# ASSESSING SELF-SIMILARITY IN REDUNDANT COMPLEX AND QUATERNION WAVELET DOMAINS: THEORY AND APPLICATIONS

A Thesis
Presented to
The Academic Faculty

by

Taewoon Kong

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Industrial and Systems Engineering

Georgia Institute of Technology
May 2019

# ASSESSING SELF-SIMILARITY IN REDUNDANT COMPLEX AND QUATERNION WAVELET DOMAINS: THEORY AND APPLICATIONS

Approved by:


Professor Brani Vidakovic, Advisor
School of Industrial and Systems Engineering
*Georgia Institute of Technology*

Professor Yajun Mei
School of Industrial and Systems Engineering
*Georgia Institute of Technology*

Professor Kamran Paynabar
School of Industrial and Systems Engineering
*Georgia Institute of Technology*

Professor Sung Ha Kang
School of Mathematics
*Georgia Institute of Technology*

Professor Kichun Lee
Department of Industrial Engineering
*Hanyang University*



Date Approved: March 15, 2019

*To my beloved family:*

*my wife (Julia Jiyul Lim),*

*my daughter (Eleanor Rahee Kong),*

*and my son (Everett Jaehee Kong).*

# ACKNOWLEDGEMENTS

While this dissertation represents the culmination of my graduate research, there are many others who deserve a large share of the credit for making this work possible.

First and foremost, I would like to express my heartfelt gratitude to Prof. Brani Vidakovic for being the best advisor, mentor, and role model. From the very beginning of my Ph.D. study, he has trusted me, given me the brilliant insights for my research, and supplied the guidance with the continuous support. Without his enlightening instruction, great kindness, special patience, unwavering enthusiasm, and trust, It would have been impossible to finish my Ph.D. study. I could not have imagined having a better advisor.

I would also like to appreciate my previous advisor in Hanyang University, Prof. Kichun Lee, for giving me the first opportunity to be exposed to research. I would not be where I am now without his advice directing me to pursue Ph.D. degree in USA. I also want to extend my gratitude to my thesis committee members, Prof. Yajun Mei, Prof. Kamran Paynabar, and Prof. Sung Ha Kang for all their careful reviewing and insightful comments that greatly improved this dissertation.

Last but not least, I would like to thank my family for their constant love and support throughout this journey. I could not have done it without them. Special thanks goes to my wife, Jiyul Lim. I could not have come this far without her kindness, patience, support, and love. She wholeheartedly supported my decision to move to USA across the ocean to follow my dreams. For four years of Ph.D. study, she always has stood by me in joy and in sorrow. Thanks Jiyul, the love of my life, for being my brightest star in the darkest of nights. Also, thanks my daughter, Rahee Kong, and my son, Jaehee Kong, for your smiles that bring happiness to us.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# SUMMARY

Theoretical self-similar processes have been an essential tool for modeling a wide range of real-world signals or images that describe phenomena in engineering, physics, medicine, biology, economics, geology, chemistry, and so on. However, it is often difficult for general modeling methods to quantify a self-similarity due to irregularities in the signals or images. Wavelet-based spectral tools have become standard solutions for such problems in signal and image processing and achieved outstanding performances in real applications.

This thesis proposes three novel wavelet-based spectral tools to improve the assessment of self-similarity.

In Chapter 2, we propose spectral tools based on non-decimated complex wavelet transforms implemented by their matrix formulation. This non-decimated complex wavelet spectra utilizes both real and imaginary parts of complex-valued wavelet coefficients via their modulus and phase. A structural redundancy in non-decimated wavelets and a componential redundancy in complex wavelets act in a synergy when extracting wavelet-based informative descriptors. In particular, we suggest an improved way of separating signals and images based on their scaling indices in terms of spectral slopes and information contained in the phase in order to improve performance of classification. It is also worth mentioning that the proposed method can handle signals of an arbitrary size and in 2-D case, rectangular images of possibly different and non-dyadic dimensions because of the matrix formulation of non-decimated wavelet transform. This is in contrast to the standard wavelet transforms where algorithms for handling objects of non-dyadic dimensions requires either data

preprocessing or customized algorithm adjustments.

Quaternion wavelets are another redundant wavelet transforms generalizing complex-valued wavelet transforms. In Chapter 3, we step into the quaternion domain and propose a matrix-formulation for non-decimated quaternion wavelet transforms and define spectral tools for use in machine learning tasks. Since quaternionic algebra is an extension of complex algebra, quaternion wavelets bring more redundancy in the components that proves beneficial in wavelet based tasks. Specifically, the wavelet coefficients in the decomposition are quaternion-valued numbers that define the modulus and three phases. We define non-decimated quaternion wavelet spectra based on the modulus and three phase-dependent statistics as low-dimensional summaries for 1-D signals or 2-D images. A structural redundancy in non-decimated wavelets and a componential redundancy in quaternion wavelets are linked to extract more informative features.

Dual relation that is an alternative representation to analyze the same problem has been used in various fields including optimization, physics, engineering, and mathematics, etc. However, we are not aware of the definition of wavelet spectra dual to a standard wavelet second-order spectra, as far as we know, thus, the duality concept is worth considering to measure a self-similarity index by novel scaling indices from a different perspective. In Chapter 4, we suggest a dual wavelet spectra based on non-decimated wavelet transform in real, complex, and quaternion domains. This spectra is derived from a new perspective that draws on the link of energies of the signal with the temporal or spatial scales in the multiscale representations. Unlike finding linear relationship between log-energies and discretized scales, that was a staple for defining standard wavelet specta (or primal specta in this context), we look at the behavior of scales as a function of quantized log-energies of all wavelet coefficients. Specifically, we would be interested in an average scale found among all coefficients of a fixed log-energy (or a log-energy from a fixed interval) in a wavelet decomposition of a signal.

Potentially, any multiscale transform utilized for definition of primal spectra can serve as a basis for defining the dual spectra. Here we advocate the use of non-decimated wavelet transforms because of their redundancy and robustness. In the implementations of dual spectra we will use non-decimated wavelet transforms implemented by matrices. This matrix formulation is beneficial for several reasons, including versatility, adaptivity, and generalizability, especially in the multidimensional formulations as fore-mentioned. We utilize the dual wavelet spectral tools for distilling additional low-dimensional descriptors in classification tasks involving signals and images. We also quantify the benefits of including such descriptors in the classification procedures. As we demonstrated, the dual descriptors increase classification performance when added to spectral descriptors derived from primal spectra. The proposed dual wavelet spectra from the real domain is extended to complex and quaternion domains utilizing matrix formulation with complex and quaternion wavelet filters.

In Chapter 5, to demonstrate the use of three defined spectral methodologies, we provide four examples of application on real-data problems: classification of visual acuity using scaling in pupil diameter dynamic in time and diagnostic, classification of sounds using scaling in high-frequency recordings over time, screening digital mammogram images using the fractality of digitized images of the background tissue, and monitoring of steel rolling process using the fractality of captured digitized images. The proposed tools are compared with the counterparts based on standard wavelet transforms in terms of computing time and evaluation metrics for classification problems.

# CHAPTER I

# INTRODUCTION

## *1.1 Self-similar Process*

Theoretical self-similar processes such as fractional Brownian motion are becoming an fundamental tool for modeling a wide range of real-world phenomena in engineering, physics, medicine, biology, economics, geology, chemistry, and so on. Time series can be examined in two complementary domains: time and scale/frequency domain. Here we focus on the second domain revealing the scaling and self-similarity properties of time series.

### 1.1.1 Examples of Self-similar Process

#### *1.1.1.1 It Started with Hurst and Nile Data*

Hurst exponent was discovered by Harold Edwin Hurst who, during his 62 years in Egypt, mainly worked to design reservoirs for the Nile River. Hurst was trying to find an optimal reservoir capacity R such that it can accept the river flow in N units of time, $X_1, X_2, \ldots, X_N$, and have a constant withdrawal of $\bar{X}$ per unit time. By inspecting historical data on Nile River flow, Hurst discovered an interesting phenomenon that is now referred to as the Hurst effect. The optimal volume of the reservoir was given by the so called adjusted range,

$$R = \max_{1 \le k \le N} (X_1 + \cdots + X_k - k\bar{X}) - \min_{1 \le k \le N} (X_1 + \cdots + X_k - k\bar{X}).$$

Hurst also reviewed other geophysical data because records of the water flow rarely lasted more than 100 years. Thus, he standardized their adjusted ranges R for comparisons with sample standard deviation

$$S = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2}$$

to obtain dimensionless ratio $R/S$ and adjusted range. On basis of more that 800 records, Hurst discovered that quantity $R/S$ scales as $N^H$, for ranging from 0.46 to 0.93, with mean 0.73 and standard deviation of 0.09 (Hurst, 1951). This result was in contrast to the fact that for independent normal random variables $H$ is 1/2 in the limit. Feller (1951) proved that the theoretical value of $R/S$ was 1/2 for independent and identically distributed random variables with a finite second moment. It was supposed that strong Markovian dependence was responsible for deviations from $H = 1/2$ until Barnard (1956) proved that limit $H = 1/2$ holds for the Markovian dependence case. Mandelbrot (1975), Mandelbrot and Ness (1968), and Mandelbrot and Wallis (1968) associated the Hurst (or Joseph) phenomenon in the presence of long memory. Figure 1 (left) presents n=512 consecutive yearly measurements from the famous Nile River Data set for the years 62-1281 A.D. Figure 1 (right) displays its wavelet spectra, demonstrating the scaling law.

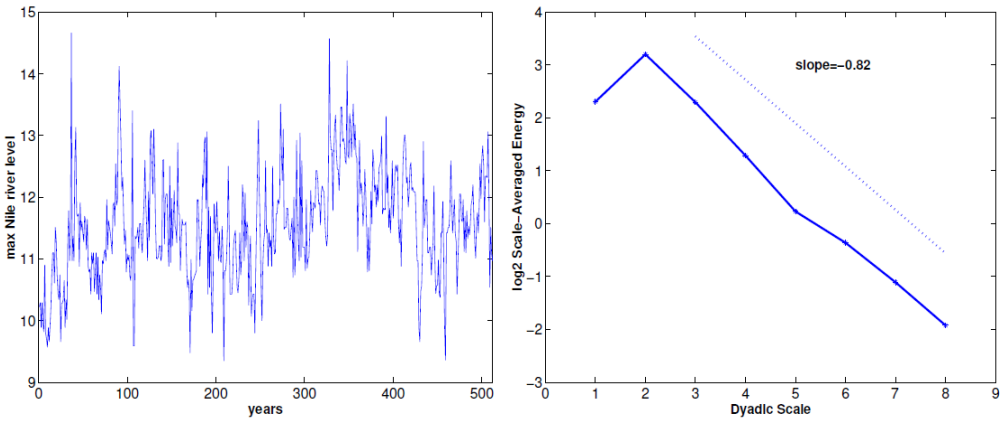



**Figure 1:** Nile yearly minimal level data (left) and its wavelet spectra (right).

#### *1.1.1.2 Rates of Exchange*

Many economic time series, such as stock market prices, exchange rates and asset returns exhibit self-similar property and long range dependence (LRD). This is in empirical contradiction to several economic theories such as random walk theory for stock market, perfect markets, and so on. Thus, several theories and models describing the scaling and LRD (such as ARFIMA, fGn, fBm, GARCH, etc) have been developed. The rates of exchange between Hong Kong Dollar (HKD) and USDollar (USD) reported by the ONADA Company between 24 March 1995 and 1 November 2000 are illustrated in Figure 2. Figure 2(a) shows the rates of exchange and panels (b) and (c) represent scaling behaviors in Fourier and wavelet domain, respectively.

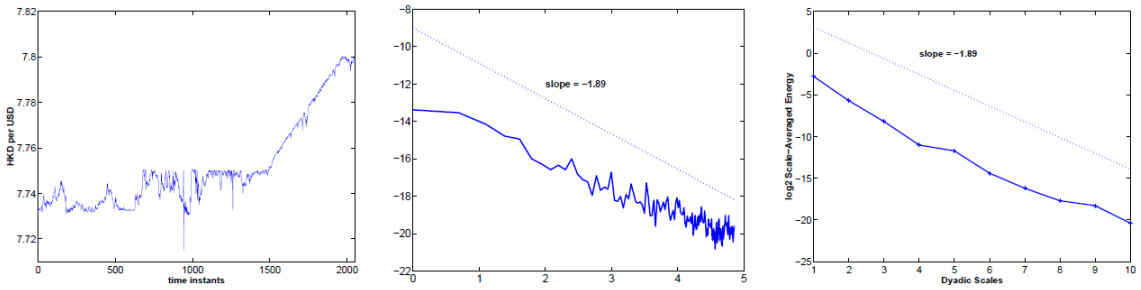


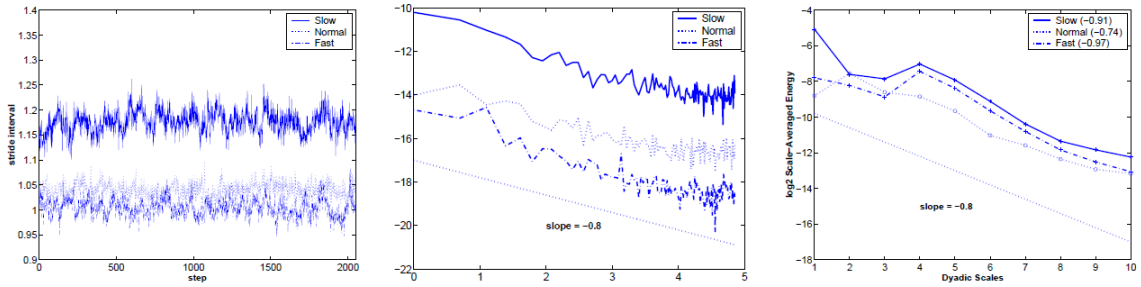

**Figure 2:** (a) Exchange Rates HKD per USD (b) scaling behavior in the Fourier domain, and (c) in the wavelet domain.

#### *1.1.1.3 Gait Data*

Scaling laws were recently detected in the apparently "noisy" variations in the stride interval (duration of the gait cycle) of human walking. Dynamic analysis of these step-to-step fluctuations revealed a self-similar pattern: Fluctuations at one time scale are statistically similar to those at multiple other time scales, at least over hundreds of steps, while healthy subjects walk at their normal rate. The experimental data are from healthy subjects walking for 1 hour at usual, slow, and fast paces.

The fluctuations of stride interval demonstrated long-range correlations with power-law decay for up to a thousand strides by every three walking rates. It is curious that during metronomically-paced walking, these long-range correlations disappeared; variations in the stride interval were anti-correlated. Experiments confirm that scaling behavior of spontaneous stride interval are normally quite robust and intrinsic to the locomotor system. Furthermore, this fractal property of neural output may be related to the higher nervous centers responsible for control of walking rhythm.

The gait data is collected from participants who had no history of any neuromuscular, respiratory or cardiovascular disorders, and were taking no medications. Average age of participants was 21.7 years and a range of age was 18-29 years. A range of height was $1.77 \pm 0.08$ meters (mean $\pm$ SD), and a range of weight was $71.8 \pm 10.7$ kg. Participants were asked to walk continuously on level ground around an obstacle free, length of either 225 or 400 meters, approximately oval path. The stride interval was collected using ultra-thin and force-sensitive switches attached inside one of shoes. Figure 3 shows 2048 data points collected for one subject. Slow and fast stride intervals have slopes of -0.91 and -0.97 while stride intervals for normal walk show scaling with -0.74 slope. Based on these we can confirm that the slopes depend on a degree of self-similarity.




(a) (b) (c)

**Figure 3:** (a) Gait timing for Slow, Normal and Fast Walk (b) scaling behavior in the Fourier domain, and (c) in the wavelet domain.

### 1.1.2 Definitions and Properties of Self-similar Process

We present the definitions and properties of self-similar processes. All stochastic processes discussed are real-valued and defined on the same parameter space.

**Definition 1.1.1.** If two stochastic processes $X(t)$ and $Y(t)$ are equal in all finite dimensional distributions, then they are denoted as $X(t) \stackrel{d}{=} Y(t)$.

The Definition 1.1.1 means that fir any selection of time, $0 \leq t_1 < t_2 < \cdots < t_k < \inf$, random vectors $(X(\omega, t_1), \ldots, (X(\omega, t_k))$ and $(Y(\omega, t_1), \ldots, (Y(\omega, t_k))$ have the same distribution. Informally, processes equal-in-distribution are statistically indistinguishable.

**Definition 1.1.2.** A stochastic processes $X(t)$ is called stochastically continuous at $t$ if $lim_{h \to 0} P(|X(t+h) - X(t)| > \epsilon) = 0$, for any fixed $\epsilon > 0$.

**Definition 1.1.3.** A stochastic processes $X(t)$, $t > 0$ is self-similar if for any $a > 0$, there exist $b > 0$ such that

$$X(at) \stackrel{d}{=} bX(t). \tag{1}$$

The Definitions 1.1.2 and 1.1.3 are building blocks for an understanding of self-similar processes. Lamperti (1962) also proved the result,

**Theorem 1.1.1.** *If a stochastic processes $X(t)$, $t > 0$ is non-trivial, stochastically continuous at $t = 0$, and self-similar, then there exists a unique $H > 0$ such that $b$ in the Definition 1.1.3 can be represented as $b = a^H$. If $X(0) = 0$, a.s. then $H > 0$.*

Then, standard definition of self-similar process is as follows.

**Definition 1.1.4.** A stochastic processes $X(t)$ is self-similar with self-similar index $H$ if and only if there exists $H > 0$ such that for any $a > 0$, $X(at) \stackrel{d}{=} a^H X(t)$.

Uniqueness of $H$ is not obvious from this definition although Lamperti's Theorem 1.1.1 guarantees uniqueness. Furthermore, it follows that $X(0) = 0$ almost surely from Definition 1.1.4.

## 1.2 Basics of Wavelets

### 1.2.1 Multiresolution Analysis

A multiresolution analysis of $L_2(\mathbb{R})$ is a nested sequence of closed subspaces, $V_j \subset L_2(\mathbb{R})$, lying in a containment hierarchy

$$V_j \subset V_{j+1}. \tag{2}$$

The interaction of the nested subspaces consists of only zeros functions and their union is dense in $L(\mathbb{R})$ such that

$$\bigcap_j V_j = \{\mathbf{0}\}, \ \overline{\bigcup_j V_j} = L_2(\mathbb{R}),$$

where $\overline{A}$ is the closure of a set $A$.

The hierarchy structure of Equation (2) is constructed based on two conditions. First, V-spaces are self-similar such that

$$f(2^j x) \in V_j \ \text{ iff } \ f(x) \in V_0.$$

Second, there is a scaling function $\phi \in V_0$ whose integer-translates span the space $V_0$ such that

$$V_0 = \Big\{ f \in L_2(\mathbb{R}) | f(x) = \sum_k c_k \phi(x-k) \Big\},$$

and for which the set $\{\phi(x-k), \ k \in \mathbb{Z}\}$ is an orthonormal basis for $V_0$ and, consequently, the set of functions

$$\phi_{j,k}(x) = 2^{\frac{j}{2}} \phi(2^j x - k)$$

is an orthonormal basis for $V_j$ where $j, k \in \mathbb{Z}$. Since $\phi \in V_0 \subset V_1$, a sequence of wavelet filter coefficients $h_k$ exists such that $\sum_{k\in\mathbb{Z}} h_k = \sqrt{2}$ and the function $\phi$ is represented as a linear combination of functions from $V_1$, i.e.,

$$\phi(x) = \sum_{k\in\mathbb{Z}} h_k \sqrt{2} \phi(2x-k). \tag{3}$$

This equation is referred to as *scaling function* and plays an fundamental role in constructing, exploring, and utilizing wavelets.

Similar to the $\phi_{j,k}(x)$, there is an orthonormal basis for $L_2(\mathbb{R})$,

$$\psi_{j,k}(x) = 2^{\frac{j}{2}}\psi(2^j x - k)$$

where $j, k \in \mathbb{Z}$ such that $\{\psi_{j,k}(x),\ j = \text{fixed},\ k \in \mathbb{Z}\}$ is an orthonormal basis for the difference space $W_j = V_{j+1} \ominus V_j$. Next, we derive $\psi(x)$ called a *wavelet function* based on the scaling function. Since $\psi \in W_0 \subset V_1$, the wavelet function can be represented as

$$\psi(x) = \sum_{k\in\mathbb{Z}} g_k \sqrt{2}\phi(2x - k), \tag{4}$$

for wavelet filter coefficients $g_k,\ k \in \mathbb{Z}$. Additionally, we can relate both $h_k$ and $g_k$, elements of wavelet filter $\mathbf{h}$ and $\mathbf{g}$ called the *quadrature mirror filters*, as follows:

$$g_k = (-1)^k h_{1-k}. \tag{5}$$

The relation in Equation (5) is well-known as the *quadrature mirror relation* in signal processing fields.

### 1.2.2 Mallat's Algorithm

After simple replacements for the indices in Equation (3) and (4) in section 1.2.1, we obtain

$$\phi_{j-1,l}(x) = \sum_{k\in\mathbb{Z}} h_{k-2l}\phi_{j,k}(x) \ \text{ and } \ \psi_{j-1,l}(x) = \sum_{k\in\mathbb{Z}} g_{k-2l}\psi_{j,k}(x). \tag{6}$$

The Equations in (6) are bases for building the cascade algorithm. Considering a multiresolution analysis in Equation (2), any function $v_j \in V_j$ can be uniquely defined as $v_j(x) = v_{j-1}(x) + w_{j-1}(x)$ where $v_{j-1} \in V_{j-1}$ and $w_{j-1} \in W_{j-1}$ since $V_j = V_{j-1} \oplus W_{j-1}$. It is usual to denote the coefficients corresponded with $\phi_{j,k}(x)$ and $\psi_{j,k}(x)$ by

$c_{j,k}$ and $d_{j,k}$, respectively. Then, we can obtain

$$\begin{aligned} v_j(x) &= \sum_k c_{j,k}\phi_{j,k}(x) \\ &= \sum_l c_{j-1,l}\phi_{j-1,l}(x) + \sum_l d_{j-1,l}\psi_{j-1,l}(x) \\ &= v_{j-1}(x) + w_{j-1}(x). \end{aligned}$$

Using the standard scaling function in Equation (6), orthogonality of $w_{j-1}(x)$ and $\phi_{j-1,l}(x)$ for any $j$ and $l$, and additivity of inner products, we can obtain the well-known fast wavelet decomposition algorithms as follows:

$$\begin{aligned} c_{j-1,l} &= \langle v_j, \phi_{j-1,l}\rangle \\ &= \langle v_j, \sum_{k\in\mathbb{Z}} h_{k-2l}\phi_{j,k}(x)\rangle \\ &= \sum_{k\in\mathbb{Z}} h_{k-2l}\langle v_j, \phi_{j,k}(x)\rangle \\ &= \sum_{k\in\mathbb{Z}} h_{k-2l}c_{j,k} \end{aligned} \tag{7}$$

and similarly $d_{j-1,l} = \sum_{k\in\mathbb{Z}} g_{k-2l}c_{j,k}$. The cascade algorithm is still effective in the reverse direction. We can obtain the coefficients located in the next finer scale corresponding to $V_j$ from the coefficients corresponding to $V_{j-1}$ and $W_{j-1}$. The equation

$$\begin{aligned} c_{j,k} &= \langle v_j, \phi_{j,k}\rangle \\ &= \sum_l c_{j-1,l}\langle \phi_{j-1,l}, \phi_{j,k}\rangle + \sum_l d_{j-1,l}\langle \psi_{j-1,l}, \phi_{j,k}\rangle \\ &= \sum_l c_{j-1,l}h_{k-2l} + \sum_l d_{j-1,l}g_{k-2l} \end{aligned}$$

represents a single step in the inverse wavelet transform algorithm for reconstruction.

## *1.3 Discrete Wavelet Transformation*

To obtain discrete outputs, we can apply discrete wavelet transformations (DWT) to the discrete datasets. The procedures of performing DWT resembles performing the fast Fourier transformation (FFT), which is the Fourier method for a set of discrete data.

**Table 1:** The analogy between Fourier and wavelet methods.

| Fourier Methods | Fourier Integrals | Fourier Series | Discrete Fourier Transformations |
|---|---|---|---|
| Wavelet Methods | Continuous Wavelet Transformations | Wavelet Series | Discrete Wavelet Transformations |

The original discrete input data in time domain can be mapped to a vector of the same size in wavelet domain using discrete wavelet transformations. Since such process is linear, it also can be implemented by a matrix multiplication. For example, a matrix of size $n \times n$ can be multiplied to one dimensional input data of size $n$ for the discrete wavelet transformation. Depending on boundary conditions, such matrices can be either orthogonal or "close" to orthogonal. When the transformation matrix is orthogonal, such process is a rotation in $\mathbb{R}^n$ space in which an input data represents coordinates of a single point. Finally, the discrete wavelet transformation of the original data is the new coordinates in the rotated space.

**Example 1.3.1.** Let an input data be $\{1, 2\}$ and let $M(1, 2)$ be the point in $\mathbb{R}^2$ with coordinates given by the data vector. DWT with the Haar wavelet basis can be performed via rotating the coordinate axes by an angle of $\frac{\pi}{4}$. The rotation matrix is

$$W = \begin{pmatrix} \cos\frac{\pi}{4} & \sin\frac{\pi}{4} \\ \cos\frac{\pi}{4} & -\sin\frac{\pi}{4} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix},$$

and the discrete wavelet transformation of $(-1, 2)'$ is $W \cdot (-1, 2)' = (\frac{1}{\sqrt{2}}, -\frac{3}{\sqrt{2}})'$. It is notable that *the energy* (squared distance of the point from the origin) is preserved, $(-1)^2 + (2)^2 = (\frac{1}{\sqrt{2}})^2 + (-\frac{1}{\sqrt{2}})^2$, since $W$ is a rotation.

**Example 1.3.2.** Let $\mathbf{y} = (1, 0, -\frac{5}{2}, \frac{3}{2}, \frac{1}{2}, -\frac{1}{2}, \frac{3}{2}, \frac{5}{2})$. The values $f(n) = y_n, n = 0, 1, \ldots, 7$ are interpolated by the father wavelet from the Haar wavelet, the vector represent the sampled piecewise constant function. It is clear that such defined

$f$ belongs to Haar's multiresolution space $V_0$. The following matrix equation links $\mathbf{y}$ and the wavelet coefficients.

$$\begin{bmatrix} 1 \\ 0 \\ -\frac{5}{2} \\ \frac{3}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \\ \frac{3}{2} \\ \frac{5}{2} \end{bmatrix} = \begin{bmatrix} \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2} & 0 & \frac{1}{\sqrt{2}} & 0 & 0 & 0 \\ \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2} & 0 & -\frac{1}{\sqrt{2}} & 0 & 0 & 0 \\ \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & -\frac{1}{2} & 0 & 0 & \frac{1}{\sqrt{2}} & 0 & 0 \\ \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & -\frac{1}{2} & 0 & 0 & -\frac{1}{\sqrt{2}} & 0 & 0 \\ \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & 0 & \frac{1}{2} & 0 & 0 & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & 0 & -\frac{1}{2} & 0 & 0 & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & 0 & -\frac{1}{2} & 0 & 0 & 0 & -\frac{1}{\sqrt{2}} \end{bmatrix} \cdot \begin{bmatrix} c_{00} \\ d_{00} \\ d_{10} \\ d_{11} \\ d_{20} \\ d_{21} \\ d_{22} \\ d_{23} \end{bmatrix}$$

The solution is

$$\begin{bmatrix} c_{00} \\ d_{00} \\ d_{10} \\ d_{11} \\ d_{20} \\ d_{21} \\ d_{22} \\ d_{23} \end{bmatrix} = \begin{bmatrix} \sqrt{2} \\ -\sqrt{2} \\ 1 \\ -2 \\ \frac{1}{\sqrt{2}} \\ -2\sqrt{2} \\ \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$$

Thus,

$$\begin{aligned} f &= \sqrt{2}\phi_{-3,0} - \sqrt{2}\psi_{-3,0} + \psi_{-2,0} - 2\psi_{-2,1} \\ &\quad + \frac{1}{\sqrt{2}}\psi_{-1,0} - 2\sqrt{2}\psi_{-1,1} + \frac{1}{\sqrt{2}}\psi_{-1,2} - \frac{1}{\sqrt{2}}\psi_{-1,3}. \end{aligned}$$

The solution is easy to prove. For example, when $x \in [0, 1)$,

$$f(x) = \sqrt{2} \cdot \frac{1}{2\sqrt{2}} - \sqrt{2} \cdot \frac{1}{2\sqrt{2}} + 1 \cdot \frac{1}{2} + \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} = \frac{1}{2} + \frac{1}{2} = 1 = y_0.$$

Such multiplication process for DWT is conceptually straightforward, but of limited practical value. Storing and manipulating transformation matrices when inputs are long ($n > 10^{15}$) may not even be feasible. However, such matrix transforms are an excellent tool when handling images of moderate size as explained in later sections.

### 1.3.1 Matrix-based Discrete Wavelet Transformation

The change of basis in $V_1$ from $\mathcal{B}_1 = \{\phi_{1k}(x), k \in Z\}$ to $\mathcal{B}_2 = \{\phi_{0k}(x), k \in Z\} \cup \{\psi_{0k}(x), k \in Z\}$ can be implemented via matrix multiplication, thus, we can define discrete wavelet transformation by matrices. Example 1.3.2 presented an example of transformation matrix corresponding to Haar's inverse transformation. Suppose that the length of the input signal is $2^J$, the wavelet filter is $\mathbf{h} = \{h_s, s \in \mathbb{Z}\}$, and $N$ is a constant. Denote $H_k$ by a matrix of size $(2^{J-k} \times 2^{J-k+1}), k = 1, \ldots$ at position $(i, j)$ with entries

$$h_s, s = (N-1) + (i-1) - 2(j-1) \text{ modulo } 2^{J-k+1}. \tag{8}$$

Note that $H_k$ is the circulant matrix, which means that its $i$th row is the circularly shifted to the right by $2(i-1)$ units. This circularity is derived from the *modulo* operator in (8).

By analogy, we can define a matrix $G_k$ by using the filter g. The $G_k$ is connected to the pre-defined $H_k$ and can be obtained by replacing $h_i$ with $(-1)^i h_{N+1-i}$. The constant $N$ is a shift parameter and affects the position of the wavelet on the time scale.

The unitary matrix $\begin{bmatrix} H_k \\ G_k \end{bmatrix}$ is a basis-change matrix in $2^{J-k+1}$ dimensional space. Therefore,

$$I_{2^{J-k}} = \begin{bmatrix} H_k' & G_k' \end{bmatrix} \begin{bmatrix} H_k \\ G_k \end{bmatrix} = H_k' \cdot H_k + G_k' \cdot G_k$$

and

$$I_{2^{J-k}} = \begin{bmatrix} H_k \\ G_k \end{bmatrix} \begin{bmatrix} H_k' & G_k' \end{bmatrix} = \begin{bmatrix} H_k \cdot H_k' & H_k \cdot G_k' \\ G_k \cdot H_k' & G_k \cdot G_k' \end{bmatrix}.$$

This implies that $H_k \cdot H_k' = I,\ G_k \cdot G_k' = I,\ G_k \cdot H_k' = H_k \cdot G_k' = 0,$ and $H_k' \cdot H_k + G_k' \cdot G_k = I$. Now, for a sequence $\mathbf{y}$ the J-step wavelet transformation is $\mathbf{d} = W_J \cdot \mathbf{y}$, where

$$W_1 = \begin{bmatrix} H_1 \\ G_1 \end{bmatrix},\ W_2 = \begin{bmatrix} \begin{bmatrix} H_2 \\ G_2 \end{bmatrix} \cdot H_1 \\ G_1 \end{bmatrix},\ W_3 = \begin{bmatrix} \begin{bmatrix} \begin{bmatrix} H_3 \\ G_3 \end{bmatrix} \cdot H_2 \\ G_2 \end{bmatrix} \cdot H_1 \\ G_1 \end{bmatrix}, \ldots$$

**Example 1.3.3.** Suppose that $\mathbf{y} = (1, 0, -\frac{5}{2}, \frac{3}{2}, \frac{1}{2}, -\frac{1}{2}, \frac{3}{2}, \frac{5}{2})$ and filter is $\mathbf{h} = (h_0, h_1, h_2, h_3) = \left(\frac{1+\sqrt{3}}{4\sqrt{2}}, \frac{3+\sqrt{3}}{4\sqrt{2}}, \frac{3-\sqrt{3}}{4\sqrt{2}}, \frac{1-\sqrt{3}}{4\sqrt{2}}\right)$. Then, $J = 3$ and matrices $H_k$ and $G_k$ are of dimension $2^{3-k} \times 2^{3-k+1}$ with

$$H_1 = \begin{bmatrix} h_1 & h_2 & h_3 & 0 & 0 & 0 & 0 & h_0 \\ 0 & h_0 & h_1 & h_2 & h_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & h_0 & h_1 & h_2 & h_3 & 0 \\ h_3 & 0 & 0 & 0 & 0 & h_0 & h_1 & h_2 \end{bmatrix}$$

and

$$G_1 = \begin{bmatrix} -h_2 & h_1 & -h_0 & 0 & 0 & 0 & 0 & h_3 \\ 0 & h_3 & -h_2 & h_1 & -h_0 & 0 & 0 & 0 \\ 0 & 0 & 0 & h_3 & -h_2 & h_1 & -h_0 & 0 \\ -h_0 & 0 & 0 & 0 & 0 & h_3 & -h_2 & h_1 \end{bmatrix}.$$

Since,

$$\begin{aligned} H_1 \cdot \mathbf{y} &= \{1.2089, -0.3239, -0.0657, 2.6818\}, \\ G_1 \cdot \mathbf{y} &= \{0.6470, 0.0947, 0.0820, -2.9451\}, \end{aligned}$$

so

$$W_1 \cdot \mathbf{y} = \{1.2089, -0.3239, -0.0657, 2.6818 \mid 0.6470, 0.0947, 0.0820, -2.9451\}.$$

Next,

$$H_2 = \begin{bmatrix} h_1 & h_2 & h_3 & h_0 \\ h_3 & h_0 & h_1 & h_2 \end{bmatrix}, \quad G_2 = \begin{bmatrix} -h_2 & h_1 & -h_0 & h_3 \\ -h_0 & h_3 & -h_2 & h_1 \end{bmatrix}.$$

In this example, we can perform discrete wavelet transformation for two steps only, $W_1$ and $W_2$ due to lengths of the filter and data.

The two-step DAUB2 discrete wavelet transformation of $\mathbf{y}$ is

$$W_2 \cdot \mathbf{y} = \{1.3125, 0.6875 \mid -0.0658, 2.6818 \mid 0.6470, 0.0947, 0.0820, -2.9451\},$$

because

$$\begin{aligned} H_2 \cdot H_1 \cdot \mathbf{y} &= H_2 \cdot \{1.2089, -0.3239, -0.0657, 2.6818\} \\ &= \{1.3125, 0.6875\} \\ G_2 \cdot H_1 \cdot \mathbf{y} &= G_2 \cdot \{1.2089, -0.3239, -0.0657, 2.6818\} \\ &= \{-0.0658, 2.6818\}. \end{aligned}$$

## *1.4 Traditional Orthogonal Wavelet Transform*

The functional representations of traditional orthogonal wavelet transform are presented in this section. Traditionally, we implement a wavelet transform with a convolution of an input data and wavelet/scaling functions. Assuming that a multiresolution framework is specified, scaling and wavelet functions are represented as $\phi$ and $\psi$, respectively. Then, we can express a data vector $\boldsymbol{y} = (y_0, y_1, \ldots, y_{m-1})$ of size $m$ as a function $f$ in context of shifts of the scaling function at some multiresolution level $J$ as follows:

$$f(x) = \sum_{k=0}^{m-1} y_k \phi_{J,k}(x)$$

where $J - 1 < \log_2 m \leq J$ and $\phi_{J,k}(x) = 2^{\frac{J}{2}} \phi(2^J x - k)$. Next, we also can represent the data interpolating function $f$ as

$$f(x) = \sum_{k=0}^{2^{J_0}-1} c_{J_0,k} \phi_{J_0,k}(x) + \sum_{j=J_0}^{J-1} \sum_{k=0}^{2^j-1} d_{j,k} \psi_{j,k}(x)$$

where

$$\phi_{J_0,k}(x) = 2^{\frac{J_0}{2}}\phi\left(2^{J_0}x - k\right),$$
$$\psi_{j,k}(x) = 2^{\frac{j}{2}}\psi\left(2^{j}x - k\right),$$

and $J_0$ is the coarsest level. Then, the discrete wavelet transforms of vector $\boldsymbol{y}$ are composed of wavelet coefficients, $\boldsymbol{c}^{(J_0)} = (c_{J_0,0}, c_{J_0,1}, \ldots, c_{J_0,2^{J_0}-1})$ and $\boldsymbol{d}^{(j)} = (d_{j,0}, d_{j,1}, \ldots, d_{j,2^j-1})$ where $j = J_0, \ldots, J-1$ for a decomposition depth $p = J - J_0$. To be specific, $\boldsymbol{c}^{(J_0)}$ contains coarse fluctuation within an input and called coarse coefficients, while $\boldsymbol{d}^{(j)}$ contains detail information within an input and called detail coefficients. Since the shift indicator $k$ depends on a level as $2^{-j}k$, the number of wavelet coefficients decreases by half as the level decreases. Consequently, the total number of wavelet coefficients is $m$ that is identical to that of original data vector $\boldsymbol{y}$.

### 1.4.1 Standard 2-D Wavelet Transform

Next, we extend the 1-D definitions to the standard and scale-mixing 2-D wavelet transforms of $f(x, y)$ where $(x, y) \in \mathbb{R}^2$. The representation of wavelets in 2-D can be implemented through one scaling function and three wavelet functions as follows:

$$\begin{aligned}
\phi(x, y) &= \phi(x)\phi(y), \\
\psi^{(h)}(x, y) &= \phi(x)\psi(y), \\
\psi^{(v)}(x, y) &= \psi(x)\phi(y), \\
\psi^{(d)}(x, y) &= \psi(x)\psi(y),
\end{aligned} \tag{9}$$

where symbols $h, v, d$ denote the horizontal, vertical, and diagonal directions, respectively.

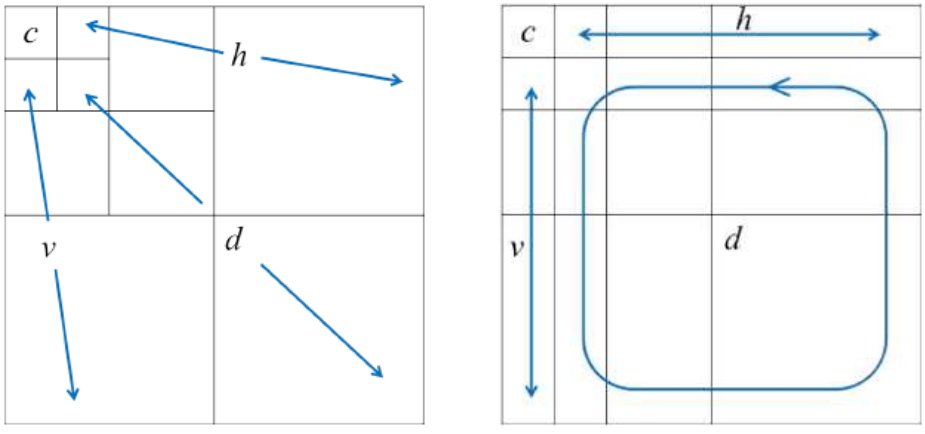
For the standard 2-D wavelet transform, we define the wavelet atoms as follows:

$$\begin{aligned}
\phi_{J_0,k_1,k_2}(x,y) &= 2^{J_0}\phi(2^{J_0}x-k_1)\phi(2^{J_0}y-k_2),\\
\psi^{(h)}_{j,k_1,k_2}(x,y) &= 2^{j}\phi(2^{j}x-k_1)\psi(2^{j}y-k_2),\\
\psi^{(v)}_{j,k_1,k_2}(x,y) &= 2^{j}\psi(2^{j}x-k_1)\phi(2^{j}y-k_2),\\
\psi^{(d)}_{j,k_1,k_2}(x,y) &= 2^{j}\psi(2^{j}x-k_1)\psi(2^{j}y-k_2),
\end{aligned}$$

where $(k_1,k_2)\in\mathbb{Z}^2$ is the location pair and $j=J_0,\ldots,J-1$ with depth $p=J-1-J_0$. Then, the wavelet coefficients of $f(x,y)$ are computed as

$$\begin{aligned}
c_{J_0,k_1,k_2} &= \iint f(x,y)\bar{\phi}_{J_0,k_1,k_2}(x,y)\,dxdy,\\
d^{(h)}_{j,k_1,k_2} &= \iint f(x,y)\bar{\psi}^{(i)}_{j,k_1,k_2}(x,y)\,dxdy,\\
d^{(v)}_{j,k_1,k_2} &= \iint f(x,y)\bar{\psi}^{(i)}_{j,k_1,k_2}(x,y)\,dxdy,\\
d^{(d)}_{j,k_1,k_2} &= \iint f(x,y)\bar{\psi}^{(i)}_{j,k_1,k_2}(x,y)\,dxdy,
\end{aligned}$$

where $\bar{\phi}$ and $\bar{\psi}$ are complex conjugates of $\phi$ and $\psi$, respectively. The tessellation of a standard 2-D wavelet transform is illustrated in Figure 4(a).




(a) (b)

**Figure 4:** (a) Tessellations for standard 2-D WT of depth 3 (b) Tessellations for scale-mixing 2-D WT of depth 3.

### 1.4.2 Scale-mixing 2-D Wavelet Transform

Compared to the traditional 2-D wavelet transform, the scale-mixing 2-D wavelet transform is better for compressibility, which is desired in dimension reduction applications (Remenyi et al, 2014). For the scale-mixing 2-D wavelet transform, we define the wavelet atoms as follows:

$$\begin{aligned}
\phi_{J_{01},J_{02},k_1,k_2}(x,y) &= 2^{\frac{J_{01}+J_{02}}{2}}\phi(2^{J_{01}}x-k_1)\phi(2^{J_{02}}y-k_2),\\
\psi^{(h)}_{J_{01},j_2,k_1,k_2}(x,y) &= 2^{\frac{J_{01}+j_2}{2}}\phi(2^{J_{01}}x-k_1)\psi(2^{j_2}y-k_2),\\
\psi^{(v)}_{j_1,J_{02},k_1,k_2}(x,y) &= 2^{\frac{j_1+J_{02}}{2}}\psi(2^{j_1}x-k_1)\phi(2^{J_{02}}y-k_2),\\
\psi^{(d)}_{j_1,j_2,k_1,k_2}(x,y) &= 2^{\frac{j_1+j_2}{2}}\psi(2^{j_1}x-k_1)\psi(2^{j_2}y-k_2),
\end{aligned}$$

where $(k_1,k_2)\in\mathbb{Z}^2$, $j_1\geq J_{01}; j_2\geq J_{02}$ and $J_{01}$ and $J_{02}$ are the coarsest decomposition levels of rows and columns. Then, any function $f\in L_2(\mathbb{R}^2)$ can be expressed as

$$\begin{aligned}
f(x,y) &= \sum_{k_1}\sum_{k_2} c_{J_{01},J_{02},k_1,k_2}\phi_{J_{01},J_{02},k_1,k_2}(x,y)\\
&+ \sum_{j_2>J_{02}}\sum_{k_1}\sum_{k_2} d_{J_{01},j_2,k_1,k_2}\psi^{(h)}_{J_{01},j_2,k_1,k_2}(x,y)\\
&+ \sum_{j_1>J_{01}}\sum_{k_1}\sum_{k_2} d_{j_1,J_{02},k_1,k_2}\psi^{(v)}_{j_1,J_{02},k_1,k_2}(x,y)\\
&+ \sum_{j_1>J_{02}}\sum_{j_2>J_{01}}\sum_{k_1}\sum_{k_2} d_{j_1,j_2,k_1,k_2}\psi^{(d)}_{j_1,j_2,k_1,k_2}(x,y),
\end{aligned} \tag{10}$$

which defines a scale-mixing wavelet transform. Unlike the traditional 2-D wavelet transform, the scale-mixing atoms mix such indices that capture the energy flux between the scales. Notice that the pair $(j_1,j_2)$ can be indexed as $(j_1,j_1+s)$ with $s\in\mathbb{Z}$ as well. Finally, the resulting scale-mixing wavelet coefficients are

$$\begin{aligned}
c_{J_{01},J_{02},k_1,k_2} &= \iint f(x,y)\phi_{J_{01},J_{02},k_1,k_2}(x,y)\,dxdy,\\
d^{(h)}_{J_{01},j_2,k_1,k_2} &= \iint f(x,y)\psi^{(h)}_{J_{01},j_2,k_1,k_2}(x,y)\,dxdy,\\
d^{(v)}_{j_1,J_{02},k_1,k_2} &= \iint f(x,y)\psi^{(v)}_{j_1,J_{02},k_1,k_2}(x,y)\,dxdy,\\
d^{(d)}_{j_1,j_2,k_1,k_2} &= \iint f(x,y)\psi^{(d)}_{j_1,j_2,k_1,k_2}(x,y)\,dxdy,
\end{aligned}$$

where $\bar{\phi}$ and $\bar{\psi}$ are complex conjugates of $\phi$ and $\psi$, respectively. Notice that we handle a pair $(j_1, j_2)$ indicating that two scales are mixed in the scale-mixing 2-D wavelet transform, while we handle a single $j$ to denote a scale in the traditional 2-D wavelet transform.

For the 2-D real-valued scale-mixing transform, $G = WAW^T$ is defined as the wavelet-transformed data of image $A$. We generalize this definition for the 2-D complex-valued scale-mixing transform here. First, let us transform the rows of $A$ through a wavelet matrix $W$, which results in $WA^\dagger$. Next, repeat a same work on the rows of $WA^\dagger$ and then we obtain the scale-mixing wavelet transform of matrix $A$ as

$$G = W(WA^\dagger)^\dagger = WAW^\dagger \tag{11}$$

where $W^\dagger$ stands for the Hermitian transpose of a complex-valued wavelet matrix $W$. Since the unitary characteristic of $W$ in Equation (11) implies $E = \text{trace}(AA^\dagger) = \text{trace}(GG^\dagger)$, the total energy, $E$, in image $A$ is preserved. Note that the relationship in Equation (11) implies a 2-D implementation of (10) for signal $f(x)$ sampled in a form of matrix $A$. The tessellation of a scale-mixing 2-D wavelet transform is illustrated in Figure 4(b) and we can identify some hierarchies of the detail spaces. The diagonal hierarchy indicates coefficients with the same component scales and has a same hierarchy in the standard 2-D transform in Figure 4(a). In addition, the hierarchies that are one level above and below the diagonal hierarchy represent detail spaces with the scales differed by 1. To be specific, the scales along $x$-axis are faced with the next coarser scale along $y$-axis for the hierarchy above the diagonal, and vice versa for the hierarchy below the diagonal.

## *1.5 Complex Wavelet Transform*

### 1.5.1 Daubechies Complex Wavelet Basis

The Daubechies complex wavelets are implemented via the symmetric Daubechies wavelets. In this section, we present some conditions of the symmetric Daubechies wavelets.

1. Compactness of the support of $\phi$: It is required for $\phi$ and $\psi$ to have a compact support within the interval $[-J, J+1]$ for integer value $J$ that is, $h_k \neq 0$ for $k = -J, -J+1, \ldots, J, J+1$.

2. Orthogonality of the $\phi(x-k)$: This constraint defines the Daubechies wavelets in a broad sense. The polynomial is written as

$$F(z) = \sum_{n=-J}^{J+1} \frac{1}{\sqrt{2}} h_n z^n, \quad \text{with} \quad F(1) = 1, \tag{12}$$

where $z$ is on the unit circle, $|z| = 1$. Then, the orthonormality constraint of the set $\{\phi_{0,k}(x), k \in \mathbb{Z}\}$ can be satisfied with the equation as

$$P(z) - P(-z) = z \tag{13}$$

where the polynormial $P(z)$ is stated as

$$P(z) = zF(z)\overline{F(z)}.$$

3. Strang-Fix condition: The vanishing of the first $J$ moments of the wavelets is required to maximize the regularity of the functions generated by $\phi$. In other words, the Strang-Fix condition can be stated as

$$F(-1) = F'(-1) = F''(-1) = \cdots = F^{(J)}(-1) = 0 \tag{14}$$

in the polynomial aspect. Note that the real polynomial solutions of the Equation (13) and (14) leads to the standard Daubechies wavelets and the degree of symmetry makes differences within them. Generally, the standard Daubechies wavelets is considered as the least asymmetric solution.

*4.* Symmetry: The following constraint have to be on the filter, i.e. $h_k = h_{1-k}$,

$$F(z) = zF(z^{-1}).$$

A solution of the full conditions was investigated by Daubechies through an inspection of the roots of the following polynomial satisfying the Equation (13) defined by

$$P_J(z) = (\frac{1+z}{2})^{2J+2} p_J(z^{-1})$$

where

$$p_J(z) = \sum_{j=0}^{2J} r_j (z+1)^{2J-j}(z-1)^j, \text{ with } \begin{cases} r_{2j} = (-1)^j 2^{-2J}\binom{2J+1}{j} \\ r_{2j+1} = 0 \end{cases}$$

where $j = 0, 1, \ldots, J$. The proof that the polynomial $P_J(z)$ satisfies the Equation (13) is provided in Appendix A.1.

The appropriacy of the polynomial $P_J(z)$ in building multiresolution bases depends on the notable relationship between $2J$ roots of $p_J(z)$ because they show distinct symmetries. To be concrete, the inverse and conjugate of a root are also roots and no root is of unit modulus. When $x_{k=1,2,\ldots,J}$ are the J roots inside the unit circle ($|x_k| < 1$) and $\bar{x}_k = x_{J+1-k}$, we can define the factorization of $p_J(z)$ as

$$p_J(z) = \prod_{k=1}^{J} \left(\frac{z - x_k}{1 - x_k}\right) \prod_{k=1}^{J} \left(\frac{z - \bar{x}_k^{-1}}{1 - \bar{x}_k^{-1}}\right)$$

and the low-pass filter $F(z)$ as

$$F(z) = \left(\frac{1+z}{2}\right)^{1+J} p(z^{-1}) \text{ with } p(z) = \prod_{m \in R} \left(\frac{z - x_m}{1 - x_m}\right) \prod_{n \in \bar{R}} \left(\frac{z - \bar{x}_n^{-1}}{1 - \bar{x}_n^{-1}}\right). \quad (15)$$

The polynomial $p(z)$ can be stated by particular roots of $p_J(z)$ with $R, \bar{R} \subset \{1, 2, \ldots, J\}$. In addition, $p_J(z) = z^J p(z^{-1})\overline{p(z)}$ is defined through the spectral factorization of $P_J(z) = zF(z)\bar{F}(z)$, which leads to the constraint on $R$ and $\bar{R}$ as follows:

$$k \in R \Leftrightarrow k \notin \bar{R}. \quad (16)$$

Note that replacing $R$ with $\bar{R}$ is equivalent to converting a wavelet filter coefficient $h_k$ to its complex conjugate. Thus, the polynomial defined by

$$m_0(w) = \left(\frac{1+e^{iw}}{2}\right) \prod_{k\in R} \left(\frac{e^{-iw}-x_k}{1-x_k}\right) \prod_{k\notin R} \left(\frac{e^{-iw}-\bar{x}_k^{-1}}{1-\bar{x}_k^{-1}}\right) \tag{17}$$

with an arbitrary choice of $R$ of roots of $p_J(z)$ implies an admissible trigonometric polynomial and consequently a compactly supported orthonormal dyadic wavelet basis. This choice of roots satisfies the conditions *1*, *2*, *3*. With the three conditions, Dabechies investigated real-valued solutions of Equation (17) for $h_k$ by a new rule derived from the constraint (16) as

$$k \in R \Leftrightarrow J-k+1 \in R \text{ and } k \notin \bar{R}, \tag{18}$$

which leads to real coefficients in $m_0(w) = \sum_k \frac{1}{\sqrt{2}} h_k e^{iwk}$. Consequently the DAUBn solution with $n = 2J+2$ leads to the usual Daubechies wavelets that have real-valued functions. Note that the number of solutions is $2^{J-1}$ that is half of the whole solutions ($2^J$) except for the Haar case ($J = 0, h_0 = \frac{\sqrt{2}}{2}, h_1 = \frac{\sqrt{2}}{2}$). We provide some solutions with the first values of $J$ with the rule (18).

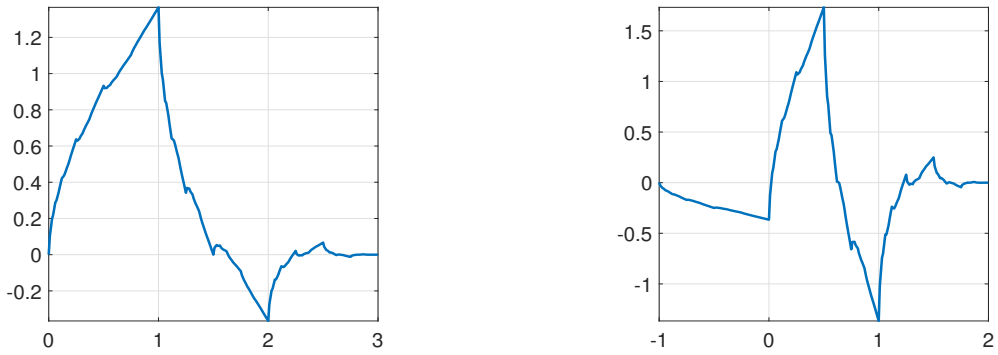
*Example 1.* J=1: The polynomial $p_1(z)$ has two roots $x_1 = r = 2-\sqrt{3}$ and $x_1^{-1}$. Then, the unique solution is

$$m_0(w) = \left(\frac{1+e^{iw}}{2}\right)^2 \left(\frac{e^{-iw}-r}{1-r}\right)$$

corresponding to the DAUB4 wavelets whose scaling and wavelet functions are illustrated in Figure 5.

*Example 2.* J=2: The polynomial $p_2(z)$ has four roots as

$$x_1 = \frac{3}{2} - i\sqrt{\frac{5}{12}} - \frac{1}{2}\sqrt{\frac{10}{3} - 2i\sqrt{15}},$$
$$x_1 = \frac{3}{2} - i\sqrt{\frac{5}{12}} + \frac{1}{2}\sqrt{\frac{10}{3} - 2i\sqrt{15}}$$




(a) (b)

**Figure 5:** (a) The scaling function $\phi$ of DAUB4 wavelets with $J = 1$. (b) The wavelet function $\psi$ of DAUB4 wavelets with $J = 1$.

and $\bar{x}_1^{-2}, \bar{x}_2^{-2}$. Two particular cases are identified: $R = \{1, 2\}$ and $R = \{1\}$. We consider the $R = \{1, 2\}$ case for real-valued wavelets here and then the $R = \{1\}$ case later in this section for complex-valued wavelets. The solution with the $R = \{1, 2\}$ is

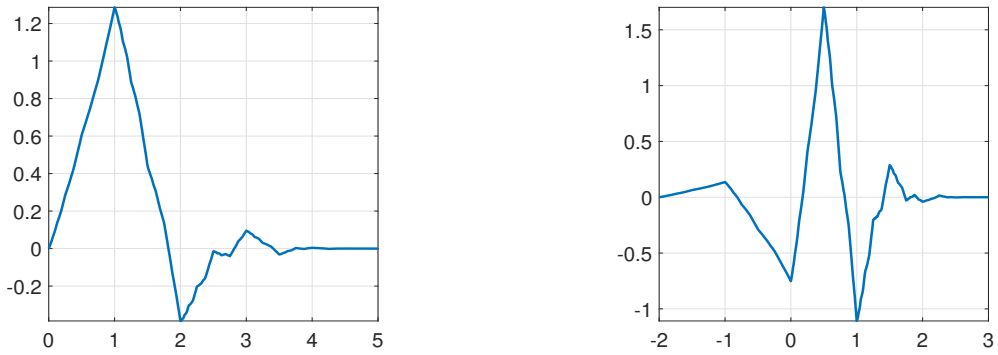
$$m_0(w) = \left(\frac{1 + e^{iw}}{2}\right)^3 \left(\frac{e^{-iw} - x_1}{1 - x_1}\right)\left(\frac{e^{-iw} - x_2}{1 - x_2}\right)$$

corresponding to the DAUB6 wavelets whose scaling and wavelet functions are illustrated in Figure 6.

With the three conditions *1*, *2*, *3* as in the above real solution cases, addition of the symmetry condition *4* provides to an other subset of solutions of Equation (15). Similar to the rule (18), an additional new selection rule of the roots can be derived as

$$k \in R \Leftrightarrow J - k + 1 \in \bar{R} \text{ and } k \notin \bar{R}. \tag{19}$$

Note that the rule (18) and the rule (19) cannot be compatible since the latter rule produces the complex-valued symmetric Daubechies wavelets. We obtain $2^{\frac{J}{2}}$ solutions in the original set of "orthogonal compactly supported regular wavelets" ($2^J$ components, complex or real) for any even value of $J$ and call them SDAUBn solutions with




(a) (b)

**Figure 6:** (a) The real scaling function $\phi$ of DAUB6 wavelets with $J = 2$ and selection of the $R = \{1, 2\}$. (b) The real wavelet function $\psi$ of DAUB6 wavelets with $J = 2$ and selection of the $R = \{1, 2\}$.

$n = 2J + 2$ leading to the symmetric Daubechies wavelets that have complex-valued functions.

Consequently, the complex scaling and wavelet functions can be defined as

$$\phi(x) = h(x) + ig(x) \quad \text{and} \quad \psi(x) = w(x) + iv(x)$$

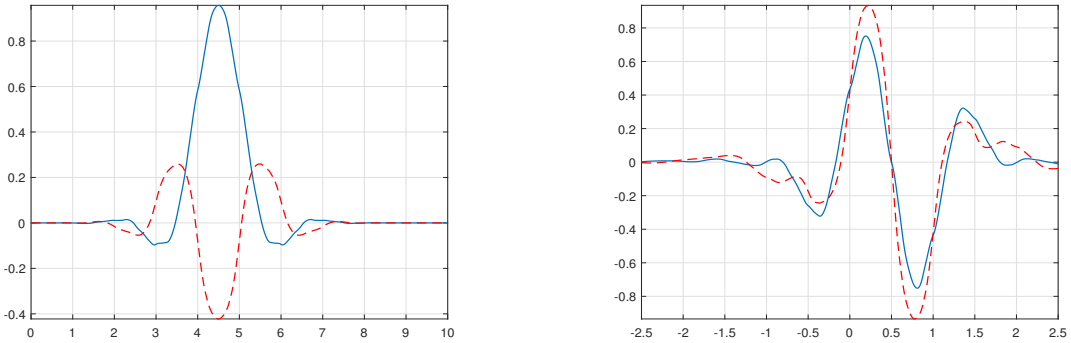
where $h, g, w$ and $v$ are all real-valued functions. Note that a complex conjugate of a scaling function is also a scaling function. We present an examined solution with the first values of $J$ with the rule (19) following the above example with the $R = \{1\}$ case.

*Example 1.* J=2: The solution with the $R = \{1\}$ is

$$m_0(w) = \left(\frac{1 + e^{iw}}{2}\right)^3 \left(\frac{e^{-iw} - x_1}{1 - x_1}\right) \left(\frac{e^{-iw} - \bar{x}_2^{-1}}{1 - \bar{x}_2 - 1}\right)$$

corresponding to the symmetric but complex-valued wavelets whose complex scaling and wavelet functions are illustrated in Figure 7. Such complex solutions can be obtained for any $J \geq 2$, however, only even values of $J$ can satisfy the symmetry condition *4*. Complex filter banks for the filter coefficients $h_k$ with $J = 2, 4, 6$ are given in Table 2. For simplicity,

we abbreviate the symmetric Daubechies Wavelets (SDAUBn) as the Complex Wavelets (CWn) from the next section.



(a) (b)

**Figure 7:** (a) The complex scaling function $\phi$ of SDAUB6 wavelets with $J = 2$ and selection of the $R = \{1\}$. (b) The complex wavelet function $\psi$ of SDAUB6 wavelets with $J = 2$ and selection of the $R = \{1\}$. The blue solid line indicates the real part and the red dashed line indicates the imaginary part for both figure (a) and (b).

**Table 2:** Complex filter banks with SDAUBn ($n = 2J + 2$) multiresolution analysis with $J = 2, 4, 6$, respectively.

| | SDAUB6, $J = 2$ | SDAUB10, $J = 4$ | SDAUB14, $J = 6$ |
|---|---|---|---|
| $h_1 = h_0$ | $0.662912 + 1.171163i$ | $0.643003 + 0.182852i$ | $0.633885 + 0.179835i$ |
| $h_2 = h_{-1}$ | $0.110485 - 0.085581i$ | $0.151379 - 0.094223i$ | $0.171512 - 0.094452i$ |
| $h_3 = h_{-2}$ | $-0.066291 - 0.085581i$ | $-0.080639 - 0.117947i$ | $-0.086478 - 0.130859i$ |
| $h_4 = h_{-3}$ | | $-0.017128 + 0.008728i$ | $-0.030746 + 0.014044i$ |
| $h_5 = h_{-4}$ | | $0.010492 + 0.020590i$ | $0.017651 + 0.037601i$ |
| $h_6 = h_{-5}$ | | | $0.0032388 - 0.001300i$ |
| $h_7 = h_{-6}$ | | | $-0.001956 - 0.004869i$ |

### 1.5.2 Mallat's Algorithm with Complex Wavelet Basis

Using one of the complex filters in Table 2, we can obtain the wavelet decomposition algorithms as follows:

$$\begin{aligned} c_{j-1,l} &= \langle v_j, \phi_{j-1,l} \rangle \\ &= \langle v_j, \sum_{k\in\mathbb{Z}} \bar{h}_{k-2l}\phi_{j,k}(x) \rangle \\ &= \sum_{k\in\mathbb{Z}} \bar{h}_{k-2l}\langle v_j, \phi_{j,k}(x) \rangle \\ &= \sum_{k\in\mathbb{Z}} \bar{h}_{k-2l} c_{j,k} \end{aligned}$$

and similarly $d_{j-1,l} = \sum_{k\in\mathbb{Z}} \bar{g}_{k-2l} c_{j,k}$. Note that the complex conjugate $\bar{h}_{k-2l}$ and $\bar{g}_{k-2l}$ are used instead of $h_{k-2l}$ and $g_{k-2l}$ in Equation (7) for traditional Mallat's algorithm since we consider complex wavelets. The cascade algorithm is still effective in the reverse direction. We can obtain the coefficients located in the next finer scale corresponding to $V_j$ from the coefficients corresponding to $V_{j-1}$ and $W_{j-1}$. The equation

$$\begin{aligned} c_{j,k} &= \langle v_j, \phi_{j,k} \rangle \\ &= \sum_l c_{j-1,l}\langle \phi_{j-1,l}, \phi_{j,k} \rangle + \sum_l d_{j-1,l}\langle \psi_{j-1,l}, \phi_{j,k} \rangle \\ &= \sum_l c_{j-1,l}\bar{h}_{k-2l} + \sum_l d_{j-1,l}\bar{g}_{k-2l} \end{aligned}$$

represents a single step in the inverse wavelet transform algorithm for reconstruction. Finally, corresponding detail complex wavelet coefficients that we will mainly handle in the later sections are obtained as

$$d_{j,k} = Re(d_{j,k}) + i \cdot Im(d_{j,k})$$

with magnitude

$$|d_{j,k}| = \sqrt{Re(d_{j,k})^2 + Im(d_{j,k})^2}$$

and phase

$$\angle d_{j,k} = \arctan\left(\frac{Im(d_{j,k})}{Re(d_{j,k})}\right).$$

## *1.6 Quaternion Wavelet Transform*

In recent decades, the traditional real-valued discrete wavelet transform (DWT) has been utilized as powerful mathematical tool in signal and image processing, in tasks of denoising, segmentation, compression, classification, and so on (Rajini, 2016). The traditional real-valued orthogonal discrete wavelet transforms (DWT) feature elegant, parsimonious, and informative representations, but have two shortcomings. The first is that DWT is not shift-invariant. Even a small shift of a signal results in complete change of wavelet coefficients, which causes problems in efficient computation and feature extraction in real-time. The second is that no phase information is encoded, unlike the Fourier representations (Chan et al, 2008). To accommodate the phase information, the complex wavelet transform was proposed in Lina (1997). We denote it as $\mathrm{WT_C}$ where c refers to complex instead of CWT that usually stands for continuous wavelet transform.

The $\mathrm{WT_C}$ is orthogonal, symmetric, and have decomposing atoms of compact size (Lina, 1997; Gao and Yan, 2011). Most notable characteristic is the phase information that $\mathrm{WT_C}$ additionally provides compared to real-valued wavelet decompositions. This phase information enables the $\mathrm{WT_C}$ to pack more information about the signal or image that it represents. Because of these merits, $\mathrm{WT_C}$ has been exploited in various wavelet-based tasks (Lina and MacGibbon, 1997; Remenyi et al, 2014; Jeon et al, 2014).

As an extension of the $\mathrm{WT_C}$, the quaternion wavelet transform (QWT) provides a richer scale-space analysis by taking into account the axioms of the quaternion algebra (Billow, 1999; Gai and Luo, 2015). This transform leads to quaternion-valued wavelet coefficients in the form of a vector of one modulus and three phases that possess symmetry properties and near shift-invariance, according to Billow's results (Billow and Sommer, 1997). The modulus reflects the outline of signal or image while the three phases encode local image shifts and represent subtle information

such as cusps, boundaries, and texture structure. Preserving the benefits of $\mathrm{WT}_{\mathbf{C}}$, the QWT can provide a more extensive redundancy with its three phases. Based on these merits, the QWT has been utilized in image denoising (Gai and Luo, 2015), texture classification (Soulard and Carre, 2011), image segmentation (Subakan and Vemuri, 2011), face recognition (Jones and Abbott, 2006), image fusion (Xue-Ni et al, 2016), etc. Note that these tasks usually have been performed with constructing quaternion wavelets utilizing four real-valued DWT: the first corresponding to the real part of the quaternion and the other three linked with the first via Hilbert transform correspond to the three imaginary parts of the quaternion wavelets. This transform possesses approximate shift invariance, abundant phase information, and limited redundancy while retaining the traditional wavelet time-frequency localization ability (Rajini, 2016). However, this original so-called QWTs were really DWT or $\mathrm{WT}_{\mathbf{C}}$ in disguise (Fletcher and Sangwine, 2017). In fact, their filter coefficients are real-valued, which means that it was technically wrong to name them QWT. In recent years, several studies have been conducted to construct a bonafide QWT that is not a conglomerate of DWTs and $\mathrm{WT}_{\mathbf{C}}$ (Carre and Denis, 2006; Hogan and Morris, 2012; Ginzberg and Walden, 2013). Ginzberg and Walden (2013) suggested true quaternion matrix-valued wavelets with quaternion-valued filter coefficients. Of the provided filters, for the analysis in later chapters, we selected non-trivial quaternion scaling and wavelet filters of length $L = 10$ and with five vanishing moments ($A = 5$) as a compromise between locality and smoothness. The selected filters correspond to non-trivial symmetric quaternion wavelet functions with compact support via a matrix-based implementation. Selected quaternion basis can address some critical problems from which other established but not fully quaternion wavelet designs had suffered. In addition, full quaternion approach leads to meaningful uniqueness and selective existence for filters of only certain lengths and numbers of vanishing moments. More details can be found in Ginzberg and Walden (2013).

### 1.6.1 Quaternion Algebra

Sir William Hamilton in 1843 first developed a quaternion algebra; the notation $\mathbb{H}$ for the field of all quaternions, is proposed after him (Gurlebeck and Sprossig, 1998). In a four-dimensional (4-D) algebra, the elements of $\mathbb{H}$ are given as linear combinations of a real scalar and three orthogonal imaginary units $i, j$, and $k$ with real coefficients as

$$\mathbb{H} = \{q = q_0 + q_1 i + q_2 j + q_3 k \mid q_0, q_1, q_2, q_3 \in \mathbb{R}\}, \tag{20}$$

where the three imaginary units $(i, j, k)$ satisfy the following non-commutative Hamilton's multiplication rules as

$$ij = -ji = k,\ jk = -kj = i,\ ki = -ik = j,\ i^2 = j^2 = k^2 = ijk = -1.$$

The conjugate of a quaternion $q$ can be written as

$$\overline{q} = q_0 - q_1 i - q_2 j - q_3 k, \tag{21}$$

and some useful properties for the conjugate as follows:

$$\overline{\overline{q}} = q,\ \overline{q + p} = \overline{q} + \overline{p},\ \overline{qp} = \overline{q}\ \overline{p}, \quad \forall\ q, p \in \mathbb{H}.$$

Since the product of a quaternion $q$ and its conjugate $\overline{q}$ in Equation (21) is

$$q\overline{q} = q_0^2 + q_1^2 + q_2^2 + q_3^2,$$

the modulus $|q|$ of a quaternion $q$ is correspondingly defined as

$$|q| = \sqrt{q\overline{q}} = \sqrt{q_0^2 + q_1^2 + q_2^2 + q_3^2}.$$

In a manner similar to complex numbers, the expression of quaternion $q$ in Equation (20) can have an alternative representation in polar form as

$$q = |q| e^{i\phi} e^{j\theta} e^{k\psi}, \tag{22}$$

where $(\phi, \theta, \psi) \in [-\pi, \pi] \times [-\frac{\pi}{2}, \frac{\pi}{2}] \times [-\frac{\pi}{4}, \frac{\pi}{4}]$. For a unit quaternion, $|q| = q\overline{q} = 1$, their corresponding three phases can be evaluated as follows:

*1.* First, compute $\psi$ as

$$\psi = -\frac{1}{2}\arcsin\Big(2(q_1q_2 - q_0q_3)\Big).$$

*2.* If $\psi \in (-\frac{\pi}{4}, \frac{\pi}{4})$, then

$$\begin{cases} \phi = \frac{1}{2}\arctan 2\big(2(q_0q_1 + q_2q_3), q_0^2 - q_1^2 + q_2^2 - q_3^2\big), \\ \theta = \frac{1}{2}\arctan 2\big(2(q_0q_2 + q_1q_3), q_0^2 + q_1^2 - q_2^2 - q_3^2\big). \end{cases}$$

*3.* If $\psi = \pm\frac{\pi}{4}$, then select either

$$\begin{cases} \phi = \frac{1}{2}\arctan 2\big(2(q_0q_1 - q_2q_3), q_0^2 - q_1^2 - q_2^2 + q_3^2\big), \\ \theta = 0. \end{cases}$$

or

$$\begin{cases} \phi = 0, \\ \theta = \frac{1}{2}\arctan 2\big(2(q_0q_2 - q_1q_3), q_0^2 - q_1^2 - q_2^2 + q_3^2\big)\big). \end{cases}$$

*4.* If $e^{i\phi}e^{j\theta}e^{k\psi} = -q$ and $\phi \geq 0$, then $\phi \to \phi - \pi$.

*5.* If $e^{i\phi}e^{j\theta}e^{k\psi} = -q$ and $\phi < 0$, then $\phi \to \phi + \pi$.

### 1.6.2 Matrix-based Non-trivial Quaternion Wavelet Basis

As we mentioned before, Ginzberg and Walden (2013) suggested a true quaternion matrix-valued wavelets with quaternion-valued filter coefficients. From the provided filters, for applications in this thesis, we selected the non-trivial quaternion scaling and wavelet filters of length $L = 10$ and with five vanishing moments ($A = 5$).

Specifically, for $L = 10$ and $A = 5$, the set of design equations is given as

$$\begin{aligned} \sum_{k=0}^{9} \mathbf{H}_k &= \sqrt{2}\mathbf{I}_4, \\ \sum_{k=0}^{9} (-1)^k k^d \mathbf{H}_k &= \mathbf{0}_4 \quad \text{for} \quad d = 0, 1, 2, 3, 4, \\ \sum_{k=0}^{9-2m} \mathbf{H}_k \mathbf{H}_{k+2m}^T &= \delta_{m,0}\mathbf{I}_4 \quad \text{for} \quad m = 1, 2, 3, 4, \end{aligned}$$

where the $\mathbf{H}_k$ each denote $4 \times 4$ matrix representations of quaternion. By solving these equations, one obtains the wavelet filters as

$$\begin{aligned} h_0 &= h_9 = C_2 i, \\ h_1 &= h_8 = -5C_1 + C_2 k, \\ h_2 &= h_7 = -7C_1 - 7C_2 i + 3C_2 k, \\ h_3 &= h_6 = 35C_1 - 5C_2 i + C_2 k, \\ h_4 &= h_5 = 105C_1 + 11C_2 i - 5C_2 k, \end{aligned}$$

where $C_1 = \frac{\sqrt{2}}{256}$ and $C_2 = \frac{\sqrt{35}}{256}$. The corresponding antisymmetric scaling filters are also described as

$$\begin{aligned} g_0 &= -g_9 = C_3(89\sqrt{35}i + 35\sqrt{2}j - 35\sqrt{35}k), \\ g_1 &= -g_8 = C_3(-480\sqrt{2} + 35\sqrt{35}i - 175\sqrt{2}j + 79\sqrt{35}k), \\ g_2 &= -g_7 = C_4(84\sqrt{2} - 91\sqrt{35}i + 35\sqrt{2}j + \sqrt{35}k), \\ g_3 &= -g_6 = C_5(35\sqrt{2} + 5\sqrt{35}i - \sqrt{35}k), \\ g_4 &= -g_5 = C_6(-5040\sqrt{2} + 577\sqrt{35}i - 245\sqrt{2}j + 5\sqrt{35}k), \end{aligned}$$

where $C_3 = \frac{1}{24576}, C_4 = \frac{1}{3072}, C_5 = \frac{1}{256}$ and $C_6 = \frac{1}{12288}$. Note that we renamed Ginzberg's $\mathbf{G}$ for $\mathbf{H}$ and $g$ for $h$ in accordance with conventional notations of $\mathbf{H}$ for low pass filter and $\mathbf{G}$ for high pass filter. The scaling and wavelet functions for all real and imaginary parts are presented in Figure 8.

## *1.7 Non-decimated Wavelet Transform*

In this section, we overview the non-decimated wavelet transform (NDWT) that is another popular version of wavelet transforms. As we look at its alternative names such as "stationary wavelet transform," "time-invariant wavelet transform," "*á trous* transform," or "maximal overlap wavelet transform", the NDWT can resolve a problem of sensitivity to translations in time and shift detected in traditional orthogonal

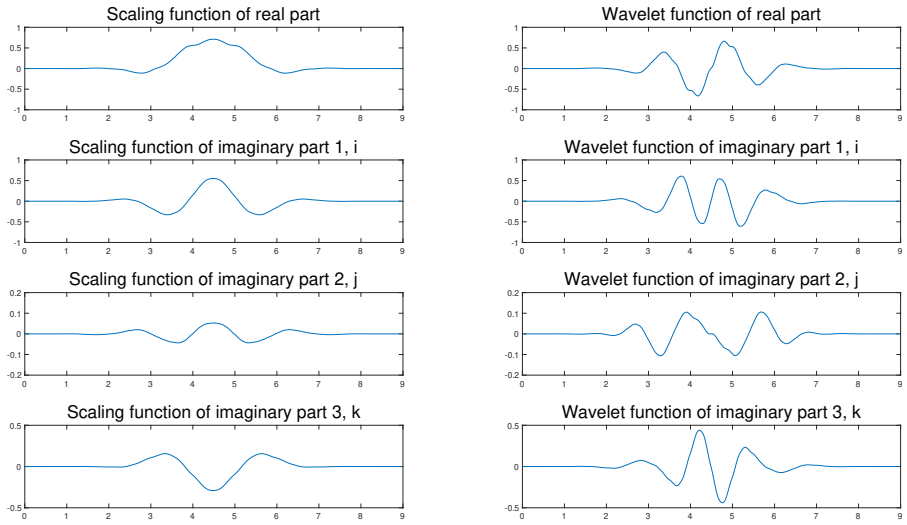



**Figure 8:** Quaternion scaling and wavelet functions for $L = 10$ and $A = 5$.

wavelet transform. Even if some researches on the names use somewhat different terms, they are inherently same with the NDWT. All of them show characteristics such as translation/shift invariance and denser approximation of continuous wavelet transform, which resulted from a redundancy property of them. This property is caused since they are performed by repeated filtering with a minimal shift (or with a maximal sampling rate) at all dyadic scales. Consequently, there are the same numbers of wavelet transformed coefficients at each multiresolution level as original data and we call this phenomenon a non-decimated transform. Even though the total size of decomposition obtained by NDWT is larger than that of the orthogonal transform, this redundancy is often preferred by practitioners in many fields. More details can be found in excellent monographs Percival and Walden (2006) and Mallat (2009).

Compared to the traditional wavelet transform, the main difference is a sampling rate. We assume that a multiresolution framework is specified and denote scaling and wavelet functions as $\phi$ and $\psi$, respectively. Then, we can represent a data vector $\boldsymbol{y} = (y_0, y_1, \ldots, y_{m-1})$ of size $m$ as a function $f$ in terms of shifts of the scaling function

at some multiresolution level $J$ as follows:

$$f(x) = \sum_{k=0}^{m-1} y_k \phi_{J,k}(x)$$

where $J - 1 < \log_2 m \leq J$ and $\phi_{J,k}(x) = 2^{\frac{J}{2}} \phi(2^J(x-k))$. Next, we also can express the data interpolating function $f$ as

$$f(x) = \sum_{k=0}^{m-1} c_{J_0,k} \phi_{J_0,k}(x) + \sum_{j=J_0}^{J-1} \sum_{k=0}^{m-1} d_{j,k} \psi_{j,k}(x) \tag{23}$$

where

$$\phi_{J_0,k}(x) = 2^{\frac{J_0}{2}} \phi\left(2^{J_0}(x-k)\right),$$
$$\psi_{j,k}(x) = 2^{\frac{j}{2}} \psi\left(2^{j}(x-k)\right),$$

and $J_0$ is the coarsest decomposition level. Then, the non-decimated wavelet transforms of vector $\boldsymbol{y}$ consist of wavelet coefficients, $\boldsymbol{c}^{(J_0)} = (c_{J_0,0}, c_{J_0,1}, \ldots, c_{J_0,m-1})$ and $\boldsymbol{d}^{(j)} = (d_{j,0}, d_{j,1}, \ldots, d_{j,m-1})$ where $j = J_0, \ldots, J-1$ for a decomposition depth $p = J - J_0$, which are redundant representations of the input data $\boldsymbol{y}$ as seen in the Equation (23). To be specific, $\boldsymbol{c}^{(J_0)}$ serves as the coarsest approximation of the input data and called coarse coefficients, while $\boldsymbol{d}^{(j)}$ contains detail information within an input and called detail coefficients. Note that we cannot take a $p$ that is larger than $\log_2 m$ due to the non-decimated property.

The principal difference between the standard- and non-decimated wavelet transform is that the shift indicator $k$ of the latter is constant at all level unlike the former whose shifts are level dependent as $2^{-j}k$. We can know that this non-decimated transform is time invariant because of the constancy of the shifts across the levels. As a result, the total number of wavelet coefficients is $(p+1) \times m$ with $p$ detail levels. Notice that the non-decimated wavelet coefficients at each decomposition level are related based on Mallat's pyramid algorithm (Mallat, 1989b,a). The low-pass wavelet filter $h$ and the high-pass wavelet filter $g$ are quadrature mirror filters and

their convolutions are performed in a cascade. Given a fully and uniquely specified $h$ through the choice of wavelet basis, the $i$th component of $g$ is $g_i = (-1)^{l-i} h_{m-s-i}$ for any fixed integer $l$ and $s$. We will further discuss the filter operators in terms of the NDWT later in this section.

### 1.7.1 Discrete Non-decimated Wavelet Transform

The functional series including wavelet and scaling functions as decomposing components have been established as a mathematical framework to describe the NDWT. However, for discrete inputs, an alternative description of the NDWT utilizing convolution operators is preferred (Nason and Silverman, 1995; Strang and Nguyen, 1996; Vidakovic, 1999).

We denote $[\uparrow 2]$ as the upsampling of a given sequence that inserts a zero between every two interfaced components of a sequence. The dilations of wavelet filters $h$ and $g$ are defined as follows:

$$\begin{aligned} h^{[0]} &= h, & g^{[0]} &= g \\ h^{[r]} &= [\uparrow 2] h^{[r-1]}, & g^{[r]} &= [\uparrow 2] g^{[r-1]}. \end{aligned} \tag{24}$$

This method is sometimes referred as Algorithm á Trous since holes (*trous*, in French) are produced by inserting zeros between each component of filter $h^{[r-1]}$ and $g^{[r-1]}$ (Shensa, 1992).

Applying convolution operators, $\mathbf{H}^{[j]}$ and $\mathbf{G}^{[j]}$, from the dilated filters $h^{[r]}$ and $g^{[r]}$ in sequence, we can complete the non-decimated wavelet transform. Then, detail and coarse coefficient vectors for each level have a same size, $m$, that is a size of the original data. Employing repeatedly the convolution operators to a coarse coefficient vector of the previous decomposition level, $J - j + 1$, we can calculate coefficient

vector of the level $J-j$ where $j \in \{1, 2, \ldots, p\}$ as follows:

$$\boldsymbol{c}^{(J-j)} = \mathbf{H}^{[j-1]}\boldsymbol{c}^{(J-j+1)}$$
$$\boldsymbol{d}^{(J-j)} = \mathbf{G}^{[j-1]}\boldsymbol{c}^{(J-j+1)},$$

where $\mathbf{H}^{[j-1]}$ and $\mathbf{G}^{[j-1]}$ are low- and high-pass filter operators based on quadrature mirror filters $h^{[j-1]}$ and $g^{[j-1]}$, respectively. Then, we can perform the NDWT by repeatedly applying the two convolution operators, $\mathbf{H}^{[j]}$ and $\mathbf{G}^{[j]}$. However, the operators have a problem of no orthogonality, so to resolve this problem, we apply two additional operators $\mathcal{D}_0$ and $\mathcal{D}_1$ to implement decimation by choosing all even and odd elements of input data. We can use these decimation operators, $\mathcal{D}_0$ and $\mathcal{D}_1$, with a signal $x$ as

$$(\mathcal{D}_0 x)_i = x_{2i},$$
$$(\mathcal{D}_1 x)_i = x_{2i+1},$$

where $i$ implies the location of an component in the signal $x$. Then, we can obtain the even and odd elements of the non-decimated wavelet coefficient vectors, $\boldsymbol{c}^{(J-j)}$ and $\boldsymbol{d}^{(J-j)}$, by employing $(\mathcal{D}_0\mathbf{H}^{[j-1]}, \mathcal{D}_0\mathbf{G}^{[j-1]})$ and $(\mathcal{D}_1\mathbf{H}^{[j-1]}, \mathcal{D}_1\mathbf{G}^{[j-1]})$ to the signal $x$. Thus, we can represent the implementation of the Equation (24) as follows:

$$(\boldsymbol{c}^{(J-j)})_{2i} = \mathcal{D}_0\mathbf{H}^{[j-1]}\boldsymbol{c}^{(J-j+1)}$$
$$(\boldsymbol{c}^{(J-j)})_{2i+1} = \mathcal{D}_1\mathbf{H}^{[j-1]}\boldsymbol{c}^{(J-j+1)}$$
$$(\boldsymbol{d}^{(J-j)})_{2i} = \mathcal{D}_0\mathbf{G}^{[j-1]}\boldsymbol{c}^{(J-j+1)}$$
$$(\boldsymbol{d}^{(J-j)})_{2i+1} = \mathcal{D}_1\mathbf{G}^{[j-1]}\boldsymbol{c}^{(J-j+1)}.$$

This NDWT process is graphically illustrated in Figure 9. Note that a shift does not make an effect on transformation since the filtering is applied twice at the even and odd locations for each decomposition level, which makes the NDWT time-invariant. This time-invariance characteristic makes it possible for the NDWT to have a smaller

mean squared error and decrease the Gibbs phenomenon in denoising (Coifman and Donoho, 1995). While such a property has the advantage, it complicates the signal reconstruction due to the non-compliance of variance preservation. In the next section, we will consider a way of lossless reconstruction using a matrix-based NDWT.

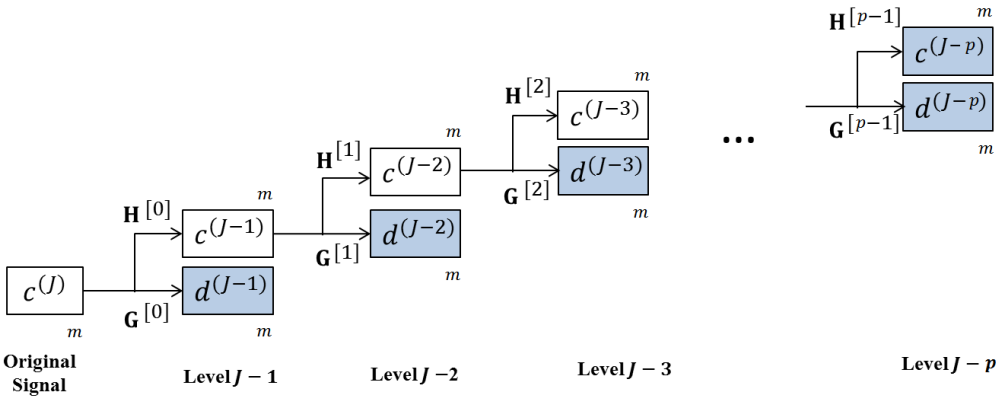



**Figure 9:** Graphical illustration of the NDWT process. The NDWT decomposes the original signal of size $m$ to $p+1$ multiresolution subspaces including $p$ levels of detail coefficients and one level of coarsest coefficients. Then we can obtain a set of coefficient vectors, $\boldsymbol{d}^{(J-1)}, \boldsymbol{d}^{(J-2)}, \ldots, \boldsymbol{d}^{(J-p)}$, and $\boldsymbol{c}^{(J-p)}$ in the shaded blocks.

### 1.7.2 Standard 2-D Non-decimated Wavelet Transform

Next, we overview the standard versions of the 2-D NDWT of $f(x, y)$ extended from the 1-D definitions where $(x, y) \in \mathbb{R}^2$. The representation of non-decimated wavelets in 2-D can be performed equivalently as Relations in (1.4.1) although $\phi(x)$ and $\psi(x)$ are different.

For the standard 2-D NDWT, we define the wavelet atoms as follows:

$$\begin{aligned}
\phi_{J_0,k_1,k_2}(x,y) &= 2^{J_0}\phi(2^{J_0}(x-k_1))\phi(2^{J_0}(y-k_2)),\\
\psi^{(h)}_{j,k_1,k_2}(x,y) &= 2^{j}\phi(2^{j}(x-k_1))\psi(2^{j}(y-k_2)),\\
\psi^{(v)}_{j,k_1,k_2}(x,y) &= 2^{j}\psi(2^{j}(x-k_1))\phi(2^{j}(y-k_2)),\\
\psi^{(d)}_{j,k_1,k_2}(x,y) &= 2^{j}\psi(2^{j}(x-k_1))\psi(2^{j}(y-k_2)),
\end{aligned}$$

where $(k_1, k_2) \in \mathbb{Z}^2$ is the location pair and $j = J_0, \ldots, J-1$ with depth $p = J-1-J_0$.

Then, the wavelet coefficients of $f(x, y)$ are obtained as

$$
\begin{aligned}
c_{J_0,k_1,k_2} &= \iint f(x,y)\phi_{J_0,k_1,k_2}(x,y)\; dxdy, \\
d^{(h)}_{j,k_1,k_2} &= \iint f(x,y)\psi^{(i)}_{j,k_1,k_2}(x,y)\; dxdy, \\
d^{(v)}_{j,k_1,k_2} &= \iint f(x,y)\psi^{(i)}_{j,k_1,k_2}(x,y)\; dxdy, \\
d^{(d)}_{j,k_1,k_2} &= \iint f(x,y)\psi^{(i)}_{j,k_1,k_2}(x,y)\; dxdy.
\end{aligned}
$$

The tessellation of a standard 2-D NDWT is presented in Figure 10(a).

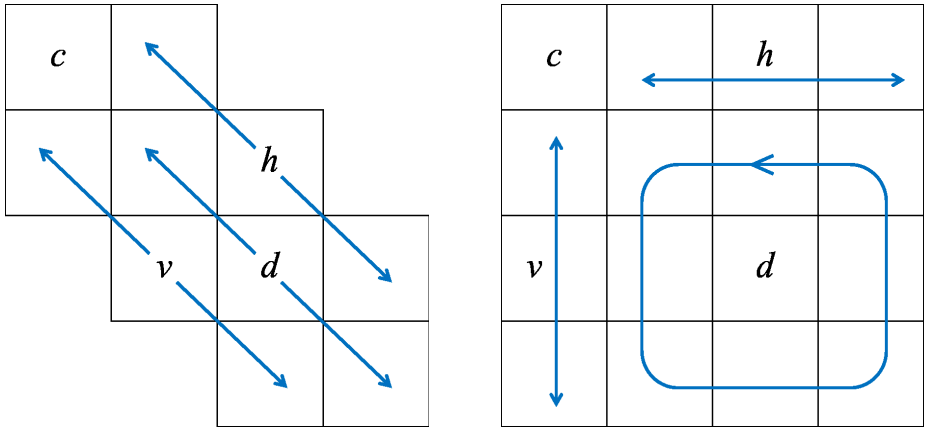



**Figure 10:** (a) Tessellations for traditional 2-D NDWT of depth 3 (b) Tessellations for scale-mixing 2-D NDWT of depth 3.

### 1.7.3 Scale-mixing 2-D Non-decimated Wavelet Transform

For the scale-mixing 2-D NDWT, we define the wavelet atoms as follows:

$$
\begin{aligned}
\phi_{J_{01},J_{02},k_1,k_2}(x,y) &= 2^{\frac{J_{01}+J_{02}}{2}}\phi(2^{J_{01}}(x-k_1))\phi(2^{J_{02}}(y-k_2)), \\
\psi^{(h)}_{J_{01},j_2,k_1,k_2}(x,y) &= 2^{\frac{J_{01}+j_2}{2}}\phi(2^{J_{01}}(x-k_1))\psi(2^{j_2}(y-k_2)), \\
\psi^{(v)}_{j_1,J_{02},k_1,k_2}(x,y) &= 2^{\frac{j_1+J_{02}}{2}}\psi(2^{j_1}(x-k_1))\phi(2^{J_{02}}(y-k_2)), \\
\psi^{(d)}_{j_1,j_2,k_1,k_2}(x,y) &= 2^{\frac{j_1+j_2}{2}}\psi(2^{j_1}(x-k_1))\psi(2^{j_2}(y-k_2)),
\end{aligned}
$$

where $(k_1, k_2) \in \mathbb{Z}^2$, $j_1 \geq J_{01}; j_2 \geq J_{02}$ and $J_{01}$ and $J_{02}$ are the coarsest decomposition levels of rows and columns. Then, any function $f \in L_2(\mathbb{R}^2)$ can be expressed as

$$\begin{aligned} f(x,y) &= \sum_{k_1}\sum_{k_2} c_{J_{01},J_{02},k_1,k_2}\phi_{J_{01},J_{02},k_1,k_2}(x,y) \\ &+ \sum_{j_2>J_{02}}\sum_{k_1}\sum_{k_2} d_{J_{01},j_2,k_1,k_2}\psi^{(h)}_{J_{01},j_2,k_1,k_2}(x,y) \\ &+ \sum_{j_1>J_{01}}\sum_{k_1}\sum_{k_2} d_{j_1,J_{02},k_1,k_2}\psi^{(v)}_{j_1,J_{02},k_1,k_2}(x,y) \\ &+ \sum_{j_1>J_{02}}\sum_{j_2>J_{01}}\sum_{k_1}\sum_{k_2} d_{j_1,j_2,k_1,k_2}\psi^{(d)}_{j_1,j_2,k_1,k_2}(x,y), \end{aligned} \tag{25}$$

which defines a scale-mixing NDWT. Unlike the standard 2-D NDWT denoting a scale as $j$, we denote mixed two scales as a pair $(j_1, j_2)$. Finally, the resulting scale-mixing wavelet coefficients are

$$\begin{aligned} c_{J_{01},J_{02},k_1,k_2} &= \iint f(x,y)\phi_{J_{01},J_{02},k_1,k_2}(x,y)\,dxdy, \\ d^{(h)}_{J_{01},j_2,k_1,k_2} &= \iint f(x,y)\psi^{(h)}_{J_{01},j_2,k_1,k_2}(x,y)\,dxdy, \\ d^{(v)}_{j_1,J_{02},k_1,k_2} &= \iint f(x,y)\psi^{(v)}_{j_1,J_{02},k_1,k_2}(x,y)\,dxdy, \\ d^{(d)}_{j_1,j_2,k_1,k_2} &= \iint f(x,y)\psi^{(d)}_{j_1,j_2,k_1,k_2}(x,y)\,dxdy. \end{aligned} \tag{26}$$

The tessellation of a scale-mixing 2-D NDWT is presented in Figure 10(b).

### 1.7.4 Matrix-based Non-decimated Wavelet Transform

#### *1.7.4.1 1-D case*

Research in Kang and Vidakovic (2016) provided how to perform the NDWT through a simple matrix multiplication instead of the multiple convolutions shown in the section 1.7.1. To be specific, we do not need to formulate the NDWT anymore by implementation of convolutions in a cascade with wavelet and scaling functions in Mallat's algorithm. Alternatively, we simplify perform the NDWT by matrix multiplication and the cascade algorithm is as follows. This matrix formulation is

beneficial for several reasons, including versatility, adaptivity, and generalizability, especially in the multidimensional formulations.

First, we repeatedly perform matrix multiplications of low- and high-pass filter matrices, $(H)$ and $(G)$, which implies the Mallat's cascade algorithm. The filter matrices $\begin{bmatrix} H_p & G_p \end{bmatrix}^T$ are constructed by the following rule.

*Rule 1.* The sizes of $H_p$ and $G_p$ for depth $p$ are the same as $m \times m$.

*Rule 2.* $h_s^{[p-1]}$ defines the $s^{th}$ component of a dilated wavelet filter $h^{[p-1]}$ that has $p-1$ zeros in between the original filter elements, $(h_1, h_2, \ldots, h_u)$, as

$$h^{[p-1]} = (h_1, \overbrace{0, \ldots, 0}^{p-1}, h_2, \overbrace{0, \ldots, 0}^{p-1}, h_3, \ldots, \overbrace{0, \ldots, 0}^{p-1}, h_u).$$

For instance, $h_1^{[p-1]} = h_1,\ h_{p+1}^{[p-1]} = h_2, \ldots,$ and, $h_{p(u-1)+1}^{[p-1]} = h_u$.

*Rule 3.* The components of $H_p$ and $G_p$ at the location $(i, j)$ are

$$h_{ij} = \frac{1}{\sqrt{2}} h_s^{[p-1]}, \quad s = N + i - j \text{ modulo } m$$
$$g_{ij} = \frac{1}{\sqrt{2}} (-1)^s h_{N+1-s}^{[p-1]}, \quad s = N + i - j \text{ modulo } m,$$

where $N$ is a shift parameter.

Then, we can obtain the NDWT matrix, $W_m^{(p)}$, that is a combination of such filter matrices, $H_p$ and $G_p$, to implement a transform of depth $p$. The $W_m^{(p)}$ is a matrix of size $m(p+1) \times m$ composed of $p+1$ stacked submatrices, $H_p$ and $G_p$, of size $m \times m$. Examples for depth 1,2, and 3 are provided as follows:

$$W_m^{(1)} = \begin{bmatrix} H_1 \\ G_1 \end{bmatrix}_{[2m \times m]} \qquad W_m^{(2)} = \begin{bmatrix} \begin{bmatrix} H_2 \\ G_2 \end{bmatrix} \cdot H_1 \\ G_1 \end{bmatrix}_{[3m \times m]} \qquad W_m^{(3)} = \begin{bmatrix} \begin{bmatrix} \begin{bmatrix} H_3 \\ G_3 \end{bmatrix} \cdot H_2 \\ G_2 \end{bmatrix} \cdot H_1 \\ G_1 \end{bmatrix}_{[4m \times m]}$$

We can perform the NDWT through the matrix formulation $W_m^{(p)}$, however, a reconstruction cannot be simply implemented by its transpose only since the provided NDWT matrix is not normalized. In fact, we need the multiplication by $\frac{1}{2}$ at each inverse transform step for a perfect reconstruction in context of Mallat's algorithm (Mallat, 2009). Thus, we provide a way of correct inverse transform using a diagonal weight matrix that re-scales the square submatrices composing the NDWT matrix. The weight matrix for $W_m^{(p)}$ is of size $m(p+1) \times m(p+1)$ and represented as

$$T_m^{(p)} = \text{diag}(\overbrace{1/2^p, \ldots, 1/2^p}^{2m}, \overbrace{1/2^{p-1}, \ldots, 1/2^{p-1}}^{m}, \overbrace{1/2^{p-2}, \ldots, 1/2^{p-2}}^{m}, \ldots, \overbrace{1/2, \ldots, 1/2}^{m}).$$

For 1-D signal $\boldsymbol{y}$ of size $m \times 1$, we transform it to a non-decimated wavelet transformed vector $\boldsymbol{d}$ with depth $p$ by multiplying $\boldsymbol{y}$ by $W_m^{(p)}$. Following the transformation, we can reconstruct $\boldsymbol{y}$ from $\boldsymbol{d}$ by multiplying $\boldsymbol{d}$ by both $(W_m^{(p)})^T$ and $T_m^{(p)}$ as follows:

$$\begin{aligned} \boldsymbol{d} &= W_m^{(p)} \times \boldsymbol{y}_{[m \times 1]} \\ \boldsymbol{y} &= (W_m^{(p)})' \times T_m^{(p)} \times \boldsymbol{d}, \end{aligned} \tag{27}$$

where $p$ and $m$ are arbitrary.

Notice that $(W_m^{(p)})' \times W_m^{(p)} \neq I_m$, while column vectors of matrix as

$$V_m^{(p)} = \left(T_m^{(p)}\right)^{1/2} W_m^{(p)}$$

build an orthonormal set, thus,

$$(V_m^{(p)})' \times V_m^{(p)} = I_m. \tag{28}$$

The reverse product, $V_m^{(p)} \times (V_m^{(p)})'$, is not an identity matrix, however,

$$\sum_i \left(V_m^{(p)} \times (V_m^{(p)})'\right)_{ij} = \left(\sum_j \left(V_m^{(p)} \times (V_m^{(p)})'\right)_{ij}\right)' = [\mathbf{1}_m, \mathbf{0}_{pm}],$$

where $[\mathbf{1}_m, \mathbf{0}_{pm}]$ implies a row vector composed of $m$ ones followed by the $pm$ zeros.

Finally, the perfect reconstruction can be performed by multiplying $\boldsymbol{d}$ by $(W_m^{(p)})' \times T_m^{(p)}$ because $I_m = (V_m^{(p)})' \times V_m^{(p)} = (W_m^{(p)})' \times T_m^{(p)} \times W_m^{(p)}$ as in Equation (27). Note

that a transform by $V_m^{(p)}$ can be considered more natural due to the relation in (28). However, the scaling of wavelet coefficients transformed by $V_m^{(p)}$ does not match the correct scaling wavelet coefficients achieved by Mallat's algorithm. Thus, we should use not $V_m^{(p)}$ but $W_m^{(p)}$ since the exact scaling of wavelet coefficients is essential in applications such as regularity assessment of signals and images.

#### *1.7.4.2 2-D case*

To obtain a 2-D wavelet transformed image $\boldsymbol{B}$ from $\boldsymbol{A}$ of size $[m \times n]$ for $p_1$- and $p_2$-depth along rows and columns through the scale-mixing NDWT, we apply the NDWT matrix multiplication to $\boldsymbol{A}$ the from the left and its transpose from the right. Then, the size of $\boldsymbol{B}$ becomes $(p_1+1)m \times (p_2+1)n$. Note that the weight matrix with $p_1$- and $p_2$-level decompositions, $T_m^{(p_1)}$ and $T_n^{(p_2)}$, are applied for a correct inverse transform. These processes can be summarized as follows:

$$\boldsymbol{B} = W_m^{(p_1)} \times \boldsymbol{A}_{[m \times n]} \times (W_n^{(p_2)})'$$
$$\boldsymbol{A} = (W_m^{(p_1)})' \times T_m^{(p_1)} \times \boldsymbol{B} \times T_n^{(p_2)} \times W_n^{(p_2)},$$

An example of a graphical illustration of a 2-D scale-mixing NDWT is provided in Figure 11. Since $W_m^{(p_1)}$ and $W_m^{(p_2)}$ are distinguished, they can be constructed by different wavelet filters, respectively.

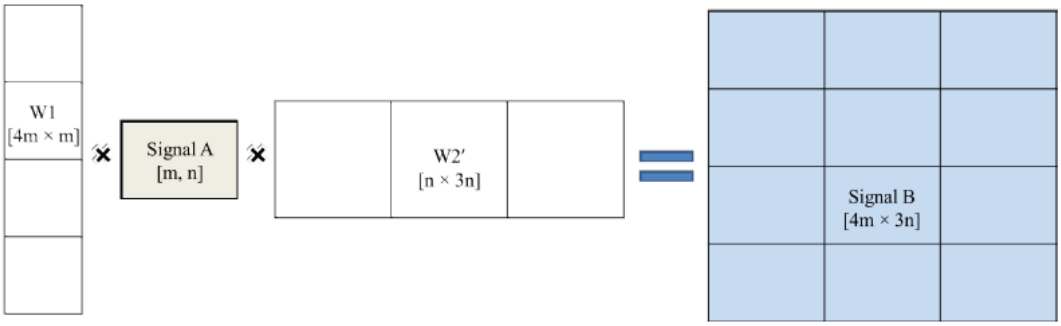
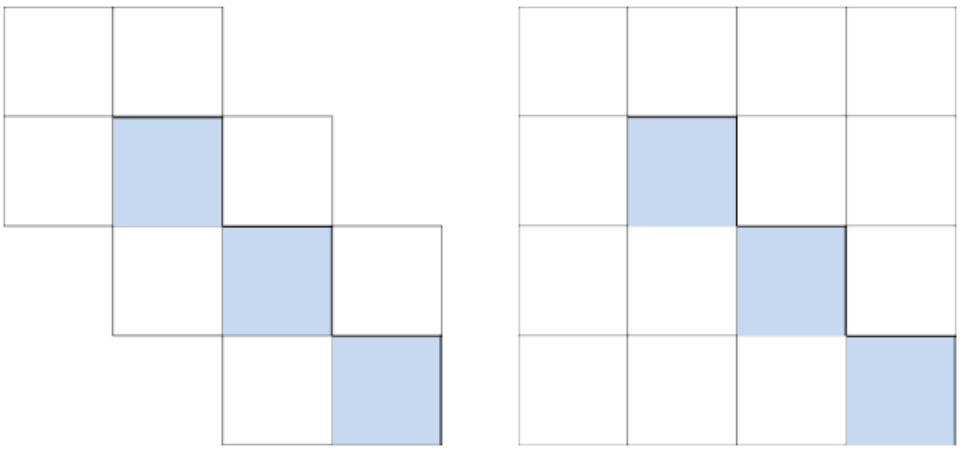



**Figure 11:** A 2-D scale-mixing NDWT process is graphically presented with 3-levels along the rows and 2-levels along the columns.

The reason why we use the 2-D scale-mixing NDWT is because it shows a remarkable compressibility. In particular, the superb performance on compression is exceptionally advantageous for wavelet-based data compression or denoising. When an image shows a some extent of smoothness, the diagonal decomposition coefficients ($d^{(d)}$ in Equation (27)) have smaller magnitudes comparing to the other three types of coefficients ($c$, $d^{(v)}$, and $d^{(h)}$ in Equation (27)). Let us consider a $p$-level decomposition example of 2-D image of size $m \times n$ by both standard NDWT and scale-mixing NDWT from matrix formulation. We can define a compressibility of wavelet transform as the proportion of $d^{(d)}$ compared with the total number of wavelet coefficients. Recall that $d^{(d)}$ indicates the decomposing atoms containing two wavelet functions, while $c$, $d^{(v)}$, and $d^{(h)}$ represent the decomposing atoms consisting of at least one scaling function. Thus, for a depth $p$, the compressibility of the standard NDWT is $pmn/((3p+1)mn)$ while that of the scale-mixing NDWT is $p^2mn/((p+1)^2mn)$, which is shown in Figure 12. Note that the latter is always larger than the former except for the case of $p = 1$ where the two values are same. Therefore, we can conclude that the scale-mixing NDWT is more compressive than the standard NDWT.



(a) (b)

**Figure 12:** (a) Tessellations for standard 2-D NDWT of depth 3 (b) Tessellations for scale-mixing 2-D NDWT of depth 3. The shaded regions indicate the diagonal wavelet coefficients, $d^{(d)}$.

More details on matrix formulation of NDWT can be found in Kang and Vidakovic (2016).

# CHAPTER II

# NON-DECIMATED COMPLEX WAVELET SPECTRAL TOOLS

Wavelets have become standard tools in signal and image processing. Of many versions of a wavelet transforms that are used in such applications, a popular version is a complex wavelet transform. We denote it as $\mathrm{WT_C}$ where c refers to complex instead of CWT that usually stands for continuous wavelet transform. In the past, the multiresolution analysis based on the complex-valued coefficients had not been widely utilized since the resulting redundant representations of real signals seemed to be uninformative (Lina, 1997). It is agreed among experts that desirable properties for basis functions in functional representation of signals and images should be orthogonality, symmetry, and compact support (Gao and Yan, 2011). Orthogonality is important because of representational parsimony (Mallat, 2009). In particular, the orthogonality is important for a coherent definition of power spectra because of energy preservation. The symmetry is especially desired when dealing with images (Antonini et al, 1992). In particular, the study in Simoncelli and Adelson (1996) showed that symmetric basis functions can prevent directional distortions via an orientation-free representation of features. Finally, functional representations should be computationally efficient and local which requires compact support for decomposing functions. These three desirable properties in the wavelet context are only available by the orthogonal complex wavelets with an odd number of vanishing moments. The Haar wavelet is an exception (Lawton, 1993). For Daubechies complex wavelets in Lina (1997), these characteristics result from the underlying differential operators defining

the complex-valued multiresolution. Even though the complex wavelets are orthogonal, the representations are redundant because of complex-valued coefficients. This provides for a potential benefit of phase information (Jeon et al, 2014). Because of this supplemental phase information the complex wavelets have been utilized in various fields including motion estimation (Magarey and Kingsbury, 1998), texture image modeling (Portilla and Simoncelli, 2000), signal denoising (Achim and Kuruoglu, 2005; Remenyi et al, 2014), NMR spectra classification (Kim et al, 2008), and mammogram images classification (Jeon et al, 2014).

Although orthogonal transforms are minimal, mathematically elegant, and easy to implement, they suffer from the Balian-Low obstacle concerning simultaneous locality in the time and scale domains. Redundant dictionaries can be constructed that preserve the ease of computation and do not suffer from the Balian-Low limitations by sacrificing the orthogonality property. As a compromise, the non-decimated wavelet transform (NDWT) is a superposition of many orthogonal transforms, and as such preserves the ease of computation but results in redundant representations. As we look at some alternative names of NDWT such as "stationary wavelet transform," "time-invariant wavelet transform," "*á trous* transform," or "maximal overlap wavelet transform", they all refer to its two properties: redundancy and translation invariance, both absent in traditional orthogonal discrete wavelet transforms (DWT). Non-decimated wavelet transform represents a dense discrete sample of coefficients from continuous wavelet transforms, which results in their structural redundancy. Operationally, non-decimated wavelet transform is performed by Mallat's algorithm without decimation: a repeated filtering with a minimal shift at all dyadic scales. Consequently, at each multiresolution level the number of wavelet coefficients is the same as the size of the original data. Although the non-decimated wavelet transform increases computational complexity, it has been widely used particularly because of the usefulness of redundancy and an easy way to adjust for the energy preservation.

More details on some additional benefits over the standard WT can be found in Kang and Vidakovic (2016).

In this chapter we propose a non-decimated complex wavelet transform ($\text{NDWT}_{\text{C}}$) that is a combination of the aforementioned two types of wavelet transform. The $\text{WT}_{\text{C}}$ produces complex-valued redundant type of wavelet coefficients and the non-decimated wavelets have a redundant structure of wavelet coefficients. We call the former componential redundancy and the latter structural redundancy. Since they represent different types of redundancy, we suggested that their combination can be beneficial in feature extraction.

A study in Jeon et al (2014) suggested a classification procedure for mammogram images based on obtained spectral slope based on the modulus and average of phases at the finest level, constructed from coefficients in $\text{WT}_{\text{C}}$. The novelty of that approach was that it calculated a descriptor based on phases of complex wavelet coefficients and used it as an additional input in machine learning tasks. The authors in Jeon et al (2014) showed that use of phase increased the precision of the classification.

We suggest that this performance can be improved by incorporating the phase information from all detail levels in the multiresolution analysis. Different levels of detail in the multiresolution hierarchy carry almost independent information on the signal behavior on various scales. Experimental evidence showed that phase information from the coarser scales can serve as useful summaries in classification algorithms. Besides, the accuracy can be increased more when the $\text{WT}_{\text{C}}$ is used with the NDWT together because of the redundancy. This is because level based summaries are obtained from large number of coefficients. One criticism could be that the increased dependence of the coefficients within the level in non-decimated transforms can be detrimental to the summary statistics. It is true that this would be an impediment for the estimation inference, but not so for the classification because the possible bias in the summaries affects the coefficients from different classes in the

same way.

The one of disadvantages of standard wavelet transforms is that they are efficiently applied only to signals and square-sized images whose dimensions are dyadic, even for the complex wavelets and convolution-based non-decimated wavelets (Lina, 1999; Percival and Walden, 2000). In practice, this is a serious limitation and to overcome it one increases the computational complexity. We construct the matrix-based NDWT in Kang and Vidakovic (2016) with complex-valued filters in order to have an automatic transform for the signals and images of arbitrary size. Thus, this property of matrix-based implementation gives us more flexibility that is necessary for tackling real-world data. We note that the use of matrix-based transform is not practical for very long 1-D signals, in which case special sparse matrix representation and operations have to be used, which ultimately boils down to the Mallat's algorithm. But for the 2-D transforms, this is not the case. If the computer can store the data matrix, then it can store the transformation matrix as well and can perform the matrix multiplication to transform. Most real-life images are of order of tens megapixels, so the matrix-type transforms are readily implementable even on modest personal computers.

This chapter is organized as follows. Section 2.1 describes the $\text{NDWT}_{\mathbf{C}}$ for 1-D and 2-D cases, respectively. For the 2-D case, we present a scale-mixing 2-D $\text{NDWT}_{\mathbf{C}}$. Section 2.2 illustrates a non-decimated complex wavelet spectra based on the modulus of the wavelet coefficients, while Section 2.3 proposes an effective way of utilizing the phase information leading to phase-based summaries enhancing discriminatory analysis of signal and images. Section 2.4 demonstrates a power of the proposed method with 1-D and 2-D applications and Section 2.5 contains some remarks and directions for future study.

## *2.1 Non-decimated Complex Wavelet Transform*

The wavelet and scaling functions for complex wavelets in Lawton (1993), Strang and Nguyen (1996), Lina (1999), and Zhang et al (1999) satisfy

$$\phi(x) = \sum_{k\in\mathbb{Z}} h_k \sqrt{2}\phi(2x-k) = h(x) + ig(x), \tag{29}$$

$$\psi(x) = \sum_{k\in\mathbb{Z}} g_k \sqrt{2}\phi(2x-k) = w(x) + iv(x), \tag{30}$$

where $h_k$ denotes the low pass filter and $g_k$ is defined as

$$g_k = (-1)^k \overline{h_{1-k}},$$

where $\overline{h_{1-k}}$ denotes a complex conjugate of $h_{1-k}$.

Using the complex wavelet bases, in this section, we define the non-decimated complex wavelet transform ($\mathrm{NDWT_C}$) separately for 1-D and 2-D cases by connecting the complex scaling and wavelet functions in non-decimated fashion.

### 2.1.1 1-D case

Suppose that a data vector $\boldsymbol{y} = (y_0, y_1, \ldots, y_{m-1})$ of size $m$ is given and that a multiresolution framework is specified. To understand the interplay between transform applied to discrete data and wavelet series representation of the function, we can link the data vector $\boldsymbol{y}$ to a function $f$ in terms of shifts of the scaling function at a multiresolution level $J$ as follows:

$$f(x) \;=\; \sum_{k=0}^{m-1} y_k \phi_{J,k}(x)$$

where $J-1 < \log_2 m \le J$, i.e. $J = \lceil \log_2 m \rceil$, and

$$\phi_{J,k}(x) = 2^{\frac{J}{2}} \phi(2^J(x-k)).$$

Since we consider the complex-valued filters in this wavelet transform, the scaling function is also complex-valued function as in Equation (29). Note that $2^J(x-k)$ is

used as an argument of scaling function, instead of $2^J x - k$ as in traditional wavelet transform, since we do not decimate.

Similarly, we can also express the data interpolating function $f$ in terms of wavelet coefficients as

$$f(x) = \sum_{k=0}^{m-1} c_{J_0,k}\phi_{J_0,k}(x) + \sum_{j=J_0}^{J-1}\sum_{k=0}^{m-1} d_{j,k}\psi_{j,k}(x)$$

where

$$\begin{aligned}\phi_{J_0,k}(x) &= 2^{\frac{J_0}{2}}\phi\left(2^{J_0}(x-k)\right),\\ \psi_{j,k}(x) &= 2^{\frac{j}{2}}\psi\left(2^{j}(x-k)\right),\end{aligned}$$

and $J_0$ is the coarsest decomposition level. Note that the non-decimated complex wavelet coefficients, $c_{J_0,k}$ and $d_{j,k}$, have both real and imaginary parts as

$$\begin{aligned}c_{J_0,k} &= \mathrm{Re}(c_{J_0,k}) + i\cdot\mathrm{Im}(c_{J_0,k}),\\ d_{j,k} &= \mathrm{Re}(d_{j,k}) + i\cdot\mathrm{Im}(d_{j,k}) \text{ for } j = J_0,\ldots,J-1.\end{aligned} \tag{31}$$

On the basis of these complex-valued wavelet coefficients we will, in the later sections, construct a wavelet spectra of modulus and as well as level-dependent phase summaries.

For a decomposition depth $p = J - J_0$, the $\mathrm{NDWT_C}$ transform of a vector $\boldsymbol{y}$ consists of a vector of "smooth" coefficients serving as a coarse approximation of $\boldsymbol{y}$,

$$\boldsymbol{c}_{(J_0)} = (c_{J_0,0}, c_{J_0,1}, \ldots, c_{J_0,m-1}),$$

and a set of "detail" coefficients containing information about the localized features in the data

$$\boldsymbol{d}_{(j)} = (d_{j,0}, d_{j,1}, \ldots, d_{j,m-1}), \quad j = J_0,\ldots,J-1.$$

The total number of coefficients of each vector is always $m$, which implies the redundancy of non-decimated transforms in contrast to the length-preserving standard

DWT. This results in total of $(p+1)\times m$ wavelet coefficients, with $p$ standing for number of levels of detail and 1 for the coarse level. The constancy of the level-wise shifts enables the $\text{NDWT}_{\text{C}}$ to be time invariant. The Mallat type algorithm for $\text{NDWT}_{\text{C}}$ is graphically illustrated in Figure 13. The coefficients in shaded boxes comprise the transform.

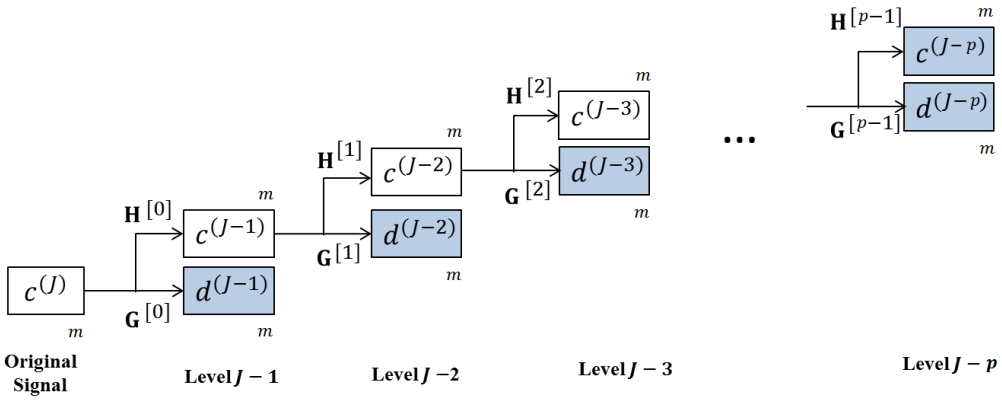



**Figure 13:** Graphical illustration of the NDCWT Mallat algorithm. The NDCWT decomposes the original signal of size $m$ to $p+1$ multiresolution subspaces including $p$ levels of detail coefficients and one level of coarse coefficients. The coefficients of the transform $\boldsymbol{d}^{(J-1)}, \boldsymbol{d}^{(J-2)}, \ldots, \boldsymbol{d}^{(J-p)}$, and $\boldsymbol{c}^{(J-p)}$ are in the shaded blocks.

Since the non-decimated wavelet transform is linear, the wavelet coefficients can be linked to the original signal by a matrix multiplication. For the proposed $\text{NDWT}_{\text{C}}$, we apply the complex scaling and wavelet filters in Equation (29) and (30) into the matrix formulation of NDWT defined in Kang and Vidakovic (2016) to obtain a matrix $W_m^{(p)}$. This matrix corresponds to a non-decimated complex wavelet transform of depth $p$, that is with $p$ levels of detail, and with $m$ as the size of input data. As we indicated earlier, the reason why we prefer the matrix-formulation is that it provides more flexibility especially in the 2-D case, with only a slight increase of computational complexity. Details for constructing the $W_m^{(p)}$ can be found in Kang and Vidakovic (2016). With use of $W_m^{(p)}$ we can transform a 1-D signal $\boldsymbol{y}$ of size $m$ to a non-decimated complex vector $\boldsymbol{d}$

$$\boldsymbol{d} \;=\; W_m^{(p)} \cdot \boldsymbol{y}$$

where $p$ is a depth of the transform and $p$ and $m$ are arbitrary. When the matrix wavelet transforms is used, one needs a weight matrix, $T_m^{(p)}$, to reconstruct back $\boldsymbol{y}$ from $\boldsymbol{d}$. The need for a weight matrix is caused by the inherent redundancy of the transform, and serves for deflation of the energy inflated by the transform. The weight matrix $T_m^{(p)}$ is defined as

$$T_m^{(p)} = \text{diag}(\overbrace{1/2^p, \ldots, 1/2^p}^{2m}, \overbrace{1/2^{p-1}, \ldots, 1/2^{p-1}}^{m}, \ldots, \overbrace{1/2, \ldots, 1/2}^{m}). \tag{32}$$

By using the weight matrix, the perfect reconstruction can be obtained as

$$\boldsymbol{y} \;=\; (W_m^{(p)})' \cdot T_m^{(p)} \cdot \boldsymbol{d}.$$

Graphical illustrations of matrix-based NDWT$_{\text{C}}$ for 1-D doppler signal is displayed in Figure 14.

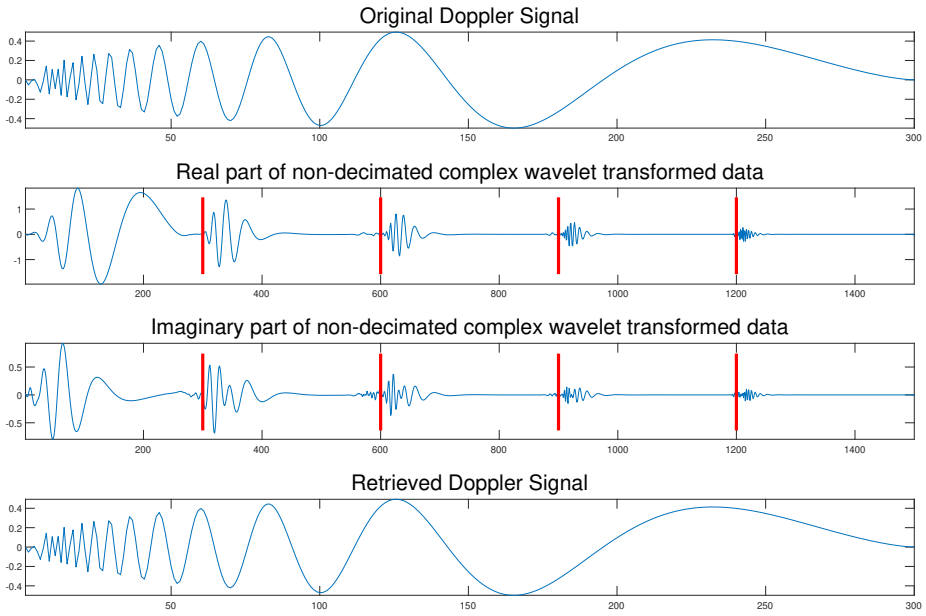



**Figure 14:** An example of matrix-based NDCWT for a Doppler signal of length 300.

### 2.1.2 2-D case

Next, we extend the 1-D definitions to the scale-mixing 2-D NDWT$_{\mathbf{C}}$ of $f(x, y)$ where $(x, y) \in \mathbb{R}^2$. The representation of non-decimated complex wavelets in 2-D can be implemented through one scaling function and three wavelet functions defined using Equations (29) and (30) as follows:

$$\begin{aligned}
\phi(x,y) &= \phi(x)\phi(y) = \Theta(x,y) + i\Psi(x,y), \\
\psi^{(h)}(x,y) &= \phi(x)\psi(y) = \xi^{(h)}(x,y) + i\zeta^{(h)}(x,y), \\
\psi^{(v)}(x,y) &= \psi(x)\phi(y) = \xi^{(v)}(x,y) + i\zeta^{(v)}(x,y), \\
\psi^{(d)}(x,y) &= \psi(x)\psi(y) = \xi^{(d)}(x,y) + i\zeta^{(d)}(x,y),
\end{aligned} \tag{33}$$

where symbols $h, v$, and $d$ denote the horizontal, vertical, and diagonal directions, respectively. This $h, v, d$ -notation is standardly used in 2-D wavelet literature and refers to directions in which the features are located in the hierarchy of multiresolution subspaces.

#### *2.1.2.1 Scale-Mixing 2-D Non-decimated Complex Wavelet Transform*

Although various versions of the 2-D WT can be constructed by appropriate tessellations of the detail spaces, here we utilize the scale-mixing 2-D NDWT$_{\mathbf{C}}$. As we will argue later, the use scale-mixing version is motivated by its remarkable flexibility, compressibility, and ease of computation.

For the scale-mixing 2-D NDWT$_{\mathbf{C}}$, we define the wavelet atoms as follows:

$$\begin{aligned}
\phi_{J_{01},J_{02},k_1,k_2}(x,y) &= \Theta_{J_{01},k_1,k_2}(x,y) + i\Psi_{J_{02},k_1,k_2}(x,y), \\
\psi^{(h)}_{J_{01},j_2,k_1,k_2}(x,y) &= \xi^{(h)}_{J_{01},k_1,k_2}(x,y) + i\zeta^{(h)}_{j_2,k_1,k_2}(x,y), \\
\psi^{(v)}_{j_1,J_{02},k_1,k_2}(x,y) &= \xi^{(v)}_{j_1,k_1,k_2}(x,y) + i\zeta^{(v)}_{J_{02},k_1,k_2}(x,y), \\
\psi^{(d)}_{j_1,j_2,k_1,k_2}(x,y) &= \xi^{(d)}_{j_1,k_1,k_2}(x,y) + i\zeta^{(d)}_{j_2,k_1,k_2}(x,y),
\end{aligned} \tag{34}$$

where $k_1 = 0, \ldots, m-1$, $k_2 = 0, \ldots, n-1$, $j_1 = J_{01}, \ldots, J-1$, $j_2 = J_{02}, \ldots, J-1$, and $J = \lceil \log_2 \min(m, n) \rceil$. Note that $J_{01}$ and $J_{02}$ are the coarsest decomposition levels of

rows and columns. Then any function $f \in L_2(\mathbb{R}^2)$ can be expressed as

$$\begin{aligned} f(x,y) &= \sum_{k_1}\sum_{k_2} c_{J_{01},J_{02},k_1,k_2}\phi_{J_{01},J_{02},k_1,k_2}(x,y) \\ &+ \sum_{j_2>J_{02}}\sum_{k_1}\sum_{k_2} d^{(h)}_{J_{01},j_2,k_1,k_2}\psi^{(h)}_{J_{01},j_2,k_1,k_2}(x,y) \\ &+ \sum_{j_1>J_{01}}\sum_{k_1}\sum_{k_2} d^{(v)}_{j_1,J_{02},k_1,k_2}\psi^{(v)}_{j_1,J_{02},k_1,k_2}(x,y) \\ &+ \sum_{j_1>J_{02}}\sum_{j_2>J_{01}}\sum_{k_1}\sum_{k_2} d^{(d)}_{j_1,j_2,k_1,k_2}\psi^{(d)}_{j_1,j_2,k_1,k_2}(x,y), \end{aligned}$$

which defines a scale-mixing NDWT$_{\mathbf{C}}$. Unlike the standard 2-D NDWT$_{\mathbf{C}}$ denoting a scale as only $j$, we denote such mixed two scales as a pair $(j_1, j_2)$ capturing the energy flux between the scales.

Finally, the resulting scale-mixing non-decimated complex wavelet coefficients are

$$\begin{aligned} c_{J_{01},J_{02},k_1,k_2} &= \iint f(x,y)\overline{\phi}_{J_{01},J_{02},k_1,k_2}(x,y)\,dxdy \\ &= \mathrm{Re}(c_{J_{01},J_{02},k_1,k_2}) + i\cdot\mathrm{Im}(c_{J_{01},J_{02},k_1,k_2}), \\ d^{(h)}_{J_{01},j_2,k_1,k_2} &= \iint f(x,y)\overline{\psi}^{(h)}_{J_{01},j_2,k_1,k_2}(x,y)\,dxdy \\ &= \mathrm{Re}(d^{(h)}_{J_{01},j_2,k_1,k_2}) + i\cdot\mathrm{Im}(d^{(h)}_{J_{01},j_2,k_1,k_2}), \\ d^{(v)}_{j_1,J_{02},k_1,k_2} &= \iint f(x,y)\overline{\psi}^{(v)}_{j_1,J_{02},k_1,k_2}(x,y)\,dxdy \\ &= \mathrm{Re}(d^{(v)}_{j_1,J_{02},k_1,k_2}) + i\cdot\mathrm{Im}(d^{(v)}_{j_1,J_{02},k_1,k_2}), \\ d^{(d)}_{j_1,j_2,k_1,k_2} &= \iint f(x,y)\overline{\psi}^{(d)}_{j_1,j_2,k_1,k_2}(x,y)\,dxdy \\ &= \mathrm{Re}(d^{(d)}_{j_1,j_2,k_1,k_2}) + i\cdot\mathrm{Im}(d^{(d)}_{j_1,j_2,k_1,k_2}). \end{aligned} \tag{35}$$

where $\overline{\phi}$ denotes the complex conjugate of $\phi$. Note that the non-decimated complex wavelet coefficients in Equation (36) have both real and imaginary parts as complex numbers.

Similar to the 1-D case, we can connect the 2-D wavelet coefficients to the original image through a matrix equation. Here we apply the complex scaling and wavelet filters in Equation (33) into the matrix formulation of NDWT to obtain $W_m^{(p_1)}$ and $W_n^{(p_2)}$ that are non-decimated complex wavelet matrices with $p_1$, $p_2$ detail levels and

$m$, $n$ size of row and column, respectively. For 2-D case, using the matrix-formulation allows to use any non-square image. More rigorous details on these matrix formulation for real-valued wavelets can be found in Kang and Vidakovic (2016).

Next, we can transform a 2-D image $\boldsymbol{A}$ of size $m \times n$ to a non-decimated complex wavelet transformed matrix $\boldsymbol{B}$ with depth $p_1$ and $p_2$ as

$$\boldsymbol{B} = W_m^{(p_1)} \cdot \boldsymbol{A} \cdot (W_n^{(p_2)})^\dagger \tag{36}$$

where $p_1, p_2, m$, and $n$ are arbitrary. The $W^\dagger$ denotes a Hermitian transpose of matrix $W$. Note that Equation (36) represents a finite-dimensional implementation of Equation (36) for $f(x)$ sampled in a form of matrix, as $f(x, y)$. Then the resulting transformed matrix $\boldsymbol{B}$ has a size of $(p_1 + 1)m \times (p_2 + 1)n$. Similar to the 1-D case, for perfect reconstruction of $\boldsymbol{A}$, we need two weight matrices, that is, $p_1$- and $p_2$-level weight matrices $T_m^{(p_1)}$ and $T_n^{(p_2)}$. The matrices are defined as in Equation (32) with different $m, n, p_1$, and $p_2$. By using the weight matrices, the perfect reconstruction can be performed as

$$\boldsymbol{A} = W_m^{(p_1)} \cdot T_m^{(p_1)} \cdot \boldsymbol{B} \cdot T_n^{(p_2)} \cdot (W_n^{(p_2)})^\dagger.$$

Graphical illustrations of matrix-based scale-mixing 2-D $\mathrm{NDWT_C}$ for a lena image is displayed in Figure 15.

## *2.2 Non-decimated Complex Wavelet Spectra*

High-frequency, time series data from various sources often possess hidden patterns that reveal the effects of underlying functional differences. Such patterns cannot be elucidated by basic descriptive statistics or trends in some real-life situations. For example, the high-frequency pupillary response behavior (PRB) data collected during computer-based interaction captures the changes in pupil diameter in response to various stimuli. Researchers found that there may be underlying unique patterns hidden within PRB data, and these patterns may reveal the intrinsic individual differences

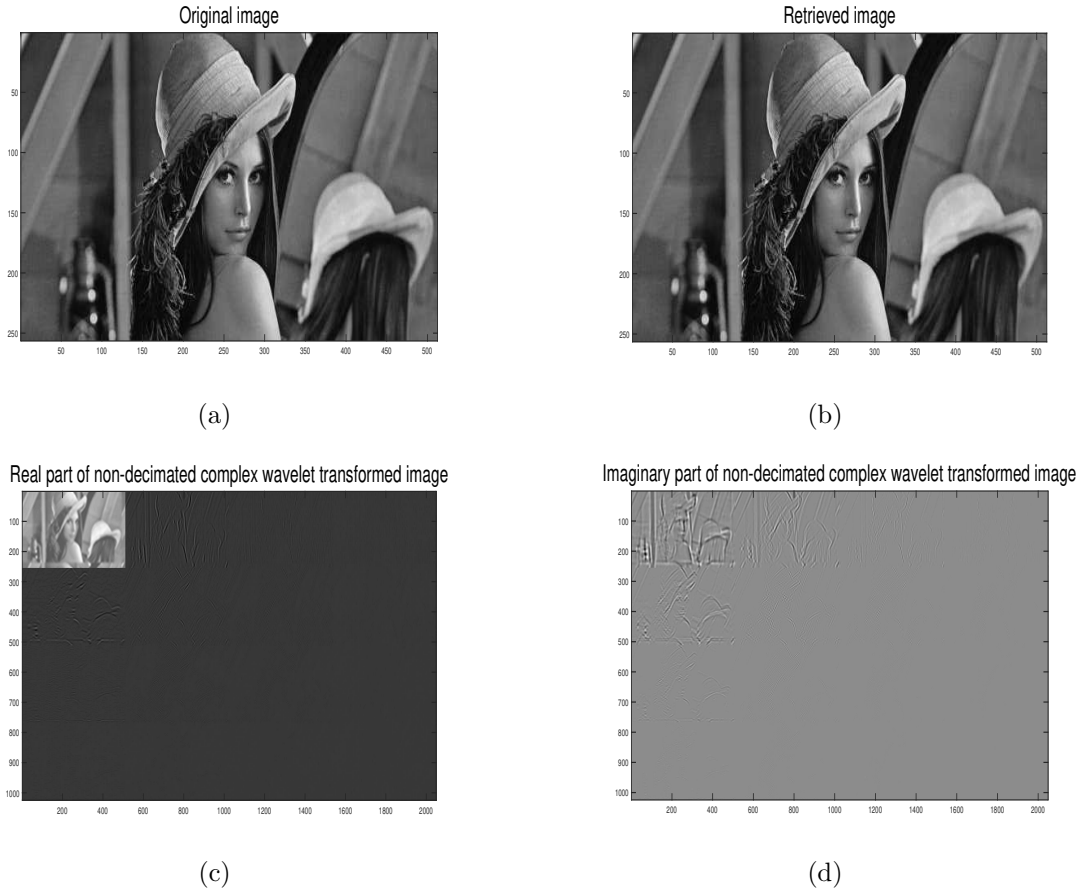



**Figure 15:** An example of matrix-based scale-mixing 2-D NDCWT for a lena image of size $256 \times 512$. (a) Original image, (b) Recovered image, (c) Real part of non-decimated complex wavelet transformed image, (d) Imaginary part of non-decimated complex wavelet transformed image.

in cognitive, sensory and motor functions (Moloney et al, 2006). Yet, such patterns cannot be explained by the trends and traditional statistical summaries, for the magnitude of the pupil diameter depends on the ambient light, not on the inherent eye function or link to the cognitive task. When the intrinsic individual functional differences cannot be modeled by statistical tools in the domain of the data acquisition, the transformed time/scale or time/frequency domains may help. High frequency data as a rule scale, and this scaling can be quantified by the Hurst exponent as an optional measure to characterize the patients.

The Hurst exponent is an informative summary of the behavior of self-similar

processes and is also related to the presence of long memory and degree of fractality in signals and images. Among many methods for estimating the Hurst exponent, the wavelet-based methods have shown to be particularly accurate. The main contribution of this chapter is a construction of the non-decimated complex wavelet spectra with extension of the method into the scale-mixing 2-D non-decimated complex wavelet spectra for 2-D case, all with the goal of assessing the Hurst exponent or its equivalent spectral slope. As a bonus, the complex valued wavelets would provide informative multiscale phase information.

Next we briefly overview the notion of self-similarity and its link with the Hurst exponent. Suppose that a random process $\{X(t), t \in \mathbb{R}\}$ for some $\lambda > 0$ satisfies

$$X(\lambda t) \stackrel{d}{=} \lambda^H X(t) \text{ for any}$$

where $\stackrel{d}{=}$ stands for equality of all joint finite-dimensional distributions, then, $X(t)$ is self-similar with self-similarity index $H$, traditionally called Hurst exponent.

If $X(t)$ is transformed in the wavelet domain and $d_{j,k}$ is the wavelet coefficient at scale $j$ and shift $k$ in standard DWT, can be shown that

$$d_{j,k} \stackrel{d}{=} 2^{-j(H+\frac{1}{2})} d_{0,k}.$$

Here the notation $\stackrel{d}{=}$ denotes the equality in distribution. For the non-decimated complex wavelets, however, $d_{j,k}$ is a complex number, as in Equation (31), and we use $|d_{j,k}|$ for a modulus of $d_{j,k}$,

$$|d_{j,k}| = \sqrt{Re(d_{j,k})^2 + Im(d_{j,k})^2}, \;\; j = J_0, \ldots, J-1.$$

The Equation (2.2) now can be re-stated as

$$|d_{j,k}| \stackrel{d}{=} 2^{-j(H+\frac{1}{2})} |d_{0,k}|, \;\; j = J_0, \ldots, J-1.$$

If the process $X(t)$ possesses stationary increments, for any $q > 0$, $E(|d_{0,k}|) = 0$ and $E(|d_{0,k}|^q) = E(|d_{0,0}|^q)$. Thus,

$$E(|d_{j,k}|^q) = C 2^{-jq(H+\frac{1}{2})}, \;\; j = J_0, \ldots, J-1 \tag{37}$$

where $C = E(|d_{0,0}|^q)$. Although $q$ could be arbitrary nonnegative, here we will use standard $q = 2$ that has "energy" interpretation. By taking logarithms on both sides in Equation (37), we can obtain the non-decimated complex wavelet spectrum of $X(t)$ as

$$S(j) = \log_2(E(|d_{j,k}|^2)) = -j(2H + 1) + C', \;\; j = J_0, \ldots, J - 1. \tag{38}$$

Note that the wavelet spectrum describes the relationship between the scales and energies at the scales. If along the scales the energies decay regularly, this indicates that there is a regular scaling in the data, and we can measure a self-similarity via a rate of energy decay. Operationally, we find the slope in regression of log energies to scale indices, as in Equation (38), and use it to estimate the Hurst exponent. For discrete observed data of size $m$, we use empirical counterpart of $S(j)$ defined as

$$\hat{S}(j) = \log_2 \frac{1}{m} \sum_{k=1}^{m} |d_{j,k}|^2 = \log_2 \overline{|d_{j,k}|^2}, \;\; j = J_0, \ldots, J - 1.$$

We can plot the set of $\hat{S}(j)$ against $j$ as $\big(j,\ \hat{S}(j)\big)$, which is called 2nd order Logscale Diagram (2-LD) and this is the wavelet spectra as displayed in Figure 16. Finally, we can estimate the slope of the spectra usually by regression methodology (an ordinary, weighted, or robust regression) and use it to estimate the Hurst exponent $H$, as $\hat{H} = -(\text{slope} + 1)/2$. More details on wavelet spectra method and its applications can be found in Veitch and Abry (1999), Mallat (2009), Ramírez and Vidakovic (2013), and Roberts et al (2017).

### 2.2.1 Scale-Mixing 2-D Non-decimated Complex Wavelet Spectra

To introduce a scale-mixing 2-D non-decimated complex wavelet spectra, consider a 2-D fractional Brownian motion (fBm) in two dimensions, $B_H(\mathbf{u})$ for $\mathbf{u} \in [0, 1] \times [0, 1]$ and $H \in (0, 1)$. The 2-D fBm, $B_H(\mathbf{u})$, is a random process with stationary zero-mean Gaussian increments leading to

$$B_H(a\mathbf{t}) \stackrel{d}{=} a^H B_H(\mathbf{t}) \;\text{ for any }\; a \geq 0.$$

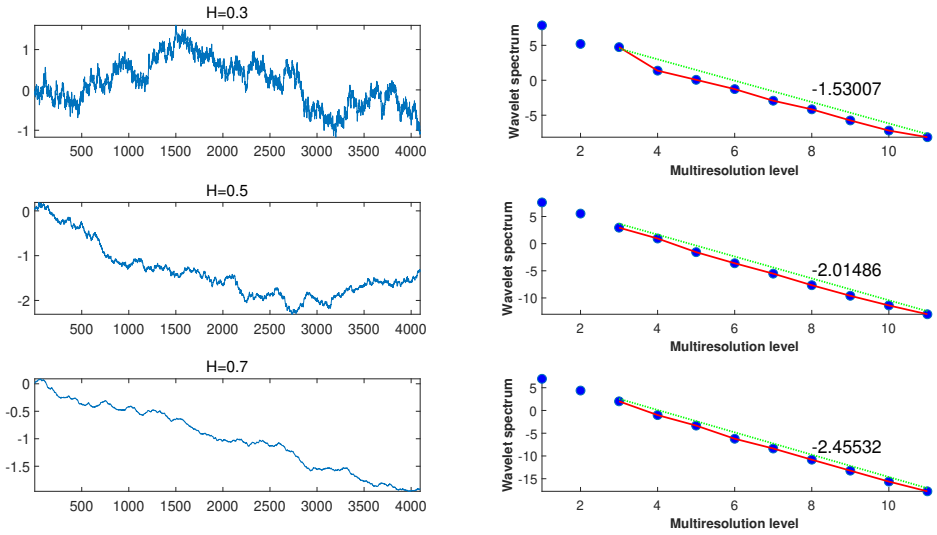



**Figure 16:** Examples of non-decimated complex wavelet spectra using the modulus of coefficients. The slopes are -1.53007, -2.01486, and -2.45532 corresponding to estimator $\hat{H} = 0.2650, 0.5074,$ and $0.7277$. The original 4096-length signals were simulated as a fBm with Hurst exponent 0.3, 0.5, and 0.7.

For this process, the scale-mixing non-decimated complex wavelet detail coefficients can be defined as

$$d_{(j_1,j_2+s,k_1,k_2)} = 2^{\frac{1}{2}(j_1+j_2+s)} \int B_H(\mathbf{u}) \overline{\psi} \left(2^{j_1}(u_1 - k_1), 2^{j_2+s}(u_2 - k_2)\right) d\mathbf{u}$$

where $\overline{\psi}$ denotes the complex conjugate of $\psi^{(d)}$ defined in Equation (34). In this chapter, we only consider the main diagonal hierarchy whose 2-D scale indices coincide as $j_1 = j_2 = j$ and thus $J_{01} = J_{02} = J_0$.

Since the $d_{(j,j+s,k_1,k_2)}$ is a complex number, we need to consider its modulus

$$|d_{(j,j+s,k_1,k_2)}| = \sqrt{Re(d_{(j,j+s,k_1,k_2)})^2 + Im(d_{(j,j+s,k_1,k_2)})^2}, \quad j = J_0, \ldots, J-1.$$

Then average of squared modulus of the coefficients is calculated as

$$\begin{aligned} \mathbb{E}\left[|d_{(j,j+s,k_1,k_2)}|^2\right] &= 2^{2j+s} \int \psi\left(2^j(u_1 - k_1), 2^{j+s}(u_2 - k_2)\right) \\ &\times \overline{\psi}\left(2^j(v_1 - k_1), 2^{j+s}(v_2 - k_2)\right) \mathbb{E}\left[B_H(\mathbf{u}) B_H(\mathbf{v})\right] d\mathbf{u}\, d\mathbf{v}. \end{aligned} \tag{39}$$

As a result, the Equation (39) can be restated as

$$\mathbb{E}\left[|d_{(j,j+s,k_1,k_2)}|^2\right] = 2^{-j(2H+2)}\ V_{\psi,s}(H), \tag{40}$$

and its proof is provided in Jeon et al (2014). Note that $V_{\psi,s}(H)$ does not depend on the scale $j$ but on $\psi$, $H$ and $s$. Finally, the scale-mixing 2-D non-decimated complex wavelet spectrum is defined by taking logarithms on both sides of the Equation (40),

$$S(j, j+s) = \log_2(\mathbb{E}(|d_{j,j+s,k_1,k_2}|^2)) = -j(2H+2) + C', \quad j = J_0, \ldots, J-1.$$

Similar to the 1-D case, its empirical counterpart is

$$\hat{S}(j, j+s) = \log_2 \frac{1}{mn} \sum_{k_1=1}^{m} \sum_{k_2=1}^{n} |d_{j,j+s,k_1,k_2}|^2 = \log_2 \overline{|d_{j,j+s,k_1,k_2}|^2}, \quad j = J_0, \ldots, J-1$$

where $m$ and $n$ are row and column sizes, respectively. The way of constructing wavelet spectra goes along the lines of the construction in 1-D case, except for the expressing the Hurst exponent from the slope. In the 2-D case $H$ is estimated as $\hat{H} = -(\text{slope} + 2)/2$.

## *2.3 Phase-based Statistics for Classification Analysis*

In the area of Fourier representations, there is a considerable of interest about the information the phase carries about signals or images (Oppenheim and Li, 1981; Levi and Stark, 1983). For complex wavelet domains, there is also an interest about information related to interactions between scales and spatial symmetries contained in the phase, as investigated by Lina (1997), Lina (1999), and Jeon et al (2014). Therefore, it is natural to explore the role of phase in the complex-valued wavelet coefficients of signals or images. Theoretically, it is known that the original signal can be reconstructed from the phase information only. We briefly describe two experiments conducted in Oppenheim and Li (1981) and Jeon et al (2014) for the Fourier and wavelet transforms, respectively. Both experiments transformed two different images of the same size to complex-valued domains and from the coefficients obtained

modulus and phases. Then the phase information was switched and images were reconstructed from the original modulus and switched phases. Surprisingly, both reconstructed images were more alike to the phase corresponding images, that is, the phase information dominated the modulus information. Motivated by these experiment results, Jeon et al (2014) proposed a way of utilizing phase information for discriminatory analysis. They suggested a summary statistic of the phases at the finest levels and demonstrated in a particular classification task the accuracy can be improved, albeit only slightly. This is because the phases from the finest level only were used. Wavelet coefficients at each level, however, have slightly different information on the given data, which is the one of advantages of their multiresolution nature. Generally, the phase information from different levels may be complementary. If we utilize phase information on the other levels, an overall accuracy would be further improved. In this section we propose more extensive phase-based modalities using $\mathrm{NDWT_C}$ for signal or image classification problems to improve an overall performance.

The phase of a non-decimated complex wavelet coefficient defined in Equation (31) is

$$\begin{aligned} \angle d_{j,k} &= \arctan\left(\frac{Im(d_{j,k})}{Re(d_{j,k})}\right), \\ \angle d_{(j,j+s,k_1,k_2)} &= \arctan\left(\frac{Im(d_{(j,j+s,k_1,k_2)})}{Re(d_{(j,j+s,k_1,k_2)})}\right) \end{aligned}$$

for 1-D and 2-D cases, respectively. Then, an average of phases at level $j$ for both cases can be calculated as

$$\begin{aligned} \angle d_j &= \frac{1}{m}\sum_{k=1}^{m} \angle d_{j,k}, \quad j = J_0, \ldots, J-1, \\ \angle d_{j,j+s} &= \frac{1}{mn}\sum_{k_1=1}^{m}\sum_{k_2=1}^{n} \angle d_{(j,j+s,k_1,k_2)}, \quad j = J_0, \ldots, J-1 \end{aligned} \tag{41}$$

for 1-D and 2-D cases, respectively. Finally, we set the averages of phases at all considered multiresolution level $j$ as new descriptors in a wavelet-based classification

analysis. Note that these descriptors do not indicate any scaling regularity, unlike the modulus, as seen in Figure 17.

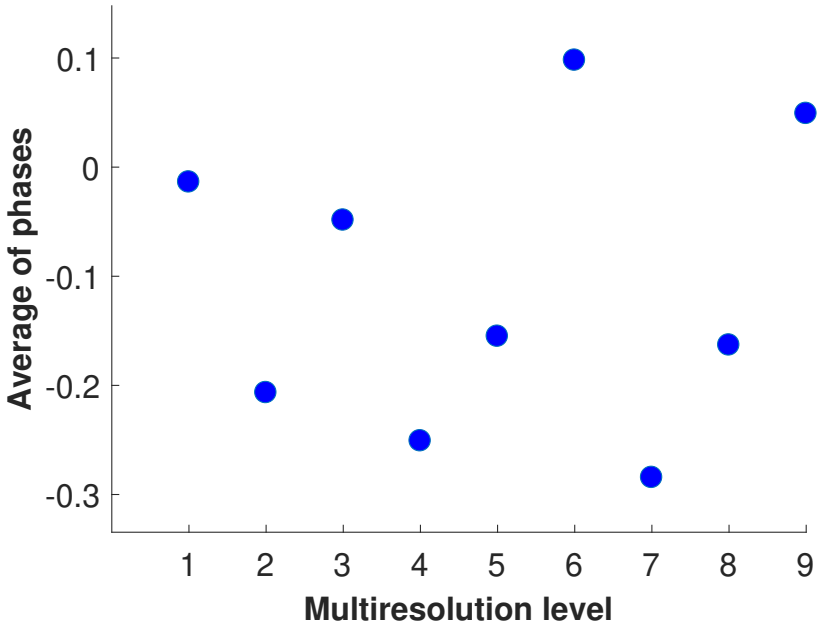



**Figure 17:** Visualization of phase averages at all multiresolution levels.

## *2.4 Applications*

Since the proposed methods in Chapter 2, 3, and 4 share the same datasets, we set aside all analysis results in Chapter 5 (APPLICATIONS). In particular, the detail explanations in terms of $\text{NDWT}_{\text{C}}$ for all four applications can be found in section 5.1.3.1, 5.2.3.1, 5.3.3.1, and 5.4.3.1.

## *2.5 Conclusions and Future Studies*

In this chapter we explored a non-decimated complex wavelet transform ($\text{NDWT}_{\text{C}}$) for both 1-D and 2-D cases. We demonstrated that the proposed spectra performs well in classification problems, with phase-based statistics improving the classification accuracy. We presented comparative simulations in four real-life applications and found that the classification procedures induced by the $\text{NDWT}_{\text{C}}$ outperforms the

$\mathrm{WT}_{\mathbf{C}}$ and NDWT. Thus, the $\mathrm{NDWT}_{\mathbf{C}}$ may be of interest to researchers seeking more efficient wavelet-based classification method for signals or images with intrinsic self-similarity.

As possible future directions we may be interested in different ways of calculating the spectral slopes, as similarly as in Hamilton et al (2011) or Feng et al (2018). Additionally, for the scale-mixing 2-D $\mathrm{NDWT}_{\mathbf{C}}$, using $d^{(h)}$ and $d^{(v)}$ in addition to $d^{(d)}$ for phase statistics could potentially improve the performance. Finally using different wavelet filters for rows and columns in the scale-mixing 2-D $\mathrm{NDWT}_{\mathbf{C}}$ would provide more modeling freedom. For instance, one can search for a wavelet, or pair of wavelets, in a library of complex-valued wavelets for which classification is optimal.

In the spirit of reproducible research we prepared an illustrative demo as a stand alone MATLAB software with solved examples. The demo is posted on the repository Jacket Wavelets `http://gtwavelet.bme.gatech.edu/`.

# CHAPTER III

# NON-DECIMATED QUATERNION WAVELET SPECTRAL TOOLS

The redundancy in $\mathrm{WT}_{\mathbf{C}}$ in Section 1.5 and the NDWT in Section 1.7 is of different nature. The Chapter 2 focused on non-decimated complex wavelet transforms ($\mathrm{NDWT}_{\mathbf{C}}$) that combine $\mathrm{WT}_{\mathbf{C}}$ and NDWT. Specifically, $\mathrm{NDWT}_{\mathbf{C}}$ produces redundant wavelet coefficients both as complex numbers and by NDWT as oversampled. The chapter suggested a way of building phase-based statistics as variables for classification problems and showed significant increase in precision of classification, compared to other existing wavelet-based methods. Furthermore, the $\mathrm{NDWT}_{\mathbf{C}}$ is more flexible than the decimated wavelet transforms because of the matrix-based implementation proposed in Kang and Vidakovic (2016). The decimated wavelet transform methods including $\mathrm{WT}_{\mathbf{C}}$ and even convolution-based NDWT can be routinely applied to signals and squared images dyadic sizes (Lina, 1999; Percival and Walden, 2006). However, a real-world data typically do not have such sizes and need pre-processing prior to application. The matrix-based NDWT enables us to directly analyze 1-D signals of an 2-D images of arbitrary size. More details can be found in Kang and Vidakovic (2016).

As an extension of the methodology suggested in Chapter 2, given the useful characteristics of the QWT in Section 1.6 and matrix-based NDWT in Section 1.7, we propose a non-decimated quaternion wavelet transform (NDQWT) and its wavelet spectra defined by quaternionic modulus and the three phases. Since the QWT is an extension of the $\mathrm{WT}_{\mathbf{C}}$, we expect that classification accuracy would improve with QWT-defined spectral tools. The modulus of the QWT behaves as the wavelet spectra

in a conventional wavelet transform, so that our main focus is on the contribution by the three phases. Several researches including Billow (1999) and Soulard and Carre (2010) also focused on exploiting phases of quaternion wavelets. However, as far as we know, there is no proposal of quaternionic phase-based levelwise statistics for use in classification problems, and wider, for machine learning. Also, as we pointed out, existing use of phase depended on the artificially constructed QWT. Taken together with non-decimated nature of the underlying transform, the proposed spectral tools are novel in terms of providing new modalities for classification problems. And finally, the goal of this study is to demonstrate superiority of the proposed method over several competing wavelet-based methods through applications in real-data.

The chapter is organized as follows. Next, Section 3.1 explains the NDQWT for 1-D and 2-D cases, respectively. For the 2-D case, we present the scale-mixing version of 2-D NDQWT. Section 3.2 describes the non-decimated quaternion wavelet-based spectra focusing on the modulus information of the coefficients, while Section 3.3 suggests construction of a three phase-based statistics as new covariates in discriminatory analysis. In Section 3.4, the proposed tools are applied on 1-D and 2-D real-data, and finally, concluding remarks and directions for future study are given in Section 3.5.

## *3.1 Non-decimated Quaternion Wavelet Transform*

The quaternion scaling and wavelet functions in Bayro-Corrochano (2005) and Chan et al (2008) satisfy

$$\phi(x) = \sum_{k\in\mathbb{Z}} h_k\sqrt{2}\phi(2x-k) = w_0(x) + i\cdot w_1(x) + j\cdot w_2(x) + k\cdot w_3(x), \quad (42)$$

$$\psi(x) = \sum_{k\in\mathbb{Z}} g_k\sqrt{2}\phi(2x-k) = v_0(x) + i\cdot v_1(x) + j\cdot v_2(x) + k\cdot v_3(x), \quad (43)$$

where $h_k$ denotes the low pass filter and $g_k$ is the high pass filter. We define the non-decimated quaternion wavelet transform (NDQWT) separately for 1-D and 2-D

cases by connecting the quaternion low- and high-pass filters with a non-decimation property.

### 3.1.1 1-D case

Given a specified a multiresolution framework and a data vector $\boldsymbol{y} = (y_0, y_1, \ldots, y_{m-1})$ of size $m$, the discrete data vector $\boldsymbol{y}$ can be connected to a function $f$ which is a linear combination of shifts of the scaling function at some decomposition level $J$,

$$f(x) = \sum_{k=0}^{m-1} y_k \phi_{J,k}(x) \tag{44}$$

where $J - 1 < \log_2 m \leq J$, i.e. $J = \lceil \log_2 m \rceil$ and

$$\phi_{J,k}(x) = 2^{\frac{J}{2}} \phi(2^J(x-k)). \tag{45}$$

For the NDQWT, the scaling functions in Equation (44) and (45) will be the quaternion-valued scaling functions from Equation (42).

Alternatively, the data interpolating function $f$ also can be represented in terms of wavelet coefficients as follows:

$$f(x) = \sum_{k=0}^{m-1} c_{J_0,k} \phi_{J_0,k}(x) + \sum_{j=J_0}^{J-1} \sum_{k=0}^{m-1} d_{j,k} \psi_{j,k}(x), \tag{46}$$

where

$$\begin{aligned} \phi_{J_0,k}(x) &= 2^{\frac{J_0}{2}} \phi\left(2^{J_0}(x-k)\right), \\ \psi_{j,k}(x) &= 2^{\frac{j}{2}} \psi\left(2^j(x-k)\right), \end{aligned} \tag{47}$$

and $J_0$ is the coarsest decomposition level. Note that $2^J(x-k)$ is used inside of the scaling function in Equation (45) and (47) instead of $2^J x - k$ for the traditional decimation in wavelet domain in order to make this wavelet decomposition non-decimated. By using $2^J(x-k)$, the shift indicator $k$ remains constant at all levels, and this corresponds to levelwise sampling rate that results in the non-decimation. In comparison,

for DWT case, the shifts are level dependent as $2^{-j}k$. After performing the NDQWT on the vector $\boldsymbol{y}$, one obtains a vector of smooth coefficients as

$$\boldsymbol{c}_{(J_0)} = (c_{J_0,0}, c_{J_0,1}, \ldots, c_{J_0,m-1}) \tag{48}$$

which corresponds to the coarsest approximation of $\boldsymbol{y}$. Likewise, the vectors of detail coefficients are given as

$$\boldsymbol{d}_{(j)} = (d_{j,0}, d_{j,1}, \ldots, d_{j,m-1}), \quad j = J_0, \ldots, J-1, \tag{49}$$

which carry fine-scale information within the input $\boldsymbol{y}$. Of course, the number of coefficients in the vectors $\boldsymbol{c}^{(J_0)}$ and $\boldsymbol{d}^{(j)}$ is always $m$ and this is due to the non-decimated property of NDQWT. As a result, we obtain total $(p+1) \times m$ wavelet coefficients, with $pm$ details and $p$ coarse coefficients. The Mallat type of algorithm for forward NDQWT is graphically illustrated in Figure 18.

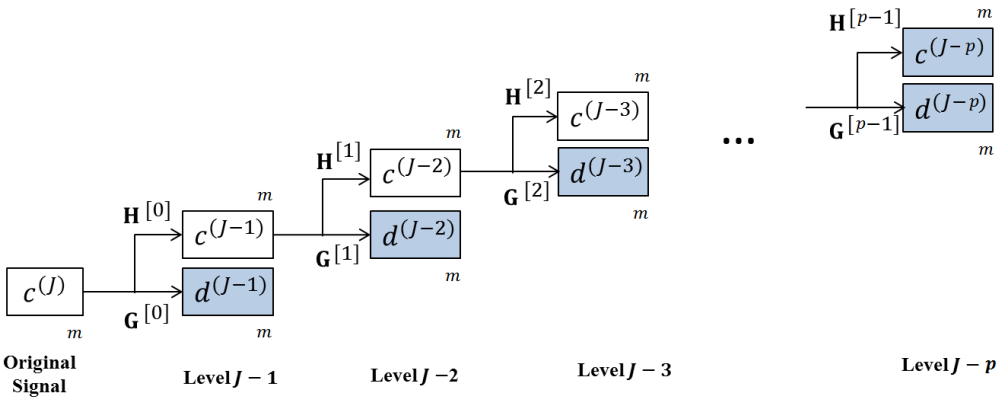



**Figure 18:** Graphical illustration of the Mallat algorithm. The NDQWT decomposes the original signal of size $m$ to $p+1$ multiresolution subspaces comprising of $p$ levels of detail coefficients and one level of coarse coefficients. The The shaded boxes represent the transformation, $\boldsymbol{d}^{(J-1)}, \boldsymbol{d}^{(J-2)}, \ldots, \boldsymbol{d}^{(J-p)}$, and $\boldsymbol{c}^{(J-p)}$.

Next we focus on each wavelet coefficient in Equation (48) and (49). Since the scaling and wavelet functions in Equation (47) are quaternion-valued, the non-decimated quaternion wavelet coefficients $c_{J_0,k}$ and $d_{j,k}$ in Equation (46) have one real and three

imaginary parts as

$$\begin{aligned} c_{J_0,k} &= \mathrm{Re}(c_{J_0,k}) + i \cdot \mathrm{Im}^i(c_{J_0,k}) + j \cdot \mathrm{Im}^j(c_{J_0,k}) + k \cdot \mathrm{Im}^k(c_{J_0,k}), \\ d_{j,k} &= \mathrm{Re}(d_{j,k}) \ + i \cdot \mathrm{Im}^i(d_{j,k}) \ + j \cdot \mathrm{Im}^j(d_{j,k}) \ + k \cdot \mathrm{Im}^k(d_{j,k}), \end{aligned} \tag{50}$$

where $j = J_0, \ldots, J-1$ and $\mathrm{Im}^i(q) = q_1$, $\mathrm{Im}^j(q) = q_2$, and $\mathrm{Im}^k(q) = q_3$. These quaternion-valued wavelet coefficients would be considered in the later sections for construction of a NDQWT-based spectra, as well as level-dependent phase summaries.

Note that the multiresolution levels and location parameters have been conventionally denoted as $j$ and $k$ in wavelet-based multiresolution analysis and this is the same for the $j$ and $k$ denoting the second and third imaginary unit in the quaternion algebra. These duplicated denotations can cause a little confusing expression of the quaternion-valued wavelet coefficients, $d_{j,k}$ in Equation (50). However, they can be so clearly and easily separated in the context that we would continuously denote both multiresolution level and second imaginary unit as the $j$, and both location parameter and third imaginary unit as the $k$.

Unlike standard convolution-based approach, the matrix-based NDWT can provide several additional features. First, the matrix-formulation allows us to use any non-dyadic size signal. Due to typical sizes of the signals and images processed, the matrix based transform does not significantly increase practical computational complexity. Thus, we incorporate the quaternion scaling and wavelet filters in Equation (42) into the matrix formulation of NDWT in order to utilize its convenient properties. We obtain a non-decimated quaternion wavelet matrix, $W_m^{(p)}$, that is formed directly from quaternion wavelet filter coefficients with $p$ detail levels and $m$ size of input data. Details for constructing $W_m^{(p)}$ are explained in Kang and Vidakovic (2016). To obtain a non-decimated quaternion wavelet transformed vector $\boldsymbol{d}$ with depth $p$ from a 1-D signal $\boldsymbol{y}$ of size $m \times 1$, we multiply $\boldsymbol{y}$ by $W_m^{(p)}$ as

$$\boldsymbol{d} \ = \ W_m^{(p)} \cdot \boldsymbol{y},$$

where $p$ and $m$ are arbitrary. For the reconstruction from $\boldsymbol{d}$ to $\boldsymbol{d}$, we need an additional weight matrix for $W_m^{(p)}$ as $T_m^{(p)}$ that is defined as

$$T_m^{(p)} = \text{diag}(\overbrace{1/2^p, \ldots, 1/2^p}^{2m}, \overbrace{1/2^{p-1}, \ldots, 1/2^{p-1}}^{m}, \ldots, \overbrace{1/2, \ldots, 1/2}^{m}). \tag{51}$$

Utilizing the weight matrix, $T_m^{(p)}$, we can perform the perfect reconstruction as

$$\boldsymbol{y} \;=\; (W_m^{(p)})^\dagger \cdot T_m^{(p)} \cdot \boldsymbol{d}$$

where the $W^\dagger$ denotes a Hermitian transpose matrix of $W$. Graphical illustrations of matrix-based NDQWT is displayed in Figure 19.

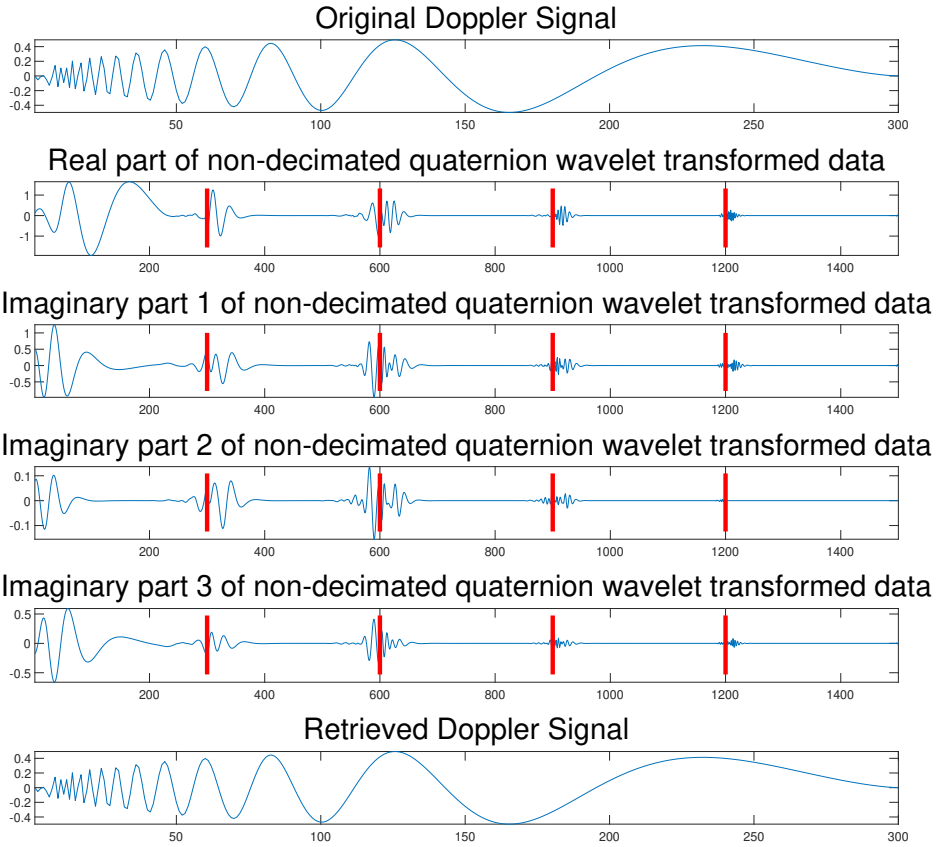



**Figure 19:** An example of matrix-based NDQWT of a Doppler signal of length 300.

### 3.1.2 2-D case

The real power of matrix implementation of wavelet transforms can be seen in 2-D cases, where the so called scale-mixing property is utilized. The scale-mixing transforms typically have lower entropy compared to traditional 2-D transforms which is beneficial in tasks of wavelet shrinkage. For the scaling analysis, scale mixing transforms enable definition of a range of spectra along the hierarchies of scale-mixing spaces. In this chapter, we focus only on diagonal hierarchy, but emphasize that the spectral tools can be further generalized.

In this section, the 1-D NDQWT from Section 3.1.1 is extended to a scale-mixing 2-D NDQWT of $f(x, y)$ where $(x, y) \in \mathbb{R}^2$. The decomposition has one scaling function and three wavelet functions defined as tensor product of functions Equations (42):

$$\begin{aligned}
\phi(x,y) &= \phi(x)\phi(y) \\
&= \mu(x,y) + i\alpha(x,y) + j\beta(x,y) + k\gamma(x,y), \\
\psi^{(h)}(x,y) &= \phi(x)\psi(y) \\
&= \xi^{(h)}(x,y) + i\zeta^{(h)}(x,y) + j\gamma^{(h)}(x,y) + k\omega^{(h)}(x,y), \qquad (52) \\
\psi^{(v)}(x,y) &= \psi(x)\phi(y) \\
&= \xi^{(v)}(x,y) + i\zeta^{(v)}(x,y) + j\gamma^{(h)}(x,y) + k\omega^{(h)}(x,y), \\
\psi^{(d)}(x,y) &= \psi(x)\psi(y) \\
&= \xi^{(d)}(x,y) + i\zeta^{(d)}(x,y) + j\gamma^{(h)}(x,y) + k\omega^{(h)}(x,y),
\end{aligned}$$

where symbols $h, v, d$ denote the horizontal, vertical, and diagonal directions, respectively.

#### *3.1.2.1 Scale-Mixing 2-D Non-decimated Quaternion Wavelet Transform*

The various versions of the 2-D WT with appropriate tessellations of the detail spaces have been considered in 2-D wavelet literature. Here we focus on the scale-mixing 2-D wavelet transform because of its remarkable flexibility, compressibility,

and ease of computation (Ramírez and Vidakovic, 2013). From the scaling and wavelet functions in Equations (52) the wavelet atoms of the scale-mixing 2-D NDQWT can be represented as

$$\begin{aligned}
\phi_{J_{01},J_{02},k_1,k_2}(x,y) &= \mu_{J_{01},k_1,k_2}(x,y) \; + \\
&\quad i \cdot \alpha_{J_{02},k_1,k_2}(x,y) \; + \; j \cdot \beta_{J_{02},k_1,k_2}(x,y) \; + \; k \cdot \gamma_{J_{02},k_1,k_2}(x,y), \\
\psi^{(h)}_{J_{01},j_2,k_1,k_2}(x,y) &= \xi^{(h)}_{J_{01},k_1,k_2}(x,y) \; + \\
&\quad i \cdot \zeta^{(h)}_{j_2,k_1,k_2}(x,y) \; + \; j \cdot \gamma^{(h)}_{j_2,k_1,k_2}(x,y) \; + \; k \cdot \omega^{(h)}_{j_2,k_1,k_2}(x,y), \\
\psi^{(v)}_{j_1,J_{02},k_1,k_2}(x,y) &= \xi^{(v)}_{j_1,k_1,k_2}(x,y) \; + \\
&\quad i \cdot \zeta^{(v)}_{J_{02},k_1,k_2}(x,y) \; + \; j \cdot \gamma^{(v)}_{j_2,k_1,k_2}(x,y) \; + \; k \cdot \omega^{(v)}_{j_2,k_1,k_2}(x,y), \\
\psi^{(d)}_{j_1,j_2,k_1,k_2}(x,y) &= \xi^{(d)}_{j_1,k_1,k_2}(x,y) \; + \\
&\quad i \cdot \zeta^{(d)}_{j_2,k_1,k_2}(x,y) \; + \; j \cdot \gamma^{(d)}_{j_2,k_1,k_2}(x,y) \; + \; k \cdot \omega^{(d)}_{j_2,k_1,k_2}(x,y),
\end{aligned} \tag{53}$$

where $k_1 = 0, \ldots, m-1$, $k_2 = 0, \ldots, n-1$, $j_1 = J_{01}, \ldots, J-1$, $j_2 = J_{02}, \ldots, J-1$, and $J = \lceil \log_2 \min(m,n) \rceil$. Notice that $J_{01}$ and $J_{02}$ indicate the coarsest decomposition levels of rows and columns, respectively. Using these definitions, we can express any function $f \in L_2(\mathbb{R}^2)$ via wavelet decomposition as

$$\begin{aligned}
f(x,y) &= \sum_{k_1}\sum_{k_2} c_{J_{01},J_{02},k_1,k_2}\phi_{J_{01},J_{02},k_1,k_2}(x,y) \\
&\quad + \sum_{j_2>J_{02}}\sum_{k_1}\sum_{k_2} d^{(h)}_{J_{01},j_2,k_1,k_2}\psi^{(h)}_{J_{01},j_2,k_1,k_2}(x,y) \\
&\quad + \sum_{j_1>J_{01}}\sum_{k_1}\sum_{k_2} d^{(v)}_{j_1,J_{02},k_1,k_2}\psi^{(v)}_{j_1,J_{02},k_1,k_2}(x,y) \\
&\quad + \sum_{j_1>J_{02}}\sum_{j_2>J_{01}}\sum_{k_1}\sum_{k_2} d^{(d)}_{j_1,j_2,k_1,k_2}\psi^{(d)}_{j_1,j_2,k_1,k_2}(x,y).
\end{aligned}$$

This defines a scale-mixing 2-D NDQWT. Unlike the standard 2-D NDQWT using a single scale denoted by $j$, here we denote by pair $(j_1, j_2)$ a mixture of two scales. The coefficients corresponding to these scale-mixed atoms in the decomposition capture the local "energy flux" between scales $j_1$ and $j_2$.

Finally, we can obtain the scale-mixing non-decimated quaternion wavelet coefficients as

$$
\begin{aligned}
c_{J_{01},J_{02},k_1,k_2} &= \iint f(x,y)\overline{\phi}_{J_{01},J_{02},k_1,k_2}(x,y)\,dxdy \\
&= \mathrm{Re}(c_{J_{01},J_{02},k_1,k_2}) + i\cdot \mathrm{Im}^i(c_{J_{01},J_{02},k_1,k_2}) \; + \\
&\quad j\cdot \mathrm{Im}^j(c_{J_{01},J_{02},k_1,k_2}) + k\cdot \mathrm{Im}^k(c_{J_{01},J_{02},k_1,k_2}) \\
d^{(h)}_{J_{01},j_2,k_1,k_2} &= \iint f(x,y)\overline{\psi}^{(h)}_{J_{01},j_2,k_1,k_2}(x,y)\,dxdy \\
&= \mathrm{Re}(d^{(h)}_{J_{01},j_2,k_1,k_2}) + i\cdot \mathrm{Im}^i(d^{(h)}_{J_{01},j_2,k_1,k_2}) \; + \\
&\quad j\cdot \mathrm{Im}^j(d^{(h)}_{J_{01},j_2,k_1,k_2}) + k\cdot \mathrm{Im}^k(d^{(h)}_{J_{01},j_2,k_1,k_2}) \\
d^{(v)}_{j_1,J_{02},k_1,k_2} &= \iint f(x,y)\overline{\psi}^{(v)}_{j_1,J_{02},k_1,k_2}(x,y)\,dxdy \\
&= \mathrm{Re}(d^{(v)}_{j_1,J_{02},k_1,k_2}) + i\cdot \mathrm{Im}^i(d^{(v)}_{j_1,J_{02},k_1,k_2}) \; + \\
&\quad j\cdot \mathrm{Im}^j(d^{(v)}_{j_1,J_{02},k_1,k_2}) + k\cdot \mathrm{Im}^k(d^{(v)}_{j_1,J_{02},k_1,k_2}) \\
d^{(d)}_{j_1,j_2,k_1,k_2} &= \iint f(x,y)\overline{\psi}^{(d)}_{j_1,j_2,k_1,k_2}(x,y)\,dxdy \\
&= \mathrm{Re}(d^{(d)}_{j_1,j_2,k_1,k_2}) + i\cdot \mathrm{Im}^i(d^{(d)}_{j_1,j_2,k_1,k_2}) \; + \\
&\quad j\cdot \mathrm{Im}^j(d^{(d)}_{j_1,j_2,k_1,k_2}) + k\cdot \mathrm{Im}^k(d^{(d)}_{j_1,j_2,k_1,k_2}),
\end{aligned}
\tag{54}
$$

where $\overline{\phi}$ denotes the quaternion conjugate of $\phi$ defined in Equation (21). As we can see, the non-decimated quaternion wavelet coefficients in Equation (55) contain one real and three imaginary parts as quaternion numbers.

Matrix formulation can be used to perform the scale-mixing 2-D NDQWT for images of any size without a preprocessing work. Transformation of a 2-D image $\boldsymbol{A}$ of size $m \times n$ into a non-decimated quaternion wavelet transformed matrix $\boldsymbol{B}$ with depth $p_1$ and $p_2$ is implemented as

$$\boldsymbol{B} = W_m^{(p_1)} \cdot \boldsymbol{A} \cdot (W_n^{(p_2)})^\dagger$$

where $p_1, p_2, m$, and $n$ are arbitrary. Also, the $W_m^{(p_1)}$ and $W_n^{(p_2)}$ with $p_1$, $p_2$ detail levels and $m$, $n$ size of input data, respectively, are constructed from the quaternion scaling

and wavelet filters in Equation (52). Then, the resulting transformed matrix $\boldsymbol{B}$ has a size of $(p_1+1)m \times (p_2+1)n$ and represents a finite-dimensional implementation of the Equation (55) for $f(x)$ sampled in a matrix form. To correctly reconstruct the original image $\boldsymbol{A}$ of size $m \times n$, we need two weight matrices $T_m^{(p_1)}$ and $T_n^{(p_2)}$ with $p_1$- and $p_2$-level weight matrices, which are equally obtained as Equation (51) with different $m, n, p_1, p_2$. Then the reconstruction can be implemented using the weight matrices as

$$\boldsymbol{A} = W_m^{(p_1)} \cdot T_m^{(p_1)} \cdot \boldsymbol{B} \cdot T_n^{(p_2)} \cdot (W_n^{(p_2)})^{\dagger}.$$

More rigorous details on these matrix formulation for real-valued wavelets can be found in Kang and Vidakovic (2016).

Graphical illustrations of matrix-based scale-mixing 2-D NDQWT for a lena image is displayed in Figure 20.

## *3.2 Non-decimated Quaternion Wavelet Spectra*

Wavelet-based spectra is an efficient tool to estimate Hurst exponent in analyzing self-similar processes, such as fractional Brownian motion. Any hierarchy of multiresolution spaces can lead to definition of spectra. Especially important is that the multiscale analysis is generated by orthogonal filters because of energy preservation and resulting unbiased spectra. The literature on different approaches to defining a spectra based on wavelets is vast.

The previous Chapter 2 suggested the non-decimated complex wavelet spectra and demonstrated that consideration of redundancy and phase information were beneficial in the tasks of signal and image classification. Here, we extend the complex-valued method into the quaternion-valued wavelet spectra retaining the non-decimation and in 2-D case scale-mixing decomposition. First, we will explain this method for 1-D case and then expand its to 2-D case, by considering the scale-mixing 2-D transforms.

A real-valued stochastic process $\{X(t), t \in \mathbb{R}\}$ is said to be self-similar with the

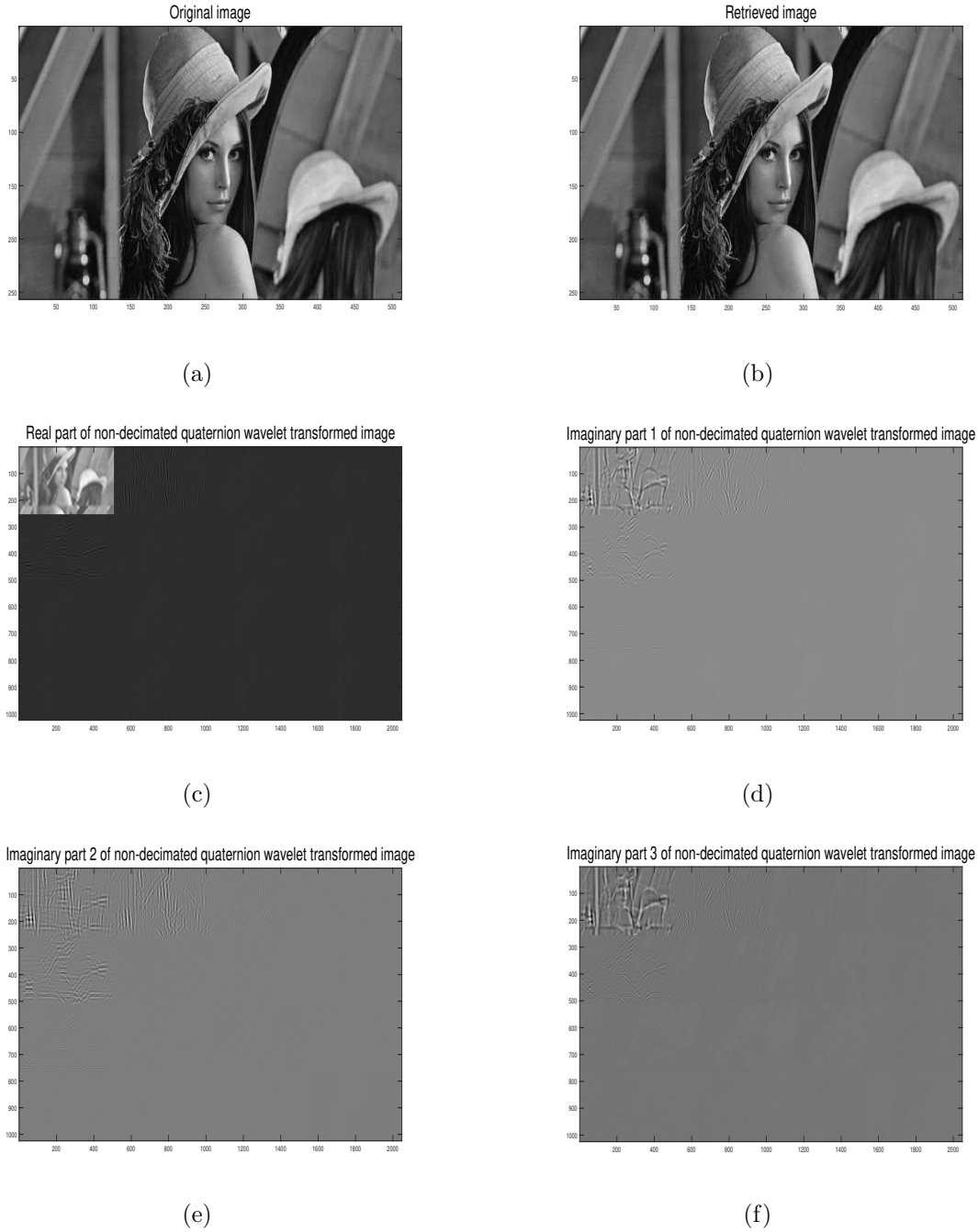



**Figure 20:** An example of matrix-based scale-mixing 2-D NDQWT for a lena image of size $256 \times 512$. (a) Original image, (b) Recovered image, (c) Real part of non-decimated quaternion wavelet transformed image, (d) Imaginary part 1 of non-decimated quaternion wavelet transformed image. (e) Imaginary part 2 of non-decimated quaternion wavelet transformed image. (f) Imaginary part 3 of non-decimated quaternion wavelet transformed image.

Hurst exponent $H$ if

$$X(\lambda t) \stackrel{d}{=} \lambda^H X(t) \text{ for any } \lambda \in \mathbb{R}, \tag{55}$$

where $\stackrel{d}{=}$ indicates equality in all joint finite-dimensional distributions. Given Equation (55), the wavelet coefficient, $d_{j,k}$, can be represented as

$$d_{j,k} \stackrel{d}{=} 2^{-j(H+\frac{1}{2})} d_{0,k} \tag{56}$$

under $L_2$ normalization in a real-valued wavelet transform at fixed dyadic scale $j$. In the NDQWT, we need to use a modulus $|d_{j,k}|$ instead of $d_{j,k}$. The $|d_{j,k}|$ is defined as

$$|d_{j,k}| = \sqrt{Re(d_{j,k})^2 + \mathrm{Im}^i(d_{j,k})^2 + \mathrm{Im}^j(d_{j,k})^2 + \mathrm{Im}^k(d_{j,k})^2}, \;\; j = J_0, \ldots, J-1.$$

Then, we can re-state Equation (56)as

$$|d_{j,k}| \stackrel{d}{=} 2^{-j(H+\frac{1}{2})} |d_{0,k}|, \;\; j = J_0, \ldots, J-1.$$

where the numbers of $k$ are all same for each $j$ because of the non-decimation property of NDQWT. Here the notation $\stackrel{d}{=}$ means the equality in all finite-dimensional distributions. When $X(t)$ shows a stationary increment, $E(|d_{0,k}|) = 0$ and $E(|d_{0,k}|^q) = E(|d_{0,0}|^q)$. This leads to

$$E(|d_{j,k}|^q) = C 2^{-jq(H+\frac{1}{2})}, \;\; j = J_0, \ldots, J-1 \tag{57}$$

where $C = E(|d_{0,0}|^q)$. Power $q$ is usually 2, corresponding to "power spectrum, or energy spectrum" and this would be used in this chapter. By taking logarithms on both sides of the Equation (57) we obtain a basis for wavelet-based estimation of $H$, as

$$S(j) = \log_2(E(|d_{j,k}|^2)) = -j(2H+1) + C', \;\; j = J_0, \ldots, J-1.$$

With all considered scaling levels as $j \in \mathbb{Z}$, a set of $S(j)$ represents a wavelet-based spectra. It describes a transition of energies along the scales. If a signal has a regular scaling, the energies would regularly decay, the plot of log-energy against the log-scale

is a straight line. The rate of energy decay, that is, the slope of the regression line, measures self-similarity of a given signal.

Operationally, we empirically estimate the wavelet-based spectrum, $S(j)$, as

$$\hat{S}(j) = \log_2 \frac{1}{m} \sum_{k=1}^{m} |d_{j,k}|^2 = \log_2 \overline{|d_{j,k}|^2}, \quad j = J_0, \ldots, J-1$$

where $m$ is the number of given data. Then, we can plot a 2nd order Logscale Diagram (2-LD) that is a set of $\hat{S}(j)$ against $j$ as $\left(j,\ \hat{S}(j)\right)$ as displayed in Figure 21. Finally, we can measure the slope of energy decay by regression methodology (an ordinary, weighted, or robust regression) and calculate the Hurst exponent $H$ based on the slope as $H = -(\text{slope} + 1)/2$. More rigorous proof and explanation of wavelet-based spectra and its applications can be found in Veitch and Abry (1999), Mallat (2009), and Ramírez and Vidakovic (2013).

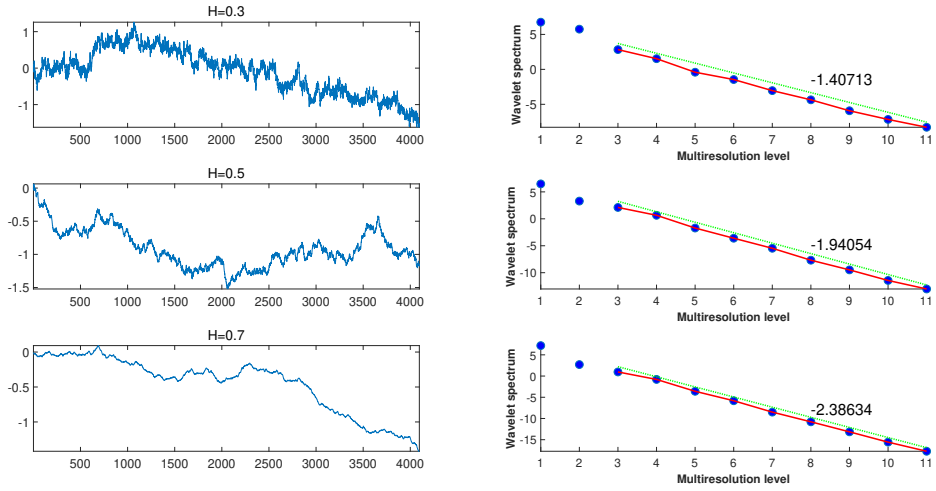



**Figure 21:** Examples of non-decimated quaternion wavelet spectra using the modulus of coefficients. The slopes are -1.40713, -1.94054, and -2.38634 corresponding to estimator $\hat{H} = 0.2035, 0.4703,$ and $0.6932$. The original 4096-length signals were simulated as a fBm with Hurst exponent 0.3, 0.5, and 0.7.

### 3.2.1 Scale-Mixing 2-D Non-decimated Quaternion Wavelet Spectra

A 2-D fractional Brownian motion (fBm) in two dimensions, $B_H(\mathbf{u})$ for $\mathbf{u} \in [0,1] \times [0,1]$ and $H \in (0,1)$, will be used as a model to explain a scale-mixing 2-D non-decimated complex wavelet spectra. The 2-D fBm, $B_H(\mathbf{u})$, is a self-similar process

$$B_H(a\mathbf{t}) \stackrel{d}{=} a^H B_H(\mathbf{t}) \text{ for any } a \in \mathbb{R},$$

with stationary zero-mean Gaussian increments. When 2-D fBm is decomposed by a scale-mixing non-decimated quaternion transform, the wavelet detail coefficients are

$$d_{(j_1,j_2+s,k1,k2)} = 2^{\frac{1}{2}(j_1+j_2+s)} \int B_H(\mathbf{u}) \overline{\psi} \left(2^{j_1}(u_1-k_1), 2^{j_2+s}(u_2-k_2)\right) d\mathbf{u},$$

where $\overline{\psi}$ is the quaternion conjugate of $\psi^{(d)}$ defined in Equation (53). We will focus on the main diagonal hierarchy where the 2-D scale indices coincide, we will use notation $j = j_1 = j_2$ and $J_0 = J_{01} = J_{02}$ in the sequel.

As in the 1-D case, we need to consider a modulus, $|d_{(j,j+s,k1,k2)}|$, instead of quaternion-valued $d_{(j,j+s,k1,k2)}$ defined as:

$$|d_{(j,j+s,k_1,k_2)}| =$$
$$\sqrt{Re(d_{(j,j+s,k_1,k_2)})^2 + Im^i(d_{(j,j+s,k_1,k_2)})^2 + Im^j(d_{(j,j+s,k_1,k_2)})^2 + Im^k(d_{(j,j+s,k_1,k_2)})^2},$$
$$j = J_0, \ldots, J-1.$$

Next, we can calculate an average of squared modulus of the coefficients as

$$\mathbb{E}\left[|d_{(j,j+s,k_1,k_2)}|^2\right] = 2^{2j+s} \int \psi\left(2^j(u_1-k_1), 2^{j+s}(u_2-k_2)\right)$$
$$\times \overline{\psi}\left(2^j(v_1-k_1), 2^{j+s}(v_2-k_2)\right) \mathbb{E}\left[B_H(\mathbf{u}) B_H(\mathbf{v})\right] d\mathbf{u}\, d\mathbf{v},$$

which be expressed as

$$\mathbb{E}\left[|d_{(j,j+s,k_1,k_2)}|^2\right] = 2^{-j(2H+2)}\, V_{\psi,s}(H). \tag{58}$$

This was proven in Jeon et al (2014) for complex wavelets and in Kang and Vidakovic (2016) for non-decimated wavelets. Here, the $V_{\psi,s}(H)$ can be treated as constant with

respect to scale $j$ but it depends on $\psi$, $H$ and $s$. Finally, we can obtain the scale-mixing 2-D non-decimated quaternion wavelet spectrum by taking logarithms on both sides of the Equation (58) as following:

$$S(j, j+s) = \log_2(\mathbb{E}(|d_{j,j+s,k_1,k_2}|^2)) = -j(2H+2) + C', \quad j = J_0, \ldots, J-1.$$

The empirical counterpart of $S(j, j+s)$ is

$$\hat{S}(j, j+s) = \log_2 \frac{1}{mn} \sum_{k_1=1}^{m} \sum_{k_2=1}^{n} |d_{j,j+s,k_1,k_2}|^2 = \log_2 \overline{|d_{j,j+s,k_1,k_2}|^2}, \quad j = J_0, \ldots, J-1$$

where $m$ is a row length and $n$ is a column length. To estimate Hurst exponent $H$ we use the spectral slope as in the 1-D case except that in 2-D case $\hat{H} = -(\text{slope} + 2)/2$ instead of $\hat{H} = -(\text{slope} + 1)/2$.

## *3.3 Phase-based Statistics for Classification Analysis*

Importance of properly utilizing phase information that is not available for the real-valued wavelets was exemplified in Jeon et al (2014) and Chapter 2 for the complex-valued wavelets. Although the spectra based on the phase information cannot be used to estimate the Hurst exponent, here we suggest the use of phase-based modalities to improve performance in classification tasks. Given the three phases in quaternion decompositions, we expect that the discriminatory power of summaries that include phase modalities would significantly increase.

First, we need to calculate three phases of non-decimated quaternion wavelet coefficient defined in Equation (50). For 1-D case, substituting the four coefficients, $\text{Re}(d_{j,k}), \text{Im}^i(d_{j,k}), \text{Im}^j(d_{j,k})$, and $\text{Im}^k(d_{j,k})$ for the $q_0, q_1, q_2$, and $q_3$, we obtain the three phases $\phi_{d_{j,k}}, \theta_{d_{j,k}}$, and $\psi_{d_{j,k}}$ as explained in Section 1.6.1. It is the same for 2-D case after replacing the $d_{j,k}$ with $d_{(j,j+s,k1,k2)}$.

Then, the three phase averages at level $j$ can be obtained as

$$\phi_j = \frac{1}{m}\sum_{k=1}^{m}\phi_{d_{j,k}}, \quad \theta_j = \frac{1}{m}\sum_{k=1}^{m}\theta_{d_{j,k}}, \quad \psi_j = \frac{1}{m}\sum_{k=1}^{m}\psi_{d_{j,k}}, \tag{59}$$

$$\phi_j = \frac{1}{mn}\sum_{k_1=1}^{m}\sum_{k_2=1}^{n}\phi_{d_{(j,j+s,k1,k2)}}, \quad \theta_j = \frac{1}{mn}\sum_{k_1=1}^{m}\sum_{k_2=1}^{n}\theta_{d_{(j,j+s,k1,k2)}}, \quad \psi_j = \frac{1}{mn}\sum_{k_1=1}^{m}\sum_{k_2=1}^{n}\psi_{d_{(j,j+s,k1,k2)}},$$

$$j = J_0, \ldots, J-1$$

for 1-D and 2-D cases, separately.

While the phases do not indicate any scaling regularity as explained at the beginning of this section and as displayed in Figure 22, the three phase averages defined in Equation (59) would improve a power of classification if used with the wavelet-based spectra described in section 3.2. In Chapter 5, we demonstrate this and show that the new modalities have surpassed the traditional wavelet-based spectra method which is based on the modulus of wavelet coefficients.

## *3.4 Applications*

To illustrate the proposed methodology, we consider four applications in tasks of supervised learning. Since the proposed methods in Chapter 2, 3, and 4 share the same datasets, we set aside all analysis results in Chapter 5 (APPLICATIONS). In particular, the detail explanations in terms of NDQWT for all four applications can be found in section 5.1.3.2, 5.2.3.2, 5.3.3.2, and 5.4.3.2.

## *3.5 Conclusions and Future Studies*

In this chapter, we suggested a non-decimated quaternion wavelet transform (NDQWT) for both 1-D and 2-D cases. We demonstrated that the proposed wavelet spectra works well in classification problems with standard spectrum based on the magnitudes is enhanced by the three quaternionic phase-based statistics. Through comparative investigation in four real-life applications, we found that the classification

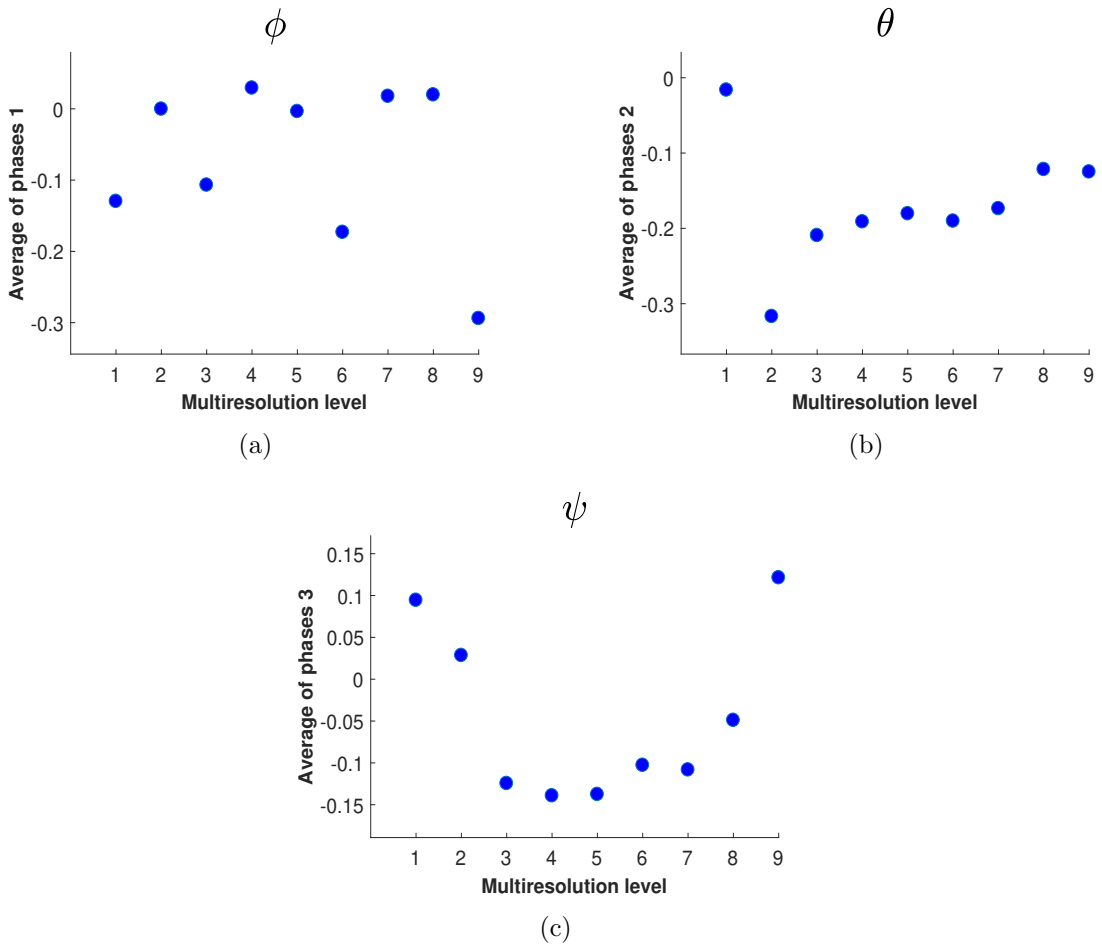



**Figure 22:** Visualization of three phase averages $(\phi, \theta, \psi)$ at all multiresolution levels.

procedure by NDQWT outperforms the QWT and NDWT. This indicated that combination of two different redundancies, structural one from NDWT and componential one from QWT benefited the performance. The NDQWT can appeal to researchers seeking more efficient wavelet-based classification methodology for signals or images with intrinsic self-similarity.

There are several directions for possible future research. The performance could be robustified if we calculate the spectral slopes in different ways as done in Hamilton et al (2011) or Feng et al (2018) where robust Theil-type regressions and trimean estimators have been proposed. For the scale-mixing 2-D NDQWT, the diagonal hierarchy of coefficients $d^{(d)}$ itself is provided good discriminatory descriptors such

as spectral slopes and phase-based statistics. By using scale mixing hierarchies in addition to $d^{(d)}$ may likely further improve the performance of classification. Finally, performing a scale-mixing 2-D NDQWT with different wavelet filters for rows and columns of pixels enables more modeling freedom. For instance, the left-hand side can be a matrix based on the quaternion-valued filter while the right-hand side can be based on real-valued filters such as Haar, Symmlet, Coiflet, and so on.

In the spirit of reproducible research, we prepared an illustrative demo as a stand alone MATLAB software with solved examples. The demo is posted on the repository Jacket Wavelets `http://gtwavelet.bme.gatech.edu/`.

# CHAPTER IV

# DUAL WAVELET SPECTRAL TOOLS

Theoretical self-similar processes such as fractional Brownian motion and processes with 1/f power-law spectra have been an essential tool for modeling a wide range of real-world signals or images that describe scaling phenomena in engineering, physics, medicine, biology, economics, geology, chemistry, and so on. However, it is often difficult for quantify a self-similarity in the signal acquisition domain due to an inherent irregularity. To solve this problem, a number of researches have focused on the modeling in the space/scale domains and focusing on the behavior of "energies". One can define the energy in the frequency/scale domain as the variance of a zero-mean signal in the original domain. This total energy can be partitioned in an ANOVA fashion to energies contained in the coefficients in a multiresolution representation of a signal, such as wavelet transform, for example. Then, the total energy is the sum of the squared coefficients (component energies) in the frequency/scale domain if the transformation is orthogonal.

Although the behavior of a signal in original domain can be chaotic, there can be regular behavior in frequency/scale domain in the terms of energies that scale predictably and we call this phenomenon a regular scaling. If the regular scaling exists in signals or images, it means that the scales or resolutions in multiresolution decomposition have average energies that behave linearly on the log-scale. This behavior is typical for a monofractal and has been widely used to quantify the degree of irregularity of signals or images.

The standard measure of self-similarity in terms of regular scaling is the Hurst exponent estimated from scaling analysis. This is an informative summary for the

behavior of self-similar process and also related to the presence of long memory and monofractality in signals and images. Of many methods for estimating the Hurst exponent, the method based on wavelets is an efficient and powerful technique. In the past, the researches mostly focused on the study of scaling in one-dimensional objects. These days, there is increased interest for exploring scaling in multidimensional objects most importantly, images.

Dual relation as an alternative representation to analyze the same problem has been used in various field including optimization, physics, engineering, and mathematics, etc. In particular, dual relation in multifractal spectra has been explored for it can provide a complementary statistical summary of singular fields (Roux and Jensen, 2004). In traditional multifractal analysis, the multifractal formalism has been proven as an excellent dimensional analysis method to solve a multitude of singular field problems. This is done by characterizing the scaling properties of a physical density $\rho$ as a function of the distance $\mathcal{L}$ (Benzi et al, 1985; Halsey et al, 1987). The singular field can be decomposed into a continuum of fractal supports depending on its singularity index, $\alpha$. Then, the multifractal spectrum can be constructed as a fractal dimension $f(\alpha)$ as a function of the singularity, $\alpha$. An alternative dual representation to analyze the same fields was proposed in Jensen (1999). While the traditional approach considers the statistical distribution of mass $m(\mathcal{L})$ over a fixed distance $\mathcal{L}$, Jensen and Roux reversely studied the distribution of distances $\mathcal{L}(m)$ associated to a fixed mass, $m$, contained in each subset. They suggest to use new scaling from the dual approach to characterize property of singular field. This alternative method was later more extensively researched; Roux and Jensen (2004) established dual relation with the traditional method. The authors applied the dual method to a multifractal cantor set as well as the turbulent GOY shell model, extending theoretical results to a realistic physical problem. It was shown that the two spectra are related through simple duality relations and that the moment scaling exponents are related to the

usual series of scaling exponents.

Even if this kind of duality concept has been studied in various fields (e.g. physics, atmosphetic sciences, etc.), we are nor aware of the definition of wavelet spectra dual to a standard wavelet second-order spectra, as far as we know. Thus, we suggest a wavelet spectral tools dual to the original wavelet spectra to characterize the self-similarity by novel scaling indices. In this context, we will call the original wavelet spectra the primal spectra.

As a dual representation of primal wavelet spectra, distributions of levels (scales) with regard to quantized energy (squared wavelet coefficients within the interval) are established as opposed to distributions of energies along the scale levels as in the primal wavelet spectra. Quantization, as a standard operation in digital signal processing, maps the squared energies from a continuous nonnegative set to a finite discrete set. The elements of this set are centers of intervals that partition the whole set of energies and the boundaries are determined by sample percentiles to ensure that the partition intervals are nonempty.

Orthogonal versions of wavelet transform produce the same number of wavelet coefficients as is the length of 1-D signal (or size of 2-D image), which is inadequate to build a robust and reliable dual wavelet spectra. Simply the number of wavelet coefficients for quantizing their energies is not sufficiently large for coarse levels to assure stability and robustness. Instead of minimal, that is orthogonal, transforms we employ a NDWT. The NDWT is time-invariant unlike traditional orthogonal transforms and provides a dense approximation to continuous wavelet transform. This invariance follows from non-decimation property that preserves location information across the scales. The term "non-decimated" is linked to implementational aspect of NDWT, in which the Mallat algorithm does not employ decimation in the signal/image filtering process, see Mallat (2009). With non-decimation property providing more dense approximation to continuous wavelet transform, we expect that dual wavelet spectra

can represent more various and elaborated information compared to primal wavelet spectra. Specifically, we can manually set the number of points on x-axis in dual spectra very dense while points on x-axis in primal spectra are the multiresolution levels that are fixed and a few. Thus, we can supposed that information from which we can derive can be different between the primal and dual spectra.

To generalize both primal and dual spectra, filters in Mallat algorithm can be complex-valued or quaternion-valued. Structural redundancy from the NDWT and componential redundancy from a complex or quaternion wavelet filters have demonstrated their usefulness and synergy effects when they are considered together as demonstrated in Chapter 2 and 3. From the dual wavelet spectra based on the non-decimated wavelet transform, we expect an additional type of redundancy, and call it dual redundancy. In addition, complex-values and quaternion-valued domains provide dual representation of phase-based statistics from $\text{NDWT}_{\text{C}}$ and NDQWT proposed in Chapter 2 and 3. Finally, we compare the proposed dual methodology with their primal versions and demonstrate that the dual representations can increase classification performance when considered jointly.

This chapter is organized as follows. Section 4.1 briefly describes a way of constructing a primal wavelet spectra with NDWT. Section 4.2 and 4.3 illustrates a dual wavelet spectra with the NDWT for 1-D and 2-D cases and extends them into complex and quaternion domains with phase information leading to dual phase-based statistics. Section 4.4 demonstrates a power of the proposed method with 1-D and 2-D applications and Section 4.5 contains some concluding remarks and directions for future study.

## *4.1 Primal Wavelet Spectra with Non-decimated Wavelet Transform*

In this section, we briefly describe a way of constructing traditional wavelet spectra with NDWT representing relationships between scale $j$ and energies at the scales.

First, an energy is defined as average of squared wavelet coefficients at each level. If energies decay regularly, it indicates that there is a regular scaling in the data, thus, we can measure a self-similarity via a rate of energy decay. Operationally, we find the slope in regression of log energies to scale indices and use it to estimate the Hurst exponent. For discrete observed data of size $m$, we use empirical wavelet spectrum defined as

$$\hat{S}(j) = \log_2 \frac{1}{m} \sum_{k=1}^{m} d_{j,k}^2 = \log_2 \overline{d_{j,k}^2}, \quad j = J_0, \ldots, J-1.$$

We can plot the set of $\hat{S}(j)$ against $j$ as $\big(j,\ \hat{S}(j)\big)$, which is called 2nd order Logscale Diagram (2-LD) and this is the wavelet spectra as displayed in Figure 23. Finally, we can estimate the slope of the spectra usually by regression methodology (an ordinary, weighted, or robust regression) and use it to estimate the Hurst exponent $H$, as $\hat{H} = -(\text{slope} + 1)/2$. More details on wavelet spectra method and its applications can be found in Veitch and Abry (1999), Mallat (2009), Ramírez and Vidakovic (2013), and Roberts et al (2017).

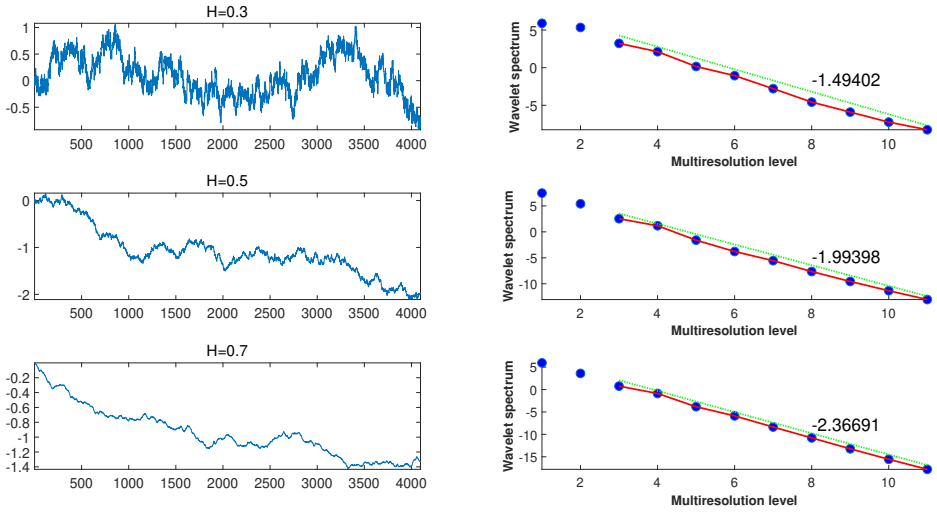



**Figure 23:** Examples of non-decimated wavelet spectra. The slopes are -1.49402, -1.99398, and -2.36691 corresponding to estimator $\hat{H} = 0.2470, 0.4970,$ and 0.6835. The original 4096-length signals were simulated as a fBm with Hurst exponent 0.3, 0.5, and 0.7.

## 4.2 1-D Dual Wavelet Spectra with Non-decimated Wavelet Transform

First, we transform a signal from time domain to the wavelet domain using the NDWT to obtain a set of non-decimated wavelet coefficients as in Section (1.7) with $J = \lceil \log_2 m \rceil$. Then, we define $\log_2 \mathrm{EN}$ that is a set of logarithm (base 2) of squared wavelet coefficients for all levels as following:

$$\log_2 \mathrm{EN} = \{\log_2(d_{j,k})^2 \ \mid \ j = J_0, \ldots, J-1, \ \ k = 0, \ldots, m-1\}.$$

To perform a role of the levels in primal wavelet spectra we need predefined intervals for $\log_2 \mathrm{EN}$ in which almost same number of wavelet coefficients exist. This is necessary to make distributions of levels at each interval consistent and robust. The number of intervals for $\log_2 \mathrm{EN}$ is denoted as $L$ and it can be manually selected depending on a size of data.

First, we define $p_i$ as

$$p_i = \frac{i}{L}, \ \ i = 0, \ldots, L.$$

and then obtain $L+1$ quantiles of $\log_2 \mathrm{EN}$ that are boundaries of $L$ intervals as

$$Q_{p_i} \ = \ \begin{cases} \log_2 \mathrm{EN}_{(1)} & \text{if } \ p_i < \frac{1}{Jm}, \\ \log_2 \mathrm{EN}_{(\lfloor Jmp_i \rfloor)} + (Jmp_i - \lfloor Jmp_i \rfloor)(\log_2 \mathrm{EN}_{(\lfloor Jmp_i \rfloor + 1)} - \log_2 \mathrm{EN}_{(\lfloor Jmp_i \rfloor)}) & \text{if } \ p_i \geq \frac{1}{Jm}, \end{cases}$$

for $i = 0, \ldots, L$, through a linear interpolation of the empirical distribution function. Then, we can find subsets of level $j$ in all $L$ individual intervals as

$$\boldsymbol{j}_{p_i, p_{i+1}} \ = \ \{ \, j \ \mid \ Q_{p_i} \leq \log_2(d_{j,k})^2 < Q_{p_{i+1}}, \ \ j = J_0, \ldots, J-1, \ \ k = 0, \ldots, m-1\}$$

for $i = 0, \ldots, L-1$. Each subset, $\boldsymbol{j}_{p_i, p_{i+1}}$, has its own distribution of level $j$ as shown in Figure 24.

As shown in Section 4.1, general wavelet spectra examines distributions of energies with regard to each level while dual wavelet spectra looks into distributions of levels with regard to each quantile intervals of $\log_2 \mathrm{EN}$. Thus, first we calculate average

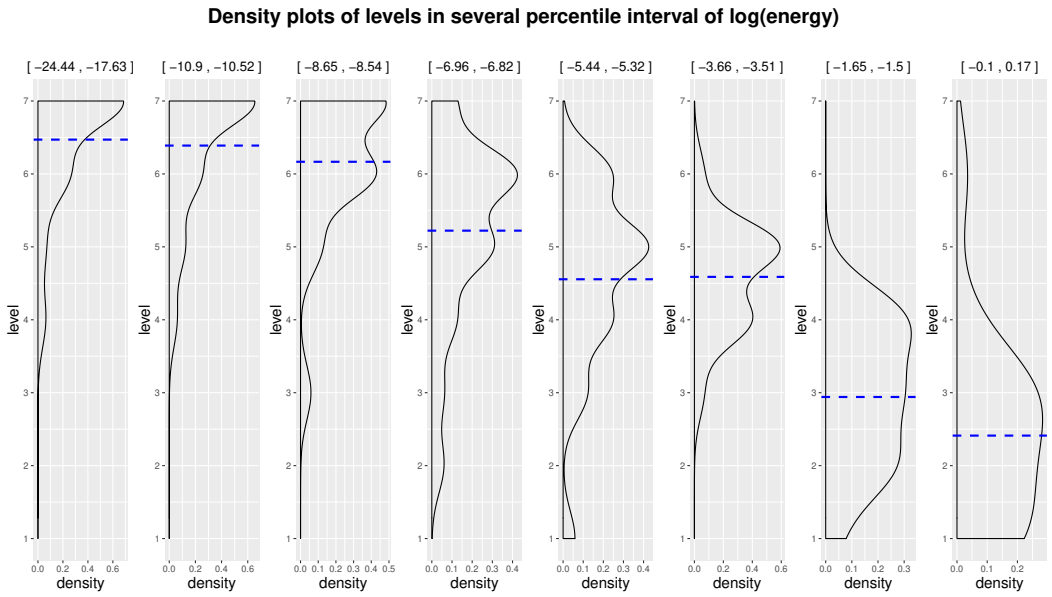



**Figure 24:** Estimated density plots of level $j$ at some selected intervals with dotted blue line for mean.

levels in each interval, the blue dotted lines in Figure 24, and center points of all intervals as

$$\begin{aligned} \overline{j}_{p_i,p_{i+1}} &= \frac{1}{n_i}\sum j, \quad \forall j \in \boldsymbol{j}_{p_i,p_{i+1}}, \\ x_{i,i+1} &= \frac{Q_{p_i} + Q_{p_{i+1}}}{2} \end{aligned}$$

for $i = 0, \ldots, L-1$, where $n_i$ is the number of coefficients in a set $\boldsymbol{j}_{p_i,p_{i+1}}$. Then, we can build the dual wavelet spectra where $x_{i,i+1}$ are on x-axis and $\overline{j}_{p_i,p_{i+1}}$ are on y-axis. Finally, we regress $\overline{j}_{p_i,p_{i+1}}$ with $x_{i,i+1}$ and find a slope to measure a self-similarity. Note that we only consider samples within 20% - 95% range of $\overline{j}_{p_i,p_{i+1}}$ to find a more robust dual spectral slope. In this dissertation the default of the $L$ is 100 for applications. Graphical examples using fractal Brownian motion (fBm) are as shown in Figure 25.

### 4.2.1 1-D Dual Wavelet Spectra with Non-decimated Complex Wavelet Transform

In the area of Fourier representations, a lot of researches as in Oppenheim and Li (1981) and Levi and Stark (1983) are interested in the phase information on signals

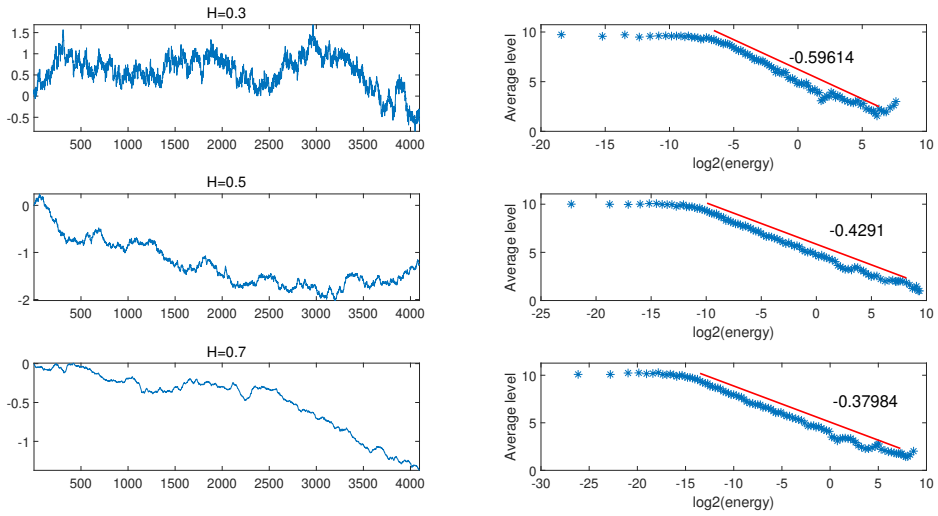



**Figure 25:** Examples of dual non-decimated wavelet spectra. The slopes are -0.59614, -0.42910, and -0.37984. The original 4096-length signals were simulated as a fBm with Hurst exponent 0.3, 0.5, and 0.7.

or images. Similarly, there is also an interest in complex wavelet domain about information related to interactions between decomposition levels and spatial symmetries contained in the phase, as investigated by Lina (1997), Lina (1999), and Jeon et al (2014). Therefore, it is natural to explore the role of phase in the complex-valued wavelet coefficients in Equation (31). The previous Chapter 2 described a way of constructing wavelet spectra with complex-valued wavelet coefficients in non-decimated fashion. Based on the non-decimated complex wavelet spectra it suggested its spectral slope to estimate the Hurst exponent and phase-based summaries as additional modalities in classification analysis. For the phase-based summaries, it did not obtain the corresponding spectral slope since the phase-based descriptors do not show a scaling property. Instead, it just utilized them as individual descriptors and demonstrated their effects with some classification applications. As a dual representation of this version, here we propose how to obtain dual spectra slope of modulus and dual phase-based summaries and use them as descriptors in classification analysis.

First, a modulus of the non-decimated complex wavelet coefficients is defined as

$$|d_{j,k}| = \sqrt{Re(d_{j,k})^2 + Im(d_{j,k})^2}, \;\; j = J_0, \ldots, J-1.$$

Then, we apply the same way of constructing dual wavelet spectra as described in Section 4.2 using $|d_{j,k}|$ instead of $d_{j,k}$.

Next, a phase of the non-decimated complex wavelet coefficients can be calculated as

$$\angle d_{j,k} \;\; = \;\; \arctan\left(\frac{Im(d_{j,k})}{Re(d_{j,k})}\right)$$

whose range is $[-\pi, \pi]$. In this case, we do not need such dense intervals in modulus case because we would extract a countable number of dual phase-based statistics instead of a spectral slope. Thus, we intuitionally divide the range of phase, $[-\pi, \pi]$, into 4 intervals with bounds, $Q_{p_i}$, defined as

$$Q_{p_i} = (p_i - 0.5)2\pi, \;\; i = 0, \ldots, 4$$

where $p_i = \frac{i}{4}$. Then, we can find subsets of level $j$ in the 4 intervals as

$$\boldsymbol{j}_{\angle d, p_i, p_{i+1}} = \{\, j \;\mid\; Q_{p_i} \leq \angle d_{j,k} < Q_{p_{i+1}}, \;\; j = J_0, \ldots, J-1, \;\; k = 0, \ldots, m-1\}$$

for $i = 0, \ldots, 3$ and calculate averages of levels in the 4 intervals as

$$\bar{j}_{\angle d, p_i, p_{i+1}} = \frac{1}{n_i} \sum j, \quad \forall j \in \boldsymbol{j}_{\angle d, p_i, p_{i+1}}$$

for $i = 0, \ldots, 3$ where $n_i$ is the number of data in a subset $\boldsymbol{j}_{\angle d, p_i, p_{i+1}}$. Finally, a dual phase-based statistics, $\bar{j}^2_{\angle d, p_i, p_{i+1}}$, are obtained as new descriptors since quadratic relations are empirically better than linear relations in our real applications.

### 4.2.2 1-D Dual Wavelet Spectra with Non-decimated Quaternion Wavelet Transform

The previous Chapter 3 extended complex wavelets in non-decimated fashion into a quaternion wavelet domain. Following emphasis on a role of phase in complex

wavelets as in Chapter 2 it supposed and demonstrated that the additional two phases can improve performances of the phase-based summaries in classification analysis. Therefore, here we suppose a similar improvement in dual representation and explore a dual wavelet spectra with non-decimated quaternion-valued wavelet coefficients in Equation (50) in terms of modulus and three phases.

Similar to the previous complex case, we can define a modulus of non-decimated quaternion wavelet coefficients as

$$|d_{j,k}| = \sqrt{Re(d_{j,k})^2 + \mathrm{Im}^i(d_{j,k})^2 + \mathrm{Im}^j(d_{j,k})^2 + \mathrm{Im}^k(d_{j,k})^2}, \;\; j = J_0, \ldots, J-1.$$

Instead of $d_{j,k}$, the $|d_{j,k}|$ can be used to build a dual wavelet spectra as explained in Section 4.2

For all three phase-based summaries, we do not find a dual spectral slope since the phase-based statistics do not show a regular decaying. Instead, we define three dual phase-based summaries as additional modalities in quaternion domain. First, we can obtain three phases of a quaternion-valued wavelet coefficient as

$$(\phi_{d_{j,k}}, \theta_{d_{j,k}}, \psi_{d_{j,k}}) \in [-\pi, \pi] \times [-\frac{\pi}{2}, \frac{\pi}{2}] \times [-\frac{\pi}{4}, \frac{\pi}{4}],$$

where $j = J_0, \ldots, J-1$ and $k = 0, \ldots, m-1$. Calculating procedures for the three phases can be found in Section 1.6.1. The dense intervals as in modulus case is not necessary here because we will not obtain a dual spectral slope. To divide the ranges of three phases, $[-\pi, \pi] \times [-\frac{\pi}{2}, \frac{\pi}{2}] \times [-\frac{\pi}{4}, \frac{\pi}{4}]$, into 4 intervals, we first define bounds, $Q_{p_i}$, as

$$Q_{p_i} = (p_i - 0.5)2\pi, \;\; i = 0, \ldots, 16$$

where $p_i = \frac{i}{16}$. With uses of $Q_{p_i}$ we can find subsets of level $j$ in the 4 intervals for all three phases as

$$\begin{aligned}
\boldsymbol{j}_{\phi,p_i,p_{i+4}} &= \{\, j \mid Q_{p_i} \le \phi_{d_{j,k}} < Q_{p_{i+4}}, \;\; j = J_0, \ldots, J-1, \;\; k = 0, \ldots, m-1\} \text{ for } i = 0, 4, 8, 12, \\
\boldsymbol{j}_{\theta,p_i,p_{i+2}} &= \{\, j \mid Q_{p_i} \le \theta_{d_{j,k}} < Q_{p_{i+2}}, \;\; j = J_0, \ldots, J-1, \;\; k = 0, \ldots, m-1\} \text{ for } i = 4, 6, 8, 10, \\
\boldsymbol{j}_{\psi,p_i,p_{i+1}} &= \{\, j \mid Q_{p_i} \le \psi_{d_{j,k}} < Q_{p_{i+1}}, \;\; j = J_0, \ldots, J-1, \;\; k = 0, \ldots, m-1\} \text{ for } i = 6, 7, 8, 9,
\end{aligned}$$

and calculate averages of levels in them as

$$\begin{aligned}
\bar{j}_{\phi,p_i,p_{i+4}} &= \frac{1}{n_i}\sum j, \quad \forall j \in \boldsymbol{j}_{\phi,p_i,p_{i+4}} \text{ for } i = 0,4,8,12,\\
\bar{j}_{\theta,p_i,p_{i+2}} &= \frac{1}{n_i}\sum j, \quad \forall j \in \boldsymbol{j}_{\theta,p_i,p_{i+2}} \text{ for } i = 4,6,8,10,\\
\bar{j}_{\psi,p_i,p_{i+1}} &= \frac{1}{n_i}\sum j, \quad \forall j \in \boldsymbol{j}_{\psi,p_i,p_{i+1}} \text{ for } i = 6,7,8,9,
\end{aligned}$$

where $n_i$ is the number of coefficients in a corresponding set of $\boldsymbol{j}_{\phi,p_i,p_{i+4}}, \boldsymbol{j}_{\theta,p_i,p_{i+2}}$, and $\boldsymbol{j}_{\psi,p_i,p_{i+1}}$. Finally, we can obtain three dual phase-based statistics, $\bar{j}^2_{\phi,p_i,p_{i+4}}, \bar{j}^2_{\theta,p_i,p_{i+2}}, \bar{j}^2_{\psi,p_i,p_{i+1}}$, as new descriptors since empirical experiments showed that quadratic relations are more reasonable than linear relations.

## *4.3 2-D Dual Wavelet Spectra with Non-decimated Wavelet Transform*

The 1-D dual wavelet spectra can be extended to its 2-D counterpart. Although various versions of the 2-D wavelet transform can be constructed by appropriate tessellations of the detail spaces, here we utilize the scale-mixing 2-D NDWT. As we argued before, the use of scale-mixing version is motivated by its remarkable flexibility, compressibility, and ease of computation. As a first step, we transform a 2-D image size of $m \times n$ from time domain to the wavelet domain using the scale-mixing 2-D NDWT to obtain a set of non-decimated wavelet coefficients in Equation (27) with $J = \lceil \log_2 \min(m,n) \rceil$. In this chapter, we only consider the main diagonal hierarchy whose 2-D scale indices coincide as $j_1 = j_2 = j$ and thus $J_{01} = J_{02} = J_0$. Next, a set of logarithm (base 2) of squared wavelet coefficients for all levels, $\log_2 \mathrm{EN}$, is defined as

$$\log_2 \mathrm{EN} = \{\log_2(d_{j,j+s,k_1,k_2})^2 \;\mid\; j = J_0,\ldots,J-1,\;\; k_1 = 0,\ldots,m-1,\;\; k_2 = 0,\ldots,n-1\}.$$

For this 2-D dual wavelet spectra, the $L$ intervals for $\log_2 \mathrm{EN}$ is also needed to have consistent distributions of levels at each interval. With use of $p_i$ defined as

$$p_i = \frac{i}{L}, \;\; i = 0,\ldots,L,$$

$L+1$ quantiles of $\log_2 \mathrm{EN}$ are obtained as

$$Q_{p_i} = \begin{cases} \log_2 \mathrm{EN}_{(1)} \quad \text{if } p_i < \frac{1}{Jmn}, \\ \log_2 \mathrm{EN}_{(\lfloor Jmnp_i \rfloor)} + (Jmnp_i - \lfloor Jmnp_i \rfloor)(\log_2 \mathrm{EN}_{(\lfloor Jmnp_i \rfloor+1)} - \log_2 \mathrm{EN}_{(\lfloor Jmnp_i \rfloor)}) \\ \quad \text{if } p_i \geq \dfrac{1}{Jmn} \end{cases}$$

for $i = 0, \ldots, L$, through a linear interpolation of the empirical distribution function. Utilizing the $Q_{p_i}$ as boundaries, we can find subsets of level $j$ in all $L$ individual intervals as

$$\begin{aligned} \boldsymbol{j}_{p_i,p_{i+1}} = \{\, j \mid\ & Q_{p_i} \leq \log_2(d_{j,j+s,k_1,k_2})^2 < Q_{p_{i+1}}, \\ & j = J_0, \ldots, J-1, \ \ k_1 = 0, \ldots, m-1, \ \ k_2 = 0, \ldots, n-1\} \end{aligned}$$

for $i = 0, \ldots, L-1$. Based on the set of $\boldsymbol{j}_{p_i,p_{i+1}}$, the ways of building 2-D dual wavelet spectra and finding a slope are the same with 1-D case explained in Section 4.2. In this dissertation we set the default of the $L$ as 100 for applications.

### 4.3.1 2-D Dual Wavelet Spectra with Non-decimated Complex Wavelet Transform

Using a complex filter, we can build a dual wavelet spectra for 2-D images with non-decimated complex-valued wavelet coefficients in Equation (36) in terms of modulus and phase.

First, we have a modulus of the non-decimated complex wavelet coefficients as

$$|d_{(j,j+s,k_1,k_2)}| = \sqrt{Re(d_{(j,j+s,k_1,k_2)})^2 + Im(d_{(j,j+s,k_1,k_2)})^2}, \ \ j = J_0, \ldots, J-1.$$

Then, dual wavelet spectra can be constructed equally as described in Section 4.3 using $|d_{(j,j+s,k_1,k_2)}|$ instead of $d_{(j,j+s,k_1,k_2)}$.

Next, we can calculate a phase of non-decimated complex wavelet coefficients as

$$\angle d_{(j,j+s,k_1,k_2)} = \arctan\left(\frac{Im(d_{(j,j+s,k_1,k_2)})}{Re(d_{(j,j+s,k_1,k_2)})}\right)$$

whose range is $[-\pi, \pi]$. Instead of dense intervals in modulus case, the range of phase, $[-\pi, \pi]$, is divided into 4 intervals with bounds, $Q_{p_i}$, defined as

$$Q_{p_i} = (p_i - 0.5)2\pi, \quad i = 0, \ldots, 4$$

where $p_i = \frac{i}{4}$. Then, 4 subsets of level $j$ in each interval can be obtained as

$$\begin{aligned} \boldsymbol{j}_{\angle d, p_i, p_{i+1}} \quad = \quad & \{\, j \mid Q_{p_i} \leq \angle d_{j,j+s,k_1,k_2} < Q_{p_{i+1}}, \\ & j = J_0, \ldots, J-1, \quad k_1 = 0, \ldots, m-1, \quad k_2 = 0, \ldots, n-1\} \end{aligned}$$

for $i = 0, \ldots, 3$ and calculate averages of levels in each subset as

$$\bar{j}_{\angle d, p_i, p_{i+1}} = \frac{1}{n_i} \sum j, \quad \forall j \in \boldsymbol{j}_{\angle d, p_i, p_{i+1}}$$

for $i = 0, \ldots, 3$ where $n_i$ is the number of data in a subset $\boldsymbol{j}_{\angle d, p_i, p_{i+1}}$. Because of the same reasons fore-mentioned in Section 4.2.1 we do not find a spectral slope and instead determine a set of dual phase-based statistics, $\bar{j}^2_{\angle d, p_i, p_{i+1}}$, as descriptors.

### 4.3.2 2-D Dual Wavelet Spectra with Non-decimated Quaternion Wavelet Transform

Using a quaternion filter, we can build a dual wavelet spectra for 2-D images with non-decimated quaternion wavelet coefficients in Equation (55) with regard to modulus and three phases.

First, a modulus of the non-decimated quaternion wavelet coefficients is defined as

$$|d_{(j,j+s,k1,k2)}| =$$

$$\sqrt{Re(d_{(j,j+s,k_1,k_2)})^2 + Im^i(d_{(j,j+s,k_1,k_2)})^2 + Im^j(d_{(j,j+s,k_1,k_2)})^2 + Im^k(d_{(j,j+s,k_1,k_2)})^2},$$

$$j = J_0, \ldots, J-1.$$

Then, we apply the same way of constructing dual wavelet spectra as described in Section 4.3 using $|d_{(j,j+s,k_1,k_2)}|$ instead of $d_{(j,j+s,k_1,k_2)}$.

First, we can obtain three phases of a quaternion-valued wavelet coefficient as

$$(\phi_{d_{(j,j+s,k_1,k_2)}}, \theta_{d_{(j,j+s,k_1,k_2)}}, \psi_{d_{(j,j+s,k_1,k_2)}}) \in [-\pi, \pi] \times [-\frac{\pi}{2}, \frac{\pi}{2}] \times [-\frac{\pi}{4}, \frac{\pi}{4}]$$

where $j = J_0, \ldots, J-1$ $k_1 = 0, \ldots, m-1$, and $k_2 = 0, \ldots, n-1$.

In this case, dense intervals in modulus case is not required because we just extract a countable dual phase-based summaries. Therefore, we divide the respective ranges of three phases, $[-\pi, \pi] \times [-\frac{\pi}{2}, \frac{\pi}{2}] \times [-\frac{\pi}{4}, \frac{\pi}{4}]$, into 4 intervals with boundaries, $Q_{p_i}$, defined as

$$Q_{p_i} = (p_i - 0.5)2\pi, \quad i = 0, \ldots, 16$$

where $p_i = \frac{i}{16}$. Then, we can find subsets of level $j$ in 4 intervals for each phase as

$$\begin{aligned}
\boldsymbol{j}_{\phi,p_i,p_{i+4}} &= \{\, j \mid Q_{p_i} \le \phi_{d_{j,j+s,k_1,k_2}} < Q_{p_{i+4}}, \\
&\quad j = J_0, \ldots, J-1, \quad k_1 = 0, \ldots, m-1, \quad k_2 = 0, \ldots, n-1\} \text{ for } i = 0, 4, 8, 12, \\
\boldsymbol{j}_{\theta,p_i,p_{i+2}} &= \{\, j \mid Q_{p_i} \le \theta_{d_{j,j+s,k_1,k_2}} < Q_{p_{i+2}}, \\
&\quad j = J_0, \ldots, J-1, \quad k_1 = 0, \ldots, m-1, \quad k_2 = 0, \ldots, n-1\} \text{ for } i = 4, 6, 8, 10, \\
\boldsymbol{j}_{\psi,p_i,p_{i+1}} &= \{\, j \mid Q_{p_i} \le \psi_{d_{j,j+s,k_1,k_2}} < Q_{p_{i+1}}, \\
&\quad j = J_0, \ldots, J-1, \quad k_1 = 0, \ldots, m-1, \quad k_2 = 0, \ldots, n-1\} \text{ for } i = 6, 7, 8, 9,
\end{aligned}$$

and calculate averages of levels in each subset of three phases as

$$\begin{aligned}
\bar{j}_{\phi,p_i,p_{i+4}} &= \frac{1}{n_i}\sum j, \quad \forall j \in \boldsymbol{j}_{\phi,p_i,p_{i+4}} \text{ for } i = 0, 4, 8, 12, \\
\bar{j}_{\theta,p_i,p_{i+2}} &= \frac{1}{n_i}\sum j, \quad \forall j \in \boldsymbol{j}_{\theta,p_i,p_{i+2}} \text{ for } i = 4, 6, 8, 10, \\
\bar{j}_{\psi,p_i,p_{i+1}} &= \frac{1}{n_i}\sum j, \quad \forall j \in \boldsymbol{j}_{\psi,p_i,p_{i+1}} \text{ for } i = 6, 7, 8, 9,
\end{aligned}$$

where $n_i$ is the number of coefficients in a corresponding set $\boldsymbol{j}$. Finally, a dual phase-based statistics, $\bar{j}^2_{\phi,p_i,p_{i+4}}, \bar{j}^2_{\theta,p_i,p_{i+2}}, \bar{j}^2_{\psi,p_i,p_{i+1}}$, are obtained as new modalities.

## 4.4 Applications

Since the proposed methods in Chapter 2, 3, and 4 share the same datasets, we set aside all analysis results in Chapter 5 (APPLICATIONS). In particular, the

detail explanations in terms of dual wavelet spectra in real, complex, and quaternion domains for all four applications can be found in section 5.1.3.3, 5.2.3.3, 5.3.3.3, and 5.4.3.3.

## *4.5 Conclusions and Future Studies*

In this chapter, we examined dual wavelet-based spectra with non-decimated wavelet transform for both 1-D and 2-D cases. Furthermore, we extended it from real domain to complex and quaternion domains with a non-decimated complex wavelet transform ($\text{NDWT}_{\text{C}}$) and a non-decimated quaternion wavelet transform (NDQWT). Through comparative simulations in four real-life applications, we demonstrated that the proposed dual method can help increase performances of the original counterpart in classification problems with additional descriptors obtained via dual relations. Therefore, the dual wavelet spectra may be of interest to researchers seeking new perspectives on wavelet-based classification methodology for signals or images with intrinsic self-similarity.

As possible future studies, using $d^{(h)}$ and $d^{(v)}$ in addition to $d^{(d)}$ from the scale-mixing 2-D wavelet transform could potentially improve the performance of classification. Specifically, asymmetric statistics in Roberts et al (2017) can be considered as new descriptors. Moreover, performing a scale-mixing 2-D wavelet transforms with different wavelet filters for rows and columns of pixels would provide more modeling freedom. For example, one can be the complex-valued filter and the other can be the real-valued filter for which classification is optimal. Finally, we may be interested in more robust ways of calculating the dual spectral slopes as suggested in Hamilton et al (2011) or Feng et al (2018).

# CHAPTER V

# APPLICATIONS

**Table 3:** Descriptions of employed wavelet-based features.

| Symbol | Method | Description |
|---|---|---|
| Slope | All versions of original WT | Spectral slope |
| Slope* | All versions of dual WT | Dual spectral slope |
| $\angle d_j$ | $\mathrm{WT_C}$ and $\mathrm{NDWT_C}$ | Average phase ($\angle d$) at level j with complex filter |
| $\overline{j}_{\angle d,p_i,p_{i+1}}$ | Dual $\mathrm{NDWT_C}$ | Average level at each phase ($\angle d$) interval with complex filter |
| $\phi_j$ | QWT and NDQWT | Average 1st phase ($\phi$) at level j with quaternion filter |
| $\theta_j$ | QWT and NDQWT | Average 2nd phase ($\theta$) at level j with quaternion filter |
| $\psi_j$ | QWT and NDQWT | Average 3rd phase ($\psi$) at level j with quaternion filter |
| $\overline{j}_{\phi,p_i,p_{i+4}}$ | Dual NDQWT | Average level at each 1st phase ($\phi$) interval with quaternion filter |
| $\overline{j}_{\theta,p_i,p_{i+2}}$ | Dual NDQWT | Average level at each 2nd phase ($\theta$) interval with quaternion filter |
| $\overline{j}_{\psi,p_i,p_{i+1}}$ | Dual NDQWT | Average level at each 3rd phase ($\psi$) interval with quaternion filter |

## *5.1 Application 1: Classifying Pupillary Signal Data*

The human computer interaction (HCI) community has been interested in evaluating and improving user performance and interaction in a variety of fields. In particular, a variety of researches have been performed to investigate the interactions of users with age-related macular degeneration (AMD) since it is one of main causes of visual impairments and blindness in people over 55 years old (The Schepens Eye Research Institute, 2002). AMD influences high resolution vision that affects abilities of people for focus-intensive tasks such as using a computer (The Center for the Study of Macular Degeneration, 2002). The research has proved that people with AMD are likely to show worse performance than ordinary people based on measures such as task times and errors on simple computer-based tasks. In this regard, mental workload due to sensory impairments is well known as a significant factor of human performance while interacting with a complicated system (Gopher and Donchin,

1986). However, only a few studies have been performed to investigate how mental workload due to sensory impairments makes effects on the performance mentioned above. Thus, we need to consider pupil diameter that is one of significant measures of workload (Loewenfeld, 1999; Andreassi, 2000). However, the pupil has such a complex control mechanism that it is difficult to find meaningful signals from considerably noised signals of pupillary activity (Barbur, 2004). Therefore, it is necessary for a strong support to develop an analytical model to analyze dynamic pupil behaviors. Note that trends in high frequency of pupil-diameter measures are not significant because other factors that are not related to the pathologies could affect them, such as a change of environmental light intensity. Instead, the scaling information can be used for the analysis since pupil-diameter measures are considered self-similar signals. Thus, we propose an comprehensive wavelet-based analytic tool based on the all methods suggested in section 2, 3, and 4.

### 5.1.1 Description of Data

The dataset consists of pupillary response signals for 24 subjects as described in Table 4.

**Table 4:** Group characterization summary.

| Group | N | Visual acuity | AMD | Number of samples |
|---|---|---|---|---|
| Control | 6 | 20/20 - 20/40 | No | 1170 |
| Case 1 | 8 | 20/20 - 20/50 | Yes | 1970 |
| Case 2 | 4 | 20/60 - 20/100 | Yes | 1928 |
| Case 3 | 6 | 20/100 | Yes | 3547 |

In this summary of data, N refers to the number of subjects for each group. Visual acuity indicates the range of visual acuity scores assessed by ETDRS of the better eye and AMD represents the presence (Yes) or absence (No) of AMD. Then data are classified into 4 groups based on the visual acuity and the presence or absence of

AMD. The visual acuity is related to an ability to resolve fine visual detail and can be measured by the protocol outlined in the Early Treatment of Diabetic Retinopathy Study (ETDRS) (Moloney et al, 2006), which means that the group of case 3 is the worst case and the group of case 1 is the weakest among the three patient groups. Data on pupil diameter are recorded in the system at a rate of 60 HZ, or 60 times per second and a scaling factor is applied the relative recorded pupil diameter to account for camera distortion of size.

Note that we segmented the signals for each individual since the number of subjects is too small due to difficulty of collecting the measurements. Another reason for the segmentation is that their lengths are not equally long. For each signal, we cut the total signal into 1024-length pieces with 100 window size. For example, we obtain total 11 dataset (segments) of 1024 length from a 2048 length pupillary signal and its visual representation is provided in Figure 26. Table 4 summarizes the finalized dataset according to this segmentation concept and finally the total number of samples is 8615.

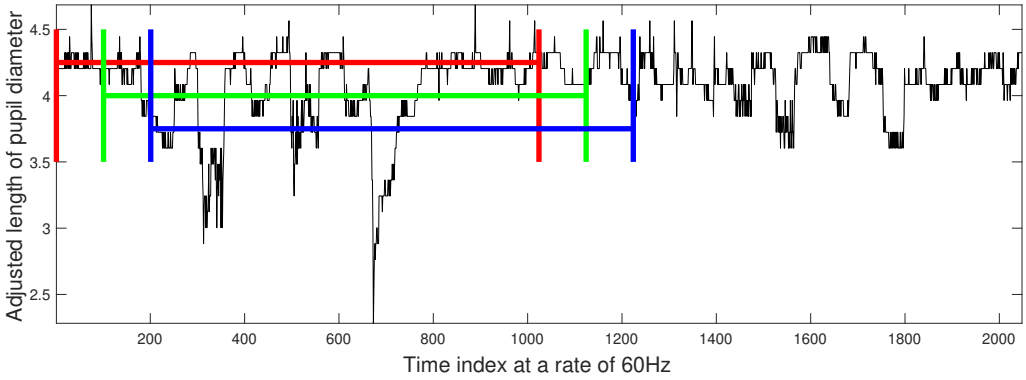



**Figure 26:** An example of 2048 length pupillary signal segmentation. The red, green, and blue intervals represent the 1st, 2nd, and 3rd segments.

### 5.1.2 Classification

In this section, we describe a way of classifying the pupillary signals based on the comprehensive wavelet-based analytic tool based on the all methods suggested

in Section 2, 3, and 4. First, we performed the all considered versions of wavelet transforms to the segmented signals found in section 5.1.1. Next, we extract all suggested features using both original and dual wavelet spectra and use them in classification analysis.

As we discussed in Section 5.1.1, segmentation of signals can increase the number of available data. However, it also induces dependence within the data for each subject. In order to quantify and remove the dependence effects within each subject, we performed a two-way nested analysis of variance (ANOVA) under the model as

$$y_{ijk} = u + \alpha_i + \beta_{j(i)} + \epsilon_{ijk}, \;\; \epsilon_{ijk} \sim N(0, \sigma^2) \tag{60}$$

with standard identifiability constraints $\sum_i \alpha_i = 0,\ \sum_j \beta_{j(i)} = 0$. For the model (60), let us consider $y_{ijk}$ as the spectral slope obtained by the $\text{NDWT}_{\text{C}}$ for each segmented pupillary signal, then it can be decomposed to a grand mean $u$, an effect of groups on the slope $\alpha_i,\ i = 1, 2, 3, 4$, an effect of subjects on the slope $\beta_{j(1)},\ j = 1, 2, \ldots, 6$, $\beta_{j(2)},\ j = 1, 2, \ldots, 8$, $\beta_{j(3)},\ j = 1, 2, \ldots, 4$, and $\beta_{j(4)},\ j = 1, 2, \ldots, 6$ for the control, case 1, case 2, and case 3, respectively, and finally an error $\epsilon_{ijk}$. The result of the two-way nested ANOVA test based on the model (60) is presented in the Table 5.

**Table 5:** The result of the two-way nested ANOVA based on the model (60).

| Source | SSE | df | MSE | F stat | Prob>F |
|---|---|---|---|---|---|
| Group | 131.5808 | 3 | 43.8603 | 498.0589 | 0 |
| Nested subject | 355.0408 | 20 | 17.7520 | 201.5848 | 0 |
| Error | 756.6321 | 8592 | 0.0881 | | |
| Total | 1243.2537 | 8614 | | | |

We can see that effects of both the groups and subjects are significantly different; the two hypotheses, $H_0 : \alpha_i = 0$ for all $i$ and $H_0 : \beta_{j(i)} = 0$ for all $i$ and $j$, are rejected. Since we are not interested in the effects of nested subjects, to represent each pupillary signal we use $y^*_{ijk} = y_{ijk} - \hat{\beta}_{j(i)}$ instead of $y_{ijk}$ for our classification

analysis where $\hat{\beta}_{j(i)} = \bar{y}_{ij.} - \bar{y}_{i..}$. All other factors such as phase averages at each level and spectral slopes from different wavelet transform methods were tested in the same way. Every test showed comparable results with the case of the spectral slope obtained by the $\text{NDWT}_{\text{C}}$. We use the $y^*_{ijk} = y_{ijk} - \hat{\beta}_{j(i)}$ instead of $y_{ijk}$ for all variables.

Using the two types of extracted descriptors with such modifications, we employed gradient boosting to classify the pupillary signals. We also considered random forest, k-NN, and SVM, however, the gradient boosting consistently outperformed the rest. In next section, we only provide the classification result by gradient boosting. The others by the unselected three machine learning methods and the comprehensive best classification result are given in Appendix B.1. For simulations, we randomly split the dataset to training and testing sets in proportion 75% to 25%, respectively. This random partition to training and testing sets was repeated $1,000$ times, and the reported prediction measures are averages over the $1,000$ runs.

### 5.1.3 Results

Since there are four labeled groups, we evaluated performances of the three suggested methods in the context of overall accuracy and sensitivities of the four groups as shown in Table 6. For comparisons, we also performed the standard DWT, $\text{WT}_{\text{C}}$, and QWT and their non-decimated versions. Moreover, the dual wavelet spectra of them were also considered to extract features. For convenience, we named the methods in order from 1 to 26 for the convenience. We recorded calculational complexity (in terms of times) for standard and dual versions of wavelets (DWT, $\text{WT}_{\text{C}}$, NDWT, $\text{NDWT}_{\text{C}}$, and NDQWT) to transform one 1024-length signal. As expected, the computation times are proportion to the overall accuracies: more accurate results take longer to calculate. It is also notable that computing times for the dual wavelet spectra are a little bit longer than the corresponding original cases. However, when they are used together to improve a classification performance as suggested, computing

time is not a summation of times for both original and dual methods since they share the same wavelet coefficients from wavelet transforms. Thus, the dual descriptors are worth considering to improve a classification performance in reasonable ranges.

**Table 6:** Gradient Boosting classification results. Total 26 methods are compared and the best result is obtained by the 26th method.

| Order | Transform | Features | Overall Accuracy rate | Sensitivity Control | Sensitivity Case 1 | Sensitivity Case 2 | Sensitivity Case 3 | Computing Time |
|---|---|---|---|---|---|---|---|---|
| 1st | DWT | Slope | 0.4458 | 0.0913 | 0.4119 | 0.0513 | 0.7966 | 0.0150s |
| 2nd | $WT_C$ | Slope | 0.3992 | 0.2066 | 0.3018 | 0.1265 | 0.6668 | |
| 3rd | | $\angle d_j$ | 0.5808 | 0.2117 | 0.2755 | 0.8395 | 0.7332 | 0.0207s |
| 4th | | Slope + $\angle d_j$ | 0.6685 | 0.3301 | 0.3728 | 0.8729 | 0.8338 | |
| 5th | QWT | Slope | 0.4082 | 0.1685 | 0.3743 | 0.1903 | 0.6249 | |
| 6th | | $\phi_j + \theta_j + \psi_j$ | 0.6915 | 0.4610 | 0.5162 | 0.7871 | 0.8121 | 0.0298s |
| 7th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.7535 | 0.5238 | 0.6647 | 0.8206 | 0.8423 | |
| 8th | NDWT | Slope | 0.4596 | 0.0916 | 0.4799 | 0.0635 | 0.7856 | 0.0192s |
| 9th | | Slope* | 0.5387 | 0.0319 | 0.0822 | 0.7725 | 0.8299 | 0.0229s |
| 10th | | 8th + 9th | 0.6177 | 0.1917 | 0.4577 | 0.7504 | 0.7747 | 0.0283s |
| 11th | $NDWT_C$ | Slope | 0.4172 | 0.1762 | 0.3936 | 0.2178 | 0.6184 | |
| 12th | | $\angle d_j$ | 0.7753 | 0.6588 | 0.6235 | 0.8755 | 0.8431 | 0.0272s |
| 13th | | Slope + $\angle d_j$ | 0.8226 | 0.6857 | 0.7263 | 0.8864 | 0.8870 | |
| 14th | | Slope* | 0.4825 | 0.1099 | 0.2185 | 0.5513 | 0.7164 | |
| 15th | | $\overline{j}_{\angle d, p_i, p_{i+1}}$ | 0.4507 | 0.1167 | 0.2460 | 0.5259 | 0.6324 | 0.0321s |
| 16th | | Slope* + $\overline{j}_{\angle d, p_i, p_{i+1}}$ | 0.5709 | 0.1282 | 0.3350 | 0.6750 | 0.7930 | |
| 17th | | 13th + 16th | 0.8879 | 0.7688 | 0.8417 | 0.9210 | 0.9348 | 0.0375s |
| 18th | NDQWT | Slope | 0.4151 | 0.1543 | 0.3887 | 0.2098 | 0.6275 | |
| 19th | | $\phi_j + \theta_j + \psi_j$ | 0.9005 | 0.8788 | 0.8102 | 0.9389 | 0.9371 | 0.0415s |
| 20th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.9278 | 0.8976 | 0.8830 | 0.9446 | 0.9537 | |
| 21th | | Slope* | 0.4596 | 0.0874 | 0.2016 | 0.5956 | 0.6535 | |
| 22th | | $\overline{j}_{\phi, p_i, p_{i+4}} + \overline{j}_{\theta, p_i, p_{i+2}} + \overline{j}_{\psi, p_i, p_{i+1}}$ | 0.7827 | 0.6000 | 0.6435 | 0.8960 | 0.8581 | 0.0620s |
| 23th | | Slope* + $\overline{j}_{\phi, p_i, p_{i+4}} + \overline{j}_{\theta, p_i, p_{i+2}} + \overline{j}_{\psi, p_i, p_{i+1}}$ | 0.8056 | 0.6403 | 0.6754 | 0.9137 | 0.8733 | |
| 24th | | 20th + 23th | 0.9523 | 0.9178 | 0.9108 | 0.9798 | 0.9717 | 0.0757s |
| 25th | | 13th + 20th | 0.9533 | 0.9216 | 0.9176 | 0.9806 | 0.9699 | 0.0687s |
| 26th | | 13th + 16th + 20th + 23th | **0.9711** | **0.9484** | **0.9361** | **0.9879** | **0.9821** | **0.1132s** |

#### *5.1.3.1 Non-decimated Complex Wavelet Transform*

Box plots of suggested non-decimated complex wavelet spectral slope and phase-based statistics at all levels are displayed in Appendix C.1.1.

Note that for the $WT_C$ and $NDWT_C$, the phase averages are more than the slopes and combinations of the two gives the best results. Another interesting finding is that classifiers without phase information tend to show low performance in classifying the control and case 2. In this context, we can conclude that information that separates the control and case 2 is located in the phase. Comparing the 4th and 8th to the 13th, one can see that the $NDWT_C$ dominates both $WT_C$ and NDWT. Therefore, the

performance improves if the wavelet spectra from $\text{NDWT}_\text{C}$ with additional descriptors based on phase are used.

#### *5.1.3.2 Non-decimated Quaternion Wavelet Transform*

Box plots of suggested non-decimated quaternion wavelet spectral slope and three phase-based statistics at all levels are displayed in Appendix C.1.2.

First, QWT (7th) outperformed DWT (1th) and $\text{WT}_\text{C}$ (4th) because they generate wavelet coefficients in a more redundant domain. Moreover, its non-decimated version, NDQWT (20th), is also superior to NDWT (8th) and $\text{NDWT}_\text{C}$ (13th) with the same trend. Although DWT (1th) and NDWT (8th) show very similar performance, $\text{NDWT}_\text{C}$ (13th) and NDQWT (20th) show better performance compared to decimated versions because of the compound effects of componential and structural redundancies. In general, classifiers with only slope show similar and inadequate performance while the others based on phase-dependent summaries perform better. In particular, classifiers with quaternion-based phase information tend to show excellent performance in classifying the control group. In this context, we can find that discriminatory information characteristic for the control group is located in the phases of quaternion-valued wavelet coefficients.

In conclusion, comparing the 7th and 8th with 20th methods, we can observe that the proposed NDQWT dominates both QWT and NDWT by its compounding positive effect of the componential and structural redundancies. Note that the best performance is shown by 25th method utilizing all features from the $\text{NDWT}_\text{C}$ and NDQWT together in one integrated model. Since this is slightly better than the NDQWT alone, we can conclude that $\text{NDWT}_\text{C}$ and NDQWT can act in a synergy with complementary interactions.

#### *5.1.3.3 Duality of Non-decimated Wavelet Transform*

Box plots of suggested features obtained by dual wavelet spectra in all real, complex, and quaternion domains are displayed in Appendix C.1.3.

The notable thing is that the dual spectral slopes in 9th, 14th, and 21th show slightly better performances compared to their original versions in 8th, 11th, and 18th. This implies that dual relations are worthy of notice in wavelet domain. In complex and quaternion versions, classifiers with only the dual phase-based statistics as in 15th and 22th tend to show lower performances compared to their original versions in 12th and 19th. However, we can demonstrate that the new descriptors with the proposed dual wavelet spectra can help classifiers with original wavelet spectra to increase performances in classification. In all of the real, complex, and quaternion domains, the performances improved when the original and dual wavelet spectra are used together as seen in 10th, 17th, and 24th. Note that the best performance is achieved when both $\text{NDWT}_{\mathbb{C}}$ and NDQWT are used together with the additional dual descriptors as in 26th method.

## *5.2 Application 2: Classifying Sounds Data*

Nowadays the air conditioners (AC) are quieter than ever. High-efficiency AC utilizes sound-dampening technology and two-stage compressors to keep noise levels below 55 decibels. So if unusual sounds come from an air conditioner during the course of normal operation, one should not ignore them as this could be a sign of malfunction or wearout. Ignoring unusual noises from AC can turn minor issues into major expenses because these noises could indicate a specific problem. The sooner we can find and resolve the cause of the noise, the better. Therefore, if an automatic noise analyzing system is available, it could increase AC's reliability and maintainability.

To develop an automatic noise analyzer, we are given three sound signals that AC (from unnamed company) could make. Since the normal sound signal is not provided,

our task is to build a classification model for the three noises named as air, sha, and water, which would be used as a prototypes. Descriptions of the noise sounds and what they may signify are as follows.

'Air' indicates hissing noise sound and implies a possible leak. So if there is a hissing sound coming from AC, it is likely either a ductwork issue or a refrigerant leak. There may be a leak in ducts allowing air or refrigerant to escape. When air is leaking the system is not running efficiently, while if the refrigerant is leaking, users may be exposed to a dangerous chemical.

'Sha' represents buzzing noise sound. Buzzing is almost always a sign of an electrical issue. If buzzing only occurs when triggering certain settings through a control panel, it is likely just because of a faulty part. But the constant buzzing is more likely to indicate a problem with the wiring, like a loose or exposed wire, causing electricity to spark within the unit. Slight humming is common and usually does not mean anything serious. On the other hand, if air conditioner is making loud buzzing, it could be a sign of loose parts or motor problems.

'Water' relates to bubbling noise sound. Although bubbling noise is not common, the problem is not likely to be a serious issue. Bubbling noise usually occurs when the condensing pump malfunctions. As condensate builds up within the AC, it drips to the bottom of the air handler where it empties into a drain pan via either gravity or a condensing pump. Water accumulation and pump malfunctioning can lead to such noises.

Since a variety of noises including the air, sha, and water require different types of professional attention, it is important task to classify them. To achieve this goal the extraction of informative features from the sound signals is critical. It is notable that trends are not significant because they usually relate to volume of the sounds. Alternatively, focusing on the scaling information may be discriminatory since these sound signals digitized at a high frequency are typically self-similar in nature.

### 5.2.1 Description of Data

Data on sound signals are recorded in the system at a rate of 0.0333MHz. The original dataset consists of three long sound signals: air, sha, and water, of unequal sizes. We segmented the signals to make the dataset convenient for our analysis. For each signal, we take subsequent non-overlapping 1024-length pieces. For instance, we obtain a total 4 dataset (segments) of length 1024 from a signal of 4096 length. We emphasize that pieces of arbitrary length can be selected to form the data set, but because of comparisons with decimated transforms, we selected length which is a power of 2.

Table 7 summarizes the finalized dataset according to this segmentation concept and finally the total number of samples is 1341.

**Table 7:** Group characterization summary.

| Group | Original length | Number of samples |
|---|---|---|
| Air | 491520 | 479 |
| Sha | 655360 | 639 |
| Water | 229376 | 223 |

### 5.2.2 Classification

In this section, we explain how to classify the sound signals. First, we performed all considered versions of wavelet transforms to the segmented signals described in Section 5.2.1. Next, we extract all suggested features using both original and dual wavelet spectra and use them in classification analysis. Finally, we chose gradient boosting to classify the sound signals. Random forest, k-NN with $k = 7$, and SVM are also considered. The gradient boosting consistently outperformed the rest. In next section, we only provide the classification result by gradient boosting. The others by the unselected three machine learning methods and the comprehensive best

classification result are given in Appendix B.2. In simulations, we randomly split the dataset into 75% part as training set, and take the remaining 25% part as testing set. This random partition to training and testing sets was repeated $1,000$ times, thus, the provided performance measures were averaged over $1,000$ runs.

### 5.2.3 Results

The performance in terms of overall accuracy and sensitivities of the three groups is shown in Table 8. For comparisons, we also performed the standard DWT, $\mathrm{WT}_{\mathbf{C}}$, and QWT and their non-decimated versions. Moreover, the dual wavelet spectra of them were also considered to extract features. For convenience, the considered methods are numbered from 1 to 26 according to transformation method and features utilized. We denoted the average of phases from $\mathrm{NDWT}_{\mathbf{C}}$ as $\angle d_j$, and the averages of three phases from NDQWT as $\phi_j, \theta_j, \psi_j$.

To compare computational costs, we recorded computing times for all considered versions of wavelets (DWT, $\mathrm{WT}_{\mathbf{C}}$, QWT, NDWT, $\mathrm{NDWT}_{\mathbf{C}}$, and NDQWT) needed to transform a single signal of length 1024. As expected, the computation times are proportional to the overall accuracies: more accurate results take longer to calculate. It is also notable that computing times for the dual wavelet spectra are a little bit longer than the corresponding original cases. However, when they are used together to improve a classification performance as suggested, computing time is not a summation of times for both original and dual methods since they share the same wavelet coefficients from wavelet transforms. Thus, the dual descriptors are worth considering to improve a classification performance in reasonable ranges.

#### *5.2.3.1 Non-decimated Complex Wavelet Transform*

Box plots of suggested non-decimated complex wavelet spectral slope and phase-based statistics at all levels are displayed in Appendix C.2.1.

Comparing the 4th and 8th to the 13th, we can demonstrate that the $\mathrm{NDWT}_{\mathbf{C}}$

**Table 8:** Gradient boosting classification results. Total 26 methods are compared and the best result is obtained by the 26th method.

| Order | Transform | Features | Overall Accuracy rate | Sensitivity Air | Sensitivity Sha | Sensitivity Water | Computing Time |
|---|---|---|---|---|---|---|---|
| 1st | DWT | Slope | 0.3841 | 0.3324 | 0.4918 | 0.1846 | 0.0150s |
| 2nd | $WT_C$ | Slope | 0.3993 | 0.3624 | 0.5221 | 0.1387 | |
| 3rd | | $\angle d_j$ | 0.4014 | 0.3767 | 0.5002 | 0.1787 | 0.0207s |
| 4th | | Slope + $\angle d_j$ | 0.4236 | 0.3969 | 0.5337 | 0.1641 | |
| 5th | QWT | Slope | 0.3947 | 0.3416 | 0.5214 | 0.1423 | |
| 6th | | $\phi_j + \theta_j + \psi_j$ | 0.6539 | 0.6208 | 0.7377 | 0.4881 | 0.0298s |
| 7th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.6580 | 0.6294 | 0.7348 | 0.5037 | |
| 8th | NDWT | Slope | 0.3968 | 0.3565 | 0.5040 | 0.1781 | 0.0192s |
| 9th | | Slope* | 0.3832 | 0.3652 | 0.4737 | 0.1636 | 0.0229s |
| 10th | | 8th + 9th | 0.4155 | 0.3744 | 0.5255 | 0.1869 | 0.0283s |
| 11th | $NDWT_C$ | Slope | 0.4026 | 0.3618 | 0.5085 | 0.1962 | |
| 12th | | $\angle d_j$ | 0.6536 | 0.7088 | 0.7422 | 0.2898 | 0.0272s |
| 13th | | Slope + $\angle d_j$ | 0.6614 | 0.7006 | 0.7504 | 0.3213 | |
| 14th | | Slope* | 0.3998 | 0.3437 | 0.5162 | 0.1885 | |
| 15th | | $\bar{j}_{\angle d, p_i, p_{i+1}}$ | 0.4482 | 0.4083 | 0.5604 | 0.2234 | 0.0321s |
| 16th | | Slope* + $\bar{j}_{\angle d, p_i, p_{i+1}}$ | 0.4725 | 0.4490 | 0.5763 | 0.2333 | |
| 17th | | 13th + 16th | 0.6852 | 0.7211 | 0.7672 | 0.3736 | 0.0375s |
| 18th | NDQWT | Slope | 0.3789 | 0.3442 | 0.4829 | 0.1606 | |
| 19th | | $\phi_j + \theta_j + \psi_j$ | 0.7558 | 0.7585 | 0.8096 | 0.5957 | 0.0415s |
| 20th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.7564 | 0.7541 | 0.8199 | 0.5860 | |
| 21th | | Slope* | 0.3936 | 0.3712 | 0.4907 | 0.1610 | |
| 22th | | $\bar{j}_{\phi, p_i, p_{i+4}} + \bar{j}_{\theta, p_i, p_{i+2}} + \bar{j}_{\psi, p_i, p_{i+1}}$ | 0.4513 | 0.4040 | 0.5815 | 0.1742 | 0.0620s |
| 23th | | Slope* + $\bar{j}_{\phi, p_i, p_{i+4}} + \bar{j}_{\theta, p_i, p_{i+2}} + \bar{j}_{\psi, p_i, p_{i+1}}$ | 0.4764 | 0.4272 | 0.6103 | 0.2030 | |
| 24th | | 20th + 23th | 0.7684 | 0.7756 | 0.8212 | 0.6027 | 0.0757s |
| 25th | | 13th + 20th | 0.7822 | 0.7895 | 0.8407 | 0.5997 | 0.0687s |
| 26th | | 13th + 16th + 20th + 23th | **0.7914** | **0.8039** | **0.8491** | **0.6014** | **0.1132s** |

dominates both $WT_C$ and NDWT. Furthermore, the classifier with only phase averages provides a better result than the one with only modulus as comparing 11th and 12th. In particular, the classifiers including phase averages (12th and 13th) show almost double sensitivity for air sound compared to the one with only modulus (11th). Thus, we can conclude that information that separates the air sound is located in the complex phase. The best performance is achieved by 13th derived from the wavelet spectra from $NDWT_C$ with additional descriptors.

#### *5.2.3.2 Non-decimated Quaternion Wavelet Transform*

Box plots of suggested non-decimated quaternion wavelet spectral slope and three phase-based statistics at all levels are displayed in Appendix C.2.2.

As expected, $WT_C$ (4th) is better than DWT (1th) and QWT (7th) is superior to

$WT_C$ (4th) because they produce wavelet coefficients in a more redundant domain. Also, their corresponding non-decimated versions, NDWT (8th), $NDWT_C$ (13th), and NDQWT (20th), show the same trend and better performance compared to decimated versions, except for the NDWT because of the compound effects of componential and structural redundancies. Generally, all methods using slopes show similar and inadequate performance while the methods that in addition use phase-based statistics perform better. In particular, classifiers without quaternion-based phase information tend to show low performance in classifying the water sound. In this context, we can conclude that discriminatory information characteristic for the water sound is located in the phases of quaternion-valued wavelet coefficients. This, of course, may not be case for arbitrary data, but the results in this case validated the use of quaternion-based phase information.

In conclusion, comparing the 7th and 8th with 20th methods, we can observe that the proposed NDQWT dominates both QWT and NDWT by its compounding positive effect of the componential and structural redundancies. Note that the best performance is achieved if we utilize all features from the $NDWT_C$ and NDQWT together in one integrated model as 25th, which is slightly better than the NDQWT alone.

#### *5.2.3.3 Duality of Non-decimated Wavelet Transform*

Box plots of suggested features obtained by dual wavelet spectra in all real, complex, and quaternion domains are displayed in Appendix C.2.3.

The dual spectral slopes in 9th, 14th, and 21th show very similar performances compared to their original versions in 8th, 11th, and 18th. This implies that dual approaches can be another options in wavelet domain. In complex and quaternion versions, classifiers with only the dual phase-based statistics as in 15th and 22th tend to show lower performances compared to their original versions in 12th and 19th.

However, the new descriptors with the proposed dual wavelet spectra slightly increase performances of primal versions as seen in 17th and 24th. This is a little significant compared to the case in Section 5.1.3.3, however, it is still worth considering because of its possibility of development. Finally, the 26th method with all descriptors shows the best performance.

## *5.3 Application 3: Screening Mammograms*

Breast cancer is the second leading cause of cancer-related death in women in the United States. The National Cancer Institute's research in Altekruse et al (2010) estimated that 1 in 8 women is likely to develop breast cancer during their lives. The U.S Department of Health and Human Services set a goal to reduce breast cancer death rate by 10% by 2020. Mammography is the one of the widely-used screening methods for early detection of breast cancer which can improve prognosis as well as lead to less invasive interventions (National Cancer Institute, 2014). However, the radiological interpretation of mammogram images is a difficult task due to the heterogeneous nature of normal breast tissue. In other words, it is difficult to classify cancerous and non-cancerous images by merely looking at them. Moreover, cancers can be of similar radiographic density as normal tissue, which can affect correct detection and decrease the sensitivity of the tests. Specificity of detection is of concern as well because it was observed that of the 5% of the mammogram images suggested for further testing, as many as 93% turned out to be false positives (Houssami et al, 2006). Therefore, it is very important to improve both the sensitivity and specificity of the mammographic diagnostics.

It is well-known fact that one of the testing modalities is a density and fine scale structure of the breast tissue. This indicates that the scaling information of the digitized images can be utilized for classification. Some previous work on mammogram classification by using a wavelet spectra can be found in Jeon et al (2014), Roberts

et al (2017), and Feng et al (2018). Since the wavelet spectra captures information contained in the background tissue of images rather than predefined templates of expected cancer morphology (tumors and microcalcifications), the spectral descriptors provide for a new and independent modality for diagnostic testing.

A study of Jeon et al (2014) suggested a classification procedure based on the estimated slope of modulus and phase average from the finest level in a $\mathrm{WT_C}$ transformed image. As mentioned in the previous section, the method showed relatively low classification accuracy in spite of better balancing specificity and sensitivity compared to other wavelet-based methods using real-valued wavelets. Another disadvantage of the method was that it only can be applied to squared images of dyadic size, since it is based on the standard DWT. In studies by Jeon et al (2014), and Roberts et al (2017) the mammogram images were manually split into 5 dyadic sub-images due to this limitation in experiments. This manual selection of sub-images is impractical for screening mammogram images and even causes a problem of multicollinearity due to overlapping. The study of Kang and Vidakovic (2016) resolved these problems by using the NDWT with non-decimation property. However, its classification result can increase more if we apply the wavelet filter in complex or quaternion as well as their dual versions. In the next section, we provide classification results using all forementioned methods in section 2, 3, and 4 and demonstrate that the three proposed methods dominates others.

### 5.3.1 Description of Data

The collection of digitized mammograms for analysis was obtained from the University of South Florida's Digital Database for Screening Mammography (DDSM), which are explained in detail in Heath et al (2001). Images from this database containing suspicious areas are accompanied gold standard true label assessed and verified

through biopsy. We selected 45 normal controls (benign) and 79 cancer cases (malignant) scanned on the HOWTEK scanner at the full 43.5 micron per pixel spatial resolution. Each case contains craniocaudal (CC) and mediolateral oblique (MLO) projection mammograms from a screening exam. We only analyze the CC projections. Note that an image containing an area outside of breast can seriously impact the result when self-similarity features are used in classification. Since the outside area is smooth the spectral slope may appear steeper. To resolve this problem, the studies in Jeon et al (2014), and Roberts et al (2017) split the mammogram images into 5 sub-images within the tissue region. This image-by-image splitting method, however, has problems since some subimages partially overlapped. Instead, the study in Kang et al (2019) used a mask-based method to remove irrelevant parts of the mammogram image, and to define self-similarity properties based on coefficients belonging only to tissue part. However, the masked images also covered the side regions of breast tissues that are unlikely to contain significant information on the cancer status. Thus, we alternatively select a single region of interest (ROI) from each mammogram image as illustrated in Figure 27. Even though we could analyze images of any size, thanks to the non-decimation property, the sub-images of size $1024 \times 1024$ were manually selected because some other methods used for comparisons require dyadic image dimensions.

### 5.3.2 Classification

In this section, we explain how to classify the mammogram images based on the comprehensive wavelet-based analytic tool based on the all methods suggested in section 2, 3, and 4. First, on the ROI images from Section 5.3.1 we applied the all versions of scale-mixing 2-D wavelet transforms with $s = 0$ for comparison.

With all of the extracted features, we employed random forest to classify the mammogram images. We also considered the logistic regression, k-NN, SVM, and gradient

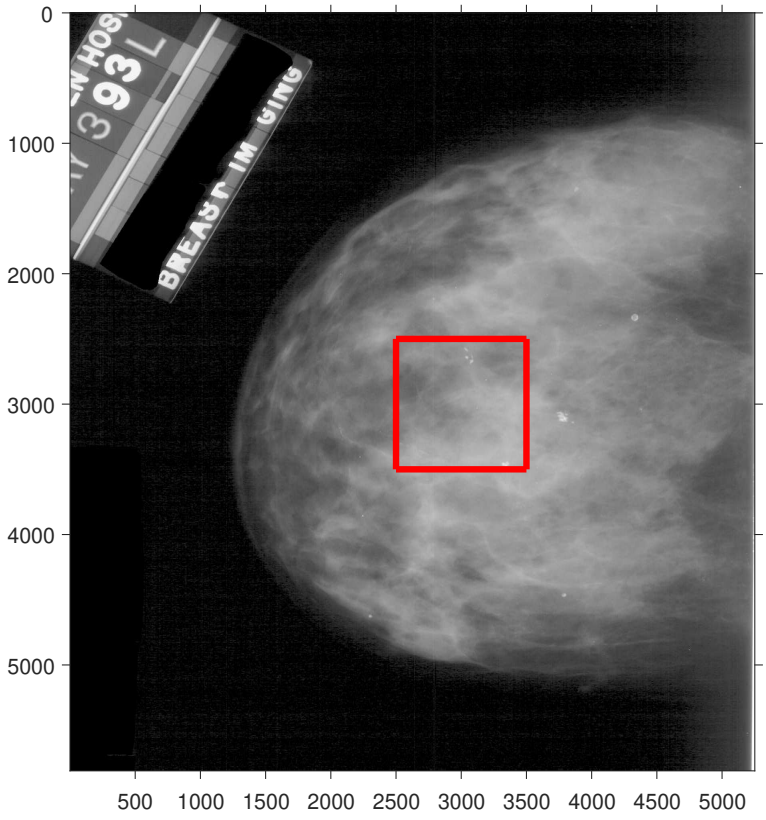



**Figure 27:** An example of mammogram image. The $1024 \times 1024$ area surrounded by red lines indicates the ROI.

boosting, however, the random forest consistently outperformed the competitors. In next section, we only provide the classification result by random forest. The others by the unselected four machine learning methods and the comprehensive best classification result are given in Appendix B.3. For a simulation, since the dataset is imbalanced and has a relatively small size, we selected 75% for training and 25% for testing at random for both control and case samples. The classification was repeated 1,000 times, and the prediction measures were obtained by averaging over 1,000 runs.

### 5.3.3 Results

We compared classification performances in the context of sensitivity, specificity, and overall accuracy rate, which are shown in Table 9. For comparisons, we also performed the standard DWT, $\mathrm{WT}_{\mathbf{C}}$, and QWT and their non-decimated versions. Moreover, the dual wavelet spectra of them were also considered to extract features. For convenience, we named the methods in order from 1 to 26 for the convenience. We denote a pair of levels as $(j)$ instead of $(j, j+s)$ because we only consider diagonal wavelet coefficients in scale-mixing 2-D wavelets.

Similar to the 1-D applications, we recorded computation times for all standard and dual versions of wavelets (DWT, $\mathrm{WT}_{\mathbf{C}}$, NDWT, $\mathrm{NDWT}_{\mathbf{C}}$, and NDQWT) to transform one $1024 \times 1024$ image. The computation times also increase with the increase of overall accuracies, as in with 1-D case, however, the rate of increase is much larger. This is because in 2-D the wavelet transform needs double matrix multiplication, compared to single in 1-D case. Although the times rapidly increase, they are still in a reasonable range, for the 25th method showing the best performance takes approximately 28 seconds per image.

In this 2-D case, computing times for dual wavelet spectra are rather longer than original ones. However, as discussed in Section 5.1.3, this can decrease significantly since computing time is not a summation of times for both original and dual methods since they share wavelet transforms.

#### *5.3.3.1 Non-decimated Complex Wavelet Transform*

Box plots of suggested non-decimated complex wavelet spectral slope and phase-based statistics at all levels are displayed in Appendix C.3.1.

Comparing the 4th and 8th with the 13th, the $\mathrm{NDWT}_{\mathbf{C}}$ dominates both $\mathrm{WT}_{\mathbf{C}}$ and NDWT, thus we can see that the $\mathrm{NDWT}_{\mathbf{C}}$ acts in a synergy. Note that the classifier with only phase averages (12) is slightly better than the one with only modulus (11).

**Table 9:** Random Forest classification results. Total 26 methods are compared and the best result is achieved by the 25th method.

| Order | Transform | Features | Accuracy rate | Specificity | Sensitivity | Computing Time |
|---|---|---|---|---|---|---|
| 1st | DWT | Slope | 0.5306 | 0.3651 | 0.6302 | 0.0724s |
| 2nd | $\text{WT}_\text{C}$ | Slope | 0.4900 | 0.3114 | 0.5991 | |
| 3rd | | $\angle d_j$ | 0.5117 | 0.3347 | 0.6175 | 0.3378s |
| 4th | | Slope + $\angle d_j$ | 0.5173 | 0.2975 | 0.6505 | |
| 5th | QWT | Slope | 0.5608 | 0.3756 | 0.6734 | |
| 6th | | $\phi_j + \theta_j + \psi_j$ | 0.6190 | 0.1723 | 0.8806 | 0.6576s |
| 7th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.6174 | 0.1685 | 0.8814 | |
| 8th | NDWT | Slope | 0.5571 | 0.3726 | 0.6694 | 2.3428s |
| 9th | | Slope* | 0.5165 | 0.3354 | 0.6257 | 7.9523s |
| 10th | | 8th + 9th | 0.5617 | 0.2908 | 0.7231 | 9.1563s |
| 11th | $\text{NDWT}_\text{C}$ | Slope | 0.5453 | 0.3667 | 0.6541 | |
| 12th | | $\angle d_j$ | 0.7456 | 0.7038 | 0.7748 | 8.2451s |
| 13th | | Slope + $\angle d_j$ | 0.7342 | 0.6954 | 0.7617 | |
| 14th | | Slope* | 0.6082 | 0.4651 | 0.6956 | |
| 15th | | $\bar{j}_{\angle d, p_i, p_{i+1}}$ | 0.6723 | 0.4135 | 0.8260 | 21.4702s |
| 16th | | Slope* + $\bar{j}_{\angle d, p_i, p_{i+1}}$ | 0.6705 | 0.4020 | 0.8303 | |
| 17th | | 13th + 16th | 0.7472 | 0.6915 | 0.7851 | 27.5641s |
| 18th | NDQWT | Slope | 0.5290 | 0.3487 | 0.6375 | |
| 19th | | $\phi_j + \theta_j + \psi_j$ | 0.7678 | 0.6666 | 0.8309 | 20.1100s |
| 20th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.7651 | 0.6639 | 0.8296 | |
| 21th | | Slope* | 0.5858 | 0.3469 | 0.7271 | |
| 22th | | $\bar{j}_{\phi, p_i, p_{i+4}} + \bar{j}_{\theta, p_i, p_{i+2}} + \bar{j}_{\psi, p_i, p_{i+1}}$ | 0.6375 | 0.3397 | 0.8152 | 54.5265s |
| 23th | | Slope* + $\bar{j}_{\phi, p_i, p_{i+4}} + \bar{j}_{\theta, p_i, p_{i+2}} + \bar{j}_{\psi, p_i, p_{i+1}}$ | 0.6499 | 0.3509 | 0.8281 | |
| 24th | | 20th + 23th | 0.7413 | 0.6132 | 0.8218 | 63.1964s |
| 25th | | 13th + 20th | **0.7830** | **0.7442** | **0.8116** | 28.3551s |
| 26th | | 13th + 16th + 20th + 23th | 0.7761 | 0.7397 | 0.8012 | **90.7605s** |

It is also notable that the phase averages dominate slopes when comparing 11th with 12th, and even the phase averages alone is slightly outperform the slope in comparing 12th with 13th. This, of course, may not be the case for other data, but these results emphasized the discriminatory power of the phase information. Note that specificity significantly increased when the phase averages of $\text{NDWT}_\text{C}$ are included. In conclusion, we can see that the best performance is achieved by 12th method which is based on $\text{NDWT}_\text{C}$ with only phase-based features.

#### *5.3.3.2 Non-decimated Quaternion Wavelet Transform*

Box plots of suggested non-decimated quaternion wavelet spectral slope and three phase-based statistics at all levels are displayed in Appendix C.3.2.

As expected the methods show almost same trend comparing to the 1-D case in Section 5.1.3 and 5.2.3. Comparing 7th, 8th, and 20th, we conclude that the NDQWT

outperforms both QWT and NDWT. In particular, the specificity of NDQWT (20th) significantly increased compared to that of QWT (7th) and NDWT (8th). This shows that phase information in non-decimated fashion can perform better compared to the primal versions. Moreover, the three phase averages as descriptors outperform the slope, when comparing 18th and 19th.

Although the total accuracy of NDQWT is higher than that of $\text{NDWT}_\text{C}$, the former shows higher specificity and the latter does higher sensitivity when compared each other. As 25th method that is a combination of both NDQWT and $\text{NDWT}_\text{C}$, shows the best performance with balanced sensitivity and specificity, we can see that there are complementary interactions between them.

#### *5.3.3.3 Duality of Non-decimated Wavelet Transform*

Box plots of suggested features obtained by dual wavelet spectra in all real, complex, and quaternion domains are displayed in Appendix C.3.3.

In real domain, the primal slope outperforms the dual slope and there is almost no improvement when used together as in 10th. In complex and quaternion domain, dual spectral slopes (14th and 21th) slightly outperform the original counterparts (11th and 18th). However, phase-dependent summaries in dual wavelet spectra for both $\text{NDWT}_\text{C}$ and NDQWT (15th and 22th), show lower performance compared to their primal versions (12th and 19th). Moreover, in this case, the additional dual descriptors do not tend to improve classification performance when used together with their primal versions as seen in 17th, 24th, and 26th. For all four applications, this is the only case that the 26th classifier with all descriptors from both primal and dual wavelet spectra, does not show the best performance. In this case, 25th method that is a combination of both NDQWT and $\text{NDWT}_\text{C}$, is the best, however, this may not be case for arbitrary data as shown in Section 5.3.3.3.

### 5.3.4 Comparison with CNN

In a final comparison, we applied CNN (Convolutional Neural Network) which is the state-of-the-art image analyzing tool. Our goal of this additional experiment was to compare CNN with the proposed method in terms of accuracy and computing time. Tensorflow 1.5.0 in Python 3.5.2 was used for CNN with 5 layers, 0.001 learning rate, 11 batch size, and 100 training epochs and MATLAB 9.1.0 is for wavelet-based methods on Intel(R) Core(TM) i7-6500U CPU at 2.50GHz with 12GM RAM. We found their computing times notably different. For the 25th method, the time for extracting features was 58 mins and then 1000 iterations of training and testing took additional 56 sec. Thus, the total processing time was approximately 58 mins 56 secs. In contrast, the CNN took 15 hours 1 min on average for its one-time training and testing. Given large size of training data, the CNN did not need multiple training because large size of testing data was also available. However, due to a limited number of mammogram images, multiple training for validation was needed. This would take approximately $15 \times 1000$ hours for 1000 iterations. Worse yet, the average accuracy for 10 iterations was 0.6250 with 0.4286 specificity and 0.7059 sensitivity; these are inferior to the 25th method showing the best performance. One explanation is the following. The information on cancerous or non-cancerous tissue is strongly related to details, which are linked to the self-similarity, as discussed before. Generally, the CNN is well known for its superb performance on classifying MNIST or CIFAR-10 where detail information is not critical. On the other hand, the wavelet-based classifiers are very useful when critical information is located not in the coarse approximations but details, such as noise dynamics, for example.

## *5.4 Application 4: Seam Detection in Steel Rolling Process*

In various manufacturing fields, fault diagnosis or anomaly detection in 2-D images have been widely deployed thanks to their low implementation cost and the rich information from a high acquisition rate of image sensors. Thus, considerable research has been conducted on inspection systems for rolling process (Jin et al, 2004), composite material fabrication (Sohn et al, 2004), liquid crystal display manufacturing (Jiang et al, 2005), fabric and textile manufacturing (Kumar, 2008), and so on. In these systems, snapshots of particular products or parts are obtained during production process, to be analyzed as digital images for detecting defects or anomalies.

A representative example is an inspection system for the subsequent rolling process, continuous casting manufacturing, in which a semi-finished billet is solidified from molten metal. Specifically, rolling process is a high-speed deformation by consistent diameters between sets of rollers to decrease the cross-sectional size in a long steel bar by applying compressive forces. One of surface defects caused during the rolling process is a seam defect that leads to stress concentration on the bulk material, which further may create failures when a steel bar is used. Thus, timely detection of such anomaly is significant to preventing such failures and for reducing overall manufacturing costs. For this purpose, advanced vision sensing systems have been developed in rolling processes to obtain high-resolution snapshots of the product surface with a high data acquisition rate of billets at short time intervals. An example of the bar surface image with seam defects of rolling process is shown in Figure 28.

Until recently a quality inspection or anomaly detection had been performed manually. However, automatic inspection systems with high speed and accuracy have been developed, since machine learning is utilized to analyze images. Since the seam defects are typically sparse as shown in Figure 28, the inherent self-similarity of the background is disturbed by the anomaly, that can be sensed well by wavelet tools. The proposed wavelet-based method considered here should not be used in practice

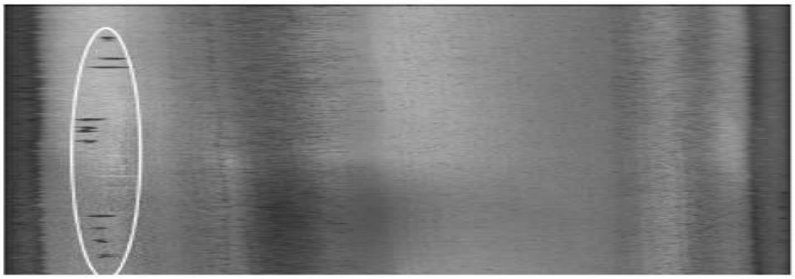


**Figure 28:** An example of surface image of steel rolling bar. The white ellipse indicates seem defects in rolling process.

alone. This should be added to a battery of other standard image recognition tools based on the image features in the original domain.

To illustrate the power of the proposed method, we set a classification problem with a test set of surface images and compare classification results in following sections.

### 5.4.1 Description of Data

One hundred surface images of size of $128 \times 512$ pixels of a rolling bar are sequentially collected. In this dataset, the images generally appear smooth in the rolling (vertical) direction. Seam defects that occur typically towards the end of the rolling bar started to appear from 76th image, so 71th to 80th images are omitted for an objective experiment. We define the first 70 images as controls without seam defects and the last 20 images as case samples with seam defects.

### 5.4.2 Classification

In this section, we describe a way of classifying the rolling surface images. First, we performed all considered scale-mixing 2-D versions of wavelet transforms to the rolling bar images. Next, we extract all suggested features using both original and dual wavelet spectra as descriptors used in classification analysis.

At last, we decided to use random forests methodology to classify the rolling surface images. In addition, we considered logistic regression, k-NN, SVM, and gradient

boosting, however, the random forests consistently outperformed the rest. In next section, we only provide the classification result by random forest. The others by the unselected four machine learning methods and the comprehensive best classification result are given in Appendix B.4. In simulations, since the dataset is imbalanced and of a relatively small size, we separately selected 75% for training and 25% for testing for both control and case samples to repeatedly measure performances 1,000 times. Thus, the prediction measures were obtained by averaging the repeated 1,000 runs.

### 5.4.3 Results

We compared classification performances in the context of sensitivity, specificity, and overall accuracy rate, which are shown in Table 10. For comparisons, we also performed the standard DWT, $\mathrm{WT}_{\mathbf{C}}$, and QWT and their non-decimated versions. Moreover, the dual wavelet spectra of them were also considered to extract features. For convenience, the considered methods are numbered from 1 to 26 according to transformation method and features utilized. We denoted the phase average from $\mathrm{NDWT}_{\mathbf{C}}$ as $\angle d_j$ and the three phase averages from NDQWT as $\phi_j, \theta_j, \psi_j$.

As in the 1-D application, we recorded computation times for all considered versions of wavelets (DWT, $\mathrm{WT}_{\mathbf{C}}$, QWT, NDWT, $\mathrm{NDWT}_{\mathbf{C}}$, and NDQWT) needed to transform a single $128 \times 512$ image. The computing times also increase with the increase of the overall accuracies, as in with 1-D case, however, the rate of increase is much steeper. This is because 2-D wavelet transform requires double matrix multiplications, as explained in Section 3.1. Although the times rapidly increase, they are still in the reasonable range; the NDQWT takes approximately 1.17 seconds per image.

In this 2-D case, computing times for dual wavelet spectra are rather longer than original ones. However, as discussed in Section 5.1.3, this can decrease significantly since computing time is not a summation of times for both original and dual methods

since they share wavelet transforms.

**Table 10:** Random Forest classification results. Total 26 methods are compared and the best result is achieved by the 26th method.

| Order | Transform | Features | Accuracy rate | Specificity | Sensitivity | Computing Time |
|---|---|---|---|---|---|---|
| 1st | DWT | Slope | 0.6057 | 0.7372 | 0.2110 | 0.0222s |
| 2nd | $WT_C$ | Slope | 0.7449 | 0.8324 | 0.4822 | |
| 3rd | | $\angle d_j$ | 0.6156 | 0.7127 | 0.3240 | 0.0937s |
| 4th | | Slope + $\angle d_j$ | 0.7811 | 0.9050 | 0.4094 | |
| 5th | QWT | Slope | 0.7222 | 0.8254 | 0.4124 | |
| 6th | | $\phi_j + \theta_j + \psi_j$ | 0.8359 | 0.9749 | 0.4188 | 0.4206s |
| 7th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.8447 | 0.9823 | 0.4320 | |
| 8th | NDWT | Slope | 0.6594 | 0.7569 | 0.3666 | 0.0696s |
| 9th | | Slope* | 0.7691 | 0.8494 | 0.5280 | 0.3593s |
| 10th | | 8th + 9th | 0.8226 | 0.9311 | 0.4968 | 0.3712s |
| 11th | $NDWT_C$ | Slope | 0.7844 | 0.8755 | 0.5108 | |
| 12th | | $\angle d_j$ | 0.8600 | 0.9664 | 0.5408 | 0.2538s |
| 13th | | Slope + $\angle d_j$ | 0.8938 | 0.9717 | 0.6598 | |
| 14th | | Slope* | 0.7612 | 0.8549 | 0.4802 | |
| 15th | | $\bar{j}_{\angle d,p_i,p_{i+1}}$ | 0.8979 | 0.9602 | 0.7108 | 0.6386s |
| 16th | | Slope* + $\bar{j}_{\angle d,p_i,p_{i+1}}$ | 0.9085 | 0.9669 | 0.7330 | |
| 17th | | 13th + 16th | 0.9263 | 0.9809 | 0.7626 | 0.7168s |
| 18th | NDQWT | Slope | 0.7483 | 0.8227 | 0.5250 | |
| 19th | | $\phi_j + \theta_j + \psi_j$ | 0.9139 | 0.9868 | 0.6950 | 1.1777s |
| 20th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.9220 | 0.9831 | 0.7384 | |
| 21th | | Slope* | 0.8214 | 0.8985 | 0.5515 | |
| 22th | | $\bar{j}_{\phi,p_i,p_{i+4}} + \bar{j}_{\theta,p_i,p_{i+2}} + \bar{j}_{\psi,p_i,p_{i+1}}$ | 0.8911 | 0.9507 | 0.6825 | 2.1205s |
| 23th | | Slope* + $\bar{j}_{\phi,p_i,p_{i+4}} + \bar{j}_{\theta,p_i,p_{i+2}} + \bar{j}_{\psi,p_i,p_{i+1}}$ | 0.9230 | 0.9719 | 0.7521 | |
| 24th | | 20th + 23th | 0.9384 | 0.9873 | 0.7674 | 2.8963s |
| 25th | | 13th + 20th | 0.9310 | 0.9819 | 0.7529 | 1.4315s |
| 26th | | 13th + 16th + 20th + 23th | **0.9523** | **0.9999** | **0.8029** | **3.6131s** |

#### *5.4.3.1 Non-decimated Complex Wavelet Transform*

Box plots of suggested non-decimated complex wavelet spectral slope and phase-based statistics at all levels are displayed in Appendix C.4.1.

Comparing the 4th and 8th to the 13th, we can see that the $NDWT_C$ dominates both $WT_C$ and NDWT. It is also notable that the phase averages dominate slopes when comparing 11th with 12th. When they are used together in one integrated model (13th), it shows the best performance. In particular, the sensitivity significantly increased.

#### *5.4.3.2 Non-decimated Quaternion Wavelet Transform*

Box plots of suggested non-decimated quaternion wavelet spectral slope and three phase-based statistics at all levels are displayed in Appendix C.4.2.

The methods show almost same trend comparing to the counterparts in Section 5.2.3; the only difference is that in this case the slopes are also informative features. Contrasting 7th and 8th to 20th, we see that the NDQWT dominates both QWT and NDWT. In particular, the sensitivity of NDQWT (20th) significantly increased compared to that of QWT (7th), which means that redundancy is beneficial in capturing information on defects. It is also notable that the three phase averages as descriptors outperform the slope, when comparing 18th and 19th cases. In conclusion, we find that the best performance is achieved by 25th method which is based on all descriptors from $\text{NDWT}_{\mathbf{C}}$ and NDQWT.

#### *5.4.3.3 Duality of Non-decimated Wavelet Transform*

Box plots of suggested features obtained by dual wavelet spectra in all real, complex, and quaternion domains are displayed in Appendix C.4.3.

In real domain, the dual slope in 9th outperforms the original one in 8th. Moreover, the dual slope can help the original counterpart improve performance when used together as in 10th. In complex and quaternion domain, it is hard to distinguish superiority between the original and dual wavelet spectra since their results are very similar. However, we can see that the dual spectra can support the original one as like the real domain case as seen in 17th and 24th. It is also notable that the phase-based statistics are better descriptors than the slope for all versions. In conclusion, we find that the best performance is achieved by 26th method based on all descriptors from $\text{NDWT}_{\mathbf{C}}$ and NDQWT. This indicates that the performance can be better if we apply the $\text{NDWT}_{\mathbf{C}}$ and NDQWT together with the additional dual descriptors.

### 5.4.4 Comparison with CNN

As a final comparison, we applied CNN (Convolutional Neural Network) which is the state-of-the-art image analyzing tool nowadays. The goal of this additional experiment is to compare CNN with the proposed method in terms of accuracy and computing time. Tensorflow 1.5.0 in Python 3.5.2 is used for CNN with 5 layers, 0.001 learning rate, 70 batch size, and 100 training epochs and MATLAB 9.1.0 is for wavelet-based methods on Intel(R) Core(TM) i7-6500U CPU at 2.50GHz with 12GM RAM. Surprisingly, we found that computing times are notably different. For 26th method showing the best performance, its time for extracting features was 5.42 mins and then 1000 iteration of training and testing took about 1.47 mins. Thus, total processing time was approximately 6.89 mins. In comparison, CNN showed 0.8987 average accuracy for 100 iterations with 0.9139 specificity and 0.8457 sensitivity, which seems to be competitive with the 26th method. Surprisingly, the CNN took 22 mins 48 secs on average for its one-time training and testing. For dataset of large size of training data, CNN does not need multiple training because a large size of testing data would be available as well. However, due to a small size of dataset, multiple training and validation runs are desired here, which will take whooping $23 \times 1000$ mins for 1000 iterations. This is the reason why the proposed method outperformed the CNN method in classification of rolling surface images in terms of both accuracy and computing time.

**Acknowledgement.** We are grateful to Prof. Andrew T. Walden, Imperial College London, and Dr. Paul Ginzberg for their kind permission to use software on quaternion wavelets they developed. Also, we acknowledge developers of Quaternion toolbox for MATLAB (Stephen and Nicolas, 2013) that facilitated calculation with matrices of quaternions in almost the same manner as with the matrices of complex numbers. Finally, we thank Dr. Seonghye Jeon and Dr. Minkyoung Kang for the

mammogram data, Dr. Bin Shi for the pupil-diameter data, and Prof. Kamran Paynabar and Ana Maria Estrada Gomez for providing the steel rolling bar image data and useful discussions.

# APPENDIX A

# PROOFS AND DERIVATIONS

## *A.1 The proof that the polynomial $P_J(z)$ satisfies the Equation (13)*

First, reform the $P_J(z)$ as

$$P_J(z) = \frac{1}{z^{2J}} \sum_{j=0}^{J} (-1)^j \binom{2J+1}{j} H(-z)^j H(z)^{2J+1-j} \tag{61}$$

where $H(z) = \frac{(1+z)^2}{4}$. Since $\binom{2J+1}{j} = \binom{2J+1}{2J+1-j}$, we can obtain

$$P_J(-z) = \frac{1}{z^{2J}} \sum_{j=J+1}^{2J+1} (-1)^{1-J} \binom{2J+1}{j} H(-z)^j H(z)^{2J+1-j} \tag{62}$$

and consequently Equation (13) by taking the difference of the last two Equation (61) and Equation (62) as follows:

$$\begin{aligned} P_J(z) - P_J(-z) &= \frac{1}{z^{2J}} \sum_{j=0}^{2J+1} (-1)^j \binom{2J+1}{j} (-H(-z))^j H(z)^{2J+1-j} \\ &= \frac{1}{z^{2J}} (H(z) - H(-z))^{2J+1} \\ &= z \end{aligned}$$

# APPENDIX B

# CLASSIFICATION RESULTS

## *B.1 Application 1: Classifying Pupillary Signal Data*

**Table 11:** Comprehensive best classification results of classifying pupillary signal data.

| Order | Transform | Features | Overall Accuracy rate | Sensitivity Control | Sensitivity Case 1 | Sensitivity Case 2 | Sensitivity Case 3 | Method |
|---|---|---|---|---|---|---|---|---|
| 1st | DWT | Slope | 0.4467 | 0.0000 | 0.1872 | 0.0000 | 0.9777 | SVM |
| 2nd | $WT_C$ | Slope | 0.5416 | 0.0000 | 0.0000 | 0.6932 | 0.9353 | SVM |
| 3rd | | $\angle d_j$ | 0.6779 | 0.2987 | 0.5121 | 0.7284 | 0.8690 | SVM |
| 4th | | Slope + $\angle d_j$ | 0.7007 | 0.3237 | 0.5399 | 0.7616 | 0.8826 | SVM |
| 5th | QWT | Slope | 0.4359 | 0.0123 | 0.1683 | 0.0000 | 0.9792 | SVM |
| 6th | | $\phi_j + \theta_j + \psi_j$ | 0.6915 | 0.4610 | 0.5162 | 0.7871 | 0.8121 | GB |
| 7th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.7535 | 0.5238 | 0.6647 | 0.8206 | 0.8423 | GB |
| 8th | NDWT | Slope | 0.4596 | 0.0916 | 0.4799 | 0.0635 | 0.7856 | GB |
| 9th | | Slope* | 0.5418 | 0.0000 | 0.0000 | 0.6931 | 0.9375 | SVM |
| 10th | | 8th + 9th | 0.6177 | 0.1917 | 0.4577 | 0.7504 | 0.7747 | GB |
| 11th | $NDWT_C$ | Slope | 0.4454 | 0.0069 | 0.1732 | 0.0000 | 0.9754 | SVM |
| 12th | | $\angle d_j$ | 0.7883 | 0.6641 | 0.6012 | 0.8821 | 0.8819 | RF |
| 13th | | Slope + $\angle d_j$ | 0.8345 | 0.6842 | 0.7400 | 0.8899 | 0.9062 | RF |
| 14th | | Slope* | 0.5222 | 0.0000 | 0.0000 | 0.6013 | 0.9389 | SVM |
| 15th | | $\overline{j}_{\angle d,p_i,\underline{p}_{i+1}}$ | 0.4890 | 0.0000 | 0.0186 | 0.4081 | 0.9601 | SVM |
| 16th | | Slope* + $\overline{j}_{\angle d,p_i,p_{i+1}}$ | 0.5723 | 0.1181 | 0.3597 | 0.7868 | 0.7215 | RF |
| 17th | | 13th + 16th | 0.8879 | 0.7688 | 0.8417 | 0.9210 | 0.9348 | GB |
| 18th | NDQWT | Slope | 0.4432 | 0.0000 | 0.1825 | 0.0000 | 0.9805 | SVM |
| 19th | | $\phi_j + \theta_j + \psi_j$ | 0.9048 | 0.9391 | 0.8253 | 0.9708 | 0.9016 | kNN |
| 20th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.9278 | 0.8976 | 0.8830 | 0.9446 | 0.9537 | GB |
| 21th | | Slope* | 0.5236 | 0.0000 | 0.0000 | 0.5984 | 0.9386 | SVM |
| 22th | | $\overline{j}_{\phi,p_i,\underline{p}_{i+4}} + \overline{j}_{\theta,p_i,\underline{p}_{i+2}} + \overline{j}_{\psi,p_i,\underline{p}_{i+1}}$ | 0.7827 | 0.6000 | 0.6435 | 0.8960 | 0.8581 | GB |
| 23th | | Slope* + $\overline{j}_{\phi,p_i,p_{i+4}} + \overline{j}_{\theta,p_i,p_{i+2}} + \overline{j}_{\psi,p_i,p_{i+1}}$ | 0.8056 | 0.6403 | 0.6754 | 0.9137 | 0.8733 | GB |
| 24th | | 20th + 23th | 0.9523 | 0.9178 | 0.9108 | 0.9798 | 0.9717 | GB |
| 25th | | 13th + 20th | 0.9533 | 0.9216 | 0.9176 | 0.9806 | 0.9699 | GB |
| 26th | | 13th + 16th + 20th + 23th | 0.9711 | 0.9484 | 0.9361 | 0.9879 | 0.9821 | GB |

**Table 12:** Random forest classification results of classifying pupillary signal data.

| Order | Transform | Features | Overall Accuracy rate | Sensitivity Control | Sensitivity Case 1 | Sensitivity Case 2 | Sensitivity Case 3 |
|---|---|---|---|---|---|---|---|
| 1st | DWT | Slope | 0.3548 | 0.1996 | 0.3268 | 0.2580 | 0.4745 |
| 2nd | $\text{WT}_\text{C}$ | Slope | 0.3604 | 0.2510 | 0.3148 | 0.2460 | 0.4840 |
| 3rd | | $\angle d_j$ | 0.6062 | 0.1560 | 0.2063 | 0.8637 | 0.8370 |
| 4th | | Slope + $\angle d_j$ | 0.6723 | 0.2912 | 0.3607 | 0.8816 | 0.8579 |
| 5th | QWT | Slope | 0.3678 | 0.2168 | 0.3558 | 0.2660 | 0.4790 |
| 6th | | $\phi_j + \theta_j + \psi_j$ | 0.6504 | 0.3073 | 0.3402 | 0.7965 | 0.8568 |
| 7th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.6997 | 0.3726 | 0.5291 | 0.7823 | 0.8568 |
| 8th | NDWT | Slope | 0.3602 | 0.2019 | 0.3414 | 0.2583 | 0.4781 |
| 9th | | Slope* | 0.4322 | 0.1957 | 0.2860 | 0.6303 | 0.4842 |
| 10th | | 8th + 9th | 0.5721 | 0.2348 | 0.4528 | 0.7382 | 0.6591 |
| 11th | $\text{NDWT}_\text{C}$ | Slope | 0.3742 | 0.2218 | 0.3487 | 0.2848 | 0.4871 |
| 12th | | $\angle d_j$ | 0.7883 | 0.6641 | 0.6012 | 0.8821 | 0.8819 |
| 13th | | Slope + $\angle d_j$ | 0.8345 | 0.6842 | 0.7400 | 0.8899 | 0.9062 |
| 14th | | Slope* | 0.4050 | 0.1631 | 0.2803 | 0.5975 | 0.4497 |
| 15th | | $\overline{j}_{\angle d, p_i, p_{i+1}}$ | 0.4785 | 0.0777 | 0.2272 | 0.5735 | 0.6984 |
| 16th | | Slope* + $\overline{j}_{\angle d, p_i, p_{i+1}}$ | 0.5723 | 0.1181 | 0.3597 | 0.7868 | 0.7215 |
| 17th | | 13th + 16th | 0.8604 | 0.7082 | 0.7699 | 0.9334 | 0.9215 |
| 18th | NDQWT | Slope | 0.3758 | 0.1976 | 0.3586 | 0.2791 | 0.4965 |
| 19th | | $\phi_j + \theta_j + \psi_j$ | 0.8667 | 0.8173 | 0.6960 | 0.9259 | 0.9453 |
| 20th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.8986 | 0.8423 | 0.8019 | 0.9281 | 0.9550 |
| 21th | | Slope* | 0.4208 | 0.1764 | 0.3020 | 0.6065 | 0.4667 |
| 22th | | $\overline{j}_{\phi, p_i, p_{i+4}} + \overline{j}_{\theta, p_i, p_{i+2}} + \overline{j}_{\psi, p_i, p_{i+1}}$ | 0.7680 | 0.5107 | 0.5648 | 0.9024 | 0.8932 |
| 23th | | Slope* + $\overline{j}_{\phi, p_i, p_{i+4}} + \overline{j}_{\theta, p_i, p_{i+2}} + \overline{j}_{\psi, p_i, p_{i+1}}$ | 0.7931 | 0.5512 | 0.6273 | 0.9224 | 0.8959 |
| 24th | | 20th + 23th | 0.9227 | 0.8530 | 0.8317 | 0.9711 | 0.9700 |
| 25th | | 13th + 20th | 0.9251 | 0.8691 | 0.8388 | 0.9673 | 0.9684 |
| 26th | | 13th + 16th + 20th + 23th | 0.9295 | 0.8665 | 0.8536 | 0.9695 | 0.9709 |

**Table 13:** Support vector machine classification results of classifying pupillary signal data.

| Order | Transform | Features | Overall Accuracy rate | Sensitivity Control | Sensitivity Case 1 | Sensitivity Case 2 | Sensitivity Case 3 |
|---|---|---|---|---|---|---|---|
| 1st | DWT | Slope | 0.4467 | 0.0000 | 0.1872 | 0.0000 | 0.9777 |
| 2nd | $\text{WT}_\text{C}$ | Slope | 0.5416 | 0.0000 | 0.0000 | 0.6932 | 0.9353 |
| 3rd | | $\angle d_j$ | 0.6778 | 0.2987 | 0.5121 | 0.7284 | 0.8690 |
| 4th | | Slope + $\angle d_j$ | 0.7006 | 0.3237 | 0.5399 | 0.7616 | 0.8826 |
| 5th | QWT | Slope | 0.4358 | 0.0123 | 0.1683 | 0.0000 | 0.9792 |
| 6th | | $\phi_j + \theta_j + \psi_j$ | 0.6279 | 0.2909 | 0.3676 | 0.7171 | 0.8351 |
| 7th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.6908 | 0.3485 | 0.5573 | 0.7358 | 0.8515 |
| 8th | NDWT | Slope | 0.4518 | 0.0000 | 0.1841 | 0.0000 | 0.9778 |
| 9th | | Slope* | 0.5417 | 0.0000 | 0.0000 | 0.6931 | 0.9375 |
| 10th | | 8th + 9th | 0.5987 | 0.0190 | 0.2981 | 0.7175 | 0.8908 |
| 11th | $\text{NDWT}_\text{C}$ | Slope | 0.4453 | 0.0069 | 0.1732 | 0.0000 | 0.9754 |
| 12th | | $\angle d_j$ | 0.7004 | 0.3367 | 0.4502 | 0.8022 | 0.9016 |
| 13th | | Slope + $\angle d_j$ | 0.7457 | 0.3909 | 0.5871 | 0.8141 | 0.9129 |
| 14th | | Slope* | 0.5222 | 0.0000 | 0.0000 | 0.6013 | 0.9389 |
| 15th | | $\overline{j}_{\angle d, p_i, p_{i+1}}$ | 0.4889 | 0.0000 | 0.0186 | 0.4081 | 0.9601 |
| 16th | | Slope* + $\overline{j}_{\angle d, p_i, p_{i+1}}$ | 0.5629 | 0.0240 | 0.0351 | 0.7307 | 0.9467 |
| 17th | | 13th + 16th | 0.7907 | 0.4739 | 0.6823 | 0.8638 | 0.9190 |
| 18th | NDQWT | Slope | 0.4431 | 0.0000 | 0.1825 | 0.0000 | 0.9805 |
| 19th | | $\phi_j + \theta_j + \psi_j$ | 0.8235 | 0.6449 | 0.6611 | 0.8676 | 0.9481 |
| 20th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.8648 | 0.6881 | 0.7596 | 0.9135 | 0.9555 |
| 21th | | Slope* | 0.5236 | 0.0000 | 0.0000 | 0.5984 | 0.9386 |
| 22th | | $\overline{j}_{\phi, p_i, p_{i+4}} + \overline{j}_{\theta, p_i, p_{i+2}} + \overline{j}_{\psi, p_i, p_{i+1}}$ | 0.6921 | 0.3500 | 0.4640 | 0.7966 | 0.8816 |
| 23th | | Slope* + $\overline{j}_{\phi, p_i, p_{i+4}} + \overline{j}_{\theta, p_i, p_{i+2}} + \overline{j}_{\psi, p_i, p_{i+1}}$ | 0.7133 | 0.3302 | 0.5111 | 0.8396 | 0.8887 |
| 24th | | 20th + 23th | 0.9049 | 0.7477 | 0.8333 | 0.9599 | 0.9669 |
| 25th | | 13th + 20th | 0.9033 | 0.7230 | 0.8324 | 0.9540 | 0.9759 |
| 26th | | 13th + 16th + 20th + 23th | 0.9171 | 0.7669 | 0.8634 | 0.9601 | 0.9741 |

**Table 14:** k-nearest neighbor ($k = 1$) classification results of classifying pupillary signal data.

| Order | Transform | Features | Overall Accuracy rate | Sensitivity Control | Sensitivity Case 1 | Sensitivity Case 2 | Sensitivity Case 3 |
|---|---|---|---|---|---|---|---|
| 1st | DWT | Slope | 0.3550 | 0.1996 | 0.3256 | 0.2601 | 0.4742 |
| 2nd | $WT_C$ | Slope | 0.3590 | 0.2530 | 0.3184 | 0.2385 | 0.4827 |
| 3rd | | $\angle d_j$ | 0.4690 | 0.2366 | 0.2498 | 0.6805 | 0.5529 |
| 4th | | Slope + $\angle d_j$ | 0.5199 | 0.2915 | 0.2989 | 0.7055 | 0.6174 |
| 5th | QWT | Slope | 0.3673 | 0.2160 | 0.3594 | 0.2674 | 0.4766 |
| 6th | | $\phi_j + \theta_j + \psi_j$ | 0.4820 | 0.3146 | 0.3303 | 0.5805 | 0.5681 |
| 7th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.5291 | 0.3514 | 0.4040 | 0.6242 | 0.6050 |
| 8th | NDWT | Slope | 0.3611 | 0.2033 | 0.3428 | 0.2610 | 0.4777 |
| 9th | | Slope* | 0.4330 | 0.1954 | 0.2855 | 0.6333 | 0.4847 |
| 10th | | 8th + 9th | 0.5248 | 0.2577 | 0.4284 | 0.6528 | 0.5965 |
| 11th | $NDWT_C$ | Slope | 0.3737 | 0.2142 | 0.3447 | 0.2852 | 0.4908 |
| 12th | | $\angle d_j$ | 0.7520 | 0.6703 | 0.6076 | 0.8604 | 0.8002 |
| 13th | | Slope + $\angle d_j$ | 0.8142 | 0.7372 | 0.7222 | 0.8821 | 0.8538 |
| 14th | | Slope* | 0.4065 | 0.1661 | 0.2823 | 0.5972 | 0.4525 |
| 15th | | $\overline{j}_{\angle d,p_i,\underline{p}_{i+1}}$ | 0.3960 | 0.1917 | 0.2804 | 0.4720 | 0.4864 |
| 16th | | Slope* + $\overline{j}_{\angle d,p_i,p_{i+1}}$ | 0.5007 | 0.2678 | 0.3669 | 0.7221 | 0.5321 |
| 17th | | 13th + 16th | 0.8428 | 0.8038 | 0.7423 | 0.9336 | 0.8624 |
| 18th | NDQWT | Slope | 0.3760 | 0.1972 | 0.3621 | 0.2813 | 0.4949 |
| 19th | | $\phi_j + \theta_j + \psi_j$ | 0.9048 | 0.9391 | 0.8253 | 0.9708 | 0.9016 |
| 20th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.9171 | 0.9514 | 0.8477 | 0.9767 | 0.9124 |
| 21th | | Slope* | 0.4195 | 0.1760 | 0.3048 | 0.6080 | 0.4601 |
| 22th | | $\overline{j}_{\phi,p_i,\underline{p}_{i+4}} + \overline{j}_{\theta,p_i,\underline{p}_{i+2}} + \overline{j}_{\psi,p_i,\underline{p}_{i+1}}$ | 0.7108 | 0.6225 | 0.5413 | 0.8711 | 0.7463 |
| 23th | | Slope* + $\overline{j}_{\phi,p_i,p_{i+4}} + \overline{j}_{\theta,p_i,p_{i+2}} + \overline{j}_{\psi,p_i,p_{i+1}}$ | 0.7461 | 0.6673 | 0.6030 | 0.8944 | 0.7707 |
| 24th | | 20th + 23th | 0.9204 | 0.9515 | 0.8457 | 0.9885 | 0.9151 |
| 25th | | 13th + 20th | 0.9371 | 0.9725 | 0.8791 | 0.9887 | 0.9296 |
| 26th | | 13th + 16th + 20th + 23th | 0.9320 | 0.9673 | 0.8673 | 0.9935 | 0.9229 |

## *B.2 Application 2: Classifying Sounds Data*

**Table 15:** Comprehensive best classification results of classifying sounds data.

| Order | Transform | Features | Overall Accuracy rate | Sensitivity Air | Sensitivity Sha | Sensitivity Water | Method |
|---|---|---|---|---|---|---|---|
| 1st | DWT | Slope | 0.4340 | 0.3142 | 0.6506 | 0.0858 | kNN |
| 2nd | $WT_C$ | Slope | 0.4339 | 0.3233 | 0.6544 | 0.0401 | kNN |
| 3rd | | $\angle d_j$ | 0.4446 | 0.3637 | 0.6323 | 0.0770 | kNN |
| 4th | | Slope + $\angle d_j$ | 0.4236 | 0.3969 | 0.5337 | 0.1641 | GB |
| 5th | QWT | Slope | 0.4336 | 0.3324 | 0.6403 | 0.0763 | kNN |
| 6th | | $\phi_j + \theta_j + \psi_j$ | 0.6654 | 0.6101 | 0.7934 | 0.4130 | RF |
| 7th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.6702 | 0.6209 | 0.7931 | 0.4244 | RF |
| 8th | NDWT | Slope | 0.4347 | 0.3189 | 0.6322 | 0.1091 | kNN |
| 9th | | Slope* | 0.4209 | 0.3209 | 0.6118 | 0.0998 | kNN |
| 10th | | 8th + 9th | 0.4326 | 0.3824 | 0.5676 | 0.1615 | RF |
| 11th | $NDWT_C$ | Slope | 0.4246 | 0.3215 | 0.6156 | 0.1060 | kNN |
| 12th | | $\angle d_j$ | 0.6757 | 0.7333 | 0.7836 | 0.2419 | RF |
| 13th | | Slope + $\angle d_j$ | 0.6837 | 0.7333 | 0.8007 | 0.2326 | RF |
| 14th | | Slope* | 0.4471 | 0.3392 | 0.6580 | 0.0917 | kNN |
| 15th | | $\overline{j}_{\angle d, p_i, p_{i+1}}$ | 0.4678 | 0.4028 | 0.6176 | 0.1793 | RF |
| 16th | | Slope* + $\overline{j}_{\angle d, p_i, p_{i+1}}$ | 0.4901 | 0.4326 | 0.6487 | 0.1672 | RF |
| 17th | | 13th + 16th | 0.6884 | 0.7222 | 0.8273 | 0.2201 | RF |
| 18th | NDQWT | Slope | 0.4302 | 0.3183 | 0.6282 | 0.1145 | kNN |
| 19th | | $\phi_j + \theta_j + \psi_j$ | 0.7655 | 0.7670 | 0.8486 | 0.5269 | RF |
| 20th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.7565 | 0.7589 | 0.8407 | 0.5216 | RF |
| 21th | | Slope* | 0.4603 | 0.3753 | 0.6720 | 0.0439 | kNN |
| 22th | | $\overline{j}_{\phi, p_i, p_{i+4}} + \overline{j}_{\theta, p_i, p_{i+2}} + \overline{j}_{\psi, p_i, p_{i+1}}$ | 0.4878 | 0.3962 | 0.6819 | 0.1376 | RF |
| 23th | | Slope* + $\overline{j}_{\phi, p_i, p_{i+4}} + \overline{j}_{\theta, p_i, p_{i+2}} + \overline{j}_{\psi, p_i, p_{i+1}}$ | 0.5113 | 0.4278 | 0.7001 | 0.1514 | RF |
| 24th | | 20th + 23th | 0.7684 | 0.7756 | 0.8212 | 0.6027 | GB |
| 25th | | 13th + 20th | 0.7822 | 0.7895 | 0.8407 | 0.5997 | GB |
| 26th | | 13th + 16th + 20th + 23th | 0.7914 | 0.8039 | 0.8491 | 0.6014 | GB |

**Table 16:** Random forest classification results of classifying sounds data.

| Order | Transform | Features | Overall Accuracy rate | Sensitivity Air | Sensitivity Sha | Sensitivity Water |
|---|---|---|---|---|---|---|
| 1st | DWT | Slope | 0.3802 | 0.3378 | 0.4833 | 0.1772 |
| 2nd | $\mathrm{WT_C}$ | Slope | 0.3893 | 0.3662 | 0.4920 | 0.1522 |
| 3rd | | $\angle d_j$ | 0.4010 | 0.3753 | 0.4915 | 0.1973 |
| 4th | | Slope + $\angle d_j$ | 0.4230 | 0.3799 | 0.5629 | 0.1238 |
| 5th | QWT | Slope | 0.3866 | 0.3449 | 0.4977 | 0.1625 |
| 6th | | $\phi_j + \theta_j + \psi_j$ | 0.6654 | 0.6101 | 0.7934 | 0.4130 |
| 7th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.6702 | 0.6209 | 0.7931 | 0.4244 |
| 8th | NDWT | Slope | 0.3961 | 0.3677 | 0.4848 | 0.2117 |
| 9th | | Slope* | 0.3900 | 0.3909 | 0.4662 | 0.1709 |
| 10th | | 8th + 9th | 0.4326 | 0.3824 | 0.5676 | 0.1615 |
| 11th | $\mathrm{NDWT_C}$ | Slope | 0.4019 | 0.3688 | 0.4928 | 0.2147 |
| 12th | | $\angle d_j$ | 0.6757 | 0.7333 | 0.7836 | 0.2419 |
| 13th | | Slope + $\angle d_j$ | 0.6837 | 0.7333 | 0.8007 | 0.2326 |
| 14th | | Slope* | 0.3958 | 0.3630 | 0.4872 | 0.2080 |
| 15th | | $\bar{j}_{\angle d, p_i, p_{i+1}}$ | 0.4678 | 0.4028 | 0.6176 | 0.1793 |
| 16th | | Slope* + $\bar{j}_{\angle d, p_i, p_{i+1}}$ | 0.4901 | 0.4326 | 0.6487 | 0.1672 |
| 17th | | 13th + 16th | 0.6884 | 0.7222 | 0.8273 | 0.2201 |
| 18th | NDQWT | Slope | 0.3740 | 0.3455 | 0.4662 | 0.1716 |
| 19th | | $\phi_j + \theta_j + \psi_j$ | 0.7655 | 0.7670 | 0.8486 | 0.5269 |
| 20th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.7565 | 0.7589 | 0.8407 | 0.5216 |
| 21th | | Slope* | 0.3906 | 0.3701 | 0.4835 | 0.1707 |
| 22th | | $\bar{j}_{\phi, p_i, p_{i+4}} + \bar{j}_{\theta, p_i, p_{i+2}} + \bar{j}_{\psi, p_i, p_{i+1}}$ | 0.4878 | 0.3962 | 0.6819 | 0.1376 |
| 23th | | Slope* + $\bar{j}_{\phi, p_i, p_{i+4}} + \bar{j}_{\theta, p_i, p_{i+2}} + \bar{j}_{\psi, p_i, p_{i+1}}$ | 0.5113 | 0.4278 | 0.7001 | 0.1514 |
| 24th | | 20th + 23th | 0.7651 | 0.7660 | 0.8560 | 0.4993 |
| 25th | | 13th + 20th | 0.7774 | 0.7898 | 0.8648 | 0.5073 |
| 26th | | 13th + 16th + 20th + 23th | 0.7762 | 0.7834 | 0.8686 | 0.4951 |

**Table 17:** Support vector machine classification results of classifying sounds data.

| Order | Transform | Features | Overall Accuracy rate | Sensitivity Air | Sensitivity Sha | Sensitivity Water |
|---|---|---|---|---|---|---|
| 1st | DWT | Slope | 0.3907 | 0.3055 | 0.5382 | 0.1501 |
| 2nd | $\text{WT}_\text{C}$ | Slope | 0.3955 | 0.3317 | 0.5367 | 0.1310 |
| 3rd | | $\angle d_j$ | 0.3886 | 0.3212 | 0.5240 | 0.1578 |
| 4th | | Slope + $\angle d_j$ | 0.3995 | 0.3610 | 0.5081 | 0.1732 |
| 5th | QWT | Slope | 0.3908 | 0.3131 | 0.5333 | 0.1561 |
| 6th | | $\phi_j + \theta_j + \psi_j$ | 0.6610 | 0.6586 | 0.7358 | 0.4572 |
| 7th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.6647 | 0.6464 | 0.7514 | 0.4622 |
| 8th | NDWT | Slope | 0.3858 | 0.3271 | 0.5105 | 0.1457 |
| 9th | | Slope* | 0.3965 | 0.3454 | 0.5279 | 0.1363 |
| 10th | | 8th + 9th | 0.3955 | 0.3670 | 0.5035 | 0.1526 |
| 11th | $\text{NDWT}_\text{C}$ | Slope | 0.3929 | 0.3458 | 0.5104 | 0.1611 |
| 12th | | $\angle d_j$ | 0.6336 | 0.6961 | 0.7333 | 0.2062 |
| 13th | | Slope + $\angle d_j$ | 0.6431 | 0.7043 | 0.7406 | 0.2441 |
| 14th | | Slope* | 0.3850 | 0.3325 | 0.5170 | 0.1179 |
| 15th | | $\overline{j}_{\angle d, p_i, p_{i+1}}$ | 0.4013 | 0.3517 | 0.5255 | 0.1461 |
| 16th | | Slope* + $\overline{j}_{\angle d, p_i, p_{i+1}}$ | 0.4104 | 0.3889 | 0.5190 | 0.1507 |
| 17th | | 13th + 16th | 0.6451 | 0.6908 | 0.7329 | 0.2972 |
| 18th | NDQWT | Slope | 0.3909 | 0.3263 | 0.5222 | 0.1571 |
| 19th | | $\phi_j + \theta_j + \psi_j$ | 0.7327 | 0.7531 | 0.7916 | 0.5261 |
| 20th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.7396 | 0.7482 | 0.8060 | 0.5314 |
| 21th | | Slope* | 0.3908 | 0.3508 | 0.5152 | 0.1269 |
| 22th | | $\overline{j}_{\phi, p_i, p_{i+4}} + \overline{j}_{\theta, p_i, p_{i+2}} + \overline{j}_{\psi, p_i, p_{i+1}}$ | 0.4145 | 0.3789 | 0.5289 | 0.1670 |
| 23th | | Slope* + $\overline{j}_{\phi, p_i, p_{i+4}} + \overline{j}_{\theta, p_i, p_{i+2}} + \overline{j}_{\psi, p_i, p_{i+1}}$ | 0.4231 | 0.3608 | 0.5613 | 0.1667 |
| 24th | | 20th + 23th | 0.7455 | 0.7587 | 0.8080 | 0.5416 |
| 25th | | 13th + 20th | 0.7725 | 0.7859 | 0.8391 | 0.5577 |
| 26th | | 13th + 16th + 20th + 23th | 0.7798 | 0.7936 | 0.8332 | 0.5943 |

**Table 18:** k-nearest neighbors ($k = 7$) classification results of classifying sounds data.

| Order | Transform | Features | Overall Accuracy rate | Sensitivity Air | Sensitivity Sha | Sensitivity Water |
|---|---|---|---|---|---|---|
| 1st | DWT | Slope | 0.4340 | 0.3142 | 0.6506 | 0.0858 |
| 2nd | $\mathrm{WT_C}$ | Slope | 0.4339 | 0.3233 | 0.6544 | 0.0401 |
| 3rd | | $\angle d_j$ | 0.4446 | 0.3637 | 0.6323 | 0.0770 |
| 4th | | Slope + $\angle d_j$ | 0.4222 | 0.3368 | 0.6124 | 0.0696 |
| 5th | QWT | Slope | 0.4336 | 0.3324 | 0.6403 | 0.0763 |
| 6th | | $\phi_j + \theta_j + \psi_j$ | 0.6190 | 0.6093 | 0.7509 | 0.2701 |
| 7th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.6137 | 0.5985 | 0.7443 | 0.2796 |
| 8th | NDWT | Slope | 0.4347 | 0.3189 | 0.6322 | 0.1091 |
| 9th | | Slope* | 0.4209 | 0.3209 | 0.6118 | 0.0998 |
| 10th | | 8th + 9th | 0.4303 | 0.3582 | 0.6022 | 0.1027 |
| 11th | $\mathrm{NDWT_C}$ | Slope | 0.4246 | 0.3215 | 0.6156 | 0.1060 |
| 12th | | $\angle d_j$ | 0.6529 | 0.6528 | 0.8226 | 0.1774 |
| 13th | | Slope + $\angle d_j$ | 0.6594 | 0.6632 | 0.8217 | 0.1873 |
| 14th | | Slope* | 0.4471 | 0.3392 | 0.6580 | 0.0917 |
| 15th | | $\overline{j}_{\angle d, p_i, p_{i+1}}$ | 0.4537 | 0.4106 | 0.6107 | 0.1094 |
| 16th | | Slope* + $\overline{j}_{\angle d, p_i, p_{i+1}}$ | 0.4758 | 0.4437 | 0.6262 | 0.1261 |
| 17th | | 13th + 16th | 0.6581 | 0.6268 | 0.8342 | 0.2160 |
| 18th | NDQWT | Slope | 0.4302 | 0.3183 | 0.6282 | 0.1145 |
| 19th | | $\phi_j + \theta_j + \psi_j$ | 0.7010 | 0.6909 | 0.8553 | 0.2878 |
| 20th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.7033 | 0.6946 | 0.8531 | 0.2950 |
| 21th | | Slope* | 0.4603 | 0.3753 | 0.6720 | 0.0439 |
| 22th | | $\overline{j}_{\phi, p_i, p_{i+4}} + \overline{j}_{\theta, p_i, p_{i+2}} + \overline{j}_{\psi, p_i, p_{i+1}}$ | 0.4576 | 0.4052 | 0.6232 | 0.0903 |
| 23th | | Slope* + $\overline{j}_{\phi, p_i, p_{i+4}} + \overline{j}_{\theta, p_i, p_{i+2}} + \overline{j}_{\psi, p_i, p_{i+1}}$ | 0.4710 | 0.4233 | 0.6362 | 0.1023 |
| 24th | | 20th + 23th | 0.7027 | 0.6916 | 0.8535 | 0.2879 |
| 25th | | 13th + 20th | 0.7190 | 0.7194 | 0.8723 | 0.2797 |
| 26th | | 13th + 16th + 20th + 23th | 0.7212 | 0.7189 | 0.8695 | 0.2910 |

## *B.3 Application 3: Screening Mammograms*

**Table 19:** Comprehensive best classification results of screening mammograms.

| Order | Transform | Features | Accuracy rate | Specificity | Sensitivity | Method |
|---|---|---|---|---|---|---|
| 1st | DWT | Slope | 0.6038 | 0.0326 | 0.9251 | LR |
| 2nd | $WT_C$ | Slope | 0.6078 | 0.0558 | 0.9183 | LR |
| 3rd | | $\angle d_j$ | 0.6006 | 0.0819 | 0.8923 | LR |
| 4th | | Slope + $\angle d_j$ | 0.6088 | 0.1491 | 0.8673 | SVM |
| 5th | QWT | Slope | 0.6376 | 0.0233 | 0.9831 | LR |
| 6th | | $\phi_j + \theta_j + \psi_j$ | 0.6665 | 0.2964 | 0.8746 | SVM |
| 7th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.6608 | 0.2927 | 0.8679 | SVM |
| 8th | NDWT | Slope | 0.6625 | 0.3194 | 0.8555 | LR |
| 9th | | Slope* | 0.6702 | 0.3533 | 0.8485 | LR |
| 10th | | 8th + 9th | 0.6578 | 0.3364 | 0.8386 | LR |
| 11th | $NDWT_C$ | Slope | 0.6546 | 0.2996 | 0.8543 | LR |
| 12th | | $\angle d_j$ | 0.7456 | 0.7038 | 0.7748 | RF |
| 13th | | Slope + $\angle d_j$ | 0.7648 | 0.7277 | 0.7858 | LR |
| 14th | | Slope* | 0.6564 | 0.2901 | 0.8625 | LR |
| 15th | | $\overline{j}_{\angle d,p_i,p_{i+1}}$ | 0.6723 | 0.4135 | 0.8260 | RF |
| 16th | | Slope* + $\overline{j}_{\angle d,p_i,p_{i+1}}$ | 0.6705 | 0.4020 | 0.8303 | RF |
| 17th | | 13th + 16th | 0.7472 | 0.6915 | 0.7851 | RF |
| 18th | NDQWT | Slope | 0.6561 | 0.3019 | 0.8553 | LR |
| 19th | | $\phi_j + \theta_j + \psi_j$ | 0.7678 | 0.6666 | 0.8309 | RF |
| 20th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.7651 | 0.6639 | 0.8296 | RF |
| 21th | | Slope* | 0.6532 | 0.2810 | 0.8626 | LR |
| 22th | | $\overline{j}_{\phi,p_i,p_{i+4}} + \overline{j}_{\theta,p_i,p_{i+2}} + \overline{j}_{\psi,p_i,p_{i+1}}$ | 0.6375 | 0.3397 | 0.8152 | RF |
| 23th | | Slope* + $\overline{j}_{\phi,p_i,p_{i+4}} + \overline{j}_{\theta,p_i,p_{i+2}} + \overline{j}_{\psi,p_i,p_{i+1}}$ | 0.6499 | 0.3509 | 0.8281 | RF |
| 24th | | 20th + 23th | 0.7413 | 0.6132 | 0.8218 | RF |
| 25th | | 13th + 20th | 0.7830 | 0.7442 | 0.8116 | RF |
| 26th | | 13th + 16th + 20th + 23th | 0.7761 | 0.7397 | 0.8012 | RF |

**Table 20:** Gradient boosting classification results of screening mammograms.

| Order | Transform | Features | Accuracy rate | Specificity | Sensitivity |
|---|---|---|---|---|---|
| 1st | DWT | Slope | 0.5560 | 0.3714 | 0.6598 |
| 2nd | $\text{WT}_\text{C}$ | Slope | 0.5036 | 0.3098 | 0.6126 |
| 3rd | | $\angle d_j$ | 0.5274 | 0.3330 | 0.6367 |
| 4th | | Slope + $\angle d_j$ | 0.4855 | 0.3061 | 0.5864 |
| 5th | QWT | Slope | 0.5679 | 0.3843 | 0.6712 |
| 6th | | $\phi_j + \theta_j + \psi_j$ | 0.5649 | 0.3340 | 0.6948 |
| 7th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.5636 | 0.3209 | 0.7001 |
| 8th | NDWT | Slope | 0.5697 | 0.3963 | 0.6672 |
| 9th | | Slope* | 0.5327 | 0.3464 | 0.6375 |
| 10th | | 8th + 9th | 0.5267 | 0.3070 | 0.6503 |
| 11th | $\text{NDWT}_\text{C}$ | Slope | 0.5623 | 0.3831 | 0.6631 |
| 12th | | $\angle d_j$ | 0.7366 | 0.6579 | 0.7809 |
| 13th | | Slope + $\angle d_j$ | 0.7290 | 0.6523 | 0.7721 |
| 14th | | Slope* | 0.5978 | 0.4377 | 0.6879 |
| 15th | | $\overline{j}_{\angle d, p_i, p_{i+1}}$ | 0.6470 | 0.4482 | 0.7588 |
| 16th | | Slope* + $\overline{j}_{\angle d, p_i, p_{i+1}}$ | 0.6528 | 0.4481 | 0.7679 |
| 17th | | 13th + 16th | 0.7266 | 0.6233 | 0.7847 |
| 18th | NDQWT | Slope | 0.5375 | 0.3397 | 0.6488 |
| 19th | | $\phi_j + \theta_j + \psi_j$ | 0.7337 | 0.6254 | 0.7946 |
| 20th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.7301 | 0.6037 | 0.8012 |
| 21th | | Slope* | 0.5790 | 0.3054 | 0.7329 |
| 22th | | $\overline{j}_{\phi, p_i, p_{i+4}} + \overline{j}_{\theta, p_i, p_{i+2}} + \overline{j}_{\psi, p_i, p_{i+1}}$ | 0.6256 | 0.4052 | 0.7496 |
| 23th | | Slope* + $\overline{j}_{\phi, p_i, p_{i+4}} + \overline{j}_{\theta, p_i, p_{i+2}} + \overline{j}_{\psi, p_i, p_{i+1}}$ | 0.6198 | 0.4039 | 0.7413 |
| 24th | | 20th + 23th | 0.7111 | 0.5886 | 0.7800 |
| 25th | | 13th + 20th | 0.7452 | 0.6707 | 0.7872 |
| 26th | | 13th + 16th + 20th + 23th | 0.7662 | 0.7084 | 0.7987 |

**Table 21:** Support vector machine classification results of screening mammograms.

| Order | Transform | Features | Accuracy rate | Specificity | Sensitivity |
|---|---|---|---|---|---|
| 1st | DWT | Slope | 0.5856 | 0.3094 | 0.7409 |
| 2nd | $\text{WT}_\text{C}$ | Slope | 0.5700 | 0.3747 | 0.6798 |
| 3rd | | $\angle d_j$ | 0.5690 | 0.3607 | 0.6863 |
| 4th | | Slope + $\angle d_j$ | 0.6088 | 0.1491 | 0.8673 |
| 5th | QWT | Slope | 0.5519 | 0.4197 | 0.6263 |
| 6th | | $\phi_j + \theta_j + \psi_j$ | 0.6665 | 0.2964 | 0.8746 |
| 7th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.6608 | 0.2927 | 0.8679 |
| 8th | NDWT | Slope | 0.5409 | 0.3117 | 0.6699 |
| 9th | | Slope* | 0.5752 | 0.3446 | 0.7049 |
| 10th | | 8th + 9th | 0.5498 | 0.2950 | 0.6932 |
| 11th | $\text{NDWT}_\text{C}$ | Slope | 0.5834 | 0.3336 | 0.7239 |
| 12th | | $\angle d_j$ | 0.7418 | 0.6981 | 0.7664 |
| 13th | | Slope + $\angle d_j$ | 0.7334 | 0.6021 | 0.8073 |
| 14th | | Slope* | 0.5888 | 0.3396 | 0.7290 |
| 15th | | $\overline{j}_{\angle d,p_i,p_{i+1}}$ | 0.5813 | 0.4393 | 0.6611 |
| 16th | | Slope* + $\overline{j}_{\angle d,p_i,p_{i+1}}$ | 0.6630 | 0.5033 | 0.7528 |
| 17th | | 13th + 16th | 0.7408 | 0.6351 | 0.8002 |
| 18th | NDQWT | Slope | 0.5761 | 0.3278 | 0.7158 |
| 19th | | $\phi_j + \theta_j + \psi_j$ | 0.7170 | 0.6448 | 0.7576 |
| 20th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.7235 | 0.6080 | 0.7884 |
| 21th | | Slope* | 0.5660 | 0.3313 | 0.6981 |
| 22th | | $\overline{j}_{\phi,p_i,p_{i+4}} + \overline{j}_{\theta,p_i,p_{i+2}} + \overline{j}_{\psi,p_i,p_{i+1}}$ | 0.5827 | 0.2011 | 0.7973 |
| 23th | | Slope* + $\overline{j}_{\phi,p_i,p_{i+4}} + \overline{j}_{\theta,p_i,p_{i+2}} + \overline{j}_{\psi,p_i,p_{i+1}}$ | 0.5898 | 0.2861 | 0.7606 |
| 24th | | 20th + 23th | 0.7138 | 0.5699 | 0.7948 |
| 25th | | 13th + 20th | 0.7412 | 0.6347 | 0.8012 |
| 26th | | 13th + 16th + 20th + 23th | 0.7576 | 0.6840 | 0.7991 |

**Table 22:** k-nearest neighbors ($k = 10$) classification results of screening mammograms.

| Order | Transform | Features | Accuracy rate | Specificity | Sensitivity |
|---|---|---|---|---|---|
| 1st | DWT | Slope | 0.4990 | 0.4965 | 0.5003 |
| 2nd | $\text{WT}_{\text{C}}$ | Slope | 0.5012 | 0.5023 | 0.5006 |
| 3rd | | $\angle d_j$ | 0.4994 | 0.5009 | 0.4986 |
| 4th | | Slope + $\angle d_j$ | 0.4994 | 0.5000 | 0.4991 |
| 5th | QWT | Slope | 0.5009 | 0.5012 | 0.5007 |
| 6th | | $\phi_j + \theta_j + \psi_j$ | 0.4546 | 0.7535 | 0.3340 |
| 7th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.4585 | 0.7362 | 0.3494 |
| 8th | NDWT | Slope | 0.5001 | 0.4987 | 0.5008 |
| 9th | | Slope* | 0.5031 | 0.5019 | 0.5038 |
| 10th | | 8th + 9th | 0.4958 | 0.4957 | 0.4959 |
| 11th | $\text{NDWT}_{\text{C}}$ | Slope | 0.5007 | 0.5000 | 0.5010 |
| 12th | | $\angle d_j$ | 0.5005 | 0.5001 | 0.5007 |
| 13th | | Slope + $\angle d_j$ | 0.5013 | 0.4989 | 0.5027 |
| 14th | | Slope* | 0.5005 | 0.5015 | 0.5000 |
| 15th | | $\bar{\bar{j}}_{\angle d, p_i, p_{i+1}}$ | 0.5001 | 0.4987 | 0.5008 |
| 16th | | Slope* + $\bar{\bar{j}}_{\angle d, p_i, p_{i+1}}$ | 0.5001 | 0.5010 | 0.4996 |
| 17th | | 13th + 16th | 0.6983 | 0.7709 | 0.6605 |
| 18th | NDQWT | Slope | 0.4987 | 0.4986 | 0.4988 |
| 19th | | $\phi_j + \theta_j + \psi_j$ | 0.6098 | 0.8429 | 0.4883 |
| 20th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.6206 | 0.8201 | 0.5167 |
| 21th | | Slope* | 0.4972 | 0.4970 | 0.4973 |
| 22th | | $\bar{\bar{j}}_{\phi, p_i, p_{i+4}} + \bar{\bar{j}}_{\theta, p_i, p_{i+2}} + \bar{\bar{j}}_{\psi, p_i, p_{i+1}}$ | 0.5490 | 0.4719 | 0.5896 |
| 23th | | Slope* + $\bar{\bar{j}}_{\phi, p_i, p_{i+4}} + \bar{\bar{j}}_{\theta, p_i, p_{i+2}} + \bar{\bar{j}}_{\psi, p_i, p_{i+1}}$ | 0.5552 | 0.4696 | 0.5999 |
| 24th | | 20th + 23th | 0.6650 | 0.7930 | 0.6061 |
| 25th | | 13th + 20th | 0.7064 | 0.8586 | 0.6364 |
| 26th | | 13th + 16th + 20th + 23th | 0.7062 | 0.8646 | 0.6333 |

**Table 23:** Logistic regression classification results of screening mammograms.

| Order | Transform | Features | Accuracy rate | Specificity | Sensitivity |
|---|---|---|---|---|---|
| 1st | DWT | Slope | 0.6038 | 0.0326 | 0.9251 |
| 2nd | $\text{WT}_\text{C}$ | Slope | 0.6078 | 0.0558 | 0.9183 |
| 3rd | | $\angle d_j$ | 0.6006 | 0.0819 | 0.8923 |
| 4th | | Slope + $\angle d_j$ | 0.5929 | 0.1896 | 0.8198 |
| 5th | QWT | Slope | 0.6376 | 0.0233 | 0.9831 |
| 6th | | $\phi_j + \theta_j + \psi_j$ | 0.6231 | 0.5296 | 0.6758 |
| 7th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.6144 | 0.5190 | 0.6680 |
| 8th | NDWT | Slope | 0.6625 | 0.3194 | 0.8555 |
| 9th | | Slope* | 0.6702 | 0.3533 | 0.8485 |
| 10th | | 8th + 9th | 0.6578 | 0.3364 | 0.8386 |
| 11th | $\text{NDWT}_\text{C}$ | Slope | 0.6546 | 0.2996 | 0.8543 |
| 12th | | $\angle d_j$ | 0.7667 | 0.7416 | 0.7809 |
| 13th | | Slope + $\angle d_j$ | 0.7648 | 0.7277 | 0.7858 |
| 14th | | Slope* | 0.6564 | 0.2901 | 0.8625 |
| 15th | | $\overline{j}_{\angle d, p_i, p_{i+1}}$ | 0.6535 | 0.4927 | 0.7439 |
| 16th | | Slope* + $\overline{j}_{\angle d, p_i, p_{i+1}}$ | 0.6696 | 0.4844 | 0.7737 |
| 17th | | 13th + 16th | 0.7337 | 0.6664 | 0.7716 |
| 18th | NDQWT | Slope | 0.6561 | 0.3019 | 0.8553 |
| 19th | | $\phi_j + \theta_j + \psi_j$ | 0.7450 | 0.6848 | 0.7789 |
| 20th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.7340 | 0.6746 | 0.7674 |
| 21th | | Slope* | 0.6532 | 0.2810 | 0.8626 |
| 22th | | $\overline{j}_{\phi, p_i, p_{i+4}} + \overline{j}_{\theta, p_i, p_{i+2}} + \overline{j}_{\psi, p_i, p_{i+1}}$ | 0.5773 | 0.3676 | 0.6953 |
| 23th | | Slope* + $\overline{j}_{\phi, p_i, p_{i+4}} + \overline{j}_{\theta, p_i, p_{i+2}} + \overline{j}_{\psi, p_i, p_{i+1}}$ | 0.5720 | 0.3476 | 0.6983 |
| 24th | | 20th + 23th | 0.7070 | 0.6232 | 0.7541 |
| 25th | | 13th + 20th | 0.7074 | 0.6229 | 0.7550 |
| 26th | | 13th + 16th + 20th + 23th | 0.7368 | 0.6974 | 0.7589 |

## *B.4 Application 4: Seam Detection in Steel Rolling Process*

**Table 24:** Comprehensive best classification results of seam detection in steel rolling process.

| Order | Transform | Features | Accuracy rate | Specificity | Sensitivity | Method |
|---|---|---|---|---|---|---|
| 1st | DWT | Slope | 0.7480 | 0.9973 | 0.0000 | LR |
| 2nd | $\text{WT}_\text{C}$ | Slope | 0.8133 | 0.9806 | 0.3114 | LR |
| 3rd | | $\angle d_j$ | 0.7499 | 0.9998 | 0.0000 | LR |
| 4th | | Slope + $\angle d_j$ | 0.8120 | 0.9779 | 0.3144 | LR |
| 5th | QWT | Slope | 0.7852 | 0.9661 | 0.2422 | SVM |
| 6th | | $\phi_j + \theta_j + \psi_j$ | 0.8359 | 0.9749 | 0.4188 | RF |
| 7th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.8447 | 0.9823 | 0.4320 | RF |
| 8th | NDWT | Slope | 0.8432 | 0.9666 | 0.4730 | LR |
| 9th | | Slope* | 0.8237 | 1.0000 | 0.2946 | LR |
| 10th | | 8th + 9th | 0.8427 | 0.9701 | 0.4602 | LR |
| 11th | $\text{NDWT}_\text{C}$ | Slope | 0.8371 | 0.9679 | 0.4448 | LR |
| 12th | | $\angle d_j$ | 0.8600 | 0.9664 | 0.5408 | RF |
| 13th | | Slope + $\angle d_j$ | 0.8938 | 0.9717 | 0.6598 | RF |
| 14th | | Slope* | 0.7674 | 1.0000 | 0.0694 | LR |
| 15th | | $\bar{j}_{\angle d,p_i,p_{i+1}}$ | 0.9265 | 0.9883 | 0.7412 | LR |
| 16th | | Slope* + $\bar{j}_{\angle d,p_i,p_{i+1}}$ | 0.9195 | 0.9831 | 0.7286 | LR |
| 17th | | 13th + 16th | 0.9263 | 0.9809 | 0.7626 | RF |
| 18th | NDQWT | Slope | 0.8406 | 0.9751 | 0.4372 | LR |
| 19th | | $\phi_j + \theta_j + \psi_j$ | 0.9139 | 0.9868 | 0.6950 | RF |
| 20th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.9220 | 0.9831 | 0.7384 | RF |
| 21th | | Slope* | 0.8214 | 0.8985 | 0.5515 | RF |
| 22th | | $\bar{j}_{\phi,p_i,p_{i+4}} + \bar{j}_{\theta,p_i,p_{i+2}} + \bar{j}_{\psi,p_i,p_{i+1}}$ | 0.8911 | 0.9507 | 0.6825 | RF |
| 23th | | Slope* + $\bar{j}_{\phi,p_i,p_{i+4}} + \bar{j}_{\theta,p_i,p_{i+2}} + \bar{j}_{\psi,p_i,p_{i+1}}$ | 0.9230 | 0.9719 | 0.7521 | RF |
| 24th | | 20th + 23th | 0.9384 | 0.9873 | 0.7674 | RF |
| 25th | | 13th + 20th | 0.9310 | 0.9819 | 0.7529 | RF |
| 26th | | 13th + 16th + 20th + 23th | 0.9523 | 0.9999 | 0.8029 | RF |

**Table 25:** Gradient boosting classification results of seam detection in steel rolling process.

| Order | Transform | Features | Accuracy rate | Specificity | Sensitivity |
|---|---|---|---|---|---|
| 1st | DWT | Slope | 0.6168 | 0.7551 | 0.2018 |
| 2nd | $WT_C$ | Slope | 0.7410 | 0.8308 | 0.4714 |
| 3rd | | $\angle d_j$ | 0.6241 | 0.7274 | 0.3140 |
| 4th | | Slope + $\angle d_j$ | 0.7632 | 0.8574 | 0.4806 |
| 5th | QWT | Slope | 0.7040 | 0.7976 | 0.4232 |
| 6th | | $\phi_j + \theta_j + \psi_j$ | 0.8022 | 0.9194 | 0.4504 |
| 7th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.8104 | 0.9197 | 0.4826 |
| 8th | NDWT | Slope | 0.6675 | 0.7593 | 0.3922 |
| 9th | | Slope* | 0.7558 | 0.8517 | 0.4680 |
| 10th | | 8th + 9th | 0.7711 | 0.8581 | 0.5098 |
| 11th | $NDWT_C$ | Slope | 0.7873 | 0.8803 | 0.5080 |
| 12th | | $\angle d_j$ | 0.8460 | 0.9237 | 0.6128 |
| 13th | | Slope + $\angle d_j$ | 0.8731 | 0.9387 | 0.6764 |
| 14th | | Slope* | 0.7487 | 0.8465 | 0.4550 |
| 15th | | $\overline{\overline{j}}_{\angle d,p_i,p_{i+1}}$ | 0.8881 | 0.9456 | 0.7156 |
| 16th | | Slope* + $\overline{\overline{j}}_{\angle d,p_i,p_{i+1}}$ | 0.8787 | 0.9388 | 0.6982 |
| 17th | | 13th + 16th | 0.9001 | 0.9523 | 0.7436 |
| 18th | NDQWT | Slope | 0.7600 | 0.8379 | 0.5262 |
| 19th | | $\phi_j + \theta_j + \psi_j$ | 0.8842 | 0.9457 | 0.6996 |
| 20th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.8728 | 0.9347 | 0.6870 |
| 21th | | Slope* | 0.8001 | 0.8738 | 0.5790 |
| 22th | | $\overline{\overline{j}}_{\phi,p_i,p_{i+4}} + \overline{\overline{j}}_{\theta,p_i,p_{i+2}} + \overline{\overline{j}}_{\psi,p_i,p_{i+1}}$ | 0.8205 | 0.8984 | 0.5868 |
| 23th | | Slope* + $\overline{\overline{j}}_{\phi,p_i,p_{i+4}} + \overline{\overline{j}}_{\theta,p_i,p_{i+2}} + \overline{\overline{j}}_{\psi,p_i,p_{i+1}}$ | 0.8418 | 0.9093 | 0.6392 |
| 24th | | 20th + 23th | 0.8733 | 0.9429 | 0.6644 |
| 25th | | 13th + 20th | 0.8646 | 0.9315 | 0.6640 |
| 26th | | 13th + 16th + 20th + 23th | 0.8991 | 0.9599 | 0.7168 |

**Table 26:** Support vector machine classification results of seam detection in steel rolling process.

| Order | Transform | Features | Accuracy rate | Specificity | Sensitivity |
|---|---|---|---|---|---|
| 1st | DWT | Slope | 0.6842 | 0.8905 | 0.0652 |
| 2nd | $\text{WT}_\text{C}$ | Slope | 0.6743 | 0.7803 | 0.3562 |
| 3rd | | $\angle d_j$ | 0.6847 | 0.8863 | 0.0796 |
| 4th | | Slope + $\angle d_j$ | 0.6962 | 0.8409 | 0.2618 |
| 5th | QWT | Slope | 0.7852 | 0.9661 | 0.2422 |
| 6th | | $\phi_j + \theta_j + \psi_j$ | 0.8273 | 0.9740 | 0.3872 |
| 7th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.8321 | 0.9773 | 0.3962 |
| 8th | NDWT | Slope | 0.6842 | 0.8084 | 0.3116 |
| 9th | | Slope* | 0.6053 | 0.7897 | 0.0522 |
| 10th | | 8th + 9th | 0.6985 | 0.8339 | 0.2922 |
| 11th | $\text{NDWT}_\text{C}$ | Slope | 0.7138 | 0.8033 | 0.4454 |
| 12th | | $\angle d_j$ | 0.7531 | 0.9934 | 0.0322 |
| 13th | | Slope + $\angle d_j$ | 0.8171 | 0.9659 | 0.3706 |
| 14th | | Slope* | 0.7311 | 0.9483 | 0.0792 |
| 15th | | $\overline{\overline{j}}_{\angle d,p_i,p_{i+1}}$ | 0.8689 | 0.9885 | 0.5100 |
| 16th | | Slope* + $\overline{\overline{j}}_{\angle d,p_i,p_{i+1}}$ | 0.8388 | 0.9715 | 0.4404 |
| 17th | | 13th + 16th | 0.8791 | 0.9834 | 0.5662 |
| 18th | NDQWT | Slope | 0.6745 | 0.7724 | 0.3808 |
| 19th | | $\phi_j + \theta_j + \psi_j$ | 0.8301 | 0.9791 | 0.3832 |
| 20th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.8446 | 0.9689 | 0.4716 |
| 21th | | Slope* | 0.7902 | 0.9983 | 0.1658 |
| 22th | | $\overline{\overline{j}}_{\phi,p_i,p_{i+4}} + \overline{\overline{j}}_{\theta,p_i,p_{i+2}} + \overline{\overline{j}}_{\psi,p_i,p_{i+1}}$ | 0.8571 | 0.9442 | 0.5956 |
| 23th | | Slope* + $\overline{\overline{j}}_{\phi,p_i,p_{i+4}} + \overline{\overline{j}}_{\theta,p_i,p_{i+2}} + \overline{\overline{j}}_{\psi,p_i,p_{i+1}}$ | 0.8614 | 0.9649 | 0.5506 |
| 24th | | 20th + 23th | 0.8766 | 0.9903 | 0.5354 |
| 25th | | 13th + 20th | 0.8422 | 0.9711 | 0.4556 |
| 26th | | 13th + 16th + 20th + 23th | 0.8892 | 0.9993 | 0.5588 |

**Table 27:** k-nearest neighbors ($k = 3$) classification results of seam detection in steel rolling process.

| Order | Transform | Features | Accuracy rate | Specificity | Sensitivity |
|---|---|---|---|---|---|
| 1st | DWT | Slope | 0.4926 | 0.4799 | 0.5320 |
| 2nd | $\text{WT}_\text{C}$ | Slope | 0.6235 | 0.6602 | 0.5106 |
| 3rd | | $\angle d_j$ | 0.4989 | 0.5036 | 0.4844 |
| 4th | | Slope + $\angle d_j$ | 0.5924 | 0.6136 | 0.5259 |
| 5th | QWT | Slope | 0.6027 | 0.6727 | 0.3867 |
| 6th | | $\phi_j + \theta_j + \psi_j$ | 0.6588 | 0.6928 | 0.5105 |
| 7th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.6827 | 0.7181 | 0.5224 |
| 8th | NDWT | Slope | 0.6841 | 0.7282 | 0.5480 |
| 9th | | Slope* | 0.6259 | 0.6954 | 0.4118 |
| 10th | | 8th + 9th | 0.6662 | 0.7108 | 0.5255 |
| 11th | $\text{NDWT}_\text{C}$ | Slope | 0.7087 | 0.7527 | 0.5730 |
| 12th | | $\angle d_j$ | 0.6874 | 0.7495 | 0.4749 |
| 13th | | Slope + $\angle d_j$ | 0.7429 | 0.7943 | 0.5626 |
| 14th | | Slope* | 0.5847 | 0.6491 | 0.3862 |
| 15th | | $\bar{\bar{j}}_{\angle d, p_i, p_{i+1}}$ | 0.7770 | 0.8084 | 0.6730 |
| 16th | | Slope* + $\bar{\bar{j}}_{\angle d, p_i, p_{i+1}}$ | 0.7655 | 0.8077 | 0.6255 |
| 17th | | 13th + 16th | 0.8229 | 0.8596 | 0.6799 |
| 18th | NDQWT | Slope | 0.6725 | 0.7116 | 0.5520 |
| 19th | | $\phi_j + \theta_j + \psi_j$ | 0.7821 | 0.8127 | 0.6465 |
| 20th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.7875 | 0.8133 | 0.6710 |
| 21th | | Slope* | 0.6485 | 0.7314 | 0.3928 |
| 22th | | $\bar{\bar{j}}_{\phi, p_i, p_{i+4}} + \bar{\bar{j}}_{\theta, p_i, p_{i+2}} + \bar{\bar{j}}_{\psi, p_i, p_{i+1}}$ | 0.8095 | 0.8277 | 0.7397 |
| 23th | | Slope* + $\bar{\bar{j}}_{\phi, p_i, p_{i+4}} + \bar{\bar{j}}_{\theta, p_i, p_{i+2}} + \bar{\bar{j}}_{\psi, p_i, p_{i+1}}$ | 0.8230 | 0.8442 | 0.7394 |
| 24th | | 20th + 23th | 0.8626 | 0.8892 | 0.7034 |
| 25th | | 13th + 20th | 0.8049 | 0.8266 | 0.6920 |
| 26th | | 13th + 16th + 20th + 23th | 0.9027 | 0.9235 | 0.7373 |

**Table 28:** Logistic regression classification results of seam detection in steel rolling process.

| Order | Transform | Features | Accuracy rate | Specificity | Sensitivity |
|---|---|---|---|---|---|
| 1st | DWT | Slope | 0.7480 | 0.9973 | 0.0000 |
| 2nd | $\mathrm{WT_C}$ | Slope | 0.8133 | 0.9806 | 0.3114 |
| 3rd | | $\angle d_j$ | 0.7499 | 0.9998 | 0.0000 |
| 4th | | Slope + $\angle d_j$ | 0.8120 | 0.9779 | 0.3144 |
| 5th | QWT | Slope | 0.7504 | 1.0000 | 0.0016 |
| 6th | | $\phi_j + \theta_j + \psi_j$ | 0.7760 | 0.8677 | 0.5008 |
| 7th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.7848 | 0.8686 | 0.5334 |
| 8th | NDWT | Slope | 0.8432 | 0.9666 | 0.4730 |
| 9th | | Slope* | 0.8237 | 1.0000 | 0.2946 |
| 10th | | 8th + 9th | 0.8427 | 0.9701 | 0.4602 |
| 11th | $\mathrm{NDWT_C}$ | Slope | 0.8371 | 0.9679 | 0.4448 |
| 12th | | $\angle d_j$ | 0.7478 | 0.9768 | 0.0608 |
| 13th | | Slope + $\angle d_j$ | 0.8342 | 0.9349 | 0.5320 |
| 14th | | Slope* | 0.7674 | 1.0000 | 0.0694 |
| 15th | | $\overline{\overline{j}}_{\angle d,p_i,p_{i+1}}$ | 0.9265 | 0.9883 | 0.7412 |
| 16th | | Slope* + $\overline{\overline{j}}_{\angle d,p_i,p_{i+1}}$ | 0.9195 | 0.9831 | 0.7286 |
| 17th | | 13th + 16th | 0.8790 | 0.9291 | 0.7288 |
| 18th | NDQWT | Slope | 0.8406 | 0.9751 | 0.4372 |
| 19th | | $\phi_j + \theta_j + \psi_j$ | 0.7840 | 0.8261 | 0.6574 |
| 20th | | Slope + $\phi_j + \theta_j + \psi_j$ | 0.7801 | 0.8199 | 0.6604 |
| 21th | | Slope* | 0.7579 | 1.0000 | 0.0314 |
| 22th | | $\overline{\overline{j}}_{\phi,p_i,p_{i+4}} + \overline{\overline{j}}_{\theta,p_i,p_{i+2}} + \overline{\overline{j}}_{\psi,p_i,p_{i+1}}$ | 0.8480 | 0.8968 | 0.7014 |
| 23th | | Slope* + $\overline{\overline{j}}_{\phi,p_i,p_{i+4}} + \overline{\overline{j}}_{\theta,p_i,p_{i+2}} + \overline{\overline{j}}_{\psi,p_i,p_{i+1}}$ | 0.8383 | 0.8921 | 0.6766 |
| 24th | | 20th + 23th | 0.7817 | 0.8327 | 0.6284 |
| 25th | | 13th + 20th | 0.7378 | 0.7791 | 0.6138 |
| 26th | | 13th + 16th + 20th + 23th | 0.8010 | 0.8370 | 0.6928 |

# APPENDIX C

# BOX PLOTS

## *C.1 Application 1: Classifying Pupillary Signal Data*

### C.1.1 Non-decimated Complex Wavelet Transform

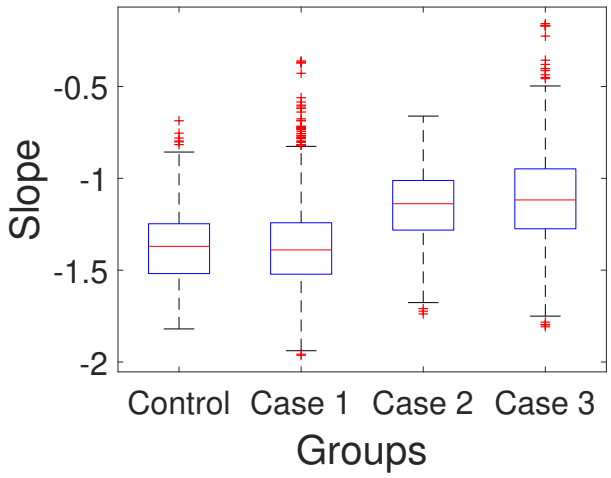



**Figure 29:** Application 1: Box plot of spectral slope obtained by NDCWT.

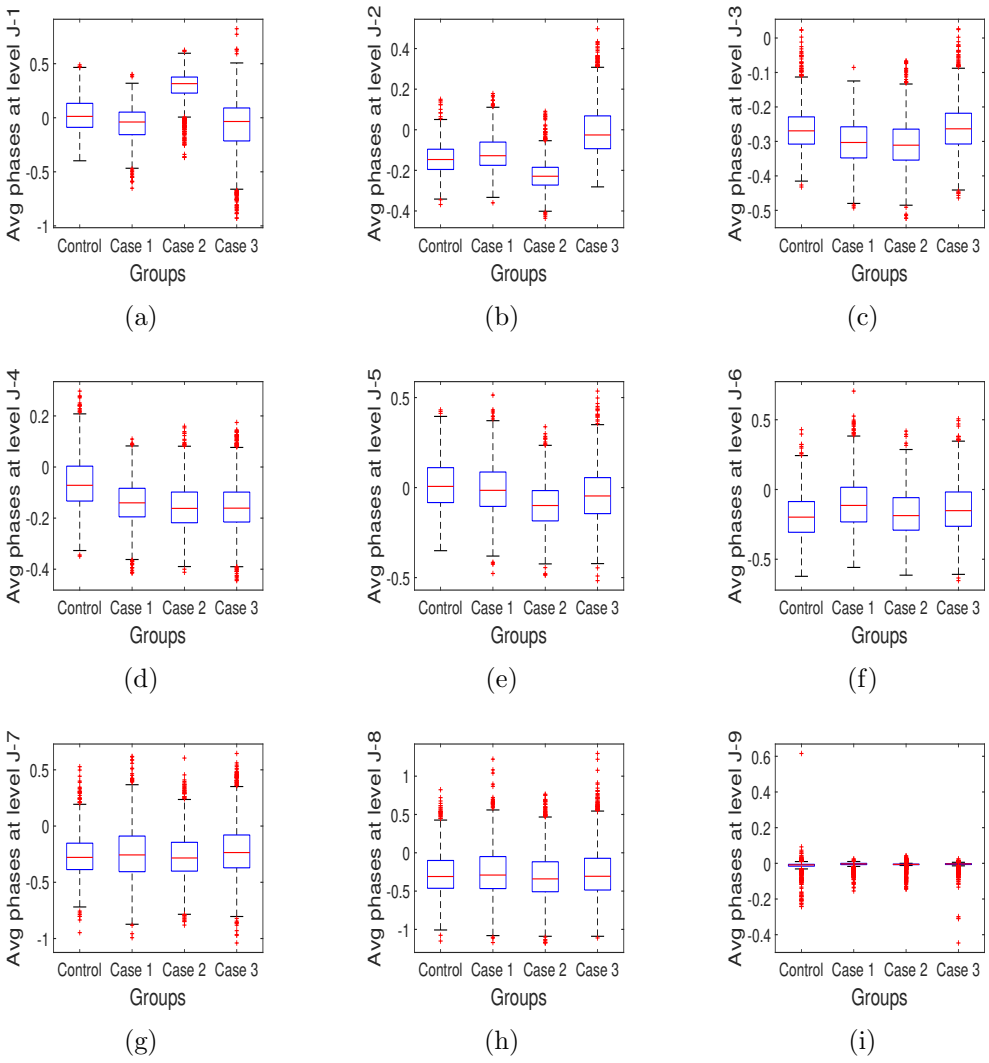



**Figure 30:** Application 1: Box plots of phase averages at all multiresolution levels obtained by NDCWT.

### C.1.2 Non-decimated Quaternion Wavelet Transform

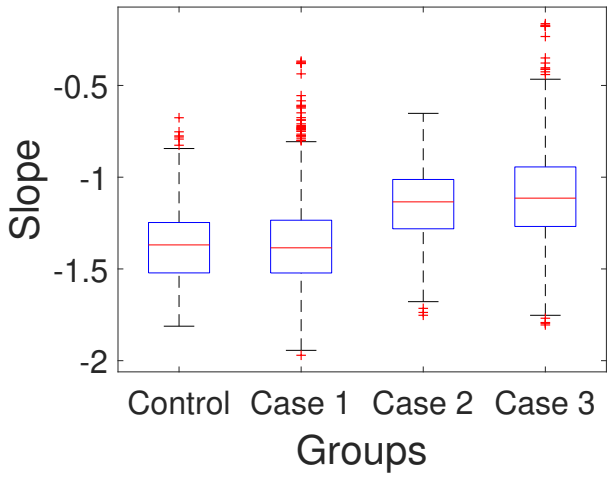



**Figure 31:** Application 1: Box plot of spectral slope obtained by NDQWT.

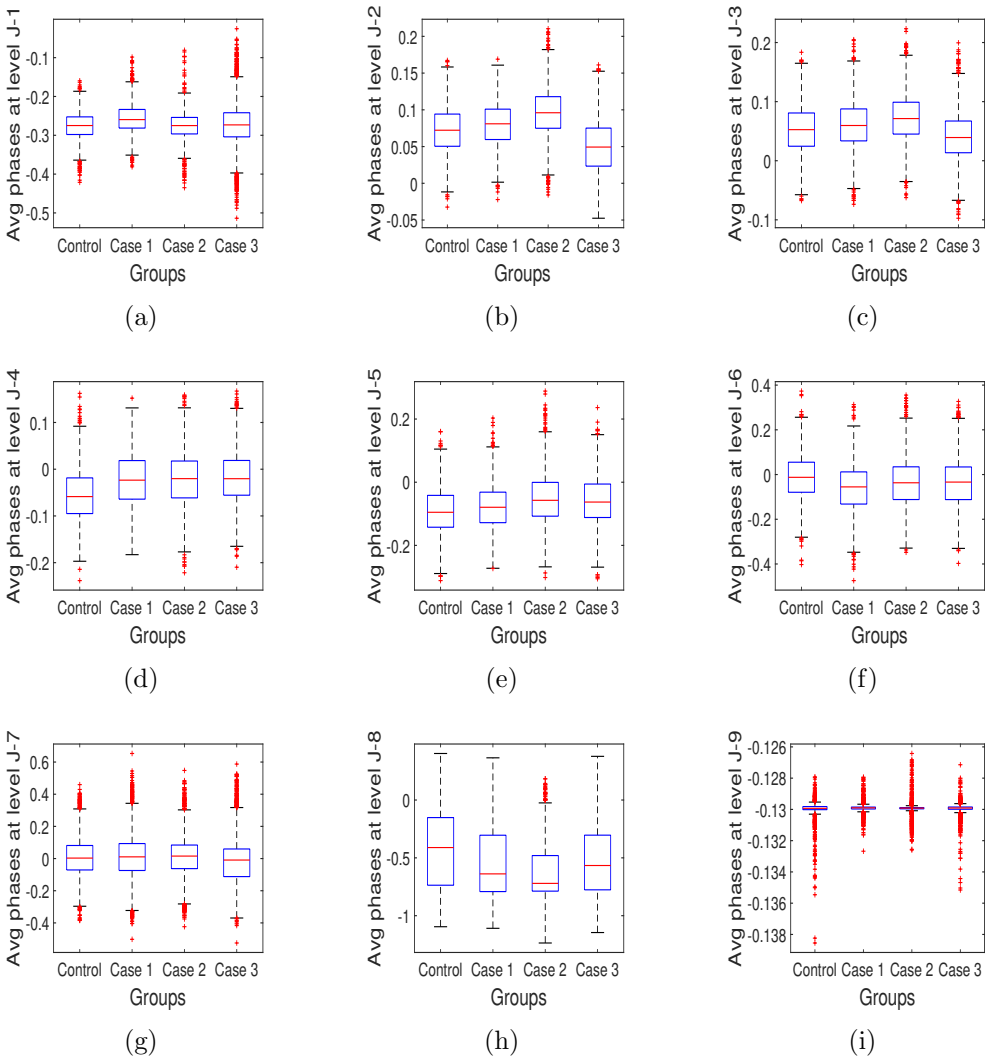



**Figure 32:** Application 1: Box plots of averages of phase $\phi$ at all multiresolution levels obtained by NDQWT.

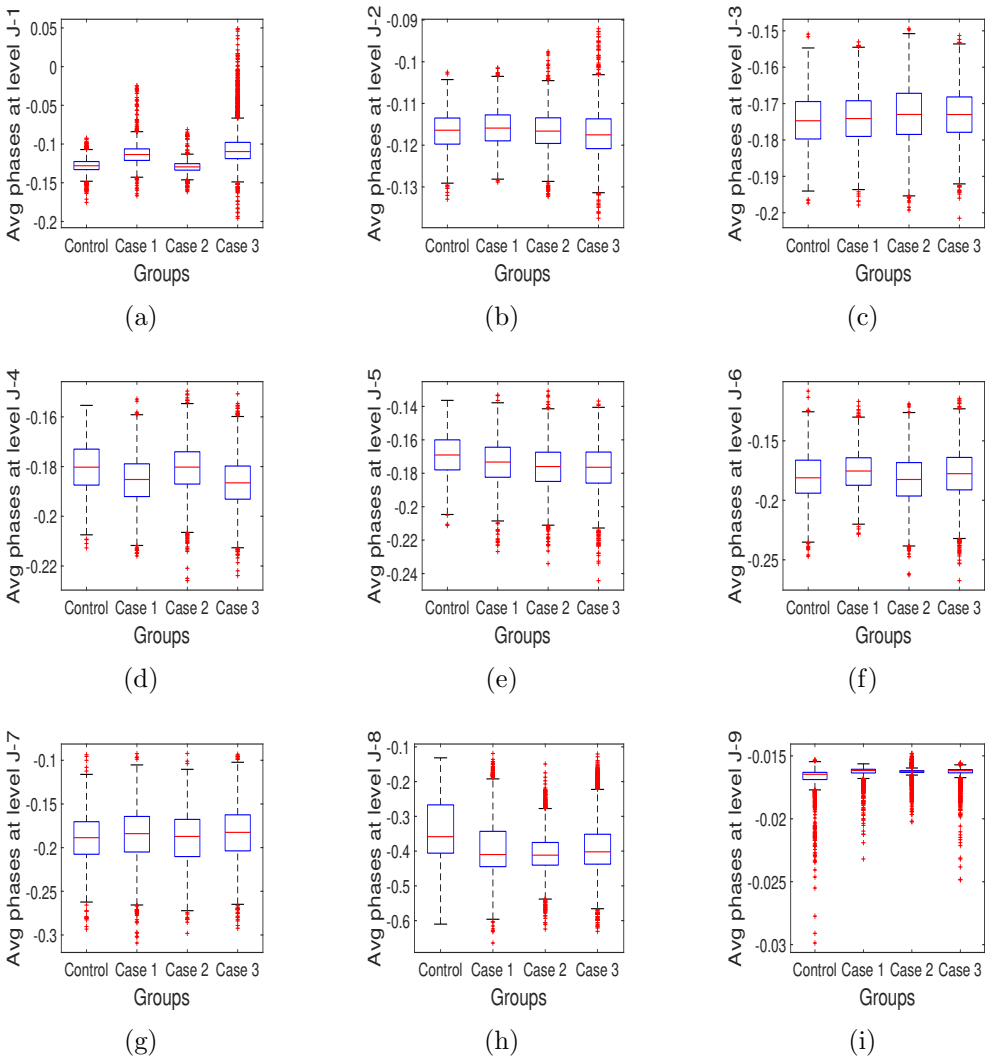



**Figure 33:** Application 1: Box plots of averages of phase $\theta$ at all multiresolution levels obtained by NDQWT.

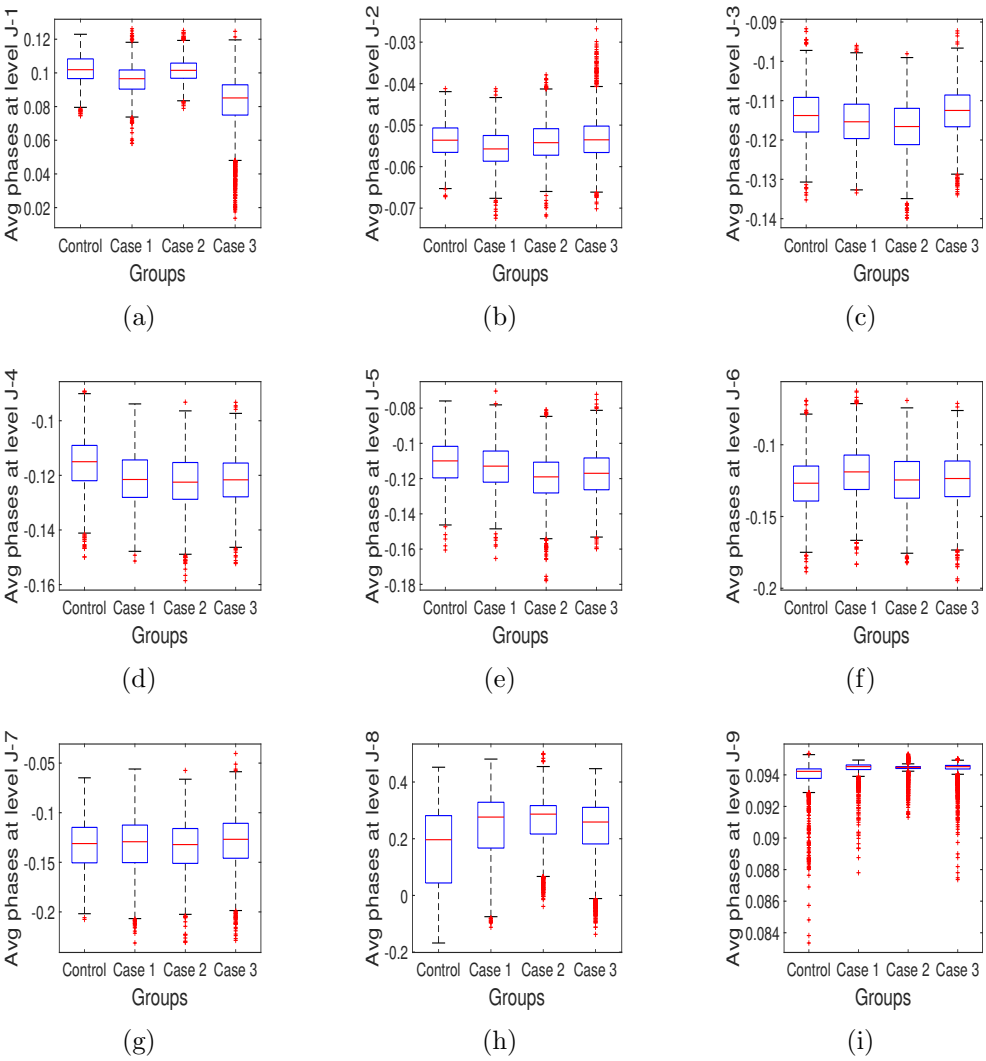
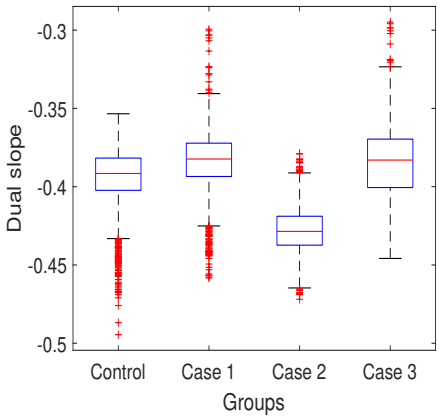
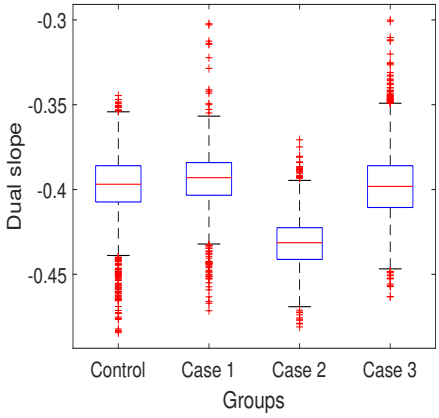



**Figure 34:** Application 1: Box plots of averages of phase $\psi$ at all multiresolution levels obtained by NDQWT.

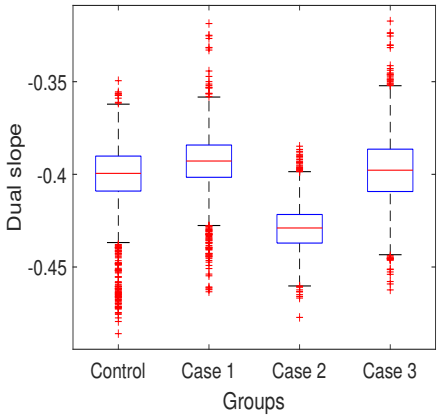
### C.1.3 Duality of Non-decimated Wavelet Transform




(a)




(b)




(c)

**Figure 35:** Application 1: Box plots of slopes of dual slopes with NDWT (a), NDCWT (b), and NDQWT (c).

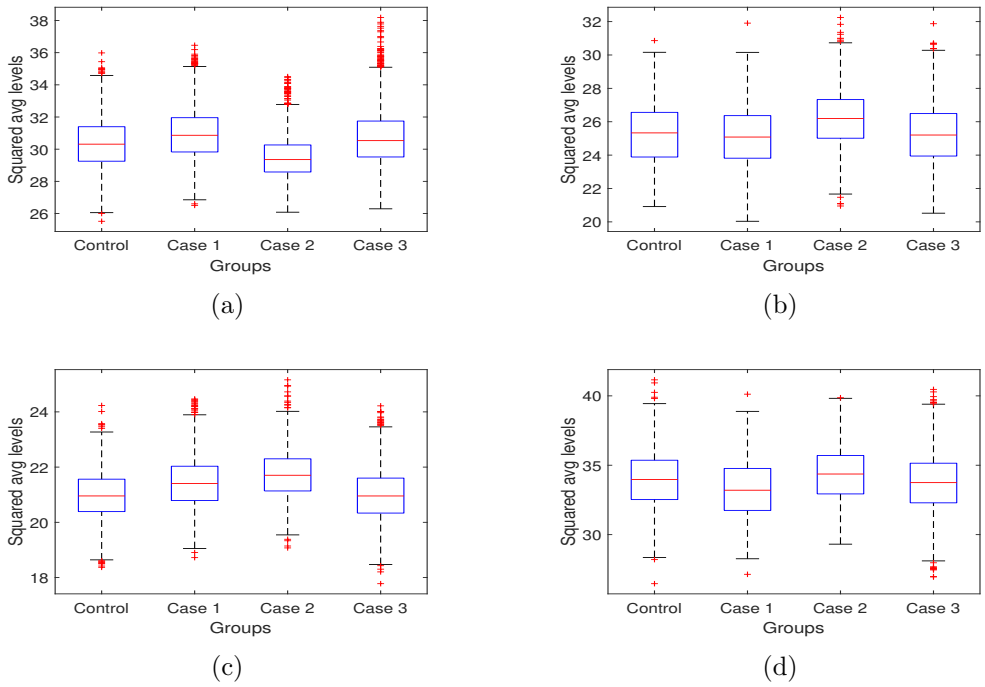



**Figure 36:** Application 1: Box plots of average squared levels from NDCWT in interval (a) $[-\pi, -\frac{1}{2}\pi)$, (b) $[-\frac{1}{2}\pi, 0)$, (c) $[0, \frac{1}{2}\pi)$, and (d) $[\frac{1}{2}\pi, \pi]$.

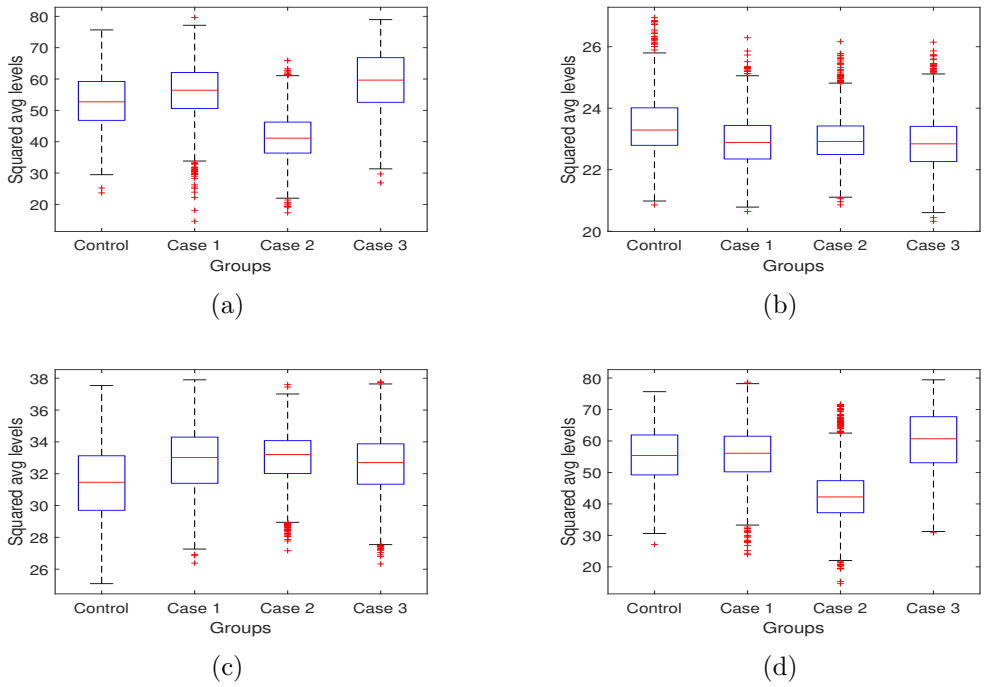



**Figure 37:** Application 1: Box plots of average squared levels from $\phi$ of NDQWT in interval (a) $[-\pi, -\frac{1}{2}\pi)$, (b) $[-\frac{1}{2}\pi, 0)$, (c) $[0, \frac{1}{2}\pi)$, and (d) $[\frac{1}{2}\pi, \pi]$.

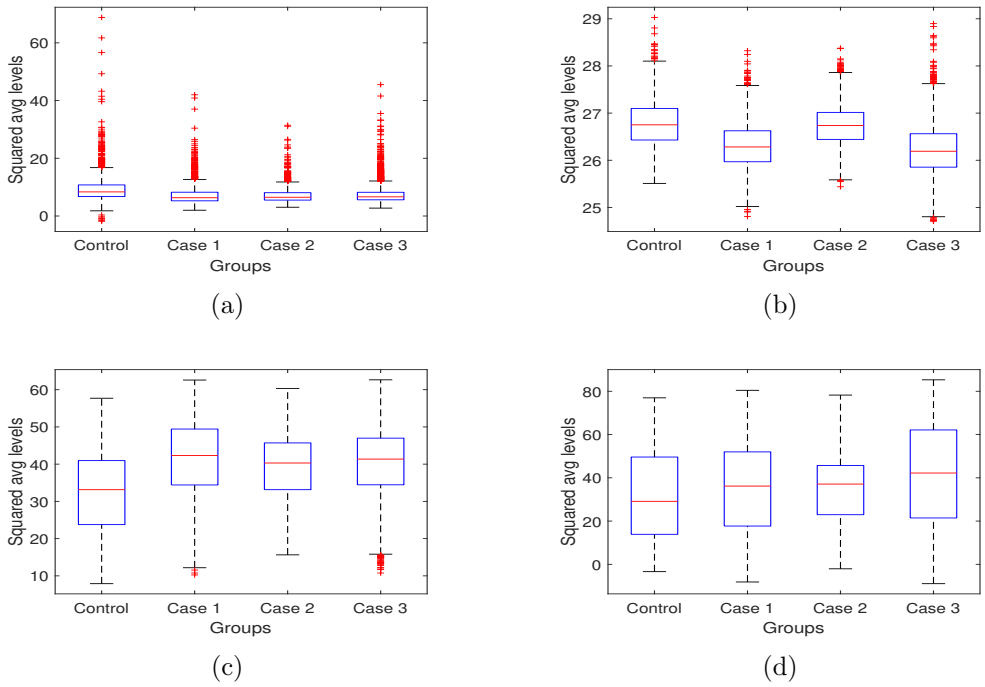



**Figure 38:** Application 1: Box plots of average squared levels from $\theta$ of NDQWT in interval (a) $[-\frac{1}{2}\pi, -\frac{1}{4}\pi)$, (b) $[-\frac{1}{4}\pi, 0)$, (c) $[0, \frac{1}{4}\pi)$, and (d) $[\frac{1}{4}\pi, \frac{1}{2}\pi]$.

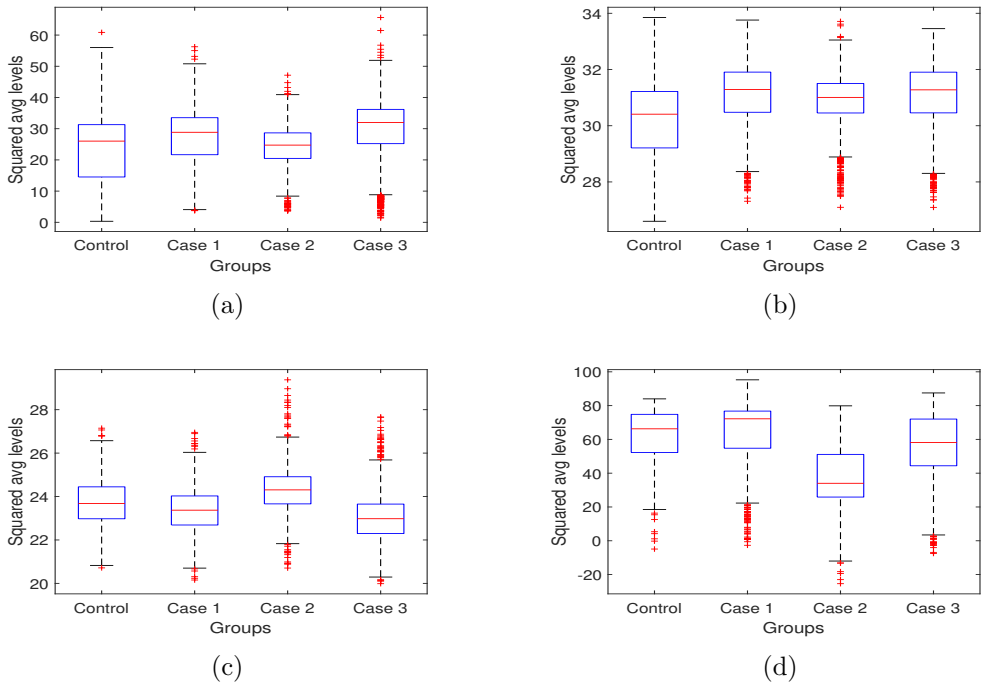



**Figure 39:** Application 1: Box plots of average squared levels from $\psi$ of NDQWT in interval (a) $[-\frac{1}{4}\pi, -\frac{1}{8}\pi)$, (b) $[-\frac{1}{8}\pi, 0)$, (c) $[0, \frac{1}{8}\pi)$, and (d) $[\frac{1}{8}\pi, \frac{1}{4}\pi]$.

## *C.2 Application 2: Classifying Sounds Data*

### C.2.1 Non-decimated Complex Wavelet Transform

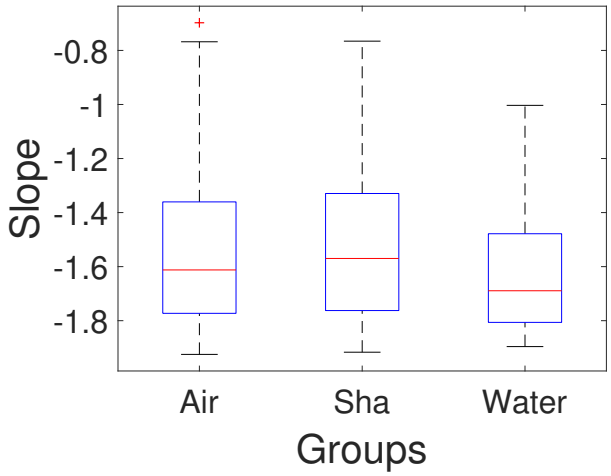



**Figure 40:** Application 2: Box plot of spectral slope obtained by NDCWT.

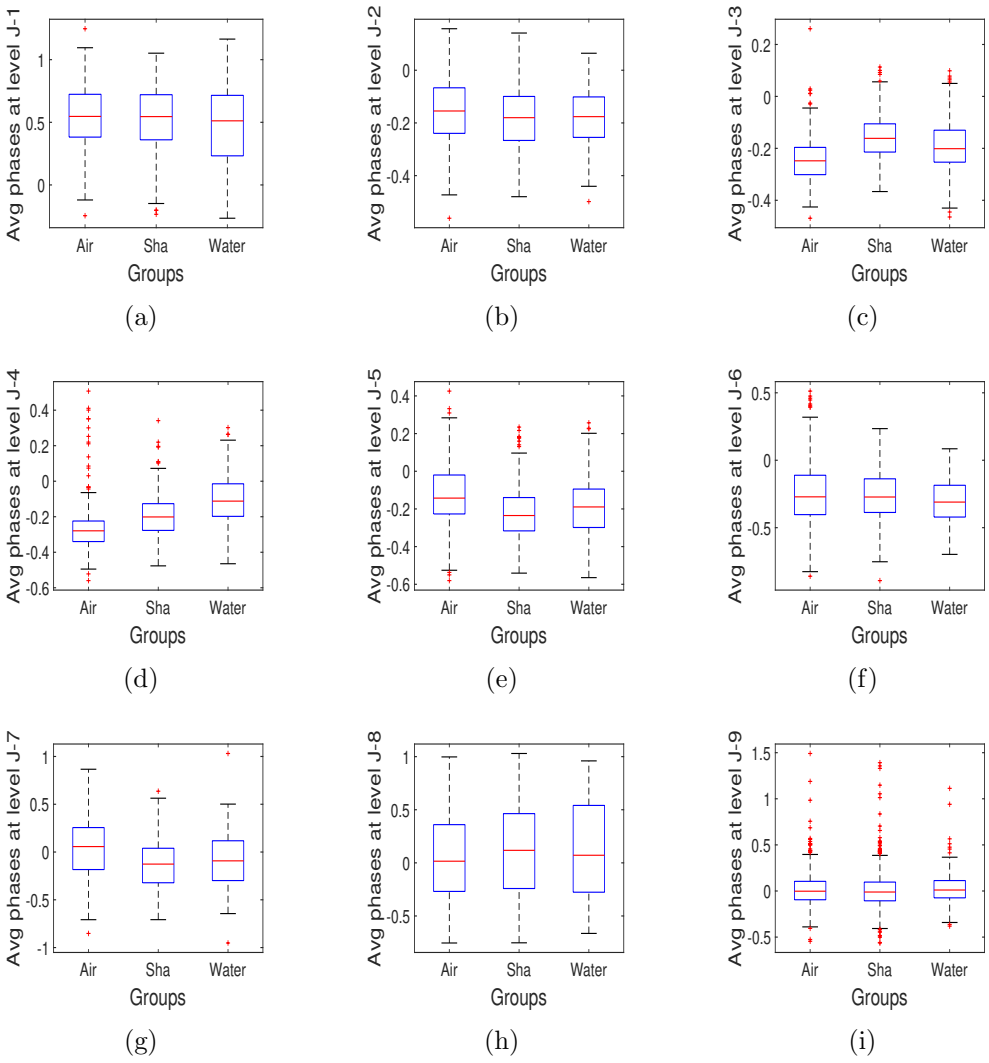



**Figure 41:** Application 2: Box plots of phase averages at all multiresolution levels obtained by NDCWT.

### C.2.2 Non-decimated Quaternion Wavelet Transform

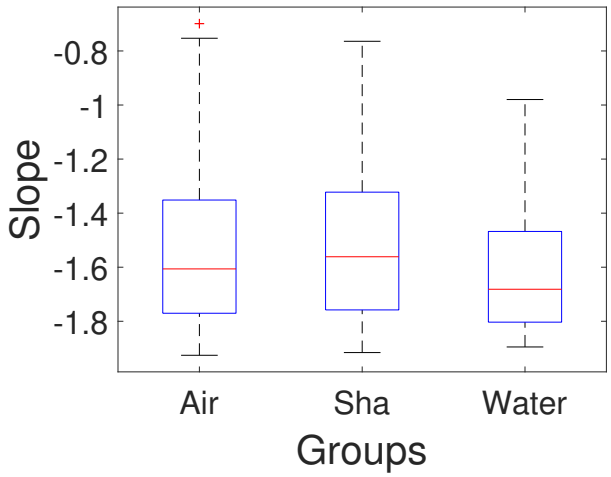



**Figure 42:** Application 2: Box plot of spectral slope obtained by NDQWT.

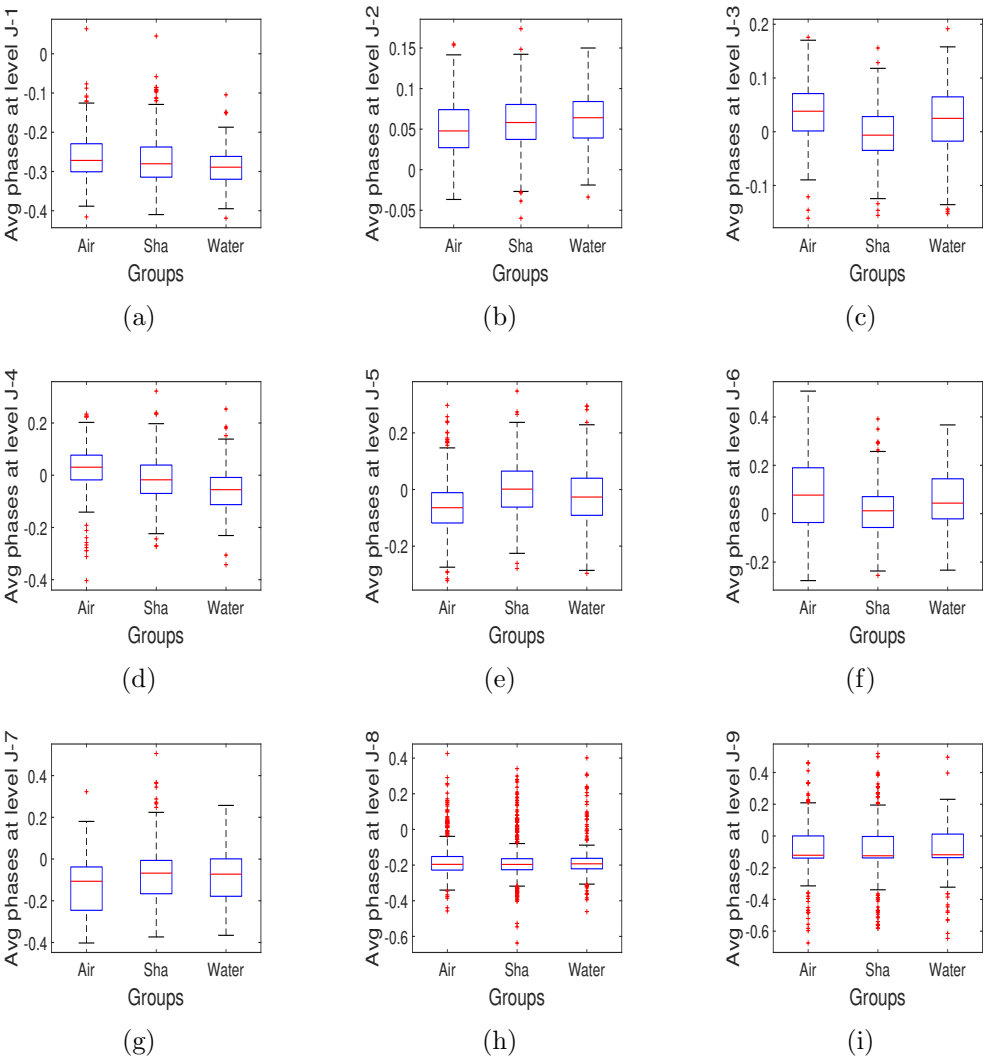



**Figure 43:** Application 2: Box plots of averages of phase $\phi$ at all multiresolution levels obtained by NDQWT.

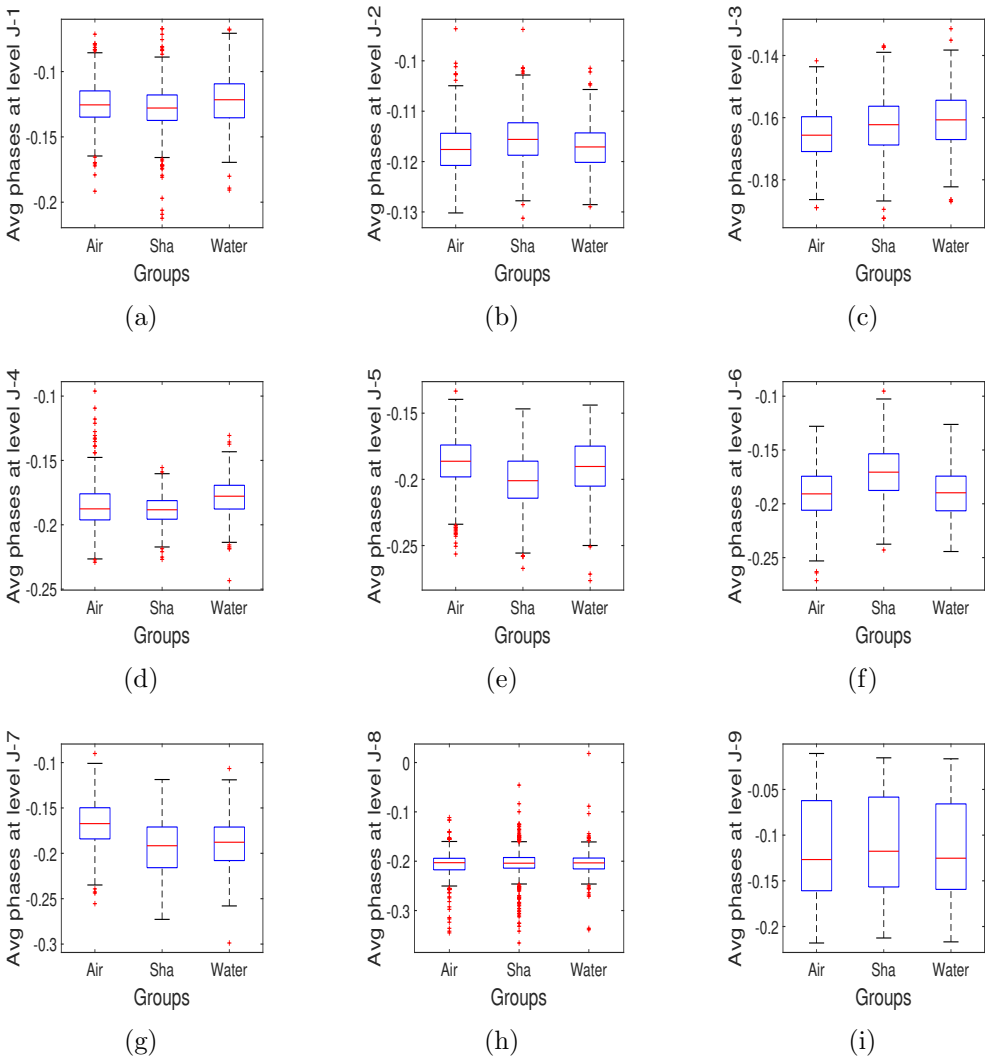



**Figure 44:** Application 2: Box plots of averages of phase $\theta$ at all multiresolution levels obtained by NDQWT.

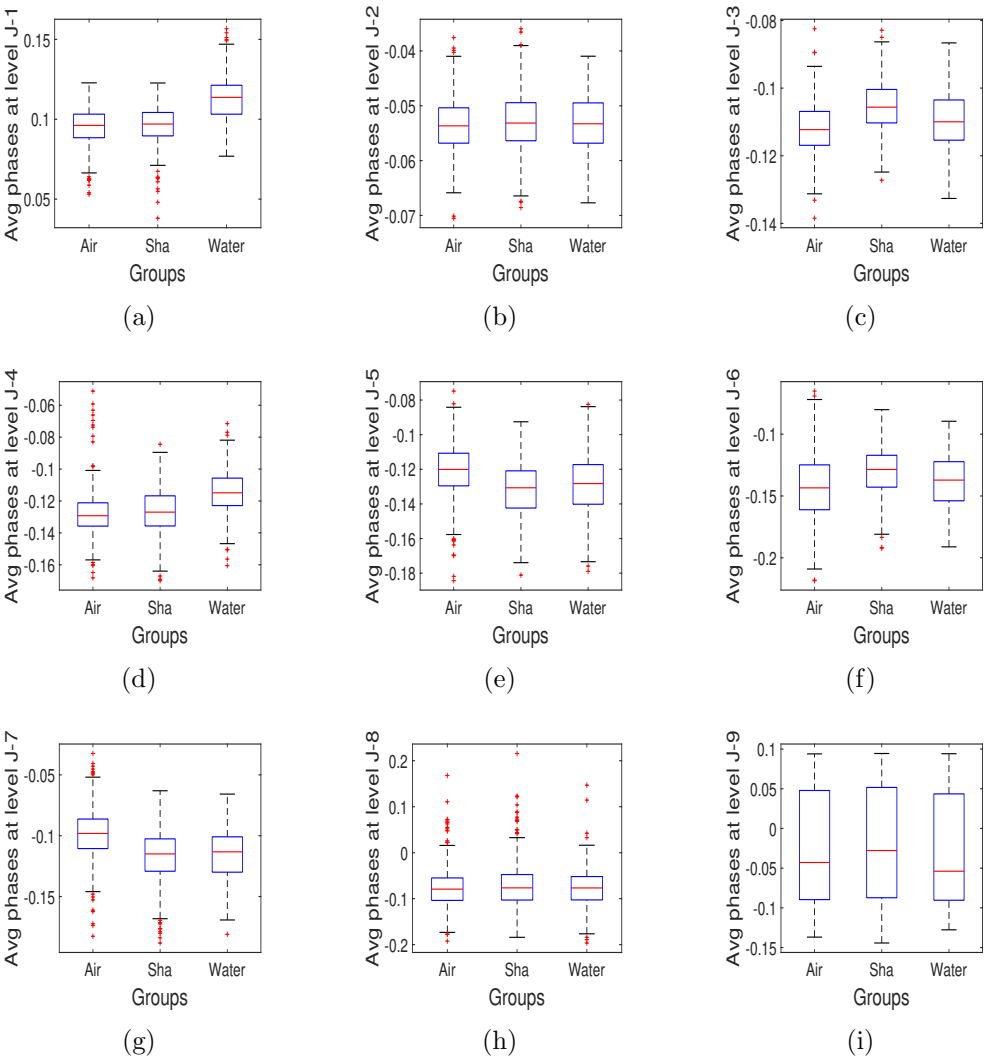
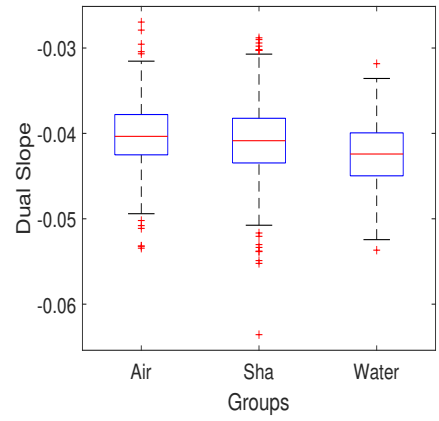
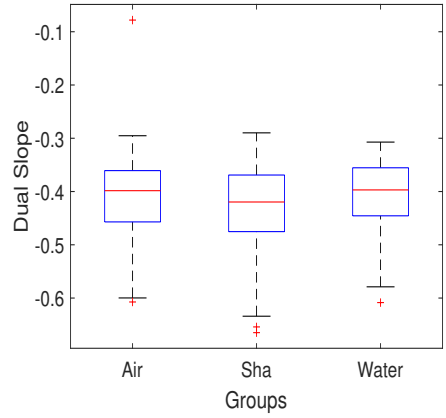



**Figure 45:** Application 2: Box plots of averages of phase $\psi$ at all multiresolution levels obtained by NDQWT.

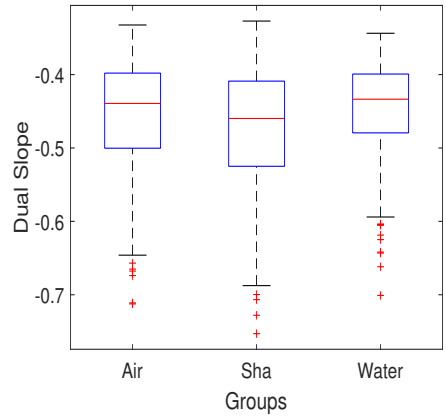
### C.2.3 Duality of Non-decimated Wavelet Transform




(a)




(b)




(c)

**Figure 46:** Application 2: Box plots of slopes of dual slopes with NDWT (a), NDCWT (b), and NDQWT (c).

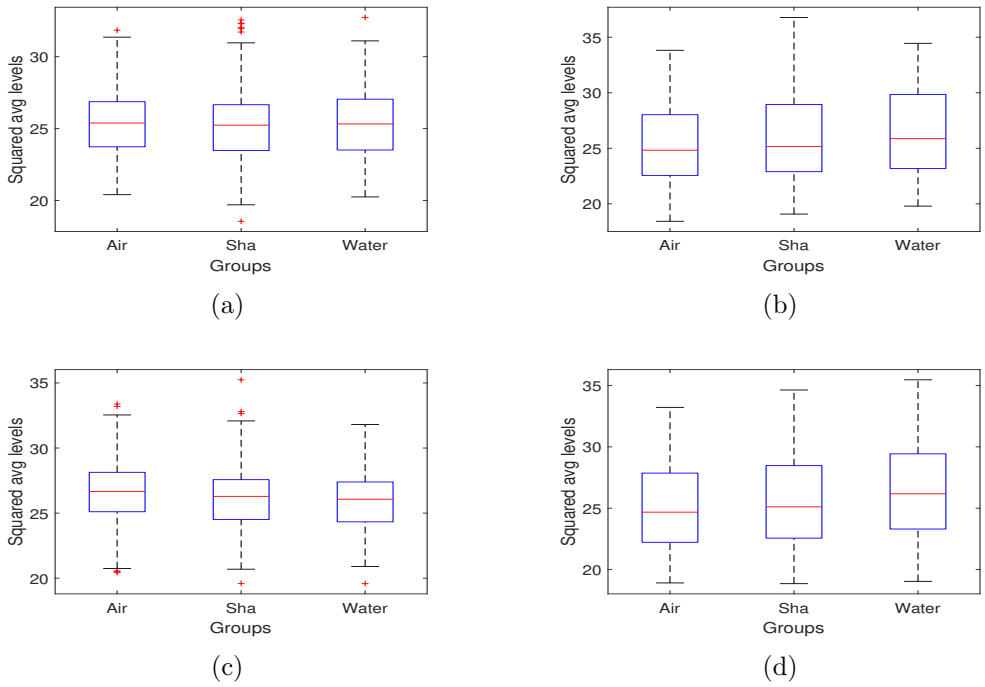



**Figure 47:** Application 2: Box plots of average squared levels from NDCWT in interval (a) $[-\pi, -\frac{1}{2}\pi)$, (b) $[-\frac{1}{2}\pi, 0)$, (c) $[0, \frac{1}{2}\pi)$, and (d) $[\frac{1}{2}\pi, \pi]$.

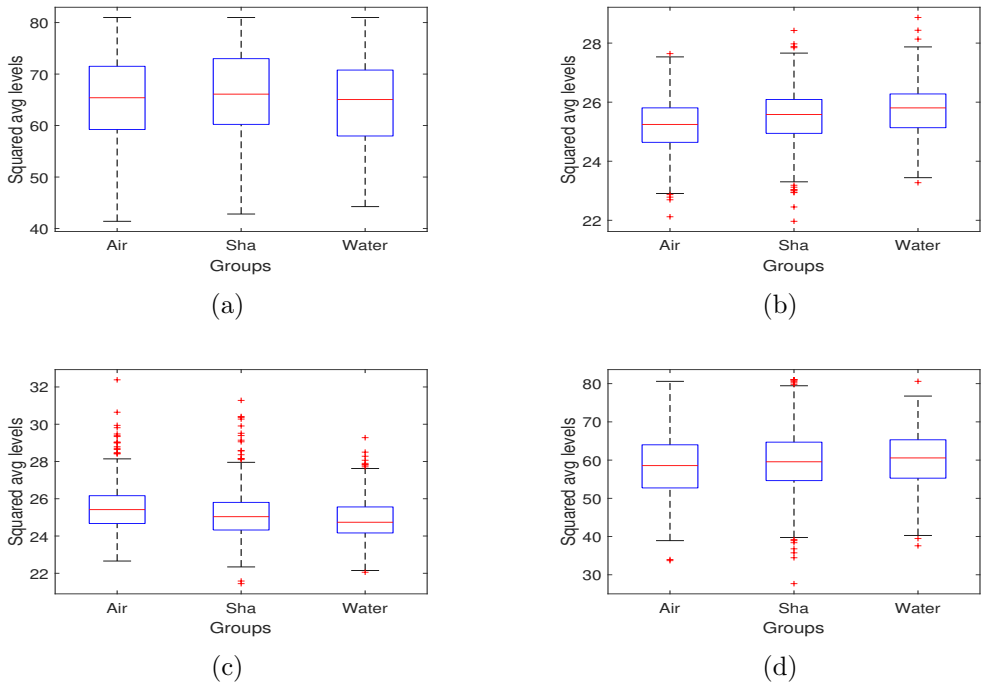



**Figure 48:** Application 2: Box plots of average squared levels from $\phi$ of NDQWT in interval (a) $[-\pi, -\frac{1}{2}\pi)$, (b) $[-\frac{1}{2}\pi, 0)$, (c) $[0, \frac{1}{2}\pi)$, and (d) $[\frac{1}{2}\pi, \pi]$.

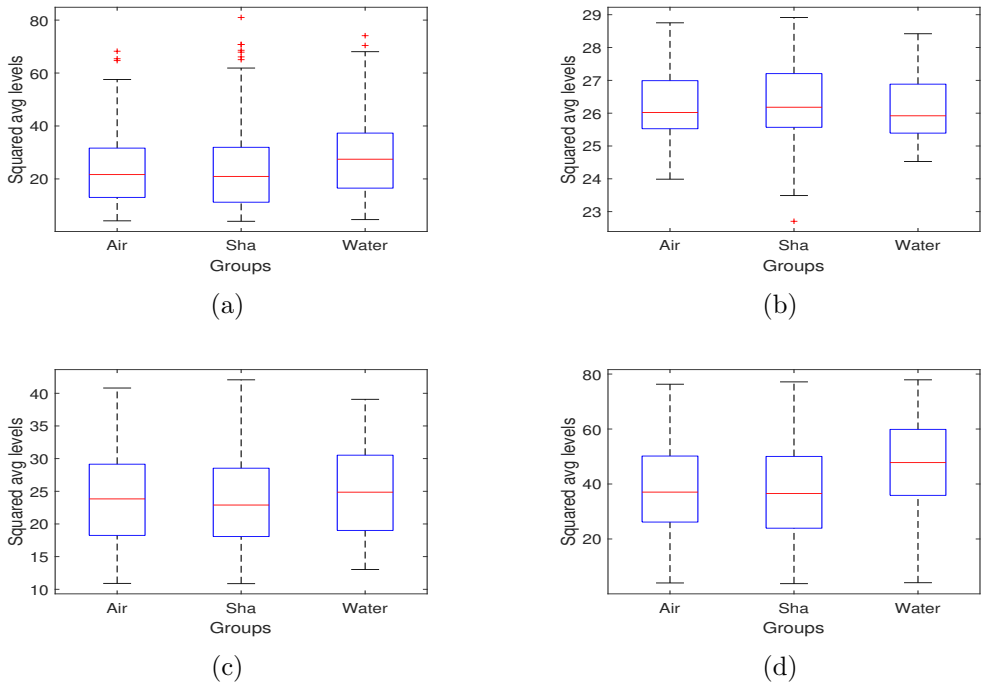



**Figure 49:** Application 2: Box plots of average squared levels from $\theta$ of NDQWT in interval (a) $[-\frac{1}{2}\pi, -\frac{1}{4}\pi)$, (b) $[-\frac{1}{4}\pi, 0)$, (c) $[0, \frac{1}{4}\pi)$, and (d) $[\frac{1}{4}\pi, \frac{1}{2}\pi]$.

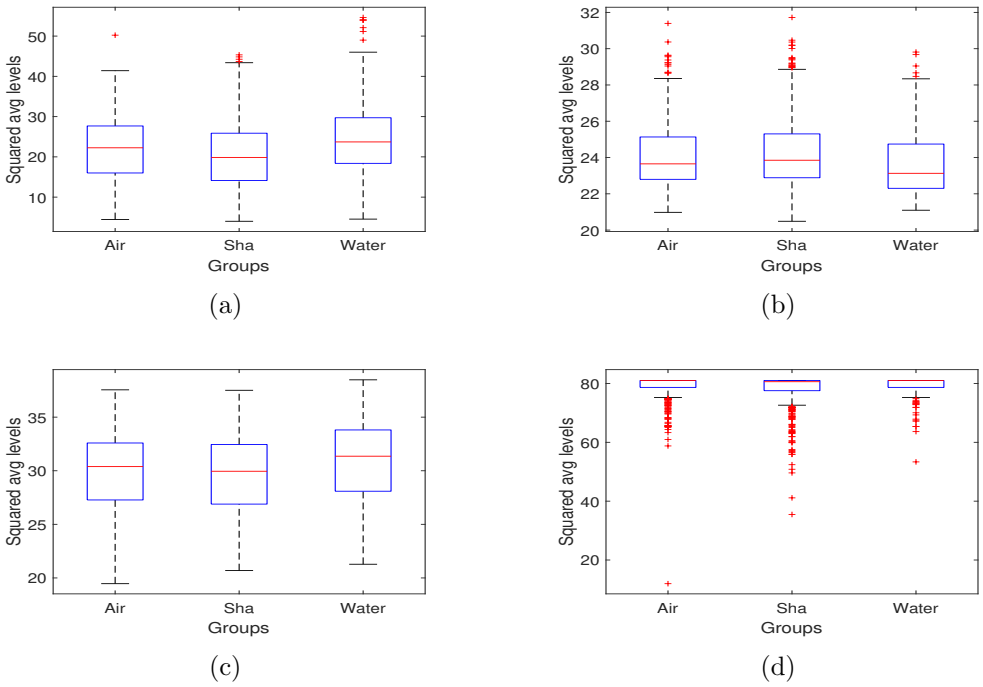



**Figure 50:** Application 2: Box plots of average squared levels from $\psi$ of NDQWT in interval (a) $[-\frac{1}{4}\pi, -\frac{1}{8}\pi)$, (b) $[-\frac{1}{8}\pi, 0)$, (c) $[0, \frac{1}{8}\pi)$, and (d) $[\frac{1}{8}\pi, \frac{1}{4}\pi]$.

## *C.3 Application 3: Screening Mammograms*

### C.3.1 Non-decimated Complex Wavelet Transform

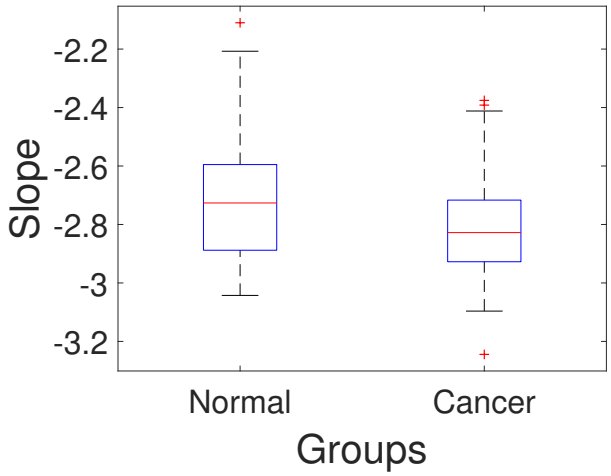



**Figure 51:** Application 3: Box plot of spectral slope obtained by NDCWT.

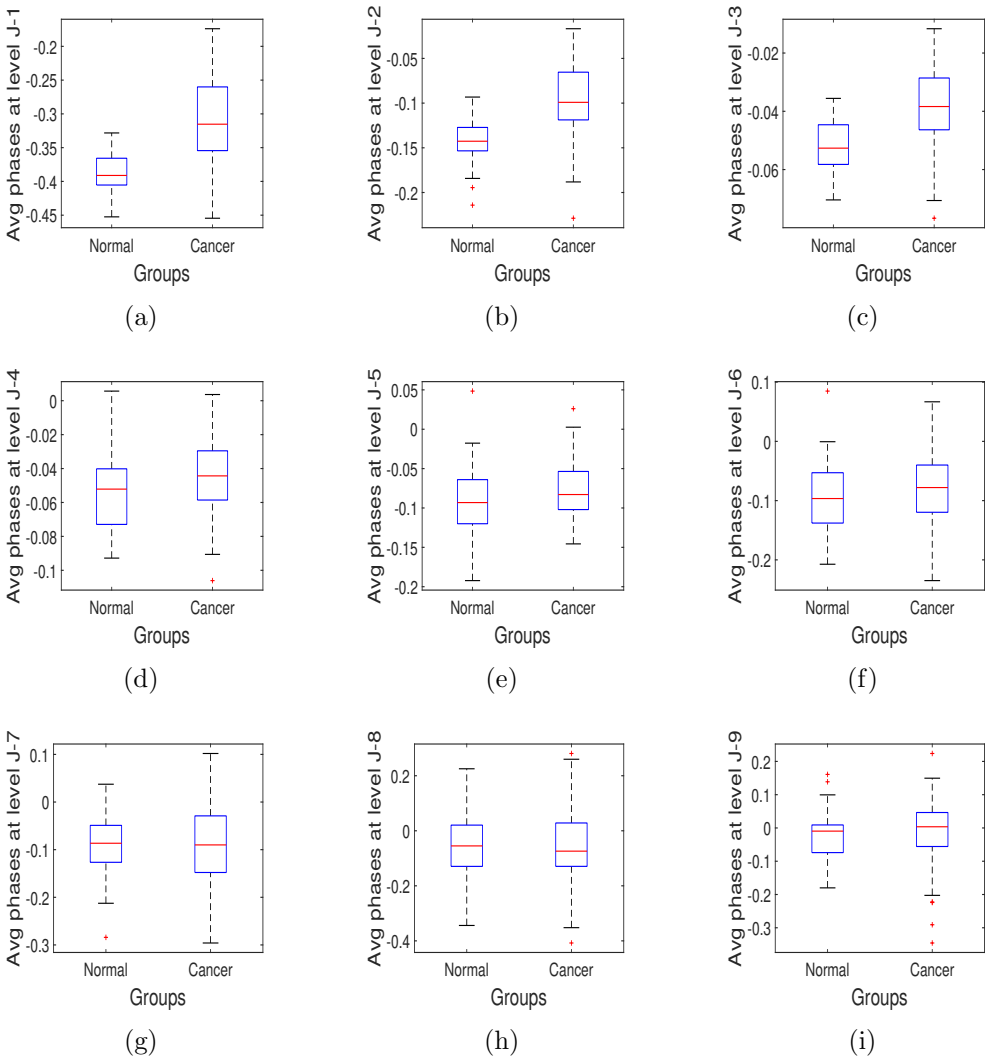



**Figure 52:** Application 3: Box plots of phase averages at all multiresolution levels obtained by NDCWT.

### C.3.2 Non-decimated Quaternion Wavelet Transform

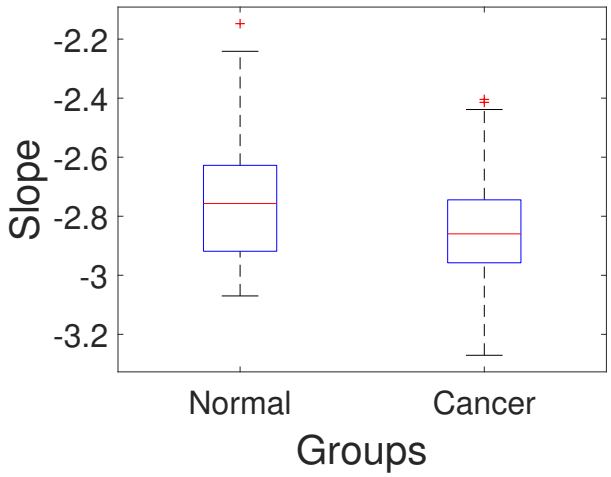



**Figure 53:** Application 3: Box plot of spectral slope obtained by NDQWT.

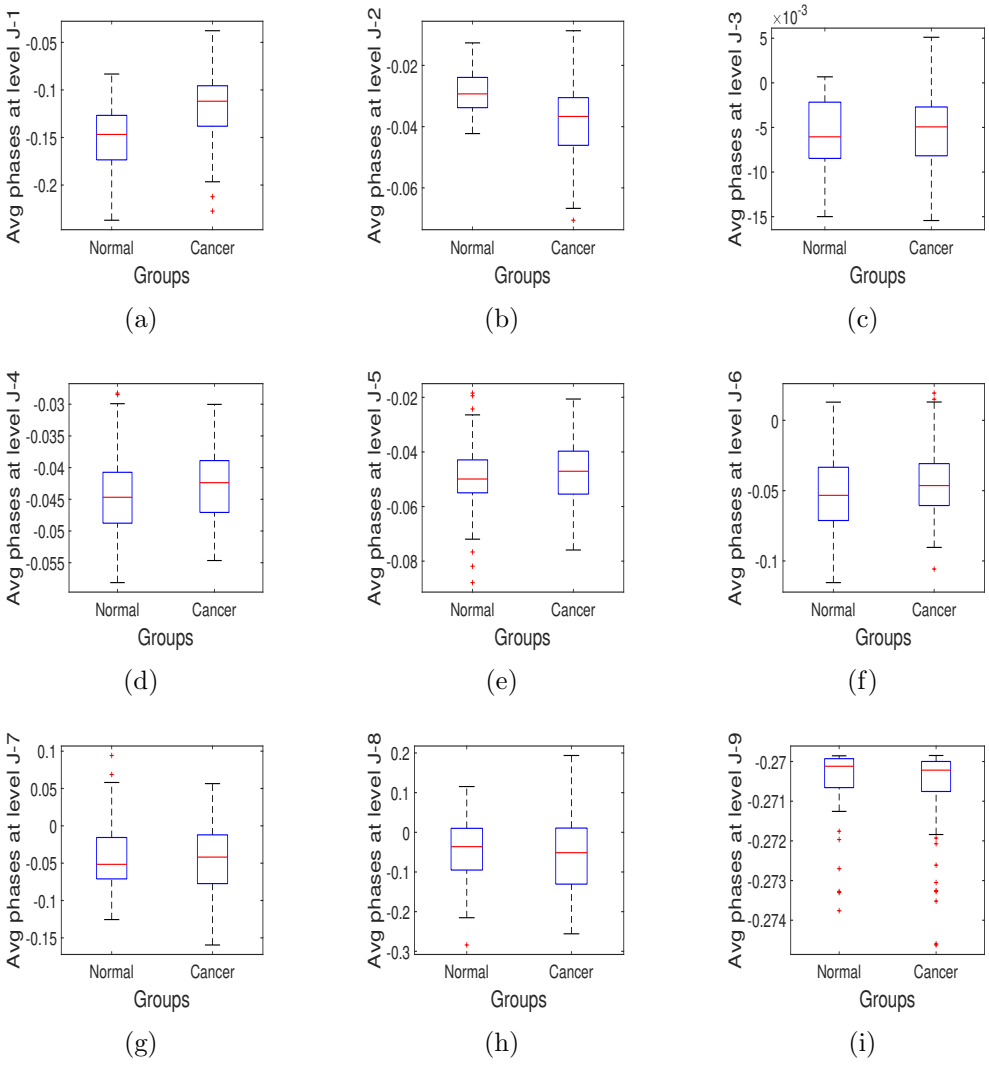



**Figure 54:** Application 3: Box plots of averages of phase $\phi$ at all multiresolution levels obtained by NDQWT.

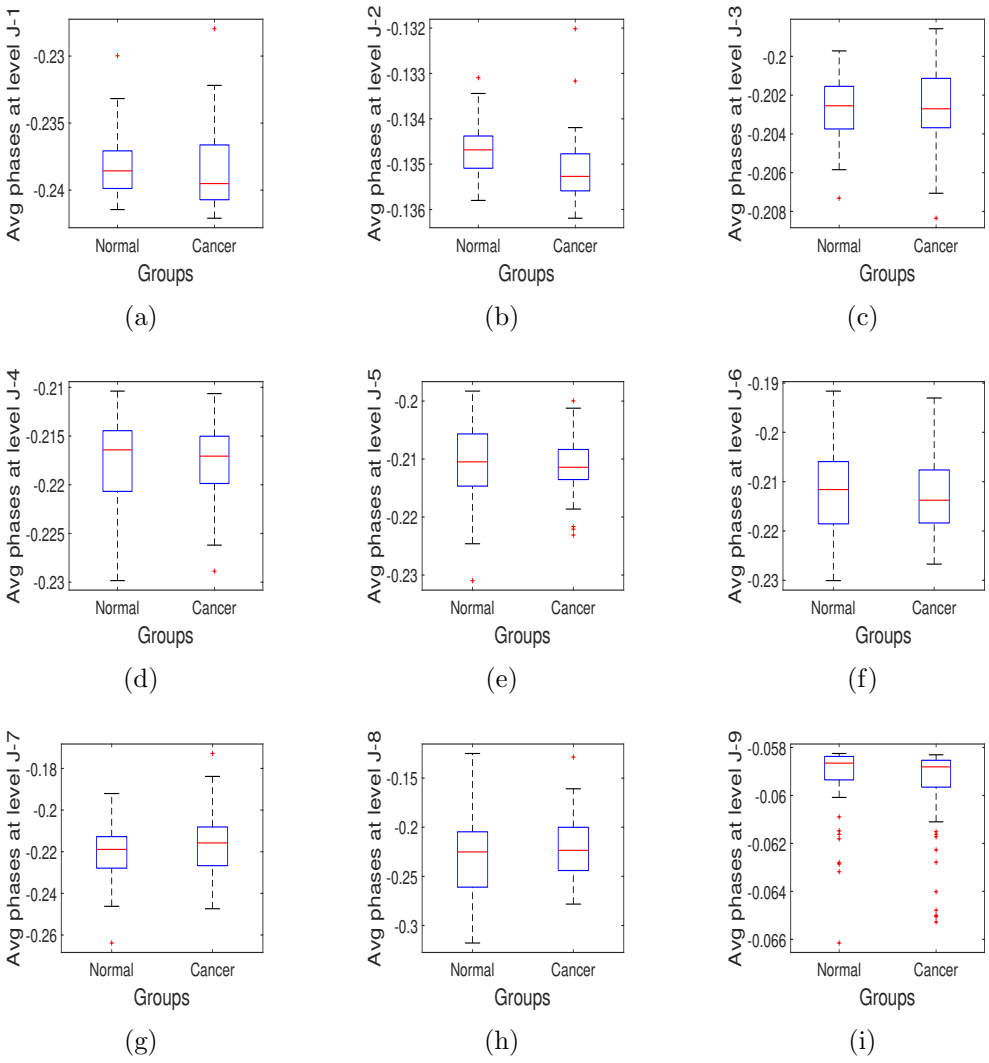



**Figure 55:** Application 3: Box plots of averages of phase $\theta$ at all multiresolution levels obtained by NDQWT.

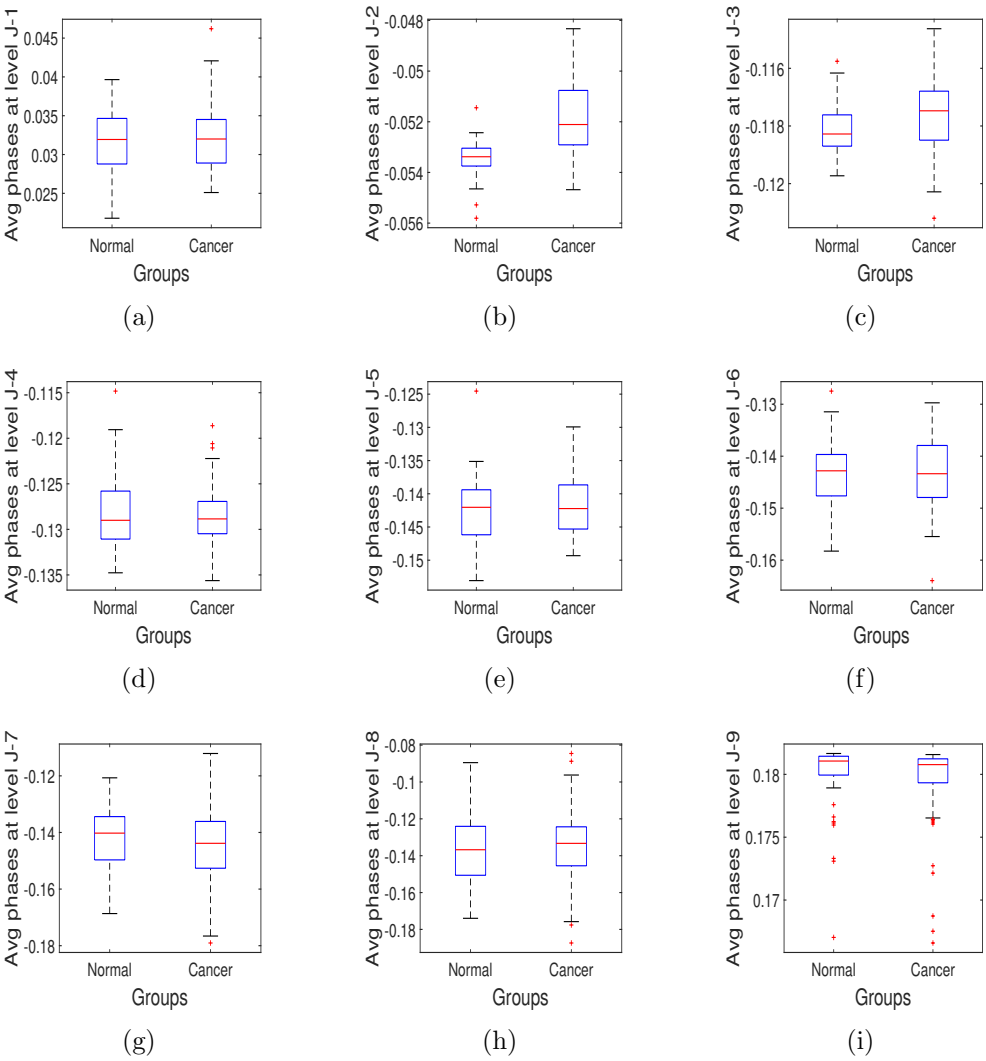
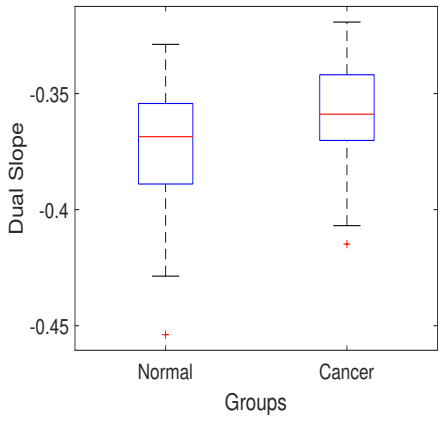
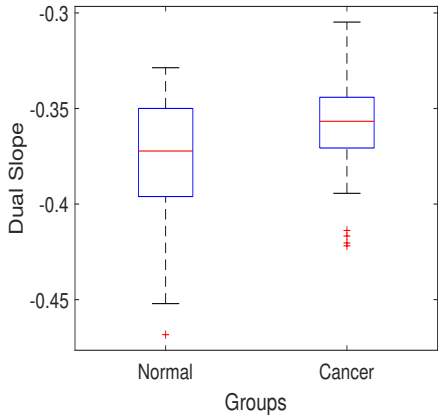



**Figure 56:** Application 3: Box plots of averages of phase $\psi$ at all multiresolution levels obtained by NDQWT.

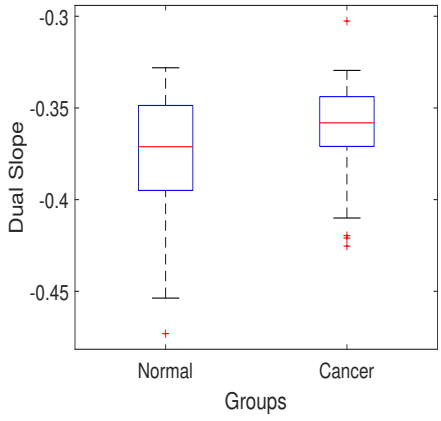
### C.3.3 Duality of Non-decimated Wavelet Transform




(a)




(b)




(c)

**Figure 57:** Application 3: Box plots of slopes of dual slopes with NDWT (a), NDCWT (b), and NDQWT (c).

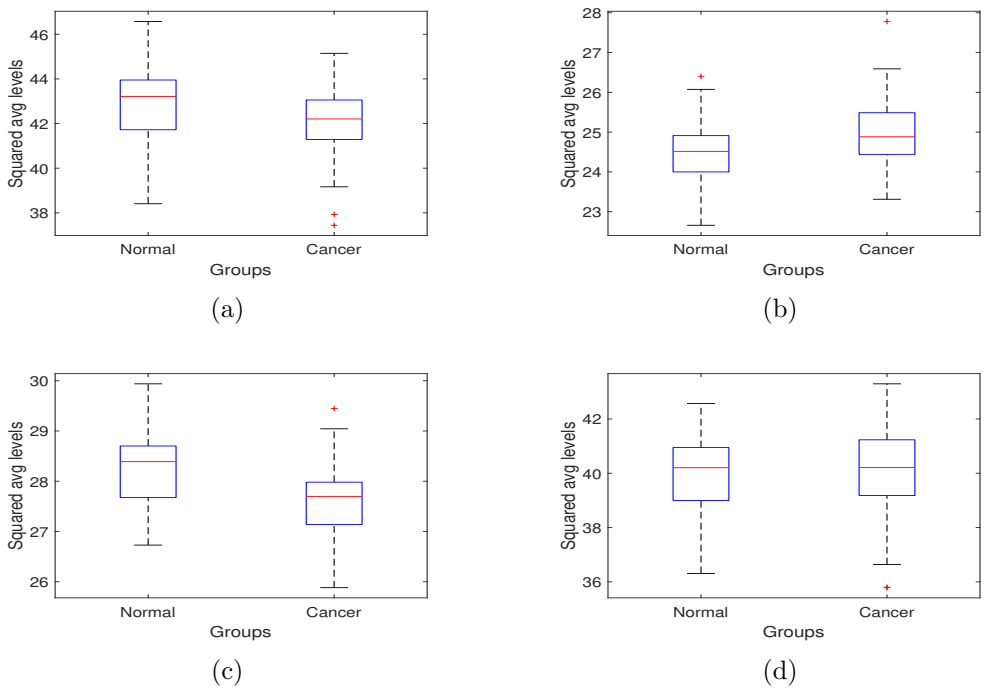



**Figure 58:** Application 3: Box plots of average squared levels from NDCWT in interval (a) $[-\pi, -\frac{1}{2}\pi)$, (b) $[-\frac{1}{2}\pi, 0)$, (c) $[0, \frac{1}{2}\pi)$, and (d) $[\frac{1}{2}\pi, \pi]$.

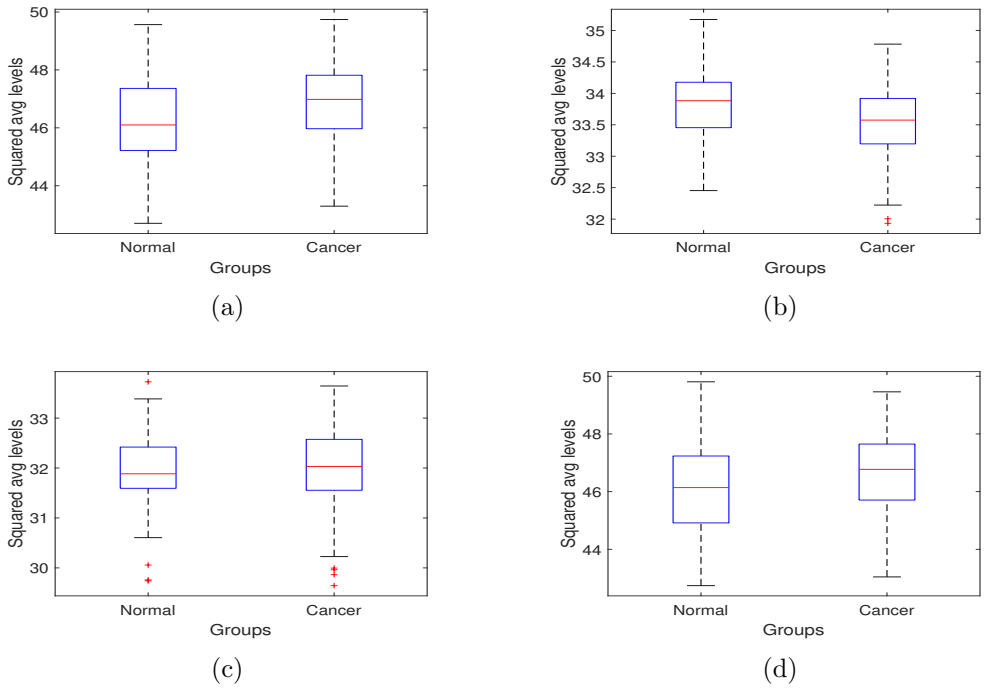



**Figure 59:** Application 3: Box plots of average squared levels from $\phi$ of NDQWT in interval (a) $[-\pi, -\frac{1}{2}\pi)$, (b) $[-\frac{1}{2}\pi, 0)$, (c) $[0, \frac{1}{2}\pi)$, and (d) $[\frac{1}{2}\pi, \pi]$.

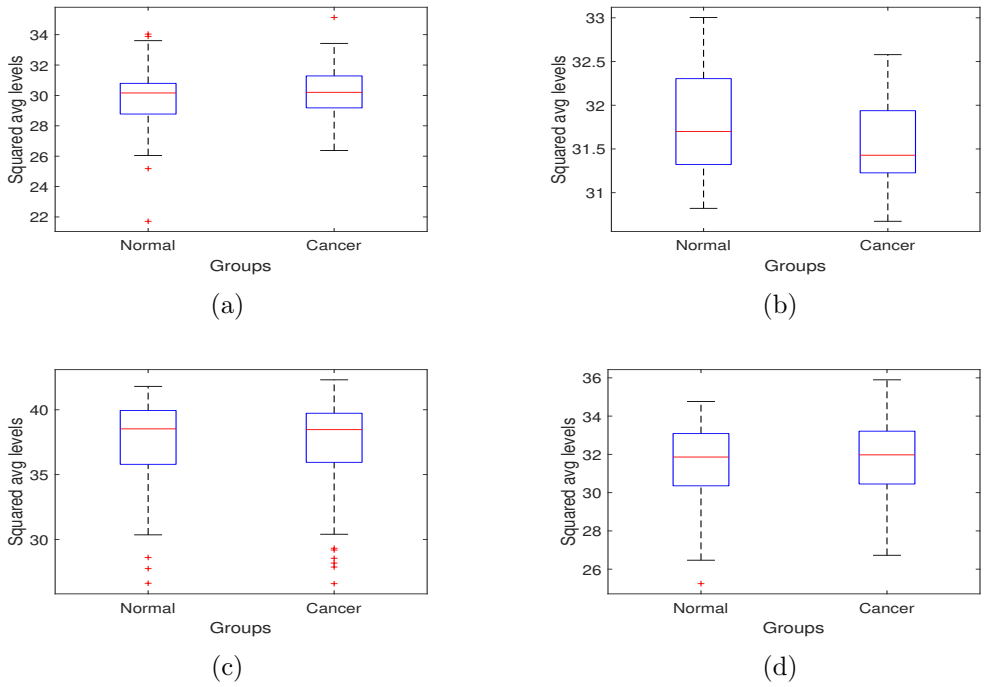



**Figure 60:** Application 3: Box plots of average squared levels from $\theta$ of NDQWT in interval (a) $[-\frac{1}{2}\pi, -\frac{1}{4}\pi)$, (b) $[-\frac{1}{4}\pi, 0)$, (c) $[0, \frac{1}{4}\pi)$, and (d) $[\frac{1}{4}\pi, \frac{1}{2}\pi]$.

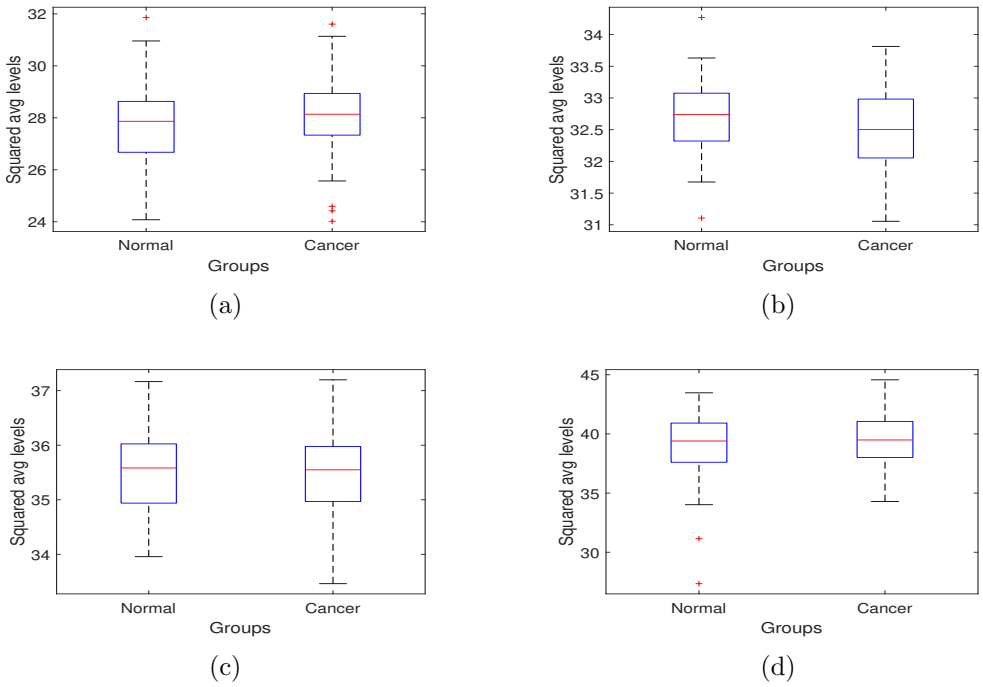



**Figure 61:** Application 3: Box plots of average squared levels from $\psi$ of NDQWT in interval (a) $[-\frac{1}{4}\pi, -\frac{1}{8}\pi)$, (b) $[-\frac{1}{8}\pi, 0)$, (c) $[0, \frac{1}{8}\pi)$, and (d) $[\frac{1}{8}\pi, \frac{1}{4}\pi]$.

## *C.4 Application 4: Seam Detection in Steel Rolling Process*

### C.4.1 Non-decimated Complex Wavelet Transform

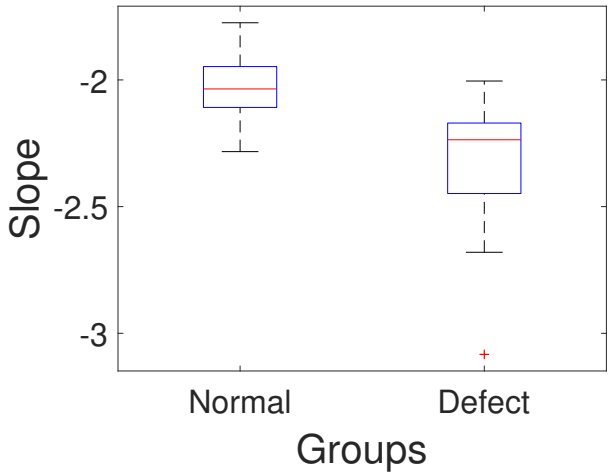



**Figure 62:** Application 4: Box plot of spectral slope obtained by NDCWT.

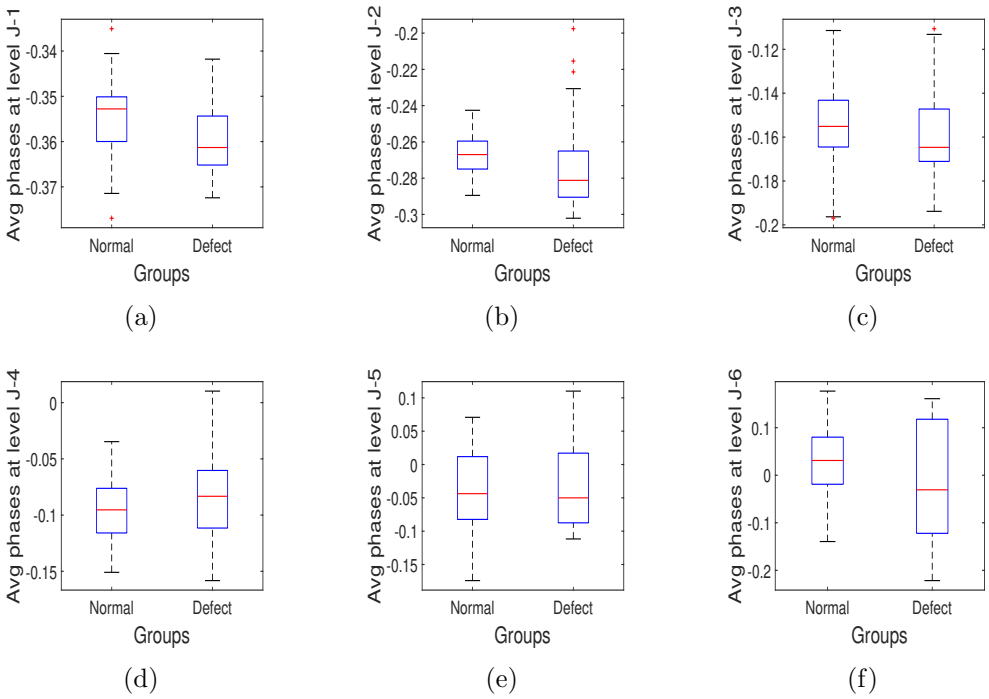



**Figure 63:** Application 4: Box plots of phase averages at all multiresolution levels obtained by NDCWT.

### C.4.2 Non-decimated Quaternion Wavelet Transform

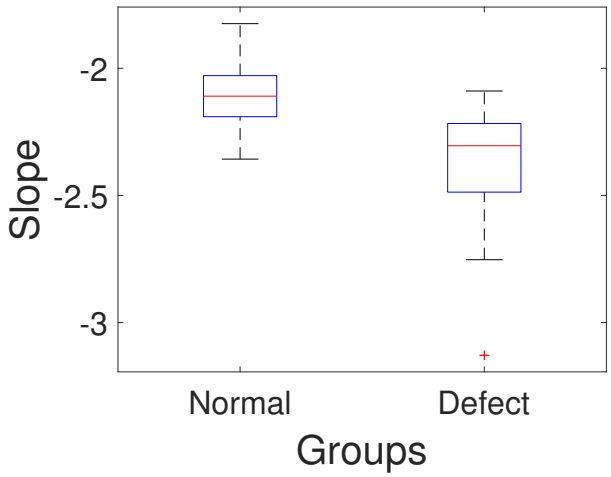



**Figure 64:** Application 4: Box plot of spectral slope obtained by NDQWT.

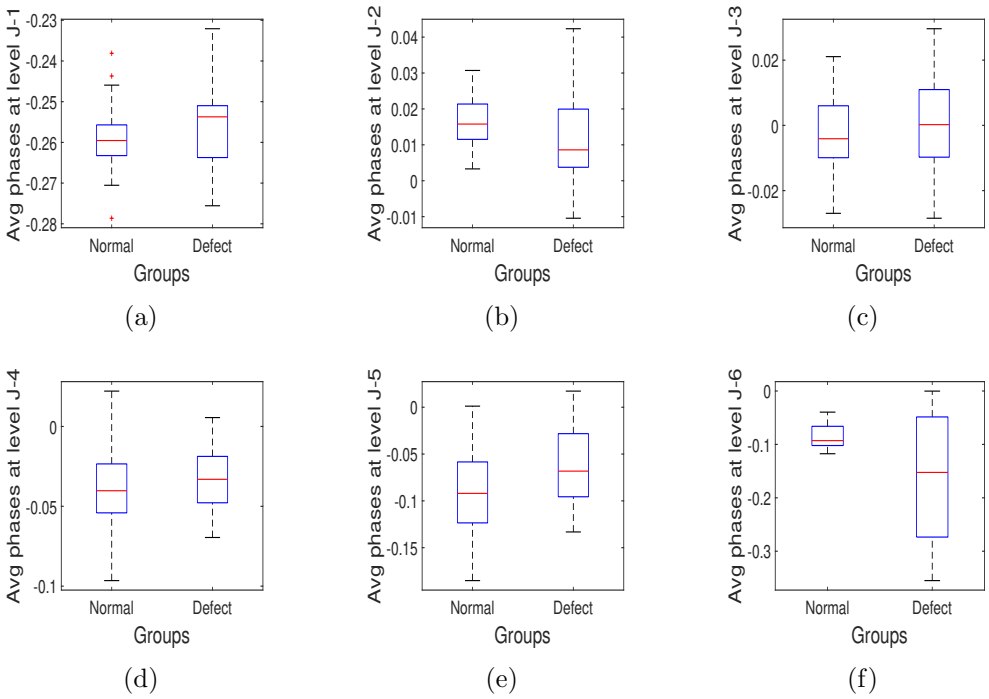



**Figure 65:** Application 4: Box plots of averages of phase $\phi$ at all multiresolution levels obtained by NDQWT.

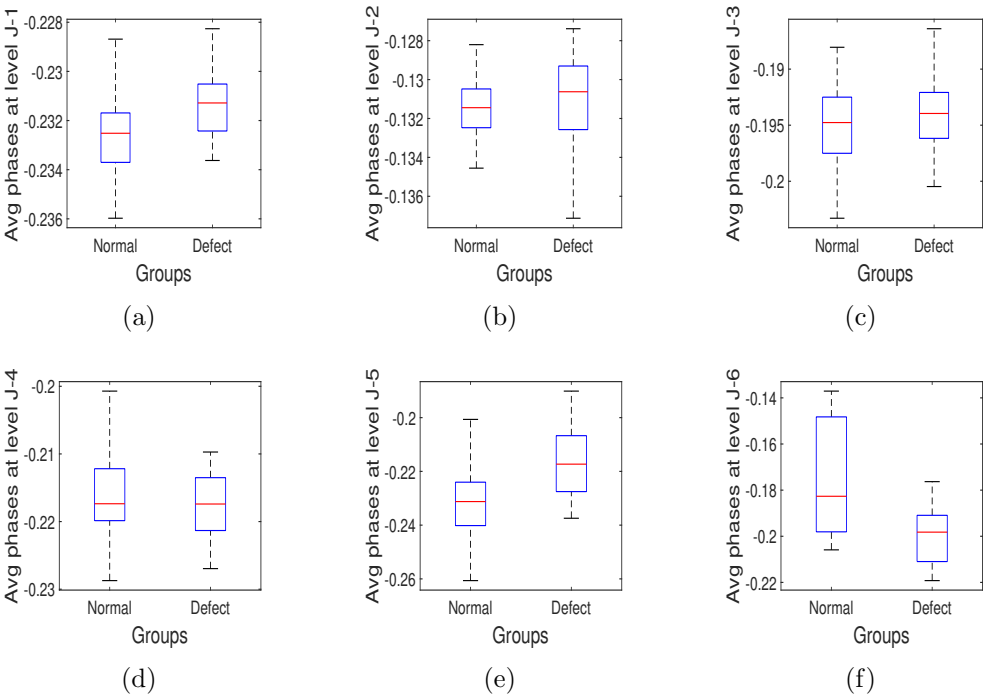



**Figure 66:** Application 4: Box plots of averages of phase $\theta$ at all multiresolution levels obtained by NDQWT.

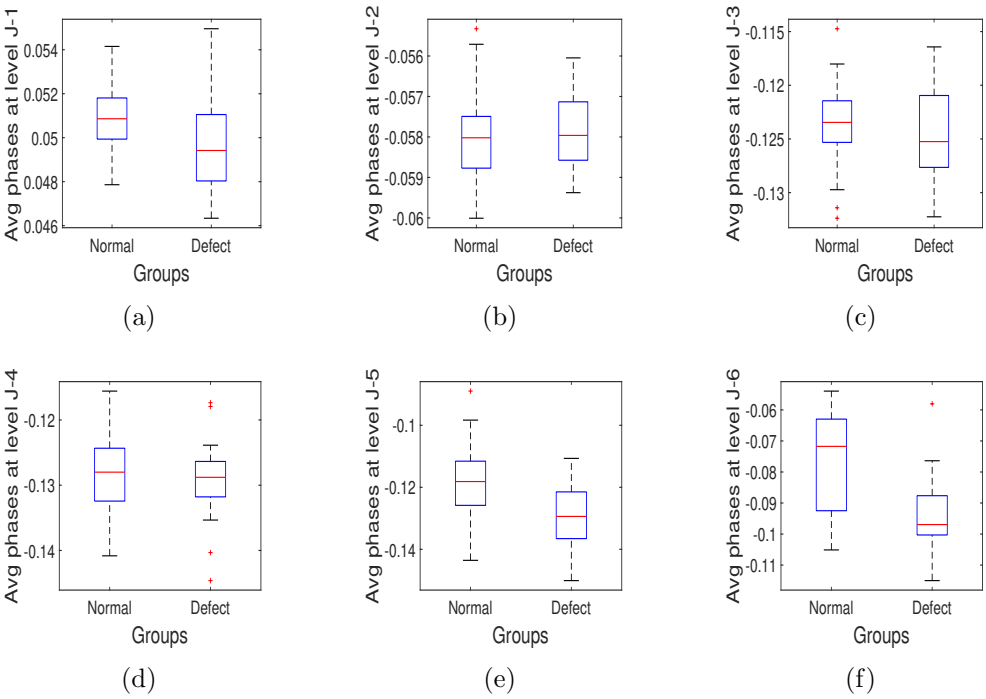
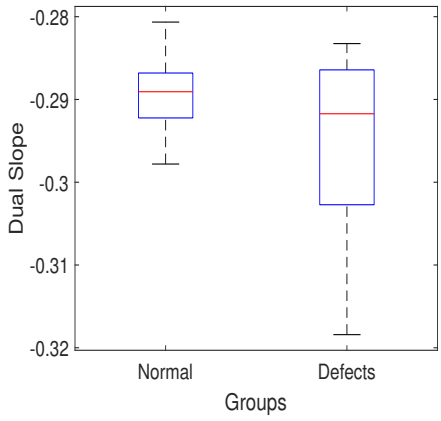
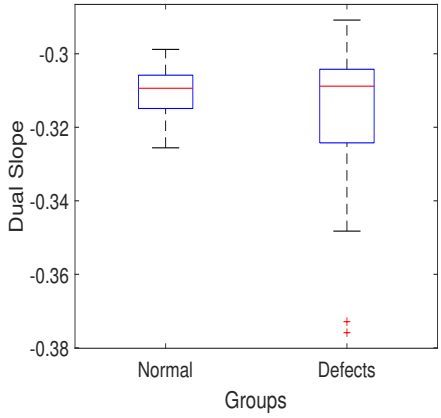



**Figure 67:** Application 4: Box plots of averages of phase $\psi$ at all multiresolution levels obtained by NDQWT.

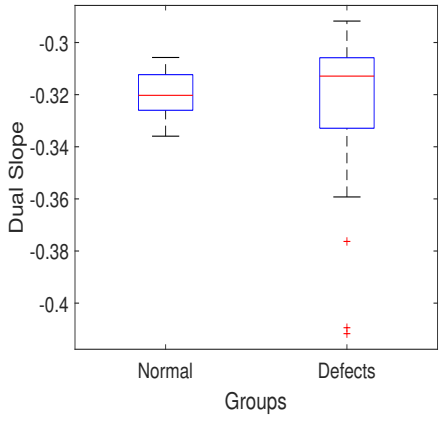
### C.4.3 Duality of Non-decimated Wavelet Transform




(a)




(b)




(c)

**Figure 68:** Application 4: Box plots of slopes of dual slopes with NDWT (a), NDCWT (b), and NDQWT (c).

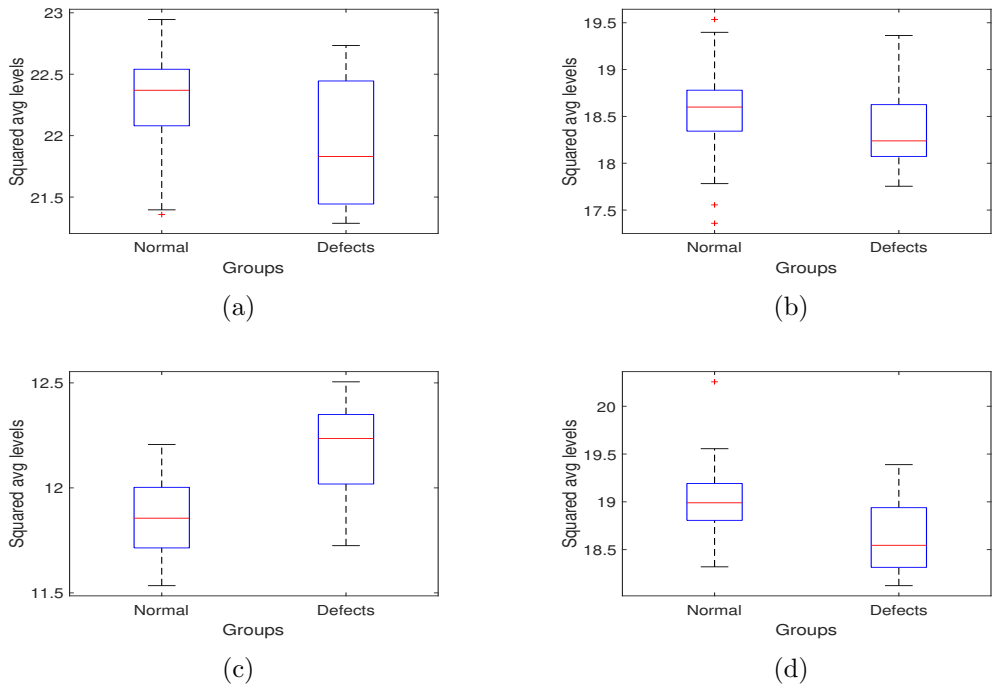



**Figure 69:** Application 4: Box plots of average squared levels from NDCWT in interval (a) $[-\pi, -\frac{1}{2}\pi)$, (b) $[-\frac{1}{2}\pi, 0)$, (c) $[0, \frac{1}{2}\pi)$, and (d) $[\frac{1}{2}\pi, \pi]$.

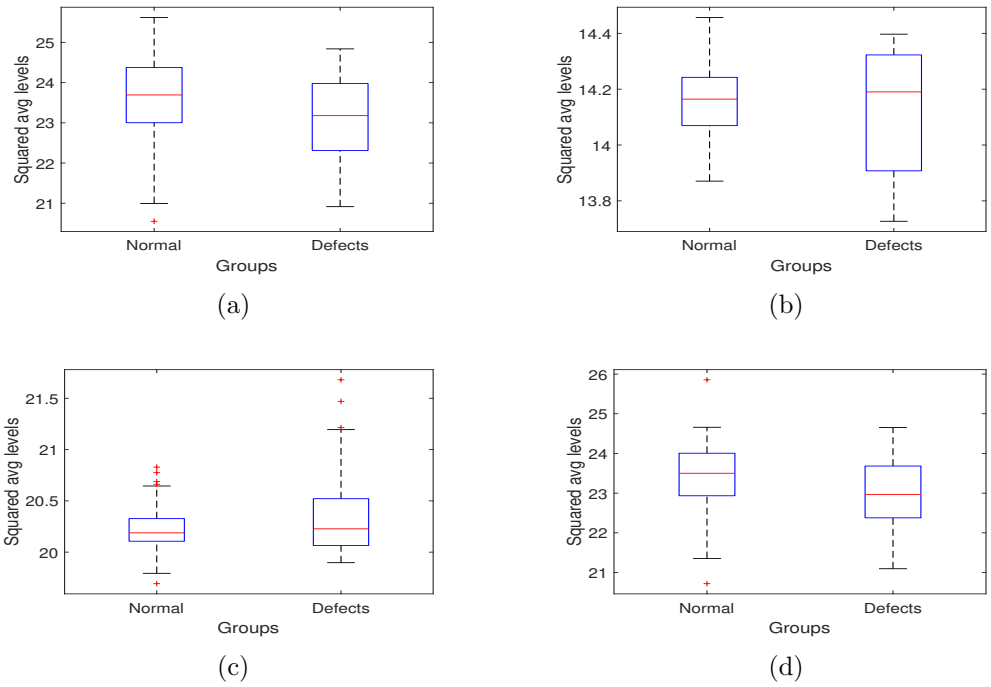



**Figure 70:** Application 4: Box plots of average squared levels from $\phi$ of NDQWT in interval (a) $[-\pi, -\frac{1}{2}\pi)$, (b) $[-\frac{1}{2}\pi, 0)$, (c) $[0, \frac{1}{2}\pi)$, and (d) $[\frac{1}{2}\pi, \pi]$.

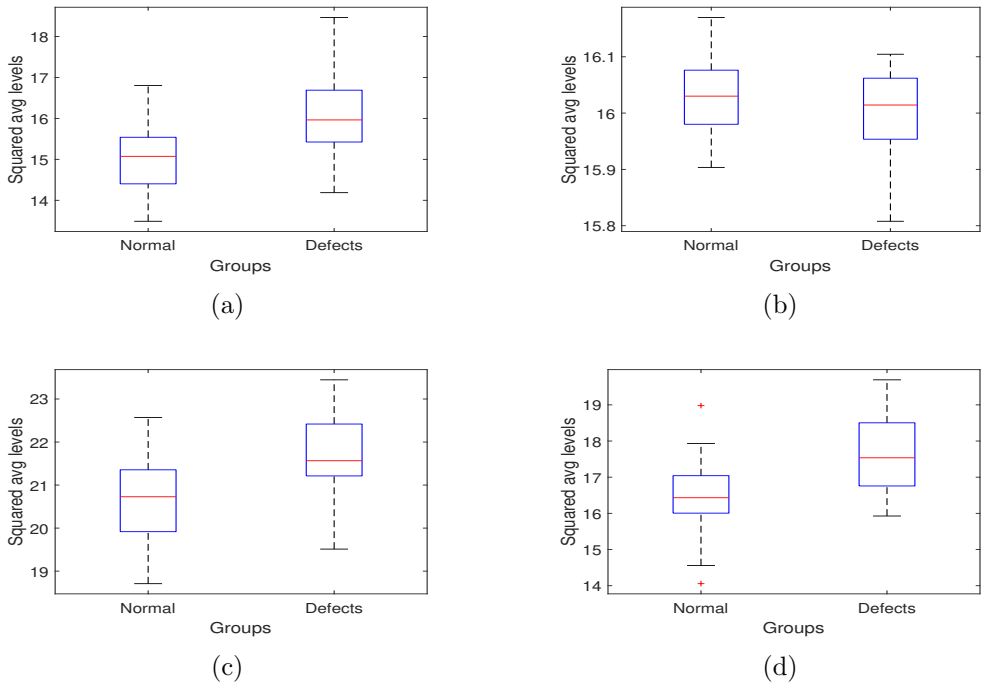



**Figure 71:** Application 4: Box plots of average squared levels from $\theta$ of NDQWT in interval (a) $[-\frac{1}{2}\pi, -\frac{1}{4}\pi)$, (b) $[-\frac{1}{4}\pi, 0)$, (c) $[0, \frac{1}{4}\pi)$, and (d) $[\frac{1}{4}\pi, \frac{1}{2}\pi]$.

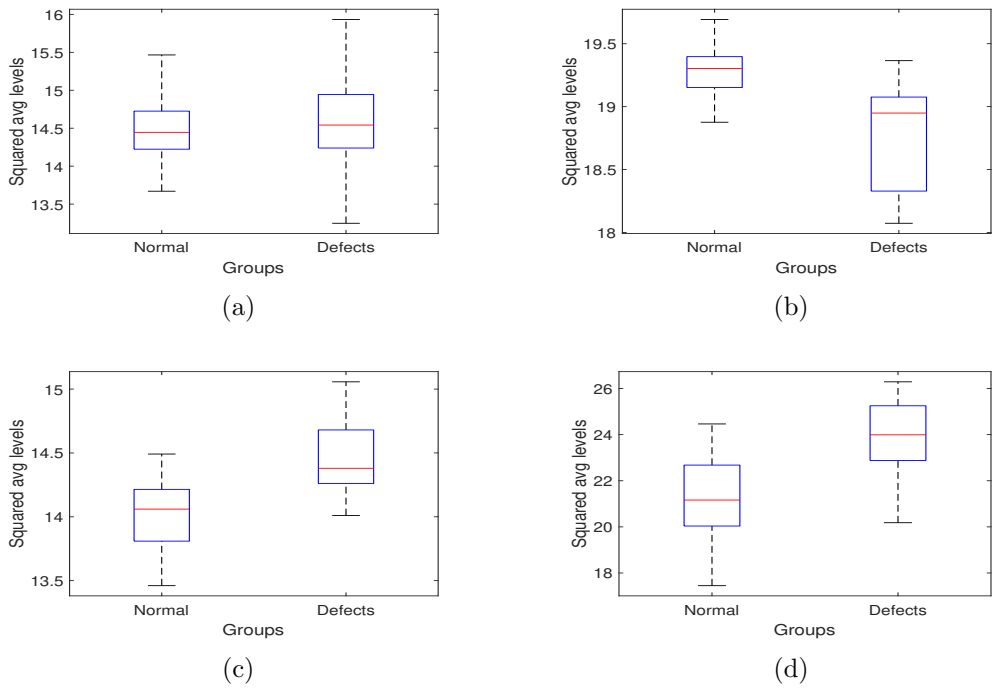



**Figure 72:** Application 4: Box plots of average squared levels from $\psi$ of NDQWT in interval (a) $[-\frac{1}{4}\pi, -\frac{1}{8}\pi)$, (b) $[-\frac{1}{8}\pi, 0)$, (c) $[0, \frac{1}{8}\pi)$, and (d) $[\frac{1}{8}\pi, \frac{1}{4}\pi]$.

# Bibliography


Achim A, Kuruoglu E (2005) Image denoising using bivariate $\alpha$-stable distributions in the complex wavelet domain. IEEE Signal Processing Letters 12(1):17–20

Altekruse S, Kosary C, Krapcho M, Neyman N, Aminou R, Waldron W (2010) Seer cancer statistics review: 1975-2007

Andreassi J (2000) Psychophysiology: Human behavior and physiological response. Psychology Press, Mahwah, NJ

Antonini M, Barlaud M, Mathieu P, Daubechies I (1992) Image coding using wavelet transform. IEEE Transactions on Image Processing 1(2):205–220

Barbur J (2004) Learning from the pupil: Studies of basic mechanisms and clinical applications, vol 1, pp 641–656

Barnard GA (1956) Discussion of hurst. Proceedings of the Institution of Civil Engineers 5:552–553

Bayro-Corrochano E (2005) Multi-resolution image analysis using the quaternion wavelet transform. Numerical Algorithms 39(1-3):35–55

Benzi R, Paladin G, Parisi G, Vulpiani A (1985) Characterisation of intermittency in chaotic systems. Journal of Physics A: Mathematical and General 18(12):2157–2165

Billow T (1999) Hypercomplex spectral signal representations for the processing and analysis of images. PhD thesis, University Christian Albrechts university of Kiel

Billow T, Sommer G (1997) Multi-dimensional signal processing using an algebraically extended signal representation, vol 1315, Springer, Berlin, Heidelberg, pp 148–163

Carre P, Denis P (2006) Quaternionic wavelet transform for colour images. In: Proceedings of SPIE - The International Society for Optical Engineering, vol 6383, pp 638,301–638,301

Chan WL, Choi H, Baraniuk RG (2008) Coherent multiscale image processing using dual-tree quaternion wavelets. IEEE Transactions on Image Processing 17(7):1069–1082

Coifman R, Donoho D (1995) Translation-invariant de-noising. In: Wavelets and Statistics, vol 103, Springer, pp 125–150

Feller W (1951) The asymptotic distribution of the range of sums of independent random variables. The Annals of Mathematical Statistics 22(3):427–432

Feng C, Mei Y, Vidakovic B (2018) Mammogram diagnostics using robust wavelet-based estimator of hurst exponent. In: Zhao Y, Chen DG (eds) New Frontiers of Biostatistics and Bioinformatics, Springer International Publishing, pp 109–140

Fletcher P, Sangwine SJ (2017) The development of the quaternion wavelet transform. Signal Processing 136:2–15

Gai S, Luo L (2015) Image denoising using normal inverse gaussian model in quaternion wavelet domain. Multimedia Tools and Applications 74(3):1107–1124

Gao R, Yan R (2011) Wavelets: Theory and Applications for Manufacturing. Springer

Ginzberg P, Walden AT (2013) Matrix-valued and quaternion wavelets. IEEE Transactions on Signal Processing 61(6):1357–1367

Gopher D, Donchin E (1986) Workload: An Examination of the Concept. John Wiley and Sons, New York, NY

Gurlebeck K, Sprossig W (1998) Quaternionic and Clifford Calculus for Physicists and Engineers. John Wiley and Sons, England

Halsey TC, Jensen MH, Kadanoff LP, Procaccia I, Shraiman BI (1987) Fractal measures and their singularities: The characterization of strange sets. Nuclear Physics B - Proceedings Supplements 2:501–511

Hamilton E, Jeon S, Ramírez P, Lee K, Vidakovic B (2011) Diagnostic classification of digital mammograms by wavelet-based spectral tools: A comparative study. In: 2011 IEEE International Conference on Bioinformatics and Biomedicine, pp 384–389

Heath M, Bowyer K, Kopans D, Moore R, Kegelmeyer W (2001) The digital database for screening mammography. Proceedings of the 5th International Workshop on Digital Mammography pp 212–218

Hogan JA, Morris AJ (2012) Quaternionic wavelets. Numerical Functional Analysis and Optimization 33(7-9):1031–1062

Houssami N, Irwig L, Ciatto S (2006) Radiological surveillance of interval breast cancers in screening programmes. The Lancet Oncology 7(3):259–265

Hurst HE (1951) Long-term storage capacity of reservoirs. Transactions of the American Society of Civil Engineers 116:770–808

Jensen MH (1999) Multiscaling and structure functions in turbulence: An alternative approach. Physical Review Letters 83:76–79

Jeon S, Nicolis O, Vidakovic B (2014) Mammogram diagnostics via 2-D complex wavelet-based self-similarity measures. Sao Paulo Journal of Mathematical Sciences 8(2):265–284

Jiang B, Wang C, Liu H (2005) Liquid crystal display surface uniformity defect inspection using analysis of variance and exponentially weighted moving average techniques. International Journal of Production Research 43(1):67–80

Jin N, Zhou S, Chang T (2004) Identification of impacting factors of surface defects in hot rolling processes using multi-level regression analysis. Transactions of the North American Manufacturing Research Institute of SME 32:557–564

Jones C, Abbott AL (2006) Color face recognition by hypercomplex gabor analysis. In: 7th International Conference on Automatic Face and Gesture Recognition (FGR06), pp 6–131

Kang M, Vidakovic B (2016) WavmatND: A MATLAB package for non-decimated wavelet transform and its applications. arXiv:1604.07098

Kang M, Auffermann W, Vidakovic B (2019) Wavelet-based scailng indices for breast cancer diagnostics. In: IC-SMHD-2016, A Festschrift in Honor of Professor Hamparsum Bozdogan, Springer, pp 1–26

Kim S, Wang Z, Oraintarac S, Temiyasathita C, Wongsawatc Y (2008) Feature selection and classification of high-resolution nmr spectra in the complex wavelet transform domain. Chemometrics and Intelligent Laboratory Systems 90:161–168

Kumar A (2008) Computer-vision-based fabric defect detection: a survey. IEEE Transactions on Industrial Electronics 55(1):348–363

Lamperti J (1962) Semi-stable stochastic processes. Transactions of the American Mathematical Society 104:62–78

Lawton W (1993) Applications of complex valued wavelet transforms to subband decomposition. IEEE Transactions on Signal Processing 41(12):3566–3568

Levi A, Stark H (1983) Signal restoration from phase by projections onto convex sets. Journal of the Optical Society of America 73(6):810–822

Lina J (1997) Image processing with complex daubechies wavelets. Journal of Mathematical Imaging and Vision 7(3):211–223

Lina J (1999) Complex dyadic multiresolution analyses. Advances in Imaging and Electron Physics 109:163–197

Lina J, MacGibbon B (1997) Nonlinear shrinkage estimation with complex daubechies wavelets. In: Wavelet Applications in Signal and Image Processing V, vol 3169, pp 67–79, DOI 10.1117/12.279680

Loewenfeld I (1999) The pupil: Anatomy, physiology, and clinical applications. Butterworth-Heinemann, Oxford, UK

Magarey J, Kingsbury N (1998) Motion estimation using a complex-valued wavelet transform. IEEE Transactions on Signal Processing 46(4):1069–1084

Mallat S (1989a) Multiresolution approximations and wavelet orthonormal bases of $L^2$ (r). Transactions of the American Mathematical Society 315(1):69–87

Mallat S (1989b) A theory for multiresolution signal decomposition: the wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence 11(7):674–693

Mallat S (2009) A Wavelet Tour of Signal Processing. Academic Press, DOI https://doi.org/10.1016/B978-0-12-374370-1.X0001-8

Mandelbrot BB (1975) Limit theorems on the self-normalized range for weakly and strongly dependent processes. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 31(4):271–285

Mandelbrot BB, Ness JV (1968) Fractional brownian motions, fractional noises and applications. SIAM Review 10(4):422–437

Mandelbrot BB, Wallis JR (1968) Noah, joseph, and operational hydrology. Water Resources Research 4:909–918

Moloney K, Jacko J, Vidakovic B, Sainfort F, Leonard K, Shi B (2006) Leveraging data complexity: Pupillary behavior of older adults with visual impairment during hci. ACM Trans Comput-Hum Interact 13:376–402

Nason G, Silverman B (1995) The stationary wavelet transform and some statistical applications. In: Wavelets and Statistics, vol 103, Springer, pp 281–299

National Cancer Institute (2014) Mammograms fact sheet. URL `http://www.cancer.gov/cancertopics/types/breast/mammograms-fact-sheet`

Oppenheim A, Li J (1981) The importance of phase in signals. IEEE Transactions on Image Processing 69(5):529–541

Percival D, Walden A (2000) Wavelet Methods for Time Series Analysis. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, DOI 10.1017/CBO9780511841040

Percival D, Walden A (2006) Wavelet Methods for Time Series Analysis, vol 4. Cambridge University Press

Portilla J, Simoncelli E (2000) A parametric texture model based on joint statistics of complex wavelet coefficients. International Journal of computer vision 40(1):49–71

Rajini GK (2016) A comprehensive review on wavelet transform and its applications. ARPN Journal of Engineering and Applied Sciences 11(19):11,713–11,723

Ramírez P, Vidakovic B (2013) A 2-D wavelet-based multiscale approach with applications to the analysis of digital mammograms. Computational Statistics & Data Analysis 58:71–81

Remenyi N, Nicolis O, Nason G, Vidakovic B (2014) Image denoising with 2-d scale-mixing complex wavelet transforms. IEEE Transactions on Image Processing 23(12):5165–5174

Roberts T, Newel M, Auffermann W, Vidakovic B (2017) Wavelet-based scaling indices for breast cancer diagnostics. Statistics in Medicine 36(12):1989–2000

Roux S, Jensen MH (2004) Dual multifractal spectra. Physical Review E 69:016,309

Shensa M (1992) The discrete wavelet transform: wedding the a trous and mallat algorithms. IEEE Transactions on Signal Processing 40(10):2464–2482

Simoncelli E, Adelson E (1996) Noise removal via bayesian wavelet coring. In: 3rd IEEE International Conference on Image Processing, IEEE Signal Processing Society, Lausanne, Switzerland, vol 1, pp 379–382

Sohn H, Park G, Wait J, Limback N, Farrar C (2004) Wavelet-based active sensing for delamination detection in composite structures. Smart Materials and Structures 13(1):153–160

Soulard R, Carre P (2010) Quaternionic wavelets for image coding. In: 2010 18th European Signal Processing Conference, pp 125–129

Soulard R, Carre P (2011) Quaternionic wavelets for texture classification. Pattern Recognition Letters 32(13):1669–1678

Stephen J, Nicolas L (2013) Quaternion Toolbox for Matlab®, version 2 with support for octonions. URL `fromhttp://qtfm.sourceforge.net`

Strang G, Nguyen T (1996) Wavelets and Filter Banks. Wellesley-Cambridge Press

Subakan O, Vemuri B (2011) A quaternion framework for color image smoothing and segmentation. International Journal of Computer Vision 91(3):233–250

The Center for the Study of Macular Degeneration (2002) Macular degeneration: Your questions answered. URL `http://www.csmd.ucsb.edu/faq/faq.html`

The Schepens Eye Research Institute (2002) Macular degeneration: Your questions answered. URL http://www.eri.harvard.edu/htmlfiles/md.html

Veitch D, Abry P (1999) A wavelet-based joint estimator of the parameters of long-range dependence. IEEE Transactions on Information Theory 45(3):878–897

Vidakovic B (1999) Statistical Modeling by Wavelets. John Wiley & Sons, Inc.

Xue-Ni Z, Xiao-Qing L, Zhan-Cheng Z, Xiao-Jun W (2016) Multi-focus image fusion using quaternion wavelet transform. In: 2016 23rd International Conference on Pattern Recognition (ICPR), pp 883–888

Zhang X, Desai M, Peng Y (1999) Orthogonal complex filter banks and wavelets: Some properties and design. IEEE Transactions on Signal Processing 47(4):1039–1048

# VITA

Taewoon Kong is a Ph.D. candidate in Industrial and Systems Engineering (ISyE) at Georgia Institute of Technology with a specialization in Statistics. He joined the Ph.D. program in 2015 and started working with Dr. Brani Vidakovic shortly thereafter. His research focus is on machine learning, time series analysis, wavelets, data science, etc. Before coming to Georgia Tech, he earned both Bachelor of Science and Master of Science degrees in Industrial Engineering at Hanyang University in 2013 and 2015.