INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

NOTE TO USERS

The original manuscript received by UMI contains pages with indistinct and/or slanted print.   Pages were microfilmed as received.

This reproduction is the best copy available

UMI

Wavelet Techniques for Chaotic and Fractal Dynamics

by

William L.B. Constantine

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Washington

1999

Approved by_____

(Chairperson of Supervisory Committee)

_____

_____

_____

_____

_____

Program Authorized
to Offer Degree_____Mechanical Engineering_____

Date___2/12/99_____

University of Washington

Abstract

Wavelet Techniques for Chaotic and Fractal Dynamics

by William L.B. Constantine

Chairperson of Supervisory Committee
Professor Per G. Reinhall
Mechanical Engineering

A novel wavelet based denoising technique is developed and shown to be more effective on average than waveshrink in denoising contaminated chaotic signals. A chaotic beam experiment is used to verify its effectiveness.

Correlation dimension (D_2) convergence issues for stochastic colored noise processes are revisited with the Gaussian Spectral Synthesis method. The inadequacies of the Phase Randomization method are discussed and is shown to be an inappropriate means to investigate D_2 convergence.

A ensemble of nonlinear and linear measures is used to assess determinism and existing fractal structure in RR intervals from over 50 heart patients. Techniques are critiqued for their ability to classify disease states of the heart and predict the onset of fatal cardiac rhythms.

TABLE OF CONTENTS

iv

LIST OF FIGURES

ix

NOTATION

► *Transforms and Algorithms*

AAFT   Theiler's Amplitude Adjusted Fourier Transform surrogate data algorithm

BDSWV   Bridge Detrended Scaled Windowed Variance

CWT   Continuous Wavelet Transform

DFT   Discrete Fourier Transform

Disp   Dispersional analysis

DSM   Direct Summation Method

DWPT   Discrete Wavelet Packet Transform

DWT   Discrete Wavelet Transform

FNN   False Nearest Neighbor

FFT   Fast Fourier Transform

FT   Theiler's Fourier Transform surrogate data algorithm

GSSM   Gaussian Spectral Synthesis Method

IDFT   Inverse Discrete Fourier Transform

KDM   Kaplan's Determinism Method

MODRA   MODWPT Optimal Deterministic Reduction Algorithm

MODWPT   Maximum Overlap Discrete Wavelet Packet Transform

MODWT   Maximum Overlap Discrete Wavelet Transform

PRM   Phase Randomization Method

STFT   Short-Time Fourier Transform

SVD   Singular Value Decomposition

| SWV | Scaled Windowed Variance |
| WS | Waveshrink or Wavelet Shrinkage |

▶ *Nonlinear Dynamics and Fractal Mathematics*

D_0	box-counting or capacity dimension
D_1	information dimension
D_2	correlation dimension
D_q	generalized fractal dimensions
D_L	Lyapunov dimension
D_e	embedding dimension
$C(r, E)$	correlation integral at scale r and embedding dimension E
τ	time lag for single variable embeddings
κ	KDM image lead
γ	neighbor trajectory lead
δ_{kap}	Kaplan rejection threshold
H	Hurst coefficient

▶ *Statistics and Mathematics*

rv	random variable
acvs	autocovariance sequence
fGn	Fractional Gaussian noise
MAD	Median Amplitude Deviation
MSE	Mean Squared Error
SNR	Signal-to-Noise Ratio
$\widehat{\sigma}_X^2$	sample variance of $\{X_t\}$
$\widehat{\sigma}_X$	sample standard deviation of $\{X_t\}$

σ_X^2	sample variance of rv$\{X\}$
σ_X	sample standard deviation of rv$\{X\}$
$\mathbf{sign}\{y\}$	sign of number y
$\|X\|$	L^2 norm of sequence $\{X_t\}$
$\mathcal{I}$	identity matrix
$E\{\cdot\}$	expected value operator
$corr\{\cdot,\cdot\}$	correlation operator
$cov\{\cdot,\cdot\}$	covariance operator
R_{XX}	autocorrelation function of sequence $\{X_t\}$
$\mathbb{R}$	set of real numbers
$\mathbb{Z}$	set of integers
$\mathbb{N}$	set of positive integers
$*$	convolution operator
$1/f^\alpha$	stochastic one-over-f process corresponding to exponent α
$S_X(f)$	power spectral density of a real valued sequence $\{X_t\}$
$L^2(\cdot)$	set of squared integrable functions
$\mathbb{R}^N$	set of real function spanning N dimensions
$\{X_t\}$	discrete real valued time series indexed by "time" t
$X(f)$	value of the function $X(\cdot)$ at f
X_t	single value of a discrete real valued time series $\{X_t\}$
$\underline{\mathbf{X}}$	matrix X
$\underline{\mathbf{X}}^{\mathbf{T}}$	transpose of matrix $\underline{\mathbf{X}}$
$\mathbf{X}$	column vector X
$\lfloor\cdot\rfloor$	round towards $-\infty$ operator
$\lceil\cdot\rceil$	round towards $+\infty$ operator

► *Wavelet Specific*

QMF	Quadrature Mirror Filter
δ_{DJ}	Donoho and Johnstone WS threshold
δ_E	Energy admissibility threshold
$\{h_l\}$	impulse response of wavelet filter
$\{g_l\}$	impulse response of scaling filter
$\{\widetilde{h}_l\}$	impulse response of maximum overlap wavelet filter
$\{\widetilde{g}_l\}$	impulse response of maximum overlap scaling filter
$\mathcal{H}(\cdot)$	wavelet filter squared gain function
$\mathcal{G}(\cdot)$	scaling filter squared gain function
$H(f)$	wavelet filter frequency response function
$G(f)$	scaling filter frequency response function
τ_j	scale at decomposition level j
j	j^{th} decomposition level
J	represents a level J decomposition
$\downarrow 2$	decimation (downsampling) by two
$\uparrow 2$	upsampling by two
V_j	scaling coefficient at index j
W_j	wavelet coefficient at index j
$W_{j,n}$	DWT (DWPT) coefficients at level j, local node index n
$\widetilde{W}_{j,n}$	MODWT (MODWPT) coefficients at level j, local node index n
$W_{j,n,t}$	DWT or DWPT coefficient at time t in node $W_{j,n}$
$\widetilde{W}_{j,n,t}$	the MODWT or MODWPT coefficient at time t in node $\widetilde{W}_{j,n}$
$\underline{\mathcal{W}}$	wavelet transform matrix
$\mathcal{D}$	DWT or DWPT detail
$\widetilde{\mathcal{D}}$	MODWT or MODWPT detail

DEDICATION

To my family. Thanks for believing in me.

ACKNOWLEDGMENTS

Chapter 1

INTRODUCTION

In recent decades, significant advances have been made in wavelet theory, fractal mathematics, and chaos theory. The purpose of this dissertation is to introduce a novel wavelet based denoising technique and to sufficiently demonstrate its applicability to nonlinear dynamics. A secondary focus is to refine the existing knowledge of the link between chaos, fractals, and stochastic modeling. Finally, an ensemble of conventional methods and novel techniques developed herein are used to further our understanding of the dynamics of a highly complex physiological system: the human heart.

1.1 Impetus of the Research

1.1.1 Wavelet Techniques

It is well known that nonlinear measures of chaotic systems are sensitive to the influence of noise contamination. Figure 1.1 illustrates the effects of noise on a common nonlinear measure of chaotic systems. In the left column, a collection of points, associated with a deterministic chaotic system, is projected into a 2-dimensional space known as the phase plane. The correlation dimension, also referred to as D_2, is a nonlinear measure used to characterize the resulting topology and is estimated by a linear approximation of the slopes of the correlation integrals (second row of Fig. 1.1). A flat, plateau-like region in D_2 indicates a convergent measure that

is nonexistent for a stochastic white noise process (right column of Fig. 1.1). A
chaotic system contaminated with a small amount of noise can have a significant
effect on the D_2 estimate, yielding nonconvergent results. Thus, it is of consider-
able interest to develop efficient and effective techniques to denoise contaminated
chaotic response. Currently there exists a powerful, wavelet based denoising tech-
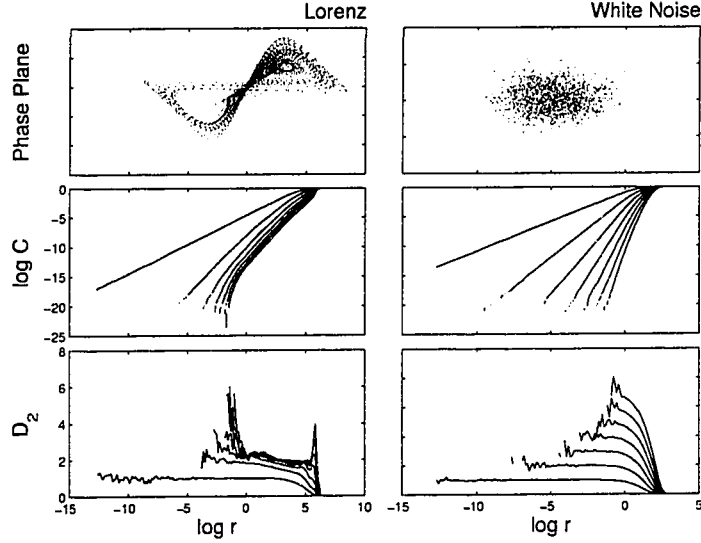
Figure 1.1: Comparison of a deterministic chaotic and a stochastic system with regard to a nonlinear
measure

nique, known as waveshrink, which has been shown to work remarkably well on a
wide variety of signals. Its ability to denoise is dependent upon the automatic selec-
tion of a wavelet coefficient rejection threshold. The threshold is estimated based on
the assumption of the type of noise which contaminates the signal, which may not be
appropriate for band limited noise or broadband chaotic response. The advantage
of the wavelet denoising technique developed herein is that it directly attempts to
extract the deterministic features of the data without making an underlying assump-
tion on the type and (broadband) extent of the noise. Our research indicates that

the new technique outperforms waveshrink on a fairly consistent basis for denoising contaminated chaotic sequences.

1.1.2 D_2 and $1/f^\alpha$ Processes

Phase randomization methods (PRMs) have been used to simulate stochastic processes whose power spectra have a $1/f^\alpha$ falloff. However, PRMs impose an unnecessary deterministic structure in $1/f^\alpha$ noise simulations and generally do a poor job in emulating true colored noise processes. Consequently, the PRM is an inappropriate means to investigate D_2 convergence issues for colored noise. In this dissertation, we revisit these issues with a technique which overcomes the inadequacies of the PRM, and extend the study for red, white, and blue noise processes. These results have implications for experimentalists interested in classifying fractal structure for sequences of unknown origin and confirms the notion that stochastic fractal processes do not yield convergent global generalized fractal dimensions if analyzed correctly.

1.1.3 Dynamics of the Human Heart

Recent advances in cardiodynamic modeling suggest that the highly complex rhythms of the human heart may potentially be described by a low order chaotic system. Understanding the dynamics of the human heart is of extreme value to the medical community, as heart disease is the number one killer in America with over $350,000$ sudden cardiac deaths occurring every year [21].

The advances made in chaos, wavelet, and fractal mathematics can play a significant role in our understanding of the heart. The task at hand is to systematically assess the validity and effectiveness of both nonlinear and linear algorithms, and to obtain a clearer picture of what physiological measures of the heart contribute most significantly to fatal rhythms such as ventricular fibrillation (VF) or ventricular tachycardia (VT).

In this dissertation, an ensemble of relevant nonlinear and linear techniques is used to study heart rate variability in both healthy (5) and sick (48) patients. The study will be performed on *RR intervals* which are defined as the differences in time between adjacent QRS complexes in an electrocardiogram (ECG) recording. The QRS complexes correspond to the rapid depolarization of the ventricles during a normal heart beat. The following questions are assessed:

1. Are RR intervals generated by a deterministic or stochastic process?

2. Is there any topological fractal structure in RR intervals, indicating the existence of an underlying low order deterministic chaotic system?

3. Do RR intervals exhibit any spectral scaling, indicating statistical self similarity common to colored random noise processes?

4. Can any of the measures be used to predict the onset of VF/VT?

5. Can cardiac arrythmias be classified by a MODWPT best basis for a representative sample of a patient's RR intervals?

1.2   Contributions to the Field

The following list the contributions to nonlinear dynamics, signal processing, systems modeling, and cardiodynamics.

1. Development of MODRA: a novel wavelet based denoising technique

2. Correct assessment of correlation dimension convergence issues for stochastic colored noise processes

3. Development of a wavelet based ECG feature extraction algorithm

4. Investigation of inability of the Wavelet Transform Maximum Modula method to characterize chaotic flow

5. Analysis of RR intervals in 53 patients, both sick and healthy, using scaled window variance, MODWT variance, approximate entropy, correlation dimension, dispersion analysis, MODWPT best basis, and Fourier techniques

6. Investigation of limitations on Kaplan's determinism method for colored noise

7. Development of a Shannon entropy functional for use with the MODWPT

8. Development of a chaotic beam experiment for verification of MODRA effectiveness

Chapter 2

INTRODUCTION TO CHAOTIC SYSTEMS AND FRACTALS

This chapter details a brief history of chaos and fractals, and provides the link between them. Examples of chaos and fractals in computer generated sequences are given, and methods used to characterize chaos and fractals are defined. Finally, a wavelet based technique used to quantify fractal structure is reviewed, and its inability to properly analyze chaotic sequences is critiqued.

2.1 Description of Chaos

While chaotic behavior was recognized long ago by scientists such as Poincaré [76], Birkhoff, Van der Pol and Van der Mark [102], Ueda [100], and Tseng and Dugundji [98], it has taken the advent of the computer to truly identify it in real systems. Many of the early observations of chaotic behavior in electro-mechanical and thermal-fluid systems were incorrectly classified as *noise* and delayed its acceptance in the dynamics community as a separate class of system response [58]. From a vibrations and dynamics perspective, *chaos* is defined as an aperiodic and nonquasi-periodic response to a periodic stimulus of a deterministic nonlinear dynamic system. For clarity, let us examine a few of the key words in the above definition more closely:

- *Deterministic* implies that the equations of motion, parameters, and initial conditions of a dynamic system are known and are not stochastic (random). Theoretically, a set of deterministic equations which accurately model a dy-

namic system can be used to predict any future state if the original state is known exactly, i.e. we have infinite precision in the initial conditions.

- *Nonlinearity* is a property of a system or mathematical operation for which the output is not linearly proportional to the input.

- *Quasi-periodic* is a vibration motion consisting of two or more incommensurate frequencies, i.e. the ratio of two response frequencies is irrational.

A hallmark of chaotic motion is that the trajectories emanating from arbitrarily close initial conditions will diverge exponentially. Thus, chaotic systems are extremely sensitive to initial conditions. The motion, however, is bounded in a global sense and long term evolution of the trajectories produces what is known as a *chaotic attractor*. An attractor is by definition a subspace of the phase space toward which a time history approaches after transients die out [58]. Examples of chaotic attractors are shown in Fig. 2.1(a,b,d) in both two and three dimensions. The coordinates used in these plots are the state variables of the system. The number of coordinates is equivalent to the *embedding dimension* used to display or visualize the trajectories in the phase space. The trajectories of a chaotic attractor remain bounded, never to escape the volume by which they are encompassed. While there is a seemingly underlying structure in the pattern created by chaotic trajectories in the phase space, it is often difficult to distinguish any sense of order in the corresponding time series (Fig. 2.2). As a result, chaotic time series are often misconstrued as being stochastic in origin. The random element of chaos is that long term predictability is lost without infinite precision in the initial conditions. In physical experiments, one is **never** able to measure the initial state of the system with infinite precision. Thus, long term prediction of chaotic response for real systems is unrealistic.

With the discovery of chaos, scientists have had to invent signal processing techniques capable of describing the patterns created in phase space embeddings and/or
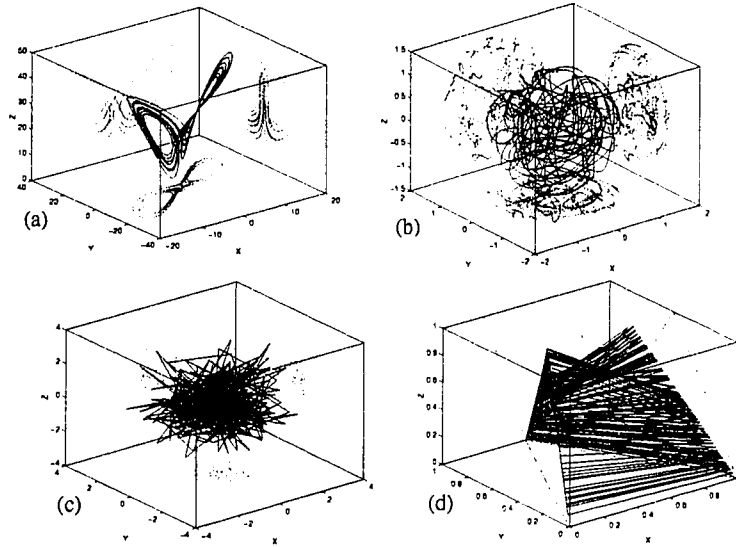
Figure 2.1: Phase space plots corresponding to the sequences shown in Figure 2.2.
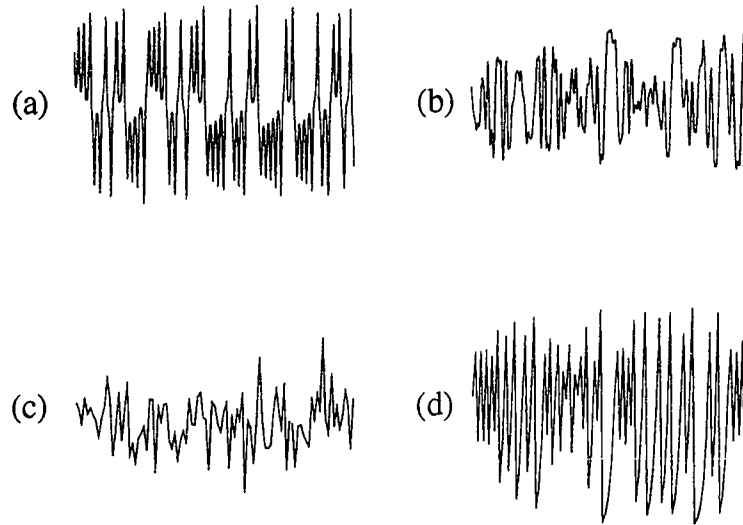
Figure 2.2: Chaotic sequences from the (a) Lorenz system, (b) Bender-Orzag system, (c) pseudorandom number generator, and (d) the tent map.

the dynamics of the trajectories which make up the patterns. Classic techniques, such as perturbation analysis, fail to characterize the dynamics of chaos because of an assumed periodicity of the solutions. Specifically, it is assumed that a periodic stimulus will result in a periodic or quasi-periodic response; quite the opposite of chaotic response to periodic stimulus. Thus, classic dynamic techniques that operate under this assumption are an inadequate means of investigating chaos. Novel methods are intended to quantify the invariant structure and flow of a chaotic attractor. Chaotic invariants fall into two main categories: dynamic and topological.

2.1.1 Dynamic Invariants: The Lyapunov Exponents

In assessing the dynamic invariants of an attractor, local flow divergence is measured and averaged across the attractor; yielding the nonlinear dynamic invariants known as the *Lyapunov exponents*. Lyapunov exponents can be well understood by first examining the phase space flow in a linear system. The eigenvalues for a linear system are invariants and represent the exponential rate of convergence or divergence of points on nearby orbits in the phase space. Consider a deterministic system described by the set of N first order equations

$$
\begin{aligned}
\dot{x}_1 &= f_1(x_1, x_2, \dots, x_N) \\
\dot{x}_2 &= f_2(x_1, x_2, \dots, x_N) \\
&\quad \cdot \\
&\quad \cdot \\
\dot{x}_N &= f_N(x_1, x_2, \dots, x_N)
\end{aligned}
\tag{2.1}
$$

The functions $f_1, \dots, f_N$ form a vector field **f**. The relative rate at which an N-dimensional volume V contracts or expands in the phase space is defined by the divergence of the vector field **f** [13]. A physical analogy is the growth or shrinkage of a three dimensional volume of compressible fluid whose change in volume is governed

by the divergence of a velocity vector. In a dynamic system defined by Eq. 2.1, the state derivatives $\dot{x}_i$ for $i = 1, \dots, N$ take the form of "phase space velocities" of an "N-dimensional fluid flow." The relative volume growth rate is related to the divergence of the vector field $\mathbf{f}$ by

$$\frac{\dot{V}(t)}{V(t)} = \nabla \mathbf{f} = \sum_{i=1}^{N} \frac{\partial f_i}{\partial x_i} \qquad (2.2)$$

where $V(t)$ is the phase space volume at time t evolved from an initial N-dimensional phase space volume V_0. If the system is linear, the solution of Eq. 2.2 is

$$V(t) = V_0 e^{(\sum \beta_i)t} \qquad (2.3)$$

where β_i are the system eigenvalues. For example, consider the system

$$\dot{x}_1 = -1.2x_1$$
$$\dot{x}_2 = x_2 \qquad (2.4)$$

The eigenvalues for this simple system are $\lambda_1 = -1.2$ and $\lambda_2 = 1$. A volume in the phase space defined by initial conditions set on the unit circle will contract in the x_1 direction and expand in the x_2 as time progresses (Fig. 2.3).

The Lyapunov exponents of a nonlinear chaotic system are analogous to the real parts of the eigenvalues for a linear system. For chaotic flows, one of the exponents must be positive, while the sum over the set of exponents must be negative. A positive Lyapunov exponent gives rise to a sensitivity to initial conditions; a hallmark of chaotic systems. A negative sum implies that the divergence of the vector field is negative and thus the system is accordingly dissipative. If it were not, unbounded trajectories would exist. It is also true that one of the exponents must be zero. A zero Lyapunov exponent can be seen by considering two test points on the attractor which are initially close to one another in time. While the distance between the orbits that evolve from these test points generally varies as a function of time, over long
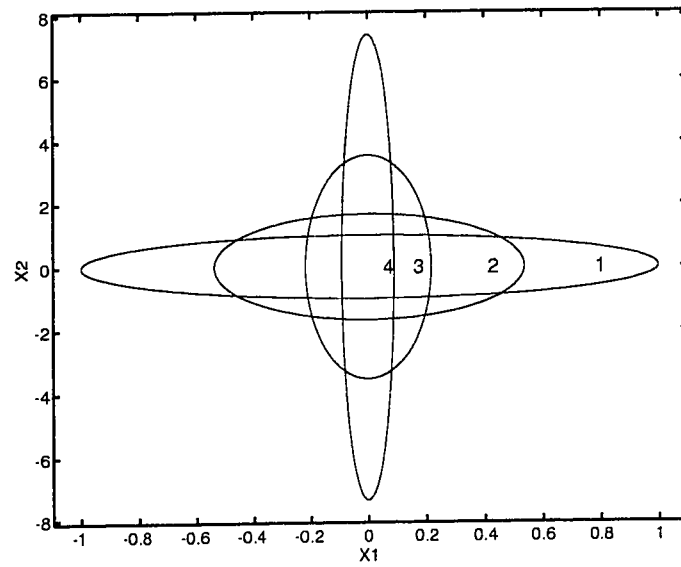
Figure 2.3: The evolution of a circular neighborhood of the unstable equilibrium at the origin of a linear system. The initial circle is denoted by marker 1, while snapshots of the forward evolution of the circle in evenly sampled time intervals are denoted by markers 2,3,4.

periods of time this separation will on average approach a constant value leading to a zero rate of exponential divergence (Lyapunov exponent approaches zero). Thus, a typical set of Lyapunov exponents for a third order chaotic system is $\{\lambda_1, 0, \lambda_3\}$ where $\lambda_1 > 0$ and $\lambda_3 < 0$ with $|\lambda_3| > \lambda_1$ since $\sum_i \lambda_i < 0$.

2.1.2 Topological Invariants: The Fractal Dimensions

A fractal dimension is by definition a quantitative property of a set of points in an n-dimensional space which measures the extent to which the points fill a subspace as the number of points becomes very large [58]. A fractal object is seen to be caught between Euclidean dimensions as it usually takes on a noninteger value. This section reviews the current techniques used to quantify the fractal dimension(s) of fractal objects.

Capacity Dimension

Perhaps the most intuitive of the methods used to estimate the dimension of a set of points is the *box counting* or *capacity* dimension. Indicative of its name, this method employs the use of "boxes" to cover a set of points in space. The boxes may in fact be balls, such that in one dimension they appear as a lines, in two dimensions as circles, in three dimensions as spheres, and in higher dimensions as *hyperspheres*. The essence of the method is to relate the minimum number of boxes N of length ε needed to cover the set as a function of ε. If a *uniform* distribution of N_0 points along some line or one-dimensional manifold is covered by boxes of size ε (Fig. 2.4), it is intuitive that the number of boxes to cover a line will scale as

$$N(\varepsilon) \propto \frac{1}{\varepsilon}.$$

This box coverage is referred to in the literature as *coarsegraining*. Similarly, if points are distributed uniformly on some two-dimensional manifold (Fig. 2.4), the
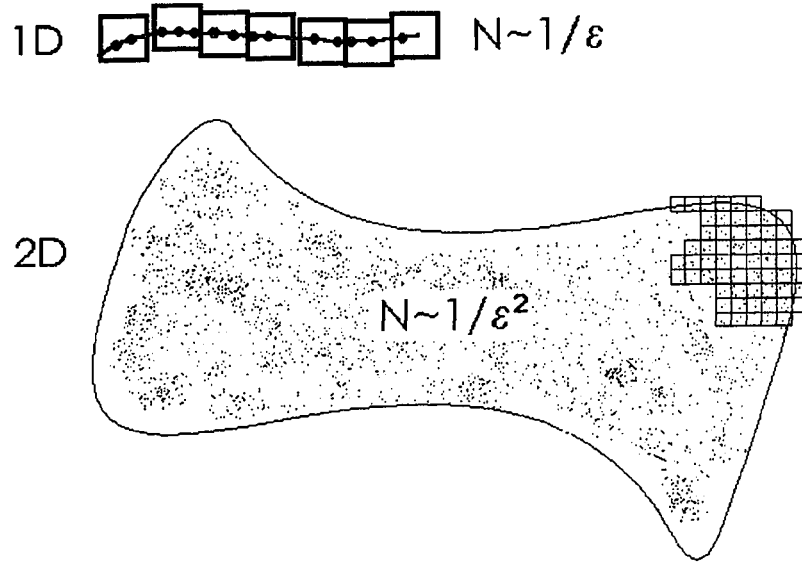
Figure 2.4: Coarsegrained coverage of a one and two dimensional distribution of $1/\varepsilon^d$ where d is the Euclidean dimension.

minimum number of boxes needed to cover the set scales as

$$N(\varepsilon) \propto \frac{1}{\varepsilon^2}.$$

Naturally, it follows to define the covering of a set of points residing on a d-dimensional manifold by the scaling law

$$N(\varepsilon) \propto \frac{1}{\varepsilon^d}. \tag{2.5}$$

Taking the logarithm of both sides of Eq. 2.5 and adding a subscript we obtain the definition of the *capacity dimension*

$$d_c = \lim_{\varepsilon \to 0} \frac{\log N(\varepsilon)}{\log (1/\varepsilon)}. \tag{2.6}$$

Implicit in this definition is the requirement that the number of points be large $(N_0 \to \infty)$. A set of points is said to be fractal if its dimension is noninteger, i.e. the scaling of the point distribution can be characterized by a *fractal dimension*.

Information Dimension

In an attempt to account for the frequency in which a trajectory visits the boxes of a coarsegrained partition, another fractal dimension was developed called the *information dimension*. It is assumed that the trajectories cover the phase space sufficiently to form an attractor whose dimension we wish to calculate.

To calculate the information dimension, the number of points N_i in each of the N boxes is counted and the probability of finding a point in that box is determined by

$$P_i \equiv \frac{N_i}{N_0}, \quad \sum_{i=1}^{N} P_i = 1 \tag{2.7}$$

where N_0 is the total number of points in the set. The *information entropy* is defined as

$$I(\varepsilon) = -\sum_{i=1}^{N} P_i \log P_i. \tag{2.8}$$

If the log function is base 2, $I(\varepsilon)$ has the units of bits. For small ε, I behaves as

$$I \approx d_I \log(1/\varepsilon),$$

so that for small ε we may define a dimension

$$d_I = \lim_{\varepsilon \to 0} \frac{I(\varepsilon)}{\log(1/\varepsilon)} = \lim_{\varepsilon \to 0} \frac{\sum_{i=1}^{N} P_i \log P_i}{\log \varepsilon}. \tag{2.9}$$

To see how d_I relates to d_c, suppose that the probabilities are equal for all boxes so that

$$P_i = \frac{N_i}{N_0} = \frac{1}{N}.$$

Then,

$$I = -\sum_{i=1}^{N} P_i \log P_i = -N P_i \log P_i = -\log(1/N) = \log N.$$

By substituting this expression into Eq. 2.9, we obtain

$$d_I = \lim_{\varepsilon \to 0} \frac{I(\varepsilon)}{\log(1/\varepsilon)} = \frac{\log N}{\log(1/\varepsilon)} = d_c. \tag{2.10}$$

Thus, the information dimension is equivalent to the capacity dimension when the probabilities are equal for all boxes covering the attractor, i.e. when the data is uniformly distributed in space. In general, it can be shown that [35]

$$d_I \le d_c.$$

The information entropy I is a measure of the *unpredictability* in a system. For a uniform probability distribution, $P_i = 1/N$, I is at a maximum. If all of the points are located in one box then we are at maximum predictability and the rate of information generation (entropy) is $I = 0$. This is seen by the calculations

$$P_i = P_1 = N_0/N_0 = 1$$

$$\begin{aligned} I &= -P_1 \log P_1 \\ &= -\log(1) \\ &= 0. \end{aligned}$$

Correlation Dimension

The correlation dimension was developed to estimate the joint probability of finding two points within a coarsegrained partition [32]. This has become the centerpiece of numerous published time series analyses. The correlation integral is defined as

$$C_m(r) = \frac{2}{N(N-1)} \sum_{i \ne j}^{N} \Theta(r - |\vec{\xi}_i - \vec{\xi}_j|) \tag{2.11}$$

where N is the number of embedded points, $\Theta(\cdot)$ is the Heaviside function, r is a Euclidean distance or scale, and $\vec{\xi}_k$ is a vector locating point k in D_e-space for

$k = 1, \ldots, N$ where D_e is the embedding dimension. If a single variable X is used to embed the data in D_e-space, $\vec{\xi}_k$ takes on the form

$$\vec{\xi}_k = (X_k, X_{k+\tau}, \ldots, X_{k+\tau(D_e-1)}) \tag{2.12}$$

where τ is an integer delay called the *time lag*. Similarly, if multiple states are used in the embedding of the data, $\vec{\xi}_k$ takes on the form

$$\vec{\xi}_k = (X_k, Y_k, Z_k, \ldots, \Psi_k) \tag{2.13}$$

where X, Y, Z, and Ψ are the state variables. The correlation dimension is defined as the slope of the correlation integrals on a $\log - \log$ scale

$$d_\nu = \lim_{r \to 0} \frac{d \log C_m(r)}{d \log r}. \tag{2.14}$$

where $d(\cdot)$ is the first derivative operator.

2.2 The Relation of D_q to the Capacity, Information, and Correlation Dimensions

For a nonuniform distribution of points in the phase space, a single fractal dimension is an incomplete description. To completely characterize a nonuniform topology, an entire spectrum of dimensions, known as the *generalized fractal dimensions* D_q is required. D_q is defined as

$$D_q \equiv \frac{1}{(q-1)} \lim_{\varepsilon \to 0} \frac{\ln \sum_{i=1}^{N(\varepsilon)} P_i^q}{\ln \varepsilon} \tag{2.15}$$

where as before P_i is the probability of the ith box in a phase space coarsegraining and $q \in \mathbb{R}$. The natural logarithm (base e) is used here, but any base is acceptable. In the following analysis, the relations between D_q and the capacity, information, and correlation dimensions are derived.

2.2.1 D_0 and the Capacity Dimension

Using $q = 0$ in Eq. 2.15, D_0 becomes

$$
\begin{aligned}
D_0 &= \lim_{\varepsilon \to 0} \frac{-\ln N(\varepsilon)}{\ln \varepsilon} \\
&= \lim_{\varepsilon \to 0} \frac{\ln N(\varepsilon)}{\ln(1/\varepsilon)}
\end{aligned}
\tag{2.16}
$$

and we find that D_0 is equivalent to the capacity dimension d_c.

2.2.2 D_1 and the Information Dimension

If $q \to 1$ and $\varepsilon \to 0$, Eq. 2.15 is ill defined. We thus take an alternative route to derive the relation between D_1 and d_I [55]. To begin, let $\mathcal{Z}$ be a partition function defined as

$$
\mathcal{Z}(q,\varepsilon) \equiv \sum_{i=1}^{N(\varepsilon)} P_i^q
\tag{2.17}
$$

where $\mathcal{Z}(q,\varepsilon) \sim \varepsilon^{(q-1)D_q}$ based on Eq. 2.15. Now,

$$
\ln \mathcal{Z}(q,\varepsilon) = \ln \sum_{i=1}^{N(\varepsilon)} P_i^q = \ln \left(\sum_{i=1}^{N(\varepsilon)} P_i^q - 1 + 1 \right)
\tag{2.18}
$$

The natural logarithm function can be approximated in a Taylor series expansion by

$$
\ln(x + 1) = x - \frac{x^2}{2!} + \frac{x^3}{3!} - \cdots
\tag{2.19}
$$

Since $\sum_i P_i = 1$, $\left(\sum_i P_i^q - 1 \right) \ll 1$ as $q \to 1$, and thus the terms $\left(\sum_i P_i^q - 1 \right)^k$ for $k > 1$ can be neglected in a Taylor series expansion of Eq. 2.18 with $x = \left(\sum_i P_i^q - 1 \right)$. Equation 2.18 is therefore approximated by the relation

$$
\ln \sum_{i=1}^{N(\varepsilon)} P_i^q \approx \sum_{i=1}^{N(\varepsilon)} P_i^q - 1
\tag{2.20}
$$

for $q \approx 1$. Now assume that $q = 1 + \delta q$ where $\delta q \ll 1$. Then

$$P_i^q = P_i^{1+\delta q} \simeq P_i e^{\delta q \ln P_i}. \tag{2.21}$$

The natural exponential function can also be expanded in a Taylor series as

$$e^y = 1 + y + \frac{y^2}{2!} + \dots \tag{2.22}$$

so that

$$e^{\delta q \ln P_i} = 1 + \delta q \ln P_i + \frac{(\delta q \ln P_i)^2}{2!} + \dots \approx 1 + \delta q \ln P_i \tag{2.23}$$

since $\delta q \ll 1$. Equation 2.21 can now be approximated as

$$P_i^q \approx P_i + P_i \delta q \ln P_i. \tag{2.24}$$

Substituting Eq. 2.24 into Eq. 2.20 we get

$$\ln \sum_{i=1}^{N(\varepsilon)} P_i^q \approx \sum_{i=1}^{N(\varepsilon)} P_i + \delta q \sum_{i=1}^{N(\varepsilon)} P_i \ln P_i - 1 \approx \delta q \sum_{i=1}^{N(\varepsilon)} P_i \ln P_i \tag{2.25}$$

since $\sum_{i=1}^{N(\varepsilon)} P_i = 1$. Rearranging Eq. 2.15 and substituting $q = 1 + \delta q$ we obtain

$$\delta q D_1 \ln(\varepsilon) \sim \ln \sum_{i=1}^{N(\varepsilon)} P_i^q \sim \delta q \sum_{i=1}^{N(\varepsilon)} P_i \ln P_i. \tag{2.26}$$

Finally, dividing through by $\delta q \ln(\varepsilon)$ reduces Eq. 2.26 to

$$D_1 = \lim_{\varepsilon \to 0} \frac{\sum_{i=1}^{N(\varepsilon)} P_i \ln P_i}{\ln(\varepsilon)}, \tag{2.27}$$

and we conclude that D_1 is equivalent to the information dimension d_I.

2.2.3 D_2 and the Correlation Dimension

The correlation integrals as defined in Eq. 2.11 represent the joint probability of finding any two points $\vec{\xi_i}$ and $\vec{\xi_j}$ ($i \neq j$) separated by a Euclidean distance less than

scale ϵ in the phase space. For the very special case where the points are *statistically independent*, $\sum_{i=1}^{N(\varepsilon)} P_i^2$ is also the probability that any two points fall into a box of size ε so

$$C(\varepsilon) = \sum_{i=1}^{N(\varepsilon)} P_i^2. \tag{2.28}$$

Using $q = 2$ in Eq. 2.15, D_2 becomes

$$D_2 = \lim_{\varepsilon \to 0} \frac{\ln\left(\sum_{i=1}^{N(\varepsilon)} P_i^2\right)}{\ln \varepsilon} \tag{2.29}$$

If the analysis is performed over statistically independent points in the phase space, then

$$D_2 = \lim_{\varepsilon \to 0} \frac{\ln C(\varepsilon)}{\ln \varepsilon} \tag{2.30}$$

and we find that D_2 is equivalent to the correlation dimension d_ν.

2.3 Examples of Chaotic Flow and Chaotic Maps

The following systems are examples of nonlinear ordinary differential equations capable of producing chaotic response. Each system is chaotic because the motion is bounded, the systems are nonlinear deterministic, and the phase space flow is characterized by at least one positive Lyapunov exponent with the appropriate selection of system parameters. A chaotic parameter space is defined after the equations of motion for each system. These chaotic parameter sets are not unique.

Lorenz system:

$$\frac{dx}{dt} = \sigma(y - x)$$

$$\frac{dy}{dt} = rx - y - xz \qquad (2.31)$$

$$\frac{dz}{dt} = xy - bz$$

where $\sigma = 10, r = 28.0$, and $b = 8/3$ [48].

Rössler system:

$$\frac{dx}{dt} = -z - y$$

$$\frac{dy}{dt} = x + ay \qquad (2.32)$$

$$\frac{dz}{dt} = bx + z(x - c)$$

where $a = 0.36, b = 0.45$, and $c = 4.5$ [82]. Figures 2.5 and 2.6 show the chaotic attractors for the Lorenz and Rössler system, respectively, for $D_e = 2$ and 3. The third example is a coupled Van der Pol oscillator system (Fig. 2.7). In each plot, the three dimensional attractor is projected onto two dimensional planes which are represented by the "walls" of the box encompassing the attractor. Each projection is used here for visualization purposes only. In general, embedding the data in different dimensions is used to help quantify the structure of the attractor (see Section 2.1.2).

Coupled Van der Pol Oscillator:

$$\ddot{x} - \varepsilon(1 - x^2)\dot{x} = \varepsilon A(y - x) + \varepsilon B(\dot{y} - \dot{x})$$

$$\tag{2.33}$$

$$\ddot{y} - \varepsilon(1 - y^2)\dot{y} = \varepsilon A(x - y) + \varepsilon B(\dot{x} - \dot{y})$$

where $\varepsilon A = 2.0305$ and $\varepsilon B = -2.5$ [80].
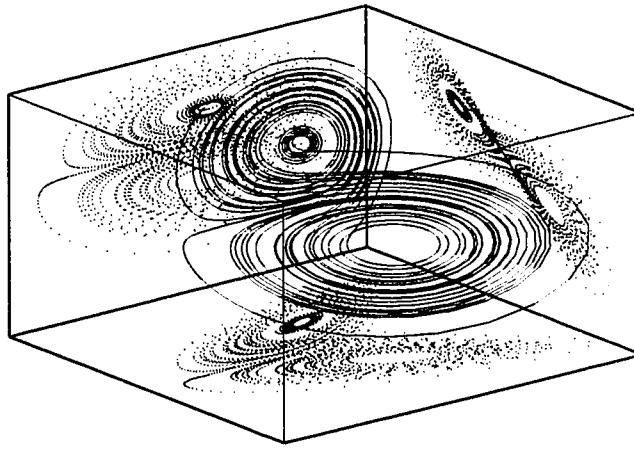
Figure 2.5: The Lorenz attractor.

Another means of creating chaos comes in the form of iterated maps. There are an infinite number of chaotic attractors that can be created with two dimensional iterated quadratic maps in the form:

$$x_{n+1} = a_1 + a_2 x_n + a_3 x_n^2 + a_4 x_n y_n + a_5 y_n + a_6 y_n^2$$

$$y_{n+1} = a_7 + a_8 x_n + a_9 x_n^2 + a_{10} x_n y_n + a_{11} y_n + a_{12} y_n^2.$$

$$\tag{2.34}$$
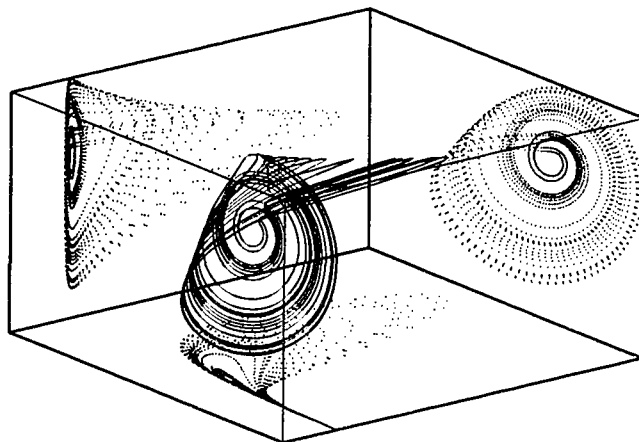
Figure 2.6: The Rössler attractor.
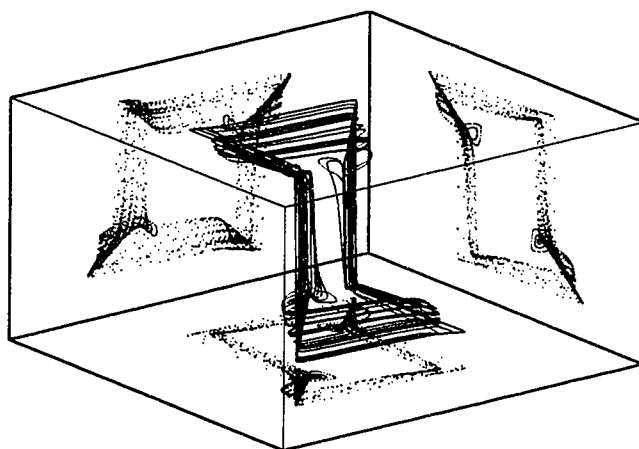
Figure 2.7: A coupled Van der Pol oscillator.
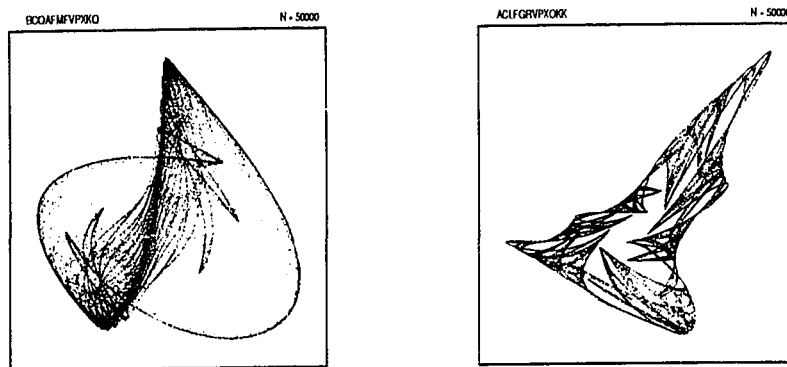
Figure 2.8: The Sea Shell and Tattered Iceberg: chaotic attractors produced by quadratic iterative maps. The letters denote the values assigned to the coefficients $a_1, a_2, \ldots, a_{12}$, where $A = -1.2, B = -1.1, \ldots, Z = 1.3$, and help to easily identify and classify a particular attractor. The abscissa and the ordinate are defined by the coordinates x_n and y_n, respectively.

Given the appropriate combination of coefficients, these maps yield chaotic attractors. They are attractors because a range of starting values x_0 and y_0, within the *basin of attraction*, yield the same eventual pattern. They are chaotic because each has at least one positive Lyapunov exponent. The time histories of the maps appear random while the structure created in the phase space has an underlying sense of order or pattern. Figure 2.8 shows two examples of chaotic attractors developed by iterating Eq. 2.34.

It is possible to generate chaos using one dimensional maps such as the (symmetric) tent map defined by

$$f(z_i) = z_{i+1} = \begin{cases} 2z_i, & 0 < z_i < 0.5 \\ 2(1 - z_i), & 0.5 \leq z_i < 1.0. \end{cases} \tag{2.35}$$

The dynamics of the symmetric tent map (Fig. 2.9) can be easily understood in
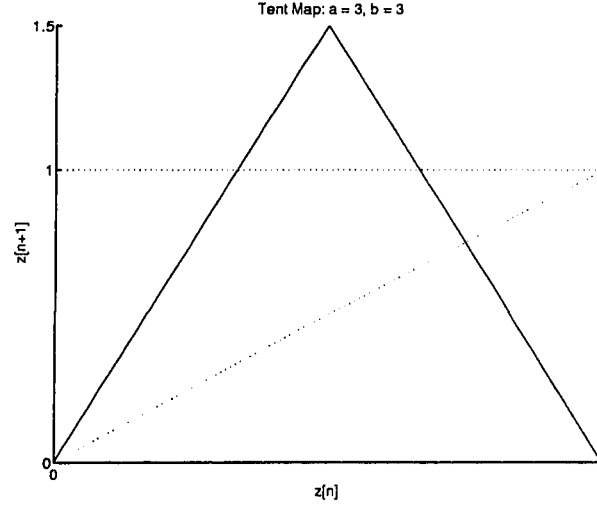
Figure 2.9: The symmetric tent map. The absolute value of the left and right face slopes is specified as a and b, respectively.

base-2 arithmetic where all parameters and numbers are written as binary strings,

$$z_n = \sum_{j=1}^{\infty} \varepsilon_i(n)/2^j = 0.\varepsilon_1(n)\varepsilon_2(n)\ldots\varepsilon_N(n)\ldots , \qquad (2.36)$$

where $\varepsilon_i(n) = 0$ or 1. The map is exactly equivalent to a binary shift map defined by the simple automaton

$$\varepsilon_i(n+1) = \begin{cases} \varepsilon_{i+1}(n), & \varepsilon_1(n) = 0 \\ 1 - \varepsilon_{i+1}(n), & \varepsilon_1(n) = 1 \end{cases} \qquad (2.37)$$

so that the initial conditions that are written as finite binary strings

$$z_0 = \sum_{j=1}^{N} \varepsilon_i(0)/2^j = 0.\varepsilon_1(0)\varepsilon_2(0)\ldots\varepsilon_N(0) \qquad (2.38)$$

truncate to zero in finite time $n = N$. As an example, the first ten iterates of the symmetric tent map with $z_0 = \sqrt{2} - 1$ as written in binary form follow exactly the automaton as described by Eq. 2.37:

$$0.0110101000001001$$
$$0.1101010000010011$$
$$0.0101011111011000$$
$$0.1010111110110000$$
$$0.1010000010011110$$
$$0.1011111011000011$$
$$0.1000001001111001$$
$$0.1111101100001100$$
$$0.0000100111100110$$
$$0.0001001111001100$$

The dynamics of the map can be understood as follows: in z_0, the information stored in any bit $\varepsilon_j(0)$ propagates to the left in the bit stream as the map is iterated [55]. After j iterations, $\varepsilon_j(0)$ becomes the most significant bit $\varepsilon_1(0)$ in the stream. Based on this type of analysis, it is easy to see how the tent map is very sensitive to initial conditions as least significant bits soon become the most significant in the bit stream. This flow of information is fundamental in chaotic systems. The tent map is chaotic because there is information flow from right to left in the bit stream as the map is iterated forward at the rate of one bit per iteration. Defining the Lyapunov exponent to be the exponential rate of divergence between two nearby orbits, i.e.

$$|\delta z_n| \sim |\delta z_0| e^{\lambda n} \quad n \gg 1, \tag{2.39}$$

where δz_n is the orbit separation after n forward iterations and δz_0 is the initial separation implicitly assumed to be very small. The result is a positive Lyapunov exponent $\lambda = \ln(2)$ for the symmetric tent map. This can proven by computing exactly the orbit of the symmetric tent map $z_0 \to z_1 \to z_2 \dots$, starting from some initial condition z_0, and linearizing about this exact orbit for some initial condition $z_0 + \delta z_0$:

$$z_n + \delta z_n = f(z_{n-1} + \delta z_{n-1}) \approx f(z_{n-1}) + f'(z_{n-1})\delta z_{n-1} + \dots \tag{2.40}$$

26

We then obtain the linear approximation to the separation of nearby orbits

$$\delta z_n \simeq f'(z_{n-1})\delta z_{n-1},$$

(2.41)

which can be solved exactly to obtain

$$\delta z_n \cong \prod_{i=0}^{n-1} f'(z_i)\delta z_0,$$

(2.42)

where δz_0 is the initial separation of the two orbits. With

$$|\delta z_n| \approx \sum_{i=0}^{n-1} |f'(z_i)||\delta z_0|$$

(2.43)

the Lyapunov exponent for an iterative map $z_{i+1} \to f(z_i)$ is defined as

$$\lambda = \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \ln |f'(z_i)|.$$

(2.44)

Since the slope of the tent map shown in Fig. 2.9 is $|f'(z_n)| = 3$, the Lyapunov exponent becomes $\lambda = 3$ by way of Eq. 2.44. If the slope $|f'(z_i)| > 1$, λ is positive and the map is chaotic.

The tent-like structure can also be found in reduced maps of chaotic flow in continuous systems. For example, by extracting the local extrema from the z state of the Lorenz system (Eq. 2.32), a tent map can be produced by plotting the extracted series in *return map* as shown in Fig. 2.10. One dimensional maps that represent bounded chaotic behavior cannot provide more than partial information about higher dimensional attractors [55]. However, a reduction in dimension via sampling schemes does not seem to affect the interpretation of the resulting motion, i.e. the sampled flows produce maps which indicate that chaos is present because a positive Lyapunov exponent exists while the motion remains bounded. These ideas are the basis for the use of Poincaré maps in analyzing chaotic systems.
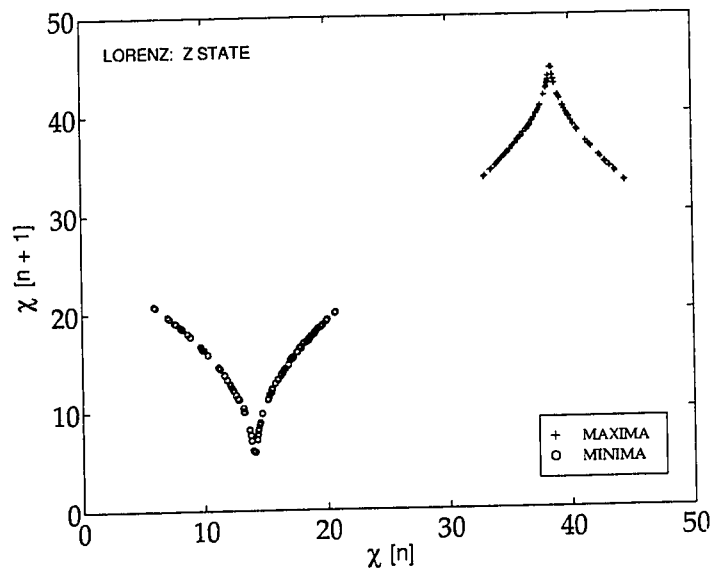
Figure 2.10: A return map of the local extrema of the z state in the Lorenz system. The result is a tent map which seems to be at the heart of many chaotic systems. The separation between the two maps implies that no local extrema exist in the approximate range [22, 32].

2.3.1 Poincaré Maps

If the system being measured is of the form Eq. 2.1, the states can be stroboscopically sampled to produce what is known as a *Poincaré map* of the original attractor. A Poincaré map is visualized by taking slices or *Poincaré sections* of the attractor such that the intersection of the orbits with the Poincaré section produces a mapping of the original attractor. Since we are, in effect, taking a slice out of the original attractor, the resulting attractor dimension is decreased by one. Another means of generating a Poincaré map is to sample specific locations in a time series. For example, given a series of measurements, one could sample the series by keeping only the local extrema. Figure 2.11 demonstrates this procedure on the X state of the Lorenz system.
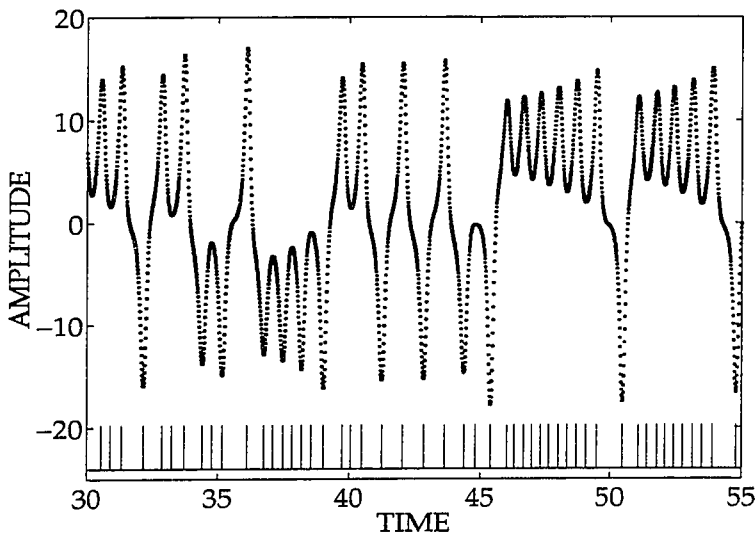
Figure 2.11: Sampling of the local extrema of the X state of the Lorenz system. The spike train denotes the sampling locations in time.

2.3.2 Interspike Intervals

Maps are often encountered in the analysis of measured biological signals such as electrocardiograms (ECGs). For example, the difference in the temporal locations of successive depolarization peaks (or R-waves) in the ECG (Fig. 2.12) are used to form a special type of sequence known as *RR intervals*. Mathematically speaking, RR intervals are a form of *interspike intervals (ISIs)* (Fig. 2.13). If the heart is considered to be a source of (cardiac) dynamics, RR intervals form a type of Poincaré mapping of the ECG signal. From these maps, one can hope to estimate chaotic invariants. The process by which interspike intervals are formed is also illustrated in Fig. 2.11 by calculating successive differences between local extrema in the X state of the Lorenz system (Eq. 2.32).
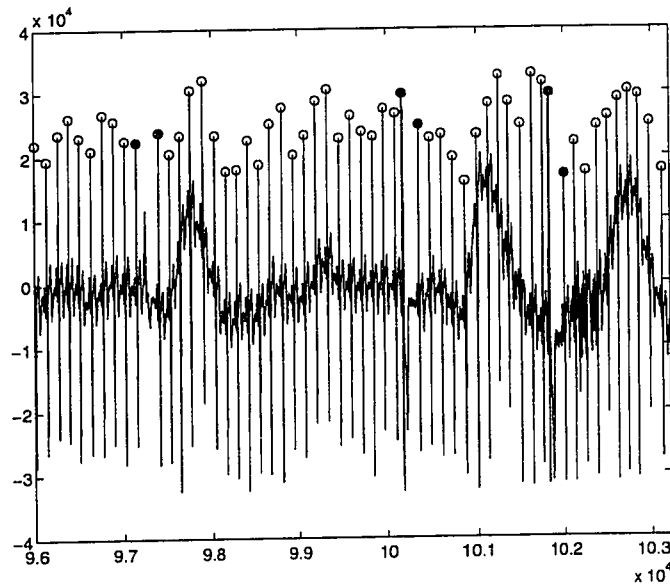
Figure 2.12: Sample electrocardiogram signal. The R-wave peaks are denoted by both filled and hollow circles.
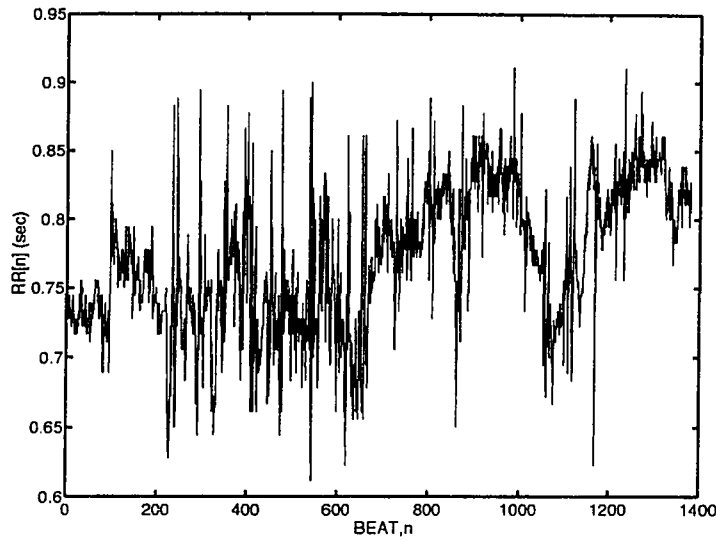
Figure 2.13: Sample ECG RR data. RR data are a form of interspike intervals.

2.4 Description of Fractals

The word *fractal* was coined by Benoit Mandelbrot in 1975 to describe an intricate set of curves, a majority of which had never been seen until the advent of the computer. Many of these curves, or fractals, are strikingly similar to shapes found in nature. Natural objects tend to have irregularities over a range of scale that cannot be adequately described by Euclidean geometry. Indeed, examination of Euclidean shapes (such as lines, circles, cones, and paraboloids) at smaller and smaller scales, reveals structures that resemble a line. i.e. structures that are smooth. Mandelbrot approached the issue of modeling natural shapes from a recursive process perspective rather than from Euclidean form and found that complicated shapes could be created with very simple recursive maps. In effect, the simplicity of the maps used to create complex shapes adds a sense of elegance to the method. Slight variations in the

change of the parameters used in the mapping functions can create vastly different fractal structures. Thus, a simple set of equations can be used to create an infinite set of fractal curves, simply by changing a few parameters. The excitement in this field is that many physical and biological processes are speculated to produce fractal structures including eddy cascades in turbulent flow, blood vessel branching, and coastline lengths to name a few [7].

Fractals are *self-similar*. Loosely interpreted, self-similarity implies that the information perceived at one scale is equivalent to the information perceived at other scales. The term *information* as it is used here may mean the shape of an object, or it may refer to a measure of an object such as surface area or length. Thus, fractals exhibit constant detail or bumpiness with an increase in magnification. One of the most popular examples of a self similar object is the Mandelbrot set produced by the iterative map [52]

$$z_{n+1} = z_n^2 + c \tag{2.45}$$

where z and c are complex variables of the form $z = x + iy$ and $c = a + ib$. In terms of real variables, the Mandelbrot map becomes

$$x_{n+1} = x_n^2 - y_n^2 + a$$

$$\tag{2.46}$$

$$y_{n+1} = 2x_n y_n + b.$$

Maps of this type can be generalized as

$$x_{n+1} = f(x_n, y_n)$$

$$\tag{2.47}$$

$$y_{n+1} = g(x_n, y_n).$$

In the case of the complex map (Eq. 2.45), $F = f + ig$ is an analytic function of

z. That is, the derivative $\frac{dF(z)}{dz}$ exists and the functions $f(x,y)$ and $g(x,y)$ satisfy

$$\frac{\partial f}{\partial x} = \frac{\partial g}{\partial y}$$

$$\frac{\partial f}{\partial y} = -\frac{\partial g}{\partial x}.$$

(2.48)

Given an appropriate choice for c, the Mandelbrot map can produce a fractal structure by means of an *escape plot* as shown in Figure 2.14. An escape plot is produced by color coding the number of iterates it takes for a given initial condition to either escape towards infinity or remain bounded. For computational efficiency, a finite number of iterations N is performed on the map and the resulting magnitude $|z_N|$ is compared to an arbitrarily selected large value z_{ROC}. If $|z_N| > z_{ROC}$, future iterates of the map are assumed to blow up or escape to infinity and the location of the initial condition in the complex plane is color coded accordingly. The light grey areas in Fig. 2.14 represent the locations of initial conditions that escape in this way. The white areas represent the initial conditions which are forever bounded or caught in a so-called *basin of attraction* and never exceed z_{ROC}. The boundary which separates the unstable and stable set of initial conditions forms a beautiful and highly fragmented curve, represented in Fig. 2.14 with darker shades of grey leading to black. This interface is fractal in the sense that it is self similar: if one zooms in on one portion of the escape plot, it resembles exactly the area from which was produced as seen at a lower resolution. Theoretically, no matter how far one zooms into the interface, one will always see a high degree of fragmentation and complexity. While such maps serve as a visual paradigm in regard to the intimate connection between dynamical systems and fractals, the physical relevance is still to be discovered [58].

Beyond the Mandelbrot set, some of the most popular examples of fractals are the Koch curve, the Cantor set, the devil's staircase, and the Sierpinski triangle. Each is produced by a simple geometric rule. For example, consider the following rule for a Sierpinski triangle: for every solid triangle, divide it up into 4 equi-sized
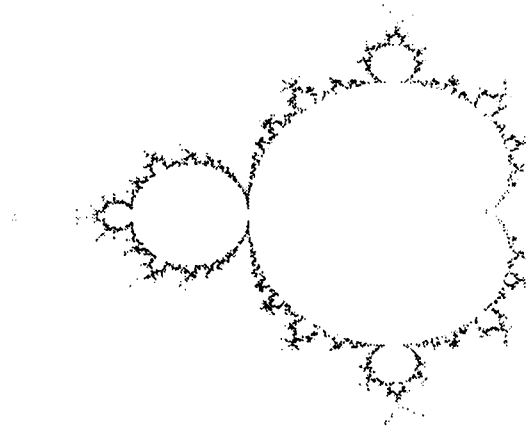
Figure 2.14: Mandelbrot set for the function $z_{n+1} = z_n^2 + c$ plotted in the complex plane.

triangles and remove the middle triangle. Thus for each iteration n of the Sierpinski triangle set, there are $N = 3^n$ segments of size $\varepsilon = 2^{-n}$ (the length of a side in the child triangle is $1/2$ that of the corresponding leg in the parent triangle). The result after many iterations is the Sierpinski triangle shown in Fig. 2.15. To create the Koch curve a geometric pattern is iterated on a line with the following rule: for each line of length L, take out the middle third of the line and replace it with two lines each of length $L/3$ as shown in Figure 2.16. Thus, for each level n in the iteration process, there are $N = 4^n$ segments of size $\varepsilon = 3^{-n}$. To generate a fractal curve, the iterations are repeated forever. The Cantor set is created in a similar fashion to the Koch curve. If the Koch curve can be considered a process of adding finer and finer length structure to an initial line segment, then the Cantor set is the complement operation of removing smaller and smaller segments from a set of points initially on a line. The rule for the Cantor set is: For each line take out the middle third

34
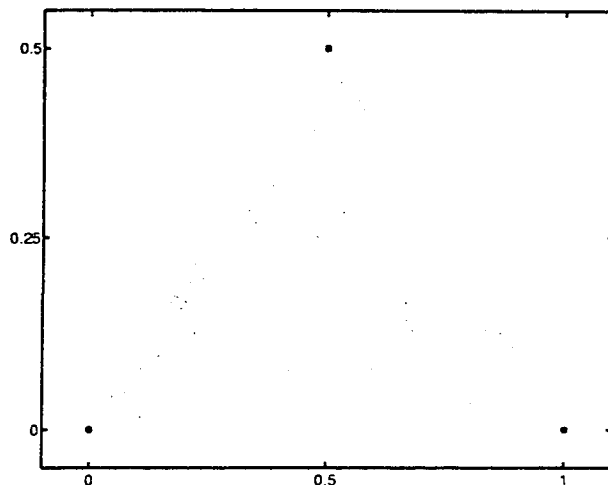
Figure 2.15: The Sierpinski triangle.

segment as shown in Figure 2.16. Thus for each iteration n of the Cantor set. there are $N = 2^n$ segments of size $\varepsilon = 3^{-n}$. The discontinuous Cantor set can be used to generate a continuous distribution function by integrating over an appropriate density function defined on the set. Suppose. for example. that mass μ_0 is uniformly distributed on the interval $0 \leq x \leq 1$ of length ε_0 such that the total mass is unity (Fig. 2.17). At each iteration level. the mass is again uniformly distributed from the previous level ($\mu_1 = \mu_2$) and the mass density. $\rho = \mu/\varepsilon$. increases for each segment of the set. Specifically. at the nth iteration level. the mass density of each interval is $\rho = (3/2)^n$. By integrating the mass density along x. we obtain the mass as a function of x:

$$f_n(x) = \int_0^x \rho_n(x)\, dx \tag{2.49}$$

In the limit as $n \to \infty$. $f_n(x)$ is called the devil's staircase for its infinite number of steps (Fig. 2.18). Realistically. there is no density function that can be associated
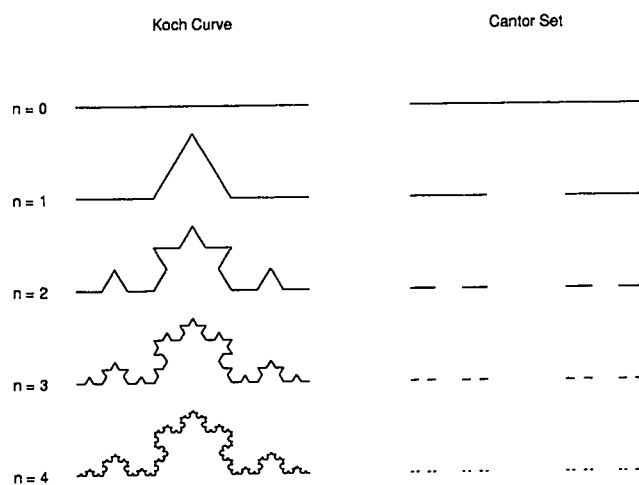
Figure 2.16: The Koch curve and the Cantor set for $n = 4$ iterations. The Cantor set can be built graphically by not displaying parts of the Koch curve which are greater than zero on the ordinate. Thus the areas where the Koch curve meets the abscissa defines the Cantor set.

with the Cantor set since the support[1] of the set approaches zero as $n \to \infty$. It is reasonable, however, to associate a distribution function $f_n(x)$ with the Cantor set. Thus far, we have arbitrarily designed both the support and the distribution in the

$n = 0$ $\varepsilon 0$ $\mu 0 = 1$

$n = 1$ $\varepsilon 1$ $\mu 1$ $\varepsilon 1$ $\mu 2$

$n = 2$ $\varepsilon 2$ $\mu 1 \mu 1$ $\varepsilon 2$ $\mu 1 \mu 2$ $\varepsilon 2$ $\mu 2 \mu 1$ $\varepsilon 2$ $\mu 2 \mu 2$
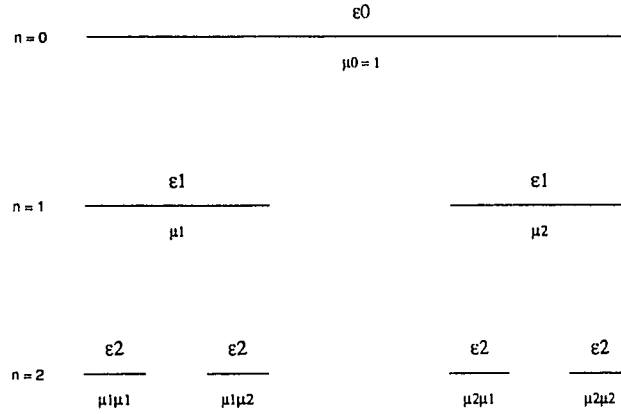
Figure 2.17: Propagation of mass μ and fundamental length scale(s) ε in the Cantor set. Here, the length scale for each iteration level is the same. An alteration in the length scales at $n = 1$ such that the left side has length ε_{11} and the right ε_{12} for $\varepsilon_{11} \neq \varepsilon_{12}$, will produce a nonuniform support. If $\mu_1 \neq \mu_2$, the mass distribution is nonuniform. In both cases, the result is the formation of a multifractal structure.

Cantor set to be uniform or *homogeneous*. If, however, we choose to distribute the mass in a nonuniform fashion ($\mu_1 \neq \mu_2$) or to develop the support of the Cantor set via the propagation of two unique length scales (ε differs for each segment in level n), then the resulting Cantor set is *multifractal*. Figures 2.19, 2.20, and 2.21 demonstrate various of combinations of (non)homogeneous distribution and support for the Cantor set. The homogeneous Cantor set with corresponding devil's staircase, Koch curve, and Sierpinski triangle are fractal sets because at any iteration

---

[1]The *support* of a set is where it is nonzero.
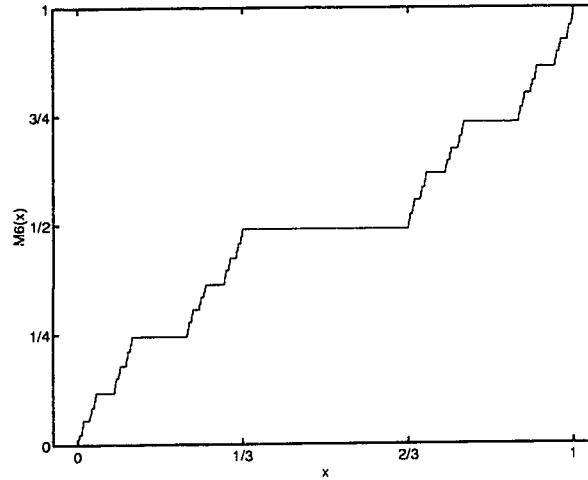
Figure 2.18: Homogeneous devil's staircase distribution function.

level one may find an exact replica of the whole. As an example, zooming in on the first 1/3 of the support of the homogeneous devil's staircase (Fig. 2.18) results in a curve that is an exact replica of the prezoomed image on the whole of its support. Typically, the zoom must be performed *anisotropically* for an exact replica to be achieved. Anisotropic zooming is achieved by dilating a portion of an object more in one direction than in the other. Mathematically, this relation is summarized as

$$f(\lambda x) = \lambda^H f(x). \tag{2.50}$$

An object or signal is said to be *self affine* if an anisotropic dilation of a portion is geometrically or statistically similar to the whole. In the homogeneous devil's staircase, the first third must be stretched by a factor of three to span the same support, and then it must be stretched along the direction of the ordinate by a factor of two to achieve the equivalent amplitude of the original curve and $H = \log 2/\log 3$ in Eq. 2.50.
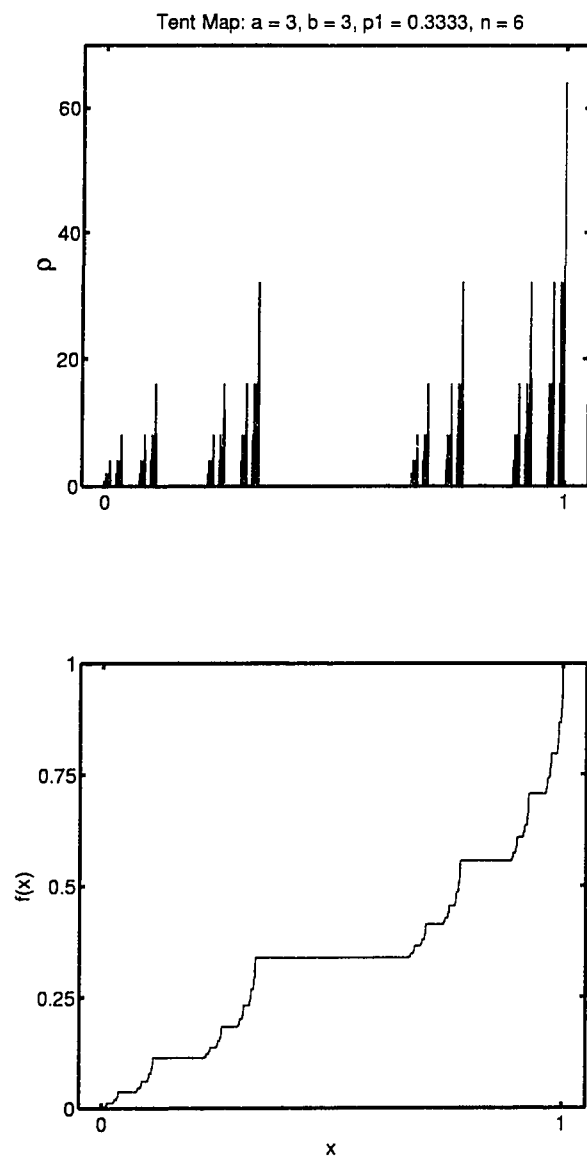
Figure 2.19: Multifractal Cantor set mass density and distribution with nonhomogeneous distribution/homogeneous support: $\mu_1 = 1/3, \mu_2 = 2/3, \varepsilon_n = 3^{-n}$, and $n = 6$.
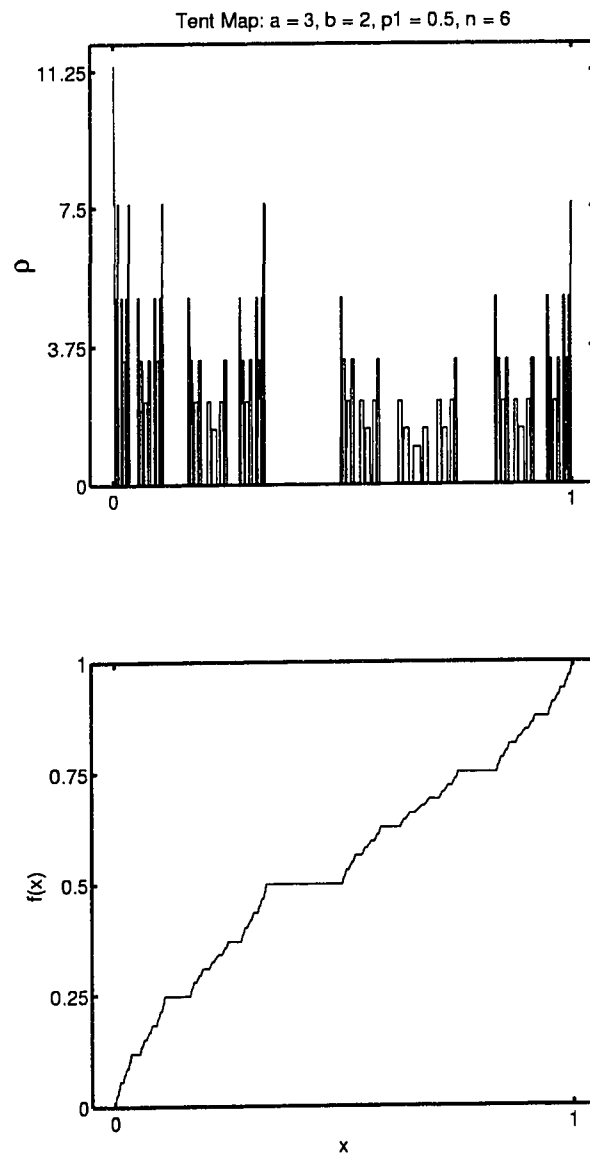
Figure 2.20: Multifractal Cantor set mass density and distribution with homogeneous distribution/nonhomogeneous support: $\mu_n = 2^{-n}, \varepsilon_{11} = 2/5, \varepsilon_{12} = 3/5$, and $n = 6$.
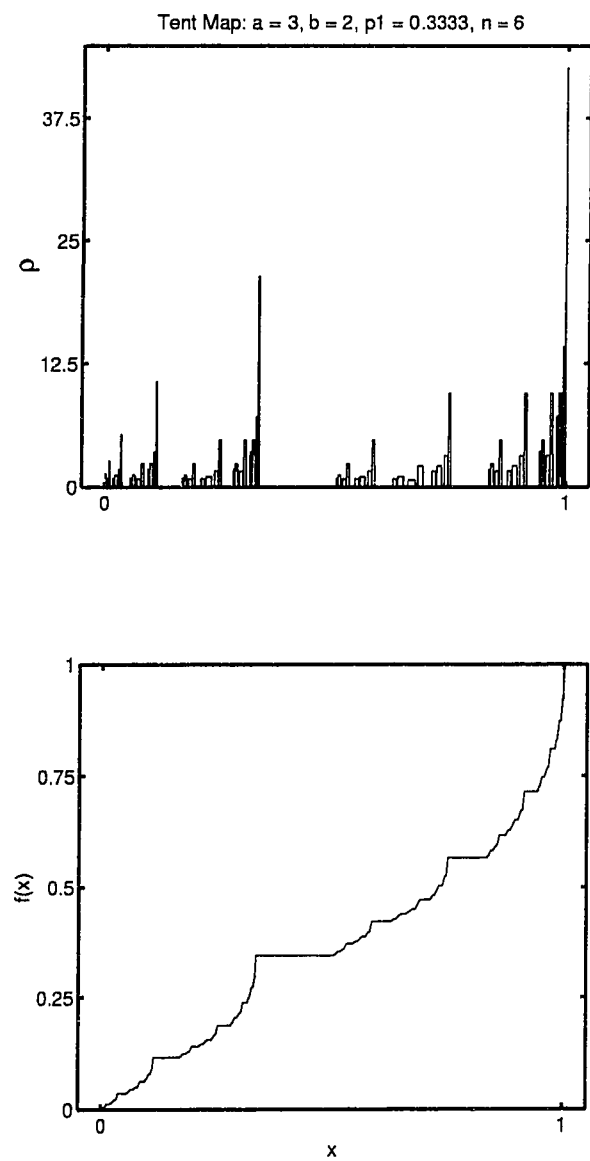
Figure 2.21: Multifractal Cantor set mass density and distribution with nonhomogeneous distribution/nonhomogeneous support: $\mu_1 = 1/3, \mu_2 = 2/3, \varepsilon_{11} = 2/5, \varepsilon_{12} = 3/5$, and $n = 6$.

In cases where the support and/or distribution of the Cantor set is nonuniform there is no one value of H that satisfies Eq. 2.50. This is primarily due to the fact that multifractal structures generally do not contain exact replicas of the whole at smaller scales. Rather, they resemble the whole in a statistical fashion. As a result, an entire spectrum of exponents H are needed to describe multifractals. We shall thoroughly discuss this issue in subsequent sections.

Multifractal Analysis

Multifractals arise when a measure defined on the support has different fractal dimensions on different parts of the support. Such phenomena cannot adequately be described by a single fractal dimension. Rather, a complete spectrum of dimensions characterizes multifractal structure. The calculation of multifractal spectra is approached from two basic perspectives: α-based and q-based methods. In what follows, each of these methods is developed and a relation between the two is formed [62]. For clarity, the techniques will be demonstrated on Cantor sets, although they are not limited to simple pedagogical examples of multifractal geometry.

Consider the geometrically self-similar triadic Cantor set with uniform distribution ($\mu_1 = \mu_2, \mu_1 + \mu_2 = 1$) as shown in Fig. 2.17. At each level of the level of the set we see that there are $N_n = 2^n$ segments of (fundamental) length $\varepsilon_n = 3^{-n}$. Using the definition of the capacity dimension (Eq. 2.6) it is easily shown that

$$
\begin{aligned}
N_n &= \varepsilon_n^{-D_0} \\
\mu_n &= \mu_0 N_n^{-1} = \varepsilon_n^{D_0}
\end{aligned}
\tag{2.51}
$$

Thus far, we have made two assumptions:

1. The support is uniform

2. The distribution in terms of some measure is uniform

If the above assumptions hold, then only one fractal dimension is needed to characterize the set. If, however, one or both of these is not upheld, then an entire spectrum of fractal dimensions is needed to characterize the set. Based on the possibility of inhomogeneity in the Cantor set, the results found in Eq. 2.51 are generalized as

$$N_n(\alpha) = \varepsilon_n(\alpha)^{-f(\alpha)}$$
$$\mu_{i,n} = \varepsilon_{i,n}{}^{\alpha_{i,n}} \tag{2.52}$$

where $\mu_{i,n}$ is the measure found at the ith segment of the nth iteration level, $\varepsilon_{i,n}$ is the corresponding length scale, and $\alpha_{i,n}$ is the corresponding scaling index. $N_n(\alpha)$ is then the number of intervals that contain like α's and $\varepsilon_n(\alpha)$ is the fundamental length of these intervals with the underlying assumption being made that intervals of like sizes have like measures on their support segments; an assumption that is true for any case we will encounter. Finally, $f(\alpha)$ is the fractal dimension of intervals that are visited with like probability (measure) $\mu = \varepsilon^{\alpha}$, i.e. $f(\alpha)$ is the fractal dimension as seen by the intervals who share a common scaling index α. Another interpretation is that the $f(\alpha)$ describes how the "histogram" $N_n(\alpha)$ varies as $\varepsilon \to 0$.

The q-based method starts off with the definition of the partition function $\mathcal{Z}$

$$\mathcal{Z}(q) \equiv \sum_{i=1}^{N_n} \mu_i^q \tag{2.53}$$

The exponent q is used to "weight" the measures on the support of the set. The denser the population on segments of the support, the more "weight" they are given with negative values of q and vice versa. Note that if $q = 0$, Eq. 2.53 is effectively the summation of line segments across the support, i.e. $\mathcal{Z}(0) = N_n$ and the population of those segments is ignored. Therefore, the partition function $\mathcal{Z}$ is specifically designed to coordinate with the generalized fractal dimensions D_q, where, as discussed in previous sections, various values of q are used to draw out the influence of various

population densities across the support of the set being analyzed. We can form a mathematical link between $\mathcal{Z}$ and D_q via

$$\mathcal{Z} \sim \varepsilon^{\tau(q)} \tag{2.54}$$

where $\tau(q) \equiv D_q(q-1)$. Rearranging terms yields

$$D_q = 1/(q-1) \lim_{\varepsilon \to 0} \frac{\log \mathcal{Z}(q)}{\log \varepsilon} \tag{2.55}$$

In Eq. 2.54 and 2.55, ε is a the fundamental "box" size used to cover the support of the set in the same sense as the box counting technique. Typically, one does not know a priori the various length scales that make up the support of a set as is obvious in the Cantor set. For this reason, a generic box size is used. The point is that the invariant set of a map can be covered arbitrarily by the N_n intervals of length ε_n, but there is an optimal partitioning that is defined by the set itself. The support of the uniform Cantor set is in fact the optimal coarsegraining to describe its fractal microstructure. Obviously, this partitioning is not usually known for more complicated sets.

The relation between the $f(\alpha)$ and D_q spectra is formed by substituting Eq. 2.52 into the definition of the partition function (Eq. 2.53):

$$
\begin{aligned}
\mathcal{Z}(q) \equiv \sum_{i=1}^{N_n} \mu_i^q &= \sum_\alpha N_n(\alpha)\left(\varepsilon_n^\alpha\right)^q \\
&= \sum_\alpha \varepsilon_n^{-f(\alpha)} \varepsilon_n^{\alpha q} \\
&= \sum_\alpha \varepsilon_n^{q\alpha - f(\alpha)}
\end{aligned}
\tag{2.56}
$$

Since $\varepsilon_n \ll 1$ as $n \to \infty$, $\mathcal{Z}(q)$ is maximized by minimizing the function

$$g(\alpha) = q\alpha - f(\alpha) \tag{2.57}$$

Eq. 2.57 is minimized when

$$q = f'(\alpha)$$
$$f''(\alpha(q)) < 0 \tag{2.58}$$

where $\alpha(q)$ is the value of α which satisfies $q = f'(\alpha)$. Note that the $f(\alpha)$ spectrum must be concave downward for a minimum of g to be achieved. The partition function can then be approximated by

$$\mathcal{Z}(q) \approx \varepsilon_n{}^{q\alpha(q) - f(\alpha(q))} \tag{2.59}$$

Since $\mathcal{Z}(q) \sim \varepsilon_n{}^{\tau(q)}$, it is apparent that

$$\tau(q) = q\alpha(q) - f(\alpha(q)). \tag{2.60}$$

Thus, $\tau(q)$ is found by taking the Legendre transform of the $f(\alpha)$ spectrum. Figure 2.22 shows the effects of homogeneity in the distribution (with homogeneous support) on the partition scaling exponents $\tau(q)$ and generalized fractal dimensions D_q. For the homogeneous set, the generalized fractal dimensions converge to one value, namely $D_q = \log(2)/\log(3)$, while the nonhomogeneous set creates a set of generalized dimensions $D_q = (\log(\mu_1^q + \mu_2^q))/((1-q)\log(3))$. The corresponding $f(\alpha)$ spectrum is shown in Fig. 2.23.

2.5 The Link Between Fractals and Chaos

The fractal dimensions used to quantify the structure of chaotic attractors are also applicable to the geometric fractals shown in Sec. 2.4. To illustrate, the capacity dimension D_0 for the Cantor set (Fig. 2.16, right column), the Koch curve (Fig. 2.16, left column), and the Sierpinski triangle 2.15 are calculated.

- **The Cantor Set**

  For each level n in the iteration process, there are $N = 2^n$ segments of size
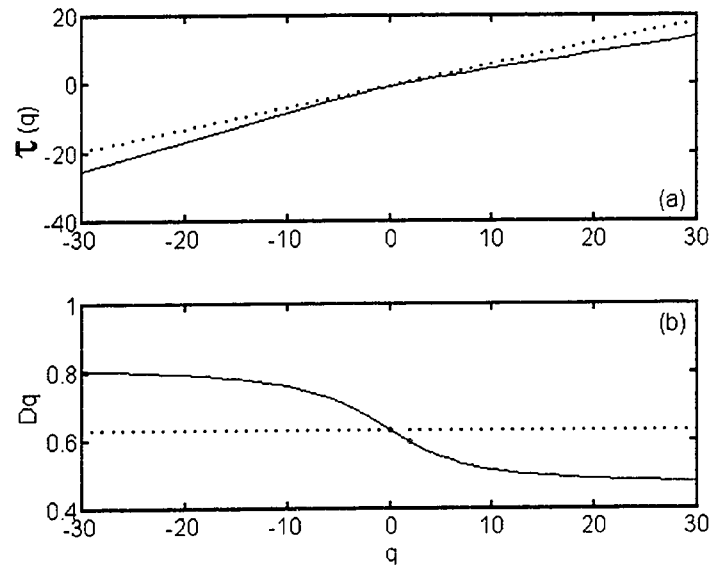
Figure 2.22: The scaling exponents (a) and generalized dimensions (b) for a homogeneous distribution $\mu_1 = \mu_2 = 0.5$ {dotted line} and a nonhomogeneous distribution $\mu_1 = 0.6, \mu_2 = 0.4$ {solid line} in the Cantor set (See Fig. 2.17 for reference).

46
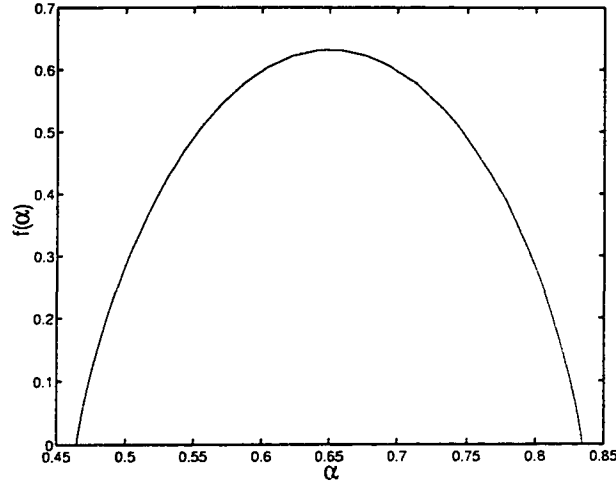
Figure 2.23: The $f(\alpha)$ spectrum for the nonhomogeneous distribution $\mu_1 = 0.6$. $\mu_2 = 0.4$ in the Cantor set. The apex of the curve is the (single point) $f(\alpha) = \log(2)/\log(3)$ for the homogeneous distribution.

$\varepsilon = 3^{-n}$. Using Eq. 2.6, we find

$$D_0^{\text{Cantor}} = \lim_{\varepsilon \to 0} \frac{\log 2^n}{\log 3^n} = \frac{\log 2}{\log 3} = 0.63092\ldots$$

• **The Koch Curve**

For each level n in the iteration process, there are $N = 4^n$ segments of size $\varepsilon = 3^{-n}$. Using Eq. 2.6, we find

$$D_0^{\text{Koch}} = \lim_{\varepsilon \to 0} \frac{\log 4^n}{\log 3^n} = \frac{\log 4}{\log 3} = 1.26185\ldots$$

• **The Sierpinski Triangle**

For each level n in the iteration process, there are $N = 3^n$ new triangles of size $\varepsilon = 2^{-n}$. Using Eq. 2.6, we find

$$D_0^{\text{Sierpinski}} = \lim_{\varepsilon \to 0} \frac{\log 3^n}{\log 2^n} = \frac{\log 3}{\log 2} = 1.58496\ldots$$

Thus,

$$D_0^{\text{Cantor}} < D_0^{\text{Koch}} < D_0^{\text{Sierpinski}}.$$

To visualize these sets, one must look at them in a minimum Euclidean dimension of $D_e = \lceil D_0 \rceil$ where $\lceil \cdot \rceil$ is an operator which rounds the value to the next highest integer. Thus, the embedding dimensions for the above examples are

$$D_e^{\text{Cantor}} = 1$$
$$D_e^{\text{Koch}} = 2$$
$$D_e^{\text{Sierpinski}} = 2$$

The closer the capacity dimension comes to the embedding dimension, the closer it comes to filling up the space in which the data are viewed. For clarity, consider an alteration to the geometric rules used to develop the Sierpinski triangle in the following way: for every solid triangle, divide it up into 4 equi-sized triangles, only this time, keep the middle triangle as well. We would expect that this rather boring rule to do nothing more than simply section off various portions of the space filled by the original solid triangle. Furthermore, one does not expect its dimension to increase since areas (or masses) are neither added nor subtracted from the original structure. In fact, the capacity dimension becomes

$$D_0^{\text{Altered Sierpinski}} = \lim_{\varepsilon \to 0} \frac{\log 4^n}{\log 2^n} = \frac{\log 4}{\log 2} = 2$$
$$= D_e^{\text{Altered Sierpinski}}$$

which demonstrates the fact that the sectioning does nothing to the original dimension of the solid triangle.

Another link between chaos and fractals stems from the fact that chaotic attractors have a (multi)fractal microstructure. While box counting techniques can be

used to estimate the fractal dimension of the attractor geometry, it is an insufficient means of characterizing an attractor which has a nonuniform probability distribution. Uniform distribution is the exception rather than the rule for chaotic attractors. Thus, chaotic attractors are generally not geometrically self similar; rather, they are statistically self similar. For example, consider the Poincaré section of the Duffing oscillator with chaotic parameters shown in Fig. 2.24. The resulting attractor is referred to as the "Dancing Newt" [13]. It is clear that in the Dancing Newt, there
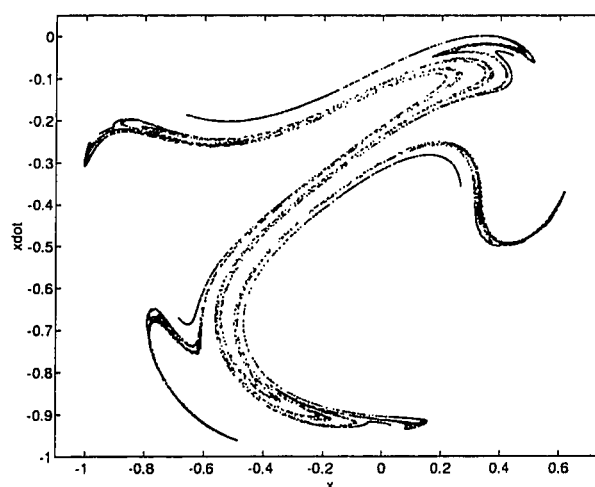
Figure 2.24: A phase plane embedding of a chaotic Duffing system whose motion is described by the nonlinear ODE: $\ddot{x} + 2\xi\dot{x} + x^3 = p\cos\Omega t$. The parameters used to produce the chaotic "Dancing Newt" are $p = 1.0, \Omega = 0.229$, and $\xi = 0.015$.

is a nonuniformity in population density across the attractor. In addition, zooming in on a portion of the attractor will result in an image which only resembles the whole, but does not replicate it exactly. The similarity is in the fine structure that exists upon zooming. Theoretically, there will exist an infinite number of gaps and fractal microstructure, regardless of the zoom factor.

As another example, consider the backwards iteration of both the symmetric and asymmetric tent map (Fig. 2.25 and Fig. 2.26, respectively). The resulting Cantor sets are the optimal means of coarsegraining the phase space, i.e. we can arbitrarily choose boxes of various sizes to estimate the generalized fractal dimensions of the attractor, but the resultant support of the Cantor set is the best choice. If the tent map is asymmetric, it implies that the support will be nonuniform. The dynamics of the map are responsible for distribution of iterates on the attracting set. It is possible to achieve a multifractal distribution whether or not the support is fractal; all that is required is that there exists an infinitely fragmented measure of the distribution.

2.6 Multifractal Analysis Using Wavelets: the Wavelet Transform Maximum Modula Method (WTMM)

The material found in this section is a replication of the work found in [62] with the exception of the comments in Section 2.6.1 where a critique of the WTMM is given.

Multifractal analysis has been used on a wide range of mathematical and natural phenomena including atmospheric turbulence, pore spaces in sandstones, and diffusion limited aggregation. In the discussion of multifractal properties for a distribution of points, a distinction must be made between a *measure* of a distribution and the geometric set or so-called *support* of a distribution. As an example, the support of the Cantor set consists of the nonzero (line) sections at a given iteration level. A measure is a real quantity, such as mass or probability, that is associated with a distribution. Furthermore, while both the support and the measure may be fractal, they do not generally have the same fractal dimension, i.e. the set is multifractal. The wavelet transform can be used to reveal D_q by uncovering the singularity strength of various portions of a fractal signal in "time" and in scale. Specifically, the portions of interest are the locations of the maximum moduli of wavelet transform coefficients in the time-scale plane [62]. With each maximum moduli line one
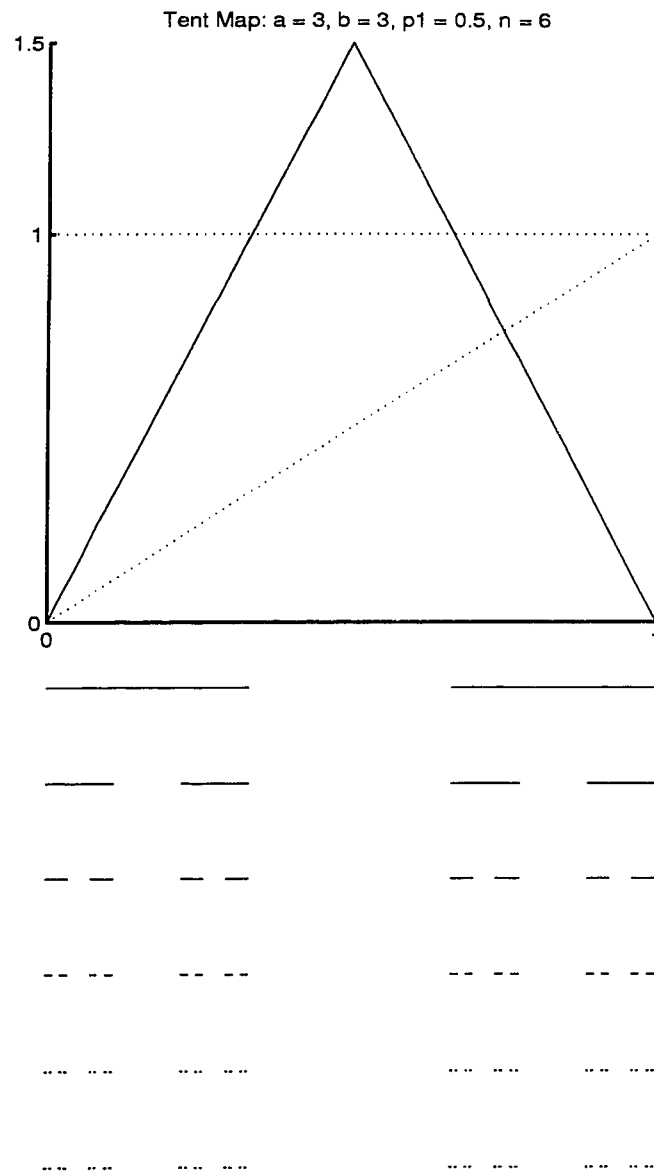
Figure 2.25: The symmetric tent map and the homogeneous Cantor set produced by the backwards iteration of the unit interval. The support of the Cantor set is directly related to the shape of the tent map while the dynamics of the map control the population of iterates on the support. For homogeneous conditions, the resulting Cantor set can be described by a single fractal dimension.
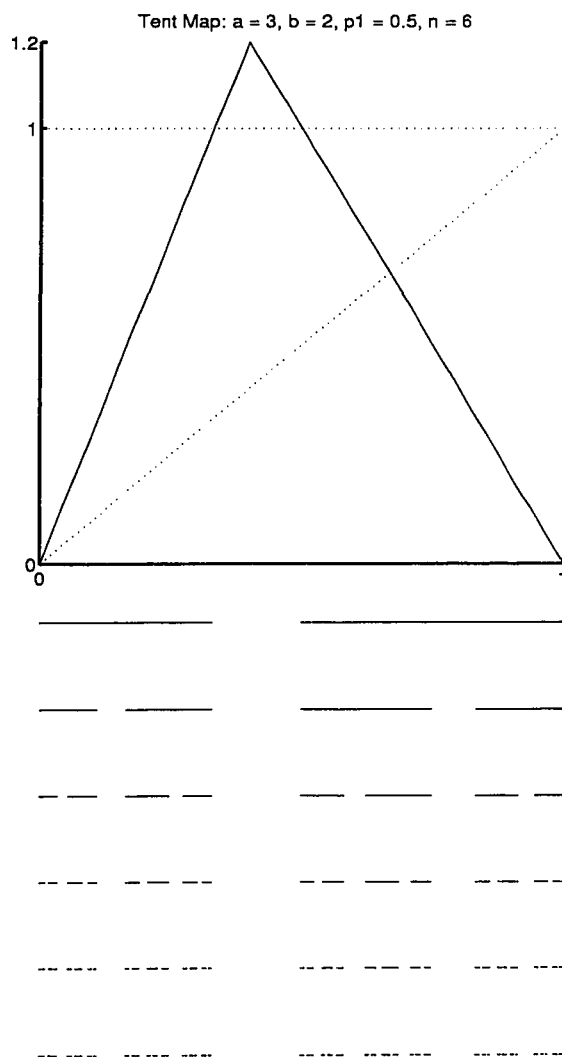
Figure 2.26: The asymmetric tent map and the nonhomogeneous Cantor set produced by the backwards iteration of the unit interval. The support of the Cantor set is directly related to the shape of the tent map while the dynamics of the map control the population of iterates on the support. For nonhomogeneous conditions, the resulting Cantor set is characterized by a spectrum of fractal dimensions.

can associate a singularity strength α that is ultimately used to form a singularity spectrum $f(\alpha)^2$. The singularity spectrum characterizes a multifractal signal because the wavelet maxima lines follow the same renormalization rules as the wavelet transform itself and thus reproduces the hierarchical structure of the signal.

The analyzing wavelets are themselves self-similar at all scales and thus can be used to expose the geometric or statistical self-similarity of a (multi)fractal sequence. The wavelets act as "boxes" that cover the distribution. The wavelet coefficients are produced by convolving a scaled wavelet (filter) with the original signal, resulting in a time-scale representation of the original signal. Mallat and Hwang [51] have shown that singular behavior around a point x_0 implies that there exist maxima lines converging towards x_0 as the box size or scale $a \to 0$. For each line a singularity strength is associated such that

$$T_\psi[f](x_0, a) \sim a^{\alpha(x_0)}, \quad a \to 0^+ \tag{2.61}$$

The number of maxima lines at scale a associated with the exponent $\alpha(x_0)$ is denoted as $N_\alpha(a)$. Bacry et al [5] have rigorously proven that $N_\alpha(a)$ scales like

$$N_\alpha(a) \sim a^{-f(\alpha)} \tag{2.62}$$

where $f(\alpha)$ is the *singularity spectrum* or *Hölder spectrum*. The singularity spectrum $f(\alpha)$ can be used to obtain estimates of the generalized dimensions D_q. To form the relation, consider the partition function

$$\mathcal{Z}(q, a) = \sum_{\ell \in \mathcal{L}(a)} |T_\psi[f](b_\ell(a), a)|^q \tag{2.63}$$

where $\mathcal{L}(a)$ is the set of all maxima lines ℓ existing at scale a, and $b_\ell(a)$ is the position, at scale a, of the maximum belonging to line ℓ. Then by using Eq. 2.61,

---

[2]The singularity strength α is associated with the Hölder exponent while the singularity spectrum $f(\alpha)$ is associated with the Hausdorff dimension.

$\mathcal{Z}(q, a)$ becomes

$$\mathcal{Z}(q, a) \sim N_\alpha(a)(a^\alpha)^q \tag{2.64}$$

and by substituting the expression of $N_\alpha(a)$ given by Eq. 2.62, we arrive at

$$
\begin{aligned}
\mathcal{Z}(q, a) &\sim \sum_\alpha a^{(q\alpha - f(\alpha))} \\
&\sim a^{min_\alpha(q\alpha - f(\alpha))}.
\end{aligned}
\tag{2.65}
$$

Thus, if $\tau(q)$ is the scaling exponent characterizing the power-law behavior of $\mathcal{Z}(q, a)$:

$$\mathcal{Z}(q, a) \sim a^{\tau(q)}, \tag{2.66}$$

then we obtain

$$\tau(q) = min_\alpha(q\alpha - f(\alpha)) \tag{2.67}$$

Finally, the generalized dimensions of the distribution can be calculated via

$$\boxed{D_q = \frac{\tau(q)}{q-1}} \tag{2.68}$$

Thus, the wavelet transform provides a natural generalization of classical box-counting techniques and can be used to calculate the generalized dimensions D_q. The process of using the wavelet transform to reveal multifractal structure is commonly referred to as the method of *wavelet transform maximum modula* or WTMM.

2.6.1 A Critique of the WTMM Method

There are many benefits in using the WTMM method over conventional box-counting techniques. For example, the fractal structure of the signal (if it exists) naturally falls out in the wavelet transform. A second benefit is that the type of wavelet can be chosen such that it ignores (is blind to) certain degrees of polynomial influence. That is, the number of vanishing moments for a given wavelet directly affects how

54

sensitive it is to polynomial behavior. It is possible to construct a mother wavelet such that the number of vanishing moments are fairly large and, in doing so, will greatly attenuate polynomial contamination. In using this technique, all of the possible generalized dimensions can be estimated at once. Even if the object is not multifractal, an estimate of the capacity dimension is returned. The WTMM is a relatively fast computational technique that yields results in a quick and efficient manner.

The reader may ponder as to why wavelet techniques were not used in the past to analyze chaotic response, especially in light of all the advantages they have to offer. Certainly, the idea of using wavelets with fractal structures seems a very natural one. The reason lies in the fact that a *single variable of a chaotic system (one coordinate of a chaotic flow) is not self affine*, a fundamental drawback that cannot be easily overcome. The level of preconditioning required to map a chaotic flow down to some self affine one dimensional form would probably outweigh the computational advantages of the WTMM. In addition, such a mapping (if it exists) would highly limit the interpretation of the data as it is impossible for maps to maintain all of the fractal characteristics of a chaotic flow [55]. Nevertheless, the WTMM method will work on any one dimensional sequence which has either a geometric or statistical self-affinity. We therefore explore some applications of the WTMM in Sec. 2.6.2.

2.6.2 Application of the WTMM Method

Here we use the WTMM method to analyze the tent map with uniform support and (non)uniform distribution to illustrate the capabilities of the WTMM. Where it is relevant, an image of the wavelet used in the analysis appears in the upper left hand corner of the plot. The wavelets used are derivatives of a Gaussian wavelet such that the Mth order Gaussian wavelet is defined as

$$\psi^{(M)}(x) = \frac{d^M}{dx^M}\left(e^{-x^2/2}\right) \tag{2.69}$$

Figure 2.27 shows Gaussian wavelets for various orders.
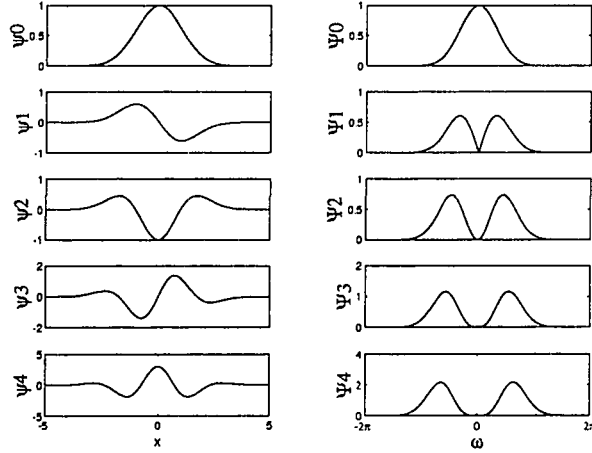
Figure 2.27: Gaussian wavelets $\psi_1, \ldots, \psi_4$ in the time and frequency domain for derivative orders $M = 1, 2, 3,$ and, 4. The original Gaussian curves used to build the wavelets are shown the top row.

Homogeneous Tent Map

The use of the tent map ties in ideas from chaos and fractals and thus serves as a nice pedagogical example. We can form a hypothetical fractal measure, $\mu(x)$, by assigning a distribution to the intervals generated by backwards iteration of the unit interval. In this case, we examine uniform conditions, i.e. both the support and the distribution are homogeneous. The parameter set is $\{\mu_1 = \mu_2 = 1/2, a = b = 3, n = 6, N = 1024\}$, where μ_1 and μ_2 are the mass distributions for the left and right intervals at iteration level $n = 1$, respectively, a and b are the absolute values of the slopes on the left and right sides of the tent map, respectively, and N is the number of points used to form a discrete approximation of the set. The resulting measure on the discretized Cantor set is $\mu(x)$ and is used as an input into the WTMM

program. Figure 2.28 shows the symmetric tent map and the corresponding Cantor set resulting from a finite number of backwards iteration of the unit interval. Figure 2.29 shows the corresponding discrete mass density and distribution function.

The continuous wavelet transform (CWT) of the measure $\mu(x)$ is shown in Fig. 2.30. The WTMM map (Fig. 2.31 upper left) is provided for comparison. The self similarity of the measure μ is reflected in the symmetric forkings of the WTMM. For each parent at scale a, two children (forks) develop at scale $a/3$. On a logarithmic scale, this split is equispaced and in this particular case, occurs at every $\log_2 3 = 1.585$ interval (on a $\log_2$ scale). Thus, we find the expected result $N_\alpha(a) = N_0 a^{-\log 2/\log 3}$ where N_0 depends on the order of the analyzing wavelet. In addition, it can be seen that the WTMM converges to the support of μ as $a \to 0$.

In an algorithmic sense, it is useful to identify the chains of the WTMM. After the chains have been identified, they are pruned of spurious maxima lines, and the result is a WTMM *skeleton map* as shown in upper right corner of Fig. 2.31. The numbers on the skeleton map are the used to identify distinct WTMM ridges. The skeleton map is used to develop the partition function as shown in Fig. 2.31 (lower left). The partition function (plotted on a log-log scale) is fit with a least squares algorithm to extract the partition scaling exponents $\tau(q)$ (Fig. 2.31 lower right). The $\tau(q)$ results seem to fit almost perfectly with the theoretical curve $\tau(q) = (q-1)\log(2)/\log(3)$. Since $D_q = \tau_q/(q-1)$, the generalized fractal dimensions collapse (theoretically) into one fractal dimension $D = \log(2)/\log(3)$. Figure 2.32 compares the theoretical value to that achieved using the WTMM method.

The results for the WTMM method on the uniform Cantor set using wavelet $\psi^{(2)}$ are shown in the Fig. 2.34. There are more skeleton lines when using $\psi^{(2)}$ than there are when using $\psi^{(0)}$. This is due to the fact that $\psi^{(2)}$ has more oscillations than $\psi^{(0)}$ and should not affect the estimates of D_q.

Tent Map: a = 3, b = 3, p1 = 0.5, n = 6
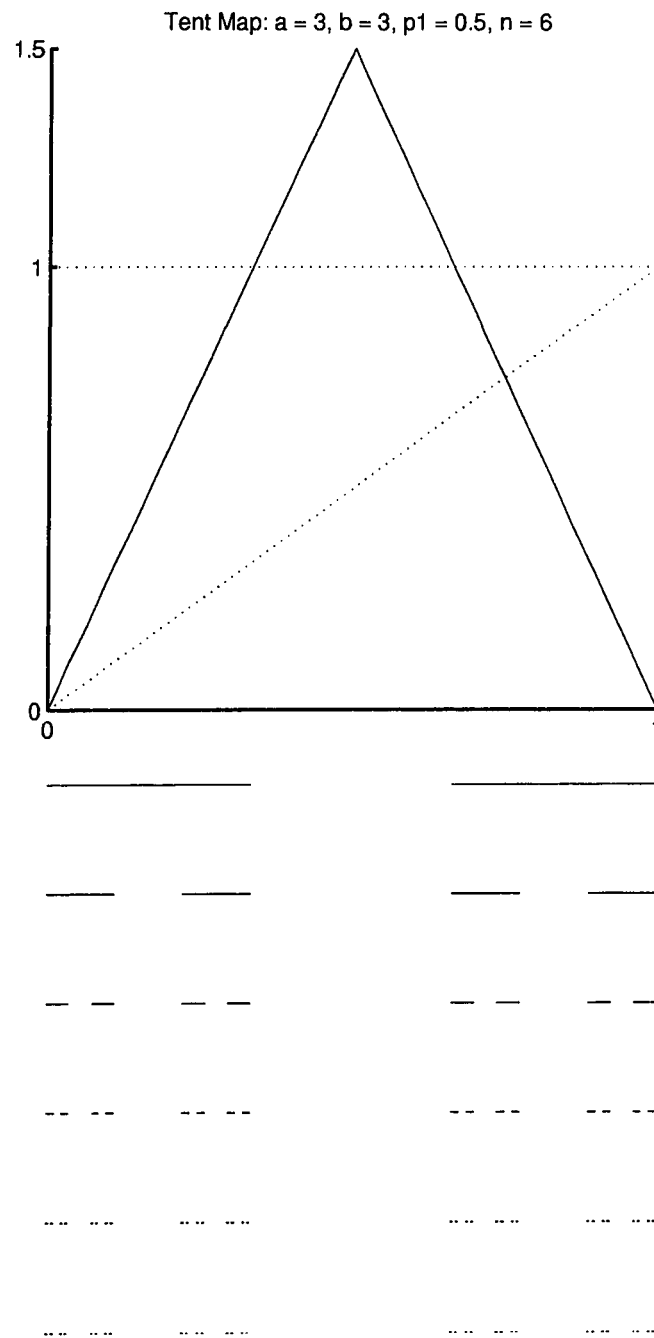
Figure 2.28: Tent map and corresponding Cantor set built by backwards iterating the unit interval of the tent map $\{\mu_1 = \mu_2 = 1/2, a = b = 3, n = 6, N = 1024\}$.
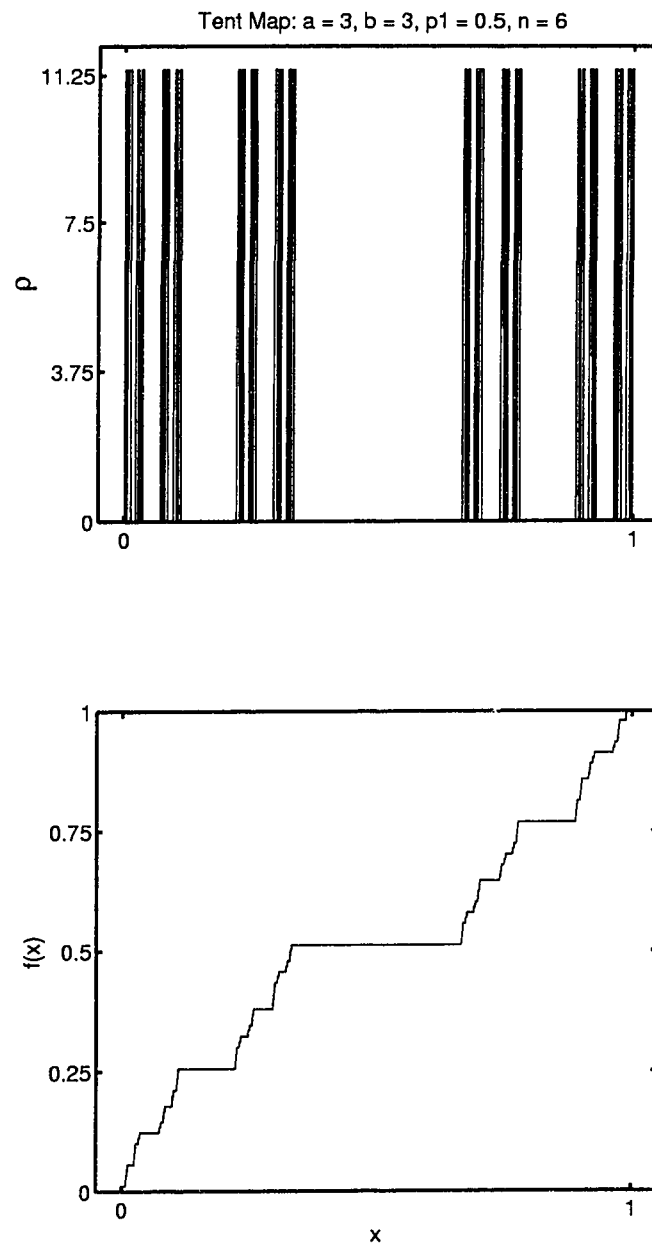
Figure 2.29: Discrete mass density and distribution function $\{\mu_1 = \mu_2 = 1/2, a = b = 3, n = 6, N = 1024\}$.
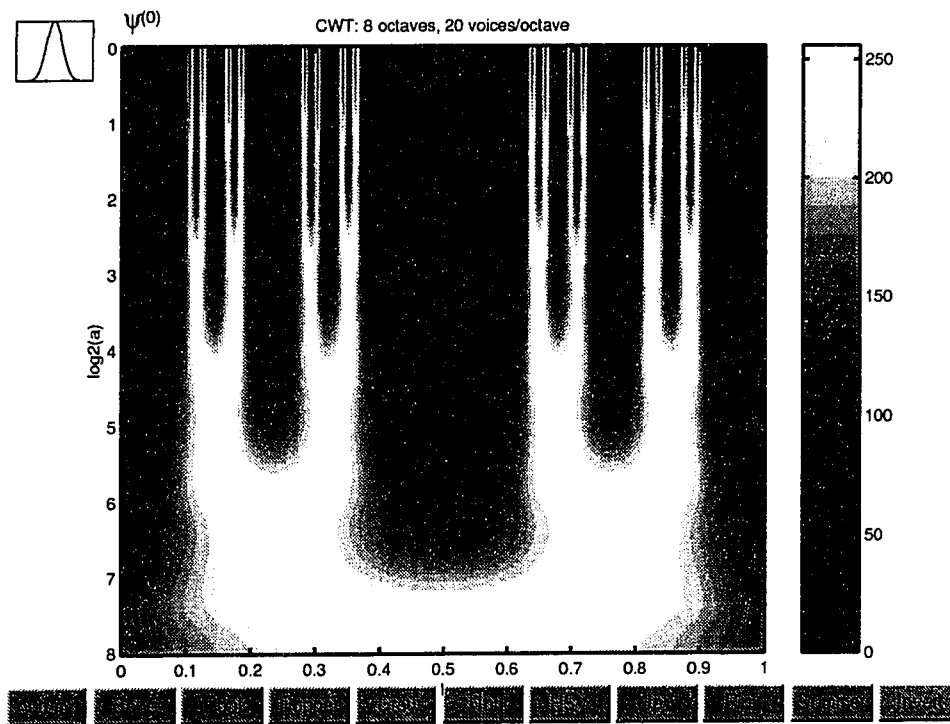
Figure 2.30: CWT of tent map using $\psi^{(0)}$.

60
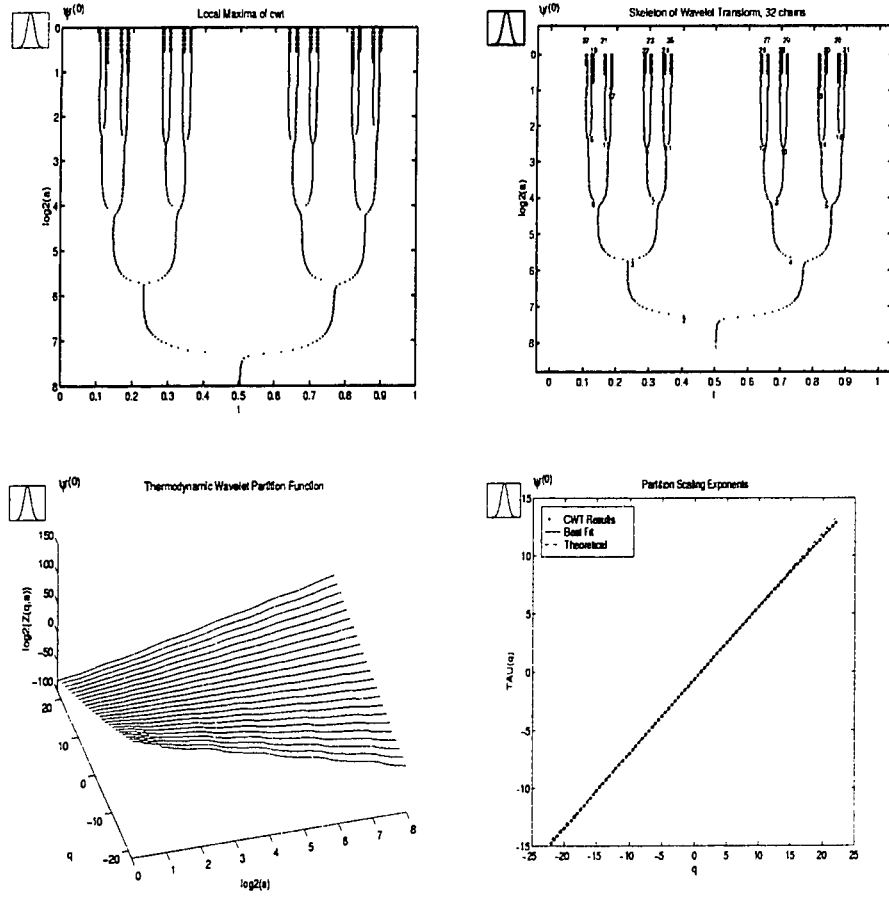
Figure 2.31: WTMM, skeleton map, partition function and scaling exponents for Cantor set with $\{\mu_1 = \mu_2 = 1/2, a = b = 3, n = 6, N = 1024\}$ using $\psi^{(0)}$
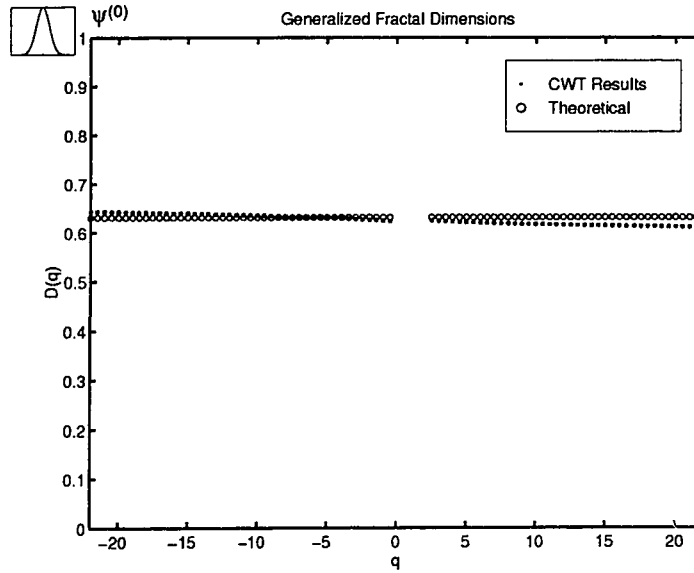
Figure 2.32: Generalized fractal dimensions $\{\mu_1 = \mu_2 = 1/2, a = b = 3, n = 6, N = 1024\}$ using $\psi^{(0)}$.

Nonhomogeneous Tent Map

For the nonhomogeneous case, we focus on the effects of a nonuniform distribution with homogeneous support in the Cantor set Fig. 2.35. The parameter set for this case is $\{\mu_1 = 0.3, \mu_2 = 0.6, a = b = 3, n = 6, N = 1024\}$. Figure 2.36 shows the results for the WTMM analysis using $\psi^{(0)}$ and $\psi^{(2)}$. In both cases, the $\psi^{(0)}$ results are significantly closer to theoretical spectra than are the $\psi^{(2)}$ results.
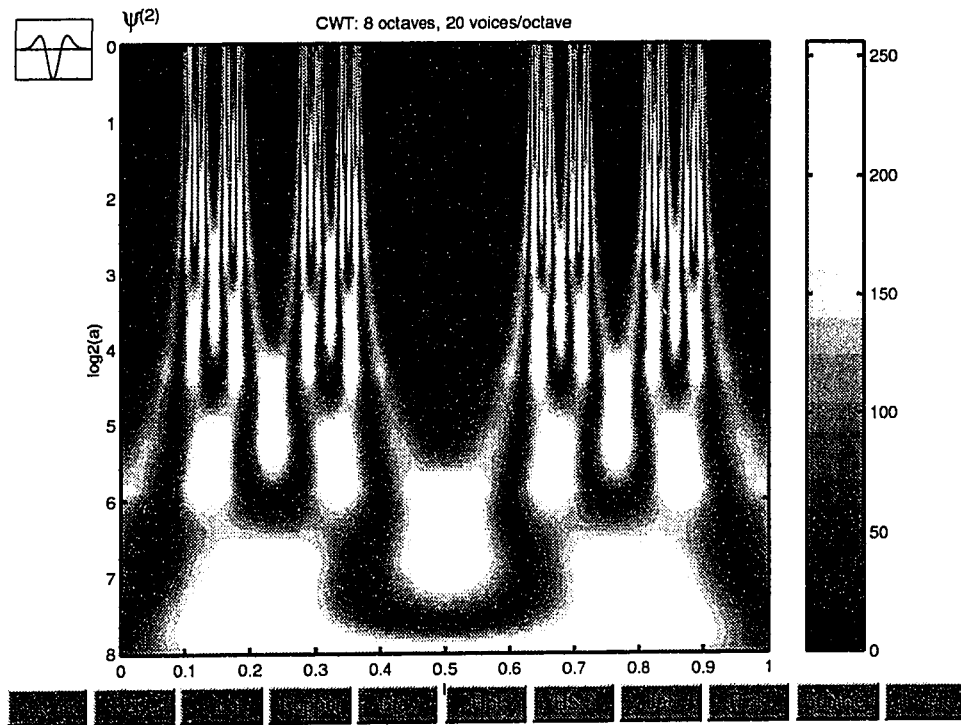
Figure 2.33: CWT of tent map using $\psi^{(2)}$.
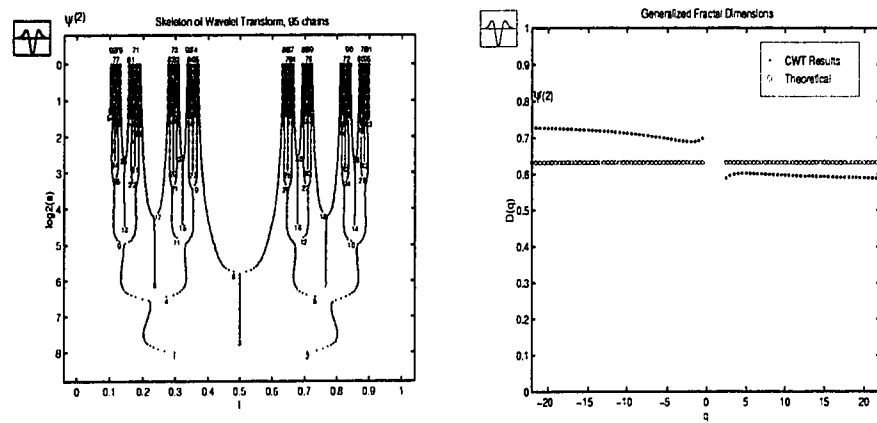
Figure 2.34: Skeleton map and D_q for uniform Cantor set with $\{\mu_1 = \mu_2 = 1/2, a = b = 3, n = 6, N = 1024\}$ using $\psi^{(2)}$.
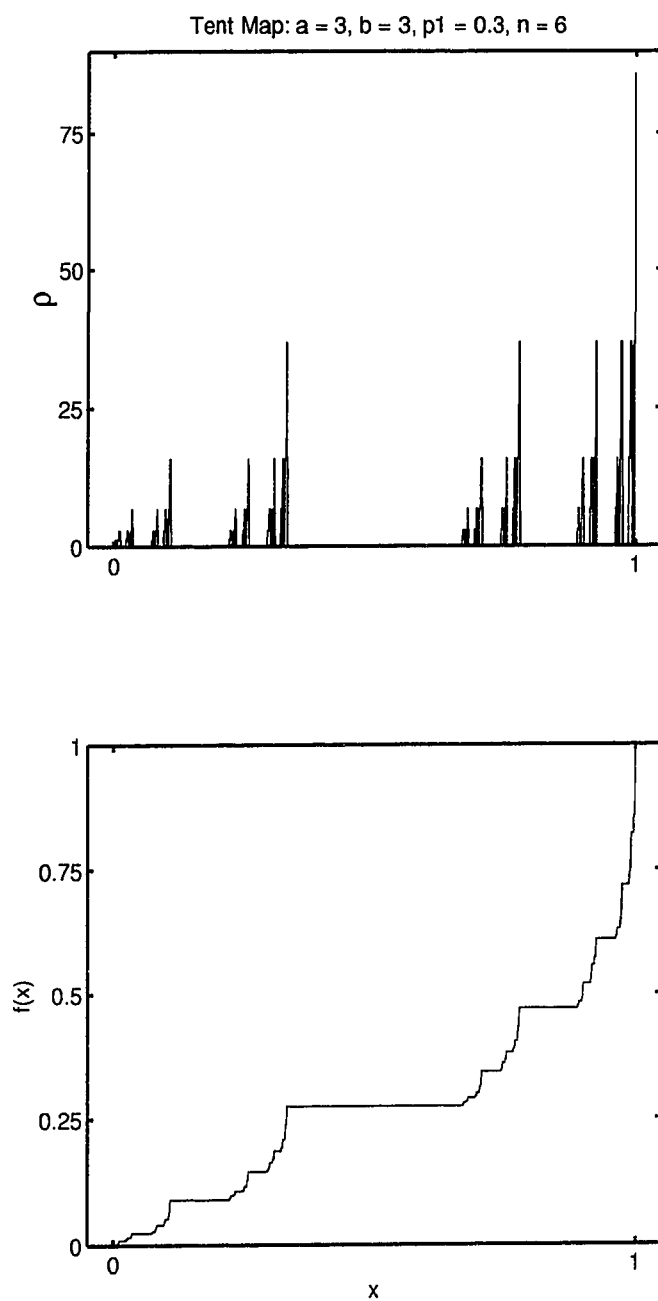
Figure 2.35: Discrete mass density and distribution function $\{\mu_1 = 0.3, \mu_2 = 0.6, a = b = 3, n = 6, N = 1024\}$.
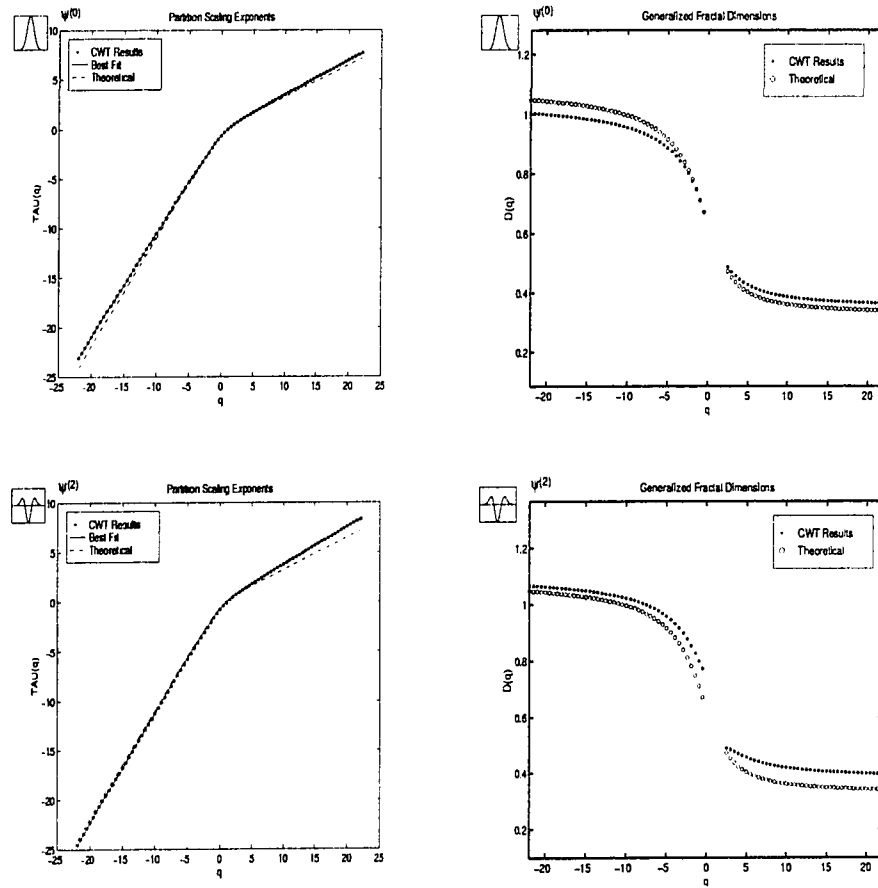
Figure 2.36: Scaling exponents and WTMM D_q estimates for $\{\mu_1 = 0.3, \mu_2 = 0.6, a = b = 3, n = 6, N = 1024\}$ using $\psi^{(0)}$ (top row) and $\psi^{(2)}$ (bottom row).

Chapter 3

DENOISING CHAOTIC SEQUENCES USING THE MAXIMUM OVERLAP DISCRETE WAVELET TRANSFORM

One facet of time series analysis is to separate deterministic from random events. A signal is said to be *denoised* when the desired time and frequency content has been sufficiently extracted from a contaminated sequence. The objective here is to denoise chaotic sequences contaminated with random noise in order to estimate the correlation dimension D_2 of the original uncomtaminated chaotic attractor. Without denoising, D_2 estimates diverge even with a small level of noise contamination.

Developed in this chapter are the key ideas behind each component of a novel denoising technique for use with, but not restricted to, contaminated chaotic sequences. Once denoised, the chaotic sequences are analyzed with conventional nonlinear dynamic techniques and compared to the results for the original sequence. The denoising algorithm is hereafter referred to as *MODRA*, which serves as a suitable acronym

Maximum overlap discrete wavelet packet transform

Optimal

Deterministic

Reduction

Algorithm.

Each element of the denoising algorithm and details concerning the nonlinear measures are described in order of their use. Section 3.1 defines a highly flexible and

useful extension of the conventional discrete wavelet transform known as the MOD-WPT [73]. The MODWPT plays a pivotal role in the denoising algorithm. In Sec. 3.2, a description of a best basis optimization is given and a novel technique for developing a MODWPT cost functional using Shannon's entropy is introduced. Section 3.3 discusses a simple technique for reducing the number of coordinates needed to sufficiently characterize the time-frequency content of the original sequence. In Sec. 3.4, Kaplan's determinism method (KDM) [37, 36] is described and opportunities for advancement identified. KDM is used on each member of the reduced coordinate set, and the coordinates found to be deterministic are synthesized to form a denoised version of the contaminated data. The synthesis process is described in Sec. 3.5. The remainder of the chapter is devoted to the description of nonlinear measures, such as the correlation dimension, and their usefulness on chaotic sequences is presented. The subtleties of obtaining quality nonlinear measures are discussed in Sec. 3.6 while the advantages of using singular value decomposition with nonlinear measures is explained in Sec. 3.7. Finally, the results are presented in Sec. 3.8 for a comparison of MODRA to a conventional wavelet denoising technique (Sec. 3.8.1), for a collection of nonlinear results for denoised contaminated chaotic sequences (Sec. 3.8.2), and for a chaotic beam experiment (Sec. 3.8.3). Each section is prefaced with a statement of purpose and concluded with a description of relevance to the method.

3.1 The Maximum Overlap Discrete Wavelet Packet Transform (Stage I)

The purpose of this section is to introduce the discrete wavelet transform (DWT), the discrete wavelet packet transform (DWPT), the maximum overlap discrete wavelet transform (MODWT), and the maximum overlap discrete wavelet packet transform (MODWPT). We will start with the DWT and gradually develop ideas key to the MODWPT, which is at the heart of the denoising algorithm for chaotic sequences.

An in-depth discussion of the DWT and the continuous wavelet transform (CWT) is presented in Appendix A and the reader is referred there for reference. The nomenclature used in this chapter is consistent with that found in Percival and Walden [73].

3.1.1 The DWT

The DWT can be explained quite easily in terms of linear filtering theory and some basic linear algebra such as *differencing* and *summation* operations. To start, consider a $N = 4$ point series $\{X_t\} = \{X_0, X_1, X_2, X_3\}$ as shown in the top level in Fig. 3.1. At each level j, the topmost left block of the previous level or *parent* block

Figure 3.1: The decomposition of a discrete sequence $\{X_t\}$ using summation and differencing operations. The decomposition level is denoted as j such that the sequence at $j = 0$ is the original signal.

is broken down into two *children* blocks V and W, each of which contains half the number of elements as the parent. These transform coefficients are identified with

a index pair (j, t) which denote the decomposition level and "time" index, respectively. By construction, the V's are proportional to the sums of adjacent averages on a scale of $\tau_j \equiv 2^{j-1}$ for $0 < j \leq J$ where $J \leq \log_2 N$. Qualitatively, the V's represent smoothed versions of portions from the original signal. The "time" span of these portions is equivalent to the decomposition scale τ_j. Conversely, the W's are proportional to the differences of adjacent averages on a scale of τ_j. The W's are large if there exists a sharp discontinuity in $\{X_t\}$ since they are the result of differencing operations on the original signal. Qualitatively then, the W's pick out the rougher aspects of the original sequence.

The transformation of the sequence $\{X_t\}$ into V's and W's is invertible, i.e. we can synthesize the original signal by recombining the V's and W's in a specified way. The coefficients used to rebuild the original signal are highlighted in Fig. 3.1 by a thick border around the corresponding blocks. The idea is to combine the smooth and rough transform coefficients at level j to approximate (or in this case obtain exactly) the smooth transform coefficients at level $j - 1$. Then the smooth and rough coefficients are used to rebuild the smooth coefficients at level $j - 2$ and so on until the original signal is returned. For the example shown in Fig. 3.1, the synthesis of $\{X_t\}$ is obtained via

$$V_{1,0} = \frac{V_{2,0} - W_{2,0}}{2}$$

$$V_{1,1} = \frac{V_{2,0} + W_{2,0}}{2}$$

$$X_0 = \frac{V_{1,0} - W_{1,0}}{2}$$

$$X_1 = \frac{V_{1,0} + W_{1,0}}{2}$$

$$X_2 = \frac{V_{1,1} - W_{1,1}}{2}$$

$$X_3 = \frac{V_{1,1} + W_{1,1}}{2}.$$

The transformation of $\{X_t\}$ can be put into matrix form $\mathbf{W} = \underline{\mathcal{W}}\mathbf{X}$:

$$\begin{Bmatrix} W_{1,0} \\ W_{1,1} \\ W_{2,0} \\ V_{2,0} \end{Bmatrix} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ -1 & -1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{Bmatrix} X_0 \\ X_1 \\ X_2 \\ X_3 \end{Bmatrix}$$

The transform matrix $\underline{\mathcal{W}}$ as depicted above is row orthogonal but is not orthonormal since $\underline{\mathcal{W}}\underline{\mathcal{W}}^{\mathbf{T}} \neq \underline{\mathcal{I}}$. However, an orthonormal transformation can be achieved by scaling the transform coefficients such that at every decomposition level j each coefficient is divided by $\sqrt{2}$ resulting in the transform

$$\begin{Bmatrix} W_{1,0} \\ W_{1,1} \\ W_{2,0} \\ V_{2,0} \end{Bmatrix} = \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{2} & 0 & 0 \\ 0 & 0 & -1/\sqrt{2} & 1/\sqrt{2} \\ -1/2 & -1/2 & 1/2 & 1/2 \\ 1/2 & 1/2 & 1/2 & 1/2 \end{bmatrix} \begin{Bmatrix} X_0 \\ X_1 \\ X_2 \\ X_3 \end{Bmatrix} \tag{3.1}$$

Orthonormality is desired because by construction an orthonormal matrix $\underline{\mathcal{W}}$ has the property $\underline{\mathcal{W}}^{-1} = \underline{\mathcal{W}}^{\mathbf{T}}$, hence $\underline{\mathcal{W}}^{\mathbf{T}}\underline{\mathcal{W}} = \underline{\mathcal{I}}$ where $\underline{\mathcal{I}}$ is the identity matrix. This is useful for inverting the transform $\mathbf{W} = \underline{\mathcal{W}}\mathbf{X}$ since $\underline{\mathcal{W}}^{\mathbf{T}}\mathbf{W} = \underline{\mathcal{W}}^{\mathbf{T}}\underline{\mathcal{W}}\mathbf{X} = \mathbf{X}$. In other words, the original signal can be rebuilt from the transform coefficients via

$$\begin{Bmatrix} X_0 \\ X_1 \\ X_2 \\ X_3 \end{Bmatrix} = \begin{bmatrix} -1/\sqrt{2} & 0 & -1/2 & 1/2 \\ 1/\sqrt{2} & 0 & -1/2 & 1/2 \\ 0 & -1/\sqrt{2} & 1/2 & 1/2 \\ 0 & 1/\sqrt{2} & 1/2 & 1/2 \end{bmatrix} \begin{Bmatrix} W_{1,0} \\ W_{1,1} \\ W_{2,0} \\ V_{2,0} \end{Bmatrix}. \tag{3.2}$$

Equation 3.1 is a special case of the discrete wavelet transform known as the Haar DWT, and Eq. 3.2 the corresponding inverse discrete wavelet transform.

The DWT is an invertible orthonormal linear transform, mapping a sequence from the time domain to the time-scale domain. An alternative view of the DWT is that it decomposes a sequence into a set of sequences nominally associated with a pass-band filter set distributed on a logarithmic frequency grid. To see this, we can define a filter h_l for $l = 0, \ldots, L-1$ where L is the length of the filter. For simplicity, let us choose the nonzero filter coefficients in the first row of the $\underline{W}$ matrix in Eq. 3.1 as our filter, i.e. let $\{h_l\} = \{1/\sqrt{2}, -1/\sqrt{2}\}$. This filter has a special name in the literature known as the Haar filter and is the most basic of the wavelet filters.

The frequency response of the Haar filter is most easily obtained via the z-transform. In general, the z-transform of $\{h_l\}$ is

$$H(z) = h_0 + h_1 z^{-1} + \ldots + h_{L-1} z^{-(L-1)} \tag{3.3}$$

where $\{h_l\}$ is the impulse response of the filter. For the Haar, the z-transform is

$$H(z) = h_0 + h_1 z^{-1}. \tag{3.4}$$

Since $z = e^{i\omega}$ maps $H(z)$ from the z domain to the Fourier domain, the frequency response function of the Haar filter is

$$H(\omega) = h_0 + h_1 e^{-i\omega}. \tag{3.5}$$

The associated squared gain function, defined as $\mathcal{H} \equiv |H(\omega)|^2$, is

$$\begin{aligned} \mathcal{H} &= (h_0 + h_1 \cos(\omega))^2 + h_1^2 \sin(\omega)^2 \\ &= h_0^2 + h_1^2 + 2h_0 h_1 \cos(\omega) \\ &= 1 - \cos(\omega) \end{aligned} \tag{3.6}$$

and H is revealed as a high pass filter whose cutoff frequency is centered about normalized frequency $f = 1/4$ (Fig. 3.2). The Haar wavelet filter used to form the W's in Eq. 3.1 is associated with a high pass filter since differencing operations expose the rougher or more discontinuous aspects of a discrete sequence. Likewise,
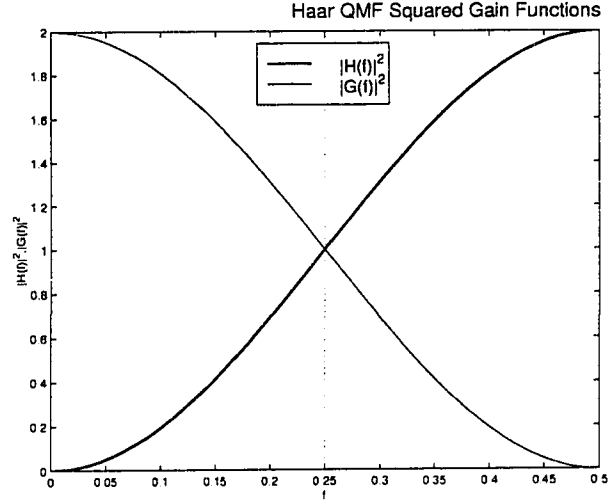
Figure 3.2: The squared gain functions for the Haar wavelet and scaling filters.

a filter defined by summation operations on adjacent averages in $\{X_t\}$ is associated with a low pass filter known as a *scaling* filter. The impulse response of the Haar scaling filter $G(f)$ is $\{g_l\} = \{1/\sqrt{2}, 1/\sqrt{2}\}$ with squared gain function

$$\mathcal{G} = 1 + \cos(\omega) \tag{3.7}$$

where $\mathcal{G} \equiv |G(f)|^2$. $G(f)$ is a low pass filter with normalized cutoff frequency centered at $f = 1/4$ (Fig. 3.2). The wavelet filter $\{h_l\}$ and the scaling filter $\{g_l\}$ are at the heart of the DWT and are related by

$$g_l = (-1)^{l+1} h_{L-1-l} \tag{3.8}$$

for $l = 0, \ldots, L - 1$. Equation 3.8 defines a *quadrature mirror filter* (QMF) pair. Given a suitable wavelet filter $\{h_l\}$ a corresponding scaling filter $\{g_l\}$ is formed through the QMF relation by reversing $\{h_l\}$ in time and negating every other coefficient starting with the first. A filter is considered to be a suitable wavelet filter if it adheres to the guidelines presented in the table below.

Wavelet Filter Conditions

Description	Time Domain	Frequency Domain
Zero Summation	$\sum_{l=0}^{L-1} h_l = 0$	$\mathcal{H}(0) = 0, \mathcal{H}(1/2) = 2$
Unit energy	$\sum_{l=0}^{L-1} h_l^2 = 1$	$\mathcal{H}(f) + \mathcal{H}(f + 1/2) = 2$
Even shift orthogonality	$\sum_{l=0}^{L-1} h_l h_{l+2k} = 0 \; (k \neq 0)$	$\mathcal{H}(f) + \mathcal{H}(f + 1/2) = 2$

The unit energy and even shift orthogonality properties are necessary to establish an orthonormal transform while the zero summation property is a characteristic of all Daubechies wavelet filters in which the Haar wavelet is a member. The corresponding requirements of the scaling filter are provided in the following table.

Scaling Filter Conditions

Description	Time Domain	Frequency Domain
QMF Pair	$g_l = (-1)^{l+1} h_{L-1-l}$	$\mathcal{G}(f) = \mathcal{H}(1/2 - f) = 2$
Positive summation	$\sum_{l=0}^{L-1} g_l = \sqrt{2}$	$\mathcal{G}(0) = 2, \mathcal{G}(1/2) = 0$
Unit energy	$\sum_{l=0}^{L-1} g_l^2 = 1$	$\mathcal{G}(f) + \mathcal{G}(f + 1/2) = 2$
Even shift orthogonality	$\sum_{l=0}^{L-1} g_l g_{l+2k} = 0 \; (k \neq 0)$	$\mathcal{G}(f) + \mathcal{G}(f + 1/2) = 2$
Cross Even shift orthog.	$\sum_{l=0}^{L-1} g_l h_{l+2k} = 0 \; (k \in \mathbb{Z})$	$\mathcal{G}(f) + \mathcal{H}(f) = 2$

From a filter bank perspective, the QMF pair can be used to produce a DWT of sequence $\{X_t\}$ with dyadic (power of two) length. The DWT filter bank algorithm is illustrated below for decomposition level $J = 2$.
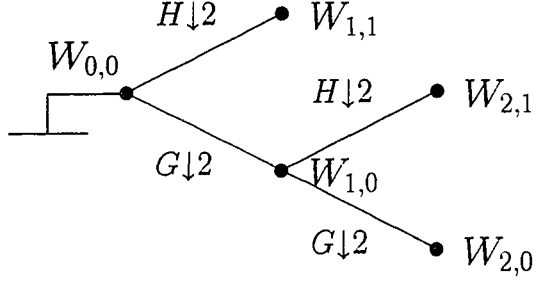
DWT Tree out to level $j = 2$.

Each node in the DWT tree is identified by an index pair (j, n) which denote the decomposition level and local node index, respectively. The algorithm begins by defining $W_{0,0} \equiv \{X_t\}$. $W_{0,0}$ is filtered with the G and H QMF pair and the result decimated by 2 (represented by the $\downarrow 2$ sign in the above diagram). The first level high pass-decimation operation is expressed in the time domain by

$$W_{1,1,t} = \sum_{l=0}^{L-1} h_l X_{2t+1-l \mod N} \tag{3.9}$$

while the low pass-decimation operation is given by

$$W_{1,0,t} = \sum_{l=0}^{L-1} g_l X_{2t+1-l \mod N} \tag{3.10}$$

for $t = 0, \ldots, N - 1$ where $W_{j,n,t}$ is the tth element of node $W_{j,n}$. The algorithm is then repeated on the low pass node $W_{1,0}$ to produce $W_{2,0}$ and $W_{2,1}$, and so on.

Since the G and H filters are approximate low pass and high pass filters, the spectrum of the original sequence $X(f)$ is essentially divided into normalized frequency ranges $[0, 1/4)$ and $[1/4, 1/2)$, respectively, in the first decomposition level.

The low pass output is then critically decimated in time to spread the spectrum out to twice its current range, and a second run through the QMF filter pair yields two nodes $W_{2,0}$ and $W_{2,1}$ associated with normalized frequency ranges $[0, 1/8)$ and $[1/8, 1/4)$, respectively.

Once decomposed, the original signal can be synthesized by performing another series of filter bank operations. For $J = 2$, synthesis involves manipulation of the coefficients found in nodes $W_{2,0}$, $W_{2,1}$, and $W_{1,1}$. If all the nodes are used, the original signal can be recovered exactly (disregarding round off error and limitations on the implementing hardware/software). If, however, a single node $W_{j,n}$ is used and the rest ignored (zeroed out) in the synthesis algorithm, the result is an approximation of $\{X_t\}$ called a *detail* signal denoted as $\mathcal{D}_{j,n}$. The set of all detail signals forms an additive decomposition known as a multiresolution analysis (MRA) of $\{X_t\}$. The synthesis algorithm and the development of the MRA using the DWT is described in detail in Appendix A.

3.1.2   The MODWT

Let $\widetilde{h}_l \equiv h_l/\sqrt{2}$ and $\widetilde{g}_l \equiv g_l/\sqrt{2}$ for $l = 0, \dots, L - 1$. The maximum overlap discrete wavelet transform for a level $J = 2$ decomposition goes as
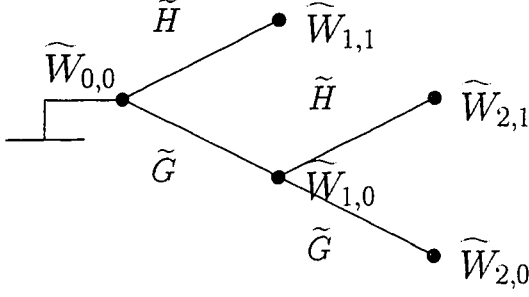
where

$$\widetilde{W}_{1,1,t} = \sum_{l=0}^{L-1} \widetilde{h}_l X_{t-l \mod N}$$

$$\widetilde{W}_{1,0,t} = \sum_{l=0}^{L-1} \widetilde{g}_l X_{t-l \mod N}$$

$$\widetilde{W}_{2,1,t} = \sum_{l=0}^{L-1} \widetilde{h}_l \widetilde{W}_{1,0,t-l \mod N}$$

$$\widetilde{W}_{2,0,t} = \sum_{l=0}^{L-1} \widetilde{g}_l \widetilde{W}_{1,0,t-l \mod N} \tag{3.11}$$

for $t = 0, \ldots, N - 1$. Notice the similarity between Eq. 3.9-3.10 and Eq. 3.11. The MODWT and DWT are in fact related by

$$W_{j,n,t} = 2^{j/2} \widetilde{W}_{j,n,2^j(t+1)-1} \quad \text{for } t = 0, \ldots, N - 1. \tag{3.12}$$

The MODWT differs in that there is no decimation operation after circular convolution with the MODWT filters. The filters $\widetilde{g}_l$ and $\widetilde{h}_l$ are still QMFs, but the orthonormality of the transform matrix is lost since $\sum_{l=0}^{L-1} \widetilde{h}_l^2 \neq 1$ and the orthogonality condition is lost because not all of the dot products between distinct rows

equal zero. Consider, for example, the transform matrix for node $\widetilde{W}_{1,0}$ using the MODWT Haar wavelet filter on a $N = 4$ point sequence $\{X_t\} = \{X_0, X_1, X_2, X_3\}$:

$$
\begin{Bmatrix} \widetilde{W}_{1,0,0} \\ \widetilde{W}_{1,0,1} \\ \widetilde{W}_{1,0,2} \\ \widetilde{W}_{1,0,3} \end{Bmatrix} = \begin{bmatrix} 1/2 & 0 & 0 & -1/2 \\ -1/2 & 1/2 & 0 & 0 \\ 0 & -1/2 & 1/2 & 0 \\ 0 & 0 & -1/2 & 1/2 \end{bmatrix} \begin{Bmatrix} X_0 \\ X_1 \\ X_2 \\ X_3 \end{Bmatrix}.
$$

The sum of the squares of each column or row is $\pm 1/4$ (eliminating the possibility of an orthonormal transform) and the transform matrix is by inspection nonorthogonal.

There are however many benefits in using the MODWT over the DWT. A comparison is made in the following chart.

MODWT/DWT Comparison		
Property	MODWT	DWT
Shift Invariant Transform	$\checkmark$	
Usable for MRA	$\checkmark$	$\checkmark$
Shift Invariant MRA	$\checkmark$	
Details are Zero Phase	$\checkmark$	
Transform Coeff. Usable for Variance Analysis	$\checkmark$	$\checkmark$
Details Usable for Variance Analysis		$\checkmark$
Operable on Arbitrary Length Sequences	$\checkmark$	
Number of Multiplication Operations	$O(N \log_2 N)$	$O(N)$

Note that, although the number of multiplication operations in the MODWT is much greater than that of the DWT, it is still on the order of that needed for the FFT and is suitably fast for most applications.

3.1.3 DWPT and MODWPT

As in the Short-Time Fourier Transform (STFT), the DWT and MODWT offer only one frequency grid pattern in which to view the data. For the STFT, windows of uniform size are used to break down a sequence which also limits its frequency content to be distributed over a uniform grid. The DWT and MODWT use varying window sizes of dyadic length and decompose a signal over a logarithmic frequency grid. Qualitatively, the DWT and MODWT offer a coarse to fine strategy in decomposing a signal where small scale filters are used to localize content in time while large scale filters are used to localize frequency content and examine slower trends of a signal. Still, the MODWT and DWT are inflexible to decomposing a sequence onto arbitrary frequency grids. For compression and detection of signal characteristics, decomposition onto an arbitrary frequency grid is of value if the frequency grid combinations can somehow be optimized in an efficient manner.

The discrete wavelet packet transform (DWPT) and the maximum overlap discrete wavelet packet transform (MODWPT) are examples of transforms which decompose a sequence using a large number of bases which can then be optimized to find an optimal distribution in the time-frequency or time-scale plane. The algorithm for the MODWPT and DWPT is exactly the same as it counterparts except that *each and every* parent node in a given decomposition level is broken down into two children nodes. The algorithm tree for the DWPT and MODWPT is shown in the following diagram for a level $J = 2$ decomposition.
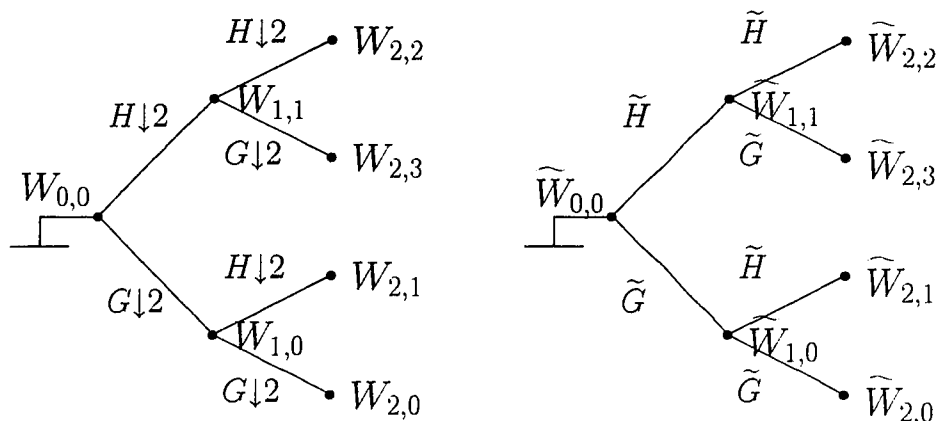
Notice the additional splitting of node $(1,1)$ in comparison to the regular DWT or MODWT. This essentially divides the normalized frequency range $[1/4, 1/2)$ into $[1/4, 3/8)$ and $[3/8, 1/2)$ which are associated with nodes $(2,2)$ and $(2,3)$, respectively. However, the roles of the low pass $G(\cdot)$ and high pass $H(\cdot)$ filters are reversed in this operation! The reason for this is that the frequencies components in the range $[1/4, 1/2)$ (associated with node $(1,1)$) are reversed due to the downsampling operation. In general, a sequence $Y_t = (\downarrow 2)X_t$ has a corresponding frequency response $Y(f) = 1/2[X(f/2) + X(f/2 + 1/2)]$. The $X(f/2 + 1/2)$ term is due entirely to aliasing and its effects are shown diagrammatically in Fig. 3.3. Since the frequencies are reversed at node $(1,1)$ so must be the order of filters used to split them. This reversal reverses again on the next iteration of the algorithm as does the filtering order. Thus, for every other iteration the filter order must be reversed to obtain the proper division in frequency content.

An example of a MODWPT is shown in Fig. 3.4 where three sine waves of various frequencies were summed to form the original sequence. The decomposition level is printed to the left of each row in the plot and the wavelet packet nodes appear from left to right at each level starting with index zero. The topmost node
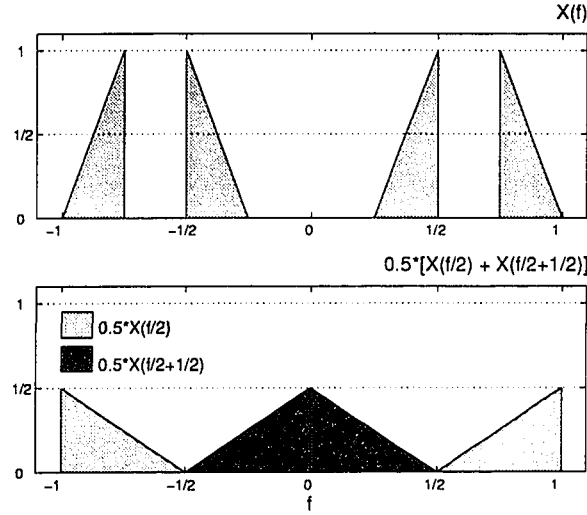
Figure 3.3: The effect of downsampling a sequence in time with a band limited spectrum. The values $X(f)$ for $f \in [1/4, 1/2]$ are attenuated in amplitude, reversed in frequency, and expanded to fill $f \in [0, 1/2]$ after decimation by two.

$\widetilde{W}_{0,0}$ is the original signal. Each node contains N coefficients for $\{X_t\}$ an N point sequence. The width of each block is commensurate with its frequency span with the frequency range of the abscissa being $f = [0, 1/2)$. Notice how the contributing frequencies of the waveform are localized as the decomposition level increases. This is expected as the filter bandwidth diminishes with an increase in j. There is some leakage amongst the nodes because the MODWPT filters are not perfect band pass filters, i.e. there is some overlap in the filters' squared gain functions with regard to nonzero frequency response.

Notice that the energy of the sine wave at $f = 1/4$ is shared between nodes $\widetilde{W}_{1,0}$ and $\widetilde{W}_{1,1}$. Node $\widetilde{W}_{1,0}$ also contains the energy of the $f = 1/64$ wave while $\widetilde{W}_{1,1}$ also contains the energy of the $f = 3/8$ wave. At node $\widetilde{W}_{2,0}$ and $\widetilde{W}_{2,1}$ the $f = 1/64$ and $f = 3/8$ waves are distinctly separated. It is not until level $j = 4$ that we see a clear division of the $f_N/2$ Hz and $3/4f_N$ Hz waves. We see by this example that the
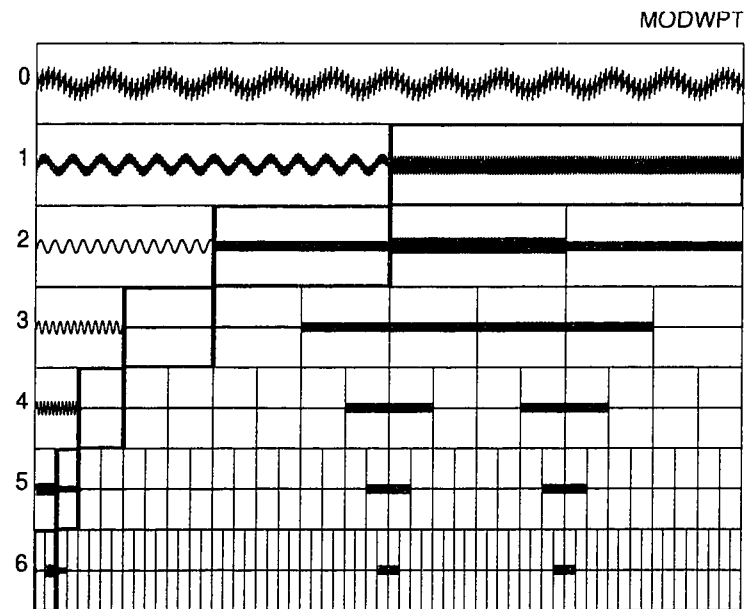
MODWPT

Figure 3.4: MODWPT of a sequence $X_t = \sum_{j=1}^{3} \sin(2\pi f_j t)$ where $f_1 = 1/64, f_2 = 1/4$, and $f_3 = 3/8$.

82

MODWPT is a flexible and powerful means of isolating time-frequency or time-scale content.

Neglected thus far is the fact that the *all* wavelet transforms are capable of decomposing nonstationary data. As an example, a spike of magnitude $A = 5.0\,\widehat{\sigma}_X$, where $\widehat{\sigma}_X$ is the sample standard deviation of $\{X_t\}$, was added to the multifrequency sine wave sequence (Fig. 3.5). Because a spike is mathematically represented as a discrete Dirac delta function $\delta(t - t_0)$, its spectrum contains nonzero components at all frequencies and thus its influence should be seen in all frequency bands. This is seen to be the case as the spike appears in each node (Fig. 3.5). It is highly localized in time at small scales, and less so as the scale increase. This is due to the fact that the singularity is averaged or smeared out as the effective window size increases with scale.
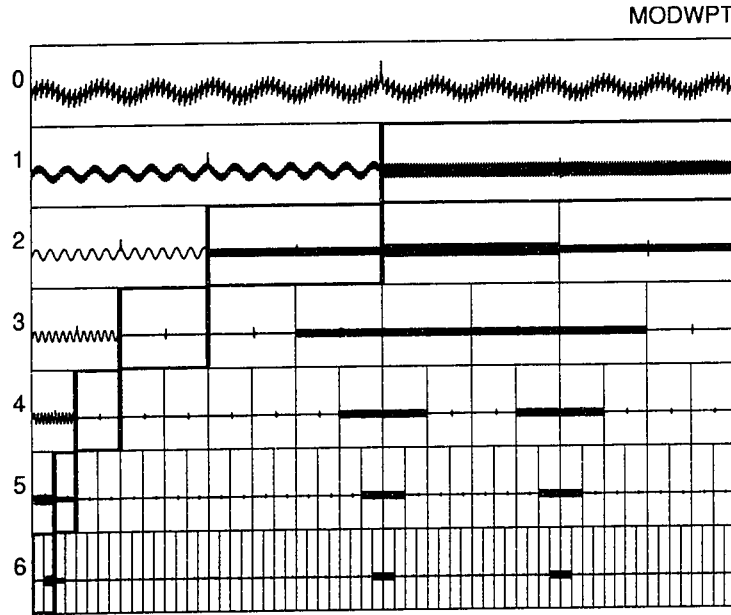
Figure 3.5: MODWPT of a sequence $\{X_t\} = \sin(2\pi t) + \sin(\pi f_N t) + \sin(3/8\pi f_N t) + 5.0\,\widehat{\sigma}_X \delta(t-512)$ where f_N is the Nyquist frequency, $N = 1024$, and $\widehat{\sigma}_X$ is the sample standard deviation of $\{X_t\}$.

Finally we note some interesting aspects of the DWPT and MODWPT. Because the DWPT is an orthonormal transform, any disjoint dyadic decomposition in the DWPT is also orthonormal. For a level $j = 2$ decomposition there are four possible dyadic decompositions

- $\mathcal{C}_1 = W_{1,0}$ and $W_{1,1}$

- $\mathcal{C}_2 = W_{1,0}, W_{2,2},$ and $W_{2,3}$

- $\mathcal{C}_3 = W_{2,0}, W_{2,1},$ and $W_{1,1}$

- $\mathcal{C}_4 = W_{2,0}, W_{2,1}, W_{2,2},$ and $W_{2,3}$.

The total energy of any disjoint dyadic MODWPT or DWPT decomposition is equal to the energy of the original sequence, i.e. $||\mathcal{C}_i||^2 = ||X||^2$ for $i = 1, \ldots, P_j$ where $\mathcal{C}_i$ is one possible disjoint DWPT decomposition and P_j is the total number of disjoint dyadic decompositions for a level j decomposition. For instance, in the example above, the third set's energy relation to the original sequence is $||\mathcal{C}_3||^2 = ||W_{2,0}||^2 + ||W_{2,1}||^2 + ||W_{1,1}||^2 = ||X||^2$. For energy conservation to hold, it must mean that the sum of children's energy be equal to that of their parent's for each and very node in the decomposition.

Relevance: The MODWPT is used to decompose a sequence into its fundamental time-frequency components.

3.2 Shannon Entropy as a Best Basis Functional (Stage II)

The purpose of this section is to introduce the DWPT best basis technique and to develop an appropriate extension to the technique for use with the MODWPT.

For a jth level DWPT decomposition, there exists more than $2^{2^{j-1}}$ possible orthonormal bases for $j \geq 1$ [84]. Coifman and Wickerhauser [16] were of the first to

pursue a *best basis* which is in some sense optimal for a given time series by forming an additive cost functional $\Gamma(\cdot)$ to be applied to each node $W_{j,n}$. It is necessary that $\Gamma(\cdot)$ be additive so that the sum of the children's costs can be legitimately compared to that of the parent. A minimization of the nodal costs $\mathcal{C}_{j,n}$ over a set of disjoint nodes spanning the normalized frequency range $f \in [0, 1/2]$ is seen to produce an "optimal" basis. This can be expressed in a mathematical formulation as

$$\min_{\mathcal{C}} \sum_{(j,n) \in \mathcal{C}} \mathcal{C}_{j,n} \tag{3.13}$$

where $\mathcal{C}$ is the space spanned by all disjoint dyadic decompositions corresponding to orthonormal DWPTs of some time series $\{X_t\}$ for $t = 0, \ldots, N - 1$ ($N = 2^j$ for $j \in \mathbb{N}$).

There are many cost functionals that have been proposed but the Shannon entropy is arguably the most widely used and is defined as

$$\mathcal{S}(p) = -\sum_i |p_i| \log |p_i| \tag{3.14}$$

where $|p_i| \log |p_i| \equiv 0$ for $p_i = 0$ [16, 77]. The Shannon entropy has been qualitatively associated with the *amount of uncertainty* in the data or as a means of measuring the *flatness of the distribution*. These qualities can be seen in a statistical sense through a gambling analogy. Imagine, for example, a roulette wheel with ten spokes, each of which is assigned an arbitrary probability p_i such that $\sum p_i = 1.0$ for $i = 1, \ldots, 10$. The most certain case is one where one spoke is assigned a probability $p_i = 1.0$ while the rest are assigned probability $p_j = 0.0$ for $j \neq i$ (see left column of Fig. 3.6). The entropy for this case is $S = 0$, meaning that the outcome of each and every spin of the wheel is completely predictable. The corresponding distribution is compact with its support spanning only a single spoke. Conversely, the most uncertain case is one in which each spoke is assigned an equal probability $p_i = 1/r$ for $r = 10$, i.e. a uniform probability distribution over the spokes. The entropy in this case is at

its maximum value, namely $S = ln(r)$. Thus, entropy can be used as a means of characterizing the uncertainty in data, spanning the range of $S = 0$ (most certain) to $S = ln(r)$ (least certain). With regard to compactness of the distribution, $S = 0$ implies a very compact support while $S = ln(r)$ (or large values of S) implies a less compact or "flat" distribution.
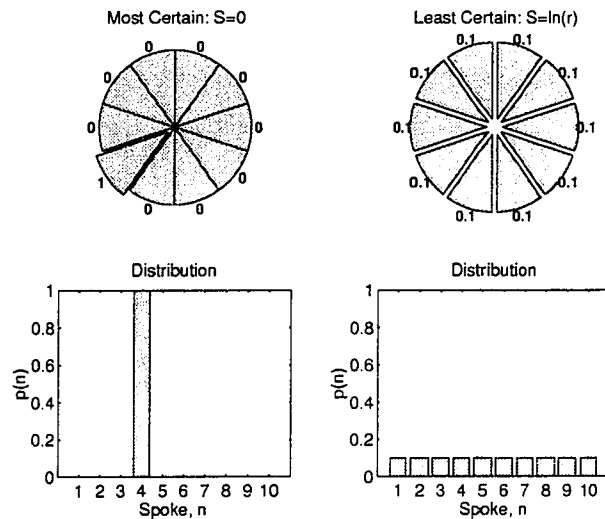
Figure 3.6: A roulette wheel to which is assigned probabilities to each spoke and its associated distribution function.

While probability is an appropriate measure for investigating the entropy of a roulette wheel trial, any relevant additive measure of the data is acceptable for discrete sequences in general. Our goal is to find the most compact means of characterizing a discrete sequence in terms of its time and frequency content. Since the DWPT coefficients are localized in frequency and in time, they are suitable measures for entropy minimization. We will in fact use the energy of the wavelet packet coefficients as our measure such that $p_i = W_i^2$ where i is the global node index representing the ith wavelet packet node. The most compact time-frequency

representation is found by Eq. 3.13 where

$$\mathcal{C}_{j,n} = -\sum_{t=0}^{N_j-1} W_{j,n,t}^2 \log W_{j,n,t}^2 \qquad (3.15)$$

is the set of entropies assigned to each node in the decomposition. As an example, the X state of the Lorenz system in the chaotic parameter regime is decomposed via a level $J = 4$ DWPT using Daubechies' LA8 filter set (Fig. 3.7). It can be seen that the lower frequencies tend to dominate the Lorenz response in this case, and thus its energy lies mainly on the left side of the transform plot. One would expect then that an optimal description would involve nodes more closely associated with a standard DWT which is highlighted by the thick boxes in Fig. 3.7. The corresponding best basis was calculated using a Shannon energy functional. The sample mean was subtracted from the original sequence $\{X_t\}$ and was normalized such that $||X||^2 = 1.0$. An L^2 normalization establishes a means of comparison on the basis of energy decomposition between arbitrary signals. In Fig. 3.8, a graphical tree is shown illustrating the best basis decomposition. Each node is labeled with its corresponding nodal position in (j, n) coordinates where j is the decomposition level and n is the local node index. At the top of the tree (location $(0, 0)$) sits the original signal with its corresponding entropy shown just below the coordinate label. Below the top node are the descendant nodes and their corresponding entropies. By virtue of the optimization process, no parent can have an entropy larger than the sum of its children's entropies. The nodes comprising the best basis lie on the lower perimeter of the tree. In this case, the best basis for the Lorenz X state is the DWT, just as surmised. If, however, the DWPT was carried out to greater scales, it is not so clear that the DWT would emerge as the transform of choice. As a second example, the DWPT decomposition of a blue noise realization generated by the Gaussian Spectral Synthesis Method (GSSM) [71] is shown in Fig. 3.9 and its corresponding best basis in Fig 3.10. Blue noise is an example of a $1/f^\alpha$ process, i.e. a stochastic process whose power spectral density goes as $S_X \sim f^{-\alpha}$. For blue noise,
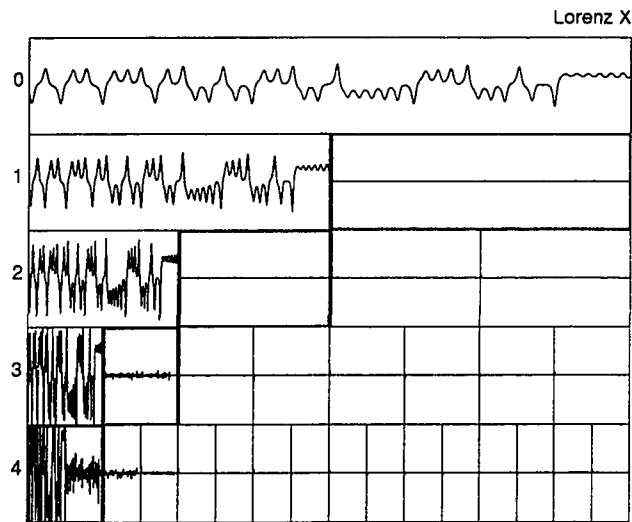
Figure 3.7: A level $J = 4$ partial DWPT decomposition of the X state of the Lorenz system in a chaotic regime using Daubechies' LA8 filters. The lower frequency content of the data dominates.
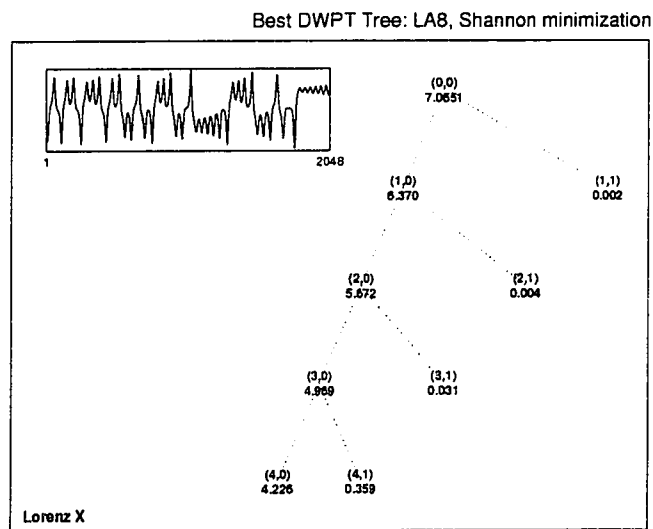
Figure 3.8: The best basis DWPT tree for the Lorenz X state with no noise.

we have $\alpha < 0.0$, yielding dominant high frequency components as can be seen in its DWPT decomposition and best basis tree which branches more to the right and away from the DWT branching.
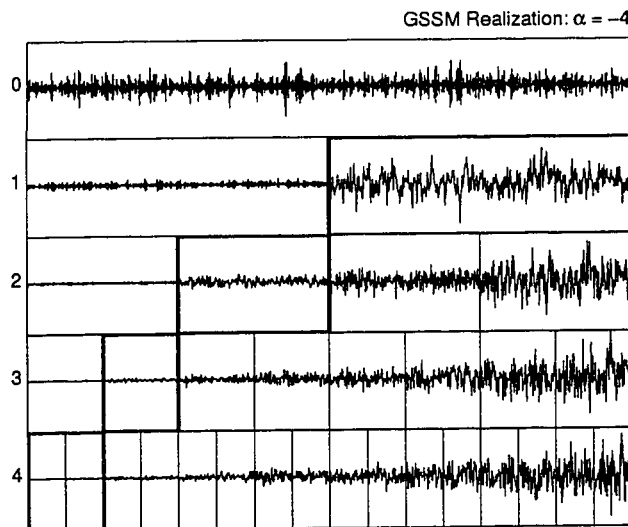
Figure 3.9: DWPT decomposition for a $1/f^{-4}$ (blue noise) GSSM realization.

3.2.1 Shannon's Entropy and the MODWPT

The MODWPT is a highly redundant nonorthogonal decomposition and cannot be used without modification to calculate a best basis via a Shannon entropy functional. The redundant use of the elements of X in the decomposition increases the overall entropy relative to that computed by the DWPT as can be seen in Fig. 3.11.

Using the fact that the MODWPT coefficients are related to those of the DWPT by Eq. 3.12, the MODWPT coefficients can be used to calculate an equivalent

Best DWPT Tree: LA8, Shannon minimization
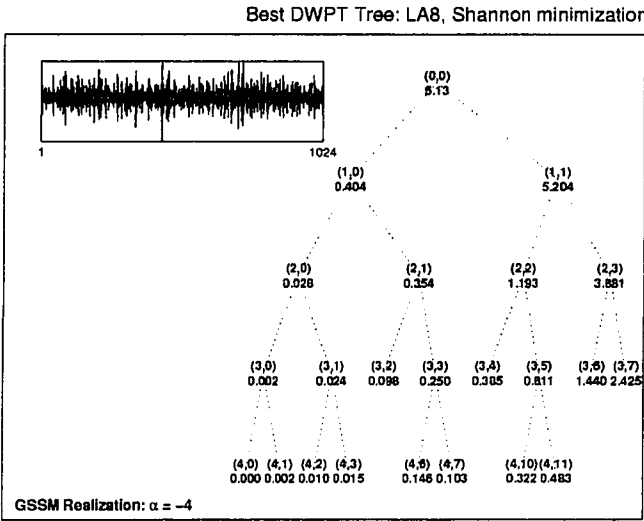
Figure 3.10: The best basis DWPT tree for a $1/f^{-4}$ (blue noise) GSSM realization using a Shannon entropy functional.
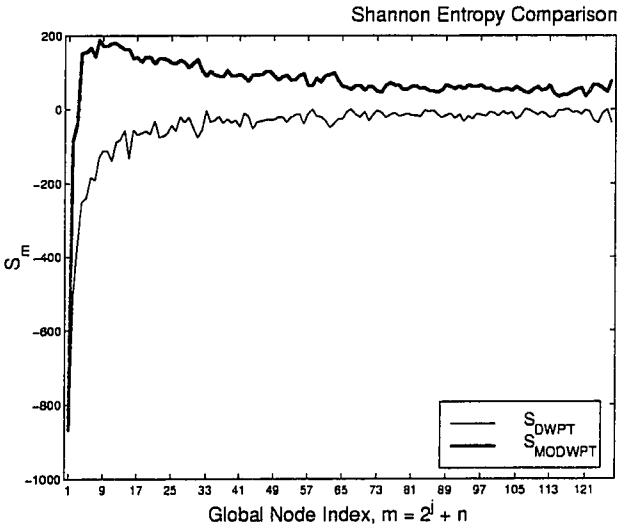
Figure 3.11: Comparison of the Shannon entropy for each node in a level $J = 6$ MODWPT and DWPT decomposition using Daubechies' LA8 filters.

DWPT Shannon entropy functional

$$
\begin{aligned}
\mathcal{C}_{j,n} &= -\sum_{t=0}^{\tilde{N}_j-1} 2^j \widetilde{W}^2_{j,n,2^j(t+1)-1} \log 2^j \widetilde{W}^2_{j,n,2^j(t+1)-1} \qquad (3.16) \\
&= -2^j \sum_{t=0}^{\tilde{N}_j-1} \widetilde{W}^2_{j,n,2^j(t+1)-1} \left(j \log 2 + 2 \log \widetilde{W}_{j,n,2^j(t+1)-1} \right)
\end{aligned}
$$

where $\tilde{N}_j = \lfloor N/2^j \rfloor$. If N is a power of 2, we obtain exactly the DWPT Shannon entropy. For arbitrary length sequences, however, this restriction may be released with the understanding that the resulting entropy will be greater than or equal to that computed via the DWPT coefficients simply because we are including more energy from the original sequence in the entropy calculation. The limitation on N in the DWPT is mainly due to the fact that at every decomposition level a critical decimation (downsampling by 2) is performed. Since there is no such restriction placed on the MODWPT sequence lengths, an "extended" entropy estimation can be carried out using the remaining coefficients beyond the dyadic length. However, there will (most likely) be a difference in the best DWPT and MODWPT bases in situations where $N - 2^{\lfloor \log_2 N \rfloor}$ is large. We can quantify this difference by a *dyadic divergence coefficient* $\delta_D(N) \equiv 2(N - 2^{\lfloor \log_2 N \rfloor})/N$ where δ_D is confined to the range $0 \leq \delta_D < 1$. If N is dyadic, $\delta_D = 0$ while for N far from $2^{\lfloor \log_2 N \rfloor}$, $\delta_D \to 1.0$. Figure 3.12 shows the difference between best trees for $\delta_D = 0.0, 0.25, 0.50, 0.99$ using a random walk sequence. The plot in the upper left corner of the figure is the same tree obtained using the DWPT while the remaining plots incorporate points neglected by the DWPT in calculating entropy values.

The best MODWPT tree can vary as a function of signal length depending upon the stationarity of $\{X_t\}$ and whether or not we are using a large enough sample size to adequately characterize its statistics. A random walk was specifically selected as a test sequence because it comes from a $1/f^\alpha$ process ($\alpha = 2.0$) which is a nonstationary process. Blue noise, however, is a stationary $1/f^\alpha$ process and its
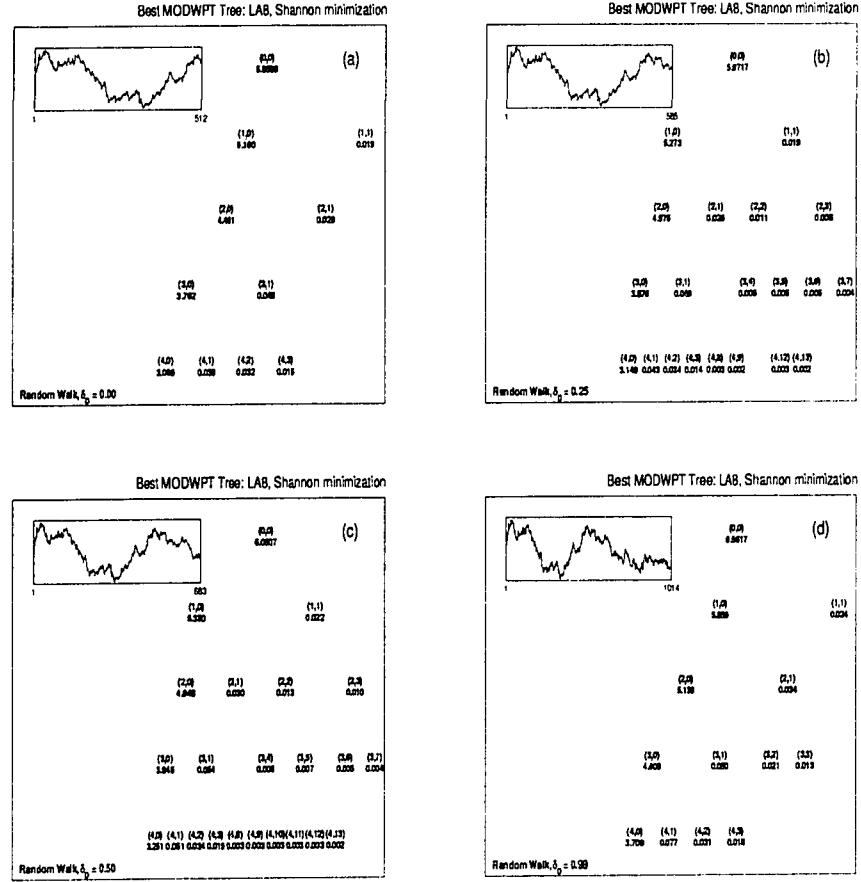
Figure 3.12: Best MODWPT trees for a random walk with (a) $\delta_D = 0.0, N = 512$ (b) $\delta_D = 0.25, N = 585$ (c) $\delta_D = 0.50, N = 683$ and (d) $\delta_D = 0.99, N = 1014$.

corresponding best tree should not (on average) be as sensitive to sequence length as are random walks. Figure 3.13 demonstrates the dependence of the MODWPT best basis on sequence length for a blue noise realization generated by the GSSM for $\alpha = -4$. By inspection, the degree of tree variability is less for the blue noise than that it is for the red noise and may be a reflection of the stationarity of the governing processes.

A more stringent mathematical argument can be made regarding the best tree variability in nonstationary colored noise realizations. It is shown in Sec. 4.5 that the variance of a random walk process $\{X_t\}$ is a function of time and therefore is by definition a nonstationary process. The entropy for $\{X_t\}$ is defined as

$$S = -\sum_t X_t^2 \log X_t^2 \qquad (3.17)$$

and therefore

$$E\{S\} = -\sum_t E\{X_t^2 \log X_t^2\}. \qquad (3.18)$$

Since $E\{S\}$ is a function of the variance, and since the variance is a function of time, the best basis must also fluctuate as a function of time.

Relevance: A best basis will be selected by optimizing a MODWPT Shannon entropy functional. The nodes corresponding to the best basis form an optimal disjoint coverage of the normalized frequency range $[0, 1/2)$.
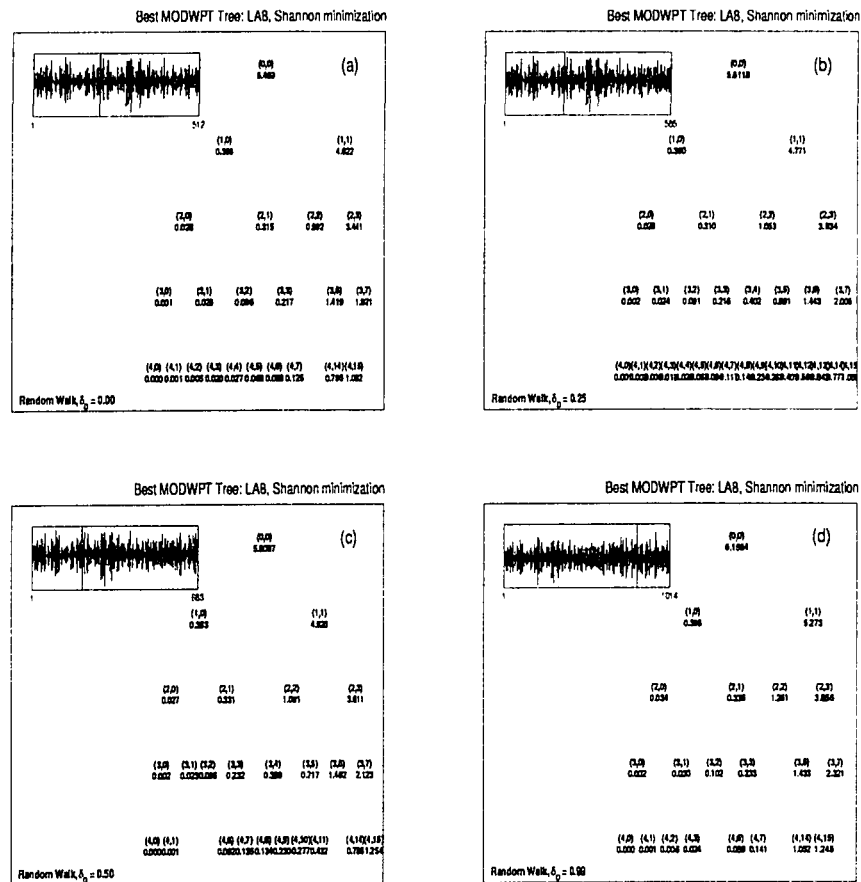
Figure 3.13: Best MODWPT trees for a blue noise GSSM realization ($\alpha = -4.0$) with (a) $\delta_D = 0.0, N = 512$ (b) $\delta_D = 0.25, N = 585$ (c) $\delta_D = 0.50, N = 683$ and (d) $\delta_D = 0.99, N = 1014$.

3.3 Obtaining an Optimized Node Set through Nodal Energy Thresholding (Stage III)

The purpose of this section is to introduce a technique which reduces the set of nodes corresponding to the best MODWPT basis. A reduction in nodes lessens the computational burden of the KDM described in Sec. 3.4. The optimization is achieved by imposing an admissible energy threshold δ_E on the nodes corresponding to the MODWPT best basis. Any node that does not exceed δ_E is ignored and, unless specified otherwise, a minimum of 3 nodes are selected for each basis creating an *optimally reduced nodal set*. As an example, Fig. 3.14 shows the nodal energies of the best basis nodes for a chaotic response from the Lorenz system contaminated with Gaussian uniform noise. The nodal energies are normalized and sorted from strongest to weakest. A threshold of 0.2 is applied and the nodes that exceed the threshold are listed in the upper right hand corner of the figure. Out of the 32 nodes which comprise the best basis, only 4 are selected for further investigation. The threshold value is somewhat arbitrary as the nodal energy depends entirely on the energy distribution of the original sequence. Allowing weaker nodes into the optimized set by lowering the energy threshold results in a heavy computational burden in the KDM (Stage IV of the MODRA) while contributing very little to the overall energy of the original signal. Conversely, setting the energy threshold at too high a level will result in an overreduction of signal energy, destroying the dynamics of the original signal and possibly invoking spurious D_2 results. Numerical tests indicate that a threshold of 0.2 is appropriate for the sequences examined herein.

Relevance: Admissible energy thresholding produces an optimal set of nodes characterizing the MODWPT best basis and reduces the number of relevant coordinates.

Figure 3.14: Normalized and ranked nodal energies corresponding to the MODWPT best basis for a noisy chaotic sequence. A thresholding of 0.2 eliminates all but 4 strong nodes which are listed in the upper right corner of the plot. The nodes for the best basis are listed on the abscissa in (j, n) form.
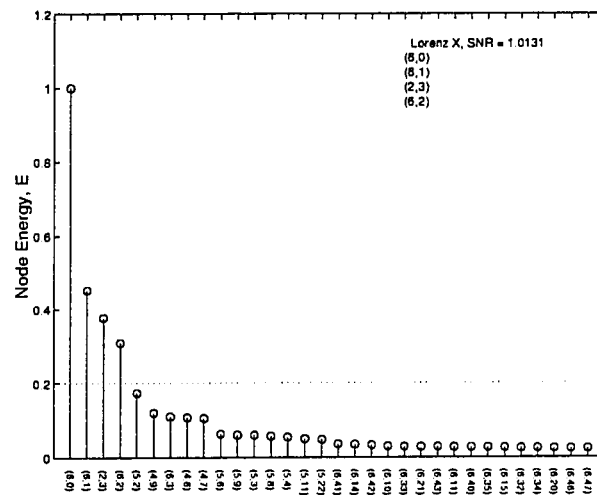
3.4   Kaplan's Determinism Test (Stage IV)

The purpose of this section is to introduce Kaplan's Determinism algorithm (KDM); a technique used to infer determinism [37, 36].

Separating deterministic dynamics from stochastic noise is an arduous task. This is especially true of chaotic sequences as it is often difficult to distinguish any sense of order or fixed pattern in chaotic time histories. As a result, chaos can be misconstrued as stochastic noise. Fig. 2.2 shows an example of four time histories, three of which are deterministic chaotic and one of which is pseudo-random Gaussian noise. When embedded in the phase space, however, the deterministic sequences are easily differentiated from random sequences as distinct patterns emerge that are not present in the embeddings of the random data (Fig. 2.1). Kaplan developed a method which exploits the phase space continuity of deterministic sequence embeddings and is referred to as *Kaplan's Determinism Method (KDM)* hereafter.

The basis of KDM works by checking whether or not the images of unique neighbors remain close on average across the attractor[1]. Kaplan poses the question, "If two points z_j and z_k are very close to one another, are their images z_{j+1} and z_{k+1} also close together?" By assuming a continuity in the flow created by a deterministic system, it is logical to assume that the average Euclidean distance between images will differ from that created by *surrogate* data [96]; that is, data that has been randomized in such a way as to provide a basis of comparison for a discriminating statistic or method (see Sec. 3.4.1 for a detailed description of Theiler's surrogate data algorithms). The difference between surrogate and raw data is more easily distinguished with neighbors that are initially "close" to one another. The neighbors considered should not include points from the same trajectory as they may be highly correlated, thereby violating a coordinate independence rule implicit in the

---

[1]An image of a reference point in space is a point located in the future along the same orbit. While images are assumed to be located closely in time to the reference point they are not necessarily close in space as fast flow can separate them significantly.

definition of D_2. Kaplan defines a Euclidean distance image test by averaging over the whole of the data embedded in space:

$$\delta_{j,k} = |z_j - z_k| \tag{3.19}$$

$$\epsilon_{j,k} = |z_{j+\kappa} - z_{k+\kappa}| \tag{3.20}$$

$$e(r) \equiv \overline{\epsilon_{j,k}} \quad \text{for } j, k \text{ s.t. } r \le \delta_{j,k} < r + \Delta r \tag{3.21}$$

where Δr is the Euclidean bin size. A cumulative average of $\epsilon(r)$ is formed by

$$E(r) \equiv \sum \overline{\epsilon(r)}. \tag{3.22}$$

If there exists a distinct separation of $E(r)$ from surrogate data, it implies that the signal is deterministic. Figure 3.15 shows a graphical representation of KDM.

An example of KDM is shown in Fig. 3.16. Each subplot displays the results for different embedding dimensions. starting from $D_e = 1, \ldots, 7$. The filled circles represent the average cumulative image distance $E(r)$ as a function of scale while the boxes represent the range of E for 50 surrogate sets using Theiler's amplitude adjusted Fourier technique (AAFT). The centerline of each box represents the mean, while the half box height is one sample standard deviation from the mean. The outliers denote the extremities of E for the surrogate data. If a box appears as a line it means that no images were found to be within a distance of r from its preimage. These blank areas tend to occur more and more with an increase in embedding dimension for random data since random embeddings tend to fill the space allotted. A coverage or overlapping of the dots with the boxes suggests that the original data cannot be differentiated from the surrogate data, and implies that the original data must be stochastic in origin. A separation implies that the original signal is deterministic. The upper right plot of Figure 3.17 summarizes the results of Fig. 3.16 through a *violation* measure. A violation V_{D_e} is the fraction of dots occluded by the surrogate boxes over the span of the dots for embedding dimension D_e. A

98
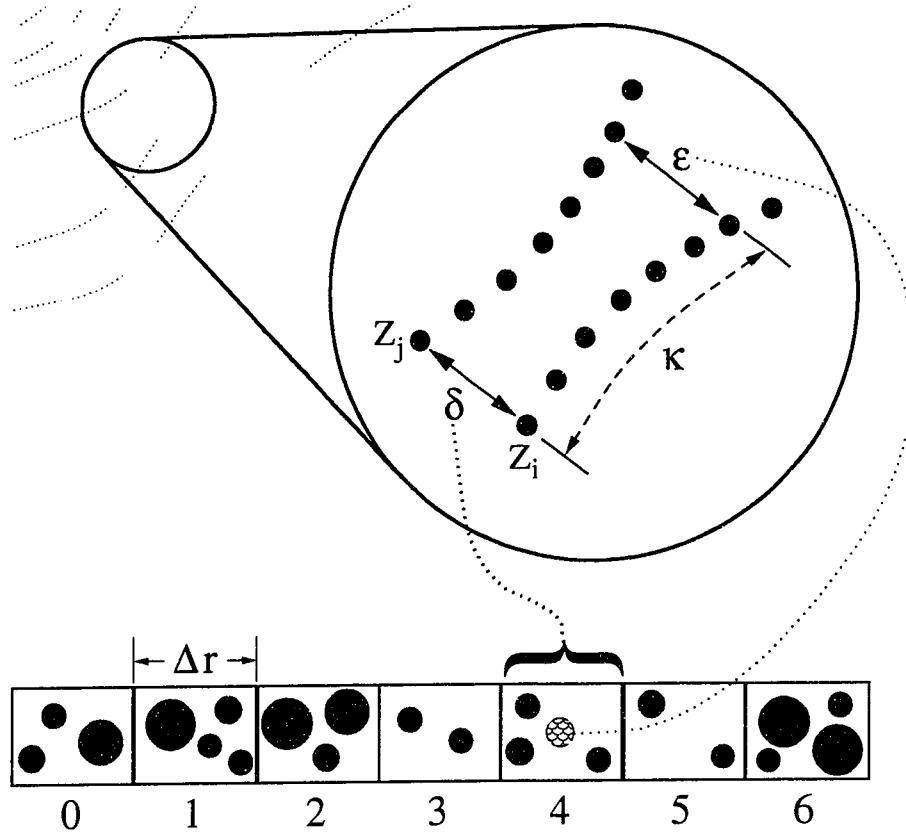
Figure 3.15: A graphical illustration of Kaplan's determinism algorithm. The distance between neighbors $\delta_{i,j}$ decides which bin should be used for recording the Euclidean distance $\epsilon_{i,j}$ between their corresponding images found κ points later in time along respective trajectories.
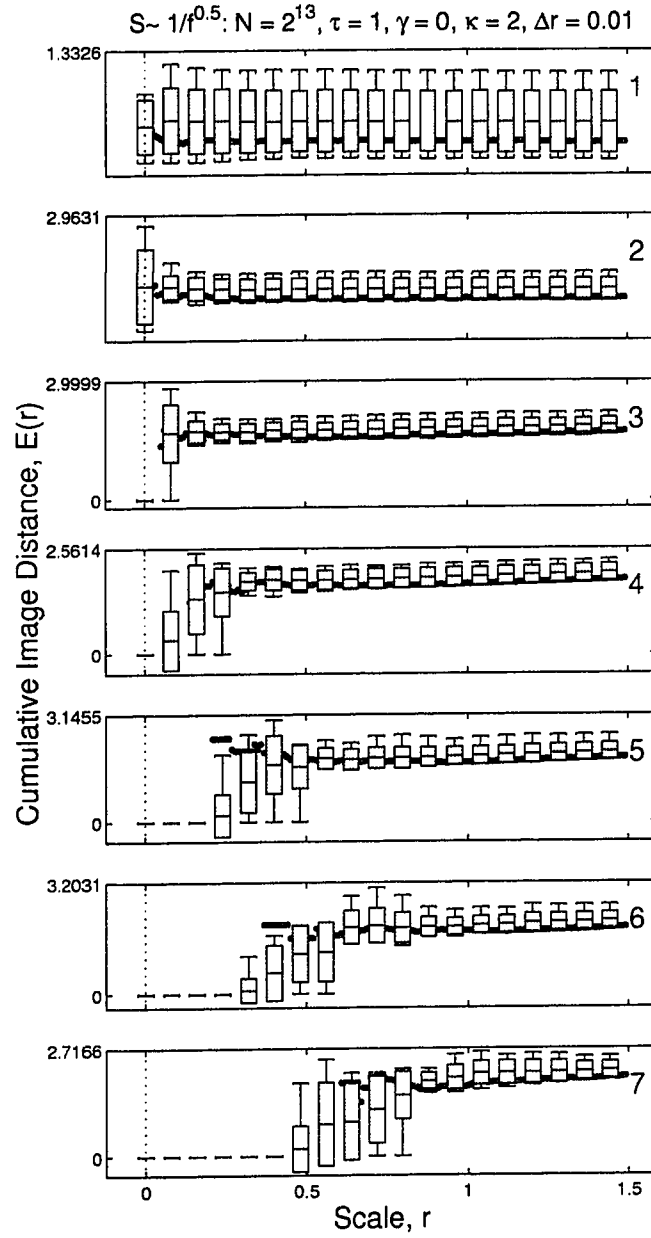
Figure 3.16: The results of KDM on blue noise generated by the Gaussian Spectral Synthesis Method for $\alpha = -0.5$ with length $N = 2^{13}$, time lag $\tau = 1$, trajectory lead $\gamma = 0$, image lead $\kappa = 2$, Euclidean bin size of $\Delta r = 0.0$, and embedding dimensions $D_e = 1, \ldots, 7$. See text for details.

violation is measured with regard to the outliers $V_{D_c}^{out}$ or one standard deviation from the mean (half height of each box) $V_{D_c}^{\sigma}$. According to Kaplan, any violation in the lower dimensions is evidence to conclude randomness in the original data [39]. However, tests conducted for this research show that KDM does not work for red noise realizations, i.e. for $1/f^\alpha$ realizations for $\alpha \geq 1.0$ (see Fig. 3.17). By definition, red noise has powerful low frequency components which tend to smooth out red noise trajectories in the phase space. Kaplan's test fails because on average red noise images *do* tend to stay close to their preimages in the phase space, tricking the technique into categorizing red noise as deterministic. The effects of colored noise on measures involving the phase space is thoroughly examined in Chapter 4.

3.4.1 Theiler's Surrogate Data Algorithms

The erratic fluctuations observed in experimental time series stem from a mix of dynamic influences: chaos, nonchaotic but still nonlinear determinism, and noise[96]. To determine the underlying dynamic process of a time series, *surrogate* data is used in an attempt to isolate one of these influences. Surrogate data is produced by generating a time series which is consistent with a *null hypothesis* or potential explanation we seek to show is inadequate in characterizing the data. A discriminating statistic is then used to quantify some aspect of the data. The null hypothesis is rejected if the discriminating statistic for the observed data sufficiently differs from that of an ensemble of surrogates.

Theiler has developed two main algorithms for producing surrogate data: the *Fourier Transform (FT)*, and the *Amplitude Adjusted Fourier Transform (AAFT)* which are summarized below [96].

Method of Fourier Transform (FT)

The null hypothesis is that the data stems from a linear Gaussian process. This method preserves the spectrum of the observed sequence. This is accomplished by
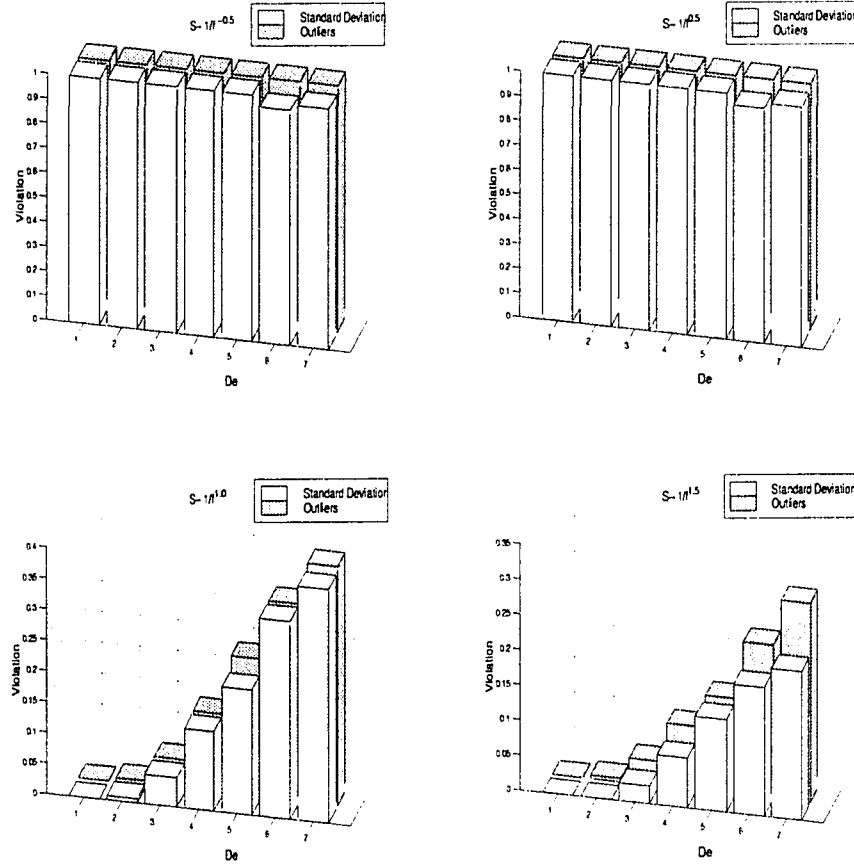
Figure 3.17: Kaplan violation summaries of $1/f^\alpha$ realizations for $\alpha = -0.5, 0.5, 1.0$, and 1.5 clockwise from upper left. The upper row is blue noise and the lower row is red noise. No violation exists for the first few embedding dimensions for red noise, and Kaplan's technique subsequently and falsely identifies it as deterministic.

multiplying the DFT coefficients of the observed data by $e^{i\phi}$, where ϕ is independently chosen for each frequency on the interval $[0,2\pi]$ as a uniformly distributed deviate. In other words, the phase of the observed data is randomized to form the surrogate data. Typically, a window is applied to the data prior to the DFT operation to attenuate spectral leakage.

Amplitude Adjusted Fourier Transform (AAFT)

The null hypothesis is the same for that of the FT method. The AAFT preserves the probability distribution of the observed time series. The algorithm begins by sorting a Gaussian random sequence such that the order of the deviates (ranked by amplitude) is commensurate with that of the original sequence. The phases are then randomized via the FT method and the resulting sequence is rescaled to return the distribution of the original time series.

An example of the Theiler's surrogate methods is shown in Fig. 3.18 for a chaotic Lorenz sequence. The values to the right of the plots are the cross-correlation coefficients which quantify the statistical similarity of the surrogate data with the original time series. A correlation value of 1.0 means that the surrogate data is equivalent to the original in regard to the corresponding measure. The discrete wavelet transform shuffle (DWTS) was developed for this dissertation and is used strictly as a means of comparison to Theiler's techniques. The DWTS works by randomly shuffling the order of wavelet coefficients followed by and inverse DWT operation.

Relevance: Theiler's AAFT algorithm is used to differentiate deterministic response from stochastic noise in both the Kaplan's determinism algorithm and the correlation dimension.
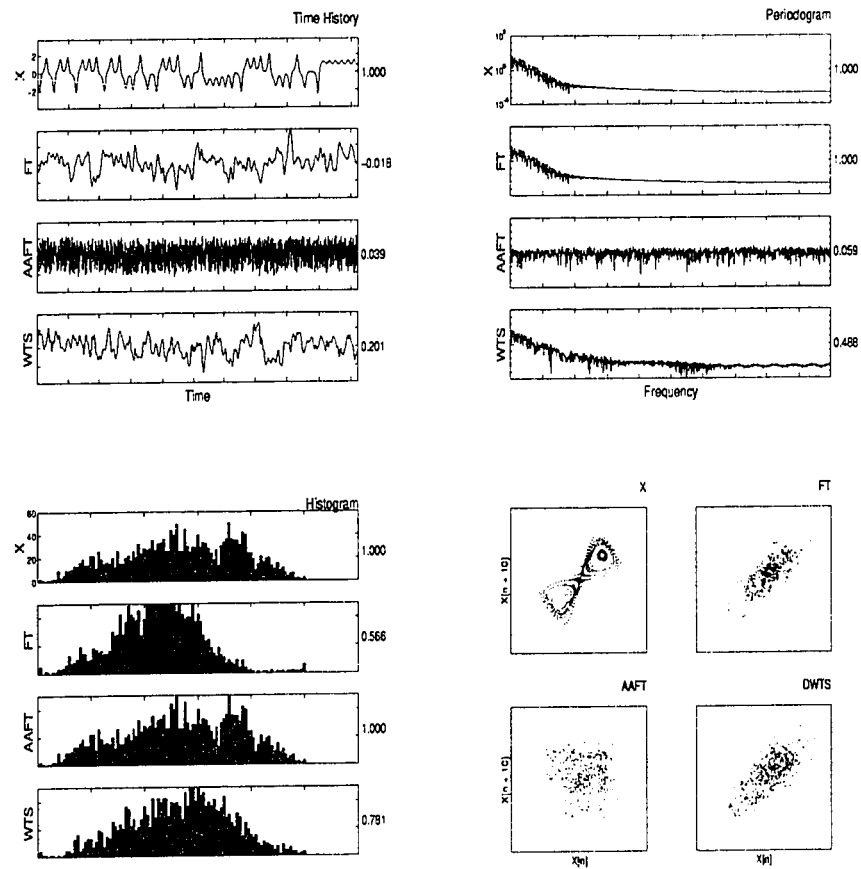
Figure 3.18: An example of Theiler's FT and AAFT surrogate data methods. Shown clockwise from the top left is the time series, periodograms, phase plane embeddings, and histograms of the original and surrogate data. See text for details.

3.4.2 KDM on MODWPT Nodes

The MODWPT nodes have two special properties which can be exploited for use with the KDM: (1) For an N point input sequence $\{X_t\}$, each node is also an N point sequence and (2) Daubechies least asymmetric wavelet and scaling filters have approximately linear phase. The approximate linear phase property means that the wavelet packet coefficients at any node $\widetilde{W}_{j,n}$ can be advanced to achieve an approximate zero phase filtering operation. This is important in that wavelet packet nodal events in time can be meaningfully lined up with those in the original signal. Furthermore, it is imperative that these phase relations remain in tact if there is to be a meaningful analysis made in the phase space (as with the KDM), i.e. if the phases were scrambled it would destroy the continuity of the phase space embedding and any further attempt at detecting determinism would be lost. Phase randomization is in fact one of the most popular methods by which surrogate data is developed. Secondly, for measures such as KDM or D_2, it is important that there be a large number of orbits in the phase space to ensure good statistics on the measure. The number of orbits we can obtain with any given wavelet packet node is maximized if no points are lost in the transformation. This is not the case, for example, with the DWT or DWPT which loses half its length at every decomposition level. The combination of obtaining an approximate zero phase transformation without losing any sequence length allows us to use the KDM on each node of the optimal node set to classify its dynamics origin. If node $\widetilde{W}_{j,n}$ is found to be deterministic it is kept; otherwise it is discarded. To automate the selection process, a rejection tolerance δ_{kap} is established and a weighted averaged violation measure μ_V formed by

$$\mu_V = \frac{\sum_i \rho_i \overline{V}_i}{\sum_i \rho_i} \quad \text{for } i = 1, \dots, D_e \tag{3.23}$$

where ρ_i is an exponential weighting function

$$\rho_i = e^{D_e - i + 1} \quad \text{for } i = 1, \dots, D_e \tag{3.24}$$

and $\overline{V}_i$ is the mean of the violation indices

$$\overline{V}_i = \frac{V_i^{out} + V_i^{\sigma}}{2} \quad \text{for } i = 1, \ldots, D_e. \tag{3.25}$$

The exponential weighting factor ρ_i ensures that more emphasis is placed on violations in lower dimension where KDM is most sensitive. The definition of $\overline{V}$ is somewhat arbitrary; here we use the average of the two violation indices. The rejection criterion is established as

$$\begin{aligned} \textbf{Rejection} &\quad \mu_V > \delta_{kap} \\ \textbf{Acceptance} &\quad \mu_V \leq \delta_{kap}. \end{aligned}$$

The Kaplan rejection threshold was determined through computer experiments where $\delta_{kap} = 0.02$ was found to be an appropriate selection. For Gaussian random noise a typical violation measure was $mu_V \approx 0.75$ for higher frequency nodes while it was substantially lower for lower frequency nodes. The rejection threshold was set just below the minimum value recorded for pure noise experiments.

Relevance: Kaplan's determinism algorithm eliminates stochastic nodes in the optimized node set.

3.5 MODWPT Synthesis (Stage V)

The final stage of the MODRA is to synthesize the MODWPT nodes which passed the KDM test. Since the MODWPT is an invertible transform, a node $\widetilde{W}_{j,n}$ can be synthesized with all other nodes ignored (zeroed) to form an approximation to $\{X_t\}$ denoted as $\widetilde{\mathcal{D}}_{j,n}$ or *detail* signal at node (j, n). The collection of detail signals for any dyadic disjoint decomposition are summable and can be used to reconstruct $\{X_t\}$. A useful property of the MODWPT details is that they are *exactly* zero phase, i.e. events in time can be lined up exactly with the original signal and the phase space
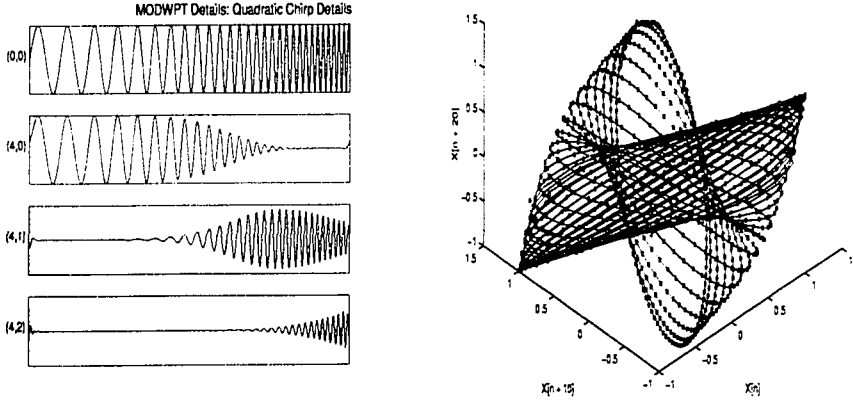
Figure 3.19: Sample MODWPT detail signals and the embedding of the chirp $\{X_t\}$ overlayed by an embedding of the synthesis sequence $\{\tilde{X}_t\} = \tilde{\mathcal{D}}_{4,0} + \tilde{\mathcal{D}}_{4,1} + \tilde{\mathcal{D}}_{4,2}$.

embedding remains in tact. As an example, Fig. 3.19 shows the details $\tilde{\mathcal{D}}_{4,0}, \tilde{\mathcal{D}}_{4,1}$, and $\tilde{\mathcal{D}}_{4,2}$ for a quadratic chirp sequence as well as an embedding of the chirp (lines) overlayed with an embedding of the synthesis $\{\widehat{X}_t\} = \tilde{\mathcal{D}}_{4,0} + \tilde{\mathcal{D}}_{4,1} + \tilde{\mathcal{D}}_{4,2}$ (dots). The chirp has a quadratic increase in frequency as a function of time so the lower frequency content appears in detail $\tilde{\mathcal{D}}_{4,0}$, the midrange frequency content in $\tilde{\mathcal{D}}_{4,1}$, and the high frequency content in $\tilde{\mathcal{D}}_{4,2}$. The oscillations in the original signal line up exactly in time with the corresponding data in the details. The embedding plot shows that the original signal is almost perfectly reconstructed with the synthesis of the three nodes.

Relevance: The deterministic MODWPT nodes are synthesized to form a denoised approximation of the original uncontaminated sequence; thereby completing the MODRA.

3.6 Considerations for Single Variable Embeddings

The MODRA output is considered to be a single variable time series since it represents only one variable of a dynamic system. Our interest is to use the MODRA output as a means to estimate the fractal dimension(s) of the original uncontaminated sequence. To do so requires the estimation of embedding parameters which have influence on the fractal measures. In this section, we discuss the subtleties involved in setting these parameters in the pursuit of robustness and statistical consistency.

There are three main difficulties one encounters in embedding a single variable time series, namely finding the proper embedding dimension, time lag, and trajectory lag. Each one can have a significant influence on the correlation dimension results.

3.6.1 Finding the Proper Embedding Dimension

Typically in experimental situations, only a few variables of a dynamic system are observable and accurately measurable. In the chaotic beam experiment (Sec. 3.8.3), for example, only the beam tip velocity was observed while the displacement and acceleration were ignored. The natural question arises, "Can the dynamics of an original attractor be recreated in the phase space using only a single observed variable of the system?"

A major difficulty with this notion is that the dimension of the original attractor, assuming one exists, is almost never known a priori in experimental situations. One of the more important applications of fractal mathematics is to allow one to determine the minimum number of first order equations needed to adequately describe the dynamics found in the original attractor. If a fractal dimension of D is accurately determined for a given system, then the minimum number of first order equations P needed to describe that system is the next highest integer relative to D. Not knowing P, we cannot know how many physical variables $\{x(t), y(t), z(t), \ldots\}$

108

to measure. Instead, a pseudo-phase space is constructed by using time-delayed measurements of one physical variable, say $x(t)$ for example. A three dimensional pseudo-phase space may be obtained using the digitized version of $x(t)$ as

$$\mathbf{x}_n = \{x(t_0 + n\tau), x(t_0 + (n+1)\tau), x(t_0 + (n+2)\tau), \dots\} \qquad (3.26)$$

where n is a integer representing a normalized time index, t_0 is a reference time, and τ is the *time lag*. To create a pseudo-phase space from a physical experiment, a discrete form of an experimental variable $x(t)$ is obtained by uniformly sampling $x(t)$ to produce $x(kT)$ where T is the sampling period and k is an integer. Then, uniformly shifted versions of $x(kT)$ are used to create the coordinates in Eq. 3.26.

In reconstructing the attractor using measurements of a single variable, the question naturally arises as to what the minimum embedding dimension must be in order to capture the dynamics of the original attractor. Takens theorem states that if the original attractor lives in E-space, then in general one must use a pseudo-embedding space of $2E + 1$ or higher [92, 63].

The dimension in which the (lagged) coordinates are viewed is the *embedding dimension*, D_e. Figure 3.20 demonstrates the embedding of a chaotic sequence in different dimensions. The appropriate D_e is found by determining the minimum dimension for which a nonlinear measure exhibits convergence. As an example, the slopes of correlation integrals are used to estimate D_2 and the dimension at which they converge is the appropriate minimum embedding dimension. Unfortunately, Takens' embedding theorem requires that single variable time series be embedded in a dimension at least twice that of the original attractor. In many cases, D_e is accordingly large ($D_e > 5$) and dramatically increases the computational onus. There are other methods which can be used to disclose the proper embedding dimension for discrete sequences including the method of False Nearest Neighbors [40] and an embedding method proposed by Mañe and Takens [63]. Both of these, however, suffer from problems similar to using correlation integral convergence. We will use
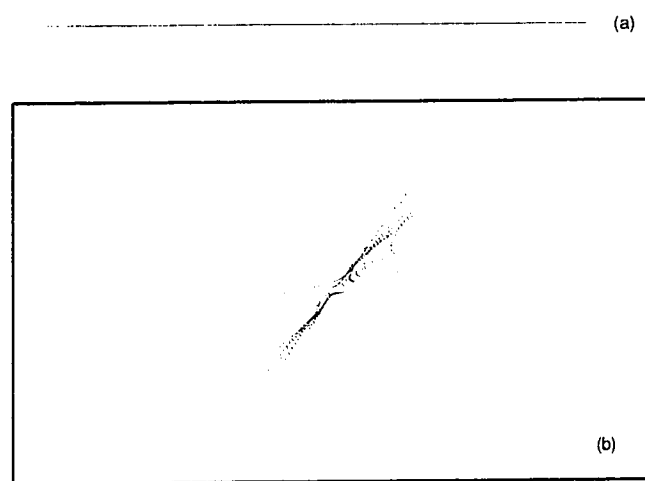
Figure 3.20: Lorenz embedding for dimensions (a) one and (b) two. While there is no discernable structure in the one dimensional embedding, a chaotic attractor is suggested in dimension two.

the correlation integrals to estimate D_2 for embedding dimensions $D_e = 1, \ldots, M$ where $M > 5$ is a suitable choice for the low dimensional chaotic systems used in this dissertation.

3.6.2 Time Lag

The time lag is used to unfold the dynamics of a system using only one observed variable. An embedding of the lagged coordinates maintains the structure of the original attractor if the embedding dimension and time lag are chosen properly. If the time lag is chosen too small, the coordinates $X_k \approx X_{k+\tau} \approx \ldots \approx X_{k+\tau(D_e-1)}$ causing the attractor to stretch along the diagonal in D_e-space and creates a highly correlated coordinate set. If τ is chosen too large, the coordinates lose their dynamic relevance as events in time become nonlocalized. Figure 3.21 demonstrates the effects of using different time lags to embed the data. The autocorrelation function and the
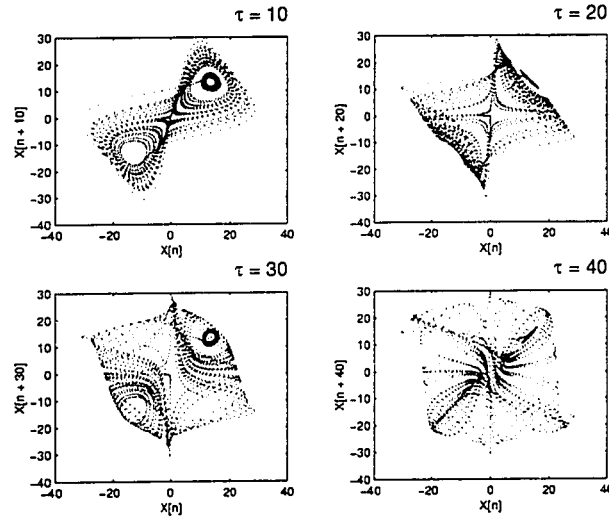
Figure 3.21: The embedding of the X state of the Lorenz system for $D_e = 3$ and $\tau = 10, 20, 30,$ and 40.

average mutual information are common techniques used identify an appropriate time lag [1, 2, 24, 88, 94, 19]. Figure 3.23 shows an example of three candidates for a chaotic sequence using the autocorrelation function method. Unfortunately, an ideal and efficient algorithm for computing an appropriate time lag has not been established. Improper lagging can have a significant influence on nonlinear measures. For example, in Fig. 3.22 the slopes of correlation integrals have a strong dependence on the time lag. For $\tau = 1$ the correlation integrals appear to converge, while for $\tau = 16$ they do not.
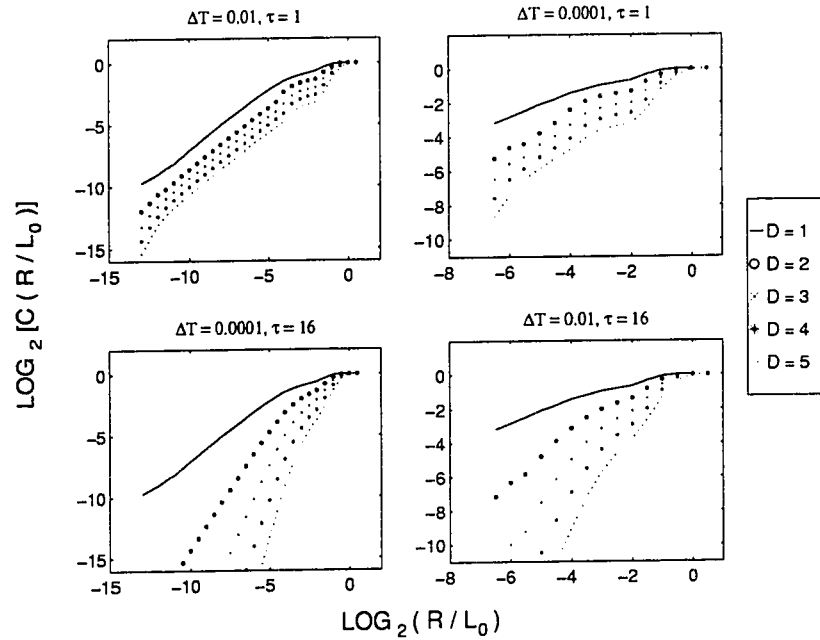
Figure 3.22: The correlation results for a Poincaré section of the embedding of the X state for the Lorenz system in pseudo-phase space for embedding dimensions $D_e = 1, 2, \ldots, 5$. These plots demonstrate the effect of sampling period ΔT and temporal lag τ on the correlation integrals. Notice how improper lagging adversely affects the slope of the correlation integrals. The upper left plot is seen to be correct.

In this dissertation, $\tau = 2.5\,T_e$ where T_e is the autocorrelation time or first $1/e$

falloff in the autocorrelation function. This is based on the suggestion of Albano et al [2] on the basis that this lag produced the most consistent correlation dimension results (in numerical experiments) and is easy to evaluate.
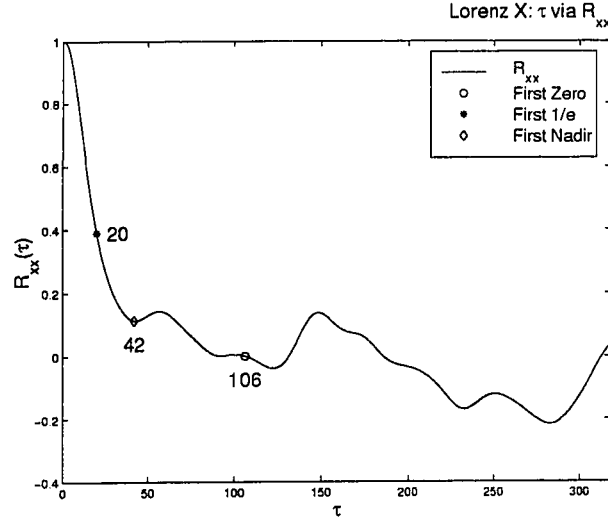
Figure 3.23: The autocorrelation function of the X state of the Lorenz system in a chaotic regime. Shown are three common time lags used to produce the pseudoembedding of single variable time series.

3.6.3 Trajectory Lag

When performing a time series analysis of a dynamic flow in the phase space, it is important to exclude neighboring points along the same trajectory. The inclusion of these images can produce a strong correlation amongst neighbors and can result in spuriously convergent correlation integrals. Visually one can estimate the proper trajectory lag by examining the pseudo-phase space or by identifying the dominant frequency in the spectrum of the series. With higher order systems the attractors can appear "cloudy" in the phase space, making the selection a more difficult process. Consequently, the trajectory lags used in this dissertation were chosen on a

signal-to-signal basis by visually examining the phase space embedding and choosing appropriately.

Relevance: The proper selection of the embedding dimension, time lag, and trajectory lag is essential for robust and consistent estimation of the generalized fractal dimensions. Particular care has been taken to avoid inappropriate selection of these parameters.

3.7 Singular Value Decomposition

The purpose of this section is to discuss the relevance and benefits of using singular value decomposition in embedding techniques.

It is well known that any MxN matrix $\underline{\mathbf{X}}$ for $M \geq N$ can be expressed as

$$\underline{\mathbf{X}} = \underline{\mathbf{U}}\underline{\Sigma}\underline{\mathbf{V}}^{\mathbf{T}} \tag{3.27}$$

where $\underline{\mathbf{V}}^{\mathbf{T}}$ is the transpose of $\underline{\mathbf{V}}$. The right hand side of Eq. 3.27 is referred to as the *singular value decomposition* of the matrix $\underline{\mathbf{X}}$ [2]. $\underline{\Sigma}$ is an MxN diagonal matrix

$$\Sigma_{i,j} = \delta_{i,j}\sigma_i \quad \text{for} \quad \begin{matrix} i = 1, \ldots, M \\ j = 1, \ldots, N \end{matrix}. \tag{3.28}$$

The elements σ_i of $\underline{\Sigma}$ are known as the *singular values* of $\underline{\mathbf{X}}$. $\underline{\mathbf{U}}$ is an MxM orthonormal matrix,

$$(\underline{\mathbf{U}}^{\mathbf{T}}\underline{\mathbf{U}})_{i,j} = \delta_{i,j} \quad \text{for} \quad \begin{matrix} i = 1, \ldots, M \\ j = 1, \ldots, M \end{matrix}. \tag{3.29}$$

The columns of $\underline{\mathbf{U}}$ are called the *left singular vectors* of $\underline{\mathbf{X}}$. $\underline{\mathbf{V}}$ is an NxN orthonormal matrix,

$$(\underline{\mathbf{V}}^{\mathbf{T}}\underline{\mathbf{V}})_{i,j} = (\underline{\mathbf{V}}\underline{\mathbf{V}}^{\mathbf{T}})_{i,j} = \delta_{i,j} \quad \text{for} \quad \begin{matrix} i = 1, \ldots, N \\ j = 1, \ldots, N \end{matrix}, \tag{3.30}$$

whose columns are called the *right singular vectors* of $\underline{\mathbf{X}}$. The number of nonzero diagonal elements of $\underline{\Sigma}$ is equivalent to the rank of $\underline{\mathbf{X}}$, hereby denoted as r.

A graphical interpretation of singular value decomposition is shown in Fig. 3.24 where a unit circle $\underline{S}$ is mapped by a matrix $\underline{X}$: $\underline{S} \rightarrow \underline{XS}$. Any real matrix $\underline{X}$ maps a unit sphere into a hypersphere of dimension r [97]. The lengths of the semimajor
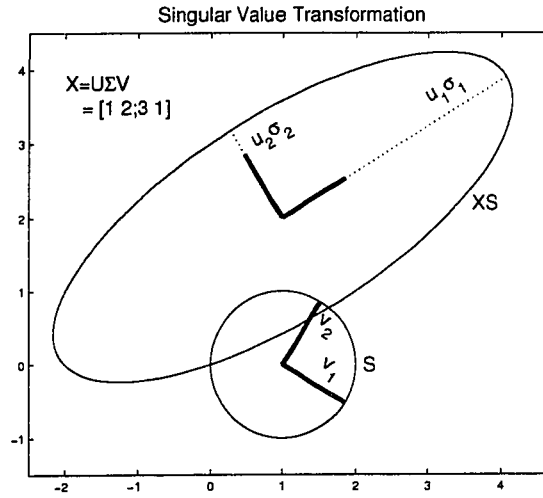
Figure 3.24: Transformation of the unit circle $\underline{S}$ by the map $\underline{S} \rightarrow \underline{XS}$. The result is an ellipse whose semimajor and semiminor axes are defined by the singular values and left singular vectors of $\underline{X}$.

and minor axes are σ_1 and σ_2, respectively. The axes directions are given by the columns of $\underline{U}$ with the first column corresponding to σ_1 and so forth.

A singular value decomposition is useful in the embedding process for two reasons: (1) it produces an independent coordinate set which spans the embedding space and (2) it can be used to reject low level noise contamination and reduce the necessary embedding dimension [2]. Let $\underline{X}$ be a trajectory matrix whose columns are the

embedding coordinates,

$$\underline{\mathbf{X}} = \frac{1}{\sqrt{M}} \begin{bmatrix} \vec{\xi}_1 \\ \vec{\xi}_2 \\ \vdots \\ \vec{\xi}_M \end{bmatrix} \tag{3.31}$$

where $\vec{\xi}_k = (X_k, X_{k+\tau}, \ldots, X_{k+\tau(D_e-1)})$ for $k = 1, \ldots, M$, τ the time lag, and D_e the embedding dimension. The columns of $\underline{\mathbf{V}}$ form an orthonormal basis of the map $\underline{\mathbf{X}} \to \underline{\mathbf{X}}' = \underline{\mathbf{X}}\underline{\mathbf{V}}$ where $\underline{\mathbf{X}}'$ is seen as a rotation of $\underline{\mathbf{X}}$ [2]. The transformation $\underline{\mathbf{X}}\underline{\mathbf{V}}$ diagonalizes the covariance matrix $\underline{\mu}$

$$\mu_{i,j} = 1/M \sum_k \xi_{i,k}\xi_{j,k} = (\underline{\mathbf{X}}^{\mathbf{T}}\underline{\mathbf{X}})_{i,j} \tag{3.32}$$

under the assumption that each column in $\underline{\mathbf{X}}$ sums to zero [2]. To see that $\underline{\mathbf{X}}\underline{\mathbf{V}}$ diagonalizes the covariance matrix, note that

$$\underline{\mathbf{X}}^{\mathbf{T}}\underline{\mathbf{X}} = \underline{\mathbf{V}}\underline{\Sigma}^{\mathbf{T}}\underline{\mathbf{U}}^{\mathbf{T}}\underline{\mathbf{U}}\underline{\Sigma}\underline{\mathbf{V}}^{\mathbf{T}} \tag{3.33}$$

$$= \underline{\mathbf{V}}\underline{\Sigma}^{\mathbf{T}}\underline{\mathcal{I}}_{\mathbf{M}}\underline{\Sigma}\underline{\mathbf{V}}^{\mathbf{T}} \tag{3.34}$$

$$= \underline{\mathbf{V}}\underline{\Sigma}^{\mathbf{2}}\underline{\mathbf{V}}^{\mathbf{T}} \tag{3.35}$$

so

$$\underline{\mathbf{V}}^{\mathbf{T}}(\underline{\mathbf{X}}^{\mathbf{T}}\underline{\mathbf{X}})\underline{\mathbf{V}} = \underline{\Sigma}^{\mathbf{2}} \tag{3.36}$$

where $\underline{\Sigma}^{\mathbf{2}}$ is a diagonal matrix whose diagonal elements σ_i^2 are the eigenvalues of $\underline{\mathbf{X}}^{\mathbf{T}}\underline{\mathbf{X}}$ while the columns of V are its eigenvectors.

By diagonalizing the covariance matrix, the off-diagonal elements of $\underline{\mu}$ are set to zero and decorrelates the columns of the trajectory matrix, i.e. the columns of $\underline{\mathbf{X}}'$ are statistically independent [2]. Implicit in the definition of the correlation dimension (Eq. 2.30) is that the coordinates be statistically independent and thus the map $\underline{\mathbf{X}} \to \underline{\mathbf{X}}' = \underline{\mathbf{X}}\underline{\mathbf{V}}$ is beneficial for use with correlation dimension estimates.

Assuming $\underline{X}$ is produced by a mix of deterministic dynamics and stochastic noise, singular value decomposition can be helpful in noise reduction and minimizing the number of coordinates needed to obtain convergent D_2 estimates. Because $\underline{X}'$ contains a set of orthogonal coordinates, the map $\underline{XV}$ rotates the data in the phase space onto its *principal axes*. The eigenvalue σ_i^2 is the variance of the ith principal coordinate. An eigenvalue $\sigma_i^2 = 0$ signifies that the ith principal axis is not visited by the rotated trajectory. In the presence of noise, the eigenvalues do not vanish. However, if the variance of the noise is sufficiently small, some eigenvalues will also be small and the corresponding eigenvectors can be ignored in the transformation. The elements along the diagonal of the eigenvalue matrix $\underline{\Sigma}^2$ are ranked such that $\sigma_1^2 > \sigma_2^2 > \ldots > \sigma_{r_{eff}}^2 > \ldots > \sigma_r^2$ where r_{eff} and r are the effective and true rank, respectively. Ignoring unenergetic eigenvalues is tantamount to using only the first r_{eff} columns of $\underline{X}'$ for the embedding. A similar reduction can be made for all weak coordinates in the transformation, thereby reducing the number of coordinates necessary to span the embedding space.

In this dissertation, an eigenvalue rejection tolerance was established (based on recommendations found in [2]) and any normalized eigenvalue

$$\widehat{\sigma}_i^2 = \frac{\sigma_i^2}{\sum_{j=1}^{D_c} \sigma_j^2} \tag{3.37}$$

that did not exceed this threshold was ignored in the embedding. The reduced coordinate set was then analyzed by the correlation dimension to quantify any existing fractal structure.

3.8 Results

The results for the MODRA are presented in this section including a comparison to an existing wavelet denoising technique (waveshrink), examples of denoised chaotic sequences and their corresponding correlation dimension estimates, and application

to a physical experiment design to provoke a chaotic response. Waveshrink was chosen because of its popularity and effectiveness as a denoising algorithm. All the results in this section are carried out using the following parameters unless specified otherwise:

▶ *MODRA and D_2 Defaults*

Sequence length	$N = 2048$
MODWPT decomposition level	$J = 6$
Daubechies' Filter Set	Least Asymmetric 8 Tap (LA8)
Nodal energy admissibility threshold	$\delta_E = 0.2$
Kaplan violation threshold	$\delta_{kap} = 0.02$
Singular value decomposition rotation	On
Time lag	$\tau = 2.5\,T_e$
Trajectory lag	$\gamma = \lfloor N/200 \rfloor$
Embedding dimension	$D_e = 1, \ldots, 7$

where T_e is the autocorrelation time.

3.8.1 MODRA versus Wavelet Shrinkage

A popular wavelet based denoising technique is the *wavelet shrinkage* or *waveshrink (WS)* method invented by Donoho and Johnstone [18]. The algorithm starts with a partial or full DWT decomposition. Next, a threshold δ_{DJ} is assigned and all wavelet coefficients whose absolute value does not exceed this threshold are zeroed. The remaining wavelet coefficients are used to synthesize the original signal. The synthesis is "denoised" by eliminating nonenergetic time-scale content. The WS method has been shown to work very well for a variety of different sequences (see [18, 73] for example).

Considering the popularity and effectiveness of the WS technique, it is of interest
to compare the denoising capability of MODRA. Shown in Fig. 3.25 is a comparison
of the denoising techniques for a chaotic Lorenz sequence contaminated with Gaus-
sian pseudo-random noise such that the signal-to-noise ratio is SNR $\equiv \hat{\sigma}_X^2/\hat{\sigma}_\epsilon^2 = 1.0$
where $\hat{\sigma}_X^2$ and $\hat{\sigma}_\epsilon^2$ are the sample variance of the chaotic sequence and noise, respec-
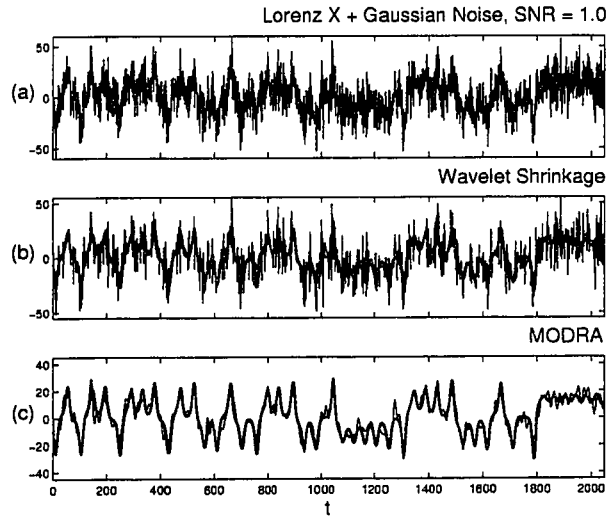tively. In this test case, it is visually apparent that MODRA does a better job in

Figure 3.25: Comparison of MODRA to WS. The subplots are (a) a chaotic Lorenz sequence
contaminated with Gaussian pseudorandom noise such that the SNR = 1.0 (b) the wavelet shrinkage
using only the top 20% of the DWT coefficients for synthesis (thin line) overlayed with the original
uncontaminated sequence (thick line), and (c) MODRA synthesis (thin line) overlayed with the
original uncontaminated sequence (thick line).

recovering the original sequence as the denoised sequence more closely conforms to
the pattern of the original uncontaminated sequence (Fig. 3.25c). To better quantify
the comparison, a series of tests were performed on the same chaotic sequence using
a wide variety of signal-to-noise (SNR) and δ_{DJ} combinations. The quality of the

synthesis was quantified with a sample *mean squared error* (MSE) value defined as

$$MSE \equiv \frac{1}{N} \sum_{t=0}^{N-1} (X_t - \widehat{X}_t)^2 \qquad (3.38)$$

where $\{\widehat{X}_t\}$ is the synthesized or denoised sequence and $\{X_t\}$ is the original uncontaminated chaotic sequence. $\{X_t\}$ was preconditioned to have zero mean and unit energy. In almost all cases, the results were favorable for MODRA (Fig. 3.26). The
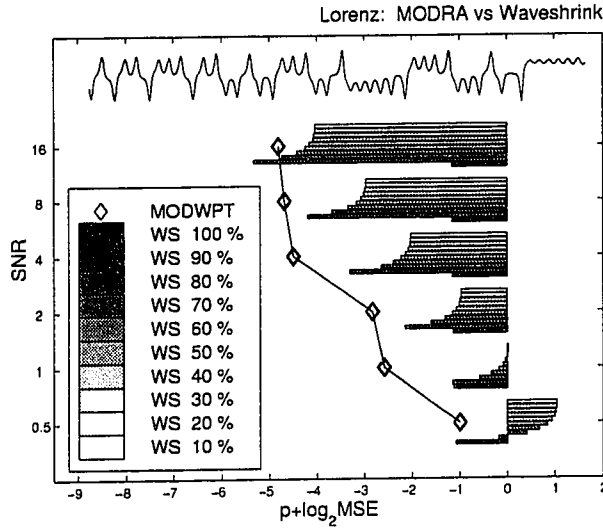

Figure 3.26: A comparison of MODRA to WS for contaminated chaotic Lorenz sequences for $\delta_{kap} = 0.2, N = 2^p = 1024, J = 6$, using Daubechies' LA8 filters. See text for details.

horizontal bars in Fig. 3.26 represent the MSE results for the wavelet shrinkage technique. The values in the legend represent the percentage of wavelet coefficients that were rejected (zeroed) in the synthesis process. There are 10 rejection percentages used per every SNR value ranging from SNR = 0.5 (very noisy) to SNR = 16 (very clean). The diamonds represent the MSE for MODRA. In almost every case, MODRA maintained a smaller MSE than the wavelet shrinkage algorithm. However,

the higher quality noise rejection comes at a computational cost mainly attributed to using KDM on every node in the best basis.

A similar test was performed on a chaotic response from the Rössler system (Fig. 3.27). Again MODRA outperformed the WS technique in terms of minimum MSE. However, MODRA did not fair as well for a chaotic tent map sequence test shown in Fig. 3.28 The tent map is the most difficult of the test sequences to denoise because
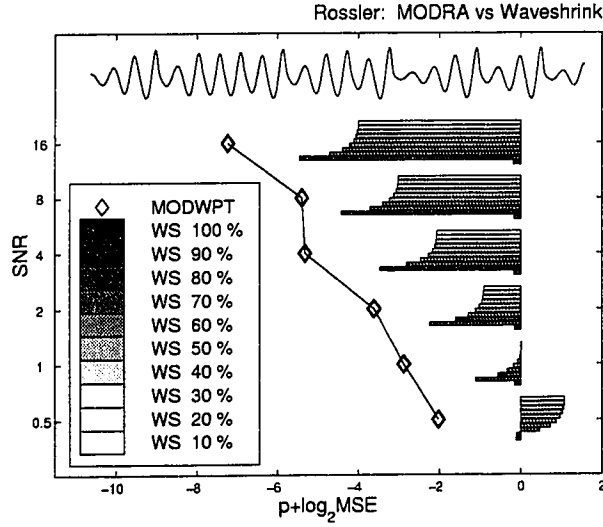

Figure 3.27: A comparison of MODRA to WS for contaminated chaotic Rössler sequences for $\delta_{kap} = 0.2, N = 2^p = 1024, J = 6$, using Daubechies' LA8 filters. See text for details.

it has a comparatively broader and flatter spectrum, much like the spectrum of the noise (Fig. 3.29). Figure 3.30 shows the synthesis and original time histories for SNR = 1.0 with δ_{DJ} set such that 80% of the DWT coefficients were rejected.

A comparison of the phase plane embeddings corresponding to Fig. 3.25 and Fig. 3.30 is shown in Fig. 3.31. With the exception of the MODRA for the Lorenz system, the denoised embeddings are fuzzy. The conditions, however, are extremely noisy in this case with the SNR = 1.0, i.e. the noise is equally as energetic as the
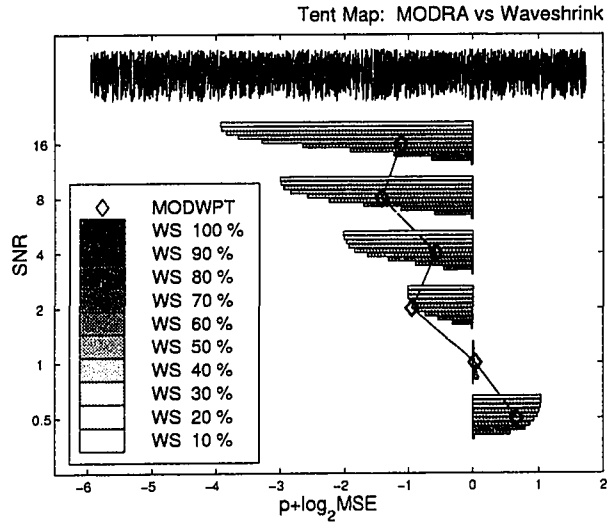
Figure 3.28: A comparison of MODRA to WS for contaminated chaotic tent map sequences for $\delta_{kap} = 0.2, N = 2^p = 1024, J = 6$, using Daubechies' LA8 filters. The MODRA has a better or equivalent MSE value for SNR $= 0.5, 1, 2$ (noisy conditions), but does relatively poorly for SNR > 2.0.
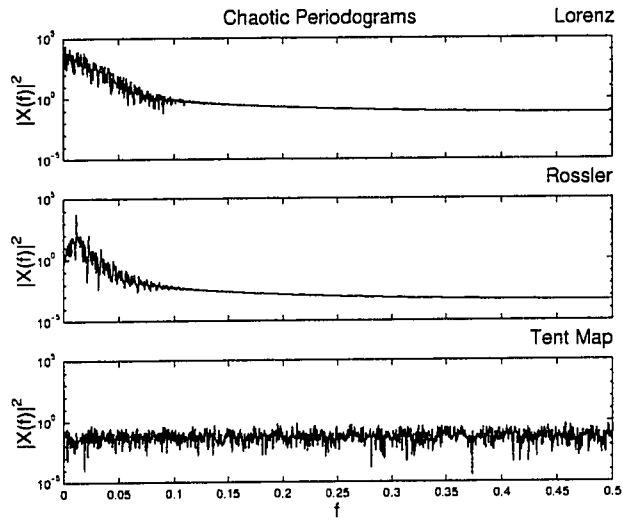

Figure 3.29: Periodograms for a chaotic Lorenz, Rössler, and tent map systems.
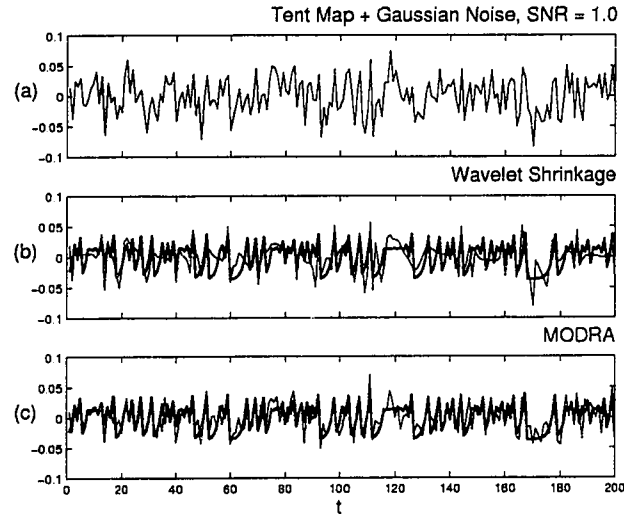
Figure 3.30: Comparison of MODRA to WS. The subplots are (a) a chaotic tent map sequence contaminated with Gaussian pseudorandom noise such that the SNR = 1.0 (b) the wavelet shrinkage using only the top 20% of the DWT wavelet coefficients for synthesis (thin line) overlayed with the original uncontaminated sequence (thick line), and (c) MODRA synthesis (thin line) overlayed with the original uncontaminated sequence (thick line). The Kaplan rejection tolerance was set at $\delta_{kap} = 0.02$, $N = 2^P = 1024$, $J = 6$, using Daubechies' LA8 filters.

chaotic sequence. The very semblance of a recognizable attractor is promising and lends credit to the MODRA.
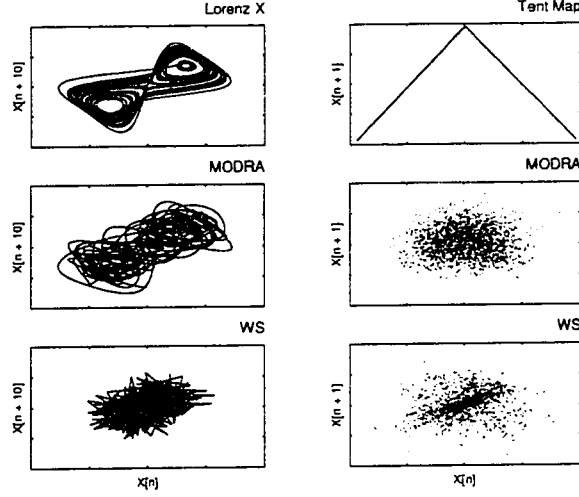

Figure 3.31: Phase plane embeddings of original and denoised chaotic sequences. The original uncontaminated sequence is plotted in the first row. The left(right) column contains a chaotic Lorenz(tent map) sequence followed by the MODRA and waveshrink results for a contaminated sequence with SNR = 1.0.

In an attempt to optimize the waveshrink process. Donoho and Johnstone developed a number of thresholding algorithms for use with the DWT coefficients. The Donoho-Johnstone threshold δ_{DJ} is defined as

$$\delta_{DJ}^2 \equiv 2\sigma_\epsilon^2 \log(N),$$

(3.39)

where σ_ϵ^2 is the variance of a zero mean Gaussian white noise process and N is the sequence length. Three suggested denoising techniques for use with the DWT wavelet nodes are shown below. $W_{j,t}$ is the wavelet coefficient at time t and decomposition level j.

Hard Thresholding

$$W_{j,t}^{(ht)} = \begin{cases} W_{j,t}, & |W_{j,t}| > \delta_{DJ} \\ 0, & \text{otherwise} \end{cases} \tag{3.40}$$

Soft Thresholding

$$W_{j,t}^{(st)} = \text{sign}\{W_{j,t}\} \left(|W_{j,t}| - \delta_{DJ}\right)_+, \tag{3.41}$$

where

$$\text{sign}\{W_{j,t}\} \equiv \begin{cases} +1, & W_{j,t} > 0 \\ 0, & W_{j,t} = 0 \\ -1, & W_{j,t} < 0 \end{cases}$$

and

$$(x)_+ = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Hard/Soft Thresholding

$$W_{j,t}^{(hst)} = \text{sign}\{W_{j,t}\} \left(|W_{j,t}| - \delta_{DJ}\right)_{++}, \tag{3.42}$$

where

$$\left(|W_{j,t}| - \delta_{DJ}\right)_{++} = \begin{cases} 2(|W_{j,t}| - \delta_{DJ})_+, & |W_{j,t}| < 2\delta_{DJ} \\ |W_{j,t}|, & \text{otherwise} \end{cases}$$

There is no manipulation of the scaling coefficients. i.e. the DWT node associated with the normalized frequency range $[0, 2^{-(J+1)})$ is left alone. For the case where σ_ϵ^2 is unknown, Donoho and Johnstone recommend the so called *MAD* standard deviation estimation by

$$\hat{\sigma}\text{MAD} \equiv \frac{\text{median}\{|W_{1,0}|, |W_{1,1}|, \ldots, |W_{1,\frac{N}{2}-1}|\}}{0.6745}. \tag{3.43}$$

The waveshrink results used thus far were a result of hard thresholding without the use of Eq. 3.39 as an estimate for the optimal threshold. It is beyond the scope of

this dissertation to assess the attributes and drawbacks of the thresholding methods. However, an excellent review is given in [73, 18] and an illustrative example shown in Fig. 3.32 where a sine wave of unit amplitude is reduced with $\delta_{DJ} = 0.5$. The soft
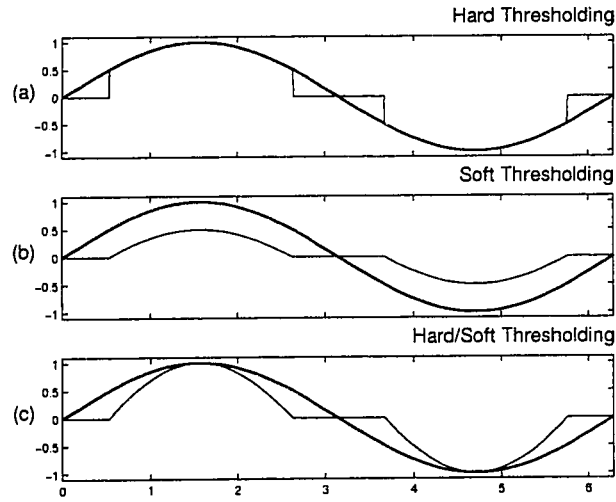

Figure 3.32: Illustration of the Donoho and Johnstone thresholding techniques on a unit amplitude sine wave using $\delta_{DJ} = 0.5$ for the (a) hard (b) soft and (c) hard soft methods.

thresholding smooths out the discontinuitous portions seen in the hard thresholding but poorly represents highly energetic coefficients. The hard/soft thresholding is smooth and more accurately portrays energetic coefficients. The Lorenz, Rössler, and tent map systems were contaminated and denoised with the WS method using hard, soft, and hard/soft thresholding in Fig. 3.33, 3.34, and 3.35. The average number of wavelet coefficients rejected during the thresholding is listed to the right of each plot at the corresponding SNR level. In every case, the rejection level is

very high with more than 95% of the coefficients thrown out during thresholding. The results indicate that in most cases, MODRA is equivalent or better than the denoising capability of waveshrink. The exception is again at low SNR values for the tent map where WS clearly dominates. However, the MODRA performs excellently on the tent map for SNR > 1.0. WS does not perform well in this regime because of the flatness of the tent map spectrum, i.e. if the tolerance is set too high, most or all of the coefficients perish in the thresholding. This is not the case for the Lorenz or Rössler spectra where an overvalued tolerance does not wipe out all of the coefficients for synthesis. For the tent map, automatic threshold selection is consequently not recommended for high SNR values.
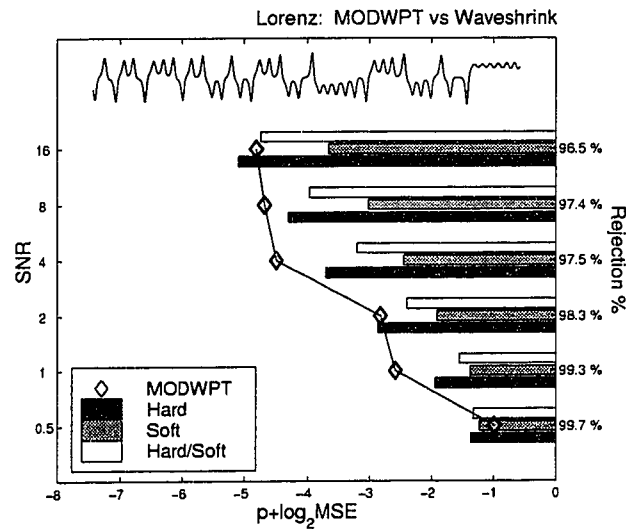

Figure 3.33: A comparison of MODRA to WS using hard, soft, and hard/soft thresholding for contaminated chaotic Lorenz sequences.
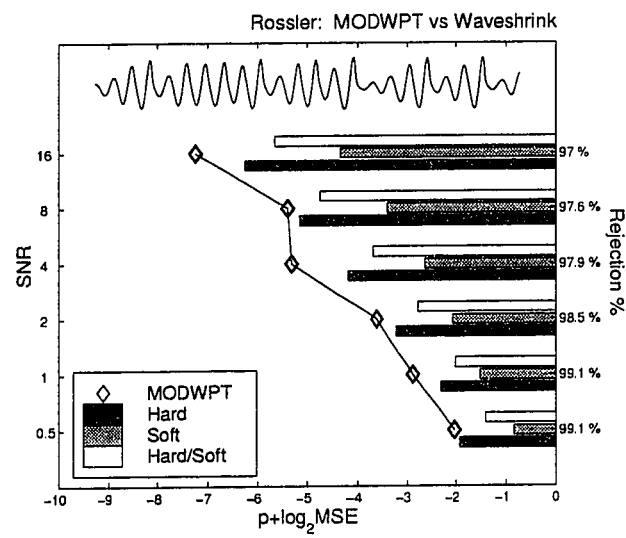
Figure 3.34: A comparison of MODRA to WS using hard. soft, and hard/soft thresholding for contaminated chaotic Rössler sequences.
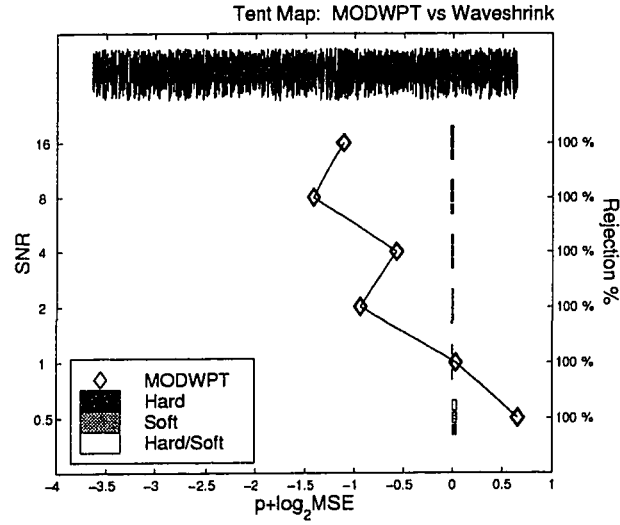
128

Figure 3.35: A comparison of MODRA to WS using hard, soft, and hard/soft thresholding for contaminated chaotic tent map sequences.

3.8.2 D_2 Results for MODRA-Denoised Chaotic Sequences

While the MODRA is not limited to use on chaotic sequences, they serve as an interesting test cases because chaotic signals exhibit complex, seemingly random response. In this section, numerical experiments are performed on the Bender-Orzag, Lorenz, Rössler, and tent map systems whose parameter space is set such that a chaotic response occurs. D_2 is used to quantify the MODRA noise reduction capability in terms of a nonlinear measure.

The mean of each sequence is removed and the result normalized to have unit energy prior to the D_2 computation, making it easier to compare results. Furthermore,

singular value decomposition is used to rotate the embedding (trajectory matrix) which maps the data onto its principal axes in the phase space [2]. Fig. 3.36 shows the correlation dimension results for the uncontaminated test sequences. The Lorenz
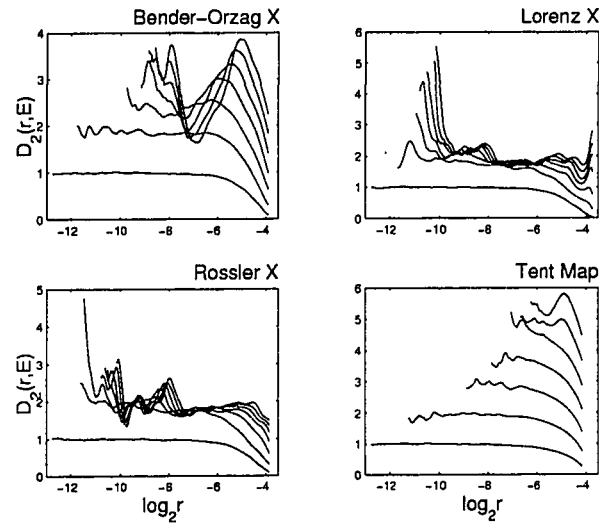

Figure 3.36: D_2 results for uncontaminated chaotic sequences.

and Rössler system have a characteristic "flat" scaling region in which the correlation dimension is estimated. The Bender-Orzag response widely fluctuates and has no apparent flat scaling region. However, the D_2 curves for the Bender-Orzag system do not rise with embedding dimension, indicating nonstochastic behavior. The interesting case is the D_2 results for the tent map as the curves do *not* converge and *do* rise with embedding dimension, even though the map is deterministic! The reason for this is because blind selection of the time lag proportional the autocorrelation

time is an incorrect approach for chaotic maps. The best lag for chaotic maps is $\tau = 1$. If $\tau \gg 1$, replications of the map muddle the phase space, giving it a random appearance (Fig. 3.37).
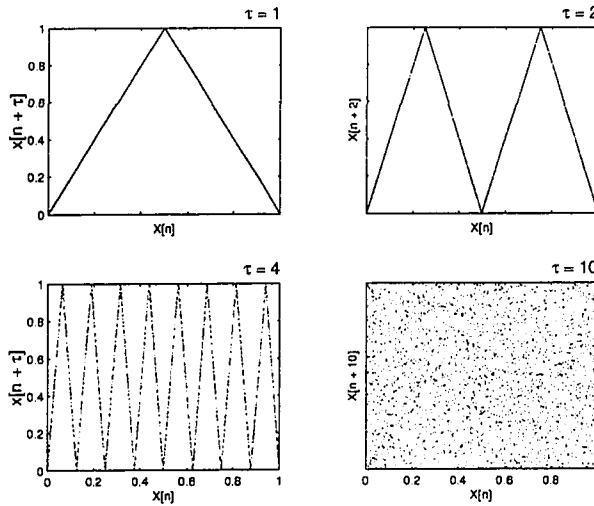

Figure 3.37: A two dimensional embedding of the tent map for time lags $\tau = 1, 2, 4, 10$. There are $2^{\tau-1}$ replications of the tent map structure for a given time lag. Selecting too high of a lag muddles the phase space, much like an embedding of a pseudo-random sequence.

The MODRA-D_2 results for the Bender-Orzag system are shown in Fig. 3.38-3.40, the Lorenz system in Fig. 3.41-3.43, the Rössler system results in Fig. 3.44-3.46, and the tent map in Fig. 3.47-3.49. For each case, an ensemble of plots is displayed, showcasing various stages of the MODRA and D_2 algorithms. The values to the right of each plot in the MODWPT details indicate the acceptance (D) or rejection (R) of the nodes subscripted by the Kaplan violation measure μ_λ. The titles above each plot in the MODRA synthesis relate the energy and entropy values

of the corresponding sequences. The lower synthesis plot also shows the reduction in these values expressed as a percentage of the measure calculated for the original sequence. Finally, the M and N_r values appearing in the top row of the D_2 plot are the number of points used to average the correlation integral values in a unit lag moving average and the total number of correlation points used to estimate D_2, respectively. A linear least squares fit is used to calculate the correlation integral slopes over each M point window, yielding an estimate of D_2 at the corresponding scale and embedding dimension. Of course, the SNR $= 0.5$ case proves to be the most difficult one and no complete D_2 convergence is achieved. However, the MODRA successfully rids the test sequences of a majority of the noise for SNR $= 1.0$. and 2.0 cases. This is most evident in the lower left plot in the bottom row of each figure panel, i.e. in the synthesis plots. The synthesis (thick line) is overlayed with the original *uncontaminated* chaotic sequence (thin line). The MODRA synthesis is impressive considering the high noise level. Finally, all of the MODRA-D_2 results do not converge for the tent map, regardless of the level of noise reduction. We wish to emphasize that this is more a function of improper lagging than it is of the MODRA as can be seen by comparing these estimates to the original uncontaminated D_2 shown in Fig. 3.36.

132
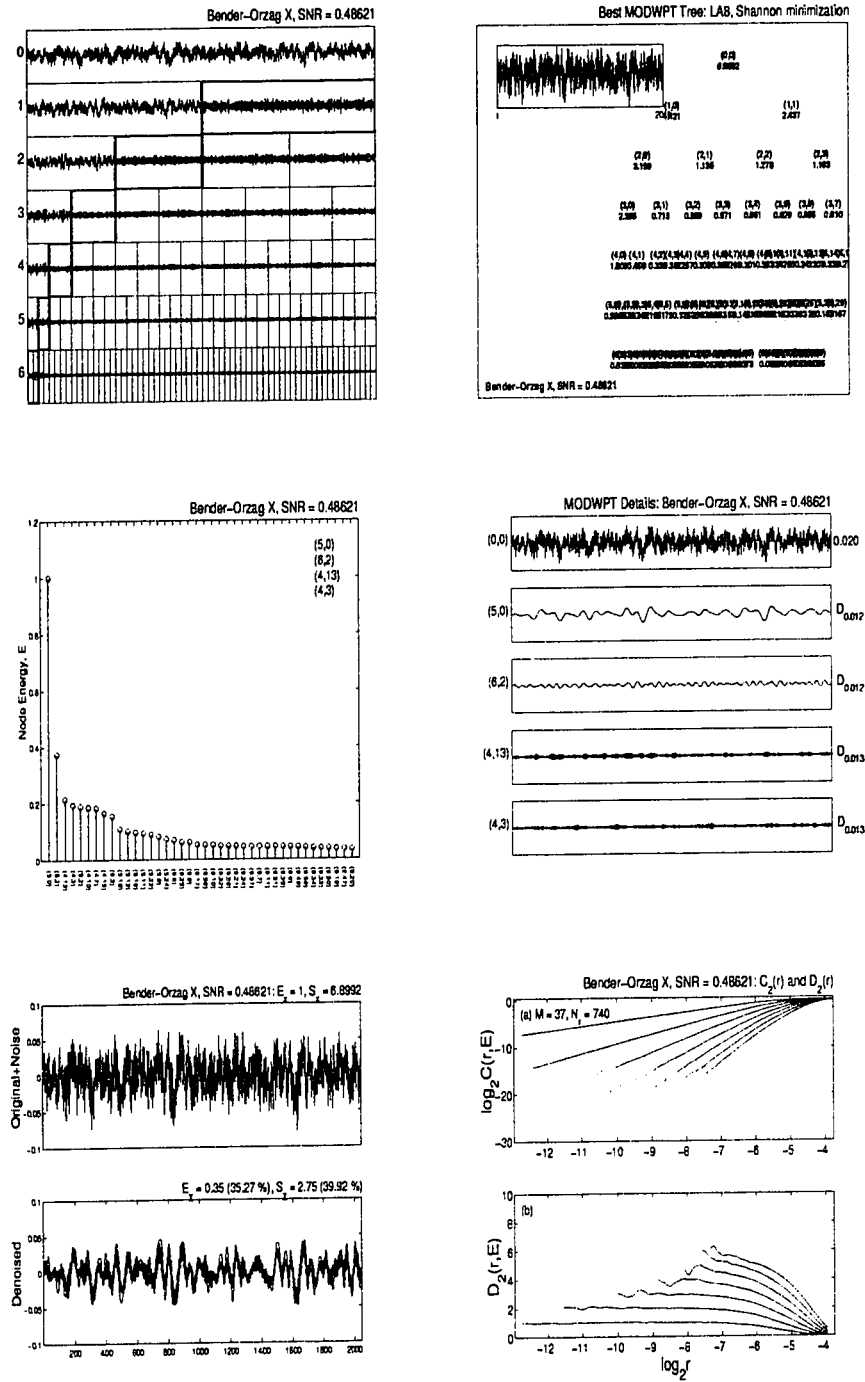
Figure 3.38: MODRA and D_2 results for the Bender-Orzag system with SNR = 0.5.
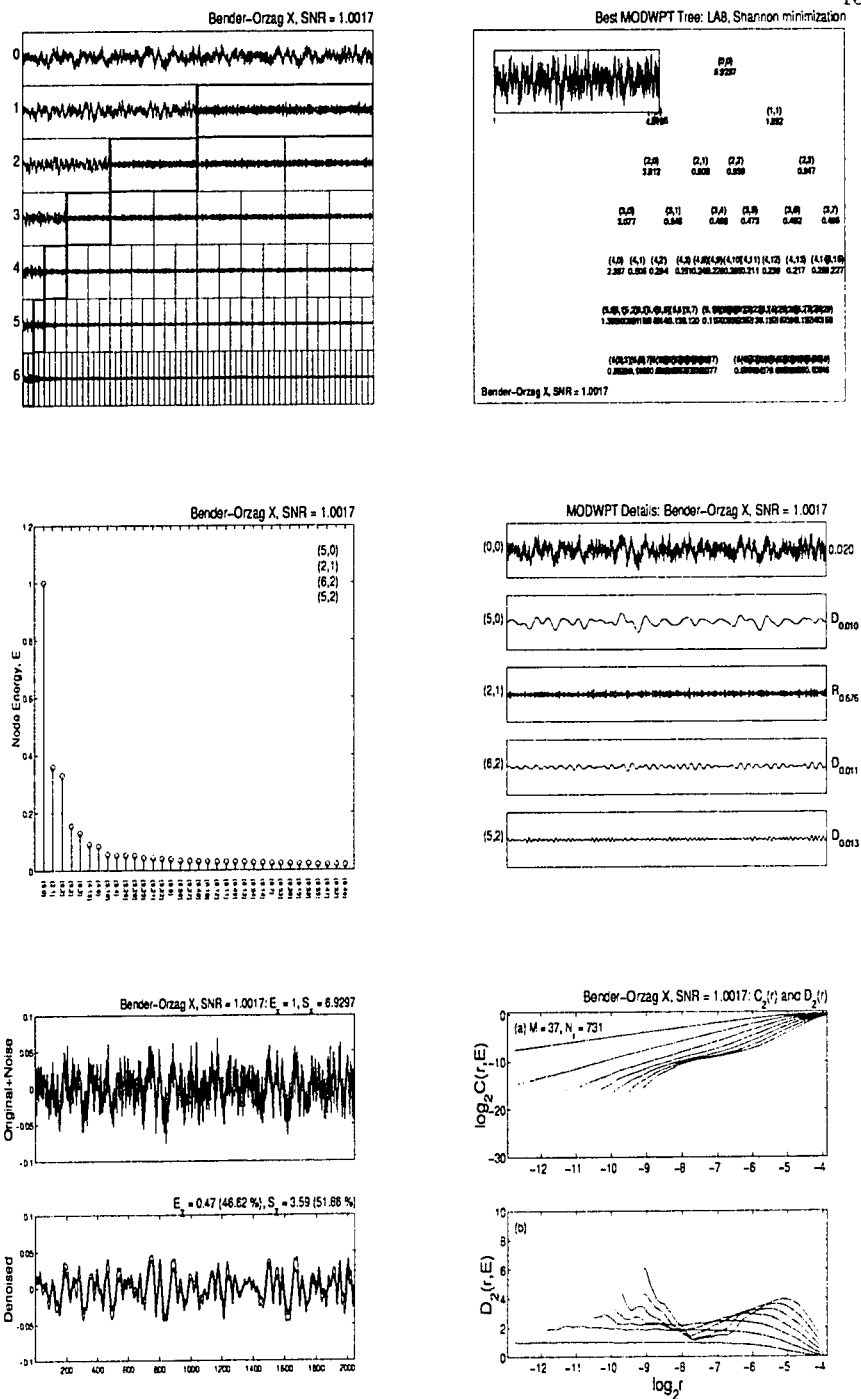
Figure 3.39: MODRA and D_2 results for the Bender-Orzag system with SNR $= 1.0$.
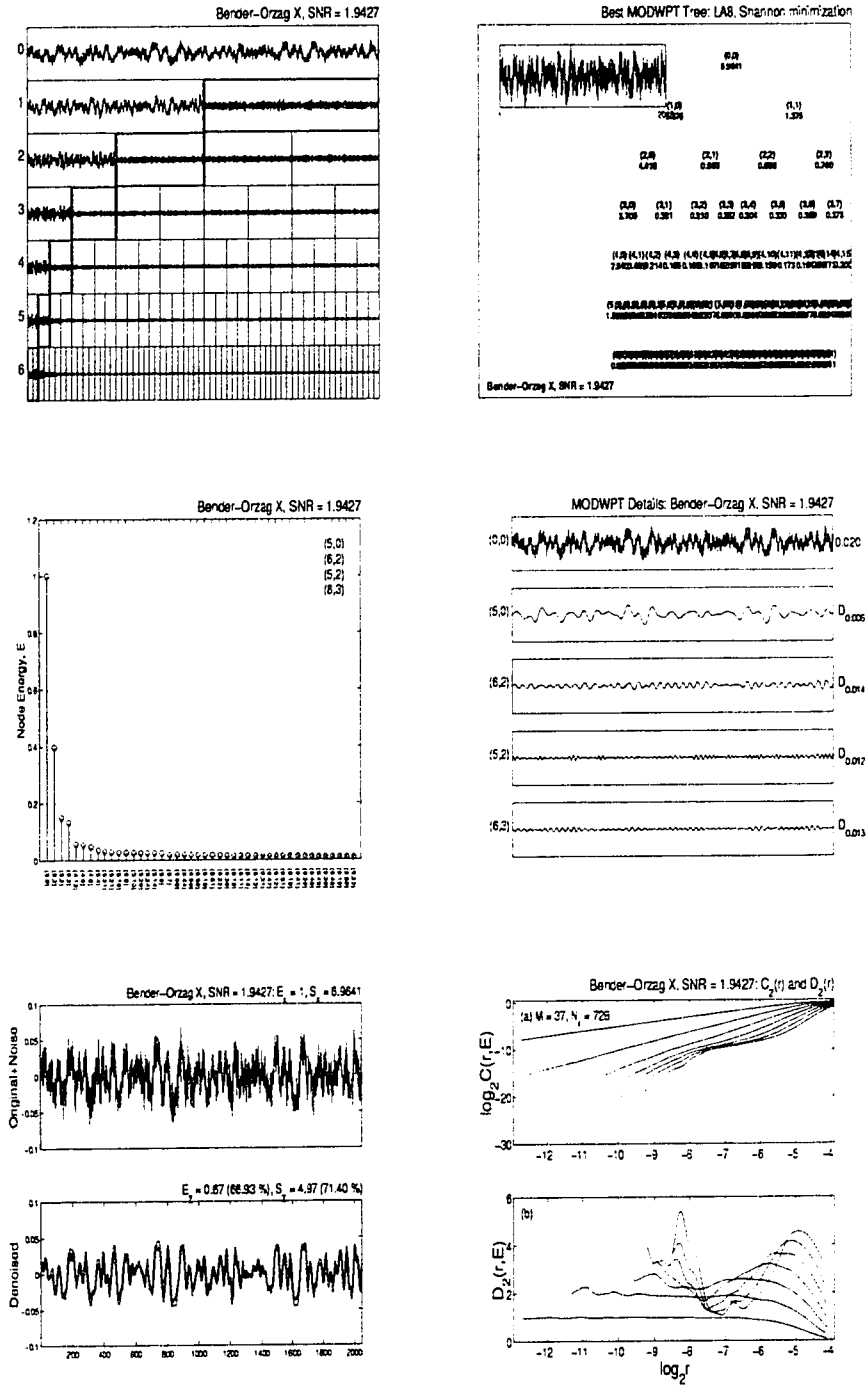
Figure 3.40: MODRA and D_2 results for the Bender-Orzag system with SNR = 2.0.

Figure 3.41: MODRA and D_2 results for the Lorenz system with SNR $= 0.5$.
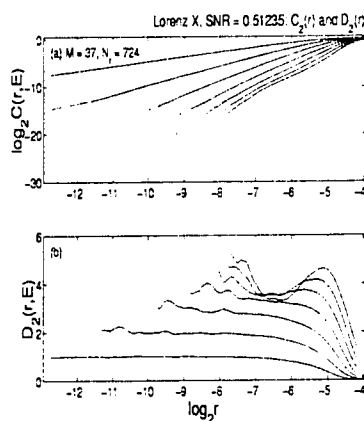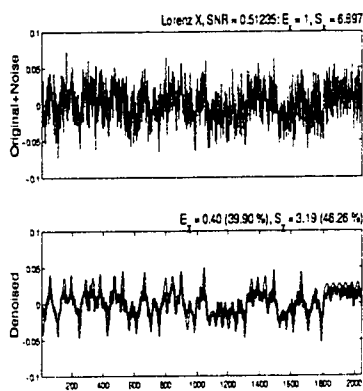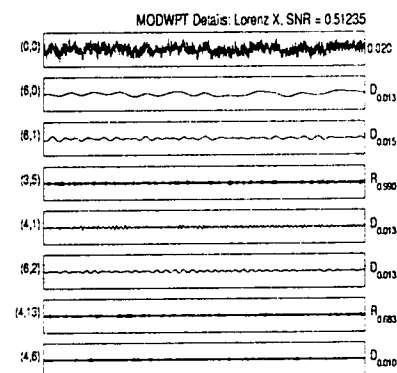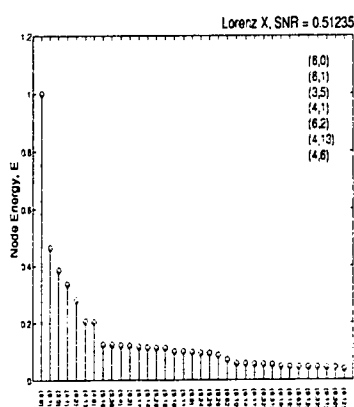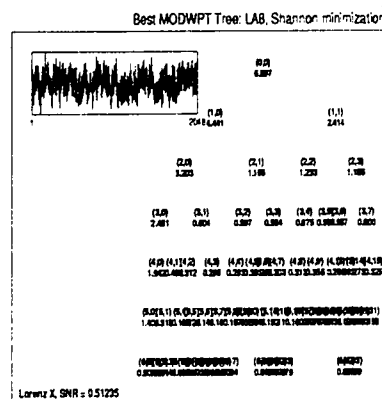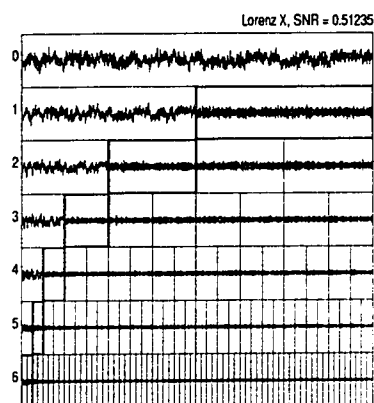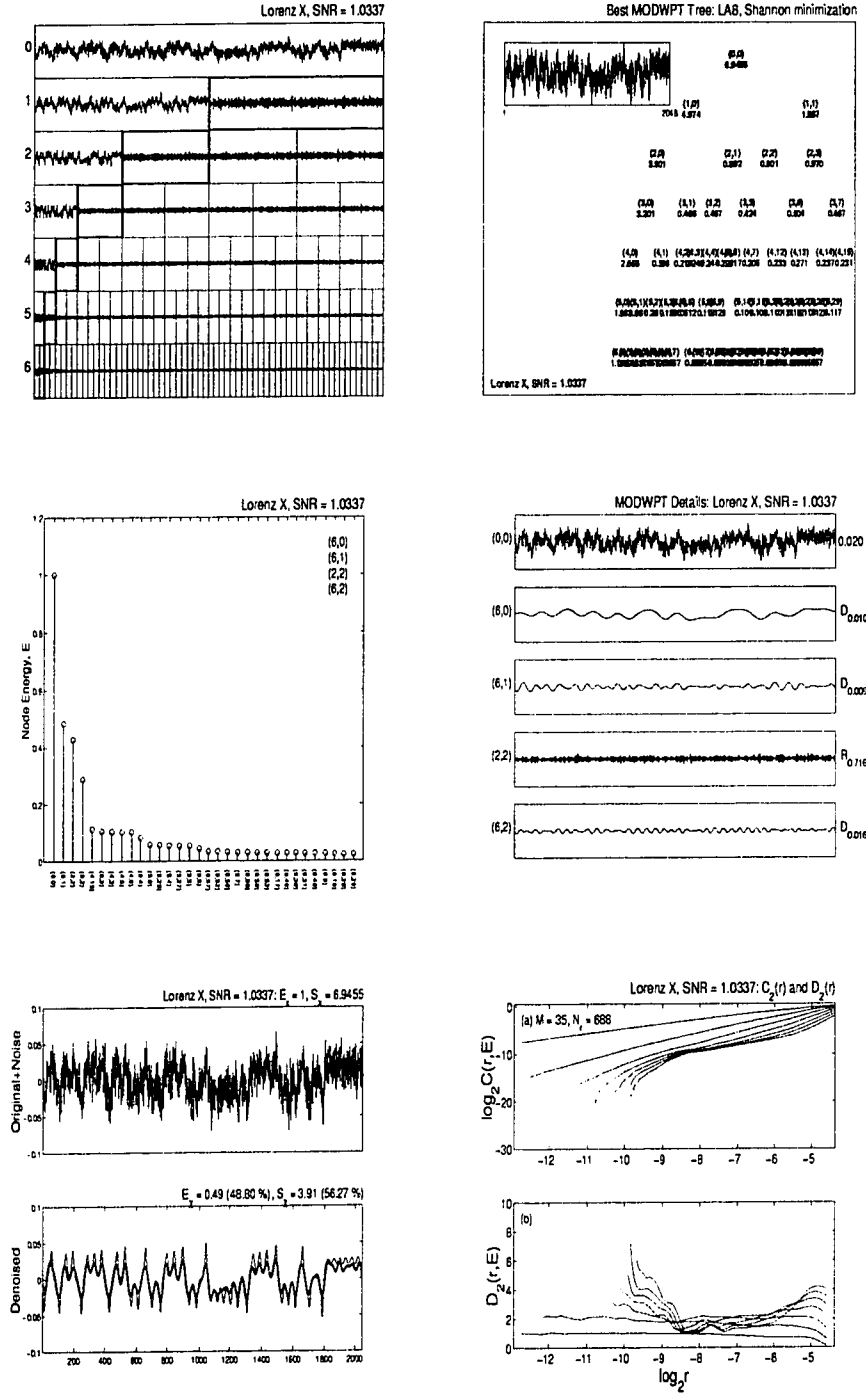
Figure 3.42: MODRA and D_2 results for the Lorenz system with SNR = 1.0. See text for details.
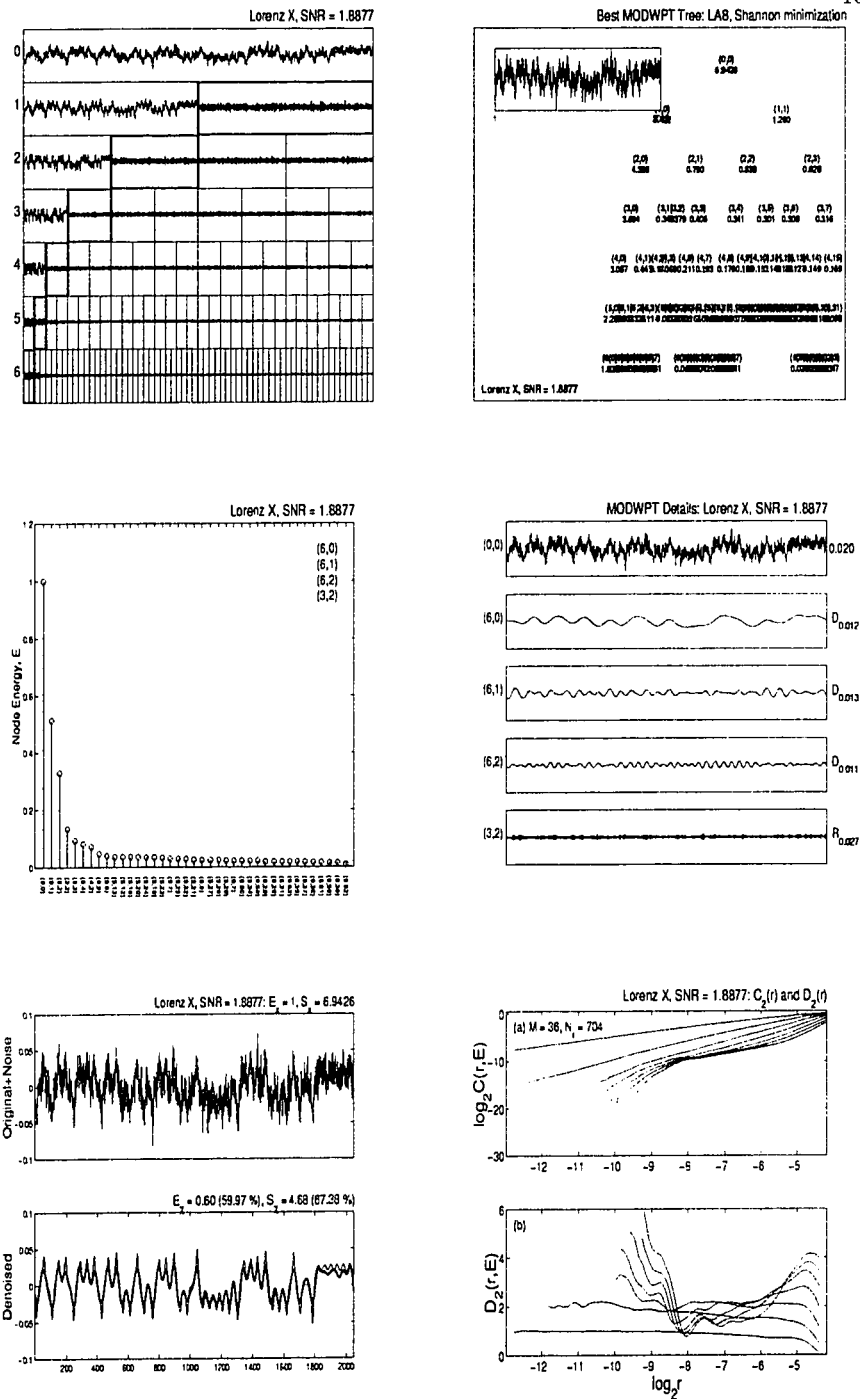
Figure 3.43: MODRA and D_2 results for the Lorenz system with SNR = 2.0.

Rossler X, SNR = 0.46163

Best MODWPT Tree: LAB, Shannon minimization

Rossler X, SNR = 0.46163

Node Energy, E

(6,1)
(6,2)
(4,5)
(4,7)

MODWPT Details: Rossler X, SNR = 0.46163

Rossler X, SNR = 0.46163: E = 1, S = 6.9206

Original+Noise

E = 0.33 (33.01 %), S = 2.70 (38.96 %)

Denoised

Rossler X, SNR = 0.46163: C₂(r) and D₂(r)

(a) M = 36, N = 709

log₂ C(r,E)

(b)

D₂(r,E)

log₂ r

Figure 3.44: MODRA and D_2 results for the Rössler system with SNR = 0.5. See text for details.

Figure 3.45: MODRA and D_2 results for the Rössler system with SNR = 1.0.

Figure 3.46: MODRA and D_2 results for the Rössler system with SNR = 2.0. See text for details.

Figure 3.47: MODRA and D_2 results for the tent map system with SNR $= 0.5$.

142
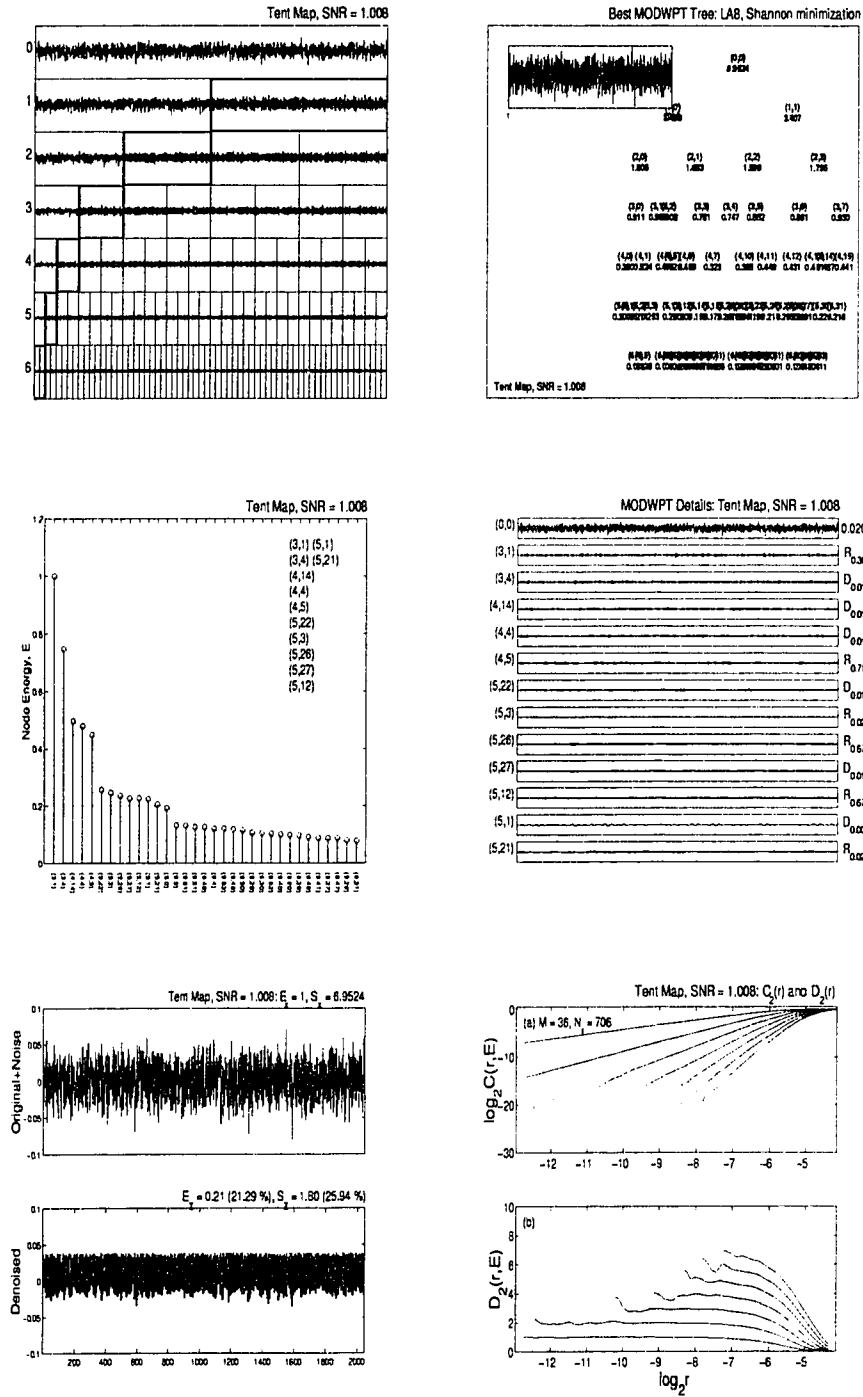
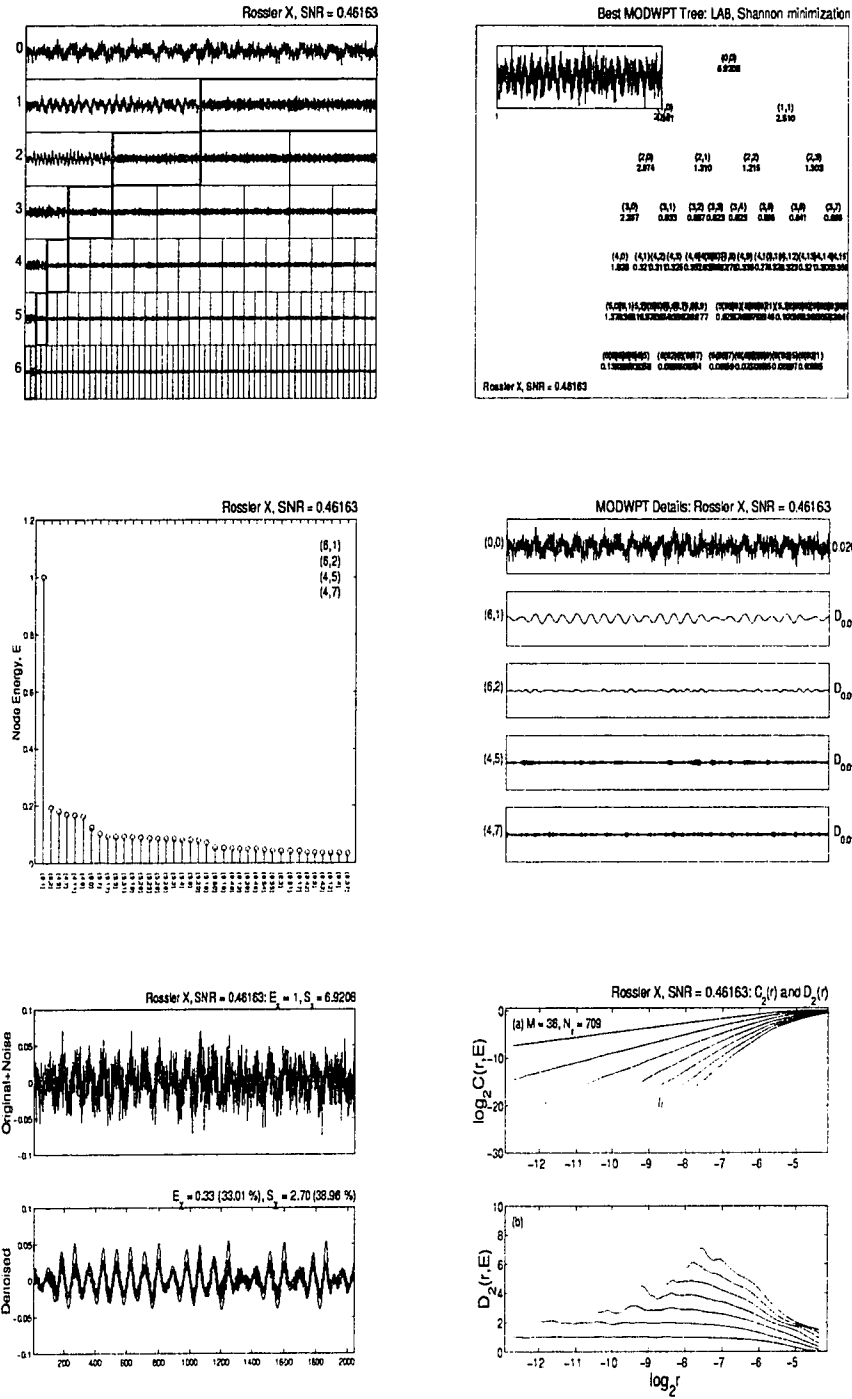Figure 3.48: MODRA and D_2 results for the tent map system with SNR = 1.0. See text for details.
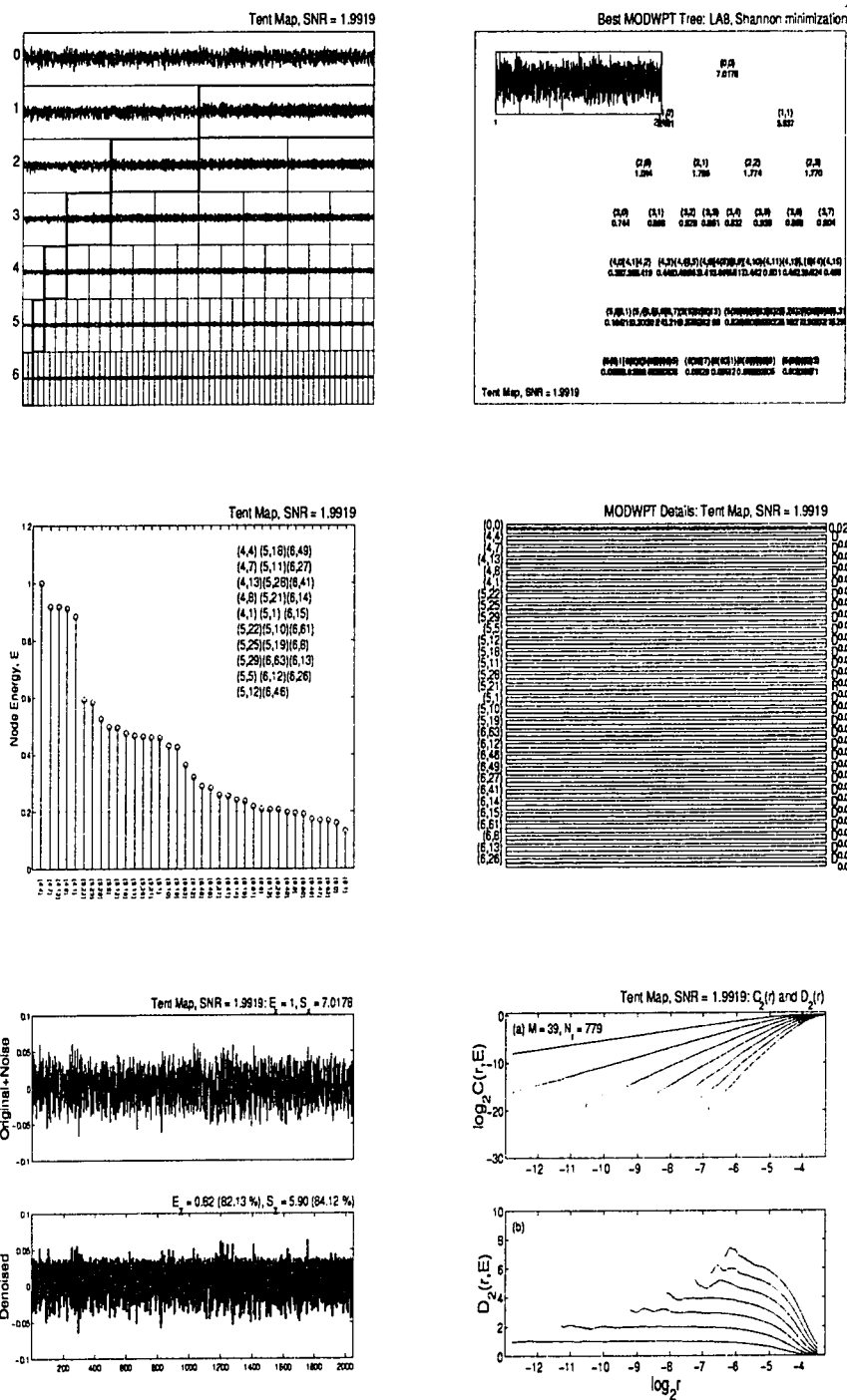
Figure 3.49: MODRA and D_2 results for the tent map system with SNR = 2.0.
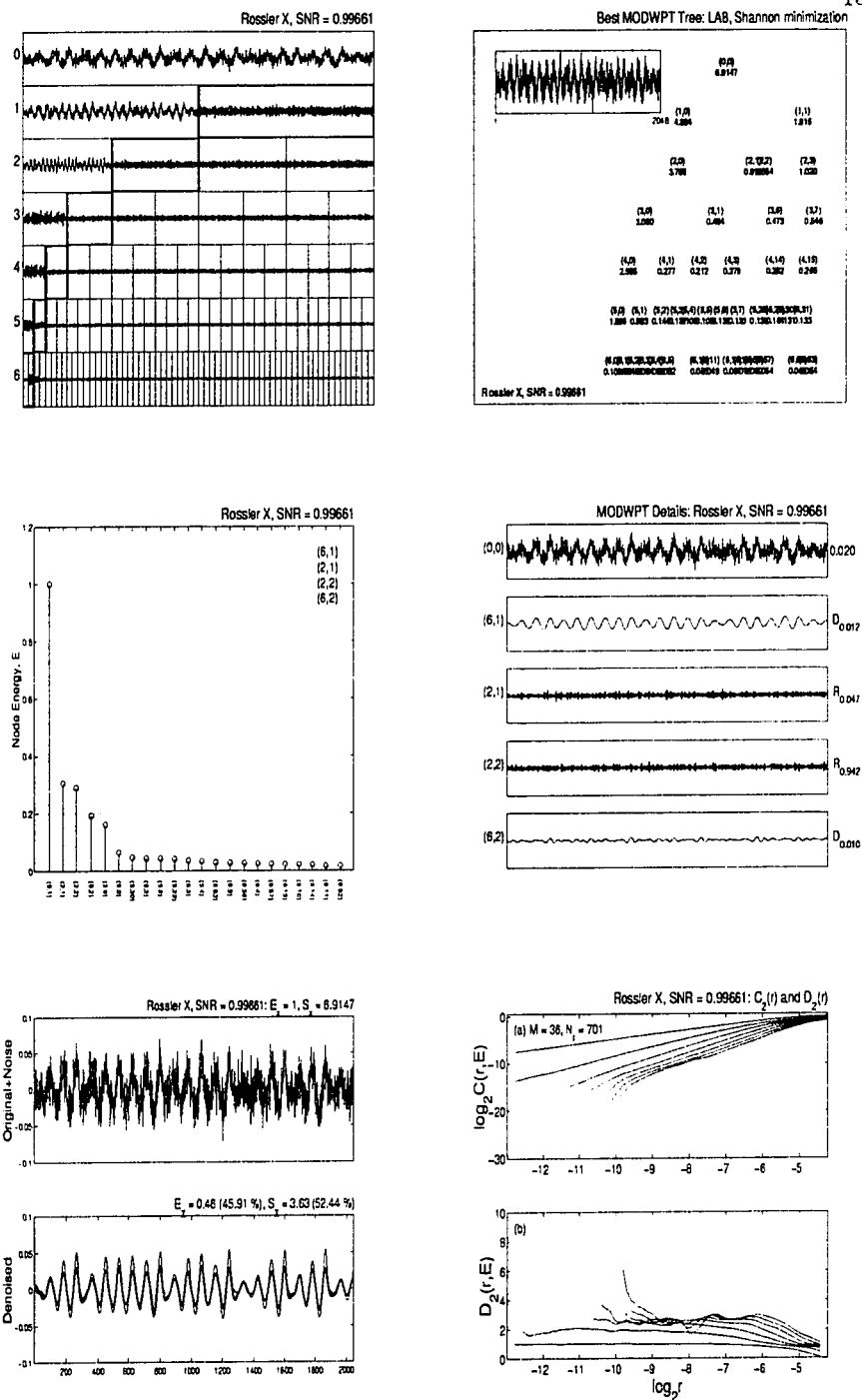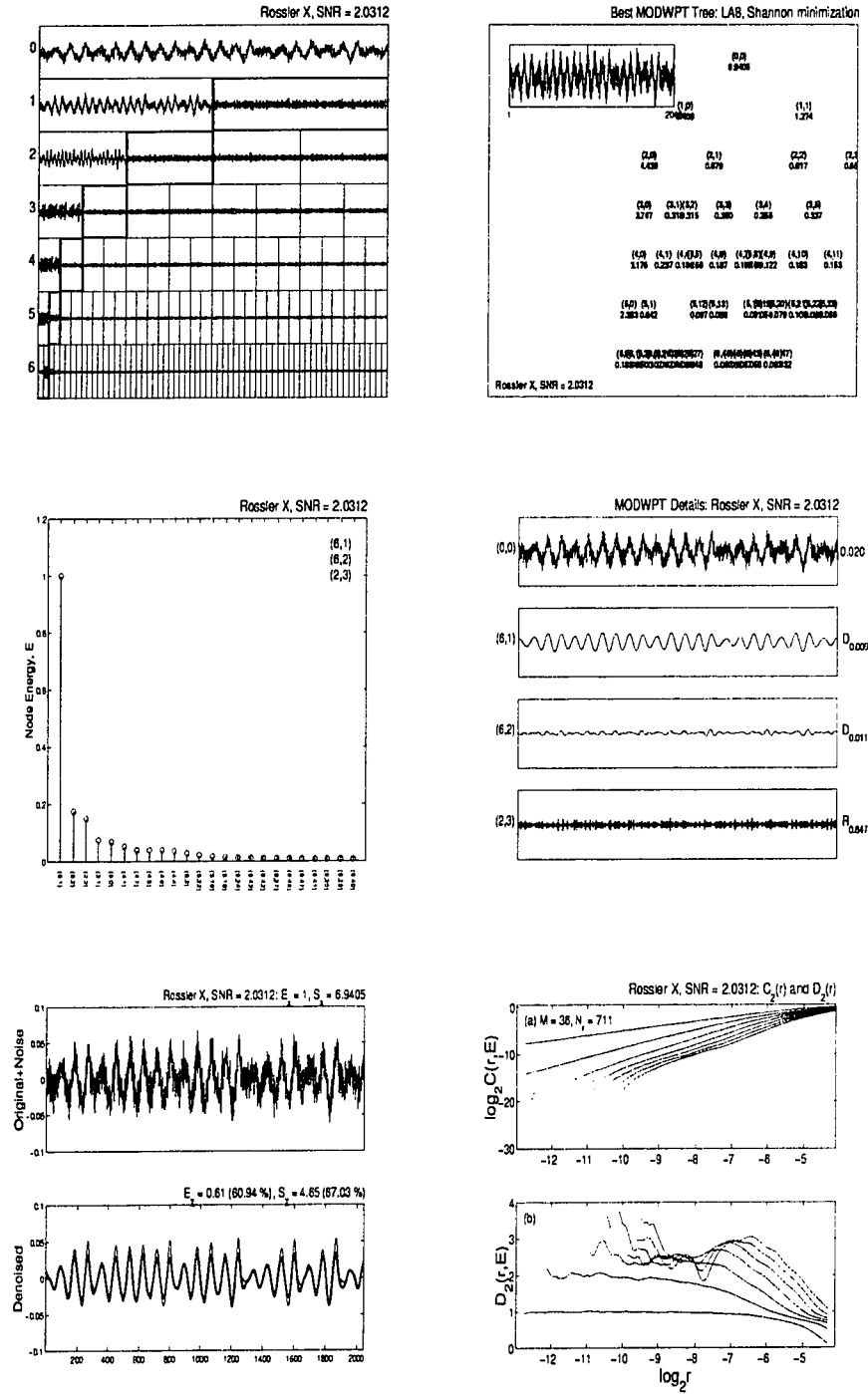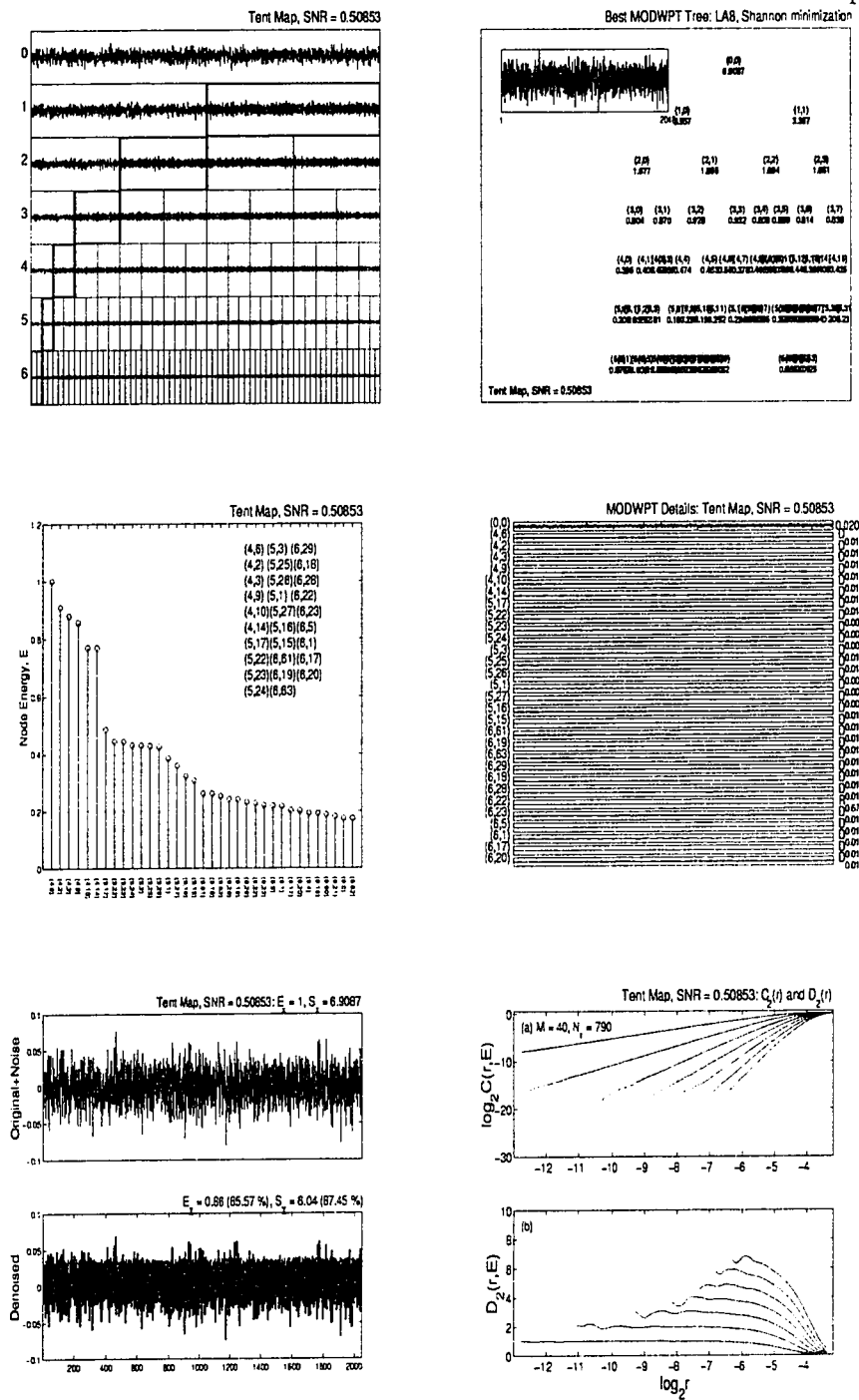
144

3.8.3   Chaotic Beam Experiment

In order to test the MODWPT denoising technique on a real system, a chaotic beam experiment was performed. The experimental setup is shown in Fig. 3.50. A flexible thin steel beam was mounted vertically to a electromechanical shaker which provided a transverse sinusoidal excitation. The beam tip was placed near two rare earth magnets so as to provide nonlinear buckling forces. The beam was treated with a viscoelastic strip adhered to one side to provide a little damping. The addition of the damping treatment helps to form a more distinguishable fractal structure in phase space embeddings [58]. A laser vibrometer was used to record the beam tip velocity and the analog signal streamed to a National Instruments data acquisition board. The data was sampled at $F_s = 1000$ Hz and displayed in a custom built LabVIEW Virtual Instrument panel which displayed a buffered time history, periodogram, and phase plane plot. The excitation frequency was chosen as the first mode frequency, which turned out to be approximately 10 Hz. Driving the beam at its first mode guarantees the most beam tip response for a given excitation amplitude. This gives the beam tip enough kinetic energy to potentially overcome the forces of the rare earth magnets and display a wide range of dynamic behavior without incurring any damage to the shaker by driving it too hard.

A model of the beam experiment which describes the dynamic motion of the
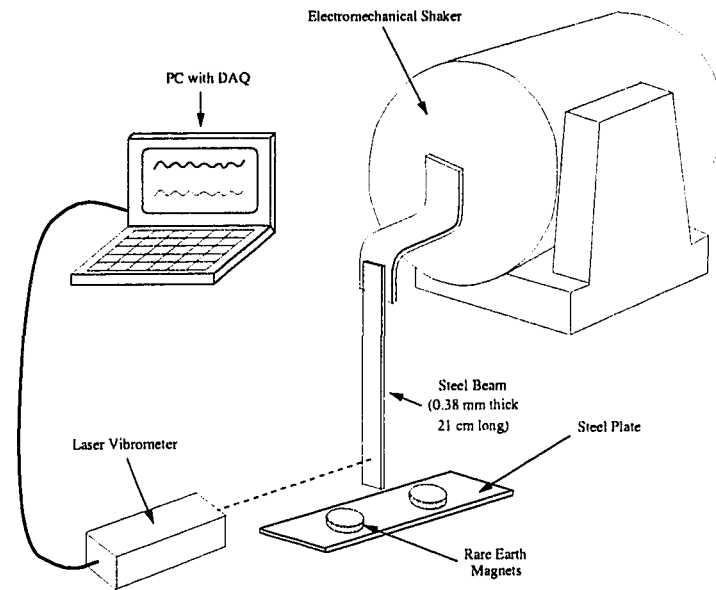
Figure 3.50: Setup for the chaotic beam experiment.

system is given by [58]

$$\ddot{x} + \gamma\dot{x} - \alpha x + \beta x^3 = A\cos(\omega t + \phi) \tag{3.44}$$

where x is the horizontal displacement of the beam tip relative to vertical, A is the excitation amplitude, ω is the excitation frequency in rad/sec, ϕ is the phase of the excitation, γ is the damping coefficient, and α and β are spring constants. The corresponding potential energy function

$$V(x) = -\alpha/2x^2 + \beta/4x^4. \tag{3.45}$$

A plot of the potential energy function is shown in Fig. 3.51. The shape of the potential function dictates the stability of the system. The critical points are located

$$V(x) = -(\alpha/2)\, x^2 + (\beta/4)\, x^4$$
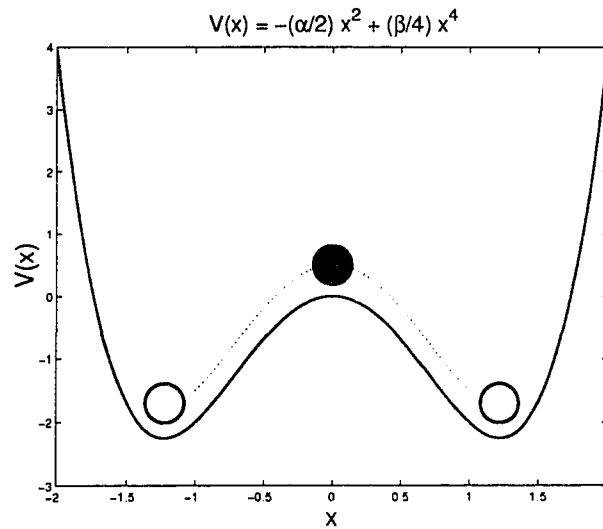
Figure 3.51: Double well potential function.

at $x = 0, \pm\sqrt{\alpha/\beta}$. The position $x = 0$ is stable iff $\alpha > 3$ while the equilibria $x = \pm\sqrt{\alpha/\beta}$ are both stable for $\alpha > 0$. The stability is analogous to a ball rolling on the surface of the potential function which forms a double well on either side of $x = 0$. The ball sitting at position $x = 0$ is obviously unstable since a small perturbation will drive it away from that location. The two wells on either side are locally stable. The beam tips reacts as does the ball in the potential curve as it is excited sinusoidally in the x-direction. If the excitation amplitude is small, the ball will oscillate about either well. In the experiment, this is analogous to the beam tip oscillating about one of the magnets. This motion is called *period 1* motion because it has only one fundamental period of oscillation. If the amplitude is increased

slightly, the ball may jump from one well to the other in a left-right pattern denoted as $LRLRLR$.... This pattern is period 2 and is analogous to the beam tip hovering about one magnet with a quick diversion towards the other magnet during each period 1 oscillation. A further increase in amplitude may produce a period 4 motion with a symbolic pattern $LRRLRRL$..., for example, and so on.

Each period doubling is referred to in the literature as a *bifurcation*. A bifurcation can occur when a parameter in the equations of motion changes. A sample bifurcation diagram is shown in Fig. 3.52. In the beam experiment, the parameter
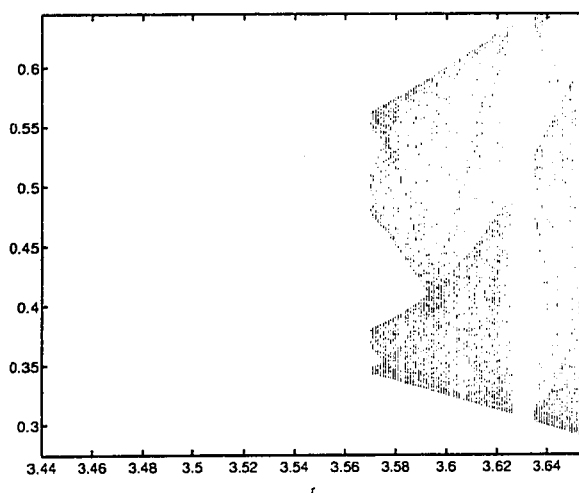
Figure 3.52: A bifurcation diagram for the Logistic map: $x_{n+1} = rx_n(1 - x_n)$ where n is the iteration index. The ordinate values are the iterate amplitudes plotted as a function of parameter r. For example, at $r = 3.44$ the system is at period one response. At $r = 3.445$, the system bifurcates into a period two response and so on. For a relatively large range of r values, the system is chaotic, represented by the wide span of extrema values.

change is through the excitation amplitude controlled by a Mcintosh analog ampli-

fier (not shown in Fig. 3.50), through an analog control. By varying the amplifier gain, a broad variety of dynamic behavior is exhibited. The system may go through a number of bifurcations before it destabilizes and goes into chaotic motion.

Figure 3.53 shows sample time histories for period 1, 2, and 4 motion as well as a chaotic response. For every period doubling, there must be a commensurate halving of the fundamental frequencies in the response spectrum. This can be seen in the periodograms shown in Fig. 3.54. As the system bifurcates towards chaos the spectrum broadens. Notice that the period 1 spectrum contains a few harmonics. This gives the phase plane embedding of the data a nonsymmetric pattern (Fig. 3.55). The phase plane is one of the easiest means of detecting bifurcations. Once the time lag is established, period 2 will exhibit twice as many embedding loops as does period 1, period 4 will have twice as many loops as period 2, and so on. A chaotic pattern emerges when the trajectories never return to their origin (never repeat). Of course, experimentally one can never verify chaotic motion absolutely, as one would need to record an infinite amount of data. What is shown as chaos in Fig. 3.55 is only speculated as such, but may in fact be a period n response where n is "a very large" integer.

To partially verify the validity of the Duffing-type model for the beam experiment, a digital force gauge with range of $0 - 100$ lbs was used to estimate the force deflection curve shown in Fig. 3.56. The beam was isolated from the influence of the magnets during the deflection test. A least squares fit of a function of the form
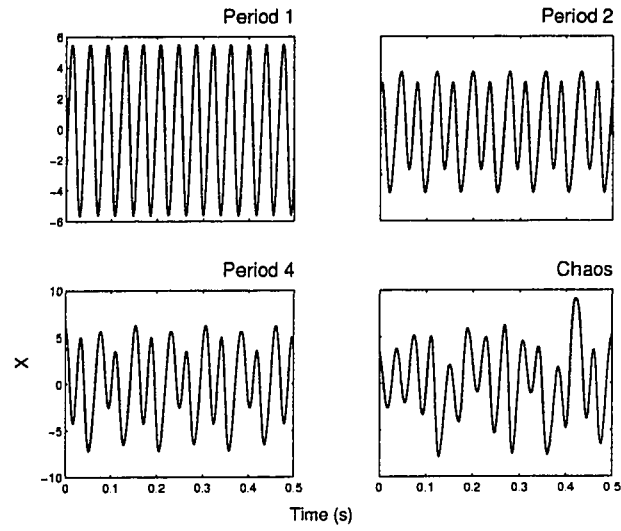
Figure 3.53: Time histories of period 1, 2, 4 and chaotic motion in the beam experiment.
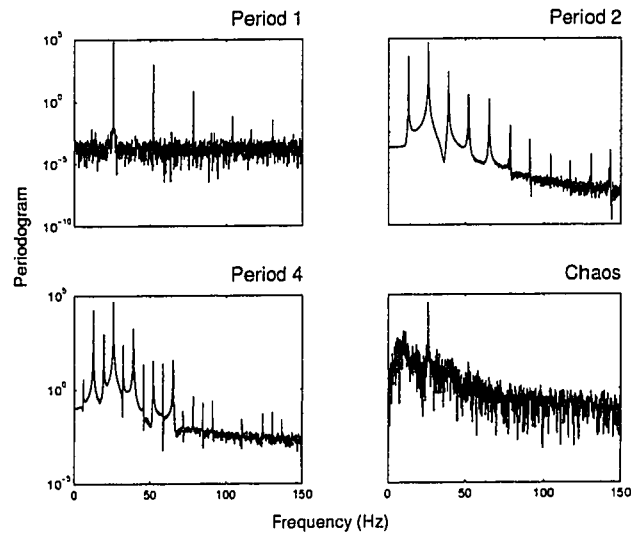
Figure 3.54: Spectra of period 1, 2, 4 and chaotic motion in the beam experiment.
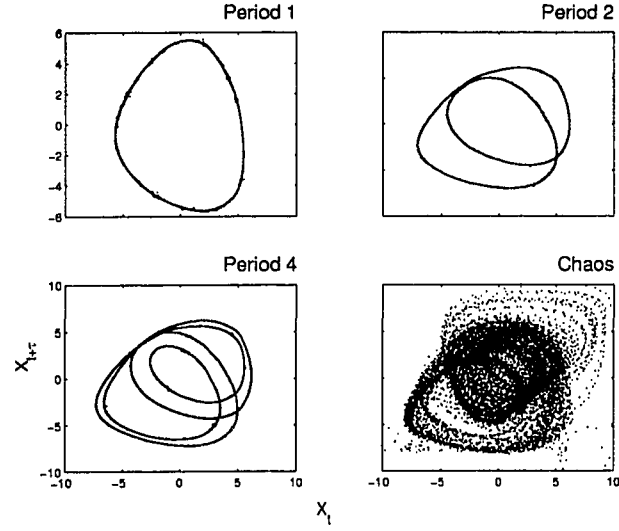
Figure 3.55: Phase plane embeddings of period 1, 2, 4 and chaotic motion in the beam experiment.

$f(x) = \alpha x + \beta x^3$ was found to be in good agreement with the raw data. After normalizing by the mass of the beam ($m \approx= 1/32^2$ slugs), the coefficients were approximated as $\alpha \approx 200$, and $\beta \approx 51$. Notice however that the linear spring constant is opposite in sign of α in Eq. 3.44. The reason for this is that at deflections close to $x = 0$ the magnets pull on the beam tip to either side, creating a repulsive force which pushes it away from equilibrium at $x = 0$. The magnetic force must offset the natural linear restoring force of the spring so that the combined linear force on the beam tip yields $\alpha > 3.0$. This assumption was validated by experimental observation of the system stability at $x = 0$.

The chaotic beam response was denoised via the MODRA and its D_2 estimate
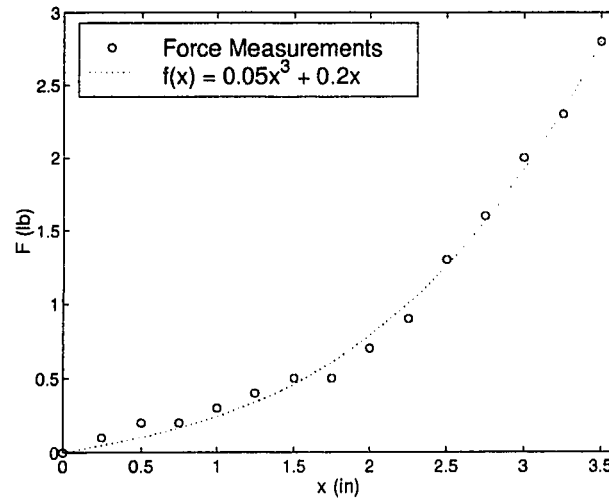
Figure 3.56: Force deflection curve due to transverse loading of the beam tip.

calculated (Fig. 3.57). Unfortunately, a high level of noise is not seen in the beam responses, so the MODRA is not tested to its full potential as a denoising scheme for observed variables in real experiments. However, the technique is still useful and valid as the correlation dimension is shown to converge to $D_2 \approx 2.5$ for the beam experiment and is in agreement with the D_2 results found without MODRA preconditioning (Fig. 3.58).
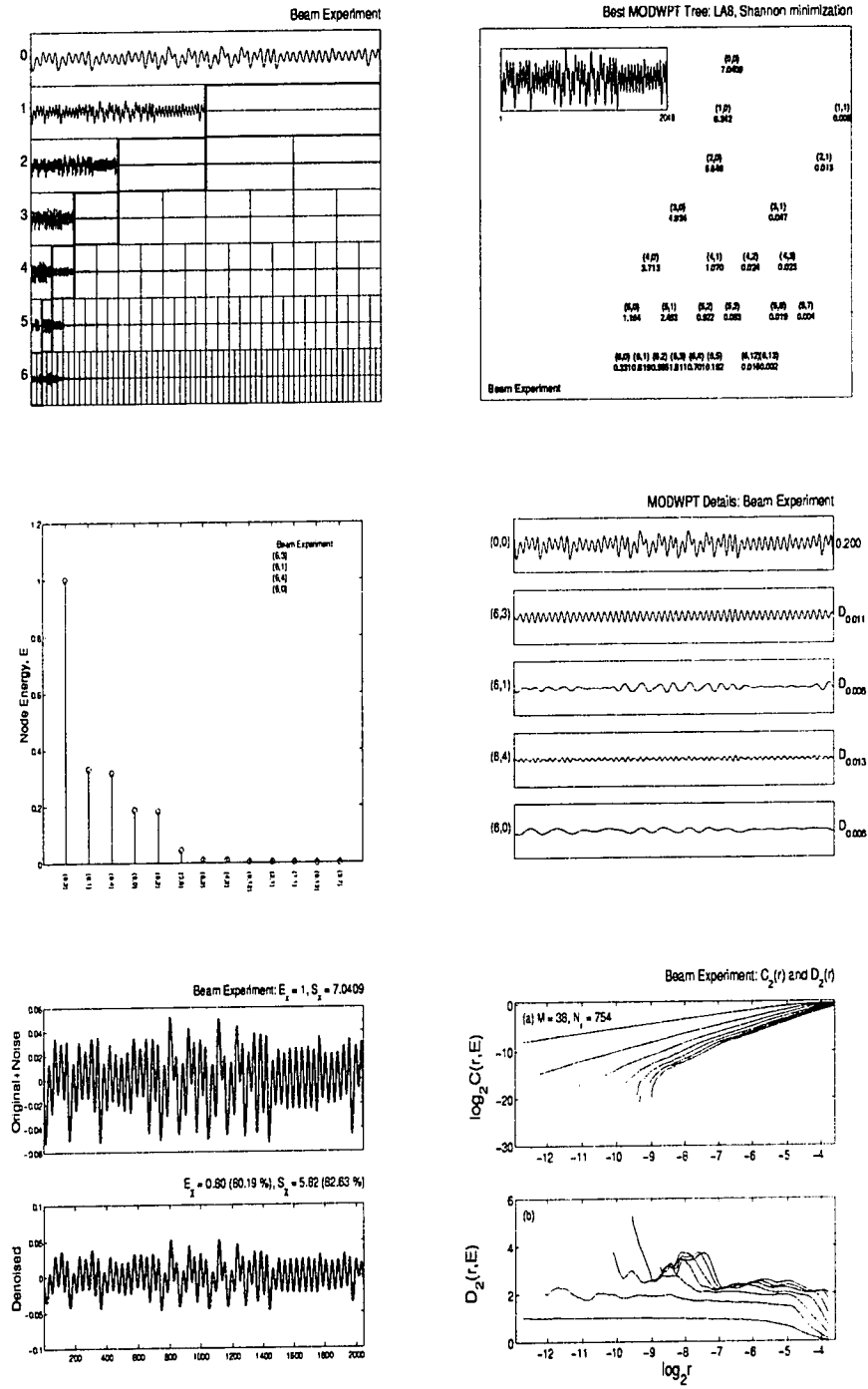
Figure 3.57: MODRA and D_2 results for the chaotic beam experiment.
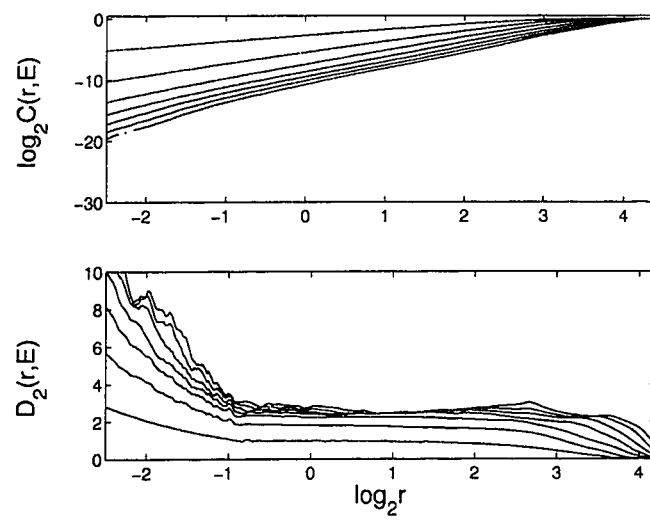
Figure 3.58: D_2 results for the chaotic beam response without MODRA denoising.

Chapter 4

SIMULATION TECHNIQUES AND D_2 FOR

$1/f^\alpha$ PROCESSES

Phase randomization techniques have often been used to simulate stochastic processes whose power spectra have a $1/f^\alpha$ falloff. We criticize phase randomization methods for their use in obtaining convergent correlation dimension results for these processes. Specifically, we show that phase randomization methods can impose an unnecessary deterministic structure in $1/f^\alpha$ noise simulations and generally do a poor job in emulating true colored noise processes. We introduce the Gaussian Spectral Synthesis method which overcomes these obstacles. We then discuss the difficulties in obtaining estimates of the global fractal dimensions of red noise processes. Specifically, we assess the probability of recurrent orbits for red noise and provide results which suggest that phase space analysis of red noise is computationally impractical. These results have implications for experimentalists interested in classifying fractal structure for sequences of unknown origin and confirms the notion that stochastic fractal processes do not yield convergent global generalized fractal dimensions if analyzed correctly.

4.1 Introduction

Dissipative nonlinear deterministic dynamic systems are capable of producing fractal topologies in the phase space commonly known as strange attractors. While there exists an infinite number of generalized fractal dimensions D_q for $q \in \mathbb{R}$ which characterize the structure of strange attractors [35], the most commonly used measure is the *correlation dimension* or D_2 due to its computational efficiency.

D_2 gained wide popularity among researchers who tried to determine if their measured, seemingly random, data were produced by a nonlinear deterministic process or a stochastic process. This was driven by the belief that D_2 would converge for data produced by a deterministic (chaotic) system and diverge for stochastic data. However, Osborne and Provenzale seemingly disproved this notion by demonstrating local D_2 convergence for stochastic processes with a power law spectral density using the phase randomization method as a colored noise simulation technique [67].

In the present work, we criticize the phase randomization method and discuss its pitfalls. Our goal is to reexamine *local D_2* convergence for $1/f^\alpha$ processes using a simulation technique whose statistical properties agree in a quantifiable sense with those of the process it seeks to emulate. As in [95], we define *local D_2* to mean the correlation dimension of a small segment of a trajectory in the phase space. We define *global D_2* to mean the correlation dimension of an embedding which contains a large number of orbits (say $> 10,000$). Using these definitions, we then address

global D_2 convergence issues for colored noise and show that D_2 does not converge for red noise for $\alpha > 1$. We then expand upon work done by Theiler [95] by examining the probabilities of nonrecurrent $1/f^\alpha$ noise for $\alpha \geq 2$. Finally, we focus on the limitations of using any $1/f^\alpha$ simulation technique when dealing with global measures in the phase space and argue that for a large class of colored noise, global phase space analysis is computationally impractical.

The outline of this chapter is as follows: In Section 4.2 we define $1/f^\alpha$ processes and in Section 4.3 we analyze three techniques used to simulate colored noise processes: the Direct Summation Methods (DSM), the Phase Randomization Method (PRM), and the Gaussian Spectral Synthesis Method (GSSM). In Section 4.4 we compare the PRM to the GSSM. In Section 4.5 we review the correlation dimension and discuss the subtleties involved in obtaining an accurate measure. In Section 4.6 we revisit the issue of local D_2 convergence using the GSSM. In Section 4.7 we provide numerical evidence of nonconvergent global correlation dimension using a combination of the GSSM and the DSM to simulate colored noise. We then address probabilities regarding the frequency of recurrent phase space orbits for $\alpha \geq 2$.

4.2 $1/f^\alpha$ Processes

If the power spectral density $S_X(f)$ of a discrete real-valued sequence X_t exhibits a power law behavior $S_X \sim 1/f^\alpha$, then X_t is referred to as $1/f^\alpha$ noise or *colored noise* (assuming $\alpha \neq 0$). For $\alpha > 0$, low frequency power dominates and is appropriately

coined *red noise*. Similarly, $\alpha < 0$ is *blue noise*, $\alpha = 0$ *white noise*, and $\alpha = 1$ *pink noise*, *flicker noise*, or simply *"one-over-f" noise*. Figure 4.1 summarizes the noise classifications for $1/f^\alpha$ processes.
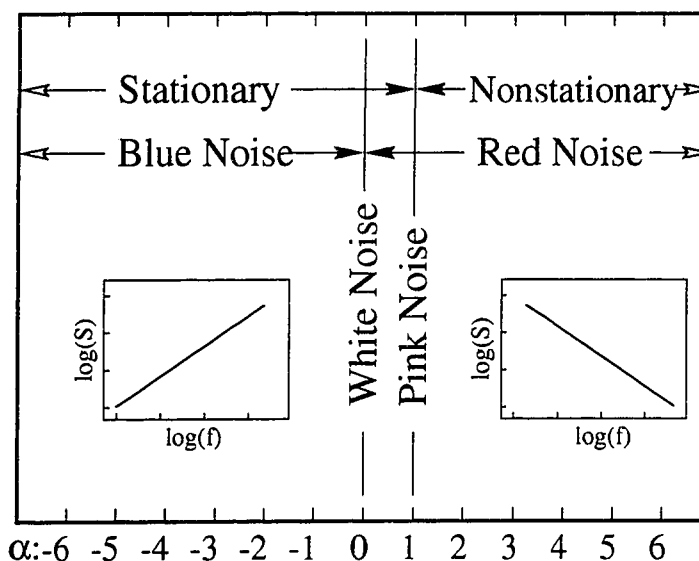
Figure 4.1: Spectrum exponent chart describing the classification and the stationarity for $1/f^\alpha$ processes as a function of the spectral coefficient α. The hollow arrowheads imply infinite extension in the direction in which they point while the solid arrowheads imply termination.

One-over-f behavior has been found in many fields of study including electronics [41, 49, 57, 87], biology [6, 14, 30, 34, 68, 83, 101], astronomy [11, 12, 31, 45, 54, 56, 59], music and speech [20, 44, 89, 99], magnetics [43, 65, 64, 78, 79], economics [17, 47, 53], transportation [15, 105], ecology [33, 74, 91], granular flow [10, 69, 86], chemistry [61], nuclear physics [3], optics [60], and fluids [70]. Red noise has been observed in fully developed turbulent flow [26, 27], blue noise in the image motion

of stars [46], and white noise as a common form of background noise.

4.3 Simulation Techniques for $1/f^\alpha$ Noise

Considering the abundance of $1/f^\alpha$ behavior reported in the literature, it is of considerable interest to develop simulation techniques that satisfactorily emulate these noise processes. In this section we discuss three simulation techniques and assess their drawbacks.

4.3.1 Direct Summation Method

Let $\{X_t : t = 0, \pm1, \pm2, \dots\}$ be a stochastic process whose d^{th} ordered backwards difference is given by

$$Y_t \equiv (1 - B)^d X_t = \sum_{k=0}^{d} \binom{d}{k} (-1)^k X_{t-k} \tag{4.1}$$

where $d \in \mathbb{N}$ and B is the backward shift operator defined by $B^k X_t \equiv X_{t-k}$. For example,

$$Y_t = \begin{cases} X_t & \text{if } d = 0, \\ X_t - X_{t-1} & \text{if } d = 1, \\ X_t - 2X_{t-1} + X_{t-2} & \text{if } d = 2. \end{cases} \tag{4.2}$$

If $\{X_t\}$ is a stationary process with power spectral density $S_X(f) \sim |f|^{-\alpha}$ for $|f| \le 1/2$ (implying that $\alpha < 1$), then $\{Y_t\}$ is a stationary process whose power spectral density is given by $S_Y(f) = \mathcal{D}^d(f) S_X(f)$ where $\mathcal{D}(f) \equiv 4\sin^2(\pi f)$ is the squared

gain function for a first order backwards difference filter. If however $\{X_t\}$ is a nonstationary process but $\{Y_t\}$ is stationary, S_X can be *defined* as $S_Y(f)/\mathcal{D}^d(f)$ [73, 104]. In this case, $\{X_t\}$ is a nonstationary process with stationary backwards differences.

A random walk sequence is one example of a nonstationary process with stationary backward differences which can be built by performing a cumulative summation over a stationary process (in this case a white noise process). Let $\{\epsilon_t\}$ be a discrete real-valued stationary white noise process with zero mean and variance σ_ϵ^2. The autocovariance sequence for $\{\epsilon_t\}$ is given by

$$s_\tau = \begin{cases} \sigma_\epsilon^2 & \text{if } \tau = 0, \\ \\ 0 & \text{otherwise.} \end{cases} \tag{4.3}$$

Using $\{\epsilon_t\}$, a *random walk process* is defined as

$$X_t \equiv \sum_0^t \epsilon_u. \tag{4.4}$$

Successive cumulative summations over the random walk sequence X_t can be performed, producing discrete $1/f^\alpha$ realizations associated with larger values of α. Let $X_{t,\alpha}$ represent a discrete $1/f^\alpha$ realization and define

$$X_{t,\alpha-2} = \sum_{u=0}^t X_{u,\alpha}. \tag{4.5}$$

This cumulative sum over $X_{t,\alpha}$ produces a sequence $X_{t,\alpha-2}$ which has a spectral density proportional to $f^{-(\alpha+2)}$. Figure 4.2 shows four $1/f^\alpha$ realizations generated with the direct summation technique.

160
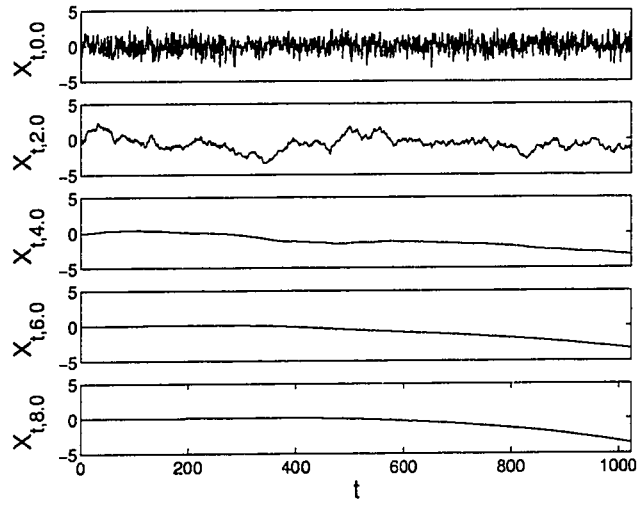

Figure 4.2: Discrete realizations of $1/f^\alpha$ processes for $\alpha = 0.2,4.6$, and 8 with $N = 1024$ using the direct summation method. Each realization is normalized by its sample standard deviation (so that the sample variance is unity) for ease of comparison with realizations created by the PRM (Fig. 4.3) and the GSSM (Fig. 4.4).

The direct summation method produces "exact" realizations of random walk processes. Discrete realizations however are never truly exact since (Gaussian) white noise is simulated via pseudo-random number generators which are in fact deterministic and periodic [103].

4.3.2 Phase Randomization Method

A disadvantage in using the direct summation method (Eq. 4.4) is that, if we limit ourselves to repetitive sums of white noise, α is limited to positive even integers. Fourier techniques can be used to approximate $1/f^\alpha$ realizations which carry no such restrictions. One such technique used in [67] utilizes the Fourier series summation

$$X_t = \sum_{t=1}^{N/2} \zeta_k \cos\left(\omega_k t + \phi_k\right), \quad t = 1, \ldots, N \tag{4.6}$$

where $\zeta_k = \sqrt{P(\omega_k)\Delta\omega}$, $P(\omega_k) = C\omega_k^{-\alpha}$, $\omega_k = k\Delta\omega$, $\Delta\omega = 2\pi/N$, and ϕ_k is a uniform random deviate on $[0, 2\pi]$ for $k = 1, \ldots, N/2$. The constant C is chosen such that the resulting time series has unit variance. An efficient means of generating realizations using Eq. 4.6 is through the explicit development of the discrete Fourier coefficients F_k followed by an inverse discrete Fourier transform operation. The following algorithm describes the Phase Randomization Method (PRM) for N an even integer.

Phase Randomization Method (PRM):

1. $F_0 = 0$

2. $F_k = \zeta_k cos(\phi_k) - i\zeta_k sin(\phi_k)$ for $k = 1, \ldots, (N/2 - 1)$

3. $F_{N/2} = \zeta_{N/2}$

4. $F_{N-k} = F_k^*$ for $k = 1, \ldots, (N/2 - 1)$

5. $X_{n,\alpha} = \mathcal{F}^{-1}(F_k) = \sum_{k=0}^{N-1} F_k e^{i2\pi kn/N}$

where $\mathcal{F}^{-1}()$ is the discrete inverse Fourier transform operator. The subscript α in $X_{n,\alpha}$ is used to emphasize that the colored noise realization is associated with a specific spectrum exponent α. Figure 4.3 shows five such realizations for various values of α (each $X_{n,\alpha}$ has been renormalized to have unit variance). A criticism of
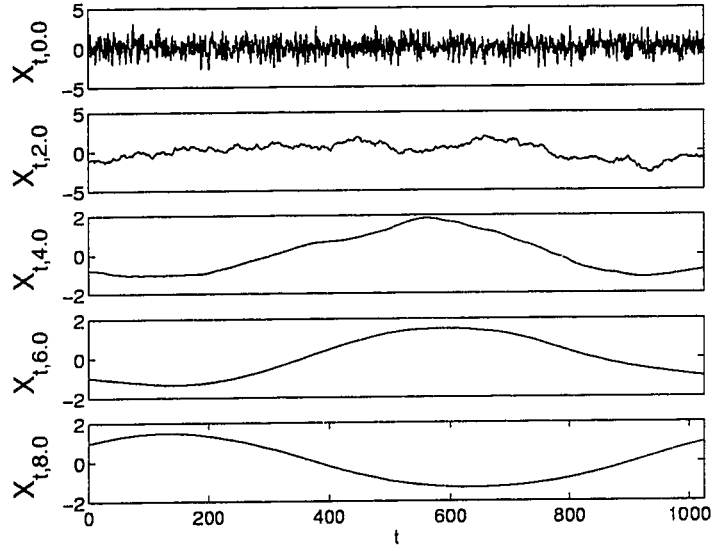

Figure 4.3: Zero mean, unit variance, stochastic PRM simulations of $1/f^\alpha$ processes for $\alpha = 0, 2, 4, 6,$ and 8 with $N = 1024$.

the PRM is that while it succeeds in generating discrete stochastic realizations with a seemingly correct power spectrum, it does a poor job in modeling the behavior of true $1/f^\alpha$ processes. An immediate observation is that the realizations are in fact periodic as each series ends where it began (see Fig. 4.3). A second criticism is that the frequency of zero crossings is higher than what is expected for larger values of α (as can be seen by comparing Fig. 4.3 with Fig. 4.2). A third criticism of the PRM is that only the phase of the spectral coefficients are randomized while the power is set *exactly* to a $1/f^\alpha$ falloff. This imposes an unnecessary deterministic structure in the data which can otherwise be avoided by using a more suitable method.

4.3.3   Gaussian Spectral Synthesis Method

A technique which overcomes the flaws of the PRM is the *Gaussian Spectral Synthesis Method (GSSM)* which explicitly randomizes both the power and phase components of the Fourier coefficients. The algorithm goes as follows:

Gaussian Spectral Synthesis Method (GSSM):

1. For a N point realization (N even), let $N' = MN$, $M \in \mathbb{N}$

2. Generate a Gaussian white noise sequence $\{Z_1, \ldots, Z_{N'}\}$

3. $C_0 = Z_1$, $C_{N'/2} = Z_{N'}$

4. $C_n = (Z_{2n} + iZ_{2n+1})/\sqrt{2}$, for $n = 1, \ldots, (N'/2) - 1$

5. For $f_n = n/N'$ define $F_n = C_n \sqrt{S_X(f_n)}$, $n = 0, \ldots, N'/2$

6. $F_n = F^*_{N'-n}$, $n > N'/2$

7. $X_{t,\alpha} = \mathcal{F}^{-1}(F_n) = \sum_{n=0}^{N'-1} F_n e^{i2\pi f_n t}$

where $S_X(f_n) = f_n^{-\alpha}$. Figure 4.4 shows GSSM realizations for various values of α. A good agreement with the direct summation method (Fig. 4.2) can be seen. For fractional Gaussian noise, recent developments in the literature show that the optimal selection of $S_X(0)$ is

$$S_X(0) = \frac{M}{2N-1} \left(\sum_{\tau=-(N-1)}^{N-1} s_{X,\tau} - \frac{1}{M} \sum_{j=1}^{M-1} S_X(f_j) \frac{\sin([2N-1]\pi f_j)}{\sin(\pi f_j)} \right)$$

(4.7)

where $s_{X,\tau} \equiv \mathrm{cov}\{X_t, X_{t+\tau}\} = \frac{\sigma_X^2}{2}(|\tau+1|^{2H} - 2|\tau|^{2H} + |\tau-1|^{2H})$, $\tau = 0, \pm 1, \pm 2$ is the acvs for a fGn with Hurst coefficient $0 < H < 1$ and $\sigma_X^2 \equiv \mathrm{var}\{X_t\} = s_{X,0}$ is the process variance assumed to be bounded [72]. Since $S_X(0)$ affects only the mean of the GSSM realizations and since D_2 is a second order statistic we arbitrarily set $S_X(0) = 0$.

4.4 Comparison of PRM and GSSM

Both the PRM and GSSM are actually stationary harmonic processes with periodic autocovariance sequences (ACVS) [71]. In the case of PRM, the period of the ACVS is N, so that the beginning and end of a time series of length N are highly correlated
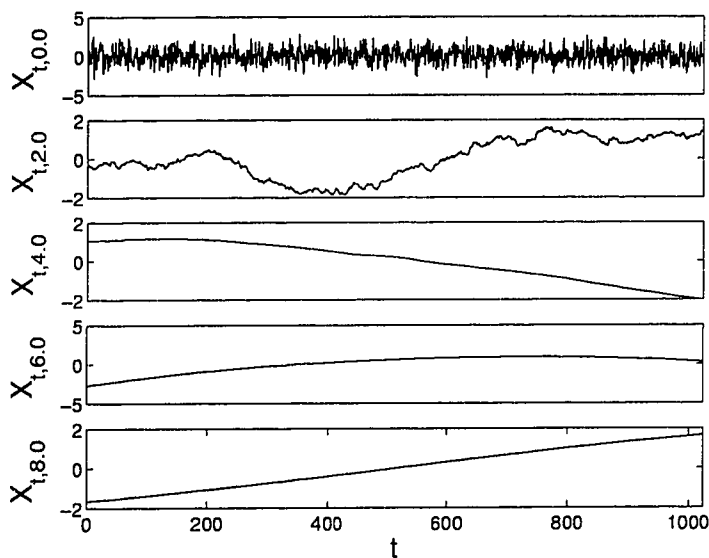
Figure 4.4: Discrete GSSM stochastic realizations of $1/f^{\alpha}$ processes for $\alpha = 2.0, 4.0, 6.0,$ and 8.0 with $N = 1024$, renormalized to have zero mean and unit variance.

with each other. In addition, the ACVSs for PRM and the process that it is intended to approximate typically differ quite significantly for lags greater than $N/2$. In the case of GSSM, the period of the ACVS is MN, but the agreement in the ACVSs between GSSM and the process to be approximated can be made arbitrarily good over lags ranging from 0 to $N - 1$ by increasing M. As a result, a proper choice of M means that the beginning and end of a length N series will have the correct covariance properties to a good approximation, which is the most serious deficiency of PRM as a simulation technique. Tests indicate that $M = 4$ yields a very good approximation for the kinds of processes considered in this paper. Finally, we note that GSSM generates Gaussian distributed deviates, whereas PRM does so only approximately

(the deviation from Gaussianity however is only noticeable for $N \leq 32$).

4.5 Correlation dimension, D_2

Fractal topologies with a nonuniform distribution of points in the phase space can be characterized by an infinite number of fractal dimensions D_q, where the subscript q denotes the probability of obtaining q points within a hypersphere with a given radius. The correlation dimension D_2 is a fractal dimension whose popularity is due to the computational efficiency and simplicity of developed algorithms [32, 93].

An important consideration for phase space analysis using a single variable time series is the selection of the time lag τ. The goal is to select τ such that the resulting coordinates are uncorrelated. Typically the lag corresponding to the first zero crossing of the average mutual information or autocorrelation function R_{xx}, or the first $1/e$ falloff of R_{xx} is used to estimate τ [1, 2, 24, 88, 94].

It has been claimed that if τ is chosen to be sufficiently large, then the D_2 estimates are independent of τ and should coincide with the results obtained using independent realizations as coordinates [67, 95]. We contend that using lagged coordinates for red noise processes is undesirable because red noise is highly correlated at large times, regardless of the lag. As an example, given that $\{X_t\}$ is a random walk process associated with a spectrum exponent $\alpha = 2$, the correlation coefficient

is given by

$$\rho_X \equiv corr\{X_t, X_{t+\tau}\} = \left\{ \frac{|t|+1}{|t+\tau|+1} \right\}^{1/2}. \tag{4.8}$$

Proof Random Walk Correlation and Stationarity

Let $\{\epsilon_t\}$ be a stationary white noise process with zero mean and variance σ_ϵ^2 and whose autocovariance sequence is given by

$$s_\tau = \begin{cases} \sigma_\epsilon^2 & \text{if } \tau = 0, \\ \\ 0 & \text{otherwise.} \end{cases} \tag{4.9}$$

Using $\{\epsilon_t\}$, a *random walk process* is defined as

$$X_t \equiv \sum_0^t \epsilon_u \tag{4.10}$$

Since $E\{\epsilon_u\} = 0$ for all u, it follows that

$$E\{X_t\} = \sum_0^t E\{\epsilon_u\} = 0 \tag{4.11}$$

for all t. Thus, the first moment of a random walk process is independent of time t as is required for a stationary process. For $\tau > 0$, the autocovariance of $\{X_t\}$ is given by

$$\begin{aligned} s_{X,\tau} \equiv cov\{X_t, X_{t+\tau}\} &= \{(X_t - E\{X_t\})(X_{t+\tau} - E\{X_t\})\} \\ &= E\{X_t X_{t+\tau}\} \\ &= \sum_{u=0}^{t} \sum_{u'=0}^{t+\tau} E\{\epsilon_u \epsilon_{u'}\} \\ &= (|t|+1)\sigma_\epsilon^2 \end{aligned} \tag{4.12}$$

since $E\{\epsilon_u \epsilon_{u'}\}$ is nonzero only if $u = u'$ which occurs exactly $t + 1$ times in the double summation. With the variance defined as $var\{X_t\} \equiv s_{X,0}$, it is clear that random walk processes are nonstationary because the second moment is a function of time t.

Using equation Eq. 4.12, the correlation coefficient ρ_X is

$$\rho_X \equiv corr\{X_t, X_{t+\tau}\} = \frac{cov\{X_t, X_{t+\tau}\}}{\sqrt{var\{X_t\}var\{X_{t+\tau}\}}} = \left\{\frac{|t| + 1}{|t + \tau| + 1}\right\}^{1/2} .$$
(4.13)

It is clear that $\rho \to 1$ as $t \to \infty$ implying that lagged random walk realizations become highly correlated as t increases, regardless of the lag τ. Higher order random walks, i.e. $1/f^\alpha$ processes associated with $\alpha > 2$, are even more correlated for a given combination of τ and t. Consider, for example, a random run process $\{Y_t\}$ associated with a spectrum exponent $\alpha = 4$. The correlation coefficient is given by

$$\rho_Y = \frac{s_{Y_t, \tau}}{\sqrt{s_{Y_t, 0} s_{Y_{t+\tau}, 0}}}$$
(4.14)

where

$$s_{Y_t, \tau} = \sum_{u=0}^{t} \left[\sum_{k=1}^{u+1} k + (|t + \tau| - u)(u + 1)\right] \sigma_\epsilon^2.$$
(4.15)

Proof Random Run Correlation and Stationarity

Let $\{X_t\}$ be a random walk process as defined in Eq. 4.10. Let $\{Y_t\}$ be a random

run process defined as

$$Y_t \equiv \sum_{u=0}^{t} X_u. \tag{4.16}$$

Since $E\{X_u\} = 0$ for all u (Eq. 4.11), it follows that

$$E\{Y_t\} = \sum_{u=0}^{t} E\{X_u\} = 0 \tag{4.17}$$

for all t. Thus, the first moment of a random run process is independent of time t.

The autocovariance of Y_t is given by

$$
\begin{aligned}
s_{Y,\tau} &= cov\{Y_t, Y_{t+\tau}\} \equiv E\{(Y_t - E\{Y_t\})(Y_{t+\tau} - E\{Y_t\})\} \\
&= E\{Y_t Y_{t+\tau}\} \\
&= \sum_{u=0}^{t} \sum_{u'=0}^{t+\tau} E\{X_u X_{u'}\} \tag{4.18}
\end{aligned}
$$

where $E\{X_t X_{t+\tau}\} = (|t| + 1)\sigma_\epsilon^2$ (Eq. 4.12). Incrementing the outer summation loop u, the following pattern evolves:

$$
\begin{aligned}
u = 0: \quad & E\{X_0 X_0\} = \sigma_\epsilon^2 \\
& E\{X_0 X_1\} = \sigma_\epsilon^2 \\
& \vdots \\
u = 1: \quad & E\{X_1 X_0\} = \sigma_\epsilon^2 \\
& E\{X_1 X_1\} = 2\sigma_\epsilon^2 \\
& E\{X_1 X_2\} = 2\sigma_\epsilon^2 \\
& \vdots
\end{aligned}
$$

$$u = 2: \quad E\{X_2 X_0\} = \sigma_\epsilon^2$$

$$E\{X_2 X_1\} = 2\sigma_\epsilon^2$$

$$E\{X_2 X_2\} = 3\sigma_\epsilon^2$$

$$E\{X_2 X_3\} = 3\sigma_\epsilon^2$$

$$\vdots$$

The sum T_u over each outer loop index u is

$$T_0 = \sigma_\epsilon^2 + (|t + \tau|)\sigma_\epsilon^2$$

$$T_1 = \sigma_\epsilon^2 + 2\sigma_\epsilon^2 + (|t + \tau| - 1)2\sigma_\epsilon^2$$

$$T_2 = \sigma_\epsilon^2 + 2\sigma_\epsilon^2 + 3\sigma_\epsilon^2 + (|t + \tau| - 2)3\sigma_\epsilon^2$$

$$T_3 = \sigma_\epsilon^2 + 2\sigma_\epsilon^2 + 3\sigma_\epsilon^2 + 4\sigma_\epsilon^2 + (|t + \tau| - 3)4\sigma_\epsilon^2$$

$$\vdots$$

$$T_u = \left[(|t + \tau| - u)(u + 1) + \sum_{k=1}^{u+1} k \right] \sigma_\epsilon^2.$$

With $s_{Y,\tau} = \sum_u T_u$ we arrive at

$$s_{Y_t,\tau} = \sum_{u=0}^{t} \left[\sum_{k=1}^{u+1} k + (|t + \tau| - u)(u + 1) \right] \sigma_\epsilon^2 \qquad (4.19)$$

Since $var\{Y_t\} \equiv s_{Y_t,0}$, ρ_Y becomes

$$\rho_Y = \frac{s_{Y_t,\tau}}{\sqrt{s_{Y_t,0} s_{Y_{t+\tau},0}}} \qquad (4.20)$$

Although not obvious, it can be shown that the random run correlation coefficient ρ_Y also tends to unity as $t \to \infty$. Figure 4.5 shows the convergence of the correlation coefficients for a random walk and random run process as a function of time for various lags. For small lags, the correlation coefficient rapidly converges toward unity
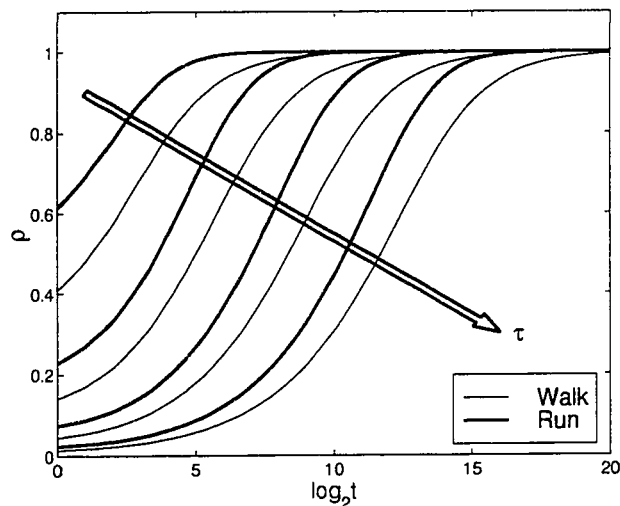

Figure 4.5: The correlation coefficient for a random walk process (thin lines) and a random run process (thick lines) as a function of time t and lag τ. The arrow indicates the direction of increasing lag, using $\tau = 10, 100, 1000,$ and 10000.

while for larger lags, the convergence is slower. It is also evident that random run sequences are more correlated than are random walks for a given t, τ combination. We expect these trends to continue for higher order random walks $(\alpha > 4)$.

Since each and every coordinate produced by the lagged embedding scheme is subject to high correlation, the neighboring points along a trajectory defined by those coordinates become highly correlated as the trajectory length grows. Care

should be taken to exclude these highly correlated neighbors when calculating the correlation integrals because $C(r, D_e)$ as defined by Eq. 2.11 are a valid means of estimating D_2 if and only if the neighbors used in the analysis are independent [55]. However, the analysis of long term dynamics for red noise processes requires "lengthy" coordinate realizations, making the avoidance of correlated neighbors a difficult task. One practical way of minimizing correlated neighbors in red noise embeddings is via a Poincaré section technique. The collection of Poincaré points is separated by large time intervals and provides a somewhat decorrelated data set for D_2 analysis. Secondly, a Poincaré section delivers a complete picture of the long term dynamics since it "captures" many periods of a given system's response, i.e. by collecting a set of N Poincaré points, it is assured that the set contains N fundamental periods of the system, whereas collecting N successive iterates does not ensure that a single fundamental period of the system has been captured.

4.6   Local D_2 for GSSM Realizations

Using lagged coordinates and independent realizations, we investigated local D_2 convergence for GSSM simulations of $1/f^\alpha$ noise. Figure 4.6 shows GSSM embeddings in the phase plane for various spectrum exponents using lagged coordinates. The points converge toward a single curvilinear path in the phase plane as α is increased. Since a uniform distribution of points along a line in the phase space results in a correlation dimension of $D_2 = 1$, we expect that, for large values of α, D_2 will

tend towards unity. Specifically, for $1/f^\alpha$ noise the *local* box counting dimension is

$D_0 = 2/(\alpha - 1)$ for $1 < \alpha < 3$ and $D_0 = 1$ for $\alpha \geq 3$ [95]. Since $D_0 > D_2$ with

$D_2 \simeq D_0$, it is expected that $D_2 \sim 2/(\alpha - 1)$ for $1 < \alpha < 3$.

As a means of comparison, phase plane embeddings of PRM realizations are

shown in Fig. 4.7. For larger values of α, true $1/f^\alpha$ embeddings will venture off into

space only to return to their origin after a large number of iterations or evolutions

in time (see Section 4.7 for more details). For PRM simulations, however, the orbits

always return to their origin regardless of α or iteration length.
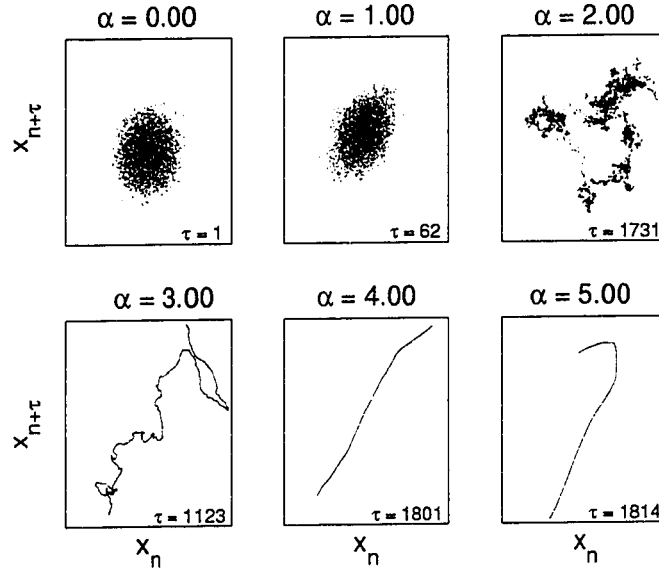

Figure 4.6: Phase plane embedding using lagged coordinates from a single GSSM $1/f^\alpha$ realization for $\alpha = 0.0, 1.0, 2.0, 3.0, 4.0$, and 5.0, with $N = 4096$. The temporal lag τ was selected based on the location of the first $1/e$ falloff in the autocorrelation function.

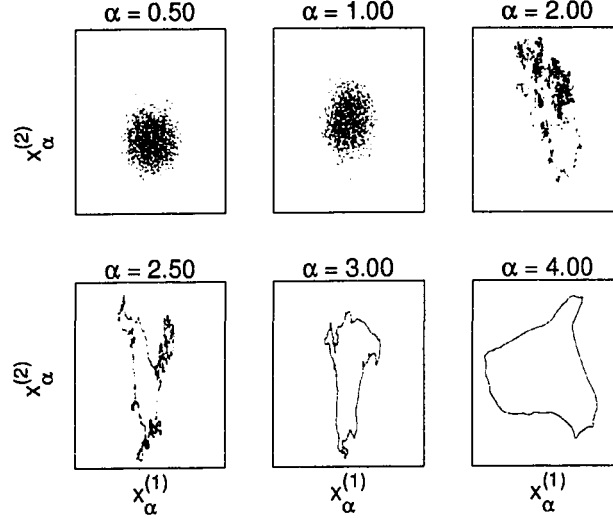Figure 4.8 shows the convergent behavior of D_2 for lagged GSSM realizations

Figure 4.7: Phase plane embedding of two PRM realizations for $\alpha = 0.0, 1.0, 2.0, 3.0, 4.0,$ and $5.0,$ with $N = 4096$. $X_\alpha^{(k)}$ represents the kth realization for a given α.

using $0.5 < \alpha < 7.5$. The outliers denote the range of D_2 estimates for a given value of α. Similar results were found using independent coordinates. The local D_2 results of the GSSM are in excellent agreement with those of the PRM [67] as well as the theoretical values. We emphasize that the GSSM does a better job in simulating $1/f^\alpha$ processes than does the PRM, and thus more diligently proves the local D_2 convergence for $1/f^\alpha$ processes. However, these results are accurate descriptions of only the short term dynamics and do not yield any relevant information regarding the long term dynamics of $1/f^\alpha$ processes.
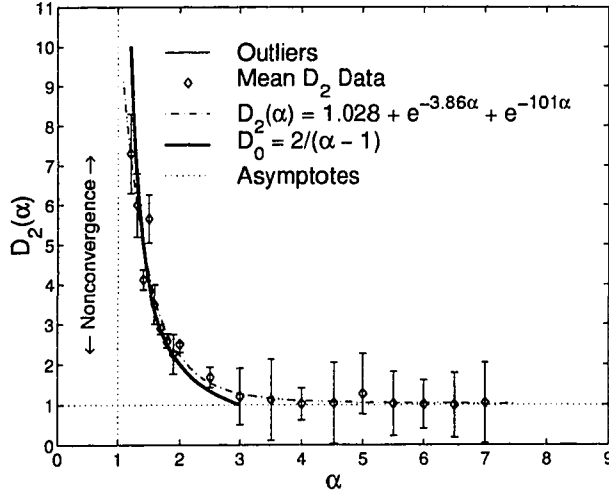
Figure 4.8: Summary of D_2 estimates for an ensemble of GSSM α's with $N = 2^{13}$, and $D_e = 1, \ldots, 15$. Shown is the estimated D_2 data, a least squares fit to the resulting curve, and the convergence asymptotes.

4.7 Global D_2 for GSSM/DSM Realizations

To obtain global D_2 estimates, the number of phase space orbits must be "large". The exact number of suitable orbits is a subjective topic of which there has been much research [7], but certainly an accurate portrayal of the global dynamics requires more than a few orbits in the phase space. The immediate difficulty one encounters in attempting to obtain many orbits for red noise realizations is that true red noise tends to wander off into space, returning (close) to its origin only after an extremely large number of iterations. To obtain a better understanding of the recurrence frequency of red noise orbits, we now examine the statistics of $1/f^\alpha$ processes for

176

$\alpha \geq 2$.

Given that a random walk is a red noise process associated with a spectrum exponent $\alpha = 2$, we will use probability theory to demonstrate the difficulty in obtaining a large number of phase space orbits for embedded random walk sequences. An excellent review of random walk statistics is given by Feller [22], and we direct the reader there for a more in-depth discussion.

We begin by examining a Poincaré section of random walk data X_n embedded in 3-space using a time lag embedding technique (Fig. 4.9). We arbitrarily choose
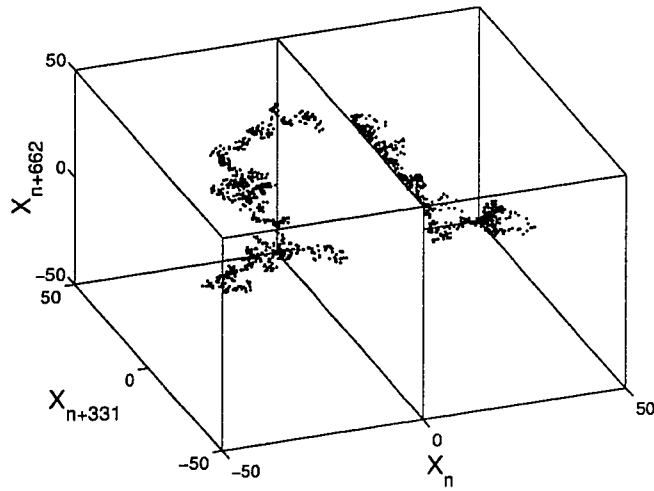

Figure 4.9: A Poincaré section illustration of a random walk realization embedded using a time lag $\tau = 331$ based on the average first $1/e$ falloff of the autocorrelation function over 100 realizations.

the $X_{n+\tau}, X_{n+2\tau}$ plane passing through $X_n = 0$ as a Poincaré section so that if a zero crossing is found at X_{100}, for example, the Poincaré intersection is taken as the

location $\{X_{100}, X_{100+\tau}, X_{100+2\tau}\} = \{0, X_{100+\tau}, X_{100+2\tau}\}$. Collecting a large number of Poincaré points refines the correlation integral structure at small scales and provides a more accurate portrayal of the long term dynamics of the system. Consequently. it is appropriate to investigate the probability theory associated with random walk zero crossings. To do so we examine two theorems, one relating the probability of the final return to zero, and the other the total number of expected zero crossings for long walks [22].

Theorem 1 The probability $A(x)$ that the last return to zero occurred on or before epoch $2k$ and did not return thereafter up to and including epoch $2n$ is

$$A(x) = \sum_{k < xn} f_{2k,2n} \approx \frac{2}{\pi} \arcsin \sqrt{x} \qquad (4.21)$$

for $0 < x < 1$, $x = k/n$, and

$$f_{2k,2n} = \binom{2k}{k}\binom{2n - 2k}{n - k} 2^{-2n}. \qquad (4.22)$$

Theorem 1 can be best illustrated through an example. Consider a thought experiment where a great many random walks are conducted simultaneously at an iteration rate of one step per second for 10 straight years (equivalent to 315.36 e6 iterations). On average, in one out of twenty walks the last return to zero will occur before 22.5 days have passed and will *not* return for the remaining 3,627.5 days. In one out of ten walks, the last return will occur before 90 days. and in one out of

5 walks it will occur before 349 days. These results are strikingly nonintuitive and suggest that obtaining a sufficient number of "hits" on a given Poincaré section may be difficult. The question then becomes, have we collected a sufficient number of Poincaré points before the trajectory ventures off into space? The following theorem provides insight into this issue.

Theorem 2 The probability ξ that fewer than $x\sqrt{n}$ sign changes occur before epoch n approaches

$$\xi(x) = \frac{2}{\sqrt{2\pi}} \int_{-\infty}^{2x} e^{-y^2/2} dy - 1 \tag{4.23}$$

as $n \to \infty$ for $x \geq 0$.

One might intuitively expect that if a given number of returns to zero, say N_0, were to occur during the first million iterations, then after the next million iterations there would be a total of approximately $2N_0$ returns to zero. This is not the case however as we see that the total number of zero crossings is proportional to $\sqrt{n}$ so that as the iteration number increases the rate of zero crossings diminishes. As an example, let us return to our 10 year experiment with the goal of obtaining N_0 returns to zero. Table 4.1 shows the probability ξ that less than N_0 zero crossings will occur for $n = 315.36\,e6$ iterations. According to Table 4.1, approximately 90% of the random walks on average will return to zero less than $15,000$ times. We put this prediction to the test by simulating a random walk process via the direct summation technique (Eq. 4.4) using a pseudo-random number generator with a

Table 4.1: The probability ξ that less than N_0 zero crossings will occur for the 10 year Gedanken experiment.

N_0	100	1,000	5,000	10,000	15,000	20,000	25,000	30,000
ξ	.0094	.0901	.427	.7401	.9089	.9757	.9951	.9993

theoretical period of 2^{1492} [23, 42]. The total number of iterates needed to achieve $N_0 = 15,000$ zero crossings was over one billion ($n = 1.1098e9$), indicating a zero crossing per every 74,000 iterations on average; a result which is commensurate with the trends shown in Table 4.1.

The implication of Theorem 1 and 2 is that obtaining a sufficient number of zero crossings is a difficult yet achievable task. Using the zero crossings as a means of identifying Poincaré points, we are now ready to explore the global D_2 convergence for $\alpha = 2$. Phase space coordinates were developed with the 15,000 point Poincaré sequence using a time delay embedding scheme. The lag was chosen as the average of the first $1/e$ crossing of the autocorrelation function for 100 realizations. Figure 4.10 shows the correlation dimension estimates, as a function of embedding dimension D_e and scale r, found by fitting the correlation integrals ($\log C$ versus $\log r$) using a linear least squares regression. Clearly, the D_2 estimates do *not* converge with embedding dimension and in fact D_2 is seen to grow in proportion to D_e as is expected for a stochastic process. The same results were found using separate realizations as

independent coordinates.

To explore global red noise D_2 convergence trends for $1 \leq \alpha < 2$, a GSSM realization corresponding to a spectrum exponent of $\alpha - 2$ was generated, followed by a discrete cumulative summation; effectively raising the spectrum exponent to α. Poincaré coordinates were produced using the $1/e$ lag scheme and D_2 was estimated (Fig. 4.10). Again we find that the D_2 estimates grow in proportion to D_e, signifying a lack of convergence.

These findings imply that the convergent D_2 shown in Fig. 4.8 are in fact anomalous global characterizations and cannot be used to quantify the long term dynamics of random walk processes.
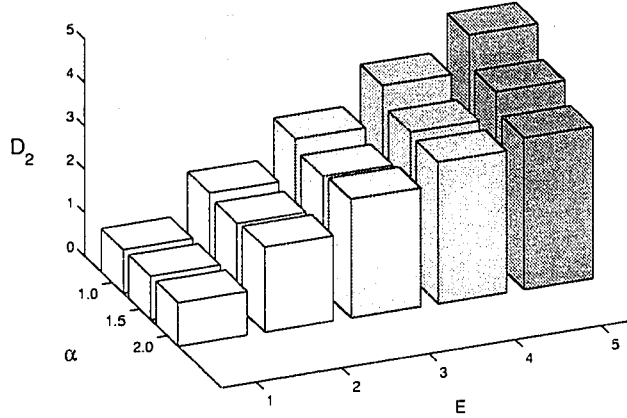
Figure 4.10: A linear least squares estimate of the slope of the correlation integrals a function of embedding dimension D_e. It is clear that the correlation integrals do not converge to a finite low dimension contrary to the results obtained for the GSSM realizations.

We assert that D_2 will not converge for larger spectrum exponents ($\alpha > 2$). It is very difficult (if not impossible) however to verify this assertion through numerical experiments because $1/f^\alpha$ processes associated with an $\alpha > 2$ have extremely small recurrency rates. We show this in the following analysis by obtaining the exact probability of a return to zero for $1/f^\alpha$ processes with $\alpha \geq 2$.

The direct summation method was used to obtain random walks of "higher order" and the exact recurrence probabilities for each series was calculated. The initial random walk sequence was restricted to step exactly ± 1 unit in amplitude at each iteration. Since there are exactly 2^n possible paths for n iterations, the probability $\xi_{n,k}$ of returning to zero is $\xi_{n,k} = N_{n,k}/2^n$ where $N_{n,k}$ is the number of paths that return exactly to zero at iteration n for a random walk order k ($k = 1$ is a random walk, $k = 2$ is a random run, and so on). Figure 4.11 shows the return to zero probability results for orders $k = 1, \ldots, 4$ and iteration levels $n = 1, \ldots, 16$. It is not possible for some walkers to return to zero at specific iteration values and as a result there are gaps in the return sequences. The equations at the end of each return sequence are the least squares approximations while the dashed lines are the corresponding curve fits. All orders exhibit have a power law falloff whose exponent decreases approximately by one as the return order is incremented.

The exact return to zero probability for a random walk sequence is given by

$$\xi_{n,1} = \binom{n}{n/2} 2^{-n} = \frac{n! 2^{-n}}{(n/2)!^2} \qquad (4.24)$$
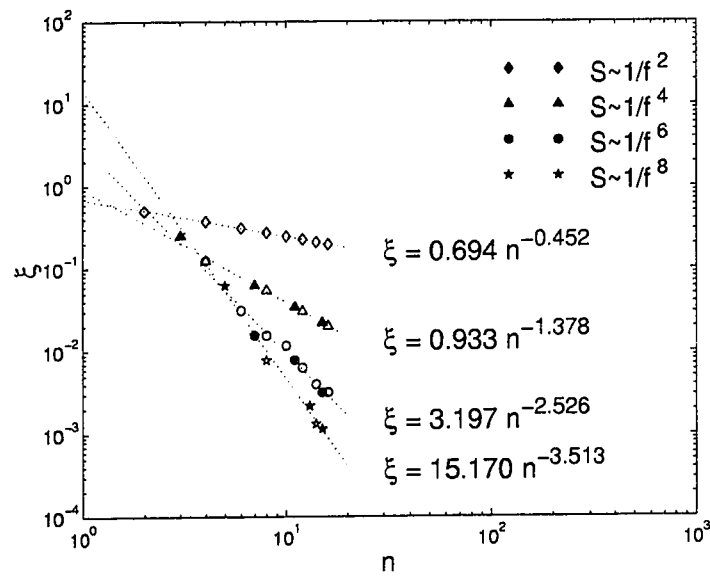
Figure 4.11: The exact probability ξ of a return to zero for a random walk process ($S \sim f^{-2}$), a random run process ($S \sim f^{-4}$), and higher order runs for the first 16 iterations. The hollow markers represent even valued returns while the solid markers represent odd valued returns. There is an apparent power law drop off in ξ for each process which can be used to estimate the return to zero probability for larger iteration values.

for n a positive even integer [22]. Using Stirling's formula: $n! \sim \sqrt{2\pi} n^{n+1} e^{-n}$, Eq. 4.24 reduces to

$$\xi_{n,1} \sim \sqrt{\frac{2}{\pi n}} = 0.7979\, n^{-1/2} \qquad (4.25)$$

which is in good agreement with the least squares fit shown in Fig. 4.11. From the results, one may speculate that an estimate of the return to zero probability for a kth order walk at iteration n is $\xi_{n,k} \sim n^{(1-2k)/2}$. Since the probability of the last return to zero at epoch n is proportional to the total returns to zero up to and including epoch n [22], one can conclude that obtaining a satisfactory number of zero crossings for orders $k > 1$ is an arduous task. Since stochastic realizations on a computer invoke the use of pseudo-random number generators which inevitably repeat themselves, obtaining a sufficient number of zero crossings, and equivalently a sufficient number of Poincaré points, is numerically impractical. We would like emphasize that the trends shown in Fig. 4.11 are exact and consequently *all* legitimate $1/f^\alpha$ simulation techniques are subject to the same probability trends.

4.8 Summary and conclusions

It is shown here that $1/f^\alpha$ stochastic fractals, which can be characterized by statistical self similarity measures, have a convergent local correlation dimension for $\alpha > 1.0$ but do not have a corresponding convergent global correlation dimension. Red noise processes maintain local self-similarity patterns in a statistical sense that

may lead one to incorrectly conclude that they have a relevant and convergent global fractal dimension.

It is shown that lagged coordinates do not become independent because the correlation coefficient for a random walk and random run process approaches unity as time $t \rightarrow \infty$, regardless of the lag τ. However, lagged coordinates developed by finding the zero crossings of red noise realizations help to sufficiently decorrelate the data and were shown to produce a nonconvergent global correlation dimension.

A similar analysis for random walks of higher order was shown to be computationally difficult due to the low recurrency trends of the phase space orbits for $\alpha \geq 2$. Specifically, we numerically demonstrated that the probability of recurrence ξ for a kth order random walk goes as $\xi \sim n^{(1-2k)/2}$ where n is the number of iterations and $k \geq 1$. Because the spectrum exponent α is related to the walk order k as $\alpha \Leftrightarrow 2k$, it becomes difficult to analyze D_2 convergence for $\alpha > 2.0$ using computer simulations.

Finally, we showed that the phase randomization method (PRM), used by [67] to explore local fractal structure, produces realizations which are highly correlated and have an underlying deterministic structure. The GSSM produces realizations which are more representative of stochastic $1/f^\alpha$ processes. Using GSSM realizations to create phase space coordinates, local D_2 estimates were shown to converge to the theoretical local box counting dimension of $D_0 = 2/(\alpha - 1)$ for $1 < \alpha < 3$ and $D_0 = 1$ for $\alpha \geq 3$.

Chapter 5

LINEAR AND NONLINEAR CLASSIFICATION OF

HUMAN CARDIAC RHYTHMS

Time series analysis is a common technique used to gain insight into biological systems. This is especially true when the system cannot be analyzed through a synthesis of behavior at the molecular. cellular. or organ level. The response of the system under normal operating conditions, or under conditions somewhat changed by the use of external stimuli. becomes the principal window into the underlying mechanisms.

In recent years, the application of concepts from nonlinear dynamics have gained popularity in time series analysis. This development has been driven by the discovery that simple nonlinear dynamic systems can produce chaotic responses. Numerous investigators have shown that mathematical models of simple biological systems can in fact produce chaotic solutions. The need to create mathematical models and to characterize the system based solely on experimental data requires the establishment of time series techniques which can discriminate between chaotic and stochastic behavior.

Time series analysis using methods from nonlinear dynamics may find an important role in modeling the human heart and in predicting complex responses. The behavior of the dynamics of the human heart and underlying mechanisms have been studied for decades with the help of the electrocardiogram (ECG). Cardiologists have studied ECGs and learned to infer a range of different disease states from subtle variations in the cardiac signal. Detection of disease states with the ECG coupled with information obtained from modern time series analyses presents a unique opportunity to further disclose the underlying mechanisms of the healthy and diseased heart.

The potential existence of nonlinear cardiac mechanisms has been demonstrated in numerous studies. Noteworthy examples of such findings are in studies by Savino et al [85] and Glass, Goldberger, and Belair [28]. Savino et al showed that the trace of successive firing intervals can exhibit period-doubling bifurcations in small samples of in vitro paced heart tissue. Glass, Goldberger, and Belair investigated an abnormal heart rhythm called parasystole, and developed a nonlinear mathematical model capable of predicting the observed behavior.

Variations in sinus rhythm have long been investigated and used in cardiac diagnostics. However, standard heart rate analysis based on linear theory may not be sufficient in characterizing the cardiac system. For example, it has recently been suggested that variations of the RR intervals are chaotic during normal sinus rhythm. This would imply that small variations in the heart beat are governed by determin-

istic rules.

In response to these issues and to the abundance of RR interval analysis reported in the literature [4, 29, 39, 90], we thoroughly examine RR interval data extracted from ECG Holter tape recordings of 30 patients with a known history of cardiac disease, from implantable cardioverter defibrillator (ICD) data of 18 patients who have experienced ventricular fibrillation (VF) and/or ventricular tachycardia (VT), and from 5 Holter tape recordings of healthy patients. We pose five main questions in our analysis:

1. Are RR intervals generated by a deterministic or stochastic process?

2. Is there any topological fractal structure in RR intervals, indicating the existence of an underlying low order deterministic chaotic system?

3. Do RR intervals exhibit any spectral scaling, indicating statistical self similarity common to colored random noise processes?

4. Can any of the measures be used to predict the onset of VF/VT?

5. Can cardiac arrythmias be classified by a MODWPT best basis for a representative sample of a patient's RR intervals?

The techniques used to explore these issues are Kaplan's determinism method (KDM), the correlation dimension (D_2), the MODWT variance $(\hat{\nu}_X^2(\tau, j))$, the approximate

entropy (ApEn), the MODWPT best basis, the bridge-detrended scaled windowed variance (BDSWV), and the dispersion analysis (Disp). Care is taken to address the subtleties involved in obtaining a good measure for each method. For Holter tape data, the sources of noise are scrutinized to avoid spurious assessment of the results.

5.1 Feature Extraction

A wavelet based feature extraction program was developed to effectively isolate the RR intervals from Holter tape ECG recordings. To improve temporal resolution, a cubic spline fit was applied to approximate the peaks of QRS complexes. The differences (in time) between successive peaks were recorded as the RR intervals. Fig. 5.1 shows a sample Holter ECG series with marked QRS peaks. Intervals suspected to be associated with either preventricle contractions (PVCs) or noise contamination were not included in the RR estimates. The hollow circles in Fig. 5.1 indicate QRS peaks of normal beats while the filled circles represent suspected areas associated with PVCs and/or noise. A PVC is characterized by a short beat followed by a long one such that the sum of the two intervals is approximately equal to the time it takes for two normal beats to occur. These are easily identified by inspection in the plot and verify that the wavelet based extraction program does a good job in isolating these components. Wavelet transforms are an excellent means of denoising ECG sequences because they effectively isolate low baseline drift due to respiration as well as most of the high frequency noise in the signal (see [73] for example). The main

steps in the algorithm are to (1) Decompose an ECG sequence in a 6 level DWT using Daubechies' least asymmetric 8 tap filters (2) Synthesize with the wavelet coefficients at decomposition levels 2-4. This effectively rids the base line drift component residing in the scaling coefficients and excludes the high frequency noise found in small scale wavelet coefficients (3) Apply conventional amplitude and amplitude-derivative based algorithms to identify the locations of the QRS complexes (4) Verify the peak locations by comparing them to the original signal. This comparison was necessary since some peaks found by the wavelet algorithm were off by one point to either side of the true peak locations.
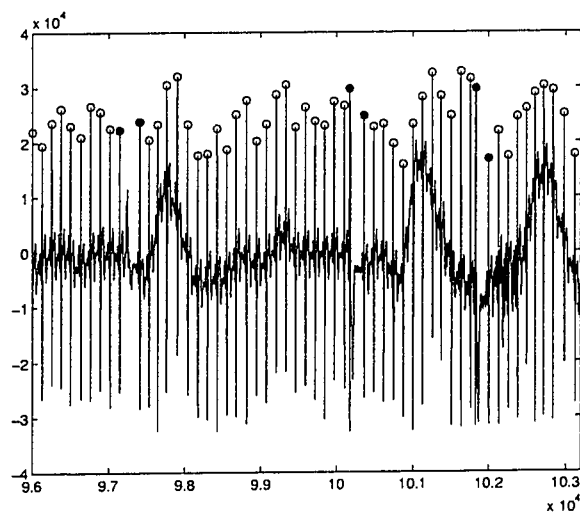
Figure 5.1: Example of RR extraction with PVC/noise detection. From a signal processing perspective, the ECG sequence is seen to be contaminated with high frequency noise and a low frequency baseline drift due to respiration.

5.2 Sources of Error

With the exception to the RR data received from the SCDHeFT program headed by Dr. Bardy of the University of Washington Medical Center, the ECG data was recorded using equipment developed by Northeast Monitoring. The comments in this section refer to Northeast Monitoring data.

There are numerous difficulties in performing accurate time series analysis on RR intervals extracted from real ECG data. Feature extraction of externally monitored ECG signals, especially ambulatory data, is a complex issue due to the abundance of noise that may be introduced during recording. The ECG can be contaminated from electrode contact noise, motion artifact, muscle contractions, baseline drift and/or ECG modulation with respiration, poorly shielded electrical devices in the signal processor, and electrosurgical noise [25]. The quality of ambulatory data is also restricted by hardware limitations affecting sampling frequency and SNR. Digital sampling of ECG data recorded on mechanical hardware such as tape drives can induce noise through quantization error, tape playback noise, low sampling rates, and nonlinear filtering operations.

Quantization effects and *nonlinear filtering* operations introduce errors in amplitude. The data was extracted from 24-hour Holter tapes via a 12-Bit A/D converter, yielding a 12.5 μV quantization error for the least significant bit (LSB) of a given sample. A nonlinear filter was applied to reduce the amount of data. The error

induced in averaging with a nonlinear filter is proportional to the difference in amplitude between adjacent samples, resulting in significant errors for large amplitude impulsive behavior such as that found in the QRS complex. The maximum filtering error for a given sample is $e_{NL} = \pm 50 * 12.5\mu V = .625mV$. A correction algorithm was used by the Northeast Monitoring Company to ensure that the error did not accumulate, affecting only $2 - 3$ successive samples in the ECG record.

Discretization of ECG signals via mechanical hardware can also produce errors in the timing of the data, known as *tape playback noise*. Mechanical tape devices used to read Holter tapes induce noise via inconsistent playback speeds during A/D conversion. Corrective measures can always be performed to adjust for tape speed variations but, depending on the quality of the correction, may not result in perfectly timed data extraction. For the system used to extract the ECG data, the temporal noise is estimated to be $\mu = \pm 1.5$ milliseconds.

Sampling rates play a large role in the accuracy of the discretization process. The ECG data was sampled at $F_s = 180Hz$, well above the Nyquist rate estimated at $F_{NYQ} \approx 25Hz$. Although aliasing is avoided, the sampling rate is seen to be deficient in that the number of points which define the QRS complex is quite small (@ $8-10$ points). For accurate estimation of R-R intervals, a curve fitting algorithm can be applied to the R-wave peaks. However, with the knowledge that the data was previously nonlinearly filtered, we are faced with an additional dilemma in estimating peak locations because samples residing near the R-wave peak are inaccurate. Points

near the base of the QRS complex are more accurate since they are less affected by the nonlinear filtering operation. To take advantage of accurate QRS base data, a cubic spline technique was chosen to estimate QRS peaks. Forcing the fit through each point of the QRS complex reduces the error that may be incurred otherwise, such as in polynomial fits. An illustrative example of the spline error using real ECG data is shown in Fig. 5.2. The curves in Fig. 5.2 represent two possible scenarios for a typical discretization of a QRS complex. The boxes denote a set of points where the nonlinear filtering has undesirably imposed a large change in amplitude between the second and third samples located at times 0.0153-0.0205 seconds. The filled circles denote a similar discretization which did not adversely impose such a gap. Communications with Northeast Monitoring has revealed that this gap is not a quantization effect but is in fact an effect of the nonlinear filtering. The specifics on the nonlinear filters used was not made available for public knowledge. The diamonds in Fig. 5.2 represent the estimated peak locations as found by the cubic spline technique. The difference in time between the peaks is denoted as δ and is seen as a source of error in resolving the location of the QRS peaks in time. If embedding techniques are used to analyze a discrete data set, any error that is inherent in the data is unfortunately accentuated by the embedding process: the error in calculating Euclidean distances between points in space grows nonlinearly with the embedding dimension (Fig. 5.3). It is essential then to understand all aspects of inherent noise in the ECG data before attempting to ascertain dynamic
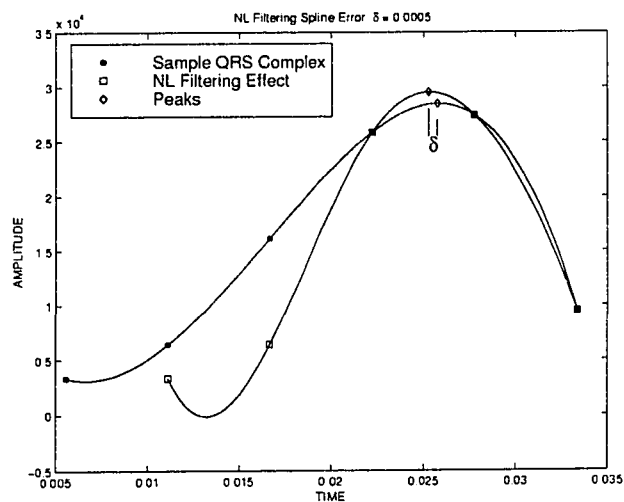
Figure 5.2: Error of cubic spline fit for QRS complex
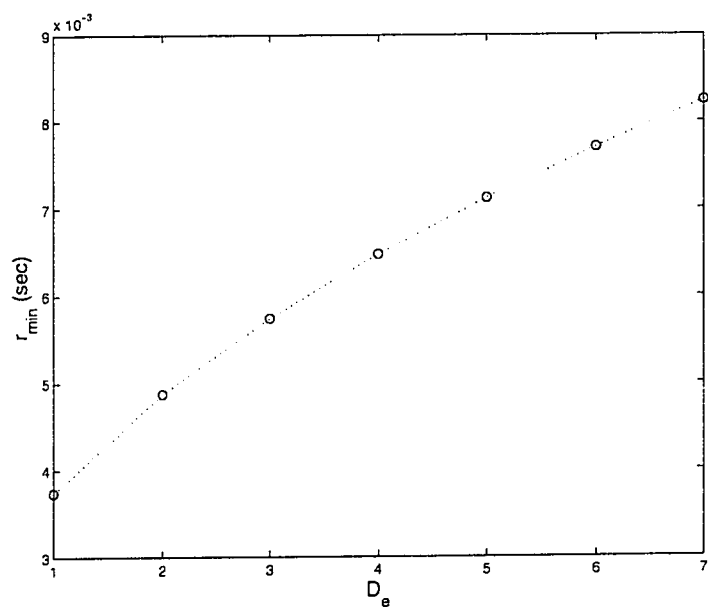
Figure 5.3: The minimum scale as a function of embedding dimension. Since the signal to be embedded is R-R interval data, the units of r are in seconds.

characteristics through embedding techniques. If the magnitude of the total error is well defined a priori, a minimum scale r_{min} can be established, which relates the smallest window in space over which the data can be viewed before the inherent error of the time series occludes measures made within the window. Combining all the possible sources of error in the data, the function r_{min} is approximated by

$$r_{min}(D_e) = \frac{N_{SP}(1/F_s + \mu)}{M} \sqrt{D_e} + \delta \qquad (5.1)$$

where

N_{SP} Number of samples in QRS complex fit with a cubic spline

D_e Embedding Dimension

F_s Sampling frequency of ECG data in Hertz

M Number of points used to fit the QRS complex using a cubic spline

μ Tape noise error in seconds

δ Offset of splined peak due to nonlinear filtering operation

The values for the parameters in Eq. 5.1 were set as: $N_{SP} = 8$, $M = 100$, $\mu = 0.003$ seconds, and $\delta = 0.001$ seconds.

5.3 Description of measures

A description of Kaplan's determinism method is found in Sec. 3.4, the *correlation dimension* D_2 is given in Sec. 2.1.2, and the MODWPT in Sec. 3.1. The remaining measures are briefly described below.

5.3.1 *MODWT variance*

The wavelet variance is a useful alternative to conventional statistical techniques used to estimate the power spectral density and process variance. It can also be used to provide insight on the stationarity of a process. For power law processes such as colored random noise, the wavelet variance is a useful descriptor in that it can be used to estimate the spectrum exponent and to verify the process stationarity. The wavelet nomenclature used in this chapter is consistent with that found in Percival and Walden [73].

Let X_t with $t = 0, \ldots, N - 1$ be a discrete real-valued time series of length $N = 2^J$ $J \in \mathbb{N}$. The discrete wavelet transform maps X_t from the time domain to the time-scale domain. *Time* denotes a monotonic progression of events with the understanding that the units of t need not be temporal. For example, t may represent a spatial progression in units of length, area, or volume. *Scale* denotes the size of the window with which the wavelet coefficients are associated.

The wavelet coefficients W_k for $k = 0, \ldots, N-1$ are proportional to the difference between adjacent averages on a scale $\tau_j \equiv 2^{j-1}$ for $j = 1, \ldots, J$. Here, τ_j is a normalized scale. In practical observations, τ_j corresponds to the physical scale $\tau_j \Delta t$ where Δt is the sampling period.

There are exactly $N/2^j$ wavelet coefficients associated with scale τ_j, the collection of which is placed into a vector $\mathbf{W_j}$ for $j = 1, \ldots, J$. As an example, for $N = 16$

$(J = 4)$, the wavelet coefficients $\mathbf{W_1} = [W_0, \ldots, W_7]$ are associated with scale

$\tau = 1$, $\mathbf{W_2} = [W_8, \ldots, W_{11}]$ with scale 2, $\mathbf{W_3} = [W_{12}, W_{13}]$ with scale 4, and

$\mathbf{W_4} = W_{14}$ with scale 8. The last wavelet coefficient W_{N-1} is placed into a vector

$\mathbf{V_J} = \mathbf{V_4} = W_{15}$ and is by definition proportional to the mean of X_t. The collection

of vectors is concatenated into one vector $\mathbf{W} = [\mathbf{W_1} \ \ \mathbf{W_2} \ \ \ldots \ \ \mathbf{W_J} \ \ \mathbf{V_J}]$.

The discrete wavelet transform is energy preserving in that

$$||\mathbf{X}||^2 = ||\mathbf{W}||^2 = \sum_{j=1}^{J} ||\mathbf{W_j}||^2 + ||\mathbf{V_J}||^2 \qquad (5.2)$$

where $||\mathbf{X}||^2 = \sum_{t=0}^{N-1} X_t^2$. Since the sample variance of X_t is defined as

$$\hat{\sigma}_X^2 \equiv \frac{1}{N} \sum_{t=0}^{N-1} (X_t - \overline{X})^2 = \sum_{t=0}^{N-1} \frac{X_t^2}{N} - \overline{X}^2 = ||\mathbf{X}||^2/N - \overline{X}^2 \qquad (5.3)$$

and since for discrete Daubechies' wavelets $\overline{X} = \mathbf{V_J}/\sqrt{N}$, it follows from Eq. 5.2

and 5.3 that

$$\hat{\sigma}_X^2 = \frac{\sum_{j=1}^{J} ||\mathbf{W_j}||^2}{N} \qquad (5.4)$$

The term $||\mathbf{W_j}||^2/N$ represents the contribution to the sample variance of X_t at scale

τ_j. We define the wavelet power spectrum as

$$P_W(\tau_j) \equiv ||\mathbf{W_j}||^2/N. \qquad (5.5)$$

Just as the power spectral density decomposes the sample variance of X_t across fre-

quencies, the wavelet variance decomposes the sample variance of X_t across scale.

The wavelet variance is therefore an appealing approach to analyzing physical pro-

cesses which exhibit variations in different scales. The problem in using the wavelet

spectrum is that it is not (circularly) shift-invariant. A technique, known as the *Maximum Overlap Discrete Wavelet Variance (MODWT)* [73] can be used to overcome this obstacle at the price of incurring redundant operations in the algorithm. Still, its computational onus is the same as the widely used fast Fourier transform algorithm and is acceptable for most time-critical processing. The MODWT coefficients are defined as

$$\widetilde{W}_{j,t} \equiv \sum_{l=0}^{L_j-1} \widetilde{h}_{j,l} X_{t-l \bmod N}, \quad t = 0, \ldots, N-1 \tag{5.6}$$

with $\widetilde{h}_{j,l} \equiv h_{j,l}/2^{j/2}$ where $h_{j,l}$ denotes the wavelet filter associated with scale $\tau_j \equiv 2^{j-1}$ for $j \in \mathbb{N}$. The length of the wavelet filter at scale τ_j is given by $L_j = (2^j - 1)(L-1) + 1$ where L is number of filter taps at scale $\tau_1 = 1$. The MODWT can be used to obtain an unbiased estimate of the wavelet variance if care is taken to not include coefficients subject to circularity. The unbiased MODWT variance is defined as

$$\nu_X^2 \equiv \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} \widetilde{W}_{j,t}^2 \tag{5.7}$$

where $M_j \equiv N - L_j + 1$.

The MODWT can be used to form an analysis of variance as a function of time and scale or as a function of scale only. The scale only version is formed by averaging the wavelet variance over time at a given scale. Fig. 5.4 and Fig. 5.5 show an example of the time dependent and time independent unbiased MODWT wavelet variance,

respectively, for interspike intervals (ISI) extracted from a chaotic Lorenz sequence. The ISI coefficients are the differences in time between adjacent local extrema.
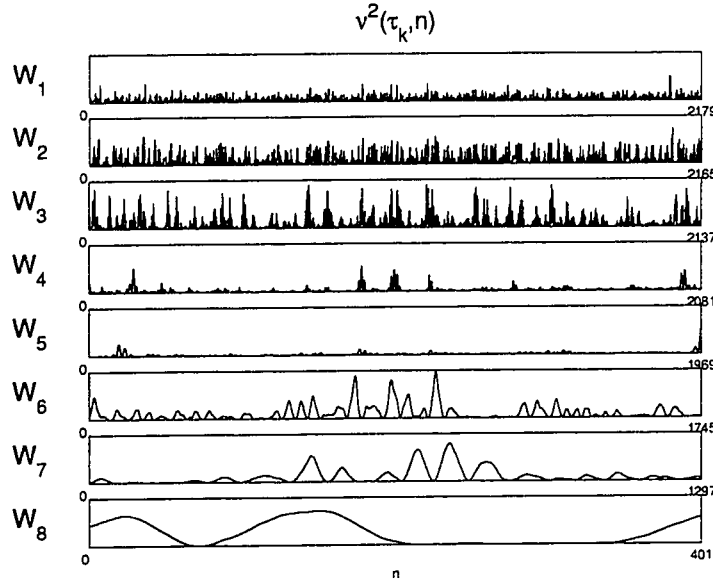
Figure 5.4: The unbiased wavelet variance as a function of time and scale for Lorenz ISI.

The wavelet variance can be used to estimate the exponent for a power law process, such as α in a stochastic $1/f^\alpha$ process (colored noise), by finding the best linear slope to fit the data in a $\nu^2(\tau_j)$ versus τ_j plot on a log-log scale [73]. For example, Fig. 5.6 shows the time history and time independent MODWT wavelet variance as a function of scale for a fractionally differenced process $\{Y_t\}$ with backwards difference order $d = 0.25$. The spectral density of $\{Y_t\}$ is a power law process with $S_Y(f) \sim f^{-2d}$. The slope of the wavelet variance plot is $\alpha \approx -1/2$, and since $\alpha = -2d$, $d \approx 1/4$ which we know is true!
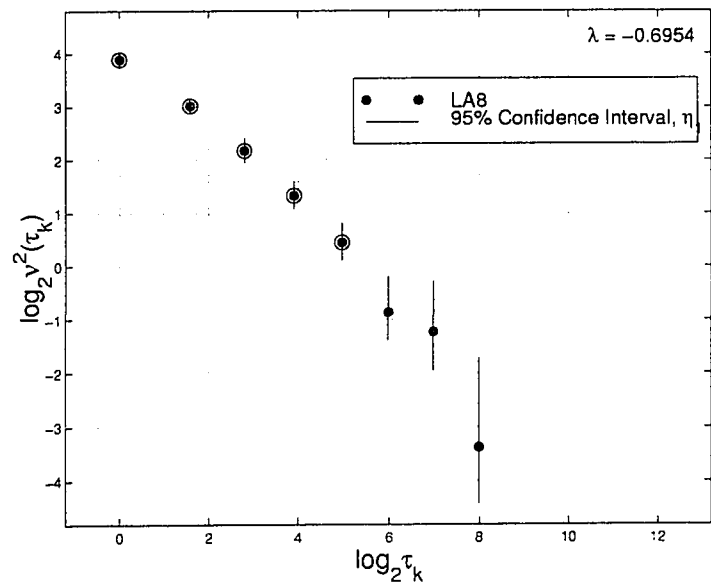
Figure 5.5: The unbiased wavelet variance as a function of scale only for Lorenz ISI. The circles surrounding the dots denote data used in the linear least squares estimation of the slope.
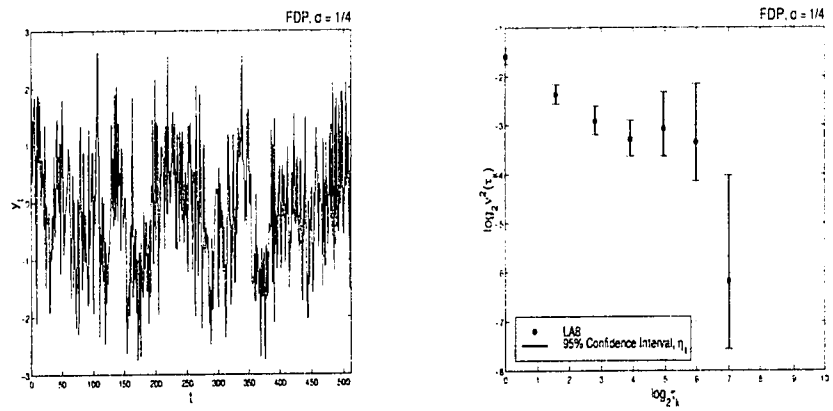
Figure 5.6: Realization of a FDP for $d = 1/4$ and the corresponding time independent unbiased MODWT variance.

5.3.2 Approximate entropy

Approximate entropy or $ApEn$ is a measure of the change in entropy between points embedded in $D_e = 2$ and $D_e = 3$. Pincus [75] developed a simple algorithm for discrete sequences. Briefly, the algorithm works by calculating the fraction $C_{i,r}$ of all pairs of points which fall within some radius r for each point pair i in $D_e = 2$ for $i = 1, \ldots, N_D$ where N_D is the number of unique pairs in the phase plane. Define Φ_2 to be

$$\Phi_{2,r} = \log\left(\frac{1}{N_D} \sum_{i=1}^{N_D} C_{i,r}\right). \tag{5.8}$$

Now repeat these steps in $D_e = 3$ to form $\Phi_{3,r}$. Then, ApEn is defined by the difference

$$ApEn \equiv \Phi_{2,r} - \Phi_{3,r}. \tag{5.9}$$

Approximate entropy is claimed to be a sensitive measure of complexity in cardio-vascular dynamics as it was shown in [38] that the ApEn of heart rate decreased with age for healthy subjects.

5.3.3 Bridge-detrended scaled windowed variance

The BDSWV method is used to estimate a statistical self-similarity exponent known as the Hurst coefficient H for fractional Brownian motion or *fBm*. Fractional Brownian motion has a definitive relationship with fractional Gaussian noise or *fGn*. Let $\{X_t\}$ be a sequence of random variables. $\{X_t\}$ is a zero mean fGn if it meets the following conditions:

1. any finite subset of $\{X_t\}$ obeys a multivariate Gaussian distribution

2. $E\{X_t\} = 0$ for all t

3. $s_{X,\tau} \equiv \text{cov}\{X_t, X_{t+\tau}\} = \frac{\sigma_X^2}{2}(|\tau+1|^{2H} - 2|\tau|^{2H} + |\tau-1|^{2H})$, $\quad \tau = 0, \pm 1, \pm 2$

where $E\{X_t\}$ is the expected value of any member in $\{X_t\}$, $s_{X,\tau}$ is the autocovariance sequence, H is the Hurst coefficient for $0 < H < 1$, and $\sigma_X^2 \equiv \text{var}\{X_t\} = s_{X,0}$ is the process variance assumed to be bounded [72]. Let $B_H(t)$ be a zero mean Gaussian process with variance $\sigma_X^2 t^{2H}$ such that $B_H(0) \equiv 0$. Then a discrete fractional Brownian motion sequence B_t can be obtained by discretely summing over an fGn realization

$$B_t \equiv B_H(t) = \sum_{u=0}^{t-1} X_t, \quad t = 1, 2, \dots \qquad (5.10)$$

The scaled windowed variance algorithm works by aggregating the data points into non-overlapping windows of dyadic length. Each window size is associated

with a scale $\tau_j \equiv 2^j$ for $j = 1, \ldots, \log_2 N$ where N is the length of the series (for convenience assume N is a power of 2). Define $S(\tau_j)$ be the scaled windowed variance at scale τ_j

$$S(\tau_j) = \frac{1}{M} \sum_{m=0}^{M-1} \hat{\sigma}_m(\tau_j) \tag{5.11}$$

where $M = N/\tau_j$ is the number of windows and $\hat{\sigma}_m(\tau_j)$ is the standard deviation of the mth window at scale τ_j. The Hurst coefficient is then estimated by fitting the $\log S(\tau_j)$ versus $\log \tau_j$ curve with a linear least squares regression. The BDSWV works in the same way, only for each window a line connecting the endpoints of the sequence is subtracted to "detrend" the data. The BDSWV works well in estimating H for fBm but fails for estimating H for fGn.

5.3.4 Dispersion analysis

Dispersion analysis, also known as standard deviation analysis, explores the effect of averaging over non-overlapping windows of dyadic length with regard to the standard deviation of the data [8, 9]. Dispersion analysis is designed for the analysis of fGn and is not suitable for analyzing fractal Brownian motion.

5.4 Results

Due to the quantity of collected data and for ease of comparison, the results will be shown in groups preceded by a summary of observations regarding the data.

5.4.1   Healthy Patients

The RR intervals were calculated using methods explained in Sec. 5.1 for 5 patients.
A sample of the RR histories is shown in Fig. 5.7. The ordinate is in units of
milliseconds while the abscissa is in units of beats. On average, the patients maintain
a resting heart rate of approximately one beat every 7/10 of a second or about 86
beats/min.
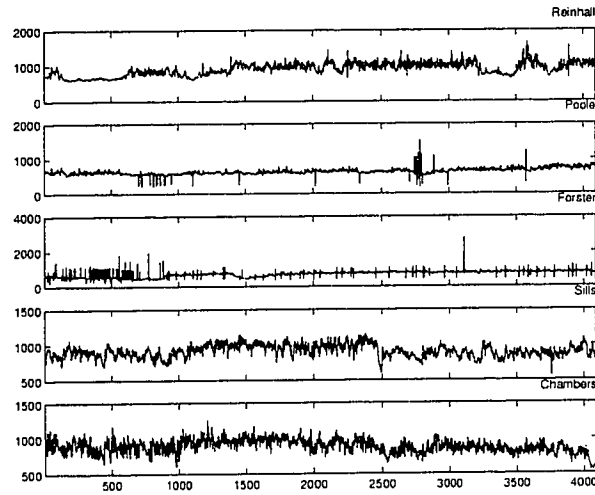
Figure 5.7: Samples of RR intervals for healthy patients.

Correlation Dimension (D₂): Figure 5.8 shows the correlation dimension re-
sults for the healthy patients. For each case, $N = 8192, D_e = 7, \tau = 1, \gamma = 0$. Fifty
surrogate sets were also run using Theiler's AAFT algorithm with a Hanning win-

dow. A representative surrogate result is shown in the lower right corner of Fig. 5.8. The D_2 curves do not have a divergent behavior indicative of random noise. However, there is no significant scaling region (plateau) for any of the patients, making interpretation difficult. So while the data does not appear to be random, there also is no apparent fractal structure.

KDM: The KDM parameter set was $N = 2^{14}, \tau = 1, \gamma = 0, \kappa = 2, \Delta r = 0.1$, and $D_e = 1, \ldots, 7$ for all tests. A sample of the KDM results is shown in Fig. 5.9. There is a distinct separation between the original RR and surrogate data in the lower dimensions, indicating that the RR intervals (appear to be) are deterministic. Figure 5.10 shows the violation results for all the patients. Without exception, the KDM violation results indicate that the RR intervals are deterministic in the healthy patients.

5.4.2 Sick: ECG Holter Recordings

In this section, the results for 30 patients with a history of heart disease are presented.

Correlation Dimension (D_2): For each test, $N = 16,384, D_e = 7, \tau = 1, \gamma = 0$. Figure 5.11 and Fig. 5.12 show the D_2 results. The results for patients $6, 7, 8, 9, 13, 16, 20$, and 27 exhibits a (semi)convergent plateau region indicates some fractal structure is present in the RR intervals for these patients. The remaining patients have the
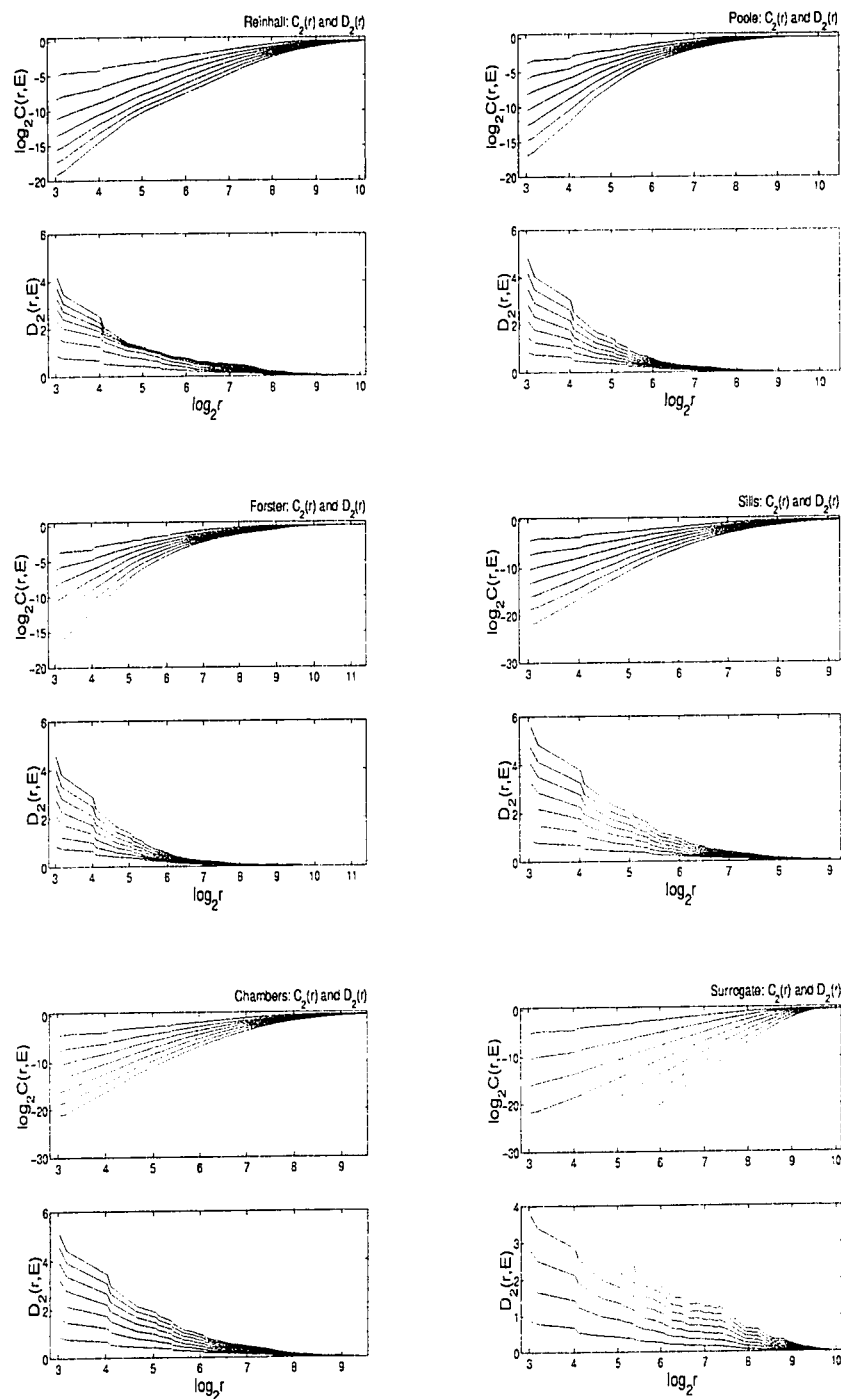
Figure 5.8: Correlation dimension results for 5 physically healthy patients.
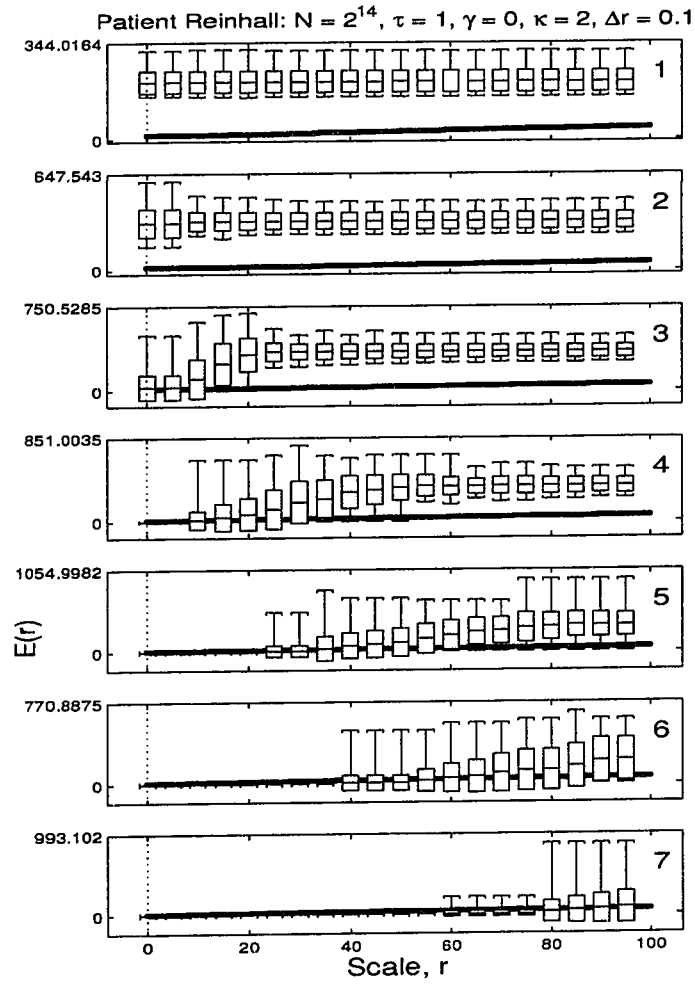
Figure 5.9: KDM results for a healthy patient: RR intervals.
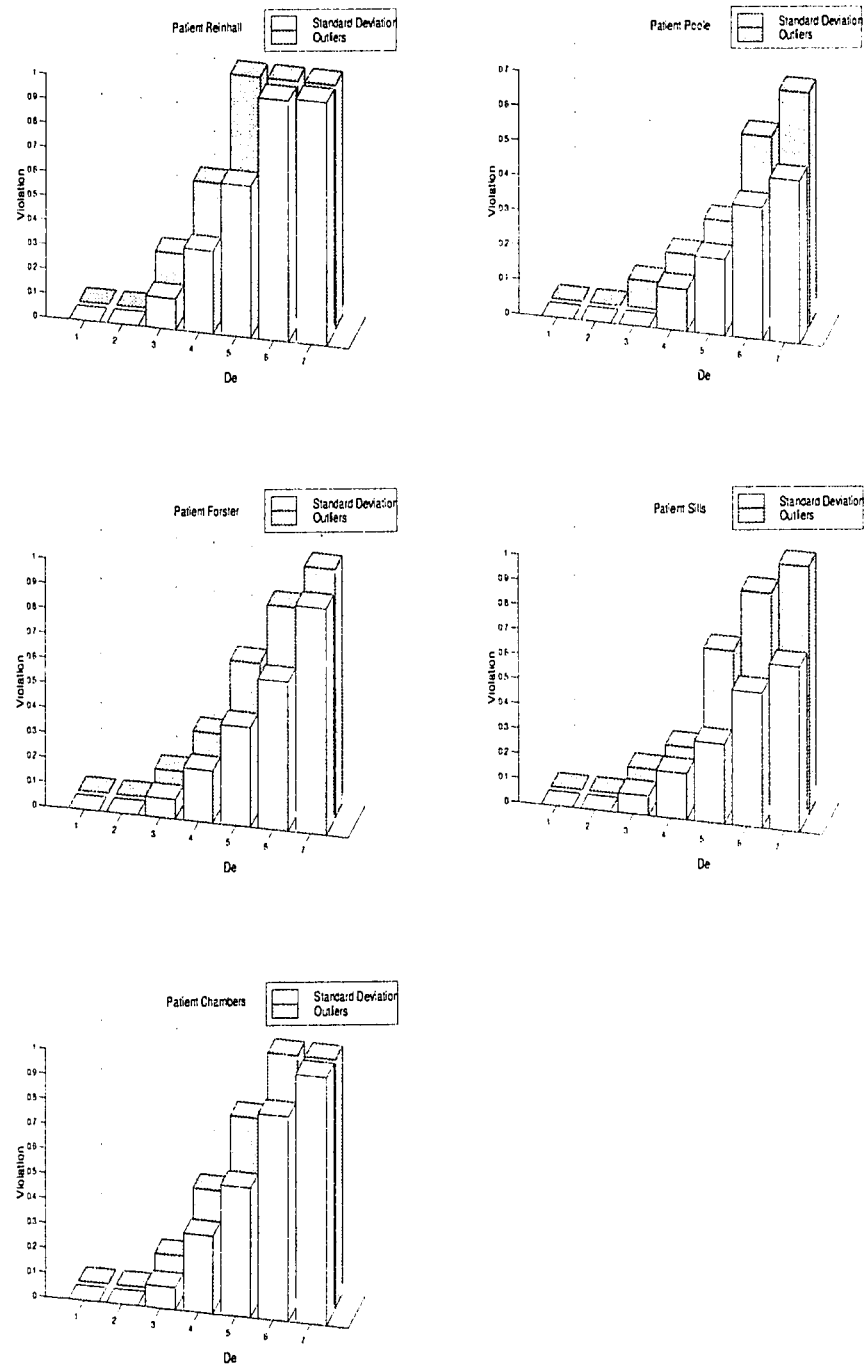
Figure 5.10: Violation summaries for 5 physically healthy patients.

same type of structure as exhibited for the healthy cases (Fig. 5.8).

MODWT variance: Figure 5.13 summarizes the time independent unbiased MODWT wavelet variance for all patients. Six patients (1, 2, 6, 10, 14, 29) of the thirty exhibit a U-shaped patten in the variance suggesting a transitional state from a stationary process (negative slope) at small scales to a nonstationary behavior (positive slope) at large scales, or vice versa. Ten patients (3, 4, 5, 8, 13, 16, 19, 21, 26, 27) exhibit a fairly linear positive upward trend in the data, indicating a nonstationary process with power law scaling. Only four patients (7, 9, 11, 18) exhibit a linear continuous downward trend, indicative of stationary behavior with power law scaling. The remaining eleven patients do not seem to exhibit any scaling structure at all. As a caveat to this technique, a pathologically straight line in the wavelet variance plots can be a result of the sidelobe behavior of the wavelet filter transfer functions and not the process itself. However, no such behavior exists in the results shown in Fig. 5.13.

5.4.3 Sick ICD Patients

Ventricular fibrillation (VF) is defined as a spasmodic fatal cardiac rhythm [21]. During VF, the heart pumps no blood and the blood pressure falls to zero within seconds, resulting in clinical death. The VF victim passes out within seconds due to lack of oxygen to the brain and is essentially rendered helpless. If not revived
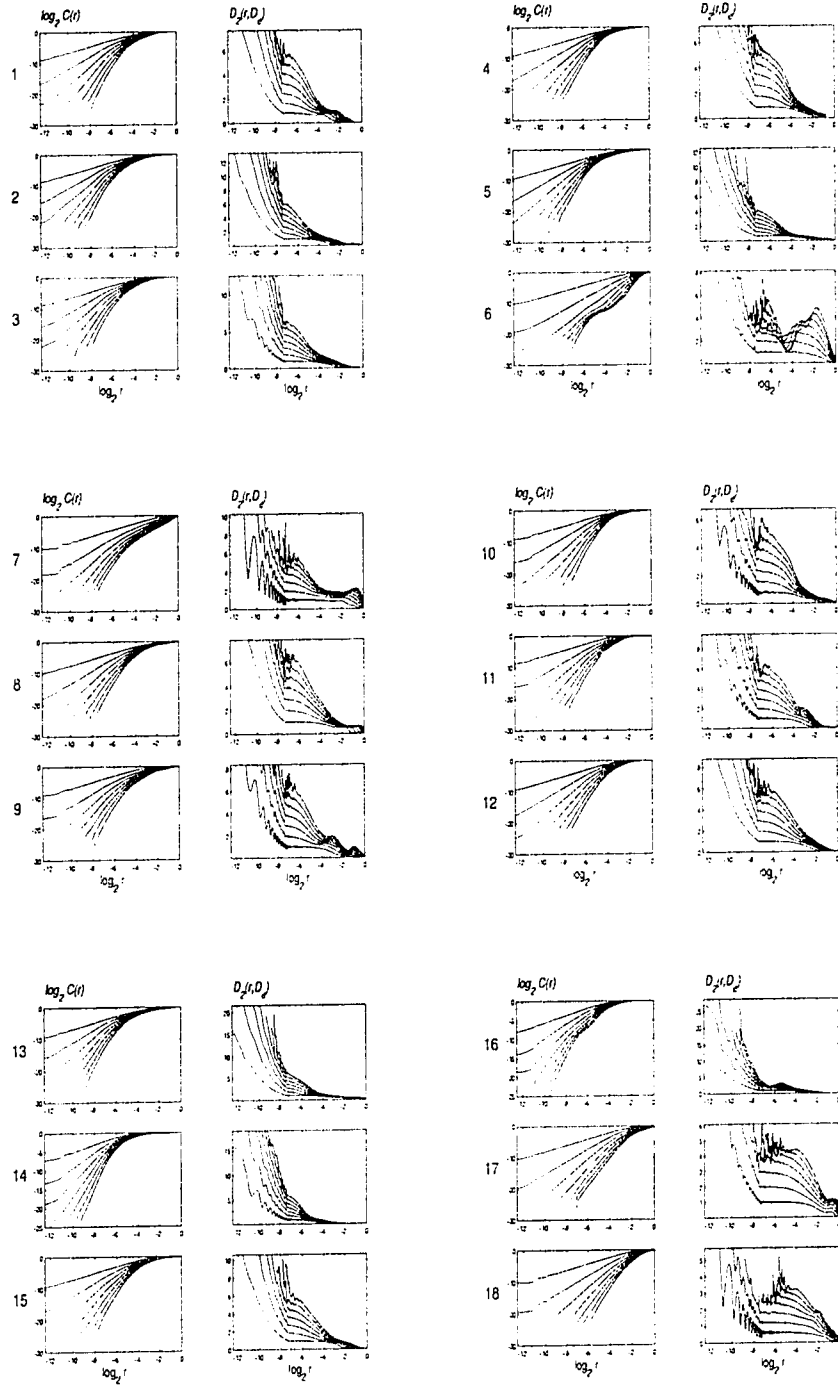
Figure 5.11: Correlation dimension results for sick Holter patients 1..... 12.
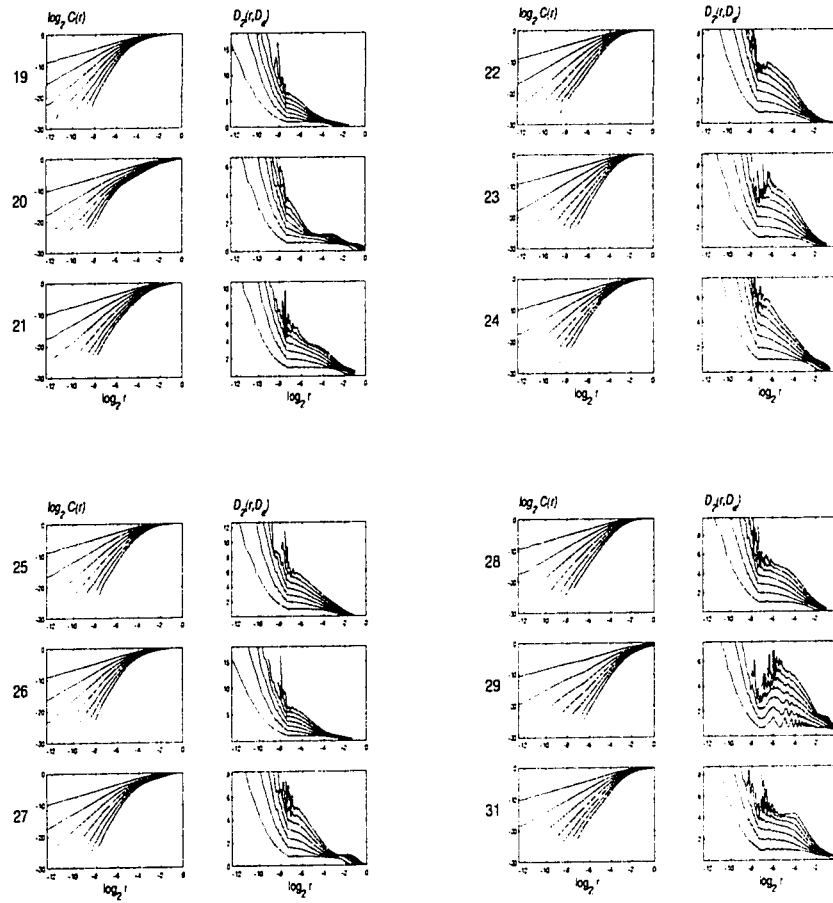
Figure 5.12: Correlation dimension results for sick Holter patients 19,.....31.
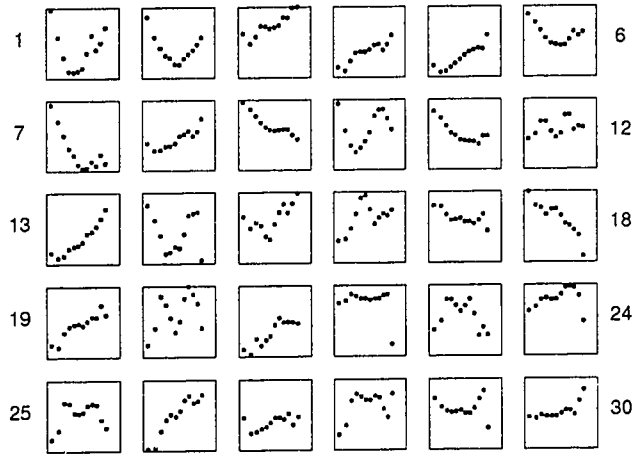
Figure 5.13: Time independent unbiased wavelet variance estimates for splined RR intervals from 30 patients. The patients are monotonically ordered in rows from left to right and from top to bottom. The patient numbers for the outer columns are noted for clarity. The abscissa of each plot is $\log_2(\tau_j)$ where $\tau_j \equiv 2^j$ is the scale while the ordinates are $\log_2(\nu^2(\tau_j))$ where ν^2 is the wavelet variance.

within minutes, biological death will occur. The patients studied in this section are unique in that they all have a recorded VF and/or VT incident. Some statistics will suffice in supporting the importance of this data: over 350,000 sudden cardiac deaths occur every year in America with 75% occuring at home. VF occurs 65% of the time while in ~ 3% of the cases, VT ensues. Finally, 20% of VF patients have *no* history of cardiac disease [21].

Consider for example the time history of ICD patient 3 shown in Fig. 5.14. A normal rhythm disrupted by two apparent PVC events at times around 7 and 15 seconds into the record, puts the patient into a VF mode. VF continues for approximately 10 seconds when the ICD kicks in and shocks the heart in an attempt to return it to a normal sinus rhythm.

VF is not the only means to a heart attack. Another fatal rhythm is ventricular tachycardia which is a sustained rapid heart rate. Consider for example the time history of ICD patient 2 shown in Fig. 5.14. During the first 12 seconds the heart appears to be in normal rhythm with the exception of a PVC event around 4 seconds. Then another PVC event occurs around 15 seconds into the record and the patient is cast into VT. The patient's heart rate has gone from approximately 90 beats/min to 240 beats/min and at that rate is fatal. As with the VF incident described earlier, the ICD unit applies (fires) a relatively high voltage across the heart in an attempt to restore its natural rhythm.

Presented below is an ensemble of results for the RR intervals leading up to a
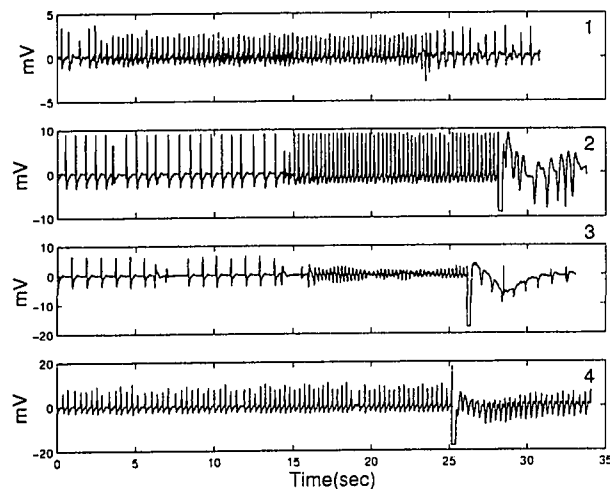
Figure 5.14: Sample ICD time histories for patients 1,....,4. The histories exhibit (1) AF or VT (2) VT (3) VF (4) AF where AF is atrial fibrillation, VF is ventricular fibrillation, and VT is ventricular tachycardia

ICD shock due to VT/VF incidents. As a note, a missing subplot in the plot frames denotes a lack of data for that measure to be used.

Periodogram, BDSWV, Dispersion Analysis: A periodogram is essentially a decomposition of the variance with respect to frequency. The envelope of a spectrum can be used to reveal any spectral scaling of the data. BDSWV and Disp are also used to explore power law scaling in the in terms with regard to the standard deviation (square root of the variance). Because of the similarity between the measures and for ease of comparison, the results are presented in a picture book format. Figure 5.15 through Fig. 5.17 show the periodogram, Fig. 5.18 through Fig. 5.20 the BDSWV.

and Fig. 5.21 through Fig. 5.23 the Disp results for the RR intervals for each of the 18 ICD patients. For each patient there are three RR records, the first of which is classified as a "normal" rhythm for which no shock occurred, while the remaining two are RR records leading up to VF and/or VT. For reference, the RR records are plotted above the measures.

The bulk of the results indicate that there is some scaling of variance within the data. With few exceptions, the BDSWV results show a negative slope (equivalent to a negative estimate of the Hurst coefficient) for the ICD RR intervals. This, however, is problematic in that the Hurst coefficient is defined on $0 < H < 1$. A Hurst coefficient close to zero may indicate that the RR intervals are an anticorrelated noise and not a fBm. According to G. Raymond at the National Simulation Resource [9], in these cases the slopes tend to be small, and indicates that the SWV is an inappropriate means of estimating H for a fGn. Dispersion analysis is the recommended measure in these cases. A box plot of the results summarizes the BDSWV and Disp data for all of the ICD RR data in Fig. 5.24. The outliers show the maximum range spanned by the measures while the box represents one standard deviation above and below the mean. Each marker summarizes the results for all patients for a particular RR record index, for example RR 1 denotes the group of normal RR intervals. A comparison between the markers indicates that these measures do not appear to be appropriate for discerning healthy from sick with regard to the RR intervals. This may in fact have more to do with the inability of RR data to sufficiently character-

ize the health of an individual since heart rate variability in general is subject to so many external factors. For example, without complimentary data, one cannot tell whether an increase in heart rate is due to some internal change in the (cardiac) system or whether the increase is due to an external stimulus from the brain. The difference is that changes in the system itself may be an indication of the health of the heart whereas stimulus from the brain to avoid being hit by an oncoming car, for example, has nothing to do with how healthy the heart is. More data is needed to address these issues.

With respect to these observations, it may be more beneficial to characterize the data with respect to time and scale as opposed to scale alone, i.e. we must employ a technique capable of handling data where characteristics evolve with time.

Optimal MODWPT Basis: The MODWPT was shown in Chapter 3 to be an effective means of characterizing nonstationary data with regard to its time-frequency or time-scale content. We seek an optimal wavelet packet node set to describe the data. To this end, the MODWPT was performed on every ICD RR record using Daubechies' LA8 filters in a level $J = 6$ decomposition. A MODWPT Shannon entropy minimization was then performed to create a best basis for data. The wavelet packet nodes corresponding to the MODWPT best basis were stored in a matrix, and a subsequent singular value decomposition was used to eliminate the weaker nodes of the set by means of a singular value thresholding. The results
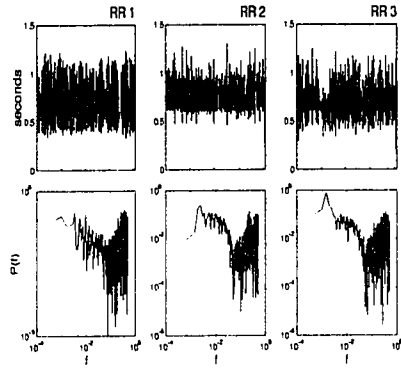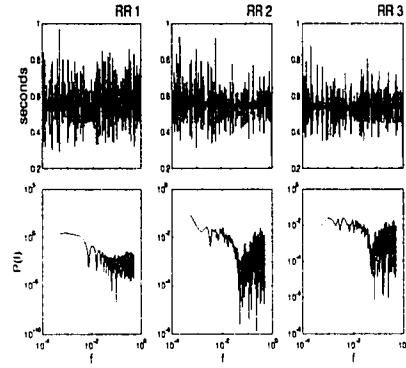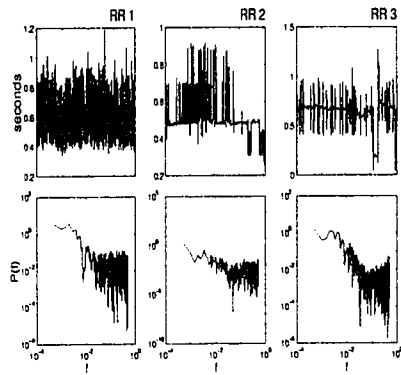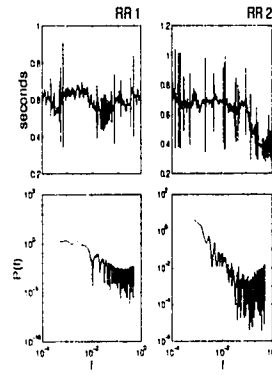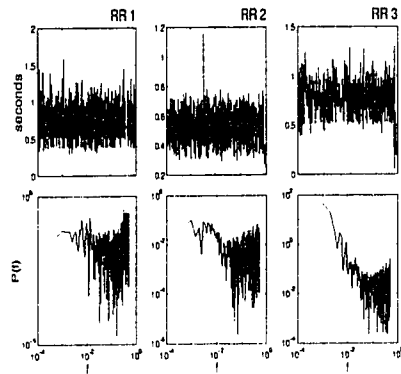
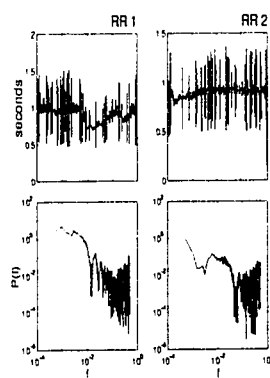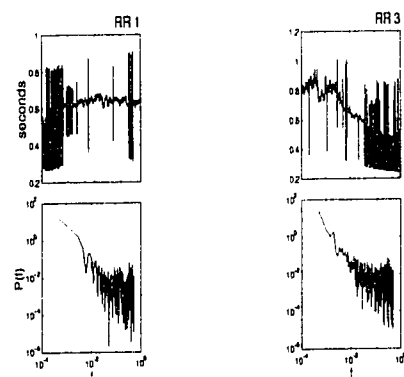Figure 5.15: Periodogram for ICD patients 1,... ,6.

Patient 01122451: Periodogram
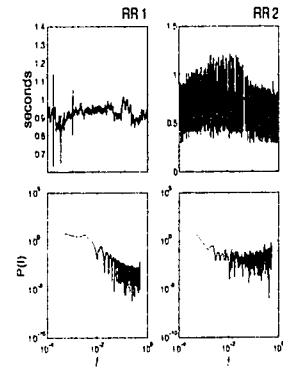
Patient 13110738: Periodogram
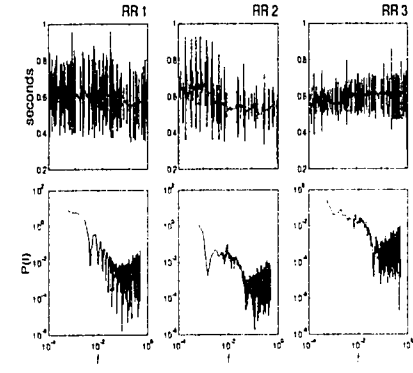
Patient 27110910: Periodogram
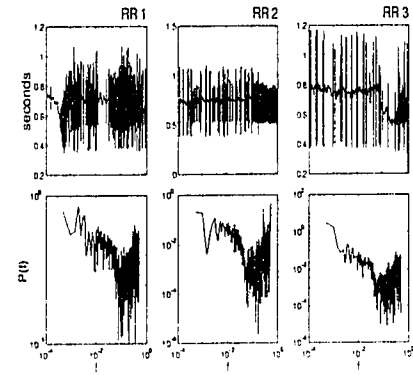
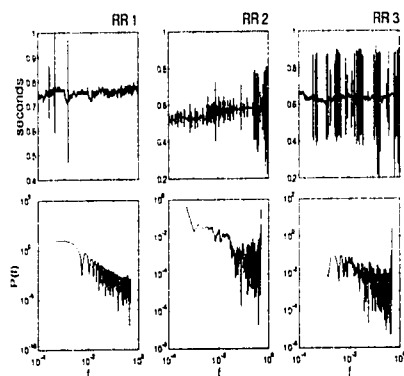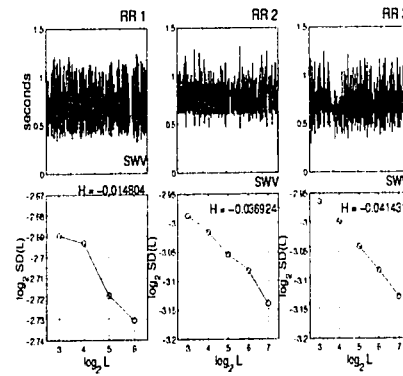Patient 01130958: Periodogram
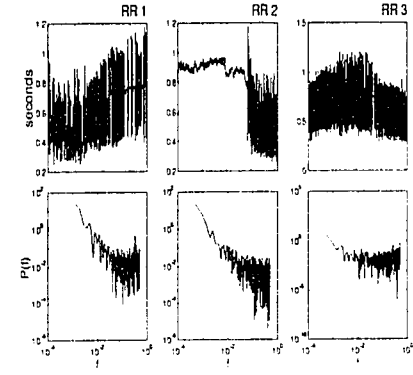
Patient 14144702: Periodogram
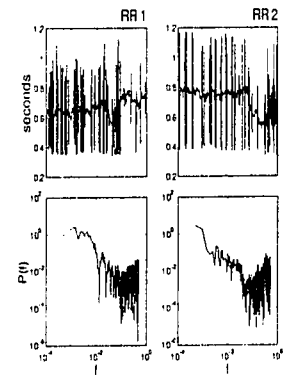
Patient 27162548: Periodogram
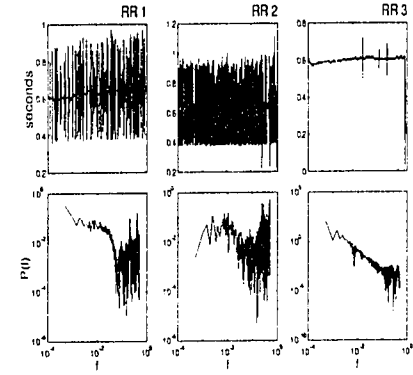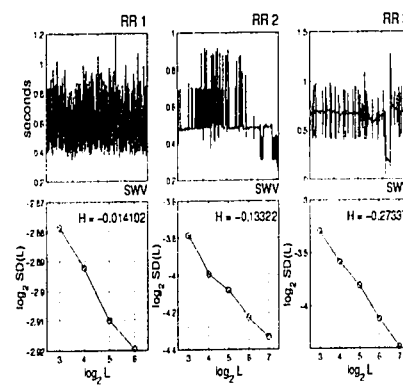
Figure 5.16: Periodogram for ICD patients 7, ... , 12.
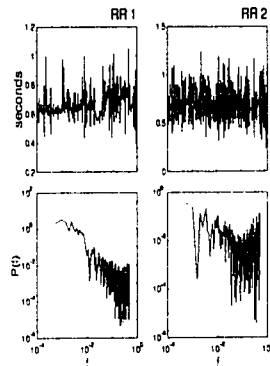
218
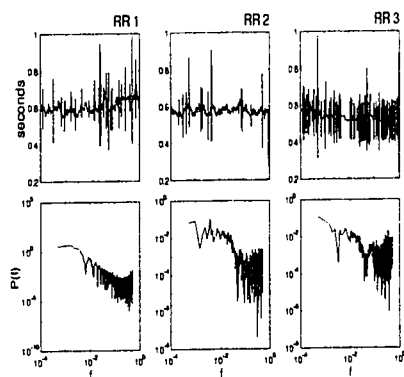
Figure 5.17: Periodogram for ICD patients 13,.....18.
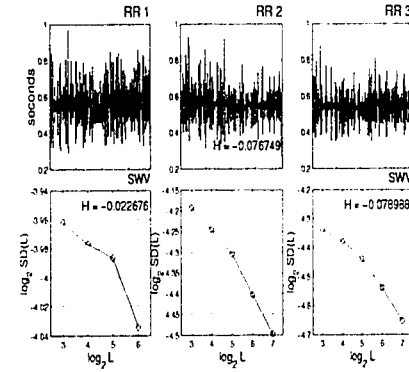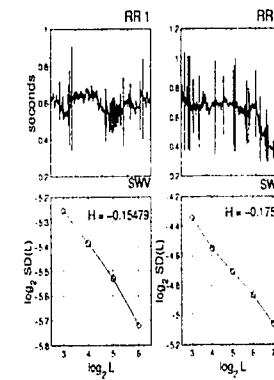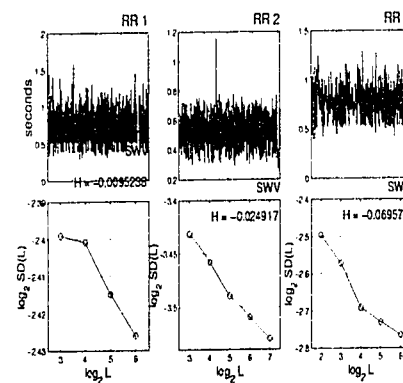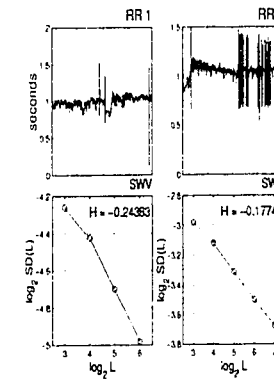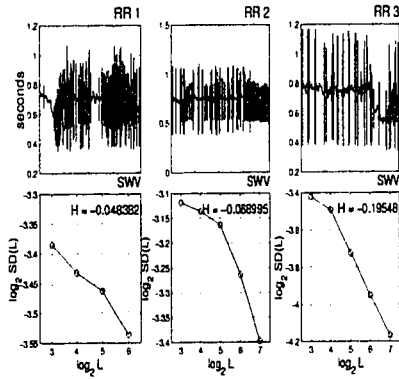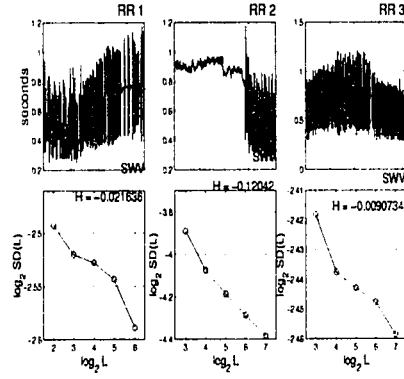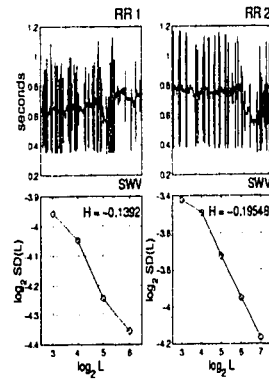
Figure 5.18: BDSWV for ICD patients 1, ... , 6.
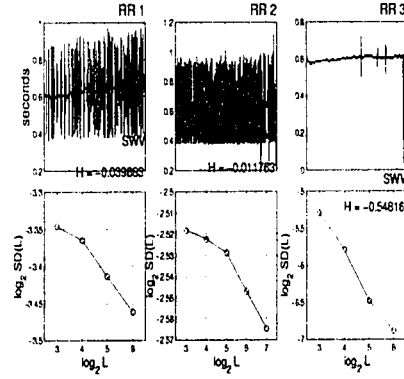
Figure 5.19: BDSWV for ICD patients 7, ... , 12.
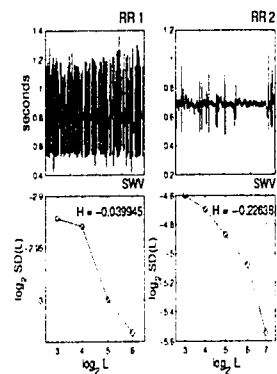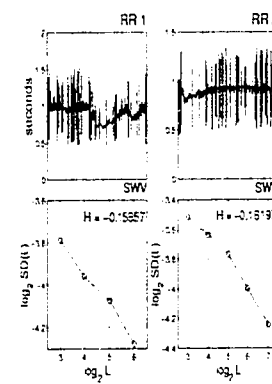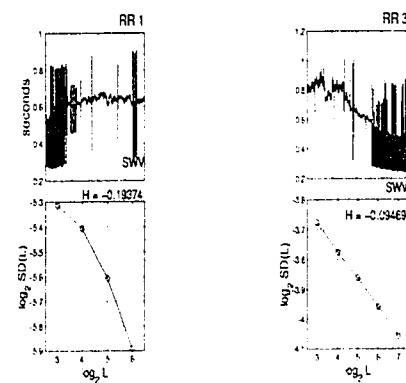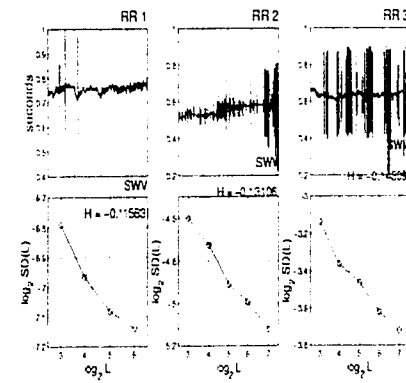
Figure 5.20: BDSWV for ICD patients 13,....,18.
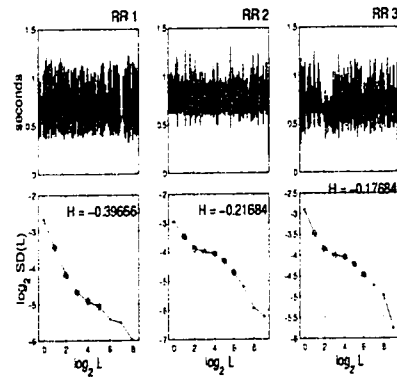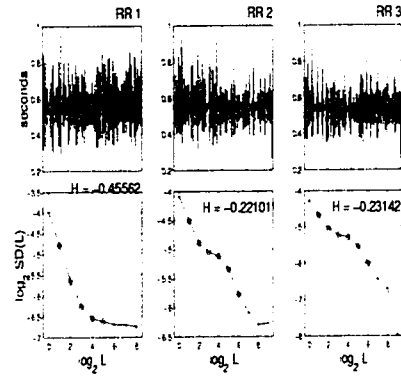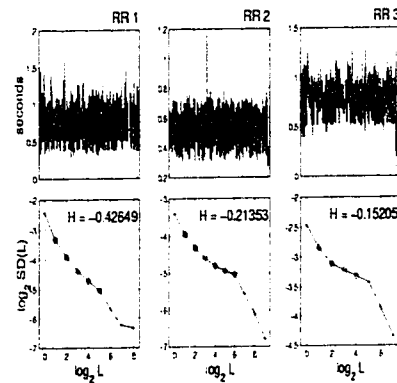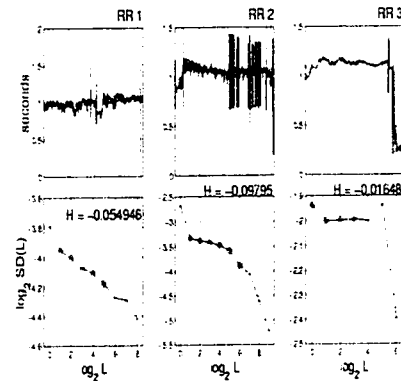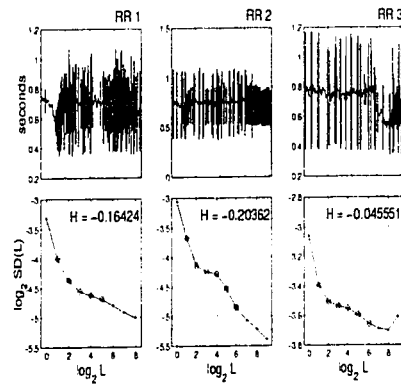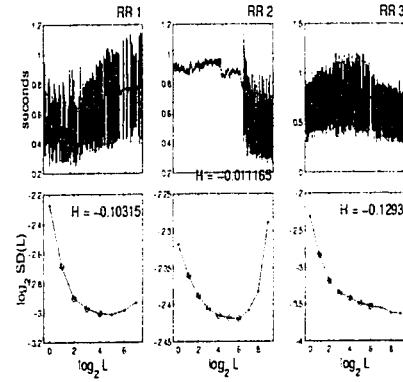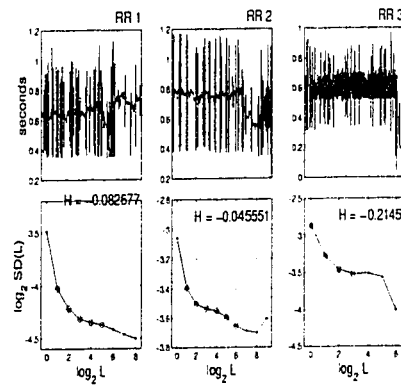
Figure 5.21: Dispersion analysis for ICD patients 1.....6.
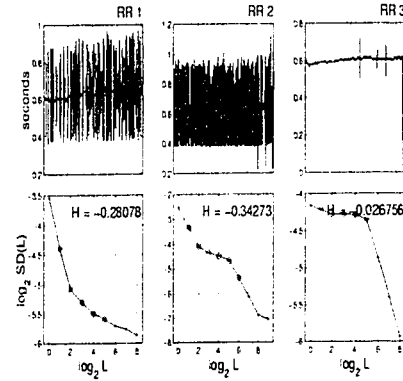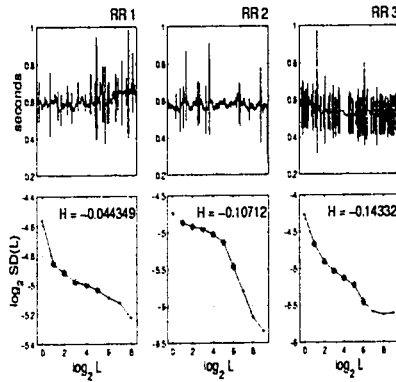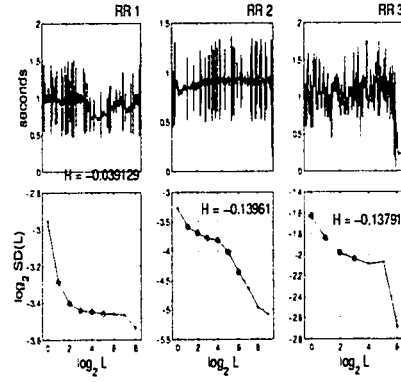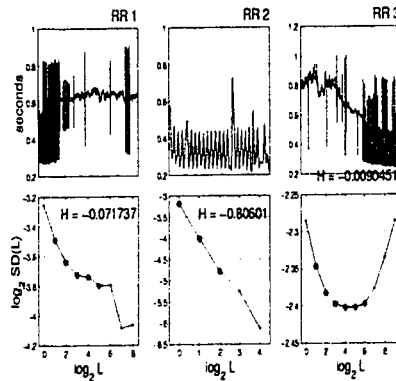
Figure 5.22: Dispersion analysis for ICD patients 7..... 12.

224
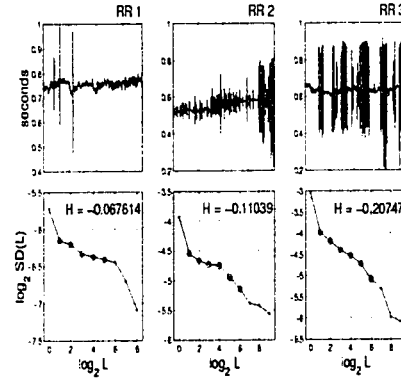
Figure 5.23: Dispersion analysis for ICD patients 13,.... 18.
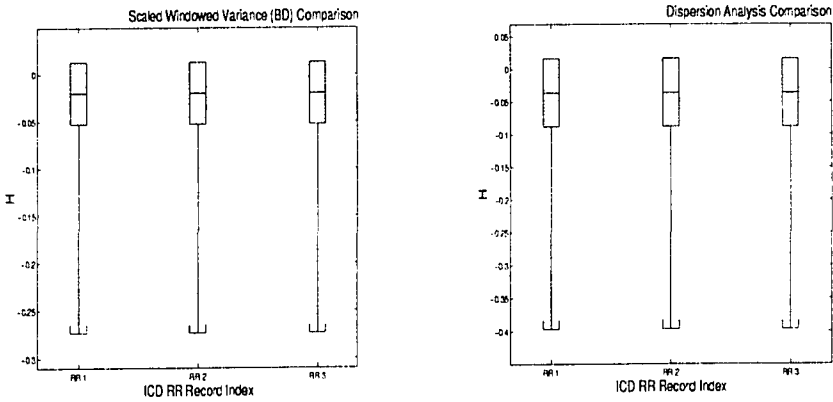
Figure 5.24: Box plot of all ICD results for the BDSWV (left) and Disp (right).

are shown in Fig. 5.25 through Fig. 5.27. Beside each node is its corresponding

frequency range in normalized frequency units $f \in [0, 1/2]$.   A summary of the

MODWPT results is shown in Fig. 5.28 and Fig. 5.29 in the form of horizontal bar

plots. The span of each bar is commensurate with the normalized frequency range of

the corresponding MODWPT node and the shading of the box indicates the nodal

energy level with the darkest boxes being the most energetic. Such a plot provides

an excellent means to compare the best MODWPT bases between the patients. The

symbol (W) appearing next to some of the plots indicates that the best basis is

a DWT. If the boxes were uniformly distributed in frequency, it would mean that

the best basis is akin to the decomposition of a STFT. For arbitrary sequences, the

MODWPT provides the flexibility to find the best basis amongst a wide variety of
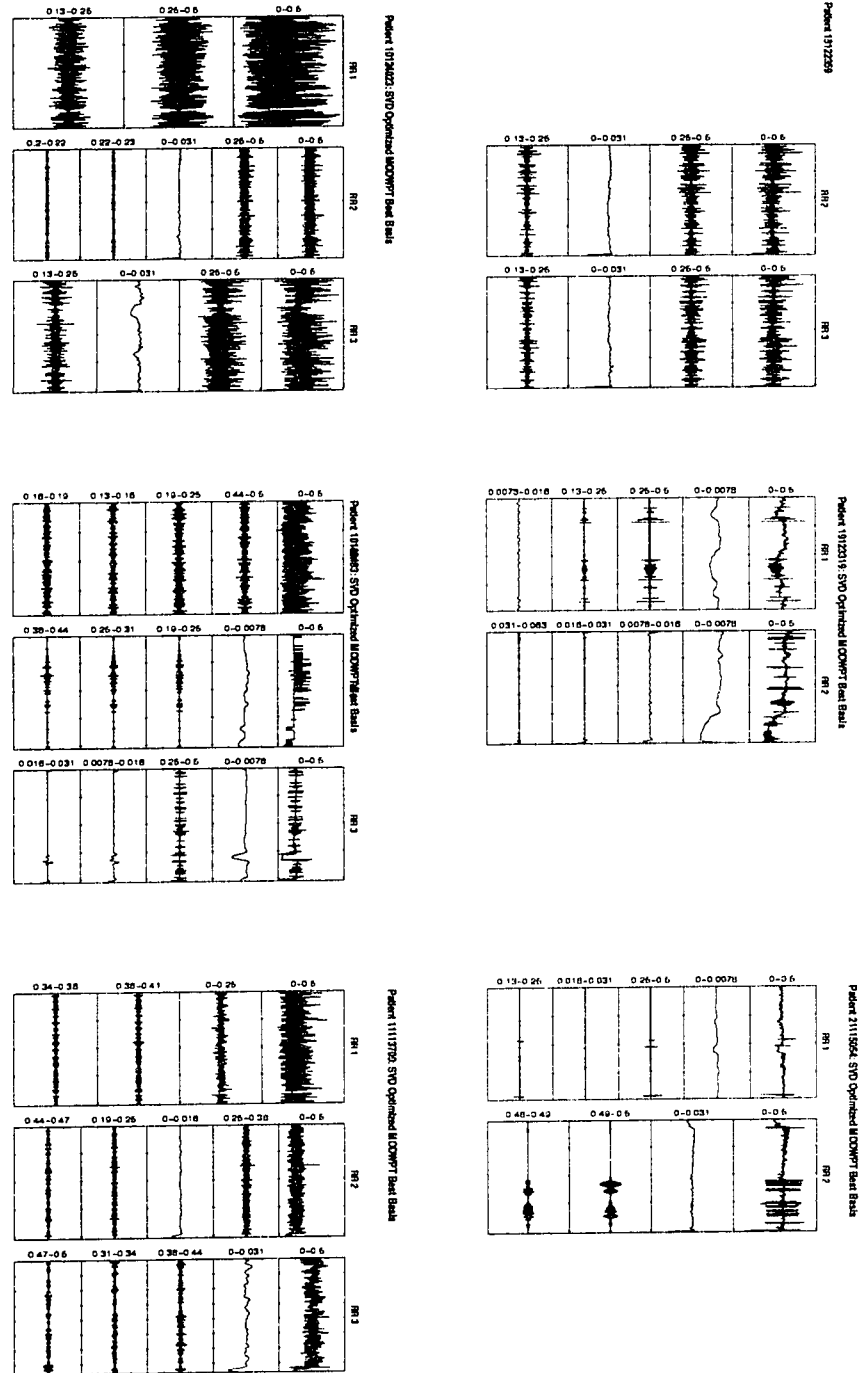
Figure 5.25: Optimal MODWPT for ICD patients 1, . . . , 6.

Figure 5.26: Optimal MODWPT for ICD patients 7, ... , 12.
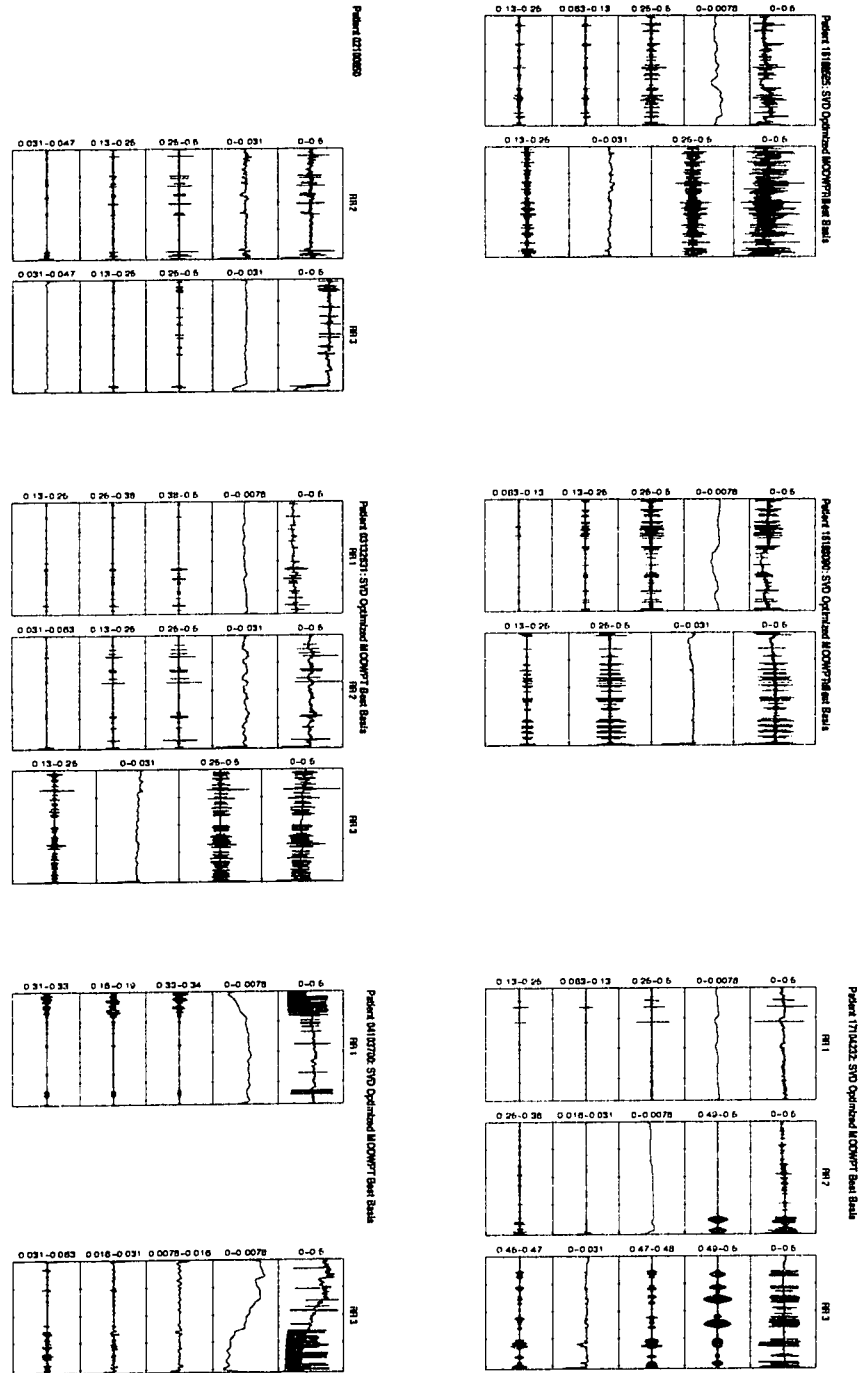
Figure 5.27: Optimal MODWPT for ICD patients 13, ... , 18.

transforms including the DWT and the DWPT (which is quite close to the STFT with regard to its partitioning of the normalized frequency interval $f \in [0, 1/2)$).

A characterization such as this could be used to help categorize disease states or an individuals health. This type of analysis is referred to in the literature as *transient signal classification*. To date, however, this type of analysis has never been performed on ICD RR interval data. The energy of the wavelet packet coefficients are a measure of the localized variance in time and could also be used to quantify changes in the character of a sequence as time progresses.
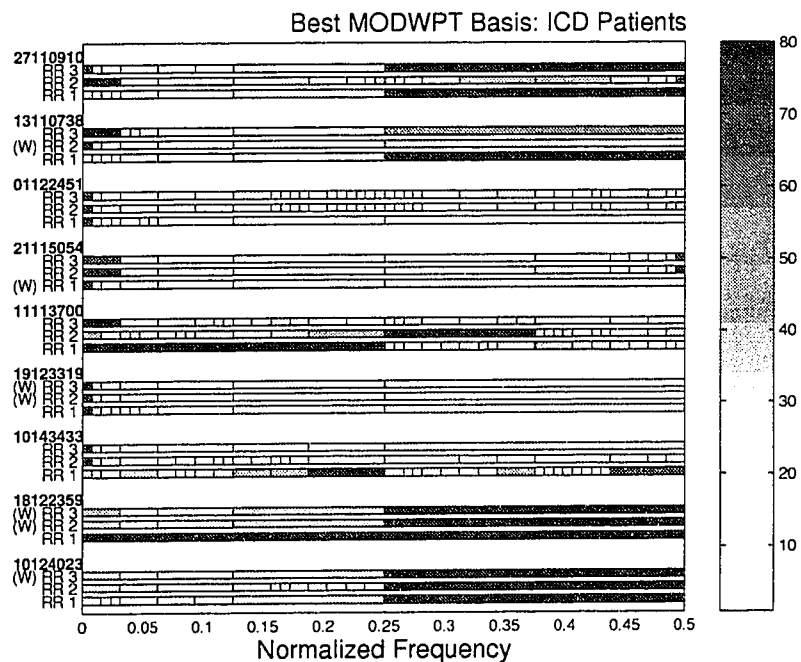
Figure 5.28: Energy distribution of optimized MODWPT nodes for ICD patients 1, . . . , 9.
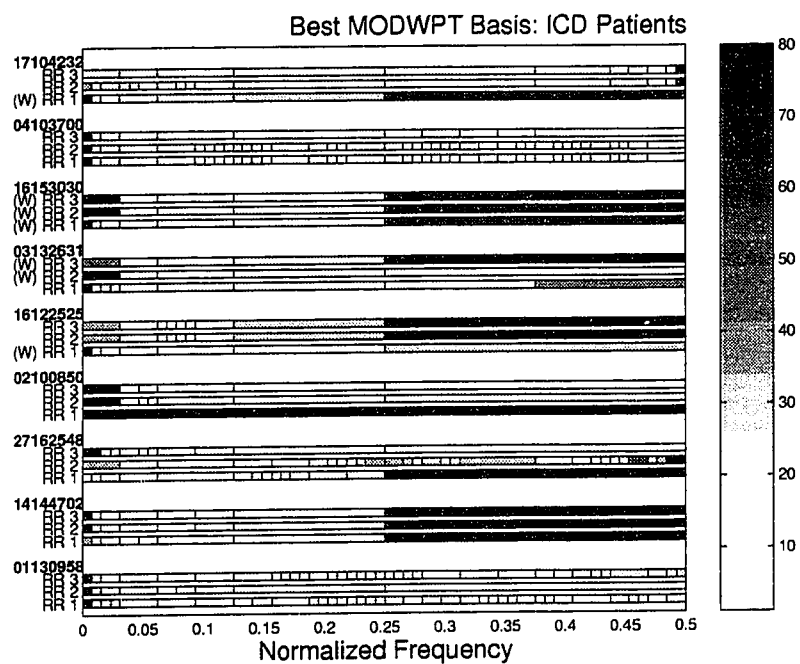
Figure 5.29: Energy distribution of optimized MODWPT nodes for ICD patients 10, ... 18.

Figure 5.30 shows an example of how the MODWPT nodal energies could be used to classify differences between RR records. An optimal nodal energy is formed such that $E_n^{(j)}$ is the energy of the strongest node in the SVD-optimized MODWPT set for patient $j = 1, \ldots, N_p$ where N_p is the number of ICD patients and $n = 1, 2, 3$ is the RR record index. Define $\beta_n^{(j)} = E_n^{(j)}/E_1^{(j)}$ for $n = 1, 2, 3$. By definition, $\beta_1^{(j)} = 1.0$ and $\beta_2^{(j)}, \beta_3^{(j)}$ are the optimized nodal energies of the VT/VF records normalized by that of the "normal" RR record. If β strays from unity it means that the optimized energy for the VT/VF record is different from the corresponding normal record. Figure 5.30 shows the results of this energy ratio test. The ratio test is produced by plotting $\beta_2^{(j)}$ versus $\beta_3^{(j)}$ (hollow circles) and comparing it to the normal energy point located at $(1.0, 1.0)$ (filled circle) in the plane. Each marker in the plot represents the energy ratio for one patient in the analysis. If a pattern of markers is clustered away from $(1.0, 1.0)$, it signifies a distinct difference between sick and healthy RR data with respect to an optimized measure of energy. Of course, a relevant criticism of this technique would be that other normal records may also shift from a comparative normal in the energy ratio plot and one must take into account sampling variability. Unfortunately, due to lack of data at the present time, such an assessment cannot be verified. The hope, of course, is that the data associated with sick records will venture far from the healthy ones. Alternatively, this method could be used to (feasibly) differentiate between different types of arrythmia, yielding an optimized energy based fingerprint of the data. The energy ratios are essentially

coordinates which form trajectory in a (cardiac) phase space. One could imagine

a scenario where a collection of points in the phase space could represent a given

aspect of the data, such as a certain arrythmia for example. The impact of such a

technique lies solely in the value of the coordinates used to develop the phase space.

For the heart, RR intervals may not be enough and other more relevant measures

may need to be added. The work shown here is thus for illustrative purposes only,

and is meant to show the potential of the technique rather than drawing critical

conclusions concerning the validity of using RR intervals as a measure of cardiac

health.
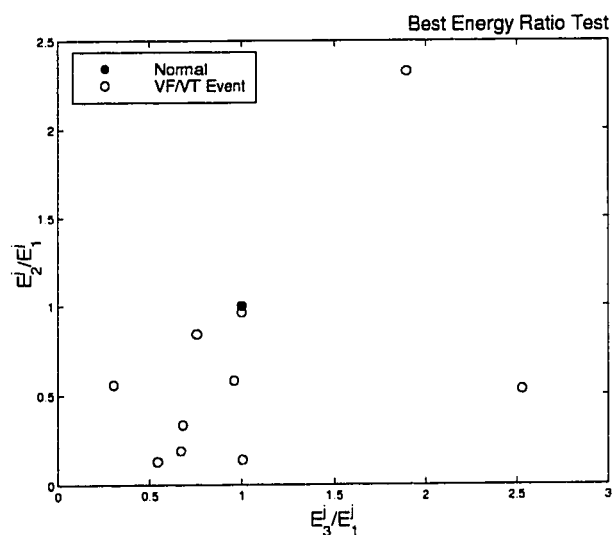
Figure 5.30: Best energy ratio test for all ICD data: a comparison between normal and VF/VT events.

Approximate entropy: Finally, ApEn was calculated on all the ICD RR records (see Fig. 5.31 to Fig. 5.33) using a 1000 point sliding window with a window shift of 10 points. In short, there is nothing in ApEn findings which suggest that it can be used as a predictive tool for the onset of VT/VF based solely on the variation of RR intervals. No definitive patterns emerge which would indicate otherwise. However, more data is needed to reach a definitive conclusion on this matter.

5.5 Summary and Conclusions

To summarize, we return to the questions posed at the beginning of this chapter:

Are RR intervals generated by a deterministic or stochastic process?

The KDM and D_2 results point to the same conclusion: for both sick and healthy patients, a majority of the RR intervals are deterministic and suggests deterministic modeling may be appropriate in modeling human heart rate variability. More data is required before a definitive conclusion can be drawn.

Is there any topological fractal structure in RR intervals, indicating the existence of an underlying low order deterministic chaotic system?

The answer to this question is *maybe*. The slopes of the correlation integrals for

Figure 5.31: Approximate entropy for ICD patients 1, ... , 6.
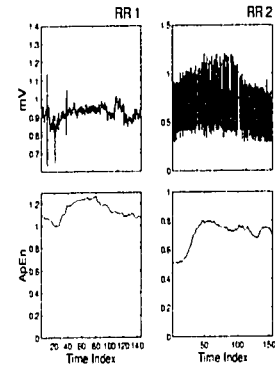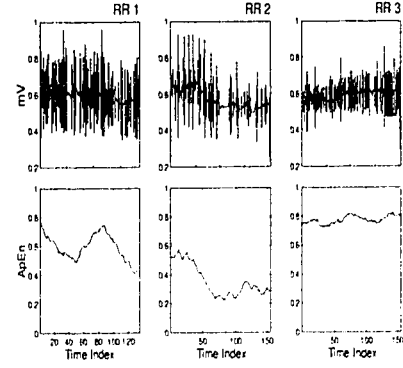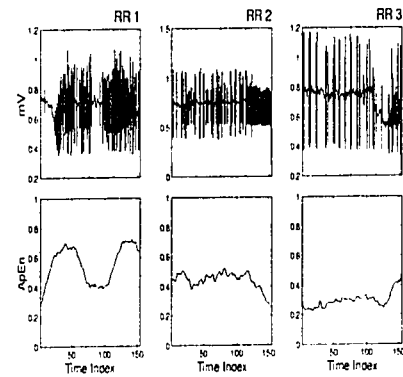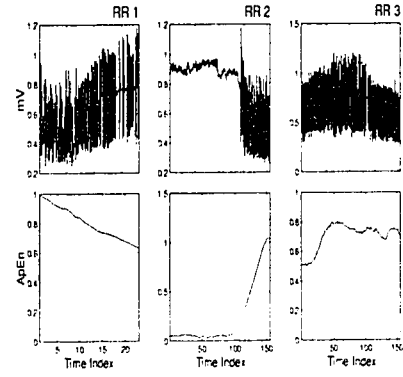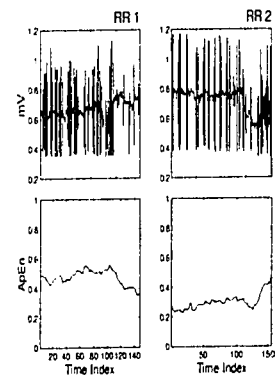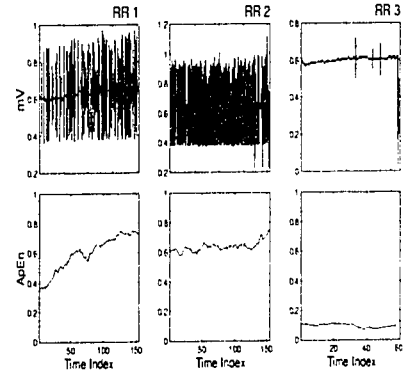
Figure 5.32: Approximate entropy for ICD patients 7, ... , 12.

236

Figure 5.33: Approximate entropy for ICD patients 13.....18.

the healthy patients do not have a substantial scaling plateau from which one may draw a decisive conclusion concerning the convergence of D_2. For the sick Holter patients, a minority of the patients produced more conclusive D_2 results, displaying obvious patterns of convergence in the correlation integrals. On the whole, the D_2 measures on RR intervals for both sick and healthy patients did not produce sufficient evidence to support the hypothesis that the dynamics of RR intervals are governed by low order deterministic chaotic equations of motion, i.e. while the D_2 results suggest that RR intervals are deterministic, they do not appear to be chaotic.

Do RR intervals exhibit any spectral scaling, indicating statistical self similarity common to colored random noise processes?

The results suggest that there exists some power law scaling in the RR intervals with regard to variance. For the bulk of the data, the estimated Hurst exponent is small and negative. However, in cases where $H < 0$ the data is uninterpretable without further analysis and cannot be classified as an fGn or fBm.

Can any of the measures be used to predict the onset of VF/VT ?

With regard to ApEn on RR intervals, the answer appears to be *no*. The only other time-dependent measure used for analysis of the RR data was the MODWPT. It is the author's opinion that use of the MODWPT offers much potential for signal classification and separation. To go beyond this and say that it should also work as

a good predictive tool would be misleading. The ability to predict the outcome of an event is significantly influenced by the data collected and not necessarily the tool used to analyze the data. If data (such as RR intervals) are not sensitive to subtle changes in the governing dynamic system from which they are measured, then it is not likely that linear transformations, such as the wavelet transform and all its varieties, will be sensitive enough to predict the future outcome of the system. Nonlinear measures are (typically) more sensitive to changes in a system, and consequently offer more predictive potential.

Can cardiac arrythmias be classified by a MODWPT best basis for a representative sample of a patient's RR intervals?

It is the opinion of the author that the MODWPT best basis has much potential in classifying disease states of the heart. RR intervals are but one aspect of a cardiac system and other relevant aspects should also be included in the analysis.

Appendix A

SIGNAL PROCESSING TECHNIQUES

The role of a transform in a mathematical sense is to alter a sequence or an equation in such a way as to pull out certain characteristics that may not otherwise be obvious. Thus, a transform may be viewed as a sort of mathematical catalyst for obtaining desired information.

Often times, it is desirable to know the spectral characteristics of a signal. For example, an engineer responsible for the dynamic stability of a bridge, car suspension system, or building must possess detailed knowledge of a structure's natural frequencies so as to avoid resonance and/or possible catastrophic failure. Frequency analysis has consequently become a very important tool in a wide variety of scientific and engineering disciplines. Since the spectral characteristics of a signal may not be obvious in the time domain, frequency analysis relies on the use of transforms which map a signal from the time domain to the frequency domain. While traditional frequency transforms work well in many situations, they often are inadequate in characterizing sharp discontinuities or abrupt transitional behavior. As a result, new techniques have been developed which are better able to deal with such signals.

240

The focus of this chapter is to present and discuss traditional frequency domain transformations and to introduce a novel transformation, the wavelet transform, which was developed in response to the limitations of classic techniques. In doing so, the concepts of orthogonality, linearity, basis functions, scale, resolution, and windowing (to name a few) will be illustrated and discussed. These sub-issues lead to the development and understanding of the wavelet transform in both its continuous and discrete version. The approach to these issues is mainly chronological, with slight diversions to discover the intricacies and subtleties of various methods.

A.1 Fourier Series Representations

In 1822, Jean Baptiste Joseph Fourier published *Théorie analytique de la chaleur* which was devoted to the mathematical theory of heat conduction. He established the partial differential equation governing heat diffusion and solved it by using an infinite series of trigonometric functions. In this he introduced the representation of a function as a series of sine and cosine waves now known as the Fourier series[1]. He asserted that any 2π-periodic function $f(x)$ can be expressed as the sum

$$a_0 + \sum_{k=1}^{\infty} [a_k \cos(kx) + b_k \sin(kx)] \tag{A.1}$$

In order to calculate the Fourier coefficients a_0, a_k, and b_k we shall use the following lemmas.

---

[1] It may interest the reader to know that *Théorie analytique de la chaleur* was not initially accepted due to lack of mathematical rigor. The ideas contained in it, however, have played a very important role in mathematics.

Lemma 1 *If n and m are distinct nonnegative integers,*

$$\int_{-\pi}^{\pi} \cos(mx)\cos(nx)dx = \int_{\pi}^{\pi} \sin(mx)\sin(nx)dx = 0$$

Lemma 2 *For any positive integers n and m,*

$$\int_{-\pi}^{\pi} \cos(mx)\sin(nx)dx = 0$$

Lemma 3 *For any positive integer n,*

$$\int_{-\pi}^{\pi} \cos^2(nx)dx = \int_{-\pi}^{\pi} \sin^2(nx)dx = \pi$$

Lemmas 1 and 2 are called *orthogonality* relationships and the functions $\cos(nx)$ and $\sin(nx)$ for $n = 1, 2, \ldots$ are said to be *orthogonal* on $[-\pi, \pi]$. Note that orthogonality implies that the inner product between two functions across a defined region is zero. Orthogonality turns out to be an important quality for frequency based transforms as we shall see later. Lemmas 1, 2, and 3 can be developed with routine integration methods and is stated here without proof.

In order to solve for the coefficients of Eq. A.1, we shall make the assumption that discrete summations and integrals are interchangeable. The first Fourier coefficient, a_0, can be solved for by integrating Eq. A.1 with respect to x from $-\pi$ to π.

$$\int_{-\pi}^{\pi} f(x)dx = a_0 \int_{-\pi}^{\pi} dx + \sum_{n=1}^{\infty} \left[a_n \int_{-\pi}^{\pi} \cos(nx)dx + b_n \int_{-\pi}^{\pi} \sin(nx)dx \right] = 2\pi a_0$$

Notice that all of the integrals in the summation are zero. Solving for a_0 we get

$$a_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)dx \qquad\qquad (A.2)$$

Now let k be a positive integer. We shall determine a_k by multiplying Eq. A.1 by $\cos(kx)$ and integrating from $-\pi$ to π. We get

$$\int_{-\pi}^{\pi} f(x)\cos(kx)dx = a_0 \int_{-\pi}^{\pi} \cos(kx)dx$$
$$+ \sum_{n=1}^{\infty}\left[a_n \int_{-\pi}^{\pi} \cos(nx)\cos(kx)dx + b_n \int_{-\pi}^{\pi} \sin(nx)\cos(kx)dx \right].$$

By Lemmas 1 and 2, all of the integrals on the right hand side are zero except the one involving $\cos(nx)\cos(kx)$ when $n = k$. The last equation therefore collapses to

$$\int_{-\pi}^{\pi} f(x)\cos(kx)dx = a_k \int_{-\pi}^{\pi} \cos(kx)\cos(kx)dx = a_k\pi$$

Solving for a_k we get

$$\boxed{a_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(kx)dx} \qquad (A.3)$$

To solve for b_k in Eq. A.1, we employ the same technique used to derive a_k by multiplying Eq. A.1 by $\sin(kx)$ and integrating from $-\pi$ to π to get

$$\int_{-\pi}^{\pi} f(x)\sin(kx)dx = a_0 \int_{-\pi}^{\pi} \sin(kx)dx$$
$$+ \sum_{n=1}^{\infty}\left[a_n \int_{-\pi}^{\pi} \cos(nx)\sin(kx)dx + b_n \int_{-\pi}^{\pi} \sin(nx)\sin(kx)dx \right].$$

which can be reduced to

$$\int_{-\pi}^{\pi} f(x)\sin(kx)dx = b_k \int_{-\pi}^{\pi} \sin(kx)\sin(kx)dx = b_k\pi$$

Solving for b_k we get

$$b_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(kx) dx \qquad (A.4)$$

Now that we have solved for the Fourier coefficients, lets take a closer look to reveal their underlying purpose in the Fourier series. In Eq. A.2, it can be seen that a_0 is simply the mean of the function $f(x)$ over the range $[-\pi, \pi]$. Electrical engineers often refer to a_0 as the direct current(DC) bias because it analogous to the voltage or current about which an electrical signal tends to oscillate. In Eq. A.3 and A.4 we see that the remaining Fourier coefficients are derived from the inner product of the function $f(x)$ with $\cos(kx)$ and $\sin(kx)$, respectively. The inner product of two functions is really a comparison of how similar each is to one another. The more that they are alike, the larger the absolute value of the inner product and thus the larger the Fourier coefficient amplitude for a given frequency k.

The sine and cosine waves of the Fourier series are referred to as *orthogonal basis functions*. A *basis* is a set of elementary building blocks that can be used to synthesize or reconstruct a function through linear combinations in a given range. An *orthogonal basis* assures us that the building blocks (basis functions) share no common or redundant information in regard to the original signal. The concept of a basis function can be more rigorously defined as follows.

Define an *n-vector* to be an n-tuple $< x_1, x_2, \ldots, x_n >$ in which each x_j is a real number. The number x_j is the jth component of $< x_1, x_2, \ldots, x_n >$. Let $\mathbb{R}^n$ be

the set of all n-vectors, $\mathbb{R}^n = \{< x_1, x_2, \ldots, x_n > \mid x_j \in \mathbb{R}, \quad j = 1, 2, \ldots, n\}$. As

an example, in planetary orbital mechanics one generally needs three coordinates to

locate an object's center of mass in space; that is to say 3-space or $\mathbb{R}^3$.

A *vector space* can contain not only a collection of real values in $\mathbb{R}^n$, but also a

collection of functions or other objects. Thus, $\mathbb{R}^n$ is a special case of a vector space.

Consider a vector space V and suppose that $< v_1, v_2, \ldots, v_n >$ are vectors in

V. A *linear combination* of $< v_1, v_2, \ldots, v_n >$ is any sum of scalars times these

vectors. Such a linear combination has the form

$$\alpha_1 \mathbf{v_1} + \alpha_2 \mathbf{v_2} + \ldots \alpha_n \mathbf{v_n}$$

If one of $< v_1, v_2, \ldots, v_n >$ is a linear combination of the others, the $< v_1, v_2, \ldots, v_n >$

are *linearly dependent*. Otherwise, they are *linearly independent*. For example,

$\sin^2(x)$, $\cos^2(x)$, and $\cos(2x)$ are linearly dependent because

$$\cos(2x) = \cos^2(x) - \sin^2(x)$$
$$= \alpha_1 \cos^2(x) + \alpha_2 \sin^2(x)$$

where $\alpha_1 = 1$, and $\alpha_2 = -1$. Given these developments, the vectors $< v_1, v_2, \ldots, v_n >$

form a *basis* for the vector space V if they are linearly independent and every vector

in V is a linear combination of these vectors. A basis does not imply that the basis

functions are orthogonal as is illustrated in the following example.

The functions 1, x, and x^2 form a basis for polynomials of degree 2 because they

are linearly independent; that is, we cannot write one of these functions as a linear

combination of the others valid for all x. However, 1, x, and x^2 do not form an orthogonal basis because there exists no range of values for x in which the all of the inner product combinations of the basis functions is equal to zero (Fig. A.1).
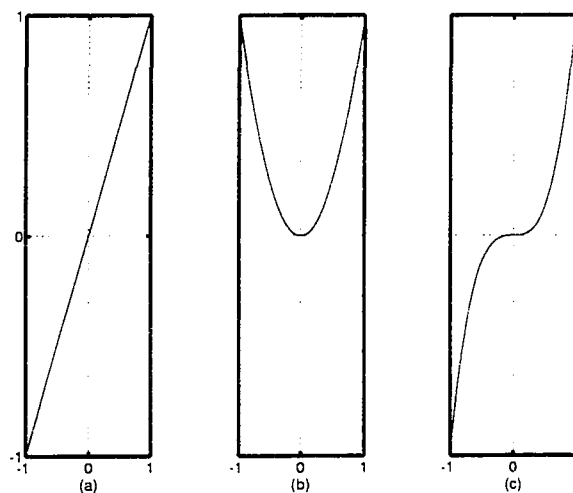
Figure A.1: The inner products of 1, x. and x^2 resulting in (a) x (b) x^2 and (c) x^3. For x on $[-L, L]$ for $L > 0$, the basis is not orthogonal because x^2 is even-symmetric. Thus, the inner product of 1 and x on $[-L, L]$ is not zero. For any other range of values for x, none of the inner product will sum to zero.

For periodic signals, the Fourier series usually does an excellent job in rebuilding the function in the time domain. In fact, it is sometimes sufficient to sum over a finite set of basis functions instead of the infinite sum as is defined in Eq. A.1. For example, consider Figure A.2 which shows various stages of development in the Fourier synthesis of a periodic saw tooth waveform. Only a relatively small number of Fourier coefficients are needed to adequately portray the original saw tooth signal.

246
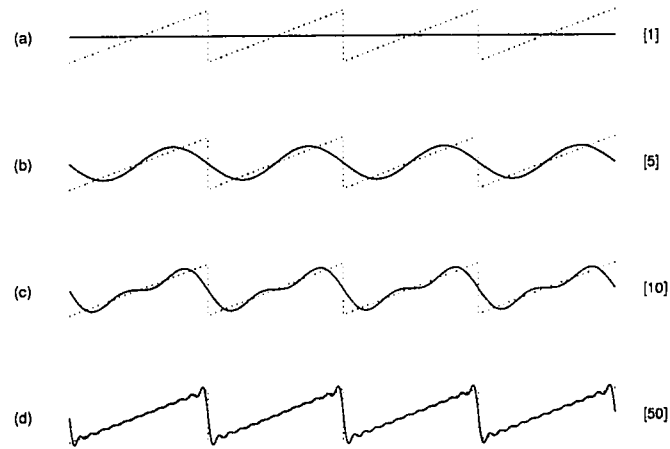
Figure A.2: The gradual Fourier synthesis of a periodic saw tooth wave. The numbers in brackets to the right of each plot indicates the number of distinct sine-cosine pairs used at the corresponding stage of development. The first Fourier coefficient, a_0, is shown in plot(a). It is easy to see that a_0 is indeed the mean of the signal (often referred to as the DC bias). The remaining plots are the partial sums $\tilde{f}(x)$ of the saw-tooth function $f(x)$ such that $\tilde{f}(x) = a_0 + \sum_{k=1}^{N} [a_k \cos(kx) + b_k \sin(kx)]$ for (b) $N = 5$ (c) $N = 10$ and (d) $N = 50$.

The Fourier series of Eq. A.1 can be extended quite easily to an interval $[-L, L]$ simply by changing scale. The change of variables $x \to \pi x / L$ converts the interval $[-L, L]$ to $[-\pi, \pi]$. If we apply this scaling to the definition of the Fourier series and Fourier coefficients the following is obtained and is shown without proof. Let f be integrable on $[-L, L]$.

1. The *Fourier coefficients of f on* $[-L, L]$ are

$$
\begin{aligned}
a_0 &= \frac{1}{2L} \int_{-L}^{L} f(x) dx \\
a_n &= \frac{1}{L} \int_{-L}^{L} f(x) \cos\left(\frac{n\pi x}{L}\right) dx \\
b_n &= \frac{1}{L} \int_{-L}^{L} f(x) \sin\left(\frac{n\pi x}{L}\right) dx
\end{aligned}
$$

for $n = 1, 2, 3, \ldots$.

2. The *Fourier series of f on* $[-L, L]$ is

$$
a_0 + \sum_{n=1}^{\infty} \left[a_n \cos\left(\frac{n\pi x}{L}\right) + b_n \sin\left(\frac{n\pi x}{L}\right) \right]
$$

in which the numbers $a_0, a_1, \ldots, b_1, \ldots$ are the Fourier coefficients of f on $[-L, L]$.

A.2 The Continuous Fourier Transform

Fourier's representation of functions as a superposition of sines and cosines has become ubiquitous for both the analytic and numerical solution of differential equations and frequency analysis in general. The Fourier series representation is the building block to the design of the *Fourier transform* which uses complex sinusoids as a basis as opposed to real sine and cosine waves used in the Fourier representation. The introduction of complex sinusoids serves as a mathematical convenience in probing a signal for oscillatory behavior which may be arbitrarily shifted in time. A sinusoidal shift in time is referred to as *phase* delay and can be readily calculated via the Fourier transform. The Fourier transform coefficients are the same coefficients defined in the Fourier series representation (Eq. A.1) and are explictly derived in the Fourier transform operation.

Definition 1 Continuous Fourier Transform

Let f be piecewise continuous on every finite interval $[-L, L]$ and $\int_{-\infty}^{\infty} |f(t)| dt$ be convergent, then define the Fourier transform $\mathcal{F}$ of f to be

$$\mathcal{F}\{f(t)\} = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt$$

Observe that $\mathcal{F}\{f(t)\}$ is a function of the frequency ω whose units are in rad/s. The Fourier transform of a lowercase function will be denoted as the uppercase form

of the same letter:

$$\mathcal{F}\{f(t)\} = F(\omega)$$

The inverse Fourier transform is given by

Definition 2   Continuous Inverse Fourier Transform

$$\mathcal{F}^{-1}\{F(\omega)\} = f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega)e^{i\omega t}dt$$

A.2.1   Properties of the Fourier Transform

A number of properties that prove useful in interpreting and using the Fourier transform are developed below.

Theorem 3   Linearity

If α and β are real numbers

$$\mathcal{F}\{\alpha f(t) + \beta g(t)\} = \alpha F(\omega) + \beta G(\omega)$$

assuming that the Fourier transform of f and g exist. Similarly,

$$\mathcal{F}^{-1}\{\alpha f(t) + \beta g(t)\} = \alpha \mathcal{F}^{-1}\{F(\omega)\} + \mathcal{F}^{-1}\{\beta G(\omega)\}$$

Theorem 4   Time Shifting

If $\mathcal{F}\{f(t)\} = F(\omega)$, then

$$\mathcal{F}\{f(t - t_0)\} = e^{-i\omega t_0} F(\omega).$$

Thus, the Fourier transform of f shifted by t_0 is obtained by multiplying the Fourier transform of f by $e^{-i\omega t_0}$.

Theorem 5   Frequency Shifting

If $\mathcal{F}\{f(t)\} = F(\omega)$, then

$$\mathcal{F}\{e^{-i\omega_0 t} f(t)\} = F(\omega - \omega_0)$$

Theorem 6   Scaling

If $\mathcal{F}\{f(t)\} = F(\omega)$ and a is a real nonzero constant, then

$$\mathcal{F}\{f(at)\} = \frac{1}{|a|} F\left(\frac{\omega}{a}\right)$$

Theorem 7   Modulation

If $\mathcal{F}\{f(t)\} = F(\omega)$, then

$$\mathcal{F}\{f(t)\cos(\omega_0 t)\} = \frac{1}{2}[F(\omega - \omega_0) + F(\omega + \omega_0)]$$

and

$$\mathcal{F}\{f(t)\sin(\omega_0 t)\} = \frac{i}{2}[F(\omega + \omega_0) - F(\omega - \omega_0)]$$

An important property of the Fourier transform is *linearity* (Theorem 3). Linearity implies that the *principle of superposition* is applicable. Thus, the spectrum of a signal composed of many sub-signals can be found by summing over the individual spectra for each sub-signal. In general, linearity is a desirable property since the influence of one component (sub-signal) in an analysis can be more easily identified.

The property of *time shifting* states that a shift of a signal in time produces a corresponding phase delay of $e^{-i\omega_0 t}$ in the Fourier transform. To see this more clearly, it is helpful to introduce the following relation.

Theorem 8   Euler's Formula

$$re^{\pm i\theta} = r[\cos(\theta) \pm i\sin(\theta)]$$

Euler's Formula is a means of relating the rectangular and polar coordinate form of complex numbers. Using Euler's formula, $e^{-i\omega t_0} = \cos(\omega t_0) - i\sin(\omega t_0)$. Now

consider a pulse modulated cosine wave of the form:

$$f(t) = \begin{cases} k\cos(\omega_0 t), & -a < t < a \\ \\ 0, & \text{otherwise} \end{cases}$$

The Fourier transform of the pulse is

$$\mathcal{F}\{f(t)\} = F(\omega) = \frac{k\sin\left[a(\omega - \omega_0)\right]}{\omega - \omega_0} + \frac{k\sin\left[a(\omega + \omega_0)\right]}{\omega + \omega_0}.$$

Notice that $F(w)$ in this case is purely real (no imaginary components exist). This result is intuitive in that one would expect a cosine wave to influence only the real portions (cosines) of the complex sinusoids used to probe $f(t)$ in the Fourier transform. This is a result of the orthogonality condition. If $f(t)$ is then temporally shifted by $\frac{\pi}{2\omega_0}$ such that $f(t) \rightarrow f(t - \frac{\pi}{2\omega_0})$ then

$$f(t) = \begin{cases} k\cos\left(\omega_0 t - \frac{\pi}{2}\right) = k\sin\left(\omega_0 t\right), & -a < t < a \\ \\ 0, & \text{otherwise} \end{cases}$$

To find the Fourier transform of the shifted modulated pulse, Theorem 4 is used:

$$\mathcal{F}\{f(t - t_0)\} = e^{-i\frac{\omega\pi}{2\omega_0}}\left[\frac{k\sin\left[a(\omega - \omega_0)\right]}{\omega - \omega_0} + \frac{k\sin\left[a(\omega + \omega_0)\right]}{\omega + \omega_0}\right].$$

In the limit as $\omega \rightarrow \omega_0$, $e^{-i\frac{\omega\pi}{2\omega_0}} \rightarrow e^{-i\frac{\pi}{2}} = -i$. Thus, $\mathcal{F}\{f(t - t_0)\}$ is purely imaginary as $\omega \rightarrow \omega_0$. This result is intuitive in that by shifting the cosine modulator to a sine wave, one expects the complex Fourier transform coefficients at the corresponding frequency ω_0 to be purely imaginary. Again, this is a result of the orthogonality condition.

Another interesting attribute of the Fourier transform is the *scaling property*. To understand it, first recall that when a function $f(t)$ is scaled:

$$f(t) \rightarrow f(at), \; where \; a > 0,$$

then it is contracted if $a > 1$ and expanded if $a < 1$ (Fig. A.3). The examination of
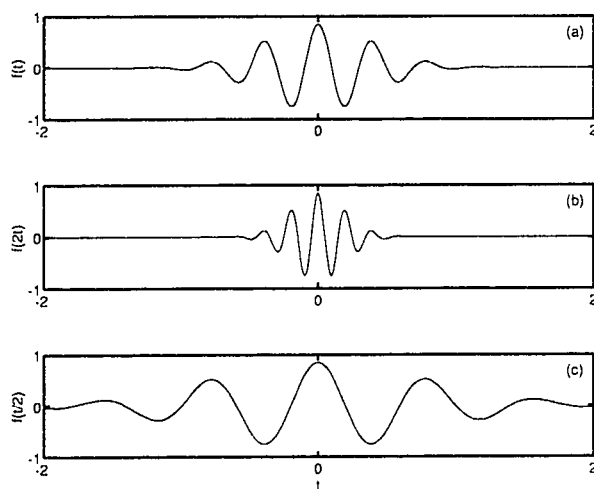
Figure A.3: Illustration of the effects of scaling a function $f(t)$ like $f(t) \rightarrow f(at)$ for (a) $a = 1$ (b) $a = 2$ and (c) $a = \frac{1}{2}$. For $a > 1$ the function $f(at)$ is contracted while for $a < 1$ $f(at)$ is expanded.

the scaling theorem (Theorem 6) reveals an interesting fact about signals or functions in general: Expansion(contraction) of a function in the time domain results in contraction(expansion) in the frequency domain. Consequently, those functions which change rapidly, such as the Dirac delta function or the Heaviside step function, exhibit wide spectra containing especially high frequencies while smooth functions

254

contain a narrow band of frequencies. This fact will be used later to show that a transform closely related to the Fourier transform is inadequate in analyzing signals that contain both rapidly changing and smooth components.

A.2.2 The Energy of a Signal

The concept of *signal energy* was introduced in the 1930s through a research effort by Littlewood, Paley, and Stein who computed the energy E_f of a signal $f(x)$ defined over the domain $[0, 2\pi]$ to be

$$E_f = \frac{1}{2} \int_0^{2\pi} |f(x)|^2 dx \qquad (A.5)$$

The term *energy* is analogous to the that of a linear spring with spring constant k. If the restoring force of the spring is $F_{sp} = kx$ where x is the displacement relative to the spring's unstretched length, then the energy of the spring is

$$E_{sp} = \frac{1}{2} k x^2 \qquad (A.6)$$

While the units of E_{sp} are often expressed in Joules or some other form of work, the units of a signal's energy E_f are generally unspecified. Based on the efforts to quantify energy in the time domain, the question naturally arises, "What happens to the energy of a signal after it has been mapped to the frequency domain via the Fourier transform?" To answer this question, consider the following analysis.

One means of filtering a signal $f(t)$ with a filter $g(t)$ is achieved through the method of convolution.

Definition 3 Convolution

If f and g satisfy the conditions $\int_{-\infty}^{\infty} |f(t)|dt$ and $\int_{-\infty}^{\infty} |g(t)|dt$ are convergent, then the convolution of f and g is denoted as $f * g$ and is defined by

$$[f * g](t) = \int_{-\infty}^{\infty} f(\tau)g(t - \tau)d\tau$$

The convolution operation works by time reversing g and sliding it along f in τ time increments. The inner product of the f and g serves to draw out the similarity between the two signals. The sequence g is often referred to as a *filter* since it used to filter out information from f. In effect, the complex sinusoidal basis of the Fourier transform is used to draw out or filter information and its ties with convolution can be derived as follows.

Theorem 9 Time/Frequency Convolution

Suppose that f and g satisfy the convergence conditions as stated in the definition of convolution, and let $\mathcal{F}\{f(t)\} = F(\omega)$ and $\mathcal{F}\{g(t)\} = G(\omega)$. Then

$$\mathcal{F}\{[f * g](t)\} = F(\omega)G(\omega) \quad \text{time convolution}$$

and

$$\mathcal{F}\{f(t)g(t)\} = \frac{1}{2\pi}[F * G](\omega) \quad \text{frequency convolution}$$

256

Proof The time convolution formula is proven below. A similar proof exists for the frequency convolution, however, it is omitted for the sake of redundancy.

$$
\begin{aligned}
\mathcal{F}\{[f * g](t)\} &= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau\right] e^{-i\omega t}dt \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\tau)g(t-\tau)e^{-i\omega t}d\tau dt \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\tau)g(t-\tau)e^{-i\omega t}dt d\tau \\
&= \int_{-\infty}^{\infty} f(\tau)\left[\int_{-\infty}^{\infty} g(t-\tau)e^{-i\omega t}dt\right] d\tau
\end{aligned}
$$

The bracketed term in the last line is the Fourier transform of $g(t-\tau)$. Apply the time shifting theorem (Theorem 4) to get

$$
\int_{-\infty}^{\infty} g(t-\tau)e^{-i\omega t}dt = e^{-i\omega\tau}\mathcal{F}\{g(t)\} = e^{-i\omega\tau}G(\omega)
$$

Therefore,

$$
\begin{aligned}
\mathcal{F}\{[f * g](t)\} &= \int_{-\infty}^{\infty} f(\tau)[G(\omega)e^{-i\omega\tau}]dt \\
&= G(\omega)\int_{-\infty}^{\infty} e^{-i\omega\tau}dt = F(\omega)G(\omega). \quad \cdot
\end{aligned}
$$

Moving closer to forming a relation between the energy of a signal in the Fourier and time domains, the frequency convolution theorem can be directly evaluated to

yield

$$\int_{-\infty}^{\infty} f(t)g(t)e^{-i\omega t}dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(v)G(\omega - v)dv$$

Let $\omega = 0$ to get

$$\int_{-\infty}^{\infty} f(t)g(t)dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(v)G(-v)dv.$$

Using the fact that $F(-\omega) = F^*(\omega)$ where $*$ denotes the complex conjugate operator.

$$\int_{-\infty}^{\infty} f(t)g(t)dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(v)G^*(v)dv. \tag{A.7}$$

Now substituting $g(t)$ for $f(t)$ in Eq. A.7 such that $\mathcal{F}\{f(t)\} = F(\omega)$ (since v is a dummy variable, we also make the substitution $v \to \omega$)

$$\int_{-\infty}^{\infty} [f(t)]^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |F(\omega)|^2 d\omega. \tag{A.8}$$

Eq. A.8 is known as *Parseval's identity* and it relates the energy of a signal in the time domain to its energy in the frequency domain. Examination of Eq. A.8 reveals that the only way for energy to be increased is through amplification of the sequence $f(t)$. Merely shifting a signal in time has no affect on the overall energy of the signal since the energy calculation is always summed over the domain of its support. Put another way, temporal shifting of a signal in time results in an energy transfer between the sine and cosine components of the Fourier coefficients. but does not affect the total energy level.

A.3 The Discrete-Time Fourier Transform

The Fourier transform is a valuable tool in signal analysis for continuous systems. However, problems can be encountered in evaluating the continuous spectrum defined as an integral over an infinite domain. Often times, it is more appropriate and efficient to approximate a continuous signal through discrete sampling methods. Discrete-time processing of continuous signals is very common in applications including communication systems, radar and sonar, speech, and video coding and enhancement [66]. Indeed, much of the work done in spectral analysis today is done in the digital domain, where computers are inherently equipped to analyze and manipulate discrete signals which may be represented as binary streams of 1's and 0's.

In analyzing discrete-time sequences, the *z-transform* is often used. The z-transform for discrete-time signals is the counterpart of the Laplace transform for continuous-time signals.

Definition 4   z-Transform

$$X(z) = \sum_{n=-\infty}^{\infty} x[n]z^{-n}$$

If we let $z = e^{i\omega}$, we obtain the discrete-time Fourier transform (DTFT).

Definition 5   Discrete-Time Fourier Transform

$$X(e^{i\omega}) = \sum_{n=-\infty}^{\infty} x[n]e^{-i\omega n}, \quad n \in \mathbb{Z}$$

The DTFT is thus seen as a special case of the z-transform. Since the z-transform is a function of a complex variable, it is convenient to interpret it using the complex z-plane. The contour $|z| = 1$ is the unit circle in the z-plane. Furthermore, since $|z| = |e^{i\omega}| = 1$, the unit circle in the z-plane corresponds to the DTFT (Fig. A.4).

A.4 The Discrete Fourier Transform

One of the most important sequences in analyzing linear processes is the discrete-time complex exponential

$$x[n] = Ae^{i\omega_0 n}$$

By analogy with continuous-time complex exponential functions, ω_0 denotes frequency while n is an integer specifying the nth value of the sequence resulting from sampling a continuous sequence $x(t)$ like $x[n] = x(nT)$ where T is some period of time not necessarily a constant. For our purposes, it will suffice to consider a uniform sampling period. There are two important differences between discrete-time and continuous time representations of complex exponentials. The first difference is revealed by considering discrete-time complex exponentials with frequencies $(\omega_0 + 2\pi r)$ for r

an integer. In this case,

$$x[n] = Ae^{i(\omega_0 + 2\pi r)n}$$

$$= Ae^{i\omega_0 n}e^{i2\pi rn}$$

$$= Ae^{i\omega_0 n}$$

It is apparent that the frequencies $(\omega_0 + 2\pi r)$ are indistinguishable in the discrete-time domain. Thus, in dealing with discrete-time exponentials, we need only consider frequencies in a frequency interval of 2π such as $-\pi < \omega_0 \leq \pi$ or $0 < \omega_0 \leq 2\pi$. The second main difference concerns the periodicity of complex exponentials. In the continuous-time case, a complex-exponential may be periodic with period equal to 2π divided by frequency. In the discrete-time case, a periodic sequence is represented as

$$x[n] = x[n + N], \quad \text{for all } n,$$

where the period N is necessarily an integer. Testing the periodicity condition for a discrete-time complex exponential yields

$$Ae^{i\omega_0 n} = Ae^{i\omega_0(n+N)}$$

$$= Ae^{i\omega_0 n}e^{i\omega_0 N}$$

The only way for this equation to hold is if

$$\omega_0 N = 2\pi k \qquad (A.9)$$

where k is an integer. As a result, complex exponential are not necessarily periodic with period $\frac{2\pi}{\omega_0}$ and, depending on the value of ω_0, may not be periodic at all. For example, for $\omega_0 = 1$, there exists no combination of integers k and N such that Eq. A.9 is satisfied. The periodicity condition is satisfied however for any frequency ω_k such that $\omega_k = \frac{2\pi k}{N}$, k is an integer.

In combining the observation that for discrete-time complex exponentials, $(\omega_0 + 2\pi r)$ are indistinguishable frequencies and the periodicity condition (Eq. A.9), we see that there are N distinguishable frequencies for which the corresponding sequences are periodic with period N. One such frequency set is $\omega_k = \frac{2\pi k}{N}$, $k = 0, 1, 2, \ldots, N-1$.

Since the basis for Fourier analysis is composed of complex exponentials, we can use the preceding results to extend the discrete-time Fourier transform to the discrete Fourier transform (DFT). The main difference between the DTFT and the DFT is that the DFT is designed to analyze finite-length sequences. In addition, with the imposed condition of assumed periodicity, the DFT utilizes the frequency set $\omega_k = \frac{2\pi k}{N}$, $k = 0, 1, 2, \ldots, N-1$ to obtain N distinguishable frequency estimates. Given these results, the discrete Fourier transform X of a continuous function $f(t)$ is defined as

Definition 6   Discrete Fourier Transform

$$X[k] = T \sum_{k=0}^{N-1} f(kT)e^{-i\frac{2\pi k}{N}n} \text{ for } n = 0, 1, 2, \ldots, N-1.$$

where k denotes the kth Fourier coefficient, T is the sampling period such that $T > 0$, and N is a positive integer.

The discrete Fourier transform (DFT) is seen as an equi-radial sampling of the discrete-time Fourier transform around the unit circle in the z-plane (See Fig. A.4). The range of frequencies span from $0 < \omega_k \leq 2\pi$. The ω_k are actually normalized frequencies in units of $\frac{rad}{sec}$ such that $\omega = 2\pi$ corresponds to the sampling frequency $f_s = \frac{1}{T}$. The normalization of the frequencies represented in the DFT is a result of sampling the DTFT in the frequency domain, much like the renormalization that occurs in time when sampling the continuous-time signal $x(t)$ to develop $x[n]$. In contrast, the spacing of sequence values $x[n]$ is always unity; i.e. the time axis is normalized by a factor T. There are many properties of the DFT. One of the most relevant properties is the fact that half of the DFT coefficients are redundant when $x[n] = f(kT)$ is a real valued sequence. The redundant coefficients reside below the *Real* axis in the z-plane (Fig. A.4). In addition, there are two main types of distortion that one generally needs to be aware of when analyzing discrete, finite length signals. These distortional effects, known as *aliasing* and *leakage* are discussed below.
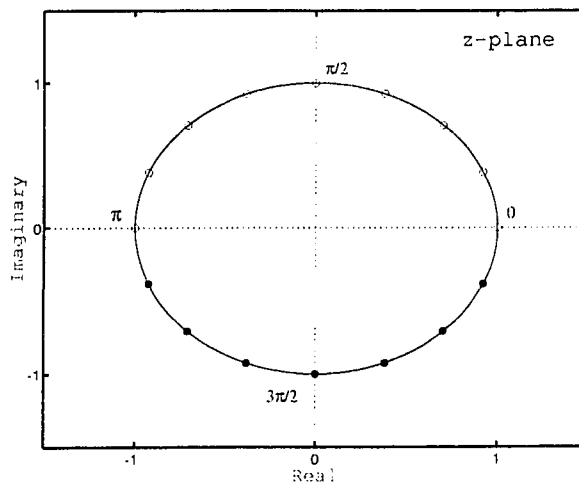
Figure A.4: Illustration of the Fourier frequencies as distributed on the unit circle in the z-plane. The filled circles denote redundant frequencies when the input $f(t)$ is purely real.

A.4.1 Aliasing: Cause and Effect

Spurious low frequency content can be introduced into the FT by undersampling a continuous signal in time. Specifically, if the highest frequency component in a continuous band limited signal, $x_c(t)$, is Ω_N, the sampling frequency, Ω_S, used to discretize $x_c(t)$ must be $\Omega_S \geq 2\Omega_N$. Ω_N is commonly referred to as the *Nyquist frequency* while the frequency $2\Omega_N$ is called the *Nyquist rate*. In order to understand the aliasing artifact it is helpful to consider the process of discretizing $x_c(t)$ through an impulse train modulation. The modulating signal $s(t)$ is a periodic impulse train

$$s(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT),$$

264

where $\delta(t)$ is the unit impulse function or Dirac delta function. Consequently, $x_S(t) = x_C(t)s(t)$, or

$$x_S(t) = x_C(t) \sum_{n=-\infty}^{\infty} \delta(t - nT).$$

Through the sifting property of the impulse function, $x_S(t)$ can be written as

$$x_S(t) = \sum_{n=-\infty}^{\infty} x_C(t)\delta(t - nT).$$

Note that this is a multiplication process performed in the time domain. By the Convolution Theorem, the process can also be performed in the frequency domain by convolving the Fourier transform of $x_S(t)$ with that of $s(t)$.

It has been shown that the FT of a periodic impulse train is a periodic impulse train [66] (Fig. A.5). Given these basic principles, consider the process by which aliasing can occur by examining Figure 1. The Fourier transform of the band limited signal $x_C(t)$ is shown in Figure 1a. Using convolution in the frequency domain, spectra of the sampled signal are produced (Fig 1c, 1d). Notice that for signals that are sampled at a frequency less than the Nyquist rate, an overlap in spectral content occurs during the convolution process, giving rise to the artificial spectral power around the difference frequency $(\Omega_S - \Omega_N)$. This type of distortion is referred to as *aliasing*. An immediate implication of aliasing is that $X_c(j\Omega)$ is no longer recoverable by low pass filtering. Based on the observation that aliasing occurs at the difference frequency $(\Omega_S - \Omega_N)$, it can be said that the part of the signal which has frequency components above $\Omega_S/2$ will appear reflected (or 'aliased') in the range of $[0, \Omega_S/2]$.
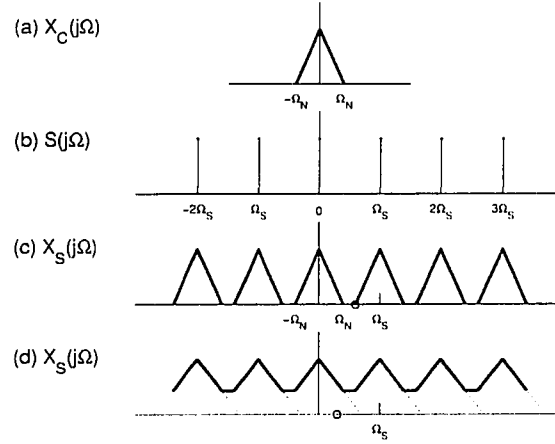
Figure A.5: The effect of aliasing in the frequency domain. (a) spectrum of the original signal (b) spectrum of the sampling function (c) spectrum of the sampled signal with $\Omega_S > 2\Omega_N$ and (d) spectrum of the sampled signal with $\Omega_S < 2\Omega_N$

A.4.2 Leakage: Cause and Effect

In the discrete Fourier transform, a finite number of spectral components are used to approximate the Fourier series representation of the input. These frequencies, denoted as *Fourier frequencies*, are specified by the equi-spaced sampling of the continuous Fourier transform around the unit circle in the z-plane (i.e. sampling $z = e^{i\theta}$). If the input contains energy at a frequency in between the Fourier frequencies, then this energy is spread out, or leaked, across the *local* Fourier frequencies. This type of distortion is referred to as *leakage*.

As an example of leakage, consider an input formed by summing equi-amplitude sinusoids at distinct frequencies (Fig. A.6). By inspection, one can easily identify the areas of the DFT where leakage has occurred: around $f = 20.4, 51.0, 142.8$, and 183.3 Hz. Notice, however, that the energy of one of the sine components is captured by only one spike, hence no leakage, in the DFT at $f = 102$ Hz. The question then arises, "What separates the 102 Hz sinusoid from the others which form the input such that it does not experience any leakage?" The answer is found in the discretization of the continuous Fourier transform (FT). The FT is broken up into N equi-spaced divisions, where $N = 256$ is the number of points in the input. With a sampling rate of $f_s = 512$ Hz, one such division is specified by $\Delta f = f_s/N = 2$ Hz. The Fourier frequencies are thus formed by

$$F_n = n\Delta f \qquad\qquad n = 0, 1, \ldots, N - 1$$

Leakage in the DFT

We can estimate the location of spectral energy in the input with respect to the F_n by forming the nondimensional ratio

$$\gamma_i = \frac{f_i}{\Delta f} = \frac{f_i N}{f_s} \qquad\qquad i = 1, 2, \ldots, 5$$

where f_i are the individual frequencies (in Hertz) that form the input:
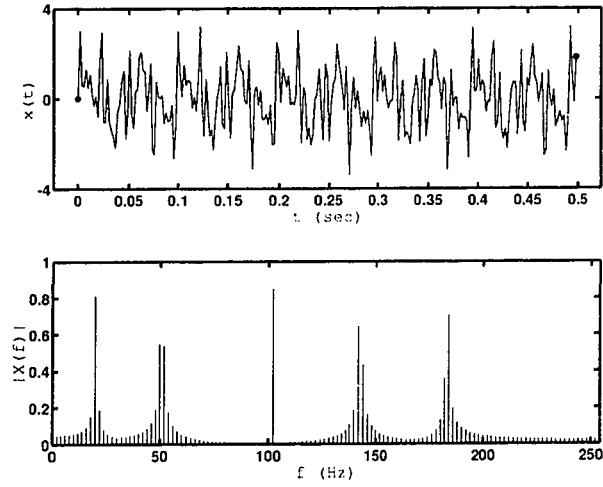
Figure A.6: The DFT of a function $x(t) = \sum_{n=1}^{5} A\sin(2\pi f_n t)$ for $A = 0.85$ and $f_n = \{20.4, 51.0, 102.0, 142.8, \text{ and } 183.3\}$ Hertz. The plots are (a) the time history of the $x(t)$ and (b) the DFT of $x(t)$.

$$\gamma_1 = \frac{20.4}{\Delta f} = 10.2$$

$$\gamma_2 = \frac{51.0}{\Delta f} = 25.5$$

$$\gamma_3 = \frac{102.0}{\Delta f} = 51 \qquad\qquad (A.10)$$

$$\gamma_4 = \frac{142.8}{\Delta f} = 71.4$$

$$\gamma_5 = \frac{183.3}{\Delta f} = 91.65$$

From Eq. A.11 we see that the only frequency that is evenly divisible by Δf is $f_3 = 102.0$ Hz, with $\gamma_3 = 51.0$. This equivocates to a Fourier frequency at exactly the same frequency of f_3. Hence, no leakage occurs at 100 Hz in the DFT (Fig. A.6b). Notice, however, that all of the other frequencies are *not* evenly divisible by Δf and therefore their energy is leaked onto local Fourier frequencies (Fig. A.6a). In addition, the powers of the other dominant frequency components are not the same as that at $f_3 = 102$ Hz, even though the input was formed by sinusoids with the same amplitude. It is clear that the worst case scenario is represented by the energy at $f_2 = 51.0$ Hz. By Eq. A.11 we see that $\gamma_2 = 25.5$, suggesting that the energy lies exactly in the middle of two Fourier frequencies. In such a case, the largest side lobe off the main spike around f_2 contains more power than those around f_1, f_4, and f_5. In general it can be said that the closer the γ_n ratio is to a whole number, the closer the main spike comes to represent the true energy of the input around that frequency.

There is one other form of leakage which is a direct consequence of using a finite length input coupled with an assumed periodicity. The cause is that the input sequence has starting and ending amplitudes which significantly differ. The effect is that spurious high frequency content is introduced into the spectrum. In effect, the jump discontinuity that exists between end points of the input must be fit by the appropriate Fourier frequencies. It is natural then that the larger the discontinuity, the larger the level of spurious energy that will inhabit the Fourier

frequencies approximating the jump. This energy spread can be seen in the upper frequency range around $f_5 = 183.3$ Hz (Fig A.6b). The upper band distortion cannot be attributed to leakage alone in that, by way of comparison, other areas in the spectrum which exhibit leakage have a rolloff of around 40 Hz (Compare rolloffs of f_2 towards f_3 and f_4 towards f_3). To this end, we can estimate the rolloff of the leakage around f_5 to be at 220 Hz. Yet, there exists energy up to and beyond this point; suggesting that a type of distortion other than leakage is the cause.

A.5 Windowing Signals in Time

In Fourier analysis, there always exists an assumed periodicity in the signal. When dealing with finite length discrete-time signals, the periodicity is of the form $x[n] = x[n + rN]$, where N is the period necessarily an integer, and r is also an integer. Often times, however, finite length signals do not abide by this rule as was discussed in the previous section and illustrated in Fig. A.6. The need for the multiplication of $x[n]$ by $w[n]$, where $w[n]$ is introduced as a *window function*, is required to force the end points $\{x[0], x[N-1]\}$ to match as well as satisfy the finite-length requirement of the DFT. Some commonly used windows are shown in Fig. (A.7) and are defined as follows:

Rectangular

$$w[n] = \begin{cases} 1, & 0 \le n \le M \\ 0, & \text{otherwise} \end{cases}$$

Bartlett (triangular)

$$w[n] \equiv \begin{cases} \frac{2n}{M}, & 0 \le n \le \frac{M}{2} \\ 2 - \frac{2n}{M}, & \frac{M}{2} < n \le M \\ 0, & \text{otherwise} \end{cases}$$

Hanning

$$w[n] = \begin{cases} 0.5 - 0.5\cos\left(\frac{2\pi n}{M}\right), & 0 \le n \le M \\ 0, & \text{otherwise} \end{cases}$$

Hamming

$$w[n] = \begin{cases} 0.54 - 0.46\cos\left(\frac{2\pi n}{M}\right), & 0 \le n \le M \\ 0, & \text{otherwise} \end{cases}$$

The choice of window is governed by the desire to have $w[n]$ be as short as possible so as to limit the computational cost of implementing the filter while having the frequency response $W(\omega)$ be close to that of an impulse in the frequency domain. The closer the $W(\omega)$ is to an impulse in the frequency domain, the less distortion it will impose of the filtered signal. This is simply due to fact that the multiplication of the $x[n]$ by $w[n]$ in the discrete-time domain is analogous to periodic convolution in the frequency domain[2]. Thus, if $W(\omega)$ were truly an impulse such that $W(\omega) = \delta(\omega)$.

---

[2]For these reasons, windowing a signal in time is a filtering operation in the frequency domain.
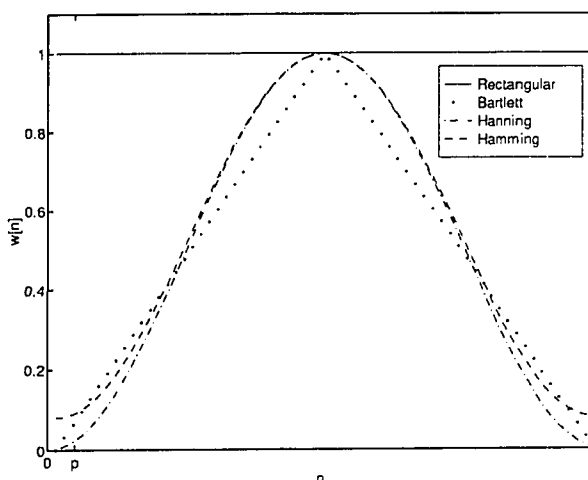
Figure A.7: Commonly used windows.

the resulting convolution would yield the original spectrum $X(\omega)$. Unfortunately, the length of a window in time is inversely proportional to its spectral resolution. Thus, as window length decreases, the main lobe of the window frequency response widens and consequently lessens frequency resolving characteristics of the window[3] (See Fig. A.8).

An ideal frequency response of a filter is a rectangular window in the frequency domain. The convergence of a window spectrum consisting of a finite number of sequence values M is identically cognate to the issue of convergence in Fourier series. Thus, the larger the number of window filter taps M, the better the approximation

---

[3]The fact that ideal temporal and frequency resolution cannot be simultaneously achieved has played an integral part in the development of new signal processing techniques such as the wavelet transform. We shall examine this topic more closely later on in the dissertation.
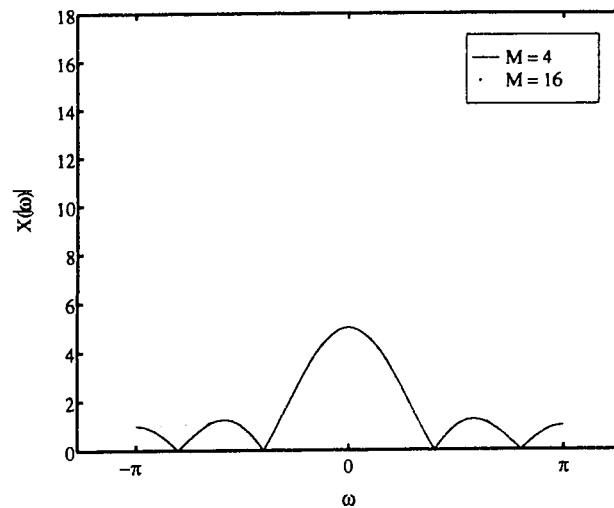
Figure A.8: Magnitude of the Fourier spectrum for a rectangular window of length $M = 4, 16$. The window length is inversely proportional to the main lobe width.

to the ideal rectangular pulse in the frequency domain. However, just as one achieves

nonuniform convergence or *Gibb's phenomena* in the time domain for a rectangular

pulse, ideal filter convergence using windows is imperfect near sharp transitions in

ideal spectra; i.e. the edges of the rectangular box in the frequency domain are hard

areas to "fit". These hard to fit regions usually result in oscillating side lobe peaks in

the frequency response of a window. In the theory of Fourier series, it is known that

the Gibb's phenomena can be lessened with a less abrupt truncation of the Fourier

series. By tapering a window smoothly to zero at each end, the side lobes of $W(\omega)$

will be attenuated. However, this comes at the cost of increasing the width of the

main lobe.

A.6   The Short-Time Fourier Transform

Time-Frequency representations have become a very important tool in the analysis of signals whose spectral components vary with time. In these cases, the Fourier Transform (FT) is directly insufficient to classify temporally localized frequency content. While the FT maps a sequence from the time domain to the frequency domain, it does not maintain a combination of the two domains. Strictly speaking, time information is available in the phase information of the FT but is not easily interpreted. An obvious solution is to divide the signal into blocks via windowing operations, and take the FT of each successive block. In this sense, we can obtain an estimate of time localized spectral content. This process is known as the Short Time Fourier Transform (STFT) and is defined in continuous-time as

Definition 7   Short-Time Fourier Transform

$$STFT(\tau,\omega) = \int_{-\infty}^{\infty} x(t)g^*(t-\tau)e^{-i\omega t}dt \qquad (A.11)$$

where the signal $x(t)$ is assumed stationary[4] when seen through the window function $g(t)$ of limited extent, centered around a time location τ. The parameter ω is similar

---

[4]The term *stationary* implies that the central moments $\mu_n = \int_{-\infty}^{\infty}(x-\overline{x})^n p(x)\,dx$ of a function $f(x)$ with probability density $p(x)$ and mean $\overline{x}$ are invariant to arbitrary shifts in time. In other words, the statistical properties of a stationary process are independent of the absolute time origin. Generally, only the first two moments need to fit these criteria before the signal is considered stationary; although technically, processes whose first two moments are the only moments to be invariant to shifts in time are referred to as *weakly stationary* processes.

274
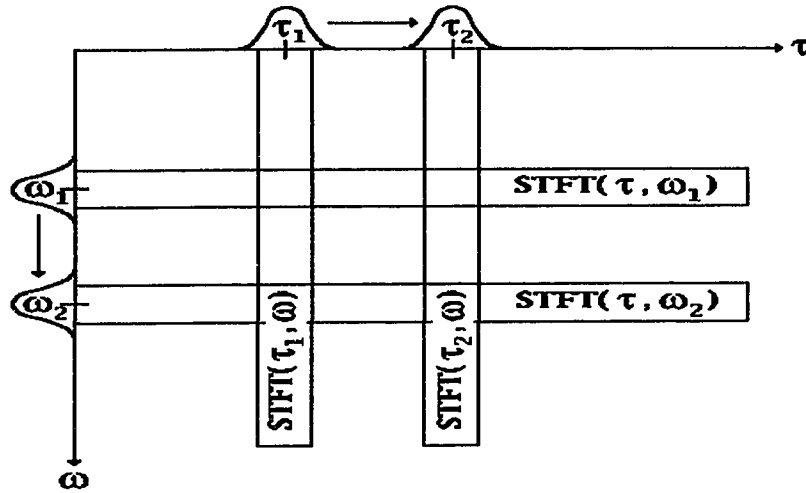
Figure A.9: The STFT in the frequency-time plane. The STFT is interpreted as either a succession of Fourier transforms of windowed segments in time (vertical bands) or as a modulated filter bank analysis (horizontal bands).

to the Fourier frequency and many properties carry over to the STFT. However, the STFT analysis depends critically on the choice of $g(t)$. The effect of the window is to localize the signal $x(t)$ in time.

There are two different interpretations of Eq. A.11. One view of the STFT is that it is a succession of Fourier transforms of windowed segments centered around various time lags τ. The vertical stripes in Fig. A.9 represent this view. An alternate view is to regard the STFT as a modulated filter bank operation. At a given frequency ω, the entire signal in time is filtered with a bandpass filter which has an impulse response modulated to that frequency. The horizontal lines in Fig. A.9 reflect this viewpoint. The "thickness" of the horizontal and vertical extensions in Fig. A.9

reflect the ability of the STFT to discriminate between two impulses in the frequency domain and time domain, respectively. Given a window function $g(t)$ and its Fourier transform $G(\omega)$, the *bandwidth* $\Delta\omega$ of the filter is defined as

$$\Delta\omega^2 = \frac{\int_{-\infty}^{\infty} \omega^2 |G(\omega)|^2 d\omega}{\int_{-\infty}^{\infty} |G(\omega)|^2 d\omega} \qquad (A.12)$$

where the denominator is the energy of the window $g(t)$. Eq. A.12 can be used to calculate the minimum frequency difference $\Delta\omega$ two sinusoids must be in order to be discriminated. i.e. if two pure sinusoids are less than $\Delta\omega$ apart, they cannot be individually isolated by the STFT. Thus, the resolution in frequency of the STFT is given by $\Delta\omega$. Similarly, the spread in time is given by Δt as

$$\Delta t^2 = \frac{\int_{-\infty}^{\infty} t^2 |g(t)|^2 dt}{\int_{-\infty}^{\infty} |g(t)|^2 dt} \qquad (A.13)$$

where again the denominator is the energy of the window $g(t)$. Here, two pulses cannot be discriminated unless they are separated by Δt in time.

Ideally, it is desirable to have good resolution in both the frequency and time domains. Unfortunately, the window size in time is inversely proportional to the bandwidth as was demonstrated in Fig. A.8. Thus, there is always a tradeoff between the resolution in the frequency and time domains. In fact, the product of $\Delta\omega$ and Δt is lower bounded such that

$$\Delta t \Delta f \geq \frac{1}{4\pi} \qquad (A.14)$$

Eq. A.14 is referred to as the *Heisenberg inequality* or the *uncertainty principle*. As a note, Gaussian windows are often used because they meet the Heisenberg bound

with equality.

A major drawback of the STFT resides in the fact that once a window size is chosen, the time-frequency resolution is fixed (Fig. A.10). Therefore, sharp transitional (bursty) quasi-stationary components can be analyzed with good temporal resolution or good frequency resolution, but not both.

Figure A.10: The time-frequency resolution of the STFT. The tiles represent the essential concentration of a given basis function in the time-frequency plane. Once a window size is selected for the STFT, the resolution for both domains remains fixed.

A.7 The Continuous Wavelet Transform

Working as a seismology engineer, Morlet recognized that the use of variable size windows could be used to attenuate the frequency-time resolution difficulties of the STFT. Since Morlet's work, researchers such as Guido Weiss, Ronald Coifman, Y.

Meyer, Stephan Mallat, and Ingrid Daubechies have helped develop a novel technique that overcomes many of the obstacles found in the STFT and other time-frequency representations (TFR). The technique is referred to as the *wavelet analysis*.

Wavelet analysis is defined such that, when viewed as a filter bank, time resolution increases with the central frequency of the analysis filters. This is achieved by imposing the condition

$$\frac{\Delta f}{f} = c \tag{A.15}$$

where c is a constant. The analysis filter bank is thus composed of bandpass filters of constant relative bandwidth (often referred to as "constant-Q" analysis in the signal processing community) (Fig. A.11). Of course, this type of analysis is also subject to the limitations of the Heisenberg inequality. Therefore, with an increase (decrease) in Δf, a corresponding decrease (increase) in Δt must occur. The resulting logarithmic resolutional structure works best on signals which contain a mix of short lived bursty components and low frequency components of long duration. Many signals such as electrocardiograms (ECG), speech, and nuclear magnetic resonance curves (to name a few) contain just such a mix of relatively short lived high and long lived low frequency components.

The *continuous wavelet transform (CWT)* encompasses the preceding ideas with one simplification: all impulse responses of the filter bank are derived from scaled and translated versions of a prototype filter known as a *mother wavelet*. Thus, a

Frequency
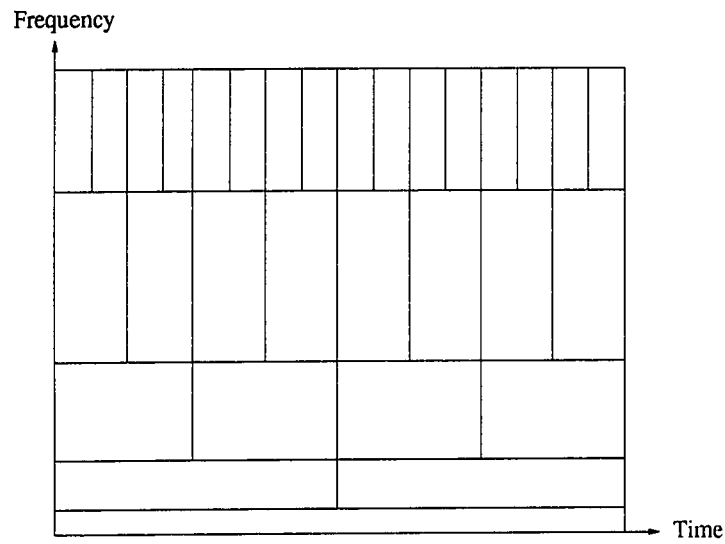
Time

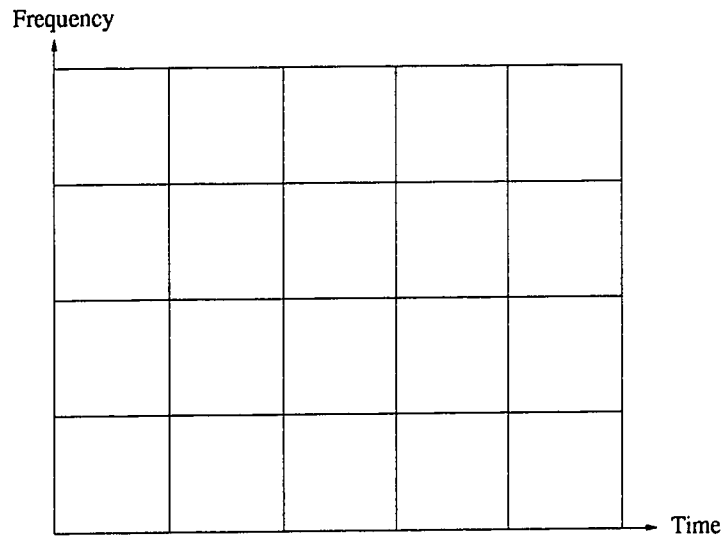Figure A.11: The time-scale resolution of the CWT. The tiles represent the essential concentration of a given basis function in the time-scale plane. Varying the window size logarithmically in time results in a logarithmic distribution in frequency concentration. With this structure, good resolution in frequency is achieved at low frequencies while good time resolution is achieved at high frequencies.

typical mother wavelet $\psi(x)$ can be used to form the children filters or *wavelets* as

$$\psi_a(x) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{x}{a}\right)$$

where $\frac{1}{\sqrt{|a|}}$ is used for energy normalization and a is called a *scale factor*. From this, the continuous wavelet transform is defined as

Definition 8   Continuous Wavelet Transform

$$T_\psi[f](b,a) = \frac{1}{\sqrt{|a|}}\int_{-\infty}^{\infty} f(x)\psi^*\left(\frac{x-b}{a}\right)dx \qquad (A.16)$$

By using a scaled version of a mother wavelet, the CWT can be used to explore bursty behavior with a small window and long term gross behavior with a large window (Fig. A.12). In effect, the wavelet analysis gives a view of the entire forest as well as the details of a single tree and thus is often referred to as a sort of mathematical microscope.

The CWT and the STFT are alike in that they are both linear transforms. However, the basis functions used in the CWT are more complicated than the sines and cosines used in the STFT. The differences in the analysis filters (basis) used in the STFT and the CWT is highlighted in Fig. A.12. Notice that the window length of the STFT for both low and high frequency content remains the same. As a consequence, analysis of relatively high frequency content for a constant window size in

280

the STFT requires an increase in the number of oscillations within the window. The wavelet transform, however, maintains the same number of oscillations in a dynamically sized window (Fig. A.12). To analyze relatively high frequency content with wavelets, the window length is decreased while the number of oscillations remains constant. Both of these effects are achieved via a scaling operation.
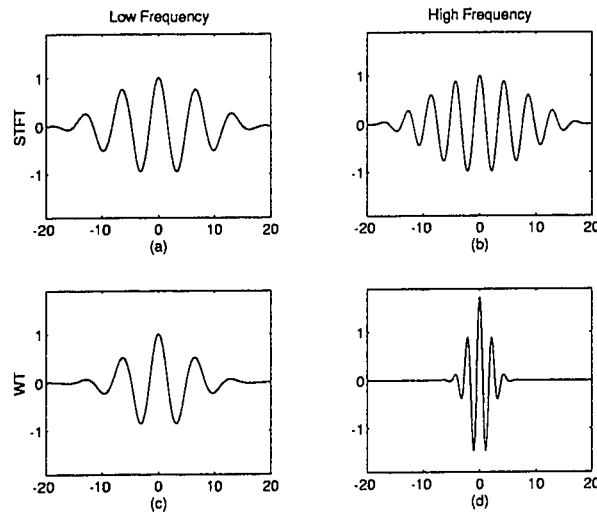
Figure A.12: Typical STFT and CWT analyzing waveforms. The STFT basis functions are windowed here with a Gaussian for (a) a low frequency and (b) high frequency waveforms. The corresponding wavelets are shown in plots (b) and (d), respectively.

A.8 The Discrete Wavelet Transform

The discrete version of the continuous wavelet transform will be approached from a a multiscale interpretation known as *multiresolution analysis (MRA)*. The goal of multiresolutional analysis is to provide a simple hierarchical framework by which

signal information can be interpreted at various resolutions. We will restrict ourselves to working with one dimensional signals in the development of discrete wavelet representations.

Let A_{2^j} be the operator which approximates a signal at resolution 2^j. We suppose that an original signal $f(x)$ is measurable and has finite energy. Thus, $f(x) \in L^2(R)$ where $L^2(R)$ denotes the vector space of measurable, square-integrable, one dimensional functions $f(x)$. To develop a multiresolution approximation we will use the following guidelines.

MRA Conditions

1. A_{2^j} is a linear operator.

2. Among all the approximated functions at resolution 2^j, $A_{2^j} f(x)$ is the function which is most similar to $f(x)$. This requires that the A_{2^j} operator is an orthogonal projection on the vector space V_{2^j}.

3. The approximation of a signal at resolution 2^{j+1} contains all of the information necessary to compute the signal at a coarser resolution 2^j. Since A_{2^j} is a projection operator on V_{2^j} this principle is equivalent to

$$\forall j \in \mathbf{Z}, \; V_{2^j} \subset V_{2^{j+1}}$$

where $\mathbf{Z}$ is the set of all integers.

4. The approximation operation is similar at all resolutions. Thus, the spaces of approximated functions should be derived from one another by scaling each approximation function by the ratio of their resolution values.

5. When computing an approximation of $f(x)$ at resolution 2^j, some information about $f(x)$ is lost. However, as the resolution increases to $+\infty$, the approximated signal converges to the original signal. Conversely, as the resolution decreases to 0, the approximated signal contains less and less information and converges to zero.

Condition 2 relates the fact that at a given resolution 2^j, the approximation to $f(x)$ given by $A_{2^j} f(x)$ is the closest one to $f(x)$ that is possible. Condition 3 specifies that a signal approximation at a given resolution can be used to develop approximations at lower resolutions. Mathematically speaking, the vector space V_{2^j} is a subset of $V_{2^{j+1}}$. For normalization purposes, we shall assume the original signal to be at resolution 1, i.e. the original signal is at the finest resolution.

In order to numerically characterize the A_{2^j} operator, an orthonormal basis of V_{2^j} must be found. The following shows that an orthonormal basis can be defined by dilating and translating a unique function $\phi(x)$.

Assumption 1   Existence of Unique a Scaling Function

Let $(V_{2^j})_{j \in Z}$ be a multiresolution approximation of $L^2(R)$. There exists a unique function $\phi(x)$ called a *scaling function*, such that for $j \in \mathbf{Z}$, (the dilation of $\phi(x)$ by 2^j), then

$$\left(\phi_{j,n}(x) = 2^{j/2} \phi(2^j x - n) \right)_{\{j,n\} \in \mathbf{Z}}$$

is an orthonormal basis of V_{2^j}.

The assumption shows that we can build any basis function on V_{2^j} by scaling a function $\phi(x)$ by 2^j and translating the resulting function on a grid whose interval is proportional to 2^{-j}. The $2^{-j/2}$ coefficient is a normalization factor for the scaling operation. For a given MRA $(V_{2^j})_{j \in Z}$, there exists a unique function $\phi(x)$ that satisfies

284

Assumption 1. Figure A.13 shows an example of a spline smoothing function which

oscillates in a symmetric exponentially decaying envelope. The Fourier transform

$\Phi(\omega)$ of $\phi(x)$ reveals the scaling function to be a low pass filter. Figure A.14 shows
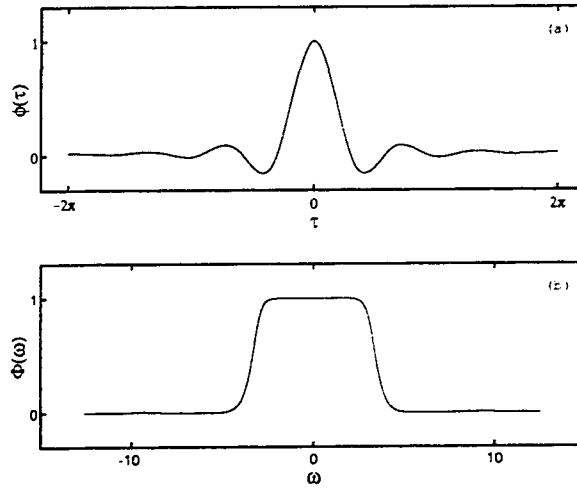
Figure A.13: A typical scaling function's (a) impulse response $o(\tau)$ and (b) frequency response $\Phi(\omega)$. The scaling function is a low pass filter.

a grid of dilations and translations of the scaling function $o(x)$ shown in Fig. A.13.

Each row can be seen as a subset of all the functions that form the orthonormal

basis in $(V_{2^j})_{j \in Z}$ for $j = -1, 0, 1$ as we descend the rows. As j decreases so does

the resolution of the scaling filter. This is apparent in the stretching of the scaling

function in time as the resolution decreases. It is this very dilation that allows the

scaling function to be used as a sort of mathematical microscope capable of zooming

in at short lived bursty behavior (high resolution) and long lived low frequency be-

havior (low resolution). Notice also that as the resolution decreases, the amplitude

of the scaling function envelope decreases as well. This is due to the fact that we

wish to keep the energy of the scaling function (which amounts to the area under

$(\phi_{j,n}(x))^2$) at a constant equal to one. In this sense, the $\mathbf{A}_{2^j}$ operator neither adds

nor diminishes the energy E_f of a signal $f(x)$. The orthogonal projection on $\mathbf{V}_{2^j}$ can
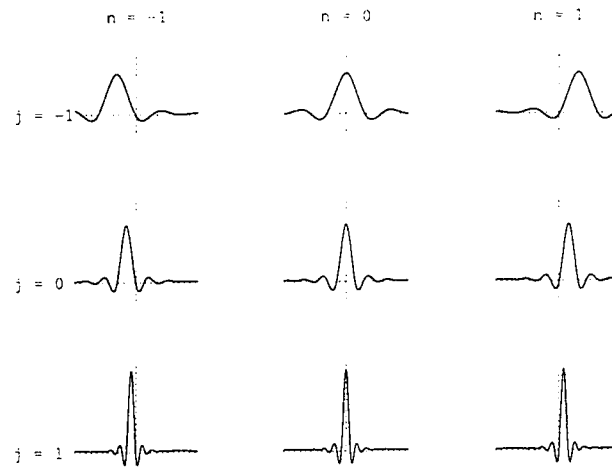
Figure A.14: Example of translations and dilations of a typical scaling function $\phi(x)$ by $\phi_{j,n}(x) = 2^{\frac{j}{2}}\phi(2^j x - n)$.

now be computed by decomposing a signal $f(x)$ on the orthonormal basis given by

Assumption 1. Specifically,

$$\forall f(x) \in \mathbf{L^2(R)},$$

$$\mathbf{A}_{2^j} f(x) = 2^{-j} \sum_{n=-\infty}^{\infty} \langle f(u), \phi_{j,n}(u) \rangle \, \phi_{j,n}(x)$$

where $\langle \cdot \rangle$ is the inner product operator. The discrete approximation of the signal $f(x)$ at the resolution 2^j, $\mathbf{A}_{2^j} f(x)$, is characterized by a set of inner products

$$\mathbf{A}_{2^j}^d f = \left(\langle f(u), \phi_{j,n}(u) \rangle \right)_{n \in Z}. \tag{A.17}$$

Each inner product can be interpreted as a convolution evaluated at point $2^{-j}n$. Thus, we can round off our analogy in recognizing that the camera itself is analogous to the smoothing function $\phi(u)$ and the dilation of the smoothing function serves as the lens. i.e. the camera $\phi(u)$ uses its lens $\phi_{j,n}(u)$ to focus in on a signal at a resolution 2^j. It has been stated that as resolution is increased, there is a natural decrease in the global view of the scenery. To capture the entire scene at finer resolutions, the cameraman must sweep the scene. This is mathematically equivalent to the convolution of the signal with a dilated version of the scaling function (Eq. A.17). In this manner, $\phi_{j,n}(u)$ absorbs information across the entire span of the signal and not just in the local frame.

To implement our discrete approximation operator we need to find a relation between the discrete approximation of $f(x)$ at resolution 2^j with that of a coarser resolution 2^{j+1}. Using the fact that $\left(\phi_{j+1,k}(x) \right)_{k \in Z}$ is an orthonormal basis of $\mathbf{V}_{2^{j+1}}$ and

that for any $n \in \mathbf{Z}$, the function $\phi_{j,n}(x)$ is a member of $\mathbf{V}_{2^j}$ which is included in $\mathbf{V}_{2^{j+1}}$, it can be shown that the relation between $\mathbf{A}_{2^j}^d f(x)$ and $\mathbf{A}_{2^{j+1}}^d f(x)$ is

$$\langle f(u), \phi_{j,n}(u) \rangle = \sum_{k=-\infty}^{\infty} \tilde{h}(2n-k) \langle f(u), \phi_{j+1,k}(u) \rangle .$$

Thus, $\mathbf{A}_{2^j}^d f(x)$ can be calculate from $\mathbf{A}_{2^{j+1}}^d f(x)$ by convolving $\mathbf{A}_{2^{j+1}}^d f(x)$ with the filter $\tilde{h}$ and keeping every other sample. Here $\tilde{H}$ is the mirror filter of H with impulse response $\tilde{h}(n)$ such that $\tilde{h}(n) = h(-n)$ where

$$\forall n \in \mathbf{Z}, \ h(n) = \langle \sqrt{2^{-1}} \phi_{1,0}(u), \phi_{0,n}(u). \rangle \tag{A.18}$$

In a similar manner to development of $\mathbf{A}_{2^j}^d f(x)$ above, the difference in information between the approximations at resolution 2^{j+1} and 2^j can be calculated by

$$\mathbf{W}_{2^j}^d f(u) = \left(\langle f(u), \psi_{j,n}(u) \rangle \right)_{n \in \mathbf{Z}}. \tag{A.19}$$

where $\mathbf{W}_{2^j}^d f(u)$ is a vector containing the *wavelet coefficients* at resolution 2^j and $\psi(x)$ is a *wavelet function*. It turns out that

$$\psi_{j,n}(x) = 2^{\frac{1}{2}} \psi(2^j x - n)$$

forms an orthonormal basis on the vector space $\mathbf{O}_{2^j}$ where $\mathbf{O}_{2^j}$ is the orthogonal complement of $\mathbf{V}_{2^j}$ in $\mathbf{V}_{2^{j+1}}$. Thus, it can be seen that the $\mathbf{W}_{2^j}^d f$ are created in exactly the same manner as the $\mathbf{A}_{2^j}^d f$. The difference is between the filter functions. Because the $\mathbf{W}_{2^j}^d f$ represent the difference in information between the approximate signals $\mathbf{A}_{2^j}^d f$ for consecutive resolutions, and because the frequency response of the

scaling function $\phi(x)$ is a low pass filter, it is intuitive that the frequency response of

the wavelet $\psi(x)$ is a band-pass filter. That is, the relatively high frequency content

lost between successive approximations comprise the detail signals.

With the same derivation steps as used to find the relation between $\mathbf{A}_{2^j}^d f(x)$ and

$\mathbf{A}_{2^{j+1}}^d f(x)$, it can be shown that $\mathbf{W}_{2^j}^d f$ can be calculated by convolving $\mathbf{A}_{2^{j+1}}^d f(x)$

with a discrete filter G whose form we will characterize.

$$\left\langle f(u), \psi_{j,n}(u) \right\rangle = \sum_{k=-\infty}^{\infty} \tilde{g}(2n - k)\left\langle f(u), \psi_{j+1,k}(u) \right\rangle.$$

where $\widetilde{G}$ is the mirror filter of G with impulse response $\tilde{g}(n)$ such that $\tilde{g}(n) = g(-n)$.

Here, $g(n)$ is discrete filter defined by

$$\forall n \in \mathbf{Z}, \; g(n) = \left\langle \sqrt{2^{-1}}\psi_{-1,0}(u), \phi_{0,n}(u). \right\rangle \tag{A.20}$$

Thus, $\mathbf{W}_{2^j}^d f(x)$ can be calculate from $\mathbf{A}_{2^{j+1}}^d f(x)$ by convolving $\mathbf{A}_{2^{j+1}}^d f(x)$ with the

filter $\tilde{g}$ and keeping every other sample.

At this stage, all of the components are in place to perform a multiresolution

signal decomposition of $f(x)$ into a series of approximate signals $\mathbf{A}_{2^j}^d f$ and wavelet

coefficient vectors $\mathbf{W}_{2^j}^d f$ for $-1 \geq j \geq J$ (Fig. A.15). The discrete wavelet anal-

ysis begins by splitting the original signal into two separate signals by convolving

$f(x) \rightarrow \mathbf{A}_1^d f(x)$ with the high and low pass filters $\tilde{g}$ and $\tilde{h}$, respectively. These

signals are then downsampled by a factor of 2 to form a new pair of signals de-

noted as $\mathbf{W}_{2^{-1}}^d f(x)$ and $\mathbf{A}_{2^{-1}}^d f(x)$, respectively. The wavelet coefficient vector and

approximate signal pair forms the *wavelet representation* at resolution $2^{-1} = \frac{1}{2}$. The
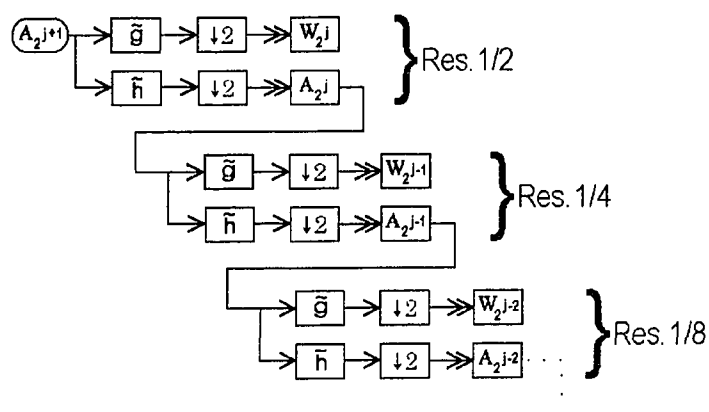
Figure A.15: Multiresolution signal decomposition algorithm.

algorithm is then reapplied to the approximated signal at resolution 2^{-1} to form the wavelet representation at resolution 2^{-2}, and so on. Figure A.16 shows the wavelet representation for a sawtooth signal in a 7-resolution split.
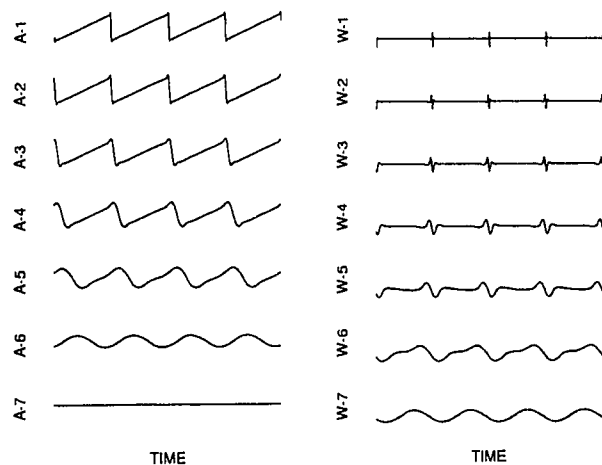
Figure A.16: The wavelet representation for a sawtooth signal in a 7-resolution split. The $\mathbf{A}_{2^j}^d f$ and $\mathbf{W}_{2^j}^d f$ are abbreviated in the graph as Aj and Wj, respectively, for $j < 0$. That is, A-3 is the approximate signal at resolution 2^{-3}. At a given resolution 2^j, the wavelet coefficient vectors Wj represent the difference in information between Aj and $A(j+1)$. Thus, W-1 represents the difference in information between A-1 and the original signal $A0$ (not shown). The program used to generate these graphs uses an algorithm similar the one shown in Figure A.15, known as the *Maximal Overlap Discrete Wavelet Transform (MODWT)*. The main difference between the MODWT and the DWT described above is that there is no decimation in time at every stage of the analysis algorithm. In this manner, temporal information is easier to line up in time across the resolutions. It is for this very reason that the MODWT was chosen over the standard DWT to display the $\mathbf{A}_{2^j}^d f$ and $\mathbf{W}_{2^j}^d f$. The concept behind both remains the same; thus the interpretation of the roles for $\mathbf{A}_{2^j}^d f$ and $\mathbf{W}_{2^j}^d f$ does not change.

In 1985. Stephane Mallat discovered the relationship between *quadrature mirror filters (QMF)* and orthonormal wavelet bases[50]. The filters $\tilde{g}$ and $\tilde{h}$ are quadrature mirror filters; so named because they comprise a pair of filters that have mirror image amplitude responses about a certain frequency, and maintain a phase difference of $\frac{\pi}{2}$ at all frequencies. Their mirror pairs, h and g, are used to reconstruct the signal $f(x)$. This process is referred to as the *synthesis operation*. To perform both a multiresolutional analysis and synthesis of a signal, four filters are needed. Once a suitable low pass filter $\tilde{h}$ is chosen, the remaining three filters are calculated by

$$
\begin{aligned}
h(n) &= \tilde{h}(-n) \\
g(n) &= (-1)^{1-n} h(1-n) \\
\tilde{g}(n) &= g(-n)
\end{aligned}
\tag{A.21}
$$

Figure A.17 shows the impulse responses for typical synthesis and analysis filters while Fig. A.18 shows the corresponding frequency responses. The analysis filters $\tilde{h}$ and $\tilde{g}$ are used to decompose the original signal into a multiresolutional analysis while the synthesis filters g and h are used to build the signal back up. The original signal can be reconstructed by building on the approximated signal at the coarsest resolution with the higher frequency content found in the wavelet coefficient vectors at finer resolutions. The original input signal can be fully reconstructed provided that all temporal portions of the signals in all the respective resolutions are used.

292
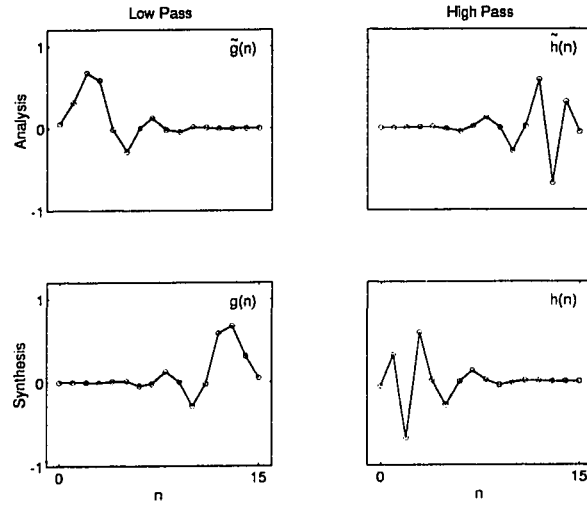
Figure A.17: The impulse responses for Daubechies' extremal phase 16 tap synthesis and analysis filters. Once $\tilde{h}$ has been specified, the remaining filters can be built via Eq. A.22. The filters in each row are quadrature mirror filters.
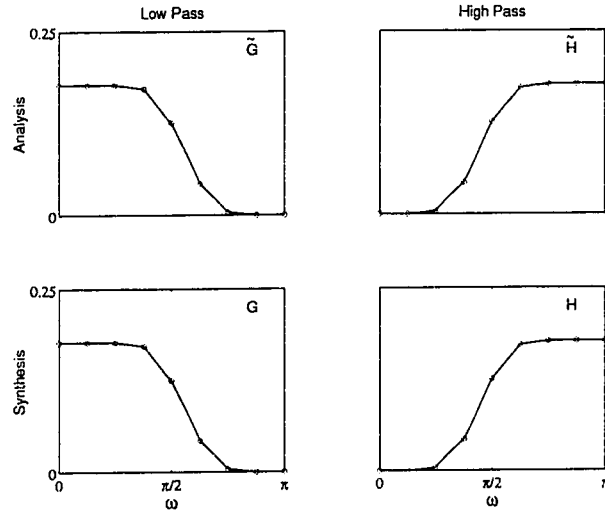
Figure A.18: The frequency responses for Daubechies' extremal phase 16 tap synthesis and analysis filters. The QMF filters have mirror image amplitude responses about a certain frequency; in this case at $\pi/2$.

The synthesis algorithm is shown in Fig. A.19. An alternative and equivalent means
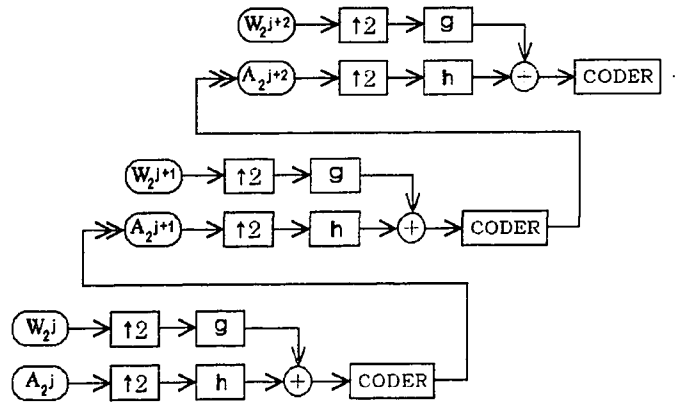
Figure A.19: Multiresolution signal synthesis.

of synthesis is to form an approximation of the original signals with each and every

wavelet coefficient vector by essentially zeroing out all the remaining energy found

in other resolutions. The resultant synthesis using only the energy of one $\mathbf{W}_{2^j}^d f$

is known as the *detail signal* from resolution 2^j and is denoted $\mathbf{D}_{2^j}^d f(x)$. These

signals mainly contain the band pass frequency information lost between resolutions;

hence the name *detail signal*. The lowest resolution approximate signal $\mathbf{A}_{2^{-J}}^d f(x)$

is often referred to as the *residue signal* since it contains the leftovers of the MRA.

Figure A.20 shows the coding scheme for a three band split. An advantage in using

this analysis/synthesis technique is that the resulting wavelet coefficient vectors and

the residue can be summed together directly to reconstruct the original signal. In

addition, at a glimpse one can isolate the effect of various features in the original

signal by panning across the scales at a given time. Using the zeroing technique
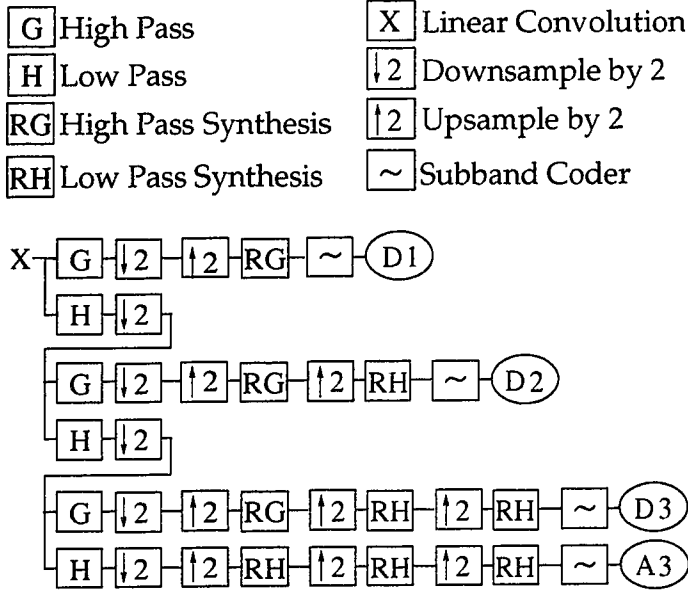
Figure A.20: The multiresolution wavelet coding scheme used to create the detail and residue signals. Three subband splits of the input X are performed. In this particular scheme, the bands $D1, D2, D3, A3$ are constructed by essentially zeroing out all the energy in the other bands. An advantage in using this synthesis technique is that the subbands can be summed together to reconstruct the original signal while doing a good job in preserving the temporal distribution of the spectral energy.

described above, the lowest resolution detail signal $\mathbf{D}_{2^{-J}}^d f(x)$ is added to the residue $\mathbf{A}_{2^{-J}}^d f(x)$ to form an approximation to the original signal. The resultant signal is denoted as a *wavelet smooth* at resolution 2^{-J+1}, or $\mathbb{R}_{2^{-J+1}}^d f(x)$. The $\mathbb{R}_{2^{-J+1}}^d f(x)$ is then added to the next highest resolution detail signal $\mathbf{D}_{2^{-J+1}}^d f(x)$ to form a better approximation $\mathbb{R}_{2^{-J+2}}^d f(x)$ to the original signal $f(x)$. It is clear that eventually the wavelet smooth will converge to the original signal; i.e. $\mathbb{R}_1^d f(x) \rightarrow f(x)$. Thus, the

principle of superposition holds in regard to the wavelet details and residue signal.

Figure A.21 demonstrates the gradual convergence of the wavelet synthesis for a

sawtooth waveform. The wavelet synthesis converges in a similar manner to the

Fourier synthesis (Compare Fig. A.2 and Fig. A.21).
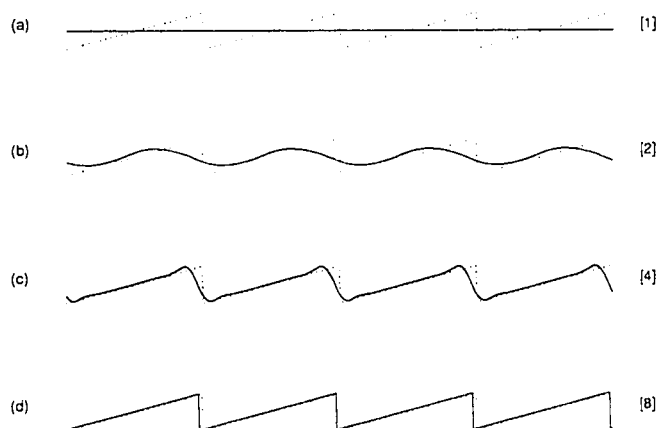
Figure A.21: The gradual reconstruction of a sawtooth wave broken down into a full scale multiresolution analysis. The numbers in brackets to the right of each plot indicates the level of synthesis. The first plot (a) is the residue $\mathbb{R}^d_{2^{j}-11}f(x)$ which is proportional to the mean of the signal. The remaining plots are the approximations $\mathbb{R}^d_{2^j}f(x)$ of the saw-tooth function $f(x)$ for resolutions (b) 2^{-8} (c) 2^{-6} and (d) 2^{-1}.

Now for a word on wavelet semantics. Because wavelets got their start from two

main scientific communities, often times there are different terms used to refer to the

same component in wavelet analysis. As an example, consider the wavelet coefficient

vectors $\mathbf{W}^d_{2^j}f$. The $\mathbf{W}^d_{2^j}f$ are orthogonal to one another and each is associated with

a distinct logarithmic frequency band or *octave*. Thus, the signals $\mathbf{W}^d_{2^j}f$ for $j \in \mathbf{Z}$

and the residue signal $\mathbf{A}_{2^{-J}}^d f$ are referred to as *subbands*. Still another interpretation

is to refer to the $\mathbf{W}_{2^j}^d f$'s and the $\mathbf{A}_{2^{-J}}^d f$ as *scales*. For example, the $\mathbf{W}_{2^{-k}}^d f$ signal

in a J-level split is either referred to as the $(J - k + 1)^{th}$ scale wavelet coefficients or

the $(J - k + 1)^{th}$ subband for $k = 1, 2, \ldots, J$. The *resolution* of a signal is linked to

its frequency content while *scale* refers to the topographical sense of the word; i.e.

scale here implies that of the scale of maps. It is often convenient to think of scale as

the size of the yardstick you are using to measure with. Scale and resolution are not

always commensurate. For example, low pass filtering a signal keeps its scale, but

reduces its resolution[81]. However, by the manner in which a MRA is implemented,

the scale and resolution of a signal *are* proportional and thus the phrases *scale* and

subband are used interchangeably in MRA.

BIBLIOGRAPHY

[1] H.D.I. Abarbanel, R. Brown, and J.B. Kadtke. Prediction in chaotic nonlinear systems: Methods for time series with broadband fourier spectra. *Physical Review A*, 41:1782–1807, 1989.

[2] A.M. Albano, J. Muench, C. Schwartz, A.I. Mees, and P.E. Rapp. Singular-value decomposition and the grassberger-procaccia algorithm. *Physical Review A*, 38:3017–26, 1988.

[3] M.A. Azhar and K. Gopala. 1/f fluctuations in beta- decay of 90y. *Physical Review A*, 44(2):1044–48, 1991.

[4] A. Babloyantz and A. Destexhe. Is the normal heart a periodic oscillator? *Biological Cybernetics*, 58:203–11, 1988.

[5] E. Bacry, J.F. Muzy, and A. Arneodo. Singularity spectrum of fractal signals from wavelet analysis: Exact results. *Journal of Statistical Physics*, 70:635–74, 1993.

[6] S. Baskaran, P.F. Stadler, and P. Schuster. Approximate scaling properties of rna free energy landscale. *Journal of Theoretical Biology*, 181:299–310, 1996.

[7] J.B. Bassingthwaighte, L.S. Liebovitch, and B.J. West. *Fractal Physiology*. Oxford University Press, New York, 1994.

[8] J.B. Bassingthwaighte and G.M. Raymond. Evaluation of the dispersional analysis method for fractal time series. *Ann. Biomed. Eng.*, 23:491–505, 1995.

[9] J.B. Bassingthwaighte and G.M. Raymond. *Signal analysis programs*. National Simulation Resource, National Simulation Resource, University of Washington, Seattle, WA 98195-7962, 1995–1997. http://nsr.bioeng.washington.edu/NSR/NSRinfo/avsoftware/NSR_SW_fractal.html.

[10] G.W. Baxter, R.P. Behringer, T. Fagert, and G.A. Johnson. Pattern formation in flowing sand. *Physical Review Letters*, 62:2825–8, 1989.

[11] T. Belloni and G. Hasinger. Variability in the noise properties of cygnus x-1. *Astronomical Astrophysics*, 227:L33–L36, 1990.

[12] L.F. Burlaga and M.L. Goldstein. Radial variations of large-scale magnetohydrodynamic fluctuations in the solar wind. *Journal of Geophysical Research*, 9:6813–17, 1984.

[13] T.D. Burton. *Introduction to Dynamic Systems Analysis*. McGraw Hill, Inc., Princeton Road. S-1, Hightstown, NJ, 1994.

[14] E.B. Cargill, H.H. Barrett, R.D. Fiete, M. Ker, D.D. Patton, and G.W. Seley. Fractal physiology and nuclear medicine scans. *Proceedings of the SPIE - The International Society for Optical Engineering*, 914:141–61, 1988.

[15] M.Y. Choi and H.Y. Lee. Traffic flow and 1/f fluctuations. *Physical Review E*, 52(6):5979–84, 1995.

[16] R. Coifman and M. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory*, 38(2):713–8, 1992.

[17] M. DishonBerkovits and R. Berkovits. Work-related tardiness: lateness incident distribution and long-range correlations. *Fractal*, 5(2):321–4, 1997.

[18] D.L. Donoho and I.M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81:425–55, 1994.

[19] J.P. Eckmann, S.O. Kamphorst, D. Ruelle, and S. Ciliberto. Liapunov exponents from time series. *Physical Review A*, 34(36):4971–9, 1986.

[20] R. Eglash. Inferring representation type from the fractal dimension of biological communication waveforms. *Journal of Social and Evolutionary Systems*, 16(4):375–99, 1993.

[21] M.S. Eisenberg. Defibrillation: the spark of life. *Scientific American*, 278(6):86–90, 1998.

300

[22] W. Feller. *An Introduction to Probability Theory and Its Applications*, volume 3. John Wiley & Sons, New York, 3 edition, 1968. pp. 67-97.

[23] G.E. Forsythe, M.A. Malcolm, and C.B. Moler. *Computer Methods for Mathematical Computations*. Prentice Hall, Englewood Cliffs, NJ, 1977. Section3.

[24] A.M. Fraser and H.L. Swinney. Independent coordinates for strange attractors from mutual information. *Physical Review A*, 33:1134-40, 1986.

[25] G.M. Friesen, T.C. Jannett, M.A. Jadallah, S.L. Yates, S.R. Quint, and H.T. Nagle. A comparison of the noise sensitivity of nine qrs detection algorithms. *IEEE Transactions on Biomedical Engineering*, 37(1):85-98, 1990.

[26] U. Frisch and G. Parisi. *Turbulence and Predictability in Geophysical Fluid Dynamics and Climate Dynamics*. NorthHolland, Amsterdam, 1985. Fully developed turbulence and intermittency and Appendix: On the singularity structure of fully developed turbulence. pp. 71-88.

[27] Y. Gagne, E.J. Hopfinger, and U. Frisch. *New Trends in Nonlinear Dynamics and PatternForming Phenomena: The Geometry of Nonequilibrium and A new universal scaling for fully developed turbulence: The distribution of velocity increments*. Plenum Press, New York, 1988. pp. 315-9.

[28] L. Glass, A.L. Goldberger, and J. Belair. Dynamics of pure parasystole. *Am J Physiol.* 251:H841–H847, 1986.

[29] L. Glass and D. Kaplan. Time series analysis of complex dynamics in physiology and medicine. *Medical Progress Through Technology*, 19:115–28, 1993.

[30] A.L. Goldberger, V. Bhargava, B.J. West, and A.J. Mandell. On the mechanism of cardiac electrical stability: The fractal hypothesis. *Biophys. Journal.* 48:525–8, 1985.

[31] M.L. Goldstein, L.F. Burlaga, and W.H. Matthaeus. Journal of geophysical research. *Power spectral signatures of interplanetary corotating and transient flows*, 89:3747–61, 1984.

[32] P. Grassberger and I. Procaccia. Measuring the strangeness of strange attractors. *Physica D,* 9:189–208, 1983.

[33] J.M Halley. Ecology, evolution and 1/f noise. *Tree,* 11(1):33–7, 1996.

[34] J.M. Hausdorff and C.K. Peng. Multiscaled randomness: a possible source of 1/f noise in biology. *Physical Review E,* 54(2):2154–55, 1996.

[35] H.G.E. Hentschel and I. Procaccia. The infinite number of generalized dimensions of fractals and strange attractors. *Physica D,* 8:435–44, 1983.

[36] D. Kaplan. Evaluating deterministic structure in maps deduced from discrete-time measurements. *International Journal of Bifurcation and Chaos in Applied Sciences and Engineering*, 3(3):617–23, 1993.

[37] D. Kaplan. Exceptional events as evidence for determinism. *Physica D*, 73(1–2):38–48, 1994.

[38] D. Kaplan, M. Furman, S. Pincus, S. Ryan, L. Lipsitz, and A. Goldberger. Aging and the complexity of cardiovascular dynamics. *Biophys. J.*, 59:945–949, 1991.

[39] D.T. Kaplan and R.J. Cohen. Fibrillation vs. random noise: A comparison using dimensionality calculations. In *IEEE Engineering in Medicine and Biology Society: 11th Annual International Conference*, pages 92–3, 1989.

[40] M.B. Kennel, R. Brown, and H.D.I. Abarbanel. Determining the dimension for phase-space reconstruction using geometrical construction. *Physical Review A*, 45(6):3403–11, 1992.

[41] L.B. Kiss, U. Klein, C.M. Muirhead, J. Smithyman, and Z. Gingl. Diffusive fluctuations, long-time and short-time cross-correlations in the motion of vortice-pancakes in different layers of ybco/pbco superlattices. *Solid State Communications*, 101, 191997.

[42] Donald E. Knuth. *The Art of Computer Programming Volume II*. Addison-Wesley Pub. Co., Reading, Mass., 1997.

[43] S.M. Kogan. 1/f noise in spin glasses and in the disordered kinetic ising model. *Solid State Communications*, 8:1015-8, 1981.

[44] G. Landini. Fractal music: the 1/f variable algorithm. Fractal report (newsletter), Reeves Telecommunications Laboratories, Ltd., 1987.

[45] A. Lawrence, M.G. Watson, K.A. Pounds, and M. Elvis. Low-frequency divergent x-ray variability in the seyfert galaxy ngc4051. *Nature*, 325:694-6, 1987.

[46] P.F. Lazorenko. Empirical image-motion spectrum. i. astronomical seeing and the atmospheric constraints on the accuracy of meridian observations. *Kinematics and Physics of Celestial Bodies*, 8(3):70–82, 1992.

[47] A. Lo and J. Haubrich. The sources and nature and long-range dependence in the business cycle. MIT working paper.

[48] E.N. Lorenz. Deterministic nonperiodic flow. *Journal of Atmospheric Science*, 20:130, 1963.

[49] K. Maeda, S. Sugita, H. Kurita, M. Uota, S. Uchida, M. Hinomaru, and

304

Y. Mera. Spatial variation of 1/f current noise in scanning tunneling microscopes. *Journal of Vacuum Science and Technology B*, 12(3):2140–3, 1994.

[50] S. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(7):674–93, 1989.

[51] S. Mallat and L.H. Hwang. Singularity detection and processing with wavelets. *IEEE Transactions on Information Theory*, 38:617–643, 1992.

[52] B. Mandelbrot. *The Fractal Geometry of Nature*. W.H. Freeman, San Francisco, CA, 1982.

[53] R.N. Mantegna and H.E. Stanley. Turbulence and financial markets. *Nature*, 383:587–8, 1996.

[54] W.H. Matthaeus and M.L. Goldstein. Low-frequency 1/f noise in the interplanetary magnetic field. *Physical Review Letters*, 57(4):495–8, 1986.

[55] J. McCauley. *Chaos, Dynamics and Fractals: an algorithmic approach to deterministic chaos*. Cambridge University Press, Cambridge, UK, 1993.

[56] I. McHardy and B. Czerny. Fractal x-ray time variability and spectral invariance of the seyfert galaxy ngc5506. *Nature*, 325:696–8, 1987.

[57] J.L. Melendez and J. Beck. The role of the insulator in determining 1/f noise in hgcdte integrating mis devices. *Journal of Electronic Materials*, 22:993–8, 1993.

[58] F.C. Moon. *Chaotic and Fractal Dynamics: An Introduction for Applied Scientists and Engineers*. John Wiley and Sons, Inc, 1992. Chaotic Elastica found on pp. 452.

[59] W.H. Munk, J. Gordon, and F. MacDonald. *The Rotation of the Earth*. Cambridge University Press, Cambridge, England, 1975.

[60] T. Musha. Power spectral density estimation of light intensity fluctuations in photon counting. *Japanese Journal of Applied Physics*, 26:2022–5, 1987.

[61] T. Musha, K. Sugita, and M. Kaneko. 1/f fluctuations in aqueous ionic solution. *Journal of the Physical Society of Japan*, 51(12):3820–5, 1982.

[62] J.F. Muzy, E. Bacry, and A. Arneodo. The multifractal formalism revisited with wavelets. *International Journal of Bifurcation and Chaos*, 4(2):245-302, 1994.

[63] R. Ma ne and F. Takens. Dynamic systems and turbulence. *Lecture Notes in Mathematics*, 898:366, 1980.

[64] M. Ocio. Observation of 1/f magnetic fluctuations in a spin glass. *Journal of Physique Letters*, 46:L647–L652, 1985.

[65] M. Ocio, H. Bouchiat, and M. Monod. Observation of 1/f magnetic fluctuations in spin glasses. *Journal of Magnetism and Magnetic Materials*, 11:54–7, 1986.

[66] A.V. Oppenheim and A.S. Willsky. *Signals and Systems*. Prentice Hall, Englewood Cliffs, NJ, 1983.

[67] A.R. Osborne and A. Provenzale. Finite correlation dimension for stochastic systems. *Physica D*, 35:357–81, 1989.

[68] C.K. Peng, J. Mietus, J.M. Hausdorff, S. Havlin, H.E. Stanley, and A.L. Goldberger. Long-range anticorrelations and non-gaussian behavior of the heartbeat. *Physical Review Letters*, 70(9):1343–6, 1993.

[69] G. Peng and H.J. Herrman. Density waves and 1/f density fluctuations in granular flow. *Physical Review E*, 51(3):1745–56, 1995.

[70] T.J.P. Penna, P.M.C. de Oliveira, J.C. Sartorelli, W.M. Goncalves, and R.D. Pinto. Long range anticorrelation and non-gaussian behavior of a leaky faucet. *Physical Review E*, 52:R2168–R2171, 1995.

[71] D.B. Percival. Simulating gaussian random processes with specified spectra. *Computing Science and Statistics*, 24:534–8, 1992.

[72] D.B. Percival, J.B. Bassingthwaighte, and G.M. Raymond. Approximation of fractional gaussian noise and related processes via spectral synthesis methods. 1998.

[73] D.B. Percival and A. Walden. *Wavelet Methods for Time Series Analysis*. Cambridge University Press, Cambridge, UK, 1999.

[74] S.L. Pimm and A. Redfearn. The variability of population densities. *Nature*, 334:613–4, 1988.

[75] S.M Pincus. Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci.*, 88:2297–2301, 1991.

[76] H. Poincaré. *The Foundation of Science: Science and Method*. The Science Press, New York. English translation.

[77] A. Quinquis. A few practical applications of wavelet packets. *Digital Signal Processing*, 8:49–60, 1998.

[78] P. Refregier, M. Ocio, and H. Bouchiat. Equilibrium magnetic fluctuations in spin glasses: temperature dependence and deviations from 1/f behaviour. *Europhysics Letters*, 3(4):503–10, 1987.

[79] W. Reim, R.H. Koch, A.P. Malozemoff, M.B. Ketchen, and H. Maletta. Mag-

308

netic equilibrium noise in spin-glasses: $eu_{0.4}sr_{0.6}s$. *Physical Review Letters*, 57(7):905–8, 1986.

[80] P. Reinhall and D. Storti. Phase-locked and chaotic behavior of coupled van der pol oscillators. To be submitted for publication, 1999.

[81] O. Rioul and M. Vetterli. Wavelets and signal processing. *IEEE SP Magazine*, pages 14–38, 1991.

[82] O.E. Rössler. *Physics Letters A*, 57(5):397–8, 1976.

[83] H. Rosu and E. Canessa. Solitons and 1/f noise in molecular chains. *Physical Review E*, 47(6):R3818–R3821, 1993.

[84] N. Saito and R. Coifman. Local discriminant bases and their applications. *Journal of Mathematical Imaging and Vision*, 5:337–58, 1995.

[85] G.V. Savino, L. Romanelli, D.L. Gonzalez, and et al. Evidence for chaotic behavior in driven ventricles. *Biophysical Journal*, 56:273–80, 1989.

[86] K.L. Schick and A.A. Vervear. 1/f noise with a low frequency white noise limit. *Nature*, 251:599–601, 1974.

[87] R.A. Schiebel. A model for 1/f noise in diffusion current based on surface

recombination velocity fluctuations and insulator trapping [hgcdte]. *IEEE Trans. Electron Dev.*, 41:768-78, 1994.

[88] J. Schuster. *Deterministic Chaos*. VCH Publishers, Inc., New York, 1988.

[89] Y. Shi. Correlations of pitches in music. *Fractal*, 4(4):547-53, 1996.

[90] M.G. Signorini, S. Cerutti, S. Guzzetti, and R. Parola. Non-linear dynamics of cardiovascular variability signals. *Methods of Information in Medicine*, 33:81--4, 1994.

[91] J.H. Steele. A comparison of terrestrial and marine ecological systems. *Nature*, 313(6001):355-8, 1985.

[92] F. Takens. Detecting strange attractors in turbulence. *Lecture Notes in Mathematics*, 898:366-81, 1981.

[93] J. Theiler. Efficient algorithm for estimating the correlation dimension from a set of discrete points. *Physical Review A*, 36(9):4456-62, 1987.

[94] J. Theiler. Estimating fractal dimension. *J. Op. Soc. Am.*, A7:1055-73, 1990.

[95] J. Theiler. Some comments on the correlation dimension of $1/f^\alpha$ noise. *Physics Letters A*, 155:480-93, 1991.

310

[96] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J.D. Farmer. Testing for nonlinearity in time series: the method of surrogate data. *Physica D*, 58:77–94, 1992.

[97] L. Trefethen and D. Bau. *Numerical Linear Algebra*. SIAM, Philadelphia,PA, 1997.

[98] W.Y. Tseng and J. Dugundji. Nonlinear vibrations of a buckled beam under harmonic excitation. *Journal of Applied Mechanics*, 38:467–76, 1971.

[99] G.R. Turner. Production of 1/f music series. Fractal report (newsletter). Reeves Telecommunications Laboratories, Ltd., 1987.

[100] Y. Ueda, H. Doumoto, and K. Nobumoto. An example of random oscillation in three-order self-restoring systems. In *Proceedings of the Electric and Electronic Communication Joint Meeting*, 1978.

[101] M. Usher, M. Stemmler, and Z. Olami. Dynamic pattern formation leads to 1/f noise in neural populations. *Physical Review Letters*, 74(2):329–9, 1995.

[102] B. Van der Pol and J. Van der Mark. Frequency demultiplication. *Nature*, 120(3019):363–4, 1927.

[103] A.T.A. Wood and G. Chan. Simulation of stationary gaussian processes in $[0, 1]^d$. *Journal of Computational and Graphical Statistics*, 3:409–32, 1996.

[104] A.M. Yaglom. Correlation theory of processes with random stationary nth increments. *American Mathematical Society Translations, Series 2*, 8:87–141, 1958.

[105] X. Zhang and G. Hu. 1/f noise in a two-lane highway traffic model. *Physical Review E*, 52(5):4664–8, 1995.

VITA

William Constantine

11737 Exeter Ave. NE, Seattle, WA 98125

wconstan@home.com

INTERESTS

Structural vibrations, system identification via digital signal and image processing, graphical user interface software development, mathematical modeling of biological systems

EDUCATION

University of Washington Seattle, WA

Mechanical Engineering, Ph.D. 1993–Present

- Developed novel wavelet based denoising technique: Maximum Overlap Deterministic Reduction Algorithm (MODRA), shown to be more effective than waveshrink for in-band contaminated chaotic sequences
- Examined convergence issues for nonlinear measures of colored noise processes
- Developed wavelet based electrocardiogram feature extraction program
- Assessed determinism, entropy fluctuation, fractal structure, and power law scaling in over 50 sick and healthy human electrocardiogram records
- Built cantilever beam experiment capable of exhibiting chaotic response for purposes of testing the MODRA denoising algorithm

Washington State University Pullman, WA

Mechanical Engineering, M.S., B.S. 1988–93

- Used wavelet techniques to analyze helicopter noise signals

EXPERIENCE

Seismic Safety Products Wenatchee, WA

Vibrations Consultant Present

- Developed earthquake simulation software to test actuated gas shutoff devices
- Consulted in redesign to meet new ASCE vibration standards

Miller Brewing Company   Milwaukee, WI
Vibrations Consultant 1995–Present
- Developed advanced vibration software: Dynamode©
- Software used to simulate shipping environments for purposes of reducing response amplitudes of transportation palettes, testing new stacking configurations, and minimizing shrink wrap

NASA Ames Research Center   Moffett Field, CA
Rotorcraft Signal Analyst 1992-93
- Designed wavelet based filter bank software to analyze rotorcraft acoustics
- Software used to isolate blade vortex interaction noise from helicopter noise field
- Software to be integrated in an adaptive control scheme to minimize the noise level in tilt rotorcraft

Univ. of Washington and Washington State Univ.   Seattle/Pullman, WA
Teaching Assistant 1993-99
- Introduction to Thermodynamics, Introduction to System Dynamics, Advanced Dynamics and Vibrations, Methods in Finite Element Analysis (two quarters)

Washington State Department of Fisheries   Olympia, WA
Engineering Aid III 1991
- Developed database and velocity profile software for flow probe analysis
- Measured power draw on pump, generator, and freezer systems

OTHER EXPERIENCE

HBI Office Interiors   Bellevue, WA
Office Installer 1990
- Assembled office cubicles and electrical hardware

Bainbridge Babies Daycare Center   Bainbridge Island, WA
Construction Worker 1989
- Designed and built barn, deck, greenhouse, and two story playhouse

Flightcraft Inc. Boeing Field: Seattle, WA
Line Service Technician 1988
- Fueled, stored, and chocked general aviation, and commercial aircraft

ACHIEVEMENTS

- Co-founder of ODIN Dynamics® - a division of Rotodyne®, Inc.: developers of advanced vibration and biomedical software
- Submitted #1 research proposal for NASA Graduate Studies Research Program in 1992
- Outstanding Paper by a Young Presenter Award for 125th Conference of the Journal of the Acoustical Society of America
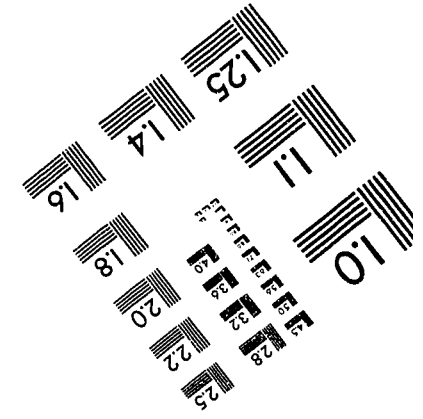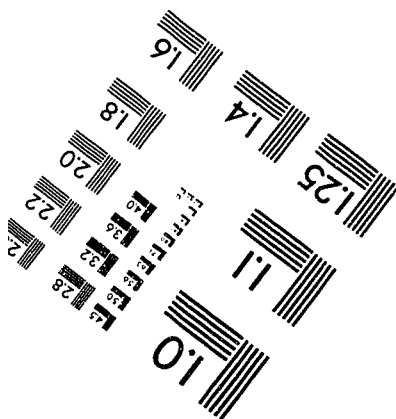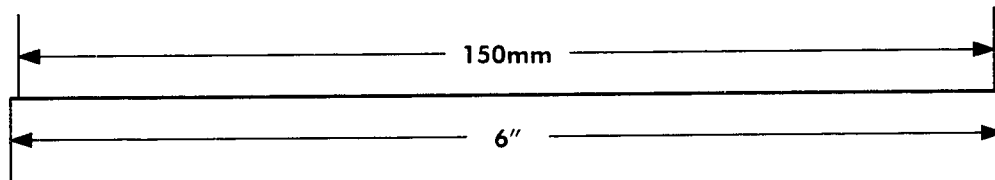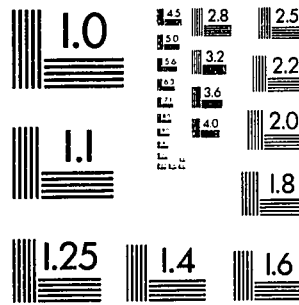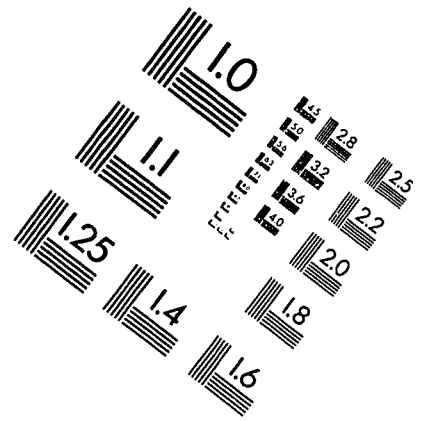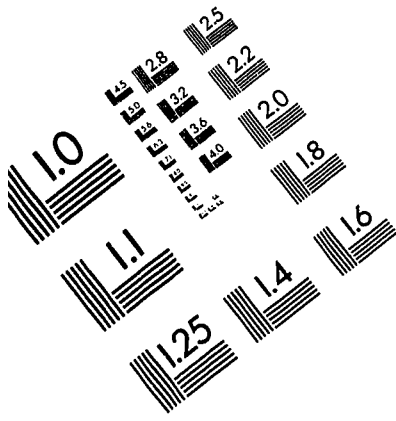- Obtained Private Pilot's License (SEL)

COMPUTER SKILLS

- Operating systems: HP UNIX, Linux, Windows NT 4.0, Windows 95/98
- Languages: C, Matlab, G, HTML, Interactive Data Language, FORTRAN
- Engineering software: ANSYS, LabVIEW, Matlab, PCTeX, LaTeX
- Other: Windows and HTML based help file development via (HTML) Help Workshop, software setup and installation development via InstallShield, C based Win32 DLL development for LabVIEW modules

PUBLICATIONS

- W. Constantine, C. Pezeshki, R. Bamberger, and M. Mosher, **Discrete Wavelet Analysis of Blade Vortex Interaction Noise**, *Journal of the Acoustical Society of America.* 97 (6), 3688-3693, Jun 1995
- **Wavelet Techniques for Chaotic and Fractal Dynamics**, Ph.D. dissertation
- W. Constantine, P. Reinhall, D. Percival, **On the Simulation Techniques and Fractal Dimension of 1/f $^{\square}$ Processes.** Submitted to *Physica D: Nonlinear Phenomena (preprints available)*
- W. Constantine, P. Reinhall, G. Bardy, and M. Gleva, **Arrhythmia Classification and Ventricular Fibrillation Prediction via Discrete Wavelet Packet Techniques.** To be submitted to *The Annals of Biomedical Engineering (in preparation)*
- W. Constantine, P. Reinhall, D. Percival, **In-band Denoising of Chaotic Sequences Using the Maximum Overlap Discrete Wavelet Packet Transform.** To be submitted to the *Journal of Bifurcation and Chaos (in preparation)*
- W. Constantine, P. Reinhall, G. Bardy, **Approximate Entropy Patterns for Heart Rate Variability Data.** To be submitted to the *Annals of Biomedical Engineering (in preparation)*

IMAGE EVALUATION
TEST TARGET (QA-3)

150mm

6"