Syracuse University

SURFACE

Dissertations - ALL SURFACE

12-2013

# Filter-based multiscale entropy analysis of complex physiological time series

Liang Zhao

Follow this and additional works at: https://surface.syr.edu/etd

Part of the Applied Mathematics Commons, and the Bioinformatics Commons

## Recommended Citation

Zhao, Liang, "Filter-based multiscale entropy analysis of complex physiological time series" (2013). *Dissertations - ALL*. 3.
https://surface.syr.edu/etd/3

This Dissertation is brought to you for free and open access by the SURFACE at SURFACE. It has been accepted for inclusion in Dissertations - ALL by an authorized administrator of SURFACE. For more information, please contact surface@syr.edu.

## ABSTRACT

The multiscale entropy (MSE) has been widely and successfully used in analyzing the complexity of physiologic time series. In this thesis, we re-interpret the averaging process in MSE as filtering a time series by a filter of a piecewise constant type. From this viewpoint, we introduce the *filter-based multiscale entropy* (FME) which filters a time series by filters to generate its multiple frequency components and then compute the *blockwise* entropy of the resulting components. By choosing filters adapted to the feature of a given time series, FME is able to better capture its multiscale information and to provide more flexibility for studying its complexity. Motivated by the heart rate turbulence theory which suggests that the human heartbeat interval time series (HHITS) can be described in piecewise linear patterns, we propose the piecewise linear filter multiscale entropy (PLFME) for the complexity analysis of the time series. Numerical results from PLFME are more robust to data of various lengths than those from MSE. We then propose wavelet packet transform entropy (WPTE) analysis. We apply WPTE analysis to HHITS using lower and higher piecewise linear filters. Numerical results show that WPTE using piecewise linear filters gives us the highest classification rates discriminating different cardiac systems among other multiscale entropy analysis. At the end, we discuss the application of FME on discrete time series. We introduce an 'eliminating' algorithm to examine and compare the complexity of coding and noncoding DNA sequences.

# Filter-Based Multiscale Entropy Analysis of Complex Physiologic Time Series

by

Liang Zhao

B.S., Jilin University - China, 2007

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Mathematics

in the Graduate School of Syracuse University

Syracuse University

December 2013

©2013, Liang Zhao

All Rights Reserved

# Acknowledgments

I would like to express my special gratitude for advice from Professor Yuesheng Xu who gave direction to this dissertation. I would like to thank Professor Lixin Shen and Uday Banerjee for their comments and suggestions. I am also indebted to my many student colleagues for providing help through years of graduate study. I am especially grateful to Feishe Chen, Jeremy Entner and Qian Huang. Last but not least, I am sincerely thankful for the help and support from my parents, my wife Jing, and my daughter Zoe who brought so much joy while I was writing this dissertation. All remaining errors are my own.

# Contents

**Acknowledgments** **iv**

**1 Introduction** **1**

**2 Filter-Based Multiscale Entropy** **8**

2.1 Filter-Based Multiscale Entropy . . . . . 10

2.2 FME for Gaussian and $1/f$ Noise . . . . . 15

2.3 Application to Human Heartbeat Interval Time Series . . . . . 40

2.4 Adaptive Filters . . . . . 50

**3 WPT Entropy Analysis** **55**

3.1 Wavelet Packet Transform . . . . . 57

3.2 Piecewise Polynomial Wavelets Filters on Invariant Sets . . . . . 60

3.3 Wavelet Packet Transform Entropy Analysis . . . . . 67

3.4 Application to Human Heartbeat Interval Time Series . . . . . 69

3.5 Multi-category classification SVM . . . . . 75

3.6 Classification for Heartbeat Interval Time Series . . . . . 81

**4 Application of FME on DNA** **85**

4.1 Sample Entropy on Discrete Time Series . . . . . 87

4.2 MSE on DNA . . . . . 92

4.3 'Eliminating' Algorithm for DNA sequences . . . . . 94

# Chapter 1

# Introduction

Heart disease is one of the top threats to people's health. According to Center for Disease Control and Prevention, about 600,000 people die of heart disease in the United States every year, that is 1 in every 4 deaths. Worldwide, an estimated 17.3 million people died from heart disease in 2008 according to World Health Organization. There are several tests to diagnose heart disease including electrocardiogram, stress testing, chest X Ray, etc. However, no single test can diagnose heart disease completely. People have been making great efforts to diagnose heart disease more accurately and effectively.

Over the past years, interest has risen in applying methods and concepts from nonlinear dynamics to problems in physiology. This is evidenced by several focus issues on cardiology and nonlinear dynamics [20, 30]. For example, it has been proposed that the normal heartbeat is associated with complex nonlinear dynamics and

chaos [62, 70]. This has opened new avenues to use methods from nonlinear dynamics as diagnostic tools for the analysis of physiological data for certain heart trouble [23, 31].

There have been several approaches coming from nonlinear dynamics to identify signatures in heart data [68]. Among these approaches, quantifying the "complexity" of physiologic signals has drawn considerable attention [19, 20, 22, 32, 56, 57]. In general, there is no precise mathematical definition for the "physiologic complexity" and no single statistical measure can be used to assess the complexity of physiologic systems [32]. Intuitively, complexity is referred to "meaningful structural richness". Several entropy metrics such as approximation entropy ( [55]) and sample entropy ( [58]) were proposed to measure the regularity of a system by quantifying the degree of predictability of a series of data points generated from the system. However, irregularity is not the same as complexity. For example, the entropy measures mentioned above assign the highest values to uncorrelated random signals (Gaussian white noise), which are highly unpredictable but not structurally "complex". Moreover, when they are applied to human heart interbeat time series (HHITS), certain pathologies including cardiac arrhythmias like atrial fibrillation are assigned higher entropy value than healthy dynamics which represent more physiologically complex, adaptive states. The reason for which these entropy matrices do not work on physiological systems is that they only measure the complexity at the single scale. While biological systems operate across multiple spatial and temporal scales, their complexity should also be measured

multiscaled. Therefore, these entropy matrices are not direct indices of physiologic complexity. In this dissertation as in [20, 22] we take the point of view that for a physiologic system, the complexity should be measured across multiple scales using entropy matrices and the higher the multiscale entropy value is, the more complex the system is.

The remarkable MSE [20, 22] takes it into account that the biological systems operate across multiple spatial and temporal scales during measuring the complexity of the physiologic time series and examines the physiologic dynamics over multiple scales. MSE gives more precise descriptions of the complexity of signals. For example, it is able to show that correlated random signals ($1/f$ noise) are more complex than uncorrelated random signals (Gaussian white noise). When applied to HHITS, MSE not only provides a meaningful measure for the complexity of the physiological time series but also shows good results in distinguishing different patterns from subjects with different aging and heart diseases. The advantages of MSE have drawn great attention since it was proposed in 2002. This work is highlighted in Nature News and Views [16], the American Institute of Physics News Update (Aug. 1, 2002), the Harvard Focus (Mar. 8, 2002). Numerical results from MSE show that cardiac system from healthy young subjects is more complex than that from healthy old subjects and subjects with heart diseases. They also suggest that cardiac system may lose more complexity from certain heart disease (congestive heart failure) than from aging. However, when applied to HHITS of various lengths, numerical results

from MSE have serval inconsistent observations, which is discussed in the second chapter of this dissertation. We will interpret MSE from the viewpoint of filters and propose filter-based multiscale entropy analysis, which gives us more robust results to data of various lengths.

A crucial step in MSE is the coarse graining procedure which assesses the entropy rate. It is achieved by an averaging process, extracting low-frequency components of the time series, at different scales. This procedure can be reinterpreted from a filter viewpoint as applying a piecewise constant low-pass filter which has a matrix representation to the time series. We shall take this point of view in studying filter-based multiscale entropy analysis.

The main purpose of this dissertation is to introduce filter-based multiscale entropy analysis. Specifically, with FME the time series is passed through desired fine-to-coarse filter matrices at different scales and a blockwise sample entropy value is calculated at each scale. On one hand, this general setting will give us insightful understanding of MSE and on the other hand, it will allow us to choose a filter that better fits the given data when certain prior information of the data is available to improve the entropy result. When prior information of the time series is not available we can develop adaptive filters which extract the main feature of the time series.

Entropy values at different scales can also be used as extracted features to classify different cardiac systems. Fish discriminant analysis and support vector machine methods were applied to the classification of heartbeat interval time series using en-

tropy values from MSE as features in [18] and [46] respectively. In terms of classification, we want to have as many distinct features as we can on different systems. While both MSE and PLFME only focus on the lower frequency components of the time series at different scales, the higher frequency components are overlooked. The entropy of the higher frequency components of a time series may also provide useful information in addition to that encoded in the lower frequency components. One example of such a system was presented in [38]. The hierarchical entropy (HE) analysis introduced in [38] defined two operators on a time series to extract both lower and higher frequency components of the time series. Wavelet packet transform [17] provides us with a systematic way to decompose the original time series into lower and higher frequency components. Wavelet packet transform has been widely used in texture classification [41], gearbox fault detection [24], embedded image coding [61], sparse approximation [64] and among others. We will introduce wavelet packet transform entropy (WPTE) analysis, as a further development of FME. In WPTE analysis we decompose the original time series using wavelet packet transform at different scales. For all of the decomposed time series obtained from wavelet packet transform, which present a hierarchical structure as in the HE analysis, we compute the (blockwise) sample entropy.

Using entropy values as extracted features and following from multi-category classification support vector machine (SVM), we develop several classifiers for human heartbeat interval time series. Classification rates will be given and compared when

we use entropy values from different multiscale entropy methods. Among these classifiers, we will see that if we use entropy values from WPTE using piecewise linear filters, we have the highest classification rates.

While most of the application of entropy methods focuses on continuous time series in current literature, such as HHITS, the application of entropy methods to discrete time series, such as DNA sequences, is overlooked. In [22], MSE is applied to DNA sequences, but there is an oscillation artifact. This is because every component of the discrete time series can only take values from a finite set, but new values will be created in the MSE analysis. FME provides us a more flexible choice of filters and it enables us design special filters for discrete time series. We will study the unique properties of sample entropy on discrete time series. Based on these properties, a multiscale 'eliminating' process is introduced via a special filter designed especially for the discrete time series. We will verify a conjecture in the medical field that noncoding DNA sequences are more complex than coding DNA sequences, which shows the applicability of FME to discrete time series.

We organize the remaining of this dissertation into three chapters. In the second chapter, we shall provide a general framework of FME and theoretical results of FME for Guassian white noise as well as $1/f$ noise. The application of FME to HHITS will also be thoroughly studied. In the third chapter, we shall introduce the WPTE analysis. Theoretical results of WPTE for Gaussian white noise will be presented. We will also apply WPTE to HHITS. Moreover, several classifiers for HHITS will be

proposed and the classification results will be compared. In the forth chapter, we study the application of FME to discrete time series. Special properties of sample entropy on discrete time series will be examined and an "eliminating" algorithm will be proposed and applied to DNA sequences.

# Chapter 2

# Filter-Based Multiscale Entropy

The complex fluctuations exhibited by a signal generated from a physiologic system contain information of underlying interacting mechanisms which regulate the system. Quantifying the "complexity" of physiologic signals has drawn considerable attention. The remarkable MSE [20, 22] takes it into account that the biological systems operate across multiple spatial and temporal scales during measuring the complexity of the physiologic time series and examines the physiologic dynamics over multiple scales.

We reinterpreted a crucial step in MSE, the coarse graining procedure, from a filter viewpoint as applying a piecewise constant low-pass filter which has a matrix representation to the time series. We shall take this point of view and propose filter-based multiscale entropy analysis.

We consider in this chapter HHITS as our main study case. Heart rate turbulence (HRT), the technique of acceleration-deceleration oscillation analysis proposed in [59],

suggests that HHITS can be described in the piecewise linear pattern. The time series generated by different heart conditions show distinguished differences in this pattern. Therefore, using piecewise linear filters for capturing this pattern is highly desirable. We apply the piecewise linear filter to HHITS before measuring their complexity and find that aging may reduce the complexity of the cardiac system more than congestive heart failure. Numerical results from PLFME are more robust to data of various lengths than those from MSE. We furthermore design an adaptive filter for HHITS (without prior information of HHITS) and use it in developing adaptive piecewise constant filter-based multiscale entropy (APCFME) analysis. In the study of HHITS, numerical performance of APCFME is comparable to that of PLFME.

We organize this chapter in seven sections. In section 1.1, we describe the coarse graining processing using filters and a blockwise sample entropy for computing the resulting filter-based multiscale entropy. We then study in section 1.2 FME for the Gaussian noise and the $1/f$ noise. We provide theoretical results and numerical results. Section 1.3 is devoted to application of PLFME to HHITS. In section 1.4, we design an adaptive piecewise constant filter and use it in developing APCFME. Numerical results of APCFME applied to the Gaussian noise, $1/f$ noise and HHITS are also presented in this section.

## 2.1 Filter-Based Multiscale Entropy

We motivate FME from a filter viewpoint which re-interprets the averaging process in MSE as filtering a time series through a lower pass filter of a *piecewise constant* type in generating its multiple frequency components. Piecewise constant filters may be suitable for signals which can be described in piecewise constant patterns but may not for others. To make MSE more robust to signals with different nature, we introduce FME which considers the meaningful structural complexity of a physiologic system over multiple spatial and temporal scales resulted from filters appropriate for the specific physiologic system. Specifically, at each scale, from finer to coarser, the time series is passed through a desired filter to capture its characteristic pattern. For example, a piecewise polynomial filter of order $k$ can be used to approximate a time series which can be intrinsically represented by such a function. When prior information of the signal is available, one can use it in the filter design and when it is not available, one may construct filters adaptively from the given signal.

A filter may be described in terms of a matrix. For example, the Haar filter is the $1 \times 2$ matrix $[\frac{1}{2}, \frac{1}{2}]$. The piecewise polynomial filter may be derived from the wavelets on invariant sets [49] and a general construction of filters of this type was discussed in [50]. From [43], the piecewise linear filter is the $2 \times 4$ matrix given by

$$A := \frac{1}{2} \begin{bmatrix} 1 & 0 & 1 & 0 \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} & \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix} \tag{2.1}$$

and the piecewise quadratic filter is the $3 \times 6$ matrix given by

$$B := \frac{1}{2}\begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} & 0 & \frac{\sqrt{3}}{2} & \frac{1}{2} & 0 \\ 0 & -\frac{\sqrt{15}}{4} & \frac{1}{4} & 0 & \frac{\sqrt{15}}{4} & \frac{1}{4} \end{bmatrix}. \tag{2.2}$$

A coarse graining process of a time series $\mathbf{x} := [x_0, \ldots, x_{N-1}]$ of real numbers can be viewed as a matrix multiplication of the vector $\mathbf{x}$ (we use the same notation $\mathbf{x}$ for the time series and the vector). Specifically, at each scale $\tau = 2, 3, \ldots$, a matrix $A^{(\tau)} \in \mathbb{R}^{p_\tau \times q_\tau}$ is chosen as a filter for $\mathbf{x}$. For matrices $P := [p_{jk}]$ and $Q$ we define the Kronecker product $P \otimes Q := [p_{jk}Q]$. By $\lfloor \cdot \rfloor$ we denote the floor function. At scale $\tau$, the coarse-grained time series is constructed by $A^{(\tau)}$ as $\mathbf{y}^\tau := (I_n \otimes A^{(\tau)})\mathbf{x}$, where $n := \lfloor \frac{N}{q_\tau} \rfloor$ and $I_n$ is the $n \times n$ identity matrix. If $N = nq_\tau + k$, for some integers $n$ and $k$ with $1 \leq k < q_\tau$, we shall drop the last $k$ components of $\mathbf{x}$ when construct the coarse-grained time series since such insignificant loss of a few components will barely affect the complexity of the whole system. Thus, in each coarse-grained procedure, the time series is partitioned into $n$ blocks with each having $q_\tau$ components and being transformed by $A^{(\tau)}$ to another block of $p_\tau$ components.

The coarse graining process can also be viewed as the application of the same filter matrix recursively to $\mathbf{x}$. For a filter matrix $A \in \mathbb{R}^{p \times q}$ and $\tau = 2, 3, \ldots$, the coarse-grained time series at the scale $\tau$ is obtained recursively by $\mathbf{y}^\tau = (I_{n_{\tau-1}} \otimes A)\mathbf{y}^{\tau-1}$ with $\mathbf{y}^1 := \mathbf{x}$ and $n_{\tau-1} := \lfloor \frac{N_{\tau-1}}{q} \rfloor$, where $N_{\tau-1}$ is the length of the time series $\mathbf{y}^{\tau-1}$. For example, for PLFME, in the above formula $A$ is the piecewise linear filter defined in

(2.1) and $q = 4$.

We recall the notion of sample entropy [58]. Let $\mathbb{Z}_m := \{0, 1, \ldots, m-1\}$ for a positive integer $m$. We denote by $\mathbf{x}(i)$ the ith component of a time series $\mathbf{x}$. For given $\mathbf{x}$, we construct a sequence $\mathbf{u}_m := \{u_m(j) : j \in \mathbb{Z}_{N-m}\}$, where $u_m(j) := [x(j+k) : k \in \mathbb{Z}_m]$ are vectors of $m$ data points, with $m$ being the length of pattern templates. The distance between $u_m(\ell)$ and $u_m(j)$ is defined as $d[u_m(\ell), u_m(j)] := \max\{|x(\ell+k) - x(j+k)| : k \in \mathbb{Z}_m\}$. For a given tolerance $r > 0$ and a fixed integer $\ell \in \mathbb{Z}_{N-m}$, we let $B_\ell^m$ denote the number of vectors $u_m(j)$ with $j > \ell$ which satisfy $d[u_m(\ell), u_m(j)] \leq r$. The number $r$ serves as the tolerance for accepting matches and $u_m(\ell)$ is called the template. Then the probability of vectors $u_m(j) \in \mathbf{u}_m$ that are near the template $u_m(\ell)$ within tolerance $r$ is given by

$$C_\ell^m(\mathbf{x}, r) := B_\ell^m/(N - m + 1).$$

Let

$$C^m(\mathbf{x}, r) := \sum_{\ell=0}^{N-m} C_\ell^m(\mathbf{x}, r). \tag{2.3}$$

Sample entropy of $\mathbf{x}$ is defined by

$$S_m(\mathbf{x}, r) := -\ln\left[\frac{C^{m+1}(\mathbf{x}, r)}{C^m(\mathbf{x}, r)}\right].$$

We propose a *blockwise sample entropy* (BSE) for the filtered time series $\mathbf{y}^\tau$ at each scale. The use of BSE (instead of the standard sample entropy) is to adjust to the size of the filters used in FME which result in output signals having blocks consisting of more than one components. We now introduce BSE for the coarse-

grained time series $\mathbf{y}^\tau$ for $\tau \geq 2$. It follows the same idea as sample entropy and is designed to suit the structure of $\mathbf{y}^\tau$. Note that $\mathbf{y}^\tau$ consists of $n$ blocks, each of which has $p_\tau$ components and is obtained from a block of $\mathbf{x}$ transformed by the same matrix $A^{(\tau)}$. Components in each block represent different information of $\mathbf{x}$, captured by different rows of the filter $A^{(\tau)}$. For example, different rows of the linear filter transform $\mathbf{x}$ to different components in the piecewise linear function basis. For this reason, in calculating BSE, we consider each block of $\mathbf{y}^\tau$ as a single unit and apply the sample entropy calculation process to $\mathbf{y}^\tau$. To this end, we write $\mathbf{y}^\tau = [\mathbf{y}_0^\tau, \ldots, \mathbf{y}_{n-1}^\tau]$, where each $\mathbf{y}_j^\tau$ is a vector with $p_\tau$ components and $[\mathbf{y}_j^\tau]_s := \mathbf{y}^\tau(jp_\tau + s)$ for each $s \in \mathbb{Z}_{p_\tau}$ with $[\mathbf{d}]_s$ denoting the sth component of a vector $\mathbf{d}$. Then the sequence corresponding to $\mathbf{u}_m$ in sample entropy is constructed as $\mathbf{u}_m^\tau := \{u_m^\tau(j) : j \in \mathbb{Z}_{n-m}\}$, where $u_m^\tau(j) := [\mathbf{y}_{j+k}^\tau : k \in \mathbb{Z}_m]$ consists of $m$ vectors. Let

$$\sigma_s(u_m^\tau(\ell), u_m^\tau(j)) := \max\{|[\mathbf{y}_{\ell+k}^\tau]_s - [\mathbf{y}_{j+k}^\tau]_s| : s \in \mathbb{Z}_m\}.$$

The distance between $u_m^\tau(\ell)$ and $u_m^\tau(j)$ is then defined as a vector of $p_\tau$ components by

$$\mathbf{d}[u_m^\tau(\ell), u_m^\tau(j)] := [\sigma_s(u_m^\tau(\ell), u_m^\tau(j)) : s \in \mathbb{Z}_{p_\tau}].$$

For a given $u_m^\tau(\ell)$, we denote by $B_\ell^{m,\tau}$ the number of vectors $u_m^\tau(j)$ with $j > \ell$ which satisfy $\sigma_s(u_m^\tau(\ell), u_m^\tau(j)) \leq r_s^\tau$ for each $s \in \mathbb{Z}_{p_\tau}$, where $r_s^\tau := \sum_{t=1}^q |A_{st}^{(\tau)}| r$. Here $r$ is the tolerance used in calculating sample entropy of $\mathbf{x}$ and $A_{st}^{(\tau)}$ is the $(s,t)$-entry of $A^{(\tau)}$. Let $r^\tau := [r_0^\tau, r_1^\tau, \ldots, r_{p^\tau-1}^\tau]$ and

$$\tilde{C}_\ell^m(\mathbf{y}^\tau, r^\tau) := \frac{B_\ell^{m,\tau}}{n-m+1}, \ell \in \mathbb{Z}_{n-m+1}.$$

BSE $\tilde{S}_m(\mathbf{y}^\tau, r^\tau)$ of $\mathbf{y}^\tau$ is defined in the same way as sample entropy with $C_\ell^m(\mathbf{x}, r)$ replaced by $\tilde{C}_\ell^m(\mathbf{y}^\tau, r)$. That is,

$$\tilde{S}_m(\mathbf{y}^\tau, r^\tau) := -\ln\left[\frac{\tilde{C}_\ell^{m+1}(\mathbf{y}^\tau, r)}{\tilde{C}_\ell^m(\mathbf{y}^\tau, r)}\right]. \tag{2.4}$$

The choice of parameter $r^\tau$ in BSE is crucial. In MSE, the same value $r$ was used for different scales. Adjusted to the decreasing variance of the filtered time series in scales, the parameter $r$ in MSE was adjusted in [51] as a certain percentage of the standard deviation of the filtered time series at each scale. In BSE, the parameter $r^\tau$ is calculated from a different viewpoint. Since the filtering process in FME is a transformation of the time series by a matrix, measuring the similarity of the components in the filtered time series should be related to the filter matrix. Thus the parameter $r^\tau$ is a vector whose components are transformed by the corresponding rows of the filter matrix. Each component $r_s^\tau$, $s \in \mathbb{Z}_{p_\tau}$, is to measure the similarity of the corresponding components among different blocks in the filtered time series.

We now elaborate the relation of MSE and FME. In MSE, the consecutive coarse-grained time series $\{\mathbf{y}^\tau\}$ is constructed according to the equation:

$$\mathbf{y}^\tau(j) = 1/\tau \sum_{k=j\tau}^{(j+1)\tau-1} \mathbf{x}(k), j \in \mathbb{Z}_{\lfloor N/\tau \rfloor}.$$

The time series is actually filtered by the $1 \times \tau$ matrix

$$C^{(\tau)} := \left[\frac{1}{\tau}, \frac{1}{\tau}, \ldots, \frac{1}{\tau}\right]$$

at scale $\tau$. That is, $\mathbf{y}^\tau = (I_{\lfloor \frac{N}{\tau} \rfloor} \otimes C^{(\tau)})\mathbf{x}$. Thus $\mathbf{y}^\tau$ has $\lfloor N/\tau \rfloor$ blocks with each having one component. In this case, BSE degenerates to the standard sample entropy.

Therefore, MSE is actually a special case of FME with the piecewise constant filter $C^{(\tau)}$ at scale $\tau$. We also remark that when $\tau = 2^k$, for $k = 1, 2, \ldots$, filtering a time series by $C^{(\tau)}$ is equivalent to filtering it by the Haar filter $k$ times recursively.

## 2.2 FME for Gaussian and $1/f$ Noise

In this section, we discuss the behavior of FME in simulating white noise, a completely irregular signal and in $1/f$ noise, a correlated signal.

**FME for Gaussian Noise.** We first apply FME to Gaussian white noise. Intuitively, complexity is associated with "meaningful structural richness" [4]. There is no straightforward correspondence between regularity and complexity. For example, uncorrelated random signals, such as Gaussian white noise, are highly unpredictable but not structurally complex. We present both theoretical and numerical results of Gaussian white noise from FME. We shall see that the entropy measure of the Gaussian white noise decreases as the scale increases with proper choices of the filter. This indicates the lack of the complexity of the Gaussian white noise.

For a positive integer $N$, let $\mathbf{x} := [x_j : j \in \mathbb{Z}_N]$ denote a random vector taking values in $\mathbb{R}^N$. When the components $x_j, j \in \mathbb{Z}_N$, are independent and have the same Gaussian distribution in the sense that they have the same mean and standard deviation, we call $\mathbf{x}$ a real Gaussian random vector, call a component of $\mathbf{x}$ a Gaussian random variable and call an instance of $\mathbf{x}$ the Gaussian white noise. In this subsection, we use $\mathbf{g} := [g_j : j \in \mathbb{Z}_N]$ for the real Gaussian random vector, with the mean of $g_j$

being 0 and the standard deviation of $g_j$ being $\delta$.

The analysis of Gaussian noise mainly relies on the statistical properties of linear combination of Gaussian random variables. Theorem 4.2.14 in [15] states that a linear combination of two independent real Gaussian random variables is also a real Gaussian random variable. This result can be easily generalized to the following lemma. The proof of the lemma is straightforward and will be omitted.

**Lemma 2.1.** *If $X_j$ are $n$ independent real Gaussian random variables with mean 0 and standard deviation $\delta_j$, then for $\alpha_j \in \mathbb{R}, j \in \mathbb{Z}_n$, $\sum_{j=0}^{n-1} \alpha_j X_j$ is a real Gaussian random variable with mean 0 and standard deviation $\tilde{\delta} := (\sum_{j=0}^{n-1} \alpha_j^2 \delta_j^2)^{1/2}$.*

As usual, the expectation of a random variable $X$ taking values in $\mathbb{R}$ is defined by

$$E(X) := \int_{\mathbb{R}} tp(t)dt,$$

where $p$ is the probability density function for the random variable $X$. Two random variables $X$ and $Y$ are independent if and only if $E(XY) = 0$. We let $\mathrm{COV}(X, Y)$ denote the covariance of $X$ and $Y$. It is known from Theorem 4.5.3 in [15] that

$$\mathrm{COV}(X, Y) = E(XY) - E(X)E(Y).$$

Given a filter $A^{(\tau)} \in \mathbb{R}^{p_\tau \times q_\tau}$, we consider BSE of the filtered Gaussian random vector,

$$\mathbf{g}^\tau := (I_n \otimes A^{(\tau)})\mathbf{g}, \tag{2.5}$$

where $n = \lfloor \frac{N}{q_\tau} \rfloor$. We write $\mathbf{g}^\tau$ as defined in (2.5) in a blockwise form

$$\mathbf{g}^\tau := [\mathbf{g}_0^\tau, \ldots, \mathbf{g}_{n-1}^\tau], \tag{2.6}$$

where $\mathbf{g}_j^\tau, j \in \mathbb{Z}_n$, is a vector with $p_\tau$ components, and for each $k \in \mathbb{Z}_{p_\tau}$. let $[\mathbf{g}_j^\tau]_k := \mathbf{g}^\tau(jp_\tau + k)$.

Recall that BSE estimates the negative natural logarithm of the conditional probability that the distance between two blocks in $\mathbf{g}^\tau$ is small (measured by $r_s$ in the description of BSE) given that the distance between the two preceding blocks is also small. This conditional probability can be analytically expressed by the probability density function of all of the blocks in $\mathbf{g}^\tau$ (Lemma 2.2) and the independence of blocks in $\mathbf{g}^\tau$ (Lemma 2.3).

**Lemma 2.2.** *If $\mathbf{g}_j^\tau$ is defined in (2.6), $j \in \mathbb{Z}_n$, then $\mathbf{g}_j^\tau$ is a $p_\tau$-variate normally distributed random vector with its mean being the zero vector and and its covariance matrix being $\Sigma$ with*

$$\Sigma_{s,t} := \sum_{k=1}^{p_\tau} A_{sk}^{(\tau)} A_{tk}^{(\tau)} \delta^2, 1 \leq s, t \leq p_\tau, \tag{2.7}$$

*where $\delta$ is the standard deviation of $g_j$.*

*Proof.* From the construction of $\mathbf{g}_j^\tau$, for each $j \in \mathbb{Z}_n$, we have that $\mathbf{g}_j^\tau = A^{(\tau)}\mathbf{g}_j$, where $\mathbf{g}_j \in \mathbb{R}^{q_\tau}$ with $[\mathbf{g}_j]_k = \mathbf{g}(q_\tau + k)$ for each $k \in \mathbb{Z}_{q_\tau}$. For each $j \in \mathbb{Z}_n$, $\mathbf{g}_j$ consists of $q_\tau$ independent real Gaussian random variables with mean 0 and standard deviation $\delta$ and $[\mathbf{g}_j^\tau]_k, k \in \mathbb{Z}_{p_\tau}$, is a linear combination of these random variables. Hence, it follows from Lemma 2.1 that for each $k \in \mathbb{Z}_{p_\tau}$, $[\mathbf{g}_j^\tau]_k$ is a Gaussian random variable

with mean 0 and standard deviation $\sqrt{\sum_{\ell=1}^{p_\tau}(A_{k\ell}^{(\tau)})^2}\delta$. Hence, for any $1 \le s, t \le p_\tau$,

$$\begin{aligned}\mathrm{COV}([\mathbf{g}_j^\tau]_s, [\mathbf{g}_j^\tau]_t) &= E([\mathbf{g}_j^\tau]_s[\mathbf{g}_j^\tau]_t) \\ &= E(A_s^{(\tau)}\mathbf{g}_j \cdot A_t^{(\tau)}\mathbf{g}_j) \\ &= \sum_{k=1}^{p_\tau} A_{sk}^{(\tau)} A_{tk}^{(\tau)} \delta^2,\end{aligned}$$

where $A_k^{(\tau)}$ is the $k$th row of $A^{(\tau)}$ for $1 \le k \le p_\tau$. The desired result follows directly from the definition of the multivariate normal distribution. □

Noting that formula (2.7) is independent of $j$, all blocks in $\mathbf{g}^\tau$ have the same probability density function. In the next lemma, we show the independence of blocks in $\mathbf{g}^\tau$. We say that two random variable vectors $\mathbf{X}$ and $\mathbf{Y}$ are independent if the elements of $\mathbf{X}$ (as a collection of random variables) are independent of the elements of $\mathbf{Y}$. Elements within $\mathbf{X}$ or $\mathbf{Y}$ need not be independent when $\mathbf{X}$ and $\mathbf{Y}$ are independent.

**Lemma 2.3.** *If $\mathbf{g}_j^\tau$ is defined as in (2.6), $j \in \mathbb{Z}_n$, then $\mathbf{g}_j^\tau$ and $\mathbf{g}_k^\tau$, $j, k \in \mathbb{Z}_n$, are independent when $j \neq k$.*

*Proof.* It suffices to prove that $[\mathbf{g}_j^\tau]_s$ and $[\mathbf{g}_k^\tau]_t$ are independent for any $s, t \in \mathbb{Z}_{p_\tau}$ if $j \neq k, j, k \in \mathbb{Z}_n$. Since $\mathbf{g}(jp_\tau + s)$ and $\mathbf{g}(kp_\tau + t)$ are independent, we have that $E[\mathbf{g}(jp_\tau + s)\mathbf{g}(kp_\tau + t)] = 0$. It follows that

$$\begin{aligned}&E([\mathbf{g}_j^\tau]_s][\mathbf{g}_k^\tau]_t) \\ =&E\left[\sum_{\ell=1}^{p_\tau} A_{s\ell}^{(\tau)}\mathbf{g}(jq_\tau + \ell - 1)\sum_{\ell=1}^{p_\tau} A_{t\ell}^{(\tau)}\mathbf{g}(kq_\tau + \ell - 1)\right] \\ =&\sum_{\ell=1}^{p_\tau}\sum_{u=1}^{p_\tau} A_{s\ell}^{(\tau)}A_{tu}^{(\tau)}E[\mathbf{g}(jq_\tau + \ell - 1)\mathbf{g}(kq_\tau + u - 1)] = 0, \qquad (2.8)\end{aligned}$$

if $j \neq k$. Thus, $[\mathbf{g}_j^\tau]_s$ and $[\mathbf{g}_k^\tau]_t$, $s, t \in \mathbb{Z}_{p_\tau}$, are independent. We then conclude that $\mathbf{g}_j^\tau$ and $\mathbf{g}_k^\tau$ are independent for any $j, k \in \mathbb{Z}_n$ with $j \neq k$. □

For notational convenience, in the remaining part of this section we use the same notation $\tilde{S}_m(\mathbf{g}^\tau, r)$ for the theoretical value of BSE of $\mathbf{g}^\tau$. For a positive integer N, let $\mathbb{Z}_N^+ := \{1, 2, \ldots, N\}$. For a given $r > 0$, we let

$$r_s^\tau := \sum_{t=1}^{q_\tau} |A_{st}^{(\tau)}| r, \quad s \in \mathbb{Z}_{p_\tau}^+.$$

For a given vector $\mathbf{y} \in \mathbb{R}^{p_\tau}$, we define

$$\Omega_{\mathbf{y}} := \big([\mathbf{y}]_1 - r_1^\tau, [\mathbf{y}]_1 + r_1^\tau\big) \times \cdots \times \big([\mathbf{y}]_{p_\tau} - r_{p_\tau}^\tau, [\mathbf{y}]_{p_\tau} + r_{p_\tau}^\tau\big).$$

Clearly, $\Omega_{\mathbf{y}} \subset \mathbb{R}^{p_\tau}$. For the standard deviation $\delta$ of $g_j$, we define the matrix $\Sigma \in \mathbb{R}^{p_\tau \times p_\tau}$ by

$$\Sigma_{s,t} := \sum_{j=1}^{p_\tau} A_{sj}^{(\tau)} A_{tj}^{(\tau)} \delta^2, 1 \leq s, t \leq p_\tau.$$

In this chapter we assume that the matrix $\Sigma$ is invertible. We showed in Lemma 2.2 that all of the blocks in $\mathbf{g}^\tau$ have the same probability density function with the covariance matrix $\Sigma$. We next present $\tilde{S}_m(\mathbf{g}^\tau, r)$ in terms of the matrix $\Sigma$. To this end, we let

$$\mathbf{I}(\Omega_{\mathbf{y}}) := \int_{\Omega_{\mathbf{y}}} \lambda_\Sigma \exp(-\frac{1}{2}\mathbf{x}^T \Sigma^{-1} \mathbf{x}) d\mathbf{x}, \tag{2.9}$$

where $\lambda_\Sigma := \frac{1}{(2\pi)^{p_\tau/2}|\Sigma|^{1/2}}$. We say that the distance between two blocks $\mathbf{g}_j^\tau$ and $\mathbf{g}_k^\tau$, $j, k \in \mathbb{Z}_n$, is less than $r^\tau$ if $[|\mathbf{g}_j^\tau - \mathbf{g}_k^\tau|]_s < r_s^\tau$ for all $s \in \mathbb{Z}_{p_\tau}$ and we write it as $|\mathbf{g}_j^\tau - \mathbf{g}_k^\tau| < r^\tau$.

**Proposition 2.4.** *If $A^{(\tau)} \in \mathbb{R}^{p_\tau \times q_\tau}$ and $\mathbf{g}^\tau$ is defined as in (2.5), then for any $r > 0$*

$$\tilde{S}_m(\mathbf{g}^\tau, r^\tau) = -\ln\left\{\int_{\mathbb{R}^{p_\tau}} \mathbf{I}(\Omega_\mathbf{y})\lambda_\Sigma \exp(-\frac{1}{2}\mathbf{y}^T\Sigma^{-1}\mathbf{y})d\mathbf{y}\right\}. \tag{2.10}$$

*Proof.* BSE $\tilde{S}_m(\mathbf{g}^\tau, r^\tau)$ is the negative natural logarithm of the conditional probability that the distance between two blocks is less than $r$ provided that the distance between the two preceding blocks is also less than $r$. We write $\mathbf{g}^\tau$ in the block form as in (2.6) and from Lemma 2.3, we know that $\mathbf{g}_j^\tau$ and $\mathbf{g}_k^\tau$ are independent if $j \neq k$. Thus, when $m = 1$ the conditional probability is

$$\begin{aligned}
&P(|\mathbf{g}_j^\tau - \mathbf{g}_k^\tau| < r^\tau || \mathbf{g}_{j-1}^\tau - \mathbf{g}_{k-1}^\tau| < r^\tau) \\
=&\frac{P(|\mathbf{g}_j^\tau - \mathbf{g}_k^\tau| < r \wedge |\mathbf{g}_{j-1}^\tau - \mathbf{g}_{k-1}^\tau| < r^\tau)}{P(|\mathbf{g}_{j-1}^\tau - \mathbf{g}_{k-1}^\tau| < r^\tau)} \\
=&\frac{P(|\mathbf{g}_j^\tau - \mathbf{g}_k^\tau| < r^\tau) \times P(|\mathbf{g}_{j-1}^\tau - \mathbf{g}_{k-1}^\tau| < r^\tau)}{P(|\mathbf{g}_{j-1}^\tau - \mathbf{g}_{k-1}^\tau| < r^\tau)} \\
=&P(|\mathbf{g}_j^\tau - \mathbf{g}_k^\tau| < r^\tau).
\end{aligned} \tag{2.11}$$

Using this approach recursively, it can be proved that this result is valid for any value of $m$.

From Lemma 2.2 and the definition of multivariate normal distribution, we know that the probability density functions of $\mathbf{g}_j^\tau, j \in \mathbb{Z}_n$, are all equal to

$$f(\mathbf{x}) := \frac{1}{(2\pi)^{p_\tau/2}|\Sigma|^{1/2}} \exp\{-\frac{1}{2}\mathbf{x}^T\Sigma^{-1}\mathbf{x}\}, \quad \mathbf{x} \in \mathbb{R}^{p_\tau}, \tag{2.12}$$

where $\Sigma$ is defined as in (2.7). Then the desired formula follows from the definition of BSE, (2.11) and (2.12). □

Formula (2.10) holds for a general filter matrix $A^{(\tau)}$. In particular, when we choose $A^{(\tau)} = C^{(\tau)}$, it recovers the theoretical results originally proved in [22] of MSE for Gaussian white noise.

We next present a special result when the rows of the the filter matrix $A$ are orthogonal. Examples of such filter matrices include the piecewise polynomial filter of order $k$. Let $A \in \mathbb{R}^{p\times q}$ be a matrix with orthogonal rows and

$$\tilde{\mathbf{g}} := (I_{\lfloor \frac{N}{q} \rfloor} \otimes A)\mathbf{g}.$$

We also write $\tilde{\mathbf{g}}$ in the blockwise form, that is,

$$\tilde{\mathbf{g}} := [\tilde{\mathbf{g}}_0, \ldots, \tilde{\mathbf{g}}_{n-1}]. \tag{2.13}$$

In this case, a set of independent variables are transformed by $A$ to another set of independent variables due to the orthogonality of the rows of $A$. Thus in addition to the independence of blocks of the filtered Gaussian white noise, elements within each block are also independent. In particular, given a real Gaussian random vector $\mathbf{g}$, we have the following result for each block $\tilde{\mathbf{g}}_j, j \in \mathbb{Z}_n$.

**Lemma 2.5.** *If $A \in \mathbb{R}^{p\times q}$ is a matrix with orthogonal rows and $\tilde{\mathbf{g}}_j$ is defined as in (2.13) via a real Gaussian random vector $\mathbf{g}$, then $\tilde{\mathbf{g}}_j$ consists of $p$ independent Gaussian random variables for each $j \in \mathbb{Z}_n$, with the mean and the standard deviation of $[\tilde{\mathbf{g}}_j]_k$ being $0$ and $\sqrt{\sum_{\ell=1}^{p}[A_{k\ell}]^2}\delta$, respectively, for $k \in \mathbb{Z}_p$, where $\delta$ is the standard deviation of $g_j$.*

*Proof.* For each $j \in \mathbb{Z}_n$ and $k \in \mathbb{Z}_p$, we know from Lemma 2.1 that $[\tilde{\mathbf{g}}_j]_k$ is a real Gaussian variable since it is a linear transformation of real Gaussian variables. The mean and standard deviation of $[\tilde{\mathbf{g}}_j]_k$ are easily obtained from Lemma 2.1.

It remains to prove that $[\tilde{\mathbf{g}}_j]_s$ and $[\tilde{\mathbf{g}}_j]_t$, $s, t \in \mathbb{Z}_p$, are independent when $s \neq t$ for each $j \in \mathbb{Z}_n$. This follows from the orthogonality of the rows of $A$. In fact,

$$\begin{aligned}
&E\left([\tilde{\mathbf{g}}_j]_s[\tilde{\mathbf{g}}_j]_t\right)\\
=&E\Big[\sum_{k=1}^{p} A_{sk}\mathbf{g}(jq+(k-1)) \cdot \sum_{k=1}^{p} A_{tk}\mathbf{g}(jq+(k-1))\Big]\\
=&\sum_{k=1}^{p} A_{sk}A_{tk}E\left[\mathbf{g}^2(jq+(k-1))\right]\\
=&\sum_{k=1}^{p} A_{sk}A_{tk}\delta = 0.
\end{aligned} \tag{2.14}$$

□

We use erf to denote the error function defined by

$$\operatorname{erf}(x) := \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dx.$$

For given $r > 0$ and matrix $A \in \mathbb{R}^{p\times q}$, we let $\delta(A, j) := \sqrt{\sum_{k=1}^{p} A_{jk}^2}\delta$ and

$$r_s := \sum_{t=1}^{q} |A_{st}| r, s \in \mathbb{Z}_p^+. \tag{2.15}$$

Then $\delta(A, j)$ is the standard deviation of the $j$th element in each block of the filtered Gaussian random vector (from Lemma 2.5). For real numbers $a$ and $b$, we define

$$E(a, b) := \operatorname{erf}\left(\frac{a+b}{\delta(A,j)}\right) - \operatorname{erf}\left(\frac{a-b}{\delta(A,j)}\right).$$

**Proposition 2.6.** *If matrix $A \in \mathbb{R}^{p\times q}$ has orthogonal rows, then for $\tilde{\mathbf{g}} := (I_{\lfloor \frac{N}{q} \rfloor} \otimes A)\mathbf{g}$ and for any $r > 0$,*

$$\tilde{S}_m(\tilde{\mathbf{g}}, r^\tau) = -\ln\Big(\prod_{j=1}^{p} \frac{1}{\sqrt{2\pi}\delta(A,j)} \int_{\mathbb{R}} E(x_j, r_j) \exp \frac{-x_j^2}{2(\delta(A,j))^2} dx_j\Big). \tag{2.16}$$

*Proof.* From Lemma 2.5, we know that the probability density function of $\mathbf{g}_j^\tau$, for each $j \in \mathbb{Z}_n$, is given by

$$f(\mathbf{x}) = \prod_{\ell=1}^{p_\tau} \frac{1}{\sqrt{2\pi \sum_{k=1}^{p} A_{\ell k}^2}\delta} \exp\big(\frac{-x_\ell^2}{2\sum_{k=1}^{p} A_{\ell k}^2 \delta^2}\big), \tag{2.17}$$

where $\mathbf{x}$ is a $p_\tau$ random vector. By applying (2.11) to (2.17) and the definition of BSE, we obtain the desired result. □

Since $\sqrt{\sum_{k=1}^{p} A_{ik}^2}\delta$ is actually the standard deviation of $[\tilde{\mathbf{g}}_j]_i$ for each block $\tilde{\mathbf{g}}_j$ in the filtered time series, BSE of $\tilde{\mathbf{g}}$ is determined by the standard deviation of $[\tilde{\mathbf{g}}_j]_i$ by Proposition 3.1. This leads us to investigate the standard deviation of the filtered time series at different scales if we apply the filter matrix $A$ recursively to $\mathbf{g}$.

Considering the Gaussian noise is a completely irregular signal, one expects that its complexity, measured by FME, decreases as the scale increases. This is the case when the filter matrix $A$ used in FME has orthogonal rows and satisfies the condition

$$A^T A = \rho^2 I_p, \tag{2.18}$$

for some constant $\rho \in (0, 1)$. This is stated in the next theorem. For this purpose, we let

$$\mathbf{g}^\tau := (I_{\lfloor \frac{N_{\tau-1}}{q} \rfloor} \otimes A)(I_{\lfloor \frac{N_{\tau-2}}{q} \rfloor} \otimes A) \cdots (I_{\lfloor \frac{N_1}{q} \rfloor} \otimes A)\mathbf{g}, \tag{2.19}$$

where $N_j$ is the length of the time series $\mathbf{g}^j, j = 2, 3, \ldots, \tau-1$ and $N_1$ is the length of $\mathbf{g}$. We also write $\mathbf{g}^\tau$ as defined in (2.19) in the blockwise form as in (2.6). By Proposition 2.4, at a fixed scale, all blocks have the same probability density function. Let $\delta_i^{(\tau)}$ denote the standard deviation of of the $i$th element in a block at scale $\tau$. If the matrix $A$ has orthogonal rows and satisfies condition (2.18), we have the following result for $\delta_i^{(\tau)}$.

**Lemma 2.7.** *If $A \in \mathbb{R}^{p\times q}$ has orthogonal rows and satisfies condition (2.18), then for $\tau = 2, 3, \ldots,$*

$$\delta_k^{(\tau)} = \rho^{\tau-1}\delta, \quad k \in \mathbb{Z}_p \tag{2.20}$$

*where $\rho$ is the constant that appears in (2.18) and $\delta$ is the standard deviation of $g_j$.*

*Proof.* We prove this result by induction on $\tau$. When $\tau = 2$, it follows from Lemma 2.5 and the fact that $A$ satisfies condition (2.18) that $\delta_k^{(2)} = \rho\delta$. By the induction hypothesis, we have that $\delta_k^{(\tau-1)} = \rho^{\tau-2}\delta$, for some $\tau \geq 3$. Following the computation similar to that used in the proof of Lemma 2.5 for computing the standard deviation, we obtain that for any $k \in \mathbb{Z}_p$,

$$\delta_k^{(\tau)} = \sqrt{\sum_{t=1}^{q} A_{kt}^2 \delta_k^{(\tau-1)}} = \rho\rho^{\tau-2}\delta = \rho^{\tau-1}\delta.$$

This completes the induction and thus the proof. □

**Theorem 2.8.** *If $\mathbf{g} := [g_j : j \in \mathbb{Z}_N]$ is a real Gaussian random vector with mean 0 and standard deviation $\delta$, $A \in \mathbb{R}^{p\times q}$ is a filter with orthogonal rows and satisfies the*

*condition (2.18), $m \geq 1$, $r > 0$ and $\mathbf{g}^\tau$ is defined as in (2.19), then for $\tau_1 > \tau_2$,*

$$\tilde{S}_m(\mathbf{g}^{\tau_1}, r^{\tau_1}) < \tilde{S}_m(\mathbf{g}^{\tau_2}, r^{\tau_2}).$$

*Proof.* The proof follows from Proposition 3.1 and the fact that the error function is strictly increasing.

We first re-express $\tilde{S}_m(\mathbf{g}^\tau, r^\tau)$ at the scale $\tau$. For each $s \in \mathbb{Z}_p$, let $r_s$ be defined in (2.15), $x_s$ be a fixed number and

$$E(x_s, r_s, \tau) := \operatorname{erf}\left(\frac{x_s + r_s}{\delta_s^{(\tau)}}\right) - \operatorname{erf}\left(\frac{x_s - r_s}{\delta_s^{(\tau)}}\right),$$

where $\delta_s^{(\tau)}$ is the standard deviation of of the $s$th element in a block of $\mathbf{g}^\tau$ at the scale $\tau$. Replacing $\delta(A, s)$ by $\delta_s^{(\tau)}$ in Proposition 3.1, we have that

$$\tilde{S}_m(\mathbf{g}^\tau, r^\tau) = -\ln \prod_{s=1}^{p} I_s(\tau),$$

where

$$I_s(\tau) := \frac{1}{\sqrt{2\pi}\delta_s^{(\tau)}} \int_{\mathbb{R}} E(x_s, r_s, \tau) \exp \frac{-x_s^2}{2(\delta_s^{(\tau)})^2} dx_s.$$

It remains to prove that for each $s \in \mathbb{Z}_p$, $I_s$ is strictly increasing. By employing Lemma 2.7 and with a change of variable, $y_s = \frac{x_s}{\rho^{\tau-1}}$, we observe that

$$I_s(\tau) = \frac{1}{\sqrt{2\pi}\delta} \int_{\mathbb{R}} \tilde{E}(y_s, r_s, \tau) \exp \frac{-y_s^2}{2\delta^2} dy_s$$

where

$$\tilde{E}(y_s, r_s, \tau) := \operatorname{erf}\left(\frac{y_s + r_s/\rho^{\tau-1}}{\delta}\right) - \operatorname{erf}\left(\frac{y_s - r_s/\rho^{\tau-1}}{\delta}\right).$$

Since erf is strictly increasing, for $0 < \rho < 1$, we have that $\tilde{E}(y_s, r_s, \tau_1) > \tilde{E}(y_s, r_s, \tau_2)$ when $\tau_1 > \tau_2$. Thus, $I_s$ is strictly increasing. □

It can be verified that the piecewise polynomial filters of order $k$ whose construction was described in [50] have orthogonal rows and satisfy the condition (2.18). Hence, the hypotheses of Theorem 2.8 are satisfied for this class of filters, and as a result, the corresponding filter based multiscale entropy of the Gaussian noise decreases as the scale increases. This fact is further confirmed by the numerical example. Numerical results from PLFME and also MSE for Gaussian white noise are presented in Fig. 2.1. Unless stated otherwise, all entropy values presented in this paper are computed by choosing $m := 2$ and $r$ being 15% percentage of the time series standard deviation. A similar pattern that entropy value decreases as the scale increases is shown in both of the methods.

**FME for $1/f$ Noise.** Now we apply FME to $1/f$ noise. Note that $1/f$ noise can be observed in various physical, chemical and biological systems [10]. It is the signal whose power spectral density is proportional to the reciprocal of its frequency. To describe $1/f$ noise, we recall complex Gaussian variables and the discrete Fourier transform. As usual, we let $i = \sqrt{-1}$ be the imaginary unit and denote the complex plane by $\mathbb{C}$. A complex variable $z := x + iy$ is called a complex Gaussian random variable if both $x$ and $y$ are real independent Gaussian variables with the mean 0 and the same standard deviation $\delta$. The corresponding probability density function for the complex Gaussian random variable $z$ is given by

$$\rho(z) := \frac{1}{\pi\delta_z^2} e^{|z|^2/\delta_z^2}, \quad z \in \mathbb{C},$$

where $\delta_z := \sqrt{2}\delta$. Given $n \in \mathbb{N}$, we let $\theta_n := \frac{2\pi}{2^n}$ and define the discrete Fourier

transform $F_n$ by a $2^n \times 2^n$ matrix

$$F_n := \frac{1}{2^n}[e^{-i\theta_n k\ell} : k \in \mathbb{Z}_{2^n}, \ell \in \mathbb{Z}_{2^n}]. \tag{2.21}$$

For a random vector $x$ taking values in $\mathbb{R}^{2^N}$, we use $\hat{\mathbf{x}}$ to denote the discrete Fourier transform of $\mathbf{x}$, that is, $\hat{\mathbf{x}} := F_N\mathbf{x}$. We write $\hat{\mathbf{x}} := [z_k : k \in \mathbb{Z}_{2^N}]^T$. It is well-known that the discrete Fourier Transform has the symmetric property

$$z_{2^{N-1}+k} = \bar{z}_{2^{N-1}-k}, \quad k \in \mathbb{Z}^+_{2^{N-1}}. \tag{2.22}$$

We need only to obtain the first $2^{N-1} + 1$ components of the vector $\hat{\mathbf{x}}$ since the remaining components may be obtained from the symmetry property.

We describe the $1/f$ noise following [2]. If $z_k$, $k \in \mathbb{Z}^+_{2^{N-1}-1}$, are independent complex Gaussian random variables with mean 0, $z_0$ and $z_{2^N-1}$ are real Gaussian random variables with mean 0, and there is a positive constant $c$ such that for all $k \in \mathbb{Z}_{2^{N-1}+1}$ the standard deviation $\delta_k$ of $z_k$ satisfy $\delta_k \leq \frac{c}{k+1}$, then we call $\mathbf{x}$ $1/f$ random vector and call an instance of $\mathbf{x}$ a $1/f$ noise. In this subsection, we use $\mathbf{f} := [f_k : k \in \mathbb{Z}_{2^N}]^T$ to denote a $1/f$ random vector and $\hat{\mathbf{f}} := [z_k : k \in \mathbb{Z}_{2^N}]^T$ to denote the discrete Fourier transform of $\mathbf{f}$. It is known that $1/f$ noise contains complex structures across multiple time scales [25,71]. We shall show that the filtered $1/f$ noise is again $1/f$ noise if the filters satisfy certain conditions. This indicates that the filtered $1/f$ signal is as complex as the original $1/f$ signal. In the remaining part of this section, we assume that the length of $\mathbf{f}$ is $2^N$ for a positive integer $N$.

We start with a simple filter $A$ which has the form

$$A := [\alpha, \beta],$$

where $\alpha, \beta \in \mathbb{R}$. Let $\mathbf{f}_A := (I_{2^{N-1}} \otimes A)\mathbf{f}$. We will show that the filtered $1/f$ signal, $\mathbf{f}_A$, is again $1/f$ signal for most of the filters of this form. We first need several technical lemmas.

From the definition of $1/f$ noise, we need to consider the Fourier transform of the filtered signal $\mathbf{f}_A$, which is

$$\hat{\mathbf{f}_A} = F_{N-1}(I_{2^{N-1}} \otimes A)\mathbf{f}. \tag{2.23}$$

Since $\mathbf{f} = F_N^{-1}F_N\mathbf{f}$, equation (3.12) may be rewritten as

$$\hat{\mathbf{f}_A} = F_{N-1}(I_{2^{N-1}} \otimes A)F_N^{-1}F_N\mathbf{f},$$

where $F_N^{-1}$ is the inverse discrete Fourier transform which has the form

$$F_N^{-1} := [e^{i\theta_N k\ell} : k \in \mathbb{Z}_{2^N}, \ell \in \mathbb{Z}_{2^N}].$$

We shall express the Fourier transform of the filtered signal in terms of the Fourier transform of the original signal. To this end, we investigate the matrix

$$\tilde{A} := F_{N-1}(I_{2^{N-1}} \otimes A)F_N^{-1}. \tag{2.24}$$

In the next lemma, we express $\tilde{A}$ in terms of two diagonal matrices

$$D^+ := \mathrm{diag}[d_k^+ : k \in \mathbb{Z}_{2^{N-1}}]$$

and

$$D^- := \mathrm{diag}[d_k^- : k \in \mathbb{Z}_{2^{N-1}}],$$

where

$$d_k^+ := \alpha + \beta e^{i\theta_N k}, \quad d_k^- := \alpha - \beta e^{i\theta_N k}, \quad k \in \mathbb{Z}_{2^{N-1}}.$$

**Lemma 2.9.** *For any positive integer $N$, there holds*

$$\tilde{A} = [D^+, D^-]. \tag{2.25}$$

*Proof.* From the definition of $F_N^{-1}$ and $A$, for $\ell \in \mathbb{Z}_{2^{N-1}}$, $k \in \mathbb{Z}_{2^N}$ it is straightforward to compute that

$$[(I_{2^{N-1}} \otimes A)F_N^{-1}]_{\ell,k} = e^{i\theta_{N-1}\ell k}(\alpha + \beta e^{i\theta_N k}). \tag{2.26}$$

Letting $D := \mathrm{diag}[D^+, D^-]$ and $\tilde{F}_N := [\alpha F_{N-1}^{-1}, \beta F_{N-1}^{-1}]$, it follows from (2.26) that

$$(I_{2^{N-1}} \otimes A)F_N^{-1} = \tilde{F}_N D. \tag{2.27}$$

From (2.27) we have that

$$\tilde{A} = [I_{N-1}, I_{N-1}]D. \tag{2.28}$$

Formula (2.25) is then obtained by substituting the expression of $D$ into (2.28). □

The discrete Fourier transform of a real vector is a complex vector. When analyzing $1/f$ noise, it is convenient to separate the real and imaginary parts of the discrete Fourier transform of a signal. Considering the symmetry property of the discrete Fourier transform, we define two operators $T_1$ and $T_2$. The operator $T_1$ projects a

vector of length $2^N$ to the vector of length $2^{N-1}$ consisting of the first $2^{N-1}$ components of the original vector and the operator $T_2$ projects a vector of length $2^N$ to the vector of length $2^{N-1}$ consisting of the last $2^{N-1}$ components of the original vector. For a real vector $\mathbf{x}$ of length $2^N$, we write its discrete Fourier transform as $\hat{\mathbf{x}} := \tilde{\mathbf{x}} + i\tilde{\mathbf{y}}$, where $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{y}}$ are two real vectors of length $2^N$. We let

$$\tilde{\mathbf{x}}_1 := T_1\tilde{\mathbf{x}}, \quad \tilde{\mathbf{x}}_2 := T_2\tilde{\mathbf{x}}, \quad \tilde{\mathbf{y}}_1 := T_1\tilde{\mathbf{y}}, \quad \tilde{\mathbf{y}}_2 := T_2\tilde{\mathbf{y}}.$$

In the next lemma, we express the discrete Fourier transform $\hat{\mathbf{x}}_A$ of $\mathbf{x}_A := (I_{2^{N-1}} \otimes A)\mathbf{x}$ in terms of $\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \tilde{\mathbf{y}}_1$ and $\tilde{\mathbf{y}}_2$. To simplify the notation, we introduce three diagonal matrices

$$A^+ := \mathrm{diag}[\alpha + \beta\cos(\theta_N k) : k \in \mathbb{Z}_{2^{N-1}}],$$

$$A^- := \mathrm{diag}[\alpha - \beta\cos(\theta_N k) : k \in \mathbb{Z}_{2^{N-1}}],$$

and

$$B = \mathrm{diag}[\beta\sin(\theta_N k) : k \in \mathbb{Z}_{2^{N-1}}].$$

Moreover, we denote by $\mathrm{Re}(\hat{\mathbf{x}}_A)$ and $\mathrm{Im}(\hat{\mathbf{x}}_A)$ the real part and imaginary part of $\hat{\mathbf{x}}_A$, respectively.

**Lemma 2.10.** *If* $\mathbf{x}$ *is a real vector of length* $2^N$*, then*

$$Re(\hat{\mathbf{x}}_A) := A^+\tilde{\mathbf{x}}_1 + A^-\tilde{\mathbf{x}}_2 - B\tilde{\mathbf{y}}_1 + B\tilde{\mathbf{y}}_2$$

*and*

$$Im(\hat{\mathbf{x}}_A) := A^+\tilde{\mathbf{y}}_1 + A^-\tilde{\mathbf{y}}_2 + B\tilde{\mathbf{x}}_1 - B\tilde{\mathbf{x}}_2.$$

*That is,*

$$\hat{\mathbf{x}}_A = Re(\hat{\mathbf{x}}_A) + iIm(\hat{\mathbf{x}}_A). \tag{2.29}$$

*Proof.* According to the definition of $\hat{\mathbf{x}}, \hat{\mathbf{x}}_A$ and $\mathbf{x}_A$, we have that

$$\hat{\mathbf{x}}_A = F_{N-1}A\mathbf{x} = F_{N-1}AF_N^{-1}\hat{\mathbf{x}} = \tilde{A}\hat{\mathbf{x}}. \tag{2.30}$$

Applying Lemma 2.9 to (2.30) yields that

$$\hat{\mathbf{x}}_A = (D^+, D^-)\hat{\mathbf{x}}. \tag{2.31}$$

We partition the real and imaginary parts of the vector $\hat{\mathbf{x}}$ as

$$\hat{\mathbf{x}} = \begin{pmatrix} \tilde{\mathbf{x}}_1 \\ \\ \tilde{\mathbf{x}}_2 \end{pmatrix} + i \begin{pmatrix} \tilde{\mathbf{y}}_1 \\ \\ \tilde{\mathbf{y}}_2 \end{pmatrix} \tag{2.32}$$

and notice that the definition of $D^+$, $D^-, A^+, A^-$ and $B$ gives us

$$D^+ = A^+ + iB, \quad D^- = A^- - iB. \tag{2.33}$$

The desired formula (2.29) may be obtained by substituting (2.33) and (2.32) into (2.31). □

The next lemma states that the independence of components of $\hat{\mathbf{x}}$ determines the independence of components of $\hat{\mathbf{x}}_A$.

**Lemma 2.11.** *Suppose that* $\mathbf{x}$ *is a real random vector of length* $2^N$*. If the first* $2^{N-1}+1$ *components of* $\hat{\mathbf{x}}$ *are independent, then the first* $2^{N-2}+1$ *components of* $\hat{\mathbf{x}}_A$ *are independent.*

Lemma 2.11 is a straightforward extension of a known result, Lemma 4.3 of [38]. We thus omit the proof.

We next recall a known result that describes the statistical property of the linear combinations of two independent complex Gaussian random variables. Its proof can be found in [38].

**Lemma 2.12.** *If $z_1$ and $z_2$ are independent complex Gaussian random variables with mean 0 and standard derivation $\delta_1$ and $\delta_2$, respectively, then for each pair of complex numbers $a := a_1 + ia_2, b := b_1 + ib_2$ with $a_1, a_2, b_1, b_2 \in \mathbb{R}$, $az_1 + bz_2$ is a complex Gaussian random variable with mean 0 and standard derivation*

$$\delta := (a_1^2\delta_1^2 + a_2^2\delta_1^2 + b_1^2\delta_2^2 + b_2^2\delta_2^2)^{1/2}. \tag{2.34}$$

In the next lemma, we describe $\hat{\mathbf{x}}_A$ in terms of $\hat{\mathbf{x}}$.

**Lemma 2.13.** *If the first $2^{N-1}+1$ components of $\hat{\mathbf{x}}$ are independent complex Gaussian random variables, then the first $2^{N-2}+1$ components of $\hat{\mathbf{x}}_A$ are independent complex (except the first and the $(2^{N-2}+1)$th components which are real) Gaussian random variables. Moreover, for each $k \in \mathbb{Z}_{2^{N-1}+1}$, if the mean of the $(k+1)$th component of $\hat{\mathbf{x}}$ is 0 and its standard deviation is $\delta_k$, then the mean of each component of $\hat{\mathbf{x}}_A$ is 0 and the standard deviation of the $(k+1)$th component of $\hat{\mathbf{x}}_A$ is*

$$\delta_{A,k} = [(\alpha^2 + \beta^2 + \gamma_k)\delta_k^2 + (\alpha^2 + \beta^2 - \gamma_k)\delta_{2^{N-1}-k}^2]^{1/2}, \tag{2.35}$$

*where $\gamma_k := 2\alpha\beta\cos(\theta_N k)$.*

*Proof.* The description of the components of $\hat{\mathbf{x}}_A$ relies on Lemma 2.10. We write $\hat{\mathbf{x}} := [z_k : k \in \mathbb{Z}_{2^N}]$ and $\hat{\mathbf{x}}_A := [z_{A,k} : k \in \mathbb{Z}_{2^{N-1}}]$. We first prove that $z_{A,0}$ and $z_{A,2^{N-2}}$ are real Gaussian random variables. By Lemma 2.10, we have that

$$z_{A,0} = \alpha(z_0 + z_{2^{N-1}}). \tag{2.36}$$

Since $z_0$ and $z_{2^{N-1}}$ are real Gaussian random variables, from (2.36) and Lemma 2.1 we know that $z_{A,0}$ is a real Gaussian random variable. Noting $2^{N-1} - 2^{N-2} = 2^{N-2}$, it also follows from Lemma 2.10 that

$$z_{A,2^{N-2}} = \alpha(x_{2^{N-2}} - y_{2^{N-2}}), \tag{2.37}$$

where $x_{2^{N-2}}$ and $y_{2^{N-2}}$ are respectively the real part and the image part of $z_{2^{N-2}}$. Hence, we conclude that $z_{A,2^{N-2}}$ is also a real Gaussian random variable. By Lemma 2.10, we have that

$$z_{A,k} = (\alpha + \beta e^{i\theta_N k}) z_k + (\alpha - \beta e^{i\theta_N k}) z_{2^{N-1}+k}. \tag{2.38}$$

Thus, by Lemma 2.12 $z_{A,k}$ is a complex Gaussian random variable. The independence of the random variables $z_{A,k}$, $k \in \mathbb{Z}_{2^{N-2}+1}$, is ensured by Lemma 2.11 from the fact that the random variables $z_k$, $k \in \mathbb{Z}_{2^{N-1}+1}$, are independent. By applying Lemma 2.12 to equation (2.38) with the symmetric property (2.22) and Lemma 2.1 to equations (2.36) and (2.37), we obtain (2.35). □

Now we are are ready to give the result that the filtered $1/f$ signal, $\mathbf{f}_A$, is again $1/f$ signal for most of the filters of this form. To present this result, we let $\delta_k, \delta_{A,k}$ denote the standard derivations of the $(k+1)$th random variable of $\hat{\mathbf{f}}$ and $\hat{A}\mathbf{f}$, respectively.

**Proposition 2.14.** *If* $\mathbf{f}$ *is a* $1/f$ *random vector and* $(\alpha, \beta)$ *satisfies the condition*

$$\alpha + \beta \neq 0, \tag{2.39}$$

*then* $A\mathbf{f}$ *is also a* $1/f$ *random vector. Moreover, if there exists a positive constant* $c$ *such that for all* $k \in \mathbb{Z}_{2^{N-1}+1}$, $\delta_k^2 \leq \frac{c}{1+k}$, *then for all* $k \in \mathbb{Z}_{2^{N-2}+1}$, $\delta_{A,k}^2 \leq \frac{c'}{k+1}$, *where* $c' = 2(\alpha^2 + \beta^2)c$.

*Proof.* We denote by $\hat{A}\mathbf{f}$ the discrete Fourier transform of $A\mathbf{f}$. To prove $\hat{A}\mathbf{f}$ is a $1/f$ random vector, we need to show that the first $2^{N-2}+1$ elements of $\hat{A}\mathbf{f}$ are independent Gaussian random variables with mean 0, and for all $k \in \mathbb{Z}_{2^{N-2}+1}$, $\delta_{A,k} \leq \frac{c'}{1+k}$ for some constant $c'$.

Since $\mathbf{f}$ is a $1/f$ random vector, from the definition of a $1/f$ random vector we know that the first $2^{N-1}+1$ elements of $\hat{\mathbf{f}}$ are independent Gaussian random variables with mean 0. Thus, from Lemma 2.13, we have that the first $2^{N-2}+1$ elements of $\hat{A}\mathbf{f}$ are independent Gaussian random variables with mean 0.

We next show that for all $k \in \mathbb{Z}_{2^{N-1}+1}$, $\delta_{A,k} \leq \frac{c'}{1+k}$ for some constant $c'$. Let

$$\gamma_k^+ := \alpha^2 + \beta^2 + 2\alpha\beta\cos(\theta_N k)$$

and

$$\gamma_k^- := \alpha^2 + \beta^2 - 2\alpha\beta\cos(\theta_N k).$$

When $\alpha + \beta \neq 0$, we have that $\gamma_k^+ \neq 0$ and $\gamma_k^- \neq 0$ for all $k \in \mathbb{Z}_{2N-1+1}$. Thus, from

(2.35), we have for all $k \in \mathbb{Z}_{2^{N-2}+1}$ that

$$\begin{aligned}\delta_{A,k}^2 &= \gamma_k^+ \delta_k^2 + \gamma_k^- \delta_{2^{N-1}-k}^2 \\ &\le \gamma_k^+ \frac{c}{1+k} + \gamma_k^- \frac{c}{1+2^{N-1}-k} \\ &= \frac{c}{1+k} \frac{(1+2^{N-1}-k)\gamma_k^+ + (1+k)\gamma_k^-}{1+2^{N-1}-k}. \end{aligned} \tag{2.40}$$

Note that $k \in \mathbb{Z}_{2^{N-2}+1}$ and thus $0 \le \cos(\theta_N k) \le 1$. Then the second fraction in the last term of formula (2.40) is an increasing function of $k$, which has the maximum value $2(\alpha^2+\beta^2)$ when $k = 2^{N-2}$. This gives the desired estimate. □

Note that the special case of Proposition 2.14 with $[\alpha, \beta] := [\frac{1}{2}, \frac{1}{2}]$ was proved in [38]. If $(\alpha, \beta)$ does not satisfy the condition (2.39), then $\alpha = -\beta$. In this case, $(I_{2^{N-1}} \otimes A)\mathbf{f}$ may not be $1/f$ noise. For example, when $A$ is the high pass Haar filter, that is, $A := [\frac{1}{2}, -\frac{1}{2}]$, it was verified in [38] by a numerical experiment that $(I_{2^{N-1}} \otimes A)\mathbf{f}$ is not $1/f$ noise and the entropy value of the filtered $1/f$ signal will decrease as the scale increases. In addition, if we take $A := [1, 0]$ (resp. $A := [0, 1]$) in Proposition 2.14, we can see that the random vector consisting of the odd components (resp. the even components) of a $1/f$ random vector is still a $1/f$ random vector. This result is summarized in the next corollary.

**Corollary 2.15.** *If* $\mathbf{f} := [f_0, f_1, \ldots, f_{N-1}]$ *is a* $1/f$ *random vector, then* $\mathbf{u} := [u_m : m \in \mathbb{Z}_{N/2+1}]$ *with* $u_m = f_{2m}$ *and* $\mathbf{v} := [v_m : m \in \mathbb{Z}_{N/2+1}]$ *with* $v_m = f_{2m+1}$ *are also* $1/f$ *random vectors.*

Though the result in Proposition 2.14 is only for filters of a simple form, it can

be utilized (together with Corollary 2.15) to establish below that a filtered $1/f$ noise is still $1/f$ noise.

**Proposition 2.16.** *If* $\mathbf{f}$ *is* $1/f$ *random vector and the matrix* $A$ *is defined as in (2.1), then* $(I_{2^{N-2}} \otimes A)\mathbf{f}$ *is also a* $1/f$ *random vector.*

To prove Proposition 2.16, we need several lemmas. We first recall a known fact of the discrete Fourier transform, whose proof may be found in [26] (Lemma 2.37).

**Lemma 2.17.** *Suppose* $M \in \mathbb{N}$*, and* $N = 2M$*. Let* $\mathbf{z} \in \ell^2(\mathbb{Z}_N)$*. Define* $\mathbf{u}, \mathbf{v} \in \ell^2(\mathbb{Z}_M)$ *by*

$$u_k = z_{2k} \quad \textit{for} \quad k \in \mathbb{Z}_M,$$

*and*

$$v_k = z_{2k+1} \quad \textit{for} \quad k \in \mathbb{Z}_M.$$

*Let* $\hat{\mathbf{z}}$ *denote the discrete Fourier transform of* $\mathbf{z}$ *defined on* $N$ *points. Let* $\hat{\mathbf{u}}$*,* $\hat{\mathbf{v}}$ *denote the discrete Fourier transform of* $\mathbf{u}$ *and* $\mathbf{v}$ *respectively, defined on* $M = N/2$ *points. Then for* $m \in \mathbb{Z}_M$*,*

$$\hat{z}_m = \hat{u}_m + e^{-2\pi i m/N}\hat{v}_m. \tag{2.41}$$

*Also, for* $m = M, M+1, M+2, \ldots, N-1$*, let* $\ell = m-N$*. Note that the corresponding values of* $\ell$ *are* $\ell = 0, 1, \ldots, M-1$*. Then*

$$\hat{z}_m = \hat{z}_{\ell+M} = \hat{u}_\ell - e^{-2\pi i m/N}\hat{v}_\ell. \tag{2.42}$$

From Lemma 2.17, we shall show in the next lemma that the vector obtained by interlacing two $1/f$ random vectors is again a $1/f$ random vector.

**Lemma 2.18.** *If* $\mathbf{g}$ *and* $\mathbf{h}$ *are* $1/f$ *random vectors of length* $n$*, then the random vector* $\mathbf{f}$ *defined by*

$$\mathbf{f} := [g_0, h_0, g_1, h_1, \ldots, g_{n-1}, h_{n-1}]$$

*is also a* $1/f$ *random vector.*

*Proof.* Since $\mathbf{g}$ and $\mathbf{h}$ are $1/f$ random vectors, by the definition of the $1/f$ random vector we know that $\hat{g}_k$ and $\hat{h}_k$, $k \in \mathbb{Z}_n$, are independent Gaussian random variables with mean 0 and there exist positive constants $c, c'$ such that their standard deviations $\delta_{g,k}$ and $\delta_{h,k}$ satisfy

$$\delta_{g,k} \leq \frac{c}{1+k} \quad \text{and} \quad \delta_{h,k} \leq \frac{c'}{1+k}, \tag{2.43}$$

for any $k \in \mathbb{Z}_n$.

It follows from Lemma 2.17 that for $m \in \mathbb{Z}_n$,

$$\hat{f}_m = \hat{g}_m + e^{-2\pi i m/N} \hat{h}_m, \tag{2.44}$$

and for $m = n, n+1, n+2, \ldots, 2n-1$,

$$\hat{f}_m = \hat{g}_{m-n} - e^{-2\pi i m/N} \hat{h}_{m-n}. \tag{2.45}$$

By applying Lemma 2.12 to equations (2.44) and (2.45) and using conditions (2.43), we know for $m \in \mathbb{Z}_{2n}$ that $\hat{f}_m$ is a Gaussian random vector with mean 0 and the standard deviation of $\hat{f}_m$, $\delta_{f,m}$, satisfies

$$\delta_{f,m} \leq \frac{c''}{1+m},$$

where $c'' = \sqrt{(1+n)(2c^2 + (c')^2)}$. The independence of $\hat{f}_m, m \in \mathbb{Z}_{2n}$ can be proved by using a similar computation used in the proof of Lemma (2.5) for proving the independence considering that $[1, e^{-2\pi im/N}]$ and $[1, -e^{-2\pi im/N}]$ are orthogonal. Therefore, $\mathbf{f}$ is a $1/f$ random vector. □

We next recall a known fact that the sum of two $1/f$ random vectors is also a $1/f$ random vector. Its proof can be found in [63].

**Lemma 2.19.** *If* $\mathbf{f}_1$ *and* $\mathbf{f}_2$ *are* $1/f$ *random vectors, then* $\mathbf{f}_1 + \mathbf{f}_2$ *is also a* $1/f$ *random vector.*

We are now ready to prove Proposition 2.16. The filtered $1/f$ random vector through the piecewise linear filter $A$ defined in (2.1) is again a $1/f$ random vector. In the remaining part of this section, $A$ is referred to the piecewise linear filter defined in (2.1).

*Proof of Proposition 2.16:* Let $A_1$ and $A_2$ denote the matrices formed respectively by the first row and the second row of $A$. Since $(I_{2^{N-2}} \otimes A)\mathbf{f}$ may be obtained by interlacing $(I_{2^{N-2}} \otimes A_1)\mathbf{f}$ and $(I_{2^{N-2}} \otimes A_2)\mathbf{f}$, according to Lemma 2.18, it suffices to prove that $(I_{2^{N-2}} \otimes A_1)\mathbf{f}$ and $(I_{2^{N-2}} \otimes A_2)\mathbf{f}$ are $1/f$ random vectors.

Let

$$\mathbf{f}_1 := [f_0, f_2, \ldots, f_{N-2}],$$

$$\mathbf{f}_2 := [f_0, f_1, f_4, f_5, \ldots, f_{N-4}, f_{N-3}]$$

and

$$\mathbf{f}_3 := [f_2, f_3, f_6, f_7, \ldots, f_{N-2}, f_{N-1}].$$

Since $\mathbf{f}$ is a $1/f$ random vector, it follows from Lemma 2.18 and Corollary 2.15 that $\mathbf{f}_1$, $\mathbf{f}_2$ and $\mathbf{f}_3$ are $1/f$ random vectors. By Proposition 2.14, $(I_{2^{N-2}} \otimes [1/2, 1/2])\mathbf{f}_1$ is a $1/f$ random vector. Since

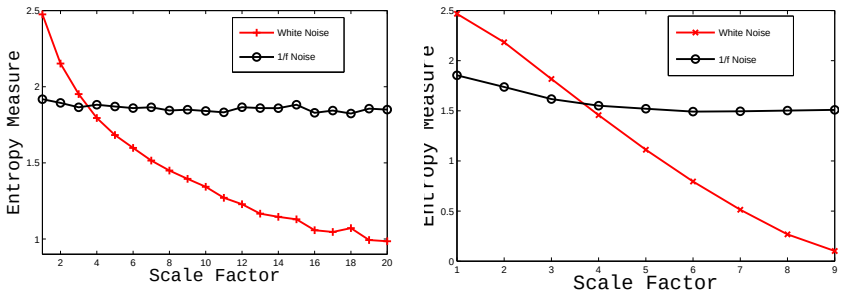
$$(I_{2^{N-2}} \otimes A_1)\mathbf{f} = (I_{2^{N-2}} \otimes [1/2, 1/2])\mathbf{f}_1,$$

we conclude that $(I_{2^{N-2}} \otimes A_1)\mathbf{f}$ is a $1/f$ random vector. Note that

$$(I_{2^{N-2}} \otimes A_2)\mathbf{f} = (I_{2^{N-2}} \otimes [-\sqrt{3}/4, 1/4])\mathbf{f}_2 + (I_{2^{N-2}} \otimes [\sqrt{3}/4, 1/4])\mathbf{f}_3. \tag{2.46}$$

By Proposition 2.14, both $(I_{2^{N-2}} \otimes [-\sqrt{3}/4, 1/4])\mathbf{f}_2$ and $(I_{2^{N-2}} \otimes [\sqrt{3}/4, 1/4])\mathbf{f}_3$ are $1/f$ random vectors. Hence, by Lemma 2.19 and (2.46), $(I_{2^{N-2}} \otimes A_2)\mathbf{f}$ is a $1/f$ random vector. □

Numerical results from MSE and PLFME for $1/f$ noise are presented in Fig. 2.1. Results from both of these methods are consistent with the fact that $1/f$ noise contains complex structures across multiple scales [25, 71].




(a) MSE (b) PLFME

Figure 2.1: MSE and PLFME for Gaussian white noise (mean 0, variance 1) and $1/f$ noise, $N = 8 \times 10^4$.

## 2.3 Application to Human Heartbeat Interval Time Series

We apply FME to HHITS to study the loss of complexity, a generic feature of pathologic dynamics. Specifically, we apply the piecewise linear filter recursively to HHITS of healthy young subjects (YOUNG), healthy old subjects (OLD), subjects with the cardiac arrhythmia, atrial fibrillation (AF) and subjects with severe congestive heart failure (CHF), and compute BSE of the resulting signals of multiple scales. We test the hypothesis that healthy interbeat interval dynamics are more complex than those with pathology. Our numerical results also suggest that aging may reduce the complexity of the heart interbeat interval more than CHF. This finding is robust to data of different lengths.

In this consideration, the use of the piecewise linear filter in FME is motivated

by a study of the biological mechanism of the cardiac system described by HRT [59]. HRT describes short-term fluctuations in the sinus cycle length that follow spontaneous ventricular premature complexes (VPCs). The physiologic pattern described in HRT consists of brief heart rate acceleration, which is followed by more gradual heart deceleration before the rate returns to a pre-ectopic level. Following singular VPCs, the HRT pattern is frequently masked by heartbeat interval time series. Consequently, HRT is usually assessed from Holter recordings as an average response to VPCs over longer periods (e.g., 24 hours). From such recordings, the VPC tachogram is constructed, by aligning and averaging sequences of heartbeat interval time series surrounding isolated VPCs. Fig. 2.2 from [8] is VPCs tachograms showing normal (left) and abnormal (right) HRT. From Fig. 2.2, we observe that different heart conditions show distinguished differences in HRT pattern.

Several parameters characterizing the HRT pattern were proposed [9, 35, 59, 69]. Among these parameters, turbulence onset (TO) and turbulence slope (TS) [59] quantify two phases of HRT, early acceleration and late deceleration. These two parameters are meaningful in the clinical use, especially in risk prediction and monitoring of disease progression in several pathologies (see [8] and the references therein). For example, compared to healthy control patients, patients with congestive heart failure have significantly depressed HRT indexes [39]. Also, increasing age is associated with a decrease in the HRT index [60]. It is reported in [67] that TO is significantly less negative before AF occurs than during the remaining part of recordings. TO and TS

represent the regression slope of the corresponding sequences of HHITS (acceleration and deceleration) surrounding isolated VPCs. The acceleration phase surrounding an isolated VPC is characterized by a negative value of TO and the deceleration phase surrounding an isolated VPC is characterized by a positive value of TS. TO and TS are constants for each sequence of HHITS surrounding VPCs and representing the acceleration and deceleration phase, respectively. This implies that regression slops of the acceleration and deceleration sequences of HHITS surrounding VPCs have a piecewise constant pattern. Therefore, sequences of HHITS surrounding VPC may have a piecewise linear pattern. This inspires us to use piecewise linear filter for the cardiac signal to capture its piecewise linear pattern. Though HRT only describes sequences of HHITS surrounding VPCs, we use piecewise linear filter for the entire HHITS assuming that the piecewise linear pattern represents the entire HHITS better than the piecewise constant pattern, which is captured by the piecewise constant filter used in MSE. In Fig. 2.3, we compare the HHITS with their piecewise linear representations, where the signals in column (b) are obtained from the original HHITS in column (a) by the linear filter $\frac{1}{4}[1, 2, 1]$. This shows that HHITS may be well represented by a piecewise linear curve.

According to the discussion above, we use PLFME for the cardiac signal. Specifically, in PLFME, the original time series is filtered by the piecewise linear filter $A$ defined in (2.1) recursively at multiple scales. That is, we define

$$\mathbf{y}^1 := \mathbf{x},\ \mathbf{y}^\tau := (I_{n_{\tau-1}} \otimes A)\mathbf{y}^{\tau-1}$$

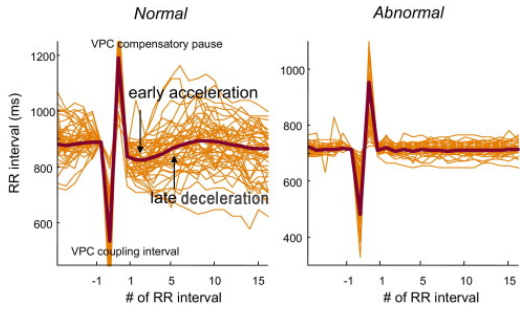


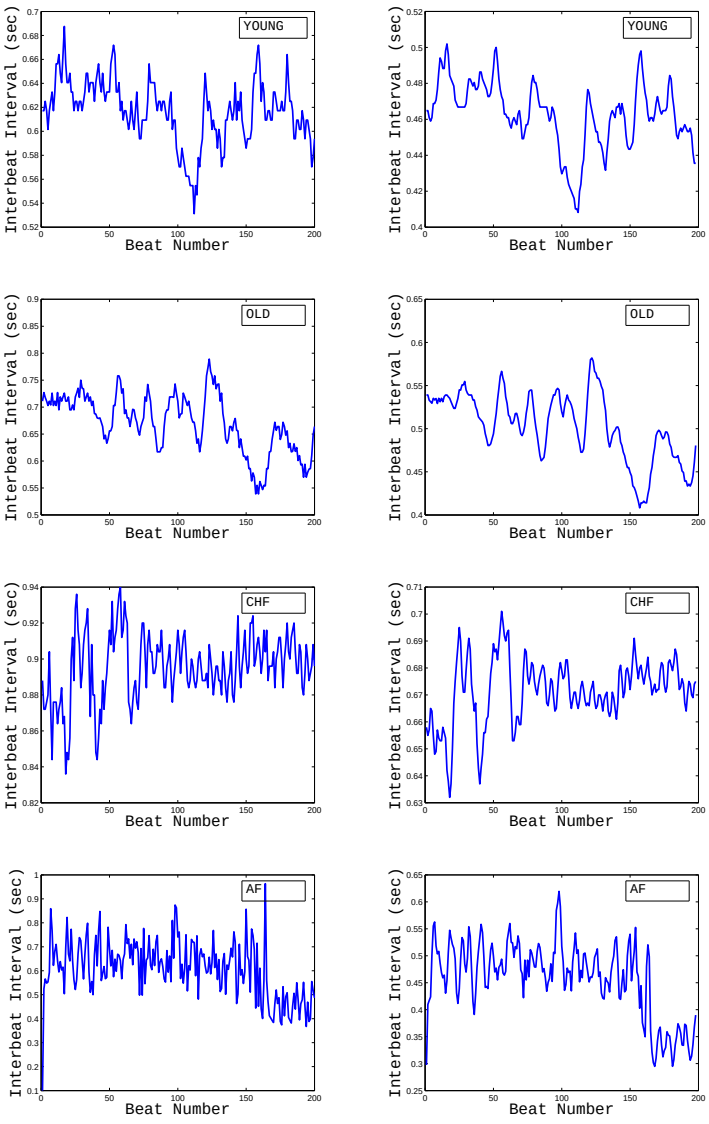

Figure 2.2: VPC tachograms showing normal (left) and abnormal (right) HRT. Orange curves show single VPC tachograms. Bold brown curves show the averaged VPC tachogram over 24 hours.

where $n_{\tau-1} := \lfloor \frac{N_{\tau-1}}{4} \rfloor$ and $N_{\tau-1}$ is the length of the time series $\mathbf{y}^{\tau-1}$.

We next present numerical results of PLFME. We compare MSE and PLFME of the time series of consecutive heart beat intervals derived from 20 YOUNG, 20 OLD, 7 AF and 20 CHF subjects of data lengths $N = t \times 10^4$ where $t = 3, 4, \ldots, 8$. The entropy value of each group shown in Fig. 2.4 and Fig. 2.5 is the mean entropy value of the group. MSE values are computed using the software provided in [28]. We see that PLFME improves the robustness of MSE to the data of different lengths.

The most significant difference between the results of MSE and PLFME occurs in the OLD and CHF groups.

We first look at the results of MSE (Fig. 2.4). When $N = 3 \times 10^4$, the same data length as used in [21, 22], we obtain the same results as those in [21, 22]. There is an inconsistency in the results for data of different lengths. As $N$ increases, the curve representing entropy values at different scales of the OLD group (BLACK)




(a) (b)

Figure 2.3: The comparison of the original signal (column (a)) and their piecewise linear representations (column (b)).

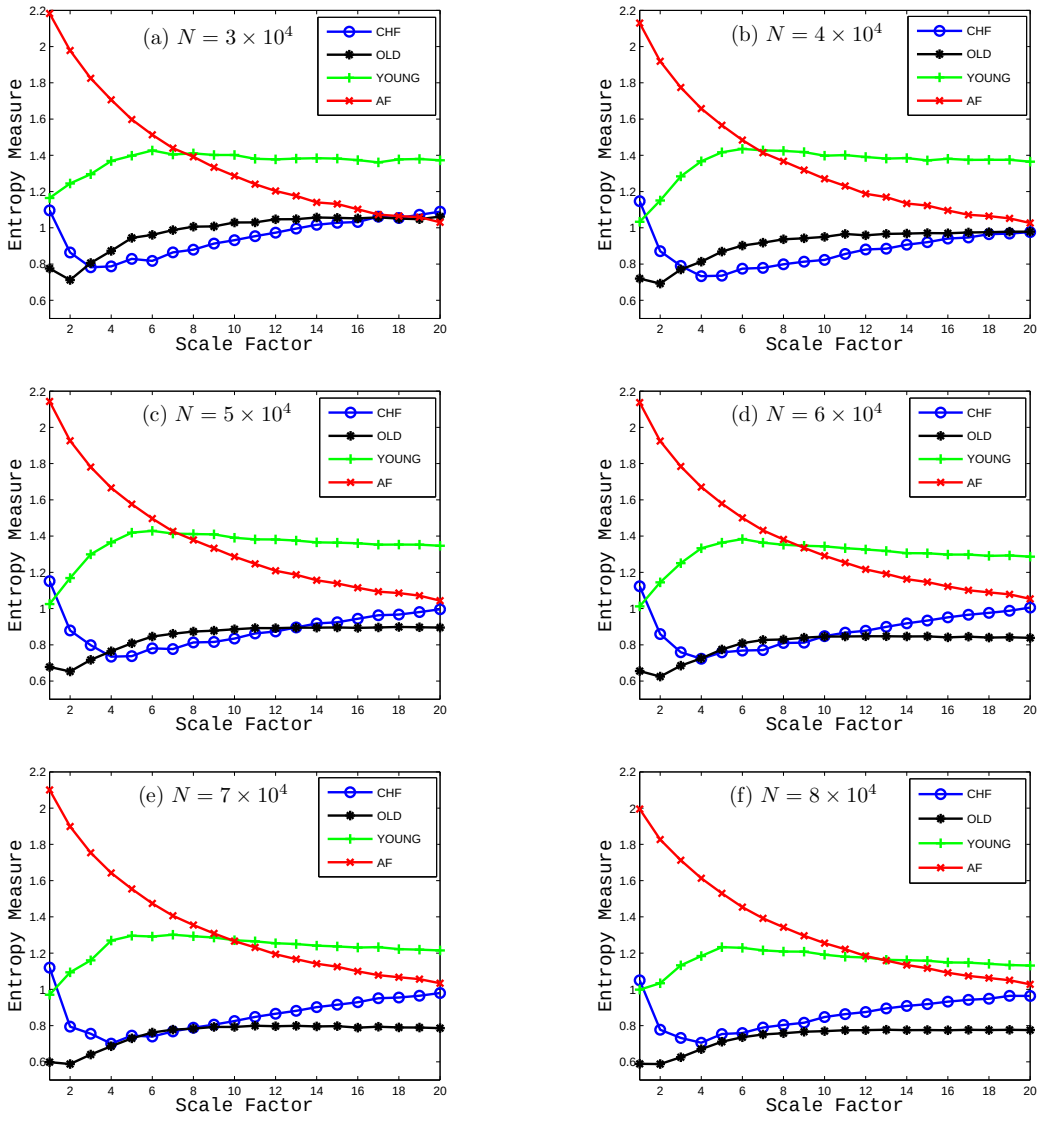



Figure 2.4: The MSE analysis for human cardiac interbeat interval time series with data of various lengths.

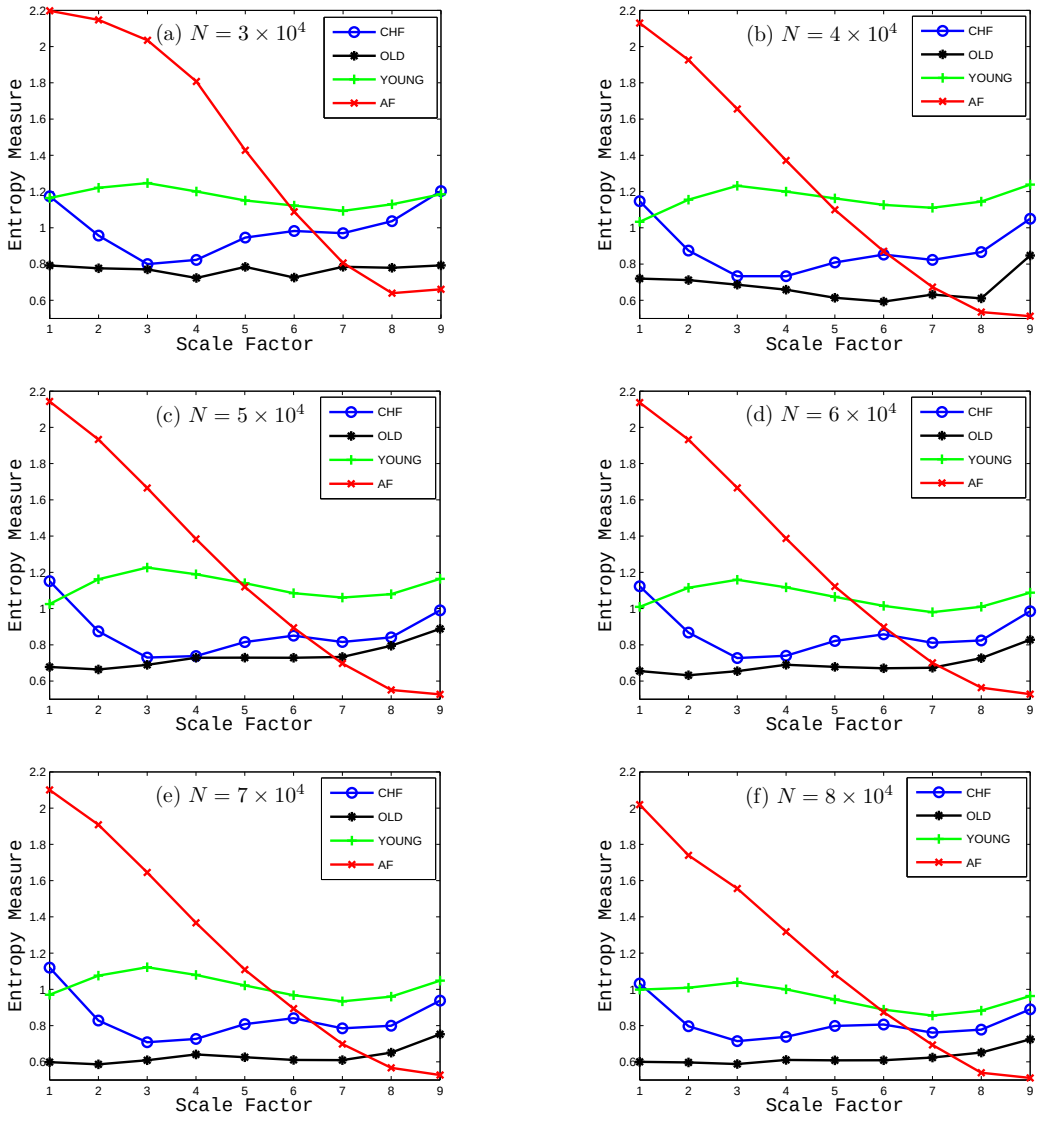



Figure 2.5: The PLFME analysis for human cardiac interbeat interval time series with data of various lengths.

gradually drops below that representing entropy values of the CHF group (BLUE). In particular, the curve BLACK is above the curve BLUE at scales 4 to 12 when $N = 5 \times 10^4$. The curve BLUE exceeds beyond the curve BLACK at all scales when $N = 8 \times 10^4$, suggesting the cardiac system from the CHF group is more complex than that from the OLD group. This contradicts the conclusion made in [21]. It is known that both disease and aging will reduce the biological complexity [29]. However, which condition loses the complexity more remains inconclusive from MSE, which provides different results from data of different lengths.

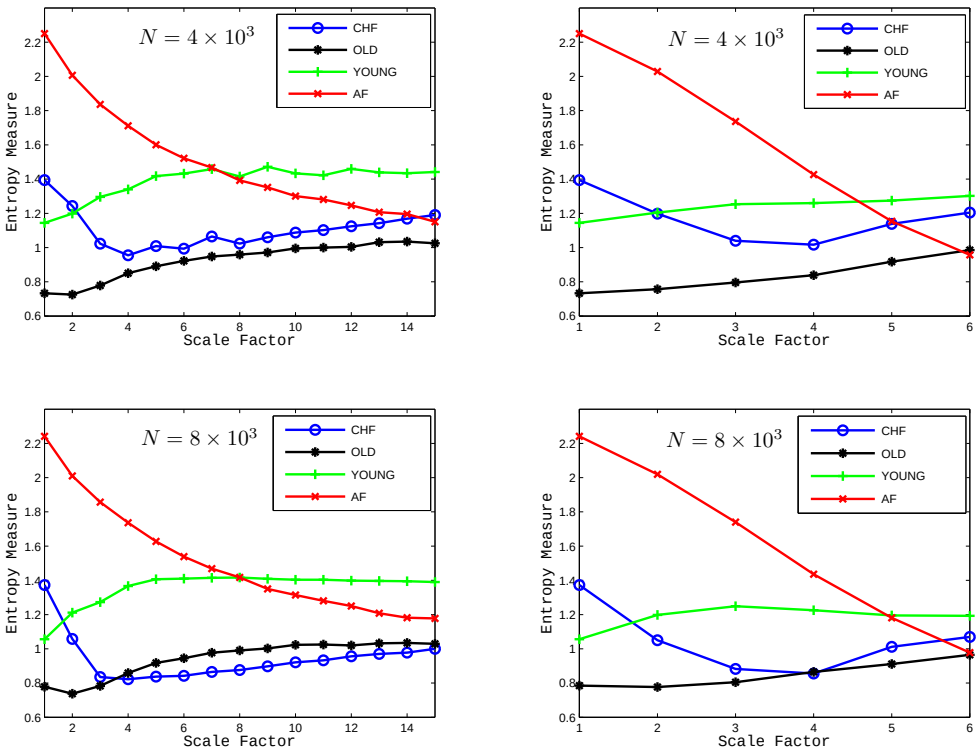
Next we look at the results of PLFME. Notice that for data of the OLD and CHF groups of different lengths, the results of FLFME are more robust than those of MSE. For all of the values of $N$ reported in Fig. 2.5, the curve BLUE is above the curve BLACK at all scales. It suggests that the cardiac interbeat intervals may lose more complexity from aging than from the CHF group. This result is consistent with that of MSE when $N = 8 \times 10^4$. Moreover, the curves representing entropy values at different scales of the OLD and CHF groups are better separated by PLFME than by MSE. This is further confirmed by the classification result presented in section .

Several additional observations for the PLFME method from the Fig. 2.5 are made: The entropy values of the AF group have a decreasing pattern similar to that of the white noise. The entropy values at all scales of the YOUNG group are constantly higher than those of other groups, except the first 5 scales in the AF group. The entropy values of the OLD group have a pattern similar to that of $1/f$

signal. Moreover, we point out that the difference of the entropy measures among each group varies on different scales in the PLFME method. For example, As seen in Fig. 2.5(f), the largest difference between the YOUNG group and the CHF group is from scales 2 to 4 while the OLD group and the CHF group are separated best at scales 5 and 6 when $N = 8 \times 10^4$. Therefore, as in the original MSE analysis, both of the entropy values and their dependence on resolution have to be taken into consideration to better characterize a physiologic process.

Obtaining long time series may be difficult and expensive in practical applications. Data analysis with short time series is highly desirable. Applications of MSE to short term physiological recordings were recently studied in [1, 65]. We compare the performance of MSE and PLFME for shorter HHITS with data lengths $N = 4, 8 \times 10^3$. The numerical results shown in Fig. 2.6 from the shorter data are consistent with those shown in Fig. 2.4 and Fig. 2.5 from the longer data.

To close this section, we compare PLFME with two refinements of MSE proposed by [36,66]. In [66], the averaging process of MSE was interpreted as the finite-impulse filter (FIR) and a refined MSE (RMSE) was proposed based on the replacement of the FIR filter with a low-pass Butterworth filter, which aims to reduces aliasing when the filtered series are downsampled. In [36], adaptive MSE (AMSE) method was proposed by using empirical mode decomposition to extract the lower frequency components of the time series at different scales. We performed both RMSE and AMSE methods on our data (Fig. 2.7). Both of these methods do not provide satisfactory numerical



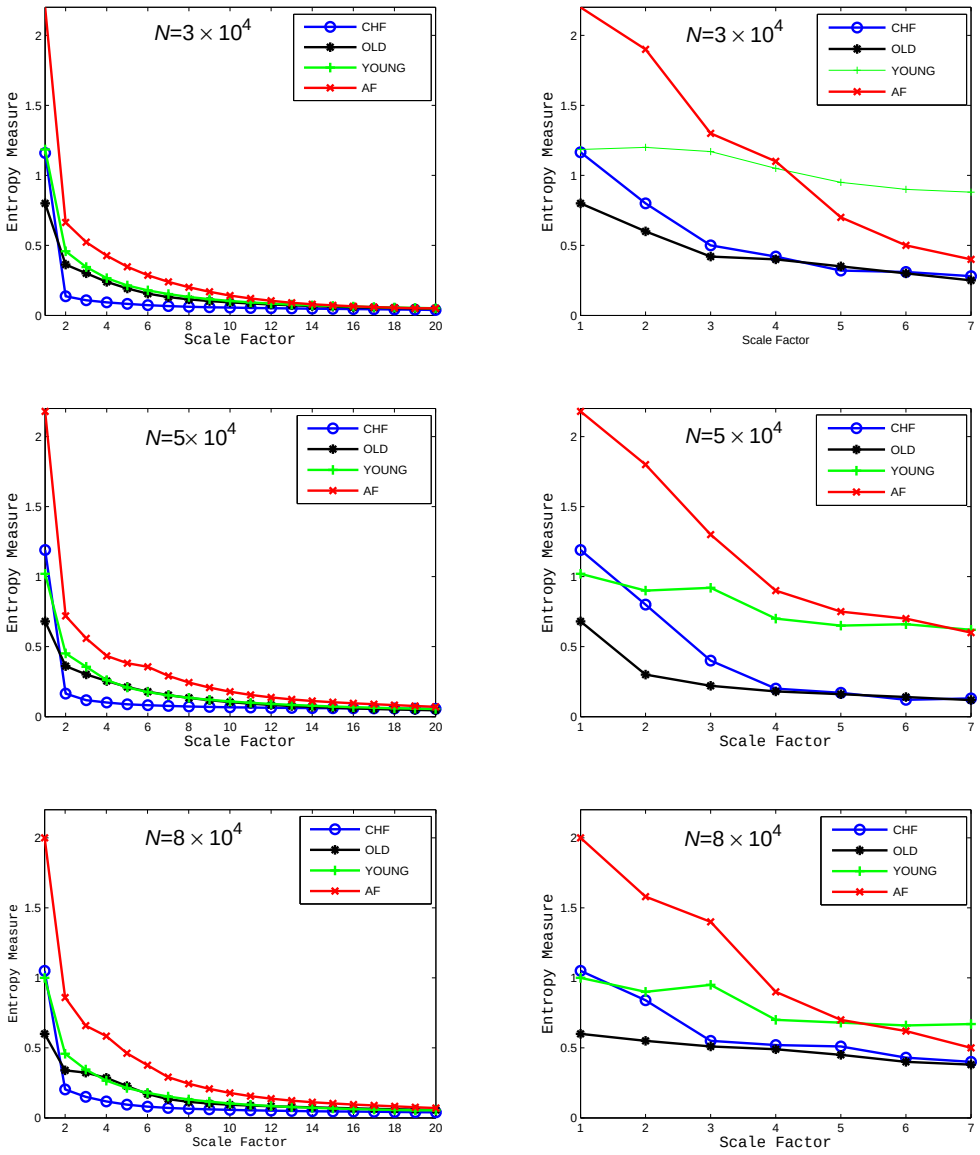

(a) MSE (b) PLFME

Figure 2.6: MSE and PLFME for shorter HHITS data.

results. Numerical results of RMSE (Fig. 2.7(a)) do not give any evidence that cardiac systems from healthy young subjects are more complex than those from pathologic subjects. Numerical results of AMSE (Fig. 2.7(b)) do not discriminate OLD and CHF groups as well as those from MSE and PLFME, especially when the length of the time series is small. The reason that PLFME and MSE give better description of the complexity of heartbeat interval time series than RMSE and AMSE may be due to the fact that PLFME and MSE filter the time series more "locally" than RMSE and AMSE.

## 2.4 Adaptive Filters

When prior knowledge of the system that generates a time series is not available, we propose to use adaptive filters constructed from the time series to compute its FME. Since traditional entropy methods quantify the degree of regularity of a time series by evaluating the appearance of its repetitive patterns and since consecutive components which are close to each other (measured by $r$ in sample entropy) are considered as a repetitive pattern, one may consider these repetitive patterns as single units which will present the regularity of the whole system. We further consider the multiscale structure of these patterns to measure the complexity of the physiological system by using adaptive filters. To illustrate the process of constructing an adaptive filter, we present below APCFME as an example. The idea is applicable to constructing adaptive piecewise polynomial filters of order $k$.




(a) RMSE (b) AMSE

Figure 2.7: RMSE and AMSE for HHITS data.

Given a time series, we group it according to its repetitive patterns. Specifically, for a time series $\mathbf{x}$ of length $N$, we write $\mathbf{x} = \{\mathbf{x}_0, \ldots, \mathbf{x}_{n-1}\}$ with disjoint $\mathbf{x}_j$, being a repetitive pattern of $\mathbf{x}$. Each $\mathbf{x}_j$ sequence consisting of several consecutive components of $\mathbf{x}$ and the distance between any two components in $\mathbf{x}_j$ is not bigger than a preselected tolerance $r$. For each $\mathbf{x}_j$, we use $|\mathbf{x}_j|$ to denote the number of its elements. A new coarse-grained time series $\mathbf{y}$ of length $n$ is generated by

$$\mathbf{y}(k) := \frac{1}{|\mathbf{x}_k|} \sum_{x(j)\in\mathbf{x}_k} x(j), \quad k \in \mathbb{Z}_n. \tag{2.47}$$

In other words, $\mathbf{x}$ is filtered by an $n \times n$ block diagonal matrix, whose $j$th diagonal block is the $1 \times |\mathbf{x}_j|$ matrix

$$\left[\frac{1}{|\mathbf{x}_j|}, \frac{1}{|\mathbf{x}_j|}, \ldots, \frac{1}{|\mathbf{x}_j|}\right].$$

This *adaptive piecewise constant filter* (APCF) is different from the piecewise constant filter used in MSE. We choose $m = 1$ and an increasing sequence $\{r_0, r_1 \ldots, \}$ at different scales to compute sample entropy. APCFME is then computed by the procedure:

(1) Compute $S_1(\mathbf{x}, r_0)$.

(2) At scale $\tau \geq 1$, a new time series $\mathbf{y}^\tau$ is generated from $\mathbf{y}^{\tau-1}$ with $\mathbf{y}^0 := \mathbf{x}$ using APCF. The parameter $r$ used to construct APCF is chosen as $r_{\tau-1}$.

(3) Compute $S_1(\mathbf{y}^\tau, r_\tau)$.

We now discuss the choice of the parameters $r_j$ used in APCFME. Constructing coarse-grained time series as described in (3.6) is equivalent to averaging several

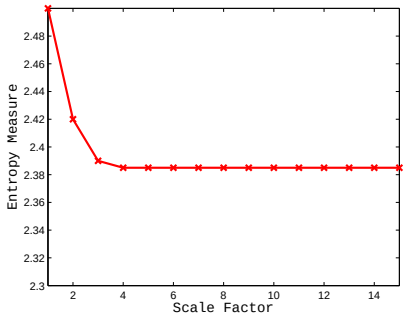
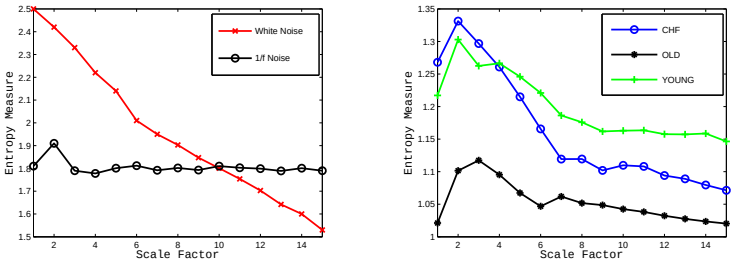



Figure 2.8: APCFME results for simulated Gaussian white noise when the parameters $r_j$ are constant.

consecutive components of the original time series among which any two components are close to each other (measured by the tolerance $r$). Thus the distance between any two components in the coarse-grained time series is potentially bigger than that in the original time series. In order to further construct the coarse time series, a bigger tolerance in the next coarse-grained procedure is desired. We present the numerical results of APCFME applied to Gaussian white noise when the parameters $r_j$ are chosen as a constant ($r_j = 0.15$) at all scales in Fig. 2.8. It shows that the entropy value of the coarse-grained time series remains a constant after scale 3, which indicates that the construction of the coarse-grained time series fails when the scale is bigger than 3.

According to the discussion above, we choose parameters $r_0 = 0.15, r_{j+1} = 1.1 \times r_j$ for $0 \leq j \leq 5$ and $r_{j+1} = 1.05 \times r_j$ for $j > 5$ in the numerical results shown in the remaining part of this section. We present in Fig. 2.9(a) results of APCFME applied to Gaussian white noise and $1/f$ noise. The results are similar to those in Fig. 2.1.




(a) (b)

Figure 2.9: APCFME results for simulated Gaussian white noise, $1/f$ noise and HHITS.

In Fig. 2.9(b), we present numerical results of APCFME applied to HHITS of length $N = 3 \times 10^4$ for the CHF, YOUNG and OLD groups. We find that YOUNG is most complex when the scale is bigger than 3. Moreover, CHF is more complex than OLD, which is consistent with the results of PLFME (Fig. 2.5). This example shows that APCFME without using any prior information is comparable to PLFME whose construction uses prior information of HHITS.

# Chapter 3

# Wavelet Packet Transform Entropy Analysis

Both MSE analysis and PLFME analysis discussed in previous chapter measure the complexity of a time series at different scales. When we apply MSE or PLFME method to the human heartbeat interval time series, we measure the complexity of the time series by looking at entropy values at different scales, which tell us which cardiac system is more complex than others at certain scales. Entropy values at different scales can also be used as extracted features to classify different cardiac systems. Fish discriminant analysis and support vector machine methods were applied to the classification of heartbeat interval time series using entropy values from MSE as features in [18] and [46] respectively.

In terms of classification, we want to have as many distinct features as we can

on different systems. Except the lower frequency components extracted by MSE or PLFME, higher frequency components of the time series may also provide useful information. The hierarchical entropy (HE) analysis introduced in [38] extracted both lower and higher frequency components of the time series by two different operators. Wavelet packet transform [17] provides us with a systematic way to decompose the original time series into lower and higher frequency components. Following the idea of HE, we propose wavelet packet transform entropy (WPTE) analysis in this chapter. In wavelet packet transform entropy analysis we decompose the original time series using wavelet packet transform at different scales and compute the (blockwise) sample entropy of each filtered time series.

We still consider human heartbeat interval time series in this chapter as our main study case. We have seen in the last chapter that human heartbeat interval time series may be well represented by a piecewise linear curve. Thus in this application, we will use lower and higher piecewise linear filters in the WPTE analysis. Piecewise polynomial wavelets filters has been applied to image denoising problems in [43]. Details constructing lower and higher piecewise polynomial filters will also be given in this chapter. Using entropy values as extracted features and following from multi-category classification support vector machine (SVM), we develop several classifiers for human heartbeat interval time series. Classification rates will be given and compared when we use entropy values from MSE, PLFME, HE and WPTE using piecewise linear filters. Among these classifiers, we will see that if we use entropy values from WPTE

using piecewise linear filters, we have the highest classification rates.

This chapter is organized into six sections. In section 2.1, we review wavelet packet transform. Section 2.2 presents the piecewise polynomial wavelet filters following from wavelets on invariant sets. Section 2.3 is devoted to describing the wavelet packet transform entropy analysis. In section 2.4, we apply wavelet packet transform entropy analysis with piecewise linear wavelet filters to HHITS. For comparison reason, theoretical results for Gaussian white noise and numerical results for both Gaussian white noise and $1/f$ noise are also given in this section. Multi-category classification SVM is reviewed in section 2.5 and at the end, classification results for human heartbeat interval time series are reported in section 2.6.

## 3.1 Wavelet Packet Transform

In this section, we first set up the general form of wavelets following from the multiresolution analysis [45]. A multiresolution analysis of the space $\mathbf{L}^2(\mathbb{R})$ consists of a sequence of nested subspaces

$$\{0\} \cdots \subset \mathbf{F}_0 \subset \mathbf{F}_1 \subset \cdots \subset \mathbf{F}_N \subset \mathbf{F}_{N+1} \subset \cdots \subset \mathbf{L}^2(\mathbb{R}). \tag{3.1}$$

Let $\mathbf{W}_k$ be the orthogonal complement of $\mathbf{F}_k$ in $\mathbf{F}_{k+1}$; symbolically we write

$$\mathbf{F}_{k+1} = \mathbf{F}_k \oplus^{\perp} \mathbf{W}_k, \quad k = 0, 1, \ldots.$$

The spaces $\mathbf{W}_k$, $k = 0, 1, \ldots$ are called the wavelets spaces. Due to the relation of spaces $\mathbf{F}_k$, $k = 0, 1, \ldots$in (3.1), we have that

$$\mathbf{L}^2(\mathbb{R}) = \mathbf{F}_0 \oplus^{\perp} \mathbf{W}_0 \oplus^{\perp} \mathbf{W}_1 \oplus^{\perp} \mathbf{W}_2 \cdots .$$

We denote the set of the orthonormal basis functions in $\mathbf{F}_k$ by $\mathbf{f}_k$ and the set of the orthonormal basis functions in $\mathbf{W}_k$ by $\mathbf{w}_k$ for $k = 0, 1, \ldots$. Since $\mathbf{F}_0 \subset \mathbf{F}_1$, each element in $\mathbf{f}_0$ can be written as a linear combination of the elements in $\mathbf{f}_1$. Therefore, we have the matrix representation

$$\mathbf{f}_0 = A\mathbf{f}_1 \tag{3.2}$$

for some $m \times n$ matrix $A$ where $m$ is the number of elements in $\mathbf{f}_0$ and $n$ is the number of elements in $\mathbf{f}_1$. Similarly, we have

$$\mathbf{w}_0 = B\mathbf{f}_1 \tag{3.3}$$

for some $m \times p$ matrix where $p$ is the number of elements in $\mathbf{w}_0$. Since $\mathbf{F}_1 = \mathbf{F}_0 \oplus \mathbf{W}_0$, we have that $p + n = m$. The matrix $A$ in (3.2) is called high-pass filter and the matrix $B$ in (3.3) is called low-pass filter. From the orthonormality of elements in $\mathbf{f}_0, \mathbf{w}_0$ and $\mathbf{f}_1$, it is easy to check that rows of the matrix $A$ are orthonormal, and so are the rows of the matrix $B$.

In the discipline of digital signal precessing , the sampled data set is passed through the low-pass filter and the high-pass filter. In other words, for a given signal $\mathbf{x}$ of length $N \in \mathbb{N}$ with $N = m \times q$ for some integer $q$, we compute

$$\mathbf{x}_A := (I_q \otimes A)\mathbf{x} \tag{3.4}$$

and

$$\mathbf{x}_B := (I_q \otimes B)\mathbf{x} \tag{3.5}$$

respectively. The high-passed filtered data set $\mathbf{x}_B$ is the wavelet transform detail coefficients and the low-pass filtered data set $\mathbf{x}_A$ is the wavelet transform approximation coefficients. Of note, the original signal $\mathbf{x}$ can be fully recovered from $\mathbf{x}_A$ and $\mathbf{x}_B$. We state this result in the following Proposition.

**Proposition 3.1.** *Let matrices $A$ and $B$ be defined in (3.2) and (3.3). Given signal $\mathbf{x}$, if $\mathbf{x}_A$ and $\mathbf{x}_B$ are defined as in (3.4) and (3.5), then*

$$\mathbf{x} = [A^T \quad B^T] \begin{bmatrix} \mathbf{x}_A \\ \\ \mathbf{x}_B \end{bmatrix}. \tag{3.6}$$

*Proof.* From (3.2) and (3.3), we have that

$$\begin{pmatrix} \mathbf{F}_0 \\ \\ \mathbf{W}_0 \end{pmatrix} = \begin{pmatrix} A \\ \\ B \end{pmatrix} \mathbf{F}_1.$$

From the orthonormality of elements in $\mathbf{F}_0$, $\mathbf{W}_0$ and $\mathbf{F}_1$, we know that the matrix $[A \quad B]^T$ is orthonormal. From (3.4) and (3.5), we have that

$$\begin{pmatrix} \mathbf{x}_A \\ \\ \mathbf{x}_B \end{pmatrix} = \begin{pmatrix} A \\ \\ B \end{pmatrix} \mathbf{x}.$$

Therefore, we obtain (3.6) from last equation. □

Both of the approximation and detail coefficients can be used as the sampled data input for another pair of wavelet filters, identical to the first pair, generating another

set of detail and approximation coefficients at the next lower level of scale. This process can be continued until the limit of the unit interval is reached. This process is called the wavelet packet transform, originally proposed in [17].

We will focus on the filters derived from wavelets on invariant sets in this chapter, especially the one dimension case.

## 3.2 Piecewise Polynomial Wavelets Filters on Invariant Sets

In this section, we present the piecewise polynomial wavelets filters on invariant sets in $\mathbb{R}^d$, initially introduced in [49]. We start from the matrix representation for construction of wavelets on invariant sets introduced in [49, 50]. For a given positive integer $m$, let

$$\phi_i : \mathbb{R}^d \to \mathbb{R}^d, \quad i \in \mathbb{Z}_m,$$

be $m$ contractive affine maps. Suppose the compact set $E \subset \mathbb{R}^d$ satisfies

$$E = \bigcup_{i=0}^{m-1} \phi_i(E)$$

and

$$m\big(\phi_i(E) \bigcap \phi_j(E)\big) = 0, \quad i \neq j,$$

where $m(E)$ denotes the measure of the set $E$. Then the set $E$ is called the invariant set with respect to $\phi_i$, $i \in \mathbb{Z}_m$. For some orthogonal matrix $Q \in \mathbb{R}^{m\times m}$, we define a

set of linear operator $T_i : \mathbf{L}^2(E) \to \mathbf{L}^2(E)$ , $i \in \mathbb{Z}_m$ by

$$(T_i g)(t) := \sum_{j=0}^{m-1} q_{ij} g(\phi_j^{-1}(t)) \chi_{E_i}(t), \quad t \in E, \tag{3.7}$$

where $\chi$ is the characteristic function. Furthermore, the adjoint operator $T_i^*$ can be verified as

$$T_i^* = \frac{1}{m} \sum_{\ell=0}^{m-1} q_{\ell i} G_\ell, \tag{3.8}$$

where the matrix $G_i$, $i \in \mathbb{Z}_m$, satisfies

$$G_i f := f \circ \phi_i, \quad i \in \mathbb{Z}_m, \quad f \in \mathbf{L}^2(E). \tag{3.9}$$

The vector valued function $\mathbf{f} := [f_0, \ldots, f_{n-1}]$ is called a refinable curve if it satisfies a refinement equation

$$G_i \mathbf{f} = A_i^T \mathbf{f}, \quad i \in \mathbb{Z}_m, \tag{3.10}$$

for some given matrices $A_i$. Then we can define a family of subspaces $\mathbf{F}_k$, $k = 0, 1, \ldots$ in $\mathbf{L}^2(E)$ recursively by

$$\mathbf{F}_{k+1} = \bigoplus_{i=0}^{m-1} T_i \mathbf{F}_k, \quad k = 0, 1, \ldots. \tag{3.11}$$

It was shown in [49] that

$$\mathbf{F}_k \subset \mathbf{F}_{k+1}, \quad k = 0, 1 \ldots$$

and

$$\overline{\bigcup_{k=0}^{\infty}} \mathbf{F}_k = \mathbf{L}^2(E).$$

Therefore, the sequence of subspaces $\{\mathbf{F}_n, n \in \mathbb{Z}_+\}$ forms a multiresolution analysis of $\mathbf{L}^2(E)$ corresponding to (3.1). Let $\mathbf{W}_{k-1}$ be an orthogonal complement of $\mathbf{F}_{k-1}$ in

$\mathbf{F}_k$, i.e.,

$$\mathbf{F}_k = \mathbf{F}_{k-1} \oplus \mathbf{W}_{k-1}.$$

Then we have

$$\mathbf{L}^2(E) = \mathbf{F}_0 \oplus^{\perp}_{k\in\mathbb{Z}_+} \mathbf{W}_{k-1},$$

which gives a multiscale decomposition for the space $\mathbf{L}^2(E)$. The subspaces $\mathbf{W}_k$ are the wavelet spaces for $\mathbf{L}^2(E)$ and in particular, the space $\mathbf{W}_0$ is called the initial wavelet space. The general construction of the initial wavelet space $\mathbf{W}_0$ was formulated in [50] in terms of a general solution of a matrix completion problem and a particular solution was given there. As usual, we denote by $\dim X$ the dimension of a finite dimensional space $X$. Since we have that

$$\mathbf{W}_{k+1} = \bigoplus_{i=0}^{m-1} T_i \mathbf{W}_k,$$

it is easy to check that $\dim \mathbf{W}_{k-1} = n(m-1)m^{k-1}$ and $\dim \mathbf{F}_{k-1} = pm^{k-1}$.

We can see that the contractive maps $\phi_i$, the scaling operators $T_i$ and the orthogonal matrix $Q$ are the three main ingredients in the abstract construction of wavelets on invariant sets. Now we consider a special case when $Q = \sqrt{m} I_m$, which will generate a set of orthonormal basis in the wavelet spaces. This result is stated in Proposition 3.3. In this case, the operator $T_i$ defined in (3.7) will be

$$(T_i g)(t) := \sqrt{m} \sum_{j=0}^{m-1} g(\phi_j^{-1}(t)) \chi_{E_i}(t), \quad t \in E. \tag{3.12}$$

In the following context, we will use $\delta_{i,j}$ to denote the Kronecker delta function,

that is,

$$\delta_{i,j} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

We also use $\mathcal{I}$ to denote the identity operator and $(f, g)$ to denote the inner product of $f$ and $g$ in $\mathbf{L}^2(E)$. In the following Proposition, we present the relation of operator $T_i^*$ and $T_j$ for any any $i, j \in \mathbb{Z}_m$.

**Proposition 3.2.** *If $Q = \sqrt{m} I_m$ in (3.7), then*

$$T_i^* T_j = \delta_{i,j} \mathcal{I} \tag{3.13}$$

*for any $i, j \in \mathbb{Z}_m$. Moreover, for any $f, g \in \mathbf{L}^2(E)$, it holds that*

$$(T_f, T_j g) = \delta_{i,j}(f, g). \tag{3.14}$$

*Proof.* For any $i, j \in \mathbb{Z}_m$ and $f \in \mathbf{L}^2(E)$, we know from (3.8) and (3.12) that

$$(T_i^* T_j f)(x) = \sqrt{m}(T_j)(\phi_i(x)).$$

Then it follows from the definition of $T_j$ that $T_i^* T_j = 0$ if $i \neq j$ and $T_i^* T_j = \mathcal{I}$ if $i = j$.

For any $f, g \in \mathbf{L}^2(E)$, it follows from 3.13 that

$$(T_f, T_j g) = (T_i^* T_j f, g) = \delta_{i,j}(f, g).$$

□

**Proposition 3.3.** *If $\mathbf{f}_0$ is as an orthonormal basis of $\mathbf{F}_0$, then the set $\mathbf{f}_{n+1}$ defined as*

$$\mathbf{f}_{n+1} := \bigcup_{i \in \mathbb{Z}_m} T_i \mathbf{f}_n, \quad n \in \mathbb{N}$$

*is an orthonormal basis of $\mathbf{F}_{n+1}$.*

*Proof.* We will prove this by induction on $n$. When $n = 0$, $\mathbf{f}_0$ is an orthonormal basis of $\mathbf{F}_0$ by hypothesis. Now we assume that $\mathbf{f}_j$ is an orthonormal basis of $\mathbf{F}_j$ for some $j \geq 1$, then $\mathbf{f}_{j+1}$ being an orthonormal basis for $\mathbf{F}_{j+1}$ follows from the way which the space $\mathbf{F}_{j+1}$ is constructed in (3.11), the orthonormal property of $\mathbf{f}_j$ and the equation (3.14). □

Now we are ready to present the piecewise polynomial wavelets filters based on the construction of wavelet on invariant sets, specifically using the matrix $Q = \sqrt{m}I_m$. Given a refinable vector field $\mathbf{f}$ satisfying (3.10) for some matrices $A_i$, we assume that elements in $\mathbf{f}$ are orthonormal. Let $\mathbf{f}_1 = \bigcup_{i\in\mathbb{Z}_m} T_i\mathbf{f}$. Then if follows from Proposition 3.3 that functions in $\mathbf{f}_1$ is an orthonormal basis of the space $\mathbf{F}_1$. From equations (3.7) and (3.9), we have that

$$\mathbf{f} = \tilde{A}\mathbf{f}_1, \tag{3.15}$$

where $\tilde{A} = \frac{1}{\sqrt{m}}[A_0^T A_1^T \cdots A_{m-1}^T]$. Let $\mathbf{w}$ be an orthonormal basis of the space $\mathbf{W}_0$. Since $\mathbf{W}_0 \subset \mathbf{F}_1$, there is an $mn \times (m-1)n$ matrix $\tilde{B}$ such that

$$\mathbf{f} = \tilde{B}\mathbf{f}_1. \tag{3.16}$$

The matrices $\tilde{A}$ and $\tilde{B}$ in (3.15) and (3.16) will be the low-pass and high-pass filters according to (3.2) and (3.3) introduced in the multiresulution analysis.

We will pay our special attention to the set $E := [0, 1]$. Note that $E$ is the invariant set with respect to the contractive affine maps $\phi_\alpha(t) = \frac{t+\alpha}{2}$, $\alpha \in \mathbb{Z}_2$. We also define the refinable vector field $\mathbf{f} := [f_j, j \in \mathbb{N}_n]$, where $f_j$ is the Legendre polynomial of

degree $j$ on $E$. We give several examples of piecewise polynomial wavelet filters of different orders.

- Piecewise Constant Wavelets Filters.

  We choose $\mathbf{f} = [f_0]$ where $f_0(t) := 1$, $t \in E$. The corresponding orthogonal matrix $Q$ is the $1 \times 1$ matrix $[\sqrt{2}]$. An orthonormal basis of $\mathbf{W}_0$ is given by

$$\mathbf{w}_0 = \begin{cases} 1, & \text{if } t \in [0, 1/2], \\ -1, & t \in (1/2, 1]. \end{cases}$$

  In this case, one can compute that the low-pass filter $\tilde{A}$ is

$$\tilde{A} := [\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}] \tag{3.17}$$

  and the high-pass filter $\tilde{B}$ is

$$\tilde{B} := [\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}]. \tag{3.18}$$

  We can see that they are actually the Haar Wavelets filters.

- Piecewise Linear Wavelets Filters.

  We choose $\mathbf{f} = [f_0, f_1]$ where $f_0(t) := 1$, and $f_1(t) := \sqrt{3}(2t-1)$, $t \in E$. Then the corresponding orthogonal matrix $Q$ is the $2 \times 2$ matrix $\sqrt{2}I_2$. An orthonormal basis of $\mathbf{W}_0$ is given by

$$\mathbf{w}_0 = \begin{cases} 1 - 6t, & t\text{if } \in [0, 1/2], \\ 5 - 6t, & \text{if } t \in (1/2, 1], \end{cases} \quad \text{and} \quad \mathbf{w}_1 = \begin{cases} \sqrt{3}(1 - 4t), & \text{if } t \in [0, 1/2], \\ \sqrt{3}(5 - 6t), & \text{if } t \in (1/2, 1]. \end{cases}$$

In this case, one can compute that the low-pass filter $\tilde{A}$ is given by

$$\tilde{A} := \frac{1}{2}\begin{bmatrix} 1 & 0 & 1 & 0 \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} & \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}, \tag{3.19}$$

and the high-pass filter $\tilde{B}$ is given by

$$\tilde{B} := \frac{1}{2}\begin{bmatrix} 0 & -1 & 0 & 1 \\ -\frac{1}{2} & -\frac{\sqrt{3}}{2} & \frac{1}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix}. \tag{3.20}$$

- Piecewise Quadratic Wavelets Filters.

  We choose $\mathbf{f} = [f_0, f_1, f_2]$ where $f_0(t) = 1$, $f_1(t) = \sqrt{3}(2t-1)$ and $f_3(t) = \sqrt{5}(6t^2 - 6t + 1)$ for any $t \in E$. The corresponding orthogonal matrix $Q$ is the $3 \times 3$ matrix $\sqrt{3}I_2$. And an orthonormal basis of $\mathbf{W}_0$ is given by

$$\mathbf{w}_0 = \begin{cases} 1 - 6t, & \text{if } t \in [0, 1/2], \\ 5 - 6t, & \text{if } t \in (1/2, 1], \end{cases} \quad \mathbf{w}_1 = \begin{cases} \frac{\sqrt{93}}{31}(240t^2 - 116t + 9), & \text{if } t \in [0, 1/2], \\ \frac{\sqrt{93}}{31}(3 - 2t), & \text{if } t \in (1/2, 1], \end{cases}$$

$$\text{and } \mathbf{w}_2 = \begin{cases} \frac{\sqrt{93}}{31}(4t - 1), & \text{if } t \in [0, 1/2], \\ \frac{\sqrt{93}}{31}(240t^2 - 364t + 133), & \text{if } t \in (1/2, 1]. \end{cases}$$

  In this case, one can compute that the low-pass filter $\tilde{A}$ is given by

$$\tilde{A} := \frac{1}{2}\begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} & 0 & \frac{\sqrt{3}}{2} & \frac{1}{2} & 0 \\ 0 & -\frac{\sqrt{15}}{4} & \frac{1}{4} & 0 & \frac{\sqrt{15}}{4} & \frac{1}{4} \end{bmatrix}, \tag{3.21}$$

and the high-pass filter $\tilde{B}$ is given by

$$\tilde{B} := \frac{1}{2}\begin{bmatrix} 0 & -1 & 0 & 0 & 1 & 0 \\ -\frac{1}{2} & -\frac{\sqrt{3}}{2} & 0 & \frac{1}{2} & -\frac{\sqrt{3}}{2} & 0 \\ 0 & -\frac{1}{4} & -\frac{\sqrt{15}}{4} & 0 & \frac{1}{4} & \frac{\sqrt{15}}{4} \end{bmatrix}. \tag{3.22}$$

## 3.3 Wavelet Packet Transform Entropy Analysis

In this section, we introduce the wavelet packet transform entropy analysis. Wavelet packet transform extracts both lower and higher frequency components of the original signal at different scales. The original signal can be fully recovered from wavelet packet transform at each scale. Thus we can examine the complexity of both lower and higher frequency components of the signal at different scales if we apply (blockwise) sample entropy to the filtered signals obtained from wavelet packet transform.

We first present a hierarchical decomposition of the signal based on wavelet packet transform. This idea was initially introduced in [38] based on lower and higher Haar wavelet filters. Given a one-dimensional time series, $\mathbf{x} := \{x_0, x_1, \ldots, x_{N-1}\}$, positive integer $m$ and positive number $r$, let $S_m(\mathbf{x}, r)$ be the sample entropy of $\mathbf{x}$ defined in (2.1). Let $A$ and $B$ be a low-pass and high-pass filter, respectively. We assume that the size of $A$ and $B$ are the same. We see from last section that all piecewise polynomial wavelet filters satisfy this assumption. We further suppose that $A, B \in \mathbb{R}^{p\times q}$ and $N = np$, $n \in \mathbb{N}$, for simplicity. We define two operators $Q_0$ and $Q_1$ on $\mathbf{x}$ where

$$Q_0(\mathbf{x}) := (A \otimes I)\mathbf{x} \tag{3.23}$$

and

$$Q_1(\mathbf{x}) := (B \otimes I)\mathbf{x}. \tag{3.24}$$

The size of the identity matrix $I$ is chosen based on the integer $p$ and the length of $\mathbf{x}$. On the one hand, for any $n \in \mathbb{Z}_{\mathbb{N}}^+$ and $[\ell_1, \ell_2, \ldots, \ell_n] \in \{0,1\}^n$, the integer $e$ defined by

$$e := \sum_{j=1}^{n} \ell_j 2^{n-1} \tag{3.25}$$

is nonnegative. On the other hand, when the integer $n$ is fixed, given a nonnegative integer $e$, there is a unique vector $[\ell_1, \ell_2, \ldots, \ell_n] \in \{0,1\}^n$ corresponding to $e$ through equation (3.25). Thus for any $n \in \mathbb{Z}_n^+$ and nonnegative integer $e$, we define the hieratical components of the time series $\mathbf{x}$ by

$$\mathbf{x}_{n,e} := Q_{\ell_n} \circ Q_{\ell_{n-1}} \circ \cdots \circ Q_{\ell_1}(\mathbf{x}). \tag{3.26}$$

For any $k \in \mathbb{Z}_{N+1}^+$, we call the signal $\mathbf{x}_{k,e}$, $e \in \mathbb{Z}_{2^k}$, the wavelet packet transform of $\mathbf{x}$ at the level $k$. From Proposition 3.1 and the way of constructing $\mathbf{x}_{n,e}$, we know that given any level $n \in \mathbb{Z}_k$, the decomposition $\mathbf{x}_{k,e}$, $e \in \mathbb{Z}_{2^k}$, gives us a full description for the original signal $\mathbf{x}$. For any $k \in \mathbb{Z}_{N+1}^+$, we define an index set $\mathbb{J}_k := \{(n,e) : n \in \mathbb{Z}_k, e \in \mathbb{Z}_{2^n}\}$. Let $\mathbf{x}_{0,0} := \mathbf{x}$. Then the signals $\mathbf{x}_{n,e}$, $(n,e) \in \mathbb{J}_k$, are called the hierarchical decomposition of the signal $\mathbf{x}$ up to $k$ levels.

With a hierarchical decomposition $\mathbf{x}_{n,e}$, $(n,e) \in \mathbb{J}_k$, in place, we compute the sample entropy of the original signal $\mathbf{x}_{0,0}$, $S_m(\mathbf{x}_{0,0}, r)$, and the blockwised sample entropy of the filtered signal $\mathbf{x}_{n,e}$, $\tilde{S}_m(\mathbf{x}_{n,e}, r)$ defined in (2.4), when $n \geq 1$ to measure

the complexity of the biological system which has the time series $\mathbf{x}$ as its output variables. We call this process the wavelet pocket transform entropy analysis. When the filters $A$ and $B$ are chosen as (3.17) and (3.18), it is the hierarchical entropy analysis introduced in [38]. Thus hierarchical entropy is a special case of the wavelet packet transform entropy analysis.

## 3.4 Application to Human Heartbeat Interval Time Series

In this section, we apply the wavelet packet transform entropy analysis to human heartbeat interval time series, the same data set we used in Section 1.3. We have seen from Section 1.3 that piecewise linear wavelets give a good description for the heartbeat time series. Thus we will choose the lower and higher piecewise linear wavelets filters in WPTE in this application.

For comparison reason, we present the results of sample wavelet packet transform entropy analysis with lower and higher piecewise linear wavelet filters for the Gaussian white noise and $1/f$ noise first. We present a theoretical result for the Gaussian white noise and numerical results for the Gaussian white noise and $1/f$ noise. In this section, the matrices $A$ and $B$ refer to the lower frequency and higher frequency piecewise linear wavelet filters defined in (3.19) and (3.22) respectively. Accordingly, operators $Q_0$ and $Q_1$ in (3.23) and (3.24) are defined from lower frequency and higher

frequency piecewise linear wavelet filters, which are further used to define the hierarchical components of the time series $\mathbf{x}$, $\mathbf{x}_{n,e}$, for any $n \in \mathbb{Z}_n^+$ and nonnegative integer $e$ via (3.26).

For the Gaussian white noise, any filtered time series in the hierarchical decomposition is still a Gaussian white noise. Specifically, we have the following result.

**Lemma 3.4.** *If for a positive integer $N$, $\mathbf{g} := [g_j : j \in \mathbb{Z}_{2^N}]$ denotes a real Gaussian random vector with mean 0 and standard deviation $\delta$, then for each $(n, e) \in \mathbb{J}_{N+1}$, $\mathbf{g}_{n,e}$ is a real Gaussian random vectors with mean 0 and standard deviation $\delta/2^{\frac{n}{2}}$.*

*Proof.* Fox a fixed positive integer $N$, we prove this result by induction on $n$ where $1 \leq n \leq N$.

Suppose that $n = 1$. Since $A^T A = \frac{1}{2} I$ and $B^T B = \frac{1}{2} I$, it was shown in Lemma 2.5 and Lemma 2.7 that $\mathbf{g}_{1,0}$ is a real Gaussian random vectors with mean 0 and standard deviation $\delta/2^{\frac{1}{2}}$. Thus the statement holds for $n = 1$.

Now suppose that the statement holds for some integer $k$, $1 \leq k < N$. Thus for each $(k, e), e \in \mathbb{Z}_{2^k}$, $\mathbf{g}_{k,e}$ is a real Gaussian random vectors with mean 0 and standard deviation $\delta/2^{\frac{k}{2}}$. Since $\mathbf{g}_{k+1,e'} = Q_i(\mathbf{g}_{k,e})$, for some $e \in \mathbb{Z}_{2^k}$ and $i \in \mathbb{Z}_2$, applying Lemma 2.5 and Lemma 2.7 to $\mathbf{g}_{k,e}$ with matrix A (when $i = 0$) and matrix B (when $i = 1$) concludes that $\mathbf{g}_{k+1,e'}$ is a real Gaussian random vectors with mean 0 and standard deviation $\delta/2^{\frac{k+1}{2}}$ for any $(k + 1, e'), e' \in \mathbb{Z}_{2^{k+1}}$. Therefore we finish the induction and the proof. □

Since each $\mathbf{g}_{n,e}$ in the hierarchical decomposition is a real Gaussian random vector,

it enables us to establish a relation of the (blockwise) sample entropy of each filtered time series. A result similar to Theorem 3.2 in [38] can be obtained.

**Theorem 3.5.** *Suppose that* $\mathbf{g} := [g_j : j \in \mathbb{Z}_{2^N}]$ *with* $N \in \mathbb{N}$ *is a real Gaussian random vector with mean 0 and standard deviation* $\delta$ *and* $m \in \mathbb{Z}^+_{2^N-1}$. *Then the following statements hold:*

1. *For all* $r > 0$ *and* $(n, e) \in \mathbb{J}_{N+1}$,

$$\tilde{S}_m(\mathbf{g}_{n,e}, r) = -\ln\Big(\frac{1}{2\delta\sqrt{2\pi}}\int_{\mathbb{R}} D(y)e^{-y^2/\delta^2}\Big)^2, \tag{3.27}$$

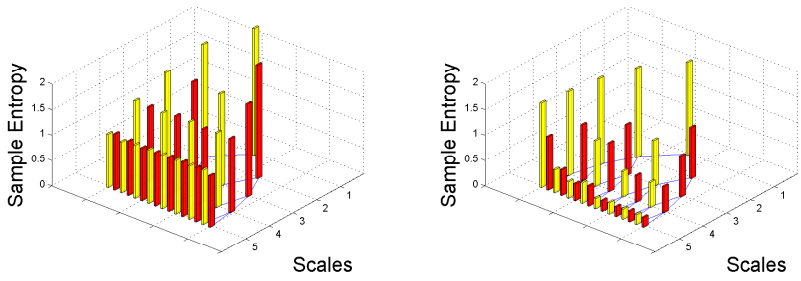
   *where* $D(y) = erf(\frac{y+2^{n/2}r}{\delta}) - erf(\frac{y-2^{n/2}r}{\delta})$.

2. *For fixed* $r$ *and* $n$, $\tilde{S}_m(\mathbf{g}_{n,e}, r) = \tilde{S}_m(\mathbf{g}_{n,e'}, r)$, *for all* $e, e' \in \mathbb{Z}_{2^n}$.

3. *For fixed* $r$, *if* $n > n'$, *then* $\tilde{S}_m(\mathbf{g}_{n,e}, r) < \tilde{S}_m(\mathbf{g}_{n',e'}, r)$ *for all* $e \in \mathbb{Z}_{2^n}$ *and* $e' \in \mathbb{Z}^{2^{n'}}$.

*Proof.* It follows from Lemma 3.4 that for each $(n, e) \in \mathbb{J}_{N+1}$, $\mathbf{g}_{n,e}$ is a real Gaussian random vectors with mean 0 and standard deviation $\delta/2^{\frac{n}{2}}$. The first two conclusions can be obtained following from the same argument used in Theorem 2.8. The last conclusion can be immediately drawn from Theorem 2.8 and the second conclusion. □

In Fig. 3.1, we present the numerical results when we apply WPTE using lower and higher piecewise linear wavelet filters to Gaussian white noise and $1/f$ noise. From Fig. 3.1(a), we can see that for a fixed scale factor $n$, the (blockwise) sample entropy



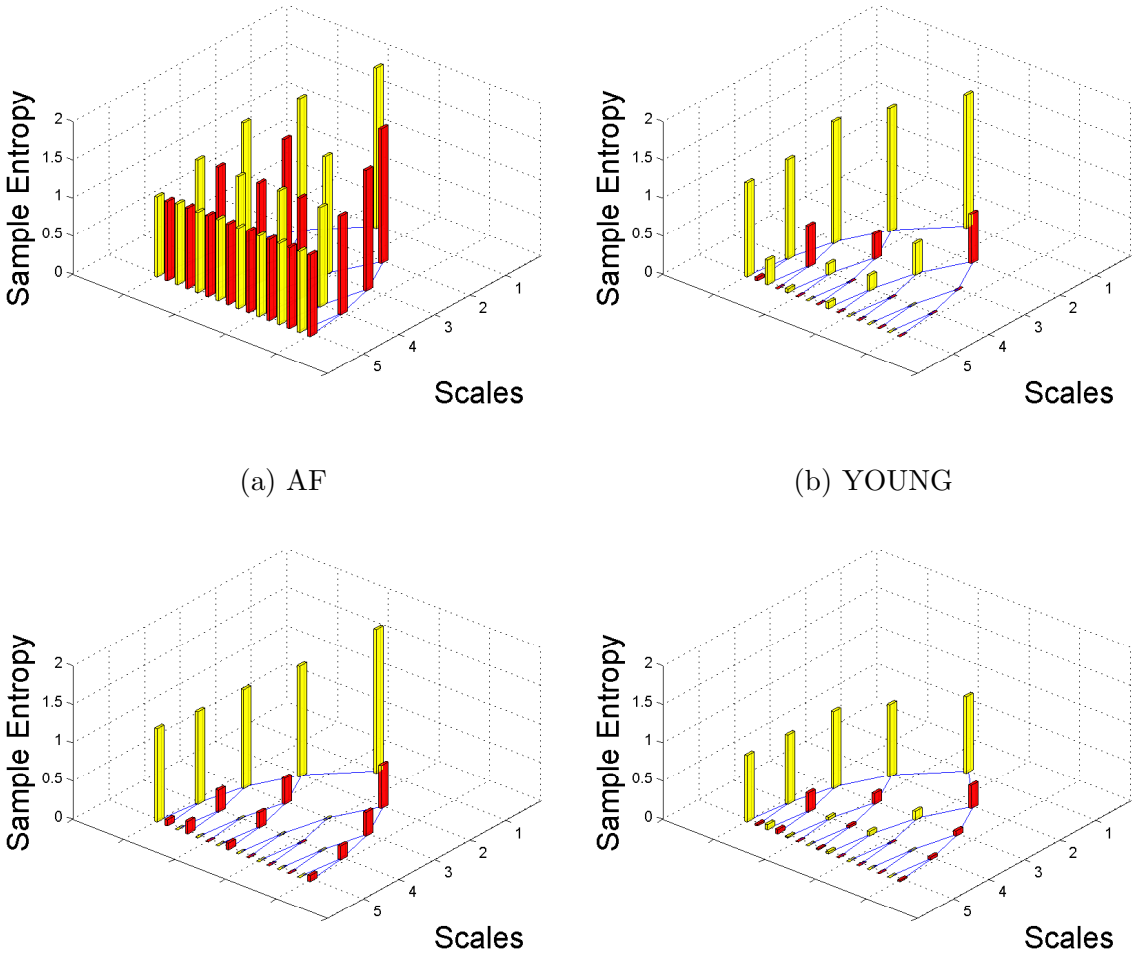

(a) Gaussian White Noise (b)$1/f$ Noise

Figure 3.1: WPTE analysis with lower and higher piecewise linear wavelet filters results for simulated Gaussian white noise and $1/f$ noise.

of $\tilde{S}_{n,e}, e \in \mathbb{Z}_{2^n}$, is a constant with respect to $e$, and the value of $\tilde{S}_{n,e}$ are decline as the scale factor $n$ increases. From Fig. 3.1(b), we see that the higher frequency components of $1/r$ noise presents the similar pattern to the Gaussian white noise in terms of blockwise sample entropy.

Now we show numerical results when we apply the wavelet pocket transform entropy analysis with lower and higher piecewise linear wavelet filters on human heartbeat interval time series. Results are shown in Fig. 3.2. For comparison reasons, we also give the numerical results that we apply hierarchical entropy analysis on the heartbeat time series [38]. The results are shown in Fig. 3.3.

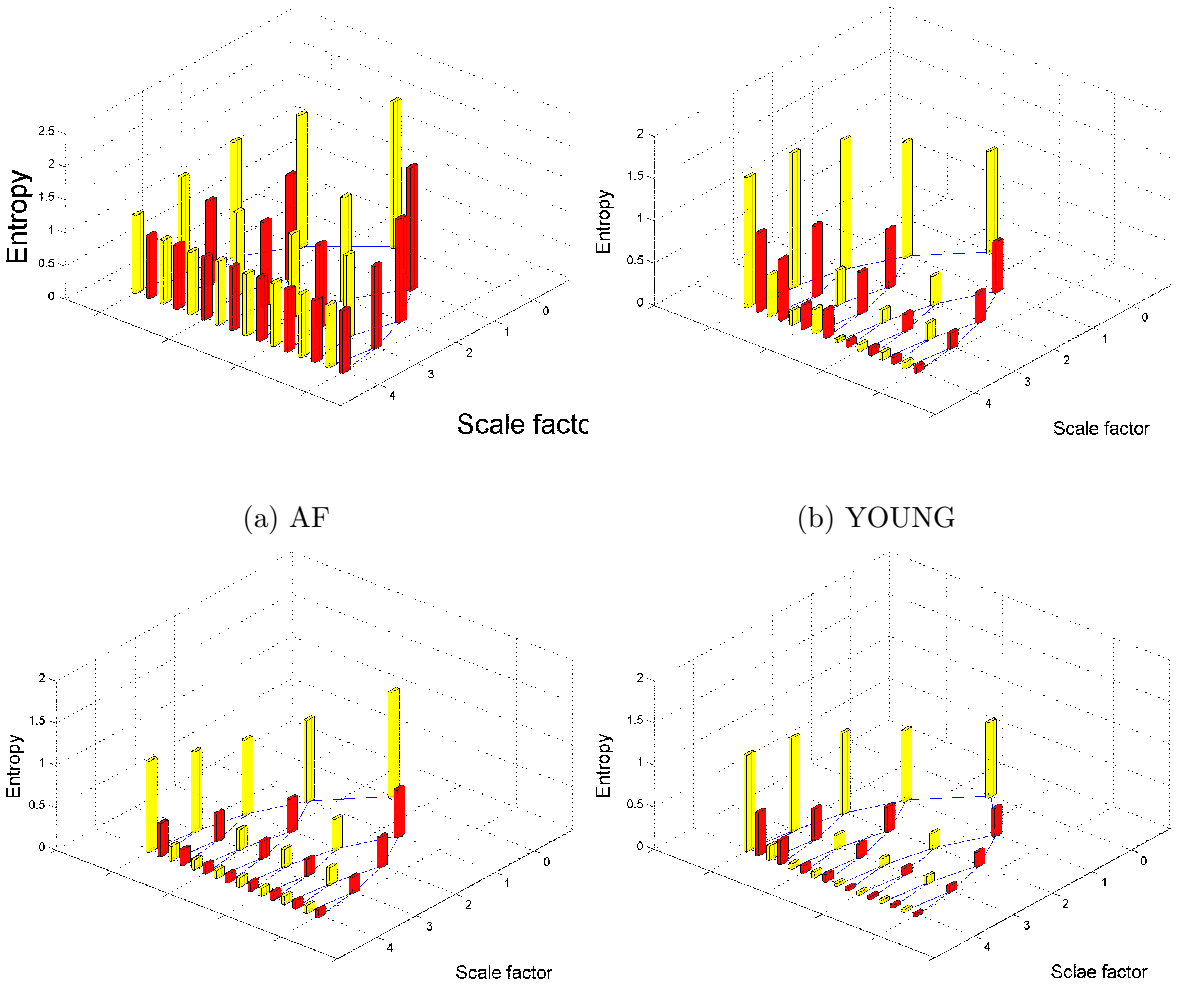
From Fig. 3.2 and 3.3, we can see that higher frequency components do provide useful information for different classes of time series. For groups CHF and OLD, which show similar patterns in both methods, piecewise linear wavelets filters give a better discrimination than piecewise constant wavelets. We will give a more a accurate result




(a) AF (b) YOUNG

(c) CHF (d) OLD

Figure 3.2: The wavelet packet transform entropy analysis with lower and higher piecewise linear filter results for human cardiac interbeat interval time series with data of length $N = 8 \times 10^4$.




(a) AF (b) YOUNG

(c) CHF (d) OLD

Figure 3.3: The hierarchical entropy analysis results for human heartbeat interval time series with data of length $N = 8 \times 10^4$.

for the classification in next section. Moreover, we can see that most of entropy values of the higher frequency components in the wavelet packet transform with piecewise linear wavelets filters are relatively small compared to those from hierarchical entropy entropy analysis, which also indicates that piecewise linear wavelets have a better approximation for the heartbeat interval time series than piecewise constant wavelets.

## 3.5 Multi-category classification SVM

In this section, we review multi-category classification. Learning can be thought of as inferring regularities from a set of training examples. There are various learning algorithms which allow the extraction of these underlying regularities, which will usually be represented in the values of some parameters of a statistic. Traditional neural network approaches for learning have suffered difficulties with generalization, producing models that can over-fit the data. Support Vector Machine (SVM) related to the statistical learning theory [48] was first introduced in [13]. SVM employs the structural risk minimization principle, which has great ability to generalize. SVM was developed firstly to solve the classification problem, but it is also applied to the domain of regression problems. It becomes popular because of its success in many applications, such as handwriting recognition [7], image clustering [44], text categorization [40], gene classification [34], protein structure prediction [33], etc.

The mathematical model of SVM is to parameterize training examples into vectors and then construct hyperplanes to separate different categories of vectors. This

is formulated and solved as an optimization problem. The solution to the related optimization problem is described as hyperplanes with maximal margins which are supported by some vectors among all vectors derived from training examples. When training vectors can not be separated by hyperplanes sometimes, their images through a feature map may separated by linear classifiers. And those vectors are separated by hyper-surfaces instead of hyperplanes. Such feature maps are related to reproducing kernels [3]. By choosing different reproducing kernels, SVM allows the construction of various hyper-surfaces. It is called a kernel trick.

We consider the following model of classification problem. There are $k$ classes of vectors in $\mathbb{R}^d$ for some integer $d > 0$, denoted by $\{\mathcal{C}_i : i \in \mathbb{Z}_k^+\}$. We want to find some functions to separate those $k$ classes from each other. A natural thought of multi-category classification based on SVM can be described as the model one-versus-rest: for a given class $\{\mathcal{C}_i\}$ for some $i \in \mathbb{Z}_k^+$, we construct a hyperplane to separate $\{\mathcal{C}_i\}$ from other $k-1$ classes. We obtain $k$ such hyperplanes. This strategy has been widely used to handle the multi-category classification problem. Another model of SVM is one-versus-one: for each pair $\{i, j\}$ such that $1 \leq i \leq j \leq k$, we construct a hyperplane to separate $\{\mathcal{C}_i\}$ and $\{\mathcal{C}_j\}$ . Totally there are $\frac{k(k-1)}{2}$ hyperplanes, see [27]. In this section we employ the one-versus-one model for two reasons: (i) If we obtain the best separation for each pair of classes locally, then we believe that we get the best separation globally; (ii) It reduces the scale of solving the related optimization problem. Each time we only consider vectors of two classes instead of vectors from

all categories.

Suppose that the training set consists of $N$ vectors $\{x_n : x \in \mathbb{Z}_N+\} \subset \mathbb{R}^d$ and these vectors belong to $d$ different classes $\{\mathcal{C}_i : i \in \mathbb{Z}_k^+\}$. We first assume that all these classes of vectors can be separated by hyperplanes. We now consider the classification problem for two classes. Given two classes $\{\mathcal{C}_i\}$ and $\{\mathcal{C}_j\}$ for some $1 \leq i \leq j \leq k$. We suppose that there exists a $w_{ij} \in \mathbb{R}^d$ and a constant $b_{ij} \in \mathbb{R}$ such that $w_{ij}^T x + b_{ij} = 0$, which is a hyperplane separating $\{\mathcal{C}_i\}$ and $\{\mathcal{C}_j\}$. We can adjust $w_{ij}$ and $b_{ij}$ so that $w_{ij}^T x + b_{ij} \geq 1$ if $x \in \mathcal{C}_i$ and $w_{ij}^T x + b_{ij} \leq 1$ if $x \in \mathcal{C}_j$. The distance between two hyperplanes $w_{ij}^T x + b_{ij} = 1$ and $w_{ij}^T x + b_{ij} = -1$ are called the directional margin between $\mathcal{C}_i$ and $\mathcal{C}_j$. In order to obtain an optimal separation, we find the maximal directional margin, which is formulated as an optimization problem as follows: Finding a $w_{ij} \in \mathbb{R}^d$ and a constant $b_{ij} \in \mathbb{R}$ to

$$\begin{aligned} &\text{maximize} && \tfrac{1}{\|w_{ij}\|} \\ &\text{subject to} && w_{ij}^T x + b_{ij} \geq 1 \quad \text{if } x_n \in \mathcal{C}_i \text{ for any } n \in \mathbb{Z}_N^+, \\ & && w_{ij}^T x + b_{ij} \leq -1 \quad \text{if } x_n \in \mathcal{C}_j \text{ for any } n \in \mathbb{Z}_N^+, \end{aligned}$$

where $\|w_{ij}\|$ is the Euclidian norm of the vector $w_{ij}$.

Now we consider all classes at one time. We want to obtain the maximal total of all directional margins, which is called the total margin. We get the following optimization problems: Finding $\{w_{ij} : 1 \leq i \leq j \leq k\} \subset \mathbb{R}^d$ and $\{b_{ij} : 1 \leq i \leq j \leq$

$k\} \subset \mathbb{R}$ to

$$\begin{aligned} \text{maximize} \quad & \textstyle\sum_{1\le i\le j\le k} \frac{1}{\|w_{ij}\|} \\ \text{subject to} \quad & w_{ij}^T x + b_{ij} \ge 1 \quad \text{if } x_n \in \mathcal{C}_i \text{ for any } n \in \mathbb{Z}_N^+, \\ & w_{ij}^T x + b_{ij} \le -1 \quad \text{if } x_n \in \mathcal{C}_j \text{ for any } n \in \mathbb{Z}_N^+, \end{aligned} \tag{3.28}$$

For any $1 \le i \le j \le k$ and $n \in \mathbb{Z}_N^+$, let

$$\delta_{i,j,n} = \begin{cases} -1 & \text{if } x_n \in \mathcal{C}_i \\ 1 & \text{if } x_n \in \mathcal{C}_j \\ 0 & \text{otherwise} \end{cases} \tag{3.29}$$

Applying Lagrange multiplier method, we obtain the Lagrangian

$$\mathcal{L} := \sum_{1\le i\le j\le k} \frac{1}{\|w_{ij}\|} + \sum_{1\le i\le j\le k} \sum_{n\in\mathbb{Z}_n^+} \lambda_{i,j,n}(\delta_{i,j,n}(w_{ij}^T x + b_{ij}) + 1), \tag{3.30}$$

where additional variables $\{\lambda_{i,j,n} : 1 \le i \le j \le k, n \in \mathbb{Z}_N^+\}$ take nonnegative values. Taking gradient of $\mathcal{L}$ and let it be the zero vector, we get

$$w_{ij} = \frac{y_{ij}}{\|y_{ij}\|^{3/2}}, \quad \text{where } y_{ij} := \sum_{n\in\mathbb{Z}_N^+} \lambda_{i,j,n}\delta_{i,j,n}x_n, \tag{3.31}$$

and

$$b_{ij} = -\delta_{i,j,n} - w_{ij}^T x_n, \quad \text{if} \lambda_{i,j,n} \ne 0 \text{ and } \delta_{i,j,n} \ne 0. \tag{3.32}$$

Each non-zero $\lambda_{i,j,n}$ indicates that the corresponding $x_n$ is a support vector in classing $\mathcal{C}_i$ and $\mathcal{C}_j$. Substituting (3.31) into (3.30), we derive the dual problem of (3.28):

Finding $\{\lambda_{i,j,n} : 1 \le i \le j \le k, n \in \mathbb{Z}_N^+\} \subset \mathbb{R}$ to

$$\begin{aligned} \text{minimize} \quad & 2\sum_{1\le i\le j\le k}\Big(\sum_{m,n\in\mathbb{Z}_N^+}\lambda_{i,j,n}\lambda_{i,j,m}\delta_{i,j,n}\delta_{i,j,m}x_n^Tx_m\Big)^{1/4} \\ & +\sum_{1\le i\le j\le k,n\in\mathbb{Z}_N^+}\lambda_{i,j,n} \qquad (3.33) \\ \text{subject to} \quad & \sum_{n\in\mathbb{Z}_N^+}\lambda_{i,j,n}\delta_{i,j,n}=0 \quad \text{for } 1\le i\le j\le k, \\ & \lambda_{i,j,n}\ge 0 \quad \text{for } 1\le i\le j\le k, n\in\mathbb{Z}_N^+. \end{aligned}$$

Suppose that the solutions to (3.33) are $\{\lambda^*_{i,j,n} : 1 \le i \le j \le k, n \in \mathbb{Z}_N^+\}$. Then the solutions $\{w^*_{ij} : 1 \le i \le j \le k\}$ to (3.28) can be given by (3.31) if we replace $\lambda_{i,j,n}$ by $\lambda^*_{i,j,n}$. And then $b^*_{ij}$ to (3.28) for any $1 \le i \le j \le k$ with $\lambda_{i,j,n} \ne 0$ and $\delta_{i,j,n} \ne 0$ can be given via (3.32). If $\lambda_{i,j,n} = 0$, $b^*_{ij}$ takes the value so that it gets the highest correction rate in classifying groups $\mathcal{C}_i$ and $\mathcal{C}_j$ by the hyperplane $w_{ij}^Tx + b_{ij} = 0$.

Note that to get the solution to (3.28) and (3.33), we assume that there exists such hyperplanes to separate those classes. If such hyperplanes do not exist, we have to modify the optimization problem (3.28) through introducing the 'slack variables' $\{\xi_n : n \in \mathbb{Z}_N^+\}$. Given a constant $C > 0$ , find $\{w_{ij} : 1 \le i \le j \le k\} \subset \mathbb{R}$, and $\{\xi_n : n \in \mathbb{Z}_N^+\} \subset \mathbb{R}$ to

$$\begin{aligned} \text{maximize} \quad & \sum_{1\le i\le j\le k}\frac{1}{\|w_{ij}\|} - C\sum_{n\in\mathbb{Z}_N^+}\xi_n, \qquad (3.34) \\ \text{subject to} \quad & \delta_{i,j,n}(w_{ij}^nx_n + b_{ij}) + 1 - \xi_n \le 0 \quad \forall 1\le i\le j\le k, n\in\mathbb{Z}_N^+, \\ & \xi_n \ge 0 \qquad\qquad \forall n\in\mathbb{Z}_N^+. \end{aligned}$$

The constant $C$ is a trade-off parameter between error and margin. The solution to (3.34) is called a soft margin. The dual optimization problem of (3.34) is: Finding

$\{\lambda_{i,j,n} : 1 \le i \le j \le k, n \in \mathbb{Z}_N^+\} \subset \mathbb{R}$ to

$$\begin{aligned} \text{minimize} \quad & 2\sum_{1\le i\le j\le k}\Big(\sum_{m,n\in\mathbb{Z}_N^+}\lambda_{i,j,n}\lambda_{i,j,m}\delta_{i,j,n}\delta_{i,j,m}x_n^Tx_m\Big)^{1/4} \\ & +\sum_{1\le i\le j\le k,n\in\mathbb{Z}_N^+}\lambda_{i,j,n} \qquad (3.35) \\ \text{subject to} \quad & \sum_{n\in\mathbb{Z}_N^+}\lambda_{i,j,n}\delta_{i,j,n}=0 \quad \text{for } 1\le i\le j\le k, \\ & 0\le\lambda_{i,j,n}\le C \quad \text{for } 1\le i\le j\le k, n\in\mathbb{Z}_N^+. \end{aligned}$$

If the solutions to (3.35) are $\{\lambda^*_{i,j,n} : 1 \le i \le j \le k, n \in \mathbb{Z}_N^+\}$, then the solutions to (3.34) are given by

$$w_{ij} = \frac{y_{ij}}{\|y_{ij}\|^{3/2}}, \quad \text{where } y_{ij} := \sum_{n\in\mathbb{Z}_N^+}\lambda_{i,j,n}\delta_{i,j,n}x_n, \qquad (3.36)$$

and $b^*_{ij}$ takes the value to get the minimum of the target function in (3.35), for all $1 \le i \le j \le k$.

If vectors $\{x_n : n \in \mathbb{Z}_N^+\}$ are mapped to a feature space, we hope in the feature space those classes can be separated better. Thus reproducing kernels come to the stage. Instead of calculating the inner product of $x_n$ and $x_n$ via $x_n^Tx_m$ in the target function in (3.35), we let the inner product of $x_n$ and $x_m$ be $K(x_n, x_m)$ where $K$ is a reproducing kernel function. This is the so-called kernel trick. The modified dual problem reads like: Finding $\{\lambda_{i,j,n} : 1 \le i \le j \le k, n \in \mathbb{Z}_N^+\} \subset \mathbb{R}$ to

$$\begin{aligned} \text{minimize} \quad & 2\sum_{1\le i\le j\le k}\Big(\sum_{m,n\in\mathbb{Z}_N^+}\lambda_{i,j,n}\lambda_{i,j,m}\delta_{i,j,n}\delta_{i,j,m}K(x_n,x_m)\Big)^{1/4} \\ & +\sum_{1\le i\le j\le k,n\in\mathbb{Z}_N^+}\lambda_{i,j,n} \qquad (3.37) \\ \text{subject to} \quad & \sum_{n\in\mathbb{Z}_N^+}\lambda_{i,j,n}\delta_{i,j,n}=0 \quad \text{for } 1\le i\le j\le k, \\ & 0\le\lambda_{i,j,n}\le C \quad \text{for } 1\le i\le j\le k, n\in\mathbb{Z}_N^+, \end{aligned}$$

where $C$ is a positive parameter. In this case, we use hyper-surfaces $K(w_{ij}^*, x)+b_{ij}^* = 0$ instead of hyperplanes $(w_{ij}^*)^T x + b_{ij}^* = 0$ for classification.

We explain specifically how to use problem (3.37) to solve multi-category classification problem. Suppose that there is a new vector $z$, we can predict the class of this new vector through those hyper-surfaces $K(w_{ij}^*, x) + b_{ij}^* = 0$ obtained by solving (3.37). First we determine whether $z$ is in $\mathcal{C}_1$. If $K(w_{1j}^*, x)+b_{1j}^* \geq 0$ for all $2 \leq j \leq k$, then we say $z \in \mathcal{C}_1$. If not, we check $K(w_{2j}^*, x) + b_{2j}^*$ for all $3 \leq j \leq k$ to see whether $z$ belongs to $\mathcal{C}_2$. If they are all nonnegative, $z \in \mathcal{C}_2$; otherwise, repeat this process until we find the class $z$ belongs to.

## 3.6 Classification for Heartbeat Interval Time Series

In this section, we develop the MSE-classifier, PLFME-classifier, HE-classifier and WPTE-classifier by using entropy values at all scales of MSE, PLFME, HE and WPTE with lower and higher piecewise linear wavelets filters respectively as features to classify different cardiac systems via the multi-category classification support vector machine. This heartbeat data set consists of 43 CHF subjects, 9 AF subjects, 46 YOUNG subjects and 26 OLD subjects. Since there are only 9 AF subjects and time series from AF subject has a clear decreasing signature as we see in previous sections, our target is to classify the CHF, YOUNG and OLD groups. We randomly select a

| | Labeled as | | |
|---|---|---|---|
| Recognized as | Old | CHF | YOUNG |
| Old | 16 | 4 | 3 |
| CHF | 2 | 18 | 0 |
| YOUNG | 2 | 1 | 10 |

Table 3.1: Classification result from the MSE-classifier.

| | Labeled as | | |
|---|---|---|---|
| Recognized as | Old | CHF | YOUNG |
| Old | 17 | 3 | 3 |
| CHF | 2 | 18 | 0 |
| YOUNG | 2 | 1 | 10 |

Table 3.2: Classification result from the PLFME-classifier.

half of time series to make the training set. The training set consists of 22 CHF, 23 OLD and 13 YOUNG. The remaining time series make up a test set. In this experiment, we use the data of length $8 \times 10^4$ and polynomial kernel $K(x, y) = (1 + x^T y)^2$ in SVM.

From Table 4.1 to 4.4, we can see that the WPTE-classifier has the best accuracy, followed by HE-classifier, PLFME-classifier and MSE-classifier. Of note, even piece-

| | Labeled as | | |
|---|---|---|---|
| Recognized as | Old | CHF | YOUNG |
| Old | 19 | 2 | 2 |
| CHF | 2 | 18 | 0 |
| YOUNG | 1 | 1 | 11 |

Table 3.3: Classification result from the HE-classifier.

| | Labeled as | | |
|---|---|---|---|
| Recognized as | Old | CHF | YOUNG |
| Old | 20 | 2 | 1 |
| CHF | 2 | 18 | 0 |
| YOUNG | 1 | 0 | 12 |

Table 3.4: Classification result from the WPTE-classifier.

wise linear wavelets have a better approximation than piecewise constant wavelets, HE-classifier is better than PLFME-classifier. Thus we can see that, in terms of classification, higher frequency components of the time series from wavelets packet transform does provide us useful information in the classification.

# Chapter 4

# Application of FME on DNA

The previous two chapters focus on filter-based sample entropy on time series derived from a continuous process. In this chapter, we discuss the application of FME on discrete time series. We propose a new scheme of FME to measure the complexity of a specific discrete time series, DNA sequences.

In both PLFME and WPTE, filtering the original time series can be viewed as approximation via continuous wavelets functions. It may not work if we approximate the discrete time series by continuous wavelets functions due to the finite number of values that each component of the discrete time series can take. For example, when we apply MSE on DNA sequences, there is an oscillation artifact [22]. We examine the unique properties of the sample entropy on discrete time series. Based on these properties, a multiscale 'eliminating' process is introduced via a special filter designed especially for the discrete time series. As usual, the sample entropy is then evaluated

on the filtered time series at multiscale.

In this chapter we use DNA sequences as our main study case. Through the dynamic process of evolution, the DNA sequence is likely to be the most sophisticated information database created by nature. The building blocks for DNA sequences are called nucleotides. Each nucleotide contains a phosphate group, a deoxyribose sugar moiety and either a purine or a pyrimidine base. Two purines and two pyrimidines are found in DNA. The two purines are adenine (A) and guanine (G); the two pyrimidines are cytosine (C) and thymine (T). Furthermore, the purine-pyrimidine rule maps bases A and G to the number 1, and bases C and T to the number -1 given the original DNA sequence. Thus each component of the time series derived from DNA sequences by the purine-pyrimidine rule only has two possible values. So we consider the discrete time series $\mathbf{x}$ such that

$$\mathbf{x}(i) \in \Theta := \{\alpha, \beta\}, \tag{4.1}$$

where $\alpha, \beta \in \mathbb{R}$ in this chapter. Discriminated by encoding protein or not, there are two different kinds of DNA sequences, coding DNA and noncoding DNA. We will examine and compare the complexity of both coding and noncoding DNA sequences.

This chapter is organized into three sections. In section 3.1, we explore the properties of sample entropy on discrete time series. Section 3.2 discusses the oscillation artifact of MSE on discrete time series. Section 3.3 is devoted to introducing a new 'eliminating' algorithm to measure the complexity of DNA sequences at multiple scales. Numerical results of 'eliminating' algorithm on DNA sequences are also given.

## 4.1 Sample Entropy on Discrete Time Series

In this section, we discuss the special properties of the sample entropy on the discrete time series $\mathbf{x}$ as defined in (4.1). To this end, we write

$$\mathbf{x} = (\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^n), \tag{4.2}$$

where $\mathbf{x}^i, i \in \mathbb{Z}^+$, is a segment from $\mathbf{x}$ and each component in $\mathbf{x}^i$ has the same value, $\alpha$ or $\beta$. For the convenience of notation, we write $\mathbf{x}^i = \mathbf{x}^j$ if elements in $\mathbf{x}^i$ and $\mathbf{x}^j$ have the same value and $x(i) \prec \mathbf{x}^j$ if the value of $x(i)$ is the same as the value of elements in $\mathbf{x}^j$. We also denote $|\mathbf{x}^j|$ by the number of elements in $\mathbf{x}^j$. We will see in the following propositions that given time series $\mathbf{x}$ as in (4.1), $S_m(\mathbf{x}, r)$ is only determined by the form (4.2) for a proper choice of $r$. Since we are only interested in small values of $m$ in real applications, we consider the the value of $S_m(\mathbf{x}, r)$ when $m = 1, 2$ in this section. From the expression of $S_m(\mathbf{x}, r)$, it is enough to compute $C^m(\mathbf{x}, r)$ when $m = 1, 2, 3$. For a given segment $\mathbf{x}^s$ and a positive number $N$, we define

$$D(\mathbf{x}^s, N) := \begin{cases} |\mathbf{x}^s| - N & \text{if } N < |\mathbf{x}^s|, \\ 0 & \text{otherwise.} \end{cases}$$

**Proposition 4.1.** *If the discrete time series* $\mathbf{x}$ *is defined as in (4.1) and* $r$ *satisfies* $r < |\alpha - \beta|$*, then*

$$C^1(\mathbf{x}, r) = \sum_{1 \le s \le n-1} \left[ \frac{|\mathbf{x}^s| D(\mathbf{x}^s, 1)}{2} + |\mathbf{x}^s| \sum_{t > s, \mathbf{x}^t = \mathbf{x}^s} |x^t| \right] + \frac{|\mathbf{x}^n| D(\mathbf{x}^n, 1)}{2}. \tag{4.3}$$

*Proof.* From the definition of $C^m(\mathbf{x}, r)$ in (2.3), we have

$$C^1(\mathbf{x}, r) = \frac{\#\{(x(i), x(j))|(x(i) = x(j))\}}{N}. \tag{4.4}$$

We consider the numerator in (4.4). If $x(i)$ and $x(j)$ are both in the segment $\mathbf{x}^s$, then the total number of pairs of $x(i), x(j)$ that satisfies $x(i) = x(j)$ is given by

$$\frac{(|\mathbf{x}^s|)(|\mathbf{x}^s| - 1|)}{2}. \tag{4.5}$$

If $x(i) \in \mathbf{x}^s$ and $x(j) \in \mathbf{x}^t$ with $s < t$, the total number of pairs of $(x(i), x(j))$ that satisfies $x(i) = x(j)$ is given by

$$|\mathbf{x}^s| \sum_{t>s, \mathbf{x}^t = \mathbf{x}^s} |x^t|. \tag{4.6}$$

Therefore, the total number of pairs of $(x(i), x(j))$ that satisfies $x(i) = x(j)$ is the sum of (4.5) and (4.6). We further make the summation for $s$, which gives us the numerator in (4.4). □

**Proposition 4.2.** *If the discrete time series* $\mathbf{x}$ *is defined as in (4.1) and* $r$ *satisfies* $r < |\alpha - \beta|$*, then*

$$C^2(\mathbf{x}, r) = \sum_{1 \le s \le n-1} \left[ \#\{\mathbf{x}^t | \mathbf{x}^t = \mathbf{x}^s, t > s\} + \frac{D(\mathbf{x}^s, 1) D(\mathbf{x}^s, 2)}{2} + D(\mathbf{x}^s, 1) \sum_{t>s, \mathbf{x}^t = \mathbf{x}^s} D(\mathbf{x}^t, 1) \right]$$
$$+ \frac{D(\mathbf{x}^n, 1) D(\mathbf{x}^n, 2)}{2}. \tag{4.7}$$

*Proof.* From the definition of $C^m(\mathbf{x}, r)$ in (2.3), we have

$$C^2(\mathbf{x}, r) = \frac{\#\{(x(i), x(j))|(x(i) = x(j)) \wedge (x(i+1) = x(j+1))\}}{N - 1}. \tag{4.8}$$

We compute the numerator in (4.8) first. Suppose $x(i)$ is fixed and $x(i) \in \mathbf{x}^s$, $1 \leq i \leq N-1, 1 \leq s \leq n$. We need to find the number of elements $x(j)$ such that $j \geq i, x(i) = x(j)$ and $x(i+1) = x(j+1)$. To satisfy the condition that $x(i) = x(j)$, we must have $x(j) \prec \mathbf{x}^s$. There are two cases when $x_{i+1} = x_{j+1}$ is also satisfied.

Case 1: $x(i+1) \neq x(i)$ and $x(j+1) \neq x(j)$. In this case, $x(i)$ must be the last component in $\mathbf{x}^s$. Suppose $x(j) \in \mathbf{x}^t$, then $x(j)$ is also the last components in $\mathbf{x}^t$. Thus $x_i, x_j$ can not be in the same segment. So in this case, the total number of pairs of $(x(i), x(j))$ that satisfies $(x(i) = x(j)) \wedge (x(i+1) = x(j+1))$ is

$$\#\{\mathbf{x}^t | \mathbf{x}^t = \mathbf{x}^s, t > s\}. \tag{4.9}$$

Case 2: $x(i+1) = x(i)$ and $x(j+1) = x(j)$. In this case, $x(i)$ can not be the last component in $\mathbf{x}^s$. Suppose $x(j) \in \mathbf{x}^t$, then $x(j)$ is not the last component in $\mathbf{x}^t$. When $s = t$, $|\mathbf{x}^s|$ can not be less than 3. The total number of pairs of $(x(i), x(j))$ that satisfies $(x(i) = x(j)) \wedge (x(i+1) = x(j+1))$ in $\mathbf{x}^s$ is

$$\frac{(|\mathbf{x}^s| - 1)(|\mathbf{x}^s| - 2)}{2}. \tag{4.10}$$

When $s < t$, the total number of pairs of $(x(i), x(j))$ that satisfies $(x_i = x_j) \wedge (x_{i+1} = x_{j+1})$ is

$$(|\mathbf{x}^s| - 1) \sum_{t>s, \mathbf{x}^t = \mathbf{x}^s} (|\mathbf{x}^t| - 1). \tag{4.11}$$

Therefore, the total number of pairs of $(x(i), x(j))$ that satisfies $(x(i) = x(j)) \wedge (x(i+1) = x(j+1))$ is the sum of (4.9), (4.10) and (4.11) for given $x(i)$. At the end, making the summation for $s$ gives us the numerator in (4.8). □

Let $N_2 := \#\{\mathbf{x}^s||\mathbf{x}^s| \geq 2\}$. We have the following proposition for the value of $C^3(\mathbf{x}, r)$.

**Proposition 4.3.** *If the discrete time series* $\mathbf{x}$ *is defined as in (4.1) and* $r$ *satisfies* $r < |\alpha - \beta|$, *then*

$$C^3(\mathbf{x}, r) = \tfrac{1}{N-2}\{\textstyle\sum_{1\leq s\leq n-1}\big[D(|\mathbf{x}^s|, 2)(\frac{D(|\mathbf{x}^s|,3)}{2} + \sum_{t>s,\mathbf{x}^t=\mathbf{x}^s} D(|x^t|, 2))\big] + \frac{D(|\mathbf{x}^n|, 1)D(|\mathbf{x}^n|, 2)}{2} + K_1 + K_2\},$$

*where* $K_1, K_2 = \frac{N_2(N_2-1)}{2}$ *or* $\frac{(N_2-1)(N_2-2)}{2}$.

*Proof.* From the definition of $C^m(\mathbf{x}, r)$ in (2.3), we have

$$C^3(\mathbf{x}, r) = \frac{\#\{(x(i), x(j))|(x(i) = x(j)) \wedge (x(i+1) = x(j+1)) \wedge (x(i+2) = x(j+2))\}}{N-2}. \tag{4.12}$$

Now we compute the numerator in (4.12). We consider the following cases.

Case 1: $x(i) = x(i+1) = x(i+2)$. In this case, we must have $x(j) = x(j+1) = x(j+2)$. It implies that $x(i), x(i+1), x(i+2)$ are in the same segment $\mathbf{x}^s$ and $(x(j), x(j+1), x(j+2))$ is also in the same segment. If $x(j) \in \mathbf{x}^s$, then the total number of $(x(i), x(j))$ that satisfies $(x(i) = x(j)) \wedge (x(i+1) = x(j+1)) \wedge (x(i+2) = x(j+2))$ is given by

$$\frac{(D(\mathbf{x}, 2))(D(\mathbf{x}, 3))}{2}. \tag{4.13}$$

If $x(j) \in \mathbf{x}^t$ with $t > s$, then the total number of $(x(i), x(j))$ that satisfies $(x(i) = x(j)) \wedge (x(i+1) = x(j+1)) \wedge (x(i+2) = x(j+2))$ is

$$\sum_{t>s,\mathbf{x}^t=\mathbf{x}^s} D(|\mathbf{x}^s|, 2)D(|\mathbf{x}^t|, 2). \tag{4.14}$$

Adding (4.13) and (4.14) together and making the summation for $s$, we have the total number of $(x(i), x(j))$ that satisfies $(x(i) = x(j)) \wedge (x(i+1) = x(j+1)) \wedge (x(i+2) = x(j+2))$ in this case is

$$\sum_{1\le s\le n-1} \big[D(|\mathbf{x}^s|, 2)(\frac{D(|\mathbf{x}^s|, 3)}{2} + \sum_{t>s, \mathbf{x}^t=\mathbf{x}^s} D(|x^t|, 2))\big] + \frac{D(|\mathbf{x}^n|, 1)D(|\mathbf{x}^n|, 2)}{2}. \quad (4.15)$$

Case 2: $x(i) = x(i+1), x(i+1) \neq x(i+2)$. In this case, we also have $x(j) = x(j+1), x(j+1) \neq x(j+2)$. Suppose $x(j) \in \mathbf{x}^t$, then we have $t > s$. Since $x(i+1) \neq x(i+2)$ and $x(j+1) \neq x(j+2)$, $x(i), x(i+1)$ must be the last two components of $\mathbf{x}^s$ and $x(j), x(j+1)$ must be the last two components of $\mathbf{x}^t$. The total number of $(x(i), x(j))$ that satisfies $(x(i) = x(j)) \wedge (x(i+1) = x(j+1)) \wedge (x(i+2) = x(j+2))$ in this case is $\frac{N_2(N_2-1)}{2}$ if $|x^1| \geq 2$ and is $\frac{N_2(N_2-1)}{2}$ if $|x^1| \leq 2$.

Case 3: $x(i) \neq x(i+1), x(i+1) = x(i+2)$. In this case, we also have $x(j) \neq x(j+1), x(j+1) = x(j+2)$. Suppose $x(j) \in \mathbf{x}^t$, then we have $t > s$. Since $x(i+1) \neq x(i)$ and $x(j+1) \neq x(j)$, $x(i+1), x(i+2)$ must be the first two components of $\mathbf{x}^s$ and $(x(j+1), x(j+2))$ must be the first two components of $\mathbf{x}^t$. The total number of $(x(i), x(j))$ that satisfies $(x(i) = x(j)) \wedge (x(i+1) = x(j+1)) \wedge (x(i+2) = x(j+2))$ in this case is $\frac{N_2(N_2-1)}{2}$ if $|x^n| \geq 2$ and is $\frac{N_2(N_2-1)}{2}$ if $|x^n| \leq 2$.

Considering all cases together, we get the numerator in (4.12). □

From Proposition 4.1, 4.2 and 4.3, we can see that the sample entropy of the discrete time series $\mathbf{x}$ is completely determined by the form $(\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^n)$.

## 4.2 MSE on DNA

Using sample entropy to explore the complexity of DNA sequences was firstly studied in [22] via the MSE analysis. There is an important artifact, oscillation, that affecters the MSE analysis for DNA sequences. This artifact was briefly discussed in [22] and we will describe it more thoroughly in this section.

For a given discrete time series, all components in the time series take value from a finite set. We call this finite set alphabet. We think the oscillation artifact in the MSE analysis is due to the increases size of alphabet as scales increases. At each scale new values, which are not in the original time series, are added to alphabet.

We consider an uncorrelated random variable, $X$, with alphabet $\Theta_1 := \{\alpha, \beta\}, \alpha, \beta \geq 0, \alpha \neq \beta$. We assume that both symbols $\alpha$ and $\beta$ occur with probability $1/2$. For simplicity, we first consider the time series with two components. There are only four possible different two-component sequences built from the binary series, which are $\alpha\alpha, \alpha\beta, \beta\alpha$ and $\beta\beta$. At the scale 2, we take the average of every two components in the original time series. Thus a new value, $(\alpha+\beta)/2$, is created. Therefore the alphabet of the coarse-grained time series corresponding to scale 2 is $\Theta_2 : \{\alpha, (\alpha+\beta)/2, \beta\}$. The probabilities associated with the occurrence of these three different values are $1/4, 1/2$ and $1/4$, respectively. If we use $r := (\alpha + \beta)/2$ to calculate the sample entropy, only the distance between the coarse-grained values $\alpha$ and $\beta$ (and not between values $\alpha$ and $(\alpha + \beta)/2$, and between values $(\alpha + \beta)/2$ and $\beta$) is higher than $r$. Therefore, the probability of distinguishing two data points randomly chosen from

the coarse-grained time series is obtained by

$$P_r(|x_a - x_b| > r) = p(\alpha) \times p(\beta) = \frac{1}{4} \times \frac{1}{4} = \frac{1}{16}.$$

Similarly, we consider the time series with there components. There are eight possible different three-component sequences built form the binary series, which are $\alpha\alpha\alpha, \alpha\alpha\beta, \alpha\beta\alpha, \beta\alpha\alpha, \beta\beta\alpha, \alpha\beta\beta, \beta\alpha\beta$ and $\beta\beta\beta$. At the scale 3, two new values, $(\alpha+\beta)/3, 2(\alpha+\beta)/3$, are created. Consequently, the alphabet of the coarse-grained time series corresponding to scale 3 is $\Theta_3 := \{\alpha, (\alpha+\beta)/3, 2(\alpha+\beta)/3, \beta\}$ and the probability associated with the occurrence of each value are $1/8, 3/8, 3/8$ and $1/8$, respectively. If we use $r := (\alpha+\beta)/2$ to calculate the sample entropy, only the distances between the coarse-grained data points 0 and $2(\alpha+\beta)/3$, $(\alpha+\beta)/3$ and $\beta$, and $\alpha$ and $\beta$ are higher that $r$. Therefore, the probability of distinguishing two data points randomly chosen from the coarse-grained time series is obtained by

$$P_r(|x_a - x_b| > r) = p(\alpha) \times p(2(\alpha+\beta)/3) + p((\alpha+\beta)/3) \times p(\beta) + p(\alpha) \times p(\beta) = \frac{7}{64}.$$

Note that the probability of distinguishing two data points of the coarse-grained time series increases from scale 2 to scale 3. As a consequence, the sample entropy also increases. For larger scales, a general result was provided in [22]. This artifact is due to the fact that the size of the alphabet of the coarse-grained time series increases with scales.

Due to the artifact we discussed above, MSE may not be suitable to DNA sequences well. In Fig. 4.2, we present the numerical results of MSE analysis on DNA

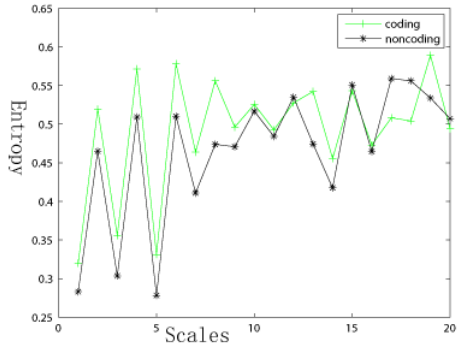



Figure 4.1: MSE results for coli DNA sequences with 4000 base pairs.

sequences. In this experiment, we take 3 coli coding DNA sequences and 3 coli non-coding DNA sequences. The entropy value of coding and noncoding DNA sequences shown in the figure is the mean entropy value of 3 DNA sequences. The oscillation artifact is shown on the MSE output curve and coding and noncoding DNA sequences are not well separated from Fig. 4.2.

In [22], several approaches are proposed to overcome this oscillation artifact. We will introduce a new algorithm to measure the complexity of DNA sequences in next section.

## 4.3 'Eliminating' Algorithm for DNA sequences

From section 2.1, we can see that the sample entropy of the discrete time series $\mathbf{x}$ is completely determined by the form $(\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^n)$. If we filter the discrete time series $\mathbf{x}$ by filters derived from continuous functions, we may have the oscillation artifact that MSE exhibits. We will introduce a new filter and a new algorithm in this section

based on the segment structure $(\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^n)$ to measure the complexity of DNA sequences at different scales and the oscillation artifact can be avoided.

We first define a 'eliminating' filter for $\mathbf{x}$, which aims to modify the segment structure (4.2).

**Definition 4.1.** Given a discrete time series $\mathbf{x}$ as defined in (4.1), we write it as the segment presentation as in (4.2). An 'eliminating' filter is defined as a block diagonal matrix as

$$A_{\mathbf{x}} := \begin{pmatrix} A_1 & & \\ & \ddots & \\ & & A_n \end{pmatrix},$$

where $A_s, s \in \mathbb{Z}^+$ is the matrix of the form $(I_{|bx_s|-1}, 0)$ if $|\mathbf{x}^s| > 1$ and is the $1 \times 1$ identity matrix if $|\mathbf{x}^s| = 1$.

**Algorithm 1:**

1. Compute $SE_1(\mathbf{x}, r)$.

2. At scale $\tau \geq 1$, a new time series $\mathbf{y}^\tau$ is generated from $\mathbf{y}^{\tau-1}$ ($\mathbf{y}^0 := \mathbf{x}$) by $\mathbf{y}^\tau := A_{\mathbf{y}^{\tau-1}}\mathbf{y}^{\tau-1}$.

3. Compute $SE_1(\mathbf{y}^\tau, r)$.

The largest number of $\tau$ we can take is $\max\{|\mathbf{x}^1|, |\mathbf{x}^2|, \ldots, |\mathbf{x}^n|\} - 1$. We next prove a decreasing property of the above algorithm. We first need the following technical lemma.

**Lemma 4.4.** *Let $a, b, a_1, b_1$ be positive numbers. If $a < b$ and $a - a_1 > b - b_1 > 0$, then $\frac{a}{b} < \frac{a_1}{b_1}$.*

*Proof.* It follows from the following calculation directly.

$$\begin{aligned} \frac{a}{b} - \frac{a_1}{b_1} &= \frac{ab_1 - a_1 b}{bb_1} = \frac{a(b_1 - b) - b(a_1 - a)}{bb_1} \\ &< \frac{a[(b_1 - b) - (a_1 - a))]}{bb_1} < 0. \end{aligned}$$

□

**Theorem 4.5.** *If $\mathbf{x}$ is a discrete time series as defined in (4.1) and let $\mathbf{y} = A_{\mathbf{x}}\mathbf{x}$, then $SE_1(\mathbf{x}, r) \geq SE_1(\mathbf{y}, r)$. Moreover, $SE_1(\mathbf{x}, r) = SE_1(\mathbf{y}, r)$ if and only if $\mathbf{x} = (\alpha, \beta, \alpha, \ldots)$ or $\mathbf{x} = (\beta, \alpha, \beta, \ldots)$.*

*Proof.* To prove $SE_1(\mathbf{x}, r) > SE_1(\mathbf{y}, r)$, it suffices to show that

$$\frac{C^2(\mathbf{x}, r)}{C^1(\mathbf{x}, r)} < \frac{C^2(\mathbf{y}, r)}{C^1(\mathbf{y}, r)}. \tag{4.16}$$

Let $A_{\mathbf{x}}$, $A_{\mathbf{y}}$ be the numerator of $C^2(\mathbf{x}, r)$ and $C^2(\mathbf{y}, r)$. Let $B_{\mathbf{x}}$, $B_{\mathbf{y}}$ be the numerator of $C^1(\mathbf{x}, r)$ and $C^1(\mathbf{y}, r)$. Then $A_{\mathbf{x}} - B_{\mathbf{x}}$ is the number of pairs $(x(i), x(j))$ such that $x(i) = x(i) \wedge x(i+1) \neq x(j+1)$ and $A_{\mathbf{y}} - B_{\mathbf{y}}$ is the number of pairs $(y(i), y(j))$ such that $y(i) = y(i) \wedge y(i+1) \neq y(j+1)$. Suppose the length of $\mathbf{x}$ is $N$ and the length of $\mathbf{y}$ is M. From the construction of $\mathbf{y}$, we have $A_{\mathbf{x}} - B_{\mathbf{x}} < y(i) = y(i) \wedge y(i+1) \neq y(j+1)$ and $N > M$. Thus

$$\frac{A_{\mathbf{x}} - B_{\mathbf{x}}}{N} < \frac{A_{\mathbf{y}} - B_{\mathbf{y}}}{M},$$

which implies

$$\frac{A_{\mathbf{x}}}{N} - \frac{A_{\mathbf{y}}}{M} < \frac{B_{\mathbf{x}}}{N} - \frac{B_{\mathbf{y}}}{M}. \tag{4.17}$$

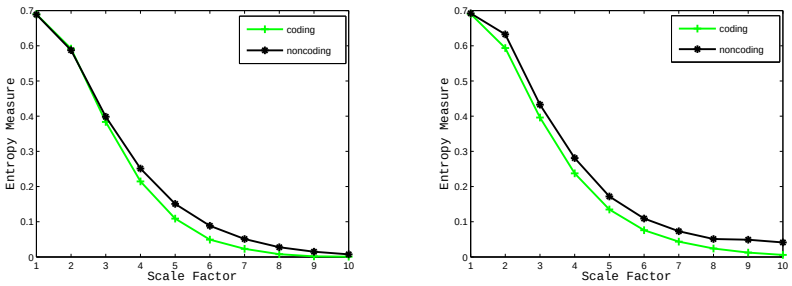
It follows from inequality (4.17), the fact $\frac{A_{\mathbf{x}}}{N} < \frac{B_{\mathbf{x}}}{N}$ and Lemma 4.4 that (4.16) holds.

From the construction of $\mathbf{y}$, we have $SE_1(\mathbf{x}, r) = SE_1(\mathbf{y}, r)$ if and only $\mathbf{x}$ and $\mathbf{y}$ are the same sequence. $\mathbf{x}$ and $\mathbf{y}$ are the same sequence if and only if $\mathbf{x} = (\alpha, \beta, \alpha, \ldots)$ or $\mathbf{x} = (\beta, \alpha, \beta, \ldots)$. □

Now we apply Algorithm 1 introduced above to the analysis of DNA sequences, likely one of the most complex natural information databases.

The role of genomic DNA sequences in coding for protein structure is well known [37]. The genomic sequence is likely to be the most sophisticated information database created by nature through the dynamic process of evolution. Equally remarkable is the precise transformation of information (duplication, decoding, etc.) that occurs in a relatively short time interval.

The DNA building units are called nucleotides. Two of them contain a purine base, adenine (A) or guanine (G), and the other two contain a pyrimidine base, cytosine (C) or thymine (T). There are different ways of mapping the DNA sequences to a numerical sequence that take into different properties of DNA sequences ( [11, 42], etc.). In this application, we use the purine-pyrimidine rule [14, 53, 54]. Recent work [52] indicates that the original purine-pyrimidine rule provides the most robust results in the study of some statistical properties of DNA sequences, probably due to the purine-pyrimidine chemical complementarity. Given the original DNA sequences, bases A and G are mapped to number 1, and bases C and T are mapped to number -1.




(a) 4000 bp (b) 6000 bp

Figure 4.2: Numerical results for DNA sequences.

In Fig.4.2, we present the numerical results for selected coding and noncoding coli sequences. The results support the view that noncoding sequences contain important biological information. As pointed out by others [5, 6, 12, 47], biological complexity and phenotype variations should related not only to proteins, which are the main effectors of cellular activity, but also to the organizational structure of the control mechanism responsible for the networking and integration of gene activity.

This verification that noncoding DNA sequences are more complex than coding DNA sequences implies the applicability of FME on discrete time series. One may design other special filters when the alphabet of the discrete time series has a bigger size.

# Bibliography

[1] L. Angelini, R. Maestri, D. Marinazzo, L. Nitti, M. Pellicoro, G.D. Pinna, S. Stramaglia, and S.A. Tupputi. Multiscale analysis of short term heart beat interval, arterial blood pressure, and instantaneous lung volume time series. *Artificial intelligence in medicine*, 41(3):237–250, 2007.

[2] T. Antal, M. Droz, G. Györgyi, and Z. Rácz. Roughness distributions for $1/f^{\alpha}$ signals. *Phys. Rev. E*, 65:046140, Apr 2002.

[3] Nachman Aronszajn. Theory of reproducing kernels. *Transactions of the American mathematical society*, 68(3):337–404, 1950.

[4] H. Atmanspacher and H. Scheingraber. *Information dynamics*, volume 1990. Plenum Pub Corp, 1991.

[5] Benjamin Audit, Claude Thermes, Cedric Vaillant, Yves d'Aubenton Carafa, Jean-Francois Muzy, and Alain Arneodo. Long-range correlations in genomic dna: a signature of the nucleosomal structure. *Physical Review Letters*, 86(11):2471–2474, 2001.

[6] Benjamin Audit, Cedric Vaillant, Alain Arneodo, Yves d'Aubenton Carafa, and Claude Thermes. Long-range correlations between dna bending sites: relation to the structure and dynamics of nucleosomes. *Journal of Molecular Biology*, 316(4):903–918, 2002.

[7] C. Bahlmann, B. Haasdonk, and H. Burkhardt. Online handwriting recognition with support vector machines-a kernel approach. In *Frontiers in Handwriting Recognition, 2002. Proceedings. Eighth International Workshop on*, pages 49–54. IEEE, 2002.

[8] Axel Bauer, Marek Malik, Georg Schmidt, Petra Barthel, Hendrik Bonnemeier, Iwona Cygankiewicz, Przemyslaw Guzik, Federico Lombardi, Alexander Muller, Ali Oto, Raphael Schneider, Mari Watanabe, Dan Wichterle, and Wojciech Zareba. Heart rate turbulence: Standards of measurement, physiological interpretation, and clinical use: International society for holter and noninvasive electrophysiology consensus. *Journal of the American College of Cardiology*, 52(17):1353 – 1365, 2008.

[9] A Berkowitsch, N Guettler, T Neumann, D Vukajlovic, B Schulte, J Neuzner, and HF Pitschner. Turbulence jump-a new descriptor of heart-rate turbulence after paced premature ventricular beats. a study in dilated cardiomyopathy patients. *Eur Heart J*, 22(547):P2941, 2001.

[10] J. Berntgen, T.L. Lim, K. Heime, W. Daumann, U. Auer, FJ Tegude, and A. Matulionis. The hooge parameter of 2deg quantum well structures and its dependence on temperature. In *Proc. of the 15th Int. Conf. on Noise in Physical Systems and 1/f Fluctuations, Hong-Kong, August 23-26*, pages 291–294, 1999.

[11] Cheryl L Berthelsen, James A Glazier, and Mark H Skolnick. Global fractal dimension of human dna sequences treated as pseudorandom walks. *Physical Review A*, 45(12):8902, 1992.

[12] Athanasius F Bompfünewerer, Christoph Flamm, Claudia Fried, Guido Fritzsch, Ivo L Hofacker, Jörg Lehmann, Kristin Missal, Axel Mosig, Bettina Müller, Sonja J Prohaska, et al. Evolutionary patterns of non-coding rnas. *Theory in Biosciences*, 123(4):301–369, 2005.

[13] B.E. Boser, I.M. Guyon, and V.N. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152. ACM, 1992.

[14] SV Buldyrev, AL Goldberger, S Havlin, RN Mantegna, ME Matsa, C-K Peng, M Simons, and HE Stanley. Long-range correlation properties of coding and noncoding dna sequences: Genbank analysis. *Physical Review E*, 51(5):5084, 1995.

[15] George Casella and Roger Berger. *Statistical Inference*. Duxbury Resource Center, June 2001.

[16] Dante R Chialvo. Physiology: unhealthy surprises. *Nature*, 419(6904):263–263, 2002.

[17] R.R. Coifman and M.V. Wickerhauser. Entropy-based algorithms for best basis selection. *Information Theory, IEEE Transactions on*, 38(2):713–718, 1992.

[18] M. Costa and J.A. Healey. Multiscale entropy analysis of complex heart rate dynamics: discrimination of age and heart failure effects. In *Computers in Cardiology, 2003*, pages 705 – 708, sept. 2003.

[19] M. Costa, C.-K. Peng, Ary L. Goldberger, and Jeffrey M. Hausdorff. Multiscale entropy analysis of human gait dynamics. *Physica A: Statistical Mechanics and its Applications*, 330(1?):53 – 60, 2003.

[20] Madalena Costa, Ary L. Goldberger, and C.-K. Peng. Multiscale entropy analysis of complex physiologic time series. *Phys. Rev. Lett.*, 89:068102, Jul 2002.

[21] Madalena Costa, Ary L. Goldberger, and C.-K. Peng. Costa, goldberger, and peng reply:. *Phys. Rev. Lett.*, 92:089804, Feb 2004.

[22] Madalena Costa, Ary L. Goldberger, and C.-K. Peng. Multiscale entropy analysis of biological signals. *Phys. Rev. E*, 71:021906, Feb 2005.

[23] Timothy A Denton, George A Diamond, Richard H Helfant, Steven Khan, and Hrayr Karagueuzian. Fascinating rhythm: a primer on chaos theory and its application to cardiology. *American heart journal*, 120(6):1419–1440, 1990.

[24] Xianfeng Fan and Ming J Zuo. Gearbox fault detection using hilbert and wavelet packet transform. *Mechanical Systems and Signal Processing*, 20(4):966–982, 2006.

[25] Hans C. Fogedby. On the phase space approach to complexity. *Journal of Statistical Physics*, 69, 1992.

[26] M.W. Frazier. *An introduction to wavelets through linear algebra*. Springer, 1999.

[27] Y. Gao. *Multi-category support vector machines*. 2006. Thesis (Ph.D.)–Syracuse University.

[28] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, Ch, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, June 2000 (June 13).

[29] A.L. Goldberger, C.K. Peng, and L.A. Lipsitz. What is physiologic complexity and how does it change with aging and disease? *Neurobiology of Aging*, 23, 2002.

[30] Ary L Goldberger. Is the normal heartbeat chaotic or homeostatic? *Physiology*, 6(2):87–91, 1991.

[31] Ary L Goldberger. Non-linear dynamics for clinicians: chaos theory, fractals, and complexity at the bedside. *The Lancet*, 347(9011):1312–1314, 1996.

[32] Ary L GOLDBERGER, C-K PENG, Lewis A LIPSITZ, David E VAILLANCOURT, and Karl M NEWELL. What is physiologic complexity and how does it change with aging and disease? authors' reply. *Neurobiology of aging*, 23(1):27–29, 2002.

[33] J. Guo, H. Chen, Z. Sun, and Y. Lin. A novel method for protein secondary structure prediction using dual-layer svm and profiles. *PROTEINS: Structure, Function, and Bioinformatics*, 54(4):738–743, 2004.

[34] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik. Gene selection for cancer classification using support vector machines. *Machine learning*, 46(1):389–422, 2002.

[35] Przemysław Guzik and Georg Schmidt. A phenomenon of heart-rate turbulence, its evaluation, and prognostic value. *Cardiac electrophysiology review*, 6(3):256–261, 2002.

[36] M. Hu and H. Liang. Adaptive multiscale entropy analysis of multivariate neural data. *Biomedical Engineering, IEEE Transactions on*, 59(1):12–15, 2012.

[37] J. Witkowski J.D. Watson, M.Gilman and M. Zoller. *Recombinant DNA*. Ssientific American Books, New York, 1992.

[38] Ying Jiang, C.-K. Peng, and Yuesheng Xu. Hierarchical entropy analysis for biological signals. *J. Comput. Appl. Math.*, 236(5):728–742, 2011.

[39] Koyama Jiro, Watanabe Jun, Yamada Aki, Koseki Yoshito, Konno Yuji, Toda Sunao, Shinozaki Tsuyoshi, Miura Masahito, Fukuchi Mitsumasa, Ninomiya Mototsugu, Kagaya Yutaka, and Shirato Kunio. Evaluation of heart-rate turbulence as a new prognostic marker in patients with chronic heart failure. *Circulation journal : official journal of the Japanese Circulation Society*, 66(10):902–907, 2002.

[40] T. Joachims. Text categorization with support vector machines: Learning with many relevant features. *Machine learning: ECML-98*, pages 137–142, 1998.

[41] Andrew Laine and Jian Fan. Texture classification by wavelet packet signatures. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 15(11):1186–1191, 1993.

[42] Wentian Li, Thomas G Marr, and Kunihiko Kaneko. Understanding long-range correlations in dna sequences. *Physica D: Nonlinear Phenomena*, 75(1):392–416, 1994.

[43] Qiaofang Lian, Lixin Shen, Yuesheng Xu, and Lihua Yang. Filters of wavelets on invariant sets for image denoising. *Appl. Anal.*, 90(8):1299–1322, 2011.

[44] R. Liu, Y. Wang, T. Baba, D. Masumoto, and S. Nagata. Svm-based active feedback in image retrieval using clustering and unlabeled data. *Pattern Recognition*, 41(8):2645–2655, 2008.

[45] Stéphane Mallat. *A wavelet tour of signal processing*. Access Online via Elsevier, 1999.

[46] Dong Mao. *Biological time series classification via reproducing kernels and sample entropy*. ProQuest LLC, Ann Arbor, MI, 2008. Thesis (Ph.D.)–Syracuse University.

[47] John S Mattick. Non-coding rnas: the architects of eukaryotic complexity. *EMBO reports*, 2(11):986–991, 2001.

[48] C.A. Micchelli, Y. Xu, and H. Zhang. Universal kernels. *The Journal of Machine Learning Research*, 7:2651–2667, 2006.

[49] Charles A. Micchelli and Yuesheng Xu. Using the matrix refinement equation for the construction of wavelets on invariant sets. *Appl. Comput. Harmon. Anal.*, 1(4):391–401, 1994.

[50] Charles A. Micchelli and Yuesheng Xu. Reconstruction and decomposition algorithms for biorthogonal multiwavelets. *Multidimens. Systems Signal Process.*, 8(1-2):31–69, 1997.

[51] Vadim V. Nikulin and Tom Brismar. Comment on "multiscale entropy analysis of complex physiologic time series". *Phys. Rev. Lett.*, 92:089803, Feb 2004.

[52] SM Ossadnik, SV Buldyrev, AL Goldberger, S Havlin, RN Mantegna, CK Peng, M Simons, and HE Stanley. Correlation approach to identify coding regions in dna sequences. *Biophysical Journal*, 67(1):64–70, 1994.

[53] C-K Peng, SV Buldyrev, AL Goldberger, S Havlin, RN Mantegna, M Simons, and HE Stanley. Statistical properties of dna sequences. *Physica A: Statistical Mechanics and its Applications*, 221(1):180–192, 1995.

[54] CK Peng, SV Buldyrev, AL Goldberger, S Havlin, F Sciortino, M Simons, HE Stanley, et al. Long-range correlations in nucleotide sequences. *Nature*, 356(6365):168–170, 1992.

[55] Steven M Pincus. Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences*, 88(6):2297–2301, 1991.

[56] Steven M Pincus. Assessing serial irregularity and its implications for health. *Annals of the New York Academy of Sciences*, 954(1):245–267, 2001.

[57] Alberto Porta, Stefano Guzzetti, Nicola Montano, Raffaello Furlan, Massimo Pagani, Alberto Malliani, and Sergio Cerutti. Entropy, entropy rate, and pattern classification as tools to typify complexity in short heart period variability series. *Biomedical Engineering, IEEE Transactions on*, 48(11):1282–1291, 2001.

[58] Joshua S. Richman and J. Randall Moorman. Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology - Heart and Circulatory Physiology*, 278(6):H2039–H2049, 2000.

[59] Georg Schmidt, Marek Malik, Petra Barthel, Raphael Schneider, Kurt Ulm, Linda Rolnitzky, A John Camm, J Thomas Bigger Jr, and Albert Schömig. Heart-rate turbulence after ventricular premature beats as a predictor of mortality after acute myocardial infarction. *The Lancet*, 353(9162):1390 – 1396, 1999.

[60] JÖG OTTO Schwab, GERRIT Eichner, NIKOLAY Shlevkov, JAN Schrickel, ALEXANDER Yang, OSMAN Balta, THORSTEN Lewalter, and BERNDT Lüeritz. Impact of age and basic heart rate on heart rate turbulence in healthy persons. *Pacing and Clinical Electrophysiology*, 28:S198–S201, 2005.

[61] Jerome M Shapiro. Embedded image coding using zerotrees of wavelet coefficients. *Signal Processing, IEEE Transactions on*, 41(12):3445–3462, 1993.

[62] George Sugihara, Walter Allan, Daniel Sobel, and Kenneth D Allan. Nonlinear control of heart rate variability in human infants. *Proceedings of the National Academy of Sciences*, 93(6):2608–2613, 1996.

[63] Hideki Takayasu. *Fractals in physical sciences*. Manchester University Press, 1990.

[64] Joel A Tropp. Greed is good: Algorithmic results for sparse approximation. *Information Theory, IEEE Transactions on*, 50(10):2231–2242, 2004.

[65] Z. Trunkvalterova, M. Javorka, I. Tonhajzerova, J. Javorkova, Z. Lazarova, K. Javorka, and M. Baumert. Reduced short-term complexity of heart rate and blood

pressure dynamics in patients with diabetes mellitus type 1: multiscale entropy analysis. *Physiological measurement*, 29(7):817, 2008.

[66] José F Valencia, Alberto Porta, Montserrat Vallverdú, Francesc Claria, Rafal Baranowski, Ewa Orlowska-Baranowska, and Pere Caminal. Refined multiscale entropy: Application to 24-h holter recordings of heart period variability in healthy and aortic stenosis subjects. *Biomedical Engineering, IEEE Transactions on*, 56(9):2202–2213, 2009.

[67] Saila Vikman, Kai Lindgren, Timo H. Mäkikallio, Sinikka Yli-Mäyry, K.E. Juhani Airaksinen, and Heikki V. Huikuri. Heart rate turbulence after atrial premature beats before spontaneous onset of atrial fibrillation. *Journal of the American College of Cardiology*, 45(2):278 – 284, 2005.

[68] Claus D Wagner and Pontus B Persson. Chaos in the cardiovascular system: an update. *Cardiovascular Research*, 40(2):257–264, 1998.

[69] Mari A Watanabe, Joseph E Marine, Robert Sheldon, and Mark E Josephson. Effects of ventricular premature stimulus coupling interval on blood pressure and heart rate turbulence. *Circulation*, 106(3):325–330, 2002.

[70] Francis X Witkowski, Katherine M Kavanagh, Patricia A Penkoske, Robert Plonsey, Mark L Spano, William L Ditto, Daniel T Kaplan, et al. Evidence for determinism in ventricular fibrillation. *Physical review letters*, 75(6):1230–1233, 1995.

[71] Yicheng Zhang. Complexity and $1/f$ noise. a phase space approach. *J. Phys. I France*, 1, 1991.

## BIOGRAPHICAL DATA

NAME OF AUTHOR: Liang Zhao

PLACE OF BIRTH: CHINA

DATE OF BIRTH: December 30, 1983

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

Jilin University

DEGREES AWARDED:

B.S., Jilin University - CHINA, 2007

M.S., Syracuse University - Syracuse, 2009

PUBLICATION:

- Haizhang Zhang and Liang Zhao: On the Inclusion Relation of Reproducing Kernel Hilbert Spaces, *Analysis and Applications* 11 (2013).

- Yuesheng Xu and Liang Zhao: Filter-Based Multiscale Entropy Analysis of Complex Physiologic Time Series, *Physical Review E* Accepted.