

Synthesis, Analysis, and Processing of Fractal Signals

RLE Technical Report No. 566

Gregory W. Wornell

October 1991

**Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139-4307**

This work was supported in part by the Defense Advanced Research Projects Agency monitored by the Office of Naval Research under Grant N00014-89-J-1489, in part by the Air Force Office of Scientific Research under Grant AFOSR 91-0034, in part by the National Science Foundation under Grant MIP 87-14969, and in part by Lockheed Sanders, Inc.



SYNTHESIS, ANALYSIS, AND PROCESSING OF FRACTAL SIGNALS

by

GREGORY WAYNE WORNELL

Submitted to the Department of Electrical Engineering and Computer Science
on August 27, 1991, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Fractal geometry arises in a truly extraordinary range of natural and man-made phenomena. The $1/f$ family of fractal random processes, in particular, are appealing candidates for data modeling in a wide variety of signal processing scenarios involving such phenomena. In contrast to the well-studied family of ARMA processes, $1/f$ processes are typically characterized by persistent long-term correlation structure. However, the mathematical intractability of such processes has largely precluded their use in signal processing. We introduce and develop a powerful Karhunen-Loëve-like representation for $1/f$ processes in terms of orthonormal wavelet bases that considerably simplifies their analysis. Wavelet-based representations yield highly convenient synthesis and whitening filters for $1/f$ processes, and allow a number of fundamental detection and estimation problems involving $1/f$ processes to be readily solved. In particular, we obtain robust and computationally efficient algorithms for parameter and signal estimation with $1/f$ signals in noisy backgrounds, coherent detection in $1/f$ backgrounds, and optimal discrimination between $1/f$ signals. Results from a variety of simulations are presented to demonstrate the viability of the algorithms.

In contrast to the statistically self-similar $1/f$ processes, homogeneous signals are governed by deterministic self-similarity. Orthonormal wavelet bases play an equally important role in the representation of these signals, and, in fact, are used to construct orthonormal "self-similar" bases. The spectral and fractal characteristics of homogeneous signals make them appealing candidates for use in a number of applications. As one potential example, we consider the use of homogeneous signal sets in a communications-based context. In particular, we develop a strategy for embedding information into a homogeneous waveform on all time-scales. The result is a unique multirate modulation strategy that is well-suited for use with noisy channels of simultaneously unknown duration and bandwidth. Computationally efficient modulators and demodulators are developed for the scheme, and the results of a preliminary performance evaluation are presented. Although not yet a fully developed protocol, "fractal modulation" represents a novel and compelling paradigm for communication.

Thesis Supervisor: Alan V. Oppenheim
Title: Distinguished Professor of Electrical Engineering

Acknowledgments

First, I wish to express my most sincere gratitude to Prof. Alan Oppenheim, whose inspired mentoring, steadfast encouragement, and, above all, friendship, have meant a great deal to me. No single individual has had a greater impact either on my professional development in general or this thesis in particular.

I am also greatly indebted to my readers Prof. Alan Willsky and Prof. William Siebert for their involvement in the thesis. Numerous thought-provoking discussions with them have strongly influenced the way I think about the topic, and the thesis has benefited enormously as a result. I am also grateful to both Prof. Ehud Weinstein of Tel Aviv University and Prof. Sidney Burrus of Rice University for their interest, support and encouragement of the work.

Thanks, too, to my many friends and colleagues who have made graduate student life so stimulating and such a lot of fun. In particular, thanks to John Hardwick, Steve Isabelle, and Jim Preisig for plenty of assistance with various parts of the thesis as well as for episodes of hiking, mountain biking and gumbo production; to Jerry Shapiro, for all the technical brainstorming, and without whom my social life would have suffered drastically; and to Steve Smith, not only for being my athletic coordinator in hockey, squash, volleyball, and softball, but also for reminding me periodically that $1/f$ noise can exist outside a computer. More generally, for a whole host of bull sessions and camaraderie, thanks to DSPG cohorts Andy Singer and John Buck, my Brazilian buddies Henrique Malvar and Daniel Cobra, and my Canadian connection Frank Kschischang and David Asano.

I also wish to thank Prof. Leonard Gould for providing the Dow Jones data used in some of the experiments described in the thesis, as well as Dr. David Rigney and his colleagues at Beth Israel Hospital in Boston for providing several large sets of heart rate data that were used in additional experiments.

The generous financial support and encouragement of the Defense Advanced Research Projects Agency, the Air Force Office of Scientific Research, the National Science Foundation, Lockheed Sanders Inc., Raytheon Co., the Woods Hole Oceanographic Institution, and the Natural Sciences and Engineering Research Council (Canada) are all gratefully acknowledged.

Finally, special thanks to my parents and my brother, Jonathan, for their unwavering love and support. And my deepest thanks to Kimberly for her boundless love and patience, and for brightening my life immeasurably. By agreeing to marry me, she has given me a lifetime of happiness to which to look forward.

Contents

1	Introduction	11
1.1	Outline of the Thesis	13
2	Wavelet Transformations	16
2.1	Wavelet Bases	18
2.2	Orthonormal Wavelet Bases	19
2.2.1	An Octave-Band Filter Bank Interpretation	22
2.2.2	Multiresolution Signal Analysis Interpretation	27
2.2.3	Discrete Wavelet Transform	34
2.2.4	Finite Data Length and Resolution Effects	35
2.2.5	Orthonormal Wavelet Basis Constructions	39
2.2.6	Examples	41
2.2.7	Non-Dyadic Orthonormal Wavelet Bases	42
3	Statistically Self-Similar Signals	44
3.1	1/ f Processes	46
3.1.1	Fractional Brownian Motion and Fractional Gaussian Noise . .	51
3.1.2	A New Mathematical Characterization of 1/ f Processes	57
3.2	Nearly-1/ f Processes	60
3.2.1	ARMA Models	60
3.2.2	Wavelet-Based Models	65
4	Detection and Estimation with 1/f Processes	80
4.1	1/ f Synthesis and Whitening Filters	83
4.2	Parameter Estimation for 1/ f Signals	86
4.2.1	Case I: $\beta, \sigma^2, \sigma_w^2$ Unknown	91
4.2.2	Case II: β, σ^2 Unknown; σ_w^2 Known	93
4.2.3	Case III: β, σ^2 Unknown; $\sigma_w^2 = 0$	94
4.2.4	Properties of the Estimators	95
4.2.5	Simulations	98
4.3	Smoothing of 1/ f Signals	101
4.3.1	Simulations	108
4.4	Coherent Detection in 1/ f Noise	110
4.5	Discriminating Between 1/ f Signals	117
4.5.1	Simulations	121

4.6 Outstanding Issues	129
5 Deterministically Self-Similar Signals	131
5.1 Energy-Dominated Homogeneous Signals	133
5.2 Power-Dominated Homogeneous Signals	142
5.3 Discrete-Time Algorithms for Processing Homogeneous Signals	148
6 Fractal Modulation	155
6.1 Transmitter Design: Modulation	160
6.2 Receiver Design: Demodulation	167
6.2.1 Minimum Mean-Square Error Demodulation	168
6.2.2 Minimum Probability-of-Error Demodulation	173
6.3 Outstanding Issues	179
7 Linear Self-Similar Systems	184
7.1 Linear Time-Invariant Systems	185
7.2 Linear Scale-Invariant Systems	186
7.2.1 Generalized Linear Scale-Invariant Systems	191
7.3 Linear Time- and Scale-Invariant Systems	192
7.3.1 Self-Similar Signals and LTSI(λ) Systems	194
7.4 Wavelet-Based LTSI(λ) Systems	196
7.4.1 Dyadic Approximations to LTSI(λ) Systems Based on Ortho-normal Wavelet Bases	200
8 Conclusions and Future Directions	205
8.1 Future Directions	207
A Derivation of the Discrete Wavelet Transform	209
A.1 Analysis Algorithm	209
A.2 Synthesis Algorithm	210
B Proofs for Chapter 3	212
B.1 Proof of Theorem 3.2	212
B.2 Proof of Theorem 3.3	215
B.3 Proof of Theorem 3.4	216
B.4 Proof of Proposition 3.5	218
B.5 Proof of Theorem 3.6	220
C The EM Parameter Estimation Algorithm	222
D Proofs for Chapter 5	226
D.1 Proof of Theorem 5.2	226
D.2 Proof of Theorem 5.3	228
D.3 Proof of Theorem 5.6	229
D.4 Proof of Theorem 5.7	231

List of Figures

2-1	Critically-sampled filter bank interpretation of an orthonormal wavelet decomposition.	23
2-2	The octave band filters corresponding to an orthonormal wavelet decomposition. The wavelet basis in this example is one due to Daubechies [1].	24
2-3	Time-frequency portraits corresponding to two signal analyses.	25
2-4	Interpretation of an orthonormal wavelet expansion as a multirate modulation scheme.	26
2-5	A single stage of discrete-wavelet transform algorithm.	36
2-6	An efficient implementation of the orthonormal wavelet transformation based on the discrete wavelet transform.	37
3-1	Sample paths of $1/f$ processes corresponding to different values of γ	48
3-2	Synthesis of fractional Brownian motion $x(t)$ in terms of fractional Gaussian noise $x'(t)$ and stationary white Gaussian noise $w(t)$	56
3-3	The time-averaged spectrum of a process synthesized from the first-order Battle-Lemarie orthonormal wavelet basis. The parameters of the nearly- $1/f$ spectrum are $\gamma = 1$ and $\sigma_U^2/\sigma_L^2 = 1.103$ in this case.	67
3-4	Along-scale correlation between wavelet coefficients for an exactly- $1/f$ process for which $\gamma = 1$. The squares \square indicate a numerical estimate of the exact magnitude of the normalized correlation between wavelet coefficients as a function of the lag l between them. The ideal bandpass wavelet was assumed in the analysis. The triangles \triangle indicate the corresponding values of the closed-form bound obtained in the text. The circles \circ show the average sample-correlation as computed from a projections of a $1/f$ process generated using Keshner's synthesis onto a 5th-order Daubechies wavelet basis.	74
3-5	Weekly Dow Jones Industrial Average data, to present.	76
3-6	Wavelet-based analysis of weekly Dow Jones Industrial Average data. The time-series is analyzed using a 5th-order Daubechies wavelet basis.	77
3-7	Heartbeat interarrival times for a healthy patient.	78
3-8	Wavelet-based analysis of the heartbeat interarrival times for a healthy patient. The time-series is analyzed using a 5th-order Daubechies wavelet basis.	79

4-1	<i>Canonical form realizations of synthesis and whitening filters for processes that are the superposition of $1/f$-type and white components, i.e., $1/f$-plus-white processes.</i>	87
4-2	<i>RMS Errors in the estimates of the signal parameters as a function of the SNR of the observations. The symbols associated with each γ mark the actual empirical measurements; dashed lines are provided as visual aides only.</i>	99
4-2	<i>Continued.</i>	100
4-3	<i>RMS Errors in the estimates of the signal parameters as a function of the data length N of the observations. Again, the symbols associated with each γ mark the actual empirical measurements; dashed lines are provided as visual aides only.</i>	102
4-3	<i>Continued.</i>	103
4-4	<i>Tracking the time-varying spectral exponent γ of a noise-free $1/f$-type signal. For the left half of the signal, $\gamma = 0.90$, while for right half, $\gamma = 1.10$.</i>	104
4-5	<i>A canonic form implementation of the optimal linear filter for estimating a $1/f$ signal $x(t)$ from noisy observations $r(t)$. The linear least-squares filter is the cascade of a whitening filter followed by an innovations filter. The intermediate innovations process $v(t)$ is stationary and white.</i>	107
4-6	<i>SNR gain (dB) of the signal estimate as a function of the SNR of the observations. Both the gains predicted by eq. (4.28) and gains actually obtained are indicated.</i>	109
4-7	<i>Optimal smoothing of a noisy $1/f$ signal.</i>	111
4-8	<i>Canonical prewhitening implementation of the optimal receiver for detection of a known $s(t)$ in the presence of both Gaussian $1/f$ and stationary white Gaussian noise, where $\kappa_w(t, \tau)$ is the kernel of the whitening filter for $1/f$-plus-white noise.</i>	114
4-9	<i>The receiver operating characteristic for the detection of a known signal in a background of $1/f$ plus white noise for various thresholds determined via the performance index d.</i>	116
4-10	<i>A canonical form implementation of the optimal receiver for discriminating between $1/f$ models with distinct parameters based on noisy observations $r(t)$.</i>	119
4-11	<i>Optimal discriminator performance as a function of SNR, as estimated via the Chernoff bound.</i>	122
4-12	<i>Optimal discriminator performance as a function of the number of samples N of noisy observations, as estimated by the Chernoff bound. The symbols \square, Δ and \diamond correspond to actual estimates; the lines are provided as visual aides only in this case.</i>	124
4-13	<i>Optimal discriminator performance as a function of the parameter separation $\Delta\gamma$ between the two hypotheses, as estimated via the Chernoff bound.</i>	125

4-14	<i>Optimal discriminator performance as a function of SNR, as estimated via both the Chernoff-bound (4.36) and the CLT-based approximation (4.37).</i>	126
4-15	<i>Optimal discriminator performance as a function of the number of samples of observed data, as estimated via both the Chernoff-bound (4.36) and the CLT-based approximation (4.37). The Δ symbols correspond to actual estimates; the lines are provided as visual aides only.</i>	127
4-16	<i>Optimal discriminator performance as a function of the parameter separation $\Delta\gamma$, as estimated via both the Chernoff-bound (4.36) and the CLT-based approximation (4.37).</i>	128
5-1	<i>The self-similar basis functions $\theta_4^H(t)$, $\theta_5^H(t)$, $\theta_6^H(t)$, and $\theta_7^H(t)$ of an orthonormal basis for E^H, $H = 0$.</i>	141
5-2	<i>Comparison between the sample path of a $1/f$ process and a power-dominated homogeneous signal. Both correspond to $\gamma = 1$ (i.e., $H = 0$).</i>	146
5-3	<i>The time-frequency portrait of a homogeneous signal of degree $H = -1/2$.</i>	149
5-4	<i>The discrete-time self-similarity identity for a characteristic sequence $p[n]$.</i>	151
5-5	<i>Iterative algorithm for the synthesis of the characteristic sequence $p[n]$ of a homogeneous signal $x(t)$ from its generating sequence $q[n]$. The notation $p^{[i]}[n]$ denotes the value of $p[n]$ at the ith iteration.</i>	153
6-1	<i>A communication system for transmitting a continuous- or discrete-amplitude data sequence $q[n]$ over a noisy and unreliable continuous-amplitude, continuous-time channel.</i>	156
6-2	<i>The channel model for a typical communications scenario.</i>	156
6-3	<i>Spectral efficiency of fractal modulation. At each bandwidth B, the solid curve indicates the maximum rate at which transmitted data can be recovered. The dashed curve, indicating the corresponding rate for UMSE modulation, represents a performance bound.</i>	162
6-4	<i>A portion of the time-frequency portrait of the transmitted signal for fractal modulation of a finite-length data vector q. The case $H = -1/2$ is shown for convenience.</i>	165
6-5	<i>Error-rate-bandwidth tradeoffs for fractal modulation in noise with the optimum receiver. The solid lines represent the performance of fractal modulation, while the dashed lines corresponds to the performance of UMSE modulation with repetition coding.</i>	174
6-5	<i>Continued.</i>	175
6-6	<i>Bit error probabilities for bit-by-bit signaling with fractal modulation over noisy channels with the optimum receiver. Solid lines indicate the performance of fractal modulation, while dashed lines indicate the performance of UMSE modulation with repetition coding.</i>	180
6-6	<i>Continued.</i>	181

7-1	<i>Wavelet-based implementation of an LTSI(λ) system.</i>	199
7-2	<i>A dyadic approximation of a LTSI(λ) system as implemented via an orthonormal wavelet basis.</i>	202

Chapter 1

Introduction

There are a wide range of contexts in which there is a need to be able to synthesize, analyze and process fractal signals. Indeed, fractal geometry abounds in nature. Fractal structure can be found, for example, in natural landscapes, in the distribution of earthquakes, in ocean waves, in turbulent flow, in the pattern of errors on communication channels, in the bronchi of the human lung, and even in fluctuations of the stock market. In many applications, we are interested in modeling the inherent fractal behavior in order that we might perform some form of signal processing. For example, there are many problems of detection, classification, smoothing, and prediction that involve fractal signals. Likewise, that fractal behavior is so prevalent suggests that fractal geometry is somehow optimal or efficient. Consequently, there is increasing interest in the design of communication, telemetry, and other engineering systems based on the use of fractal signals.

This thesis is about the development and exploitation of a framework for representing and characterizing fractal signals. But what is a fractal signal? Most generally, a fractal signal is a function possessing structure at every scale of detail. However, the fractals of most interest, and those to which we restrict our attention in this thesis, are those for which the detail at each scale is similar. In this case we say that the fractal is self-similar or, alternatively, scale-invariant, reflecting the fact that the signal has no absolute scale of reference. Fractal signals may be classified into one of two broad categories: those in which the self-similarity is statistical, and those in

which it is deterministic. For statistically self-similar signals, the detail at all scales have the same statistics, while for deterministically self-similar signals the detail at all scales is identical.

Various representations for self-similar signals can be found in the literature; however, none have been particularly suitable for engineering applications either for reasons of mathematical intractability or computationally complexity. In this thesis, we introduce and develop highly efficient representations for some important classes of fractal signals based on the use of orthonormal wavelet bases. Orthonormal wavelet bases, having the property that all basis functions are dilations and translations of some prototype function, are in many respects ideally suited for use with self-similar signals. In fact, as will become apparent in the ensuing chapters, wavelet-based representations are as convenient and natural for self-similar signals as Fourier-based representations are for stationary and periodic signals. Furthermore, because wavelet transformations can be implemented in a computationally efficient manner, the wavelet transform is not only a theoretically important tool, but a practical one as well.

We specifically consider two families of self-similar signals. The first is the family of $1/f$ processes. These statistically self-similar processes, specifically, are important candidates for modeling a wide range of natural and man-made phenomena. Due to their generally nonstationary character, $1/f$ processes have properties that are rather distinct from the traditional models used in signal processing. In contrast to the well-studied family of ARMA processes, for example, $1/f$ processes typically exhibit long-term statistical dependence. Yet despite their apparent applicability in many contexts, $1/f$ models have not enjoyed widespread use in the signal processing community. In large part, this has been due to the lack of a sufficiently convenient mathematical characterization. The introduction of wavelet-based representations for $1/f$ processes in this thesis allows us to address a wide range of signal modeling and signal processing problems involving $1/f$ processes in a highly straightforward manner. In particular, we are able to obtain computationally efficient algorithms both for classifying, parametrizing, and isolating $1/f$ signals. Furthermore, in contrast to

previous algorithms, those we develop are robust with respect to both measurement noise and modeling errors.

The second family of self-similar signals we consider are *homogeneous* signals, which we characterize in terms of a novel deterministic self-similarity relation. Homogeneous signals, too, have highly efficient wavelet-based representations, and are potentially useful in a wide range of engineering applications. As an example of one promising direction for applications, we consider the use of homogeneous signal sets in a communications-based context. Specifically, we develop an approach for embedding information into homogeneous waveforms which we term “fractal modulation.” Because the resulting waveforms have the property that the information can be recovered with either arbitrarily little duration or arbitrarily little bandwidth, we are able to show that such signals are well-suited for transmission over noisy channels of simultaneously unknown duration and bandwidth. Not only is this a reasonable model for many physical channels, but also of the receiver constraints inherent in many point-to-point and broadcast communication scenarios. As a consequence of its special fractal and spectral properties, fractal modulation is potentially useful in a range of military and commercial communication contexts. Indeed, the concepts underlying fractal modulation may ultimately lead to novel and important approaches for both low probability of intercept and shared-spectrum communications.

1.1 Outline of the Thesis

The detailed structure of the thesis is as follows. Chapter 2 is a review of wavelet theory. In addition to establishing notation and summarizing the important results, this review provides a particular perspective on wavelets and their relationship to signal processing that is central to the thesis. Orthonormal wavelet basis signal decompositions are interpreted first in terms of an octave-band filter bank structure that is familiar to signal processors, and then in terms of a multiresolution signal analysis from which new insights are obtained. In particular, we show how this interpretation leads naturally to the computationally efficient discrete-time implementation via the

discrete wavelet transform.

Chapter 3 reviews the $1/f$ -family of statistically self-similar random processes, and develops some important new models for $1/f$ -like behavior in signals. In particular, early in the chapter, we introduce a novel and useful frequency-based characterization for $1/f$ processes, while in the latter half of the chapter we develop wavelet-based representations for $1/f$ processes. Specifically, we demonstrate that orthonormal wavelet basis expansions are Karhunen-Lo  e-like expansions for $1/f$ -type processes, *i.e.*, when $1/f$ processes are expanded in terms of orthonormal wavelet bases, the coefficients of the expansion are effectively uncorrelated. This powerful result is supported both theoretically and empirically, and examples involving both simulated and real data are included.

Exploiting the efficiency of wavelet basis expansions for $1/f$ processes, Chapter 4 develops solutions to some fundamental problems of detection and estimation involving $1/f$ -type signals. In particular, we develop both maximum likelihood parameter estimation algorithms and minimum mean-square error signal estimation algorithms for $1/f$ processes embedded in white measurement noise. Additionally, we address the problem of coherent detection in $1/f$ backgrounds, as well as the problem of discriminating between $1/f$ signals with different parameters. In each case, we provide useful interpretations of the solutions to these problems in terms of wavelet-based synthesis and whitening filters for $1/f$ processes. Results from a variety of simulations are presented.

Chapter 5 introduces and develops our new family of homogeneous signals defined in terms of a dyadic scale-invariance property. We distinguish between two classes: energy-dominated and power-dominated, and develop their spectral properties. We show that orthonormal self-similar bases can be constructed for homogeneous signals using wavelets. Using these representations, we then derive highly efficient discrete-time algorithms for synthesizing and analyzing homogeneous signals.

Chapter 6 develops the concept of fractal modulation. In particular, we use the orthonormal self-similar basis expansions derived in Chapter 5 to develop an approach for modulating discrete- or continuous-valued information sequences onto homoge-

neous signals. After developing the corresponding optimal receivers, we evaluate the performance of the resulting scheme in the context of a particular channel model. Our analysis includes comparisons to more traditional forms of modulation.

Chapter 7 represents a rather preliminary and cursory investigation into the system theoretic foundations of the thesis. In particular, after defining scale-invariant systems, we explore the relationships between such systems, self-similar signals, and the wavelet transform. We observe that synthesis filters for the self-similar signals we consider in the thesis are exactly or approximately linear jointly time- and scale-invariant systems. Furthermore, we demonstrate that while the Laplace and Fourier representations are natural for linear time-invariant systems, and while the Mellin representation is natural for linear scale-invariant systems, it is the wavelet transform that is most natural for linear systems that are jointly time- and scale-invariant. We show, in fact, that wavelet representations lead to some very efficient and practical computational structures for characterizing and implementing such systems.

Finally, Chapter 8 summarizes the principal contributions of the thesis and suggests some interesting and potentially important directions for future research.

Chapter 2

Wavelet Transformations

Wavelet transformations play a central role in the study of self-similar signals and systems. Indeed, as we shall see, the wavelet transform constitutes as natural a tool for the manipulation of self-similar or scale-invariant signals as the Fourier transform does for translation-invariant signals such as stationary, cyclostationary, and periodic signals. Furthermore, just as the discovery of fast Fourier transform (FFT) algorithms dramatically increased the viability the Fourier-based processing of translation-invariant signals in real systems, the existence of fast discrete wavelet transform (DWT) algorithms for implementing wavelet transformations means that wavelet-based representations of self-similar signals are also of great practical significance.

The theory of wavelet transformations dates back to the work of Grossmann and Morlet [2], and was motivated by applications in seismic data analysis [3]. Many key results in the theory of nonorthogonal wavelet expansions are described by Daubechies in [4]. In this thesis, however, we shall be primarily interested in *orthonormal* wavelet bases. The development of such bases, and their interpretation in the context of multiresolution signal analysis, is generally attributed to Meyer [5] and Mallat [6]. However, it was Daubechies who introduced the first highly practical families of orthonormal wavelet bases in her landmark paper [1].

Yet although wavelet theory is rather new, it is important to note at the outset that many of the ideas underlying wavelets are not new. Indeed, wavelet theory can

be viewed as a convenient and useful mathematical framework for formalizing and relating some well-established methodologies from a number of diverse areas within mathematics, physics, and engineering. Examples include:

- pyramidal image decompositions in computer vision [7],
- multigrid methods in the solution of partial-differential and integral equations [8],
- spectrogram methods in speech recognition [9],
- progressive transmission algorithms and embedded coding in communications [10] [11], and
- multirate filtering algorithms in digital audio [12], speech and image coding [13], voice scrambling [12], and frequency division data multiplexing [14].

In fact, wavelet transformations are closely associated with a number of topics that have been extensively explored in the signal processing literature in particular, including constant- Q filter banks and time-frequency analysis [15], and quadrature mirror and conjugate quadrature filter banks [12].

This chapter is designed as a self-contained overview of wavelet transformations in general and of orthonormal wavelet transformations in particular. Although it presents essentially no new results, it serves three main purposes. First, it establishes the notational conventions for wavelets we adopt for the thesis. Second, it summarizes the key results from wavelet theory we shall exploit in the applications in subsequent chapters of the thesis. However, the third purpose of the chapter is to introduce wavelet transformations from a signal processing perspective, and it is this objective which has lead to the rather tutorial style of this chapter. While a number of excellent introductions to wavelet theory tutorials can be found in the literature—see, *e.g.*, [6] [4] [16] [17]—we stress that the one presented here emphasizes a perspective that is particularly important in light of the applications we consider in this thesis.

2.1 Wavelet Bases

Most generally, the wavelet transformation of a signal $x(t)$

$$x(t) \longleftrightarrow X_\nu^\mu$$

is defined in terms of projections of $x(t)$ onto a family of functions that are all normalized dilations and translations of a prototype “wavelet” function $\psi(t)$, *i.e.*,

$$\mathcal{W}\{x(t)\} = X_\nu^\mu = \int_{-\infty}^{\infty} x(t) \psi_\nu^\mu(t) dt \quad (2.1)$$

where

$$\psi_\nu^\mu(t) = |\mu|^{-1/2} \psi\left(\frac{t-\nu}{\mu}\right).$$

In this notation, μ and ν are the continuous dilation and translation parameters, respectively, and take values in the range $-\infty < \mu, \nu < \infty$, $\mu \neq 0$. A necessary and sufficient condition for this transformation to be invertible is that $\psi(t)$ satisfy the *admissibility condition*

$$\int_{-\infty}^{\infty} |\Psi(\omega)|^2 |\omega|^{-1} d\omega = C_\psi < \infty, \quad (2.2)$$

where $\Psi(\omega)$ is the wavelet’s Fourier transform. Provided $\psi(t)$ has reasonable decay at infinity, (2.2) is equivalent to the admissibility condition

$$\int_{-\infty}^{\infty} \psi(t) dt = 0. \quad (2.3a)$$

For any admissible $\psi(t)$, the synthesis formula corresponding to the analysis formula (2.1) is then

$$x(t) = \mathcal{W}^{-1}\{X_\nu^\mu\} = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} X_\nu^\mu \psi_\nu^\mu(t) \mu^{-2} d\mu d\nu. \quad (2.4)$$

Under certain circumstances, it is also possible to reconstruct $x(t)$ solely from

samples of K_ν^μ on some lattice defined by

$$\begin{aligned}\mu &= a^{-m} \\ \nu &= n b a^{-m}\end{aligned}$$

where $-\infty < m < \infty$ and $-\infty < n < \infty$ are the integer dilation and translation indices, respectively, and a and b are the corresponding dilation and translation increments. In such cases, the collection of samples is termed a “frame.” A general theory and some iterative reconstruction algorithms are presented in [4]. However, it is also possible to construct wavelets and lattices such that the resulting transformation is not only invertible, but *orthonormal* as well. In general, orthonormal transformations are extremely convenient analytically, and possess very nice numerical properties. Consequently, it is this class of wavelet transformations that is of primary interest in this work, and the theory is summarized in the sequel.

2.2 Orthonormal Wavelet Bases

Our focus in this section is on the particular case of *dyadic* orthonormal wavelet bases, corresponding to the case $a = 2$ and $b = 1$ for which the theory is comparatively better developed. In Section 2.2.7, however, we construct a simple family of orthonormal wavelet bases corresponding to lattices defined by $a = (L + 1)/L$ and $b = L$ where $L \geq 1$ is an integer.

An orthonormal wavelet transformation of a signal $x(t)$

$$x(t) \longleftrightarrow x_n^m$$

can be described in terms of the synthesis/analysis equations

$$x(t) = \mathcal{W}_d^{-1} \{x_n^m\} = \sum_m \sum_n x_n^m \psi_n^m(t) \quad (2.5a)$$

$$x_n^m = \mathcal{W}_d \{x(t)\} = \int_{-\infty}^{\infty} x(t) \psi_n^m(t) dt \quad (2.5b)$$

and has the special property that the orthogonal basis functions are all dilations and translations of a single function referred to as the *basic wavelet* $\psi(t)$. In particular,

$$\psi_n^m(t) = 2^{m/2} \psi(2^m t - n) \quad (2.6)$$

where m and n are the dilation and translation indices, respectively.

An important example of a wavelet basis, and one to which we will refer on numerous occasions throughout the thesis, is that derived from the ideal bandpass wavelet $\tilde{\psi}(t)$. This wavelet is the impulse response of an ideal bandpass filter with frequency response

$$\tilde{\Psi}(\omega) = \begin{cases} 1 & \pi < |\omega| \leq 2\pi \\ 0 & \text{otherwise} \end{cases} \quad (2.7)$$

It is straightforward to verify that the dilations and translations of $\tilde{\psi}(t)$ constitute an orthonormal basis for the space of finite energy functions, $L^2(\mathbf{R})$. However, there are many other examples of orthonormal wavelet bases.

The basic (or “mother”) wavelet, $\psi(t)$ typically has a Fourier transform $\Psi(\omega)$ that satisfies several more general properties. First, because, for a fixed m the $\{\psi_n^m(t)\}$ constitute an orthonormal set we get the Poisson formula

$$\sum_k |\Psi(\omega - 2\pi k)|^2 = 1$$

whence

$$|\Psi(\omega)| \leq 1. \quad (2.8a)$$

Moreover, from (2.3) we have immediately

$$\Psi(0) = 0. \quad (2.8b)$$

Finally, we are generally interested in *regular* bases, *i.e.*, bases comprised of regular basis functions. Regularity is a measure of the smoothness of a function. In particular, a function $f(t)$ will be said to be R th-order regular if its Fourier transform $F(\omega)$ decays

according to¹

$$F(\omega) \sim \mathcal{O}(|\omega|^{-R}), \quad |\omega| \rightarrow \infty.$$

We use the term “regular” to denote a function that is at least first-order regular, and note that an R th-order regular function has $R - 1$ regular derivatives. Consequently, in order for our wavelet basis to be regular we require that

$$\Psi(\omega) \sim \mathcal{O}(|\omega|^{-1}), \quad |\omega| \rightarrow \infty. \quad (2.8c)$$

As implied by (2.8a)–(2.8c), $\psi(t)$ is often the impulse response of an at least roughly bandpass filter. Consequently, the wavelet transformation can usually be interpreted either in terms of a generalized constant- Q (specifically, octave-band) filter bank, or, as we shall see later, in terms of a multiresolution signal analysis. While we will restrict our attention to this class of wavelet bases, it is important to remark, however, that wavelets need not correspond to either an octave-band filter bank or a multiresolution analysis. For example, the following wavelet due to Mallat [18]

$$\Psi(\omega) = \begin{cases} 1 & \text{if } 4\pi/7 < |\omega| \leq \pi \text{ or } 4\pi < |\omega| \leq 32\pi/7 \\ 0 & \text{otherwise} \end{cases}$$

generates a perfectly valid orthonormal wavelet basis.

¹The order notation $\mathcal{O}(\cdot)$ used in this thesis is to be understood in the following sense. If

$$F(\omega) = \mathcal{O}(G(\omega)), \quad \omega \rightarrow \infty$$

then

$$\lim_{\omega \rightarrow \infty} \frac{F(\omega)}{G(\omega)} < \infty.$$

2.2.1 An Octave-Band Filter Bank Interpretation

The filter bank interpretation of the wavelet transform arises by viewing the analysis equation (2.5b) as a *filter-and-sample operation*, *viz.*,

$$x_n^m = \{x(t) * \psi_0^m(-t)\}|_{t=2^{-m}n}.$$

Although the interpretation applies more generally, it is often convenient to visualize the basis associated with the ideal bandpass wavelet (2.7). In this case, the output of each filter in the bank is sampled at the corresponding Nyquist rate. More generally, we say that the filter bank is *critically-sampled* [15], in that reconstruction is not possible if any of the sampling rates are reduced regardless of the choice of wavelet. The critically-sampled filter bank corresponding to the wavelet decomposition is depicted in Fig. 2-1.

For a particular choice of wavelet basis, the magnitude of the frequency response of the filters in such a filter bank is portrayed in Fig. 2-2. As this figure illustrates, there can be significant spectral overlap in the magnitude responses while preserving the orthogonality of the decomposition. In essence, while the frequency response magnitudes are not supported on disjoint frequency intervals, aliasing is avoided—*i.e.*, perfect reconstruction and orthogonality are achieved—due to the characteristics of the phase in the filters. However, it is possible to construct wavelet bases such that the spectral overlap between channels is much smaller in applications where this is important.

A filter bank decomposition is closely related to the notion of a local time-frequency analysis. Provided the filters are reasonably bandpass in character, the output of each filter in the bank is an estimate of the frequency content in the signal localized to the corresponding frequency band. Likewise, provided the filter impulse responses are localized in time, the sequence of output samples from each filter gives a picture of the time-evolution of frequency content within the corresponding frequency band. In the case of the wavelet decomposition, $(x_n^m)^2$ represents an estimate of the energy of the signal $x(t)$ in the vicinity of $t \sim 2^{-m}n$, and for a band of frequencies

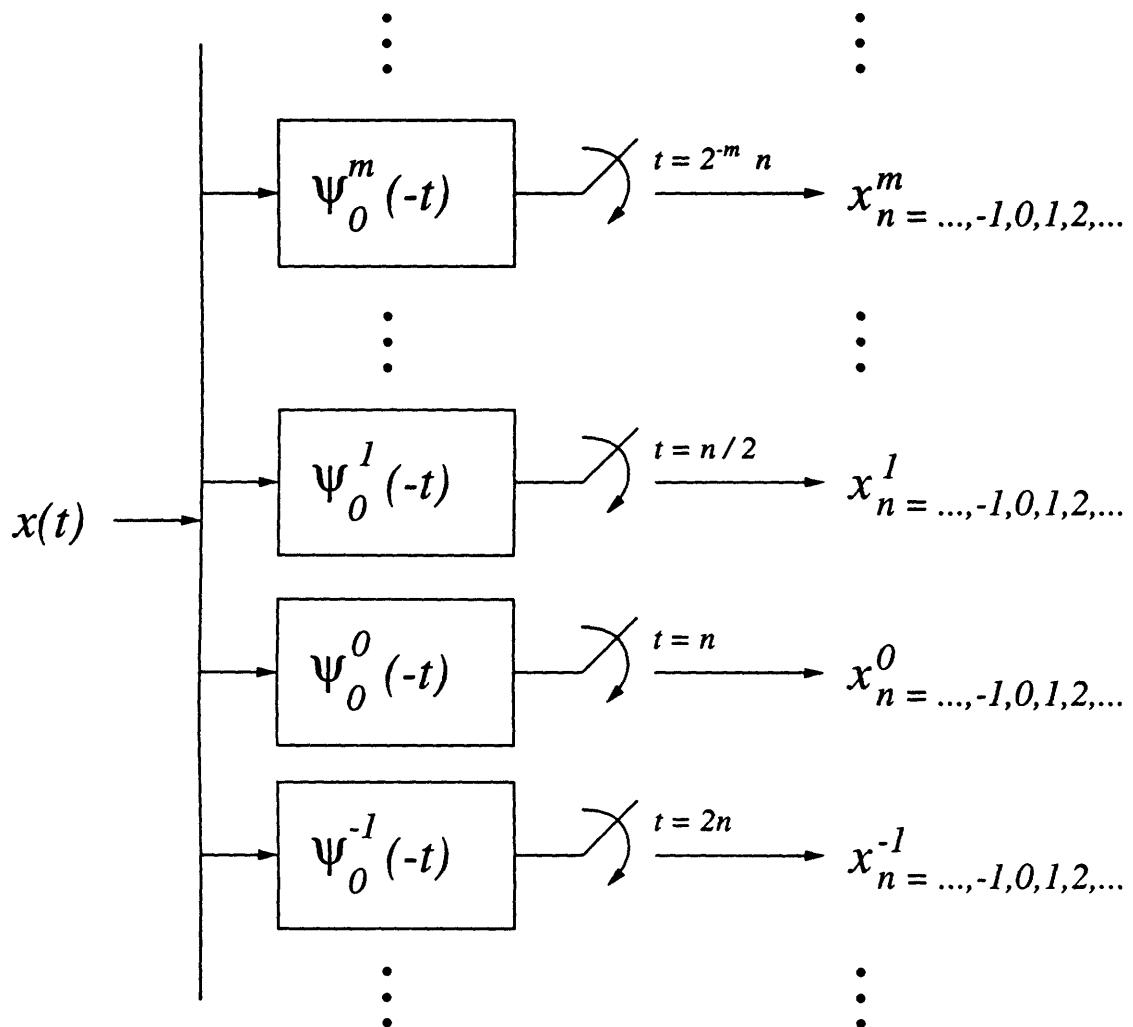


Figure 2-1: Critically-sampled filter bank interpretation of an orthonormal wavelet decomposition.

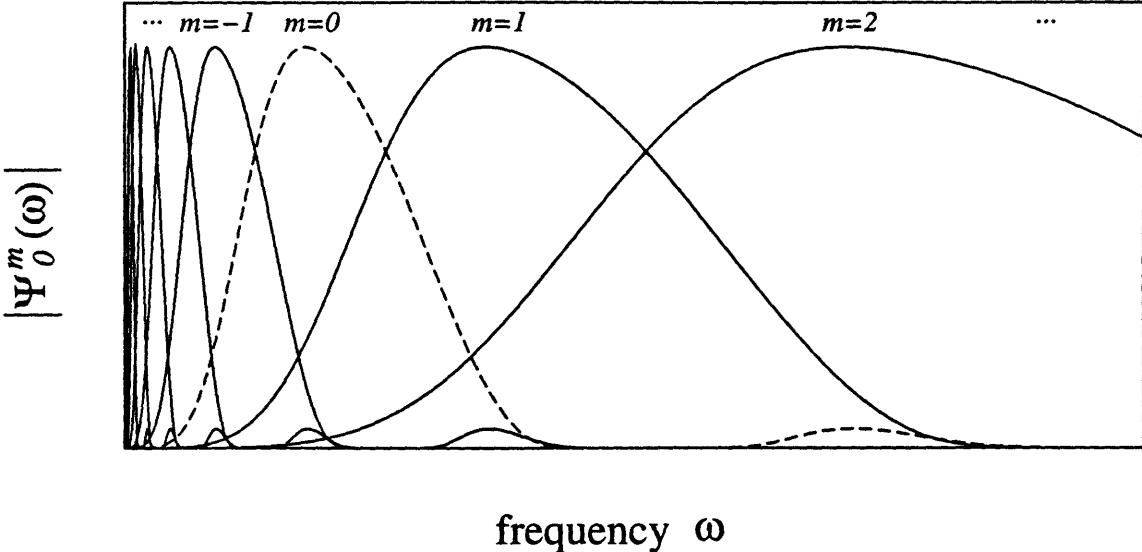
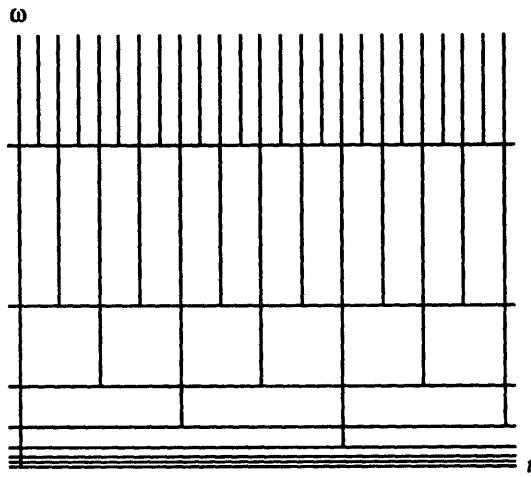


Figure 2-2: The octave band filters corresponding to an orthonormal wavelet decomposition. The wavelet basis in this example is one due to Daubechies [1].

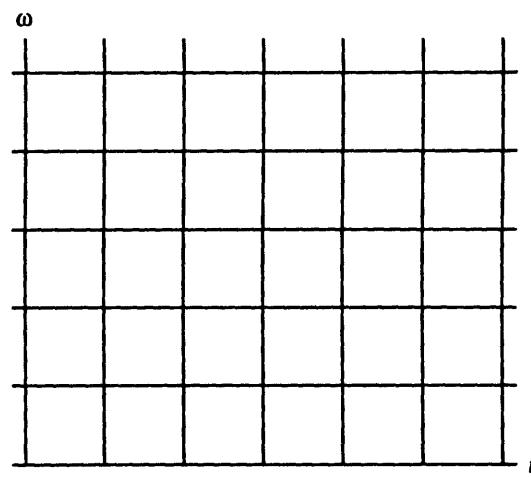
in the neighborhood of $\omega \sim 2^m\pi$. This is graphically depicted in the time-frequency plane of Fig. 2-3(a). Note that the octave-band frequency partitioning leads to a partitioning of the time axis that is finer in the higher (and wider) frequency bands. We emphasize that the partitioning in this figure is idealized: in accordance with the Fourier transform uncertainty principle, one cannot have perfect localization in both time and frequency. Nevertheless, one can construct wavelet bases whose basis functions have their energy concentrated at least roughly according to this partitioning.

In contrast to the wavelet transform, the familiar short-time Fourier transform (STFT) representation of a signal corresponds to a filter bank in which the filters are modulated versions of one another and, hence, have equal bandwidth. As a consequence, the outputs are sampled at identical rates, and the corresponding time-frequency analysis is one in which there is uniform partitioning of both the time and frequency axes in the time-frequency plane, as depicted in Fig. 2-3(b).

While the wavelet transform analysis equation (2.5b) can be interpreted in terms of a filter bank decomposition, the corresponding *synthesis* equation (2.5a) may be interpreted, as depicted in Fig. 2-4, as multirate modulation in which for a given m each sequence of coefficients x_n^m is modulated onto the corresponding wavelet dilate



(a) Wavelet transformation



(b) Short-time Fourier transformation

Figure 2-3: Time-frequency portraits corresponding to two signal analyses.

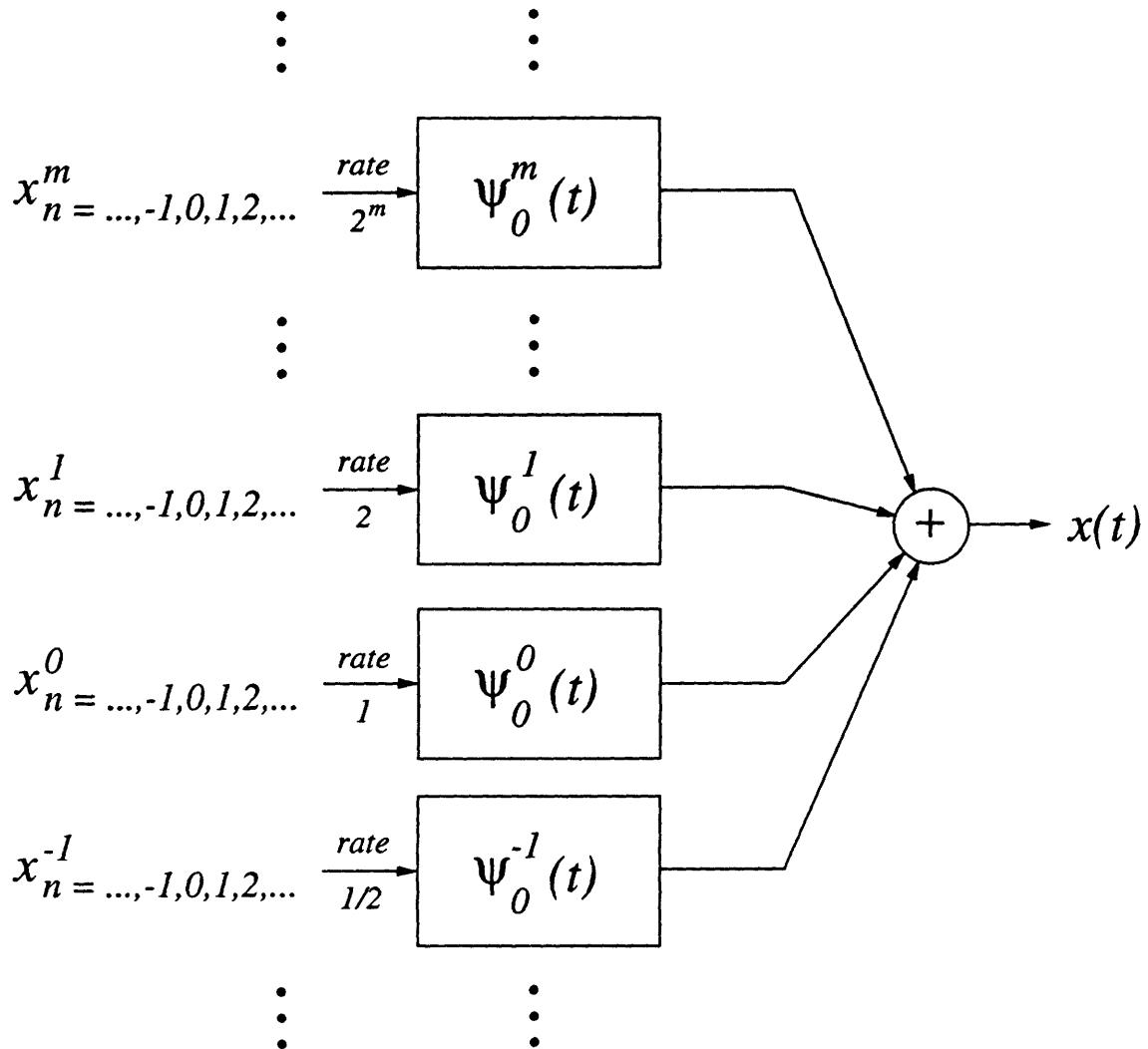


Figure 2-4: Interpretation of an orthonormal wavelet expansion as a multirate modulation scheme.

$\psi_0^m(t)$ at rate 2^m . For the case of the ideal bandpass wavelet, this corresponds to modulating each such sequence x_n^m into the distinct octave frequency band $2^m\pi < \omega \leq 2^{m+1}\pi$.

The filter bank interpretation allows us to readily derive the following useful identity

$$\sum_m |\Psi(2^{-m}\omega)|^2 = 1 \quad (2.9)$$

valid for all orthonormal wavelet bases and any $\omega \neq 0$. Specifically, consider an arbitrary finite-energy signal $x(t)$ with Fourier transform $X(\omega)$, which is decomposed

into an orthonormal wavelet basis via the filter bank of Fig. 2-1, then immediately resynthesized according to the filter bank of Fig. 2-4. It is a straightforward application of sampling theory to show that the Fourier transform of the output of this cascade can be expressed as

$$X(\omega) \sum_m |\Psi(2^{-m}\omega)|^2 + \sum_{k \neq 0} \sum_m X(\omega - 2\pi k 2^m) \Psi(2^{-m}\omega) \Psi^*(2^{-m}\omega - 2\pi k).$$

Since this must be equal to $X(\omega)$, the terms on the right must all be zero, while the factor multiplying $X(\omega)$ must be unity, yielding the identity (2.9) as desired.

While the filter bank interpretation provides a natural, convenient, and familiar framework in which to view orthonormal wavelet transformations, it is also possible to view the transformation in the context of a multiresolution signal analysis framework [5] [6] [1]. This perspective, which we consider next, provides a number of additional, rich insights into orthonormal wavelet bases.

2.2.2 Multiresolution Signal Analysis Interpretation

In general, a multiresolution signal analysis is a framework for analyzing signals based on isolating variations in the signal that occur on different temporal or spatial scales. This strategy underlies a variety of diverse signal processing algorithms including pyramidal methods used in the solution of computer vision problems [19] and multi-grid methods used in the solution of boundary value problems [8]. The basic analysis algorithm involves approximating the signal at successively coarser scales through repeated application of a smoothing or averaging operator. At each stage, a differencing operation is used to extract a detail signal capturing the information *between* consecutive approximations. The matching synthesis algorithm involves a successive refinement procedure in which, starting from some coarsest scale approximation, detail signals are accumulated in order to generate successively finer scale signal approximations.

Specifically, orthonormal wavelet bases can be interpreted in the context of a particular class of linear multiresolution signal analyses in which signal approximations

at all resolutions of the form 2^m (for m an integer) are defined. In describing this class, we begin formally by considering the Hilbert space of square-integrable signals $\mathbf{V} = \mathbf{L}^2(\mathbf{R})$. A multiresolution signal analysis is then defined as a decomposition of this signal space \mathbf{V} into a sequence of subspaces

$$\dots, \mathbf{V}_{-1}, \mathbf{V}_0, \mathbf{V}_1, \mathbf{V}_2, \dots$$

such that each \mathbf{V}_m defines signal approximations at a resolution 2^m . Associated with each \mathbf{V}_m is a linear operator A_m that defines projections from anywhere in \mathbf{V} onto \mathbf{V}_m . That is, for each signal $x(t) \in \mathbf{V}$, the projection $A_m x(t) \in \mathbf{V}_m$ defines the closest signal of resolution 2^m to $x(t)$,

$$A_m x(t) = \arg \min_{v(t) \in \mathbf{V}_m} \|x(t) - v(t)\|.$$

Central to the concept of multiresolution analysis is the notion of being able to construct successively coarser resolution approximations by repeated application of a smoothing operator. Mathematically, this characteristic is obtained by imposing the nesting or causality relation

$$\mathbf{V}_m \subset \mathbf{V}_{m+1}, \quad (2.10a)$$

which specifically ensures that the approximation of a signal at resolution 2^{m+1} contains all the information necessary to approximate the signal at the coarser resolution 2^m :

$$A_m \{A_{m+1} x(t)\} = A_m x(t).$$

The relations

$$\overline{\bigcup_{m=-\infty}^{\infty} \mathbf{V}_m} = \mathbf{V} \quad (2.10b)$$

$$\bigcap_{m=-\infty}^{\infty} \mathbf{V}_m = \{0\} \quad (2.10c)$$

ensure that a complete range of approximations is defined by the analysis. In the

process, these completeness relations define arbitrarily good and arbitrarily poor approximations that are consistent with any intuitive notion of resolution, *i.e.*,

$$\begin{aligned}\lim_{m \rightarrow \infty} A_m x(t) &= x(t) \\ \lim_{m \rightarrow -\infty} A_m x(t) &= 0.\end{aligned}$$

An additional relation is required to fully define the notion of resolution: signals in \mathbf{V}_m must be characterized by 2^m samples per unit length. Mathematically, this can be interpreted as requiring that there exist an isometry between each space of functions \mathbf{V}_m and the space of square-summable sequences $I = l^2(\mathbf{Z})$

$$\mathbf{V}_m \xleftrightarrow{\text{isom}} I \quad (2.10d)$$

such that each sequence represents samples of the corresponding signal following some potentially rather arbitrary linear processing:

$$x(t) \in \mathbf{V}_m \Rightarrow \varphi_m x(t)|_{t=2^{-m}n} \in I \quad (2.10e)$$

where φ_m is a linear operator.

In general, eqs. (2.10a) – (2.10e) are adequate to define a multiresolution signal analysis. However, imposing two additional constraints leads to an analysis with some nice structure. The first is a translation-invariance constraint, *viz.*,

$$x(t) \in \mathbf{V}_m \Leftrightarrow x(t - 2^{-m}n) \in \mathbf{V}_m \quad (2.10f)$$

which ensures that the nature of the approximation of the signal $x(t)$ is the same for any time interval. It is this condition that shall lead to the translational relationships among basis functions in the corresponding wavelet expansion. The second is a scale-invariance constraint

$$x(t) \in \mathbf{V}_m \Leftrightarrow x(2t) \in \mathbf{V}_{m+1} \quad (2.10g)$$

which ensures that the nature of the approximation at each resolution is the same. In turn, it is this condition that shall give rise to the dilational relationships among basis functions in the corresponding wavelet expansion.

It can be shown [18] that every multiresolution analysis of $L^2(\mathbf{R})$, *i.e.*, every collection of subspaces V_m defined in accordance with (2.10a) – (2.10g), is completely characterized in terms of a *scaling function* (or “father” wavelet) $\phi(t)$. Consequently, from the scaling function one can construct an orthonormal basis for each V_m , and, hence, the approximation operator A_m for each of these subspaces. In particular, for each m ,

$$\dots, \phi_{-1}^m(t), \phi_0^m(t), \phi_1^m(t), \phi_2^m(t), \dots$$

constitutes an orthonormal basis for V_m , where the basis functions, as a consequence of the invariance constraints (2.10f) and (2.10g) imposed on the multiresolution analysis, are all dilations and translations of one another, *i.e.*,

$$\phi_n^m(t) = 2^{m/2} \phi(2^m t - n). \quad (2.11)$$

The corresponding resolution- 2^m approximation of a signal $x(t)$ is then obtained as the projection of $x(t)$ onto V_m , which, exploiting the convenience of an orthonormal basis expansion, is expressed as

$$A_m x(t) = \sum_n a_n^m \phi_n^m(t) \quad (2.12)$$

with the coefficients a_n^m computed according to the individual projections

$$a_n^m = \int_{-\infty}^{\infty} x(t) \phi_n^m(t) dt. \quad (2.13)$$

In general, $\phi(t)$ has a Fourier transform $\Phi(\omega)$ that is at least roughly lowpass. Using an argument similar to that which led to (2.8a), orthonormality of the basis $\{\phi_n^m(t)\}_n$ implies that

$$|\Phi(\omega)| \leq 1. \quad (2.14a)$$

Additionally, because this basis for \mathbf{V}_m is asymptotically complete in \mathbf{V} (*cf.* (2.10b)) we have

$$|\Phi(0)| = 1. \quad (2.14b)$$

Finally, since we are, again, generally interested in *regular* bases, we must have

$$\Phi(\omega) \sim \mathcal{O}(|\omega|^{-1}), \quad \omega \rightarrow \infty. \quad (2.14c)$$

Collectively, the properties (2.14a) – (2.14c) describe a scaling function that is consistent with the notion that A_m is an approximation or smoothing operator. Consequently, we may interpret the projection (2.13) as a lowpass-like filter-and-sample operation, *viz.*,

$$a_n^m = \{x(t) * \phi_0^m(t)\}|_{t=2^{-m}n}. \quad (2.15)$$

Moreover, (2.12) can be interpreted as a modulation of these samples onto a lowpass-like waveform.

In fact, one example of a multiresolution analysis is generated from the ideal lowpass scaling function $\tilde{\phi}(t)$, whose Fourier transform is the frequency response of an ideal lowpass filter, *i.e.*,

$$\tilde{\Phi}(\omega) = \begin{cases} 1 & |\omega| \leq \pi \\ 0 & |\omega| > \pi \end{cases}. \quad (2.16)$$

In this case, the corresponding multiresolution analysis is based upon perfectly bandlimited signal approximations. Specifically, for a signal $x(t)$, $A_m x(t)$ represents $x(t)$ bandlimited to $\omega = 2^m \pi$. Furthermore, we may interpret (2.15) and (2.12) in the context of classical sampling theory [20]. In particular, $\phi(t)$ in (2.15) plays the role of an anti-aliasing filter [21], while (2.12) is the interpolation formula associated with the sampling theorem.

Of course, there are practical difficulties associated with the implementation of a multiresolution analysis based upon perfectly bandlimited approximations, foremost of which is that the sampling and reconstruction filters, *i.e.*, the $\phi_0^m(t)$, are unrealizable.

able. For this reason, this analysis is more of pedagogical than practical interest.

To derive the wavelet basis associated with each multiresolution analysis defined via (2.10), we now shift our attention from the sequence of increasingly coarse scale *approximation* signals $A_m x(t)$ to the *detail* signals representing the information lost at each stage as the resolution is halved. The collection of resolution-limited signal approximations constitutes a highly redundant representation of the signal. By contrast, the collection of detail signals constitutes a much more efficient representation. Formally, we proceed by decomposing each space V_{m+1} into the subspace V_m and its orthogonal complement subspace O_m , *i.e.*, O_m satisfies

$$O_m \perp V_m \quad (2.17a)$$

$$O_m \oplus V_m = V_{m+1} \quad (2.17b)$$

where we recognize that it is in this orthogonal complement subspace that the detail signal resides.

Associated with every multiresolution analysis is a basic wavelet $\psi(t)$ which yields the following orthonormal basis for each O_m :

$$\dots, \psi_{-1}^m(t), \psi_0^m(t), \psi_1^m(t), \psi_2^m(t), \dots$$

where $\psi_n^m(t)$ is as defined in terms of dilations and translations of $\psi(t)$ as per (2.6).

In turn, this leads to a convenient description of the projection operator D_m from anywhere in V onto O_m as

$$D_m x(t) = \sum_n x_n^m \psi_n^m(t)$$

in terms of the individual projections (*cf.* (2.5b))

$$x_n^m = \int_{-\infty}^{\infty} x(t) \psi_n^m(t) dt.$$

Hence, we have the interpretation that the wavelet coefficients x_n^m for a fixed m corre-

spond to the detail signal $D_m x(t)$ at scale 2^m , or, more specifically, to the information in the signal $x(t)$ between the resolution- 2^m and resolution- 2^{m+1} approximations, *i.e.*,

$$D_m x(t) = A_{m+1}x(t) - A_m x(t).$$

At this point, we recognize the wavelet associated with the bandlimited multiresolution analysis defined via (2.16) to be the ideal bandpass wavelet (2.7); it suffices to consider a frequency domain perspective. To complete the discussion, we observe that via (2.17) we can recursively decompose any of the approximation subspaces V_M , for some M , into the direct sum of a sequence of orthogonal subspaces, *i.e.*,

$$V_M = O_{M-1} \oplus V_{M-1} = O_{M-1} \oplus (O_{M-2} \oplus V_{M-2}) = \dots = \bigoplus_{m < M} O_m. \quad (2.18)$$

from which we see that for every $x(t)$

$$A_M x(t) = \sum_{m < M} D_m x(t) = \sum_{m < M} \sum_n x_n^m \psi_n^m(t). \quad (2.19)$$

This leads naturally to the interpretation of $A_M x(t)$ as an approximation in which details on scales smaller than 2^M are discarded. Letting $M \rightarrow \infty$ in (2.19) yields

$$x(t) = \sum_m \sum_n x_n^m \psi_n^m(t),$$

the synthesis formula (2.5a), and corresponds to the subspace decomposition

$$V = \bigoplus_{m=-\infty}^{\infty} O_m.$$

This completes our interpretation of an orthonormal wavelet basis as a multiresolution signal analysis.

2.2.3 Discrete Wavelet Transform

The discrete wavelet transform (DWT) refers to a discrete-time framework for implementing the orthonormal wavelet transform. The basic notion is that rather than implementing the analysis directly as a sequence of continuous-time filter-and-sample operations according to (2.2.1), one can reformulate the analysis into a single continuous-to-discrete conversion procedure followed by some iterative discrete-time processing. Likewise, the synthesis can be reformulated from a series of conventional modulations (2.5a) into an iterative discrete-time procedure followed by a single discrete-to-continuous conversion.

The implementation is based upon the discrete-time filters

$$h[n] = \int_{-\infty}^{\infty} \phi_n^1(t) \phi_0^0(t) dt \quad (2.20a)$$

$$g[n] = \int_{-\infty}^{\infty} \phi_n^1(t) \psi_0^0(t) dt. \quad (2.20b)$$

Typically, $h[n]$ and $g[n]$ have Fourier transforms $H(\omega)$ and $G(\omega)$ that have roughly halfband lowpass and highpass characteristics, respectively. In fact, for the case of the bandlimited multiresolution signal analysis, $h[n]$ and $g[n]$ are *ideal* lowpass and highpass filters, specifically

$$\begin{aligned} H(\omega) &= \begin{cases} 1 & 0 < |\omega| \leq \pi/2 \\ 0 & \pi/2 < |\omega| \leq \pi \end{cases} \\ G(\omega) &= \begin{cases} 0 & 0 < |\omega| \leq \pi/2 \\ 1 & \pi/2 < |\omega| \leq \pi \end{cases}. \end{aligned}$$

More generally, as we shall see, the filters $h[n]$ and $g[n]$ form a quadrature mirror filter (QMF) or conjugate quadrature filter (CQF) pair.

The analysis algorithm is structured as follows. Given a signal $x(t) \in V$ from which we would like to extract x_n^m for $m \leq M$, we can obtain the approximation coefficients a_n^{M+1} via the filter-and-sample procedure of (2.15), then recursively apply

the following filter-downsample algorithm

$$a_n^m = \sum_l h[l - 2n] a_l^{m+1} \quad (2.21a)$$

$$x_n^m = \sum_l g[l - 2n] a_l^{m+1} \quad (2.21b)$$

to extract the transform coefficients x_n^m corresponding to successively coarser scales m . A detailed derivation of this algorithm is presented in Appendix A.

The synthesis algorithm is structured in a complementary fashion. In particular, to reconstruct $x(t)$ to resolution 2^{M+1} from x_n^m for $m \leq M$, we can recursively apply the upsample-filter-merge algorithm

$$a_n^{m+1} = \sum_l \{h[n - 2l] a_l^m + g[n - 2l] x_l^m\} \quad (2.21c)$$

to compute the coefficients a_n^m of successively finer scale approximations until level $m = M$ is reached, after which $A_{M+1}x(t)$ may be constructed by modulating a_n^M according to (2.12). A detailed derivation of this algorithm is also left to Appendix A.

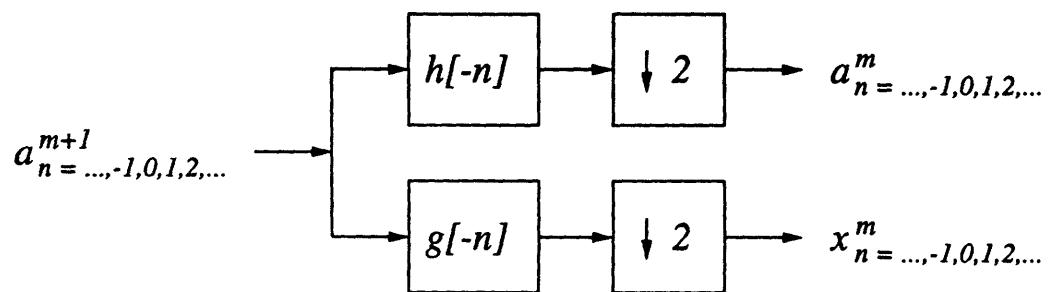
Fig. 2-5 depicts the discrete-time relationships between approximation and detail coefficients corresponding to adjacent scales. The complete algorithm for computing wavelet coefficients based on the discrete wavelet transform is depicted in Fig. 2-6.

The DWT may be computed extremely efficiently using polyphase forms [17]. Indeed, if the filters $h[n]$ and $g[n]$ have length L , an implementation of the DWT via an FFT-based algorithm generally has an asymptotic computational complexity of $\mathcal{O}(\log L)$ per input sample [22]. However, as discussed in [17] this figure can be somewhat misleading as there are many subtle issues associated with measuring complexity of the algorithm.

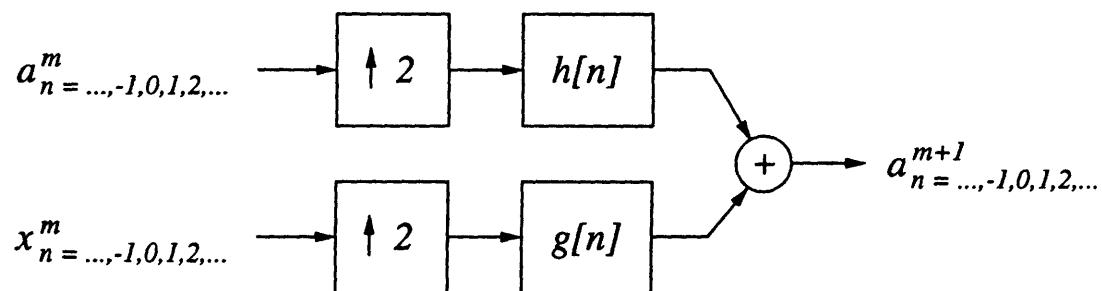
2.2.4 Finite Data Length and Resolution Effects

In most applications, the data consists of a finite collection of samples

$$x[n], \quad n = 0, 1, \dots, N.$$

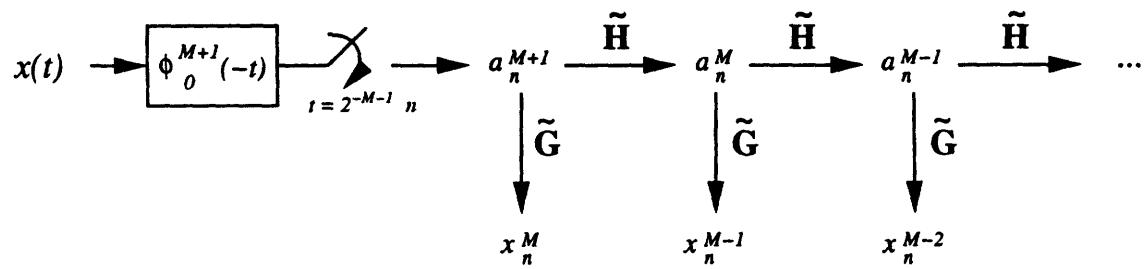


(a) The analysis step: filter-downsample.

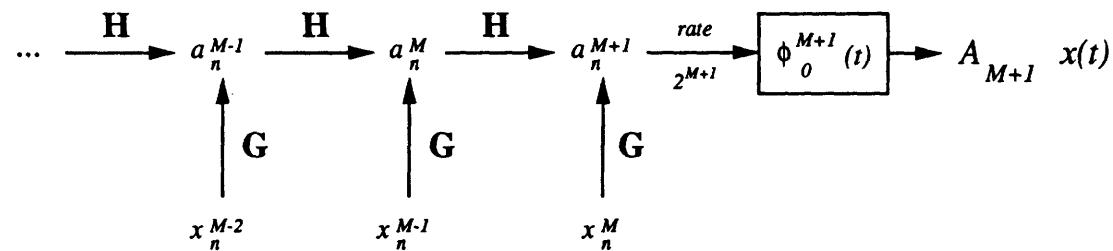


(b) The synthesis step: upsample-filter-merge.

Figure 2-5: A single stage of discrete-wavelet transform algorithm.



(a) The analysis algorithm.



(b) The synthesis algorithm

Figure 2-6: An efficient implementation of the orthonormal wavelet transformation based on the discrete wavelet transform.

While it is usually assumed that the $x[n]$ correspond to samples of a *resolution-limited approximation* of a continuous-time signal $x(t)$, i.e.,

$$x[n] = a_n^{M+1} = \{\phi_n^{M+1} * x(t)\} \Big|_{t=2^{-(M+1)}n}$$

for some M , this cannot always be justified. Nevertheless, if the signal $x(t)$ was processed by an a typical anti-aliasing filter prior to sampling, then it is often a useful approximation, particularly if the anti-aliasing filter has characteristics similar to that of the smoothing filter $\phi_0^{M+1}(t)$ associated with the approximation operator.

Note that while the discrete-time nature of the data limits access to the finer scales of detail, the length of the observations limits access to the coarser scales of detail. Hence, in practice we typically have access to wavelet coefficients over a finite range of scales for a given signal. Moreover, because the effective width of the wavelet basis functions halves at each finer scale, we expect roughly a doubling of the number of available coefficients at each successively finer scale. In a typical scenario, for a data record of $N = N_0 2^M$ samples, we would expect to be able to extract x_n^m corresponding to

$$\begin{aligned} m &= 1, 2, \dots, M \\ n &= 0, 1, \dots, N_0 2^{m-1} - 1 \end{aligned}$$

via the DWT, where N_0 is a constant that depends on the particular wavelet basis.

Note that while there are a number of ways to handle the unusual data windowing problem inherent in the wavelet decomposition, an assumption that the data is periodic outside the observation window leads to a computationally convenient implementation, and one we shall use in the context of this thesis. See [23] for some discussion and alternative approaches for addressing issues of windowing.

2.2.5 Orthonormal Wavelet Basis Constructions

As we have indicated, for every multiresolution analysis characterized by a scaling function, there exists an associated wavelet basis. In fact, it is possible to exploit the structure of the discrete wavelet transform to show how the wavelet $\psi(t)$ may always be derived directly from the scaling function $\phi(t)$. In this section we describe how this is accomplished. More generally, we show how one can construct a family of orthonormal wavelet bases directly from a class of discrete-time filters.

We begin by observing that there are a number of properties that the discrete-time filters $h[n]$ and $g[n]$ corresponding to a multiresolution signal must satisfy. For instance, as a consequence of orthogonality constraints between the $\{\psi_n^m(t)\}$ and $\{\phi_n^m(t)\}$, one can show [18] that $h[n]$ and $g[n]$ must be related by

$$g[n] = (-1)^n h[1 - n]$$

which, expressed in the frequency domain, is

$$G(\omega) = e^{-j\omega} H^*(\omega + \pi). \quad (2.22)$$

Furthermore, orthonormality of the $\{\phi_n^m(t)\}$ require that $h[n]$ satisfy

$$|H(0)|^2 = 2 \quad (2.23a)$$

$$|H(\omega)|^2 + |H(\omega + \pi)|^2 = 2. \quad (2.23b)$$

Filter pairs that satisfy both (2.22) and (2.23) are termed conjugate quadrature or quadrature mirror filters and have been discussed extensively in the signal processing literature [12].

Note that (2.22) leads immediately to an algorithm for constructing the wavelet corresponding to a particular scaling function: one can generate $h[n]$ from $\phi(t)$ via (2.20a), $g[n]$ from $h[n]$ via (2.22), then $\psi(t)$ from $g[n]$ and $\phi(t)$ via (A.1b).

However, note that $h[n]$ alone is also sufficient to fully characterize a wavelet basis

through a multiresolution analysis. Indeed, given $h[n]$, the dilation equation² (A.1a) can be solved for the corresponding scaling function $\phi(t)$. In particular, $\phi(t)$ has Fourier transform

$$\Phi(\omega) = \prod_{m=1}^{\infty} [2^{-1/2}H(2^{-m}\omega)] \quad (2.24)$$

which is intuitively reasonable from a recursive decomposition of the corresponding frequency domain equation, *viz.*, (A.2a).

In fact, the conditions (2.23) on $h[n]$ are necessary but not sufficient for (2.24) to generate a regular wavelet basis. However, choosing $h[n]$ to satisfy both (2.23) and to have a Fourier transform $H(\omega)$ with R zeros at $\omega = \pi$, *i.e.*,

$$H^{(r)}(0) = 0, \quad r = 0, 1, \dots, R - 1$$

is sufficient to generate a wavelet basis with R th-order regularity. Moreover, in this case, we find, via (A.2a), that the wavelet has R vanishing moments:

$$\int_{-\infty}^{\infty} t^r \psi(t) dt = (j)^r \Psi^{(r)}(0) = 0, \quad r = 0, 1, \dots, R - 1.$$

This vanishing moment property has been exploited in applications involving the implementation of linear operators [24] as well as in image coding. In the context of this work, we will provide evidence to suggest that this property may also be important when wavelet bases are used in representations for self-similar signals. It is important to note, however, that the vanishing moment condition is not necessary for regularity. For a more detailed discussion of necessary and sufficient conditions, see, *e.g.*, [25].

A variety of useful wavelet bases have been constructed from filter formulations of this type. In fact, this approach has been extremely useful in designing orthonormal wavelets with compact support, *i.e.*, wavelets for which

$$\psi(t) = 0, \quad |t| > T$$

²For a further discussion of dilation equations, see, *e.g.*, [16].

for some $0 < T < \infty$. This is a consequence of the natural correspondence between compactly-supported wavelets and the extensively developed theory of finite impulse response (FIR) digital filters. A more detailed discussion of relationships between wavelet theory and filter bank theory can be found in [17].

2.2.6 Examples

In this section, we briefly review some standard examples of wavelet bases. Thus far, we have discussed only one example, the wavelet basis corresponding to the ideal bandpass wavelet (2.7). This basis has excellent frequency localization properties, but very poor time-domain localization. Indeed, the corresponding wavelet $\psi(t)$ decays only like $1/t$ for large t , and the QMF filters $h[n]$ and $g[n]$ decay only like $1/n$ for large n . More seriously, this basis is unrealizable.

At the other extreme, consider a Haar-based multiresolution analysis in which the approximations at resolution 2^m are piecewise constant on intervals of length 2^{-m} . Here the scaling function is given by

$$\phi(t) = \begin{cases} 1 & 0 < t < 1 \\ 0 & \text{otherwise} \end{cases}$$

and the corresponding wavelet is

$$\psi(t) = \begin{cases} 1 & 0 < t < 1/2 \\ -1 & 1/2 \leq t < 1 \\ 0 & \text{otherwise} \end{cases} .$$

This analysis is realizable and exhibits excellent time localization but very poor frequency localization due to the abrupt time-domain transitions of the approximations. Indeed, $\Psi(\omega)$ falls off only like $1/\omega$ for $\omega \rightarrow \infty$.

More generally, we can consider the family of Battle-Lemarie wavelet bases [18] [1]. These bases may be derived from a multiresolution analysis based upon orthogonalized P th-order spline functions. For these bases, the corresponding scaling function is

given by

$$\Phi(\omega) = \frac{1}{\omega^{P+1}} \left[\sum_k \frac{1}{(\omega + 2\pi k)^{2(P+1)}} \right]^{-1/2}.$$

For example, the first-order ($P = 1$) Battle-Lemarie multiresolution analysis corresponds to piecewise-linear but continuous signal approximations. In this context, it is trivial to show that the Haar-based wavelet basis we have discussed corresponds to the case $P = 0$. Similarly, using an argument based on the central limit theorem, it is possible to show that the bandpass wavelet basis corresponds to $P \rightarrow \infty$. The Battle-Lemarie bases have very reasonable localization properties: they are characterized by exponential decay in the time domain and decay like $1/|\omega|^{P+1}$ in the frequency domain. Hence while they are, strictly-speaking, unrealizable, the exponential decay property ensures that good approximations may be realized via truncation.

Daubechies has designed an important class of compactly-supported wavelet bases [1] based upon discrete-time FIR filters. In addition to fulfilling a practical requirement of having finite-extent basis functions, these bases exhibit good localization in both time and frequency. The R th-order Daubechies basis is characterized by QMF filters $h[n]$ and $g[n]$ of length $2R$ for $R = 1, 2, \dots$, where the case $R = 1$ corresponds to the Haar-based wavelet basis. Moreover, the basis functions are *maximally-regular*, in the sense that they have the maximum number of vanishing moments (R) for a given order.

In general, the development of other families of wavelet-based multiresolution analyses continues to receive considerable attention in the literature, as described in, *e.g.*, [17].

2.2.7 Non-Dyadic Orthonormal Wavelet Bases

While we have focussed largely upon dyadic wavelet bases, for which the dilation and translation increments are $a = 2$ and $b = 1$, there are many other non-dyadic choices. In many applications, including those within the context of this thesis, such generalizations are potentially very useful particularly for $1 < a < 2$. This is because these correspond to an analysis with finer frequency resolution on the logarithmic

frequency scale. For instance, to have flexibility in choosing from among a family of bases corresponding to the lattice

$$a = (L+1)/L$$

$$b = L,$$

where $L = 1, 2, \dots$ is a parameter, would be highly convenient. An at least conceptually useful class of such bases arises out of a generalization of the bandpass basis defined by

$$\Psi(\omega) = \begin{cases} \sqrt{L} & \pi < |\omega| \leq \frac{L+1}{L}\pi \\ 0 & \text{otherwise} \end{cases}$$

where the case $L = 1$ corresponds to the usual bandpass basis. It is a straightforward exercise in analysis to verify that for each L the corresponding set $\{\psi_n^m(t)\}$, for which

$$\psi_n^m(t) = \left(\frac{L+1}{L}\right)^{m/2} \psi\left(\left(\frac{L+1}{L}\right)^m t - nL\right),$$

is complete and orthonormal. Unfortunately, however, the poor time-domain localization of these bases considerably reduces their practical value.

Chapter 3

Statistically Self-Similar Signals

Some of the most prevalent forms of fractal geometry in nature arise out of statistical scaling behavior in the underlying physical phenomena. In this chapter, we study an important class of statistically scale-invariant or self-similar random processes known as $1/f$ processes. These empirically defined processes, in particular, model a wide range of natural signals.

In first half of this chapter we first review the empirical properties of $1/f$ processes and a traditional mathematical model for $1/f$ behavior based on the fractional Brownian motion framework of Mandelbrot and Van Ness [26]. We then introduce and study a new and potentially important mathematical characterization for $1/f$ processes. The novelty and power of this characterization is its basis in the frequency domain, which admits a broader range of Fourier tools in the analysis of $1/f$ processes. In addition, we are able to show that our characterization includes the models of Mandelbrot and Van Ness, yet appears to avoid some of their limitations.

The latter half of the chapter develops models for the more broadly defined class of *nearly- $1/f$* models, which constitute equally useful models for many natural signals. For completeness, we first review some well-known ARMA-based constructions for nearly- $1/f$ processes. However, the principal focus in this section is on developing some new, powerful and efficient *wavelet-based* nearly- $1/f$ models. Using our frequency-based characterization of $1/f$ processes, we are able to show that a rather broad class of wavelet bases yield Karhunen-Loève-like expansions for nearly- $1/f$

processes. As a consequence, it is reasonable to model $1/f$ processes as orthonormal wavelet basis expansions in terms of uncorrelated coefficients. This suggests that wavelet-based analysis of $1/f$ -type behavior is not only convenient, but, in some sense, statistically optimal. In fact, in Chapter 4, we show how these wavelet-based representations are extremely useful in addressing problems of optimum detection and estimation involving $1/f$ -type signals.

We begin with a rather universally accepted definition. A random process $x(t)$ defined on $-\infty < t < \infty$ is said to be *statistically self-similar* if its statistics are invariant to dilations and compressions of the waveform in time. More specifically, a random process $x(t)$ is statistically self-similar with parameter H if for any real $a > 0$ it obeys the scaling relation

$$x(t) \stackrel{\mathcal{P}}{=} a^{-H} x(at) \quad (3.1)$$

where $\stackrel{\mathcal{P}}{=}$ denotes equality in a statistical sense. For *strict-sense* self-similar processes, this equality is in the sense of all finite-dimensional joint probability distributions. For *wide-sense* self-similar processes, the equality may be interpreted in the sense of second-order statistics, *i.e.*, mean and covariance functions. In this latter case, the self-similarity relation (3.1) may be alternately expressed as

$$M_x(t) \triangleq E[x(t)] = a^{-H} M_x(at) \quad (3.2a)$$

$$R_x(t, s) \triangleq E[x(t)x(s)] = a^{-2H} R_x(at, as). \quad (3.2b)$$

We will restrict our attention to Gaussian processes, for which the two definitions are, of course, equivalent. Furthermore, we will consider only zero-mean processes.

Even Gaussian processes satisfying (3.1) can exhibit great diversity in behavior. Some are stationary, as is the case with the classical generalized process $w(t)$ corresponding to zero-mean, stationary, white Gaussian noise. This process, whose auto-correlation function is an impulse, is self-similar with parameter $H = -1/2$. More typically, though, self-similar processes are nonstationary. For example, the Wiener

process (Brownian motion) $z(t)$ related to $w(t)$ by¹

$$z(t) = \int_0^t w(\tau) d\tau, \quad (3.3)$$

and extended to $t < 0$ through the convention

$$\int_0^t \triangleq - \int_t^0. \quad (3.4)$$

for all t is statistically self-similar with $H = 1/2$ and nonstationary but, evidently, has a stationary derivative. As a final example, the Gaussian process

$$x(t) = |t|^{H_0-1/2} z(t) \quad (3.5)$$

is self-similar with parameter H_0 for all values of H_0 , is nonstationary, and has a nonstationary derivative except for $H = 1/2$. In fact, when $x(t)$ in (3.5) is filtered by virtually any non-trivial linear time-invariant filter, the output is a nonstationary process. However, while most physical processes that exhibit self-similarity are fundamentally nonstationary, they retain a stationary quality to them. For this reason, processes such as (3.5) generally constitute rather poor models for such phenomena. By contrast, perhaps the most important class of models for such phenomena are the empirically characterized “ $1/f$ processes.”

3.1 $1/f$ Processes

The $1/f$ family of statistically self-similar random processes are generally defined as processes having measured power spectra obeying a power law relationship of the

¹Throughout this chapter, integrals with respect to the differential element $w(t) dt$, where $w(t)$ is a stationary white Gaussian noise, should be interpreted more precisely as integrals with respect to the differential element $dz(t)$, where $z(t)$ is the corresponding Wiener process. While it is customary to consider $w(t)$ to be the derivative of $z(t)$, recall that the non-differentiability of $z(t)$ means that $w(t)$ is its derivative only in a generalized sense. It is for this reason that an ordinary Riemann integral is technically inadequate, and the corresponding Riemann-Stieltjes is required. Nevertheless, we retain the notation $w(t) dt$ for conceptual convenience.

form

$$S_x(\omega) \sim \frac{\sigma_x^2}{|\omega|^\gamma} \quad (3.6)$$

for some spectral parameter γ related to H according to

$$\gamma = 2H + 1. \quad (3.7)$$

Generally, the power law relationship (3.6) extends over several decades of frequency. While data length generally limits access to spectral information at lower frequencies, data resolution limits access to spectral content at higher frequencies. Nevertheless, there are many examples of phenomena for which arbitrarily large data records justify a $1/f$ spectrum of the form (3.6) over all accessible frequencies. However, (3.6) is not integrable and, hence, strictly speaking, does not constitute a valid power spectrum in the theory of stationary random processes. As a consequence, there have been numerous attempts to attach an interpretation to such spectra based on notions of generalized spectra [27] [26] [28] [29].

As a consequence of their inherent self-similarity, the sample paths of $1/f$ processes are typically fractals, and have a fractal dimension in the sense of Mandelbrot [30] which characterizes their roughness. In fact, for the fractional Brownian motion models for $1/f$ processes we discuss in Section 3.1.1, the dependence of the fractal dimension D on the self-similarity parameter H can be derived analytically. More generally, the relationship between D and H , whereby increasing the parameter H yields a decrease in the fractal dimension D , may be obtained empirically. This is intuitively reasonable: an increase in H corresponds to an increase in γ , which, in turn, reflects a redistribution of power from high to low frequencies and leads to sample functions that are increasingly smooth in appearance. Fig. 3-1 illustrates some sample paths of $1/f$ processes corresponding to various values of γ . Note that what we have plotted are, in some sense, bandpass filtered versions of the sample functions, since the finite data length constrains the lowest accessible frequency and the discretization of the time-axis constrains the highest accessible frequency.

A truly enormous and tremendously varied collection of natural phenomena ex-

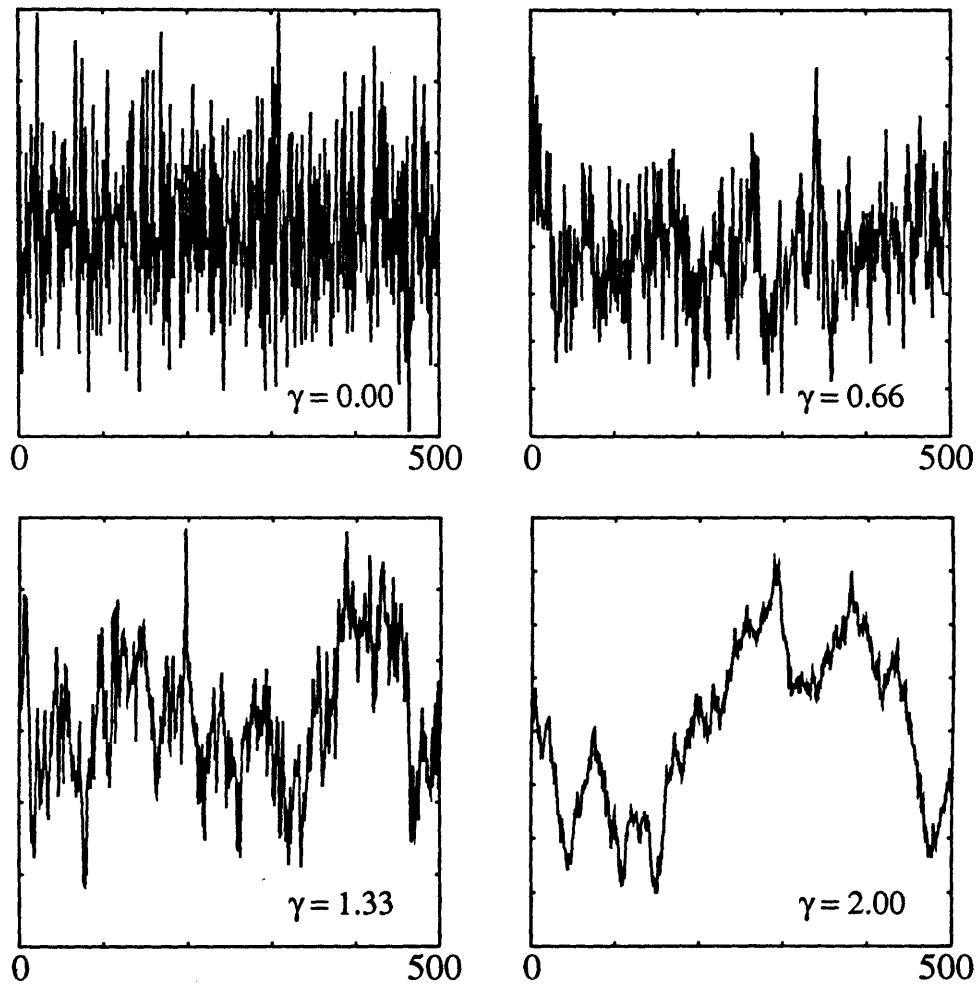


Figure 3-1: Sample paths of $1/f$ processes corresponding to different values of γ .

hibit $1/f$ -type spectral behavior over many decades of frequency. A partial list includes (see, *e.g.*, [28] [31] [32] [30] [33] [34] and the references therein):

- geophysical time series such as variation in temperature and rainfall records, measurements of oceanic flows, flood level variation in the Nile river, wobble in the Earth's axis, frequency variation in the Earth's rotation, and sunspot variations;
- economic time series such as the Dow Jones Industrial Average;
- physiological time series such as instantaneous heart rate records for healthy patients, EEG variations under pleasing stimuli, and insulin uptake rate data for diabetics;
- biological time series such as voltages across nerve and synthetic membranes;
- electromagnetic fluctuations such as in galactic radiation noise, the intensity of light sources, and flux flow in superconductors;
- electronic device noises in field effect and bipolar transistors, vacuum tubes, and Schottky, Zener and tunnel diodes;
- resistance fluctuations in metal film, semiconductor films and contacts, germanium filaments in carbon and aqueous solution, thermocells, and concentrations cells;
- frequency variation in hourglasses, quartz crystal oscillators, atomic clocks, and superconducting cavity resonators;
- man-induced phenomena including variations in traffic flow and amplitude and frequency variation in Western, African, Asian and Indian music, both modern and traditional;
- generation of perceptually-pleasing physiological stimuli, such as artificial music and breezes;

- burst errors on communication channels;
- texture variation in natural terrain, landscapes, and cloud formations.

While $\gamma \approx 1$ in many of these examples, more generally $0 \leq \gamma \leq 2$. However, there are many examples of phenomena in which γ lies well outside this range. For $\gamma \geq 1$, the lack of integrability of (3.6) in a neighborhood of the spectral origin reflects the preponderance of low-frequency energy in the corresponding processes. This phenomenon is termed the infrared (IR) catastrophe. For many physical phenomena, measurements corresponding to very small frequencies show no low-frequency roll off, which is usually understood to reveal an inherent nonstationarity in the underlying process. Such is the case for the Wiener process discussed earlier. For $\gamma \leq 1$, the lack of integrability in the tails of the spectrum reflects a preponderance of high-frequency energy and is termed the ultraviolet (UV) catastrophe. Such behavior is familiar for generalized Gaussian processes such as stationary white Gaussian noise and its usual derivatives.

An important property of $1/f$ processes is their persistent statistical dependence. Indeed, the generalized Fourier pair [35]

$$\frac{|\tau|^{\gamma-1}}{2\Gamma(\gamma)\cos(\gamma\pi/2)} \longleftrightarrow \frac{1}{|\omega|^\gamma} \quad (3.8)$$

valid for $\gamma > 0$ but $\gamma \neq 1, 2, 3, \dots$, suggests that the autocorrelation $R_x(\tau)$ associated with the spectrum (3.6) for $0 < \gamma < 1$ is characterized by slow decay of the form

$$R_x(\tau) \sim |\tau|^{\gamma-1}.$$

This power law decay in correlation structure distinguishes $1/f$ processes from many traditional models for time series analysis. Indeed, the well-studied family of autoregressive moving-average (ARMA) models have a correlation structure invariably characterized by *exponential* decay. As a consequence, ARMA models are generally inadequate for capturing long-term dependence in data.

Perhaps the most important families of $1/f$ processes are those which are non-

Gaussian. Indeed, a number of rich and interesting examples of non-Gaussian self-similar behavior can be constructed by exploiting the theory of stable distributions [26] [36] [37] [38]. Nevertheless, Gaussian models are generally applicable in a broad range of contexts, and are analytically highly tractable. For these reasons, we focus principally on Gaussian $1/f$ processes in this thesis. In the next section, we review what are perhaps the most popular mathematical models for Gaussian $1/f$ processes: fractional Brownian motion and fractional Gaussian noise.

3.1.1 Fractional Brownian Motion and Fractional Gaussian Noise

It is generally agreed [38] [26] [30] that fractional Brownian motion (fBm) and fractional Gaussian noise (fGn) models were first proposed by Kolmogorov, although their current popularity is undoubtedly due to Mandelbrot who independently derived the theory with Van Ness in [26] and promoted their use in numerous subsequent publications (see, *e.g.*, the references in [30]). An extensive bibliographic guide to various subsequent developments of the theory, principally in the mathematics literature, is presented by Taqqu in [38].

In this framework, processes corresponding to $1 < \gamma < 3$, for which there is infinite low-frequency power, are developed as nonstationary random processes having finite-power in any finite time interval. These processes are the fractional Brownian motions, and classical Brownian motion is a special case corresponding to $\gamma = 2$. By contrast, processes corresponding to $-1 < \gamma < 1$, for which there is infinite high-frequency power, are developed as generalized stationary Gaussian processes corresponding to the derivative of a fractional Brownian motion. These processes are the fractional Gaussian noises, and stationary white Gaussian noise is a special case corresponding to $\gamma = 0$. The theory does not accommodate the cases $\gamma > 3$ and $\gamma < -1$. Furthermore, the models are degenerate for the cases $\gamma = -1$, $\gamma = 1$, and $\gamma = 3$.

A reasonable approach for developing $1/f$ models would appear to be to consider

driving stationary white Gaussian noise through a linear time-invariant system with impulse response

$$v(t) = \frac{1}{\Gamma(H + 1/2)} t^{H-1/2} u(t) \quad (3.9)$$

where $\Gamma(\cdot)$ is the gamma function. Indeed, (3.9) has the generalized Laplace transform [39]

$$\Upsilon(s) = \frac{1}{s^{H+1/2}},$$

which suggests that if the input has spectral density σ_x^2 , the output will have a power spectrum, in some sense, of the form (3.6) where γ is given via (3.7). As we will discuss in Chapter 7, (3.9) represents an example of a linear jointly time- and scale-invariant system of degree $H + 1/2$. However, the system defined via (3.9) is unstable except for the degenerate case $H = -1/2$. Consequently, the convolution

$$x(t) = v(t) * w(t) = \frac{1}{\Gamma(H + 1/2)} \int_{-\infty}^t (t - \tau)^{H-1/2} w(\tau) d\tau \quad (3.10)$$

is not well-defined.

In developing their $1/f$ model, Barnes and Allan [40] addressed this dilemma by keying the integration in (3.10) to the time origin, defining their self-similar process by

$$x(t) = \frac{1}{\Gamma(H + 1/2)} \int_0^t |t - \tau|^{H-1/2} w(\tau) d\tau \quad (3.11)$$

where this definition is extended for $t < 0$ through the convention (3.4). It is interesting to remark that (3.11) is familiar in mathematics as the Riemann-Liouville integral of $w(\tau)$ over the interval $0 < \tau < t$. In fractional calculus theory [41], it often is used to define the fractional integral of $w(t)$ of order $\lambda = H + 1/2 > 0$, usually denoted

$$x(t) = \frac{d^{-\lambda}}{dt^{-\lambda}} w(t).$$

The resulting process is well-defined, satisfies $x(0) = 0$, and is statistically self-

similar with parameter H , *i.e.*,

$$R_x(t, s) = \int_0^{\max(t, s)} (\max(t, s) - \tau)^{H-1/2} d\tau = a^{-2H} R_x(at, as) \quad (3.12)$$

for $a > 0$. However, the Barnes-Allan process constitutes a rather poor model for $1/f$ behavior. In fact, it lacks any kind of stationary quality. For instance, the increment process

$$\Delta x(t; \varepsilon) \triangleq \frac{x(t + \varepsilon) - x(t)}{\varepsilon} \quad (3.13)$$

while statistically self-similar, satisfying

$$\Delta x(t; \varepsilon) \stackrel{\mathcal{P}}{=} a^{-(H-1)} \Delta x(at; a\varepsilon) \quad (3.14)$$

for every $\varepsilon > 0$, is nonstationary. Consequently, one cannot associate a stationary generalized derivative with the process. In effect, the underlying problem is that the Barnes-Allan process places too much emphasis on the time origin [26].

Fractional Brownian motion represents a very useful refinement of the Barnes-Allan process. Specifically, fractional Brownian motion is a nonstationary Gaussian self-similar process $x(t)$ also satisfying $x(0) = 0$, but defined in such a way that its corresponding increment process $\Delta x(t; \varepsilon)$ is self-similar *and* stationary for every $\varepsilon > 0$. Imposing these constraints on increments of fractional Brownian motion leads to a comparatively better model for $1/f$ behavior.

A convenient though specialized definition of fractional Brownian motion is given by [42]

$$\begin{aligned} x(t) \triangleq & \frac{1}{\Gamma(H + 1/2)} \left[\int_{-\infty}^0 (|t - \tau|^{H-1/2} - |\tau|^{H-1/2}) w(\tau) d\tau \right. \\ & \left. + \int_0^t |t - \tau|^{H-1/2} w(\tau) d\tau \right] \end{aligned} \quad (3.15)$$

for $0 < H < 1$, where $w(t)$ is a zero-mean, stationary white Gaussian noise process with unit spectral density. Again, for $t < 0$, $x(t)$ is defined through the convention (3.4). Note that with $H = 1/2$, (3.15) specializes to the Wiener process (3.3), *i.e.*,

classical Brownian motion.

Fractional Brownian motions are, in fact, fractals. Specifically, sample functions of fractional Brownian motions whose self-similarity parameters lie in the range $0 < H < 1$ (*i.e.*, $1 < \gamma < 3$) have a fractal (Hausdorff-Besicovitch) dimension given by [30]

$$D = 2 - H$$

that gives a quantitative measure of their roughness.

The correlation function for fractional Brownian motion can be readily derived as

$$R_x(t, s) = E[x(t)x(s)] = \frac{\sigma_H^2}{2} (|s|^{2H} + |t|^{2H} - |t - s|^{2H}) \quad (3.16)$$

where

$$\sigma_H^2 = \text{Var } x(1) = \Gamma(1 - 2H) \frac{\cos(\pi H)}{\pi H}, \quad (3.17)$$

from which it is straightforward to verify that the process is statistically self-similar with parameter H .

It is likewise straightforward to verify that the normalized increments of fractional Brownian motion are stationary and self-similar, and have the autocorrelation

$$\begin{aligned} R_{\Delta x}(\tau; \varepsilon) &\stackrel{\Delta}{=} E[\Delta x(t; \varepsilon)\Delta x(t - \tau; \varepsilon)] \\ &= \frac{\sigma_H^2 \varepsilon^{2H-2}}{2} \left[\left(\frac{|\tau|}{\varepsilon} + 1 \right)^{2H} - 2 \left(\frac{|\tau|}{\varepsilon} \right)^{2H} + \left(\frac{|\tau|}{\varepsilon} - 1 \right)^{2H} \right]. \end{aligned} \quad (3.18)$$

At large lags ($|\tau| \gg \varepsilon$), the correlation is asymptotically given by

$$R_{\Delta x}(\tau) \approx \sigma_H^2 H(2H - 1)|\tau|^{2H-2}. \quad (3.19)$$

Letting $\varepsilon \rightarrow 0$, and defining

$$H' = H - 1, \quad (3.20)$$

we can reason that fractional Brownian motion has the generalized derivative [42]

$$x'(t) = \frac{d}{dt}x(t) = \lim_{\varepsilon \rightarrow 0} \Delta x(t; \varepsilon) = \frac{1}{\Gamma(H' + 1/2)} \int_{-\infty}^t |t - \tau|^{H'-1/2} w(\tau) d\tau \quad (3.21)$$

which is termed fractional Gaussian noise. Note that (3.21) is precisely a convolution of the form (3.10), for which we now have an interpretation. Furthermore, from this observation we deduce that the derivative process $x'(t)$ is stationary and statistically self-similar with parameter H' .

From (3.18) it is apparent that the character of $x'(t)$ depends strongly on the value of H . Note that the right side of (3.19) has the same algebraic sign as $H - 1/2$. Hence, for $1/2 < H < 1$ this derivative process exhibits long-term dependence, *i.e.*, persistent correlation structure. For $H = 1/2$, this derivative is the usual stationary white Gaussian noise, which has no correlation, while for $0 < H < 1/2$, the derivative exhibits persistent anti-correlation. For $1/2 < H < 1$, $x'(t)$ is zero-mean and stationary with covariance

$$R_{x'}(\tau) = E[x'(t)x'(t - \tau)] = \sigma_H^2 (H' + 1)(2H' + 1)|\tau|^{2H'} \quad (3.22)$$

and we note that the generalized Fourier pair (3.8) suggests that the corresponding power spectral density of the derivative process can be expressed, for $\omega \neq 0$, as

$$S_{x'}(\omega) = \frac{1}{|\omega|^{\gamma'}}, \quad (3.23)$$

where

$$\gamma' = 2H' + 1.$$

A conceptually useful synthesis for fractional Brownian motion is depicted in Fig. 3-2. In particular, driving a linear time-invariant system with impulse response

$$v(t) = \frac{1}{\Gamma(H - 1/2)} t^{H-3/2} u(t)$$

with stationary white Gaussian noise $w(t)$ generates a fractional Gaussian noise $x'(t)$,

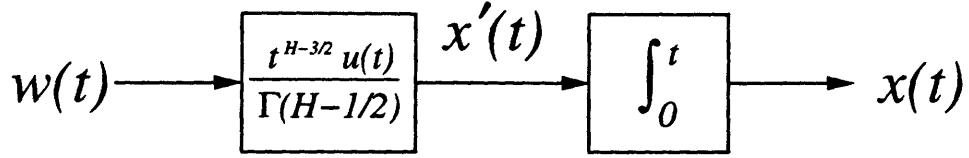


Figure 3-2: *Synthesis of fractional Brownian motion $x(t)$ in terms of fractional Gaussian noise $x'(t)$ and stationary white Gaussian noise $w(t)$.*

from which fractional Brownian motion $x(t)$ is obtained by routine integration:

$$x(t) = \int_0^t x'(t) dt.$$

The fractional Brownian motion framework provides a useful construction for some models of $1/f$ -type spectral behavior corresponding to spectral exponents in the range $-1 < \gamma < 1$ and $1 < \gamma < 3$. In fact, they uniquely model certain classes of $1/f$ behavior. One can show, for instance, that for $0 < H < 1$, fractional Brownian motion constitutes the *only* statistically self-similar, zero-mean, mean-square continuous, finite-variance, Gaussian random process satisfying $x(0) = 0$ and having stationary increments. While these are somewhat restrictive conditions, this framework has, in general, become a popular one for modeling a variety of phenomena with $1/f$ behavior—see, *e.g.*, [30] [32] [43] [42]. However, fractional Brownian motion and fractional Gaussian noise are not the only models for $1/f$ behavior, even within the respective parameter ranges $-1 < \gamma < 1$ and $1 < \gamma < 3$. In fact, in some cases they constitute rather poor models for $1/f$ behavior.

One unsatisfying characteristic of fractional Brownian motion is its pronounced time origin. Indeed, fractional Brownian motion satisfies not only $x(0) = 0$, but also power law growth in variance as a function of time, *i.e.*,

$$\text{Var } x(t) = \sigma_H^2 |t|^{2H}.$$

In modeling many physical phenomena having empirical spectra corresponding to γ in this range, the notion of such a time origin is not only rarely observed, but rather

unnatural as well.

Additionally, the fractional Brownian motion framework has a number of limitations. In particular, it does not lead to useful models for $1/f$ processes corresponding to $\gamma \leq -1$, $\gamma \geq 3$, and perhaps the most important and ubiquitous case, $\gamma = 1$. Indeed, for $\gamma = 3$ ($H = 1$), fractional Brownian motion as defined by (3.15) degenerates to a process whose sample paths are all lines through the origin, *viz.*,

$$x(t) = |t|x(1).$$

For $\gamma = 1$ ($H = 0$), fractional Brownian motion degenerates to the trivial process

$$x(t) \equiv 0.$$

More generally, choosing $H < 0$ in (3.15) leads to processes that are not mean-square continuous, while choosing $H > 1$ in (3.15) leads to processes whose increments are not stationary [26] [37].

In the next section, we consider a more general but non-constructive model for $1/f$ processes that includes both fractional Brownian motions and fractional Gaussian noises, yet appears to avoid some of the restrictions imposed by the fractional Brownian motion framework.

3.1.2 A New Mathematical Characterization of $1/f$ Processes

The notion that measurements of spectra for physical processes can only be obtained over a range of frequencies governed by data length and resolution limitations suggests a potentially useful approach for *defining* $1/f$ processes. In particular, it would appear reasonable to define $1/f$ processes in terms of their characteristics under bandpass filtering. As a consequence, we propose the following definition.

Definition 3.1 *A wide-sense statistically self-similar zero-mean random process $x(t)$ shall be said to be a $1/f$ process if when $x(t)$ is filtered by an ideal bandpass filter with*

frequency response

$$B_1(\omega) = \begin{cases} 1 & \pi < |\omega| \leq 2\pi \\ 0 & \text{otherwise} \end{cases} \quad (3.24)$$

the resulting process $y_1(t)$ is wide-sense stationary and has finite variance.

In fact, the choice of passband $\pi < |\omega| \leq 2\pi$ in this definition is rather arbitrary. More generally, replacing this passband with any other passband that contains neither $\omega = 0$ nor $\omega = \infty$ constitutes an equivalent characterization for this class of self-similar processes. Furthermore, one might speculate that choosing an ideal bandpass filter in this definition is not critical. In particular, it might suffice to choose any filter whose frequency response $B(\omega)$ with sufficient decay as $\omega \rightarrow 0$ and $\omega \rightarrow \infty$. Nevertheless, our choice is at least a convenient as we shall see. The appeal of Definition 3.1 as a characterization for $1/f$ processes is its basis in the frequency-domain. As a consequence, this allows us to extend the well-established tools of Fourier analysis to this important class of nonstationary processes. In turn, this allows us to derive a number of new properties of $1/f$ processes in a highly efficient manner.

The following theorem justifies designating processes satisfying Definition 3.1 as $1/f$ processes, and leads to an important interpretation of the spectrum (3.6) for $1/f$ processes. A detailed but straightforward proof is provided in Appendix B.1.

Theorem 3.2 *A $1/f$ process $x(t)$, when filtered by an ideal bandpass filter with frequency response*

$$B(\omega) = \begin{cases} 1 & \omega_L < |\omega| \leq \omega_U \\ 0 & \text{otherwise} \end{cases} \quad (3.25)$$

for arbitrary $0 < \omega_L < \omega_U < \infty$, yields a wide-sense stationary random process $y(t)$ with finite variance and having power spectrum

$$S_y(\omega) = \begin{cases} \sigma_x^2 / |\omega|^\gamma & \omega_L < |\omega| \leq \omega_U \\ 0 & \text{otherwise} \end{cases} \quad (3.26)$$

for some $\sigma_x^2 > 0$, and where the spectral exponent γ is related to the self-similarity parameter H according to (3.7).

An important question that must be addressed concerns whether there exist any non-trivial random processes satisfying Definition 3.1. The following theorem constitutes such an existence proof by verifying that Definition 3.1 is non-degenerate for at least some values of γ . In particular, it establishes that it is possible to construct families of Gaussian processes that satisfy this definition. A straightforward proof is provided in Appendix B.2.

Theorem 3.3 *Fractional Brownian motions corresponding to $0 < H < 1$ and the associated fractional Gaussian noises are $1/f$ processes.*

We remark that, based on our discussion of Section 3.1.1, an immediate corollary is that the Wiener process and stationary white Gaussian noise are also $1/f$ processes. However, the Barnes-Allan process we described at the outset of Section 3.1.1 is not a $1/f$ process in the sense of Definition 3.1. This is to be expected, given the shortcomings of the Barnes-Allan process in modeling $1/f$ -type behavior.

The other question that naturally arises concerns whether there are any other Gaussian $1/f$ processes besides those of Theorem 3.3. For instance, is it possible to construct non-trivial Gaussian processes that satisfy Definition 3.1 for values of H outside $0 < H < 1$? And, are there other Gaussian processes satisfying this definition for $0 < H < 1$? The very recent work of Ramanathan and Zeitouni [44] indirectly suggests that the answer may perhaps be negative, although they do not prove such a result. In effect, these authors show that if we were to replace the bandpass filter (3.24) in Definition 3.1 with a roughly bandpass filter whose frequency response is differentiable, has a simple zero at $\omega = 0$, and decays sufficiently quickly as $\omega \rightarrow \infty$, then the definition uniquely characterizes fractional Brownian motion. However, the constraints on the filter they consider are overly restrictive to answer our questions. Furthermore, it may be that the technical definition of a random process they consider is too narrowly chosen to accommodate $1/f$ behavior.

In any event, a practical difficulty with both the fractional Brownian motion framework and our new frequency-based characterization for $1/f$ processes is that while mathematically well-defined, neither is analytically convenient in many contexts. In fact, there are many very basic signal processing problems that are effectively intractable using these models.

To address this limitation, we next consider some more general classes of $1/f$ -*like* models. While these models typically do not give rise to *exactly-1/f* spectra, they give rise to spectra that are *nearly-1/f*. As we shall see, these processes retain most of the fundamental characteristics of $1/f$ processes, yet are considerably more amenable to analysis.

3.2 Nearly- $1/f$ Processes

Perhaps the best-known class of nearly- $1/f$ processes have been those based upon a generalized, infinite-order autoregressive moving-average (ARMA) framework. We shall review two such formulations before developing a wavelet-based model for $1/f$ -like behavior that constitutes the principle contribution of the chapter.

3.2.1 ARMA Models

There have been a variety of attempts to exploit a generalized autoregressive moving-average framework in modeling $1/f$ processes. Perhaps the earliest such framework, based on a “distribution of time constants” formulation, arose in the physics literature and dates back at least to the work of Bernamont [45]. However, it was really the seminal paper of van der Ziel [46] that sparked substantial interest in this approach, and much subsequent development.

Van der Ziel’s basic approach was to model a $1/f$ process as the weighted superposition of an infinite number of uncorrelated random processes, each governed by a distinct characteristic time-constant $1/\alpha > 0$. Each of these random processes has correlation function

$$R_\alpha(\tau) = e^{-\alpha|\tau|}$$

corresponding to a Lorenzian spectra of the form

$$S_\alpha(\omega) = \frac{2\alpha}{\alpha^2 + \omega^2}$$

and can be modeled as the output of a causal LTI filter with system function

$$\Upsilon_\alpha(s) = \frac{\sqrt{2\alpha}}{s + \alpha}$$

driven by an independent stationary white noise source. The weighted superposition of a continuum of such processes has an effective spectrum

$$S_x(\omega) = \int_0^\infty S_\alpha(\omega) f(\alpha) d\alpha \quad (3.27)$$

where the weights $f(\alpha)$ correspond to the density of poles or, equivalently, relaxation times. If an unnormalizable, scale-invariant density of the form

$$f(\alpha) = \alpha^{-\gamma} \quad (3.28)$$

for $0 < \gamma < 2$ is chosen, the resulting spectrum (3.27) is $1/f$, i.e.,

$$S_x(\omega) \propto \frac{1}{|\omega|^\gamma}.$$

This mathematical identity suggests a useful and practical approach to modeling $1/f$ -like behavior using the superposition of a *countable* collection of single time-constant processes whose poles are appropriately distributed. In fact, the density (3.28) implies that the poles should be uniformly distributed along a logarithmic frequency axis. The resulting process $x(t)$ synthesized in this manner then has a nearly- $1/f$ spectrum in the following sense: when $x(t)$ is filtered by any bandpass filter of the form (3.25) the result is a stationary process whose spectrum within the passband is $1/f$ with superimposed ripple that is uniform-spaced and of uniform amplitude on a log-log frequency plot.

As an example, consider exponentially-spaced poles according to

$$\alpha_m = \Delta^m, \quad -\infty < m < \infty, \quad (3.29)$$

for some $1 < \Delta < \infty$. Then the limiting spectrum

$$S_x(\omega) = \sum_m \frac{\Delta^{(2-\gamma)m}}{\omega^2 + \Delta^{2m}} \quad (3.30)$$

is bounded according to

$$\frac{\sigma_L^2}{|\omega|^\gamma} \leq S_x(\omega) \leq \frac{\sigma_U^2}{|\omega|^\gamma} \quad (3.31)$$

for some $-\infty < \sigma_L^2 \leq \sigma_U^2 < \infty$, and has ripple such that for all integers k

$$|\omega|^\gamma S_x(\omega) = |\Delta^k \omega|^\gamma S_x(\Delta^k \omega). \quad (3.32)$$

As Δ is chosen closer to unity, the pole spacing decreases, which results in a decrease in both the amplitude and spacing of the spectral ripple on a log-log plot.

Note that we may interpret the $1/f$ model that results from this discretization as an infinite-order ARMA process. That is, $x(t)$ can be viewed as the output of a rational LTI system with a countably infinite number of both poles and zeros driven by a stationary white noise source. There has been a substantial body of literature that has attempted to exploit this distribution-of-time-constants model in an effort to explain the ubiquity of $1/f$ spectra in nature. In essence, [47] [48] [49] construct mathematical arguments to the effect that $1/f$ spectra are the result of large, complex systems in nature favoring scale-invariant time-constant distributions of the form (3.28).

Somewhat more recently, Keshner [28] [31] developed an alternative ARMA-based model for $1/f$ -like behavior from an engineering perspective. This approach, which has also received considerable attention in the literature, is based on the observation that an infinite-length continuous RC transmission line when driven with a stationary white noise current $i(t)$ yields a measured voltage $v(t)$ whose power spectrum is of the form (3.6) for $0 < \gamma < 2$. That is, in some sense, the impedance function of the

line is of the form

$$\frac{V(s)}{I(s)} \propto \frac{1}{s^{\gamma/2}}.$$

By considering an infinite-length, *lumped*-parameter RC line as an approximation, Keshner showed that this gave rise to nearly- $1/f$ behavior in much the same manner as was obtained in the van der Ziel model. It is possible to interpret $1/f$ processes obtained in this manner as the result of driving stationary white noise through an LTI system with a rational system function

$$\Upsilon(s) = \prod_{m=-\infty}^{\infty} \left[\frac{s + \Delta^{m+\gamma/2}}{s + \Delta^m} \right] \quad (3.33)$$

which, in turn, leads to a spectrum of the form

$$S_x(\omega) \propto \prod_{m=-\infty}^{\infty} \left[\frac{\omega^2 + \Delta^{2m+\gamma}}{\omega^2 + \Delta^{2m}} \right]. \quad (3.34)$$

This nearly- $1/f$ spectrum has the same properties as the van der Ziel spectrum, satisfying both (3.31) and (3.32). In fact, comparing the spectra (3.34) and (3.30), we see that the pole placement strategy for both is identical. However, the zeros in the two models are, in general, distributed much differently. This, in turn, leads to differing degrees of ripple amplitude for a given pole spacing for the two different models [50].

It is interesting to remark that the ideal of using infinite lumped RC line to realize such systems was independently developed in the context of fractional calculus. Indeed, Oldham and Spanier [41] describe precisely such an implementation for fractional integration operators of the type used in the construction of the Barnes-Allan process (3.11).

The structure of the system function (3.33) of Keshner's synthesis filter provides additional insights into $1/f$ -like behavior. For example, the infinity of poles suggests that a state space characterization of $1/f$ processes would generally require uncountably many state variables, consistent with the notion of long term correlation structure in such processes.

Also, this system function lends useful insight into the limiting behavior of $1/f$ processes as $\gamma \rightarrow 0$ and $\gamma \rightarrow 1$. Note that the poles and zeros of (3.33) lie along the negative real axis in the s -plane. On a logarithmic scale, the poles and zeros are each spaced uniformly along this half line. Furthermore, in general, to the left of each pole in the s -plane lies a matching zero, so that poles and zeros are alternating along the half line. However, for certain values of γ , pole-zero cancellation takes place. In particular, as $\gamma \rightarrow 2$, the zero pattern shifts left canceling all poles except the limiting pole at $s = 0$. The resulting system is therefore an integrator, characterized by a single state variable, and generates a Wiener process as anticipated. By contrast, as $\gamma \rightarrow 0$, the zero pattern shifts right canceling all poles. The resulting system is therefore a multiple of the identity system, requires no state variables, and generates stationary white noise as anticipated.

Finally, note that the model may be interpreted in terms of a Bode plot. In general, stable, rational system functions comprised of real poles and zeros are only capable of generating transfer functions whose Bode plots have slopes that are integer multiples of

$$20 \log_{10} 2 \approx 6$$

dB per octave. However, a $1/f$ synthesis filter must fall off at

$$10\gamma \log_{10} 2 \approx 3\gamma$$

dB per octave where $0 < \gamma < 2$ is generally not an integer. To accommodate such slopes using rational system functions requires an alternating sequence of poles and zeros to generate a stepped approximation to a -3γ dB/octave slope from segments that alternate between slopes of -6 dB/octave and 0 dB/octave.

Unfortunately, neither of the ARMA-based models have been particularly useful in addressing basic problems of detection and estimation involving $1/f$ processes. However, both have been used extensively as $1/f$ noise simulators. A discrete-time implementation of the van der Ziel model is described in [51], while details of a discrete-time implementation of Keshner's model due to Corsini and Saletti appears

in [52]. A comparison of the two approaches is presented in [50]. In virtually all the simulations described in this thesis, the Corsini-Saletti implementation of Keshner's model will be used to synthesize $1/f$ processes. In particular, the $1/f$ sample paths of Fig. 3-1 are obtained via this algorithm.

3.2.2 Wavelet-Based Models

In this section, we explore the relationship between orthonormal wavelet bases and nearly- $1/f$ models. In particular, we shall show that wavelet basis expansions are both natural and convenient representations for processes exhibiting $1/f$ -like behavior. Our main result is that orthonormal wavelet basis expansions play the role of Karhunen-Loëve-like expansions for $1/f$ -type processes. That is, wavelet basis expansions in terms of uncorrelated random variables constitute good models for $1/f$ -type behavior.

Synthesis

In this section, we demonstrate that nearly- $1/f$ behavior may be generated from orthonormal wavelet basis expansions in terms of a collection of uncorrelated wavelet coefficients. In particular we establish the following theorem, an earlier version of which appears in [53], and whose proof is provided in Appendix B.3.

Theorem 3.4 *Consider any orthonormal wavelet basis with R th-order regularity for some $R \geq 1$. Then the random process constructed via the expansion*

$$x(t) = \sum_m \sum_n x_n^m \psi_n^m(t), \quad (3.35)$$

where the x_n^m are a collection of mutually uncorrelated, zero-mean random variables with variances

$$\text{Var } x_n^m = \sigma^2 2^{-\gamma m}$$

for some parameter $0 < \gamma < 2R$, has a time-averaged spectrum

$$S_x(\omega) = \sigma^2 \sum_m 2^{-\gamma m} |\Psi(2^{-m}\omega)|^2 \quad (3.36)$$

that is nearly- $1/f$, i.e.,

$$\frac{\sigma_L^2}{|\omega|^\gamma} \leq S_x(\omega) \leq \frac{\sigma_U^2}{|\omega|^\gamma} \quad (3.37)$$

for some $0 < \sigma_L^2 \leq \sigma_U^2 < \infty$, and has octave-spaced ripple, i.e., for any integer k

$$|\omega|^\gamma S_x(\omega) = |2^k \omega|^\gamma S_x(2^k \omega). \quad (3.38)$$

In Fig. 3-3 we illustrate the time-averaged spectrum of a process constructed in the manner of this theorem for $\gamma = 1$ using the first order Battle-Lemarie wavelet basis. Note the characteristic octave-spaced ripple. The bounding constants in this case correspond to $\sigma_U^2/\sigma_L^2 = 1.103$.

The result established by this theorem is certainly an intuitively reasonable one if, for example, we view the orthonormal wavelet decomposition as a generalized octave-band filter bank as described in Section 2.2.1. In fact, for the case of the ideal bandpass wavelet basis, it can be readily established from simple geometric arguments that the tightest bounding constants are

$$\begin{aligned} \sigma_L^2 &= \sigma^2 \pi^\gamma \\ \sigma_U^2 &= \sigma^2 (2\pi)^\gamma. \end{aligned}$$

Note, too, the special interpretation that may be derived from the model for the case $\gamma = 1$, arguably the most prevalent of the $1/f$ -type processes. Here the choice of the variance progression

$$\text{Var } x_n^m = \sigma^2 2^{-m}$$

corresponds to distributing power equally among the detail signals at all resolution scales, since we have for each m

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} P_m(\omega) |\Psi(2^{-m}\omega)|^2 d\omega = 1. \quad (3.39)$$

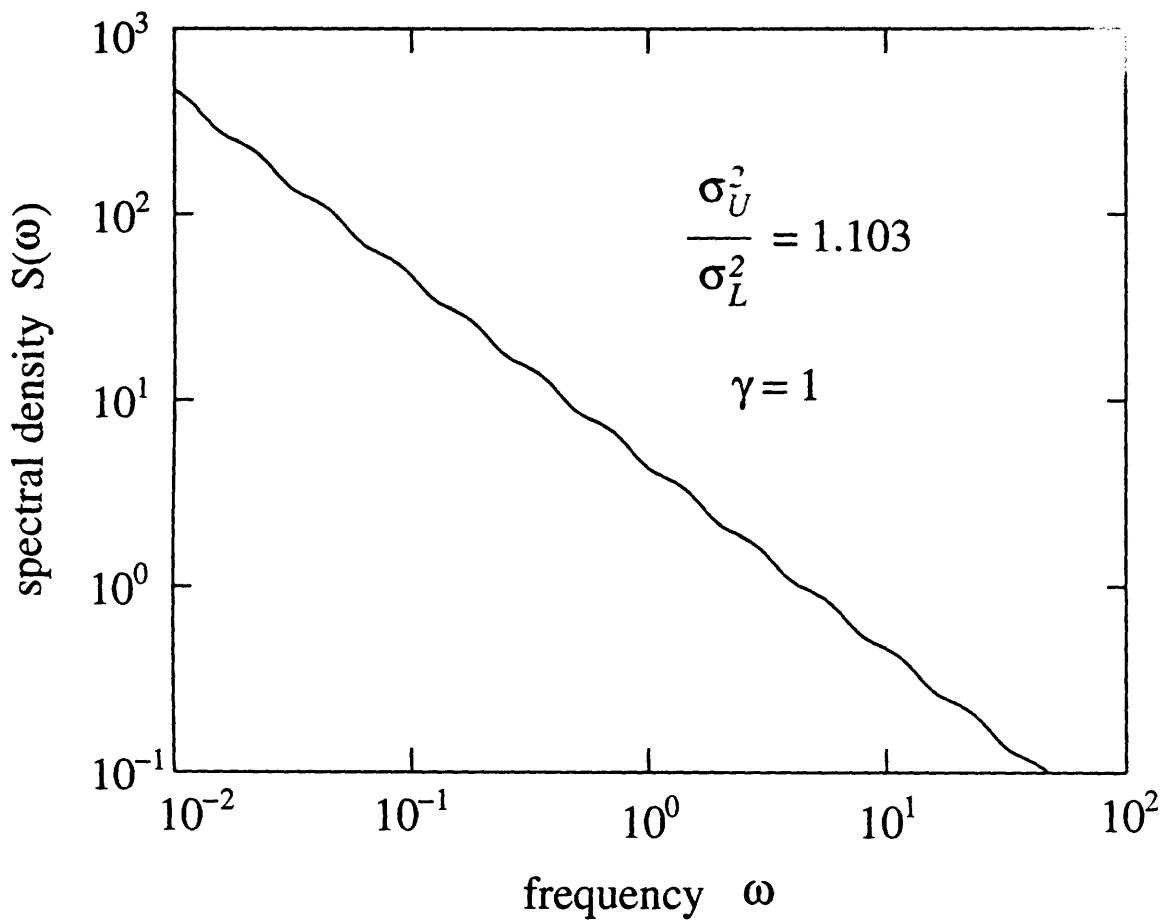


Figure 3-3: The time-averaged spectrum of a process synthesized from the first-order Battle-Lemarie orthonormal wavelet basis. The parameters of the nearly- $1/f$ spectrum are $\gamma = 1$ and $\sigma_U^2/\sigma_L^2 = 1.103$ in this case.

There are two aspects of this theorem that warrant further discussion. First, the nearly- $1/f$ spectrum (3.36) is to be interpreted in the same manner that the $1/f$ spectrum (3.6) is for exactly- $1/f$ processes. That is, if $x(t)$ is filtered by an ideal bandpass filter with frequency response of the form (3.25), the output of the filter will have finite-power and correspond to a spectrum of the form (3.36) over the passband $\omega_L < |\omega| \leq \omega_U$. However, it is important to emphasize that this spectrum is a *time-averaged* one. Indeed, the output of such a bandpass filter will *not*, in general, be stationary in any sense, which is a consequence of the discrete nature of the synthesis. This behavior is in contrast to the ARMA-based nearly- $1/f$ processes discussed in Section 3.2.1, which, when bandpass filtered, yield *stationary* processes with nearly- $1/f$ spectra.

One approach to extending this model so as to incorporate this property of stationarity is to add phase jitter in the synthesis process. Specifically, we may consider randomizing the time-origin of our processes generated via (3.35) by applying a random (positive or negative) delay to the process. In fact, this is one way of interpreting (3.36) as the generalized spectrum of a stationary process. However, the random process $\tilde{x}(t)$ constructed in this way is not ergodic. Furthermore, if the coefficients x_n^m in Theorem 3.4 are chosen to be Gaussian, $x(t)$ will be necessarily a Gaussian process, but $\tilde{x}(t)$ will not. For these reasons, the phase-jittered process, while perhaps useful for synthesizing $1/f$ -like behavior, is difficult to exploit in analyzing $1/f$ -like behavior.

Some remarks concerning the conditions on the wavelet basis are also appropriate. We begin by noting that to generate $1/f$ -like behavior for $0 < \gamma < 2$, it suffices to use a wavelet basis for which the corresponding multiresolution analysis is at least *regular*. Again, virtually any practical wavelet basis satisfies this condition, even the Haar basis. However, the theorem implies that to generate $1/f$ -like behavior for $\gamma > 2$, higher regularity ($R > 1$) is required. This can be verified experimentally as well. We find, for instance, that when we attempt to synthesis $1/f$ -like behavior for $\gamma = 5$ using bases with $R \geq 3$, the sample functions are characterized by a smoothness consistent with the decay in their spectra. However, when bases corresponding to $R <$

3 are used in the synthesis, the sample functions lose their characteristic smoothness. Specifically, using a Haar-based synthesis ($R = 1$), the sample functions exhibit abrupt discontinuities, while using a second-order ($R = 2$) Daubechies basis leads to sample functions exhibiting abrupt discontinuities in their derivatives. In effect, unless there is sufficient regularity, the characteristics of the basis functions manifest themselves in the sample functions generated by the expansion. However, at least in this context, there would appear to be no benefit to using bases that have more regularity than required by the theorem.

We also remark that a much stronger theorem holds for the case $\gamma = 0$ in which the coefficients not are not only uncorrelated but have identical variances. In this case, constructing an expansion from such a collection of random variables in *any* orthonormal basis yields stationary white noise whose spectral density is the variance of the coefficients. In particular, for any wavelet basis we have

$$S_x(\omega) = \sigma^2 = \sigma^2 \sum_m |\Psi(2^{-m})|^2$$

when $\gamma = 0$ where the last equality is a restatement of the identity (2.9) and demonstrates the consistency of this case with (3.36).

- Finally, we remark that Theorem 3.4 may, in principle, be extended to $\gamma < 0$ provided the wavelet basis used in the synthesis has a sufficient number of vanishing moments. This can be deduced from the proof in Appendix B.3. However, this extension to the theorem was not incorporated principally because there would appear to be relatively few, if any, physical processes corresponding to negative γ .

Analysis

In this section, we derive a collection of complementary results to suggest that wavelet bases are equally useful in the analysis of $1/f$ processes. In particular, we provide both theoretical and empirical evidence suggesting that when $1/f$ -like processes are expanded in terms of orthonormal wavelet bases, the resulting wavelet coefficients are typically rather weakly correlated, particularly in contrast to the rather strong

correlation present in the original process. These results, combined with those of the last section, provide evidence of that such wavelet-based representations are robust characterizations of $1/f$ -like behavior.

Virtually all the results we obtain in this section are derived conveniently and efficiently in the frequency-domain. In anticipation of these derivations, we first establish the following proposition, whose proof is outlined in Appendix B.4.

Proposition 3.5 *Let $x(t)$ be a $1/f$ process whose spectral parameters, in the sense of Theorem 3.2, are σ_x^2 and $\gamma > 0$. Furthermore, let the wavelet coefficients x_n^m be the projections of $x(t)$ onto some orthonormal wavelet basis. Then the correlation between an arbitrary pair of such coefficients x_n^m and $x_{n'}^{m'}$ is given by*

$$E[x_n^m x_{n'}^{m'}] = \frac{2^{-(m+m')/2}}{2\pi} \int_{-\infty}^{\infty} \frac{\sigma_x^2}{|\omega|^\gamma} \Psi(2^{-m}\omega) \Psi^*(2^{-m'}\omega) e^{-j(n2^{-m}-n'2^{-m'})\omega} d\omega \quad (3.40)$$

for any choice of $\psi(t)$ and γ such that this integral is convergent.

This principal shortcoming of this proposition is that it fails to establish conditions on the wavelet basis and γ under which (3.40) is defined. Nevertheless, we may generally use Proposition 3.5 to derive properties of the second-order statistics of wavelet coefficients of $1/f$ processes for $\gamma > 0$. For instance, an immediate consequence of the proposition is that we can show the variance of the x_n^m to be of the form

$$\text{Var } x_n^m = \sigma^2 2^{-\gamma m}$$

where

$$\sigma^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\sigma_x^2}{|\omega|^\gamma} |\Psi(\omega)|^2 d\omega.$$

To obtain this result, it suffices to let $m' = m$ and $n' = n$ in (3.40) and effect a change of variables.

Defining

$$\rho_{n,n'}^{m,m'} \triangleq \frac{E[x_n^m x_{n'}^{m'}]}{\sqrt{(\text{Var } x_n^m)(\text{Var } x_{n'}^{m'})}} \quad (3.41)$$

as the *normalized* wavelet correlation, a second consequence is that the wavelet co-

efficients are wide-sense stationary at each scale, *i.e.*, for a fixed scale m , $\rho_{n,n'}^{m,m}$ is a function only of $n - n'$. Specifically, we may readily establish that

$$\rho_{n,n'}^{m,m} = \frac{1}{2\pi\sigma^2} \int_{-\infty}^{\infty} \frac{\sigma_x^2}{|\omega|^\gamma} |\Psi(\omega)|^2 e^{-j(n-n')\omega} d\omega. \quad (3.42)$$

Again, this result may be obtained by specializing (3.40) to the case $m' = m$ and effecting a change of variables.

We can also show that the normalized wavelet coefficients possess a kind of stationarity *across* scales as well. Recalling from Section 2.2.1 the critically-sampled filter bank interpretation of the wavelet decomposition, whereby the output of the m th filter was sampled at rate $t = 2^{-m}n$ for $n = \dots, -1, 0, 1, 2, \dots$, we note that a pair of wavelet coefficients x_n^m and $x_{n'}^{m'}$ at distinct scales m and m' correspond to synchronous time-instants precisely when

$$2^{-m}n = 2^{-m'}n'. \quad (3.43)$$

Our stationarity result in this case is that the normalized correlation among time-synchronous wavelet coefficients corresponding to scales m and m' is a function only of $m - m'$. More precisely, we can show that whenever (3.43) holds,

$$\rho_{n,n'}^{m,m'} = \frac{1}{2\pi\sigma^2} 2^{-(m-m')/2} \int_{-\infty}^{\infty} \frac{\sigma_x^2}{|\omega|^\gamma} \Psi(2^{-(m-m')}\omega) \Psi(\omega) d\omega. \quad (3.44)$$

Again, this result follows from specializing (3.40) and effecting a change of variables.

The above results verify that the wavelet coefficients of $1/f$ processes obey the variance progression anticipated from the synthesis result. Moreover, the stationarity results provide insight into the correlation structure among wavelet coefficients. However, what we seek ideally are good bounds on the magnitude of the correlation among wavelet coefficients both in the case that they reside at the same scale, and in the case they reside at distinct scales. Certainly, as we will see, there is strong empirical evidence that the correlation among coefficients is rather small and, in most cases, negligible. The following theorem provides some theoretical evidence by establishing

an asymptotic result. A proof is provided in Appendix B.5.

Theorem 3.6 *Consider an orthonormal wavelet basis such that $\psi(t)$ has R vanishing moments, i.e.,*

$$\Psi^{(r)}(\omega) = 0, \quad r = 0, 1, \dots, R - 1 \quad (3.45)$$

for some integer $R \geq 1$. Then provided $0 < \gamma < 2R$, the wavelet coefficients obtained by projecting a $1/f$ process onto this basis have a correlation whose magnitude decays according to²

$$|\rho_{n,n'}^{m,m'}| \sim \mathcal{O}(|2^{-m}n - 2^{-m'}n'|^{-\lceil 2R-\gamma \rceil}) \quad (3.46)$$

as

$$|2^{-m}n - 2^{-m'}n'| \rightarrow \infty.$$

While this theorem makes an interesting statement about the *relative* correlation among some wavelet coefficients well-separated in (m, n) -space, we must avoid inferring some stronger statements. First, it says nothing about the correlation among time-synchronous wavelet coefficients (*i.e.*, those satisfying (3.43)), regardless of how well-separated they are. Furthermore, while plausible, the theorem itself does not assert that choosing an analysis wavelet with a larger number of vanishing moments can reduce the correlation among wavelet coefficients in the analysis of $1/f$ processes. Likewise, the theorem does not actually validate the reasonable hypothesis that choosing a wavelet with an insufficient number of vanishing moments will lead to strong correlation among the wavelet coefficients of $1/f$ processes. In fact, the theorem identifies neither a range of m, m', n, n' over which (3.46) holds, nor a leading multiplicative constant in (3.46). Consequently, this precludes us from inferring anything about the *absolute* correlation between *any* particular pair of coefficients.

For the case of the ideal bandpass wavelet basis, however, we may obtain some more useful bounds on the correlation among wavelet coefficients. In this case, the basis functions corresponding to distinct scales have non-overlapping frequency support. Hence, carefully exploiting the stationarity properties of $1/f$ processes developed in

²The ceiling function $\lceil x \rceil$ denotes the smallest integer greater than or equal to x .

Theorem 3.2, we conclude that the wavelet coefficients corresponding to distinct scales are uncorrelated. However, at a given scale the correlation at integral lag $l \geq 0$ is non-zero and may be expressed as

$$\rho_{n,n-l}^{m,m} = \frac{\sigma_x^2}{\pi\sigma^2} \int_{-\pi}^{2\pi} \omega^{-\gamma} \cos(\omega l) d\omega \quad (3.47)$$

where

$$\frac{\sigma_x^2}{\sigma^2} = \begin{cases} (2^{1-\gamma} - 1)/(\pi^\gamma(1 - \gamma)) & \gamma \neq 1 \\ (\ln 2)/(\pi) & \gamma = 1 \end{cases}. \quad (3.48)$$

While (3.47) cannot be evaluated in closed form, integrating by parts twice and using the triangle inequality gives the useful closed-form bound

$$|\rho_{n,n-l}^{m,m}| \leq \frac{\sigma_x^2}{\sigma^2} \frac{\gamma}{l^2 \pi^{2+\gamma}} \left\{ 1 + \frac{1}{2^{1+\gamma}} + \frac{1+\gamma}{l\pi} \left[1 - \frac{1}{2^{2+\gamma}} \right] \right\} \quad (3.49)$$

valid for $\gamma \geq 0$ and integer-valued $l \geq 1$.

In Fig. 3-4, we plot the exact magnitude of the normalized correlation (3.47) obtained by numerical integration as a function of lag l together with the bound (3.49). Note that correlation among wavelet coefficients is extremely small: adjacent coefficients have a correlation coefficient of less than 15 percent, and more widely separated coefficients have a correlation coefficient less than three percent. Hence, it is not an unreasonable approximation to neglect the inter-coefficient correlation in any analysis using this wavelet basis.

On the same plot we superimpose the average along-scale sample-correlation between wavelet coefficients obtained from a $1/f$ -type process generated using Keshner's synthesis. In this simulation, a 65 536-sample segment of a $1/f$ process was generated for $\gamma = 1$ and analyzed using Daubechies 5th-order wavelet basis. Here, the sample-correlation function of the coefficients at each scale was computed, and averaged appropriately with the sample-correlation functions at the other scales. That the experimental result so closely matches the exact result for the bandlimited basis suggests that our analysis result for the bandlimited basis may, in fact, be more broadly applicable.

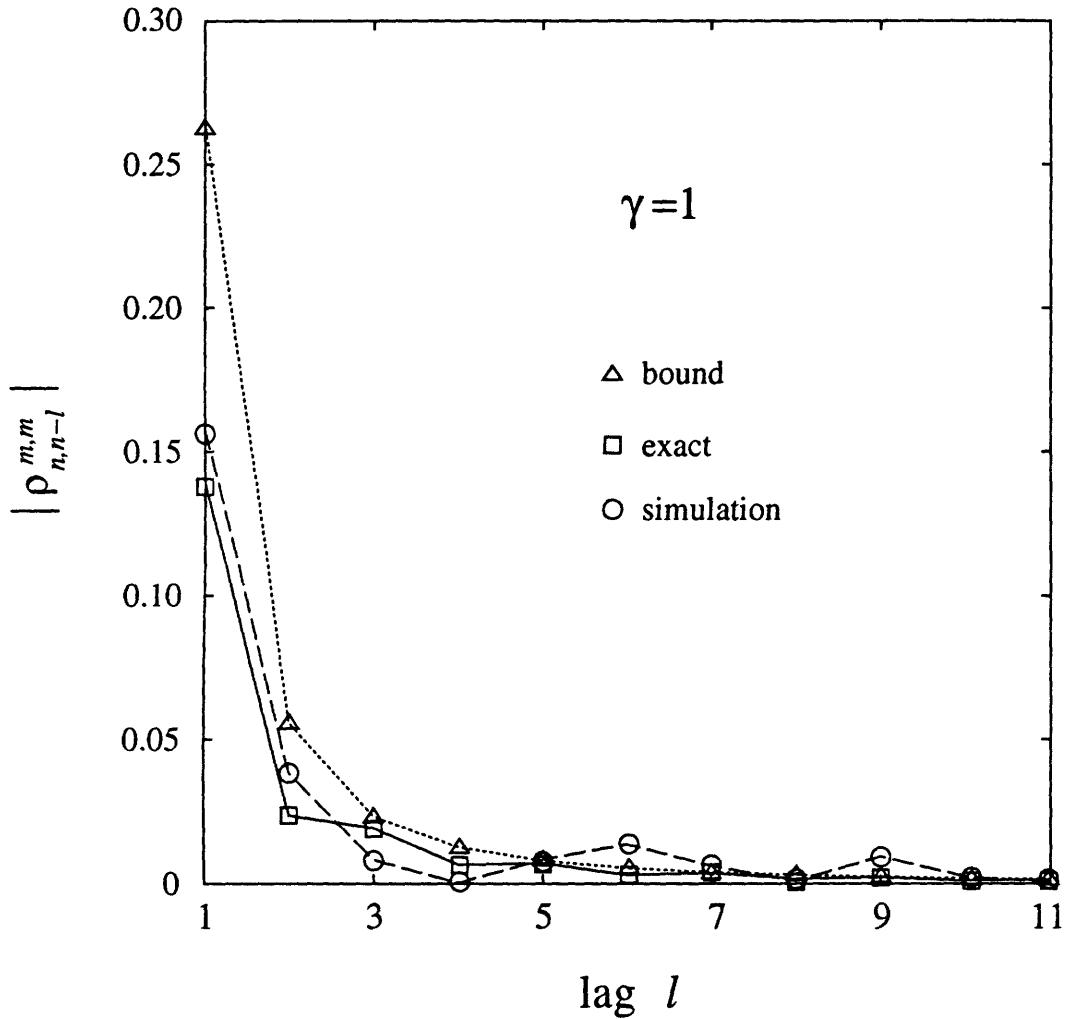


Figure 3-4: Along-scale correlation between wavelet coefficients for an exactly- $1/f$ process for which $\gamma = 1$. The squares \square indicate a numerical estimate of the exact magnitude of the normalized correlation between wavelet coefficients as a function of the lag l between them. The ideal bandpass wavelet was assumed in the analysis. The triangles \triangle indicate the corresponding values of the closed-form bound obtained in the text. The circles \circ show the average sample-correlation as computed from a projections of a $1/f$ process generated using Keshner's synthesis onto a 5th-order Daubechies wavelet basis.

Before concluding this analysis section, it is appropriate to note that several of the results we have described herein have been derived independently for the particular case of fractional Brownian motion using a time-domain approach. For instance, the stationarity of the wavelet coefficients at a fixed scale was first established by Flandrin in [29], while the inter-scale stationarity property was described by Flandrin (after Vergassola and Frisch [54]) in the recent work [55]. Likewise, the expression for the asymptotic rate-of-decay of correlation among wavelet coefficients is essentially the same as that first derived by Tewfik and Kim [56]. We also mention that Flandrin [55] is able to provide stronger statements about the correlation among wavelet coefficients of fractional Brownian motion for the specific case of the Haar wavelet basis. Finally, we remark that it ought to be possible to interpret the decorrelation results presented both here and in the works of the above authors in the context of more general results that have emerged concerning the effectiveness of wavelet decompositions in decorrelating a broad class of smooth covariance kernels [24].

Finally, it is useful to remark that through the octave-band filter bank interpretation of wavelet bases, we may view wavelet-based analysis as spectral analysis on a logarithmic frequency scale. The results of this section, together with our observations of the spectral characteristics of $1/f$ processes earlier in the chapter, suggest that this kind of spectral analysis is, in some sense, ideally matched to $1/f$ -type behavior. In the final section of this chapter, we undertake such a log-based spectral analysis of some real data sets using wavelets and show additional evidence that such analysis is potentially both useful and important in these cases.

Experiments

In this section, we undertake a very preliminary investigation of the properties of wavelet coefficients derived from some physical data sets. In the process, we identify two instances of time series that would appear to be well-modeled as $1/f$ processes. The first is an example involving economic data. Fig. 3-5 shows the time series corresponding to raw weekly Dow Jones Industrial Average data accumulated over the past 80 years. As shown in Fig. 3-6(a), the sample-variance of wavelet coefficients from

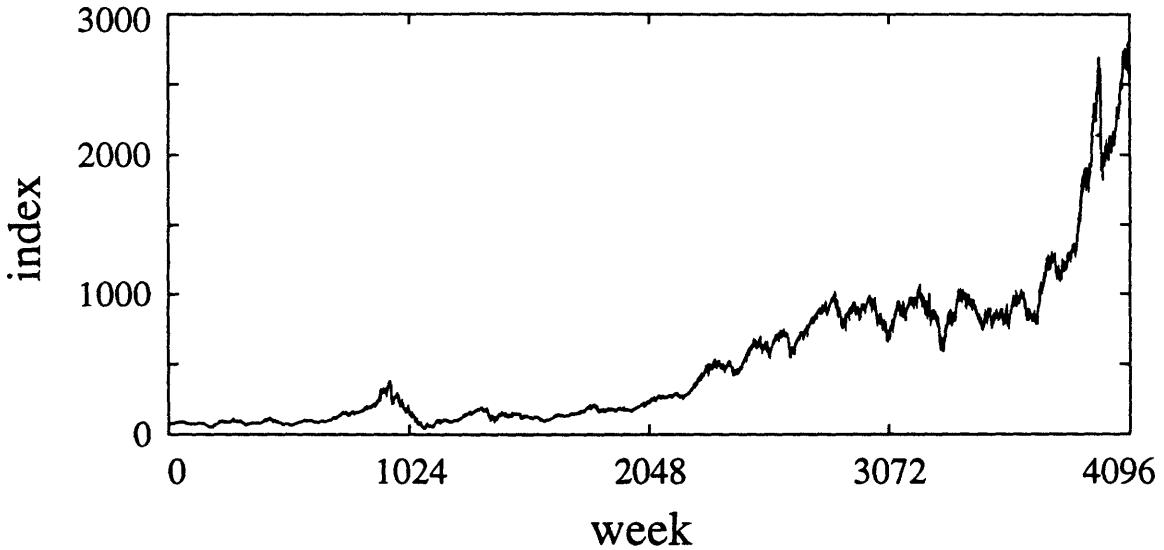
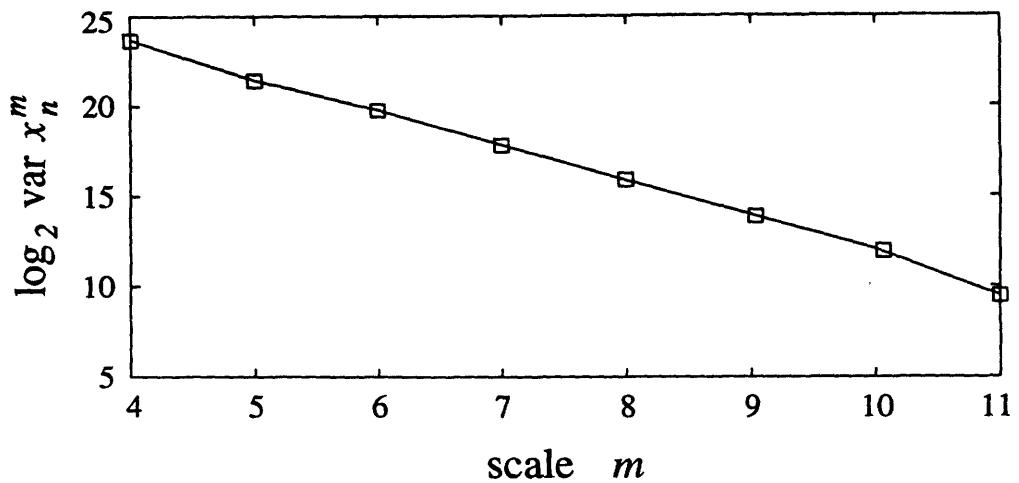


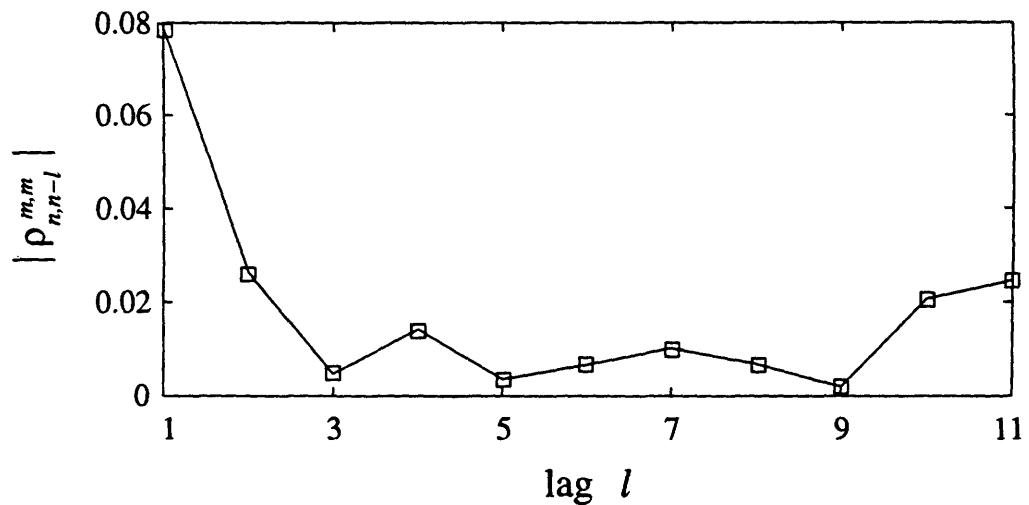
Figure 3-5: Weekly Dow Jones Industrial Average data, to present.

scale to scale obeys a geometric progression consistent with a $1/f$ process for which $\gamma \sim 2$. In Fig. 3-6(b), we see that the average along-scale sample-correlation among wavelet coefficients is rather weak. Since adjacent coefficients have a correlation of less than 8 percent, and more widely separated coefficients have a correlation of less than 3 percent, it would appear reasonable to neglect the inter-coefficient correlation in the analysis of such data. While this behavior is also consistent with a $1/f$ -type model for the data, we note that to justify such a model more fully, it would be necessary to study the correlation among coefficients between scales as well.

The second example involves physiological data. Fig. 3-7 shows a record of heart beat interarrival times for a healthy human patient corresponding to approximately 11 hours of continuously-acquired data. The quantization levels of the interarrival times are spaced four milliseconds apart. In this example, as shown in Fig. 3-8(a) the sample-variances of wavelet coefficients from scale to scale obeys a geometric progression consistent with a $1/f$ process of $\gamma \sim 1$. When viewing these progressions it is important to note that the number of samples available to make a variance estimate doubles at each successively finer scale. Hence, the standard deviation of the sample-variance measurement decreases by a factor of $\sqrt{2}$ for each successive increase in m . As a result, $1/f$ behavior manifests itself in the form of log-variance characteristic



(a) Scale-to-scale wavelet coefficient sample-variance progression.



(b) Average magnitude of the normalized along-scale sample-correlation between wavelet coefficients.

Figure 3-6: Wavelet-based analysis of weekly Dow Jones Industrial Average data. The time-series is analyzed using a 5th-order Daubechies wavelet basis.

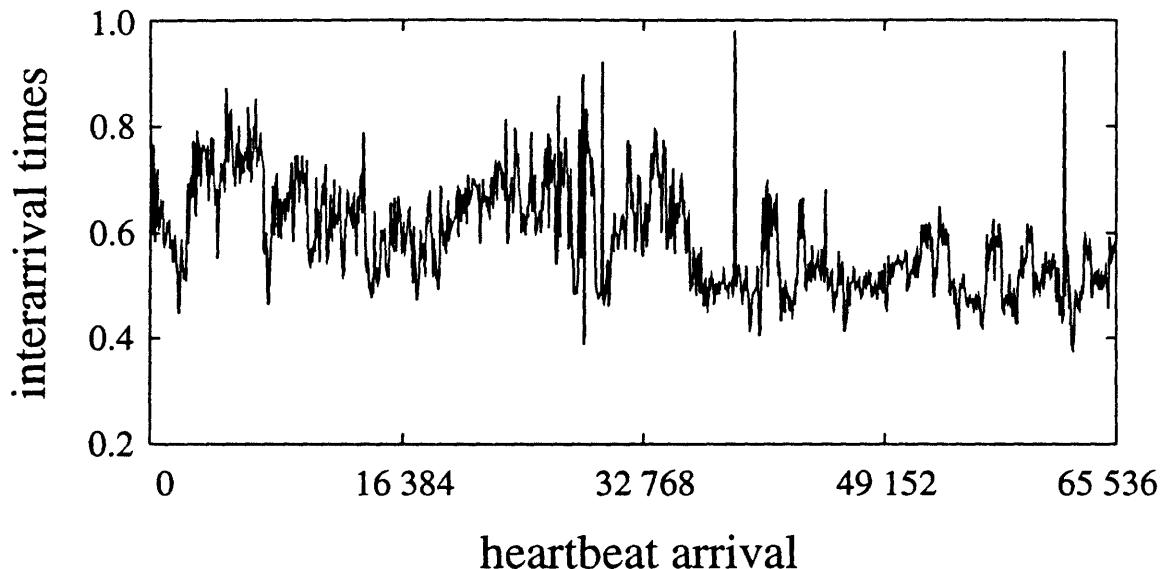
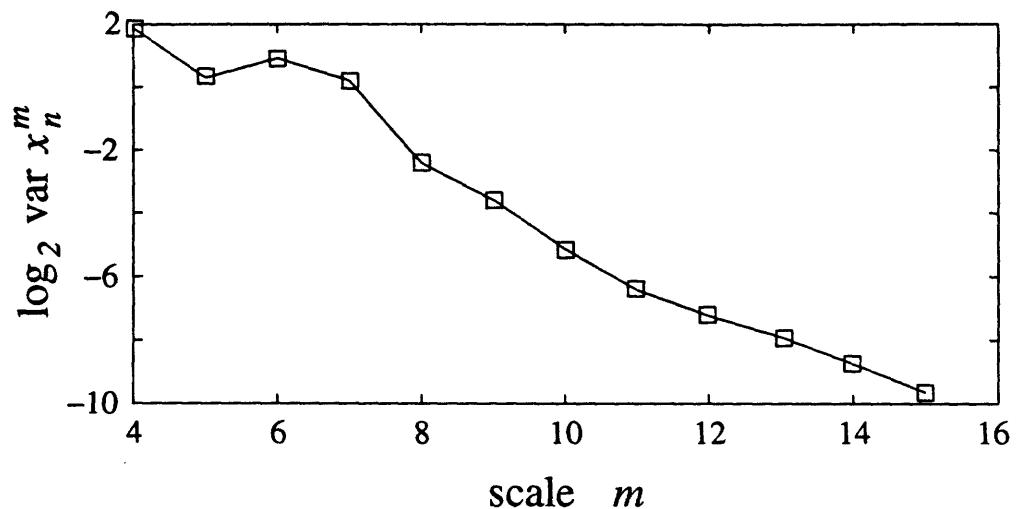


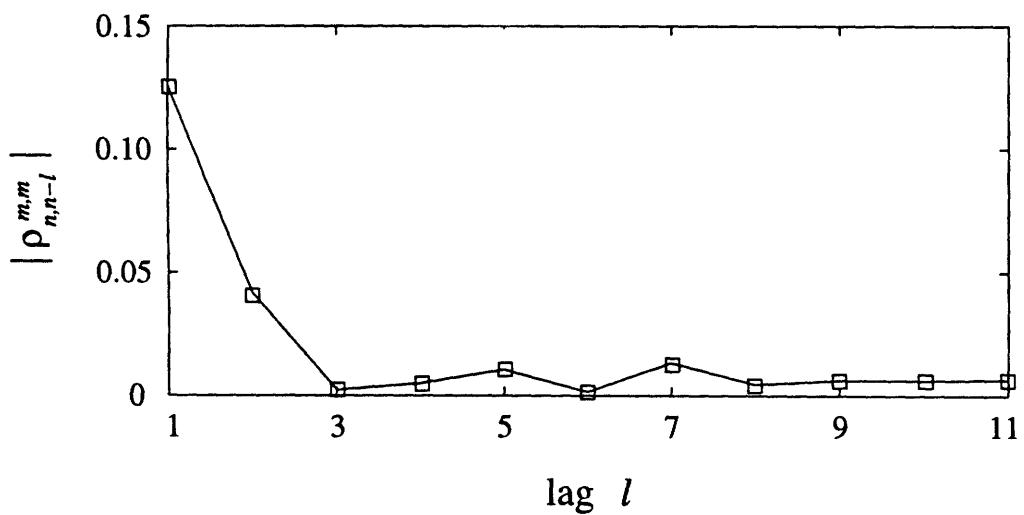
Figure 3-7: Heartbeat interarrival times for a healthy patient.

that must be asymptotically linear in the limit of large m . In Fig. 3-8(b), we show the weak average along-scale sample-correlation between wavelet coefficients. In this case, coefficients separated by lags of two or more are correlated less than 2 percent, again suggesting that it is reasonable to neglect such inter-coefficient correlation in any wavelet-based analysis. Again, we caution that no attempt was made to study the correlation structure among coefficients between scales.

Together, our theoretical and empirical results suggest that the orthonormal wavelet transform is a potentially useful and convenient tool in the synthesis and analysis of $1/f$ -type processes. In the next chapter we will explore how the wavelet transform plays an equally valuable role in the *processing* of such signals.



(a) Scale-to-scale wavelet coefficient sample-variance progression.



(b) Average magnitude of the normalized along-scale sample-correlation between wavelet coefficients.

Figure 3-8: Wavelet-based analysis of the heartbeat interarrival times for a health patient. The time-series is analyzed using a 5th-order Daubechies wavelet basis.

Chapter 4

Detection and Estimation with $1/f$ Processes

Given the ubiquity of physical signals exhibiting $1/f$ -type behavior, there are many application contexts in which one is interested in developing efficient algorithms for processing such signals. For instance, one is frequently interested in problems of

- detection and classification
- characterization and parametrization
- prediction and interpolation, and
- separation of $1/f$ signals both from one another as well as from other types of known or partially-known signals.

In some cases, the $1/f$ signal itself is of primary interest. An example would be the problem of modeling stock market data such as the Dow Jones Industrial Average as a $1/f$ process. In other cases, the $1/f$ signal represents a noise process obscuring some other signal of interest. This is more likely to be the case in optical and electronic systems, for example, where $1/f$ is a predominant form of background noise.

Even when the $1/f$ signal is of primary interest, one rarely has perfect access to such signals. Typically, our observations are incomplete. Indeed, they will generally be time-limited and resolution-limited. More generally, the observations may contain

gaps, or there may be multiple observations. Additionally, any observations of $1/f$ signals will invariably be corrupted by some degree of broadband measurement noise¹. It is important to both recognize and accommodate such measurement noise in any algorithms for processing such data. Indeed, because $1/f$ signals have a spectral density that vanishes for sufficiently high frequencies, there necessarily exists some frequency threshold beyond which the broadband noise is predominant. As a consequence, such noise can strongly affect the performance of signal processing algorithms.

In this chapter, we develop some optimal algorithms for addressing a number of basic signal processing problems involving detection and estimation with $1/f$ -type signals. Our basic approach is to exploit the properties of wavelet-based representations of $1/f$ -type processes. In particular, based upon the synthesis result of Theorem 3.4 and supported by the subsequent analysis results, our model for $1/f$ signals will be signals which when expanded into an orthonormal wavelet basis, yield coefficients that are effectively uncorrelated and obey the appropriate variance progression. That is, we will exploit the role of the wavelet expansion as a Karhunen-Loève-like expansion for $1/f$ -type processes. Because extremely efficient algorithms exist for computing orthonormal wavelet transformations as discussed in Section 2.2.3, this approach is not only analytically convenient for solving these signal processing problems, but leads to computationally highly efficient structures for implementing the resulting algorithms.

Throughout we will be careful to incorporate additive stationary white measurement noise for ensuring the robustness of our models as described above. Furthermore, most of the algorithms we develop will be designed specifically for the case of Gaussian $1/f$ processes and Gaussian measurement noises. While this requirement is principally motivated by tractability requirements, there are, in fact, many contexts in which this is a physically reasonable assumption. Furthermore, several of the algorithms we develop retain many of their important properties in the more general non-Gaussian case.

¹Actually, the coexistence of $1/f$ and white noises in electronic and optical systems is well-documented. In electronic systems, for instance, the predominant noise is $1/f$ noise at frequencies below about 1 kHz, while at higher frequencies, it is white noise in the form of thermal (*i.e.*, Johnson) and shot noise [57].

Accompanying each of the algorithms we develop is an evaluation of its performance in various settings. It is important to realize that such evaluations are necessarily of a highly preliminary nature, and are not meant to be interpreted as in any way comprehensive. Their primary function is to establish the basic viability of the algorithms, and to reveal some of their basic properties. Many of these performance studies involve Monte Carlo simulations with synthetic data. For these simulations, we generate $1/f$ processes using the Corsini-Saletti implementation of Keshner's synthesis described in [52]. Because this synthesis is fundamentally different from a wavelet-based synthesis, such simulations play an important role in verifying the robustness of the wavelet-based algorithms with respect to our particular model for $1/f$ -type behavior. However, by their design, these simulations will generally not enable us to isolate the effects of modeling error alone.

Additionally, there are a large number of wavelet bases from which to select for our algorithms. However, given the apparent insensitivity of the wavelet-based model for $1/f$ -type behavior to the choice of basis, for our simulations we choose, somewhat arbitrarily, to use the basis corresponding to Daubechies' 5th-order finite-extent maximally-regular wavelet for which the corresponding conjugate quadrature filters have 10 non-zero coefficients. We remark that in addition to being realizable, this basis satisfies the conditions of the theorems of Section 3.2.2 concerning the synthesis and analysis of $1/f$ -type behavior using wavelets. Specifically, the basis has more than enough vanishing moments to accommodate spectral parameters in our principal range of interest, $0 < \gamma < 2$.

Until recently, such problems of detection and estimation involving $1/f$ -type processes received relatively little attention in the literature. However, there has been strongly increasing interest in the topic and a number of interesting and useful related results have been developed. An important instance is the elegant work of Barton and Poor [42], who consider problems of detection in the presence of fractional Gaussian noise using reproducing kernel Hilbert space theory. Using this framework, the authors are able to develop both infinite- and finite-interval whitening filters for this class of $1/f$ noises, which, in turn, allow them to obtain some important results on

the detection of deterministic and Gaussian signals in the presence of such noise.

There is also a substantial and growing body of recent literature on the general topic of multiresolution stochastic processes, systems, and signal processing. An excellent overview of work in this area is contained in [58] and the references therein. In this work, the authors develop a clever tree- or lattice-based framework for modeling multiscale processes and problems, and introduce some novel notions of “stationarity in scale” for such processes. Treating multiscale processes as “dynamical systems in scale,” lead to several highly efficient algorithms for addressing a variety of problems involving parameter estimation, signal smoothing and interpolation, and data fusion. The $1/f$ -type models we exploit in this chapter constitute a special class of the multiresolution stochastic processes developed by these authors. In particular, they are examples of processes characterized by a “Markov scale-to-scale” property. As a consequence, many of the multiresolution signal processing algorithms they develop are directly applicable to $1/f$ processes as shown, *e.g.*, in [23].

There are also interesting parallels between the results of this chapter and recent work by Tewfik and Kim. For example, in [59], these authors develop the notion suggested by Beylkin, Coifman and Rokhlin in [24], that wavelet-based and other, more general, filter bank decompositions are useful in transforming the correlation structure in a broad class of stationary and nonstationary processes (including, *e.g.*, a class of $1/f$ processes) into a form more convenient for signal processing. Specifically, they show how the structure of the correlation in such decompositions leads directly to computationally efficient, general purpose algorithms for signal processing. As an example, [60] summarizes how their results can be applied to the problem of signal detection in the presence of fractional Brownian motion.

4.1 $1/f$ Synthesis and Whitening Filters

Many of the results on detection and estimation we will derive in this chapter are conveniently interpreted in a canonical form through the concept of a reversible (or invertible) whitening filter for $1/f$ processes. In this section, we derive such whitening

filters and their inverses for the particular wavelet-based model for $1/f$ -type behavior which we intend to exploit in this chapter.

To begin, if $x(t)$ is a $1/f$ signal corresponding to some spectral exponent γ , we model the corresponding wavelet coefficients x_n^m as zero-mean random variables having negligible correlation and obeying a variance progression of the form

$$\text{Var } x_n^m = \sigma^2 \beta^{-m}$$

where, for notational convenience, we define

$$\beta = 2^\gamma. \quad (4.1)$$

In turn, we may express the x_n^m as

$$x_n^m = \sigma \beta^{-m/2} v_n^m$$

where the v_n^m are then zero-mean, unit-variance, uncorrelated random variables. Hence, the process $v(t)$ defined according to

$$v(t) = \sum_m \sum_n v_n^m \psi_n^m(t)$$

is a wide-sense stationary white noise process since the $\psi_n^m(t)$ constitute a complete orthonormal set. This suggests that we may model $x(t)$ as the output of a linear system driven by stationary white noise $v(t)$. In particular, the system performs an orthonormal wavelet transform on the input $v(t)$, scales each of the resulting coefficients v_n^m by a factor

$$k_n^m = \sigma \beta^{-m/2}$$

then inverse wavelet transforms the resulting x_n^m to generate the output process $x(t)$. This $1/f$ synthesis filter, defined via

$$x(t) = \mathcal{W}_d^{-1} \left\{ \sigma \beta^{-m/2} \mathcal{W}_d \{ v(t) \} \right\} \quad (4.2)$$

is a linear filter whose kernel² is

$$\kappa_s^\circ(t, \tau) = \sum_m \sum_n \psi_n^m(t) \sigma \beta^{-m/2} \psi_n^m(\tau). \quad (4.3a)$$

We emphasize that viewing $x(t)$ as the output of a linear system with kernel (4.3a) driven by stationary white noise $v(t)$ is especially useful in the Gaussian scenario, in which case $w(t)$ is a stationary white Gaussian process. Nevertheless, for non-Gaussian processes this characterization remains useful at least insofar as modeling the second-order properties of $x(t)$ is concerned.

From the wavelet-based characterization of the synthesis filter (4.2) we readily deduce that this filter is invertible, and that its inverse has kernel

$$\kappa_w^\circ(t, \tau) = \sum_m \sum_n \psi_n^m(t) \frac{1}{\sigma \beta^{-m/2}} \psi_n^m(\tau). \quad (4.3b)$$

This is, therefore, the corresponding *whitening filter* for our model of $1/f$ -type behavior. Indeed, when this filter is driven by a process obtained as the output of our $1/f$ synthesis filter, the output is, evidently, a wide-sense stationary white process. When driven by an exactly- $1/f$ process, the properties of the output are readily described in terms of the analysis results of Section 3.2.2.

As discussed at the outset of the chapter, any $1/f$ -type process we consider shall invariably be accompanied by an additive stationary white observation noise component. Consequently, we shall frequently find the notion of synthesis and whitening filters for the combined $1/f$ -plus-white processes convenient in interpreting our algorithms. These filters are, of course, closely related to the filters derived above. In fact, it is straightforward to establish that synthesis and whitening filters for $1/f$ -

²In our notation, the kernel $k(t, \tau)$ of a linear system defines the response of the system at time t to a unit impulse at time τ . Consequently the response of the system to a suitable input $x(t)$ is expressed as

$$y(t) = \int_{-\infty}^{\infty} x(\tau) k(t, \tau) d\tau.$$

plus-white processes are characterized by the respective kernels

$$\kappa_s(t, \tau) = \sum_m \sum_n \psi_n^m(t) \sigma_m \psi_n^m(\tau) \quad (4.4a)$$

$$\kappa_w(t, \tau) = \sum_m \sum_n \psi_n^m(t) \frac{1}{\sigma_m} \psi_n^m(\tau) \quad (4.4b)$$

where $\sigma_m > 0$ is defined by

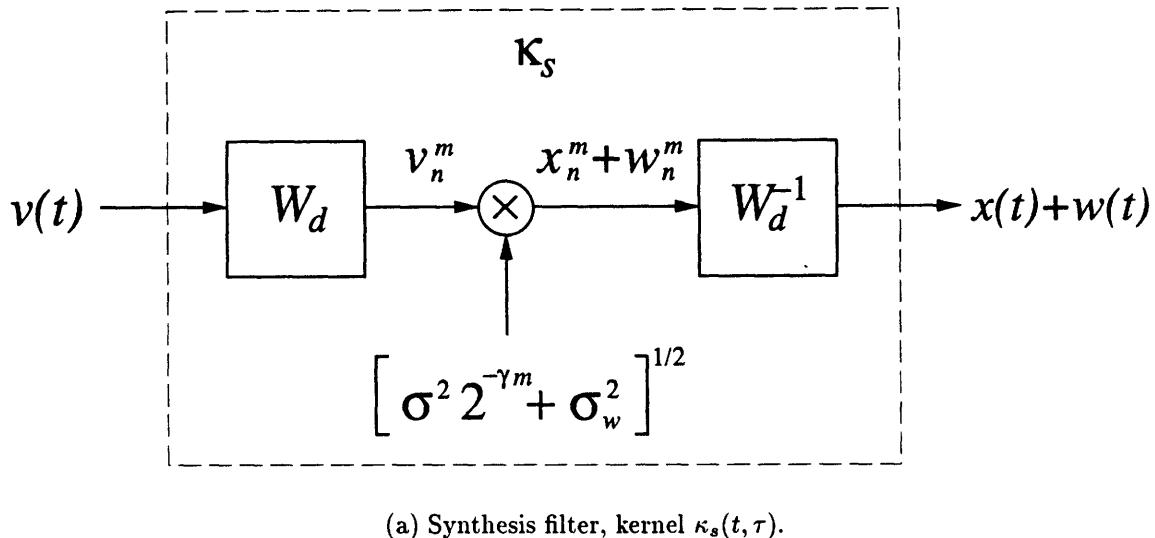
$$\sigma_m^2 = \sigma^2 \beta^{-m} + \sigma_w^2 \quad (4.5)$$

and σ_w^2 is the spectral density of the white noise component. The canonical wavelet-based realization of these filters is depicted in Fig. 4-1.

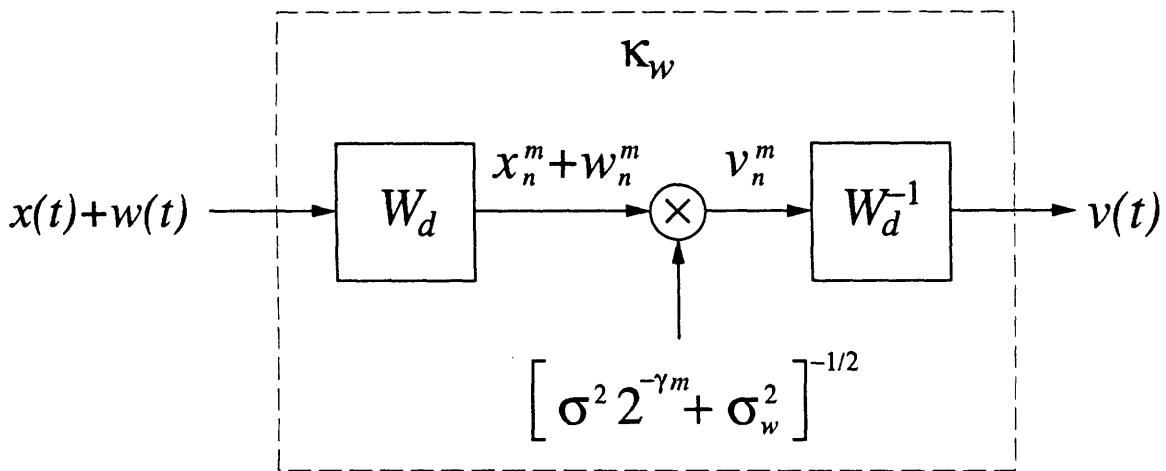
4.2 Parameter Estimation for $1/f$ Signals

In this section, we consider the problem of estimating the parameters of a Gaussian $1/f$ signal from observations corrupted by stationary white Gaussian noise. Since, typically, we lack *a priori* knowledge of the spectral density of the noise, we consider, more generally, the problem of jointly estimating signal and noise parameters for this scenario.

Such parameter estimates, in addition to providing a solution to the associated $1/f$ spectrum estimation problem, are frequently of interest in their own right. Indeed, from the parameter estimates, we can directly compute the fractal dimension of the underlying signal, when defined. Robust estimation of the fractal dimension of $1/f$ processes is important in a number of applications such as in signal detection and classification. For example, in image processing, where 2-D extensions of $1/f$ processes are used to model natural terrain and other patterns and textures [32] [43], fractal dimension can be of use in distinguishing among various man-made and natural objects. While several approaches to the fractal dimension estimation problem have been presented previously in the literature (see [43], [61], [62], and the references therein), to date none has been able to adequately handle the presence of broadband noise in the observation data. In fact, the quality of the estimates generally deteriorates dramatically in the presence of such noise even at high SNR [43]. Since noise is



(a) Synthesis filter, kernel $\kappa_s(t, \tau)$.



(b) Whitening filter, kernel $\kappa_w(t, \tau)$.

Figure 4-1: Canonical form realizations of synthesis and whitening filters for processes that are the superposition of $1/f$ -type and white components, i.e., $1/f$ -plus-white processes.

inherently present in any real data, this lack of robustness has limited the usefulness of these algorithms. In this section we obtain, indirectly, fractal dimension estimators for Gaussian $1/f$ processes that explicitly take into account the presence of additive white Gaussian observation noise. The resulting iterative estimation algorithms are computationally efficient, robust, and statistically consistent.

Our basic approach is to apply the method of Maximum Likelihood (ML) estimation, exploiting the wavelet-based characterization of our $1/f$ model. And, while we specifically consider the case of Gaussian $1/f$ processes corrupted by additive stationary white Gaussian measurement noise in our formulation of the problem, we emphasize that the resulting estimators are, in fact, applicable to a broader class of non-Gaussian $1/f$ processes and measurement noise models, and retain many desirable properties.

We formulate our problem as follows. Suppose we have observations $r(t)$ of a zero-mean Gaussian $1/f$ process $x(t)$ embedded in zero-mean additive stationary white Gaussian noise $w(t)$ that is statistically independent of $x(t)$, so

$$r(t) = x(t) + w(t), \quad -\infty < t < \infty. \quad (4.6)$$

From this continuous-time data, we assume we have extracted a number of wavelet coefficients, r_n^m . In theory, we may assume these coefficients are obtained by projecting the wavelet basis functions onto the observed data:

$$r_n^m = \int_{-\infty}^{\infty} \psi_n^m(t) r(t) dt.$$

However, in practice, these coefficients can be obtained by applying the computationally efficient DWT to the samples of a segment of data which is both time-limited and resolution-limited, as described in Section 2.2.3. Let us assume that the finite set of available distinct scales, \mathcal{M} , is, in increasing order,

$$\mathcal{M} = \{m_1, m_2, \dots, m_M\}, \quad (4.7a)$$

and that at each scale m the set of available coefficients $\mathcal{N}(m)$ is³

$$\mathcal{N}(m) = \{n_1(m), n_2(m), \dots, n_{N(m)}(m)\}. \quad (4.7b)$$

Hence, the data available to the estimation algorithm is

$$\mathbf{r} = \{r_n^m \in \mathcal{R}\} = \{r_n^m, m \in \mathcal{M}, n \in \mathcal{N}(m)\}. \quad (4.8)$$

We remark before proceeding that, based on the discussion in Section 2.2.4, for an implementation via the DWT with $N = N_0 2^M$ samples of observed data, we have, typically,

$$\mathcal{M} = \{1, 2, \dots, M\} \quad (4.9a)$$

$$\mathcal{N}(m) = \{1, 2, \dots, N_0 2^{m-1}\}, \quad (4.9b)$$

where N_0 is a constant that depends on the length of the filter $h[n]$. Consequently, while many of the results we derive will be applicable to the more general scenario, we will frequently specialize our results to this case.

Exploiting the Karhunen-Loève-like properties of the wavelet decomposition for $1/f$ -type processes, and using the fact that the w_n^m are independent of the x_n^m and are decorrelated for any wavelet basis, the resulting observation coefficients

$$r_n^m = x_n^m + w_n^m$$

can be modeled as mutually-independent zero-mean, Gaussian random variables with variance

$$\text{Var } r_n^m = \sigma_m^2 = \sigma^2 \beta^{-m} + \sigma_w^2$$

where β is defined in terms of the spectral exponent γ of the $1/f$ process according

³Note that, without loss of generality we may assume $\mathcal{N}(m) \neq \emptyset$, any m , or else the corresponding scale m could be deleted from \mathcal{M} .

to (4.1). Hence, it is the parameter set

$$\Theta = (\beta, \sigma^2, \sigma_w^2)$$

we wish to estimate. As discussed at the outset, it is often the case that only β or some function of β such as the spectral exponent γ , the fractal dimension D , or the self-similarity parameter H , is of interest. Nevertheless, σ^2 and σ_w^2 will still need to be estimated simultaneously as they are rarely known *a priori*. Furthermore, ML estimates of γ, D, H are readily derived from the ML estimate $\hat{\beta}_{\text{ML}}$. Indeed, since each of these parameters is related to β through an invertible transformation, we have

$$\hat{\gamma}_{\text{ML}} = \log_2 \hat{\beta}_{\text{ML}} \quad (4.10a)$$

$$\hat{D}_{\text{ML}} = (5 - \hat{\gamma}_{\text{ML}})/2 \quad (4.10b)$$

$$\hat{H}_{\text{ML}} = (\hat{\gamma}_{\text{ML}} - 1)/2. \quad (4.10c)$$

Proceeding, we express the likelihood as a function of the parameters by

$$\mathcal{L}(\Theta) = p_{\mathbf{r}}(\mathbf{r}; \Theta) = \prod_{m,n \in \mathcal{R}} \frac{1}{\sqrt{2\pi\sigma_m^2}} \exp \left[-\frac{(r_n^m)^2}{2\sigma_m^2} \right]$$

for which the log-likelihood function is

$$L(\Theta) = \ln p_{\mathbf{r}}(\mathbf{r}; \Theta) = -\frac{1}{2} \sum_{m,n \in \mathcal{R}} \left\{ \frac{1}{\sigma_m^2} (r_n^m)^2 + \ln(2\pi\sigma_m^2) \right\}.$$

Equivalently,

$$L(\Theta) = -\frac{1}{2} \sum_{m \in \mathcal{M}} N(m) \left\{ \frac{\hat{\sigma}_m^2}{\sigma_m^2} + \ln(2\pi\sigma_m^2) \right\} \quad (4.11)$$

where the M sample-variances

$$\hat{\sigma}_m^2 = \frac{1}{N(m)} \sum_{n \in \mathcal{N}(m)} (r_n^m)^2 \quad (4.12)$$

summarize the aspects of the data required in the estimation. It is straightforward to show that the likelihood function in this case is well-behaved and bounded from above on

$$\beta \geq 0, \sigma^2 \geq 0, \sigma_w^2 \geq 0$$

so that, indeed, maximizing the likelihood function is reasonable.

While we shall assume that $\beta, \sigma^2, \sigma_w^2$ are all unknown, it will be appropriate during the development to also specialize results to the case in which σ_w^2 is known. Still more specific results will be described when $\sigma_w^2 = 0$, corresponding to the case of noise-free observations. We may also assume, where necessary, that all $m \in \mathcal{M}$ are positive without loss of generality. Indeed, if, for example, $m_1 < 0$, then we could define new parameters through the invertible transformation

$$\begin{aligned}\tilde{\sigma}_w^2 &= \sigma_w^2 \\ \tilde{\sigma}^2 &= \sigma^2 \beta^{m_1 - 1} \\ \tilde{\beta} &= \beta\end{aligned}$$

for which the observations correspond to positive scales

$$\tilde{\mathcal{M}} = \{1, m_2 - m_1 + 1, \dots, m_M - m_1 + 1\}$$

and which lead to the same ML estimates for $\beta, \sigma^2, \sigma_w^2$.

4.2.1 Case I: $\beta, \sigma^2, \sigma_w^2$ Unknown

Differentiating $L(\Theta)$ with respect to σ_w^2, σ^2 , and β , respectively, it follows that the stationary points of $L(\Theta)$ are given as the solutions to the equations

$$\begin{aligned}\sum_{m \in \mathcal{M}} T_m &= 0 \\ \sum_{m \in \mathcal{M}} \beta^{-m} T_m &= 0 \\ \sum_{m \in \mathcal{M}} m \beta^{-m} T_m &= 0\end{aligned}$$

where

$$T_m \triangleq \frac{N(m)}{\sigma_m^2} \left[1 - \frac{\hat{\sigma}_m^2}{\sigma_m^2} \right].$$

However, these equations are difficult to solve, except in special cases. Consequently, we utilize an iterative estimate-maximize (EM) algorithm [63].

A detailed development of the EM algorithm for our problem is given in Appendix C. The essential steps of the algorithm are summarized below, where we denote the estimates of the parameters $\beta, \sigma^2, \sigma_w^2$ generated on the l th iteration by $\hat{\beta}^{[l]}, \hat{\sigma}^{2[l]}, \hat{\sigma}_w^{2[l]}$.

E step: As shown in Appendix C, this step reduces to estimating the noise and signal portions of the wavelet coefficient variances at each scale $m \in \mathcal{M}$ using current estimates of the parameters $\hat{\beta}^{[l]}, \hat{\sigma}^{2[l]}, \hat{\sigma}_w^{2[l]}$:

$$S_m^w(\hat{\Theta}^{[l]}) = A_m(\hat{\Theta}^{[l]}) + B_m^w(\hat{\Theta}^{[l]})\hat{\sigma}_m^2 \quad (4.13a)$$

$$S_m^x(\hat{\Theta}^{[l]}) = A_m(\hat{\Theta}^{[l]}) + B_m^x(\hat{\Theta}^{[l]})\hat{\sigma}_m^2 \quad (4.13b)$$

where

$$A_m(\hat{\Theta}^{[l]}) = \frac{\hat{\sigma}_w^{2[l]} \cdot \hat{\sigma}^{2[l]} [\hat{\beta}^{[l]}]^{-m}}{\hat{\sigma}_w^{2[l]} + \hat{\sigma}^{2[l]} [\hat{\beta}^{[l]}]^{-m}} \quad (4.14a)$$

$$B_m^w(\hat{\Theta}^{[l]}) = \left(\frac{\hat{\sigma}_w^{2[l]}}{\hat{\sigma}_w^{2[l]} + \hat{\sigma}^{2[l]} [\hat{\beta}^{[l]}]^{-m}} \right)^2 \quad (4.14b)$$

$$B_m^x(\hat{\Theta}^{[l]}) = \left(\frac{\hat{\sigma}^{2[l]} [\hat{\beta}^{[l]}]^{-m}}{\hat{\sigma}_w^{2[l]} + \hat{\sigma}^{2[l]} [\hat{\beta}^{[l]}]^{-m}} \right)^2. \quad (4.14c)$$

M step: This step reduces to using these signal and noise variance estimates to obtain the new parameter estimates $\hat{\beta}^{[l+1]}, \hat{\sigma}^{2[l+1]}, \hat{\sigma}_w^{2[l+1]}$:

$$\hat{\beta}^{[l+1]} \leftarrow \sum_{m \in \mathcal{M}} C_m N(m) S_m^x(\hat{\Theta}^{[l]}) \beta^m = 0 \quad (4.15a)$$

$$\hat{\sigma}^{2[l+1]} = \frac{\sum_{m \in \mathcal{M}} N(m) S_m^x(\hat{\Theta}^{[l]}) [\hat{\beta}^{[l+1]}]^m}{\sum_{m \in \mathcal{M}} N(m)} \quad (4.15b)$$

$$\hat{\sigma}_w^{2[l+1]} = \frac{\sum_{m \in \mathcal{M}} N(m) S_m^w(\hat{\Theta}^{[l]})}{\sum_{m \in \mathcal{M}} N(m)} \quad (4.15c)$$

where

$$C_m \triangleq \frac{m}{\sum_{m \in \mathcal{M}} m N(m)} - \frac{1}{\sum_{m \in \mathcal{M}} N(m)}. \quad (4.16)$$

4.2.2 Case II: β, σ^2 Unknown; σ_w^2 Known

If σ_w^2 is known, the above algorithm simplifies somewhat. In particular, we may omit the estimation (4.15c) and replace occurrences of $\hat{\sigma}_w^{2[l]}$ in the algorithm with the true value σ_w^2 . This eliminates the need to compute $S_m^w(\hat{\Theta}^{[l]})$ and, hence, $B_m^w(\hat{\Theta}^{[l]})$. The resulting algorithm is as follows.

E step: Estimate the signal portion of the wavelet coefficient variances at each scale $m \in \mathcal{M}$ using current estimates of the parameters $\hat{\beta}^{[l]}, \hat{\sigma}^{2[l]}$:

$$S_m^x(\hat{\Theta}^{[l]}) = A_m(\hat{\Theta}^{[l]}) + B_m^x(\hat{\Theta}^{[l]}) \hat{\sigma}_m^2 \quad (4.17)$$

where

$$A_m(\hat{\Theta}^{[l]}) = \frac{\sigma_w^2 \cdot \hat{\sigma}^{2[l]} [\hat{\beta}^{[l]}]^{-m}}{\sigma_w^2 + \hat{\sigma}^{2[l]} [\hat{\beta}^{[l]}]^{-m}} \quad (4.18a)$$

$$B_m^x(\hat{\Theta}^{[l]}) = \left(\frac{\hat{\sigma}^{2[l]} [\hat{\beta}^{[l]}]^{-m}}{\sigma_w^2 + \hat{\sigma}^{2[l]} [\hat{\beta}^{[l]}]^{-m}} \right)^2. \quad (4.18b)$$

M step: Use these signal variance estimates to obtain the new parameter estimates $\hat{\beta}^{[l+1]}, \hat{\sigma}^{2[l+1]}$:

$$\hat{\beta}^{[l+1]} \leftarrow \sum_{m \in \mathcal{M}} C_m N(m) S_m^x(\hat{\Theta}^{[l]}) \beta^m = 0 \quad (4.19a)$$

$$\hat{\sigma}^2[l+1] = \frac{\sum_{m \in \mathcal{M}} N(m) S_m^x(\hat{\Theta}^{[l]}) [\hat{\beta}^{[l+1]}]^m}{\sum_{m \in \mathcal{M}} N(m)} \quad (4.19b)$$

where C_m is as in (4.16).

4.2.3 Case III: β, σ^2 Unknown; $\sigma_w^2 = 0$

If σ_w^2 is known (or assumed) to be zero, the EM algorithm becomes unnecessary as the likelihood may be maximized directly. Specifically, with $\sigma_w^2 = 0$, the signal variance estimates are available directly as $\hat{\sigma}_m^2$. Hence the estimation simplifies to the following:

$$\hat{\beta}_{ML} \leftarrow \sum_{m \in \mathcal{M}} C_m N(m) \hat{\sigma}_m^2 \beta^m = 0 \quad (4.20a)$$

$$\hat{\sigma}_{ML}^2 = \frac{\sum_{m \in \mathcal{M}} N(m) \hat{\sigma}_m^2 [\hat{\beta}_{ML}]^m}{\sum_{m \in \mathcal{M}} N(m)} \quad (4.20b)$$

with C_m still as in (4.16).

It is worth discussing this special case in more detail not only for its own sake, but also because it characterizes one of the components of each iteration of the EM algorithm. The derivation of the parameter estimates in this case is essentially the same as the derivation of the M step in the Appendix. We begin by differentiating the likelihood function to find equations for its stationary points. This leads to a pair of equations in terms of σ^2 and β . Eliminating σ^2 from these equations is straightforward and gives (4.20a) directly. Having determined $\hat{\beta}_{ML}$ as the solution to this polynomial equation, $\hat{\sigma}_{ML}^2$ is obtained by back substitution.

From Lemma C.1 in Appendix C, it is apparent that (4.20a) has exactly one positive real solution, which is the ML estimate $\hat{\beta}_{ML}$. Hence, L has a unique local and hence global maximum. Moreover, we may use bisection as a method to find the solution to this equation, provided we start with an initial interval containing $\hat{\beta}_{ML}$. For instance, when we expect $0 < \gamma < 2$, an appropriate initial interval is

$1 < \beta < 4$. Naturally, with some caution, Newton iterations may be used to accelerate convergence.

Again, since solving equations of the form of (4.20) constitutes the M step of the iterative algorithm for the more general problem, the above remarks are equally applicable in those contexts.

4.2.4 Properties of the Estimators

In this section, we consider two principal issues:

- how the parameter estimates of the EM algorithm converge to the ML parameter estimates, and
- how the ML parameter estimates converge to the true parameter values.

Regarding the first of these issues, we are assured that the EM algorithm always adjusts the parameter estimates at each iteration so as to increase the likelihood function until a stationary point is reached. It can be shown that in our problem, the likelihood function has multiple stationary points, one of which corresponds to the desired ML parameter estimates. Others correspond to rather pathological saddle points of the likelihood function at the boundaries of the parameter space:

$$\begin{aligned}\hat{\beta} &= \hat{\beta}_{\text{ML}} \Big|_{\sigma_w^2=0} \\ \hat{\sigma}^2 &= \hat{\sigma}_{\text{ML}}^2 \Big|_{\sigma_w^2=0} \\ \hat{\sigma}_w^2 &= 0\end{aligned}$$

and

$$\begin{aligned}\hat{\beta} &: \text{ arbitrary} \\ \hat{\sigma}^2 &= 0 \\ \hat{\sigma}_w^2 &= \frac{\sum_{m \in \mathcal{M}} N(m) \hat{\sigma}_m^2}{\sum_{m \in \mathcal{M}} N(m)}.\end{aligned}$$

That they are saddle points is rather fortunate, for the only way they are reached is if the starting value for any one of $\hat{\beta}$, $\hat{\sigma}^2$, $\hat{\sigma}_w^2$ is chosen to be exactly zero. Given arbitrarily small positive choices for these initial parameters, the algorithm will iterate towards the ML parameters.

The preceding discussion suggests that the EM algorithm is fundamentally rather robust in this application. However, the selection of the initial parameter values will naturally affect the rate of convergence of the algorithm. Moreover, it should be noted that the EM algorithm converges substantially faster for the case in which σ_w^2 is known. In essence, for the general algorithm much of the iteration is spent locating the noise threshold in the data.

We now turn to a discussion of the properties of the ML estimates themselves. It is well-known that ML estimates are generally asymptotically efficient and consistent. This, specifically, turns out to be the case here. It is also the case that at least in some higher signal-to-noise ratio (SNR) scenarios, the Cramér-Rao bounds closely approximate the true estimation error variances.

To compute the Cramér-Rao bounds for the estimates of γ , σ^2 , and σ_w^2 , we construct the corresponding Fisher matrix

$$\mathbf{I} = \sum_{m \in \mathcal{M}} \frac{N(m)}{2(\sigma_m^2)^2} \begin{bmatrix} [\ln 2^m \sigma^2 \beta^{-m}]^2 & -\ln 2^m \sigma^2 [\beta^{-m}]^2 & -\ln 2^m \sigma^2 \beta^{-m} \\ -\ln 2^m \sigma^2 [\beta^{-m}]^2 & [\beta^{-m}]^2 & \beta^{-m} \\ -\ln 2^m \sigma^2 \beta^{-m} & \beta^{-m} & 1 \end{bmatrix} \quad (4.21)$$

from which we get

$$\begin{aligned} \text{Var } \hat{\gamma} &\geq I^{11} \\ \text{Var } \hat{\sigma}^2 &\geq I^{22} \\ \text{Var } \hat{\sigma}_w^2 &\geq I^{33} \end{aligned}$$

for any unbiased estimates $\hat{\gamma}$, $\hat{\sigma}^2$, $\hat{\sigma}_w^2$, and where I^{kk} is the k th element on the diagonal of \mathbf{I}^{-1} . However, local bounds such as these are of limited value in general both because our estimates are biased and because the bounds involve the true parameter

values, which are unknown.

When σ_w^2 is known, the Fisher information matrix simplifies to the upper submatrix

$$\mathbf{I} = \sum_{m \in \mathcal{M}} \frac{N(m)[\beta^{-m}]^2}{2(\sigma_m^2)^2} \begin{bmatrix} [\ln 2^m \sigma^2]^2 & -\ln 2^m \sigma^2 \\ -\ln 2^m \sigma^2 & 1 \end{bmatrix} \quad (4.22)$$

from which we get

$$\begin{aligned} \text{Var } \hat{\gamma} &\geq I^{11} \\ \text{Var } \hat{\sigma}^2 &\geq I^{22}. \end{aligned}$$

As one would expect, both the actual error variances and the Cramér-Rao bounds are smaller for this case. Note that because the bounds are still a function of the parameters in this case, their usefulness remains limited. Nevertheless, except in very low SNR settings, the estimate biases are small in a relative sense and the estimation error variance is reasonably-well approximated by these bounds. Hence, the bounds are at least useful in reflecting the quality of estimation that can be expected in various scenarios.

When $\sigma_w^2 = 0$, we get still further simplification, and we can write

$$\mathbf{I} = \begin{bmatrix} (\ln 2)^2/2 \sum_{m \in \mathcal{M}} m^2 N(m) & -(\ln 2)/(2\sigma^2) \sum_{m \in \mathcal{M}} m N(m) \\ -(\ln 2)/(2\sigma^2) \sum_{m \in \mathcal{M}} m N(m) & 1/(2\sigma^4) \sum_{m \in \mathcal{M}} N(m) \end{bmatrix} \quad (4.23)$$

from which we get

$$\begin{aligned} \text{Var } \hat{\gamma} &\geq 2 / \left[(\ln 2)^2 J \right] \sum_{m \in \mathcal{M}} N(m) \\ \text{Var } (\hat{\sigma}^2 / \sigma^2) &\geq 2 / J \sum_{m \in \mathcal{M}} m^2 N(m) \end{aligned}$$

where

$$J = \left[\sum_{m \in \mathcal{M}} m^2 N(m) \right] \left[\sum_{m \in \mathcal{M}} N(m) \right] - \left[\sum_{m \in \mathcal{M}} m N(m) \right]^2.$$

In this case, the bounds no longer depend on the parameters. Moreover, in practice, these expressions give an excellent approximation to the variances of the ML estimates. Evaluating the Cramér-Rao bounds asymptotically for the usual implementation scenario described by (4.9), we get

$$\text{Var} \hat{\gamma}_{\text{ML}} \sim 2 / [(\ln 2)^2 N] \quad (4.24\text{a})$$

$$\text{Var} (\hat{\sigma}_{\text{ML}}^2 / \sigma^2) \sim 2(\log_2 N)^2 / N \quad (4.24\text{b})$$

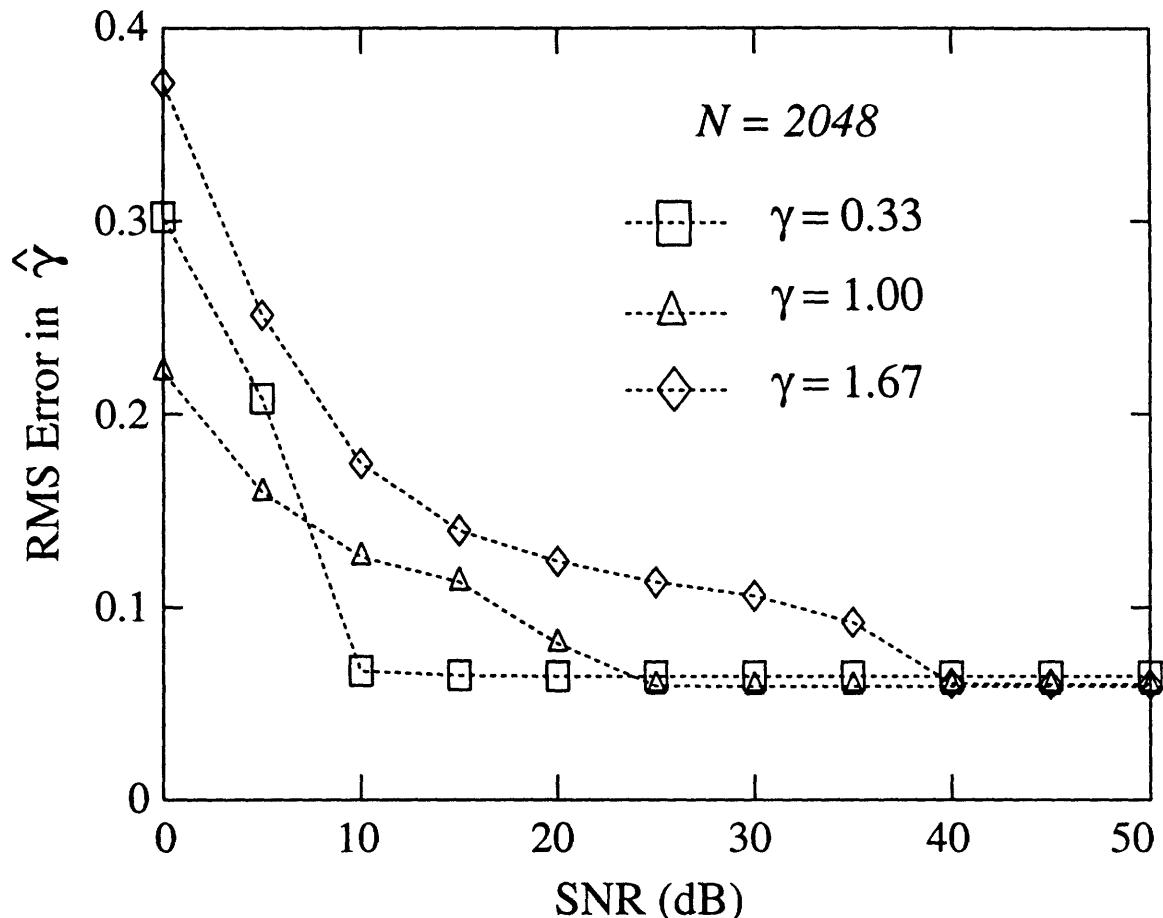
where N is the number of observation samples.

4.2.5 Simulations

For the Monte Carlo simulations of this section, we synthesize discrete samples of resolution-limited Gaussian $1/f$ processes embedded in stationary white Gaussian noise. In general, we vary the length N and SNR of the observations sequence as well as the spectral exponent γ of the underlying $1/f$ processes. We then perform parameter estimation using algorithms for the most general scenario, corresponding to the case in which all signal and noise parameters $\beta, \sigma^2, \sigma_w^2$ are unknown.

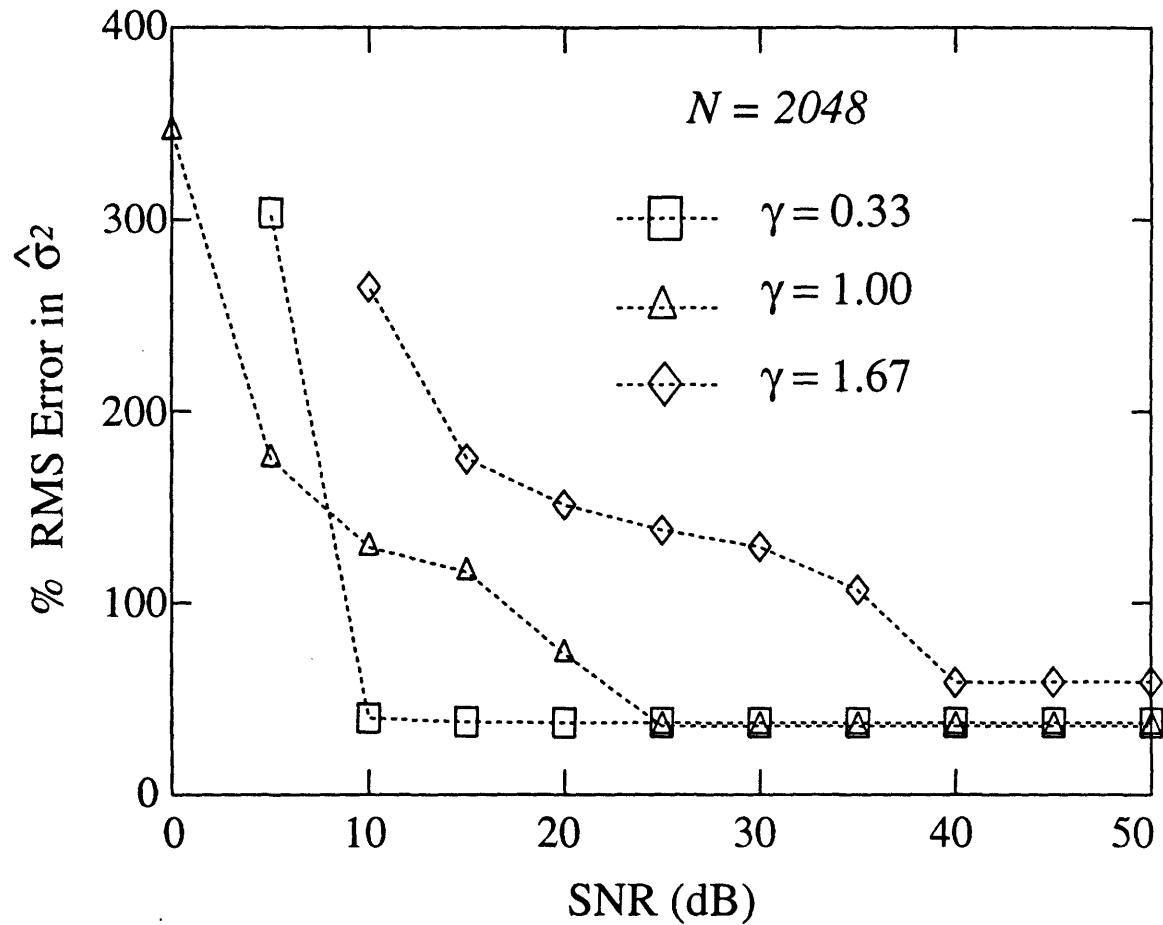
In Fig. 4-2, we plot the RMS error of the estimates of γ and σ^2 for various values of γ as a function of SNR where the observation sequence length is fixed to $N = 2048$. The results from 64 trials were averaged to obtain the error estimates shown. As the results suggest, the quality of the estimates of both parameters is bounded as a consequence of the finite length of the observations. Moreover, the bounds are virtually independent of the value of γ and are achieved asymptotically. For increasing values of γ , the bounds would appear to be attained at increasing SNR thresholds.

In Fig. 4-3, we plot the RMS error of the estimates of γ and σ^2 for various values of γ as a function of observation sequence length N where the SNR is fixed to 20 dB. Again, results from 64 trials were averaged to obtain the error estimates shown. While the results show that the estimation error decreases with data length as expected, they also suggest, particularly for the case of σ^2 , that the convergence toward the true parameters can be rather slow. Note, too, that a rather large amount of data is



(a) Absolute RMS error in $\hat{\gamma}_{\text{ML}}$.

Figure 4-2: RMS Errors in the estimates of the signal parameters as a function of the SNR of the observations. The symbols associated with each γ mark the actual empirical measurements; dashed lines are provided as visual aides only.



(b) Percentage RMS error in $\hat{\sigma}_{ML}^2$.

Figure 4-2: *Continued.*

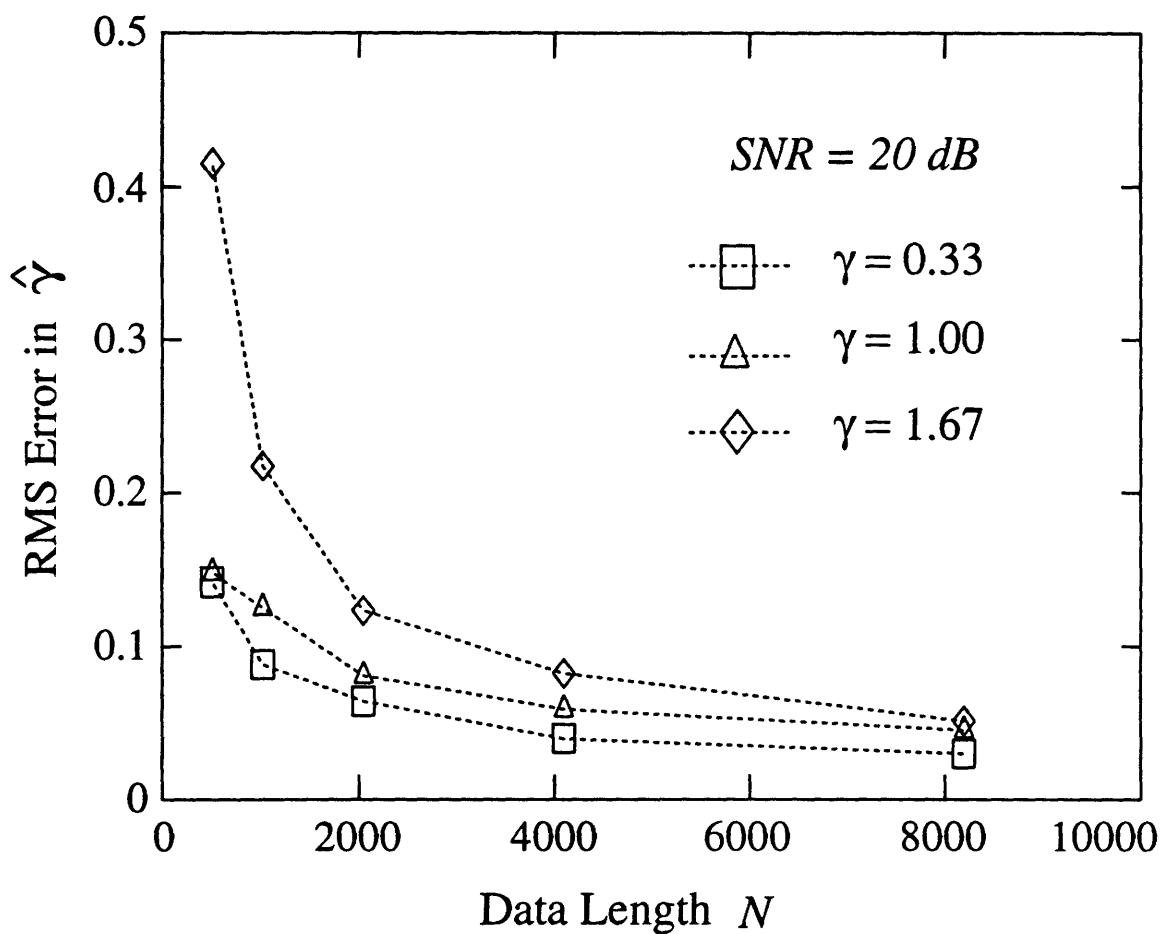
required before the relative estimation error in σ^2 can be made reasonably small.

We conclude this section with a demonstration of the tracking capabilities of the parameter estimation algorithm. Specifically, Fig 4-4 illustrates the performance of the parameter estimation in tracking a step-change in the spectral exponent γ of a noise-free $1/f$ signal. The signal was constructed such that the left and right halves of the signal correspond to $\gamma = 0.90$ and $\gamma = 1.10$, respectively, but identical variances. Local estimates of γ are computed by applying the Case III parameter estimation algorithm to the signal under a sliding window of length 16 384 centered about the point of interest. Note that the algorithm not only accurately resolves the appropriate spectral exponents, but accurately locates the point of transition as well. Finally, we remark that, as in any such tracking algorithm, using a wider estimation window would reduce the variance in the parameter estimates within each half of the waveform, but at the expense of an increase in the width of the transition zone.

4.3 Smoothing of $1/f$ Signals

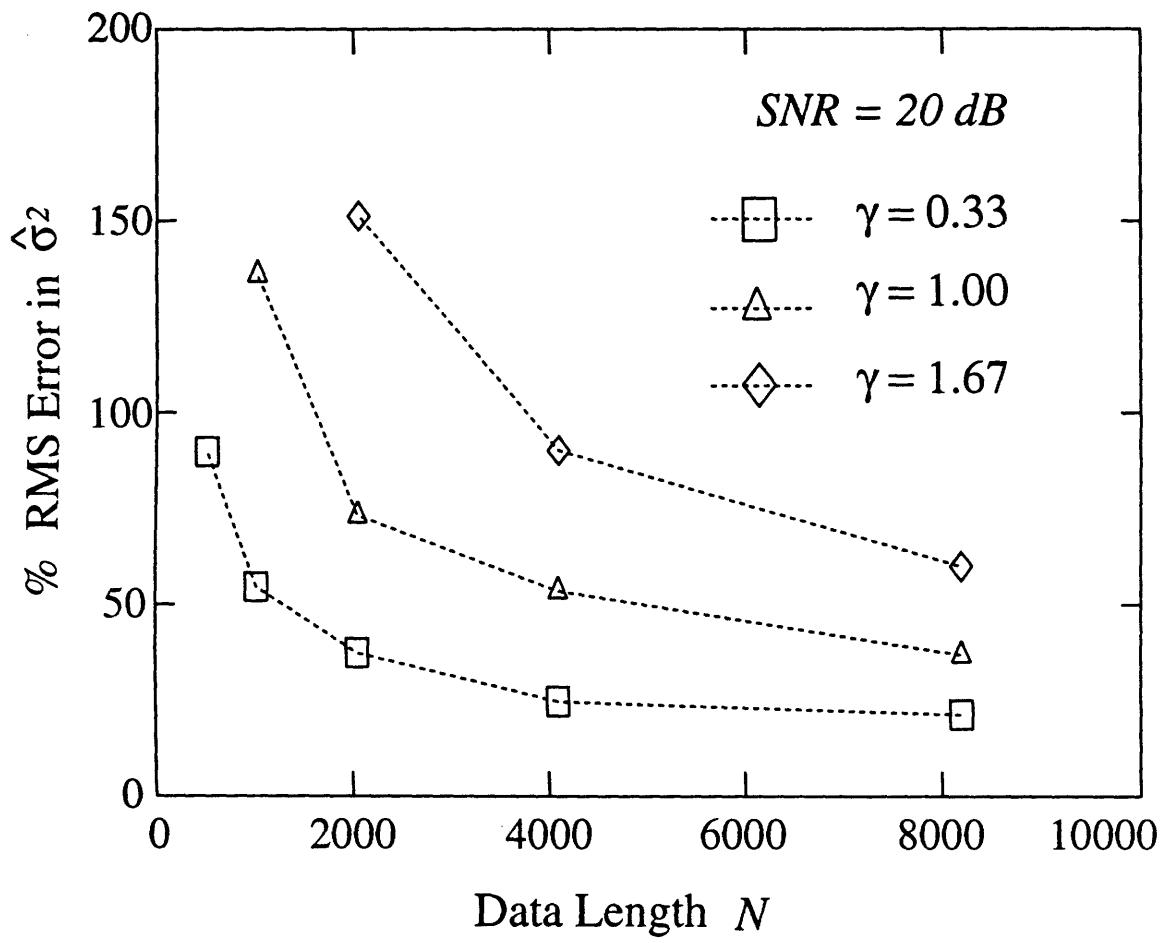
In this section, we consider the problem of extracting a $1/f$ signal from a background of additive stationary white noise. There are many potential problems involving signal enhancement and restoration to which the resulting smoothing algorithms can be applied. For this signal estimation problem, we use a Bayesian framework to derive algorithms that are optimal with respect to a mean-square error criterion. We specifically consider the Gaussian case, for which the resulting algorithms not only yield estimates having the minimum possible mean-square error, but correspond to *linear* data processors as well. However, more generally, for non-Gaussian scenarios the estimators we derive are optimal in a *linear least-squares* sense, *i.e.*, no other linear data processor will be capable of yielding signal estimates with a smaller mean-square error [64].

While we do not specifically derive our signal estimation in terms of Wiener filtering in the frequency domain, interpretations in this domain provide useful insight. In particular, it is clear that at high frequencies the white noise spectrum will dom-



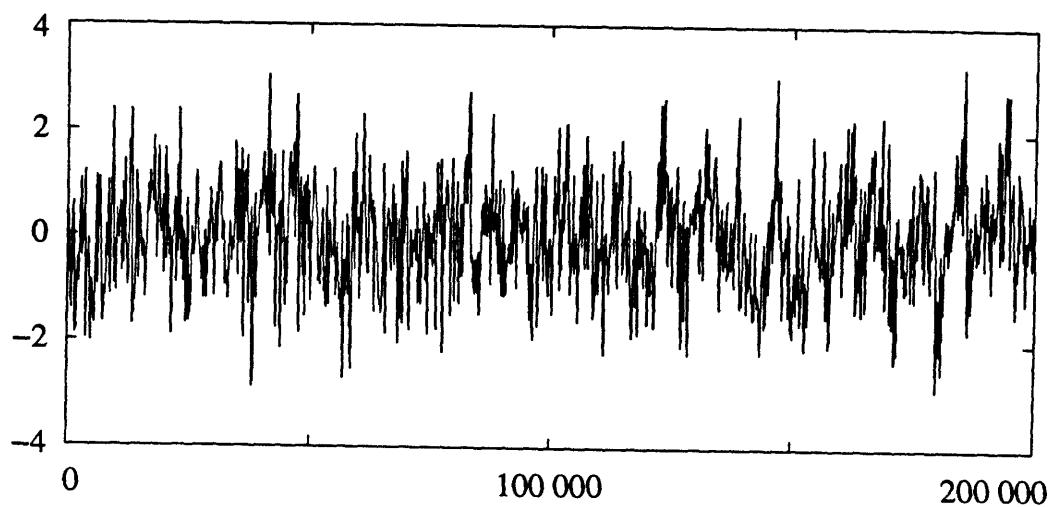
(a) Absolute RMS error in $\hat{\gamma}_{ML}$.

Figure 4-3: RMS Errors in the estimates of the signal parameters as a function of the data length N of the observations. Again, the symbols associated with each γ mark the actual empirical measurements; dashed lines are provided as visual aides only.

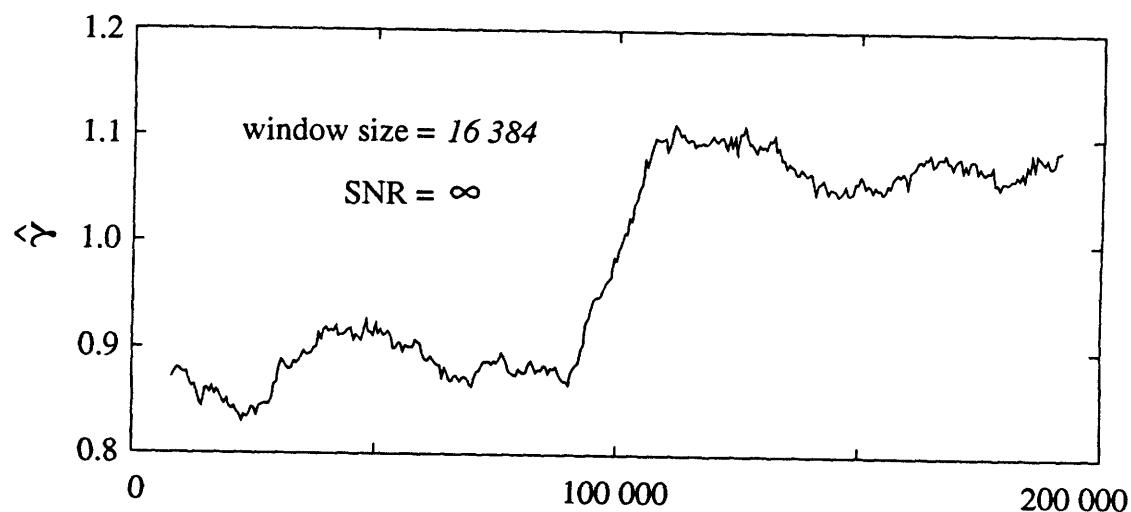


(b) Percentage RMS error in $\hat{\sigma}_{\text{ML}}^2$.

Figure 4-3: *Continued.*



(a) $1/f$ signal with step change in γ .



(b) Estimate of γ under a time-limited window.

Figure 4-4: Tracking the time-varying spectral exponent γ of a noise-free $1/f$ -type signal. For the left half of the signal, $\gamma = 0.90$, while for right half, $\gamma = 1.10$.

inate, while at low frequencies the $1/f$ signal spectrum will dominate. In fact, at sufficiently low frequencies, there will always be arbitrarily high SNR regardless of the noise threshold. Consequently, Wiener filtering for this problem involves a form of low-pass filtering, where the exact filter shape and “cut-off” are governed by the particular parameters of the noise and signal spectra.

Our basic formulation is to consider the estimation of a $1/f$ signal $x(t)$ from noisy observations $r(t)$ of the form (4.6), *viz.*,

$$r(t) = x(t) + w(t)$$

where $w(t)$ is stationary white noise, and where we still consider zero-mean processes. We shall assume in our derivation that the signal and noise parameters $\beta, \sigma^2, \sigma_w^2$ are all known, though, in practice they are typically estimated using the parameter algorithms of the last section. In fact, the parameter and signal estimation problems are quite closely coupled. Indeed it will become apparent in our subsequent development that smoothing was inherently involved in the parameter estimation process as well.

We, again, exploit the wavelet decomposition to obtain our results. Specifically, we begin with the set of wavelet coefficients (4.8). Then, since

$$r_n^m = x_n^m + w_n^m$$

where the x_n^m and w_n^m are all mutually independent with variances $\sigma^2\beta^{-m}$ and σ_w^2 respectively, it follows immediately using classical estimation theory that the estimate of x_n^m that minimizes the mean-square estimation error is given by

$$\hat{x}_n^m = E [x_n^m | r] = \begin{cases} E [x_n^m | r_n^m] & m, n \in \mathcal{R} \\ 0 & \text{otherwise} \end{cases}.$$

Furthermore, when x_n^m and r_n^m are jointly Gaussian, it is straightforward to establish

that the least-squares estimates are linear and given by

$$E[x_n^m | r_n^m] = \left[\frac{\sigma^2 \beta^{-m}}{\sigma^2 \beta^{-m} + \sigma_w^2} \right] r_n^m. \quad (4.25)$$

From these estimates, we can express our optimal estimate of the $1/f$ signal as

$$\hat{x}(t) = \sum_{m,n} \hat{x}_n^m \psi_n^m(t) = \sum_{m,n \in \mathcal{R}} \left[\frac{\sigma^2 \beta^{-m}}{\sigma^2 \beta^{-m} + \sigma_w^2} \right] r_n^m \psi_n^m(t). \quad (4.26)$$

Note that, consistent with our earlier discussion of Wiener filtering for this problem, the smoothing factor

$$\frac{\sigma^2 \beta^{-m}}{\sigma^2 \beta^{-m} + \sigma_w^2}$$

in (4.26) has a thresholding role: at coarser scales where the signal predominates the coefficients are retained, while at finer scales where noise predominates, the coefficients are discarded. Note, too, that this factor appears in (4.14c), which allows us to interpret (4.13b) in terms of sample-variance estimates of the *smoothed* data. Evidently, smoothing is inherently involved in the parameter estimation problem.

Interpreting the optimal estimator (4.26) in terms of the whitening filters of Section 4.1 leads to a conceptually convenient and familiar realization. In particular, as depicted in Fig. 4-5, the optimal linear processor consists of two stages. In the first stage, the noisy observations $r(t)$ are processed by a whitening filter with kernel $\kappa_w(t, \tau)$ given by (4.4b) to generate an intermediate white *innovations* process $v(t)$ whose wavelet coefficients are

$$v_n^m = \frac{r_n^m}{\sigma_m}.$$

In the second stage, $v(t)$ is processed by an *innovations filter* with kernel

$$\kappa_i(t, \tau) = \sum_m \sum_n \psi_n^m(t) \left[\frac{\sigma^2 \beta^{-m}}{\sigma_m} \right] \psi_n^m(\tau) \quad (4.27)$$

to generate the optimal estimate $\hat{x}(t)$ with wavelet coefficients given by (4.25). This innovations-based implementation is a classical estimation structure [65].

In practice, good performance is achieved by these estimators even in very poor

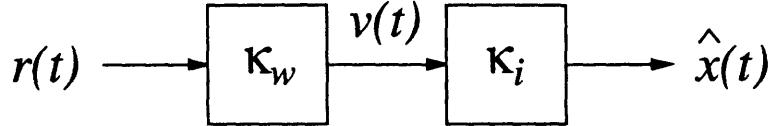


Figure 4-5: A canonic form implementation of the optimal linear filter for estimating a $1/f$ signal $x(t)$ from noisy observations $r(t)$. The linear least-squares filter is the cascade of a whitening filter followed by an innovations filter. The intermediate innovations process $v(t)$ is stationary and white.

SNR scenarios. This is not surprising given the preponderance of energy at low frequencies (coarse scales) in $1/f$ -type processes. Let us, then, turn to a quantitative analysis of the estimation error. First, we note that because our set of observations is finite the total mean-square estimation error

$$\int_{-\infty}^{\infty} E \left[(\hat{x}(t) - x(t))^2 \right] dt$$

is infinite. Nevertheless, when we define

$$\tilde{x}(t) = \sum_{m,n \in \mathcal{R}} x_n^m \psi_n^m(t)$$

as the best possible approximation to $x(t)$ from the finite data set, we can express the total mean-square error in our estimate with respect to $\tilde{x}(t)$ as

$$\begin{aligned} \varepsilon &= \int_{-\infty}^{\infty} E \left[(\hat{x}(t) - \tilde{x}(t))^2 \right] dt \\ &= \sum_{m,n \in \mathcal{R}} E \left[(\hat{x}_n^m - x_n^m)^2 \right] \\ &= \sum_{m,n \in \mathcal{R}} E \left[\text{Var} (x_n^m | r_n^m) \right] \end{aligned}$$

which, through routine manipulation, reduces to

$$\varepsilon = \sum_{m \in \mathcal{M}} N(m) \left[\frac{\sigma^2 \beta^{-m} \cdot \sigma_w^2}{\sigma^2 \beta^{-m} + \sigma_w^2} \right]. \quad (4.28)$$

4.3.1 Simulations

For the simulations of this section, we synthesize discrete samples of resolution-limited Gaussian $1/f$ processes embedded in Gaussian white noise. In general, we vary the SNR of the observations sequence as well as the spectral exponent γ of the underlying $1/f$ processes. We then perform parameter estimation, followed by signal estimation, using algorithms for the most general scenario, corresponding to the case in which all signal and noise parameters $\beta, \sigma^2, \sigma_w^2$ are unknown. Note that by using the estimated parameters in the signal estimation algorithm, our experiments do not allow us to distinguish between those components of signal estimation error due to errors in the estimated parameter values and those due to the smoothing process itself. However, it turns out that the quality of the signal estimation is generally rather insensitive to errors in the parameter estimates used.

In Fig. 4-6, we plot the SNR gain of our smoothed signal estimates for various values of γ as a function of the SNR of the observations where the sequence length is fixed to $N = 2048$. Again, results from 64 trials were averaged to obtain the error estimates shown. The SNR gains predicted by the total mean-square error formula (4.28) are also superimposed on each plot. As the results indicate, the actual SNR gain is typically no more than 1 dB below the predicted gain, as would be expected. However, under some circumstances the deviation can be more than 3 dB. Worse, the SNR gain can be negative, *i.e.*, the net effect of smoothing can be to increase the overall distortion in the signal. Such degradations in performance are due primarily to limitations on the accuracy to which the wavelet coefficients at coarser scales can be extracted via the DWT. In particular, they arise as a result of undesired effects introduced by modeling the data outside the observation interval as periodic to accommodate the inherent data-windowing problem. By contrast, error in the parameter estimates is a much less significant factor in these degradations at reasonably high SNR. The plots also indicate that better gains are achieved for larger values of γ for a given SNR. This is to be expected since for larger values of γ there is more signal energy at coarser scales and correspondingly less at finer scales where the noise predominates and the most attenuation takes place.

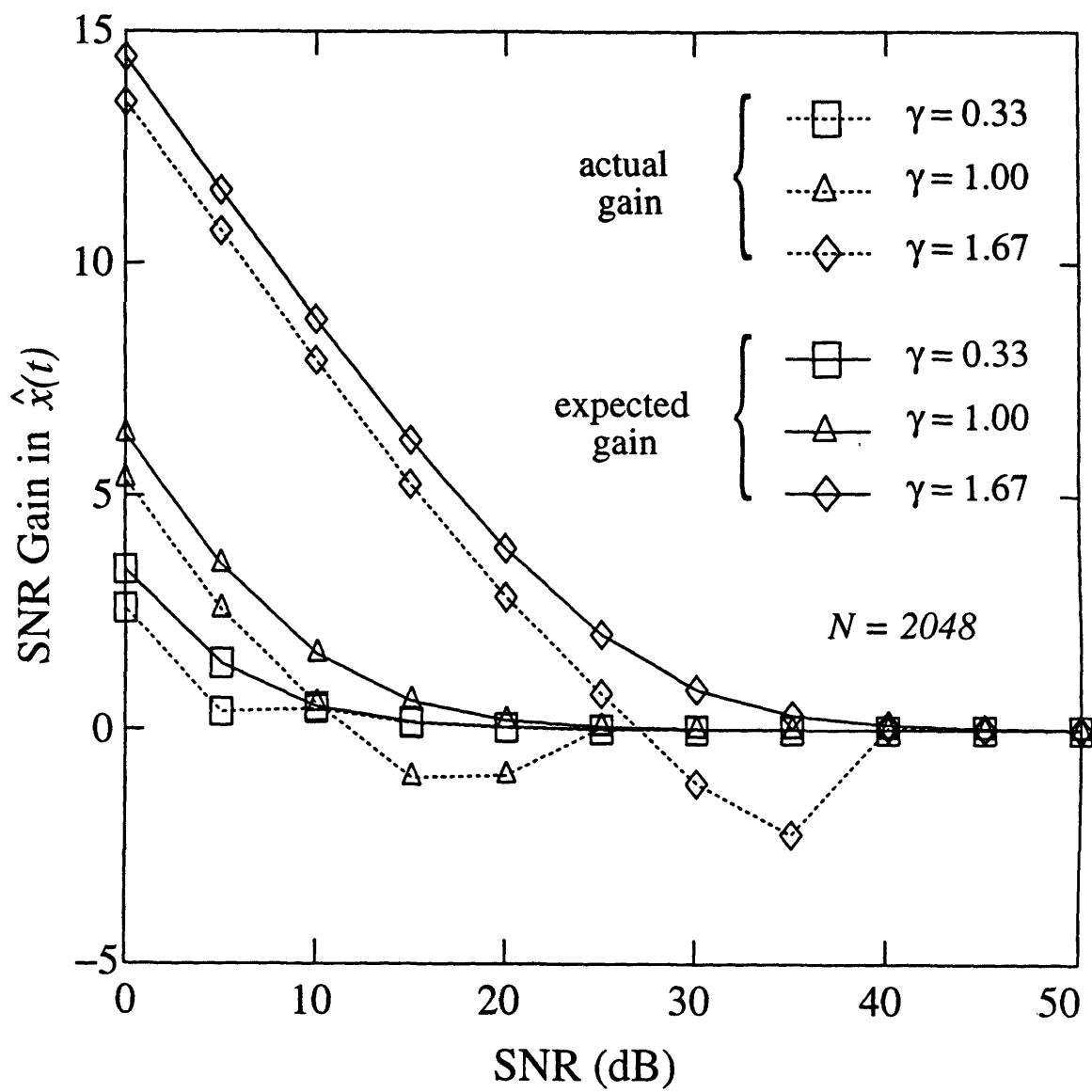


Figure 4-6: SNR gain (dB) of the signal estimate as a function of the SNR of the observations. Both the gains predicted by eq. (4.28) and gains actually obtained are indicated.

We conclude this section with a signal estimation example. In Fig. 4-7 we show a segment of a 65 536-sample $1/f$ signal, the same signal embedded in noise, and the signal estimate. In this example, the spectral exponent is $\gamma = 1.67$, and the SNR in the observations of 0 dB. The estimated spectral exponent is $\hat{\gamma}_{\text{ML}} = 1.66$, and the SNR gain of the signal estimate is 13.9 dB. As anticipated, the signal estimate effectively preserves detail at the coarse scales where the SNR was high, while detail at fine scales is lost where the SNR was low.

4.4 Coherent Detection in $1/f$ Noise

In this section we consider the problem of detecting a known signal of finite energy in a background of Gaussian $1/f$ and white noise. In general, the detection algorithms we shall derive are applicable to many potential problems involving synchronous communication and pattern recognition.

This rather fundamental problem has received some prior attention in the literature. Indeed, Barton and Poor considered the detection of known signals in fractional Gaussian noise backgrounds in [42]. Nevertheless, our formulation has a few distinguishing features, an example of which is that we consider $1/f$ processes corresponding to a much broader range of spectral exponents γ . However, perhaps the principal distinction, apart from one of approach, is the inclusion of stationary white Gaussian measurement noise in our model. In general, this refinement improves the robustness properties of the resulting algorithms and precludes certain singular detection scenarios. Furthermore, it shall become apparent that the wavelet-based approach we follow is not only analytically and conceptually convenient, but leads to practical implementation structures as well.

Let us pose our detection problem in terms of a binary hypothesis test with a Neyman-Pearson optimality criterion. Given noisy observations $r(t)$, we wish to determine a rule for deciding whether or not a known signal is present in the observations. For our test formulation, under hypothesis H_1 we observe a signal of energy $E_0 > 0$ against a background of Gaussian $1/f$ and white noise, while under hypothesis

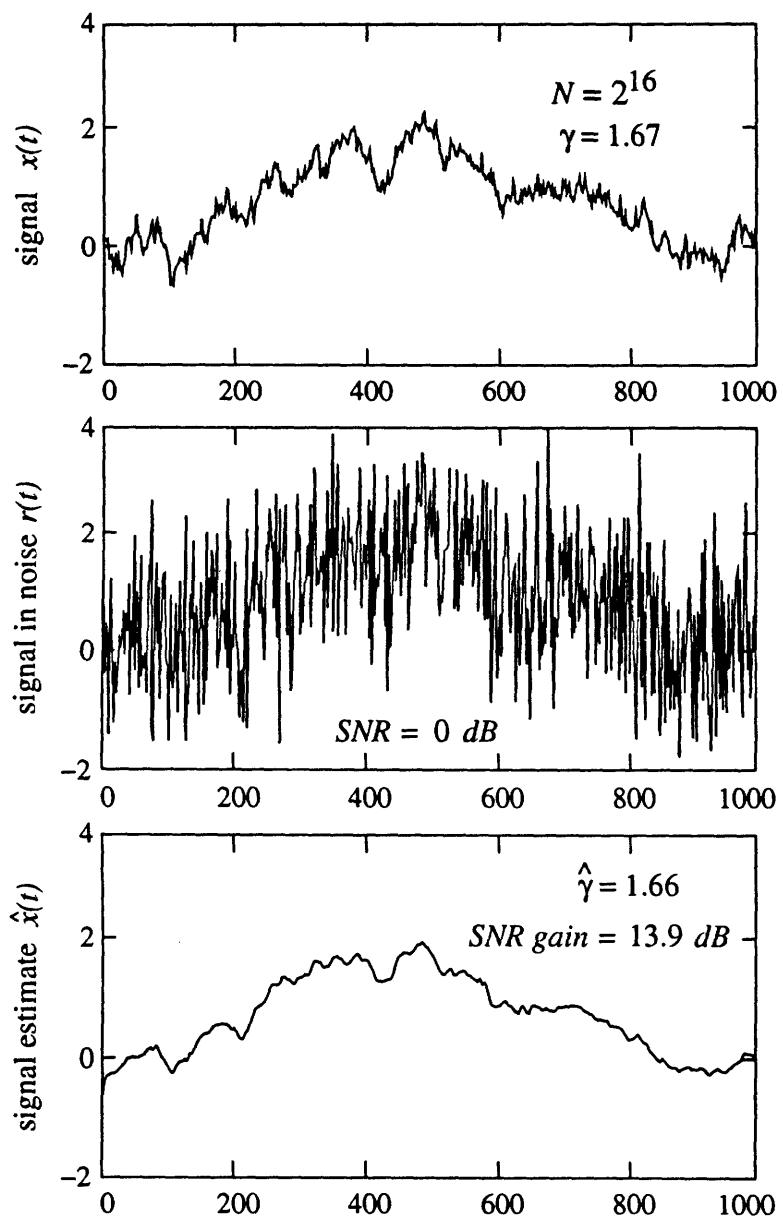


Figure 4-7: Optimal smoothing of a noisy $1/f$ signal.

H_0 we observe only the background noise, *i.e.*,

$$\begin{aligned} H_1 : r(t) &= \sqrt{E_0}s(t) + x(t) + w(t) \\ H_0 : r(t) &= x(t) + w(t) \end{aligned}$$

where $w(t)$ is stationary white Gaussian noise and $x(t)$ is Gaussian $1/f$ -type noise, and $s(t)$ is a unit energy signal:

$$\int_{-\infty}^{\infty} s^2(t) dt = 1.$$

We shall assume $w(t)$ and $x(t)$ to be statistically independent processes under either hypothesis. Let us further assume our observations generally extend over the infinite interval $-\infty < t < \infty$. The problem is then to design a decision rule that maximizes the probability of detecting $s(t)$

$$P_D = \Pr(\text{decide } H_1 \mid H_1 \text{ true})$$

subject to a constraint on the maximum allowable false alarm probability

$$P_F = \Pr(\text{decide } H_1 \mid H_0 \text{ true}).$$

As is well-known, the solution to this problem takes the form of a likelihood ratio test [64].

An equivalent hypothesis test can be constructed in terms of observations of the respective wavelet coefficients

$$\mathbf{r} = \{r_n^m\}$$

as

$$\begin{aligned} H_1 : r_n^m &= \sqrt{E_0}s_n^m + x_n^m + w_n^m \\ H_0 : r_n^m &= x_n^m + w_n^m \end{aligned}$$

for $-\infty < m < \infty$ and $-\infty < n < \infty$. According to our model, under each hypothesis, the coefficients w_n^m and x_n^m are all statistically independent, and have variances σ_w^2 and $\sigma^2\beta^{-m}$, respectively.

In this case, since joint distributions of the observations under the respective hypotheses are

$$\begin{aligned} p_{\mathbf{r}|H_1}(\mathbf{r}) &= \prod_{m,n} \frac{1}{\sqrt{2\pi\sigma_m^2}} \exp \left[-\frac{(r_n^m - \sqrt{E_0}s_n^m)^2}{2\sigma_m^2} \right] \\ p_{\mathbf{r}|H_0}(\mathbf{r}) &= \prod_{m,n} \frac{1}{\sqrt{2\pi\sigma_m^2}} \exp \left[-\frac{(r_n^m)^2}{2\sigma_m^2} \right], \end{aligned}$$

the likelihood ratio

$$\frac{p_{\mathbf{r}|H_1}(\mathbf{r})}{p_{\mathbf{r}|H_0}(\mathbf{r})}$$

can be simplified substantially to yield a test of the form

$$\ell = \sum_m \sum_n \frac{r_n^m s_n^m}{\sigma_m^2} \begin{matrix} H_1 \\ \gtrless \\ H_0 \end{matrix} \quad (4.29)$$

where α is the threshold of the test.

This optimal detector may be realized using a whitening filter based implementation as shown in Fig. 4-8. The statistic ℓ is constructed by processing both $r(t)$ and $\sqrt{E_0}s(t)$ with a *prewhitening filter* whose kernel is given by (4.4b), and correlating the respective outputs $r_*(t)$ and $s_*(t)$. It is straightforward to verify this implementation: since the prewhitened signals $r_*(t)$ and $s_*(t)$ have wavelet coefficients r_n^m/σ_m and $\sqrt{E_0}s_n^m/\sigma_m$, respectively, it suffices to recognize the expression for ℓ in (4.29) as the inner product between $s_*(t)/\sqrt{E_0}$ and $r_*(t)$, which allows us to rewrite (4.29) as

$$\ell = \int_{-\infty}^{\infty} r_*(t) s_*(t) / \sqrt{E_0} dt \begin{matrix} H_1 \\ \gtrless \\ H_0 \end{matrix}$$

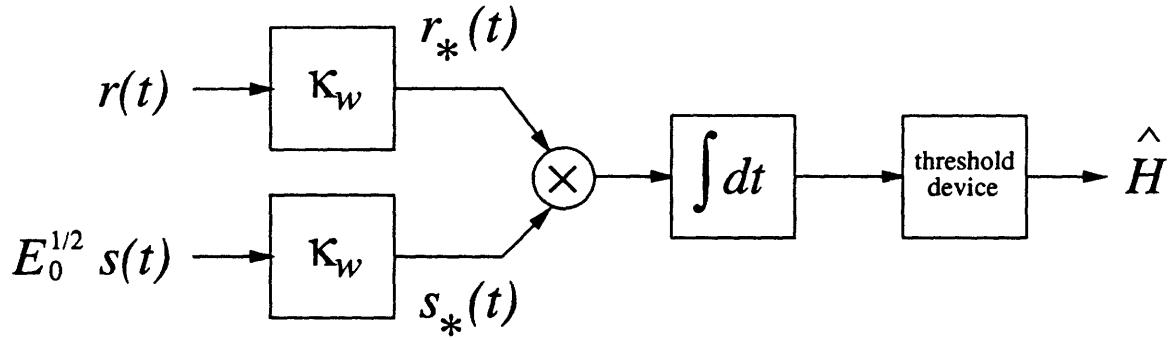


Figure 4-8: Canonical prewhitening implementation of the optimal receiver for detection of a known $s(t)$ in the presence of both Gaussian $1/f$ and stationary white Gaussian noise, where $\kappa_w(t, \tau)$ is the kernel of the whitening filter for $1/f$ -plus-white noise.

This is, of course, a canonical form receiver for optimal detection in the presence of colored noise as described in [64].

Let us turn now to a discussion of performance of this optimal receiver. Via the implementation of this receiver in terms of the whitened observations $r_*(t)$, we note that the performance is necessarily equivalent to that of an optimal detector for $s_*(t)$ in the presence of stationary white Gaussian noise of unit variance. Indeed, if we define the performance index d according to

$$d^2 \triangleq \int_{-\infty}^{\infty} s_*^2(t) dt = E_0 \sum_m \sum_n \frac{(s_n^m)^2}{\sigma_m^2} \quad (4.30)$$

then

$$\begin{aligned} E[\ell|H_0] &= 0 \\ E[\ell|H_1] &= d^2/\sqrt{E_0} \\ \text{Var}\{\ell|H_0\} = \text{Var}\{\ell|H_1\} &= d^2/E_0. \end{aligned}$$

Hence, expressing our arbitrary threshold in the form

$$\alpha = \frac{d}{\sqrt{E_0}} \left[\frac{\ln \eta}{d} + \frac{d}{2} \right]$$

for some $-\infty < \eta < \infty$, the performance of the test can be described in terms of the detection and false alarm probabilities, respectively

$$P_D = Q\left(\frac{\ln \eta}{d} - \frac{d}{2}\right) \quad (4.31a)$$

$$P_F = Q\left(\frac{\ln \eta}{d} + \frac{d}{2}\right) \quad (4.31b)$$

where

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-v^2/2} dv. \quad (4.32)$$

The familiar receiver operating characteristic (ROC) associated with such Gaussian detection problems is as shown in Fig. 4-9 for various values of d . However, while such performance projections are generally useful and suggest the viability of detection in $1/f$ -type backgrounds, there is need for substantiation of these results through a comprehensive set of Monte Carlo simulations. Indeed, it is only through such an evaluation that we can ultimately verify the anticipated insensitivity of the algorithms to our particular $1/f$ model.

In concluding this section, we make some brief remarks on the problem of optimum signal design for use in $1/f$ -plus-white backgrounds. Based on our analysis, it is apparent that we can optimize performance if we choose $s(t)$, or equivalently s_n^m , to maximize d^2 in (4.30) subject to the energy constraint

$$\int_{-\infty}^{\infty} s^2(t) dt = \sum_m \sum_n (s_n^m)^2 = 1.$$

However, this signal optimization problem is not well-posed. Indeed, because of the spectral distribution of the background noise, the optimization will attempt to construct a signal whose energy is at frequencies sufficiently high that the $1/f$ noise is negligible compared to the white component. Consequently, to preclude the generation of an arbitrarily high frequency signal, generally some form of bandwidth constraint is necessary. For an example of how this is accommodated in a communications scenario, see [66].

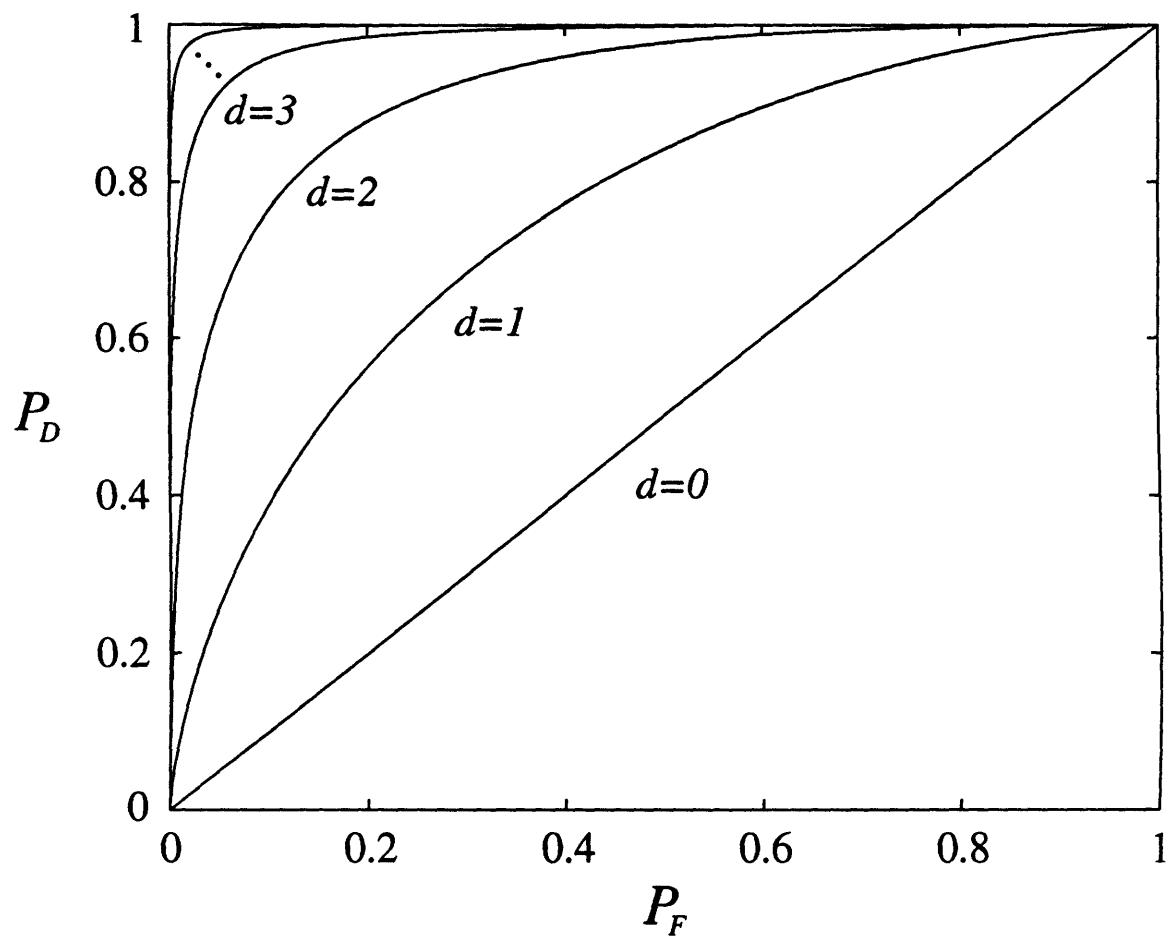


Figure 4-9: The receiver operating characteristic for the detection of a known signal in a background of $1/f$ plus white noise for various thresholds determined via the performance index d .

4.5 Discriminating Between $1/f$ Signals

In this section, we consider the ability of an optimal Bayesian detector to discriminate between Gaussian $1/f$ processes of distinct parameters in a background of stationary white Gaussian noise. The signal classification algorithms we derive are useful in a variety of potential applications. The problem of distinguishing $1/f$ processes is, of course, very closely related to the parameter estimation problem treated in Section 4.2. Indeed, parameter estimation can be viewed as distinguishing among an arbitrarily large number of $1/f$ processes with incrementally different parameters. Nevertheless, as we shall see, approaching the problem from a detection perspective affords a number of new and useful insights.

It is, again, convenient to formulate our problem in terms of a binary hypothesis test in which under each hypothesis we have noisy observations $r(t)$ of distinct $1/f$ signals. Specifically, we have as our two hypotheses

$$H_0 : r(t) = \dot{x}(t) + w(t) \quad (4.33a)$$

$$H_1 : r(t) = \ddot{x}(t) + w(t) \quad (4.33b)$$

where $\dot{x}(t)$ and $\ddot{x}(t)$ are Gaussian $1/f$ processes⁴ with distinct parameters and $w(t)$ is a white measurement noise, statistically independent of $\dot{x}(t)$ or $\ddot{x}(t)$, whose variance is the same under both hypotheses. For this test we will develop a minimum probability of error ($\text{Pr}(\varepsilon)$) decision rule under the assumption of equally likely hypotheses.

Once again, our optimum receiver is best developed and analyzed in the wavelet domain. Rewriting the hypothesis test in terms of the corresponding wavelet coefficients as

$$H_0 : r_n^m = \dot{x}_n^m + w_n^m$$

$$H_1 : r_n^m = \ddot{x}_n^m + w_n^m,$$

⁴We use the notation ‘ and “ to distinguish the $1/f$ processes and their respective parameters under the two hypotheses. These symbols should not be confused with differentiation operators, for which we have generally reserved the notation ‘ and “.

we model the r_n^m under each hypothesis as a collection of zero-mean statistically independent Gaussian random variables with variances

$$\text{Var}\{r_n^m|H_0\} = \dot{\sigma}_m^2 = \dot{\sigma}^2 \dot{\beta}^{-m} + \sigma_w^2 \quad (4.34a)$$

$$\text{Var}\{r_n^m|H_1\} = \ddot{\sigma}_m^2 = \ddot{\sigma}^2 \ddot{\beta}^{-m} + \sigma_w^2 \quad (4.34b)$$

where

$$\dot{\beta} = 2^\gamma$$

$$\ddot{\beta} = 2^{\tilde{\gamma}}.$$

In our derivation we shall assume that, in general, only a *finite* collection of observation coefficients of the form

$$\mathbf{r} = \{r_n^m \in \mathcal{R}\} = \{r_n^m, m \in \mathcal{M}, n \in \mathcal{N}(m)\},$$

where \mathcal{M} and $\mathcal{N}(m)$ are as defined in (4.7), are available. In fact, as we shall see, the problem turns out to be singular (*i.e.*, perfect detection is achievable) if complete observations over the infinite interval are available. In our simulations, we shall assume our observation set \mathcal{R} to be of the particular form (4.9), which corresponds to the collection of coefficients generally available from time- and resolution-limited observations of $r(t)$ via a DWT algorithm as discussed in Section 2.2.4

The likelihood ratio test for this problem can be simplified to a test of the form

$$\ell = \frac{1}{2} \sum_{m \in \mathcal{M}} N(m) \left\{ \left[\frac{1}{\dot{\sigma}_m^2} - \frac{1}{\ddot{\sigma}_m^2} \right] (r_n^m)^2 - \ln \frac{\ddot{\sigma}_m^2}{\dot{\sigma}_m^2} \right\} \begin{matrix} H_1 \\ \geq 0 \\ H_0 \end{matrix} \quad (4.35)$$

where the σ_m^2 are, again, the sample-variances defined via (4.12) which summarize the aspects of interest in the data. It is straightforward to show that this test can be implemented in the canonical form shown in Fig. 4-10. Here the observations $r(t)$

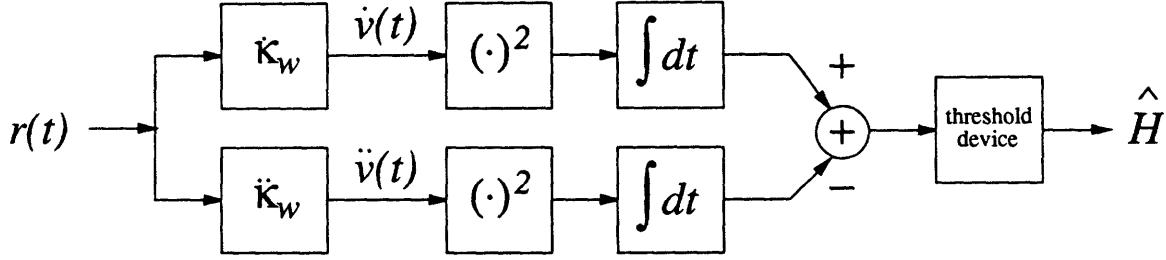


Figure 4-10: A canonical form implementation of the optimal receiver for discriminating between $1/f$ models with distinct parameters based on noisy observations $r(t)$.

are processed by $1/f$ -plus-white whitening filters corresponding to each hypothesis, for which the respective kernels are

$$\begin{aligned}\kappa_w(t, \tau) &= \sum_m \sum_n \psi_n^m(t) \frac{1}{\sigma_m} \psi_n^m(\tau) \\ \ddot{\kappa}_w(t, \tau) &= \sum_m \sum_n \psi_n^m(t) \frac{1}{\ddot{\sigma}_m} \psi_n^m(\tau).\end{aligned}$$

Consequently, only one of the residual processes $\dot{v}(t)$ and $\ddot{v}(t)$ is white, depending on which hypothesis is true. To decide between the two hypotheses, the receiver computes the difference in energy in the two residuals and compares it to the appropriate threshold.

Although the detection problem is Gaussian, it is apparent that the log-likelihood ℓ is not conditionally Gaussian under either hypothesis. Consequently, evaluating the performance of such receivers is rather difficult in general. Nevertheless, it is possible to obtain approximate performance results by exploiting a procedure described in [64] based upon the use of the Chernoff bound. Specifically, defining

$$\mu(s) \triangleq E[e^{s\ell}|H_0],$$

for an arbitrary real parameter s , we can bound the performance of our optimal detector according to

$$\Pr(\varepsilon) \leq \frac{1}{2} e^{\mu(s_*)} \quad (4.36)$$

where s_* is the parameter value yielding the best possible bound, *i.e.*,

$$s_* = \arg \min_s \mu(s).$$

When there are sufficiently many observations to justify modeling ℓ as Gaussian via a central limit theorem (CLT) argument, we can also obtain the following asymptotic expression for the error probability

$$\Pr(\varepsilon) \approx \frac{1}{2s_*(1-s_*)\sqrt{2\pi\mu''(s_*)}} e^{\mu(s_*)}, \quad (4.37)$$

which is a more optimistic and accurate estimate of the achievable performance [64]

In our case, $\mu(s)$ and its first two derivatives are given by

$$\mu(s) = \frac{1}{2} \sum_{m \in \mathcal{M}} N(m) \left\{ s \ln \frac{\dot{\sigma}_m^2}{\ddot{\sigma}_m^2} - \ln \left[s \frac{\dot{\sigma}_m^2}{\ddot{\sigma}_m^2} + (1-s) \right] \right\} \quad (4.38a)$$

$$\mu'(s) = \frac{1}{2} \sum_{m \in \mathcal{M}} N(m) \left\{ \ln \frac{\dot{\sigma}_m^2}{\ddot{\sigma}_m^2} - \left[\frac{\frac{\dot{\sigma}_m^2}{\ddot{\sigma}_m^2} - 1}{s \frac{\dot{\sigma}_m^2}{\ddot{\sigma}_m^2} + (1-s)} \right] \right\} \quad (4.38b)$$

$$\mu''(s) = \frac{1}{2} \sum_{m \in \mathcal{M}} N(m) \left\{ \frac{\frac{\dot{\sigma}_m^2}{\ddot{\sigma}_m^2} - 1}{s \frac{\dot{\sigma}_m^2}{\ddot{\sigma}_m^2} + (1-s)} \right\}^2. \quad (4.38c)$$

It is generally not possible to derive a closed form expression for the minimum value of $\mu(s)$ via either (4.38a) or (4.38b) for this asymmetric detection problem. Fortunately, though, a numerical optimization is reasonable: it suffices to consider a numerical search over values of s within a limited range. Indeed, since

$$\mu(0) = \mu(1) = 0,$$

and since from (4.38c) we have that $\mu(s)$ is convex

$$\mu''(s) \geq 0,$$

it follows that the minimum of $\mu(s)$ can be found in the range $0 \leq s \leq 1$.

4.5.1 Simulations

In this section, we obtain, via (4.36) and (4.37), numerical estimates of the probability of error performance. For our scenario, we assume that coefficients r_n^m are available in the range (4.9) consistent with what could be extracted from $N = 2^M$ samples of time- and resolution-limited observations via a DWT algorithm. In our simulations, we consider the ability of the optimal receiver to distinguish between $1/f$ processes of different spectral exponents γ (or, equivalently, fractal dimensions D , or self-similarity parameters H). In particular, we do not consider the capabilities of the algorithms to discriminate on the basis of variance differences. Consequently, in all our tests, we choose the variance parameters σ^2 and $\tilde{\sigma}^2$ such that the variance of the observations is identical under either hypothesis.

In our first set of simulations, the bound (4.36) is used as an estimate of the probability of error performance of an optimal detector in discriminating between two equal-variance $1/f$ processes whose spectral exponents differ by $\Delta\gamma$ based on noisy observations of length N corresponding to a prescribed SNR. In the tests, three different spectral exponent regimes are considered, corresponding to $\gamma = 0.33$, $\gamma = 1.00$, and $\gamma = 1.67$.

In Fig. 4-11, performance is measured as a function of SNR for noisy observations of length $N = 128$ and a parameter separation $\Delta\gamma = 0.1$. Note that there is a threshold phenomenon: above a certain γ -dependent SNR, $\Pr(\varepsilon)$ drops dramatically. Moreover, the threshold is lower for larger values of γ . This is to be expected since larger values of γ correspond to an effective $1/f$ power spectrum that is increasingly peaked at the $\omega = 0$, so that a correspondingly greater proportion of the total signal power is not masked by the white observation noise. Beyond this threshold performance saturates as the data is essentially noise-free. However, note that there is crossover behavior: at SNR values above the thresholds, better performance is obtained for smaller values of γ . In subsequent tests, we restrict our attention to performance in this high SNR regime.

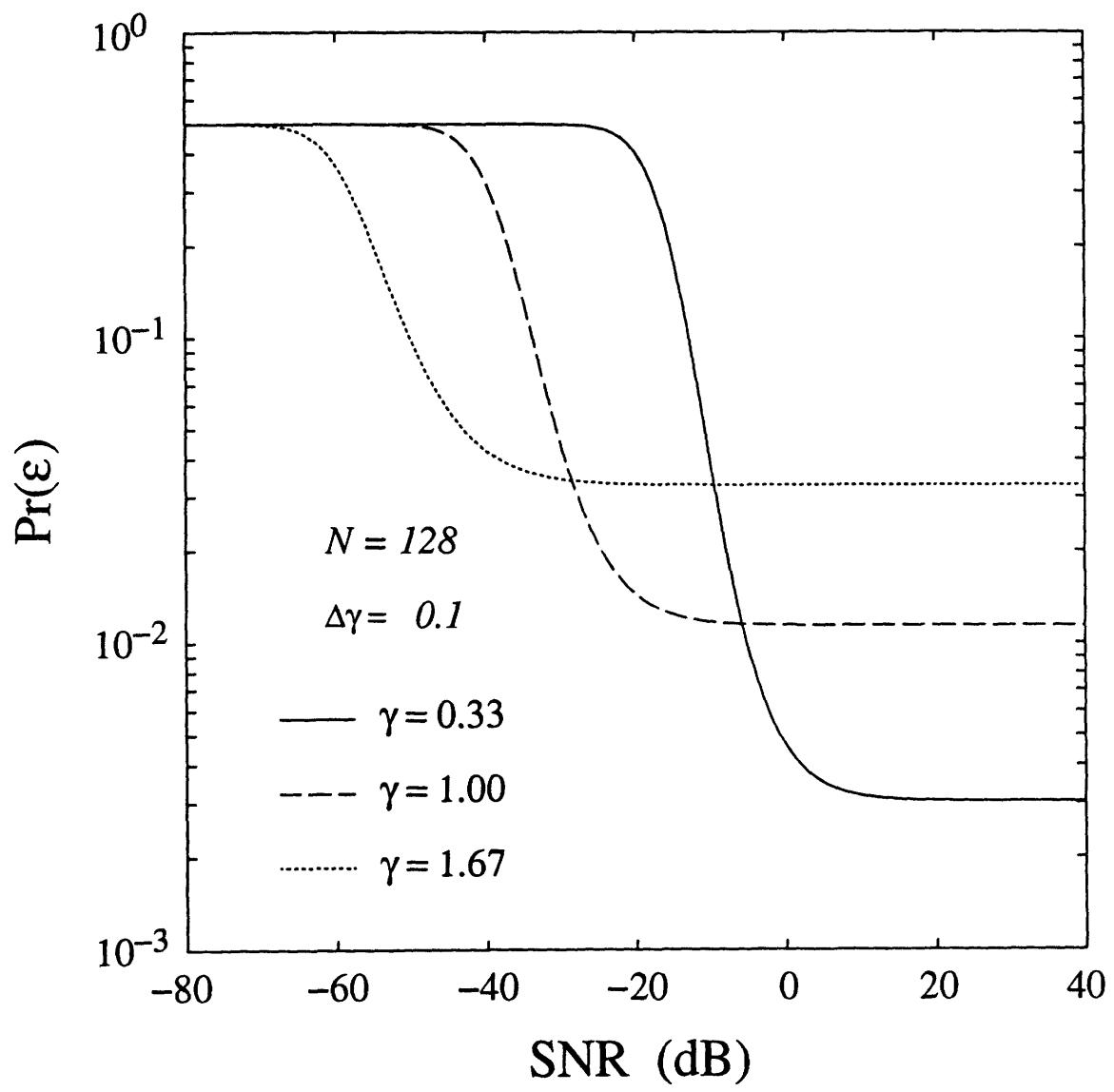


Figure 4-11: Optimal discriminator performance as a function of SNR, as estimated via the Chernoff bound.

In Fig. 4-12, performance is plotted as a function of the number of samples N of observed data corresponding to an SNR of 20 dB and hypotheses whose parameter separation is $\Delta\gamma = 0.1$. In this case, there is thresholding behavior as well. For data lengths beyond a critical order-of-magnitude we get strongly increasing performance as a function of data length. Again, because we are in the high SNR regime, we observe that the best performance is achieved for the smallest values of γ .

Finally, in Fig. 4-13, performance is plotted as a function of the separation between the two hypotheses—specifically, the difference between the spectral parameters for noisy observations of length $N = 128$ corresponding to an SNR of 20 dB. As we would expect, the results illustrate that the larger the distinction between the hypotheses, the better the performance achievable by the receiver. Again, as we are in the high SNR regime, we find that the best performance is achieved for the smallest values of γ .

Whenever the probability of error is low—*i.e.*, either when the SNR is high, large data lengths are involved, or the hypotheses are well-separated—it turns out that the CLT-based approximation (4.37) represents a more optimistic estimate of performance than does (4.36). However, in high $\text{Pr}(\varepsilon)$ scenarios, (4.36) constitutes a more useful measure of system performance than does (4.36). This behavior is illustrated in Figs. 4-14, 4-15, and 4-16 for hypotheses in the $\gamma = 1$ regime. Note that only at sufficiently high SNR, data lengths, and parameter separations does the CLT-based approximation actually yield a $\text{Pr}(\varepsilon)$ estimate that is below the bound. We cannot, of course, assess whether the CLT-based approximation is overly optimistic in the high SNR regime. In general, we can only expect the estimate to be asymptotically accurate as $N \rightarrow \infty$. Nevertheless, the fact that the rate of change of $\text{Pr}(\varepsilon)$ with respect to SNR, data length N , and parameter separation $\Delta\gamma$ has a similar form for both the bound and the approximation gives us additional confidence that the earlier plots of $\text{Pr}(\varepsilon)$ do suggest the correct form of the functional dependence upon these parameters.

It is important to emphasize once again that the simulations of this section constitute a highly preliminary performance study. Indeed, any comprehensive evaluation of

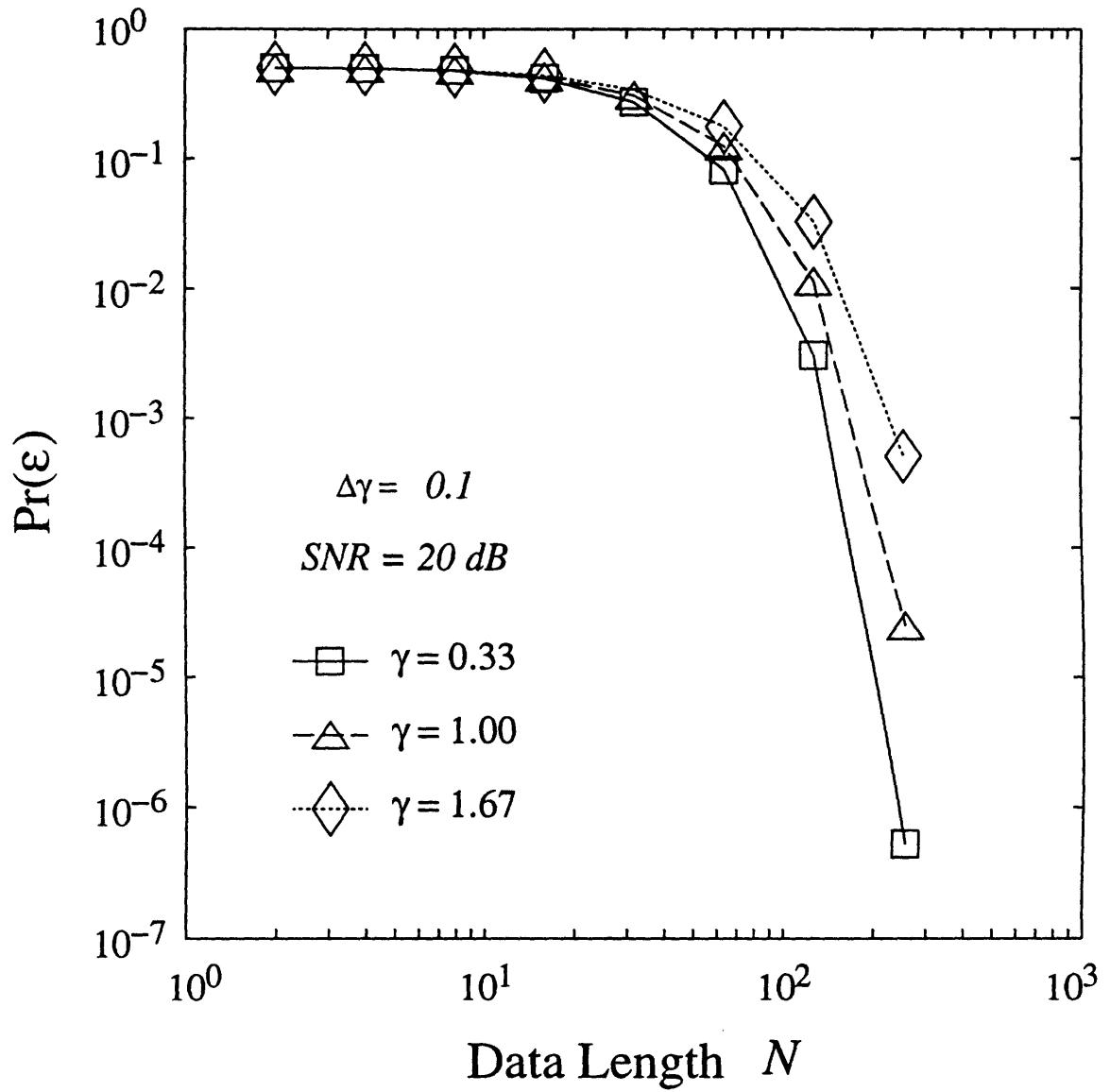


Figure 4-12: Optimal discriminator performance as a function of the number of samples N of noisy observations, as estimated by the Chernoff bound. The symbols \square , \triangle and \diamond correspond to actual estimates; the lines are provided as visual aides only in this case.

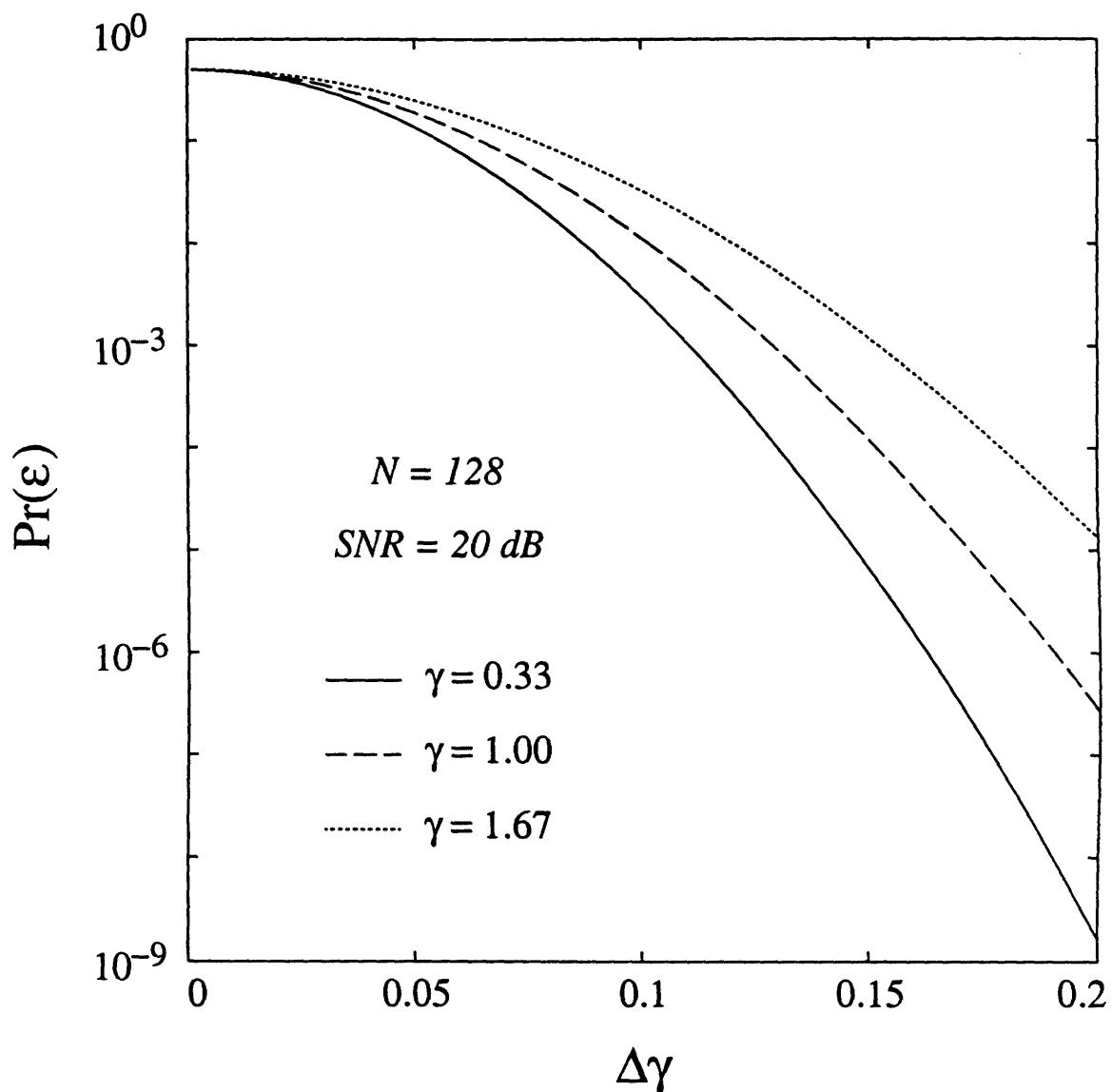


Figure 4-13: Optimal discriminator performance as a function of the parameter separation $\Delta\gamma$ between the two hypotheses, as estimated via the Chernoff bound.

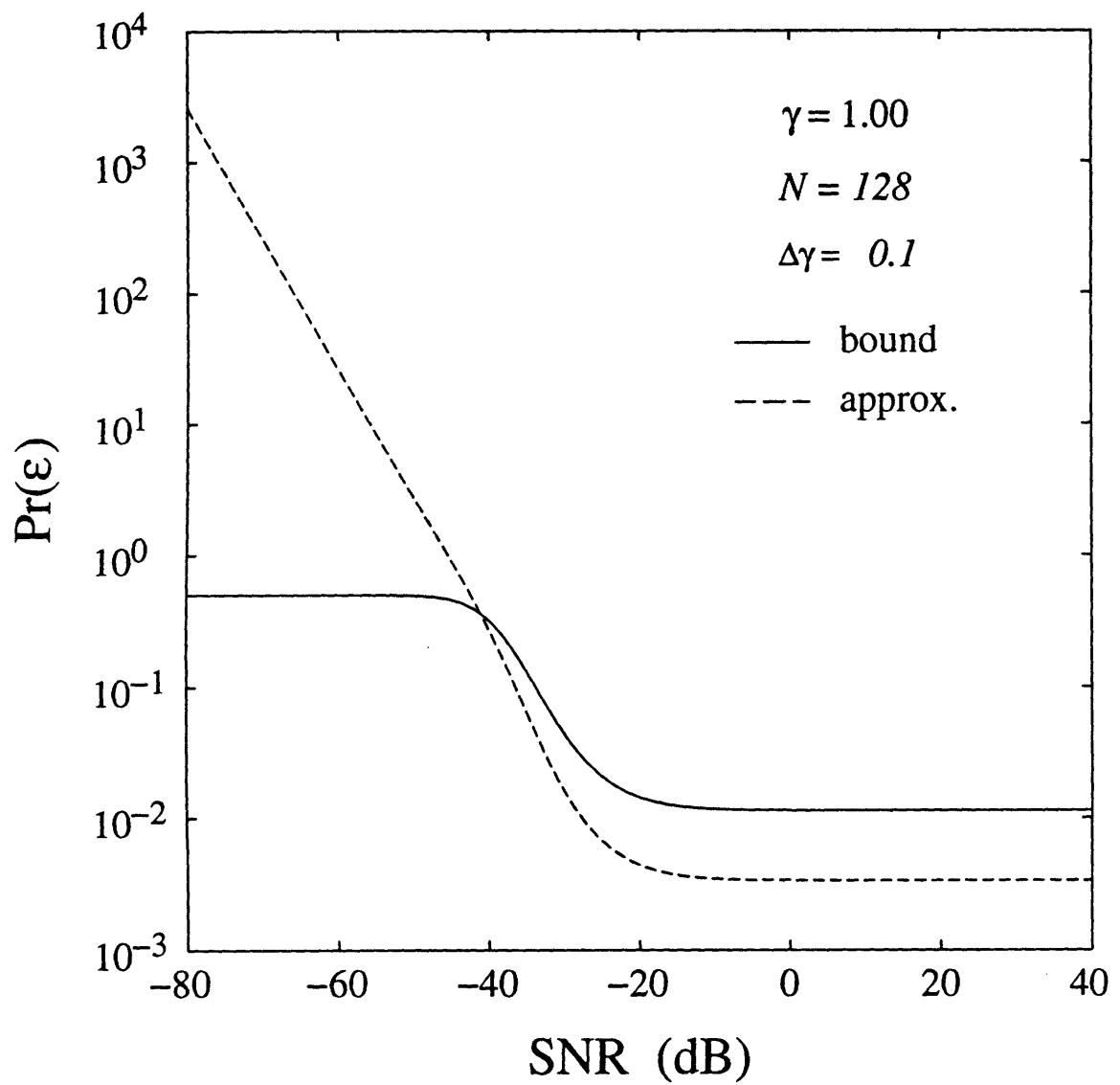


Figure 4-14: Optimal discriminator performance as a function of SNR, as estimated via both the Chernoff-bound (4.36) and the CLT-based approximation (4.37).

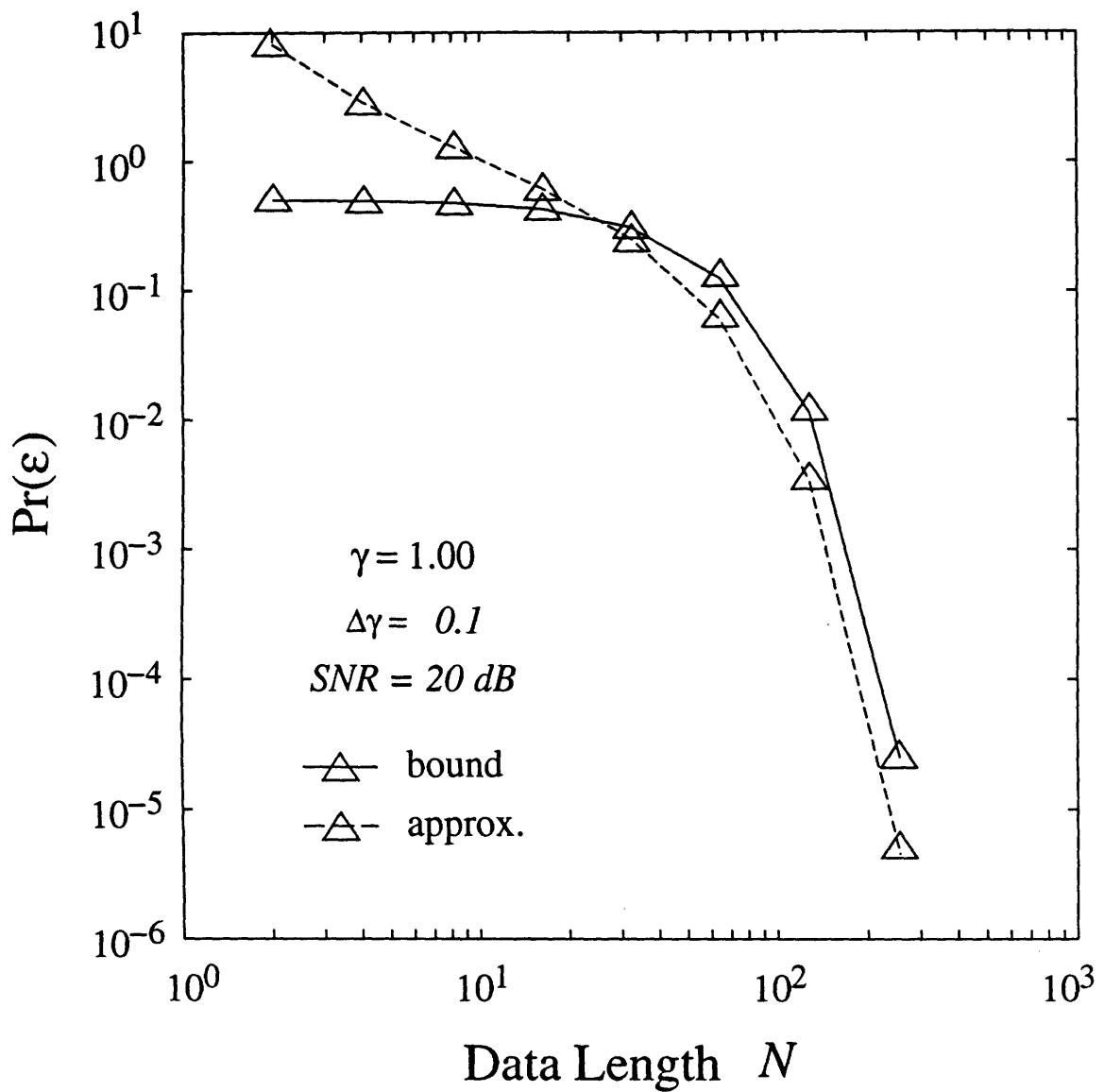


Figure 4-15: Optimal discriminator performance as a function of the number of samples of observed data, as estimated via both the Chernoff-bound (4.36) and the CLT-based approximation (4.37). The \triangle symbols correspond to actual estimates; the lines are provided as visual aides only.

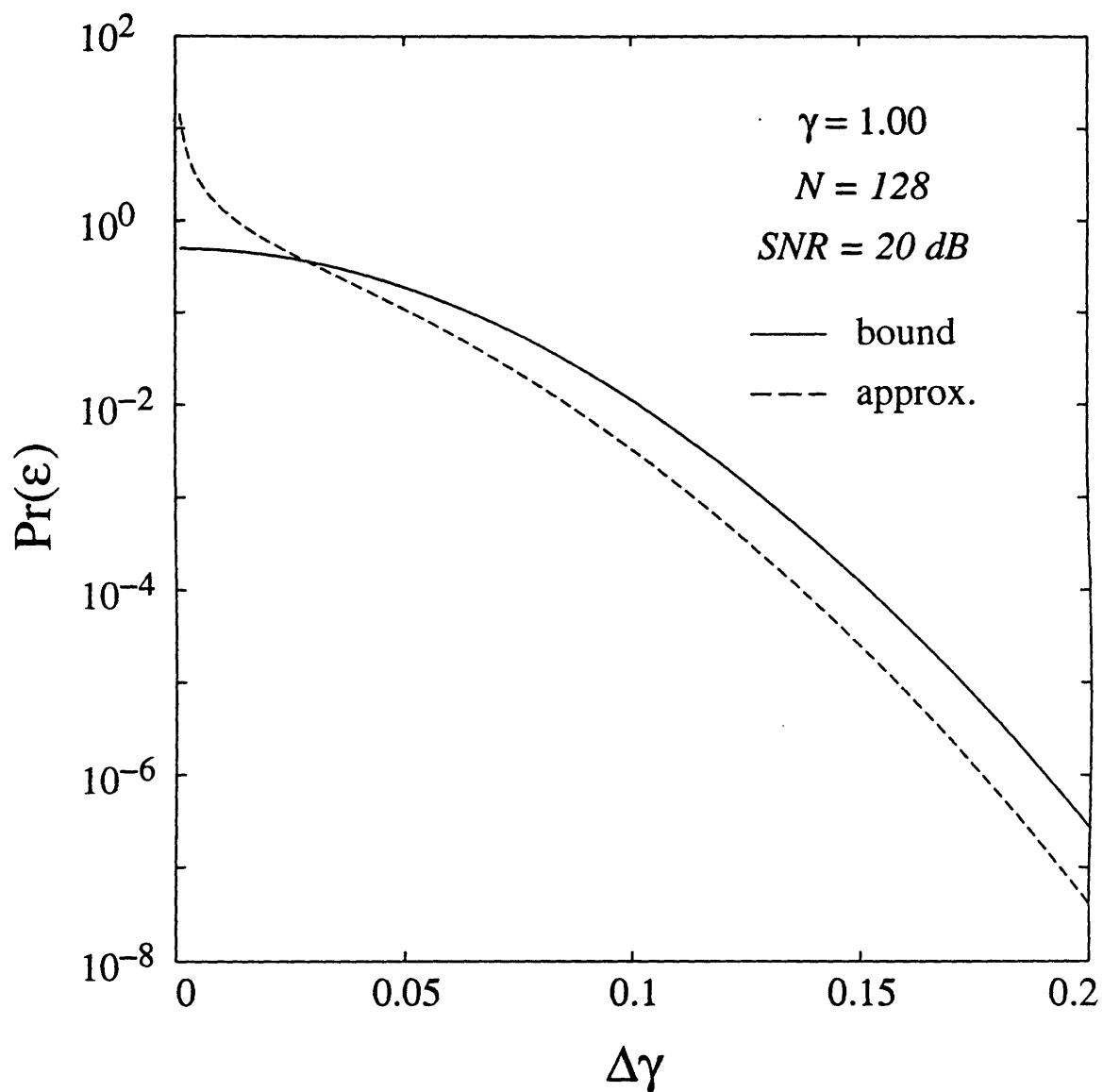


Figure 4-16: Optimal discriminator performance as a function of the parameter separation $\Delta\gamma$, as estimated via both the Chernoff-bound (4.36) and the CLT-based approximation (4.37).

the system must ultimately involve a series of Monte Carlo simulations using synthetic and real data. Such tests are critical to exploring issues pertaining to robustness of the algorithms to modeling errors. Nevertheless, the numerical results presented here lend some valuable insights into the performance potential and anticipated behavior of optimal discriminators for $1/f$ -type processes.

Before concluding this section, we consider a potentially useful and practical refinement of the optimal discrimination problem. There are a number of application contexts in which we would be more interested in distinguishing $1/f$ processes strictly on the basis of their spectral exponents, fractal dimensions, or self-similarity parameters. This would correspond to a hypothesis test (4.33) in which $\dot{\sigma}^2$, $\ddot{\sigma}^2$ and σ_w^2 would be *unwanted* parameters of the problem. In this case, a solution could be obtained using a *generalized* likelihood ratio test [64] of the form

$$\frac{\max_{\dot{\sigma}^2, \sigma_w^2} \prod_{m,n \in \mathcal{R}} \frac{1}{\sqrt{2\pi\ddot{\sigma}_m^2}} \exp\left[-\frac{(r_n^m)^2}{2\ddot{\sigma}_m^2}\right]}{\max_{\dot{\sigma}^2, \sigma_w^2} \prod_{m,n \in \mathcal{R}} \frac{1}{\sqrt{2\pi\dot{\sigma}_m^2}} \exp\left[-\frac{(r_n^m)^2}{2\dot{\sigma}_m^2}\right]} \stackrel{H_1}{>} 1. \quad (4.39)$$

In general, expressions for the maxima involved in the construction of the likelihood function of (4.39) cannot be obtained in closed form. However, a practical implementation of this receiver could potentially exploit an EM algorithm of the general type developed in Section 4.2. In terms of performance, we would anticipate that, in general, it would only be possible to adequately evaluate such a receiver through Monte Carlo simulations.

4.6 Outstanding Issues

A number of outstanding issues remain to be addressed concerning wavelet-based representations of $1/f$ processes and their use in signal processing. As an example, while we rely on both theoretical and empirical evidence that the wavelet coefficients of $1/f$ processes are decorrelated, we recall that the theoretical justification of this analysis result was rather weak. An important contribution would be to establish

both tight bounds on the correlation between wavelet coefficients and conditions on the wavelet basis underwhich such bounds are valid.

A complementary issue concerns our synthesis result, which established that wavelet expansions in terms of uncorrelated random variables give rise to nearly- $1/f$ spectra. An interesting and potentially important question concerns whether there exist wavelets giving rise to an exactly- $1/f$ spectrum. While the answer may well be negative, perhaps sequences of dyadic or non-dyadic wavelets can chosen to approach a $1/f$ spectrum arbitrarily closely.

Another issue pertains to the unusual data-windowing problem inherent in the wavelet decomposition. In the experiments described in this work, we have avoided the problem by modeling the data as periodic outside the finite observation interval during computation of the DWT. However, this leads to a number of rather undesirable effects, some of which manifested themselves in the smoothing simulations as we noted. More effective approaches to accommodating observations on the finite interval need to be developed.

Finally, a number of interesting, straightforward and useful extensions to this work are suggested by the approaches described here. Specifically, the problem of distinguishing and isolating two superimposed fractal signals is, in principle, readily solved by the methods of this work. In addition, the separable extension of the results presented herein to two and higher dimensions is likewise straightforward. In each case, we anticipate that a number of powerful yet practical algorithms can be developed.

Although a general treatment of incoherent detection problems involving $1/f$ backgrounds is beyond the scope of this thesis, a variety of such problems can be addressed using some straightforward extensions to the approaches of the chapter. In general, to model these scenarios we can consider the detection of signals known to within a collection of parameters. Optimal detection schemes for such scenarios typically involve generalized likelihood ratio tests in which there are both parameter estimation and signal detection components. An example of such an extension was outlined at the end of Section 4.5.1.

Chapter 5

Deterministically Self-Similar Signals

Signals $x(t)$ satisfying the deterministic scale-invariance property

$$x(t) = a^{-H}x(at) \quad (5.1)$$

for all $a > 0$, are generally referred to in mathematics as *homogeneous* functions (of degree H). Homogeneous functions can be regular or nearly so, for example $x(t) = 1$ or $x(t) = u(t)$, or they can be generalized functions, such as $x(t) = \delta(t)$. In any case, as shown by Gel'fand [67], homogeneous functions can be parameterized with only a few constants. As such, they constitute a rather limited class of signal models in many contexts.

A comparatively richer class of signal models is obtained by considering waveforms which are required to satisfy (5.1) only for values of a that are integer powers of two. This broader class of homogeneous signals then satisfy the dyadic self-similarity property

$$x(t) = 2^{-kH}x(2^k t) \quad (5.2)$$

for all integers k . It is this, more general, family of homogeneous signals of degree H whose properties and characterizations we study in this chapter. When there is risk of confusion in our subsequent development, we will denote signals satisfying (5.1) as

strict-sense homogeneous, and signals satisfying (5.2) as *wide-sense* homogeneous.

Homogeneous signals constitute an interesting and potentially important class of signals for use in a number of communications-based applications. In Chapter 6 we shall see that when we develop a strategy for embedding information into a waveform “on all time scales,” the resulting waveforms are homogeneous signals. As a consequence of intrinsic self-similarity, these modulated waveforms have the property that an arbitrarily short duration time-segment is sufficient to recover the entire waveform, and hence the information, given adequate bandwidth. Likewise an arbitrarily low-bandwidth approximation to the waveform is sufficient to recover the undistorted waveform, and hence the information, given adequate duration. Furthermore, we will see that these homogeneous waveforms have spectral characteristics very much like those of $1/f$ processes, and, in fact, have fractal properties as well.

Collectively, such properties make this modulation scheme an intriguing paradigm for communication over highly unreliable channels of uncertain duration, bandwidth, and SNR, as well as in a variety of other contexts. We shall explore these and other issues in the next chapter. In the meantime, we turn our attention to developing a convenient mathematical framework for characterizing homogeneous signals that shall prove useful in the sequel.

While all non-trivial homogeneous signals have infinite energy, and many have infinite power, there are nevertheless some such signals with which one can associate a generalized $1/f$ -like Fourier transform, and others with which one can associate a generalized $1/f$ -like power spectrum. We distinguish between these two classes of homogeneous signals in our subsequent treatment, denoting them *energy-dominated* and *power-dominated* homogeneous signals, respectively. We begin our theoretical development by more rigorously defining the notion of an energy-dominated homogeneous signal, and constructing some Hilbert space characterizations. This will lead to some useful constructions for orthonormal “self-similar bases” for homogeneous signals. It will become apparent that, as in the case of statistically self-similar $1/f$ -type processes, orthonormal wavelet basis expansions constitute natural and efficient representations for these signals as well.

5.1 Energy-Dominated Homogeneous Signals

Our definition of an energy-dominated homogeneous signal is reminiscent of the one we proposed for $1/f$ processes in Section 3.1. Specifically,

Definition 5.1 *A wide-sense homogeneous signal $x(t)$ is said to be energy-dominated if when $x(t)$ is filtered by an ideal bandpass filter with frequency response*

$$B_0(\omega) = \begin{cases} 1 & \pi < |\omega| \leq 2\pi \\ 0 & \text{otherwise} \end{cases} \quad (5.3)$$

the resulting signal $\tilde{x}_0(t)$ has finite-energy, i.e.,

$$\int_{-\infty}^{\infty} \tilde{x}_0^2(t) dt < \infty.$$

As with $1/f$ processes, the choice of passband edges at π and 2π in our definition is somewhat arbitrary. In particular, substituting in the definition any passband that does not include $\omega = 0$ or $\omega = \infty$ but includes one entire frequency octave leads to precisely the same class of signals. Nevertheless, our particular choice is both sufficient and convenient.

We remark that the class of energy-dominated homogeneous signals includes both reasonably regular functions, such as the constant $x(t) = 1$, the ramp $x(t) = t$, the time-warped sinusoid $x(t) = \cos[2\pi \log_2 t]$, and the unit step function $x(t) = u(t)$, as well as singular functions, such as $x(t) = \delta(t)$ and its derivatives. However, although we will not always be able to actually “plot” signals of this class, we will be able to suitably characterize such functions in some useful ways. Let us begin by using E^H to denote the collection of all energy-dominated homogeneous signals of degree H . The following theorem allows us to interpret the notion of spectra for such signals. A straightforward but detailed proof is provided in Appendix D.1.

Theorem 5.2 When an energy-dominated homogeneous signal $x(t)$ is filtered by an ideal bandpass filter with frequency response

$$B(\omega) = \begin{cases} 1 & \omega_L < |\omega| \leq \omega_U \\ 0 & \text{otherwise} \end{cases} \quad (5.4)$$

for arbitrary $0 < \omega_L < \omega_U < \infty$, the resulting signal $y(t)$ has finite energy and a Fourier transform of the form

$$Y(\omega) = \begin{cases} X(\omega) & \omega_L < |\omega| \leq \omega_U \\ 0 & \text{otherwise} \end{cases} \quad (5.5)$$

where $X(\omega)$ is some function that is independent of ω_L and ω_U and has octave-spaced ripple, i.e., for all integers k ,

$$|\omega|^{H+1} X(\omega) = |2^k \omega|^{H+1} X(2^k \omega). \quad (5.6)$$

Note that since $X(\omega)$ in this theorem does not depend on ω_L or ω_U , this function may be interpreted as the generalized Fourier transform of $x(t)$. Furthermore, (5.6) implies that the generalized Fourier transform of signals in E^H obeys a $1/f$ -like (power-law) relationship, *viz.*,

$$|X(\omega)| \sim \frac{1}{|\omega|^{H+1}}.$$

However, we shall continue to reserve the term “ $1/f$ process” or “ $1/f$ signal” for the statistically self-similar random processes defined in Chapter 3.

We also remark that $X(\omega)$ does not uniquely specify $x(t) \in E^H$, i.e., the mapping

$$x(t) \longleftrightarrow X(\omega)$$

is not one to one. As an example, $x(t) = 1$ and $x(t) = 2$ are both in E^H for $H = 0$, yet both have $X(\omega) = 0$ for $\omega > 0$. In order to accommodate this pathological

behavior in our subsequent theoretical development we shall exploit the notion of an equivalence class. In particular, we shall consider all signals having a common $X(\omega)$ as equivalent or indistinguishable.

Indirectly, Theorem 5.2 suggests that an energy-dominated homogeneous signal $x(t)$ has a convenient representation in terms of the ideal bandpass wavelet basis. In particular, if we sample the output $\tilde{x}_0(t)$ of the filter in Definition 5.1 at unit rate we obtain the sequence

$$\tilde{q}[n] = \int_{-\infty}^{\infty} x(t) b_0(t - n) dt = \int_{-\infty}^{\infty} x(t) \tilde{\psi}_n^0(t) dt = x_n^0$$

where $\tilde{\psi}(t)$ is the ideal bandpass wavelet whose frequency response is given by (2.7). Furthermore, the self-similarity of $x(t)$ according to (5.2) implies

$$\beta^{-m/2} \tilde{q}[n] = \int_{-\infty}^{\infty} x(t) \tilde{\psi}_n^m(t) dt = x_n^m$$

where, as in earlier chapters, β is defined by

$$\beta = 2^{2H+1} = 2^\gamma. \quad (5.7)$$

Consequently, using the orthonormal wavelet synthesis formula (2.5a), $x(t)$ can be expressed as¹

$$x(t) = \sum_m \sum_n \beta^{-m/2} \tilde{q}[n] \tilde{\psi}_n^m(t), \quad (5.8)$$

from which we see that $x(t)$ is completely specified in terms of $\tilde{q}[n]$. We term $\tilde{q}[n]$ a *generating sequence* for $x(t)$ since, as we shall see, this representation leads to techniques for synthesizing useful approximations to homogeneous signals in practice.

More generally, as we shall see, (5.8) is a useful expansion for a broad family of homogeneous functions. However, a homogeneous function $x(t)$ defined by (5.8) is,

¹Henceforth, we shall assume in this chapter, and in the associated appendix, that all summations over m and n extend from $-\infty$ to ∞ unless otherwise noted.

specifically, energy-dominated if and only if $\tilde{q}[n]$ has finite energy, *i.e.*,

$$\sum_n \tilde{q}^2[n] < \infty.$$

This follows from the fact that since $\tilde{x}_0(t)$ in Definition 5.1 has the orthonormal expansion

$$\tilde{x}_0(t) = \sum_n \tilde{q}[n] \tilde{\psi}_n^0(t) \quad (5.9)$$

we have

$$\int_{-\infty}^{\infty} \tilde{x}_0^2(t) dt = \sum_n \tilde{q}^2[n]. \quad (5.10)$$

In turn, this observation implies that we may use (5.10) to define the following norm on \mathbf{E}^H :

$$\|x\|_{\tilde{\psi}} = \int_{-\infty}^{\infty} \tilde{x}_0^2(t) dt = \sum_n \tilde{q}^2[n]. \quad (5.11)$$

This is a valid norm provided we adopt the convention that two homogeneous functions $f(t)$ and $g(t)$ in \mathbf{E}^H are *equivalent* if their generalized Fourier transforms $F(\omega)$ and $G(\omega)$, respectively, are identical on $0 < \omega < \infty$. In turn, this implies that $f(t)$ and $g(t)$ are equivalent if they differ by a homogeneous function whose frequency content is concentrated at the origin $\omega = 0$. In the case that H is an integer $k \geq 0$, this would correspond to

$$f(t) - g(t) = Ct^k, \quad (5.12)$$

for an arbitrary constant C . Whether (5.12) characterizes all equivalent functions for $H = k \geq 0$, and whether there are equivalent functions for other values of H are open questions. In any case, with this notion of an equivalence class, \mathbf{E}^H is, in fact, complete with respect to the norm (5.11) and, hence, constitutes a Banach space. This is an immediate consequence of the isomorphism between \mathbf{E}^H and $L^2(\mathbb{Z})$.

More generally, we may define an inner product between two energy-dominated homogeneous signals $f(t)$ and $g(t)$, whose generating sequences under the bandpass

basis are $\tilde{a}[n]$ and $\tilde{b}[n]$, respectively, as

$$\langle f, g \rangle_{\tilde{\psi}} \triangleq \int_{-\infty}^{\infty} f_0(t) g_0(t) dt = \sum_n \tilde{a}[n] \tilde{b}[n] \quad (5.13)$$

where $f_0(t)$ and $g_0(t)$ are the outputs of the bandpass filter (5.3), and where the last equality is a consequence, again, of the orthonormality of the expansion (5.9). With this inner product, the induced norm is, of course, (5.11). Since E^H is complete, E^H therefore constitutes a Hilbert space. Furthermore, one can readily construct “self-similar” bases within E^H . Indeed, (5.8) immediately provides an orthonormal basis for E^H . In particular, for any $x(t) \in E^H$, we have the synthesis and analysis pair

$$x(t) = \sum_n \tilde{q}[n] \tilde{\theta}_n^H(t) \quad (5.14a)$$

$$\tilde{q}[n] = \langle x, \tilde{\theta}_n^H \rangle_{\tilde{\psi}} \quad (5.14b)$$

where one can easily verify that the basis functions

$$\tilde{\theta}_n^H(t) = \sum_m \beta^{-m/2} \tilde{\psi}_n^m(t) \quad (5.15)$$

are self-similar, orthogonal, and have unit norm.

The fact that the ideal bandpass basis is unrealizable means that (5.14a) is not a practical mechanism for synthesizing or analyzing homogeneous signals. However, as we will show, more practical wavelet bases are equally suitable for defining an inner product for the Hilbert space E^H . Specifically, we next show, that a broad class of wavelet bases can be used to construct such inner products, and that, as a consequence, some highly efficient algorithms arise for processing homogeneous signals.

In order to determine which orthonormal wavelet bases can be used to define inner products for \mathbf{E}^H , we must determine for which wavelets $\psi(t)$ the homogeneous signal

$$x(t) = \sum_m \sum_n \beta^{-m/2} q[n] \psi_n^m(t) \quad (5.16)$$

is energy-dominated when $q[n]$ has finite energy and, simultaneously, the sequence

$$q[n] = \int_{-\infty}^{\infty} x(t) \psi_n^0(t) dt$$

has finite energy when $x(t)$ is an energy-dominated homogeneous signal. In other words, we seek conditions on a wavelet basis such that $q[n] \in \ell^2(\mathbb{Z})$ will be isomorphic to \mathbf{E}^H . Our main result is a corollary of the following theorem. A proof of this theorem is provided in Appendix D.2.

Theorem 5.3 *Consider an orthonormal wavelet basis such that $\psi(t)$ has R vanishing moments for some integer $R \geq 1$, i.e.,*

$$\Psi^{(r)}(0) = 0, \quad r = 0, 1, \dots, R-1. \quad (5.17)$$

Then when any homogeneous signal $x(t)$ whose degree H is such that $\gamma = 2H + 1$ satisfies $0 < \gamma < 2R$ is filtered by an LTI system with impulse response $\psi(-t)$, the resulting process $x_0(t)$ has finite energy if and only if $x(t)$ is energy-dominated.

When the output $x_0(t)$ of the filter in Theorem 5.3 is sampled at unit rate, we obtain the sequence of wavelet coefficients

$$q[n] = \int_{-\infty}^{\infty} x(t) \psi_n^0(t) dt = x_n^0. \quad (5.18)$$

Furthermore, from the self-similarity of $x(t)$ we have also

$$\beta^{-m/2} q[n] = \int_{-\infty}^{\infty} x(t) \psi_n^m(t) dt = x_n^m. \quad (5.19)$$

Clearly then, associated with each $\psi(t)$ satisfying (5.17) is a generating sequence $q[n]$ for the homogeneous signal $x(t)$. Also, again because of the orthogonality of $\psi(t)$ with respect to its unit translates, we have

$$\int_{-\infty}^{\infty} x_0^2(t) dt = \sum_n q^2[n].$$

Combining this result with that of Theorem 5.3 allows us to immediately deduce our main result:

Corollary 5.4 *For any wavelet basis satisfying the conditions of Theorem 5.3, the homogeneous signal $x(t)$*

$$x(t) = \sum_m \sum_n \beta^{-m/2} q[n] \psi_n^m(t) \quad (5.20)$$

is energy-dominated (i.e., $x(t) \in \mathbf{E}^H$) if and only if $q[n]$ has finite energy (i.e., $q[n] \in L^2(\mathbb{Z})$).

This corollary implies that we may choose for our Hilbert space \mathbf{E}^H from among a large number of inner products whose induced norms are all equivalent. In particular, for any wavelet $\psi(t)$ with sufficient vanishing moments, we may define the inner product between two functions $f(t)$ and $g(t)$ in \mathbf{E}^H whose generating sequences are $a[n]$ and $b[n]$, respectively, as

$$\langle f, g \rangle_\psi = \sum_n a[n] b[n]. \quad (5.21)$$

It is important to emphasize that this collection of inner products is not necessarily exhaustive. Even for wavelet-based inner products, Corollary 5.4 asserts only that the vanishing moment condition is sufficient to ensure that the inner product generates an equivalent norm. It is conceivable, however, that the vanishing moment condition is not a necessary condition.

In any case, each wavelet-based inner product leads immediately to an orthonormal self-similar basis for \mathbf{E}^H : if $x(t) \in \mathbf{E}^H$, then

$$x(t) = \sum_n q[n] \theta_n^H(t) \quad (5.22a)$$

$$q[n] = \langle x, \theta_n^H \rangle_\psi \quad (5.22b)$$

where, again, the basis functions

$$\theta_n^H(t) = \sum_m \beta^{-m/2} \psi_n^m(t) \quad (5.23)$$

are all self-similar, mutually orthogonal, and have unit norm. For the case $H = 0$, Fig. 5-1 depicts the self-similar basis functions $\theta_4^0(t)$, $\theta_5^0(t)$, $\theta_6^0(t)$, and $\theta_7^0(t)$ corresponding to the Daubechies 5th-order compactly-supported wavelet basis. These functions were generated by evaluating the summation (5.23) over a large but finite range of scales m . We emphasize that $q[n]$ is only a unique characterization of $x(t)$ when we associate it with a particular choice of wavelet $\psi(t)$. In general, every different wavelet decomposition of $x(t)$ will yield a different $q[n]$, though all will have finite-energy.

For an arbitrary non-homogeneous signal $x(t)$, the sequence

$$q[n] = \langle x, \theta_n^H \rangle_\psi$$

defines the projections of $x(t)$ onto \mathbf{E}^H , so that

$$\hat{x}(t) = \int_{-\infty}^{\infty} q[n] \theta_n^H(t) dt$$

represents the closest homogeneous signal to $x(t)$

$$\hat{x}(t) = \arg \min_{y(t) \in \mathbf{E}^H} \|y - x\|_\psi$$

with respect to the induced norm $\|\cdot\|_\psi$. In Chapter 6, it will be apparent how such projections arise rather naturally in treating problems of estimation with homoge-

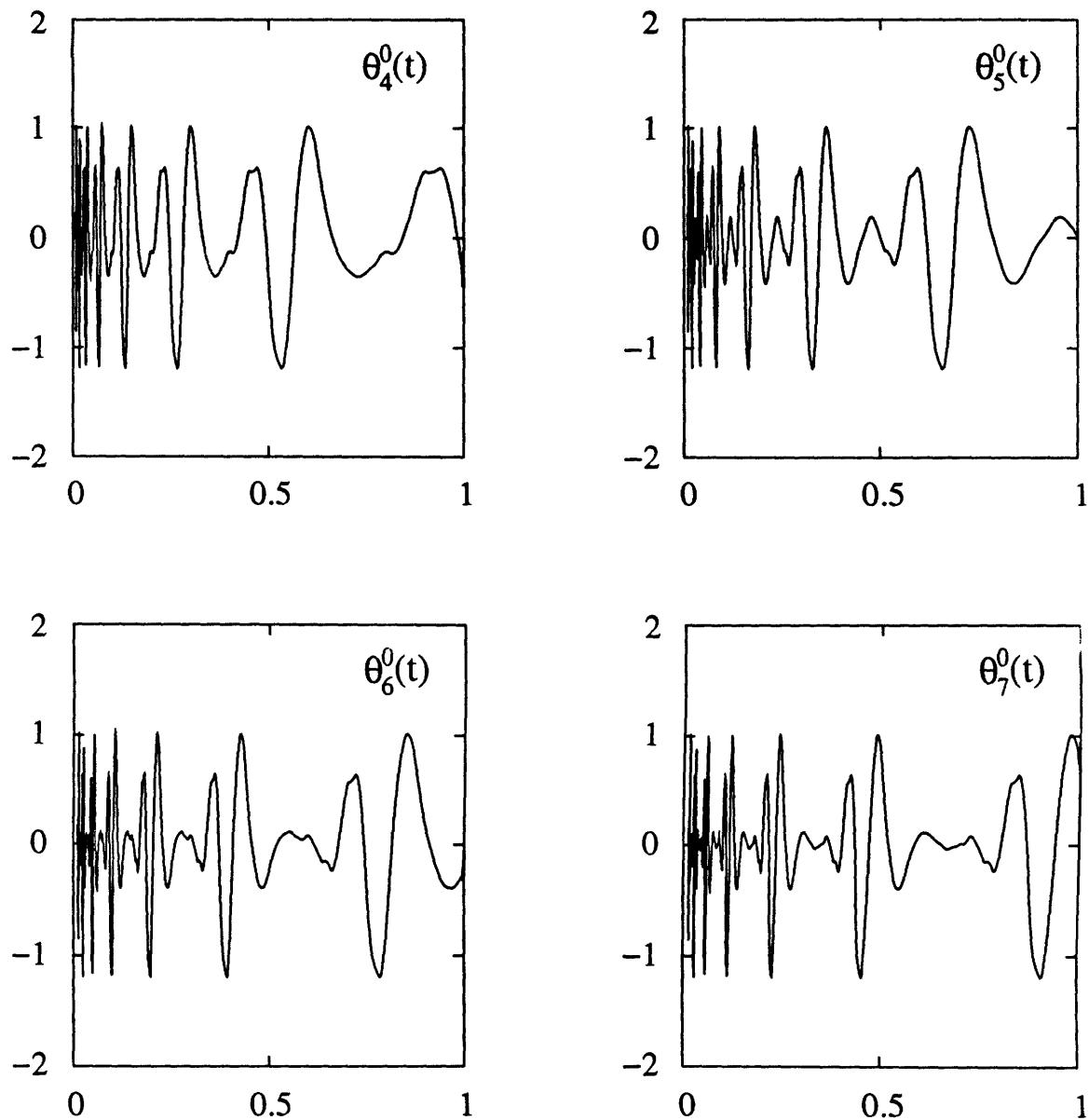


Figure 5-1: The self-similar basis functions $\theta_4^H(t)$, $\theta_5^H(t)$, $\theta_6^H(t)$, and $\theta_7^H(t)$ of an orthonormal basis for E^H , $H = 0$.

neous signals.

Finally, we remark that wavelet-based characterizations give rise to a convenient expression for the generalized Fourier transform of an energy-dominated homogeneous signal, $x(t)$. In particular, if we take the Fourier transform of (5.16) we get, via some routine algebra,

$$X(\omega) = \sum_m 2^{-(H+1)m} \Psi(2^{-m}\omega) Q(2^{-m}\omega) \quad (5.24)$$

where $Q(\omega)$ is the Fourier transform of $q[n]$. This spectrum is to be interpreted in the sense of Theorem 5.1, *i.e.*, $X(\omega)$ defines the spectral content of the output of a bandpass filter at every frequency ω within the passband.

In summary, we have shown that a broad class of wavelet-based norms are equivalent for \mathbf{E}^H in a mathematical sense, and that each of these norms is associated with a particular inner product. This leads one to speculate whether *every* equivalent norm for \mathbf{E}^H can be associated with a wavelet basis, in which case the basis functions associated with every orthonormal basis for \mathbf{E}^H could be expressed in terms of some wavelet according to (5.23). This issue remains unresolved. In any case, regardless of whether the collection of inner products we construct is exhaustive or not, they at least constitute an highly convenient and practical collection from which to choose in any given application involving the use of homogeneous signals.

5.2 Power-Dominated Homogeneous Signals

Energy-dominated homogeneous signals have infinite energy. In fact, most have infinite power as well. However, there are other infinite power homogeneous signals that are not energy-dominated. In this section, we consider a more general class of infinite-power homogeneous signals that will be of interest in Chapter 6. We begin with what is a natural definition.

Definition 5.5 *A wide-sense homogeneous signal $x(t)$ is said to be power-dominated if when $x(t)$ is filtered by an ideal bandpass filter with frequency response (5.3) the*

resulting signal $\tilde{x}_0(t)$ has finite power, i.e.,

$$\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \tilde{x}_0^2(t) dt < \infty.$$

We use the notation P^H to designate the class of power-dominated homogeneous signals of degree H . Note that while our definition necessarily includes the energy-dominated signals, which have zero power, insofar as our discussion is concerned they constitute a degenerate case.

Before proceeding, we recall some basic definitions and useful relationships for deterministic finite-power signals. Reference [68] contains a more thorough exposition of these results. A finite-power signal $f(t)$ has a (deterministic) autocorrelation

$$R_f(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T f(t) f(t - \tau) dt$$

whose Fourier transform is the (deterministic) power spectrum

$$\begin{aligned} S_f(\omega) &= \int_{-\infty}^{\infty} R_f(\tau) e^{-j\omega\tau} d\tau \\ &= \lim_{T \rightarrow \infty} \frac{1}{2T} \left| \int_{-T}^T f(t) e^{-j\omega t} dt \right|^2. \end{aligned}$$

Analogous formulae are obtained for finite power sequences. The deterministic autocorrelation and power spectrum satisfy many of the LTI filtering and sampling properties satisfied by the corresponding quantities for stationary random processes.

Analogous to the energy-dominated case, we can establish the following theorem describing the spectral properties of power-dominated homogeneous signals.

Theorem 5.6 *When a power-dominated homogeneous signal $x(t)$ is filtered by an ideal bandpass filter with frequency response (5.4), the resulting signal $y(t)$ has finite power and a power spectrum of the form*

$$S_y(\omega) = \begin{cases} S_x(\omega) & \omega_L < |\omega| \leq \omega_U \\ 0 & \text{otherwise} \end{cases} \quad (5.25)$$

where $S_x(\omega)$ is some function that is independent of ω_L and ω_U and has octave-spaced ripple, i.e., for all integers k ,

$$|\omega|^{2H+1} S_x(\omega) = |2^k \omega|^{2H+1} S_x(2^k \omega). \quad (5.26)$$

The details of the proof of this theorem are contained in Appendix D.3, although it is identical in style to the proof of its counterpart, Theorem 5.2. Note that since $S_x(\omega)$ in this theorem does not depend on ω_L or ω_U , this function may be interpreted as the generalized power spectrum of $x(t)$. Furthermore, the relation (5.26) implies that signals in \mathbf{P}^H have a generalized *time-averaged* power spectrum that is $1/f$ -like, i.e.

$$S_x(\omega) \sim \frac{1}{|\omega|^\gamma}.$$

where γ is as defined in (5.7). However, we again emphasize that we reserve the term “ $1/f$ process” or “ $1/f$ signal” for the *statistically* self-similar random processes defined in Chapter 3. Note too that there is no notion of time-averaging in the spectrum defining a $1/f$ process.

In turn, Theorem 5.6 directly implies that a homogeneous signal $x(t)$ is power-dominated if and only if its *generating sequence* $\tilde{q}[n]$ in the ideal bandpass wavelet basis has finite power, i.e.,

$$\lim_{L \rightarrow \infty} \frac{1}{2L + 1} \sum_{n=-L}^L \tilde{q}^2[n] < \infty.$$

Similarly we can readily deduce from the results of Section 5.1 that, in fact, for any orthonormal wavelet basis with sufficient vanishing moments R that $0 < \gamma < 2R$ (where $\gamma = 2H + 1$), the *generating sequence* for a homogeneous signal of degree H in that basis has finite power if and only if the signal is power-dominated. This implies that, for such bases, when we use (5.22a) to synthesize a homogeneous signal $x(t)$ using an arbitrary finite power sequence $q[n]$, we are assured that $x(t) \in \mathbf{P}^H$. Likewise, when we use (5.22b) to analyze any signal $x(t) \in \mathbf{P}^H$, we are assured that $q[n]$ has finite power. Again we emphasize that different choices of wavelet basis in

such an analysis will yield different generating sequences, all of which, however, have finite power.

Energy-dominated homogeneous signals of arbitrary degree H can be highly regular, at least away from $t = 0$. In contrast, power-dominated homogeneous signals typically have a fractal structure similar to $1/f$ processes of corresponding degree H . In fact, it is a reasonable conjecture that the Hausdorff-Besicovitch dimension, when defined, is identical for the two types of signals. Indeed, despite their obvious structural differences, power-dominated homogeneous signals and $1/f$ processes “look” remarkably similar in a qualitative sense. This is apparent in Fig. 5-2, where we depict the sample path of a $1/f$ process along side a randomly-generated power-dominated homogeneous signal of the same degree. We reiterate that in Fig. 5-2(a), the self-similarity of the $1/f$ process is statistical, while in Fig. 5-2(b), the self-similarity of the homogeneous signal is deterministic.

We can quantify the similarity between the two types of signals through an observation about their spectra. In general, we remarked that for a given H , both exhibit power law spectral relationships with the same parameter γ . The following theorem further substantiates this for the case of randomly-generated power-dominated homogeneous signals. The details of the proof are contained in Appendix D.4.

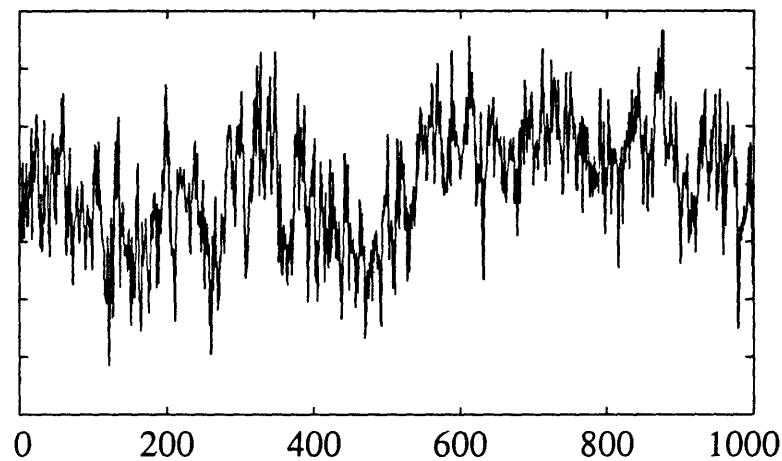
Theorem 5.7 *For any orthonormal wavelet basis in which $\psi(t)$ has R th order regularity for some $R \geq 1$, the random process $x(t)$ synthesized according to*

$$x(t) = \sum_m \sum_n \beta^{-m/2} q[n] \psi_n^m(t), \quad (5.27)$$

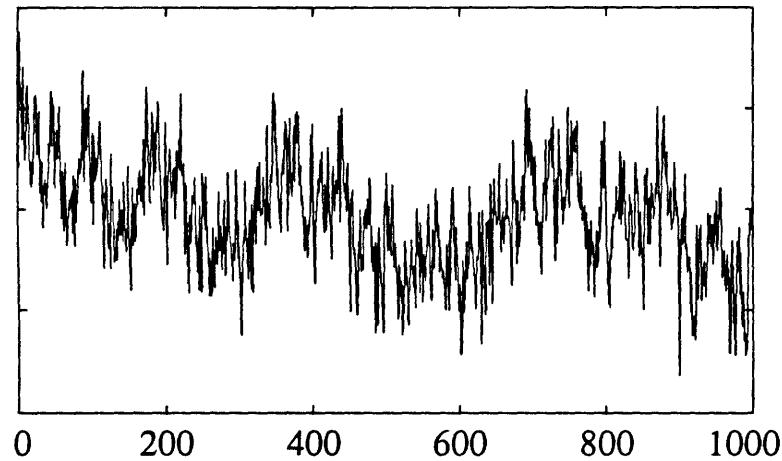
using a correlation-ergodic (e.g., Gaussian), zero-mean, stationary white random sequence $q[n]$ of variance σ^2 , has a generalized time-averaged power spectrum of the form

$$S_x(\omega) = \sigma^2 \sum_m 2^{-\gamma m} |\Psi(2^{-m}\omega)|^2. \quad (5.28)$$

Note that the time-averaged spectrum (5.28) is identical to the time-averaged spectrum (3.36) for the wavelet-based synthesis of $1/f$ processes described in Section 3.2.2. However, we must be careful not to misinterpret this result. It does not



(a) A sample function of a $1/f$ process.



(b) A randomly-generated power-dominated homogeneous signal.

Figure 5-2: Comparison between the sample path of a $1/f$ process and a power-dominated homogeneous signal. Both correspond to $\gamma = 1$ (i.e., $H = 0$).

suggest that (5.27) is a reasonable approach for synthesizing $1/f$ processes. Indeed, it would constitute a very poor model for $1/f$ -type behavior based on the analysis results of Section 3.2.2: when $1/f$ processes are decomposed into wavelet bases we get statistical rather than deterministic similarity from scale to scale. Instead, the theorem remarks that the *time-averaged* second order statistics of the two types of signals are the same. Consequently, one would anticipate that distinguishing $1/f$ processes from power-dominated homogeneous signals based on spectral analysis alone would be rather difficult. Nevertheless, the tremendous structural differences between the two means that they may be readily distinguished using other techniques such as, for example, wavelet-based analysis.

Equation (5.28) corresponds to the superposition of the spectra associated with each scale or octave-band in the wavelet-based synthesis. In general, we would expect the spectrum of $x(t)$ to be the superposition of the spectra of the individual channels together with their cross-spectra. However, the time-averaged cross-spectra in this scenario are zero, which is a consequence of the fact that the white sequence $q[n]$ is modulated at different rates in each channel. Indeed, the time-averaged correlation is zero between $q[n]$ and $q[2^m n]$ for any $m \geq 1$; that is, white noise is uncorrelated with dilated and compressed versions of itself.

Because (5.28) and (3.36) are identical, we can use Theorem 3.4 to conclude that the spectra of a class of randomly generated power-dominated homogeneous signals are bounded on any finite interval of the frequency axis that does not include $\omega = 0$. However, there are many $x(t) \in \mathbf{P}^H$ whose spectra is not bounded in this manner. An interesting and potentially important subclass of power-dominated homogeneous signals with such unbounded spectra are those for which $\tilde{x}(t)$ as defined in Definition 5.5 is *periodic*. This class of power-dominated homogeneous signals will be referred to as *periodicity-dominated*. It is straightforward to establish that these homogeneous signals have the property that when passed through an arbitrary bandpass filter of the form (5.4) the output is periodic as well. These processes have a power spectrum consisting of impulses (spikes) whose *areas* decay according to a $1/|\omega|^\gamma$ relationship. An important class of periodicity-dominated homogeneous signals can be generated

through a wavelet-based synthesis of the form (5.22) in which the generating sequence $q[n]$ is periodic.

5.3 Discrete-Time Algorithms for Processing Homogeneous Signals

Orthonormal wavelet representations provide some useful insights into homogeneous signals $x(t)$. For instance, because $q[n]$ is replicated at each scale in the representation (5.20), the detail signals $D_m x(t)$, which represent $q[n]$ modulated into a particular octave-band, are simply amplitude-scaled and time-dilated or compressed versions of one another. The corresponding time-frequency portrait of a homogeneous signal (of degree $H = -1/2$, for convenience) is depicted in Fig. 5-3, from which the scaling properties are apparent. We emphasize again that the partitioning in such time-frequency portraits is idealized. In general, there is both spectral and temporal overlap between cells.

Wavelet representations also lead to some highly efficient algorithms for synthesizing, analyzing, and processing homogeneous signals, just as they did in the case of $1/f$ processes. The signal processing structures we develop in this section are a consequence of applying the DWT algorithm to the highly structured form of the wavelet coefficients of homogeneous signals.

We have already encountered one discrete-time representation for a homogeneous signal $x(t)$, namely that in terms of a generating sequence $q[n]$

$$x(t) \longleftrightarrow q[n] \quad (5.29)$$

which corresponds to the coefficients of the expansion of $x(t)$ in an orthonormal basis $\{\theta_n^H(t)\}$ for E^H . When the $\theta_n^H(t)$ are derived from a wavelet basis, a second discrete-time representation for $x(t)$ is available, which we now discuss.

Consider the coefficients a_n^m characterizing the resolution- 2^m approximation of a homogeneous signal $x(t)$ with respect to a particular multiresolution signal analysis.

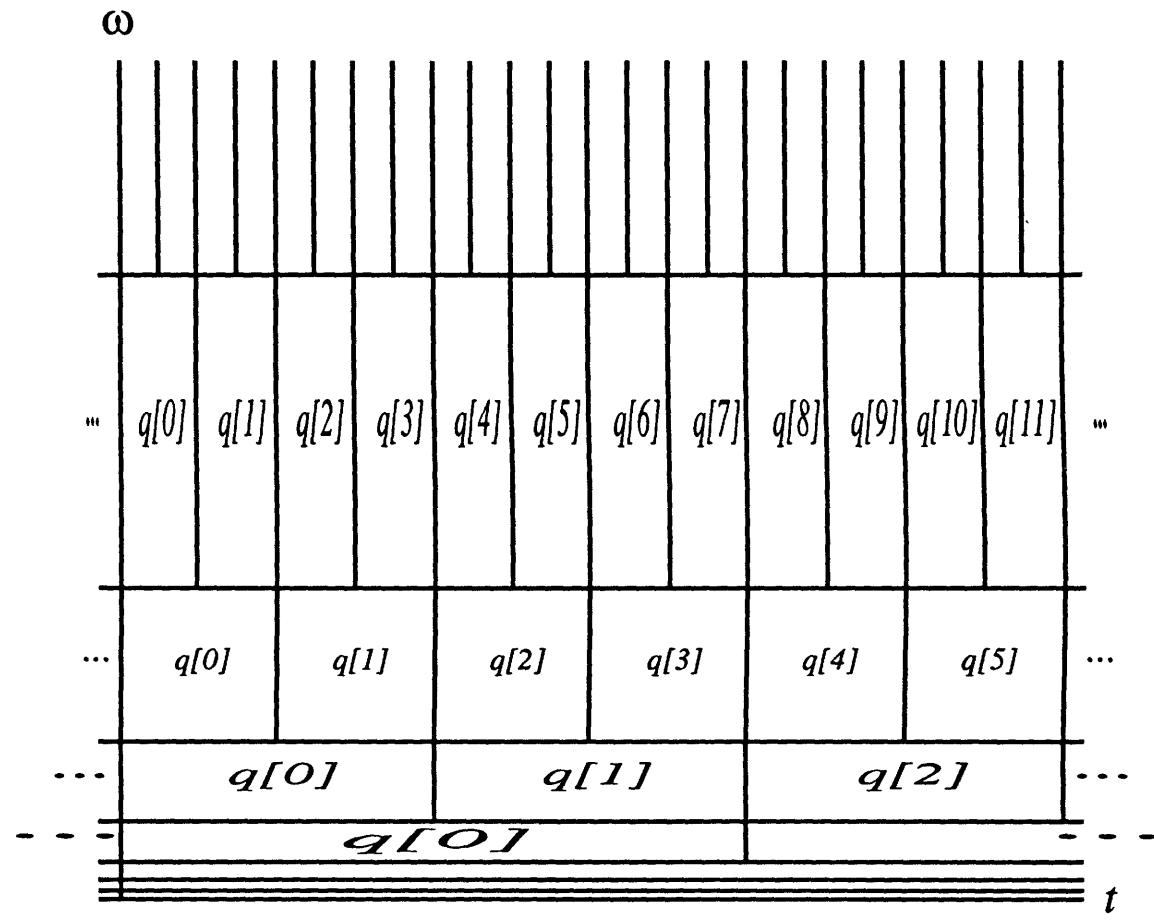


Figure 5-3: The time-frequency portrait of a homogeneous signal of degree $H = -1/2$.

If $\phi(t)$ is the scaling function associated with this analysis, then these coefficients are defined through the projections (2.13), *viz.*,

$$a_n^m = \int_{-\infty}^{\infty} x(t) \phi_n^m(t) dt.$$

Using the self-similarity of $x(t)$ it is straightforward to show that these coefficients are identical at all scales to within a scale factor, *i.e.*,

$$a_n^m = \beta^{-m/2} a_n^0. \quad (5.30)$$

Consequently, the *characteristic sequence* associated with $x(t)$, defined via

$$p[n] = a_n^0 \quad (5.31)$$

is an alternative discrete-time characterization of $x(t)$, since knowledge of $p[n]$ is sufficient to reconstruct $x(t)$ to arbitrary accuracy. We stress that the characteristic sequence associated with $x(t)$ is not unique: distinct multiresolution signal analyses generally yield different characteristic sequences for any given homogenous signal. Furthermore, we shall require the wavelet associated with any multiresolution analysis we consider to have sufficient vanishing moments that it meets the conditions of Theorem 5.3.

The characteristic sequence $p[n]$ is associated with a resolution-limited approximation to a homogeneous signal $x(t)$. Specifically, $p[n]$ represents unit-rate samples of the output of a roughly lowpass filter with frequency response $\Phi^*(\omega)$ driven by $x(t)$. Because frequencies in the neighborhood of the spectral origin, where there is a preponderance of energy, are passed by such a filter, $p[n]$ will often have infinite energy or, worse, infinite power, even when $q[n]$ has finite energy ($x(t) \in E^H$). Still more severe, there will be cases in which $p[n]$ is actually singular and, hence, cannot be “plotted.” In fact, the characteristic sequence $p[n]$ is typically a *generalized sequence* in the same sense that its corresponding homogeneous signal $x(t)$ is a generalized function.

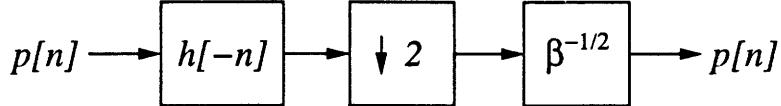


Figure 5-4: The discrete-time self-similarity identity for a characteristic sequence $p[n]$.

The characteristic sequence can be viewed as a *discrete-time* homogeneous signal. As such it satisfies the discrete-time self-similarity relation²

$$\beta^{1/2}p[n] = \sum_k h[k - 2n]p[k] \quad (5.32)$$

which is readily obtained using (5.30) with (5.31) in the DWT analysis equation (2.21a). Indeed, (5.32) is a statement that when $p[n]$ is lowpass filtered with the conjugate filter whose unit-sample response is $h[-n]$ and then downsampled, we get back an amplitude-scaled version of $p[n]$, as depicted in Fig. 5-4. And although characteristic sequences are generalized sequences, when highpass filtered with the corresponding conjugate highpass filter whose unit-sample response is $g[-n]$, the output is a finite energy or finite power sequence, depending on whether $p[n]$ corresponds to a homogeneous signal $x(t)$ that is energy-dominated or power-dominated, respectively. Consequently, we can analogously classify $p[n]$ as energy-dominated in the former case, and power-dominated in the latter case. In fact, when the output of the highpass filter is downsampled at rate two, we recover the characteristic sequence $q[n]$ associated with the decomposition of $x(t)$ into the corresponding wavelet basis, *i.e.*,

$$\beta^{1/2}q[n] = \sum_k g[k - 2n]p[k]. \quad (5.33)$$

This can be readily verified by substituting (5.18) with (5.19), and (5.31) with (5.30), into the DWT analysis equation (2.21b).

From a different perspective, (5.33) provides a convenient mechanism for obtaining the representation for a homogeneous signal $x(t)$ in terms of its generating sequence

²Relations of this type may be considered discrete-time counterparts of the *dilation equations* considered by Strang in [16].

$q[n]$ from one in terms of its corresponding characteristic sequence $p[n]$, *i.e.*,

$$p[n] \longrightarrow q[n].$$

To obtain the reverse mapping

$$q[n] \longrightarrow p[n]$$

is less straightforward. For an arbitrary sequence $q[n]$, the associated characteristic sequence $p[n]$ is the solution to the linear equation

$$\beta^{-1/2}p[n] - \sum_k h[n-2k]p[k] = \sum_k g[n-2k]q[k], \quad (5.34)$$

as can be verified by specializing the DWT synthesis equation (2.21c) to the case of homogeneous signals. There appears to be no direct method for solving this equation; even in the frequency domain, where, for finite energy $q[n]$, (5.34) becomes

$$\beta^{-1/2}P(\omega) = H(\omega)P(2\omega) + G(\omega)Q(2\omega)$$

with $P(\omega)$ the generalized Fourier transform of $p[n]$, the problem appears intractable. However, the DWT synthesis pyramid of Fig. 2-6(b) suggests a convenient and efficient iterative algorithm for constructing $p[n]$ from $q[n]$. In particular, denoting the estimate of $p[n]$ on the i th iteration by $p^{[i]}[n]$, the algorithm is

$$p^{[0]}[n] = 0 \quad (5.35a)$$

$$p^{[i+1]}[n] = \beta^{1/2} \sum_k \{h[n-2k]p^{[i]}[k] + g[n-2k]q[k]\}. \quad (5.35b)$$

This recursive upsample-filter-merge algorithm, depicted in Fig. 5-5, can be interpreted as repeatedly modulating $q[n]$ with the appropriate gain into successively lower octave bands of the frequency interval $0 \leq |\omega| \leq \pi$, and we note that the

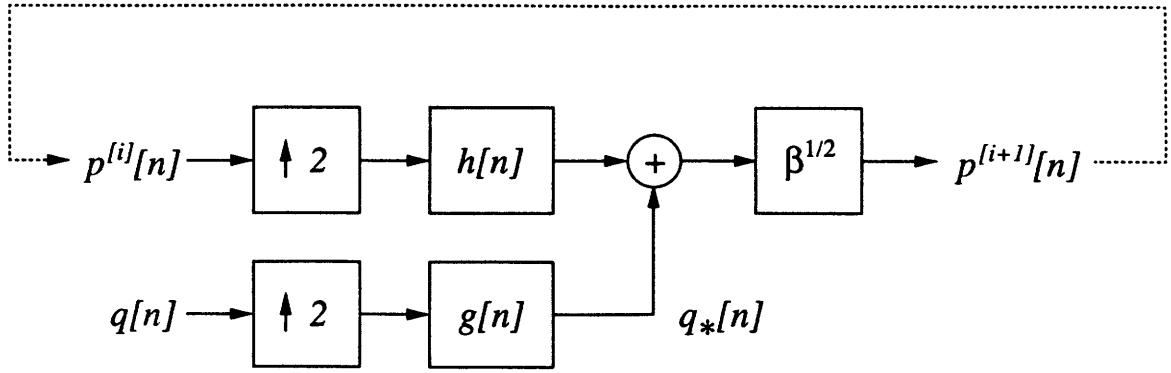


Figure 5-5: Iterative algorithm for the synthesis of the characteristic sequence $p[n]$ of a homogeneous signal $x(t)$ from its generating sequence $q[n]$. The notation $p^{[i]}[n]$ denotes the value of $p[n]$ at the i th iteration.

precomputable quantity

$$q_*[n] = \sum_k g[n - 2k]q[k]$$

represents the $q[n]$ modulated into essentially the upper half band of frequencies.

The modulation-based interpretation of the algorithm suggests that issues of convergence as $i \rightarrow \infty$ are ultimately tied to issues of whether or not $p[n]$ is a *regular* sequence. In particular, when $p[n]$ is well-behaved, the algorithm exhibits rapid convergence. The Fourier transform of the error between iterations is readily derived as

$$\begin{aligned} P^{[i+1]}(\omega) - P^{[i]}(\omega) &= \beta^{(i+1)/2} \left[\prod_{k=0}^{i-1} H(2^k \omega) \right] G(2^i \omega) Q(2^{i+1} \omega) \\ &= (2\beta)^{i/2} \beta^{1/2} \Phi(2^i \omega) G(2^i \omega) Q(2^{i+1} \omega) / \Phi(\omega) \end{aligned}$$

for $|\omega| < \pi$, where the last inequality follows from the repeated application of (A.2a). In turn, using Parseval's relation, the total energy difference between iterates can be obtained as

$$\sum_n |p^{[i+1]}[n] - p^{[i]}[n]|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |P^{[i+1]}(\omega) - P^{[i]}(\omega)|^2 d\omega.$$

However, it is important to keep in mind that any practical application of homogeneous signals will ultimately only involve scaling behavior over a *finite* range of

scales, corresponding to $q[n]$ modulated into a finite range of adjacent octave bands. Consequently, only a finite number of iterations would be used with the algorithm (5.35). More generally, this also means that many of the theoretical issues associated with homogeneous signals concerning singularities and convergence do not present practical difficulties in the application of deterministically self-similar signals, as will be apparent in our developments of the next chapter.

In our closing remarks, we mention that there would appear to be important connections to be explored between the self-similar signal theory described here and the work of Barnsley, *et al.*, [69] on deterministically self-affine one-dimensional and multi-dimensional signals. Interestingly, the recent work of Malassenet and Mersereau [70] has shown that these *iterated function systems* have efficient representations in terms of wavelet bases as well.

Chapter 6

Fractal Modulation

There are a number of interesting potential applications for the wide-sense homogeneous signal theory developed in the last chapter. In this chapter, we consider a specific one as an indication of the direction that some applications may take. In particular, we explore the use of homogeneous signals as a modulating waveforms in a communication-based context. We begin by considering an idealized but general channel model, and proceed to demonstrate that the use of homogeneous waveforms in such channels is at least natural, if not optimal, and leads to a novel multirate modulation strategy in which data is transmitted simultaneously at multiple rates. We remark at the outset that while multirate modulation has received attention in the recent literature (see, *e.g.*, [14] [71] [72] [73]), the work is, at best, peripherally related.

Consider the problem of designing a communication system for the transmission of continuous- or discrete-valued data sequences over a noisy and unreliable continuous-amplitude, continuous-time channel. As depicted in Fig. 6-1, the classical structure for such a system involves a modulator at the transmitter that embeds the data sequence $q[n]$ into a signal $x(t)$ which is sent over the channel. At the receiver, a demodulator processes the distorted signal $r(t)$ to extract an optimal estimate of the data sequence $\hat{q}[n]$.

In a typical scenario, the channel would be “open” for some time-interval T , during which it has a particular bandwidth W and signal-to-noise ratio (SNR). In

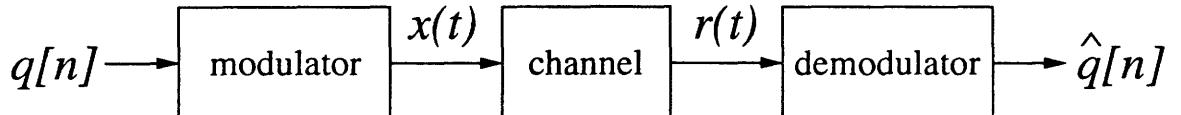


Figure 6-1: A communication system for transmitting a continuous- or discrete-amplitude data sequence $q[n]$ over a noisy and unreliable continuous-amplitude, continuous-time channel.

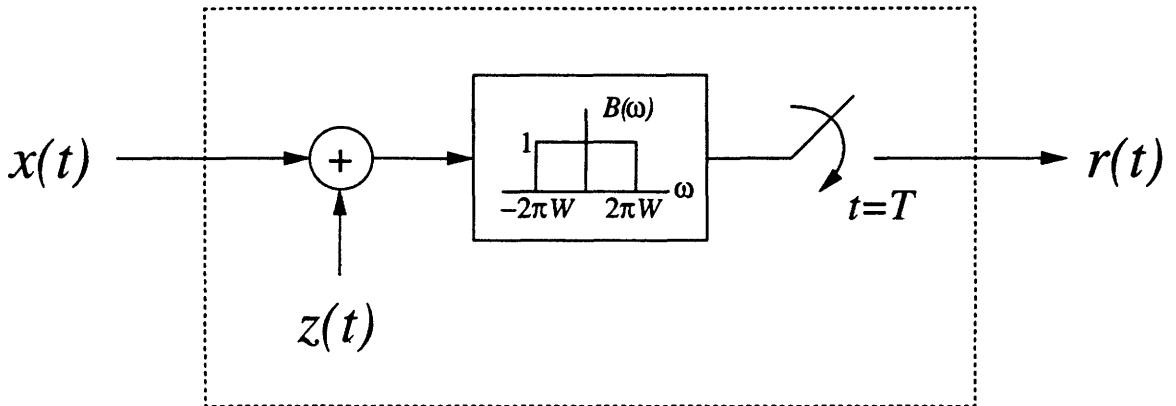


Figure 6-2: The channel model for a typical communications scenario.

general, this rather generic model can be used to capture both characteristics of the transmission medium and constraints inherent in one or more receivers. When the noise characteristics are additive, the overall channel model is as depicted in Fig. 6-2, where $z(t)$ represents the noise process.

When either the bandwidth or duration parameters of the channel are known *a priori*, there are many well-established approaches for designing an optimum communication system for transmitting $q[n]$ reliably. However, there are a variety of military and commercial communication contexts in which *both* the bandwidth and duration parameters are either unknown or at least inaccessible to the transmitter¹. This case, by contrast, has received comparatively less attention in the communications literature, although it encompasses a range of both point-to-point and broadcast communication scenarios, including

- channels subject to hostile jamming

¹Equivalently, we might be interested in designing systems whose performance is optimal over a range of possible bandwidth and duration parameter combinations.

- broadcast communication involving disparate receivers
- multiple access channels involving time-, frequency-, or code-division multiplexing (*i.e.*, TDMA, FDMA, or CDMA)
- packet-switching
- fading channels, such as meteor burst, ocean acoustic, or mobile radio channels
- low probability of intercept (LPI) communication

In these situations, it is often desirable to have a system with the following performance characteristics:

1. Given a duration-bandwidth product $T \times W$ that exceeds some threshold, we must be able to transmit $q[n]$ without error in the absence of noise ($z(t) = 0$).
2. Given increasing duration-bandwidth product in excess of this threshold, we must be able to transmit $q[n]$ with increasing fidelity in the presence of noise. Furthermore, in the limit of infinite duration-bandwidth product, perfect transmission should be achievable at any finite SNR.

Note that the first of these requirements implies we ought to be able, at least in principle, to recover $q[n]$ from arbitrarily little bandwidth given sufficient duration, or, alternatively, from arbitrarily little duration given sufficient bandwidth. The second requirement implies that we ought to be able to obtain better estimates of $q[n]$ the longer the receiver is able to listen, or the greater the bandwidth it has available. Consequently, the modulation must contain redundancy to exploit for this error correction capability. As we shall see, the use of homogeneous signals for transmission is naturally suited to fulfilling these requirements.

In general, we would anticipate that the time-bandwidth threshold involved would be a function of the length L of the data sequence. For this reason, it is more convenient to phrase the discussion in terms of a rate-bandwidth ratio R/W . Indeed, when the duration constraint T is transformed into a symbol rate constraint R defined

as

$$R = L/T. \quad (6.1)$$

our time-bandwidth threshold is transformed into a rate-bandwidth ratio R/W threshold that is independent of sequence length.

In a conventional communication system in which the bandwidth is prescribed, the information-bearing sequence $q[n]$ is modulated into the available bandwidth W at some rate R such that in the absence of noise, perfect recovery of $q[n]$ is possible. The *spectral efficiency* η of such a system is usually defined as the maximum rate R achievable for a given bandwidth W , *i.e.*,

$$\eta = \max \frac{R}{W} \quad (6.2)$$

which is measured in symbols/sec/Hz. More efficient systems can achieve a higher rate for a given bandwidth, or, equivalently, support a given rate with less bandwidth.

A reasonable approach to spectrally efficient transmission can be constructed as follows. A transmitter modulates the data sequence $q[n]$ onto a lowpass waveform by expanding it into an orthonormal basis to generate the lowpass transmitted waveform

$$x(t) = \sum_n q[n] \sqrt{R_0} \operatorname{sinc}(R_0 t - n)$$

where

$$\operatorname{sinc}(t) = \begin{cases} 1 & t = 0 \\ \frac{\sin \pi t}{\pi t} & \text{otherwise} \end{cases},$$

and where R_0 is a fixed parameter of the system. In turn, the receiver recovers $q[n]$ from the projections

$$q[n] = \int_{-\infty}^{\infty} x(t) \sqrt{R_0} \operatorname{sinc}(R_0 t - n) dt$$

which are implemented as a sequence of filter-and-sample operations. Since

$$\mathcal{F}\left\{\sqrt{R_0}\text{sinc}(R_0t)\right\} = \begin{cases} 1/\sqrt{R_0} & |\omega| \leq \pi R_0 \\ 0 & \text{otherwise} \end{cases}$$

this system achieves a rate of $R = R_0$ symbols/sec using a (double-sided) bandwidth of $W = R_0$ Hz. Thus, the system is characterized by a spectral efficiency of

$$\eta_0 = 1 \text{ symbol/sec/Hz.} \quad (6.3)$$

Furthermore, it is power efficient in the presence of additive stationary white noise. Needless to say, however, such a system is not practical. Indeed, the lack of temporal localization in $\text{sinc } t$ makes the system not only unrealizable, but have infinite delay as well. Nevertheless, useful approximations can be implemented using other bases with better time localization properties and with little penalty in spectral characteristics. Consequently, the idealized system is a useful baseline for the purposes of comparison, and (6.3) is the corresponding performance benchmark. We shall refer to this modulation as *uniformly most spectrally efficient* (UMSE). Indeed, it represents the potential performance of a system in which the transmitter has perfect knowledge of the rate-bandwidth characteristics of the channel. Since under our requirements the transmitter is constrained to have no knowledge of these parameters, and, hence, cannot reconfigure itself accordingly, UMSE modulation provides a useful performance bound in our analysis.

Note that choosing a different value of R_0 effects a time-scaling of the system. The new system has the same efficiency but attains it for a different combination of rate and bandwidth. However, any one such system is only efficient when operating at the rate and bandwidth for which it was designed. That is, given a doubling of the available bandwidth, the system will still function properly, though it will not be able to exploit the potential to double the symbol rate afforded by the excess bandwidth. To do this would require a different, re-scaled system. Because this involves a re-designed transmitter, this change requires that the transmitter have knowledge of the

bandwidth available.

We now turn our attention to the design of a transmission scheme based upon the concept of embedding the data to be transmitted into a wide-sense homogeneous signal. As we shall see, this leads to a systems that maintains its spectral efficiency over a broad range of rate-bandwidth combinations using a fixed transmitter configuration. Due to the fractal properties of the transmitted signals, we refer to the resulting scheme as *fractal modulation*.

6.1 Transmitter Design: Modulation

Consider a transmitter that expands $q[n]$ in an orthonormal self-similar basis $\tilde{\theta}_n^H(t)$ of arbitrary degree H

$$x(t) = \sum_n q[n] \tilde{\theta}_n^H(t)$$

where the basis is constructed from the ideal bandpass wavelet $\tilde{\psi}(t)$ according to (5.15), *viz.*,

$$\tilde{\theta}_n^H(t) = \sum_m \beta^{-m/2} \tilde{\psi}_n^m(t).$$

When $q[n]$ has finite-power, as we shall generally assume, the resulting $x(t)$ is a power-dominated homogeneous signal and corresponds to a time-frequency portrait of the form depicted in Fig. 5-3. More generally, we recognize this as a *multirate modulation* of $q[n]$: in the m th channel, $q[n]$ is modulated at rate 2^m using a (double-sided) bandwidth of 2^m Hz. Furthermore, the energy per symbol used in successively higher channels (m) scales by

$$\beta = 2^\gamma = 2^{2H+1}.$$

It is apparent that from such a transmission, a suitably designed receiver can recover $q[n]$ at rate 2^m using a *baseband* bandwidth of 2^{m+1} Hz, from which we deduce that this modulation has a spectral efficiency

$$\eta_F = (1/2) \text{ symbol/sec/Hz},$$

which is half that of our baseline system². In effect, this loss in efficiency is the price paid to enable a receiver to use any of an arbitrarily broad range of rate-bandwidth combinations to accommodate the characteristics of the channel or its own internal processing constraints.

In Fig. 6-3, we illustrate the rate-bandwidth tradeoffs that can be made at the receiver in demodulating the transmitted data. In particular, in the absence of noise a receiver can, in principle, perfectly recover the information sequence $q[n]$ using rate-bandwidth combinations that lie on or below the solid curve. Note that, according to our definition, the rate corresponding to $B = 1$ Hz is $R = \eta_F$, *i.e.*, the spectral efficiency of the modulation scheme. The stepped character of this curve reflects the fact that only rates of the form 2^m can be accommodated, and that full octave increases in signal bandwidth are required to enable $q[n]$ to be demodulated at successively higher rates. For reference, the performance bound corresponding to UMSE modulation is superimposed on this plot as a dashed line. We emphasize that with UMSE modulation, the transmitter has perfect knowledge of the rate-bandwidth characteristics of the channel, while with fractal modulation the transmitter has no knowledge.

Clearly, this fractal modulation scheme is unrealizable for several reasons. First, the basis functions used at the transmitter again have infinite length. Hence, not only can such a transmitter not be implemented, but a corresponding receiver would involve infinite delay. However, we may more generally replace the ideal bandpass wavelet used to synthesize the orthonormal self-similar basis with one having comparable frequency domain characteristics but better time localization properties. As we have discussed, there are many suitable wavelets to choose from, among which are those due to Daubechies [1].

Another reason that fractal modulation is impractical is associated with the fact that $q[n]$ is modulated into an infinite number of octave-bandwidth channels. As a consequence, such a transmitter requires infinite power. However, in a practical implementation, only a finite collection of contiguous channels \mathcal{M} would be used by

²We emphasize that it is the baseband bandwidth that is important in defining the spectral efficiency in accordance with our channel model of Fig. 6-2, since it defines the highest frequency available at the receiver.

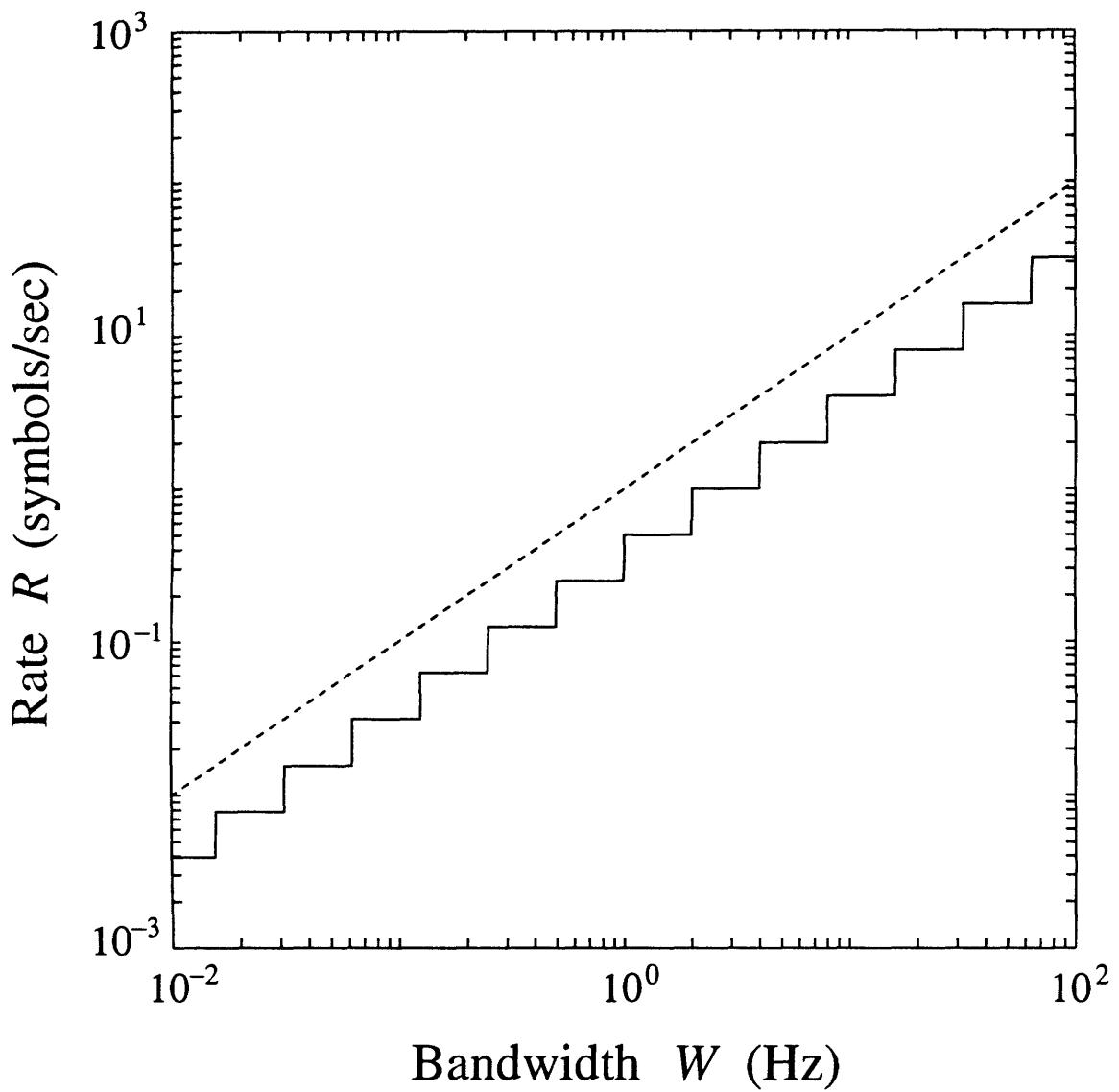


Figure 6-3: Spectral efficiency of fractal modulation. At each bandwidth B , the solid curve indicates the maximum rate at which transmitted data can be recovered. The dashed curve, indicating the corresponding rate for UMSE modulation, represents a performance bound.

the transmitter. The resulting modulated waveform

$$x(t) = \sum_n q[n] \sum_{m \in \mathcal{M}} \beta^{-m/2} \psi_n^m(t) \quad (6.4)$$

then exhibits self-similarity only over a range of scales, and the data can be demodulated at one of only a finite number of rates. In terms of Fig. 6-3, the rate-bandwidth characteristic would extend over a finite range of bandwidths which have been chosen to cover extremes anticipated for the system.

It is important to note that the fractal modulation transmitter can be implemented in a computationally highly efficient manner. Indeed, much of the processing can be performed using discrete-time algorithms of Section 5.3. Assuming the number of scales to be used is M , the signal $x(t)$ defined via (6.4) is synthesized from $q[n]$ as follows. First, the algorithm (5.35) is iterated M times to generate $p^{[M]}[n]$ using the QMF filter pair $h[n], g[n]$ appropriate to the wavelet basis. This modulates $q[n]$ into a sequence of M contiguous discrete-time channels. Then the resulting sequence $p^{[M]}[n]$ is modulated into the desired bandwidth via the appropriate scaling function according to

$$x(t) = \sum_n p^{[M]}[n] \phi_n^M(t).$$

Here we have assumed, for convenience, the collection of scales

$$\mathcal{M} = \{0, 1, \dots, M - 1\}$$

where M is some positive integer. It is important to point out that because a batch-iterative algorithm is employed, potentially large amounts of data buffering may be required. Hence, while the algorithm may be computationally efficient, it may be considerably less so in terms of storage requirements. However, in the event that $q[n]$ is *finite length*, it is conceivable that the algorithm may be modified so as to be memory-efficient as well.

In fact, the transmission of finite length sequences raises another issue: in transmitting finite length messages, a direct implementation of fractal modulation is rather

inefficient, principally because the successively higher channels are increasingly underutilized. In particular, as is apparent from the time-frequency portrait of Fig. 5-3, if $q[n]$ has finite length, *e.g.*,

$$q[n] = 0, \quad n < 0, \quad n > L - 1$$

then the m th channel will have completed its transmission and go idle in half the time it took the $(m - 1)$ st channel to complete its transmission, and so forth. However, finite length messages may be accommodated efficiently by modulating their periodic extensions

$$\tilde{q}[n] = q[n \bmod L],$$

thereby generating a transmitted waveform

$$x(t) = \sum_n \tilde{q}[n] \theta_n^H(t)$$

which constitutes a periodicity-dominated homogeneous signal of the type discussed in Section 5.2. If we let

$$\mathbf{q} = \{q[0] \ q[1] \ \cdots \ q[L - 1]\}$$

denote the data vector, then the time-frequency portrait associated with this signal is shown in Fig. 6-4. Note that not only do we retain the ability to make various rate-bandwidth tradeoffs at the receiver with this modification, but we acquire a certain flexibility in our choice of time origin as well. Specifically, as is apparent from Fig. 6-4, the receiver need not begin demodulating the data at $t = 0$, but may choose a time-origin that is some multiple of LR when operating at rate R . Additionally, this strategy can, in principle, be extended to accommodate blocking of the data.

Before considering the problem of optimum demodulation, we consider one final aspect of transmitter configuration. Having assumed an arbitrary choice for the parameter H , we now consider how this parameter might be appropriately chosen for a given operating scenario. From the point of view of spectral efficiency, it is apparent that this parameter has no effect. Indeed, the rate-bandwidth tradeoffs

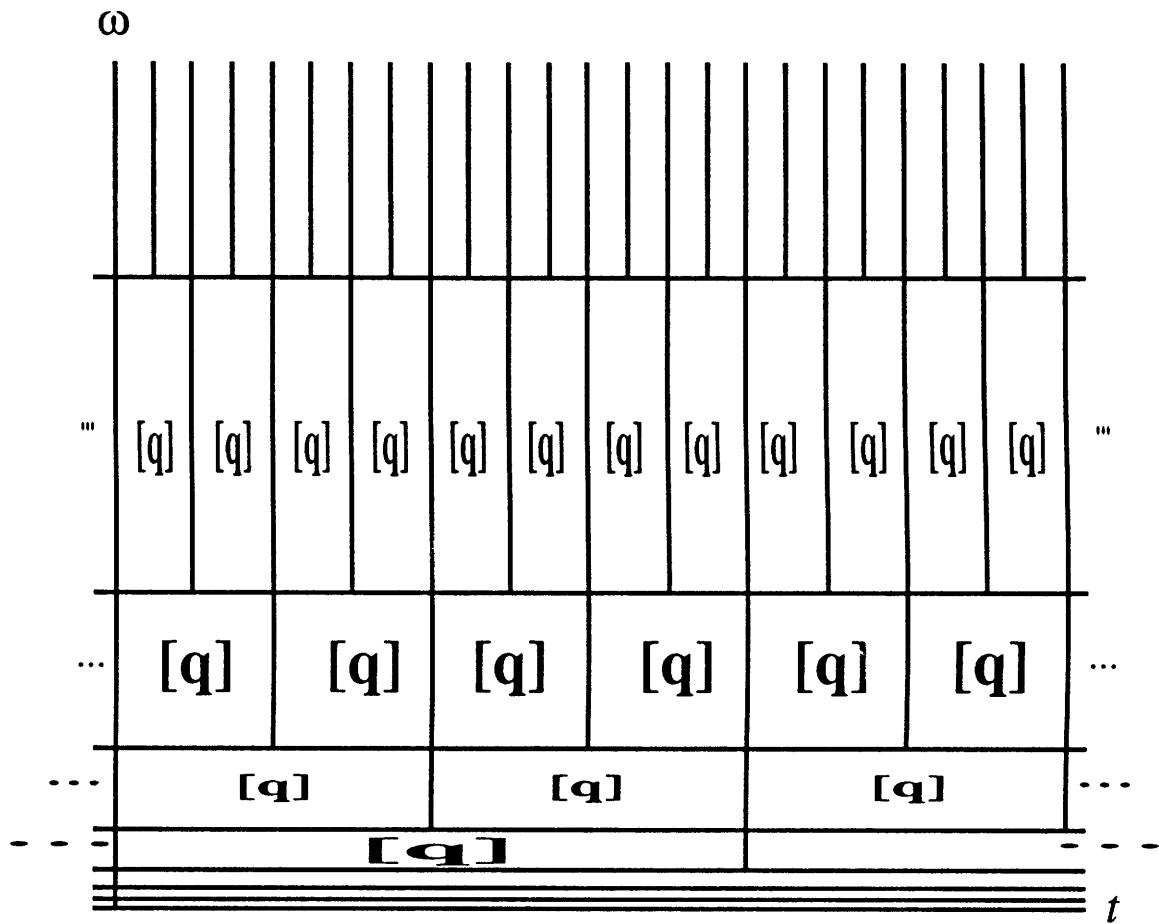


Figure 6-4: A portion of the time-frequency portrait of the transmitted signal for fractal modulation of a finite-length data vector \mathbf{q} . The case $H = -1/2$ is shown for convenience.

available to a receiver are independent of the value of H . However, H controls the relative distribution of power between channels. As such it affects the spectrum of the fractal modulation signal. As we noted in our discussion of homogeneous signals in Chapter 5, such power-dominated homogeneous signals have a power spectrum of the form

$$S_x(\omega) \sim \frac{1}{|\omega|^\gamma} \quad (6.5)$$

where $\gamma = 2H + 1$. Hence, the selection of H becomes important when we consider the presence of additive noise $z(t)$ in the channel of Fig. 6-2.

For more traditional additive stationary Gaussian noise channels in which, for example, the bandwidth is known *a priori*, the appropriate spectral shaping of the transmitted signal is governed by a “water-filling” procedure [14] [71]. In fact, this is the method by which the capacity of such channels is computed [74]. Through this procedure, the available signal power is distributed in such a way that proportionally more power is located at frequencies where the noise power is smaller. The graphical interpretation is one in which the available signal power is “poured” onto the noise spectrum within the available bandwidth, so that the spectral density of the signal at a given frequency is given by the distance from the noise floor to the “water level.”

However, when there is uncertainty in the available bandwidth, the water-filling approach is a less desirable strategy. Imagine a scenario in which the noise power is spectrally flat except in some frequency band $0 < \omega_L < \omega_U < \infty$, where it is zero. Then a water-filling procedure will locate the signal power predominantly within this frequency band. Consequently, when the channel bandwidth is such that it includes this band, the SNR in the channel will be much higher than if the bandwidth did not include this band. Hence, the system performance would strongly depend on the available bandwidth. A more reasonable power allocation strategy for the channel of Fig. 6-2, therefore, is to distribute power according to a spectral-matching rule, *i.e.*, so as to maintain an SNR that is independent of frequency. This leads to system performance that is uniform with variations in bandwidth, as would naturally arise out of a min-max type performance criterion. Note, too, that the spectral-matching

rule leads to a transmitted signal that is potentially well-suited for covert and LPI communication.

Since power-dominated homogeneous signals have a power spectrum of the form of eq. (6.5), the spectral-matching rule suggests that fractal modulation may be naturally matched to channels with additive $1/f$ -type noise. As discussed in Chapter 3, this rather broad class of statistically self-similar noises includes both classical white Gaussian noise and Brownian motion, as well as some more general and rather prevalent nonstationary processes characterized by strong long-term statistical dependence.

In this section, we have developed a strategy for embedding an information sequence $q[n]$ into a homogeneous signal $x(t)$ that satisfies the first of the two system requirements described at the outset of the chapter. Next, we turn our attention to the problem of designing receivers for fractal modulation, such that the second of our system requirements are satisfied.

6.2 Receiver Design: Demodulation

In this section, we consider the problem of recovering the information sequence $q[n]$ from the observations $r(t)$ available to the receiver. In general, $r(t)$ is assumed to be a band-limited, time-limited, and noisy version of $x(t)$ consistent with our channel model of Fig. 6-2, and where the noise $z(t)$ is Gaussian $1/f$ -type noise. We shall assume the degree H of the transmitted homogeneous signal has been chosen according to our spectral-matching rule, *i.e.*,

$$H = H_{\text{signal}} = H_{\text{noise}}. \quad (6.6)$$

We remark that if it is necessary that the transmitter measure H_{noise} in order to perform this spectral matching, the robust and efficient parameter estimation algorithms for $1/f$ processes developed in Section 4.2 may be exploited.

We shall also assume in our analysis that the ideal bandpass wavelet basis has been used to synthesize $x(t)$. While we recognize that realizable wavelet bases will be

used in practice, we anticipate that the results we obtain will be generally applicable to many of the approximately bandpass wavelet bases, such as those of Daubechies. Finally, we shall assume $q[n]$ to be of finite length L , and that its periodic extension $q[n \bmod L]$ is modulated as discussed in the previous section.

In our performance analyses exploring rate, bandwidth and SNR tradeoffs, we shall compare fractal modulation to the performance of UMSE modulation with repetition-coding. That is, we make comparisons to UMSE modulation where redundancy is provided by transmitting each sample of sequence $q[n]$ some number K times in succession. This simple redundancy scheme provides for error correction capability that is comparable to that of fractal modulation. However, with this coding, UMSE modulation performance does not constitute a performance bound. Indeed, it is not “uniformly most power efficient” in two respects. For one, using UMSE modulation corresponds to distributing signal power uniformly over the available bandwidth, which is inefficient except in the presence of stationary white noise ($H = -1/2$). This follows from the water-filling rule for power allocation. Secondly, even for the case of white noise, there are much more effective (*i.e.*, more power efficient) redundancy schemes available for use with channels of known bandwidth; see, *e.g.*, [75]. Nevertheless, with these caveats in mind, such comparisons do lend some insight into the *relative* power efficiency of fractal modulation.

6.2.1 Minimum Mean-Square Error Demodulation

In this section, we assume that $q[n]$ is a continuous-valued sequence of independent, identically-distributed, zero-mean Gaussian random variables, each with variance

$$\text{Var } q[n] = \sigma_x^2,$$

and develop a receiver yielding the minimum mean-square error (MSE) estimate of $q[n]$ based on our corrupted observations $r(t)$. We begin by projecting our observations onto the ideal bandpass wavelet basis from which $x(t)$ was synthesized, so that

our observations may be expressed as

$$r_n^m = \beta^{-m/2} q[n \bmod L] + z_n^m$$

where the z_n^m are the Gaussian wavelet coefficients of the noise process $z(t)$. Consistent with the wavelet-based models of $1/f$ processes we developed in Section 3.2.2, we may reasonably model the z_n^m as independent zero-mean random variables with variances

$$\text{Var } z_n^m = \sigma_z^2 \beta^{-m} \quad (6.7)$$

for some variance parameter $\sigma_z^2 > 0$.

If the channel is bandlimited to 2^{M_U} Hz for some integer M_U , this precludes access to the observation coefficients r_n^m at scales finer than $m = M_U$. Simultaneously, if we translate the time-limiting in the channel into a constraint on the minimum allowable rate of 2^{M_L} symbols/sec for some integer M_L , this precludes access to the observation coefficients at scales coarser than $m = M_L$. Hence, the observation coefficients r_n^m available at the receiver will correspond to the range of scales

$$M_L \leq m \leq M_U.$$

We have assumed $M_U \geq M_L$, which corresponds to the case in which we have at least enough time-bandwidth product to recover $q[n]$ in the absence of noise.

Consistent with both our rate constraint and the fact that we modulate the periodic extension of $q[n]$, we have available

$$K = 2^{M_U - M_L + 1} - 1 \quad (6.8)$$

measurements of each of the L non-zero samples of the sequence $q[n]$ from which to compute an optimal estimate. This can be inferred from Fig. 6-4. In particular, at scale m we have 2^{m-M_L} copies of $q[n]$, so that there is a doubling of the number of available copies at successive scales leading to the geometric sum (6.8). The specific

relationship between rate, bandwidth and K in terms of the spectral efficiency of fractal modulation can, therefore, be expressed as

$$\frac{R}{W} = \frac{2\eta_F}{K+1}, \quad (6.9)$$

where we note that the case $K = 1$, for which $M_U = M_L$ and a single copy of $q[n]$ is available at the receiver, corresponds to the minimum time-bandwidth product for which reconstruction is possible. In the case of the ideal bandpass wavelet, we established earlier that $\eta_F = 1/2$. Using more general wavelets, the spectral bandwidth and hence efficiency of fractal modulation is less well-defined. Nevertheless, for any reasonable definition of bandwidth, practical wavelets may be chosen for use with fractal modulation that yield a spectral efficiency close to $1/2$. We shall therefore assume in general that $\eta_F \approx 1/2$ in our subsequent discussion.

It is a straightforward to derive minimum MSE estimates of $q[n]$ for each n from the collection of independent, noisy observations

$$\mathbf{r} = \{r_n^m, m \in \mathcal{M}, n \in \mathcal{N}(m)\}$$

where

$$\mathcal{M} = \{M_L, M_L + 1, \dots, M_U\} \quad (6.10)$$

$$\mathcal{N}(m) = \{0, 1, \dots, L 2^{m-M_L} - 1\}. \quad (6.11)$$

The optimum estimator is given by

$$\hat{q}[n] = E[q[n]|\mathbf{r}] = \frac{1}{K + (1/\sigma_c^2)} \sum_{m=M_L}^{M_U} \beta^{m/2} \sum_{l=0}^{2^m - M_L - 1} r_{n+lL}^m \quad (6.12)$$

where σ_c^2 is the SNR in the channel, defined by

$$\sigma_c^2 = \frac{\sigma_x^2}{\sigma_z^2}.$$

In general the optimum estimate represents a blend of *a priori* information about $q[n]$, and information obtained from the observations. At high SNR ($\sigma_c^2 \gg 1/K$), the *a priori* information is essentially ignored, and the resulting estimator specializes to the maximum likelihood estimator. At low SNR ($\sigma_c^2 \ll 1/K$), the observations are essentially ignored, and the estimator approaches the *a priori* estimate

$$\hat{q}[n] = E[q[n]] = 0.$$

We note that, as in the case of the transmitter, the receiver has a convenient, computationally-efficient, hierarchical structure based on the DWT. Assuming $r(t)$ is bandlimited to resolution 2^{M_U} , it may be sampled at rate 2^{M_U} , then successively filtered and downsampled to level $m = M_L$ according to the wavelet decomposition tree of Fig. 2-6(a). To produce the desired estimate, at each level m , the terms from the detail sequence r_n^m corresponding to the same value of the $q[n]$ are collected together, weighted by

$$\frac{\beta^{m/2}}{K + 1/\sigma_c^2}, \quad (6.13)$$

and accumulated with the weighted r_n^m from previous stages. Again, however, this is a batch algorithm, and while computationally efficient, may not be efficient in terms of storage requirements.

Finally, we remark that the receiver (6.12) we have designed is a linear data processor, as would be anticipated since we have restricted the discussion to Gaussian sequences and Gaussian noise. In non-Gaussian scenarios, the receivers we have developed are the best *linear* data processors, *i.e.*, no other linear data processor is capable of generating an estimate of $q[n]$ with a smaller mean-square error.

Performance

The normalized MSE associated with the optimum receiver (6.12) can be readily derived as

$$\epsilon^2 = \frac{E[(q[n] - \hat{q}[n])^2]}{E[(q[n])^2]} = \frac{E[\text{Var}(q[n]|\mathbf{r})]}{\text{Var} q[n]} = \frac{1}{1 + K\sigma_c^2}. \quad (6.14)$$

Generally, it is convenient to substitute for K in (6.14) via (6.9) to get

$$\epsilon^2 = \frac{1}{1 + \sigma_c^2 \left[\frac{2\eta_F}{R/W} - 1 \right]} \quad (6.15)$$

where $\eta_F \approx 1/2$, and where $R/W \leq \eta_F$ by virtue of our definition of η_F . From (6.15) we see, then, that for $R/W \ll \eta_F$, the MSE is given asymptotically by

$$\epsilon^2 \sim \frac{1}{\sigma_c^2 \frac{2\eta_F}{R/W}}. \quad (6.16)$$

Note that fractal modulation performance (6.15) is independent of the parameter H when we use spectral matching.

For comparison, let us consider the MSE performance of UMSE modulation with repetition-coding in the presence of stationary white Gaussian noise. In this case, incorporating redundancy reduces the effective rate-bandwidth ratio by a factor of K , *i.e.*,

$$\frac{R}{W} = \frac{\eta_0}{K}, \quad (6.17)$$

where R is the rate at which the symbols $q[n]$ are transmitted, and where η_0 is the efficiency of the modulation without coding, which is unity. The optimum Bayesian receiver for this scheme, using a minimum MSE criterion, demodulates the repeated sequence, and averages the terms corresponding to the same value of $q[n]$ to generate $\hat{q}[n]$. Hence this K -fold redundancy leads to a normalized MSE of

$$\epsilon^2 = \frac{E[(q[n] - \hat{q}[n])^2]}{E[(q[n])^2]} = \frac{1}{1 + \sigma_c^2 K} \quad (6.18)$$

where σ_c^2 is the SNR, *i.e.*, the ratio of the power in $q[n]$ to the density of the white noise power spectrum. Combining (6.18) with (6.17) we get

$$\epsilon^2 = \frac{1}{1 + \sigma_c^2 \frac{\eta_0}{R/W}} \quad (6.19)$$

whenever $R/W \leq \eta_0$. By comparison with (6.16) we see that when $R/W \ll \eta_0$,

$$\epsilon^2 \sim \frac{1}{\sigma_c^2 \frac{\eta_0}{R/W}}, \quad (6.20)$$

which is essentially (6.16) since the achievable η_0 is unity and $\eta_F \approx 1/2$ as discussed earlier. This means that, at least asymptotically, the performance between the two schemes is comparable in the presence of white noise.

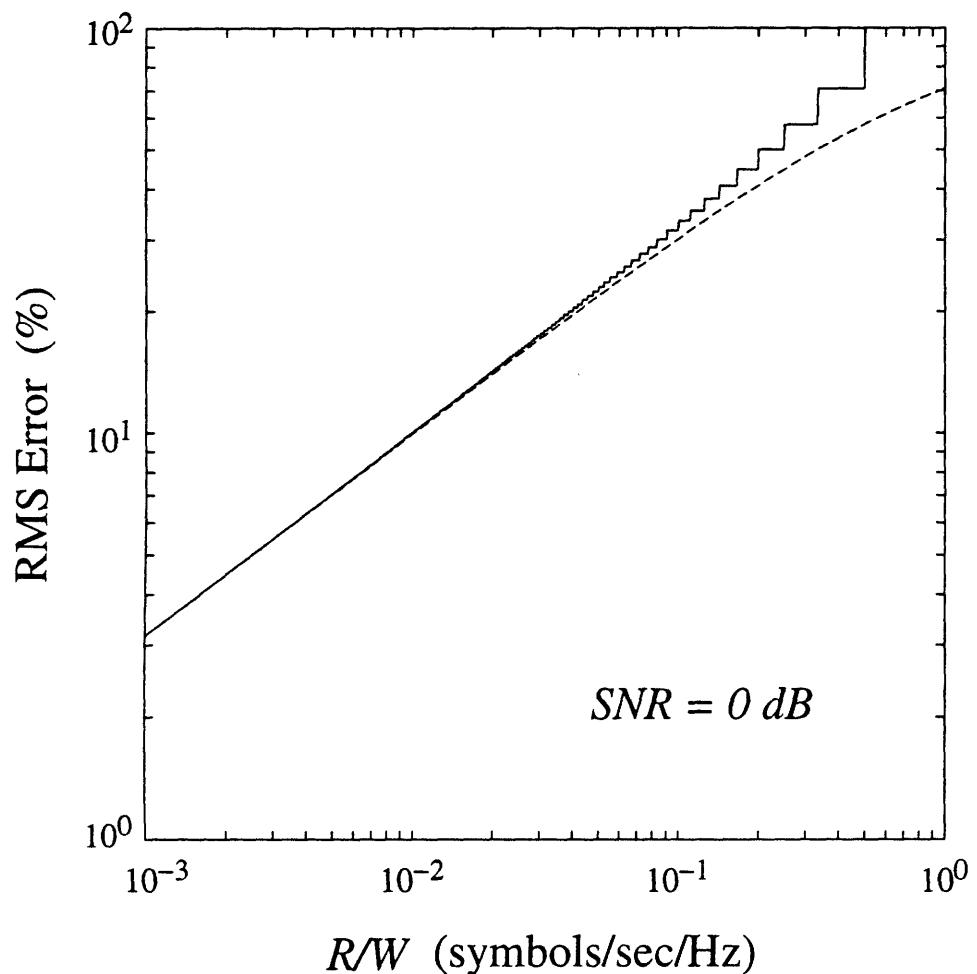
This behavior is reflected in the performance curves for both fractal modulation and UMSE modulation with repetition coding of Fig. 6-5. In Fig. 6-5(a), MSE is shown as a function of R/W at a fixed SNR of 0 dB ($\sigma_c^2 = 1$), while in Fig. 6-5(b), MSE is shown as a function of SNR at a fixed $R/W = 0.1$ symbols/sec/Hz. As we expect, the longer the channel is open, or the greater the available bandwidth in the channel, the better the performance of fractal modulation. Although comparisons between the two modulation schemes is appropriate only for the special case of additive white Gaussian noise channels, we reiterate that the performance of fractal modulation (6.15) is independent of the spectral exponent of the $1/f$ noise. By contrast, we would not, in general, expect (6.19) to describe the performance of UMSE modulation with repetition-coding in the presence $1/f$ noise.

6.2.2 Minimum Probability-of-Error Demodulation

In this section, we address the closely related problem of designing and evaluating optimal receivers for bit-by-bit signaling using fractal modulation. Specifically, let us consider the transmission of a random binary data stream via the bi-valued sequence³ $q[n]$ with average energy E_0 per bit:

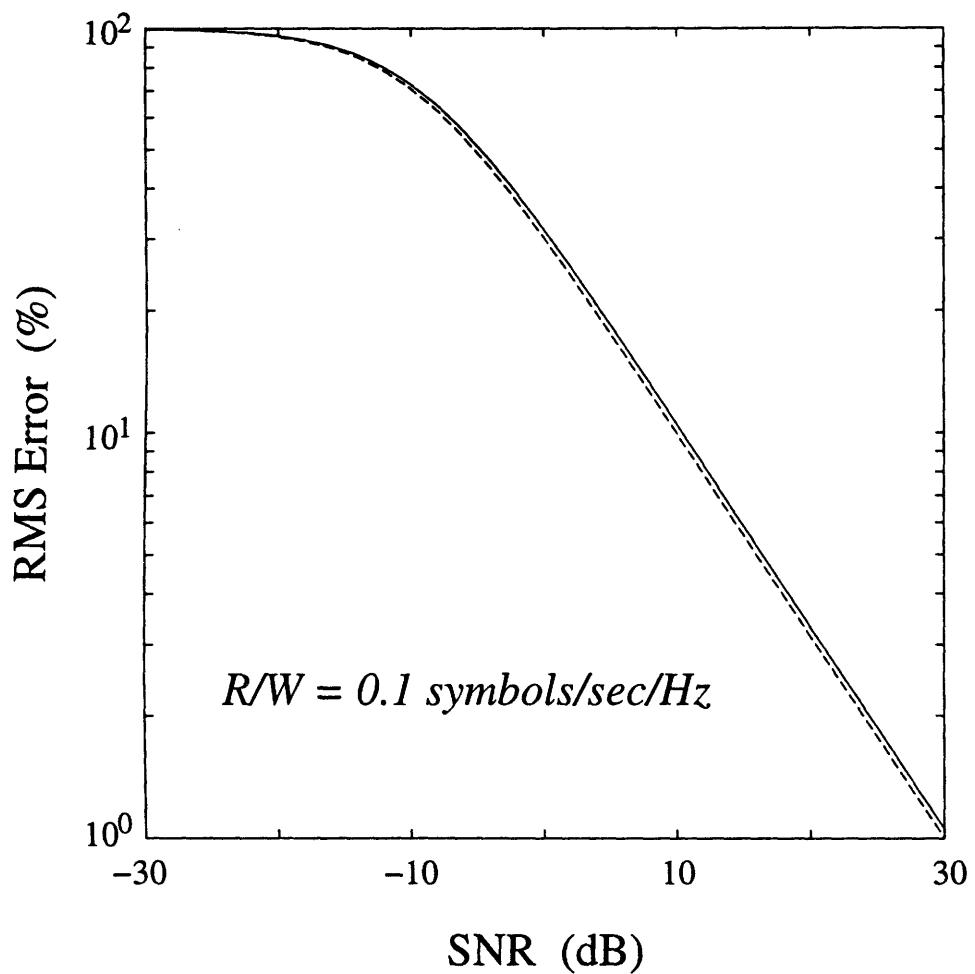
$$q[n] \in \left\{ +\sqrt{E_0}, -\sqrt{E_0} \right\}, \quad (6.21)$$

³Although we do not explicitly consider it here, the extension to the case of L -ary signaling is likewise straightforward.



(a) MSE ϵ^2 as a function of Rate/Bandwidth ratio R/W at 0 dB SNR.

Figure 6-5: *Error-rate-bandwidth tradeoffs for fractal modulation in noise with the optimum receiver. The solid lines represent the performance of fractal modulation, while the dashed lines corresponds to the performance of UMSE modulation with repetition coding.*



(b) MSE ϵ^2 as a function of SNR at $R/W = 0.1$ symbols/sec/Hz.

Figure 6-5: *Continued.*

representing a 1-bit and a 0-bit, respectively. Moreover, as before, we consider a channel characterized by additive Gaussian $1/f$ noise.

The problem of optimally decoding each bit can be described in terms of a binary hypothesis test on the observations $r(t)$, or equivalently, the coefficients r_n^m of these observations decomposed into the wavelet basis used in the synthesis of $x(t)$. Choosing the latter of the options, our hypotheses are

$$H_0 : r_n^m = -\sqrt{E_0}\beta^{-m/2} + z_n^m \quad (6.22a)$$

$$H_1 : r_n^m = +\sqrt{E_0}\beta^{-m/2} + z_n^m \quad (6.22b)$$

where, again, z_n^m are the wavelet coefficients of the noise process. Again according to our wavelet-based models for $1/f$ noise of Section 3.2.2, under each hypothesis we may model the z_n^m as independent, zero-mean Gaussian random variables with variances given by (6.7). Likewise, based on rate and bandwidth limitations in the channel, the available observation coefficients are

$$\mathbf{r} = \{r_n^m, m \in \mathcal{M}, n \in \mathcal{N}(m)\}$$

where the ranges \mathcal{M} and $\mathcal{N}(m)$ are given by (6.11). We shall also assume in our analysis that the ideal bandpass wavelet basis was used in the synthesis of $x(t)$, with the implication that the analysis applies more generally to a broader class of wavelets.

The likelihood ratio

$$\frac{p_{\mathbf{r}|H_1}(\mathbf{r})}{p_{\mathbf{r}|H_0}(\mathbf{r})}$$

associated with this hypothesis testing problem is readily derived, and can be reduced to the sufficient statistic

$$\ell = \sum_{m=M_L}^{M_U} \sum_{l=0}^{2^m-M_L-1} \frac{r_{n+lK}^m \cdot \sqrt{E_0}\beta^{-m/2}}{\sigma_z^2 \beta^{-m}}. \quad (6.23)$$

For a minimum probability of error $\Pr(\varepsilon)$ receiver and a random bit stream (*i.e.*,

equally likely hypotheses), the optimum test can be derived as

$$\begin{array}{c} H_1 \\ \ell \gtrsim 0, \\ H_0 \end{array}$$

which is intuitively reasonable from symmetry considerations. Indeed, ℓ is conditionally Gaussian under each hypothesis, where the conditional means are

$$E[\ell|H_1] = -E[\ell|H_0] = K\sigma_c^2 \quad (6.24)$$

and the conditional variances are

$$\text{Var}[\ell|H_1] = \text{Var}[\ell|H_0] = K\sigma_c^2, \quad (6.25)$$

where σ_c^2 is the SNR, *viz.*,

$$\sigma_c^2 = \frac{E_0}{\sigma_z^2}.$$

Performance

The bit-error probability associated with this optimal receiver can be readily derived as

$$\Pr(\varepsilon) = \Pr(\ell > 0|H_0) = Q\left(\frac{1}{2}\sqrt{K\sigma_c^2}\right) \quad (6.26)$$

where $Q(\cdot)$ is defined, as in Section 4.4, by

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-v^2/2} dv,$$

and where we have exploited (6.24) and (6.25). Substituting for K in (6.26) via (6.9) we can express this error probability in terms of the rate-bandwidth ratio as

$$\Pr(\varepsilon) = Q\left(\frac{1}{2}\sqrt{\sigma_c^2 \left[\frac{2\eta_F}{R/W} - 1\right]}\right), \quad (6.27)$$

where $\eta_F \approx 1/2$. Again, the performance of fractal modulation is independent of the spectral exponent of the noise process when we use spectral matching. Note, too, that for $R/W \ll \eta_F$, the bit error probability is asymptotically

$$\Pr(\varepsilon) = Q\left(\frac{1}{2}\sqrt{\sigma_c^2 \left[\frac{2\eta_F}{R/W}\right]}\right). \quad (6.28)$$

For UMSE modulation with repetition coding in the presence of stationary white Gaussian noise, to decode a bit, the optimal receiver naturally demodulates the received data and averages together the K symbols associated with the transmitted bit, thereby generating a sufficient statistic. When this statistic is positive, the receiver decodes a 1-bit, and a 0-bit otherwise. The corresponding performance is, therefore, given by

$$\Pr(\varepsilon) = Q\left(\frac{1}{2}\sqrt{\sigma_c^2 K}\right).$$

Substituting for K via (6.17) yields the error expression

$$\Pr(\varepsilon) = Q\left(\frac{1}{2}\sqrt{\sigma_c^2 \left[\frac{\eta_0}{R/W}\right]}\right) \quad (6.29)$$

in terms of the rate-bandwidth ratio, where $\eta_0 = 1$. Comparing (6.29) with (6.28), we note that again since $\eta_0 \approx 2\eta_F$, the asymptotic performance of the two schemes is effectively equivalent.

This is apparent in Fig. 6-6, where we plot the bit-error performance of fractal modulation in bit-by-bit signaling along side corresponding performance of UMSE modulation with repetition coding. In Fig. 6-6(a), $\Pr(\varepsilon)$ is shown as a function of R/W at a fixed SNR of 0 dB ($\sigma_c^2 = 1$), while in Fig. 6-5(b), $\Pr(\varepsilon)$ is shown as a function of SNR at a fixed $R/W = 0.1$ bits/sec/Hz. Both these plots reveal strong thresholding behavior whereby the error probability falls off dramatically at high SNR and low R/W . Although the comparison between the two modulations is generally appropriate only for the special case of additive white Gaussian noise channels, we reiterate that the performance of fractal modulation (6.27) is independent of the

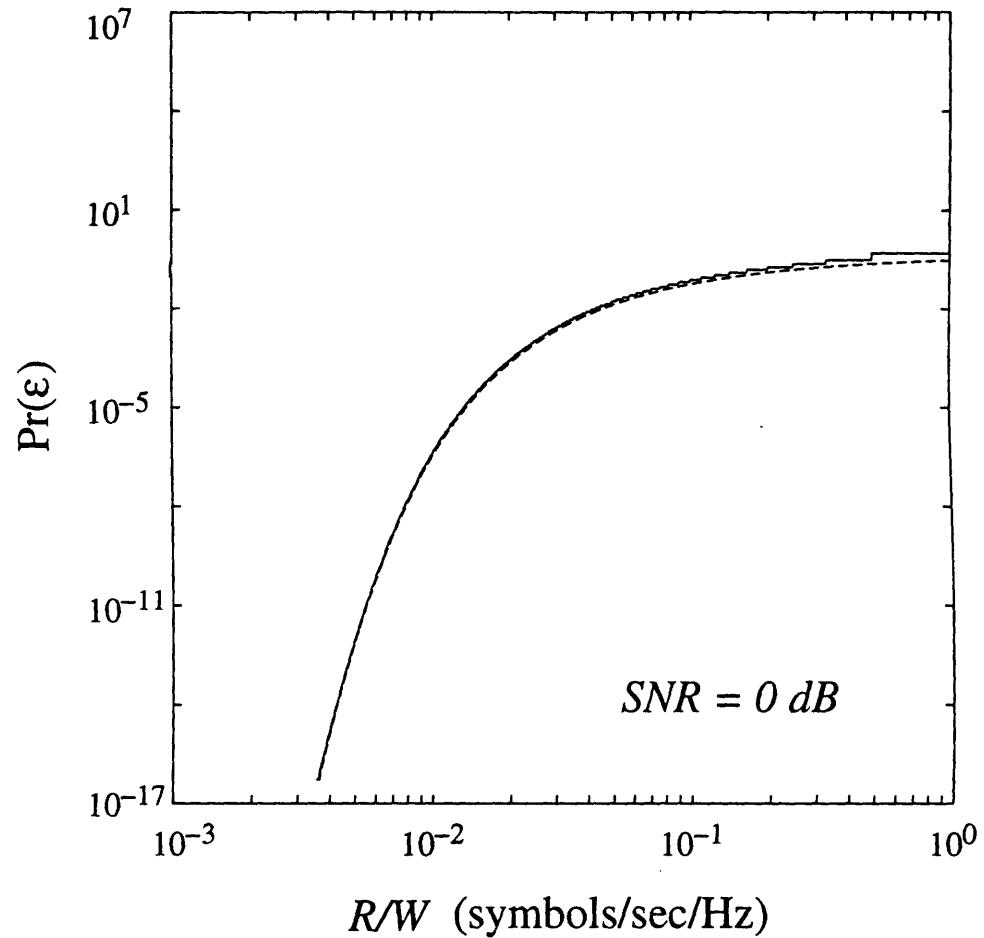
spectral exponent of the $1/f$ noise. By contrast, we would not expect (6.29) to describe the performance of UMSE modulation with repetition-coding in the presence $1/f$ noise in general.

6.3 Outstanding Issues

This chapter constitutes a highly preliminary development of fractal modulation. There are several aspects of the modulation that require further study and evaluation. For example, of significant practical interest is the robustness of fractal modulation with respect to channel modeling errors and numerical implementation errors. A comprehensive evaluation of sensitivity to such errors would involve both Monte Carlo simulations with synthetic data, and experimental implementations with real channels.

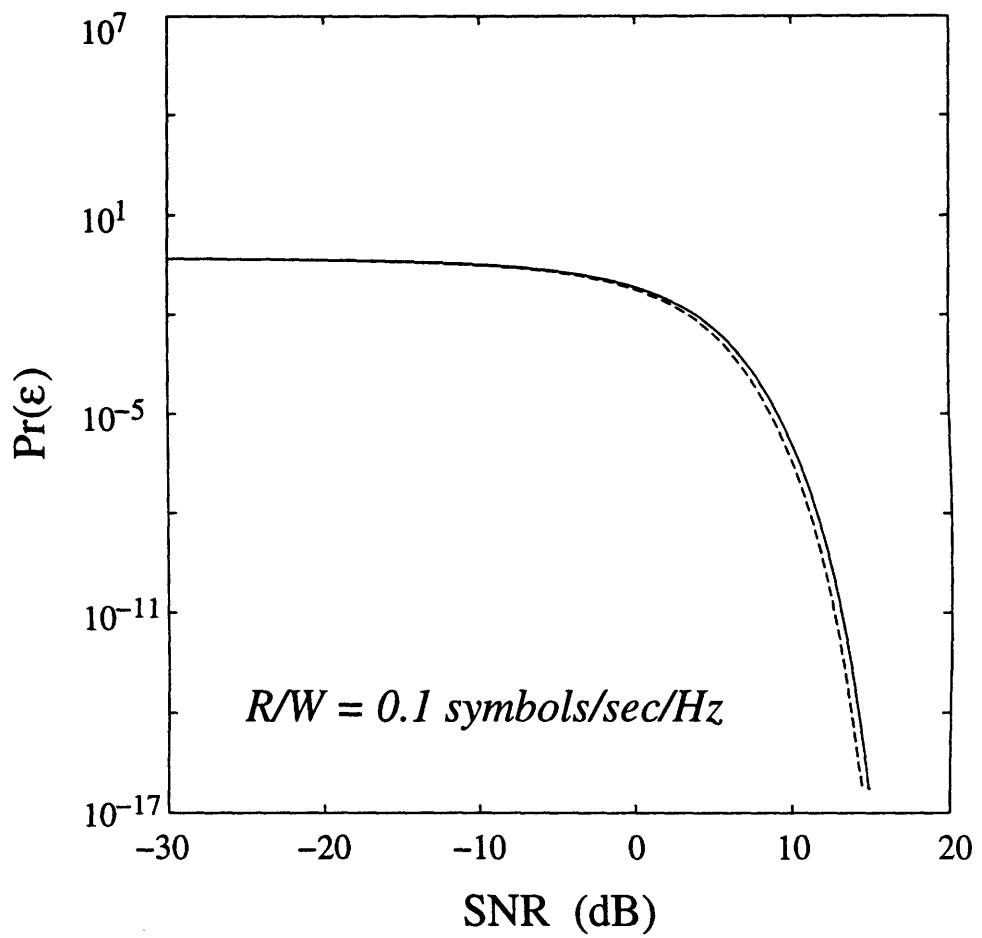
It is also important to consider whether fractal modulation constitutes an optimal modulation strategy in some sense. Specifically, while the scheme is apparently well-suited to channels of the type illustrated in Fig. 6-2 in which the bandwidth and duration are unavailable to the transmitter, we have not established that it is optimal with respect to any particular performance criterion. This question certainly deserves attention, as does the associated problem of computing the capacity of channels of this type.

In addition, there are a number of issues pertaining to fractal modulation that need to be addressed before it can be applied in various contexts. For example, accurate synchronization between transmitter and receiver would appear to be critical in the scheme, perhaps more so than in other schemes. Consequently, we would anticipate that the development of techniques for achieving the necessary synchronization would be important for any practical implementation. Another issue pertains to the use of fractal modulation in LPI applications. While we have argued that the second-order statistics of homogeneous signals can be effectively indistinguishable from those of $1/f$ noises, a more comprehensive study of the detectability of homogeneous signals is warranted. In particular, it would be important to study the vulnerability of



(a) Probability of error $\text{Pr}(\varepsilon)$ as a function of Rate/Bandwidth ratio R/W at 0 dB SNR.

Figure 6-6: Bit error probabilities for bit-by-bit signaling with fractal modulation over noisy channels with the optimum receiver. Solid lines indicate the performance of fractal modulation, while dashed lines indicate the performance of UMSE modulation with repetition coding.



(b) Probability of error $\text{Pr}(\epsilon)$ as a function of SNR at $R/W = 0.1 \text{ symbols/sec/Hz}$.

Figure 6-6: *Continued.*

fractal modulation to traditional nonlinear feature detection techniques including square-and-correlate processing [76].

This leads naturally to a consideration of some potentially important extensions to the basic fractal modulation strategy. For example, in many spread spectrum scenarios, it is important to obscure the spreading technique. In direct-sequence spread spectrum, not only is each symbol in the information sequence repeated K times prior to transmission, but the redundant stream is premultiplied by a pseudorandom bit stream known to both transmitter and receiver. This makes the transmitted sequence appear like white (or at least broadband) noise to a listener without detailed knowledge of the pseudorandom sequence. For fractal modulation, the analogous processing might involve premultiplying the entire wavelet coefficient field x_n^m of the transmitted signal $x(t)$ by a pseudorandom bit field known to both transmitter and receiver. In turn, this would appear to have the effect of making $x(t)$ appear like $1/f$ -type noise to a listener lacking detailed knowledge of the pseudorandom bit field. Such an extension may well warrant investigation.

Another extension worth considering concerns the incorporation of efficient coding techniques for use with fractal modulation. As developed, the redundancy in the transmitted signal takes the form of multirate repetition in the time-frequency plane. Perhaps block or trellis coding techniques [75] can be exploited in improving the power efficiency of fractal modulation. At first glance, it would seem that coding of this type cannot be incorporated without sacrificing properties of the transmission scheme. Nevertheless, it would be important to clearly establish the tradeoffs involved.

Finally, fractal modulation would be of potential interest in a greater range of applications were it possible to accommodate a finer lattice of rate-bandwidth tradeoffs than those prescribed by the dyadic grid we have considered. Indeed, from a communications perspective, it is rare to have channels whose bandwidth spans an octave, much less multiple octaves. However, it is reasonable to expect that should practical families of non-dyadic orthonormal wavelet bases of the type described in Section 2.2.7 emerge, the concept of fractal modulation could be accordingly extended.

In summary, while fractal modulation is not yet sufficiently developed that it

constitutes a viable communication system in its own right, it does represent an interesting, novel and potentially important paradigm for communication in many contexts.

Chapter 7

Linear Self-Similar Systems

This thesis has explored some important classes of statistically and deterministically self-similar signals. This chapter represents a preliminary investigation into the relationships between self-similar signals and self-similar systems. In particular, we explore not only how we may interpret some of our methods for synthesizing self-similar signals in the context of driven self-similar systems, but also the role that the wavelet transform plays in characterizing such systems. In the end, this leads to some interesting and potentially important insights and perspectives into the results of the thesis, and suggests some promising future directions for work in this area.

The self-similar systems we discuss in this chapter have the property that they are linear and jointly time- and scale-invariant. In the first half of the chapter we define this class of systems, develop several properties, and show how both the Laplace and Mellin transform can be used in studying these systems. In the latter half of the chapter we develop wavelet-based characterizations of this class of systems to illustrate that the wavelet transform is in some sense best matched to these systems—that such characterizations are as natural and as useful for these systems as Fourier-based characterizations are for linear time-invariant systems. We first briefly review some results in the theory of linear time-invariant systems. For a comprehensive treatment, see, *e.g.*, [77] or [78].

7.1 Linear Time-Invariant Systems

Suppose $y(t)$, $y_1(t)$, and $y_2(t)$ are the responses of a system $\mathcal{S}\{\cdot\}$ to arbitrary inputs $x(t)$, $x_1(t)$, and $x_2(t)$, respectively. Then the system is linear when it satisfies the superposition principle

$$\mathcal{S}\{ax_1(t) + bx_2(t)\} = ay_1(t) + by_2(t) \quad (7.1)$$

for any a and b , and time-invariant when it satisfies

$$\mathcal{S}\{x(t - \tau)\} = y(t - \tau) \quad (7.2)$$

for any constant τ . Collectively the properties (7.1) and (7.2) characterize a linear time-invariant (LTI) system.

A linear system is time-invariant if and only if its kernel

$$\kappa(t, \tau) = \mathcal{S}\{\delta(t - \tau)\}$$

satisfies

$$\kappa(t, \tau) = \kappa(t - b, \tau - b) \quad (7.3)$$

for any b . For this class of systems, the kernel has the form

$$\kappa(t, \tau) = v(t - \tau)$$

where $v(t)$ is the impulse response of the system. Furthermore, the corresponding input-output relation is, of course, the usual convolution,

$$y(t) = \int_0^\infty x(\tau) v(t - \tau) d\tau \triangleq x(t) * \xi(t).$$

The eigenfunctions of LTI systems are complex exponentials of the form e^{st} , from which we get that the Laplace transform possesses a convolution property, *i.e.*, for

signals $x(t)$ and $y(t)$ with Laplace transforms $X(s)$ and $Y(s)$, respectively, we have

$$x(t) * y(t) \longleftrightarrow X(s)Y(s).$$

7.2 Linear Scale-Invariant Systems

In contrast to linear time-invariant systems, linear *scale*-invariant system theory has been comparatively less explored, though it has received occasional attention in the systems [79] and pattern recognition [80] literature, and in the broader mathematics literature in connection with the Mellin transform [81] [82] [83].

Suppose $y(t)$ is the response of a system $\mathcal{S}\{\cdot\}$ to an arbitrary input $x(t)$. Then a system $\mathcal{S}\{\cdot\}$ is said to be scale-invariant whenever

$$\mathcal{S}\{x(t/\tau)\} = y(t/\tau) \quad (7.4)$$

for any constant $\tau > 0$. When there is risk of ambiguity, we will refer to such systems as *strict-sense* scale-invariant systems to distinguish them from generalized scale-invariant systems we will develop subsequently. A system satisfying both (7.1) and (7.4) will be referred to as a *linear scale-invariant* (LSI) system.

It is straightforward to show that a necessary and sufficient condition for the kernel $\kappa(t, \tau)$ of a linear system to correspond to a scale-invariant system is that it satisfy

$$\kappa(t, \tau) = a\kappa(at, a\tau) \quad (7.5)$$

for any $a \geq 0$.

A linear scale-invariant system is generally characterized in terms of the *lagged-impulse response* pair

$$\begin{aligned} \xi_+(t) &= \mathcal{S}\{\delta(t - 1)\} \\ \xi_-(t) &= \mathcal{S}\{\delta(t + 1)\}. \end{aligned}$$

Indeed, when an input $x(t)$ can be decomposed, except at $t = 0$, as

$$\begin{aligned} x(t) &= \int_{-\infty}^{\infty} x(\tau) \delta\left(\frac{t}{\tau} - 1\right) \frac{d\tau}{\tau} \\ &= \int_0^{\infty} x(\tau) \delta\left(\frac{t}{\tau} - 1\right) \frac{d\tau}{\tau} + \int_0^{\infty} x(-\tau) \delta\left(\frac{t}{\tau} + 1\right) \frac{d\tau}{\tau} \end{aligned} \quad (7.6)$$

we can exploit the superposition principle (7.1) together with (7.4) to get the following input-output relation

$$y(t) = \int_0^{\infty} x(\tau) \xi_+ \left(\frac{t}{\tau} \right) \frac{d\tau}{\tau} + \int_0^{\infty} x(-\tau) \xi_- \left(\frac{t}{\tau} \right) \frac{d\tau}{\tau}. \quad (7.7)$$

For simplicity of exposition, we will restrict our subsequent discussion to the case of causal inputs

$$x(t) = 0, \quad t \leq 0$$

and LSI systems whose outputs are causal

$$y(t) = \mathcal{S}\{x(t)\} = 0, \quad t \leq 0.$$

From the development, it will be apparent how to accommodate the more general scenario of (7.7). For causal signals, the input-output relation (7.7) simplifies to

$$y(t) = \int_0^{\infty} x(\tau) \xi \left(\frac{t}{\tau} \right) \frac{d\tau}{\tau} \triangleq x(t) \star \xi(t) \quad (7.8)$$

where we let $\xi(t) = \xi_+(t)$ to simplify our notation, and where we use the symbol \star to distinguish this convolutional relationship from the usual convolution $*$ associated with LTI systems. Note that for these LSI systems the kernel is

$$\kappa(t, \tau) = \frac{1}{\tau} \xi \left(\frac{t}{\tau} \right).$$

This new convolution operation possesses many of the properties of the usual convolution operation. For example, it is straightforward to show that it is commutative

for well-behaved operands, *i.e.*,

$$x(t) \star \xi(t) = \xi(t) \star x(t) = \int_0^\infty x\left(\frac{t}{\tau}\right) \xi(\tau) \frac{d\tau}{\tau}. \quad (7.9)$$

As a consequence, the cascade of two LSI systems with lagged-impulse responses $\xi_1(t)$ and $\xi_2(t)$, respectively, is equivalent to a single system with lagged-impulse response $\xi_1(t) \star \xi_2(t)$. Furthermore, the systems may be cascaded in either order without changing the overall system.

Likewise, it is straightforward to show that the new convolution operation is distributive for well-behaved operands, *i.e.*,

$$x(t) \star \{\xi_1(t) + \xi_2(t)\} = x(t) \star \xi_1(t) + x(t) \star \xi_2(t). \quad (7.10)$$

Hence, the parallel connection of two LSI systems with lagged-impulse responses $\xi_1(t)$ and $\xi_2(t)$, respectively, is equivalent to a single system with lagged-impulse response $\xi_1(t) + \xi_2(t)$.

The eigenfunctions of linear scale-invariant systems are homogeneous functions of degree s ; specifically they are the complex power functions defined by

$$x(t) = t^s, \quad (7.11)$$

where s is a complex number. Indeed, from (7.8) and (7.9) the response of an LSI system to (7.11) is readily obtained as

$$y(t) = \check{\Xi}(s) t^s$$

with the associated complex eigenvalue given by

$$\check{\Xi}(s) = \int_{-\infty}^\infty \xi(\tau) \tau^{-s-1} d\tau \quad (7.12)$$

whenever this integral converges. Eq. (7.12) is referred to as the Mellin transform of the signal $\xi(t)$.

The eigenfunction property of the complex power functions implies that the Mellin transform constitutes an important tool in the analysis of LSI systems¹. Indeed, it is particularly convenient to compute the response of an LSI system to any input that is the superposition of eigenfunctions. Fortunately, a broad class of signals $x(t)$ can be expressed as a superposition of eigenfunctions of LSI systems according to

$$x(t) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \check{X}(s) t^s ds \quad (7.13a)$$

for $t > 0$, where

$$\check{X}(s) = \int_0^\infty x(\tau) \tau^{-s-1} d\tau \quad (7.13b)$$

and c is in the region of convergence of $\check{\Xi}(s)$. Eqns. (7.13) collectively constitute the Mellin representation of a signal $x(t)$: the Mellin inverse formula (7.13a) is the synthesis relation, while the Mellin transform (7.13b) is the analysis formula. Interestingly, we may interpret the Mellin transformation as a representation of $x(t)$ by its “fractional” moments.

The Mellin inverse formula implies that a broad class of linear scale-invariant systems are completely characterized by the Mellin transform of their lagged-impulse response, $\check{\Xi}(s)$. Consequently, we can refer to this quantity as the *system function* associated with the LSI system. As a consequence of the eigenfunction property of the complex power functions, the input-output relation for a linear scale-invariant system with system function $\check{\Xi}(s)$ can be expressed in the Mellin domain as

$$\check{Y}(s) = \check{\Xi}(s) \check{X}(s) \quad (7.14)$$

whenever both terms on the right-hand side have a common region of convergence. Hence, via the Mellin transform, we can map our convolution operation (7.8) into a convenient multiplicative operation (7.14).

The Mellin transform, its inversion formula, properties, and numerous transform

¹Actually, we have chosen a slight but inconsequential variant of the Mellin transform—the usual Mellin transform has s replaced by $-s$ in our definition.

pairs are well-documented in the literature. See, *e.g.*, [83] [84] [85]. One basic Mellin transform pair is given by

$$t^{-s_0} u(t-1) \longleftrightarrow \frac{1}{s+s_0}, \quad Re(s) > -s_0 \quad (7.15)$$

for arbitrary s_0 .

From this pair we are able to show that the Mellin transform plays an important role in the solution of a class of scale-differential equations that give rise to linear scale-invariant systems. We begin by quantifying the notion of a “derivative operator in scale.” A reasonable definition of the derivative in scale of a signal $x(t)$ is given by

$$\nabla_s x(t) \triangleq \lim_{\varepsilon \rightarrow 1} \frac{x(\varepsilon t) - x(t)}{\ln \varepsilon}.$$

One can readily interpret this definition in the context of traditional derivatives as

$$\nabla_s x(t) = \frac{d}{d \ln t} x(t) = t \frac{d}{dt} x(t).$$

Differentiation in scale corresponds to a multiplication by s in the Mellin domain, which suggests that the Mellin transform can be used to efficiently solve what can be described as a class of “dynamical systems in scale.”² Consider the following N th-order linear constant-coefficient scale-differential equation

$$\sum_{k=0}^N a_k \nabla_s^k y(t) = \sum_{k=0}^M b_k \nabla_s^k x(t),$$

where we denote the k th derivative in scale, obtained by iterative application of the derivative operator, by ∇_s^k . Then, via the convolution property of the Mellin transform, we obtain

$$\check{Y}(s) = \check{\Xi}(s) \check{X}(s)$$

²We remark that this development raises some interesting questions regarding connections to the more general literature that is evolving on multiscale systems [86] [58]. Such relationships, however, have yet to be explored.

where $\check{\Xi}(s)$ is rational, i.e.,

$$\check{\Xi}(s) = \frac{\prod_{k=0}^M b_k s^k}{\prod_{k=0}^N a_k s^k},$$

in the corresponding region of convergence. The usual partial fraction expansion approach, together with Mellin pairs of the form (7.15), can be used to derive $y(t)$ from its Mellin transform.

It is interesting to note that in the 1950s, Gerardi [79] developed such an approach for the synthesis and analysis of time-varying networks governed by scale-differential and Euler-Cauchy equations, although he did not recognize the relationship to a linear scale-invariant system theory. In addition, he also derived the convolution relationship (7.8).

Before we turn our attention to a more broadly defined class of LSI systems, we remark that there is, in fact, a natural homomorphism between linear scale-invariant and linear time-invariant (LTI) systems. This relationship allows us to derive virtually all the results described in this section, in addition to many others, by mapping corresponding properties from the theory of LTI systems. Specifically, by replacing time t with exponential time e^t , we find, for example, that LSI systems become LTI systems, complex power functions become complex exponentials, the Mellin transform becomes the bilateral Laplace transform, and linear constant-coefficient scale-differential equations become familiar linear constant-coefficient differential equations.

7.2.1 Generalized Linear Scale-Invariant Systems

Suppose $y(t)$ is the response of a system $\mathcal{S}\{\cdot\}$ to an arbitrary input $x(t)$. Then we shall say the system $\mathcal{S}\{\cdot\}$ is *scale-invariant with parameter λ* whenever

$$\mathcal{S}\{x(t/\tau)\} = \tau^\lambda y(t/\tau) \tag{7.16}$$

for any constant $\tau > 0$. We will denote systems that satisfy the superposition principle (7.1) and the generalized scale-invariance relation (7.16) as $\text{LSI}(\lambda)$ systems.

Obviously, strict-sense LSI systems correspond to the special case $\lambda = 0$. It can be easily established that a necessary and sufficient condition for a linear system to be scale-invariant with parameter λ is that the kernel satisfy

$$\kappa(t, \tau) = a^{-(\lambda-1)} \kappa(at, a\tau) \quad (7.17)$$

for any $a > 0$.

Such generalized linear scale-invariant systems are also completely characterized in terms of their lagged-impulse response pair

$$\begin{aligned} \xi_+(t) &= \mathcal{S}\{\delta(t-1)\} \\ \xi_-(t) &= \mathcal{S}\{\delta(t+1)\}. \end{aligned}$$

Again, when we are able to decompose our input according to (7.6) and restrict our attention to the case of causal signals, we can exploit (7.1) and (7.16) to get the following input-output relation

$$y(t) = \int_0^\infty x(\tau) \xi\left(\frac{t}{\tau}\right) \frac{d\tau}{\tau^{1-\lambda}}. \quad (7.18)$$

Rewriting (7.18) as

$$y(t) = \int_0^\infty \{x(\tau)\tau^\lambda\} \xi\left(\frac{t}{\tau}\right) \frac{d\tau}{\tau}$$

we observe that any $\text{LSI}(\lambda)$ system can be implemented as the cascade of a system which multiplies the input by $|t|^\lambda$, followed by a strict-sense LSI system with lagged-impulse response $\xi(t)$. However, in many cases, this may not be a particularly convenient implementation, either conceptually or practically.

7.3 Linear Time- and Scale-Invariant Systems

We will say that a system is *linear time- and scale-invariant with parameter λ* , denoted $\text{LTSI}(\lambda)$, whenever it jointly satisfies the properties of superposition (7.1), time-invariance (7.2) and generalized scale-invariance (7.16). In this case, the time-

invariance constraint (7.3) requires the kernel to be of the form

$$\kappa(t, \tau) = v(t - \tau)$$

for some impulse response $v(\cdot)$, while the scale-invariance constraint (7.17) imposes that the impulse response be a generalized homogeneous function of degree $\lambda - 1$, *i.e.*,

$$v(t) = a^{-(\lambda-1)} v(at)$$

for all t and all $a > 0$. Following Gel'fand, *et al.* [67], we can parameterize the entire class of impulse responses for such systems. In particular, provided $\lambda \neq 0, -1, -2, \dots$, we get that $v(t)$ takes the form

$$v(t) = C_1|t|^{\lambda-1} u(t) + C_2|t|^{\lambda-1} u(-t). \quad (7.19a)$$

For the special case $\lambda = -n$ for $n = 0, 1, 2, \dots$,

$$v(t) = C_3|t|^{-(n-1)} u(t) + C_4|t|^{-(n-1)} u(-t) + C_5\delta^{(n)}(t) \quad (7.19b)$$

where $\delta^{(n)}(t)$ denotes the n th derivative of the unit impulse and $u(t)$ the unit step function. In both cases, the C_1, \dots, C_5 are arbitrary constants.

There are many familiar LTSI(λ) systems. For example, the identity system, for which

$$v(t) = \delta(t)$$

corresponds to $\lambda = 0$, $C_3 = C_4 = 0$ and $C_5 = 1$. In fact, as is apparent from the parameterizations (7.19), the identity system is the *only* stable LTSI(λ) system. A second example is the integrator. This system has a regular impulse response

$$v(t) = u(t)$$

and corresponds to $\lambda = 1$, $C_1 = 1$, and $C_2 = 0$. As a final example, consider a

differentiator, which has for an impulse response the unit doublet

$$v(t) = \delta'(t).$$

This choice corresponds to $\lambda = -1$, $C_3 = C_4 = 0$, and $C_5 = 1$.

Linear time- and scale-invariant systems are natural candidates for modeling and processing self-similar signals as we begin to show in the next section.

7.3.1 Self-Similar Signals and $\text{LTSI}(\lambda)$ Systems

In this section, we explore some relationships between self-similar signals and systems. In particular, we show how $\text{LTSI}(\lambda)$ systems preserve the time-invariance and scale-invariance of their inputs, and point out how these properties have been exploited in some of the models for self-similar signals described earlier in the thesis.

Our result in the deterministic case is as follows. Let $v(t)$ be the impulse response of an $\text{LTSI}(\lambda)$ system, so that $v(t)$ is homogeneous of degree $\lambda - 1$, *i.e.*, for any $a > 0$

$$v(t) = a^{-(\lambda-1)} v(at),$$

and consider driving the system with a scale-invariant input signal $x(t)$ that is homogeneous of degree H . Then it is straightforward to establish that the output $y(t)$ of the system

$$y(t) = \int_{-\infty}^{\infty} x(\tau) v(t - \tau) d\tau$$

when well-defined, is scale-invariant as well. In fact, it is homogeneous of degree $H + \lambda$, so that, for any $a > 0$

$$y(t) = a^{-(H+\lambda)} y(at). \quad (7.20)$$

There are two obvious special cases. The first corresponds to the case in which the system is the identity system ($\lambda = 0$). Here the output and input are identical, and (7.20) yields the appropriate result. The second corresponds to case in which

the input is an impulse ($H = -1$). Here, the output is $v(t)$, and, again, (7.20) yields the correct result. This, of course, suggest that at least one synthesis for a class of homogeneous signals is in terms of an $\text{LTSI}(\lambda)$ system driven by an impulse.

Note that we can derive analogous results for deterministically *time*-invariant inputs. However, this time the results are somewhat degenerate. In particular, except in trivial cases, for a time-invariant (*i.e.*, constant) input, the output of such a system is only well-defined if the system is an identity system since any other $\text{LTSI}(\lambda)$ system is unstable. Nevertheless, in this unique case, the output is, obviously, time-invariant as well.

Consider, next, the case of an input that is either wide-sense or strict-sense statistically scale-invariant as defined in Chapter 3. In this case, it is also straightforward to show that the output, when well-defined, is also statistically scale-invariant and satisfies

$$y(t) \xrightarrow{\mathcal{P}} a^{-(H+\lambda)} y(at)$$

with equality in the corresponding sense.

For wide- or strict-sense stationary (*i.e.*, statistically time-invariant) inputs, the outputs, when well-defined, are also stationary. This is, of course, a well-known result from LTI system theory. Note, however, that, again from stability considerations, the only non-trivial system for which the output is well-defined is the identity system. This implies, for instance, that, in general, when driven with stationary white noise, the outputs of such systems are not well-defined.

Many of these issues surfaced in Chapter 3, where we considered the modeling of $1/f$ processes through a synthesis filter formulation. Specifically, the system with impulse response (3.9) we first proposed as a synthesis filter for $1/f$ processes is precisely an example of an $\text{LTSI}(\lambda)$ system with $\lambda = H + 1/2$. A similar filter with $\lambda = H - 1/2$ appears in the conceptual synthesis for fractional Brownian motion illustrated in Fig. 3-2. Furthermore, the fractional integrator used in the Barnes-Allan synthesis for $1/f$ processes, which we described in Section 3-2, has properties similar to those of an $\text{LTSI}(\lambda)$ system. More generally, there would appear to be a number of potentially

important connections between the operators of fractional calculus [41] and linear jointly time- and scale-invariant system theory. Finally, the ARMA filter used by Keshner in his synthesis of $1/f$ -like behavior, which we discussed in Section 3.2.1, can be viewed as an approximation to an $\text{LTSI}(\lambda)$ system. Specifically, this filter is linear and time-invariant, but satisfies the scale-invariance relation (7.16) only for dilation factors τ of the form $\tau = \Delta^m$.

The impulse-response constitutes one important means of characterizing $\text{LTSI}(\lambda)$ systems. However, from an implementational perspective, it is not always most convenient. In the next section, we develop a canonical representation for a class of $\text{LTSI}(\lambda)$ systems in terms of wavelet bases. As we shall see, this characterization not only provides additional insight into such systems, but ultimately leads to some important techniques for realizing and approximating such systems. In fact, we will see that it is possible to interpret many of the wavelet-based representations for self-similar signals we derived in Chapters 3 and 5 in the context of these results.

7.4 Wavelet-Based $\text{LTSI}(\lambda)$ Systems

Consider a system which computes the wavelet transform of the input $x(t)$ via (2.1), multiplies the resulting field X_ν^μ by some regular field K_ν^μ in time-scale space, then inverts the result according to the synthesis formula (2.4), so that

$$y(t) = \mathcal{W}^{-1} \{ K_\nu^\mu \mathcal{W} \{ x(t) \} \}. \quad (7.21)$$

It is straightforward to establish that such a linear system has a kernel

$$\kappa(t, \tau) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \psi_\nu^\mu(t) K_\nu^\mu \psi_\nu^\mu(\tau) \mu^{-2} d\mu d\nu. \quad (7.22)$$

The structure of this kernel imposes certain constraints on the linear system; for example, such systems are symmetric, *i.e.*,

$$k(t, \tau) = k(\tau, t). \quad (7.23)$$

However, the structure of the kernel is sufficiently general that one can implement LTI, LSI, or LTSI systems using this framework.

For instance, using the readily derived identity

$$\psi_\nu^\mu(t - b) = \psi_{\nu+b}^\mu(t)$$

valid for all b , the system will be time-invariant (*i.e.*, satisfy (7.2)) whenever the multiplier field satisfies

$$K_\nu^\mu = K_{\nu-b}^\mu$$

for all b . In other words, (7.21) implements an LTI system whenever the field K_ν^μ is independent of ν . In this case, K_ν^μ can be expressed as

$$K_\nu^\mu = k(\mu) \tag{7.24}$$

for some regular function of scale $k(\cdot)$.

Likewise, using the identity

$$\psi_\nu^\mu(at) = |a|^{-1/2} \psi_{\nu/a}^{\mu/a}(t) \tag{7.25}$$

valid for a , the system will be scale-invariant with parameter λ (*i.e.*, satisfy (7.16)) whenever the multiplier field satisfies

$$K_\nu^\mu = a^{-\lambda} K_{a\nu}^{a\mu} \tag{7.26}$$

for all $a > 0$.

For the system to be jointly time- and scale-invariant with parameter λ , (7.24) and (7.26) require that

$$k(\mu) = a^{-\lambda} k(a\mu),$$

i.e., that $k(\cdot)$ be homogeneous of degree λ . The imposition of regularity on $k(\cdot)$ precludes it from containing impulses or derivatives of impulses. Again using Gel'fand's

parameterization of the homogeneous functions, we conclude that the system (7.21) is LTSI(λ) whenever the multiplier field has the form

$$K_\nu^\mu = k(\mu) = C_1|\mu|^\lambda u(\mu) + C_2|\mu|^\lambda u(-\mu) \quad (7.27)$$

for some constants C_1 and C_2 . Note that even if these constants are chosen so that $k(\cdot)$ is asymmetric, the impulse response $v(t)$ of the resulting system will be even:

$$v(t) = v(-t). \quad (7.28)$$

This is a consequence of the symmetry constraint (7.23). In fact, since we can rewrite (7.22) using (7.25) as

$$\kappa(t, \tau) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} d\nu \int_0^{\infty} \psi_\nu^\mu(t) [k(\nu) + k(-\mu)] \psi_\nu^\mu(\tau) \mu^{-2} d\mu,$$

we see that the kernel of the system is really only a function of the *even* part of the function $k(\cdot)$. Hence, without loss of generality we may set $C_2 = 0$ in (7.27) and choose

$$K_\nu^\mu = k(\mu) = C\mu^\lambda u(\mu) \quad (7.29)$$

where C is an arbitrary constant.

Finally, combining (7.28) with (7.19), we can conclude that whenever (7.21) implements an LTSI(λ) system, *i.e.*, whenever $k(\cdot)$ is chosen according to (7.29), the impulse response of (7.21) must take the form

$$v(t) = C_r|t|^{\lambda-1}$$

whenever $\lambda \neq 0, -2, -4, \dots$, or the form

$$v(t) = C_r|t|^{-(n-1)} + C_s\delta^{(n)}(t)$$

whenever $\lambda = -n$ for $n = 0, 2, 4, \dots$. In both cases, C_r and C_s are parameters

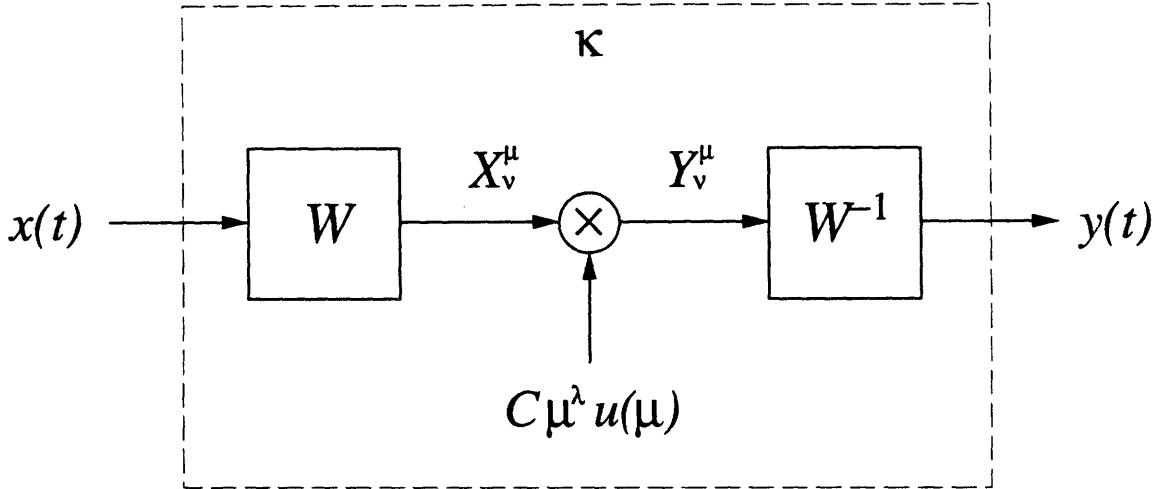


Figure 7-1: Wavelet-based implementation of an $LTSI(\lambda)$ system.

determined by the constant C in (7.29). Note, in particular, that, at least for the case $\lambda = 0$ we must have

$$C_s = C$$

$$C_r = 0.$$

This follows from the fact that, since $K_v^\mu \equiv C$, the overall system (7.21) is just a scaled identity system. Fig. 7-1 summarizes the resulting wavelet-based realization of a linear jointly time- and scale-invariant system with parameter λ . Note that this is analogous to implementing an LTI system by computing the Fourier transform of the input, multiplying by some frequency response, and applying the inverse Fourier transform to the result. As is the case for Fourier-based implementations of LTI systems, not all $LTSI(\lambda)$ systems may be realized using the wavelet-based implementation of Fig. 7-1. For example, the symmetry constraint (7.28) precludes us from being able to implement either the differentiator or integrator system examples discussed in Section 7.3 since these systems have impulse responses that are odd.

As a final remark, it is important to emphasize that the actual choice of wavelet basis plays no significant role in the representation of $LTSI(\lambda)$ systems discussed in this section. However, while the choice of basis does not enter into the theoretical de-

velopment, it is reasonable to expect it to be a factor in any practical implementation. In the next section, we consider a strategy for approximating $\text{LTSI}(\lambda)$ systems that exploits orthonormal wavelet bases. As we shall see, these quasi- $\text{LTSI}(\lambda)$ systems are particularly convenient to implement and can be made computationally efficient.

7.4.1 Dyadic Approximations to $\text{LTSI}(\lambda)$ Systems Based on Orthonormal Wavelet Bases

A practical approximation to a linear time-scale invariant system can be constructed via orthonormal wavelet bases of the type described in Section 2.2. Because signal reconstructions in terms of such bases require only a countable collection of samples of the wavelet coefficient field, the system turns out to be fundamentally more practical from an implementational perspective. In addition, using an implementation based on the DWT, the system can be made computationally highly efficient as well. In fact, in some sense, using the discrete wavelet transform to implement an $\text{LTSI}(\lambda)$ system is analogous to implementing an LTI system using the discrete Fourier transform (DFT).

Consider a system which computes the orthonormal wavelet decomposition of the input $x(t)$ according to the analysis formula (2.5b), *i.e.*,

$$x_n^m = \int_{-\infty}^{\infty} x(t) \psi_n^m(t) dt,$$

scales the resulting collection of wavelet coefficients by a factor k_n^m ,

$$y_n^m = k_n^m x_n^m$$

then re-synthesizes a signal from these modified coefficients to generate an output according to the synthesis formula (2.5a), *i.e.*,

$$y(t) = \sum_m \sum_n x_n^m \psi_n^m(t).$$

It is a straightforward exercise to show that the overall system, described via

$$y(t) = \mathcal{W}_d^{-1} \{ k_n^m \mathcal{W}_d \{ x(t) \} \}, \quad (7.30)$$

corresponds to a symmetric linear system with kernel

$$\tilde{\kappa}(t, \tau) = \sum_m \sum_n \psi_n^m(t) k_n^m \psi_n^m(\tau).$$

A close inspection reveals that, as a consequence of the nature of the discretization inherent in the system, one cannot choose the multiplier coefficients k_n^m such that the resulting system is time-invariant. Likewise, one cannot choose the coefficients so that the overall system is scale-invariant for any degree λ . However, we can show that if the k_n^m are chosen in a manner consistent with the discussion of the previous section, *viz.*,

$$k_n^m = C \mu^\lambda \Big|_{\mu=2^{-m}} = C 2^{-\lambda m}, \quad (7.31)$$

then the system defined via (7.30) obeys some associated notions of time- and scale-invariance.

We begin by noting that this system, which is depicted in Fig. 7-2, has a kernel satisfying

$$\tilde{\kappa}(t, \tau) = 2^{-(\lambda-1)m} \tilde{\kappa}(2^m t, 2^m \tau) \quad (7.32)$$

for any m , where we have used the identity

$$\psi_n^m(2^i t) = 2^{-i/2} \psi_n^{m+i}(t)$$

valid for any integer i . However, since (7.32) can be restated in terms of the generalized scale invariance condition (7.17) according to

$$\tilde{\kappa}(t, \tau) = a^{-(\lambda-1)} \tilde{\kappa}(at, a\tau), \quad a = 2^m$$

we see that the system obeys a weaker, *dyadic* scale invariance condition. In partic-

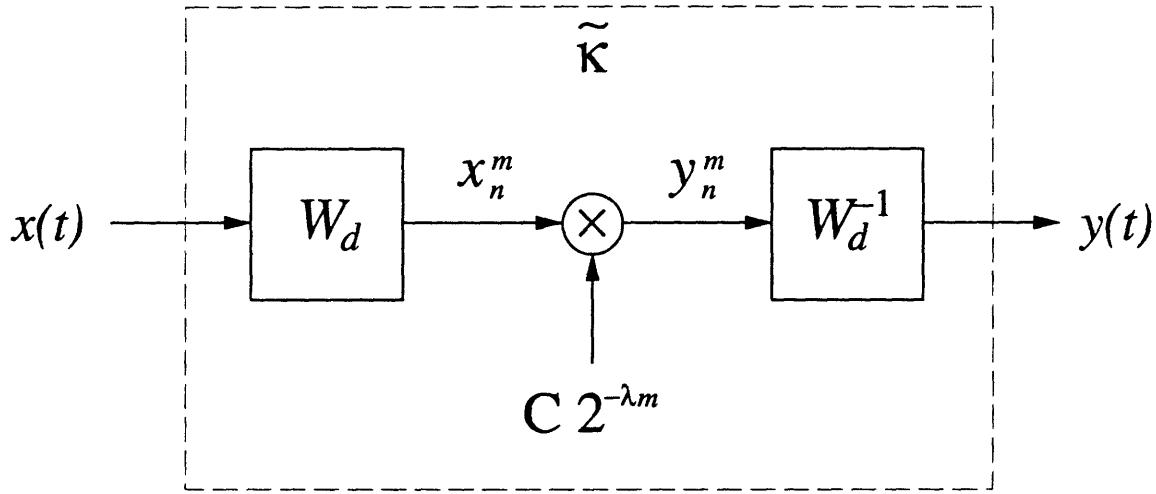


Figure 7-2: A dyadic approximation of a LTSI(λ) system as implemented via an orthonormal wavelet basis.

ular, the system satisfies (7.16) only for dilations factors τ of the form

$$\tau = 2^m$$

for integers m .

Likewise, the system obeys a somewhat weaker time-invariance property. Consider a class of input signals $x(t)$ to the system which have no detail at scales coarser than 2^M for some integer M , so that

$$x_n^m = 0, \quad m < M.$$

Then, the multiplier coefficients k_n^m for $m < M$ for the system are irrelevant and we may arbitrarily assume them to be zero. For this class of inputs then, the effective kernel is

$$\tilde{\kappa}_{\text{eff}}(t, \tau) = \sum_{m \geq M} \sum_n \psi_n^m(t) C 2^{-\lambda m} \psi_n^m(\tau).$$

Using the identity

$$\psi_n^m(t - 2^l - M) = \psi_{n+l2^{m-M}}^m(t)$$

valid whenever for $m \geq M$ and l an integer, we see that this kernel satisfies

$$\tilde{\kappa}(t, \tau) = \tilde{\kappa}(t - l2^{-M}, \tau - l2^{-M}) \quad (7.33)$$

for all integers l . Since (7.33) can be re-expressed as

$$\tilde{\kappa}(t, \tau) = \tilde{\kappa}(t - b, \tau - b), \quad b = l2^{-M}$$

we see that for this class of input signals the system is *periodically time-invariant*³ i.e., satisfies (7.2) for any shift factor τ of the form

$$\tau = l2^{-M}, \quad l = \dots, -1, 0, 1, 2, \dots$$

Note that the actual choice of wavelet affects the characteristics of the overall system in contrast to the wavelet-based systems discussed in the previous section. Indeed, with respect to scaling behavior, the choice of wavelet will affect how the system behaves under non-dyadic scale-changes at the input. Furthermore, the choice of wavelet will affect the class of inputs for which our time-invariance relation is applicable, as well as the behavior of the system under input translations that are not multiples of 2^{-M} .

In summary, this chapter has used a system theory perspective in to suggest a unifying framework for viewing various results of the thesis. For example, our results on LTSI(λ) systems, their properties, and their wavelet-based representations suggest that wavelet-based synthesis, analysis, and processing of self-similar signals is rather natural. Indeed, the $1/f$ synthesis and whitening filters described in Section 4.1 are specific examples of the quasi-LTSI(λ) systems developed in this section. As discussed, these filters play an important role in addressing problems of optimal detection and estimation involving $1/f$ processes. It is conceivable that the transmitter and receiver structures for fractal modulation discussed in Chapter 6 can also

³Actually, the multirate systems literature typically refers to such systems as periodically time-varying since they can be modeled as LTI systems whose parameters vary periodically in time.

be interpreted in terms of such quasi-LTSI(λ) systems. Such an interpretation could lead to some potentially useful additional insights into both homogeneous signals and fractal modulation.

More generally, a system theory perspective provides some novel insights into the relationships between Laplace, Fourier, Mellin and wavelet transformations, both as signal analysis tools and as representations for characterizing linear systems. In particular, we have shown results suggesting that while Laplace transforms are naturally suited to the analysis of linear time-invariant systems, and while Mellin transforms are naturally suited to the analysis of scale-invariant systems, it is the wavelet transform that plays the corresponding role for linear systems that are simultaneously time- and scale-invariant. However, this chapter represents a rather preliminary exploration of self-similar systems and many topics and issues remain to be explored. Indeed, these represent some interesting and potentially fruitful directions for further research.

Chapter 8

Conclusions and Future Directions

In this thesis we have developed robust, efficient, and convenient wavelet-based representations for two distinct families of fractal signals, and explored some potentially important applications.

For the $1/f$ family of fractal signals, we established that orthonormal wavelet expansions generally play the role of Karhunen-Loève-like expansions. That is, when $1/f$ processes are expanded in terms of orthonormal wavelet bases, the coefficients of the expansion tend to be very weakly correlated. This conclusion was supported both with theoretical evidence and with empirical evidence based upon real and simulated data. Many of the results exploited a new frequency-based characterization for $1/f$ processes introduced in the thesis.

Using this Karhunen-Loève-type representation for $1/f$ processes we were able to solve many fundamental but previously intractable problems of detection and estimation involving such processes. For example, we were able to develop maximum likelihood parameter estimation algorithms for $1/f$ signals embedded in white background noise. We emphasize that while other parameter estimation algorithms for $1/f$ processes exist, our work is the first we are aware of to explicitly take into account the presence of background noise. As we pointed out, such noise is important in ensuring that the algorithms are robust with respect to both measurement noise and modeling errors. The resulting parameter estimation algorithms, which compute both fractal dimension parameters and the accompanying signal and noise variance

parameters, can be used for signal and texture classification or for signal detection in a number of possible applications.

We were also able to develop what appear to be the first minimum mean-square error signal estimation (smoothing) algorithms for $1/f$ processes corrupted by white noise. There are many potential problems involving signal restoration and enhancement to which the resulting wavelet-based Wiener filters may be applied.

As examples of other signal processing problems that can be addressed using our new representation, we developed algorithms for coherent detection in the presence of $1/f$ and white noise, and algorithms for discriminating between distinct noise-corrupted $1/f$ signals. The detection algorithms are potentially important in a number of communication and telemetry applications, while the discrimination algorithms can be used for signal classification in a number of application contexts.

While a considerable amount of additional analysis and testing of these various algorithms remains to be done, our preliminary theoretical and empirical studies suggest that they are not only robust, but highly computationally efficient as well. Indeed, the existence of high-speed discrete wavelet transform algorithms means that wavelet representations are of both significant theoretical and practical importance.

In the second half of the thesis, we introduced and developed a novel family of homogeneous signals that were defined in terms of a dyadic scale-invariance property. We distinguished between two classes, energy-dominated and power-dominated, and established that orthonormal wavelet basis representations are as important for these signals as they are for $1/f$ signals. In particular, choosing a suitable wavelet-based inner product, we showed that the energy-dominated homogeneous signals constitute a Hilbert space for which we were able to construct wavelet-based orthonormal “self-similar” bases. We discussed the spectral and fractal properties of homogeneous signals, and developed efficient discrete-time algorithms for their synthesis and analysis.

As one potential application direction, we considered the use of homogeneous signal sets in a communications-based context. In particular, we developed a strategy for embedding information sequences into homogeneous waveforms that we termed

fractal modulation. With fractal modulation, the transmitted waveforms have the property that the information can be recovered given either arbitrarily little duration or arbitrarily little bandwidth. More generally, we were able to show that the resulting multirate modulation scheme is well-suited for use with noisy channels of simultaneously uncertain duration and bandwidth. As we discussed, this model is useful not only in describing a variety of physical channels, but also in capturing receiver constraints inherent in many point-to-point and broadcast communication scenarios. We developed both minimum mean-square error receivers for demodulating continuous-amplitude information and minimum probability-of-error receivers for demodulating discrete-amplitude information. In each case, we evaluated the performance of the overall communication scheme and made comparisons to more traditional modulation strategies.

While our development of fractal modulation considered many issues, including finite message length effects and computational complexity, many other issues, such as synchronization and buffering, remain to be adequately explored. As a consequence, fractal modulation does not yet constitute a fully-developed communication protocol. Nevertheless, based on the results of our preliminary investigation, it does represent a novel, compelling and potentially important paradigm for communication in a variety of military and commercial communication contexts.

8.1 Future Directions

In general, there are numerous outstanding issues pertaining to the topics treated in this thesis. Indeed, many of these issues have been identified within the appropriate chapters of the thesis. Still others have yet to be identified. Such issues suggest important future directions for work on the topic.

In the case of Chapter 7, which discussed connections between self-similar signals and self-similar systems, our treatment was particularly cursory. Much work remains to be done in developing a unified framework from which to view the various results of the thesis, and in understanding the nature of the relationship between the Laplace,

Fourier, Mellin, and wavelet transforms.

An example of a problem not considered within this thesis, but warranting investigation, is that of optimal prediction of $1/f$ processes. Naturally, one can envision numerous applications involving economic, geophysical and other time series where optimal forecasting algorithms could be exploited. However, given the particular correlation structure within $1/f$ processes, we might reasonably speculate that such processes are fundamentally unpredictable, *i.e.*, that accurate prediction is impossible from observations corresponding to typical data lengths and SNRs. In this case, a quantitative result on the predictability of $1/f$ processes would be an equally important result. It is certainly plausible that at least indirectly our wavelet-based representations for $1/f$ processes can be used to obtain both prediction algorithms and performance bounds for this class of signals.

The development of efficient representations for non-Gaussian fractal signal models represents another important direction for future research. In this thesis, we have restricted our attention principally to Gaussian $1/f$ processes, largely for reasons of tractability. And while Gaussian $1/f$ processes are valuable models for many real signals, considerably richer behavior can be derived from non-Gaussian $1/f$ models generated by nonlinear mechanisms. Consequently, techniques for modeling non-Gaussian $1/f$ signals as well as algorithms for processing such signals would be potentially useful in an apparently wide range of application contexts.

Finally, we remark that there would appear to be many other potential applications for the classes of homogeneous signals we have introduced in this thesis. Even within the context of communications, one can imagine a whole host of additional uses of homogeneous signals beyond the particular fractal modulation application we consider in this thesis. In many respects, identifying and exploring new and potentially promising applications represents perhaps the most exciting direction for future research suggested by this thesis.

Appendix A

Derivation of the Discrete Wavelet Transform

A.1 Analysis Algorithm

The key to developing an efficient discrete-time implementation of the wavelet decomposition lies in recognizing a useful recursion. Because

$$\phi_0^0(t), \psi_0^0(t) \in V_0 \subset V_1,$$

there exists a pair of sequences $h[n]$ and $g[n]$ such that we can express these functions in terms of a basis for V_1 , *i.e.*,

$$\phi_0^0(t) = \sum_l h[l] \phi_l^1(t) \tag{A.1a}$$

$$\psi_0^0(t) = \sum_l g[l] \phi_l^1(t) \tag{A.1b}$$

where the coefficient $h[n]$ and $g[n]$ are given by the appropriate projections, *viz.*, (2.20). Equivalently, we may express (A.1) in the frequency domain as

$$\Phi(\omega) = 2^{-1/2} H(\omega/2) \Phi(\omega/2) \tag{A.2a}$$

$$\Psi(\omega) = 2^{-1/2} G(\omega/2) \Phi(\omega/2). \tag{A.2b}$$

In any case, multiplying both sides of (A.1) by $2^{m/2}$, replacing t with $2^m t - n$, and effecting a change of variables we get, more generally,

$$\phi_n^m(t) = \sum_l h[l - 2n] \phi_l^{m+1}(t) \quad (\text{A.3a})$$

$$\psi_n^m(t) = \sum_l g[l - 2n] \phi_l^{m+1}(t) \quad (\text{A.3b})$$

where, in turn, we may rewrite (2.20) as

$$h[l - 2n] = \int_{-\infty}^{\infty} \phi_n^m(t) \phi_l^{m+1}(t) dt, \quad (\text{A.4a})$$

$$g[l - 2n] = \int_{-\infty}^{\infty} \psi_n^m(t) \phi_l^{m+1}(t) dt. \quad (\text{A.4b})$$

The discrete-time algorithm for the fine-to-coarse decomposition associated with the analysis follows readily. Specifically, substituting (A.3a) into (2.13) and (A.3b) into (2.5b), we get, for each m , the filter-downsample relations (2.21a) and (2.21b) defining the algorithm.

A.2 Synthesis Algorithm

The coarse-to-fine refinement algorithm associated with the synthesis can be derived in a complementary manner. Since

$$\phi_n^{m+1}(t) \in \{V_m \oplus O_m\},$$

we can write

$$\begin{aligned} \phi_n^{m+1}(t) &= A_m \phi_n^m(t) + D_m \phi_n^m(t) \\ &= \sum_l \{h[n - 2l] \phi_l^m(t) + g[n - 2l] \psi_l^m(t)\} \end{aligned} \quad (\text{A.5})$$

where the last equality follows by recognizing the projections in the respective expansions as (A.4). The upsample-filter-merge relation (2.21c) then follows immediately

by substituting (A.5) into

$$a_n^{m+1} = \int_{-\infty}^{\infty} x(t) \phi_n^{m+1}(t) dt.$$

Appendix B

Proofs for Chapter 3

B.1 Proof of Theorem 3.2

We first establish the following useful lemma.

Lemma B.1 *When a $1/f$ process $x(t)$ is passed through a filter with frequency response*

$$B_a(\omega) = \begin{cases} 1 & \pi a < |\omega| \leq 2\pi a \\ 0 & \text{otherwise} \end{cases} \quad (\text{B.1})$$

for any $a > 0$, the output $y_a(t)$ is wide-sense stationary, has finite variance and has an autocorrelation satisfying

$$R_{y_a}(\tau) = E[y_a(t)y_a(t-\tau)] = a^{-2H}R_{y_1}(a\tau) \quad (\text{B.2})$$

for all $a > 0$. Furthermore, for any distinct integers m and k , the processes $y_{2^m}(t)$ and $y_{2^k}(t)$ are jointly wide-sense stationary.

Proof:

First, from Definition 3.1 we have immediately that $y_1(t)$ is wide-sense stationary. More generally, consider the case $a > 0$. Let $b_a(t)$ be the impulse response of the filter with frequency response (B.1). To establish (B.2), it suffices to note that $y_a(t)$ has correlation function

$$R_{y_a}(t, s) = E[y_a(t)y_a(s)]$$

$$\begin{aligned}
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} b_a(t - \alpha) b_a(s - \beta) R_x(\alpha, \beta) d\alpha d\beta \\
&= a^{-2H} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} b_1(at - \alpha) b_1(as - \beta) R_x(\alpha, \beta) d\alpha d\beta \\
&= a^{-2H} R_{y_1}(at, as)
\end{aligned} \tag{B.3}$$

where we have exploited the identities (3.2b) and

$$b_a(t) = ab_1(at).$$

However, since $y_1(t)$ is wide-sense stationary, the right side of (B.3) is a function only of $t - s$. Hence, $y_a(t)$ is wide-sense stationary and (B.2) follows. Furthermore, $y_a(t)$ has variance

$$R_{y_a}(0, 0) = a^{-2H} R_{y_1}(0, 0) < \infty$$

where the inequality is a consequence of Definition 3.1. To establish our final result, since $B_{2^m}(\omega)$ and $B_{2^k}(\omega)$ occupy disjoint frequency intervals for $m \neq k$, the spectra of $y_{2^m}(t)$ and $y_{2^k}(t)$ likewise occupy disjoint frequency intervals. Thus, $y_{2^m}(t)$ and $y_{2^k}(t)$ are uncorrelated, and, hence, jointly wide-sense stationary as well ■

We now proceed to a proof of our main theorem.

First, let us establish that $y(t)$ is wide-sense stationary. Let M_L and M_U be any pair of integers such that

$$2^{M_L}\pi < \omega_L < \omega_U < 2^{M_U+1}\pi.$$

and consider preceding the filter (3.25) with a filter whose frequency response is

$$\tilde{B}(\omega) = \begin{cases} 1 & 2^{M_L}\pi < |\omega| \leq 2^{M_U+1}\pi \\ 0 & \text{otherwise} \end{cases} \tag{B.4}$$

since this will not affect the output $y(t)$.

Let $\tilde{y}(t)$ be the output of the filter (B.4) when driven by $x(t)$. Then since

$$\tilde{B}(\omega) = \sum_{m=M_L}^{M_U} B_{2^m}(\omega)$$

where $B_{2^m}(\omega)$ is as defined in (B.1) of Lemma B.1, we can decompose $\tilde{y}(t)$ according to

$$\tilde{y}(t) = \sum_{m=M_L}^{M_U} y_{2^m}(t) \tag{B.5}$$

where $y_{2^m}(t)$ is the response of the filter with frequency response $B_{2^m}(\omega)$ to $x(t)$.

From Lemma B.1 we have that the

$$\{y_{2^m}(t)\}, \quad M_L \leq m \leq M_U$$

are jointly wide-sense stationary. Hence, from (B.5) we see that $\tilde{y}(t)$ is wide-sense stationary. Then since $y(t)$ is obtained from $\tilde{y}(t)$ through the filter (3.25), the stationarity of $y(t)$ is an immediate consequence of the stationarity of $\tilde{y}(t)$ [65].

Let us now derive the form of the spectrum of $y(t)$, *i.e.*, (3.26). We begin by rewriting (B.2) of Lemma B.1 in the frequency domain as

$$S_{y_1}(a\omega) = a^{-(2H+1)} S_{y_a}(\omega) \quad (\text{B.6})$$

where $S_{y_a}(\omega)$ is the power spectrum associated with $y_a(t)$. For $1 < a < 2$, we observe that $S_{y_1}(\omega)$ and $S_{y_a}(\omega)$ have spectral overlap in the frequency range $a\pi < |\omega| < 2\pi$, and can therefore conclude that the two spectra must be identical in this range. The reasoning is as follows. If we pass either $y_a(t)$ or $y_1(t)$ through the band pass filter with frequency response

$$B^\dagger(\omega) = \begin{cases} 1 & a\pi < |\omega| \leq 2\pi \\ 0 & \text{otherwise} \end{cases}$$

whose impulse response is $b^\dagger(t)$, the outputs must be identical, *i.e.*,

$$b^\dagger(t) * y_a(t) = b^\dagger(t) * y_1(t) = b^\dagger(t) * x(t).$$

Since $y_a(t)$, and $y_1(t)$ are jointly wide-sense stationary, we then conclude

$$S_{y_a}(\omega) |B^\dagger(\omega)|^2 = S_{y_1}(\omega) |B^\dagger(\omega)|^2$$

whence

$$S_{y_a}(\omega) = S_{y_1}(\omega), \quad a\pi < |\omega| < 2\pi. \quad (\text{B.7})$$

Combining (B.7) with (B.6) we get

$$S_{y_1}(a\omega) = a^{-(2H+1)} S_{y_1}(\omega), \quad a\pi < |\omega| < 2\pi \quad (\text{B.8})$$

for any $1 < a < 2$. Differentiating (B.8) with respect to a and letting $a \rightarrow 1+$, we find that

$$\omega S'_{y_1}(\omega) = -(2H+1)S_{y_1}(\omega), \quad \pi < \omega < 2\pi,$$

and note that all positive, even, regular solutions to this equation are of the form

$$S_{y_1}(\omega) = \sigma_x^2 / |\omega|^\gamma, \quad \pi < |\omega| \leq 2\pi \quad (\text{B.9})$$

for some $\sigma_x^2 > 0$ and γ given by (3.7). Using (B.9) with (B.6) we find, further, that

$$S_{y_{2^m}}(\omega) = \begin{cases} \sigma_x^2 / |\omega|^\gamma & 2^m\pi < |\omega| \leq 2^{m+1}\pi \\ 0 & \text{otherwise} \end{cases}$$

Via Lemma B.1, the $y_{2^m}(t)$ are uncorrelated, so we deduce that $\tilde{y}(t)$ has spectrum

$$S_{\tilde{y}}(\omega) = \sum_{m=M_L}^{M_U} S_{y_{2^m}}(\omega) = \begin{cases} \sigma_x^2 / |\omega|^\gamma & 2^{M_L}\pi < |\omega| \leq 2^{M_U+1}\pi \\ 0 & \text{otherwise} \end{cases}$$

Finally, since

$$S_y(\omega) = |B(\omega)|^2 S_{\tilde{y}}(\omega)$$

our desired result (3.26) follows ■

B.2 Proof of Theorem 3.3

To show that a fractional Brownian motion $x(t)$, for $0 < H < 1$, is a $1/f$ process according to Definition 3.1, it suffices to consider the effect on $x(t)$ of *any* LTI filter with a regular finite-energy impulse response $b(t)$ and frequency response $B(\omega)$ satisfying $B(\omega) = 0$. In particular, since $x(t)$ has correlation given by (3.16), the output

of the filter

$$y(t) = \int_{-\infty}^{\infty} b(t - \tau) x(\tau) d\tau \quad (\text{B.10})$$

has autocorrelation

$$\begin{aligned} R_y(t, s) &= E[y(t)y(s)] \\ &= \frac{\sigma_H^2}{2} \int_{-\infty}^{\infty} b(v) dv \int_{-\infty}^{\infty} |t - s + u - v|^{2H} b(u) du \end{aligned}$$

as first shown by Flandrin [29]. Since $R_y(t, s)$ is a function only of $t - s$, the process is stationary, and has spectrum

$$S_y(\omega) = |B(\omega)|^2 \cdot \frac{1}{|\omega|^{2H+1}}.$$

When we restrict our attention to the case in which $B(\omega)$ is the ideal bandpass filter (3.24), we see that $y(t)$ is not only stationary, but has finite variance. This establishes that any fractional Brownian motion $x(t)$ satisfies the definition of a $1/f$ process.

That the generalized derivative, fractional Gaussian noise $x'(t)$, is also a $1/f$ process follows almost immediately. Indeed, when $x'(t)$ is processed by the LTI filter with impulse response $b(t)$ described above, the output is $y'(t)$, the derivative of (B.10). Since $y(t)$ is stationary, so is $y'(t)$. Moreover, $y'(t)$ has spectrum

$$S_{y'}(\omega) = |B(\omega)|^2 \cdot \frac{1}{|\omega|^{2H'+1}}.$$

where H' is as given by (3.20). Again, when $B(\omega)$ is given by (3.24), $y'(t)$ is not only stationary, but has finite variance, which is our desired result ■

B.3 Proof of Theorem 3.4

Without loss of generality, let us assume $\sigma^2 = 1$. Next, we define

$$x_M(t) = \sum_{m=-M}^M \sum_n x_n^m \psi_n^m(t) \quad (\text{B.11})$$

as a resolution-limited approximation to $x(t)$ in which information at resolutions coarser than 2^{-M} and finer than 2^M is discarded, so

$$x(t) = \lim_{M \rightarrow \infty} x_M(t) = \sum_m \sum_n x_n^m \psi_n^m(t).$$

Since for each m the wavelet coefficient sequence x_n^m is wide-sense stationary with spectrum $2^{-\gamma m}$, the approximation $x_M(t)$ is cyclostationary [65] with period 2^M , has finite variance, and has the associated time-averaged spectrum

$$S_M(\omega) = \sum_{m=-M}^M 2^{-\gamma m} |\Psi(2^{-m}\omega)|^2. \quad (\text{B.12})$$

The limiting time-averaged spectrum

$$S_x(\omega) = \lim_{M \rightarrow \infty} S_M(\omega)$$

gives the desired spectrum expression (3.36), and corresponds to the time-averaged spectrum of $x(t)$ as measured at the output of a bandpass filter for each frequency ω in the passband. The desired octave-spaced ripple relation (3.38) for arbitrary integer k follows immediately from (3.36).

To establish (3.37), we begin by noting that, given ω , we can choose m_0 and ω_0 such that $\omega = 2^{m_0}\omega_0$ and $1 \leq |\omega_0| < 2$. Hence, using (3.38) we see

$$S_x(\omega) = 2^{-m_0\gamma} S_x(\omega_0)$$

from which it follows that

$$\left[\inf_{1 \leq |\omega_0| < 2} S_x(\omega_0) \right] \frac{1}{|\omega|^\gamma} \leq S_x(\omega) \leq \left[\sup_{1 \leq |\omega_0| < 2} S_x(\omega_0) \right] \frac{2^\gamma}{|\omega|^\gamma}.$$

It suffices, therefore, to find upper and lower bounds for $S_x(\omega_0)$ on $1 \leq |\omega_0| < 2$.

Since $\psi(t)$ is R th-order regular, $\Psi(\omega)$ decays at least as fast as $1/\omega^R$ as $\omega \rightarrow \infty$.

This, together with the fact that $\Psi(\omega)$ is bounded according to (2.8a), implies that

$$|\Psi(\omega)| \leq \frac{C}{1 + |\omega|^R}.$$

for some $C \geq 1$. Using this with (2.14a) in (3.36) leads to the upper bound

$$S_x(\omega_0) \leq \sum_{m=0}^{\infty} 2^{-\gamma m} + \sum_{m=1}^{\infty} 2^{\gamma m} C^2 2^{-2Rm} < \infty.$$

To establish the lower bound it suffices to show $S_x(\omega) > 0$ for every $1 \leq \omega \leq 2$, which we establish by contradiction.

Suppose for some $1 \leq \omega_0 \leq 2$,

$$S_x(\omega_0) = \sum_m 2^{-\gamma m} |\Psi(2^{-m}\omega_0)|^2 = 0$$

Then since all the terms in the sum are non-negative, this would imply that each term is zero, from which we could conclude

$$\sum_m |\Psi(2^{-m}\omega_0)|^2 = 0.$$

However, this contradicts the wavelet basis identity (2.9). Hence, we must have that $S(\omega) > 0$ for every $\pi \leq \omega_0 \leq 2\pi$. The complete theorem follows ■

B.4 Proof of Proposition 3.5

We begin by defining the process $x_K(t)$ as the result of filtering $x(t)$ with the ideal bandpass filter whose frequency response is given by

$$B_K(\omega) = \begin{cases} 1 & 2^{-K} < |\omega| \leq 2^K \\ 0 & \text{otherwise} \end{cases}$$

so that

$$\lim_{K \rightarrow \infty} x_K(t) = x(t).$$

Then by Theorem 3.2, $x_K(t)$ is wide-sense stationary and has power spectrum

$$S_K(\omega) = \begin{cases} \sigma_x^2 / |\omega|^\gamma & 2^{-K} < |\omega| \leq 2^K \\ 0 & \text{otherwise} \end{cases}$$

If we denote its corresponding autocorrelation by

$$R_K(\tau) = E[x_K(t)x_K(t - \tau)]$$

and its wavelet coefficients by

$$x_n^m(K) = \int_{-\infty}^{\infty} x_K(t) \psi_n^m(t) dt,$$

the correlation between wavelet coefficients may be expressed as

$$\begin{aligned} E[x_n^m(K)x_{n'}^{m'}(K)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \psi_n^m(t) R_K(t - \tau) \psi_{n'}^{m'}(\tau) dt d\tau \\ &= \int_{-\infty}^{\infty} \psi_n^m(t) \cdot [R_K(t) * \psi_{n'}^{m'}(t)] dt. \end{aligned} \quad (\text{B.13})$$

Applying Parseval's theorem and exploiting (B.4), we may rewrite (B.13) in the frequency domain as

$$\begin{aligned} E[x_n^m(K)x_{n'}^{m'}(K)] &= \frac{2^{-(m+m')/2}}{2\pi} \left\{ \int_{-2^K}^{-2^{-K}} \frac{\sigma_x^2}{|\omega|^\gamma} \Psi(2^{-m}\omega) \Psi^*(2^{-m'}\omega) d\omega \right. \\ &\quad \left. + \int_{2^{-K}}^{2^K} \frac{\sigma_x^2}{|\omega|^\gamma} \Psi(2^{-m}\omega) \Psi^*(2^{-m'}\omega) d\omega \right\}. \end{aligned} \quad (\text{B.14})$$

Interchanging limits, we get

$$x_n^m = \lim_{K \rightarrow \infty} x_n^m(K)$$

and, in turn,

$$E[x_n^m x_{n'}^{m'}] = \lim_{K \rightarrow \infty} E[x_n^m(K)x_{n'}^{m'}(K)]. \quad (\text{B.15})$$

Finally, substituting (B.14) into (B.15) yields (3.40) as desired ■

B.5 Proof of Theorem 3.6

Let us define

$$\Delta = 2^{-m}n - 2^{-m'}n'$$

and

$$\Xi(\omega) = \omega^{-\gamma} \Psi(2^{-m}\omega) \Psi^*(2^{-m'}\omega)$$

for $\omega > 0$, so that (3.41) may be expressed, via (3.40), as

$$\rho_{n,n'}^{m,m'} = \frac{\sigma_x^2}{\pi\sigma^2} \operatorname{Re} I(\Delta) \quad (\text{B.16})$$

where

$$I(\Delta) = \int_0^\infty \Xi(\omega) e^{-j\Delta\omega} d\omega. \quad (\text{B.17})$$

Thus, to establish the desired result, it suffices to show that $I(\Delta)$ has the appropriate decay.

We first note that if $\gamma \geq 2R + 1$, then we cannot even guarantee that $I(\Delta)$ converges for any Δ . Indeed, since

$$\Xi(\omega) \sim \mathcal{O}(\omega^{2R-\gamma}), \quad \omega \rightarrow 0$$

we see that $I(\Delta)$ is not absolutely integrable. However, provided $\gamma \leq 2R$, $I(\Delta)$ is absolutely integrable, *i.e.*,

$$\int_0^\infty |\Xi^{(Q)}(\omega)| d\omega < \infty.$$

In this case, we have, by the Riemann-Lebesgue lemma [35], that

$$I(\Delta) \rightarrow 0, \quad \Delta \rightarrow \infty.$$

When $0 < \gamma < 2R$, we may integrate (B.17) by parts Q times, for some positive

integer ζ , to obtain

$$\begin{aligned} I(\Delta) &= \frac{1}{(j\Delta)^Q} \int_0^\infty \Xi^{(Q)}(\omega) e^{-j\Delta\omega} d\omega \\ &+ \sum_{q=0}^{Q-1} \frac{1}{(j\Delta)^q} \left\{ \lim_{\omega \rightarrow 0} [\Xi^{(q)}(\omega) e^{-j\Delta\omega}] - \lim_{\omega \rightarrow \infty} [\Xi^{(q)}(\omega) e^{-j\Delta\omega}] \right\}. \end{aligned} \quad (\text{B.18})$$

Due to the vanishing moments of the wavelet we have

$$\Xi^{(q)}(\omega) \sim \mathcal{O}(\omega^{2R-\gamma-q}), \quad \omega \rightarrow 0 \quad (\text{L}^{\infty})$$

while due to the regularity of the wavelet, $\Psi(\omega)$ decays at least as fast as $1/\omega^R$ as $\omega \rightarrow \infty$, whence

$$\Xi^{(q)}(\omega) \sim \mathcal{O}(\omega^{-2R-\gamma-q}), \quad \omega \rightarrow \infty. \quad (\text{B.20})$$

Hence, the limit terms in (B.18) for which $-2R - \gamma < q < 2R - \gamma$ all vanish.

Moreover, when we substitute $q = Q$, (B.19) and (B.20) imply that $\Xi^{(Q)}(\omega)$ is absolutely integrable, *i.e.*,

$$\int_0^\infty |\Xi^{(Q)}(\omega)| d\omega < \infty, \quad (\text{B.21})$$

whenever $-2R - \gamma + 1 < Q < 2R - \gamma + 1$, which implies, again via the Riemann-Lebesgue lemma, that the integral in (B.18) vanishes asymptotically, *i.e.*,

$$\int_0^\infty \Xi^{(Q)}(\omega) e^{-j\Delta\omega} d\omega \rightarrow 0, \quad \Delta \rightarrow \infty. \quad (\text{B.22})$$

Hence, choosing $Q = \lceil 2R - \gamma \rceil$ in (B.18) (so $2R - \gamma \leq Q < 2R - \gamma + 1$) allows us to conclude

$$I \sim \mathcal{O}(\Delta^{-\lceil 2R - \gamma \rceil}), \quad \Delta \rightarrow \infty. \quad (\text{B.23})$$

Substituting (B.23) into (B.16) then yields the desired result ■

Appendix C

The EM Parameter Estimation Algorithm

In this appendix, we derive the EM algorithm for the estimation of the signal and noise parameters $\Theta = \{\beta, \sigma^2, \sigma_w^2\}$ for the scenario described in Section 4.2.

We begin by defining our observed (incomplete) data to be

$$\mathbf{r} = \{r_n^m, m, n \in \mathcal{R}\},$$

and our complete data to be (\mathbf{x}, \mathbf{r}) where

$$\mathbf{x} = \{x_n^m, m, n \in \mathcal{R}\}.$$

Consequently, the EM algorithm for the problem is defined as [63]

E step: Compute

$$U(\Theta, \hat{\Theta}^{[l]})$$

M step:

$$\max_{\Theta} U(\Theta, \hat{\Theta}^{[l]}) \rightarrow \hat{\Theta}^{[l+1]}$$

where

$$U(\Theta, \tilde{\Theta}) \triangleq E \left[\ln p_{\mathbf{r}, \mathbf{x}}(\mathbf{r}, \mathbf{x}; \Theta) | \mathbf{r}; \tilde{\Theta} \right].$$

For our case, U is obtained conveniently via

$$U(\Theta, \tilde{\Theta}) = E \left[\ln p_{\mathbf{r}|\mathbf{x}}(\mathbf{r}|\mathbf{x}; \Theta) + \ln p_{\mathbf{x}}(\mathbf{x}; \Theta) | \mathbf{r}; \tilde{\Theta} \right]$$

with

$$p_{\mathbf{r}|\mathbf{x}}(\mathbf{r}|\mathbf{x}; \Theta) = \prod_{m,n \in \mathcal{R}} \frac{1}{\sqrt{2\pi\sigma_w^2}} \exp -\frac{(r_n^m - x_n^m)^2}{2\sigma_w^2}$$

and

$$p_{\mathbf{x}}(\mathbf{x}; \Theta) = \prod_{m,n \in \mathcal{R}} \frac{1}{\sqrt{2\pi\sigma^2\beta^{-m}}} \exp -\frac{(x_n^m)^2}{2\sigma^2\beta^{-m}}.$$

Then

$$U(\Theta, \tilde{\Theta}) = -\frac{1}{2} \sum_{m \in \mathcal{M}} N(m) \left\{ \frac{1}{\sigma_w^2} S_m^w(\tilde{\Theta}) + \ln 2\pi\sigma_w^2 + \frac{1}{\sigma^2\beta^{-m}} S_m^x(\tilde{\Theta}) + \ln 2\pi\sigma^2\beta^{-m} \right\} \quad (\text{C.1})$$

where

$$\begin{aligned} S_m^w(\Theta) &= \frac{1}{N(m)} \sum_{n \in \mathcal{N}(m)} E \left[(w_n^m)^2 | r_n^m; \Theta \right] \\ S_m^x(\Theta) &= \frac{1}{N(m)} \sum_{n \in \mathcal{N}(m)} E \left[(x_n^m)^2 | r_n^m; \Theta \right] \end{aligned}$$

are (quasi) conditional sample-variance estimates from the data based upon the model parameters Θ . Evaluating the expectations we get

$$\begin{aligned} S_m^w(\Theta) &= A_m(\Theta) + B_m^w(\Theta)\hat{\sigma}_m^2 \\ S_m^x(\Theta) &= A_m(\Theta) + B_m^x(\Theta)\hat{\sigma}_m^2 \end{aligned}$$

where

$$A_m(\Theta) = \frac{\sigma_w^2 \cdot \sigma^2\beta^{-m}}{\sigma_w^2 + \sigma^2\beta^{-m}}$$

$$\begin{aligned} B_m^w(\Theta) &= \left(\frac{\sigma_w^2}{\sigma_w^2 + \sigma^2 \beta^{-m}} \right)^2 \\ B_m^x(\Theta) &= \left(\frac{\sigma^2 \beta^{-m}}{\sigma_w^2 + \sigma^2 \beta^{-m}} \right)^2, \end{aligned}$$

which completes our derivation of the E step.

To derive the structure of the M step, we maximize $U(\Theta, \tilde{\Theta})$ as given by (C.1). This maximization is always well-defined as $U(\Theta, \tilde{\Theta}) \leq L(\Theta)$ for any $\Theta, \tilde{\Theta}$.

The local extrema are obtained by differentiating $U(\Theta, \tilde{\Theta})$ with respect to each of the parameters of Θ . Since (C.1) expresses $U(\Theta, \tilde{\Theta})$ as the sum of two terms, one of which depends only on σ_w^2 and the other of which depends only on β and σ^2 , the maximization can be broken down into two independent parts.

Considering first our maximization over σ_w^2 , we readily obtain the maximizing $\hat{\sigma}_w^2$ as the sample-average

$$\hat{\sigma}_w^2 = \frac{\sum_{m \in \mathcal{M}} N(m) S_m^w(\tilde{\Theta})}{\sum_{m \in \mathcal{M}} N(m)}.$$

Turning next to β and σ^2 , we find that the maximizing parameters $\hat{\beta}$ and $\hat{\sigma}^2$ satisfy

$$\sum_{m \in \mathcal{M}} N(m) S_m^x(\tilde{\Theta}) \beta^m = \sigma^2 \sum_{m \in \mathcal{M}} N(m) \quad (\text{C.2a})$$

$$\sum_{m \in \mathcal{M}} m N(m) S_m^x(\tilde{\Theta}) \beta^m = \sigma^2 \sum_{m \in \mathcal{M}} m N(m) \quad (\text{C.2b})$$

Eliminating σ^2 we obtain that $\hat{\beta}$ is the solution of the polynomial equation

$$\sum_{m \in \mathcal{M}} C_m N(m) S_m^x(\tilde{\Theta}) \beta^m = 0, \quad (\text{C.3})$$

where C_m is as defined in (4.16). The eliminated variable $\hat{\sigma}^2$ is trivially obtained by back-substitution:

$$\hat{\sigma}^2 = \frac{\sum_{m \in \mathcal{M}} N(m) S_m^x(\tilde{\Theta}) \hat{\beta}^m}{\sum_{m \in \mathcal{M}} N(m)}.$$

Finally, to show that the maximizing parameters are the only solution to (C.2) it suffices to show that the solution to (C.3) is unique, which we establish via the following lemma.

Lemma C.1 *Any polynomial equation of the form*

$$\sum_{m \in \mathcal{M}} C_m K_m \beta^m = 0 \quad (\text{C.4})$$

where C_m is given by (4.16) and $K_m \geq 0$ has a unique positive real solution provided $M \geq 2$ and not all K_m are zero.

Proof:

Let

$$m_* = \frac{\sum_{m \in \mathcal{M}} m N(m)}{\sum_{m \in \mathcal{M}} N(m)}$$

be a weighted average of the $m \in \mathcal{M}$, so $m_1 < m_* < m_M$. Then, from (4.16), for $m > m_*$, $C_m > 0$, while for $m < m_*$, $C_m < 0$. Hence, $C_m(m - m_*) \geq 0$ with strict inequality for at least two values of $m \in \mathcal{M}$ from our hypothesis. Now let $f(\beta)$ be the left-hand side of (C.4), and observe that

$$\tilde{f}(\beta) \stackrel{\Delta}{=} f(\beta)\beta^{-m_*}$$

is increasing for $\beta > 0$, i.e.,

$$\tilde{f}'(\beta) = \sum_{m \in \mathcal{M}} C_m (m - m_*) N(m) \hat{\sigma}_m^2 \beta^{m-m_*-1} > 0.$$

Then, since $\tilde{f}(0) = -\infty$ and $\tilde{f}(\infty) = \infty$, we see $\tilde{f}(\beta)$ has a single real root on $\beta > 0$. Since $f(\beta)$ shares the same roots on $\beta > 0$, we have the desired result. ■

This completes our derivation for the M step. The complete algorithm follows directly.

Appendix D

Proofs for Chapter 5

D.1 Proof of Theorem 5.2

We begin by noting that we may insert any ideal bandpass filter with a passband that includes $\omega_L < |\omega| < \omega_U$ before the filter specified by (5.4) without changing the overall system. We can construct such a filter as the finite parallel combination of filters with adjacent, octave width passbands. In particular, let $\tilde{x}_m(t)$ be the output, when driven by $x(t)$, of the filter with frequency response

$$B_m(\omega) = \begin{cases} 1 & 2^m\pi < |\omega| \leq 2^{m+1}\pi \\ 0 & \text{otherwise} \end{cases}, \quad (\text{D.1})$$

and corresponding impulse response $b_m(t)$. By definition of $x(t)$ we then know that $\tilde{x}_0(t)$ has finite energy. Furthermore, we get

$$\tilde{x}_m(t) = \int_{-\infty}^{\infty} x(\alpha) b_m(t - \alpha) d\alpha = 2^{-mH} \tilde{x}_0(2^m t) \quad (\text{D.2})$$

where the last inequality results from using a change of variables together with the self-similarity relation (5.2) and the identity

$$b_m(t) = 2^m b_0(2^m t).$$

Hence, every $\tilde{x}_m(t)$ has finite energy.

Next, choose finite integers M_L and M_U such that $2^{M_L}\pi < \omega_L$ and $\omega_U < 2^{M_U+1}\pi$.

Then

$$\tilde{b}(t) = \sum_{m=M_L}^{M_U} b_m(t)$$

is the impulse response of an ideal bandpass filter with passband $2^{M_L}\pi < |\omega| < 2^{M_U+1}\pi$. When this filter is driven by $x(t)$ the output is

$$z(t) = \sum_{m=M_L}^{M_U} \tilde{x}_m(t).$$

Because the $\tilde{x}_m(t)$ have Fourier transforms supported on disjoint frequency intervals, they are orthogonal. Hence, $z(t)$ has energy

$$\int_{-\infty}^{\infty} z^2(t) dt = \sum_{m=M_L}^{M_U} \int_{-\infty}^{\infty} \tilde{x}_m^2(t) dt$$

which is finite since each of the $\tilde{x}_m(t)$ has finite energy.

Finally, when we cascade the filter having impulse response $\tilde{b}(t)$ with the filter having frequency response (5.4), the overall frequency response of the cascade is, of course, (5.4) by our choice of M_L and M_U . Hence, when the cascade is driven by $x(t)$, the intermediate signal is $z(t)$ and the final output is necessarily $y(t)$ as defined in the statement of the theorem. We can then conclude that $y(t)$ has finite energy since

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} |Y(\omega)|^2 d\omega = \frac{1}{\pi} \int_{\omega_L}^{\omega_U} |Z(\omega)|^2 d\omega \leq \frac{1}{\pi} \int_0^{\infty} |Z(\omega)|^2 d\omega < \infty.$$

To show the spectrum relation (5.5), it suffices to note that

$$Y(\omega) = B(\omega) \sum_m \tilde{X}_m(\omega) = \begin{cases} X(\omega) & \omega_L < |\omega| < \omega_U \\ 0 & \text{otherwise} \end{cases}$$

where

$$X(\omega) \triangleq \sum_{m=-\infty}^{\infty} \tilde{X}_m(\omega).$$

Note that this sum is convergent because for each ω only one term in the sum is non-zero. Finally, applying the self-similarity relation (D.2) gives

$$X(\omega) = \sum_m 2^{-m(H+1)} \tilde{X}_0(2^{-m}\omega),$$

which, as one can readily verify, satisfies (5.6) ■

D.2 Proof of Theorem 5.3

We begin by noting that the energy in the output $x_0(t)$ can be expressed as

$$E \triangleq \int_{-\infty}^{\infty} x_0^2(t) dt = \frac{1}{\pi} \int_0^{\infty} |X_0(\omega)|^2 d\omega = \sum_m \frac{1}{\pi} \int_{2^m \pi}^{2^{m+1} \pi} |X(\omega)|^2 |\Psi(\omega)|^2 d\omega$$

where $X_0(\omega)$ is the Fourier transform of $x_0(t)$, and where $X(\omega)$ is as defined in Theorem 5.2. Exploiting the ripple relation (5.6), we get, through a change of variables,

$$E = \frac{1}{\pi} \int_{\pi}^{2\pi} |X(\omega)|^2 S(\omega) d\omega$$

where the non-negative spectrum $S(\omega)$ is defined as

$$S(\omega) = \sum_m 2^{-\gamma m} |\Psi(2^m \omega)|^2. \quad (\text{D.3})$$

Let us first show that $x_0(t)$ has finite energy if $x(t)$ is energy-dominated. Since,

$$E \leq \left[\sup_{\pi \leq \omega \leq 2\pi} S(\omega) \right] \frac{1}{\pi} \int_{\pi}^{2\pi} |X(\omega)|^2 d\omega$$

it suffices to show that $S(\omega) < \infty$ on $\pi \leq \omega \leq 2\pi$. The vanishing moment condition (5.17), together the fact that $\Psi(\omega)$ is bounded, implies there exists a $C > 0$ such that

$$|\Psi(\omega)| \leq \frac{C}{1 + |\omega|^{-R}}$$

for all ω . Using this with (2.8a) in (D.3) gives the desired bound

$$S(\omega) \leq \sum_{m=0}^{\infty} 2^{-\gamma m} + \sum_{m=1}^{\infty} 2^{\gamma m} C^2 2^{-2Rm} (2\pi)^{2R} < \infty$$

for $\omega < 2\pi$ and $0 < \gamma < 2R$.

Let us now show the converse, *i.e.*, that $x(t)$ is energy dominated if $x_0(t)$ has finite energy. Since

$$E \geq \left[\inf_{\pi \leq \omega \leq 2\pi} S(\omega) \right] \frac{1}{\pi} \int_{\pi}^{2\pi} |X(\omega)|^2 d\omega$$

it suffices to show $S(\omega) > 0$ for every $\pi \leq \omega \leq 2\pi$, which we establish by contradiction.

Suppose for some $\pi \leq \omega_0 \leq 2\pi$,

$$S(\omega_0) = \sum_m 2^{-\gamma m} |\Psi(2^m \omega_0)|^2 = 0.$$

Then since all the terms in the sum are non-negative, this would imply that each term is zero, from which we could conclude

$$\sum_m |\Psi(2^m \omega_0)|^2 = 0.$$

However, this contradicts the wavelet basis identity (2.9). Hence, we must have that $S(\omega) > 0$ for every $\pi \leq \omega \leq 2\pi$. The theorem follows ■

D.3 Proof of Theorem 5.6

Following the approach of the proof of Theorem 5.2 in Appendix D.1, we begin by noting that we may insert any ideal bandpass filter with a passband that includes $\omega_L < |\omega| < \omega_U$ before the filter specified by (5.4) without changing the overall system. We can construct such a filter as the finite parallel combination of filters with adjacent, octave width passbands. In particular, let $\tilde{x}_m(t)$ be the output of the filter with frequency response (D.1) and corresponding impulse response $b_m(t)$, when driven by $x(t)$. By definition of $x(t)$ we then know that $\tilde{x}_0(t)$ has finite power. Furthermore, from (D.2) we get that every $\tilde{x}_m(t)$ has finite power.

Next, again choose finite integers M_L and M_U such that $2^{M_L}\pi < \omega_L$ and $\omega_U < 2^{M_U+1}\pi$. Then

$$\tilde{b}(t) = \sum_{m=M_L}^{M_U} b_m(t)$$

is the impulse response of an ideal bandpass filter with passband $2^{M_L}\pi < |\omega| < 2^{M_U+1}\pi$. When this filter is driven by $x(t)$ the output is

$$z(t) = \sum_{m=M_L}^{M_U} \tilde{x}_m(t).$$

Because the $\tilde{x}_m(t)$ have power spectra supported on disjoint frequency intervals, their cross-correlations are zero. Hence, $z(t)$ has power

$$\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T z^2(t) dt = \sum_{m=M_L}^{M_U} \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \tilde{x}_m^2(t) dt$$

which is finite since each of the $\tilde{x}_m(t)$ has finite power.

When we cascade the filter with impulse response $\tilde{b}(t)$ with the filter whose frequency response is (5.4), the overall frequency response of the cascade is course (5.4) by our choice of M_L and M_U . Hence, when the cascade is driven by $x(t)$, the intermediate signal is $z(t)$ and the final output is necessarily $y(t)$ as defined in the statement of the theorem. We can then conclude that $y(t)$ has power

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} S_y(\omega) d\omega = \frac{1}{\pi} \int_{\omega_L}^{\omega_U} S_z(\omega) d\omega \leq \frac{1}{\pi} \int_0^{\infty} S_z(\omega) d\omega$$

which is finite because the last term on the right is the power in $z(t)$.

To show the spectrum relation (5.25), it suffices to note that

$$S_y(\omega) = |B(\omega)|^2 \sum_m S_{\tilde{x}_m}(\omega) = \begin{cases} X(\omega) & \omega_L < |\omega| < \omega_U \\ 0 & \text{otherwise} \end{cases}$$

where

$$S_x(\omega) \triangleq \sum_m S_{\tilde{x}_m}(\omega).$$

Again, since for each value of ω only one term is non-zero, the sum is convergent. Finally, applying the self-similarity relation (D.2) gives

$$S_x(\omega) = \sum_m 2^{-\gamma m} S_{\hat{x}_0}(2^{-m}\omega)$$

which, as one can readily verify, satisfies (5.26) ■

D.4 Proof of Theorem 5.7

We first establish some notation. Let us denote the cross-correlation between two finite-power signals $f(t)$ and $g(t)$ by

$$R_{f,g}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T f(t) g(t - \tau) dt.$$

Its Fourier transform is the corresponding cross-spectrum $S_{f,g}(\omega)$. Similarly

$$R_{a,b}[k] = \lim_{L \rightarrow \infty} \frac{1}{2L + 1} \sum_{n=-L}^L a[n] b[n - k]$$

will denote the cross-correlation between two finite-power sequences $a[n]$ and $b[n]$.

We begin by expressing $x(t)$ as

$$x(t) = \sum_m x_m(t)$$

where

$$x_m(t) = \beta^{-m/2} \sum_n q[n] \psi_n^m(t).$$

Then the deterministic power spectrum of $x(t)$ is given by

$$S_x(\omega) = \sum_m \sum_{m'} S_{x_m, x_{m'}}(\omega). \quad (\text{D.4})$$

We shall proceed to evaluate these various terms. Because of the dilational relation-

ships among the $x_m(t)$, *viz.*,

$$x_m(t) = 2^{m/2} \beta^{-m/2} x_0(2^m t),$$

it shall suffice to consider a single term of the form $S_{v_0, v_m}(t)$, for some $m \geq 0$.

Hence, let

$$v_m(t) = \beta^{-m/2} \sum_n q[n] \delta(t - 2^{-m}n)$$

and note that

$$v_0(t) = \sum_n \tilde{q}[n] \delta(t - 2^{-m}n)$$

where $\tilde{q}[n]$ is an upsampled version of $q[n]$, *i.e.*,

$$\tilde{q}[n] = \begin{cases} q[2^{-m}n] & n = 2^m l, \quad l = \dots, -1, 0, 1, 2, \dots \\ 0 & \text{otherwise} \end{cases}$$

Hence,

$$R_{v_0, v_m}(\tau) = \sum_k R_{\tilde{q}, q}[k] \delta(t - 2^{-m}k)$$

where

$$R_{\tilde{q}, q}[k] = \lim_{L \rightarrow \infty} \frac{1}{2L+1} \sum_{|n| \leq L, n=2^m l} q[2^{-m}n] q[n-k] = \lim_{L \rightarrow \infty} \frac{1}{2L+1} \sum_{|l| \leq 2^{-m}L} q[l] q[2^m l - k].$$

Since $q[n]$ is correlation-ergodic, we may replace this correlation with its expected value:

$$R_{\tilde{q}, q}[k] = \lim_{L \rightarrow \infty} \frac{1}{2L+1} \sum_{|l| \leq 2^{-m}L} \delta[(2^m - 1)l - k] = \begin{cases} \delta[k] & m = 0 \\ 0 & \text{otherwise} \end{cases}$$

Hence,

$$S_{v_0, v_m}(\omega) = \begin{cases} 1 & m = 0 \\ 0 & \text{otherwise} \end{cases}$$

where, without loss of generality, we have set $\sigma^2 = 1$. Then, using

$$S_{x_0, x_m}(\omega) = (2\beta)^{-m/2} \Psi(\omega) \Psi^*(2^{-m}\omega) S_{v_0, v_m}(\omega)$$

we get that

$$S_{x_0, x_m}(\omega) = \begin{cases} |\Psi(\omega)|^2 & m = 0 \\ 0 & \text{otherwise} \end{cases} \quad (\text{D.5})$$

Finally, we note that

$$S_{x_m, x_{m'}}(\omega) = S_{x_{m'}, x_m}^*(\omega)$$

and that

$$S_{x_m, x_{m'}}(\omega) = \beta^{-m'} S_{x_0, x_{m-m'}}(2^{-m'}\omega).$$

Using these identities together with (D.5) in (D.4) yields

$$S_x(\omega) = \sum_m \beta^{-m} |\Psi(2^{-m}\omega)|^2$$

as desired ■

Bibliography

- [1] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Commun. Pure Appl. Math.*, vol. 41, pp. 909–996, Nov. 1988.
- [2] A. Grossmann and J. Morlet, "Decompositions of Hardy functions into square integrable wavelets of constant shape," *SIAM J. Math.*, vol. 15, pp. 723–736, Jan. 1984.
- [3] P. Goupillaud, A. Grossmann, and J. Morlet, "Cycle-octave and related transforms in seismic signal analysis," *Geoexploration*, vol. 23, pp. 85–102, 1985.
- [4] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Trans. Inform. Theory*, vol. IT-36, pp. 961–1005, Sept. 1990.
- [5] Y. Meyer, "Ondelettes et fonctions splines," in *Sem. Equations aux Dérivées Partielles*, Ecole Polytechnique, Paris, France, Dec. 1986.
- [6] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-11, pp. 674–693, July 1989.
- [7] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. COM-31, pp. 532–540, Apr. 1983.
- [8] S. F. McCormick, ed., *Multigrid Methods*. Philadelphia, PA: SIAM, 1987.
- [9] L. R. Rabiner and R. W. Shafer, *Digital Processing of Speech Signals*. Englewood Cliffs, NJ: Prentice-Hall, 1978.
- [10] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*. Englewood Cliffs, NJ: Prentice-Hall, 1984.
- [11] W. H. R. Equitz, *Successive Refinement of Information*. PhD thesis, Stanford Univ., Stanford, CA, June 1989.
- [12] P. Vaidyanathan, "Multirate digital filters, filter banks, polyphase networks, and applications: A tutorial," *Proc. IEEE*, vol. 78, pp. 56–93, Jan. 1990.
- [13] E. P. Simoncelli and E. H. Adelson, "Subband transforms," in *Subband Image Coding* (J. W. Woods, ed.), Norwell, MA: Kluwer Academic Press, 1990.

- [14] J. A. C. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Comm. Mag.*, pp. 5–14, May 1990.
- [15] R. E. Crochiere and L. R. Rabiner, *Multirate Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1983.
- [16] G. Strang, "Wavelets and dilation equations: A brief introduction," *SIAM Rev.*, vol. 31, pp. 614–627, Dec. 1989.
- [17] M. Vetterli, "Wavelets and filter banks for discrete-time signal processing," Tech. Rep. CU/CTR/TR 218-91-48, Columbia Univ., Jan. 1991.
- [18] S. G. Mallat, *Multiresolution Representations and Wavelets*. PhD thesis, Univ. Penn., Philadelphia, PA, Aug. 1988.
- [19] D. H. Ballard and C. M. Brown, *Computer Vision*. Englewood Cliffs, NJ: Prentice-Hall, 1982.
- [20] A. J. Jerri, "The Shannon sampling theorem—its various extensions and applications: A tutorial review," *Proc. IEEE*, vol. 65, pp. 1565–1596, Nov. 1977.
- [21] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
- [22] M. Vetterli and C. Herley, "Wavelets and filter banks: Relationships and new results," in *Proc. Int. Conf. Acoust. Speech, Signal Processing*, 1990.
- [23] K. C. Chou, *A Stochastic Modeling Approach to Multiscale Signal Processing*. PhD thesis, M. I. T., Cambridge, MA, June 1991.
- [24] G. Beylkin, R. Coifman, and V. Rokhlin, "Fast wavelet transforms and numerical algorithms I," *Commun. Pure Appl. Math.*, Dec. 1989. Submitted for publication.
- [25] W. M. Lawton, "Necessary and sufficient conditions for constructing orthonormal wavelet bases," tech. rep., Aware, Inc., Cambridge, MA, Apr. 1990.
- [26] B. B. Mandelbrot and H. W. Van Ness, "Fractional Brownian motions, fractional noises and applications," *SIAM Rev.*, vol. 10, pp. 422–436, Oct. 1968.
- [27] B. Mandelbrot, "Some noises with $1/f$ spectrum: A bridge between direct current and white noise," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 289–298, Apr. 1967.
- [28] M. S. Keshner, *Renewal Process and Diffusion Models of $1/f$ Noise*. PhD thesis, M. I. T., Cambridge, MA, May 1979.
- [29] P. Flandrin, "On the spectrum of fractional Brownian motions," *IEEE Trans. Inform. Theory*, vol. IT-35, pp. 197–199, Jan. 1989.

- [30] B. B. Mandelbrot, *The Fractal Geometry of Nature*. San Francisco, CA: Freeman, 1982.
- [31] M. S. Keshner, “ $1/f$ noise,” *Proc. IEEE*, vol. 70, pp. 212–218, Mar. 1982.
- [32] A. P. Pentland, “Fractal-based description of natural scenes,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 661–674, Nov. 1984.
- [33] R. F. Voss, “ $1/f$ (flicker) noise: A brief review,” in *Proc. Ann. Symp. Freq. Contr.*, pp. 40–46, 1979.
- [34] A. van der Ziel, “Unified presentation of $1/f$ noise in electronic devices: Fundamental $1/f$ noise sources,” *Proc. IEEE*, pp. 233–258, 1988.
- [35] D. C. Champeney, *A Handbook of Fourier Theorems*. Cambridge, England: Cambridge Univ. Press, 1987.
- [36] V. M. Zolotarev, *One-dimensional Stable Distributions*. Providence, RI: Am. Math. Soc., 1986.
- [37] M. S. Taqqu, “Self-similar processes and related ultraviolet and infrared catastrophes,” in *Random Fields: Rigorous Results in Statistical Mechanics and Quantum Field Theory* (J. Fritz, J. L. Lebowitz, and D. Szasz, eds.), pp. 1057–1096, Amsterdam: North-Holland, 1981.
- [38] M. S. Taqqu, “A bibliographic guide to self-similar processes and long-range dependence,” in *Dependence in Probability and Statistics* (E. Eberlein and M. S. Taqqu, eds.), Boston, MA: Birkhauser, 1986.
- [39] M. Abramowitz and I. A. Stegun, eds., *Handbook of Mathematical Functions*. New York, NY: Dover, 1965.
- [40] J. A. Barnes and D. W. Allan, “A statistical model of flicker noise,” *Proc. IEEE*, vol. 54, pp. 176–178, Feb. 1966.
- [41] K. B. Oldham and J. Spanier, *The Fractional Calculus*. New York: Academic Press, 1974.
- [42] R. J. Barton and V. H. Poor, “Signal detection in fractional Gaussian noise,” *IEEE Trans. Inform. Theory*, vol. IT-34, pp. 943–959, Sept. 1988.
- [43] T. Lundahl, W. J. Ohley, S. M. Kay, and R. Siffert, “Fractional Brownian motion: A maximum likelihood estimator and its application to image texture,” *IEEE Trans. on Medical Imaging*, vol. MI-5, pp. 152–161, Sept. 1986.
- [44] J. Ramanathan and O. Zeitouni, “On the wavelet transform of fractional Brownian motion,” *IEEE Trans. Inform. Theory*, vol. IT-37, no. 4, pp. 1156–1158, 1991.

- [45] J. Bernamont, "Fluctuations in the resistance of thin films," *Proc. Phys. Soc.*, vol. 49, pp. 138–139, 1937.
- [46] A. van der Ziel, "On the noise spectra of semi-conductor noise and of flicker effect," *Physica*, vol. 16, no. 4, pp. 359–372, 1950.
- [47] S. Machlup, "Earthquakes, thunderstorms and other $1/f$ noises," in *Noise in Physical Systems* (P. H. E. Meijer, R. D. Mountain, and R. J. Soulen, Jr., eds.), pp. 157–160, Washington, DC: National Bureau of Standards, 1981. special publ. no. 614.
- [48] B. J. West and M. F. Shlesinger, "On the ubiquity of $1/f$ noise," *Int. J. Mod. Phys.*, vol. 3, no. 6, pp. 795–819, 1989.
- [49] E. W. Montroll and M. F. Shlesinger, "On $1/f$ noise and other distributions with long tails," *Proc. Natl. Acad. Sci.*, vol. 79, pp. 3380–3383, May 1982.
- [50] R. Saletti, "A comparison between two methods to generate $1/f^\gamma$ noise," *Proc. IEEE*, vol. 74, pp. 1595–1596, Nov. 1986.
- [51] B. Pellegrini, R. Saletti, B. Neri, and P. Terreni, " $1/f^\nu$ noise generators," in *Noise in Physical Systems and $1/f$ Noise* (A. D'Amico and P. Mazzetti, eds.), pp. 425–428, Amsterdam: North-Holland, 1986.
- [52] G. Corsini and R. Saletti, "Design of a digital $1/f^\nu$ noise simulator," in *Noise in Physical Systems and $1/f$ Noise* (C. M. V. Vliet, ed.), pp. 82–86, Singapore: World Scientific, 1987.
- [53] G. W. Wornell, "A Karhunen-Loève-like expansion for $1/f$ processes via wavelets," *IEEE Trans. Inform. Theory*, vol. IT-36, pp. 859–861, July 1990.
- [54] M. Vergassola and U. Frisch, "Scaling exponents for global and local self-similar processes." Preprint, Aug. 1990.
- [55] P. Flandrin, "Wavelet analysis and synthesis of fractional Brownian motion," *IEEE Trans. Inform. Theory*, Jan. 1991. Submitted for publication.
- [56] A. H. Tewfik and M. Kim, "Correlation structure of the discrete wavelet coefficients of fractional Brownian motions," *IEEE Trans. Inform. Theory*, Oct. 1990. Submitted for publication.
- [57] S. D. Senturia and B. D. Wedlock, *Electronic Circuits and Applications*. New York, NY: John Wiley and Sons, 1975.
- [58] M. Basseville, A. Benveniste, K. C. Chou, S. A. Golden, R. Nikoukhah, and A. S. Willsky, "Modeling and estimation of multiresolution stochastic processes," *IEEE Trans. Inform. Theory*, Feb. 1992. To appear.
- [59] A. H. Tewfik and M. Kim, "Fast multiscale statistical signal processing algorithms," *IEEE Trans. Signal Processing*, Apr. 1991. Submitted for publication.

- [60] M. Kim and A. H. Tewfik, "Fast multiscale detection in the presence of fractional brownian motions," in *Proc. SPIE Conf. Advanced Algorithms and Architectures for Signal Processing V*, 1990.
- [61] Y. Yajima, "On estimation of long-memory time-series models," *Austral. J. Statist.*, vol. 27, no. 3, pp. 321–325, 1985.
- [62] R. Fox and M. S. Taqqu, "Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series," *Ann. Stat.*, vol. 14, no. 2, pp. 517–532, 1986.
- [63] N. M. Laird, A. P. Dempster, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Ann. Roy. Stat. Soc.*, pp. 1–38, Dec. 1977.
- [64] H. L. V. Trees, *Detection, Estimation, and Modulation Theory, Part I*. New York, NY: John Wiley and Sons, 1968.
- [65] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York, NY: McGraw-Hill, second ed., 1984.
- [66] G. W. Wornell, "Communication over fractal channels," in *Proc. Int. Conf. Acoust. Speech, Signal Processing*, May 1991.
- [67] I. M. Gel'fand, G. E. Shilov, N. Y. Vilenkin, and M. I. Graev, *Generalized Functions*. New York, NY: Academic Press, 1964.
- [68] A. Papoulis, *The Fourier Integral and its Applications*. New York, NY: McGraw-Hill, 1962.
- [69] M. F. Barnsley, *Fractals Everywhere*. New York, NY: Academic Press, 1988.
- [70] F. J. Malassenet and R. M. Mersereau, "Wavelet representations and coding of self-affine signals," in *Proc. Int. Conf. Acoust. Speech, Signal Processing*, 1991.
- [71] I. Kalet, "The multitone channel," *IEEE Trans. Commun.*, vol. COM-37, pp. 119–124, Feb. 1989.
- [72] E. Feig, "Practical aspects of DFT-based frequency division multiplexing for data transmission," *IEEE Trans. Commun.*, vol. COM-38, no. 7, pp. 929–932, 1990.
- [73] B. Hirosaki, "An orthogonally multiplexed QAM system using the discrete Fourier transform," *IEEE Trans. Commun.*, vol. COM-29, no. 7, pp. 982–989, 1981.
- [74] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY: John Wiley and Sons, 1968.

- [75] G. D. Forney, Jr., R. G. Gallager, G. R. Lang, F. M. Longstaff, and S. U. Qureshi, "Efficient modulation for band-limited channels," *IEEE J. Select. Areas Commun.*, vol. SAC-2, pp. 632–647, Sept. 1984.
- [76] R. A. Dillard and G. M. Dillard, *Detectability of Spread-Spectrum Signals*. Norwood, MA: Artech House, 1989.
- [77] A. V. Oppenheim and A. S. Willsky, with I. T. Young, *Signals and Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1983.
- [78] W. M. Siebert, *Circuits, Signals and Systems*. Cambridge, MA: MIT Press, 1986.
- [79] F. R. Gerardi, "Application of Mellin and Hankel transforms to networks with time-varying parameters," *IRE Trans. Circuit Theory*, vol. CT-6, pp. 197–208, 1959.
- [80] D. Casasent, "Coherent optical pattern recognition," *Proc. IEEE*, vol. 67, pp. 813–825, May 1979.
- [81] P. G. Rooney, "A survey of Mellin multipliers," in *Fractional Calculus* (A. C. McBride and G. F. Roach, eds.), London, England: Pitman, 1989.
- [82] A. C. McBride, "A Mellin transform approach to fractional calculus on $(0, \infty)$," in *Fractional Calculus* (A. C. McBride and G. F. Roach, eds.), London, England: Pitman, 1985.
- [83] F. Oberhettinger, *Tables of Mellin Transforms*. Berlin: Springer-Verlag, 1974.
- [84] R. N. Bracewell, *The Fourier Transform and its Applications*. New York, NY: McGraw-Hill, 1965.
- [85] A. Erdélyi, ed., *Tables of Integral Transforms*, vol. I. New York, NY: McGraw-Hill, 1954.
- [86] A. Benveniste, R. Nikoukhah, and A. S. Willsky, "Multiscale system theory," in *Proc. IEEE Conf. on Dec. and Contr.*, Dec. 1990.

