors converting signals to electrical signals. For example, a microphone converts sound signals (vibration) to electrical signals that can be measured as seen in Figure 4.1.

The sample/hold circuit consists of a capacitor and two switches both of which are initially open. At every sampling period kT , $k \in \mathbb{Z}$, the switch connecting the external voltage to the capacitor is closed, thereby charging the capacitor to the signal $y(t)$. Afterwards the switch closes and the other switch opens and the capacitor discharges to the analogue-to-digital converter. The discharge period is forced to be slower than the charging period by circuit manipulation, creating the “hold” effect. The effect of the sample/hold causes the input signal to be measured every sampling period and then the measured value is held for a time period.

The analogue-to-digital converter is more complex and thus a more in-depth block diagram is presented in Figure 4.2.

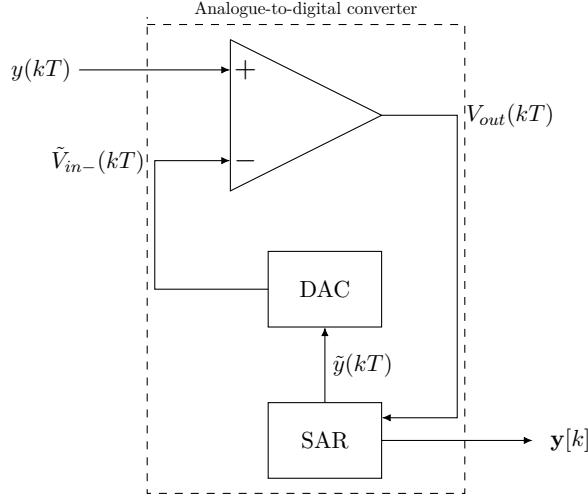


Figure 4.2: Block diagram of the analogue-to-digital circuit.

The measured voltage of the sample/hold circuit $y(kT) = V_{in+}(kT)$ is the positive input of the operational amplifier (comparator), which then outputs a voltage that is temporarily stored as a bit value in the Successive Approximation Register (SAR). The SAR sends the digital representation (the bit value) of the measured value to the digital-to-analogue converter (DAC), which converts the signal to an analogue voltage $\tilde{V}_{in-}(kT)$. The voltage $\tilde{V}_{in-}(kT)$ is then compared to the measured voltage $V_{in+}(kT)$ by the comparator. Depending on the result of the comparison, the bit value of $\tilde{V}_{in-}(kT)$ is either reduced or increased. This process is repeated until the bit representation of $y(kT)$ is as accurate as possible with the given bit depth of the SAR, and then the digitalised measurement is saved as $y[k]$.

When a continuous signal is being sampled, some information can be lost if the signal is not sampled with a sufficiently high sampling frequency. More information can be obtained by decreasing the sampling period T which corresponds to increasing the sampling frequency f_s .

If the sampling frequency is too low a phenomenon called aliasing can occur. Aliasing occurs when sampling with a sufficiently low sampling frequency that causes different signals to be indistinguishable.

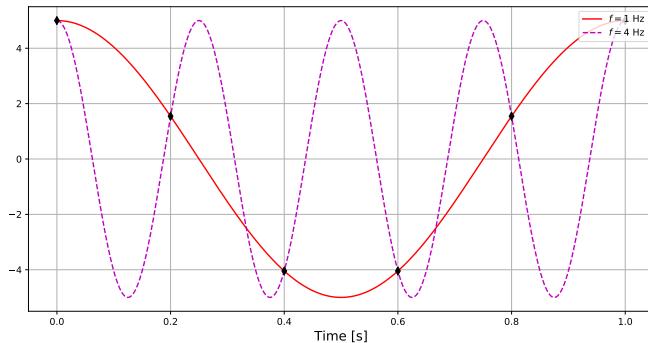


Figure 4.3: Illustration of aliasing. The sampling frequency of the signals is $f_s = 6$ Hz.

In Figure 4.3 an example of aliasing is illustrated. The aliasing is caused by the sampling frequency $f_s = 6$ Hz being too low to uniquely determine the sampled signal (magenta

curve) and thus the sampling process instead presents the signal as a sine wave with a frequency $f = 1$ Hz (red curve).

To avoid aliasing, a sufficiently high sampling frequency must be used. The Nyquist-Shannon theorem describes how this sampling frequency is determined.

Theorem 4.1 (Nyquist-Shannon Sampling Theorem)

Let $x_c(t)$ be a bandlimited signal with Fourier-transform $\mathcal{F}(x_c(t))(\omega) = 0, \forall |\omega| > \omega_N$. Then $x_c(t)$ is uniquely determined by the samples $\mathbf{x}[n] = x(nT)$ if

$$\omega_s = \frac{2\pi}{T} \geq 2\omega_N,$$

where ω_s is the angular sampling frequency and ω_N is the Nyquist-Shannon frequency.

[Oppenheim, 1989, p.146]

A bandlimited signal, is a signal which frequency content is limited to a certain interval. By the Nyquist-Shannon Theorem 4.1 the upper limit of a bandlimited signal must be half the sampling frequency at most to avoid aliasing. The interval of the frequency spectrum is $[a; b]$ where $0 \leq a$ and $b \leq \frac{f_s}{2}$. The bandwidth of such a signal is the width of the frequency spectrum given as $B = b - a$, where b is the upper limit and a is the lower limit of the spectrum.

A method for characterising the content of a signal is the energy of a signal.

Definition 4.5 (Energy of a Finite Sequence)

Let \mathbf{x} be a signal with $\text{card}(\mathbf{x}) = N$, the energy $E_{\mathbf{x}}$ of \mathbf{x} is then defined as

$$E_{\mathbf{x}} = \sum_{k=0}^{N-1} \|\mathbf{x}[k]\|^2.$$

[Dr. Shrivastava, 2017, p.11]

4.3 Sound Sensors

A microphone is a sound sensor that converts sounds from being vibrations in the air into an electrical signal. In Figure 4.4 an illustration of an electret microphone is shown.

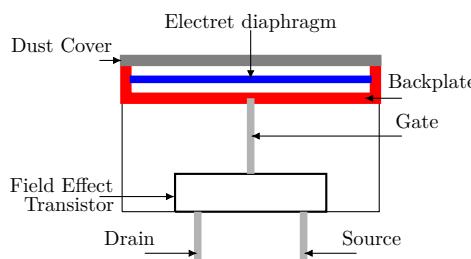


Figure 4.4: Simplification of an electret microphone

This type of microphone converts vibrations to an analogue electrical signal through the use of an electret diaphragm and a backplate. The two components have opposite charge which, due to the small distance between them, act as a capacitor. The capacitance changes when the distance between the components change, as will the voltage between the components. The voltage is received by the field effect transistor.

When the vibrations from a sound signal hits the electret diaphragm it will begin to vibrate correspondingly thus causing a change in the voltage received by the field effect transistor. The field effect transistor then amplifies the received voltage in order to represent the sound signal. The amplified signal emitted by the drain is then received by the sample/hold circuit as $y(t)$, which is seen in Figure 4.1. [Fraden, 2016, p.495]

5 | Convolution

5.1 Introduction

A discrete convolution is a mathematical operator that operates on two sequences and produces a third sequence that resembles how the shape of one sequence is modified by the other. Convolution is an operation used when applying a filter to a signal which is described in Chapter 6. A convolution can be used to correlate two signals called cross-correlation by reverting one of the signals. An application of cross-correlation is to determine the phase shift of two signals.

This project is confined to discrete convolution and cross-correlation as it is only applied to discrete signals. This chapter is based on “Discrete-Time Signal Processing”. [Chitode, 2009]

5.2 Discrete Convolution

Convolution can either be linear or circular. Linear convolution is the simplest convolution, whereas circular convolution is used for periodic signals or if the cardinality of the output must be the same as the input sequences.

Definition 5.1 (Discrete Linear Convolution)

Let a discrete signal $\mathbf{h} \in \mathbb{R}^M$ and a discrete signal $\mathbf{x} \in \mathbb{R}^N$, where $M \leq N$. The discrete convolution of \mathbf{h} and \mathbf{x} is then

$$\mathbf{y}[n] = \sum_{k=0}^{M-1} \mathbf{h}[n] \mathbf{x}[n - k],$$

where \mathbf{y} is the output with $\text{card}(\mathbf{y}) = M + N - 1$.

The convolution operation is denoted $\mathbf{y} = \mathbf{h} * \mathbf{x}$.

[Oppenheim, 1989, p.23]

It might not be advantageous that the cardinality of \mathbf{y} is $M + N - 1$, this can be avoided by using circular convolution. Circular convolution is simplified by ensuring that the input sequences have the same cardinality. The shortest sequence is extended by adding zeros until it has the same cardinality as the other sequence. The output sequence of a circular convolution has the same cardinality as both of the sequences.

Circular convolution of the inputs is given by shifting the sequence \mathbf{x} with $\text{card}(\mathbf{x}) = N$ circularly. The signal \mathbf{x} is constructed as a periodic sequence with period N . The circular shifting of $\mathbf{x}[n]$ shifted by k is then

$$\mathbf{x}[n - k + pN], \quad k \in \mathbb{Z}_+, \quad \text{and} \quad p = \begin{cases} 0, & \text{if } k = 0 \\ 1, & \text{if } k > 0 \end{cases}.$$

Definition 5.2 (Discrete Circular Convolution)

Let \mathbf{x} be a real-valued periodic sequence with period N and \mathbf{h} be a real-valued signal with $\text{card}(\mathbf{h}) = M$. Both sequences must have length N , thus \mathbf{h} is extended by $N - M$ zeros. The circular convolution is then

$$\mathbf{y}[n] = \sum_{k=0}^{N-1} \mathbf{h}[k] \mathbf{x}[n - k + pN] \quad p = \begin{cases} 0, & \text{if } k = 0 \\ 1, & \text{if } k > 0 \end{cases}$$

where \mathbf{y} is the output with $\text{card}(\mathbf{y}) = N$.

The circular convolution operation is denoted $\mathbf{y} = \mathbf{h} \circledast \mathbf{x}$. [Chitode, 2009, p.36]

A numerical example of circular convolution is seen in Example 5.1.

Example 5.1

Let \mathbf{x} and \mathbf{h} be discrete sequences with $\text{card}(\mathbf{x}) = 4$ and $\text{card}(\mathbf{h}) = 2$.

$$\mathbf{h} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 7 \\ 11 \end{bmatrix}.$$

The input sequence \mathbf{h} is extended with zeroes, so $\text{card}(\mathbf{x}) = \text{card}(\mathbf{h})$:

$$\mathbf{h} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 7 \\ 11 \end{bmatrix}.$$

By Definition 5.2 the output components are given by $\mathbf{h} \circledast \mathbf{x}$:

$$\begin{aligned} \mathbf{y}[0] &= \sum_{k=0}^3 \mathbf{h}[k] \mathbf{x}[0 - k + p4] = 1 \cdot 1 + 1 \cdot 11 + 0 \cdot 7 + 0 \cdot 2 = 12 \\ \mathbf{y}[1] &= \sum_{k=0}^3 \mathbf{h}[k] \mathbf{x}[1 - k + p4] = 1 \cdot 2 + 1 \cdot 1 + 0 \cdot 11 + 0 \cdot 7 = 3 \\ \mathbf{y}[2] &= \sum_{k=0}^3 \mathbf{h}[k] \mathbf{x}[2 - k + p4] = 1 \cdot 7 + 1 \cdot 2 + 0 \cdot 1 + 0 \cdot 11 = 9 \\ \mathbf{y}[3] &= \sum_{k=0}^3 \mathbf{h}[k] \mathbf{x}[3 - k + p4] = 1 \cdot 11 + 1 \cdot 7 + 0 \cdot 2 + 0 \cdot 1 = 18. \end{aligned}$$

The output sequence \mathbf{y} is then:

$$\mathbf{y} = \mathbf{h} \circledast \mathbf{x} = \begin{bmatrix} 12 \\ 3 \\ 9 \\ 18 \end{bmatrix}.$$

5.2.1 Cross-Correlation

In signal processing, cross-correlation is a method to compute the similarities of two sequences.

Definition 5.3 (Discrete Cross-Correlation)

Let $\mathbf{x}_1 \in \mathbb{R}^M$ and $\mathbf{x}_2 \in \mathbb{R}^N$ be discrete sequences. The cross-correlation of \mathbf{x}_1 and \mathbf{x}_2 is then

$$\mathbf{z}[n] = \sum_{k=0}^L \mathbf{x}_1[n]\mathbf{x}_2[n+k], \quad L = \min(M, N),$$

where $\mathbf{z}[n]$ is the output with $\text{card}(\mathbf{z}) = M + N - 1$.

The cross-correlation operation is denoted $\mathbf{z} = \mathbf{x}_1 \star \mathbf{x}_2$.

[Sundararajan, 2016, p.2i]

Notice how cross correlation is similar to the discrete linear convolution (Definition 5.1), with the difference being that one of the signals is reverted.

6 | Filters

6.1 Introduction

Signals can be expressed as a sum of high and low frequency components. These components can be distinguished through the use of filters. low-pass filters allow only sufficiently low frequencies to pass as the amplitude of high frequencies are diminished, whereas high-pass filters diminish amplitudes of low frequencies.

In discrete signal processing, filters can be represented mathematically by vectors, where the components are the filter coefficients that determine how the signal is filtered.

This project focuses on time-invariant filters. Such a filter is a time-invariant operator whose output \mathbf{y} is the discrete convolution of an input vector \mathbf{x} and a vector that contains the filter coefficients \mathbf{h} . This project uses discrete circular convolution (Definition 5.2) when applying a filter to a signal. This chapter describes filters and filter banks as matrices. When working with matrices discrete circular convolution is advantageous as the cardinality of output and input is the same.

This chapter is based on the book “Wavelets and Filter Banks”. [Strang, 1996]

6.2 Low-pass and High-pass Filters

To study filters, a simple low-pass and high-pass filter is used. The filters are given respectively by:

$$\begin{aligned}\mathbf{y}[n] &= \frac{1}{2}\mathbf{x}[n] + \frac{1}{2}\mathbf{x}[n - 1] && \text{(low-pass filter)} \\ \mathbf{y}[n] &= \frac{1}{2}\mathbf{x}[n] - \frac{1}{2}\mathbf{x}[n - 1] && \text{(high-pass filter),}\end{aligned}\tag{6.1}$$

where $n \in \mathbb{Z}_+$. This specific low-pass filter is a moving average as the output to time $t = n$ is the average of the current input $\mathbf{x}[n]$ and the previous input $\mathbf{x}[n - 1]$. Thus equation (6.1) can be described as a combination of two operators that operate on the input \mathbf{x} : The identity $\mathbf{x}[n]$ yields output = input, and the delay $\mathbf{x}[n - 1]$ yields the input one sample earlier.

This specific high-pass filter computes moving differences. The output to time $t = n$ is the difference between the identity $\mathbf{x}[n]$ and the delay $\mathbf{x}[n - 1]$.

The low-pass and high-pass filters can each be described by a vector containing the filter coefficients. These vectors are respectively denoted \mathbf{h} and \mathbf{g} . The two filters have similar filter coefficients for the identity $\mathbf{h}[0] = \mathbf{g}[0] = \frac{1}{2}$ whereas the filter coefficients for

the delay have inverted signs $\mathbf{h}[1] = \frac{1}{2}$, $\mathbf{g}[1] = -\frac{1}{2}$:

$$\mathbf{h} = \frac{1}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix} \quad (\text{low-pass filter coefficients}),$$

$$\mathbf{g} = \frac{1}{2} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix} \quad (\text{high-pass filter coefficients}).$$

The output attained by a filter acting on an input signal is given by Definition 5.2. For the low-pass filter this can be constructed as the matrix-vector product $\mathbf{y} = H\mathbf{x}$, where the matrix H is the filter matrix containing the filter coefficients of \mathbf{h} , and \mathbf{x} is the input signal. The matrix-vector product is given by

$$\begin{bmatrix} \mathbf{y}[0] \\ \mathbf{y}[1] \\ \mathbf{y}[2] \\ \mathbf{y}[3] \end{bmatrix} = \begin{bmatrix} \mathbf{h}[0] & 0 & 0 & \mathbf{h}[1] \\ \mathbf{h}[1] & \mathbf{h}[0] & 0 & 0 \\ 0 & \mathbf{h}[1] & \mathbf{h}[0] & 0 \\ 0 & 0 & \mathbf{h}[1] & \mathbf{h}[0] \end{bmatrix} \begin{bmatrix} \mathbf{x}[0] \\ \mathbf{x}[1] \\ \mathbf{x}[2] \\ \mathbf{x}[3] \end{bmatrix} = \begin{bmatrix} \mathbf{x}[0]\mathbf{h}[0] + \mathbf{x}[3]\mathbf{h}[1] \\ \mathbf{x}[0]\mathbf{h}[1] + \mathbf{x}[1]\mathbf{h}[0] \\ \mathbf{x}[1]\mathbf{h}[1] + \mathbf{x}[2]\mathbf{h}[0] \\ \mathbf{x}[2]\mathbf{h}[1] + \mathbf{x}[3]\mathbf{h}[0] \end{bmatrix}.$$

Note that the size of the matrix can be extended to any signal with cardinality N .

6.3 Frequency Response

To derive the frequency response of the filters, the input \mathbf{x} is replaced by

$$\mathbf{x}[n] = e^{in\omega}, \quad \omega \in \mathbb{R}_+,$$

where ω is the angular frequency of the signal.

The frequency response $H(\omega)$ of the low-pass filter (equation (6.1)) describes the scaling of an arbitrary input at different angular frequencies. It is derived by:

$$\begin{aligned} \mathbf{y}[n] &= \mathbf{h}[0]\mathbf{x}[n] + \mathbf{h}[1]\mathbf{x}[n-1] \\ &= \mathbf{h}[0]e^{in\omega} + \mathbf{h}[1]e^{i(n-1)\omega} \\ &= \underbrace{(\mathbf{h}[0] + \mathbf{h}[1]e^{-i\omega})}_{H(\omega)} e^{in\omega}. \end{aligned}$$

For $\omega = 0$ the frequency response is $H(\omega) = 1$. When $\omega \rightarrow 0$ the frequency response $H(\omega) \rightarrow 1$, and the amplitude of the signal is barely affected.

The frequency response of the high-pass filter is

$$G(\omega) = \mathbf{g}[0] + \mathbf{g}[1]e^{-i\omega}.$$

For $\omega \rightarrow \infty$ the high-pass response $G(\omega) \rightarrow \frac{1}{2}$, thus for high angular frequencies amplitudes are scaled by approximately $\frac{1}{2}$.

The frequency response of the low-pass and high-pass filters are plotted in Figure 6.1.

Generally for any filter \mathbf{p} with M coefficients, the frequency response is

$$P(\omega) = \sum_{n=0}^{M-1} \mathbf{p}[n]e^{-in\omega}. \quad (6.2)$$

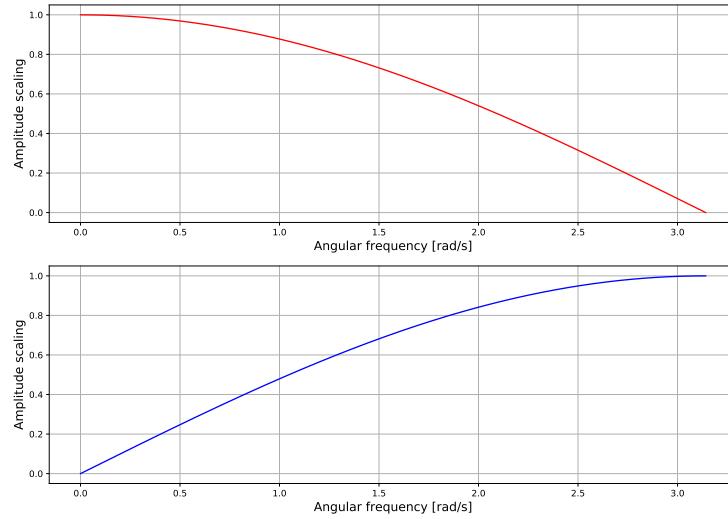


Figure 6.1: The frequency response of the magnitude of the low-pass and high-pass filter. The red curve is $|H(\omega)|$ of the low-pass filter and the blue curve is $|G(\omega)|$ of the high-pass filter.

6.4 Filter Bank

A filter bank is an arrangement of filters. This could for instance be an arrangement of a low-pass and a high-pass filter. A low-pass filter with the frequency response H diminishes the highest frequencies and the high-pass filter with the frequency response G diminishes the lowest frequencies. They separate a signal into frequency bands.

The cardinality of the output signal depends on the amount of filters in the filter bank and the cardinality of the input signal. Since the filtering is computed by circular convolution (Definition 5.2) each filter contributes to an output cardinality corresponding to the cardinality of the input. Thus applying the filter coefficients of \mathbf{h} and \mathbf{g} in the filter bank to an input signal \mathbf{x} with $\text{card}(\mathbf{x}) = N$, the cardinality of the output signal \mathbf{y} is $2N$.

In practice it is desired that $\text{card}(\mathbf{y}) = \text{card}(\mathbf{x})$, and therefore downsampling is applied. Downsampling is an important part of the filter bank, and it is conducted by removing the odd-numbered components from the outputs of the separate filters ($(\downarrow 2)$ means downsampling by 2):

$$(\downarrow 2)\mathbf{y} = \begin{bmatrix} \mathbf{y}[0] \\ \mathbf{y}[2] \\ \mathbf{y}[4] \\ \vdots \\ \mathbf{y}[2n] \end{bmatrix}.$$

As compensation for removing half of the components, normalisation is applied; the remainder $(\downarrow 2)\mathbf{y}$ gets multiplied by a normalising factor $\sqrt{2}$ to keep the energy of \mathbf{y}

constant, so that

$$\mathbf{c} = \sqrt{2}\mathbf{h} = \sqrt{2} \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

$$\mathbf{d} = \sqrt{2}\mathbf{g} = \sqrt{2} \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

When normalisation is applied, the frequency response of the low-pass and the high-pass filter becomes

$$H(\omega)\sqrt{2} = C(\omega),$$

$$G(\omega)\sqrt{2} = D(\omega).$$

The combination of low-pass filtering and downsampling can be done with a single matrix L . Since downsampling removes the odd-numbered components the matrix L is achieved by removing the odd-numbered rows in C , where C is the matrix containing the normalised low-pass filter coefficients of \mathbf{c} .

$$L = (\downarrow 2)C = (\downarrow 2) \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

Similarly the filtering with the high-pass filter while downsampling can be achieved with a single matrix B , where D is the matrix containing the normalised high-pass filter coefficients of \mathbf{d} .

$$B = (\downarrow 2)D = (\downarrow 2) \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \\ 1 & 0 & 0 & 0 & 0 & -1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix}$$

Matrices containing downsampling and normalised filter coefficients are from now on called channels. The low-pass channel L and the high-pass channel B can be presented in one matrix:

$$\begin{bmatrix} L \\ B \end{bmatrix} = \begin{bmatrix} (\downarrow 2)C \\ (\downarrow 2)D \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix}.$$

The matrix containing both the low-pass channel L and high-pass channel B is called the analysis filter bank. The two-channel analysis filter bank can be represented by a block diagram, where the blocks are linear operators. The analysis filter bank creates an output vector containing information from the low-pass and high-pass filters respectively. The output vector is of cardinality N , where N is the cardinality of the input signal.

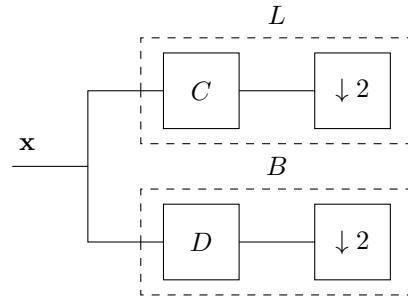


Figure 6.2: Block diagram of the analysis filter bank.

To reconstruct the signal the synthesis filter bank is necessary. The synthesis filter bank is the inverse of the analysis filter bank. The analysis filter bank consists of two steps, filtering and downsampling, and the synthesis filter bank also consists of two steps, upsampling and filtering. However, the order is reversed hence the synthesis bank is the inverse.

To restore the cardinality of the signal upsampling is used. Upsampling is achieved by returning the odd-numbered components as zeroes.

$$(\uparrow 2)\mathbf{y} = \begin{bmatrix} \mathbf{y}[0] \\ 0 \\ \mathbf{y}[2] \\ 0 \\ \vdots \\ \mathbf{y}[2n] \\ 0 \end{bmatrix}.$$

The next step in the synthesis bank is filtering. The synthesis filters are constructed from the inverse of the analysis filters C and D . These matrices are orthogonal as a result of \mathbf{c} and \mathbf{d} being orthonormal. By Theorem 3.13(c) the synthesis filters are found by transposing the analysis filters:

$$[C^{-1}] = [C^\top] = \sqrt{2} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

$$[D^{-1}] = [D^\top] = \sqrt{2} \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 1 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}.$$

By upsampling these filters and combining them, the two-channel synthesis filter bank is

obtained.

$$\begin{bmatrix} F \\ G \end{bmatrix} = \begin{bmatrix} (\uparrow 2)C^{-1} \\ (\uparrow 2)D^{-1} \end{bmatrix} = \begin{bmatrix} L^{-1} \\ B^{-1} \end{bmatrix} = \begin{bmatrix} L^\top \\ B^\top \end{bmatrix} = \sqrt{2} \begin{bmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}.$$

A block diagram representation of the synthesis filter bank is seen in Figure 6.3.

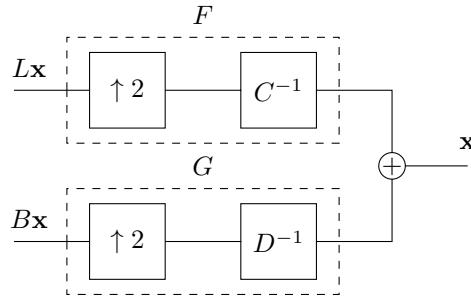


Figure 6.3: Block diagram of the synthesis filter bank.

Note that all of the presented matrices are extended to $N \times N$ matrices for a signal \mathbf{x} with $\text{card}(\mathbf{x}) = N$.

6.5 Half-band Filters

The low-pass and high-pass filters can in practice be treated as half-band filters. The bandwidth of a signal is halved by applying an ideal half-band filter. Consider a signal \mathbf{x} sampled with a sampling frequency f_s . The passband of \mathbf{x} is then $[0; f_N]$ where f_N is the Nyquist-Shannon frequency (Theorem 4.1). Due to the downsampling the Nyquist-Shannon frequency is halved for every application. Applying the low-pass filter to \mathbf{x} results in a filtered signal \mathbf{x}_L which then has the passband $[0; \frac{f_N}{2}]$. The same applies for the high-pass filter, for which the signal \mathbf{x}_H has the passband $[\frac{f_N}{2}; f_N]$. The downsampling causes aliasing as the highest frequency that can be represented is $f = \frac{f_s}{4} = \frac{f_N}{2}$. As a result all frequencies higher than $\frac{f_N}{2}$ are mirrored into the low frequencies $[0; \frac{f_N}{2}]$. Thus the passband for the high-pass filter is regarded as $[f_N; \frac{f_N}{2}]$ since the high frequencies are mirrored into the lowest frequencies. This effect cascades down as the filters are repeatedly applied. An illustration can be found in Figure 6.4.

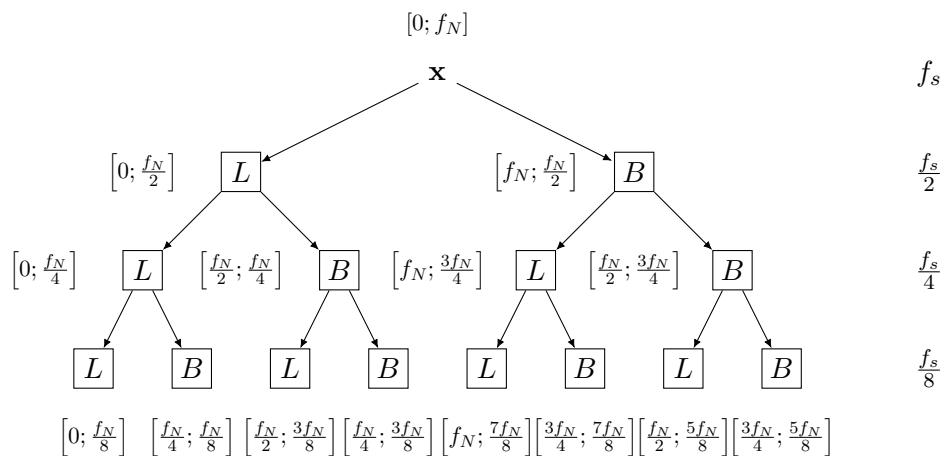


Figure 6.4: The passbands of a signal when half-band filters are applied.

7 | The Discrete Wavelet Transform

7.1 Introduction

The discrete Wavelet transform (DWT) is a mathematical tool which has several applications in engineering and mathematics. Most notably, it is used for signal processing. The project aims to triangulate the position of a faulty component in a wind turbine. The faulty component emits a signal that can be filtered and denoised by the DWT. This facilitates a time delay estimation of the recorded signals from the experiment, so the position of the fault can be determined.

This chapter is an application of the theory of linear algebra and filters, which is presented in Chapter 3 and 6 respectively.

7.2 Matrix Formulation

The discrete Wavelet transform is an application of filter banks. It is a convolution of a signal with filters, which can be represented as a matrix transformation. The DWT uses a filter bank as a matrix transformation consisting of a low-pass and a high-pass filter.

The transformation is applied in levels and the single level transformation matrix is defined in Definition 7.1. The notion of levels refer to a decomposition of the input signal. For a signal with cardinality $N = 2^J$ there are $J = \log_2(N)$ levels, with

$$j = J - 1, J - 2, \dots, 1, 0$$

levels of decompositions possible.

Definition 7.1 (Single Level Transformation Matrix)

Let \mathbf{x} with $\text{card}(\mathbf{x}) = N$ be the input vector for the transformation matrix $H_{1,j}$, and let L and B be the $N/2 \times N$ low-pass and high-pass channels respectively. The $N \times N$ single level transformation matrix with filter coefficients \mathbf{h} and \mathbf{g} is then

$$H_{1,j} = \begin{bmatrix} L \\ B \end{bmatrix} = \begin{bmatrix} \ddots & \ddots & \ddots & \dots & 0 & 0 & \dots & \dots \\ \mathbf{h}[M-1] & \mathbf{h}[M-2] & \dots & \mathbf{h}[0] & 0 & 0 & \dots & \dots \\ 0 & 0 & \mathbf{h}[M-1] & \mathbf{h}[M-2] & \dots & \mathbf{h}[0] & 0 & \dots \\ 0 & 0 & 0 & 0 & \mathbf{h}[M-1] & \mathbf{h}[M-2] & \dots & \mathbf{h}[0] \\ \ddots & \ddots & \ddots & \dots & 0 & 0 & \vdots & \dots \\ \mathbf{g}[M-1] & \mathbf{g}[M-2] & \dots & \mathbf{g}[0] & 0 & 0 & \dots & \dots \\ 0 & 0 & \mathbf{g}[M-1] & \mathbf{g}[M-2] & \dots & \mathbf{g}[0] & 0 & \dots \\ 0 & 0 & 0 & 0 & \mathbf{g}[M-1] & \mathbf{g}[M-2] & \dots & \mathbf{g}[0] \end{bmatrix},$$

where $M = \text{card}(\mathbf{h}) = \text{card}(\mathbf{g})$. The subscript j of $H_{1,j}$ indicates the level of the signal after the transformation.

Definition 7.2 (*i*-level Decomposition)

Let $\mathbf{x} \in \mathbb{R}^N$ be a signal in the time domain with $\text{card}(\mathbf{x}) = N = 2^{i+j}$ where $i, j \in \mathbb{N}_0$, then the *i*-level decomposition is given by

$$\mathbf{y} = H_{i,j}\mathbf{x},$$

where $H_{i,j}$ is the $N \times N$ transformation matrix containing the low-pass and high-pass filters that decomposes \mathbf{x} , *i* times such that the signal is at level *j*. The output $\mathbf{y} \in \mathbb{R}^N$ with cardinality *N* contains 2^i subsignals. Each half of the output \mathbf{y} respectively express the approximation and the detail components.

In Definition 7.2 the output \mathbf{y} is the input \mathbf{x} expressed in terms of the bases for the approximation and detail subspaces at level *j*. The *i*-level transformation matrix is called the analysis transformation matrix as it decomposes the signal into increasingly fine frequency components, allowing analysis of the signal.

Applying Theorem 3.8, an *i*-level decomposition transformation matrix is attained by multiplying the single level transformation with itself *i* times. As such the *i*-level transformation matrix is given by

$$H_{i,j} = H_{1,j}^i.$$

Theorem 7.1 (Reconstruction)

Let $\mathbf{x} \in \mathbb{R}^N$ be the original signal and $\mathbf{y} \in \mathbb{R}^N$ the decomposed output. Assume that the inverse of $H_{i,j}$ exists, then the original signal \mathbf{x} can be reconstructed as

$$\mathbf{x} = H_{i,j}^{-1}\mathbf{y}.$$

Proof 7.1

The output \mathbf{y} is given by Definition 7.2 as

$$\mathbf{y} = H_{i,j}\mathbf{x}.$$

Multiplication by $H_{i,j}^{-1}$ yields

$$H_{i,j}^{-1}\mathbf{y} = H_{i,j}^{-1}(H_{i,j}\mathbf{x}).$$

In accordance with Theorem 3.7

$$H_{i,j}^{-1}\mathbf{y} = H_{i,j}^{-1}(H_{i,j}\mathbf{x}) = (H_{i,j}^{-1}H_{i,j})\mathbf{x}.$$

Confer Definition 3.20 the equation is reduced to

$$\begin{aligned} H_{i,j}^{-1}\mathbf{y} &= I_N\mathbf{x} = \mathbf{x} \\ \mathbf{x} &= H_{i,j}^{-1}\mathbf{y}, \end{aligned}$$

thereby completing the proof. ■

As a result of orthogonality the synthesis transformation matrix is the transpose of the analysis transformation matrix according to Theorem 3.13

$$H_{i,j}^{-1} = H_{i,j}^T,$$

thus the original signal \mathbf{x} can be reconstructed as

$$\mathbf{x} = H_{i,j}^T \mathbf{y}.$$

Therefore orthogonality is a desired filter property as it guarantees the existence of an inverse transformation. Furthermore, the inverse transformation is easily derived from the transformation matrix.

Example 7.1

Let $\mathbf{x} = \{3, 5, 2, -8, 9, 2, 1, -3\}$ be a signal with $\text{card}(\mathbf{x}) = 8$. The 2-level decomposition transformation matrix using the Haar filters for $\text{card}(\mathbf{x}) = 8$ is then

$$H_{2,1} = H_{1,2}^2 = \left(\frac{1}{\sqrt{2}}\right)^2 \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \end{bmatrix} \\ = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & -1 & 1 \end{bmatrix}.$$

With the 2-level decomposition transformation matrix, the approximation and detail components at level 1 can be computed:

$$\mathbf{y} = H_{2,1} \mathbf{x} = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \\ 2 \\ -8 \\ 9 \\ -11 \\ -14 \\ -13 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 2 \\ 9 \\ -8 \\ -11 \\ -14 \\ -13 \\ -12 \\ 3 \end{bmatrix}.$$

By Theorem 7.1 \mathbf{x} can be reconstructed, and since the Haar filter is orthogonal, the

reconstruction is perfect:

$$\frac{1}{4} \begin{bmatrix} 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 \\ 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & -1 & 0 & -1 \\ 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 9 \\ -8 \\ -11 \\ -14 \\ -13 \\ -12 \\ 3 \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 12 \\ 20 \\ 8 \\ -32 \\ 36 \\ 8 \\ 4 \\ -12 \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \\ 2 \\ -8 \\ 9 \\ 2 \\ 1 \\ -3 \end{bmatrix}.$$

In practice the DWT can be used in wavelet packet analysis, in which multiple levels of decompositions are analysed simultaneously, also called a multiresolution analysis. [Strang, 1996, p.72] This creates a pyramid structure of the approximation and detail components, seen in Figure 7.1. This analysis enables both time and frequency analysis of the original signal, where the resolution of either depends on the amount of level decomposition. A higher level decomposition resulting in a lower level leads to an increased frequency resolution, whereas a lower level decomposition resulting in a higher level provides a more accurate resolution in time.

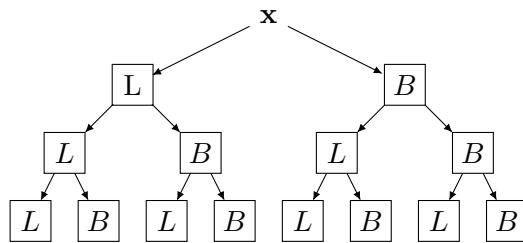


Figure 7.1: Pyramid structure of wavelet packet analysis.

The filters are treated as half-band filters and thus the specific approximation or detail components containing desired information can be determined before decomposing the signal, allowing for a less memory heavy analysis that only computes those components. This analysis can be advantageous as it decreases the amount of calculations. This analysis is called a wavelet analysis, and it computes the least amount of approximation and detail components needed to reach a desired component. In comparison to the wavelet packet analysis, the structure of the wavelet analysis is illustrated in Figure 7.2. Note that the detail components belonging to the approximations are computed and saved, as these contain necessary information for reconstruction.

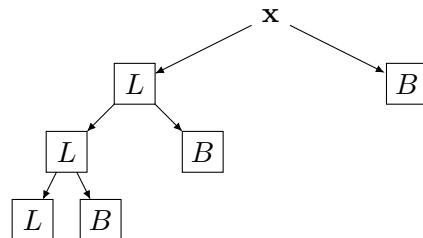


Figure 7.2: Structure of wavelet analysis, illustrating the approximation branch.

7.3 Properties of the Transform

In multiresolution analysis a signal is decomposed into levels, the lower the level, the better the frequency resolution will be. As such it is required to find a balance between time and frequency resolution for the specific signal in question.

When decomposing signals through linear operations such as a matrix transformation it is natural to describe the output of the transformation in terms of subspaces. The subspace denoted by V_J encompasses the space for the signal \mathbf{x} ($\mathbf{x} \in V_J$). The single level DWT produces two sequences which respectively belong to the approximation subspace and detail subspace, denoted as V_{J-1} and W_{J-1} . The subspace V_{J-1} contains less information than V_J , but no information is lost, since the remaining information is contained by W_{J-1} . [Strang, 1996, p. 175-176]

$$V_J = V_{J-1} \oplus W_{J-1}.$$

Notice that performing a single level DWT multiple times results in

$$V_0 \subset V_1 \subset V_2 \cdots \subset V_J$$

The number of subspaces n formed by a wavelet packet analysis can be expressed as

$$n = \sum_{k=0}^i 2^k$$

subspaces, where i is the amount of decomposition levels.

7.4 The Transform and Subspaces

For a single level transformation the output belongs to two subspaces, for which the direct sum is equal to the space the input belongs to. The basis for each of these subspaces is respectively the row space of the upper and lower half of the transformation matrix $H_{1,j}$. To display this property, an input $\mathbf{x} \in \mathbb{R}^8$ with $\text{card}(\mathbf{x}) = 8$ is used. The corresponding transformation matrix using the Haar filters is

$$H_{1,2} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix}.$$

The transformation translates the input \mathbf{x} to the bases \mathcal{B}_{V_2} and \mathcal{B}_{W_2} . To obtain perfect reconstruction and express the components in terms of the standard basis, \mathcal{B}_{V_2} and \mathcal{B}_{W_2}

are applied in the synthesis matrix. The basis for each subspace is

$$\mathcal{B}_{V_2} = \left\{ \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \right\}$$

$$\mathcal{B}_{W_2} = \left\{ \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 0 \\ -1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ -1 \\ 1 \\ 0 \end{bmatrix}, \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\},$$

where \mathcal{B}_{V_2} is a basis for the V_2 subspace and \mathcal{B}_{W_2} is a basis for the W_2 subspace. Note that the vectors in each basis are all linearly independent (a requirement for basis) and their union forms a basis for \mathbb{R}^8 . It is clear that the bases of the subspaces \mathcal{B}_{V_2} and \mathcal{B}_{W_2} formed by the Haar filters are both orthonormal. Orthonormality is a desired property as it maintains the energy of the input and the inverse transform is given by the transpose. The bases for V_1 and W_1 are given by the row space of the upper and lower half of the two level transformation matrix

$$H_{2,1} = H_{1,2}^2.$$

This can be done recursively for any desired subspace of V_J . In this specific example $J = 3$.

8 | Selected Filters

8.1 Introduction

The filters used to decompose the signal through the DWT are essential. For accurate estimation of specific properties in a signal, it is important that the filters resemble those features. If it is desired to synthesise impulses only, then it is advantageous to select a filter that resembles an impulse. This chapter describes the selected filters for this project.

8.2 Daubechies 16

The Daubechies wavelets are a family of orthogonal maxflat wavelets. [Strang, 1996, Chapter 5.5] The family is based on the work of Ingrid Daubechies. One of the features of Daubechies wavelets are their ability to represent polynomials. The Daubechies 16 filter with the acronym db16, is a filter with 32 coefficients for the low-pass and high-pass filter respectively. Impulses resemble polynomials of high degree as they oscillate rapidly with limited support. Due to the 32 coefficients, it is hard to properly list the numerical coefficients, instead the filters are displayed in Figure 8.1.

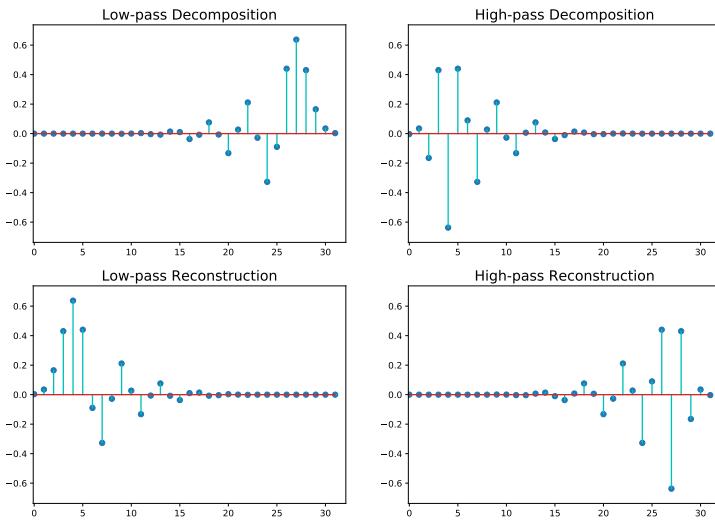


Figure 8.1: Filter coefficients of the Daubechies 16 filter.

The structure of the filters resemble an impulse and therefore it is suitable for detecting and representing impulses.

8.2.1 Frequency Response

In practice the normalised filter coefficients are used. The frequency responses of the filters are derived by equation (6.2).

The frequency response of the decomposition and reconstruction filters are identical as they are simply inverted. The frequency response of the filters are seen in Figure 8.2.

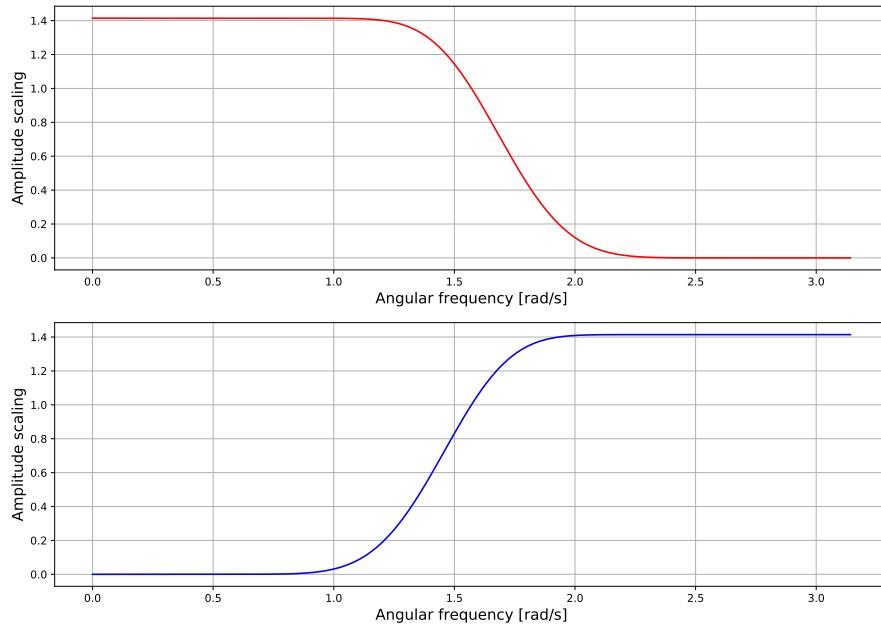


Figure 8.2: Frequency response of the db16 filters. The low-pass filter is the red curve and the high-pass filter is the blue curve.

From the frequency response it is apparent that the filters are half-band filters, albeit not ideal half-band filters. It is also clear that within the filters bandwidth, the amplitudes are amplified by a notable amount.

The frequency response of the Daubechies filters are similar. The higher the cardinality of the filters, the steeper the frequency response becomes near the cutoff frequency.

9 | Filtering

9.1 Introduction

In modelling and ideal examples, the signals can be represented perfectly. But as a consequence of sampling analogue signals, the signal will be affected by disturbances that obscure the signal of interest. These disturbances in the signal are called noise, and it is inevitable in all signal processing. On top of noise, it can be of interest to filter out other frequency components of the signal. In this chapter the DWT is used to filter noise as well as redundant frequency components of a synthetic signal.

9.2 Filtering with the DWT

Noise is unwanted disturbances in the studied signal. With sound, noise can be caused by reverberation as well as ambient sounds. The process of removing noise is called denoising. There exists a plethora of methods to denoise a signal. One method to denoise a signal is the DWT. An application of DWT is pinpointing the wavelet packet with most energy assuming the noise has less energy than the signal, and then synthesising only that packet to remove noise. To illustrate this, a simple sine wave is generated with added brown noise, seen in Figure 9.1.

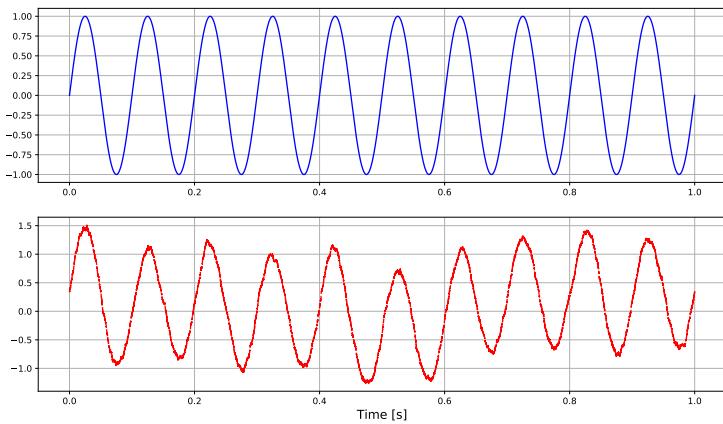


Figure 9.1: The first plot shows a single sine wave. The second plot shows the single sine wave with added brown noise.

To denoise the signal, a wavelet packet analysis is conducted with different levels of decompositions, where the wavelet packet with most energy is synthesised. At few levels of

decompositions, the noise still remains in the signal, but when the level of decompositions is increased, the bandwidth becomes narrower for the wavelet packets and thus when synthesising, the noise is gradually removed. The signal synthesised from different levels of decomposition can be seen in Figure 9.2. Due to the low frequencies of brown noise, not all of them are filtered since the signal has a low frequency as well.

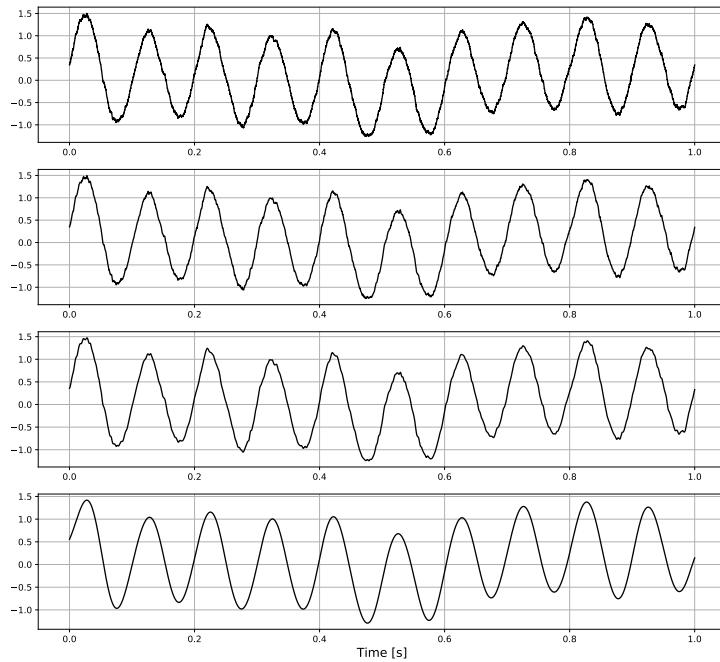


Figure 9.2: Signal denoised with a level 4, level 7, level 9 and level 13 DWT decomposition.

The DWT will now be used to filter redundant frequency components as well as noise. The frequency components are removed by setting the components of the packets with the frequencies, to zero. It can be complicated to differentiate between frequencies if their values are too close. If two frequencies have almost the same frequency, then a very fine bandwidth is necessary.

The signal with and without noise is seen in Figure 9.3.

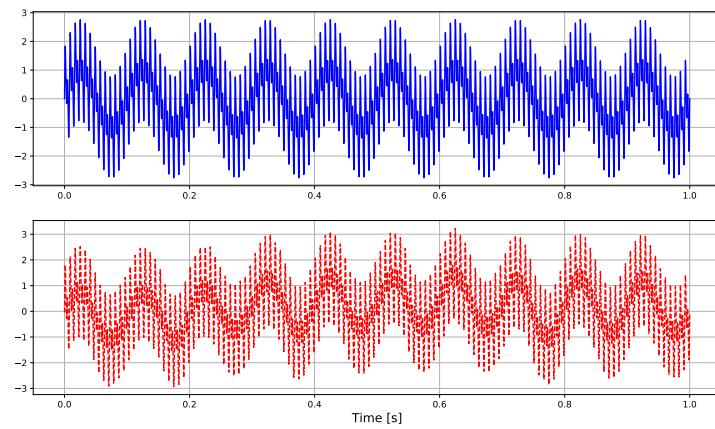


Figure 9.3: The first plot shows a signal consisting of three frequencies. The second plot shows the signal with added white noise.

This slightly more complicated signal is denoised by the same method. The levels of decomposition are from level 2 to level 6, and the denoising effect is visualized in Figure 9.4.

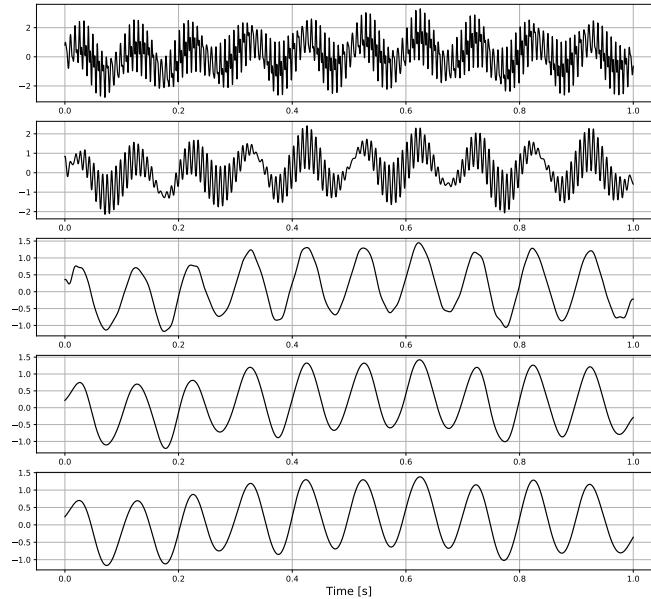


Figure 9.4: Signal filtered and denoised with a level 9 to level 13 DWT decomposition.

10 | The Discrete Fourier Transform

10.1 Introduction

When recording continuous signals in the time-domain a problem arises: it is rarely useful to know a signal only in the time-domain. An important tool from signal processing is the Fourier transform, which is a linear operator that transforms a signal from the time-domain to the frequency-domain, thereby highlighting aspects of the signal different from those of the time-domain.

This chapter is based on “Fourier and Laplace Transforms”. [Beerends, 2003]

10.2 Prerequisite

In general, the Fourier theorems and the Fourier transform work with signals that are periodic, which is a property defined in Definition 4.2. By transforming into the frequency domain, certain operations such as determining or modifying specific frequency-components become a matter of addition and multiplication. The general idea behind the Fourier transform and the subsequent Fourier Series Theorem is that certain periodic functions can be decomposed into infinite sums of scaled sinusoidal functions. The constituent parts of these infinite sums are known as Fourier Series and will be discussed in the following section.

10.3 Fourier Series

It is often of interest to analyse a periodic signal in terms of its frequency spectrum, this however, is difficult to do in the time domain. The Fourier series approximates the periodic signal with weighted sinusoids, thus the frequency spectrum is easily analysed. The complex Fourier series transforms a periodic signal in time into a complex function of frequency.

Definition 10.1 (Complex Fourier Series)

The Complex Fourier Series of a periodic signal $x_c : \mathbb{R} \rightarrow \mathbb{C}$ with period $T \in \mathbb{R}$ and fundamental frequency $\omega_0 = 2\pi/T$ is

$$x_c(t) = \sum_{k=-\infty}^{\infty} c_k e^{ik\omega_0 t},$$

where all $c_k \in \mathbb{C}$ are called the Fourier coefficients of the signal x_c .

[Beerends, 2003, p.69]

The Fourier coefficients c_k of a periodic function signal x_c mentioned in Definition 10.1 can be calculated as

$$c_k = \frac{1}{T} \int_0^T x_c(t) e^{-ik\omega_0 t} dt.$$

By applying the well-known trapezoidal rule with integrand $g(t) = x_c(t)e^{-ik\omega_0 t}$ with $\omega_0 = \frac{2\pi}{T}$, the approximation

$$c_k \approx \frac{1}{T} \frac{T}{N} \sum_{n=0}^{N-1} \mathbf{x}[n] e^{-ik\omega_0 n T/N} = \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{x}[n] e^{-2\pi i n k / N} \quad (10.1)$$

for the Fourier coefficient c_k is obtained. It follows from (10.1) that

$$c_k \approx \frac{1}{N} \mathcal{F}(\mathbf{x})[k].$$

[Beerends, 2003, p.358]

Further study of Fourier Series is beyond the scope of this project and will not be used in subsequent chapters.

10.4 Discrete Fourier Transform

The right-hand side of equation (10.1) is important enough to warrant its own definition, which can be seen in Definition 10.2.

Definition 10.2 (Discrete Fourier Transform)

Let \mathbf{x} be an arbitrary periodic discrete-time signal with period $N \in \mathbb{N}$. The output sequence $\mathcal{F}(\mathbf{x})$ called the N -point discrete Fourier transform of \mathbf{x} is then for all $k \in \mathbb{Z}$ defined by

$$\mathcal{F}(\mathbf{x})[k] = \sum_{n=0}^{N-1} \mathbf{x}[n] e^{-2\pi i n k / N}.$$

[Beerends, 2003, p.360]

In practise when calculating the discrete Fourier transform (DFT) of an input signal \mathbf{x} , Definition 10.2 is rarely used. Instead, more efficient algorithms – such as the fast Fourier transform – are used.

The discrete Fourier transformation is invertible and thus \mathbf{x} can be completely recovered if $\mathcal{F}(\mathbf{x})$ is known. This is defined as the inverse discrete Fourier transform (IDFT).

Definition 10.3 (Inverse Discrete Fourier Transform)

Let \mathbf{x} be a periodic discrete-time signal with period N and $\mathcal{F}(\mathbf{x})$ given by Definition 10.2. Then $\mathbf{x}[n]$ is given for all $n \in \mathbb{Z}$ as

$$\mathbf{x}[n] = \frac{1}{N} \sum_{k=0}^{N-1} \mathcal{F}(\mathbf{x})[k] e^{2\pi i n k / N}.$$

[Beerends, 2003, p.362]

The sequence $\mathcal{F}(\mathbf{x})$ consists of complex values. The modulus $|\mathcal{F}(\mathbf{x})|$ of the spectrum $\mathcal{F}(\mathbf{x})$ is called the amplitude spectrum of \mathbf{x} . The phase for $\mathcal{F}(\mathbf{x})$ is given by the argument of the complex values.

11 | Identifying a New Frequency

11.1 Introduction

The discrete Fourier transform approximates a signal by a sum of weighted sinusoids and can because of that, describe the signal in terms of frequency components. This chapter displays the application of the DFT in this project. For two similar signals that share some frequency components, the difference between their amplitude spectra can indicate which frequencies are present in one signal but not the other.

11.2 Identifying a New Frequency through the DFT

A change of the frequency components of a signal in the time domain can be detected by the difference of the amplitude spectra of the signal before and after the change. The difference is computed by subtracting the first signal from the latter signal. This is performed on two synthetically generated signals consisting of sinusoids and brown noise. The first signal represents a signal during normal operation where a fault has yet to occur and the second represents the signal containing the sound of the faulty component. Both signals are seen in Figure 11.1. Note that only one frequency component of the signal has changed.

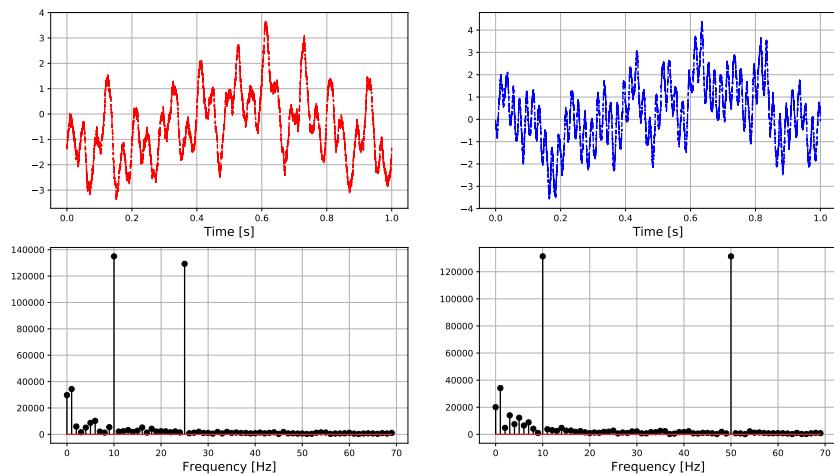


Figure 11.1: Two synthetically generated signals each consisting of two weighted sinusoids and brown noise.

The sound of the faulty component is identified by the difference of the two amplitude spectra, which generates a new amplitude spectrum, where the frequency with the

maximum positive value represents the sound of the faulty component. All values of the amplitude spectrum that are not the identified frequency are set to zero, and then the IDFT (Definition 10.3) is performed on the spectrum, resulting in a signal in the time domain consisting solely of the sound of the faulty component.

The amplitude spectrum where the sound of the faulty component is identified, as well as the sine wave with the corresponding frequency, are seen in Figure 11.2.

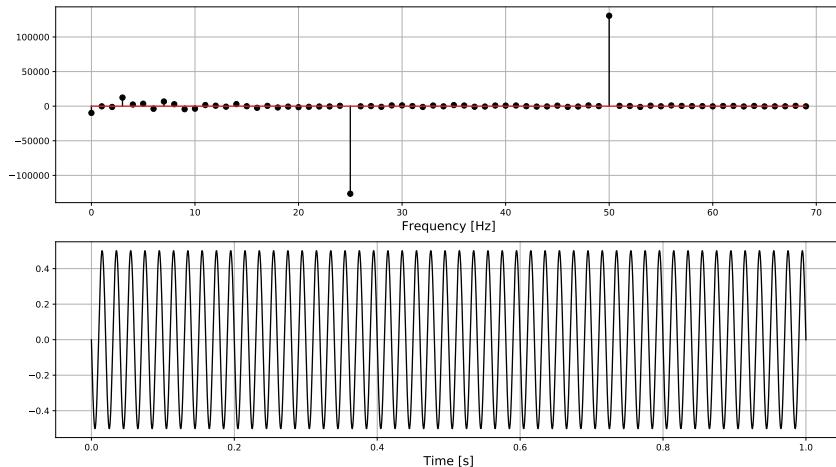


Figure 11.2: The amplitude spectrum containing the sound of the faulty component and the corresponding signal in the time-domain.

The added brown noise does not have enough energy to dominate the generated frequencies and thus the frequency is accurately identified.

12 | Localisation through Time Delay Estimation

12.1 Introduction

It is possible to approximate the location of the source of an emitted sound in two dimensions, under the assumption that the sound has been recorded by three or more synchronised microphones from different known positions. The different positions of the microphones causes the sound to arrive at certain points in time, thus making it possible to estimate a time delay between the microphones, which can then be used to triangulate the position of the sound emitting source.

The methods used in this project for time delay estimation and triangulation are demonstrated in this chapter.

12.2 Time Delay Estimation through Cross-correlation

To demonstrate how cross-correlation can facilitate a time delay estimation, some synthetic data is created. Three identical signals \mathbf{x}_1 , \mathbf{x}_2 , and \mathbf{x}_3 are generated, after which \mathbf{x}_2 is shifted 50 samples and \mathbf{x}_3 is shifted 100 samples. The three signals are illustrated in Figure 12.1.

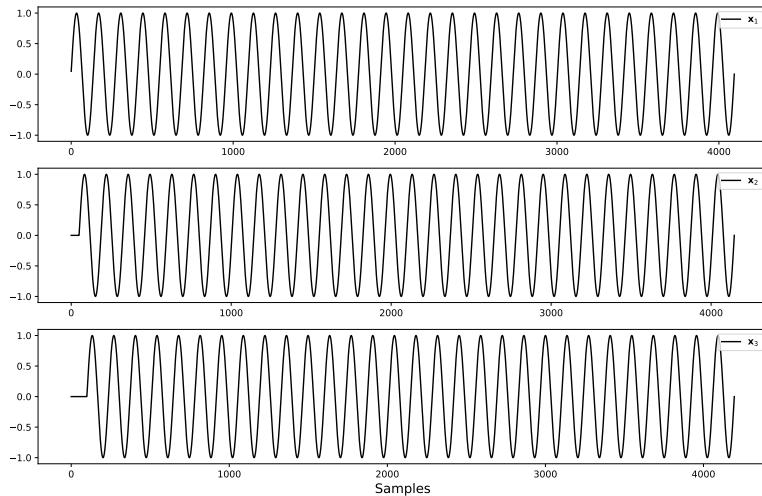


Figure 12.1: Three identical signals. \mathbf{x}_2 is shifted 50 samples, and \mathbf{x}_3 is shifted 100 samples.

The synthetic signals will now be cross-correlated in pairs respectively $\mathbf{x}_1 \star \mathbf{x}_2$, $\mathbf{x}_1 \star \mathbf{x}_3$, and then $\mathbf{x}_2 \star \mathbf{x}_3$. The three cross-correlations can be seen in Figure 12.2.

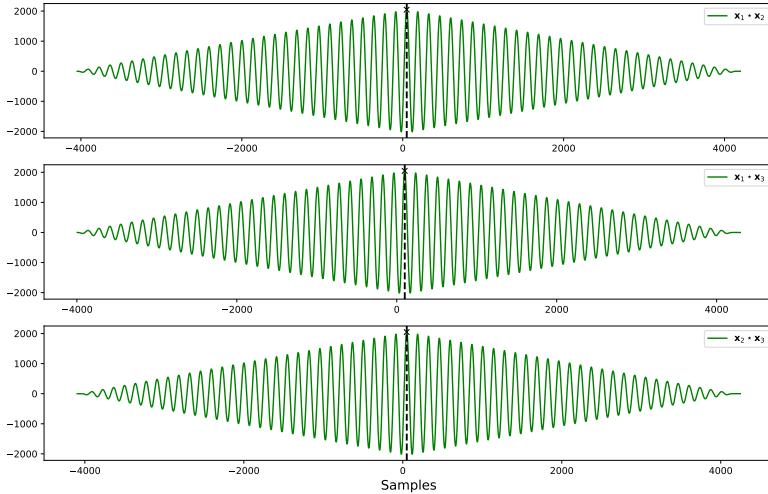


Figure 12.2: The three cross-correlations.

The cross-correlations yield sequences where the sample with the largest value is used to compute the sample delays d . The values of d are seen in Table 12.1.

Cross-correlations	$\mathbf{x}_1 \star \mathbf{x}_2$	$\mathbf{x}_1 \star \mathbf{x}_3$	$\mathbf{x}_2 \star \mathbf{x}_3$
d (Samples)	50	100	50

Table 12.1: The sample delay of the signals.

The dashed lines in Figure 12.2 indicate the amount of samples one signal is shifted relative to the other.

12.3 Triangulation of Position in Two Dimensions

To perform triangulation in two dimensions it is necessary to estimate the sample delays between three microphones. The sample delay d of each pair of microphones can be restated as a delay in time Δt as follows:

$$\Delta t = \frac{d}{f_s},$$

where f_s is the sampling frequency.

This time-difference can be expressed as a difference in distance Δs

$$\Delta s = \Delta tv = \frac{dv}{f_s}, \quad (12.1)$$

where v is the speed of sound in the given medium.

Since Δs is the difference in distance from the sound emitting source to the microphones, then through the use of the Pythagorean theorem, the following must hold true:

$$\Delta s = \sqrt{(x_1 - x_s)^2 + (y_1 - y_s)^2} - \sqrt{(x_2 - x_s)^2 + (y_2 - y_s)^2}, \quad (12.2)$$

where (x_1, y_1) is the coordinates of the first microphone, (x_2, y_2) is the coordinates of the second microphone, and (x_s, y_s) is the coordinates of the sound emitting source.

This leads to the conclusion that a pair of microphones limits the position of the sound emitting source to a certain hyperbola. By using three microphones the position of the sound emitting source is limited to the intersection(s) of two such hyperbolas.

12.3.1 Practical Procedure

In order to triangulate the location of a sound emitting source in the plane, at least three microphones must be used. The procedure of triangulation is now described from a practical point of view, and in this description of the procedure, three microphones are used and positioned in an equilateral triangle. The three microphones are called α, β and γ and the sound emitting source is referred to as a speaker.

The first step in the procedure is to calculate the sample delays between the three microphones. The next step is to calculate a difference in distance, which is given by equation (12.1) based on the estimated sample delays. The approximated x_s and y_s coordinates of the speaker is given by equation (12.2) for each pair of microphones:

$$\Delta s_{\alpha,\beta} = \sqrt{(x_\beta - x_s)^2 + (y_\beta - y_s)^2} - \sqrt{(x_\alpha - x_s)^2 + (y_\alpha - y_s)^2} \quad (12.3)$$

$$\Delta s_{\beta,\gamma} = \sqrt{(x_\gamma - x_s)^2 + (y_\gamma - y_s)^2} - \sqrt{(x_\beta - x_s)^2 + (y_\beta - y_s)^2} \quad (12.4)$$

$$\Delta s_{\alpha,\gamma} = \sqrt{(x_\gamma - x_s)^2 + (y_\gamma - y_s)^2} - \sqrt{(x_\alpha - x_s)^2 + (y_\alpha - y_s)^2}, \quad (12.5)$$

where x_α and y_α are the coordinates for microphone α . The coordinates of β and γ are denoted likewise.

The location of the speaker is given in relation to a reference point P_R for the microphones, which is set as the midpoint between the microphones α and β . The coordinates of the midpoint is given by

$$P_R = \left(x_\beta + \frac{x_\alpha - x_\beta}{2}, y_\beta + \frac{y_\alpha - y_\beta}{2} \right).$$

A visualisation of the microphones and the reference point is seen in Figure 12.3.

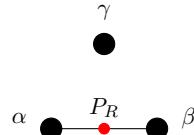


Figure 12.3: The black dots represent the microphones, and the red dot is the reference point.

The location of the speaker is confined to the hyperbolas. If there are any intersections of the hyperbolas, the location can be limited to these intersections. In theory only one intersection of the three hyperbolas should exist, but since the intersections are in practice calculated based on discrete time delays, and an experiment is conducted with some uncertainty, it is highly unlikely that this will be the case. Instead multiple intersections can occur. In the case of there being no intersections the point where two hyperbolas are closest to intersect is treated as an intersection. Due to the fact that multiple intersections suggest that there can be more than one possible position of the speaker, it is more relevant to state the location of the sound source as a single angle or an angle interval.

From P_R two vectors are formed, one from P_R to microphone α called \mathbf{r} , and the other from P_R to an intersection called \mathbf{v}_1 . In the case of two intersections an additional vector \mathbf{v}_2 is formed from P_R to the second intersection.

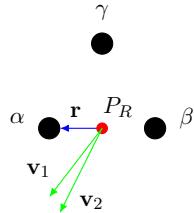


Figure 12.4: Vectors used to calculate the angle of the speaker.

If the hyperbolas have a single intersection then the location of the sound source is in the angle of \mathbf{v}_1 relative to \mathbf{r} . If the hyperbolas have two intersections, the speaker is confined to the interval between the angle of \mathbf{v}_1 relative to \mathbf{r} and the angle of \mathbf{v}_2 relative to \mathbf{r} . If the hyperbolas does not have an intersection, then the point where the two hyperbolas are closest to intersect is treated as an intersection, and the angle is calculated as for the hyperbolas with a single intersection.

13 | Practical Experiment

13.1 Introduction

The purpose of the experiment is to validate that a practical implementation of the methods for filtering, identifying a new frequency, time-delay estimation, and triangulation yields sufficient results.

This chapter explains the setup and the procedure of the experiment.

13.2 Setup

In the project scope the project is limited to four components: The pitch system, brake, gearbox, and generator. The setup of the experiment is based on the interior of a wind turbine nacelle. The positions of the four components in a nacelle are displayed in Figure 13.1.

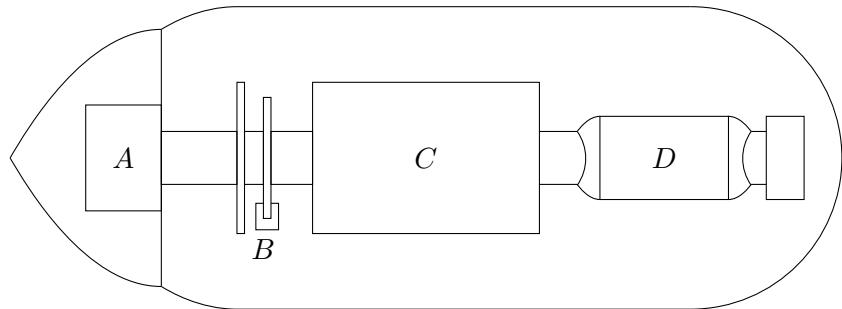


Figure 13.1: Sketch of a wind turbine’s nacelle. *A* is the pitch system, *B* is the brake, *C* is the gearbox, and *D* is the generator.

For the experiment four speakers are used, each representing a component in the nacelle. The equipment used for the setup can be seen in Appendix A.1. The experiment is conducted in a room with the dimensions $7.762 \text{ m} \times 4.11 \text{ m}$. Three microphones are positioned in an equilateral triangle 40 cm apart, in the same plane as the speakers. The height of all components is adjusted with a Stanley laser to the height 1.10 m, and all speakers are turned towards the microphones. The exact placement of the microphones and speakers are seen in Figure 13.2. In Appendix A.3, photos of the practical setup can be seen.

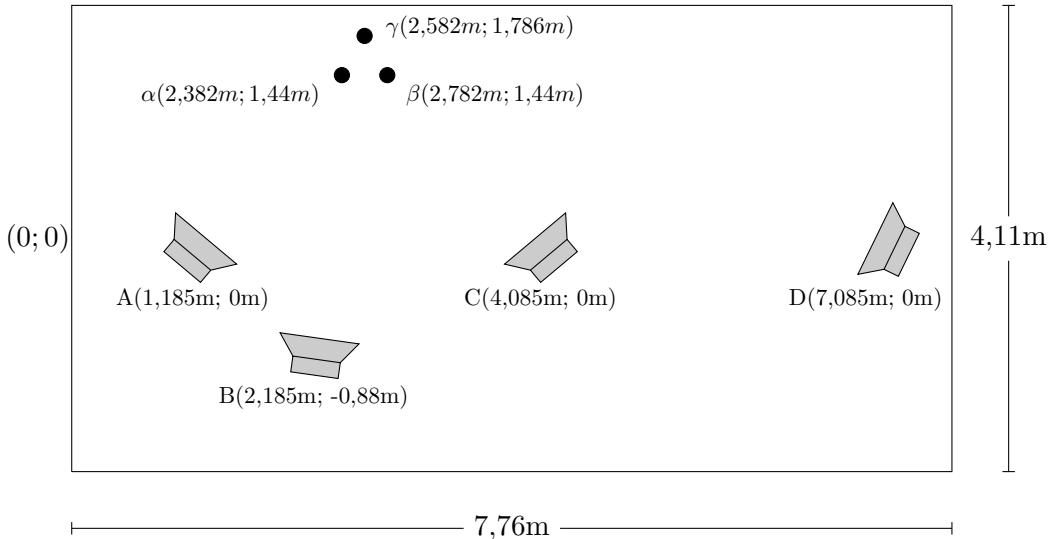


Figure 13.2: Arrangement of the equipment. A, B, C and D represent the speakers, whereas α , β and γ represent the microphones

13.3 Procedure

Eight recordings are conducted. For the first four recordings the four speakers emit a constant frequency for thirty seconds. After 15 seconds, a fault sound is introduced as an impulse added to the sound emitted by one of the speakers. The speaker emitting the fault sound is changed to a new speaker for each recording. The impulse sound imitates a cracked component.

For the last four recordings the speakers emit a constant frequency for thirty seconds. When the fault sound is introduced after 15 seconds, one of the speakers changes frequency. The frequency is either doubled or halved. The speaker emitting the fault sound is changed to a new speaker for each recording. The altering in frequency is simulating wear on the component.

The exact frequencies emitted can be seen in the table of conducted recordings in Appendix A.2. The emitted sounds are generated with Audacity.

All of the eight recordings are conducted with noise added to the sounds emitted by the speakers for the entire duration of the recordings. The purpose of this noise is to simulate the noise caused by the remaining machinery. The amplitude spectrum of the added noise can be seen in Figure 13.3. The distribution of the noise is weighted towards the low frequencies.

The added noise is generated based on the amplitude spectrum presented in Figure 1.4, which contains the frequencies of wind turbine machinery.

All of the recordings are sampled with a sampling frequency $f_s = 48,000$ Hz, which causes the sound to move $\Delta s = \frac{343 \text{ [m/s]}}{48,000 \text{ [s}^{-1}\text{]}} = 0.00715$ m per sampling period. The sampling frequency allows for a precision of 0.7 cm. A higher sampling frequency would result in an even better precision, but would also result in larger data sets. As such a sampling frequency of 48,000 Hz is assessed to be sufficient.

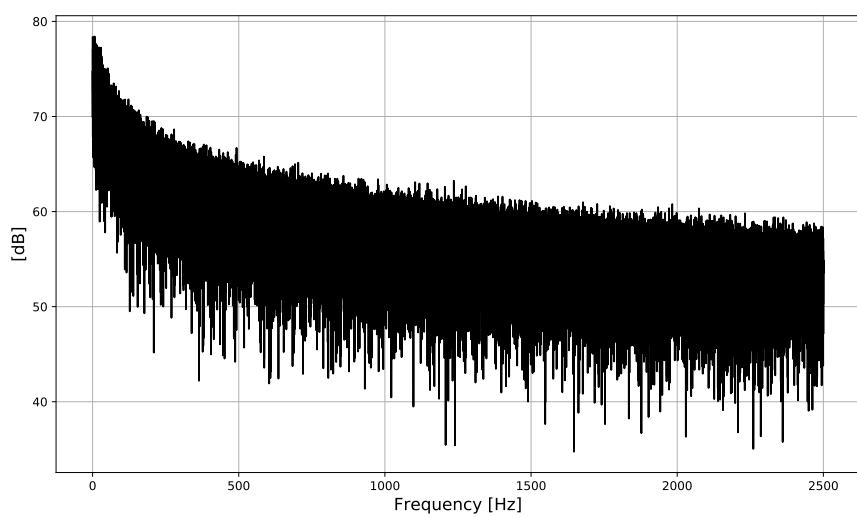


Figure 13.3: Amplitude spectrum of the added noise.

14 | Processing of Data from Experiment

14.1 Introduction

In this chapter the data collected from the practical experiment is analysed using numerical software developed in Python 3.6.5 and Maple 2018. The duration of the recorded signals are 30 seconds, which with a sampling frequency of 48,000 Hz amounts to 1,440,000 samples for each recorded signal. The data collection as well as the concepts for the signal processing, is outlined in Figure 14.1.

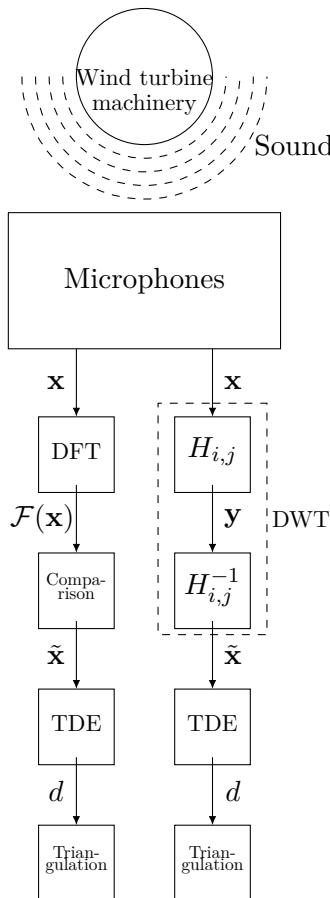


Figure 14.1: Block diagram of the methods applied to locate a change in the sound signal.

14.2 DWT Analysis

14.2.1 Description of the DWT Analysis Procedure

The first step in the signal processing is choosing a segment of the raw data containing the fault. This is conducted for each signal of the three microphones. The segments are denoted \mathbf{x}_α , \mathbf{x}_β , and \mathbf{x}_γ . Each segment has 2^{19} samples, thus $\text{card}(\mathbf{x}_\alpha) = \text{card}(\mathbf{x}_\beta) = \text{card}(\mathbf{x}_\gamma) = 2^{19} = 524,288$. To avoid numerical errors, the three signals are standardised by dividing the values with the standard deviation.

The next step is a wavelet packet analysis of each signal. The Daubechies 16 filter is used in the analyses. The signals \mathbf{x}_α , \mathbf{x}_β , and \mathbf{x}_γ are separately decomposed i levels through a wavelet packet analysis to obtain a sufficient frequency resolution. Too few levels of decomposition causes the frequency resolution to be inadequate, and too many levels of decomposition significantly increases the amount of computations. The i level decomposition results in 2^i packets each with cardinality $M = \frac{N}{2^i}$.

The next step is discarding the packets that contain the frequencies which are present during normal operation (when a fault has yet to occur), thus filtering the signal and isolating the transients. In the software a packet is discarded by setting all the components of the packet to zero. The frequencies present during normal operation are known for the experiment, and they are denoted f_1 , f_2 , f_3 , and f_4 . For a real installation of the concept in a wind turbine nacelle, the frequencies would be unknown. These could be determined by a DFT calibration.

The discarding is executed by treating the filters as ideal halfband filters, and then discarding the packets whose passband includes either f_1 , f_2 , f_3 , or f_4 . To account for the filters not being ideal halfband filters, the limits of the passbands are artificially extended by a specific amount before discarding. The amount is referred to as the threshold. The threshold ensures that a sufficient amount of the energy contained in the signals from f_1 , f_2 , f_3 and f_4 is removed.

To detect and ensure that the signals contain a sound of a fault, the total energy of the remaining packets for each of the signals is examined. If the total energy is still large relative to the energy in the original signal, then the original signal must contain

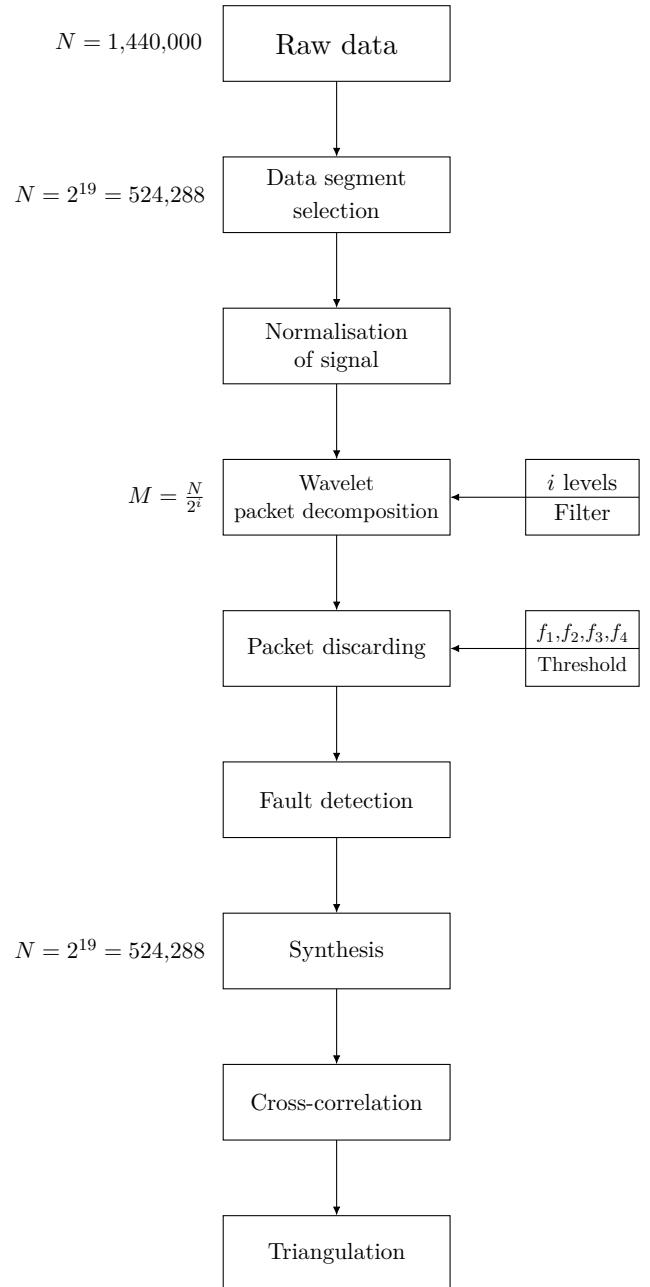


Figure 14.2: Block diagram illustrating the signal processing using the DWT.

information apart from f_1, f_2, f_3 , and f_4 , hence the signal must contain information of a fault.

The syntheses are based on the wavelet packet analysis for one of the signals, for instance \mathbf{x}_α . A corresponding wavelet packet analysis is conducted for the other two signals \mathbf{x}_β and \mathbf{x}_γ by choosing the packets at the same indices as in the decomposition of the first signal. The signals $\tilde{\mathbf{x}}_\alpha, \tilde{\mathbf{x}}_\beta$, and $\tilde{\mathbf{x}}_\gamma$ are the resulting signals from the synthesis of the remaining packets of each signal.

Afterwards, the three synthesised signals are cross-correlated, resulting in the amount of samples d the fault is shifted. From these samples the location of the faulty component will be triangulated, and lastly a deviation from the actual location will be computed.

14.2.2 Analysis of Impulses

The procedure for the analysis is described and is now conducted for one of the four recordings which has impulses as the sound indicating a fault. The process of the analysis is detailed for one recording. For the remaining recordings, the results are merely presented. The conducted analysis is of Recording 6 (A.2) where the faulty sound is emitted by Speaker B after 15 seconds. A segment of the data containing the sound of the fault is chosen for the signal of each microphone as seen in Figure 14.3. The segments start at sample number 800,000 and extend up to and including sample number 1,324,287.

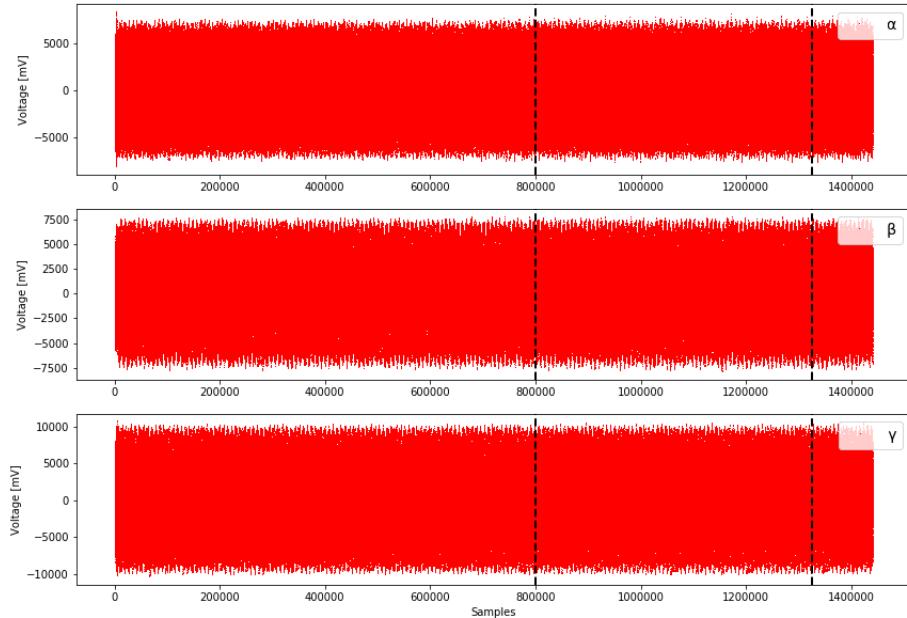


Figure 14.3: Sound signal of microphone α , β , and γ from Recording 6, where a segment of data is selected with size $2^{19} = 524288$ samples.

The segments of the recordings for each of the microphones is chosen to be of size $2^{19} = 524288$ samples, which makes a wavelet packet analysis of several levels possible. To avoid numerical errors, the three signals are standardised before they are decomposed. A 10 level wavelet packet decomposition is then executed resulting in $2^{10} = 1024$ packets

with a cardinality of $524288/2^{10} = 512$ samples at level 9 for each of the signals. The filter used for the decompositions is the Daubechies 16 filter.

To determine which packets need to be discarded, the passband of the packets at level 9 for one of the signals is compared with the calibrated frequencies (125 Hz, 240 Hz, 737 Hz, and 1000 Hz). In this case, it is signal \mathbf{x}_α . The packets whose passband contain the calibrated frequencies are then discarded, and the packets at the same indices are removed from the two other signals \mathbf{x}_β and \mathbf{x}_γ .

Since the impulses consists mainly of high frequencies, and the calibrated frequencies are similar to the frequencies containing the most energy of the added noise 13.3, efficient denoising can be performed by setting a high threshold. The threshold artificially extends the passband limits of the packets. Setting the threshold sufficiently high, ensures that all low frequencies including the low-frequent noise are filtered.

The remaining packets from each of the signals are then synthesised resulting in three reconstructed signals $\tilde{\mathbf{x}}_\alpha$, $\tilde{\mathbf{x}}_\beta$, and $\tilde{\mathbf{x}}_\gamma$, which can be seen in Figure 14.4.

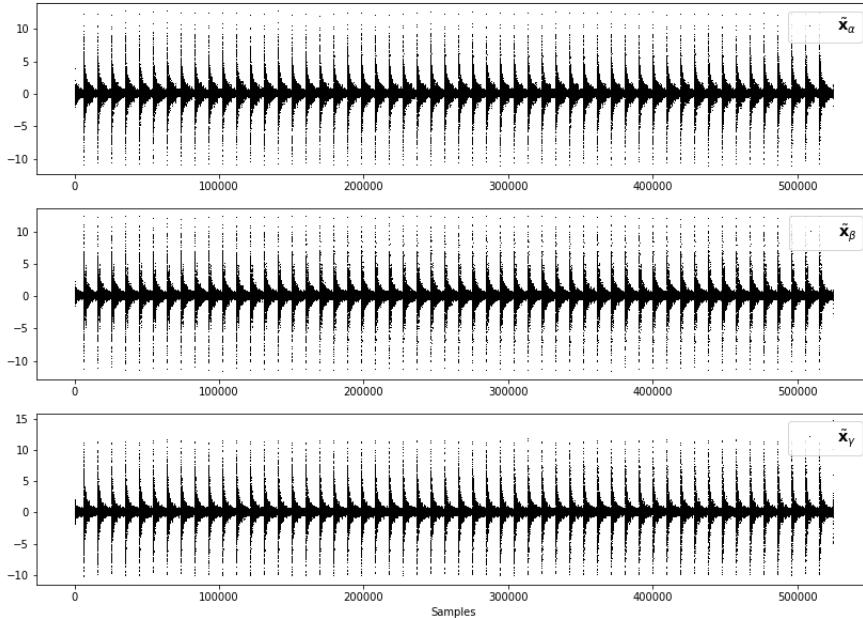


Figure 14.4: Reconstructed signals from the remaining packets.

The three signals have now been denoised and filtered which enables an estimation of the sample delay between each pair of microphones. This is performed by cross-correlation. The three cross-correlations are illustrated in Figure 14.5.

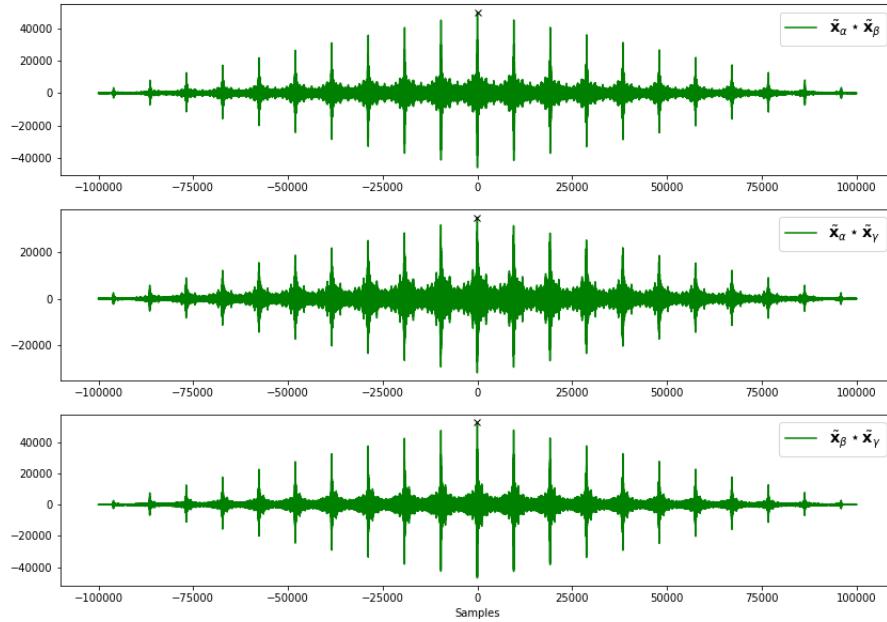


Figure 14.5: Cross-correlations of the three synthesised signals \tilde{x}_α , \tilde{x}_β , and \tilde{x}_γ .

Since the three reconstructed signals are periodic, it is not necessary to cross-correlate the full length of the signals. Therefore, the cross-correlations are performed on 100,000 samples from the same indices of each signal, in order to reduce the amount of calculations, thereby reducing the computation time. The results of the sample delay estimation of Recording 6 is displayed in Figure 14.1 alongside the three other similar experiments.

Recording	Speaker	$d_{\tilde{x}_\alpha, \tilde{x}_\beta}$	$d_{\tilde{x}_\alpha, \tilde{x}_\gamma}$	$d_{\tilde{x}_\beta, \tilde{x}_\gamma}$
5	A	38	124	18
6	B	9	52	42
7	C	-40	15	55
8	D	-704	-10	-18

Table 14.1: The sample delays d of Recording 5-8.

The final step in the analysis is the localisation of the speaker emitting the sound of a fault. This is conducted by a triangulation based on the sample delay between the microphones. The procedure of the triangulation is outlined in Section 12.3.1. The placement of the speaker and microphones are treated as coordinates in the plane as seen in Figure 13.2.

A difference in distance given by equation (12.1) is calculated based on the estimated sample delays for Recording 6. The approximated x_s and y_s coordinates of the speaker is given by the three expressions from equation (12.3), (12.4) and (12.5). The solutions to the three equations are a distinct hyperbola for each equation.

The hyperbolas calculated for Recording 6 are plotted pairwise to estimate the angle of the speaker relative to the reference point P_R (Figure 12.3). The plots are seen in Figure 14.6a, 14.6b and 14.6c.

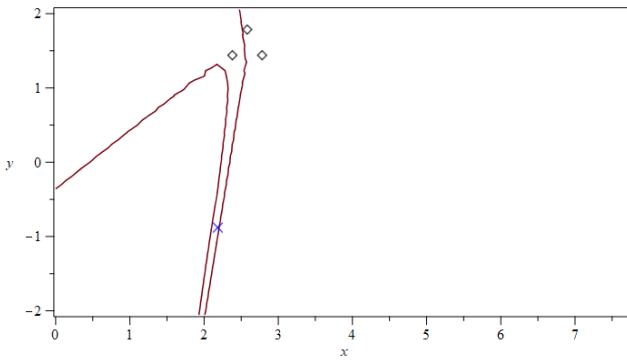
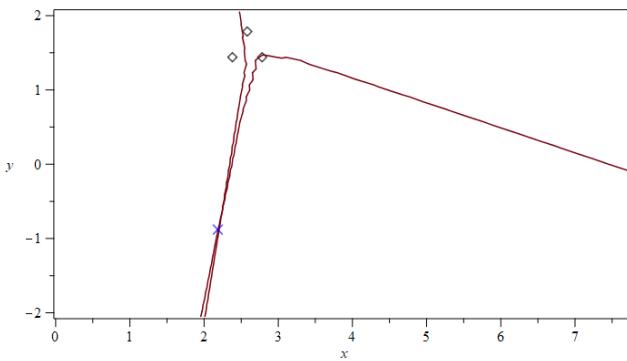
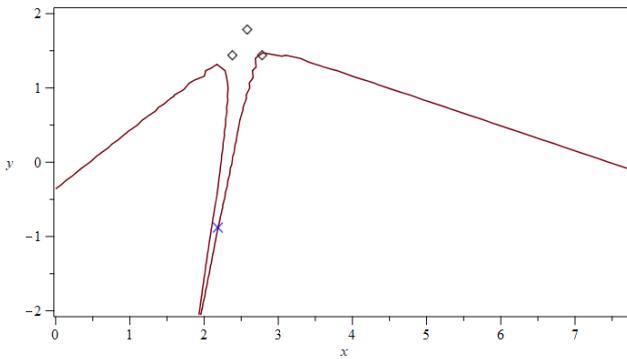

 (a) The hyperbolas of $\Delta s_{\alpha,\beta}$ and $\Delta s_{\alpha,\gamma}$

 (b) The hyperbolas of $\Delta s_{\alpha,\beta}$ and $\Delta s_{\beta,\gamma}$.

 (c) The hyperbolas of $\Delta s_{\alpha,\gamma}$ and $\Delta s_{\beta,\gamma}$.

Figure 14.6: The white diamonds show the positions of the microphones. The blue cross shows the true position of the speaker.

In Figure 12.4 the vectors used for calculating the angles are presented. The angles estimated for the recordings (5–8) of the experiment are given in Table 14.2. The angles that are not presented in the table can not be calculated as a result of the sample delays, and are denoted by the “–”.

Recording	Speaker	$\angle \Delta s_{\alpha,\beta}, \Delta s_{\alpha,\gamma}$ [°]	$\angle \Delta s_{\alpha,\beta}, \Delta s_{\beta,\gamma}$ [°]	$\angle \Delta s_{\alpha,\gamma}, \Delta s_{\beta,\gamma}$ [°]
5	A	—	45.44	—
6	B	80.72	80.68	79.67
7	C	[135.9;142.14]	[135.9;142.14]	[135.9;142.14]
8	D	—	—	—

Table 14.2: Calculated angles.

In order to visualise the results, the calculated angles from Table 14.2 are illustrated in Figure 14.7. The angles span an area in which the speaker is approximated to be.

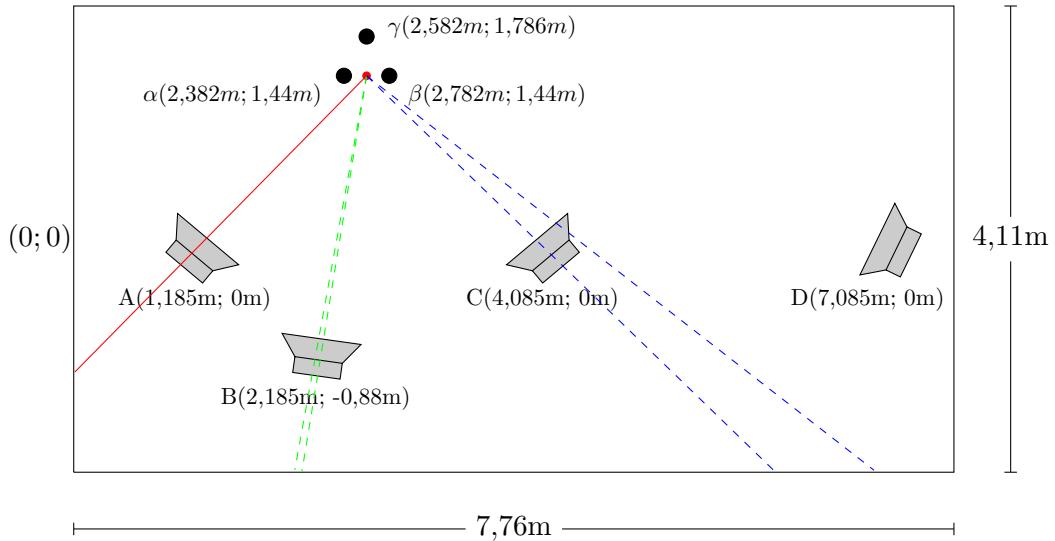


Figure 14.7: The calculated angles formed from the reference point.

In Figure 14.7 the accuracy of the approximated angles are illustrated. The angles used to locate Speaker A, B, and C appear quite precise. Some angles could not be approximated, since the hyperbolas gave no clear indication of where the speaker was located. This is due to the sample delay estimation not yielding sufficiently precise delays. The reasons for this and the precision of the approximated angles are discussed in the subsequent chapter.

14.3 DFT Analysis

14.3.1 Description of the DFT Analysis Procedure

Initially two subsets of the recorded data from each of the three microphones are chosen such that each segment of interest contains $2^{19} = 524,288$ data points. For each microphone, one of the segments $\mathbf{x}_{\text{prior}}$ contains data about the system prior to a fault, while the other segment $\mathbf{x}_{\text{fault}}$ contains information about the system after a fault has occurred, thereby including the fault sound. The two segments for α are referenced as $\mathbf{x}_{\text{prior},\alpha}$ and $\mathbf{x}_{\text{fault},\alpha}$, the notation is likewise for β and γ .

Frequency analyses of the segments are conducted using the DFT. In a real installation in a wind turbine, it is not certain that a fault would be present in two consecutive signals. Assuming a fault has occurred, then it can be detected by comparing the frequencies of two consecutive signals. If a fault is detected in form of a new frequency, then the frequency of the fault needs to be identified.

After applying the DFT to all pairs of $\mathbf{x}_{\text{prior}}$ and $\mathbf{x}_{\text{fault}}$, then the difference $|\mathcal{F}(\mathbf{x}_{\text{fault}})| - |\mathcal{F}(\mathbf{x}_{\text{prior}})|$ is computed for said pairs, and is denoted $\tilde{\mathbf{x}}$. The fault is then identified as the frequency of the largest positive value.

The phase of the identified frequency is computed as a shift in samples. The sample delays d between the microphones are then calculated as the difference of the shift in samples for the three pairs of α , β , and γ . The location of the sound emitting source can then be triangulated based on the sample delays.

14.3.2 Analysis of Frequency Altering

The procedure for the analysis is now conducted in detail for one of the four recordings, which has an altering in frequency as the sound indicating a fault. The conducted analysis is of Recording 2 (A.2), where the sound of the fault is emitted by Speaker B.

A segment $\mathbf{x}_{\text{prior}}$ and a segment $\mathbf{x}_{\text{fault}}$ is chosen for the signal of each of the microphones. This is seen in Figure 14.9.

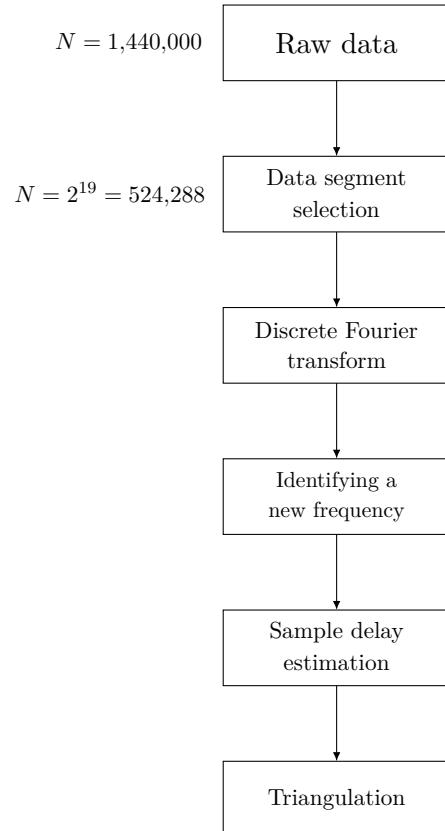


Figure 14.8: Block diagram illustrating the signal processing using the DFT.

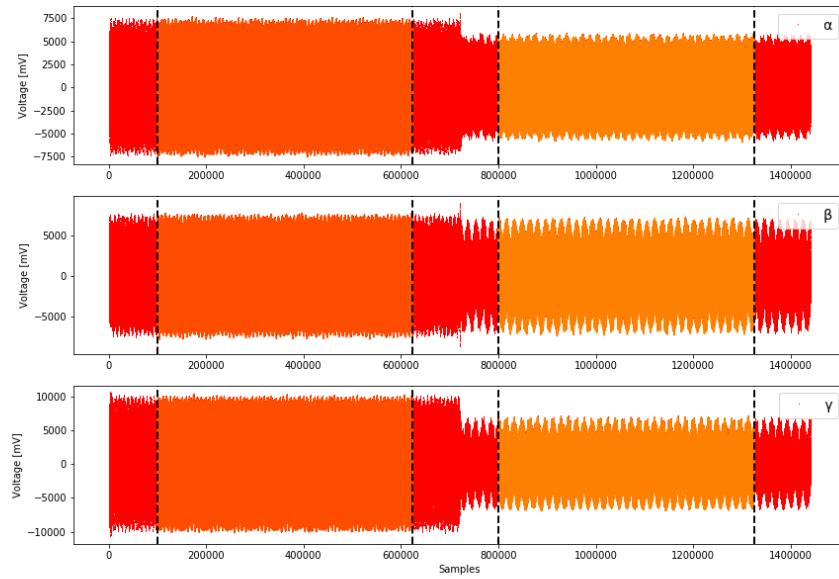


Figure 14.9: Sound signal of microphone α , β , and γ from Recording 2.

It is clear that the two segments are different, thus the emitted sounds are altered at some point between the two segments.

The next step in the analysis is applying the DFT to the chosen segments, which enables a comparison of the frequency content of $\mathbf{x}_{\text{prior}}$ and $\mathbf{x}_{\text{fault}}$ for the signal of each microphone. The amplitude spectra of the transformed signals are seen in Figure 14.10.

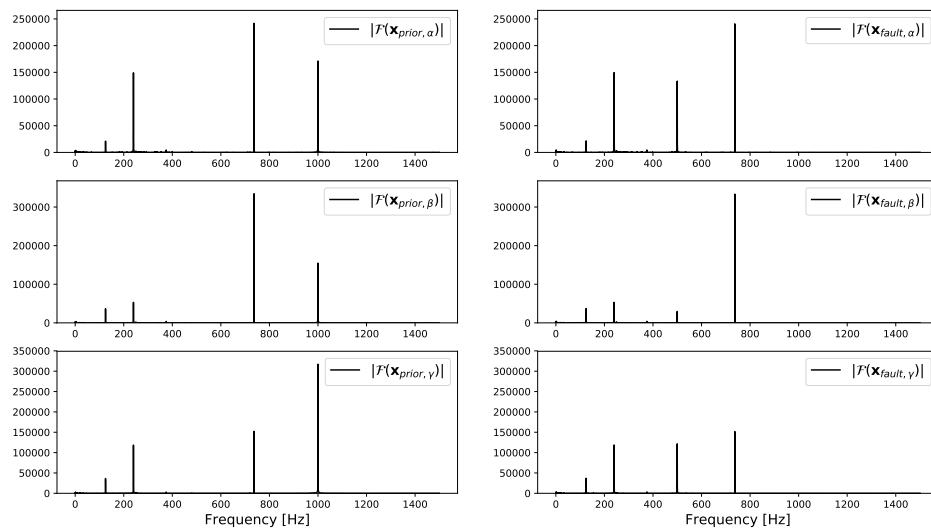


Figure 14.10: The amplitude spectra of segment $\mathbf{x}_{\text{prior}}$ and $\mathbf{x}_{\text{fault}}$ for the three microphones from Recording 2.

The amplitude spectra has four frequencies with notable amplitude, however, the amplitude spectra of $\mathbf{x}_{\text{fault}}$ displays four frequencies, where one of the frequencies is different from the amplitude spectra of $\mathbf{x}_{\text{prior}}$.

The identification of the frequency of the fault is now conducted by computing $|\mathcal{F}(\mathbf{x}_{\text{fault}})| - |\mathcal{F}(\mathbf{x}_{\text{prior}})|$ for each microphone. This can be seen in Figure 14.11.

The frequency of the fault sound is identified as the frequency with the largest positive value, while the frequency with the largest negative value is the sound prior to the fault.

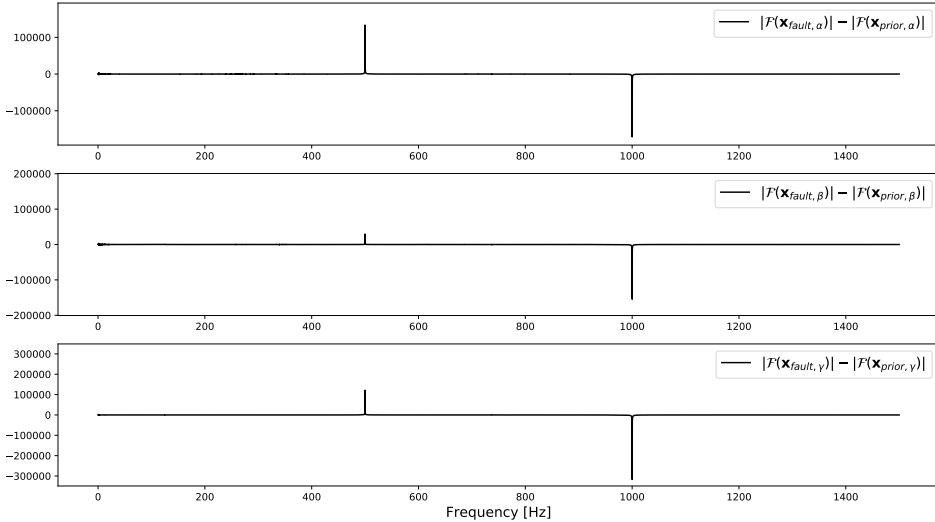


Figure 14.11: The spectrum $|\mathcal{F}(\mathbf{x}_{\text{fault}})| - |\mathcal{F}(\mathbf{x}_{\text{prior}})|$ for microphone α , β , and γ .

The identified frequency of the fault is 500 Hz, which is a change of the emitted sound with an identified frequency of 1000 Hz.

The phases ϕ of the identified frequencies can then be used to calculate a shift in samples for the signals $\tilde{\mathbf{x}}_\alpha$, $\tilde{\mathbf{x}}_\beta$, and $\tilde{\mathbf{x}}_\gamma$, given by

$$d = \frac{\phi}{2\pi f_{\text{fault}}} f_s,$$

where $f_{\text{fault}} = 500$ Hz for Recording 2 and $f_s = 48,000$ Hz.

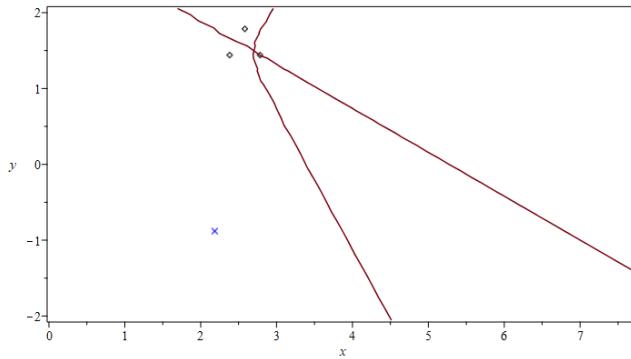
Based on these samples the sample delays in Table 14.3 are computed.

Recording	Speaker	$d_{\tilde{\mathbf{x}}_\alpha, \tilde{\mathbf{x}}_\beta}$	$d_{\tilde{\mathbf{x}}_\alpha, \tilde{\mathbf{x}}_\gamma}$	$d_{\tilde{\mathbf{x}}_\beta, \tilde{\mathbf{x}}_\gamma}$
1	A	-4	-7	-3
2	B	-27	0	27
3	C	50	-16	63
4	D	-27	11	38

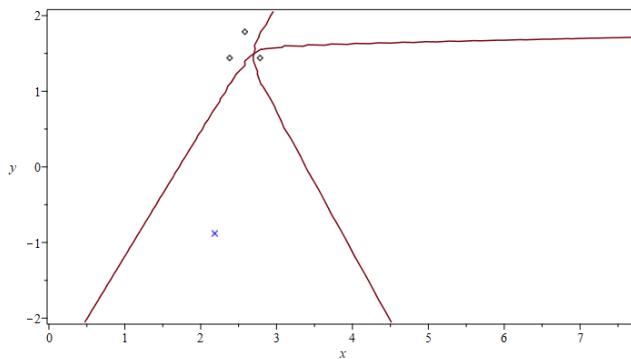
Table 14.3: The sample delays d of Recording 1–4.

From the calculated sample delays, the position of the speaker emitting the fault sound is triangulated. The procedure for the triangulation is outlined in Chapter 12. This is performed by computing the solutions to equations (12.3), (12.4), and (12.5) with the approximated sample delays. The solution sets form hyperolas, and these can be used to

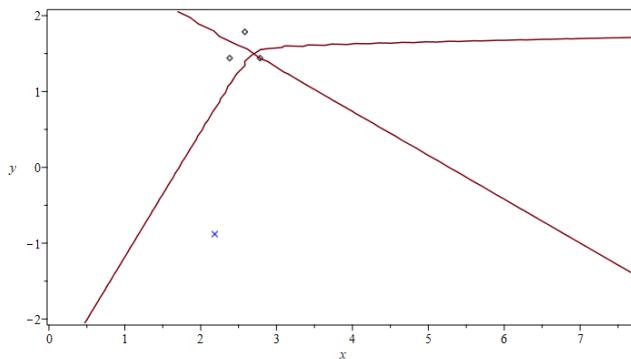
estimate the angle of the speaker relative to the reference point P_R . The three pairs of hyperbolas for Recording 2 are given in Figures 14.12a, 14.12b, and 14.12c.



(a) The hyperbolas of $\Delta s_{\alpha,\beta}$ and $\Delta s_{\alpha,\gamma}$



(b) The hyperbolas of $\Delta s_{\alpha,\beta}$ and $\Delta s_{\beta,\gamma}$.



(c) The hyperbolas of $\Delta s_{\alpha,\gamma}$ and $\Delta s_{\beta,\gamma}$.

Figure 14.12: The white diamonds show the positions of the microphones. The blue cross shows the true position of the speaker.

The intersections of the hyperbolas are used to calculate the angles for locating Speaker B. These intersections clearly yield incorrect angles, since the hyperbolas in Figures 14.12a, 14.12b, and 14.12c do not have any intersections near the true position of Speaker B. The reasons for the inaccurate results are discussed in the following chapter.

15 | Discussion

15.1 Introduction

In this chapter the possible reasons for deviations occurring when comparing the results of the DWT and DFT analyses to the true angles are discussed. Furthermore, it is discussed why some of the angles are approximated inaccurately. Before discussing the results, the sources of error for the experiment are discussed.

During the discussion, the sample delays and angles calculated in the analyses are referred to as the approximated delays and angles. The sample delays and angles based on the measured distances between the equipment are referred to as the true delays and angles.

15.2 Experimental Uncertainties and Errors

During the setup and execution of the experiment, multiple sources of error emerges which can affect the recordings.

The equipment in the experiment is aligned vertically and horizontally with a laser and the distance between the equipment is measured with a measuring tape. As such the coordinates of the equipment are not exact. Photos of the setup can be seen in Appendix A.3.

The speakers are placed on a speaker stand where each speaker is adjusted by hand to a height measured with the laser. The speakers are all set at a preset tilt, which is slightly forwards or backwards for each speaker. Furthermore, the speakers are all rotated around the vertical axis to face the setup of the microphones, which is done without the use of any instruments. These factors can cause a minuscule change to the distance from the speakers to the microphones compared to the expected measured distance.

The microphones are positioned in an equilateral triangle, and they each face a specific direction in the plane as seen in Figures A.3 and A.4. Additionally, the microphones are not perfectly omnidirectional. Therefore, the orientation may have a slight impact on the recorded sound signal.

The reverb also affects the recordings as the reflected sound interferes with the desired signal, especially due to the short distance between the microphones and the wall.

As mentioned in the experimental setup, the sampling frequency affects the precision of the localisation as it dictates the distance between two consecutive samples.

Overall, the effect of these uncertainties and errors are expected to be minimal, however, they are taken into consideration when evaluating the results of the DWT and DFT analyses.

15.3 Results from the DWT Analyses

The approximated sample delays from the DWT analyses are seen alongside the true sample delays in Table 15.1.

Recording	Speaker	Approx.			True		
		$d_{\tilde{\mathbf{x}}_\alpha, \tilde{\mathbf{x}}_\beta}$	$d_{\tilde{\mathbf{x}}_\alpha, \tilde{\mathbf{x}}_\gamma}$	$d_{\tilde{\mathbf{x}}_\beta, \tilde{\mathbf{x}}_\gamma}$	$d_{\tilde{\mathbf{x}}_\alpha, \tilde{\mathbf{x}}_\beta}$	$d_{\tilde{\mathbf{x}}_\alpha, \tilde{\mathbf{x}}_\gamma}$	$d_{\tilde{\mathbf{x}}_\beta, \tilde{\mathbf{x}}_\gamma}$
5	A	38	124	18	38.9	55.3	16.4
6	B	9	52	42	9.4	51.4	42.0
7	C	-40	15	55	-40.3	14.6	54.9
8	D	-704	-10	-18	-53.3	-10.4	42.9

Table 15.1: The approximated and true sample delays d of Recording 5–8.

The approximated angles from the DWT analyses and the true angles of the speakers relative to the reference point P_R are shown in Table 15.2.

Recording	Speaker	Approx. Angle [°]		True Angle [°]
5	A	—	45.44	45.87
6	B	80.72	80.68	80.29
7	C	[135.9;142.14]	[135.9;142.14]	136.23
8	D	—	—	162.27

Table 15.2: Approximated and true angles for the speakers.

The hyperbolas used to calculate the angles can be seen for Recording 5, 7, and 8 in Appendix B and Recording 6 can be seen in Figures 14.6a, 14.6b, and 14.6c.

15.3.1 Successful Localisation

For Speaker A, B, and C the approximated angles are very similar when compared with the true angles. However, the true angles are calculated based on the distance between the equipment in the setup of the experiment, which is measured by hand. The true angles are therefore not exact, but a good reference point for validating the approximated angles.

The approximated angles for Speaker A and B only deviate with at most angles 0.62 degrees from the true angles. The uncertainties of the measurements used to calculate the true angles are unknown, but they may cause a similar deviation. Therefore, it is not possible to determine whether the approximated angles or the true angles are the most precise for Speaker A and B. The speaker emitting the fault sound can be determined accurately by the approximated angles.

The approximated angles for Speaker C are intervals, since the three pairs of hyperbolas all have two intersections. With two intersections there can only be one intersection that is the real location of the speaker. The real location can be determined by using the true angle as a reference point, and thus the real approximated angle for Speaker C is 135.9 degrees, as it barely differs from the true angle. However, the deviation for the angle of the other intersection is 5.93 degrees, which is still accurate enough to determine the correct speaker, which is Speaker C.

15.3.2 Unsuccessful Localisation

For Speaker A it is not possible to approximate angles with any accuracy for $\Delta s_{\alpha,\beta}, \Delta s_{\alpha,\gamma}$ and $\Delta s_{\alpha,\gamma}, \Delta s_{\beta,\gamma}$, since the hyperbolas do not have any intersections. This is due to the

solution set for $\Delta s_{\alpha,\gamma}$ not being defined for $0 \leq x \leq 7.76$ and $-2.055 \leq y \leq 2.055$ (the dimensions of the room). The inaccuracy of $\Delta s_{\alpha,\gamma}$ is due to the sample delay between microphone α and γ , which is seen in Table 15.1. The sample delay is $d_{\alpha,\gamma} = 124$ samples, which is clearly inaccurate since the maximum sample delay between two microphones is $\frac{0.40 \text{ [m]} 48,000 \text{ s}^{-1}}{343 \text{ [m/s]}} \approx 56$ samples.

The inaccurate sample delay $d_{\alpha,\gamma} = 124$ is possibly caused by the reverb of the sound from Speaker A, since the position of microphone γ is close to the wall. The frequency of the impulses and the reverberated impulses is identical, and thus the wavelet packet analysis is not removing the reverberated impulses, which then can disturb the cross-correlation.

It is not possible to approximate any accurate angles for Speaker D from the hyperbolas. This is partly due to the solution set for $\Delta s_{\alpha,\beta}$ not being defined for $0 \leq x \leq 7.76$ and $-2.055 \leq y \leq 2.055$, and due to $\Delta s_{\beta,\gamma}$ being imprecise. The cause is the inaccuracy of sample delay $d_{\tilde{\mathbf{x}}_\alpha, \tilde{\mathbf{x}}_\beta} = -704$ and $d_{\tilde{\mathbf{x}}_\beta, \tilde{\mathbf{x}}_\gamma} = -18$.

A probable reason for the inaccuracy of the results for Speaker A and D is the orientation of the microphones in relation to the speakers. As seen in Appendix A.3 microphone α and γ are facing the same direction, which is towards Speaker D, while β is facing Speaker A. For the recording of Speaker A, the delay $d_{\alpha,\gamma} = 124$ is based on the recordings of microphone α and γ , both of which are oriented at an approximately ninety degree angle relative to Speaker A. This may influence the signal such that the synthesised signals slightly differs in resemblance and thereby disturb the cross-correlation. The same phenomenon may be affecting the recordings for Speaker D, where the inaccurate sample delays are $d_{\beta,\gamma} = -704$ and $d_{\alpha,\beta} = -18$. For these sample delays microphone β is oriented at an approximately ninety degree angle relative to Speaker D.

15.4 Results from the DFT Analysis

The approximated sample delays from the DFT analyses are seen alongside the true sample delays in Table 15.3.

Recording	Speaker	Approx.			True		
		$d_{\tilde{\mathbf{x}}_\alpha, \tilde{\mathbf{x}}_\beta}$	$d_{\tilde{\mathbf{x}}_\alpha, \tilde{\mathbf{x}}_\gamma}$	$d_{\tilde{\mathbf{x}}_\beta, \tilde{\mathbf{x}}_\gamma}$	$d_{\tilde{\mathbf{x}}_\alpha, \tilde{\mathbf{x}}_\beta}$	$d_{\tilde{\mathbf{x}}_\alpha, \tilde{\mathbf{x}}_\gamma}$	$d_{\tilde{\mathbf{x}}_\beta, \tilde{\mathbf{x}}_\gamma}$
1	A	-4	-7	-3	38.9	55.3	16.4
2	B	-27	0	27	9.4	51.4	42.0
3	C	50	-16	63	-40.3	14.6	54.9
4	D	-27	11	38	-53.3	-10.4	42.9

Table 15.3: The approximated and true sample delays d of Recording 1–4.

The hyperbolas for the sample delays can be seen for Recording 1, 3, and 4 in Appendix B and Recording 2 can be seen in Figures 14.12a, 14.12b, and 14.12c.

15.4.1 Errors of the Localisation

The calculated sample delays deviate drastically from the true sample delays, causing the intersection(s) of the hyperbolas to indicate significantly incorrect angles for the speakers. The sample delays for each signal are calculated from the phase shifts of the frequency of the fault sound. It is highly probable, that the phase is affected by the reverberated fault sound, as this causes the same frequency to be recorded multiple times, but with different delays. This is especially likely since the microphones are positioned near the wall, and as a result the impact of the reverberated sounds is greater. For a signal consisting of

multiple sinusoids with the same frequency but unknown delays, the phase information of the DFT of the signal combines the phases and amplitudes, thereby making it impossible to determine the phase and amplitude of each specific sinusoid. This poses a problem, since the phase returned by the DFT is obviously affected by the reverberated fault sound. Thus this is likely the cause of error for the applied method.

16 | Conclusion

This project applies the discrete Wavelet transform and discrete Fourier transform to facilitate the localisation of the speaker emitting a fault sound in an experiment. The experiment simulates a mechanical fault occurring in a wind turbine nacelle. The analyses of the recordings are conducted by applying developed numerical software.

Three recorded signals are filtered and denoised using the discrete Wavelet transform. The filtered signals are cross-correlated in order to estimate time delays. Then the estimated time delays are used to triangulate the position of the faulty component emitting an impulse sound.

This project shows that the discrete Wavelet transform can facilitate the localisation of potential mechanical failures in wind turbine machinery. However, the discrete Wavelet transform analyses prove not to be consistent, as one of the speakers can not be located.

Likewise, a localisation using the discrete Fourier transform is attempted. Three recorded signals are analysed using the discrete Fourier transform. More specifically, the signals are analysed prior to, and after a fault has occurred, thereby yielding the frequency content of the two segments for each signal. Subsequently, the two amplitude spectra of each signal are compared in order to detect a possible fault. If a fault has occurred the frequency of the fault sound is determined.

The project concludes that, by applying the discrete Fourier transform, it is possible to detect an error in mechanical machinery, which is indicated by a change in frequency. Nevertheless, the obtained phase information is supposedly incorrect due to reverberation, and thus the components can not be located.

Bibliography

- Beerends, R.J.; Morsche, H. v. d. B. J. v. d. V. E. (2003). *Fourier and Laplace Transforms*. Cambridge University Press, 1st edition.
- Chitode, J. (2009). *Digital Signal Processing*. Technical Publications Pune, 3rd edition.
- CWIF (2018). Summary of wind turbine accident data to 31 december 2018. <http://www.caithnesswindfarms.co.uk/accidents.pdf>. [Online; accessed 02-May-2019].
- Dr. Shrivastava, R. P. (2017). *Principles of Digital Signal Processing*. VIKAS Publishing House.
- Feuer, A. (1996). *Sampling in Digital Signal Processing and Control*. Birkhäuser.
- Fraden, J. (2016). *Handbook of Modern Sensors*. Springer, 5th edition. [Online; accessed 22-May-2019].
- GL, D. (2017). Definitions of availability terms for the wind industry. <https://www.ourenergypolicy.org/wp-content/uploads/2017/08/Definitions-of-availability-terms-for-the-wind-industry-white-paper-09-08-2017.pdf>. [Online; accessed 02-April-2019].
- GWEC (2017). Global wind report - annual market update 2017. [Online; accessed 03-April-2019].
- IRENA (2018). Renewable power generation cost in 2017. https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2018/Jan/IRENA_2017_Power_Costs_2018.pdf. [Online; accessed 04-April-2019].
- Lehto, P. (2014). Signal analysis of wind turbine acoustic noise. https://aaltodoc.aalto.fi/bitstream/handle/123456789/14832/master_Lehto_Panu_2014.pdf?sequence=2&isAllowed=y. [Online; accessed 22-May-2019].
- NASA (2009). The causes of climate change. <https://climate.nasa.gov/causes/>. [Online; accessed 25-April-2019].
- Oppenheim, A. V. (1989). *Discrete-Time Signal Processing*. Prentice-hall, inc.
- Qiao, Wei; Lu, D. (2015). A survey on wind turbine condition monitoring and fault diagnosis. <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7084135>. [Online; accessed 08-April-2019].
- Rabbath, C.A.; Léchevin, N. (2013). *Discrete-Time Control System Design with Applications*. Springer.

- S. Faulstich, Berthold Hahn, P. T. (2011). Wind turbine downtime and its importance for offshore deployment. https://www.researchgate.net/publication/48776663_Wind_turbine_downtime_and_its_importance_for_offshore_deployment. [Online; accessed 23-April-2019].
- Spence E., Lawrence; Insel J., A. F. H. S. (2008). *Elementary Linear Algebra*. Pearson, 2nd edition.
- Strang, Gilbert; Nguyen, T. (1996). *Wavelets and Filter Banks*. Wellesley - Cambridge Press.
- Sundararajan, D. (2016). *Discrete Wavelet Transform: A Signal Processing Approach*. John Wiley & Sons, Incorporated.
- Vujicic, M. (2008). *Linear Algebra Thoroughly Explained*. Springer.

Appendices

A | Experiment

A.1 Equipment

1. 4 x Nokalon Trawlnet Ball (34758) (Speaker)
2. Fireface UCX + Frontend (Audio interface)
3. 3 x GRAS 26CC 1/4" CCP Standard Preamplifier with SMB Connector (Amplifier)
4. 3 x GRAS 40AZ 1/2" Prepolarized Free-field Microphone, Low Frequency (Microphone)
5. Stanley CL-E (Laser)

A.2 Table of Conducted Recordings

Recording 1 - Frequency altering				
Speaker	Sound emitted (Hz)	Time interval (s)	Fault sound (Hz)	Time interval (s)
A	125+ noise	0-15	62,5 + noise	15-30
B	1000 + noise	0-30		
C	737 + noise	0-30		
D	240 + noise	0-30		

Recording 2 - Frequency altering				
Speaker	Sound emitted (Hz)	Time interval (s)	Fault sound (Hz)	Time interval (s)
A	125+ noise	0-30		
B	1000 + noise	0-15	500 + noise	15-30
C	737 + noise	0-30		
D	240 + noise	0-30		

Recording 3 - Frequency altering				
Speaker	Sound emitted (Hz)	Time interval (s)	Fault sound (Hz)	Time interval (s)
A	125+ noise	0-30		
B	1000 + noise	0-30		
C	737 + noise	0-15	368,5 + noise	15-30
D	240 + noise	0-30		

Recording 4 - Frequency altering				
Speaker	Sound emitted (Hz)	Time interval (s)	Fault sound (Hz)	Time interval (s)
A	125 + noise	0-30		
B	1000 + noise	0-30		
C	737 + noise	0-30		
D	240 + noise	0-15	480 + noise	15-30

Recording 5 - Impulse				
Speaker	Sound emitted (Hz)	Time interval (s)	Fault sound (Hz)	Time interval (s)
A	125+ noise	0-30	5 + noise	15-30
B	1000 + noise	0-30		
C	737 + noise	0-30		
D	240 + noise	0-30		
Recording 6 - Impulse				
Speaker	Sound emitted (Hz)	Time interval (s)	Fault sound (Hz)	Time interval (s)
A	125+ noise	0-30		
B	1000 + noise	0-30	5 + noise	15-30
C	737 + noise	0-30		
D	240 + noise	0-30		
Recording 7 - Impulse				
Speaker	Sound emitted (Hz)	Time interval (s)	Fault sound (Hz)	Time interval (s)
A	125+ noise	0-30		
B	1000 + noise	0-30		
C	737 + noise	0-30	5 + noise	15-30
D	240 + noise	0-30		
Recording 8 - Impulse				
Speaker	Sound emitted (Hz)	Time interval (s)	Fault sound (Hz)	Time interval (s)
A	125+ noise	0-30		
B	1000 + noise	0-30		
C	737 + noise	0-30		
D	240 + noise	0-30	5 + noise	15-30

A.3 Practical Setup



Figure A.1: Setup of the experiment.



Figure A.2: Setup of the experiment.



Figure A.3: Setup of the microphones.



Figure A.4: Setup of the microphones.

B | Hyperbolas from Data Processing

B.1 Recording 1

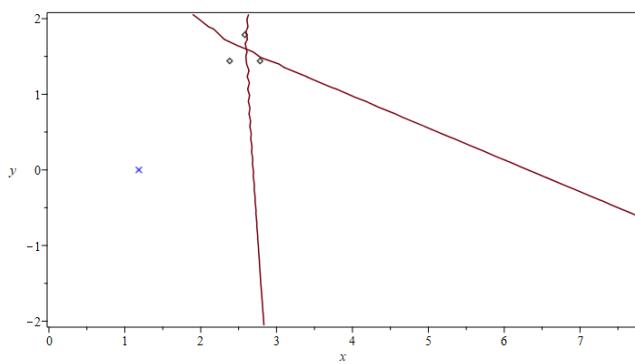


Figure B.1: The hyperbolas of $\Delta s_{\alpha,\beta}$ and $\Delta s_{\alpha,\gamma}$.

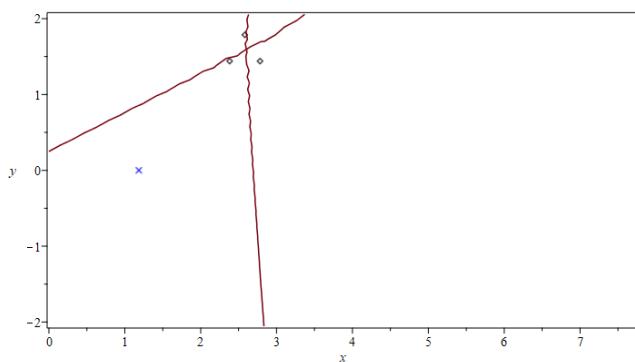
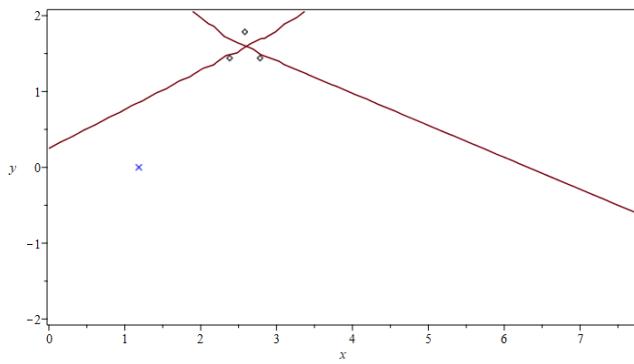
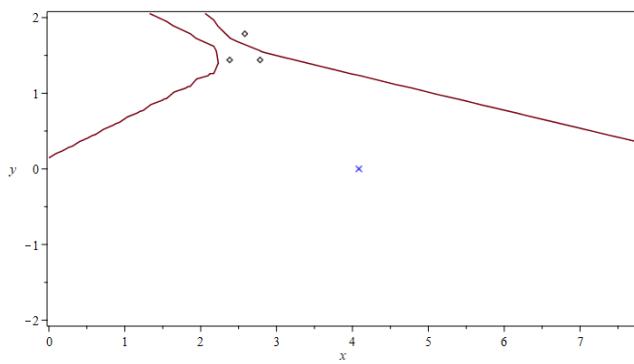
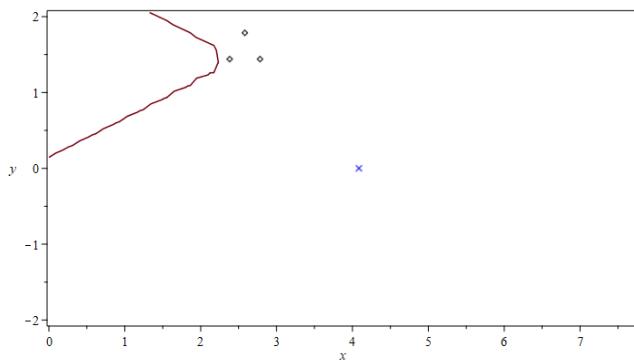
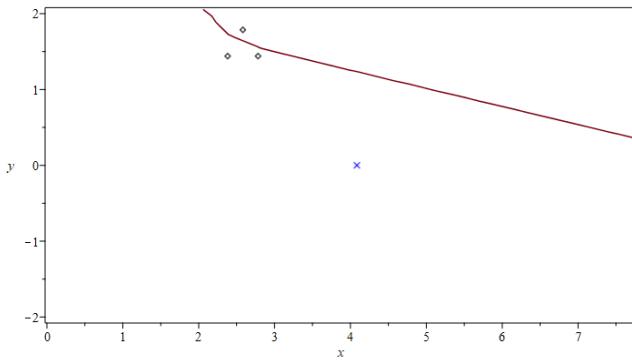


Figure B.2: The hyperbolas of $\Delta s_{\alpha,\beta}$ and $\Delta s_{\beta,\gamma}$.

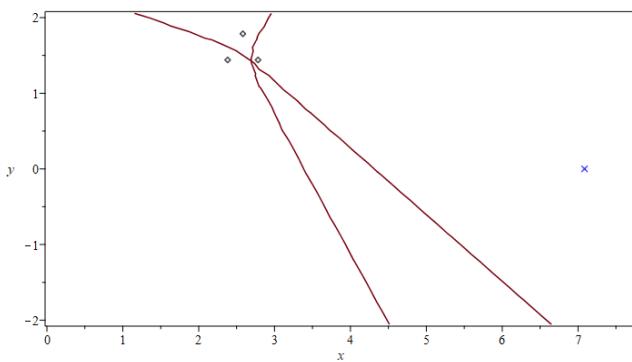
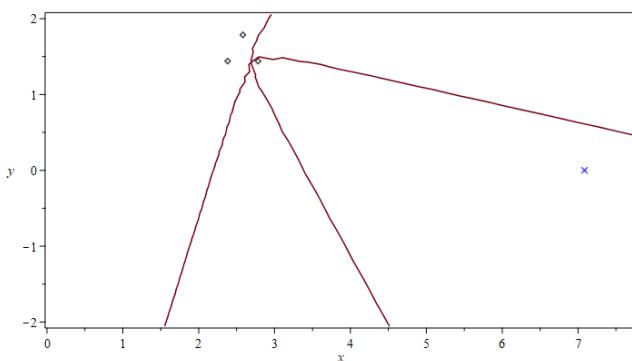
Figure B.3: The hyperbolas of $\Delta s_{\alpha,\gamma}$ and $\Delta s_{\beta,\gamma}$.

B.2 Recording 3

Figure B.4: The hyperbolas of $\Delta s_{\alpha,\beta}$ and $\Delta s_{\alpha,\gamma}$.Figure B.5: The hyperbolas of $\Delta s_{\alpha,\beta}$ and $\Delta s_{\beta,\gamma}$.

Figure B.6: The hyperbolas of $\Delta s_{\alpha,\gamma}$ and $\Delta s_{\beta,\gamma}$.

B.3 Recording 4

Figure B.7: The hyperbolas of $\Delta s_{\alpha,\beta}$ and $\Delta s_{\alpha,\gamma}$.Figure B.8: The hyperbolas of $\Delta s_{\alpha,\beta}$ and $\Delta s_{\beta,\gamma}$.

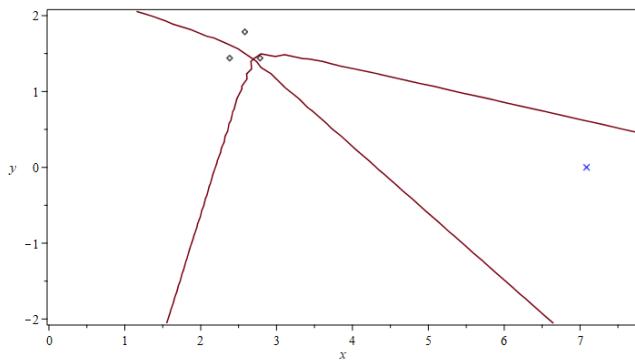


Figure B.9: The hyperbolas of $\Delta s_{\alpha,\gamma}$ and $\Delta s_{\beta,\gamma}$.

B.4 Recording 5

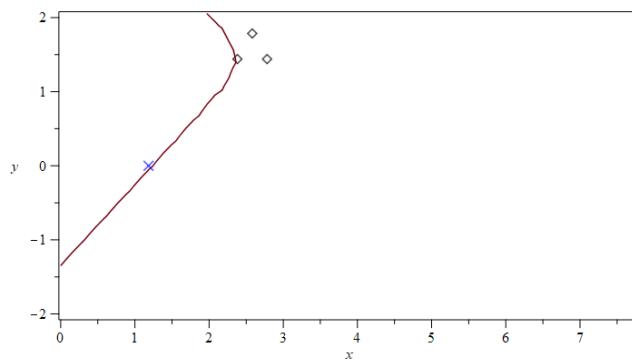


Figure B.10: The hyperbolas of $\Delta s_{\alpha,\beta}$ and $\Delta s_{\alpha,\gamma}$.

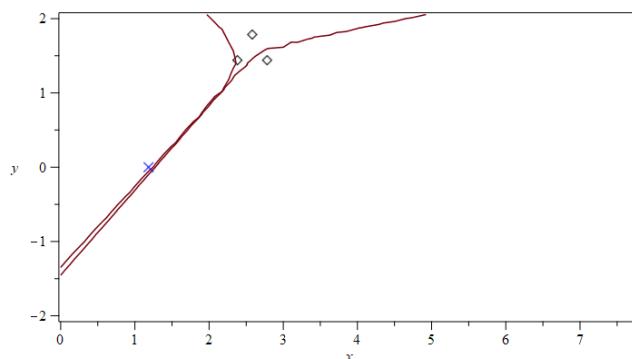
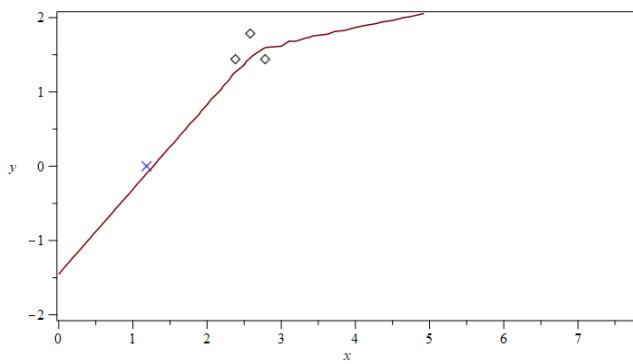
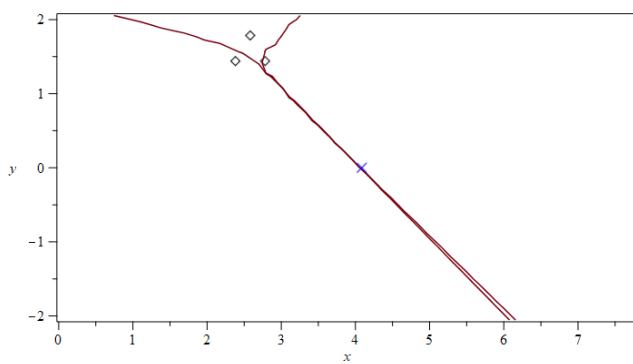
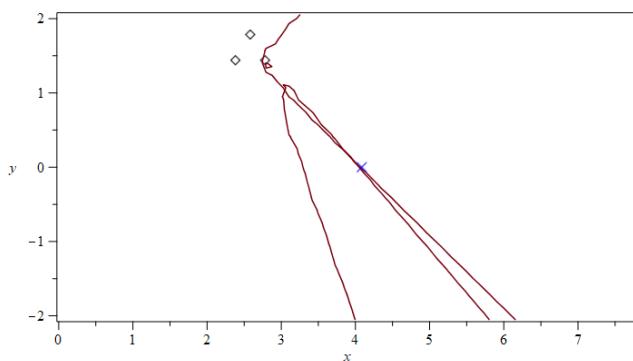


Figure B.11: The hyperbolas of $\Delta s_{\alpha,\beta}$ and $\Delta s_{\beta,\gamma}$.

Figure B.12: The hyperbolas of $\Delta s_{\alpha,\gamma}$ and $\Delta s_{\beta,\gamma}$.

B.5 Recording 7

Figure B.13: The hyperbolas of $\Delta s_{\alpha,\beta}$ and $\Delta s_{\alpha,\gamma}$.Figure B.14: The hyperbolas of $\Delta s_{\alpha,\beta}$ and $\Delta s_{\beta,\gamma}$.

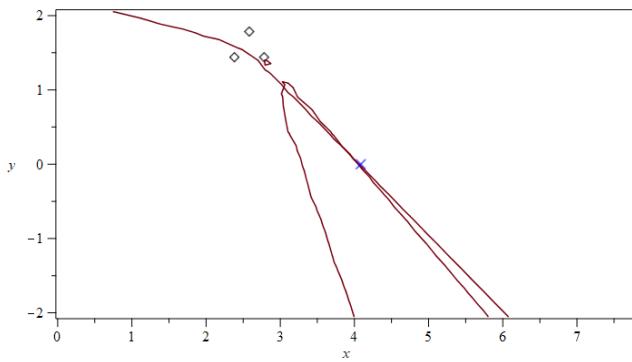


Figure B.15: The hyperbolas of $\Delta s_{\alpha,\gamma}$ and $\Delta s_{\beta,\gamma}$.

B.6 Recording 8

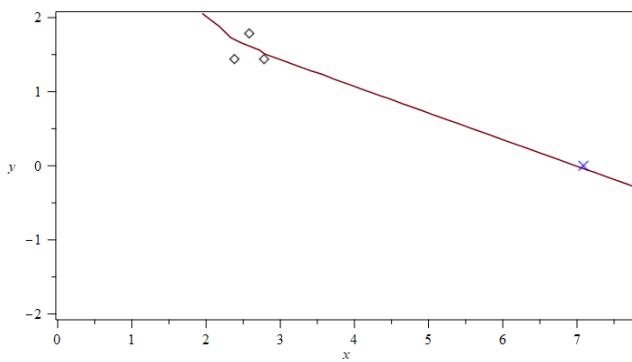


Figure B.16: The hyperbolas of $\Delta s_{\alpha,\beta}$ and $\Delta s_{\alpha,\gamma}$.

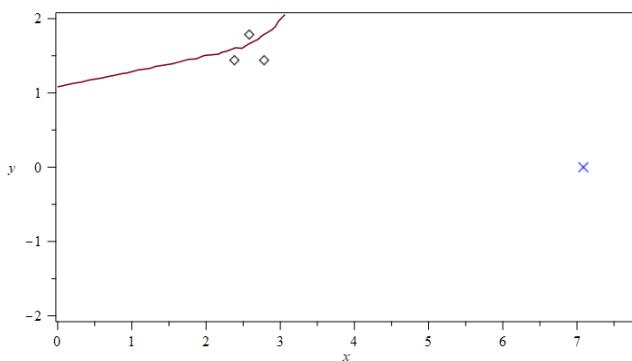


Figure B.17: The hyperbolas of $\Delta s_{\alpha,\beta}$ and $\Delta s_{\beta,\gamma}$.

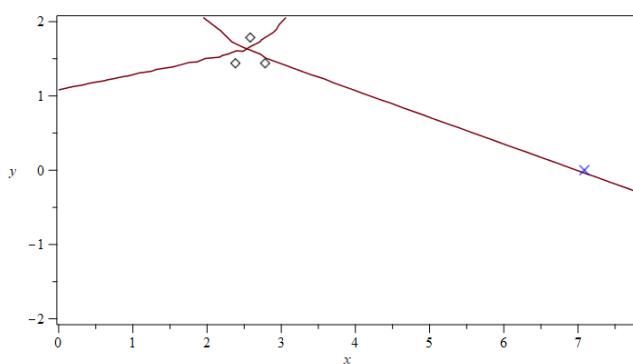


Figure B.18: The hyperbolas of $\Delta s_{\alpha,\gamma}$ and $\Delta s_{\beta,\gamma}$.