FUNCTIONAL DATA MINING WITH MULTISCALE STATISTICAL PROCEDURES

A Thesis
Presented to
The Academic Faculty

by

Kichun Lee

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 2010

FUNCTIONAL DATA MINING WITH MULTISCALE STATISTICAL PROCEDURES

Approved by:

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

Professor Xiaoming Huo
School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Paul Kvam
School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Justin Romberg
School of Electrical and Computer Engineering
Georgia Institute of Technology

Professor Ming Yuan
School of Industrial and Systems Engineering
Georgia Institute of Technology

Date Approved: August 2010

To my family:

Hanbyeol, Joseph, and Lumin.

ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my advisor Prof. Brani Vidakovic for the continuous support of my Ph.D study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D study.

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Xiaoming Huo, Prof. Justin Romberg, Prof. Paul Kvam, and Prof. Ming Yuan for their encouragement, insightful comments, and hard questions.

My sincere thanks also goes to Prof. F. DuBois Bowman, Prof. Suzette LaRoche, Prof. Youngja Park, and Prof. Sheryl Gabram-Mendola in Emory; Prof. Charlene Bayer, Prof. Chen Zhou, Prof. Alexander Gray, and Prof. Barbara Fasse in GaTech; and Prof. Gabriel Katul in Duke for giving me opportunities to work with them and helping me grow.

Also I thank my friends: Gordana Derado, Pepa Ramirez Cobo, Seonghye Jeon, Hinkyeol Woo, Kiersten Petersen, and Jorge Arroyo.

Last but not the least, I would like to thank my family: my wife Hanbyeol Cho, my first son Joseph Lee, and my second son Lumin Lee for letting me know how much God loves me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

The discipline of statistics has been challenged by advances of technical capability in processing and storing real-life measurements. The devices of modern technology compile vast amounts of data that are functional and multivariate in nature (for example, time-dependent functional responses from multiple sources). An increasingly important approach to make sense of such functional and high-dimensional data is dimension reduction via assessment of regularity (also called as Hurst exponent and related to fractal dimension) of functional paths or methodologies in machine learning.

In this thesis we tackle two approaches for dimension reduction: (1) development of informative and descriptive summaries based on the regularity of data, and (2) use of topologies in which data "live" to enhance dimension-reduction process. The common theme in the thesis is multiscale, and we use wavelet domains for either tool-building or for illustration.

The indices of the regularity (summarized in multifractal spectrum) and various wavelet-based spectra have been studied for their theoretical properties in mathematics and probability and for their practical use in applied disciplines. One application is worth emphasizing: When the data consist of high-frequency bio-medical responses, numerous studies have been conducted to link human conditions and diagnostics with the regularity of the responses. It was shown that properly assessed indices of regularity in measurements carry significant diagnostic information about the patient.

Many of real-life measurements can be collected as labeled and unlabeled. Such data are often observed as a mixture. For developing classifiers traditionally only labeled data are used and some recent research efforts suggest the use of unlabeled

data to enhance the classification method. These are referred as semi-supervised learning. In this thesis we develop methodology based on semi-supervised learning to enhance some multiscale procedures, notably wavelet shrinkage.

The thesis is organized as follows.

In the second chapter, we propose a stable method to extract the time-varying regularity and the noise level of the data. Theoretical contributions of the proposed model is simultaneous treatment of high-frequency process and an additive white noise. We illustrate this methodology on EEG data, in which medical researchers look for the onset and signature of vasospasm, a potentially fatal neurological episode.

In the third chapter, we develop a method of testing whether the high-frequency data set is consistent with monofractality using wavelet-generated multifractal descriptors. We discuss theoretical properties of the descriptors, their computational implementation, the use in data mining, and the effectiveness in the context of simulations. Applications include turbulence and analysis of coding/noncoding regions in DNA sequences.

In the fourth chapter, we propose a novel denoising methodology that combines classical wavelet shrinkage methods with state-of-art machine-learning techniques. This methodology takes advantage of geometric structure of wavelet coefficients. Theoretical optimality properties of the proposed method are discussed and its performance is demonstrated in a comparative manner.

CHAPTER I

INTRODUCTION

Classes of random processes which are intrinsically invariant to changes in scale are increasingly finding their way into many fields: geoscience, medicine, economics, physics, electrical and computer engineering.

We look at such processes from the standpoint of statistical modeling and introduce several traditional models. We also introduce multiscale methods (wavelets, wavelet-like decompositions, general time/frequency representations) as tools and environments to analyze and model such processes and to unify several related phenomena including fractality, multifractality, and long range dependence. Finally, we present the overview of the remainder of this thesis.

1.1 Self-similar Processes

We start with various examples of self-similar processes.

1.1.1 Examples Scaling Processes

Statistically self-similar processes (such as fractional Brownian motion) and $1/f$ processes with power law spectra are becoming fundamental in modeling of wide-range of real-world phenomena from the fields of engineering, physics, medicine, biology, engineering, art, economics, astronomy, chemistry, etc.

1.1.1.1 It Started with Hurst and Nile Data

British hydrologist Harold Edwin Hurst spent 62 years in Egypt and mostly worked on design and construction of reservoirs along the Nile River. By inspecting historical data on the Nile River flows, Hurst discovered phenomenon (now called Hurst effect).

Hurst was trying to find an optimal reservoir capacity R such that it can accept the river flow in N units of time,$X_1, X_2, \ldots X_N$, and have a constant withdrawal of $\bar{X}$ per unit time. The optimal volume of the reservoir was given by the so called adjusted range,

$$R = \max_{1 \le k \le N} (X_1 + \cdots + X_k - k\bar{X}) - \min_{1 \le k \le N} (X_1 + \cdots + X_k - k\bar{X}) \tag{1.1.1}$$

Since the records for the waterflow rarely exceeded 100 years Hurst inspected other geophysical data and in order to compare them, he standardized their adjusted ranges R, with sample standard deviation

$$S = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})^2} \ , \tag{1.1.2}$$

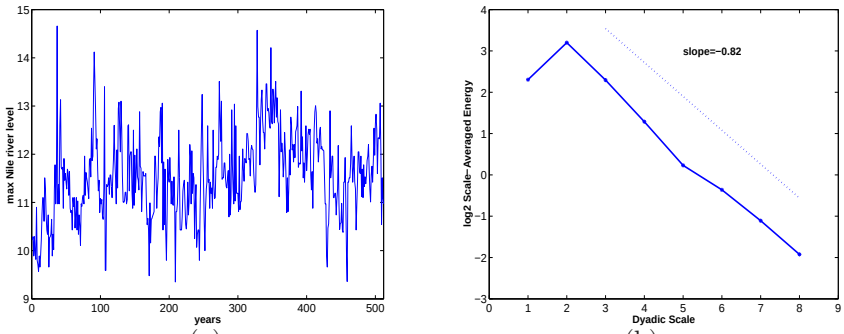
and obtained dimensionless ratio R/S - rescaled and adjusted range. On basis of more that 800 records, he found [45] that quantity R/S scales as N^H, for ranging from 0.46 to 0.93, with mean 0.73 and standard deviation of 0.09.

This result was is contrast with the fact that for independent normal random variables H is 1/2 in limit. Feller proved that the limit is 1/2 for independent identically distributed random variables with finite second moment, this limit was 1/2. It was believed that strong Markovian dependence was responsible for this deviation untill Barnard [8] proved that limit $H = 1/2$ holds for the Markovian dependence case.

It was the work of Mandelbrot [58], Mandelbrot and Van Ness [59], and Mandelbrot and Wallis [60] who associated the Hurst (or Joseph) phenomenon on the presence of long-memory. Figure 1(a) gives n=512 consecutive yearly measurements from the famous Nile River Data set for the years 62-1281 A.D. Panel (b) gives its wavelet spectra demonstrating the scaling law.

1.1.1.2 ESCA Spectrum

The ESCA spectrum used in this example was provided by J.P. Bibérian, of the Université de Marseille – Luminy. This set is one of the template data sets in WaveLab

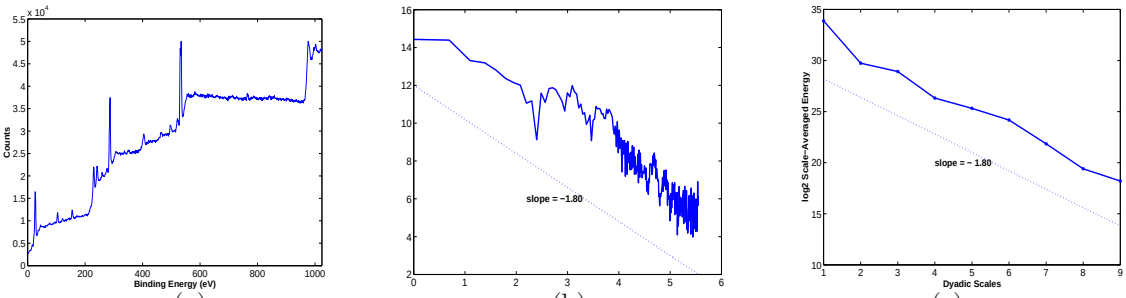


(a) (b)

Figure 1: (a) Nile yearly minimal level data; (b) its Wavelet log spectra

802 and it is sampled at 1024 equally spaced values. Here is short description of ESCA spectrum methodology.

Electron Spectroscopy for Chemical Analysis (ESCA), also referred to as X-ray Photoelectron Spectroscopy (XPS), irradiates the sample surface with a soft (low energy) X-ray. This X-ray excites the electrons of the sample atoms, and if their binding energy is lower than the X-ray energy, they will be emitted from the parent atom as a photoelectron. Only the photoelectrons at the extreme outer surface (10-100 Angstroms (Å); $1\mathring{A} = 10^{-10}m$) can escape the sample surface, making this methodology a surface analysis technique.



(a) (b) (c)

Figure 2: (a) ESCA spectrum; (b) Scaling behavior in the Fourier domain; (c) and in the wavelet domain.

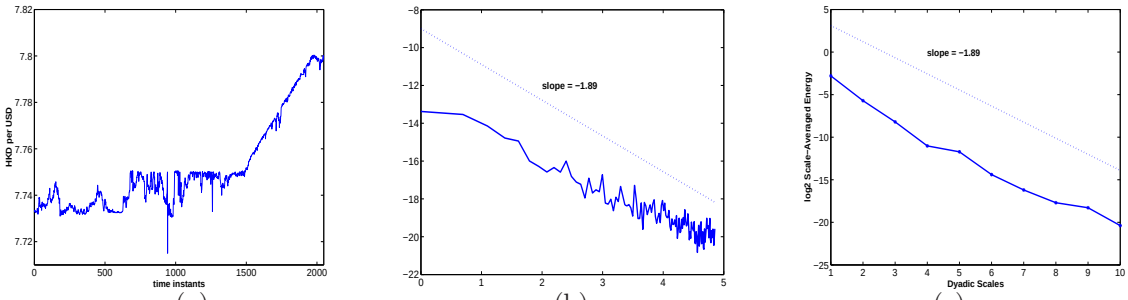
An ESCA spectrum consists of a series of peaks corresponding the the binding energies of the photoelectrons that produced these peaks. ESCA analysis not only provides elemental information, but because the technique is detecting the binding energy of emitted electrons, it can also provide some chemical bonding information.

Depending on what elements the parent atom is bound to, the binding energy of the emitted photoelectrons may shift slightly. Figure 2(a) shows the ESCA spectrum (usual ordering of energy in eV is opposite than in Panel (a), it ranges from large to small). Panels (b) and (c) represent log spectrum and log wavelet spectra. Clear power law with the slope of - 1.80 (indicated in Panels (b) and (c)) by dotted lines) is notable.

1.1.1.3 Exchange Rates

Many economic time series, such as stock market prices, exchange rates and asset returns exhibit scaling laws and long range dependence. This is in empirical contradiction to several economic theories (random walk theory for stock market, perfect markets, etc) and gave rise to several theories and models describing the scaling and LRD (such as ARFIMA, fGn, fBm, GARCH, etc).

The rates of exchange between Hong Kong Dollar (HKD) and USDollar (USD) as reported by the ONADA Company between 24 March 1995 and 1 November 2000. Figure 3(a) shows the rates of exchange. Panels (b) and (c) represent log spectrum and log wavelet spectra, which show clear power law with the slope of -1.89.



(a) (b) (c)

Figure 3: (a) Exchange Rates HKD per US$; (b) scaling behavior in the Fourier domain, and (c) in the wavelet domain.

1.1.2 Definitions and Properties of Self-similar Processes

We define the notions of self-similar and long-range dependent processes.

We assume that all processes discusses are real valued and defined on the same parameter space. Two processes $X(t)$ and $Y(t)$, *equal in all finite dimensional distributions*, will be denoted as $X(t) \stackrel{d}{=} Y(t)$. This means that for any selection of "times" $0 \leq t_1 < t_2 < \ldots t_k < \infty$ random vectors $(X(\omega, t_1), \ldots, X(\omega, t_k))$ and $(Y(\omega, t_1), \ldots, Y(\omega, t_k))$ have the same distribution. Informally, processes equal-in-distribution are statistically indistinguishable.

Random process $X(t)$ is called stochastically continuous at t_0 if $\lim_{h \to 0} P(|X(t_0 + h) - X(t_0)| > \epsilon) = 0$, for any fixed $\epsilon > 0$. Additionally, we consider processes not to be trivial. [1]

Definition 1.1.1. A random process $X(t), t > 0$ is called *self-similar* if for any $a > 0$, there exists $b > 0$ such that

$$X(at) \stackrel{d}{=} bX(t). \tag{1.1.3}$$

Lamperti [54] proved the result,

Theorem 1.1.1. *(Lamperti, 62.) If random process $X(t), t \geq 0$ is nontrivial, stochastically continuous at 0, and self-similar, then there exists unique $H \geq 0$ such that $b = a^H$. If $X(0) = 0, a.s.$ then $H > 0$.*

Standard definition of self-similar processes is as follows,

Definition 1.1.2. Process $X(t),\ t \geq 0$ is self-similar, with self-similarity index H (H-ss) if and only if there exists $H > 0$ such that for any $a > 0$, $X(at) \stackrel{d}{=} a^H X(t)$.

Uniqueness of H is not obvious from this definition, although, H is unique by the Lamperti's theorem. Also, from Definition 1.1.2 it follows $X(0) = 0$.

Example 1.1.1. Standard Brownian Motion $B(t)$ is 1/2-ss. Indeed, the process $W(t) = 1/\sqrt{a}B(at)$ is standard Brownian motion, as well.

[1] Process $X(t)$ is trivial if the distribution of random variable $X(\omega, t),\ t$ fixed is a point mass measure. For example, $X(t) = \text{const}$ or $X(t) = \sin(t)$ would be examples of trivial processes.

Let $\{X(t),\ t \in \mathbb{R}\}$ be a random process such that the autocovariance function

$$\gamma_X(r,s) = \mathcal{COV}(X(r), X(s)) = E(X(r) - EX(r))(X(s) - EX(s)) \tag{1.1.4}$$

is finite for any pair $r, s \in \mathbb{R}$ The random process (time series) $\{X(t),\ t \in \mathbb{R}\}$ is said to be (weakly, second-order, or wide-sense) stationary if

(i) $E|X(t)|^2 < \infty$,

(ii) $EX(t) = m$, for all $t \in \mathbb{R}$, and

(iii) $\gamma_X(r,s) = \gamma_X(r+t, s+t)$ for all r, s, and $t \in \mathbb{R}$.

The stationarity condition (iii) is often given in the form,

$$\mathcal{COV}(X_{t+h}, X_t) = \gamma_X(h),$$

emphasizing the independence of t. When it is clear what the underlying process is, we will write $\gamma(h)$ instead of $\gamma_X(h)$. If the index space for parameter t is not $\mathbb{R}$ but the set of integers, $\mathbb{Z}$ random process $X(t)$ is called random sequence or time series and often indexed as $X_t, t \in \mathbb{Z}$.

Example 1.1.2. (i) *White noise* is a stationary sequence Z_t such that $EZ_t = 0$ and $\gamma(h) = \sigma^2 \cdot \delta_h$; in notation, $Z_t \sim \mathcal{WN}(0, \sigma^2)$.

(ii) The *moving average* MA(q) process, defined as

$$X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots + \theta_q Z_{t-q},\ Z_t \sim \mathcal{WN}(0, \sigma^2),$$

has autocovariance function

$$\gamma(h) = \begin{cases} \sigma^2 \sum_{j=0}^{q-|h|} \theta_j \theta_{j+|h|}, & |h| \le q \\ 0 & |h| > q. \end{cases} \tag{1.1.5}$$

(iii) The *autoregressive* AR(1) process, $X_t - \phi X_{t-1} = Z_t,\ Z_t \sim \mathcal{WN}(0, \sigma^2)$, has autocovariance function

$$\gamma(h) = \sigma^2 \frac{\phi^{|h|}}{1 - \phi^2}. \tag{1.1.6}$$

Autocorrelation function is defined as normalized version of autocovariance function, $\rho(h) = \gamma(h)/\gamma(0)$. We note that $\gamma(0) = Var(X(t))$ is constant for stationary processes.

Fourier transformation of autocorrelation (or autocovariance) function leads to spectral density $f(\omega)$

$$f(\omega) = \int_R \gamma(h) e^{-ih\omega} dh, \tag{1.1.7}$$

which is non-negative by Wiener-Khinchine theorem. Properly normalized [2] indeed represents a density in a probabilistic sense. The function $f(\omega)$ is also called power-spectrum since $E|X(t)|^2 = 1/(2\pi)\int_R f(\omega)d\omega$, and $E|X(t)|^2$ represents the "power" of zero-mean signal $X(t)$.

It is possible define a counterpart of a spectral density of a nonstationary process if, for example, linear filtering will produce a stationary process. In such cases, we define a pseudo (quasi) spectral density as a function of spectral density of filtered stationary process and transfer function of a filter.

Now we are ready for the definition of long range dependence and $1/f$ processes.

A stationary process $Y(t)$ is called long-range dependent (LRD) process if its autocorrelation function or spectral density behave as

$$\gamma_Y(h) \sim C_\gamma |h|^{\alpha-1}, \ h \to \infty, \alpha \in (0,1), \tag{1.1.8}$$

or

$$f_Y(\omega) \sim C_f |\omega|^{-\alpha}, \ \omega \to 0, \alpha \in (0,1), \tag{1.1.9}$$

where C_γ and C_f are two related constants. These two relations are equivalent, subject to mild asymptotic monotonicity assumptions on γ.

Next, we will make link between self-similarity and LRD.

[2] Depending on the definition of Fourier transformation, in our case f should be divided by $2\pi\gamma(0)$

Let $X(t), t \in R$ be $H-ss$ process. If its increments are stationary, i.e, if the distribution of $X(t+h)-X(t)$ is independent of t, it will be called $H-sssi$ process. The following theorem gives the form of autocorrelation function of any $H-sssi$ process with finite second moment.

Theorem 1.1.2. *Let $X(t), t \in \mathbb{R}$ be an $H-sssi$ process for which $E|X(1)|^2 < \infty$. Then,*

$$\gamma(t,s) = EX(t)X(s) = \frac{E|X(1)|^2}{2}\left[|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right]. \tag{1.1.10}$$

Proof: From H-ss and stationarity of increments property,

$$\begin{aligned} EX(t)X(s) &= \frac{1}{2}\left[E(X(t)^2) + E(X(s)^2) - E[X(t)-X(s)]^2\right] \\ &= \frac{1}{2}\left[t^{2H}E(X(1)^2) + s^{2H}E(X(1)^2) - E[X(|t-s|)-X(0)]^2\right] \\ &= \frac{E|X(1)|^2}{2}\left[t^{2H} + s^{2H} - |t-s|^{2H}\right]. \quad \square \end{aligned}$$

Let $X(t)$ be an H-sssi process with $0 < H < 1$ and $E|X(1)|^2 < \infty$. Define stationary sequence of random variables $Y(n)$ as

$$Y(n) = X(n+1) - X(n).$$

If $\gamma_Y(n)$ is the autocorrelation function for $Y(n)$, i.e., $\gamma_Y(n) = EY(n)Y(0)$, then if $H = 1/2$, $\gamma_Y(n) = 0$, for $n \geq 1$ and if $H \neq 1/2$, it is possible to find an explicit expression for $\gamma(n)$. Using the fact that $X(0) = 0$ we find,

$$\begin{aligned} \gamma_Y(n) = EY(1)Y(n) &= EX(1)(X(n+1) - X(n)) \\ &= E(X(1)X(n+1) - E(X(1)X(n)) \qquad (1.1.11) \\ &= \frac{E|X(1)|^2}{2}\left[(n+1)^{2H} - n^{2H} + (n-1)^{2H}\right]. \end{aligned}$$

If in (1.1.11) the expressions $(n \pm 1)^{2H}$ are replaced by their polynomial expansions $n^{2H} \pm 2Hn^{2H-1} + H(2H-1)n^{2H-1} + \ldots$, the following asymptotic result holds

$$\lim_{n\to\infty} \frac{\gamma_Y(n)}{H(2H-1)E|X(1)|^2\, n^{2H-2}} = 1. \tag{1.1.12}$$

In other words, $\gamma_Y(n) = O(n^{2H-2})$. Note that series $\sum_n |\gamma_Y(n)|$ converges if $2-2H > 1$ or, equivalently, if $0 < H < 1/2$. For such H, expression $2H-1$ in (1.1.12) is negative, and correlations $\gamma_Y(n)$ are negative. If $1/2 < H < 1$ correlations $\gamma(n)$ are positive, but $\sum_n |\gamma_Y(n)| = \infty$, since $2 - 2H < 1$.

Long range dependent process $Y(n)$ is asymptotically second-order self-similar, i.e., the second order moments of Y_n and aggregated time series $Y^{(m)}$ coincide. The series $Y^{(m)}(k)$ is defined as series of averages of non-overlapping blocks of size m from the sequence $Y(n)$,

$$Y^{(m)}(k) = \frac{Y(km - m + 1) + \cdots + Y(km)}{m}, \tag{1.1.13}$$

It is easy to see that if $1/2 < H < 1$, the asymptotic behavior of $VarY^{(m)}$ is influenced by asymptotic behavior of $\gamma_Y(n)$. Indeed,

$$VarY^{(m)} \sim 1/m\ \gamma_Y(0) + \sum_{k=1}^{m-1} k^{2H-2}(m-k) \sim m^{2H-2}.$$

Informally, $Y^{(m)}(k)$ and $Y(n)$ look similar at all scales, and we will see later that this asymptotic behavior of the variance of aggregated process, $VarY^{(m)} \sim m^{2H-2}$, can be used for inference about H.

1.1.3 Fractional Brownian Motion (fBm) and Fractional Gaussian Noise (fGn)

Fractional Brownian motion (fBm)is generalization of Brownian motion (Wienner Process). Brownian motion $B(t)$ is standardly defined as random process satisfying: (i) B(0)=0, (ii) for any choice n and $0 \le t_1 < t_2 < \cdots < t_n$, the increments $B(t_2) - B(t_1), \ldots, B(t_n) - B(t_{n-1})$ are independent and stationary; (iii) B(t) is Gaussian random variable with zero mean and variance t, and (iv) $B(t)$ is a continuous function of t, a.s. It is easy to check that Brownian motion is an 1/2-sssi process, since $W(t) = a^{-1/2}B(at)$ satisfies properties (i)-(iv). Covariance of Brownian motion is, because of Theorem 1.1.2, $EB(t)B(s) = 1/2(t + s - |t - s|) = \min\{t, s\}$.

Brownian motion is Gausian process and Gaussian processes are fully determined by their second order properties. Therefore, Brownian motion is unique Gaussian process having covariance function $\gamma(t,s) = \min\{t,s\}$. The theorem 1.1.2 gave the covariance structure for an H-sssi process. If such process is gaussian it is unique an it is called fractional Brownian motion.

Definition 1.1.3. A zero mean gaussian process $B_H(t)$ is called fractional Brownian motion with Hurst exponent H, if

$$EX(t)X(s) = \frac{E|X(1)|^2}{2}\left[|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right],$$

where $E|X(1)|^2 = \frac{\Gamma(2-2H)\cos(\pi H)}{\pi H(1-2H)}$.

The process $B_H(t)$ is unique, in the sense that class of all fractional Brownian motions with exponent H coincides with the class of all Gaussian $H-ss$ processes. However, a Gaussian process is $H-ss$ with independent increments, if and only if it $H = 1/2$, i.e., if it is a Brownian motion.

The difference process, $Y(n) = B_H(n+1) - B_H(n)$ is called fractional Gaussian noise (fGn). As for more general H-sssi processes, covariance function of fGn is

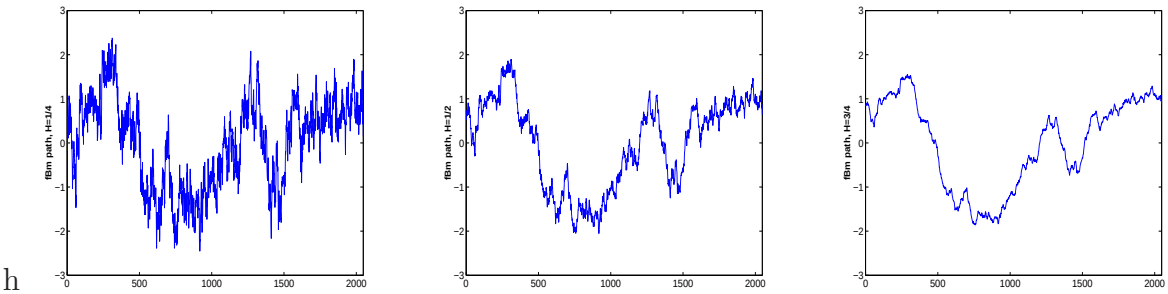
$$\gamma(h) = \frac{E|X(1)|^2}{2}\left[(h+1)^{2H} - h^{2H} + (h-1)^{2H}\right]. \tag{1.1.14}$$

An alternative definition of fractional Brownian motion can be given via stochastic integration. Mandelbrot and Van Ness [59], Taqqu defined fBm as the process for which:

(i) $B_H(0) = 0,$ and

$$(ii)\quad B_H(t) = 1/C_H \cdot \left[\int_{-\infty}^{0}[(t-s)^{H-1/2} - (-s)^{H-1/2}]B(ds) + \int_0^t (t-s)^{H-1/2}B(ds)\right], \tag{1.1.15}$$

where $B(dt)$ is the Wiener measure, and $1/C_H = \Gamma(H+1/2)/(\Gamma(2H+1)\sin(\pi H))^{1/2}$.



h

(a) (b) (c)

Figure 4: Simulated paths of fractional Brownian motion, (a) $H = 1/4$, (b) $H = 1/2$, and (c) $H = 3/4$.

This representation can be discretized, in the sense that discrete counterpart of $B(dt)$ is normal noise, and as such it provides a way to simulate fBm.

Sample paths of fractional Brownian motion are behaving similarly to those of standard Brownian motion. See Figure 4 for simulated paths of several values of H. They are continuous almost surely for all $H \in (0, 1)$ and nowhere differentiable. The fractal (Hausdorff) dimension of sample paths is $D = 2 - H$. That means that for small H (say, $H < 0.5$) the sample paths are quite irregular and *space-filling.* It is interesting that sample paths of fractional Brownian motions are also continuous in H, a result of Peltier and Lévy-Véhel [64, 65].

Let $\Delta(i) = B_H(\frac{i+1}{n}) - B_H(\frac{i}{n})$, for $i = 1, \ldots, n-1$. Let $\Delta_{1:n} \leq \ldots \Delta_{n:n}$ be the corresponding order statistics. Define the polynomial $VB_{H,n}(t) = \sum_{i=0}^{[nt]-1} \Delta_{i:n} + (nt - [nt])\Delta_{[nt]:n}$. Phillpe and Thilly [66] demonstrated that

$$\frac{VB_{H,n}(t)}{n^{1-H}\sqrt{C}} \to L(t), \tag{1.1.16}$$

where $L(t) = -\frac{1}{\sqrt{2\pi}} \exp -\frac{1}{2}\Phi^{-1}(t)$, and Φ is the standard Gaussian cdf. This result can be utilized to estimate H. Similarly to integral representation (1.1.15), fBm allows the so called harmonizable representation,

$$B_H(t) = \int_{\mathbb{R}} \frac{e^{it\omega} - 1}{|\omega|^{H+1/2}} B(d\omega), \tag{1.1.17}$$

where $B(d\omega)$ is the Wiener measure.

1.1.4 Autoregressive, Fractionally Integrated, Moving Average Processes (ARFIMA)

The autoregressive fractionally integrated moving average model, denoted ARFIMA (p,d,q), can also be used for the statistical modeling of time series with long memory. The more familiar ARMA models (when $d = 0$) or ARIMA (when d is a positive integer) are special cases. ARFIMA models were defined by Granger and Joyeux [41] and Hosking [44].

Define the back-lag operator B as $B^k Y(n) = Y(n-k),\ k = 0, 1, \ldots$. We write the ARFIMA(p,d,q) model as:

$$\Phi(B)(1-B)^d Y(n) = \Theta(B)\epsilon(n), n = 1, 2, \ldots \tag{1.1.18}$$

where $\Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p)$ is the autoregressive polynomial and $\Theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$ is the moving average polynomial in the back-lag operator B; p and q are integers, d is real. The innovations $\epsilon(n)$ are assumed i.i.d. normal with zero mean and variance σ^2. The fractional difference operator can be expresses by the following binomial expansion:

$$(1-B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k,$$

where $\binom{d}{k}$ is defined as $\frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)}$.

For $d > 1/2$ the ARFIMA process is not stationary, although it can be diferenced to a stationary process. For $-0.5 < d < 0$ the process is called intermediate memory or *overdifferenced*, see Brockwell and Davis [12]. The ARFIMA model exhibit long memory when when $0 < d < 1/2$.

In particular, ARFIMA(0,d,0) is the model model of special interest since it is close to fractional Gaussian noise with parameter $H = d + 1/2$. Defined as $(1-B)^d Y(n) = \epsilon(n)$, it is responsible for the LRD behavior in $Y(n)$. Indeed, $Y(n)$ can be represented as infinite moving average process, since

$$Y(n) = (1-B)^{-d}\epsilon(n). \tag{1.1.19}$$

Thus,

$$\begin{aligned} Y(n) &= \sum_{k=0}^{\infty} \binom{-d}{k} (-1)^k \epsilon(n-k) \\ &= \sum_{k=0} \frac{\Gamma(k+d)}{\Gamma(k+1)\Gamma(d)} \, \epsilon(n-k). \end{aligned}$$

Covariance function of $Y(t)$ is given as

$$\begin{aligned} \gamma_Y(h) &= E(Y(n)Y(n+h)) = \sigma^2 \frac{(-1)^h(-2d)!}{(h-d)!(-h-d)!} \\ &= \gamma(0) \frac{\Gamma(1-d)\Gamma(h+d)}{\Gamma(d)\Gamma(h+1-d)}, \end{aligned} \tag{1.1.20}$$

where $\gamma(0) = \sigma^2 \frac{(-d)!}{((-d)!)^2}$.

If $\sigma^2 = 1$, i.e., if the noise is independent standard normal, one has $\gamma(h) = \frac{(-1)^h(-2d)!}{(h-d)!(-h-d)!}$. Autocorrelation function can be represented as

$$\rho(h) = \gamma(h)/\gamma(0) = \frac{d(1+d)\dots(h-1+d)}{(1-d)(2-d)\dots(h-d)}, \; h = 1, 2, 3, \dots.$$

By Sheppard's Formula, $\rho(h) \sim h^{-2d-1}$. For details see Deriche and Tewfik [29]. In fact, an ARFIMA(p,d,q) can be decomposed on two components, ARFIMA (0,d,0)-part, responsible for the long memory and an ARMA(p,q)-part responsible for the short-range memory. In fact, one way to generate an ARFIMA(p,d,q) process is to generate ARMA(p,q) process with ARFIMA(0,d,0) innovations.

Example 1.1.3. An ARFIMA(0, 1+d, 0) [an integrated ARFIMA(0,d,0)] process can be thought as a discrete, sampled process from an $fBm(H)$ process, for $H = a + 1/2$.

1.1.5 Multifractional Brownian Motion (mBm)

Not all scalings are perfect or universal. Scaling exponent in fBm is a global parameter. Holder regularity of paths of fBm are constant and in fact equal to the Hurst exponent.

A model that accommodates change of parameter H over time is multifractional Brownian motion, mBm. Definition of multifractional Brownian motion can be given

via stochastic integration. follows that of Mandelbrot and Van Ness [59] for fBm as with H replaced by $H(t)$.

Definition 1.1.4. Multifractional Brownian Motion is a process for which:

(i) $W(0) = 0$, and

$$(ii)\quad W(t) = \frac{1}{\Gamma(H_t+1/2)}\left[\int_{-\infty}^{0}[(t-s)^{H_t-1/2}-(-s)^{H_t-1/2}]dB(s) + \int_0^t (t-s)^{H_t-1/2}dB(s)\right],$$

where $B(t)$ is standard Brownian motion. The function $H_t : [0,\infty) \mapsto [a,b] \subset (0,1)$ is assumed to be Holder function of exponent $\beta > 0$.

The frequency definition of mBm(H) is achieved by [10]

$$W(t) = \int_{\mathbb{R}} \frac{e^{it\omega}-1}{|\omega|^{H(t)+1/2}} B(d\omega), \tag{1.1.21}$$

We discuss details of process W in the second chapter.

1.2 Basic on Wavelets

The first theoretical results in wavelets are connected with continuous wavelet decompositions of $\mathbb{L}_2$ functions and go back to the early 1980s. Papers of Morlet *et al.* [61] and Grossmann and Morlet [62] were among the first on this subject.

Let $\psi_{a,b}(x),\ a \in \mathbb{R}\backslash\{0\}, b \in \mathbb{R}$ be a family of functions defined as translations and re-scales of a single function $\psi(x) \in \mathbb{L}_2(\mathbb{R})$,

$$\psi_{a,b}(x) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{x-b}{a}\right). \tag{1.2.1}$$

Normalization by $\frac{1}{\sqrt{|a|}}$ ensures that $||\psi_{a,b}(x)||$ is independent of a and b. The function ψ (called *the wavelet function* or *the mother wavelet*) is assumed to satisfy the *admissibility condition,*

$$C_\psi = \int_{\mathbb{R}} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty, \tag{1.2.2}$$

where $\Psi(\omega) = \int_R \psi(x) e^{-ix\omega} dx$ is the Fourier transformation of $\psi(x)$. The admissibility condition (1.2.2) implies

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

Also, if $\int \psi(x)dx = 0$ and $\int (1 + |x|^\alpha)|\psi(x)|dx < \infty$ for some $\alpha > 0$, then $C_\psi < \infty$. Wavelet functions are usually normalized to "have unit energy", i.e., $||\psi_{a,b}(x)|| = 1$.

For any $\mathbb{L}_2$ function $f(x)$, the continuous wavelet transformation is defined as a function of two variables

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

Here the dilation and translation parameters, a and b, respectively, vary continuously over $\mathbb{R}\backslash\{0\} \times \mathbb{R}$.

Resolution of Identity. When the admissibility condition is satisfied, i.e., $C_\psi < \infty$, it is possible to find the inverse continuous transformation via the relation known as *resolution of identity* or *Calderón's reproducing identity,*

$$f(x) = \frac{1}{C_\psi} \int_{\mathbb{R}^2} \mathcal{CWT}_f(a,b) \psi_{a,b}(x) \frac{da\ db}{a^2}.$$

If a is restricted to $\mathbb{R}^+$, which is natural since a can be interpreted as a reciprocal of frequency, (1.2.2) becomes

$$C_\psi = \int_0^\infty \frac{|\Psi(\omega)|^2}{\omega} d\omega < \infty, \tag{1.2.3}$$

and the *resolution of identity* relation takes the form

$$f(x) = \frac{1}{C_\psi} \int_{-\infty}^\infty \int_0^\infty \mathcal{CWT}_f(a,b) \psi_{a,b}(x) \frac{1}{a^2} da\ db. \tag{1.2.4}$$

Next, we list a few important properties of continuous wavelet transformations.

Shifting Property. If $f(x)$ has a continuous wavelet transformation $\mathcal{CWT}_f(a,b)$, then $g(x) = f(x - \beta)$ has the continuous wavelet transformation $\mathcal{CWT}_g(a,b) =$

$\mathcal{CWT}_f(a, b-\beta)$.

Scaling Property. If $f(x)$ has a continuous wavelet transformation $\mathcal{CWT}_f(a,b)$, then $g(x) = \frac{1}{\sqrt{s}}\ f\left(\frac{x}{s}\right)$ has the continuous wavelet transformation $\mathcal{CWT}_g(a,b) = \mathcal{CWT}_f\left(\frac{a}{s}, \frac{b}{s}\right)$.

Both the shifting property and the scaling property are simple consequences of changing variables under the integral sign.

Energy Conservation. From (1.2.4),

$$\int_{-\infty}^{\infty} |f(x)|^2 dx = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_0^{\infty} |\mathcal{CWT}_f(a,b)|^2 \frac{1}{a^2} da\ db.$$

Localization. Let $f(x) = \delta(x - x_0)$ be the Dirac pulse at the point x_0. Then, $\mathcal{CWT}_f(a,b) = \frac{1}{\sqrt{a}}\psi(\frac{x_0 - b}{a})$.

Reproducing Kernel Property. Define $\mathbb{IK}(u,v;a,b) = \langle \psi_{u,v}, \psi_{a,b} \rangle$. Then, if $F(u,v)$ is a continuous wavelet transformation of $f(x)$,

$$F(u,v) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_0^{\infty} \mathbb{IK}(u,v;a,b) F(a,b) \frac{1}{a^2} da\ db,$$

i.e., $\mathbb{IK}$ is a reproducing kernel. The associated reproducing kernel Hilbert space (RKHS) is defined as a $\mathcal{CWT}$ image of $\mathbb{L}_2(\mathbb{R})$ – the space of all complex-valued functions F on $\mathbb{R}^2$ for which $\frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_0^{\infty} |F(a,b)|^2 \frac{da\ db}{a^2}$ is finite.

Characterization of Regularity. Let $\int (1 + |x|)\ |\psi(x)|\, dx < \infty$ and let $\Psi(0) = 0$. If $f \in \mathbb{C}^\alpha$(Hölder space with exponent α), then

$$|\mathcal{CWT}_f(a,b)| \leq C|a|^{\alpha + 1/2}. \tag{1.2.5}$$

Conversely, if a continuous and bounded function f satisfies (1.2.5), then $f \in \mathbb{C}^\alpha$.

Example 1.2.1. Mexican hat or Marr's wavelet. The function

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

is a wavelet [known as the "Mexican hat" or Marr's wavelet.

By direct calculation one may obtain $C_\psi = 2\pi$.

Example 1.2.2. Poisson wavelet. The function $\psi(x) = -(1 + \frac{d}{dx})\frac{1}{\pi}\ \frac{1}{1+x^2}$ is a wavelet [known as the Poisson wavelet. The analysis of functions with respect to this wavelet is related to the boundary value problem of the Laplace operator.

The continuous wavelet transformation of a function of one variable is a function of two variables. Clearly, the transformation is redundant. To "minimize" the transformation one can select discrete values of a and b and still have a transformation that is invertible. However, sampling that preserves all information about the decomposed function cannot be coarser than the *critical sampling.*

The critical sampling (Fig. 5) defined by

$$a = 2^{-j},\ b = k2^{-j},\ j, k \in \mathbb{Z}, \tag{1.2.6}$$

will produce the minimal basis. Any coarser sampling will not give a unique inverse transformation; that is, the original function will not be uniquely recoverable. Moreover under mild conditions on the wavelet function ψ, such sampling produces an orthogonal basis $\{\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k),\ j, k \in \mathbb{Z}\}$.

There are other discretization choices. For example, selecting $a = 2^{-j},\ b = k$ will lead to non-decimated (or stationary) wavelets. For more general sampling, given by

$$a = a_0^{-j},\ b = k\, b_0\, a_0^{-j},\ j, k \in \mathbb{Z},\ a_0 > 1, b_0 > 0, \tag{1.2.7}$$

numerically stable reconstructions are possible if the system $\{\psi_{jk},\ j, k \in \mathbb{Z}\}$ constitutes a frame. Here

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

is (1.2.1) evaluated at (1.2.7).

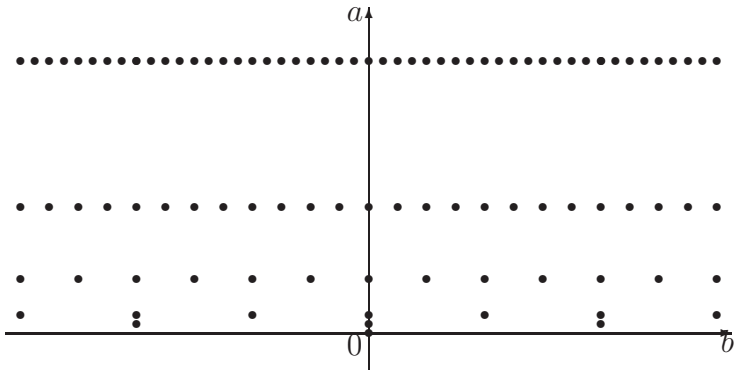


Figure 5: Critical Sampling in $\mathbb{R} \times \mathbb{R}^+$ half-plane ($a = 2^{-j}$ and $b = k\ 2^{-j}$).

Next, we consider wavelet transformations (wavelet series expansions) for values of a and b given by (1.2.6). An elegant theoretical framework for critically sampled wavelet transformation is *Mallat's Multiresolution Analysis* [55, 56, 57].

1.2.1 Multiresolution Analysis

A multiresolution analysis (MRA) is a sequence of closed subspaces $V_n, n \in \mathbb{Z}$ in $\mathbb{L}_2(\mathbb{R})$ such that they lie in a containment hierarchy

$$\cdots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \cdots . \tag{1.2.8}$$

The nested spaces have an intersection that contains the zero function only and a union that is dense in $\mathbb{L}(\mathbb{R})$,

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

[With $\overline{A}$ we denoted the closure of a set A]. The hierarchy (1.2.8) is constructed such that (i) V-spaces are self-similar,

$$f(2^j x) \in V_j \text{ iff } f(x) \in V_0. \tag{1.2.9}$$

and (ii) there exists *a scaling function* $\phi \in V_0$ whose integer-translates span the space V_0,

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

and for which the set $\{\phi(\bullet - k),\ k \in \mathbb{Z}\}$ is an orthonormal basis.[3]

Mild technical conditions on ϕ are necessary for future developments. It is assumed $\int \phi(x)dx \neq 0$. Since $V_0 \subset V_1$, the function $\phi(x) \in V_0$ can be represented as a linear combination of functions from V_1, i.e.,

$$\phi(x) = \sum_{k\in\mathbb{Z}} h_k\ \sqrt{2}\phi(2x - k), \tag{1.2.10}$$

for some coefficients $h_k,\ k \in \mathbb{Z}$. This equation is called the *scaling equation* (or two-scale equation) and it is fundamental in constructing,exploring, and utilizing wavelets.

In the wavelet literature, the reader may encounter an indexing of the multiresolution subspaces, which is the reverse of that in (1.2.8),

$$\cdots \subset V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \subset \cdots . \tag{1.2.11}$$

The coefficients h_n in (1.2.10) are important in connecting the MRA to the theory of signal processing. The (possibly infinite) vector $\mathbf{h} = \{h_n,\ n \in \mathbb{Z}\}$ will be called a *wavelet filter.* It is a low-pass (averaging) filter as will become clear later by considerations in the Fourier domain.

To further explore properties of multiresolution analysis subspaces and their bases, we will often work in the Fourier domain. Define the function m_0 as follows:

$$m_0(\omega) = \frac{1}{\sqrt{2}} \sum_{k\in\mathbb{Z}} h_k e^{-ik\omega} = \frac{1}{\sqrt{2}} H(\omega). \tag{1.2.12}$$

The function in (1.2.12) is sometimes called the *transfer function* and it describes the behavior of the associated filter $\mathbf{h}$ in the Fourier domain. Notice that the function m_0 is periodic with the period 2π and that the filter taps $\{h_n,\ n \in \mathbb{Z}\}$ are the Fourier coefficients of the function $H(\omega) = \sqrt{2}\ m_0(\omega)$. In the Fourier domain, the relation (1.2.10) becomes

$$\Phi(\omega) = m_0\left(\frac{\omega}{2}\right) \Phi\left(\frac{\omega}{2}\right), \tag{1.2.13}$$

[3]It is possible to relax the orthogonality requirement. It is sufficient to assume that the system of functions $\{\phi(\bullet - k),\ k \in \mathbb{Z}\}$ constitutes a Riesz basis for V_0.

where $\Phi(\omega)$ is the Fourier transformation of $\phi(x)$. Indeed,

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

By iterating (1.2.13), one gets

$$\Phi(\omega) = \prod_{n=1}^{\infty} m_0\left(\frac{\omega}{2^n}\right), \tag{1.2.14}$$

which is convergent under very mild conditions on rates of decay of the scaling function ϕ. There are several sufficient conditions for convergence of the product in (1.2.14). For instance, the uniform convergence on compact sets is assured if (i) $m_0(\omega) = 1$ and (ii) $|m_0(\omega) - 1| < C|\omega|^{\epsilon}$, for some positive C and ϵ. See also Theorem 1.2.3.

Next, we prove two important properties of wavelet filters associated with an orthogonal multiresolution analysis, *normalization* and *orthogonality.*

Normalization.

$$\sum_{k\in\mathbb{Z}} h_k = \sqrt{2}. \tag{1.2.15}$$

Proof:

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

Since $\int \phi(x)dx \neq 0$ by assumption, (1.2.15) follows. $\square$

This result also follows from $m_0(0) = 1$, since $\int \phi(x)dx \neq 0$ and $\phi \in \mathbb{L}_1(\mathbb{R})$ in the time domain translate to $\Phi(0) \neq 0$ and $\Phi(\omega) \in \mathbb{L}_\infty$ in the Fourier domain.

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

$$\sum_k h_k h_{k-2l} = \delta_l. \tag{1.2.16}$$

Proof: Notice first that from the scaling equation (1.2.10) it follows that

$$\begin{aligned} \phi(x)\phi(x-l) &= \sqrt{2} \sum_k h_k \phi(2x-k)\phi(x-l) \\ &= \sqrt{2} \sum_k h_k \phi(2x-k) \sqrt{2} \sum_m h_m \phi(2(x-l)-m). \end{aligned} \tag{1.2.17}$$

By integrating the both sides in (1.2.17) we obtain

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

The last line is obtained by taking $k = 2l + m$. An important special case is $l = 0$ for which (1.2.16) becomes

$$\sum_k h_k^2 = 1. \tag{1.2.18}$$

One consequence of the orthogonality condition (1.2.16) is the following: the convolution of filter $\mathbf{h}$ with itself, $\mathbf{f} = \mathbf{h} \star \mathbf{h}$, is an *à trous.*[4]

The fact that the system $\{\phi(\bullet - k),\ k \in \mathbb{Z}\}$ constitutes an orthonormal basis for V_0 can be expressed in the Fourier domain in terms of either $\Phi(\omega)$ or $m_0(\omega)$.

(**a**) In terms of $\Phi(\omega)$:

$$\sum_{l=-\infty}^{\infty} |\Phi(\omega + 2\pi l)|^2 = 1. \tag{1.2.19}$$

[4]The attribute *à trous (Fr.)* ($\equiv$ with holes) comes from the property $f_{2n} = \delta_n$, i.e., each tap on even position in $\mathbf{f}$ is 0, except the tap f_0. Such filters are also called half-band filters.

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

$$\begin{aligned}\delta_k &= \int_{\mathbb{R}} \phi(x)\overline{\phi(x-k)}dx \\ &= \frac{1}{2\pi}\int_{\mathbb{R}} \Phi(\omega)\overline{\Phi(\omega)}e^{i\omega k}d\omega \\ &= \frac{1}{2\pi}\int_0^{2\pi}\sum_{l=-\infty}^{\infty}|\Phi(\omega+2\pi l)|^2 e^{i\omega k}d\omega. \end{aligned} \tag{1.2.20}$$

The last line in (1.2.20) is the Fourier coefficient a_k in the Fourier series decomposition of

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

Due to the uniqueness of Fourier representation, $f(\omega) = 1$.

Remark. Utilizing the identity (1.2.19), any set of independent functions spanning V_0, $\{\phi(x-k), k \in \mathbb{Z}\}$, can be orthogonalized in the Fourier domain. The orthonormal basis is generated by integer-shifts of the function

$$\mathcal{F}^{-1}\left[\frac{\Phi(\omega)}{\sqrt{\sum_{l=-\infty}^{\infty}|\Phi(\omega+2\pi l)|^2}}\right]. \tag{1.2.21}$$

This normalization in the Fourier domain is used in constructing of some wavelet bases.

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

$$|m_0(\omega)|^2 + |m_0(\omega+\pi)|^2 = 1. \tag{1.2.22}$$

Since $\sum_{l=-\infty}^{\infty}|\Phi(2\omega+2l\pi)|^2 = 1$, then by (1.2.13)

$$\sum_{l=-\infty}^{\infty}|m_0(\omega+l\pi)|^2|\Phi(\omega+l\pi)|^2 = 1. \tag{1.2.23}$$

Now split the sum in (1.2.23) into two sums – one with odd and the other with even indices, i.e.,

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

To simplify the above expression, we use relation (1.2.19) and the 2π-periodicity of $m_0(\omega)$.

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

Whenever a sequence of subspaces satisfies MRA properties, there exists (though not unique) an orthonormal basis for $\mathbb{L}_2(\mathbb{R})$,

$$\{\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k),\ j,k \in \mathbb{Z}\} \tag{1.2.24}$$

such that $\{\psi_{jk}(x),\ j\text{-fixed},\ k \in \mathbb{Z}\}$ is an orthonormal basis of the "difference space" $W_j = V_{j+1} \ominus V_j$. The function $\psi(x) = \psi_{00}(x)$ is called a *wavelet function* or informally *the mother wavelet*.

Next, we detail the derivation of a wavelet function from the scaling function. Since $\psi(x) \in V_1$ (because of the containment $W_0 \subset V_1$), it can be represented as

$$\psi(x) = \sum_{k\in\mathbb{Z}} g_k\ \sqrt{2}\phi(2x-k), \tag{1.2.25}$$

for some coefficients $g_k,\ k \in \mathbb{Z}$. Define

$$m_1(\omega) = \frac{1}{\sqrt{2}}\ \sum_k g_k e^{-ik\omega}. \tag{1.2.26}$$

By mimicking what was done with m_0, we obtain the Fourier counterpart of (1.2.25),

$$\Psi(\omega) = m_1(\frac{\omega}{2})\Phi(\frac{\omega}{2}). \tag{1.2.27}$$

The spaces W_0 and V_0 are orthogonal by construction. Therefore,

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

By repeating the Fourier series argument, as in (1.2.19), we conclude

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

By taking into account the definitions of m_0 and m_1, and by mimicking the derivation of (1.2.22), we find

$$m_1(\omega)\overline{m_0(\omega)} + m_1(\omega+\pi)\overline{m_0(\omega+\pi)} = 0. \tag{1.2.28}$$

From (1.2.28), we conclude that there exists a function $\lambda(\omega)$ such that

$$(m_1(\omega),\ m_1(\omega+\pi)\ ) = \lambda(\omega)\left(\overline{m_0(\omega+\pi)},\ -\overline{m_0(\omega)}\ \right). \tag{1.2.29}$$

By substituting $\xi = \omega+\pi$ and by using the 2π-periodicity of m_0 and m_1, we conclude that

$$\begin{aligned} \lambda(\omega) &= -\lambda(\omega+\pi),\ \text{and} \\ \lambda(\omega) &\quad \text{is } 2\pi\text{-periodic.} \end{aligned} \tag{1.2.30}$$

Any function $\lambda(\omega)$ of the form $e^{\pm i\omega}S(2\omega)$, where S is an $\mathbb{L}_2([0,2\pi])$, 2π-periodic function, will satisfy (1.2.28); however, only the functions for which $|\lambda(\omega)| = 1$ will define an orthogonal basis ψ_{jk} of $\mathbb{L}_2(\mathbb{R})$.

To summarize, we choose $\lambda(\omega)$ such that

(i) $\lambda(\omega)$ is 2π-periodic,

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

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

Standard choices for $\lambda(\omega)$ are $-e^{-i\omega}$, $e^{-i\omega}$, and $e^{i\omega}$; however, any other function satisfying (i)-(iii) will generate a valid m_1. We choose to define $m_1(\omega)$ as

$$m_1(\omega) = -e^{-i\omega}\overline{m_0(\omega+\pi)}. \tag{1.2.31}$$

since it leads to a convenient and standard connection between the filters $\mathbf{h}$ and $\mathbf{g}$.

The form of m_1 and the equation (1.2.19) imply that $\{\psi(\bullet - k),\ k \in \mathbb{Z}\}$ is an orthonormal basis for W_0. Since $|m_1(\omega)| = |m_0(\omega+\pi)|$, the orthogonality condition (1.2.22) can be rewritten as

$$|m_0(\omega)|^2 + |m_1(\omega)|^2 = 1. \tag{1.2.32}$$

By comparing the definition of m_1 in (1.2.26) with

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

we relate g_n and h_n as

$$g_n = (-1)^n\ h_{1-n}. \tag{1.2.33}$$

In signal processing literature, the relation (1.2.33) is known as the *quadrature mirror relation* and the filters $\mathbf{h}$ and $\mathbf{g}$ as *quadrature mirror filters.*

Remark. Choosing $\lambda(\omega) = e^{i\omega}$ leads to the rarely used high-pass filter $g_n = (-1)^{n-1}$ h_{-1-n}. It is sometimes convenient to define g_n as $(-1)^n h_{1-n+M}$, where M is a "shift constant." Such re-indexing of $\mathbf{g}$ affects only the shift-location of the wavelet function.

1.2.2 Haar Wavelets

In addition to their simplicity and formidable applicability, Haar wavelets have tremendous educational value. Here we illustrate some of the relations discussed in the Section 1.2.1 using the Haar wavelet. We start with $\phi(x) = \mathbf{1}(0 \le x \le 1)$ and pretend that everything else is unknown.

The scaling equation (1.2.10) is very simple for the Haar case. By inspection of simple graphs of two scaled Haar wavelets $\phi(2x)$ and $\phi(2x+1)$ stuck to each other, we conclude that the scaling equation is

$$\begin{aligned}\phi(x) &= \phi(2x) + \phi(2x-1) \\ &= \frac{1}{\sqrt{2}}\sqrt{2}\phi(2x) + \frac{1}{\sqrt{2}}\sqrt{2}\phi(2x-1),\end{aligned} \tag{1.2.34}$$

which yields the wavelet filter coefficients:

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

Now, 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

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

Notice that $m_0(\omega) = |m_0(\omega)|e^{i\varphi(\omega)} = \cos\frac{\omega}{2}\cdot e^{-i\omega/2}$ (after $\cos x = \frac{e^{ix}+e^{-ix}}{2}$). Since $\varphi(\omega) = -\frac{\omega}{2}$, Haar's wavelet has *linear phase,* i.e., the scaling function is symmetric in the time domain. The orthogonality condition $|m_0(\omega)|^2 + |m_1(\omega)|^2 = 1$ is easily verified, as well. Relation (1.2.27) 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 by applying the inverse Fourier transformation we obtain

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

in the time-domain. Therefore we "discovered" the Haar wavelet function ψ. From the expression for m_1 or by inspecting the representation of $\psi(x)$ by $\phi(2x)$ and $\phi(2x-1)$, we "conclude" that $g_0 = -g_{-1} = \frac{1}{\sqrt{2}}$.

The Haar basis is not an appropriate basis for all applications for several reasons. The building blocks in Haar's decomposition are discontinuous functions that obviously are not effective in approximating smooth functions. Although the Haar wavelets are well localized in the time domain, in the frequency domain they decay at the slow rate of $O(\frac{1}{n})$.

1.2.3 Daubechies' Compactly Supported Wavelets

Daubechies was first to construct compactly supported orthogonal wavelets with a preassigned degree of smoothness. Here we present the idea of Daubechies, omitting some technical details. Detailed treatment of this topic can be found in the monograph Daubechies [27], Chapters 6 and 7.

Suppose that ψ has N (≥ 2) vanishing moments, i.e., $\int x^n \psi(x)dx = 0$, $n = 0, 1, \ldots, N-1$. Then by Theorem 1.2.3, $m_0(\omega)$ has the form:

$$m_0(\omega) = \left(\frac{1+e^{-i\omega}}{2}\right)^N \mathcal{L}(\omega), \tag{1.2.35}$$

where $\mathcal{L}(\omega)$ is a trigonometric polynomial. When

$$M_0(\omega) = |m_0(\omega)|^2 = \left(\cos^2\frac{\omega}{2}\right)^N \cdot |\mathcal{L}(\omega)|^2,$$

the orthogonality condition (1.2.22) becomes

$$M_0(\omega) + M_0(\omega+\pi) = 1. \tag{1.2.36}$$

$|\mathcal{L}(\omega)|^2$ is a polynomial in $\cos\omega$. It can be re-expressed as a polynomial in $\sin^2\frac{\omega}{2}$ since $\cos\omega = 1 - 2\sin^2\frac{\omega}{2}$. Denote this polynomial by $P(\sin^2\frac{\omega}{2})$. In terms of P the orthogonality condition (1.2.36) becomes

$$(1-y)^N P(y) + y^N P(1-y) = 1, \quad (y = \sin^2\frac{\omega}{2}). \tag{1.2.37}$$

By Bezout's result (outlined below), there exists a unique solution of the functional equation (1.2.37). It can be found by the Euclidean algorithm since the polynomials $(1-y)^N$ and y^N are relatively prime.

Lemma 1.2.1. *(Bezout) If p_1 and p_2 are two polynomials of degree n_1 and n_2, respectively, with no common zeroes, then there exist unique polynomials q_1 and q_2 of degree $n_2 - 1$ and $n_1 - 1$, respectively, so that*

$$p_1(x)q_1(x) + p_2(x)q_2(x) = 1.$$

For the proof of the lemma, we direct the reader to Daubechies ([27], 169-170). The unique solution of (1.2.37) with degree $deg(P(y)) \leq N - 1$ is

$$\sum_{k=0}^{N-1} \binom{N+k-1}{k} y^k, \quad y = \sin^2 \frac{\omega}{2}, \tag{1.2.38}$$

and since it is positive for $y \in [0, 1]$, it does not contradict the positivity of $|\mathcal{L}(\omega)|^2$.

Remark: If the degree of a solution is not required to be minimal then any other polynomial $Q(y) = P(y) + y^N R(\frac{1}{2} - y)$ where R is an odd polynomial preserving the positivity of Q, will lead to a different solution for $m_0(\omega)$. By choosing $R \neq 0$, one can generalize the standard Daubechies family, to construct symmlets, complex Daubechies wavelets, coiflets, etc.

The function $|m_0(\omega)|^2$ is now completely determined. To finish the construction we have to find its square root. A result of Riesz, known as the *spectral factorization lemma*, makes this possible.

Lemma 1.2.2. *(Riesz) Let A be a positive trigonometric polynomial with the property $A(-x) = A(x)$. Then, A is necessarily of the form*

$$A(x) = \sum_{m=1}^{M} u_m \cos mx.$$

In addition, there exists a polynomial B of the same order $B(x) = \sum_{m=1}^{M} v_m e^{imx}$ such that $|B(x)|^2 = A(x)$. If the coefficients u_m are real, then B can be chosen so that the coefficients v_m are also real.

We first represent $|\mathcal{L}(\omega)|^2$ as the polynomial

$$\frac{a_0}{2} + \sum_{k=1}^{N-1} a_k \cos^k \omega,$$

by replacing $\sin^2\frac{\omega}{2}$ in (1.2.38) by $\frac{1-\cos\omega}{2}$. An auxiliary polynomial P_A, such that $|\mathcal{L}(e^{-i\omega})|^2 = |P_A(e^{-i\omega})|$, is formed.

If $z = e^{-i\omega}$, then $\cos\omega = \frac{z+z^{-1}}{2}$ and one such auxiliary polynomial is

$$P_A(z) = \frac{1}{2}\sum_{k=1-N}^{N-1} a_{|k|} z^{N-1+k}. \tag{1.2.39}$$

Since $P_A(z) = z^{2N-2}P_A(\frac{1}{z})$, the zeroes of $P_A(z)$ appear in reciprocal pairs if real, and quadruples $(z_i, \bar{z}_i, z_i^{-1}, \bar{z}_i^{-1})$ if complex. Without loss of generality we assume that $z_j, \bar{z}_j$ and r_j lie outside the unit circle in the complex plane. Of course, then $z_j^{-1}, \bar{z}_j^{-1}$ and r_j^{-1} lie inside the unit circle. The factorized polynomial P_A can be written as

$$\begin{aligned} P_A(z) &= \frac{1}{2}a_{N-1}\left[\prod_{i=1}^{I}(z-r_i)(z-\frac{1}{r_i})\right] \\ &\quad \left[\prod_{j=1}^{J}(z-z_j)(z-\bar{z}_j)(z-z_j^{-1})(z-\bar{z}_j^{-1})\right]. \end{aligned} \tag{1.2.40}$$

Here $r_1, r_2, \ldots, r_I$ are real and non zero, and $z_1, \ldots, z_J$ are complex; $I + 2J = N - 1$.

The goal is to take a square root from $|P_A(z)|$ and the following simple substitution puts $|P_A(z)|$ in a convenient form. Since $z = e^{-i\omega}$, we replace $|(z-z_j)(z-\bar{z}_j^{-1})|$ by $|z_j|^{-1}|z-z_j|^2$, and the polynomial $|P_A|$ becomes

$$\frac{1}{2}|a_{N-1}|\prod_{i=1}^{I}|r_i^{-1}|\prod_{j=1}^{J}|z_j|^{-2}\cdot|\prod_{i=1}^{I}(z-r_i)\prod_{j=1}^{J}(z-z_j)(z-\bar{z}_j)|^2.$$

Now, $\mathcal{L}(\omega)$ becomes

$$\begin{aligned} &\pm(\frac{1}{2}|a_{N-1}|\prod_{i=1}^{I}|r_i^{-1}|\prod_{j=1}^{J}|z_j|^{-2})^{\frac{1}{2}} \\ &\quad\cdot|\prod_{i=1}^{I}(z-r_i)\prod_{j=1}^{J}(z-z_j)(z-\bar{z}_j)|,\ z = e^{-i\omega}, \end{aligned} \tag{1.2.41}$$

where the sign is chosen so that $m_0(0) = \mathcal{L}(0) = 1$. Note that $deg[P_A(z)] = deg[|\mathcal{L}(z)|^2]$ $= N - 1$. Finally, the coefficients $h_0, h_1, \ldots, h_{2N-1}$ in the polynomial $\sqrt{2}\ m_0(\omega)$ are the desired wavelet filter coefficients.

Example 1.2.3. We will find m_0 for $N = 2$.

$|\mathcal{L}(\omega)|^2 = \sum_{k=0}^{2-1} \binom{2+k-1}{k} \sin^2 \frac{k\omega}{2} = 1 + 2\frac{1-\cos\omega}{2} = \frac{1}{2}4 - 1 \cdot \cos\omega$ gives $a_0 = 4$ and $a_1 = -1$.

The auxiliary polynomial P_A is

$$\begin{aligned} P_A(z) &= \frac{1}{2}\sum_{k=-1}^{1} a_{|k|} z^{1+k} \\ &= \frac{1}{2}(-1 + 4z - z^2) \\ &= -\frac{1}{2}\left(z - (2+\sqrt{3})\right)\left(z - (2-\sqrt{3})\right). \end{aligned}$$

One square root from the above polynomial is

$$\begin{aligned} \sqrt{\frac{1}{2}\,(|-1|)\,\frac{1}{2+\sqrt{3}}}\,\left(z - (2+\sqrt{3})\right) &= \frac{1}{\sqrt{2}}\sqrt{2-\sqrt{3}}\,\left(z - (2+\sqrt{3})\right) \\ &= \frac{1}{2}\left((\sqrt{3}-1)z - (1+\sqrt{3})\right). \end{aligned}$$

The change in sign in the expression above is necessary, since the expression should have the value of 1 at $z = 1$ or equivalently at $\omega = 0$. Finally,

$$\begin{aligned} m_0(\omega) &= \left(\frac{1+e^{-i\omega}}{2}\right)^2 \frac{1}{2}\left((1-\sqrt{3})e^{-i\omega} + (1+\sqrt{3})\right) \\ &= \frac{1}{\sqrt{2}}\left(\frac{1+\sqrt{3}}{4\sqrt{2}} + \frac{3+\sqrt{3}}{4\sqrt{2}}e^{-i\omega} + \frac{3-\sqrt{3}}{4\sqrt{2}}e^{-2i\omega} + \frac{1-\sqrt{3}}{4\sqrt{2}}e^{-3i\omega}\right). \end{aligned}$$

Table 1 gives **h**-filters for DAUB2 - DAUB10 wavelets.

1.2.4 Regularity of Wavelets

There is at least continuum many different wavelet bases. An appealing property of wavelets is diversity in their properties. One can construct wavelets with different smoothness, symmetry, oscillatory, support, etc. properties. Sometimes the requirements can be conflicting since some of the properties are exclusive. For example, there is no symmetric real-valued wavelet with a compact support. Similarly, there is no $\mathbb{C}^\infty$-wavelet function with an exponential decay, etc.

Table 1: The **h** filters for 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

Scaling functions and wavelets can be constructed with desired degree of smoothness. The regularity (smoothness) of wavelets is connected with the rate of decay of scaling functions and ultimately with the number of vanishing moments of scaling and wavelet functions. For instance, the Haar wavelet has only the "zeroth" vanishing moment (as a consequence of the admissibility condition) resulting in a discontinuous wavelet function.

Theorem 1.2.3 is important in connecting the regularity of wavelets, the number of vanishing moments, and the form of the transfer function $m_0(\omega)$. The proof is based on the Taylor series argument and the scaling properties of wavelet functions. For details, see Daubechies [27], pp 153–155. Let

$$\mathcal{M}_k = \int x^k \phi(x) dx \quad \text{and} \quad \mathcal{N}_k = \int x^k \psi(x) dx,$$

be the kth moments of the scaling and wavelet functions, respectively.

Theorem 1.2.3. *Let $\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k),\ j,k \in \mathbb{Z}$ be an orthonormal system of functions in $\mathbb{L}_2(\mathbb{R})$,*

$$|\psi(x)| \leq \frac{C_1}{(1+|x|)^\alpha},\ \alpha > N,$$

and $\psi \in \mathbb{C}^{N-1}(\mathbb{R})$, where the derivatives $\psi^{(k)}(x)$ are bounded for $k \leq N-1$.

Then, ψ has N vanishing moments,

$$\mathcal{N}_k = 0, \quad 0 \leq k \leq N-1.$$

If, in addition,

$$|\phi(x)| \leq \frac{C_2}{(1+|x|)^\alpha},\ \alpha > N$$

then, the associated function $m_0(\omega)$ is necessarily of the form

$$m_0(\omega) = \left(\frac{1+e^{-i\omega}}{2}\right)^N \cdot \mathcal{L}(\omega), \tag{1.2.42}$$

where $\mathcal{L}$ is a 2π-periodic, $\mathbb{C}^{N-1}$-function.

The following definition of regularity is often used,

Definition 1.2.1. The multiresolution analysis (or, the scaling function) is said to be r-regular if, for any $\alpha \in \mathbb{Z}$,

$$|\phi^{(k)}(x)| \leq \frac{C}{(1+|x|)^{\alpha}},$$

for $k = 0, 1, \ldots, r$.

The requirement that ψ possesses N vanishing moments can be expressed in terms of Ψ, m_0, or equivalently, in terms of the filter $\mathbf{h}$.

Assume that a wavelet function $\psi(x)$ has N vanishing moments, i.e.,

$$\mathcal{N}_k = 0, \ k = 0, 1, \ldots, N-1. \tag{1.2.43}$$

By basic property of Fourier transformations, the requirement (1.2.43) corresponds to

$$\left.\frac{d^k \Psi(\omega)}{d\omega^k}\right|_{\omega=0} = 0, \ k = 0, 1, \ldots, N-1,$$

which implies

$$m_1^{(k)}(\omega)\,|_{\omega=0} = m_1^{(k)}(0) = 0, \ k = 0, 1, \ldots, N-1. \tag{1.2.44}$$

It is easy to check that in terms of m_0, relation (1.2.44) becomes

$$m_0^{(k)}(\omega)\,|_{\omega=\pi} = m_0^{(k)}(\pi) = 0, \ k = 0, 1, \ldots, N-1. \tag{1.2.45}$$

The argument is inductive. The case $k = 0$ follows from $\Psi(0) = m_1(0)\Phi(0)$ [(1.2.27) evaluated at $\omega = 0$] and the fact that $\Phi(0) = 1$. Since $\Psi'(0) = \frac{1}{2}m_1'(0)\Psi(0) + \frac{1}{2}m_1(0)\Psi'(0)$ it follows that $m_1'(0) = 0$, as well. Then, $m_1^{(N-1)}(0) = 0$ follows by induction.

The condition $m_1^{(k)}(0) = 0, \ k = 0, 1, \ldots, N-1$ translates to a constraint on the wavelet-filter coefficients

$$\sum_{n\in\mathbb{Z}} n^k g_n = \sum_{n\in\mathbb{Z}} (-1)^n n^k h_n = 0, \quad k = 0, 1, \ldots, N-1. \tag{1.2.46}$$

How smooth are the wavelets from the Daubechies family? There is an apparent trade-off between the length of support and the regularity index of scaling functions. Daubechies [24] and Daubechies and Lagarias [25, 26], obtained regularity exponents for wavelets in the Daubechies family.

Let ϕ be the DAUBN scaling function. There are two popular measures of regularity of ϕ: Sobolev and Hölder regularity exponents. Let α_N^* be the supremum of β such that

$$\int (1+|\omega|)^{\beta} |\Phi(\omega)| d\omega < \infty,$$

and let α_N be the exponent of the Hölder space $\mathbb{C}^{\alpha_N}$ to which the scaling function ϕ belongs.

Table 2: Sobolev α_N^* and Hölder α_N regularity exponents of Daubechies' scaling functions.

N	1	2	3	4	5	6	7	8	9	10
α_N^*	0.5	1	1.415	1.775	2.096	2.388	2.658	2.914	3.161	3.402
α_N		0.550	0.915	1.275	1.596	1.888	2.158	2.415	2.661	2.902

The following result describes the limiting behavior of α_N.

Theorem 1.2.4.

$$\lim_{N\to\infty} \alpha_N = N\left(1 - \frac{\log 3}{2\log 2}\right) + O(\frac{\ln N}{N}).$$

From Table 2, we see that DAUB4 is the first differentiable wavelet, since $\alpha > 1$. More precise bounds on α_N yield that ϕ from the DAUB3 family is, in fact, the first differentiable scaling function ($\alpha_3 = 1.0878$), even though it seams to have a peak at 1. See also Daubechies [27], page 239, for the discussion.

Remark. the Sobolev and Hölder regularities are related, thus, Theorem 1.2.4 holds for the exponent α_N^*, as well.

1.2.4.1 Moment Conditions Determine Filters

We saw that the requirement that the wavelet function possesses N-vanishing moments was expressed in terms of Φ, m_0, or $\mathbf{h}$.

Suppose that we wish to design a wavelet filter $\mathbf{h} = \{h_0, \ldots, h_{2N-1}\}$ only by considering properties of its filter taps. Assume that

$$\mathcal{N}_k = \int_{\mathbb{R}} x^k \psi(x) dx = 0, \text{ for } k = 0, 1, \ldots, N-1. \tag{1.2.47}$$

As it was discussed in Section 1.2.1, some relevant properties of a multiresolution analysis can be expressed as relations involving coefficients of the filter $\mathbf{h}$. For example, the normalization property gave

$$\sum_{i=0}^{2N-1} h_i = \sqrt{2},$$

the requirement for vanishing moments on ψ led to

$$\sum_{i=0}^{2N-1} (-1)^i i^k h_i = 0, \ k = 0, 1, \ldots, N-1,$$

and, finally, the orthogonality property reflected to

$$\sum_{i=0}^{2N-1} h_i h_{i+2k} = \delta_k, \ k = 0, 1, \ldots, N-1.$$

That defines $2N+1$ equations with $2N$ unknowns; however the system is solvable since the equations are not linearly independent. For example, the equation

$$h_0 - h_1 + h_2 - \cdots - h_{2N-1} = 0,$$

can be expressed as a linear combination of the others.

Example 1.2.4. For $N = 2$, we obtain the system:

$$\begin{cases} h_0 + h_1 + h_2 + h_3 = \sqrt{2} \\ h_0^2 + h_1^2 + h_2^2 + h_3^2 = 1 \\ -h_1 + 2h_2 - 3h_3 = 0 \\ h_0 \ h_2 + h_1 \ h_3 = 0 \end{cases},$$

which has the familiar solution $h_0 = \frac{1+\sqrt{3}}{4\sqrt{2}}, h_1 = \frac{3+\sqrt{3}}{4\sqrt{2}}, h_2 = \frac{3-\sqrt{3}}{4\sqrt{2}}$, and $h_3 = \frac{1-\sqrt{3}}{4\sqrt{2}}$.

For $N = 4$, the system is

$$\begin{cases} h_0 + h_1 + h_2 + h_3 + h_4 + h_5 + h_6 + h_7 = \sqrt{2} \\ h_0^2 + h_1^2 + h_2^2 + h_3^2 + h_4^2 + h_5^2 + h_6^2 + h_7^2 = 1 \\ h_0 - h_1 + h_2 - h_3 + h_4 - h_5 + h_6 - h_7 = 0 \\ h_0h_2 + h_1h_3 + h_2h_4 + h_3h_5 + h_4h_6 + h_5h_7 = 0 \\ h_0h_4 + h_1h_5 + h_2h_6 + h_3h_7 = 0 \\ h_0h_6 + h_1h_7 = 0 \\ 0h_0 - 1h_1 + 2h_2 - 3h_3 + 4h_4 - 5h_5 + 6h_6 - 7h_7 = 0 \\ 0h_0 - 1h_1 + 4h_2 - 9h_3 + 16h_4 - 25h_5 + 36h_6 - 49h_7 = 0 \\ 0h_0 - 1h_1 + 8h_2 - 27h_3 + 64h_4 - 125h_5 + 216h_6 - 343h_7 = 0. \end{cases}$$

The above systems can easily be solved by a symbolic software package such as `Maple` or `Mathematica`.

1.2.5 Discrete Wavelet Transformations

Discrete wavelet transformations (DWT) are applied to the discrete data sets to produce discrete outputs. Transforming signals and data vectors by DWT is a process that resembles the fast Fourier transformation (FFT), the Fourier method applied to a set of discrete measurements.

Table 3: 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

Discrete wavelet transformations map data from the time domain (the original or input data, signal vector) to the wavelet domain. The result is a vector of the same size. Wavelet transformations are linear and they can be defined by matrices of dimension $n \times n$ if they are applied to inputs of size n. Depending on boundary

conditions, such matrices can be either orthogonal or "close" to orthogonal. When the matrix is orthogonal, the corresponding transformation is a rotation in $\mathbb{R}^n$ space in which the signal vectors represent coordinates of a single point. The coordinates of the point in the new, rotated space comprise the discrete wavelet transformation of the original coordinates.

Example 1.2.5. Let the vector be $\{1, 2\}$ and let $M(1, 2)$ be the point in $\mathbb{R}^2$ with coordinates given by the data vector. The rotation of the coordinate axes by an angle of $\frac{\pi}{4}$ can be interpreted as a DWT in the Haar wavelet basis. 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 $W \cdot (1, 2)' = (\frac{3}{\sqrt{2}}, -\frac{1}{\sqrt{2}})'$. Notice that *the energy* (squared distance of the point from the origin) is preserved, $1^2 + 2^2 = (\frac{1}{2})^2 + (\frac{\sqrt{3}}{2})^2$, since W is a rotation.

Example 1.2.6. Let $\mathbf{y} = (1, 0, -3, 2, 1, 0, 1, 2)$. If Haar wavelet is used, the values $f(n) = y_n,\ n = 0, 1, \ldots, 7$ are interpolated by the father wavelet, the vector represent the sampled piecewise constant function. It is obvious that such defined f belongs to Haar's multiresolution space V_0.

The following matrix equation gives the connection between $\mathbf{y}$ and the wavelet coefficients (data in the wavelet domain).

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

The solution is

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

Thus,

$$\begin{aligned} f &= \sqrt{2}\phi_{-3,0} - \sqrt{2}\psi_{-3,0} + \psi_{-2,0} - \psi_{-2,1} \\ &\quad + \frac{1}{\sqrt{2}}\psi_{-1,0} - \frac{5}{\sqrt{2}}\psi_{-1,1} + \frac{1}{\sqrt{2}}\psi_{-1,2} - \frac{1}{\sqrt{2}}\psi_{-1,3}. \end{aligned} \tag{1.2.48}$$

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

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

Performing wavelet transformations by multiplying the input vector with an appropriate orthogonal matrix is conceptually straightforward, but of limited practical value. Storing and manipulating transformation matrices when inputs are long

(> 2000) may not even be feasible.

In the context of image processing, Burt and Adelson [15] developed orthogonal and biorthogonal pyramid algorithms. Pyramid or cascade procedures process an image at different scales, ranging from fine to coarse, in a tree-like algorithm. The images can be denoised, enhanced or compressed by appropriate scale-wise treatments.

Mallat [55] was the first to link wavelets, multiresolution analyses and cascade algorithms in a formal way. Mallat's cascade algorithm gives a constructive and efficient recipe for performing the discrete wavelet transformation. It relates the wavelet coefficients from different levels in the transformation by filtering with $\mathbf{h}$ and $\mathbf{g}$. Mallat's algorithm can be viewed as a wavelet counterpart of Danielson-Lanczos algorithm in fast Fourier transformations.

It is convenient to link the original signal with the space coefficients from the space V_J, for some J. Such link is exact for interpolating wavelets (Haar, Shannon, some biorthogonal and halfband-filter wavelets) and close to exact for other wavelets, notably coiflets. Then, coarser smooth and complementing detail spaces are (V_{J-1}, W_{J-1}), (V_{J-2}, W_{J-2}), etc. Decreasing the index in V-spaces is equivalent to coarsening the approximation to the data.

By a straightforward substitution of indices in the scaling equations (1.2.10) and (1.2.25), one obtains

$$\phi_{j-1,l}(x) = \sum_{k\in\mathbb{Z}} h_{k-2l}\phi_{jk}(x) \quad \text{and} \quad \psi_{j-1,l}(x) = \sum_{k\in\mathbb{Z}} g_{k-2l}\phi_{jk}(x). \tag{1.2.49}$$

The relations in (1.2.49) are fundamental in developing the cascade algorithm.

Consider a multiresolution analysis $\cdots \subset V_{j-1} \subset V_j \subset V_{j+1} \subset \ldots$. 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 customary to denote the coefficients associated with $\phi_{jk}(x)$ and $\psi_{jk}(x)$ by c_{jk} and d_{jk}, respectively.

Thus,

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

By using the general scaling equations (1.2.49), orthogonality of $w_{j-1}(x)$ and $\phi_{j-1,l}(x)$ for any j and l, and additivity of inner products, we obtain

$$\begin{aligned} c_{j-1,l} &= \langle v_j, \phi_{j-1,l}\rangle \\ &= \langle v_j, \sum_k h_{k-2l}\phi_{j,k}\rangle \\ &= \sum_k h_{k-2l}\langle v_j, \phi_{j,k}\rangle \\ &= \sum_k h_{k-2l}c_{j,k}. \end{aligned} \tag{1.2.50}$$

Similarly $d_{j-1,l} = \sum_k g_{k-2l}c_{j,k}$. The cascade algorithm works in the reverse direction as well. Coefficients in the next finer scale corresponding to V_j can be obtained from the coefficients corresponding to V_{j-1} and W_{j-1}. The relation

$$\begin{aligned} c_{j,k} &= \langle v_j, \phi_{j,k}\rangle \\ &= \sum_l c_{j-1,l}\langle \phi_{j-1,l}, \phi_{j,k}\rangle + \sum_l d_{j-1,l}\langle \psi_{j-1,l}, \phi_{j,k}\rangle \\ &= \sum_l c_{j-1,l}h_{k-2l} + \sum_l d_{j-1,l}g_{k-2l}, \end{aligned} \tag{1.2.51}$$

describes a single step in the reconstruction algorithm.

Example 1.2.7. For DAUB2, the scaling equation at integers is

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

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

Since $\phi(0) = \sqrt{2}h_0\phi(0)$ and $\sqrt{2}h_0 \neq 1$, it follows that $\phi(0) = 0$. Also, $\phi(3) = 0$. For $\phi(1)$ and $\phi(2)$ we obtain the system

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

From $\sum_k \phi(x-k) = 1$ it follows that $\phi(1) + \phi(2) = 1$. Solving for $\phi(1)$ and $\phi(2)$ we 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,

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

etc.

1.2.5.1 Discrete Wavelet Transformations as Linear Transformations

The change of basis in V_1 from $\mathcal{B}_1 = \{\phi_{1k}(x), k \in Z\}$ to $\mathcal{B}_2 = \{\phi_{0k}, k \in Z\} \cup \{\psi_{0k}, k \in Z\}$ can be performed by matrix multiplication, therefore, it is possible to define

discrete wavelet transformation by matrices. We have already seen a transformation matrix corresponding to Haar's inverse transformation in Example 1.2.6.

Let the length of the input signal be 2^J, and let $\mathbf{h} = \{h_s,\ s \in \mathbb{Z}\}$ be the wavelet filter and let N be an appropriately chosen constant. Denote by H_k is a matrix of size $(2^{J-k} \times 2^{J-k+1}),\ k = 1, \ldots$ with entries

$$h_s, \quad s = (N-1) + (i-1) - 2(j-1) \text{ modulo } 2^{J-k+1}, \tag{1.2.52}$$

at the position (i, j). Note that H_k is a circulant matrix, its ith row is 1st row circularly shifted to the right by $2(i-1)$ units. This circularity is a consequence of using the *modulo* operator in (1.2.52).

By analogy, define a matrix G_k by using the filter $\mathbf{g}$. A version of G_k corresponding to the already defined H_k can be obtained by changing h_i by $(-1)^i h_{N+1-i}$. The constant N is a shift parameter and affects the position of the wavelet on the time scale. For filters from the Daubechies family, standard choice for N is the number of vanishing moments. See also Remark 1.2.1.

The matrix $\begin{bmatrix} H_k \\ G_k \end{bmatrix}$ is a basis-change matrix in 2^{J-k+1} dimensional space; consequently, it is unitary. 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 1.2.8. Suppose that $\mathbf{y} = \{1, 0, -3, 2, 1, 0, 1, 2\}$ 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,

$$\begin{aligned} H_1 \cdot \mathbf{y} &= \{2.19067, -2.19067, 1.67303, 1.15539\} \\ G_1 \cdot \mathbf{y} &= \{0.96593, 1.86250, -0.96593, 0.96593\}. \end{aligned}$$

$W_1\mathbf{y} = \{2.19067, -2.19067, 1.67303, 1.15539 \mid 0.96593, 1.86250, -0.96593, 0.96593\}$.

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

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

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

$W_2 \cdot \mathbf{y} = \{1.68301, 0.31699 \mid -3.28109, -0.18301 \mid 0.96593, 1.86250, -0.96593, 0.96593\}$, because

$$\begin{aligned} H_2 \cdot H_1 \cdot \mathbf{y} &= H_2 \cdot \{2.19067, -2.19067, 1.67303, 1.15539\} \\ &= \{1.68301, 0.31699\} \\ G_2 \cdot H_1 \cdot \mathbf{y} &= G_1 \cdot \{2.19067, -2.19067, 1.67303, 1.15539\} \\ &= \{-3.28109, -0.18301\}. \end{aligned}$$

For quadrature mirror wavelet filters $\mathbf{h}$ and $\mathbf{g}$, we define recursively up-sampled filters $\mathbf{h}^{[r]}$ and $\mathbf{g}^{[r]}$

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

In practice, the dilated filter $\mathbf{h}^{[r]}$ is obtained by inserting zeroes between the taps in $\mathbf{h}^{[r-1]}$. Let $\mathbf{H}^{[r]}$ and $\mathbf{G}^{[r]}$ be convolution operators with filters $\mathbf{h}^{[r]}$ and $\mathbf{h}^{[r]}$, respectively. A non-decimated wavelet transformation, NDWT, is defined as a sequential application of operators (convolutions) $\mathbf{H}^{[j]}$ and $\mathbf{G}^{[j]}$ on a given time series.

Definition 1.2.2. *Let* $\mathbf{a}^{(J)} = \mathbf{c}^{(J)}$ *and*

$$\begin{aligned} \mathbf{a}^{(j-1)} &= \mathbf{H}^{[J-j]}\mathbf{a}^{(j)}, \\ \mathbf{b}^{(j-1)} &= \mathbf{G}^{[J-j]}\mathbf{a}^{(j)}. \end{aligned}$$

The non-decimated wavelet transformation of $\mathbf{c}^{(J)}$ *is* $\mathbf{b}^{(J-1)}$, $\mathbf{b}^{(J-2)}$, $\ldots$, $\mathbf{b}^{(J-j)}$, $\mathbf{a}^{(J-j)}$, *for some* $j \in \{1, 2, \ldots, J\}$ *the depth of the transformation.*

If the length of an input vector $\mathbf{c}^{(J)}$ is 2^J, then for any $0 \leq m < J$, $\mathbf{a}^{(m)}$ and $\mathbf{b}^{(m)}$ are of the same length. Let $\phi_j(x) = \phi_{j,0}(x)$ and $\psi_j(x) = \psi_{j,0}(s)$. If the measurement

sequence $\mathbf{c}^{(J)}$ is associated with the function $f(x) = \sum_k c_k^{(J)} \phi_J(x - 2^{-J}k)$ then the kth coordinate of $\mathbf{b}^{(j)}$ is equal to

$$b_{jk} = \int \psi_j(x - 2^{-J}k) f(x) dx.$$

Thus, the coefficient b_{jk} provides information at scale 2^{J-j} and location k. One can think of a nondecimated wavelet transformation as sampled continuous wavelet transformation $\langle f(x), \frac{1}{\sqrt{a}} \psi\left(\frac{x-b}{a}\right) \rangle$ for $a = 2^{-j}$, and $b = k$.

1.2.6 Wavelets and Self-Similar Processes

We already saw that wavelets are a capable tool in detecting self-similarity in the signals. In this section we discuss some properties of self-similar signals in the wavelet domain – it would be a prelude to part of this handout that talks about inference in the wavelet domain. Some important pioneering work in this area was done by Flandrin and his collaborators [Flandrin 1989a,b, 1992a,b; Flandrin and Gonçlavès, 1993; Abry, Gonçlavès and Flandrin 1993]. The body of recent literature is quite large.

1.2.6.1 Wavelets and Stationary Processes

Wavelets and stationary processes are well researched area. Highlights of the research include several topics: preservation of stationarity, whitening property of wavelets and assessing correlations in the wavelet domain, wavelet shrinkage in the presence of stationary noise, to list a few.

Let $X(t)$ be a second order process with autocorrelation function $\gamma(t, s)$. The discrete wavelet transformation of $X(t)$ is a discrete random field

$$\{d_{jk},\ j, k \in \mathbb{Z}\} = \left\{ \int_{\mathbb{R}} X(t) \psi_{jk}(t)\, dt,\ j, k \in \mathbb{Z} \right\}, \tag{1.2.53}$$

which is well defined if the path integrals in (1.2.53) are defined and

$$\int_{\mathbb{R}} \sqrt{\gamma(t,t)}\ |\psi_{jk}(t)|\, dt < \infty. \tag{1.2.54}$$

Thus, when (1.2.54) is satisfied

$$Ed_{jk}d_{j'k'} = \iint_{\mathbb{R}^2} \gamma(t,s)\psi_{jk}(t)\psi_{j'k'}(s)\,dt\,ds. \tag{1.2.55}$$

If the process $X(t)$ is stationary, then (1.2.55) becomes

$$Ed_{jk}d_{j'k'} = \iint_{\mathbb{R}^2} \gamma(h)\psi_{jk}(s)\psi_{j'k'}(s+h)\,ds\,dh, \tag{1.2.56}$$

or in the Fourier domain,

$$\begin{aligned} Ed_{jk}d_{j'k'} &= \frac{1}{2\pi}\int_{\mathbb{R}} f(\omega)\Psi\left(\frac{\omega}{2^j}\right)\overline{\Psi\left(\frac{\omega}{2^{j'}}\right)} \\ &\quad \cdot e^{-i\omega k 2^{-j}} e^{i\omega k' 2^{-j'}} 2^{-j/2} 2^{-j'/2}\,d\omega, \end{aligned} \tag{1.2.57}$$

where Ψ and f are the Fourier transformations of ψ and γ, respectively.

Relations (1.2.56) and (1.2.57) are critical in assessing the whitening property of wavelet transformations for in case of stationary processes as well as the second order properties of wavelet coefficients of self-similar processes.

1.2.6.2 Approximation and Production of Stationarity

For a stationary process $X(t)$, $t \in \mathbb{R}$ and its wavelet approximation $X_m(t) = \sum_k c_{mk}\phi_{mk}(t)$, holds

$$E|X_m(t) - X(t)|^2 \to 0 \text{ when } m \to \infty, \tag{1.2.58}$$

i.e., degraded process $X_m(t)$ goes in the mean-square to the original.

Indeed, for r-regular wavelet ϕ and a wavelet-based reproducing kernel of V_m, $\mathbb{IK}_m(t,s) = 2^m \sum_k \phi(2^m t - k)\phi(2^m s - k)$,

$$\begin{aligned} E|X(t) - X_m(t)|^2 &= EX(t)^2 - 2EX(t)X_m(t) + EX_m(t)^2 \\ &= \gamma(0) - 2\int \gamma(t-s)\mathbb{IK}_m(t,s)\,ds \\ &+ \iint \gamma(u-s)\mathbb{IK}_m(s,t)\mathbb{IK}_m(t,u)\,du\,ds \end{aligned}$$

and

$$\int \gamma(u-s)\mathbb{K}_m(t,s)\,ds \to \gamma(u-t),$$

uniformly on bounded sets, implying (1.2.58). In general case $X_m(t)$ is not projection of $X(t)$ on V_m since the sample paths may not be $\mathbb{L}_2$ integrable.

It is well known that wavelet transformations of stationary processes and sequences yield level-wise stationary sequences of coefficients. More generally, wavelet transformation of a process with stationary increments yields a stationary sequences of wavelet coefficients in any fixed level j. The following lemma is straightforward.

Lemma 1.2.5. *Let $X(t),\ t \in \mathbb{R}$ be a process with stationary (in strong sense) increments, i.e., for all finite-dimensional vectors $(X(t+h_1)-X(t),\ldots,X(t+h_k)-X(t)) \stackrel{d}{=} (X(h_1)-X(0),\ldots,X(h_k)-X(0))$, for all $t \in \mathbb{R}$. Then for a fixed j, $d_{j,k}$ is a stationary sequence.*

Proof: Select arbitrary $m \in \mathbb{Z}$ and fix j. Then

$$\begin{aligned}
d_{j,k} &= \int_{\mathbb{R}} X(t)\psi_{jk}(t)dt \\
&= \int_{\mathbb{R}} X(t-2^{-j}m)\psi_{jk}(t-2^{-j}m)dt \\
&= \int_{\mathbb{R}} X(t-2^{-j}m)\cdot 2^{j/2}\psi(2^j(t-2^{-j}m)-k)dt \qquad (1.2.59)\\
&= \int_{\mathbb{R}} \left(X(t-2^{-j}m)-X(2^{-j}m)\right)\psi_{j,k+m}(t)dt \quad [\text{ because } \int \psi_{jk}=0] \\
&\stackrel{d}{=} \int_{\mathbb{R}} (X(t)-X(0))\,\psi_{j,k+m}(t)dt \quad [\text{ because increments are stationary}] \\
&= d_{j,k+m}. \quad \square
\end{aligned}$$

Standard arguments probabilistic arguments involving characteristic functions of finite linear combinations of d's (see Abry, Flandrin, Taqqu, and Veitch, [3]) lead to conclusion that finite-dimensional distributions are free of m, implying that the sequence $\{d_{j,k}, k \in \mathbb{Z}\}$ is stationary.

1.2.7 Wavelet Analysis of Scaling Processes

Let $\{X(t), t \in \mathbb{R}\}$ be a H-ss process. Then for a fixed level j,

$$d_{jk} \stackrel{d}{=} 2^{-j(H+1/2)}\ d_{0,k}. \tag{1.2.60}$$

Indeed,

$$\begin{aligned} d_{jk} &= \int X(t) 2^{j/2} \psi(2^j t - k) dt \\ &= \int X(2^{-j}u) 2^{j/2} \psi(u-k) d(2^{-j}u) \\ &\stackrel{d}{=} 2^{-jH-j/2} \int X(u)\psi(u-k)du \\ &= 2^{-j(H+1/2)}\ d_{0,k}. \end{aligned}$$

Note that $1/2$ in the exponent is because of $\mathbb{L}_2$ normalizing of wavelets, for $\mathbb{L}_1$ normalized wavelets $\psi_{j,k} = 2^j\psi(2^j - k)$ the scaling is $(2^{-j})^H$.

Let $X(t)$ be a H-sssi process. Then $d_{0,k}, k \in \mathbb{Z}$ is stationary sequence and for any k,$\mathbb{E}d_{0k} = 0$ and $\mathbb{E}d_{0k}^2 = \mathbb{E}d_{00}^2$. Then,

$$\mathbb{E}d_{jk}^2 = C\ 2^{-j(2H+1)}, \tag{1.2.61}$$

with $C = \mathbb{E}d_{00}^2$. The re-expression of equation (1.2.61) as

$$\log_2 \mathbb{E}d_{jk}^2 = -(2H+1)\cdot j + C', \tag{1.2.62}$$

is a basis for wavelet based estimation of H, as we will see later.

If the process is LRD, i.e., its spectra is behaving as $|\omega|^{-\alpha}$ at 0, then from (1.2.57)

$$\begin{aligned} \mathbb{E}d_{jk}^2 &= \frac{2^{-j}}{2\pi} \int f(\omega) |\Psi(2^{-j}\omega)|^2 d\omega \\ &\sim c \cdot 2^{-\alpha j} \int \frac{|\Psi(\omega)|^2}{|\omega|^\alpha} d\omega, \end{aligned}$$

This relation is a basis for estimating α (or $H = \frac{1+\alpha}{2}$) by taking logarithms of both sides.

If in (1.2.57), $j = j'$, we can explore covariance structure of wavelet coefficients belonging to a single level and separated by distance of $|k - k'|$.

$$\mathbb{E}d_{jk}d_{j'k'} = \frac{1}{2\pi}\int_{\mathbb{R}} f(\omega)\left|\Psi\left(\frac{\omega}{2^j}\right)\right|^2 \cdot e^{-i\omega(k-k')2^{-j}}\, d\omega, \tag{1.2.63}$$

The correlation $\mathbb{E}d_{jk}d_{j'k'}$ is thus a function of the difference $(k - k')$, and the asymptotic behavior of integral in (1.2.63) when $|k-k'| \to \infty$ is influenced by behavior of $f(\omega)\left|\Psi\left(\frac{\omega}{2^j}\right)\right|^2$ as $\omega \to 0$.

As we saw before, if the wavelet has exactly N vanishing moments, then its Fourier transformation is differentiable N times at the origin, and for $0 \leq i \leq N-1$, $\Psi^{(i)}(0) = 0$, and $\Psi^{(N)}(0) \neq 0$. By Taylor series argument, the behavior of $|\Psi(\omega)|$ matches that of $|\omega|^N \cdot |\Psi^{(N)}(0)|$ at origin. On the other end, the spectra of LRD process is singular at zero and behaves as $|\omega|^{-\alpha}$. Thus, when the decomposing wavelet has N vanishing moments, singularity at 0 of spectra is compensated; the behavior at $\omega \sim 0$ of $f(\omega)\left|\Psi\left(\frac{\omega}{2^j}\right)\right|^2$ is $|\omega|^{2N-\alpha}$.

Two comments. (i) If $N > \alpha/2$ then sequence of wavelet coefficients is not LRD, eventhough the original signal is, and (ii) looked as Fourier pair, the autocorrelations $\mathbb{E}d_{jk}d_{j'k'}$ behave as $|k-k'|^{\alpha-N-1}$ when $|k-k'| \to \infty$. This means, the LRD processes are better decorrelated with wavelets possessing more vanishing moments.

1.2.7.1 LogScale Diagrams or Scalograms

The totality of squared wavelet coefficients represents the energy content of the zero-mean signal and expected levelwise energies form wavelet counterpart of Fourier spectra. The following definitions are as in Abry, Flandrin, Taqqu, and Veitch [2, 3].

Definition 1.2.3. Theoretical Wavelet Spectra of a process $X(t)$ with stationary increments is the sequence

$$e(j) = \mathbb{E}(d_{j,\bullet}), \tag{1.2.64}$$

where because of stationarity, $d_{j,\bullet}$ stands for an arbitrary coefficient from the level j. The plot of $(j, e(j))$ is referred as *Theoretical Logscale Diagram*. Let $E(j)$ be an estimator of $e(j)$. Then, the plot of (j, E(j)) is called simply Logscale Diagram or Scalogram.

Because of linearity of wavelet transformation, the wavelet spectra gives complete second-order description of the random process. The logscale diagram was originally proposed and utilized as a tool for detecting and estimating periodicities, since, as in the case of Fourier counterparts, energy is large at dominant scales. In our context, logscale diagram is used to estimate the scaling exponent of a signal.

Definition 1.2.4. Let, for some $q \in \mathbb{R}$

$$e_q(j) = \mathbb{E}|d_{j,\bullet}|^q, \tag{1.2.65}$$

The plot of $(j, e(j))$ is referred as q-th Order Theoretical Logscale Diagram. Let $E_q(j)$ be an estimator of $e_q(j)$. Then, the plot of $(j, E_q(j))$ is called q-th Order Logscale Diagram or Scalogram.

1.3 Overview of the Thesis

This section pertains to the overview of the remainder of this thesis. We introduce background and delineate challenges for multiscale statistical methods in classification of high-frequency data and classification methods in estimation of true signals on multiscale domains.

1.3.1 Multiscale Methodology in Classification of Multifractionality

Fractional Brownian motion (fBm) has been commonly used to characterize a wide range of complex structures in natural phenomena that exhibit self-similarity and long-range dependence. The fractal dimension capturing the regularity of the signal, also called the Hurst exponent, is extended to vary with time to model realistic situations occurring in various fields such as biomechanics, stock market, and turbulence,

to name a few. A stochastic process $W(t)$, called multifractional Brownian motion (mBm), has recently been proposed [59] to model both a long-range dependence and path regularity varying with time.

A Gaussian process $\big(W(t)\big)_{t\geq 0}$ is called multifractional Brownian motion (mBm) with Hurst function $H(t)$ and its scaling factor C if its covariance function is represented as

$$\mathbb{E}[W(t)W(s)] = \frac{C}{2} g(H(t), H(s)) \Big\{ |t|^{H(t)+H(s)} + |s|^{H(t)+H(s)} - |t-s|^{H(t)+H(s)} \Big\} \quad (1.3.1)$$

for $H \in \mathbb{C}^{\eta}\big([0,1]\big)$, $s, t \in [0,1]$, and

$$g\big(H(t), H(s)\big) = \frac{\sqrt{K\big(2H(t)\big) K\big(2H(s)\big)}}{K\big(H(t)+H(s)\big)}. \quad (1.3.2)$$

The process is square-integrable if function $H(t)$ is Hölderian of order $0 < \eta \leq 1$ on $[0,1]$: $H \in \mathbb{C}^{\eta}\big([0,1]\big)$. From (1.3.1), we have $\mathbb{E}[W(t)^2] = Ct^{H(t)}$ and consequently, $Var[W(1)] = C$: in this sense, C is called the variance level of the process in the second chapter.

In addition to the traditional time domain representation of mBm that involves covariance function, [10] proposed a spectral representation of mBm as follows:

$$W(t) = \sqrt{C}\sqrt{K\big(2H(t)\big)/2} \int_{\mathbb{R}} \frac{e^{jt\lambda} - 1}{|\lambda|^{H(t)+1/2}} dB(\lambda), \quad (1.3.3)$$

where C is a constant scale parameter, B is the standard Brownian motion, and K is a function defined by

$$K(\alpha) = \Gamma(\alpha+1) \sin\left(\frac{\alpha\pi}{2}\right)/\pi, \qquad \text{for } 0 < \alpha < 2.$$

In [10], process $W(t)$ was modeled with a piecewise constant Hurst exponent and continuous paths. A local version of quadratic variations (QVs) to estimate the constant Hurst exponent was studied by [46], [50], and [20]. The application of a local version of QVs to estimate $H(t)$ and C in mBm was discussed in [21], who treated

C separately in each sample path. In the second chapter, we extend the QV method to estimate $H(t)$ and C simultaneously for mBm by applying a local k-variation to all sampled paths (all sub-sample paths from a given sample path) and aggregating all local conditions that follow from the previous step. This method includes filtering all sampled paths with a filter possessing a sufficient number of vanishing moments and generating stationary outputs. The method further calculates empirical moments of the filtered signals and then estimates $H(t)$ and C simultaneously in a regression setup specified by the empirical moments.

The main objective of the second chapter is to develop an estimation procedure for unknown parameters $\big(H(t), C\big)$ given a path of $W(t)$ in the presence of independent white noise. The contribution of the second chapter is twofold. One is to provide the stable and simultaneous estimation of variance level C and time-changing Hurst exponent $H(t)$ in a general mBm model with independent white noise. Previous approaches by Coeurjolly and Istas relied on local sample paths that resulted in estimators of C sensitive to the sampled paths. The other contribution is to propose a method that measures the white noise level in the mBm model since it is widely accepted that noise occurs from a variety of sources. One example is the noise of measuring instruments.

1.3.2 Multiscale Methodology in Classification of Multifractality

The third chapter is concerned with assessing the deviation from monofractality in measured high-frequency signals. It has been observed that a wide range of complex structures in nature is characterized by seemingly irregular behavior. Examples of such irregular signals in both time and scale are abundant in medicine, physics, economics, and geosciences, to list a few. Although irregular, such signals can be well modeled by multifractal processes. Concepts of fractal dimension and self-similarity

have been used to quantify the multifractal behavior. The key idea is to quantify statistical similarity of patterns at many different scales. The regularity index describes the strength of the similarity. The scaling is usually stochastically complex and may include inhomogeneity of patterns in both time and scale. Multifractal formalism has been developed in order to quantify the irregular scaling [37, 47].

The essence of multifractal analysis is to assess fractal dimensions of self-similar structures with varying regularities and to produce the distribution of indices of regularity, which constitutes the multifractal spectrum (MFS). The MFS describes the "richness" of local singularities in the signal. The multifractal formalism relates the MFS to the partition function measuring high-order dependencies in the data. In recent years, the multifractal formalism has been implemented with wavelets [6, 76]. This approach is very amenable to computation and estimation in practice. The advantages of using the wavelet-based MFS are availability of fast algorithms for wavelet transform, the locality of wavelet representations in both time and scale, and intrinsic dyadic self-similarity of basis functions.

Rigorous mathematical foundations of the multifractal process and wavelet-based approaches have been studied by several researchers [67, 76]. Many applications to dynamics of the multifractal processes [68, 18, 73], such as TCP/IP traffic data and financial data, can be found. In addition, the wavelet-based fractal analysis is a pervasive concept in the medical fields; many medical images, treated as signals, demonstrate a certain degree of self-similarity over a range of scales, driving the development of data mining algorithms based on fractal analysis of those images. A wavelet transform modulus maxima method combined with a multifractal analysis was used to detect tumors as well as microcalcifications [53]. A classification technique based on features derived from the fractal description of mammograms was used [28]. The wavelet-based multifractal discrimination model was proposed to determine ocular pathology based on the pupillary response behaviors exhibited by older adults

with and without ocular disease during the performance of a computer-based task [70].

The presence of multifractality in real-life signals is difficult to assess due to finite signal sizes and numerical instability of assessing tools. Veitch and Abry [72] used a collection of regularities on blocks of the signal, and Veitch *et al.* [73] reviewed the evidence for multifractal behavior of aggregate TCP traffic using wavelet-based logscale diagram. In most of the approaches for the assessment, level-wise analysis of $\mathbb{L}^p$ norms of wavelet coefficients was utilized. However, the slopes in this scaling behavior could be misleading because multifractal signals may result in a perfect linear decay of energies. Also the slopes are sensitive to the exponent in the partition function. Extraction of meaningful multifractal characteristics for effectively assessing deviation from monofractality based on the MFS has not received much attention in the literature. The main contribution of the third chapter is the development of a test for monofractality of a signal based on relevant multifractal descriptors from the wavelet-based MFS. We demonstrate effectiveness of this test in simulations and real-life examples that include turbulence and DNA nucleotide sequences.

1.3.3 Classification in Multiscale Domains

In recent years statistical wavelet modeling has attracted the attention of both theoretical and applied statisticians. The most important property of wavelets is their adaptive locality in both time and frequency, which helps in dealing with phenomena that change rapidly in both domains.

Shrinkage in the wavelet domain is a simple, yet powerful tool in nonparametric statistical modeling. It utilizes the fact that wavelet transforms are energy-compressing; that is, most of the signal variance is described by only a few wavelet coefficients. Many approaches for wavelet shrinkage are suggested in the literature: shrinkage by Bayes' rules, unbiased estimator of the risk, multiple hypothesis testing,

and cross-validation techniques, to name just a few [5]. In almost all of the cases there is an underlying statistical model on the wavelet coefficients and the shrinkage rule represents the optimal action in the adopted statistical paradigm. Many wavelet shrinkage methods based on those approaches are proper thresholding rules, meaning that the inclusion of a wavelet coefficient in the model takes place if its magnitude exceeds a particular threshold. Perturbations on the threshold level always affect the selection of wavelet coefficients and proper strategies are needed to ensure that the model is neither over-fitted nor under-fitted.

In the fourth chapter, we propose a semi-supervised wavelet shrinkage. For the coefficients whose magnitudes are close to the adopted threshold, we seek additional information to decide if they are going to be retained in the model or not. We put this task in the framework of statistical learning and introduce labeled and unlabeled wavelet coefficients. For the labeled coefficients the membership in the model is decided, while for the unlabeled coefficients the decision is not clear and hence additional information is needed. The unlabeled coefficients are processed under a semi-supervised learning paradigm that incorporates information from the labeled neighboring coefficients.

The semi-supervised learning has become very popular in the area of machine learning. It comprises a wide range of methods aimed to enhance learning from both labeled and unlabeled data and provide better inference (usually in classification and clustering tasks). While labeled data can be expensive or time-consuming to obtain, unlabeled data is usually easy to collect and may carry information useful for the inference, such as in the development of classifiers. In semi-supervised learning, information contained in unlabeled data can be incorporated by a variety of techniques, such as expectation maximization (EM) algorithm, transductive support vector machines (SVMs), graph regularization, and others [78].

We start with two thresholds λ_1 and λ_2, $\lambda_1 \leq \lambda_2$. The labeled coefficients would

have two labels, 0 or 1, depending on whether they are excluded or included in the model. The exclusion is decided by a thresholding rule using the threshold λ_1. The coefficients whose magnitude is less than λ_1 carry the label 0, while the coefficients greater than threshold λ_2 in magnitude carry the label 1. The labels of some wavelet coefficients with their magnitude between λ_1 and λ_2 remain unassigned or undetermined. These will be treated as unlabeled.

By taking the approach of manifold regularization, the decisions for the unlabeled coefficients are based on the neighborhood content of manifold structure. We demonstrate that semi-supervised (SS) shrinkage based on background shrinkages with two different thresholds λ_1 and λ_2 possesses optimal properties of their background shrinkages and improves their performance.

CHAPTER II

MULTISCALE METHODOLOGY IN CLASSIFICATION OF MULTIFRACTIONALITY

We apply multiscale statistical methods to classify and characterize high-frequency data in the model of noised multifractional Brownian motions.

2.1 Local Variations of Multifractional Brownian Motion

Let us consider a case in which a discretized sample path ($\mathbf{W}'$) is given by

$$W'(i/N) = W(i/N) + \sigma\varepsilon(i/N), \qquad i = 1, \ldots, N, \tag{2.1.1}$$

in which W is as in 1.3.1, and $\varepsilon(i/N)$ is independent white noise of level σ. Hurst function $H(t)$ generated by $W(i/N)$ is assumed to be Hölderian function $\mathbb{C}^{\eta}\big([0,1]\big)$, where $0 < \eta \leq 1$. Additionally, noise magnitude σ is assumed to be sufficiently small compared to the variance of mBm. The covariance function of ($\mathbf{W}'$) is

$$\begin{aligned}\mathbb{E}[W'(t)W'(s)] =& \frac{C}{2} g(H(t), H(s)) \Big\{ |t|^{H(t)+H(s)} + |s|^{H(t)+H(s)} - |t-s|^{H(t)+H(s)} \Big\} \\ &+ \sigma^2 \mathbf{1}(t = s),\end{aligned} \tag{2.1.2}$$

where $\mathbf{1}(A)$ is an indicator of relation A. The entries in (2.1.2) generate covariance matrix $\mathbf{\Sigma}$ that depends on unknown parameters $\theta := \big(H(t), C, \sigma\big) \in R^{N+2}$. The covariance matrix consists of $N(N+1)/2$ parameters (due to symmetry) that can be organized into an $N(N+1)/2 \times 1$ vector $\mathbf{\Gamma}(\theta)$. Model (2.1.2) is locally identifiable almost everywhere if Jacobian matrix $\partial\mathbf{\Gamma}(\theta)/\partial\theta'$, which is $N(N+1)/2 \times (N+2)$, has full column rank [69].

In order to weaken the dependence in $W'(t)$ in (2.1.1), a differencing filter $\mathbf{a}$ of length $l+1$ and of order $p > 1$ (the number of vanishing moments) is applied. Filter

$\mathbf{a}$ is defined by its taps, $(a_0, \ldots, a_l)$, such that

$$\begin{aligned} \sum_{q=0}^{l} a_q q^i &= 0, \qquad \text{for } i = 0, \ldots, p-1, \\ \sum_{q=0}^{l} a_q q^i &\neq 0, \qquad \text{otherwise.} \end{aligned} \tag{2.1.3}$$

Let $(\mathbf{V'_a})$ be the process consisting of $(\mathbf{W}')$ filtered by $\mathbf{a}$, that is,

$$V'_{\mathbf{a}}\left(\frac{j}{N}\right) = \sum_{q=0}^{l} a_q W'\left(\frac{j-q}{N}\right), \qquad \text{for } j = l+1, \ldots, N. \tag{2.1.4}$$

For example, when a is $(1, -2, 1)$, the filter is of order 2, and (V'_a) represents the second-order differences of (W'). Furthermore, we can choose a as high-pass wavelet filters corresponding to orthogonal wavelets such as Daubechies and Symlet wavelets. A detailed discussion of wavelets can be found in [27] and [75]. The process $(\mathbf{V_a})$ is defined similarly with W instead of W'. Then, the process $(\mathbf{V'_a})$ is stationary due to the vanishing moment property of filter $\mathbf{a}$. Let $\nu(t)$ be an index set of a neighborhood of t, defined as

$$\nu(t) = \{j \in \mathbb{Z} \mid l < j \leq N, |j/N - t| \leq \epsilon\} \tag{2.1.5}$$

for a parameter $\epsilon > 0$. Let $|\nu(t)|$ be the cardinal number of $\nu(t)$. We set ϵ to be a function of N in such a way that $\epsilon \to 0,\ N\epsilon \to \infty,$ as $N \to \infty$. In other words, for a sufficiently large N, the size of one neighbor becomes sufficiently small while maintaining the summation of the sizes of all the neighbors that are sufficiently large. More precisely, the following specific form for ϵ is suggested:

$$\epsilon \propto N^{-\alpha}(\log(N))^{\beta}, \qquad \text{for } 0 < \alpha < 1,\ \beta \in \mathbb{R}, \tag{2.1.6}$$

which results in $\epsilon^{\eta} \log(N) \to 0$ as $N \to \infty$.

Proposition 2.1.1. *Let $j \in \nu(t)$, $j' \in \nu(t')$. Then, $V'_{\mathbf{a}}(\frac{j}{N})$ in (2.1.4) is weakly stationary, and its covariance $\mathbb{E}\left[V'_{\mathbf{a}}\left(\frac{j}{N}\right) V'_{\mathbf{a}}\left(\frac{j'}{N}\right)\right]$ depends on $j - j'$ as follows:*

$$\mathbb{E}\left[V'_{\mathbf{a}}\left(\frac{j}{N}\right)V'_{\mathbf{a}}\left(\frac{j'}{N}\right)\right] = \frac{Cg\left(H(t),H(t')\right)}{N^{H(t)+H(t')}}\pi_{\mathbf{a},H(t)/2+H(t')/2}(j'-j) + \sigma^2\sum_{q-q'=j-j'}a_q a'_q + \mathcal{O}(\epsilon^{\eta}\log N),$$

where

$$\pi_{\mathbf{a},h}(k) = -\frac{1}{2}\sum_{q,q'=0}^{l} a_q a_{q'}|q-q'+k|^{2h}.$$

This result is similar as Proposition 1 of [21] except that it has a different normalizing factor for the first term and the second term from the white noise in the model.

Let us define the second empirical moment of the filtered signal $V'_{\mathbf{a}}$ as follows:

$$S'(t,\mathbf{a}) = \frac{1}{|\nu(t)|}\sum_{j\in\nu(t)} V'_{\mathbf{a}}\left(\frac{j}{N}\right)^2, \qquad \text{for } t\in[0,1], \tag{2.1.7}$$

which represents the squared energy of the averaged filtered signal in the neighbor of t. Now, define a statistic $V(t,\mathbf{a})$, called the local variation of $(\mathbf{W})$ as:

$$V(t,\mathbf{a}) = \frac{1}{|\nu(t)|}\sum_{j\in\nu(t)}\Big\{\frac{V_{\mathbf{a}}(j/N)^2}{\mathbb{E}[V_{\mathbf{a}}(j/N)^2]} - 1\Big\}. \tag{2.1.8}$$

The statistic $V(t,\mathbf{a})$ can be interpreted as the local $\mathbb{H}_2$-variations of a certain Gaussian process. It captures the amount of deviations of the filtered signal from the standard normal distribution near t. The definition of local k variations is extended using the kth order Hermite polynomial in the summation of (2.1.8): H_2 is the second Hermite polynomial defined by $\mathbb{H}_2(t) = t^2 - 1$. In this chapter, we use local $\mathbb{H}_2$-variations as the minimum asymptotic variance estimators as shown in Coeurjolly [20]; however, one may set k to be different from 2. Notice that $V(t,\mathbf{a})$ and $\nu(t)$ are dependent on N and ϵ. The local variation $V(t,\mathbf{a})$ has many desirable properties as pointed out by [46] and [50]. Next, we connect the local variation $V(t,\mathbf{a})$ with the empirical moment $S'(t,\mathbf{a})$ through the following relationship:

Proposition 2.1.2. *Let local variation $V(t, \mathbf{a})$ and the empirical moment $S'(t, \mathbf{a})$ are defined by (2.1.8) and (2.1.7), respectively, given $W'(t)$ as above and $\mathbf{a}$ of order > 1. Then*

$$\log S'(t, \mathbf{a}) = \log \mathbb{E}\big[S'(t, \mathbf{a})\big] + V(t, \mathbf{a})\big(1 + o(1)\big), \qquad as \;\; N \to \infty. \tag{2.1.9}$$

The proposition connects the empirical moment $S'(t, \mathbf{a})$ and the log of its expectation through the local variation $V(t, \mathbf{a})$. Since the local variation converges to 0 almost surely and its distribution follows normal distribution asymptotically [21], the above relationship can be regarded as a regression setup. In the following section, we discuss distribution of $V(t, \mathbf{a})$ and the relationship between $\log \mathbb{E}\big[S'(t, \mathbf{a})\big]$ and the parameters $(H(t), C, \sigma)$. We also adapt $\mathbf{a}$, so that it can be used to estimate the regression parameters.

2.2 Estimation of Hurst Function and Scaling Factor

Let us introduce $\mathbf{a}^m$, the filter defined by

$$a_i^{(m)} = \begin{cases} a_{i/m}, & i/m \;\text{ is an integer,} \\ 0, & \text{otherwise.} \end{cases} \tag{2.2.1}$$

Notice $\mathbf{a}^m$ is the filter $\mathbf{a}$ dilated m times and captures a resolution with low frequency as m increases. By definition, $\mathbf{a}^1 = \mathbf{a}$. For example, for a second order filter $\mathbf{a} := (1, -2, 1)$, $\mathbf{a}^2$ becomes $(1, 0, -2, 0, 1)$. We are interested in distribution of local variation $V(t, \mathbf{a})$ with $\mathbf{a}$ replaced by $\mathbf{a}^m$.

Proposition 2.2.1. *Let $i \in \nu(t_i)$, $j \in \nu(t_j)$. Then, we have*

$$\Big(\sqrt{n(t_i)}V(t_i, \mathbf{a}^m), \sqrt{n(t_j)}V(t_j, \mathbf{a}^u)\Big)^T \xrightarrow{d} \big(\mathbb{G}_{\mathbf{a}^m}(t_i), \mathbb{G}_{\mathbf{a}^u}(t_j)\big)^T,$$

where $\big(\mathbb{G}(t_i), \mathbb{G}(t_j)\big)^T$ *is a centered Gaussian vector, such that*

$$\mathbb{E}[\mathbb{G}_{\mathbf{a}^m}(t_i), \mathbb{G}_{\mathbf{a}^u}(t_j)] = \begin{cases} 2\sum_{k\in\mathbb{Z}} \left[\frac{\pi_{\mathbf{a}^m,\mathbf{a}^u,H(t_i)}(k)}{\pi_{\mathbf{a},H(t_i)}(0)}\right]^2, & i = j, \\ 0, & \textit{otherwise}, \end{cases} \tag{2.2.2}$$

for

$$\pi_{\mathbf{a}^m,\mathbf{a}^u,h}(k) = -\frac{1}{2}\sum_{q=0}^{l}\sum_{q'=0}^{l} a_q a_{q'} |mq - uq' + k|^{2h}. \tag{2.2.3}$$

The proof of the proposition relies on Lemma 2 of [21; 22] with modification of π function as in (2.2.3).

Note that a filter of order at least 2 ensures asymptotic normality for all the values of the function $H(t)$. For a filter of order 1, this convergence is available if and only if $0 < \sup_t H(t) < 3/4$. The above proposition shows that a joint distribution of local variations $V(t, \mathbf{a}^m)$ for any two time points t_i and t_j; and any two dilations m and u converges to normal distribution with a blocky diagonal covariance in a sense that non-zero elements are shown only for the entries corresponding to the same time point. Next, we derive relationship between the log of the expectation of the empirical moment $S'(t, \mathbf{a}^m)$ and the parameters of interest $H(t)$, C, and σ in the light of the proposition 2.1.2.

Proposition 2.2.2. *Let* $d > 1$ *and let* $t_1, \ldots, t_d \in [0, 1]$. *Then, we have*

$$\log \mathbb{E}[S(t_i, \mathbf{a}^m)] = \log\Big(C\left(\frac{m}{N}\right)^{2H(t_i)} \pi_{\mathbf{a},H(t_i)}(0) + \sigma^2 \sum_q a_q^2\Big) + \mathcal{O}\big(\epsilon^\eta \log(N)\big), \tag{2.2.4}$$

for all $i \in \{1, \ldots, d\}$ *and all* $m \in \{1, \ldots, M\}$.

The proof of this proposition follows from Proposition 2.1.1 when $\mathbf{a}$ is replaced by $\mathbf{a}^m$ and from stationarity of $V'(j/N)$. The above proposition implies that $\log \mathbb{E}[S(t_i, \mathbf{a}^m)]$ is a nonlinear function of the unknown parameters. Considering Propositions 2.1.2

and 2.2.2, we obtain a regression model for $\log S(t_i, \mathbf{a}^m)$ as $N \to \infty$:

$$\log S(t_i, \mathbf{a}^m) \sim \log\Big(C \left(\frac{m}{N}\right)^{2H(t_i)} \pi_{\mathbf{a},H(t_i)}(0) + \sigma^2 \sum_q a_q^2\Big), \quad \forall i, m, \tag{2.2.5}$$

which is nonlinear with respect to $H(t_i)$, C, and σ. Its covariance matrix $\Sigma_{\mathbf{H},C}$ is given by the equation (2.2.2) in Proposition (2.2.1). In particular, when the noise level σ is zero, the regression model simplifies to

$$\log S(t_i, \mathbf{a}^m) \sim \log C + 2H(t_i) \log\left(\frac{m}{N}\right) + \log\big(\pi_{\mathbf{a},H(t_i)}(0)\big) + \mathcal{O}\big(\epsilon^\eta \log(N)\big), \quad \forall i, m, \tag{2.2.6}$$

which turns out be linear with respect to $H(t_i)$ and $\log C$, if $\log\big(\pi_{\mathbf{a},H(t_i)}(0)\big)$ is negligible.

The following ordinary least square (OLS) estimator of $\big(H(t_1), \ldots, H(t_d), C, \sigma\big) = (\mathbf{H}, C, \sigma)$ is introduced.

$$(\hat{\mathbf{H}}_{ols}, \hat{C}_{ols}, \hat{\sigma}_{ols}) = \underset{(\mathbf{H},C,\sigma)}{\arg\min} \sum_{i=1}^{d} \sum_{m=1}^{M} \Big(\log S(t_i, \mathbf{a}^m) - \log \mathbb{E}[S(t_i, \mathbf{a}^m)]\Big)^2. \tag{2.2.7}$$

One may use a weighted least square (WLS) estimator using the covariance matrix. In practice, we use the OLS estimator due to its computational convenience and negligible loss of precision compared to the use of WLS. We note that the gain of the WLS estimator is extremely small compared with the computational cost involved in the calculation of the covariance matrix.

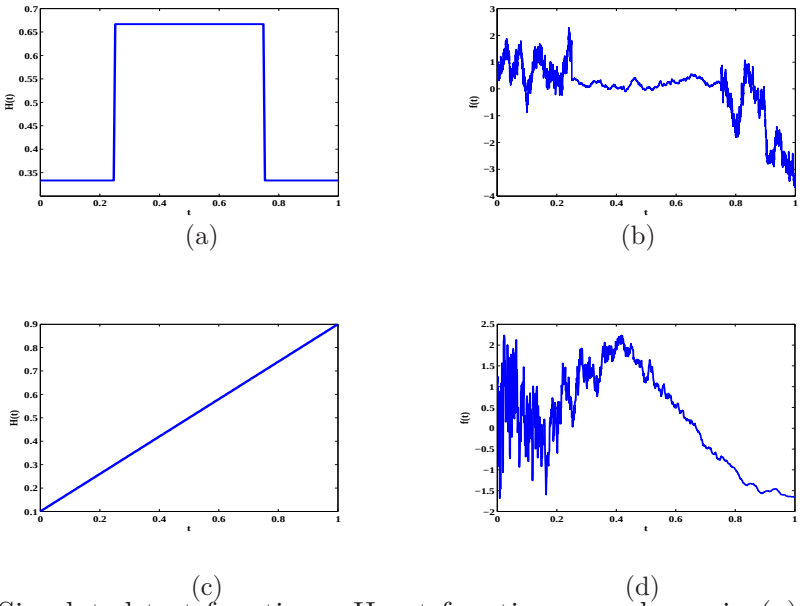
2.3 Simulations and Comparisons

We present here a simulation study of the performance of the approach suggested in this chapter (*Simultaneous K-var*). The simulation is done with "known truth" of Hurst function $H(t)$, and controlled signal variance and signal-to-noise (SNR) ratio. Test functions are shown in Figure 6 with *step-function* for $H(t)$ in Figure 6(a) and *straight-line* function in Figure 6(c). We also compare the average mean squared error (AMSE) performance with several popular methods such as Local Spectra Slope which

is summarized in [39]; k-variation for known variance, which is denoted by *K-var*, and k-variation of variance corrected version in [20], which is denoted by *K-var (VC)*. We will use hereafter the following notations regarding filters:

- Diff.i: denotes the filter of differences of order i (i vanishing moments).
- Db.i: denotes a Daubechies' wavelet filter of order i (i vanishing moments).
- Sym.i: denotes a Symmlet' wavelet filter of order i (i vanishing moments).

To simulate a sample path from a fBm on $[0, 1]$, we used the method of [77]. One can simulate a standard mBm W with covariance matrix $C_{H(\cdot)}$ by generating $Z \sim N(0, I_N)$ and estimating $W := C_{H(\cdot)}^{1/2} Z$. This method is exact in theory and sufficiently fast for reasonable sample size N. We generate 1000 series of length $N = 4096$ for step-

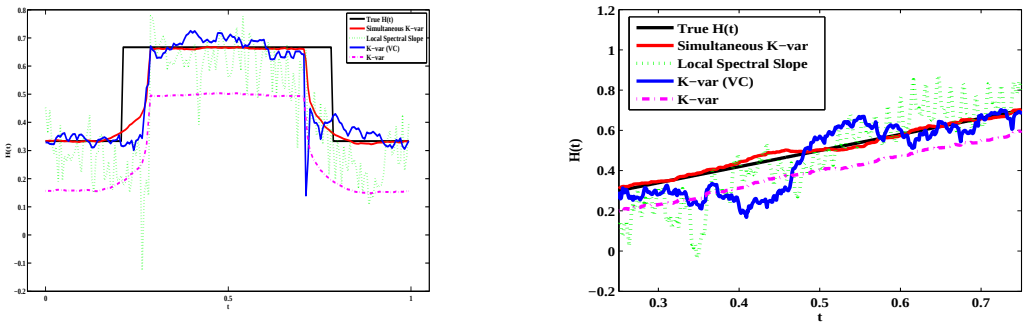


(a) (b) (c) (d)

Figure 6: Simulated test functions. Hurst functions are shown in (a) and (c); their illustrations of signal are shown in (b) and (d), correspondingly.

function $H(t)$ and $N = 1024$ for straight-line function $H(t)$, respectively. A Symmlet filter of order 8 (Sym.8) was used for step-function $H(t)$ and a simple difference filter $[\,1 \;\; -2 \;\; 1\,]$ (Diff.2) was used for straight-line $H(t)$. For Local Spectra Slope, the length of sub-signal was set to be 512 to be sufficient for its numerical stability and the two levels, by which spectral slopes are calculated, were 3 and 6. The size of a neighborhood of t, $\nu(t)$ in (2.1.5), is set to be 50 for Simultaneous K-var, K-var,

and K-var (VC). Illustrations of the estimators under no noise are shown in Figure 7. Estimators from K-var without considering scale parameter C deviate from true $H(t)$. We note that the distances between the K-var estimators and true $H(t)$ are connected to C. Estimators by Local Spectral Slope are bumpy because it assumes the sub-signal in its processing is fBm without considering the connection of $H(t)$. Also it is observed that K-var (VC) is more unstable than Simultaneous K-var: it will be pointed out in terms of AMSEs of $\hat{H}(t)$ and $\hat{C}$.

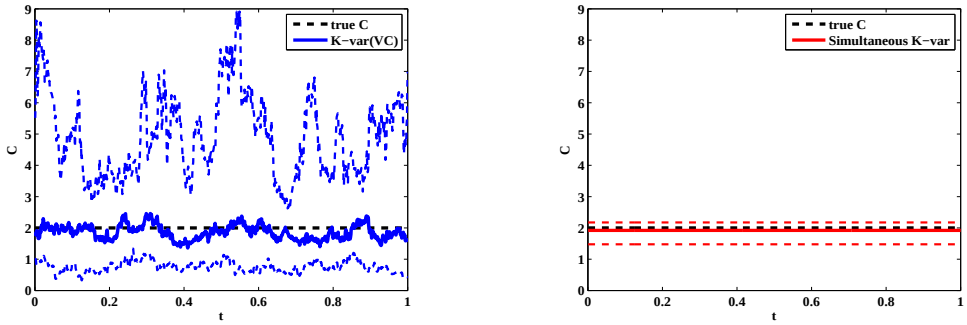


(a) (b)

Figure 7: Illustrations of the estimators; in panel (a), variance C was 2 and in (b) variance C was 4;the method of Simultaneous K-var matches the true $H(t)$ best among the four.

Regarding estimation of C, the comparison between K-var (VC) and Simultaneous K-var is shown in Figure 8, in which two empirical confidence intervals for the true $C = 2$ are shown. We sampled 1000 series of $W'(t)$ with $C = 2$ and straight-line $H(t)$ under no noise, $\sigma = 0$. In Figure 8(a), the confidence interval from K-var (VC) is not only dependent on time, but also it is wide and bumpy, while the confidence interval by Simultaneous K-var is constant in time and also sharp.

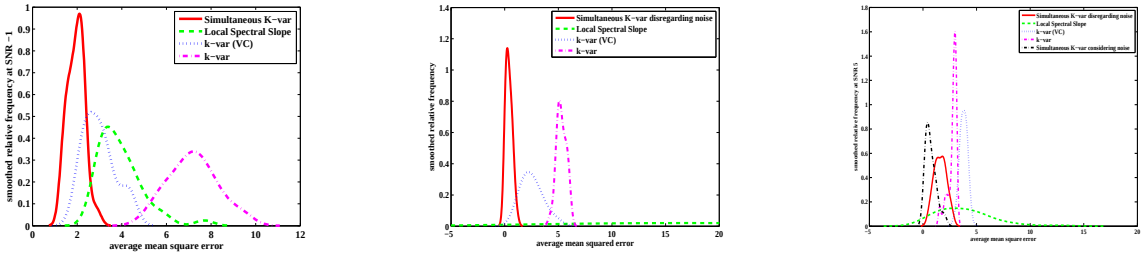
The comparison between K-var (VC) and Simultaneous K-var is made also in terms of AMSE. We calculated AMSEs for each $H(t)$ and generated histograms of AMSEs that were smoothed by kernel density estimation. Step-function $H(t)$ was used for Figure 9(a) and straight-line $H(t)$ was used for (b) and (c). The levels of variance C were 1, 4, and 4 for Figure 9(a), (b), and (c), respectively. In Figure 9(c)



(a) (b)

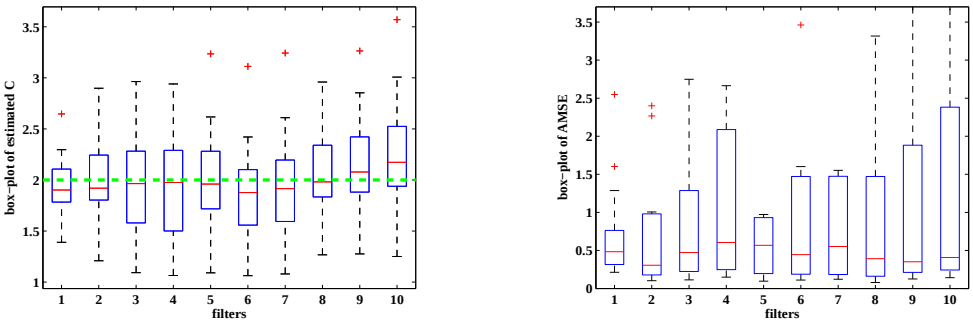
Figure 8: In panel (a), estimation of C with its empirical 95% confidence interval in blue by K-var (VC) when C = 2; in (b), estimation of C with simultaneous K-var in red. Simultaneous K-var gives a more stable and shorter confidence interval.

only, white noise was added with signal-to-noise (SNR) 7. The picture shows that Simultaneous K-var has the sharpest and smallest among other methods. In Figure 9(c), Simultaneous K-var was divided into two methods, the one considering white noise and the other ignoring the white noise in (2.2.4). Simultaneous K-var considering white noise was superior to others. The effects of adapted filters are summarized in Figure 10. The experiments were done with straight-line $H(t)$, variance $C = 2$, and SNR = 7. We observe that the performance of Simultaneous K-var on the estimation of C does not change much depending on the filter it uses. However, we mention that the variance of AMSEs tends to increase according to the filter size.



(a) (b) (c)

Figure 9: Smoothed histogram of average mean squared square for step-function $H(t)$ and variance $C = 1$ in (a) and straight-line $H(t)$ and variance $C = 4$ in (b) and (c); SNR 7 in (c); the method of Simultaneous K-var has sharper and smaller distribution of AMSE.



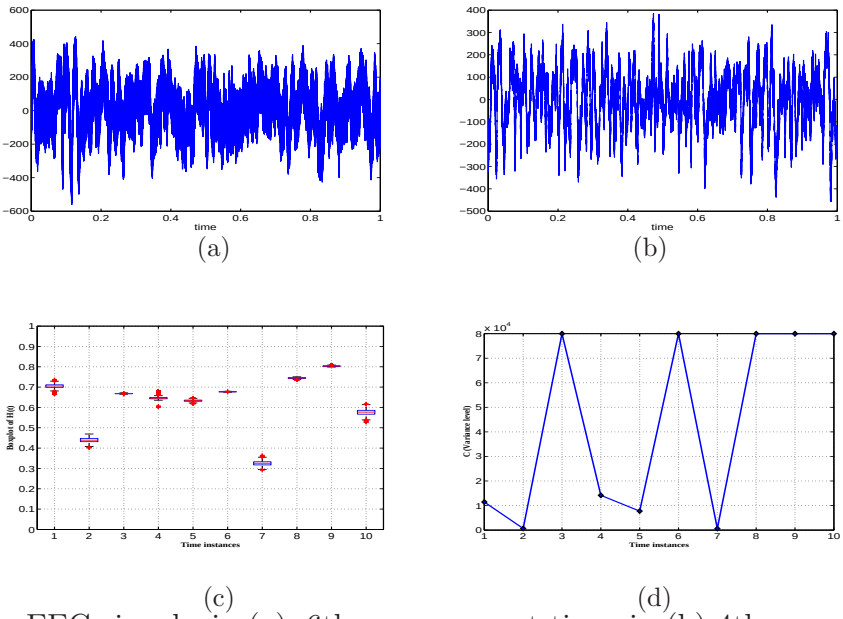
(a) (b)

Figure 10: The performances according to different filters with straight-line $H(t)$; in the panel (a), box plots for estimations of $C = 2$; in (b), box plots for AMSE of $H(t)$; filter numbers represent: 1, Diff.1; 2, Diff.2; 3, Diff.3; 4, Diff.4; 5, Sym.4; 6, Sym.6; 7, Sym.8; 8, Db.2; 9, Db.3; 10, Db.4.

2.3.1 An Example in EEG Data

To illustrate applications of the method proposed here, we estimate the time-changing hurst exponent and the variance factor in the noisy measurements from Electroencephalography (EEG) data set. EEG signals are the recording of electrical activity along the scalp produced by the firing of neurons within the brain. Continuous EEG monitoring is an increasingly utilized tool for the detection of secondary injury in neuro-critical care patients. Here we applied the suggested approach to correlate focal changes in regularity with onset of vasospasm. The EEG data for one vasospasm patient were collected over 7 measurement days. Each day, the signals were recorded over 1 minute period at 16 different channels of his/her scalp with a 500 Hz frequency. In Figure 11(a) and 11(b), we see two EEG signals at the 5th channel on 6th and 7th measurement days. In Figure 11(c), the longitudinal plot of boxplots of $H(t)$ show that two big jumps of regularity from 2nd to 3rd day and from 7th to 8th day. The biggest jump of the latter coincided with onset of vasospasm of the patient. In 11(d) the variance factor C according to the measurement days are shown. The search interval for C in (2.2.7) was set $(0, 80000)$ and the estimators on several days reached to the maximum of it, which brings suspicion of the mBm models to those signals.

The noise levels were all under 10^{-2} which confirms the assumption of sufficiently small noise magnitude compared to the signal. We admit that we are in the middle of seeking more patients to extract meaningful descriptors out of EEG signals with onset of vasospasm and their $H(t)$ and to validate their correlation.



(a) (b)

(c) (d)

Figure 11: EEG signals; in (a), 6th measurement time; in (b) 4th measure time. In (d), boxplots of estimated $H(t)$; in (c), estimated variance levles.

CHAPTER III

MULTISCALE METHODOLOGY IN CLASSIFICATION OF MULTIFRACTALITY

We apply multiscale statistical methods and extract meaningful multifractal descriptors to discriminate high-frequency data posessing monofractality from those which are inconsistenct with monofractality.

3.1 Monofractality

In this section we examine monofractality of a process by using the properties of singularity and scaling in wavelet transforms. By inspecting decay of wavelet coefficients, we can detect singularity and scaling simultaneously. We will discuss possible deviations from monofractality at the end of this section.

3.1.1 Singularity and Scaling

A signal, or a process $Y(t)$ is regular if it can be locally approximated by a polynomial. Referring to observed paths of a random process, a process will be interchangeably used with a signal. An irregular signal features local singularities. The singularity behavior of a process $Y(t)$ at time t_0 is characterized by Hölder exponent H_{t_0} (Lipschitz exponent): H_{t_0} is defined as the largest h such that there exists a polynomial P satisfying $|Y(t) - P(t)| \leq C|t - t_0|^h$ for t sufficiently close to t_0. Roughly speaking, saying that $Y(t)$ has exponent h at t_0 means that, around t_0, the process Y is bounded by the curves of $Y(t_0) + C|t - t_0|^h$ and $Y(t_0) - C|t - t_0|^h$ (see Figure 12 for graphical interpretation). If h is close to 0, the wide boundary from the two curves allows for large variations. As H_t approaches 1, the process becomes regular or smooth at the point t. The Hölder exponent of $Y(t)$ over an interval $[a \;\; b]$ is the sup of H_t over that

interval.

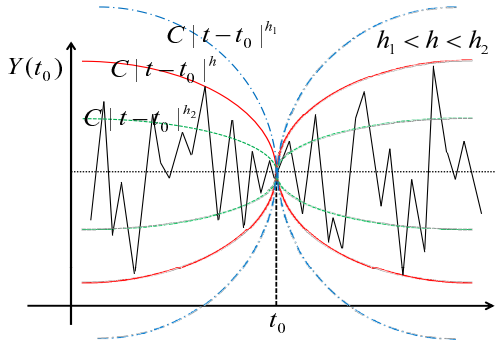


Figure 12: Graphical interpretation of Hölder exponent h of a process $Y(t)$ at a point t_0. Note that smaller h corresponds to a wider boundary within which the process is allowed to vary.

A process scales if its distributional properties are intrinsically invariant to changes of a scale. A process $Y(t)$ is self-similar with self-similarity index $H > 0$ (H-ss) if $Y(at) \stackrel{d}{=} a^H Y(a)$. Here $\stackrel{d}{=}$ denotes equality in all finite-dimensional distributions. An H-ss process with stationary increments exhibits long range dependence (LRD) when $H > 1/2$. A zero mean Gaussian process $B_H(t)$ with stationary self-similar increments is called fractional Brownian motion (fBm) with Hurst exponent H (fBm$_H$) if $B_H(t) \sim N(0, \sigma^2|t|^{2H})$, and

$$B_H(t+\tau) - B_H(t) \stackrel{d}{=} B_H(\tau) - B_H(0) \stackrel{d}{=} \tau^H B_H(1). \tag{3.1.1}$$

As a zero mean Gaussian process, $B_H(t)$ could be alternatively defined via its covariance structures:

$$\mathbb{E}\big[B_H(t)B_H(s)\big] = \frac{\sigma^2}{2}\Big[|t|^{2H} + |s|^{2H} - |t-s|^{2H}\Big]. \tag{3.1.2}$$

The scaling behavior of a signal is tightly related to singularity of wavelet coefficients [48, 36]. The singularity (Hölder exponent) and the self-similarity (Hurst exponent) are obtainable through multi-scale analysis of wavelet transforms. We will discuss wavelet transforms only at the level needed to introduce the concepts of singularity and self-similarity.

3.1.2 Wavelets: Detecting Singularity and Scaling

To detect the phenomena of singularity and self-similarity using wavelets, let us consider an $\mathbb{L}^1$-normalized orthogonal wavelet basis comprised of $\psi_{j,k}(t) = 2^j\psi(2^jt - k)$. Wavelet functions $\psi_{j,k}(t)$ are generated from wavelet function $\psi(t)$ by dilation by a scale factor 2^{-j} and translation of $2^{-j}k$. We assume that the $\psi(t)$ has $\mathcal{R}$ vanishing moments: $\int t^r\psi(t)dt = 0, r = 0, \ldots, \mathcal{R} - 1$. The coefficients of discrete wavelet transform of a process Y are defined by

$$d_{j,k} = \int_{-\infty}^{\infty} Y(t)\psi_{j,k}(t)dt, \tag{3.1.3}$$

which carries information on the local difference of the process near to the position k on a dyadic scale j. Let $k2^{-j} \to t$ means that $t \in [k2^{-j}, (k+1)2^{-j}[$ and $j \to \infty$. The result of Jaffard [see 48, p. 291] and Gonçalves [40] concerns detecting singularity of a signal: if $Y(t)$ is of Hölder exponent H, then

$$|d_{j,k}| = \mathcal{O}(2^{-jH}), \qquad \text{as} \quad k2^{-j} \to t \tag{3.1.4}$$

for any wavelet with $\mathcal{R} > H$. This means that the decay of the local differences of a process is related to the singularity of the signal, provided that the decomposing wavelet is more regular than the process.

Wavelets also enable us to detect the self-similarity of a signal. For an H-ss process with stationary increments (H-sssi), it can be shown that

$$d_{j,k} \stackrel{d}{=} 2^{-jH}d_{0,k} \stackrel{d}{=} 2^{-jH}d_{0,0}, \qquad \forall k, \tag{3.1.5}$$

which leads to the same order of $|d_{j,k}|$ as in (3.1.4). Note that $\mathbb{L}^2$ normalization is used in computations for the sake of computational simplicity, and $\mathbb{L}^1$ normalization is selected in discussions to simplify the rate of the decay: for $\mathbb{L}^2$-normalized wavelets, $d_{j,k} \stackrel{d}{=} 2^{-j(H+1/2)}d_{0,0}$. The equation (3.1.5) also serves as a basis for wavelet based estimation of H:

$$\log_2 \mathbb{E}|d_{j,k}|^q = -jqH + C_q, \tag{3.1.6}$$

where C_q is a constant depending on q, wavelet function ψ, and the magnitude of the signal. The *partition function*

$$T(q) = \lim_{j\to\infty}(-1/j)\log_2 \mathbb{E}|d_{j,k}|^q$$

measures the scaling of the higher order dependencies and the singularity structure of the process at the exponent q. Index k is arbitrary, given $d_{j,k}$ within the level j, that the partition function does not depend on k. In particular, for the H-sssi signal or the signal with Hölder exponent H signal, equation (3.1.6) rewrites: $\log_2 \mathbb{E}|d_{j,k}|^q = -jT(q) + C$, where C is a constant.

A practical estimation of H is based on empirical moments of the wavelet coefficients at dyadic scale j:

$$\hat{S}_j(q) = \frac{1}{n_j}\sum_k |d_{j,k}|^q, \tag{3.1.7}$$

where n_j is the number of $d_{j,k}$ available at dyadic scale j. We assume that the wavelet coefficients are uncorrelated, and hence independent, as has been approximately the case in various contexts (see [2] for a review and [4] for numerical simulations). A plot of the logarithm of the estimates $\hat{S}_j(q)$ against j, $\left(j, \log_2 \hat{S}_j(q)\right)$, is called as *qth order Logscale Diagram* (q-LD): it is also a wavelet spectrum for q=2. These diagrams result in straight lines with slopes of $-qH$, or $-T(q)$, for the fBm_H or signals with Hölder exponent H over the interval. Straight lines in q-LDs provide empirical evidence for monotone scaling. Partition function $T(q)$ is estimated as the slope in the following regression:

$$\log_2 \hat{S}_j(q) = -jT(q) + \varepsilon_j, \tag{3.1.8}$$

where the error term ε_j is introduced by the moment matching method when the true moments are replaced with the empirical ones. Simple ordinary least square (OLS) is the most convenient choice to estimate the partition function. Figure 13 shows wavelet spectra from three simulated fBm_H with different slopes under L^2 normalized Haar wavelet. In what follows, the Haar wavelet was used unless mentioned otherwise.

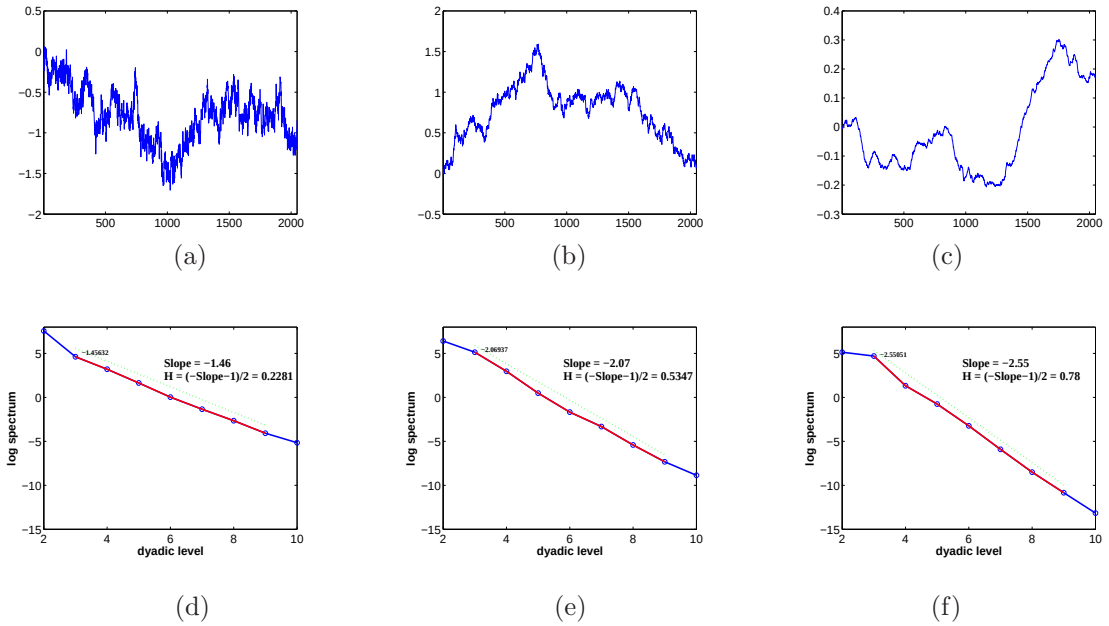


Figure 13: Simulations of fBm with (a) $H = 0.33$, (b) $H = 0.50$, and (c) $H = 0.80$; in the lower, the corresponding wavelet spectra are shown; as H gets larger, the spectrum line gets steeper.

Next, we will analyze the concept of deviation from monofractality by relating it to q-LDs.

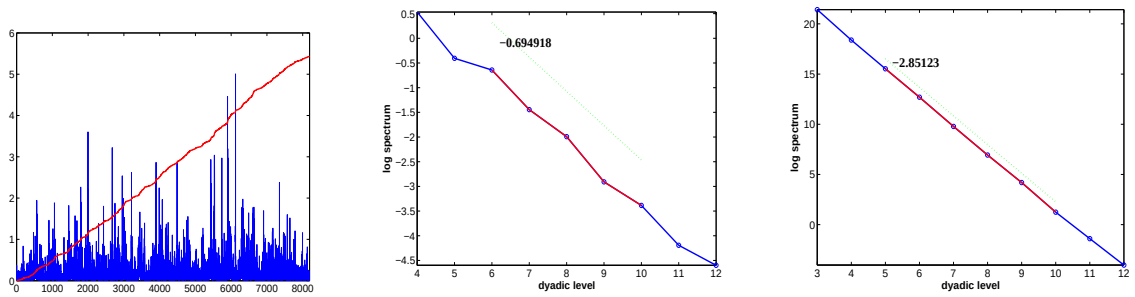
3.1.3 Deviation from Monofractality

We consider deviations from the linearity of $\log_2 \mathbb{E}|d_{j,k}|^q$ over dyadic scale j as *evidence of deviation from monofractality*. We make a distinction between deviation from monofractality and evidence for multifractality as the two are not synonymous: multifractality is richer as a form of scaling behavior since it is associated with randomness and distribution. Multifractal signals possess rich scaling behavior and deviation from linearity of the spectrum is not sufficient to characterize multifractal signals. For instance, multifractal signals can have perfectly linear 2nd order spectra as exemplified later. Distribution of local singularity is required to assess multifractality.

Linear scaling behavior at q-LDs does not necessarily suggest evidence for the presence of monofractality since multifractal signals can show linear scaling behavior. Figure 14(a) shows a realization of the multifractal wavelet model (MWM) synthesis

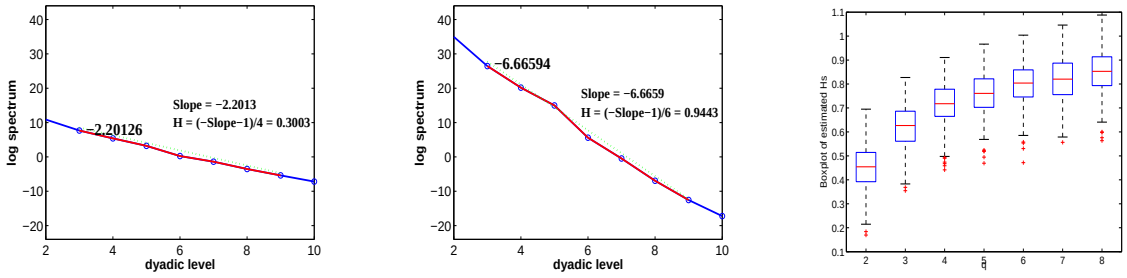
and its cumulative sum of the signal [68]. Since the signal, generated to be non-negative, was regarded to be comparable to fractional Gaussian noise, we took the cumulative sum, which reveals a more stable scaling behavior. Indeed, the Hurst exponent, 0.9255, from the cumulative sum was reasonable. The wavelet spectra of the two signals are shown in Figure 14(b), which reveals that the spectra are linear while the signal is multifractal. MWM is a multifractal extension of traditional fBm models and the MWM synthesis is a multiplicative and coarse-to-fine construction of scaling coefficients for positive and stationary LRD signals. It models the wavelet coefficients of a signal as $d_{j,k} = a_{j,k}u_{j,k}$ with the multiplier $a_{j,k}$ being independent random variables on $[-1\ 1]$ and $u_{j,k}$ being an approximation of the signal at dyadic scale j. The simulation in Figure 14(a) was done by β multifractal wavelet model using beta distribution as the multiplier $a_{j,k}$ and $\text{fBm}_{0.8}$ as the initial approximation of the signal $u_{j,k}$ in the coarsest level of smooth. This observation that the multifractal signal has linear spectrum indicates a weakness of spectral slopes in characterization of deviation from monofractality.

Moreover, scaling behavior is sensitive to the exponent q in q-LD. Figure 15 shows different scaling behavior over the exponent for simulated signals from $\text{fBm}_{0.3}$ + $\text{fBm}_{0.7}$. The spectral slopes from 2-LD in 15(a) and 6-LD in 15(b) were -2.2013 and -6.659, respectively, which resulted in different Hurst exponents, 0.6626 and 0.7454. Figure 15(c) plots boxplots of $1,000$ estimators of H for different exponent q. Thus, looking at an isolated q will not be sufficient to assess monofractality. This also emphasizes the shortcoming of spectral slopes and motivates the MFS to consider different scaling behaviors relevantly. Instead of making scaling inferences on the spectral slopes, one can adopt an empirical approximate of the MFS from wavelet-based partition functions that include information on the spectral slopes. Now, we propose a testing procedure to distinguish signals of monofractality from those that deviate from monofractality based on the MFS.



(a) (b) (c)

Figure 14: (a) One realization (in blue) of the multifractal wavelet model synthesis overlapping with its cumulative sum (in red) scaled by 1/200; (b) the wavelet spectrum of the signal; (c) the wavelet spectrum of the cumulative sum, which shows a clearly linear scaling behavior.



(a) (b) (c)

Figure 15: (a) 2th order logscale diagram for a $\mathrm{fBm}_{0.3}$ + $\mathrm{fBm}_{0.7}$; (b) 6th order logscale diagram for the signal; (c) boxplots of estimated Hurst exponents over the scaling exponent q; linear spectrum decays at an exponent q is insufficient to link to monofractality; the estimated Hurst exponents would not vary with exponent q if the process was monofractal.

3.2 Multifractal Spectrum

MFS of a process is a summary of its scaling and singularity properties. Here we describe MFS and discuss how to apply it in measuring deviations from monofractality.

Let us consider the local singularity strength of wavelet coefficients as follows [40]:

$$\alpha(t) = \lim_{k2^{-j}\to t} -\frac{1}{j}\log_2 |d_{j,k}|. \tag{3.2.1}$$

The local singularity strength measure (3.2.1) converges to the local Hölder exponent of the process at time t. Small values of $\alpha(t)$ reflect the more irregular behavior at time t. Any inhomogeneous process has a collection of local singularity strength

measures and their distribution $f(\alpha)$ forms the MFS. A direct way to obtain this spectrum is to use the counting technique,

$$f(\alpha) = \lim_{\epsilon \to 0} \lim_{j \to \infty} \frac{1}{j} \log_2 \Big(2^{-j} \#\{k \,:\, 2^{-j(\alpha+\epsilon)} < |d_{j,k}| < 2^{-j(\alpha-\epsilon)}\} \Big), \tag{3.2.2}$$

which captures the limiting frequency of occurrences of a given singularity α and ranges from -1 to 0. It relates to the distribution of the local singularities. Smaller $f(\alpha)$ implies fewer points in t behave with singularity α. If all points in t behave with singularity α^*, then $f(\alpha^*) = 0$.

Although it is feasible to estimate the MFS using (3.2.1) and (3.2.2), the method is not practicable due to the computational difficulty of approximating the limit. The multifractal formalism enables MFS f to be calculated by taking Legendre transform f_L of partition function T, $f_L(\alpha) := \inf_q \{q\alpha - T(q)\}$: using the theory of large deviations, one can show that $f_L(\alpha)$ converges to the true MFS $f(\alpha)$ [68, 67]. Because of the log-convex property of the moment generating function and concavity of $T(q)$, f_L is obtained as follows:

$$f_L(\alpha) = q\alpha - T(q) \quad \text{at} \quad \alpha = T'(q).$$

Using the estimator $\hat{T}(q)$ of $T(q)$ in (3.1.8), for equally spaced q_i with spacing $q_0 = q_i - q_{i-1}$, we estimate $f_L(\alpha)$ as follows [40]:

$$\begin{aligned} \hat{\alpha}_i &= [\hat{T}(q_{i+1}) - \hat{T}(q_i)]/q_0, \\ \hat{f}_L(\hat{\alpha}_i) &= q_i \hat{\alpha}_i - \hat{T}(q_i). \end{aligned} \tag{3.2.3}$$

Inspecting the MFS of monofractals is beneficial to build intuition on the variety of shapes of the Legendre transform based MFS of different signals.

Example 3.2.1. For a fBm_H in (3.1.1), it is easy to show that $\mathbb{L}^1$-normalized wavelet coefficients are

$$d_{j,k} \sim N(0, \sigma_\psi 2^{-2jH}), \tag{3.2.4}$$

where σ_ψ is a constant that depends on the wavelet function ψ and the magnitude of the signal, hence the partition function T and the MFS as Legendre transform f_L of T become

$$T(q) = \begin{cases} -\infty, & q \leq -1, \\ qH, & q > -1, \end{cases} \quad \text{and} \quad f_L(\alpha) = \begin{cases} -\infty, & \alpha < H, \\ 0, & \alpha = H, \\ H - \alpha, & \alpha > H. \end{cases} \tag{3.2.5}$$

Figure 16 depicts the theoretical partition function and the corresponding MFS for $\text{fBm}_{0.3}$. Note that both are the Legendre transform of each other: the slope -1 and the intercept $(0, H)$ in Figure 16(b) of MFS correspond to a point of $(-1, -H)$ in Figure 16(a) of partition function $T(q)$.

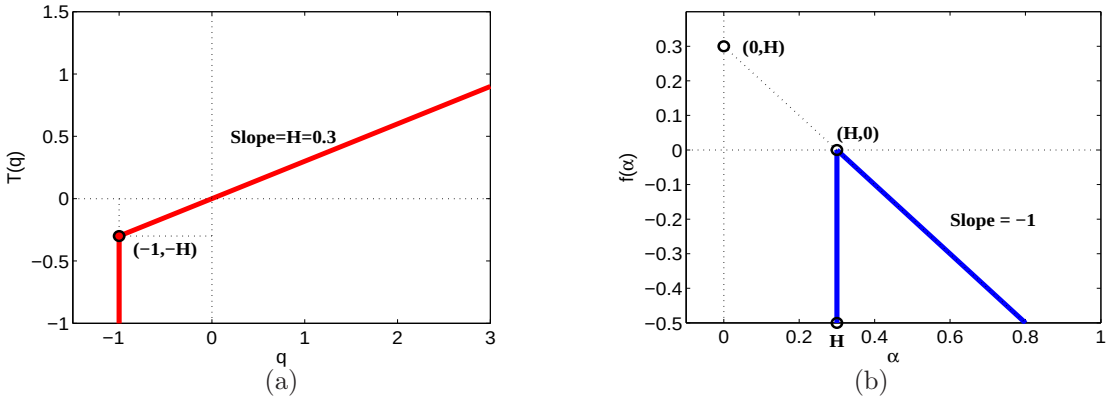


Figure 16: (a) Partition function $T(q)$ of fBm(0.3); (b) MFS $f(\alpha)$ of the signal.

Example 3.2.2. Suppose we observe different fBm processes varying with time intervals: $X(t)$ is given by

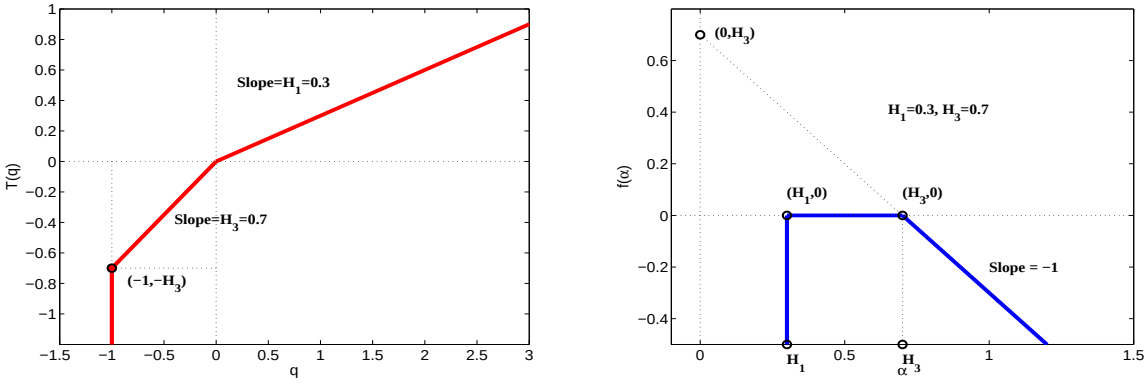
$$X(t) = B_{H_k}(t), \qquad t \in [t_{k-1}, t_k[, \tag{3.2.6}$$

for $k = 1, 2, 3$, and $H_1 < H_2 < H_3$. Since $T(q)$ is determined by the minimum of

Hurst exponents when $q > 0$ and by the maximum when $q < 0$ [71], we have

$$T(q) = \begin{cases} -\infty, & q \leq -1, \\ qH_3, & -1 < q < 0, \\ qH_1, & q \geq 0, \end{cases} \quad \text{and} \quad f_L(\alpha) = \begin{cases} -\infty, & \alpha < H_1, \\ 0, & H_1 \leq \alpha \leq H_3, \\ H_3 - \alpha, & \alpha > H_3. \end{cases} \tag{3.2.7}$$

The illustration of (3.2.7) is shown in Figure 17 for $H_1 = 0.3$, $H_1 = 0.5$, and $H_3 = 0.7$. It is worth mentioning that the MFS is flat in the interval between $\min\{H_i\}$ and $\max\{H_i\}$ and that information on regularities between $\min\{H_i\}$ and $\max\{H_i\}$, which is H_2 in this example, is lost in $T(q)$ and $f_L(\alpha)$. We will see in the next section that some of the low-dimensional descriptors of MFS are consistent with the deviation from monofractality.



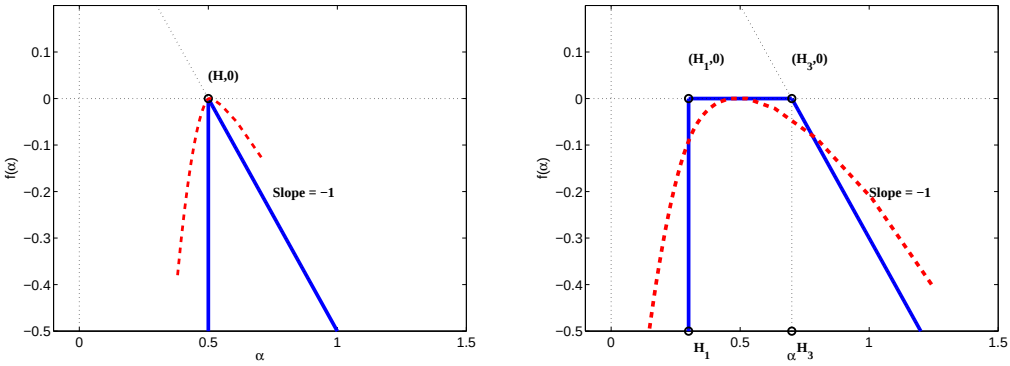
(a) (b)

Figure 17: (a) Partition function $T(q)$ of $X(t)$; (b) MFS $f(\alpha)$ of the signal.

3.2.1 Multifractal Descriptors

Rather than operating with MFS as a function (density), we summarize it by a small number of meaningful descriptors. These descriptors are interpreted in terms of location and shape of MFS because they are calibrated by the counterpart of MFS of monofractal signals. Theoretically, the MFS of fBm (a representative of monofractal) consists of three geometric parts: the vertical line, the maximum point, and the right slope, as is shown in Figure 16(b). However, it is rare to obtain such

a perfect spectrum in practice. Even for the well simulated fBm, due to error of estimation (most of them are due to the partition function estimation and derivative calculation as presented above), the MFS deviates from the theoretical form, as shown in Figure 18. Panels (a) and (b) of Figure 18 show theoretical MFS as a blue solid line and empirical MFS as a red dashed line for $\text{fBm}_{0.5}$ and $X(t)$ in the example 3.2.2, respectively. Notice that the maximum or mode is well approximated, but the slope exhibits discrepancy between theoretical and empirical MFS due to numerical instability.

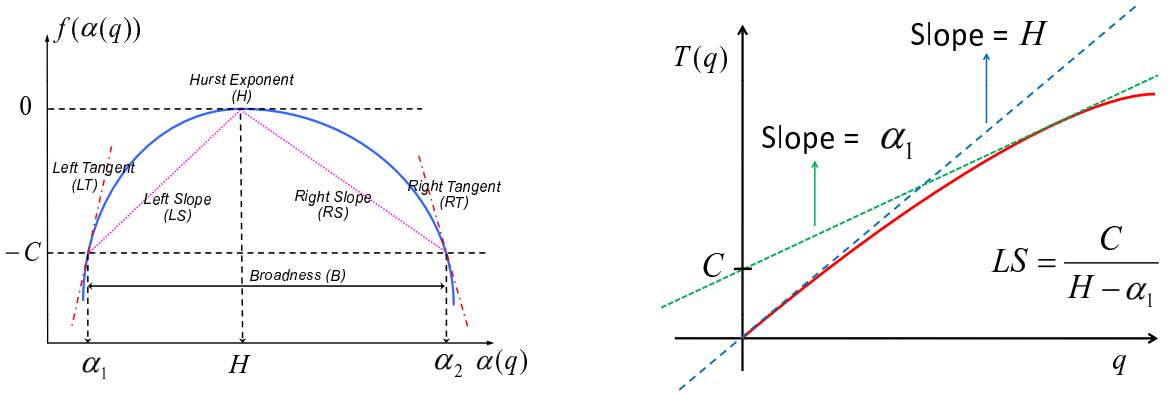


(a) (b)

Figure 18: Theoretical MFS (blue solid line) and an empirical MFS (red dash line); (a) for $\text{fBm}_{0.5}$; (b) for $X(t)$ in the example 3.2.2; empirical MFSs deviate from theoretical ones.

Despite the existence of estimation error, the MFS can be approximately summarized by 3 canonical descriptors (multifractal descriptors) without loss of the discriminant information. The proposed summaries are (1) the spectral mode (Hurst exponent, H), (2) left slope (LS) or left tangent (LT) and (3) width spread (broadness, B) or right slope (RS) or right tangent (RT). A typical MFS can be quantitatively described as shown in Figure 19(a). There are many ways to define the broadness (B). These descriptors have been successfully used in classification procedures as in [28] and [70]. In this chapter, we select the following definition [70].

Definition 3.2.1. Suppose that α_1 and α_2 are two roots which satisfy the equation $f(\alpha) + C = 0$ and $\alpha_1 < \alpha_2$. The broadness (B) of MFS is defined as $B = \alpha_2 - \alpha_1$.



(a) (b)

Figure 19: (a) Illustration of geometric descriptors of MFS. Note that the horizontal axis represents values of Hölder exponent $\alpha(q)$, while the vertical axis represents values proportional to the relative frequency of these indices, $f\big(\alpha(q)\big)$; (b) interpretation of left slope (LS) with partition function $T(q)$; LS is obtained by the two slopes (H and α_1) of the two tangent lines; LS is adopted as a measure of deviation from the straight line passing through the origin.

The suggested multifractal descriptors are graphically presented in Figure 19(a). Figure 19(b) shows the link of descriptor LS to the configuration of partition function $T(q)$ since LS is determined by two points $(-C, \alpha_1)$ and $(H, 0)$ in Figure 19(a), which correspond to the tangent line passing through $(0, C)$ in green and the tangent line passing through the origin in blue in Figure 19(b). It should be noted that the threshold value C in the definition could be adjusted empirically in the practical analysis to ensure that this measure is well computed for signals under analysis. In practical implementations, we use $C = 0.15$ or 0.2.

Another difficulty in computation is caused by the discreteness of $\alpha(q)$. The problem is that it may be hard to find the exact roots of the equation $f(\alpha) + C = 0$ among the discrete values of α's. To get around this, we first find the two closest points $(\alpha_i^l,\ f(\alpha_i^l))$ and $(\alpha_i^u,\ f(\alpha_i^u))$ for each i such that

$$f(\alpha_i^l) < -C \quad \text{and} \quad f(\alpha_i^u) > -C, \quad i = 1, 2,$$

and then obtain the two solutions α_1, α_2 by interpolation. The slopes LS and RS and tangents LT and RT can be obtained using the interpolation technique, as computed by

$$\begin{aligned} LS = C/(H-\alpha_1) \quad \text{and} \quad RS = -C/(\alpha_2 - H), \\ LT = (f(\alpha_1^u) - f(\alpha_1^l))/(\alpha_1^u - \alpha_1^l) \quad \text{and} \quad RT = (f(\alpha_2^u) - f(\alpha_2^l))/(\alpha_2^u - \alpha_2^l). \end{aligned} \tag{3.2.8}$$

Interpretation of H and LS (or LT) is straightforward. The apex of the spectrum or the most common Hölder exponent α found within the signal represents Hurst exponent H. The slope of the distribution produced by the collection of Hölder exponent α with smaller values of the mode (H) represents LS (or LT).

In this study, we selected the LS as the multifractal characteristic for measuring deviation from monofractality because the monofractality theoretically corresponds to a vertical line at H, infinite LS, in MFS. We related the extent of deviation from the vertical line to the characterization of monofractality, which is explored more in the following section.

3.3 Test for Deviation from Monofractality

In this section, we analyze the MFS summaries as possible statistics for assessing deviation from monofractality. For this goal, the LS turns out to be an informative index.

3.3.1 Left Slope in MFS as a Measure of Deviation from Monofractality

We start with intuitive interpretation of LS, connecting it with the partition function $T(q)$. Geometrically, α_1 in Figure 19(a) is the slope of the tangent line whose intercept is C in Figure 19(b). In addition, H in Figure 19(a) is the slope of the tangent line that passes through the origin. Theoretically, the expectation of $T(q)$ is linear in an ideal case of fBm, which leads to a perfect vertical line at the Hurst exponent as in Figure

16(b) and thus the infinite LS. Empirically, the wavelet-based estimator $\hat{T}(q)$ of $T(q)$ in (3.2.3) deviates from the straight line because of the finite approximation to the moments and numerical deviations. This causes LS to be finite for empirical fBms. As a result, LS incorporates information on the shape of the partition function. It reflects deviation from the straight line passing through the origin: the more linear the partition function, the larger LS. The process $X(t)$, which is a synthetic superposition of the three fBms, in the example 3.2.2, lead the flat segment between $\min\{H_i\}$ and $\max\{H_i\}$ (as the blue line in Figure 18(b)) and much wider breadth compared to MFS of the individual fBm (as the blue line in Figure 18(a)). The theoretical MFS has a wide breadth (as the blue line in Figure 18(b)) leading to small LSs for its empirical processes (as the dotted red line in Figure 18(b)) in comparison with those for empirical fBms (as the dotted red line in Figure 18(a)). We attribute this decrease to deviation from monofractality of the signal.

To emphasize this point, consider a multifractional Brownian motion (mBm) with time varying Hurst exponent. A mBm with $H(t)$ ($\text{mBm}_{H(t)}$) is a zero mean Gaussian process defined as in (3.1.2) by replacing H with a time-varying $H(t)$ [10, 21]. Specifically, we consider a $\text{mBm}_{H(t)}$ with $H(t)$ given as, for $T = 2^{11}$,

$$H(t) = \frac{0.6}{T}t + 0.2, \qquad t \in [\,0 \;\; T\,].$$

Next we compare this $\text{mBm}_{H(t)}$ with a standard Brownian motion, $\text{fBm}_{0.5}$. In Figure 20(a) and Figure 20(b), simulated signals of the two processes and the corresponding MFS are shown. The LS of $\text{mBm}_{H(t)}$ was smaller than that of $\text{fBm}_{0.5}$ (0.48 compared to 1.09). In Figure 20(c) and 20(d), the partition functions from the two signals are shown, respectively. The shapes of the two partition function that carry information equivalent to MFS are strikingly different. The LS reflects the difference of the two tangent lines (red and green) for each partition function: the larger the difference of the two slopes, the smaller the LS. We observe that the partition function in 20(d) deviates from the theoretical partition function (the straight red line) more severely

than that in 20(c). The slopes of the two tangent lines (the green lines) of the two partition functions that pass through the point $(0, 0.2)$ are different due to different shapes of the two empirical partition functions.

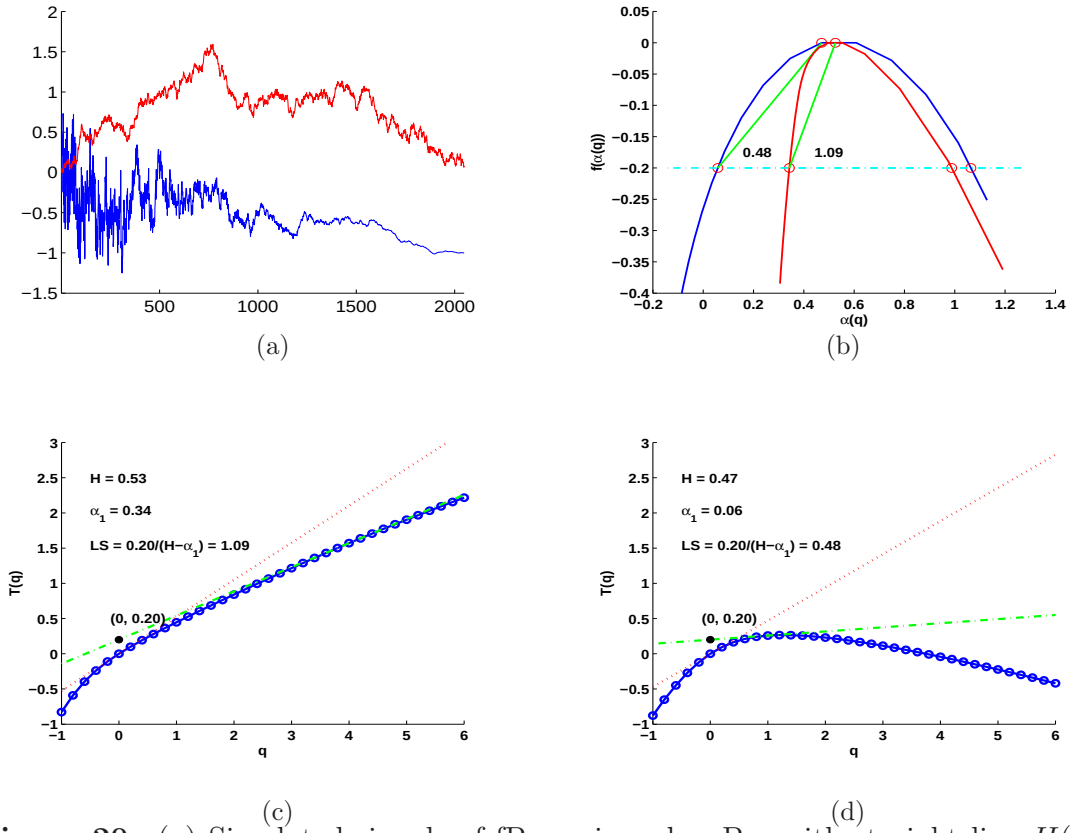


Figure 20: (a) Simulated signals of $\text{fBm}_{0.5}$ in red, mBm with straight line $H(t) = \frac{0.6}{T}t + 0.2, (T = 2^{11})$ in blue; (b) MFS of the $\text{fBm}_{0.5}$ in red and of the mBm in blue; (c) partition function for the $\text{fBm}_{0.5}$; (d) partition function for the mBm.

This behavior of LS is consistent whenever the monofractality of the signal is violated. Using this observation, we propose a testing procedure described in the next section, which tests the monofractality of a signal based on LS. The details of the testing procedure and its applications are provided.

3.3.2 Parametric Bootstrap Test

Bootstrapping is a computer-based method for assigning measures of accuracy to statistical estimates with sampling from an approximating distribution [34]. The advantage of bootstrapping is that it is straightforward to apply the bootstrap to derive

estimates of complex estimators such as percentile points, proportions, odds ratio, and correlation coefficients. The bootstrap method may also be used for constructing hypothesis tests as an alternative to inference based on parametric assumptions. In the case in which exact distributions are unknown or analytic procedures are too complex to obtain, even an approximation to the distribution, the bootstrap techniques are employed. In our case, the distribution of LS for monofractality of fixed size, wavelet basis, and precision settings of MFS calculation are overly complex.

With LS as a measure of deviation from monofractality, we propose a new testing procedure to check if a signal is monofractal; H_0: the signal is monofractal vs. H_1: not H_0. This type of hypothesis is a goodness-of-fit type. Not rejecting H_0 leads to the conclusion that the signal is consistent with assumption of monofractality. Rejecting H_0 does not indicate multifractality, but just a violation of monofractality or inconsistency of the monofractality assumption. The proposed test is conducted with parametric bootstrap which is outlined in Figure 21. We start with an observed signal and a wavelet basis with a sufficient number of vanishing moments; and also fix C as in Definition 3.2.1 and q_i in (3.2.3). The following steps describe the testing algorithm:

[1] Calculate $\hat{LS}$ and $\hat{H}$ as estimators of LS and H, respectively, for an input.

- *(a)* Calculate wavelet coefficients $d_{j,k}$ as in (3.1.3).
- *(b)* Estimate the partition function $T(q)$ with $d_{j,k}$ as in (3.1.8).
- *(c)* Estimate the MFS $f(\alpha)$ with $T(q)$ as in (3.2.3).
- *(d)* Estimate H as the maximizer of $f(\alpha)$ and find LS as in (3.2.8).

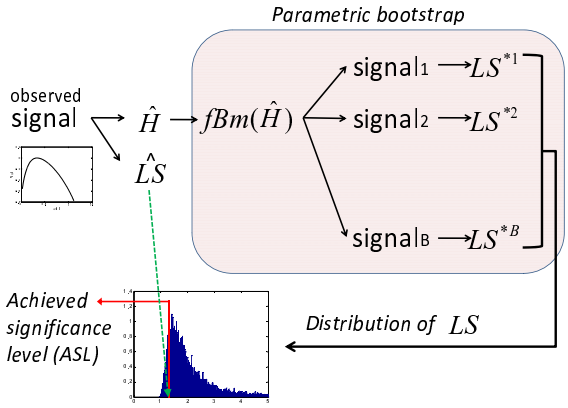
[2] Generate B copies of $\text{fBm}_{\hat{H}}$ and for each copy (realization) find LS^{*b}, $b = 1, \ldots, B$; this is the parametric bootstrap step.

[3] Construct a bootstrap distribution of $\hat{\mathbf{LS}}$ using bootstrap replicates; find empirical 0.05 quantile ($q_{0.05}$).

[4] If the LS is less than $q_{0.05}$, reject H_0.

To simulate a sample path from a fBm, we used the method of Wood and Chan, which

is based on Fourier transform [77]. We construct an empirical distribution of $\hat{LS}$ as a surrogate of the true distribution of the true LS from B number of replicates of $\text{fBm}_{\hat{LS}}$. Since signals of monofractality have high LS values, the achieved significance level (ASL) of the test is the proportion of the number of replicates for which the left slope (LS^{*b}) is less than $\hat{LS}$ to the total number of replicates (B). In hypothesis testing with bootstrapping, ASL is the counterpart of the p-value in the classical hypothesis testing. We can also adjust the quantile of $q_{0.05}$ to be different from that of 0.05.



(a)

Figure 21: Parametric bootstrapping for testing if a signal observed is monofractal; the achieved significant level (ASL) of the test is the area of the bootstrap distribution enclosed with the red line since monofractal signals have high values of LS.

3.3.3 Experimental Result

We perform a simulation experiment to test the following non-monofractal signal $X(t)$ for monofractality:

$$X(t) = B_{H_k}(t), \qquad t \in [t_k + 1, t_k + 2^{10}[, \tag{3.3.1}$$

for $t_k = (k-1)2^{10}$, $k = 1, \ldots, 4$, and $H_1 = H_3 = 0.3, H_2 = H_4 = 0.7$. We chose B as 5000, C as 0.15, and q_i as equi-spaced with size 0.2 on $(-1\ 6]$; and tested 3000 samples of $X(t)$. Obviously, the signal is not monofractal since regularity is

not constant across the time. We want to test if H_0: $X(t)$ is monofractal vs. H_1: not H_0. An illustration of $X(t)$, its wavelet spectrum, its MFS, and the empirical distribution of LSs are shown in Figure 22. The wavelet spectrum in Figure 22(b) shows a monotone decay across the dyadic scales, and yet the LS was 0.38 in 22(c), which is indicating irregular scaling. We show the normalized histogram (bootstrap

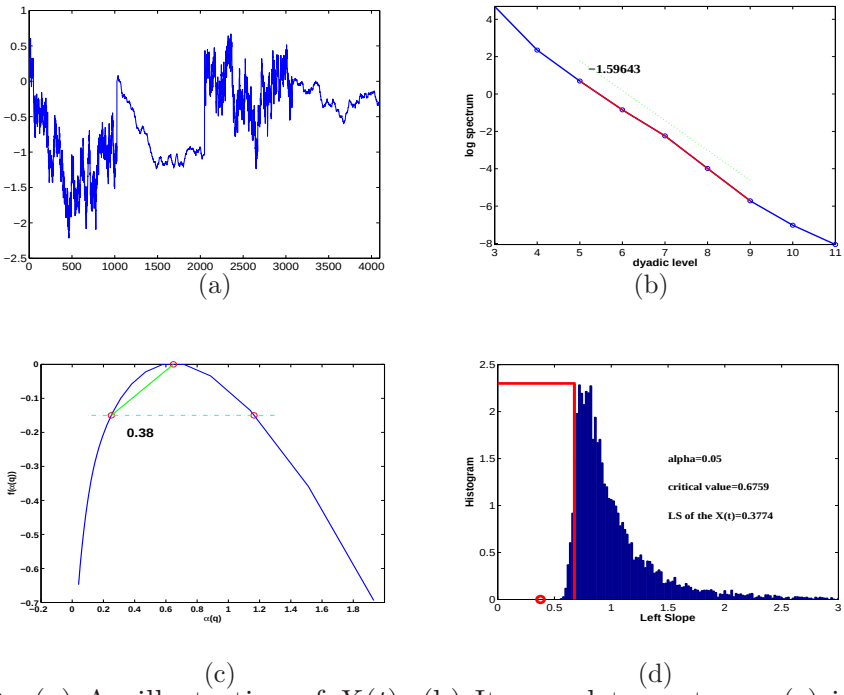


Figure 22: (a) An illustration of $X(t)$; (b) Its wavelet spectrum; (c) its MFS; (d) bootstrap distribution of LS^{*b} with the LS (0.38) for the signal (in red circle) and a rejection region of 95% achieved significance level (within the red line).

distribution) of LS^{*b} from the 5000 bootstrap fBm simulations and the critical region of level 0.05 in Figure 22(d). The LS clearly falls in the rejection region and H_0 is rejected: The ASL of this test was 0. Out of the 3000 tested signals, 2838 signals were concluded not be monofractal: The rate, or ability to recognize the true non-monofractal signals, was 0.946. Next we apply this to real-life examples.

3.3.4 Turbulence and DNA Examples

To illustrate the test procedure in a real-world example, we compared a turbulence signal with a $\text{fBm}_{1/3}$. Understanding the properties of turbulence is a major problem

of modern physics, which remains mostly open despite intense research efforts from 1941 when Kolmogorov formulated a statistical theory of turbulence [52]. Kolmogorov introduced his theory, often referred to as **K41** theory, for locally isotropic turbulence. The velocity field is modeled as a process $U(x)$ with increments having the following structure function of order p:

$$E\big[|U(x+r) - U(x)|^p\big] \propto (\epsilon r)^{\frac{p}{3}}.$$

Parameter ϵ, energy per unit of fluid mass per unit time, describes the energy transmission from large eddies, where the energy is injected, to small eddies, where the energy is converted to heat by viscosity. The **K41** theory states that a one-dimensional longitudinal trace of a three-dimensional velocity field is a fractal noise process with constant Hurst exponent 1/3 and models turbulence as a monofractal. Though the theory was verified in many empirical observations possessing the property of monotone spectral decay, it does not take into account the existence of coherent structures such as vortices and helicity. Kolmogorov [51] refined the homogeneity assumption of ϵ to be a location-varying dissipation rate $\epsilon(x)$, which leads to the model of multifractional Brownian motion. This turbulence model is not monofractal.

We tested a turbulence signal of length 2^{14} from velocity measurements on July 12, 1997 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 to check if the turbulence is monofractal. In Figure 23(a), the turbulence signal and $\text{fBm}_{1/3}$ are in black and red, respectively. The average log spectrum of squared wavelet coefficients for the two signals are shown in Figure 23(b). The spectral slopes are indistinguishable; the two signals do not differ with respect to their second order properties. The wavelet spectrum of the turbulence signal is shifted upwards from that of $\text{fBm}_{1/3}$ because of difference in their energies.

The two MFS along with their descriptors are shown in Figure 23(c): the MFS of the turbulence signal is wider than that of the $\text{fBm}_{1/3}$ signal. To quantify the degree of

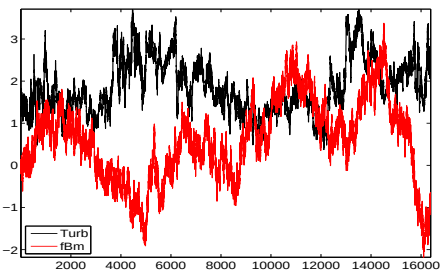
deviation of the turbulence signal from monofractality, $10,000$ samples of the $\text{fBm}_{1/3}$ and the empirical distribution of estimated LSs were obtained as is shown in Figure 23(d). Two circles represent LSs of the turbulence (left in black) and the $\text{fBm}_{1/3}$ (right in red). The critical point at 95% is highlighted with the red line in Figure 23(d). The black circle, corresponding to the turbulence signal falls in the critical region, leading to rejection of the null hypothesis. We concluded the turbulence signal was not monofractal.

Next, we demonstrate our method in an analysis of DNA sequences. In the analysis of DNA sequences, one of the most important tasks is to study whether two sequences are related. This is studied by using a scoring system to rank the possible relations between the sequences and by considering statistical methods to evaluate the significance of such relations [33]. Often the sequences of nucleotides (A, C, G, and T) are coded as functions or DNA walks, and fractal properties of these associated functions can be informative for functional properties of DNA segments [7].

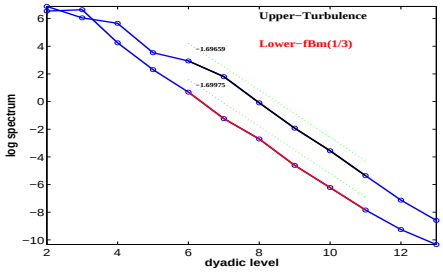
The analysis of DNA walks is influenced by the presence of a global linear trend induced by the excess of purines over pyramidines. In all eukaryotic species, a DNA molecule consists of a long complementary double helix of purine nucleotides (denoted as A and G) and pyrimidine nucleotides (denoted as C and T). A single strain of this DNA can be represented as a long word that corresponds to a random walk. Depending on the letter at position i in the word, the random walk gets a cumulative sum of increments of $x(i) = 1$ for A and G, and $x(i) = -1$ for C and T. Hence, the corresponding random walk is defined as $s(n) = \sum_{i=1}^{n} x(i)$, in which n is an index smaller than the length of the sequence.

In Figure 24(a), we show an 8196-long DNA random walk for a spider from the EMBL Nucleotide Sequence Database, which is also known as the EMBL-Bank. The wavelet spectrum and MFS of the signal are shown in Figure 24(a) and 24(b), respectively. The estimated Hurst exponent was 0.648 and the left slope was 1.47. We

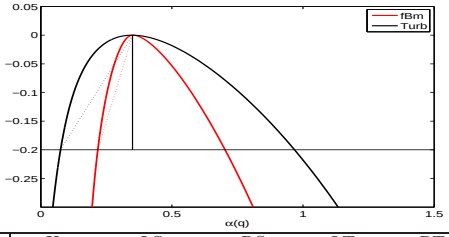
noticed that the MFS yielded only the left part from the mode because the partition function was flat for negative exponents and made the right part of the MFS computationally unobtainable. The empirical distribution of LS^{*b} from $10,000$ simulations of $\text{fBm}_{0.648}$ is shown in Figure 24(d). The ASL of the observed LS (1.47) was greater than 0.05, by which we conclude that the signal is monofractal. This conclusion is in accordance with the observation made by Arneodo that the DNA sequences are the most perfect monofractals found in nature (personal communications).



(a)

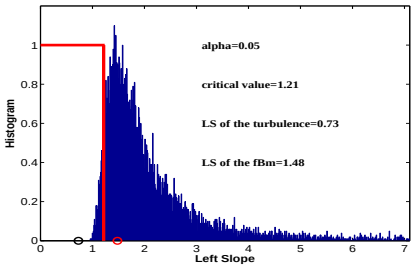


(b)



Signal	H	LS	RS	LT	RT	B
Turb	0.3382	0.7331	-0.3229	2.7500	-0.5400	0.8922
fBm	0.3422	1.4845	-0.5704	3.7400	-0.8700	0.4853

(c)



(d)

Figure 23: Comparison of the turbulence and $\text{fBm}_{1/3}$ signals; (a) turbulence in black and $\text{fBm}_{1/3}$ in red are indistinguishable with respect to their second order properties; (b) log spectra for the two signals with two spectral slopes produced identical slopes; (c) MFS and the descriptors for (a); (d) the bootstrap distribution of LS^{*b} along with two circles (left in back for $LS, 0.73$, of the turbulence signal, right in red for $LS, 1.43$, of the $\text{fBm}_{1/3}$) signal and a rejection region of 95% achieved significance level (within the red line).

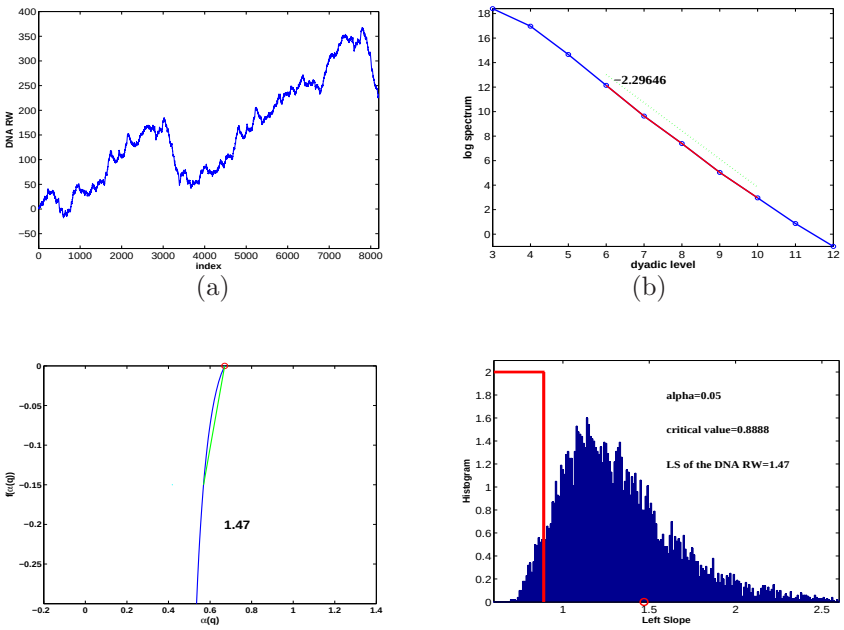


Figure 24: Demonstration of the test of monofractality to a DNA random walk: (a) 8196-long DNA random walk for a spider from the EMBL Nucleotide Sequence Database; (b) wavelet scaling with slope -2.296 and estimated Hurst exponent 0.648; (c) MFS with left slope 1.47; only the left part from the mode was computationally available due to a straight line in the partition function of negative exponents; (d) the distribution of LS^{*b} with the LS (1.47) in a red circle.

CHAPTER IV

CLASSIFICATION IN MULTISCALE DOMAINS

This chapter concerns a machine learning based classification method in function estimation on multiscale domains.

4.1 Wavelet Shrinkage

If wavelet transform $\mathcal{W}$ is applied to noisy measurements $y_i = g_i + \epsilon_i, i = 1, \ldots, N$, or in vector notation $\mathbf{y} = \mathbf{g} + \boldsymbol{\epsilon}$, with a possibly multivariate regression function g and normal noise $\boldsymbol{\epsilon}$, the transformed noise $\mathcal{W}\boldsymbol{\epsilon}$ is normal, as well. This linear and orthogonal transformation can be described for the discrete inputs by an orthogonal matrix $\mathcal{W}$ of dimension $N \times N$. With $\mathbf{d} = \mathcal{W}\mathbf{y}$ and $\boldsymbol{\theta} = \mathcal{W}\mathbf{g}$, the noise model in the time domain can be reformulated as $\mathbf{d} = \boldsymbol{\theta} + \boldsymbol{\epsilon}$ in the wavelet domain. Wavelet shrinkage methods, now widely utilized in nonparametric function estimation, estimate $\boldsymbol{\theta}$ using noisy observations $\mathbf{d}$. The simplest nonlinear wavelet shrinkage technique is thresholding. The two most common thresholding policies are *hard* and *soft* thresholding rules with corresponding rules given by:

$$\theta^{hard}(d, \lambda) = d\mathbf{1}\big(|d| > \lambda\big),$$
$$\theta^{soft}(d, \lambda) = \big(d - \mathrm{sgn}(d)\lambda\big)\mathbf{1}\big(|d| > \lambda\big),$$

where $\mathbf{1}(A)$ is an indicator of relation A.

These two shrinkage mechanisms with properly selected thresholds are characterized by exceptional statistical properties especially under i.i.d. Gaussian noise model [30]. Under the correlated Gaussian noise model, they also have near optimal behavior in a wide range of function classes [49]. Bayesian approaches [19, 74, 35] and minimax approaches [32] among others are suggested for accurate estimation of the

true signal. The wavelet thresholding based on semi-supervised learning, which is developed in this chapter, aims to exploit the hierarchical dependence structure of wavelet coefficients for more accurate model selection.

4.2 Semi-supervised Learning

In the real world, we encounter both labeled and unlabeled observations. When we estimate the true link between a label (response) and attributes (variables), it is beneficial to incorporate the unlabeled observations conditional on whether labeled and unlabeled attributes come from the same population. This requires a formal model that is capable of handling both kinds.

In this section, *Laplacian* kernel, associated with *manifold regularization* [9], was used to incorporate the information contained in the unlabeled observations. This information refers to low dimensional manifolds in the attribute space spanned by both labeled and unlabeled data. When such manifolds are well estimated, the assignment of labels can be more precise.

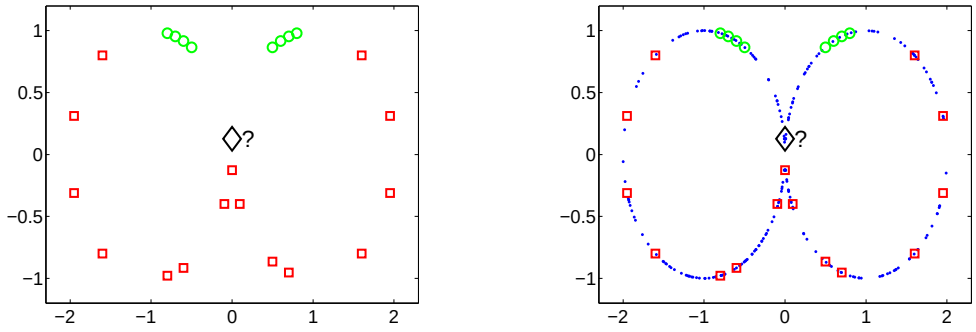


Figure 25: Illustration of semi-supervised learning where the new observation (the black diamond with question mark) is to be classified in the presence of two labels (a red square and a green circle): (a) when only labeled data are considered, the classification is a red square; (b) when both labeled and unlabeled data are considered, the classification is a green circle.

For the illustration of our point, consider Figure 25. In panel (a) only the labeled observations are present and they suggest that the best decision for the new observation denoted by the question mark is the red rectangle. However, when unlabeled observations are added (blue dots) the intrinsic low-dimensional topology of the attributes suggests that the new observation in question should be the green circle.

Suppose that empirical data are given as

$$\begin{aligned} (x_1, r_1), \ldots, (x_l, r_l) &\in \boldsymbol{X} \times \{0, 1\}, \\ x_{l+1}, \ldots, x_{l+u} &\in \boldsymbol{X}, \end{aligned} \tag{4.2.1}$$

where $\boldsymbol{X}$ is the domain of attributes x_i and the r_i is the *label*. There are l labeled and u unlabeled observations in (4.2.1).

The goal is to find a function f with domain $\boldsymbol{X}$ and range $\{0, 1\}$:$f : \boldsymbol{X} \mapsto \{0, 1\}$, which is well fitted to the labeled observations and regularized both in the ambient space (data space) and the intrinsic space (low-dimensional space of data) of $\boldsymbol{X}$. The labels selected are 0 and 1, but in principle the values can be arbitrary. An estimate of f, $\hat{f}$, is given by

$$\hat{f} = \arg\min_f \frac{1}{l} \sum_{i=1}^{l} V(r_i, f) + \|f\|_I^2,$$

where V is a loss function. Possible choices for V are squared error loss $\big(r_i - f(x_i)\big)^2$ for Ordinary Least Square (OLS), or hinge loss function $\max\{0, 1 - r_i f(x_i)\}$ for SVMs. The norm $\| \ \|_I$ controls the complexity of the solution in the intrinsic geometry of $\boldsymbol{X}$. The graph Laplacian L is constructed with an adjacency matrix W with entries W_{ij} representing the closeness between x_i and x_j and a diagonal matrix D with $D_{ii} = \sum_{j=1}^{l+u} W_{ij}$: $L = D - W$. While several natural choices for the norm $\| \ \|_I$ are possible, the graph Laplacian has an intuitive interpretation, in which the penalty is

proportional to the adjacency W_{ij} between x_i and x_j when the two are mapped apart.

$$\begin{aligned}\hat{f} &= \arg\min_f \frac{1}{l}\sum_{i=1}^{l} V(r_i, f) + \sum_{i\neq j}^{l+u}\big(f(x_i) - f(x_j)\big)^2 W_{ij} \\ &= \arg\min_f \frac{1}{l}\sum_{i=1}^{l} V(r_i, f) + f^T L f.\end{aligned} \tag{4.2.2}$$

Here the estimate for the labeled observations is the same as the labels: $f(x_i) = r_i, i = 1, \ldots, l$. We have $V(r_i, f) = 0$ and the minimization of (4.2.2) involves only unlabeled f_u:

$$\begin{aligned}\hat{f} &= \begin{bmatrix} f^{(l)} \\ \hat{f}^{(u)} \end{bmatrix} = \arg\min_f f^T L f \\ &= \arg\min_{f^{(u)}} [\ (f^{(l)})^T\ (f^{(u)})^T\ ] \begin{pmatrix} L_1 & L_2 \\ L_2^T & L_3 \end{pmatrix} \begin{bmatrix} f^{(l)} \\ f^{(u)} \end{bmatrix} = \begin{bmatrix} f^{(l)} \\ -L_3^{-1} L_2^T f^{(l)} \end{bmatrix},\end{aligned} \tag{4.2.3}$$

where L_1, L_2, and L_3 are l by l, l by u, and u by u, respectively. The quantity $-L_3^{-1} L_2^T f^{(l)}$ is interpreted as a confidence measure of support of label 1 against label 0, which can be justified by the approach of optimal graph cut, as in [9].

4.3 Semi-supervised Wavelet Shrinkage

Although wavelet transforms are whitening, in most cases the wavelet coefficients still show inter-dependence. There are several shrinkage mechanisms that utilize the hierarchical dependence of coefficients. Some of the approaches are [42] and [16]. The underlying idea is that the inclusion/exclusion of a coefficient should depend not only on its magnitude, but also on magnitudes of its neighbors. The proposed procedure for semi-supervised shrinkage uses the geometry of coefficients implied by their hierarchical structure. The methodology of semi-supervised wavelet shrinkage is outlined next.

In what follows, the double index jk representing the scale/shift indices in d_{jk} is omitted and a "typical" wavelet coefficient is denoted as d. Assume that d is modeled

as

$$d = \theta + \epsilon, \qquad \epsilon \sim \mathcal{N}(0, \sigma^2), \tag{4.3.1}$$

where we are interested in estimating the location θ. The SS shrinkage consists of the following three steps.

Step-1. Specify a thresholding rule of $\delta(d, \lambda) = d\mathbf{1}(|d| > \lambda)$. Other choices of thresholding policies (soft, semisoft[13], garrotte[38], *et al.*) can be used as well.

Step-2. Find an estimate $\hat{\sigma}$ of the standard deviation σ of d, and set two threshold levels λ_1 and λ_2 such that $\lambda_1 \leq \lambda_2$. This can be done by selecting a (τ_1, τ_2), where $\tau_1 \leq \tau_2$, so that the two threshold levels are

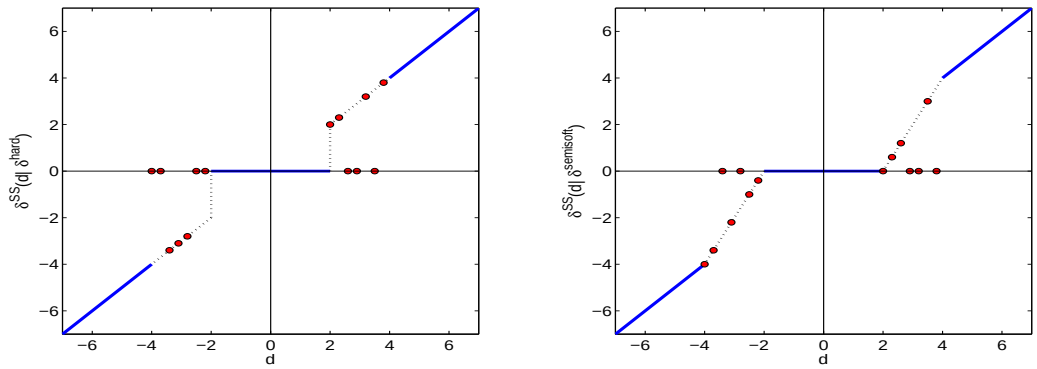
$$\begin{aligned} \lambda_1 &= \hat{\sigma}\sqrt{(2+\tau_1)\log N}, \\ \lambda_2 &= \hat{\sigma}\sqrt{(2+\tau_2)\log N}. \end{aligned} \tag{4.3.2}$$

Step-3. Estimate θ by θ^{SS} as

$$\theta^{SS}(d|\delta, \lambda_1, \lambda_2) = \begin{cases} 0, & \text{if } |d| < \lambda_1, \\ d\, f(\underset{\sim}{d}), & \text{if } \lambda_1 \leq |d| < \lambda_2, \\ d, & \text{if } |d| \geq \lambda_2, \end{cases} \tag{4.3.3}$$

where f is a decision function depending on neighbors $\underset{\sim}{d}$ of d that takes values 0 or 1 indicating whether d is excluded or included, respectively. We can generalize θ^{SS} with other choices of thresholding policy by applying the rule to the right-hand side of (4.3.3) accordingly.

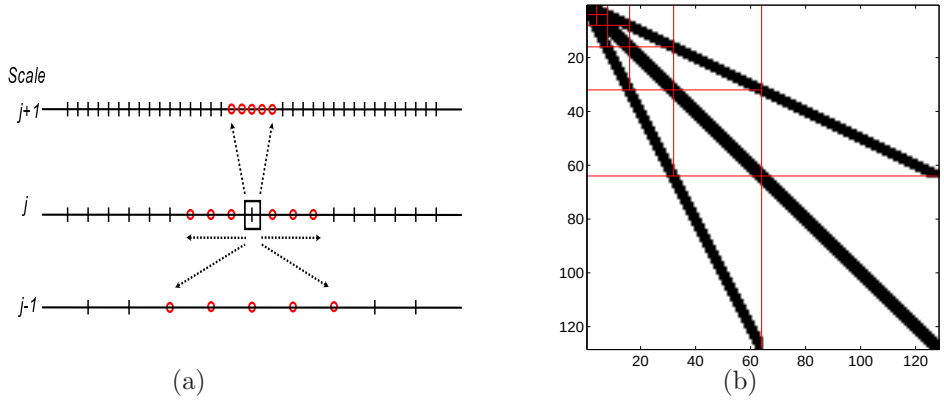
The coefficient-by-coefficient estimator at *Step-1* with thresholding level will be background thresholding of the SS rule. The rule in (4.3.3) defines labeled (*included* and *excluded*) and unlabeled (*undetermined*) coefficients. An illustration of θ^{SS} is given in Figure 26(a) where δ is hard thresholding and in Figure 26(b) where δ is semisoft thresholding.



(a) (b)

2.5in1.9in

Figure 26: Illustration of estimator δ^{SS} when the background shrinkage is (a) hard thresholding, δ^{hard} and (b) semisoft thresholding, $\delta^{semisoft}$; estimators for undetermined coefficients are red points. In the panel (c), a neighborhood for a wavelet coefficient in a rectangle is illustrated. It contains 6 neighbors at the same level and 5 neighbors each at the upper and lower levels.



(a) (b)

Figure 27: (a) A neighborhood for a wavelet coefficient in a rectangle is illustrated. It contains 6 neighbors at the same level and 5 neighbors each at the upper and lower levels. (b) A two-dimensional representation ($N = 128$) of the neighborhood in (a) with the column and row being wavelet coefficients d. The upper, middle, and lower diagonal (black) lines represent neighbors at the upper, same, and lower levels, respectively.

To obtain the estimate $\hat{f}$ of the decision function, the following semi-supervised learning mechanism is adopted. As discussed before, we treat included and excluded coefficients as labeled and coefficients for which $\lambda_1 \leq |d| \leq \lambda_2$ as unlabeled. Let s_1, s_2, and q be the number of included, excluded, and undetermined coefficients, respectively ($N = s_1 + s_2 + q$). Also let I_i and I_e be the index sets of indices for included and excluded coefficients, respectively. The learning mechanism consists of

the following three steps.

Step-1. For N wavelet coefficients, define an adjacency matrix W. The size of the matrix is N by N and possible entries are given as follows. For a constant $\rho \in \mathbb{R}^+$,

$$W_{nm} = \begin{cases} \exp\left(-\frac{(d_m - d_n)^2}{\rho}\right), & d_n \text{ and } d_m \in \text{the same } k \text{ nearest} \\ & \text{neighborhood,} \\ 0, & \text{otherwise.} \end{cases} \tag{4.3.4}$$

The parameter ρ is called the diffusion parameter and its choice depends on the assumption of how much the mutual influence between two coefficients diminishes as their distances increase. When ρ is ∞, W_{nm} simply becomes 1 if the two wavelet coefficients are in the same neighborhood and 0 otherwise. We note that the binary adjacency ($\rho = \infty$) performed well in practice. We fix the number of neighbors to be k (the k nearest neighbors). In the Figure 27(a), neighbors for one unlabeled coefficient are depicted: The neighbors are located at the same level, at its upper level and lower level. Figure 27(b) shows the corresponding neighborhood structure when $N = 126$. We will discuss how to set the value of k in Section 4.4.1.

Step-2. Construct the graph Laplacian matrix L (N by N). Rearrange W so that L has a block of L_1 ($s_1 + s_2$ by $s_1 + s_2$) corresponding to the included and the excluded coefficients and L_3 (q by q) corresponding to the undetermined coefficients:

$$L := \text{diag}\big(\sum_i W_{1i}, \ldots, \sum_i W_{Ni}\big) - W = \begin{pmatrix} L_1 & L_2 \\ L_2^T & L_3 \end{pmatrix}. \tag{4.3.5}$$

Step-3. Obtain an estimate $\hat{f}_j^{(u)}$ for the undetermined coefficients by

$$\hat{f}_j^{(u)} = \mathbf{1}\left(\sum_p \Delta_{jp} f_p^{(included)} > \frac{1}{2}\sum_p \Delta_{jp}\right), \tag{4.3.6}$$

where Δ is $-L_3^{-1}\, L_2^T$, and $f^{(included)}$ is $(s_1 + s_2)$ by 1 column vector with elements $f_p^{(included)} := \mathbf{1}(p \in I_i)$. Derivation of the decision rule (4.3.6) is provided in the Appendix A.3.

4.3.1 Interpretation of SS Rule

In the following section, properties of SS shrinkage are discussed.

K-nearest neighbor (k-NN) algorithm. The SS rule for an undetermined coefficient can be interpreted as *k-NN* algorithm when its neighborhood consists of labeled coefficients only. In this case, the decision condition (4.3.6) proves to be a *k-NN* algorithm that compares the number of included neighbors with the number of excluded neighbors when adjacency between the two neighbors is constant.

Theorem 4.3.1. *Suppose that undetermined coefficients d have neighbors that are either included or excluded with constant adjacency between the undetermined coefficients and their neighbors. Then, the SS estimator at undetermined coefficients d in (4.3.6) becomes a k-NN rule;*

$$\theta^{SS} = d\ \mathbf{1}\big(\#\{\textit{included neighbors of } d\} > \#\{\textit{excluded neighbors of } d\}\big).$$

The proof is provided in the Appendix. When the adjacency is not constant but varies across its neighbors, the rule becomes a weighted *k-NN* algorithm. When an undetermined coefficient has another undetermined coefficient as its neighbor, the SS shrinkage can not be expressed simply as *k-NN* but as an estimator based on geometry of all coefficients imposed by the complete adjacency structure. By adjusting the neighborhood in (4.3.4), and hence the adjacency matrix, the SS shrinkage generalizes the nearest neighborhood algorithms.

4.3.2 Optimality and Risk Analysis

Asymptotic near-optimality. We demonstrate that the SS rule achieves the same asymptotic convergence rate as its background thresholding rule. Without loss of generality, assume that δ is a hard thresholding estimator, δ^{hard}, that is proven to be near-optimal as in (4.3.9). Under the model in (4.3.1) and the squared error loss, the risk of diagonal projection (DP), $\hat{\theta} = d\,\mathbf{1}(|\theta| \geq \sigma)$, is

$$R(\hat{\theta}, \theta) = E||\hat{\theta} - \theta||_2^2 = \sum_{i=1}^{N} \min(\theta_i^2, \sigma^2) := R_{oracle}(\mathrm{DP}, \theta). \tag{4.3.7}$$

The risk is called the oracle risk (R_{oracle}) since it is unachievable because the true σ and θ are unknown. [30] prove that traditional hard thresholding estimators with threshold λ exhibit good asymptotic optimality when λ is sufficiently close to $\sigma\sqrt{2\log N}$, which means that for some $\gamma > 0$,

$$(1-\gamma)\log\log N \leq (\lambda/\sigma)^2 - 2\log N \leq o(\log N). \tag{4.3.8}$$

Following this result, it will be shown that when risks of $\delta^{hard}(d, \lambda_1)$ and $\delta^{hard}(d, \lambda_2)$ are within the $\log N$ factor of the ideal risk, the same holds for $\delta^{SS}(d|\delta^{hard}, \lambda_1, \lambda_2)$ rule.

Theorem 4.3.2. *Under the model in (4.3.1), the SS estimator $\delta^{SS}(d|\delta^{hard}, \lambda_1, \lambda_2)$, as defined in (4.3.3) above, satisfies the inequality*

$$R(\delta^{SS}, \theta) \leq M\Big\{\sigma^2 + \sum_{i=1}^{n} \min(\theta_i^2, \sigma^2)\Big\} \tag{4.3.9}$$

for an $M \sim \log N$ and all $\theta \in \mathbb{R}^N$, where λ_1 and λ_2 are sufficiently close to $\sigma\sqrt{2\log N}$ by (4.3.8).

The proof of the theorem, given in the Appendix A.5, is based on bounding $\delta^{SS}(d|\delta^{hard}, \lambda_1, \lambda_2)$ with $\delta^{hard}(d, \lambda_1)$ and $\delta^{hard}(d, \lambda_2)$. In the 'Oracle' notation of (4.3.7), the equation (4.3.9) is

$$R(\delta^{SS}, \theta) \leq L\Big\{\sigma^2 + R_{oracle}(\mathrm{DP}, \theta)\Big\}.$$

The inequality indicates that the SS estimator can mimic, within the $\log N$ factor, the performance of an oracle plus one extra parameter induced by unbiased estimation of θ. The theorem can be extended to any λ_1 and λ_2 that satisfy asymptotic near-optimality in the sense of (4.3.9). For example, the threshold $\hat{\sigma}\sqrt{(2+\tau)\log N}$ as in (4.3.2) satisfies the condition (4.3.8) whenever $\tau \sim o(1)$.

4.4 Examples

In this section we apply the proposed thresholding rules. First, we discuss the neighborhood structure and the selection of parameters λ_1 and λ_2. This consideration helps in automating the methodology. Next, we compare performance of the proposed rules to eight other popular methods (both global and adaptive). In the simulations we set the primary resolution level j_0 to be $\log_2(\log(N)) + 1$, following the asymptotic considerations given in Chapter 10 of [43].

4.4.1 Selection of Parameters

In any shrinkage task, selection of the parameters is essential for satisfactory performance of the model. It is also preferable to have guidelines that indicate how to select the parameters, thus providing automaticity to the shrinkage procedure (i.e., user intervention is not required). In SS shrinkage, the neighborhood structure for a coefficient and two threshold parameters should be preset. The characteristics of the signal can guide the shape and size of the neighborhood, while the selection of two threshold levels can be guided by the variance of the signal and the signal-to-noise ratio.

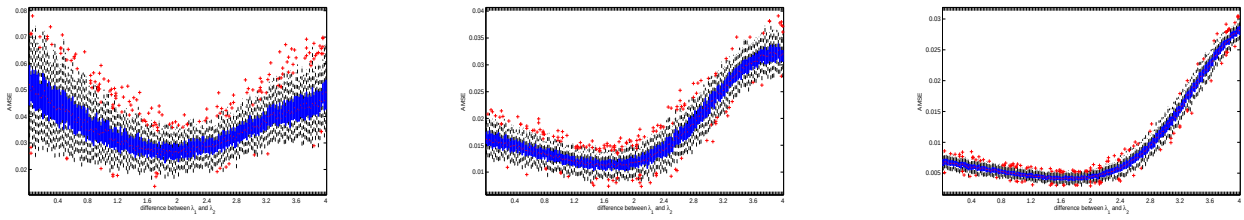
Number of neighbors. The number of neighbors at the upper and the lower levels should depend on the support of the decomposing wavelet. The support size is linked to the cone of influence of a wavelet coefficient [57]. Therefore, it is preferable to use wavelets with compact support and a sufficient number of vanishing moments.

As for the number of neighbors at the upper and lower levels, the suggestion is as follows: the more irregular the signal is locally, the more vertical neighbors the local wavelet coefficient should have. Prior information about smoothness of the signal can be useful in guiding this setup. It is also recommended that the two numbers of neighbors at the upper and lower levels be set to be equal (by the cone of influence of a wavelet coefficient), and in practice one horizontal level, one upper level, and one lower level will be sufficient for the levels of neighbors. Notice that neighbors of which positions (k) are not within indices of the levels (j) are truncated and that neighbors at the upper and lower levels of a coefficient can be asymmetric by one coefficient depending on the position of the coefficient.

Threshold levels. We recommend selection of λ_1 and λ_2 by computing optimal lower thresholds for a fixed background thresholding δ from the perspective of the average mean squared error (AMSE), as in [13], when true signals are available. In practice, the λ_2 can be selected as an under-fitting threshold. As an example, the background thresholding δ can be hard thresholding with the upper threshold level λ_2 as $\sqrt{2\log N}$. The choice for λ_2 is motivated by an argument used for universal thresholding. It was observed that the universal threshold is under-fitting, or larger than the optimal threshold. The λ_1 can be selected as an over-fitting threshold: an example is a threshold set by the cross-validation approach [63]. Other examples for λ_1 are as follows: λ_1 as $\arg\min_{0\le\lambda\le\sigma\sqrt{2\log N}} \mathrm{SURE}(\lambda, \mathbf{d})$ by SURE shrinkage paradigm[31] or as $\sigma\sqrt{2(1-\varepsilon)\log N}$ by a minimax approach [30].

For the sake of simulation, parameters λ_1 and λ_2, equivalently τ_1 and τ_2 as in (4.3.2), can be set to minimize AMSE according to the types of signal and the length. The lower threshold λ_1 can also be chosen accordingly when the upper threshold λ_2 is set to be $\sqrt{u\log N}$ for $u = 2, 3, 4$. Figure 28 shows the Piecewise-Regular signal at signal-to-noise ratio (SNR) = 5 and sample size $n = 256, 1024$, and 4096, smoothed by SS shrinkage with hard thresholding as its background shrinkage for

various values of (τ_1, τ_2) with τ_2 being 1. The optimal parameters for the signals of various sizes can be selected by minimizing AMSE. As the distance between λ_1 and λ_2 increases, the AMSE first decreases and then increases. The parameters are tabulated in Table 4 according to the signal length. Additional information about the numerical experiments is found in Section 4.4.2. It is noted that the optimal distance between two threshold levels to minimize AMSE varies as the length of signal changes.



(a) (b) (c)

Figure 28: AMSE for Piecewise-Regular signal of (a) length of 256, (b) length of 1024, and (c) length of 4096 at SNR = 5; The λ_1 is selected so that the distance between λ_1 and λ_2 is the minimizer of AMSE.

Table 4: Optimal parameters for semi-supervised shrinkage: column $\mathrm{SS}^{(2)}$ has $\lambda_2 = \sqrt{2 \log N}$; column $\mathrm{SS}^{(3)}$, $\lambda_2 = \sqrt{3 \log N}$; and column $\mathrm{SS}^{(4)}$, $\lambda_2 = \sqrt{4 \log N}$.

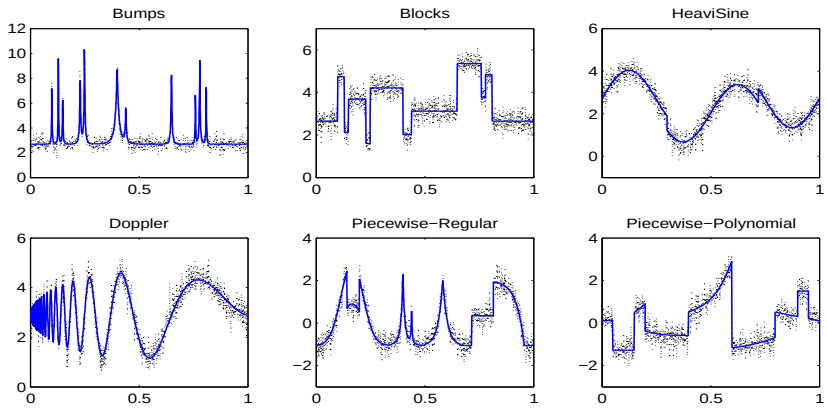
	$\mathrm{SS}^{(2)}$		$\mathrm{SS}^{(3)}$		$\mathrm{SS}^{(4)}$	
N	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2
256	2.1402	3.3302	2.1687	4.7096	3.3996	4.7096
512	2.6222	3.5322	2.6961	4.3261	3.4853	4.9953
1024	2.8333	3.7233	2.9701	4.5601	3.4555	5.2655
2048	3.0150	3.9050	3.1527	4.7827	3.6125	5.5225
4096	3.1887	4.0787	3.3653	4.9953	3.6581	5.7681

4.4.2 Simulations and Comparisons

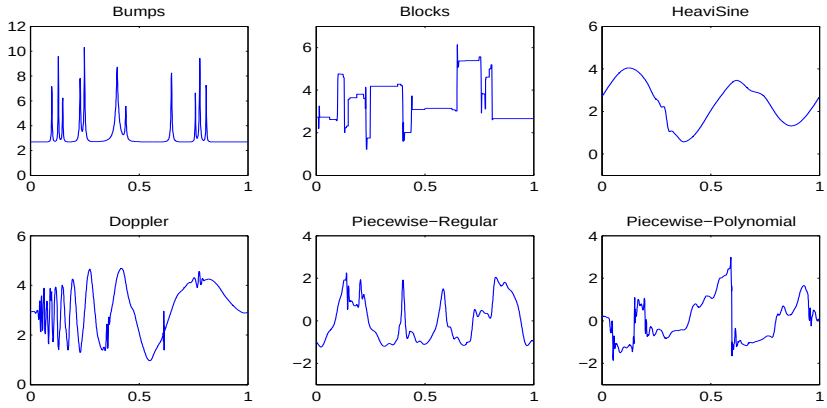
We present a simulation study of the performance of the SS method. The simulation is done with the "known truth", that is, with test functions and SNR specified. We compare AMSEs of SS and several other popular methods. Comparisons between the performance of SS and its own background thresholding method are also made.

Simulation setting. For our simulation study, six standard test functions (`Bumps, Blocks, HeaviSine, Doppler, Piecewise-Regular, Piecewise-Polynomial`) were perturbed by scaled normal noise in order to produce a preassigned SNR. For each method, the six test functions were generated at equally-spaced N points on the unit interval. The true and noisy signals are shown in Figure 29(a) at $N = 1024$ and SNR $= 4$. Denoised signals with SS shrinkage based on VisuShrink are presented in Figure 29(b). The decomposing wavelets were chosen in a standard way: Symmlet 8 for `HeaviSine, Doppler`, and `Piecewise-Polynomial`; Daubechies 6 for `Bumps` and `Piecewise-Regular`; and Haar for `Blocks`. The neighborhood structure and thresholding levels were set as follows: the number of horizontal neighbors was 4; the number of vertical neighbors was 12; the diffusion parameter ρ was 20; (τ_1, τ_2) was $(-0.2593, 0)$ for the SS rule based on Hybrid-SureShrink, while it was $(-0.4222, 0)$ for the SS rule based on VisuShrink; VisuShrink was set with the hard thresholding option. The accuracies of the estimated signals were measured by an AMSE over 1000 simulation runs. All computations were carried out using MATLAB with the Wavelab toolbox [14].

Comparisons with other methods. We compare SS shrinkage with several established wavelet-based denoising methods. In particular, we consider the classical term-by-term estimators *VisuShrink* of [30]; *Hybrid-SureShrink* of [31]; the scale invariant term-by-term Bayesian *ABE* method of [35]; *LPM* of [23]; the "leave-out-half" version of the *Cross-Validation* method of [63]; the term-by-term False Discovery Rate (*FDR*) method of [1]; the term-by-term Bayesian estimator *BAMS* of [74]; *NeighCoeff* of [17]; and finally *BlockJS* of Cai (1999). NeighCoeff and BlockJS represent classical estimators that incorporate the blocking procedure to achieve a better performance. We consider the CrossValidation method with the hard thresholding policy, the BlockJS with the option 'Augment' [5], and SS shrinkage with Hybrid-SureShrink as its background shrinkage.



(a)



(b)

Figure 29: All true signals (in blue line) and noised signals (in black dots) for simulation at the panel (a); estimated signals with SS rule based on VisuShrink at the panel (b).

Figure 30 presents the boxplots of the AMSE computed for the above 9 methods with $N = 1024$ at SNR $= 4$. We observe that the SS shrinkage performed comparably to other established methods. For some signals, it outperformed several methods such as CrossValidation, FDR, and BlockJS. In particular, it performed best for the HeaviSine signal and surpassed the background threshold Hybrid-SureShrink for all the six signals in terms of the sample average of AMSEs. Its ratios are shown in Figure 32(a).

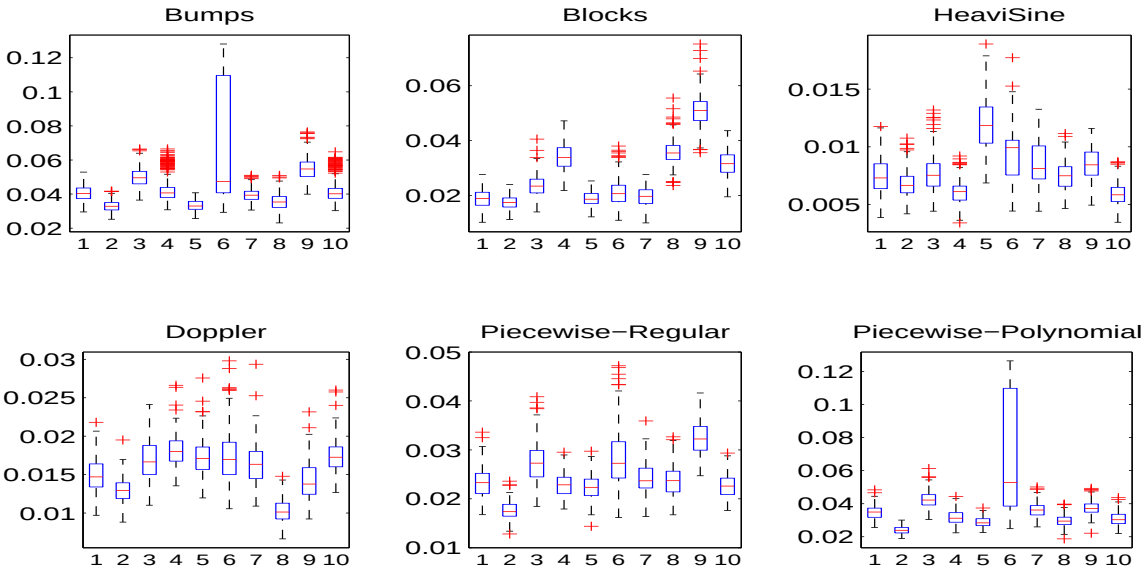
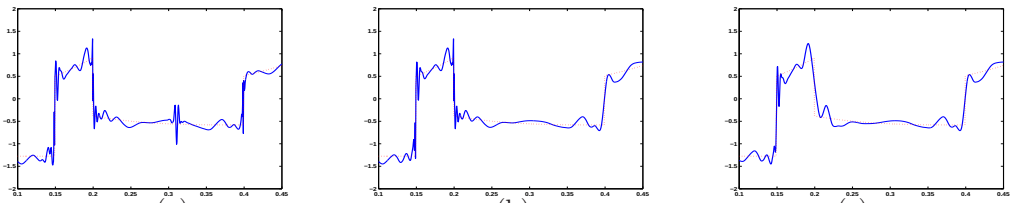


Figure 30: Boxplots of AMSE for (1) LPM (GAMMARULE), (2) BAMS, (3) VisuShrink, (4) Hybrid-SureShrink, (5) ABE, (6) CV, (7) FDR, (8) NC (9) BJS (10) SS rule based on Hybrid-SureShrink, with $n = 1024$ at SNR=4.



(a) (b) (c)

Figure 31: Estimated signals by semi-supervised shrinkage and its background shrinkage, where the true signal is Blocks with SNR $= 3$, $\lambda_1 = \sqrt{(2 - 0.9300)\log N}$, and $\lambda_2 = \sqrt{(2+1)\log N}$; (a) for hard thresholding with λ_1, $\delta^{hard}_{\lambda_1}$, (b) for semi-supervised shrinkage based on hard thresholding, $\delta^{SS}_{\lambda_1,\lambda_2}$, and (c) for hard thresholding with λ_2, $\delta^{hard}_{\lambda_2}$; Note that the signal in panel (b) compromises between the two signals in panel (a) and (b).

Comparisons with background thresholding. Figure 31 illustrates the comparisons between SS shrinkage and its background shrinkage (hard thresholding) with a part of standard "Blocks" signal and SNR $= 3$. Panels (a) and (c) present background estimators for $\lambda_1 = \sqrt{(2 - 0.9300)\log N}$ and $\lambda_2 = \sqrt{(2+1)\log N}$, respectively, while panel (b) gives an estimator from the associated SS shrinkage. In contrast to each

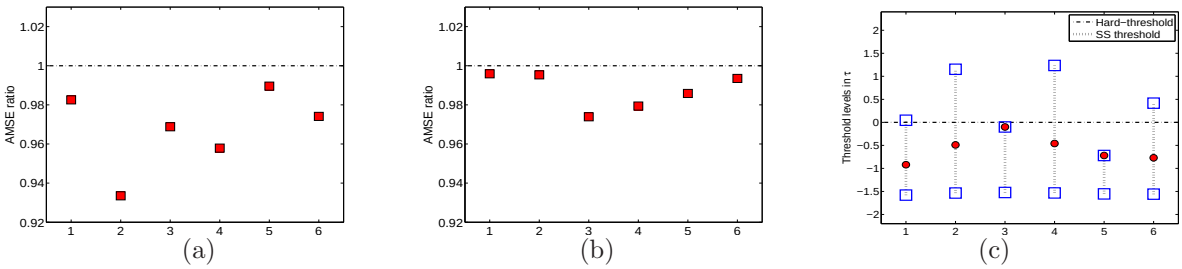
of the background estimators, the estimation by semi-supervised shrinkage compromises between the two signals in panel (a) and (b) and is more sensitive to the overall geometry of the true signal, which is evident by signal features and AMSE.

We observed that SS shrinkage outperformed its own background thresholding method. To demonstrate this point, the thresholding level λ^* for hard thresholding was preset in such a way that it minimized AMSE from 1000 simulations. The simulation was conducted using signals of size $N = 2048$ and with noise such that SNR is 5. Then, the AMSE for each signal was computed based on 1000 simulations. The neighborhood structure was characterized as follows: the number of horizontal neighbors was 20; the number of vertical neighbors was 20; the diffusion parameter ρ was 5000. The threshold levels for the SS rule were obtained by perturbing λ^* with estimated noise level $\hat{\sigma}$: $\lambda_1 = \lambda^* - k_1\hat{\sigma}$ and $\lambda_1 = \lambda^* + k_2\hat{\sigma}$; $k_1 = k_2 = 0.1$ for Bumps; $k_1 = k_2 = 0.2$ for Blocks, Doppler, and Piecewise-Polynomial; $k_1 = 0.5, k_2 = 0$ for HeaviSine; $k_1 = 0.3, k_2 = 0$ for Piecewise-Regular. The estimate of the noise level was calculated by the median absolute deviation of the wavelet coefficients in the finest level of detail. The threshold levels in terms of τ_1 and τ_2 are shown in Figure 32(c) and AMSE ratios of the SS rule compared to the background method are shown in Figure 32(b). This simulation shows that the SS rule outperformed the background method in terms of AMSE for the six test signals.

4.4.3 An Example in Atomic Force Microscopy

To illustrate the performance of the SS shrinkage method proposed here, we used measurements in atomic force microscopy (AFM).

AFM is a type of scanned proximity probe microscopy (SPM) that can measure the adhesion strength between two materials at the nanonewton scale [11]. In AFM, a cantilever beam is adjusted until it bonds with the surface of a sample, and then the



(a) (b) (c)

Figure 32: (a) Comparison (AMSE ratio) of the SS rule with its background thresholding (Hybrid-SureShrink as for Figure 30); (b) comparison (AMSE ratio) of the SS rule with its background thresholding (hard thresholding with a threshold level that minimizes AMSE) at $N = 2048$ and SNR=5; the SS rule outperformed its background thresholding for all test signals; (c) threshold levels of the background hard thresholding and the SS rule in terms of τ; numbers represent: 1, Bumps; 2, Blocks; 3, HeaviSine; 4, Doppler; 5, Piecewise-Regular; 6, Piecewise-Polynomial.

force required to separate the beam and sample is measured from the beam deflection. Beam vibration can be caused by factors such as thermal energy of the surrounding air or the footsteps of someone outside the laboratory. The vibration of a beam acts as noise on the deflection signal; in order for the data to be useful this noise must be removed.

The AFM data from the adhesion measurements between carbohydrate and the cell adhesion molecule (CAM) E-Selectin was collected by Bryan Marshal from the BME Department at Georgia Institute of Technology. The detailed technical description is provided in Marshall, McEver, and Zhu (2001). Researchers are interested in precise slope at the ramp (dotted line in Figure 33).

In Figure 33 the top panel shows the original noisy data. We consider the SS with VisShrink. The middle panel shows the SS estimate with $(\tau_1, \tau_2) = (0.5, 1.5)$, while the bottom panel shows LPM estimate with $(\tau_1, \tau_2) = (0.7, 1.5)$. The sample size was $n = 2^{11}$ and Symmlet 8-tap filter was used to obtain the estimate. We observe that the latter estimate exhibits slightly smoother behavior, especially in the long-middle part without oversmoothing the "ramp-like" structure which is the feature of interest here.

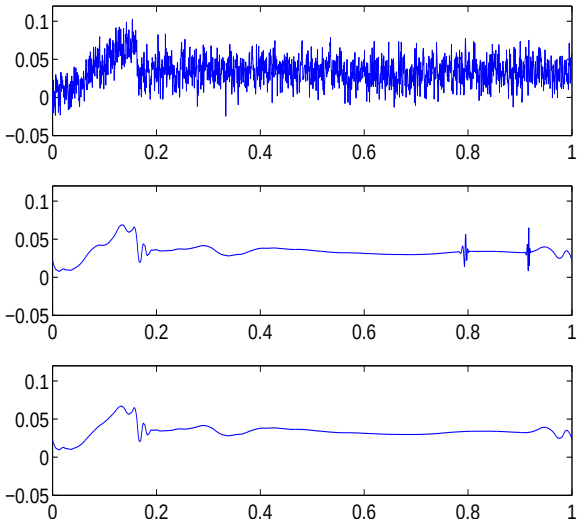


Figure 33: Original AFM measurements (top), SS estimator based on VisuShrink with $(\tau_1, \tau_2) = (0.5, 1.5)$ (middle), SS estimator based on VisuShrink with $(\tau_1, \tau_2) = (0.7, 1.5)$ (bottom).

CHAPTER V

CONCLUSIONS

Firstly, we derived joint estimators of time-changing Hurst exponent $H(t)$ and its variance coefficient C for mBm under independent white noise. Properties of the estimator such as bias and asymptotic distribution are shown utilizing generalized regression theory. The effectiveness of the approach was verified through numerical experiments in comparison with several other approaches. Its application to real EEG data set was conducted supporting clinical observations of the patient.

Secondly, evidence for deviation from monofractality, which is contained in the partition function and MFS, can be used to develop a paradigm for formal testing for deviation from monofractality. In the third chapter, we introduced a measure of deviation from monofractality by using left slope as one of the descriptors of wavelet-based MFS of a signal. We constructed a test procedure based on parametric bootstrapping of sampled fBm signals which are monofractals. This produces a distribution of left slopes consistent with the assumption of monofractality. Our simulation results indicate that the testing procedure effectively separates multifractal Brownian motion signals from fBm signals. Its effectiveness is also shown in the real life example of turbulence and DNA sequences, first as an example of multifractal, and second as an example of monofractal.

Thirdly, we developed a new method for wavelet-filtering of noisy signals that combines semi-supervised learning with wavelet thresholding. Construction of labeled and unlabeled coefficients for semi-supervised learning and its justification is provided. Both interpretations of the method and its theoretical properties have been explored. We have demonstrated that the performance of semi-supervised shrinkage

is comparable to several existing shrinkage methods in terms of average mean-squared error. Evidence has also been provided to show that the method preserves geometries of signals better than underlying background shrinkage method and that it improves its own background shrinkage method. For future research, we envision combining different decision rules of a general type, not necessarily thresholding, with machine learning algorithms. The implementation is fast in computations for sample sizes up to 4096 and feasible for 8192. We want to explore computationally efficient algorithms for samples of larger sizes.

APPENDIX A

PROOFS AND DERIVATION

A.1 Proof of Proposition 2.1.1

Let us denote $G(t) = \sqrt{C}K\big(2H(t)\big)/\sqrt{2}$ for the sake of simplicity. We have the covariance $\mathbb{E}\Big[V_{\mathbf{a}}(\frac{j}{N})V_{\mathbf{a}}(\frac{j'}{N})\Big]$ be

$$\begin{aligned}
&\sum_{q,q'} a_q a_{q'} \left(G(\tfrac{j-q}{N}) G(\tfrac{j'-q'}{N}) \int \frac{\exp(i\frac{j-q}{N}\lambda) - 1}{|\lambda|^{H(\frac{j-q}{N})+1/2}} \times \frac{\exp(-i\frac{j'-q'}{N}\lambda) - 1}{|\lambda|^{H(\frac{j'-q'}{N})+1/2}} d\lambda + \sigma^2 \mathbb{E}[\varepsilon(\tfrac{j-q}{N})\varepsilon(\tfrac{j'-q'}{N})] \right) \\
= \; & \sum_{q,q'} a_q a_{q'} \frac{G\big(\frac{j-q}{N}\big) G\big(\frac{j'-q'}{N}\big)}{N^{H(\frac{j-q}{N})+H(\frac{j'-q'}{N})}} \int \frac{\exp\big(i(j-q)u\big) - 1}{|u|^{H(\frac{j-q}{N})+1/2}} \times \frac{\exp\big(-i(j'-q')u\big) - 1}{|u|^{H(\frac{j'-q'}{N})+1/2}} du \\
& + \sigma^2 \sum_{q-q'=j-j'} a_q a'_q \qquad (\text{by } u = \frac{\lambda}{N}).
\end{aligned}$$

By Taylor's expansion and Hölderian order η of $H(t)$, we approximate

$$\begin{aligned}
& G\big(\frac{j-q}{N}\big) = G(t) + \mathcal{O}(\epsilon^\eta), \qquad\qquad G\big(\frac{j'-q'}{N}\big) = G(t') + \mathcal{O}(\epsilon^\eta), \\
& \frac{1}{N^{H(\frac{j-q}{N})+H(\frac{j'-q'}{N})}} = \frac{1}{N^{H(t)+H(t')}}\big(1 + \mathcal{O}(\epsilon^\eta \log N)\big), \\
& \frac{1}{|u|^{H(\frac{j-q}{N})+H(\frac{j'-q'}{N})}} = \frac{1}{N^{H(t)+H(t')}}\big(1 + \mathcal{O}(\epsilon^\eta \log u)\big).
\end{aligned}$$

Using also $\epsilon^\eta \to 0$ as N goes to infinity, the covariance becomes

$$\sum_{q,q'} a_q a_{q'} \frac{G(t)G(t')}{N^{H(t)+H(t')}} \int \frac{\exp\big(i(j-q)u\big) - 1}{|u|^{H(t)+1/2}} \times \frac{\exp\big(-i(j'-q')u\big) - 1}{|u|^{H(t')+1/2}} du + \sigma^2 \sum_{q-q'=j-j'} a_q a'_q + \mathcal{O}(\epsilon^\eta \log N). \quad \text{(A.1.1)}$$

By the property of $\mathbf{a}$, the equation (A.1.1) becomes

$$\sum_{q,q'} a_q a_{q'} \frac{G(t)G(t')}{N^{H(t)+H(t')}} \int \frac{\cos\big((j-q-j'+q')u\big) - 1}{|u|^{H(t)+H(t')+1}} du + \sigma^2 \sum_{q-q'=j-j'} a_q a'_q + \mathcal{O}(\epsilon^\eta \log N).$$

Since

$$K(\alpha) \int \frac{1 - \cos\big(\kappa u\big)}{|u|^{\alpha+1}} du = |\kappa|^\alpha, \qquad \forall \kappa,\ 0 < \alpha < 2,$$

we write:

$$\begin{aligned}
\mathbb{E}\Big[V_{\mathbf{a}}(\frac{j}{N})V_{\mathbf{a}}(\frac{j'}{N})\Big] &= -\sum_{q,q'} a_q a_{q'} \frac{G(t)G(t')}{N^{H(t)+H(t')}} \frac{|j-q-j'+q'|^{H(t)+H(t')}}{K\big(H(t)+H(t')\big)} + \sigma^2 \sum_{q-q'=j-j'} a_q a'_q + \mathcal{O}(\epsilon^\eta \log N) \\
&= -\frac{Cg\big(H(t),H(t')\big)}{N^{H(t)+H(t')}} \sum_{q,q'} \frac{a_q a_{q'}}{2} |j-q-j'+q'|^{H(t)+H(t')} + \sigma^2 \sum_{q-q'=j-j'} a_q a'_q + \mathcal{O}(\epsilon^\eta \log N) \\
&= \frac{Cg(H(t),H(t'))}{N^{H(t)+H(t')}} \pi_{\mathbf{a},H(t)/2+H(t')/2}(j'-j) + \sigma^2 \sum_{q-q'=j-j'} a_q a'_q + \mathcal{O}(\epsilon^\eta \log N).
\end{aligned}$$

A.2 Proof of Proposition 2.1.2

By denoting $A = \frac{1}{|\nu(t)|}\sum_{j\in\nu(t)}\sum_{q,q'}^{l} a_q a'_q W(\frac{j-q}{N})\varepsilon(\frac{j-q'}{N})$ and $B = \frac{1}{|\nu(t)|}\sum_{j\in\nu(t)}\big(\sum_{q}^{l} a_q \varepsilon(\frac{j-q}{N})\big)^2$, $\log S'(t,\mathbf{a})$ can be written as

$$\begin{aligned}
\log S'(t,\mathbf{a}) &= \log\frac{1}{|\nu(t)|}\sum_{j\in\nu(t)} V'_{\mathbf{a}}(\tfrac{j}{N})^2 \\
&= \log\frac{1}{|\nu(t)|}\sum_{j\in\nu(t)}\Big((\sum_{q}^{l} a_q W(\tfrac{j-q}{N}))^2 + 2\sigma\sum_{q,q'}^{l} a_q a'_q W(\tfrac{j-q}{N})\varepsilon(\tfrac{j-q'}{N}) + \sigma^2(\sum_{q}^{l} a_q\varepsilon(\tfrac{j-q}{N}))^2\Big) \\
&= \log S(t,\mathbf{a}) + \Big(\sigma\frac{A}{S(t,\mathbf{a})} + \sigma^2\frac{B}{S(t,\mathbf{a})}\Big)\big(1+o(1)\big). \qquad \text{(by Taylor's expansion)}
\end{aligned}$$

Also, $\log\mathbb{E}\big[S'(t,\mathbf{a})\big]$ is expressed as

$$\begin{aligned}
\log\mathbb{E}[S'(t,\mathbf{a})] &= \log\big(\mathbb{E}[S(t,\mathbf{a})] + \sigma^2\sum_{q}^{l} a_q^2\big) \\
&= \log\big(\mathbb{E}[S(t,\mathbf{a})]\big) + \Big(\frac{\sigma^2\sum_{q}^{l} a_q^2}{\mathbb{E}\big[S(t,\mathbf{a})\big]}\Big)\big(1+o(1)\big). \qquad \text{(by Taylor's expansion)}
\end{aligned}$$

Using independence of $W(t)$ and $\varepsilon(t)$, properties of white noise $\varepsilon(t)$, and convergence of $S(t,\mathbf{a})$ to $\mathbb{E}\big[S(t,\mathbf{a})\big]$ almost surely as $N\to\infty$, we have consequently $A\to 0$, $B\to\sum_{q}^{l} a_q^2$, and

$$\begin{aligned}
\log S'(t,\mathbf{a}) - \log\mathbb{E}[S'(t,\mathbf{a})] &= \log S(t,\mathbf{a}) - \log\mathbb{E}[S(t,\mathbf{a})] + o(1) \\
&= \log\Big(\frac{1}{|\nu(t)|}\sum_{j\in\nu(t)}\frac{V_{\mathbf{a}}(\frac{j}{N})^2}{\mathbb{E}\big[S(t,\mathbf{a})\big]}\Big) + o(1) \qquad \text{(by def. of } S(t,\mathbf{a})\text{)} \\
&= \log\Big(\frac{1}{|\nu(t)|}\sum_{j\in\nu(t)}\frac{V_{\mathbf{a}}(\frac{j}{N})^2}{\mathbb{E}\big[V_{\mathbf{a}}(\frac{j}{N})^2\big]}\Big) + o(1) \qquad \text{(by stationarity of } V_{\mathbf{a}}\text{)} \\
&= \left(\Big(\frac{1}{|\nu(t)|}\sum_{j\in\nu(t)}\frac{V_{\mathbf{a}}(\frac{j}{N})^2}{\mathbb{E}\big[V_{\mathbf{a}}(\frac{j}{N})^2\big]}\Big) - 1\right)\big(1+o(1)\big) \qquad \text{(by Taylor's expansion)} \\
&= V(t,\mathbf{a})\big(1+o(1)\big).
\end{aligned}$$

A.3 Derivation of Equation 4.3.6

Define $f^{(included)}$ and $f^{(excluded)}$ as (s_1+s_2) by 1 column vector, respectively,

$$f_p^{(included)} := \begin{cases} 1, & \text{if } p\in I_i, \\ 0, & \text{if not,} \end{cases} \quad \text{and} \quad f_p^{(excluded)} := \begin{cases} 1, & \text{if } p\in I_e, \\ 0, & \text{if not.} \end{cases} \tag{A.3.1}$$

By equation (4.2.3), $-L_3^{-1}L_2^T f^{(l)}$ is a measure to support the label $f^{(l)}$ against all the other labels. Then, the jth undetermined coefficient is to be included if

$$[-L_3^{-1}L_2^T f^{(included)}]_j > [-L_3^{-1}L_2^T f^{(excluded)}]_j, \tag{A.3.2}$$

or, equivalently, (recall Δ is denoted to be $-L_3^{-1}L_2^T$)

$$\sum_p \Delta_{jp} f_p^{(included)} > \sum_p \Delta_{jp} f_p^{(excluded)}. \tag{A.3.3}$$

By definition of $f^{(included)}$ and $f^{(excluded)}$ in equation (A.3.1), it follows

$$f_p^{(included)} = 1 - f_p^{(excluded)} \quad \text{for all } p \in I_i \cup I_e. \tag{A.3.4}$$

From equations (A.3.4) and (A.3.3), equation (4.3.6) follows.

A.4 Proof of Theorem 4.1

Since neighbors of undetermined coefficients are either included or excluded, the matrix L_3 in (4.3.5) becomes a diagonal matrix: $L_3 = \text{diag}(\sum_i W_{ij})$. By definition, the elements of L_2 become $-W_{ij}$. The right hand side of the decision condition for the b-th undetermined wavelet coefficient in (A.3.2) writes:

$$[-L_3^{-1}\ L_2^T\ f^{(included)}]_b = -\sum_{a\in I_i}\sum_h W_{hb}(-W_{ab}) = \sum_h W_{hb}\sum_{a\in I_i} W_{ab}.$$

Similarly, the right-hand side of (A.3.2) becomes $\sum_h W_{hb}\sum_{a\in I_e} W_{ab}$. Consequently, the decision rule for the b-th undetermined wavelet coefficient is

$$\hat{f}_b^{(u)} = \mathbf{1}\Big(\sum_{a\in I_i} W_{ab} > \sum_{a\in I_e} W_{ab}\Big),$$

which can be simplified into a *k-NN* algorithm because of constant adjacency among neighbors (W_{ab} = constant, $\forall a, b$):

$$\hat{f}_b^{(u)} = \mathbf{1}\big(\#\{\text{included neighbors of } b\} > \#\{\text{excluded neighbors of } b\}\big).$$

A.5 Proof of Theorem 4.2

Let $\delta_{\lambda_1}^{hard}$ and $\delta_{\lambda_2}^{hard}$ be two hard thresholding estimators, respectively. By the definition of δ^{SS} at (4.3.3), for all d we have

$$|\delta_{\lambda_2}^{hard}(d)| < |\delta^{SS}(d|\delta^{hard}, \lambda_1, \lambda_2)| < |\delta_{\lambda_1}^{hard}(d)|. \tag{A.5.1}$$

Then, by (A.5.1), the square error loss of δ^{SS} at d is bounded,

$$\left(\theta - \delta^{SS}(d|\delta^{hard}, \lambda_1, \lambda_2)\right)^2 < \max\left\{\left(\theta - \delta^h_{\lambda_1}(d)\right)^2, \left(\theta - \delta^{hard}_{\lambda_2}(d)\right)^2\right\}.$$

The risk of δ^{SS} is bounded;

$$R(\delta^{SS}, \theta) < R(\delta^{hard}_{\lambda_1}, \theta) + R(\delta^{hard}_{\lambda_2}, \theta).$$

Since risks of $\delta^{hard}_{\lambda_1}$ and $\delta^{hard}_{\lambda_2}$ are within the $\log N$ factor of the ideal risk, it follows that $R(\delta^{SS}, \theta)$ is within the $\log N$ factor of it.

REFERENCES

[1] Abramovich, F. and Benjamini, Y., "Thresholding of wavelet coefficients as multiple hypotheses testing procedure," in *In Wavelets and Statistics, Lecture Notes in Statistics 103, Antoniadis*, pp. 5–14, Springer-Verlag, 1995.

[2] Abry, P., Flandrin, P., Taqqu, M., and Veitch, D., "Wavelets for the analysis, estimation and synthesis of scaling data," *Self-Similar Network Traffic and Performance Evaluation*, pp. 39–88, 2000.

[3] Abry, P., Flandrin, P., Taqqu, M., and Veitch, D., "Self-similarity and long-range dependence through the wavelet lens," *Long-range dependence: theory and applications*, pp. 527–56, 2002.

[4] Abry, P. and Veitch, D., "Wavelet analysis of long-range-dependent traffic," *IEEE transactions on information theory*, vol. 44, no. 1, pp. 2–15, 1998.

[5] Antoniadis, A., Bigot, J., and Sapatinas, T., "Wavelet estimators in nonparametric regression: A comparative simulation study," *Journal of Statistical Software*, vol. 6, no. i06, pp. 1–83, 2001.

[6] Arneodo, A., Bacry, E., Jaffard, S., and Muzy, J., "Singularity spectrum of multifractal functions involving oscillating singularities," *Journal of fourier analysis and applications*, vol. 4, no. 2, pp. 159–174, 1998.

[7] Arneodo, A., d'Aubenton Carafa, Y., Bacry, E., Graves, P., Muzy, J., and Thermes, C., "Wavelet based fractal analysis of DNA sequences," *Physica D: Nonlinear Phenomena*, vol. 96, no. 1-4, pp. 291–320, 1996.

[8] Barnard, G., "Discussion of Hurst," *Proc. Inst. Civ. Eng*, vol. 5, no. 5, pp. 552–553, 1956.

[9] Belkin, M. and Niyogi, P., "Semi-supervised learning on riemannian manifolds," *Machine Learning*, vol. 56, no. 1-3, pp. 209–239, 2004.

[10] Benassi, A., Cohen, S., and Istas, J., "Identifying the multifractional function of a gaussian process," *Statist. Probab. Lett.*, vol. 39, pp. 337–345, 1998.

[11] Binnig, G., Quate, C. F., and Gerber, C., "Atomic force microscope," *Phys. Rev. Lett.*, vol. 56, pp. 930–933, Mar 1986.

[12] Brockwell, P. and Davis, R., *Time series: theory and methods*. Springer Verlag, 2009.

[13] Bruce, A. and Gao, H.-Y., "Waveshrink: Shrinkage functions and thresholds," in *Proc. SPIE*, p. proc., 1995.

[14] Buckheit, J. and Donoho, D., "Wavelab and reproducible research," pp. 55–81, Springer-Verlag, 1995.

[15] Burt, P. and Adelson, E., "The Laplacian pyramid as a compact image code," *IEEE Transactions on communications*, vol. 31, no. 4, pp. 532–540, 1983.

[16] Cai, T. T., "Adaptive wavelet estimation: a block thresholding and oracle inequality approach," *The Annals of Statistics*, vol. 27, pp. 898–924, 1999.

[17] Cai, T. and Silverman, B., "Incorporating information on neighbouring coefficients into wavelet estimation," *Sankhyā: The Indian Journal of Statistics, Series B*, vol. 63, no. 2, pp. 127–148, 2001.

[18] Calvet, L. and Fisher, A., "Multifractality in asset returns: theory and evidence," *Review of Economics and Statistics*, vol. 84, no. 3, pp. 381–406, 2002.

[19] Chipman, H. A., Kolaczyk, E. D., and Mcculloch, R. E., "Adaptive bayesian wavelet shrinkage," *Journal of the American Statistical Association*, vol. 92, pp. 1413–1421, 1997.

[20] Coeurjolly, J.-F., "Estimating the parameters of a fractional brownian motion by discrete variations of its sample paths," *Statistical Inference for Stochastic Processes*, vol. 4, pp. 199–227, May 2001.

[21] Coeurjolly, J.-F., "Identification of multifractional brownian motion," *Bernoulli*, vol. 11, pp. 987–1008, 2005.

[22] Coeurjolly, J., "Erratum: Identification of multifractional Brownian motion," *Bernoulli*, vol. 12, no. 2, p. 381, 2006.

[23] Cutillo, L., Jung, Y., Ruggeri, F., and Vidakovic, B., "Larger posterior mode wavelet thresholding and applications," *Journal of Statistical Planning and Inference*, vol. 138, no. 12, pp. 3758–3773, 2008.

[24] Daubechies, I., "Orthonormal bases of compactly supported wavelets," *Communications on pure and applied mathematics*, vol. 41, no. 7, pp. 909–996, 1988.

[25] Daubechies, I. and Lagarias, J., "Two-scale difference equations. I. Existence and global regularity of solutions," *SIAM Journal on Mathematical Analysis*, vol. 22, p. 1388, 1991.

[26] Daubechies, I. and Lagarias, J., "Two-scale difference equations II. Local regularity, infinite products of matrices and fractals," *SIAM Journal on Mathematical Analysis*, vol. 23, p. 1031, 1992.

[27] Daubechies, I., *Ten Lectures on Wavelets*. Soc for Industrial & Applied Math, December 1992.

[28] Derado, G., Lee, K., Nicolis, O., Bowman, F., Newell, M., Rugger, F., and Vidakovic, B., "Wavelet-based 3-D Multifractal Spectrum with Applications in Breast MRI Images," *Bioinformatics Research and Applications*, vol. 4983, pp. 281–292, 2008.

[29] Deriche, M. and Tewfik, A., "Signal modeling with filtered discrete fractional noise processes," *IEEE Transactions on Signal Processing*, vol. 41, no. 9, pp. 2839–2849, 1993.

[30] Donoho, D. L. and Johnstone, I. M., "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.

[31] Donoho, D. L. and Johnstone, I. M., "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association*, vol. 90, pp. 1200–1224, 1995.

[32] Donoho, D. and Johnstone, I., "Minimax estimation via wavelet shrinkage," *Annals of statistics*, pp. 879–921, 1998.

[33] Durbin, R., Eddy, S., Krogh, A., and Mitchison, G., *Biological sequence analysis: Probabilistic models of proteins and nucleic acids*. Cambridge Univ Pr, 1998.

[34] Efron, B. and Tibshirani, R., *An introduction to the bootstrap*. Chapman & Hall, 1997.

[35] Figueiredo, M. and Nowak, R., "Wavelet-based image estimation: an empirical Bayes approach using Jeffrey's noninformative prior," *IEEE Transactions on Image Processing*, vol. 10, no. 9, pp. 1322–1331, 2001.

[36] Flandrin, P., "On the spectrum of fractional Brownian motions," *IEEE Transactions on Information Theory*, vol. 35, no. 1, pp. 197–199, 1989.

[37] Frisch, U., "Fully developed turbulence and intermittency," *Annals of the New York Academy of Sciences*, vol. 357, no. 1, pp. 359–367, 1980.

[38] Gao, H., "Wavelet shrinkage denoising using the non-negative garrote," *Journal of Computational and Graphical Statistics*, vol. 7, no. 4, pp. 469–488, 1998.

[39] Gao, J., Cao, Y., Tung, W., and Hu, J., *Multiscale analysis of complex time series: integration of chaos and random fractal theory, and beyond*. Wiley-Interscience, 2007.

[40] Gonçalves, P., Riedi, R., and Baraniuk, R., "A simple statistical analysis of wavelet-based multifractal spectrum estimation," in *Proceedings 32nd Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 287–291, Citeseer, 1998.

[41] Granger, C. and Joyeux, R., "An introduction to long-memory time series models and fractional differencing," *Journal of time series analysis*, vol. 1, no. 1, pp. 15–29, 1980.

[42] Hall, P., Kerkyacharian, G., and Picard, D., "Block threshold rules for curve estimation using kernel and wavelet methods," *Annals of Statistics*, vol. 26, pp. 922–942, 1998.

[43] Hardle, W., Kerkyacharian, G., and Tsybakov, A. B., *Wavelets, Approximation and Statistical Applications*, vol. 129 of *Lecture Notes in Statistics*. New York: Springer-Verlag, 1998.

[44] Hosking, J., "Fractional differencing," *Biometrika*, vol. 68, no. 1, p. 165, 1981.

[45] Hurst, H., "Long-term storage capacity of reservoirs," *Transactions of the American Society of Civil Engineers*, vol. 116, no. 770-799, p. 35, 1951.

[46] Istas, J. and Lang, G., "Quadratic variations and estimation of the local hölder index of a gaussian process," *Ann. Inst. H. Poincaré Probab. Statist.*, vol. 33, pp. 407–436, 1997.

[47] Ivanov, P., Amaral, L., Goldberger, A., Havlin, S., Rosenblum, M., Struzik, Z., and Stanley, H., "Multifractality in human heartbeat dynamics," *Letters to Nature*, vol. 399, pp. 461–465, 1999.

[48] Jaffard, S., "Local behavior of Riemann's function," *Contemporary Mathematics*, vol. 189, pp. 287–287, 1995.

[49] Johnstone, I. M. and Silverman, B. W., "Wavelet threshold estimators for data with correlated noise," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 59, pp. 319–351, 1997.

[50] Kent, J. and Wood, A., "Estimating the fractal dimension of a locally self-similar gaussian process by using increments," *Journal of the Royal Statistical Society. Series B. Methodological*, vol. 59, pp. 679–699, 1997.

[51] Kolmogorov, A., "A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high Reynolds number," *J. Fluid Mech*, vol. 13, pp. 82–85, 1962.

[52] Kolmogorov, A., "The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers," *Proceedings: Mathematical and Physical Sciences*, pp. 9–13, 1991.

[53] Kuklinski, W., "Utilization of fractal image models in medical image processing," *Fractals*, vol. 2, pp. 363–369, 1994.

[54] Lamperti, J., "Semi-stable processes," *Transactions of the American Mathematical Society*, p. 62, 1962.

[55] Mallat, S., "Multiresolution approximations and wavelet orthonormal bases of L 2 (R)," *Transactions of the American Mathematical Society*, vol. 315, no. 1, pp. 69–87, 1989.

[56] Mallat, S. and others, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 11, no. 7, pp. 674–693, 1989.

[57] Mallat, S., *A Wavelet Tour of Signal Processing, Second Edition (Wavelet Analysis & Its Applications)*. Academic Press, September 1999.

[58] Mandelbrot, B., "Limit theorems on the self-normalized range for weakly and strongly dependent processes," *Probability Theory and Related Fields*, vol. 31, no. 4, pp. 271–285, 1975.

[59] Mandelbrot, B. and Van Ness, J., "Fractional Brownian motions, fractional noises and applications," *SIAM review*, vol. 10, no. 4, pp. 422–437, 1968.

[60] Mandelbrot, B. and Wallis, J., "Noah, Joseph, and operational hydrology," *Water Resources Research*, vol. 4, no. 5, pp. 909–918, 1968.

[61] Morlet, J., Arens, G., Fourgeau, E., and Glard, D., "Wave propagation and sampling theory?Part I: Complex signal and scattering in multilayered media," *Geophysics*, vol. 47, p. 203, 1982.

[62] Morlet, J. and Grossman, A., "Decomposition of Hardy functions into square integrable wavelets of constant shape," *SIAM Journal on Mathematical Analysis*, vol. 15, pp. 723–736, 1984.

[63] Nason, G. P., "Wavelet shrinkage using cross-validation," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, pp. 463–479, 1996.

[64] Peltier, R. and Levy Vehel, J., "A new method for estimating the parameter of fractional Brownian motion," *Rapport de Recherche n. 2396*, 1994.

[65] Peltier, R., Véhel, J., and others, "Multifractional Brownian motion: definition and preliminary results," *Rapport de Recherche n. 2645*, 1995.

[66] Philippe, A. and Thilly, E., "Identification of a locally self-similar Gaussian process by using convex rearrangements," *Methodology and Computing in Applied Probability*, vol. 4, no. 2, pp. 195–209, 2002.

[67] Riedi, R., "Multifractal processes," *Theory and applications of long-range dependence*, pp. 625–716, 2003.

[68] Riedi, R., Crouse, M., Ribeiro, V., and Baraniuk, R., "A multifractal wavelet model with application to network traffic," *IEEE Transactions on Information Theory*, vol. 45, no. 3, pp. 992–1018, 1999.

[69] Shapiro, A., "Statistical inference of moment structures," *Handbook of Latent Variable and Related Models*, pp. 229–260, 2007.

[70] Shi, B., Moloney, K., Leonard, K. V., Jacko, J., and Sainfort, F., "Multifractal discrimination model (mdm) of high-frequency pupil diameter measurements for human-computer interaction," in *Quantitative Medical Data Analysis Using Mathematical Tools and Statistical Techniques* (Hong, D. and Shyr, Y., eds.), ch. 16, pp. 333–350, World Scientific Publications, 2007.

[71] Vehel, J. and Riedi, R., "Fractional Brownian motion and data traffic modeling: The other end of the spectrum," *Fractals in Engineering*, pp. 185–202, 1997.

[72] Veitch, D. and Abry, P., "A statistical test for the time constancy of scaling exponents," *IEEE Transactions on Signal Processing*, vol. 49, no. 10, pp. 2325–2334, 2001.

[73] Veitch, D., Hohn, N., and Abry, P., "Multifractality in TCP/IP traffic: the case against," *Computer Networks*, vol. 48, no. 3, pp. 293–313, 2005.

[74] Vidakovic, B. and Ruggeri, F., "BAMS method: Theory and simulations," *Sankhyā: The Indian Journal of Statistics, Series B*, pp. 234–249, 2001.

[75] Vidakovic, B., *Statistical Modeling By Wavelets*. Wiley, 1999.

[76] Wendt, H., Roux, S., Jaffard, S., and Abry, P., "Wavelet leaders and bootstrap for multifractal analysis of images," *Signal Processing*, vol. 89, no. 6, pp. 1100–1114, 2009.

[77] Wood, A. and Chan, G., "Simulation of stationary Gaussian processes in [0, 1] d," *Journal of computational and graphical statistics*, vol. 3, no. 4, pp. 409–432, 1994.

[78] Zhu, X., "Semi-supervised learning literature survey," Tech. Rep. 1530, Computer Sciences, University of Wisconsin-Madison, 2005.

VITA

Kichun Lee is also called "Sky" or Kichun Sky Lee. It is because the name in Korean has the meaning of "foundation" for "Ki" and "sky" for "Chun". Now, he is a Ph.D. candidate in statistics in Industrial and Systems Engineering Department, Georgia Institute of Technology. He joined the Ph.D program in August 2006 after working for Tmax Soft and Samsung Networks as software researcher for 5 years in system software and RFID applications. He is interested in computational, statistical models in a variety of theoretical and practical problems with various data mining techniques.