ESTIMATION AND CONTROL OF PAPER MACHINE VARIABLES USING WAVELET PACKET ANALYSIS

by

JOHN BYUNG-KYU CHUN

B.A.Sc.(E.E.), University of British Columbia, 1995

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

March 1998

© John Chun, 1998
In presenting this thesis in partial fulfillment of the requirements of an advanced
degree at the University of British Columbia, I agree that the library shall make it free
available for reference and study. I further agree that permission for extensive copying of
this thesis for scholarly purposes may be granted by the Head of my department or by his
or her representatives. It is understood that copying or publication of this thesis for finan­
cial shall not be allowed without my written permission.

Department of Electrical and Computer Engineering

The University of British Columbia
2356 Main Mall
Vancouver, BC
V6T 1Z4 Canada

Date: March 1998
Abstract

The thesis discusses a new method in paper machine signal analysis using the wavelet packet transformation. It has been shown that this technique can be applied to three main areas in paper machine control such as the paper machine data analysis including estimation and compression, control of basis weight or moisture variations, and control performance monitoring. Compared to the traditional windowed Fourier transform, other filtering techniques, and wavelet transform, this method offers greater flexibility in analyzing signals with suitable time-frequency resolution. The Best Basis Algorithm has been introduced to obtain a best decomposition tree structure in decomposition of a given signal.

The paper machine model for use in CD control has been derived from experimental bump test data. It has been clearly shown that the new method along with wavelet estimator produces results superior to standard analysis methods by reducing the error of estimation. In estimating more complex signals, however, the wavelet packet method is superior than the wavelet analysis since the best decomposition tree structure is chosen enabling the estimation to be more precise in the regions where frequency changes. When applied to industrial data, the wavelet packet estimation has provided more visual results by distinguishing the bump locations more evidently. Moreover, it has been shown that wavelet packet compression is more efficient than the wavelet case since it has achieved more compression ratio by yielding the combination of more zero coefficients and lesser mean square error.

The thesis also presents the control scheme of the paper machine as well as an efficient technique to monitor the control performance on-line using both wavelets and wavelet packets. It has been shown that the calculation of performance index using the wavelet packet method is more precise since it offers finer frequency resolution.
Table of Contents

Abstract .................................................................................................................. ii

List of Table .......................................................................................................... vi

List of Figures ......................................................................................................... vii

Acknowledgments .................................................................................................. x

1 Introduction ......................................................................................................... 1

1.1 Overview of Estimation Methods ........................................................................ 3

1.2 Motivation ......................................................................................................... 4

1.3 Outline of Thesis ............................................................................................... 5

2 Paper Machine Model for Use in CD Control ................................................... 6

2.1 Modeling of a Paper Machine ........................................................................... 6

2.2 Identification .................................................................................................... 9

2.2.1 Estimating the Interaction Matrix $G$ ......................................................... 11

2.2.2 Identification of Dynamic and Time Delay ............................................... 12

3 Estimation Theory ............................................................................................... 13

3.1 Brief History of Wavelet Theory ....................................................................... 13

3.2 Wavelet Analysis ............................................................................................. 14

3.2.1 Wavelet Bases ............................................................................................ 16

3.2.2 Properties .................................................................................................. 18

3.2.3 Discrete Wavelet transform and Multi-Resolution Analysis ....................... 19

3.3 Wavelet Packets .............................................................................................. 21

3.3.1 Wavelet Packet Functions .......................................................................... 21
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.3.2 The Best Basis Algorithm</td>
<td>23</td>
</tr>
<tr>
<td>3.4 Two-Dimensional Analysis</td>
<td>27</td>
</tr>
<tr>
<td>3.5 Some Practical Issues</td>
<td>27</td>
</tr>
<tr>
<td>3.6 De-Noising</td>
<td>29</td>
</tr>
<tr>
<td>3.6.1 Threshold Rules</td>
<td>30</td>
</tr>
<tr>
<td>3.6.1.1 Threshold Selection Methods</td>
<td>31</td>
</tr>
<tr>
<td>3.7 Data Compression</td>
<td>32</td>
</tr>
<tr>
<td>4 Data Analysis</td>
<td>34</td>
</tr>
<tr>
<td>4.1 Estimation of Simulated Data</td>
<td>35</td>
</tr>
<tr>
<td>4.2 Estimation of Two-Dimensional Paper Machine Data</td>
<td>42</td>
</tr>
<tr>
<td>4.2.1 Smooth Waves</td>
<td>42</td>
</tr>
<tr>
<td>4.2.1.1 Estimation</td>
<td>44</td>
</tr>
<tr>
<td>4.2.2 Signal Containing Several Frequency Components</td>
<td>49</td>
</tr>
<tr>
<td>4.2.2.1 Estimation</td>
<td>50</td>
</tr>
<tr>
<td>4.2.3 Signal with Bumps and a Streak</td>
<td>55</td>
</tr>
<tr>
<td>4.2.3.1 Estimation</td>
<td>56</td>
</tr>
<tr>
<td>4.2.4 Industrial Data Analysis</td>
<td>61</td>
</tr>
<tr>
<td>4.3 Data Compression</td>
<td>66</td>
</tr>
<tr>
<td>5 CD Control of Basis Weight</td>
<td>69</td>
</tr>
<tr>
<td>5.1 Dahlin Controller</td>
<td>69</td>
</tr>
<tr>
<td>5.2 CD Profile Assessment</td>
<td>71</td>
</tr>
<tr>
<td>5.2.1 Performance Index</td>
<td>73</td>
</tr>
<tr>
<td>5.2.1.1 Frequency Characteristics of Wavelet and Wavelet Packet</td>
<td>73</td>
</tr>
</tbody>
</table>
5.3 CD Basis Weight Control ................................................................. 77

6 Conclusion .................................................................................. 80
References ...................................................................................... 83
Appendix A. Gram Polynomial Estimation ..................................... 87
List of Tables

3.1 Wavelet families and associated properties ........................................... 17
3.2 Minimax thresholds for various sample size $n$ (Donoho and Johnstone [17]) .......... 32
4.1 Best MSE of four signal using different methods ........................................... 36
4.2 MSE with the best option sets of wavelets .................................................... 36
4.3 MSE with the best option sets of wavelet packets ......................................... 36
4.4 MSE for the wavelet estimation of the smooth waves ...................................... 45
4.5 MSE for the wavelet packet estimation of the smooth waves .............................. 45
4.6 MSE with best option sets for wavelet estimation (signal containing several frequency components) ................................................................. 52
4.7 MSE with best option sets for wavelet packet estimation (signal containing several frequency components) ................................................................. 52
4.8 MSE with best option sets for the wavelet estimation (signal with bumps and a streak) .... 58
4.9 MSE with best option sets for the wavelet packet estimation (signal with bumps and a streak) 58
4.10 Data compression results ........................................................................... 68
List of Figures

1.1 A schematic of a paper machine ................................................................. 1

1.2 Scanning path on stationary sheet .................................................................. 2

1.3 An example of Gram-polynomial estimation: the solid line is the basis weight and the
dashed line is the estimated ............................................................................... 4

2.1 Bump responses of different paper quality ....................................................... 9

2.2 (a) up: Actuator set point changes (b) bottom: bump responses data .......... 10

2.3 (a) Aligned bump responses (b) CD actuator position and averaged bump response ...... 11

2.4 Estimated first order dynamic response .......................................................... 12

3.1 (a) Gaussian window, (b) Time-frequency resolution of the STFT (c) Time-frequency
resolution of the wavelet analysis, (d) Randomly chosen time-frequency resolution of the
wavelet packet analysis. (Each rectangle corresponds to the localization of a basis
function.) .............................................................................................................. 15

3.2 Haar mother wavelet and its translated and scaled version ............................. 16

3.3 Discrete wavelet families. The filter coefficients are used to generate decomposition (or
reconstruction) highpass (or lowpass) filters ...................................................... 17

3.4 Block diagram of one-dimensional DWT ......................................................... 20

3.5 Wavelet decomposition tree ........................................................................... 22

3.6 Wavelet coefficients indiscrete and continuous transform .............................. 22

3.7 The Haar wavelet packets ............................................................................. 23

3.8 Wavelet packet decomposition of sine with jumps .......................................... 26

3.9 Block diagram of two-dimensional DWT ......................................................... 28

3.10 Decomposition using DWT and DWPT of two-dimensional image ............... 28
3.11 Data analysis using wavelets ................................................................. 30
3.12 Hard and soft thresholding with $\lambda = 0.5$ ..................................................... 31
3.13 Block diagram of compression and decompression schemes ......................... 33
4.1 One-dimensional signals to be estimated. (a) Bumps (b) Pure sine (c) Signal containing
different frequency components (d) Blocks ..................................................... 35
4.2 Gram-polynomial estimation ....................................................................... 37
4.3 Wavelet and wavelet packet estimation .......................................................... 39
4.4 Wavelet tree structure for signal c ............................................................... 39
4.5 First 50 low MSE resulted using wavelets ...................................................... 40
4.6 First 50 low MSE resulted using wavelet packets ............................................ 41
4.7 Initial actuator settings and resulting data set (Smooth waves) ......................... 43
4.8 Generated paper machine basis weight data .................................................. 43
4.9 MSE for estimation using different mother wavelets ........................................ 46
4.10 The smooth wave estimation profiles ........................................................... 47
4.11 The smooth wave estimation profiles (Images) .............................................. 48
4.12 Cumulative error along CD for the smooth wave ........................................... 49
4.13 Initial actuator settings and resulting data set (Signal containing several different frequency
components) ....................................................................................................... 51
4.14 Generated paper machine basis weight data .................................................. 51
4.15 The estimated profiles of the signal containing several frequency components .... 53
4.16 The estimated profiles of the signal containing several frequency components (Images) .... 54
4.17 Cumulative error along CD for the signal containing several frequency components ...... 55
4.18 Initial actuator settings and resulting data set (Signal with bumps and a streak) ...... 57
Acknowledgments

I would like to thank my thesis supervisor Professor Micael S. Davies for his kind support and guidance during my research. I also would like to thank Professor G. A. Dumont and friends in Pulp and Paper Center as well as in Electrical Engineering Department, in particular, Ming Zhang, Lahoucine Ettaleb, Roger Shirt, Huiping Chen.

I also thank Jahan Ghosfraniha and Zoran Nesic for their sharing of valuable knowledge. Thanks to Tae-Kyung Kim, Sinclair Chi, Chris Kim, and Peter Kim for their friendship and encouragements.

Special thanks to my parents who endeavored with me and helped me in everything during my research. Finally, I thank my God.
CHAPTER 1

Introduction

A paper machine, with several stages, continuously produces a sheet of paper with a width up to 10 m wide and at a speed of up to 90 km/h [37]. The mixture of about 5% fibre and 95% water by weight is ejected from the headbox unto the moving wire as uniformly as possible. It passes over the moving wire, through the press section, and to the dryer section. The mixture becomes paper as water is removed from the fibre. At the dry end, where a sheet is wound onto the reel, the water to fibre ratio will be reversed to roughly 5% moisture and 95% fibre. A simplified schematic of a paper machine is shown in Figure 1.1.
A traversing sensor mounted on an O-frame measures three sheet properties: the basis weight defined as weight per unit area, the moisture content defined as percent water by weight, and the thickness of the manufactured paper or caliper. In this thesis, the measurement always refers to dry basis weight. As shown in Figure 1.1, the sensor moves to and fro over the rapidly moving sheet and in each scan collects up to 3000 points of data across the sheet in about 20 seconds. As a result of the sensor motion across the moving sheet, measurements are taken at points that would be traced in a zig-zag pattern on the stationary sheet (See Figure 1.2). The actual pattern of points would make an angle of $1^\circ$ with the sheet edge.

Burkhard and Wrist classified basis weight variations into three components, machine direction (MD), cross direction (CD), and random residual variations (R) [5]. The MD component, varying with respect to time, is the variation in the machine direction. It depends mainly on the headbox consistency and pressure changes, which influences the entire width of the sheet. The CD component is the variation in the cross machine direction, assumed to be time-invariant, and changes slower than the MD variation. The final component, R, contains measurement errors. For CD control, the CD profile is extracted from the raw measurement and it is used to manipulate the actuator positions accordingly, which ideally make a uniform distribution of basis weight across the sheet. The scan average of MD samples during each scan is subtracted from the raw data to leave zero mean CD profiles and noise. This method is called scan...
average estimation. [28].

1.1 Overview of Estimation Methods

The computer control of paper machine requires an efficient estimation method and a proper control scheme. In industrial mills, the most common estimation method is an exponential filtering (EXPO) technique that can be illustrated as follows [11].

\[
y^n(t) = (1 - a)y^n(t-1) + ax^n(t)
\]

(1.1)

where \(x^n(t)\) is the zero mean raw measurement and \(y^n(t)\) is the filtered CD profile obtained at CD position \(n\) during scan \(t\) and \(a\) is the weighting factor, \(0 < a < 1\). The MD profile can be extracted by averaging \(y^n(t)\) over one scan period. Generally, the separation of the CD and MD using this method is slow due to the low value of \(a\) that is used. For example, if \(a = 0.2\) is used, 20% of the most recent measurement will be added to 80% of the historical value to obtain the trended measurement [36]. Consequently, it would take 10 scans to register 90% of a step-wise profile change.

More advanced filtering techniques were also introduced by Chen [6] and Goodwin and Sin [20]. Algorithms that use a modified least-squares parameter identifier for CD estimation and a Kalman filter for MD estimation are presented by Dumont [18], Wang [37], and Natarajan [31].

Modeling profiles with discrete orthogonal polynomials has been used in Sendzimir steel rolling mills [19]. More recently, Kristinsson also presented an estimation method using Gram polynomials, which are defined discrete equidistant orthogonal polynomials with equal weighting on every data points [25]. An example in Figure 1.3 shows the estimation of industrial paper basis weight data using Gram-polynomials. The data used in the estimation is CD profile and noise after the scan average of 45.19 \(g/m^2\) is subtracted from the raw profile. The details of this method are given in the Appendix A. In this thesis,
the EXPO method and the Gram polynomial method are adapted in order to compare them with the wavelet and wavelet packet methods.

\[ \text{Estimation of basis weight using Gram-polynomial} \]

![Graph showing basis weight and estimate](image)

Figure 1.3 An example of Gram-polynomial estimation: the solid line is the basis weight and the dashed line is the estimate.

### 1.2 Motivation

On a paper machine, the CD control provides benefits such as increased production, improved paper quality, reduction of raw material usage, and a decreased number of sheet breaks [5][10][25]. The development of computer and networking technologies with fast computation have enabled more powerful estimation systems to be set on-line with sophisticated control loops. In addition, as higher resolution sensors appeared in the market place, measurements of paper qualities with shorter sampling intervals have become available. As a result, the need for more precise estimation and control methods has become evident. On the other hand, data compression becomes more necessary in order to store the large amount of industrial data for future analysis. The performance of the control loops has to be monitored on a regular basis so that an operator acquires suitable information and is able to react to a sudden change of manufacturing conditions effectively. The main objectives of this thesis are first to develop a novel method to
estimate CD components of the paper basis weight from noisy raw data, second to study the control of a paper machine using the identified paper machine model, and finally to create an efficient technique to monitor the control performance on-line.

1.3 Outline of Thesis

In Chapter 2, a paper machine model is developed and then parameters of the model are identified from the actual input-output measurements. Estimation theory using wavelets and wavelet packets are presented in chapter 3. Also in this chapter, the noise reduction procedure and the compression method will be illustrated. Chapter 4 presents the error analysis of simulated data with different methods such as the EXPO analysis, wavelet estimation, and wavelet packet estimation. The estimation and compression of industrial basis weight data will also be displayed. Chapter 5 presents the control scheme of the paper machine as well as the CD profile performance assessment methods using both wavelets and wavelet packets. Finally, conclusions are drawn in chapter 6.
CHAPTER 2

Paper Machine Model for Use in CD Control

A CD basis weight control system uses the measured profile read at the sensor to manipulate the slice lip actuator settings. The computer analyzes the scanned data, calculates proper control actions, and then generates the actuator set point changes accordingly. The control simulations in later chapters are carried out using the paper machine model developed in this section.

2.1 Modeling of a Paper Machine

The paper machine model which gives a relationship between the actuator settings and the CD basis weight responses can be derived mathematically using experimental bump test data. The discrete-time mathematical model for a paper machine includes actuator dynamics, time-delay, and interactions [25]. The major assumptions made in modeling the actuator response are

(i) The responses to all actuators have the same dynamics (Wilhelm and Field, 1983)

(ii) The response shape is symmetric.

(iii) The total profile response is the sum of the responses to each other.

(iv) A sampling time is known as $T_s$.

Usually, all actuators are represented by the same first order lag dynamic with the gain $k_a$ and the time constant $p_a$ as follows

$$\frac{k_a}{(1 - p_a q^{-1})}$$ (2.1)

Since the basis weight scans are taken some distance down the machine direction from the actuators, the
model should include a time-delay $q^{-d}$ corresponding to $d \cdot T_s$. It is assumed that the speed of a paper machine is constant as is the time delay.

Movement of an actuator causes the basis weight to change along the profile. This profile response spreads to the location of neighboring actuators resulting a coupling between actuator spaces. Therefore, assuming the same spatial response for all actuators, a paper machine should be modeled by a banded diagonal matrix called the interaction matrix $G$ as shown in equation (2.2). Except near edges of the matrix, each column consists of zeros and the bump response data streams $[p_k \ldots p_1, p_0, p_1, \ldots p_k]$. The height of the matrix is the number of measurements, and the width of the matrix is equal to the number of actuators.

$$G = \begin{bmatrix}
p_0 & p_1 & \cdots & p_k & 0 & \cdots & 0 \\
p_1 & p_0 & p_1 & \cdots & \cdots & \cdots & \cdots \\
\vdots & p_1 & \cdots & \cdots & \cdots & \cdots & 0 \\
p_k & \cdots & \cdots & \cdots & p_k & \cdots & \cdots \\
0 & \cdots & \cdots & p_1 & \cdots & \cdots & \cdots \\
\cdots & \cdots & p_1 & p_0 & p_1 & \cdots & \cdots \\
0 & \cdots & 0 & p_k & \cdots & p_1 & p_0
\end{bmatrix}$$

(2.2)

As discussed in [25] and [34], some problems exist in building the interaction matrix $G$. The interaction matrix should be constructed to be nonsingular, otherwise the condition number will grow infinite and it can lead to a poor performance of CD control. The condition number is defined as the measure of the severity of expected control difficulty [26]. More specifically, Kristinn defines the condition number of $G$ as the ratio between its largest and smallest singular value [25]. Singular values are obtained from:

$$\sqrt{\text{eig}(G^T \cdot G)}$$

(2.3)

It can be shown that if the control number grows to infinite, the machine is uncontrollable because
corresponding actuator positions to the response profiles cannot be determined. (If singular, \( G \) is not invertible.) The interaction matrix should be built with care avoiding its singularity for better control performance.

The model parameter uncertainties exist due to the changes occurred in machine speed, pulpwood quality, and so on during normal operation. This uncertainty can also lead to poor control performance [26]. The measurements are all aligned with the actuator locations, only if the number of measurements is the integer multiple of the actuator number.

Consequently, the overall dynamic model for the paper machine is given as

\[
y(t) = G\frac{k_d q^{-d}}{1 - p_d q^{-1}} u(t)
\]

(2.4)

The actuator responses vary with fibre quality basis and machine speed. Figure 2.1 displays how a unit change of the actuator affects the basis weight response for different types of paper machines. As the paper weight increases from newsprint to board, the response peak has smaller amplitude, however, the coupling region becomes much greater. The different CD responses downstream from \( p_0 \) to \( p_k \) in equation (2.2) are reported in [26].
2.2 Identification

Basis weight profile, measured at the O-frame, is controlled by the slice lip actuators at the headbox, where control actions are influenced by several control loops. Identification of paper machine response model is desirable before building any control loops. The parameters of the model in equation (2.4), which depend on machine speed and grade, can be identified using the input-output measurements of bump test data. As mentioned before, actuator dynamics, time delay and interaction matrix are three components that make up a paper machine model and they are to be identified.

The deposition of coating material on to a base sheet is controlled by screw adjustments to shape the coater blade, and leads to a similar response model to that of basis weight response due to headbox slice lip adjustments. In this thesis, the model is first identified using the bump response data from the coater, and the identified model is used as a paper machine model. However, as explained later, it is required to change a parameter value of time-delay so that the identified model from the coater gives a similar response to that from the headbox. The following bump test data, taken as provided by Measurex Devron Inc. is used as an example to identify the response model that will be used in the simulation.
The parameters of test are as follows:

(i) Number of minislice (high resolution) weight measurement points: 300

(ii) Number of date scans gathered: 47

(iii) Number of actuators: 100

(iv) Actuator spacing: 3 mm

(v) Coat weight target 7 gsm

(vi) Positive bump actuator locations: 10, 30, 50, 70, 90

(vii) Negative bump actuator locations: 20, 40, 60, 80

In the coater, the set points of ten actuators were manipulated by one step at the specified locations for 30 scan period. The sensor collected 300 measurement points per scan for 47 scans. Five positive and four negative steps at the actuators, as shown in Figure 2.2 (a), created 10 bumps at the output measurement. As Figure 2.2 (b) reveals, the profiles of all bump responses have similar shapes.

Figure 2.2 (a) Up: actuator set point changes. (b) Bottom: bump responses data
2.2.1 Estimating the Interaction Matrix $G$

As shown in Equation (2.2), the interaction matrix $G$ describes the cross machine direction steady state actuators coupling in one simple matrix. The coefficients of $G$ can be approximated by averaging the positive bump responses. Each bump in Figure 2.2 (b) is relocated so that the centers of the responses are aligned, and they are drawn in Figure 2.3 (a). The databases reveal that the peaks of every bump responses are moved to left by 1 measurement point from the bump locations. All bump responses are then averaged. The resulting averaged bump response is plotted in Figure 2.3 (b) by assuming the actuator position at zero. It is seen that the single actuator movement affects about 5 actuator positions. The figure reveals that the coefficients of $G$ in this case are $[p_0, p, \ldots, p_k] = [1.71, 1.42, 1.09, 0.57, 0.1, 0.01, -0.15]$.

![Figure 2.3 (a) The aligned bump responses. (b) CD actuator position and averaged bump response](image-url)
2.2.2 Identification of Dynamic and Time Delay

In MD, the distance between the location of actuators and the sensor creates the time delay. The MD step responses are extracted at the CD location corresponding to the peak bump response magnitude. Then, they are averaged and approximated by the first order dynamic. The result gives the value of time delay $d = 1$ scans, $K_a = 0.0384$, and $p_a = 0.732$. The first order approximation is very close to the real shape of the step response as shown in Figure 2.4. The time-delay from the coater to the sensor is found to be less than one since the coater is located near to the sensor. The identification was performed under the assumption that the coater bump responses are similar to those from headbox actuators. Because of the longer distance from a headbox to a sensor, the time-delay of 1 is not an appropriate value for the paper machine model. To generate more sounding responses, the time-delay $d = 2$ will be employed in the simulation.

Overall, the paper machine model is identified using the coat weight data as

$$y(t) = G \frac{(0.0384)q^{-2}}{1 - (0.732)q^{-1}}u(t)$$

(2.5)

![Step response estimated in first order](image)

Figure 2.4 Estimated first order dynamic response
CHAPTER 3

Estimation Theory

On a paper machine, a sensor mounted on an O-frame collects measurements as it travels to and fro over the moving paper. The most common measurements are the dry basis weight and the moisture content of sheet. In addition to the measured value, the basis weight data contains high frequency sensor noise as well as random environmental disturbance. Control action and other operations are dependent on accurate measured data, and it is important to obtain data with a minimum of noise. More precise sensors and faster computer systems provide efficient ways of acquiring correct data. With the help of improved data measurement and processing techniques, including de-noising and compression techniques, it is possible to improve our knowledge of process variations. In this chapter, estimation and compression methods using wavelets and wavelet packets are presented. Traditional methods, for example Gram-polynomial and Expo-filter approximations, are described in the Appendix A and in the previous chapter respectively.

3.1 Brief History of Wavelet Theory

The underlying theory of wavelets can be traced back to the nineteenth century when Joseph Fourier presented the frequency analysis method that is now known as the Fourier analysis [27]. Later the short-time Fourier transform (STFT) was developed for time-varying frequency domain analysis. More recently, wavelet analysis, recognized as a simple tool with a great variety of possible applications, has been applied in many different fields of mathematics, physics, and engineering [14]. The term ‘wavelet’ was first introduced by Alfred Haar in 1909 [33]. However, it was not until 1982, when Morlet published his work, that wavelets became a method for signal analysis [29]. Grossmann and Morlet [21] in 1984 carried out the detailed mathematical study of the continuous wavelet transform, followed by the work in the discrete case by Daubechies, Grossmann, and Meyer [13]. In 1988, Daubechies [12] constructed a set of wavelet orthonormal base functions that drew much academic attention to the subject of wavelets.
Moreover, wavelet packets were constructed by Coifman and Meyer in 1991. More references for wavelet packets can be found in Coifman, Meyer, and Wickenhausen [7] and in [32]. Currently, researchers are actively searching for different applications in various areas of signal processing, image processing, data compression, and so on.

3.2 Wavelet Analysis

The classical Fourier analysis represents a signal in terms of trigonometric sine and cosine functions in different frequencies, allowing analysis in the frequency domain. Since the trigonometric basis functions extend over all time, this analysis is unsuitable for dealing with signals whose frequency contents are changing in time. To address this problem, Gabor (1946) proposed the Short-time Fourier transform, or windowed transform [33], as

$$STFT(\tau, f) = \int x(t)g^*(t-\tau)e^{-2j\pi ft}dt$$

(3.1)

where $g(t)$ is a window that truncates sinusoids to form basis functions with a particular width. For example, Figure 3.1 (a) shows a Gaussian window $g(t)$ that is given as

$$g_\alpha(t) = \frac{1}{\sqrt{2\alpha}}e^{-t^2/(4\alpha)}$$

(3.2)

This representation allows changes in signal spectrum to be tracked over time. For the STFT, the choice of $g(t)$ fixes the resolution over the entire time-frequency plane since the same window is used at all frequencies. Figure 3.1 (b) shows the resolution of the STFT in a time-frequency plot. Here, it is clear that the time and the frequency resolutions are fixed in all ranges. To compare resolution with the wavelet and the wavelet packet bases in Figure 3.1 (c) and (d), some definitions are to be made. Given the window function $g(t)$ and its Fourier transform $G(f)$, the time and the frequency resolutions are defined as

$$\Delta f^2 = \frac{\int f^2|G(f)|^2df}{\int|G(f)|^2df}$$

(3.3)
and \( \Delta t^2 = \frac{\int t^2 |g(t)|^2 dt}{\int |g(t)|^2 dt} \) (3.4)

where the denominator is the energy of \( g(t) \). \( \Delta t \) and \( \Delta f \) are related through an uncertainty principle [33], similar to the Heisenberg inequality which illustrates that the product of time and frequency spreads are lower bounded as

\[
\Delta t \cdot \Delta f \geq \frac{1}{4\pi} \tag{3.5}
\]

As seen from equation (3.5), the area of each rectangle in Figure 3.1 (b), (c) and (d) which is proportional to \( \Delta t \cdot \Delta f \) is near constant. This implies that a signal can be analyzed either with good time resolution or with good frequency resolution, but not both.

As shown in Figure 3.1 (c), for wavelet analysis the resolutions \( \Delta t \) and \( \Delta f \) vary in the time-frequency plane, resulting in time resolution that is finer at high frequency and frequency resolution that becomes better at low frequency. Therefore, in comparison to the Fourier transform, the wavelet transform
has the flexibility to estimate a signal with time domain discontinuity and sharp spikes, giving better time-frequency localization. The wavelet packet case will be discussed later in this chapter.

### 3.2.1 Wavelet Bases

Wavelet analysis uses wavelet functions as basis functions, where the STFT employs truncated sinusoids. Wavelets are fixed basis functions, localized in time and last for a few cycles. In wavelet decomposition, a mother wavelet which is defined below is dilated and shifted to represent various signals of different frequency and shape. For example, Haar wavelets in different scales and positions, as shown in Figure 3.2, are constructed from the basic building block, $\Psi(x)$, known as the “mother wavelet”. The discrete operation uses the dyadic scaling index $j$ and integer translation index $k$, which can be expressed as

$$\Psi_{j,k}(x) = 2^{-j/2}\Psi(2^{-j}x - k), \quad j \in \mathbb{Z}, \quad k \in \mathbb{Z}$$ (3.6)

The good time localization of Haar wavelet is traded with the poor localization property in the frequency domain. Several smoother wavelet bases, that result in better frequency localization, have been introduced. For example, Meyer, Daubechies, Symlets, and Coiflets wavelet families are shown in Figure 3.3. Similar to Equation (3.6), the mother wavelet in continuous analysis is defined in [30]. Wavelet analysis in this thesis refers to only discrete operation, unless it is mentioned otherwise.

![Wavelet Examples](image.png)

(a) Haar basis function $\Psi(x)$
(b) Haar wavelet examples

Figure 3.2 Haar mother wavelet and its translated and scaled version
Figure 3.3 Discrete wavelet filter families. Those filter coefficients are used to generate decomposition (or reconstruction) highpass (or lowpass) filters.

Table 3.1 Wavelet families and associated properties (N is the order of wavelet e.g. N = 3 for db3.)

<table>
<thead>
<tr>
<th></th>
<th>Meyer</th>
<th>Haar</th>
<th>Daubechies</th>
<th>Symlets</th>
<th>Coiflets</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Filter length</strong></td>
<td>2</td>
<td>2N</td>
<td>2N</td>
<td>6N</td>
<td></td>
</tr>
<tr>
<td>Compactely supported</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>orthogonal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Symmetry</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Asymmetry</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Near symmetry</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Smoothness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># of vanishing moments of</td>
<td></td>
<td></td>
<td></td>
<td>2N</td>
<td></td>
</tr>
<tr>
<td>(\Psi(t))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Exact reconstruction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.1 Wavelet families and associated properties (N is the order of wavelet e.g. N = 3 for db3.)
3.2.2 Properties

The qualities of different types of wavelets vary one from another, and a suitable basis wavelet should be selected to obtain the best results in estimation. The wavelet families and their associated properties are illustrated in Table 3.1. In this section, those wavelet properties are defined as follows;

- **Compact support.** If the wavelet $\Psi(x)$ is compactly supported, the wavelet filters, that will be explained in the next section, are finite impulse response (FIR) filters. Consequently, the wavelet transform becomes finite.

- **Symmetry.** Symmetry is essential for the filters to have linear phase; this property is required to avoid phase distortion.

- **Smoothness.** The smoothness of wavelets plays an important role in compression and estimation (de-noising) applications. If the wavelet is not smooth, the error can be easily detected from the filtered signal except for the signal which has sharp edges. $M$ degrees of smoothness implies that the $M^{th}$ derivative of a function is continuous at all points. A higher degree of smoothness offers better frequency localization of the filters.

- **Number of vanishing moments.** A sequence $\{\psi_k\}$ is said to have $M$ vanishing moment if

  $$\sum_k \psi_k \cdot k^l = 0, \text{ for } 0 \leq l < M.$$  

  (3.7)

Large number of vanishing moments give smoother $\Psi(t)$. Thus, this property and the smoothness are related each other. This property for each mother wavelet is defined in Table 3.1.

- **Time-frequency localization.** If $\Psi(t)$ decays rapidly, the wavelet has good time localization. Good frequency localization is achieved if $\hat{\Psi}(\omega)$ decays fast as $\omega \rightarrow \infty$ and if $\hat{\Psi}(\omega)$ is sufficiently flat near $\omega = 0$.

- **Orthogonality.** The wavelets given in equation (3.6) are orthogonal to each other in all scale:
CHAPTER 3 Estimation Theory

\[ \sum \Psi_{j,k}(t)\Psi_{j,k}(t) = \delta(j - J) \cdot \delta(k - K) \]  \hspace{1cm} (3.8)

This property links the \( L^2 \) norm of a function directly to the norm of its wavelet coefficients by

\[ \| f \| = \sqrt{\sum_{j,k} c_{j,k}} \]  \hspace{1cm} (3.9)

If orthogonal wavelets are used, the fast wavelet transform is a unitary transformation. Under this condition as in equation (3.8), numerical calculation is stable since the transformation will not increase error in initial data. This is an important property in numerical calculation.

More details on wavelet properties are presented in [24] [27] [30] and [38].

3.2.3 Discrete wavelet transform and Multi-Resolution Analysis

For analyzing discrete-time samples in paper machine data, the discrete wavelet transform (DWT) is used in this thesis with Mallet's multi-resolution analysis (MRA). In MRA, the mother wavelet \( \Psi(x) \) is determined by the highpass filter \( G \), which produces the details of the wavelet decomposition. Moreover, the scaling function \( \Phi(x) \), determined by the lowpass filter \( H \), gives the approximations of the wavelet decomposition. Thus, at level \( j \), the signal \( c_{j-1,l} \) is decomposed into the approximation \( c_{j,k} \) and the detail \( d_{j,k} \) is 1 to the length of the signal.

\[ c_{j,k} = \langle f(x), \Phi_{j,k}(x) \rangle = \sum_l h_{l-2k} \cdot c_{j-1,l} \]  \hspace{1cm} (3.10)

\[ d_{j,k} = \langle f(x), \Psi_{j,k}(x) \rangle = \sum_l g_{l-2k} \cdot c_{j-1,l} \]  \hspace{1cm} (3.11)

where the filter coefficients are related as \( g_k = (-1)^k \cdot h_{1-k} \). The inner product is defined as

\[ \langle f(x), \Psi_{j,k}(x) \rangle = \int f(x)\Psi_{j,k}(x)dx \]  \hspace{1cm} (3.12)
Figure 3.4 illustrates the multi-level decomposition and reconstruction schemes of DWT. In the
decomposition scheme, the sequence \( f(n) \) with a length \( n \) is first filtered using \( H \) and \( G \) and then down-
sampled by a factor of 2. One level decomposition produces the \( n/2 \) approximation coefficients \( c_{1,k} \) and
the \( n/2 \) detail coefficients \( d_{1,k} \).

![Block diagram of one-dimensional DWT](image)

With details and approximations, perfect reconstruction is achievable using the quadrature mirror
reconstruction filters \( H^* \) and \( G^* \), which is summarized in the following equation,

\[
c_{j,k} = \sum_{l \in Z} h_{k-2l}c_{j+1,l} + g_{k-2l}d_{j+1,l}
\]  

(3.13)

Here, a system formed by the low and high pass decomposition filters \( H \) and \( G \), together with their
associated reconstruction filters \( H^* \) and \( G^* \) is called quadrature mirror filter. The down-sampling in
decomposition introduces a distortion called aliasing. However, if the filters are chosen carefully to relate
the decomposition and the reconstruction, it is possible to cancel out the effects of aliasing. Thus, perfect
reconstruction is possible. Those chosen filters are quadratic mirror filters. This breakthrough was made
possible by the work of Ingrid Daubechies and others. The shape of the wavelets in Figure 3.3 were
determined by various kinds of quadrature mirror decomposition filters. The technical details of how to
design these filters are discussed in [35].
As an example, a sine wave with two jumps in Figure 3.6 (a) is decomposed using the MRA. The decomposition tree, in this case up to level 5, is shown in Figure 3.5. The detail coefficients at each level are plotted in Figure 3.6 (b) after the DWT is performed to level 5. At level 1, the wavelet coefficients contain the high frequency part of the signal. The analysis clearly located the jumps at this level. At higher levels, lower frequency components are extracted as shown in the figure. In Figure 3.6 (c), the wavelet coefficients of the continuous wavelet transform are displayed just to compare it with the discrete case.

3.3 Wavelet Packets

As shown in Figure 3.1 (b) and (c), the time-frequency resolution for different basis functions such as the STFT basis and the wavelet basis are predetermined. For efficient estimation of different types of signals, the basis of the signal, and so the time-frequency resolution, should be selected in a signal-dependent manner. Discrete wavelet packet transform (DWPT) to be discussed in this section, the generalization of wavelets bases, offers greater flexibility because it allows the time-frequency resolution to change in a signal decomposition. Figure 3.1 (d) shows one possible wavelet packet basis among many possible options. An adaptive procedure for choosing the most suitable set of basis functions is described later in this section.

3.3.1 Wavelet packet functions

This section introduces wavelet packets following the appearance of [27]. A wavelet packet function is indexed by three parameters: \( W_{j,m,k}(x) \). As in wavelets, \( j \) can be interpreted as a scale parameter and \( k \) as a position parameter:

\[
W_{j,m,k}(x) = 2^{-j/2} W_m(2^{-j}x - k)
\]  

(3.14)

The extra index \( m = 0, 1, \ldots \) is called the oscillation parameter. The wavelet packet functions \( W_m(x) \) is defined as
Figure 3.5 Wavelet decomposition tree

(a) Sine wave with jumps to be analyzed.

(b) Discrete Transform, wavelet coefficients.

(c) Continuous Transform, wavelet coefficients.

Figure 3.6 Wavelet coefficients in discrete and continuous transform
\[ W_{2m}(x) = 2 \sum_{k=0}^{2N-1} h_k W_m(2x-k) \]  
(3.15)

\[ W_{2m+1}(x) = 2 \sum_{k=0}^{2N-1} g_k W_m(2x-k) \]  
(3.16)

where \( W_0(x) = \Phi(x) \) is the scaling function and \( W_1(x) = \Psi(x) \) is the wavelet function. \( 2N \) is the length of the filters, \( h_k \) and \( g_k \). The Haar wavelet packets for \( m = 0 \) to 7 are computed and plotted in Figure 3.7. There are many possible combinations of grouping wavelet packet bases as a result of the added degree of freedom. The criteria for the selection of the best basis and wavelet coefficients will now be discussed.

![Figure 3.7 The Haar wavelet packets](image)

### 3.3.2 The Best Basis Algorithm

In wavelet analysis, a signal is split into an approximation and a detail component. Then the approximation itself is split into the next level approximation and the detail. This procedure is repeated
until it reaches the desired level as illustrated by the tree structure in Figure 3.5. In wavelet packet analysis, however, both the detail and the approximation can be decomposed. The completed decomposition tree structure is shown in Figure 3.8 (a). Coifman and Wickerhauser introduced the Best Basis Algorithm to suggest the best decomposition for a particular signal [8]. This algorithm finds the most efficient tree structure for a given signal in the sense that the signal will be most economically represented. Before splitting at each node, the algorithm calculates an entropy value (to be defined in this section) and decides whether further splits are necessary in order to yield a minimum-entropy decomposition tree. Entropy is a measure of concentration, describing information related property for an accurate representation of a given signal [27]. If entropy value is small, more coefficients are negligible. An example of this procedure is given at the end of the section. The information cost functions that calculate entropy are now defined; other entropy definitions are presented in [2] and [41]. Four alternative entropy-like measure of a sequence are:

(i) **Shannon entropy.** The non-normalized entropy of a sequence \( s = \{ s_i \} \) depends upon the logarithm of the squared value of each signal sample and defined by

\[
H_s = - \sum x_i^2 \log (x_i^2)
\]

where \( x_i = \frac{|s_i|^2}{\|s\|^2} \) and if \( x = 0 \) then set \( x \log x = 0 \). It has been shown that \( \exp H_s \) is proportional to the number of coefficients needed to represent the signal within a fixed mean square error [9] [41].

(ii) **The number of samples above a threshold level.** The number of samples in a sequence \( s = \{ s_i \} \) whose absolute value exceeds chosen threshold \( \lambda \) is counted, giving the number of coefficients needed to transmit the signal to a precision \( \lambda \).

(iii) **Concentration in \( l^p \) norm.** The \( l^p \) norm with \( 1 \leq p < 2 \) is calculated. For a sequence \( s = \{ s_i \} \), the \( l^p \) norm with chosen value of \( p \) is \( \sum |s_i|^p \). The smaller the \( l^p \) norm of a function for a given energy level, the more concentrated that energy is into a few coefficients [41].
(iv) **Logarithm of energy.** The log energy is defined as

\[ H_s = \sum \log(s_i^2), \]

(3.18)

setting \( \log 0 = 0 \) whenever necessary. This is an additivity-type of function.

For example, using the complete tree representation of the wavelet packet decomposition shown in Figure 3.8 (a), the sine wave with jumps is decomposed up to level 3. Using 8 terminal nodes from (3,0) to (3,7), the wavelet coefficients are displayed in Figure 3.8 (b) with the high frequency coefficients at the top and the low frequency towards the bottom. Using Shannon entropy, the calculated entropy values are shown at each node in Figure 3.8 (a). The Best Basis Algorithm first splits all approximation nodes to the desired level. Here, the approximation nodes are (1,0), (2,0) and (3,0). Then at each detail node where a split is possible, (1,1) and (2,1) in this case, the entropy value is compared with the sum of the next level entropy values. At (1,1), for example, the entropy 0.3007 is compared with the value 0.3385 that is the sum of entropies at (2,2) and (2,3). Since the entropy at (1,1) is smaller than the compared entropy, the algorithm will not split this node. Similarly all detail nodes are examined. The final result of the Best Basis Algorithm is given in Figure 3.8 (c) with the corresponding wavelet coefficient plot. In this figure, it is clear that fewer coefficients are used in the representation of the signal using wavelet packets. For an \( n \) sample signal, there exist \( 2^N \) different possible tree structures or bases in the library. In Chapter 4, it will be shown that the Best Basis decomposition result is a more compressed representation following thresholding.
CHAPTER 3 Estimation Theory

(a) Complete wavelet packet decomposition tree with entropy

(b) Wavelet packet coefficients

(c) The best tree and its decomposed coefficients

Figure 3.8 Wavelet packet decomposition of sine with jumps
3.4 Two-dimensional Analysis

The two-dimensional DWT and the DWPT algorithm can be constructed by two separate one-dimensional transforms. As shown in Figure 3.9, the rows of the two-dimensional matrix \( c_j \) are decomposed first and then the columns of the resulted sets are decomposed, resulting 4 submatrices. In Figure 3.10, the wood image is decomposed to level 3 using DWT and DWPT. At the first level of decomposition, the transformation gives 4 sets of coefficients. As mentioned before, at the next stage the DWT operates on the approximation only, where the DWPT is performed using both the approximation and the details. Thus, the fixed pattern of decomposition in the left hand image of Figure 3.10 is replaced by the right hand decomposition in which component of the mosaic may be further transformed.

3.5 Some Practical Issues

In practice, the computational complexity of the fast wavelet decomposition or reconstruction algorithm is \( O(n) \) for an \( n \) sample signal, while the fast Fourier transform requires \( O(n \log^2 n) \) operations. In the wavelet packet case, the computational cost increases to \( O(n \log^2 n) \) and the search for the best basis uses an additional \( O(n) \) operations [30].

Since the length of the filters \( H \) and \( G \) is larger than 2 and the signal sequence is finite, a boundary problem occurs. The filter lengths for different wavelets are defined in Table 3.1. In the convolution of the signal and the filters, some sequence values have to be padded at the ends of signal. There are three usual ways to handle the boundaries: symmetric, periodic, and zero padding. The symmetric boundary approach involves reflecting the function of interest about the boundaries. In the periodic case, the signal is repeated at the edge. Finally, the simplest way is to pad zeros at the edge. Periodic and zero padding solutions may introduce a discontinuity at the edge, which leads to poor estimation results. In this thesis, the first approach will be used. Other solutions to the boundary problem, such as Meyer's boundary wavelets, dyadic boundary wavelet, and orthogonal wavelets on the interval, are discussed in [14] [24] and [32].

There is no formal criteria exist for selecting the best mother wavelet. Nevertheless, a priori knowledge of the nature of the signal may be used to select the mother wavelet based on its smoothness.
[2].

Figure 3.9 Block diagram of two-dimensional DWT

Figure 3.10 Decomposition using DWT and DWPT of two-dimensional wood image
3.6 De-Noising

Traditional noise removing techniques remove high frequency noise by low-pass filtering, which smooths the function and also removes high frequency signal components. As discussed in Section 3.2 and 3.3, wavelet and wavelet packet analyses are superior to these de-noising methods, especially in estimating the signals with jumps, spikes and other non-smooth features. The wavelet-based approach to de-noising is non-linear since it is based on thresholding of coefficients. In the wavelet packet framework, the de-noising approach is identical to that of the wavelet framework, and is illustrated as follows.

The model for the noisy signal is assumed to be

\[ s_i = f_i + \sigma \epsilon_i, \quad i = 0, \ldots, n-1 \]  \hspace{1cm} (3.19)

where \( \epsilon_i \sim \text{iid } N(0,1) \) is a Gaussian white noise, and \( \sigma \) is a noise level.

In extracting the signal \( f_i \) from noise, the following two goals are suggested in the approximation, \( \hat{f} \):

- To represent \( f_i \) in terms of only a relatively small number of wavelet coefficients (i.e. achieve compression)
- To minimize the expected mean-square error

\[
R(\hat{f}, f) = \frac{1}{n} \sum_{i=0}^{n-1} E(\hat{f}_i - f_i)^2. \quad (3.20)
\]

Donoho and Johnstone in 1992 have proposed a de-noising procedure in three steps [17]. First, decompose the discrete signal \( f = \{f_1, \ldots, f_n\} \) into wavelet coefficients \( \omega \) at level N. Second, apply the thresholding technique to the detail coefficients and produce the set \( \omega^* \). Third, reconstruct \( \hat{f} \) using \( \omega^* \) which consists of the original approximation coefficients of level N and the modified detail coefficients of level from 1 to N. The procedure is also displayed in Figure 3.11.
They explain their motivation to propose this principle "selective wavelet reconstruction" as follows [17]. First, the decomposition of the real signal \( f \) is concentrated into small number leaving large value of nonzero coefficients. Second, the wavelet transform of the white noise given in Equation 3.19 is also a white noise which is evenly spread over all coefficients.

![Data analysis using wavelets](image)

**3.6.1 Threshold Rules**

Wavelet decomposition produces two sets of wavelet coefficients: approximations and details (See Figure 3.5). Under the assumption given in Equation (3.19), deleting small coefficients in details does not affect the original signal significantly. Thresholding refers to a way of cleaning out unimportant details considered to be noise (i.e. white noise). Figure 3.12 illustrates hard and soft thresholding of the straight line \( s \) for a fixed threshold \( \lambda = 0.5 \). The hard and soft thresholding strategies are:

\[
\begin{align*}
    d_{jk}^{\text{hard}} &= \begin{cases} 
    0, & d_{jk} < \lambda \\
    d_{jk}, & d_{jk} \geq \lambda 
    \end{cases} 
\end{align*}
\]

\[
\begin{align*}
    d_{jk}^{\text{soft}} &= \begin{cases} 
    \text{sign}(d_{jk})(|d_{jk}| - \lambda), & \text{if } |d_{jk}| > \lambda \\
    0, & \text{if } |d_{jk}| \leq \lambda 
    \end{cases} 
\end{align*}
\]

(3.21) (3.22)
3.6.1.1 Threshold Selection Methods

Most of the work on thresholding is due to Donoho and Johnstone, who have introduced several threshold selection methods that can be grouped into two categories: *global thresholding* and *level-dependent thresholding* [27] [32]. The former chooses a single value $\Lambda$ to be applied globally to all detail coefficients. The latter adapts a different threshold value $\Lambda_j$ for each wavelet level $j$. This data adaptive wavelet thresholding is used in handling of a nonwhite noise sequence. The thresholding value $\Lambda$ is defined by

$$\Lambda_j = \hat{\sigma}_j \cdot \lambda$$  \hspace{1cm} (3.23)

where the median of absolute deviation (MAD) is given as

$$\hat{\sigma}_j = \frac{\text{Median}(|d_{jk} - \text{Median}(d_{jk})|)}{0.6745}$$  \hspace{1cm} (3.24)

and $j$ is fixed to 1 for global thresholding. The following threshold selection methods, implemented in Matlab™ toolbox, are used to find $\lambda$. The methods are given as:

![Figure 3.12 Hard and soft thresholding with $\Lambda = 0.5$](image)
CHAPTER 3  Estimation Theory

(i) **Minimax thresholding** depends on the sample size \( n \). \( \lambda \) is chosen to minimize the upper bound of mean square error involved in estimating a signal [32]. No closed form is exist for this method. Its numerical approximation for various sample sizes are given in Table 3.2.

<table>
<thead>
<tr>
<th>( n )</th>
<th>( \lambda )</th>
<th>( n )</th>
<th>( \lambda )</th>
</tr>
</thead>
<tbody>
<tr>
<td>64</td>
<td>1.474</td>
<td>2048</td>
<td>2.414</td>
</tr>
<tr>
<td>128</td>
<td>1.669</td>
<td>4096</td>
<td>2.594</td>
</tr>
<tr>
<td>256</td>
<td>1.860</td>
<td>8192</td>
<td>2.773</td>
</tr>
<tr>
<td>512</td>
<td>2.047</td>
<td>16384</td>
<td>2.952</td>
</tr>
<tr>
<td>1024</td>
<td>2.232</td>
<td>32768</td>
<td>3.131</td>
</tr>
</tbody>
</table>

Table 3.2 Minimax thresholds for various sample size \( n \) (Donoho and Johnstone [17])

(ii) **Universal thresholding** The thresholding value is known as \( \lambda = \sqrt{2 \log(n)} \) where \( n \) is the sample size. The universal threshold value \( \lambda \) is substantially larger than that in the minimax case, so the reconstruction is smoother. This method is also called VisuShrink.

(iii) **SURE thresholding** is based on Stein's Unbiased Estimation of Risk. The threshold value \( \lambda \) is derived by minimizing and estimating the expected error value [32]. (See Equation (3.20)).

(iv) **Hybrid SURE thresholding** If the signal to noise ratio is small, the SURE estimate may be very noisy. If such a situation is present, universal thresholding is used; otherwise the SURE criterion is used [32].

As will be seen in Chapter 4, the differences in results using those methods when applied to paper machine data will not be significant and the threshold method is possibly fixed to a particular case in simulations.

3.7 Data Compression

In industry, it is desirable that the estimated paper machine data be compressed and stored for a future analysis. The wavelet-based data compression schemes have been reported by researchers, ranging
from simple entropy coding to more complex techniques such as vector quantization [2] [23]. In the wavelet transform, the compression feature appears naturally due to the large number of zero coefficient that are present after thresholding. The fact that a signal can be approximated with the decreased number of wavelet coefficients leads to the wavelet data compression. The wavelet compression of a sequence \( s = \{s_i\} \) is shown in Figure 3.13. The compression is accomplished by applying a wavelet transform or a wavelet packet transform with a thresholding method as described in the previous section. Coefficient amplitude quantization is optional, but leads to much higher compression ratio. The resulting coefficients are encoded to be stored in a data file. The compressed signal is recovered to form a sequence \( \hat{s} = \{\hat{s}_i\} \) by the procedure illustrated in bottom of Figure 3.13.

![Figure 3.13 Block diagram of compression and decompression schemes](image)

In the simulations described here, the quantization and dequantization and the encoding and decoding will not be implemented. However, in this thesis, to measure the compression performance, non-zero coefficients are counted after thresholding the data, and then compression ratio is calculated. In the reconstruction process, the mean square error, \( R \), defined in equation (3.20) is also used to assess the performance. The compression procedure using wavelet packets is identical to that of wavelet analysis. The simulation results are displayed in the next chapter.
CHAPTER 4
Data Analysis

Paper machine data includes bumps, spikes, and smooth variations at different frequencies. In this section, various methods for estimation of scanned paper machine data are examined and compared. Data compression using wavelet and wavelet packet analyses will also be discussed. In one-dimensional analysis, three different methods are considered: Gram-polynomial basis function, wavelet, and wavelet packet. The EXPO method is compared with wavelet analysis in the two-dimensional estimation. For simulation, these methods are tested using Matlab™. Several software packages are available for Matlab™ wavelet use including: WaveLab (Stanford University) [3], Wavelet Toolbox by the Math Work Inc. [27], Rice Wavelet Toolbox (Rice university), and the UBC PPC wavelet toolbox [38]. The Matlab Wavelet Toolbox is used in this thesis and has been adapted to process two-dimensional paper machine data for estimation and compression.

Estimation using the EXPO and Gram-polynomial methods is simple since only one parameter need be manipulated; a weighting factor in the EXPO algorithm and the number of polynomial order for gram decomposition. However, wavelet and wavelet packet analyses offer many different options to examine different signals. The parameters used in simulation for both wavelet and wavelet packet analyses are given briefly as follows:

- Boundary handling: symmetrical padding at the edge
- Level of decomposition: 1 to 5
- Threshold type: Hard, Soft
- Threshold selection: Minimax, Universal, SURE, SURE hybrid
- Thresholding level: global threshold, level dependent threshold

- Wavelet basis function selection: 'Haar', 'Daubechies', 'Symlets', 'Coiflets'

For the wavelet packet analysis, there is one more option, that is:

- Best basis algorithm: Shannon, threshold, norm, log energy, SURE, no best basis algorithm

4.1 Estimation of One-Dimensional Data

In this simulation, four signals with different characteristics are considered as shown in Figure 4.1. The first signal contains bumps and high frequency disturbances of the type often found in paper machine data. The second signal, a sine wave, is estimated and compared with the third signal which contains three different frequencies. The last signal is piece wise constant. To measure the error occurred in each estimation, the mean square error (MSE) defined in Equation (3.20) is employed. White noise is added to each signal before estimation carried out.

Figure 4.1 One-dimensional signals to be estimated. (a) Bumps (b) Pure sine (c) Signal containing different frequency components (d) Blocks
Table 4.1 displays the mean square errors between the estimated signals and the original signals. The first row shows the mean square errors calculated using the raw signals before estimation. Note that the noise variation is higher for the two middle cases. The numbers shown inside the parenthesis in the Gram-polynomial case are the polynomial orders used in the estimations. As mentioned above, wavelet and wavelet packet have many option choices. For the wavelet and wavelet packet analyses, an exhaustive search of 5280 estimates was used to test nearly all possibilities of combined options. The best results are listed in Table 4.2 and Table 4.3.

<table>
<thead>
<tr>
<th>Estimation Methods</th>
<th>Signal a</th>
<th>Signal b</th>
<th>Signal c</th>
<th>Signal d</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Estimation</td>
<td>10.3</td>
<td>41.3</td>
<td>40.2</td>
<td>10.3</td>
</tr>
<tr>
<td>Gram-polynomial</td>
<td>7.7 (118)</td>
<td>1.6 (16)</td>
<td>92.4 (148)</td>
<td>7.5 (118)</td>
</tr>
<tr>
<td>Wavelet</td>
<td>5.1</td>
<td>2.0</td>
<td>12.0</td>
<td>2.5</td>
</tr>
<tr>
<td>Wavelet/packet</td>
<td>6.4</td>
<td>2.0</td>
<td>9.7</td>
<td>2.6</td>
</tr>
</tbody>
</table>

Table 4.1 Best MSE of four signals using different methods (unit \(10^{-3}\))

<table>
<thead>
<tr>
<th>Signal Types</th>
<th>Levels</th>
<th>Noise Methods</th>
<th>Threshold Types</th>
<th>Threshold level</th>
<th>Wavelet Types</th>
<th>MSE (10^{-3})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal a</td>
<td>3</td>
<td>SURE</td>
<td>Hard</td>
<td>level dependent</td>
<td>sym7</td>
<td>5.1</td>
</tr>
<tr>
<td>Signal b</td>
<td>5</td>
<td>SURE</td>
<td>Hard</td>
<td>global</td>
<td>db9</td>
<td>2.0</td>
</tr>
<tr>
<td>Signal c</td>
<td>4</td>
<td>Universal</td>
<td>Soft</td>
<td>level dependent</td>
<td>sym5</td>
<td>12.0</td>
</tr>
<tr>
<td>Signal d</td>
<td>5</td>
<td>SURE hybrid</td>
<td>Soft</td>
<td>level dependent</td>
<td>haar</td>
<td>2.5</td>
</tr>
</tbody>
</table>

Table 4.2 MSE with the best choice of wavelets

<table>
<thead>
<tr>
<th>Signal Types</th>
<th>Levels</th>
<th>Noise Methods</th>
<th>Best Tree Algorithm</th>
<th>Threshold level</th>
<th>Wavelet Types</th>
<th>MSE (10^{-3})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal a</td>
<td>2</td>
<td>SURE</td>
<td>Shannon</td>
<td>level dependent</td>
<td>sym7</td>
<td>6.4</td>
</tr>
<tr>
<td>Signal b</td>
<td>5</td>
<td>SURE</td>
<td>Threshold</td>
<td>global</td>
<td>db7</td>
<td>2.0</td>
</tr>
<tr>
<td>Signal c</td>
<td>4</td>
<td>SURE</td>
<td>Threshold</td>
<td>global</td>
<td>coif2</td>
<td>9.7</td>
</tr>
<tr>
<td>Signal d</td>
<td>5</td>
<td>SURE</td>
<td>Shannon</td>
<td>level dependent</td>
<td>haar</td>
<td>2.6</td>
</tr>
</tbody>
</table>

Table 4.3 MSE with the best choice of wavelet packets
Now, a comparison of the estimations using the three methods is given more in detail. In Table 4.1, it is evident that after estimation, the mean square errors for all signals have been decreased significantly except in the Gram polynomial case where the high order polynomials are tracking the high frequency noise. The right side of Figure 4.2 shows how the MSE changes as the order of polynomial increases. For the signals a, c, and d, MSE decreases abruptly at low orders and decreases gradually at high orders. Using this method, high order Gram-polynomials, which require more computation, must be used to reduce the error closer to the level achieved by wavelet analysis. In spite of using high order Gram-polynomials, the estimation cannot obtain the MSE as low as the value resulting from the wavelet method. The Gram-polynomial method works particularly poorly in estimating the signal c, consisting of sine waves of different frequencies. As expected, however, the Gram-polynomial method gives the lowest MSE for the simple sine wave because the basis function has a similar shape. The left column in Figure 4.2 shows the estimated signals for the order marked by ‘X’ in the MSE vs. order plot. The dashed line indicates the original signal, and the bold line is the estimate.
Generally, the wavelet and wavelet packet methods give reasonable mean square errors for all of the signals, as shown in Table 4.1. For signals a, b, and d, both wavelet and wavelet packet work similarly and give similar results. Theoretically wavelet packet works better than wavelet analysis. However, the results from wavelet analysis for signals a and d are just slightly better in Table 4.1 since some options are fixed for the simulations using wavelet packet method. Table 4.2 and Table 4.3 shows that these methods use the same mother wavelets and similar thresholding. In estimating the signal c, the wavelet packet method is superior to the other methods since the best decomposition tree structure is chosen enabling the estimation to be more precise in regions where the frequency changes. In the wavelet estimation, however, the tree structure is fixed as shown in Figure 3.5. Figure 4.4 displays the tree structure for the wavelet packet estimation of the signal c. In general, the wavelet packet method works better in more complex situations because more detailed analysis at a particular frequency region is possible. In Donoho's paper [16], the mathematical analysis comparing efficiencies of those two methods is given.

As explained, in wavelet and wavelet packet analyses, the combination of different options results in different MSE's. To examine which option sets yield the best estimation for the wavelet method, the best 50 - 70 combinations that result in the lowest MSE's are selected from among 5280 estimation trials. The number of each option usage in the 50 - 70 best estimations is counted and plotted in Figure 4.5. The left upper columns of 5 figures display the examples for signal a. Here, the plots show that the best choice in estimating the first signal for decomposition level, noise method, threshold type, threshold level type, and wavelet type are level 3, SURE, Hard, Level Dependent Thresholding, and Symmlet 7 respectively. The details of different options are explained at the beginning of this section. The cases for signals b, c, and d are also shown in the bottom and left columns of Figure 4.5. The best level of decomposition varies from 2 to 5 in all signals. It is also clearly shown that the first noise method (SURE), hard thresholding, and level dependent thresholding are best for the most signals generally, allowing these options to be set at fixed value. The choice of mother wavelet is also an important factor. For blocks, the Haar mother wavelet performs best due to its similarity in shape.
For the wavelet packet estimation, the results are similar and shown in Figure 4.6. The additional option, the best basis criterion, has clear preferred setting for signals a and d. In the analysis of two-dimensional data, since more computational effort is needed, some options are fixed and only the significant variables are manipulated in the next section.

![Wavelet Estimation](image1)

![Wavelet packet Estimation](image2)

**Figure 4.3 Wavelet and wavelet packet estimation**

```
(0,0)

(1,0)       (1,1)

(2,0)       (2,1)

(3,0)       (3,1)       (3,2)       (3,3)

(4,0)       (4,1)       (4,2)       (4,3)       (4,6)       (4,7)
```

**Figure 4.4 Wavelet packet tree structure for signal c**
Figure 4.5 Count of options used by first 50 ~ 70 best results that give lowest MSE using wavelets
Figure 4.6 Count of options used by first 50 ~ 70 best results that give lowest MSE using wavelet packets.
4.2 Estimation of Two-Dimensional Paper Machine Data

After being sampled by a sensor, the basis weight measurements are stored into a matrix. Each row of the matrix contains the data collected during one scan and each column stores the profile measurement at an actuator position. By adjusting actuator settings in the paper machine model given in Chapter 2, various types of two-dimensional data sets can be created for error analysis of the different estimation methods. In this section, three different paper machine data sets are created from the model and then they are estimated using the three different filtering techniques: exponential filter, two dimensional wavelet method, and two dimensional wavelet packet method. Then, the estimated or de-noised signal is compared with the desired signal which is the original signal without measurement noise. Similar to the one-dimensional analysis, mean square error between these two signals is used to assess the efficiency of each method.

4.2.1 Smooth waves

The first data set that has been generated from the paper machine model is made up with a smooth sine wave across the sheet, a fluctuation in MD, and Gaussian noise. More specifically, during the scanning period 1 to 40, actuators are adjusted to a sine wave as shown in Figure 4.7 (a). From scan 41 onward, all actuators are set to zero. A two dimensional view of the actuator setting is depicted in Figure 4.7 (b). The basis weight profiles, as shown in Figure 4.7 (c), result from the time delay of 2-scans as well as the first order dynamic in the machine direction. For the cross direction, there are interactions between actuators, approximated by the interaction matrix G as given in equation 2.2. The MD signal in Figure 4.8 (a) is added to the previously obtained pure CD profile as shown in Figure 4.7 (c). As mentioned before, the machine direction fluctuation of basis weight results mainly from variations of consistency in the headbox, and from the pump speed changes. The MD signal is approximately periodic waves. Adding white Gaussian noise $z_{i_1,i_2} \sim N(0, \sigma^2)$ where $\sigma^2 = (0.2)^2$ to the combination of MD and CD yields the final paper machine data set that will be used in the simulation (See figure 4.8 (c)).
CHAPTER 4 Data Analysis

Figure 4.7 Initial actuator settings and resulting data set (Smooth waves)

Figure 4.8 Generated paper machine basis weight data
4.2.1.1. Estimation

The raw data given in Figure 4.8 (c) is processed using different wavelet and wavelet packet estimators, in addition to the EXPO method. Before estimation, the MD components are extracted from the raw data by subtracting scan averages of MD samples. Thus, the CD components combined with noise are used in the estimations. The simulations using various options in wavelet and wavelet packet estimations have yielded the results, in terms of MSE, as shown in Table 4.4 and Table 4.5. The lowest MSE is given at the top of the tables, and the result from use of the EXPO is given at the bottom. In comparison to the EXPO, it is shown that both wavelet and wavelet packet methods decrease the errors to less than 5% of that for the EXPO. However, there are no significant differences in mean square error between the wavelet and the wavelet packet methods.

For both methods, the choice of wavelets is an important factor in decreasing the estimation errors, as shown in Figure 4.9. It is not surprising that the smoother wavelets perform better for the estimation of the smooth profile. In effect, the best results using the Haar wavelet are about 5 times larger than the estimation errors using the smooth wavelets. The threshold values are automatically chosen according to the noise level of the signal. Figure 4.10 displays the clean CD profile and the estimated profiles in 3D view. The images of the estimated data, as shown in Figure 4.11, also illustrate that the wavelet and wavelet packet methods are superior to the EXPO method.

Figure 4.12 presents the cumulative estimated errors along the cross direction for each method. Here, errors along MD are added together to show these error profiles. The signal amplitude is higher in the error profile from the EXPO method. These figures also reveal that the error signals for both wavelet and wavelet packet have the saw-tooth form. This comes about because the original signal that is used for estimation has been created using the interaction matrix G that introduces high frequency (saw-tooth like) actuator responses. Error calculation that is obtained by subtracting the estimated signal from the original signal does not remove this high frequency component since the estimated signal using wavelets and
wavelet packets has a smooth surface. Consequently, the error signals in Figure 4.12 show this saw-tooth effect. It can also be seen that the edge effects are present in the wavelet and wavelet packet estimations.

<table>
<thead>
<tr>
<th>Levels</th>
<th>Noise Methods</th>
<th>Threshold Types</th>
<th>Threshold level</th>
<th>Wavelet Types</th>
<th>MSE ($10^{-3}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>SURE</td>
<td>Soft</td>
<td>Level dependent</td>
<td>sym6</td>
<td>4.3</td>
</tr>
<tr>
<td>3</td>
<td>SURE</td>
<td>Soft</td>
<td>Level dependent</td>
<td>sym6</td>
<td>4.6</td>
</tr>
<tr>
<td>2</td>
<td>SURE</td>
<td>Soft</td>
<td>Global</td>
<td>sym6</td>
<td>5.6</td>
</tr>
<tr>
<td>4</td>
<td>SURE hybrid</td>
<td>Hard</td>
<td>Level dependent</td>
<td>coif4</td>
<td>6.5</td>
</tr>
<tr>
<td>3</td>
<td>SURE</td>
<td>Soft</td>
<td>Global</td>
<td>db5</td>
<td>5.1</td>
</tr>
<tr>
<td>4</td>
<td>Universal</td>
<td>Hard</td>
<td>Global</td>
<td>db8</td>
<td>9.7</td>
</tr>
<tr>
<td>4</td>
<td>SURE</td>
<td>Soft</td>
<td>Level dependent</td>
<td>haar</td>
<td>26.6</td>
</tr>
</tbody>
</table>

Exponential filtering estimation | 124.5

Table 4.4 MSE for the wavelet estimation of the smooth waves

<table>
<thead>
<tr>
<th>Levels</th>
<th>Noise Methods</th>
<th>Best Tree Algorithm</th>
<th>Threshold level</th>
<th>Wavelet Types</th>
<th>MSE ($10^{-3}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>SURE</td>
<td>Shannon</td>
<td>Global</td>
<td>sym5</td>
<td>4.2</td>
</tr>
<tr>
<td>3</td>
<td>SURE</td>
<td>Threshold</td>
<td>Global</td>
<td>sym5</td>
<td>4.6</td>
</tr>
<tr>
<td>3</td>
<td>SURE</td>
<td>Norm</td>
<td>Global</td>
<td>sym5</td>
<td>6.4</td>
</tr>
<tr>
<td>3</td>
<td>SURE</td>
<td>Log energy</td>
<td>Global</td>
<td>db4</td>
<td>6.5</td>
</tr>
<tr>
<td>4</td>
<td>SURE</td>
<td>Shannon</td>
<td>Global</td>
<td>sym7</td>
<td>4.6</td>
</tr>
<tr>
<td>3</td>
<td>SURE</td>
<td>Shannon</td>
<td>Global</td>
<td>haar</td>
<td>19.9</td>
</tr>
</tbody>
</table>

Exponential filtering estimation | 124.5

Table 4.5 MSE for the wavelet packet estimation of the smooth waves
Figure 4.9 MSE for estimations using different wavelets.
CHAPTER 4 Data Analysis

(a) Clean CD Profile

(b) $\text{MSE}_{\text{expo}} = 0.1245$

Figure 4.10 Estimated profiles of the smooth waves

(c) $\text{MSE}_{\text{wavelet}} = 0.0045$

(d) $\text{MSE}_{\text{packet}} = 0.0043$

Figure 4.10 Estimated profiles of the smooth waves
Figure 4.11 Estimated profiles of the smooth waves (Images)
4.2.2 Signal Containing Several Frequency Components

The paper machine data set, depicted in Figure 4.14 (c), has been created to test the performance of the different methods for estimating signal containing several frequency components. The CD and MD profiles given in Equation 4.1 and Equation 4.2 are corrupted by the white Gaussian noise $z_{i_1, i_2} \sim N(0, \sigma^2)$ where $\sigma^2 = (0.2)^2$. This MD profile is shown in Figure 4.14 (a).

$$CD = \sin\theta_{CD1} + \sin\theta_{CD2} + \sin\theta_{CD3}$$  \hspace{1cm} (4.1)
where $\theta_{CD1} \in [0, 2\pi]$ for the actuator positions $[1, 100]$,

$\theta_{CD2} \in [0, 30\pi]$ for the actuator positions $[70, 90]$,

$\theta_{CD3} \in [0, 50\pi]$ for the actuator positions $[20, 40]$

$$MD = \sin \theta_{MD} \quad \text{where} \quad \theta_{MD} = [0, 2\pi] \quad (4.2)$$

The corresponding actuator positions for generating the CD profile (Equation (4.1)) are shown in Figure 4.13 (a) and (b). Note that there are three different frequency components existing in the actuator setting. Two different high frequency components are added during the actuator positions from 20 to 40 and from 70 to 90. The resulting data set without noise is also given in Figure 4.13 (c).

### 4.2.2.1. Estimation

The estimations have been processed by using the wavelet, wavelet packet, and EXPO methods and their results are presented in Table 4.6 and Table 4.7. Compared to MSE from the EXPO estimation, MSE for the wavelet estimation has been decreased to 15\% and that for the wavelet packet to 11\%. The wavelet packet estimation decompose a signal according to its frequency characteristics using the Best Basis Algorithm as presented in Section 3.3.2. Thus, this method analyzes the signal with different frequencies in a more efficient way. As a result, the overall MSE's for the wavelet packet estimation are lower than those found from other methods. The resulting data sets estimated from those three methods are displayed in Figure 4.15 (3D view) and Figure 4.16 (images). The cumulative error along CD, as shown in Figure 4.17, reveals that the EXPO method is not able to detect the frequency changes in the signal. Consequently, the errors at the locations where the signal has a different frequency content, are high and seem to be affected by the amplitude of the original signal. In Figure 4.17 (b), the wavelet analysis has decreased the errors significantly. Nevertheless, the wavelet packet analysis is superior to the wavelet method in detecting the frequency differences in the signal. The wavelet packet almost flattened the errors.
CHAPTER 4 Data Analysis

(a) Actuator set point changes

(b) Actuator adjustments in a different view

(c) Resulting data set

Figure 4.13 Initial actuator settings and resulting data set (Signal containing several frequency components)

(a) MD

(b) Clean MD + CD

(c) Noisy signal to be estimated

Figure 4.14 Generated paper machine basis weight data
throughout the CD range, as illustrated in Figure 4.17 (c).

<table>
<thead>
<tr>
<th>Levels</th>
<th>Noise Methods</th>
<th>Threshold Types</th>
<th>Threshold level</th>
<th>Wavelet Types</th>
<th>MSE $(10^{-3})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>SURE</td>
<td>Hard</td>
<td>Level dependent</td>
<td>coif5</td>
<td>9.1</td>
</tr>
<tr>
<td>3</td>
<td>SURE</td>
<td>Hard</td>
<td>Level dependent</td>
<td>coif5</td>
<td>11.4</td>
</tr>
<tr>
<td>4</td>
<td>SURE</td>
<td>Hard</td>
<td>Level dependent</td>
<td>coif5</td>
<td>11.9</td>
</tr>
<tr>
<td>2</td>
<td>SURE hybrid</td>
<td>Soft</td>
<td>Global</td>
<td>db6</td>
<td>10.9</td>
</tr>
<tr>
<td>2</td>
<td>SURE</td>
<td>Hard</td>
<td>Level dependent</td>
<td>sym8</td>
<td>9.4</td>
</tr>
<tr>
<td>2</td>
<td>SURE hybrid</td>
<td>Soft</td>
<td>Level dependent</td>
<td>haar</td>
<td>18.6</td>
</tr>
</tbody>
</table>

Exponential filtering estimation

Table 4.6 MSE with the best option sets for wavelet estimation (signal containing several frequency components)

<table>
<thead>
<tr>
<th>Levels</th>
<th>Noise Methods</th>
<th>Best Tree Algorithm</th>
<th>Threshold level</th>
<th>Wavelet Types</th>
<th>MSE $(10^{-3})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>SURE</td>
<td>SURE</td>
<td>Global</td>
<td>coif4</td>
<td>6.7</td>
</tr>
<tr>
<td>4</td>
<td>SURE hybrid</td>
<td>SURE</td>
<td>Global</td>
<td>coif4</td>
<td>6.7</td>
</tr>
<tr>
<td>3</td>
<td>SURE</td>
<td>Log energy</td>
<td>Global</td>
<td>coif4</td>
<td>7.2</td>
</tr>
<tr>
<td>4</td>
<td>SURE</td>
<td>Log energy</td>
<td>Global</td>
<td>db6</td>
<td>7.5</td>
</tr>
<tr>
<td>4</td>
<td>SURE hybrid</td>
<td>SURE</td>
<td>Global</td>
<td>sym7</td>
<td>7.6</td>
</tr>
<tr>
<td>3</td>
<td>SURE hybrid</td>
<td>Shannon</td>
<td>Global</td>
<td>haar</td>
<td>18.3</td>
</tr>
</tbody>
</table>

Exponential filtering estimation

Table 4.7 MSE with the best option sets for wavelet packet estimation (signal containing several frequency components)
Figure 4.15 The estimated profiles of the signal containing several frequency components
CHAPTER 4 Data Analysis

(a) Noisy CD (b) EXPO estimation

(c) Wavelet estimation (d) Wavelet Packet estimation

Figure 4.16 The estimated profiles of the signal containing several frequency components (Images).
4.2.3 Signal with Bumps and a Streak

On a paper machine, the scanned data often contains response to actuator step changes and streaks resulting when a few adjacent actuators are manipulated. To test the estimation of data containing such bump responses and streaks, the actuator settings shown in Figure 4.18 (a) and (b) are used. The resulting CD profile (Figure 4.18 (c)) from these actuator settings is added with the MD profile (Figure 4.19 (a)) together with the white Gaussian noise $z_{i_1, i_2} \sim N(0, \sigma^2)$ where $\sigma^2 = (0.2)^2$. Using a paper machine model, two positive bumps, two negative bumps and one streak are created at the following locations.
• Positive bumps: (amplitude 1)
  actuator number 23-26 during scan 30-40
  actuator number 74-77 during scan 10-20

• Negative bumps: (amplitude -1)
  actuator number 23-26 during scan 10-20
  actuator number 74-77 during scan 30-40

• Streak: (amplitude 0.5)
  actuator number 49-51 during scan 1-60

Figure 4.19 (c) displays the resulting paper machine data used in the simulation.

4.2.3.1. Estimation

Table 4.8 and Table 4.9 show results obtained from the simulation. Again, the best results are shown at the top of the tables. Since the data set in Figure 4.19 (c) has null values in most locations except those where bumps and a streak occur, the errors from the EXPO estimation are relatively low when compared with the previous results from this method. In estimating bumps and a streak, the wavelet packet estimation is slightly better than the wavelet case. Moreover, both smooth and non-smooth (Haar) wavelets accomplish estimations that give reasonable mean square errors as a result of the shapes of bumps and a streak, which appear sharp in the cross direction and smooth in the machine direction. The estimated profiles resulting from simulations are displayed in Figure 4.20 (3D view) and Figure 4.21 (images). The cumulative error along CD, as shown in Figure 4.22, shows that the EXPO method gives high estimation errors along the measurement locations, and the error is especially large at the location of the streak; that is between measurement points 147 and 153. Nevertheless, the errors from the wavelet and wavelet packet estimations are lower.
Figure 4.18 Initial actuator settings and resulting data set (Signal with bumps and a streak)

Figure 4.19 Generated paper machine basis weight data
Table 4.8 MSE with the best option sets for wavelet estimation (signal with bumps and a streak)

<table>
<thead>
<tr>
<th>Levels</th>
<th>Noise Methods</th>
<th>Threshold Types</th>
<th>Threshold level</th>
<th>Wavelet Types</th>
<th>MSE $(10^{-3})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>SURE</td>
<td>Soft</td>
<td>Global</td>
<td>db4</td>
<td>5.4</td>
</tr>
<tr>
<td>2</td>
<td>SURE</td>
<td>Soft</td>
<td>Global</td>
<td>db4</td>
<td>6.3</td>
</tr>
<tr>
<td>4</td>
<td>SURE</td>
<td>Soft</td>
<td>Global</td>
<td>db4</td>
<td>8.7</td>
</tr>
<tr>
<td>3</td>
<td>Minimax</td>
<td>Hard</td>
<td>Level dependent</td>
<td>sym5</td>
<td>6.1</td>
</tr>
<tr>
<td>3</td>
<td>Minimax</td>
<td>Hard</td>
<td>Level dependent</td>
<td>coif2</td>
<td>6.2</td>
</tr>
<tr>
<td>5</td>
<td>SURE</td>
<td>Hard</td>
<td>Level dependent</td>
<td>haar</td>
<td>8.7</td>
</tr>
</tbody>
</table>

Exponential filtering estimation 38.7

Table 4.9 MSE with the best option sets for wavelet packet estimation (signal with bumps and a streak)

<table>
<thead>
<tr>
<th>Levels</th>
<th>Noise Methods</th>
<th>Best Tree Algorithm</th>
<th>Threshold level</th>
<th>Wavelet Types</th>
<th>MSE $(10^{-3})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>SURE</td>
<td>Threshold</td>
<td>Global</td>
<td>db4</td>
<td>5.2</td>
</tr>
<tr>
<td>3</td>
<td>SURE hybrid</td>
<td>Threshold</td>
<td>Global</td>
<td>db4</td>
<td>5.2</td>
</tr>
<tr>
<td>3</td>
<td>SURE</td>
<td>Shannon</td>
<td>Global</td>
<td>db4</td>
<td>7.0</td>
</tr>
<tr>
<td>3</td>
<td>SURE</td>
<td>Norm</td>
<td>Global</td>
<td>sym7</td>
<td>6.1</td>
</tr>
<tr>
<td>3</td>
<td>SURE hybrid</td>
<td>Log energy</td>
<td>Global</td>
<td>Coif3</td>
<td>6.2</td>
</tr>
<tr>
<td>3</td>
<td>SURE</td>
<td>Shannon</td>
<td>Global</td>
<td>haar</td>
<td>8.5</td>
</tr>
</tbody>
</table>

Exponential filtering estimation 38.7

Table 4.9 MSE with the best option sets for wavelet packet estimation (signal with bumps and a streak)
Figure 4.20 The estimated profiles of the signal with bumps and a streak
Figure 4.21 The estimated profiles of the signal with bumps and a streak (Images)
4.2.4 Industrial Data Analysis

The industrial data set, presented in Figure 4.23 (b), has been obtained from a coater with several actuator manipulations during 15 scanning periods. The data represents changes in sheet basis weight as a result of the addition of coating material. The actuator set point changes are shown in Figure 4.23 (a). The specifications of this data are given as follows:

- # of high resolution weight measurements: 360
- # of scans: 37
• # of actuators: 78

• Coat weight target: 7gsm

• Four bumps during scan 10-26 resulted from following actuator settings:
  Positive bumps: (amplitude shown inside parenthesis)
  actuator number 37 (0.9), 38 (2.2), and 39 (0.9)
  actuator number 67 (1.9), 68 (2.2), and 69 (0.9)

  Negative bumps: (amplitude shown inside parenthesis)
  actuator number 22 (-0.9), 23 (-2.2), and 24 (-0.9)
  actuator number 52 (-0.9), 53 (-2.2), and 54 (-0.9)

The estimation of the noisy data has been conducted using three different methods, the EXPO, wavelet, and wavelet packet analyses. The images in Figure 4.23 illustrate that it is hard to distinguish the locations of bumps in the EXPO estimation, while the wavelet and wavelet packet methods clearly identify two positive bumps and two negative bumps. The simulation results are also shown in Figure 4.24 in 3D plots. Note that the basis weight value is normalized in the range of 0 to 1 in this figure.

As mentioned before, the wavelet packet method gives results that are similar to the wavelet method. However, the wavelet packet method is more flexible since it selects the best decomposition tree structure among many possible options. For example, the Best Basis Algorithm using five different entropy types generates the five different best tree structures as shown in Figure 4.25 (b) to (f). If the options 'threshold', 'norm', and 'SURE' are selected in entropy calculation, the decomposition tree structures will differ depending upon what parameter values are chosen. In Figure 4.25 (c), the threshold value 1.5 was chosen and the tree structure becomes the same as that for the wavelet method as displayed in Figure 4.25 (a). For 'norm' and 'SURE', the values 1.8 and 15 are used respectively. The different tree structures result in different estimations. In this examples using the above tree structures, the difference in accuracy of the
Figure 4.23 The estimated profiles of the industrial data (Images)
estimated data is not comparable in this case. It is clear from Figure 4.25 that the wavelet packet tree structure analyzes the high frequency region more precisely than the wavelet.

Figure 4.24 The estimated profiles of the industrial data
CHAPTER 4 Data Analysis

(a) Wavelet method  
(b) Shannon method

(c) Threshold method: 1.5  
(d) Norm method: 1.8

(e) Log energy method  
(f) Sure method: 15

Figure 4.25 Decomposition trees resulted from the Best Basis Algorithm
4.3 Data Compression

In investigating data compression, the estimated data sets obtained in the previous section are tested with the compression techniques described in Section 3.7. The data compression procedures are summarized into three steps: decomposition, thresholding of wavelet coefficients, and reconstruction. In the decomposition process, the wavelet and wavelet packet analysis transforms the data into fewer wavelet coefficients. For example, the estimated smooth wave with the size 60 by 300 is decomposed and its wavelet coefficients are displayed in Figure 4.26. Figure 4.27, shown with a different axis scale, illustrates that first 1000 coefficients out of total coefficients fluctuate in amplitude within the range ±20, and these concentrated coefficients have larger values compared to the remaining coefficients. The thresholding reduces to zero all coefficients between the dashed lines in Figure 4.26 (a). All coefficients except those that have relatively large values become null values, providing the compression. In this thesis, the compression ratio and the mean square error in Equation (3.20) are used to determine how well data is compressed. To achieve a high compression ratio, it is desired to have more zero coefficients and small mean square error after recovering signal from data compression.

Table 4.10 illustrates the efficiency of the data compression in terms of compression ratio and mean square error between the signals before and after the compression, using different options for the given data sets. For all cases, the wavelet packet compressions achieve higher compression ratios with lower mean square errors.

The original data sets before data compression and the recovered signals after compression are shown in Figure 4.28 in an image view. The data compression is lossy since the recovered data has lost some values in the thresholding process. Nevertheless, the recovered signals have retained most of their energy while keeping the compression ratio high so that the recovered signals in Figure 4.28 show very similar images to the original signals.
Figure 4.26 Truncated wavelet coefficients before and after thresholding

Figure 4.27 First 2500 coefficients before thresholding
Table 4.10 Data compression results using wavelet and wavelet packet methods

<table>
<thead>
<tr>
<th>Signal used</th>
<th>Estimation methods</th>
<th>Wavelet types</th>
<th>Compression ratio</th>
<th>MSE (10^{-3})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smooth wave</td>
<td>wavelet</td>
<td>db3</td>
<td>38.81 : 1</td>
<td>9.4</td>
</tr>
<tr>
<td></td>
<td>wavelet packet</td>
<td>coif3</td>
<td>47.14 : 1</td>
<td>2.5</td>
</tr>
<tr>
<td>Signal with different</td>
<td>wavelet</td>
<td>sym4</td>
<td>19.58 : 1</td>
<td>15.3</td>
</tr>
<tr>
<td>frequencies</td>
<td>wavelet packet</td>
<td>coif5</td>
<td>37.39 : 1</td>
<td>5.1</td>
</tr>
<tr>
<td>Signal with bumps and</td>
<td>wavelet</td>
<td>sym4</td>
<td>37.03 : 1</td>
<td>5.4</td>
</tr>
<tr>
<td>streak</td>
<td>wavelet packet</td>
<td>db4</td>
<td>59.23 : 1</td>
<td>1.5</td>
</tr>
<tr>
<td>Industrial data</td>
<td>wavelet</td>
<td>sym5</td>
<td>31.73 : 1</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>wavelet packet</td>
<td>sym5</td>
<td>64.00 : 1</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Figure 4.28 Images before and after the data compression
Chapter 5

CD Control of Basis Weight

After the estimation of the true profile using data that is measured by a sensor, actuators in a headbox are manipulated in order to achieve a uniform distribution of basis weight in the cross direction of a paper sheet. It has been shown in chapter 2 that the process model has a first-order response with a time-delay, and that the couplings between actuators may be handled by the interaction matrix G. Using these characteristics, the Dahlin controller, which is a dead time compensator, is implemented because of its simple design. Later in this chapter, a method for profile control performance assessment is also discussed.

5.1 Dahlin Controller

The Dahlin algorithm can be applied to an arbitrary process, forcing the closed loop response of the feedback system to be a unit gain first-order process with dead-time [25]. The closed loop dead-time is equal to the open loop value. This algorithm has the only one tuning parameter, \( p \), that is the closed loop pole location and should be positive and less than one.

It is assumed that the process dynamic of a discrete plant, \( H_{OL}(q^{-1}) \), can be described by

\[
H_{OL}(q^{-1}) = \frac{B(q^{-1})}{A(q^{-1})} = \frac{b_0(1 + b_1q^{-1} + \cdots + b_{N_b}q^{-N_b})}{1 + a_1q^{-1} + \cdots + a_{N_a}q^{-N_a}} q^{-d}
\]  

(5.1)

A controller, \( C(q^{-1}) = \frac{B_c(q^{-1})}{A_c(q^{-1})} \) can be designed using feedback as shown in Figure 5.1 so that the closed loop system becomes a unit gain first-order process with a time delay. The controller first cancels out all the stable poles and zeros of \( H_{OL} \) which means that the process has to be stable. Since the model
used for basis weight control has a stable first-order dynamic with a delay, this controller can be used. The desired closed loop system is given as

$$H_{CL}(q^{-1}) = \frac{1 - p}{1 - pq^{-1}}q^{-d} \tag{5.2}$$

where $p$ is a tuning parameter. From Figure 5.1, the closed loop system can be derived to be

$$H_{CL}(q^{-1}) = \frac{H_{OL} \cdot C}{1 + H_{OL} \cdot C} = \frac{BB_c}{AA_c + BB_cq^{-d}}q^{-d} = \frac{1 - p}{1 - pq^{-1}}q^{-d} \tag{5.3}$$

Using this equation, the controller is obtained:

$$C(q^{-1}) = \frac{B_c(q^{-1})}{A_c(q^{-1})} = \frac{1 - p}{1 - pq^{-1} - (1 - p)q^{-d}} \cdot \frac{A(q^{-1})}{B(q^{-1})} \tag{5.4}$$

Assume that the plant model is given as a first-order process with delay of $d$, as

$$H_{OL}(q^{-1}) = \frac{b}{1 - aq^{-1}}q^{-d} \tag{5.5}$$

Then, the controller becomes

$$C(q^{-1}) = \frac{(1 - p)}{1 - pq^{-1} - (1 - p)q^{-d}} \cdot \frac{1 - aq^{-1}}{b} \tag{5.6}$$

![Figure 5.1 Dahlin controller](image_url)
For a time delay equal to one, the controller in recursive form can be written as:

$$u_c(t) = \frac{1-p}{b} [e(t) - ae(t-1)] + u_c(t-1) \quad (5.7)$$

and if a delay is two, the controller becomes

$$u_c(t) = \frac{1-p}{b} [e(t) - ae(t-1)] + pu_c(t-1) + (1-p)u_c(t-2) \quad (5.8)$$

5.2 CD Profile Assessment

In two-dimensional wavelet analysis, the number of scans required for the estimation should be greater than 2 to 3 filter lengths [38]. In particular, the filter length for Daubechies 4 as shown in Figure 3.3 is eight samples. Hence, the required number of scans is in the range of 16 to 24 scans. The previous data boxes that include 16 to 24 scans should be estimated and then the most recent estimates are taken in the calculation of control action.

On the basis of the current profile, the next actuator settings are determined automatically in the computer control setting. However, it is necessary for an operator to monitor those processes continuously in order to respond to the changing conditions adequately. Therefore, a method with which an operator can assess the control performance is developed in this section. The industrial settings used in this simulation are as follows: The basis weight profile data were collected on a newsprint machine for 130 scans. The paper machine is equipped with $N_m = 350$ measurements, $N_a = 130$ actuators, and 3.25 inches of the actuator width.

The profile measurements contain various frequency components that can be divided into three categories: controllable components, uncontrollable components, and noise. If there exist controllable frequency components in the profile, control actions are required to remove those components. The uncontrollable frequency components and noise are those variations that can not be removed by manipul-
The spatial frequency that divides the controllable and the uncontrollable regions is defined as a cut-off frequency. This frequency is determined by the actuator width and the single actuator response for a particular machine. It was found that the spatial cut-off frequency should be less than 0.5 or that the wavelength be greater than 2 [40]. Here, the unit wavelength is defined to be an actuator width. The frequency is inversely proportional to the wavelength. In Figure 5.2, the dotted lines display actuator widths and the solid lines display the sine wave for the cut-off frequencies which are identified in the spectrum plot. The sine wave is plotted only up to 40 measurement points. In plotting these figures, the industrial settings given at the beginning were employed. The top two plots show the cut-off frequency 0.38 Hz or the wavelength 2.7 when the single actuator response is broader than the given actuator width. Also, two plots in the bottom illustrate the sine wave when the spatial cut-off frequency is 0.5 or the wavelength 2. This is the case when the cut-off frequency is equal to one actuator width. In the simulation, the wavelength value in the later case is used to calculate a cut-off frequency.

![Figure 5.2 Different cut-off frequencies or wavelength for a given actuator width.](image)
5.2.1 Performance Index

In the wavelet and the wavelet packet analyses, some features such as their abilities to decompose a signal into different frequencies and orthonormal characteristics of the decomposed coefficients, can be used to build the performance index that measures the control performance. Let $\sigma^2_{pr}$ be the noise-free overall profile variance and $\sigma^2_{cont}$ be the variance of the controllable component. Then, the control performance index, $C$, is defined [40] as

$$C = \frac{\sigma^2_{pr}}{\sigma^2_{pr} - \sigma^2_{cont}}$$ (5.9)

If $C = 1$, it denotes that the control action has removed all controllable components. $C > 1$ indicates that the controllable variations remain in the profile. $C < 1$ is not possible.

5.2.1.1 Frequency Characteristics of Wavelet and Wavelet Packet

To implement the performance index in Equation (5.9), it is crucial to divide the profile data into the controllable and uncontrollable components using the given cut-off frequency. For an example, the profile from the data specified at the beginning of Section 5.2 is decomposed down to level 3 using wavelet analysis. The profile is shown at top of Figure 5.3. Also the approximation and details with their spectrums in the frequency domain are displayed in this figure. The spectra at the right hand side show that the wavelet decomposition divides the data into different frequency regions. The dashed line shows the expected ranges of frequency that is resulted from the decomposition. Moreover, the wavelength ranges of the signals are displayed between the signal and the spectrum plots. Similarly, the wavelet packet decomposition using the full tree structure, as shown in Figure 3.8 (a), results in Figure 5.4. It is clear that the wavelet packet analysis divides the signal into smaller frequency components. Hence, the wavelet
packet provides the more accurate value in calculating the performance index. For example, the performance index calculation using the wavelet analysis gives $C = 4.94$ while that for the wavelet packet analysis yields $C = 7.81$. The lower value of $C$, 4.94, in the wavelet case implies better control performance than the value 7.81. However, this lower value resulted from less accurate frequency division. More specifically, using the spatial cut-off frequency 0.5, the controllable components in the wavelet decomposition are located at (3,0) and (3,1) as shown in Figure 5.3. Here, some controllable components in (2,1) are not included in the calculation. However, the controllable variations from the wavelet packet analysis includes the components at (3,2) in the addition to those at (3,0) and (3,1). It is clear that the wavelet packet yields a more precise calculation.
Figure 5.3 Approximation and details from wavelet decomposition and their spectra
Figure 5.4 Approximation and details from wavelet packet decomposition and their spectra.
5.3 CD Basis Weight Control

The control strategy for CD basis weight control using wavelets or wavelet packets is shown in block diagram in Figure 5.5. The inverse of the interaction matrix, $G$, is used to estimate present actuator settings before the estimated output proceeds to the controller, $C$. The parameter values for $G$ as found in Chapter 2 are used. The collected profile data by a sensor is de-noised using wavelet and wavelet packet methods that were presented in Chapter 3 and 4. For assessment of control performance, the performance index defined in Section 5.2 is calculated with the help of wavelet and wavelet packet decompositions.

![Figure 5.5 Block diagram of basis weight control scheme using wavelet analysis](image)

In control simulations, it is assumed that the control action takes place at every scan even though industrial practice is to take action every 2 to 3 scans. In [25], it is explained that the results from those two cases have small differences. To simplify the estimation process in this simulation, the estimator analyzes one-dimensional data in every scan instead of estimating the two-dimensional data as illustrated in Chapter 4.

In the simulation of CD control, the paper machine data is first generated. The actuators are fixed as the shape shown in Figure 4.7 (a) for 9 scan periods to generate the basis weight data. From the scan 9, the controller is turned on to calculate appropriate actuator adjustments. The simulation result is displayed in Figure 5.6. The basis weight reaches its peak at scan 9 and then gradually decreases toward zero after
the controller is turned on. Figure 5.7 (a) and (b) show the calculation of the performance index based on wavelet analysis and on wavelet packet analysis respectively. The final performance index value reaches around to 1.9 if wavelet analysis is used and around to 2.4 for wavelet packet. It seems that the wavelet analysis yields more desirable control performance. However, as mentioned before, the performance index calculation based on the wavelet packet analysis yields more accurate results due to its ability to decompose the signal in finer frequency resolution.

As a result, the control actions of the designed controller have reduced the basis weight variation significantly and the value of performance index coefficient has been decreased dramatically after a controller has been turned on at scan 9.

![Estimated profile data]

Figure 5.6 Control of CD Profile: control action took place from scan 9
Performance Index using (a) Wavelet (b) Wavelet Packet

Figure 5.7 Performance index during CD control
CHAPTER 6

Conclusion

6.1 Summary

The thesis discusses a new method for paper machine signal analysis using the wavelet packet transformation. It has been shown that this technique can be applied to three main areas in paper machine control: paper machine data analysis including estimation and compression, control of basis weight or moisture variations, and control performance monitoring.

The paper machine model for use in CD control has been derived from experimental bump test data. This model gives a relationship between the actuator settings and the CD basis weight response. It has been shown that the model consists of actuator dynamic, time-delay, and interaction. Coupling between actuator spaces has been approximated using an interaction matrix.

To improve our knowledge of process variation from noisy sensor measurements, there exist several traditional methods; Expo-filter and Gram-polynomial approximation, and more recently wavelet estimation. A new approach, discrete wavelet packet transform (DWPT), introduced in this thesis, offers greater flexibility because it allows time-frequency resolution to change in a signal decomposition where other methods use fixed time-frequency resolution. The Best Basis Algorithm has been introduced to obtain a best decomposition tree structure in decomposition of a given signal.

It has been shown clearly that the new method, along with the wavelet estimator, produces results superior to standard analysis methods by reducing the error of estimation. In estimating a more complex signal, however, the wavelet packet method is superior to the wavelet analysis since the best decomposition tree structure is chosen enabling the estimation to be more precise in the regions where frequency changes. From the exhaustive search in the estimation of scanned paper machine data, it has been found that options
in wavelet and wavelet packet analyses can be fixed depending on the characteristics of a given signal. When applied to industrial data, the wavelet packet estimation has provided more visual results by distinguishing the bump locations more evidently. Moreover, it has been shown that wavelet packet compression is more efficient than the wavelet case since it has achieved a higher compression ratio by yielding the combination of more zero coefficients and lower mean square error.

The control scheme has been developed using Dahlin controller. Also the control performance assessment tool developed by using wavelet and wavelet packet has been introduced. Wavelet packet provide more precise values in calculating the performance index because this method can divide the frequency range of a signal with more finer resolution. It has been shown that the control actions of the designed controller have reduced the basis weight variation significantly and the value of performance index coefficient has been decreased dramatically after a controller has been turned on.

6.2 Future Work

There are some possible areas for further research. Some topics are given as follows:

- In this thesis, only white noise was used as a noise model, but the paper machine measurements can be corrupted by various types of noise (i.e. colored noise). It is necessary to analyze the effect of estimation and compression carried out using different noise models.

- On the assumption that the basis weight variations include time-varying components, the time-varying wavelet packet estimator can be adapted, which would achieve better estimation with reduced error. This method can also be applied to data compression.

- More detailed analysis can be carried out in boundary handling problems using wavelet and wavelet packet methods.
• Data compression based on the wavelet technique can be further developed in paper machine data compression. For example, many researchers reported wavelet-based data compression using vector quantization, adaptive transforms, tree encodings, and edge-based coding.
References


References


Appendix A. Gram Polynomial Estimation

By definition, Gram polynomials are the discrete equidistance orthogonal polynomials with equal weighting on every data points. This technique is used to approximate the noisy measurements. The general form of the polynomial is described as

\[ P_m(i) = c_{mN} \sum_{j=0}^{m} \frac{(-1)^j (m+j)^{(2j)} i^j}{(j!)^2 (N-1)^{(j)}} \]  

(EQ 1)

The polynomial order is given as \( m \) and \( i \) is defined over the interval \([0, N-1] \). When \( c_{mN} = 1 \), the Gram polynomials in a recursive form are written as

\[ P_m(i) = \frac{(N-1)(2m-1)}{m(N-m)} \left( 1 - \frac{2i}{N-1} \right) P_{m-1} - \frac{(m-1)(N-1+m)}{m(N-m)} P_{m-2}(i) \]  

(EQ 2)

and set \( P_0(i) = 1 \) and \( P_{-1}(i) = 0 \).

The \( m \) orthogonal polynomial evaluated at \( N \) points can be constructed and it is defined as

\[ X_0 = \begin{bmatrix} P_1(0) & P_2(0) & \cdots & P_m(0) \\ P_1(1) & P_2(1) & \cdots & P_m(0) \\ \vdots & \vdots & \cdots & \vdots \\ P_1(N-1) & P_2(N-1) & \cdots & P_m(N-1) \end{bmatrix} \]  

(EQ 3)

Now, let the profile \( y(t) \) be a sequence measured at \( i = 0, \ldots, N-1 \). The profile then can be expressed in terms of Gram-polynomial as \( y = X^T_0 y_C + \varepsilon \) where \( y_C \) is the polynomial coefficients and \( \varepsilon \) the estimation residue that is consists of terms of order \( m + 1 \) and higher. The coefficients \( y_C \) is obtained
from the least square estimation, that is

$$\hat{y}_c = (X_0X_0^T)^{-1}X_0y$$

(EQ 4)