NON-DECIMATED WAVELET TRANSFORM IN STATISTICAL ASSESSMENT OF SCALING: THEORY AND APPLICATIONS

A Thesis
Presented to
The Academic Faculty

by

Minkyoung Kang

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

Georgia Institute of Technology
August 2016

NON-DECIMATED WAVELET TRANSFORM IN STATISTICAL ASSESSMENT OF SCALING: THEORY AND APPLICATIONS

Approved by:

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

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

Professor David Goldsman
School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Eberhard O. Voit
Department of Biomedical engineering
Georgia Institute of Technology

Professor Ben Haaland
School of Industrial and Systems Engineering
Georgia Institute of Technology

Date Approved: May 6, 2016

To my family: my father (Kang, ShinSul), my mother (Kim, JeongJa), and my sister (Kang, SooJung)

ACKNOWLEDGEMENTS

First and foremost, I would like to express my appreciation to Prof. Brani Vidakovic for his support, extreme patience, kindness, and brilliance. I am very fortunate to have an opportunity to work under his advisement through my Ph.D. I also would like to appreciate my thesis committee: Prof. David Goldsman, Prof. Kamran Paynabar, Prof. Eberhard O. Voit, and Prof. Ben Haaland for their insightful comments and questions that are very helpful. I appreciate Prof. William Auffermann for valuable comments.

In addition, I am very thankful to Jane Chisholm for her classes and English corrections. She supported me to become a better English writer. I thank all of my internship supervisors who provided me with great opportunities so that I could work with interesting real-world data sets in locations around the world. I've been fortunate to have my Ph.D. friends who suported me through this journey and filled my daily life with good laughs. I am very grateful for having CRC near my office and staff members who provided me with great services everyday.

Last but not least, my sincere appreciation to my family members who give me endless love and support, whenever and wherever.

Indeed, for the last four years, Georgia Institute of Technology provided me with the most exciting adventures in the world. Thank you Georgia Tech!

Contents

List of Tables

List of Figures

SUMMARY

The advancement of sensor technology enables us to collect a massive amount of data and at the same time, poses a challenge of summarizing such data in useful features. In this thesis, the focus is on the summary of complex real-life signals that possess self-similarity, which indicates that the signal behaves similarly in a range of scales, or resolutions. Such signals can be well characterized with a scaling index, or self-similarity index, that represents essential scaling characteristics.

This thesis proposes four novel methods that facilitate and improve the assessment of scaling in signals based on non-decimated wavelet transform (NDWT). NDWT's are preferred to the standard orthogonal wavelet transforms in a number of data analytic tasks because of their time-invariance and redundancy. To facilitate NDWT, Chapter 2 of this thesis devises an NDWT matrix that efficiently maps an input signal from an acquisition domain to the wavelet domain with a simple matrix multiplication. Applying the proposed NDWT matrix provides four advantages: It is compressive in summarizing information, faster in computation, and flexible in processing inputs of any size. Such advantages of an NDWT matrix are illustrated with various example applications. An NDWT matrix is used for all NDWT transforms in Chapters 3 to 5. Chapter 3 introduces a method for scaling estimation based on an NDWT and its wavelet spectrum. The method utilizes a distinctive character of an NDWT that does not decimate wavelet coefficients, which enables us to obtain local spectra and more accurate scaling estimators. The method applied to simulated signals for which scaling is known in advance yields estimators with lower mean squared errors. An example application with mammographic images for breast cancer detection yields the best diagnostic accuracy in excess of 80%. In Chapter 4, we shift the focus to

some real-life signals for which theoretical scaling index is known and fixed. Based on Bayesian statistics, the method proposed in the chapter incorporates prior information about Hurst exponent H of signals to a likelihood function and estimate H with maximum a posteriori (MAP) estimation. The method yields estimators with lower mean squared errors even when the mean value of the prior distribution is slightly different from the true theoretical value. In the assessment of scaling of one-dimensional data based on NDWT, the regression method, which is standardly used, yields biased estimators because of autocorrelations present within wavelet coefficients. This autocorrelation is a result of redundancy of NDWT. Chapter 5 illustrates two robust methods for estimation of scaling that decrease the autocorrelation with a logarithmic transform and a resampling approach. The proposed methods yield lower mean square errors with decreased bias and the resulting estimators are asymptotically normal and unbiased with variance that is independent of the multiresolution levels.

Chapter I

INTRODUCTION

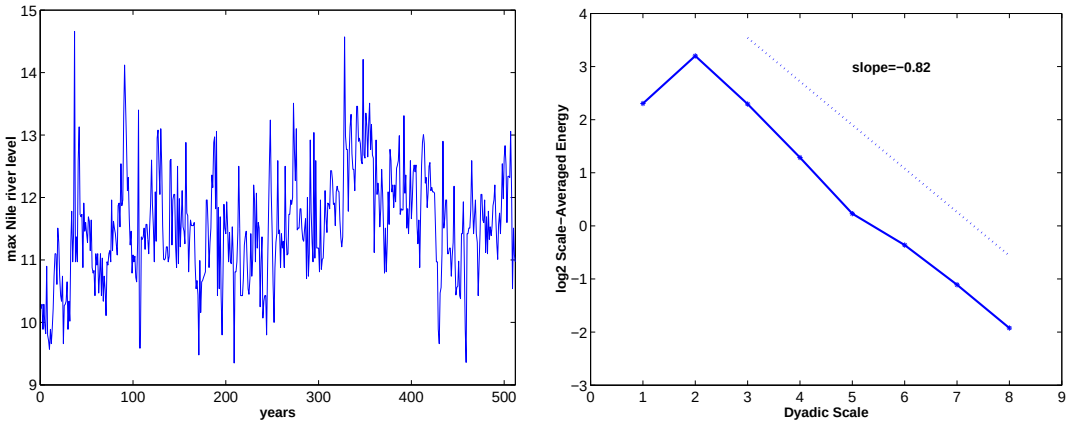
This chapter overviews statistical models and transforms that are pertinent for an understanding of the subsequent chapters. A number of models of random processes provide useful tools for analysis in various fields. We discuss some of traditional random processes that scale and are fundamental in assessing performance of methods developed in the following chapters. In addition, we overview multiscale analysis methods such as multiresolution analysis, wavelet transforms, and general time/frequency representations that will be domains in which the proposed estimators are defined.

1.1 Self-similar Processes

Theoretical self-similar processes such as fractional Brownian motion and processes with $1/f$ power-law spectra have become an essential tool for modeling a wide range of real-world signals that describe phenomena in engineering, physics, medicine, biology, economics, geology, chemistry, and so on. We often summarize the behavior of self-similar processes with the Hurst exponent, a self-similarity index that quantifies a measure of self-similarity in time series.

1.1.1 Hurst exponent

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's goal was to identify an optimal reservoir capacity R so that the reservoir could accept river flow in N units of time, $X_1, X_2, \ldots X_N$, and emit a constant flow 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. He obtained the optimal



(a) minimal level of Nile (yearly data) (b) wavelet log spectra of (a)

Figure 1: (a) presents consecutive yearly measurements (n=512) from the Nile River dataset for the years 62-1281 A.D. while (b) shows the wavelet spectra of (a), which demonstrates the scaling behavior.

volume of the reservoir, given by the so-called adjusted range

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

Hurst also reviewed other geophysical data because records of the water flow rarely lasted more than 100 years. To compare data from different sources, he standardized their adjusted ranges R with a sample standard deviation

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

and obtained dimensionless ratio R/S, which is a rescaled and adjusted range. Based on more than 800 records, Hurst discovered that the value of R/S scales at N^H, ranging from 0.46 to 0.93 with a mean of 0.73 and a standard deviation of 0.09. In contrast to Hurst's findings, Feller proved that the theoretical value of R/S was $1/2$ for independent and identically distributed random variables with a finite second moment (Feller (1951)). It was assumed that strong Markovian dependence was responsible for the deviation, which Hursts results showed. Later on Barnard proved that $H = 1/2$ holds for Markovian dependence cases (Barnard (1956)). Mandelbrot,

Mandelbrot, B.B. and Van Ness, and Mandelbrot and Wallis associated the Hurst (or Joseph) phenomenon in the presence of long memory (Mandelbrot, 1975; Mandelbrot and J. W. Van Ness, 1968; Mandelbrot and Wallis, 1969). The Hurst exponent is closely related to self-similar processes.

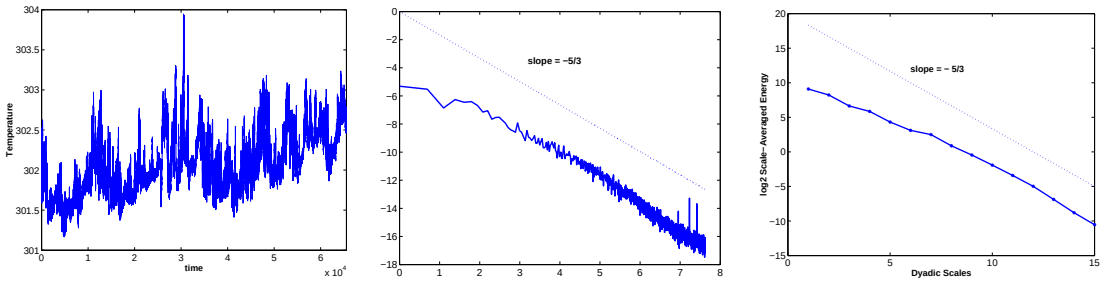
1.1.2 Examples of self-similar processes

We present some self-similar processes that scale in daily life. The degree of self-similarity of such processes is well represented with Hurst exponent H.

1.1.2.1 Turbulence

Here, we analyze an example of a real-life self-similar process, turbulence, which is measurements in turbulent phenomena. From July 12 to 16, 1995, the velocity and air temperature were measured at 5.2 m above the ground surface over an *Alta Fescue* grass site at the Blackwood division of the Duke Forest in Durham, North Carolina. A heat wave was present in North Carolina for several days when these measurements were taken until it swept from the Midwest to the East Coast. During the experiment in Durham, the maximum average air temperature was as high as $38°C$. The sky was clear with low to moderate winds during the measurement period. The measurement took place in a 480 m-by-305 m grass-covered forest clearing ($36°2'N$ $79°8'W$, elevation = 163 m) and a triaxial sonic anemometer mounted on a mast situated 250 m and 160 m north and west, respectively, of the edge of the 12 m tall Loblolly pine forest. The triaxial sonic anemometer (Gill Instruments/1012R2) collected the measurements of three velocity components (U, V, W) and air temperature T with sampling frequency (f_s) of 56 Hz and period (T_p) of 19.5 minutes, which yielded $N = 65,536$ measurements per velocity component per run.

The analysis of the measurement data collected in North Carolina was based on Kolmogorov's **K41** theory. In 1941, Kolmogorov devised the theory, which was based on the idea of energy cascading by Richardson, for the analysis of locally isotropic and



(a) Temperature (b) Scaling behavior in the Fourier domain (c) Scaling behavior in the wavelet domain

Figure 2: Temperature in different domains

fully developed turbulences. Since then, several models related to energy cascades have been proposed. Such models consider energy dissipation ϵ, [*energy per unit of fluid mass per unit time*], a key parameter that depends on the viscosity of kinematics, the components of velocity, and the position of moving fluid. A description of this transmission process follows: Energy is transmitted from large eddies, the source of energy injection, to small eddies, where the energy is converted into heat by viscosity and parameter ϵ. For the mathematical derivations of **K41** theory, let $\mathbf{x} = (x_1, x_2, x_3)$ as the position vectors and $\mathbf{u} = (u_1(\mathbf{x}), u_2(\mathbf{x}), u_3(\mathbf{x}))$ as the velocity components. When the flow in which the probability distribution of the relative velocity differences

$$\Delta\mathbf{u}(\mathbf{r}) = \mathbf{u}(\mathbf{x} + \mathbf{r}) - u(\mathbf{x}),$$

is independent of time and invariant under translations, reflections, and rotations, it can be referred to as locally isotropic turbulence. The principal objects of **K41** theory are *structure functions*. Structure functions are closely related to the correlations of two-point velocity differences,

$$\langle \Delta\mathbf{u}(\mathbf{r})^2 \rangle = 2\sigma_u^2(1 - \rho_u(r)).$$

The definition of a (longitudinal) structure function of order p is

$$D_p(r) = \langle ||\Delta\mathbf{u}(\mathbf{r})||^p \rangle,$$

where the angular brackets represent time averaging. We derive a functional description for the moments of velocity differences, or structure functions, with *dimensional analysis* and obtain

$$D_p(r) = C_p[\langle \epsilon \rangle \ r]^{\frac{p}{3}}, \tag{3}$$

where C_p is a universal constant. From Navier-Stokes equations, we can directly infer that $C_3 = -\frac{4}{5}$ for the third-order structure function. From (3), we know that the structure functions exhibit scaling behavior. Then,

$$D_p(r) \propto r^{\zeta_p},$$

where the exponent ζ_p is called the *scaling exponent* and $\propto$ means "proportional to." The **K41** theory yields the simple model $\zeta_p = \frac{p}{3}$. Similarly, with respect to the structure functions, the energy of the turbulent fluctuations per unit of mass of fluid in scales r can be derived from the hypotheses. By dimensional analysis,

$$E_r \propto (r)^{\frac{2}{3}}. \tag{4}$$

With the Fourier transform of E_r, which yields the spectral density $\phi(k)$, we obtain the introduced "$-\frac{5}{3}$ law" for the power spectrum, as shown in Figure 2

$$\begin{aligned} E_k &= 2R^{-1}k^2\phi(k) \\ &\propto k^{-\frac{5}{3}}. \end{aligned}$$

From this,

$$\log E_k = -\frac{5}{3}\log k + C,$$

where C is a constant independent of k. We will see later that this is connected with Hurst exponent $H = 1/3$.

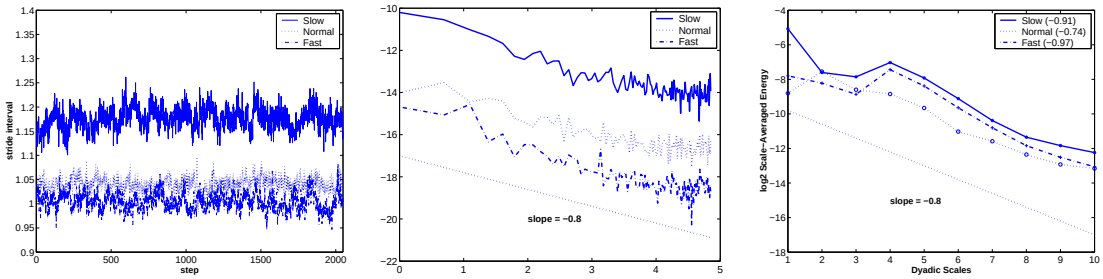
1.1.2.2 Gait Data

Dynamical analysis of gait data revealed that the stride interval (duration of the gait cycle) of human walking exhibit scaling behaviors. The analysis indicates that fluctuations at one time scale are statistically similar to those at multiple other time scales, when more than hundreds steps are measured from healthy human subjects walking 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. Interestingly, long-range correlations disappeared during metronomically-paced walking, that is, variations in the stride interval were anti-correlated. Other experiments confirmed the result that scaling behavior of spontaneous stride interval are normally quite robust and intrinsic to the locomotor system. In addition, scaling behavior of neural output may have some connection to the higher nervous system responsible for walking rhythm.

The gait data is collected from participants who are not under any medication and do not have medical history of neuromuscular, respiratory or cardiovascular disorders. Average age of participants was 21.7 years and a range of age was 18-29 years. A range of height was 1.77 ± 0.08 meters, and a range of weight was 71.8 ± 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 indicates 2048 data points collected from one of participants. We can easily find that slow and fast stride intervals have slopes of -0.91 and -0.97, respectively, while stride intervals for normal walk has a slope of -0.74. Such slopes are closely related to the degree of self-similarity, Hurst exponent H. We will explain the relationship of slopes and H later.



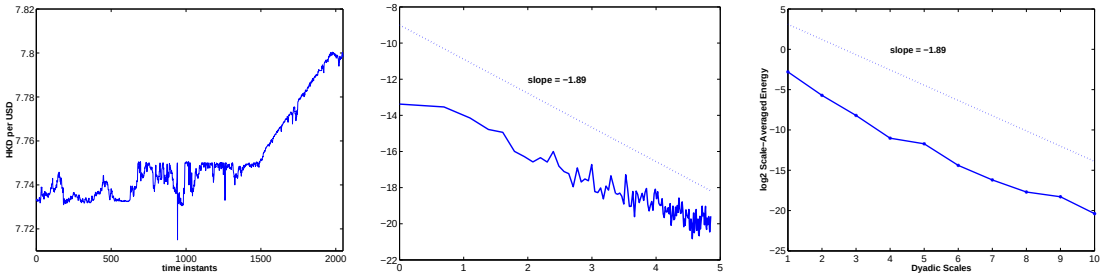
(a) (b) (c)

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

1.1.2.3 Rate of exchange

A number of economic time series, such as stock market prices, exchange rates, and asset returns possess self-similar property and long range dependence (LRD). Such fact is in empirical contradiction to several economic theories such as random walk theory for stock market, perfect markets, and so on. Thus, several modeling tools for scaling and LRD processes such as ARFIMA, fGn, fBm, GARCH, and so on. are developed.

We present rates of exchange between Hong Kong Dollar (HKD) and US Dollar (USD) reported by the ONADA Company between 24 March 1995 and 1 November 2000. In Figure 4, we can find the scaling behavior of the exchange rate.



(a) (b) (c)

Figure 4: (a) Exchange Rates HKD per US$; (b) scaling behavior in the Fourier domain, and (c) in the wavelet domain.

1.1.2.4 DNA Random Walks

One of principal tasks in the analysis of DNA sequences is finding if two sequences are related. This is studied by using a scoring system, which ranks the possible relations between the sequences and consider statistical significance of each relation (Durbin et al., 1998).

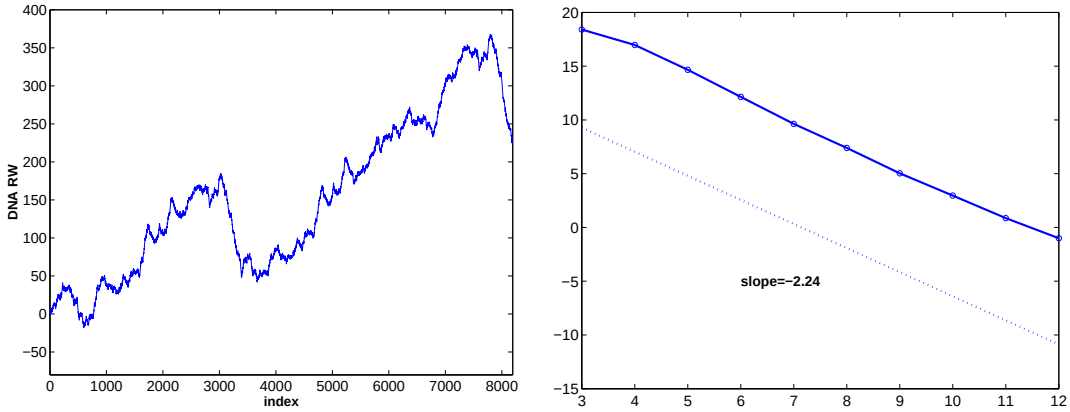
Bacry et al. (1996) wrote that the analysis of DNA walks is affected by the presence of a global linear trend induced by the excess of purines over pyramidines. Such trend is a singular process that can be assessed for long-range correlations and related properties.

A DNA molecule consists of long complementary double helix of purine nucleotides (denoted as A and G) and pyrimidine nucleotides (denoted as C and T). We can translate the a strain of DNA as a sequence of alphabet that corresponds to a random walk. By the letter at position i in the sequence, the random walk increase $x(i) = 1$ for A and G, and decrease $x(i) = -1$ for C and T. The random walk is defined as a cumulative sum of that increment or decrement $s(n) = \sum_{i=1}^{n} x(i)$, where n is an index smaller than the length of the sequence. Peng et al. (1994) first proposed DNA walks.

Bacry et al. (1996) studied self-similarity and fractality of DNA walks for humans. It is found that that the Hurst exponent for introns (non-coding sequences) is about 0.6, while for exons (coding sequences), the Hurst exponent is close to 0.5.

1.1.3 Basics of Self-similar Processes

We provide the definitions and properties of self-similar processes. Then we dicuss types of analysis that estimate the scaling and self-similarity properties. To understand self-similar processes, we begin by an discussion of some basic concepts and definitions. All the stochastic processes presented are real-valued and defined in a general probability space $(\Omega, \mathcal{F}, P)$.



(a) (b)

Figure 5: (a) 8196-long DNA random walk for a spider monkey, from EMLB Nucleotide sequence alignment DNA database; (b) wavelet scaling with the slope -2.24.

Definition 1.1.1. $X(t) \stackrel{d}{=} Y(t)$ represents equality in all finite dimensional distributions of stochastic processes $X(t)$ and $Y(t)$.

The following definitions are building blocks for an understanding of self-similar processes.

Definition 1.1.2. If $\lim_{h\to 0} P(|X(t+h) - X(t)| > \epsilon) = 0$, for any $\epsilon > 0$, then the process $X(t), t \geq 0$ is stochastically continuous at t.

Definition 1.1.3. A stochastic process $X(t)$, $t \geq 0$ is self-similar if for any $a > 0$, there exists $b > 0$ such that

$$X(at) \stackrel{d}{=} bX(t) \tag{5}$$

A stochastic process is trivial when its distribution is defined as a point measure. For example, any single real-valued random-variable process $X(t) = const$ is a trivial stochastic process. Following Lamperti (1962),

Theorem 1.1.1. If a stochastic process $X(t), t > 0$ is non-trivial, stochastically continuous at $t = 0$, and self-similar, then there exists a unique $H > 0$ such that b

in Definition 1.1.3 can be expressed as $b = a^H$. In addition, $H > 0$ if and only if $X(0) = 0$ (almost surely)

A self-similar process is generally defined as follows.

Definition 1.1.4. If there exists $H > 0$ such that for any choice of $a > 0$, $X(at) \stackrel{d}{=} bX(t)$. Then it follows that $X(0) = 0$ almost surely and the stochastic process $X(t)$ is self-similar.

In this definition, the uniqueness of H is not explicitly given even though Theorem 1.1.1 guarantees uniqueness. The phenomena referred to as long-range dependence is closely related to self-similarity.
To understand long-range dependence, we start by introducing the stationarity of a process.

Definition 1.1.5. A stochastic process $X(t)$ is strictly stationary if given n time point $t_1, \dots t_n$ for any n and lag $k > 0$,

$$X(t_1), \dots, X(t_n) \stackrel{d}{=} X(t_1 + k), \dots, X(t_n + k) \tag{6}$$

This condition of the process is strong because it forces all moments of all degrees of the process such as expectations, variances, and third- or higher-order quantities to be the same at any time point. One can also express a stationary process with a spectral representation and the auto-covariance of the process in terms of the spectrum of the process based on the Wiener-Khinchine theorem. Based on the assumption that $\mathbb{E}|X(t)|^2 < \infty$, the strict stationarity of $X(t)$ implies the following:

(i) $\mathbb{E}(X(t)) = \mu$ for all $t \in \mathbb{R}$,

(ii) $COV(X(t+h), X(t)) = \gamma_X(h) = \mathbb{E}\big[(X(t+h) - \mathbb{E}[X(t+h)])(X(t) - \mathbb{E}[X(t)])\big]$, which is independent of time t but dependent on temporal difference h,

(iii) the auto-covariance function of $X(t)$ is $\gamma_X(h) = COV(X(t+h), X(t))$,

(iv) the autocorrelation at lag τ is $\rho(\tau) = \gamma(\tau)/\gamma(0)$, where $\gamma(0) = cov(X(t), X(t)) = var(X(t))$.

When $X(t), t \geq 0$ is self-similar with Hurst exponent H and when the distributions of $X(t+h) - X(t)$ are independent of t (i.e., stationary increment), $X(t)$ is referred to as $H - sssi$, for short. Relaxing the previous conditions, stochastic processes that satisfy (i) and (ii) are referred to as weakly stationary or second-order stationary.

Definition 1.1.6. A second-order stationary stochastic process $X(t)$ has long-range dependence with parameter α, if for a constant $c_\gamma > 0$, its autocovariance function is

$$\gamma_x(k) \sim c_\gamma |k|^{-(1-\alpha)}, \alpha \in (0, 1)$$

Intuitively, when a stationary process X has long-range dependence, long memory, or strong dependence, its auto-covariance function, $\gamma(\cdot)$, has slow decays that

$$\sum_{i=0}^{\infty} \gamma(i) = \infty.$$

Unlike short-range dependence, a process with long-range dependence has a significant amount of cumulative correlation effect. Generally speaking, self-similarity and long-range dependence are not identical concepts but under the condition that $1/2 < H < 1$, long-range dependence becomes a equivalent concept to asymptotic second-order self-similarity for stationary processes.

1.2 Fractional Brownian Motion

Fractional Brownian motions (fBms) and fractional Brownian fields (fBfs) are Gaussian $H-sssi$ processes with stationary increments. They scale regularly and represent examples of monofractal objects with a singular scaling index, Hurst exponent H. For a mathematical representation, let us denote path of a one-dimensional (1-D) fBm process with Hurst exponent H as $\{B_H(t),\, t \in \mathbb{R}\}$. By definition of self-similarity of

fBm, it holds that $B_H(at)$ is equal in distribution to $a^H B_H(t),\ a \geq 0$. The covariance function of $B_H(t)$ is

$$\gamma_{B_H}(t,s) = \mathbb{E}\{B_H(t)B_H(s)\} = \frac{\sigma_H^2}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right),\ s,\ t \in \mathbb{R}, \qquad (7)$$

where $\sigma_H^2 = \Gamma(1-2H)\frac{cos(\pi H)}{\pi H}$, and $0 \leq H \leq 1$. Because $\gamma_{B_H}(t,s)$ is not a function of $|t-s|$ only, the fBm is non-stationary. We cannot obtain a spectrum of fBm by a direct definition. However, we can indirectly deduce a pseudo-spectrum from the fact that increments of fBm are stationary Reed et al. (1995)

$$S_{B_H}(\omega) \propto |\omega|^{-2H-1},$$

where ω indicates frequency in the power spectrum. This definition extends to a two-dimensional (2-D) fractional Brownian motion, or equivalently, fractional Brownian field (fBf), $B_H(\boldsymbol{u})$, where $\boldsymbol{u}$ and $\boldsymbol{v}$ are points in 2-D space $[0,1]\times[0,1]$. The covariance function of $B_H(\boldsymbol{u})$ is

$$\gamma_{B_H}(\boldsymbol{u},\boldsymbol{v}) = \mathbb{E}\{B_H(\boldsymbol{u})B_H(\boldsymbol{v})\} = \frac{\sigma_H^2}{2}[\ ||\boldsymbol{u}||^{2H} + ||\boldsymbol{v}||^{2H} - ||\boldsymbol{u}-\boldsymbol{v}||^{2H}\ ], \qquad (8)$$

where $||\cdot||$ represents the Euclidean norm, and $\sigma_H^2 = \frac{2^{-(1+2H)}\Gamma(1-H)}{\pi H\Gamma(1+H)}$. As a result, the relationship between the power spectrum and Hurst exponent H is

$$S_{B_H}(\omega) \propto ||\omega||^{-2H-2}.$$

1.3 Basics of Wavelets

The first theoretical results in wavelets were reported in the early 1980s in connection with the continuous wavelet decompositions of square integrable functions. We will denote the space of all squared integrable functions $\mathbb{L}_2$, that is, functions for which $\int f^2 < \infty$. The seminal papers on these subjects are Grossmann and Morlet (1985); Morlet et al. (1982). We define $\psi_{a,b}(x),\ a \in \mathbb{R}\backslash\{0\}, b \in \mathbb{R}$ as a family of functions that are translations and re-scales of a single function, $\psi(x) \in \mathbb{L}_2(\mathbb{R})$,

$$\psi_{a,b}(x) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{x-b}{a}\right). \qquad (9)$$

With a normalization factor, $\frac{1}{\sqrt{|a|}}$, $||\psi_{a,b}(x)||$ is valid for any choice of a and b. The wavelet function ψ (also known as *the mother wavelet*) is assumed to meet the *admissibility condition*,

$$C_\psi = \int_{\mathbb{R}} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty, \tag{10}$$

where $\Psi(\omega) = \int_R \psi(x) e^{-ix\omega} dx$ is the Fourier transformation of $\psi(x)$. An implication of the admissibility condition (10) is

$$0 = \Psi(0) = \int \psi(x) dx.$$

In addition, if $\int \psi(x)dx = 0$ and $\int (1 + |x|^\alpha)|\psi(x)|dx < \infty$ for some $\alpha > 0$, then $C_\psi < \infty$.

Using the continuous wavelet transformation, we can define any function $f(x)$ in $\mathbb{L}_2$ as a function of two variables

$$\mathcal{CWT}_f(a,b) = \langle f, \psi_{a,b} \rangle = \int f(x) \overline{\psi_{a,b}(x)} dx.$$

In the above equation, dilation (a) and translation (b) parameters can vary continuously in $\mathbb{R}\backslash\{0\} \times \mathbb{R}$. The following paragraphs present several important properties of the continuous wavelet transformation.

Shifting Property. The shift in a function in the time domain does not affect continuous wavelet transformation. For example, if $f(x)$ undergoes continuous wavelet transformation $\mathcal{CWT}_f(a,b)$, then the shifted function of $f(x)$, $g(x) = f(x-\beta)$ also has the same continuous wavelet transformation $\mathcal{CWT}_g(a,b) = \mathcal{CWT}_f(a, b-\beta)$.
Scaling Property. The scaling of a function in the time domain does not affect continuous wavelet transformation. For example, if $f(x)$ undergoes continuous wavelet transformation $\mathcal{CWT}_f(a,b)$, then $g(x) = \frac{1}{\sqrt{s}} f\left(\frac{x}{s}\right)$ has the same continuous wavelet transformation $\mathcal{CWT}_g(a,b) = \mathcal{CWT}_f\left(\frac{a}{s}, \frac{b}{s}\right)$.

One can simply prove the shifting and the scaling property by changes in variables before performing integration for the continuous wavelet transform.

Example 1.3.1. The "Mexican hat" or "Marr's" wavelet is defined as

$$\psi(x) = \frac{d^2}{dx^2}[-e^{-x^2/2}] = (1 - x^2)e^{-x^2/2},$$

where $C_\psi = 2\pi$ can be obtained by direct calculation.

When a function of one variable is transformed by continuous wavelet transformation, the resulting function consists of two variables indicating redundancy in the transformation. To control the amount of information produced from the transformation, which maintains invertibility, one can choose the discrete values of a and b, which are used in the transformation. However, a sampling rate that preserves all information about the decomposed function should be greater than *critical sampling.* The critical sampling (Figure 6) that produces the minimal basis is defined by

$$a = 2^{-j},\ b = k2^{-j},\ j, k \in \mathbb{Z}. \tag{11}$$

If the sampling rate becomes coarser than the critical sampling rate, a unique inverse transformation does not exist, which means that one cannot fully recover the original function. In addition, the given sampling rate yields an orthogonal basis $\{\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k),\ j, k \in \mathbb{Z}\}$ under some conditions on wavelet function ψ. Other options for the sampling rate exist. As an example, by selecting $a = 2^{-j},\ b = k$, one can obtain non-decimated (or stationary) wavelets. More general sampling rates are given by

$$a = a_0^{-j},\ b = k\, b_0\, a_0^{-j},\ j, k \in \mathbb{Z},\ a_0 > 1, b_0 > 0. \tag{12}$$

Reconstructions become numerically stable when the system $\{\psi_{jk},\ j, k \in \mathbb{Z}\}$ constitutes a frame. At (12), we evaluate (9) as

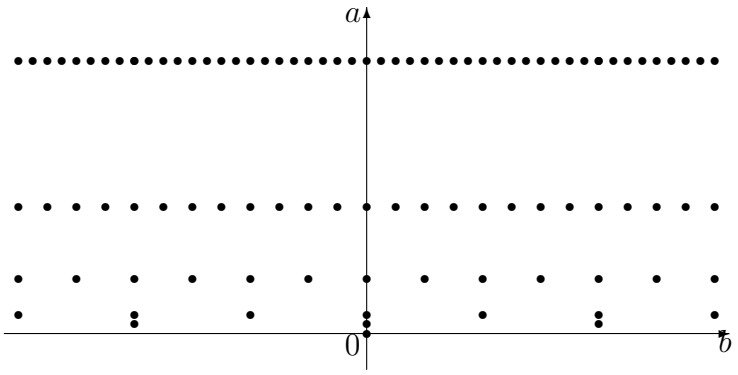


Figure 6: Critical Sampling in the $\mathbb{R} \times \mathbb{R}^+$ half-plane ($a = 2^{-j}$ and $b = k\ 2^{-j}$).

$$\psi_{jk}(x) = a_0^{j/2}\,\psi\left(\frac{x - k\,b_0\,a_0^{-j}}{a_0^{-j}}\right) = a_0^{j/2}\psi(a_0^j x - k\,b_0).$$

1.4 Multiresolution Analysis

A multiresolution analysis (MRA) is the nested sequence of closed subspaces $V_n, n \in \mathbb{Z}$ in $\mathbb{L}_2(\mathbb{R})$ with a containment hierarchy

$$\cdots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \cdots . \tag{13}$$

The intersection of such nested spaces contains only zero functions and their union is dense in $\mathbb{L}(\mathbb{R})$,

$$\cap_n V_j = \{\mathbf{0}\}, \quad \overline{\cup_j V_j} = \mathbb{L}_2(\mathbb{R}),$$

where the closure of set A is denoted as $\overline{A}$. The hierarchy written in (13) is constructed so that the following two conditions are met: First, V-spaces are self-similar, which means that

$$f(2^j x) \in V_j \text{ iff } f(x) \in V_0. \tag{14}$$

Second, *a scaling function* $\phi \in V_0$, whose integer-translates span the space V_0, exists as

$$V_0 = \left\{ f \in \mathbb{L}_2(\mathbb{R}) |\ \ f(x) = \sum_k c_k \phi(x-k) \right\},$$

where the set $\{\phi(\bullet - k),\ k \in \mathbb{Z}\}$ is an orthonormal basis for V_0. We present some technical conditions on ϕ. We assume that $\int \phi(x)dx \geq 0$, and this integral, in fact, equals 1, which is proven in the subsequent section. The function $\phi(x) \in V_0$ is represented as the linear combination of functions from V_1, i.e.,

$$\phi(x) = \sum_{k\in\mathbb{Z}} h_k\ \sqrt{2}\phi(2x-k), \tag{15}$$

for some coefficients $h_k,\ k \in \mathbb{Z}$, because $V_0 \subset V_1$. This equation, referred to as *scaling equation* (or two-scale equation), is indeed a fundamental element of wavelets for construction and utilization. Note that one may find another indexing of multi-resolution subspaces, that is, the reverse of that in (13)

$$\cdots \subset V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \subset \cdots . \tag{16}$$

We further explore the properties of multiresolution analysis subspaces and their bases in the Fourier domain. We define the function m_0 as

$$m_0(\omega) = \frac{1}{\sqrt{2}} \sum_{k\in\mathbb{Z}} h_k e^{-ik\omega} = \frac{1}{\sqrt{2}} H(\omega). \tag{17}$$

(17) is sometimes referred to as the *transfer function*, which describes the behavior of associated filter $\mathbf{h} = \{h_k, k \in \mathbb{Z}\}$ in the Fourier domain. The period of function m_0 is 2π and the Fourier coefficients of function $H(\omega) = \sqrt{2}\ m_0(\omega)$ are filter taps $\{h_n,\ n \in \mathbb{Z}\}$. In addition, (15) becomes

$$\Phi(\omega) = m_0\left(\frac{\omega}{2}\right) \Phi\left(\frac{\omega}{2}\right), \tag{18}$$

where $\Phi(\omega)$ is the Fourier transformation of $\phi(x)$ in the Fourier domain. Because

$$\begin{aligned}
\Phi(\omega) &= \int_{-\infty}^{\infty} \phi(x) e^{-i\omega x} dx \\
&= \sum_k \sqrt{2}\, h_k \int_{-\infty}^{\infty} \phi(2x-k) e^{-i\omega x} dx \\
&= \sum_k \frac{h_k}{\sqrt{2}} e^{-ik\omega/2} \quad \int_{-\infty}^{\infty} \phi(2x-k) e^{-i(2x-k)\omega/2} d(2x-k) \\
&= \sum_k \frac{h_k}{\sqrt{2}} e^{-ik\omega/2} \quad \Phi\left(\frac{\omega}{2}\right) \\
&= m_0\left(\frac{\omega}{2}\right) \Phi\left(\frac{\omega}{2}\right).
\end{aligned}$$

We iterate (18) and obtain

$$\Phi(\omega) = \prod_{n=1}^{\infty} m_0\left(\frac{\omega}{2^n}\right). \tag{19}$$

(19) converges under some mild conditions on the rates of decay of scaling function ϕ.

In connecting the MRA to the theory of signal processing, coefficient h_k in (15) are important. Vector $\mathbf{h} = \{h_k, k \in \mathbb{Z}\}$ is referred to as a wavelet filter, that is, a low-pass (averaging) filter. We prove the *normalization* and *orthogonality* properties of wavelet filters.

Normalization.

$$\sum_{k\in\mathbb{Z}} h_k = \sqrt{2}. \tag{20}$$

Proof:

$$\begin{aligned}
\int \phi(x) dx &= \sqrt{2} \sum_k h_k \int \phi(2x-k) dx \\
&= \sqrt{2} \sum_k h_k \, \frac{1}{2} \int \phi(2x-k) d(2x-k) \\
&= \frac{\sqrt{2}}{2} \sum_k h_k \int \phi(x) dx.
\end{aligned}$$

Because $\int \phi(x)dx \neq 0$, which is based on the assumption, (20) follows.

Orthogonality. For any $l \in \mathbb{Z}$,

$$\sum_k h_k h_{k-2l} = \delta_l. \tag{21}$$

Proof: From the scaling equation, (15), it follows that

$$\begin{aligned} \phi(x)\phi(x-l) &= \sqrt{2}\ \sum_k h_k \phi(2x-k)\phi(x-l) \\ &= \sqrt{2}\ \sum_k h_k \phi(2x-k)\ \sqrt{2}\ \sum_m h_m \phi(2(x-l)-m). \end{aligned} \tag{22}$$

We integrate both sides of (22) to obtain

$$\begin{aligned} \delta_l &= 2\sum_k h_k \left[\sum_m h_m \frac{1}{2}\int \phi(2x-k)\phi(2x-2l-m)\, d(2x)\right] \\ &= \sum_k \sum_m h_k h_m \delta_{k,2l+m} \\ &= \sum_k h_k h_{k-2l}. \end{aligned}$$

By taking $k = 2l+m$, we obtain the last line. When $l = 0$, (21) becomes an important and special case,

$$\sum_k h_k^2 = 1. \tag{23}$$

In the Fourier domain, we can show that system $\{\phi(\bullet - k),\ k \in \mathbb{Z}\}$ constitutes an orthonormal basis for V_0 with respect to either $\Phi(\omega)$ or $m_0(\omega)$.

(**a**) Using $\Phi(\omega)$, we express

$$\sum_{l=-\infty}^{\infty} |\Phi(\omega + 2\pi l)|^2 = 1. \tag{24}$$

The [PAR] property of the Fourier transformation and the 2π-periodicity of $e^{i\omega k}$ yields

$$\begin{aligned} \delta_k &= \int_{\mathbb{R}} \phi(x)\overline{\phi(x-k)}dx \\ &= \frac{1}{2\pi}\int_{\mathbb{R}} \Phi(\omega)\overline{\Phi(\omega)}e^{i\omega k}d\omega \\ &= \frac{1}{2\pi}\int_0^{2\pi} \sum_{l=-\infty}^{\infty} |\Phi(\omega+2\pi l)|^2 e^{i\omega k}d\omega. \end{aligned} \tag{25}$$

The final line in (25) is Fourier coefficient a_k when one performs the Fourier series decomposition of

$$f(\omega) = \sum_{l=-\infty}^{\infty} |\Phi(\omega + 2\pi l)|^2.$$

The uniqueness of Fourier representation ensures that $f(\omega) = 1$.

(**b**) In terms of m_0 :

$$|m_0(\omega)|^2 + |m_0(\omega + \pi)|^2 = 1. \tag{26}$$

Because $\sum_{l=-\infty}^{\infty} |\Phi(2\omega + 2l\pi)|^2 = 1$, by (18),

$$\sum_{l=-\infty}^{\infty} |m_0(\omega + l\pi)|^2 |\Phi(\omega + l\pi)|^2 = 1. \tag{27}$$

We divide the sum of (27) into two sums – the first sum with odd and the second sum with even indices, which yield

$$\begin{aligned} 1 \;\; = \;\; & \sum_{k=-\infty}^{\infty} |m_0(\omega + 2k\pi)|^2 |\Phi(\omega + 2k\pi)|^2 + \\ & \sum_{k=-\infty}^{\infty} |m_0(\omega + (2k+1)\pi)|^2 |\Phi(\omega + (2k+1)\pi)|^2. \end{aligned}$$

We use relation (24) and the 2π-periodicity of $m_0(\omega)$ and simplify the above expression.

$$\begin{aligned} 1 \;\; &= \;\; |m_0(\omega)|^2 \sum_{k=-\infty}^{\infty} |\Phi(\omega + 2k\pi)|^2 + |m_0(\omega + \pi)|^2 \sum_{k=-\infty}^{\infty} |\Phi((\omega + \pi) + 2k\pi)|^2 \\ &= \;\; |m_0(\omega)|^2 + |m_0(\omega + \pi)|^2. \end{aligned}$$

Based on the properties of the wavelet filter, we derive wavelet functions. If a sequence of subspaces satisfies MRA properties, at least one orthonormal basis for $\mathbb{L}_2(\mathbb{R})$ (though not unique) exists as

$$\{\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k), \;\; j, k \in \mathbb{Z}\}. \tag{28}$$

$\psi_{jk}(x)$, where j is fixed and $k \in \mathbb{Z}$, is the orthonormal basis of the "difference space" $W_j = V_{j+1} \ominus V_j$. The function $\psi(x) = \psi_{00}(x)$ is formally referred to as a *wavelet function* or casually as *mother wavelet*. A wavelet function is derived from the scaling function. Because of the containment $W_0 \subset V_1$, $\psi(x) \in V_1$, the wavelet function is represented as

$$\psi(x) = \sum_{k\in\mathbb{Z}} g_k \ \sqrt{2}\phi(2x-k), \tag{29}$$

for some coefficients $g_k,\ k \in \mathbb{Z}$. Let us define

$$m_1(\omega) = \frac{1}{\sqrt{2}} \ \sum_k g_k e^{-ik\omega}. \tag{30}$$

We obtain the Fourier counterpart of (29) by following the same procedure of m_0,

$$\Psi(\omega) = m_1(\frac{\omega}{2})\Phi(\frac{\omega}{2}). \tag{31}$$

By the rules of construction, the spaces W_0 and V_0 are orthogonal. Therefore,

$$\begin{aligned} 0 = \int \psi(x)\phi(x-k)dx &= \frac{1}{2\pi}\int \Psi(\omega)\overline{\Phi(\omega)}e^{i\omega k}d\omega \\ &= \frac{1}{2\pi}\int_0^{2\pi} \sum_{l=-\infty}^{\infty} \Psi(\omega+2l\pi)\overline{\Phi(\omega+2l\pi)}e^{i\omega k}d\omega. \end{aligned}$$

Following the same logic as in (24), we conclude that

$$\sum_{l=-\infty}^{\infty} \Psi(\omega+2l\pi)\overline{\Phi(\omega+2l\pi)} = 0.$$

Considering the definitions of m_0 and m_1 and following the same steps for deriving (26), we find

$$m_1(\omega)\overline{m_0(\omega)} + m_1(\omega+\pi)\overline{m_0(\omega+\pi)} = 0. \tag{32}$$

Based on (32), we conclude that there exists a function $\lambda(\omega)$ such that

$$(m_1(\omega),\ m_1(\omega+\pi)\ ) = \lambda(\omega)\left(\overline{m_0(\omega+\pi)},\ -\overline{m_0(\omega)}\ \right). \tag{33}$$

We substitute $\xi = \omega + \pi$ and use the 2π-periodicity of m_0 and m_1 to conclude that

$$\begin{aligned} \lambda(\omega) &= -\lambda(\omega + \pi), \text{ and} \\ \lambda(\omega) &\quad \text{is } 2\pi\text{-periodic.} \end{aligned} \tag{34}$$

Any function $\lambda(\omega)$ can satisfy (32) as long as their forms are in $e^{\pm i\omega} S(2\omega)$, where S is an $\mathbb{L}_2([0, 2\pi])$ and a 2π-periodic function. However, only functions with $|\lambda(\omega)| = 1$ can define an orthogonal basis ψ_{jk} of $\mathbb{L}_2(\mathbb{R})$. The summary of the conditions of $\lambda(\omega)$ are

(i) $\lambda(\omega)$ is 2π-periodic,

(ii) $\lambda(\omega) = -\lambda(\omega + \pi)$, and

(iii) $|\lambda(\omega)|^2 = 1$.

Although any function satisfying the above three conditions can create a valid m_1, standard choices for $\lambda(\omega)$ are $-e^{-i\omega}$, $e^{-i\omega}$, and $e^{i\omega}$. We define $m_1(\omega)$ as

$$m_1(\omega) = -e^{-i\omega} \overline{m_0(\omega + \pi)} \tag{35}$$

so that it creates a convenient and standard connection between filters $\mathbf{h}$ and $\mathbf{g}$. $\{\psi(\bullet - k),\ k \in \mathbb{Z}\}$ is an orthonormal basis for W_0, which is implied in the form of m_1 and equation (24). Using $|m_1(\omega)| = |m_0(\omega + \pi)|$, we can rewrite the orthogonality condition (26) as

$$|m_0(\omega)|^2 + |m_1(\omega)|^2 = 1. \tag{36}$$

We compare the definition of m_1 in (30) with

$$\begin{aligned} m_1(\omega) &= -e^{-i\omega} \frac{1}{\sqrt{2}} \sum_k h_k e^{i(\omega+\pi)k} \\ &= \frac{1}{\sqrt{2}} \sum_k (-1)^{1-k} h_k e^{-i\omega(1-k)} \\ &= \frac{1}{\sqrt{2}} \sum_n (-1)^n h_{1-n} e^{-i\omega n}. \end{aligned}$$

As a result, we can relate the elements of filters **h** and **g** as

$$g_n = (-1)^n \; h_{1-n}. \tag{37}$$

Such a relation (37) is often referred to as the *quadrature mirror relation*, and filters **h** and **g** are referred to as *quadrature mirror filters* in the signal processing literature.

Remark 1.4.1. Selecting $\lambda(\omega) = e^{i\omega}$ yields high-pass filter $g_n = (-1)^{n-1} \; h_{-1-n}$, which is rarely used. Defining g_n as $(-1)^n h_{1-n+M}$, where M is a "shift constant," is sometimes convenient. Since such a re-definition of **g** affects only the shift-location of the wavelet function.

1.5 Mallat's Algorithm

Burt and Adelson invented orthogonal and biorthogonal pyramid algorithms for image processing Burt and Adelson (1983a,b). In a tree-like fashion, pyramid algorithm process an image at a sequence of scales which range from fine to coarse in cascade. In each scale, processed images can be denoised, improved, or compressed through different applications.

Mallat was a pioneer in formally linking wavelets, multiresolution analyses, and cascade algorithms Mallat (1989b,a). Mallat's cascade algorithm elucidate a constructive and efficient recipe that performs the discrete wavelet transformation. It filters data with **h** and **g** in cascade, so that the wavelet coefficients from different scales in the transformation can be related to one another. In fast Fourier transformations, Danielson-Lanczos algorithm takes a similar role as Mallat's algorithm in wavelet transform.

It is useful to understand the relationship between the original signal with the space coefficients from the space V_J, for some J. Two complementing spaces, smooth and detail spaces are pairs of (V_{J-1}, W_{J-1}), (V_{J-2}, W_{J-2}), and etc. Decreasing the index in V-spaces coarsens the approximation of the data.

By a direct substitution of indices in the scaling equations (15) and (29), one obtains

$$\phi_{j-1,l}(x) = \sum_{k\in\mathbb{Z}} h_{k-2l}\phi_{jk}(x) \quad \text{and} \quad \psi_{j-1,l}(x) = \sum_{k\in\mathbb{Z}} g_{k-2l}\phi_{jk}(x). \tag{38}$$

This relationship becomes a building block of the cascade algorithm.

In a multiresolution analysis $\cdots \subset V_{j-1} \subset V_j \subset V_{j+1} \subset \ldots$, because Since $V_j = V_{j-1} \oplus W_{j-1}$, any function $v_j \in V_j$ can be represented uniquely as $v_j(x) = v_{j-1}(x) + w_{j-1}(x)$, where $v_{j-1} \in V_{j-1}$ and $w_{j-1} \in W_{j-1}$. It is common to denote the coefficients related to $\phi_{jk}(x)$ and $\psi_{jk}(x)$ by c_{jk} and d_{jk}, respectively.

Therefore,

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

We use the general scaling equations (38), orthogonality of $w_{j-1}(x)$ and $\phi_{j-1,l}(x)$ for any j and l, and additivity of inner products, to obtain

$$\begin{aligned} c_{j-1,l} &= \langle v_j, \phi_{j-1,l}\rangle \\ &= \langle v_j, \sum_k h_{k-2l}\phi_{j,k}\rangle \\ &= \sum_k h_{k-2l}\langle v_j, \phi_{j,k}\rangle \\ &= \sum_k h_{k-2l}c_{j,k}. \end{aligned} \tag{39}$$

Similarly $d_{j-1,l} = \sum_k g_{k-2l}c_{j,k}$.

The cascade algorithm is still valid in the reverse direction. Coefficients that belong to the next detailed scale in V_j can be obtained from the coefficients corresponding to V_{j-1} and W_{j-1}. The relation below specify a single iteration of the

reconstruction algorithm.

$$\begin{aligned} c_{j,k} &= \langle v_j, \phi_{j,k} \rangle \\ &= \sum_l c_{j-1,l} \langle \phi_{j-1,l}, \phi_{j,k} \rangle + \sum_l d_{j-1,l} \langle \psi_{j-1,l}, \phi_{j,k} \rangle \\ &= \sum_l c_{j-1,l} h_{k-2l} + \sum_l d_{j-1,l} g_{k-2l}, \end{aligned} \tag{40}$$

Example 1.5.1. Considering DAUB2, the scaling equation at integer k is

$$\phi(n) = \sum_{k=0}^{3} h_k \sqrt{2} \phi(2n-k).$$

Recall that $\mathbf{h} = \{h_0, h_1, h_2, h_3\} = \{\frac{1+\sqrt{3}}{4\sqrt{2}}, \frac{3-\sqrt{3}}{4\sqrt{2}}, \frac{3+\sqrt{3}}{4\sqrt{2}}, \frac{1-\sqrt{3}}{4\sqrt{2}}\}$.

Because $\phi(0) = \sqrt{2} h_0 \phi(0)$ and $\sqrt{2} h_0 \neq 1$, it becomes clear that $\phi(0) = 0$. Using $\phi(3) = 0$, we obtain the system for $\phi(1)$ and $\phi(2)$

$$\begin{bmatrix} \phi(1) \\ \phi(2) \end{bmatrix} = \sqrt{2} \cdot \begin{bmatrix} h_1 & h_0 \\ h_3 & h_2 \end{bmatrix} \cdot \begin{bmatrix} \phi(1) \\ \phi(2) \end{bmatrix}.$$

Using $\sum_k \phi(x-k) = 1$, we have that $\phi(1) + \phi(2) = 1$. We solve for $\phi(1)$ and $\phi(2)$ to obtain

$$\phi(1) = \frac{1+\sqrt{3}}{2} \text{ and } \phi(2) = \frac{1-\sqrt{3}}{2}.$$

Now, one can refine ϕ,

$$\begin{aligned} \phi\left(\frac{1}{2}\right) &= \sum_k h_k \sqrt{2} \phi(1-k) = h_0 \sqrt{2} \phi(1) = \frac{2+\sqrt{3}}{4}, \\ \phi\left(\frac{3}{2}\right) &= \sum_k h_k \sqrt{2} \phi(3-k) = h_1 \sqrt{2} \phi(2) + h_2 \sqrt{2} \phi(1) \\ &= \frac{3+\sqrt{3}}{4} \cdot \frac{1-\sqrt{3}}{2} + \frac{3-\sqrt{3}}{4} \cdot \frac{1+\sqrt{3}}{2} = 0, \\ \phi\left(\frac{5}{2}\right) &= \sum_k h_k \sqrt{2} \phi(5-k) = h_3 \sqrt{2} \phi(2) = \frac{2-\sqrt{3}}{4}, \end{aligned}$$

or ψ,

$$\begin{aligned}
\psi(-1) &= \psi(2) = 0, \\
\psi\left(-\frac{1}{2}\right) &= \sum_k g_k \sqrt{2}\phi(-1-k) = h_1\sqrt{2}\phi(1) = -\frac{1}{4}, \qquad [g_n = (-1)^n h_{1-n}] \\
\psi(0) &= \sum_k g_k \sqrt{2}\phi(0-k) = g_{-2}\sqrt{2}\phi(2) + g_{-1}\sqrt{2}\phi(1) \\
&= -h_2\sqrt{2}\phi(1) = -\frac{\sqrt{3}}{4}, \\
&\text{etc.}
\end{aligned}$$

1.6 Wavelet Filters

In wavelet transforms, users can choose various wavelet filters that are appropriate for the purpose of the transform. We presents several commonly used wavelet filters.

1.6.1 Haar Wavelet

The Haar wavelet is a simple and widely applicable wavelet. In addition, it has immense educational value in the introduction of wavelets. In this section, we use the Haar wavelet to exhibit some of the properties discussed in Section 1.4. Assuming that everything else is unknown, we begin with $\phi(x) = \mathbf{1}(0 \le x \le 1)$. The scaling equation (15) for the Haar wavelet is very simple. By observing simple graphs of two scaled Haar wavelet $\phi(2x)$ and $\phi(2x+1)$ placed next to each other, we conclude that the scaling equation is

$$\begin{aligned}
\phi(x) &= \phi(2x) + \phi(2x-1) \\
&= \frac{1}{\sqrt{2}}\sqrt{2}\phi(2x) + \frac{1}{\sqrt{2}}\sqrt{2}\phi(2x-1). \qquad (41)
\end{aligned}$$

Therefore, the wavelet filter coefficients are

$$h_0 = h_1 = \frac{1}{\sqrt{2}}.$$

The transfer functions become

$$m_0(\omega) = \frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}e^{-i\omega 0}\right) + \frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}e^{-i\omega 1}\right) = \frac{1+e^{-i\omega}}{2},$$

and

$$m_1(\omega) = -e^{-i\omega}\,\overline{m_0(\omega+\pi)} = -e^{-i\omega}\left(\frac{1}{2} - \frac{1}{2}e^{i\omega}\right) = \frac{1-e^{-i\omega}}{2}.$$

Note $m_0(\omega) = |m_0(\omega)|e^{i\varphi(\omega)} = \cos\frac{\omega}{2}\cdot e^{-i\omega/2}$ (after $\cos x = \frac{e^{ix}+e^{-ix}}{2}$). Because $\varphi(\omega) = -\frac{\omega}{2}$, Haar's wavelet has the scaling function, which is symmetric in the time domain. One can easily verify the orthogonality condition $|m_0(\omega)|^2 + |m_1(\omega)|^2 = 1$. Relation (31) becomes

$$\Psi(\omega) = \frac{1-e^{-i\omega/2}}{2}\Phi\left(\frac{\omega}{2}\right) = \frac{1}{2}\Phi\left(\frac{\omega}{2}\right) - \frac{1}{2}\Phi\left(\frac{\omega}{2}\right)e^{-i\omega/2},$$

and in the time-domain, we apply the inverse Fourier transformation to obtain

$$\psi(x) = \phi(2x) - \phi(2x-1).$$

Following such steps, we finally obtain the Haar wavelet function, ψ. By inspecting the expression of m_1 or of $\psi(x)$, which consist of $\phi(2x)$ and $\phi(2x-1)$, respectively, we conclude that $g_0 = -g_{-1} = \frac{1}{\sqrt{2}}$. However, the Haar basis is not appropriate for all applications for several reasons. The fundamental components in Haar's decomposition are discontinuous functions that are not sufficient for approximating smooth functions. While Haar wavelet are well localized in the time domain, they decay at a slow rate of $O(\frac{1}{n})$ in the time/frequency domain.

1.6.2 Other Wavelets

In this section, we overview Daubechies, Symlet, and Coiflet wavelet filters. Here we briefly discuss features and provide filter tabs for several wavelets from each family.

1.6.2.1 Daubechies

Ingrid Daubechies was the first in inventing a family of wavelets that are compactly supported orthogonal wavelets with extremal phase. Daubechies are indexed with the number of vanishing moments N, or the number of filter tabs that amounts to $2N$. Table 1 gives filter coefficients for the DAUB2, ..., and DAUB10 (indexed with vanishing moments) filters.

1.6.2.2 Symlet

While the compactly supported wavelets cannot be symmetric, one can form a wavelet with compact support that are close to symmetry. Proposed by I. Daubechies, the symlet wavelets are modified Daubechies' family wavelets for least asymmetry with compact support. Thus, Symlet and Daubechies wavelets share similar properties. The naming convention and the number of filter tabs of Symlet follow the same rule as Daubechies. Table 2 gives filter coefficients for the SYM4, SYM5, and SYM6 (indexed with vanishing moments) filters.

1.6.2.3 Coiflet

Coiflets were requested by R. Coifman and developed by I. Daubechies. It is less asymmetric than the wavelets from the Daubechies or Symlet families, which was possible by the price of a larger support. Coiflets have high vanishing moments of both the scaling and the wavelet functions. The vanishing moments of scaling functions minimizes aliasing error occurred in discretization of continuous function. Coiflets are indexed with the number of vanishing moments N for both the wavelet and scaling functions while the number of filter tabs amounts to $3N$. Table 3 gives filter coefficients for the COIF2, COIF4, and COIF6 (indexed with vanishing moments) filters.

Table 1: The filter coefficients of Daubechies' wavelets for $N = 2, \ldots, 10$ vanishing moments.

k	DAUB2	DAUB3	DAUB4
0	0.4829629131445342	0.3326705529500827	0.2303778133088966
1	0.8365163037378080	0.8068915093110930	0.7148465705529161
2	0.2241438680420134	0.4598775021184915	0.6308807679298592
3	-0.1294095225512604	-0.1350110200102548	-0.0279837694168604
4		-0.0854412738820267	-0.1870348117190935
5		0.0352262918857096	0.0308413818355607
6			0.0328830116668852
7			-0.0105974017850690
k	DAUB5	DAUB6	DAUB7
0	0.1601023979741926	0.1115407433501095	0.0778520540850092
1	0.6038292697971887	0.4946238903984531	0.3965393194819173
2	0.7243085284377723	0.7511339080210954	0.7291320908462351
3	0.1384281459013216	0.3152503517091976	0.4697822874051931
4	-0.2422948870663808	-0.2262646939654398	-0.1439060039285650
5	-0.0322448695846383	-0.1297668675672619	-0.2240361849938750
6	0.0775714938400454	0.0975016055873230	0.0713092192668303
7	-0.0062414902127983	0.0275228655303057	0.0806126091510831
8	-0.0125807519990819	-0.0315820393174860	-0.0380299369350144
9	0.0033357252854738	0.0005538422011615	-0.0165745416306669
10		0.0047772575109455	0.0125509985560998
11		-0.0010773010853085	0.0004295779729214
12			-0.0018016407040475
13			0.0003537137999745
k	DAUB8	DAUB9	DAUB10
0	0.0544158422431070	0.0380779473638881	0.0266700579005487
1	0.3128715909143165	0.2438346746126514	0.1881768000776480
2	0.6756307362973218	0.6048231236902548	0.5272011889316280
3	0.5853546836542239	0.6572880780514298	0.6884590394535462
4	-0.0158291052563724	0.1331973858249681	0.2811723436606982
5	-0.2840155429615815	-0.2932737832793372	-0.2498464243271048
6	0.0004724845739030	-0.0968407832230689	-0.1959462743773243
7	0.1287474266204823	0.1485407493381040	0.1273693403356940
8	-0.0173693010018109	0.0307256814793158	0.0930573646035142
9	-0.0440882539307979	-0.0676328290613591	-0.0713941471663802
10	0.0139810279173996	0.0002509471148278	-0.0294575368218849
11	0.0087460940474065	0.0223616621236844	0.0332126740593155
12	-0.0048703529934519	-0.0047232047577528	0.0036065535669515
13	-0.0003917403733769	-0.0042815036824646	-0.0107331754833277
14	0.0006754494064506	0.0018476468830567	0.0013953517470513
15	-0.0001174767841248	0.0002303857635232	0.0019924052951842
16		-0.0002519631889428	-0.0006858566949593
17		0.0000393473203163	-0.0001164668551292
18			0.0000935886703200
19			-0.0000132642028945

Table 2: Symlet coefficients for $N = 4,\ 5,$ and 6 vanishing moments.

k	SYM4	SYM5	SYM6
0	0.032223100604052	0.019538882735250	0.015404109327045
1	-0.012603967262031	-0.021101834024689	0.003490712084222
2	-0.099219543576634	-0.175328089908056	-0.117990111148520
3	0.297857795605306	0.016602105764511	-0.048311742585698
4	0.803738751805133	0.633978963456791	0.491055941927974
5	0.497618667632775	0.723407690404040	0.787641141028651
6	-0.029635527646003	0.199397533976856	0.337929421728166
7	-0.075765714789502	-0.039134249302313	-0.072637522786377
8		0.029519490925706	-0.021060292512371
9		0.027333068344999	0.044724901770781
10			0.001767711864254
11			-0.007800708325032

Table 3: Coiflet coefficients for $N = 2,\ 4,$ and 6 vanishing moments.

k	COIF2	COIF4	COIF6
-6	0	0	-0.003793512864381
-5	0	0	0.007782596425673
-4	0	0.016387336463204	0.023452696142077
-3	0	-0.041464936786872	-0.065771911281469
-2	$(\sqrt{15}-3)\cdot\sqrt{2}/32$	-0.067372554723724	-0.061123390002973
-1	$(1-\sqrt{15})\cdot\sqrt{2}/32$	0.386110066822763	0.405176902409118
0	$(3-\sqrt{15})\cdot\sqrt{2}/16$	0.812723635449413	0.793777222626087
1	$(\sqrt{15}+3)\cdot\sqrt{2}/16$	0.417005184423239	0.428483476377370
2	$(\sqrt{15}+3)\cdot\sqrt{2}/32$	-0.076488599078281	-0.071799821619155
3	$(9-\sqrt{15})\cdot\sqrt{2}/32$	-0.050594344186464	-0.082301927106300
4	0	0.023680171946848	0.034555027573298
5	0	0.005611434819369	0.015880544863669
6	0	-0.001823208870911	-0.009007976136731
7	0	-0.000720549445520	-0.002574517688137
8	0	0	0.001117518770831
9	0	0	0.000466216959821
10	0	0	-0.000070983302506
11	0	0	-0.000034599773197

1.7 Discrete Wavelet Transformations (DWT)

One applies discrete wavelet transformations (DWT) to the discrete datasets and yields discrete outputs. The procedures of performing DWT is similar to performing fast Fourier transformation (FFT), which is the Fourier method for a set of discrete data.

Table 4: 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 time domain discrete input data is mapped to the wavelet domain as a vector of the same size using discrete wavelet transformations. Such process is linear and thus can be performed by a matrix multiplication as well. For example, we multiply a matrix of size $n \times n$ to one dimensional input data of size n for the discrete wavelet transformation. Such transformation matrices can be either orthogonal or "close" to orthogonal depending on boundary conditions. When the transformation matrix is orthogonal, we can regard such process as a rotation in $\mathbb{R}^n$ space in which an input data represent coordinates of a single point. New coordinates in the rotated space is the discrete wavelet transformation of the original data.

Example 1.7.1. 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 is the same as 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 $W \cdot (-1, 2)' = (\frac{1}{\sqrt{2}}, -\frac{3}{\sqrt{2}})'$. Note that *the energy* (squared distance of the point from the origin) is preserved, $(-1)^2 + (2)^2 = (\frac{1}{\sqrt{2}})^2 + (\frac{3}{\sqrt{2}})^2$, since W is a rotation.

Example 1.7.2. Let $\mathbf{y} = (1, 0, -5/2, 3/2, 1/2, -1/2, 3/2, 5/2)$. We interpolate the values $f(n) = y_n,\ n = 0, 1, \ldots, 7$ by the father wavelet from the Haar wavelet, the vector represent the sampled piecewise constant function. It is obvious that such defined f belongs to Haar's multi-resolution space V_0. The following matrix equation gives the connection between $\mathbf{y}$ and the wavelet coefficients (data in the wavelet domain).

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

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,

$$\begin{aligned} f &= \sqrt{2}\phi_{-3,0} - \sqrt{2}\psi_{-3,0} + \psi_{-2,0} - 2\psi_{-2,1} \\ &\quad + \frac{1}{\sqrt{2}}\psi_{-1,0} - 2\sqrt{2}\psi_{-1,1} + \frac{1}{\sqrt{2}}\psi_{-1,2} - \frac{1}{\sqrt{2}}\psi_{-1,3}. \end{aligned} \tag{42}$$

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

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

Such multiplication process for wavelet transformation is easy to understand at conceptual level, but has some limitation in practical utility. The size of a matrix that can transform an input of size ($n > 10^{15}$) is immense ($10^{15} \times 10^{15}$) which makes it difficult to be stored and computed for large signals. However, as we will see later, when dealing with images of moderate size, matrix transforms are an excellent approach.

1.7.1 Matrix-based DWT

We can perform the change of basis in V_1 from $\mathcal{B}_1 = \{\phi_{1k}(x), k \in Z\}$ to $\mathcal{B}_2 = \{\phi_{0k}, k \in Z\} \cup \{\psi_{0k}, k \in Z\}$ by matrix multiplication. Therefore, as mentioned before, we can define a discrete wavelet transformation matrix. Example 1.7.2 already showed a transformation matrix corresponding to Haar's.

Let the length of the input signal be 2^J, $\mathbf{h} = \{h_s,\ s \in \mathbb{Z}\}$ are the wavelet filter, and N is some constant. Denote H_k a matrix of size $(2^{J-k} \times 2^{J-k+1}),\ k = 1, \ldots$ with an i^{th} row and j^{th} column element

$$h_s, \quad s = (N-1) + (i-1) - 2(j-1) \text{ modulo } 2^{J-k+1}. \tag{43}$$

Note that H_k is the circulant matrix, which means that its i^{th} row is the circularly shifted version of the first row by $2(i-1)$ units, which is derived from the *modulo* operator in (43).

By analogy, we define a matrix G_k based on the filter $\mathbf{g}$. To obtain a G_k that corresponds to the already defined H_k, we change h_i by $(-1)^i h_{N+1-i}$. The constant N is a shift parameter, which 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,

$$I_{2^{J-k}} = [H_k' \; G_k'] \begin{bmatrix} H_k \\ G_k \end{bmatrix} = H_k' \cdot H_k + G_k' \cdot G_k.$$

and

$$I = \begin{bmatrix} H_k \\ G_k \end{bmatrix} \cdot [H_k' \; G_k'] = \begin{bmatrix} H_k \cdot H_k' & H_k \cdot G_k' \\ G_k \cdot H_k' & G_k \cdot G_k' \end{bmatrix}.$$

That implies,

$$H_k \cdot H_k' = I, \; G_k \cdot G_k' = I, \; G_k \cdot H_k' = H_k \cdot G_k' = 0, \text{ and } H_k' \cdot H_k + G_k' \cdot G_k = I.$$

Now, for a sequence y the J-step wavelet transformation is $\mathbf{d} = W_J \cdot \mathbf{y}$, where

$$W_1 = \begin{bmatrix} H_1 \\ G_1 \end{bmatrix}, \quad W_2 = \begin{bmatrix} \begin{bmatrix} H_2 \\ G_2 \end{bmatrix} \cdot H_1 \\ G_1 \end{bmatrix},$$

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

Example 1.7.3. Suppose that $\mathbf{y} = \{1, 0, -5/2, 3/2, 1/2, -1/2, 3/2, 5/2\}$ 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 $2^{3-k} \times 2^{3-k+1}$.

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

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

Since,

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

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

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

Because of short lengths of the filter and data, we are only able to perform discrete wavelet transformation for two steps, W_1 and W_2.

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

$W_2 \cdot \mathbf{y} = \{1.3125, 0.6875| - 0.0658, 2.6818|0.64700.09470.0820, -2.9451\}$, because

$$\begin{aligned} H_2 \cdot H_1 \cdot \mathbf{y} &= H_2 \cdot \{1.2089, -0.3239, -0.0657, 2.6818\} \\ &= \{1.3125, 0.6875\} \\ G_2 \cdot H_1 \cdot \mathbf{y} &= G_1 \cdot \{1.2089, -0.3239, -0.0657, 2.6818\} \\ &= \{-0.0658, 2.6818\}. \end{aligned}$$

We can easily extend such 1-D matrix-based DWT to a 2-D matrix-based DWT. Let us consider an example of transforming a 2-D input $\boldsymbol{X}$ of size $(m \times m)$ with DWT of depth 3. We form a matrix W_3 for an input of size m and use it for both column and row decomposition.

$$W_3 \cdot \boldsymbol{X} \cdot W_3'$$

Note that the first W_3 is for a row-wise decomposition of $\boldsymbol{X}$, while the second W_3 is for a column-wise decomposition.

1.8 Traditional orthogonal Wavelet Transform

In this section, we provide functional representations of standard orthogonal wavelet transform (DWT). We assume that a multiresolution framework is specified, and we denote ϕ and ψ scaling and wavelet functions, respectively. A data vector $\boldsymbol{y} = (y_0, y_1, \ldots, y_{m-1})$ of size m can be expressed as function f in terms of shifts of the scaling function at some multiresolution level J such that $J - 1 < \log_2 m \leq J$, as

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

where $\phi_{J,k}(x) = 2^{J/2}\phi\left(2^J x - k\right)$. We can express the data interpolating function f as

$$f(x) = \sum_{k=0}^{m-1} c_{J_0,k}\phi_{J_0,k}(x) + \sum_{j=J_0}^{J-1}\sum_{k=0}^{2^n-1} d_{jk}2^{j/2}\psi\left(2^j x - k\right), \tag{44}$$

where

$$\begin{aligned}
\phi_{J_0,k}(x) &= 2^{J_0/2}\phi\left(2^{J_0}x - k\right), \\
\psi_{jk}(x) &= 2^{j/2}\psi\left(2^j x - k\right), \\
&\quad j = J_0, \ldots, J-1;\ k = 0, 1, \ldots, m-1.
\end{aligned}$$

The DWT of vector $\boldsymbol{y}$ consist of wavelet coefficients, $c_{J_0,k}, k = 0, \ldots 2^{J_0} - 1$ and $d_{jk}, j = J_0, \ldots, J-1; k = 0, \ldots, 2^j - 1$,.

Note that shift k is level dependent, $2^{-j}k$. Thus, as the level decreases, number of wavelet coefficients are reduced by half. The resulting wavelet coefficients are $\boldsymbol{c}^{(J_0)} = \left(c_{J_0,0}, \ldots, c_{J_0,2^{J_0}-1}\right)$ and $\boldsymbol{d}^{(j)} = \left(d_{j,0}, \ldots, d_{j,2^j-1}\right)$, $j = J_0, \ldots, J-1$, for $p = J - J_0$. Since the number of wavelet coefficients decreases by half in each decomposition level $p < J$.

Coefficients in $\boldsymbol{d}^{(j)}$ captures detail information within an input and referred to as detail coefficients, while coefficients in $\boldsymbol{c}^{(J_0)}$ captures coarse fluctuation within an input and referred to as coarse coefficients. For any given decomposition depth p,

because of decimation, the total number of wavelet coefficients is the same as the original data vector, m.

Such 1-D definitions can be easily extended to 2-D transforms. We describe standard and a scale-mixing 2-D DWTs of $f(x, y)$, where $(x, y) \in \mathbb{R}^2$. For the standard 2-D DWT, the wavelet atoms are

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

where $\boldsymbol{k} = (k_1, k_2)$ is the location pair, and $j = J_0, \ldots, J-1$ is the scale. The depth of the transform is $p = J - 1 - J_0$. The wavelet coefficients of $f(x, y)$ are calculated as

$$\begin{aligned}
c_{J_0;\boldsymbol{k}} &= 2^{J_0} \iint f(x,y)\phi_{J_0;\boldsymbol{k}}(x,y)\, dxdy, \\
d^{(i)}_{j;\boldsymbol{k}} &= 2^{j} \iint f(x,y)\psi^{(i)}_{j;\boldsymbol{k}}(x,y)\, dxdy,
\end{aligned}$$

where J_0 is the coarsest decomposition level, and $i \in \{h, v, d\}$ represents the "orientation" of detail coefficients as horizontal, vertical, and diagonal (e.g., Vidakovic (1999), p. 155). Figure 7(a) presents the tessellation of a standard 2-D DWT.

For the scale-mixing 2-D DWT, the wavelet atoms are

$$\begin{aligned}
\phi_{J_{01},J_{02};\boldsymbol{k}}(x,y) &= 2^{(J_{01}+J_{02})/2}\phi(2^{J_{01}}x - k_1)\phi(2^{J_{02}}y - k_2), \\
\psi_{J_{01},j_2;\boldsymbol{k}}(x,y) &= 2^{(J_{01}+j_2)/2}\phi(2^{J_{01}}x - k_1)\psi(2^{j_2}y - k_2), \\
\psi_{j_1,J_{02};\boldsymbol{k}}(x,y) &= 2^{(j_1+J_{02})/2}\psi(2^{j_1}x - k_1)\phi(2^{J_{02}}y - k_2), \\
\psi_{j_1,j_2;\boldsymbol{k}}(x,y) &= 2^{(j_1+j_2)/2}\psi(2^{j_1}x - k_1)\psi(2^{j_2}y - k_2),
\end{aligned}$$

where J_{01} and J_{02} are coarsest levels, and $j_1 \geq J_{01}; j_2 \geq J_{02},$. The resulting wavelet

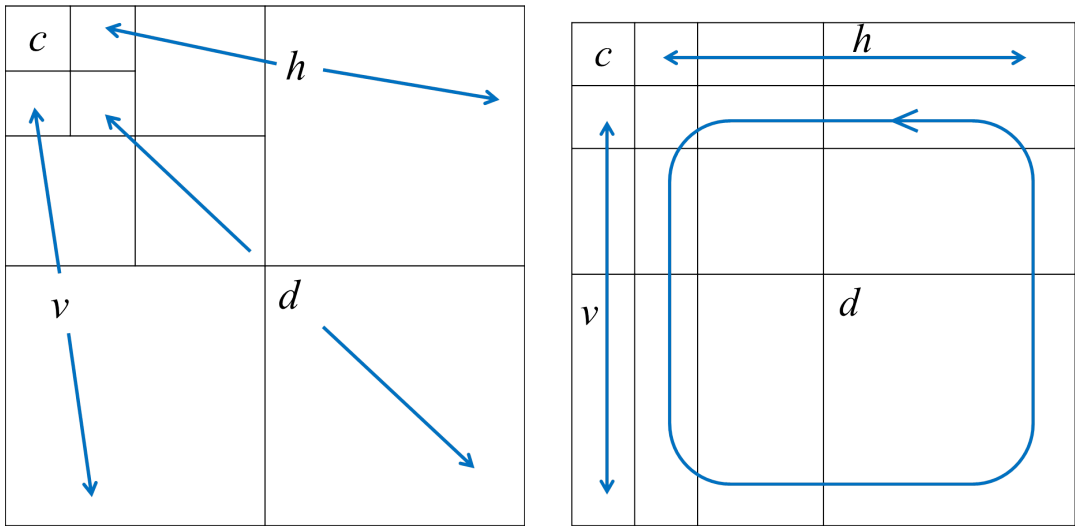


Figure 7: Locations of four types of wavelet coefficients in the tessellation of 3-level decomposition with the standard (a) and scale-mixing (b) 2-D DWT's. Different types of coefficients are defined in (45).

coefficients for $f(x, y)$ from the scale-mixing DWT are

$$\begin{aligned} c_{J_{01},J_{02};\boldsymbol{k}} &= \iint f(x,y)\phi_{J_{01},J_{02};\boldsymbol{k}}(x,y)\,dxdy, \\ h_{J_{01},j_2;\boldsymbol{k}} &= \iint f(x,y)\psi_{J_{01},j_2;\boldsymbol{k}}(x,y)\,dxdy, \\ v_{j_1,J_{02};\boldsymbol{k}} &= \iint f(x,y)\psi_{j_1,J_{02};\boldsymbol{k}}(x,y)\,dxdy, \\ d_{j_1,j_2;\boldsymbol{k}} &= \iint f(x,y)\psi_{j_1,j_2;\boldsymbol{k}}(x,y)\,dxdy. \end{aligned} \tag{45}$$

Notice that in the standard DWT, we use common j to denote a scale, while in the scale-mixing DWT, we use a pair (j_1, j_2), which indicates that two scales are mixed. Figure 7(b) illustrates the tessellation of coefficients of scale-mixing 2-D DWT.

1.9 Non-decimated Wavelet Transform (NDWT)

Each version of a wavelet transform has characteristics that are useful in certain applications. A popular version is a non-decimated wavelet transform (NDWT), which can overcome sensitivity to translations in time and shift found in standard

orthogonal wavelet transform.

In early 1990s, NDWT appeared in literature with various names for a number of applications and purposes as the following. Mallat (1991) proposed a method that approximates continuous wavelet transform with an iterative algorithm, which turned out to be equivalent to a shift-invariant representation. Shensa (1992); Beylkin (1992) describes an efficient algorithm with $O(n \log_2(n))$ complexity for calculating wavelet coefficients that are shift-invariant, that is, yielding redundant wavelet coefficients at all N circulant shift for an input signal size of N. Pesquet et al. (1996) extended a wavelet packet decomposition for time invariance and applied it to estimation and detection problems. Unser (1995) uses an overcomplete wavelet decomposition, which is referred to as "discrete wavelet frame", for classification of texture. Coifman and Donoho (1995); Lang et al. (1995) applied translation invariant transform to thresholding for noise reduction. Nason and Silverman (1995) describes "stationary wavelet transform" with example applications for local spectra estimation. Liang and Parks (1996) applied a translation-invariant wavelet algorithm for data compression.

Such proposed methods are described with slightly different terms but essentially are NDWT and utilize properties such as translation/shift invariance and more dense approximation of continuous wavelet transform that are results of redundancy of NDWT. The NDWT is a redundant transform because it is performed by repeated filtering with a minimal shift, or a maximal sampling rate, at all dyadic scales. Subsequently, the transformed signal contains the same number of coefficients as the original signal at each multiresolution level. We start by describing algorithmic procedure of 1-D NDWT. Traditionally, we perform a wavelet transformation with a convolution of an input data and wavelet and scaling functions. A principal difference between NDWT and DWT is a sampling rate. This will be illustrated in the following derivations. With an assumption that a multiresolution framework is specified and that ϕ and ψ are scaling and wavelet functions respectively. We 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 such that $J - 1 < \log_2 m \leq J$, as

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

where $\phi_{J,k}(x) = 2^{J/2}\phi\left(2^J(x-k)\right)$. The data interpolating function f can be re-expressed as

$$f(x) = \sum_{k=0}^{m-1} c_{J_0,k}\phi_{J_0,k}(x) + \sum_{j=J_0}^{J-1}\sum_{k=0}^{m-1} d_{jk}2^{j/2}\psi\left(2^j(x-k)\right), \tag{46}$$

where

$$\phi_{J_0,k}(x) = 2^{J_0/2}\phi\left(2^{J_0}(x-k)\right),$$
$$\psi_{jk}(x) = 2^{j/2}\psi\left(2^j(x-k)\right),$$
$$j = J_0, \ldots, J-1;\ k = 0, 1, \ldots, m-1.$$

The coefficients, $c_{J_0,k}, k = 0, \ldots m-1$ and $d_{jk}, j = J_0, \ldots, J-1; k = 0, \ldots, m-1$, represent the NDWT of vector $\boldsymbol{y}$.

Notice that a shift, k, is constant at all levels, unlike the traditional orthogonal wavelet transform in which the shifts are level dependent, $2^{-j}k$. This constancy of the shifts across the levels in NDWT indicates that the transform is time invariant. As we see from equation (52), the NDWT produces a redundant representation of the data. For an original signal of size m transformed into p decomposition levels (the depth of transform is p), the resulting non-decimated wavelet coefficients are $\boldsymbol{c}^{(J_0)} = (c_{J_0,0}, \ldots, c_{J_0,m-1})$ and $\boldsymbol{d}^{(j)} = (d_{j,0}, \ldots, d_{j,m-1}),\ j = J_0, \ldots, J-1$, for $p = J - J_0$. Since NDWT does not decimate, nothing stops the user from taking p larger than $\lceil \log_2 m \rceil$. For such p coarse levels of details become zero-vectors.

Coefficients $\boldsymbol{d}^{(j)}$ serve as the detail coefficients while coefficients $\boldsymbol{c}^{(J_0)}$ serves as the coarsest approximation of the data. Later, we will refer to these coefficients as d-type and c-type coefficients. With p detail levels, the total number of wavelet coefficients

is $(p+1) \times m$. Such wavelet coefficients at different decomposition levels are related to one another by Mallat's pyramid algorithm (Mallat (1989b), Mallat (1989a)) in which convolutions of low- and high-pass wavelet filters, (h) and (g), respectively, take place in a cascade. The filters h and g are known as quadrature mirror filters. As we discussed in Remark 1.4.1, given a low-pass wavelet filter $h = (h_0, \ldots, h_M)$, fully and uniquely specified the choice of wavelet basis, the i^{th} entry of the high-pass counterpart g is $g_i = (-1)^{l-i} \cdot h_{M-s-i}$, for arbitrary but fixed integers l and s. We will further discuss the filter operators in the context of NDWT later in this section.

Expanding on the 1-D definitions, we overview a 2-D NDWT of $f(x, y)$, where $(x, y) \in \mathbb{R}^2$. Several versions of 2-D NDWT exist but we focus on the standard and a scale-mixing versions. For the standard 2-D NDWT, the wavelet atoms are

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

where (k_1, k_2) is the location pair, and $j = J_0, \ldots, J-1$ is the scale. The depth of the transform is $p = J - 1 - J_0$. The wavelet coefficients of $f(x, y)$ are calculated as

$$\begin{aligned}
c_{J_0;k_1,k_2} &= 2^{J_0}\iint f(x,y)\phi_{J_0;k_1,k_2}(x,y)\,dxdy,\\
d^{(i)}_{j;k_1,k_2} &= 2^{j}\iint f(x,y)\psi^{(i)}_{j;k_1,k_2}(x,y)\,dxdy,
\end{aligned}$$

where J_0 is the coarsest decomposition level, and $i \in \{h, v, d\}$ indicates the "orientation" of detail coefficients as horizontal, vertical, and diagonal (e.g., Vidakovic (1999), p. 155). The tessellation to a standard 2-D NDWT is presented in Figure 14(a).

For the scale-mixing 2-D NDWT, the wavelet atoms are

$$\begin{aligned}
\phi_{J_{01},J_{02};\boldsymbol{k}}(x,y) &= 2^{(J_{01}+J_{02})/2}\phi(2^{J_{01}}(x-k_1))\phi(2^{J_{02}}(y-k_2)),\\
\psi_{J_{01},j_2;\boldsymbol{k}}(x,y) &= 2^{(J_{01}+j_2)/2}\phi(2^{J_{01}}(x-k_1))\psi(2^{j_2}(y-k_2)),\\
\psi_{j_1,J_{02};\boldsymbol{k}}(x,y) &= 2^{(j_1+J_{02})/2}\psi(2^{j_1}(x-k_1))\phi(2^{J_{02}}(y-k_2)),\\
\psi_{j_1,j_2;\boldsymbol{k}}(x,y) &= 2^{(j_1+j_2)/2}\psi(2^{j_1}(x-k_1))\psi(2^{j_2}(y-k_2)),
\end{aligned}$$

where J_{01} and J_{02} are coarsest levels, $j_1 \geq J_{01}; j_2 \geq J_{02}$, and $\boldsymbol{k} = (k_1, k_2)$. As a result, we obtain wavelet coefficients for $f(x,y)$ from the scale-mixing NDWT as

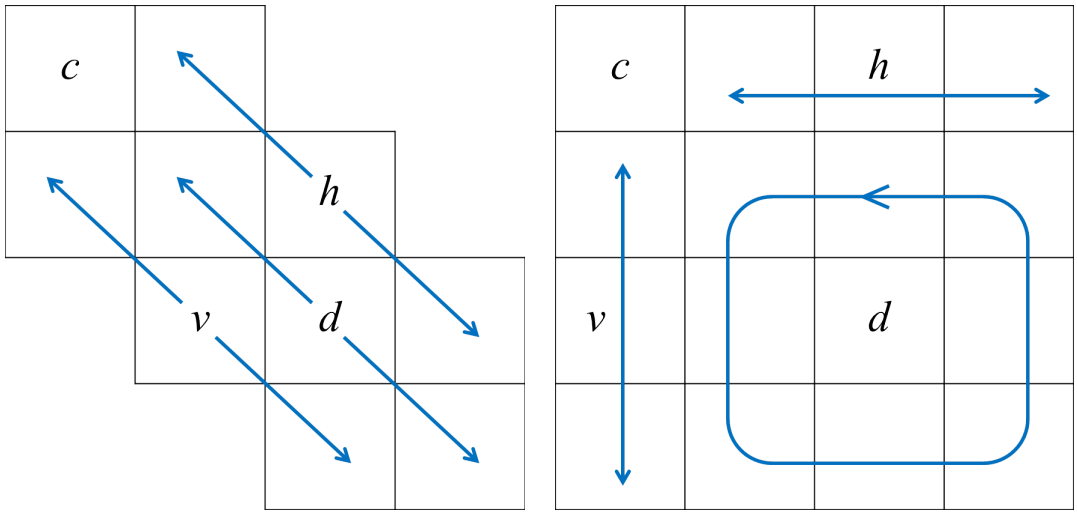
$$\begin{aligned}
c_{J_{01},J_{02};\boldsymbol{k}} &= \iint f(x,y)\phi_{J_{01},J_{02};\boldsymbol{k}}(x,y)\,dxdy,\\
h_{J_{01},j_2;\boldsymbol{k}} &= \iint f(x,y)\psi_{J_{01},j_2;\boldsymbol{k}}(x,y)\,dxdy,\\
v_{j_1,J_{02};\boldsymbol{k}} &= \iint f(x,y)\psi_{j_1,J_{02};\boldsymbol{k}}(x,y)\,dxdy,\\
d_{j_1,j_2;\boldsymbol{k}} &= \iint f(x,y)\psi_{j_1,j_2;\boldsymbol{k}}(x,y)\,dxdy.
\end{aligned} \tag{47}$$

Notice that in the standard NDWT, we use common j to denote a scale, while in the scale-mixing NDWT, we use a pair (j_1, j_2), which indicates that two scales are mixed. Figure 14(b) illustrates the tessellation of coefficients of scale-mixing 2-D NDWT. In Section 2.2.1 we will refer to coefficients from (53) as c-, h-, v-, and d-type coefficients.

1.9.1 Discrete Version of NDWT

While the functional series involving wavelet and scaling functions as decomposing atoms is an established mathematical framework for describing the NDWT, we provide an alternative description of NDWT using convolution operators (Nason and Silverman (1995), Vidakovic (1999), Strang and Nguyen (1996)). Such a description is preferred for discrete inputs.

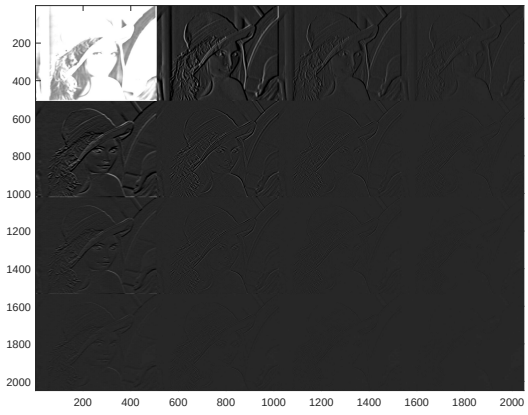
Let $[\uparrow 2]$ denote the upsampling of a given sequence by inserting a zero between every two neighboring elements of a sequence. We define the dilations of wavelet



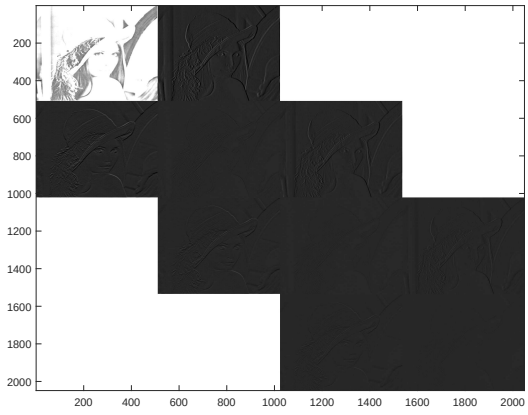
(a) (b)

Figure 8: Four types of wavelet coefficients with their locations in the tessellation of a 2-D standard (a) and scale-mixing (b) NDWT of depth 3. Coefficients c represent the coarsest approximation, h and v represent the mix of coarse and detail information, and d carry information about details only.

(a) Original lena image



(b) Transformed by 2-D scale-mixing NDWT



(c) Transformed by standard 2-D NDWT

Figure 9: Lena image in different domains

filters h and g as

$$\begin{aligned} h^{[0]} &= h, & g^{[0]} &= g \\ h^{[r]} &= [\uparrow 2]h^{[r-1]}, & g^{[r]} &= [\uparrow 2]g^{[r-1]}. \end{aligned} \tag{48}$$

Inserting zeros between each element of filters $h^{[r-1]}$ and $g^{[r-1]}$ creates holes (*trous*, in French), which is why this approach is sometimes called Algorithm á Trous.

A non-decimated wavelet transform is completed by applying convolution operators, $\mathbf{H}^{[j]}$ and $\mathbf{G}^{[j]}$, which come from dilated filters $h^{[r]}$ and $g^{[r]}$ in sequence. Detail

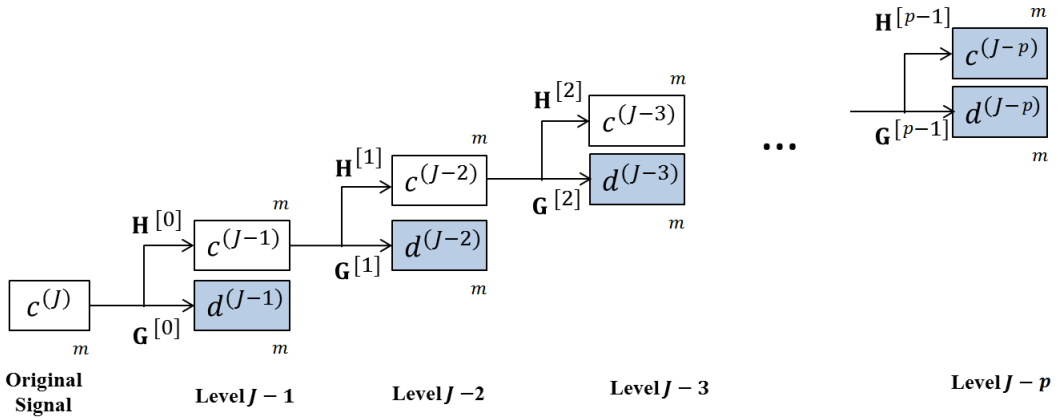


Figure 10: Graphical explanation of the NDWT process. Given signal a^J of size m, we decompose the signal to $p+1$ multi-resolution subspaces that include p levels of detail coefficients and one level of scaling coefficients, resulting in a set of coefficient vectors, $\boldsymbol{d}^{(J-1)}, \boldsymbol{d}^{(J-2)}, \ldots, \boldsymbol{d}^{(J-p)}$, and $\boldsymbol{c}^{(J-p)}$ in shaded blocks.

and coarse coefficients generated from each level have an identical size, m, which is the same as that of the original signal. To obtain coefficients at decomposition level $J-j$, where $j \in \{1, 2, \ldots, p\}$, we repeatedly apply convolution operators to a coarse coefficient vector from the previous decomposition level, $J-j+1$

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

where $\mathbf{H}^{[j-1]}$ and $\mathbf{G}^{[j-1]}$ are filter operators that perform low- and high-pass filtering using quadrature mirror filters $h^{[j-1]}$ and $g^{[j-1]}$, respectively. The NDWT is the result

of repeated applications of two filter operators, $\mathbf{H}^{[j]}$ and $\mathbf{G}^{[j]}$. Operators ($\mathbf{H}^{[j]}$, $\mathbf{G}^{[j]}$) do not have an orthogonality property, so to obtain such a property, we utilize two additional operators $\mathcal{D}_0$ and $\mathcal{D}_1$, which perform decimation by selecting every even and odd member of an input signal. An example of the use of the decimation operator $\mathcal{D}$ with a signal x is

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

where i indicates the position of an element in the signal x. We apply ($\mathcal{D}_0\mathbf{H}^{[j-1]}$, $\mathcal{D}_0\mathbf{G}^{[j-1]}$) and ($\mathcal{D}_1\mathbf{H}^{[j-1]}$, $\mathcal{D}_1\mathbf{G}^{[j-1]}$) to a given signal and obtain the even and odd elements of NDWT wavelet coefficient vectors, $\boldsymbol{c}^{(J-j)}$ and $\boldsymbol{d}^{(J-j)}$, respectively. Thus, equation (54) is, in fact, performed as the following process

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

We apply the filtering twice at the even and odd positions for each decomposition level, so a shift does not affect transformation results, which means that the NDWT is time-invariant. Such time-invariance property of the NDWT yields a smaller mean-squared-error and reduces the Gibbs phenomenon in de-noising Coifman and Donoho (1995). However, the violation of variance preservation in the NDWT complicates the signal reconstruction.

1.9.2 Scaling, Anisotropy, and Wavelet Spectrum

We present several properties that self-similar processes exhibit in the wavelet domain. Some seminal papers in this field are written by Flandrin and his collaborators

(Abry et al., 1993; Flandrin, 1989b,a, 1992). A rich amount of recent literature continues with this topic. The self-similarity is an inherent property in a number of high-frequency real-life signals and images. Wavelets, which are local and adaptive functional bases, are suitable for assessing the degree of self-similarity in such data Vidakovic (1999). The literature on assessing the Hurst exponent is rich and the monograph Doukhan et al. (2003) provides a comprehensive overview.

To quantify characteristics within 2-D inputs with a wavelet spectrum, we consider two types of descriptors: scaling and asymmetry measures. Defined in a time/scale domain, a wavelet spectrum represents the distribution of energies within an original signal along the range of scales (i.e., resolution levels). In the wavelet jargon, the term "energy" stands for a squared wavelet coefficient. For each 2-D resolution level indexed by $\boldsymbol{j} = (j_1, j_2)$, $\overline{|d_{\boldsymbol{j}}|^2}$ represents the average level energy. The wavelet spectrum refers to a sequence of logarithms of average level energies along a hierarchy that can be selected in various ways. Figure 11 demonstrates three possible hierarchies in a tessellation of 2-D scale mixing NDWT. In this thesis, we focus on only the main diagonal hierarchy whose 2-D scale indices coincide, i.e., $j_1 = j_2$. We denote a set of levels that belong to the main diagonal hierarchy $\boldsymbol{j}_s = (j, j),$ where $\max(j_{02}, j_{01}) \leq j \leq J - 1$. Wavelet coefficients obtained from an NDWT possess spatial location invariance across the level spaces. Thus, once the area of interest in an original signal is selected, one can easily identify wavelet coefficients in each resolution level that correspond to the selected area. Therefore, when calculating a wavelet spectrum, one can use either all wavelet coefficients or only the portion of coefficients corresponding to an area of interest. Such local spectra are natural for NDWT, unlike the orthogonal transforms that decimate. Right panel in Figure 13 exemplifies this feature. For the estimation of the scaling, we measure the rate of average energy decrease along $\boldsymbol{j}_s$ in the main diagonal hierarchy. When this decrease of energy is regular, it can be connected to the degree of self-similarity in signals and defines a commonly used scaling index, Hurst

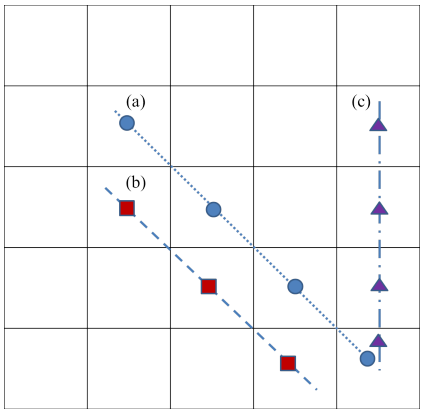


Figure 11: Three hierarchies of detail spaces in the tessellation of 2-D scale mixing NDWT of depth 4, where we index detail spaces with the pair of indices (j_1, j_2). (a) A main diagonal hierarchy whose scale indices satisfy $j_1 = j_2$, (b) a hierarchy whose scale index j_2 is fixed to the finest resolution level, and (c) a hierarchy whose scale indices satisfy $j_1 = j_2 + 1$.

exponent H. The relationship between energies in the wavelet spectra and Hurst exponent H is captured by a simple equation. To explain the equation in detail, we consider one example with a 2-D fBf $B_H(x, y) \in \mathbb{R}^2$ of size $(2^J \times 2^J)$. We perform a 2-D scale-mixing NDWT to $B_H(x, y)$ with the lowest resolution levels for column and row decompositions of $B_H(x, y)$ as j_{01} and j_{02}, respectively. Figure 12 demonstrates the resulting resolution space and wavelet coefficients yielded from this transform. The wavelet spectrum from the main diagonal hierarchy is defined as a set of pairs

$$\left(j,\ \log_2\left(\overline{|d_{(j,j);\boldsymbol{k}}|^2}\right)\right), \tag{49}$$

where $(j, j) \in \boldsymbol{j}_s$ and $\max(j_{02}, j_{01}) \leq j < J$. Wavelet coefficients $d_{(j,j);\boldsymbol{k}}$ are approximately independent and identically distributed Gaussian random variables with zero mean and a variance dependent on level j, Heneghan et al. (1996). The expected energy at each level in the main diagonal hierarchy is

$$\begin{aligned} E\left[|d_{(j,j);\boldsymbol{k}}|^2\right] &= 2^{2j} \iint \psi\left(2^j(\mathbf{v} - \boldsymbol{k})\right) \psi\left(2^j(\mathbf{u} - \boldsymbol{k})\right) E\left[B_H(\mathbf{v}) B_H(\mathbf{u})\right] d\mathbf{v} d\mathbf{u} \\ &= \frac{\sigma_H^2}{2} V_\psi 2^{-(2H+2)j}, \end{aligned} \tag{50}$$

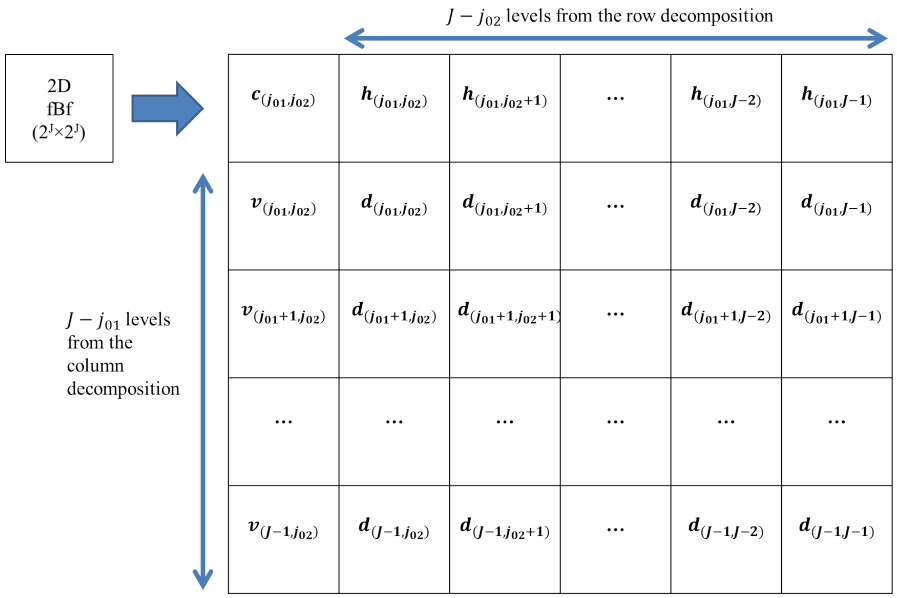


Figure 12: The location of wavelet coefficients that belong to various resolution levels when we perform a 2-D scale-mixing NDWT to an image of the size $(2^J \times 2^J)$ with the lowest resolution levels for columns and rows of j_{01} and j_{02}, respectively.

where $(j, j) \in \boldsymbol{j}_s$, $\mathbf{u}$,$\mathbf{v} \in \mathbb{R}^2$, $\boldsymbol{k} = (k_1, k_2)$, and

$$V_\psi = -\iint \psi(\boldsymbol{p}+\boldsymbol{q})\psi(\boldsymbol{q})|\boldsymbol{p}|^{2H} d\boldsymbol{p} d\boldsymbol{q}.$$

Expression V_ψ is independent of scale j but dependent on wavelet function ψ and Hurst exponent H. Details on derivation of (50) are deferred to Appendix A.1. Taking the logarithm of both sides in (50) yields

$$\log_2 E\left[|d_{(j,j);\boldsymbol{k}}|^2\right] = -(2H+2)j + C, \tag{51}$$

where $(j, j) \in \boldsymbol{j}_s$ and C is a constant that does not depend on j but possibly on wavelet function ψ and H. Figure 13 provides a graphic of a wavelet spectrum. In the left panel of Figure 13, a wavelet spectrum of log average energies across the scales is represented by a solid line, and its linear fit is represented by a dotted line. The right panel of Figure 13 shows a 2-D NDWT of depth 5 of a image that has a selected area of interest. Red marked areas represent wavelet coefficients that belong to the area of interest and are used for scaling estimation from the main diagonal

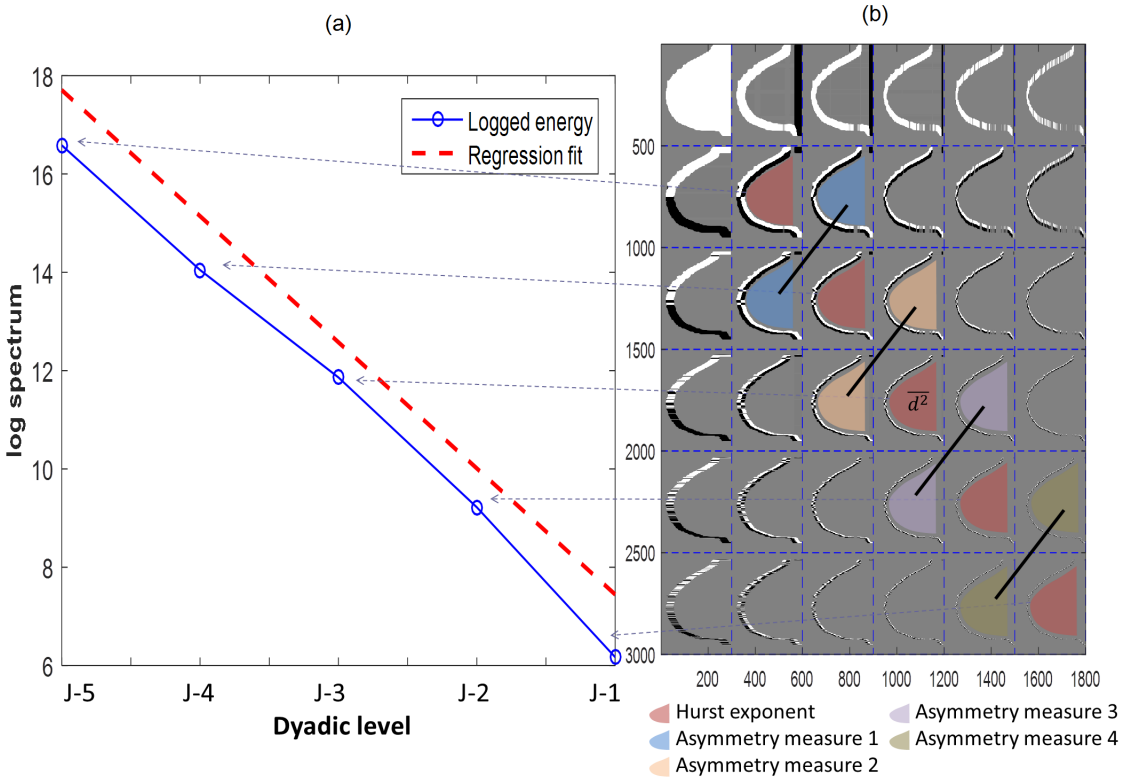


Figure 13: (a): Wavelet spectrum of log average energies across the scales represented by a solid line and its linear fit represented by a dotted line. (b): A 2-D NDWT of depth 5 of an image that has a selected area of interest(colored) with the diagonal hierarchy of levels and pairs of subspaces for anisotropy calculations.

hierarchy. The left top image in the matrix represents the coarsest features of an original image. Note that for calculation of the scaling descriptors, we use only wavelet coefficients located in the colored areas, i.e., corresponding to the area of interest since we are interested in local spatial characteristics. At each resolution level, we can readily identify wavelet coefficients that correspond to a selected area of interest in the original image. This spatial location invariance across the scales is distinctive feature of NDWT's, and enables the construction of local spectra.

Once we calculate the slope $\hat{\beta}$ in the linear fit of pairs in (49), according to (51), the Hurst exponent is estimated as $\hat{H} = -\hat{\beta}/2 - 1$.

For the estimation of the degree of horizontal/vertical anisotropy A_j, we calculate

the asymmetry ratio of two average energies that are adjacent to the main diagonal hierarchy and whose scale indices differ by 1. Thus, an asymmetry measure at level j is $A_j = \overline{|d_{(j+1,j)}|^2}/\overline{|d_{(j,j+1)}|^2}$, where $(j,j) \in \boldsymbol{j}_s$. If an input image exhibits isotropy in horizontal and vertical directions, the ratio is close to 1. Figure 13 visually describes the linked pairs of subspaces used for calculation of four asymmetry measures. Introduced scaling and anisotropy measures will be used in Chapter 3. In the next chapter, we will introduce a matrix-based NDWT that can facilitate NDWT in various ways.

Chapter II

MATRIX FOR NON-DECIMATED WAVELET TRANSFORM

A traditional way of performing NDWT involves convolution of scaling and wavelet functions with an input data, or a function. However, as all wavelet transforms are linear, NDWT can be described as a matrix multiplication. This chapter develops an algorithm that performs an NDWT as a matrix multiplication and provides the algorithm as a MATLAB package, **WavmatND**, which efficiently performs NDWT. The proposed package has three novel features. The first feature is that instead of using convolution-based Mallat's pyramid algorithm Mallat (1989a), we perform the NDWT by matrix multiplication. The matrix is formed directly from wavelet filter coefficients. Remenyi et al. (2014) also performed the NDWT using a matrix-based approach; however, their rules of constructing a matrix were based on Mallat's algorithm. Percival and Walden (2006) provide a matrix construction rule for NDWT, but the construction requires a convolution of filters in defining entries of the matrix, which is, essentially, Mallat's algorithm. The proposed method explicitly defines each entry of the transform matrix directly from the filter elements. With its simple construction rules, the proposed matrix-based NDWT requires significantly less time for computation compared to the convolution-based NDWT when the input signals are of a moderate size.

The second feature is that inverse transform matrix differs from the transpose of direct transform matrix up to a multiplicative rescaling matrix. Rescaling of submatrices of a NDWT matrix is needed to both obtain resulting wavelet coefficients

in their proper scales and retrieve the original signal without loss of information. Unlike the matrix for the orthogonal wavelet transform, which is a square matrix, a NDWT matrix for a p-depth decomposition of a signal of size m consists of $(p+1)$ square ($[m \times m]$) submatrices, each of which corresponds to one decomposition level. For a perfect reconstruction, the proposed process utilizes a weight matrix of size $[(p+1) \cdot m \times (p+1) \cdot m]$ that enables lossless reconstruction. The multiplication of the transposed NDWT matrix, the weight matrix, and the NDWT matrix, in that order, yields an identity matrix of size $[m \times m]$, which guarantees a lossless inverse transform. The matrix of Percival and Walden (2006) can retrieve an input signal but the resulting wavelet coefficients are down-scaled because of insisting on the energy preservation in redundant transform. With the proposed two-stage process, we can obtain the wavelet coefficients in their correct scale and then we can utilize a weight matrix if the inverse transform is necessary.

The third feature is that the package can handle one- or two-dimensional (1-D or 2-D) signals of an arbitrary size, and even the rectangular shapes in the case of a 2-D transform. This property is not shared by critically sampled wavelet transforms that require an input of dyadic size. In addition, one can perform a 2-D NDWT with two different wavelet bases, one base acting on the rows and another acting on the columns of the 2-D input signal, which allows for more modeling freedom in the case of spatially anisotropic 2-D signals.

2.1 Non-decimated Wavelet Transforms

Unique characteristics of the NDWT are well captured by its alternative names such as "stationary wavelet transform," "time-invariant wavelet transform," "*á trous* transform," or "maximal overlap wavelet transform." In this section, we will overview the features of the NDWT that motivate such names, beginning with a description of a

one-dimensional NDWT for a discrete input.

Assume that a multiresolution framework is specified and that ϕ and ψ are scaling and wavelet functions respectively. We 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 such that $J - 1 < \log_2 m \leq J$, as

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

where $\phi_{J,k}(x) = 2^{J/2}\phi\left(2^J(x-k)\right)$. The data interpolating function f can be re-expressed as

$$f(x) = \sum_{k=0}^{m-1} c_{J_0,k}\phi_{J_0,k}(x) + \sum_{j=J_0}^{J-1}\sum_{k=0}^{2^n-1} d_{jk}2^{j/2}\psi\left(2^j(x-k)\right), \tag{52}$$

where

$$\begin{aligned}
\phi_{J_0,k}(x) &= 2^{J_0/2}\phi\left(2^{J_0}(x-k)\right), \\
\psi_{jk}(x) &= 2^{j/2}\psi\left(2^j(x-k)\right), \\
&\qquad j = J_0, \ldots, J-1;\ k = 0, 1, \ldots, m-1.
\end{aligned}$$

The coefficients, $c_{J_0,k}, k = 0, \ldots m-1$ and $d_{jk}, j = J_0, \ldots, J-1; k = 0, \ldots, m-1$, comprise the NDWT of vector $\boldsymbol{y}$.

Notice that a shift, k, is constant at all levels, unlike the traditional orthogonal wavelet transform in which the shifts are level dependent, $2^{-j}k$. This constancy of the shifts across the levels in NDWT indicates that the transform is time invariant. As we see from equation (52), the NDWT produces a redundant representation of the data. For an original signal of size m transformed into p decomposition levels (the depth of transform is p), the resulting non-decimated wavelet coefficients are $\boldsymbol{c}^{(J_0)} = (c_{J_0,0}, \ldots, c_{J_0,m-1})$ and $\boldsymbol{d}^{(j)} = (d_{j,0}, \ldots, d_{j,m-1})$, $j = J_0, \ldots, J-1$, for $p = J - J_0$. Since NDWT does not decimate, nothing stops the user from taking p larger than $\lceil \log_2 m \rceil$. For such p coarse levels of detail become zero-vectors.

Coefficients in $\boldsymbol{d}^{(j)}$ serve as the detail coefficients while coefficients in $\boldsymbol{c}^{(J_0)}$ serve as the coarsest approximation of the data. Later, we will refer to these coefficients as d-type and c-type coefficients. With p detail levels, the total number of wavelet coefficients is $(p+1) \times m$. Such wavelet coefficients at different decomposition levels are related to one another by Mallat's pyramid algorithm (Mallat (1989b), Mallat (1989a)) in which convolutions of low- and high-pass wavelet filters, (h) and (g), respectively, take place in a cascade. The filters h and g are known as quadrature mirror filters. Given a low-pass wavelet filter $h = (h_0, \ldots, h_M)$, fully and uniquely specified by the choice of wavelet basis, the i^{th} entry of the high-pass counterpart g is $g_i = (-1)^{l-i} \cdot h_{M-s-i}$, for arbitrary but fixed integers l and s. We will further discuss the filter operators in the context of NDWT later in this section.

Expanding on the 1-D definitions, we overview a 2-D NDWT of $f(x, y)$, where $(x, y) \in \mathbb{R}^2$. Several versions of 2-D NDWT exist but we focus on the standard and a scale-mixing versions. For the standard 2-D NDWT, the wavelet atoms are

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

where (k_1, k_2) is the location pair, and $j = J_0, \ldots, J-1$ is the scale. The depth of the transform is $p = J - 1 - J_0$. The wavelet coefficients of $f(x, y)$ are calculated as

$$\begin{aligned}
c_{J_0;k_1,k_2} &= 2^{J_0}\iint f(x,y)\phi_{J_0;k_1,k_2}(x,y)\,dxdy,\\
d^{(i)}_{j;k_1,k_2} &= 2^{j}\iint f(x,y)\psi^{(i)}_{j;k_1,k_2}(x,y)\,dxdy,
\end{aligned}$$

where J_0 is the coarsest decomposition level, and $i \in \{h, v, d\}$ indicates the "orientation" of detail coefficients as horizontal, vertical, and diagonal (e.g., Vidakovic (1999), p. 155). The tessellation to a standard 2-D NDWT is presented in Figure 14(a).

For the scale-mixing 2-D NDWT, the wavelet atoms are

$$\begin{aligned}
\phi_{J_{01},J_{02};\boldsymbol{k}}(x,y) &= 2^{(J_{01}+J_{02})/2}\phi(2^{J_{01}}(x-k_1))\phi(2^{J_{02}}(y-k_2)),\\
\psi_{J_{01},j_2;\boldsymbol{k}}(x,y) &= 2^{(J_{01}+j_2)/2}\phi(2^{J_{01}}(x-k_1))\psi(2^{j_2}(y-k_2)),\\
\psi_{j_1,J_{02};\boldsymbol{k}}(x,y) &= 2^{(j_1+J_{02})/2}\psi(2^{j_1}(x-k_1))\phi(2^{J_{02}}(y-k_2)),\\
\psi_{j_1,j_2;\boldsymbol{k}}(x,y) &= 2^{(j_1+j_2)/2}\psi(2^{j_1}(x-k_1))\psi(2^{j_2}(y-k_2)),
\end{aligned}$$

where J_{01} and J_{02} are coarsest levels, $j_1 \geq J_{01}; j_2 \geq J_{02}$, and $\boldsymbol{k} = (k_1, k_2)$. As a result, we obtain wavelet coefficients for $f(x,y)$ from the scale-mixing NDWT as

$$\begin{aligned}
c_{J_{01},J_{02};\boldsymbol{k}} &= \iint f(x,y)\phi_{J_{01},J_{02};\boldsymbol{k}}(x,y)\,dxdy,\\
h_{J_{01},j_2;\boldsymbol{k}} &= \iint f(x,y)\psi_{J_{01},j_2;\boldsymbol{k}}(x,y)\,dxdy,\\
v_{j_1,J_{02};\boldsymbol{k}} &= \iint f(x,y)\psi_{j_1,J_{02};\boldsymbol{k}}(x,y)\,dxdy,\\
d_{j_1,j_2;\boldsymbol{k}} &= \iint f(x,y)\psi_{j_1,j_2;\boldsymbol{k}}(x,y)\,dxdy.
\end{aligned} \tag{53}$$

Notice that in the standard NDWT, we use common j to denote a scale, while in the scale-mixing NDWT, we use a pair (j_1, j_2), which indicates that two scales are mixed. Figure 14(b) illustrates the tessellation of coefficients of scale-mixing 2-D NDWT. In Section 2.2.1 we will refer to coefficients from (53) as c-,h-, v-, and d-type coefficients.

While the functional series involving wavelet and scaling functions as decomposing atoms is an established mathematical framework for describing the NDWT, we provide an alternative description of NDWT using convolution operators (Nason and Silverman (1995), Strang and Nguyen (1996), Vidakovic (1999)). Such a description is preferred for discrete inputs.

Let $[\uparrow 2]$ denote the upsampling of a given sequence by inserting a zero between every two neighboring elements of a sequence. We define the dilations of wavelet

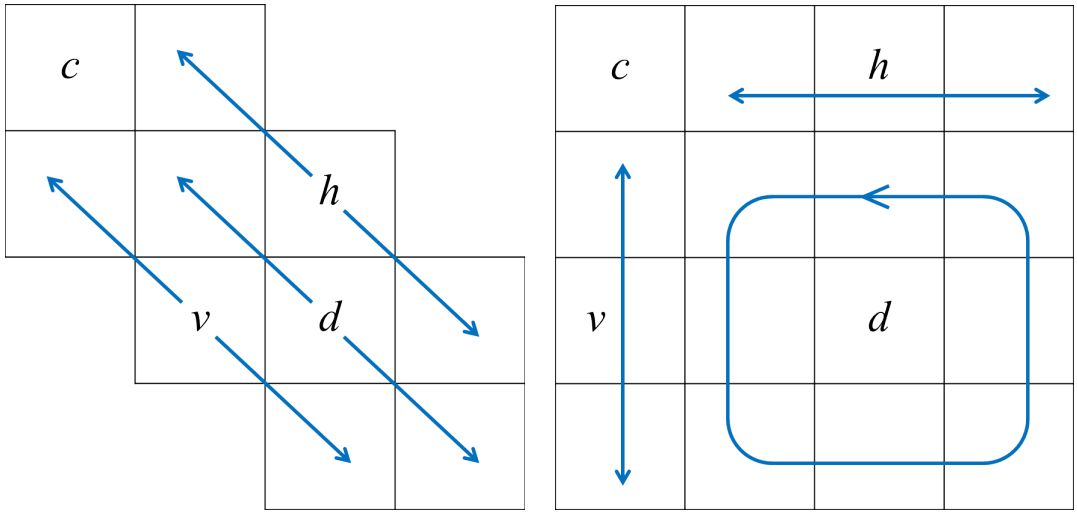


Figure 14: Locations of four types of wavelet coefficients in the tessellation of 3-level decomposition with the standard and scale-mixing 2-D NDWT's. Different types of coefficients are defined in (53).

filters h and g as

$$\begin{aligned} h^{[0]} &= h, & g^{[0]} &= g \\ h^{[r]} &= [\uparrow 2]h^{[r-1]}, & g^{[r]} &= [\uparrow 2]g^{[r-1]}. \end{aligned} \tag{54}$$

Inserting zeros between each element of filters $h^{[r-1]}$ and $g^{[r-1]}$ creates holes (*trous*, in French), which is why this approach is sometimes called Algorithm á Trous, see Shensa (1992).

A non-decimated wavelet transform is completed by applying convolution operators, $\mathbf{H}^{[j]}$ and $\mathbf{G}^{[j]}$, which come from dilated filters $h^{[r]}$ and $g^{[r]}$ in sequence. Detail and coarse coefficients generated from each level have an identical size, m, which is the same as that of the original signal. To obtain coefficients at decomposition level $J-j$, where $j \in \{1, 2, \ldots, p\}$, we repeatedly apply convolution operators to a coarse coefficient vector from the previous decomposition level, $J-j+1$

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

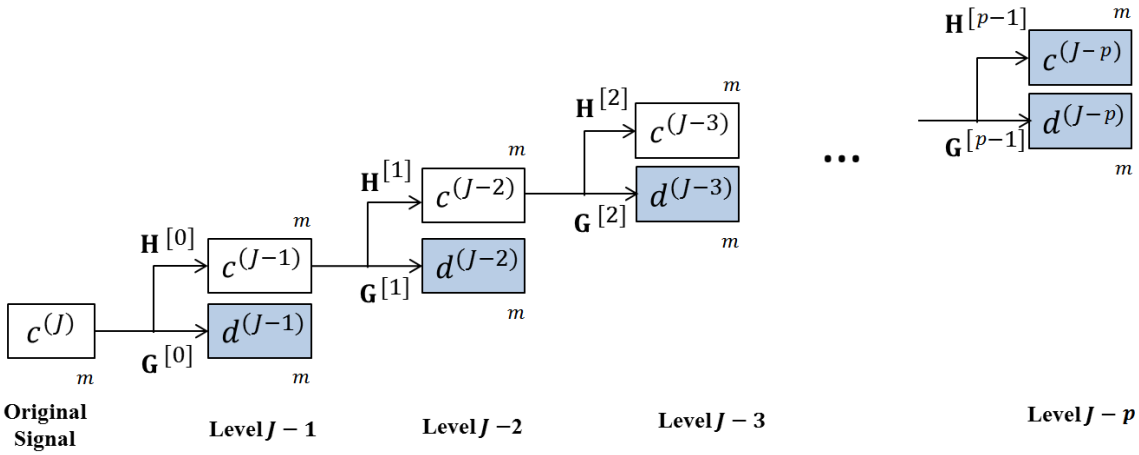


Figure 15: Graphical explanation of the NDWT process. Given signal a^J of size m, we decompose the signal to $p+1$ multi-resolution subspaces that include p levels of detail coefficients and one level of scaling coefficients, resulting in a set of coefficient vectors, $\boldsymbol{d}^{(J-1)}, \boldsymbol{d}^{(J-2)}, \ldots, \boldsymbol{d}^{(J-p)}$, and $\boldsymbol{c}^{(J-p)}$ in shaded blocks.

where $\mathbf{H}^{[j-1]}$ and $\mathbf{G}^{[j-1]}$ are filter operators that perform low- and high-pass filtering using quadrature mirror filters $h^{[j-1]}$ and $g^{[j-1]}$, respectively. The NDWT is the result of repeated applications of two filter operators, $\mathbf{H}^{[j]}$ and $\mathbf{G}^{[j]}$. Operators ($\mathbf{H}^{[j]}$, $\mathbf{G}^{[j]}$) do not have an orthogonality property, so to obtain such a property, we utilize two additional operators $\mathcal{D}_0$ and $\mathcal{D}_1$, which perform decimation by selecting every even and odd member of an input signal. An example of the use of the decimation operator $\mathcal{D}$ with a signal x is

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

where i indicates the position of an element in the signal x. We apply ($\mathcal{D}_0\mathbf{H}^{[j-1]}$, $\mathcal{D}_0\mathbf{G}^{[j-1]}$) and ($\mathcal{D}_1\mathbf{H}^{[j-1]}$, $\mathcal{D}_1\mathbf{G}^{[j-1]}$) to a given signal and obtain the even and odd elements of NDWT wavelet coefficient vectors, $\boldsymbol{c}^{(J-j)}$ and $\boldsymbol{d}^{(J-j)}$, respectively. Thus,

equation (54) is, in fact, performed as the following process

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

We apply the filtering twice at the even and odd positions for each decomposition level, so a shift does not affect transformation results, which means that the NDWT is time-invariant. Such time-invariance property of the NDWT yields a smaller mean squared error and reduces the Gibbs phenomenon in de-noising applications (Coifman and Donoho, 1995). However, the violation of variance preservation in the NDWT complicates the signal reconstruction. In the following section we will discuss how to perform lossless reconstruction of an original image using a matrix-based NDWT.

2.2 Matrix Formulation of NDWT

In this section, we translate multiple convolutions in the NDWT into a simple matrix multiplication. In Mallat's algorithm, scaling and wavelet functions are convolved in a cascade. Instead of performing convolutions with wavelet and scaling functions, we formulate the NDWT as matrix multiplication. We simplify the cascade algorithm as follows. With filtering matrices, Mallat's cascade algorithm is implicit in repeated matrix multiplications of low- and high-pass filter matrices, (H) and (G), respectively. The following matrices illustrate combination of component filter matrices, to achieve transforms of depth 1, 2, and 3.

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

Filter matrices $\begin{bmatrix} H_p & G_p \end{bmatrix}^T$ as submatrices of $W_m^{(p)}$ are formed by simple rules. The sizes of H_p and G_p for $p \in \{1, 2, \ldots\}$ are the same, $m \times m$, and their entries at the position (i, j) are

$$\begin{aligned} a_{ij} = &\quad \tfrac{1}{\sqrt{2}} h_s^{[p-1]}, &&s = N + i - j \ modulo \ m \\ b_{ij} = &\quad \tfrac{1}{\sqrt{2}} (-1)^s h_{N+1-s}^{[p-1]}, &&s = N + i - j \ modulo \ m, \end{aligned}$$

respectively, where N is a shift parameter and $h_s^{[p-1]}$ is the s^{th} element of a dilated wavelet filter h with $p-1$ zeros in between the original components $(h_1, h_2, \ldots, h_u)$,

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

For example, $h_1^{[p-1]} = h_1$, $h_{p+1}^{[p-1]} = h_2$, ..., and, $h_{p(u-1)+1}^{[p-1]} = h_u$. Following such construction rules, $W_m^{(p)}$ becomes a matrix of size $\big(m(p+1) \times m\big)$ consisting of $p+1$ stacked submatrices of size $[m \times m]$. The NDWT matrix formed in the described process is not normalized and signal reconstruction cannot be done by using its transpose only. Indeed, in terms of Mallat's algorithm, for the inverse transform, at each step the multiplication by $1/2$ is needed for perfect reconstruction (see Mallat, 1999, Proposition 5.6).

Thus, we construct a diagonal weight matrix that rescales the square submatrices comprising the NDWT matrix, to be used when performing the inverse transform.

The weight matrix for $W_m^{(p)}$ has size $(m(p+1) \times m(p+1))$ and is defined as

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

A 1-D signal $\boldsymbol{y}$ of size $[m \times 1]$ is transformed in a p-level decomposition to a vector $\boldsymbol{d}$ by multiplication by wavelet matrix $W_m^{(p)}$. The original signal is then reconstructed by multiplying $\boldsymbol{d}$ by $(W_m^{(p)})'$ rescaled by the weight matrix $T_m^{(p)}$.

$$\begin{aligned} \boldsymbol{d} &= W_m^{(p)} \times \boldsymbol{y}_{[m \times 1]} \\ \boldsymbol{y} &= (W_m^{(p)})' \times T_m^{(p)} \times \boldsymbol{d}, \end{aligned} \tag{55}$$

where p and m are arbitrary.

Note that $(W_m^{(p)})' \times W_m^{(p)} \neq I_m$. On the other hand, column vectors of matrix $V_m^{(p)} = \left(T_m^{(p)}\right)^{1/2} W_m^{(p)}$ form an orthonormal set, that is,

$$(V_m^{(p)})' \times V_m^{(p)} = I_m \tag{56}$$

The product $V_m^{(p)} \times (V_m^{(p)})'$ cannot be an identity matrix, but

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

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

Since $I_m = (V_m^{(p)})' \times V_m^{(p)} = (W_m^{(p)})' \times T_m^{(p)} \times W_m^{(p)}$, the perfect reconstruction is achieved by $(W_m^{(p)})' \times T_m^{(p)}$ applied on the vector transformed by $W_m^{(p)}$, as in (55).

Although transformation by $V_m^{(p)}$ looks more natural because of (56), the scaling of wavelet coefficients when transformed by $V_m^{(p)}$ is not matching the correct scaling produced by Mallat's algorithm, or equivalently, by integrals in (52). The correct scaling of wavelet coefficients is important in applications involving regularity assessment of signals and images, as we will see in the mammogram example from Section 2.4.

2.2.1 Scale-Mixing 2-D NDWT

A 2-D signal $\boldsymbol{A}$ of size $[m \times n]$ for p_1- and p_2-level decomposition along rows and columns, respectively, is obtained by NDWT matrix multiplication from the left and its transpose from the right. The transform results in a 2-D signal $\boldsymbol{B}$ of size $(p_1 + 1)m \times (p_2 + 1)n$. The inverse transform applies the rescaling matrices $T_m^{(p_1)}$ and $T_n^{(p_2)}$ on the corresponding NDWT matrices,

$$\begin{aligned} \boldsymbol{B} &= W_m^{(p_1)} \times \boldsymbol{A}_{[m\times n]} \times (W_n^{(p_2)})' \\ \boldsymbol{A} &= (W_m^{(p_1)})' \times T_m^{(p_1)} \times \boldsymbol{B} \times T_n^{(p_2)} \times W_n^{(p_2)}, \end{aligned} \tag{57}$$

Here p_1, p_2, m, and n can take any integer value, and $W_m^{(p_1)}$ and $W_m^{(p_2)}$ could be constructed using possibly different wavelet filters.

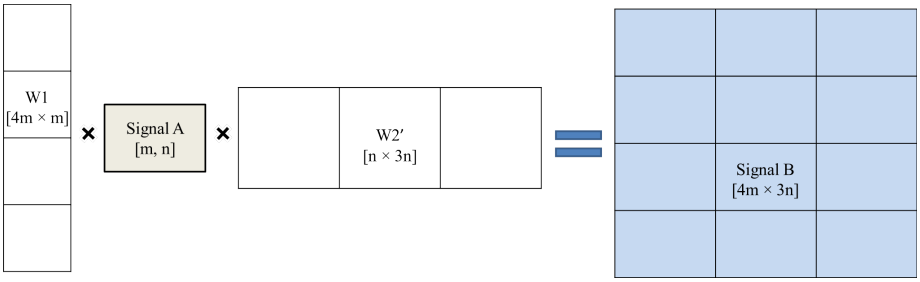


Figure 16: Graphical illustration of a 2-D NDWT scale-mixing transform with 3-levels along the columns and 2-levels along the rows. The NDWT matrices W_1 and W_2 can be constructed by possibly different wavelet filters.

One of the advantages of the scale-mixing 2-D NDWT is its superior compressibility. Wavelet transforms act as approximate Karhunen-Loève transforms and compressibility in the wavelet domain is beneficial in tasks wavelet-based data compression and denoising. When an image possesses a certain degree of smoothness, the coefficients corresponding to diagonal decomposition atoms [d-coefficients in (53)] tend to be smaller in magnitude compared to the c-, v- or h-type coefficients in (53). As an example, consider performing a p-level decomposition of a 2-D image of size $[m\times n]$ with the both NDWT matrix (as scale-mixing) and standard 2-D NDWT. The

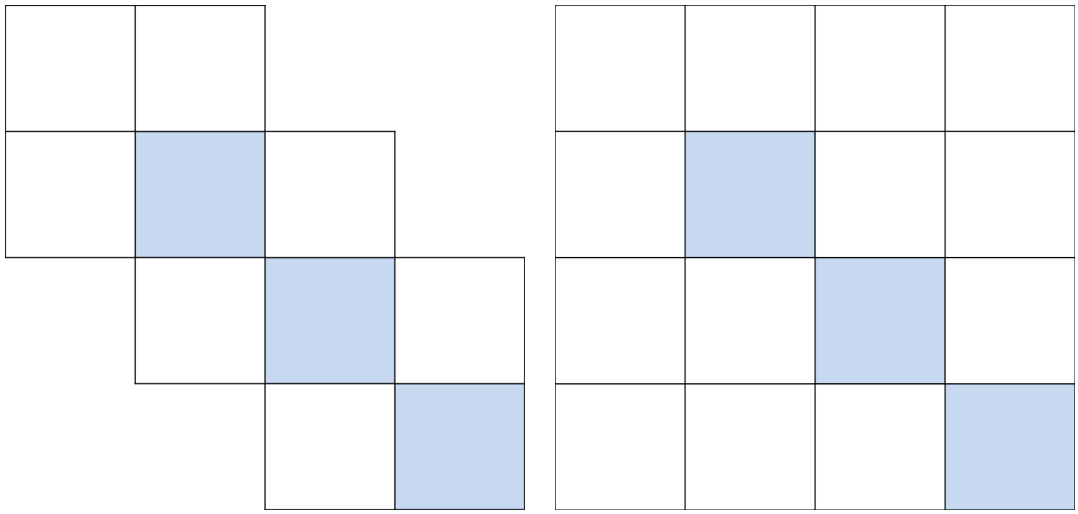
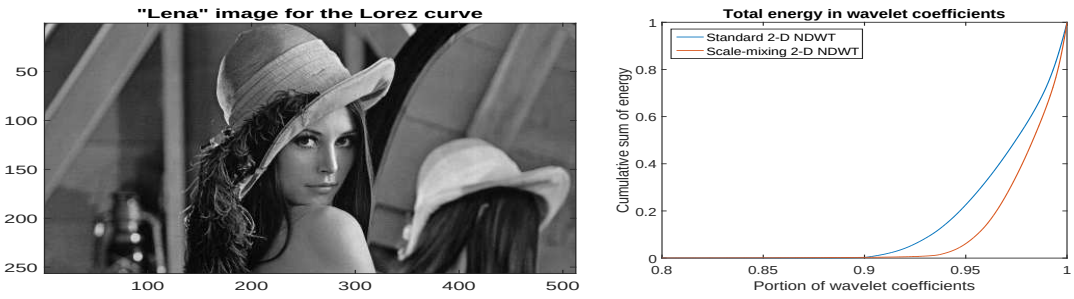


Figure 17: Tesselation of 3-level decomposition with standard (left) and scale-mixing (right) 2-D NDWT. Shadded areas correspond to d-type wavelet coefficients

compressibility of transform can be defined as the proportion of diagonal-type coefficients divided by the total number of wavelet coefficients. As we mentioned before, d-coefficients correspond to decomposing atoms consisting of two wavelet functions, while the atoms of c-, v- or h-type coefficients contain at least one scaling function. In the scale-mixing NDWT of depth p, $p^2mn/((p+1)^2mn)$ is the proportion of d-type coefficients, while in the standard 2-D NDWT this proportion is $pmn/((3p+1)mn)$ (see Figure 17). The former is always greater than the later, except when $p = 1$, in which case the two proportions coincide. Thus, the scale-mixing 2-D NDWT tends to be more compressive compared to the standard 2-D NDWT.

As an illustration, we transform a noiseless "Lena" image of size 256×512 (Figure 18(a)) with both the standard and scale-mixing 2-D NDWT in a 3-level decomposition using the Haar wavelet. To compare the compressibility, we calculate and contrast Lorenz curves. For the Lorenz curve, we normalize all squared wavelet coefficients as $p_k = d_k^2/(\sum_i d_i^2)$, sort p_k in an increasing order, and obtain the cumulative sum of sorted p_k. This cumulative sum of (normalized) energy for the two transforms is



(a) (b)

Figure 18: Image in panel (a) is transformed with both scale-mixing and standard 2-D NDWT into 3 decomposition levels based on a Haar wavelet filter. A detail of the Lorenz curve in panel (b) indicates that the 2-D scale-mixing NDWT compresses the information in image more efficiently compared to the traditional transform.

shown Figure 18b). The curves are plotted against the portion of wavelet coefficients used in the cumulative sum. At top right corner of Figure 18b, the curves meet, since for both curves $\sum p_i = 1$. However, the blue curve (standard NDWT) uniformly dominates the red curve (scale-mixing NDWT). This means that the compressibility of the scale-mixing NDWT is higher. In simple terms, the scale-mixing NDWT requires smaller portion of the wavelet coefficients to preserve the same relative "energy." To numerically quantify this compressibility, we think of p_k's as the probabilities and calculate entropies of their distributions. Calculating the normalized Shannon entropy, $(\sum_{i=1}^{n} p_i \log p_i)/\log n$, we obtain 0.7994 for the scale-mixing NDWT and 0.8196 for the standard NDWT. The scale-mixing NDWT has lower entropy, which confirms its superior compressibility. Although demonstrated here only on "Lena" image, this superiority in compression for scale-mixing transforms holds generally, see Remenyi et al. (2014).

2.3 Computational Efficiency of the NDWT Matrix

Next we discuss several features of NDWT matrix, so that users are aware of its advantages as well as limitations.

Principal advantages of a NDWT matrix are compressibility, computational speed,

and flexiblity in size of an input signal. We already discussed the better compressibility when NDWT matrices are used for 2-D scale-mixing transforms.

Next, we compare the computation time of the matrix-based NDWT to that of the convolution-based NDWT. The NDWT matrix performs a transform faster than the convolution-based NDWT. This statement is conditional on the software used for the computation. We used MATLAB version 8.6.0.267246 (R2015b, 64-bit) on a laptop with quad-core CPU running at 1,200 MHz with 8GB of RAM.

At first glance, improving the speed of calculation by using matrix multiplication over convolutions looks counterintuitive. The asymptotic computational complexity for convolutions is much lower than the complexity of matrix multiplication. The NDWT based on Mallat's algorithm has calculational complexity of $O(n \log n)$, while the (naïve) matrix multiplication has the complexity of $O(n^3)$. The complexity of matrix multiplication could be improved by the Le Gall (2014) algorithm to $O(n^{2.3729})$, with a theoretical lower bound of $O(n^2 \log n)$, still inferior to convolutions. However, the "devil is in the constants." For signals of moderate size, the calculational overhead that manages repeated filtering operations in convolution-based approach slows down the computation and direct matrix multiplication turns out to be faster.

As an illustration, we simulated 100 2-D fractional Brownian fields (fBf) of size $(2^{10} \times 2^{10})$ with the Hurst exponent $H = 0.5$ and performed the eight-level decomposition NDWT with four wavelets: Haar, Daubechies (4 and 6 tabs), and Coiflet. For the NDWT of a single signal, the computation time of the matrix-based NDWT was on average of 9.02 seconds while that of the convolution-based NDWT was on average of 17.26 seconds. In addition, about 40 % of the computation time of the matrix-based NDWT was spent on constructing an NDWT matrix that could be used repeatedly in simulation for the same type of NDWT once generated. Thus, as the NDWT is repeated on the input signals of the same size using the same wavelet filter, the difference in computation time becomes even greater. For a 1-D signal transform, the

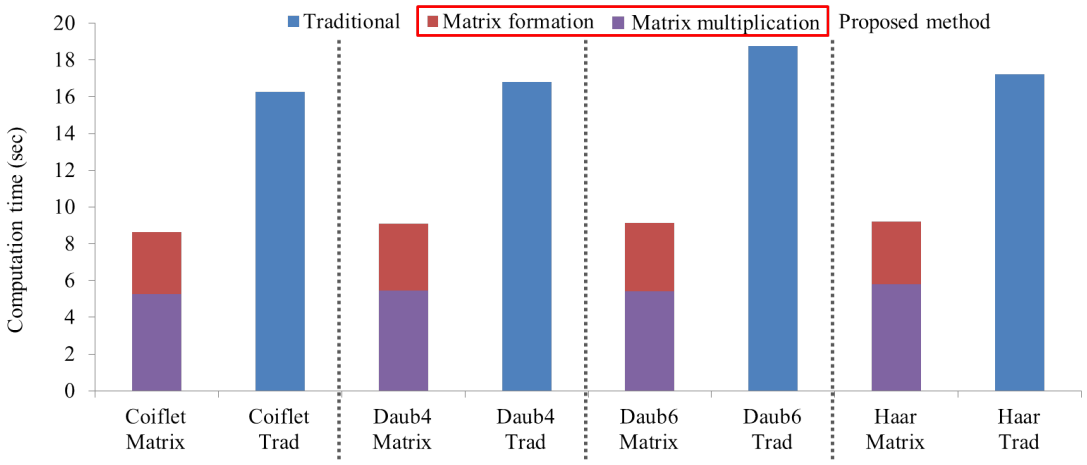


Figure 19: Average computation time (in seconds) to perform the matrix-based and convolution-based NDWT for 8-level decompositions along both rows and columns using Coiflet, Daubechies 4, Daubechies 6, and Haar wavelets. The size of inputs is $2^{10} \times 2^{10}$ and the computation time is averaged over 100 repetitions.

matrix-based NDWT is approximately twice as fast as the convolution-based NDWT under the given conditions, but this factor increases to three for the NDWT of 100 signals having the same size and transformed using the same matrix.

While NDWT matrices reduce the computation time by storing all entries of the matrices used in convolution for each decomposition level in a single matrix, such property can limit the usage of NDWT matrices. When the size of an input is large, a computer with standard specifications may not have enough memory to store a NDWT matrix of appropriate size. This issue affects mostly the cases of 1-D signals. For 2-D transforms, if the computer can store an image, it can most likely store the NDWT matrix, since the matrix is only $(p+1)$ times larger, and p is typically small. To find a limit on the size of an 1-D input, we repeatedly constructed NDWT matrices for one-level decomposition increasing the size of an input by 500 in each trial. We found that as the size of an input signal exceeded $35,000$, matrix construction was not possible because of limited memory capacity.

The matrix-based NDWT can be applied to signals of a non-dyadic length and for 2-D applications, to rectangular signals of possibly non-dyadic sides. Typically, the

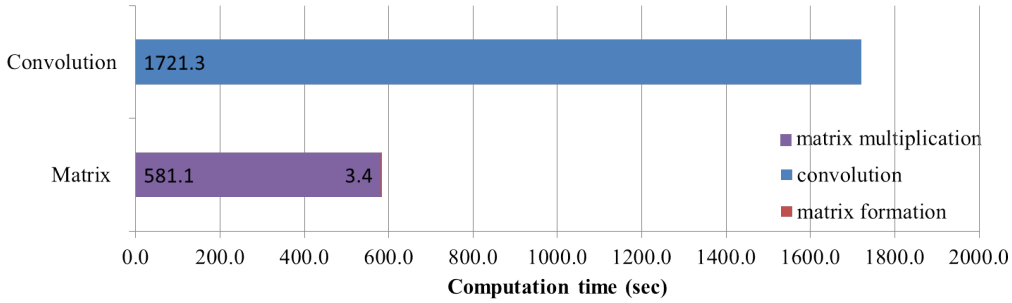


Figure 20: Computation time (in seconds) of the matrix- and convolution-based NDWT for 8-level decomposition evaluated for 100 2-D signals of the size $(1,024 \times 1,024)$ using Haar wavelet. The matrix was pre-constructed to perform the same type of transform.

standard convolution-based NDWT can only manage dyadic or squared 2-D input signals of dyadic scale (e.g., Wavelab).

2.4 Two Examples of Application

In this section, we provide two applications in which the package **WavmatND** is used. In the first application we apply our matrix-based NDWT to obtain a scaling index from the background of a mammogram image. The scaling index of an image is measured by Hurst exponent, a dimensionless constant in interval $[0, 1]$. For locally isotropic medical images, the Hurst exponent is known to be useful for diagnostic purposes (Ramirez and Vidakovic (2007), Nicolis et al. (2011), Jeon et al. (2014)). Wavelet-based spectra of an image is defined on a selected hierarchy of multiresolution spaces in a wavelet representation as a set of pairs $(j, S(j))$, where j is the multiresolution level and $S(j)$ is the logarithm of the average of squared wavelet coefficients at that level. The Hurst exponent, as a measure of regularity of the image, is functionally connected with the slope of a linear fit on pairs $(j, S(j))$. Any type of wavelet decomposition can serve as a generator of wavelet spectra, and in this application we look at 2-D scale-mixing NDWT of a digital mammogram.

The digital mammogram analyzed comes from the Digital Database for Screening Mammography (DDSM) at the University of South Florida. The image is digitized

by HOWTEK scanner at the full 43.5-micron per pixel spatial resolution and features craniocaudal (CC) projection. A detailed description of the data can be found in Bowyer et al. (1996). Figure 21 shows the location of the region of interest (ROI)

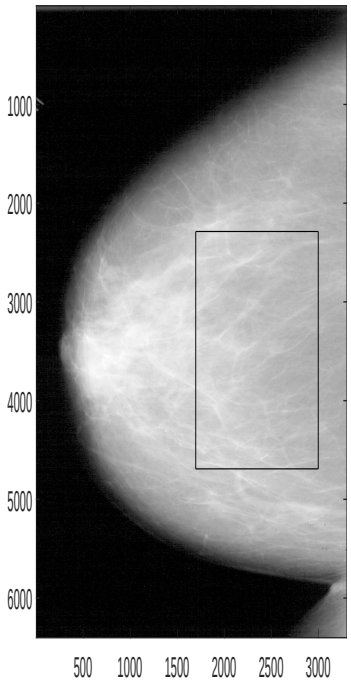


Figure 21: A ROI in a mammogram image selected for the estimation of scaling.

within the mammogram. We selected the ROI of size 2401×1301 and transformed it to a scale-mixing 2-D non-decimated wavelet domain. The spectral slope is estimated from the levelwise log-average squared coefficients along the diagonal hierarchy of multiresolution spaces, comprising the wavelet spectra, as in Figure 22. The slope of -2.6722 gives the Hurst exponent of $-(slope + 2)/2 = 0.3361$. Details can be found in 3 who use the Hurst exponent estimators to classify the mammograms from the DDSM data base for breast cancer detection.

In the second example, we denoise a signal captured by an atomic force microscope. The atomic force microscopy (AFM) is a type of scanned proximity probe microscopy that measures the adhesion strength between two materials at the nanonewton scale. The AFM data from the adhesion measurements between carbohydrate and the cell

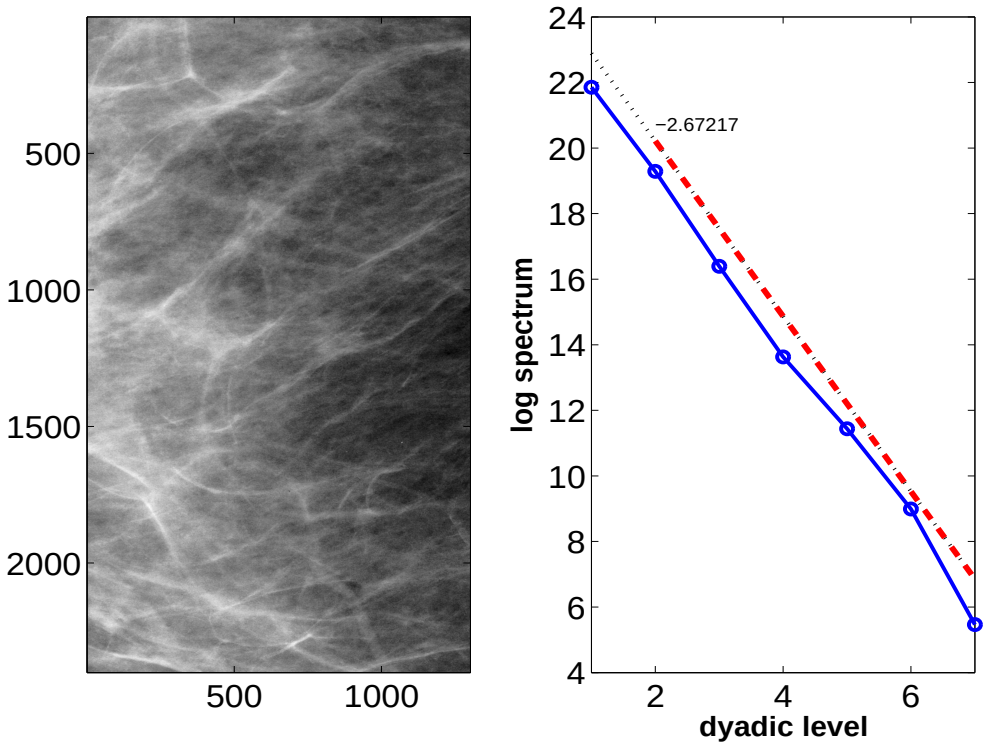


Figure 22: Left panel: The selected region of interest (ROI) where the Hurst exponent is estimated. Right panel: The dash-dotted line represents 2-D non-decimated wavelet spectra of the ROI from the left panel. The dashed line shows the regression result using the corresponding energy levels.

adhesion molecule (CAM) E-Selectin was collected by Bryan Marshall from the Department of Biomedical Engineering at Georgia Institute of Technology. The technical description and details are provided in Marshall et al. (2005).

In AFM, a cantilever beam is adjusted until it bonds with the surface of a sample, and then, the force required to separate the beam and sample is measured from the beam deflection. Beam vibration can be caused by external factors such as thermal energy of the surrounding air and the footsteps of someone outside the laboratory. The vibration of a beam shows as noise on the deflection signal. For denoising purposes, we decomposed AFM signal of size 3,000 into 10 decomposition levels using the NDWT with a 6-tab Daubechies wavelet (3 vanishing moments) and applied hard thresholding on wavelet coefficients. The threshold for this process is set as $\sqrt{2\log m}\,\hat{\sigma}$, where $\hat{\sigma}$ is an estimator of standard deviation of noise present in the wavelet coefficients at the

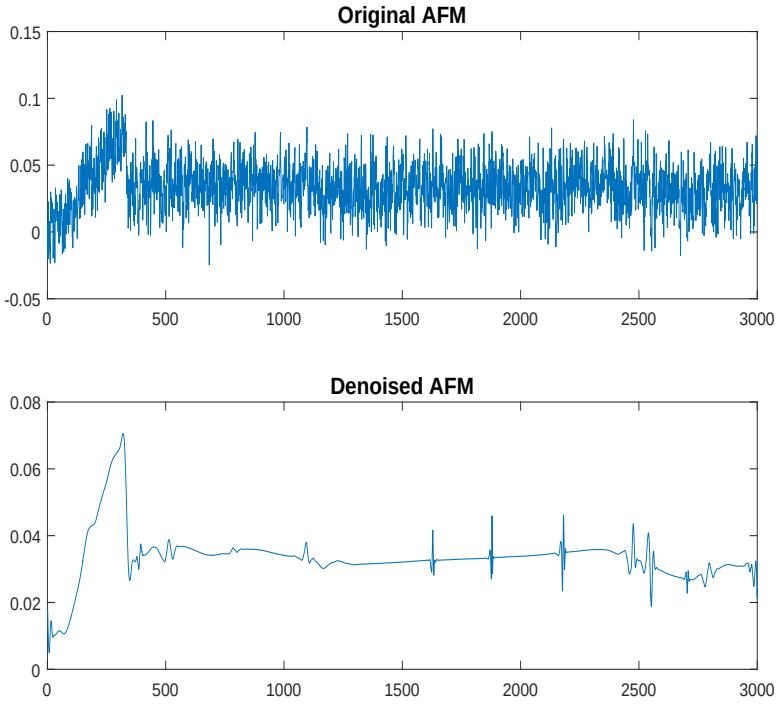


Figure 23: Denoising of AFM by hard-thresholding on NDWT coefficients with 6-tab Daubechies wavelet.

finest level of detail, and m is the size of the original signal. Given the redundancy of the transform, we estimate $\hat{\sigma}$ by averaging two estimators, $\hat{\sigma}_o$ and $\hat{\sigma}_e$ as $\sqrt{(\hat{\sigma}_o + \hat{\sigma}_e)/2}$. Estimators $\hat{\sigma}_o$ and $\hat{\sigma}_e$ are sample standard deviations of wavelet coefficients at every odd and even locations, respectively, within the finest level of detail. Figure 23 shows the noisy AFM signal and its denoised version. The researchers are particularly interested in the shape of the signal for the first 350 observations of an AFM signal, prior to cantilever detachment.

2.5 Package Description and Demos

The MATLAB package, **WavmatND**, includes two core functions, several additional functions, and data sets needed for illustrative examples and demos.

2.5.1 Core Functions

`WavmatND()` is a core function that generates a transform matrix. Inputs to this function are a wavelet filter, size of an input signal, the depth of transformation, and a shift. The shift corresponds to parameter s in the definition of quadrature mirror filter $g_i = (-1)^{l-i} h_{M-s-i}, \;\; i = 0, \ldots, M$, and is usually taken as 0 or 2.

`weight()` generates a weight matrix that rescales every submatrix in the inverse wavelet transform. This matrix is necessary for the lossless inverse transform, as in (57). It assigns different weights to each submatrix, as described in Section 2.2. Inputs to this function are the size of the original signal and the depth of the transform.

2.5.2 Other Functions and Data Sets Included

For the illustration purposes, we include a custom made function `WaveletSpectra2NDM.m` for assessing the scaling in images based on 2-D NDWT.

`WaveletSpectra2NDM()` estimates a scaling index of an image using the diagonal hieararchy of nested multiresolution spaces in a 2-D scale-mixing NDWT. It returns the average level-energies for a specified range of levels, scaling slope, and a graph showing linear regression fit of log energies on the selected levels. The inputs are 2-D data/image, the depth of transform, a wavelet filter, a range of levels used for the regression, and an option for showing the plot (1 for a plot and 0 for no plot).

`NDWT2D()` is a function that performs a standard 2-D NDWT using NDWT matrices. It returns c-, h-, v-, and d-types of wavelet coefficients. In this transform there is no scale mixing and x-scale is the same as the y-scale. Inputs to this function are an image, a wavelet filter, the depth of transform, and a shift.

`filters.m` contains some commonly used wavelet filters needed for construction of a NDWT matrix. It provides Haar, Daubechies 4-20, Symmlet 8-20, and Coiflet 6,

12, and 18 filters with high accuracy. Users can choose an appropriate wavelet filter based on type of analysis and input data, compromising between the smoothness and locality.

`afm.mat` and `tissue.mat` are data sets used in the two applications. Interested readers can load the data sets for further analysis. We also included the code used to generate the results in the thesis at `exampleApplications.m`.

`lena.mat` is well-known image of Lena Söderberg, one of the most used images in signal processing community. This image is utilized in DEMO 1 explained in the next section.

2.5.3 DEMO 1: Transform and reconstruction

As we discussed earlier, a matrix-based NDWT maps an original data set into a time-scale domain with efficient and simple steps. In the following code, we load image `lena`, of size (256×512) and create two NDWT matrices `W1` and `W2` that perform the NDWT on image by columns and rows, respectively. We use the Haar wavelet and perform a p-depth NDWT in both columns and rows for $p = \log(\min(m, n)) - 1 = 7$.

```
load lena; [n m]=size(lena);
p=floor(log(min(m,n)))-2; shift=0;
h = [1/sqrt(2) 1/sqrt(2)];
W1=WavmatND(h,n,p,shift); W2=WavmatND(h,m,p,shift);
tlena=W1*lena*W2';
```

The reconstruction of the transformed lena `tlena` is simple. We generate weight matrices, `T1` and `T2`, of the sizes compactible with `W1` and `W2`, respectively, and reconstruct the signal sa follows:

```
T1=weight(n,p); T2=weight(m,p);
rlena=W1'*T1*tlena*T2*W2;
```

The reconstructed signal is `rlena`. The transformation and reconstruction are illustrated in Figure 24.

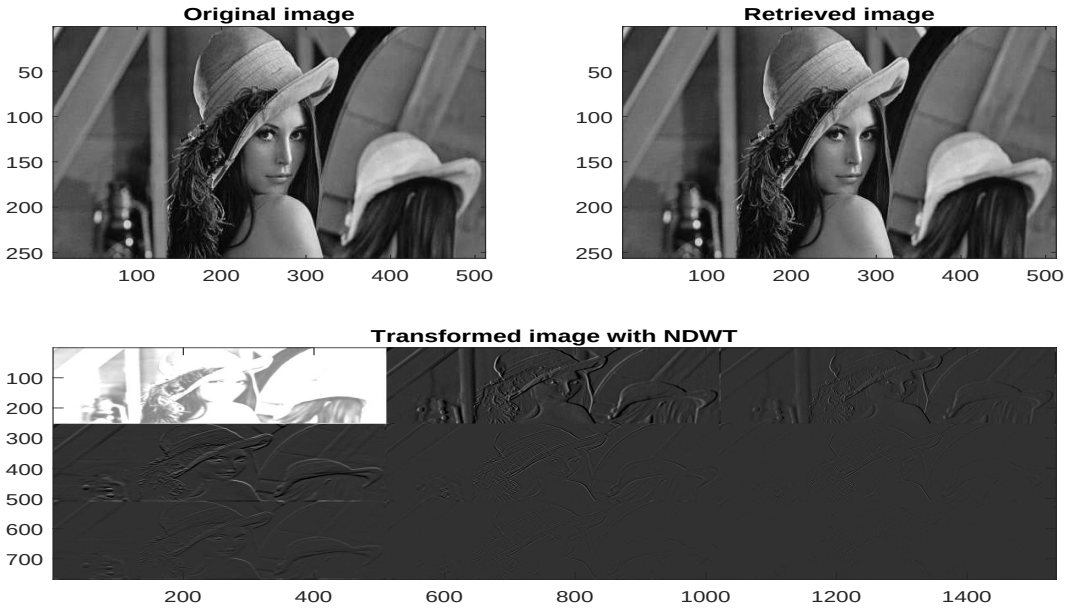


Figure 24: Lena image in the original and wavelet domains from Demo 1.

2.5.4 DEMO 2: Denoising of Doppler Signal

In this demo, we first generate a Doppler signal `s` of size 1,000 and a matching signal `noise` consisting of i.i.d normal variables with mean 0 and variance 0.05^2. The sum of `s` and `noise` constitutes a noisy signal `sn` with signal-to-noise ratio of 5.78.

```
sigma=0.05; m=250;
t = linspace(1/m,1,m);
s = sqrt(t.*(1-t)).*sin((2*pi*1.05) ./(t+.05));
noise=normrnd(0,1,size(s))*sigma;
sn=s+noise;
```

Next, with a Haar wavelet, we generate the NDWT matrix, `W`, which decomposes the signal into $\big(\lfloor \log(1000) \rfloor - 1\big)$ decomposition levels. The resulting wavelet coefficients are in `tsn`.

```
J=floor(log2(m)); k=J-1;
```

```
qmf = [1/sqrt(2) 1/sqrt(2)];
W = WavmatND(qmf,m,k,0);
T = weight(m, k);
tsn=W*sn';
```

Then, we apply hard thresholding for denoising. Hard thresholding is applied to all detail level subspaces, and the threshold is set to be $\sqrt{2\log(m)}\hat{\sigma}$, where $\hat{\sigma}$ is the square root of average of variances of wavelet coefficients at odd and even positions at the finest level of detail, and m is the length of the original signal.

```
sigma2hat=(var(tsn(end-m+1:2:end))+var(tsn(end-m:2:end)))/2;
threshold=sqrt(2*log(k*m)*sigma2hat);
snt= tsn(m+1:end).*(abs(tsn(m+1:end))>threshold);
rs=W'*T*[tsn(1:m); snt];
```

The reconstructed denoised signal is `rs`.

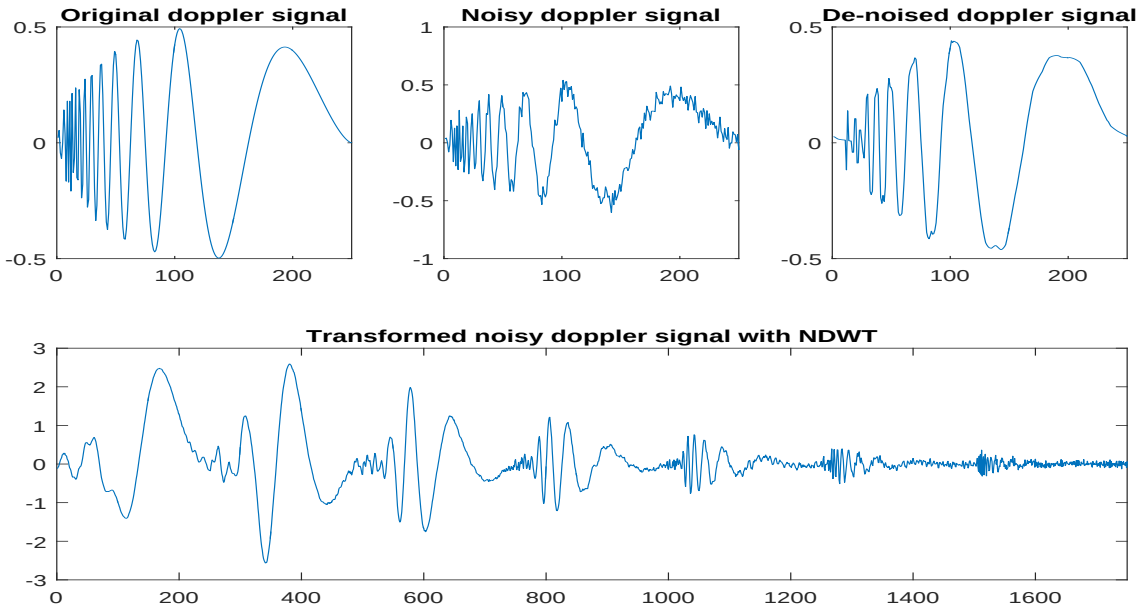


Figure 25: Doppler signals in time and wavelet domains for denoising from Demo 2.

2.6 Discussion

The non-decimated wavelet transform (NDWT) possesses properties beneficial in various wavelet applications. We developed MATLAB package,**WavmatND**, which performs the NDWT in one or two dimensions. Instead of repeated convolutions that are standardly performed, the NDWT is performed by matrix multiplication. This significantly decreased the computation time in simulations when performed in MATLAB computing environment. This reduction in computation time is additionally augmented when we applied the NDWT repeatedly to signals of the same size, decomposition level, and choice of wavelet basis. In 2-D case, the NDWT matrix yields a scale-mixing NDWT, which turns out to be more compressive compared to the standard 2-D NDWT. For lossless retrieval of an original signal, we utilize a weight matrix. We also relax the constraint on the size of input signals so that the NDWT could be performed on signals of non-dyadic size in one or two dimensions. We hope that this stand-alone MATLAB package will be a useful tool for practitioners interested in various aspects of signal and image processing.

The package **WavmatND** can be downloaded from Jacket's Wavelets WWW repository site `http://gtwavelet.bme.gatech.edu/`.

Chapter III

NON-DECIMATED WAVELET SPECTRA AND ITS APPLICATION

In this chapter, we apply the NDWT-based scaling estimation method introduced in section 1.9.2 to simulated and mammographic images and improve scaling parameter estimation.

The method is illustrated with an application of the breast cancer detection. Breast cancer is the most common form of cancer in terms of incidence and the second most common form of cancer with regards to cancer mortality in women in the United States. The early detection of the breast cancer is crucial for patients' survival because the survival rates significantly decrease as the breast cancer metastasizes American-CancerSociety (2014). Mammographic screening is the most common means of breast cancer screening for the early detection of breast cancer. However, even experienced radiologists misdiagnose up to 30% of mammograms because breast tissue is complex and signatures of disease can be subtle Martin et al. (1979). In addition, multiple mammographic screenings might be required to confirm the cancer and each screening is costly and stressful to the subjects. A number of existing computer methods for breast cancer detection focus on the detection of specific markers, such as tumors or micro-calcifications. In this chapter, we characterize the self-similar properties of normal breast tissue with NDWT, where non-normal tissue is a potential marker of breast cancer.

To assess the health conditions based on mammographic images, we utilize a method based on non-decimated wavelet spectra that is introduced in section 1.9.2. The parameters describing spectral regularity form a battery of spectral summaries

which describe the self-similarity and the degree of fractality present in mammogram images. Diagnostics of breast cancer based on scaling measures of mammograms obtained with orthogonal wavelet transform (DWT) and linear regression can be found in Nicolis et al. (2011). For the same task, multifractal spectral tools have been used in Ramírez-Cobo and Vidakovic (2013) while the complex wavelets have been utilized by Jeon et al. (2014). Extending on the aforementioned results, we develop a wavelet spectral scaling estimation method based on the non-decimated wavelet transform (NDWT). The NDWT provides two features that improve on the scaling estimation: First, the NDWT as a redundant transform, yields wavelet coefficients at a maximal sampling rate independently of the multiresolution level. Thus, we obtain the maximum number of wavelet coefficients at all levels, which improves the stability and accuracy of estimation. Second, the size of an input signal is maintained at each resolution level. This enables us to localize wavelet coefficients corresponding to a region of interest (ROI) at any level in the wavelet domain. We highlight such features in a simulation study and an application for breast cancer detection with mammogram images. In addition, non-decimated transforms do not require dyadic size inputs, which is a constraint for wavelet transforms that decimate. To understand benefits of NDWT in scaling estimation, we start by estimating simulated signals and compare the estimation performance of the methods.

3.1 Simulated cases

In a simulation study, we compare the performance of NDWT- and DWT-based scaling estimation methods. We simulate three sets of 500 2-D fBm signals of size $2^9 \times 2^9$ with Hurst exponents 0.3, 0.5, and 0.7, respectively. Next, we transform each signal into 8 multiresolution subspaces for both row and column decompositions with a 2-D scale-mixing NDWT based on four wavelets: Daubechies 6, Symmlet 8, Coiflet 6, and Haar. The wavelets are indexed by number of filter components and

not by the number of vanishing moments. We found that the estimation performance is robust with respect to choice of wavelet filter. In estimating the scaling, we take logarithm on part of the main diagonal wavelet spectrum which includes level (j, j) where $j = 2, 3, \ldots, 6$, and then fit a linear regression model on log average level energies. The slope of this linear regression leads the estimator of Hurst exponent. We evaluate the performance of both NDWT- and DWT- based estimation methods, by comparing their means, variances, and biases.

Unlike the decimated case in which the number of coefficients differs at each level, and correct linear fitting procedure involves weighted regression Veitch and Abry (1999), here the ordinary least square (OLS) regression provides theoretically correct and satisfactory fit. Because of redundancy, dependence among neighboring wavelet coefficients within the same level is much more pronounced than in the case of orthogonal wavelets. Although this dependence is not biasing estimators, the variances of estimators are affected. Another factor that influences variance is the choice of wavelet basis, and more local wavelet bases are preferred.

Tables 5-7 summarize the estimation results under various settings. An average of $\hat{H}$, its mean-square error, variance, and bias-squared, based on 300 2-D fBm's when true Hurst exponents, $H = 0.3, 0.5$ and 0.7, are provided. Symmlet 8, Daubechies 6, Coiflet 6, and Haar wavelet bases are used in non-decimated and orthogonal versions of wavelet spectra. Figure 26-28 show box-and-whisker plots of simulations described in Tables 5-7.

Because NDWT produces the maximum number of wavelet coefficients at each resolution level, we are able to obtain more accurate estimates of energies with more coefficients to average at each level. Thus, the NDWT-based method yields estimators with lower mean squared errors compared to the DWT-based method. As it can be seen in Tables 5-7 and Figures 26-28, the empirical variances are influenced by the choice of wavelet. The redundancy of NDWT turned out to be beneficial despite some

Table 5: An average of $\hat{H}$, its mean-square error, variance, and bias-squared, based on 300 2-D fBm's when true $H = 0.3$ obtained by various wavelet-bases and transform choices, i.e., non-decimated and orthogonal.

Wavelets	Symmlet 8		Daubechies 6		Coiflet 6		Haar	
	ND	Ortho	ND	Ortho	ND	Ortho	ND	Ortho
Mean	0.2946	0.2945	0.2955	0.2939	0.2963	0.2939	0.2959	0.2632
MSE	0.0017	0.0043	0.0016	0.0041	0.0015	0.0041	0.0012	0.0047
Variance	0.0017	0.0043	0.0015	0.0041	0.0015	0.0041	0.0012	0.0034
Bias	2.2E-5	1.3E-5	1.4E-5	2.2E-5	7.9E-6	2.2E-05	1.2E-5	0.0013

Table 6: As in Table 5 but for $H = 0.5$.

Wavelets	Symmlet 8		Daubechies 6		Coiflet 6		Haar	
	ND	Ortho	ND	Ortho	ND	Ortho	ND	Ortho
Mean	0.5115	0.5096	0.5109	0.5112	0.5109	0.5125	0.4703	0.5153
MSE	0.005	0.002	0.0046	0.0019	0.0046	0.0018	0.0044	0.0018
Variance	0.0049	0.0019	0.0045	0.0018	0.0045	0.0017	0.0035	0.0016
Bias	0.0001	8.5E-5	9.9E-5	0.0001	9.9E-5	0.0001	0.0009	0.0002

Table 7: As in Table 5 but for $H = 0.7$.

Wavelets	Symmlet 8		Daubechies 6		Coiflet 6		Haar	
	ND	Ortho	ND	Ortho	ND	Ortho	ND	Ortho
Mean	0.7212	0.727	0.7279	0.6688	0.7237	0.7267	0.7256	0.7267
MSE	0.0026	0.0065	0.0028	0.0045	0.0026	0.0058	0.0026	0.0058
Variance	0.0022	0.0058	0.002	0.0035	0.002	0.0051	0.002	0.0051
Bias	0.0004	0.0007	0.0008	0.001	0.0006	0.0007	0.0006	0.0007

Figure 26: Boxplots of $\hat{H}$ from 300 simulations of 2-D fBm's when $H = 0.3$ with various wavelet bases and non-decimated and orthogonal transform.

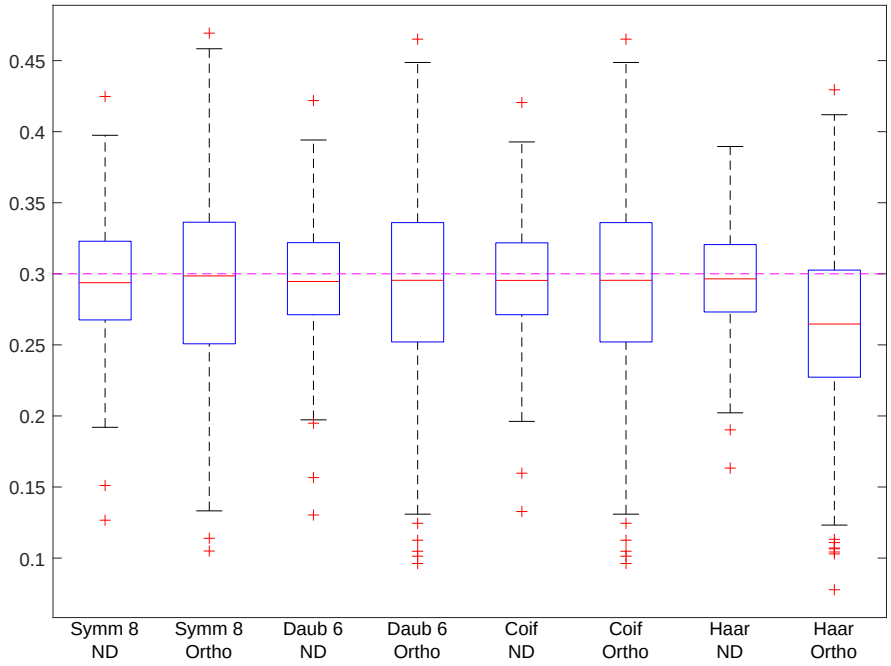


Figure 27: As in Figure 26, but for $H = 0.5$.

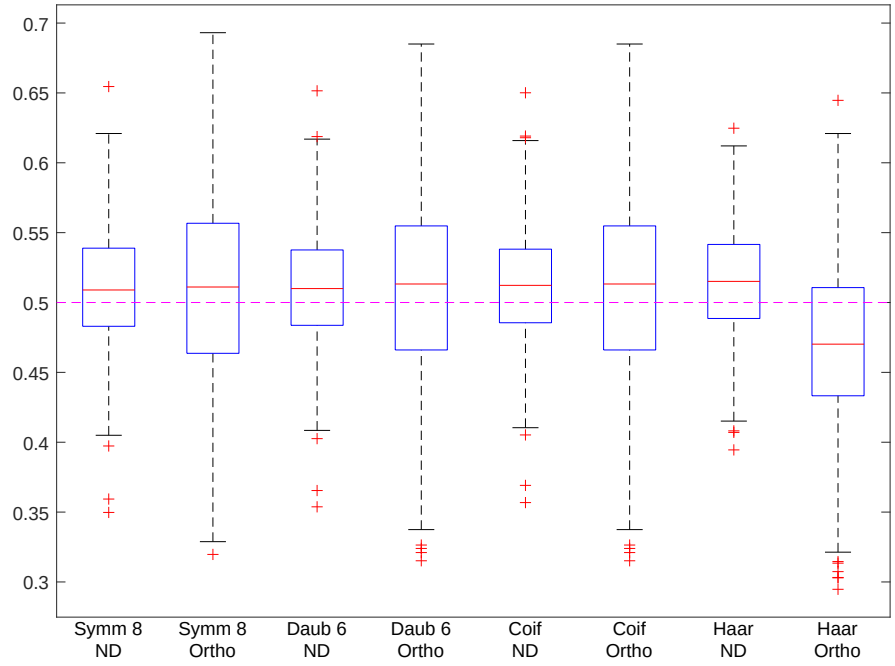
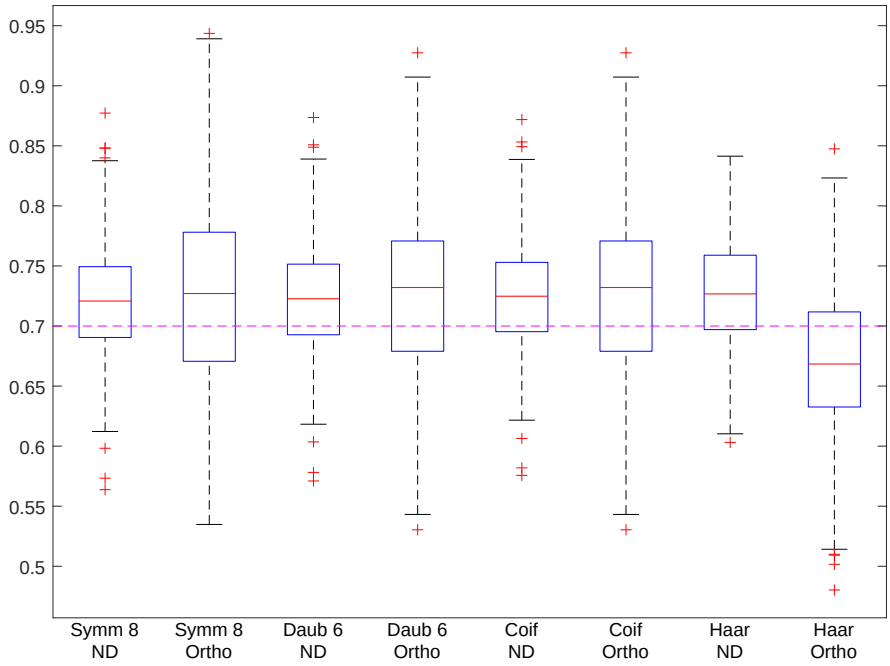


Figure 28: As in Figure 26, but for $H = 0.7$.



negative effect of increased levelwise dependence among the coefficients.

3.2 Application in Mammogram Diagnostic

We apply the 2-D scale-mixing NDWT-based method to digital mammograms with the goal of identifying wavelet features suggestive of breast cancer.

3.2.1 Source of Data

We obtain the mammographic images from Digital Database for Screening Mammography (DDSM) at the University of South Florida Bowyer et al. (1996). All cases examined had biopsy results which served as ground truth. University researchers used the HOWTEK scanner at the full 43.5-micron per pixel spatial resolution to scan 45 mammograms from patients with normal studies and 79 from patients with confirmed breast cancer. Each case contains two images of each breast in cranio-caudal (CC) and mediolateral oblique (MLO) projections. From these images, we

obtain 124 CC projection images and identify background tissue area of a breast in each mammographic image so that we can extract wavelet features that indicate the health conditions of patients based on the identified tissue areas without affects of background non-breast area. Figure 29 shows the mammogram and the mask image, which is a binary image that takes value 1 if the location belongs to a breast tissue area or 0 otherwise. In a subsequent classification process, we use the mask image to filter out numerical values (i.e., wavelet coefficients) from NDWT that are irrelevant to defining self-similar properties of breast tissues.

3.2.2 Diagnostic Classification

For breast cancer diagnostics, we performed 2-D scale-mixing NDWT of depth 6 for each mammogram. The mammograms had various sizes and the location of a breast tissue area within a mammogram also varied. As we were interested exclusively in the scaling characteristics of the breast tissue, we first identified the wavelet coefficients, which belong to the breast tissue area using a two-step process. We began by orienting all mammograms from left to right, so that a breast tissue area starts at the right-hand side of mammogram. Then, for each row, we defined the boundary of the breast tissue area. This was done by an algorithm that selected the left-most-pixel for which the average intensity of the 64 subsequent pixels decreased below a predefined threshold λ. We averaged a sequence of pixel intensities so that noisy fluctuations among pixel intensities in a row are smoothed, to prevent multiple boundary points in a single row of pixels. Once the boundary point in each row of pixels was identified, we adjust for the "cone-of-influence" effect typical for all wavelet transforms. The cone-of-influence effect refers to a blurring effect of wavelet filters when applied in a sequential manner, like in Mallat's algorithm. Because of this blurring effect, a local feature of a signal propagates along the multiresolution spaces in a shape of a cone. The longer the wavelet filter, the wider the cone. To eliminate this effect, that is,

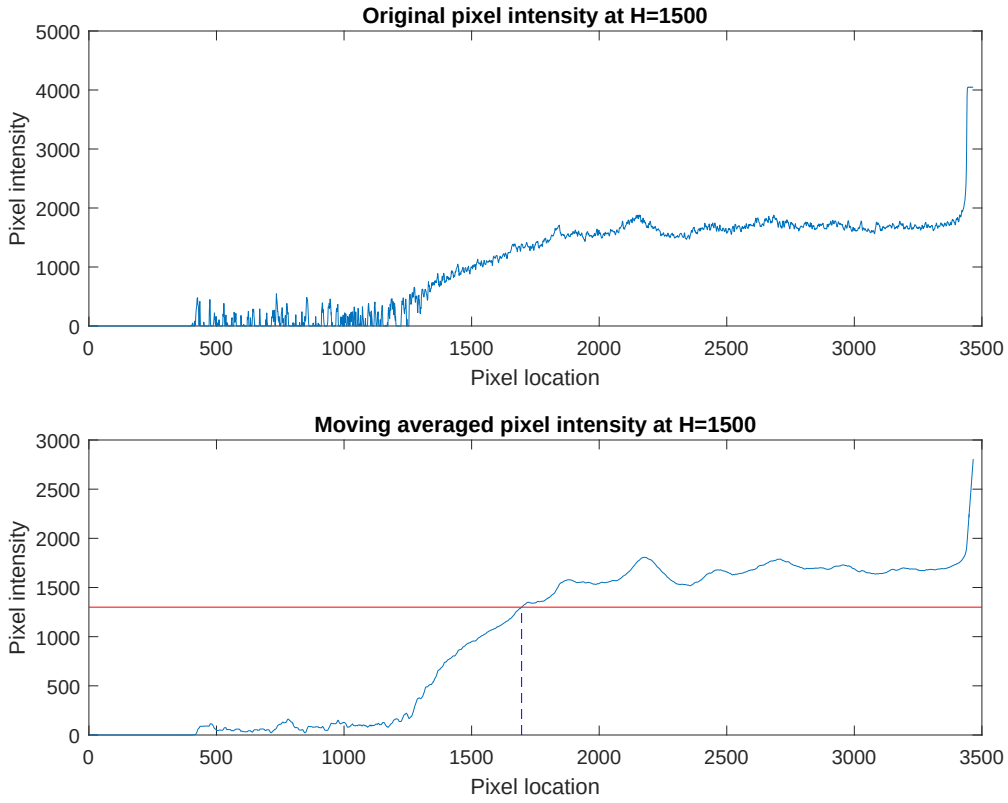


Figure 29: The top panel depicts an exemplary row of pixel intensities from an original image that exhibits fluctuations. The bottom panel shows the pixel intensity smoothed by a moving average, from which we are able to identify a single boundary pixel.

to eliminate influence of non-tissue pixels on the local wavelet spectra, we shift m pixel-locations to the right from the original boundary to form an updated boundary.

To emphasize locality, we used Haar wavelet, which produces the most narrow cone. The maximum length of a Haar wavelet filter convolved over the 6-level non-decimated decomposition is bounded by 2^6 which is an approximation to the maximal width of the cone. For comparisons and robustness assessment, in further analysis we selected three m values, 0, 2^6, and 2^7.

Based on the boundary construction rules for each row, we form a 0-1 image (mask) so that we can select wavelet coefficients corresponding to the ROI through a simple multiplication.

An entry at row i and column j of a binary image (mask) $\boldsymbol{\mu}$ is

$$\mu_{ij} = \begin{cases} 1, & \frac{1}{m}\sum_{k=j+1}^{j+m} I_{ik} > \lambda \quad \text{and} \quad j \geq m+1 \\ 0, & \text{else} \end{cases}$$

where I_{ik} indicates a pixel intensity in a mammogram at position (i,k), and m the length of a shift. The mask $\boldsymbol{\mu}$ has the same size as the mammogram. The entry of $\boldsymbol{\mu}$ at location (i,j) is an indicator that is 1 if pixel (i,j) belongs to a breast tissue region and is unaffected by the cone of influence, and 0 otherwise. We perform Hadamard (element-by-element) multiplication of a mask image and wavelet coefficients at each resolution level. With such multiplication, only wavelet coefficients that belong to the breast tissue region at each level are selected. From those selected wavelet coefficients, we find five descriptors: a scaling measure and four asymmetry measures, as in Panel (b) of Figure 13.

As we discussed, the asymmetry measure, compares horizontal vs vertical isotropy of in breast tissue.

The scaling descriptor, is calculated by the equation (51) from coefficients in the main diagonal hierarchy, $\overline{|d_{(J-5,J-5)}|^2}$, $\overline{|d_{(J-4,J-4)}|^2}$, $\overline{|d_{(J-3,J-3)}|^2}$, $\overline{|d_{(J-2,J-2)}|^2}$, and $\overline{|d_{(J-1,J-1)}|^2}$. The choice of diagonal hierarchy provides the most information about the regularity of breast tissue Nicolis et al. (2011).

The asymmetry measures are the energy ratios of two adjacent levels to the main diagonal hierarchy (j,j), indexed by $(j,j-1)$ and $(j-1,j)$ for $j = J-4, J-3, J-2$, and $J-1$. For example, at scale $j = J-3$, an asymmetry measure is defined as $\overline{|d_{(J-3,J-4)}|^2}/\overline{|d_{(J-4,J-3)}|^2}$.

We found that the five scales of finest detail were most effective in classification of the health conditions of patients because disease signatures are mostly captured in subtle variations within the tissue area. Note that the energy at each level is calculated with only the wavelet coefficients located in the breast tissue area.

We use the obtained five features of all mammograms as inputs to three classifiers:

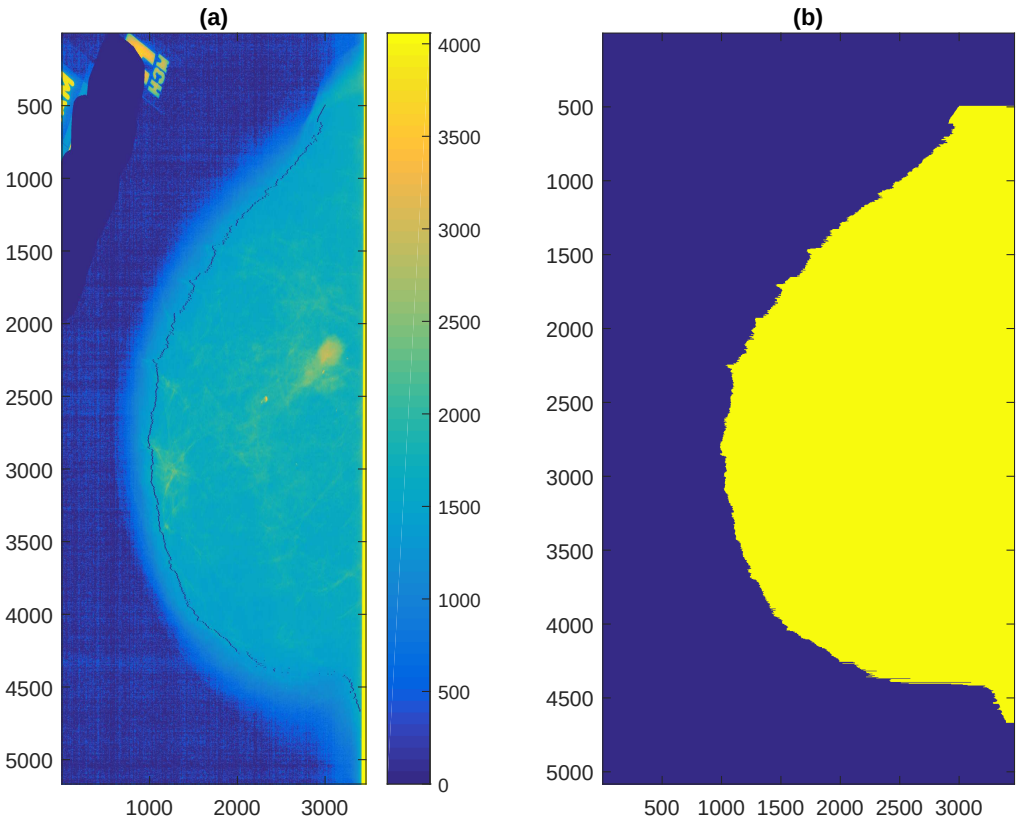


Figure 30: An original mammogram (a) and a binary mask (b) indicating the domain of wavelet coefficients used in the analysis. The black line in the original image represents the boundary detected by the algorithm. Panel (b) shows the mask image in which white corresponds to 1 and black to 0.

logistic regression, support vector machine, and random forest algorithm. In each iteration, we use four-fold cross validation, which randomly divides the data into four sets and then uses three sets as training data and the remaining set as test data. We repeat this random division of training and testing data sets 200 times, and the report prediction accuracies averaged over these 200 repetitions.

We present and compare the performance of classification in terms of sensitivity, specificity, and the overall classification accuracy, which are shown in Tables 8, 9, and 10.

Table 8: Sensitivity with three classifiers. All algorithms show strong diagnostic power in identifying cancerous mammograms.

m value	Logistic regression	SVM	Random forest
0	0.7354	0.6811	0.8511
2^6	0.7692	0.7104	0.8721
2^7	0.7703	0.7213	0.8739

Table 9: Specificity with three classifiers.

m value	Logistic regression	SVM	Random forest
0	0.6293	0.585	0.585
2^6	0.6642	0.5865	0.5865
2^7	0.6572	0.5954	0.5954

Table 10: Classification accuracy with three classifiers. Random forest algorithm shows the best diagnostic accuracy exceeding 80%.

m value	Logistic regression	SVM	Random forest
0	0.692	0.6474	0.7975
2^6	0.7264	0.6655	0.8272
2^7	0.7256	0.6753	0.8335

3.3 Conclusions

Most existing computer aided breast cancer detection methods focus on identifying markers of breast cancer in specific regions. The diagnostic use of information contained in the background tissue is often ignored modality. This chapter relates the

degree of self-similarity and anisotropy of patterns in breast tissue areas of a mammogram to the presence of breast cancer. We develop a 2-D scale-mixing NDWT based method that estimates the degree of scaling behavior and anisotropy of breast background tissue. We first assess the scaling estimation performance of the proposed method in simulated cases with 2-D fBm's. In the simulations, the proposed method yields, on average, scaling estimators closer to the target values and with lower mean square errors. Then, we apply the NDWT method to publicly available mammographic images from University of South Florida Bowyer et al. (1996) for the detection of breast cancer. The selected classifiers use five descriptors: one self-similarity measure and four asymmetry measures. Computation of those descriptors benefited from two distinctive characteristics of non-decimated wavelet transforms. First, the redundancy of transform produced estimators with smaller variance without inducing additional bias, and the second, the spatial invariance of the transform enabled calculation of local spectra so that coefficients not corresponding to breast tissue were excluded from the analysis. With the five descriptors described in this chapter, we achieved an average diagnostic accuracy in excess of 80%.

One of the valid criticisms for the clinical use of this methodology is that the accuracy rate is not high enough. Indeed, this would be the case if the proposed method is to be used by itself. However, even the classifiers, "slightly better than flipping a coin," can improve accuracy when added to a battery of other independent testing modalities. In this respect, our findings provide an opportunity for significant improvement of existing mammogram classification procedures and can assist the radiologist.

Chapter IV

BAYESIAN APPROACH TO ESTIMATION OF SCALING WITH APPLICATIONS

A number of phenomena in various fields such as geology, atmospheric sciences, economics, to list a few, can be modeled as a fractional Brownian motion indexed by Hurst exponent H. This exponent is related to the degree of regularity and self-similarity present in the signal, and it often captures important characteristics useful in various applications. Given its importance, a number of methods have been developed for the estimation of the Hurst exponent. Typically, the proposed methods do not utilize prior information about scaling of a signal.

Several examples in which the Hurst exponent is well localized are as follows. For locally isotropic and fully developed turbulences, Kolmogorov introduced K41 theory. Following his theory, the Hurst exponent H of turbulence processes is 1/3. For physical particles, the asymptotic behavior of some Brownian motions that interact through collisions on a real line converges to an fBm with Hurst exponent $H = 1/4$ Nourdin et al. (2009); Peligrad and Sethuraman (2007); Swanson (2011). In a study of DNA sequences, Arneodo et al. mapped nucleotide sequences onto a "DNA walk" and determined that non-coding regions can be well modeled by a fractional Brownian motions with a Hurst exponent close to 0.6 Arneodo et al. (1996). For atmospheric turbulence, wave fronts become fractal surfaces behaving as an fBm with Hurst parameter $H = 5/6$ once they are degraded by turbulence Schwartz et al. (1994); Ribak (1997); Pérez et al. (2004). In addition, other refined models for turbulence yield various Hurst exponent values different from 1/3, but instead, a value that can be estimated by the local power law Nelkin (1975); Biskamp (1994);

Horbury et al. (2008). Such real-life phenomena are just a few examples in which we have prior information about the Hurst exponent prior to observing the data.

Thus, we develop a Bayesian scaling estimation method with non-decimated wavelet transform (NDWT) motivated by real-life signals that are known to possess a certain theoretical degree of self-similarity. Bayesian approaches have been previously employed in this context. The Hurst exponent for Gaussian data was estimated with a Bayesian model in Makarava et al. (2011); Benmehdi et al. (2011); Conti et al. (2004). Holan et al. Holan et al. (2009) developed a hierarchical Bayesian model to estimate the parameter of stationary long-memory processes. A Baysian model for the parameter estimation of auto-regressive fractionally integrated moving average (ARFIMA) processes Hosking (1981) are discussed in Graves et al. (2015); Ravishanker and Ray (1997); Pai and Ravishanker (1998). These models are based on time domain data. However, the de-correlation property of wavelet transforms facilitates a simplified model construction, and multiple wavelet-based Bayesian techniques has been developed. Based on a Bayesian approach, Vannucci and Corradi Vannucci and Corradi (1999) estimated parameters for long memory process with a recursive algorithm and Markov chain Monte Carlo (MCMC) sampling. A Baysian wavelet model for ARFIMA processes is illustrated in Ko and Vannucci (2006).

In this chapter, we estimate Hurst exponent of a fractional Brownian motion (fBm) with wavelet coefficients from non-decimated wavelet transform (NDWT) and a Bayesian approach that incorporates information about the theoretical value of Hurst exponent via the location of a prior distribution. We combine the likelihood function and the prior distribution on $(H,\ \sigma^2)$ to obtain non-normalized posterior distribution. Because we want to estimate the most likely H value of an input signal given prior information and wavelet coefficients, we calculate $\hat{H}$, which maximizes the non-normalized posterior distribution. This is equivalent to estimating the mode of the posterior distribution, also referred to as maximum a posteriori (MAP) estimation.

In addition, MAP estimation method results in an optimization problem that can be solved in various ways and yields an estimator optimal under a zero-one loss function. We apply the proposed method to simulated signals for the estimation of Hurst exponent H based on prior distributions with approximately correct mean values. The results indicate that averaged mean squared error (MSE) of estimators significantly decreases with a prior distribution with a mean that matches the value of a true Hurst exponent. Moreover, when a slightly biased mean value of a prior distribution is provided, the averaged mean squared errors of the estimators from the proposed method are still lower than those from the regression-based method.

4.1 Method

We applied a Bayesian model to wavelet coefficients in the domain of non-decimated wavelet transforms (NDWT). In multiresolution analysis of a m-dimensional fBm $B_H(\mathbf{t})$ with Hurst exponent H, a coefficient d_j from multiresolution subspace at level j, is related to a coefficient d_0 from a subspace at level 0, as Flandrin (1992)

$$d_j \stackrel{d}{=} 2^{-(H+m/2)j} d_0,\ d_0 \sim N(0, \sigma^2).$$

As wavelet coefficients at each multiresolution subspace follow a normal distribution with mean zeros and common variance, an average of the squared wavelet coefficients, under the assumption of independence, follows a chi-square distribution. The number of degrees of this distribution is equal to the size of the original data. Based on such properties, we establish the following lemma:

Lemma 4.1.1. Let y_j be the average of squared wavelet coefficients, $\overline{d_j^2}$, in a wavelet subspace at level j. Then the distribution of y_j is

$$g(y_j) = \left(\frac{1}{\Gamma(2^{mJ-1})}\right)\left(\frac{2^{(2H+m)j+mJ}}{2\sigma^2}\right)^{2^{mJ-1}} \quad (y_j)^{2^{mJ-1}-1} \exp\Big(-\tfrac{2^{mJ}}{2\sigma^2} y_j 2^{(2H+m)j}\Big),$$

where m is the dimension of the signal, H is the Hurst exponent, J is an integer part of $\log_2 n$, and n is the size of the input signal.

The likelihood function of (H, σ^2) conditional on observations of averaged energies from levels $j_1, \ldots, j_2$ is

$$\mathcal{L}(H, \sigma^2 | y_{j_1}, \ldots, y_{j_2}) = \prod_{i=j_1}^{j_2} g(y_i).$$

We use beta distribution and non-informative prior $1/\sigma^2$ as independent priors on H and σ^2, respectively,

$$\pi(H, \sigma^2) = \frac{H^{\alpha-1}(1-H)^{\beta-1}}{\mathcal{B}(\alpha, \beta)} \times \frac{1}{\sigma^2}.$$

The hyperparameters in beta distribution, α and β are calibrated by considering the impact of effective sample size (ESS) and the mean of the beta distribution, $\frac{\alpha}{\alpha+\beta}$, which is linked to the Hurst exponent of an input signal. The ESS for the beta(α, β) prior is approximated with $\alpha + \beta$ and is closely related to the performance of the Bayesian estimation. For example, when ESS is large, the posterior distribution is dominated by the prior Morita et al. (2008). Based on simulations, we selected the ESS to be approximately 50% the original data size, but the ESS can be calibrated based on the level of certainty about H. The larger the ESS is, the more confident we are about the mean of a prior, that is, in the "true" value of H.

Theorem 4.1.1. The maximum a posteriori (MAP) estimator of H is a solution to the following non-linear system:

$$\begin{cases} \frac{\partial \pi(H,\sigma^2|y_{j_1},\ldots,y_{j_2})}{\partial \sigma^2} = -\left(\frac{bc+2}{2}\right)\frac{1}{\sigma^2} + \frac{b}{2\sigma^4}\sum_{j=j_1}^{j_2} y_j 2^{(2H+m)j} = 0 \\ \\ \frac{\partial \pi(H,\sigma^2|y_{j_1},\ldots,y_{j_2})}{\partial H} = \frac{\alpha-1}{H} - \frac{\beta-1}{1-H} + b\ln\ln 2 \sum_{j=j_1}^{j_2} j - \frac{\ln 2 \sum_{j=j_1}^{j_2} y_j j 2^{(2H+m)j}}{\sum_{j=j_1}^{j_2} y_j 2^{(2H+m)j}}(bc+2) = 0 \end{cases} \tag{58}$$

Details of derivation and solution of (58) are deferred to Appendix. As the closed form solution that satisfies the non-linear system (58) is not available and given that the value of H ranges only from 0 to 1, we approximately solve the equations by inserting sequentially increasing H from 0 to 1 with increments of 10^{-7}.

4.2 Simulations

In this section, we compare the estimation performance of the proposed method to that of non-decimated wavelet transform-based method that uses no prior information on H and estimates scaling by regression, which is standardly performed. The mean, variance, mean squared error, and squared bias are reported. We simulated three sets of two hundred one-dimensional (1-D) fractional Brownian motions (fBm's) of size 2^{11} with Hurst exponents 0.3, 0.5, and 0.7 each. Next, we estimated the Hurst exponent of each signal using the proposed method and the traditional regression-based method. We perform an NDWT of depth 8 using Haar wavelet and analyze resulting wavelet coefficients on the 4^{th}, 5^{th}, and 6^{th} levels, noting that resolution increases with the level index and that the finest level of detail is 10. The prior distribution for H is the beta with specified hyperparameters. For each set, we use three sets of prior hyperparameter settings. The prior means are taken the same as the real (used for simulation) value, and 0.05 higher or lower than the real value, so that the effect of prior robustness can be observed. The parameters of different prior distribution settings are in Table 11.

Tables 12-14 summarize the estimation results in terms of mean, variance, MSE, and

Table 11: Setting of the parameters in the simulation study. Prior mean is μ and (α,β) are parameters for beta prior.

μ	0.25	0.3	0.35	0.45	0.5	0.55	0.65	0.7	0.75
α	256	307.2	358.4	460.8	512	563.2	665.6	716.8	768
β	768	716.8	665.6	563.2	512	460.8	358.4	307.2	256

squared bias. Figure 31 shows the estimation results as box-and-whisker plots. The proposed method yields estimators with lower MSE compared to the regression-based method under various prior settings. The estimation performance is robust to slight deviations in parameters of the prior. Even if the mean of a prior differs from the

Table 12: Estimation performance comparison under various prior settings with simulated 200 1-D fBm's of size 2^{11} when Hurst exponent $H = 0.3$.

	Prior mean			Regression
	0.25	0.3	0.35	
Mean	0.2756	0.3043	0.3316	0.3100
Variance	0.0013	0.0013	0.0013	0.0068
MSE	0.0018	0.0013	0.0023	0.0068
Squared bias	0.0006	1.45E-5	0.0010	1.71E-5

Table 13: As in Table 12, but for $H = 0.5$.

	Prior mean			Regression
	0.45	0.5	0.55	
Mean	0.4669	0.4922	0.5176	0.4863
Variance	0.0010	0.0010	0.0010	0.0043
MSE	0.0023	0.0011	0.0012	0.0047
Squared bias	0.0013	0.0001	0.0002	0.0004

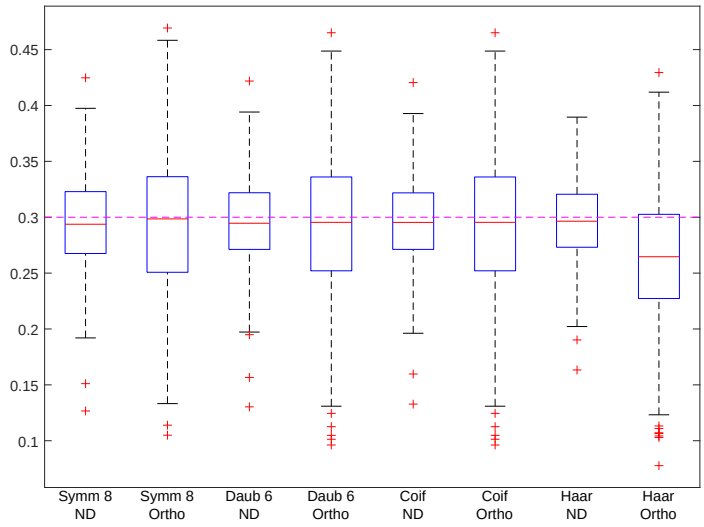
Table 14: As in Table 12, but for $H = 0.7$.

	Prior mean			Regression
	0.65	0.7	0.75	
Mean	0.6280	0.6561	0.6858	0.5502
Variance	0.0014	0.0014	0.0015	0.0062
MSE	0.0059	0.0029	0.0015	0.0255
Squared bias	0.0045	0.0015	0.0001	0.0193

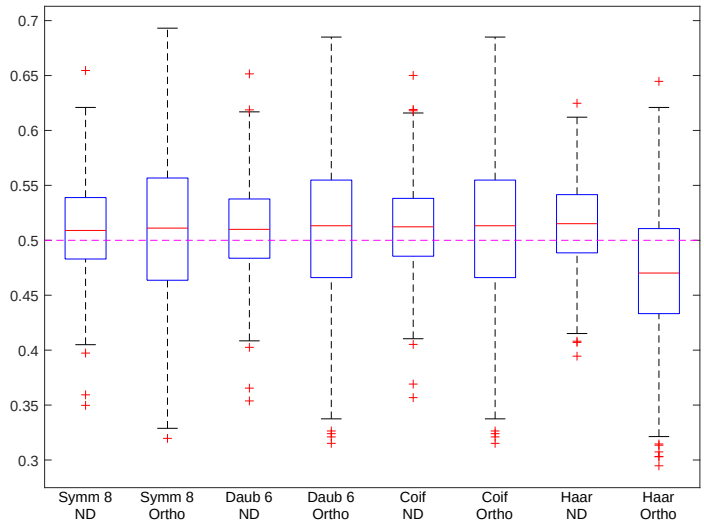
value of a true Hurst exponent, estimation performance is better than the regression-based method. A correct prior mean settings significantly enhance the estimation performance. We noticed, that due to autocorrelations among the NDWT wavelet coefficients, regression-based scaling estimation suffers from bias for Hurst exponents exceeding 1/2. Such bias is substantially alleviated by the proposed method.

4.3 An Application

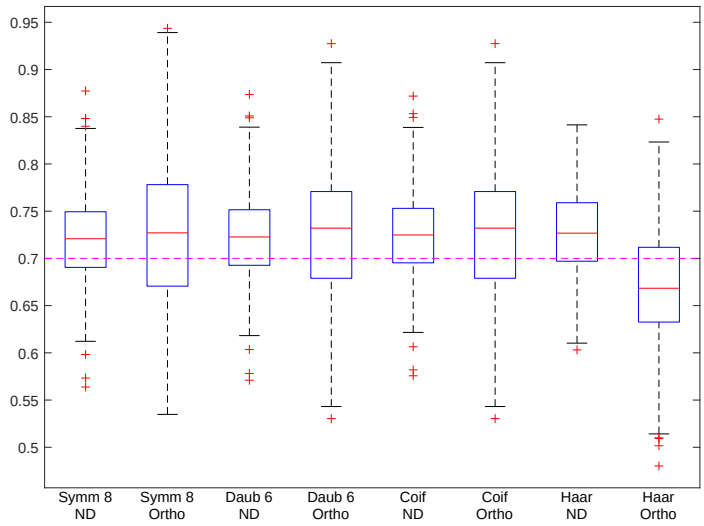
As an example with a real-life measurements that scale, we apply the proposed method to a dataset that traces the velocity components of turbulence. Measurements are taken with sampling frequency (f_s) of 56 Hz and period (T_p) of 19.5 minutes at Duke Forrest (Durham, NC) on July 12, 1995. The data set was from a triaxial sonic



(a) $H = 0.3$

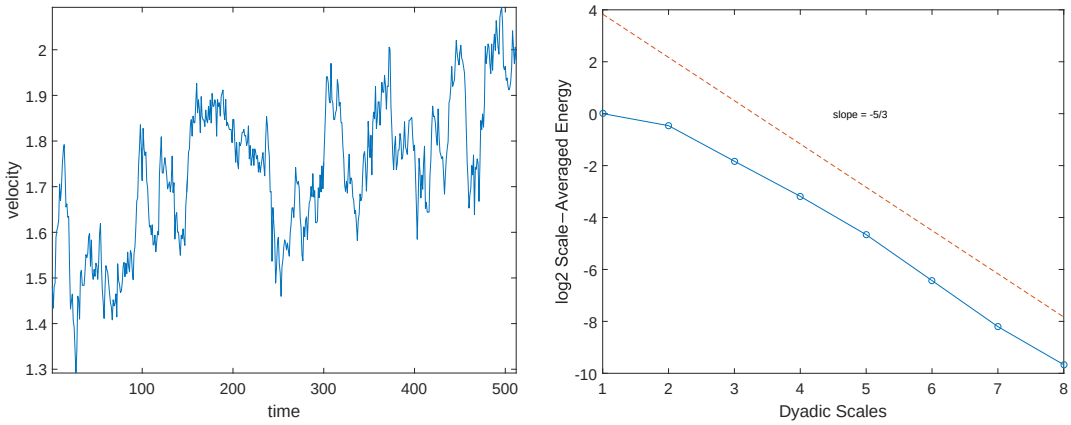


(b) $H = 0.5$



(c) $H = 0.7$

Figure 31: Estimation results of simulated 200 1-D fBm's with size 2^{11} when Hurst exponent is 0.3, 0.5, and 0.7 under various prior settings.



(a) U velocity component of turbulence (b) Scaling behavior in the wavelet domain

Figure 32: U velocity component of turbulence in time and wavelet domains.

anemometer (Gill Instruments/1012R2) mounted on a mast 5.2 m above the ground surface over an *Alta Fescue* grass site. We select the U component of the velocity with size 2^9 and use it to compare the estimators from the proposed and regression-based methods. Based on Kolmogorov's **K41** theory, we know that measurements of velocity components should have Hurst exponent close to $H = 1/3$. Therefore, for the proposed method, we set the prior distribution to be the beta distribution with parameters, $\alpha = 85.3$ and $\beta = 170.7$, which is apriori centered at $1/3$. We perform NDWT of depth 8 on the input signal and use wavelet coefficients from the eighth to the fifth level for calculations in both methods. We obtain $\hat{H} = 0.341$ with the regression-based method while $\hat{H} = 0.335$ with the proposed method. Figure 32 depicts the input turbulence signal in time domain and its wavelet spectrum by an NDWT.

4.4 Conclusions

A theoretical value of Hurst exponent H is available for some signals, but standard scaling estimation methods do not utilize such information. We proposed a Bayesian scaling estimation method that incorporates theoretical scaling information via a prior distribution and estimates H with a MAP principle. The proposed method

yields lower mean squared errors in simulations, and such performance was robust to small misspecification in the prior location. The method applied to a turbulence velocity signal yields an estimator of H close to the theoretical value.

Chapter V

METHODS FOR ASSESSMENT OF SCALING BY MEDIANS OF LOG-SQUARED NONDECIMATED WAVELET COEFFICIENTS:MEDL AND MEDLA

At first glance, data that scale look like noisy observations, and often the large-scale features (e.g. basic descriptive statistics, trends, smoothed functional estimates, and so on) carry no useful information. For example, the pupil diameter in humans fluctuates at a high frequency (hundreds of Hz), and prolonged monitoring of this yields massive data sets. Researchers found that the dynamics of change in the diameter are informative of eye pathologies, e.g., macular degeneration (Moloney et al., 2006). Yet, the trends and traditional summaries of the data are clinically irrelevant because the magnitude of the diameter depends on ambient light and not on the inherent eye pathology.

Formally, a deterministic function $f(\mathbf{t})$ of a d-dimensional argument $\mathbf{t}$ is said to be self-similar if $f(\lambda\mathbf{t}) = \lambda^{-H} f(\lambda\mathbf{t})$, for some choice of Hurst exponent H, and for all dilation factors λ. The notion of self-similarity has been extended to random processes. Specifically, a stochastic process $\{X(\mathbf{t}),\ \mathbf{t} \in R^d\}$ is self-similar with scaling exponent (or *Hurst exponent*) H if for any $\lambda \in R^+$,

$$X(\lambda\mathbf{t}) \stackrel{d}{=} \lambda^H X(\mathbf{t}), \tag{59}$$

where the relation "$\stackrel{d}{=}$" is the equality in all finite dimensional distributions.

In this chapter, we are concerned with the precise estimation of scaling exponent H in one-dimensional setting. A number of estimation methods for H exist: re-scaled

range calculation (R/S), Fourier-spectra methods, variance plots, quadratic variations, zero-level crossings, and so on. For a comprehensive description, please refer to Beran (1994), Doukhan et al. (2003), and Abry et al. (2013). Wavelet transforms are especially suitable for modeling self-similar phenomena, as is reflected by vibrant researches. An overview is provided in Abry et al. (2000a).

If processes possess a stochastic structure (e.g. Gaussianity, stationary increments), the scaling exponent H becomes a parameter in a well-defined statistical model and can be estimated as such. Fractional Brownian motion (fBm) is important and well-known model for data that scale. Its importance follows from the fact that fBm is a unique Gaussian process with stationary increments that is self-similar in the sense of (59).

A fBm has a (pseudo)-spectrum of the form $S(\omega) \propto |\omega|^{-(2H+1)}$, and thus the log-magnitudes of detail coefficients at different resolutions in a wavelet decomposition exhibit a linear relationship. Leveraging on this linearity and using non-decimated wavelet domains constitute the staple of this chapter.

In p-level decomposition of an input signal of size m, NDWT yields $m \times (p+1)$ wavelet coefficients, while DWT yields m wavelet coefficients independent of p. Each decomposition level in NDWT contains the same number of coefficients as the size of the original signal. This multiplicity of coefficients contributes to the accuracy of estimators of H. However, given the redundancy of NDWT, reducing the bias induced by level-wise correlation among the coefficients becomes an important issue. The two estimators we propose are based on logarithm-first approach connecting Hurst exponent with a robust location and resampling techniques.

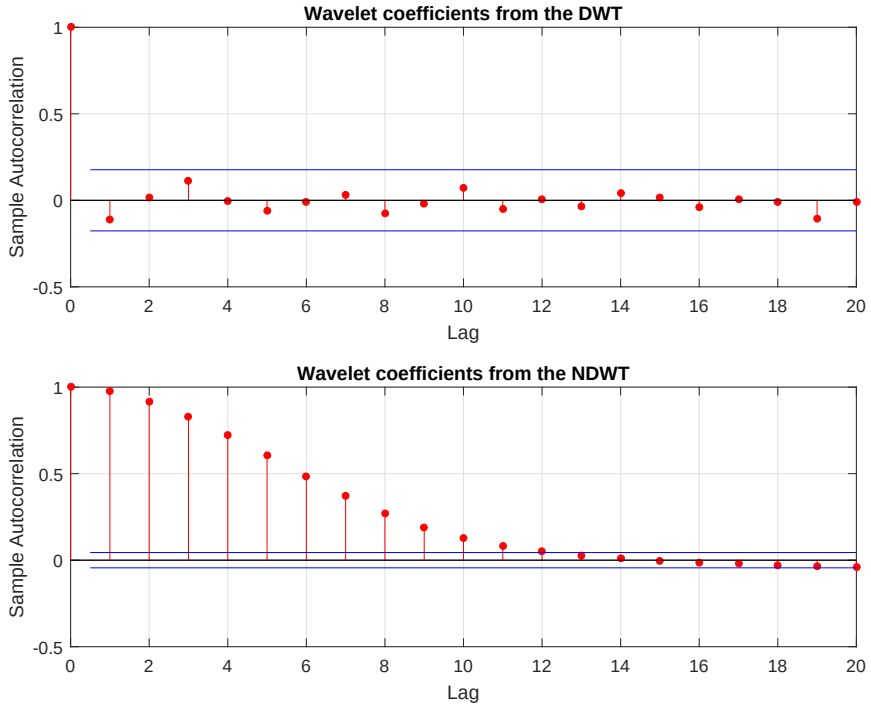


Figure 33: The autocorrelation present in wavelet coefficients from the DWT and the NDWT.

Figure 33 illustrates the autocorrelation within wavelet coefficients in the level $J-4$ (the level of finest detail is $J-1$, so $J-4$ is 4th "most detailed" level) in DWT and NDWT. Haar wavelet was used on a Brownian motion path of size 2^{11}. As we note, the coefficients from the NDWT are highly correlated while such correlation is not present among the DWT coefficients.

The two methods introduced in the following section reduce the effect of the correlation among wavelet coefficients, while maintaining redundancy and invariance as desirable threads of NDWT.

5.1 Method

We start by an overview of properties of wavelet coefficients and existing methods in literature based on which we develop the proposed methods.

Only detail wavelet coefficients are used for defining a wavelet spectrum, and subsequently, for estimating H. When an fBm with Hurst exponent H is mapped to the wavelet domain by DWT, resulting detail wavelet coefficients satisfy the following properties (Tewfik and Kim, 1992; Abry et al., 1995; Flandrin, 1992):

(i) d_j, a detail wavelet coefficient from level j, follows the Gaussian distribution with a mean of 0 and a variance of $\sigma_0^2 2^{-j(2H+1)}$, where σ_0^2 is a variance of a detail coefficient at level 0,

(ii) a sequence of wavelet coefficients from level j is stationary, and

(iii) the covariance between two coefficients from any level of detail decreases exponentially as the distance between them increases.

From the property (i), the relationship between detail wavelet coefficients and Hurst exponent H is

$$\log_2 \mathbb{E}\{d_j^2\} = -j(2H+1) + 2\log_2 \sigma_0.$$

Abry et al. (2000b) calculate a sample variance of wavelet coefficients to estimate $\mathbb{E}\{d_j^2\}$ assuming i.i.d. Gaussianity of coefficients on level j. Empirically, we look at the levelwise average of squared coefficients,

$$\overline{d_j^2} = \frac{1}{n_j} \sum_{i=1}^{n_j} d_{j,k}^2,$$

where n_j is the number of wavelet coefficients at level j. The relationship between $\overline{d_j^2}$ and H is

$$\log_2 \overline{d_j^2} \overset{d}{\approx} -(2H+1)j - \log_2 C - \log \chi_{n_j}^2 / \log 2,$$

where $\overset{d}{\approx}$ indicates approximate equality in distribution, $\chi_{n_j}^2$ follows a chi-square distribution with n_j degrees of freedom, and C is a constant. The method of Abry et al. (2000b) is affected by the non-normality of $\log_2 \overline{d_j^2}$ and correlation among detail

wavelet coefficients, which results in biases of weighted least squares estimates. To reduce the bias, Soltani et al. (2004) defined "mid-coefficients," as

$$D_{j,k} = \frac{d_{j,k}^2 + d_{j,k+n_j/2}^2}{2}.$$

According to this approach, each multiresolution level is split on two equal parts and corresponding coefficients from each part are paired, squared, and averaged. This produces a quasi-decorrelation effect. Soltani et al. (2004) show that level-wise averages of $\log_2 D_{j,k}$ are asymptotically normal with the mean $-(2H+1)j+C$, which is used to estimate H by regression.

The estimators in Soltani et al. (2004) consistently outperform the estimators in Abry et al. (2000b) under various settings. In addition, Shen et al. (2007) shows that the method of Soltani et al. (2004) yields more accurate estimators since it takes the logarithm of a mid-coefficient, and then averages. Thus, averaging logged wavelet coefficients, rather than taking logarithm of averaged wavelet coefficients, yields more precise estimators and this approach will be pursued in this chapter.

For both proposed methods, we first take the logarithm of a squared wavelet coefficient or an average of two squared wavelet coefficients, and then we derive the distribution of such logarithms under the assumption of independence. Next, we calculate the median of the derived distribution instead of the mean. The medians are more robust to potential outliers that can occur when logarithmic transform of a squared wavelet coefficient is taken and the magnitude of coefficient is close to zero. This numerical instability may increase the bias and variance of sample means. However, since the logarithms are monotone, the variability of the sample medians will not be affected.

The first proposed method is based on the relationship between the median of the logarithm of squared wavelet coefficients and the Hurst exponent. We use acronym "MEDL" to refer to this method. In MEDL, the logarithmic transform reduces the autocorrelation, while the number of coefficients remains the same. The second method derives the relationship between the median of the logarithm of average of two squared

wavelet coefficients and the Hurst exponent. We use acronym "MEDLA" to refer to this method. The MEDLA method is similar in concept to approach of Soltani et al. (2004) who paired and averaged wavelet coefficients prior to taking logarithm. Then the mean of logarithms was conected to H. Instead, we repeatedly sample with replacement N random pairs keeping distance between them at least q_j. Then, as in Soltani et al. (2004) we find the logarithm of pair's average and then connect the Hurst exponent with the median of the logarithms. As we relax the constraints on the distance between members of each pair, we obtain a larger amount of distinct samples and selecting only N samples out of such sample population further reduces the correlation.

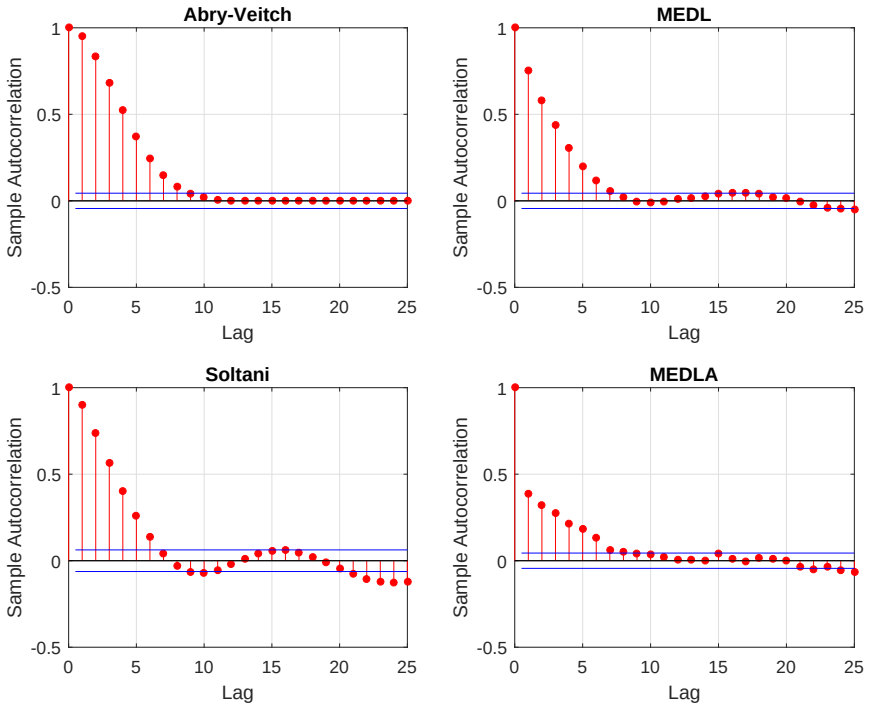


Figure 34: Autocorrelation of variables used in four methods.

To illustrate the decorrelation effects of the proposed methods, in Figure 34, we compare the autocorrelation present in variables that are averaged: means of d_{jk}^2 for traditional method, means of $\log_2 \left[\frac{d_{jk}^2+d_{jk+N/2}^2}{2}\right]$ for Soltani-like method, medians of $\log d_{jk}^2$ for MEDL, and medians of sampled $\log \left[\frac{d_{jk_1}^2+d_{jk_2}^2}{2}\right]$ for MEDLA method. The

two default methods exhibit higher amount of autocorrelation that decreases at a slower rate. MEDLA shows substantial reduction in correlation.

For formal distributional assessment of the two proposed methods, we start with an arbitrary wavelet coefficient from decomposition level j at location k, d_{jk}, resulting from a non-decimated wavelet transform of a one-dimensional fBm $B_H(\omega, t), t \in \mathbb{R}$,

$$d_j = \int_{\mathbb{R}} B_H(\omega, t)\psi_{jk}(t)dt, \text{for some fixed } k.$$

As Flandrin (1992) showed, the distribution of a single wavelet coefficient is

$$d_j \stackrel{d}{=} 2^{-(H+1/2)j}\sigma Z, \tag{60}$$

where Z follows a standard normal distribution, and σ^2 is the variance of wavelet coefficients at level 0. We will use (60) repeatedly for the derivations that follow.

5.1.1 MEDL Method

For the median of the logarithm of squared wavelet coefficients (MEDL) method, we derive the relationship between the median of the logarithm on an arbitrary squared wavelet coefficient from decomposition level j and Hurst exponent H. The following theorem serves as a basis for the MEDL estimator:

Theorem 5.1.1. Let y_j^* be the median of $\log d_j^2$, where d_j is an arbitrary wavelet coefficient from level j in a NDWT of an fBm with Hurst exponent H. Then, the population median is

$$y_j^* = -\log 2\,(2H+1)j + C, \tag{61}$$

where C is a constant independent of j. The Hurst exponent can be estimated as

$$\widehat{H} = -\frac{\widehat{\beta}}{2\log 2} - \frac{1}{2}, \tag{62}$$

where $\widehat{\beta}$ is the slope in least squares linear regression on pairs $(j, \hat{y}_j^*)$, and $\hat{y}_j^*$ is the sample median.

The proof of Theorem 5.1.1 is deferred to Appendix A.3. We estimate y_j^* by taking sample median of all logged squared wavelet coefficients at each level. The use of ordinary linear regression is justified by the fact that variances of the sample medians $\hat{y}_j^*$ are constant in j, that is,

Lemma 5.1.1. The variance of sample median $\hat{y}_j^*$ at level j is approximately

$$\frac{\pi e^Q}{2NQ}$$

where N is the sample size and $Q = (\Phi^{-1}(3/4))^2$.

The theorem is stating that the logarithm acts as a variance stabilizing operator; the variance of the sample median is independent of level j, and ordinary regression to find slope β in Theorem 5.1.1 is fully justified. Note that the use of ordinary regression simplifies approaches used in DWT, where the weighted regression is needed to account for levelwise heteroscedasticity.

The levelwise variance is approximately $5.4418/N$, independent of H and σ^2. The proof of Theorem 5.1.1 is deferred to Appendix A.3. In addition, we find the distribution of $\widehat{H}$:

Theorem 5.1.2. The MEDL estimator $\widehat{H}$ follows the asymptotic normal distribution

$$\widehat{H} \overset{approx}{\sim} \mathcal{N}\left(H, \frac{3A}{Nm(m^2-1)(\log 2)^2}\right),$$

where $A = \pi e^Q/(2Q) \cong 5.4418$, N is the sample size, and m is the number of levels used for the spectrum.

The proof of Theorem 5.1.2 is deferred to Appendix A.3. To verify Theorem 5.1.2, we perform an NDWT of depth 10 on simulated fBm's with $H = 0.3, 0.5$, and 0.7. We use resulting wavelet coefficients from levels $J-7$ to $J-2$ inclusive (i.e., six levels) to estimate H with MEDL. Following Theorem 5.1.2, $\widehat{H}$ of MEDL in the simulation follows a normal distribution with mean H and variance 7.9007×10^{-5}, which is illustrated in Figure 35.

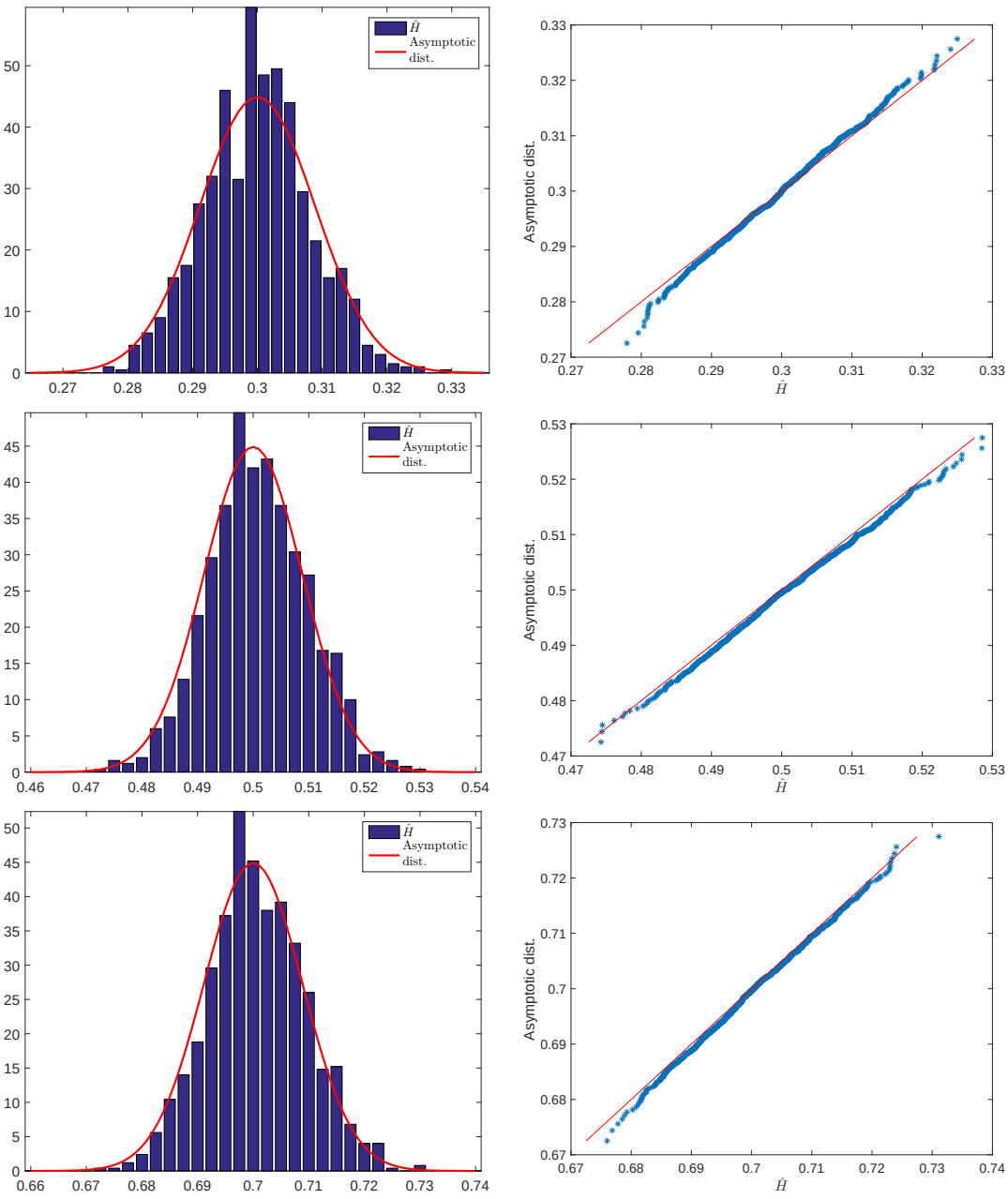


Figure 35: Panels on the right are histograms of $\widehat{H}$ and panels on the left are q-q plots of $\widehat{H}$ versus the quantiles asymptotic distribution when $H = 0.3, 0.5$, and 0.7, respectively.

5.1.2 MEDLA Method

For the median of the logarithm of averaged squared wavelet coefficients (MEDLA) method, we derive the relationship between logarithm of an average of two squared wavelet coefficients and Hurst exponent H. Soltani et al. (2004) proposed a method that quasi-decorrelates wavelet coefficients by dividing all wavelet coefficients on one level into left and right sections and pairing every coefficient in the left section with its counterpart in the right section, maintaining the equal distance to its partner (i.e., members in each pair are $N/2$ apart when N is the number of wavelet coefficients on that level). Then, Soltani et al. (2004) averaged every pair of squared wavelet coefficients and took logarithm from each average. We follow similar idea except that instead of fixing the combinations of pairs, which amounts to $N/2$ pairs in Soltani et al. (2004), we randomly sample with replacement N pairs whose members are at least q_j apart. Based on sample autocorrelation graphs, we define $q_j = 2^{J-j}$ that increase in each level j because the finer the subspace (i.e., larger j), the lower the correlation among wavelet coefficients. Then, we propose an estimator of H according to the following theorem.

Theorem 5.1.3. Let d_{jk_1} and d_{jk_2} be two wavelet coefficients from level j, at positions k_1 and k_2, respectively, from a NDWT of a fBm with Hurst exponent H. Assume that $|k_1 - k_2| > q_j$, where q_j is the minimum separation distance that depends on level j and selected wavelet base. Let y_j^* be the median of $\log\left[\frac{d_{jk_1}^2 + d_{jk_2}^2}{2}\right]$. Then, as in Theorem 5.1.1, results (61) and (62) hold.

The proof of Theorem 5.1.3 is deferred to Appendix A.4. To estimate y_j^*, we first repeatedly sample N pairs of wavelet coefficients with replacement from all pairs that are at least q_j apart. Then, we take logarithm of pair's average energy and take the median. As in Theorem 5.1.1, the variances of sample medians $\hat{y}_j^*$ are free of j.

Lemma 5.1.2. The variance of the sample median $\hat{y_j^*}$ at level j is approximated by

$$\frac{1}{N(\log 2)^2},$$

where N is the sample size.

The proof is straightforward and given in Appendix A.4. Thus, the variance of $\hat{y}_j^*$ is constant over levels. We find that MEDLA estimator of H indeed follows a normal distribution with a mean and a variance as the following.

Theorem 5.1.4. The estimator $\widehat{H}$ of MEDLA follows the asymptotic normal distribution

$$\widehat{H} \overset{approx}{\sim} \mathcal{N}\left(H, \frac{3}{Nm(m^2-1)(\log 2)^4}\right),$$

where N is the sample size, and m is the number of levels used for the spectrum.

The proof of Theorem 5.1.4 is deferred to Appendix A.4. To verify Theorem 5.1.4, we use the same wavelet coefficients from the simulation in section 5.1.1. Following Theorem 5.1.4, $\widehat{H}$ of MEDLA in the simulation follows a normal distribution with mean H and variance 7.9007×10^{-5}, which is verified in Figure 36.

5.2 Simulation

We simulate three sets of three hundred one-dimensional fractional Brownian motion (1-D fBm) signals of size 2^{11} with Hurst exponents 0.3, 0.5, and 0.7 for each set. Then, we perform an NDWT of depth 10 with a Haar wavelet on each simulated signal and obtain wavelet coefficients to which we applied the proposed methods for the estimation of Hurst exponent. For all methods and estimations, we used wavelet coefficients from levels $J-7$ to $J-2$ for the regression. We compare the estimation performance of the proposed methods to two standard methods: a method of Veitch and Abry (1999) and a method of Soltani et al. (2004), both in the context of NDWT. We present the estimation performance in terms of mean, variance, bias-squared, and

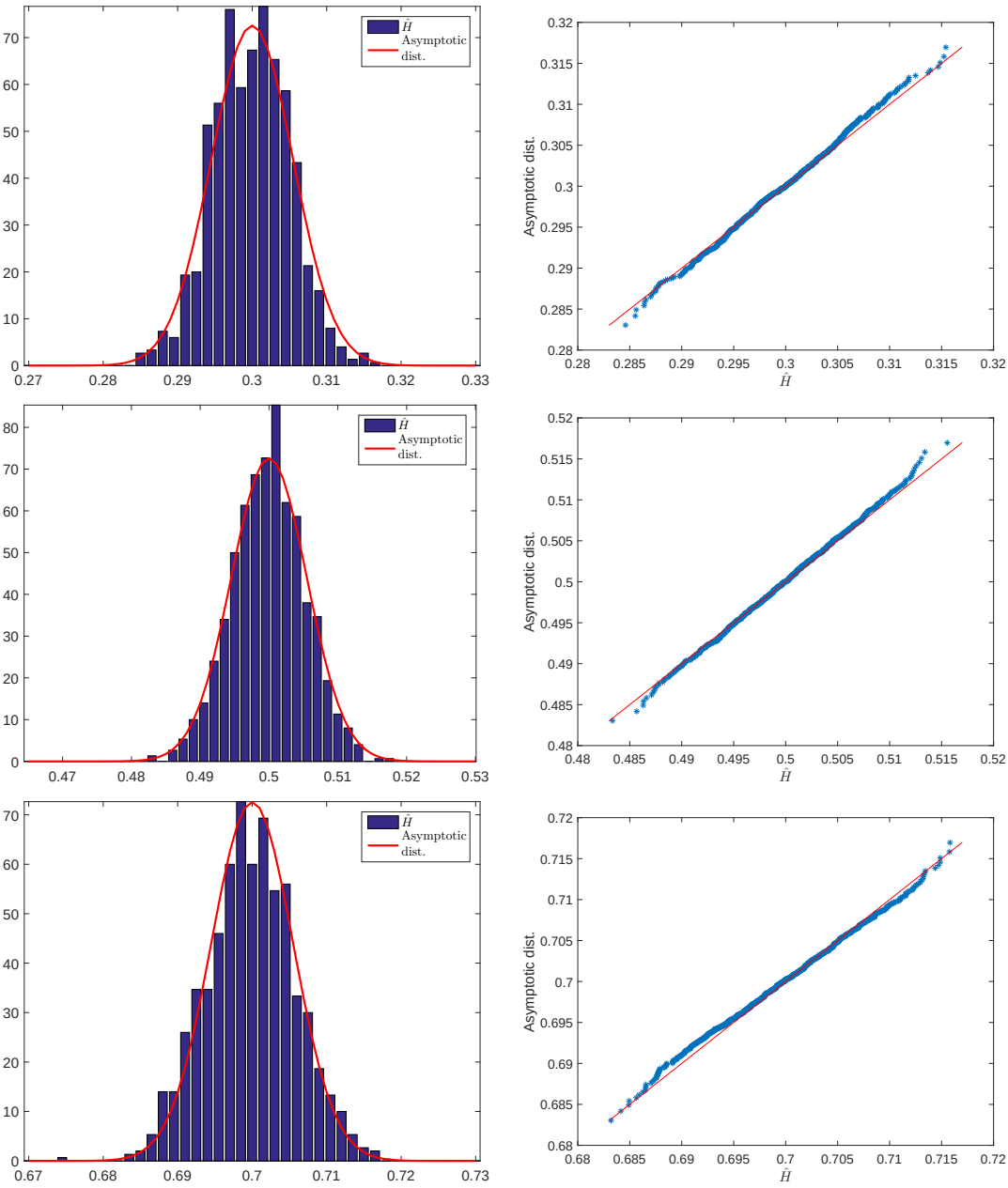


Figure 36: Panels on the right are histograms of $\widehat{H}$ and panels on the left are q-q plots of $\widehat{H}$ versus the quantiles of asymptotic distribution when $H = 0.3, 0.5$, and 0.7, respectively.

mean squared error, based on 300 simulations for each case. Table 15 and Figure 37 indicate that as H increases, the proposed methods outperform the standard methods. For smaller H, the estimation performance of all methods were comparable.

Table 15: Estimation of H with 300 simulated 1-D fBm signals of size 2^{11} when H=0.3, 0.5, and 0.7 by four methods

H=0.3

Method	Traditional	Soltani	MEDL	MEDLA
Mean	0.2864	0.2849	0.2778	0.2783
Variance	0.0017	0.0015	0.0021	0.0016
Bias-squared	0.0002	0.0003	0.0005	0.0005
MSE	0.0019	0.0018	0.0026	0.0021

H=0.5

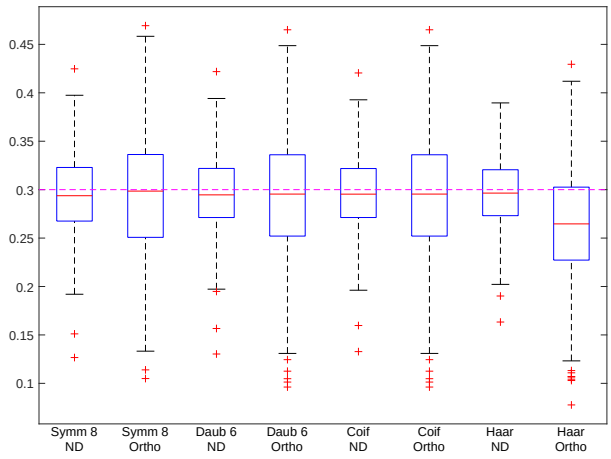
Method	Traditional	Soltani	MEDL	MEDLA
Mean	0.475	0.5091	0.4966	0.4982
Variance	0.0012	0.0022	0.0023	0.0017
Bias-squared	0.0006	6.7E-5	4.1E-6	1.3E-6
MSE	0.0018	0.0023	0.0023	0.0017

H=0.7

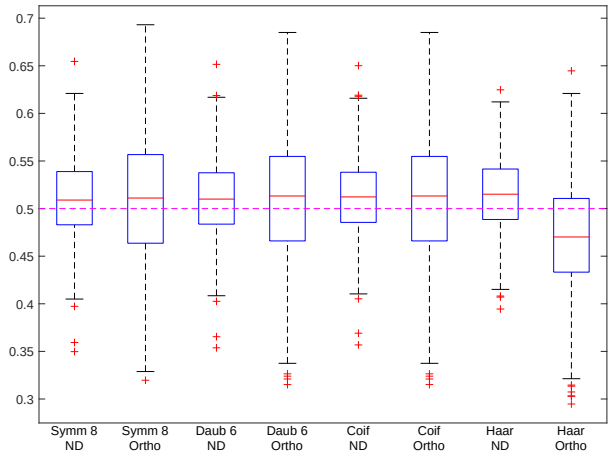
Method	Traditional	Soltani	MEDL	MEDLA
Mean	0.5524	0.7286	0.7065	0.7084
Variance	0.0039	0.0028	0.0033	0.0024
Bias-squared	0.0217	0.0008	3.3E-5	6.2E-5
MSE	0.0256	0.0036	0.0033	0.0024

5.3 Conclusions

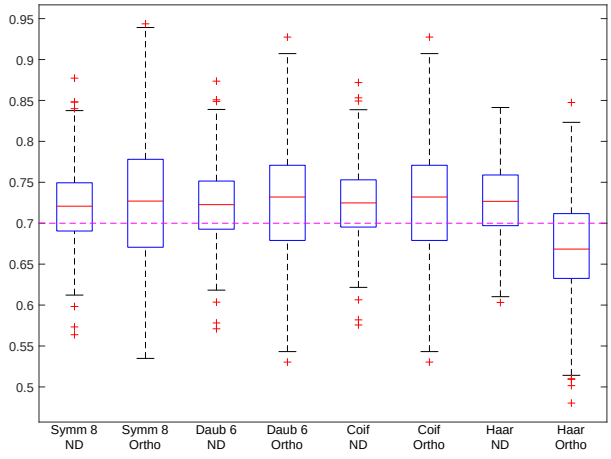
We proposed two methods for robust estimation of Hurst exponent in one-dimensional signals that scale. Unlike the standard methods, the proposed methods are based on NDWT. The motivation for using NDWT was its redundancy and time-invariance. However, the redundancy, which was useful for the stability of estimation, introduces autocorrelations among the wavelet coefficients. The proposed methods lower the present autocorrelation by (i) taking logarithm of the squared wavelet coefficients



(a) H=0.3



(b) H=0.5



(c) H=0.7

Figure 37: Boxplots of $\widehat{H}$ by four methods with 300 simulated 1-D fBm signals of size 2^{11} when H=0.3, 0.5, and 0.7

prior to averaging, (ii) relating the Hurst exponent to the median of the model distribution, rather than the mean, and (iii) resampling the coefficients.

The method is compared to a standard approach and gives estimators with smaller MSE for a range of input conditions.

Chapter VI

CONCLUSIONS

In this thesis, we introduced four novel methods that facilitate the scaling estimation based on NDWT. Chapter 2 introduced an NDWT matrix which is used to perform an NDWT in one or two dimensions. The use of matrix significantly decreased the computation time when 2-D inputs of moderate size are transformed under MATLAB environment, and such reduction of computation time was augmented when the same type of NDWT is performed repeatedly. With 2-D inputs, an NDWT matrix yielded a scale-mixing NDWT, which is more compressive than the standard 2-D NDWT. The retrieval of an original signal after the transform was possible with a weight matrix. An NDWT matrix can handle signals of non-dyadic sizes in one or two dimensions. The proposed NDWT matrix was used for the transforms in Chapters 3-5. Chapter 3 introduced a method for scaling estimation based on a non-decimated wavelet spectrum. A distinctive feature of NDWT, redundancy, enables us to obtain local spectra and improves the accuracy of scaling estimation. For simulated signals with known H values, the method yields estimators of H with lower mean squared errors. We characterized mammographic images with the proposed scaling estimator and anisotropy measures from non-decimated wavelet spectra for breast cancer detection, and obtained the best diagnostic accuracy in excess of 80%. Some real-life signals are known to possess a theoretical value of the Hurst exponent. Chapter 4 described a Bayesian scaling estimation method that utilizes the value of a theoretical scaling index as a mean of prior distribution and estimates H with MAP estimation. The accuracy of estimators from the proposed method is robust to small misspecification of the prior mean. We applied the method to a turbulence velocity signal and yielded

an estimator of H close to the theoretical value. Chapter 5 proposed two methods based on NDWT for robust estimation of Hurst exponent H of 1-D self-similar signals. The redundancy of NDWT, which improved the accuracy of estimation, introduced autocorrelations within the wavelet coefficients. With the two proposed methods, we alleviated the autocorrelation in three ways: taking the logarithm prior to taking the median, relating Hurst exponent to the median instead of mean of the model distribution, and resampling the coefficients.

Appendix A

PROOFS AND DERIVATIONS

A.1 2-D scale-mixing non-decimated wavelet transform

We derive expression (51). For 2-D fBf $B_H(x,y) \in \mathbb{R}^2$, detailed wavelet coefficients obtained by an NDWT located in one level from the main diagonal hierarchy is

$$\begin{aligned} d_{(j,j);k_1,k_2} &= \int\int B_H(x,y)\psi_{j;k_1}(x)\psi_{j;k_2}(y)\,dx\,dy \\ &= 2^j \int\int B_H(x,y)\psi(2^j(x-k_1))\psi(2^j(y-k_2))\,dx\,dy, \end{aligned} \tag{63}$$

where $(j,j) \in \boldsymbol{j}_s$. We can simplify (63) by letting $\boldsymbol{k} = (k_1, k_2)$ and $\mathbf{v}= (x,y)$.

$$d_{(j,j);\boldsymbol{k}} = 2^j \int\int B_H(\mathbf{v})\psi\big(2^j(\mathbf{v}-\boldsymbol{k})\big)\,d\mathbf{v}.$$

The energy of each decomposition level is the variance of the detailed wavelet coefficients $d_{(j,j);\boldsymbol{k}}$ Heneghan et al. (1996)

$$\begin{aligned} E\big[|d_{(j,j);\boldsymbol{k}}|^2\big] &= 2^{2j}\int\int \psi\big(2^j(\mathbf{v}-\boldsymbol{k})\big)\,\psi\big(2^j(\mathbf{u}-\boldsymbol{k})\big)E\big[B_H(\mathbf{v})B_H(\mathbf{u})\big]\,d\mathbf{v}d\mathbf{u} \\ &= \frac{\sigma_H^2}{2}2^{2j}\Big[\int\int \psi\big(2^j(\mathbf{v}-\boldsymbol{k})\big)\,\psi\big(2^j(\mathbf{u}-\boldsymbol{k})\big)|\mathbf{v}|^{2H}\,d\mathbf{v}d\mathbf{u} \\ &+ \int\int \psi\big(2^j(\mathbf{v}-\boldsymbol{k})\big)\,\psi\big(2^j(\mathbf{u}-\boldsymbol{k})\big)|\mathbf{u}|^{2H}\,d\mathbf{v}d\mathbf{u} \\ &- \int\int \psi\big(2^j(\mathbf{v}-\boldsymbol{k})\big)\,\psi\big(2^j(\mathbf{u}-\boldsymbol{k})\big)|\mathbf{v}-\mathbf{u}|^{2H}\,d\mathbf{v}d\mathbf{u}\Big]. \end{aligned} \tag{64}$$

By the property of wavelet filters, we know that

$$\int \psi(2^j(\mathbf{v}-\boldsymbol{k}))d\mathbf{v} = \int \psi(2^j(\mathbf{u}-\boldsymbol{k}))d\mathbf{u} = 0.$$

Thus we can simplify (64) as

$$E\big[|d_{(j,j);\boldsymbol{k}}|^2\big] = -\tfrac{\sigma_H^2}{2}2^{2j}\int\int \psi(2^j(\mathbf{v}-\boldsymbol{k}))\psi(2^j(\mathbf{u}-\boldsymbol{k}))|\mathbf{v}-\mathbf{u}|^{2H}d\mathbf{v}d\mathbf{u}.$$

We substitute $\boldsymbol{p} = 2^j(\mathbf{v} - \mathbf{u})$ and $\boldsymbol{q} = 2^j(\mathbf{u} - \boldsymbol{k})$ and obtain

$$\begin{aligned} E\left[|d_{(j,j);\boldsymbol{k}}|^2\right] &= -\frac{\sigma_H^2}{2} 2^{2j} \int\int \psi(\boldsymbol{p}+\boldsymbol{q})\psi(\boldsymbol{q}) 2^{-2Hj} |\boldsymbol{p}|^{2H} 2^{-4j} d\boldsymbol{p} d\boldsymbol{q} \\ &= -\frac{\sigma_H^2}{2} 2^{-j(2H+2)} \int\int \psi(\boldsymbol{p}+\boldsymbol{q})\psi(\boldsymbol{q}) |\boldsymbol{p}|^{2H} d\boldsymbol{p} d\boldsymbol{q} \\ &= \frac{\sigma_H^2}{2} V_\psi 2^{-j(2H+2)}, \end{aligned}$$

where $V_\psi = -\int\int \psi(\boldsymbol{p}+\boldsymbol{q})\psi(\boldsymbol{q})|\boldsymbol{p}|^{2H} d\boldsymbol{p} d\boldsymbol{q}$ which is dependent on wavelet function ψ and Hurst exponent H, but independent of j. σ_H^2 is given in (8). By taking logarithm on the energy, we obtain the relationship between wavelet coefficients and the Hurst exponent H.

$$\log_2 E\left[|d_{(j,j),\boldsymbol{k}}|^2\right] = -(2H+2)j + C.$$

A.2 Derivation of Bayesian estimation of scaling

Let $d_j = d_{j\boldsymbol{k}}$ be an arbitrary (w.r.t. $\boldsymbol{k}$) wavelet coefficient from the j^{th} level of the non-decimated wavelet decomposition of the m-dimensional fractional Brownian motion $B_H(\omega, \boldsymbol{t}), \boldsymbol{t} \in \mathbb{R}^m$,

$$d_j = \int_{\mathbb{R}^m} B_H(\omega, \boldsymbol{t}) \psi^*_{j\boldsymbol{k}}(\boldsymbol{t}) dt, \quad \text{for some fixed } \boldsymbol{k} = (k_1, \ldots, k_m)$$

Here $\psi^*_{j\boldsymbol{k}}(\boldsymbol{t}) = \prod_{i=1}^k \psi^*_{jk_i}(t_i)$ where ψ^* is either ψ or ϕ, but in the product there is at least one ψ. It is known that Flandrin (1992)

$$d_j \stackrel{d}{=} 2^{-(H+m/2)j}\ d_0,$$

where d_0 is a coefficient from the level $j = 0$, and $\stackrel{d}{=}$ means equality in distributions. Coefficient d_j is a random variable with expectation

$$\mathbb{E}d_j = 0, \quad \text{Var}d_j = \mathbb{E}d_j^2 = 2^{-(2H+m)j}\ \sigma^2,$$

where $\sigma^2 = \text{Var}d_0^2$.

The fBm $B_H(\boldsymbol{t})$ is a Gaussian m-dimensional field, thus

$$d_j \sim \mathcal{N}(0, 2^{-(2H+m)j}\sigma^2).$$

The rescaled "energy" is

$$\frac{2^{(2H+m)j}}{\sigma^2} d_j^2 \sim \chi_1^2$$

while, under assumption of independence,

$$\frac{2^{(2H+m)j}}{\sigma^2} \sum_{\boldsymbol{k} \in j\text{th level}} d_{j\boldsymbol{k}}^2 = \frac{2^{(2H+m)j+mJ}}{\sigma^2}\ \overline{d_j^2}$$

has $\chi^2_{2^{mJ}}$ distribution. Here J is the integer part of the logarithm for base 2 of the size of the signal.

Here, $\overline{d_j^2}$ is the average energy in j^{th} level.

Thus,

$$\overline{d_j^2} \stackrel{d}{=} 2^{-(2H+m)j-mJ}\sigma^2\chi^2_{2^{mJ}}.$$

From this,

$$\mathbb{E}\overline{d_j^2} = \sigma^2 2^{-(2H+m)j-mJ}\mathbb{E}\chi^2_{2^{mJ}} = 2^{-(2Hj+mj)}\sigma^2,$$

and

$$\mathrm{Var}\overline{d_j^2} = \sigma^4 2^{-(4H+2m)j-2mJ} \times 2 \cdot 2^{mJ} = 2^{-4Hj-2mj-mJ+1}\sigma^4.$$

The density of $\overline{d_j^2}$ for fixed H, j, m, and σ^2 is

$$g(y_j) = \left(\frac{1}{\Gamma(2^{mJ-1})}\right)\left(\frac{2^{(2H+m)j+mJ}}{2\sigma^2}\right)^{2^{mJ-1}} \quad (y_j)^{2^{mJ-1}-1} \exp\Big(-\tfrac{2^{mJ}}{2\sigma^2} y_j 2^{(2H+m)j}\Big)$$

Indeed, the cdf of $\overline{d_j^2}$ is

$$G(y_j) = \mathbb{P}(\overline{d_j^2} \le y_j) = \mathbb{P}\left(\chi^2_{2^{mJ}} \le \frac{2^{(2H+m)j+mJ}}{\sigma^2} y_j\right).$$

Then,

$$g(y) = G'(y) = f(h(y))\ |h'(y)|,$$

with $h(y_j) = \frac{2^{(2H+m)j+mJ}}{\sigma^2} y_j$ and $f(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} \exp\{-x/2\}$, $x \geq 0$, for $n = 2^{mJ}$. Once the energy at each level j, y_j, is calculated, we can obtain the likelihood:

$$\mathcal{L}(H, \sigma^2 | y_{j_1}, \ldots, y_{j_2}) = \prod_{j=j_1}^{j_2} g(y_j) = \left(\frac{(2\sigma^2)^{-2^{mJ-1}}}{\Gamma(2^{mJ-1})} \right)^{(j_2-j_1+1)} \times$$

$$\prod_{j=j_1}^{j_2} \left(2^{(2H+m)j+mJ}\right)^{2^{mJ-1}} \left(y_j\right)^{2^{mJ-1}-1} \times \exp\Big(- \sum_{j=j_1}^{j_2} \frac{2^{mJ}}{2\sigma^2} y_j 2^{(2H+m)j} \Big) =$$

$$\left(\frac{(2\sigma^2)^{-b/2}}{\Gamma(2^{b/2})} \right)^{c} \prod_{j=j_1}^{j_2} \left(2^{(2H+m)j} b\right)^{b/2} \left(y_j\right)^{b/2-1} \times \exp\Big(- \frac{b}{2\sigma^2} \sum_{j=j_1}^{j_2} y_j 2^{(2H+m)j} \Big)$$

where $b = 2^{mJ}$ and $c = j_2 - j_1 + 1$.

To obtain an expression proportional to the posterior distribution, we multiply likelihood function with a prior distribution, $\pi(H, \sigma^2)$,

$$\mathcal{L}(H, \sigma^2 | y_{j_1}, \ldots, y_{j_2}) \times \pi(H, \sigma^2).$$

As the Hurst exponent is supported on $(0, 1)$ interval, we selected beta(α, β) distribution as the prior on H. For the prior distribution of σ^2, we selected a non-informative (improper) prior $\frac{1}{\sigma^2}$. The parameters H and σ^2 are considered apriori independent, so their joint prior is

$$\pi(H, \sigma^2) = \frac{1}{\sigma^2} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} H^{\alpha-1}(1-H)^{\beta-1}.$$

A non-normalized posterior is

$$F = \pi(H, \sigma^2)\mathcal{L}(H, \sigma^2 | y_{j_1}, \ldots, y_{j_2}) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} H^{\alpha-1}(1-H)^{\beta-1} \Big(\frac{1}{\Gamma(b/2)} \Big)^{c} \Big(\frac{b}{2} \Big)^{bc/2} \times$$

$$\Big(\frac{1}{\sigma^2} \Big)^{bc/2+1} \prod_{j=j_1}^{j_2} 2^{(2H+m)jb/2} \left(y_j\right)^{b/2-1} \exp\Big(- \frac{b}{2\sigma^2} \sum_{j=j_1}^{j_2} y_j 2^{(2H+m)j} \Big). \tag{65}$$

Taking logarithm of (65) yields

$$\ln F = -\frac{bc+2}{2}\ln\sigma^2 + \sum_{j=j_1}^{j_2}\left[\frac{(2H+m)jb}{2}\ln 2 + \frac{b-2}{2}\ln y_j\right] - \frac{b}{2\sigma^2}\sum_{j=j_1}^{j_2} y_j 2^{(2H+m)j} + \ln\left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\right] + (\alpha-1)\ln H + (\beta-1)\ln(1-H) - c\ln\left[\Gamma(b/2)\right] + \frac{bc}{2}\ln[b/2]. \tag{66}$$

For the estimator that maximizes the posterior, maximizes its non-normalized version as well, and we proceed with finding the MAP estimator. First, we obtain σ^2 that maximize the likelihood by taking derivative

$$\frac{\partial \ln F}{\partial \sigma^2} = -\Big(\frac{bc+2}{2}\Big)\frac{1}{\sigma^2} + \frac{b}{2\sigma^4}\sum_{j=j_1}^{j_2} y_j 2^{(2H+m)j} = 0 \tag{67}$$

$$\sigma^2 = \frac{b\sum_{j=j_1}^{j_2} y_j 2^{(2H+m)j}}{bc+2} \tag{68}$$

Using (68) obtained, we can express (66) with respect to H only and take derivative to obtain H that maximize the likelihood.

$$\begin{aligned}\ln F &= -\frac{bc+2}{2}\ln\left[\frac{b\sum_{j=j_1}^{j_2} y_j 2^{(2H+m)j}}{bc+2}\right] + \sum_{j=j_1}^{j_2}\left[\frac{(2H+m)jb}{2}\ln 2 + \frac{b-2}{2}\ln y_j\right] \\ &- \frac{bc+2}{2} + \ln\left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\right] + (\alpha-1)\ln H + (\beta-1)\ln(1-H) \\ &- c\ln\left[\Gamma(b/2)\right] + \frac{bc}{2}\ln[b/2]. \end{aligned} \tag{69}$$

$$\begin{aligned}\frac{\partial \ln F}{\partial H} &= -(bc+2)\frac{\ln 2\sum_{j=j_1}^{j_2} y_j j 2^{(2H+m)j}}{\sum_{j=j_1}^{j_2} y_j 2^{(2H+m)j}} + b\ln 2\sum_{j=j_1}^{j_2} j + \frac{\alpha-1}{H} \\ &- \frac{\beta-1}{1-H} = 0. \end{aligned} \tag{70}$$

There is no closed form solution for H, se we numerically approximate its value by solving equations in (70) numerically.

A.3 Derivation of MEDL

Proof of Theorem 5.1.1

A single wavelet coefficient in a non-decimated wavelet transform of fBm(H) is normally distributed, with variance depending on its level j,

$$d_j \stackrel{d}{=} \mathcal{N}(0, 2^{-(2H+1)j}\sigma^2).$$

Its rescaled square is χ^2 with one degree of freedom,

$$\delta = \frac{2^{(2H+1)j}}{\sigma^2} d_j^2 \stackrel{d}{=} \chi_1^2,$$

with density

$$\frac{\delta^{1/2-1}(\frac{1}{2})^{1/2}}{\Gamma(\frac{1}{2})} e^{-\delta/2} = \frac{e^{-\delta/2}}{\sqrt{2\delta}\Gamma(\frac{1}{2})}.$$

The pdf of d_j^2 is

$$f(d_j^2) = \frac{e^{-c_j d_j^2/2}}{\sqrt{2c_j d_j^2}\Gamma(1/2)} c_j,$$

where $c_j = \frac{2^{(2H+1)j}}{\sigma^2}$. Let $y = \log d_j^2$, then $d_j^2 = e^y$ and $\frac{\partial d_j^2}{\partial y} = e^y$. The pdf of y is

$$f(y) = \frac{c_j e^{\frac{-c_j e^y}{2}}}{\sqrt{2c_j e^y}\Gamma(1/2)} e^y = \frac{\sqrt{c_j} e^{-\frac{c_j e^y}{2}} e^{y/2}}{\sqrt{2}\Gamma(1/2)} = \sqrt{\frac{c_j}{2\pi}} e^{-\frac{c_j e^y}{2}} e^{y/2},$$

The cdf of y is

$$F(y) = \int_{-\infty}^{y} f(t)dt = 2\Phi\left(\sqrt{c_j} e^{y/2}\right) - 1,$$

where Φ is the cdf of standard normal distribution. Let y^* be the median of the distribution of y. We obtain the expression of $y*$ by solving $F(y^*) = 1/2$. This results in

$$y^* = 2\log\left[\frac{1}{\sqrt{c_j}}\Phi^{-1}(3/4)\right]$$

From this equation, we can find a link between y^* and the Hurst exponent H by substituting c_j,

$$\begin{aligned} y^* &= 2\log[\Phi^{-1}(3/4)] - \log c_j \\ &= -\log 2\,(2H+1)j + \log\sigma^2 + 2\log[\Phi^{-1}(3/4)] \qquad (71) \\ &= -\log 2\,(2H+1)j + C, \end{aligned}$$

where C is a constant independent of the level j.

Proof of Lemma 5.1.1

An approximation of variance of sample median $\hat{y}_j^*$ is obtained using normal approximation to a quantile of absolutely continuous distributions,

$$Var(\hat{y}_j^*) \approx \frac{1}{4N(f(y_j^*))^2}.$$

After substituting the expression for y^* we obtain Lemma 5.1.1

$$Var(\hat{y}_j^*) \approx \frac{\pi e^Q}{2NQ}, \qquad Q = \left[\Phi^{-1}(3/4)\right]^2 \approx 0.4549.$$

Thus the variance of the sample median is approximately $5.4418/N$.

Proof of Theorem 5.1.2

An NDWT-based spectrum uses the pairs

$$\left(j,\ \hat{y}_j^*\right),\ j = J-m-a-1, \ldots, J-a-1.$$

from m decomposition levels, starting with a coarse $j = J-m-a$ level and ending with finer level $j = J-1-a$. Here a is an arbitrary integer between 0 and $J-3$. When $a = 0$, the finest level $j = J-1$ until level $J-1-m$ are used.

Then the spectral slope is

$$\hat{\beta} = \frac{12}{m(m^2-1)} \sum_{j=J-m-a-1}^{J-1-a} (j - J - a - (m+1)/2)\ \hat{y}_j^*.$$

The estimator $\widehat{\beta}$ is unbiased,

$$\begin{aligned} E\hat{\beta} &= \frac{12}{m(m^2-1)} \sum_{j=J-m-a-1}^{J-1-a} (j - J - a - (m+1)/2)\ \ (-\log 2\,(2H+1)j + C) \\ &= -\log 2\,(2H+1), \end{aligned}$$

where C is a constant and H is the theoretical Hurst exponent.

By substituting $\mathrm{Var}(\hat{y}_j^*) = A/N$ from Theorem 5.1.1 we find

$$\mathrm{Var}(\widehat{\beta}) = \frac{12A}{Nm(m^2-1)},\ \text{and} \quad \mathrm{Var}(\widehat{H}) = \frac{3A}{Nm(m^2-1)(\log 2)^2},$$

for $\widehat{H} = -\widehat{\beta}/(2\log 2) - 1/2$.

Thus, the MEDL estimator $\widehat{H}$ is approximately normal with mean H and variance $3A/(Nm(m^2-1)(\log 2)^2)$, where $A \cong 5.4418$, N is the sample size, and m is the number of levels used for the spectrum.

A.4 Derivation of MEDLA

Proof of Theorem 5.1.3.

We begin by selecting the pair of wavelet coefficients that follow a normal distribution with a zero mean and a variance dependent on level j, from which the wavelet coefficients are sampled.

$$d_{j,k_1}, d_{j,k_2} \sim \mathcal{N}(0, 2^{-(2H+1)j}\sigma^2),$$

where σ is the standard deviation of wavelet coefficients from level 0, k_1 and k_2 are positions of wavelet coefficients in level j, and H is the Hurst exponent. We also assume that coefficients d_{j,k_1} and d_{j,k_2} are independent, which is a reasonable assumption when the distance $|k_1 - k_2| > q_j = 2^{J-j}$. Then, we define δ as

$$\delta = \frac{2^{(2H+1)j}}{\sigma^2}(d_{j,k_1}^2 + d_{j,k_2}^2) = C_j \cdot a,$$

for $C_j = \frac{2\cdot 2^{(2H+1)j}}{\sigma^2}$ and $a = \frac{d_1^2+d_2^2}{2}$. Since δ follow χ_2^2 distribution, the pdf of the average of two squared wavelet coefficients a is

$$f(a) = \frac{C_j}{2}e^{-\frac{C_j a}{2}}.$$

Denote $y = \log a$. The pdf and cdf of y are

$$\begin{aligned} f(y) &= \frac{C_j}{2}e^{-\frac{C_j e^y}{2}}e^y \\ F(y) &= 1 - e^{-C_j e^y/2}. \end{aligned}$$

Let y^* be the median of y. The expression for y^* is obtained by solving $F(y^*) = 1 - e^{-C_j e^{y^*}/2} = 1/2$,

$$y^* = \log(\log 4) - \log C_j.$$

After replacing C_j with $\frac{2 \cdot 2^{(2H+1)j}}{\sigma^2}$, the median becomes

$$y^* \quad = -\log 2\,(2H+1)j + \log \sigma^2 + \log(\log 2),$$

similarly as in (71) in the MEDL method.

Proof of Lemma 5.1.2

An approximation of variance of sample median is obtained as

$$Var(\hat{y}_j^*) \approx \frac{1}{4n(f(y_j^*))^2}.$$

After plugging in the expression for y^* into $\frac{1}{4n(f(y_j^*))^2}$, we obtain

$$Var(\hat{y}_j^*) \approx \frac{1}{N(\log 2)^2},$$

Thus the variance of the sample median in MEDLA method is approximately $2.08/N$.

Proof of Theorem 5.1.4

For the distribution of $\widehat{H}$ from MEDLA, we follow the same regression steps on pair $(j, \hat{y_j^*})$ as in Appendix A.3. By substituting $\mathrm{Var}(\hat{y}_j^*) = \frac{1}{N(\log 2)^2}$ from (5.1.2) we find

$$\mathrm{Var}(\widehat{\beta}) = \frac{12}{Nm(m^2-1)(\log 2)^2}, \text{ and } \quad \mathrm{Var}(\widehat{H}) = \frac{3}{Nm(m^2-1)(\log 2)^4},$$

for $\widehat{H} = -\hat{\beta}/(2\log 2) - 1/2$. Thus, the MEDLA estimator $\hat{H}$ is approximately normal with mean H and variance $3/(Nm(m^2-1)(\log 2)^2)$, where N is the sample size, and m is the number of levels used for the spectrum.

Bibliography

P. Abry, P. Gonçlavès, and P. Flandrin. Wavelet-based spectral analysis of 1/f processes. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 3, pages 237–240, 1993.

P. Abry, P. Flandrin, M. Taqqu, and D. Veitch. Wavelets for the analysis, estimation, and synthesis of scaling data. In K. Park and W. Willinger, editors, *Self-Similar Network Traffic and Performance Evaluation*, pages 39–88. Wiley, 2000a.

P. Abry, P. Gonçalvés, and J. L. Vehel. *Scaling, Fractals and Wavelets*. Wiley-ISTE, 2013. ISBN 978-1-118-62290-2.

Patrice Abry, Paulo Gonçalvés, and Patrick Flandrin. Wavelets, spectrum analysis and 1/f processes. In *Wavelets and statistics*, pages 15–29. Springer, 1995.

Patrice Abry, Patrick Flandrin, Murad S Taqqu, and Darryl Veitch. Wavelets for the analysis, estimation and synthesis of scaling data. *Self-similar network traffic and performance evaluation*, pages 39–88, 2000b.

AmericanCancerSociety. Cancer facts & figures 2014. *Atlanta: American Cancer Society*, 2014.

A Arneodo, Y d'Aubenton Carafa, E Bacry, PV Graves, JF Muzy, and C Thermes. Wavelet based fractal analysis of dna sequences. *Physica D: Nonlinear Phenomena*, 96(1):291–320, 1996.

E. Bacry, A. Arneodo, J.F. Muzy, and P.V. Graves. A wavelet based multifractal formalism: application to the study of dna sequences. *J. Tech. Phys.*, 37(3-4): 281–284, 1996.

G.A. Barnard. Discussion of hurst. *Proc. Inst. Civ. Eng.*, 5:552–553, 1956.

S Benmehdi, N Makarava, N Benhamidouche, and M Holschneider. Bayesian estimation of the self-similarity exponent of the Nile river fluctuation. *Nonlinear Processes in Geophysics*, 18(3):441–446, 2011.

Jan Beran. *Statistics for Long-memory Processes*. Chapman & Hall, 1994.

Gregory Beylkin. On the representation of operators in bases of compactly supported wavelets. *SIAM Journal on Numerical Analysis*, 29(6):1716–1740, 1992.

D Biskamp. Cascade models for magnetohydrodynamic turbulence. *Physical Review E*, 50(4):2702, 1994.

K Bowyer, D Kopans, WP Kegelmeyer, R Moore, M Sallam, K Chang, and K Woods. The digital database for screening mammography. In *Third International Workshop on Digital Mammography*, volume 58, 1996.

P. J. Burt and E. H. Adelson. A multiresolution spline with applications to image mosaic. *ACM Trans. Graphics*, 2:217–236, 1983a.

P. J. Burt and E. H. Adelson. The Laplacian pyramid as a compact image code. *IEEE Trans. Comm.*, COM-31:532–540, 1983b.

Ronald R Coifman and David L Donoho. *Translation-invariant de-noising*. Springer, 1995.

PL Conti, A Lijoi, and F Ruggeri. A Bayesian approach to the analysis of telecommunication systems performance. *Applied Stochastic Models in Business and Industry*, 20:305–321, 2004.

Paul Doukhan, George Oppenheim, and Murad S Taqqu. *Theory and applications of long-range dependence*. Springer Science & Business Media, 2003.

Richard Durbin, Sean R Eddy, Anders Krogh, and Graeme Mitchison. *Biological sequence analysis: probabilistic models of proteins and nucleic acids*. Cambridge university press, 1998.

W. Feller. The asymptotic distribution of the range of sums of independent random variables. *Ann. Math. Statist.*, 22:427–432, 1951.

P. Flandrin. Some aspects of non-stationary signal processing with emphasis on time-frequency and time-scale methods. In J. M. Combes, A. Grossmann, and Ph. Tchamitchian, editors, *Wavelets Time-Frequency Methods and Phase Space*, pages 68–98, New York, 1989a. Springer-Verlag.

P. Flandrin. On the spectrum of fractional Brownian motions. *IEEE Trans. Inform. Theory*, 35(1):197–199, 1989b.

Patrick Flandrin. Wavelet analysis and synthesis of fractional Brownian motion. *Information Theory, IEEE Transactions on*, 38(2):910–917, 1992.

T Graves, RB Gramacy, CLE Franzke, and NW Watkins. Efficient Bayesian inference for ARFIMA processes. *Nonlinear Processes in Geophysics Discussions*, 2:573–618, 2015.

A. Grossmann and J. Morlet. Decomposition of functions into wavelets of constant shape and related transforms. In L. Streit, editor, *Mathematics and physics, lectures on recent results*. World Scientific, River Edge, NJ, 1985.

C Heneghan, SB Lowen, and Malvin C Teich. Two-dimensional fractional Brownian motion: wavelet analysis and synthesis. In *Image Analysis and Interpretation, 1996., Proceedings of the IEEE Southwest Symposium on*, pages 213–217. IEEE, 1996.

Scott Holan, Tucker McElroy, Sounak Chakraborty, et al. A Bayesian approach to estimating the long memory parameter. *Bayesian Analysis*, 4(1):159–190, 2009.

Timothy S Horbury, Miriam Forman, and Sean Oughton. Anisotropic scaling of magnetohydrodynamic turbulence. *Physical Review Letters*, 101(17):175005, 2008.

Jonathan RM Hosking. Fractional differencing. *Biometrika*, 68(1):165–176, 1981.

Seonghye Jeon, Orietta Nicolis, and Brani Vidakovic. Mammogram diagnostics via 2-D complex wavelet-based self-similarity measures. *São Paulo Journal of Mathematical Sciences*, 8(2):256–284, 2014.

Kyungduk Ko and Marina Vannucci. Bayesian wavelet analysis of autoregressive fractionally integrated moving-average processes. *Journal of Statistical Planning and Inference*, 136(10):3415–3434, 2006.

John Lamperti. Semi-stable stochastic processes. *Transactions of the American mathematical Society*, pages 62–78, 1962.

Markus Lang, Haitao Guo, Jan E Odegard, C Sidney Burrus, and Raymond O Wells Jr. Nonlinear processing of a shift-invariant discrete wavelet transform (dwt) for noise reduction. In *SPIE's 1995 Symposium on OE/Aerospace Sensing and Dual Use Photonics*, pages 640–651. International Society for Optics and Photonics, 1995.

Franccois Le Gall. Powers of tensors and fast matrix multiplication. In *Proceedings of the 39th International Symposium on Symbolic and Algebraic Computation*, volume arXiv:1401.7714, 2014.

Jie Liang and Thomas W Parks. A translation-invariant wavelet representation algorithm with applications. *Signal Processing, IEEE Transactions on*, 44(2):225–232, 1996.

Natallia Makarava, Sabah Benmehdi, and Matthias Holschneider. Bayesian estimation of self-similarity exponent. *Physical Review E*, 84(2):021109, 2011.

Stephane Mallat. Zero-crossings of a wavelet transform. *Information Theory, IEEE Transactions on*, 37(4):1019–1033, 1991.

Stéphane Mallat. *A Wavelet Tour of Signal Processing, Second Edition*. Academic Press, 1999.

Stephane G Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 11(7):674–693, 1989a.

Stephane G Mallat. Multiresolution approximations and wavelet orthonormal bases of L^2 (R). *Transactions of the American mathematical society*, 315(1):69–87, 1989b.

B.B. Mandelbrot. *Les objects fractals: forme, hasard et dimension*. Flammarion, Paris, 1975.

B.B. Mandelbrot and J. W. J. W. Van Ness. Fractional Brownian motions, fractional noises and applications. *SIAM rev.*, 10(4):422–437, 1968.

B.B. Mandelbrot and J. R. Wallis. Some long-run properties of geophysical records. *Water Resources Res. 5*, 5:321–340, 1969.

Bryan T Marshall, Krishna K Sarangapani, Jizhong Lou, Rodger P McEver, and Cheng Zhu. Force history dependence of receptor-ligand dissociation. *Biophysical journal*, 88(2):1458–1466, 2005.

J. Martin, M. Moskowitz, and J. Milbrath. Breast cancer missed by mammography. *American Journal of Roentgenology*, 37(2):142–162, 1979.

K. P. Moloney, J. A. Jacko, B. Vidakovic, F. Sainfort, V. K. Leonard, and B. Shi. Leveraging data complexity: Pupillary behavior of older adults with visual impairment during HCI. *Journal ACM Transactions on Computer-Human Interaction (TOCHI)*, 13(3):376–402, 2006.

Satoshi Morita, Peter F Thall, and Peter Müller. Determining the effective sample size of a parametric prior. *Biometrics*, 64(2):595–602, 2008.

J. Morlet, G. Arens, I. Fourgeau, and D. Giard. Wave propagation and sampling theory. *Geophysics*, 47:203–236, 1982.

Guy P Nason and Bernard W Silverman. The stationary wavelet transform and some statistical applications. In *Wavelets and statistics*, pages 281–299. Springer, 1995.

Mark Nelkin. Scaling theory of hydrodynamic turbulence. *Physical Review A*, 11(5): 1737, 1975.

Orietta Nicolis, Pepa Ramírez-Cobo, and Brani Vidakovic. 2D wavelet-based spectra with applications. *Computational Statistics & Data Analysis*, 55(1):738–751, 2011.

Ivan Nourdin, Anthony Réveillac, et al. Asymptotic behavior of weighted quadratic variations of fractional Brownian motion: the critical case h= 1/4. *The Annals of Probability*, 37(6):2200–2230, 2009.

Jeffrey S Pai and Nalini Ravishanker. Bayesian analysis of autoregressive fractionally integrated moving-average processes. *Journal of Time Series Analysis*, 19(1):99–112, 1998.

Magda Peligrad and Sunder Sethuraman. On fractional Brownian motion limits in one dimensional nearest-neighbor symmetric simple exclusion. *arXiv preprint arXiv:0711.0017*, 2007.

C.-K. Peng, S.V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A.L. Goldberger. Mosaic organization of dna nucleotides. *Phys. Rev. E*, 49:1685–1689, 1994.

Donald B Percival and Andrew T Walden. *Wavelet methods for time series analysis*, volume 4. Cambridge university press, 2006.

Darıo G Pérez, Luciano Zunino, and Mario Garavaglia. Modeling turbulent wavefront phase as a fractional Brownian motion: a new approach. *JOSA A*, 21(10): 1962–1969, 2004.

J-C Pesquet, Hamid Krim, and Hervé Carfantan. Time-invariant orthonormal wavelet representations. *Signal Processing, IEEE Transactions on*, 44(8):1964–1970, 1996.

Pepa Ramirez and Brani Vidakovic. Wavelet-based 2D multifractal spectrum with applications in analysis of digital mammography images. *Technical Report 02-2007*, 2007.

Pepa Ramírez-Cobo and Brani Vidakovic. A 2D wavelet-based multiscale approach with applications to the analysis of digital mammograms. *Computational Statistics & Data Analysis*, 58(The Third Special Issue on Statistical Signal Extraction and Filtering):71–81, 2013.

Nalini Ravishanker and Bonnie K Ray. Bayesian analysis of vector arfima processes. *Australian Journal of Statistics*, 39(3):295–311, 1997.

Irving S Reed, PC Lee, and Trieu-Kien Truong. Spectral representation of fractional Brownian motion in n dimensions and its properties. *Information Theory, IEEE Transactions on*, 41(5):1439–1451, 1995.

Norbert Remenyi, Orietta Nicolis, Guy Nason, and Brani Vidakovic. Image denoising with 2-D scale-mixing complex wavelet transforms. 2014.

Erez Ribak. Atmospheric turbulence, speckle, and adaptive optics. *Annals of the New York Academy of Sciences*, 808(1):193–204, 1997.

C Schwartz, G Baum, and EN Ribak. Turbulence-degraded wave fronts as fractal surfaces. *JOSA A*, 11(1):444–451, 1994.

Haipeng Shen, Zhengyuan Zhu, and Thomas CM Lee. Robust estimation of the self-similarity parameter in network traffic using wavelet transform. *Signal Processing*, 87:2111–2124, 2007.

Mark J Shensa. The discrete wavelet transform: wedding the a trous and mallat algorithms. *Signal Processing, IEEE Transactions on*, 40(10):2464–2482, 1992.

Skander Soltani, Patrice Simard, and Daniel Boichu. Estimation of the self-similarity parameter using the wavelet transform. *Signal Processing*, 84(1):117–123, 2004.

Gilbert Strang and Truong Nguyen. *Wavelets and filter banks*. SIAM, 1996.

Jason Swanson. Fluctuations of the empirical quantiles of independent Brownian motions. *Stochastic Processes and their Applications*, 2011.

Ahmed H Tewfik and M Kim. Correlation structure of the discrete wavelet coefficients of fractional Brownian motion. *IEEE transactions on information theory*, 38(2): 904–909, 1992.

Michael Unser. Texture classification and segmentation using wavelet frames. *Image Processing, IEEE Transactions on*, 4(11):1549–1560, 1995.

Marina Vannucci and Fabio Corradi. Modeling dependence in the wavelet domain. In *Bayesian inference in wavelet-based models*, pages 173–186. Springer, 1999.

Darryl Veitch and Patrice Abry. A wavelet-based joint estimator of the parameters of long-range dependence. *Information Theory, IEEE Transactions on*, 45(3):878–897, 1999.

Brani Vidakovic. *Statistical modeling by wavelets*, volume 503. John Wiley & Sons, 1999.

VITA

Minkyoung Kang received her B.S. in industrial and systems engineering from KAIST in South Korea. She worked as a research intern at Telefonica Research in Madrid while earning her M.S. in telecommunications and network engineering from the Technical University of Madrid. She is currently a Ph.D. candidate investigating parameter estimation in wavelet domains. During her doctoral studies, she interned and analyzed real-world data related to credit history, equipment monitoring, and employees performance history.